

V0.82

Forecasting Interface User Guide



This document is under periodical revisions, contents and screenshots may change.

Table of Contents

1	Introduction.....	3
2	Getting started	3
3	H2G connections	4
3.1	To get data.....	4
3.1.1	The “Nodes” tab	5
3.1.2	The “Files” tab	5
3.1.3	Operations	6
3.1.4	Making groups for ‘ <i>sum</i> ’ or ‘ <i>swap</i> ’ operations	7
3.1.5	Other settings	8
3.1.6	Controls at the bottom	9
3.1.7	Important notes	10
3.2	Creating models and timeseries outputs.....	10
4	Forecasting interface	11
4.1	Control Area.....	12
4.2	Visualization Area	15
4.3	Scheduling of model executions.	16
5	Modelling interface.....	18
5.1	Modelling Area	19
5.2	Management Area	22
5.2.1	Mean Absolute Percentage Error.....	24
5.2.2	Correlation Coefficient.....	24
5.2.3	Root Mean Square Error	24
5.3	Example	25
	Annex A: data input format.....	27
	Annex B: References	28
	Annex C: Known Restrictions	29
	Annex D: Code errors	30

1 Introduction

The Forecasting web application is a client application that uses different forecasting techniques available as web services. These web services are hosted in an external server, property of the UdG.

This web client allows the end-user to upload their own CSV files containing data, create a forecasting model, perform consumption forecasting, and show the results. Mainly, the web application is composed by 2 interfaces: modelling and forecasting. Next sections explore a brief guide to use these interfaces with a practical example.

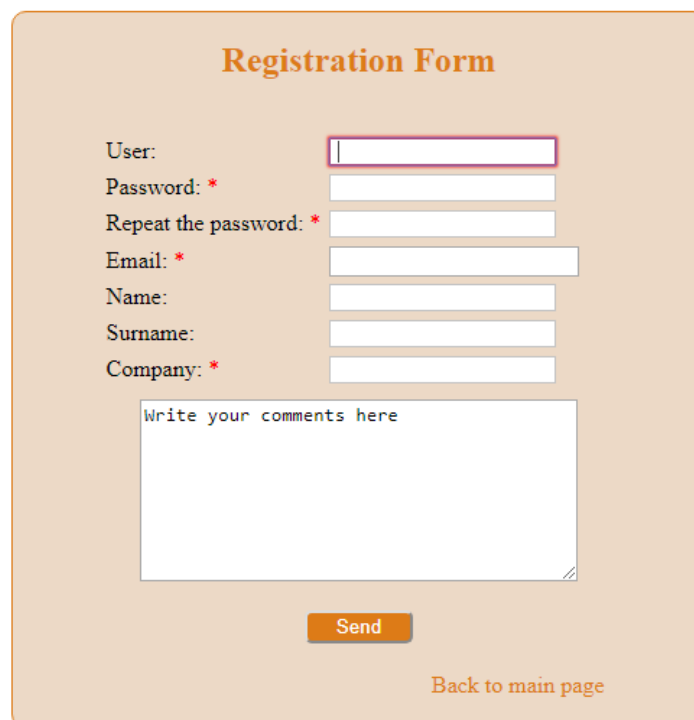
NOTE: This application is currently under development. Disposition of controls and colours can vary during the development stage.

IMPORTANT: it is convenient to know and understand the input data format, since an error in the format may cause subsequent errors with the wrong conclusion "the application does not work". See Annex A.

2 Getting started

Currently, to enter the application, the user must be identified by an ID and password. Thus, previously, the user must fill the registration form (Figure 16). When the information is sent, the administrator will be notified by email in order to enable the account before the user can use the application.

NOTE: Don't forget the password, since the password reset/reminder does not work yet.



The registration form is titled "Registration Form" in orange text. It contains several input fields: "User:" (text), "Password: *" (password), "Repeat the password: *" (password), "Email: *" (text), "Name:" (text), "Surname:" (text), and "Company: *" (text). Below these is a larger text area labeled "Write your comments here". At the bottom, there is an orange "Send" button and a link "Back to main page" in orange text.

Figure 1. Registration form

After login (Figure 2), the user is addressed to the forecasting page (see XXX). Also, depending on the user permissions, the user will be able to access to the modelling interface (Figure 16).

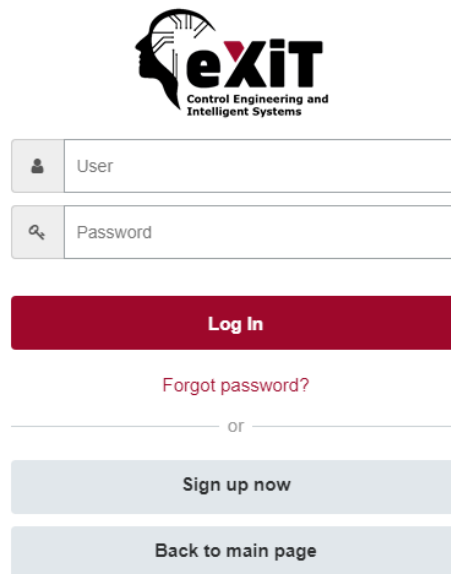
The login window features the eXiT logo at the top center. Below it are two input fields: the first is labeled "User" with a person icon, and the second is labeled "Password" with a key icon. A red "Log In" button is positioned below the password field. Underneath the button is a red link that says "Forgot password?". A horizontal line with the word "or" in the center separates this from the bottom section, which contains two light blue buttons: "Sign up now" and "Back to main page".

Figure 2. Login window

3 H2G connections

From the forecasting or modelling interface it is possible to visualize a modal interface that will allow us to do some basic operations with the H2G core. These are:

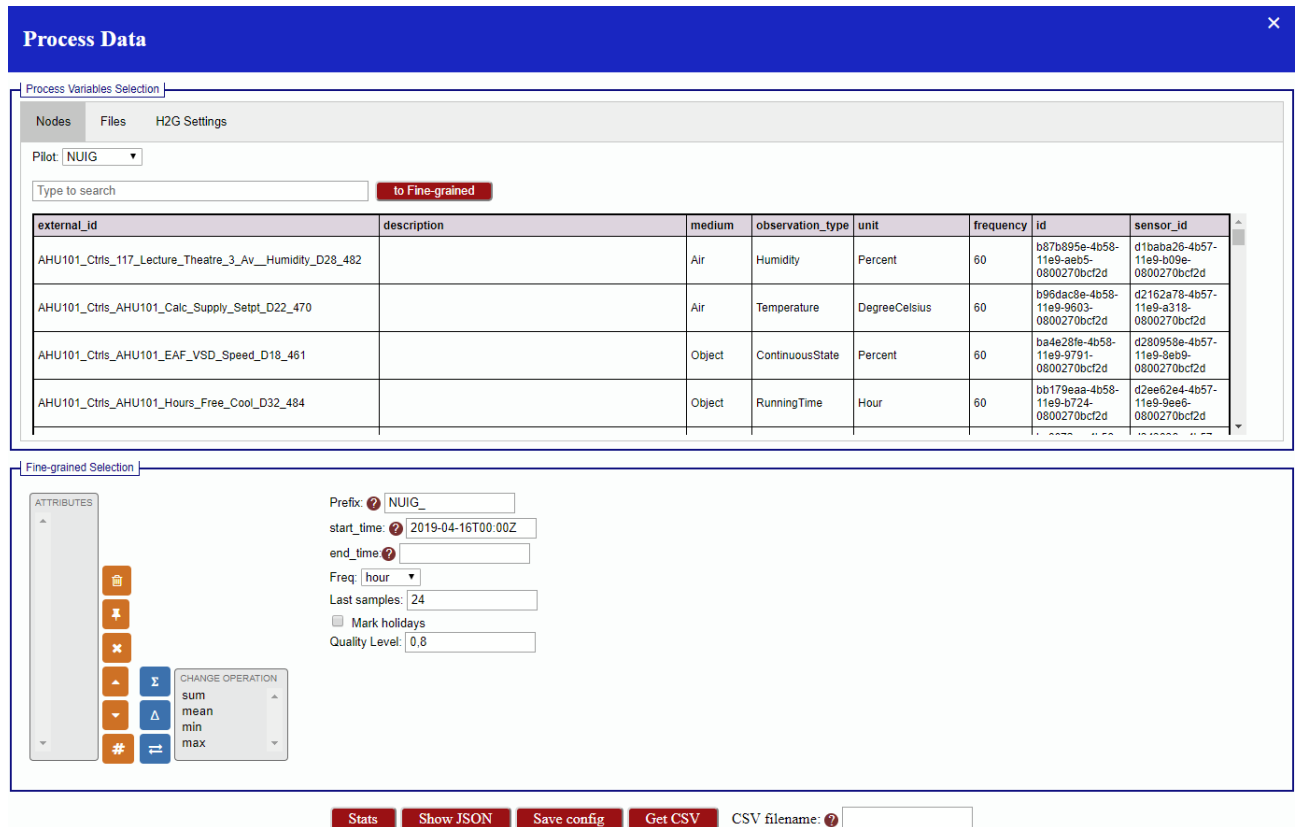
- To get data.
- To create models.
- To create timeseries outputs.

3.1 To get data.

It is possible to get data from the “Nodes” and “Files” tabs (Figure 3 and Figure 4). Next subsections describe the available operations.

3.1.1 The “Nodes” tab

Mainly, the “Nodes” tab (Figure 3) will be the one we must use for the first time. It allows selecting the variables and operations, the start and end time, the sampling time, and the name of the resulting CSV file. If the operation is done successfully, as a result it will be created two files, the data file and a data configuration file, containing the list of variables and settings. It is very important to get this file, since it will allow to configure the automatic execution of models.



Process Data

Process Variables Selection

Nodes Files H2G Settings

Pilot: NUIG

Type to search to Fine-grained

external_id	description	medium	observation_type	unit	frequency	id	sensor_id
AHU101_Ctrls_117_Lecture_Theatre_3_Av__Humidity_D28_482		Air	Humidity	Percent	60	b87b895e-4b58-11e9-9a0b-0800270bcf2d	d1baba26-4b57-11e9-b09e-0800270bcf2d
AHU101_Ctrls_AHU101_Calc_Supply_Setpt_D22_470		Air	Temperature	DegreeCelsius	60	b96dac8e-4b58-11e9-9a0b-0800270bcf2d	d2162a78-4b57-11e9-a31b-0800270bcf2d
AHU101_Ctrls_AHU101_EAF_VSD_Speed_D18_461		Object	ContinuousState	Percent	60	ba4e281e-4b58-11e9-9791-0800270bcf2d	d280958e-4b57-11e9-9a0b-0800270bcf2d
AHU101_Ctrls_AHU101_Hours_Free_Cool_D32_484		Object	RunningTime	Hour	60	bb179eaa-4b58-11e9-b724-0800270bcf2d	d2ae62e4-4b57-11e9-9a0b-0800270bcf2d

Fine-grained Selection

Prefix: NUIG_

start_time: 2019-04-16T00:00Z

end_time:

Freq: hour

Last samples: 24

☐ Mark holidays

Quality Level: 0.8

CHANGE OPERATION

sum
mean
min
max

Stats Show JSON Save config Get CSV CSV filename:

Figure 3 Getting data from the "Nodes" tab.

3.1.2 The “Files” tab

Because the settings of a list of variables could be a long and tedious task, it exist the “Files” tab. Through this tab, it is possible to get data selecting a model or a data configuration file previously created. Here the user can easily modify their configurations for the created models.

NOTE: if the user changes the data configuration file, it may be necessary to create a new model if the variables are different from those used by the model.

Process Data
×

Process Variables Selection
Nodes
Files
H2G Settings

List of models

AAA

Challenger_diana2018w3

Challenger_diana2018w3r02

Challenger_diana2018w3r02

CHA_Elect732019

Config data

AGORE.csv.json

CHAlights.csv.json

CHAL_test_TN_F0_select_mso2018.csv.json

CHA_1h500rAll.csv.json

esgr.csv.json

to Fine-grained

Summary ?

Fine-grained Selection

ATTRIBUTES

Prefix ?

start_time ?

end_time ?

Freq:

Last samples:

☐ Mark holidays

Quality Level:

CHANGE OPERATION
 sum
 mean
 min
 max

Show JSON
Save config
Get CSV
CSV filename: ?

Figure 4 Getting data from the "Files" tab.

3.1.3 Operations

At the bottom of both tabs there is a list with the selected variables (Figure 5) and, on the right, there are some buttons to manage the list and to set each variable individually.

ATTRIBUTES

WIL_BMS31_LEN_RB3_FC_rEnergiaCzynnaPobierana	max	Σ	Δ
WIL_BMS31_LEN_RB3_GN_rEnergiaCzynnaPobierana	max	Σ	Δ
WIL_BMS31_LEN_RB3_OSW_rEnergiaCzynnaPobierana	max	Σ	Δ
WIL_BMS31_LEN_RK3_GN_rEnergiaCzynnaPobierana	max	Σ	Δ
WIL_BMS41_LEN_RB4_NAWILZACZ_rEnergiaCzynnaPobierana	max	Σ	Δ
WIL_BMS41_LEN_RB4_NW1_rEnergiaCzynnaPobierana	max	Σ	Δ
WIL_BMS03_LEN_AGR_FC_rEnergiaCzynna	max	Σ	Δ
WIL_BMS03_LEN_AGR_NW1_rEnergiaCzynna	max	Σ	Δ
OWM_6695624_forecast_temperature	mean		

CHANGE OPERATION

sum







mean

min

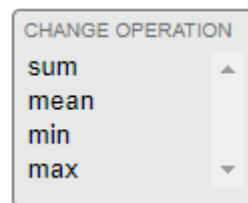
max

Figure 5 List of variables and available operations




To manage the list there exist some buttons in orange.

	Remove all the elements from the table
	Select all the elements of the table
	Remove the selected element from the table
	Move up the selected element. Useful to make groups.
	Move down the selected element. Useful to make groups.
	Insert a separator. Useful to make groups.

The operations performed by the H2G core are documented in the H2G API documentation. It is necessary to assign one operation to each variable, selecting the attribute and then one of the four operations. By default the operation is “max”.



In addition, there are available other useful operations performed by the UdG server. These are selectable with the blue buttons.

	Sum variables. The CSV file will contain an extra column named “S”. The user have to select individually all the attributes to sum, and then select this button. It will appear the Sigma symbol.
	The resulting value will be the difference between samplings. Useful when the measures are meter readings. It is necessary an extra value at the beginning of the desired period.
	Swap operation. It is necessary to have 2 attributes using this operation. The value will be the first attribute by default, but if this does not exist, the value will be replaced by the second attribute but using the name of the first one.

3.1.4 Making groups for ‘sum’ or ‘swap’ operations

It is possible to configure different groups of variables to sum or swap (blue buttons). To do this it is essential to separate the groups in the list using the separator. For example, let’s consider the case of Figure 6. This configuration produces 4 attributes, although the resultant CSV will contain 5 variables. These are:

1. The variable 2.
2. The variable 3.
3. The variables 4, 5 and 6 use the 'sum' option, so this 3 variables generate a unique attribute named by default Agg1.
4. The variable 7 will be the fourth attribute. If some value it was no present in the BEMserver it will be filled by the value of the variable 8, if it exists, since they use the "swap" operator. The CSV also contain this last variable.

ATTRIBUTES		
AHU101_Ctrls_117_Lecture_Theatre_3_Av_Humidity_D28_482	mean	
AHU101_Ctrls_TE_147_1_Lecture_Theatre_3_Temp_D11_469	mean	
AHU101_Ctrls_TE_147_2_Lecture_Theatre_3_Temp_D8_465	mean	
#####		
EnC_EM_008_Chiller_Elec_Accum_Total_D10_204	max	Σ Δ
EnC_EM_008_Chiller_Elec_Daily_Total_D7_215	max	Σ Δ
EnC_EM_008_Chiller_Elec_Hourly_Rate_D6_206	max	Σ Δ
#####		
2964179_current_temperature	mean	⇌
2964179_forecast_temperature	mean	⇌

Figure 6 Example of configuration

3.1.5 Other settings

On the right side of the operations controls, there are some additional inputs to set up data collection (Figure 7).

Prefix: ? WIL_

start_time: ? 2019-04-16T00:00Z

end_time: ? 2019-04-16T00:00Z

Freq: hour ▼

Last samples:

☐ Mark holidays

Quality Level: 0,8

Figure 7 Additional set up for data collection

Prefix: it could be necessary to add a prefix to the tag names. Check the names in the list.

Start_time: initial timestamp to download data. It must have the same format as indicated in the Figure 7.

End_time: final time to download data. This value is optional if a value is entered in the "Last samples" control. See final note.

Freq: sampling time.

Last samples: it indicates how many samples to download, from the beginning of the day (00:00Z).

Mark holidays: it will create an additional column (called ‘*Holidays*’) indicating 0 (working days) or 1 (non-working days). National holidays are based on information from www.timeanddate.com.

Quality Level: it is the minimum quality level to accept the value of the attribute, otherwise it will be discarded.

3.1.6 Controls at the bottom

At the bottom of the tabs there are different controls listed below:

Stats: it calls the API *stats* for the selected variable. It is useful to explore the existence of this variable in the BEMserver.

Show JSON: it shows the resulting JSON according to the chosen configuration. Useful for developers.

Save config: it saves the configuration file (JSON) without downloading the data. It is useful when the user wants to change the name of the resulting aggregated variable, by default *AggN* (see third attribute obtained in the example of Section 3.1.4). To do this, select the file in the “File” tab, and change the name in the “Summary” control (Figure 8). To be effective this choice the download must be done from this tab. Use the buttons under the “Summary” control to save the modified configuration file and/or download new data.

Get CSV: to download the data as a CSV file. A JSON configuration file will also be saved.

CSV filename: the name of the resultant CSV file, without the “.csv” extension.

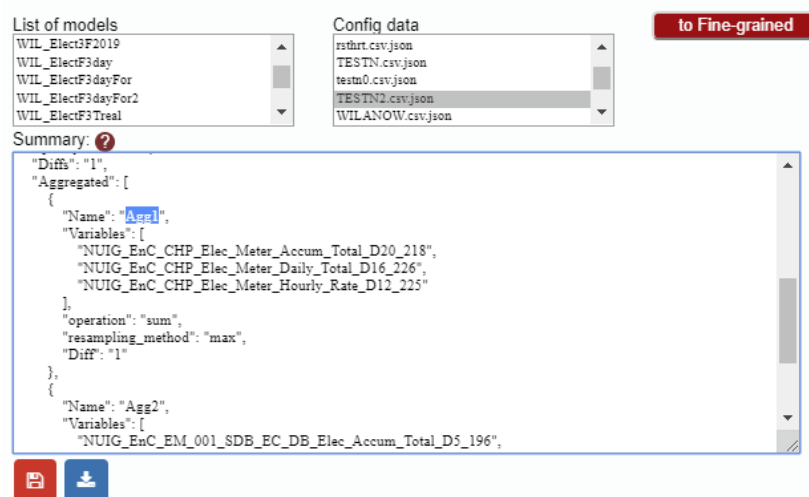


Figure 8 Changing names for aggregated variables.

3.1.7 Important notes

Please, consider:

- Only selected attributes will be downloaded.
- Separators must not be selected.
- If the “difference” operator is used within a group, all the variables must be marked with the symbol Δ (see example in Figure 6).
- If both timestamps are introduced (start and end) these values will be used to download data. If no value is introduced in *end_time*, the start and end time will be calculated according to the ‘*last samples*’ value.
- Finally, a JSON containing the settings will be created. This file will only contain the selected variables. This file will be attached to the model during its creation.

3.2 Creating models and timeseries outputs.

To push results to the H2G core it is mandatory to create (register) first a model, and then create a time series output for this model. The H2G provide the API to do it, but the tab “H2G Settings” (Figure 9) facilitates the user a way to do it. Here there are the following controls:

Models

List of models: they are shown the built local models.

Description: this is a field to enter some description regarding the selected model.

Blue button (register model in the core):

Registered Models:

Red button (delete model from core):

Outputs

External ID: enter a string as a ID for the new output. It is suggested to use a prefix to help the recognition of the origin of this output.

Localization: choose a localization.

Sampling (s): enter the sampling for the created time series in seconds

Unit: select a unit for the created time series.

Blue button (register output in the core):

Registered Outputs: shows the time series outputs created in the H2G core for a selected model.

Red button (delete output from core): remove the selected output from the H2G core.

Process Data

Process Variables Selection

Nodes

Model


H2G Settings

Models

List of models


AAA
Challenger_diarANNh1w3
Challenger_diarANNh1w3v02
Challenger_diarANNh1w3V2
wiTEST_NEW
WIL_ElectF3day
WIL_ElectF3dayFor
WIL_ElectF3dayFor2

Description:



Registered Models

WIL_ElectF31real
WIL_ElectF3dayFor2




Outputs

External ID:

Localization: Challenger

Sampling (s):

Unit: Kilowatthour



Registered Outputs




Figure 9 The H2G settings tab

4 Forecasting interface

The forecasting page (Figure 10) is the main interface of the web application. It is accessible to all users and it allows to forecast energy consumptions once a forecasting model has been created.

Visually, this interface is organized in two areas separated vertically. The left area correspond to the control of the interface, and the right one corresponds to the visualization of the data.

4.1 Control Area

The steps to start the energy consumption forecasting (currently on manual request) are enumerated in this area. Next, these steps are reviewed in more detail.

1a. Upload a file

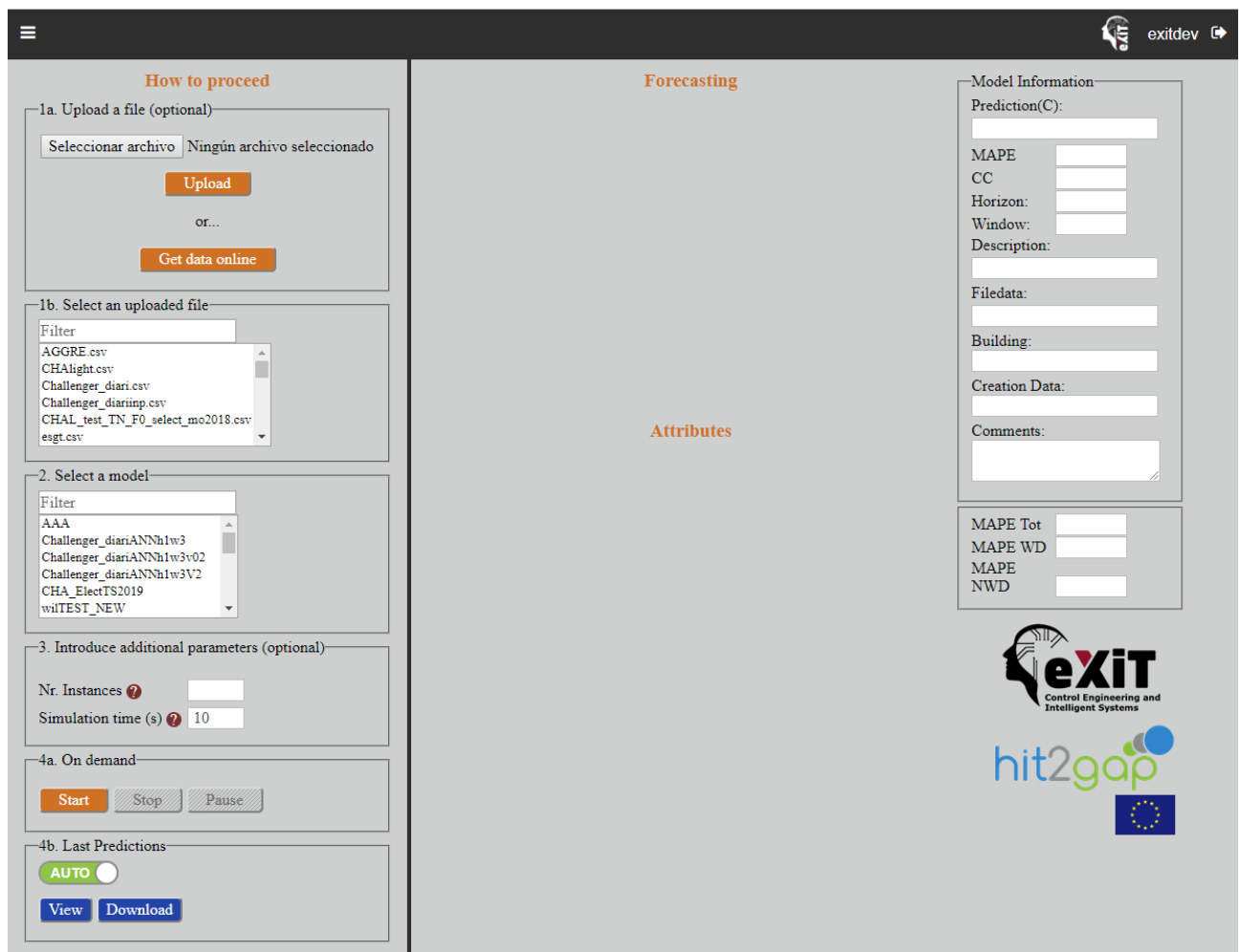
This first step allows the user to upload a CSV file containing data to calculate forecasted consumptions.

Another option is to get data online (see section 3.1)

These steps are optional, if the file has been previously uploaded.

1b. Select a file

Here the user chooses one of the uploaded files to calculate forecasted consumptions, between the available files shown in the 'Files uploaded' control.



The screenshot shows the 'Forecasting' interface of the eXiT system. The interface is divided into several sections:

- How to proceed:**
 - 1a. Upload a file (optional):** Includes a file selection button 'Seleccionar archivo', a status 'Ningún archivo seleccionado', an 'Upload' button, and a 'Get data online' button.
 - 1b. Select an uploaded file:** A dropdown menu with a 'Filter' field. The list includes: AGGRE.csv, CHAight.csv, Challenger_diari.csv, Challenger_diariinp.csv, CHAL_test_TN_F0_select_mo2018.csv, and esgt.csv.
 - 2. Select a model:** A dropdown menu with a 'Filter' field. The list includes: AAA, Challenger_diariANNh1w3, Challenger_diariANNh1w3v02, Challenger_diariANNh1w3V2, CHA_ElectTS2019, and witTEST_NEW.
 - 3. Introduce additional parameters (optional):** Includes input fields for 'Nr. Instances' and 'Simulation time (s)' (set to 10).
 - 4a. On demand:** Includes 'Start', 'Stop', and 'Pause' buttons.
 - 4b. Last Predictions:** Includes an 'AUTO' toggle switch and 'View' and 'Download' buttons.
- Forecasting:** A large central area labeled 'Attributes'.
- Model Information:**
 - Prediction(C):** A text input field.
 - MAPE:** A text input field.
 - CC:** A text input field.
 - Horizon:** A text input field.
 - Window:** A text input field.
 - Description:** A text input field.
 - Filed data:** A text input field.
 - Building:** A text input field.
 - Creation Data:** A text input field.
 - Comments:** A text input field.
 - Summary:** Includes 'MAPE Tot', 'MAPE WD', 'MAPE', and 'NWD' with corresponding text input fields.

The interface also features the eXiT logo, 'hit2gap' logo, and the European Union flag in the bottom right corner.

Figure 10 Forecasting interface

2. Select a model

Here the user chooses one of the created models for forecasting, between the available models shown in the '*List of models*' control.

3. Introduce additional parameters (optional)

Although these parameters are not strictly necessary for the execution of forecasting algorithms, they are important for the obtained result. Please, choose carefully:

- **Number of instances:** it is the number of forecasted values plus the number of previous energy consumption values needed in autoregressive models (parameter "Window"). By default, Number of instances = Window + 1. It means that the next value of energy consumption is forecasted using the previous "Window" recorded energy consumptions.
- **Simulation time:** it is the time between two consecutive calls (in seconds), that is, the time between the executions of two consecutive forecast calculation.

4a. On demand

This section contains the following 3 buttons:

- **Start:** initializes the forecasting process. Also, it enables the buttons "Stop" and "Pause". Then, the interface requests to the forecasting web services the forecast according to the current settings. The server will return the results and they will displayed in the charts.
- **Stop:** stops the forecasting process. The buttons "Stop" and "Pause" will be disabled again.
- **Pause:** it pauses the current forecasting process. That is, the interface will not send any data to the forecasting web services until the button "Pause" will be pressed again.

The steps to calculate a forecasting on demand are:

Select a model and a CSV containing the necessary data for that model.

In the 'Nr Instances' box the minimum number of instances needed to forecast a value is shown. This number is calculated as the sum of the window and horizon parameters (see section 4.2), in the example of Figure 11 is 4. If more than one prediction are desired it should be changed the 'Nr Instances' value. In the same example of the Figure 11 this number has been set to 7, since the window is 3, the algorithm will return the forecast of the next 4 instances.

Previously to start the forecasting, the simulation time could be changed. Usually it depends of the quality of the user's internet connection.

In this mode the algorithm will be running while there are new data in the CSV files. Since this file can contain a large amount of data (is supplied to the algorithm in small groups of instances, 7 in this case), it is necessary to stop the running by clicking on the button "Stop".

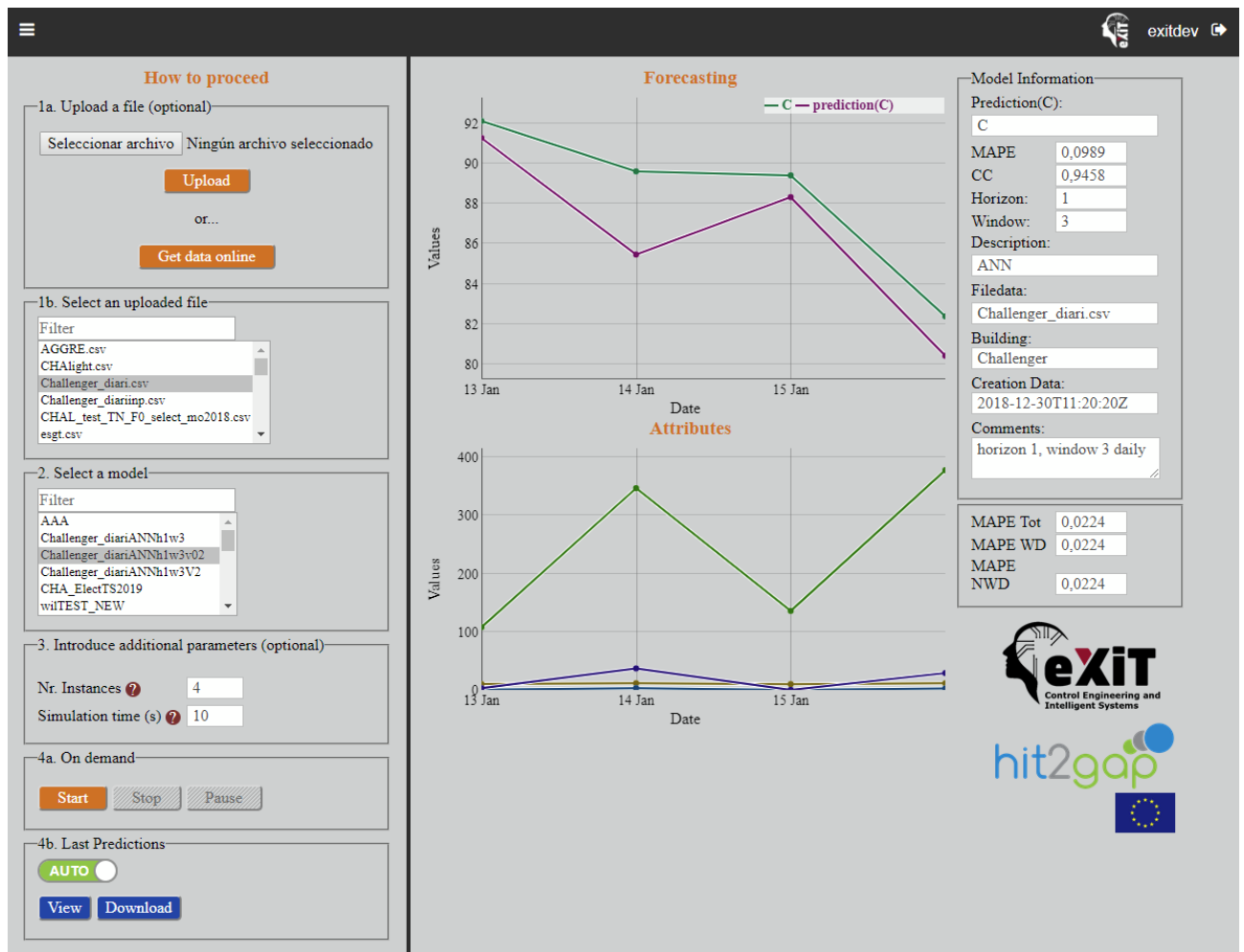


Figure 11 Example of forecasting

4b. Last Predictions

It is possible to visualize the results of previous executions on the graphs (see section 4.2). The UdG server stores the last result for each model, and there may be two results according to the way they were calculated:

AUTO: the forecasting was calculated on a scheduled way (see section 4.3).

MAN: the forecasting was calculated on demand (see above).

To graph the results, click the button “View”. The MAPE indexes will be recalculated again according to the points shown (Figure 12).

Alternatively, it is possible to download the results clicking on the button “Download”.



Figure 12 Visualization of previous results

4.2 Visualization Area

The right side of the forecasting interface contains two charts to display the results, and some fields showing useful parameters about the selected model (Figure 13).

The Model Information area shows some values concerning the settings of the model, such as window and horizon parameters, the used method, the quality indexes obtained by the model and some comments that the user could introduce to help in the identification of the created models.

Below the model information area, the MAPE index is displayed according to the forecasting chart shown. This index is recalculated each time new data is presented in the graph. For the calculation, all the shown data will be used, but only when there is real consumption. The points (predictions) without real consumption will be discarded.

The new MAPE value will be shown in the "MAPE Tot" control. In addition, if the data file contains the column "Holidays", the MAPE corresponding to the working days (MAPE WD) and the holidays (MAPE NWD) will also be calculated. If not, the three controls will show the same value.

See section 5.2 for more information about quality indexes.

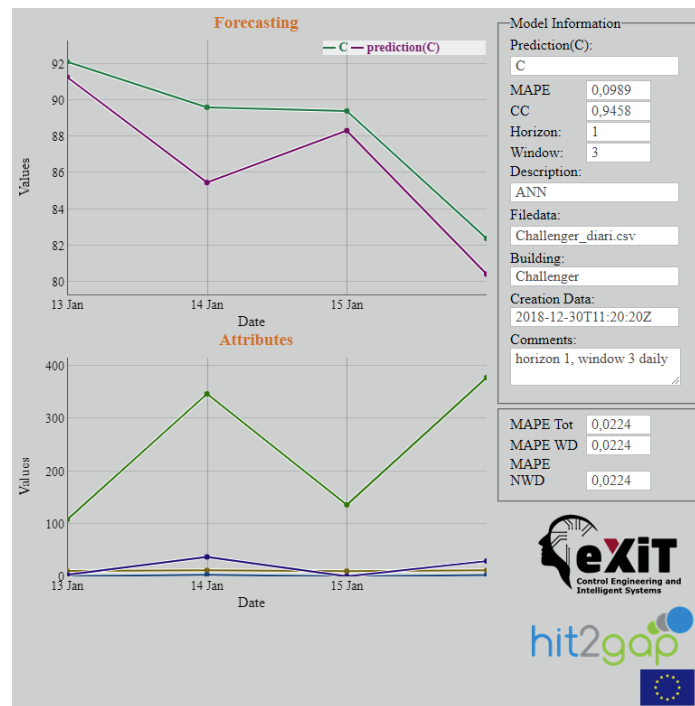


Figure 13 Results presentation

Regarding the graphs, the interpretation of data displayed on them is the following:

- **Graph 1 (upper graph):** The upper graph shows the values of energy consumption predictions: prediction (C). In this example, as the real consumptions are available, they are also graphed (C – green curve).
- **Graph 2 (lower graph):** The lower graph shows the values of the variables used to predict the consumption. In this example: degree days, global radiation, sunshine hours and sunshine rate.

4.3 Scheduling of model executions.

In addition to the mode on demand, it is possible to schedule the executions of models (here called event). To do this, click on the menu button on the top left and select the option “Schedule events” (Figure 14). The modal window shown in Figure 15 will be displayed.

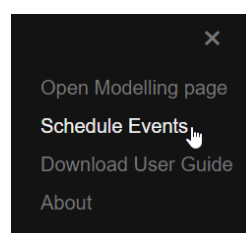


Figure 14 Options accessible from the menu button

Events ✕

Services Calls

List of models

- AAA
- Challenger_diaANNh1w3
- Challenger_diaANNh1w3v02
- Challenger_diaANNh1w3V2
- witTEST_NEW
- WIL_ElectF3day
- WIL_ElectF3dayFor
- WIL_ElectF3dayFor2
- WIL_ElectF3Treat
- WIL_MLR**
- ZZZ

start_time (HH:MM): :

Repeat (DD:HH): :

Summary: ?

```
{
  "IDmodel": "WIL_MLR",
  "Originalfile": "WIL_ElectF3dayFor.csv",
  "Scenario": {
    "horizon": "1",
    "window": "0"
  },
  "Method": {
    "MLR": {
      "feature_selection": "greedy"
    }
  },
  "Building": "Wilanow",
  "ForecastTime": "1 days",
  "IDprediction": "S",
  "Attributes": [
    "6695624_forecast_temperature",
    "S"
  ],
  "TimestampModel": "2018-09-25T10:06:06Z",
}
```

Save event

Events

Active events

- WIL_MLR=07:15/24

Edited events

- WIL_MLR

Start events
Stop events

Summary: ?

```
{
  "IDmodel": "WIL_MLR",
  "Variables": [
    "6695624_forecast_temperature",
    "S"
  ],
  "IDfile": "ZZWIL_MLR.csv",
  "Calls": {
    "Start_time": "07:15",
    "Periodicity": "24"
  }
}
```

Figure 15 Scheduling model executions

The steps to schedule an event are:

1. Select the model from the list.
2. Choose the start time to trigger the execution.
3. Enter the period to repeat the executions in days and hours.
4. Save the event (click on the button).
5. Select the event from the "Edited events" select box.
6. Add the event to the "Active events" select box through the arrow buttons. You can also remove active events from this selector with the right arrow button.
7. Click on the "Start events" button to start all events, or "Stop events" to stop them.
8. Currently, because the UdG server is able to execute automatically the active events, simply click on the "Start events" button and immediately click on the "Stop events". This procedure it is enough to save the active events. The UdG server will recognize them and executes them.

Forecasting user guide

17/30

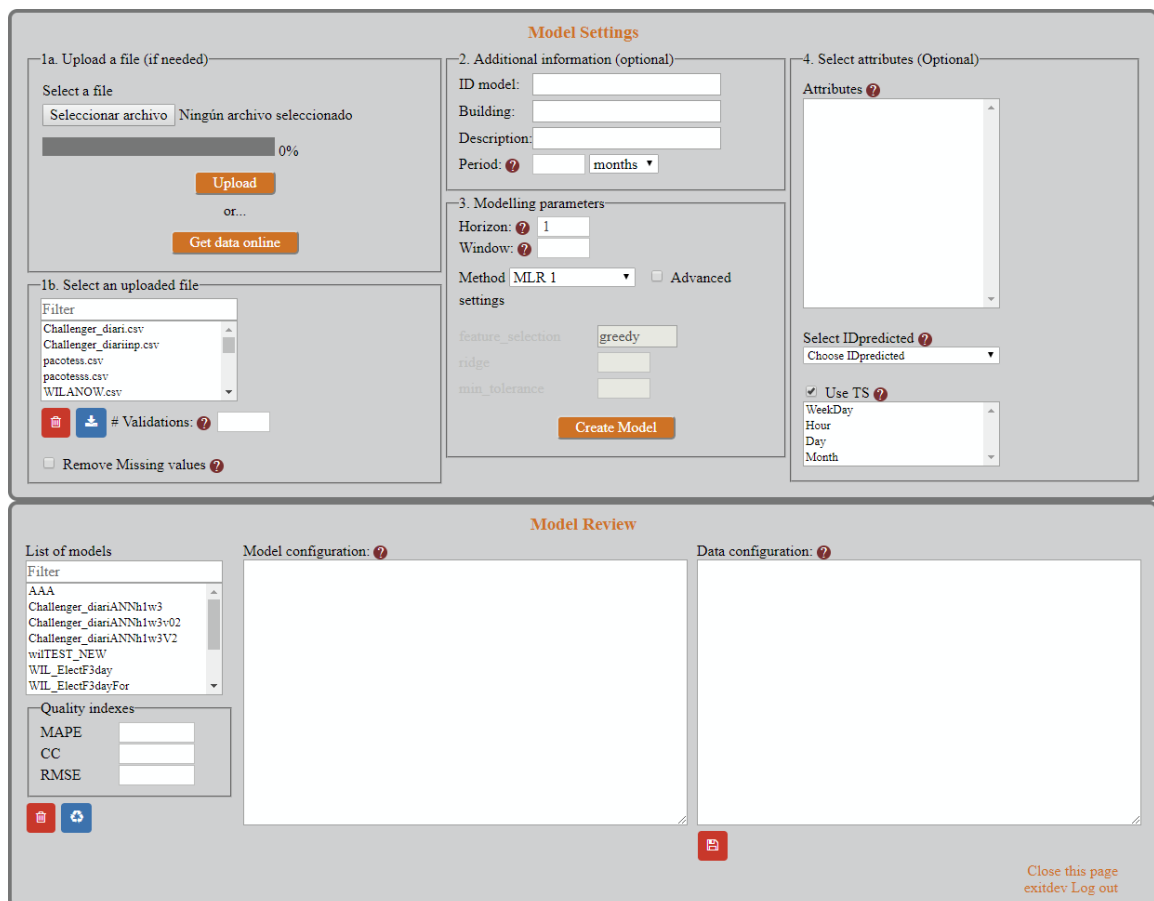
This mode uses the JSON file containing variables settings generated by the application (section 3.1.6), since during the automatic execution new data will be downloaded from the H2G core. This file was attached to the model when it was built (Figure 21 and Figure 22). The schedule won't run without this file.

NOTES:

- Only the active events will be triggered.
- If the scheduling is active the mode on demand is disabled.
- It is necessary to keep the application open in the browser so that the scheduled mode works. This not applies if the models are executed automatically by the UdG server.

5 Modelling interface

The modelling interface is shown in Figure 16. Visually, this interface is organized in two areas separated horizontally. The upper area (Model Settings) corresponds to the modelling parameters, and the lower one (Model Review) corresponds to the management of the models.



The interface is divided into two main horizontal sections: **Model Settings** (top) and **Model Review** (bottom).

Model Settings section:

- 1a. Upload a file (if needed):** Includes a file selection area with a button "Seleccionar archivo", a progress bar at 0%, an "Upload" button, and a "Get data online" button.
- 1b. Select an uploaded file:** Includes a list of files (Challenger_diani.csv, Challenger_dianimp.csv, pacotessa.csv, WILANOW.csv) and a "Remove Missing values" checkbox.
- 2. Additional information (optional):** Includes input fields for "ID model:", "Building:", "Description:", and "Period:" (with a dropdown for "months").
- 3. Modelling parameters:** Includes "Horizon:" (set to 1), "Window:" (set to 1), "Method:" (set to MLR 1), and checkboxes for "Advanced settings", "feature_selection" (set to greedy), "ridge", and "min_tolerance". A "Create Model" button is at the bottom.
- 4. Select attributes (Optional):** Includes an "Attributes" list, a "Select IDpredicted" dropdown (set to Choose IDpredicted), and a "Use TS" checkbox (checked) with a dropdown for "WeekDay", "Hour", "Day", and "Month".

Model Review section:

- List of models:** Includes a filter and a list of models (AAA, Challenger_dianANNh1w3, Challenger_dianANNh1w3v02, Challenger_dianANNh1w3V2, wiTEST_NEW, WIL_ElectF3day, WIL_ElectF3dayFor).
- Quality indexes:** Includes input fields for MAPE, CC, and RMSE.
- Model configuration:** A large empty box for model details.
- Data configuration:** A large empty box for data details.
- Bottom right:** A "Close this page" button and a "Log out" link.

Figure 16. User interface for forecasting modelling

5.1 Modelling Area

The steps to create a model are enumerated in the interface. Next, these steps are reviewed in more detail.

1a. Upload a file

This first step allows the user to upload a CSV file (see format in Annex A) containing historical data to create the model.

Another option is to get data online (see section 3.1)

These steps are optional, if the file has been previously uploaded.

1b. Select a file

Here the user chooses one of the uploaded files to create a model, among the available files shown in the 'Files uploaded' control.

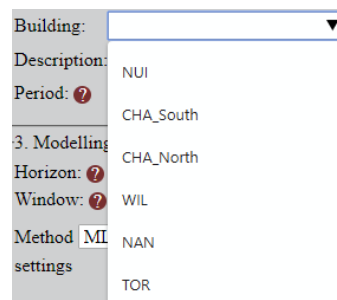
Validations: An optional parameter is the number of subsets for the cross-validation. This is an optional value, it is not necessary to put any value in the box, in this case the default value is 10.

2. Additional information

This section allows the user to introduce some additional information, not necessary for modelling, but useful to describe the model. These are:

IDmodel: identifier of the model. If missing, it will be assigned by the web service. However, it is recommended that the user assign a name to recognize it easily. Valid identifiers must be written as character vector of alphanumerics (A–Z, a–z, 0–9) and underscores.

Building: here the user have to introduce the building related with the model. It is mandatory to choose between the 6 options shown in Figure 17 and do not change the names.



Building:	Options
	NUI
	CHA_South
	CHA_North
	WIL
	NAN
	TOR

Figure 17 Valid options for building input

Description: here the user can introduce other useful information or details about the construction of the model (optional).

Period: this is the time between predictions, according to the periodicity of the input data. It will be used to push data to the H2G server, thus it is a mandatory requirement.

3. Modelling parameters

This section allows the user to introduce some parameters for modelling. The server will save this information in order to use it during exploitation of forecasting.

Horizon: is the forecasted time range in number of instances. For example, if data are gathered every hour and horizon is 24, the forecasting during exploitation is performed for the next 24 hours.

Window: is the number of past instances used in autoregressive models. For example, if data are gathered every hour and window is 6, the forecasting during exploitation for each hour is performed using the consumption of the previous 6 hours.

Method: is the forecasting technique (Figure 18) used to create a forecasting model from the selected historical data. Since different forecasting techniques have been implemented, the user must select from the following options which one s/he wants to use:

a) **MLR** (Multiple Linear Regression)

There are two available options for feature selection. Both using what is known as the Akaike Information Criterion (AIC) [1] to select features for the linear regression:

- **Greedy (MLR 1):** The 'greedy' feature selection however selects a random attribute (that's why it is called greedy) and checks if removing it improves the AIC. This is also repeated until no attribute is dropped anymore.
- **M5 prime (MLR 2):** In each iteration the 'M5 prime' feature selection selects the attribute with the smallest standardized coefficient, removes it and performs another regression. If the result improves AIC, the attribute is dropped. This is repeated until no attribute is dropped anymore.

b) **ANN** (Artificial Neural Network)

c) **SVM** (Support Vector Machine)

It supports various kernel types including:

- **Dot:** The dot kernel is defined by $k(x,y)=x*y$ i.e. it is inner product of x and y .
- **Radial:** The radial kernel is defined by $\exp(-g ||x-y||^2)$ where g is the gamma, it is specified by the kernel gamma parameter. The adjustable parameter gamma plays a major role in the performance of the kernel, and should be carefully tuned to the problem at hand.
- **Polynomial:** The polynomial kernel is defined by $k(x,y)=(x*y+1)^d$ where d is the degree of polynomial and it is specified by the kernel degree parameter. The polynomial kernels are well suited for problems where all the training data is normalized.

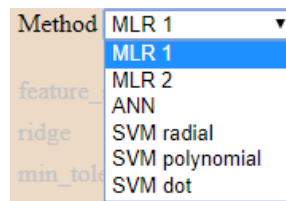


Figure 18 List of available methods

There are also some parameters related to the chosen forecasting method, so they change depending of the method. These parameters are only accessible if the box “Advanced settings” is checked. The complete list of available parameters is:

- MLR
 - **Feature selection:** there are two options, ‘Greedy’ and ‘M5 prime’, explained above.
 - **Ridge:** specifies the ridge parameter for using in ridge regression. **Values range:** 0-100. Default Value: 1E-8
 - **Min tolerance:** it specifies the minimum tolerance for eliminating collinear features. **Values range:** 0-100. Default Value: 0.05
- ANN:
 - **Training cycles:** specifies the number of training cycles used for the neural network training. In back-propagation the output values are compared with the correct answer to compute the value of some predefined error-function. The error is then fed back through the network. Using this information, the algorithm adjusts the weights of each connection in order to reduce the value of the error function by some small amount. This process is repeated n number of times. **Values range:** 100-2500. Default Value: 1200
 - **Learning rate:** This parameter determines how much we change the weights at each step. It should not be 0. **Values range:** (0-0’9]. Default Value: 0.03
 - **Momentum:** adds a fraction of the previous weight update to the current one. This prevents local maxima and smoothes optimization directions. **Values range:** 0-0’9. Default Value: 0.055
- SVM (**Kernel type: radial**)
 - **C:** This is the SVM complexity constant which sets the tolerance for misclassification, where higher C values allow for ‘softer’ boundaries and lower values create ‘harder’ boundaries. A complexity constant that is too large can lead to over-fitting, while values that are too small may result in over-generalization. **Values range:** 0-100. Default Value: 70
 - **Kernel Gamma:** This is the SVM kernel parameter gamma. **Values range:** 0-10. Default Value: 0.9.
- SVM (**Kernel type: polynomial**)
 - **C:** See above. **Values range:** 0-100. Default Value: 70
 - **Kernel Degree:** This is the SVM kernel parameter degree. **Values range:** 1-10. Default Value: 1

- SVM (Kernel type: dot)
 - C: See above. Values range: 0-100. Default Value: 70

NOTE: It is convenient to move around the default values if you do not want to significantly increase the training time.

4. Select Attributes

The users can select if they want to use all the attributes contained in the data or select some of them. For this, the list of available variables within the selected CSV file, is shown in the “Attributes” control. Do not mark any attribute to use all of them in the modelling process, or mark all the attributes you want to use.

In addition, there is a checkbox named “Use TS”. The modelling process creates internally 4 new attributes from the timestamp (listed below the checkbox), but only if the timestamp meets the specified format in Annex A. These attributes will be used in the modelling process by default, unless the user unmarks this checkbox. Otherwise, the user also can select some of these attributes shown in the list.

Finally, it is mandatory to select the variable to be forecasted using the “Select IDpredicted” control

NOTE: the selected attributes can modify significantly the quality of the obtained model.

5. Creating the model

Once introduced all the parameters, click the “Create model” button. The modelling process lasts from a few seconds until several hours, depending of the chosen parameters.

IMPORTANT:

- Creation of a new model using an existent identifier will overwrite the old one.
- The server aborts modelling execution after 2 hours. Contact the administrator to obtain complex models.

5.2 Management Area

The lower area of the interface shows the available models in the “List of models” control. Selecting one of them, it allows the user:

- To view the quality indexes (Figure 19) of the model for evaluation of forecast accuracy (Table 1).
- To view the settings of the model (Figure 20).
- To delete the model definitively (Figure 20, red button).

- To apply its settings (Figure 20, blue button) to the upper area of the interface to create a new model with minor variations.
- To view the variables configuration (Figure 21) if it exist (see note in section 3.1.5). Here it is possible to modify this file and save it again. This action is useful, for example, to remove some variable not used by the model, reducing time in scheduled mode.

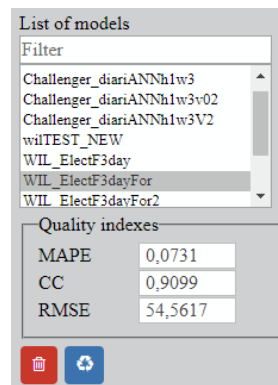


Figure 19 Selection of a created model and quality indexes

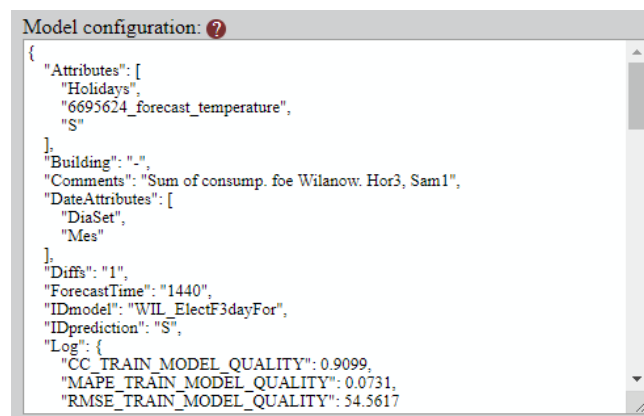


Figure 20 Setting of the model (JSON format)

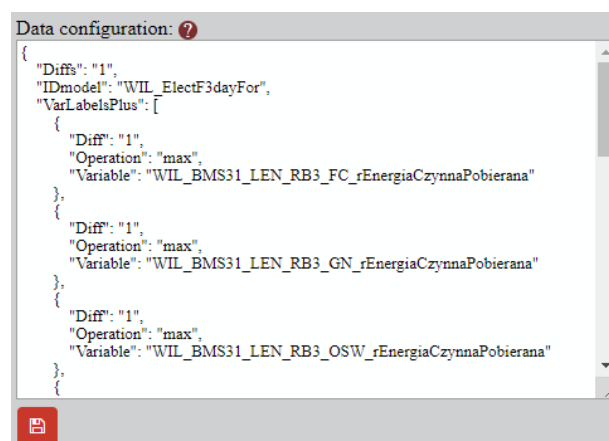


Figure 21 Configuration variables (JSON format)

Table 1. Quality indexes to evaluate forecast accuracy.

Quality Indexes	
MAPE	Mean Absolute Percentage Error
CC	Correlation Coefficient
RMSE	Root Mean Square Error

5.2.1 Mean Absolute Percentage Error

The Mean Absolute Percentage Error (MAPE) of a forecasted variable y is defined as:

$$\text{MAPE (\%)} = 100 \cdot \frac{\sum_{i=1}^N \left| \frac{y_{m(i)} - y_{p(i)}}{y_{m(i)}} \right|}{N} \quad (\text{Eq. 1})$$

where:

N is the number of predictions for the variable y

$y_{m(i)}$ is the i -th actual value (measurement) of the variable y .

$y_{p(i)}$ is the i -th forecasted value (prediction) of the variable y .

MAPE is widely used to evaluate the performance of load forecasting. It does not depend on the magnitude of the unit of measurement, and is used to compare models. If the value of MAPE is small, the model is accurate.

5.2.2 Correlation Coefficient

The Correlation Coefficient (CC) is a measure of the linear relation between two variables, in this case, between the actual value (measurement) and the forecasted value (prediction). Then, it is a measure of how well the predicted values from a forecast model fit the real values.

CC = 1 if total linear correlation exists: accuracy of forecast model is 100%.

5.2.3 Root Mean Square Error

The Root Mean Square Error (RMSE) is the measure of the differences between values predicted by a model and the values actually observed. The RMSE represents the sample standard deviation of the differences between predicted values and measured values.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (y_{m(i)} - y_{p(i)})^2}{N}} \quad (\text{Eq. 2})$$

where:

N is the number of predictions for the variable y

$y_{m(i)}$ is the i -th actual value (measurement) of the variable y .

$y_{p(i)}$ is the i -th forecasted value (prediction) of the variable y .

5.3 Example

Figure 22 shows the modelling interface after the creation of a new model. In this case there is a variable called S in the CSV file resulting from the sum of some other variables contained in the file. This sum was configured as described in section 3.1. Thus, this new variable representing the aggregate consumption was chosen as the variable to forecast.

In the example, only attributes “*Holidays*” and “*6695624_forecast_temperature*” from the CSV file were used to build the model. However, the rest of the variables are necessary to calculate the aggregate consumption ‘ S ’, so it is not convenient to remove them from the variable configuration file (observe the label “*VarLabelsSum*” in the Data configuration control).

Also, note as in the list of generated attributes based on the timestamp, only the Weekday and Month attributes were used. The addition of the other variables generates a model with worse quality.

Model Settings

1a. Upload a file (if needed)

Select a file

Seleccionar archivo Ningún archivo seleccionado

0%

[Upload](#)

or...

[Get data online](#)

2. Additional information (optional)

ID model: WIL_ElectF3Treal

Building: WILANOW - 3rd floor

Description: Electricity kWh + Treal + feb/s

Period: 1440 minutes

4. Select attributes (Optional)

Attributes

Holidays

BMS31_LEN_RB3_FC_rEnergiaCzynnaPobierana

BMS31_LEN_RB3_GN_rEnergiaCzynnaPobierana

BMS31_LEN_RB3_OSW_rEnergiaCzynnaPobierana

BMS31_LEN_RB3_GN_rEnergiaCzynnaPobierana

BMS41_LEN_RB4_NAWILZACZ_rEnergiaCzynnaPobierana

BMS41_LEN_RB4_NW1_rEnergiaCzynnaPobierana

BMS03_LEN_AGR_FC_rEnergiaCzynna

BMS03_LEN_AGR_NW1_rEnergiaCzynna

BMS41_StacjaPogodowa_rTemperatura

6695624_forecast_temperature

S

1b. Select an uploaded file

Filter

WIL_ElectF3dayinp.csv

WIL_ElectF3dayPA.csv

WIL_ElectF3Treal.csv

WIL_Heatinginp.csv

WIL_Heatinginp.csv

Validations: 0

☐ Remove Missing values

3. Modelling parameters

Horizon: 3

Window: 0

Method: ANN

☐ Advanced settings

training_cycles

learning_rate

momentum

[Create Model](#)

Select IDpredicted: S

☒ Use TS

WeekDay

Hour

Day

Month

Model Review

List of models

Filter

wtTEST_NEW

WIL_ElectF3day

WIL_ElectF3dayFor

WIL_ElectF3dayFor2

WIL_ElectF3Treal

WIL_MLR

ZZZ

Quality indexes

MAPE	0,0704
CC	0,9563
RMSE	47,0057

Model configuration:

```

{
  "Attributes": [
    "Holidays",
    "6695624_forecast_temperature",
    "S"
  ],
  "Building": "WILANOW - 3rd floor",
  "Comments": "Electricity kWh + Treal + feb/sep",
  "DataAttributes": [
    "DiaSet",
    "Mes"
  ],
  "Diff": "1",
  "ForecastTime": "1440",
  "Holidays": "Wilanow",
  "IDmodel": "WIL_ElectF3Treal",
  "IDprediction": "S",
  "Log": {
    "CC_TRAIN_MODEL_QUALITY": 0.9563,
    "MAPE_TRAIN_MODEL_QUALITY": 0.0704,
  }
}

```

Data configuration:

```

{
  "Operation": "mean",
  "Variable": "OWM_6695624_forecast_temperature"
},
{
  "VarLabelsSum": [
    "WIL_BMS31_LEN_RB3_FC_rEnergiaCzynnaPobierana",
    "WIL_BMS31_LEN_RB3_GN_rEnergiaCzynnaPobierana",
    "WIL_BMS31_LEN_RB3_OSW_rEnergiaCzynnaPobierana",
    "WIL_BMS31_LEN_RB3_GN_rEnergiaCzynnaPobierana",
    "WIL_BMS41_LEN_RB4_NAWILZACZ_rEnergiaCzynnaPobierana",
    "WIL_BMS41_LEN_RB4_NW1_rEnergiaCzynnaPobierana",
    "WIL_BMS03_LEN_AGR_FC_rEnergiaCzynna",
    "WIL_BMS03_LEN_AGR_NW1_rEnergiaCzynna"
  ],
  "end_time": "2018-09-28T00:00Z",
  "freq": "day",
  "start_time": "2018-01-31T00:00Z"
}

```

[Close this page](#)
[exitdev](#) [Log out](#)

Figure 22 Example of modelling execution

Annex A: data input format

The FDD algorithms have been implemented with the input data in CSV format, with the first row containing the identifiers of the variables and timestamps.

Regarding the format of this first row, it is mandatory:

- Identifier of timestamp must be: *TS*
- Other identifiers: a valid identifier is a character vector of alphanumerics (A–Z, a–z, 0–9) and underscores, such that the first character is a letter.

Regarding data, it is mandatory:

- Decimal separator: decimal point.
- Timestamp is a string using the ISO standard ISO-8601, the format must be:
 - yyyy-mm-ddThh:MM:ssZ. It is UTC time.
 - yyyy-mm-ddThh:MM:ss+00:00. It is UTC time.
 - yyyy-mm-dd hh:MM:ssZ. It is UTC time.
 - yyyy-mm-ddThh:MM:ss+00:00. It is UTC time.
 - Also, the following formats are allowed:
 - yyyy-mm-dd hh:MM:ss. It is local time.
 - yyyy-mm-dd hh:MM. It is local time.
 - yyyy-mm-dd. It is local time.
- Column separator: semicolon.
- All the variables have the same sampling time and in each row all the samples have the same monitoring time.
- Every measurement must correspond to a disaggregated value, for example, the energy consumed every hour.

Missing values & Outliers

Pre-processing of data is not one of the goals of this module, so it is highly recommended that the input data have been previously pre-processed. However, in order to not cause a program interruption, the missing values will be replaced by the average value. The outliers will not be treated.

Example:

```
TS;AHU101_CO_147_1_3_CO2_DATALOG10;AHU101_CO_147_2_3_CO2_DATALOG7
```

```
2017-01-01T00:00:00Z;422.798278809;416.026885986
```

```
2017-01-01T00:01:00Z;422.334899902;414.005615234
```

```
2017-01-01T00:02:00Z;422.206390381;412.897644043
```

Annex B: References

1. Akaike H. A new look at statistical model identification. IEEE Transactions on Automatic Control. 1974;19(6):716–723.

Annex C: Known Restrictions

Not supported in Internet Explorer 9 and earlier versions.

Annex D: Code errors

The web services will return a code error if something fails. Some of these errors and their possible solution are outlined next:

Status	400
Text	ERROR 002: Generation exception: 'Offending attribute: 'TSdate', Expression: 'date_parse_custom(TS, "yyyy-MM-dd HH:mm")', Error: 'Error during evaluation: Invalid argument type for 'date_parse_custom', first argument must be valid date string Error during evaluation: Invalid argument type for 'date_parse_custom', first argument must be valid date string
Solution	using a not valid format for timestamps will cause the interruption of the execution. Revise the format for all timestamps.

Status	400
Text	"ERROR 003: An attribute attX was specified for col... Y, but this column does not exist in input data."
Solution	The header of the CSV file contains N columns, but at least one row of data contains less than N columns. Review the input CSV file.

Status	400
Text	The operator XXXX does not have sufficient capabilities for the given data set: polynomial attributes not supported
Solution	Some attribute is unknown for the algorithm, probably the CSV file does not contain the tag "TS" for the timestamp data.