Meteorological Ensemble Forecast Processor (MEFP) HEFS Release 0.1.1 User's Manual

Last updated: 06/07/2012

National Weather Service
Office of Hydrologic Development

Table of Contents

1 Overview

This section is pending transition of HEFS from beta-testing to operational use, and will be supplied at a later date. Once completed, it is anticipated that this section will provide an overview of the user's manual.

1.1 Software Components

1.2 Terminology

1.3 Notation

1.4 Manual Layout

2 Science of MEFP

2.1 Introduction

The Meteorological Ensemble Forecast Processor (MEFP) is a key component of the U.S. National Weather Service (NWS) software system for the Hydrologic Ensemble Forecast Service (HEFS). In the HEFS, the total uncertainty in streamflow forecasting is decomposed into atmospheric uncertainty and hydrologic uncertainty. The atmospheric uncertainty here refers to the uncertainty of the input forcing for the hydrologic model, whereas hydrologic uncertainty refers to the uncertainty associated with the structure, parameters, initial conditions, etc., of the hydrologic model. The MEFP component is used to quantify the atmospheric uncertainty. The MEFP extracts information from single-valued and ensemble meteorological forecasts from short, medium, and long range numerical weather prediction (NWP) models produced by a number of U.S. weather and climate forecast centers. The extracted forecast information is then used to generate forcing ensembles, which quantify the meteorological forcing uncertainty for the Ensemble Streamflow Prediction (ESP) processor, which comprises a set of hydrological models.

Currently, the MEFP can extract forecast information from three sources:

- Single-valued forecasts generated by the NWS River Forecast Centers (RFC) from Hydrometeorological Prediction Center (HPC) guidance.
- Ensemble forecasts generated by the Global Forecast System (GFS) developed at National Centers for Environmental Prediction (NCEP).
- Single-valued forecasts generated by the Climate Forecast System - version 1 (CVSv1) and version 2 (CFSv2) - of NCEP.

The MEFP can be used to produce ensemble forecasts from any feasible combination of the HPC/RFC, GFS, and CFS forecasts. In addition, it can be used to generate climatology ensemble traces with either historical observations or a sample of statistically smoothed climatology from historical observations for forecast periods up to 1-year. An important feature of the system is that it can also be used to generate hindcast ensembles for system verification and validation.

The motivations for pre-processing these singled-valued and ensemble forecasts through the MEFP include:

1. The HPC/RFC forecasts are adjusted by human forecasters and therefore may have additional skill.
2. Raw precipitation and temperature ensemble forecasts from NWP models are, in general, biased in the mean, spread, and higher moments.
3. Forcing ensembles are required to be coherent in space and time among forecast areas and forecast variables.
4. Forecast signals may be extracted and maximized through the use of multiple time scales.

The MEFP aims to produce reliable precipitation and temperature ensemble forecasts that capture the skill in the original forecasts for generating reliable and skillful hydrologic ensemble forecasts. The scientific foundation for the basic algorithms used in MEFP is described in Herr and Krzysztofowicz (2005), Schaake et al. (2007), and Wu et al. (2011).

2.2 Scientific Requirements

The creation of the MEFP has been driven by the need for uncertainty quantification in hydrologic forecasting. The MEFP has been designed with a range of forecast horizons for applications in local emergency management, flood control management, reservoir regulations, and water supply planning. The MEFP has been developed to meet the requirements of the operational HEFS, which are described in the NWS OHD report titled "Requirements for the Hydrologic Ensemble Forecast Service". The ensemble hydrologic, hydraulic, and water management system forecasts to be produced and delivered by the HEFS must:

- Span lead times from one hour to one year or more (defaulting to climatology) with seamless transitions between lead time regimes (e.g., weather to climate, short to medium to seasonal range).
- Be calibrated from a probabilistic standpoint for relevant forecast periods.
- Be spatially and temporally consistent, thus linkable (routable) across RFC domains.
- Effectively capture the information available from current operational weather to climate forecast systems by utilizing meteorological ensemble forecasts (e.g., precipitation and temperature) that are calibrated from a probabilistic standpoint for relevant forecast periods.
- Be methodologically consistent with retrospective forecast ensembles that are used for verification and training/optimization of user decision support tools.
- Be verified via a comprehensive verification system that can generate products qualifying the expected performance of the output streamflow ensembles.

There are also other important considerations specific to MEFP for precipitation and temperature ensemble forecasts:

1. The ensemble members must have the same climatology as used to calibrate hydrologic forecast models.
2. The ensemble members must be consistent over areas as large as the Mississippi river basin so that the upstream ensemble hydrograph members for different tributaries are derived from consistent precipitation and temperature ensemble members.

It may be possible to relax these requirements for some short-term forecast applications where the spatial range of influence for future events is limited. An example is flash-flood forecasting. But consistency in ensemble precipitation and temperature forecasts for RFC forecasts beyond a few days is a necessary requirement.

In calibrating the MEFP, availability of a large archive of historical observations and corresponding forecasts (or reforecasts) of a weather or climate forecast system is crucial. An increase in archive shall result in inclusion of more extreme values and reduced sampling errors in parameter estimation. A large archive is also needed for verifying the MEFP ensembles. Furthermore, in downstream applications, a large precipitation and temperature archive is required for the ESP calibration and verification, and for developing decision support systems and their performance evaluation.

2.3 Forecast Sources

The MEFP uses raw precipitation and temperature forecasts from several sources as inputs to generate ensemble forecasts of lead times from 1 day to about 9 months. The MEFP forecast horizon can be extended to about a year (limited by software constraints) with the use of precipitation and/or temperature climatology. The forecast sources currently used in the MEFP are the following.

- For short range forecasting, the HPC/RFC operational single-valued forecasts of precipitation and temperature are used. Depending on forecast locations, the precipitation forecasts are available from 1 to 5 forecast lead days, and the temperature are available up to 7 forecast lead days. The HPC/RFC forecasts are generated with adjustments by human forecasters. Some studies show that these forecasts have added skill over the guidance in some cases.
- For medium range forecasting, the mean of the GFS ensemble forecasts (the 1998 frozen version) is used by the MEFP to cover a forecast lead time of up to 14 days. The 1998 GFS model has a resolution of T62 (about 200-km grid spacing). A 15-member ensemble is produced operationally daily by the GFS from 00 UTC initial conditions, and the 1st member is from the control run. The GFS ensemble forecasts extend to15 days, with data archived at a 12-hour time step. Because these forecasts are unreliable – biased in the mean, spread, and higher moments, in the MEFP, the mean of the GFS ensembles is used to derive reliable forcing ensembles. Overall, the biased-corrected GFS precipitation forecasts have moderate skill over the CONUS.
- For long range forecasting, the CFSv1 or CFSv2 forecasts are used. It is expected that the CFSv1 will phase out and be replaced completely by CFSv2. The CFSv2 has a resolution of T126 (about 100-km grid spacing at the Equator) and a forecast horizon of about 10 months in 6-hour time steps. The CFSv2 reforecast data sets are available for 3 different forecast horizons: 9 months, 3 months, and 45 days. The MEFP uses the 9-month data set for calibration. The 9-month reforecasts cover 29 years (1982–2010). Beginning from Jan.1 each year, the 9-month reforecast runs are initiated every 5th day, 4 times per day at 0, 6, 12, and 18 UTC. This large data set of 29 years is important to the proper calibration (bias correction) of the MEFP to generate seasonal forecasts operationally. The MEFP uses the mean of lagged ensembles, constructed from a window of past and current CFSv2 reforecasts or forecasts, in estimating parameters with the reforecasts or generating forcing ensembles with the forecasts, respectively.

2.4 Methodology

The main statistical procedures implemented in the MEFP are based on the meta-Gaussian model of the bivariate probability distribution between the observed and the corresponding single-valued forecast, whereby the forecast ensembles are generated from the conditional distribution of the model given the single-valued forecast. The MEFP generates precipitation and temperature ensembles from estimated forecast probability distributions for a number of future events that span the forecast period. Subsection **2.4.1** gives an overview of the general strategy implemented in the MEFP. The other subsections describe the major concepts and techniques in more detail.

2.4.1 General Strategy

The MEFP is a complicated system involving applications of several mathematical algorithms and statistical procedures to data from multiple sources. It can be viewed loosely as a 5-step process.

- The 1st step is to calculate aggregated values for different forecast time scales useful for the application under consideration. A forecast time scale is defined here as a time period consisting of multiple basic time steps (6 hours for precipitation and 24 hours for temperature) within the forecast time horizon. Hereafter, a forecast time scale will also be referred to as a canonical event. Forecast uncertainty and skill are time-scale dependent. Even though the forecast skill at the individual basic time steps may be limited, especially for long lead times, the skill of forecasts aggregated over multiple time steps is more likely to be useful and needs to be extracted and preserved. The aggregation involves calculating precipitation accumulations and temperature averages for all predefined time scales for the single-valued forecasts and corresponding observations.
- The 2nd step is to estimate the parameters. In this step, for a given variable, basin, forecast source, calibration start time in the year, and canonical event, the joint probability distribution between the single-valued forecasts and the corresponding observations is modeled using Meta-Gaussian distribution, with model parameters generated. This step aims to build a probability model that approximates the joint empirical distribution of the forecast and observed so that the skill in the single-valued forecasts can be captured and forecast uncertainty quantified.
- The 3rd step is to sample the conditional distribution of the joint distribution model. In this step, a sample is drawn from the conditional distribution of the joint distribution model for a given single-value forecast. The sample size is chosen as the number of years of archived historical observations so that the Schaake shuffle technique can be applied at a later step.
- The 4th step is to rank the canonical events for a forecast source according to the correlation of the forecast and corresponding observed for these events.
- The last step is to apply the Schaake shuffle. The Schaake shuffle is a simple and efficient scheme used to preserve the space-time statistical properties of climatology among multiple hydro-meteorological variables for multiple forecast locations. Ensemble merging occurs naturally as ensemble traces derived from different sources of raw forecasts get generated sequentially.

Precipitation and temperature are processed slightly differently because precipitation is intermittent and its distribution is highly skewed whereas temperature is continuous and its distribution is nearly Gaussian. For temperature, after generating ensembles of daily maximum temperatures and minimum temperatures, the MEFP merges these ensembles to produce end ensembles at 6-hour time steps. This disaggregation of the daily temperature ensemble traces is done using the same interpolation procedure for estimating historical temperature time series.

2.4.2 Canonical Events

A canonical event is defined as the average value of a forecast variable over some number of future time steps. The term canonical event is motivated by the analogy to the procedure called canonical correlation analysis that can be used to reduce the number of components in a relationship between a set of predictors and a set of predictands. This idea is used here because generated ensemble members may be needed for hundreds of future 6-hour time steps. Most of the information content of the atmospheric forecasts can be represented by a set of basis functions with a much smaller number of components. The basis functions used in the MEFP are called canonical events.

A different set of canonical event files is used for precipitation and temperature. These are each defined by a set of two files: a base event file and a modulation event file. Each event is defined by a start time and an end time (as an offset from the forecast creation time). The base event file defines a set of events that are concatenated end-to-end to fill the total future period of any atmospheric forecast model that might be used as an input to the MEFP. Typically, base events have durations that tend to increase with time to the start of the event – reflecting the fact that forecast information content for short duration events decrease with forecast lead time. Modulation events define events that are aggregates of base events (i.e. they span two or more base events). Modulation events provide a way to assure that the MEFP can account for effects of time scale dependence in the variability and forecast uncertainty of future events. In an operational setting, there is little reason to modify the canonical events. The MEFP is delivered with a set of default canonical event files. But for certain research studies it will be necessary to modify these files.

2.4.3 The Meta-Gaussian Model

The bivariate meta-Gaussian distribution is used in MEFP to model the joint distribution of the forecast and the observed. The model has been used successfully in the fields of meteorology and hydrology for many years. The model has several desired properties:

1) Each of its two marginal distributions can be specified with any continuous distribution model that fits the data well.
2) Its conditional distributions have an analytical form and can be easily estimated numerically.
3) Under the standard-normal condition, the model fits the underlying joint distribution exactly. Note that since precipitation is discontinuous due to intermittency, a special algorithm is used to treat this discontinuity in the MEFP.

Can we replace the meta-Gaussian model by a simple linear regression model in the MEFP? A linear regression model may work for temperature, but not for precipitation. There are several assumptions that need to be met for a linear regression model to work well. One of them is that errors in the predictand are normally distributed; another is that the predictand has constant variance. Since precipitation amounts cannot drop below 0 and vary over multiple orders of scale, these two assumptions are unlikely to hold for precipitation for the time scales considered in the MEFP.

2.4.4 Precipitation Intermittency

The meta-Gaussian model requires its marginal distributions to be continuous. Precipitation is however intermittent and therefore not continuous. There are two distinct statistical procedures available to treat precipitation intermittency. One procedure (explicit treatment) is described in Herr and Krzysztofowicz (2005) and Wu et al. (2011), which decomposes the joint distribution of the forecast and the observed and then models the continuous-continuous (wet-wet) component by the meta-Gaussian distribution. This approach works better for time scales for which probability of precipitation is low. The other procedure (implicit treatment) is described in Schaake et al. (2007) and Wu et al. (2011). It implicitly accounts for precipitation intermittency by modeling each of the marginal distributions as a combination of continuous distributions, resulting in continuous conditional distributions for sampling ensemble members. The ensemble members less than a threshold value are set to zero. This implicit approach may work better for large time scales and wet periods for which probability of precipitation is high.

These two procedures also differ in another aspect: dependence structure modeling. For the implicit treatment the dependence structure of the transformed space is still bivariate standard normal, but only partially specified. Thus, the correlation coefficient of the dependence structure is estimated by a weighted average of the Pearson product-moment correlation coefficient of the untransformed variates (including zeros) and that of the transformed variates. The explicit treatment procedure simply uses Pearson's correlation coefficient for the dependence structure of the continuous-continuous part of the bivariate distribution. The MEFP allows the user to choose between these two procedures.

2.4.5 The Schaake Shuffle

The Schaake shuffle is a simple and efficient scheme used to preserve the space-time statistical properties of climatology among multiple hydro-meteorological variables for multiple forecast locations (Clark et al., 2004) in ensemble forecasting. Suppose that a set of historical observations from the years 1948 – 1998 (41 years) is available for calibration. In the ensemble generation step, a sample of 41 points is drawn from the forecast distribution, which is the conditional distribution from the joint distribution model. Then the sample points are arranged in such a way that the largest sample point sits in the same position in the ensemble as the largest historical member does, the same is done for the second largest sample point, and so on so forth. For an illustrative description of the scheme, see Appendix H of the EPP3 User's Manual.

2.4.6 Ensemble Merging

How does the MEFP blend ensemble forecasts from the HPC/RFC, GFS, and CFS? The MEFP uses each of these sources of forecast information to generate forecast probability distributions of future canonical events. The decision as to which forecast sources to use is controlled by the user through the MEFP input control file. Only one forecast distribution from a certain source is used for each canonical event in the final step of ensemble generation. The MEFP does not include options to merge forecast distributions from different sources into a multi-model forecast for a given event. Nor does MEFP include an option to automatically select the "best" source based on the one that correlates best with past observations. The selected forecast probability distributions are used with the Schaake shuffle to generate ensemble members.

2.5 Scientific Considerations

2.5.1 Calibration

The MEFP "calibrates" the raw forecasts from NWP models. What does this mean and why is it necessary? For example, why can't we just use the forecasts of the Multi-Model Ensemble Forecast System (MMEFS)? The MEFP assures that the climatology of all of the members produced by the MEFP is the same as the climatology of the atmospheric forcing used to calibrate the hydrologic forecast model used in the ESP for each forecast basin/segment area. This is especially important in the mountainous west, but it is very important everywhere. The MEFP also assures that the ensemble forcing is conditionally unbiased, conditioned on the forecast. Finally, the MEFP does a very good job of accounting for the predictive uncertainty in the occurrence of future events. To do all of this, the MEFP must be calibrated using historical single-valued forecasts and corresponding observations. Obviously, we can "just do the MMEFS". But the MMEFS forecasts do not have the same climatology as used to calibrate hydrologic forecast models, they are not necessarily unbiased and they do not necessarily account for the predictive uncertainty of future events. Ideally, the MMEFS forecasts could be calibrated so that they could be used as input to the MEFP to do this. But that would require MMEFS reforecasts. Reforecasts of the GEFS part of the MMEFS are being developed so that it will be possible to integrate these into the MEFP in the future.

2.5.2 HPC/RFC Archive

How much of a record of the HPC/RFC forecasts is needed? Does it matter that these forecasts are not temporally consistent? The length of record required to estimate the MEFP parameters to use the HPC/RFC forecasts depends a lot on the skill of the RFC forecasts. If the RFC forecasts were almost perfectly correlated with observations, only a few observations would be needed to define the relationship between the forecasts and observations. If the HPC/RFC forecasts have very little skill (as is often the case in summer), a lot of data are needed to define the climatology of the forecasts. As a rule of thumb, it is good to have at least 4 years of RFC forecasts and corresponding observations.

2.5.3 Why Use GFS Ensemble Mean

There are two reasons why the GFS ensemble mean is used rather than the full ensemble. Firstly, the ensemble mean contains most of the predictive information in an ensemble forecast. The ensemble mean has been shown to be more skillful than the single-value "control" forecast in some studies. Hamill and Whitaker examined the relationship between the ensemble spread and the variability of the differences between the ensemble mean and the corresponding observation for the 1998 ensemble GFS forecasts. They could not find a clear relationship. Therefore, the MEFP uses ensemble means rather than the full ensemble. There may be more forecast uncertainty information in newer ensemble forecast systems. More research is needed to understand how to make practical use of this information. Secondly, a seamless approach to weather and climate prediction is needed to meet user needs for hydrological ensemble forecasts over a wide range of forecast lead times. Meeting this need involves using atmospheric forecast information from several sources as input to the MEFP/ESP hydrologic ensemble forecast process. It is not clear how individual forecast members from different atmospheric models could be used to meet this need.

2.5.4 Strategies to Reduce Sampling Uncertainty

The MEFP estimates parameters for every day of the year. Large data samples (much larger than the number of years of historical data) are needed to reduce the sampling uncertainty in the estimation of model parameters. If parameters for each day of the year were estimated using only the historical forecast and observation pairs for that given day of the year, the sample size would be very small and the resulting parameter values would be very noisy as a result of large sampling error. Therefore, some degree of data pooling for days before and after the day for which parameters are being estimated is needed. Because precipitation is much more variable than temperature, larger samples are needed to estimate precipitation parameters than temperature parameters. There is a basic uncertainty principle associated with estimation of all the MEFP parameters. This uncertainty principal involves a trade-off between how much can be known about the "real" climatological value of a parameter in any given day of the year and how much can be known about the intra-annual variability of the parameter during the year. The intra-annual variability varies gradually from day to day. This trade-off is especially important in parts of the country where there is a strong average seasonal variation in precipitation amount. It is also important everywhere for estimation of temperature MEFP parameters for the HPC/RFC forecasts because of the limited length of RFC forecast archives. This trade-off between uncertainty in daily parameter values and uncertainty in the annual variability of parameter values is an important issue for the MEFP application because:

- When the correlation between forecasts and observations is weak, the forecast distribution tends toward the climatological probability distribution of the data used to estimate model parameters for the given forecast day of the year.
- As the sampling uncertainty of model parameters increases, the day to day variability in the MEFP estimates of the climatological distributions implied by these parameters also increases. This means that when the correlation between forecasts and observations is weak, large day to day variations in forecast distributions can occur if parameter values are highly uncertain. This may cause very large day to day variation in probability estimates of the occurrence of extreme events even when there is little day to day variability in the raw forecasts.
- As the width of the data window increases the uncertainty in the MEFP parameters decreases. But the data window may begin to include days with very different climatological distributions (if the window gets wide enough). This means that the MEFP estimated climatological distributions may be very different than the true climatological distributions. This is especially important in parts of the West where precipitation is highly seasonal, most of the precipitation occurs in as little two months and rapid transition between times of very little precipitation and a lot of precipitation also occurs over periods as small as two months.
- When the correlation between forecasts and observations is very strong, the forecast distributions have little spread and the amount of data required to characterize the joint relationship between forecasts and observations is much less than for weak forecasts. The control file includes options to allow the user to specify how the MEFP defines the data window on any given day of the year. The MEFP centers the data window on the current day and chooses a width in days that can be controlled by parameters that can be specified in the control file. The MEFP has default values for these control file options, and it is recommended that the user allow the MEFP programs to use these default values unless the user fully understands how changing these values affects the results.

- Control file options that affect the width of the data window are slightly different for precipitation and temperature parameter estimation. The MEFP parameter estimation programs produce parameter values for every day of the year by linearly interpolating parameter values between days when parameters are estimated from data available in a data window that spans many days on either side of the target day. The MEFP does not need to make estimates for every day because the data windows assure that parameters tend to vary smoothly during the year. Therefore, the control file includes a parameter (par_interval: number of days between days for which parameters are estimated, defaulting to 5) that can be used to reduce the computer time required to estimate parameter values. Many successful experimental runs have been made with par_interval set to 15.
- During very dry periods, the length of the data window may reach maxwin (default is 61) and there may not be enough data to estimate all of the parameters. In that case, if the total number of observations equals or exceeds minobs (default is 90), the program computes a value of the precipitation threshold and estimates values of the probability of precipitation for both observations and forecasts. Otherwise, the MEFP estimates parameter values for the current day to be the same as for the last day when there were sufficient data for parameter estimation. If the total number of observations or forecasts is less than minposobs (default is 30) or minposfcst (default is 30), parameter values controlling the wet part of the distribution (including the correlation parameter) are set equal to the values for the last day when parameters were successfully estimated. This approach to deal with very limited data seems to work well for locations in California where it hasn't rained in the last 12 years on any given day of the year.

2.5.5 No Skill Forecasts

What if the raw forecasts have no skill? Does the MEFP make the forecasts worse? If the forecasts have no skill, the MEFP will produce forecasts based on climatology. So, the MEFP will not make forecasts consistently worse. It is always possible, in an anecdotal sense, that a particular single-valued forecast is closer to an observation than, say, the MEFP ensemble mean. But that can only happen by chance and cannot be predicted prior to observing the event. This is just an example of the adage that all the careful planning in the world can't beat a little dumb luck.

2.5.6 The Use of Historical MAP and MAT

The MEFP uses the climatology of the historical MAP and MAT data in three ways:
1) It uses the MAP and MAT data in its parameter estimation.
2) It ascribes the space-time patterns of the MAP and MAT data to generated ensembles.
3) It uses the MAP and MAT data to generate climatological ensemble forcings for ESP.

2.5.7 Evaluation

The MEFP was tested extensively for the HPC/RFC- and GFS-based forecasts as it was developed over the last 10 years. It was used experimentally in CNRFC operations during this entire period. At CNRFC it is now being used experimentally at about 100 locations. It appears to be working well. The MEFP has been implemented at about 400 locations for CBRFC and was used to make predictions for the 2011 April – July runoff volume. Results so far look very promising. The MEFP was applied to data for a selection of 24 unregulated river basins distributed throughout the U.S. to illustrate the potential forecast skill of the GFS and CFSv1 forecasts for different forecast lead times and different times of year at each location. The results of this study are presented in sections 7 and 8 of the EPP3 Science Documentation. The results of the MEFP validation are largely unpublished. Published results can be found in the following papers: For ensembles generated from HPC/RFC forecasts, verifications results are given in Schaake et al. (2007) for precipitation and temperature, Demargne et al. (2007) for precipitation ensembles and corresponding streamflow ensembles, and Wu et al. (2010) for precipitation ensembles. For GFS-based MEFP ensemble forecasts, verification results are described in Schaake et al. (2007) and Demargne et al. (2010). Evaluation of the MEFP should include not only verification studies of the MEFP ensemble output, but also studies in conjunction with the ESP hindcasting and evaluation. The capacity to produce these hindcasts easily is just becoming available through the development of CHPS. Like all forecast systems, the MEFP has limitations. These limitations need to be studied so that improvements can be made on the system. The current version of the MEFP is a good starting point for NWS to produce ensemble streamflow predictions to meet a wide range of user needs.

2.6 Summary

The MEFP is a key component of the HEFS. The MEFP can be used to extract information from single-valued, as well as ensemble precipitation and temperature forecasts, to produce calibrated forcing ensembles for the ESP. For the short forecast range, the system can process HPC/RFC operational single-valued forecasts to produce forcing ensembles for lead times up to 5 days for precipitation and 7 days for temperature. For the medium forecast range, it processes GFS ensemble forecasts to produce forcing ensembles for lead times up to 14 days. For the long forecast range, it processes CFSv1 or CFSv2 forecasts to produce forcing ensembles for lead times up to 9 months. The MEFP can also be used to process any feasible combination of HPC/RFC, GFS, and CFS forecasts. In addition, it can be used to generate climatology ensemble traces with either historical observations or a sample of statistically smoothed climatology from historical observations, for forecast periods up to 1-year.

The MEFP aims to produce reliable forcing ensembles that preserve the skill in the original forecasts. The system applies the Schaake Shuffle technique to preserve space-time coherence in the forcing ensembles. The system also provides a way to extract and maximize forecast information through the use of multiple time scales.

References

Clark, M., Gangopadhyay, S., Hay, L., Rajagopalan, B., Wilby, R., 2004. The Schaake Shuffle: A method for reconstructing space-time variability in forecasted precipitation and temperature fields. Journal of Hydrometeorology 5 (1), 243-262.

Demargne, J., L. Wu, D.-J. Seo, and J.C. Schaake, 2007. Experimental hydrometeorological and hydrologic ensemble forecasts and their verification in the U.S. National Weather Service. *IAHS Pub. Series (Red Books)*, 313, 177-187.

Demargne, J., Brown, J. D., Liu, Y., Seo, D.-J., Wu, L., Toth, Z., and Zhu, Y., 2010: Diagnostic verification of hydrometeorological and hydrologic ensembles. *Atmos. Sci. Let*., 11(2), 114-122.

Herr, H.D., and R. Krzysztofowicz, 2005: Generic probability distribution of rainfall in space: the bivariate model. *J. Hydrol*., 306, 234–263.

Schaake, J., J. Demargne, R. Hartman, M. Mullusky, E. Welles, L. Wu, H. Herr, X. Fan, and D.-J. Seo, 2007: Precipitation and temperature ensemble forecasts from single-value forecasts. *Hydrol. Earth Syst. Sci. Discuss*., 4, 655-717.

Wu, L., D.-J. Seo, J. Demargne, J. Brown, S. Cong, S., and J. Schaake, J., 2011: Generation of ensemble precipitation forecast from single-valued quantitative precipitation forecast via meta-Gaussian distribution models. *J. Hydrol*., 399(3-4), 281-298.

3 MEFPPE Reference Manual

3.1 Overview

The Meteorological Ensemble Forecast Processor (MEFP) Parameter Estimator (PE) computes parameters used by the MEFP to generate ensembles operationally. Those parameters are zipped and stored in FEWS module data set files, being exported via exportDataSetActivity entries in the MEFP module configuration files. The MEFPPE guides the user through a step-by-step estimation process that includes setup, acquiring archive and reforecast data files, estimating parameters, and accepting/zipping those parameters. It runs as a FEWS explorer plug-in, being seamlessly integrated within the CHPS/FEWS interface, and provides diagnostic capabilities to enable the user to more easily control the quality of forecast source data and the estimated parameters.

This section of the manual describes how to use the MEFPPE software interface to accomplish parameter estimation and provides details about all interface components. It is recommended that users read Section 3.2, Getting Started, prior to using the software, and refer to the other sections as needed while using the software. This manual is available via the MEFPPE help functionality.

3.1.1 Terminology

The following important terms are used throughout this section:

- *active estimation data type*: The current data type for which estimation will be performed; either precipitation or temperature. It is controlled by a choice box in the **Location Summary Panel**.
- *CHPS locationId*: The locationId used in the CHPS configuration files to specify a location.
- *CHPS parameterId*: The parameterId used in the CHPS configuration files to specify a data type. Common parameterIds are as follows:
    - MAP/FMAT: observed/forecast 6h accumulated precipitation
    - MAT/FMAT: observed/forecast 6h instantaneous temperature
    - TFMX/TMAX: observed 24h maximum temperature
    - TFMN/TMIN: observed 24h minimum temperature
- *forecast data source*: A source of forecasts for which the MEFPPE is to compute parameters and which the MEFP uses as input to generate ensembles operationally. Current forecast sources include HPC/RFC QPF and QTF, NCEP GFS, and NCEP CFSv2 forecasts.
- *MEFP location*: A location for which the MEFP is to be executed and parameters are to be estimated. An MEFP location is defined by a CHPS locationId and parameterId and will sometimes be referred to by its *identifier* within this manual, which is "<locationId> (<parameterId>)". For example "NFDC1HUF (MAP)".
- *parameter zip group*: A group of locations for which the MEFP will be executed at one time (via one model adapter execution) and, therefore, for which the estimated parameters must be zipped together.

3.1.2 Notation

The following notation is used:

- Important terms are displayed in *italics* the first time they are used and defined.
- Graphics user interface components are displayed in **Bold**.
- List items, such as available plug-ins or allowed parameter settings, will be in "quotes".
- Column names in tables will be in 'single quotes'.
- Parameter names are displayed as normal text.
- Text which is to be entered at a command line or into an ASCII text file (including XML files) is denoted in `this font`.

3.2 Getting Started

The MEFPPE is used to estimate parameters for the MEFP to generate ensembles of precipitation and temperature time series. It guides the users through a step-by-step procedure outlined in Section 3.2.1, providing tools to allow for quality controlling data and analyzing the parameters.

This section provides basic background material pertinent to the understanding of the MEFPPE software in order to get started using the software. It explains:

1) How to run MEFPPE.
2) The parameter estimation procedure through which the MEFPPE guides the users and how that procedure connects to the interface components.
3) General concepts that are core to understanding and using the MEFPPE.

3.2.1 Input Data Requirements

To use the MEFPPE, historical precipitation (MAP) and temperature (MAT) time series must be made available to it. See Section 3.5.1 and 3.5.4 for how to make that data available. Furthermore, archived forecasts or reforecasts must be available for each forecast source for which parameters are to be estimated. The forecast sources and required data are as follows:

- RFC Forecast Source: HPC/RFC QPF/QTF forecasts are required along with corresponding observations. See Section 3.7.3 for information on how to provide that data to the MEFPPE: either via the vfypairs table of the archive database or via importing files constructed outside of the MEFPPE.
- GFS Forecast Source: GFS reforecasts are required. Those reforecasts are automatically acquired as needed by MEFPPE via SFTP. See Section 3.8.3.1. No preparation by the user is required.
- CFSv2 Forecast Source: CFSv2 reforecasts are required. Those reforecasts are automatically acquired as needed by the MEFPPE via SFTP. See Section 3.9.3.1. No preparation by the user is required.
- Climatology (Historical) Forecast Source: The aforementioned historical data is used to estimate parameters for this forecast source.

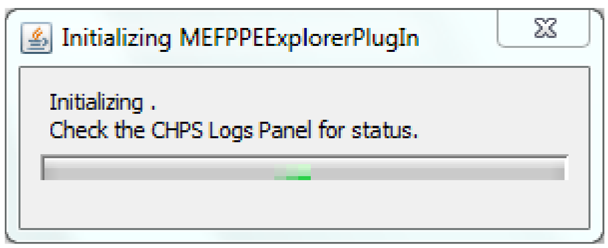
All other information is acquired from CHPS via PI-XML files that are exported, including location ids, parameter ids, and coordinates.

3.2.2 Running MEFPPE

To use MEFPPE, you must install it in a CHPS stand-alone as described in the *HEFS Release Install Notes* and then start the CHPS session. After starting CHPS, the main toolbar will include an **MEFPPE Button**:

Click on that button to run MEFPPE.

Upon starting, MEFPPE reads historical XML/fastInfoset files in order to identify for which locations parameter estimation will be performed. This may take a while if there are many such files available. A progress dialog will be visible while the information is being read and will disappear upon completion:



Log messages generated while initializing the interface will be displayed in the standard CHPS **Logs Panel**.

3.2.3 The Parameter Estimation Procedure

The MEFP parameter estimation step procedure is provided below. With each step, the sections describing how to use components of the MEFPPE to perform the step are referred to.

1. *Setup*
   Acquire historical MAP/MAT data from the CHPS database and create historical data files for the MEFPPE to use. The time series in those files specify the locations for which the MEFP will be executed. The time series acquired are the same MAP and MAT time series that are used to drive the standard ESP forecasts at an RFC. The historical data files can be constructed manually; see the section on MEFPPE installation in the *HEFS Release Install Notes*.

   See: Section 3.5.4 for how to setup for parameter estimation using the **Setup Subpanel** of the **Estimation Steps Panel**.

2. *Process historical data and generate binary files*
   Create faster-access binary files containing historical data, to be stored with the estimated parameters in the module data set .zip file for access during operational ensemble generation. Also, some data processing is performed for the historical temperature data, computing the

historical 24h minimum and maximum temperature values. For precipitation, the historical data is used without change.

See: Section 3.6.3.1 for how to create the binary files using the **Historical Data Subpanel** of the **Estimation Steps Panel**.

3. *Acquire/Create HPC/RFC QPF and QTF archive forecast data files*
   Either create RFC archived data files containing past QPF/QTF and corresponding observations based on data in the vfypairs table of the archive database, or copy the RFC archived data files created by the user. Archives of past QPF/QTF along with corresponding observed values are necessary to estimate the MEFP parameters for the RFC forecast data source.

   See: Section 3.7.3 for how to create or import RFC archived data files using the **RFC Forecast Subpanel** of the **Estimation Steps Panel**, including creating and using QPF/QTF forecast-observed pairs to construct archived data files and importing files constructed outside of MEFPPE if the archive database cannot be used.

4. *Acquire GFS reforecast data files*
   Acquire the reforecast data files for the GFS forecast source. Reforecasts (forecasts for past dates) are necessary in order to estimate the MEFP parameters.

   See: Section 3.8.3.1 for how to acquire GFS reforecast files via SFTP using the **GFS Subpanel** of the **Estimation Steps Panel**.

5. *Acquire CFSv2 reforecast data files*
   Acquire the reforecast data files for the CFS version 2 forecast source. Reforecasts (forecasts for past dates) are necessary in order to estimate MEFP parameters.

   See: Section 3.9.3.1 for how to acquire CFSv2 reforecast files via SFTP using the **CFSv2 Subpanel** of the **Estimation Steps Panel**.

6. *Estimate parameters*
   Specify estimation options and estimate the parameters of the MEFP for whichever forecast sources will be used to generate ensembles operationally. Examine the quality of the estimated parameters to determine their acceptability.

   See: Section 3.10.4.1 for how estimate parameters and Section 3.10.4.2 for how to view diagnostics related to estimated parameters, both using the **Estimation Subpanel** of the **Estimation Steps Panel**.

7. *Accept (zip) parameter files*
   Create zip files of parameters to be exported during operational ensemble generation. Since the MEFP is typically executed for a group of locations at one time, and only one exportDataSetActivity can be specified in a module configuration file, all parameters for a group must be zipped together.

See: Section 3.11.5.1 for how to define parameter zip groups and Section 3.11.5.2 for how to construct module data set zip files using the **Acceptance Subpanel** of the **Estimation Steps Panel**.

3.2.4 Core Concepts

This section discusses several concepts that are core the operations of the MEFPPE.

3.2.4.1 The MEFPPE Run Area

The MEFPPE runs using files stored on the file system within the CHPS region directory:

Models/hefs/mefppeRunArea

Files stored under that directory include run-time information files, historical data files, archive or reforecast data files, and parameter files. The user should never modify anything within the MEFPE run area unless specifically instructed to do so (as in Section 3.2.4.3).

3.2.4.2 Run-time Information

The MEFPPE run-time information includes any information necessary for the MEFPPE to execute and that needs to be remembered whenever the MEFPPE is closed so that the user can pick-up where he or she left off upon restarting the MEFPPE. That run-time information includes the following:

- The MEFP location information, including mapped location ids and coordinates
- Defined canonical events
- Defined parameter zip groups

All other information, including step status, is determined at run-time based on the contents of the MEFPPE run area.

The run-time information is stored in a file underneath of the system files directory within the MEFPPE run area:

.systemFiles/runTimeInformation.xml

Do not modify this file unless you are told to do so by an OHD developer while debugging an issue. The file is updated once per minute while the MEFPPE is running and whenever the MEFPPE is closed.

3.2.4.3 Importing From XEFS EPP3

The MEFPPE is capable of importing index (location and station mapping) information and canonical events from its predecessor, the eXperimental Ensemble Forecast System (XEFS) Ensemble Pre-Processor 3 (EPP3) software. To do so, before starting the MEFPPE, place the files to import under the import directory within the MEFPPE run area.

Models/hefs/mefppeRunArea/import

The following files beneath the import directory will be loaded (description of file contents is provided in parentheses):

control/…
- base_events_precip_v2.txt (base canonical events for precipitation)
- base_events_temp_cfsv2.txt (base canonical events for temperature)
- modulation_events_precip_v2.txt (modulation canonical events for precipitation)
- modulation_events_temp_cfsv2.txt (modulation canonical events for temperature)

index/…
- map_ts_index.txt (station mapping index information for precipitation locations)
- map_area_loc.txt (location coordinate information for precipitation locations)
- mat_ts_index.txt (station mapping index information for temperature locations)
- stationlocs.txt (location coordinate information for temperature locations)

For older version of EPP3, the names of the files may not be identical to those listed above. Rename the files to be imported as needed. Furthermore, the XEFS EPP3 software used lower case letters to denote EPP3 locations (e.g., “nfdc1huf”), whereas CHPS usually uses all upper case letters (e.g., “NFDC1HUF”). Edit the index files prior to importing so that all EPP3 location ids are upper case to match the CHPS location ids.

3.2.4.4 FEWS PI-service Connection

The MEFPPE acquires historical MAP and MAT time series via the FEWS PI-service, and, in order to use the FEWS PI-service, the connection port number must be identified. After the CHPS interface has started, check the **Logs Panel** for lines similar to the following:

```
11-04-2010 11:16:01 INFO - Started FewsPiServiceImpl on localHost : 8101
11-04-2010 11:16:01 WARN - Failed to start: SocketListener0@0.0.0.0:8100
```

The line that begins with “Started FewsPiServiceImpl…” indicates the port number (as highlighted above) of the FEWS PI-service session initialized for the currently running session of CHPS. This is the PI-service to which the MEFPPE should connect. If the port number is not 8100 (the default) or is not the value which was setup during installation, then the MEFPPE must be directed to the correct port number. See Section 3.5.1.1 for details on how to change the port number in MEFPPE.

3.2.4.5 Canonical Events

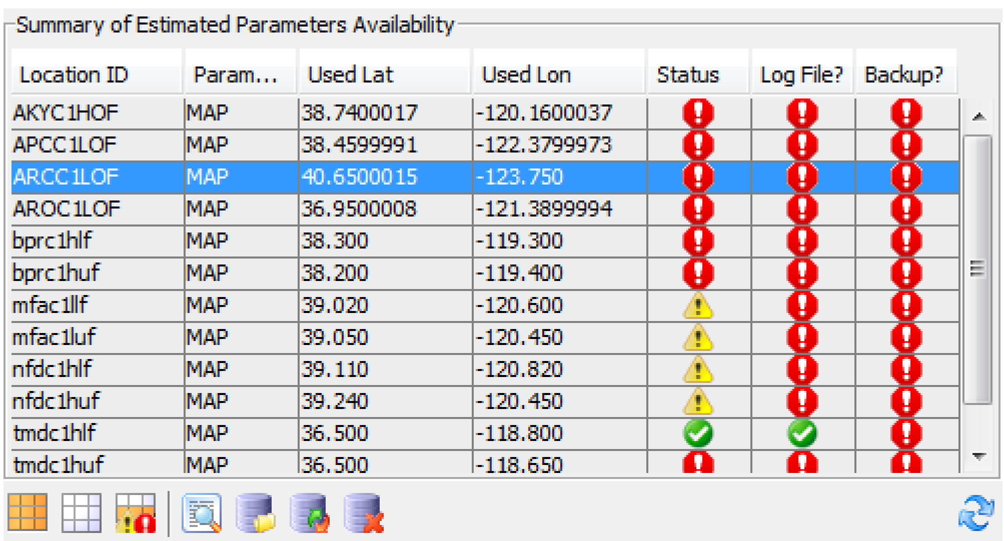
A detailed definition of canonical events is provided in Section 2. To summarize, a canonical event defines an aggregation period for forecast time series. It is defined by a start period (positive integer) and end period (positive integer). For precipitation data, the periods are defined in units of 6 hours, while for temperature it is 24 hours. For example, a canonical event for periods 1 – 2 for precipitation covers the first 12 hours of the forecast time series, while periods 120 – 240 covers the second thirty-day period, generally referred to as month 2. For every day of the year for which they are computed, parameters are estimated for each canonical event.

3.2.5 General Graphical User Interface Components

Some graphical user interface components are used many times within the MEFPPE and are described below.

3.2.5.1 Generic Summary Table

Various panels within the MEFPPE make use of a **Generic Summary Table**, which provides information about the MEFP locations and the status of steps performed. For example:


Summary of Estimated Parameters Availability

Location ID	Param...	Used Lat	Used Lon	Status	Log File?	Backup?
AKYC1HOF	MAP	38.7400017	-120.1600037			
APCC1LOF	MAP	38.4599991	-122.3799973			
ARCC1LOF	MAP	40.6500015	-123.750			
AROC1LOF	MAP	36.9500008	-121.3899994			
bprc1hlf	MAP	38.300	-119.300			
bprc1huf	MAP	38.200	-119.400			
mfac1llf	MAP	39.020	-120.600			
mfac1luf	MAP	39.050	-120.450			
nfdc1hlf	MAP	39.110	-120.820			
nfdc1huf	MAP	39.240	-120.450			
tmdc1hlf	MAP	36.500	-118.800			
tmdc1huf	MAP	36.500	-118.650			

Underneath the table is a toolbar containing buttons that are panel specific; the example above is for the **Estimation Subpanel** (Section 3.4). Four buttons, however, are common to all **Generic Summary Tables**:

- **Select All Button**: Selects all rows of the table.
- **Unselect All Button**: Unselects all rows, clearing the table selection.
- **Select Rows That Need Processing Button**: Select all rows for which the status in the primary status column is not a check mark: or . These are the rows indicating locations for which the associated step needs to be performed or updated.
- **Refresh Button**: Refresh the table, determining the status of the rows from scratch. Clicking this button is usually not necessary, but may be required if the user changes files in the MEFPPE run area.

When this table is used within panel, it will be referred to as a **Generic Summary Table** associated with a specific step described in Section 3.2.1 and its panel specific buttons will be described.

3.2.5.2 Table Delete/Add and Status Columns

Many tables used within the MEFPPE include a leading column that allows for deleting or adding rows, or status columns indicating the status of steps performed. Those columns display icons as follows:

Delete/Add Column:

- **Delete Icon**: Click to delete the row from the associated table. Sometimes this will cause a dialog popup confirming the delete.
- **Add Icon**: Click to add a row to the associated table.

Status Column:

- **Bad Status Icon**: Indicates that as step has not been performed or an error of some kind occurred while performing some other action.
- **Warning Status Icon**: Indicates that a step has been performed but needs to be updated (performed again).
- / **Good Status Icon**: Indicates that a step has been performed or some other action was successful. The icon is usually used to indicate success, but in some cases a is used; the difference between the two icons are explained in later sections when needed.

For all status icons, a tool tip will display further information, such as the cause of failures or why a step needs to be updated. To see the message, leave the mouse cursor over the icon without moving it for a few seconds. If a table within the MEFPPE uses either a delete/add or status column, it will be stated in the description of that table. All **Generic Summary Tables** use a status column.

3.2.5.3 Archive/Reforecast Time Series Diagnostic Display Panel

For forecast sources that use archived or reforecast time series to estimate parameters (RFC, GFS, and CFSv2), a standard panel is used to display the archive/reforecast time series. Shown in Figure 1, the panel is designed to display one year of time series at a time and has the following components:

- **Time Series Chart**: The chart occupies the left part of the panel and displays the RFC archived forecasts and the GFS/CFSv2 reforecasts, depending on the subpanel of the **Estimation Steps Panel** that was used to create the diagnostic. For RFC forecasts, which provide their own observation time series, an observed time series is also display in dark gray.
- **Show All Checkbox**: Click to see all time series in a single plot. It may take a long time to render the chart, so checking this box is generally not recommended.
- **T0 List**: A selectable list of the forecast times, or T0s, for which reforecasts or archived forecasts are available. When one or more items in the list are selected, the time series for those forecast times are "emphasized"; they are highlighted in red on the chart and

displayed in front of the other time series. For temperature data, the region between corresponding 24-hour minimum and maximum values is shaded.

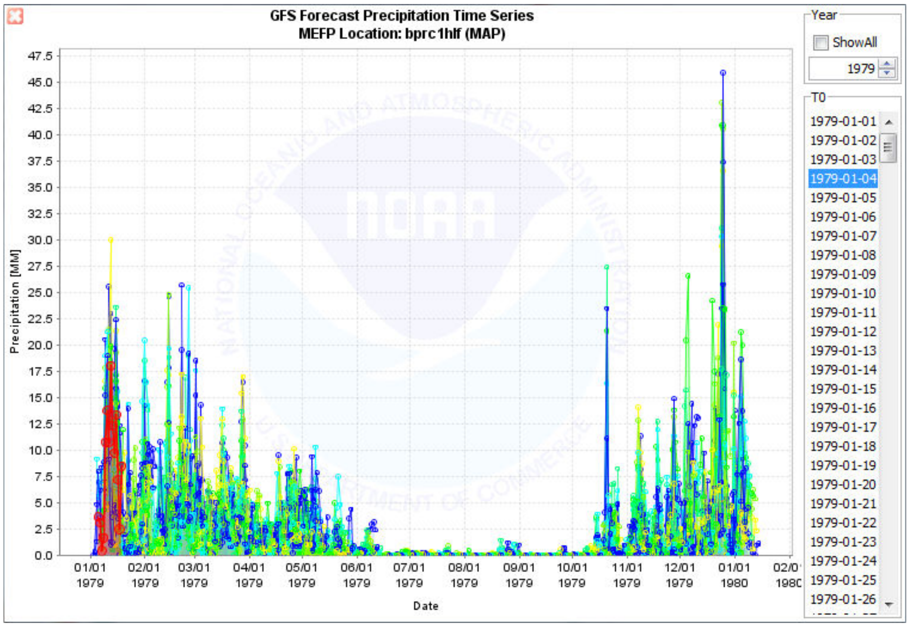


Figure 1: Example diagnostic display panel displaying GFS reforecast time series.

3.2.6 Format of this Reference Manual

The Sections 3.4 – 3.13 are provided as a reference for the MEFPPE component panels of the MEFPPE interface. Each section provides the following information:

- A description of the component panel to which the section applies.
- Any special considerations required for the panel.
- A listing of the interface components, including buttons, tables, lists, etc.
- Instructions for how to perform basic tasks using the components.

3.3 MEFPPE Main Panel

Shown in Figure 2-1, the MEFPPE Main Panel is displayed as a plug-in to CHPS after initialization is completed. It includes three components:

- **Estimation Steps Panel**: Guides the user through the steps outlined in Section 3.2.3. A tabbed panel is provided for each of the steps.
- **Location Summary Panel**: Summarizes the status of the steps for each of the MEFP locations. Also provides for the ability to run all steps for selected locations and select the active estimation data type (precipitation or temperature).
- **Diagnostics Display Panel**: Displays diagnostics that assist the user in quality controlling the data, deciding on options to use for estimation, and quality controlling and accepting the estimated parameters.

*NOTE: The **Select Type of Data for Estimation Choice Box** displayed in the **Location Summary Panel** controls the active estimation data type for the entire interface. This selection affects most of the panels in the **Estimation Steps Panel**.*

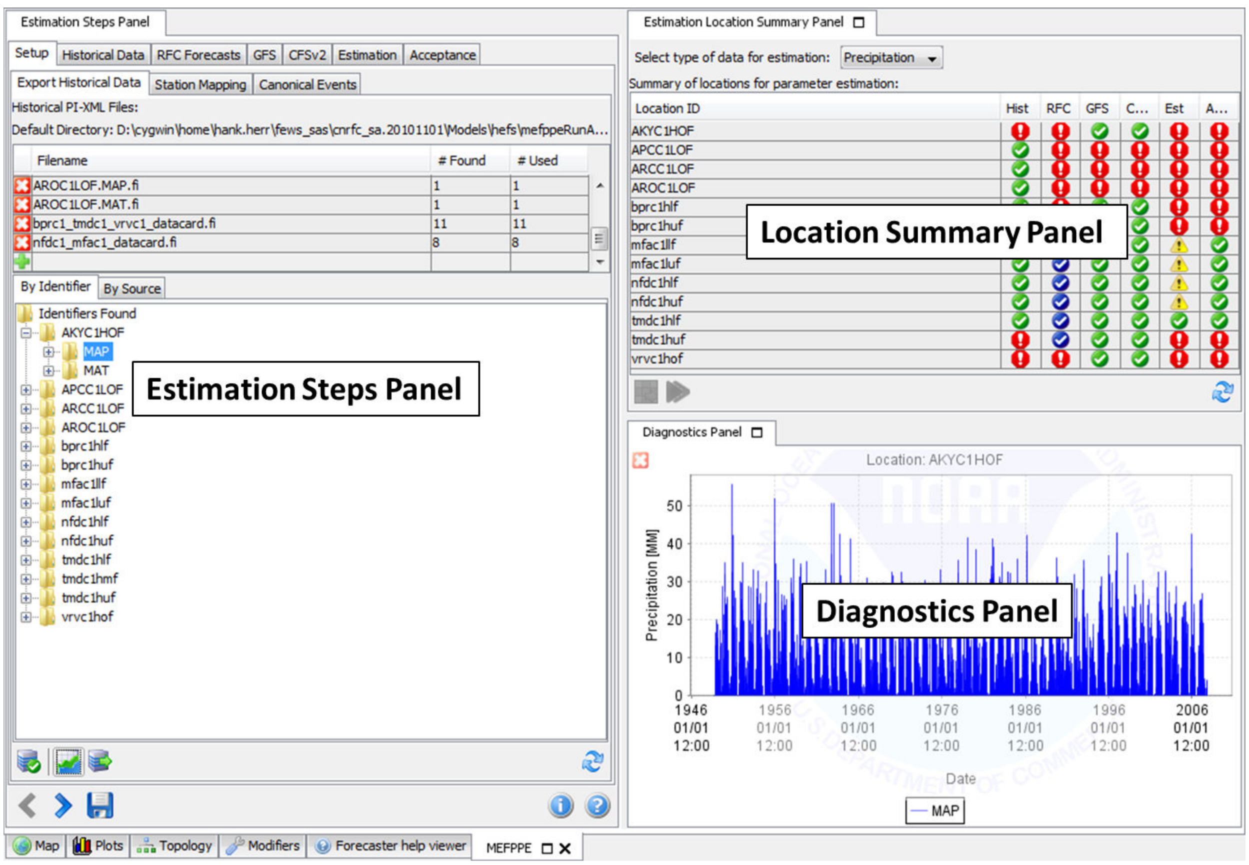
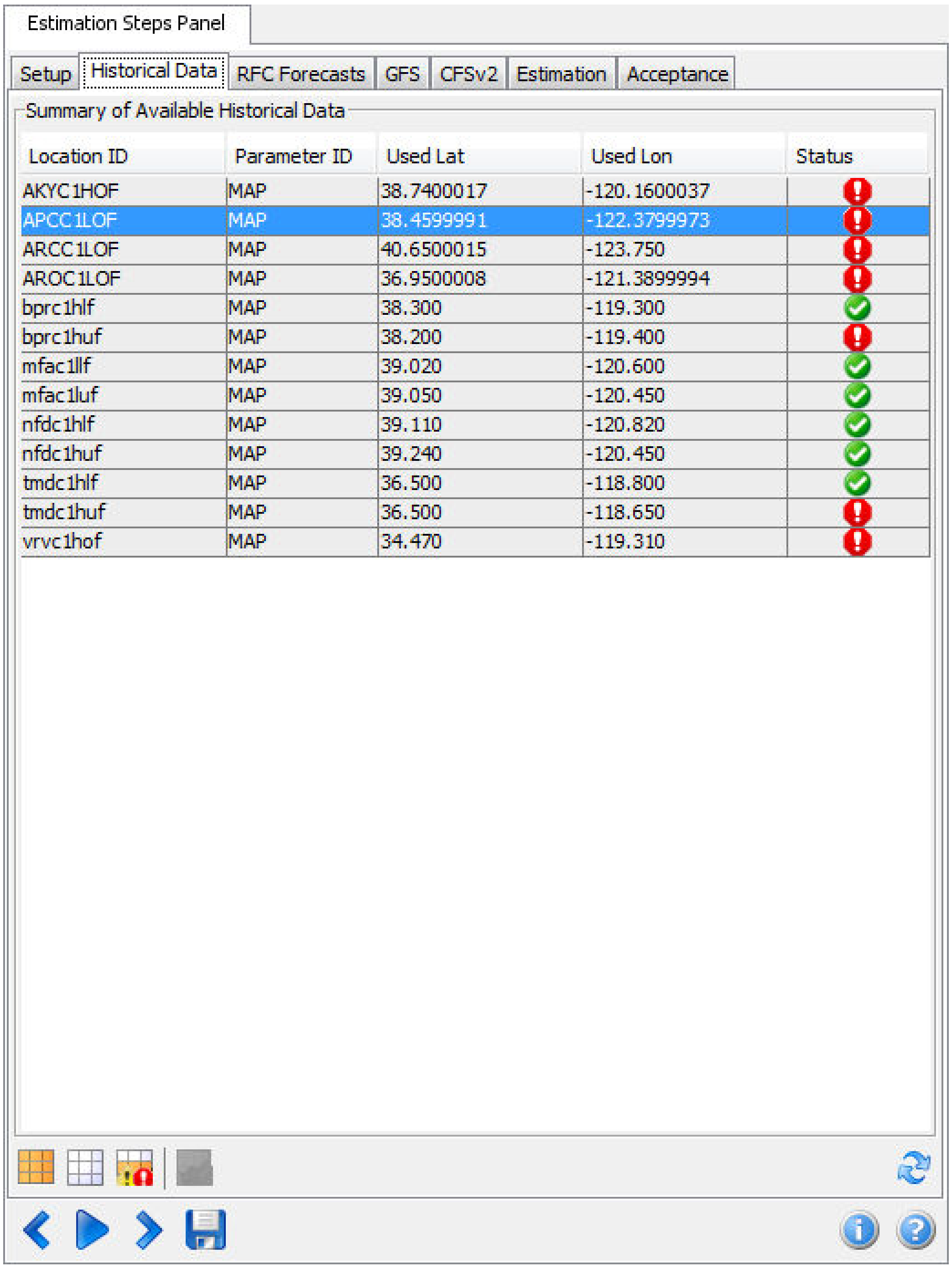


Figure 2: The **MEFPPE Main Panel**, displayed upon start-up of the MEFPPE.

3.4 Estimation Steps Panel

The **Estimation Steps Panel**, shown in Figure 3, is positioned on the left-hand side of the **MEFPPE Main Panel** and displays tabbed subpanels that correspond to the steps of the MEFP parameter estimation process. All of the tabbed subpanels are described in sections that follow. Also provided are buttons that facilitate navigating the tabbed subpanels, an information button, and a help button.



Location ID	Parameter ID	Used Lat	Used Lon	Status
AKYC1HOF	MAP	38.7400017	-120.1600037	
APCC1LOF	MAP	38.4599991	-122.3799973	
ARCC1LOF	MAP	40.6500015	-123.750	
AROC1LOF	MAP	36.9500008	-121.3899994	
bprc1hlf	MAP	38.300	-119.300	
bprc1huf	MAP	38.200	-119.400	
mfac1llf	MAP	39.020	-120.600	
mfac1luf	MAP	39.050	-120.450	
nfdc1hlf	MAP	39.110	-120.820	
nfdc1huf	MAP	39.240	-120.450	
tmdc1hlf	MAP	36.500	-118.800	
tmdc1huf	MAP	36.500	-118.650	
vrvc1hof	MAP	34.470	-119.310	

Figure 3: The **Estimation Steps Panel**.

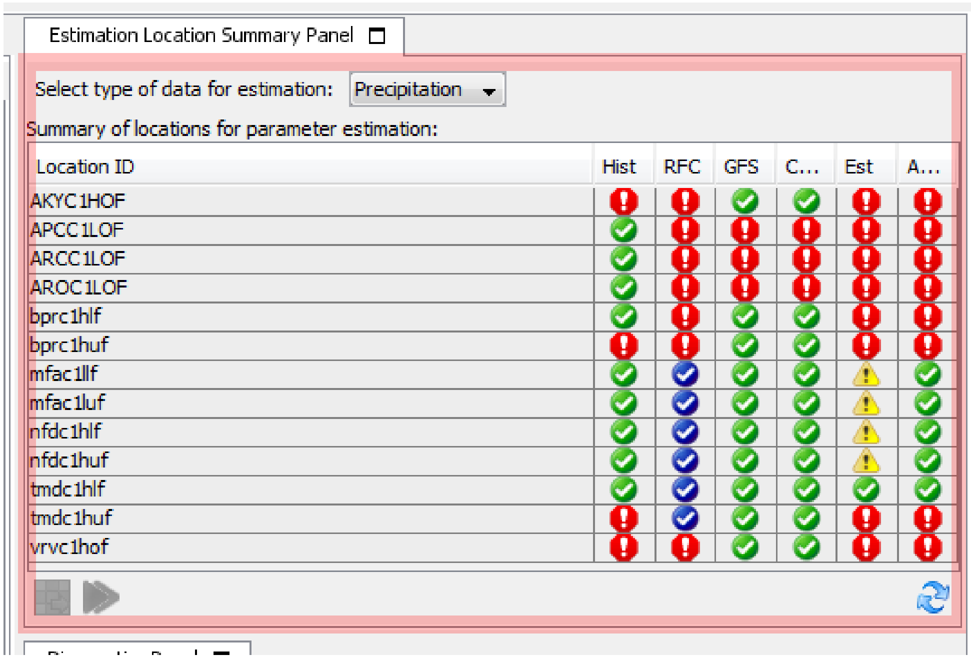
3.4.1 Components

The following describes the **Estimation Steps Panel** components:

- **Estimation Steps Tabbed Subpanels**: One tabbed subpanel is displayed for each of the estimation steps discussed in Section 3.2.3:

  To move between step subpanels, either the tabs can be clicked, or the **Back** and **Next Buttons**, described below, can be used. A subpanel is said to be 'active' if its tab is selected and its contents are current viewed. For example, in the image above, the **Historical Data Subpanel** is active.
-  **Back/Next Buttons**: Click to navigate to the previous or the next step tabbed subpanel. The buttons are disabled if there is no previous or next subpanel.
-  **Perform Step Button** (**Run Button**): Click to run the step corresponding to the active tabbed subpanel. If there is no step to perform, as for the **Setup Tabbed Panel**, then this button will not be present. The button is enabled only if one or more MEFP locations for which to perform the step are selected in the tabbed subpanel (see the description for the individual steps subpanels provided in following sections). A description of how to perform a step is presented below in Section 3.4.2.1. This button is not available in the **Setup Subpanel**.
-  **Save Run-Time Information Button**: Click to force an immediate save of the run-time information. The MEFPPE saves run-time information to a file that is loaded whenever it starts, enabling it to remember user settings. The file is saved once per minute while MEFPPE is running, when MEFPPE is closed, and when this button is clicked.
-  **About Button**: Click to display a dialog providing version information for the MEFPPE.
-  **Help Button**: Click to active help mode. When in help mode, the interface cannot be interacted with. Rather, the user can click on a component of the interface to receive help information tailored for the clicked component. The information is extracted directly from this manual and is displayed in the system's default internet browser. The component for which help will be provided is highlighted by a faded red box; for example:

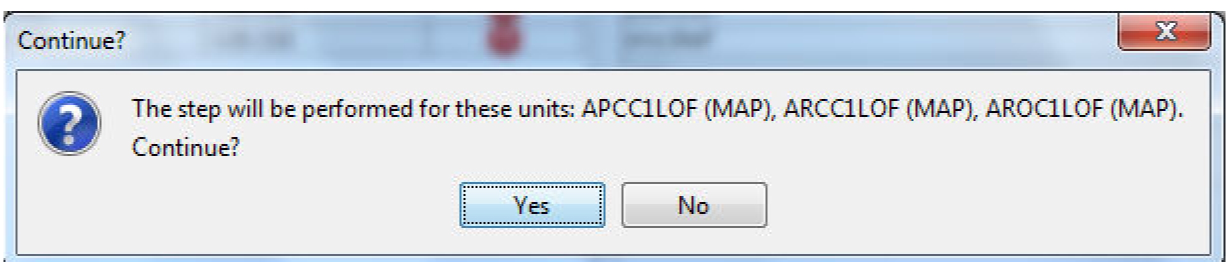


.

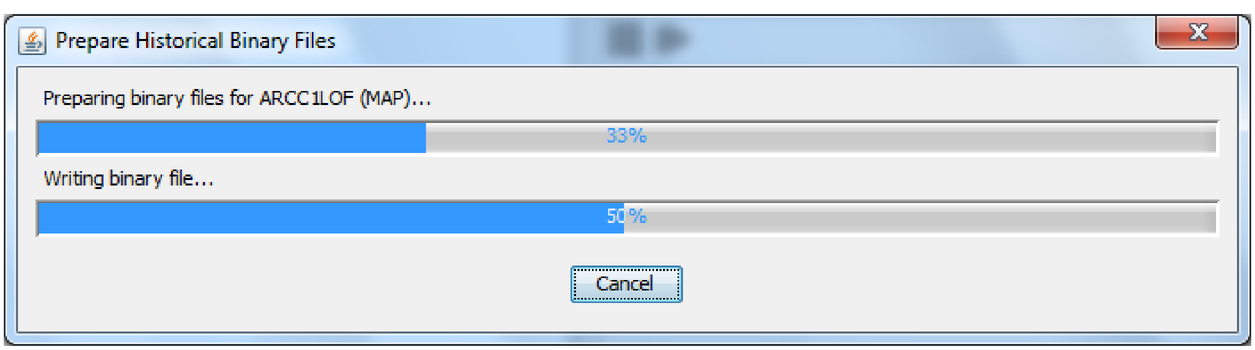
3.4.2 Usage

3.4.2.1 Performing a Parameter Estimation Step

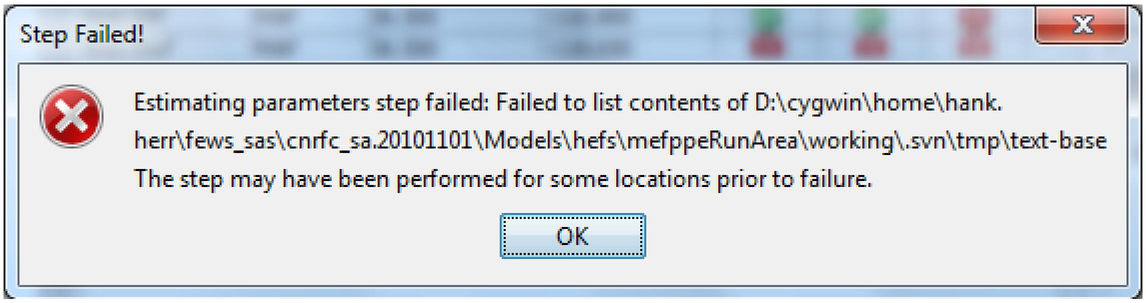
A step is performed by making the corresponding step subpanel active, selecting the MEFP locations for which to perform the step, and clicking on the **Perform Step Button**. Upon clicking, a **Continue Dialog** will be displayed allowing the user to confirm or cancel the run; for example:



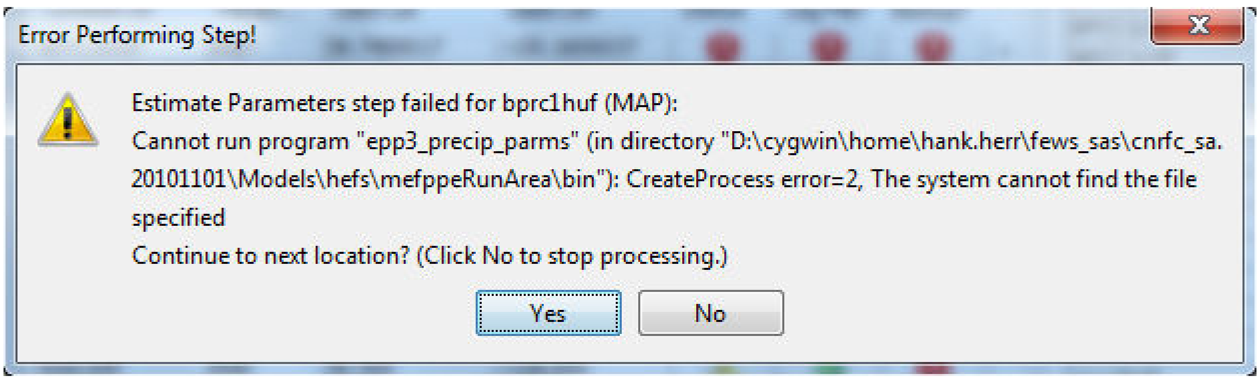
Click on **Yes** to continue or **No** to cancel. If **Yes** is clicked, a **Step Progress Dialog** will be displayed providing the ability to cancel the step via a **Cancel Button**:



If the step fails for any reason, including if it was canceled, an error dialog will be displayed. If the step is only being performed for one MEFP location or it is the last of multiple locations for which the step failed, then a **Step Failed Dialog** will be displayed explaining the cause of the failure:



Otherwise, an **Error Performing Step Dialog** will be displayed, giving the user the option to continue to the next selected MEFP location (click **Yes** to continue, **No** to stop):



If the step is successful, the progress dialog will close with no additional dialog displayed.

If a step is canceled by clicking on the **Cancel Button** in the **Step Progress Dialog**, the MEFPPE may not immediately cancel the step. Rather, it will wait until the step can be canceled cleanly without causing any problems. Upon clicking **Cancel**, the button will be disabled until the step can be canceled.

3.5 Setup Subpanel

The **Setup Subpanel** of the **Estimation Steps Panel**, shown in Figure 4, allows the user to setup the MEFP locations for which parameters are to be estimated. There are three items to setup for the MEFPPE:

- Export required historical data: The MEFPPE executes from its run area based on files therein. In order to compute parameters for a location, historical data must be present for that location within a FEWS pi-timeseries compliant XML or fastInfoset file in the MEFPPE run area. Those files can be created manually by the user (see Section 3.5.4.2) or by acquiring time series via the FEWS PI-service and creating files from those time series. Exporting and verifying the available historical data is done during the setup phase.
- Setup station mapping: In some cases, it may be necessary to use data for one location as a proxy for data for another location. For example, if temperature forecasts are derived from MDL MOS forecasts based on locationIds that do not match CHPS locationIds, a station mapping may be used. The station mapping is defined during the setup phase.
- Setup canonical events: The MEFPPE estimates parameters for canonical events, described in Section 3.2.4.5. Those canonical events are defined during the setup phase.

Subpanels within the **Setup Subpanel** are defined for each of these three setup steps and are described below. The usage section explains how to perform each of these steps using the interface components provided.

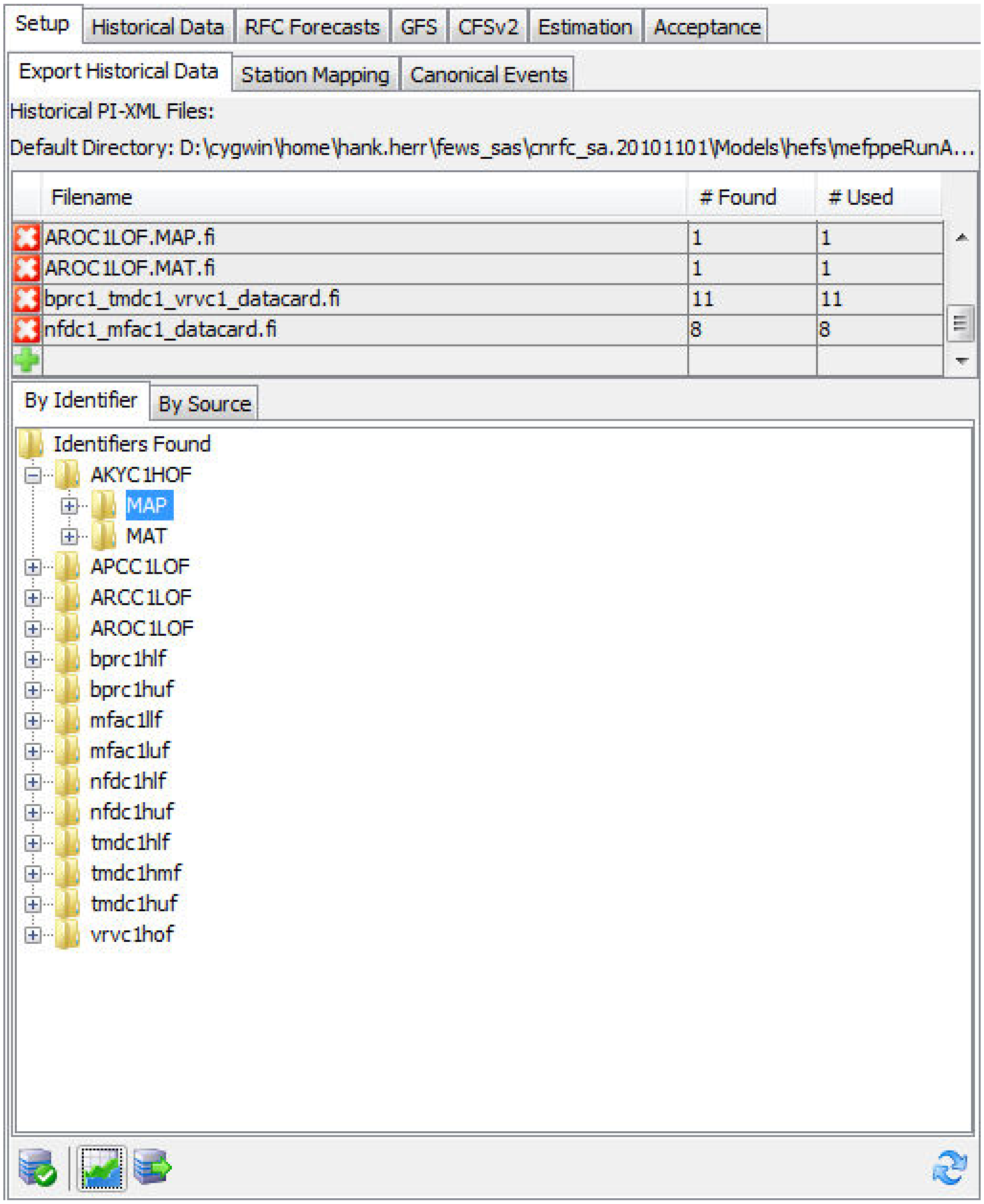


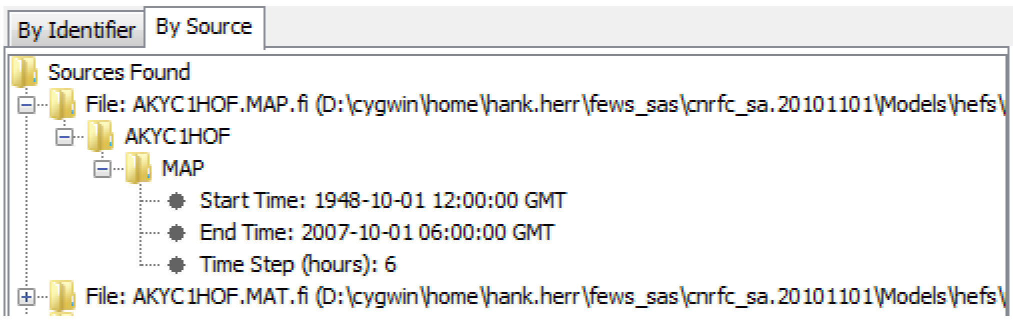
Figure 4: The **Setup Subpanel** of the **Estimation Steps Panel**.

3.5.1 Export Historical Data Subpanel

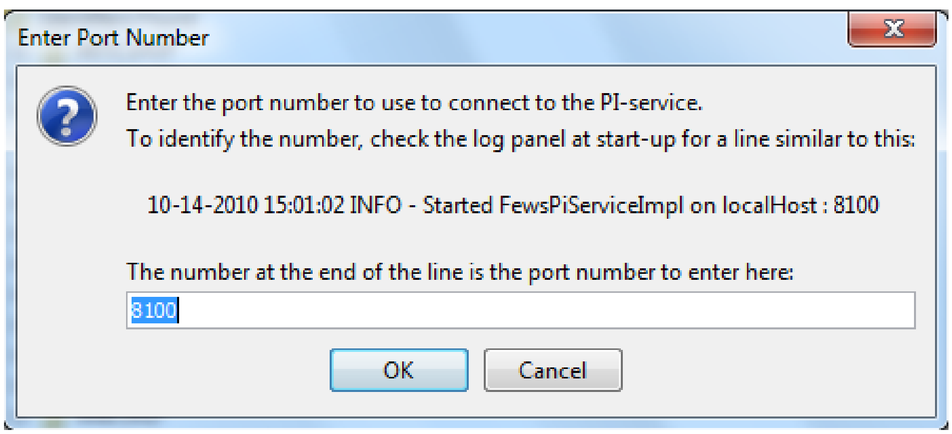
The **Export Historical Data Subpanel**, shown above in Figure 4, facilitates creating time series XML files for use by the MEFPPE based on time series acquired via the FEWS PI-service. It also facilitates viewing and quality controlling that data. The time series are gathered by examining files in the directory 'historicalData' within the MEFPPE run area (Section 3.2.4.1). Available data for both precipitation and temperature are displayed.

3.5.1.1 Components

- **Historical PI-XML Files Table**: Displays which files were found specifying historical time series. A delete column is included which, when clicked, removes the clicked row's file from the file system. A confirmation dialog will be shown before the file is deleted.
- **By Identifier Tree/By Source Tree**: The contents of the files found are listed in two trees displayed via a tabbed panel: the **By Identifier Tree** displays the time series first by the MEFP location identifier (locationId and parameterId), while the **By Source Tree** displays the time series first by source file found. The information provided in the tree includes locationId, parameterId, source file, and the start time, end time, and time step of the time series found. Both trees are selectable. Expand the tree nodes in order to view this information. For example:



-  **Reconnect to CHPS PI-service Button**: Click to open an **Enter Port Number Dialog** that allows for entering a port number to use for connecting to the PI-service:



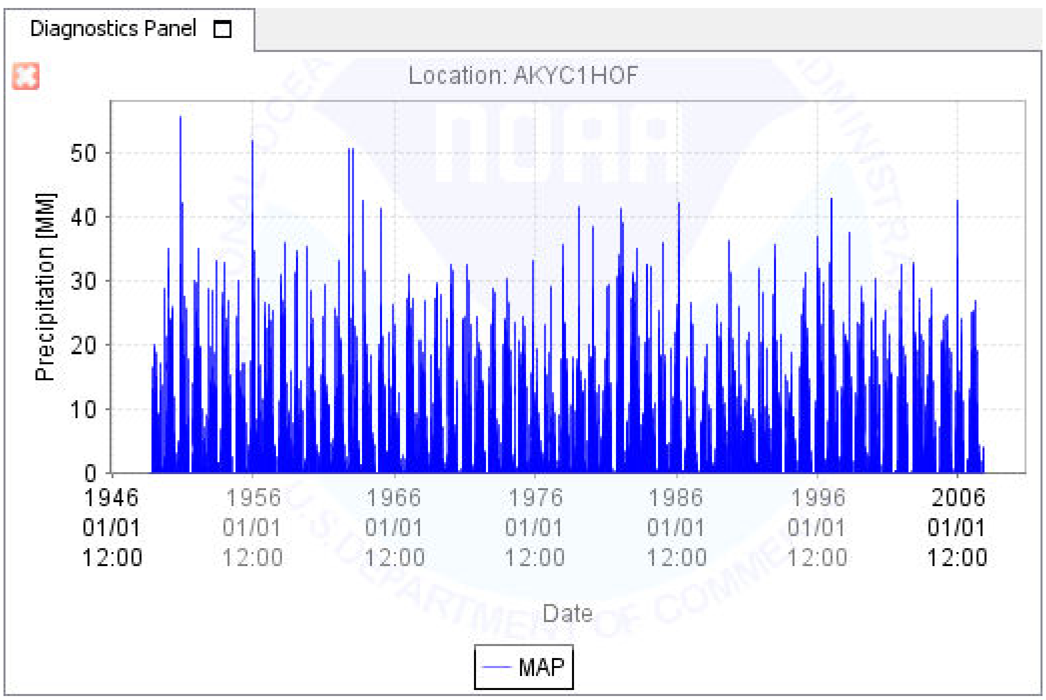
This is useful if the port number used previously (default is 8100) failed to yield a connection or connected to the wrong PI-service. A status icon is included within the larger icon,  indicates a good connections, while  indicates a bad connection. When the connection is bad, the **Export Time Series from CHPS DB Button** will be disabled. Detailed instructions for identifying this port number are provided in Section 3.2.4.4.

NOTE: It is possible for the button to indicate a good connection even though it connected to the wrong PI-service.

-  **View Button**: Click to view time series selected from either the **By Identifier Tree** or **By Source Tree**. To view a time series, all selected nodes or leaves in the tree must be for the same CHPS locationId and the same data type.
-  **Refresh Button**: Reread the files in the historicalData directory of the MEFPPE run area and reconstruct the two trees. This needs to be clicked only if the files in the historicalData are modified manually while running the MEFPPE, as in Section 3.5.4.2.

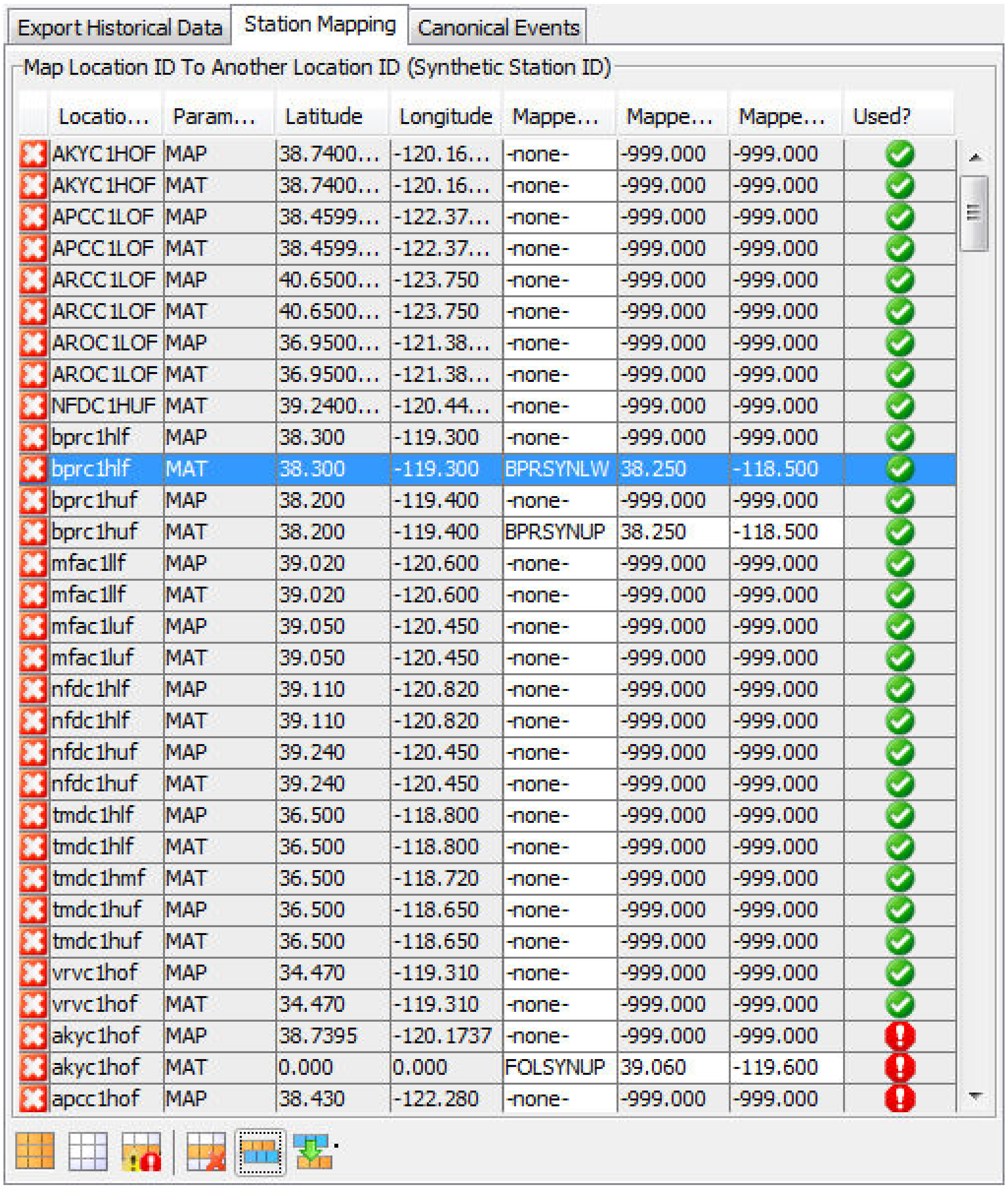
3.5.1.2 Diagnostics

The diagnostics displayed for this subpanel are the time series as provided in the XML or fastInfoset files. For both precipitation and temperature data, the time series are displayed as blue lines:



3.5.2 Station Mapping Subpanel

The **Station Mapping Subpanel**, shown in Figure 5, allows the user to define station mappings, or MEFP locations for which data for other locations are to be used. This is most useful if an MEFP location is to use RFC QPF/QTF data from an external provider (e.g., MDP MOS forecasts of temperature) and the location for that external provider not identical to the MEFP location, either in locationId or coordinates. In Figure 5, the selected row is a station mapping indicating that for MEFP location bprc1hlf (MAT), data for MOS forecast station BPRSYNLW, with coordinates (38.250, -118.50), will be used. For RFC QPF/QTF, then, the location id 'BPRSYNLW' will be used to find the archive forecast data files. Furthermore, the mapped coordinates will be used to identify which GFS and CFSv2 reforecast files must be acquired. The station mappings for both precipitation and temperature are defined in a single table.



	Locatio...	Param...	Latitude	Longitude	Mappe...	Mappe...	Mappe...	Used?
	AKYC1HOF	MAP	38.7400...	-120.16...	-none-	-999.000	-999.000	
	AKYC1HOF	MAT	38.7400...	-120.16...	-none-	-999.000	-999.000	
	APCC1LOF	MAP	38.4599...	-122.37...	-none-	-999.000	-999.000	
	APCC1LOF	MAT	38.4599...	-122.37...	-none-	-999.000	-999.000	
	ARCC1LOF	MAP	40.6500...	-123.750	-none-	-999.000	-999.000	
	ARCC1LOF	MAT	40.6500...	-123.750	-none-	-999.000	-999.000	
	AROC1LOF	MAP	36.9500...	-121.38...	-none-	-999.000	-999.000	
	AROC1LOF	MAT	36.9500...	-121.38...	-none-	-999.000	-999.000	
	NFDC1HUF	MAT	39.2400...	-120.44...	-none-	-999.000	-999.000	
	bprc1hlf	MAP	38.300	-119.300	-none-	-999.000	-999.000	
	bprc1hlf	MAT	38.300	-119.300	BPRSYNLW	38.250	-118.500	
	bprc1huf	MAP	38.200	-119.400	-none-	-999.000	-999.000	
	bprc1huf	MAT	38.200	-119.400	BPRSYNUP	38.250	-118.500	
	mfac1llf	MAP	39.020	-120.600	-none-	-999.000	-999.000	
	mfac1llf	MAT	39.020	-120.600	-none-	-999.000	-999.000	
	mfac1luf	MAP	39.050	-120.450	-none-	-999.000	-999.000	
	mfac1luf	MAT	39.050	-120.450	-none-	-999.000	-999.000	
	nfdc1hlf	MAP	39.110	-120.820	-none-	-999.000	-999.000	
	nfdc1hlf	MAT	39.110	-120.820	-none-	-999.000	-999.000	
	nfdc1huf	MAP	39.240	-120.450	-none-	-999.000	-999.000	
	nfdc1huf	MAT	39.240	-120.450	-none-	-999.000	-999.000	
	tmdc1hlf	MAP	36.500	-118.800	-none-	-999.000	-999.000	
	tmdc1hlf	MAT	36.500	-118.800	-none-	-999.000	-999.000	
	tmdc1hmf	MAT	36.500	-118.720	-none-	-999.000	-999.000	
	tmdc1huf	MAP	36.500	-118.650	-none-	-999.000	-999.000	
	tmdc1huf	MAT	36.500	-118.650	-none-	-999.000	-999.000	
	vrvc1hof	MAP	34.470	-119.310	-none-	-999.000	-999.000	
	vrvc1hof	MAT	34.470	-119.310	-none-	-999.000	-999.000	
	akyc1hof	MAP	38.7395	-120.1737	-none-	-999.000	-999.000	
	akyc1hof	MAT	0.000	0.000	FOLSYNUP	39.060	-119.600	
	apcc1hof	MAP	38.430	-122.280	-none-	-999.000	-999.000	

Figure 5: The **Station Mapping Subpanel** of the **Setup Subpanel**.

3.5.2.1 Components

- **Mapping Table**: Labeled “Map Location ID To Another Location ID (Synthetic Station ID)”, this table allows for defining station mappings. It is a standard **Generic Summary Table** (see Section 3.2.5.1) with a status column titled ‘Used?’ and a delete column (see the **Delete Selected Rows Button** below for which deletions can be made). The ‘Used?’ column displays if the MEFP location for a row has historical data corresponding to it, so that parameters can be estimated (i.e., it is displayed in the **Location Summary Panel**). There are three additional editable columns: ‘Mapped Station Id’, ‘Mapped Lat’, and ‘Mapped Lon’. The first allows for defining a mapped station id. The latter two allow for specifying the coordinates of the mapped station id and can only be edited if the mapped station id is not “-none-”. If the mapped station id is defined as one that already exists for another MEFP location, the mapped coordinates will be set to the values defined for that other location and the coordinates will always be kept in sync. To clear a station mapping, edit the ‘Mapped Station Id’ column, clear it (delete all characters), and press <Enter>.

NOTE: If a mapped station id is defined, its coordinates must be provided or GFS and CFSv2 reforecast files will not be acquired.

- **Delete Selected Rows Button**: Removes the selected row from the table. Only rows for which the ‘Used?’ column displays can be removed. This is generally useful if you imported the station mappings (Section 3.2.4.3) and many mappings were brought in for unused MEFP locations.
- **Copy the Mapped Location Information**: Copies the mapped station id and coordinates into memory.
- **Paste the Copied Mapped Location Info to Selected Rows Button**: Applies the copied station id and coordinates to the MEFP locations for the selected rows.

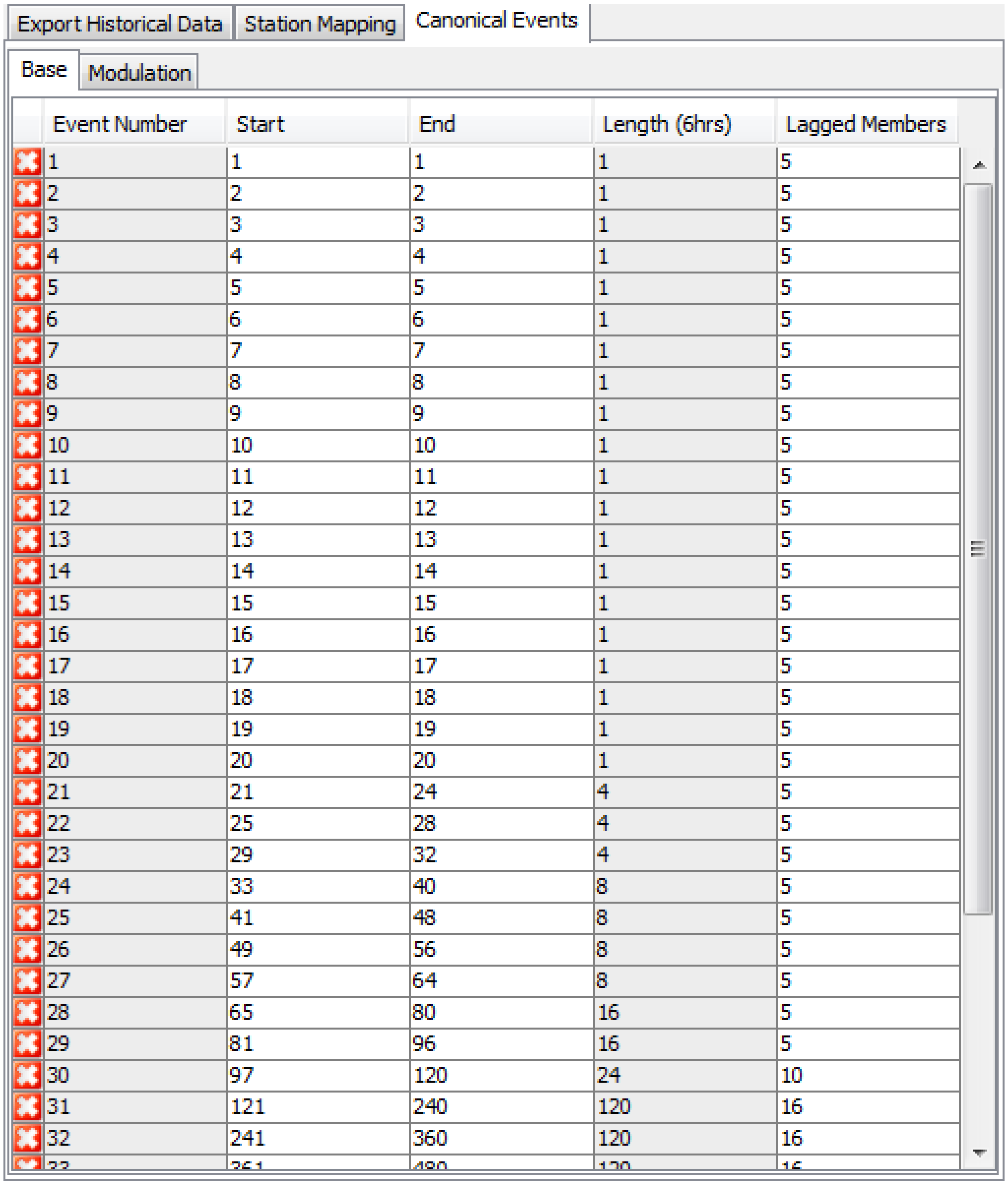
3.5.3 Canonical Events Subpanel

The **Canonical Events Subpanel**, shown in Figure 6, allows the user to modify the base and modulation canonical events, described fully in the scientific documentation accompanying this software and summarized in Section 3.2.4.5. A canonical event is defined by a start period and an end period. Also, specific to the CFSv2 forecast data source, the number of lagged ensembles to use can be specified.

3.5.3.1 Components

- **Base/Modulation Tables**: The **Base Table** and the **Modulation Table**, each in its own tab, display the current canonical events and allow the user add new events or modify existing events. The columns are as follows:
    - ‘Event Number’: The number of the event in sorted (by end period) order, or the row number.
    - ‘Start’: The start period of the event. This column is editable (click and type).
    - ‘End’: The end period of the event. This column is editable (click and type).
    - ‘Length’: The length of the canonical event in periods, with the period unit provided in the column header.
    - ‘Lagged Members’: The number of lagged ensemble members to use for the canonical event when estimating parameters for CFSv2. This column is editable (click and type).

  Whenever a change is made to any canonical event, the table is resorted and the event numbers reassigned. The table includes an add/delete column for adding new canonical events (click on at the bottom of the table) or removing existing ones.


Export Historical Data | Station Mapping | Canonical Events

Base | Modulation

Event Number	Start	End	Length (6hrs)	Lagged Members
1	1	1	1	5
2	2	2	1	5
3	3	3	1	5
4	4	4	1	5
5	5	5	1	5
6	6	6	1	5
7	7	7	1	5
8	8	8	1	5
9	9	9	1	5
10	10	10	1	5
11	11	11	1	5
12	12	12	1	5
13	13	13	1	5
14	14	14	1	5
15	15	15	1	5
16	16	16	1	5
17	17	17	1	5
18	18	18	1	5
19	19	19	1	5
20	20	20	1	5
21	21	24	4	5
22	25	28	4	5
23	29	32	4	5
24	33	40	8	5
25	41	48	8	5
26	49	56	8	5
27	57	64	8	5
28	65	80	16	5
29	81	96	16	5
30	97	120	24	10
31	121	240	120	16
32	241	360	120	16
33	361	480	120	16

Figure 6: The **Canonical Events Subpanel** of the **Setup Subpanel**.

3.5.4 Usage

3.5.4.1 Configuring the FEWS PI-service

MEFPPE uses the FEWS PI-service to acquire historical time series from the CHPS database. The query performed is defined in the file:

Config/PiServiceConfigFiles/MEFPPE.xml

with a query id "All Historical Data". The query must be configured to return all MAP and MAT time series that will be used as historical data for the MEFP parameter estimation. Here is an example:

```
<fewsPiServiceConfig xmlns="http://www.wldelft.nl/fews" xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance" xsi:schemaLocation="http://www.wldelft.nl/fews
http://fews.wldelft.nl/schemas/version1.0/fewsPiServiceConfig.xsd">
  <general>
    <importIdMap>IdImportPiService</importIdMap>
    <exportIdMap>IdExportPiService</exportIdMap>
  </general>

  <timeSeries>
    <id>All Historical Data</id>
    <timeSeriesSet>
        <moduleInstanceId>ImportDataCard</moduleInstanceId>
        <valueType>scalar</valueType>
        <parameterId>MAP</parameterId>
        <locationSetId>Catchments</locationSetId>
        <timeSeriesType>external historical</timeSeriesType>
        <timeStep unit="hour" multiplier="6" timeZone="GMT-6"/>
        <readWriteMode>read only</readWriteMode>
        <synchLevel>1</synchLevel>
    </timeSeriesSet>
    <timeSeriesSet>
        <moduleInstanceId>ImportDataCard</moduleInstanceId>
        <valueType>scalar</valueType>
        <parameterId>MAT</parameterId>
        <locationSetId>Catchments</locationSetId>
        <timeSeriesType>external historical</timeSeriesType>
        <timeStep unit="hour" multiplier="6" timeZone="GMT-6"/>
        <readWriteMode>read only</readWriteMode>
        <synchLevel>1</synchLevel>
    </timeSeriesSet>
  </timeSeries>
</fewsPiServiceConfig>
```

Instructions for configuring this file are provided in the *HEFS Release Install Notes*.

3.5.4.2 Exporting Historical Data Files Manually

The MEFPPE searches the directory 'historicalData' within the MEFPPE run area (Section 3.2.4.1) for XML and fastInfoset files specifying historical MAP and MAT time series. To specify those files manually, create a FEWS pi-timeseries compliant XML or fastInfoset file containing those time series an place it within the directory:

Models/hefs/mefppeRunArea/historicalData

The files can be manually copied or exported directly into the directory. Any file in the directory that cannot be read or is not following an appropriate schema will be skipped and a warning message will be displayed in the CHPS **Logs Panel**. If the directory contents are modified while the MEFPPE is running, click on the **Refresh Button** in the **Export Historical Data Subpanel** (see Section 3.5.1.1).

3.5.4.3 Defining a Station Mapping

To define a station mapping for an existing MEFP location, do the following:

1. Make the **Station Mapping Subpanel** active by clicking on the tab.
2. In the **Mapping Table**, find the row for which the 'Location ID' and 'Parameter ID' match that for the MEFP location. Click on the cell for the 'Mapped Station ID' column and type in the mapped station id and press <Enter>. For example, here AKYC1HOF (MAT) has been mapped to location id "STATION":

Map Location ID To Another Location ID (Synthetic Station ID)

Location ID	Parameter ID	Latitude	Longitude	Mapped S...	Mapped Lat	Mapped Lon	Used?
AKYC1HOF	MAP	38.7400017	-120.1600...	-none-	-999.000	-999.000	✓
AKYC1HOF	MAT	38.7400017	-120.1600...	STATION	0.000	0.000	✓
APCC1LOF	MAP	38.4599991	-122.3799...	-none-	-999.000	-999.000	✓
APCC1LOF	MAT	38.4599991	-122.3799	-none-	-999.000	-999.000	✓

Upon pressing <Enter>, the 'Mapped Station ID' cell value will change to the typed in value, the 'Mapped Lat' and 'Mapped Lon' cell values for the row will be set to 0.000, and the next row in the table will be selected. If the 'Mapped Station ID' matches that for another row in the table, its coordinate cell values will be set to the values for that other matching row.

3. Edit the 'Mapped Lat' and 'Mapped Lon' cell values appropriately (click and type).

   To clear the defined station mapping, clear the 'Mapped Station ID' cell value and press <Enter>. Upon pressing <Enter> the cell values for the row will indicate no mapped station:

Map Location ID To Another Location ID (Synthetic Station ID)

Location ID	Parameter ID	Latitude	Longitude	Mapped S...	Mapped Lat	Mapped Lon	Used?
AKYC1HOF	MAP	38.7400017	-120.1600...	-none-	-999.000	-999.000	✓
AKYC1HOF	MAT	38.7400017	-120.1600...	-none-	-999.000	-999.000	✓
APCC1LOF	MAP	38.4599991	-122.3799...	-none-	-999.000	-999.000	✓
APCC1LOF	MAT	38.4599991	-122.3799	-none-	-999.000	-999.000	✓

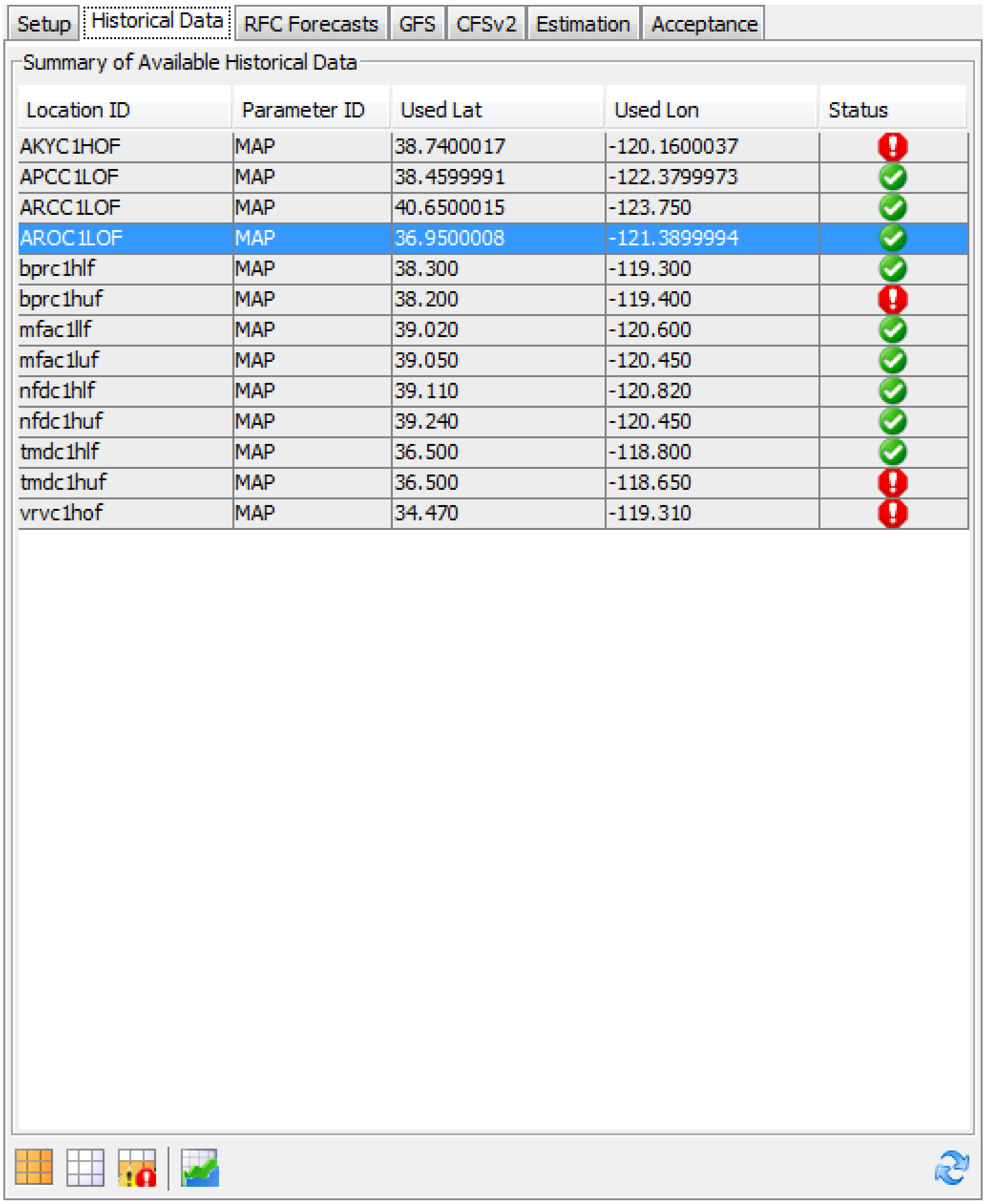
3.5.4.4 Adding a New Canonical Event

To define a new canonical event, do the following:

1. Click on at the bottom of the table in the add/delete column. You may need to scroll down. A new row will be added to the table with a 'Start' cell value of -1 and an 'End' cell value equal to one larger than the largest 'End' cell value in the table, so that the new row sorts to the bottom of the table.
2. Edit the 'Start', 'End', and 'Lagged Members' columns appropriately. For each, click on the cell, type the new value, and press <Enter>. Upon pressing <Enter> for the 'End' cell value, the table will resort and the row will move appropriately and will remain selected.

3.6 Historical Data Subpanel

The **Historical Data Subpanel** of the **Estimation Steps Panel**, shown in Figure 7 is used to perform Step 2 of the parameter estimation procedure in Section 3.2.3: processing historical data and generating binary files for fast access during operational ensemble generation.



Location ID	Parameter ID	Used Lat	Used Lon	Status
AKYC1HOF	MAP	38.7400017	-120.1600037	
APCC1LOF	MAP	38.4599991	-122.3799973	
ARCC1LOF	MAP	40.6500015	-123.750	
AROC1LOF	MAP	36.9500008	-121.3899994	
bprc1hlf	MAP	38.300	-119.300	
bprc1huf	MAP	38.200	-119.400	
mfac1llf	MAP	39.020	-120.600	
mfac1luf	MAP	39.050	-120.450	
nfdc1hlf	MAP	39.110	-120.820	
nfdc1huf	MAP	39.240	-120.450	
tmdc1hlf	MAP	36.500	-118.800	
tmdc1huf	MAP	36.500	-118.650	
vrvc1hof	MAP	34.470	-119.310	

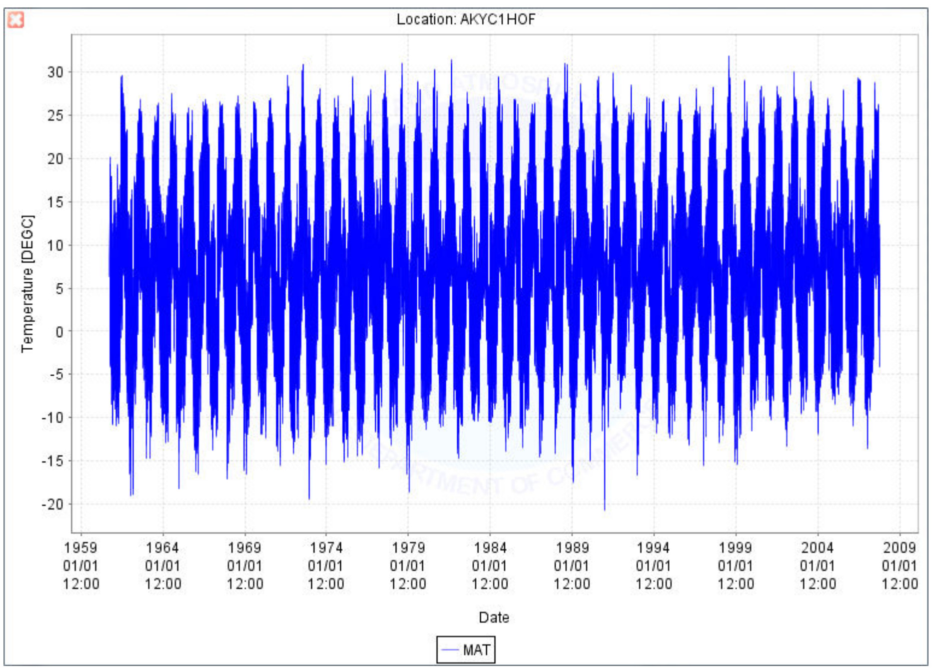
Figure 7: The **Historical Data Subpanel** of the **Estimation Steps Panel**.

3.6.1 Components

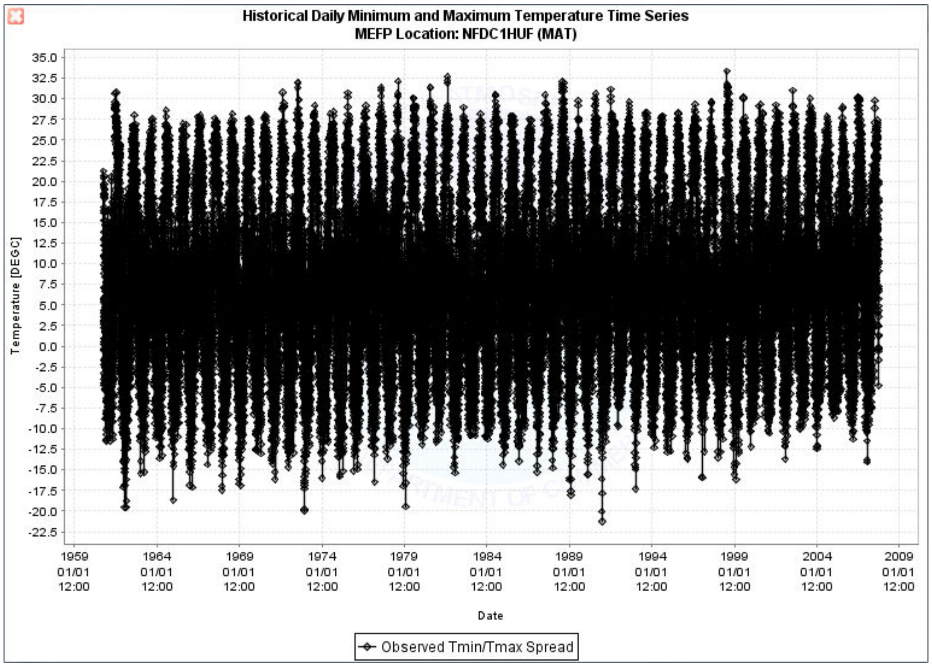
- **Summary of Available Historical Data Table**: Allows the user to select the MEFP locations for which to perform the step and view for which locations the step has been performed. It is a **Generic Summary Table** (Section 3.2.5.1) and includes all the standard buttons.
-  **View Button**: Displays diagnostics for a single MEFP location selected in the **Summary of Available Historical Data Table**.

3.6.2 Diagnostics

For precipitation, the diagnostic displayed when the **View Button** is clicked are the raw precipitation time series, and should be identical to those displayed for the **Export Historical Data Subpanel** (Section 3.5.1.2):



For temperature, the diagnostic displayed will show the 24-hour maximum and minimum temperatures computed from the historical temperature data defined in the **Export Historical Data Subpanel**. For example:



3.6.3 Usage

3.6.3.1 Generating Binary Data Files

To create binary data files to be kept with the MEFP parameters and used operationally when generating ensemble, select one or more locations from the **Summary of Available Historical Data Table** and click on the **Perform Step Button**. See Section 3.4.2.1 for details on how to perform a step using the **Perform Step Button** of the **Estimation Steps Panel**.

3.7 RFC Forecasts Subpanel

The **RFC Forecasts Subpanel** of the **Estimation Steps Panel**, shown in Figure 8, is used to perform Step 3 of the parameter estimation procedure in Section 3.2.3: acquire or create HPS/RFC QPF/QTF archive forecast files. The RFC archive forecast files are put in place by the MEFPPE in one of two ways:

1. Importing (copying) files constructed by RFC users.
2. Constructing the files based on data extracted from the vfypairs table of the archive database. See the Usage section below for more details.

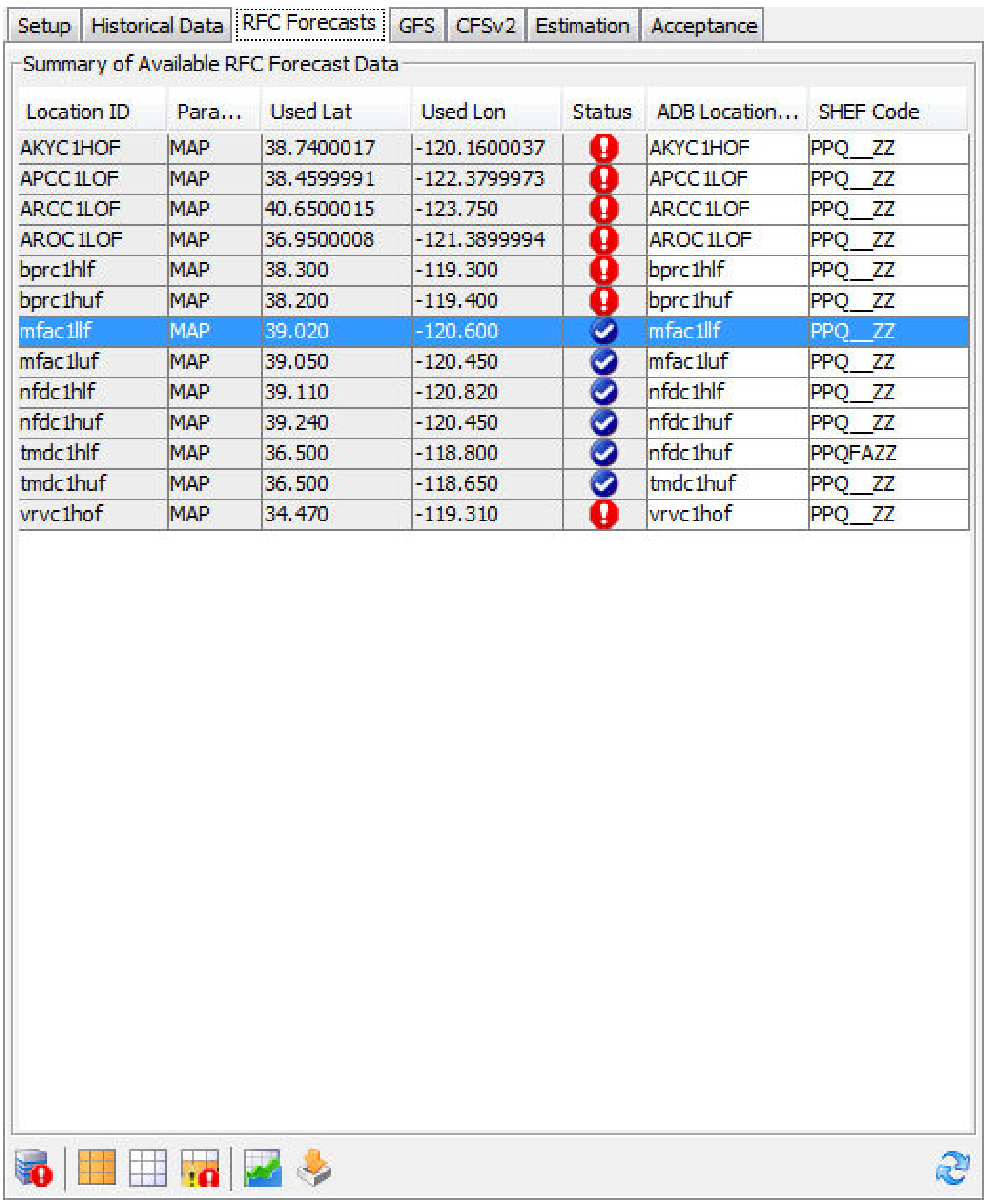
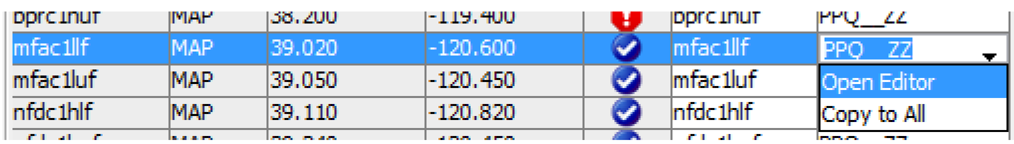


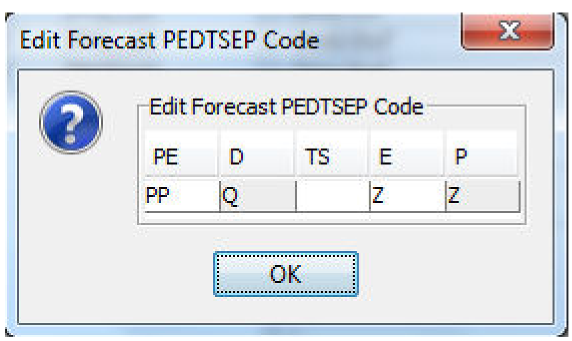
Figure 8: The **RFC Forecasts Subpanel** of the **Estimation Steps Panel**.

3.7.1 Components

- **Summary of Available RFC Forecast Data Table**: Allows the user to select the MEFP locations for which to perform the step and view for which locations the step has been performed. It is a **Generic Summary Table** (Section 3.2.5.1) and includes all the standard columns and buttons. Note that, for the 'Status' column, a indicates that the archived data files were imported, whereas indicates that the archived data files were constructed by MEFPPE using archive database data. The following additional editable columns are included: the 'ADB Location ID' and one (precipitation) or two (temperature) 'SHEF Code' columns. For temperature, the names of the SHEF code columns are 'TMIN SHEF Code' and 'TMAX SHEF Code'. The extra columns identify the rows of the vfypairs table within the archive database that will be extracted and used to create the archive forecast file. Cell values for the additional columns can be edited directly (click and type). The SHEF code columns can also be edited via the **Edit Forecast PEDTSEP Code Dialog** (see below).
- **Edit Forecast PEDTSEP Code Dialog**: Allows the user to specify a SHEF code column value for cell. It is accessed by clicking on a cell to edit and selecting the "Open Editor" menu item:

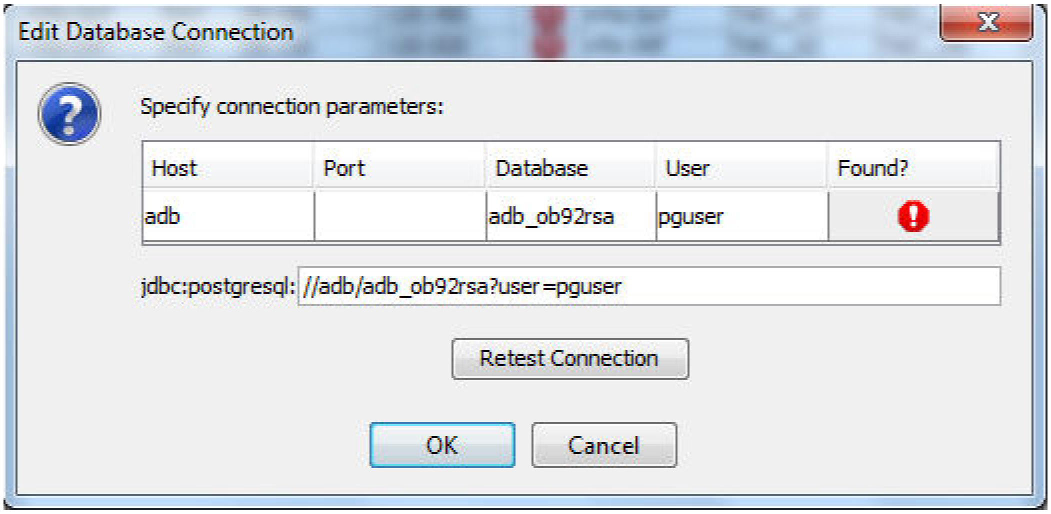


The dialog appears as follows:



The 'SHEF Code' cell value is edited by editing the contents of the table provided in the dialog (click in a cell and type the entry). The 'D'uration and 'P'robability cell values cannot be edited. The duration must always be Q (6-hourly) for precipitation data and D (daily) for temperature data, while the probability must always be Z. See Section 3.7.3.3 for more details.

- **Edit Archive Database Connection Button**: Allows the user to edit the connection used for the archive database. A status icon is included within the larger button icon: indicates a successful connection, while indicates a bad connection. The archive database cannot be accessed in order to construct archive forecast files unless the connection is successful. Clicking on the button opens up an **Edit Database Connection Dialog** to edit the connection parameters:



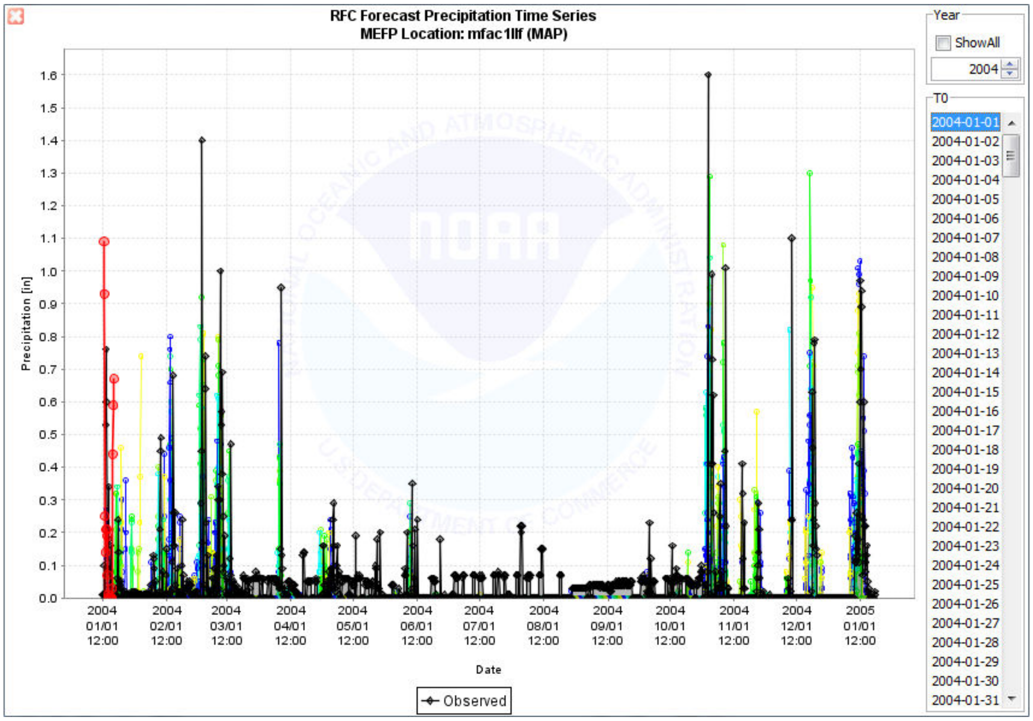
By default, all cells of the **Specify Connection Parameters Table** will be empty, except for the 'Database' cell which will be filled with a default value of "DATABASENAME". For more details on how to connect to the archive database, see Section 3.7.3.1.

- **View Button**: Displays diagnostics for a single MEFP location selected in the **Summary of Available RFC Forecast Data Table**.
- **Import Prepared RFC Forecast Files Button**: Click to import externally constructed RFC archive forecast files. When clicked, a **Choose Directory Dialog** will be opened. Navigate to the directory to import and click **Import Directory**. For more details, see Section 3.7.3.4.

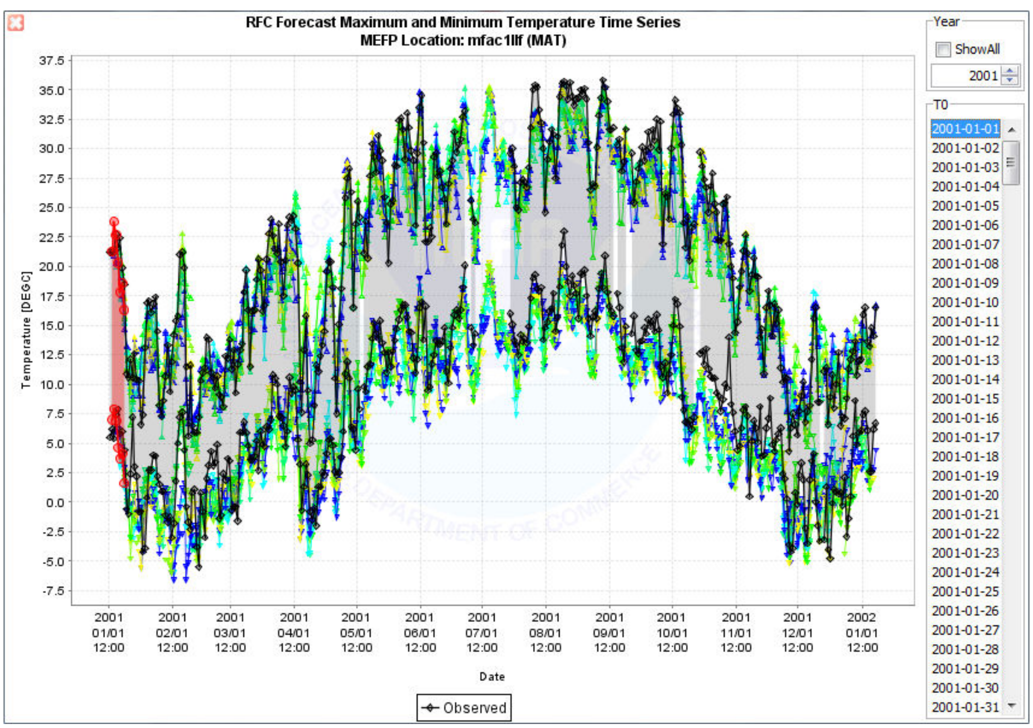
3.7.2 Diagnostics

The current diagnostic displays provided when the **View Button** is clicked for the RFC archived data are designed primarily for data quality control. Both display the archived forecast time series along with archived observed time series, all provided in the RFC archived data files. The chart is displayed within a standard **Archive/Reforecast Time Series Diagnostic Display Panel** (Section 3.2.5.3).

For precipitation, the displayed forecast and observed time series are 6-hour accumulated precipitation time series. For example:



For temperature, the displayed forecast and observed time series are 24-hour minimum and maximum temperatures. The area between corresponding minimum and maximum values is shaded for the observed data (dark gray), as well as selected (i.e., emphasized) time series. For example:



3.7.3 Usage

3.7.3.1 Using IVP to Pair RFC QPF/QTF Forecasts

The MEFPPE acquires archived forecast and observed values, including precipitation and minimum and maximum temperatures, from the vfypairs table of the archive database. The vfypairs table is populated by the Interactive Verification Program (IVP), delivered with AWIPS. Documentation is made available on the NWS website:

http://www.nws.noaa.gov/oh/hrl/verification/verification_doc_ob82.php

The general steps to be following to populate the vfypairs table of the archive database so that MEFPPE can find the required archived forecast is as follows:

1. Identify the archive database connection information. This is best done by determining the values of apps-defaults tokens as follows:

   Host machine: identify the value of the apps-defaults token rax_pghost; typically “ax”.
   Database name: identify the value of the apps-defaults token adb_name; typically “adb_ob” followed by text identifying the AWIPS release and three-letter RFC code. For example, adb_ob7rsa.
   User name: typically “pguser”.

2. Import all necessary archive database into the archive database. This should be done via the standard SHEF importing tools. The data required is as follows:

   Precipitation: 6-hour accumulated precipitation amounts, both forecast and observed. The SHEF code must include a duration code of “Q” and a probability code of “Z”. The physical element will likely be “PP” while the extremum will be “Z”. The type sources will vary between forecast and observed.

   Temperature: 24-hour minimum and maximum temperature values, both forecast and observed. The SHEF code must include a duration code of “D” and a probability code of “Z”. The physical element will likely be “” while the extremum will be “N” for minimum and “X” for maximum temperatures. The type sources will vary between forecast and observed.

   Note the SHEF codes used for both the forecast and observed data, as they are needed in order to configure IVP to populate the vfypairs table.

3. Log onto the archive database host machine.

4. Use the Vfyruninfo Editor and add verification locations in order to build pairs of forecasts and observations. For instructions on using the Vfyruninfo Editor, check Section 7 of the user’s manual:

   http://www.nws.noaa.gov/oh/hrl/verification/ob8/VfyruninfoEditorGUI.pdf

5. Construct a pairing batch file for the IVP Batch Program. The following is an example that will construct all pairs for locations AAAAA, BBBBB, and CCCCC and physical elements PP and TA where the type source of the 6-hour accumulated precipitation forecasts and 24-hour minimum and maximum temperatures is FF:

```
# Define the pairing parameters for a run for 12 years, turning
# persistence off as it is not needed for MEFPPE.
START_TIME = "2000-01-01 00:00:00"
END_TIME = "2011-12-31 00:00:00"
PERSISTENCE = OFF

# Specify locations for pairing and execute the pairing for
# precipitation.
LOCATION = "AAAAA, BBBBB, CCCCC"
FCST_TS = "FF"
PE = "PP"
DUR = "Q"
EXTREMUM = "Z"
BUILD_PAIRS = true

# Specify locations for pairing and execute the pairing for
# temperature.
LOCATION = "AAAAA, BBBBB, CCCCC"
FCST_TS = "FF"
PE = "TA"
DUR = "Q"
EXTREMUM = "N,X"
BUILD_PAIRS = true
```

6. Execute the IVP Batch Program to build the pairs. For more information, check the manual:

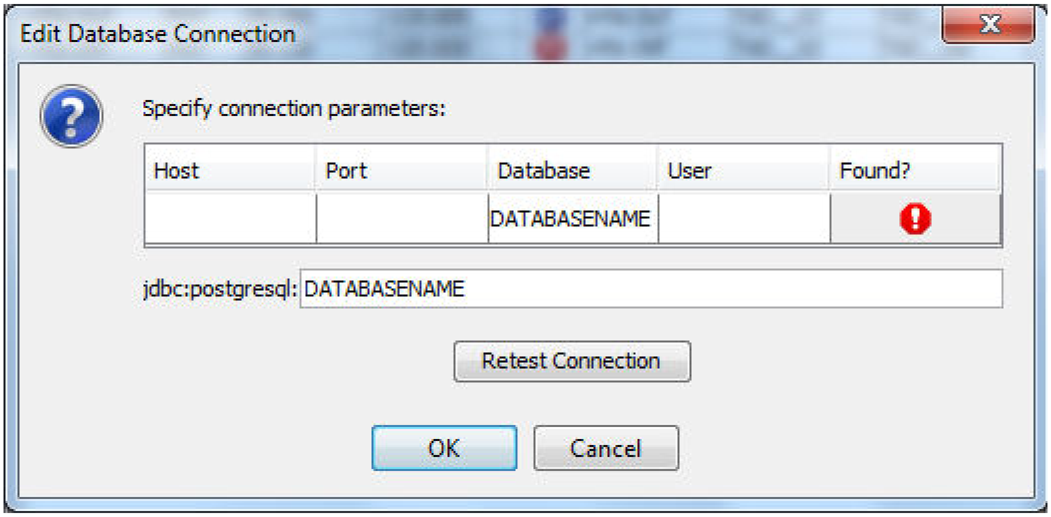
   http://www.nws.noaa.gov/oh/hrl/verification/ob8/PairingBatch.pdf

   Upon completion of Step 6, the vfypairs table should be adequately populated for use by MEFPPE in constructing RFC archived forecast data files.

3.7.3.2 Specifying the Archive Database Connection

To specify the archive database connection so that the MEFPPE can acquire the needed data from the archived database, do the following:

1. Click on the **Edit Archive Database Connection Button**. The **Edit Database Connection Dialog** with open with either empty connection settings or settings that were defined previously:



2. Enter (click, type, and press <Enter>) the archive database (or RAX) host machine name in the 'Host' cell. Typically this should be set to "ax".
3. Leave the 'Port' cell empty.
4. Enter the archive database name in the 'Database' cell. Typically this should be set to something similar to "adb_ob###".
5. Enter a valid database user into the 'User' cell. Typically this should be set to pguser. Upon pressing <Enter> after entering the user, MEFPPE will attempt to make a connection. If successful, the 'Found?' cell image will display . If unsuccessful, it will display .
6. Click **OK** to accept the database connection settings.

To retest the connection at any time, click the **Retest Connection Button**.

3.7.3.3 Constructing Archive Files Using Archive Database Data

In order for the MEFPPE to build the RFC archived data files, the QPF/QTF forecast-observed pairs must be available in the archive (RAX) database vfypairs table (see Section 3.7.3.1). If that is the case, then identify the lid and SHEF code used to store the data in the archive database. Then do the following in order to make MEFPPE generate the files:

1. In the **Summary of Available RFC Forecast Data Table**, select the row for the location for which RFC archived files will be generated.
2. Click on the cell for 'ADB Location ID', and enter the lid used in the archive database vfypairs table for the data to use. Press <Enter> when done.
3. Click on the cell for 'SHEF Code' and edit it to be the SHEF code used in the archive database vfypairs table for the data to use. The cell can be modified by editing it directly (type in the code and press <Enter>) or via the **Edit Forecast PEDTSEP Code Dialog** (see Section 3.7.1). For precipitation, one 'SHEF Code' cell must be modified. For temperature, two 'SHEF Code' cell must be modified: one for 24-hour minimum temperature (TMIN), and one for 24-hour maximum temperature (TMAX).
4. Ensure that the archive database connection properly specified (Section 3.7.3.2).
5. Click on the **Perform Step Button** of the **Estimation Steps Panel**. See Section 3.4.2.1 for details.

3.7.3.4 Importing Files Constructed Externally

In order to import RFC archived files, properly formatted files must be constructed and placed in an acceptable directory structure. The process begins by choosing a directory to be the <import directory> and creating this directory structure:

```
<import directory>/…
    rfc_pfcst06
    rfc_pobs06
    rfc_tfcst
    rfc_tobs
```

Note that this directory structure may already exist for those who have previously used the eXperimental Ensemble Forecast System (XEFS) Ensemble Pre-Processor 3 (EPP3) software upon which MEFP is based. The first two directories store forecast and observed 6-hour accumulated precipitation amount files, respectively, one file in each directory per location. The latter two directories store forecast and observed 24-hour minimum and maximum temperature files, respectively, two files for each location per directory. All of the needed files must be ASCII and must follow a specific format described below. The names of the files must also follow a specific pattern described below.

After importing, for RFC archive data files to be considered available or ready by MEFPPE for an MEFP location, *all* files must be present for that MEFP location's CHPS location id and parameter id. For precipitation locations, a forecast file under rfc_pfcst06 and an observed file under rfc_pobs06 must be present. For temperature, minimum and maximum forecast files under rfc_tfcst and minimum and maximum observed files under rfc_tobs must be present.

The pages that follow described required file formats for the different file types.

6-hour Accumulated Precipitation Forecast Files:
Directory: <import directory>/rfc_pfcst06
File Naming Convention: <MEFP locationId>.pfcst06
Example: NFDC1HUF.pfcst06
Header:
The header of the file is optional and specifies properties of the data. The properties are provided in lines of this format:

```
<key> = <value>
```

If the header exists, it must be closed with an end line:

```
end
```

The following properties are valid (default values used if the property is not provided are specified in parentheses):

- units: The units of the measurements, either "in" or "mm". (default: "in"; data is converted to mm when read in prior to estimating parameters).
- nfcstdays: The number of days of forecast values. A positive integer not greater than 5. There is no default: the parameter is required!
- iyr: The initial year of data available, a four digit year. (default: determined from the data in the file)
- lyr: The last year of data available, a four digit year. (default: determined from the data in the file)

Example:

```
units = (in/mm)
nfcstdays = 5
end
```

Data:
Every line of the file following the header must specify a single forecast time series (<> are used to separate the fields; do not include them in the text):

```
<T0: YYYYMMDD><####.##><####.##><####.##>...<####.##>
```

The first field is the T0 forecast day (YYYYMMDD format) and it is followed by one space. Starting at the 10th character, forecast values are provided that are 7 characters wide allowing for two decimal places. There must be a number of forecast values equal to nfcstdays multiplied by 4. The forecast time, or basis time, of each forecast time series must be 12Z, with the first value of the line having a valid time of 18Z. Time steps between values must be 6 hours, so that the valid times of following values are 0Z, 6Z, 12Z, etc.

Example:

```
19991231  -99.99 -99.99 -99.99 -99.99 -99.99 -99.99 -99.99 -99.99 -
99.99 -99.99 -99.99 -99.99   0.00   0.00   0.00   0.00   0.00   0.00
0.00   0.00
20000101    0.00   0.00   0.00   0.00   0.00   0.00   0.00   0.00
0.00   0.00   0.00   0.00   0.00   0.07   0.07   0.00   0.00   0.00
0.00   0.00
...
```

Full File Example:

```
units = in
nfcstdays = 5
end
19991231  -99.99 -99.99 -99.99 -99.99 -99.99 -99.99 -99.99 -99.99 -99.99 -
99.99 -99.99 -99.99   0.00   0.00   0.00   0.00   0.00   0.00   0.00   0.00
20000101    0.00   0.00   0.00   0.00   0.00   0.00   0.00   0.00   0.00
0.00   0.00   0.00   0.00   0.07   0.07   0.00   0.00   0.00   0.00   0.00
20000102    0.00   0.00   0.00   0.00   0.00   0.00   0.00   0.00   0.18
0.09   0.00   0.00   0.00   0.00   0.00   0.00   0.00   0.00   0.00   0.00
20000103    0.00   0.00   0.00   0.00   0.18   0.09   0.00   0.00   0.00
0.00   0.00   0.00   0.00   0.00   0.00   0.00   0.00   0.00   0.00   0.00
...
```

6-hour Accumulated Precipitation Observed Files:
Directory: <import directory>/rfc_pobs06
File Naming Convention: <MEFP locationId>.pobs06
Example: NFDC1HUF.pobs06
Header:
The header of the file is optional and specifies properties of the data. The properties are provided in lines of this format:

```
<key> = <value>
```

If the header exists, it must be closed with an end line:

```
end
```

The following properties are valid (default values used if the property is not provided are specified in parentheses:

- units: The units of the measurements, either “in” or “mm”. (default: “in”).
- iyr: The initial year of data available, a four digit year. (default: determined from the data in the file)
- iyr: The last year of data available, a four digit year. (default: determined from the data in the file)

Example:

```
units = in
end
```

Data :
Every line of the file must specify one day of observed data:

```
<T0: YYYYMMDD> <####.##><####.##><####.##><####.##>
```

The first field is the observation day (YYYYMMDD format) of the last value of the line (or the end of the 24-hour period). It is followed by one space. Then, starting at the 10th character, four observed values are provided that are 7 characters wide allowing for two decimal places. The observation times of the four values must be 18Z (of the day preceding the observation day shown in field 1), 0Z, 6Z, and 12Z.

In the last line of the example below, the value 0.08 in was observed at 18Z on 1/3/2000, while 0.06 in was observed on 12Z of 1/4/2000.

Example:

```
20000101  -99.99 -99.99 -99.99 -99.99
20000102  -99.99 -99.99 -99.99 -99.99
20000103  -99.99 -99.99 -99.99 -99.99
20000104    0.08   0.03   0.00   0.06
```

Full File Example:

```
units = in
end
20000101  -99.99 -99.99 -99.99 -99.99
20000102  -99.99 -99.99 -99.99 -99.99
20000103  -99.99 -99.99 -99.99 -99.99
20000104    0.08   0.03   0.00   0.06
...
```

24-hour Minimum and Maximum Temperature Forecast Files:
Directory: <import directory>/rfc_tfcst
File Naming Convention: <MEFP locationId>.rfctmnfcst (minimum temperature)
<MEFP locationId>.rfctmxfcst (maximum temperature)
Examples: NFDC1HUF.rfctmnfcst, NFDC1HUF.rfctmxfcst
Header:
Each file includes a two line header that must obey strict column widths:

Header Line 1:
Columns 1 – 8: location identifier (ex: APCSYN)
9 – 10: spaces
11 – 15: number of RFC forecast days (ex: "   7").
16 – 24: longitude (4 decimal places and do not include a negative sign; ex: "122.4700").
25 – 33: latitude (4 decimal places; ex: " 38.5000").
34 – 39: elevation (missing is allowed; ex: "-9999").
40 – 41: spaces
42 – 61: location descriptive name (ex: "NAPA SYN TEMP   ").
Header Line 2:
1 – 8: Begin date of data (YYYYMMDD; ex: "20010101")
9 – 10: spaces
11 – 18: End date of data (YYYYMMDD; ex: "20100222")

Example:
```
APCSYN        7 122.4700  38.5000 -9999  NAPA SYN TEMP
20010101  20100222
```

Data:
Every line of the file must specify a single forecast time series (<> are used to separate the fields; do not include them in the text):

```
<T0: YYYYMMDD><####.#><####.#><####.#>...<####.#>
```

The first field is the T0 forecast day (YYYYMMDD format). Then, starting at the 9th character, forecast values are provided that are 6 characters wide allowing for one decimal place. Each line must specify a number of values equal to the number of RFC forecast days specified in the header. The values listed are 24-hour minimum or maximum values, where a day is defined as 12Z – 12Z.

Example:
```
20010101 -12.0  -9.3 -11.2  -9.3  -9.1  -9.8  -9.6
20010102 -10.5 -10.8 -10.2  -8.0  -7.7  -6.7  -8.5
```

Full File Example:

```
APCSYN        7 122.4700  38.5000 -9999  NAPA SYN TEMP
20010101  20100222
20010101 -12.0  -9.3 -11.2  -9.3  -9.1  -9.8  -9.6
20010102 -10.5 -10.8 -10.2  -8.0  -7.7  -6.7  -8.5
...
```

24-hour Minimum and Maximum Temperature Observed Files:
Directory: <import directory>/rfc_tobs
File Naming Convention: <MEFP locationId>.rfctmnobs (minimum temperature)
<MEFP locationId>.rfctmxobs(maximum temperature)
Examples: NFDC1HUF.rfctmnobs, NFDC1HUF.rfctmxobs
Header:
Each file includes a two line header that must obey strict column widths:

Header Line 1:
Columns 1 – 8: location identifier (ex: APCSYN)
9 – 10: spaces
11 – 19: longitude (4 decimal places and do not include a negative sign; ex: "122.4700").
20 – 28: latitude (4 decimal places; ex: " 38.5000").
29 – 34: elevation (missing is allowed; ex: "-9999").
35 – 36: spaces
37 – 56: location descriptive name (ex: "NAPA SYN TEMP   ").
Header Line 2:
1 – 8: Begin date of data (YYYYMMDD; ex: "20010101")
9 – 10: spaces
11 – 18: End date of data (YYYYMMDD; ex: "20100222")

Example:
```
APCSYN     122.4700  38.5000 -9999  NAPA SYN TEMP
20010101  20100222
```

Data:
Every line of the file must specify a single forecast time series (<> are used to separate the fields; do not include them in the text):

```
<T0: YYYYMMDD><####.#>
```

The first field is the observation day (YYYYMMDD format), which is the day marking the end of the observed 12Z – 12Z period. Then, starting at the 9^{th} character, one observed value is provided that is 6 characters wide allowing for one decimal place.

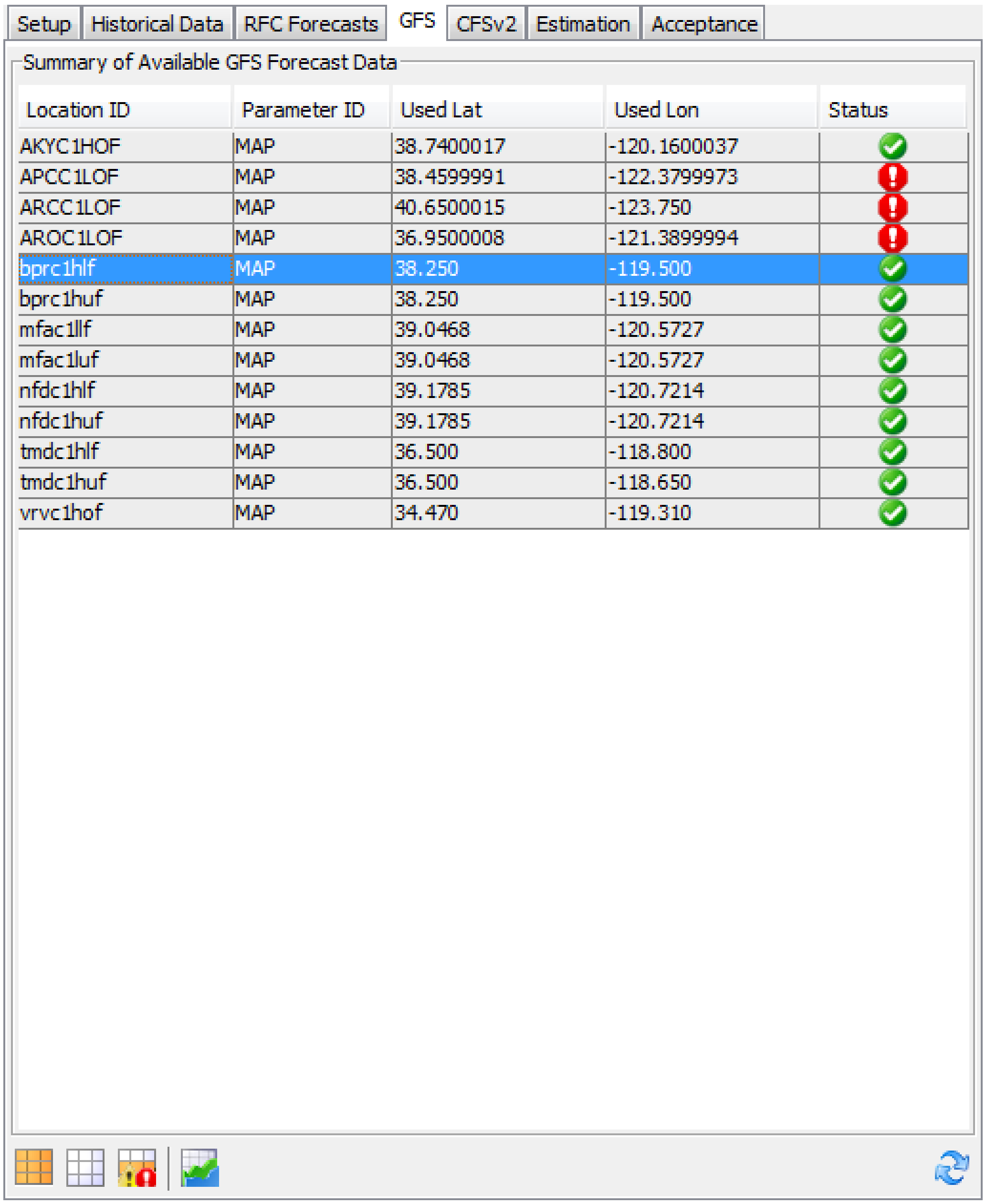
Example:
```
20010101 -13.8
20010102 -13.2
20010103 -14.9
```

Full File Example:

```
APCSYN     122.4700  38.5000 -9999  NAPA SYN TEMP
20010101  20100222
20010101 -13.8
20010102 -13.2
20010103 -14.9
...
```

3.8 GFS Subpanel

The **GFS Subpanel** of the **Estimation Steps Panel**, shown in Figure 9, is used to perform Step 4 of the parameter estimation procedure in Section 3.2.3: acquire GFS reforecast data files. For precipitation, the GFS reforecast files specify the ensemble mean of the 24-hour (12Z – 12Z) accumulated precipitation time series for 14-day forecasts. For temperature, the GFS reforecast files specify the ensemble mean of the 24-hour (12Z – 12Z) minimum and maximum temperature time series for 14-day forecasts.



Location ID	Parameter ID	Used Lat	Used Lon	Status
AKYC1HOF	MAP	38.7400017	-120.1600037	
APCC1LOF	MAP	38.4599991	-122.3799973	
ARCC1LOF	MAP	40.6500015	-123.750	
AROC1LOF	MAP	36.9500008	-121.3899994	
bprc1hlf	MAP	38.250	-119.500	
bprc1huf	MAP	38.250	-119.500	
mfac1llf	MAP	39.0468	-120.5727	
mfac1luf	MAP	39.0468	-120.5727	
nfdc1hlf	MAP	39.1785	-120.7214	
nfdc1huf	MAP	39.1785	-120.7214	
tmdc1hlf	MAP	36.500	-118.800	
tmdc1huf	MAP	36.500	-118.650	
vrvc1hof	MAP	34.470	-119.310	

Figure 9: The **GFS Subpanel** of the **Estimation Steps Panel**.

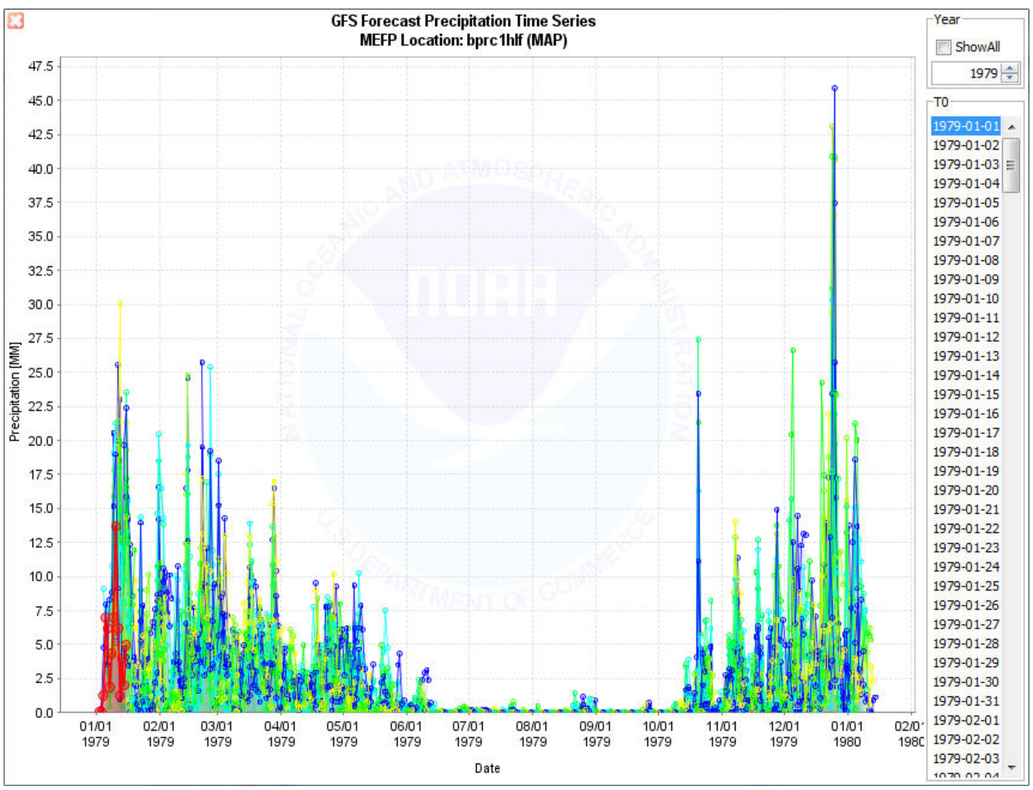
3.8.1 Components

- **Summary of Available GFS Forecast Data Table**: Allows the user to select the MEFP locations for which to perform the step and view for which locations the step has been performed. It is a **Generic Summary Table** (Section 3.2.5.1) and includes all the standard buttons.
- **View Button**: Displays diagnostics for a single MEFP location selected in the **Summary of Available GFS Forecast Data Table**.

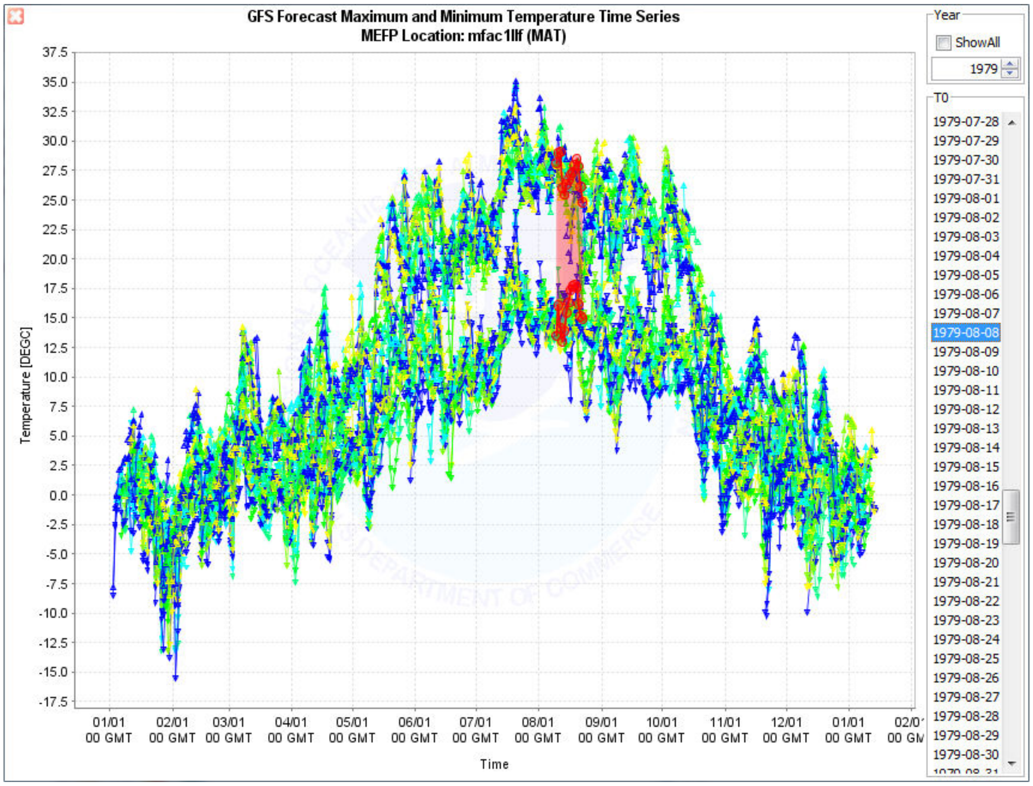
3.8.2 Diagnostics

The current diagnostic displays provided when the **View Button** is clicked for the GFS reforecast data are designed primarily for data quality control. Both diagnostics display the reforecast time series only (no observations are included). The charts are displayed within a standard **Archive/Reforecast Time Series Diagnostic Display Panel** (Section 3.2.5.3).

For precipitation, the displayed forecast and observed time series are 6-hour accumulated precipitation time series. For example:



For temperature, the displayed forecast and observed time series are 24-hour minimum and maximum temperatures. The area between corresponding minimum and maximum values is shaded only for selected (i.e., emphasized) time series. For example:



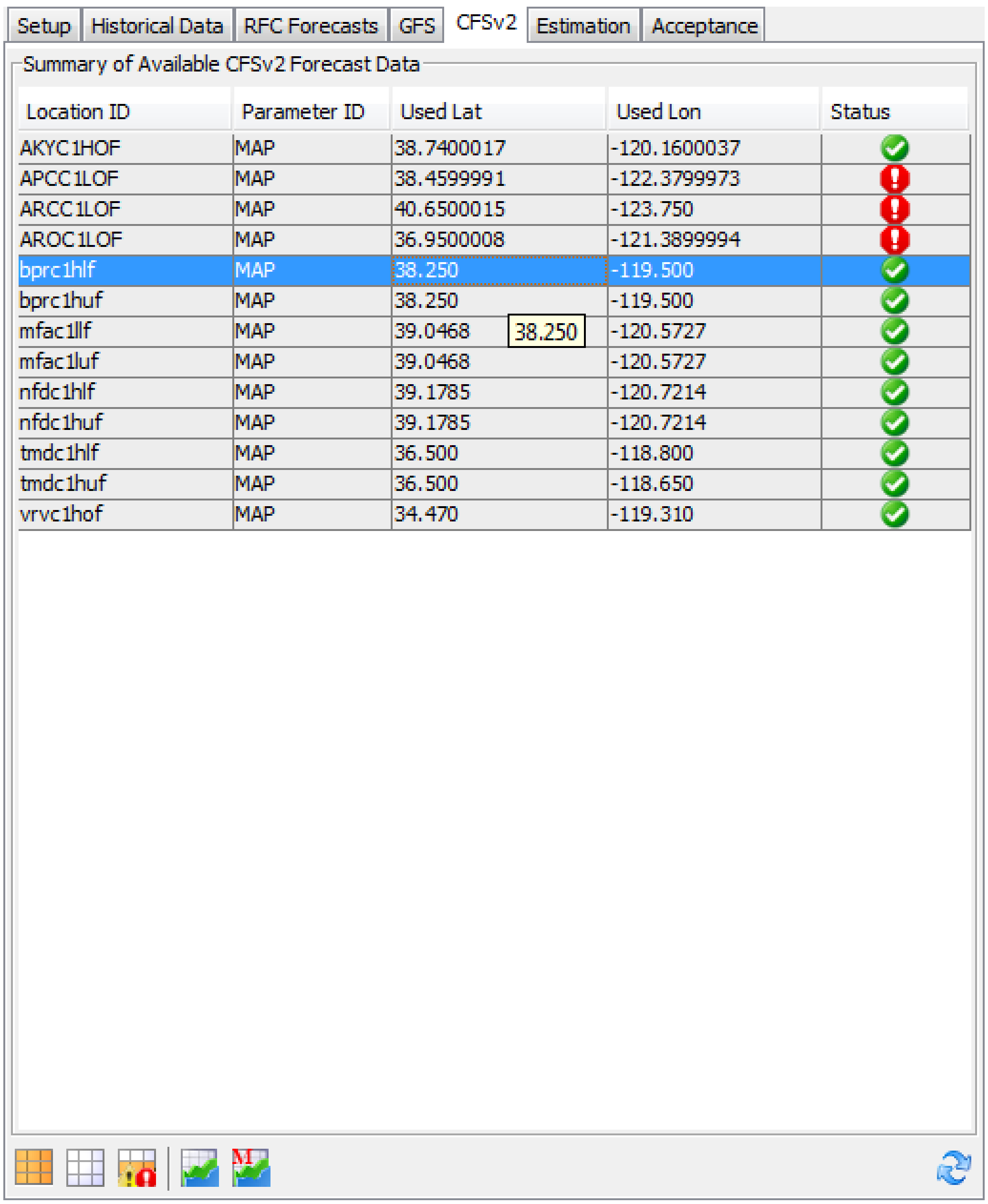
3.8.3 Usage

3.8.3.1 Downloading GFS Reforecast Files

GFS reforecast data is acquired via SFTP from IP address 165.92.28.30, which is an SFTP server behind the AWIPS firewall and should be accessible to all chps and lx machines at an RFC. To download files, select one or more locations from the **Summary of Available GFS Forecast Data Table** and click on the **Perform Step Button**. See Section 3.4.2.1 for details on how to perform a step using the **Perform Step Button** of the **Estimation Steps Panel**.

3.9 CFSv2 Subpanel

The **CFSv2 Subpanel** of the **Estimation Steps Panel**, shown in Figure 10, is used to perform Step 5 of the parameter estimation procedure in Section 3.2.3: acquire CFSv2 reforecast data files. There are two types of reforecast files: submonthly files specifying 60 days of 6-hour single-valued forecasts, and monthly files specifying 30-day accumulated values for the entire lagged ensemble used for each reforecast time. For precipitation, the CFSv2 reforecast files specify the time series of accumulated precipitation amounts (6-hourly for submonthly files, 30-days for monthly files). For temperature, the CFSv2 reforecast files specify the time series of minimum and maximum temperatures (6-hourly for submonthly files, 30-day for monthly files).



Location ID	Parameter ID	Used Lat	Used Lon	Status
AKYC1HOF	MAP	38.7400017	-120.1600037	
APCC1LOF	MAP	38.4599991	-122.3799973	
ARCC1LOF	MAP	40.6500015	-123.750	
AROC1LOF	MAP	36.9500008	-121.3899994	
bprc1hlf	MAP	38.250	-119.500	
bprc1huf	MAP	38.250	-119.500	
mfac1llf	MAP	39.0468	-120.5727	
mfac1luf	MAP	39.0468	-120.5727	
nfdc1hlf	MAP	39.1785	-120.7214	
nfdc1huf	MAP	39.1785	-120.7214	
tmdc1hlf	MAP	36.500	-118.800	
tmdc1huf	MAP	36.500	-118.650	
vrvc1hof	MAP	34.470	-119.310	

Figure 10: The **CFSv2 Subpanel** of the **Estimation Steps Panel**.

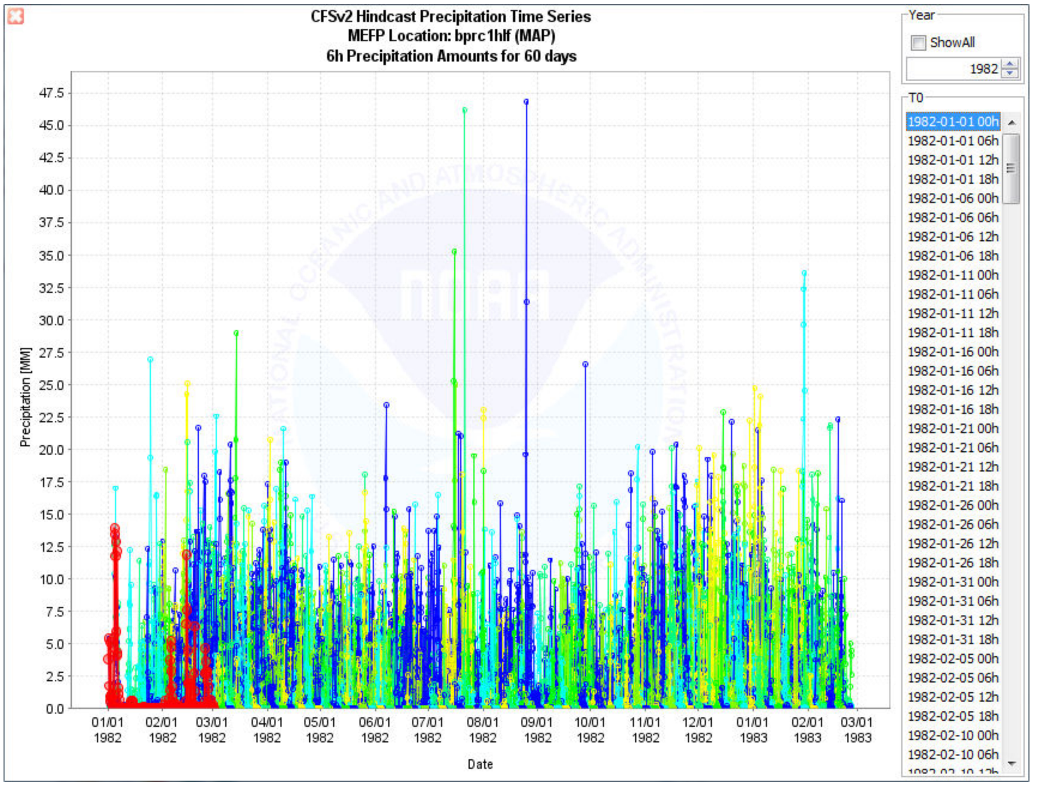
3.9.1 Components

- **Summary of Available CFSv2 Forecast Data Table**: Allows the user to select the MEFP locations for which to perform the step and view for which locations the step has been performed. It is a **Generic Summary Table** (Section 3.2.5.1) and includes all the standard buttons.
- **View Submonthly Button**: Displays submonthly diagnostics for a single MEFP location selected in the **Summary of Available CFSv2 Forecast Data Table**.
- **View Monthly Button**: Displays monthly diagnostics for a single MEFP location selected in the **Summary of Available GFS Forecast Data Table**.

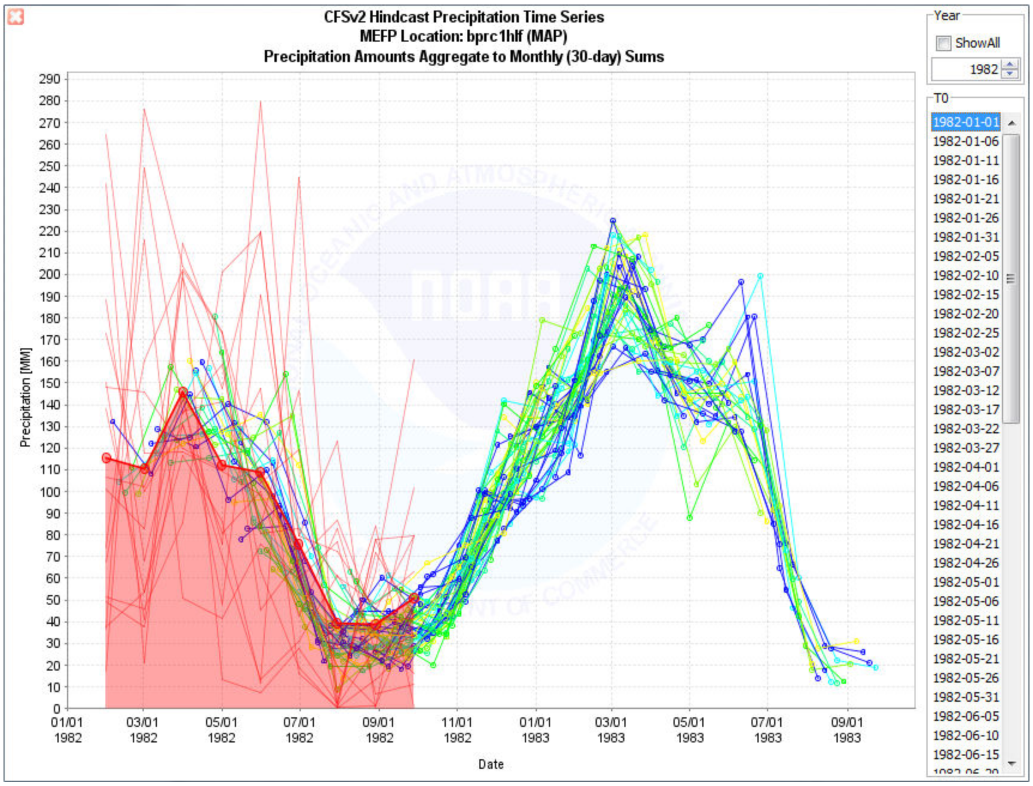
3.9.2 Diagnostics

The current diagnostic displays provided when the **View Button** is clicked for the CFSv2 reforecast data are designed primarily for data quality control. All diagnostics display the reforecast time series only (no observations are included). The diagnostic charts are displayed within a standard **Archive/Reforecast Time Series Diagnostic Display Panel** (Section 3.2.5.3).

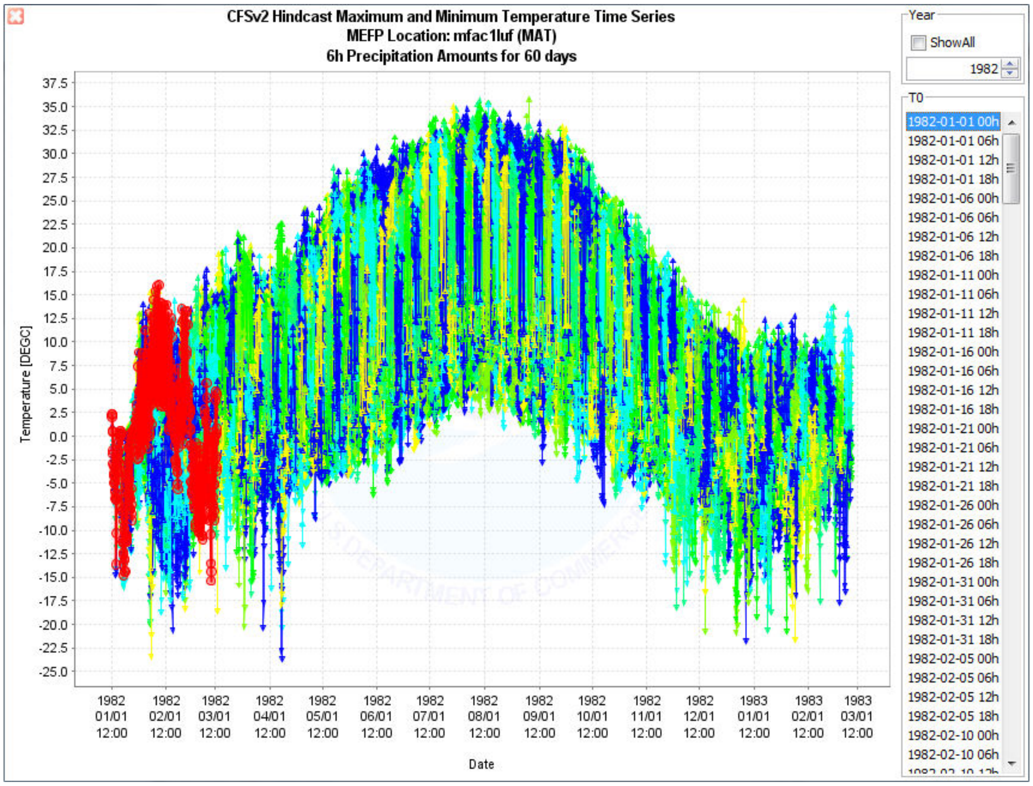
For precipitation submonthly diagnostics, the displayed forecast time series are 6-hour accumulated precipitation time series. For example:



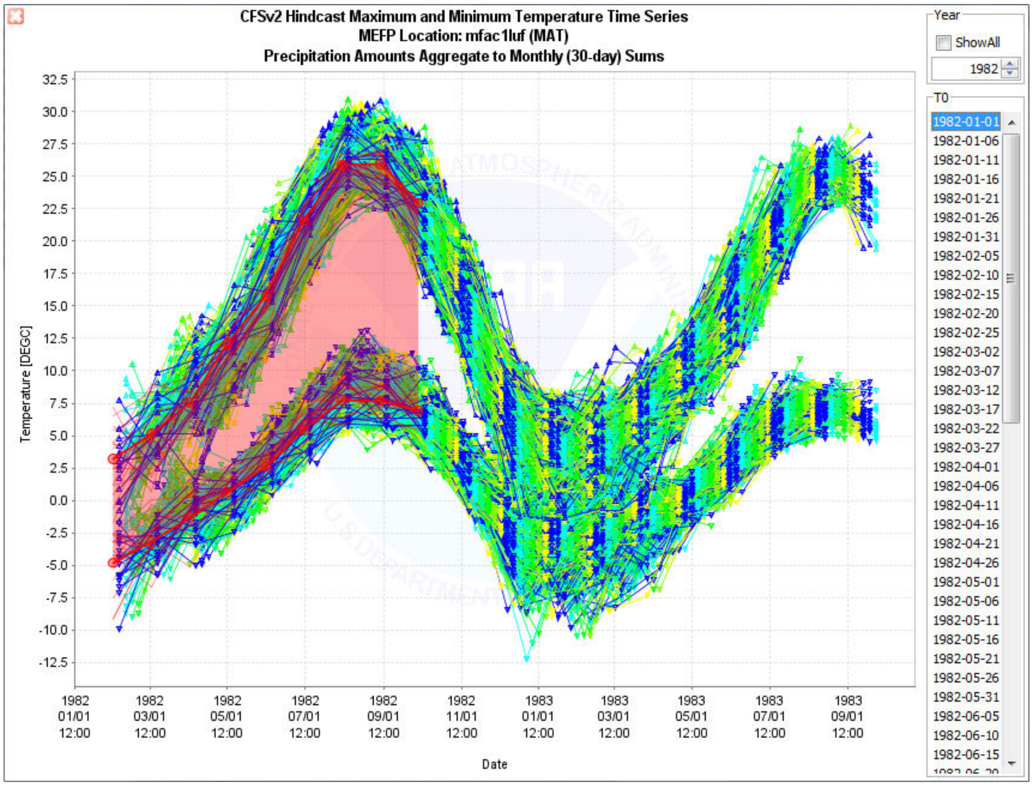
For monthly diagnostics, the displayed forecast time series are 30-day accumulated precipitation time series. For a selected (i.e., emphasized) reforecast T0, the entire lagged ensemble is shown as thin red lines and the ensemble mean is shown as a thick read line with a shaded region. For example:



For temperature, the displayed forecast time series are 6-hour minimum and maximum temperature time series. The area between corresponding minimum and maximum values is shaded only for selected (i.e., emphasized) time series. For example:



For monthly diagnostics, the displayed forecast time series are 30-day accumulated minimum and maximum temperature time series. For a selected (i.e., emphasized) reforecast T0, the entire lagged ensembles for both minimum and maximum are shown as thin red lines and the ensemble means are shown as a thick red lines with a shaded region between corresponding minimum and maximum ensemble mean values. For example:



3.9.3 Usage

3.9.3.1 Downloading CFSv2 Reforecast Files

CFSv2 reforecast data is acquired via SFTP from IP address 165.92.28.30, which is an SFTP server behind the AWIPS firewall and should be accessible to all chps and lx machines at an RFC. To download files, select one or more locations from the **Summary of Available CFSv2 Forecast Data Table** and click on the **Perform Step Button**. See Section 3.4.2.1 for details on how to perform a step using the **Perform Step Button** of the **Estimation Steps Panel**.

3.10 Estimation Subpanel

The **Estimation Subpanel** of the **Estimation Steps Panel**, shown in Figure 11, is used to perform Step 6 of the parameter estimation procedure in Section 3.2.3: estimate parameters. The panel includes two subpanels: the **Locations Summary Subpanel** and **Estimation Options Subpanel**. Each subpanel and components are described below.

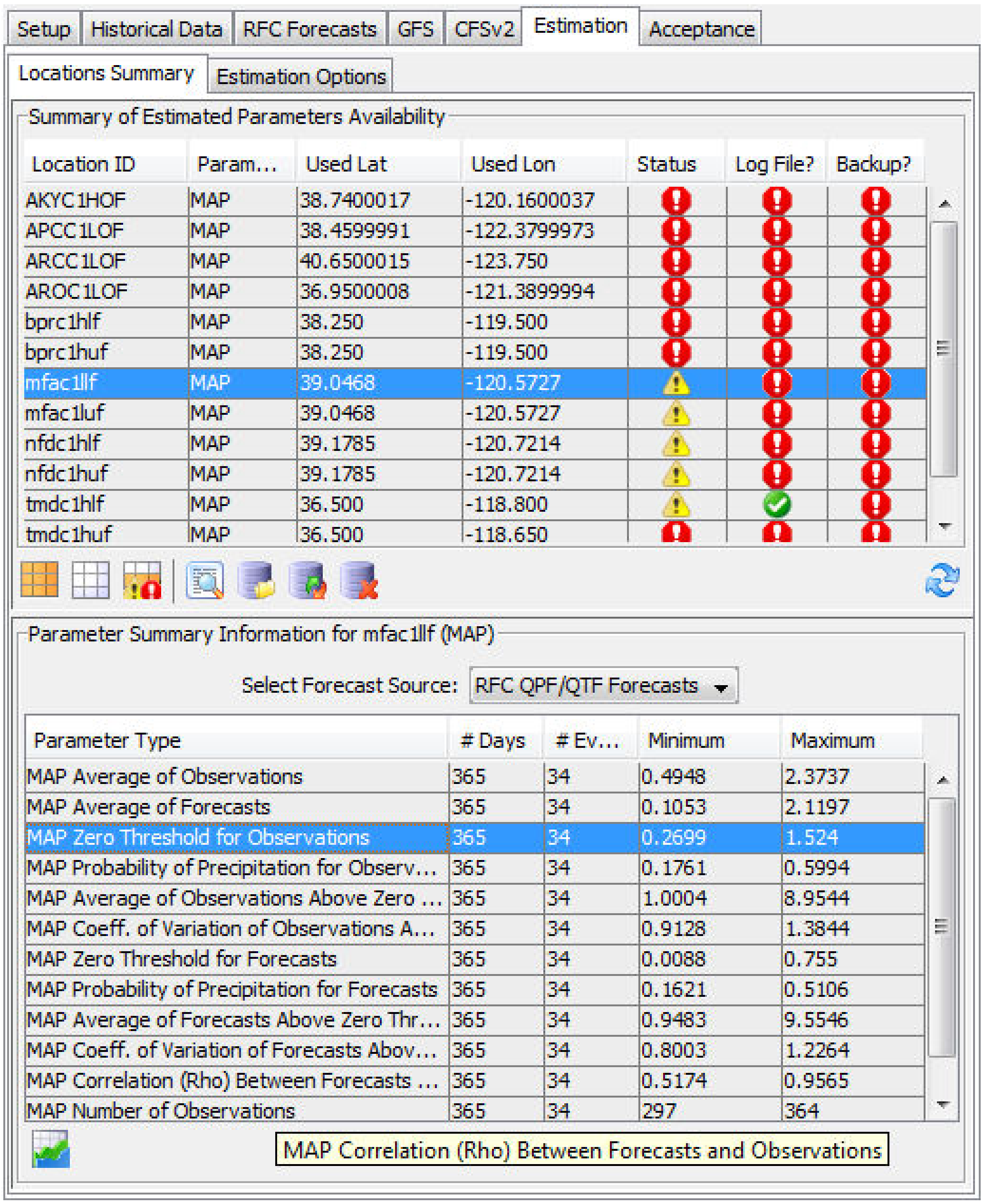


Figure 11: The **Estimation Subpanel** of the **Estimation Steps Panel**.

3.10.1 Locations Summary Subpanel

The **Locations Summary Subpanel**, shown in Figure 11, summarizes the status of estimation for all locations for the active estimation data type and allows the user view log files, delete parameters, backup parameters, restore parameters, and select diagnostics to display.

3.10.1.1 Parameter File Backups

The MEFPPE allows for *one* set of backup parameters per location. Whenever estimation is performed for the selected MEFP locations, if parameters have already been estimated for those locations, they will be backed-up, while any parameters that were backed-up will be discarded. Those newly backed-up parameters can later be restored if the new active parameters just estimated prove to be less desirable.

3.10.1.2 Components

- **Summary of Estimated Parameters Availability Table**: Allows the user to select the MEFP locations for which to perform the step and view for which locations the step has been performed. It is a **Generic Summary Table** (Section 3.2.5.1) and includes all the standard buttons. Furthermore, it includes two additional columns: the 'Log File?' column indicates if a log file is present for the current estimated parameters; the 'Backup?' column indicates if backup parameters exist for the selected location.
- **View Log File Button**: Allows the user to view the contents of the log file for the estimated parameters. It opens up an **Estimation Log File Dialog** displaying the contents.

NOTE: For Development Release 1, do not view log files for estimated parameters for temperature. This is due to memory limitations and the vast amount of information in the log files. Instead, the log files can be viewed directly under Models/hefs/mefppeRunArea/parameters/logFiles.

- **Load Parameters Button**: Click to load parameters for one selected location from the **Summary of Estimated Parameters Availability Table**. Upon loading, the **Parameters Summary Information Table** (below) will be updated to reflect the loaded parameters.
- **Restore Parameters Button**: Click to restore backup parameters for the selected MEFP locations. For the selected locations, the active and backed-up parameters will be swapped, making the backup parameters active and vice-versa. A **Continue? Dialog** will be opened allowing the user to confirm the restore.
- **Remove Parameters Button**: Click to remove the active parameters. A **Backup Parameters? Dialog** will open asking if the user wants to make the parameters backup parameters. If **Yes** is clicked, the parameters are backed-up. If **No** is clicked, the parameters are discarded. If **Cancel** is clicked, the remove is not performed.
- **Select Forecast Source Choice Box**: Allows the user to select the forecast source for which to view parameter summary information within the **Parameter Summary Information Table** (below).

- **Parameter Summary Information Table**: Displays the parameters loaded from the parameter files. The table allows for multiple selections. The following columns are provided:
    - ‘Parameter Type’: A descriptive name of the parameter.
    - ‘# Days’: The number of days of the year for which parameters were found.
    - ‘# Events’: The number of canonical events for which parameters were found.
    - ‘Minimum’: The smallest overall value found for all days of the year and canonical events for the corresponding parameter.
    - ‘Maximum’: The largest overall value found for all days of the year and canonical events for the corresponding parameter.
-  **View Button**: Click to view the parameters for selected rows from the **Parameter Summary Information Table**.

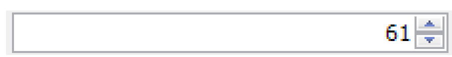
3.10.2 Estimation Options Subpanel

The **Estimation Options Subpanel** allows for users to specify options for the parameter estimation algorithm. There are three groups of parameters:

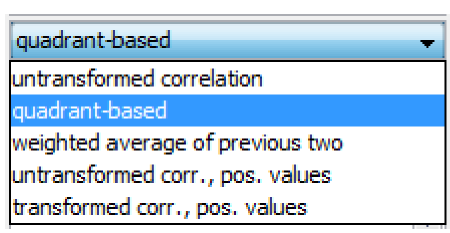
1. General options used to estimate parameters for the implicit precipitation treatment algorithm.
2. Parameters specific to the explicit precipitation treatment (EPT) algorithm
3. Source-specific parameters.

Each is described below, providing a screenshot of the associated components.

All numerical parameters are edited using text fields with spinners; for example:



The number can be edited by clicking and typing or by clicking on the up or down arrows. All choice-based parameters are edited via choice boxes; for example:



3.10.2.1 General Options

The general options for the MEFP parameter estimation algorithm used to estimate the parameters of the implicit precipitation treatment algorithm are shown in Figure 12. All options are considered advanced, meaning that it is recommended users change the options only after gaining significant experience with the software. Advanced options are described in Section 2.

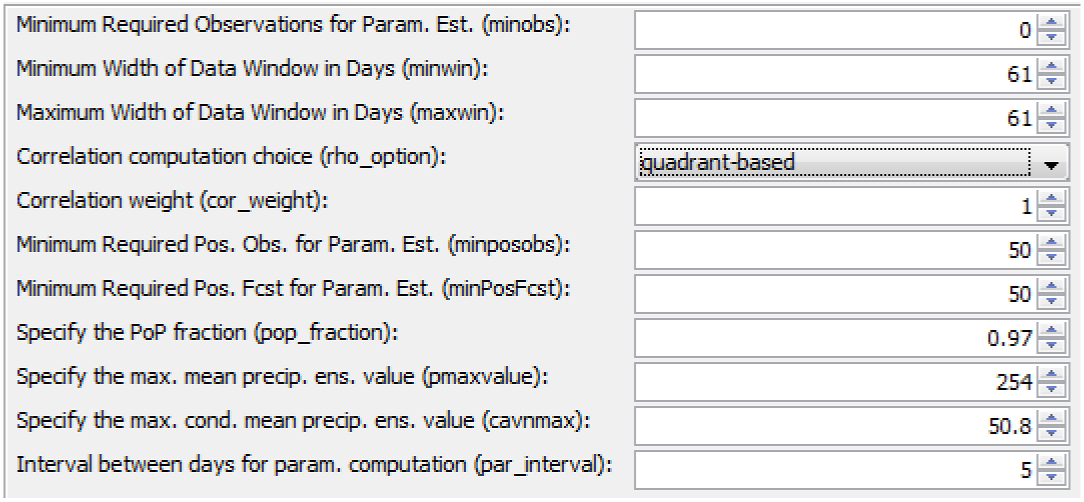


Figure 12: General options of the MEFP parameter estimation algorithm for estimating the parameters of the implicit precipitation treatment algorithm.

3.10.2.2 EPT Options

The options of the MEFP parameter estimation algorithm used to estimate parameters of the explicit precipitation treatment algorithm are shown in Figure 13. Parameters are estimated only if the **Estimate EPT Parameter Checkbox** is checked. All options are considered advanced, meaning that it is recommended users change the options only after gaining significant experience with the software. Advanced options are described in Section 2.

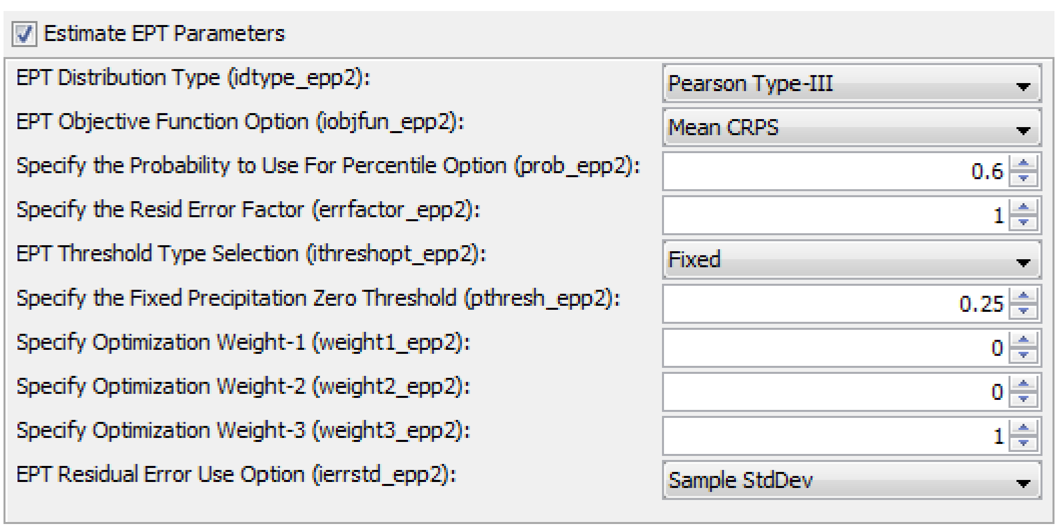


Figure 13: Options of the MEFP parameter estimation algorithm for estimating the parameters of the explicit precipitation treatment (EPT) algorithm.

3.10.2.3 Source Parameters

All forecast sources have source-specific options that are used in the estimation of parameters for both the implicit precipitation treatment algorithm and explicit precipitation treatment algorithm. For example, the components to edit the source specific options for GFS are shown in Figure 14. Distribution parameters, used for all forecast sources, are considered advanced, meaning that it is recommended users change the options only after gaining significant experience with the software. Advanced options are described in Section 2.

The first three options are not advanced and are common to all forecast sources, though the exact names may differ slightly:

- Number of <Source> Days: The number of forecast lead days for which to estimate parameters. This number must not exceed the number of forecast lead days available:
    - RFC QPF/QTF: 5 days for QPF, 7 days for QTF
    - GFS: 14 days
    - CFSv2: 270 days
- Initial Year of Parameter Estimation: The first year of data included in parameter estimation. The data available for the forecast sources is as follows:
    - RFC QPF/QTF: Varies by RFC
    - GFS: 1979 - 2006
    - CFSv2: 1982 - 2010
- Last Year of Parameter Estimation: The last year of data included in parameter estimation.

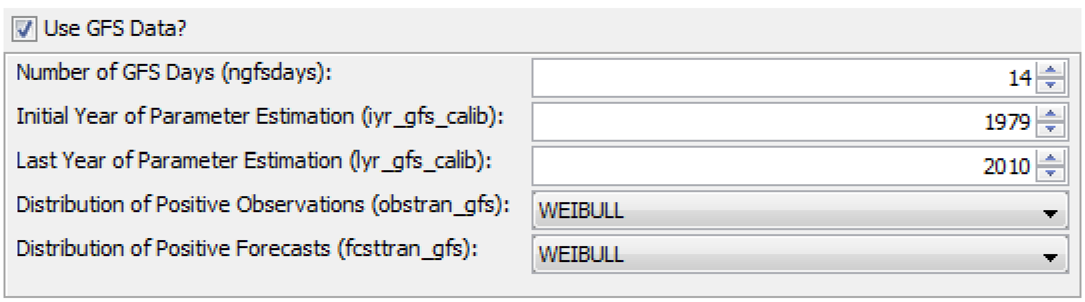
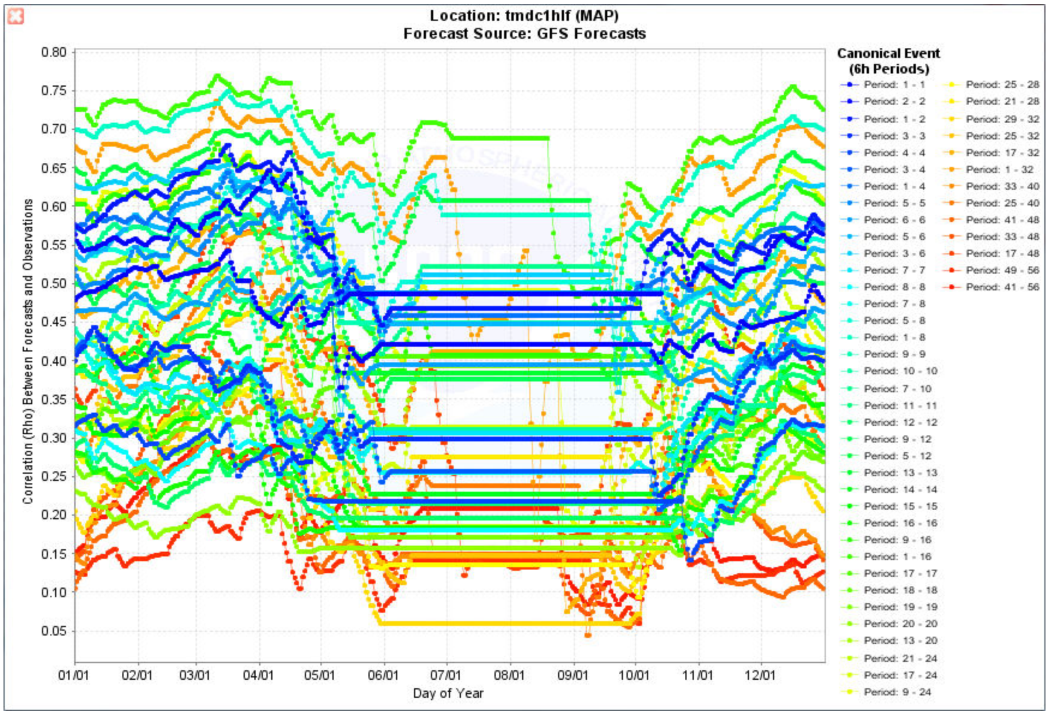


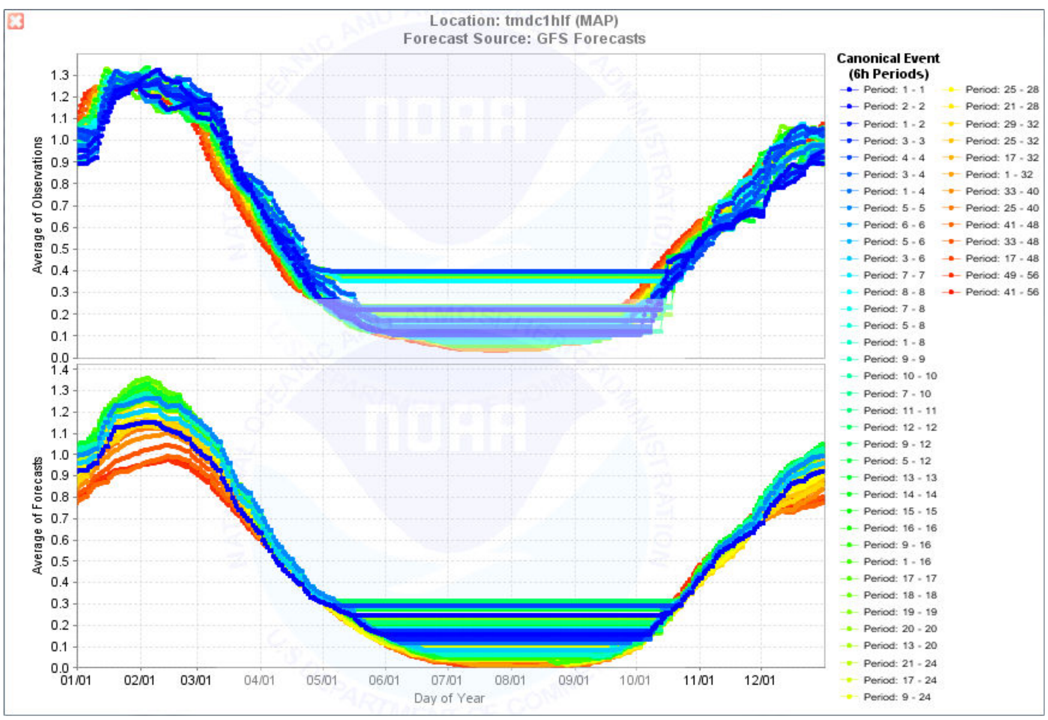
Figure 14: Source specific options of the MEFP parameter estimation algorithm.

3.10.3 Diagnostics

The current diagnostic display provided when the **View Button** is clicked for the estimated parameters are designed primarily for data viewing and quality control. The diagnostic displays parameters selected within the **Parameter Summary InformationTable**. The parameter values are displayed against the day of the year (1 – 365) and a series is included for each applicable canonical event. Because the legend can become large, it is recommended that the **Diagnostics Display Panel** be undocked and expanded prior to viewing.

Provided below are two examples of diagnostic displays. The first is for a single parameter (the correlation coefficient between the forecasts and observations) while the second is for multiple parameters (the average of the observations and forecasts).





The diagnostics should be viewed relative to user experience with the locations involved. The following are basic tips for checking parameter diagnostics:

1. Are parameter values computed for observations in line with what is expected for the location involved?
2. Does the seasonal pattern for the observed parameter values appear reasonable?
3. Does the seasonal pattern for the forecast parameter values follow that of the observed?

4. Does the curve pattern for the correlation coefficients look right for the forecast areas? For example, for test basins in MARFC, the correlation between the observed and the forecast values were higher during winter than summer.

3.10.4 Usage

3.10.4.1 Estimating Parameters

To estimate parameters, select one or more MEFP locations from the **Summary of Estimated Parameters Availability Table** and click on the **Perform Step Button**. See Section 3.4.2.1 for details on how to perform a step using the **Perform Step Button** of the **Estimation Steps Panel**.

3.10.4.2 Loading Parameters and Viewing Diagnostics

To load parameters and view the diagnostics, do the following:

1. Select one MEFP location (row) from the **Summary of Estimated Parameters Availability Table** of the **Locations Summary Subpanel**.
2. Click on the **Load Parameters Button**. A progress dialog will be displayed indicating that parameters are being loaded. This may take a minute or two. After completion, the **Parameter Summary Information Table** will update displaying the parameters loaded.
3. Select the forecast source for which you want to view parameters from the **Select Forecast Source Choice Box**. If the table is empty, then no parameters were loaded.
4. Select the estimated parameters (rows) that you wish to view from the **Parameter Summary Information Table**.
5. Click on the **View Button**.

3.10.4.3 Backing-Up Parameters and Restoring Backup Parameters

To backup parameters, select one or more MEFP locations from the **Summary of Estimated Parameters Availability Table** and click on the **Remove Parameters Button**. When the **Backup Parameters? Dialog** opens, click on **Yes**. Upon completion, the 'Status' column of the **Summary of Estimated Parameters Availability Table** will display a indicating that no active parameters are available, while the 'Backup?' column will display a indicating that backup parameters are available.

To restore backup parameters, select one or more MEFP locations from the **Summary of Estimated Parameters Availability Table** and click on the **Restore Parameters Button**. When the **Continue? Dialog** opens, click on **Yes**. Upon completion, the 'Status' column of the **Summary of Estimated Parameters Availability Table** will display a or indicating that parameters are available, while the 'Backup?' column will display either a if there were active parameters when the button was clicked or if there were no active parameters.

3.11 Acceptance Subpanel

The **Acceptance Subpanel** of the **Estimation Steps Panel**, shown in Figure 15, is used to perform Step 7 of the parameter estimation procedure in Section 3.2.3: accept (zip) parameter files. The panel includes three subpanels: the **Edit Groups Subpanel**, the **Prepare By Location Subpanel**, and the **Prepare By Group Subpanel**. Each subpanel and components are described below.

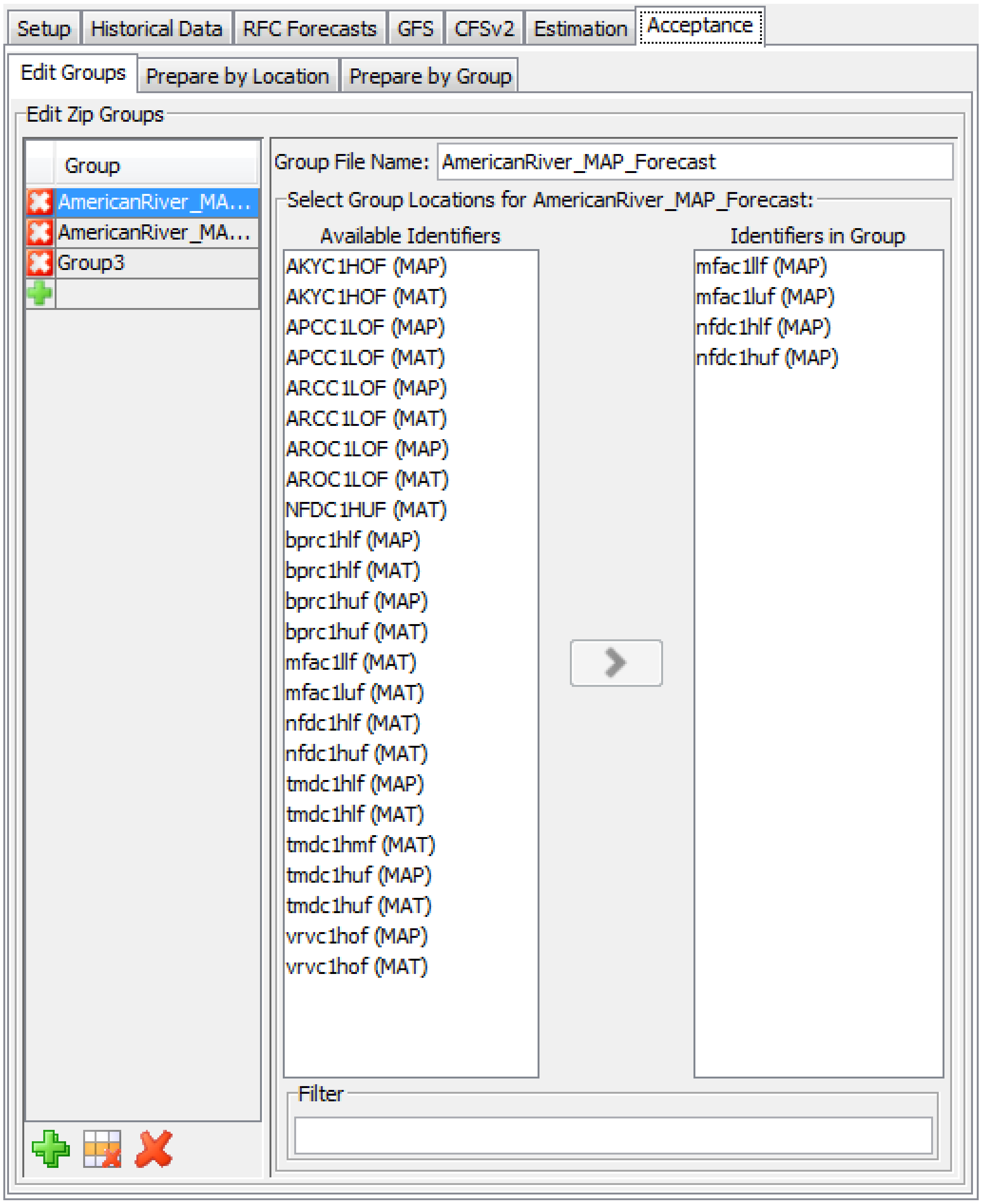


Figure 15: The **Acceptance Subpanel** of the **Estimation Steps Panel**.

3.11.1 About Parameter Zip Groups and Files

As defined in Section 3.1.1, parameter zip groups are groups of MEFP locations for which parameters are to be zipped together. The need for these groups is driven by CHPS and FEWS. Operationally, the MEFP can (and should) be executed for multiple locations at one time via a single model adapter execution, as defined in a single module configuration file. The parameters of such a run are exported for use by the model adapter via an exportDataSetActivity; for example:

```
<exportDataSetActivity>
    <moduleInstanceId>MEFP_AmericanRiver_MAP_Forecast</moduleInstanceId>
</exportDataSetActivity>
```

There can be only a single exportDataSetActivity within a module configuration file. Hence, the data set to which it refers must contain all parameter files necessary for executing the MEFP at all desired MEFP locations. Thus, parameter zip groups are required.

Furthermore, the moduleInstanceId referred to in the exportDataSetActivity must be a valid module instance id (i.e., defined in the ModuleInstanceDescriptors.xml file) and typically should match the id of the module in which the exportDataSetActivity is included. For example, the above exportDataSetActivity would be referred to in a module configuration file with the name MEFP_AmericanRiver_MAP_Forecast.xml. Thus the following rule of thumb is recommended:

Parameter zip groups should be named based on the id of the module configuration file that will be exporting it.

The zip file for a parameter zip group will be placed in the directory

Config/ModuleDataSetFiles/hefs

and its file name will match the group name with ".zip" added as a file extension.

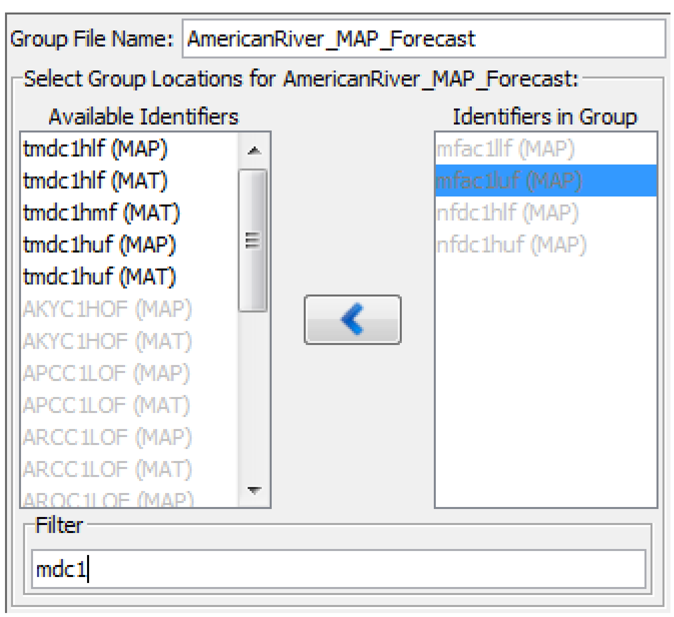
3.11.2 Edit Groups Subpanel

The **Edit Groups Subpanel** of the **Acceptance Subpanel**, shown in Figure 15, allows for editing parameter zip groups, specifying the names of the groups and the MEFP locations included in them. The subpanel operates independently of the active estimation data type, allowing for groups to contain parameters for both precipitation and temperature.

3.11.2.1 Components

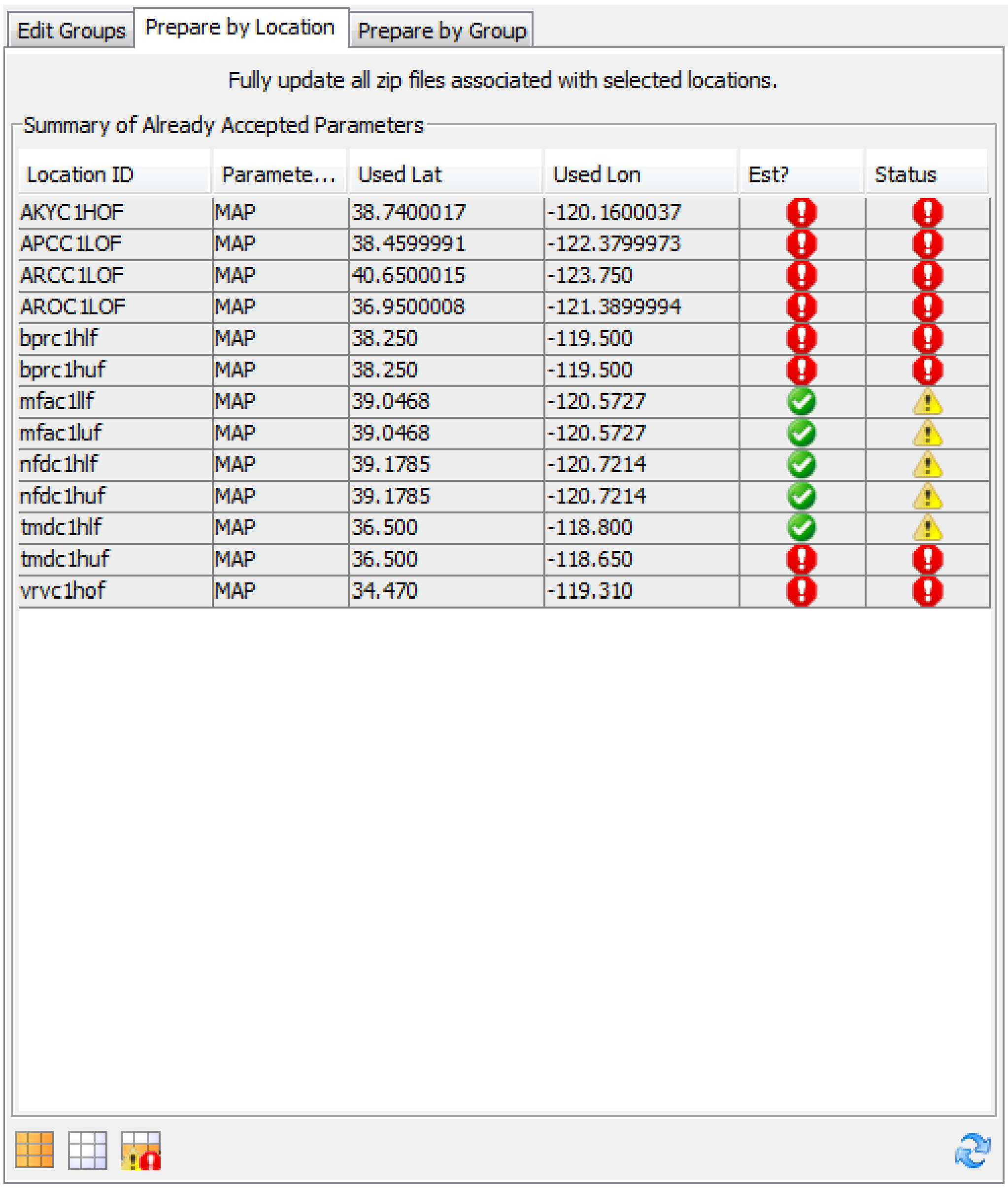
- **Group List**: Provides a list of the already defined parameter zip groups by name. The list includes an add/delete column (Section 3.2.5.2) allowing for groups to be added or removed. When a group is added by clicking , it will be given a default name of "Group#", where # is the smallest number (starting at 1) for which no group by that name exists (two groups may not have the same name). The **Group List** is selectable and, when a single group is selected, the **Group File Name Text Field**, **Available Identifiers List**, and **Identifiers in Group List** will be updated to match the selected group.

- **Create One Group for Each Location Button**: Click to add one group for each MEFP location. When clicked, a **Group Name Prefix Dialog** will open in which a name prefix to use for all created groups can be provided. The groups created with have the name "<prefix><MEFP locationId>_<MEFP parameterId>".
- **Remove Selected Groups Button**: Click to remove all groups selected in the **Group List**.
- **Remove All Groups Button**: Click to remove all groups in the **Group List**.
- **Group File Name Text Field**: Displays the name of the selected group in the **Group List**. The field can be modified in order to rename the group.
- **Available Identifiers List**: Displays the list of all MEFP locations by identifier that have not yet been added to the group selected in the **Group List**. This list allows for multiple selections.
- **Identifiers in Group List**: Displays a list of all MEFP locations by identifier that are included in the group selected in the **Group List**. This list allows for multiple selections.
- / **Add/Remove Button**: Click to add or remove the MEFP locations to or from the group selected in the **Group List**. If selections are made in the **Available Identifiers List**, when the button is clicked, those MEFP locations will be added to the group, being removed from the **Available Identifiers List** and added to the **Identifiers in Group List**. If selections are made in the **Identifiers in Group List**, when the button is clicked, those MEFP locations will be removed from the group, being removed from the **Identifiers in Group List** and added to the **Available Identifiers List**.
- **Filter Text Field**: Type in a value in order to sort the **Available Identifiers List** and the **Identifiers in Group List**. When text is typed into the field, the two lists will be sorted so that all MEFP location identifiers in the lists that include the filter text will be sorted to the top. Those MEFP location identifiers that do not include the filter text will be grayed-out (but are still selectable). For example:



3.11.3 Prepare by Location Subpanel

The **Prepare by Location Subpanel** of the **Acceptance Subpanel**, shown in Figure 16, allows for preparing estimation parameter zip files by MEFP location. Whenever a parameter zip file is updated for an MEFP location, the entire file will be updated. The subpanel only displays the MEFP locations for the active estimation data type.



Location ID	Paramete...	Used Lat	Used Lon	Est?	Status
AKYC1HOF	MAP	38.7400017	-120.1600037		
APCC1LOF	MAP	38.4599991	-122.3799973		
ARCC1LOF	MAP	40.6500015	-123.750		
AROC1LOF	MAP	36.9500008	-121.3899994		
bprc1hlf	MAP	38.250	-119.500		
bprc1huf	MAP	38.250	-119.500		
mfac1llf	MAP	39.0468	-120.5727		
mfac1luf	MAP	39.0468	-120.5727		
nfdc1hlf	MAP	39.1785	-120.7214		
nfdc1huf	MAP	39.1785	-120.7214		
tmdc1hlf	MAP	36.500	-118.800		
tmdc1huf	MAP	36.500	-118.650		
vrvc1hof	MAP	34.470	-119.310		

Figure 16: The **Prepare by Location Subpanel** of the **Acceptance Subpanel**.

3.11.3.1 Components

- **Summary of Already Accepted Parameters Table**: Allows the user to select the MEFP locations for which to perform the step (generate zip files) and view for which locations the step has been performed. It is a **Generic Summary Table** (Section 3.2.5.1) and includes all the standard buttons. The 'Status' column displays the status of the MEFP location relative to all group zip files that should contain the location's parameter files. An additional column is included before the 'Status' column called 'Est?'. That column indicates for which MEFP locations parameters have been estimated.

*NOTE: When the **Perform Step Button** is clicked, for each selected MEFP location, zip files associated with all parameter zip groups that the location is a part of will be fully updated.*

3.11.4 Prepare by Group Subpanel

The **Prepare by Group Subpanel** of the **Acceptance Subpanel**, shown in Figure 17, allows for preparing estimation parameter zip files by parameter zip group.

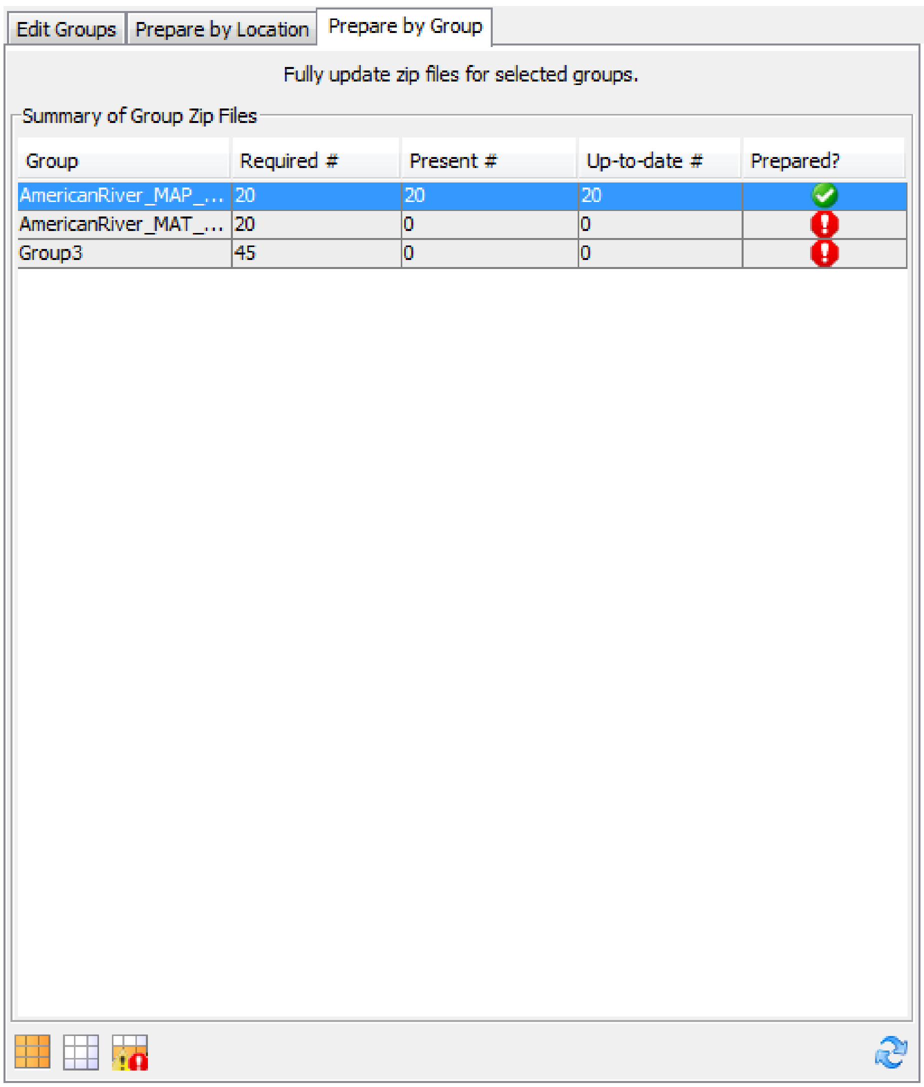


Figure 17: The **Prepare by Group Subpanel** of the **Acceptance Subpanel**.

3.11.4.1 Components

- **Summary of Group Zip Files Table**: Allows the user to select parameter zip groups for which to perform the step (generate zip files) and view for which groups the step has been performed. It is a **Generic Summary Table** (Section 3.2.5.1) and includes all the standard buttons, but different columns. The columns are as follows:
    - 'Group': The name of the group.
    - 'Required #': The number of parameter files that must be in the zip group for the zip file to be complete.
    - 'Present #': The number of parameter files found in the zip file.
    - 'Up-to-date #': The number of parameter files found in the zip file that are up-to-date relative to the estimated parameter files in the MEFP run area. A file is up to date if it is not older than the corresponding file in the MEFP run area.
    - 'Prepared?': A status column indicating if the zip file is present and up-to-date relative to all the parameter files to be contained within it.

*NOTE: When the **Perform Step Button** is clicked, for each selected parameter zip group, the zip file will be fully updated.*

3.11.5 Usage

3.11.5.1 Defining Parameter Zip Groups

To define a parameter zip group, make sure the **Edit Groups Subpanel** is active by clicking on its tab and do the following:

1. In the **Group List**, click on the icon to add a group.
2. In the **Group File Name Text Field**, modify the name as desired. The name of the group should match the name of the module in which the created module data set zip file will be exported. See Section 3.11.1.
3. Using the <ctrl> and <shift> keys, select rows in the **Available Identifiers List** for MEFP locations to add to the group. If needed, use the **Filter Text Field** to more easily find desired MEFP locations in the list.
4. Click on the **Add Button** to add the MEFP locations to the group. The selected MEFP location identifiers will move to the **Identifiers in Group List**.
5. Repeat Steps 3and 4 until all desired MEFP locations are in the group.

3.11.5.2 Generating Group Zip Files

Group zip files can be created and updated by either MEFP location or parameter zip group. Which method to use depends on how many MEFP location parameters must be zipped for and how many group zip files must be updated. For example, if the parameters for one MEFP location have been modified and that location is in multiple parameter zip groups all of which must be updated, then do the following:

1. Click on the **Prepare by Location Subpanel** tab.
2. Select the MEFP location from the **Summary of Already Accepted Parameters Table**.
3. Click on the **Perform Step Button**.

If the parameters for one or more MEFP locations have been modified and those locations are all in a one group, then do the following:

1. Click on the **Prepare by Group Subpanel** tab.
2. Select the group from the **Summary of Group Zip files Table**.
3. Click on the **Perform Step Button**.

In either case, see Section 3.4.2.1 for details on how to perform a step using the **Perform Step Button** of the **Estimation Steps Panel**.

3.12 Location Summary Panel

The **Location Summary Panel**, shown in Figure 18, summarizes the status of all steps to perform described in Section 3.2.3 for all MEFP locations. The panel also includes a **Select Type of Data for Estimation Choice Box** for selecting the active estimation data type, a **Goto Step Panel Button** to facilitate quickly navigating the **Estimation Steps Panel**, and a **Run All Steps Button** to allow for performing all steps for multiple selected locations.

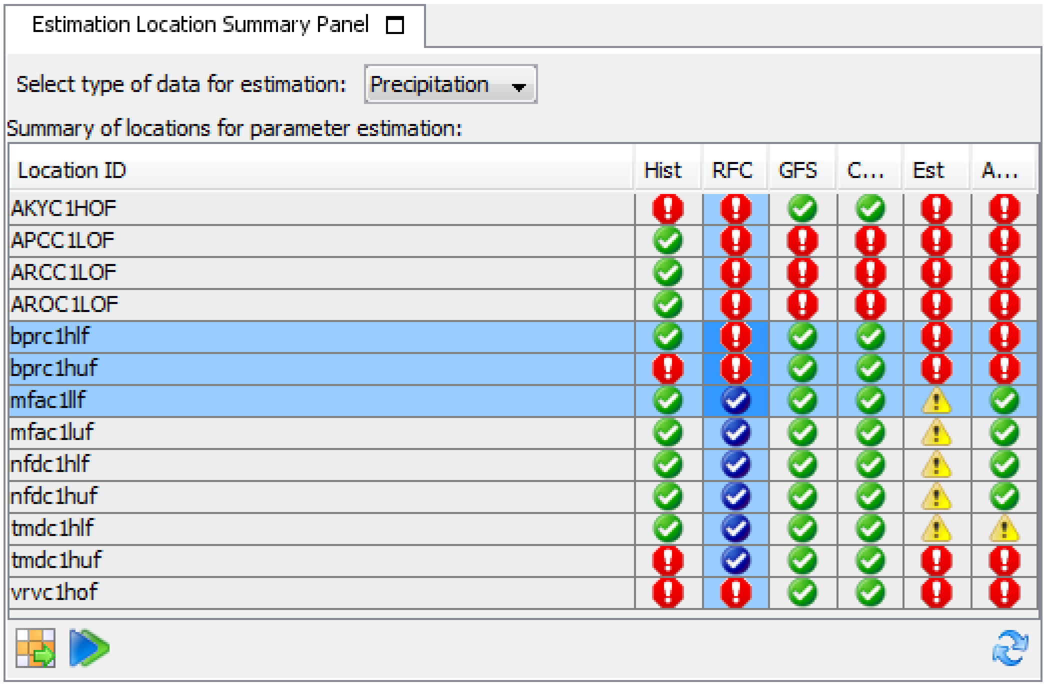


Figure 18: The **Location Summary Panel**.

3.12.1.1 Components

- **Select Type of Data for Estimation Choice Box**: Allows for selecting the active estimate data type, either "Precipitation" or "Temperature". When a change is made to the choice, many of the tables in the MEFPPE, including the **Summary of Locations for Parameter Estimation Table**, will update to reflect the change.
- **Summary of Locations for Parameter Estimation Table**: Summarizes the status of all steps to perform (except the setup step) for all MEFP locations for the active estimation data type. The columns are as follows:
    - 'Location ID': The location id of the MEFP location.
    - 'Hist': Displays the status of Step 2: process historical data and generate binary files.
    - 'RFC': Displays the status of Step 3: create RFC archive forecast data files.
    - 'GFS': Displays the status of Step 4: acquire GFS reforecast files.
    - 'CFSv2': Displays the status of Step 5: acquire CFSv2 reforecast files.
    - 'Est': Displays the status of Step 6: estimate parameters.
    - 'Accept': Displays the status of Step 7: accept (zip) parameter files.

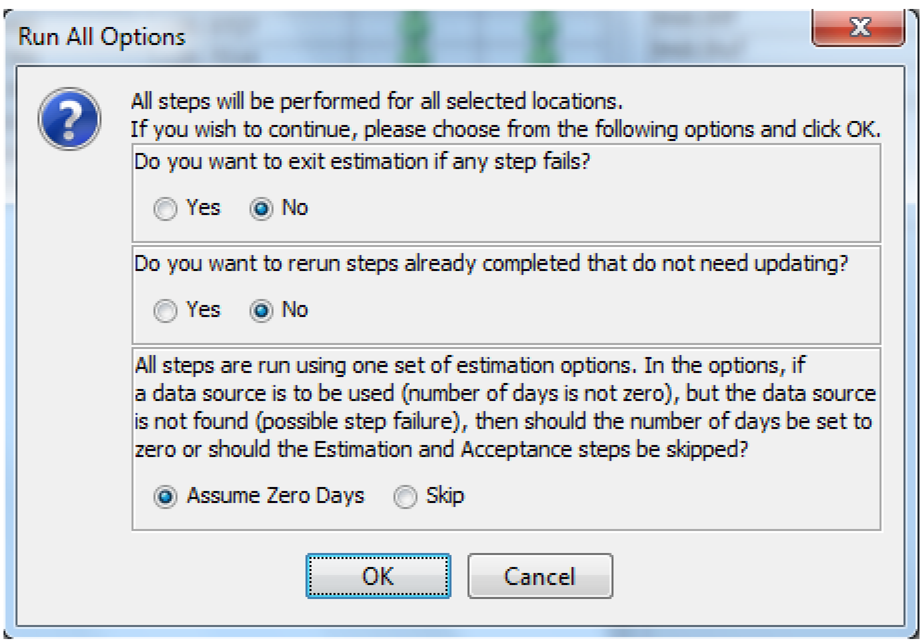
- **Goto Step Panel Button**: Click to make the step subpanel corresponding to the selected column of the **Summary of Locations for Parameter Estimation Table** active within the **Estimation Steps Panel**. Also, if appropriate, it selects rows of the **Generic Summary Table** for that subpanel for the MEFP locations selected within the **Summary of Locations for Parameter Estimation Table**. After clicking this button, the user should be able to click on the **Perform Step Button** to perform the step for all selected locations.
- **Run All Steps Button**: Click to run all of the steps for the selected MEFP locations. Upon clicking a **Run All Options Dialog** will be displayed. Click on **OK** to continue with the run all or **Cancel** to cancel it. See Section 3.12.2.1 for more details.
- **Refresh Button**: Click to refresh the status columns of the **Summary of Locations for Parameter Estimation Table**. This will also trigger a refresh of status columns in all subpanels of the **Estimation Steps Panel**.

3.12.2 Usage

3.12.2.1 Running All Steps for Multiple MEFP Locations

To perform all steps for desired MEFP locations, do the following:

1. Select the rows for the desired MEFP locations from the **Summary of Locations for Parameter Estimation Table**.
2. Click on the **Run All Steps Button**. A **Run All Options Dialog** will open:

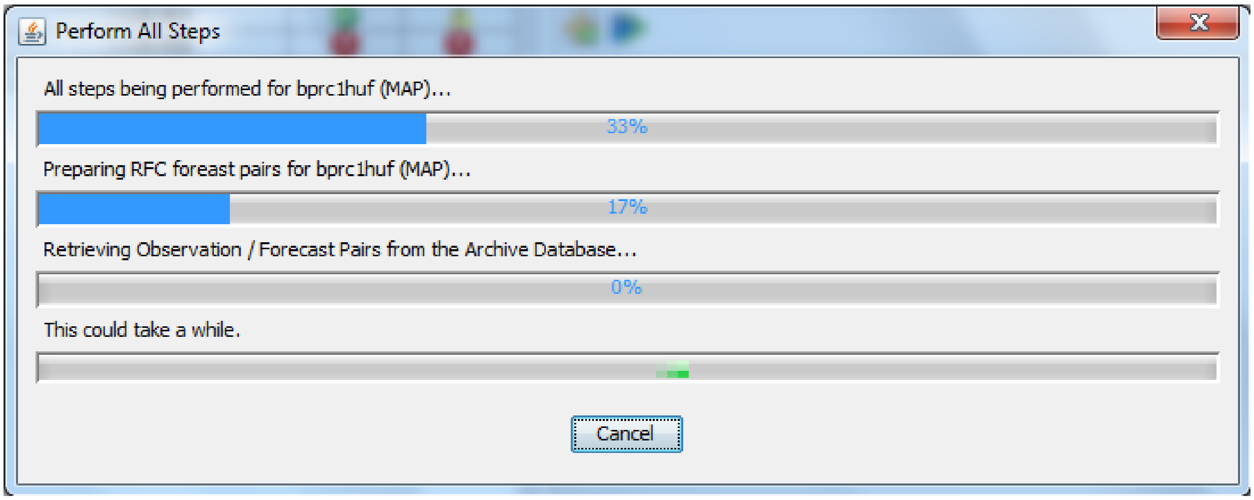


There are three options that the user can set:

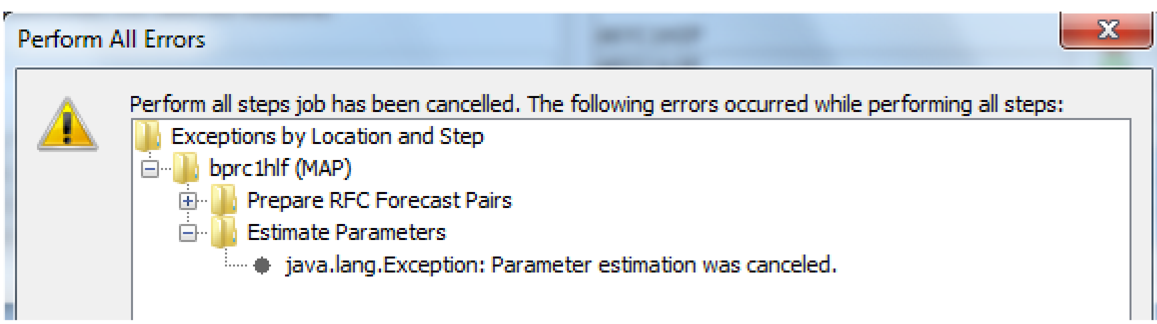
- Specify if the run all is to exit as soon as any step fails for any reason.
- Specify if the run all should re-run already completed steps.

- Specify the behavior of the MEFPPE if a source forecast is not available for a MEFP location but it is to be used (i.e., the number of days for the forecast source is not 0). The options are to set the number of days to be 0 so that the parameters are not estimated for that forecast source, or skip parameter estimation for that MEFP location.

3. Set the options as desired and click **OK** (click **Cancel** to stop the run all). A progress dialog will open:



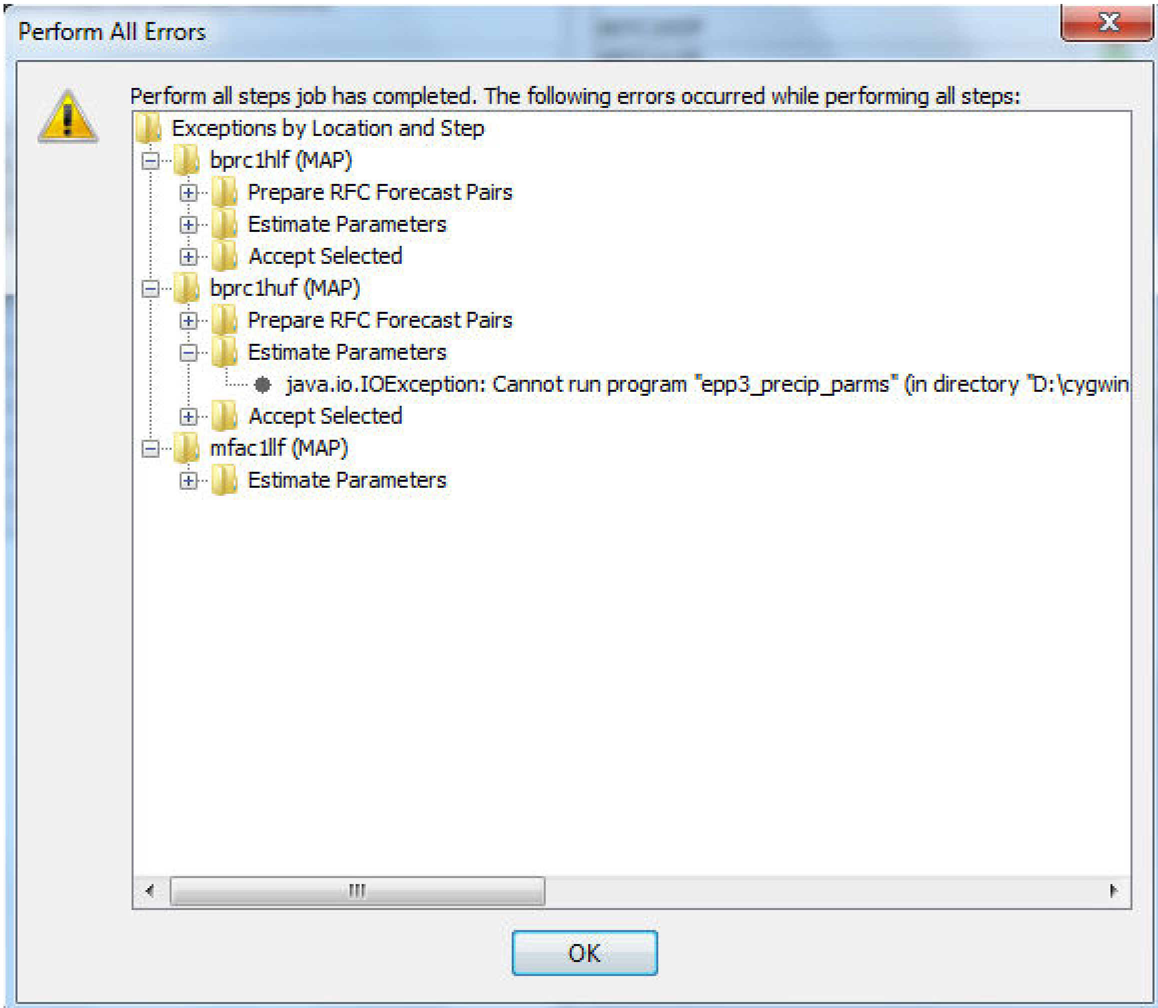
The run all can be canceled at any time by clicking on the **Cancel Button**. The run all will stop at the next opportunity and a **Perform All Errors Dialog** will be displayed where the last leaf will indicate that the user canceled the run:



Click **OK** to close the dialog.

NOTE: When a run all is canceled, the steps performed prior to the cancelation will not be undone.

4. When the run all is completed, a **Perform All Errors Dialog** will be displayed indicating any errors that occurred during the run:



Expand the tree nodes in order to identify the errors that occurred. If no errors occurred, a message will be displayed indicating that no errors occurred.

5. Click **OK** to close the dialog.

3.13 Diagnostic Display Panel

The **Diagnostic Display Panel**, an example of which is shown in Figure 19, displays diagnostics as selected by the user via subpanels of the **Estimation Steps Panel**. The diagnostics that can be displayed are described with the subpanels. Also, a general diagnostic display panel framework is described in Section 3.2.5.3.

Some diagnostics can require a significant amount of time draw. Furthermore, it will need to be redrawn whenever the **Diagnostic Display Panel** is resized. To prevent a slowdown in the software resulting from spending too much time rendering displays, a **Dispose Button** is included in the upper left corner of the panel for all diagnostics to display. Click on the button to clear the panel.

NOTE: The chart displayed in the **Diagnostic Display Panel** will not change until another diagnostic is selected to be displayed or the **Dispose Button** is clicked.

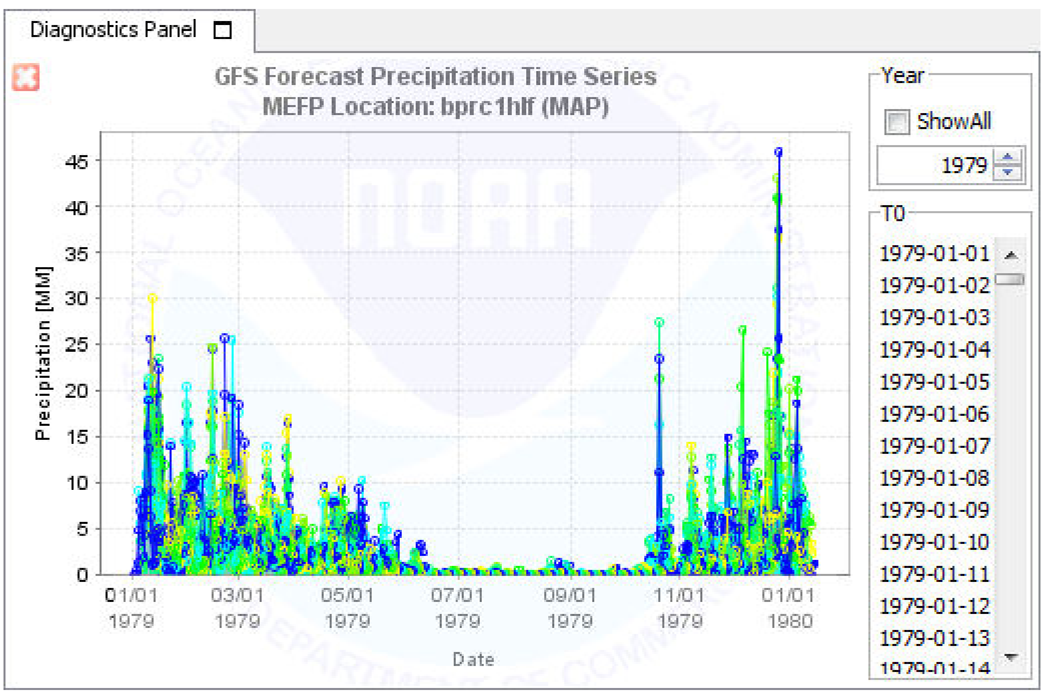


Figure 19: The **Diagnostics Display Panel**.

3.13.1.1 Components

- **Dispose Button**: Click on this button to dispose of the current diagnostics. The button is only visible if a diagnostic is currently displayed.

4 MEFP Operational Reference Manual

This section is pending transition of HEFS from beta-testing to operational use, and will be supplied at a later date. Once completed, it is anticipated that this section will serve as a reference for configuring the MEFP model adapters for execution within a CHPS workflow.

4.1 Overview

4.1.1 Using CHPS to Import and View GFS and CFSv2 Grid Data

4.1.2 Using CHPS to Generate MEFP Ensembles and View Results

4.2 MEFP Grid Data Model Adapter

4.2.1 Description

4.2.2 Model Parameters

4.2.3 Model Time Series

4.2.4 Notes on Configuration

4.3 MEFP Model Adapter

4.3.1 Description

4.3.2 Model Parameters

4.3.3 Model Time Series

4.3.4 Notes on Configuration

4.4 Graphics Generator Products

4.4.1 Required Settings


UNDER
CONSTRUCTION