

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/339448304>

# The Forecasting Toolbox of the MATLAB-Toolbox SciXMiner – Short manual

Technical Report · February 2020

DOI: 10.13140/RG.2.2.31678.84801/1

CITATIONS

0

READS

154

2 authors:



**Jorge Ángel González Ordiano**  
Colorado State University

20 PUBLICATIONS 126 CITATIONS

[SEE PROFILE](#)



**Ralf Mikut**  
Karlsruhe Institute of Technology

383 PUBLICATIONS 3,007 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Energy System 2050 [View project](#)



I-CARE [View project](#)

# The Forecasting Toolbox of the MATLAB-Toolbox

## SciXMiner

### Short manual

Jorge Ángel González Ordiano, Ralf Mikut

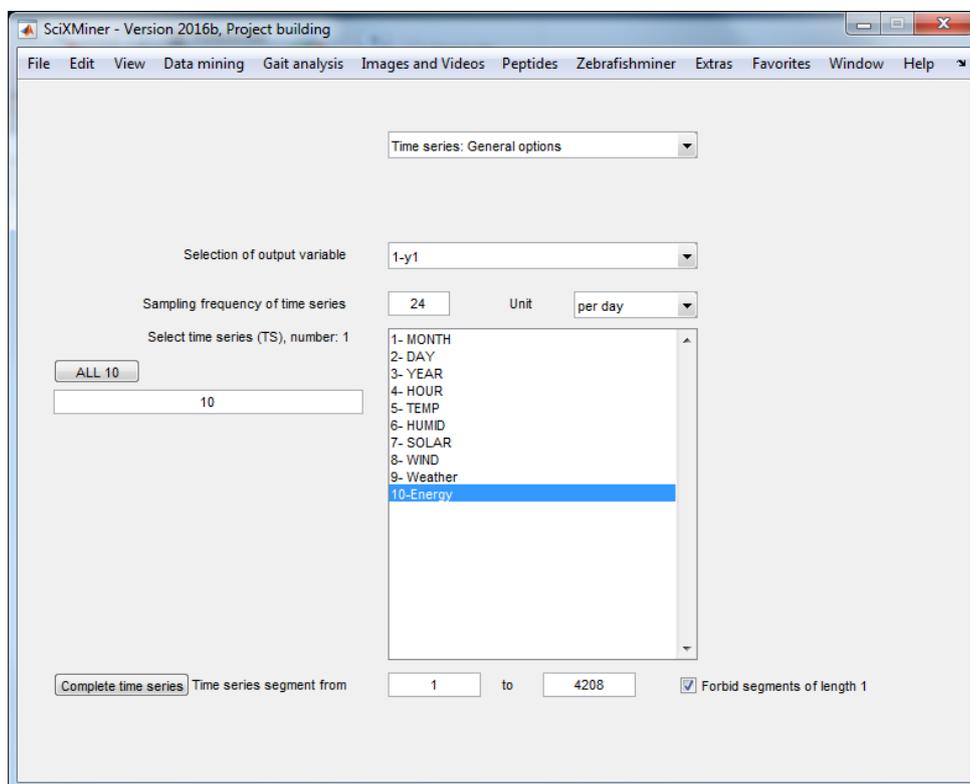
Karlsruhe Institute of Technology (KIT), Institute for Automation and Applied Informatics

P.O. Box 3640, D-76021 Karlsruhe, Germany

Phone: ++49/721/608-25731, Fax: ++49/721/608-25702

Email: ralf.mikut@kit.edu

Beta version: Version 2020a (24.02.2020)





# Contents

<b>Contents</b>	<b>iii</b>
<b>1 Motivation</b>	<b>1</b>
<b>2 Installation</b>	<b>2</b>
<b>3 General remarks</b>	<b>3</b>
3.1 Getting started . . . . .	3
3.2 Variables . . . . .	3
3.2.1 <code>regr_single</code> . . . . .	3
3.2.2 <code>cdf_bestfit</code> . . . . .	3
3.2.3 <code>scenarioForecast</code> . . . . .	4
3.3 Demo Projects . . . . .	4
3.4 Use cases . . . . .	4
3.5 Versions . . . . .	5
<b>4 Menu items and Control elements</b>	<b>6</b>
4.1 Menu items 'Forecasting' . . . . .	6
4.1.1 Forecasting Model (Regression) . . . . .	6
4.1.2 Interval Forecast . . . . .	6
4.1.3 Scenario Forecast . . . . .	6
4.1.4 Parametric Probabilistic Forecast . . . . .	7
4.1.5 Hierarchical Probabilistic Forecast . . . . .	7
4.1.6 Validate Forecasting Model . . . . .	7
4.1.7 Help . . . . .	7
4.2 Control elements for 'Forecasting' . . . . .	7
<b>5 Plugins</b>	<b>28</b>
<b>Bibliography</b>	<b>30</b>

# 1 Motivation

This forecasting toolbox offers different time series forecasting methods with a focus on special needs of energy time series [7]. A short description of the toolbox can be found in [5], the quantile forecasting methodology is explained in [5, 6]. This is a preliminary beta version that will be improved in the future. The present contribution is supported by the Helmholtz Association under the Joint Initiative "Energy System 2050 — A Contribution of the Research Field Energy".

SciXMiner [9] is open source software. The download page is

<http://sourceforge.net/projects/SciXMiner/>.

It is licensed under the conditions of the GNU General Public License (GNU-GPL) of The Free Software Foundation (see <http://www.fsf.org/>).

This manual is organized as follows: Chapter 2 explains the installation procedure. Chapter 3 outlines the implemented functionality followed by some recommendations for working with the toolbox SciXMiner. Detailed information for the use of menu items and control elements (Chapter 4) follow.

## 2 Installation

The zipped toolbox has to be extracted in the directory `application_specials` of the SciXMiner directory. The subdirectory structure must be preserved, leading to a new subdirectory `forecasting`. The extension packages can be switched on and off by *Extras - Choose application-specific extension packages....*

The toolbox requires at least SciXMiner Version 2020a (24.02.2020) .

## 3 General remarks

The forecasting toolbox offers a plethora of time series forecasting methods. These methods include point, probabilistic, and scenario forecasting approaches. Additionally, options specifically designed to be used during the forecasting of energy time series can also be found within the toolbox.

### 3.1 Getting started

The forecasting toolbox can be used on any SciXMiner project containing time series. To understand how to create SciXMiner projects, please refer to the SciXMiner related documentation.

### 3.2 Variables

The forecasting toolbox uses the SciXMiner variable `d_orgs`. The dimensions of `d_orgs` represent the following: the first, represents the dimension of a given time series (i.e. if a time series is univariate or multivariate), the second, is the number measurements that form the time series, and the last, represents the number of time series available in the project. It is important to mention that the toolbox functionality depends on the existence of `d_orgs`.

Furthermore, the new toolbox not only modifies some existing variables, but also defines some new ones. More in detail information about this is given in the next sections.

#### 3.2.1 `regr_single`

The variable `regr_single` is the variable used in SciXMiner to save the information of an already trained regression model. Therefore, the forecasting toolbox uses this variable to save the information of all regression based forecasting models. However, since the forecasting toolbox modifies the structure of some the information saved within the variable it is recommended that the users **DO NOT** save and load regression based forecasting models as they would normally do with regression models. Instead the load and save functionality available within the new toolbox should be used. When saving a regression based forecasting model, the toolbox creates a file with a `.fmodel` extension that contains the `regr_single` variable.

#### 3.2.2 `cdf_bestfit`

The variable `cdf_bestfit` is a structure array containing the parametric probabilistic forecasts obtained by the toolbox's method. The variable contains the following fields:

- `sample`: Contains the type and parameters of the parametric distribution functions estimated at each timestep. Additional information corresponding to each one of the estimated distributions is also contained in this field.
- `rank`: This field contains a rank of the distributions tested. The distributions are ranked from the best fitting one to the worse. Additionally, the difference between the metric of each distribution and the best fitting one are also contained in this field.

### 3.2.3 `scenarioForecast`

The variable `scenarioForecast` is a structure array containing the scenario forecasts obtained by the toolbox's method. The variable contains the following fields:

- `parameters`: This field contains the parameters used to create the scenario forecasts.
- `values`: The field contains all of the scenario forecasts created.

## 3.3 Demo Projects

Two demos projects can be found in the `projects` folder of the subfolder containing this extension. The demo project called `hierarchicalForecastExample.prjz` is used to test the implemented hierarchical probabilistic forecasting method, while the demo project named `Settlement_experiments_M45_S45.prjz` is used to test all other approaches. Testing the forecasting toolbox consists in running the `test_forecasting_toolbox.batch` batch file, which uses specific macros for testing each one of the available methods. Note that the macros are named using the following convention, i.e. `test_{forecasting model being tested}`.

## 3.4 Use cases

Selected SciXMiner applications are listed in Table 3.1.

Application	References
Campus load forecasting	[5]
Optimization for household batteries	[2]
Optimization for batteries of electric vehicles	[4]
Optimization for batteries of an industrial campus	[3, 1]
Solar power forecasting without weather data	[8, 10]

Table 3.1: Selected SciXMiner applications

## **3.5 Versions**

- Beta version Version 2020a (24.02.2020)

## 4 Menu items and Control elements

### 4.1 Menu items 'Forecasting'

#### 4.1.1 Forecasting Model (Regression)

Menu element containing options related forecasting models based on regressions.

- **Create:**  
Click to create forecasting models based on regressions.
- **Apply:**  
Click to apply forecasting models based on regressions.
- **Create & Apply:**  
Click to create and apply forecasting models based on regressions.
- **Save:**  
Click to save forecasting models based on regressions.
- **Load:**  
Click to load forecasting models based on regressions.
- **View:**  
Click to view the forecasting models based on **polynomial based regressions**.

#### 4.1.2 Interval Forecast

Menu element containing options related to interval forecasts.

- **Plot Interval Forecast:**  
Click to plot interval forecasts.

#### 4.1.3 Scenario Forecast

Menu element containing options related to scenario forecasts.

- **Create Scenario Forecasts with QR:**  
Click to create scenario forecasts based on quantile regressions.
- **View Scenario Forecasts:**  
Click to view the scenario forecasts created.

#### 4.1.4 Parametric Probabilistic Forecast

Click to obtain parametric probabilistic forecasts.

#### 4.1.5 Hierarchical Probabilistic Forecast

Menu element containing options related to hierarchical probabilistic forecasts.

- **Create:**  
Menu element containing options for creating hierarchical probabilistic forecasts.
- **Apply:**  
Menu element containing options for applying hierarchical probabilistic forecasts.
  - **Apply QR Hierarchical Probabilistic Forecasts:**  
Click to apply hierarchical probabilistic forecasts based on quantile regressions.

#### 4.1.6 Validate Forecasting Model

Validates a forecasting model with a selected proportion of training and test data. The proportion is selected by *Control element: Forecasting - Percentage of training data for validation*. Here, a model is created on the training data (begin of the time series) and validated on the test data (end of the time series). The results are shown in the MATLAB workspace and plotted into files if *Control element: Forecasting - Plot validation results* is selected.

#### 4.1.7 Help

Shows the help file for the Forecasting toolbox.

### 4.2 Control elements for 'Forecasting'

- **Forecast type:**  
Use to select the forecasting model to be created
- **Energy forecasting:**  
Check in case specific energy forecasting options are to be used.
- **Regression technique:**  
Use to select the data mining technique to be used (e.g., Polynomial or ANN)
- **Desired output time series:**  
Use to select the time series for which a forecast is to be obtained.
- **Forecast horizon:**  
Use the field to write the models forecast horizon in timesteps.

The screenshot shows a control panel for forecasting. At the top, a dropdown menu is set to 'Forecasting'. Below it, the 'Forecast type' is 'Point Forecast', 'Regression technique' is 'Polynomial', and 'Desired output time series' is '1- MONTH'. The 'Forecast horizon' is set to 24, and 'Input time series' is 1. The 'Lags' field contains '{0:1:96}'. There are checkboxes for 'Energy forecasting' (checked) and 'Feature selection' (unchecked). The 'Energy forecasting' dropdown is set to 'Photovoltaic Power', with a 'Degree' of 2. There are checkboxes for 'P > 0' (unchecked) and 'Set a0 to zero' (unchecked). The 'Features' field is set to 4. The 'Eliminate night values' dropdown is set to '-'. At the bottom, 'Percentage of training data for validation' is 50, and 'Plot validation results' is checked.

Figure 4.1: Control elements for Forecasting - Point Forecast- Polynomial

- **Input time series:**

Use the field to write the indexes that correspond—within the project—to the time series that will be used as input.

- **Feature selection:**

Check if a Wrapper feature selection is to be applied during the training.

- **Lags:**

Use the field to write the lags in timesteps of the input time series' past values that the models are going to use as features. For example, if values with lags ranging from 0 to 24 are used the following has to be written in the edit field {0:1:24}. On the other hand, if values with specific lags are required this can be written for instance, as {0,12,24} or {0}. The former specifies that the values used are those with lags of 0, 12, and 24 and the later states that only values with the specific lag of 0 are used. Additionally, this edit field also allows for independent definition of features for each input time series used. For example writing {0:1:24};{0,12,24};0 specifies different lags to be taken from the first, second and third time series. It needs to be mentioned, that if the independent definition is used, the lags of each input time series have to be explicitly defined.

- **Energy forecasting:**

Use to select the type of energy time series being forecast (e.g., photovoltaic power, load, wind power).

- **Degree:**  
Use the field to write the maximal degree that the polynomial models are allowed to have.
- **P > 0:**  
Check if a constraint assuring only positive forecast values is to be applied.
- **Features:**  
Use the field to write the number of features that will be selected.
- **Set a0 to zero:**  
Check if the offset of the polynomial models is to be set equal to zero.
- **Eliminate night values:**  
Use to select the approach for identifying night values. The selected approach is used to remove night values from the training set and to set them automatically to zero during the models application.
- **Percentage of training data for validation:**  
Defines the percentage of training data for a model validation by *Forecasting - Validate Forecasting Model*.
- **Plot validation results:**  
Switches the plotting of results of model validation by *Forecasting - Validate Forecasting Model*.

The screenshot displays a control panel for forecasting. At the top, a dropdown menu is set to 'Forecasting'. Below it, several configuration options are listed:

- Forecast type:** Point Forecast (dropdown), with a checked checkbox for 'Energy forecasting'.
- Regression technique:** Artificial Neural Network (MLP) (dropdown).
- Desired output time series:** 1- MONTH (dropdown).
- Forecast horizon:** 24 (text input).
- Input time series:** 1 (text input), with an unchecked checkbox for 'Feature selection'.
- Lags:** {0:1:96} (text input).
- Energy forecasting:** Photovoltaic Power (dropdown), with an unchecked checkbox for 'P > 0'.
- Hidden layers:** 1 (text input).
- Neurons:** 6 (text input).
- Features:** 4 (text input).
- Eliminate night values:** - (dropdown).
- Percentage of training data for validation:** 50 (text input), with a checked checkbox for 'Plot validation results'.

Figure 4.2: Control elements for Forecasting - Point Forecast- Artificial Neural Network (MLP)

- **Forecast type:**  
Use to select the forecasting model to be created
- **Energy forecasting:**  
Check in case specific energy forecasting options are to be used.
- **Regression technique:**  
Use to select the data mining technique to be used (e.g., Polynomial or ANN)
- **Desired output time series:**  
Use to select the time series for which a forecast is to be obtained.
- **Forecast horizon:**  
Use the field to write the models forecast horizon in timesteps.
- **Input time series:**  
Use the field to write the indexes that correspond—within the project—to the time series that will be used as input.
- **Feature selection:**  
Check if a Wrapper feature selection is to be applied during the training.
- **Lags:**  
Use the field to write the lags in timesteps of the input time series' past values that the models

are going to use as features. For example, if values with lags ranging from 0 to 24 are used the following has to be written in the edit field {0:1:24}. On the other hand, if values with specific lags are required this can be written for instance, as {0,12,24} or {0}. The former specifies that the values used are those with lags of 0, 12, and 24 and the later states that only values with the specific lag of 0 are used. Additionally, this edit field also allows for independent definition of features for each input time series used. For example writing {0:1:24};{0,12,24};0 specifies different lags to be taken from the first, second and third time series. It needs to be mentioned, that if the independent definition is used, the lags of each input time series have to be explicitly defined.

- **Energy forecasting:**  
Use to select the type of energy time series being forecast (e.g., photovoltaic power, load, wind power).
- **Hidden layers:**  
Use the field to write the number of hidden layers of the neural networks to be trained.
- **Neurons:**  
Use the field to write the number of neurons within the hidden layers of the neural networks to be trained.
- **P > 0:**  
Check if a constraint assuring only positive forecast values is to be applied.
- **Features:**  
Use the field to write the number of features that will be selected.
- **Eliminate night values:**  
Use to select the approach for identifying night values. The selected approach is used to remove night values from the training set and to set them automatically to zero during the models application.
- **Percentage of training data for validation:**  
Defines the percentage of training data for a model validation by *Forecasting - Validate Forecasting Model*.
- **Plot validation results:**  
Switches the plotting of results of model validation by *Forecasting - Validate Forecasting Model*.

The screenshot shows a control panel for forecasting. At the top, there is a dropdown menu set to 'Forecasting'. Below it, the 'Forecast type' is 'Polynomial Quantile Regression (Q...)' with a checked 'Energy forecasting' checkbox. The 'Desired output time series' is '1- MONTH'. The 'Forecast horizon' is '24'. The 'Input time series' is '1' with an unchecked 'Feature selection' checkbox. The 'Lags' field contains '{0:1:96}'. Under 'Energy forecasting', the dropdown is 'Photovoltaic Power', 'Degree' is '2', and there is an unchecked 'P > 0' checkbox. The 'Features' field is '4' with an unchecked 'Set a0 to zero' checkbox. The 'Eliminate night values' dropdown is '-'. The 'Quantiles' field contains '0.01 0.02 0.03 0.04' with an unchecked 'Quantiles constraint' checkbox and a 'Reg:' field. At the bottom, 'Percentage of training data for validation' is '50' with a checked 'Plot validation results' checkbox.

Figure 4.3: Control elements for Forecasting - Polynomial Quantile Regression (QR) w. Pinball Loss-Energy forecasting

- **Forecast type:**  
Use to select the forecasting model to be created
- **Energy forecasting:**  
Check in case specific energy forecasting options are to be used.
- **Desired output time series:**  
Use to select the time series for which a forecast is to be obtained.
- **Forecast horizon:**  
Use the field to write the models forecast horizon in timesteps.
- **Input time series:**  
Use the field to write the indexes that correspond—within the project—to the time series that will be used as input.
- **Feature selection:**  
Check if a Wrapper feature selection is to be applied during the training.
- **Lags:**  
Use the field to write the lags in timesteps of the input time series' past values that the models are going to use as features. For example, if values with lags ranging from 0 to 24 are used the following has to be written in the edit field {0:1:24}. On the other hand, if values with specific

lags are required this can be written for instance, as {0,12,24} or {0}. The former specifies that the values used are those with lags of 0, 12, and 24 and the later states that only values with the specific lag of 0 are used. Additionally, this edit field also allows for independent definition of features for each input time series used. For example writing {0:1:24};{0,12,24};0 specifies different lags to be taken from the first, second and third time series. It needs to be mentioned, that if the independent definition is used, the lags of each input time series have to be explicitly defined.

- **Energy forecasting:**  
Use to select the type of energy time series being forecast (e.g., photovoltaic power, load, wind power).
- **Degree:**  
Use the field to write the maximal degree that the polynomial models are allowed to have.
- **P > 0:**  
Check if a constraint assuring only positive forecast values is to be applied.
- **Features:**  
Use the field to write the number of features that will be selected.
- **Set a0 to zero:**  
Check if the offset of the polynomial models is to be set equal to zero.
- **Eliminate night values:**  
Use to select the approach for identifying night values. The selected approach is used to remove night values from the training set and to set them automatically to zero during the models application.
- **Quantiles:**  
Use the field to write the probabilities that correspond to the quantile regressions to be estimated (written values have to be between 0 and 1).
- **Quantiles constraint:**  
Check if constraints for avoiding quantile crossing are to be applied.
- **Reg:**  
Use to write the regularization value used during the application of the constraints for avoiding quantile crossing. If the field is left empty the value is set to  $10^{-5}$ .
- **Percentage of training data for validation:**  
Defines the percentage of training data for a model validation by *Forecasting - Validate Forecasting Model*.
- **Plot validation results:**  
Switches the plotting of results of model validation by *Forecasting - Validate Forecasting Model*.

The screenshot shows a configuration window for forecasting. At the top, a dropdown menu is set to 'Forecasting'. Below it, several parameters are defined:

- Forecast type:** 'Quantile Regressions (QR) with N...' (dropdown), with a checked checkbox for 'Energy forecasting'.
- Regression technique:** 'Polynomial' (dropdown).
- Desired output time series:** '1- MONTH' (dropdown).
- Forecast horizon:** '24' (text input).
- Input time series:** '1' (text input), with an unchecked checkbox for 'Feature selection'.
- Lags:** '{0:1:96}' (text input), with an unchecked checkbox for 'Quantile features'.
- Energy forecasting:** 'Photovoltaic Power' (dropdown), with an unchecked checkbox for 'P > 0'.
- Eliminate night values:** '-' (dropdown).
- Degree:** '2' (text input).
- Features:** '4' (text input), with an unchecked checkbox for 'Set a0 to zero'.
- k:** '50' (text input).
- Distance:** 'Euclidean' (dropdown).
- Weights:** 'None' (dropdown).
- Quantiles:** '0.01 0.02 0.03 0.04' (text input), with an unchecked checkbox for 'Quantiles constraint'.
- Reg:** (empty text input).
- Percentage of training data for validation:** '50' (text input), with a checked checkbox for 'Plot validation results'.

Figure 4.4: Control elements for Forecasting - Quantile Regressions (QR) with NNQF- Polynomial

- Forecast type:**  
 Use to select the forecasting model to be created
- Energy forecasting:**  
 Check in case specific energy forecasting options are to be used.
- Regression technique:**  
 Use to select the data mining technique to be used (e.g., Polynomial or ANN)
- Desired output time series:**  
 Use to select the time series for which a forecast is to be obtained.
- Forecast horizon:**  
 Use the field to write the models forecast horizon in timesteps.
- Input time series:**  
 Use the field to write the indexes that correspond—within the project—to the time series that will be used as input.
- Feature selection:**  
 Check if a Wrapper feature selection is to be applied during the training.
- Lags:**  
 Use the field to write the lags in timesteps of the input time series' past values that the models

are going to use as features. For example, if values with lags ranging from 0 to 24 are used the following has to be written in the edit field {0:1:24}. On the other hand, if values with specific lags are required this can be written for instance, as {0,12,24} or {0}. The former specifies that the values used are those with lags of 0, 12, and 24 and the later states that only values with the specific lag of 0 are used. Additionally, this edit field also allows for independent definition of features for each input time series used. For example writing {0:1:24};{0,12,24};0 specifies different lags to be taken from the first, second and third time series. It needs to be mentioned, that if the independent definition is used, the lags of each input time series have to be explicitly defined.

- **Quantile features:**  
Check if an independent feature selection for each quantile regression is to be applied.
- **Energy forecasting:**  
Use to select the type of energy time series being forecast (e.g., photovoltaic power, load, wind power).
- **Degree:**  
Use the field to write the maximal degree that the polynomial models are allowed to have.
- **P > 0:**  
Check if a constraint assuring only positive forecast values is to be applied.
- **Features:**  
Use the field to write the number of features that will be selected.
- **Set a0 to zero:**  
Check if the offset of the polynomial models is to be set equal to zero.
- **Eliminate night values:**  
Use to select the approach for identifying night values. The selected approach is used to remove night values from the training set and to set them automatically to zero during the models application.
- **k:**  
Use the field to write number of nearest neighbors used by the nearest neighbors quantile filter.
- **Distance:**  
Defines the distance function used by the nearest neighbors quantile filter.
- **Weights:**  
Defines the weights applied to the features during the distance calculation of the nearest neighbors quantile filter.
- **Quantiles:**  
Use the field to write the probabilities that correspond to the quantile regressions to be estimated (written values have to be between 0 and 1).
- **Quantiles constraint:**  
Check if constraints for avoiding quantile crossing are to be applied.
- **Reg:**  
Use to write the regularization value used during the application of the constraints for avoiding quantile crossing. If the field is left empty the value is set to  $10^{-5}$ .

- **Percentage of training data for validation:**  
Defines the percentage of training data for a model validation by *Forecasting - Validate Forecasting Model*.
- **Plot validation results:**  
Switches the plotting of results of model validation by *Forecasting - Validate Forecasting Model*.

The screenshot displays a web-based control interface for forecasting. At the top, a dropdown menu is set to 'Forecasting'. Below this, several configuration options are listed:

- Forecast type:** 'Quantile Regressions (QR) with N...' (dropdown), with a checked box for 'Energy forecasting'.
- Regression technique:** 'Artificial Neural Network (MLP)' (dropdown).
- Desired output time series:** '1- MONTH' (dropdown).
- Forecast horizon:** '24' (text input).
- Input time series:** '1' (text input), with an unchecked box for 'Feature selection'.
- Lags:** '{0:1:96}' (text input), with an unchecked box for 'Quantile features'.
- Energy forecasting:** 'Photovoltaic Power' (dropdown), with an unchecked box for 'P > 0'.
- Eliminate night values:** '-' (dropdown).
- Hidden layers:** '1' (text input), **Neurons:** '6' (text input).
- Features:** '4' (text input), **k:** '50' (text input).
- Distance:** 'Euclidean' (dropdown), **Weights:** 'None' (dropdown).
- Quantiles:** '0.01 0.02 0.03 0.04' (text input), with an unchecked box for 'Quantiles constraint' and a 'Reg:' text input.
- Percentage of training data for validator:** '50' (text input), with a checked box for 'Plot validation results'.

Figure 4.5: Control elements for Forecasting - Quantile Regressions (QR) with NNQF- Artificial Neural Network (MLP)

- **Forecast type:**  
Use to select the forecasting model to be created
- **Energy forecasting:**  
Check in case specific energy forecasting options are to be used.
- **Regression technique:**  
Use to select the data mining technique to be used (e.g., Polynomial or ANN)
- **Desired output time series:**  
Use to select the time series for which a forecast is to be obtained.
- **Forecast horizon:**  
Use the field to write the models forecast horizon in timesteps.
- **Input time series:**  
Use the field to write the indexes that correspond—within the project—to the time series that will be used as input.
- **Feature selection:**  
Check if a Wrapper feature selection is to be applied during the training.
- **Lags:**  
Use the field to write the lags in timesteps of the input time series' past values that the models

are going to use as features. For example, if values with lags ranging from 0 to 24 are used the following has to be written in the edit field {0:1:24}. On the other hand, if values with specific lags are required this can be written for instance, as {0,12,24} or {0}. The former specifies that the values used are those with lags of 0, 12, and 24 and the later states that only values with the specific lag of 0 are used. Additionally, this edit field also allows for independent definition of features for each input time series used. For example writing {0:1:24};{0,12,24};0 specifies different lags to be taken from the first, second and third time series. It needs to be mentioned, that if the independent definition is used, the lags of each input time series have to be explicitly defined.

- **Quantile features:**

Check if an independent feature selection for each quantile regression is to be applied.

- **Energy forecasting:**

Use to select the type of energy time series being forecast (e.g., photovoltaic power, load, wind power).

- **Hidden layers:**

Use the field to write the number of hidden layers of the neural networks to be trained.

- **Neurons:**

Use the field to write the number of neurons within the hidden layers of the neural networks to be trained.

- **P > 0:**

Check if a constraint assuring only positive forecast values is to be applied.

- **Features:**

Use the field to write the number of features that will be selected.

- **Eliminate night values:**

Use to select the approach for identifying night values. The selected approach is used to remove night values from the training set and to set them automatically to zero during the models application.

- **k:**

Use the field to write number of nearest neighbors used by the nearest neighbors quantile filter.

- **Distance:**

Defines the distance function used by the nearest neighbors quantile filter.

- **Weights:**

Defines the weights applied to the features during the distance calculation of the nearest neighbors quantile filter.

- **Quantiles:**

Use the field to write the probabilities that correspond to the quantile regressions to be estimated (written values have to be between 0 and 1).

- **Quantiles constraint:**

Check if constraints for avoiding quantile crossing are to be applied.

- **Reg:**  
Use to write the regularization value used during the application of the constraints for avoiding quantile crossing. If the field is left empty the value is set to  $10^{-5}$ .
- **Percentage of training data for validation:**  
Defines the percentage of training data for a model validation by *Forecasting - Validate Forecasting Model*.
- **Plot validation results:**  
Switches the plotting of results of model validation by *Forecasting - Validate Forecasting Model*.

The screenshot shows a control panel for forecasting. At the top, there is a dropdown menu labeled 'Forecasting' with 'Forecasting' selected. Below it, the 'Forecast type' is set to 'Interval Forecasts'. The 'Desired output time series' is '1- MONTH'. The 'Lower bound time series' is '1', the 'Upper bound time series' is '1', and the 'Corresponding interval probabilities' is '0.8'. The 'Starting timestep' is '1' and the 'Ending timestep' is '100'. At the bottom, the 'Percentage of training data for validation' is '50' and the 'Plot validation results' checkbox is checked.

Figure 4.6: Control elements for Forecasting - Interval Forecasts

- **Forecast type:**  
Use to select the forecasting model to be created
- **Desired output time series:**  
Use to select the time series for which a forecast is to be obtained.
- **Lower bound time series:**  
Use the field to write the indexes of the time series that will form the lower bounds of the interval forecasts.
- **Upper bound time series:**  
Use the field to write the indexes of the time series that will form the upper bounds of the interval forecasts.
- **Corresponding interval probabilities:**  
Use the field to write the probabilities of the interval forecasts being created.
- **Starting timestep:**  
Use the field to write the timestep in which the interval forecasts begin.
- **Ending timestep:**  
Use the field to write the timestep in which the interval forecasts end.

- **Percentage of training data for validation:**  
Defines the percentage of training data for a model validation by *Forecasting - Validate Forecasting Model*.
- **Plot validation results:**  
Switches the plotting of results of model validation by *Forecasting - Validate Forecasting Model*.

Forecasting

Forecast type: Parametric Probabilistic Forecast ...

TSs containing the quantile estimates: 1 2

Max CDF argument:

Min CDF argument:

Corresponding CDF values: 0.01 0.02

Parametric CDF to be fitted: Find Best Fit

Technique to find best fit: Least Squares

TS representing an indicator function:

Apply quantiles of the parametric CDF fitted

All quantiles <=:

All quantiles >=:

Compare to optimal fit

Percentage of training data for validation: 50  Plot validation results

Figure 4.7: Control elements for Forecasting - Parametric Probabilistic Forecast with QRs

- **Forecast type:**  
Use to select the forecasting model to be created
- **TSs containing the quantile estimates:**  
Use to write the indexes that correspond to the time series containing the quantile estimates that will be used to find the best fitting parametric distributions
- **Max CDF argument:**  
Notice that the method in which this option is available requires estimates of non-parametric CDFs!  
Use this field to define the maximal argument that the non-parametric CDFs are allowed to have. If left empty this value is calculated independently in a timestep-wise basis (i.e. the recommended setting).
- **Min CDF argument:**  
Notice that the method in which this option is available requires estimates of non-parametric CDFs!  
Use this field to define the minimal argument that the non-parametric CDFs are allowed to have. If left empty this value is calculated independently in a timestep-wise basis (i.e. the recommended setting).
- **Corresponding CDF values:**  
Use the field to write the corresponding probabilities of the quantile estimates that will be used to find the best fitting parametric distributions.

- **Parametric CDF to be fitted:**

The method containing this option tests for every timestep various parametric distributions and selects the best fitting one (i.e. the default setting *Find Best Fit*). However, if the user requires the same parametric distribution for every timestep, this option can be used. The option allows the selection of the parametric distribution, whose parameters are to be estimated in a timestep-wise basis using a method that is similar to the method of moments.
- **Technique to find best fit:**

Use to select the method the parametric CDFs that best fit the quantile estimates in a timestep-wise basis. The default option is *Least Squares*, which bases its selection on the mean squared error.
- **TS representing an indicator function:**

Use the field to write the index of a time series indicating the timesteps in which the estimation of the uncertainty is not required (e.g., night values of a photovoltaic power time series). The indicator time series should only contain values between zero and one and it should be one for the timesteps in which no uncertainty description is needed. The field can be left empty if an uncertainty description is to be obtained for all timesteps.
- **Apply quantiles of the parametric CDF fitted:**

Check if new time series containing the quantiles of the parametric CDFs estimated at each timestep are to be obtained and saved.
- **All quantiles  $\leq$ :**

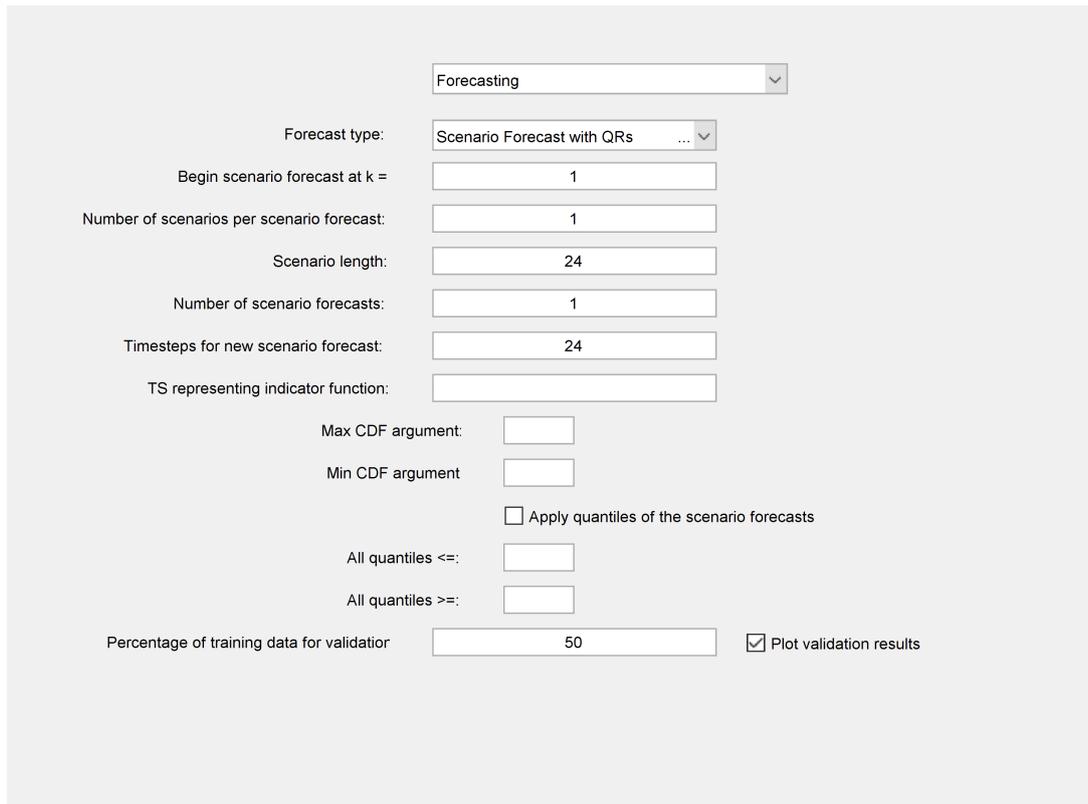
Use the field to write the maximal allowed value for all quantile estimates. In other words, all estimates that are greater than the given threshold are replaced by the threshold. If the field is left empty no constraint is applied.
- **All quantiles  $\geq$ :**

Use the field to write the minimal allowed value for all quantile estimates. In other words, all estimates that are lower than the given threshold are replaced by the threshold. If the field is left empty no constraint is applied.
- **Compare to optimal fit:**

Check if a comparison between the estimated parametric CDFs and those that could be obtained by fitting the CDFs to the quantiles used should be done.
- **Percentage of training data for validation:**

Defines the percentage of training data for a model validation by *Forecasting - Validate Forecasting Model*.
- **Plot validation results:**

Switches the plotting of results of model validation by *Forecasting - Validate Forecasting Model*.



Forecasting

Forecast type: Scenario Forecast with QRs

Begin scenario forecast at k = 1

Number of scenarios per scenario forecast: 1

Scenario length: 24

Number of scenario forecasts: 1

Timesteps for new scenario forecast: 24

TS representing indicator function:

Max CDF argument:

Min CDF argument:

Apply quantiles of the scenario forecasts

All quantiles <=:

All quantiles >=:

Percentage of training data for validation: 50  Plot validation results

Figure 4.8: Control elements for Forecasting - Scenario Forecast with QRs

- **Forecast type:**  
Use to select the forecasting model to be created
- **Begin scenario forecast at k =:**  
Use the field to write the timestep in which the first scenario forecast should begin.
- **Number of scenarios per scenario forecast:**  
Use the field to write the number of scenarios that will form the scenario forecasts (A scenario is a possible realization of a time series' future; A scenario forecast is a collection of possible scenarios)
- **Scenario length:**  
Use the field to write the length (in timesteps) of the scenarios that will be created.
- **Number of scenario forecasts:**  
Use the field to write the number of scenario forecasts that will be calculated.
- **Timesteps for new scenario forecast:**  
Use the field to the number of timesteps between each of the calculated scenario forecasts. In other words, after how many timesteps a new scenario forecast is to be obtained.
- **TS representing indicator function:**  
Use the field to write the index of a time series indicating the timesteps in which the estimation of the uncertainty is not required (e.g., night values of a photovoltaic power time series). The

indicator time series should only contain values between zero and one and it should be one for the timesteps in which no uncertainty description is needed. The field can be left empty if an uncertainty description is to be obtained for all timesteps.

- **Max CDF argument:**  
Notice that the method in which this option is available requires estimates of non-parametric CDFs! Use this field to define the maximal argument that the non-parametric CDFs are allowed to have. If left empty this value is calculated independently in a timestep-wise basis (i.e. the recommended setting).
- **Min CDF argument:**  
Notice that the method in which this option is available requires estimates of non-parametric CDFs! Use this field to define the minimal argument that the non-parametric CDFs are allowed to have. If left empty this value is calculated independently in a timestep-wise basis (i.e. the recommended setting).
- **Apply quantiles of the scenario forecasts:**  
Check if new time series containing the empirical quantiles of the scenario values at each timestep are to be obtained and saved.
- **All quantiles  $\leq$ :**  
Use the field to write the maximal allowed value for all quantile estimates. In other words, all estimates that are greater than the given threshold are replaced by the threshold. If the field is left empty no constraint is applied.
- **All quantiles  $\geq$ :**  
Use the field to write the minimal allowed value for all quantile estimates. In other words, all estimates that are lower than the given threshold are replaced by the threshold. If the field is left empty no constraint is applied.
- **Percentage of training data for validation:**  
Defines the percentage of training data for a model validation by *Forecasting - Validate Forecasting Model*.
- **Plot validation results:**  
Switches the plotting of results of model validation by *Forecasting - Validate Forecasting Model*.

Forecasting

Forecast type: Hierarchical Probabilistic Forecast...

Folder to save regressions:

Time series to be forecast: 1

Correlation threshold: 0

Correlation models regression technique: Polynomial

Correlation model parameters: 1

Percentage of training data for validation: 50  Plot validation results

Figure 4.9: Control elements for Forecasting - Hierarchical Probabilistic Forecasts with QRs

- **Forecast type:**  
Use to select the forecasting model to be created
- **Folder to save regressions:**  
Use the field to write the folder in which all regression models (i.e. forecasting models) trained are going to be saved.
- **Time series to be forecast:**  
Use the field to write the indexes of all time series for which their aggregated probabilistic forecast is required.
- **Correlation threshold:**  
Use the field to write the minimal correlation that two time series should have for them to be assumed as dependent. If the field is left empty all considered time series' are assumed to be dependent.
- **Correlation models regression technique:**  
Use to select the data mining technique for estimating the correlation models. These models are regressions that describe the relationship between the future values of correlated time series.
- **Correlation model parameters:**  
Use the field to write the hyperparameters necessary for training the correlation models. In the case of polynomials two numbers are needed; the first is the maximal allowed degree and the second is the number of features to be selected. Analogously, in the case of ANNs two numbers are needed;

the first is the number of hidden layers and the second is the number of hidden neurons in each hidden layer.

- **Percentage of training data for validation:**  
Defines the percentage of training data for a model validation by *Forecasting - Validate Forecasting Model*.
- **Plot validation results:**  
Switches the plotting of results of model validation by *Forecasting - Validate Forecasting Model*.

## 5 Plugins

- **Negativ -> Null or NaN (Neg2ZeroNaN):** Replace negative values with 0 or NaNs
  - Function name: plugin\_ts\_negativ\_to\_zeroNaN.m
  - Type: TS
  - Time series: 1 inputs, 1 outputs, Segments possible: none
  - Single features: 0 inputs, 0 outputs
  - Images: 0 inputs, 0 outputs
  - Direct callback: none
  - Number of parameters: 1
    - \* Parameter: Value for replacing the negative values (1: replace with zero, 2: replace with NaN)
- **Periodic Integral (PeriodInt):** Integrate a time series for a given period of time
  - Function name: plugin\_ts\_integral\_over\_period.m
  - Type: TS
  - Time series: 1 inputs, 1 outputs, Segments possible: yes
  - Single features: 0 inputs, 0 outputs
  - Images: 0 inputs, 0 outputs
  - Direct callback: none
  - Number of parameters: 1
    - \* Parameter: Period, Reset Integral (Number of timesteps defining a period, Boolean defining if the integral has to be set to zero again after the given period)
- **Remove NaNs with linear Interpolation (NaNfree):** Replaces NaN values with values estimated using a linear interpolation
  - Function name: plugin\_nan\_behandlung.m
  - Type: TS
  - Time series: 1 inputs, 1 outputs, Segments possible: none
  - Single features: 0 inputs, 0 outputs
  - Images: 0 inputs, 0 outputs
  - Direct callback: none
  - Number of parameters: 0

- **Shift Min -> Zero (ShiftMin2Zero):** Add bias to make the time series have zero as a minimum
  - Function name: `plugin_ts_shift_min2zero.m`
  - Type: TS
  - Time series: 1 inputs, 1 outputs, Segments possible: none
  - Single features: 0 inputs, 0 outputs
  - Images: 0 inputs, 0 outputs
  - Direct callback: none
  - Number of parameters: 0

# Bibliography

- [1] R. R. Appino. *Scheduling of Energy Storage using Probabilistic Forecasts and Energy-based Aggregated Models*. PhD thesis, Karlsruhe Institute of Technology, 2019.
- [2] R. R. Appino, J. Á. González Ordiano, R. Mikut, T. Faulwasser, and V. Hagenmeyer. On the use of probabilistic forecasts in scheduling of renewable energy sources coupled to storages. *Applied Energy*, 210:1207 – 1218, 2018.
- [3] R. R. Appino, J. Á. González Ordiano, N. Munzke, R. Mikut, T. Faulwasser, and V. Hagenmeyer. Assessment of a scheduling strategy for dispatching prosumption of an industrial campus. In *Proc., Internationaler ETG-Kongress 2019, 08. – 09.05.2019, Esslingen am Neckar*, pages 289–294. VDE-Verlag, 2019.
- [4] R. R. Appino, M. Muñoz-Ortiz, J. Á. González Ordiano, R. Mikut, V. Hagenmeyer, and T. Faulwasser. Reliable dispatch of renewable generation via charging of time-varying pev populations. *IEEE Transactions on Power Systems*, 34(2):1558–1568, 2019.
- [5] J. Á. González Ordiano. *New Data-Driven Probabilistic Forecasting Methods with Applications in Energy Systems*. PhD thesis, Karlsruher Institut für Technologie (KIT), 2019.
- [6] J. Á. González Ordiano, L. Gröll, R. Mikut, and V. Hagenmeyer. Probabilistic energy forecasting using the nearest neighbors quantile filter and quantile regression. *International Journal of Forecasting*, 2019.
- [7] J. Á. González Ordiano, S. Waczowicz, V. Hagenmeyer, and R. Mikut. Energy forecasting tools and services. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(2):e1235, 2018.
- [8] J. Á. González Ordiano, S. Waczowicz, M. Reischl, R. Mikut, and V. Hagenmeyer. Photovoltaic power forecasting using simple data-driven models without weather data. *Computer Science - Research and Development*, 32(1-2):237–246, 2017.
- [9] R. Mikut, A. Bartschat, W. Doneit, J. Á. González Ordiano, B. Schott, J. Stegmaier, S. Waczowicz, and M. Reischl. The MATLAB toolbox SciXMiner: User’s manual and programmer’s guide. Technical report, arXiv:1704.03298, 2017.
- [10] V. Sharma, U. Cali, V. Hagenmeyer, R. Mikut, and J. Á. González Ordiano. Numerical weather prediction data free solar power forecasting with neural networks. In *Proc., Ninth International Conference on Future Energy Systems, Karlsruhe, Germany*, pages 604–609. ACM, New York, NY, USA, 2018.