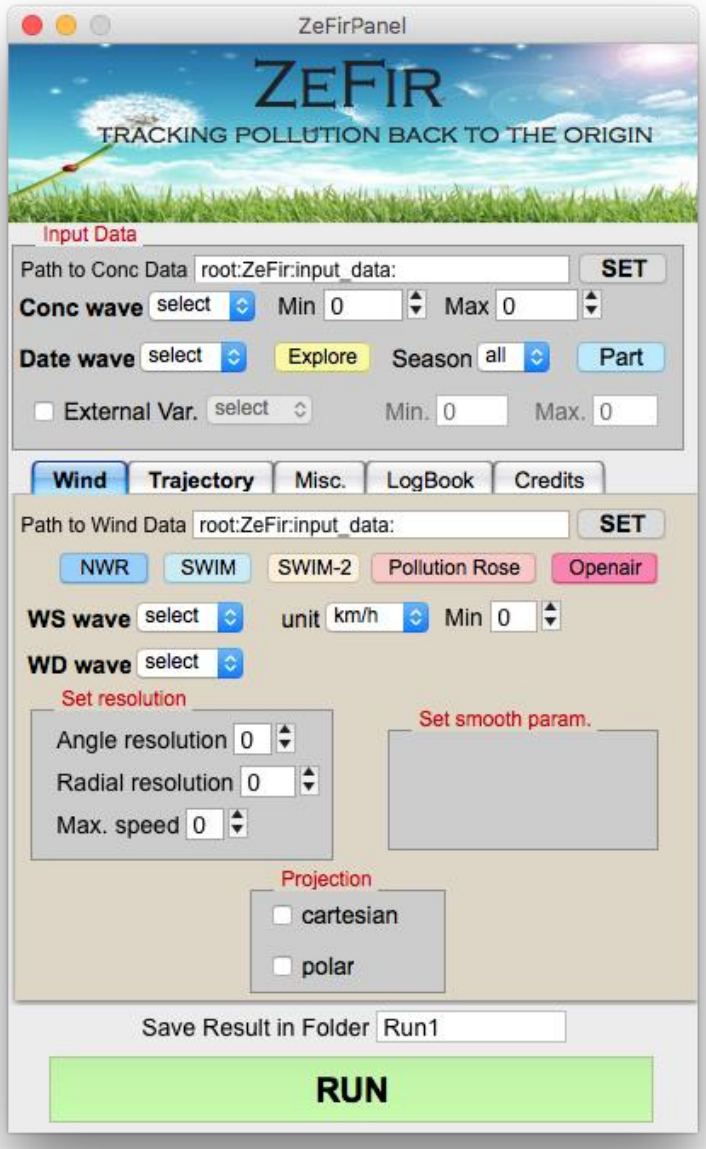
ZeFir

An Igor tool for geographical origins of atmospheric pollution



Guide prepared by Jean-Eudes PETIT.

For any bug report, comments or suggestions, feel free to contact: contact.zefir@gmail.com

ZeFir website: http://sites.google.com/site/zefirproject

Petit, J.-E., Favez, O., Albinet, A., Canonaco, F. : A user-friendly tool for comprehensive evaluation of the geographical origins of atmospheric pollution: wind and trajectory analyses, Environmental Modelling and Software, 88, 183-187, doi: 10.1016/j.envsoft.2016.11.022, 2017.

REQUIREMENTS:

- Igor Pro v.6.3 or higher (compatibility with Igor Pro 7 is not ensured yet).
- PC-based version of HYSPLIT for cluster analysis and backtrajectory calculation.
- R package with Openair installed for Openair calculation.

This manual refers to ZeFir v3.70

Table of content

1. Credits/Terms of use

ZeFir has been developed by J-E Petit.

Special thanks to:

- The Igor community for precious help
- Alexandre Albinet & Laurent Poulain for extensive testing and various suggestions.

Petit, J.-E., Favez, O., Albinet, A., Canonaco, F.: A user-friendly tool for comprehensive evaluation of the geographical origins of atmospheric pollution: wind and trajectory analyses. Environmental Modelling and Software, 88C, 183-187, doi: 10.1016/j.envsoft.2016.11.022, 2017.

ZeFir has been developed on Igor Pro v.6.37. Full compatibility with v6.2. or any prior version is not ensured, as well as Igor Pro 7; please contact zefir.contact(at)gmail.com for details. The Igor procedure is editable, meaning that the source code is available; however, it is not allowed to share a modified version of ZeFir without the consent of the developer. Any custom improvement in the code has to be apprised to zefir.contact(at)gmail.com

2. Introduction

2.1. Non-parametric Wind Regressions

Non-parametric Wind Regression techniques couples wind data (direction and/or speed) and pollutant concentration to alternatively highlight wind sectors that are associated with high measured concentrations.

The general principle is to smooth the data over a fine grid, so that a concentration could be estimated by any wind direction/wind speed couple. The smoothing is based on a weighing average where the weighing coefficients are determined by Kernel functions. The aim of the calculation is, for each angle θ (and speed υ), to give weight to concentrations that are associated with wind characteristics relatively close to θ or υ. First described by Henry et al. (2009)[1], NWR is defined as:

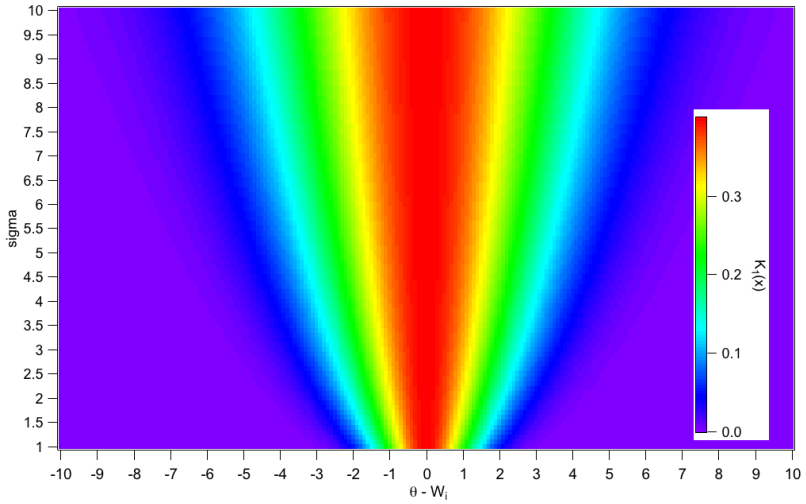
$$E(\theta|\upsilon) = \frac{\sum_{i=1}^{N} K_1\left(\frac{\theta - W_i}{\sigma}\right) \cdot K_2\left(\frac{\upsilon - \Upsilon_i}{h}\right) \cdot C_i}{\sum_{i=1}^{N} K_1\left(\frac{\theta - W_i}{\sigma}\right) \cdot K_2\left(\frac{\upsilon - \Upsilon_i}{h}\right)}$$

$$K_1(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-0.5 \cdot x^2}$$

$$K_2(x) = 0.75 \cdot (1 - x^2)$$

This concept of "relatively close" is endorsed by the smoothing parameters (σ and h) that widen or narrow the Kernel functions. In other words, the closer W_i is to θ, the higher the weighing coefficient will be, giving more weight to the associated concentration C_i. The figure below illustrates the weighing coefficients between ($\theta - W_i$) = {-10°; 10} and σ = {1; 10}.

[1] Henry, R. et al.: Source Region Identification Using Kernel Smoothing, Environ. Sci. Technol., 43(11), 4090–4097, doi:10.1021/es8011723, 2009.



It is important to note that that the coefficients obtained through Kernel K_1 is dependent of a difference between two angles, and thus corresponds to the acute or obtuse angle (i.e. below 180°). It has to be calculated in radians to take data sphericity into account (the angle difference between a_1=355° and a_2=15° is 20°, and not 240°).

$$a_1^{deg} - a_2^{deg} = \frac{180}{\pi} \cdot cos^{-1}\left(cos\left(a_1^{rad}\right) \cdot cos\left(a_2^{rad}\right) + sin\left(a_1^{rad}\right) \cdot sin\left(a_2^{rad}\right)\right)$$

One improvement of NWR, called SWIM-2 (Olson et al., 2012)[2], is based on the same equation, but a scalar weight is applied to C_i, following:

$$E(\theta|\upsilon) = \frac{\sum_{i=1}^{N} K_1\left(\frac{\theta - W_i}{\sigma}\right) \cdot K_2\left(\frac{\upsilon - \Upsilon_i}{h}\right) \cdot C_i \cdot S_i}{\sum_{i=1}^{N} K_1\left(\frac{\theta - W_i}{\sigma}\right) \cdot K_2\left(\frac{\upsilon - \Upsilon_i}{h}\right)}$$

Where $S_i = \frac{C_i \Upsilon_i}{\max(C_i \Upsilon_i)} \cdot \frac{\bar{\delta}}{\delta_i}$, δ being the wind direction standard deviation, $\bar{\delta}$ is the median. This scalar is used to downweight concentrations associated with unstable wind conditions.

[2] Olson, D. A. et al. : Determining source impacts near roadways using wind regression and organic source markers, Atmos. Environ., 47, 261–268, doi:10.1016/j.atmosenv.2011.11.003, 2012.

Finally, another approach, called SWIM (Vedantham et al., 2012)[3], uses wind direction and speed standard deviations at each t_i in place of fixed smoothing parameters.

$$E(\theta|\upsilon) = \frac{\sum_{i=1}^{N} K_1\left(\frac{\theta - W_i}{\delta_i}\right) \cdot K_2\left(\frac{\upsilon - \Upsilon_i}{\gamma_i}\right) \cdot C_i}{\sum_{i=1}^{N} K_1\left(\frac{\theta - W_i}{\delta_i}\right) \cdot K_2\left(\frac{\upsilon - \Upsilon_i}{\gamma_i}\right)}$$

Where γ_i is the wind speed standard deviation at t_i.

2.2. Openair

"Polarplot" is one of the most popular function of Openair, as it easily couples concentration and wind data. Several statistics are available, such as Conditional Bivariate Probability Function (CBPF, Uria-Tellaetxe and Carslaw, 2014)[4] and regular bivariate plots (Carslaw and Ropkins, 2012)[5]. The general principles of the calculation is to associate the concentration data into wind direction and wind speed bins. Then the grid is interpolated through with a Generalized Additive Model:

$$\sqrt{C_i} = \beta_0 + s(u_i, v_i) + \epsilon_i$$

where C_i is the i^{th} concentration, β_0 the overall mean of the response, s is an isotropic smooth function, u_i and v_i the wind components, and ϵ_i the residual.

2.3. Potential Source Contribution Function / Concentration-Weighted Trajectory / Concentration Field

Coupling concentration data and air mass history can be a powerful approach for the investigation of potentially advected pollution over large geographical scales. Most of available

[3] Vedantham, et al.: Combining continuous near–road monitoring and inverse modeling to isolate the effect of highway expansion on a school in Las Vegas, Atmospheric Pollut. Res., 3(1), 105–111, doi:10.5094/APR.2012.010, 2012.

[4] Uria-Tellaetxe, I. and Carslaw, D. C.: Conditional bivariate probability function for source identification, Environ. Model. Softw., 59, 1–9, doi:10.1016/j.envsoft.2014.05.002, 2014.

[5] Carslaw, D. C. and Ropkins, K.: openair — An R package for air quality data analysis, Environ. Model. Softw., 27-28, 52–61, doi:10.1016/j.envsoft.2011.09.008, 2012

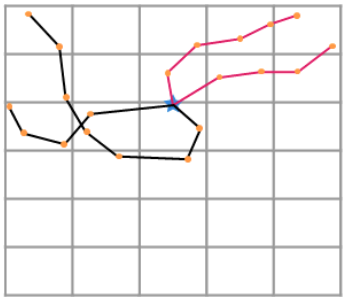
approach use a gridded map and trajectory residence time to evaluate potential zones responsible for high concentrations measured at a receptor site.

- *Potential Source Contribution Function (PSCF)*

PSCF specifically investigate the probability of an air parcel to be responsible for measured concentrations at the receptor site above a user-defined criterion. This criterion is generally a percentile value (75th or 90th):

$$PSCF_{ij} = \frac{m_{ij}}{n_{ij}}$$

Where n_{ij} is the total count of trajectory endpoints in the ijth cell, and m_{ij} the count of trajectory endpoints in the ijth cell associated to concentrations above the defined threshold.



matrix n

2	0	1	2	2
1	0	1	2	1
2	2	5	0	0
0	1	1	0	0
0	0	0	0	0
0	0	0	0	0

matrix m

0	0	1	2	2
0	0	1	2	1
0	0	2	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0

- *Concentration-Weighted Trajectory*

Instead of adding "1" in grid cells, another approach consists in summing the concentration values, weighted by the residence time.

$$\bar{C}_{ij} = \frac{1}{\tau_{ijk}} \cdot \sum_{k=1}^{N} C_k \cdot \tau_{ijk}$$

- *Concentration Field*

Finally, instead of using the arithmetic mean, concentration field uses the logarithmic mean of concentration.

$$log_{10}(\bar{C}_{ij}) = \frac{1}{\tau_{ijk}} \cdot \sum_{k=1}^{N} log_{10}(C_k) \cdot \tau_{ijk}$$

- Multi-site

Combining several sites together can be a very useful approach, from different sampling sites, leading to higher trajectory density values (Biegalski and Hopke, 2004)[6], or from different trajectory analyses (Han et al., 2007)[7].

$$MS_{ij} = \frac{\sum_l m_{ij}^l}{\sum_l n_{ij}^l}$$

[6] Biegalski, S. R. and Hopke, P. K.: Total Potential Source Contribution Function Analysis of Trace Elements Determined in Aerosol Samples Collected near Lake Huron, Environ. Sci. Technol., 38(16), 4276–4284, doi:10.1021/es035196s, 2004.

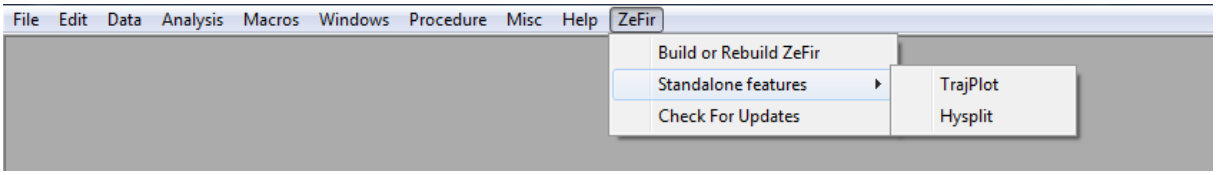
[7] Han, Y.-J. et al.: Estimation of source locations of total gaseous mercury measured in New York State using trajectory-based models, Atmos. Environ., 41(28), 6033–6047, doi:10.1016/j.atmosenv.2007.03.027, 2007.

3. ZeFir

The user needs Igor Pro installed on the computer. For more information about Igor, please refer to http://www.wavemetrics.net.

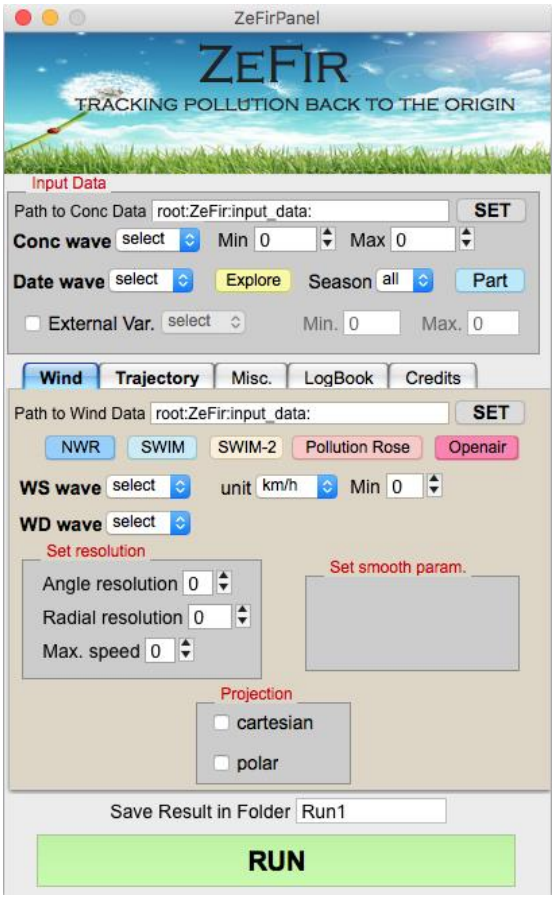
ZeFir can be loaded into Igor by double clicking the .ipf procedure file, or by placing a shortcut of the file (or of the folder containing it) in C:\Program Files\Wavemetrics\Igor Pro Folder\Igor Procedures\

Once loaded, ZeFir main panel can be brought up by clicking on the corresponding choice from the upper tabs ZeFir menu.



How to bring up ZeFir panel

Check for updates: In order to keep up with the latest version of ZeFir, the user can click on "Check For Updates" menu. An alert tab will pop up, displaying the version of ZeFir used. Then, Igor opens ZeFir website to display the latest version available. The user can manually check at https://sites.google.com/site/zefirproject/check-for-updates.



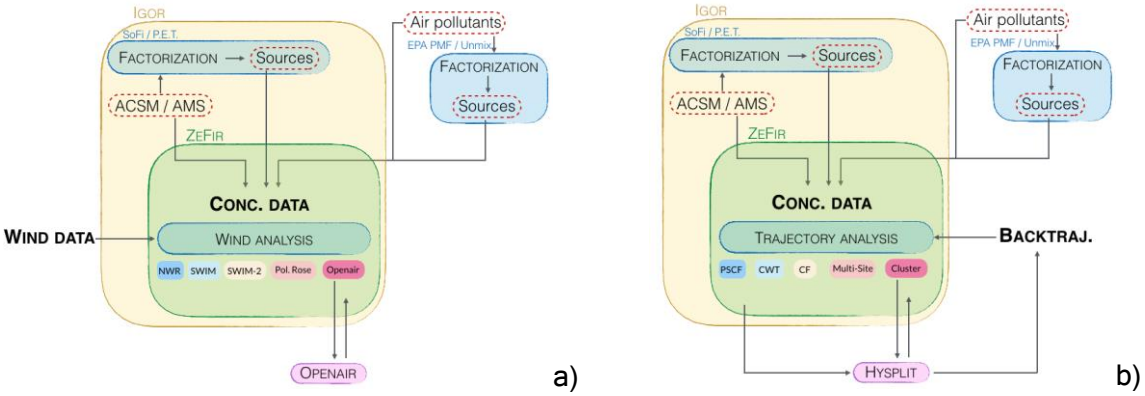
ZeFir main panel

ZeFir is divided into 2 distinct parts. The first section is dedicated to the input data, while the second one is fragmented through tab navigation.

This structure simplifies the use of global features that can be used whatever the analysis. Indeed, within the Input data groupbox, the user can set a bunch of thresholds (concentration and/or temporal and/or through an external variable). These options are pushed through wind and trajectory analysis. Specific features are available in each type of analysis because they can't be used in the other one. Then, the Run and other bottom buttons are also global and control any functions from the tab options.

In details, the structure of ZeFir is illustrated in the figure below. Igor and ZeFir scopes of action are respectively in yellow and green. Concentration data can be provided through i) data already present within Igor experiments (e.g. aerosol mass spectrometer chemical composition, or outputs from SourceFinder or PET); or ii) external data which can be manually imported, regarding air pollutants, or outputs from other source apportionment softwares. Manual import includes regular copy-and-paste operations, as well as dedicated built-in Igor panels

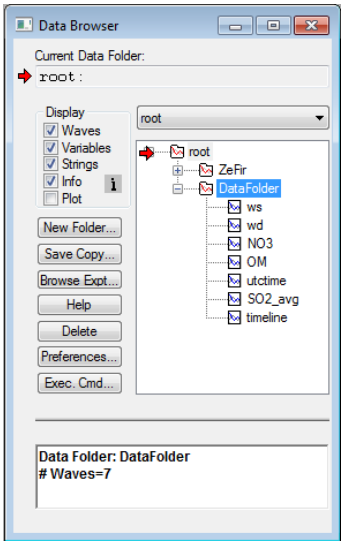
which allow the import of various types of files (e.g. Excel spreadsheets, text or NetCDF files).



Structure of ZeFir data process for a) wind analysis; b) trajectory analysis

4. Input Data

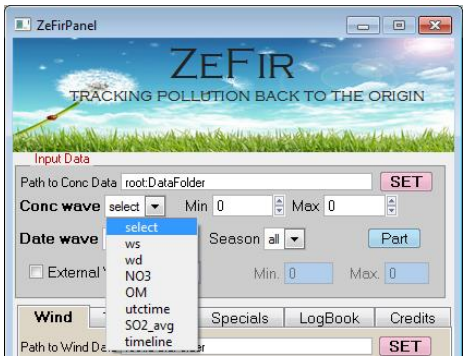
The user is first invited to select the folder where the input data are stored. To do so, the data browser needs to be displayed; if it is not the case, it can be brought up by selecting "Data Browser" in the "Data" menu. On the data browser, click on the folder to highlight it (see example below).



Data browser where the folder containing concentration data is highlighted

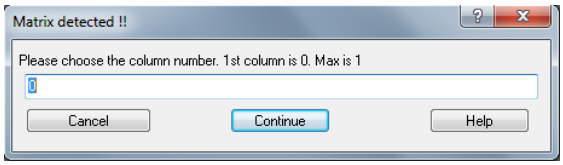
Press the "Set" button on the ZeFir panel. The correct path should be displayed on the panel.

When the path to the correct datafolder is set, the user can choose the concentration wave to use from the drop-down menu.



Drop-down menu for concentration wave selection

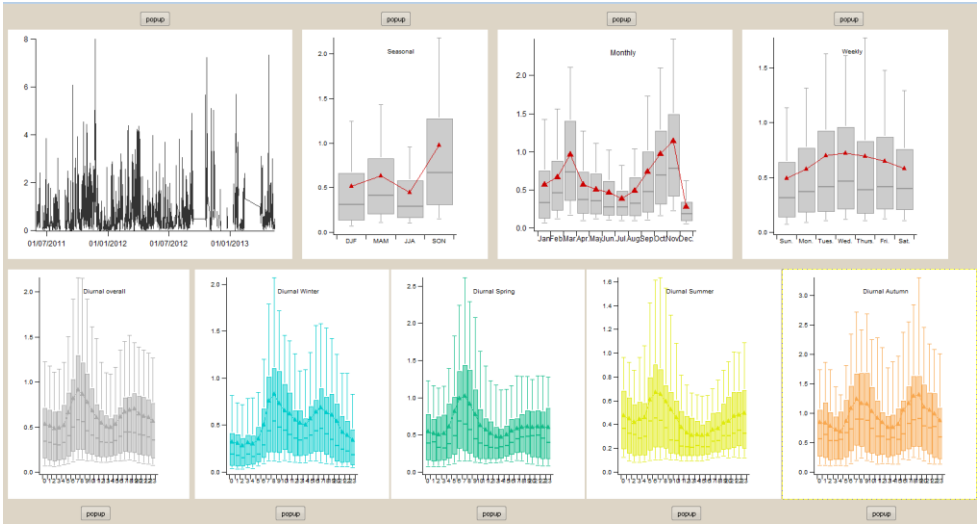
If the concentration wave is a matrix (i.e. several columns), then a message pops up and allows to choose the column that is needed for the analysis. This column is then treated as an independent 1-D wave.



The user can then set upper/lower limits for the data to use. When a concentration wave is selected, the thresholds displayed by default are the minimum and the maximum of the wave. By default, numbers should be entered, but it is possible to ask ZeFir to calculate percentile values. To do so, enter “pX” (without the “), where X is the centile to use (between 0 and 100). For example, to set the min threshold to the 10^{th} centile, type p10.

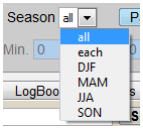
Similarly, the date/time wave needs to be selected. Once it is done, the user can explore the dataset by clicking the “Explore” button. It consists in calculating a bunch of statistical distributions of the data following different patterns (seasonally, monthly, weekly, and hourly). A dedicated panel pops up. 10^{th}, 25^{th}, 50^{th}, 75^{th}, and 90^{th} percentiles are used for the whiskers & box plots. Triangle markers are the average. Dedicated buttons allow to pop up each graph separately.

All the data are stored in root:ZeFir:ExploreVar:



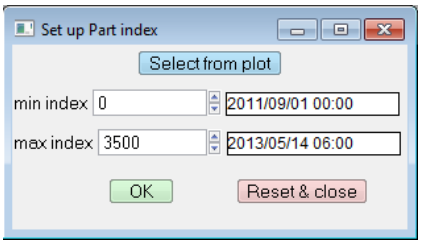
Explore Panel

For large dataset, it may be useful to cut the dataset by seasons. They are divided by month: DJF, MAM, JJA, SON.



By selecting “each”, calculation will be performed on the four seasons individually. Note that, so far, the four seasons have to be present in the dataset; otherwise, errors are expected.

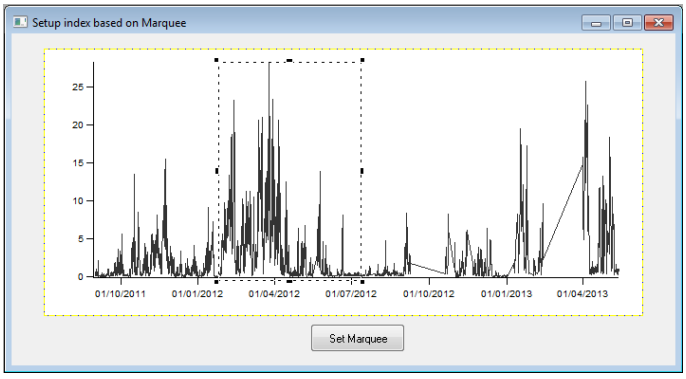
Another solution is to manually set a specific temporal window; this can be done by clicking on the “Part” button. It is independent from the seasonal threshold option, and they cannot be combined: if one is activated, the other is automatically disabled. A new panel pops up:



Part panel

The user can manually go through minimum and maximum indices (corresponding dates will be displayed). By clicking “OK”, the specified indices are saved, and will be used in the calculations. By clicking “Reset & close”, the indices are re-initialized to their original values.

Clicking the “Select from plot” button will display a new panel showing the concentration data as a graph. The user can then draw a selection rectangle representing the limits of the temporal window. Clicking the “Set Marquee” button will pull the corresponding indices to ZeFir.



“Select from plot” panel. The marquee drawn delimits the temporal window that will be used.

IMPORTANT: Indices are re-initialized as soon as the “Part” button is pressed

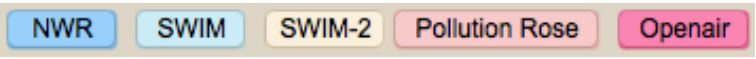
Note that when indices differ from their default values, the “Part” button is colored in red, meaning that this feature is active. When indices equal default values, “Part” is colored back in blue.

Not active	Active
	

5. Wind analysis

5.1. Type of analysis

The user needs to select the type of analysis to perform, by clicking on the corresponding button to activate specific parameters.



Depending on the analysis, as different parameters are needed, ZeFir adapts the Wind analysis sub-panel.

5.2. Wind data

By default, the path to the wind data is the same as the concentration data. If it is not the case, follow the same procedure described previously.

In case of NWR, SWIM, SWIM-2 & Openair, a wind speed (ws) wave needs to be selected from the drop-down menu. The unit of ws can be selected (km/h or m/s). Please note that the results will always be plotted in km/h. A minimum speed threshold can be entered and should have the same unit as the wind speed wave. This will filter the input data so that only data associated with wind speed > threshold will be used in the calculation.

In case of Pollution Rose, no wind speed is needed.

Select the wind direction (wd) wave. The data must be in degrees (conversion from radian: wave_deg=wave_rad*pi()/180), and 0° (360°) represents the North azimuth (rotate the data: mod(wave+rotation_angle,360)).

5.3. NWR / SWIM / SWIM-2 parameters

5.3.1. Set resolution

The user is invited to set the angle resolution. It basically sets how many angle values will be used in the calculation: an angle resolution of 0.5° corresponds to 720 angle values.

Similarly, the radial (speed) resolution and maximum speed (in km/h) should be set.

Be aware that the higher the resolution, the more time it will take to perform the calculation.

5.3.2. Set Kernel parameters

In case of NWR and SWIM-2, fixed smoothing parameters (σ and h) have to be set. Their values can be empirically set, as small changes should not lead to discrepancies in the interpretation of the results. Otherwise, an estimate of both parameters can be suggested by clicking on the « suggest estimate » button. The estimation calculation is based on the standard deviation of the wind data (Yamartino method); but usually leads to over-smoothed results.

IMPORTANT. **Be aware to use smoothing parameters > 1.**

In case of SWIM-2, the wind direction standard deviation wave must be selected.

In case of SWIM, σ and h are replaced by wind standard deviation data. The corresponding waves needs thus to be selected.

5.3.3. Projection

Check the box for which kind of plot should be displayed. (polar and/or cartesian coordinates).

5.4. Pollution Rose parameters

For pollution roses, only 3 parameters are needed. First, the width angle bins; regular values typically range from 10° to 30°. Then, the width of the concentration bin; this determines the

number of concentration classes, and the maximum concentration sets the last concentration bin.

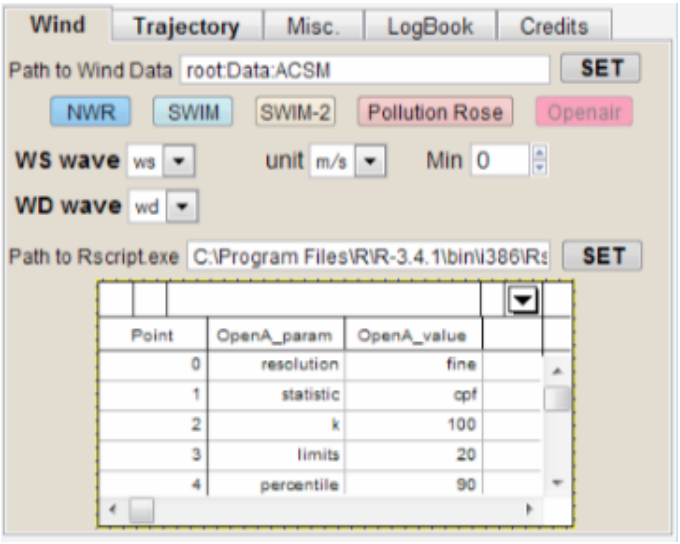
5.5. Openair

ZeFir allows the possibility to use the "Polarplot" function of Openair through R-scripting.

SPECIFIC REQUIREMENTS: R must be installed on the computer. Regular R package should include the Rscript.exe program (see https://www.r-project.org for details). Openair package must also be installed (see http://www.openair-project.org for details). Finally, **Igor needs to be launched in administrator mode**.

Similarly to NWR/SWIM/SWIM-2, wind speed and direction waves need to be selected. Then, the path to the Rscript executable must be set. For 32-bit computers, it is generally located in C:/Program Files/R/R-3.4.1/bin/i386/ .

The table underneath summarizes the parameters which the user can change. Changes should be made in the "OpenA_value" column.



resolution: "fine" or "normal"

statistic: "mean", "median", "max", "weighted.mean", "cpf" shall be used

k: any number (100 is recommended)

limits: maximum wind speed value (in km/h)

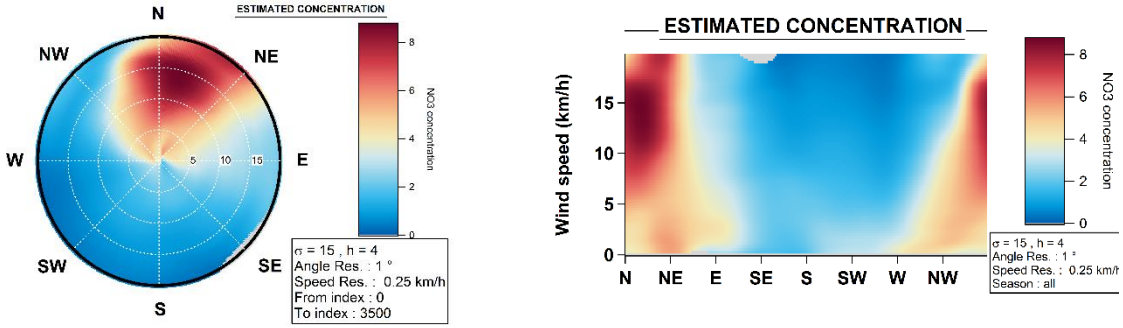
percentile: can be a single value, e.g. “75”, or a range, e.g. “c(50,90)”

5.6. Results

5.6.1. NWR/SWIM/SWIM-2

After the calculation, two graphs are generated, the estimated concentration and the joint probability.

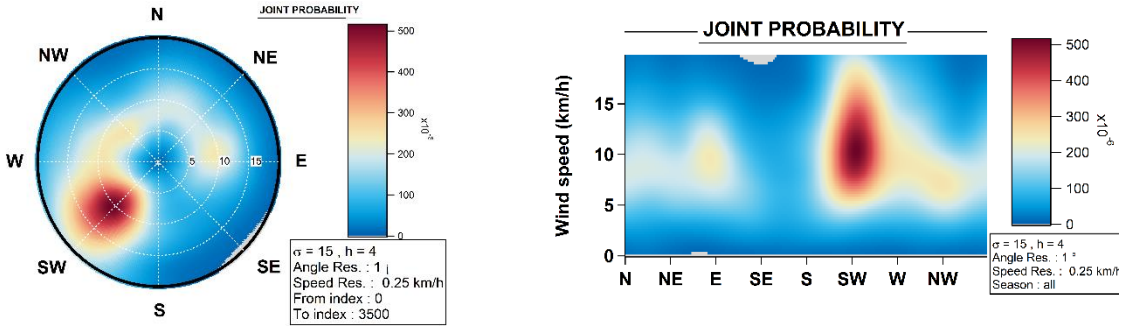
The estimated concentration:



a) polar graph b) rectangle graph

This one represents the actual results of the NWR calculation, with estimated concentration for any wind speed and direction.

The joint probability:

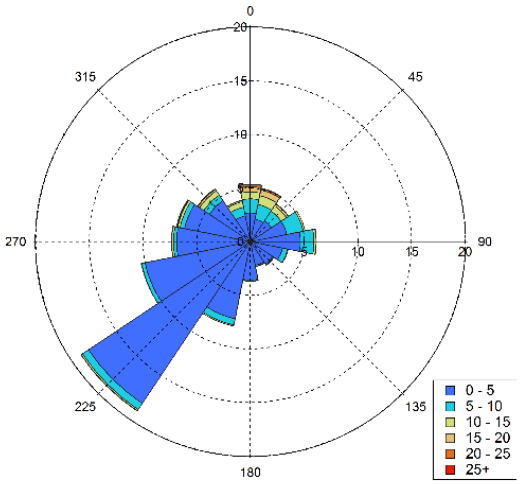


a) polar graph b) rectangle graph

This one represents the joint probability of the wind data, and is thus an equivalent to a wind rose, although the absolute values do not correspond to an occurrence.

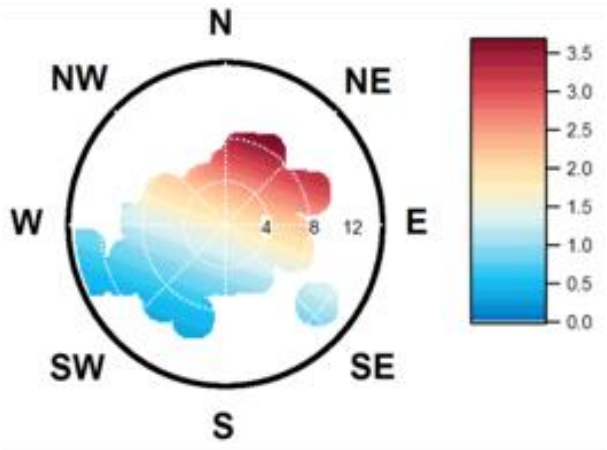
5.6.2. Pollution roses

For pollution roses, only one graph is generated.



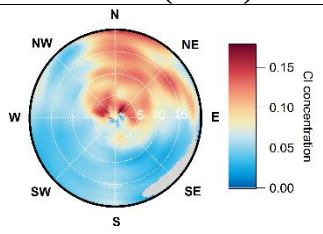
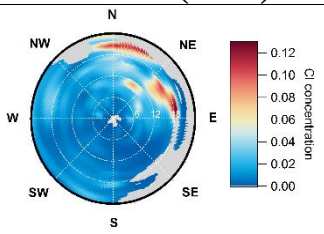
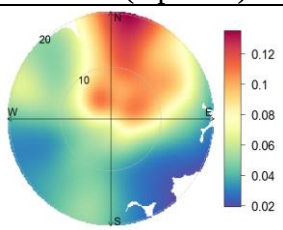
5.6.3. Openair

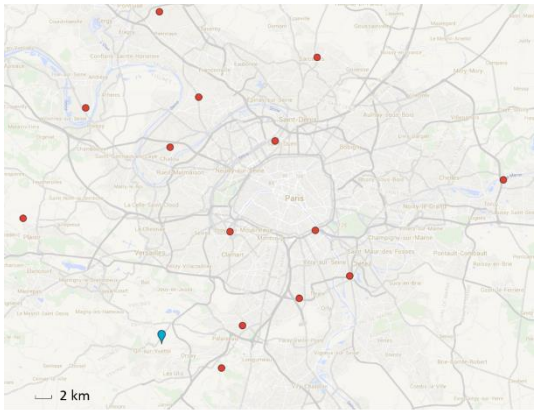
For Openair calculation, ZeFir gets the result data back into Igor, and plots it following the same procedure as a NWR polar plot. The joint probability (wind rose) is not automatically plotted here, but is accessible by using the “frequency” statistic in another run.



5.7. Wind analysis examples

1- Petit et al. (2017). The following example illustrates different wind analyses applied to 3-h non refractory Chloride concentrations measured between 2011 and 2013 in Paris, France, by ACSM measurements. NWR, SWIM-2 and CBPF (Openair) have been applied.

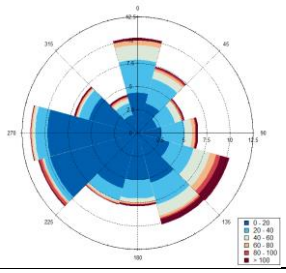
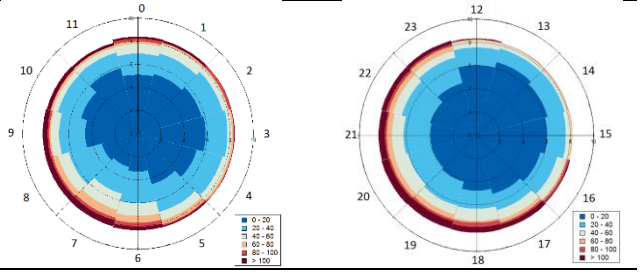
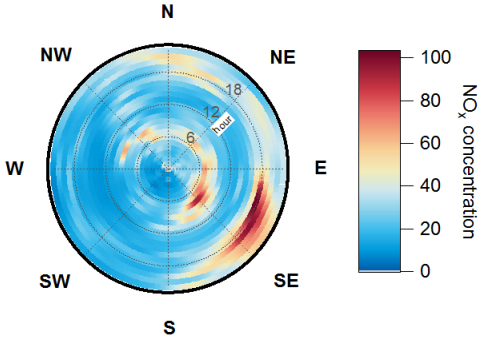
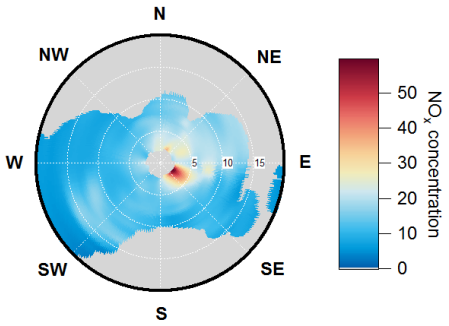
NWR (ZeFir)	SWIM-2 (ZeFir)	CBPF (Openair)
		



Location of incinerators around Paris (red dots), and sampling site (blue marker).

The discrepancies observed on the 3 methods underline the necessity of having several methodologies available.

2- Four ways to plot 1-h NO_x data from Metz, France, in 2015.

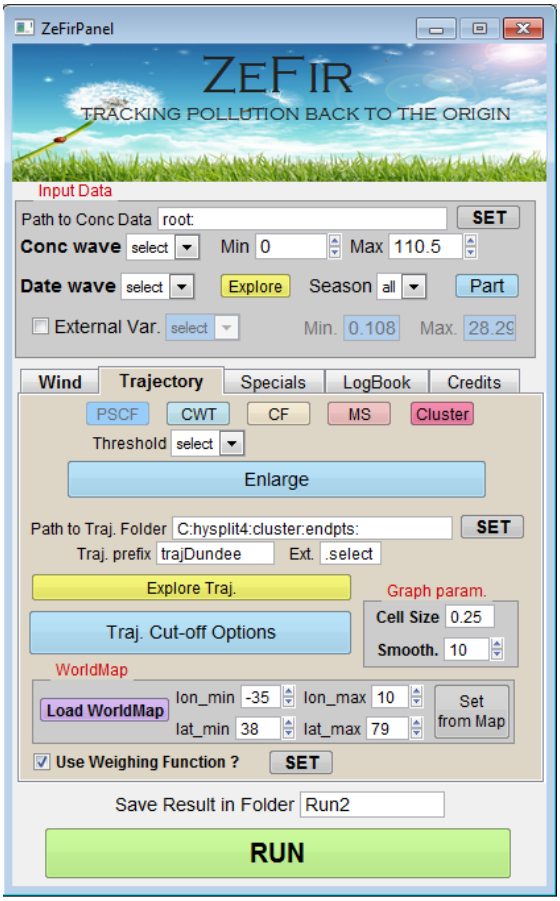
Pollution rose	Clock rose
	
Annular plot	**NWR**
	

The Clock rose is obtained by using the “Pollution Rose” statistic, with WD Wave being angles representing hours (i.e. here, WD=mod(hour*30,360)). The dataset have been divided into subsets following an external variable, being the hour values between 0 to 11h, and 12h to 23h.

The Annular plot is similar to what can be done with Openair. It is a NWR calculation, where wind speed is replaced by the hours. There is thus information about diurnal variation, and wind direction dependence.

This example puts the emphasis on the fact that they are many ways possible to explore any kind of dataset, some of them being available in ZeFir. The user should not hesitate to go beyond the proposed features. For example, ambient temperature or RH can be used as a “pollutant” to illustrate its wind dependence, or as bivariate variables using Cartesian coordinates.

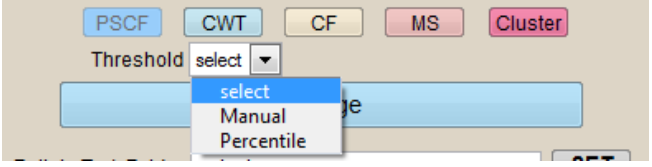
6. Trajectory analysis



Trajectory analysis sub-panel

6.1. Type of analysis

The user needs first to click on the analysis button of interest. If PSCF is selected, then, a dropdown menu pops up, and allows the user to choose between a manual, or a percentile threshold.

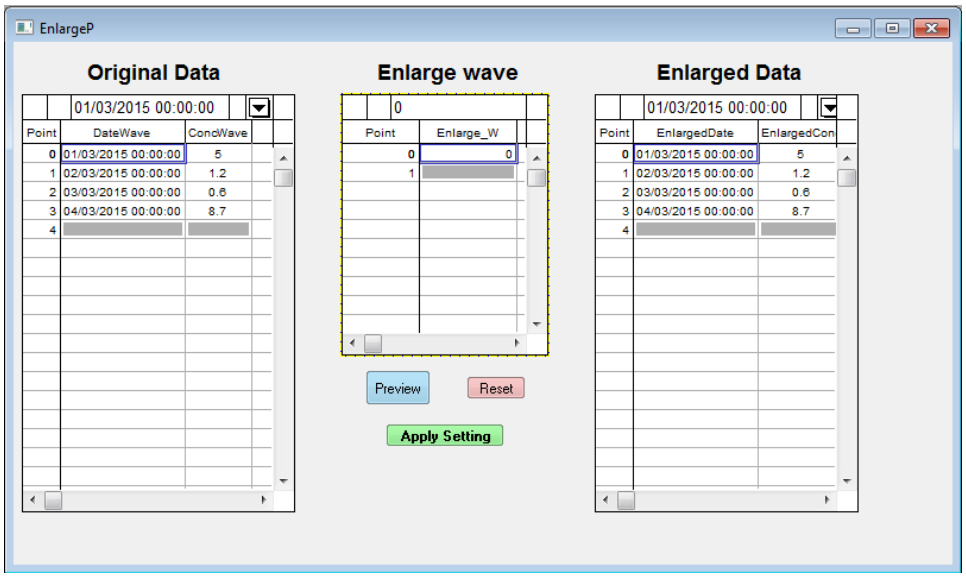


6.2. PSCF / CWT / CF

6.2.1. Input Data

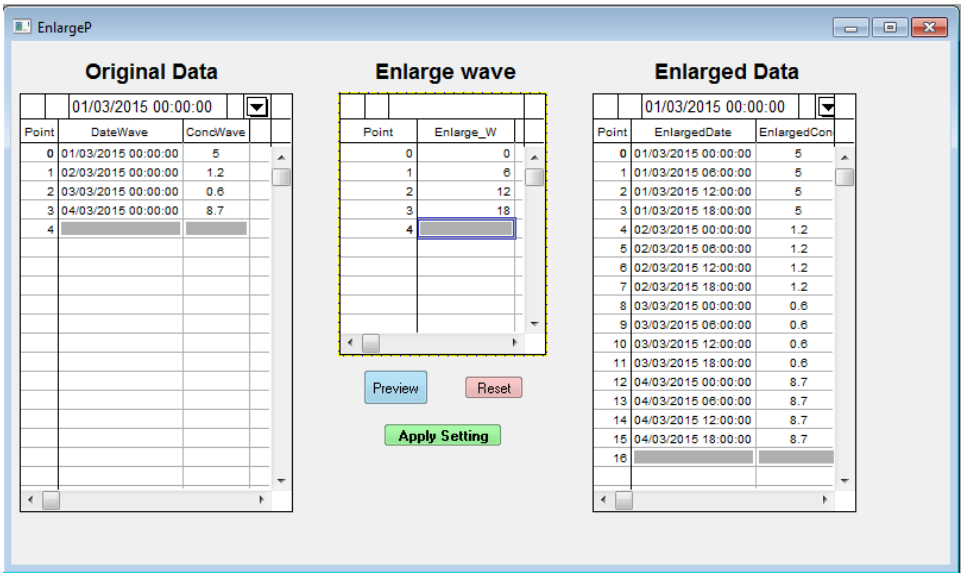
The user has the possibility to enlarge the dataset if its temporal resolution is low. This could be very helpful for e.g. 24-h filter sampling, allowing to take more backtrajectories into account. By clicking on the “Enlarge” button, a panel is displayed to edit the coefficients to use. These coefficients are expressed in hour.

By default, the enlarge wave has only one row initialized to 0. The panel displays the original dataset (left), the enlarge wave (middle), and the enlarged data (right). In the following example, the original dataset consists in daily measurements during 4 days, between March 1st and March 4th of 2015.



Enlarge Panel when brought up.

In order to take more backtrajectories during the day (and not just the one arriving at midnight) into account, the enlarge wave can be set to e.g. {0, 6, 12, 18}. See the results below.



Enlarge Panel with customized enlarge wave.

Negative values are accepted: if the datetime of the original dataset represents the mean datetime, the enlarge wave accepts negative values in order to be centered at 0: eg {-6; -3; 0; 3; 6}

After editing the enlarge wave, a preview of the enlargement can be seen by clicking the "preview" button. "Reset" will re-initialized the enlarge wave; "Apply Settings" closes the panel.

Again, the button color moves to red as soon as it is activated.

Not active	Active
Enlarge	Enlarge

The path of the folder containing the trajectory files has to be chosen by clicking the SET button. A windows browser pops up. These files are the raw outputs of Hysplit.

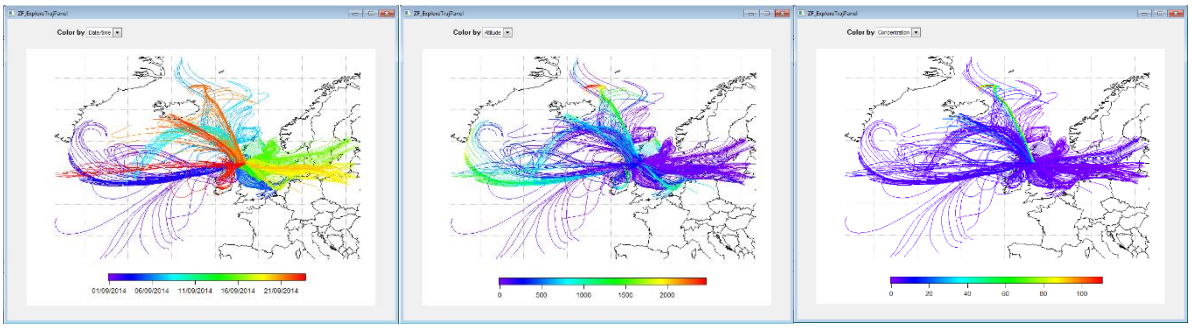
Trajectory file names are basically based on a fixed prefix and incremented date of arrival. Exemple: the file "randomtraj_15010406", where the prefix is "randomtraj_", correspond to the backtrajectory arriving on the 04/01/2015 at 06:00. Following this, the traj. Prefix needs to be set.

6.2.2. Explore Trajectories

If necessary, the trajectories corresponding to the dataset can be loaded within the Igor experiment (they won't be used during trajectory analysis), and plotted.

All different filters previously set apply, except for seasons, where "each" is not accepted.

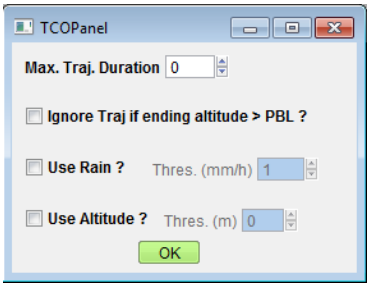
When "Explore traj." is pressed, a panel pops up and shows all backtrajectories, color-coded following date/time. The colorcode can be changed to concentration or altitude.



Example of trajectory plot with date/time (left), altitude (middle) and SO_2 concentration (right) as a color code.

Manual trajectory import can be done with the TrajPlot Tool, included in ZeFir (see section 9 in this manual).

6.2.3. Trajectory cut-off



“Max. Traj. Duration” allows to select parts of input trajectories depending on a maximum temporal duration (in hour). For examples, if 120-h backtrajectory files are available, it is possible to keep only the 2 first days by entering “48”.

“Ignore traj if ending altitude > PBL”. This option allows to consider only trajectories whose ending altitude is lower than the ending PBL value. For example, a trajectory would be excluded if its ending altitude is 500m and the ending PBL is 231m. Note that PBL data must be present in the trajectory file (“MIXDEPTH” column).

Rain data can be added in the trajectory files prior the Hysplit calculation. To do so, check the “precipitation” box in “Advanced -> Configuration Setup -> Trajectory -> Add METEOROLOGY output along trajectory” from the HYSPLIT menus. ZeFir automatically detects the rain column, as long as the “RAINFALL” header is present in the file (is written by default). The default threshold is 1 mm/h (0.1 has been shown to be a bit too drastic). If “Use rain” is checked, but the column is not found, ZeFir does not apply any restriction on the corresponding trajectory.

Similarly, cut-off can be performed by setting a maximum altitude.

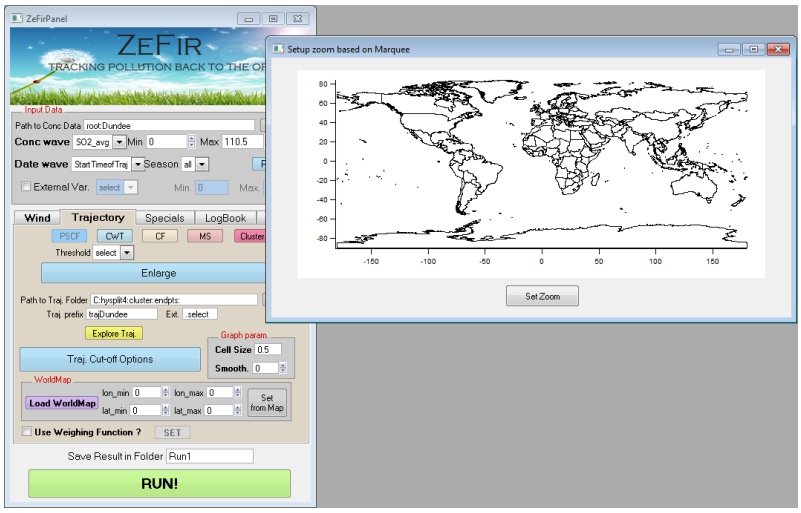
6.2.4. Graph parameters

The cell size should be set in degrees, and represents the size of each air parcel. Typical values range from 0.1° to 2°.

The smooth factor is the strength of the Gaussian smoothing performed after calculation (row-wise and column-wise). Typical values range from 0 to 40, but essentially depend on the cell size.

Then, a WorldMap must be loaded. ZeFir comes with 2 sets of WorldMaps taken from NaturalEarth Data. The user can use customized data, but must be in the form of a 2D Wave called “W_Geometries” located in root:ZeFir:Traj_A:variables:. For North-American users, where State and Province boundaries can be important information, another file (CountryStateProvince.itx.zip) can be retrieved directly from the download page of ZeFir website (https://sites.google.com/site/zefirproject/download).

As the geographical window of interest may be narrower than the entire world, a zoom can be set by directly entering the desired coordinates, or by interactively selecting them from the WorlMap by clicking on the “Set From Map” button. Simply draw a rectangle on the WorldMap, and click “Set Zoom”. The marquee variables will thus be pulled to the main panel.

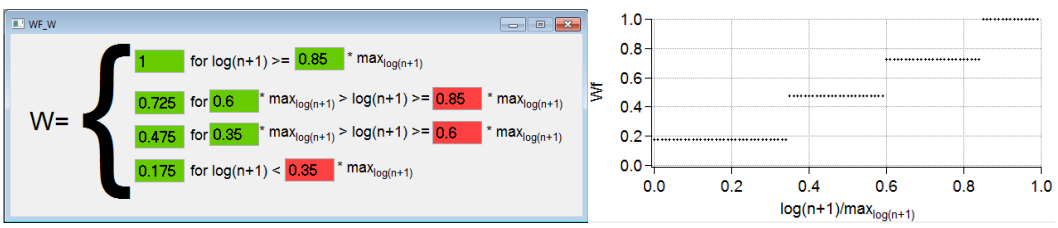


WorldMap displayed to set the zoom

6.2.5. Weighing function

In order to consider only air parcels with good representativeness, a weighing function can be added to the calculation to downweight cells associated with low values of n. The parametrization of the different weights is empirical, and thus usually increases the time required to get to final results.

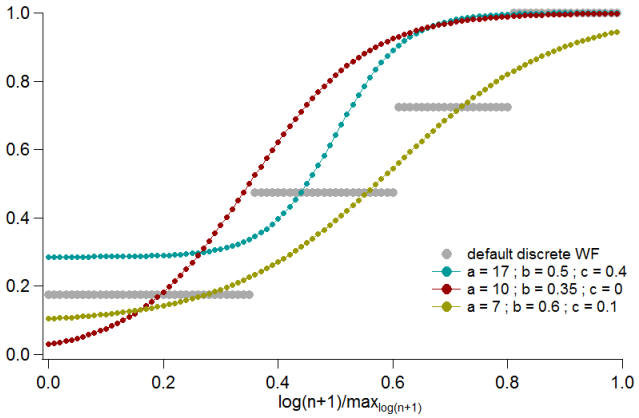
Original approaches are based on the average value of n. But as described in Waked et al. (2013) and Bressi et al. (2014), the weighing function can be instead based on the trajectory density, by calculating log(n+1). As of now, discrete functions are used, where fixed (user-defined) coefficients are set over user-defined ranges of n (or log(n+1)) values, like illustrated in the figure below, which is the default discrete weighing function in ZeFir.



This approach results in the empirical determination of 7 parameters if 4 ranges are used, which is usually found in the literature.

Another solution is to apply a continuous function that would have similar "shape", but with less parameters to control. Sigmoid functions have an interesting potential as they describe the transition between two states. The following equation describes the sigmoid function used in ZeFir, where a, b and c are the only 3 coefficients to determine.

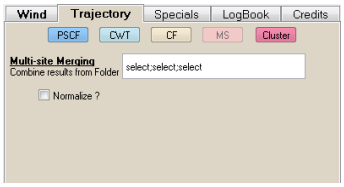
$$W = \frac{1}{(1+c)\cdot(1+e^{-a\cdot(x-b)})} + \frac{c}{1+c}$$



a represents the "strength" of the transition slope; b the x position of half transition ($(1+y_0)/2$), and c controls the y-intercept y_0.

The user can thus build various weighing functions, similarly to what is generally done in the literature, but with only 3 parameters.

6.3. Multisite merging



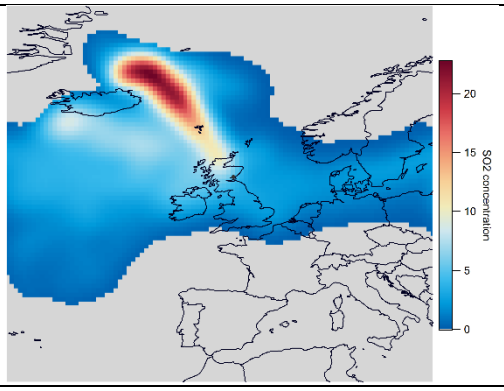
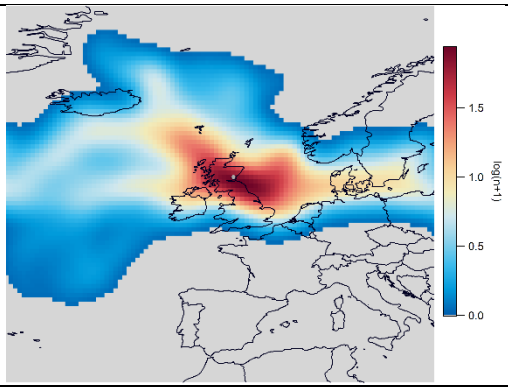
Multi-site merging can be a powerful solution to explore the bigger picture of source-receptor approaches, by combining several sites all together. This is made very easy within ZeFir.

First, individual calculations for each site need to be performed, and saved in specified result folders. Then, the user has to fill the “Combine results from Folder” list by the names of all result folders. If the “Normalize” button is checked, individual results are normalized by the maximum before merging.

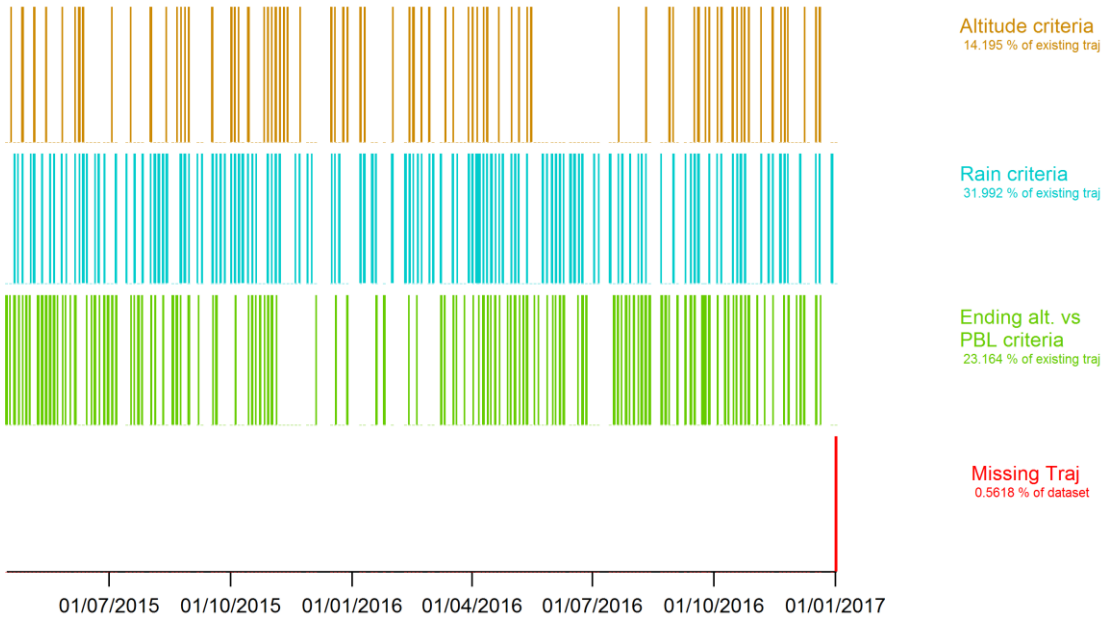
6.4. Results

After all the parameters have been setup, calculation may be eventually performed by clicking on the “RUN” button.

After calculation, two graphs are produced: the trajectory density, and the PSCF/CWT/CF map.

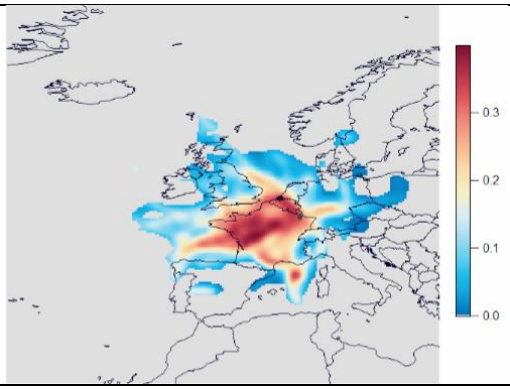
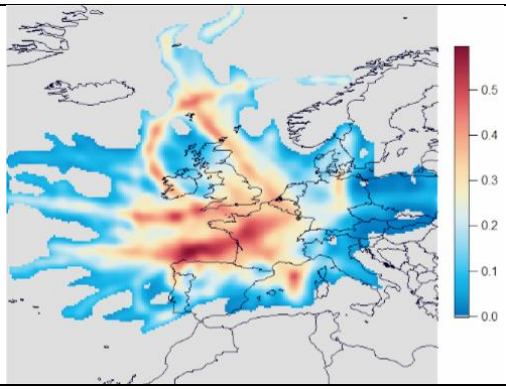
CWT result	Trajectory density
	

A third graph is automatically created, and sums up the occurrence of the criteria that have been set in Trajectory Options. This provides a good overview of what ZeFir is actually doing during the analysis.



6.5. Examples

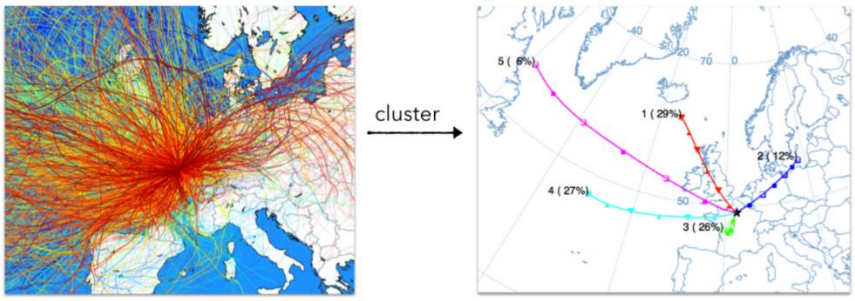
The following example illustrates the impact of data enlargement and the importance of statistical representativeness of backtrajectories. Sodium ion concentrations have been obtained from high volume filter sampling in Metz, East of France. Sampling has been performed daily, 1d/3, during 2015. From previous studies, we can reasonably expect sodium to mainly originate from sea salt, and thus to show a clear oceanic origin. However, when using the original dataset, this origin is not obvious, as results show highest concentrations in French inlands. But adding backtrajectories at +3h, +6h, +9h, 12h, +15h, +18h and +21h greatly helped to highlight the oceanic origin of sodium ion.

Without data enlargement	With data enlargement
	

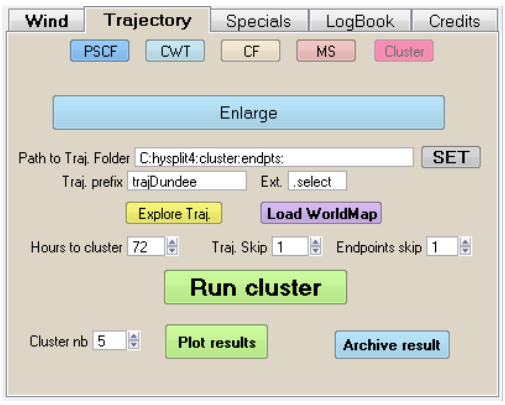
Petit et al. (2017)

6.6. Cluster analysis

Cluster analysis aims at regrouping backtrajectories that have similar pathways. The example below shows, on the left, 3-h backtrajectories ending in Paris between 2011 and 2013. The result is quite messy, and no interpretation can be done at this stage. On the left, clustering analysis led to the identification of 5 clusters, and describes more globally the air mass variability over Paris.



Cluster analysis is enabled in ZeFir through the “Trajectory” sub-panel. It is important to note that cluster calculation is not performed within the Igor experiment. Just like the ME-2 engine, ZeFir sets up batch files and uses C scripts available through the PC-based version of HYSPLIT. So the same calculation can be performed through the HYSPLIT GUI; but ZeFir automatically loads, processes and plots the associated results in Igor.



Similarly to “regular” trajectory analyses, the path to the folder containing trajectory files needs to be first set. **However, here, space character is not allowed**. For example, the path “C:Users:je.petit:Hysplit traj:” will not work. That is why it is strongly recommended for cluster analysis to place the trajectories in C:/Hysplit4/cluster/endpts/.

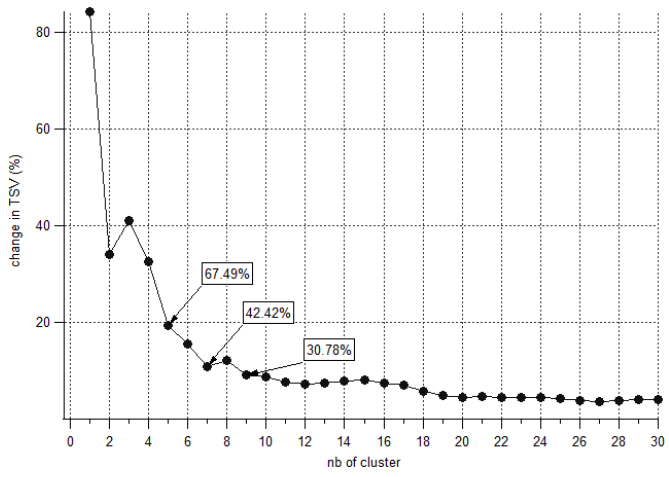
The other parameters can be set, similarly to how it is done on the HYSPLIT GUI. From the HYSPLIT user guide:

“*Hours to cluster: Trajectory durations up to the given hour are used.*

Traj. skip: Identifies which trajectories in a folder to use. A value of 1 means every trajectory will be used; 2 means every other trajectory; 5 every 5th trajectory, etc. Useful with very large sets of trajectories

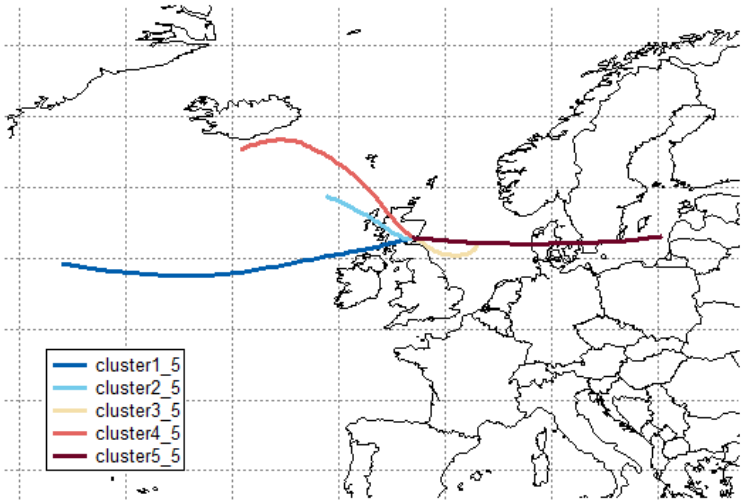
Enpoints skip: Identifies which endpoints along a trajectory to use. Typically every hourly endpoint is used (1). For long trajectories, skipping endpoints will save computational time.”

When pressing “Run Cluster”, ZeFir builds all the necessary batch files and launches the calculation. At the end, a graph pops up in Igor showing the variations of Total Space Variance (TSV) following the number of clusters used. The choice of the optimal number of cluster is indicated by tags.

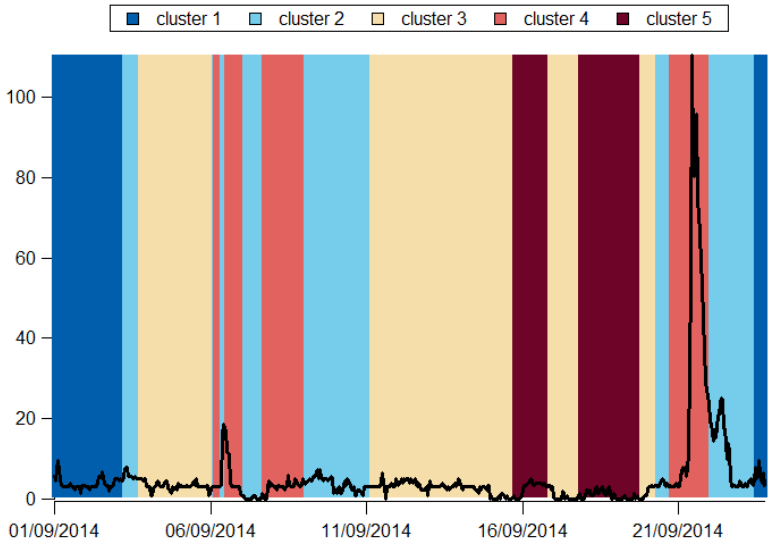


The number of clusters to use needs then to be entered; and the button “plot results” shall be pressed. After calculation, 3 graphs appear.

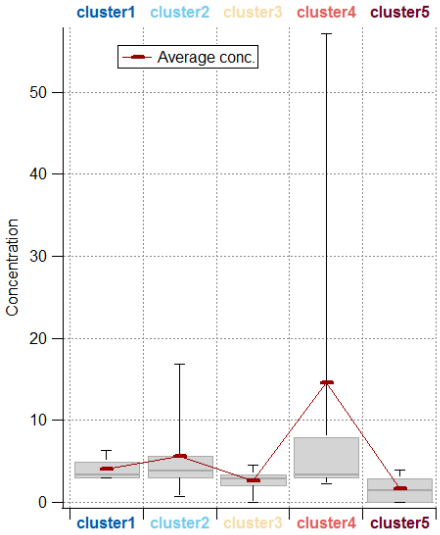
1) Mean trajectories. Colors are consistent throughout the results.



2) Concentration timeseries with the background color depending on the cluster.

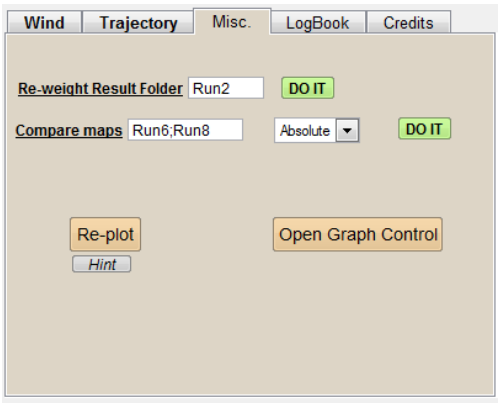


3) Average and Concentration distributions for the different clusters (10^{th}, 25^{th}, 50^{th}, 75^{th}, 90^{th} percentiles).



After calculation, by pressing the archive button, all the result waves are stored in the specified result folder.

7. Specials



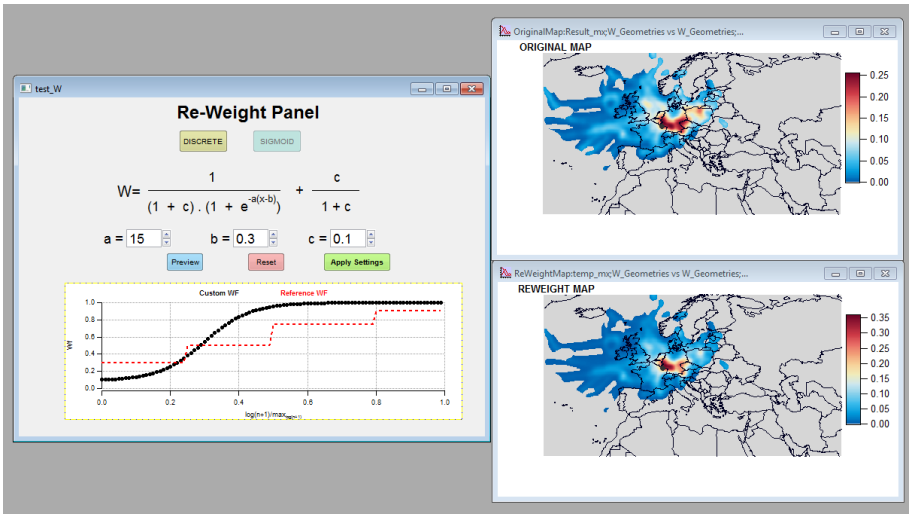
7.1. Re-weight Result

As the weighing function used in trajectory analysis is empirical, it can take a lot of time and runs to get a weighing that fits the need of the analysis. It would be much easier to run a calculation once, and adjust the weighing function “in real time”. Thus, a re-weighing panel has been implemented in order to greatly facilitate this work.

Enter the name of the result folder, and press “DO IT”. A new panel pops up, as well as 2 independent graphs showing the original run and the modified map.

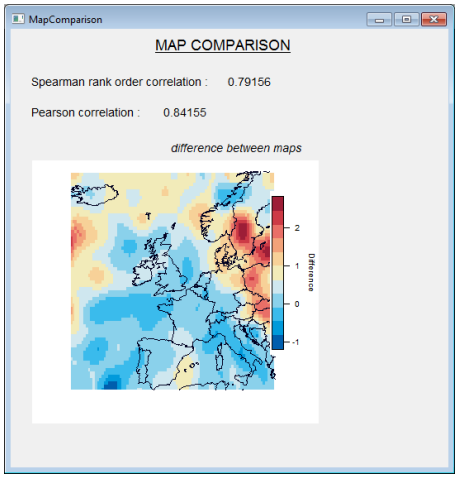
The user can change from discrete to sigmoidal weighing function. The “Preview” button enables to view the result of any modification, both on the weighing function graph, and on the reweight map. Note that the Original map should not change. The “Reset” button re-initialized the weighing parameters to their original values. If no weighing function have been used in the original run, the function is set to 1.

“Apply Settings” replaces the original parameters with the ones set in the reweight panel, and eventually closes all related windows.



7.2. Compare Maps

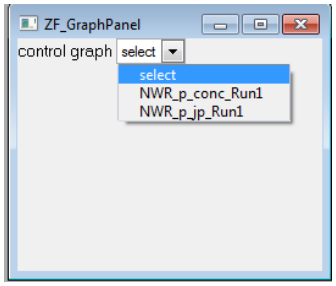
It might be useful sometimes to quantify the change in a result with an adjustment of a parameter, or simply to compare two different results. ZeFir has an implemented feature to perform such things. By entering the name of two result folders, ZeFir plots the difference of the two maps. The Spearman and Pearson correlation coefficients are also displayed.



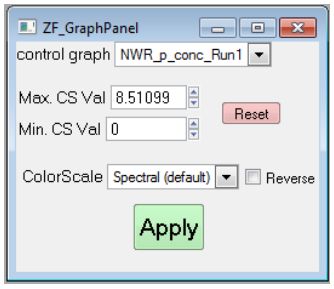
Note that the two maps need to have the same dimension: same resolution, and same lon/lat frame.

7.3. Graph Control

By default, a customized colorscale is used (interpolated from the Igor built-in EOSSpectrum11 colortable), with limits corresponding to the miminum and maximum values of the matrix. It can happen that these default settings are not adapted to specific work/result. Changing these parameters is made very easy with ZeFir. To do so, click on the "Open Graph Control" button.

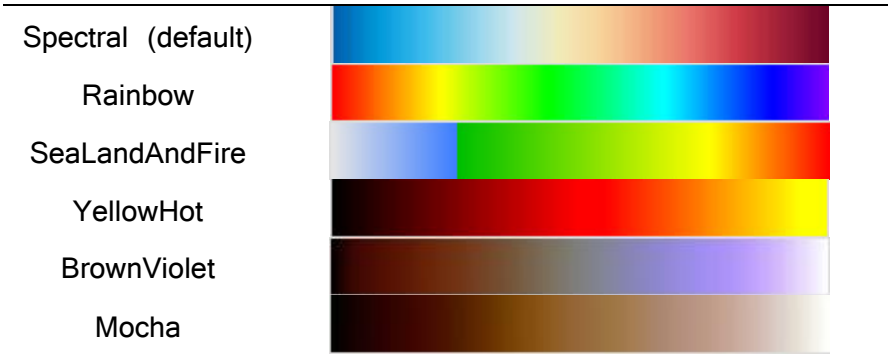


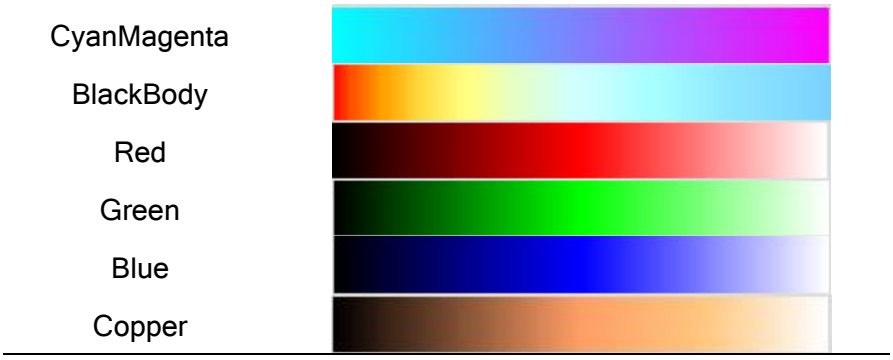
When selecting a graph, control options are displayed on the panel.



The user can enter manual max and min values for graphical considerations, and press "Apply". To reset the values by default, press "Reset" and then "Apply".

Then, a selection of colortables (see below) is available for change if the default one is not adapted. Select the colortable, and press "Apply". To reverse the selected colortable, check the "reverse" box. Note that the spectral (default) colortable is not reversible.





IMPORTANT. Controlling the colormap for pollution roses is so far not possible. It has to be done manually from the polar graph panel. Similarly for cluster results, the ClusterCS wave can only be changed manually.

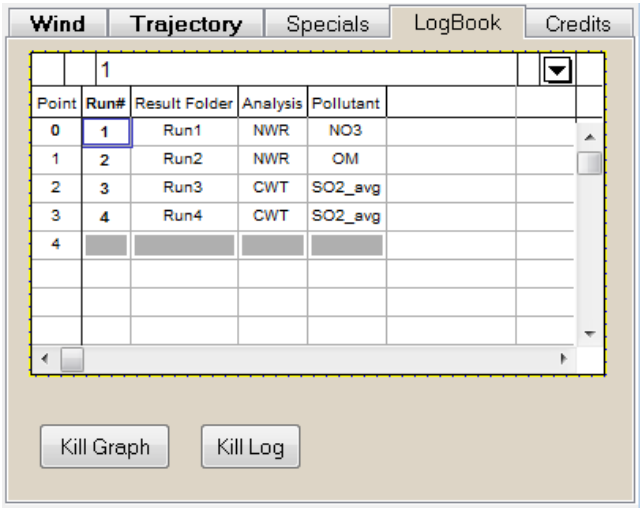
7.4. Re-plot previous calculation

It can happen that for some reason the user needs to replot results from previous run. All the results are stored within root:ZeFir:results. In the data browser, slide the red arrow to the result folder of interest. And click “Re-plot”.

ZeFir automatically detect the type of analysis that was performed.

8. Logbook

A logtable is displayed under the Logbook tab, and shows all the runs performed with some parameters used.



IMPORTANT: When clicking on "Kill log", the logtable is deleted as well as all the results stored in the "root:ZeFir:results:" data folder

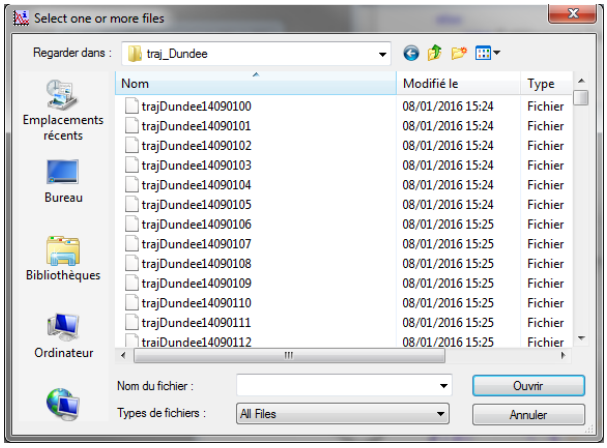
The generated graphs don't have to be closed before another calculation. This is especially useful if the user needs to compare the results from different seasons/pollutants. Closing all ZeFir-related graphs can be performed by clicking the "Kill Graph" button. This will not kill other graphs.

9. TrajPlot

TrajPlot is an Igor tool to interactively plot trajectories from HYSPLIT.

To build the panel, select “Build Panel” from the “TrajPlot” menu bar.

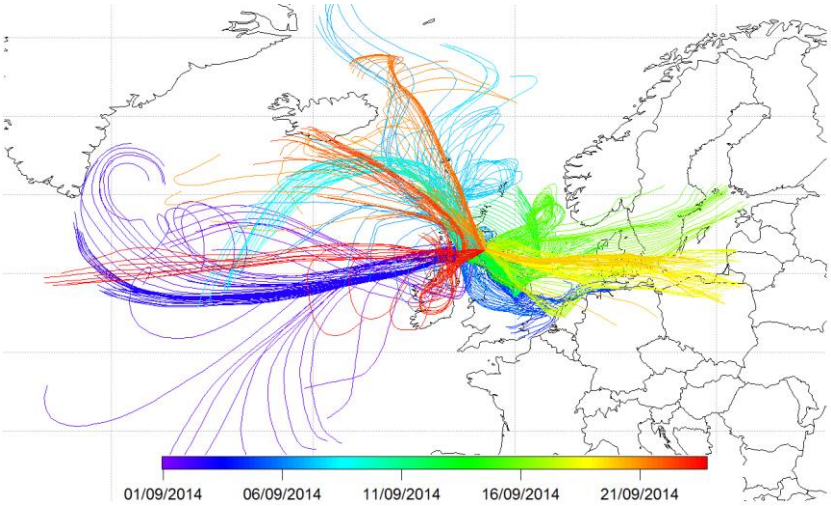
While loading, you can select one or several files at once. Each traj will be stored within the Igor experiment.

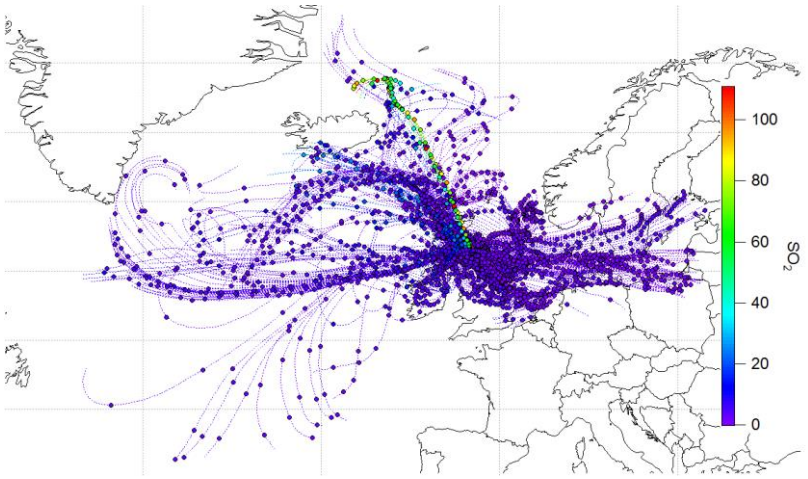


Similarly to ZeFir, a WorldMap needs to be loaded. The same wave (W_Geometries wave, or WorldMap_50m.itx) can be used.

Then, the trajectories can be colored following time, altitude, or an external variable.

Some examples below.


01/09/2014
06/09/2014
11/09/2014
16/09/2014
21/09/2014

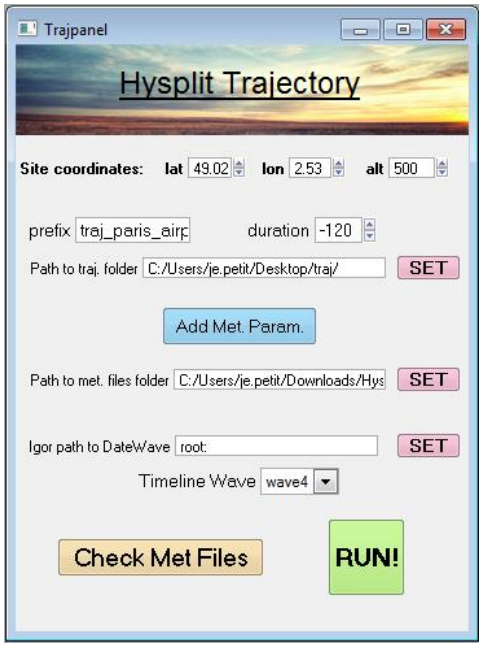

100
80
60
40
20
0
SO2

10.HYSPLIT

This module aims at calculating HYSPLIT backtrajectories from user-friendly panels directly within Igor.

HYSPLIT PC-based version must be installed on the computer, because the module doesn't calculate backtrajectories on its own; it only uses the scripts provided by HYSPLIT software package. It is also limited to the use of regular GDAS meteorological field data files, which must be physically present on the computer.

In-depth control of backtrajectory calculations should be performed from the HYSPLIT software. This module allows basic runs over a long-time period (e.g. several months), whereas HYSPLIT software can perform automated calculation at specific hours over one month only. This module runs the model at any date/time (e.g. an entire year), as long as the corresponding GDAS files are present.

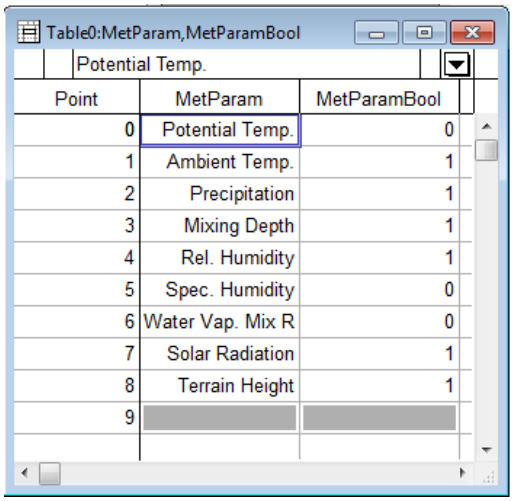


HYSPLIT module panel.

The site coordinates correspond to the arriving location of each backtrajectory.

The path to traj folder is the folder on the computer where the calculated backtrajectories will be stored. Press “Set” in order to select the folder from a pop-up window.

The user has the possibility to add additional parameters within the output files. It can be e.g. precipitation data along trajectory, which can be of great interest when doing trajectory analyses.



Point	MetParam	MetParamBool
0	Potential Temp.	0
1	Ambient Temp.	1
2	Precipitation	1
3	Mixing Depth	1
4	Rel. Humidity	1
5	Spec. Humidity	0
6	Water Vap. Mix R	0
7	Solar Radiation	1
8	Terrain Height	1
9		

Additional parameter table.

The path to met. Files is the folder on the computer which contains the GDAS meteorological data files. Press “Set” in order to select the folder from a pop-up window.

Igor path to date wave is the folder within the Igor experiment containing the date/time information. From the data browser, click on the right folder to highlight its name, and click on “Set”.

Press “Check Met Files” to make sure that the GDAS files present in the folder effectively cover the time period.

Press Run to start calculation.