Meteorological Ensemble Forecast Processor (MEFP) User's Manual

Last Updated: September 23, 2013

National Weather Service
Office of Hydrologic Development

Change History

Date	Version	Description	Author
Sep 4, 2013	1.0.1	Initial version.	Hank Herr

Table of Contents

1 Overview 1
1.1 Manual Layout 1
2 Science of MEFP 2
2.1 Introduction 2
2.2 Scientific Requirements 3
2.3 Forecast Sources 4
2.4 Methodology 4
2.4.1 General Strategy 5
2.4.2 Canonical Events 6
2.4.3 The Meta-Gaussian Model 6
2.4.4 Precipitation Intermittency 7
2.4.5 The Schaake Shuffle 7
2.4.6 Ensemble Merging 7
2.5 Scientific Considerations 8
2.5.1 Calibration 8
2.5.2 HPC/RFC Archive 8
2.5.3 Why Use GEFS Ensemble Mean 8
2.5.4 Strategies to Reduce Sampling Uncertainty 9
2.5.5 No Skill Forecasts 10
2.5.6 The Use of Historical MAP and MAT 12
2.5.7 Evaluation 12
2.6 Summary 8
References 14
3 MEFPPE Reference Manual 15
3.1 Overview 15
3.1.1 Terminology 15
3.1.2 Notation 16
3.1.3 Directories of Note 16
3.2 Getting Started 16
3.2.1 Input Data Requirements 16
3.2.2 Running MEFPPE 17
3.2.3 The Parameter Estimation Procedure 18
3.2.4 Core Concepts 19
3.2.4.1 The MEFPPE Run Area 19
3.2.4.2 Run-time Information 20
3.2.4.3 Importing from XEFS EPP3 20
3.2.4.4 FEWS PI-service Connection 21
3.2.4.5 Canonical Events 21
3.2.4.6 Questionable Data 21
3.2.5 General Graphical User Interface Components 22
3.2.5.1 Generic Summary Table 22
3.2.5.2 Table Delete/Add and Status Columns 23
3.2.6 Format of this Reference Manual 24

3.3 MEFPPE Main Panel 25
3.4 Estimation Steps Panel 26
3.4.1 Components 27
3.4.2 Usage 28
3.4.2.1 Performing a Parameter Estimation Step 28
3.5 Setup Subpanel 30
3.5.1 Export Historical Data Subpanel 31
3.5.1.1 Components 31
3.5.1.2 Diagnostics 32
3.5.2 Canonical Events Subpanel 33
3.5.2.1 Components 33
3.5.3 Usage 34
3.5.3.1 Configuring the FEWS PI-service 34
3.5.3.2 Exporting Historical Data via the FEWS PI-service 36
3.5.3.3 Exporting Historical Data Files Manually 37
3.5.3.4 Adding a New Canonical Event 37
3.6 Historical Data Subpanel 38
3.6.1 Components 38
3.6.2 Diagnostics 39
3.6.3 Usage 40
3.6.3.1 Generating Binary Data and Questionable Message Files 40
3.7 RFC Forecasts Subpanel 41
3.7.1 Components 42
3.7.2 Diagnostics 43
3.7.3 Usage 44
3.7.3.1 Using IVP to Pair RFC QPF/QTF Forecasts 44
3.7.3.2 Specifying the Archive Database Connection 46
3.7.3.3 Constructing Archive Files Using Archive Database Data 47
3.7.3.4 Importing Files Constructed Externally 48
3.8 GFS Subpanel 55
3.8.1 Components 56
3.8.2 Diagnostics 56
3.8.3 Usage 57
3.8.3.1 Downloading GFS Reforecast Files 57
3.9 GEFS Subpanel 58
3.9.1 Components 59
3.9.2 Diagnostics 59
3.9.3 Usage 60
3.9.3.1 Downloading GEFS Reforecast Files 60
3.10 CFSv2 Subpanel 61
3.10.1 Components 62
3.10.2 Diagnostics 62
3.10.3 Usage 65
3.10.3.1 Downloading CFSv2 Reforecast Files 65
3.11 Estimation Subpanel 66
3.11.1 Locations Summary Subpanel 67

3.11.1.1 Parameter File Backups ... 67
3.11.1.2 Components ... 67
3.11.2 Estimation Options Subpanel ... 68
3.11.2.1 General Options ... 68
3.11.2.2 IPT Options ... 69
3.11.2.3 EPT Options ... 69
3.11.2.4 Source Parameters ... 69
3.11.3 Standard Diagnostics Panel ... 70
3.11.4 Parameter Diagnostics Panel ... 72
3.11.4.1 Components ... 73
3.11.4.2 Features ... 74
3.11.4.3 Parameter Summary Panel ... 78
3.11.4.4 Using the Parameter Diagnostics Panel ... 78
3.11.5 Usage ... 79
3.11.5.1 Estimating Parameters ... 79
3.11.5.2 Loading Parameters and Viewing Diagnostics ... 79
3.11.5.3 Backing-Up Parameters and Restoring Backup Parameters ... 79
3.12 Acceptance Subpanel ... 81
3.12.1 Components ... 82
3.12.2 Diagnostics ... 82
3.12.3 Usage ... 82
3.12.3.1 Accepting Parameters ... 82
3.13 Location Summary Panel ... 83
3.13.1.1 Components ... 83
3.13.2 Usage ... 84
3.13.2.1 Running All Steps for Multiple MEFP Locations ... 84
3.14 Diagnostics Panel ... 87
3.14.1 General Components ... 88
3.14.2 Time Series Diagnostics Panel Components ... 88
3.14.3 Time Series Diagnostics Panel Features ... 90
3.14.4 Using Time Series Diagnostics Panels ... 93
4 MEFP Operational Reference Manual ... 95
4.1 MEFP Configuration ... 96
4.1.1 Data Ingest ... 96
4.1.2 Operational Forecasting ... 96
4.1.3 Hindcasting ... 97
4.2 TimeSeriesExporterModelAdapter Reference Manual ... 99
4.2.1 Overview ... 99
4.2.1.1 Example Configuration ... 99
4.2.2 Model Parameters ... 101
4.2.3 Model Run File Properties ... 101
4.2.3.1 fileNamePattern Property ... 102
4.2.3.2 t0ComputationAdjustmentFactorFromFirstDataValue Property ... 103
4.2.4 Model Input Time Series ... 104
4.2.5 Model Execution ... 104
4.2.6 Model Output Time Series ... 104

4.2.7 Model Description 104
4.2.7.1 Example File Names 105
4.3 CFSv2LaggedEnsembleModelAdapter Reference Manual 106
4.3.1 Overview 106
4.3.1.1 Example Configuration 107
4.3.2 Model Parameters 108
4.3.3 Model Run File Properties 109
4.3.4 Model Input Time Series 110
4.3.5 Model Execution 110
4.3.6 Model Output Time Series 111
4.3.7 Model Description 111
4.4 MEFPEnsembleGeneratorModelAdapter Reference Manual 113
4.4.1 Overview 113
4.4.1.1 Example Configuration 114
4.4.2 Model Parameters 119
4.4.3 Model Run File Properties 119
4.4.4 Model Input Time Series 125
4.4.5 Model Execution 126
4.4.6 Model Output Time Series 126
4.4.7 Model Description 128
4.4.7.1 Model Errors 128
4.4.8 Notes on Configuration 129
4.4.8.1 Base Ensemble Member Construction using memberIndexingYear 129
4.4.8.2 Appending Raw Climatology 130

1 Overview

The Meteorological Ensemble Forecast Program (MEFP) is a statistical model that combines information from different forecast sources and generates a single forecast ensemble of meteorological inputs to streamflow models. It requires parameters that must be estimated prior to executing the model. The parameters are estimated via the MEFP Parameter Estimator (MEFPPE) and are stored in a central directory for access by the model when executed. The model is executed via the MEFPEnsembleGeneratorModelAdapter, which is executed as any other CHPS module within a workflow.

Read the *HEFS Overview and Getting Started Manual* for an introduction to HEFS as a whole and MEFP in particular.

1.1 Manual Layout

This document follows this outline:

Section 2: A description of the science underlying the MEFP.

Section 3: A reference manual describing all of the components and features of the MEFPPE.

Section 4: A description of how to execute the MEFP operationally to generate forecast ensembles and configuration reference manuals for the three model adapters delivered with MEFP.

2 Science of MEFP

2.1 Introduction

The Meteorological Ensemble Forecast Processor (MEFP) is a key component of the 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 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 Ensemble Forecast System (GEFS) developed at National Centers for Environmental Prediction (NCEP).
- Single-valued forecasts generated by the Climate Forecast System version 2 (CFSv2) of NCEP.

The MEFP can be used to produce ensemble forecasts from any feasible combination of the HPC/RFC, GFS, GEFS, and CFSv2 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 over a year 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 GEFS ensemble forecasts is used by the MEFP to cover a forecast lead time of up to 15 days. The GEFS model has a resolution of 1 degree. A 16-member ensemble is produced operationally daily by the GEFS from using 0, 6, 12, and 18 UTC initial conditions, and the 1st member is from the control run. The GEFS ensemble forecasts extend to 16 days, with data archived at a 6-hour time step. Because these forecasts are unreliable – biased in the mean, spread, and higher moments, in the MEFP, the mean of the GEFS ensembles is used to derive reliable forcing ensembles. Overall, the biased-corrected GEFS precipitation forecasts have moderate skill over the CONUS.
- For long range forecasting, the CFSv2 forecasts are used. 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. Section 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.

- *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.
- *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.
- *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.
- *Rank the canonical events for a forecast source according to the correlation of the forecast and corresponding observed for these events.*
- *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 which of these two procedures are applied operationally.

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. Currently, the only reforecasts available are the GEFS part of the MMEFS, which is one of the forecast sources used by the MEFP.

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 GEFS Ensemble Mean

There are two reasons why the GEFS 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 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 archived years, the sample size would be very small and the resulting parameter values would be very noisy as a result of large sampling error. Therefore, data pooling in a time window 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 as 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 parameters include 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 control parameters. The MEFP has default values for these control parameters, 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.
- 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 parameters include an option that specifies the number of days between days for which parameters are estimated, defaulting to 5. This option should only be changed if parameter estimation computational time becomes excessive; increasing the value reduces the number of parameter estimations that are performed, thereby improving performance. Note that when applied operationally, the MEFP selects parameters for the day of the year that is closest to the operational day of the year, meaning that the control parameter should not be made very large (it is restricted to a maximum of 10 days in the MEFP parameter estimation software).
- During very dry periods, the length of the data window may reach a maximum window width, specified by a control parameter (default is 61), and there may not be enough data to estimate all of the parameters. In that case, 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 positive observations or forecasts is less than minimums, specified by control parameters independently for observations (default is 30) and forecasts (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 handling 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 EPT Model Diagnostics

The MEFP implements two different models in treating precipitation intermittency (see section 2.4.4 above). One of them separates rain and no-rain situations and is call Explicit Precipitation Intermittency Treatment (EPT) model. The other treats precipitation intermittency in an implicit way and is called Implicit Precipitation Intermittency Treatment (IPT) model. It is recommended that the EPT model be used in most situations unless the probability of precipitation (PoP) for the climatology of both forecast and observation is high.

A good understanding of the EPT model is essential in diagnosing problems with its parameter estimation. The EPT model basically decomposes the relationship between precipitation forecast and observation into several components. One can easily see that there are a total of four cases for forecast-observation pairing with wet or dry situations. They are (dry, dry), (dry, wet), (wet, dry), and (wet, wet), with the first element in the pairs representing forecast, and the second observation. These pairings provide a partition of the probability space of the bivariate forecast-

observation relationship. The related parameters shown for the EPT model within the **Diagnostics Panel** (see Sections 3.11.3and 3.11.4) of the MEFPPE are:

- Prob of No Precip for Fcst and Obs (EPT)
- Prob of Positive Fcst and Zero Obs (EPT)
- Prob of Positive Obs and Zero Fcst (EPT)

The probability of positive forecast and observation is given by one minus the sum of the above three probabilities.

Given a single-valued QPF of zero amount, an ensemble is derived for the (dry, dry) and (dry, wet) situations. Given a positive amount of QPF, an ensemble is derived for the (wet, dry) and (wet, wet) situations. The most important consideration in the EPT model diagnostics is the sample sizes for these four pairings. Sample sizes of these pairings can be small, even zero, under certain conditions. For example, during the summer months, the DOSC1HLF basin in CNRFC is very dry, resulting in small samples for the (wet, wet) pairing. The EPT model is implemented in such a way that it is "robust" in treating pairings of (dry, dry), (dry, wet), and (wet, wet), in the sense that the algorithm will not break down for extreme small samples, even when they are empty. This robustness is achieved through using different strategies in calibration. With large samples when sampling uncertainty is reduced, the EPT model fits the data to parametric distributions. With small samples, where fitting data to a parametric distribution is infeasible, the EPT model simply generates ensembles by directly sampling the historical observed data. Potentially, the (wet, dry) pairing can be troublesome, when its sample is too small. An enhanced IPT model should be explored with the aim of overcoming this difficulty.

In the following, each of the four pairings is considered one by one when the sample is small.

- There are a number of situations from which a small sample of (dry, dry) can result. One of them is that the forecast location is in a wet area. Another one is that the forecast period is in a wet season. A third one is that the forecast time scale for which precipitation amount is accumulated is large. What the situation may be, a small (dry, dry) sample does not impede the performance of the EPT model.
- The next pairing is (dry, wet). A small (dry, wet) sample means that the positive portion of an ensemble generated conditional on a zero QPF can be highly unreliable, or biased in the mean. A related issue with ensembles generated from zero single-valued QPFs is that the size of the positive portion of the ensemble is so small that the resulting ensembles are biased for the positive portion, no matter how large the (dry, wet) sample is.
- The third pairing is (wet, dry). The sample size of this pairing determines how well the probability of precipitation (PoP) for the ensembles can be estimated. If the (wet, dry) sample is too small, a poor ensemble PoP may result.
- Lastly, we have the (wet, wet) pairing. A small (wet, wet) sample can occur in a dry area or season. When this is the case, unable to generate an ensemble conditional on a specific QPF, the EPT model simply samples the entire data set for the pairing to produce positive ensemble members. If this happens, the positive portion of the ensemble reflects what historically happened under the condition of a positive single-valued QPF being issued.

The skill of the MEFP ensembles from the EPT model for the wet events is reflected in the correlation between the positive forecast and observation. The value of this correlation is provided by the MEFPPE parameter "Correlation when Obs and Fcst are Positive (EPT)" in the **Diagnostics Panel**. It is worth mentioning that the values of this parameter can be misleading when the (wet, wet) sample is small. Occasionally, medium-to-high correlation is given by the **Diagnostics Panel** when the sample consists of just a few points.

The following MEFPPE parameters are provided for diagnosing EPT model by the developers. The RFC users can ignore them.

- Probability of precipiation for Forecasts
- Probability of precipiation for Observations
- Fcst Marginal Used for PDF (EPT)
- Fcst Cond on zero Obs (EPT)
- Fcst Marginal of Bivar Dist (EPT)
- Obs Cond on Zero Fcst (EPT)
- Obs Marginal of Bivar Dist (EPT)

For the current version, the RFC users can also ignore the following two MEFPPE parameters, which are used in a research setting:

- Optimal Linear Coefficient Value (EPT)
- Optimal Median Adj Error Factor (EPT)

2.5.6 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.

2.5.7 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 (converted to 24-hour minimum and maximum temperature) data in estimating model parameters.
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 forcing for ESP.

2.5.8 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 for over 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. An ongoing activity is to evaluate the MEFP within the HEFS end-to-end verification and validation, where precipitation and temperature ensembles based on GFS, CFSv2, and GEFS forecasts are ingested to hydrological models to produce streamflow ensembles with post-processing. Basins in AB-, CB-, CN-, MA-, and NERFCs are used in this study. The part of the study involving GFS forcing has been completed.

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 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 contained in gzipped tar files (.tgz) located under the <*mefp_root_dir*> defined within the *MEFP Installation Guide: Data Ingest Components*. 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 those parameters by placing them in a central area for use in generating ensemble forecasts. Alternatively, the user is also able to perform all steps in a hands-off mode via a run-all feature. 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 determine 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 referred to in MEFPPE 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 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, NCEP GEFS, 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)". MEFP locations correspond to catchments within CHPS. The parameterId is indicates if the location is a precipitation (MAP) or temperature location (MAT).

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.
- Directory and file names will be denoted in this font.

3.1.3 Directories of Note

The following directories will be referred to in this manual:

- <*region_dir*>: The *parameter estimation stand-alone* (see the *MEFPPE Installation Guide*) region home directory, typically “##rfc_sa”.
- <*configuration_dir*>: The *parameter estimation stand-alone* configuration directory, typically <*region_dir*>/Config.
- <*mefp_root_dir*>: The directory selected to hold CFSv2 location time series files and MEFP parameter files; see the *MEFP Installation Guide: Data Ingest Components*.

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 user 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 (TMIN/TMAX) time series must be made available to it. See Section 3.5.1 and 3.5.3 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 (1998-frozen version) 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.
- GEFS Forecast Source: GEFS reforecasts are required. Those reforecasts are automatically acquired as needed by MEFPPE via SFTP. See Section 3.9.3.1. 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 3.10.3.1. No preparation by the user is required.
- Climatology (Historical) Forecast Source: The aforementioned historical data is used to estimate parameters for this forecast source.

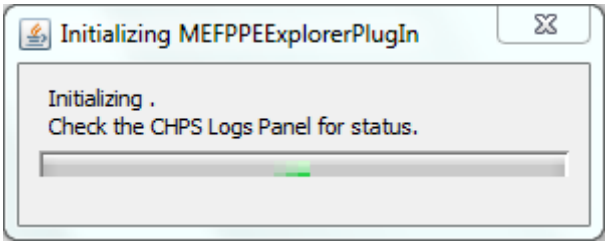
All other information is acquired from CHPS via PI-XML files that are exported, including location ids, parameter ids, and coordinates.

3.2.2 Running MEFPPE

To use MEFPPE, you must install the software as described in the *HEFS Install Notes*, configure it to run in a CHPS stand-alone as described in the *MEFPPE Configuration Guide*, 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 close 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 6-hour MAP and 24-hour TMIN/TMAX 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 historical data files can be constructed manually; see the section on MEFPPE installation in the *HEFS Release Install Notes*.

   See: Section 3.5.3 for how to setup for parameter estimation using the **Setup Subpanel** of the **Estimation Steps Panel**.

   NOTE: *The TMIN/TMAX data is computed from 6-hour instantaneous MAT time series during importing of data required for the MEFPPE. See the section on MEFPPE installation in the HEFS Install Notes.*

2. *Process historical data and generate binary files*
   Create faster-access binary files containing historical data, to be stored within the estimated parameters .tgz file. Examine the data and check for badly specified values.

   See: Section 3.6.3.1 for how to create the binary files using the **Historical Data Subpanel** of the **Estimation Steps Panel**. See Sections 3.6.2 and 3.14 for how to apply the **Diagnostics Panel** in quality controlling historical data.

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. Examine the data and check for badly specified values.

   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. See Sections 3.7.2 and 3.14 for how to apply the **Diagnostics Panel** in quality controlling the RFC data.

4. *Acquire GFS reforecast data files*
   Acquire the reforecast data files for the 1998-frozen 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 GEFS reforecast data files*
   Acquire the reforecast data files for the GEFS forecast source. Reforecasts (forecasts for past dates) are necessary in order to estimate the MEFP parameters.

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

6. *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 3.10.3.1 for how to acquire CFSv2 reforecast files via SFTP using the **CFSv2 Subpanel** of the **Estimation Steps Panel**.

7. *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 3.11.5.1 for how estimate parameters and Section 3.11.5.2 for how to view diagnostics related to estimated parameters, both using the **Estimation Subpanel** of the **Estimation Steps Panel**.

8. *Accept parameter files*
   Copy the parameter files from the MEFPPE run area to permanent storage under the <*mefp_root_dir*> directory.

   See: Section 3.12.3.1 for how to accept parameters 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 underneath the HEFS models directory, pointed to by the global property HEFSMODELSDIR (typically <*region_dir*>/Models/hefs):

$HEFSMODELSDIR$/mefppeRunArea

Files stored under that directory include run-time information files, historical data files, archive or reforecast data files, and parameter files.

NOTE: Never *modify anything within the MEFPPE run area unless specifically instructed to do so within this manual (as in Section 3.5.3.2) or in order to debug a problem.*

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 location ids and coordinates.
- Canonical events defined via the interface.
- Latest estimation options specified via the interface.

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:

…/mefppeRunArea/.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 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.

…/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)

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; MEFPPE will only import the four files specified in the list above.

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 2.4.2. 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.

3.2.4.6 Questionable Data

The **Diagnostics Panel** within the MEFPPE interface is used for examining forecast source data and estimated parameters. Of particular interest are questionable data values that may lead to poor parameter estimates. While the reforecast data for the GFS, GEFS, and CFSv2 forecast sources have been quality checked by OHD prior to making them available for RFCs, this is not true of historical data or the RFC forecast source. In those two cases, questionable data is identified when the step is performed, appropriate messages are stored in a questionable message file (no file is created if no questionable data exists), and the questionable values are indicated within the components of the **Diagnostics Panel** for that source's subpanel.

A time series value is considered questionable if one of the following is true:

- It is missing with non-missing values both before and after it within the time series.

- It exceeds gross range limits. For precipitation, if the value is negative or exceeds 100mm it is considered questionable, while for temperature data outside the range -100 degC to 100 degC is considered questionable.
- The range of temperature values for a day (i.e., maximum temperature – minimum temperature) is either negative (minimum exceeds maximum) or exceeds 50 degC.

For temperature data, if either the minimum or maximum temperature is considered questionable, both are considered questionable and will be highlighted as such within tools of the **Diagnostics Panel**.

If questionable data is found, it must be fixed outside of MEFPPE. For historical data the two choices are the following:

1. Modify the original source of the data (e.g., datacard files), reimport the data, and perform the step in the **Export Historical Data Subpanel** (see Section 3.5.3.2) again.
2. Use tools provided by CHPS, such as the FEWS **Time Series Editor**, to change the time series in the localDataStore. Once done, perform the step in the **Export Historical Data Subpanel** (see Section 3.5.3.2) again. Using the FEWS **Time Series Editor** may require additional configuration file changes.

Two choices also exist for RFC forecast source data depending the source of the data:

1. Modify the data in the archive database, reconstruct the vfypairs, and perform the step in the **RFC Forecasts Subpanel** again. See Section 3.7.3 for instructions pertaining to each of these substeps.
2. Manually modify the files that were imported as per Section 3.7.3.4 and re-import the data.

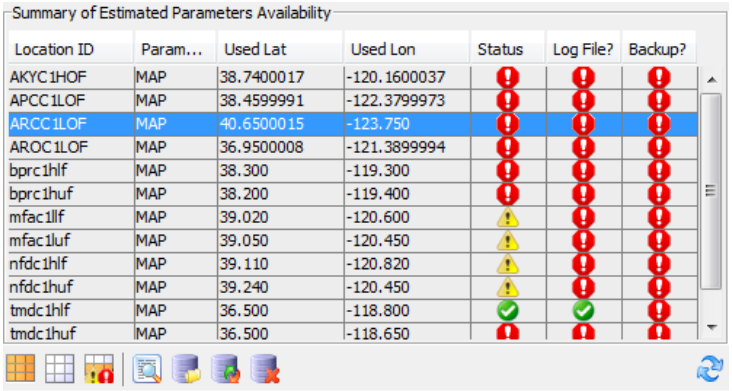
3.2.5 General Graphical User Interface Components

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:

NOTE: The MEFP locations are identified in the table by 'Location ID' and 'Parameter ID' columns. The 'Parameter ID' column identifies the type of data for which parameters are estimated by an abbreviation based on CHPS parameters: MAP (precipitation) or MAT (temperature).



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 manually changes files in the MEFPPE run area.

When this table is used within a 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:

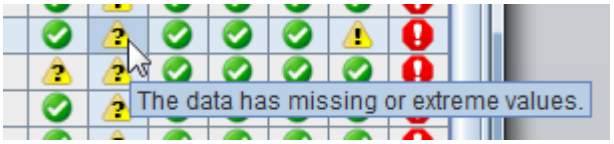
- **Delete Icon**: Click to delete the row from the associated table. Sometimes this will cause a dialog popup confirming the delete.
- **Add Icon**: Click to add a row to the associated table.

Status Column:

- **Bad Status Icon**: Indicates that a step has not been performed or an error of some kind occurred while performing the step.
- **Warning Status Icon**: Indicates that a step has been performed but needs to be updated (performed again).

- **Questionable Status Icon**: Indicates that the step was performed but some of the resulting data is questionable. For example missing data within a time series is considered questionable.
- **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 as needed.

The **Questionable Status Icon** only appears in the **Generic Summary Panels** displayed within the **Historical Data Subpanel** (Section 3.6) and **RFC Forecasts Subpanel** (Section 3.7), as well as the columns for those two sources within the **Location Summary Panel** (3.13). 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.

3.2.6 Format of this Reference Manual

The Sections 3.4 – 3.14 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.
- A description of the **Diagnostics Panel** used for that component panel.
- Instructions for how to perform basic tasks using the components.

3.3 MEFPPE Main Panel

Shown in Figure 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 Panel**: Displays diagnostics that assist the user in quality controlling the data, deciding on options to use for estimation, and quality controlling 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 subpanels in the* **Estimation Steps Panel**.

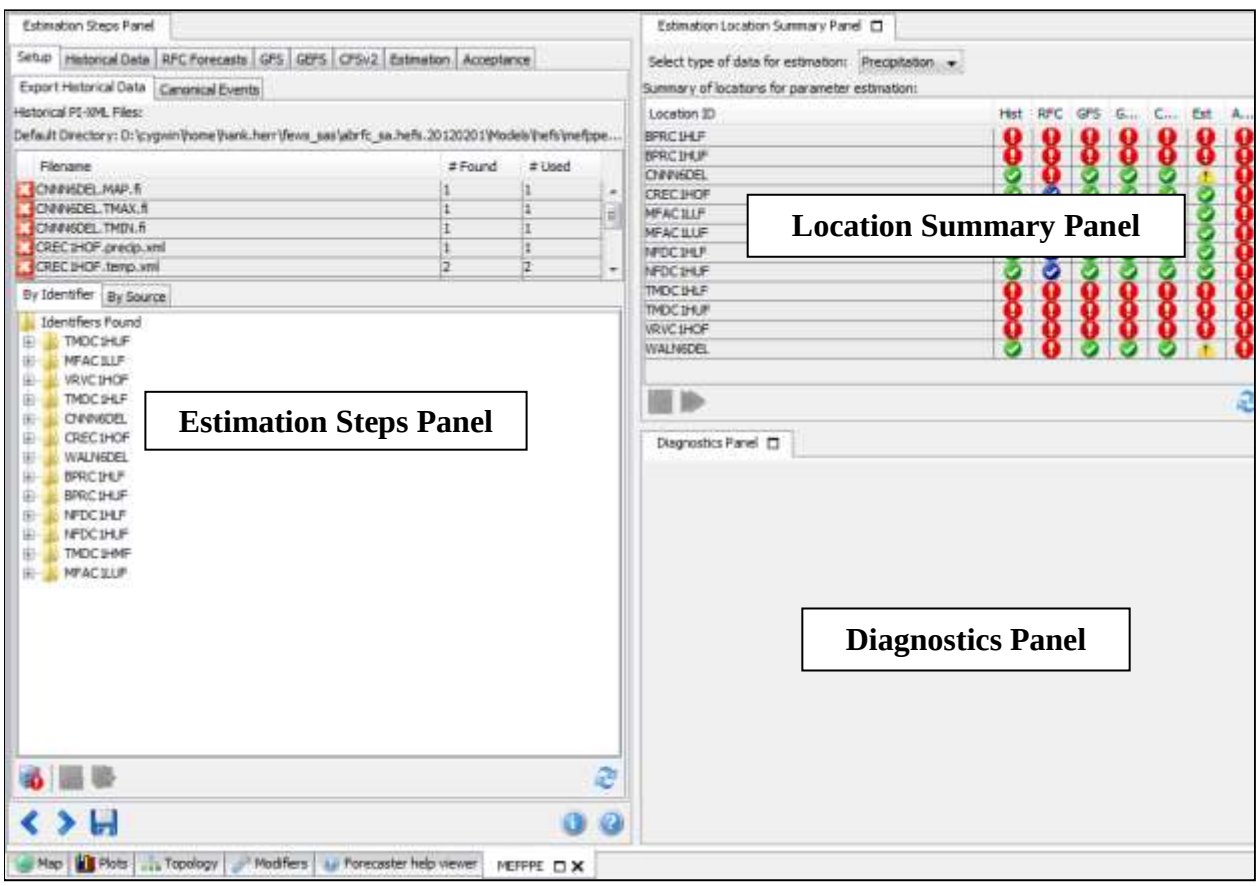


Figure 1: The **MEFPPE Main Panel**, displayed upon start-up of the MEFPPE.

3.4 Estimation Steps Panel

The **Estimation Steps Panel**, shown in Figure 2, 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 **About Button**, and a **Help Button**.

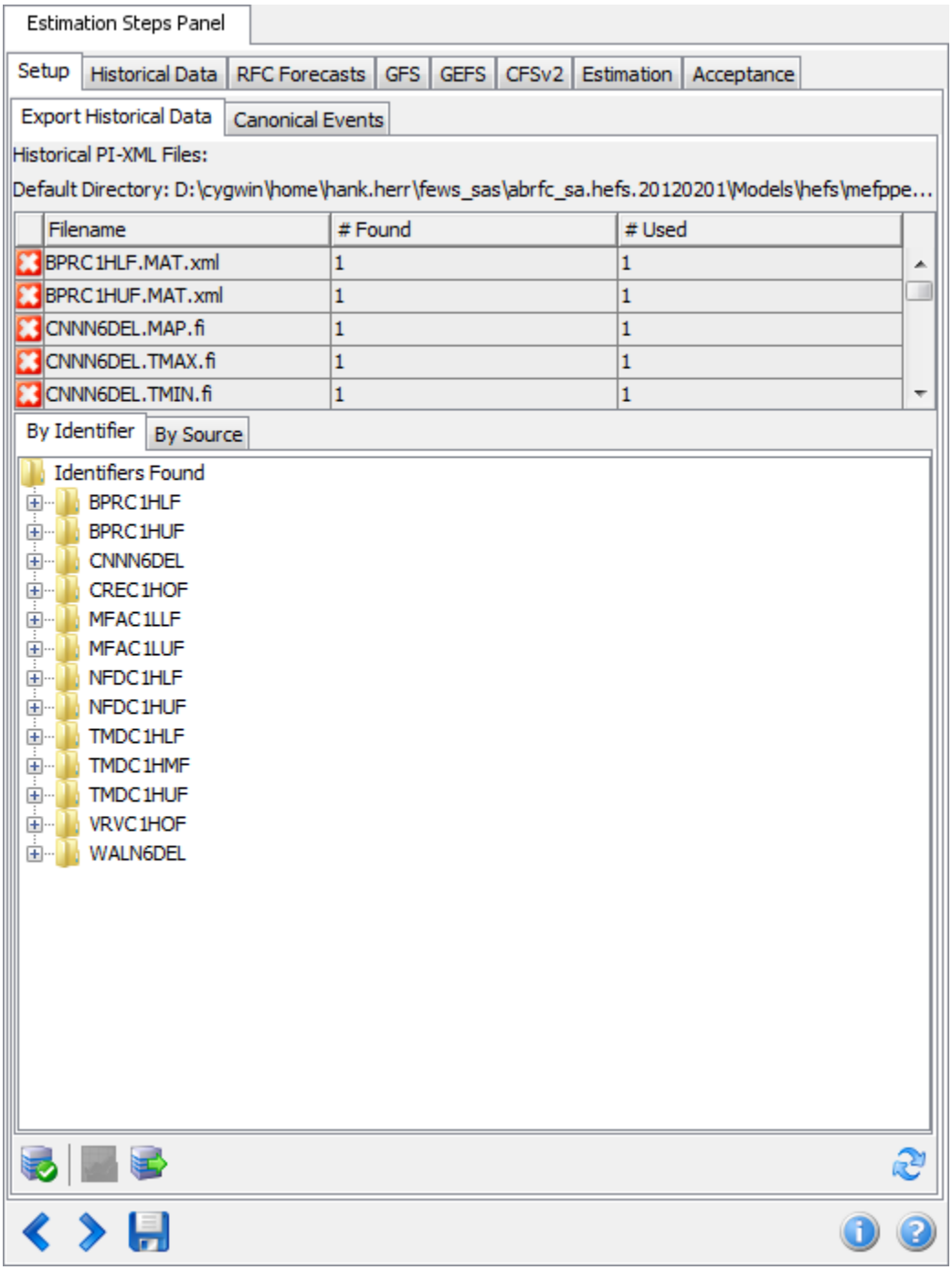
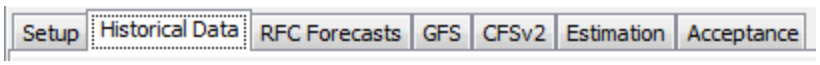


Figure 2: The **Estimation Steps Panel**.

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.

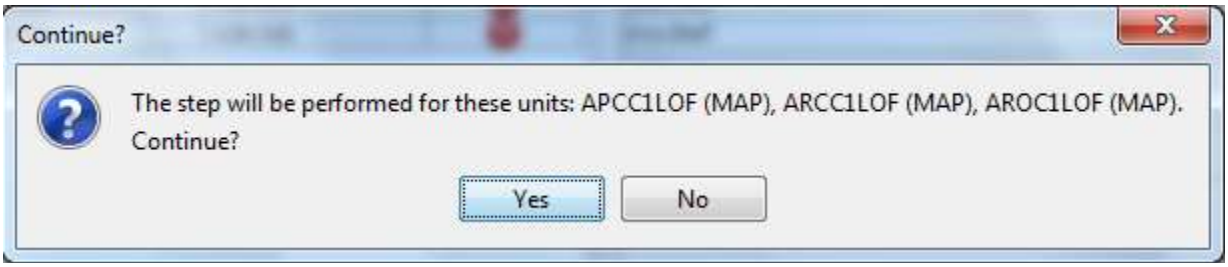
  NOTE: A subpanel is said to be active *if its tab is selected and its contents are current viewed. For example, in the image above, the* **Historical Data Subpanel** *is active.*

- / **Back/Next Buttons**: Click to navigate to the previous or the next step tabbed subpanel. The **Back** or **Next Button** will be disabled if there is no previous or next subpanel, respectively.
- **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.
- **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 every 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.
- **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 a dialog that pops up. The component for which help will be provided is highlighted by a faded red box; for example:

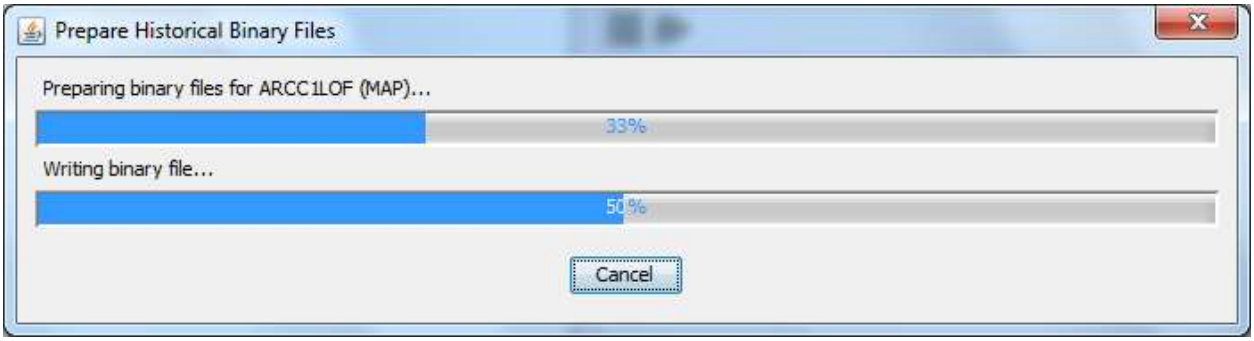
3.4.2 Usage

3.4.2.1 Performing a Parameter Estimation Step

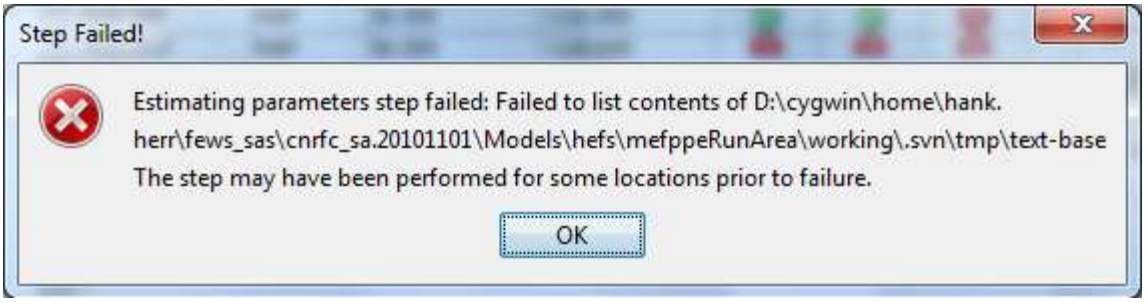
A step is performed by making the corresponding step subpanel active, selecting the MEFP locations for which to perform the step, and clicking on the **Perform Step Button**. Upon clicking, a **Continue Dialog** will be displayed showing the MEFP locations for which the step will be performed and allowing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



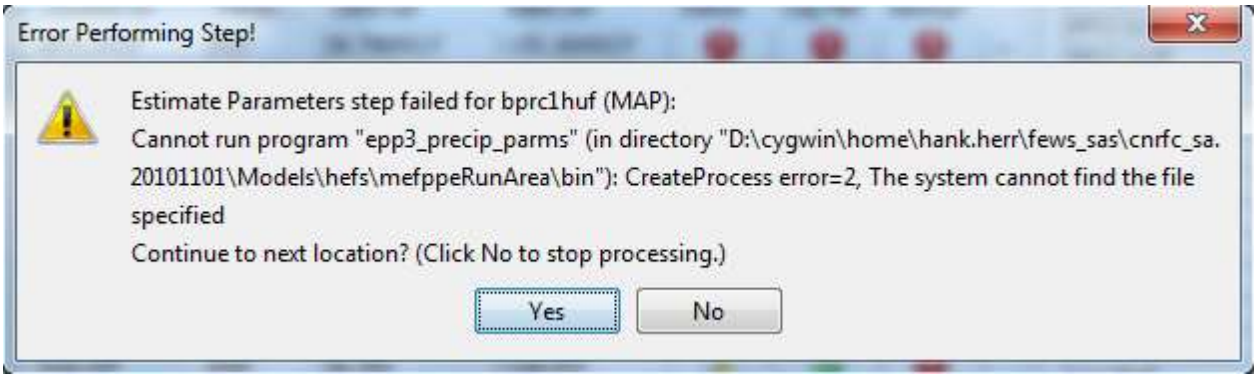
Click on **Yes** to continue or **No** to cancel. If **Yes** is clicked, a **Step Progress Dialog** will be displayed providing the ability to cancel the step via a **Cancel Button**:



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 3, 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 XML or fastInfoset file in the MEFPPE run area. Those files can be created manually by the user (see Section 3.5.3.2) 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 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.

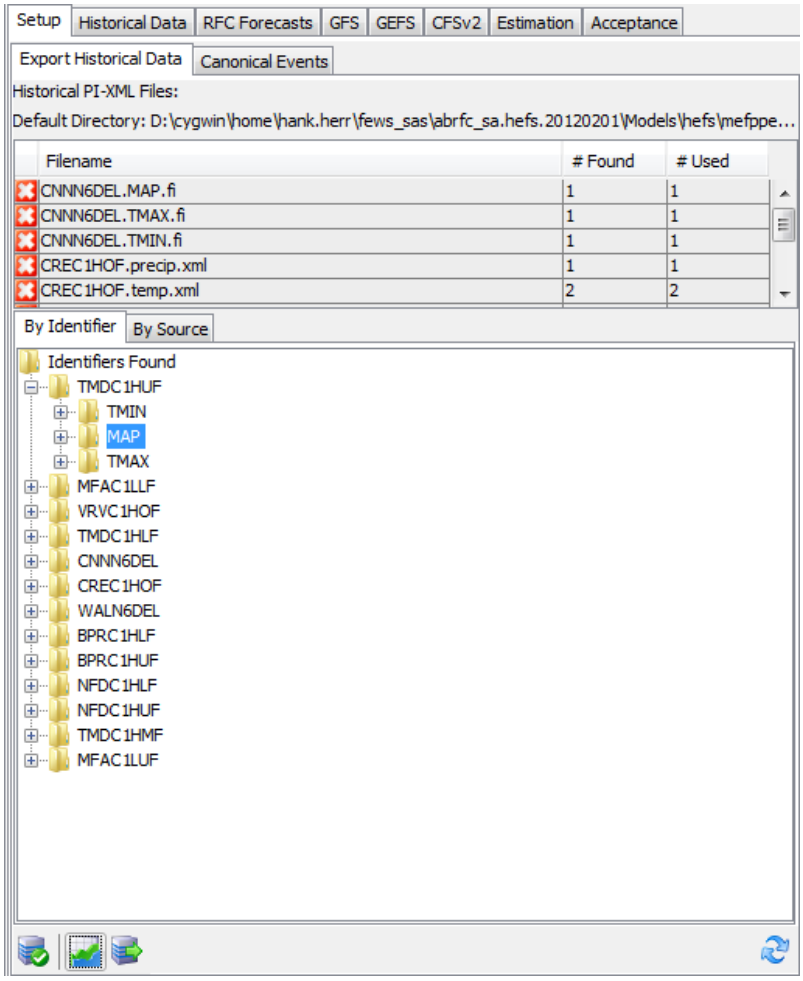


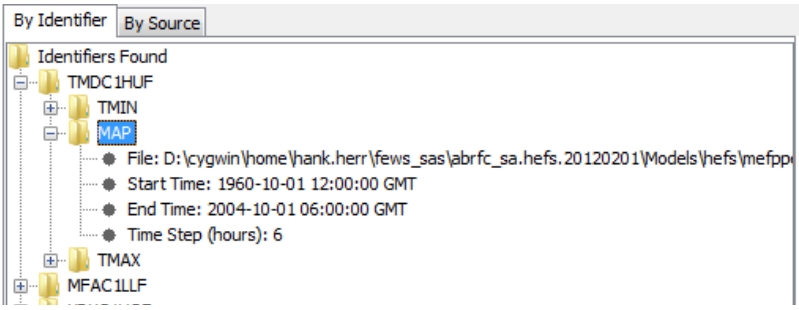
Figure 3: The **Setup Subpanel** of the **Estimation Steps Panel**.

3.5.1 Export Historical Data Subpanel

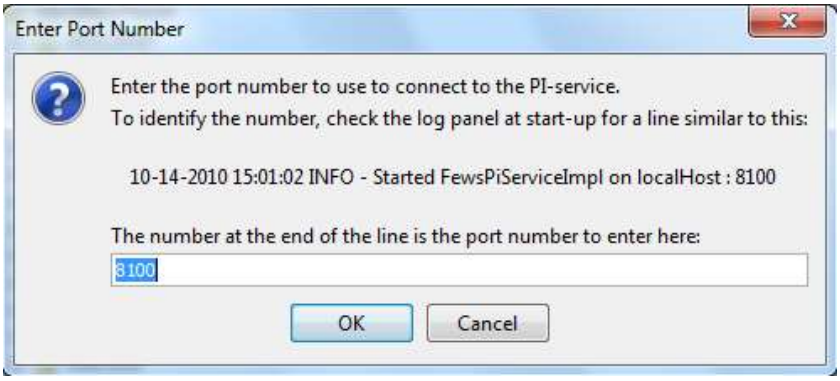
The **Export Historical Data Subpanel**, shown above in Figure 3, facilitates creating time series XML files for use by the MEFPPE based on time series acquired via the FEWS PI-service. It also facilitates viewing the data. The time series are gathered by examining files in the directory 'historicalData' 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.
- **Export Time Series from CHPS DB Button**: Click to extract time series from the localDataStore via the FEWS PI-service and export those time series to files on the file system for use by MEFPPE. See Section 3.5.3.2.
- **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 3.5.3.2.

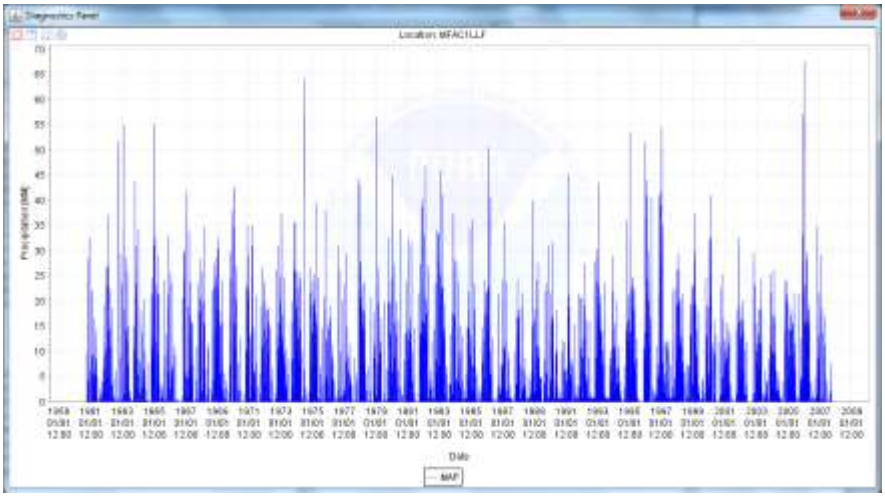
3.5.1.2 Diagnostics

This uses a standard time series **Diagnostics Panel** (Section 3.14), but does not include checks for questionable data. As such, this panel should only be used to verify that the historical data was read in properly from its source XML or fastInfoset file.

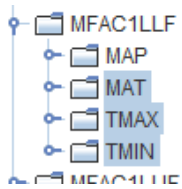
NOTE: Use the **Historical Data Subpanel** *to examine and quality control the data.*

If only one node within the **Tree** is selected

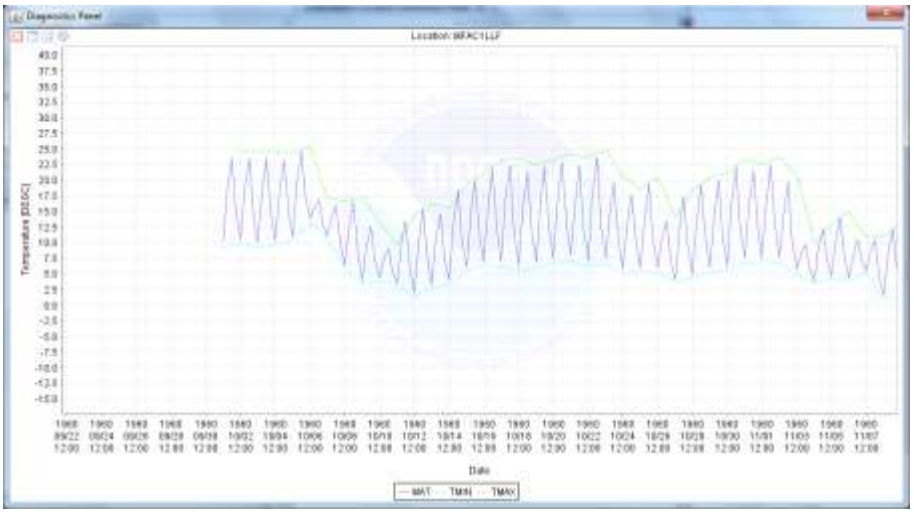
then the one selected time series will be displayed in blue:



For temperature data, however, a user can select multiple time series. For example, if MAT, TMAX, and TMIN are selected



all three will be plotted on the same chart as shown here (zoomed in to show details):



Time series can be selected for only one location at a time.

3.5.2 Canonical Events Subpanel

The **Canonical Events Subpanel**, shown in Figure 4, 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.

3.5.2.1 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 | 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	60	12	5
27	61	76	16	5
28	77	96	20	5
29	97	120	24	10

Figure 4: The **Canonical Events Subpanel** of the **Setup Subpanel**.

3.5.3 Usage

3.5.3.1 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:

<*configuration_dir*>/PiServiceConfigFiles/MEFPPE.xml

with a query id “All Historical Data”. The query must be configured to return all MAP and TMIN/TMAX time series (MAT data is optional) 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>
      <exportIdMap>IdExportMEFPPE</exportIdMap>
   </general>

   <timeSeries>
      <id>All Historical Data</id>
      <timeSeriesSet>
         <moduleInstanceId>MEFP_MAP_to_GMT</moduleInstanceId>
         <valueType>scalar</valueType>
         <parameterId>MAP</parameterId>
         <locationSetId>Catchments_HEFS</locationSetId>
         <timeSeriesType>external historical</timeSeriesType>
         <timeStep unit="hour" multiplier="6"/>
         <readWriteMode>read complete forecast</readWriteMode>
      </timeSeriesSet>
      <timeSeriesSet>
         <moduleInstanceId>MEFP_MAT_to_TAMN_TAMX</moduleInstanceId>
         <valueType>scalar</valueType>
         <parameterId>MAT</parameterId>
         <qualifierId>GMT sampled</qualifierId>
         <locationSetId>Catchments_HEFS</locationSetId>
         <timeSeriesType>external historical</timeSeriesType>
         <timeStep unit="hour" multiplier="6"/>
         <readWriteMode>read only</readWriteMode>
      </timeSeriesSet>
      <timeSeriesSet>
         <moduleInstanceId>MEFP_MAT_to_TAMN_TAMX</moduleInstanceId>
         <valueType>scalar</valueType>
         <parameterId>TAMX</parameterId>
         <locationSetId>Catchments_HEFS</locationSetId>
         <timeSeriesType>external historical</timeSeriesType>
         <timeStep id="12Z"/>
         <relativeViewPeriod unit="day" start="-36500" startOverrulable="true" end="0" endOverrulable="true"/>
         <readWriteMode>read only</readWriteMode>
      </timeSeriesSet>
      <timeSeriesSet>
         <moduleInstanceId>MEFP_MAT_to_TAMN_TAMX</moduleInstanceId>
         <valueType>scalar</valueType>
         <parameterId>TAMN</parameterId>
         <locationSetId>Catchments_HEFS</locationSetId>
         <timeSeriesType>external historical</timeSeriesType>
         <timeStep id="12Z"/>
         <relativeViewPeriod unit="day" start="-36500" startOverrulable="true" end="0" endOverrulable="true"/>
         <readWriteMode>read only</readWriteMode>
      </timeSeriesSet>
   </timeSeries>
</fewsPiServiceConfig>
```

Note that the time series for temperature have parameterIds TAMN and TAMX, whereas MEFPPE expects TMIN and TMAX, respectively. An id-mapping file is used to account for this discrepancy. For example, the configuration file above uses the following id-mapping file (IdExportMEFPPE.xml in the example above):

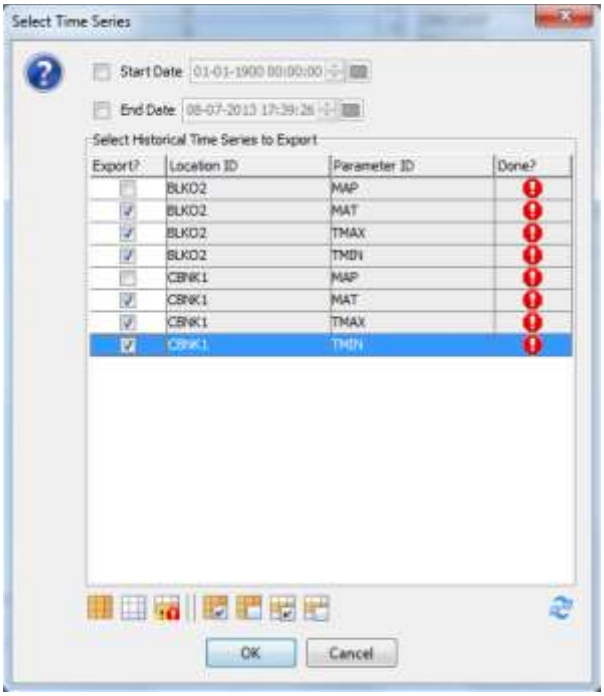
```
<?xml version="1.0" encoding="UTF-8"?>
<idMap version="1.1" xmlns="http://www.wldelft.nl/fews" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.wldelft.nl/fews http://fews.wldelft.nl/schemas/version1.0/idMap.xsd">

        <parameter external="TMAX" internal="TAMX"/>
        <parameter external="TMIN" internal="TAMN"/>
        <enableOneToOneMapping/>
</idMap>
```

Instructions for configuring this file are provided in the *MEFPPE Configuration Guide*, but typically no changes should be necessary to the file delivered with the release.

3.5.3.2 Exporting Historical Data via the FEWS PI-service

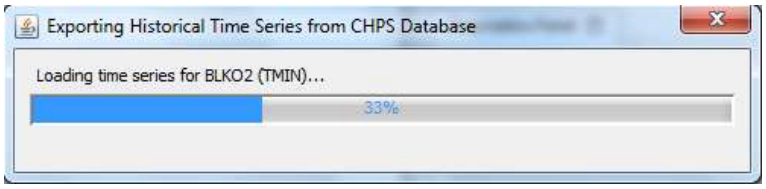
The standard way of extract historical data from the localDataStore for use in estimating MEFP parameters is via the FEWS PI-service. To do so, click on the **Export Time Series from CHPS DB Button**. Upon clicking, MEFPPE will query the FEWS PI-service for a list of time series from which the user can make selections, displaying a progress dialog while the query is being performed. When the PI-service returns the needed information, the following **Select Time Series Dialog** will open:



The 'Export?' check box column for each row of the **Select Historical Time Series to Export Table** indicates if the time series will be acquired from the PI-service. The 'Done?' column indicates if data already exists in the MEFPPE run area for that location.

Select the time series to export by checking on the checkbox in the 'Export?' column of the rows for each desired location and parameter. The buttons in the toolbar below the table can be used

to check/uncheck all rows or check/uncheck all selected rows. Specify the date range of the data to acquire by checking on the **Start Date** and/or **End Date Checkboxes** and specifying the start and/or end dates of the acquired time series in the date fields. When all selections are made and the date range specified (if necessary), click on the **OK Button**. The MEFPPE will then extract the time series and export them for use by MEFPPE, displaying a progress dialog:



Upon completion, the progress dialog will close and the **By Identifier Tree/By Source Tree** within the **Historical Data SubPanel** will update to reflect the new files. Also, any new locations for which time series required for parameter estimation are now available will appear in the **Location Summary Panel** in the upper left corner of the **MEFPPE Main Panel**.

NOTE: If only one of the date range checkboxes is checked, then one end of the acquired time series will not be limited. For example, if the **Start Date Checkbox** *is checked by not the* **End Date Checkbox**, *MEFPPE will acquire all data* <u>*after*</u> *the specified start date.*

3.5.3.3 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 and place it within the directory:

…/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).

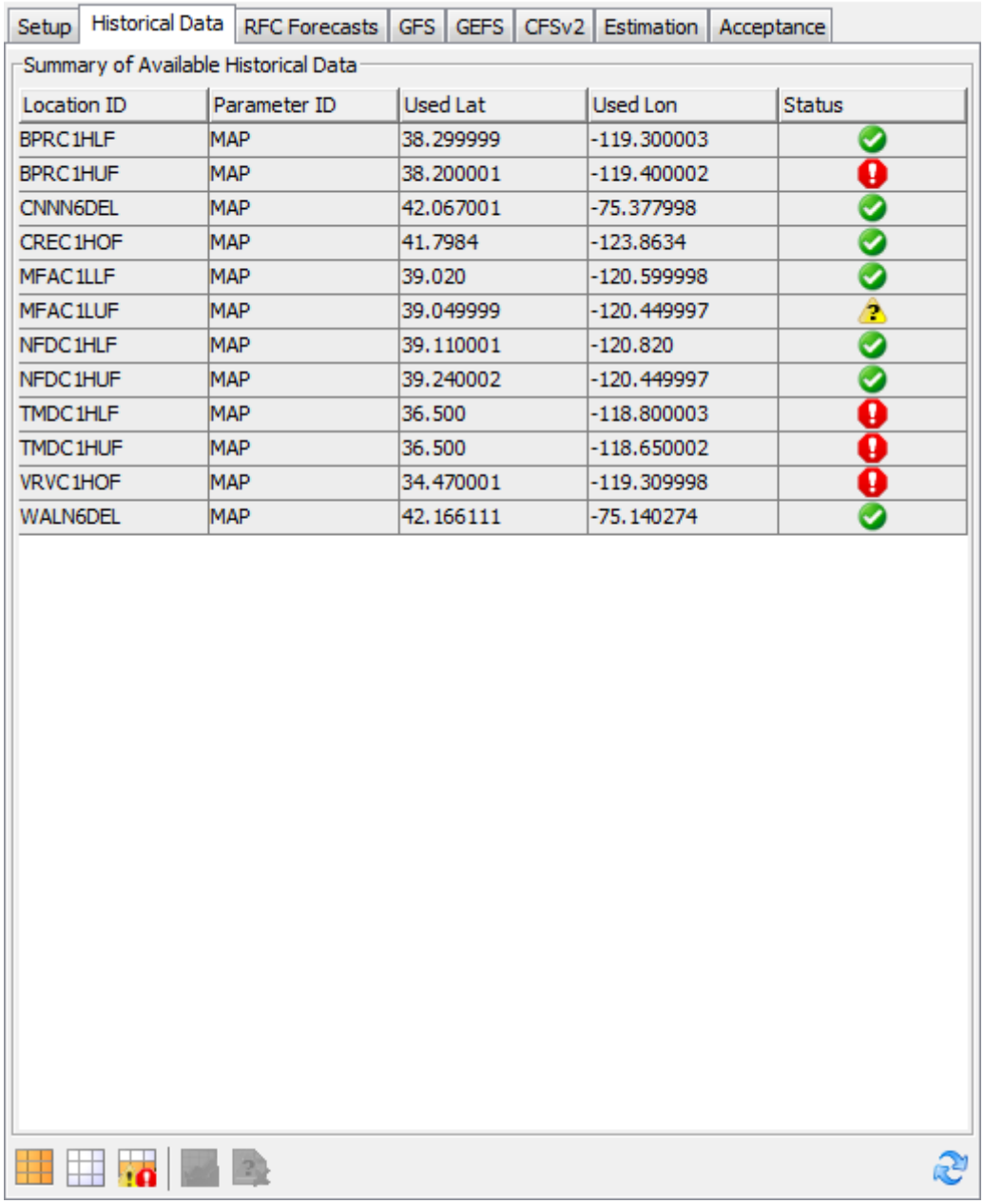
3.5.3.4 Adding a New Canonical Event

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 5 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
BPRC1HLF	MAP	38.299999	-119.300003	
BPRC1HUF	MAP	38.200001	-119.400002	
CNNN6DEL	MAP	42.067001	-75.377998	
CREC1HOF	MAP	41.7984	-123.8634	
MFAC1LLF	MAP	39.020	-120.599998	
MFAC1LUF	MAP	39.049999	-120.449997	
NFDC1HLF	MAP	39.110001	-120.820	
NFDC1HUF	MAP	39.240002	-120.449997	
TMDC1HLF	MAP	36.500	-118.800003	
TMDC1HUF	MAP	36.500	-118.650002	
VRVC1HOF	MAP	34.470001	-119.309998	
WALN6DEL	MAP	42.166111	-75.140274	

Figure 5: 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

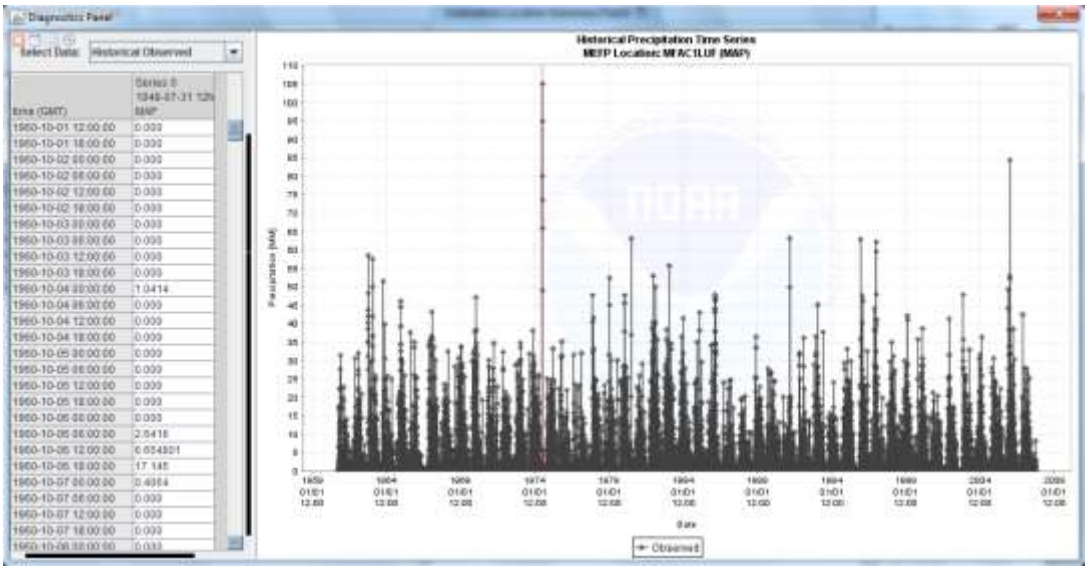
performed. It is a **Generic Summary Table** (Section 3.2.5.1) and includes all the standard buttons.

- **View Button**: Displays diagnostics for a single MEFP location selected in the **Summary of Available Historical Data Table**.
- **Remove Questionable Data File Button**: Removes the questionable message file (see Section 3.2.4.6), if one exists, for each of the selected locations so that those locations are no longer marked by a questionable status icon, . This button is enabled only when at least one selected location is marked as questionable. Click on this button when it has been determined that all questionable data for a location is acceptable.

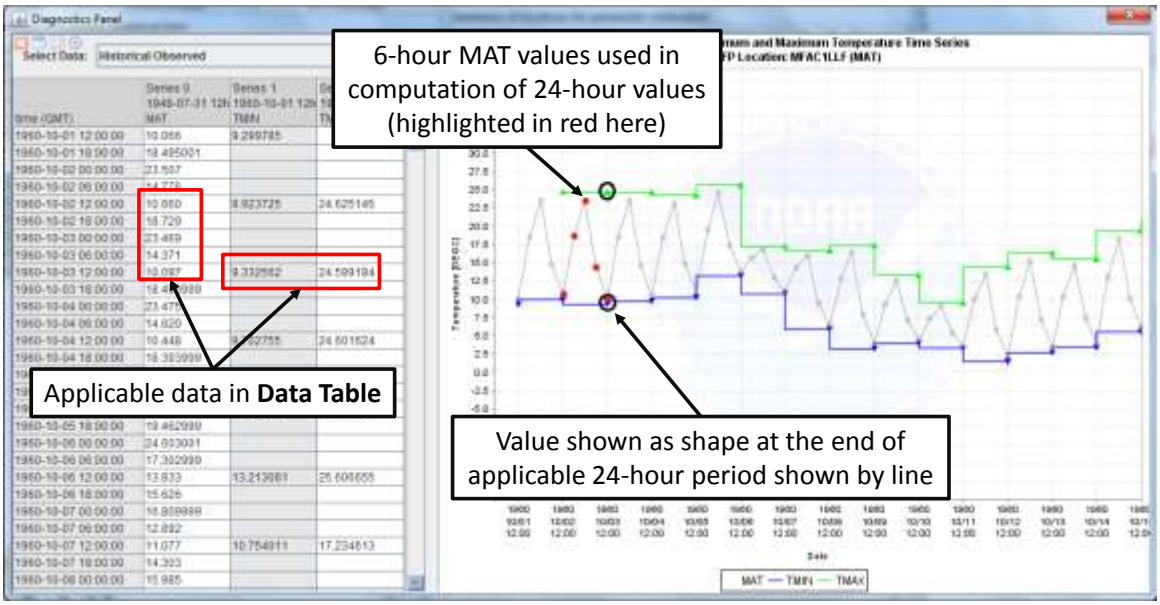
3.6.2 Diagnostics

This uses a standard time series **Diagnostics Panel** (Section 3.14) displaying the same data as that in the **Export Historical Data Subpanel** (Section 3.5.1.2), but with questionable data shown and no components for selecting years and forecast time (since reforecast data is not shown). This panel is designed to aid the user in identifying questionable data and determining if it is acceptable. See Section 3.2.4.6 for information on questionable data and how to correct problematic time series values. See Section 3.14.4 for a process to follow.

For precipitation, the **Diagnostics Panel** shows the historical precipitation time series:



For temperature, 24-hour minimum and maximum temperatures are displayed as a step plot, with a shape noting the 24-hour data value and the line (step) preceding it indicating the 24-hour period that applies to that value. If 6-hour MAT data is available (acquired via the **Export Historical Data Subpanel**, see Section 3.5.1), it will also be shown as light gray circles connected by lines. For example (zoomed in):



3.6.3 Usage

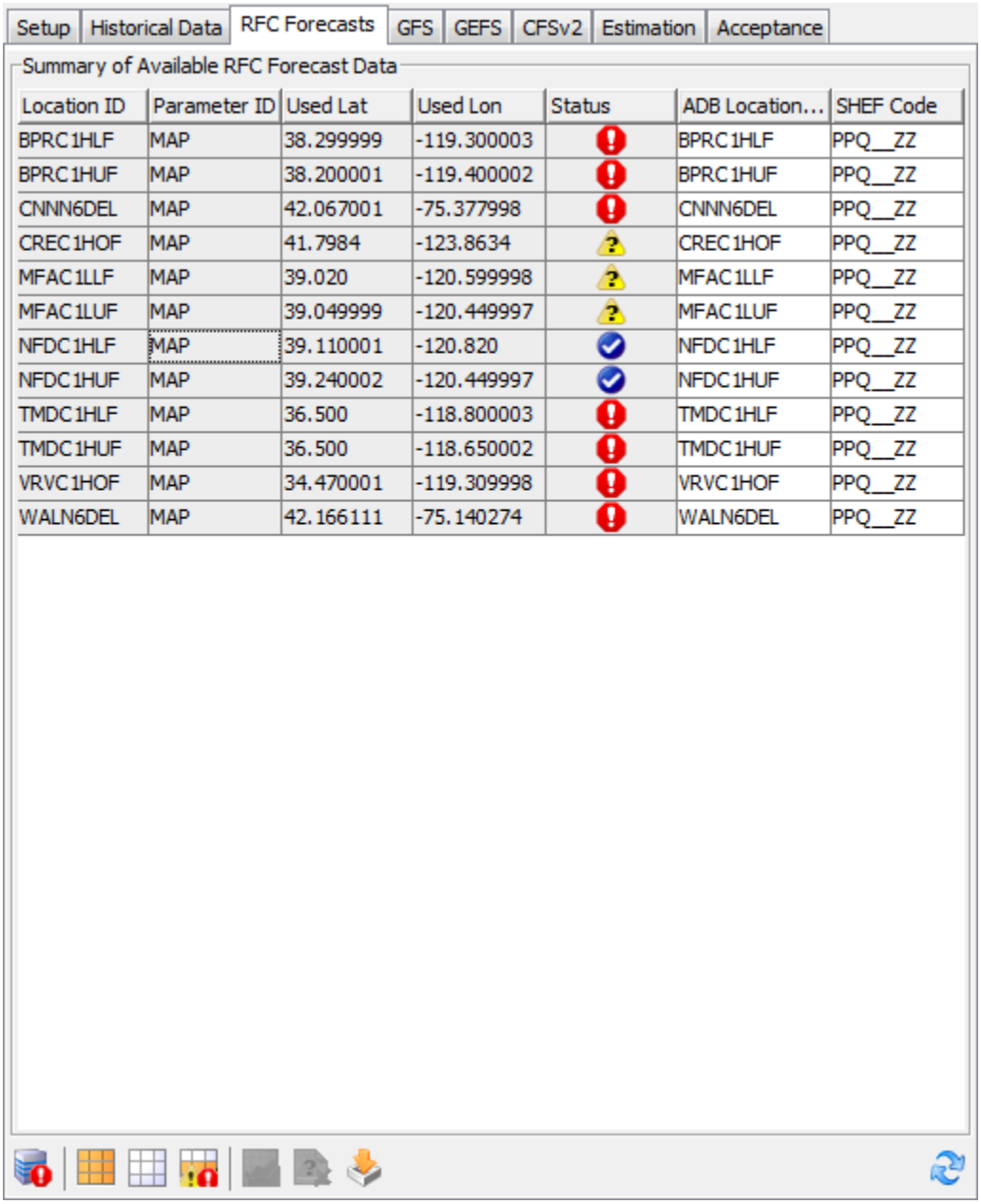
3.6.3.1 Generating Binary Data and Questionable Message Files

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**. Upon creating each binary file, the time series are examined by MEFPPE to identify questionable data values and a questionable message file is created if necessary (see Section 3.2.4.6).

3.7 RFC Forecasts Subpanel

The **RFC Forecasts Subpanel** of the **Estimation Steps Panel**, shown in Figure 6, 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.

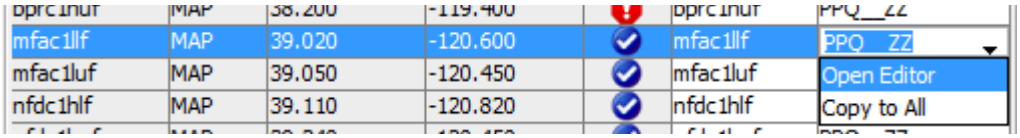


Location ID	Parameter ID	Used Lat	Used Lon	Status	ADB Location...	SHEF Code
BPRC1HLF	MAP	38.299999	-119.300003		BPRC1HLF	PPQ__ZZ
BPRC1HUF	MAP	38.200001	-119.400002		BPRC1HUF	PPQ__ZZ
CNNN6DEL	MAP	42.067001	-75.377998		CNNN6DEL	PPQ__ZZ
CREC1HOF	MAP	41.7984	-123.8634		CREC1HOF	PPQ__ZZ
MFAC1LLF	MAP	39.020	-120.599998		MFAC1LLF	PPQ__ZZ
MFAC1LUF	MAP	39.049999	-120.449997		MFAC1LUF	PPQ__ZZ
NFDC1HLF	MAP	39.110001	-120.820		NFDC1HLF	PPQ__ZZ
NFDC1HUF	MAP	39.240002	-120.449997		NFDC1HUF	PPQ__ZZ
TMDC1HLF	MAP	36.500	-118.800003		TMDC1HLF	PPQ__ZZ
TMDC1HUF	MAP	36.500	-118.650002		TMDC1HUF	PPQ__ZZ
VRVC1HOF	MAP	34.470001	-119.309998		VRVC1HOF	PPQ__ZZ
WALN6DEL	MAP	42.166111	-75.140274		WALN6DEL	PPQ__ZZ

Figure 6: The **RFC Forecasts Subpanel** of the **Estimation Steps Panel**.

3.7.1 Components

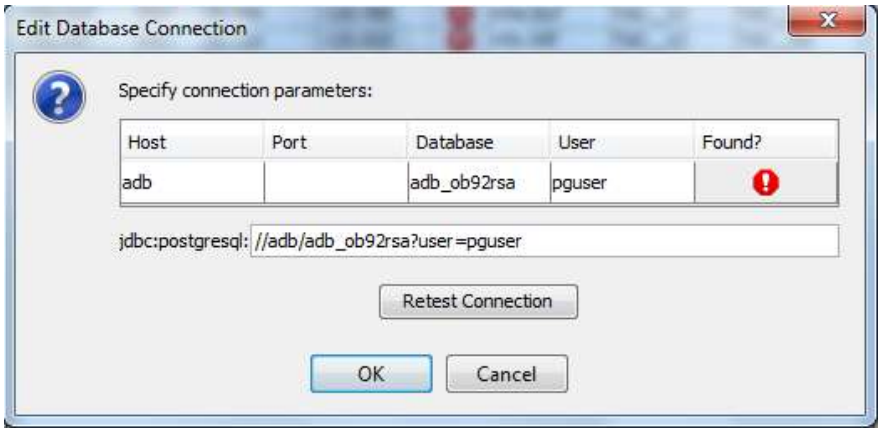
- **Summary of Available RFC Forecast Data Table**: Allows the user to select the MEFP locations for which to perform the step and view for which locations the step has been performed. It is a **Generic Summary Table** (Section 3.2.5.1) and includes all the standard columns and buttons. Note that, for 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:



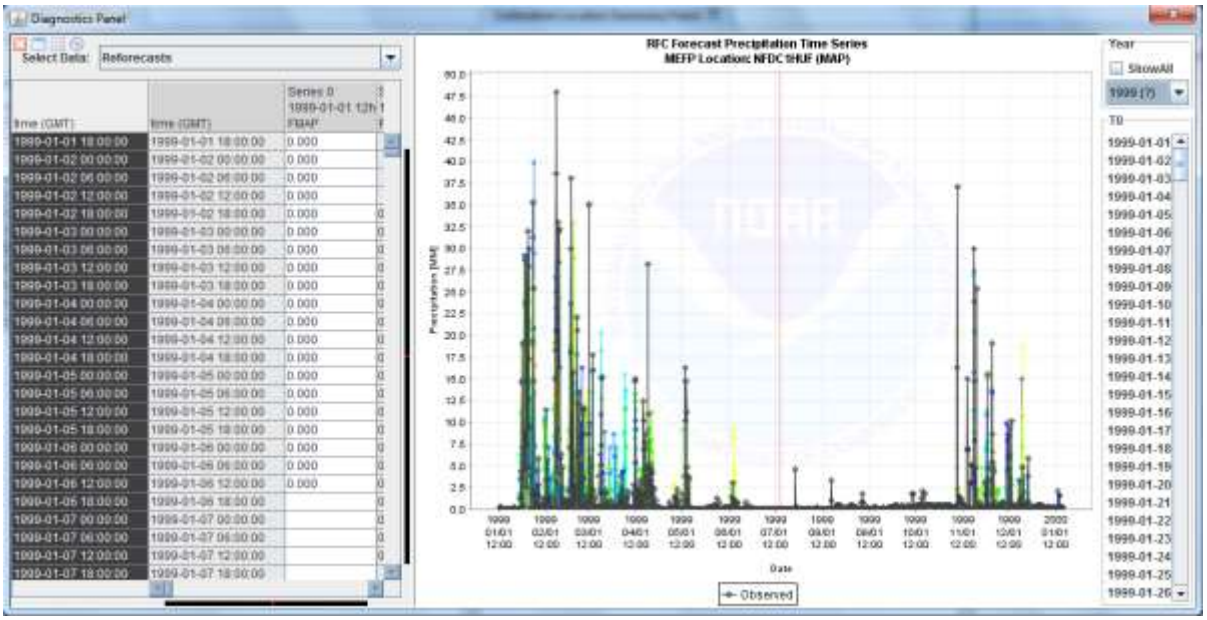
By default, all cells of the **Specify Connection Parameters Table** will be empty, except for the 'Database' cell which will be filled with a default value of "DATABASENAME". For more details on how to connect to the archive database, see Section 3.7.3.2.

- **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.
- **Remove Questionable Data File Button**: Removes the file that stores information about questionable data, if one exists, for each of the selected locations so that those locations are no longer marked by a questionable status icon, . This button is enabled only when at least one selected location is marked as questionable. Click on this button when it has been determined that all questionable data for a location is acceptable.

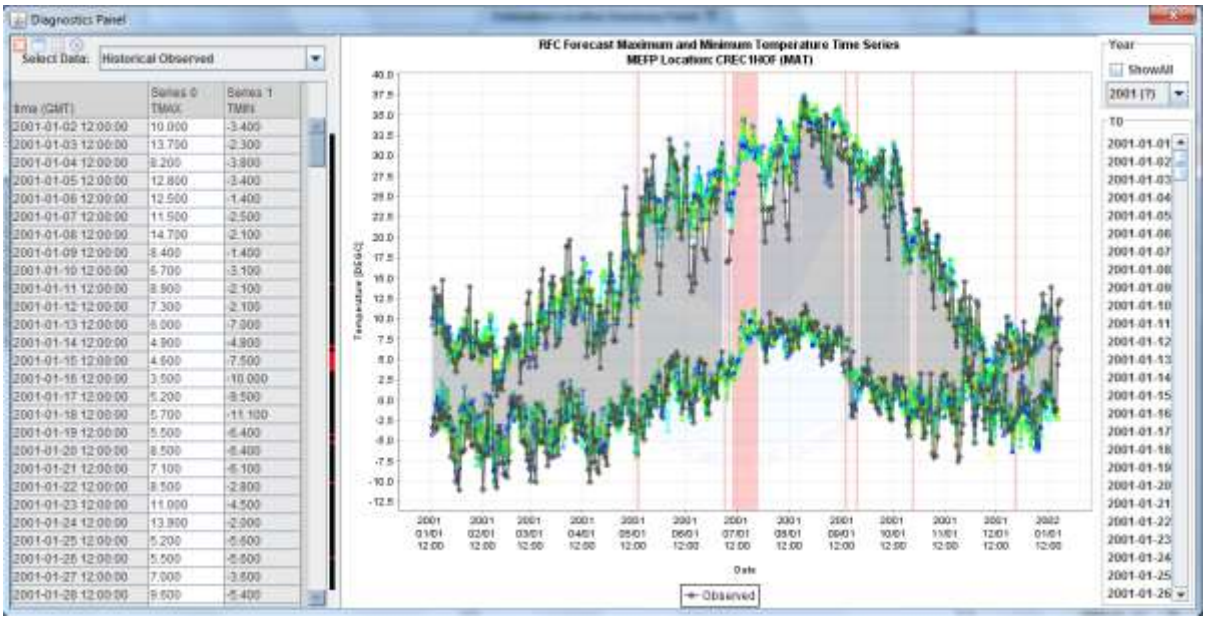
3.7.2 Diagnostics

This is a standard time series **Diagnostics Panel** (Section 3.14) that, for a selected location, displays the archived RFC forecasts and observed data either acquired from the archive database or imported from manually constructed files. This panel is designed to aid the user in identifying questionable data and determining if it is acceptable. See Section 3.2.4.6 for information on questionable data and how to correct problematic time series values. See Section 3.14.4 for a process to follow.

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



For temperature, the displayed forecast and observed time series are 24-hour minimum and maximum temperatures:



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 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 data 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 "TA" 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 to allow for building pairs of forecasts and observations. For instructions on using the Vfyruninfo Editor, see 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 starting Jan 1, 2000, 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:

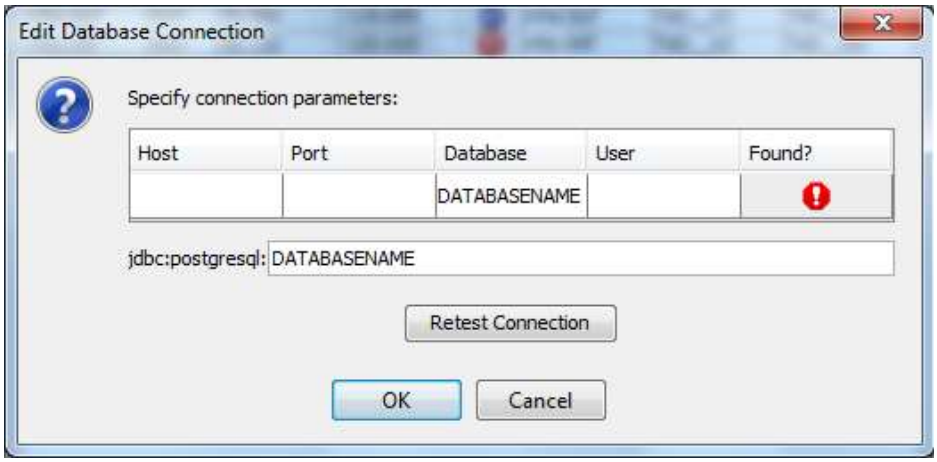
   http://www.nws.noaa.gov/oh/hrl/verification/ob8/PairingBatch.pdf

   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, and 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' cells 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. As each file is created, if all necessary files are available for the location and

current data type (precipitation or temperature), those files are examined by MEFPPE to identify questionable data values and a questionable message file is created if necessary (see Section 3.2.4.6).

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.

When the files are created, click on the **Import Prepared RFC Forecast Files Button** to import the files. Each file is imported by copying the file, one at a time, to its location within the MEFPPE run area. Upon copying each file, if the all of the observed and archived forecast files necessary for that location and data type are available, then those files are examined by MEFPPE to identify questionable data values and a questionable message file is created if necessary (see Section 3.2.4.6).

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 describe 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 required 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" (this property is required; there is no default value).
- `nfcstdays`: The number of days of forecast values. A positive integer not greater than 5. (this property is required; there is no default value).
- `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
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 a decimal and two decimal places. There must be a number of forecast values equal to `nfcstdays` (see above) 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   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   0.07   0.07   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
0.09   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
...
```

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" (there is no default; this 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
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 a decimal and 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 (can be copied-and-pasted and modified to setup header lines):

```
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 **<u>in degrees Celsius</u>** (<> 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 a decimal and 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 (can be copied-and-pasted and modified to setup header lines):

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

Data:
Every line of the file must specify a single forecast time series **<u>in degrees Celsius</u>** (<> 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 9th character, one observed value is provided that is 6 characters wide allowing for a decimal and one decimal place.

Example:

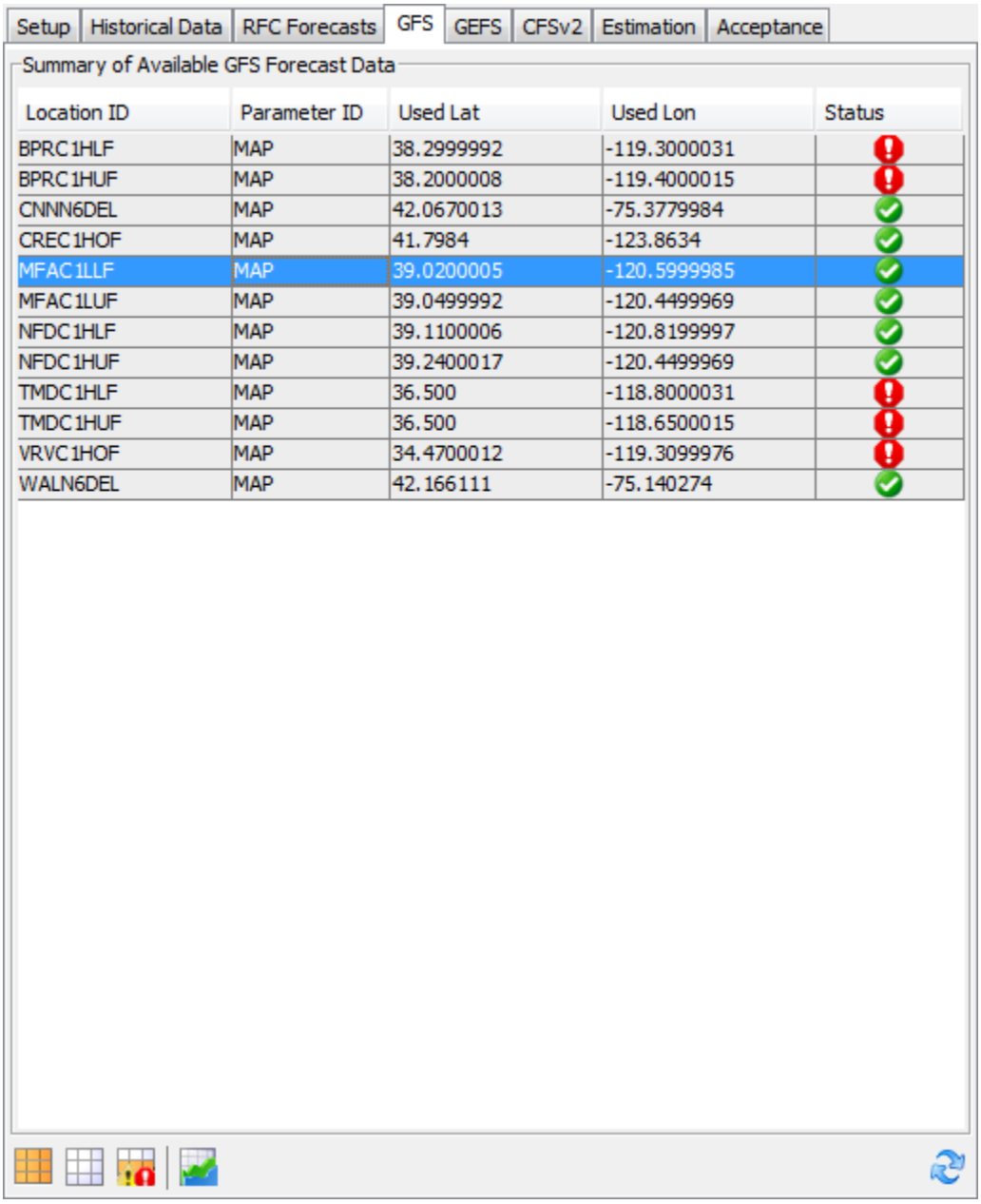
```
20010101 -13.8
20010102 -13.2
20010103 -14.9
```

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 7, 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
BPRC1HLF	MAP	38.2999992	-119.3000031	!
BPRC1HUF	MAP	38.2000008	-119.4000015	!
CNNN6DEL	MAP	42.0670013	-75.3779984	✓
CREC1HOF	MAP	41.7984	-123.8634	✓
MFAC1LLF	MAP	39.0200005	-120.5999985	✓
MFAC1LUF	MAP	39.0499992	-120.4499969	✓
NFDC1HLF	MAP	39.1100006	-120.8199997	✓
NFDC1HUF	MAP	39.2400017	-120.4499969	✓
TMDC1HLF	MAP	36.500	-118.8000031	!
TMDC1HUF	MAP	36.500	-118.6500015	!
VRVC1HOF	MAP	34.4700012	-119.3099976	!
WALN6DEL	MAP	42.166111	-75.140274	✓

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

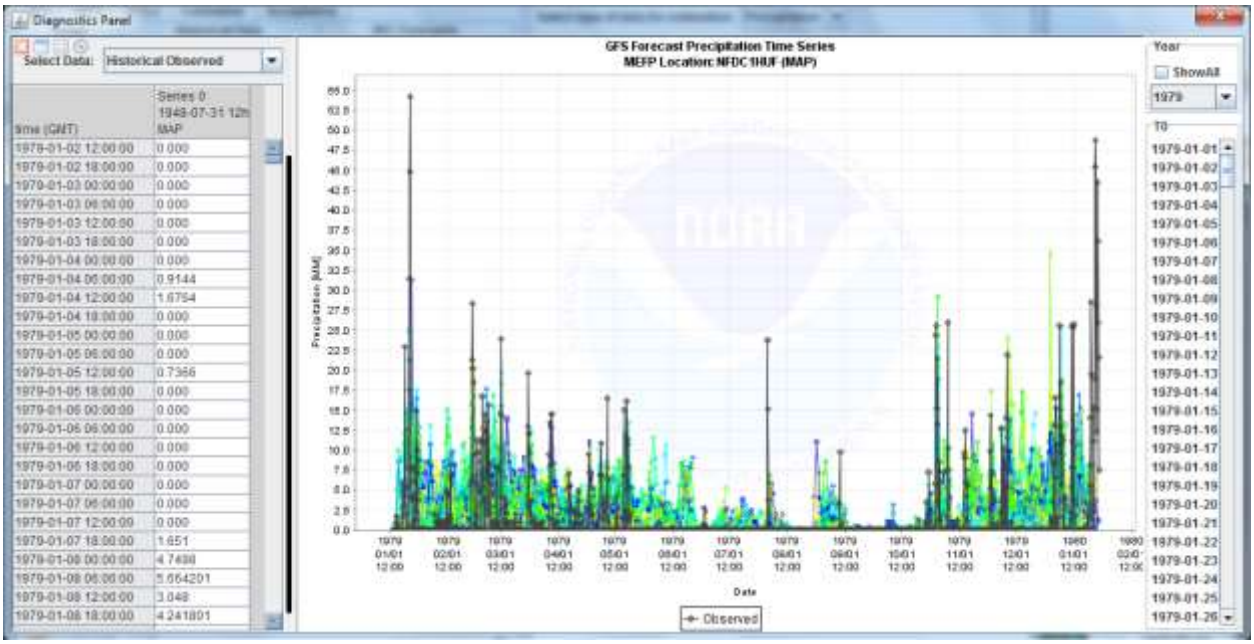
3.8.1 Components

- **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**.

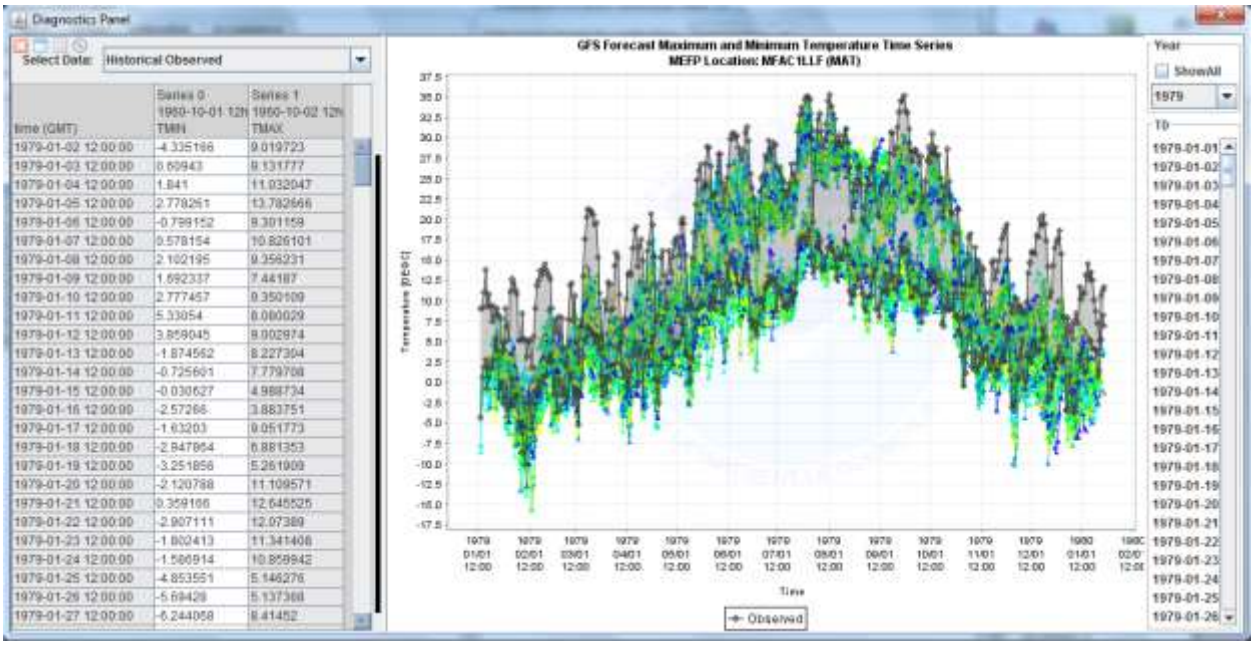
3.8.2 Diagnostics

This uses a standard time series **Diagnostics Panel** (Section 3.14) that does not include checks for questionable data; the data is quality controlled at OHD prior to making it available to RFCs. As with all of the standard time series **Diagnostics Panels**, historical time series are displayed in addition to the reforecast time series. The historical data is the same as that which is used in estimating parameters: it is a combination of the historical data shown in the **Diagnostics Panel** within the **Historical Data Subpanel** and the observed data specified with RFC QPF/QTF and shown in the **RFC Forecast Subpanel**. The RFC observed data, if available, extends the historical data and is used to fill in missing values. Use the **Diagnostics Panel** in those two subpanels to quality control the historical.

For example, this is a **Diagnostics Panel** for precipitation data:



For temperature, the displayed forecast and observed time series are 24-hour minimum and maximum temperatures:



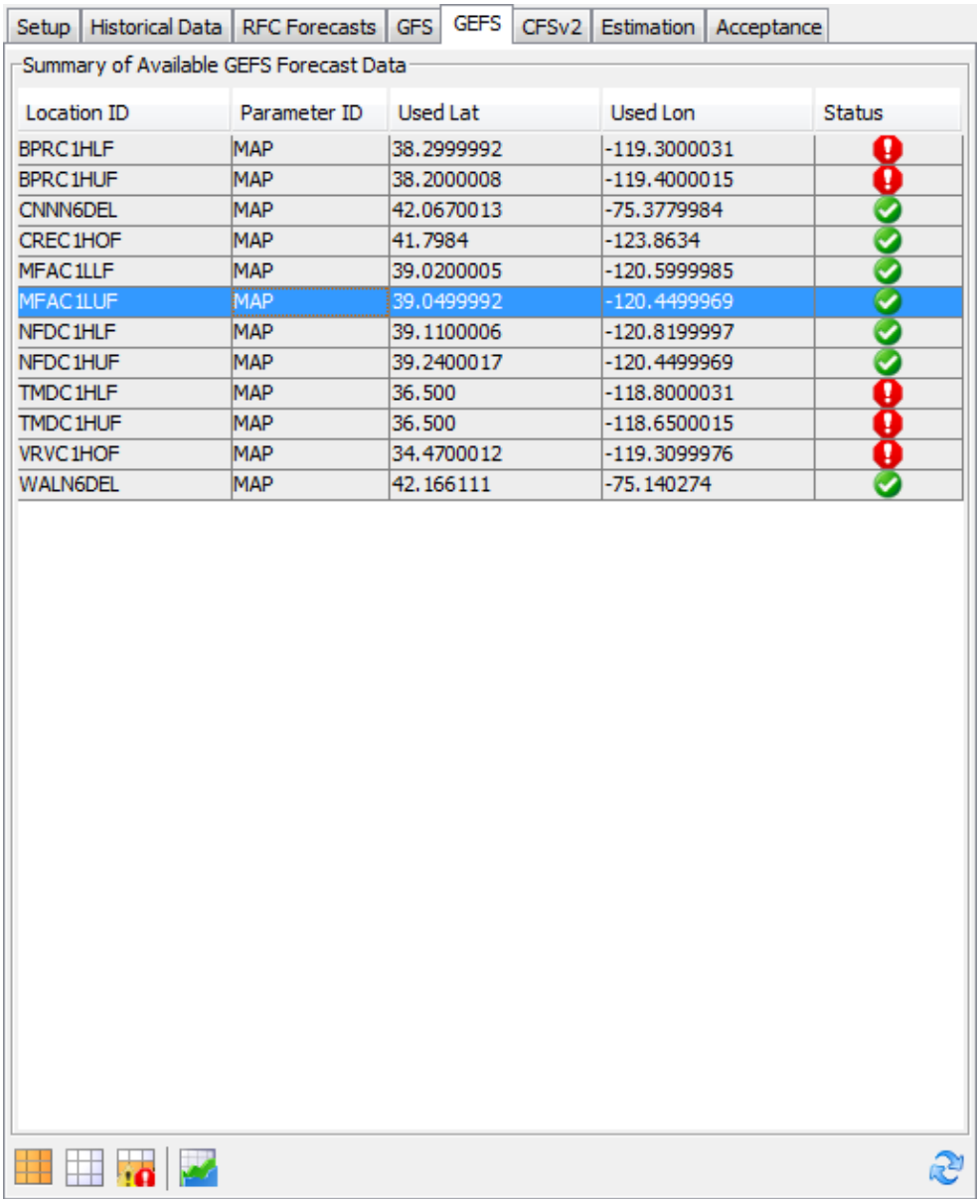
3.8.3 Usage

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 GEFS Subpanel

The **GEFS Subpanel** of the **Estimation Steps Panel**, shown in Figure 7, is used to perform Step 5 of the parameter estimation procedure in Section 3.2.3: acquire GEFS reforecast data files. For precipitation, the GEFS reforecast files specify the complete reforecast ensembles, which are for the accumulated precipitation amounts at 6-hour time steps generated at 0Z of each reforecast date and for 16 lead days. From this ensemble, an ensemble mean is computed for a 12Z forecast, so that only 15 lead days are available. For temperature, the GEFS reforecast files specify the same ensembles, except that the variables are 6-hour minimum and maximum temperature. From this ensemble, a mean is computed for a 12Z forecast of the 24-hour minimum and maximum temperature, so that only 15 days are available.



Location ID	Parameter ID	Used Lat	Used Lon	Status
BPRC1HLF	MAP	38.2999992	-119.3000031	!
BPRC1HUF	MAP	38.2000008	-119.4000015	!
CNNN6DEL	MAP	42.0670013	-75.3779984	✓
CREC1HOF	MAP	41.7984	-123.8634	✓
MFAC1LLF	MAP	39.0200005	-120.5999985	✓
MFAC1LUF	MAP	39.0499992	-120.4499969	✓
NFDC1HLF	MAP	39.1100006	-120.8199997	✓
NFDC1HUF	MAP	39.2400017	-120.4499969	✓
TMDC1HLF	MAP	36.500	-118.8000031	!
TMDC1HUF	MAP	36.500	-118.6500015	!
VRVC1HOF	MAP	34.4700012	-119.3099976	!
WALN6DEL	MAP	42.166111	-75.140274	✓

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

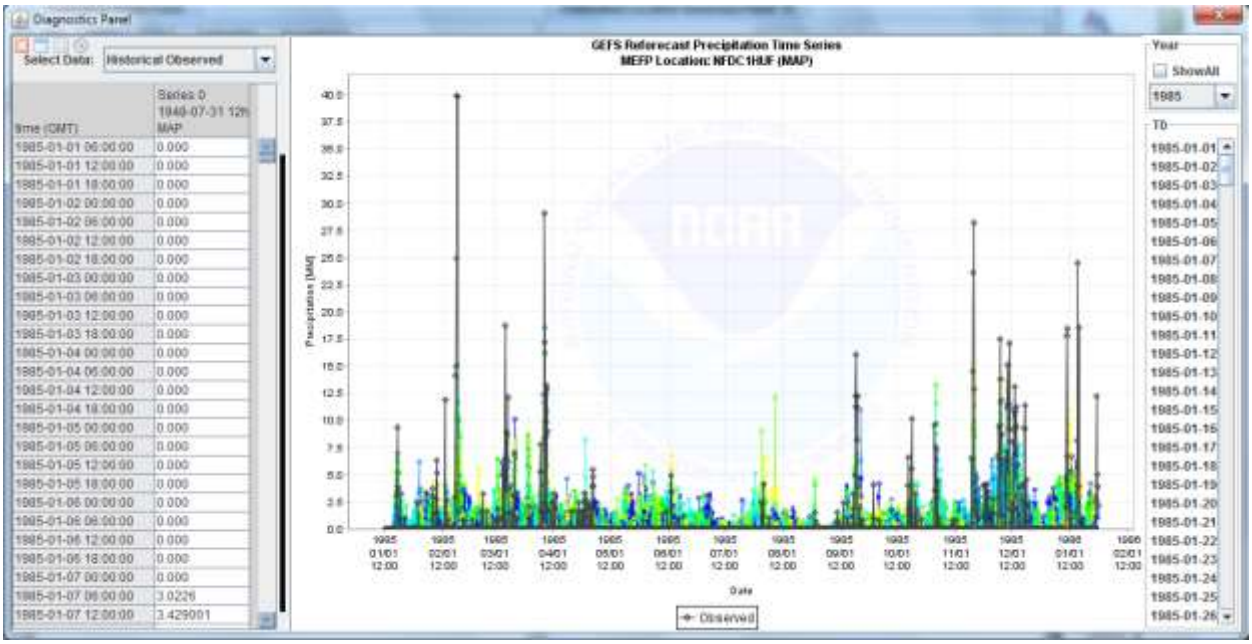
3.9.1 Components

- **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**.

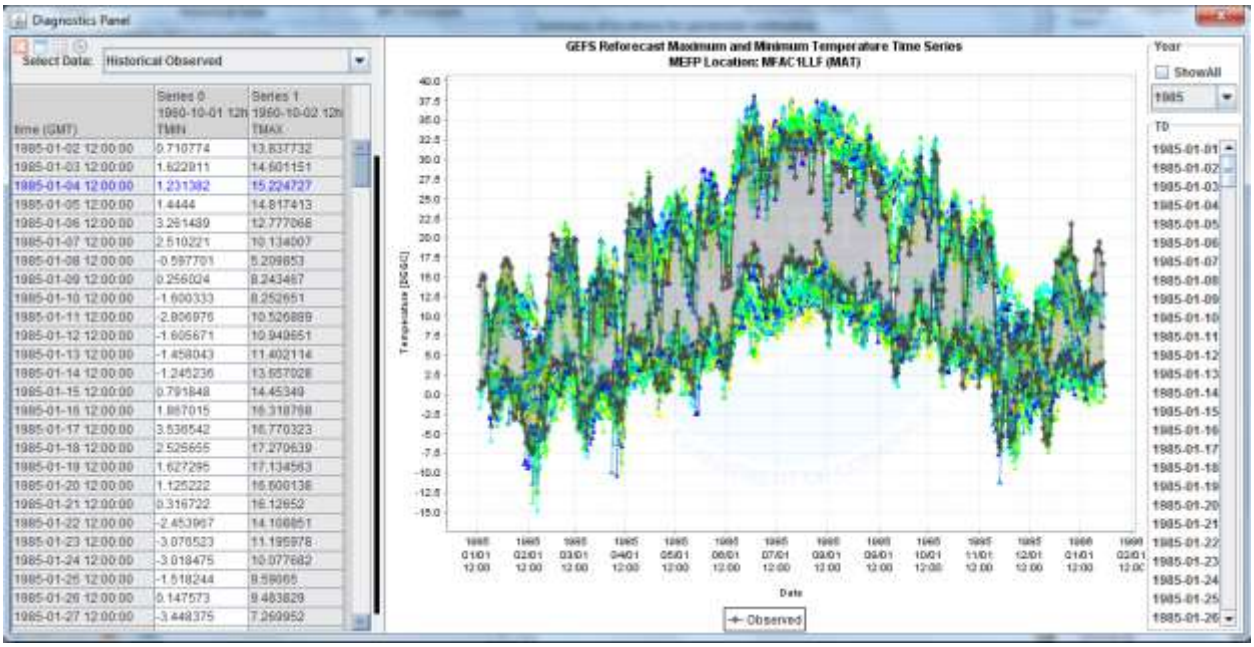
3.9.2 Diagnostics

This uses a standard time series **Diagnostics Panel** (Section 3.14) that does not include checks for questionable data; the data is quality controlled at OHD prior to making it available to RFCs. As with all of the standard time series **Diagnostics Panels**, historical time series are displayed in addition to the reforecast time series. The historical data is the same as that which is used in estimating parameters: it is a combination of the historical data shown in the **Diagnostics Panel** within the **Historical Data Subpanel** and the observed data specified with RFC QPF/QTF and shown in the **RFC Forecast Subpanel**. The RFC observed data, if available, extends the historical data and is used to fill in missing values. Use the **Diagnostics Panel** in those two subpanels to quality control the historical data.

For example, this is a **Diagnostics Panel** for precipitation data:



For temperature, the displayed forecast and observed time series are 24-hour minimum and maximum temperatures:



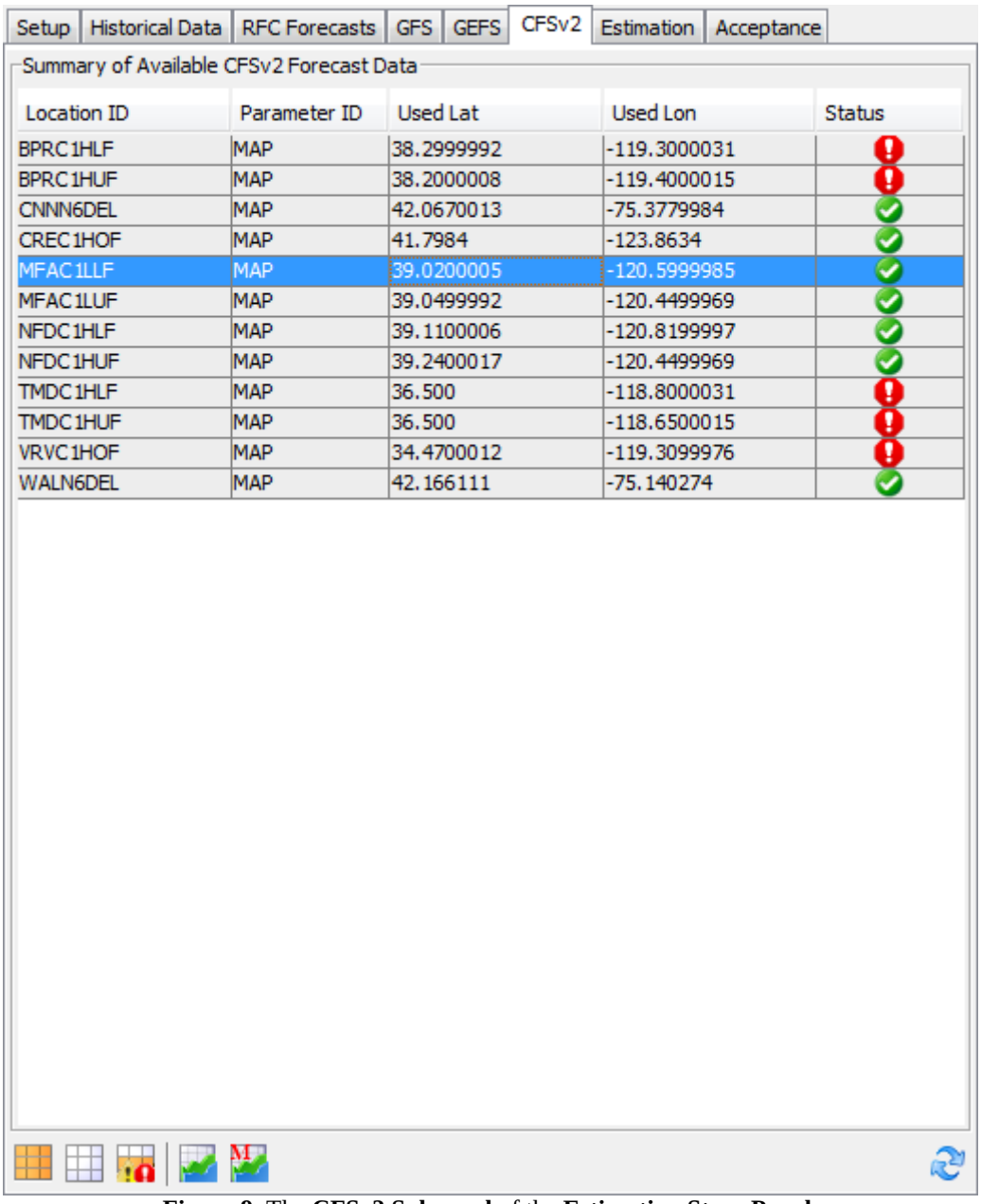
3.9.3 Usage

3.9.3.1 Downloading GEFS Reforecast Files

GEFS 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 GEFS 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 CFSv2 Subpanel

The **CFSv2 Subpanel** of the **Estimation Steps Panel**, shown in Figure 9, is used to perform Step 6 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. For temperature, the CFSv2 reforecast files specify the time series of minimum and maximum temperatures. For both data types, submonthly data is provided at a 6-hour time step and monthly data at a 30-day time step.



Location ID	Parameter ID	Used Lat	Used Lon	Status
BPRC1HLF	MAP	38.2999992	-119.3000031	!
BPRC1HUF	MAP	38.2000008	-119.4000015	!
CNNN6DEL	MAP	42.0670013	-75.3779984	✓
CREC1HOF	MAP	41.7984	-123.8634	✓
MFAC1LLF	MAP	39.0200005	-120.5999985	✓
MFAC1LUF	MAP	39.0499992	-120.4499969	✓
NFDC1HLF	MAP	39.1100006	-120.8199997	✓
NFDC1HUF	MAP	39.2400017	-120.4499969	✓
TMDC1HLF	MAP	36.500	-118.8000031	!
TMDC1HUF	MAP	36.500	-118.6500015	!
VRVC1HOF	MAP	34.4700012	-119.3099976	!
WALN6DEL	MAP	42.166111	-75.140274	✓

Figure 9: The **CFSv2 Subpanel** of the **Estimation Steps Panel**.

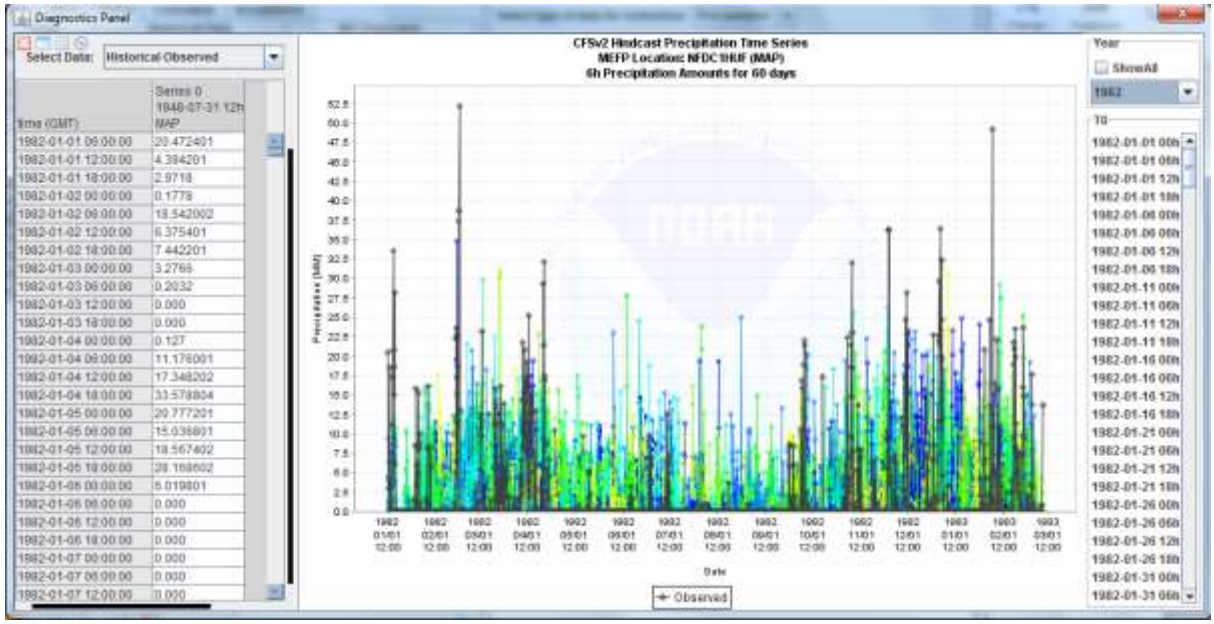
3.10.1 Components

- **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**.

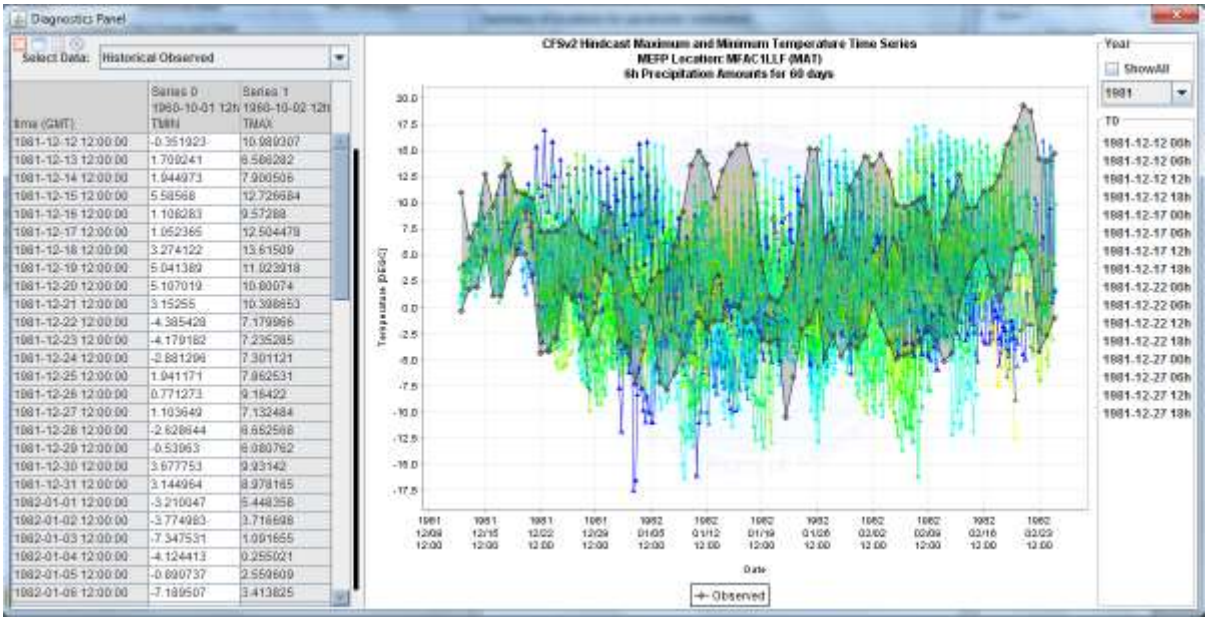
3.10.2 Diagnostics

There are two types of **Diagnostics Panels** available for CFSv2: a panel to view submonthly time series (reforecasts displayed in original time step out to 30 days), and a panel to view monthly time series (reforecasts aggregated to 30-day time steps out to 270 days).

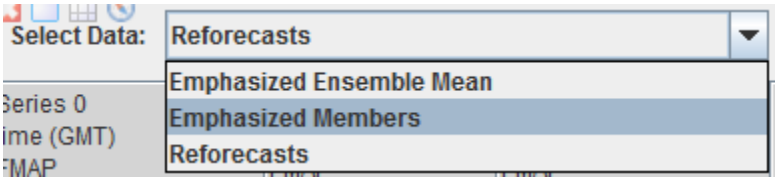
The **View Submonthly Button** opens a submonthly **Diagnostics Panel**, which is a standard time series **Diagnostics Panel** (Section 3.14) that does not include checks for questionable data; the data is quality controlled at OHD prior to making it available to RFCs. As with all of the standard time series **Diagnostics Panels**, historical time series are displayed in addition to the reforecast time series. The historical data is the same as that which is used in estimating parameters: it is a combination of the historical data shown in the **Diagnostics Panel** within the **Historical Data Subpanel** and the observed data specified with RFC QPF/QTF and shown in the **RFC Forecast Subpanel**. The RFC observed data, if available, extends the historical data and is used to fill in missing values. Use the **Diagnostics Panel** in those two subpanels to quality control the historical data. The submonthly reforecast time series are available every 5 days and four times each day (hours 0, 6, 12, and 18Z) so that the list of available forecast times will include the hour in GMT. Here is an example for precipitation data:



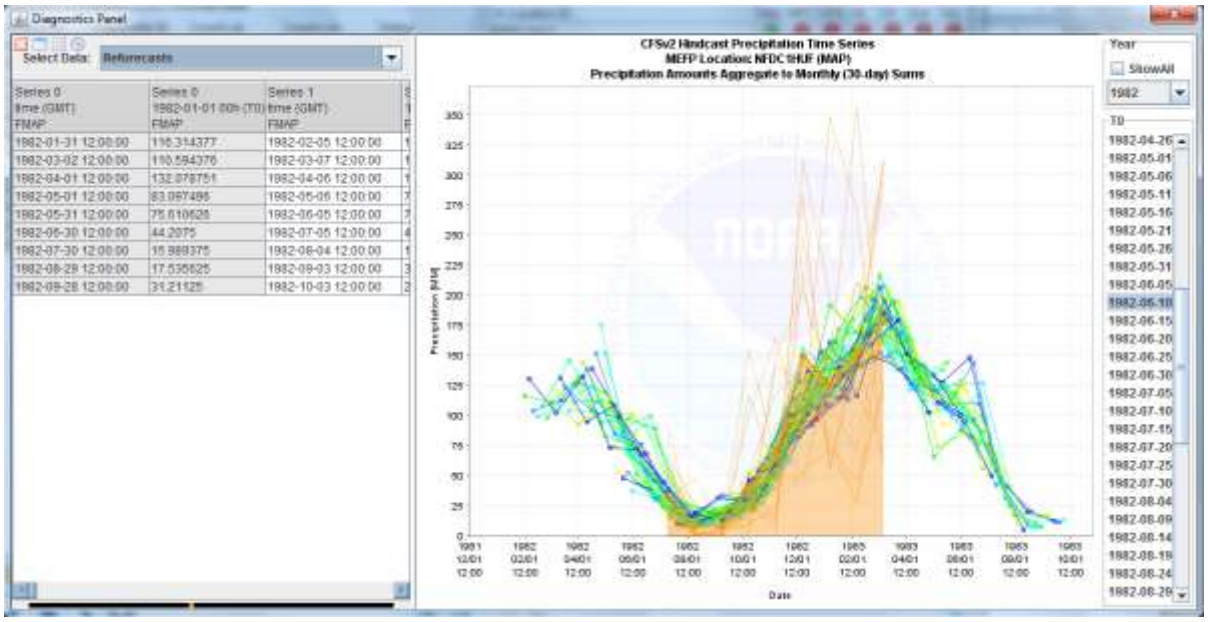
Here is an example for temperature data:



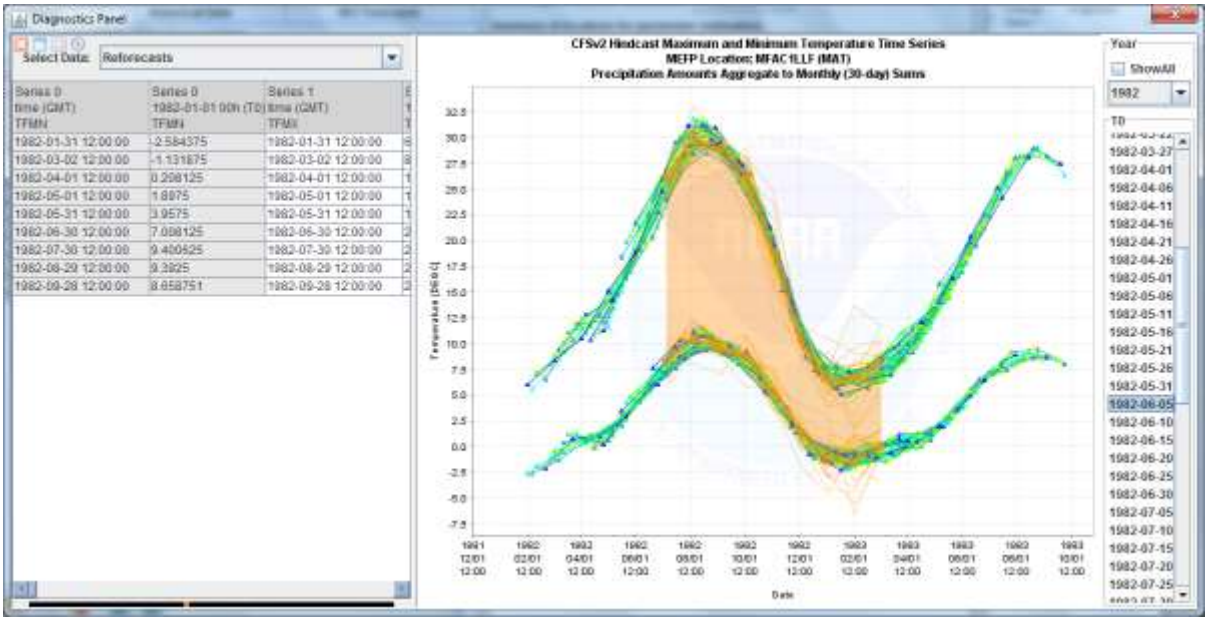
The **View Monthly Button** opens a monthly **Diagnostics Panel**, which is a standard time series **Diagnostics Panel** (Section 3.14) that does not include checks for questionable data; the data is quality controlled at OHD prior to making it available to RFCs. Unlike other time series **Diagnostics Panels**, historical time series are not displayed due to performance concerns. An additional feature provided in this plot, however, is that when a series is emphasized, the 15-members of the lagged ensemble used to compute the ensemble mean for that forecast time are shown; the examples below will include emphasized time series. Furthermore, the **Select Data Choice Box** will make the members available in the **Data Table** via the choice "Emphasized Members":



For precipitation, the monthly **Diagnostics Panel** displays 30-day accumulated precipitation forecast time series. For example:



For temperature, the monthly **Diagnostics Panel** displays 30-day overall minimum and maximum temperature time series. For example:



3.10.3 Usage

3.10.3.1 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**.

3.11 Estimation Subpanel

The **Estimation Subpanel** of the **Estimation Steps Panel**, shown in Figure 10, is used to perform Step 7 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.

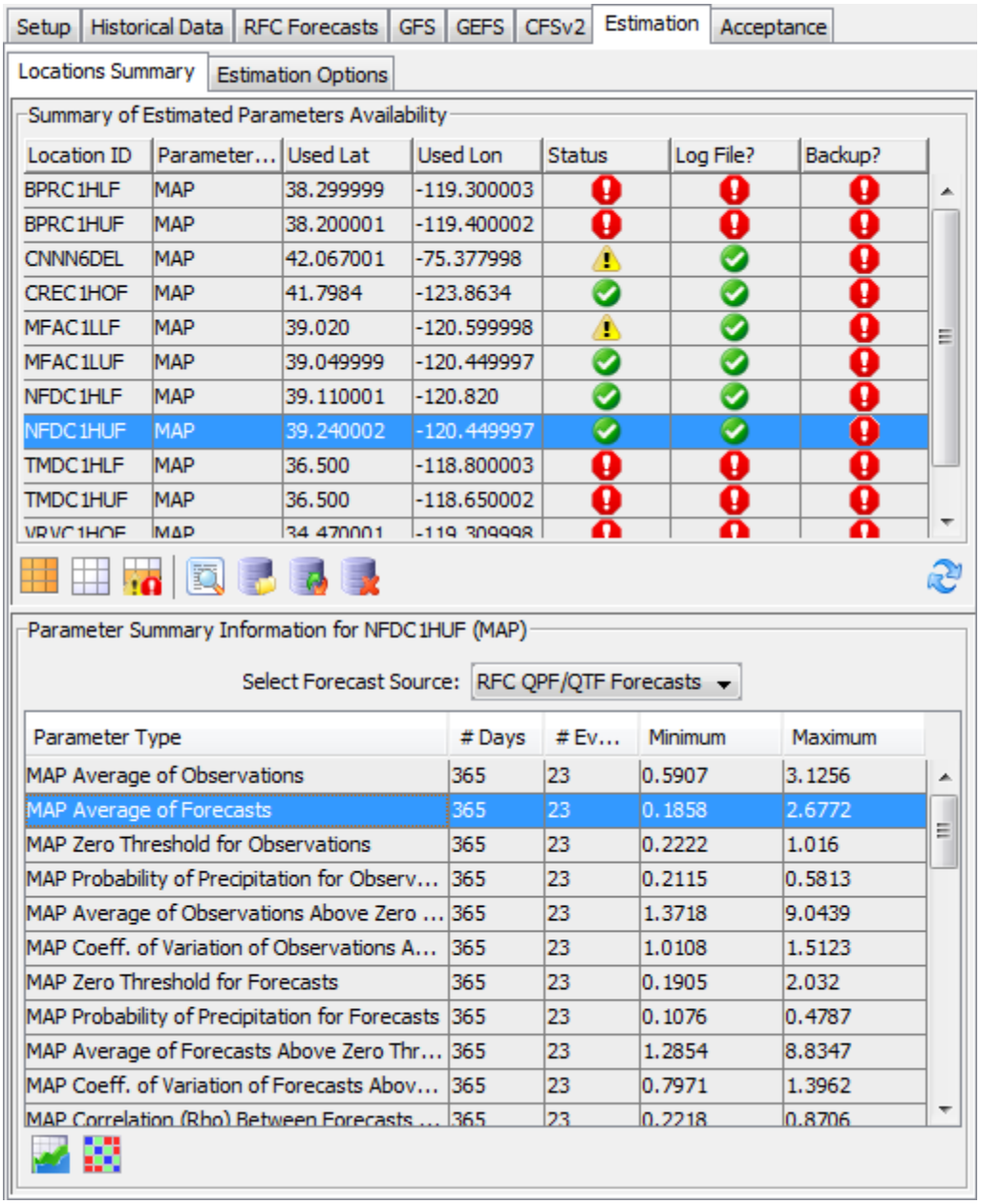


Figure 10: The **Estimation Subpanel** of the **Estimation Steps Panel**.

3.11.1 Locations Summary Subpanel

The **Locations Summary Subpanel**, shown in Figure 10, 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.

3.11.1.1 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 or used incorrect options.

3.11.1.2 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.
- **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 open a **Standard Diagnostics Panel** for the parameters currently selected in the **Parameter Summary Information Table**.
- **View Parameter Diagnostics Button**: Click to open a **Parameters Diagnostics Panel** for the currently loaded parameters.

3.11.2 Estimation Options Subpanel

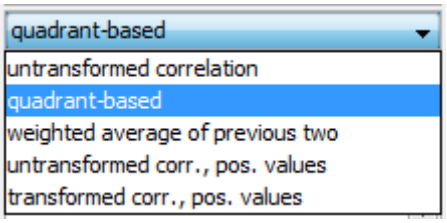
The **Estimation Options Subpanel** allows for users to specify options for the parameter estimation algorithm. There are three groups of parameters:

1. General options used to estimate parameters for the implicit precipitation treatment algorithm.
2. Parameters specific to the type of model used to handle the rain/no-rain precipitation events: the implicit precipitation treatment (IPT) algorithm and the explicit precipitation treatment (EPT) algorithm (see Section 2.4.4).
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:



3.11.2.1 General Options

The general options for the MEFP parameter estimation algorithm that are used to estimate the parameters for both the IPT or EPT models are shown in Figure 11. 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.

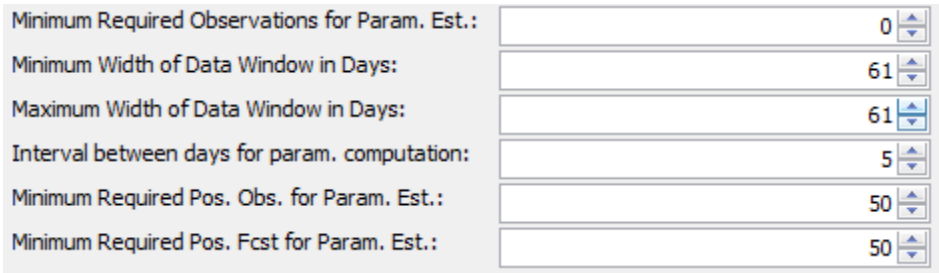


Figure 11: General options of the MEFP parameter estimation algorithm for estimating the parameters of the implicit precipitation treatment algorithm.

3.11.2.2 IPT Options

The options of the MEFP parameter estimation algorithm used to estimate parameters for the IPT algorithm (see Section 2.4.4) are shown in Figure 12. Both options are considered advanced, meaning that it is recommended users change the options only after gaining significant experience with the software.

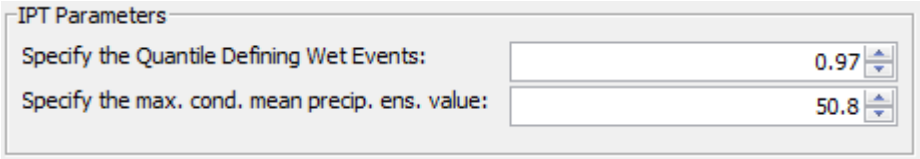


Figure 12: Options of the MEFP parameter estimation algorithm for estimating the parameters of the explicit precipitation treatment (EPT) algorithm.

3.11.2.3 EPT Options

The options of the MEFP parameter estimation algorithm used to estimate parameters of the EPT algorithm (see Section 2.4.4) are shown in Figure 13. Parameters are estimated only if the EPT model is allowed for that forecast source, meaning for RFC, GFS, and GEFS. Both options are considered advanced.

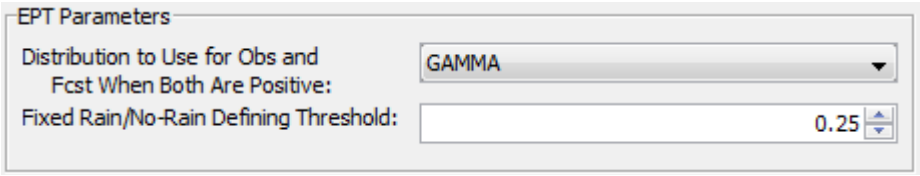


Figure 13: Options of the MEFP parameter estimation algorithm for estimating the parameters of the explicit precipitation treatment (EPT) algorithm.

3.11.2.4 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:

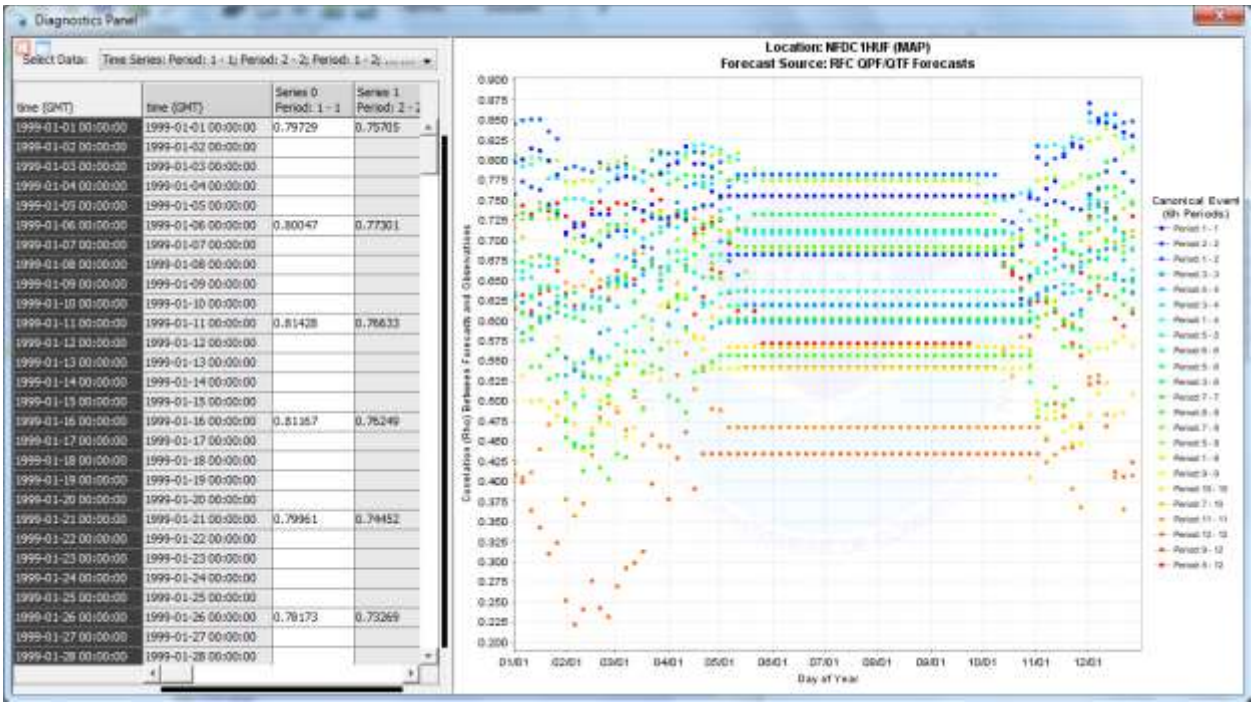
- Number of <*source*> Days: The number of forecast lead days for which to estimate parameters. This number must not exceed the number of forecast lead days available:
    - RFC QPF/QTF: usually 5 days for QPF, 7 days for QTF
    - GFS: 14 days
    - GEFS: 15 days
    - CFSv2: 270 days
- Initial Year of Parameter Estimation: The first year of data to include in parameter estimation. The data available for the forecast sources is as follows:
    - RFC QPF/QTF: Varies by RFC
    - GFS: 1979 – 2006
    - GEFS: 1985 – 2012 (ends July 30, 2012)
    - CFSv2: 1982 - 2010
- Last Year of Parameter Estimation: The last year of data to include in parameter estimation.

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

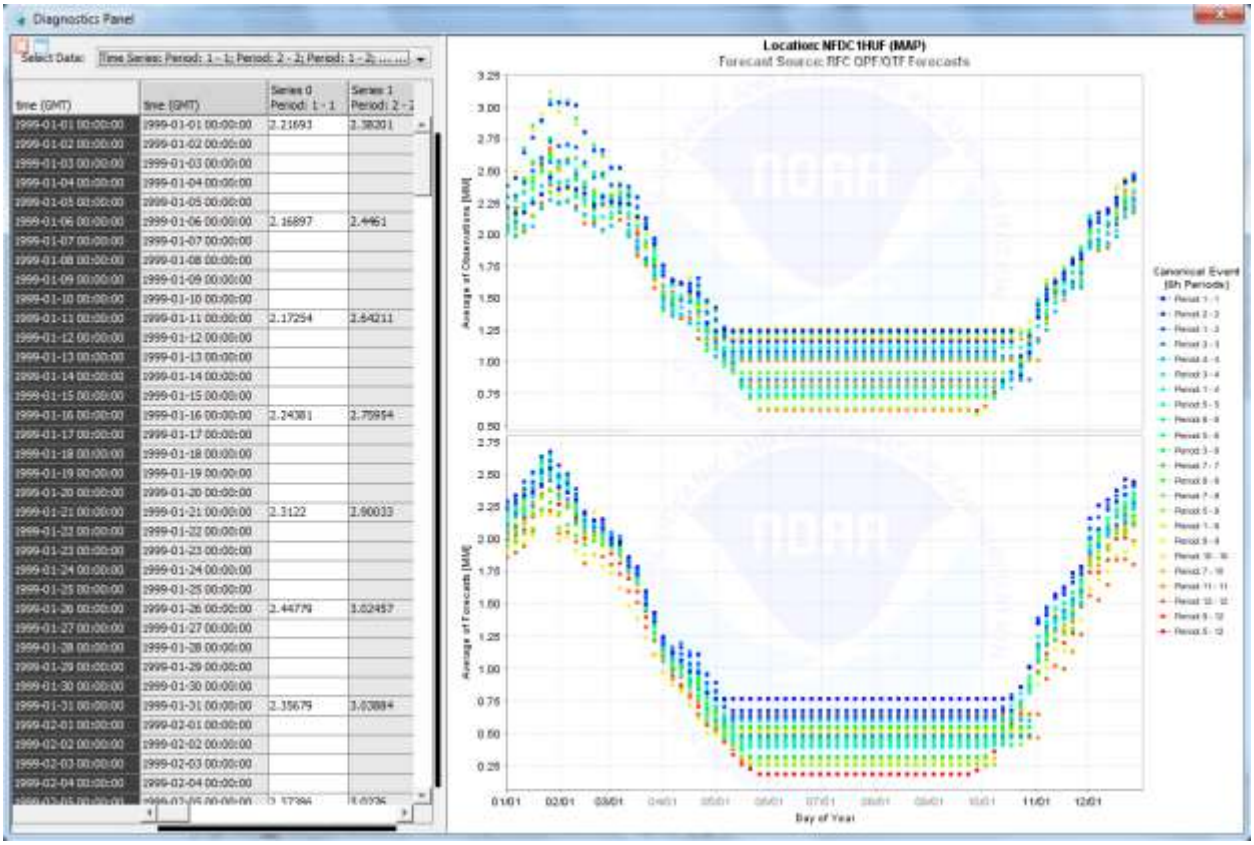
3.11.3 Standard Diagnostics Panel

The panel opened when the **View Button** is clicked for the estimated parameters is a standard Graphics Generator product viewing panel. It is similar to a **Time Series Chart Diagnostics Panel** (see Section 3.14), except that there is no **Toggle Visibility of Data Table Button** and no **Navigation Panel** that can be displayed. To display the table, check the **Show Table Menu Item** in the right-click popup menu. The diagnostic displays parameters selected within the **Parameter Summary Information Table** as time series against the day of the year (1 – 365 at 5 day time steps, by default; see the option "Interval between days for param. comp." in Figure 11). One series is included for each applicable canonical event. Because the legend can become large, it is recommended that the **Diagnostics 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):



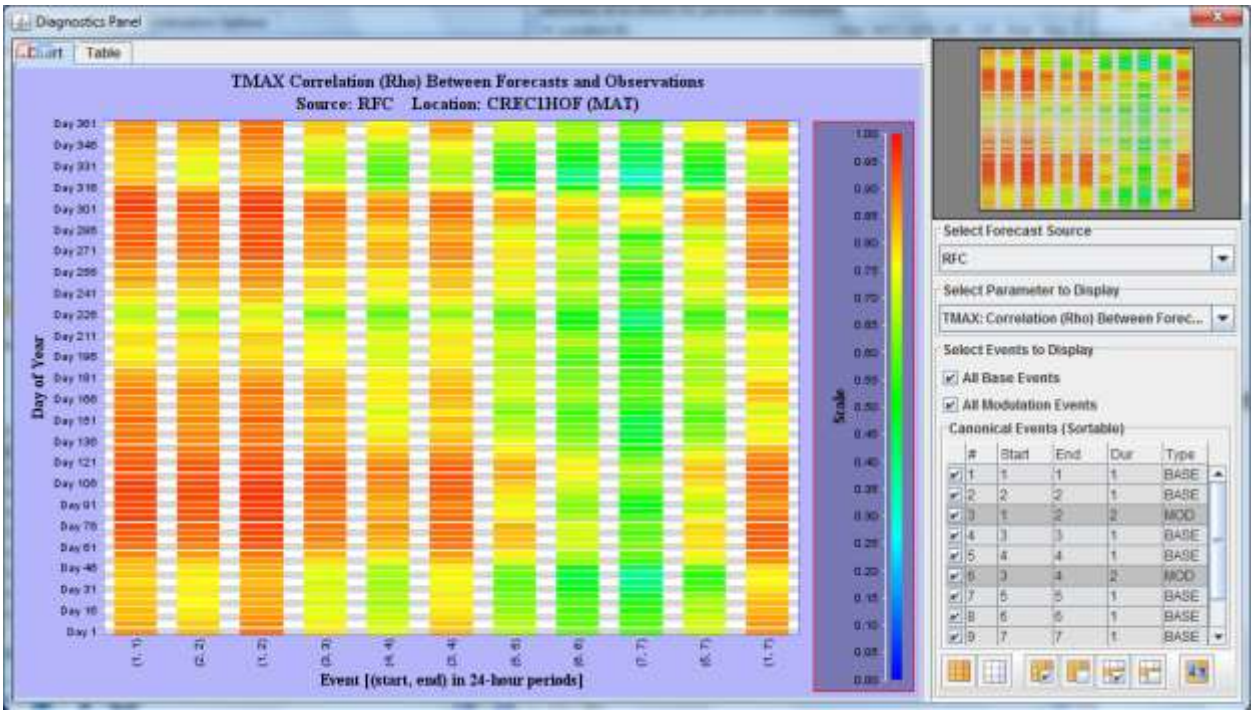
time (GMT)	time (GMT)	Series 0 Period: 1 - 1	Series 1 Period: 2 - 2
1999-01-01 00:00:00	1999-01-01 00:00:00	0.79729	0.75705
1999-01-02 00:00:00	1999-01-02 00:00:00		
1999-01-03 00:00:00	1999-01-03 00:00:00		
1999-01-04 00:00:00	1999-01-04 00:00:00		
1999-01-05 00:00:00	1999-01-05 00:00:00		
1999-01-06 00:00:00	1999-01-06 00:00:00	0.80047	0.77301
1999-01-07 00:00:00	1999-01-07 00:00:00		
1999-01-08 00:00:00	1999-01-08 00:00:00		
1999-01-09 00:00:00	1999-01-09 00:00:00		
1999-01-10 00:00:00	1999-01-10 00:00:00		
1999-01-11 00:00:00	1999-01-11 00:00:00	0.81428	0.78633
1999-01-12 00:00:00	1999-01-12 00:00:00		
1999-01-13 00:00:00	1999-01-13 00:00:00		
1999-01-14 00:00:00	1999-01-14 00:00:00		
1999-01-15 00:00:00	1999-01-15 00:00:00		
1999-01-16 00:00:00	1999-01-16 00:00:00	0.81167	0.78249
1999-01-17 00:00:00	1999-01-17 00:00:00		
1999-01-18 00:00:00	1999-01-18 00:00:00		
1999-01-19 00:00:00	1999-01-19 00:00:00		
1999-01-20 00:00:00	1999-01-20 00:00:00		
1999-01-21 00:00:00	1999-01-21 00:00:00	0.79961	0.74452
1999-01-22 00:00:00	1999-01-22 00:00:00		
1999-01-23 00:00:00	1999-01-23 00:00:00		
1999-01-24 00:00:00	1999-01-24 00:00:00		
1999-01-25 00:00:00	1999-01-25 00:00:00		
1999-01-26 00:00:00	1999-01-26 00:00:00	0.78173	0.73269
1999-01-27 00:00:00	1999-01-27 00:00:00		
1999-01-28 00:00:00	1999-01-28 00:00:00		



time (GMT)	time (GMT)	Series 0 Period: 1 - 1	Series 1 Period: 2 - 2
1999-01-01 00:00:00	1999-01-01 00:00:00	2.21693	2.38201
1999-01-02 00:00:00	1999-01-02 00:00:00		
1999-01-03 00:00:00	1999-01-03 00:00:00		
1999-01-04 00:00:00	1999-01-04 00:00:00		
1999-01-05 00:00:00	1999-01-05 00:00:00		
1999-01-06 00:00:00	1999-01-06 00:00:00	2.16897	2.4461
1999-01-07 00:00:00	1999-01-07 00:00:00		
1999-01-08 00:00:00	1999-01-08 00:00:00		
1999-01-09 00:00:00	1999-01-09 00:00:00		
1999-01-10 00:00:00	1999-01-10 00:00:00		
1999-01-11 00:00:00	1999-01-11 00:00:00	2.17254	2.64211
1999-01-12 00:00:00	1999-01-12 00:00:00		
1999-01-13 00:00:00	1999-01-13 00:00:00		
1999-01-14 00:00:00	1999-01-14 00:00:00		
1999-01-15 00:00:00	1999-01-15 00:00:00		
1999-01-16 00:00:00	1999-01-16 00:00:00	2.24381	2.75954
1999-01-17 00:00:00	1999-01-17 00:00:00		
1999-01-18 00:00:00	1999-01-18 00:00:00		
1999-01-19 00:00:00	1999-01-19 00:00:00		
1999-01-20 00:00:00	1999-01-20 00:00:00		
1999-01-21 00:00:00	1999-01-21 00:00:00	2.3122	2.90033
1999-01-22 00:00:00	1999-01-22 00:00:00		
1999-01-23 00:00:00	1999-01-23 00:00:00		
1999-01-24 00:00:00	1999-01-24 00:00:00		
1999-01-25 00:00:00	1999-01-25 00:00:00		
1999-01-26 00:00:00	1999-01-26 00:00:00	2.44779	3.02457
1999-01-27 00:00:00	1999-01-27 00:00:00		
1999-01-28 00:00:00	1999-01-28 00:00:00		
1999-01-29 00:00:00	1999-01-29 00:00:00		
1999-01-30 00:00:00	1999-01-30 00:00:00		
1999-01-31 00:00:00	1999-01-31 00:00:00	2.35679	3.03884
1999-02-01 00:00:00	1999-02-01 00:00:00		
1999-02-02 00:00:00	1999-02-02 00:00:00		
1999-02-03 00:00:00	1999-02-03 00:00:00		
1999-02-04 00:00:00	1999-02-04 00:00:00		
1999-02-05 00:00:00	1999-02-05 00:00:00	2.17284	3.0235

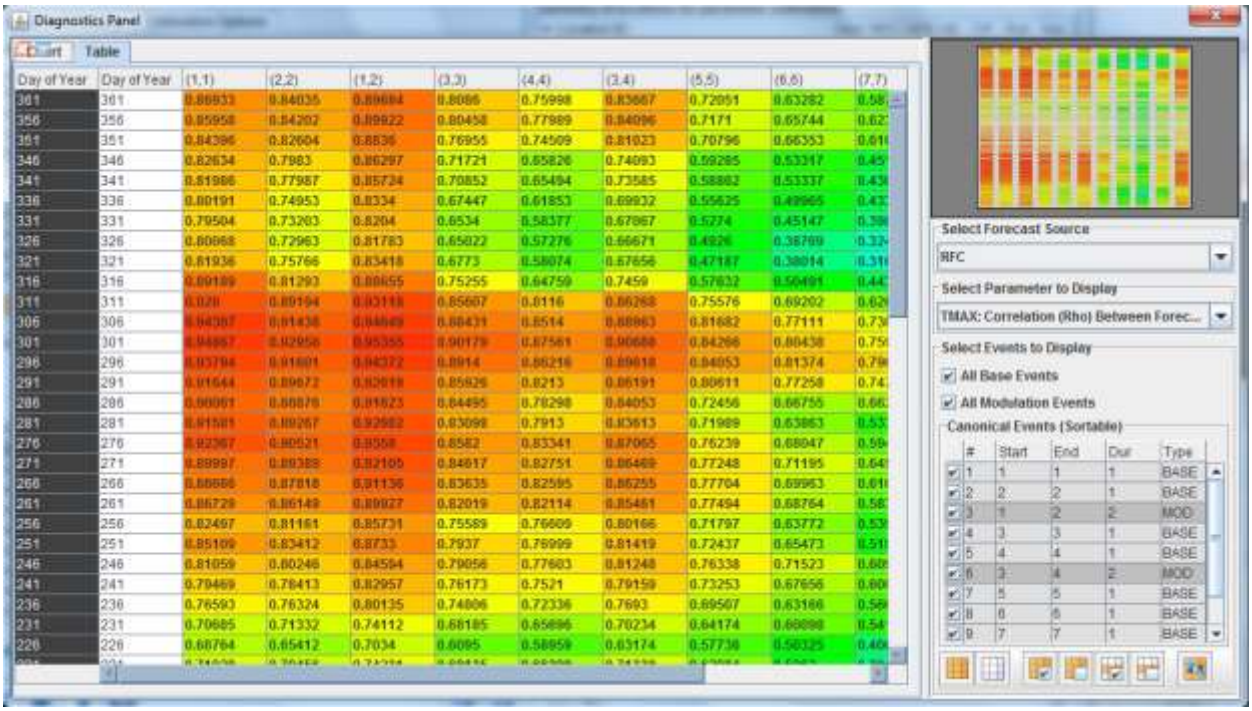
3.11.4 Parameter Diagnostics Panel

The **Parameter Diagnostics Panel** is opened by clicking on the **View Parameter Diagnostics Button**. It is a panel designed for quality control and analysis of the parameters estimated by the MEFPPE. By default, the parameters are displayed in a **Chart Panel** that displays a block plot; for example:



The domain axis shows the canonical event, the range axis shows the day-of-the-year, and the color displayed in a rectangle corresponds to the parameter value based on the **Scale** shown to the right of the plot. A **Table Panel** is also provided that can be accessed by clicking on the **Table Tab**,

For example:



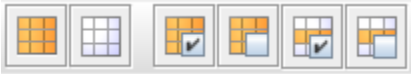
The first column is a row header column that equals the column of the table labed 'Day of Year'. It is always visible regardless of the scrolled position of the table view. Each cell displays the value of the parameter with a background color that equals the background color displayed for that value in the **Chart**.

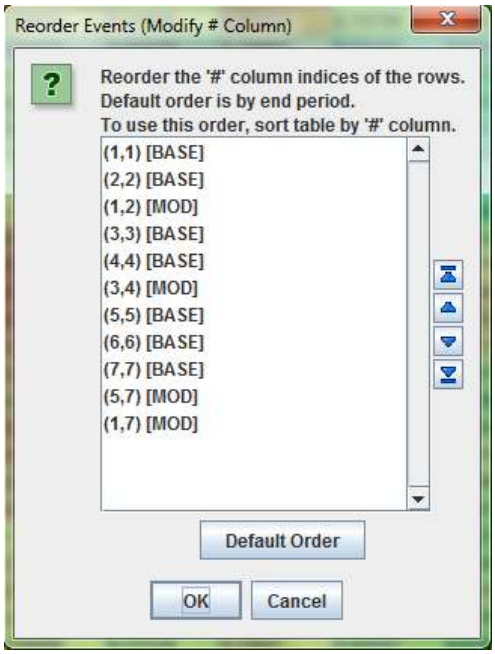
3.11.4.1 Components

The following components are displayed to the right of the **Chart** and **Table Panels**:

- **Navigation Panel**: Displays a miniaturized version of the block chart. When the block chart is zoomed in (using standard CHPS time series plots mechanisms; right click and use the popup menu to enable zooming and draw a rectangle), the zoomed axis limits will be drawn as a rectangle in the **Navigation Panel**.
- **Select Forecast Source Choice Box**: Allows for selecting which forecast source for which to view parameters.
- **Select Parameter to Display Choice Box**: Allows for selecting which parameters to view, described below. The correlation parameter (IPT correlation for precipitation) is selected initially.
- **All Base Events Checkbox**: When checked, the rows for all base events in the **Canonical Events Table** (below) will be checked.
- **All Modulation Events Checkbox**: When checked, the rows for all modulation events in the **Canonical Events Table** (below) will be checked.
- **Canonical Events Table**: Controls which events are included in the **Chart** and **Table Panels** and their display order. The leading checkbox indicates if the canonical event is displayed, while the remaining columns identify the event and if it is a base event or

modulation event. The columns of the table are sortable by clicking on the column header. If the row order is changed, the **Chart** and **Table Panels** will update immediately to reflect the new ordering.

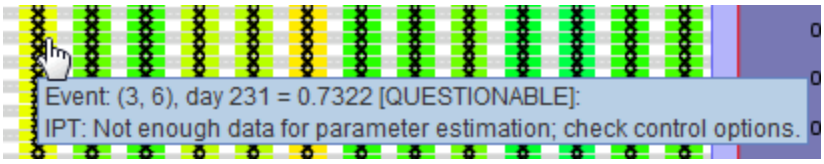
-  **Button Toolbar**: Provides various buttons to do the following (in order from left to right): select all rows, deselect all rows, check all rows, uncheck all rows, check selected rows, and uncheck selected rows.
- **Reorder Button**: Opens a **Reorder Event Dialog** to allow for customizing the ordering of the rows:



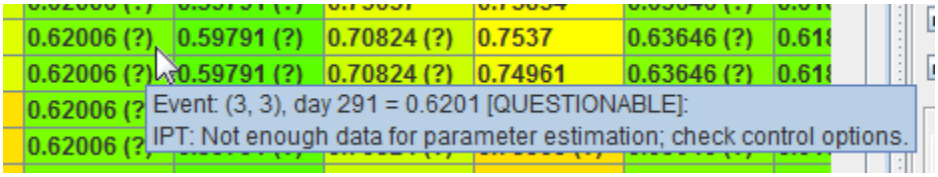
To change the ordering, select list items to move and use the up, , and down, , buttons on the left to move the items. When opened, the events will follow a default order: by canonical event end period (ascending), first, and then by start period (descending). If the order is modified, click on the **Default Order** button to reset the event order to the default order. Click on **OK** to accept the changes or **Cancel** to discard the changes. If **OK** is clicked, the **Canonical Event Table** will update to reflect the new order, and the **Chart** and **Table Panels** will update accordingly.

3.11.4.2 Features

- *Questionable Parameters*: Questionable parameter values, due to insufficient data or negative correlation coefficients, are displayed with a black **X** in the **Chart Panel** block or "(?)" appended to the **Table Panel** cell. Letting the mouse hover over the block or cell will result in a tool tip that explains why the parameters are questionable. An example for the **Chart Panel**:

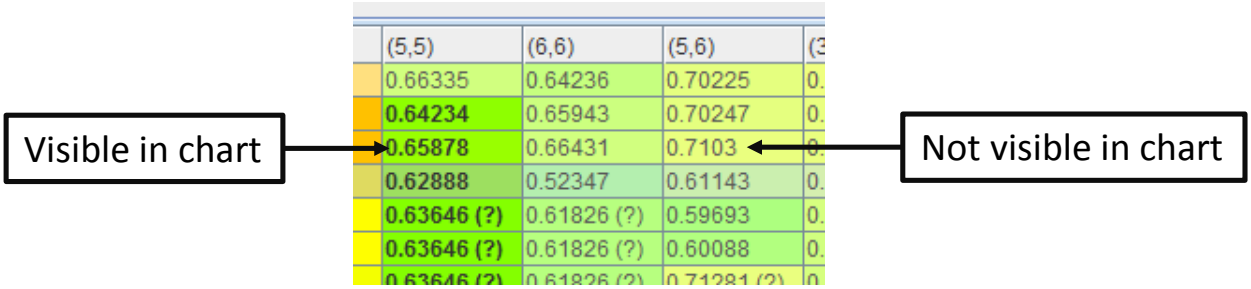


At example for the **Table Panel**:

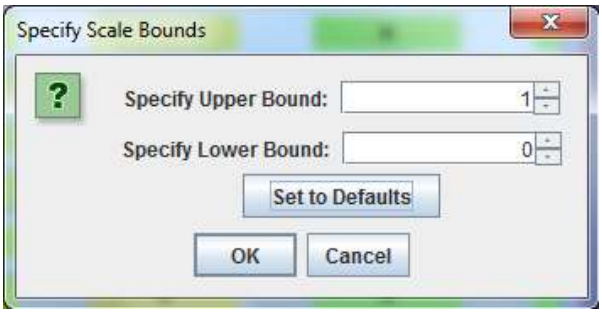


- *Zooming/Panning*: The **Chart Panel** allows for zooming through the same mechanism as a standard CHPS time series plot or Graphics Generator chart viewer (i.e., by right clicking to open a popup menu, check the allow zoom options, and drawing a rectangle). Panning is also allowed by dragging the chart using the right-mouse button. For the **Table Panel**, the viewed area can be moved by the panning the table; just press the left mouse button over the table and drag the table.

- *Table Updates With Axis Limits*: Whenever the axis limits of the chart are changed via zooming or by dragging the rectangle in the **Navigation Panel**, the viewed portion of the **Table Panel** will move to show the upper left most cell visible within the new axis limits. Furthermore, the background and foreground of cells not visible within the axis limits will be faded; for example:



- *Color Scale Editing*: The bounds of the color scale used to draw the blocks in a **Chart Panel** and determine the background color in the **Table Panel** can be changed by clicking on the "Scale" legend displayed in the **Chart Panel**. A **Specify Scale Bounds Dialog** will be displayed:

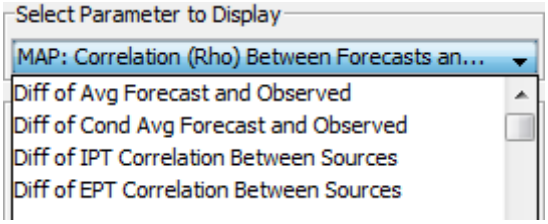


Modify the bounds by changing the **Specify Upper Bound Field** and **Specify Lower Bound Field**. The **Set to Defaults Button** can be clicked to revert to default, automatically computed values. Click on **OK** to accept the changes and **Cancel** to cancel the changes. If **OK** is clicked, both the **Chart** and **Table Panels** will change to use the new bounds on the paint scale.

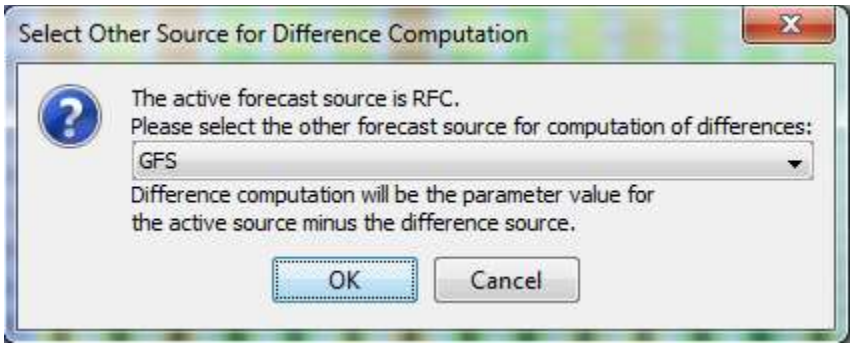
NOTE: The colors used cannot be modified. The color scheme will always go from blue to green to red with appropriate colors in between.

- *Available Parameters*: The parameters that can be chosen in the **Select Parameter to Display Choice Box** include all the standard estimated parameters for the loaded parameters data type, as well as special additional parameters that depend on data type and forecast source.

  For precipitation, the additional parameters are as shown:



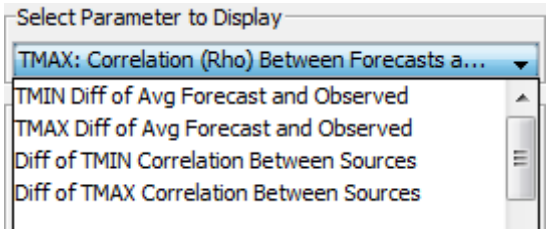
  - Diff of Avg Forecast and Observed: The difference between the parameters “Average of Forecasts” and “Average of Observations”.
  - Diff of Cond Avg Forecast and Observed: The difference between the parameters “Average of Forecasts Above Zero Threshold” and “Average of Observations Above Zero Threshold”.
  - Diff of IPT Correlation Between Sources: The difference between the parameter “Correlation (Rho) Between Forecasts and Observations” between two different sources. When selected, a **Select Other Source Dialog** will be opened:



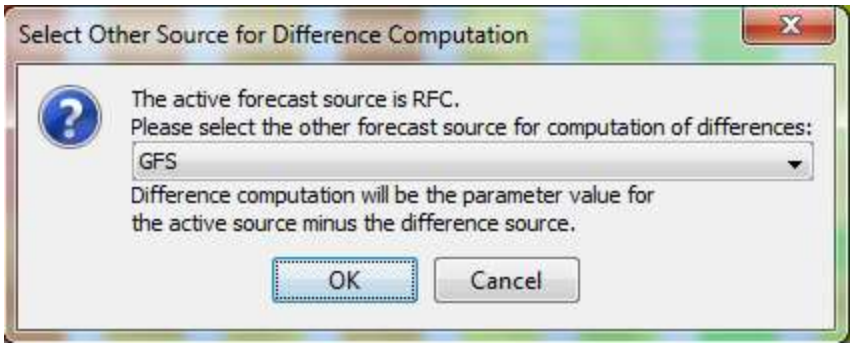
Select the source for which to compute the difference values and click **OK**. Click **Cancel** to cancel the parameter change.

- Diff of EPT Correlation Between Sources: The difference between the parameter "Correlation When Obs and Fcst are Positive (EPT)" between two different sources. When selected, a **Select Other Source Dialog** will be opened, but only sources for which EPT parameters can be computed will be listed. This parameter is only available if EPT parameters were estimated for the forecast source.

For temperature, the additional parameters are as shown:



- TMIN/TMAX Diff of Avg Forecast and Observed: The difference between the parameters "Average of Forecasts" and "Average of Observations" for the TMIN/TMAX datatype.
- Diff of TMIN/TMAX Correlation Between Sources: The difference between the parameter "Correlation (Rho) Between Forecasts and Observations" between two different sources. When selected, a **Select Other Source Dialog** will be opened:

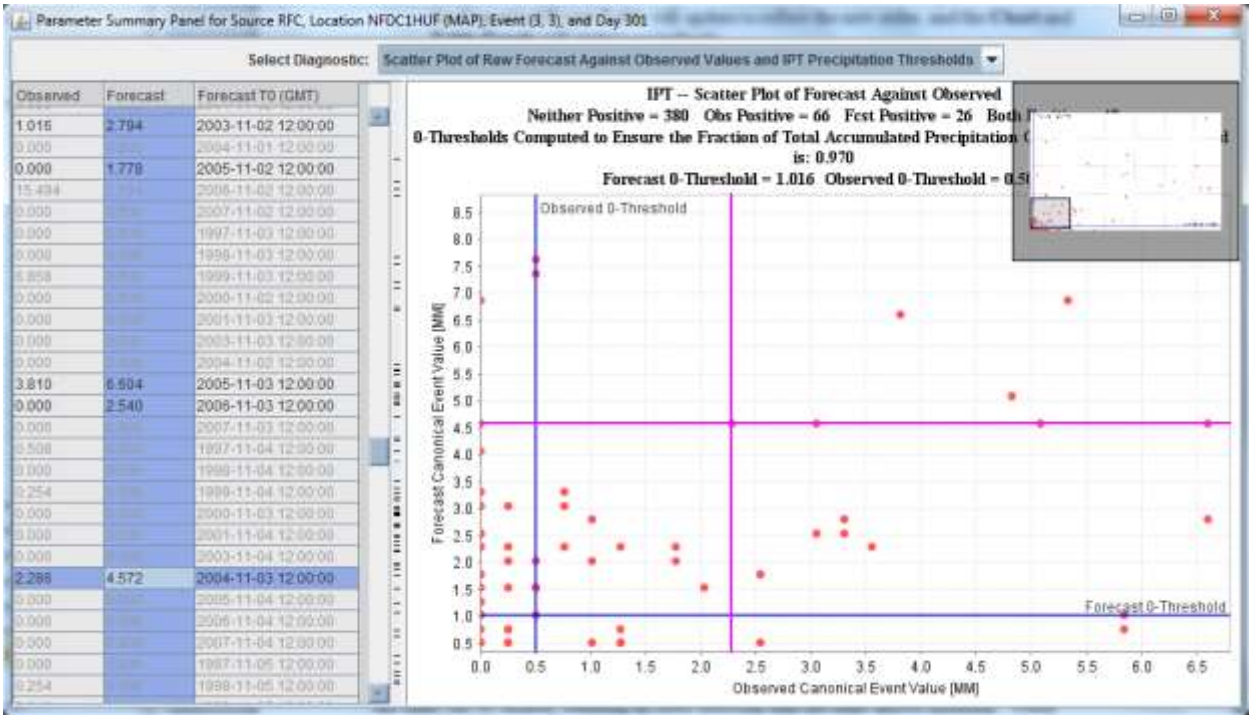


Select the source for which to compute the difference values and click **OK**. Click **Cancel** to cancel the parameter change.

- *Viewing Forecast-Observed Event Values*: Clicking on a block in the **Chart Panel** or selecting a cell in the **Table Panel** will open a **Parameter Summary Panel**, described below, which displays the forecast and observed canonical event values used to estimate the parameters for the corresponding canonical event and day-of-year.

3.11.4.3 Parameter Summary Panel

If a cell within the **Table Panel** or a block within the **Chart Panel** is clicked, a **Parameter Summary Panel** is displayed in a dialog for the currently selected parameter. For example (the chart has been zoomed in on for illustration purposes):



The panel displays a scatter plot in a standard Graphics Generator viewing panel. That means that a **Data Table** can be displayed via the right click popup menu; cells in the table can be selected and points on the chart can be clicked to yield crosshairs; and the chart allows zooming. When zoomed, a **Mark Panel** to the right of the table scroll bar indicates which rows are within the current axis limits. A non-standard feature is that, when zoomed, a **Navigation Panel** is displayed in the upper right corner that includes the same features as those described above for the **Parameter Diagnostics Panel** (i.e., a rectangle displays the current zoomed region and the rectangle can be dragged to change the axis limits).

3.11.4.4 Using the Parameter Diagnostics Panel

This section will be written in a future release.

3.11.5 Usage

3.11.5.1 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**.

3.11.5.2 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. For the **Standard Diagnostics Panel**, select the estimated parameters (rows) that you wish to view from the **Parameter Summary Information Table** and click on the **View Button**.
5. For the **Parameter Diagnostics Panel**, click on the **View Parameter Diagnostics Button.**

3.11.5.3 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.

3.12 Acceptance Subpanel

The **Acceptance Subpanel** of the **Estimation Steps Panel**, shown in Figure 15, is used to perform Step 8 of the parameter estimation procedure in Section 3.2.3: copy the parameter files from the MEFPPE run area to permanent storage. All estimated parameters for a single location and data type (precipitation or temperature) are stored in a single gzipped tarball (.tgz) file. The location of the permanent storage of those file is the subdirectory mefpParameters within the directory specified by the global property MEFP_ROOT_DIR, which is setup to point to *<mefp_root_dir>* during installation of the data ingest components; see *MEFP Configuration Guide: Data Ingest Components*. All parameter .tgz files are stored in the same directory.

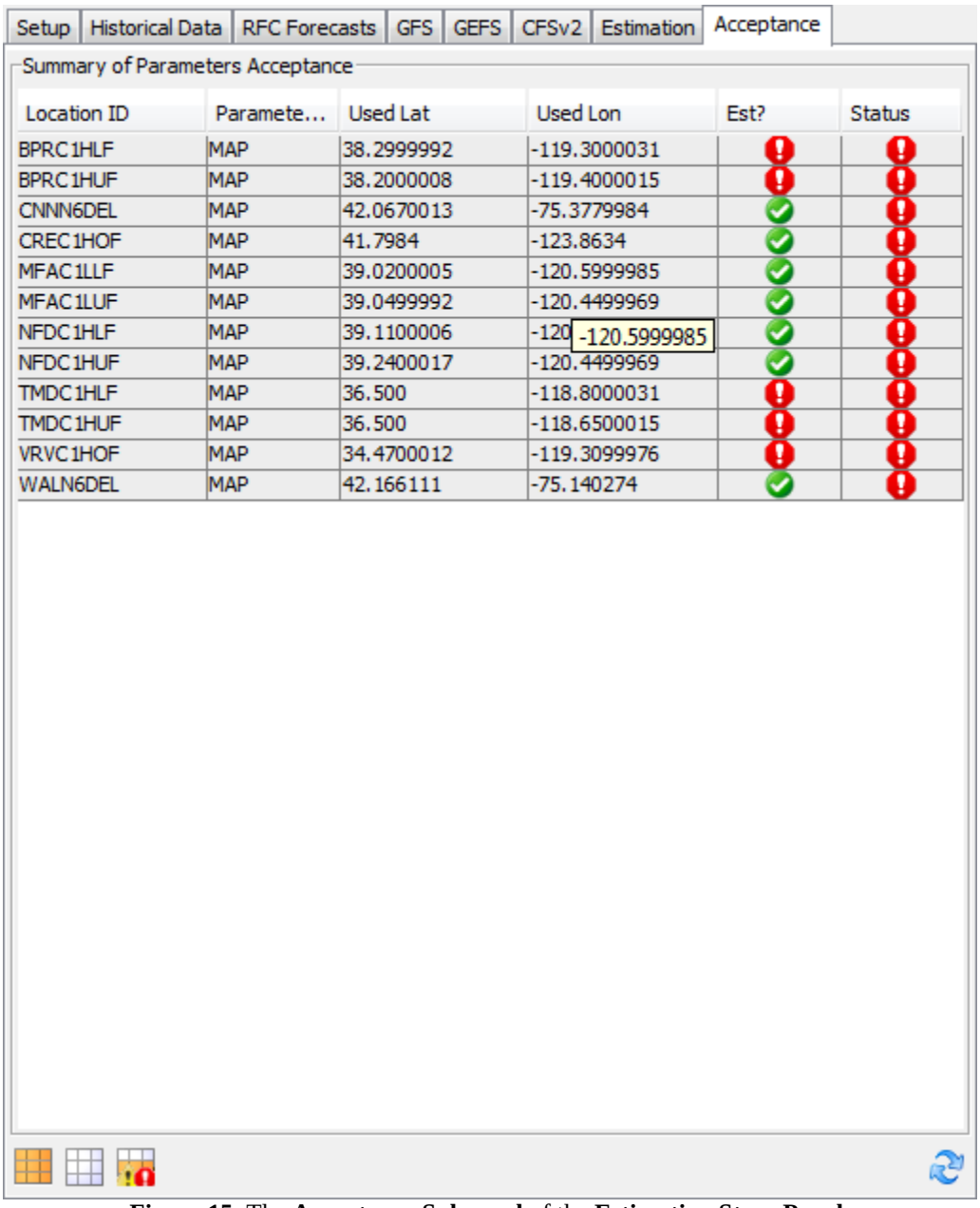


Figure 15: The **Acceptance Subpanel** of the **Estimation Steps Panel**.

3.12.1 Components

- **Summary of Parameters Acceptance 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.

3.12.2 Diagnostics

There are no diagnostics available for panel.

3.12.3 Usage

3.12.3.1 Accepting Parameters

To accept parameters, copying the parameter files to their permanent storage location where they can be accessed during MEFP adapter execution, select one or more locations from the **Summary of Parameters Acceptance 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**.

3.13 Location Summary Panel

The **Location Summary Panel**, shown in Figure 16, 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.

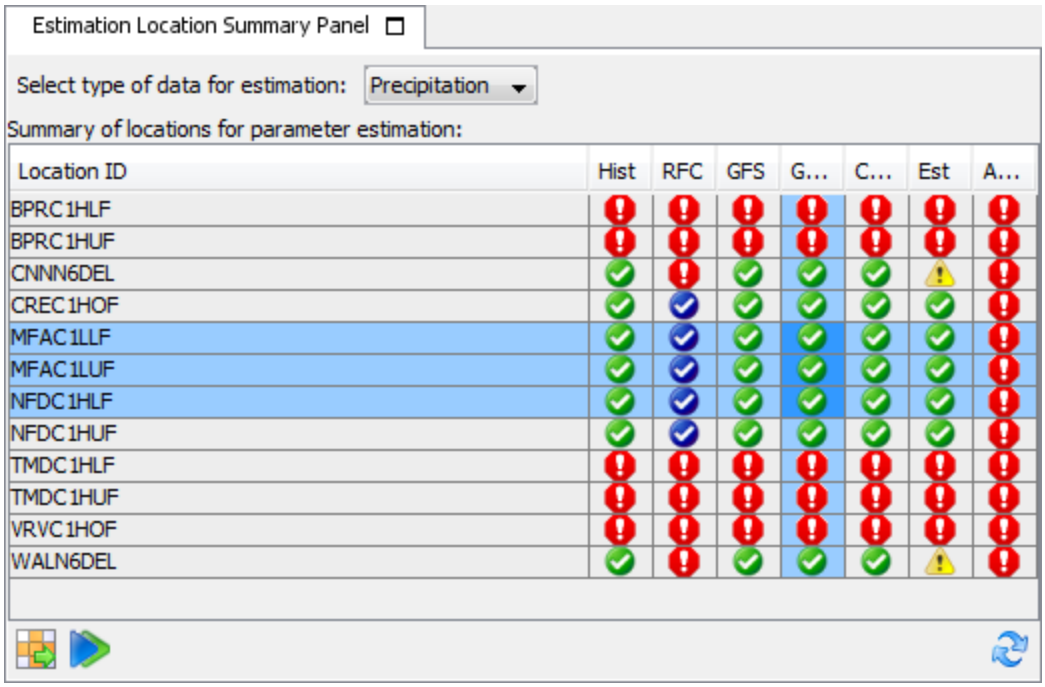


Figure 16: The **Location Summary Panel**.

3.13.1.1 Components

- **Select Type of Data for Estimation Choice Box**: Allows for selecting the active estimation data type, either "Precipitation" or "Temperature". When a change is made to the choice, nearly all of the tables in the MEFPPE, including the **Summary of Locations for Parameter Estimation Table**, will update to reflect the change of data type.
- **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 6: acquire CFSv2 reforecast files.
    - 'Est': Displays the status of Step 7: estimate parameters.
    - 'Accept': Displays the status of Step 8: accept (zip) parameter files.

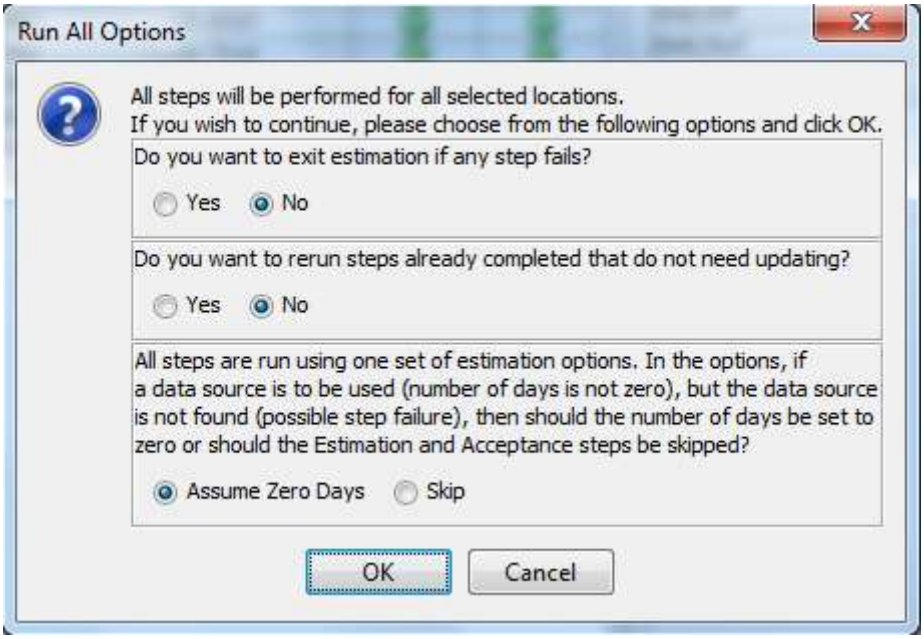
- **Goto Step Panel Button**: Click to make the step subpanel corresponding to the selected column of the **Summary of Locations for Parameter Estimation Table** active within the **Estimation Steps Panel**. Also, if appropriate, it selects rows of the **Generic Summary Table** for that subpanel for 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 3.13.2.1 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**.

3.13.2 Usage

3.13.2.1 Running All Steps for Multiple MEFP Locations

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

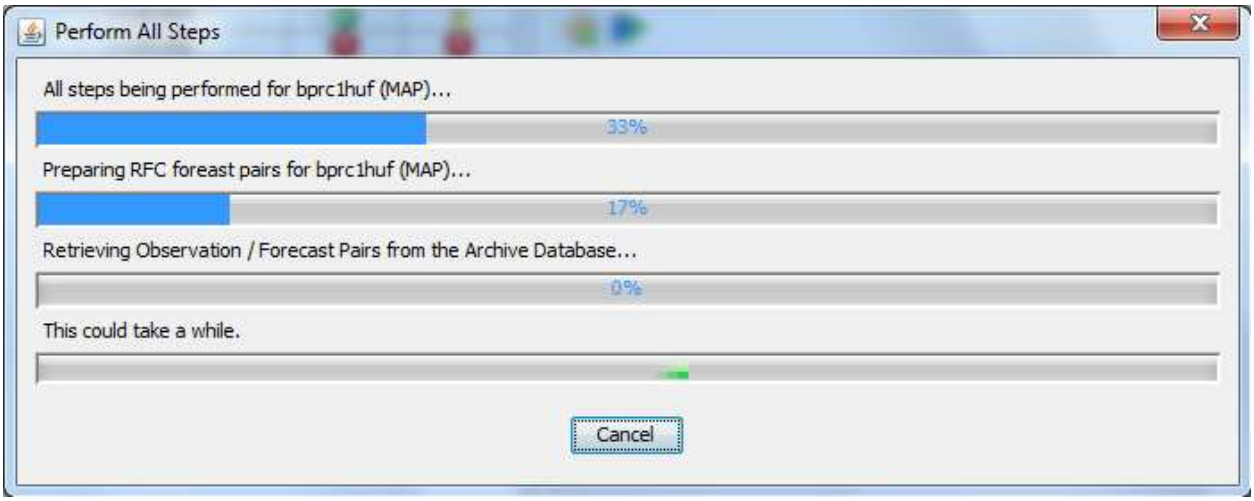
1. Select the rows for the desired MEFP locations from the **Summary of Locations for Parameter Estimation Table**.
2. Click on the **Run All Steps Button**. A **Run All Options Dialog** will open:



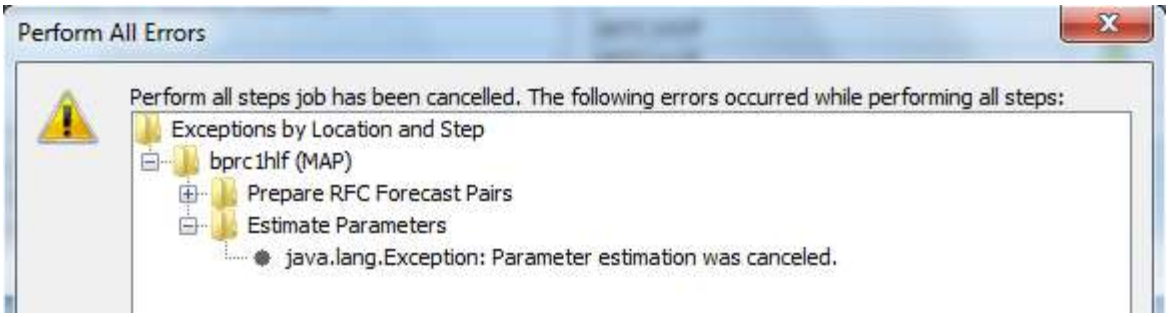
There are three options that the user can set:

- 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 based on the estimation option (i.e., **Use Checkbox** for the source in the **Estimation Options Panel** is checked and the number of forecast days for the forecast source is more than 0). The options are to set the number of days to be 0 so that the parameters are not estimated for that forecast source but still estimated for the location for other sources, 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:



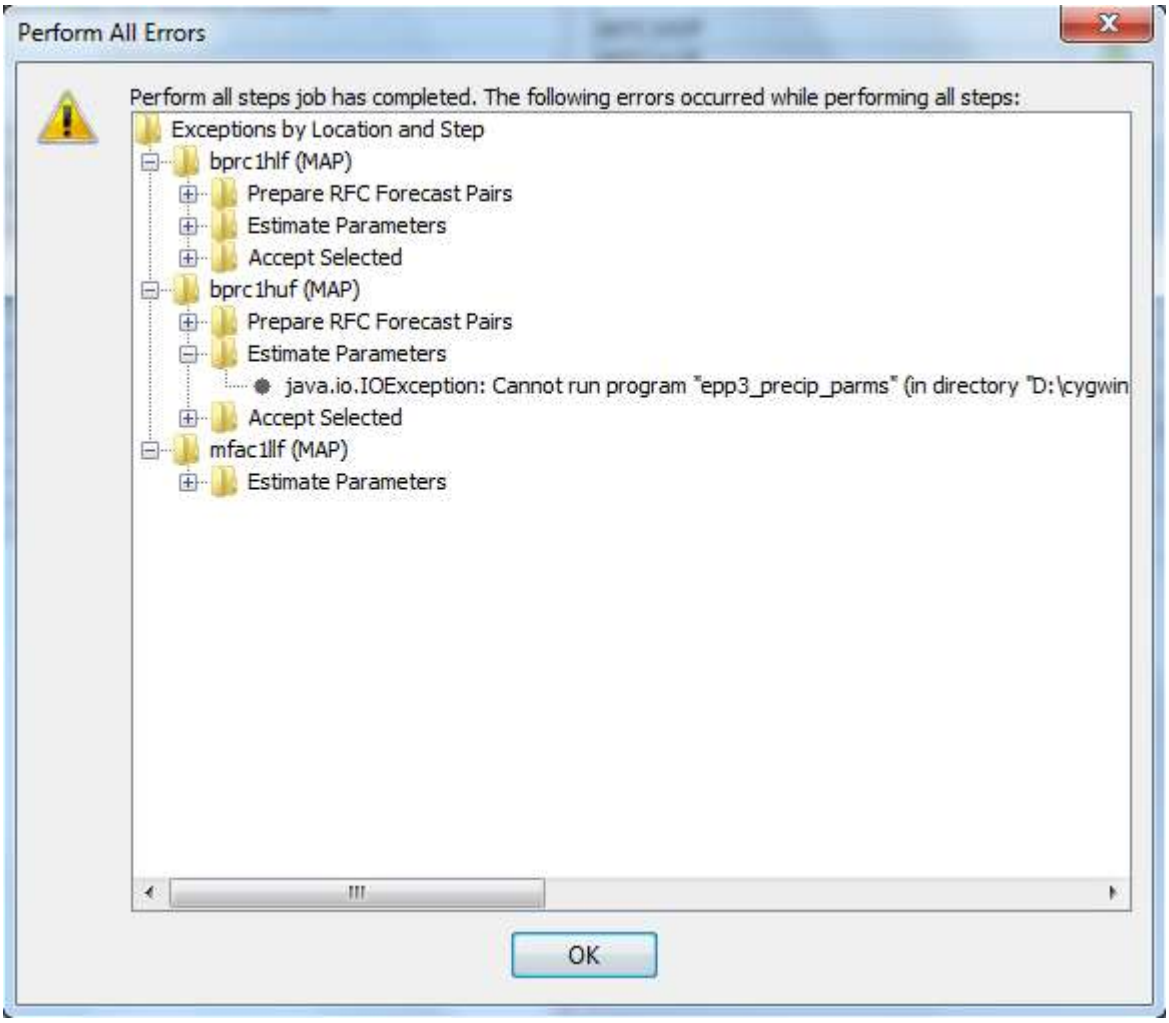
The run all can be canceled at any time by clicking on the **Cancel Button**. The run all will stop at the next opportunity and a **Perform All Errors Dialog** will be displayed where the last leaf will indicate that the user canceled the run:



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:



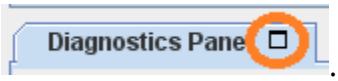
Expand the tree nodes in order to identify the errors that occurred. If no errors occurred, a message will be displayed indicating that no errors occurred.

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

3.14 Diagnostics Panel

The **Diagnostics Panel**, an example of which is shown in Figure 17, 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, but include two three types: time series **Diagnostics Panels** (described below), and the **Estimation Subpanel**'s **Standard Diagnostics Panel** (Section 3.11.3) and **Parameter Diagnostics Panel** (Section 3.11.4).

Some diagnostics can require a significant amount of time draw. Furthermore, it will need to be redrawn whenever the **Diagnostics 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. Furthermore, it is recommended that the panel be detached from the **MEFPPE Panel** and enlarged prior to viewing diagnostics so no time needs to be spent rendering the chart while resizing the panel's window. To detach the panel, click on the undock icon in the tab displayed for the **Diagnostics Panel**:



NOTE: The chart displayed in the **Diagnostics Panel** *will* <u>*not*</u> *change until another diagnostic is selected to be displayed or the* **Dispose Button** *is clicked.*

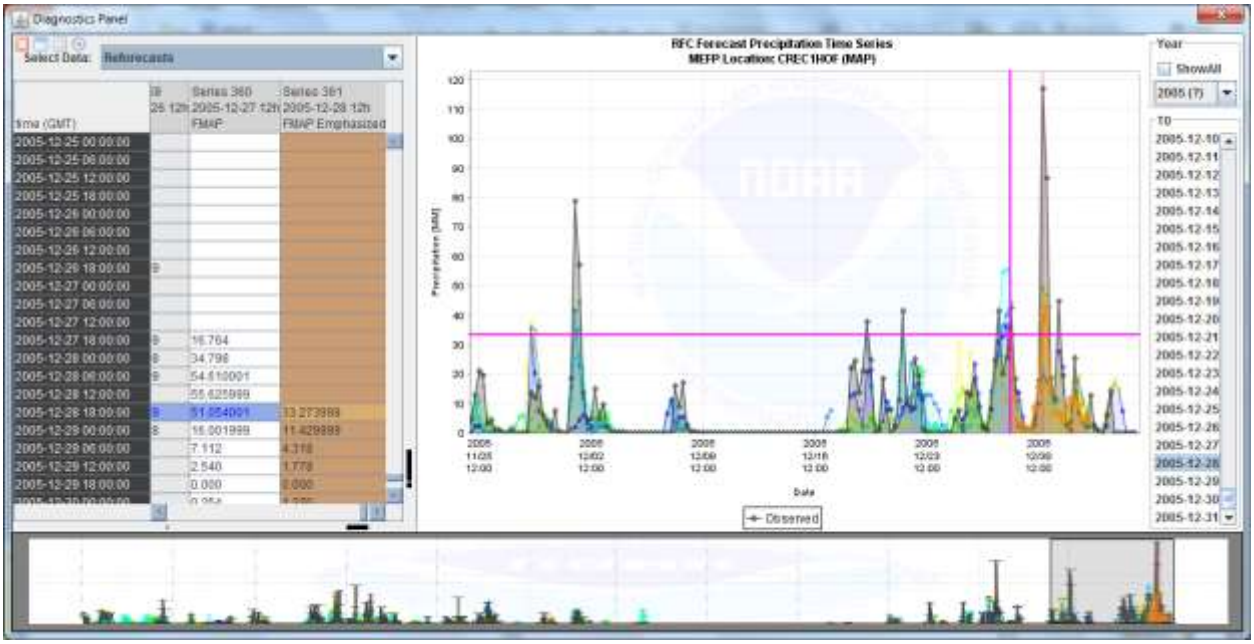


Figure 17: Example time series **Diagnostics Panel** for precipitation data. See Section 3.11.3 for an example of an estimated parameters **Diagnostics Panel**.

3.14.1 General Components

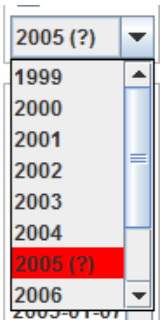
The only general components of a **Diagnostics Panel** are the buttons in the upper left corner, as shown in Figure 18. The buttons are only visible if a diagnostic is currently being displayed. The buttons are as follows:

- **Dispose Button**: Click on this button to dispose of the current diagnostics.
- **Move Diagnostic to Window Button**: Moves the diagnostic display to a separate window. This allows for multiple diagnostics to be displayed at the same time; after creating the first, click on this button to move the display to a window, allowing for a second to be drawn. This is not the same as undocking the **Diagnostics Panel**, which does not create a new panel, but, instead, detaches the existing panel.
- **Toggle Visibility of Data Table Button**: Shows or hides the **Data Table Panel**, described below. This is only visible in time series **Diagnostics Panels**.
- **Toggle Visibility of Navigation Panel Button**: Shows or hides the **Navigation Panel**, described below. This is only visible in time series **Diagnostics Panels**.

3.14.2 Time Series Diagnostics Panel Components

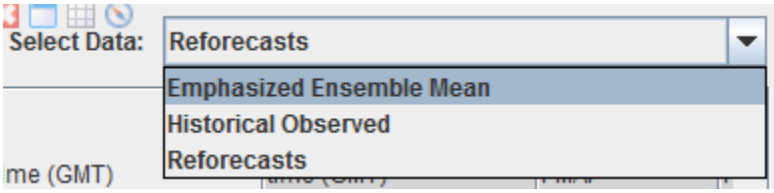
All additional components described below (shown in Figure 18) are general to **Diagnostics Panels** for displaying historical and reforecast/archived time series, which applies to the forecast source **Estimation Steps Tabbed Subpanels** described in Sections 3.6 – 3.10. The **Diagnostics Panels** for the **Estimation Subpanel** are described in Sections 3.11.3 and 3.11.4.

- **Diagnostics Chart**: A chart displaying observed and reforecast (if applicable) time series. The chart is always displayed in the **Diagnostics Panel** (unlike the **Data Table**) and includes basic CHPS time series plot controls (e.g.: popup menu and panning capabilities). If reforecasts or archived forecasts are available for a data source (i.e., for all forecast source subpanels but the **Historical Data Subpanel**), then data is displayed for only a single year of those forecasts at a time The following are components associated with the **Diagnostics Chart** when reforecasts are available:
  - **Show All Checkbox**: When checked, all of the time series, historical and reforecast, will be displayed in a single chart. It is <u>not</u> recommended that this checkbox be checked, as the chart rendering time can be very slow.
  - **Year Choice Box**: Allows for selecting the year of data to display. Years with questionable data will be followed by a "(?)" and displayed with a red background:

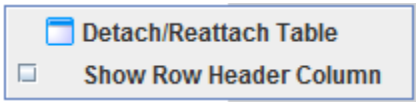


- **Forecast Times List**: Allows for selecting forecast times to emphasize in the **Diagnostic Chart** and **Data Table** (described below). Date listed are in GMT. Click on a list item to emphasize the corresponding time series and click again to de-emphasize. See below for the effects of emphasizing a time series.

- **Data Table**: A table displaying data shown in the chart within a scrollable panel. Its visibility is controlled by the **Toggle Visibility of Navigation Panel Button**. A popup menu that appears by right clicking on the table allows for detaching this table; see below. The scrolled view area of the **Data Table** is moved whenever the axis limits of the **Diagnostics Chart** changes; for example, when zooming or panning. The following are components associated with the **Data Table** when reforecasts are available:
  - **Column Header**: Shows the names of the column within the **Data Table**. The header is always visible regardless of the scrolled view in the table.
  - **Row Header**: Displays the first column of the **Data Table**, which shows the times for the data values. The visibility of the **Row Header** is controlled via a popup menu described below. When displayed, the **Row Header** is always visible regardless of the scrolled view of the table.
  - **Select Data Choice Box**: Displays a list of selectable chart series. For example:



    For the **Historical Data Subpanel**, only the "Historical Observed" item is available. The "Emphasized Ensemble Mean" item is only available when a date is selected within the **Forecast Times List**. See Section 3.10.2 for an additional option available when viewing CFSv2 monthly reforecasts.
  - **Mark Panel**: Two panels that display marks below the bottom scrollbar and to the right of the right-hand scrollbar. Marks are used to display the location of columns and rows for which data is within axis limits, the cell is selected, the corresponding time series is emphasized (see below), and the data is questionable (see below).
  - **Data Table Popup Menu**: The popup menu is the following (exact appearance depends on the operating system defaults):



    The **Detach/Reattach Menu Item** allows for detach/re-attaching the **Data Table** from the **Diagnostics Panel**. When attached, the table is displayed to the left of the chart with a resizing bar in the middle. When detached, the table is displayed within another, independently controlled frame. The **Show Row Header Column** controls if the **Row Header** is displayed.

- **Navigation Panel**: A panel that displays a miniaturized version of the **Diagnostics Chart** image with default axis limits and a navigation window (drawn as a black rectangle) that shows the portion of that chart that is visible within the current axis limits. This panel is useful when zooming in on a chart. The mouse can be used to drag the navigation window, changing the axis limits when the mouse button is released to display the region of the chart corresponding to the new position of rectangle. The window will also move when panning the chart.

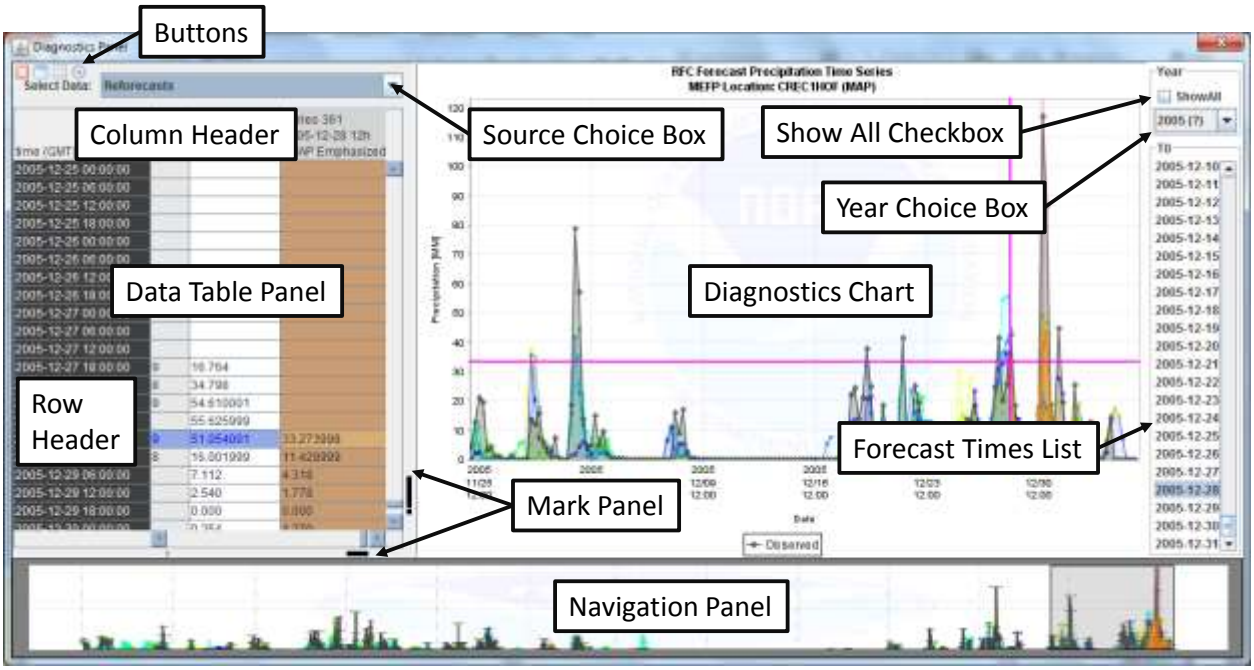


Figure 18: Components of the **Diagnostics Panel**.

3.14.3 Time Series Diagnostics Panel Features

Figure 19 and Figure 20 diagram features of the **Diagnostics Panel** when viewing historical and reforecast time series. The following features are provided:

- *Historical time series*: Historical time series are shown in black. The area under the precipitation values and between minimum and maximum values is shaded in partially transparent gray.

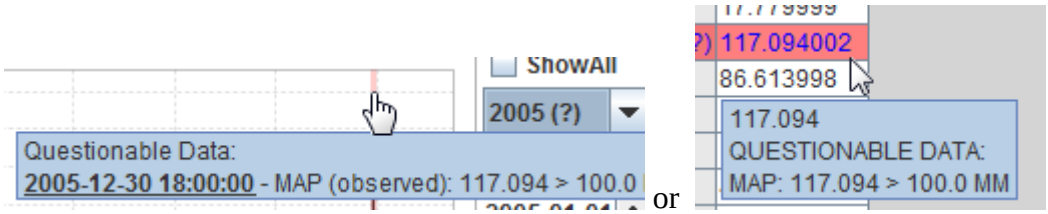
- *Reforecast/Archived forecast time series*: Reforecasts and archived forecasts (for RFC data) are displayed as blue, green, and yellow lines and shapes. They are displayed within the **Data Table** against a single date column covering all time series.

- *Selected data point*: A value in the chart can be selected by clicking on the point. When the mouse pointer is over a point that can be clicked, the cursor will be a pointing hand, . Upon selecting a point in the **Diagnostic Chart**, the corresponding chart series

within the **Select Data Choice Box** and cell in the **Data Table** will be selected. That causes crosshairs to be drawn on the chart (see Figure 19).

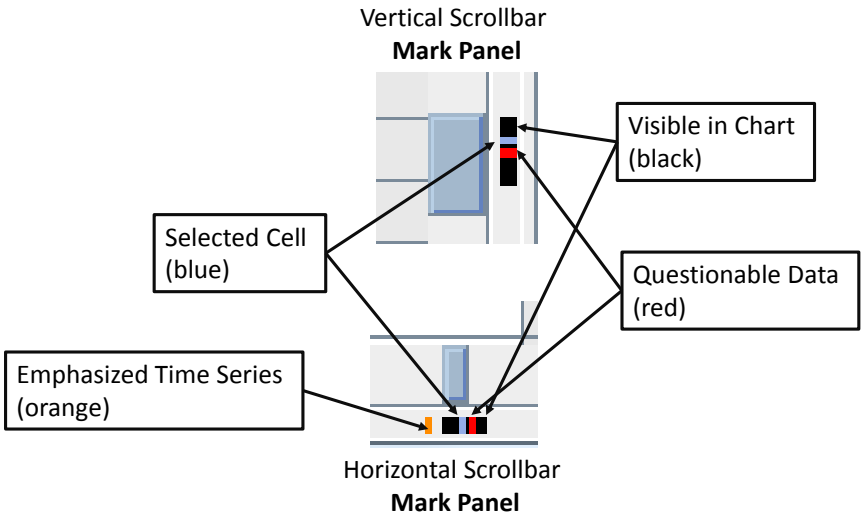
- *Emphasized reforecast time series*: Reforecast time series can be emphasized by selecting the appropriate item from the **Forecast Times List** or by <ctrl>-clicking (left mouse button) on the corresponding column in the **Data Table.** When emphasized, the time series is shown in orange in both the **Diagnostic Chart** and **Data Table** (cell background color). For temperature data, the region between the minimum and maximum time series is shaded in partially transparent orange.

- *Questionable data*: Questionable data values are only applicable to the **Historical Data Subpanel** and **RFC Forecast Subpanel**. They are marked by red marks or zones displayed within the **Diagnostic Chart** and cells with red backgrounds in the **Data Table**. By letting the mouse pointer hover over a questionable data mark/zone within the chart or cell within the table, a tool tip will display why the data is considered questionable. For example:



For temperature data, within the **Data Table**, the minimum and maximum values are treated as one cell for determining if the data is considered questionable: if either is considered questionable, both are considered questionable and will be highlighted accordingly.

- **Mark Panel** *marks*: Marks are provided in the two **Mark Panels** in order to allow users to more quickly navigate to cells. The types of marks are as follows (shown below):



- Rows and columns with at least one data value within the current axis limits of the chart are shown as a black mark.
- The position of the selected cell is shown as a blue mark.
- Rows and columns with at least one questionable data value are shown in red.
- The column corresponding the currently emphasized time series is shown in orange.

- *Zooming/panning*: Whenever the axis limits change due to zooming or panning of the chart, the (scrollable) viewed area within the **Data Table** will move to show the upper left most cell for data that is visible within the new axis limits. When panning, the displayed view change occurs only after the mouse button is released, ending the panning. For the **Data Table**, the viewed area can be moved by the panning the table; just press the left mouse button over the table and drag the table.

Additional features of time series **Diagnostics Panels** are described with each subpanel as needed.

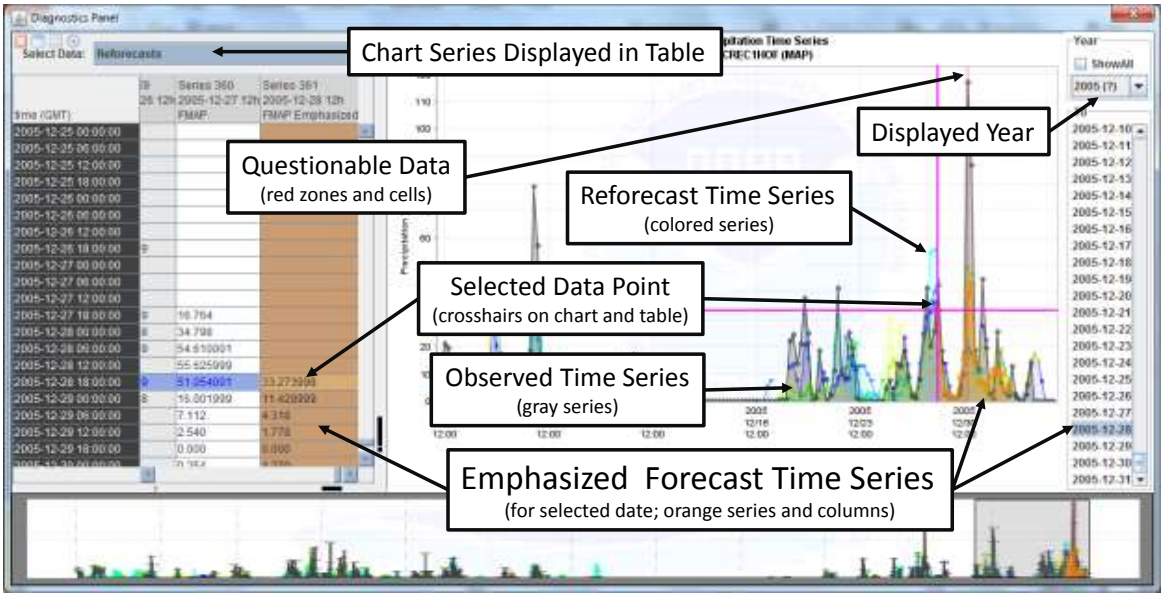


Figure 19: Features of the **Diagnostics Panel** for precipitation data.

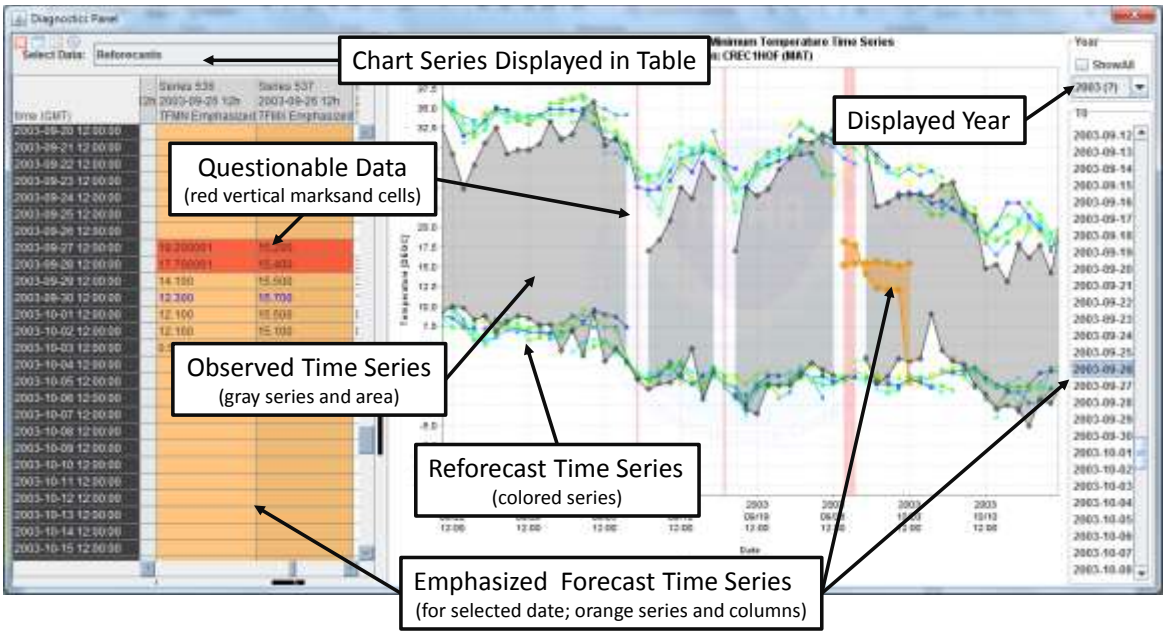


Figure 20: Features of the **Diagnostics Panel** for temperature data.

3.14.4 Using Time Series Diagnostics Panels

The **Diagnostics Panel** for displaying historical and reforecast/archived forecast time series is intended to allow the user to gain a general notion of forecast performance and identify data that may be problematic and lead to poorly estimated parameters. The general process should be as follows:

1. After gathering data for each forecast source, the use the **Generic Summary Panel** for that source's subpanel to identify parameter estimation locations for which questionable data was found by MEFPPE. Such locations are identified by a **Questionable Status Icon**, . See Section 3.2.4.6 for a description of the checks MEFPPE performs to determine if a time series value is considered questionable.

2. Select the row of a location with questionable data and open the **Diagnostics Panel** for that source by clicking on the view button, which typically uses the icon .

3. Identify the years for which questionable data was found by examining the **Year Choice Box**. Such years are displayed with a red background and a "(?)" in the label; for example 2005 (?). This does not apply to the **Historical Data Subpanel**'s **Diagnostics Panel**, since it does not include reforecast data.

4. Identify and locate data values considered questionable by MEFPPE. For each year with questionable data, select the year and use the features and components of the **Diagnostics Panel** (see the previous two sections) to locate the questionable data values and identify the reason. Questionable thresholds can be hovered over to view tool tips and marks

within the **Mark Panels** of the **Data Table** allow for quickly locating the corresponding data values within the table.

5. Determine if the questionable data value is acceptable or not. If yes, then move on to the next questionable data value. If no, then the data source must be modified and the associated MEFPPE forecast source step must be performed, again, using the new data. See Section 3.2.4.6 for instructions on modifying the problematic values and incorporating the changed value in MEFPPE.

6. Visually inspect the remaining data to determine if any other values may be questionable that were not detected by the MEFPPE automated algorithm. It is up to the user to identify such values. If any are found, repeat Step 5, above.

7. Once all of the questionable data has been checked and, if necessary, corrected, delete the questionable message file associated with that forecast source and location by selecting the location's row within the appropriate **Generic Summary Table** and clicking on the **Remove Questionable Data Files Button**, , in the source subpanel. This will make it so that the status icon used for that location and that forecast source will not be a .

With all questionable data corrected, parameters can now be estimated for that forecast source following the instructions provided in Section 3.11.

4 MEFP Operational Reference Manual

This manual describes the configuration of the MEFP workflows and modules. First, an overview of the configuration of the MEFP, including data-ingest components and forecast components, is provided. Then, a detailed reference guide for each of the three model adapters created for use with MEFP is provided:

- TimeSeriesExporterModelAdapter: A generic adapter for exporting multiple time series to multiple files, where the file names are defined by a pattern.
- CFSv2LaggedEnsembleModelAdapter: An MEFP-specific model adapter for building CFSv2 lagged ensembles for input to the MEFPEnsembleGeneratorModelAdapter. A lagged ensemble is a specific type of ensemble described in Section 4.3.1.
- MEFPEnsembleGeneratorModelAdapter: An MEFP-specific model adapter that executes the MEFP algorithm, generating forecast ensembles of precipitation and minimum/maximum temperature.

Configuration files associated with the above adapters are typically located below the directory …/Config/ModuleConfigFiles/hefs/<*forecast group*>, following instructions in the *MEFP Configuration Guides*.

Given the configuration delivered with the release of the MEFP, the MEFP configuration files associated with each of the three model adapters should not need to be significantly edited except to create new versions of the files for new forecast groups and to specify the time series that provides the RFC forecast source QPF and QTF time series for input to MEFP. Performing those tasks is described in the *MEFP Configuration Guides*. Nonetheless, a complete reference manual is provided for each adapter herein.

See the configuration guides for instructions on how to use these components operationally:

- Importing and viewing grids and exporting CFSv2 location-specific forecast time series is described in the confirmation section of the *MEFP Configuration Guide: Data Ingest Components*.
- Generating CFSv2 lagged ensembles, executing the MEFP, and viewing the generated ensembles is described in the confirmation section of the *MEFP Configuration Guide: Forecast Components*.

The next section describes the configuration of the MEFP for data ingest and operational forecasting, and each of the three adapters listed above are described in the sections that follow.

4.1 MEFP Configuration

A diagram presenting the workflows, modules, and time series data flow employed to execute the MEFP operationally is provided at the end of this section. The process consists of two phases: (1) ingesting gridded forecasts via scheduled workflows, and (2) executing MEFP to generate ensemble forecasts.

NOTE: Though available in parameter estimation and hindcasting, the GFS (1998 frozen) version can no longer be used in operational forecasting. The operational forecast grids are no longer available. All references to GFS have been removed below.

4.1.1 Data Ingest

Importing the gridded forecasts required for MEFP execution is configured as described in *MEFP Configuration Guide: Data Ingest*. Each of the two types of gridded forecasts, GEFS and CFSv2, are ingested independently via a scheduled workflow. The process for each is as follows:

- *GEFS*
  1. Import the GEFS gridded forecast ensemble time series covering the entire world.
  2. Interpolate the world-grid to a gridded ensemble mean time series covering CONUS.
- *CFSv2*
  1. Import the CFSv2 single-valued gridded forecast time series covering the entire world.
  2. Interpolate the world-grid to a gridded forecast times series covering CONUS.
  3. For each catchment for which MEFP is to execute, spatially interpolate a forecast time series from the gridded forecast time series using nearest neighbor (relative to the catchment centroid) algorithm.
  4. Export a PI-timeseries XML file containing the forecast time series interpolated for each catchment. This is necessary for running the CFSv2LaggedEnsembleModelAdapter.

The diagram at the end of this section displays the workflows and modules involved, as well as the flow of time series.

4.1.2 Operational Forecasting

Operational forecasting uses as input the output from the data ingest process

- *GEFS*: Gridded ensemble mean time series
- *CFSv2*: PI-timeseries XML files containing forecast time series for each location

In addition, it also uses as input RFC 6-hour single-valued QPF (FMAP) and QTF (FMAT) time series. Each source of input must be preprocessed prior to executing the MEFP, ensuring it is the correct type of data and in the correct time step:

- *RFC and GEFS*: A single time series must be created which is 6-hour FMAP for precipitation and 24-hour TFMN/TFMX for temperature.
- *CFSv2*: A lagged ensemble must be created which is 6-hour FMAP for precipitation and 24-hour TFMN/TFMX for temperature.

Preprocessing workflows, as shown it the diagram, have been created for each forecast source to perform the necessary conversion tasks.

The overall MEFP operational forecast workflow process is as follows:

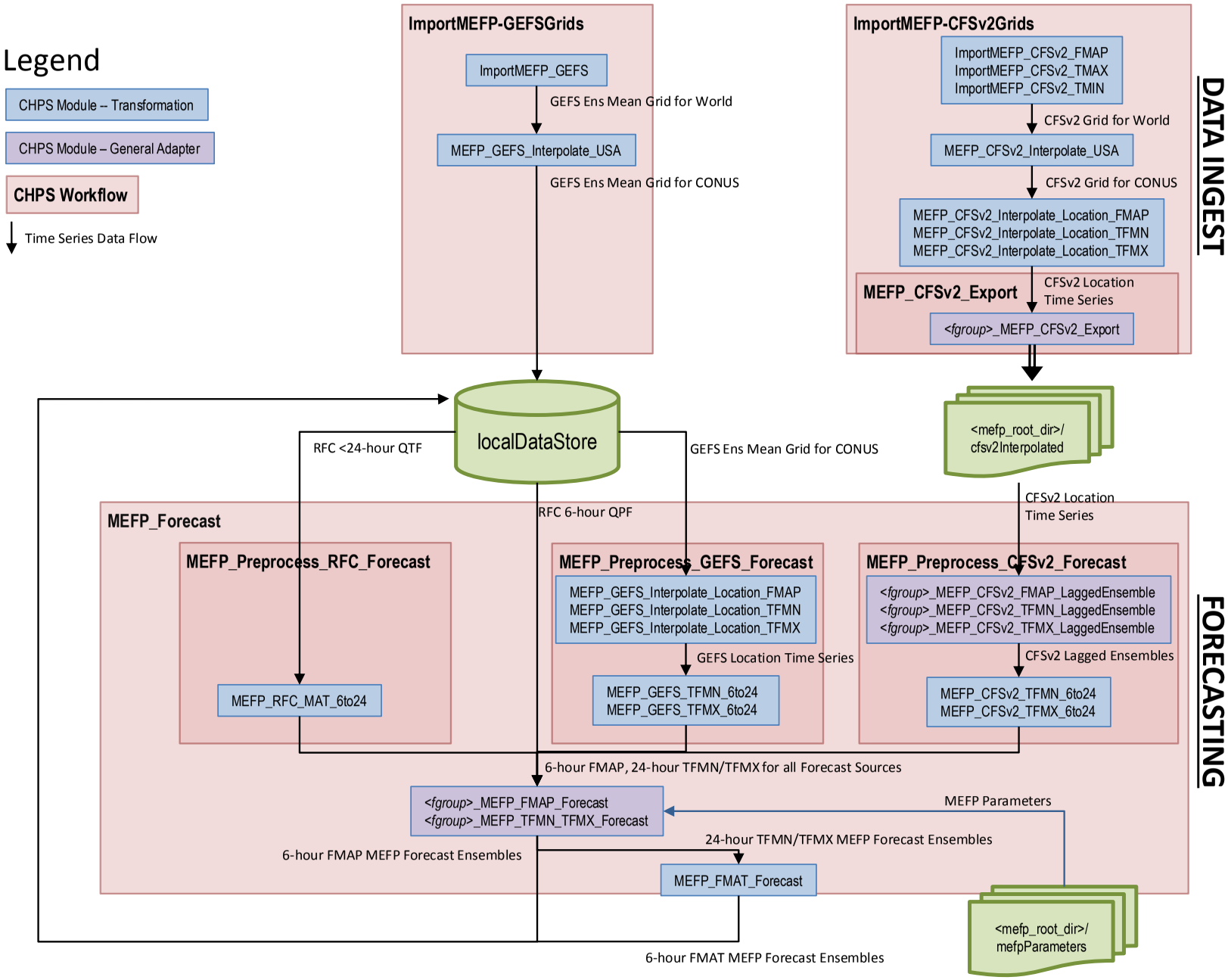
1. Preprocess RFC Forecasts
    a. Convert 6-hour QTF to 24-hour TFMN/TFMX
2. Preprocess GEFS Forecasts
    a. For each catchment for which MEFP is to execute, spatially interpolate a forecast time series from the gridded forecast ensemble mean time series.
    b. Convert the 6-hour TFMN/TFMX forecasts to 24-hour forecasts
3. Preprocess CFSv2 Forecasts
    a. Construct lagged ensembles of FMAP, TFMN, and TFMX for each catchment.
    b. Convert the 6-hour TFMN/TFMX ensembles to 24-hour ensembles
4. Execute the MEFPEnsembleGeneratorModelAdapter using the pre-processed RFC, GEFS, and CFSv2 forecasts, to generate 6-hour FMAP and 24-hour TFMN/TFMX forecast ensembles.
5. Convert the 24-hour TFMN/TFMX forecast ensembles to 6-hour FMAT ensembles applying a diurnal pattern

4.1.3 Hindcasting

Two changes to an operational configuration are required to execute MEFP in hindcast mode without errors:

- Comment out or remove the execution of <u>all</u> MEFP preprocessing forecast source workflows (Steps 1 – 4 in the previous section): MEFP_Preprocess_???_Forecast, where ??? is RFC, GEFS, and CFSv2.
- Set the MEFPEnsembleGeneratorModelAdapter run file property hindcasting to “true”. See Section 4.4.3.

More details are provided in the *MEFP Configuration Guide: Forecast Components*.


Legend
CHPS Module -- Transformation
CHPS Module – General Adapter
CHPS Workflow
Time Series Data Flow
DATA INGEST
ImportMEFP-GEFSGrids
ImportMEFP_GEFS
GEFS Ens Mean Grid for World
MEFP_GEFS_Interpolate_USA
GEFS Ens Mean Grid for CONUS
ImportMEFP-CFSv2Grids
ImportMEFP_CFSv2_FMAP
ImportMEFP_CFSv2_TMAX
ImportMEFP_CFSv2_TMIN
CFSv2 Grid for World
MEFP_CFSv2_Interpolate_USA
CFSv2 Grid for CONUS
MEFP_CFSv2_Interpolate_Location_FMAP
MEFP_CFSv2_Interpolate_Location_TFMN
MEFP_CFSv2_Interpolate_Location_TFMX
MEFP_CFSv2_Export
CFSv2 Location Time Series
<fgroup>_MEFP_CFSv2_Export
<mefp_root_dir>/ cfsv2Interpolated
localDataStore
RFC <24-hour QTF
GEFS Ens Mean Grid for CONUS
RFC 6-hour QPF
CFSv2 Location Time Series
FORECASTING
MEFP_Forecast
MEFP_Preprocess_RFC_Forecast
MEFP_RFC_MAT_6to24
MEFP_Preprocess_GEFS_Forecast
MEFP_GEFS_Interpolate_Location_FMAP
MEFP_GEFS_Interpolate_Location_TFMN
MEFP_GEFS_Interpolate_Location_TFMX
GEFS Location Time Series
MEFP_GEFS_TFMN_6to24
MEFP_GEFS_TFMX_6to24
MEFP_Preprocess_CFSv2_Forecast
<fgroup>_MEFP_CFSv2_FMAP_LaggedEnsemble
<fgroup>_MEFP_CFSv2_TFMN_LaggedEnsemble
<fgroup>_MEFP_CFSv2_TFMX_LaggedEnsemble
CFSv2 Lagged Ensembles
MEFP_CFSv2_TFMN_6to24
MEFP_CFSv2_TFMX_6to24
6-hour FMAP, 24-hour TFMN/TFMX for all Forecast Sources
<fgroup>_MEFP_FMAP_Forecast
<fgroup>_MEFP_TFMN_TFMX_Forecast
MEFP Parameters
24-hour TFMN/TFMX MEFP Forecast Ensembles
6-hour FMAP MEFP Forecast Ensembles
MEFP_FMAT_Forecast
<mefp_root_dir>/ mefpParameters
6-hour FMAT MEFP Forecast Ensembles

4.2 TimeSeriesExporterModelAdapter Reference Manual

The TimeSeriesExporterModelAdapter is used to export one or more time series to one or more files within a single run of the adapter. This section provides a description of the adapter relative to alternatives provided with FEWS and detailed instructions for how to configure the adapter.

4.2.1 Overview

The existing export capability in FEWS, using the timeSeriesExportRun module, allows for exporting time series by defining export XML elements, each of which specifies a single file to generate and the time series to output to that file. This can lead to excessive amounts of XML configuration. For example, in order for MEFP to use the CFSv2 forecast source, a file providing the spatially interpolated, single-valued forecast time series derived from the CFSv2 forecast grid must be created for each day and each catchment for which MEFP will generate ensembles. For precipitation, one such file must be generated, whereas for temperature, two such files must be generated (one for the maximum 24-h temperature and one for the minimum). Thus, using the timeSeriesExportRun module provided with FEWS, if a forecast group includes 20 catchments, 60 export files must be generated, so that 60 different export XML elements would need to be added to the timeSeriesExportRun XML element in the export configuration file.

The TimeSeriesExporterModelAdapter allows for file names to be constructed for each time series provided, so that all 60 files are created as defined by only a single set of input time series and single run of this adapter. The weakness of the TimeSeriesExporterModelAdapter, however, is that it will be more susceptible to running out of memory than the timeSeriesExportRun module provided with FEWS, as it must read in the entire set of input time series before generating files.

4.2.1.1 Example Configuration

The following is an example of a TimeSeriesExporterModelAdapter configuration file created for MEFP to generate location-specific CFSv2 forecast time series files:

```
<generalAdapterRun 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/generalAdapterRun.xsd">

  <!-- This module can run at any time relative to the CFSv2 data's external forecast time.  The forecast time assumed in output
     time series is computed based on provided data, as discussed below (see
     t0ComputationAdjustmentFactorFromFirstDataValue run info property). -->

  <general>
     <description>CFSv2 Interpolated Time Series Exporter</description>
     <piVersion>1.8</piVersion>
     <rootDir>%TEMP_DIR%</rootDir>
     <workDir>%ROOT_DIR%/work</workDir>
     <exportDir>%ROOT_DIR%/input</exportDir>
     <exportDataSetDir>%ROOT_DIR%</exportDataSetDir>
     <importDir>%ROOT_DIR%/output</importDir>
     <dumpFileDir>$GA_DUMPFILEDIR$</dumpFileDir>
```

```
    <dumpDir>%ROOT_DIR%</dumpDir>
    <diagnosticFile>%ROOT_DIR%/output/diag.xml</diagnosticFile>
  </general>
  <activities>

    <exportActivities>
      <exportTimeSeriesActivity>
        <exportFile>inputs.xml</exportFile>
        <timeSeriesSets>
          <timeSeriesSet>
            <moduleInstanceId>MEFP_CFSv2_Interpolate_Location_FMAP</moduleInstanceId>
            <valueType>scalar</valueType>
            <parameterId>FMAP</parameterId>
            <qualifierId>CFSv2</qualifierId>
            <locationSetId>Catchments_HEFS_FGroup_Export</locationSetId>
            <timeSeriesType>external forecasting</timeSeriesType>
            <timeStep unit="hour" multiplier="6"/>
            <readWriteMode>read complete forecast</readWriteMode>
            <ensembleId>main</ensembleId>
          </timeSeriesSet>
          <timeSeriesSet>
            <moduleInstanceId>MEFP_CFSv2_Interpolate_Location_TFMN</moduleInstanceId>
            <valueType>scalar</valueType>
            <parameterId>TFMN</parameterId>
            <qualifierId>CFSv2</qualifierId>
            <locationSetId>Catchments_HEFS_FGroup_Export</locationSetId>
            <timeSeriesType>external forecasting</timeSeriesType>
            <timeStep unit="hour" multiplier="6"/>
            <readWriteMode>read complete forecast</readWriteMode>
            <ensembleId>main</ensembleId>
          </timeSeriesSet>
          <timeSeriesSet>
            <moduleInstanceId>MEFP_CFSv2_Interpolate_Location_TFMX</moduleInstanceId>
            <valueType>scalar</valueType>
            <parameterId>TFMX</parameterId>
            <qualifierId>CFSv2</qualifierId>
            <locationSetId>Catchments_HEFS_FGroup_Export</locationSetId>
            <timeSeriesType>external forecasting</timeSeriesType>
            <timeStep unit="hour" multiplier="6"/>
            <readWriteMode>read complete forecast</readWriteMode>
            <ensembleId>main</ensembleId>
          </timeSeriesSet>
        </timeSeriesSets>
      </exportTimeSeriesActivity>

      <exportRunFileActivity>
        <exportFile>%ROOT_DIR%/run_info.xml</exportFile>
        <properties>
          <!-- Valid arguments to put within '@' symbols are locationId,
               parameterId, ensembleId, handbook5Id, and the forecastDateT0
               argument function which takes two parameters: date format
               and time zone.  This uses standard Graphics Generator
               arguments syntax, so refer to its documentation for more
               information. -->
          <!-- DO NOT CHANGE THE FILE NAME: only the extension can be changed
               (either .xml or .fi is valid). The CFSv2LaggedEnsembleModelAdapter
```

```
            model assumes the file name matches this pattern. -->
        <string key="fileNamePattern"
                value="@locationId@/@locationId@.@parameterId@.@forecastDateT0(yyyyMMddHH;GMT)@.xml"/>

        <!-- The base directory for output files.  Subdirectories  based
            on the file names are created as needed. -->
        <string key="exportDir" value="$MEFP_ROOT_DIR$/cfsv2Interpolated/archive"/>

        <!-- The external forecast time FEWS associates with the grid
            does not match the file name.  Hence, I use this mechanism
            to compute the forecast time relative to the data start time.-->
        <int key="t0ComputationAdjustmentFactorFromFirstDataValue" value="-6"/>
      </properties>
    </exportRunFileActivity>
  </exportActivities>

  <executeActivities>
    <executeActivity>
      <command>
        <className>ohd.hseb.hefs.exporter.adapter.TimeSeriesExporterModelAdapter</className>
      </command>
      <arguments>
        <argument>%ROOT_DIR%/run_info.xml</argument>
      </arguments>
      <timeOut>300000</timeOut>
    </executeActivity>
  </executeActivities>

 </activities>
</generalAdapterRun>
```

4.2.2 Model Parameters

The TimeSeriesExporterModelAdapter does not use parameters.

4.2.3 Model Run File Properties

The following exportRunFileActivity properties can be defined (for the Default Value column, "(required)" indicates that the property is required, so that there is no default value):

Property Name	Type	Default Value	Description
exportDir	text	(required)	Specifies the base directory where generated files will be placed. If it does not start with a '/', then the directory is located relative to the active directory of CHPS at run time.
fileNamePattern	text	(required)	Specifies the pattern to use to construct the names of the export files for each time series provided as input to the adapter. See Section 4.2.3.1.

t0ComputationAdjustment FactorFromFirstDataValue (one word)	Integer (negative values allowed)	Standard forecast time of adapter execution	For the purposes of building the file name using the fileNamePattern, if the forecast time is to be determined from the time series (i.e., the standard forecast time of the run of the adapter cannot be used), this provides an adjustment factor that is *added* to the time of the first non-missing data value in the time series. See Section 4.2.3.2.

Additional information for the latter two properties is provided next.

4.2.3.1 fileNamePattern Property

The fileNamePattern property uses Graphics Generator *arguments* (see the *Graphics Generator Getting Started* manual) in order to define the name of a file. An *argument* within the file name is preceded and followed by the character '@'. In the example in Section 4.2.1.1, the fileNamePattern is:

@locationId@/@locationId@.@parameterId@.@forecastDateT0(yyyyMMddHH;GMT)@.xml

The *arguments* used are "locationId", "parameterId", and "forecastDateT0(…)". At run time, the adapter replaces the *arguments* with values, as defined below, in order to determine file names.

The *argument* forecastDateT0 is referred to as an *argument function*, which is an *argument* that accepts parameters in order to determine its value. In this case, there are two parameters: the format of the date to output and the time zone.

NOTE: It is recommended that the fileNamePattern property always incorporate the forecastDateT0 argument function in some manner, otherwise the adapter may end up overwriting files from an earlier run.

The following table lists the acceptable *arguments* and *argument functions*:

Argument Name	Description
ensembleId	The ensembleId specified by the input time series.
forecastDateT0*	The system time in a given format and time zone (see below).
handbook5Id	The first five letters of the locationId specified by the input time series. For example, if the locationId is "WALN6DEL", the handbook5Id argument will have value "WALN6".
locationId	The locationId specified by the input time series.
parameterId	The parameterId specified by the input time series.

* Indicates an argument function.

The argument function forecastDateT0 includes the following parameters (in order, both are required), which must be separated by semicolons, ';':

Parameter Name	Description
Date Format	Date format: Text indicating the format of the date; see the Table below for valid characters to use. For example, for January 1, 2010 at 12 GMT, if the date format is "yyyyMMddHH", the output date will be "2010010112".
Time Zone	A valid Graphics Generator time zone string. Valid time zones include the following: "GMT", "AST", "AST/AKDT", "CST", "CST/CDT", "EST", "EST/EDT", "HST", "HST/HADT", "MST", "MST/MDT", "PST", "PST/PDT". Any "*ST" time zone is fixed to standard time, whereas a "*ST/*DT" time zone will vary based on daylight savings time. A more general format is also available: "Etc/GMT <+/–> #" where <+/–> is either "+" or "–" and # is a number of hours. For example, "Etc/GMT – 5" is equivalent to "EST".

Table: Valid characters for use in the Date Format parameter of the forecastDateT0 *argument function*.

Letter	Date or Time Component	Presentation	Examples
G	Era designator	Text	AD
y	Year	Year	1996; 96
M	Month in year	Month	July; Jul; 07
w	Week in year	Number	27
W	Week in month	Number	2
D	Day in year	Number	189
d	Day in month	Number	10
F	Day of week in month	Number	2
E	Day in week	Text	Tuesday; Tue
a	AM/PM marker	Text	PM
H	Hour in day (0-23)	Number	0
k	Hour in day (1-24)	Number	24
K	Hour in AM/PM (0-11)	Number	0
h	Hour in AM/PM (1-12)	Number	12
m	Minute in hour	Number	30
s	Second in minute	Number	55
S	Millisecond	Number	978
z	Time zone	General time zone	Pacific Standard Time; PST; GMT-08:00
Z	Time zone	RFC 822 time zone	-800

4.2.3.2 t0ComputationAdjustmentFactorFromFirstDataValue Property

The t0ComputationAdjustmentFactorFromFirstDataValue property defines the system time which is used to determine date strings based on the foreastDateT0 *argument function* within the fileNamePattern property. If the property is defined, then its value is added to the time of the

earliest non-missing data value in the time series. This allows for the file name to be determined based on the data instead of the system time (T0) of the model adapter run.

For example, if the property has value "-6" and the first non-missing data value in a time series is for January 1, 2010 at 18 GMT, then the forecast time used by the forecastDateT0 *argument function* within the fileNamePattern property will be January 1, 2010 at 12 GMT.

4.2.4 Model Input Time Series

The exportTimeSeriesActivity XML element in the configuration file specifies the time series to be exported. It is parsed, with each time series assigned to a file, using the process described in Section 4.2.7. There are no constraints on the time series, except for memory limitations.

4.2.5 Model Execution

The following executeActivity XML element should always be included within the executeActivities XML element in the configuration file (see the example in Section 4.2.1.1):

```
<executeActivity>
   <command>
      <className>ohd.hseb.hefs.exporter.adapter.TimeSeriesExporterModelAdapter</className>
   </command>
   <arguments>
      <argument>%ROOT_DIR%/run_info.xml</argument>
   </arguments>
   <timeOut>300000</timeOut>
</executeActivity>
```

Only the timeOut XML element, set to 5 minutes (300,000 milliseconds) above, should be modified if it is determined that it is not sufficient for the configured adapter run.

4.2.6 Model Output Time Series

This adapter does not generate any output time series. The configuration file should not include an importTimeSeriesActivity XML element.

4.2.7 Model Description

The TimeSeriesExporterModelAdapter algorithm follows this process:

1. Construct a file name for each input series it is provided. The file name is constructed using the fileNamePattern property (see Section 4.2.3.1).
2. Gather a list of unique file names (remove repeat file names) in order to determine which files must be generated.
3. For each file name…
    a. Gather all of the time series for which the file name matched that file.
    b. Output the time series to that file relative to the directory specified by the exportDir property. If the extension of the file is ".fi", the generated file will be in the

FastInfoset binary XML format. Otherwise, the generated file will be in standard ASCII XML.

This allows for any number of time series to be output to one file, but all those time series must yield the same file name when the fileNamePattern property is applied.

NOTE: The model adapter is fully compatible with FEWS id-mapping capabilities. The locationId, paramterId, and ensembleId of all input time series, which is used to determine file names via the fileNamePattern*, are the external identifiers defined in the id-mapping file applied to the module.*

4.2.7.1 Example File Names

File names are defined using the fileNamePattern run file property described Section 4.2.3.1 in conjunction with the t0ComputationAdjustmentFactorFromFirstDataValue described in Section 4.2.3.2.

For example, suppose the first non-missing value in a CFSv2 forecast time series with locationId "AAAAA" and parameterId "FMAP" is for 18Z on January 1, 2010.

Based on the configuration example in Section 4.2.1.1, suppose these properties are used:

```
<string key="fileNamePattern"
        value="@locationId@/@locationId@.@parameterId@.@forecastDateT0(yyyyMMddHH;GMT)@.xml"/>
<string key="exportDir" value="baseDir"/>
<int key="t0ComputationAdjustmentFactorFromFirstDataValue" value="-6"/>
```

Example 1: locationId "AAAAA", parameterId "FMAP", and first non-missing data value at 18Z on January 1, 2010 – file name is "baseDir/AAAAA/AAAAA.FMAP.20100112.xml".

Example 2: locationId "AAAAA", parameterId "FMAP", and first non-missing data value at 12Z on January 1, 2010 – file name is "baseDir/AAAAA/AAAAA.FMAP.20100106.xml".

Same scenario, but suppose the fileNamePattern is defined as follows:

```
<string key="fileNamePattern"
        value="@locationId@/@locationId@.@parameterId@.@forecastDateT0(yyyyMMddHH_z;GMT)@.xml"/>
```

Example 3: locationId "AAAAA", parameterId "FMAP", and first non-missing data value at 18Z on January 1, 2010 – file name is "baseDir/AAAAA/AAAAA.FMAP.20100112_GMT.xml".

4.3 CFSv2LaggedEnsembleModelAdapter Reference Manual

The CFSv2LaggedEnsembleModelAdapter is an MEFP-specific module used to generate lagged ensembles from location-specific CFSv2 forecast time series files stored under <*mefp_root_dir*>/cfsv2Interpolated/archive. The output from this adapter is used as input to the MEFPEnsembleGeneratorModelAdapter, described in Section 4.4.

NOTE: The only reason for the editing the existing CFSv2LaggedEnsembleModelAdapter module configuration files is to make them forecast group specific. For that purpose, focus on the moduleInstanceId *and* locationSetId *XML elements within the* exportTimeSeriesActivity *and* importTimeSeriesActivity *XML elements of the module configuration file.*

4.3.1 Overview

A lagged ensemble is constructed from multiple single-valued (one time series) forecasts generated at different forecast times, on which have a forecast time at T0. The ensemble is built by identifying a forecast time period: <*start date*> to <*end date*>. Then, for each time series used as input, the data spanning the forecast period is extracted, yielding a sub-series of the input time series. The sub-series are combined into a single ensemble spanning the time period from <*start date*> to <*end date*>. That ensemble is referred to as the lagged ensemble, since the source time series all have different forecast times (i.e., are lagged in time relative to T0).

For a given forecast group with name <*fgroup*>, the CFSv2LaggedEnsembleModelAdapter constructs lagged ensembles using data that is output by the TimeSeriesExporterModelAdapter module:

<*configuration_dir*>/ModuleConfigFiles/hefs/<*fgroup*>/<*fgroup*>_MEFP_CFSv2_Export.xml

The export files are location and forecast time specific containing CFSv2 forecast time series, as spatially interpolated from CFSv2 forecast grids. Each file must follow the naming convention dictated by the properties of the TimeSeriesExporterModelAdapter module:

<*locationId*>/<*locationId*>.<*parameterId*>.<*forecast date*>.xml

where the <*locationId*> and <*parameterId*> are derived from the time series, and the <*forecast date*> is the date/time in GMT 6-hours before the first non-missing value in the CFSv2 forecast time series (i.e., the forecast time of the CFSv2 grids from which the time series are interpolated) and has format “yyyyMMddHH” (for example: Jan 1, 2010, at 12Z is “2010010112”).

The output from CFSv2LaggedEnsembleModelAdapter is an ensemble of time series, either FMAT, TFMN, or TFMX, that has exactly 16 members, some of which may be empty (i.e., all missing) depending on data availability. The algorithm for constructing the lagged ensemble is described in detail in Section 4.3.7.

4.3.1.1 Example Configuration

The following is an example of a CFSv2LaggedEnsembleModelAdapter configuration file for generating lagged ensembles of FMAP forecast time series (the example is for the forecast group KEYINF at ABRFC; lines that may need to be modified for other forecast groups are highlighted):

```
<generalAdapterRun 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/generalAdapterRun.xsd">
  <general>
    <description>CFSv2 Lagged Ensemble Generator</description>
    <piVersion>1.8</piVersion>
    <rootDir>%TEMP_DIR%</rootDir>
    <workDir>%ROOT_DIR%/work</workDir>
    <exportDir>%ROOT_DIR%/input</exportDir>
    <exportDataSetDir>%ROOT_DIR%</exportDataSetDir>
    <importDir>%ROOT_DIR%/output</importDir>
    <dumpFileDir>$GA_DUMPFILEDIR$</dumpFileDir>
    <dumpDir>%ROOT_DIR%</dumpDir>
    <diagnosticFile>%ROOT_DIR%/output/diag3.xml</diagnosticFile>
  </general>
  <activities>

    <startUpActivities>
      <purgeActivity>
        <filter>%ROOT_DIR%/work/*</filter>
      </purgeActivity>
      <purgeActivity>
        <filter>%ROOT_DIR%/input/*</filter>
      </purgeActivity>
      <purgeActivity>
        <filter>%ROOT_DIR%/output/*</filter>
      </purgeActivity>
      <purgeActivity>
        <filter>%ROOT_DIR%/run_info.xml</filter>
      </purgeActivity>
    </startUpActivities>

    <exportActivities>
      <exportTimeSeriesActivity>
        <exportFile>inputs.xml</exportFile>
        <timeSeriesSets>

          <!-- Input series are used to identify locations for which to build lagged
               ensemble, as well as by CHPS to determine T0 -->
          <timeSeriesSet>
            <moduleInstanceId>MEFP_CFSv2_Interpolate_Location_FMAP</moduleInstanceId>
            <valueType>scalar</valueType>
            <parameterId>FMAP</parameterId>
            <qualifierId>CFSv2</qualifierId>
            <locationSetId>Catchments_HEFS_KEYINF</locationSetId>
            <timeSeriesType>external forecasting</timeSeriesType>
            <timeStep unit="hour" multiplier="6"/>
```

```
            <readWriteMode>read complete forecast</readWriteMode>
            <ensembleId>main</ensembleId>
          </timeSeriesSet>

        </timeSeriesSets>
      </exportTimeSeriesActivity>

      <exportRunFileActivity>
        <exportFile>%ROOT_DIR%/run_info.xml</exportFile>
        <properties>
          <!-- Must match the directory to which files are exported in the export module. -->
          <string key="cfsv2OperationalArchiveDirectory" value="$MEFP_ROOT_DIR$/cfsv2Interpolated/archive"/>
        </properties>
      </exportRunFileActivity>
    </exportActivities>

    <executeActivities>
      <executeActivity>
        <command>
          <className>ohd.hseb.hefs.mefp.adapter.CFSv2LaggedEnsembleModelAdapter</className>
          <binDir>$HEFSBINDIR$</binDir>
        </command>
        <arguments>
          <argument>%ROOT_DIR%/run_info.xml</argument>
        </arguments>
        <timeOut>3000000</timeOut>
      </executeActivity>
    </executeActivities>

    <importActivities>
      <importTimeSeriesActivity>
        <importFile>outputs.xml</importFile>
        <timeSeriesSets>
          <timeSeriesSet>
            <moduleInstanceId>KEYINF_MEFP_CFSv2_FMAP_LaggedEnsemble</moduleInstanceId>
            <valueType>scalar</valueType>
            <parameterId>FMAP</parameterId>
            <qualifierId>CFSv2</qualifierId>
            <locationSetId>Catchments_HEFS_KEYINF</locationSetId>
            <timeSeriesType>external forecasting</timeSeriesType>
            <timeStep unit="hour" multiplier="6"/>
            <readWriteMode>add originals</readWriteMode>
            <ensembleId>CFSv2</ensembleId>
          </timeSeriesSet>
        </timeSeriesSets>
      </importTimeSeriesActivity>
    </importActivities>

  </activities>
</generalAdapterRun>
```

4.3.2 Model Parameters

The CFSv2LaggedEnsembleModelAdapter does not use parameters.

4.3.3 Model Run File Properties

The following exportRunFileActivity properties can be defined (for the Default Value column, "(required)" indicates that the property is required, so that there is no default value):

Property Name	Type	Default Value	Description
cfsv2OperationalArchiveDirectory	text	(required)	Specifies the base directory for finding CFSv2 forecast time series files. This should always be set to the following: $MEFP_ROOT_DIR$/cfsv2Interpolated/archive
filePurgeWindowDays	integer	-1	If positive, then after constructing all lagged ensembles, if no errors occur, it will search the subdirectories under $MEFP_ROOT_DIR$/cfsv2Interpolated/archive that correspond to locations for which a lagged ensembles were created. In each directory, it will scan the files, parse the dates in the names of the files, and remove files that have a date OLDER than the current forecast time minus this window (in days). **By default, purging is done using a purge script delivered with the MEFP data ingest components and executed once per day.** See the *MEFP Configuration Guide: Data Ingest Components*. The default value of -1 for this property indicates that file purging, or removal, will not be done by this adapter.
firstFileSearchWindowHours	positive integer	24	Specifies the width of the window of time, ending at model run system time (T0), in which the first time series to be used to construct the lagged ensemble must be found. If no CFSv2 forecast time series file can be found for a time within that time window, the adapter will error out with an appropriate error message. It may be necessary to modify this property if this adapter must be run earlier than when 24-hour old CFSv2 data becomes available. In such a case, increase its value to an appropriate number (e.g., 48).

4.3.4 Model Input Time Series

The exportTimeSeriesActivity XML element in the configuration file exports time series that indicate for which locations (locationId) lagged ensembles need to be constructed. Otherwise, the data in the input time series is not used. The following exportTimeSeriesActivity XML element should always be used to generate precipitation (FMAP) lagged ensembles, replacing *fgroup* with the appropriate forecast group name:

```
<exportTimeSeriesActivity>
  <exportFile>inputs.xml</exportFile>
  <timeSeriesSets>
    <!-- Input series are used to identify locations for which to build lagged
         ensemble, as well as by CHPS to determine T0 -->
    <timeSeriesSet>
      <moduleInstanceId>MEFP_CFSv2_Interpolate_Location_FMAP</moduleInstanceId>
      <valueType>scalar</valueType>
      <parameterId>FMAP</parameterId>
      <qualifierId>CFSv2</qualifierId>
      <locationSetId>Catchments_HEFS_fgroup</locationSetId>
      <timeSeriesType>external forecasting</timeSeriesType>
      <timeStep unit="hour" multiplier="6"/>
      <readWriteMode>read complete forecast</readWriteMode>
      <ensembleId>main</ensembleId>
    </timeSeriesSet>
  </timeSeriesSets>
</exportTimeSeriesActivity>
```

To generate minimum temperature or maximum temperature lagged ensembles, the exportTimeSeriesActivity XML element is the same, but the two highlighted instances of "FMAP" are replaced by "TFMN" or "TFMX", respectively.

NOTE: Do not change the exportFile *XML element's value.*

4.3.5 Model Execution

The following executeActivity XML element should always be included within the executeActivities XML element in the configuration file (see the example in Section 4.2.1.1):

```
<executeActivity>
  <command>
    <className>ohd.hseb.hefs.mefp.adapter.CFSv2LaggedEnsembleModelAdapter</className>
  </command>
  <arguments>
    <argument>%ROOT_DIR%/run_info.xml</argument>
  </arguments>
  <timeOut>300000</timeOut>
</executeActivity>
```

Only the timeOut XML element, set to 5 minutes (300,000 milliseconds) above, should be modified if it is determined that it is not sufficient for the configured adapter run.

4.3.6 Model Output Time Series

This adapter outputs one lagged ensemble per location. The following importTimeSeriesActivity XML element should always be used when importing precipitation (FMAP) lagged ensembles, replacing *fgroup* with the appropriate forecast group name:

```
<importActivities>
  <importTimeSeriesActivity>
    <importFile>outputs.xml</importFile>
    <timeSeriesSets>
      <timeSeriesSet>
        <moduleInstanceId>fgroup_MEFP_CFSv2_FMAP_LaggedEnsemble</moduleInstanceId>
        <valueType>scalar</valueType>
        <parameterId>FMAP</parameterId>
        <qualifierId>CFSv2</qualifierId>
        <locationSetId>Catchments_HEFS_fgroup</locationSetId>
        <timeSeriesType>external forecasting</timeSeriesType>
        <timeStep unit="hour" multiplier="6"/>
        <readWriteMode>add originals</readWriteMode>
        <ensembleId>CFSv2</ensembleId>
      </timeSeriesSet>
    </timeSeriesSets>
  </importTimeSeriesActivity>
</importActivities>
```

To import generated minimum temperature or maximum temperature lagged ensembles, the importTimeSeriesActivity XML element is the same, but the two highlighted instances of "FMAP" are replaced by "TFMN" or "TFMX", respectively.

The time series imported here are used as input to either the MEFPEnsembleGeneratorModelAdapter (for precipitation) or a module to aggregate the 6-hour minimum and maximum temperature time series to the 24-hour minimum and maximum temperature time series (see the *MEFP Configuration Guide: Forecast Components*). Regardless, it is critical that all of the XML elements are identical to what is provided above (qualifierId, ensembleId, etc.), except for changes related to data type or forecast group name as described above.

NOTE: Do not change the importFile *XML element's value.*

4.3.7 Model Description

The CFSv2LaggedEnsembleModelAdapter algorithm follows this process:

1. Identify the current forecast time of the workflow run.
2. Define a time window using the value of the firstFileSearchWindowHours run file property.
3. For each time series in the input time series…
    a. Identify its locationId and parameterId.

b. Starting from the identified current forecast time, check if the file *<locationId>*/*<locationId>*.*<parameterId>*.*<forecast date>*.xml exists under the directory pointed to by the cfsv2OperationalArchiveDirectory run file property. If the file is not found, subtract 6-hours from the working forecast time just checked and try again. Repeat until a file is found or until the working forecast time is outside the time window computed above. If it is outside the time window, error out.
c. With the first CFSv2 forecast time series file identified, the other 15 forecast time files are identified by their forecast times as follows:

Member Index	Assumed forecast time relative to the current forecast time
2	current forecast time – 5 days at 18Z
3	current forecast time – 5 days at 12Z
4	current forecast time – 5 days at 6Z
5	current forecast time – 5 days at 0Z
6	current forecast time – 10 days at 18Z
7	current forecast time – 10 days at 12Z
8	current forecast time – 10 days at 6Z
9	current forecast time – 10 days at 0Z
10	current forecast time – 15 days at 18Z
11	current forecast time – 15 days at 12Z
12	current forecast time – 15 days at 6Z
13	current forecast time – 15 days at 0Z
14	current forecast time – 20 days at 18Z
15	current forecast time – 20 days at 12Z
16	current forecast time – 20 days at 6Z

d. For each time series file, extract the subseries from T0 through T0 + 270 days. Add the subseries to the lagged ensemble. If the time series file for a particular member does not exist, then an ensemble member with all missing values is included in the lagged ensemble.

4. IF the filePurgeWindowDays run file property is positive, then for each time series in the input time series…
   a. Identify its locationId.
   b. Remove all files under *<mefp_root_dir>*/cfsv2Interpolated/archive/<locationId> that have a date in the file name that is old; see filePurgeWindowDays in Section 4.3.3 for more details.

The lagged ensemble resulting from the steps above is output to the outputs.xml file so that CHPS can import it as instructed in the module configuration file.

NOTE: The model adapter is fully compatible with FEWS id-mapping capabilities. However, given that the exportTimeSeriesActivity *specifies* timeSeriesSet *XML elements that match those used to export the CFSv2 forecast time series files, an id-mapping should not be necessary unless the TimeSeriesExporterModelAdapter module used to export the files also uses an identical id-mapping.*

4.4 MEFPEnsembleGeneratorModelAdapter Reference Manual

The MEFPEnsembleGeneratorModelAdapter is an MEFP-specific module used to generate forecast ensembles of precipitation (CHPS parameterId FMAP) and minimum and maximum temperature (CHPS parameterIds TFMN and TFMX). It executes the MEFP algorithm described in Section 2.

NOTE: There are two reasons to modify the MEFPEnsembleGeneratorModelAdatper module configuration files provided with the release. The first reason is to specify the time series that provide the RFC forecast source QPF and QTF time series. For that purpose, focus on the RFC forecast source time series exported first in the exportTimeSeriesActivity *XML element. The second reason is to make the file forecast group specific. For that purpose, focus on the* moduleInstanceId *and* locationSetId *XML elements within the* exportTimeSeriesActivity *and* importTimeSeriesActivity *XML elements of the module configuration file and see the MEFP Configuration Guide: Forecast Components.*

4.4.1 Overview

The MEFP model requires input time series for each forecast source that will be used as input to the MEFP algorithm; See Section 2.3. The time series required for each MEFP location by source and data type are as follows:

Source	Precipitation	Temperature
RFC (QPF/QTF)	6-hour FMAP (usually) time series up to 5 days in length specifying the QPF for an MEFP location.	24-hour TFMN and TFMX time series up to 7 days in length specifying the QTF.
GEFS	Ensemble mean of 6-hour FMAP time series spatially interpolated by MEFP location from ensemble forecast grids.	Ensemble mean of 24-hour TFMN and TFMX time series spatially interpolated by MEFP location from ensemble forecast grids.
CFSv2	Complete lagged ensemble of 6-hour FMAP time series, each having been spatially interpolated from forecast grids.	Complete lagged ensemble of 24-hour TFMN and TFMX time series, each having been spatially interpolated from forecast grids.

Note the following:

- CHPS/FEWS modules are provided with the release of MEFP for calculating the 24-hour TFMN and TFMX time series for each forecast source.
- For the GEFS forecast source, the GEFS forecast grids specify the ensemble mean directly; no additional computations/transformations are required.
- For the CFSv2 forecast source, the entire lagged ensemble must be provided, since MEFP must compute ensemble mean values for canonical events (see Section 2.4.2) itself.

Throughout the algorithm described in Section 2, given run-time options set as run-file properties via the module configuration file, the MEFP generates a forecast ensemble of either 6-hour FMAP time series or 24-hour TFMN and TFMX time series.

4.4.1.1 Example Configuration

The following two example configuration files are for execution of the MEFPEnsembleGeneratorModelAdapter to generate precipitation (Example 1) and temperature (Example 2) forecast ensembles. Both configuration files match those delivered by default with the release of MEFP. The portions highlighted are RFC-specific and must be modified as described in the *MEFP Configuration Guide: Forecast Components*.

Example 1: Default module configuration file for generating precipitation ensembles.

```
<generalAdapterRun 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/generalAdapterRun.xsd">
  <general>
    <description>MEFP Ensemble Generator</description>
    <piVersion>1.8</piVersion>
    <rootDir>%TEMP_DIR%</rootDir>
    <workDir>%ROOT_DIR%/work</workDir>
    <exportDir>%ROOT_DIR%/input</exportDir>
    <exportDataSetDir>%ROOT_DIR%</exportDataSetDir>
    <exportIdMap>IdExportMEFPMAP</exportIdMap>
    <importDir>%ROOT_DIR%/output</importDir>
    <dumpFileDir>$GA_DUMPFILEDIR$</dumpFileDir>
    <dumpDir>%ROOT_DIR%</dumpDir>
    <diagnosticFile>%ROOT_DIR%/output/diag.xml</diagnosticFile>
  </general>
  <activities>

    <startUpActivities>
    </startUpActivities>

    <exportActivities>
      <exportTimeSeriesActivity>
        <exportFile>inputs.xml</exportFile>
        <timeSeriesSets>

          <!-- RFC FMAP (future MAP, in this case) time series -->
          <timeSeriesSet>
            <moduleInstanceId>FMAP_PreProcessing_QPF</moduleInstanceId>
            <valueType>scalar</valueType>
            <parameterId>FMAP</parameterId>
            <locationSetId>Catchments_HEFS_FGroup</locationSetId>
            <timeSeriesType>simulated forecasting</timeSeriesType>
            <timeStep unit="hour" multiplier="6"/>
            <relativeViewPeriod unit="day" start="0" startOverrulable="true" end="5" endOverrulable="false"/>
            <readWriteMode>read only</readWriteMode>
          </timeSeriesSet>
```

```
        <!-- GEFS FMAP Ensemble Mean -->
        <timeSeriesSet>
          <moduleInstanceId>MEFP_GEFS_Interpolate_Location_FMAP</moduleInstanceId>
          <valueType>scalar</valueType>
          <parameterId>FMAP</parameterId>
          <qualifierId>GEFS</qualifierId>
          <locationSetId>Catchments_HEFS_FGroup</locationSetId>
          <timeSeriesType>external forecasting</timeSeriesType>
          <timeStep unit="hour" multiplier="6"/>
          <relativeViewPeriod unit="day" start="0" startOverrulable="true" end="16" endOverrulable="false"/>
          <readWriteMode>read only</readWriteMode>
        </timeSeriesSet>

        <!-- CFSv2 entire lagged ensemble -->
        <timeSeriesSet>
          <moduleInstanceId>FGroup_MEFP_CFSv2_FMAP_LaggedEnsemble</moduleInstanceId>
          <valueType>scalar</valueType>
          <parameterId>FMAP</parameterId>
          <qualifierId>CFSv2</qualifierId>
          <locationSetId>Catchments_HEFS_FGroup</locationSetId>
          <timeSeriesType>external forecasting</timeSeriesType>
          <timeStep unit="hour" multiplier="6"/>
          <relativeViewPeriod unit="day" start="0" startOverrulable="true" end="270" endOverrulable="false"/>
          <readWriteMode>read only</readWriteMode>
          <ensembleId>CFSv2</ensembleId>
        </timeSeriesSet>

      </timeSeriesSets>
    </exportTimeSeriesActivity>

    <exportRunFileActivity>
      <exportFile>%ROOT_DIR%/run_info.xml</exportFile>
      <properties>
        <int key="printDebugInfo" value="0"/>
        <string key="parameterDir" value="$MEFP_ROOT_DIR$/mefpParameters"/>
        <int key="rfcNumberOfForecastDays" value="0"/>
        <int key="gefsNumberOfForecastDays" value="15"/>
        <int key="cfsv2NumberOfForecastDays" value="270"/>
        <int key="climatologyNumberOfForecastDays" value="330"/>
        <string key="rfcUseEPT" value="true"/>
        <string key="gefsUseEPT" value="true"/>
        <string key="useResampledClimatology" value="true"/>
        <int key="initialEnsembleYear" value="1961"/>
        <int key="lastEnsembleYear" value="1997"/>
        <string key="eptUseStratifiedSampling" value="true"/>
      </properties>
    </exportRunFileActivity>
  </exportActivities>

  <executeActivities>
    <executeActivity>
      <command>
        <className>ohd.hseb.hefs.mefp.adapter.MEFPEnsembleGeneratorModelAdapter</className>
        <binDir>$HEFSBINDIR$</binDir>
      </command>
      <arguments>
```

```
            <argument>%ROOT_DIR%/run_info.xml</argument>
          </arguments>
          <timeOut>300000</timeOut>
        </executeActivity>
      </executeActivities>

      <importActivities>
        <importTimeSeriesActivity>
          <importFile>outputs.xml</importFile>
          <timeSeriesSets>
            <timeSeriesSet>
              <moduleInstanceId>FGroup_MEFP_FMAP_Forecast</moduleInstanceId>
              <valueType>scalar</valueType>
              <parameterId>FMAP</parameterId>
              <locationSetId>Catchments_HEFS_FGroup</locationSetId>
              <timeSeriesType>external forecasting</timeSeriesType>
              <timeStep unit="hour" multiplier="6"/>
              <readWriteMode>add originals</readWriteMode>
              <ensembleId>MEFP</ensembleId>
            </timeSeriesSet>
          </timeSeriesSets>
        </importTimeSeriesActivity>
      </importActivities>

    </activities>
</generalAdapterRun>
```

Example 2: Default module configuration file for generating temperature ensembles.

```
<generalAdapterRun 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/generalAdapterRun.xsd">
    <general>
      <description>MEFP Ensemble Generator</description>
      <piVersion>1.8</piVersion>
      <rootDir>%TEMP_DIR%</rootDir>
      <workDir>%ROOT_DIR%/work</workDir>
      <exportDir>%ROOT_DIR%/input</exportDir>
      <exportDataSetDir>%ROOT_DIR%</exportDataSetDir>
      <exportIdMap>IdExportMEFPMAP</exportIdMap>
      <importDir>%ROOT_DIR%/output</importDir>
      <dumpFileDir>$GA_DUMPFILEDIR$</dumpFileDir>
      <dumpDir>%ROOT_DIR%</dumpDir>
      <diagnosticFile>%ROOT_DIR%/output/diag.xml</diagnosticFile>
    </general>
    <activities>

      <startUpActivities>
      </startUpActivities>
      <exportActivities>
        <exportTimeSeriesActivity>
          <exportFile>inputs.xml</exportFile>
          <timeSeriesSets>
```

```
<!-- RFC FMAT (future MAT, in this case) time series -->
<timeSeriesSet>
   <moduleInstanceId>MEFP_RFC_MAT_6to24</moduleInstanceId>
   <valueType>scalar</valueType>
   <parameterId>TFMN</parameterId>
   <locationSetId>Catchments_HEFS_FGroup</locationSetId>
   <timeSeriesType>simulated forecasting</timeSeriesType>
   <timeStep id="12Z"/>
   <relativeViewPeriod unit="day" start="0" startOverrulable="true" end="5" endOverrulable="false"/>
   <readWriteMode>read only</readWriteMode>
</timeSeriesSet>
<timeSeriesSet>
   <moduleInstanceId>MEFP_RFC_MAT_6to24</moduleInstanceId>
   <valueType>scalar</valueType>
   <parameterId>TFMX</parameterId>
   <locationSetId>Catchments_HEFS_FGroup</locationSetId>
   <timeSeriesType>simulated forecasting</timeSeriesType>
   <timeStep id="12Z"/>
   <relativeViewPeriod unit="day" start="0" startOverrulable="true" end="5" endOverrulable="false"/>
   <readWriteMode>read only</readWriteMode>
</timeSeriesSet>

<!-- GEFS TFMX/TFMN -->
<timeSeriesSet>
   <moduleInstanceId>MEFP_GEFS_TFMN_6to24</moduleInstanceId>
   <valueType>scalar</valueType>
   <parameterId>TFMN</parameterId>
   <qualifierId>GEFS</qualifierId>
   <locationSetId>Catchments_HEFS_FGroup</locationSetId>
   <timeSeriesType>external forecasting</timeSeriesType>
   <timeStep times="12:00"/>
   <relativeViewPeriod unit="day" start="0" startOverrulable="true" end="15" endOverrulable="false"/>
   <readWriteMode>read only</readWriteMode>
</timeSeriesSet>
<timeSeriesSet>
   <moduleInstanceId>MEFP_GEFS_TFMX_6to24</moduleInstanceId>
   <valueType>scalar</valueType>
   <parameterId>TFMX</parameterId>
   <qualifierId>GEFS</qualifierId>
   <locationSetId>Catchments_HEFS_FGroup</locationSetId>
   <timeSeriesType>external forecasting</timeSeriesType>
   <timeStep times="12:00"/>
   <relativeViewPeriod unit="day" start="0" startOverrulable="true" end="15" endOverrulable="false"/>
   <readWriteMode>read only</readWriteMode>
</timeSeriesSet>

<!-- CFSv2 TFMN/TFMX Lagged Ensemble -->
<timeSeriesSet>
   <moduleInstanceId>MEFP_CFSv2_TFMN_6to24</moduleInstanceId>
   <valueType>scalar</valueType>
   <parameterId>TFMN</parameterId>
   <qualifierId>CFSv2</qualifierId>
   <locationSetId>Catchments_HEFS_FGroup</locationSetId>
   <timeSeriesType>external forecasting</timeSeriesType>
   <timeStep id="12Z"/>
   <relativeViewPeriod unit="day" start="0" startOverrulable="true" end="270" endOverrulable="true"/>
```

```
            <readWriteMode>read only</readWriteMode>
            <ensembleId>CFSv2</ensembleId>
         </timeSeriesSet>
         <timeSeriesSet>
            <moduleInstanceId>MEFP_CFSv2_TFMX_6to24</moduleInstanceId>
            <valueType>scalar</valueType>
            <parameterId>TFMX</parameterId>
            <qualifierId>CFSv2</qualifierId>
            <locationSetId>Catchments_HEFS_FGroup</locationSetId>
            <timeSeriesType>external forecasting</timeSeriesType>
            <timeStep id="12Z"/>
            <relativeViewPeriod unit="day" start="0" startOverrulable="true" end="270" endOverrulable="true"/>
            <readWriteMode>read only</readWriteMode>
            <ensembleId>CFSv2</ensembleId>
         </timeSeriesSet>
      </timeSeriesSets>
   </exportTimeSeriesActivity>

   <exportRunFileActivity>
      <exportFile>%ROOT_DIR%/run_info.xml</exportFile>
      <properties>
         <int key="printDebugInfo" value="0"/>
         <string key="parameterDir" value="$MEFP_ROOT_DIR$/mefpParameters"/>
         <int key="rfcNumberOfForecastDays" value="0"/>
         <int key="gefsNumberOfForecastDays" value="15"/>
         <int key="cfsv2NumberOfForecastDays" value="270"/>
         <int key="climatologyNumberOfForecastDays" value="330"/>
         <string key="rfcUseEPT" value="true"/>
         <string key="gefsUseEPT" value="true"/>
         <string key="useResampledClimatology" value="true"/>
         <int key="initialEnsembleYear" value="1961"/>
         <int key="lastEnsembleYear" value="1997"/>
      </properties>
   </exportRunFileActivity>
</exportActivities>

<executeActivities>
   <executeActivity>
      <command>
         <className>ohd.hseb.hefs.mefp.adapter.MEFPEnsembleGeneratorModelAdapter</className>
         <binDir>$HEFSBINDIR$</binDir>
      </command>
      <arguments>
         <argument>%ROOT_DIR%/run_info.xml</argument>
      </arguments>
      <timeOut>300000</timeOut>
   </executeActivity>
</executeActivities>

<importActivities>
   <importTimeSeriesActivity>
      <importFile>outputs.xml</importFile>
      <timeSeriesSets>
         <timeSeriesSet>
            <moduleInstanceId>FGroup_MEFP_TFMN_TFMX_Forecast</moduleInstanceId>
            <valueType>scalar</valueType>
```

```
            <parameterId>TFMN</parameterId>
            <locationSetId>Catchments_HEFS_FGroup</locationSetId>
            <timeSeriesType>external forecasting</timeSeriesType>
            <timeStep times="12:00"/>
            <readWriteMode>add originals</readWriteMode>
            <ensembleId>MEFP</ensembleId>
          </timeSeriesSet>
          <timeSeriesSet>
            <moduleInstanceId>FGroup_MEFP_TFMN_TFMX_Forecast</moduleInstanceId>
            <valueType>scalar</valueType>
            <parameterId>TFMX</parameterId>
            <locationSetId>Catchments_HEFS_FGroup</locationSetId>
            <timeSeriesType>external forecasting</timeSeriesType>
            <timeStep times="12:00"/>
            <readWriteMode>add originals</readWriteMode>
            <ensembleId>MEFP</ensembleId>
          </timeSeriesSet>
        </timeSeriesSets>
      </importTimeSeriesActivity>
    </importActivities>
  </activities>
</generalAdapterRun>
```

4.4.2 Model Parameters

The MEFPEnsembleGeneratorModelAdapter does not use parameters.

4.4.3 Model Run File Properties

Some model run file properties can be defined for each forecast source. In such a case, the name of the forecast source in lower case letters is the prefix of the property name; for example:

```
<int key="rfcNumberOfForecastDays" value="5"/>
```

In this example, the number of forecast days for the RFC source is defined to be 5 days, by default (for all MEFP locations). Valid source prefixes include "rfc", "gefs", and "cfsv2". The name of each forecast source-specific property is denoted in the table below with the prefix "<source>".

Some model run file properties can be defined by MEFP location, overriding default properties defined for a source. In such cases, the location for which the property applies is specified as a suffix after a '.' within the name of the property; for example:

```
<int key="rfcNumberOfForecastDays.CREC1HOF" value="3"/>
```

In this example, the number of forecast days for the RFC source is defined to be 3 days for the location CREC1HOF (a CNRFC example). The name of each location specific property is denoted in the table below with the suffix "[.<locationId>]".

The properties are listed in two tables below based on whether or not they are considered advanced properties. Advanced properties should only be set after gaining enough familiarity with the algorithm to understand their impact.

The following exportRunFileActivity properties can be defined (for the Default Value column, "(required)" indicates that the property is required, so that there is no default value):

Basic Properties:

NOTE: For all properties of type "Boolean", if the value of the property is "true", then the property is true. Any other value is treated as false.

NOTE: A hydrologic water year starts on Oct 1of the preceding year and ends Sep 30 of. For example, hydrologic water year 1997 is from 10/1/1996 through 9/30/1997. To check which years of data are available, use the diagnostics of the **Historical Data Subpanel** *within the MEFPPE (see Section 3.6).*

Property Name	Type	Default Value	Description
<source>NumberOfForecastDays[.<locationId>]	Integer (zero or positive)	(required)	Identifies the number of forecast days for which to apply a forecast source when generating ensembles. For an MEFP location, if this property is not specified for that location and there is no default defined (i.e., no property defined without the ".<locationId>" suffix), then that source is not used. The final output ensemble generated by MEFP is a merging of forecast ensembles generated from all used forecast sources. For a given lead time, the ensemble member values used is controlled by the source assigned to that lead time. The order of precedence for assigning sources to lead times is as follows: RFC, GEFS, CFSv2, climatology. For example, if the number of days for RFC is set to 3 and GEFS is set to 15, then theRFC-based results will be used through day 3 and GEFS-based results will be used in the final generated ensembles for days 4-15.

behaviorIfEventMissing	String (see desc.)	“errorOut”	Must be one of three values: “errorOut”, “fillMissing”, “fillClimatology”. If “errorOut”, then if a canonical event value for an input cannot be computed, the adapter will generate an error and quit. If “fillClimatology”, then if a canonical event value cannot be computed, the resulting ensemble will use raw climatology for the times that correspond to that event value. If “fillMissing”, then if a canonical event value cannot be computed, the resulting forecast ensemble will contain missing values for the times that correspond to that event value. For “fillMissing” and “fillClimatology”, if another canonical event modifies the same time period later in the algorithm, then it may overwrite the effect of this property for the current event.
hindcasting	Boolean	false	If true, then the MEFP is executed in hindcasting mode, which means that, for an MEFP location, the canonical event values used to generate a forecast ensemble are pulled from the parameter file for that location instead of based on provided input time series. See Section 2.4.2. In the release, the template file provided will make use of a global property MEFP_HINDCASTING in defining this property. If used, to turn on hindcasting for all MEFPEnsembleGeneratorModelAdapter modules, set the global property to “true”. **NOTE**: The input time series are still required, even if they are all missing, when running in hindcast mode.
initialEnsembleYear	Integer (valid year)	(required)	The first historical, hydrologic water year to use to generate an ensemble member (defines the smallest member index of the generated ensemble). The smallest usable value is the <u>first</u> year such that 10/1/<year> is within the historical record for all generated MEFP catchments.

lastEnsembleYear	Integer (valid year)	(required)	The last historical, hydrologic water year to use to generate an ensemble member (defines the largest member index of the generated ensemble). The largest usable value is the <u>last</u> year such that 9/30/<year + 1> is within the historical record for the MEFP catchment. <year + 1> is used to allow for forecast up to 1 year in length. If longer forecasts are desired, change it to <year + 2>, which allows for forecasts up to 2 years in length.
parameterDir	Text	The work directory for the module run	Specifies the base directory under which all MEFP parameter files generated by MEFP can be found. Though the default is defined to use the work directory, this property should typically be set to the following (see the examples): $MEFP_ROOT_DIR$/mefpParameters MEFP_ROOT_DIR is a global property defined during configuration; see the *MEFP Configuration Guide: Data Ingest Components*. Only change this property value if told to in order to debug a problem.
useResampledClimatology	Boolean	true	If true, resampled climatology (see Section 2) will be used for the climatology portion (based on the number of forecast days defined for the climatology forecast source). If false, raw climatology will be used, meaning that the ensemble members will directly reflect observations in the historical record. *NOTE: For temperature, raw climatology refers to the 24-hour TMIN and TMAX values provided during parameter estimation. Diurnal computations performed after MEFP executes will not match the original 6-hour historical MAT values.* *NOTE: It is recommended that if raw climatology is desired, then do not include climatology in the MEFP generated ensemble, and, instead, append climatology to the ensemble generated by MEFP afterwards via sample and merge transformations; see the MEFP Configuration Guide: Forecast Components.*

Advanced Properties:

Property Name	Type	Default Value	Description
<source>UseEPT	Boolean	false	If true, then the EPT model is used for that forecast source; if false, the IPT model is used. This property is used if set for the RFC or GEFS forecast sources, but is ignored for CFSv2 and climatology forecast sources. See Section 2.4.4.
<source>ExcludeBaseEvents	Boolean	false	If true, then all base canonical events are not used for that forecast source in generating a forecast ensemble. See Section 2.4.2. This can be used in conjunction with both "…ExcludeEventsWithDur…" properties.
<source>ExcludeModulationEvents	Boolean	true	If true, then all modulation canonical events are not used for that forecast source in generating a forecast ensemble. See Section 2.4.2. This can be used in conjunction with both "…ExcludeEventsWithDur…" properties.
<source>ExcludeEventsWithDurLessThan	Integer	(not used)	If present, then only those events whose duration (end period – start period + 1) equals or exceeds the provided number are used for that forecast source in generating a forecast ensemble. This can be used in conjunction with the "…ExcludeEventsWithDurMoreThan" property. **NOTE:** The unit of this property is data type dependent (6-hour period for precipitation, 1 day for temperature). As such, if this property is specified, then the adapter should be configured to only generate forecast ensemble of one data type for that run (see below).
<source>ExcludeEventsWithDurMoreThan	Integer	(not used)	If present, then only those events whose duration (end period – start period + 1) equals or exceeds the provided number are used for that forecast source in generating a forecast ensemble. This can be used in conjunction with the "…ExcludeEventsWithDurLessThan" property. **NOTE:** The unit of this property is data type dependent (6-hour period for precipitation, 1 day for temperature). As such, if this property is specified, then the adapter should be configured to only generate forecast ensemble of one data type for that run (see below).

<source>OneEventOnly	Integer, Integer	(not used)	If present, the only one canonical event is used for that forecast source in generating a forecast ensemble. The format of the property value must be the following: <start>,<end> where <start> defines the start period of the canonical event to use and <end> defines the end period. **NOTE:** The unit of this property is data type dependent (6-hour period for precipitation, 1 day for temperature). As such, if this property is specified, then the adapter should be configured to only generate forecast ensembles of one data type for that run (see below).
condCoeffVarMax	Float	2.0	Specifies the maximum allowable value for the coefficient of variation used to generate an ensemble. If the coefficient of variation computed by MEFPPE and provided in the parameter file exceeds this amount, it will be set to this amount during ensemble generation. See Section 2.4.4.
eptUseStratifiedSampling	Boolean	false	Used for testing: forces the MEFP to use stratified sampling within the EPT algorithm when sampling from a forecast distribution to generate an ensemble. This makes the output from the EPT predictable. When set to false, it uses random sampling. See Section 2.4.4.
memberIndexingYear	String (see desc.)	standardHydroWaterYear	Defines the year used to determine the calendar year corresponding to a member index when constructing a base ensemble; see Section 4.4.8.1. There are three possible values: "standardHydroWaterYear" – The year used is the standard ESP hydrologic water year, which starts on 10/1 and ends on 9/30. For example, the 1950 water year starts on 10/1/1949 and ends 9/30/1950. "calendarYear" – The year used is a normal calendar year: 1/1 – 12/31. "--mm-dd" – A string that matches the format of the startWaterYear parameter of a sample-historical FEWS transformation. It defines the day of the year that *starts* the water year. For example, if it is set to "--07-04", then for member index 1950, the years starts on 7/4/1949 and ends 7/3/1950. The value "--10-01" is equivalent to "standardHydroWaterYear".

testing	Boolean	false	Used for testing: seeds the random number generator used in the Schaake shuffle so that it always produces the same results This makes the output from the MEFP predictable. When set to false, it uses a random seed. See Section 2.4.4

4.4.4 Model Input Time Series

The exportTimeSeriesActivity XML element in the configuration file exports time series that indicate for which locations (locationId) forecast ensembles need to be generated and provides inputs for each forecast source to be used. See the examples for the default timeSeriesSet XML elements provided with the release of the MEFP.

For generating forecast ensembles of precipitation (FMAP) time series, the input time series must satisfy these requirements by forecast source:

- Must have parameterId of FMAP.
- *RFC QPF*: 6-hour precipitation time series. The start time must be 0 and the end time should be set to a number of days equal to the number of forecast days for the RFC source set as a run file property (see Section 4.4.3). The end must <u>not</u> be overrulable. The qualifierId XML element must be either not defined or "RFC". See the highlighted section in the precipitation example in Section 4.4.1.1.
- *GEFS*: 6-hour location-specific mean precipitation time series provided by the GEFS forecast grids (spatially interpolated through FEWS transformations). The start time must be 0 and the end time must be 16 days. The end must <u>not</u> be overrulable. The qualifierId XML element must be "GEFS".
- *CFSv2*: 6-hour location-specific lagged ensemble output by the CFSv2LaggedEnsembleModelAdapter. The start time must be 0 and the end time must be 270 days. The end must <u>not</u> be overrulable. The qualifierId XML element must be "CFSv2".

NOTE: Only the RFC QPF portion of the exportTimeSeriesActivity *XML element should be modified upon configuration, specifying appropriate RFC-specific time series to use. Do NOT modify the* timeSeriesSet *XML elements defining the other three sources, except to define the* locationSetId *XML element (highlighted in the example) to be for the appropriate forecast group.*

For generating forecast ensembles of temperature (TFMN and TFMX) time series, the input series must satisfy these requirements:

- All sources to be used must include two time series, one with parameterId TFMN and one with TFMX.
- *RFC QTF*: 24-hour precipitation time series generated by the module MEFP_RFC_MAT_6to24 (see the *MEFP Configuration Guide: Forecast Components*). The start time must be 0 and the end time should be set to a number of days equal to the number of forecast days for the RFC source set as a run file property (see Section 4.4.3). The end must <u>not</u> be overrulable. Do NOT modify the RFC timeSeriesSet XML

elements; see NOTE below. The qualifierId XML element must not be either not defined or "RFC".

- *GEFS*: 24-hour location-specific mean TFMN and TFMX time series provided by the GEFS forecast grids (spatially interpolated through FEWS transformations). The start time must be 0 and the end time must be 16 days. The end must not be overrulable. The qualifierId XML element must be "GEFS".
- *CFSv2*: 24-hour location-specific lagged ensemble output by the CFSv2LaggedEnsembleModelAdapter. The start time must be 0 and the end time must be 270 days. The end must not be overrulable. The qualifierId XML element must be "CFSv2".

NOTE: Do not modify any of these time series sets, except to specify the locationSetId *XML element (highlighted in the example, above) to be for the appropriate forecast group.*

NOTE: The RFC QTF time series are specified in the module MEFP_RFC_MAT_6to24 *which is found in the file:*

<*configuration_dir*>/ModuleConfigFiles/hefs/preprocessionMEFP/MEFP_RFC_MAT_6to24.xml

See the MEFP Configuration Guide: Forecast Components.

For either precipitation or temperature, if time series are not provided for a forecast source, then that forecast source cannot be used to generate ensembles; i.e., the corresponding source number of forecast days must either be 0 or not specified.

4.4.5 Model Execution

The following executeActivity XML element should always be included within the executeActivities XML element in the configuration file (see the example in Section 4.2.1.1):

```
<executeActivity>
  <command>
    <className>ohd.hseb.hefs.mefp.adapter.MEFPEnsembleGeneratorModelAdapter</className>
  </command>
  <arguments>
    <argument>%ROOT_DIR%/run_info.xml</argument>
  </arguments>
  <timeOut>300000</timeOut>
</executeActivity>
```

Only the timeOut XML element, set to 5 minutes (300,000 milliseconds) above, should be modified if it is determined that it is not sufficient for the configured adapter run.

4.4.6 Model Output Time Series

This adapter outputs a forecast ensemble of FMAP for each location for which a precipitation forecast ensemble is to be generated, and forecast ensembles of TFMN and TFMX for each location for which a temperature forecast ensemble is to be generated. The following

importActivities XML element should always be used when importing FMAP forecast ensembles, replacing *fgroup* with the appropriate forecast group name:

```
<importActivities>
  <importTimeSeriesActivity>
    <importFile>outputs.xml</importFile>
    <timeSeriesSets>
      <timeSeriesSet>
        <moduleInstanceId>FGroup_MEFP_FMAP_Forecast</moduleInstanceId>
        <valueType>scalar</valueType>
        <parameterId>FMAP</parameterId>
        <locationSetId>Catchments_HEFS_fgroup</locationSetId>
        <timeSeriesType>external forecasting</timeSeriesType>
        <timeStep unit="hour" multiplier="6"/>
        <readWriteMode>add originals</readWriteMode>
        <ensembleId>MEFP</ensembleId>
      </timeSeriesSet>
    </timeSeriesSets>
  </importTimeSeriesActivity>
</importActivities>
```

The following importActivities XML element should always be used when importing TFMN and TFMX forecast ensembles, replacing *fgroup* with the appropriate forecast group name:

```
<importActivities>
  <importTimeSeriesActivity>
    <importFile>outputs.xml</importFile>
    <timeSeriesSets>
      <timeSeriesSet>
        <moduleInstanceId>FGroup_MEFP_TFMN_TFMX_Forecast</moduleInstanceId>
        <valueType>scalar</valueType>
        <parameterId>TFMN</parameterId>
        <locationSetId>Catchments_HEFS_fgroup</locationSetId>
        <timeSeriesType>external forecasting</timeSeriesType>
        <timeStep times="12:00"/>
        <readWriteMode>add originals</readWriteMode>
        <ensembleId>MEFP</ensembleId>
      </timeSeriesSet>
      <timeSeriesSet>
        <moduleInstanceId>FGroup_MEFP_TFMN_TFMX_Forecast</moduleInstanceId>
        <valueType>scalar</valueType>
        <parameterId>TFMX</parameterId>
        <locationSetId>Catchments_HEFS_fgroup</locationSetId>
        <timeSeriesType>external forecasting</timeSeriesType>
        <timeStep times="12:00"/>
        <readWriteMode>add originals</readWriteMode>
        <ensembleId>MEFP</ensembleId>
      </timeSeriesSet>
    </timeSeriesSets>
  </importTimeSeriesActivity>
</importActivities>
```

It is possible for MEFP to generate ensembles for both precipitation and temperature in a single run of the model adapter. In such a case, the importActivities XML element will need to be defined combining the two above appropriately.

NOTE: Do not change the importFile *XML element's value in the* importTimeSeriesActivity.

4.4.7 Model Description

The MEFPEnsembleGeneratorModelAdapter algorithm follows this general process in order to gather time series for locations and data types and call the MEFP algorithm:

1. Parse the run file properties.
2. Parse the provided time series in order to identify for which locations and data types, precipitation or temperature, the MEFP algorithm is to execute and generate a forecast ensemble. Store the found time series by location, data type, and forecast source. Note that TFMN and TFMX time series are stored together for the temperature data type.
3. For each location and data type…
   a. Load the parameters for that location and data type. The parameter file is found here (<parameterDir> is the value of the parameterDir run file property; <locationId> is the time series location id; <data type> is either "precipitation" or "temperature"):

      <parameterDir>/<locationId>.<data type>.mefp.paramters.tgz

   b. Execute the data type appropriate MEFP algorithm. Ensembles are generated using forecast sources for which the number of forecast days is defined and is positive (defined either as the location specific or default number of forecast days; see Section 4.4.3). One ensemble is generated per parameterId associated with a data type; for precipitation data type, the parameterId is "FMAP" and for temperature data type, the parameterIds are "TFMN" and "TFMX".
   c. Record the generated ensemble in a results list.
4. Output the results list to the outputs.xml file so that CHPS can import it as instructed in the module configuration file.

NOTE: The model adapter is fully compatible with FEWS id-mapping capabilities. The parameter file searched for will be determined based on the external locationId and parameterId defined in the id-mapping.

4.4.7.1 Model Errors

The MEFPEnsembleGeneratorModelAdapter will generate an error and stop under the following conditions (the conditions below are due primarily to configuration or data issues; other errors that may occur, such as parameter errors, may not be covered in the list below):

- A run file property value is invalid (wrong name or type).
- A provided input time series is not valid, possibly due to an invalid parameterId XML element or an invalid unit (relative to the parameter).

- When parsing the input time series, if any time series is has a qualifierId XML element that is not recognized.
- The parameter file cannot be loaded for any location and data type.
- If hindcasting and canonical event values cannot be found for the require hindcast date in the parameter file. This should only occur if the MEFPPE did not use data for that hindcast date when estimating parameters.
- The largest number of forecast days for any source for a location is 0, indicating a configuration error.
- The <source>NumberOfForecastDays.<locationId> run file property is not provided for any source and any of the MEFP locations for which to generate an ensemble, and the default value (i.e., the property defined within any .<locationId> suffix) is also not defined. This is equivalent to setting the number of forecast days for all sources to 0 (see the preceding error).
- For a specific location and the temperature data type, the input time series provided are either too few or do not include both required parameterIds TFMN and TFMX.
- For a specific location and the precipitation data type, the input time series provided does not have the parameterId FMAP.
- An incorrect number of time series are provided for a forecast source. One time series is expected for each required parameterId (see above) for RFC and GEFS forecast sources. 16 time series are expected for each required parameterId for CFSv2.
- Any time series to be used is all missing. The one exception is for the CFSv2 lagged ensemble members
- Any time series provided for a location and data type does not appear to be for the correct forecast time. MEFP looks at the first value (missing or not) in the time series. If that value is for a time that is not within one time step of the forecast time (system time or T0) specified in the run file, an error occurs.
- Any time series provided does not include enough data in it to cover the number of forecast days specified for the applicable forecast source and location.
- The specified run file properties initialEnsembleYear or lastEnsembleYear lead to data requirements that cannot be met by the historical record for the MEFP catchment. See Section 4.4.3.
- The behaviorIfEventMissing is set to “errorOut” (its default value) and missing data is found preventing the computation of a canonical event value applied during model execution.

4.4.8 Notes on Configuration

4.4.8.1 Base Ensemble Member Construction using memberIndexingYear

Every application of the MEFP ensemble generation model begins with the computation of a base climatology ensemble from the historical time series available in the parameters. For a given member, the portion of the historical time series that serves as the base member is determined based on the value of the property memberIndexingYear (see Section 4.4.3) and the member index of the member. There are two possibilities:

- If the memberIndexingYear is defined as a standard calendar year, then the portion of the historical time series used for a member starts from the forecast time (T0) month and day within the calendar year exactly matching the member index. For example, for member index 1950, the first day of the base ensemble member is the T0 month and day and year 1950.

- If the memberIndexingYear is defined as a user specified day and month (e.g., "--10-01", which is the standard hydrologic water year), then the portion of the historical time series used for a member depends upon the T0 month and day:
    - If the T0 month and day precede the user specified day and month, then the portion is pulled from the year matching the member index. For example, for a standard hydrologic model year, member index 1950, and a T0 before October 1, the first day of the base ensemble member is the T0 month and day and year 1950.
    - If the T0 month and day are after or on the user specified day and month, then the portion is pulled from the year matching the member index – 1. For example, , for a standard hydrologic model year, member index 1950, and a T0 after or on October 1, the first day of the base ensemble member is the T0 month and day and year 1949 (member index – 1).

By default, if the memberIndexingYear property is not defined, then a standard hydrologic water year is used; see (see Section 4.4.3). If a calendar year is to be used, then set the property to "calendarYear". It is recommended that either the hydrologic water year or calendar year be used.

NOTE: Most RFCs use a standard hydrologic water year (10/1 – 9/30) when sampling for the generation of a standard ESP climatology-based ensemble.

4.4.8.2 Appending Raw Climatology

As mentioned in Sections 2 and 4.4.3, there is an option available in the MEFPEnsembleGeneratorModelAdapter for appending raw climatology to the end of an MEFP generated ensemble. However, this option will be removed in the near future for two reasons. First, it does not generate true raw climatology for temperature (FMAT) ensembles. Rather, the ensembles use TMIN/TMAX historical data (computed during parameter estimation) and a diurnal pattern which executes after MEFPEnsembleGeneratorModelAdapter is applied within the MEFP forecast workflow to output 6-hour FMAT data (module MEFP_FMAT_Forecast shown in the diagram in Section 4.1). Thus the output 6-hour FMAT from the workflow will not equal the 6-hour MAT used as input to the parameter estimation process. Second, raw climatology can be added to the end of an MEFP generated ensemble using standard CHPS transformation (sample-historical and merge; see below). Removing raw climatology removed unnecessary duplicated functionality.

For raw climatology, therefore, it is recommended that standard CHPS transformations be used. Specifically, transformation modules named SampleESP* are typically used to create raw climatology forcing precipitation and temperature ensembles for use in generating standard ESP

climatology-based streamflow forecasts. To append raw-climatology to the end of the MEFP generated ensemble, use a FEWS merge transformation to append the output from the appropriate SampleESP* module to the output from the MEFPEnsembleGeneratorModelAdapter, specifically the output from modules <*fgroup*>_MEFP_FMAP_Forecast and <*fgroup*>_MEFP_TFMN_TFMX_Forecast (see the diagram in Section 4.1).