

---

# AtmoSwing Documentation

*Release 3.0.x*

Pascal Horton

Jul 10, 2023



# CONTENT

<b>1</b>	<b>Getting started</b>	<b>3</b>
1.1	Installation . . . . .	3
1.2	Some analog methods . . . . .	3
1.3	Using AtmoSwing . . . . .	7
1.4	Parameters file . . . . .	7
1.5	Data . . . . .	16
1.6	Preprocessing routines . . . . .	42
1.7	Analogy criteria . . . . .	43
1.8	Versioning scheme and migration . . . . .	43
1.9	Release notes . . . . .	44
<b>2</b>	<b>The Forecaster</b>	<b>65</b>
2.1	Usage . . . . .	65
2.2	Preferences . . . . .	69
2.3	Forecaster parameters file . . . . .	73
<b>3</b>	<b>The Viewer</b>	<b>79</b>
3.1	Usage . . . . .	79
3.2	Preferences . . . . .	87
<b>4</b>	<b>The Downscaler</b>	<b>91</b>
4.1	Usage . . . . .	91
4.2	Preferences . . . . .	93
4.3	Downscaler parameters file . . . . .	96
<b>5</b>	<b>The Optimizer</b>	<b>101</b>
5.1	Usage . . . . .	101
5.2	Classic calibration . . . . .	106
5.3	Genetic algorithms . . . . .	107
5.4	Preferences . . . . .	110
5.5	Verification scores . . . . .	113
5.6	Optimizer parameters file . . . . .	114



*Analog methods (AMs)* are statistical weather forecasting methods, which use synoptic-scale predictors to search in the past for similar days to a target day in order to infer the predictand of interest, such as daily precipitation. They can rely on outputs of numerical weather prediction (NWP) models in the context of operational forecasting or outputs of climate models in the context of climate impact studies. AMs require low computing capacity and have demonstrated useful potential for application in several contexts.

AtmoSwing is open-source software written in C++ that implements AMs in a flexible way so that different variants can be handled dynamically. It comprises four modules: the Forecaster for use in operational forecasting, the Viewer to display the results, the Downscaler for climate studies, and the Optimizer to establish the relationship between predictands and predictors.

*The Forecaster* handles every required processing internally, such as NWP output downloading (when possible) and reading as well as grid interpolation. No external scripts or file conversion are required. The processing of a forecast requires low computing efforts and can even run on a Raspberry Pi computer. It provides valuable results, as revealed by multiple years of operational precipitation forecasting in the Swiss Alps.

*The Viewer* displays the forecasts in an interactive GIS environment with several levels of synthesis and detail. This provides a quick overview of the potential critical situations in the upcoming days, as well as the possibility for the user to delve into the details of the forecasted predictand and criteria distributions.

*The Downscaler* allows for the use of AMs in a climatic context, either for climate reconstruction or for climate change impact studies. When used for future climate studies, it is necessary to pay close attention to the selected predictors so that they contain the climate change signal.

*The Optimizer* implements different optimization techniques, such as a semiautomatic sequential approach, Monte Carlo simulations, and a global optimization technique using genetic algorithms. Establishing a statistical relationship between predictors and predictands is computationally intensive because it requires numerous assessments over decades. To this end, the code was highly optimized for computation efficiency, is parallelized (using multiple threads), and scales well on CPU or GPU clusters. This procedure is only required to establish the statistical relationship, which can then be used for forecasting or downscaling at a low computing cost.

**Reference article:** Horton, P.: AtmoSwing: Analog Technique Model for Statistical Weather forecastING and downscalING (v2.1.0), Geosci. Model Dev., 12, 2915-2940, <https://doi.org/10.5194/gmd-12-2915-2019>, 2019.



## GETTING STARTED

This section contains documentation on how to get started with AtmoSwing. This document can also be downloaded as a pdf or epub using the bottom left menu.

This section is divided into the following topics:

### 1.1 Installation

Executables can be downloaded on [AtmoSwing release page](#). There are two versions available:

- The desktop version: contains all four tools with a graphical user interface (GUI). This version is suited for desktop PCs.
- The server version: contains the Forecaster, the Downscaler, and the Optimizer without GUI. The Viewer is not included. This version is suited for use on servers through a command-line interface.

#### 1.1.1 Windows

The .msi is a Windows installer, and the .zip contains the executable without installer.

#### 1.1.2 Linux

AtmoSwing needs to be compiled on Linux. Instructions are provided on the [AtmoSwing wiki](#).

### 1.2 Some analog methods

Different variants of AMs are presented here. These methods have varying degrees of complexity and comprise one or more subsequent levels of analogy with predictor variables of different kinds. The following content comes from Horton and Brönnimann [4].

The first method developed with NCEP/NCAR Reanalysis 1 by Bontron [1] is based on the analogy of synoptic circulation on the geopotential height at two pressure levels (Z1000 at +12 h and Z500 at +24 h) and is known here as 2Z.

The 2Z method consists of the following steps: firstly, to cope with seasonal effects, candidate dates are extracted from the archive period within four months centered around the target date, for every year of the archive (PC: preselection on a calendar basis in Table below). Then, the analogy of the atmospheric circulation of a target date with every day from the preselection set

(excluding +30 days around the target date and an eventual validation period) is assessed by processing the S1 criterion [2, 8], which is a comparison of gradients, over a defined spatial window (the domain on which the predictors are compared). S1 is processed on each level, and the average is then considered, here with the same weights.

$$S1 = 100 \frac{\sum_i |\Delta \hat{z}_i - \Delta z_i|}{\sum_i \max\{|\Delta \hat{z}_i|, |\Delta z_i|\}}$$

where  $\Delta \hat{z}_i$  is the difference in geopotential height between the  $i$ -th pair of adjacent points of gridded data describing the target situation, and  $\Delta z_i$  is the corresponding observed geopotential height difference in the candidate situation. The smaller the values S1 are, the more similar the pressure fields.

The N1 dates, where N1 is a parameter to be calibrated, with the lowest values of S1 are considered as analogs to the target date. Then, the daily observed precipitation values of the N1 selected dates provide the empirical conditional distribution, considered as the probabilistic prediction for the target date.

Another method relying only on atmospheric circulation is based on the geopotential height on four combinations of pressure levels and temporal windows (4Z, Table below) at levels that were automatically selected by genetic algorithms for the upper Rhone catchment in Switzerland [6]. The 4Z method was shown to outperform 2Z by exploiting more information from the geopotential height and by taking advantage of additional degrees of freedom, such as different spatial windows between the pressure levels and the introduction of weighting between them. However, the 4Z method presented here is a simplification of the results from Horton *et al.* [6], and only the selection of the optimal pressure levels and temporal windows are shown (Z1000 at +06 h and +30 h, Z700 at +24 h, and Z500 at +12 h).

The other methods considered hereafter add a second, or more, subsequent level(s) of analogy after the analogy of the atmospheric circulation, in a stepwise manner.

The next method adds a second level of analogy with moisture variables (method 2Z-2MI, Table below), using a moisture index (MI), which is the product of the total precipitable water (TPW) and the relative humidity at 850 hPa (RH850) [1]. When adding a second level of analogy, N2 dates are subsampled from the N1 analogs of the atmospheric circulation, to end up with a smaller number of analog situations. When this second level of analogy is added, a higher number of analogs N1 is kept at the first level.

Similar to the 4Z method, the 4Z-2MI is a simplification of the methods optimized by genetic algorithms in Horton *et al.* [6]. It consists of a first level of analogy on the geopotential height at four pressure levels (Z1000 at +30 h, Z850 at +12 h, Z700 at +24 h, and Z400 at +12 h), different from 4Z, followed by the moisture index (MI) at two pressure levels (MI700 at +24 h and MI600 at +12 h).

To constrain the seasonal effect, Ben Daoud *et al.* [9] replaced the calendar preselection (+/- 60 days around the target date) by a preselection based on similarity of air temperature (T925 at +36 h and T600 at +12 h, at the nearest grid point). It allows a more dynamic screening of similar situations in terms of air masses as the seasonal signal is also present in the temperature data. The undesired mixing of spring and autumn situations is discussed in Caillouet *et al.* [3]. The number of preselected dates (N0) is equivalent to the number of days selected with the calendar approach and thus depends on the archive size. In this method, named PT-2Z-4MI, the analogy of the atmospheric circulation is the same as in the 2Z method, but the moisture analogy is different (MI925 and MI700 at +12 h and 24 h).



Subsequently, Ben Daoud *et al.* [9] introduced an additional level of analogy between the circulation and the moisture analogy (PT-2Z-4W-4MI, Table below), based on the vertical velocity at 850 hPa (W850). This AM, named “SANDHY” for Stepwise Analogue Downscaling method for Hydrology [3, 9], was primarily developed for large and relatively flat/lowland catchments in France (Saône, Seine). It has also been applied to the whole France territory by Radanovics *et al.* [7] with ERA-40 and by Caillouet *et al.* [3] with 20CR-V2b.

Table 1: Some analog methods listed by increasing complexity. P0 is the preselection (PC: on a calendar basis, that is, +60 days around the target date), L1, L2, and L3 are the subsequent levels of analogy. The meteorological variables are: SLP – mean sea level pressure, Z – geopotential height, T – air temperature, W – vertical velocity, MI – moisture index, which is the product of the relative humidity at the given pressure level and the total water column. The analogy criterion is S1 for SLP and Z and RMSE for the other variables.

Method	P0	L1	L2	L3
2Z	PC	Z1000@12h Z500@24h		
4Z	PC	Z1000@06h Z1000@30h Z700@24h Z500@12h		
2Z-2MI	PC	Z1000@12h Z500@24h	MI850@12+24h	
4Z-2MI	PC	Z1000@30h Z850@12h Z700@24h Z400@12h	MI700@24h MI600@12h	
PT-2Z-4MI	T925@36h T600@12h	Z1000@12h Z500@24h	MI925@12+24h  MI700@12+24h	
PT-2Z-4W-4MI	T925@36h T600@12h	Z1000@12h Z500@24h	W850@06-24h  MI700@12+24h	MI925@12+24h

## 1.3 Using AtmoSwing

AtmoSwing can be used for different purposes:

- in operational forecasting (the Forecaster),
- to downscale climate data (the Downscaler),
- to identify predictors for a certain predictand (the Optimizer).

For any use of AtmoSwing some data need to be prepared, and the method(s) must be calibrated. The following steps are usually required:

1. Get a dataset for the predictor archive. It is often a *reanalysis dataset*. The impact of the choice of the dataset is discussed in Horton and Brönnimann [4].
2. Get a dataset for the target situation (NWP output for operational forecasting or GCM/RCM outputs for climate downscaling)
3. *Format the predictand data.*
4. Calibrate/optimize the method using the Optimizer.

The datasets, predictor variables, number of analogy levels, spatial windows and most other parameters used for the analogue method are defined in the *parameters files*.

## 1.4 Parameters file

The parameters files are essential elements in AtmoSwing as they define the structure of the analog method to use. It is where one defines the datasets, the predictor variables, the number of analogy levels, the spatial windows, etc... Some elements vary between the Forecaster, the Optimizer, and the Downscaler, and thus the target module must be specified in the root tag (ex: <atmoswing version="1.0" target="optimizer">).

The common basic structure is illustrated below for the 2Z-2MI method (first level of analogy on the geopotential height and second on the moisture index) and explained later.

```
<?xml version="1.0" encoding="UTF-8"?>
<atmoswing version="1.0" target="optimizer">
  <description>
    <method_id>2Z-2MI</method_id>
    <method_id_display>Classic humidity</method_id_display>
    <specific_tag>CH</specific_tag>
    <specific_tag_display>all stations</specific_tag_display>
    <description>Classic analogy on the atmospheric circulation and the moisture</
↪description>
  </description>
  <time_properties>
    <archive_period>
      <start>01.01.1981</start>
      <end>31.12.2010</end>
    </archive_period>
    <time_step>24</time_step>
    <time_array_target>
```

(continues on next page)

(continued from previous page)

```

    <time_array>simple</time_array>
  </time_array_target>
  <time_array_analogs>
    <time_array>days_interval</time_array>
    <interval_days>60</interval_days>
    <exclude_days>30</exclude_days>
  </time_array_analogs>
</time_properties>
<analog_dates>
  <analogs_number>75</analogs_number>
  <predictor>
    <preload>1</preload>
    <dataset_id>ECMWF_ERA_interim</dataset_id>
    <data_id>press/hgt</data_id>
    <level>1000</level>
    <time>12</time>
    <spatial_window>
      <x_min>-2.25</x_min>
      <x_points_nb>22</x_points_nb>
      <x_step>0.75</x_step>
      <y_min>40.5</y_min>
      <y_points_nb>11</y_points_nb>
      <y_step>0.75</y_step>
    </spatial_window>
    <criteria>S1</criteria>
    <weight>0.5</weight>
  </predictor>
  <predictor>
    <preload>1</preload>
    <dataset_id>ECMWF_ERA_interim</dataset_id>
    <data_id>press/hgt</data_id>
    <level>500</level>
    <time>24</time>
    <spatial_window>
      <x_min>-2.25</x_min>
      <x_points_nb>22</x_points_nb>
      <x_step>0.75</x_step>
      <y_min>40.5</y_min>
      <y_points_nb>11</y_points_nb>
      <y_step>0.75</y_step>
    </spatial_window>
    <criteria>S1</criteria>
    <weight>0.5</weight>
  </predictor>
</analog_dates>
<analog_dates>
  <analogs_number>25</analogs_number>
  <predictor>
    <preload>1</preload>

```

(continues on next page)

(continued from previous page)

```

<preprocessing>
  <preprocessing_method>HumidityIndex</preprocessing_method>
  <preprocessing_data>
    <dataset_id>ECMWF_ERA_interim</dataset_id>
    <data_id>press/rh</data_id>
    <level>850</level>
    <time>12</time>
  </preprocessing_data>
  <preprocessing_data>
    <dataset_id>ECMWF_ERA_interim</dataset_id>
    <data_id>surf/tcw</data_id>
    <level>0</level>
    <time>12</time>
  </preprocessing_data>
</preprocessing>
<spatial_window>
  <x_min>6</x_min>
  <x_points_nb>1</x_points_nb>
  <x_step>0.75</x_step>
  <y_min>45.75</y_min>
  <y_points_nb>1</y_points_nb>
  <y_step>0.75</y_step>
</spatial_window>
<criteria>RMSE</criteria>
<weight>0.5</weight>
</predictor>
<predictor>
  <preload>1</preload>
  <preprocessing>
    <preprocessing_method>HumidityIndex</preprocessing_method>
    <preprocessing_data>
      <dataset_id>ECMWF_ERA_interim</dataset_id>
      <data_id>press/rh</data_id>
      <level>850</level>
      <time>24</time>
    </preprocessing_data>
    <preprocessing_data>
      <dataset_id>ECMWF_ERA_interim</dataset_id>
      <data_id>surf/tcw</data_id>
      <level>0</level>
      <time>24</time>
    </preprocessing_data>
  </preprocessing>
  <spatial_window>
    <x_min>6</x_min>
    <x_points_nb>1</x_points_nb>
    <x_step>0.75</x_step>
    <y_min>45.75</y_min>
    <y_points_nb>1</y_points_nb>

```

(continues on next page)

(continued from previous page)

```
<y_step>0.75</y_step>
</spatial_window>
<criteria>RMSE</criteria>
<weight>0.5</weight>
</predictor>
</analog_dates>
<analog_values>
  <predictand>
    <station_id>1,2,3,4,5</station_id>
  </predictand>
</analog_values>
</atmoswing>
```

### 1.4.1 Content 'description'

The content of <description> is mainly relevant for the Forecaster, where results are aggregated by methods for different regions.

- <method\_id>: An ID given to the method (ex: 2Z)
- <method\_id\_display>: The name of the method (ex: Classic circulation)
- <specific\_tag>: A tag to identify the specific settings. This is usually the name of the region for which the method was calibrated.
- <specific\_tag\_display>: A more descriptive text for the provided <specific\_tag> above.
- <description>: A possible additionnal description.

Example:

```
<description>
  <method_id>2Z-2MI</method_id>
  <method_id_display>Classic humidity</method_id_display>
  <specific_tag>CH</specific_tag>
  <specific_tag_display>all stations</specific_tag_display>
  <description>Classic analogy on the atmospheric circulation and the moisture</
↪description>
</description>
```

### 1.4.2 Content 'time\_properties'

The content of <time\_properties> defines the archive period and other temporal properties.

### Element 'archive\_period'

The content of <archive\_period> defines the archive period where the analog situations will be extracted. Two options are possible:

Providing the exact dates:

- <start>: Start of the archive period (ex: 01.01.1981)
- <end>: End of the archive period (ex: 31.12.2010)

Alternatively, only providing the years:

- <start\_year>: Start of the archive period (ex: 1962)
- <end\_year>: End of the archive period (ex: 2008)

### Element 'time\_step'

The <time\_step> is time step in hours of the predictand (ex: 24), and thus of the method.

### Element 'time\_array\_target'

The content of <time\_array\_target> defines the temporal properties of the target.

- <time\_array>: Type of time selection.

The type can be one of the following:

- simple: standard full temporal selection
- DJF: Winter (Dec-Feb)
- MAM: Spring (Mar-May)
- JJA: Summer (Jun-Aug)
- SON: Fall (Sep-Nov)
- predictand\_thresholds: Selection of days based on a predictand threshold (ex: days with more than x mm of precipitation)
- Month\_to\_Month: Flexible selection of a period from one month to another month (ex: April\_to\_July, January\_to\_September, September\_to\_March)

In case of a selection based on a predictand threshold (predictand\_thresholds), the following elements must be specified:

- <predictand\_serie\_name>: selection of the raw (data\_raw) or normalized data (data\_normalized)
- <predictand\_min\_threshold>: Minimum threshold value
- <predictand\_max\_threshold>: Maximum threshold value

## Element 'time\_array\_analogs'

The content of `<time_array_analogs>` defines the temporal properties of the analog selection. It is usually defined as an interval of  $\pm 60$  days around the target date (four months centered around the target date). When working in a perfect prognosis framework, 30 days around the target date are excluded for the same year as the target date.

- `<time_array>`: Type of time selection.
- `<interval_days>`: Number of days to select around the target date for every year when using `days_interval`. Usually: 60
- `<exclude_days>`: Number of days to exclude around the target date for the same year. It has to be provided whatever type of time array is selected. Usually: 30

The type can be one of the following:

- `days_interval`: the standard selection of  $\pm x$  days around the target date.
- `simple`: full temporal selection
- `DJF`: Winter (Dec-Feb)
- `MAM`: Spring (Mar-May)
- `JJA`: Summer (Jun-Aug)
- `SON`: Fall (Sep-Nov)
- `Month_to_Month`: Flexible selection of a period from one month to another month (ex: `April_to_July`, `January_to_September`, `September_to_March`)

However, when selecting a season or a specific period for the analogs time array, the same option must be selected for the *Element 'time\_array\_target'*.

Example:

```
<time_properties>
  <archive_period>
    <start>01.01.1981</start>
    <end>31.12.2010</end>
  </archive_period>
  <time_step>24</time_step>
  <time_array_target>
    <time_array>simple</time_array>
  </time_array_target>
  <time_array_analogs>
    <time_array>days_interval</time_array>
    <interval_days>60</interval_days>
    <exclude_days>30</exclude_days>
  </time_array_analogs>
</time_properties>
```

Another example using a specific period:

```
<time_array_target>
  <time_array>April_to_September</time_array>
```

(continues on next page)



(continued from previous page)

```

</time_array_target>
<time_array_analogs>
  <time_array>April_to_September</time_array>
  <exclude_days>30</exclude_days>
</time_array_analogs>

```

### 1.4.3 Content 'analog\_dates'

The content of <analog\_dates> defines the analogy for a given analogy level. Several analogy levels can be successively defined by adding more <analog\_dates> blocks one after the other. Every new level (block) will then subsample from the previous level.

```

<analog_dates>
  ...
</analog_dates>
<analog_dates>
  ...
</analog_dates>

```

### Element 'analogs\_number'

The tag <analogs\_number> defined the number of analogs to select for the given analogy level.

### Element 'predictor'

The element <predictor> provides information about the predictor, the spatial window, and the criteria to use. Several predictors can be used together in an analogy level and are provided by adding more <predictor> blocks one after the other.

Example:

```

<predictor>
  <preload>1</preload>
  <dataset_id>ECMWF_ERA_interim</dataset_id>
  <data_id>press/hgt</data_id>
  <level>1000</level>
  <time>12</time>
  <spatial_window>
    <x_min>-2.25</x_min>
    <x_points_nb>22</x_points_nb>
    <x_step>0.75</x_step>
    <y_min>40.5</y_min>
    <y_points_nb>11</y_points_nb>
    <y_step>0.75</y_step>
  </spatial_window>
  <criteria>S1</criteria>
  <weight>0.5</weight>

```

(continues on next page)

(continued from previous page)

```

</predictor>
<predictor>
  <preload>1</preload>
  <dataset_id>ECMWF_ERA_interim</dataset_id>
  <data_id>press/hgt</data_id>
  <level>500</level>
  <time>24</time>
  <spatial_window>
    <x_min>-2.25</x_min>
    <x_points_nb>22</x_points_nb>
    <x_step>0.75</x_step>
    <y_min>40.5</y_min>
    <y_points_nb>11</y_points_nb>
    <y_step>0.75</y_step>
  </spatial_window>
  <criteria>S1</criteria>
  <weight>0.5</weight>
</predictor>

```

The <predictor> element must define:

- <preload>: Defines if the data must be loaded in memory or not (0/1; optional)
- <dataset\_id>: Defines the dataset to be used (*see the reanalyses list*)
- <data\_id>: Defines the variable to be used (*see the variables list*)
- <level>: Selection of the predictor level (ex: 500 for 500 hPa or 0 for surface)
- <time>: Selection of the predictor time (ex: 12 for 12h UTC)
- <members>: Number of members to select (optional; only for ensemble datasets)
- <criteria>: Criteria to use (ex: S1; *see the list of criteria*)
- <weight>: Weight to give to the predictor when averaging the different criteria values from the different predictors of a level of analogy (ex: 0.6). Optional: if not provided, equal weight is given to all predictors.
- <spatial\_window>: The spatial window on which the predictor variable is compared by means on the criterion. The window is defined by its minimum X (<x\_min>) and Y (<y\_min>) coordinates, the number of points in the direction of higher values (<x\_points\_nb> and <y\_points\_nb>) and the desired resolution (<x\_step> and <y\_step>).

When using an elaborated predictor, the data must go through a preprocessing routine. In this case, the structure is a bit different and can look like this (several predictors can also be defined one after the other):

```

<predictor>
  <preload>1</preload>
  <preprocessing>
    <preprocessing_method>HumidityIndex</preprocessing_method>
    <preprocessing_data>
      <dataset_id>ECMWF_ERA_interim</dataset_id>

```

(continues on next page)

(continued from previous page)

```

    <data_id>press/rh</data_id>
    <level>850</level>
    <time>24</time>
  </preprocessing_data>
  <preprocessing_data>
    <dataset_id>ECMWF_ERA_interim</dataset_id>
    <data_id>surf/tcw</data_id>
    <level>0</level>
    <time>24</time>
  </preprocessing_data>
</preprocessing>
<spatial_window>
  <x_min>6</x_min>
  <x_points_nb>1</x_points_nb>
  <x_step>0.75</x_step>
  <y_min>45.75</y_min>
  <y_points_nb>1</y_points_nb>
  <y_step>0.75</y_step>
</spatial_window>
<criteria>RMSE</criteria>
<weight>0.5</weight>
</predictor>

```

In this case, there is a new `<preprocessing>` block containing:

- `<preprocessing_method>`: The *preprocessing method* to use.
- `<preprocessing_data>`: The predictor data to preprocess defined by `<dataset_id>`, `<data_id>`, `<level>`, and `<time>`.
- The rest of the parameters `<spatial_window>`, `<criteria>`, and `<weight>` are common for the preprocessed predictors.

#### 1.4.4 Content 'analog\_values'

##### Element 'predictand'

Specification of the target predictand timeseries:

- `<station_id>`: *Station ID* or list of station IDs (separated by commas)

Example:

```

<analog_values>
  <predictand>
    <station_id>1,2,3,4,5</station_id>
  </predictand>
</analog_values>

```

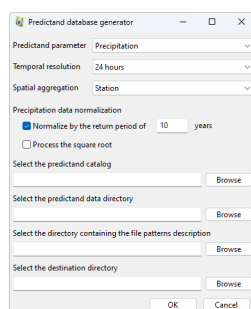
## 1.5 Data

This section describes the data required for AtmoSwing. It is divided into the following topics:

### 1.5.1 Predictand DB

The predictand data (e.g., precipitation) must be prepared for use in AtmoSwing. A tool provided along AtmoSwing allows for the ‘database’ generation. Database generation consists of extracting data from text files and combining them into a NetCDF file.

The interface for generating the precipitation database is accessible through the ‘Tools / Build predictand DB’ menu from any AtmoSwing module.



The first step consists in choosing the type of database to generate. The options are as follows:

- **Predictand parameter:**
  - Precipitation
  - Temperature
  - Lightning
  - Other
- **Temporal resolution:**
  - 24 hours
  - 6 hours
  - 1-hr MTW (moving time window, see Horton *et al.* [6])
  - 3-hr MTW (moving time window, see Horton *et al.* [6])
  - 6-hr MTW (moving time window, see Horton *et al.* [6])
  - 12-hr MTW (moving time window, see Horton *et al.* [6])
- **Spatial aggregation:**
  - Station
  - Groupment
  - Catchment
  - Region

The resulting file will contain the raw predictand data, but can also contain transformed data depending on the predictand variable (see section [Transformed data](#)).

Next, a couple of paths must be provided:

- The predictand catalog: [a catalog \(xml file\)](#) is required to provide some metadata on the different time series.
- The predictand data directory: the directory containing the original predictand files.
- The file patterns description directory: the directory containing [files defining the structure](#) of the original predictand files.
- The destination directory: where the resulting file will be saved.

After execution, a NetCDF file containing all the data necessary for the software is created in the chosen directory. This file also includes the station properties and time series.

## Transformed data

In addition to the raw predictand data, the resulting NetCDF file can also contain transformed data. The available options depend on the predictand variable. It is important to note that this normalized data will be used for the calibration of the method, but the raw data is used in the Forecaster and the Downscaler.

## Precipitation

During the data processing, Gumbel adjustments are automatically made to calculate the precipitation values corresponding to different return periods. The possible transformations are then:

- A normalization of the precipitation amount by a chosen return period value (often 10 years)
- A transformation by the square root of the normalized precipitation amounts (see Bontron [1])

## Lightning

The lightning data can be transformed using:  $\log_{10}(N + 1)$

## Predictand catalog

The predictand catalog provides some metadata on the predictand time series. It is only required to create the predictand database and is not used afterwards. It consists of an xml file such as:

```
<?xml version="1.0" encoding="UTF-8" ?>

<atmoswing version="1.0">
  <dataset>
    <id>MeteoSwiss-Switzerland</id>
```

(continues on next page)

(continued from previous page)

```

<name>MeteoSwiss daily rainfall measurements for Switzerland</name>
<description>Daily precipitation measurements for Switzerland</description>
<parameter>Precipitation</parameter>
<unit>mm</unit>
<temporal_resolution>Daily</temporal_resolution>
<spatial_aggregation>Station</spatial_aggregation>
<time_zone>0</time_zone>
<start note="first value of all stations">01/01/1864</start>
<end note="last value of all stations">31/12/2015</end>
<first_time_step>0</first_time_step>
<nan>-</nan>
<coordinate_system>EPSG: 21781</coordinate_system>

<stations>
  <station>
    <id>1</id>
    <name>Aarberg</name>
    <official_id>ABE</official_id>
    <x_coordinate>588051</x_coordinate>
    <y_coordinate>209518</y_coordinate>
    <height>493</height>
    <file_name>ABE_rre150d0_data.txt</file_name>
    <file_pattern>MeteoSwiss_IDAweb_bulletin</file_pattern>
    <start>01/01/1900</start>
    <end>31/12/2015</end>
  </station>
  <station>
    <id>2</id>
    <name>L'Abergement</name>
    <official_id>ABG</official_id>
    <x_coordinate>527540</x_coordinate>
    <y_coordinate>178770</y_coordinate>
    <height>645</height>
    <file_name>ABG_rre150d0_data.txt</file_name>
    <file_pattern>MeteoSwiss_IDAweb_bulletin</file_pattern>
    <start>01/01/1961</start>
    <end>31/12/2015</end>
  </station>

  ...

  <station>
    <id>301</id>
    <name>Zweisimmen</name>
    <official_id>ZWE</official_id>
    <x_coordinate>595870</x_coordinate>
    <y_coordinate>155480</y_coordinate>
    <height>1015</height>
    <file_name>ZWE_rre150d0_data.txt</file_name>

```

(continues on next page)

(continued from previous page)

```

    <file_pattern>MeteoSwiss_IDAweb_bulletin</file_pattern>
    <start>01/01/1901</start>
    <end>31/12/2015</end>
  </station>

</stations>
</dataset>

</atmoswing>

```

The dataset is described by the following data:

- <id>: a name given to the dataset (short)
- <name>: the full name of the dataset
- <description>: additional description (optional)
- <parameter>: the measured parameter (see above; ex: Precipitation)
- <unit>: the unit of the measure (ex: mm)
- <temporal\_resolution>: the temporal resolution (see above; ex: Daily)
- <spatial\_aggregation>: the spatial aggregation (see above; ex: Station)
- <time\_zone>: time zone of the measurement time (ex: 0 for UTC)
- <start>: first time step of all stations DD/MM/YYYY (ex: 01/01/1864)
- <end>: last time step of all stations DD/MM/YYYY (ex: 31/12/2015)
- <first\_time\_step>: first time step (ex: 0)
- <nan>: representation of nan values
- <coordinate\_system>: the EPSG code of the coordinate system (ex: EPSG:21781)

Then, every time series (named here 'station', even if they are not representing a station) must be detailed under a <station> tag in the <stations> list. For every station, the following information can/must be provided:

- <id>: a unique ID (integer) to identify the station in AtmoSwing (unique number for the considered catalog)
- <name>: the name of the station
- <official\_id>: an official ID (optional)
- <x\_coordinate>: the X coordinate (in the coordinate system defined for the dataset)
- <y\_coordinate>: the Y coordinate (in the coordinate system defined for the dataset)
- <height>: the station height (optional)
- <file\_name>: the file name containing the original data
- <file\_pattern>: *the pattern describing the data file structure*
- <start>: the start of the time series DD/MM/YYYY
- <end>: the end of the time series DD/MM/YYYY

## Predictand file structure

It was chosen not to force users to transform their data into a specific format but to let them describe the structure of the files. It must be done using files describing the structure of the original time series. The structure file defines how to read data files and is described in text files. However, a restriction is imposed: a data file must contain only one time series (single station), or structure files must be defined to read a specific column.

Files can be structured (<structure\_type>) in two common types:

- Delimited by tabs (tabdelimited)
- With constant width (constantwidth)

An option (<parsetime>) allows reading the date/time entries or ignore them. Reading this information allows an automatic check of the correct consistency of the dates between what the software thinks it is reading and what it is actually reading. It is therefore recommended to enable this option (<parse\_time>1</parse\_time>).

Several files are installed with AtmoSwing and can be found in the 'data/file\_patterns' directory, or [online in the repository](#).

## Example of the type 'delimited by tabs'

This type straightforwardly defines structured files, eventually with date entries (year, month, day) and data separated by tabs (see example below). The files contain exactly two header lines:

```
Timeseries processed by XY for the YZ project.
average from station(s) : 8080 8010
1961    1        1        6.6
1961    1        2        2.6
1961    1        3       21.5
1961    1        4        4.5
1961    1        5        0.0
1961    1        6        5.8
1961    1        7        0.0
1961    1        8        3.4
1961    1        9        0.0
1961    1       10        1.5
```

The structure describing such data files is as follows:

```
<?xml version="1.0" encoding="UTF-8" ?>
<atmoswing version="1.0">
  <pattern id="Basic_daily" name="Basic daily file structure with tabs">
    <structure_type>tabs_delimited</structure_type>
    <header_lines>2</header_lines>
    <parse_time>1</parse_time>
    <time>
      <year column="1"/>
      <month column="2"/>
```

(continues on next page)



(continued from previous page)

```

    <day column="3"/>
  </time>
  <data>
    <value column="4"/>
  </data>
</pattern>
</atmoswing>

```

### Example of the type 'constant width'

Predictand files are here structured as constant columns width.

```

...
...
stn    time                rre150d0
ABG    19610101             4.3
ABG    19610102            10.2
ABG    19610103            14.1
ABG    19610104             2.4
ABG    19610105             2.6
ABG    19610106             3.7
ABG    19610107             0.0
ABG    19610108             6.5
ABG    19610109             0.0
ABG    19610110            10.7

```

The structure describing such data files is as follows:

```

<?xml version="1.0" encoding="UTF-8" ?>
<atmoswing version="1.0">
  <pattern id="MeteoSwiss_IDAweb_bulletin" name="MeteoSwiss data pattern from_
↪ IDAweb">
    <structure_type>constant_width</structure_type>
    <header_lines>3</header_lines>
    <parse_time>1</parse_time>
    <time>
      <year char_start="8" char_end="11"/>
      <month char_start="12" char_end="13"/>
      <day char_start="14" char_end="15"/>
    </time>
    <data>
      <value char_start="25" char_end="30"/>
    </data>
  </pattern>
</atmoswing>

```

### 1.5.2 Reanalyses

A description of the different reanalyses and the assessment of the impact of the dataset selection are presented in Horton and Brönnimann [4]. AtmoSwing can read the following reanalyses (with the corresponding ID to use):

- **NCEP/NCAR Reanalysis 1: NCEP\_R1**
  - Use the original file names (yearly files)
- **NCEP/DOE Reanalysis 1: NCEP\_R2**
  - Use the original file names (yearly files)
- **NCEP CFSR original Grib files: NCEP\_CFSR**
  - Use the original file names and directory structures
- **NCEP CFSR subset from NCAR/UCAR as netcdf files: NCEP\_CFSR\_subset**
  - Use the original file names (yearly files)
- **ECMWF ERA5: ECMWF\_ERA5**
  - Use the original variable name in the file name (ex: z for geopotential)
  - Can be structured as a single file (ex: z.nc)
  - Can be structured as yearly files (ex: z.1998.nc)
  - Can be structured differently, but must be isolated in a directory.
- **ECMWF ERA-interim: ECMWF\_ERA\_interim**
  - Use the original variable name in the file name (ex: z for geopotential)
  - Can be structured as a single file (ex: z.nc)
  - Can be structured as yearly files (ex: z.1998.nc)
  - Can be structured differently, but must be isolated in a directory.
- **ECMWF ERA 20th Century: ECMWF\_ERA\_20C**
  - Must be 1 file per variable with the original variable name in the file name (ex: z.nc)
- **ECMWF Coupled ERA 20th Century: ECMWF\_CERA\_20C**
  - Must be yearly files with the original variable name in the file name (ex: z.1998.nc)
- **NASA MERRA-2 (Modern-Era Retrospective analysis for Research and Applications, Version 2): NASA\_MERRA\_2**
  - Use the original file names and directory structures
- **NASA MERRA-2 subset from MDISC: NASA\_MERRA\_2\_subset**
  - Use the original file names
- **JMA Japanese 55-year Reanalysis subset from NCAR/UCAR: JMA\_JRA\_55\_subset**
  - Use the original file names

- **JMA Japanese 55-year Reanalysis Conventional subset from NCAR/UCAR:** `JMA_JRA_55C_subset`
  - Use the original file names
- **NOAA Twentieth Century Reanalysis v2c:** `NOAA_20CR_v2c`
  - Use the original file names (yearly files – ex: `z.1956.nc`)
- **NOAA Twentieth Century Reanalysis v2c ensemble:** `NOAA_20CR_v2c_ens`
  - Use the original file names (yearly files – ex: `z500_1956.nc`)
- **NOAA Optimum Interpolation Sea Surface Temperature, version 2:** `NOAA_OISST_v2`
  - Use the original file names (single files – ex: `sst_1deg.nc`)
- **A generic NetCDF dataset:** `Generic[___]`
  - Use the original variable name in the file name (ex: `z` for geopotential)
  - Can be structured as a single file (ex: `z.nc`)
  - Can be structured as yearly files (ex: `z.1998.nc`)
  - Can be structured differently, but must be isolated in a directory.
  - *See details [here](#)*

### 1.5.3 Generic dataset

In addition to the multiple reanalyses supported, AtmoSwing allows using a generic dataset as NetCDF files. The dataset can be named as follows:

- **Generic:** simple name to be used mainly when the generic dataset is issued from a single source.
- **Generic[\_\_\_]:** the name can also be composed of the term `Generic` and any other suffix. It allows specifying the source of the dataset (for example `Generic_ERA5` or `Generic_CFSR`). This is mainly useful for the Viewer to differentiate the datasets.

The files can be organized as follows (AtmoSwing will look for possible matches):

- A single file with the name of the desired predictor variable (ex: `rh.nc`)
- Yearly files with the year in the file name
- All the files in a directory. The files need to be named in order to be sorted chronologically.

### File structure

The file needs to have four dimensions:

- `lat:` latitude
- `lon:` longitude
- `level:` needs to be defined even if there is no level (set to 0 in this case)
- `time:` needs to be expressed as Modified Julian Date (days since 1858-11-17 00:00:00.0 +0:00)

The generic dataset is different from the other datasets as there are no predefined variables. The name given to the variable must also be used in the xml parameters file.

### Generate generic datasets

The [AtmoSwing Python toolbox repository](#) contains scripts to convert some reanalyses to the generic format (for example [scripts/createGenericFromERA5.py](#)).

One must specify:

- `dir_origin`: the directory containing the original files
- `dir_target`: the directory to save the resulting files
- `files_list`: a list of the files to transform containing the following entries: 1) the subdirectory containing the original files (for example `pressure` or `surface_analysis`), 2) the name of the variable in the original files, 3) the name given to the variable in the resulting file.

The scripts might need to be modified to suit other input formats or structures.

### 1.5.4 Forecaster specific data

AtmoSwing Forecaster can read the following NWP outputs (with the corresponding ID to use):

- NWS GFS forecasts (automatically downloaded from [NOMADS](#)): `NWS_GFS`
- NWS GFS forecasts, but without download: `NWS_GFS_Local`
- ECMWF IFS forecasts (no automatic download): `ECMWF_IFS`
- Météo-France ARPEGE forecasts (no automatic download): `MF_ARPEGE`

### Specify the domain to download

**Warning:** When choosing GFS, the predictors are downloaded for the European domain by default: longitude from -32° to 42° and latitude from 20° to 72°. The domain can be changed in the `AtmoSwingForecaster.ini` file.

The default domain and levels to be downloaded are specified in the URLs in `AtmoSwingForecaster.ini` (under your profile directory; typically `C:\Users\<username>\AppData\Roaming\AtmoSwing` for Windows or `.AtmoSwing` for Linux/OSX). The domain is adjusted by changing the values of `leftlon=-32&rightlon=42&toplat=72&bottomlat=20`, and the levels with `lev_XXX_mb=on`.

```
[PredictorsUrl]
[PredictorsUrl/GFS]
hgt=https://nomads.ncep.noaa.gov/cgi-bin/filter_gfs_0p50.pl?file=gfs.
↪t[CURRENTDATE-hh]z.pgrb2full.0p50.f[LEADTIME-hhh]&lev_300_mb=on&lev_400_mb=on&
↪lev_500_mb=on&lev_600_mb=on&lev_700_mb=on&lev_850_mb=on&lev_925_mb=on&lev_1000_
↪mb=on&var_HGT=on&subregion=&leftlon=-32&rightlon=42&toplat=72&bottomlat=20&dir=
```

(continues on next page)

(continued from previous page)

```

↪%2Fgfs.[CURRENTDATE-YYYYMMDD]/[CURRENTDATE-hh]
temp=https://nomads.ncep.noaa.gov/cgi-bin/filter_gfs_0p50.pl?file=gfs.
↪t[CURRENTDATE-hh]z.pgrb2full.0p50.f[LEADTIME-hhh]&lev_300_mb=on&lev_400_mb=on&
↪lev_500_mb=on&lev_600_mb=on&lev_700_mb=on&lev_850_mb=on&lev_925_mb=on&lev_1000_
↪mb=on&var_TMP=on&subregion=&leftlon=-32&rightlon=42&toplat=72&bottomlat=20&dir=
↪%2Fgfs.[CURRENTDATE-YYYYMMDD]/[CURRENTDATE-hh]
vvel=https://nomads.ncep.noaa.gov/cgi-bin/filter_gfs_0p50.pl?file=gfs.
↪t[CURRENTDATE-hh]z.pgrb2full.0p50.f[LEADTIME-hhh]&lev_300_mb=on&lev_400_mb=on&
↪lev_500_mb=on&lev_600_mb=on&lev_700_mb=on&lev_850_mb=on&lev_925_mb=on&lev_1000_
↪mb=on&var_VVEL=on&subregion=&leftlon=-32&rightlon=42&toplat=72&bottomlat=20&dir=
↪%2Fgfs.[CURRENTDATE-YYYYMMDD]/[CURRENTDATE-hh]
rh=https://nomads.ncep.noaa.gov/cgi-bin/filter_gfs_0p50.pl?file=gfs.t[CURRENTDATE-
↪hh]z.pgrb2full.0p50.f[LEADTIME-hhh]&lev_300_mb=on&lev_400_mb=on&lev_500_mb=on&
↪lev_600_mb=on&lev_700_mb=on&lev_850_mb=on&lev_925_mb=on&lev_1000_mb=on&var_
↪RH=on&subregion=&leftlon=-32&rightlon=42&toplat=72&bottomlat=20&dir=%2Fgfs.
↪[CURRENTDATE-YYYYMMDD]/[CURRENTDATE-hh]
uwnd=https://nomads.ncep.noaa.gov/cgi-bin/filter_gfs_0p50.pl?file=gfs.
↪t[CURRENTDATE-hh]z.pgrb2full.0p50.f[LEADTIME-hhh]&lev_300_mb=on&lev_400_mb=on&
↪lev_500_mb=on&lev_600_mb=on&lev_700_mb=on&lev_850_mb=on&lev_925_mb=on&lev_1000_
↪mb=on&var_UGRD=on&subregion=&leftlon=-32&rightlon=42&toplat=72&bottomlat=20&dir=
↪%2Fgfs.[CURRENTDATE-YYYYMMDD]/[CURRENTDATE-hh]
vwnd=https://nomads.ncep.noaa.gov/cgi-bin/filter_gfs_0p50.pl?file=gfs.
↪t[CURRENTDATE-hh]z.pgrb2full.0p50.f[LEADTIME-hhh]&lev_300_mb=on&lev_400_mb=on&
↪lev_500_mb=on&lev_600_mb=on&lev_700_mb=on&lev_850_mb=on&lev_925_mb=on&lev_1000_
↪mb=on&var_VGRD=on&subregion=&leftlon=-32&rightlon=42&toplat=72&bottomlat=20&dir=
↪%2Fgfs.[CURRENTDATE-YYYYMMDD]/[CURRENTDATE-hh]
pwat=https://nomads.ncep.noaa.gov/cgi-bin/filter_gfs_0p50.pl?file=gfs.
↪t[CURRENTDATE-hh]z.pgrb2full.0p50.f[LEADTIME-hhh]&lev_entire_atmosphere_%5C
↪%28considered_as_a_single_layer%5C%29=on&var_PWAT=on&subregion=&leftlon=-32&
↪rightlon=42&toplat=72&bottomlat=20&dir=%2Fgfs.[CURRENTDATE-YYYYMMDD]/
↪[CURRENTDATE-hh]

```

### 1.5.5 Downscaler specific data

AtmoSwing Downscaler can read the following climate data in addition to reanalyses datasets (with the corresponding ID to use):

- CMIP5 (Coupled Model Intercomparison Project Phase 5): CMIP5
- CORDEX (Coordinated Regional Climate Downscaling Experiment): CORDEX

### 1.5.6 Predictor variables

The different existing datasets do not name the meteorological variables the same way. AtmoSwing can interact with many datasets and has several of their variables defined. These datasets often provide different products, for example, for different types of levels (isobaric, isentropic, surface, etc.). As some variables can be defined for different levels (e.g., temperature), the product/level may need to be specified. Thus, the selection of a variable in the parameters file can be defined as level/variable, such as pressure/z for the geopotential on pressure levels, or isentropic/t for the temperature on isentropic levels. As different level types and variables are present for different datasets, they can be named differently in AtmoSwing. For example pressure/z, pl/h, and isobaric/hgt are identical. The name of the level (or product) has to be the same as the corresponding subfolder name. For example, someone using NCEP/NCAR Reanalysis 1 and having a subfolder 'pressure' should then use pressure/z for the geopotential height.

The following sections define first the different levels, then some common variables to most datasets, and finally, list the available variables for the different datasets.

#### Content of this section

- *Differentiating subsets/levels*
- *Common variables*
- *Variables from datasets*
  - *NCEP/NCAR Reanalysis 1*
  - *NCEP/DOE Reanalysis 2*
  - *NCEP CFSR*
  - *NCEP CFSR subset*
  - *ECMWF ERA5*
  - *ECMWF ERA-interim*
  - *ECMWF ERA 20th Century*
  - *ECMWF Coupled ERA 20th Century*
  - *ECMWF IFS*
  - *NWS GFS*
  - *NASA MERRA-2*
  - *NASA MERRA-2 subset*
  - *JMA JRA-55 subset*
  - *JMA JRA-55C subset*
  - *NOAA 20CR v2c*
  - *NOAA 20CR v2c ensemble*
  - *NOAA OISST v2*
  - *METEO-FRANCE ARPEGE*

## Differentiating subsets/levels

The level can usually be defined using different terms. The provided terms are interchangeable.

### Pressure level:

- pressure\_level
- pressure\_levels
- pressure
- press
- isobaric
- pl
- pgbh
- pgbhnl
- pgb

### Isentropic level:

- isentropic\_level
- isentropic
- potential\_temperature
- pt
- ipvh
- ipv

### Surface:

- surface
- surf
- ground
- sfc
- sf

### Flux:

- surface\_fluxes
- fluxes
- flux
- flxf06
- flx

### Column:

- total\_column
- column
- tc
- entire\_atmosphere
- ea

**Potential vorticity:**

- potential\_vorticity
- pv
- pv\_surface
- epv

**Common variables**

Some variables are defined with different names (aliases) across the datasets:

**Geopotential:**

- z
- h
- zg

When loading geopotential data, AtmoSwing automatically converts it to geopotential height.

**Geopotential height:**

- z
- h
- zg
- hgt

**Air temperature:**

- t
- temp
- tmp
- ta
- air

**Relative humidity:**

- rh
- rhum
- hur
- r



**Specific humidity:**

- sh
- shum
- hus
- q
- qv

**Vertical velocity:**

- w
- vvel
- wap
- omega

**Precipitable water:**

- pwat
- p\_wat
- pr\_wtr
- prwtr

**Total column water:**

- tcw

**Total column water vapour:**

- tcwv

**Pressure:**

- pressure
- press
- pres

**Sea level pressure:**

- slp
- mslp
- ps1
- prms1
- msl

**U wind component:**

- u
- ua
- ugrd

- u\_grd
- uwnd

**V wind component:**

- v
- va
- vgrd
- v\_grd
- vwnd

**PV:**

- pv
- pvort
- epv

**Total precipitation:**

- tp
- prectot

## Variables from datasets

The following variables are defined in AtmoSwing for the different reanalyses. When no data ID is provided, one can use one of the variants define above.

### NCEP/NCAR Reanalysis 1

**For pressure levels:**

- Air Temperature
- Geopotential height
- Relative Humidity
- Specific Humidity
- Vertical Velocity
- U-Wind
- V-Wind

**For the surface:**

- Air Temperature
- Surface lifted index (1ftx)
- Best (4-layer) lifted index (1ftx4)
- Vertical velocity

- Potential temperature (pottmp)
- Precipitable water
- Pressure
- Relative humidity
- Sea level pressure
- U-wind
- V-wind

**For the surface fluxes:**

- Air Temperature 2m
- Potential evaporation rate (pevpr)
- Specific humidity at 2m (shum)
- Soil moisture (0-10cm) (soilw0-10)
- Soil moisture (10-200cm) (soilw10-200)
- Skin Temperature (sktmp)
- Temperature of 0-10cm layer (tmp0-10)
- Temperature of 10-200cm layer (tmp10-200)
- Temperature at 300cm (tmp300)
- U-wind at 10 m
- V-wind at 10 m
- Water equiv. of snow dept (weasd)
- Maximum temperature at 2m (tmax2m)
- Minimum temperature at 2m (tmin2m)
- Cloud forcing net longwave flux (cfnlw)
- Cloud forcing net solar flux (cfnsf)
- Convective precipitation rate (cprat)
- Clear sky downward longwave flux (csdlw)
- Clear sky downward solar flux (csdsf)
- Clear sky upward solar flux at surface (csusf)
- Downward longwave radiation flux (dlwrf)
- Downward solar radiation flux (dswrf)
- Ground heat flux (gflux)
- Latent heat net flux (lhtfl)
- Near IR beam downward solar flux (nbdsf)
- Near IR diffuse downward solar flux (nddsf)

- Net longwave radiation (nlwrs)
- Net shortwave radiation (nswrs)
- Precipitation rate
- Sensible heat net flux (shtfl)
- Momentum flux (zonal) (uflx)
- Zonal gravity wave stress (ugwd)
- Upward Longwave Radiation Flux (ulwrf)
- Upward Solar Radiation Flux (uswrf)
- Visible Beam Downward Solar Flux (vbdsf)
- Visible Diffuse Downward Solar Flux (vddsf)
- Momentum Flux, v-component (vflx)
- Meridional Gravity Wave Stress (vgwd)

## NCEP/DOE Reanalysis 2

### For pressure levels:

- Air Temperature
- Geopotential height
- Relative Humidity
- Vertical Velocity
- U-Wind
- V-Wind

### For the surface:

- Precipitable water
- Pressure
- Sea level pressure

### For the surface fluxes:

- Air Temperature 2m
- Specific humidity at 2m (shum)
- Maximum temperature at 2m (tmax2m)
- Minimum temperature at 2m (tmin2m)
- Skin Temperature (sktmp)
- Soil moisture (0-10cm) (soilw0-10)
- Soil moisture (10-200cm) (soilw10-200)
- Temperature of 0-10cm layer (tmp0-10)

- Temperature of 10-200cm layer (tmp10-200)
- U-wind at 10 m
- V-wind at 10 m
- Water equiv. of snow dept (weasd)
- Convective precipitation rate (cprat)
- Downward longwave radiation flux (dlwrf)
- Downward solar radiation flux (dswrf)
- Ground heat flux (gflux)
- Latent heat net flux (lhtfl)
- Potential evaporation rate (pevpr)
- Precipitation rate
- Sensible heat net flux (shtfl)
- Total cloud cover (tcdc)
- Momentum flux (zonal) (uf1x)
- Zonal gravity wave stress (ugwd)
- Upward Longwave Radiation Flux (ulwrf)
- Upward Solar Radiation Flux (uswrf)
- Momentum Flux (meridional) (vf1x)
- Meridional Gravity Wave Stress (vgwd)

## NCEP CFSR

**For pressure levels (pgbh):**

- Geopotential height
- Precipitable water
- Mean sea level pressure
- Relative humidity
- Temperature

## NCEP CFSR subset

### For pressure levels:

- Geopotential height
- Geopotential height anomaly (gpa)
- Relative humidity
- Specific humidity
- Temperature
- Vertical Velocity
- Eastward wind
- Northward wind
- Atmosphere horizontal velocity potential (vpot)
- 5-wave geopotential height (5wavh)
- 5-wave geopotential height anomaly (5wava)
- Atmosphere absolute vorticity (absv)
- Cloud water mixing ratio (clwmr)
- Atmosphere horizontal streamfunction (strm)

### For the total column:

- Relative humidity
- Cloud water (cwat)
- Atmosphere water vapor content (pwat)

## ECMWF ERA5

### For pressure levels:

- Divergence (d)
- Potential vorticity
- Specific humidity
- Relative humidity
- Temperature
- U component of wind
- V component of wind
- Vorticity (relative) (vo)
- Vertical velocity
- Geopotential

**For surface:**

- 2 meter dewpoint temperature (d2m)
- Sea level pressure (msl)
- Snow depth (sd)
- Sea surface temperature (sst)
- 2 meter temperature (t2m)
- Total column water (tcw)
- Total column water vapor (tcwv)
- 10 meter U wind component (u10)
- 10 meter V wind component (v10)
- Total precipitation (tp)
- Convective available potential energy (cape)
- Instantaneous moisture flux (ie)
- Surface net solar radiation (ssr)
- Surface solar radiation downwards (ssrd)
- Surface net thermal radiation (str)
- Surface thermal radiation downwards (strd)

**ECMWF ERA-interim****For pressure levels:**

- Divergence (d)
- Potential vorticity
- Specific humidity
- Relative humidity
- Temperature
- U component of wind
- V component of wind
- Vorticity (relative) (vo)
- Vertical velocity
- Geopotential

**For isentropic levels:**

- Divergence (d)
- Montgomery potential (mont)
- Pressure

- Potential vorticity
- Specific humidity
- U component of wind
- V component of wind
- Vorticity (relative) (vo)

**For surface:**

- 2 meter dewpoint temperature (d2m)
- Sea level pressure (msl)
- Snow depth (sd)
- Sea surface temperature (sst)
- 2 meter temperature (t2m)
- Total column water (tcw)
- Total column water vapor (tcwv)
- 10 meter U wind component (u10)
- 10 meter V wind component (v10)
- Total precipitation
- Convective available potential energy (cape)
- Instantaneous moisture flux (ie)
- Surface net solar radiation (ssr)
- Surface solar radiation downwards (ssrd)
- Surface net thermal radiation (str)
- Surface thermal radiation downwards (strd)

**For PV levels:**

- Pressure
- Potential temperature (pt)
- U component of wind
- V component of wind
- Geopotential



## ECMWF ERA 20th Century

### For pressure levels:

- Geopotential
- Temperature
- Relative humidity
- Vertical velocity

### For surface:

- Total column water
- Total precipitation
- Sea level pressure

## ECMWF Coupled ERA 20th Century

### For pressure levels:

- Geopotential
- Temperature
- Relative humidity
- Vertical velocity

### For surface:

- Total column water
- Total precipitation
- Sea level pressure

## ECMWF IFS

- Geopotential (z)
- Geopotential height (gh)
- Air temperature
- Vertical velocity
- Relative humidity
- Specific humidity
- U component of wind
- V component of wind
- Theta E (thetaE)
- Theta ES (thetaES)

- Total column water vapour (tcwv)
- Precipitable water

## NWS GFS

- Geopotential height
- Air temperature
- Vertical velocity
- Relative humidity
- U component of wind
- V component of wind
- Precipitable water

## NASA MERRA-2

For product 'inst6\_3d\_ana\_Np':

- Geopotential height
- Air temperature
- Sea-level pressure

## NASA MERRA-2 subset

For product 'inst6\_3d\_ana\_Np':

- Geopotential height
- Specific humidity
- Air temperature
- Sea-level pressure
- Eastward wind component
- Northward wind component
- Surface pressure (ps)

For product 'inst3\_3d\_asm\_Np':

- Ertel's potential vorticity
- Vertical pressure velocity
- Relative humidity after moist
- Sea level pressure
- Air temperature

**For product ‘inst1\_2d\_int\_Nx’:**

- Total precipitable ice water (tqi)
- Total precipitable liquid water (tql)
- Total precipitable water vapor (tqv)

**For product ‘inst1\_2d\_asm\_Nx’:**

- Total precipitable ice water (tqi)
- Total precipitable liquid water (tql)
- Total precipitable water vapor (tqv)
- 10-meter air temperature (t10m)

**For product ‘avg1\_2d\_flx\_Nx’:**

- Total surface precipitation flux

**For product ‘avg1\_2d\_lnd\_Nx’:**

- Total precipitation land; bias corrected

### **JMA JRA-55 subset**

**For pressure levels (anl\_p125):**

- Geopotential height
- Relative humidity
- Temperature
- Vertical velocity

**For the surface (anl\_surf125):**

- Pressure reduced to MSL

**For the total column (anl\_column125):**

- Precipitable water

**For isentropic levels:**

- Potential vorticity
- Geopotential Height

**For product ‘fcst\_phy2m125’:**

- Total precipitation 3h (tprat3h)
- Total precipitation 6h (tprat6h)

### JMA JRA-55C subset

**For pressure levels (anl\_p125):**

- Geopotential height
- Relative humidity
- Temperature
- Vertical velocity

**For the surface (anl\_surf125):**

- Pressure reduced to MSL

**For the total column (anl\_column125):**

- Precipitable water

**For isentropic levels:**

- Potential vorticity
- Geopotential Height

**For product 'fcst\_phy2m125':**

- Total precipitation 3h (tprat3h)
- Total precipitation 6h (tprat6h)

### NOAA 20CR v2c

**For pressure levels:**

- Air Temperature
- Geopotential
- Vertical velocity
- Relative Humidity
- Specific Humidity
- Specific Humidity
- U-Wind
- V-Wind

**For the surface:**

- Precipitable water
- Sea level pressure

**For surface fluxes:**

- Precipitation rate

## NOAA 20CR v2c ensemble

### For the product 'analysis':

- Sea level pressure
- Precipitable water
- Vertical velocity at 500 hPa (omega500)
- Relative Humidity at 850 hPa (rh850)
- Relative Humidity at the pressure level 0.995 times the surface pressure (rh9950)
- Air Temperature at 850 hPa (t850)
- Air Temperature at the pressure level 0.995 times the surface pressure (t9950)
- Geopotential height at 200 hPa (z200)
- Geopotential height at 500 hPa (z500)
- Geopotential height at 1000 hPa (z1000)

### For the product 'first\_guess':

- Precipitation rate

## NOAA OISST v2

There are no product/level definition for this dataset.

- Sea Surface Temperature (sst)
- Sea Surface Temperature Anomaly (sst\_anom)

## METEO-FRANCE ARPEGE

- Geopotential
- Relative humidity
- Total column water vapour
- Air temperature
- Vertical velocity

## Generic NetCDF

The generic dataset is different from the other datasets. There is no predefined variables. The variable name needs to be used in the NetCDF file.

## 1.6 Preprocessing routines

Some elaborated variables need to be preprocessed, such as the moisture index. Thus, different preprocessing routines are available in AtmoSwing.

The main preprocessing routines implemented are:

- Addition (Addition): add (point-wise) all the provided predictors.
- Average (Average): average (point-wise) all the provided predictors.
- Difference (Difference): process the difference (point-wise) between 2 predictor grids.
- Multiplication (Multiplication or Multiply): multiply (point-wise) all the provided predictors.
- Humidity (or moisture) index (HumidityIndex): multiply the relative humidity and the precipitable water
- Humidity (or moisture) flux (HumidityFlux): process the multiplication of the wind and the humidity index. Requires four predictors in the following order: 1) U wind, 2) V wind, 3) relative humidity and 4) precipitable water
- Wind speed (WindSpeed): process the wind speed with the provided U wind and V wind components.

The following preprocessing routines are usually not used directly (or are automatically handled by AtmoSwing):

- Simple gradients (SimpleGradients): processing of differences between adjacent grid points by ignoring the horizontal distance.
- Real gradients (RealGradients): processing of real gradients between adjacent grid points (using the horizontal distance). This preprocessing method is automatically used when the analogy criterion is S1.
- Simple gradients with Gaussian weights (SimpleGradientsWithGaussianWeights): same as before, but with a weighting of the spatial field by a gaussian 'hat'-shaped pattern.
- Real gradients with Gaussian weights (RealGradientsWithGaussianWeights): same as before, but with a weighting of the spatial field by a gaussian 'hat'-shaped pattern.
- Simple curvature (SimpleCurvature): processing of the 'curvature' between adjacent grid points by ignoring the horizontal distance.
- Real curvature (RealCurvature): processing of real 'curvature' between adjacent grid points (using the horizontal distance). This preprocessing method is automatically used when the analogy criterion is S2.
- Simple curvature with Gaussian weights (SimpleCurvatureWithGaussianWeights): same as before, but with a weighting of the spatial field by a gaussian 'hat'-shaped pattern.
- Real curvature with Gaussian weights (RealCurvatureWithGaussianWeights): same as before, but with a weighting of the spatial field by a gaussian 'hat'-shaped pattern.

## 1.7 Analogy criteria

The analogy criteria available in AtmoSwing are the following:

- Teweles-Wobus (S1)
- Teweles-Wobus with Gaussian weights (S1G)
- Teweles-Wobus on raw data (S0)
- Derivative of Teweles-Wobus (S2)
- Sum of absolute differences (SAD)
- Mean absolute difference (MD)
- Root mean square error (RMSE)
- Difference in mean value (nonspatial) (DMV)
- Difference in standard deviation (nonspatial) (DSD)

## 1.8 Versioning scheme and migration

### 1.8.1 Versioning scheme

AtmoSwing follows a versioning structure with version numbers as MAJOR.MINOR.PATCH. The rules are the following:

1. MAJOR version: significant changes in the overall structure or addition of a new tool,
2. MINOR version: addition of functionality that might be backward-incompatible,
3. PATCH version: backward-compatible bug fixes and small improvements.

### 1.8.2 Support Policy

When a new version is published, the development of previous versions is not continued, except for LTS (Long Term Support) releases, where bug fixes are provided for a longer period.

Version	Release	Bug Fixes Until
2.0	19.11.2018	01.07.2019
2.1	23.05.2019	01.07.2023
3.0 LTS	17.04.2023	01.07.2027

## 1.8.3 Upgrade Guide

### v2.1 to v3.0

#### Parameters files

- Changes to the predictors dataset\_id for the Forecaster. Check the new values under the [Forecaster data page](#).

#### Predictand DB

- The predictand DB has been updated to include the projection information. The DB must be generated again with the new version of the software.

#### Forecaster batch file

- The extension of the batch file has changed to .xml.

### v2.0 to v2.1

#### Parameters files

- Changes to the reanalyses dataset\_id. Check the new values under the [reanalyses page](#).

## 1.9 Release notes

The following release notes list the main changes to AtmoSwing. Many small changes are not listed and can be tracked down in the commits.

### 1.9.1 v3.0.7

20 May 2023

#### Fixed

- Fixing issue with 6-hourly forecasts.
- Fixing missing level value to display ARPEGE data.

### 1.9.2 v3.0.6

16 May 2023

#### Added

- New management of the parameters listing in the Forecaster.
- Adding access to the parameter information in the Forecaster.
- Adding menu entries for recent workspaces and batch files.
- Adding buttons to reset the plots zoom extent in the Viewer.

#### Changed

- Replacing icons with SVG versions.



**Fixed**

- Fixing proxy parsing.

### 1.9.3 v3.0.5

12 May 2023

**Fixed**

- The Forecaster returns a non 0 code when at least one forecast failed.
- Fixing run time step of Arpege.

### 1.9.4 v3.0.4

09 May 2023

**Fixed**

- Addition of missing translations related to vroomgis.

### 1.9.5 v3.0.3

08 May 2023

**Added**

- Adding the location of the target in the predictors mapping.
- Adding the batch file name in the Forecaster frame.

**Fixed**

- Fixing date issue in the predictors mapping.
- Translating the frame labels.

### 1.9.6 v3.0.2

18 Apr 2023

**Fixed**

- Fixing issue with the predictors mapping.

### 1.9.7 v3.0.1

17 Apr 2023

#### Fixed

- Fixing issue with the predictors mapping.

### 1.9.8 v3.0.0

17 Apr 2023

#### Added

- **A new frame has been added to visualize the predictor fields (from NWP models and reanalyses).**  
It allows comparing the predicted situation with the analog situations.
- Real support of the 6-hourly forecasts in the Viewer.
- Translation of all modules in French.
- Adding path to ecCodes definitions in the preferences.
- Docker: adding ecCodes definitions path to the environment variables.
- Adding column water vapor variable.
- Adding support and tests for the ARPEGE model outputs.
- Adding options in the workspace to limit the time series length to be displayed.

#### Changed

- Removing inefficient mini-batch assessment on full period.
- Improving mini-batches optimisation end.
- Changing Forecaster batch file extension to xml.
- Viewer workspaces now support relative paths for GIS layers.
- Using `std::runtime_error` instead of custom exception macro.
- Addition of predictor properties in the forecast result files.
- Renaming the operational predictor dataset IDs.
- Removing custom NaN management approach for C++ built-in approach.
- Removing parallel download as it is disliked by data providers.

#### Fixed

- Fixing CUDA options issues.
- Time units were missing in the predictand db.
- Fixing issue with the epoch number after restoration.
- Improving code based on Codacy analysis.
- Fixing a potential crash in the Optimizer (when the predictor levels and times vary).
- Fixing results csv export from plot.

- Fixing issues with 6-hourly forecasts.

### 1.9.9 v2.1.6

22 Nov 2022

#### Added

- A mini-batch approach has been implemented for Genetic Algorithms.
- Adding a warning if the preload option if not enabled for calibration.
- Adding local GFS dataset class to avoid downloading.

#### Changed

- Improvement of time units description in generated nc files.
- Removing usage of temporary storage in GAs optimization.
- Curl: disable certificate checks on Windows as not supported.
- Code formatting.

#### Removed

- Removing the history approach in GAs optimisations (not efficient).

### 1.9.10 v2.1.5

11 Oct 2022

#### Changed

- Moved the dependencies management to conan
- Simplified the areas management by removing the composite approach.
- Some code clean up and code formatting.
- Moved all CI workflows to GitHub actions.

#### Fixed

- The weights computed by the optimizer cannot take negative values.
- Addition of S0 and S1 with normalization by the reference value.
- GFS urls on nomads have been fixed.
- Optimizer: fixed convergence check when previous results are loaded.
- Optimizer: relaxed criteria for convergence in GAs (using tolerance).
- Optimizer: fixed an issue with latitude values  $> 90^\circ$  that were not corrected.

### 1.9.11 v2.1.4

09 Oct 2020

#### Changed

- Refactored time axis management in predictors loading.

#### Fixed

- Fixed an issue with missing dates in the FVG dataset.
- Fixed an issue with NaNs in the standardization.

### 1.9.12 v2.1.3

13 Jul 2020

#### Added

- Addition of Dockerfiles for the creation of Docker images.
- Addition of the mean and the sd in the parameters for standardisation.
- Standardisation for operational forecasting.
- Handling a single ‘number of analogs’ for operational forecasting.
- Handling forecasts without reference axis.
- Addition of the mean of the analogs to the aggregated forecast xml.
- Addition of the lead time hour to the aggregated results.
- Addition of synthetic text (csv) export of the forecast.
- Allowing a percentage of missing predictor data.

#### Changed

- Simplification of resulting parameters storage.
- Reduction of optimization memory footprint.
- Reduction of padding in structures to save memory.
- Disabling using GAs history by default.
- Changes in dependencies management.
- Refactoring predictor files download.
- Enabling optional downloading of predictor files.
- Standardisation after data loading.
- Changing output extension of forecast files to nc.
- Changing the specification of the forecast synthetic exports.
- Allowing NaNs to propagate in the forecast.
- Updating to Proj 7.
- Changing indentation back to 4.

**Fixed**

- Fixing GFS urls.
- Avoid crash and display an error when desired latitude was not found.
- Addition of the standardisation specification in resulting parameters.
- Fixing issue when the reference axis is NaN.
- Fixing lead times in plots.

**1.9.13 v2.1.2**

02 Dec 2019

**Added**

- Efficient GPU processing with CUDA.
- Ability to fully resume optimizations with GAs operators values.
- Parameters that were already assessed during the optimization are not assessed again.
- Parameters that are close to other parameters with poor skills are not assessed again.
- Addition of the Google benchmark framework for testing CUDA implementation.

**Changed**

- The optimization workflow has been simplified.
- Check for previous optimization convergence before loading data.
- Transitioning to Google code style.
- The dataset “NCEP\_Reanalysis\_v1” has been renamed to “NCEP\_R1”.
- The dataset “NCEP\_Reanalysis\_v2” has been renamed to “NCEP\_R2”.
- Some redundant tests have been removed.
- Addition of tests for the datasets.
- Changes in some libraries versions.

**Fixed**

- Error with weights in optimization initialization (when all weights = 0).
- Optimization resuming might have started from an older file.
- Some (rarely used) variables definition in some reanalyses were wrong.
- Fixed an issue with latitudes axis when resampling.

### 1.9.14 v2.1.1

17 Jul 2019

#### Added

- Addition of predictor data dumping to binary files (to reduce RAM usage).
- Allow loading from dumped predictor data (and keep in RAM).
- Option for replacing NaNs by -9999 (save processing time).
- Addition of a bash script to install libraries on a server.

#### Changed

- Refactoring of the CUDA implementation.
- Updating GFS urls.
- Improving Eigen usage.
- Testing downscaling parameters.
- Speeding up Grib files loading.
- Adding information to error messages.

#### Fixed

- Fixing Viewer issues to find the forecast files.
- Fixing missing node (on\_mean) when parsing xml files for the calibrator.
- Fixing a log path issue.
- Fixing a memory leak due to ecCodes index not deleted.
- Fixing a bug when interval days are not used.

### 1.9.15 v2.1.0

23 May 2019

#### Added

- Support of GRIB1 files with ecCodes.
- Adding a generic NetCDF predictor class.
- Addition of real gradients processing.
- Addition of S1 variants: S0 and S2.
- Addition of other nonspatial criteria.
- Support of IFS outputs.
- Addition of the ERA5 dataset.
- Addition of custom period definition (selection of months).
- Adding analog dates-only extraction method.
- Get preloaded data from another level if not available.

- Adding options for seasons definition.
- Addition of 2D Gauss function for predictor weighting.
- Implementing time dimension for grib files.
- Addition of lightnings data normalization.
- Adding on-the-fly standardization.
- Support non trivial time arrays for data loading and missing files.
- Supporting more complex predictor hours.

### **Changed**

- Migrating from g2clib to ecCodes.
- Improving support for NaNs.
- Handling resulting files with many stations.
- Allow Optimizer to not have a validation period.
- Allow for negative MJD values.
- Allow for retention of more analogs than the defined number.
- Adding common variables definitions between predictors.
- Addition of new reanalyses variables.
- Allowing different file structures for ERA-interim.
- Using more C++ features.
- Improving GRIB parsing efficiency.
- Heavy refactoring of the time arrays.
- Adding command-line logo.
- Updating the predictand DB tool GUI.
- Better management of missing files.
- Getting rid of the pseudo-normalized criteria.
- Refactoring the management of the time reference.
- Removing custom exception class.
- Logs refactoring.
- Removing call to Forecaster from the Viewer.
- Improving use of config.
- Auto generate the dependencies for Debian packages.

### **Fixed**

- Fixing minimum domain size for S1 and S2.
- Fixing time issue with the 6-hrly time step.
- Fix an issue related to GAs crossover on the criteria.

- Fixing issue with a Google layer projection.
- Fix a bug in multithreaded downloads.
- Fix command line usage of builds with GUIs.

### 1.9.16 v2.0.1

12 Dec 2018

#### Added

- Adding definition of a continuous validation period.

#### Changed

- Using https on nomads.
- Setting the installer creation as optional.

#### Fixed

- Fixing About panel size and Ubuntu dependencies.
- Fixing CMake issues.

### 1.9.17 v2.0.0

19 Nov 2018

#### Added

- Addition (merge) of the code of the optimization with genetic algorithms to the main repository.
- Creation of the Downscaler.
- Addition of the NOAA 20CR-v2c ensemble dataset.
- Addition of the CERA-20C dataset.
- Addition of the CMIP5 dataset.
- Addition of CORDEX data
- Transforming geopotential into geopotential height.
- Adding other MTW time steps.
- Adding an option to specify different time steps for the calibration / archive periods.
- Adding a time properties to take into account temporal shift in the predictand.
- Handling of both 3h and 6h ERA-20C dataset.
- Specification of the number of members in the parameters file.
- Adding an option to remove duplicate date from members.
- GFS urls are now configurables.
- Getting predictor time step from files.



- Getting the spatial resolution from file.
- Adding capacity to read some unregistered predictor variables.
- Adding GAs presets.

### **Changed**

- Code moved to GitHub.
- Adding continuous integration (Travis CI and AppVeyor).
- Adding code coverage of the tests.
- New MSI installer with WiX.
- Getting some libraries through external projects.
- Simplification of the CRPS calculation.
- Speeding up data loading.
- Adding possibility to skip data normalization.
- Removing the slow coefficient approach in criteria calculation.
- Removing the slower processing version.
- Heavy refactoring to simplify class names.
- Refactoring parameters files.
- Refactoring processor code.
- Reduce time for assessing the number of analogues.
- Improving parameters file parsing.
- Fix a bug when transforming Geopotential variable.
- Better acceptance of NaNs in the predictand values.
- Using initialization lists everywhere.
- CMake files are now organized by folder.
- Improving Forecaster messages.
- Changing the predictor files listing approach.
- New predictor area implementation.
- Improving and simplifying GUIs.
- The predictand DB build tool is accessible from anywhere.
- Stopping the calculations when there is not enough potential analogs.
- Limit the relevance map extension.
- Allowing the duplicate dates by default.
- Defaulting to 1 member.
- Saving results from Monte Carlo analysis.

### **Fixed**

- Fix archive length check with ensemble datasets.
- Fixing an issue of grid resolution when loading data.
- Fix issues with VS.
- Fixing link issues with cURL on Linux.
- Fixing new GFS files parsing.
- Fix compiler warnings under Windows.
- Correctly closing grib files.
- Fixing screen resolution issue on Linux.
- Adding missing CL help entries.
- Force unlock weights when  $\text{sum} > 1$ .
- Fixing Monte Carlo analysis.
- Fixing background color.

### 1.9.18 v1.5.0

30 Jan 2017

#### Added

- Addition of the CFSR v2 dataset.
- Addition of the MERRA2 dataset.
- Addition of the JRA-55 subset data.
- Addition of the JRA-55C subset.
- Addition of the 20CR v2c dataset.
- Addition of the ERA-20C dataset.
- Allow for both relative and absolute paths for predictors.
- Addition of the possibility to define the station id as parameter.
- Addition of the addition preprocessing.
- Addition of the average preprocessing.
- Addition of the Monte-Carlo approach from the Optimizer.

#### Changed

- Refactoring predictor data classes.
- Addition of support for the T382 truncature.
- Renaming level type to product.
- Split up of the big CMake file in smaller files.
- Allowing preload of humidity index data.
- Testing and improving preprocessing methods.

- Improving preprocessing parameters handling.
- Refactoring parameters loading.
- Addition of a tolerance in the area matching.
- Refactoring Classic Calibration.
- Refactoring saving and loading results.
- Addition of compression to optimizer results.
- Improving handling of Gaussian grids in the classic calibration.
- Saving both results details of calibration and validation.
- Predictor file paths can now contain wildcards!
- Refactoring logging.
- Improvement of the predictor files lookup.
- Changes in the “Classic +” method.
- Better handling of intermediate resulting files.
- Improving predictor datasets reading.

#### **Fixed**

- Fix of a bug when the area is 2 points wide.
- Fix of a bug for regular and irregular grids.
- Fix of a minor memory leak.
- Fix some issues related to new predictors.
- Fix loading of previous runs in the Optimizer.
- Fix of an issue of precision when looking for time values in an array.

### **1.9.19 v1.4.3**

21 Jul 2016

#### **Added**

- The new NCEP R1 archive format is now supported.
- Preloading of multiple data IDs.
- Addition of predictor data loading with threads.
- Handling null pointers in the preloaded data.
- Adding normalized criteria.
- Sharing data pointers across analogy steps.
- Addition of ERA-interim.
- Improving notifications when loading failed.
- NCEP R2 tested.

### Changed

- Renaming Calibrator into Optimizer.
- Parsing NaNs as string to handle non-numerical cases for predictands.
- Migrating from UnitTest++ to Google Test.
- Skip gradients preprocessing when there are multiple criteria.
- Using pointers to the parameter sets in order to keep changes in level selection.
- Replacing ERA40 by ERA-interim.
- Changes in the reanalysis datasets management.
- Simplification of the meteorological parameter types.
- Significant changes in netcdf files loading.
- Addition of a functionality in the composite areas in order to handle the row lon = 360 = 0 degrees.
- Addition of a method to remove duplicate row on multiple composites.
- New management of predictor data for realtime series.
- Using Grib2c instead of GDAL for Grib files, and data loading refactoring.

### Fixed

- Fixed unit test issues.
- Applying code inspection recommendations.
- Fix of a segmentation fault in the optimizer.

## 1.9.20 v1.4.2

18 Aug 2015

### Added

- Addition of the 300hPa level for GFS.
- Highlight the optimal method for the station in the distribution plot and the analogs list.

### Changed

- Newly created workspace now opens automatically.
- Do not load already loaded forecasts.
- Do not display the station height when null.
- Handle file opening when double-clicking.
- Improving CL usage.
- Reload forecasts previously processed if an export is required.

### Fixed

- Removal of a forecast from the GUI now works as expected.
- Removing Projection specification from WMS files.

- Past forecasts do load again.
- Fix of a bug in data preloading.

### 1.9.21 v1.4.1

19 Mar 2015

#### Added

- Addition of the export options to the command line configuration.
- Addition of an overall progress display.

#### Changed

- No need to reload forecasts after processing.
- Improving the display of former forecast files.
- TreeCtrl images of different resolutions.
- Change of every image/icon for a custom design.
- Full support implemented for high resolution screens.
- Updating the command line interface.
- Forcing UTF-8 in the netCDF files.
- Changing file version specification into major.minor

#### Fixed

- Debugging accents issue under Linux.
- Removing « using namespace std » in order to keep a clean namespace resolution.
- Removing asFrameXmlEditor.
- Fix of a crash when no forecast is opened.
- Replacing printf with wxPrintf.
- Removing unnecessary .c\_str() conversion on strings.
- Fix of a corruption in the wxFormbuilder project.
- Debugging netcdf issues under Linux.
- Fixing namespace issues.

### 1.9.22 v1.4.0

02 Mar 2015

#### Added

- Addition of the export of a synthetic xml file.
- Addition of a tree control for the forecasts in the viewer.
- Addition of an automatic methods aggregation in the viewer.

- Creation of methods IDs.
- Specification of the station IDs for specific parameters files.
- New xml format for most files.

#### Changed

- Update to the new GFS URLs and format.
- Adding a message in order to specify the selected models.
- Removal of the coordinate system specification for the predictors.
- Removing TiCPP in order to use the native xml handling from wxWidgets.
- Merging asCatalog and asCatalogPredictands.

#### Fixed

- Cleaning up processing and use of quantiles.
- No need to reload forecasts after processing.

### 1.9.23 v1.3.3

12 Jan 2015

#### Added

- Addition of buttons in the viewer frame to go backward and forward in time.
- Workspaces can now be saved to an xml file.
- Addition of a wizard to create the workspace.
- Addition of a control on the changes of the workspace to save before closing.
- Addition of a configuration option in the forecaster.

#### Changed

- Separation of the preferences.
- Definition of the preferences in the workspace.
- Change of the configuration option by using a given batch file.
- The loading of predictor data has significantly changed.
- Better handles user errors in the parameters files.
- Hide the elevation information when not available.
- Changing the name of the U/V axis into X/Y to help users.

#### Fixed

- Cleanup of the forecaster config options.
- Cleanup of the calibrator config options.
- Correction of the path to the WMS layers.
- Bug fix of unspecified directories for the archive predictors.

- Limiting the number of parallel downloads.
- Fix of the cURL hang with parallel downloading.
- Removal of the definition of the analogs number on the forecast score.
- Fix of an issue with the colors storage in the workspace.
- Now keeps the same model selection when opening new forecasts.
- Now keeps the same lead time when opening new forecasts.

### 1.9.24 v1.3.2

01 Sep 2014

#### Added

- Introduction of workspaces for the viewer.
- Addition of WMS basemaps layers.
- Merging the two viewer frames into one with a new lead time switcher.
- Addition of the ability to optimize on multiple time series together.
- Addition of the CRPS reliability skill score and removal of F0 loading methods.

#### Changed

- Improvement of the rank histogram with bootstrapping.
- Increase of bootstrapping to 10'000 for the rank histogram.
- Reduction in time for the assessment of all scores.
- Improving performance by reducing reallocation.
- Changing the MergeCouplesAndMultiply method into FormerHumidityIndex.

#### Fixed

- Fix of the paths for CUDA files.
- Fix of a linking issue with the viewer.
- Fix of a bug related to gradient preprocessing in validation.
- Minor bug fix on the evaluation of all forecasting scores.
- Removing of the S1 weighting method.
- Bug fix in the preloading option for the classic calibration parameters.
- Fix of a bug on the single instance checker.
- Limitation of the zoom level to avoid the memory issue related to GDAL caching mechanism.

### 1.9.25 v1.3.1

13 May 2014

#### Changed

- Merge of the various CMake files into one project.

#### Fixed

- Debugging the new build process under Linux.

### 1.9.26 v1.3.0

06 May 2014

#### Added

- Implementation of GPU processing
- Addition of a predictand pattern file.
- Addition of compression to the forecast files.
- Addition of CPack files.
- Addition of a unit test on gradients preprocessing.

#### Changed

- The archive and calibration periods can now start in the middle of a year.
- Better check the requested time limits when loading predictor data.

#### Fixed

- Removing a memory leak when aborting the app initialization.
- Correction of the data ordering in the forecast results.
- Bug fix in the time arrays intervals construction.
- Fix of a bug in the validation processing with a partial final year.
- Correction of the rank histogram.
- Reduced cURL outputs and fix of the multithreaded downloads.
- Adding a missing MSVC dll in the installation package.

### 1.9.27 v1.2.0

26 Feb 2014

#### Added

- Addition of the Calibrator source code.
- Addition of the rank histogram (Talagrand diagram)
- Addition of CRPS decomposition after Hersbach (2000).
- Addition of the generation of xml parameters files after the calibration.



### Changed

- The catalogs were removed for the predictors classes and new specific data classes were generated.
- Removing predictand database properties from parameters for calibration.
- Changing predictors file names.
- Changes in unit test filenames for more clarity.
- Better initialization of the scrolled window.
- Check fields in the parameters file of the forecaster and the calibrator.
- Change of the version message in CL.

### Fixed

- Fix of format errors in the GFS urls.
- Fix of an issue related to preprocessed predictors.
- Logging of the url was discarded due to formatting issues leading to crashes.
- Correction of bugs related to unit tests from the calibrator.
- Fix of errors related to Eigen vectors.
- Minor memory leaks were removed.
- Removal of compilation warnings.
- Casing fixed in the netCDF files.
- The logging in unit tests was improved.
- Fix of display issues in the sidebar.
- Simplification of the time management.
- Fix of errors related to optional parameters.
- Removal of false warnings.
- Resolving some unit tests failures.
- The precipitation predictand class has been secured for RowMajor and Colmajor.
- Removing the exhaustive calibration.
- Removal of intermediate results printing.

## 1.9.28 v1.1.0

30 Oct 2013

### Added

- The predictand DB is now generalized to data other than precipitation.
- The Forecaster is now working with various predictands.
- Addition of the option to build the Forecaster in CL without GUI.

### Changed

- Improvement of the CMake build process.
- Better management of the NaNs during processing.
- Significant changes in order to generalize the predictand DB class.

#### **Fixed**

- Fix of a change in GDAL regarding GRIB2 files origin.
- Changing the order of includes in the asFileNetcdf class.
- Unwanted slashes in paths under Linux were removed.
- The viewer is now building again.
- Fix of some bugs in unit tests.

### **1.9.29 v1.0.3**

18 Oct 2012

#### **Added**

- Addition of data preloading functionality and data pointer sharing
- Preprocessing of the humidity flux and other variables combination.
- Addition of multithreading in the 2nd and following levels of analogy.
- Addition of functionalities to the threads manager.
- Handling of the NCEP reanalysis 2 dataset.
- Handling of the NOAA OI-SST dataset and addition of adapted criteria.
- Addition of the possibility to account for an axis shift in the predictor dataset.
- Addition of the others predictand and creation of a generic instance function.
- Addition of an option to stop calculation when there is NaN in data.
- Addition of bad allocation catching.

#### **Changed**

- Some clean-up of unused code.
- Simplification of the file names of intermediate results.
- Better management of the threads.
- Improvement of the multithreading option management.
- Better clean-up after processing.
- Addition of typedefs.
- Creation of 2 separate log files for the viewer and the forecaster.
- Improvement of the CMake files.
- Small improvements to the time series plots.
- Insertion of many supplementary assertions.

- Clean-up of config paths default values.

**Fixed**

- Addition of a critical section on the config pointer.
- Addition of critical sections for TiCPP.
- Addition of critical sections for NetCDF.
- Coordinates automatic fix was bugged in the parameters class.
- Fix of a bug when trying to sort array with size of 1.
- Bug fix in temporary file names creation.
- Bug fixed in the enumeration of units
- NetCDF file class may have badly estimated the array size.
- Fix of memory filling by logging in the time array class.

**1.9.30 v1.0.2**

27 Jan 2012

**Added**

- Better control of the log targets in the command-line mode.

**Changed**

- Data import from netCDF files is less sensitive to the data type.
- Much faster import of forecast files.

**Fixed**

- The forecasting launch from the viewer has been fixed.
- Removal of the message box in the CL forecast.

**1.9.31 v1.0.1**

22 Nov 2011

**Added**

- Export of forecast text files from the time series plot.
- Possibility to cancel the current forecast processing.

**Changed**

- Faster check of previously existing forecast files: load predictand DB only when needed.
- Change from q30 to q20 in the precipitation distribution
- Display of the considered quantile and return period for the alarms panel
- Better frame size restoration with maximization detection.

**Fixed**

- An error in the proxy port was fixed.
- Preference « /Plot/PastDaysNb » was sometimes 3 or 5. Set 3 everywhere.
- Do not load the same past forecasts twice in time series plots.

### 1.9.32 v1.0.0

09 Nov 2011

First official release

## THE FORECASTER

This section contains the documentation for AtmoSwing Forecaster. This document can also be downloaded as a pdf or epub using the bottom left menu.

The Forecaster manual is divided into the following topics:

### 2.1 Usage

The Forecaster allows processing an analog method based on the latest available NWP outputs or for a given date. The real-time forecast calculation can be performed by a standard computer without specific resources. The forecast can be executed on demand through a graphical user interface (GUI) or can be automated through the use of a command-line interface (CLI). Both aspects are explained hereafter.

By default, the Forecaster processes the latest available forecast. However, it can also process a forecast for a given day in the past, provided the NWP outputs are available locally or remotely, or it can be executed to process the last x days. These options can also be executed once a day by a task manager to fill eventual gaps in the previous days. If the forecasts are already present, no computing resources are used.

A batch file can be provided to the GUI or the CLI. It contains the data and export paths as well as the analog methods to be applied (defined themselves in the xml parameter files). Automatic tasks can thus execute the Forecaster successively with different options, for example, for different regions. The results can be saved in different directories, or the same ones.

#### 2.1.1 Requirements

The Forecaster needs:

- *A predictand database*
- *Operational predictors from NWP outputs*
- *A predictors archive (e.g. a reanalysis dataset)*
- *Parameters files defining the properties of the method.*

## 2.1.2 Outputs

The Forecaster produces compressed **NetCDF files** containing:

- The values of the predictand for the different lead times
- The analog dates
- The values of the analogy criteria
- The target dates (lead times)
- The number of analogs for the different lead times
- Some reference values (e.g., precipitation for different return periods)
- Some station metadata (ID, name, coordinates, height)
- Some information about the predictors used along with the spatial windows

There is one file per variant of the analog method containing data for all stations of the database.

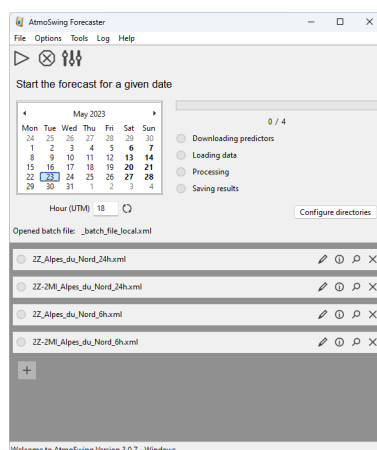
Some **exports** can also be generated (e.g. synthetic xml or csv files) to ease the integration of synthetic data on a web platform, for example.

The files are saved in a date-based directory structure (for example, 2019/04/23). The Viewer follows the same rules to look for new forecasts automatically. The output directory can be synchronized by means of a file-sharing client to distribute the forecasts.




## 2.1.3 Graphical user interface

This tool allows to do the following actions:



- Process the latest forecast or for a given date
- Define the list of methods to be performed automatically (also in command line mode)






The toolbar allows the following actions:



-  Run the forecast for the chosen date and time.
-  Stop the current calculations.
-  Define the preferences.





To perform a forecast, one must:

1. Choose the date from the calendar and the time below. When the software is started, the date and time are set to the current values. The  icon allows updating to the current values.
2. Start the forecast calculation by clicking on the  icon in the toolbar.

The processed analog methods are those listed in the lower half of the interface. The methods can differ in terms of structure or parameters, for example, to be adapted for a subregion. The methods are executed one after the other. A  icon means that the model is being processed,  that the calculations have been successfully performed and  that the forecast has failed for this method.

### Define the list of methods

The list of analog methods can be completed, or methods can be deleted. A method here is a specific parameterization of an analog method optimal for a lead time or a region. It is represented in the graphical user interface by choice of a parameters file. An entry can be removed with the  icon, and new ones can be added using the  icon below the list.

The entries can be edited by clicking on the  icon. A dialog box appears to set the parameters file name (only the name, not the full path). The parameters file must be located in the directory defined in the preferences. It is an XML file that can be edited with a text editor. The structure of the file is described in the [parameters file section](#). If the file is not found, the  icon is shown. The  icon allows displaying the description of the parameters file in a tooltip. The  icon allows displaying the content of the parameters file in a new frame (the content cannot be edited).

**Warning:** When the list of methods has been modified and should be kept as default, it is necessary to save it (menu 'File / Save batch file'); otherwise, the list will be reset at the software restart.

### 2.1.4 Command line interface

The Forecaster also has a command-line interface, which makes it possible to automate forecasts on a server. A scheduled task can then be added on a server (ex: Task Scheduler on Windows or Cron task on Linux). The options are as follows:

- h, --help**                Displays the help of command-line options
- c, --config**            Configure the forecaster
- v, --version**           Displays the software version
- f, --batch-file= <file>** Batch file to use for the forecast (full path)
- n, --forecast-now**      Run forecast for the latest available data
- p, --forecast-past= <nb\_days>** Run forecast for the given number of past days
- d, --forecast-date= <YYYYMMDDHH>** Run forecast for a specified date
- l, --log-level= <n>**    Set the log level (0: minimum, 1: errors, 2: warnings (default), 3: verbose)

`--proxy=<host[:port]>` Use proxy on given port  
`--proxy-user=<user[:password]>` Proxy user and password

### 2.1.5 Docker image

A Docker image is available on DockerHub: <https://hub.docker.com/r/atmoswing/forecaster>

Get it with: `docker pull atmoswing/forecaster`

The docker container for AtmoSwing Forecaster uses the same options than the *command line interface* (to the exception of the `--config` option). However, different directories need to be mounted in the docker container to allow AtmoSwing accessing the data and saving outputs.

The recommended way to use the docker container is to create a docker-compose file. An example is provided below:

```
version: "3"
services:
  atmoswing_forecaster:
    container_name: forecaster
    image: "atmoswing/forecaster:latest"
    working_dir: /app/home
    user: "1000:1000"
    volumes:
      - /home/atmoswing/:/app/mount/
      - /home/atmoswing/scratch/home:/app/home/
      - /home/atmoswing/scratch/tmp:/tmp/
    command: -f /app/mount/params/batch-file.xml -n
    environment:
      HOME: /app/home
      ECCODES_DEFINITION_PATH: /usr/share/eccodes/definitions
    network_mode: bridge
```

Then, the batch file needs to contain the mounted directories in the docker container. If you changed the target directories above, you need to adapt them below as well. The batch file should look like:

```
<?xml version="1.0" encoding="UTF-8"?>
  <atmoswing version="1.0" target="forecaster">
    <forecasts_output_directory>/app/mount/forecasts</forecasts_output_directory>
    <exports_output_directory>/app/mount/forecasts</exports_output_directory>
    <parameters_files_directory>/app/mount/params</parameters_files_directory>
    <predictors_archive_directory>/app/mount/predictors/archive/NR1</predictors_
↵archive_directory>
    <predictors_realtime_directory>/app/mount/predictors/realtime/GFS</predictors_
↵realtime_directory>
    <predictand_db_directory>/app/mount/predictands</predictand_db_directory>
    <export_synthetic_xml>0</export_synthetic_xml>
    <forecasts>
      <filename>2Z_Alpes_du_Nord_24h.xml</filename>
      <filename>2Z-2MI_Alpes_du_Nord_24h.xml</filename>
```

(continues on next page)



(continued from previous page)

```
<filename>2Z_Alpes_du_Nord_6h.xml</filename>
<filename>2Z-2MI_Alpes_du_Nord_6h.xml</filename>
</forecasts>
</atmoswing>
```

## 2.2 Preferences

The preferences frame can be opened from the menu Options / Preferences of the main interface or with a button from the toolbar. The options related to the batch file are saved in the batch file itself, while the other options are saved in the user preferences file. The preferences are saved when the software is closed. The preferences are automatically loaded when the software is started. AtmoSwing will first look for a preference file in the same directory as the executable, then in the user directory.

### 2.2.1 Batch file properties

Different directories must be specified for operational forecasting:

- Directory containing the parameters files: contains the *xml parameters files (defining by the method parameters)*
- Directory containing the predictand DB: contains the *NetCDF file generated from the predictand data*.
- Directory containing the archive predictors: predictor describing the candidate situations (for example *reanalyses*).
- Directory to save downloaded real-time predictors: the operational NWP model outputs. Subdirectories are automatically created for each date according to a year/month/day structure (respectively \ on Windows).
- Directory to save the forecast outputs (NetCDF): location where the results are stored. Subdirectories are also generated for each date.
- Directory to save the forecast exports (xml/csv): location where the synthesis xml file is saved.

In addition, the export format can be selected (for example, full XML or small CSV).

Directory containing the parameters files

Directory containing the predictand DB

Directory containing archive predictors

Directory to save downloaded real-time predictors (GCM forecasts)

Directory to save forecast outputs (netCDF)

Directory to save forecast exports (xml)

Export: Full XML

## 2.2.2 General options

### Language

The language can be changed from the drop-down list. The change is effective after restarting the software.

Language

English

Restart AtmoSwing for the change to take effect.

### Log options

Log preferences control the display of the log window and the level (importance) of reported messages:

1. Errors only - recommended as long as everything goes as planned.
2. Errors and warnings - some warnings help to be vigilant in case of problems. However, there might be warnings even during regular operation.
3. Verbose - all operations are reported, which is an unnecessary amount of information during regular operation. As a result, important messages might be hidden in a large number of messages. This option is only useful to identify where AtmoSwing is failing.

Logs

☒ Errors only (recommended) ☐ Display window

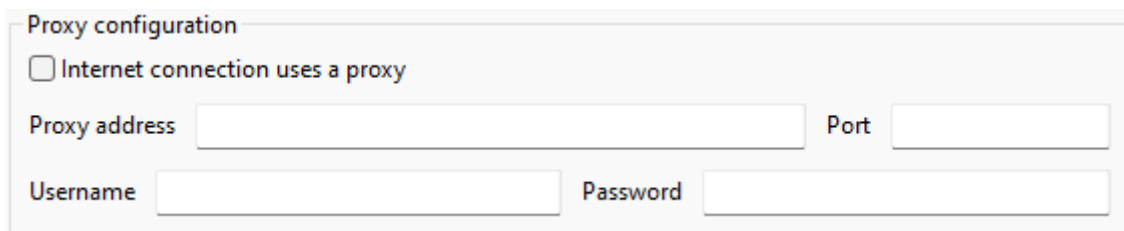
☐ Errors and warnings ☒ Save to a file

☐ Verbose

## Internet proxy

The proxy configuration contains the following elements:

- Proxy activation - if necessary.
- Proxy properties - the proxy address, port, the user name, and password.



Proxy configuration

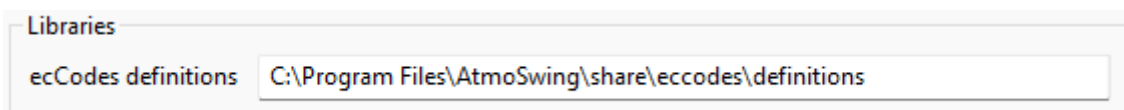
☐ Internet connection uses a proxy

Proxy address  Port

Username  Password

## Libraries options

The path to the ecCodes definitions can be specified. These files are used to decode the GRIB files. The default path is automatically set during installation.



Libraries

ecCodes definitions

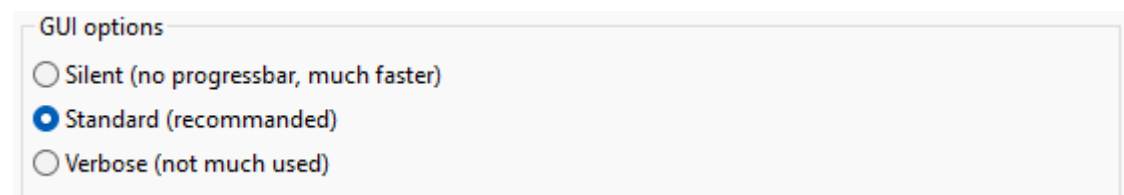
### 2.2.3 Advanced options

The advanced options are automatically configured and usually do not need to be changed.

#### Graphical User Interface (GUI) options

Options to control the behavior of the interface and its means of communication with the user. Saves calculation time by disabling dialog boxes. The choices are:

- Silent - no dialog box is displayed. The user must, therefore, keep an eye on the log window to detect any error. This option is only useful for very intensive calculations.
- Standard - recommended for regular use.
- Verbose - a bit more verbose than the Standard option.



GUI options

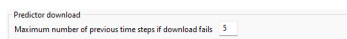
☐ Silent (no progressbar, much faster)

☒ Standard (recommended)

☐ Verbose (not much used)

## Predictor download

Options for automatic predictor download. The “maximum number of previous time steps if the download fails” corresponds to the number of attempts to retrieve data issued earlier if the desired data cannot be downloaded; this number must be greater than 2 to ensure proper operation of the Forecaster. Indeed, the latest NWP outputs are never instantly accessible, but after a certain delay. In this case, the most recent available data must be used.



## Advanced options

Different options can control the software’s behavior:

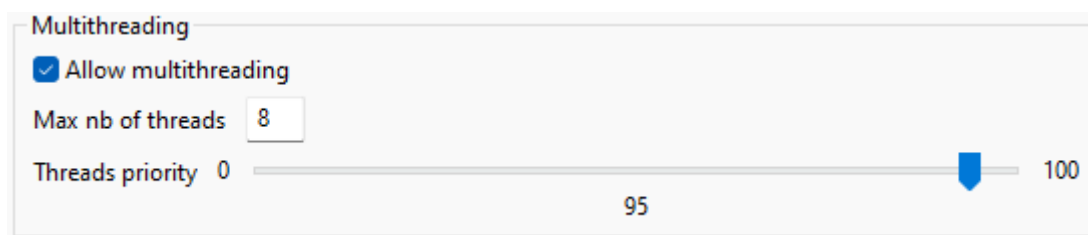
- Enable or disable the reactivity of the graphical interface. This option must always be enabled for regular use.
- Allow multiple instances of the Forecaster.



## Multithreading

Options for using multiple threads:

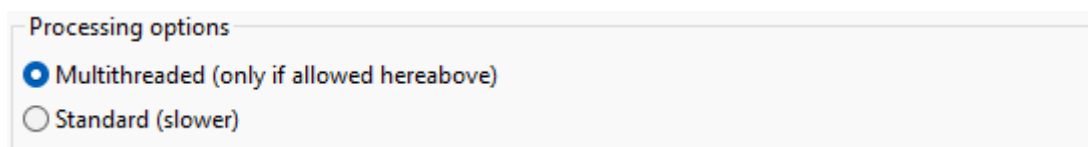
- Allow multithreading - allows a gain in performance for the calculations (recommended).
- Maximum number of threads - to be chosen according to the computer’s CPU. The software automatically proposes the optimal number of threads for the computer in use.
- Threads priority - allows defining the priority of the threads in competition with other ongoing processes.



## Processing options

These options influence how the search for analogs is performed, which has no impact on the results, only on the calculation time. The choices are as follows:

- Multithreaded - as long as the computer has at least two processors. It is the fastest version.
- Standard - this option is slower than the previous one as it uses a single thread.



## User specific paths

Some paths automatically defined according to the OS: working directory, log file, and preferences file.

### User specific paths

```
User working directory: C:\Users\Pascal\AppData\Roaming\AtmoSwing\
Log file:               C:\Users\Pascal\AppData\Local\Temp\AtmoSwingForecaster.log
Preferences file:       C:\Users\Pascal\AppData\Roaming\AtmoSwing\AtmoSwingForecaster.ini
```

## 2.3 Forecaster parameters file

The parameters files for the Forecaster are based on the *generic parameters file for AtmoSwing*, but has some specific elements for the operational forecasting as detailed below. A full example is provided at the bottom of the page.

### 2.3.1 Root node

The target property of the root node needs to be forecaster.

```
<atmoswing version="1.0" target="forecaster">
```

### 2.3.2 Content 'time\_properties'

The `<time_properties>` property has no `<time_array_target>` element, but a `<lead_time_days>` element listing the lead times (in days) to process.

```
<time_properties>
...
<lead_time>
  <lead_time_days>0,1,2,3,4,5</lead_time_days>
</lead_time>
...
</time_properties>
```

### 2.3.3 Content 'analog\_dates'

The content of `<analog_dates>` is mostly similar to the basic parameters file structure. The differences are:

- `<analog_number>`: The number of analogs to extract is not a single number as in the generic parameters file, but one value has to be provided for each lead time. The number of analogs to extract can thus increase with the lead time to take into account the increasing uncertainty of the NWP outputs.

For the operational forecasting, two predictor datasets are needed: the NWP outputs and the archive dataset:

- <realtime\_dataset\_id>: NWP outputs dataset
- <realtime\_data\_id>: Name of the variable in the NWP outputs dataset
- <archive\_dataset\_id>: Archive dataset
- <archive\_data\_id>: Name of the variable in the archive dataset

```
<analog_dates>
  <analog_number>50,50,70,110,230,490</analog_number>
  <predictor>
    <realtime_dataset_id>NWS_GFS_Forecast</realtime_dataset_id>
    <realtime_data_id>hgt</realtime_data_id>
    <archive_dataset_id>NCEP_Reanalysis_v1</archive_dataset_id>
    <archive_data_id>hgt</archive_data_id>
    ...
  </predictor>
  ...
</analog_dates>
```

### 2.3.4 Full example

Example of a parameters file for the Forecaster:

```
<?xml version="1.0" encoding="UTF-8" ?>
<atmoswing version="1.0" target="forecaster">
  <description>
    <method_id>PC-AZ4o-AHI2o</method_id>
    <method_id_display>Humidité optimisée</method_id_display>
    <specific_tag>Cretes_sud</specific_tag>
    <specific_tag_display>Crêtes du sud</specific_tag_display>
    <description>Analogie de circulation atmopshérique humidité optimisée</
↪description>
  </description>
  <time_properties>
    <archive_period>
      <start_year>1961</start_year>
      <end_year>2008</end_year>
    </archive_period>
    <lead_time>
      <lead_time_days>0,1,2,3,4,5</lead_time_days>
    </lead_time>
    <time_step>24</time_step>
    <time_array_analogs>
      <time_array>days_interval</time_array>
      <interval_days>60</interval_days>
      <exclude_days>60</exclude_days>
    </time_array_analogs>
  </time_properties>
  <analog_dates>
    <analog_number>50,50,70,110,230,490</analog_number>
```

(continues on next page)

(continued from previous page)

```

<predictor>
  <realtime_dataset_id>NWS_GFS_Forecast</realtime_dataset_id>
  <realtime_data_id>hgt</realtime_data_id>
  <archive_dataset_id>NCEP_Reanalysis_v1</archive_dataset_id>
  <archive_data_id>hgt</archive_data_id>
  <level>400</level>
  <time>18</time>
  <spatial_window>
    <grid_type>regular</grid_type>
    <x_min>2.5</x_min>
    <x_points_nb>5</x_points_nb>
    <x_step>2.5</x_step>
    <y_min>42.5</y_min>
    <y_points_nb>4</y_points_nb>
    <y_step>2.5</y_step>
  </spatial_window>
  <criteria>S1</criteria>
  <weight>0.33</weight>
</predictor>
<predictor>
  <realtime_dataset_id>NWS_GFS_Forecast</realtime_dataset_id>
  <realtime_data_id>hgt</realtime_data_id>
  <archive_dataset_id>NCEP_Reanalysis_v1</archive_dataset_id>
  <archive_data_id>hgt</archive_data_id>
  <level>700</level>
  <time>24</time>
  <spatial_window>
    <grid_type>regular</grid_type>
    <x_min>2.5</x_min>
    <x_points_nb>6</x_points_nb>
    <x_step>2.5</x_step>
    <y_min>42.5</y_min>
    <y_points_nb>2</y_points_nb>
    <y_step>2.5</y_step>
  </spatial_window>
  <criteria>S1</criteria>
  <weight>0.21</weight>
</predictor>
<predictor>
  <realtime_dataset_id>NWS_GFS_Forecast</realtime_dataset_id>
  <realtime_data_id>hgt</realtime_data_id>
  <archive_dataset_id>NCEP_Reanalysis_v1</archive_dataset_id>
  <archive_data_id>hgt</archive_data_id>
  <level>850</level>
  <time>12</time>
  <spatial_window>
    <grid_type>regular</grid_type>
    <x_min>-2.5</x_min>
    <x_points_nb>9</x_points_nb>

```

(continues on next page)

(continued from previous page)

```

    <x_step>2.5</x_step>
    <y_min>40</y_min>
    <y_points_nb>4</y_points_nb>
    <y_step>2.5</y_step>
  </spatial_window>
  <criteria>S1</criteria>
  <weight>0.24</weight>
</predictor>
<predictor>
  <realtime_dataset_id>NWS_GFS_Forecast</realtime_dataset_id>
  <realtime_data_id>hgt</realtime_data_id>
  <archive_dataset_id>NCEP_Reanalysis_v1</archive_dataset_id>
  <archive_data_id>hgt</archive_data_id>
  <level>1000</level>
  <time>30</time>
  <spatial_window>
    <grid_type>regular</grid_type>
    <x_min>0</x_min>
    <x_points_nb>8</x_points_nb>
    <x_step>2.5</x_step>
    <y_min>37.5</y_min>
    <y_points_nb>5</y_points_nb>
    <y_step>2.5</y_step>
  </spatial_window>
  <criteria>S1</criteria>
  <weight>0.22</weight>
</predictor>
</analog_dates>
<analog_dates>
  <analog_number>24, 24, 34, 54, 164, 344</analog_number>
  <predictor>
    <preprocessing>
      <preprocessing_method>HumidityIndex</preprocessing_method>
      <preprocessing_data>
        <realtime_dataset_id>NWS_GFS_Forecast</realtime_dataset_id>
        <realtime_data_id>rhum</realtime_data_id>
        <archive_dataset_id>NCEP_Reanalysis_v1</archive_dataset_id>
        <archive_data_id>rhum</archive_data_id>
        <level>600</level>
        <time>18</time>
      </preprocessing_data>
      <preprocessing_data>
        <realtime_dataset_id>NWS_GFS_Forecast</realtime_dataset_id>
        <realtime_data_id>prwtr</realtime_data_id>
        <archive_dataset_id>NCEP_Reanalysis_v1</archive_dataset_id>
        <archive_data_id>prwtr</archive_data_id>
        <level>0</level>
        <time>18</time>
      </preprocessing_data>
    </predictor>
  </analog_dates>

```

(continues on next page)



(continued from previous page)

```

</preprocessing>
<spatial_window>
  <grid_type>regular</grid_type>
  <x_min>5</x_min>
  <x_points_nb>3</x_points_nb>
  <x_step>2.5</x_step>
  <y_min>45</y_min>
  <y_points_nb>1</y_points_nb>
  <y_step>2.5</y_step>
</spatial_window>
<criteria>RMSE</criteria>
<weight>0.39</weight>
</predictor>
<predictor>
  <preprocessing>
    <preprocessing_method>HumidityIndex</preprocessing_method>
    <preprocessing_data>
      <realtime_dataset_id>NWS_GFS_Forecast</realtime_dataset_id>
      <realtime_data_id>rhum</realtime_data_id>
      <archive_dataset_id>NCEP_Reanalysis_v1</archive_dataset_id>
      <archive_data_id>rhum</archive_data_id>
      <level>700</level>
      <time>12</time>
    </preprocessing_data>
    <preprocessing_data>
      <realtime_dataset_id>NWS_GFS_Forecast</realtime_dataset_id>
      <realtime_data_id>prwtr</realtime_data_id>
      <archive_dataset_id>NCEP_Reanalysis_v1</archive_dataset_id>
      <archive_data_id>prwtr</archive_data_id>
      <level>0</level>
      <time>12</time>
    </preprocessing_data>
  </preprocessing>
  <spatial_window>
    <grid_type>regular</grid_type>
    <x_min>5</x_min>
    <x_points_nb>3</x_points_nb>
    <x_step>2.5</x_step>
    <y_min>45</y_min>
    <y_points_nb>2</y_points_nb>
    <y_step>2.5</y_step>
  </spatial_window>
  <criteria>RMSE</criteria>
  <weight>0.61</weight>
</predictor>
</analog_dates>
<analog_values>
  <predictand>
    <database>Precipitation-Daily-Station-MeteoSwiss-Rhone.nc</database>

```

(continues on next page)

(continued from previous page)

```
<station_ids>27,43,44</station_ids>
</predictand>
</analog_values>
</atmoswing>
```

## THE VIEWER

This section contains the documentation for AtmoSwing Viewer. This document can also be downloaded as a pdf or epub using the bottom left menu.

The Viewer manual is divided into the following topics:

### 3.1 Usage

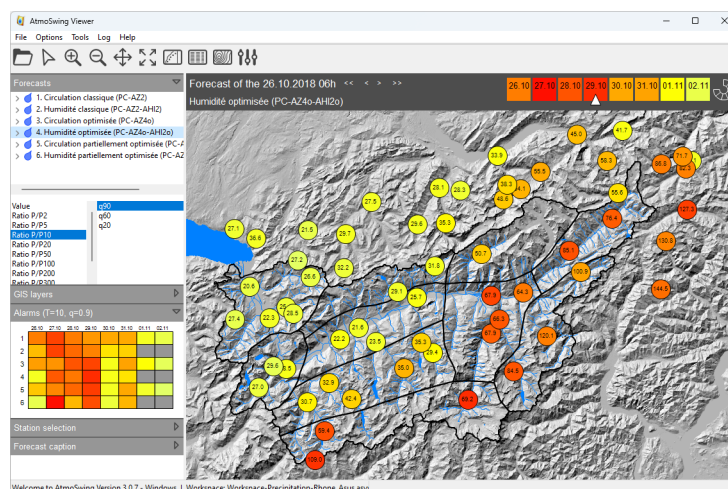
The Viewer relies on *workspaces* (defined in xml files), where the path to the forecast files is specified as well as the GIS layers and other display options. It is thus possible to quickly change from one region to another.

For a given region, the forecasts can be separated/gathered based on an *identifier specified in the parameters files*. (<method\_id>). The Viewer will manage this information as follows:

- <method\_id>: all forecast files with the same <method\_id> will be aggregated in the Viewer and labeled according to <method\_id\_display>.
- <specific\_tag>: all subregion-specific methods (defined by <specific\_tag> and labelled according to <specific\_tag\_display>) but having the same <method\_id> will be listed under the same heading. These specific forecasts for a subregion/basin are defined by specific parameters, and therefore have analog dates that may differ from other subregions. When one of these subregions is selected in the Viewer, the forecast for that specific method is displayed and the precipitation values for the other subregions (but for the same method) are grayed-out. When the header of the list is selected (i.e., the method), a summary of the optimal forecasts for each subregion is displayed (i.e., with analog dates that may differ from one subregion to another).

#### 3.1.1 Main interface

AtmoSwing Viewer has the interface shown below. It allows viewing the files produced by the Forecaster. The forecasts are first explored through the main interface on a map. Different graphs are then available to display more details.

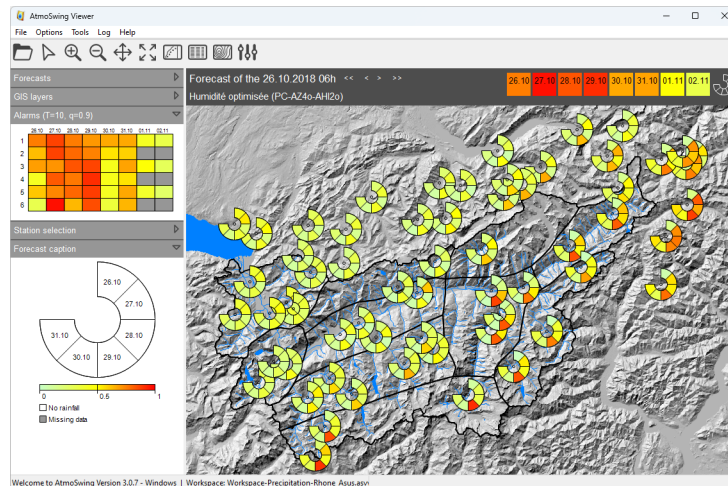


The toolbar allows the following actions:

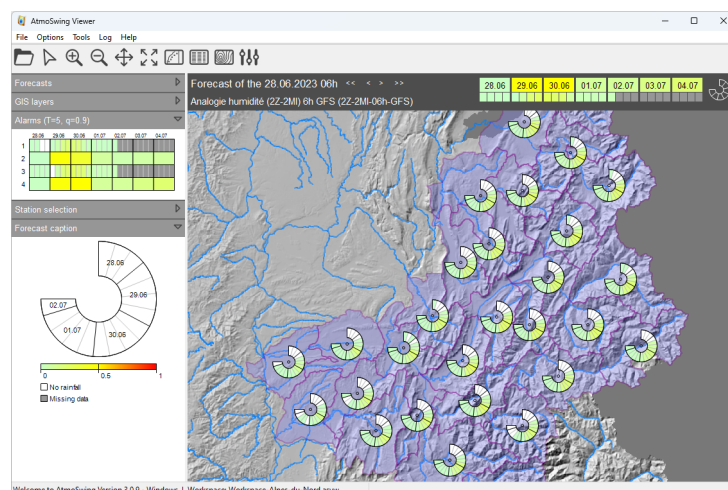
- Open a forecast file
- Select a station on the map
- Zoom in
- Zoom out
- Pan the map
- Adjust the map to all stations
- Open the distribution charts
- Open the list of analogues
- Open the predictors map
- Open the preferences

It is possible to open several forecast files, but they must have the same reference date (issued at the same lead time). If forecasts corresponding to another date (or time) are already opened, they will be closed.

The first visualization is a summary of the next lead times (Figure below). At the location of each station, a circular band is displayed with one box per lead time. The colors assigned to each slice illustrate the expected value for the lead time in question. This representation provides a quick overview of the coming days and makes it easy to identify a possible event. The same map can also be displayed for a specific lead time (Figure above). The choice of the lead time is made using the cursor above the map.



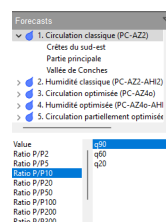
AtmoSwing Viewer offers the capability to display forecasts for sub-daily time steps. Users can view forecasts for various time steps simultaneously. In the example below, a combination of 6-hourly and daily time steps is shown. When a sub-daily forecast is opened, the interface adjusts accordingly by introducing subdivisions to the lead time selection, alarm panel, and circular bands on the map.



## Sidebar

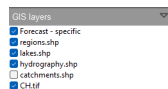
The sidebar on the left-hand side of the map provides some information and controls.

**Forecasts:** Choice of method and its representation. All open methods are displayed in the upper box. The lower left box changes the color scale, and the lower right one offers the choice of the quantile to display (20%, 60%, or 90%).

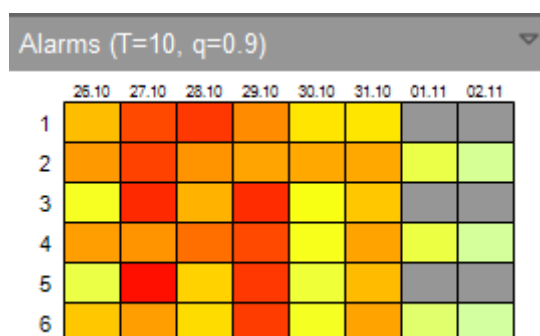


**GIS layers:** Display options for the opened GIS layers (right-click). It is possible to change the following properties of the different layers, if available:

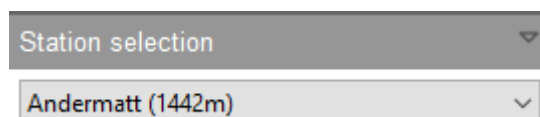
- Visibility - whether the layer is visible or not.
- Transparency - 0% is opaque and 100% is fully transparent.
- Color - it is applied to the layer features (lines & fill).
- The thickness of the line - for the lines or perimeters of polygons.



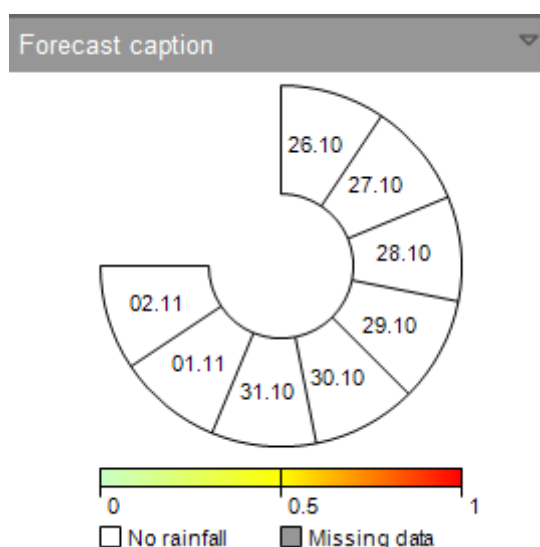
**Alarms:** Shows the potentially critical situations. The different rows of the grid correspond to the different open methods and the columns to the available lead times. The displayed colors correspond to the highest value of all stations (worst case; precipitation value normalized by the rain of a selected return period).



**Station selection:** Select the station to display the time series. An alternative to selection on the map.



**Forecast caption:** Display of the dates corresponding to the different boxes of the circular bands, as well as the color bar.

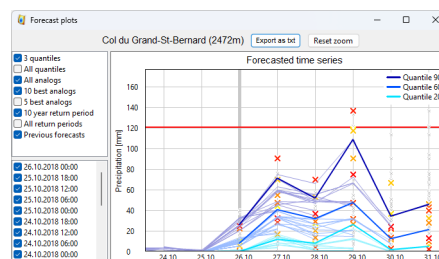


## Display of the time series

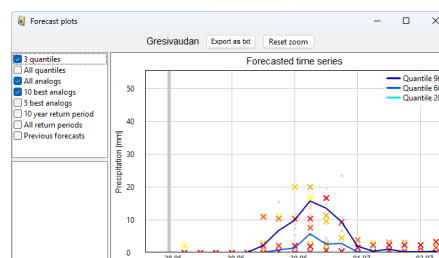
When a station is selected (on the map or in the drop-down list on the sidebar), the graph of the expected time series is displayed (Figure below). This graph corresponds to the selected method when clicking on the station. The display options (in the upper left panel) are as follows:

- The three classical quantiles (20%, 60%, 90%) in the form of thick blue lines.
- Every tenth of a quantile as a grey surface
- All analogs in the form of small grey crosses
- The ten best analogs as yellow - orange - red crosses. The colors are proportional to the rows of the analogs. Red is for the best analog and yellow for the 10th.
- The five best analogs according to the same principle.
- The value for a ten years return period in the form of a red line.
- The values of different return periods.
- Traces of the previous forecasts.


The lower left panel is the list of the previous forecasts. These are automatically imported when displaying the graph.

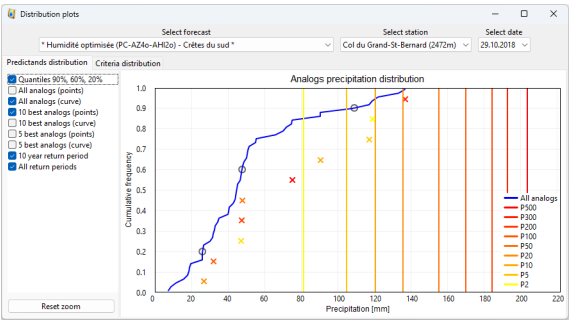


The time series can also be displayed for sub-daily time steps, such as in the example below for a 6-hourly forecast. The display options are the same as for the daily time steps.

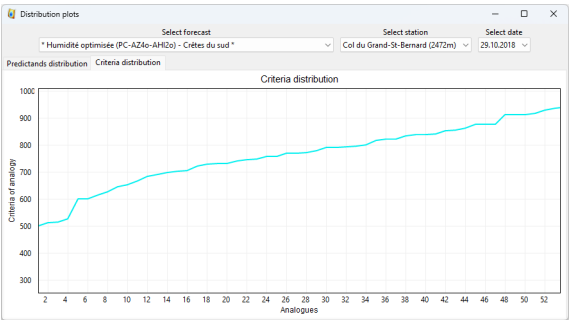


## Display of the distributions


The predictand distributions are accessible via the  button on the toolbar. It allows plotting the distribution of all analogs comparatively to the best 5 or 10 (Figure below). The display options are similar to those of the time series (previous section). The method, station, and lead time must be selected in the upper part of the window.



The distribution of the analogy criterion is also available in the same window through the tab located above the graph (Figure below). This distribution makes it possible to identify possible discontinuities in the values of the criteria and thus to increase the confidence in the best analogs.



### Display of the analog dates


The details of the analog dates and their associated criterion and predictand values (Figure below) are available through the  button on the toolbar. The method, station, and lead time must be selected in the upper part of the window. It is possible to sort the list by each of the columns by clicking on it. The first click ranks the values in ascending order and the second in descending order.


The screenshot shows the 'Analog details' window. It displays a table with 4 columns: 'Analog', 'Date', 'Value', and 'Criteria'. The table lists 18 analogs. The top of the window shows 'Forecast' (Humidité optimisée (PC-AZ4o-AH2o) - Crêtes du sud), 'Station' (Col du Grand-St-Bernard (2472m)), and 'Lead time' (29.10.2018).

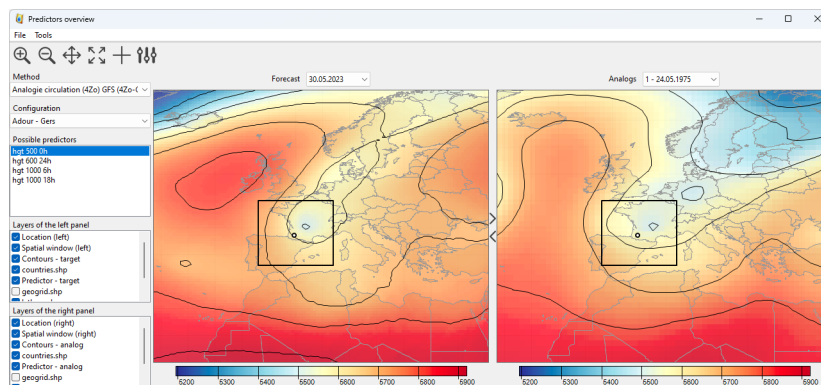
Analog	Date	Value	Criteria
1	15.10.2000	75.0	499.995
2	05.11.1994	137.0	511.379
3	14.10.1979	47.2	514.716
4	05.10.1992	32.1	525.721
5	24.11.2002	47.6	599.832
6	26.10.1976	26.7	601.085
7	07.11.1962	90.7	613.309
8	11.10.1987	117.3	625.625
9	15.10.1996	47.1	645.426
10	14.10.2000	119.0	651.304
11	16.10.1975	26.4	666.191
12	01.12.1988	19.0	683.580
13	13.10.2000	50.1	689.779
14	14.09.2006	31.5	697.044
15	20.09.2000	133.5	702.894
16	08.12.1977	48.7	703.729
17	03.10.1992	17.9	722.235
18	09.11.1971	44.6	729.535



## Display of the predictor fields

Maps of the predictor fields (Figure below) are accessible via the  button on the toolbar. This map allows comparing the predictors for the target situation and the analog dates. The options are as follows:

- The method, configuration, and predictors must be selected in the left panel.
- The lead time and the analog date are selected above the maps.
- The color scales and the map navigation are synchronized between both maps. The navigation synchronization can be disabled under the Tools menu (synchronize tools).
- The navigation tools are the same as for the main interface.
- An additional tool  is provided to display marks on both maps to help compare the location of some features.
- Each panel can be hidden to display only one map using the toggle buttons in the gap between the maps.
- The color tables can be changed in the preferences.
- The path to the predictor data must be specified in the preferences.



### 3.1.2 Workspaces

AtmoSwing Viewer relies on workspaces to allow changing of region or configuration. A workspace is an xml file containing the paths to the GIS layers and their rendering options, as well as the forecast files location and other properties that can be defined in the preferences.

The menu File allows to:

- Open a workspace
- Save the workspace
- Save the workspace as
- Create a new workspace

The creation of a new workspace is guided by a wizard to define:

1. The location of the workspace file
2. The path to the forecasts directory

### 3. The base map for the project

The base map options are:

- Custom layers
- Terrain from Google maps
- Map from Google maps
- Map from OpenStreetMap
- Map from ArcGIS Mapserver
- Satellite imagery from Google maps
- Satellite imagery from VirtualEarth

The workspace file does not need to be directly edited. The changes can be made through the interface and the workspace saved again. Its structure is as follows:

```
<?xml version="1.0" encoding="UTF-8"?>
<atmoswing version="1.0" target="viewer">
  <coordinate_system>EPSG:21781</coordinate_system>
  <forecast_directory>D:\_OwnCloud\AtmoSwing\CH-VS</forecast_directory>
  <colorbar_max_value>50</colorbar_max_value>
  <plot_time_series_past_days_nb>5</plot_time_series_past_days_nb>
  <panel_alarms_return_period>10</panel_alarms_return_period>
  <panel_alarms_quantile>0.900000</panel_alarms_quantile>
  <layers>
    <layer>
      <path>D:\AtmoSwing\GIS data\CH - Rhone\regions.shp</path>
      <type>vector</type>
      <transparency>0</transparency>
      <visibility>1</visibility>
      <line_width>2</line_width>
      <line_color>rgb(0, 0, 0)</line_color>
      <fill_color>rgb(0, 0, 0)</fill_color>
      <brush_style>106</brush_style>
    </layer>
    <layer>
      <path>D:\AtmoSwing\GIS data\CH - Rhone\lakes.shp</path>
      <type>vector</type>
      <transparency>0</transparency>
      <visibility>1</visibility>
      <line_width>1</line_width>
      <line_color>rgb(0, 128, 255)</line_color>
      <fill_color>rgb(0, 128, 255)</fill_color>
      <brush_style>100</brush_style>
    </layer>
    <layer>
      <path>D:\AtmoSwing\GIS data\CH - Rhone\hydrography.shp</path>
      <type>vector</type>
      <transparency>0</transparency>
      <visibility>1</visibility>
```

(continues on next page)

(continued from previous page)

```

    <line_width>1</line_width>
    <line_color>rgb(0, 128, 255)</line_color>
    <fill_color>rgb(0, 0, 0)</fill_color>
    <brush_style>100</brush_style>
  </layer>
  <layer>
    <path>D:\AtmoSwing\GIS data\CH - Rhone\catchments.shp</path>
    <type>vector</type>
    <transparency>0</transparency>
    <visibility>0</visibility>
    <line_width>1</line_width>
    <line_color>rgb(255, 255, 0)</line_color>
    <fill_color>rgb(0, 0, 0)</fill_color>
    <brush_style>106</brush_style>
  </layer>
  <layer>
    <path>D:\AtmoSwing\GIS data\CH - Rhone\SRTM\CH.tif</path>
    <type>raster</type>
    <transparency>0</transparency>
    <visibility>1</visibility>
    <line_width>0</line_width>
    <line_color></line_color>
    <fill_color></fill_color>
    <brush_style>0</brush_style>
  </layer>
</layers>
</atmoswing>

```

## 3.2 Preferences

The preferences frame can be opened from the menu Options / Preferences of the main interface or with a button from the toolbar. The options related to the workspace are saved in the workspace file itself, while the other options are saved in the user preferences file. The preferences are saved when the software is closed. The preferences are automatically loaded when the software is started. AtmoSwing will first look for a preference file in the same directory as the executable, then in the user directory.

### 3.2.1 Workspace

The properties defined under this tab are specific for the workspace currently opened.

The directory where the forecasts are stored must be provided:

Directory containing the forecasts

D:\AtmoSwing\Forecasts\ZAP Browse

The maximum predictand value for the color range defines the value for which red is assigned when colors are not standardized. The second option is the number of days of previous forecasts

to be displayed as traces in the time series plots. Three to five days are recommended.

Set the maximum rainfall value for the colorbar  mm/d

Number of past days to display on the timeseries

Alarm panel options: The alarm panel displays the highest value of all stations by lead time and method. The options are the return period for normalization and the quantile to be displayed. The quantile 0.9 allows remaining vigilant as to the possible occurrence of important events.

Alarms panel

Return period to display  years

Quantile to display  (in between 0 - 1)

Maximum length of time series to display: The time series plots can be limited to a maximum number of lead times to hide too uncertain lead times. This limit has to be specified in a number of days for forecasts with a daily time step and hours for forecasts with a sub-daily time step. By default, no limit is applied.

Maximum length of time series to display

Requires a restart or opening new forecasts.

Daily forecasts  days

Sub-daily forecasts  hours

### 3.2.2 Paths

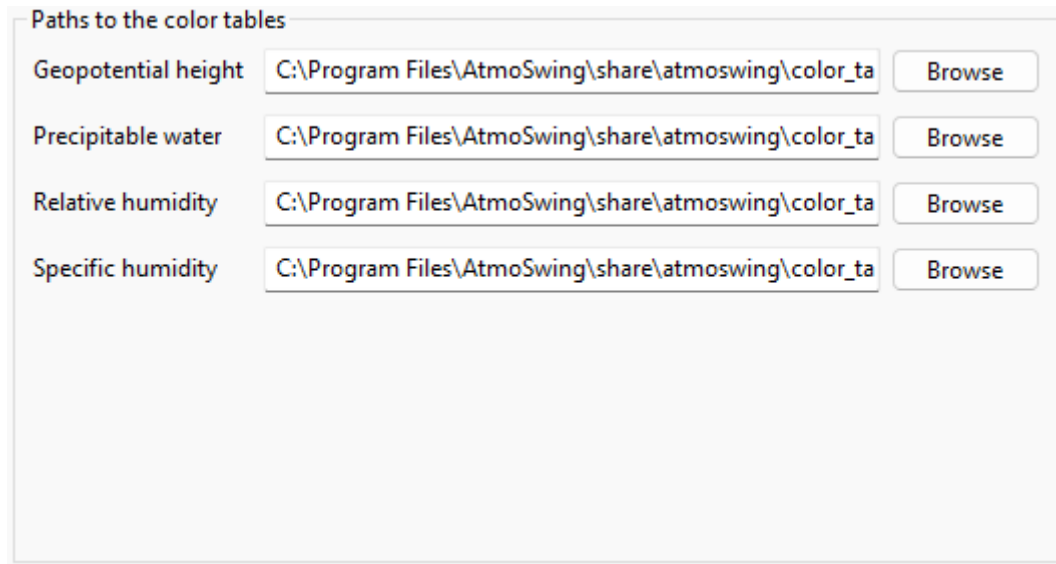
The paths to the predictor datasets must be provided along with their corresponding IDs to be found for the predictor maps.

Path to the predictor datasets

Dataset ID	Path to the directory	
<input type="text" value="Generic_ERA5"/>	<input type="text" value="D:\AtmoSwing\Reanalysis\ERA5-generic-1.00"/>	<input type="button" value="Browse"/>
<input type="text" value="Generic_NCEP_R1"/>	<input type="text" value="D:\AtmoSwing\Reanalysis\NR-1"/>	<input type="button" value="Browse"/>
<input type="text" value="NWS_GFS"/>	<input type="text" value="D:\AtmoSwing\Forecasts\gfs"/>	<input type="button" value="Browse"/>
<input type="text" value="ECMWF_IFS"/>	<input type="text" value="D:\AtmoSwing\Forecasts\ifs"/>	<input type="button" value="Browse"/>
<input type="text"/>	<input type="text"/>	<input type="button" value="Browse"/>
<input type="text"/>	<input type="text"/>	<input type="button" value="Browse"/>
<input type="text"/>	<input type="text"/>	<input type="button" value="Browse"/>

### 3.2.3 Colors

The path to the color palette files can be changed for each type of predictor. The default color tables are provided in the AtmoSwing installation directory. AtmoSwing can parse .rbg and .act files.

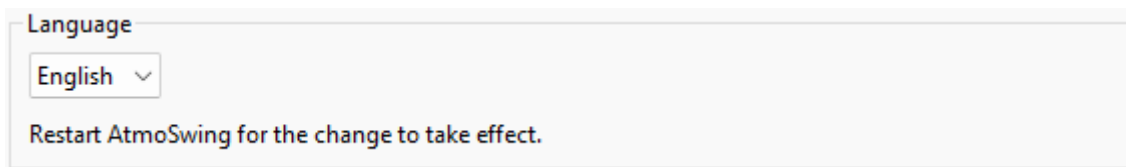


Paths to the color tables		
Geopotential height	C:\Program Files\AtmoSwing\share\atmoswing\color_ta	Browse
Precipitable water	C:\Program Files\AtmoSwing\share\atmoswing\color_ta	Browse
Relative humidity	C:\Program Files\AtmoSwing\share\atmoswing\color_ta	Browse
Specific humidity	C:\Program Files\AtmoSwing\share\atmoswing\color_ta	Browse

### 3.2.4 General

#### Language

The language can be changed from the drop-down list. The change is effective after restarting the software.



Language

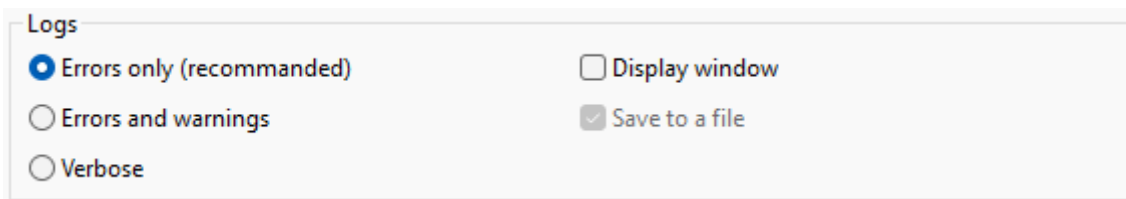
English ▾

Restart AtmoSwing for the change to take effect.

#### Log options

Log preferences control the display of the log window and the level (importance) of reported messages:

1. Errors only - recommended as long as everything goes as planned.
2. Errors and warnings - some warnings help to be vigilant in case of problems. However, there might be warnings even during regular operation.
3. Verbose - all operations are reported, which is an unnecessary amount of information during regular operation. As a result, important messages might be hidden in a large number of messages. This option is only useful to identify where AtmoSwing is failing.

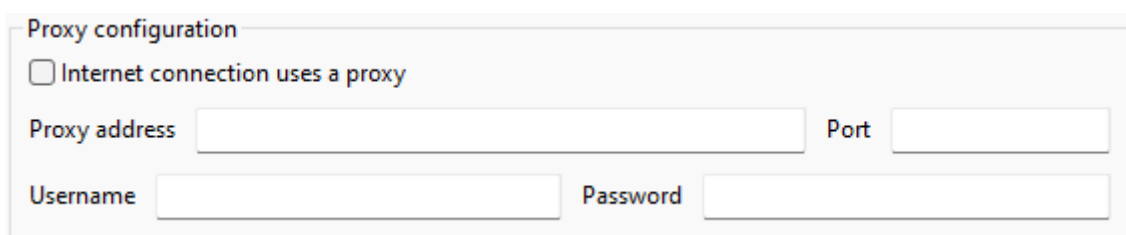


The screenshot shows a configuration panel titled "Logs". It contains three radio buttons for log levels: "Errors only (recommended)" (selected), "Errors and warnings", and "Verbose". To the right, there are two checkboxes: "Display window" (unchecked) and "Save to a file" (checked).

## Internet proxy

If required, proxy options can be provided. The proxy configuration contains the following elements:

- Proxy activation - if necessary.
- Proxy properties - the proxy address, port, the user name, and password.



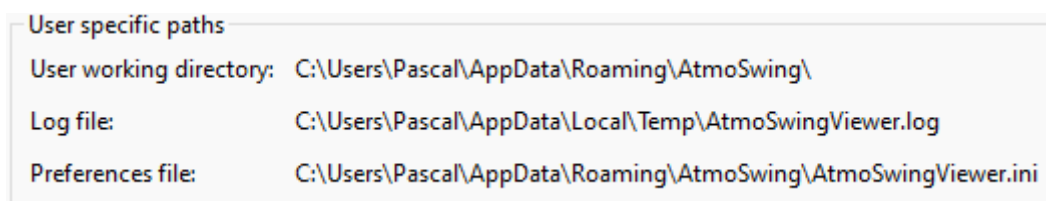
The screenshot shows a configuration panel titled "Proxy configuration". It starts with an unchecked checkbox "Internet connection uses a proxy". Below this are four text input fields: "Proxy address", "Port", "Username", and "Password".

## 3.2.5 Advanced

Allow multiple instances of the Viewer, for example to display and compare different forecasts in separate windows.

☒ Allow multiple instances of the viewer

Some paths automatically defined according to the OS: working directory, log file, and preferences file.



The screenshot shows a panel titled "User specific paths". It lists three paths with their corresponding file names: "User working directory: C:\Users\Pascal\AppData\Roaming\AtmoSwing\", "Log file: C:\Users\Pascal\AppData\Local\Temp\AtmoSwingViewer.log", and "Preferences file: C:\Users\Pascal\AppData\Roaming\AtmoSwing\AtmoSwingViewer.ini".

## THE DOWNSCALER

This section contains the documentation for AtmoSwing Downscaler. This document can also be downloaded as a pdf or epub using the bottom left menu.

The Downscaler manual is divided into the following topics:

### 4.1 Usage

The Downscaler allows the AMs to be used in a climatic context, either for climate reconstruction or for climate change impact studies. When used for future climate impact, the user must pay close attention to the selected predictors, so that they can represent the climate change signal. It is a relatively new field of application of AMs.

**Warning:** The Downscaler is the last addition to AtmoSwing and has therefore not been tested/used as intensively as the other tools. Some bugs might be encountered. Please report them as [GitHub issues](#) or contact the developer directly.

#### 4.1.1 Requirements

The Downscaler needs:

- *A predictand database*
- *Predictors from the target period*
- *A predictors archive (e.g. a reanalysis dataset)*
- *Parameters file defining the properties of the method.*

#### 4.1.2 Outputs

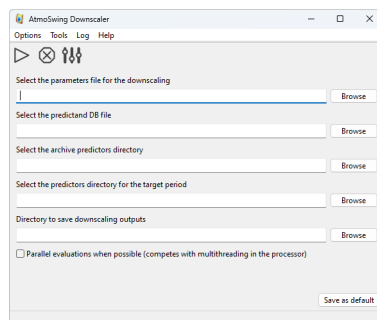
The Downscaler produces compressed **NetCDF files** containing:

- The values of the predictand for the target period
- The analog dates
- The values of the analogy criteria
- The target dates

- Some reference values (e.g., precipitation for different return periods)
- Some station metadata (id, name, coordinates, height)

### 4.1.3 Graphical user interface

The main interface of the Downscaler is as follows.



The toolbar allows the following actions:

- Start the calculations.
- Stop the current calculations.
- Define the preferences.

What is needed:

- *The parameters file for downscaling*
- The predictand DB
- The directory containing the predictors for the archive period
- The directory containing the predictors for the target period
- The directory to save the downscaling results

### 4.1.4 Command line interface

The Downscaler also has a command line interface. The options are as follows:

- h, --help** Displays the help of command line options
- v, --version** Displays the software version
- r, --run-number=<nb>** A given run number
- f, --file-parameters=<file>** File containing the downscaling parameters
- predictand-db=<file>** The predictand DB
- station-id=<id>** The predictand station ID
- dir-archive-predictors=<dir>** The archive predictors directory
- dir-scenario-predictors=<dir>** The scenario predictors directory
- downscaling-method=<method>** Choice of the downscaling method (classic:  
classic downscaling)



- n, --threads-nb=<n>** Number of threads to use
- s, --silent** Silent mode
- l, --local** Work in local directory
- log-level=<n>** Set the log level (0: minimum, 1: errors, 2: warnings (default), 3: verbose)

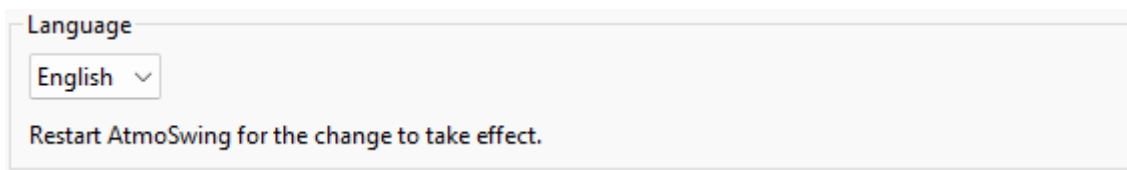
## 4.2 Preferences

The preferences frame can be opened from the menu Options / Preferences of the main interface or with a button from the toolbar. The preferences are saved when the software is closed. The preferences are automatically loaded when the software is started. AtmoSwing will first look for a preference file in the same directory as the executable, then in the user directory.

### 4.2.1 General options

#### Language

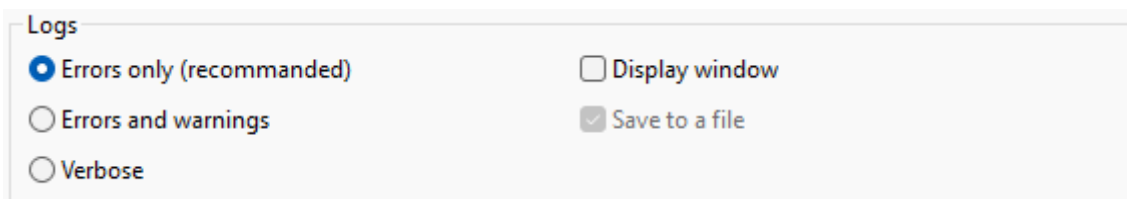
The language can be changed from the drop-down list. The change is effective after restarting the software.



#### Log options

Log preferences control the display of the log window and the level (importance) of reported messages:

1. Errors only - recommended as long as everything goes as planned.
2. Errors and warnings - some warnings help to be vigilant in case of problems. However, there might be warnings even during regular operation.
3. Verbose - all operations are reported, which is an unnecessary amount of information during regular operation. As a result, important messages might be hidden in a large number of messages. This option is only useful to identify where AtmoSwing is failing.



## Directories

Different directories must be specified:

- Directory containing the archive predictors: predictor describing the candidate situations (for example *reanalyses*).
- Directory containing the scenario predictors: predictor describing the target situations (for example *CMIP5* or *CORDEX*).
- Directory containing the predictand DB (*NetCDF file generated from the predictand data*).



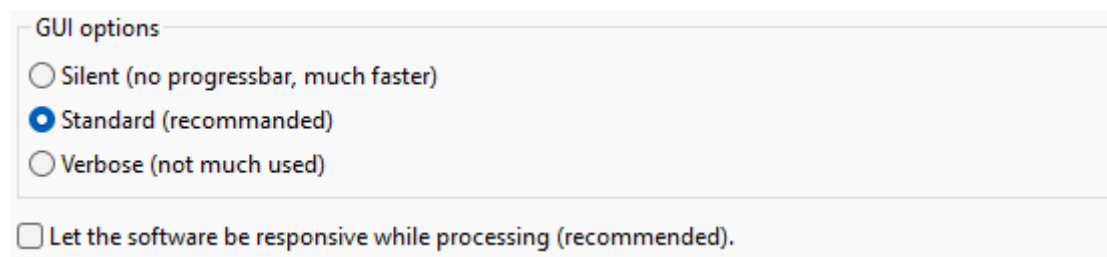
The screenshot shows a window titled "Directories". It contains three rows, each with a text input field and a "Browse" button. The first row is labeled "Directory containing archive predictors", the second "Directory containing scenario predictors", and the third "Default predictand DB directory".

### 4.2.2 Advanced options

#### Graphical User Interface (GUI) options

Options to control the behavior of the interface and its means of communication with the user. Saves calculation time by disabling dialog boxes. The choices are:

- Silent - no dialog box is displayed. The user must, therefore, keep an eye on the log window to detect any error. This option is only useful for very intensive calculations.
- Standard - recommended for regular use.
- Verbose - a bit more verbose than the Standard option.
- Enable or disable the reactivity of the graphical interface.

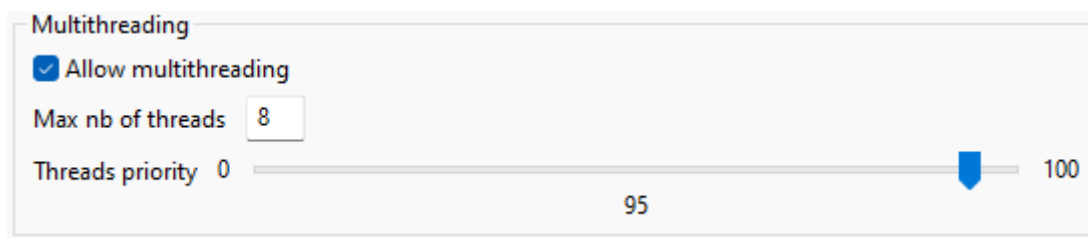


The screenshot shows a window titled "GUI options". It contains three radio button options: "Silent (no progressbar, much faster)", "Standard (recommended)" (which is selected), and "Verbose (not much used)". Below these is a checkbox labeled "Let the software be responsive while processing (recommended)".

## Multithreading

Options for using multiple threads:

- Allow multithreading - allows a gain in performance for the calculations (recommended).
- Maximum number of threads - to be chosen according to the computer's CPU. The software automatically proposes the optimal number of threads for the computer in use.
- Threads priority - allows defining the priority of the threads in competition with other ongoing processes.

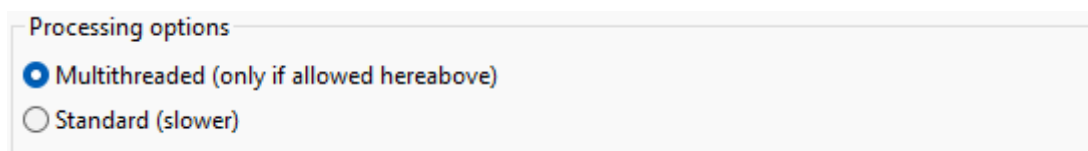


The screenshot shows a settings panel titled "Multithreading". It contains three options: a checked checkbox for "Allow multithreading", a text input field for "Max nb of threads" with the value "8", and a slider for "Threads priority" ranging from 0 to 100, with the current value set at 95.

## Processing options

These options influence how the search for analogs is performed, which has no impact on the results, only on the calculation time. The choices are as follows:

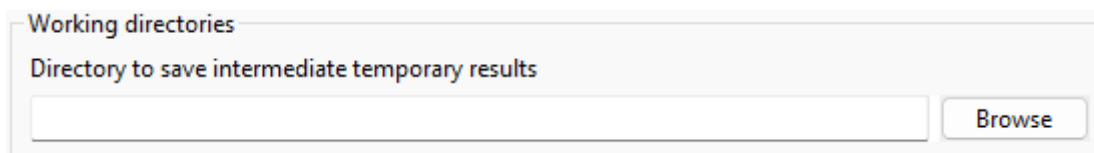
- Multithreaded - as long as the computer has at least two processors. It is the fastest version.
- Standard - this option is slower than the previous one as it uses a single thread.



The screenshot shows a settings panel titled "Processing options". It contains two radio button options: "Multithreaded (only if allowed hereabove)" which is selected, and "Standard (slower)".

## Working directory

A working directory can be specified to save optional intermediate files, for example between levels of analogy.



The screenshot shows a settings panel titled "Working directories". It contains a text input field labeled "Directory to save intermediate temporary results" and a "Browse" button to the right of the field.

## User-specific paths

Some paths automatically defined according to the OS: working directory, log file, and preferences file.

```
User specific paths
User working directory: C:\Users\Pascal\AppData\Roaming\AtmoSwing\
Log file:               C:\Users\Pascal\AppData\Local\Temp\AtmoSwingDownscaler.log
Preferences file:       C:\Users\Pascal\AppData\Roaming\AtmoSwing\AtmoSwingDownscaler.ini
```

## 4.3 Downscaler parameters file

The parameters file for the Downscaler is based on the *generic parameters file for AtmoSwing*, but has some specific elements as detailed below. A *full example* is provided at the bottom of the page.

### 4.3.1 Root node

The target property of the root node needs to be downscaler.

```
<atmoswing version="1.0" target="downscaler">
```

### 4.3.2 Content 'time\_properties'

The content of <time\_properties> is globally similar to the standard parameters file, to the exception of a new <downscaling\_period> element.

#### Element 'downscaling\_period'

The content of <downscaling\_period> defines the target period. It works as the <target\_period> element of the standard parameters file. Two options are possible:

Providing the years :

- <start\_year>: Start of the target period (ex: 2050)
- <end\_year>: End of the target period (ex: 2099)

Or the exact dates:

- <start>: Start of the target period (ex: 01.10.2050)
- <end>: End of the target period (ex: 31.09.2099)

### 4.3.3 Content 'analog\_dates'

The content of <analog\_dates> is mostly similar to the basic parameters file structure. The difference is that two predictor datasets are needed:

- <proj\_dataset\_id>: Climate dataset
- <proj\_data\_id>: Name of the variable in the climate dataset
- <archive\_dataset\_id>: Archive dataset
- <archive\_data\_id>: Name of the variable in the archive dataset

```
<analog_dates>
  <analog_number>30</analog_number>
  <predictor>
    <proj_dataset_id>CORDEX</proj_dataset_id>
    <proj_data_id>hgt</proj_data_id>
    <archive_dataset_id>NCEP_Reanalysis_v1</archive_dataset_id>
    <archive_data_id>hgt</archive_data_id>
    ...
  </predictor>
  ...
</analog_dates>
```

### 4.3.4 Full example

Example of a parameters file for the Downscaler:

```
<?xml version="1.0" encoding="UTF-8" ?>
<atmoswing version="1.0" target="downscaler">
  <description>
    <method_id>2Z</method_id>
    <method_id_display>Classic circulation</method_id_display>
    <specific_tag>somewhere</specific_tag>
    <specific_tag_display>groupment of interest</specific_tag_display>
    <description>Classic analogy on the atmospheric circulation (2 levels)</
  </description>
  </description>
  <time_properties>
    <archive_period>
      <start_year>1962</start_year>
      <end_year>2008</end_year>
    </archive_period>
    <downscaling_period>
      <start_year>2050</start_year>
      <end_year>2099</end_year>
    </downscaling_period>
    <time_step>24</time_step>
    <time_array_target>
      <time_array>simple</time_array>
    </time_array_target>
```

(continues on next page)

(continued from previous page)

```

<time_array_analogs>
  <time_array>days_interval</time_array>
  <interval_days>60</interval_days>
  <exclude_days>60</exclude_days>
</time_array_analogs>
</time_properties>
<analog_dates>
  <analog_number>100</analog_number>
  <predictor>
    <preload>1</preload>
    <proj_dataset_id>CORDEX</proj_dataset_id>
    <proj_data_id>zg500</proj_data_id>
    <archive_dataset_id>NCEP_Reanalysis_v1</archive_dataset_id>
    <archive_data_id>hgt</archive_data_id>
    <level>500</level>
    <time>24</time>
    <spatial_window>
      <x_min>-10</x_min>
      <x_points_nb>9</x_points_nb>
      <x_step>2.5</x_step>
      <y_min>30</y_min>
      <y_points_nb>5</y_points_nb>
      <y_step>2.5</y_step>
    </spatial_window>
    <criteria>S1</criteria>
    <weight>0.6</weight>
  </predictor>
  <predictor>
    <preprocessing>
      <preprocessing_method>SimpleGradients</preprocessing_method>
      <preprocessing_data>
        <proj_dataset_id>CORDEX</proj_dataset_id>
        <proj_data_id>zg850</proj_data_id>
        <archive_dataset_id>NCEP_Reanalysis_v1</archive_dataset_id>
        <archive_data_id>hgt</archive_data_id>
        <level>850</level>
        <time>12</time>
      </preprocessing_data>
    </preprocessing>
    <spatial_window>
      <x_min>-15</x_min>
      <x_points_nb>11</x_points_nb>
      <x_step>2.5</x_step>
      <y_min>35</y_min>
      <y_points_nb>7</y_points_nb>
      <y_step>2.5</y_step>
    </spatial_window>
    <criteria>S1</criteria>
    <weight>0.4</weight>
  </predictor>

```

(continues on next page)

(continued from previous page)

```
</predictor>
</analog_dates>
<analog_values>
  <predictand>
    <station_id>40</station_id>
  </predictand>
</analog_values>
</atmoswing>
```





## THE OPTIMIZER

This section contains the documentation for AtmoSwing Optimizer. This document can also be downloaded as a pdf or epub using the bottom left menu.

The Optimizer manual is divided into the following topics:

### 5.1 Usage

The Optimizer implements different optimization techniques, such as a semi-automatic sequential approach, Monte-Carlo simulations, and a global optimization technique using genetic algorithms. Establishing a statistical relationship between predictors and predictands is computationally intensive because it requires numerous assessments over decades.

The calibration of the AM is usually performed in a perfect prognosis framework. Perfect prognosis uses observed or reanalyzed data to calibrate the relationship between predictors and predictands. As a result, perfect prognosis yields relationships that are as close as possible to the natural links between predictors and predictands. However, there are no perfect models, and even reanalysis data may contain biases that cannot be ignored. Thus, the considered predictors should be as robust as possible, i.e., they should have minimal dependency on the model.

A statistical relationship is established with a trial and error approach by processing a forecast for every day of a calibration period. A certain number of days close to the target date are excluded to consider only independent candidate days. Validating the parameters of AMs on an independent validation period is very important to avoid over-parametrization and to ensure that the statistical relationship is valid for another period. In order to account for climate change and the evolution of measuring techniques, it is recommended that a noncontinuous period for validation should be used, distributed over the entire archive (for example, every 5th year). AtmoSwing users can thus specify a list of the years to set apart for the validation that are removed from possible candidate situations. At the end of the optimization, the validation score is processed automatically. The different verification scores available are detailed [here](#).

### 5.1.1 Requirements

The Optimizer needs:

- *A predictand database*
- *A predictors archive (e.g. a reanalysis dataset)*
- *Parameters files defining the properties of the method.*

### 5.1.2 Calibration methods

The Optimizer provides different approaches listed below.

#### Evaluation-only approaches

These methods do not seek to improve the parameters of the AM. They allow some assessments using the provided parameters. The different verification scores available are detailed [here](#).

- **Single assessment:** This approach processes the AM as described in the provided parameters file and assesses the defined score.
- **Evaluation of all scores:** Similar as above, but assesses all the implemented scores.
- **Only predictand values:** Does not process a skill score, but processes the AM and save the analog values into a file.
- **Only analog dates (and criteria):** Does not process a skill score, neither assign predictand values to the analog dates. It only saves the analog dates identified by the defined AM into a file.

#### Based on the classic calibration

The classic calibration is detailed [here](#).

- **Classic calibration:** The *classic calibration*
- **Classic+ calibration:** A *variant of the classic calibration* with some improvements.
- **Variables exploration Classic+:** Using the classic+ calibration to systematically explore a list of variables / levels / hours.

#### Global optimization

- **Genetic algorithms:** The optimization using *genetic algorithms*.

## Random exploration of the parameters space

- **Monte-Carlo simulations:** An exploration of the parameters space using Monte-Carlo simulations. These simulations were found to be limited in terms of ability to find reasonable parameters sets for even moderately complex AMs (1-2 levels of analogy with a few predictors).

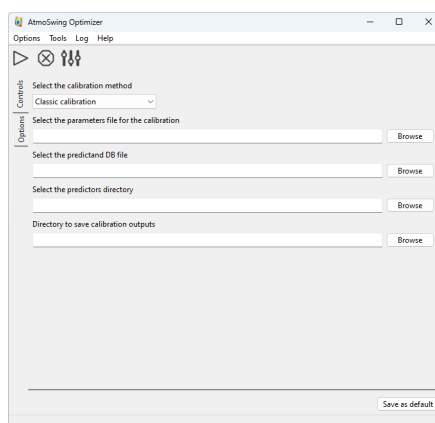
### 5.1.3 Outputs

The Optimizer produces different files:

- A text file with the resulting best parameters set and the skill score ([...]best\_parameters.txt).
- A text file with all the assessed parameters set and their corresponding skill score ([...]tested\_parameters.txt).
- An xml file with the best parameters set (to be used further by AtmoSwing Forecaster/Downscaler; [...]best\_parameters.xml).
- A NetCDF file containing the analog dates (AnalogDates[...].nc) both for the calibration and validation periods.
- A NetCDF file containing the analog values (AnalogValues[...].nc) both for the calibration and validation periods.
- A NetCDF file containing the skill scores (Scores[...].nc) both for the calibration and validation periods.

### 5.1.4 Graphical user interface

The main interface of the Optimizer is as follows.



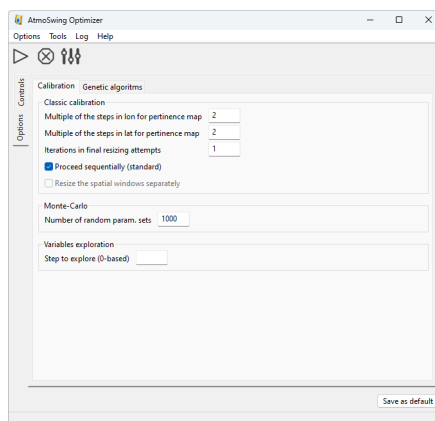
The toolbar allows the following actions:

- Start the optimization.
- Stop the current calculations.
- Define the preferences.

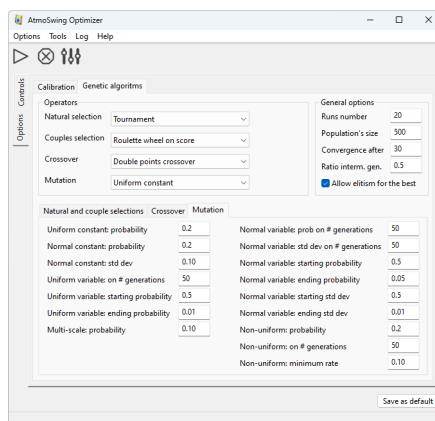
What is needed:

- Select one of the *calibration methods*
- *The parameters file*
- *The predictand database*
- The directory containing the *predictors for the archive period*
- The directory to save the results
- All the options for the selected calibration method (in the Options tab; see below)

There is one tab to define the options of the classic calibration, the variables exploration, and the Monte-Carlo simulations. The details of the options are given [here](#).



The other tab provides numerous options for genetic algorithms. The details of the options are given on the page of the *genetic algorithms*.



### 5.1.5 Command line interface

The Optimizer also has a command line interface, which is the preferred way of using it. The options are as follows:

- |                                   |   |
|-----------------------------------|---|
| <b>-h, --help</b>                 | Displays the help of command line options |
| <b>-v, --version</b>              | Displays the software version             |
| <b>-s, --silent</b>               | Silent mode                               |
| <b>-l, --local</b>                | Work in local directory                   |
| <b>-n, --threads-nb=&lt;n&gt;</b> | Number of threads to use                  |

**-g, --gpus-nb=<n>** Number of gpus to use

**-r, --run-number=<nb>** A given run number

**-f, --file-parameters=<file>** File containing the parameters

**--predictand-db=<file>** The predictand DB

**--station-id=<id>** The predictand station ID

**--dir-predictors=<dir>** The predictors directory

**--skip-valid** Skip the validation calculation

**--no-duplicate-dates** Do not allow to keep several times the same analog dates (e.g. for ensembles)

**--dump-predictor-data** Dump predictor data to binary files to reduce RAM usage

**--load-from-dumped-data** Load dumped predictor data into RAM (faster load)

**--replace-nans** Option to replace NaNs with -9999 (faster processing)

**--skip-nans-check** Do not check for NaNs (faster processing)

**--calibration-method=<method>** Choice of the calibration method:

- single: single assessment
- classic: classic calibration
- classicp: classic+ calibration
- varexplocp: variables exploration classic+
- montecarlo: Monte Carlo
- ga: genetic algorithms
- evalscores: evaluate all scores
- onlyvalues: get only the analog values
- onlydates: get only the analog dates

**--cp-resizing-iteration=<int>** Classic+: resizing iteration

**--cp-lat-step=<step>** Classic+: steps in latitudes for the relevance map

**--cp-lon-step=<step>** Classic+: steps in longitudes for the relevance map

**--cp-proceed-sequentially** Classic+: proceed sequentially

**--ve-step=<step\_nb>** Variables exploration: step to process

**--mc-runs-nb=<runs\_nb>** Monte Carlo: number of runs

**--ga-xxxxx=<value>** All GAs options are described on the [genetic algorithms](#) page

**--log-level=<n>** Set the log level (0: minimum, 1: errors, 2: warnings (default), 3: verbose)

## 5.2 Classic calibration

The calibration procedure that we call “sequential” or “classic” was elaborated by Bontron [1] (see also Radanovics *et al.* [7] and Ben Daoud *et al.* [9]). It is a semi-automatic procedure that optimizes the spatial windows in which the predictors are compared and the number of analogs are systematically assessed for every level of analogy. The different analogy levels (e.g., the atmospheric circulation or moisture index) are calibrated sequentially. The procedure consists of the following steps [1]:

1. Manual selection of the following parameters:
  - Meteorological variable
  - Pressure level
  - Temporal window (hour of the day)
  - Number of analogs
2. For every level of analogy:
  - Identification of the most skilled unitary cell (four points for the geopotential height when using the S1 criteria and one point otherwise) of the predictor data over a large domain. Every point or cell of the full domain is assessed based on the predictors of the current level of analogy.
  - From this most skilled cell, the spatial window is expanded by successive iterations in the direction of the most significant performance gain until no further improvement is possible (up to the user-specified maximum spatial window).
  - The number of analog situations  $N_i$ , which was initially set to an arbitrary value, is then reconsidered and optimized for the current level of analogy.
3. A new level of analogy can then be added based on other variables such as the moisture index at chosen pressure levels and hours of the day. The procedure starts again from step 2 (calibration of the spatial window and the number of analogs) for the new level. The parameters calibrated for the previous analogy levels are fixed and do not change.
4. Finally, the numbers of analogs for the different levels of analogy are reassessed. It is performed iteratively by varying the number of analogs of each level systematically.

The calibration is performed in successive steps for a limited number of parameters to minimize error functions or maximize skill scores. Except for the number of analogs, previously calibrated parameters are generally not reassessed. The benefit of this method is that it is relatively fast, it provides acceptable results, and it has low computing requirements.

### 5.2.1 Classic+ calibration

Small improvements were incorporated into this method in AtmoSwing Optimizer, then termed as “classic+”, by allowing the spatial windows to perform other moves, such as:

- (1) an increase in 2 simultaneous directions,
- (2) a decrease in 1 or 2 simultaneous directions,
- (3) an expansion or contraction (in every direction),
- (4) a shift of the window (without resizing) in 8 directions (including diagonals), and finally

- (5) all the above, but with a factor of 2, 3, or more. For example, an increase by two grid points in one (or more) direction is assessed. It allows skipping one size that may not be optimal.

These additional steps often result in spatial windows that are slightly different. The performance gain is rather marginal for reanalyses with a low resolution such as NR-1, but might be more consistent for reanalyses with higher resolutions due to the presence of more local minima.

## 5.3 Genetic algorithms

The sequential or classic calibration has substantial limitations: (i) it cannot automatically choose the pressure levels and temporal windows (hour of the day) for a given meteorological variable, (ii) it cannot handle dependencies between parameters, and (iii) it cannot easily handle new degrees of freedom. For this reason, genetic algorithms (GAs) were implemented in AtmoSwing Optimizer to perform a global optimization of AMs. It allows for the optimization of all parameters jointly in a fully automatic and objective way. See more details in Horton *et al.* [5].

Genetic algorithms are powerful but very demanding in terms of computational capacity. They require thousands of assessments to evolve towards the global optimum and should thus be used on a cluster rather than a single computer.

Many options and operator variants are available for genetic algorithms. Based on systematic tests detailed in Horton *et al.* [5], some presets were established in order to ease the use of genetic algorithms in AtmoSwing Optimizer. These presets are listed below, and all options are provided further down.

### 5.3.1 General options

Some general options are provided below:

- The population size is the number of individuals in the population.
- The intermediate generation ratio is the ratio of individuals that are kept after natural selection.
- The convergence steps is the number of generations without improvement after which the algorithm stops.
- The use of batches allows to reduce the computational time by using batches of data instead of the whole dataset.
- The batch size is the number of days in each batch.
- The number of epochs is the number of times the whole dataset is used for the optimization.
- The batch size and the number of epochs are only used if batches are used (in which case the number of generations for convergence is not used).

`--ga-pop-size=<int>` Size of the population (default 500)

`--ga-interm-gen=<0-1>` Ratio of the intermediate generation (default 0.5)

`--ga-conv-steps=<int>` Number of generations for convergence (default 30)

`--ga-use-batches=<0/1>` Use batches  
`--ga-batch-size=<int>` Size of the batches  
`--ga-number-epochs=<int>` Number of epochs if using batches

### 5.3.2 Operator options presets

Many options and operator variants control the optimization by genetic algorithms. Recommended configurations were predefined in presets.

`--ga-config=<1-5>` Operators presets

For all presets, the following options are identical: `--ga-conv-steps=30` `--ga-pop-size=500` `--ga-interm-gen=0.5` `--ga-ope-nat-sel=0` `--ga-ope-coup-sel=2` `--ga-ope-cross=7` `--ga-cross-bin-pt-nb=2` `--ga-cross-bin-share-b=1`

The difference between the presets concerns the mutation operator. Providing the following numbers (1-5) to the option `--ga-config` is equivalent to these corresponding presets:

1. Chromosome of adaptive search radius: `--ga-ope-mut=8`
2. Multiscale mutation: `--ga-ope-mut=9` `--ga-mut-multi-scale-p=0.1`
3. Nonuniform mutation (`pmut=0.1`, `Gmr=50`, `w=0.1`): `--ga-ope-mut=4`  
`--ga-mut-non-uni-p=0.1` `--ga-mut-non-uni-gens=50` `--ga-mut-non-uni-min-r=0.1`
4. Nonuniform mutation (`pmut=0.1`, `Gmr=100`, `w=0.1`): `--ga-ope-mut=4`  
`--ga-mut-non-uni-p=0.1` `--ga-mut-non-uni-gens=100` `--ga-mut-non-uni-min-r=0.1`
5. Nonuniform mutation (`pmut=0.2`, `Gmr=100`, `w=0.1`): `--ga-ope-mut=4`  
`--ga-mut-non-uni-p=0.2` `--ga-mut-non-uni-gens=100` `--ga-mut-non-uni-min-r=0.1`

Any of these options can be overridden by specifying it along with `--ga-config`.

### 5.3.3 Operator options

The different operators can be controlled with the following options:

`--ga-ope-nat-sel=<0-1>` Operator choice for natural selection:

0. ratio elitism (default)
1. tournament selection

`--ga-ope-coup-sel=<0-4>` Operator choice for couples selection:

0. rank pairing
1. random
2. roulette wheel on rank (default)
3. roulette wheel on score
4. tournament competition

`--ga-ope-cross=<0-9>` Operator choice for chromosome crossover:



0. single point crossover
1. double points crossover
2. multiple points crossover
3. uniform crossover
4. limited blending
5. linear crossover
6. heuristic crossover
7. binary-like crossover (default)
8. linear interpolation
9. free interpolation

**--ga-ope-mut=<0-10>** Operator choice for mutation:

0. random uniform constant
1. random uniform variable
2. random normal constant
3. random normal variable
4. non-uniform
5. self-adaptation rate
6. self-adaptation radius
7. self-adaptation rate chromosome
8. self-adaptation radius chromosome
9. multi-scale
10. no mutation

**--ga-nat-sel-tour-p=<0-1>** Natural selection - tournament probability (default 0.9)

**--ga-coup-sel-tour-nb=<2/3>** Couples selection - tournament candidates (2/3) (default 3)

**--ga-cross-mult-pt-nb=<int>** Standard crossover - number of points (default 3)

**--ga-cross-blen-pt-nb=<int>** Blending crossover - number of points (default 3)

**--ga-cross-blen-share-b=<1/0>** Blending crossover - beta shared (1/0) (default 1)

**--ga-cross-lin-pt-nb=<int>** Linear crossover - number of points (default 2)

**--ga-cross-heur-pt-nb=<int>** Heuristic crossover - number of points (default 2)

**--ga-cross-heur-share-b=<1/0>** Heuristic crossover - beta shared (1/0) (default 1)

**--ga-cross-bin-pt-nb=<int>** Binary-like crossover - number of points (default 2)

**--ga-cross-bin-share-b=<1/0>** Binary-like crossover - beta shared (1/0) (default 1)

**--ga-mut-unif-cst-p=<0-1>** Uniform mutation - probability (default 0.2)

**--ga-mut-norm-cst-p=<0-1>** Normal mutation - probability (default 0.2)

**--ga-mut-norm-cst-dev=<sd>** Normal mutation - standard deviation (default 0.1)

**--ga-mut-unif-var-gens=<int>** Variable uniform mutation - generations nb (default 50)

**--ga-mut-unif-var-p-strt=<0-1>** Variable uniform mutation - starting probability (default 0.5)

**--ga-mut-unif-var-p-end=<0-1>** Variable uniform mutation - end probability (default 0.01)

**--ga-mut-norm-var-gens-p=<int>** Variable normal mutation - generations nb for probability (default 50)

**--ga-mut-norm-var-gens-d=<int>** Variable normal mutation - generations nb for std deviation (default 50)

**--ga-mut-norm-var-p-strt=<0-1>** Variable normal mutation - starting probability (default 0.5)

**--ga-mut-norm-var-p-end=<0-1>** Variable normal mutation - end probability (default 0.05)

**--ga-mut-norm-var-d-strt=<sd>** Variable normal mutation - starting std deviation (default 0.5)

**--ga-mut-norm-var-d-end=<sd>** Variable normal mutation - end std deviation (default 0.01)

**--ga-mut-non-uni-p=<0-1>** Non uniform mutation - probability (default 0.1-0.2)

**--ga-mut-non-uni-gens=<int>** Non uniform mutation - generations nb (default 50-100)

**--ga-mut-non-uni-min-r=<0-1>** Non uniform mutation - minimum rate (default 0.1)

**--ga-mut-multi-scale-p=<0-1>** Multi-scale mutation - probability (default 0.1)

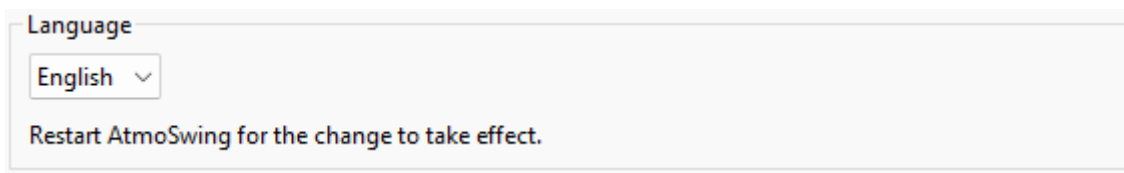
## 5.4 Preferences

The preferences frame can be opened from the menu Options / Preferences of the main interface or with a button from the toolbar. The preferences are saved when the software is closed. The preferences are automatically loaded when the software is started. AtmoSwing will first look for a preference file in the same directory as the executable, then in the user directory.

### 5.4.1 General options

#### Language

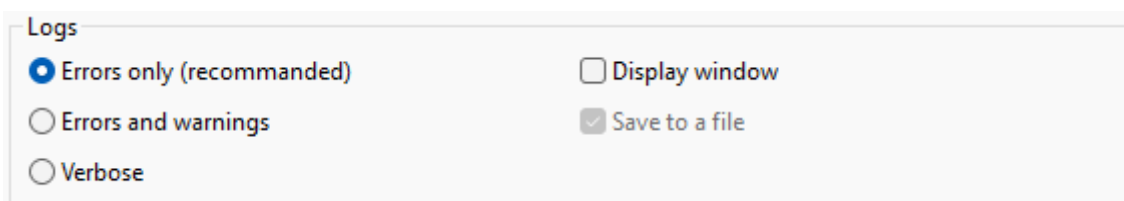
The language can be changed from the drop-down list. The change is effective after restarting the software.



#### Log options

Log preferences control the display of the log window and the level (importance) of reported messages:

1. Errors only - recommended as long as everything goes as planned.
2. Errors and warnings - some warnings help to be vigilant in case of problems. However, there might be warnings even during regular operation.
3. Verbose - all operations are reported, which is an unnecessary amount of information during regular operation. As a result, important messages might be hidden in a large number of messages. This option is only useful to identify where AtmoSwing is failing.



#### Directories

Different directories must be specified:

- Directory containing the archive predictors (for example *reanalyses*).
- Directory containing the predictand DB (*NetCDF file generated from the predictand data*).

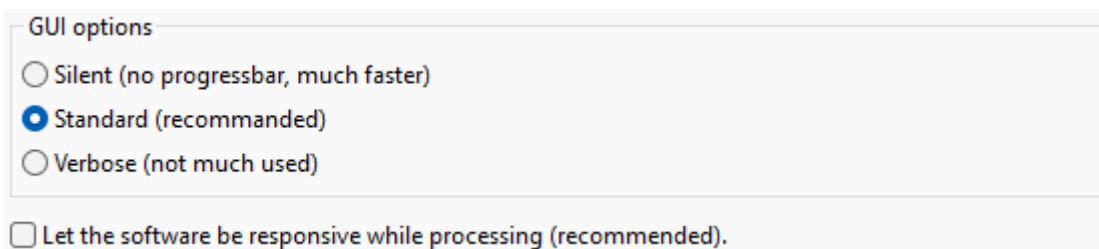


## 5.4.2 Advanced options

### Graphical User Interface (GUI) options

Options to control the behavior of the interface and its means of communication with the user. Saves calculation time by disabling dialog boxes. The choices are:

- Silent - no dialog box is displayed. The user must, therefore, keep an eye on the log window to detect any error. This option is only useful for very intensive calculations.
- Standard - recommended for regular use.
- Verbose - a bit more verbose than the Standard option.
- Enable or disable the reactivity of the graphical interface.

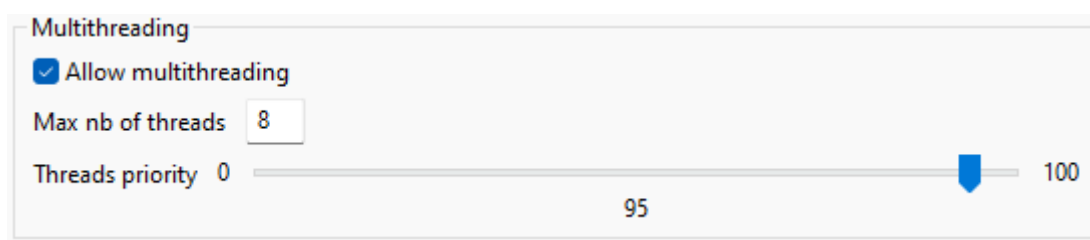


The screenshot shows a dialog box titled "GUI options". It contains three radio button options: "Silent (no progressbar, much faster)", "Standard (recommended)" (which is selected), and "Verbose (not much used)". Below these is a checkbox labeled "Let the software be responsive while processing (recommended)." which is currently unchecked.

### Multithreading

Options for using multiple threads:

- Allow multithreading - allows a gain in performance for the calculations (recommended).
- Maximum number of threads - to be chosen according to the computer's CPU. The software automatically proposes the optimal number of threads for the computer in use.
- Threads priority - allows defining the priority of the threads in competition with other ongoing processes.



The screenshot shows a dialog box titled "Multithreading". It contains a checked checkbox "Allow multithreading". Below it is a text input field for "Max nb of threads" with the value "8". At the bottom is a slider for "Threads priority" ranging from 0 to 100, with a blue arrow marker positioned at 95.

### Processing options

These options influence how the search for analogs is performed, which has no impact on the results, only on the calculation time. The choices are as follows:

- Multithreaded - as long as the computer has at least two processors. It is the fastest version.
- Standard - this option is slower than the previous one as it uses a single thread.

Processing options

☒ Multithreaded (only if allowed hereabove)

☐ Standard (slower)

## Working directory

A working directory can be specified to save optional intermediate files, for example between levels of analogy.

Working directories

Directory to save intermediate temporary results

## User-specific paths

Some paths automatically defined according to the OS: working directory, log file, and preferences file.

User specific paths

User working directory: C:\Users\Pascal\AppData\Roaming\AtmoSwing\

Log file: C:\Users\Pascal\AppData\Local\Temp\AtmoSwingOptimizer.log

Preferences file: C:\Users\Pascal\AppData\Roaming\AtmoSwing\AtmoSwingOptimizer.ini

## 5.5 Verification scores

The Optimizer has different verification scores implemented. The most often used for precipitation is the CRPS. The options are:

- CRPSS (Continuous ranked probability skill score): CRPSS
- CRPS (Continuous ranked probability score): CRPS
- CRPS accuracy: CRPSaccuracy
- CRPS sharpness: CRPSsharpness
- CRPS reliability: CRPSreliability
- CRPS potential: CRPSpotential
- Absolute difference of the frequency of null values: DF0
- Proportion correct (contingency table): PC
- Threat score (contingency table): TS
- Bias (contingency table): BIAS
- False alarm ratio (contingency table): FARA
- Hit rate (contingency table): H

- False alarm rate (contingency table): F
- Heidke skill score (contingency table): HSS
- Pierce skill score (contingency table): PSS
- Gilbert skill score (contingency table): GSS
- Mean absolute error: MAE
- Mean squared error: MSE
- Root mean squared error: RMSE
- Brier score: BS
- Brier skill score: BSS
- Stable equitable error in probability space: SEEPS
- The Verification Rank Histogram (Talagrand Diagram): RankHistogram
- Reliability of the Verification Rank Histogram (Talagrand Diagram): RankHistogramReliability

## 5.6 Optimizer parameters file

The Optimizer has two different kinds of parameter files. One type for the classic and classic+ calibration (see [Classic calibration parameters file](#)) and another for the optimization techniques using genetic algorithms or Monte-Carlo simulations (see [optimization parameters file](#)). However, some attributes are identical for both of them.

### 5.6.1 Shared properties

#### Content 'time\_properties'

The content of <time\_properties> is globally similar to the [standard parameters file](#), to the exception of a <calibration\_period> and a <validation\_period> element.

#### Element 'calibration\_period'

The content of <calibration\_period> defines the target period on which the method is trained. It works as the <archive\_period> element. Two options are possible:

Providing the exact dates:

- <start>: Start of the calibration period (ex: 01.01.1981)
- <end>: End of the calibration period (ex: 31.12.2010)

Alternatively, only providing the years:

- <start\_year>: Start of the calibration period (ex: 1981)
- <end\_year>: End of the calibration period (ex: 2010)

### Element 'validation\_period'

The content of <validation\_period> defines the independent period that is reserved for validation.

**Warning:** The validation period must be covered by the calibration period (to correctly load the required data). AtmoSwing then handles the validation period by excluding it from the calibration period. Once again: the validation period must not provide dates that are not covered by the calibration period.

Two options are possible:

Providing a list of years (distributed over the calibration period):

- <years>: List of years (ex: 1985, 1990, 1995, 2000, 2005, 2010)

Alternatively, providing a range of years:

- <start\_year>: Start of the validation period (ex: 2001)
- <end\_year>: End of the validation period (ex: 2010)

### Content 'evaluation'

The Optimizer has a <evaluation> node to specify the score to use and other characteristics:

```
<evaluation>
  <score>CRPS</score>
  <time_array>simple</time_array>
</evaluation>
```

### Element 'score'

The name of the score to use must be provided. Options are listed here: [verification scores](#)

### Element 'time\_array'

The <time\_array> element must be provided. It should however be set to simple as there are no other options available yet.

## Other optional properties

Other options can be added to specify how the score should be processed.

- `<threshold>`: Threshold to use for the processing of scores relying on thresholds exceedance, such as the contingency table. Example: `<threshold>0.5</threshold>`
- `<quantile>`: Quantile to use for the processing of scores relying on a single value rather than the distribution. Example: `<quantile>0.9</quantile>`
- `<on_mean>1</on_mean>`: Specifies to process the score on the mean of the analogs rather than a quantile (see above). For example, when used with the score MSE. The value 1 has no meaning other than “true”.

### 5.6.2 Classic calibration parameters file

The calibration parameters file defines the parameters to be calibrated and the range of the authorized values. There are different methods possible to define these options:

- minmax (the most used): you must provide a min value of the parameter, a max value and a step value. For example:

```
<x_min min="-10" max="10" step="0.25" method="minmax"></x_min>
```

- array: the content of the node will then be parsed as an array. For example:

```
<station_id method="array">1,5,6,12,35</station_id>
```

- fixed: should not be calibrated even if min and max values are defined. A value for the parameter must then be provided.

## Root node

The target property of the root node needs to be calibrator.

```
<atmoswing version="1.0" target="calibrator">
```

## Content ‘analog\_dates’

The content of `<analog_dates>` is mostly similar to the *standard parameters file structure*, but the parameters to calibrate need to be specified.



### Element 'analog\_number'

This element varies the number of analog situations  $N_i$  to consider, for example:

```
<analog_number min="10" max="100" step="5" method="minmax"></analog_number>
```

### Element 'spatial\_window'

This element defines the extent of the spatial window for each predictor. `<x_min>` and `<y_min>` are the minimum longitude and latitude respectively, `<x_points_nb>` and `<y_points_nb>` are the number of points from the predictor grid to use on the longitude and latitude axes, respectively. `<x_step>` and `<y_step>` are the longitude/latitude resolutions to use. This window needs to be defined for each predictor.

```
<spatial_window>
  <x_min min="-10" max="10" step="0.25" method="minmax"></x_min>
  <x_points_nb min="2" max="120" step="1" method="minmax"></x_points_nb>
  <x_step>0.25</x_step>
  <y_min min="35" max="50" step="0.25" method="minmax"></y_min>
  <y_points_nb min="2" max="90" step="1" method="minmax"></y_points_nb>
  <y_step>0.25</y_step>
</spatial_window>
```

### Full example

```
<?xml version="1.0" encoding="UTF-8" ?>
<atmoswing version="1.0" target="calibrator">
  <description>
    <method_id>PD-A4Z</method_id>
    <method_id_display>Enhanced circulation</method_id_display>
    <specific_tag>CH</specific_tag>
    <specific_tag_display>all stations</specific_tag_display>
    <description>Enhanced analogy on the atmospheric circulation (4 levels)</
  <description>
</description>
  <time_properties>
    <archive_period>
      <start_year>1981</start_year>
      <end_year>2010</end_year>
    </archive_period>
    <calibration_period>
      <start_year>1981</start_year>
      <end_year>2010</end_year>
    </calibration_period>
    <validation_period>
      <years>1985, 1990, 1995, 2000, 2005, 2010</years>
    </validation_period>
    <time_step>24</time_step>
```

(continues on next page)

(continued from previous page)

```

<time_array_target>
  <time_array>simple</time_array>
</time_array_target>
<time_array_analogs>
  <time_array>days_interval</time_array>
  <interval_days method="fixed">60</interval_days>
  <exclude_days>30</exclude_days>
</time_array_analogs>
</time_properties>
<analog_dates>
  <analog_number min="10" max="100" step="5" method="minmax"></analog_number>
  <predictor>
    <preload>1</preload>
    <preprocessing>
      <preprocessing_method>SimpleGradients</preprocessing_method>
      <preprocessing_data>
        <dataset_id>ECMWF_ERA5</dataset_id>
        <data_id>pressure/z</data_id>
        <level>1000</level>
        <time>6</time>
      </preprocessing_data>
    </preprocessing>
    <spatial_window>
      <x_min min="-10" max="10" step="0.25" method="minmax"></x_min>
      <x_points_nb min="2" max="120" step="1" method="minmax"></x_points_nb>
      <x_step>0.25</x_step>
      <y_min min="35" max="50" step="0.25" method="minmax"></y_min>
      <y_points_nb min="2" max="90" step="1" method="minmax"></y_points_nb>
      <y_step>0.25</y_step>
    </spatial_window>
    <criteria>S1</criteria>
  </predictor>
  <predictor>
    <preload>1</preload>
    <preprocessing>
      <preprocessing_method>SimpleGradients</preprocessing_method>
      <preprocessing_data>
        <dataset_id>ECMWF_ERA5</dataset_id>
        <data_id>pressure/z</data_id>
        <level>1000</level>
        <time>30</time>
      </preprocessing_data>
    </preprocessing>
    <spatial_window>
      <x_min min="-10" max="10" step="0.25" method="minmax"></x_min>
      <x_points_nb min="2" max="120" step="1" method="minmax"></x_points_nb>
      <x_step>0.25</x_step>
      <y_min min="35" max="50" step="0.25" method="minmax"></y_min>
      <y_points_nb min="2" max="90" step="1" method="minmax"></y_points_nb>

```

(continues on next page)

(continued from previous page)

```

    <y_step>0.25</y_step>
  </spatial_window>
  <criteria>S1</criteria>
</predictor>
<predictor>
  <preload>1</preload>
  <preprocessing>
    <preprocessing_method>SimpleGradients</preprocessing_method>
    <preprocessing_data>
      <dataset_id>ECMWF_ERA5</dataset_id>
      <data_id>pressure/z</data_id>
      <level>700</level>
      <time>24</time>
    </preprocessing_data>
  </preprocessing>
  <spatial_window>
    <x_min min="-10" max="10" step="0.25" method="minmax"></x_min>
    <x_points_nb min="2" max="120" step="1" method="minmax"></x_points_nb>
    <x_step>0.25</x_step>
    <y_min min="35" max="50" step="0.25" method="minmax"></y_min>
    <y_points_nb min="2" max="90" step="1" method="minmax"></y_points_nb>
    <y_step>0.25</y_step>
  </spatial_window>
  <criteria>S1</criteria>
</predictor>
<predictor>
  <preload>1</preload>
  <preprocessing>
    <preprocessing_method>SimpleGradients</preprocessing_method>
    <preprocessing_data>
      <dataset_id>ECMWF_ERA5</dataset_id>
      <data_id>pressure/z</data_id>
      <level>500</level>
      <time>12</time>
    </preprocessing_data>
  </preprocessing>
  <spatial_window>
    <x_min min="-10" max="10" step="0.25" method="minmax"></x_min>
    <x_points_nb min="2" max="120" step="1" method="minmax"></x_points_nb>
    <x_step>0.25</x_step>
    <y_min min="35" max="50" step="0.25" method="minmax"></y_min>
    <y_points_nb min="2" max="90" step="1" method="minmax"></y_points_nb>
    <y_step>0.25</y_step>
  </spatial_window>
  <criteria>S1</criteria>
</predictor>
</analog_dates>
<analog_values>
  <predictand>

```

(continues on next page)

(continued from previous page)

```
<station_id min="1" max="100" step="1" method="minmax"></station_id>
</predictand>
</analog_values>
<evaluation>
  <score>CRPS</score>
  <time_array>simple</time_array>
</evaluation>
</atmoswing>
```

### 5.6.3 Optimization parameters file

The optimization parameters file used for genetic algorithms or Monte Carlo simulations defines the parameters to be optimized and the range of the authorized values. There are different aspects to this:

- Defining the range of numerical values: a `lowerlimit` value of the parameter, an `upperlimit` value, and an `iteration` value must be provided. For example:

```
<x_min lowerlimit="-10.5" upperlimit="10.5" iteration="0.75" lock="0"></x_min>
```

- Defining a list of values: the content of the node will then be parsed as an array. For example:

```
<level method="array" lock="0">0, 300, 500, 850, 1000</level>
```

- The `lock` property allows to fix a value that will not be optimized even if `lowerlimit` and `upperlimit` values are defined, for example. You then need to provide a value for the parameter. It can be used to temporarily disable the calibration of a parameter. For example:

```
<time lowerlimit="0" upperlimit="30" iteration="6" lock="1">18</time>
```

#### Root node

The `target` property of the root node needs to be `optimizer`.

```
<atmoswing version="1.0" target="optimizer">
```

#### Content 'analog\_dates'

The content of `<analog_dates>` is mostly similar to the *standard parameters file structure*, but the parameters to calibrate need to be specified.

## Element 'analogs\_number'

This element varies the number of analog situations  $N_i$  to consider, for example:

```
<analogs_number lowerlimit="5" upperlimit="80" iteration="1" lock="0"></analogs_
↪number>
```

## Element 'predictor'

This element is used to define the parameters to calibrate related to the predictors. The structure is mostly similar to the *standard parameters file structure*. The preload option specifies that the data should be preloaded prior to the calibration, which saves a substantial amount of time. For example:

Example:

```
<predictor>
  <preload>1</preload>
  <dataset_id>GenericNetcdf</dataset_id>
    <data_id method="array" lock="0">pl/r, pl/t, pl/u, pl/v, pl/z, sfa/msl</
↪data_id>
  <level method="array" lock="0">0, 300, 500, 850, 1000</level>
    <time lowerlimit="0" upperlimit="30" iteration="6" lock="0"></time>
  <spatial_window>
    <x_min lowerlimit="-10.5" upperlimit="10.5" iteration="0.75" lock="0"></x_min>
    <x_points_nb lowerlimit="1" upperlimit="40" iteration="1" lock="0"></x_points_
↪nb>
    <x_step>0.75</x_step>
    <y_min lowerlimit="35.25" upperlimit="50.25" iteration="0.75" lock="0"></y_
↪min>
    <y_points_nb lowerlimit="1" upperlimit="30" iteration="1" lock="0"></y_points_
↪nb>
    <y_step>0.75</y_step>
  </spatial_window>
  <criteria method="array" lock="0">MD, RMSE, S0, S1, S2</criteria>
</predictor>
```

## Full example

```
<?xml version="1.0" encoding="UTF-8" ?>
<atmoswing version="1.0" target="optimizer">
  <description>
    <method_id>PD-A1Xo</method_id>
    <method_id_display>1 level optimized</method_id_display>
    <specific_tag>CH</specific_tag>
    <specific_tag_display>1 station</specific_tag_display>
    <description>1 level optimize</description>
  </description>
```

(continues on next page)

(continued from previous page)

```

<time_properties>
  <archive_period>
    <start_year>1981</start_year>
    <end_year>2010</end_year>
  </archive_period>
  <calibration_period>
    <start_year>1981</start_year>
    <end_year>2010</end_year>
  </calibration_period>
  <validation_period>
    <years>1985, 1990, 1995, 2000, 2005, 2010</years>
  </validation_period>
  <time_step>24</time_step>
  <time_array_target>
    <time_array>simple</time_array>
  </time_array_target>
  <time_array_analogs>
    <time_array>days_interval</time_array>
    <interval_days lock="1">60</interval_days>
    <exclude_days>30</exclude_days>
  </time_array_analogs>
</time_properties>
<analog_dates>
  <analogs_number lowerlimit="5" upperlimit="150" iteration="1" lock="0"></
↪analogs_number>
  <predictor>
    <preload>1</preload>
    <dataset_id>GenericNetcdf</dataset_id>
    <data_id method="array" lock="0">pl/r, pl/t, pl/u, pl/v, pl/z, sfa/msl</
↪data_id>
    <level method="array" lock="0">0, 300, 500, 850, 1000</level>
    <time lowerlimit="0" upperlimit="30" iteration="6" lock="0"></time>
    <spatial_window>
      <x_min lowerlimit="-10.5" upperlimit="10.5" iteration="0.75" lock="0"></x_
↪min>
      <x_points_nb lowerlimit="1" upperlimit="40" iteration="1" lock="0"></x_
↪points_nb>
      <x_step>0.75</x_step>
      <y_min lowerlimit="35.25" upperlimit="50.25" iteration="0.75" lock="0"></
↪y_min>
      <y_points_nb lowerlimit="1" upperlimit="30" iteration="1" lock="0"></y_
↪points_nb>
      <y_step>0.75</y_step>
    </spatial_window>
    <criteria method="array" lock="0">MD, RMSE, S0, S1, S2</criteria>
  </predictor>
</analog_dates>
<analog_dates>
  <analogs_number lowerlimit="5" upperlimit="80" iteration="1" lock="0"></

```

(continues on next page)

(continued from previous page)

```

↪analogs_number>
  <predictor>
    <preload>1</preload>
    <dataset_id>GenericNetcdf</dataset_id>
    <data_id method="array" lock="0">pl/r, pl/t, pl/u, pl/v, pl/z, sfa/msl</
↪data_id>
    <level method="array" lock="0">0, 300, 500, 850, 1000</level>
    <time lowerlimit="0" upperlimit="30" iteration="6" lock="0"></time>
    <spatial_window>
      <x_min lowerlimit="-10.5" upperlimit="10.5" iteration="0.75" lock="0"></x_
↪min>
      <x_points_nb lowerlimit="1" upperlimit="40" iteration="1" lock="0"></x_
↪points_nb>
      <x_step>0.75</x_step>
      <y_min lowerlimit="35.25" upperlimit="50.25" iteration="0.75" lock="0"></
↪y_min>
      <y_points_nb lowerlimit="1" upperlimit="30" iteration="1" lock="0"></y_
↪points_nb>
      <y_step>0.75</y_step>
    </spatial_window>
    <criteria method="array" lock="0">MD, RMSE, S0, S1, S2</criteria>
  </predictor>
</analog_dates>
<analog_values>
  <predictand>
    <station_id>0</station_id>
  </predictand>
</analog_values>
<evaluation>
  <score>CRPS</score>
  <time_array>simple</time_array>
</evaluation>
</atmoswing>

```