EcoImpactMapper Tutorial

Andy Stock, E-IPER, Stanford University

Version 1.2, September 30, 2016

Table of Contents

Introduction

The EcoImpactMapper is an open-source software tool for mapping human impacts on marine ecosystems. It is based on an additive model commonly used for this purpose[1]. Since version 1.2 (late September 2016), it also supports alternative models as well as sensitivity and uncertainty analyses[2].

This tutorial will guide you through the EcoImpactMapper's functions from a user's perspective. It uses the example data that can be downloaded with the software's executable. Please see the software paper[3] (open access) and references therein for technical details of the original model and software, and Stock & Micheli (2016, footnote 2) for details on the sensitivity and uncertainty analysis methods. The EcoImpactMapper, its source code and this tutorial are published under MIT license. Please cite the software paper for proper attribution.

New in version 1.2

- A status window showing detailed information about ongoing operations replaces the progress bar.
- Automatic logging.
- Exporting preprocessed data.
- New preprocessing functions: Cut top x%, Percentile (CDF) transformation, replace values with presence-absence.
- Alternative models of multiple stressor effects: Dominant effects, antagonistic effects.
- New uncertainty and sensitivity analysis functions: Monte Carlo simulations, Elementary Effects method.
- Parallel processing for uncertainty and sensitivity analyses.

Starting the EcoImpactMapper

You can start the EcoImpactMapper on most operating systems by double clicking on the file EcoImpactMapper.jar.

---

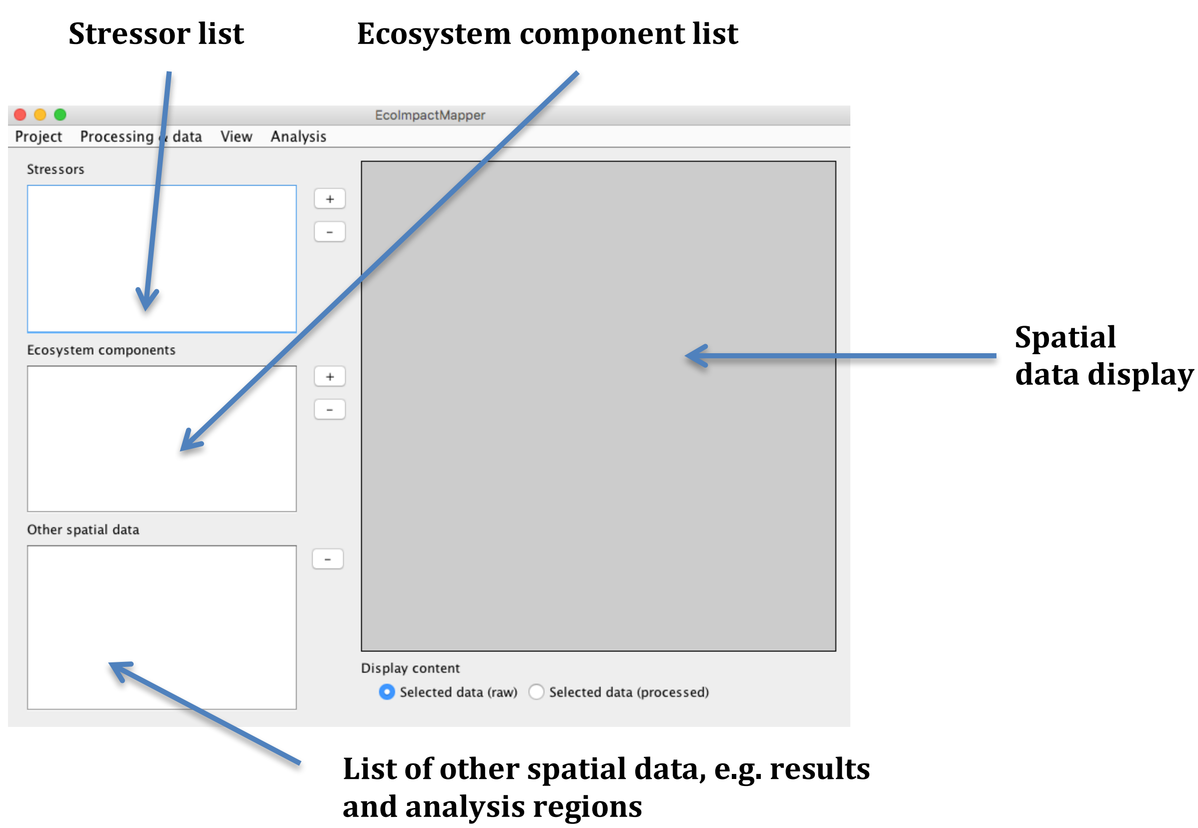
[1] Halpern, B. S., Walbridge, S., Selkoe, K. A., Kappel, C. V., Micheli, F., D'Agrosa, C., ... & Watson, R. (2008). A global map of human impact on marine ecosystems. *Science*, *319*(5865), 948-952.
[2] Stock, A., & Micheli, F. (2016). Effects of model assumptions and data quality on spatial cumulative human impact assessments. *Global Ecology and Biogeography*.
[3] Stock, A. (2016). Open source software for mapping human impacts on marine ecosystems with an additive model. *Journal of Open Research Software*, 4(1), p.e21. DOI: http://doi.org/10.5334/jors.88.

- ***Double-click on EcoImpactMapper.jar***
- ***If you cannot start the EcoImpactMapper, make sure that you have a recent version of the Java Runtime Environment (JRE) installed.***
- ***The JRE has a default cap on the amount of memory it will use, depending on your computer system. This does not matter for the tutorial, but if you plan to work with large data sets, you should increase the JRE's maximum heap size. For this purpose, instead of starting the EcoImpactMapper with a double click, start the Terminal (Mac, Linux) or Command Prompt (Windows), navigate to the folder containing EcoImpactMapper.jar, and type:***

```
java –jar –Xmx12g EcoImpactMapper.jar
```

The "12g" stands for 12 gigabytes, which should be enough for most applications (other than global mapping and very high spatial resolutions). Adjust this depending on how much RAM your computer has and how much memory you expect to need. Adjusting the maximum heap size is not necessary for this tutorial.

The EcoImpactMapper's main window (Fig. 1) should appear within a few seconds:

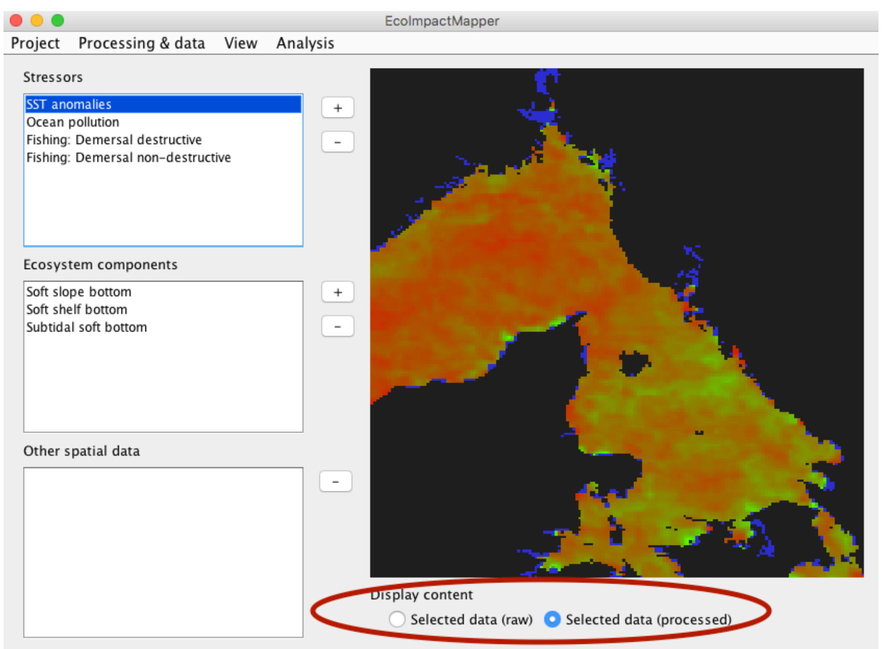


Loading a mapping project

- *Select Project/Load from the menu on top of the main window.*
- *Open the file "exampleproject.csv"*

Exploring the example project

- *You can look at any loaded spatial data layer by clicking on it in the stressors, ecosystem components or "other spatial data" list. The latter contains for example model outputs.*
- *Select View/Color scale in the menu bar on top of the window, then click on one of the colored rectangles to change the top, middle, and bottom color of the color scale.*
- *Under View/Color stretch, you can select if colors from that scale are stretched linearly between the currently displayed layer's minimum and maximum, or assigned according to percentiles instead of the grid cells' actual values.*
- *Per default, stressor data are log(x+1)-transformed and rescaled so that the maximum is 1, as suggested by Halpern et al.'s original Science paper. You can explore the effects of this transformation by switching between the "Selected data (raw)" and "Selected data (processed)" views below the data display. For ecosystem components, there is no transformation by default, and thus the raw and processed views are the same. Also note that log-transformation and rescaling do not change the order of values, and thus, you won't see a difference if you use the percentile (as opposed to the linear) color stretch.*



Note that this is just an example – while based on real data for the Kattegat and Skagerrak in northern Europe, in reality, there are many more stressors and ecosystem components to consider. The example data and project only illustrate how the EcoImpactMapper works. They should not be seen as telling anything real about that region.

Creating a human impact mapping project from scratch

Loading stressor and ecosystem component data

The EcoImpactMapper reads all stressor and ecosystem component data as regular grids from comma-separated value (CSV) files. They are interpreted as tables and must have at least three columns:

- X coordinate of grid cell center
- Y coordinate of grid cell center
- One or more columns representing stressor intensities or ecosystem components. You can have as many stressors and ecosystem components in the same CSV file as you like.

Columns must be separated by comma, and the decimal separator must be a point.

The example data contain a file called example_input_data_spatial.txt. It contains data for four stressors and three ecosystem components in the correct format[4]. Please note that it is possible to split the stressor and ecosystem component data into several CSV files. It would, for example, be possible to have one file for each stressor and one for each ecosystem component. However, the X and Y coordinates must be repeated in each file, and must represent the same regular grid (i.e. it is not possible to have different spatial resolutions, extents or reference systems). All cells must contain a number >=0.

The following table describes the example input file's contents:

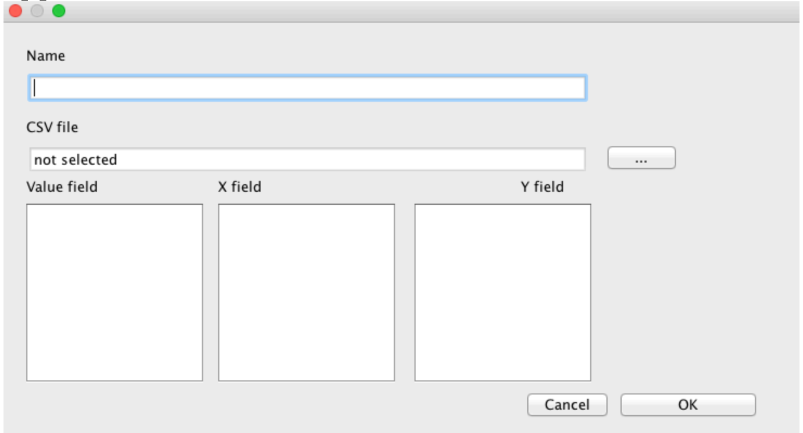
Column	Contents
soft_slope	Ecosystem component: Soft slope bottom (presence-absence)
soft_shelf	Ecosystem component: Soft shelf bottom (presence-absence)
s_t_s_bott	Ecosystem component: Subtidal soft bottom (presence-absence)
sst	Stressor: Sea-surface temperature anomalies
ocean_poll	Stressor: Ocean pollution
dem_non_hi	Stressor: Demersal non-destructive high-bycatch fishing
dem_dest	Stressor: demersal destructive fishing

[4] These data are a subset of global data downloaded from https://www.nceas.ucsb.edu/globalmarine.

Start a new project if you still have loaded the example project.

To load stressors from the example file:

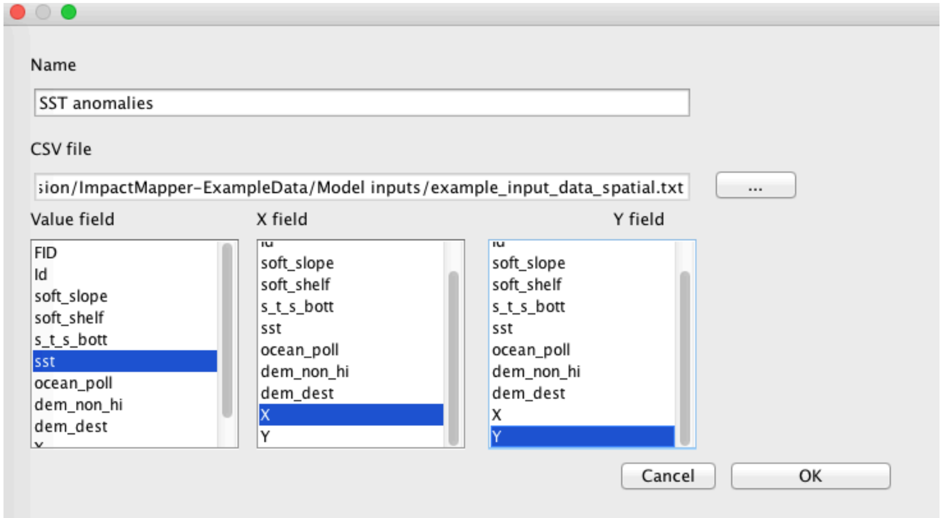
- ***Click the "+" button next to the stressor list. The data loading dialog will appear:***



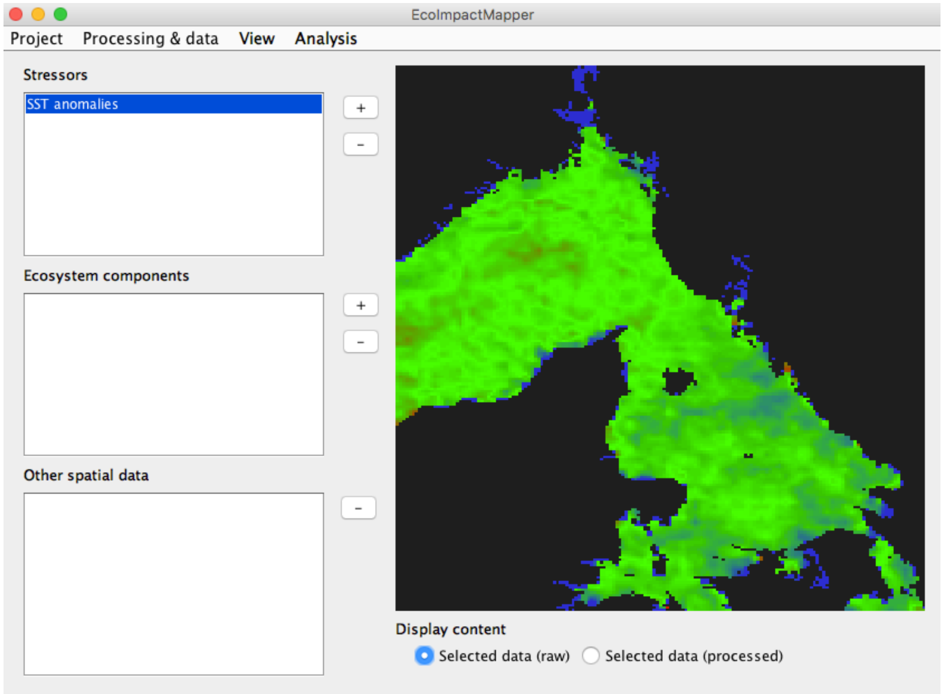
- ***Enter the name "SST anomalies".***

- ***Click on the "..." button next to the "CSV file" text field and locate the file example_input_data_spatial.txt on your hard disk.***

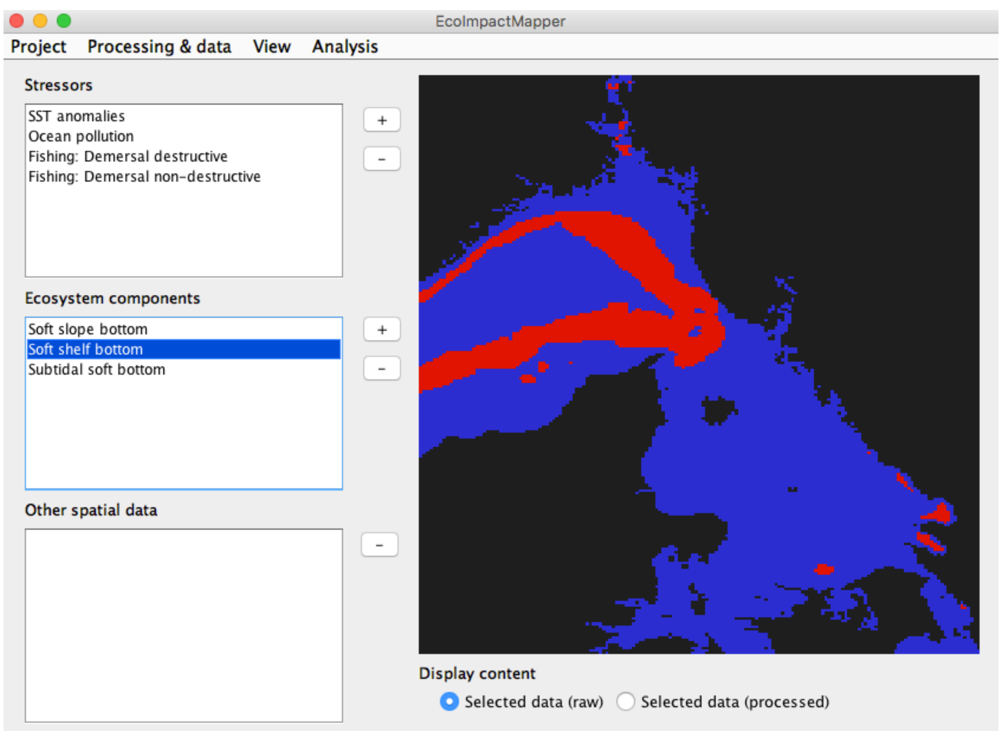
- ***The value field, x field, and y field lists will now show the names of the columns in the selected file. Select "sst" as value field, "X" as X field, and "Y" as Y field:***



- ***Click "OK". Congratulations, you've loaded the first data layer:***



- ***Repeat for the other three stressors. Make sure that the names match the ones in the screenshot below. Then add the three ecosystem components (use the "+" button next to the ecosystem component list for this), again with matching names:***



Loading sensitivity weights

The last kind of data required by Halpern et al.'s model are sensitivity weights, numbers representing the sensitivity of each ecosystem component to each stressor. In the EcoImpactMapper, they are loaded from a CSV file where columns represent ecosystem components and rows represent stressors. **IT IS IMPORTANT THAT THE DATA LAYER NAMES MATCH THE ROW AND COLUMN NAMES IN THE SENSITIVTY WEIGHT TABLE EXACTLY!** For the example data, the table (with made up sensitivity weights) looks as follows:

SensitivtyScores	Soft slope bottom	Soft shelf bottom	Subtidal soft bottom
SST anomalies	1	1	2
Ocean pollution	2	3	3
Fishing: Demersal destructive	4	3	4
Fishing: Demersal non-destructive	3	3	3

- ***Select Processing & data/Load sensitivity weights from the menu on top of the main window. A file selection dialog will appear.***
- ***Select sensitivtyscores.csv from the example model inputs folder.***

Note that stressors and ecosystem components that are missing in the sensitivity weights table will not be included in weighted stressor indices and impact indices. When working with own data, make sure that the sensitivity weight table is complete. An easy way to do this is to first load all data, then copy the names of all stressors and ecosystem components from the project file (see below).

Calculating human impact indices and other results

The EcoImpactMapper can calculate different indices and other results. The following table summarizes the data requirements for each of these. Uncertainty and sensitivity analyses will be described in the next section.

	Stressors	**Ecosystem comp.**	**Sensitivity weights**	**Regions**
Ecol. diversity index		X		
Ecol. sensitivity index		X	X	
Unweighted stressor index	X			
Weighted stressor index	X		X	
Impact index	X	X	X	
Stressor concentration	X	X		
Uncertainty analysis	X	X	X	X
Sensitivity analysis	X	X	X	X

Calculating an ecological diversity index

To calculate a diversity index, ecosystem components must be loaded, but no stressor data or sensitivity weights are needed. The index is simply the sum of all ecosystem component data layers. The index is calculated according to the software paper, Eq. 2.

- ***Select Analysis/Indices/Ecological diversity index from the menu on top of the main window. A "Save file as" dialog will appear.***
- ***Select a location and filename for saving the index and click ok.***

The index will be added to the project and saved as a CSV file with three columns: X, Y (same regular grid as input data) and value. However, because in the example data, exactly one ecosystem component is present in each grid cell, the diversity index is 1 in all locations.

Calculating an ecological sensitivity index

To calculate a sensitivity index, ecosystem components and sensitivity weights must be loaded, but no stressor data are needed. The index is like an ecological diversity index, but each ecosystem component is weighted with the mean of its sensitivity weights for all stressors. A high ecological sensitivity index means that there are many ecosystem components that are sensitive to many stressors.

- ***Select Analysis/Indices/Ecological sensitivity index from the menu on top of the main window. A "Save file as" dialog will appear.***
- ***Select a location and filename for saving the index and click ok.***

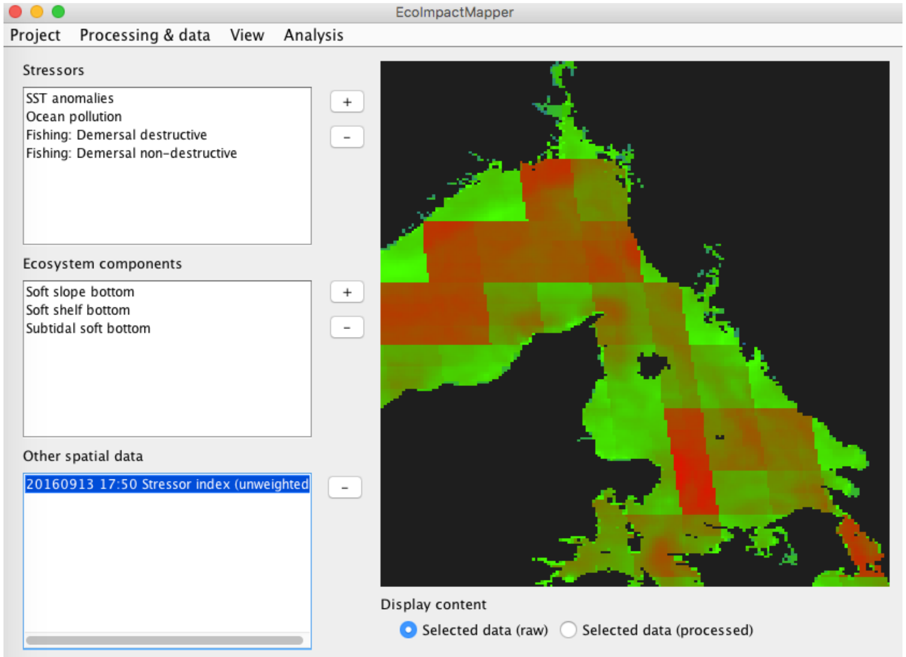
The index will be added to the project and saved as a CSV file with three columns: X, Y (same regular grid as input data) and value.

Calculating an unweighted or weighted stressor index

To calculate an unweighted stressor index, only stressor data need to be loaded. No ecosystem component data or sensitivity weights are needed. The index is the simple sum of all **processed** stressor data layers. See the software paper, Eq. 3 for the equation, and below for information about data pre-processing steps.

- ***Select Analysis/Indices/Stressor index (unweighted) from the menu on top of the main window. A "Save file as" dialog will appear.***
- ***Select a location and filename for saving the index and click ok.***

The index will be added to the project and saved as a CSV file with three columns: X, Y (same regular grid as input data) and value.



Similarly, you can calculate a weighted stressor index from the Analysis/Indices menu. This is also a sum of processed stressor data layers, but each stressor is weighted with the mean of all ecosystem components' sensitivity weights for this stressor. Thus, both stressor data and sensitivity weights must be loaded. See the software paper, Eq. 4.

Calculating a human impact index and stressor/ecosystem component contributions

The EcoImpactMapper can calculate six different human impact indices; it supports three different models for aggregating the effects of multiple stressors on a given ecosystem component, and two different ways to aggregate the impacts on several ecosystem components.

The three models for multiple stressor effects are:

- Additive effect: For each ecosystem component, the impacts from each stressor simply add up, as in Halpern et al.'s original paper (2008, footnote 1 on page 2).
- Dominant effect: For each ecosystem component, only the stressor having the largest impact on that ecosystem component is considered in each grid cell (i.e. the dominant stressor can be a different one in different places).
- Antagonistic effect: For each ecosystem component, the impacts from all stressors add up, but each additional stressor is multiplied with a smaller weight. For example, in a grid cell with three stressors having an impact on a given ecosystem component, the stressor with the highest impact would be

multiplied by 1. The stressor with the second highest impact would be multiplied by 2/3, and the stressor with the third highest impact by 1/3.

These three models of multiple stressor effects are described in detail elsewhere (see footnote 2 on page 2). Once the impacts on each ecosystem component are calculated in one of these three ways, they can be aggregated in two ways:

- As sum of impacts on all ecosystem components in a grid cell, as in the original model (footnote 1 on page 2).
- As mean of impacts on all ecosystem components in a grid cell (as in some later papers).

To calculate an impact index in any of these ways, stressor data, ecosystem component data and sensitivity weights must be loaded. Because there is exactly one ecosystem component per grid cell in the example data, the sum and mean impact indices for these data are identical.

- ***Select the appropriate type of index from Analysis/Indices/Impact index from the menu on top of the main window. For example, for an impact index calculated with an additive model and as sum of impacts on ecosystem components (as in the original model), select "Additive model, sum". A "Save file as" dialog will appear.***
- ***Select a location and filename for saving the index and click ok.***

The index will be added to the project and saved as a CSV file with three columns: X, Y (same regular grid as input data) and value. In addition to the impact index saved under the selected filename, the EcoImpactMapper will create a file ending with "_contributions.csv". For example, if you chose to save the index as "impactindex_sum.csv", there will be an additional file called "impactindex_sum_contributions.csv":

stressor	ecocomp	contribution
SST anomalies	Soft slope bottom	1664.546681
Ocean pollution	Soft slope bottom	2206.872649
Fishing: Demersal destructive	Soft slope bottom	5413.490059
Fishing: Demersal non-destructi	Soft slope bottom	2292.765051
SST anomalies	Soft shelf bottom	1755.742196
Ocean pollution	Soft shelf bottom	3906.390938
Fishing: Demersal destructive	Soft shelf bottom	3691.792788
Fishing: Demersal non-destructi	Soft shelf bottom	2313.047581
SST anomalies	Subtidal soft bottom	14501.59432
Ocean pollution	Subtidal soft bottom	19693.07384
Fishing: Demersal destructive	Subtidal soft bottom	16342.27304
Fishing: Demersal non-destructi	Subtidal soft bottom	10254.28194

The "contribution" column contains the sum over all grid cells of modeled impacts for each stressor-ecosystem component combination. This information can e.g. be used to rank stressors based on their contribution to the impact index, a result that studies using Halpern et al.'s model frequently report.

Analyzing stressor concentration in high-diversity areas

This function is available from the menu on top of the main window under "Analysis/Stressors/Overlay with diversity index". It creates a CSV table with the following columns:

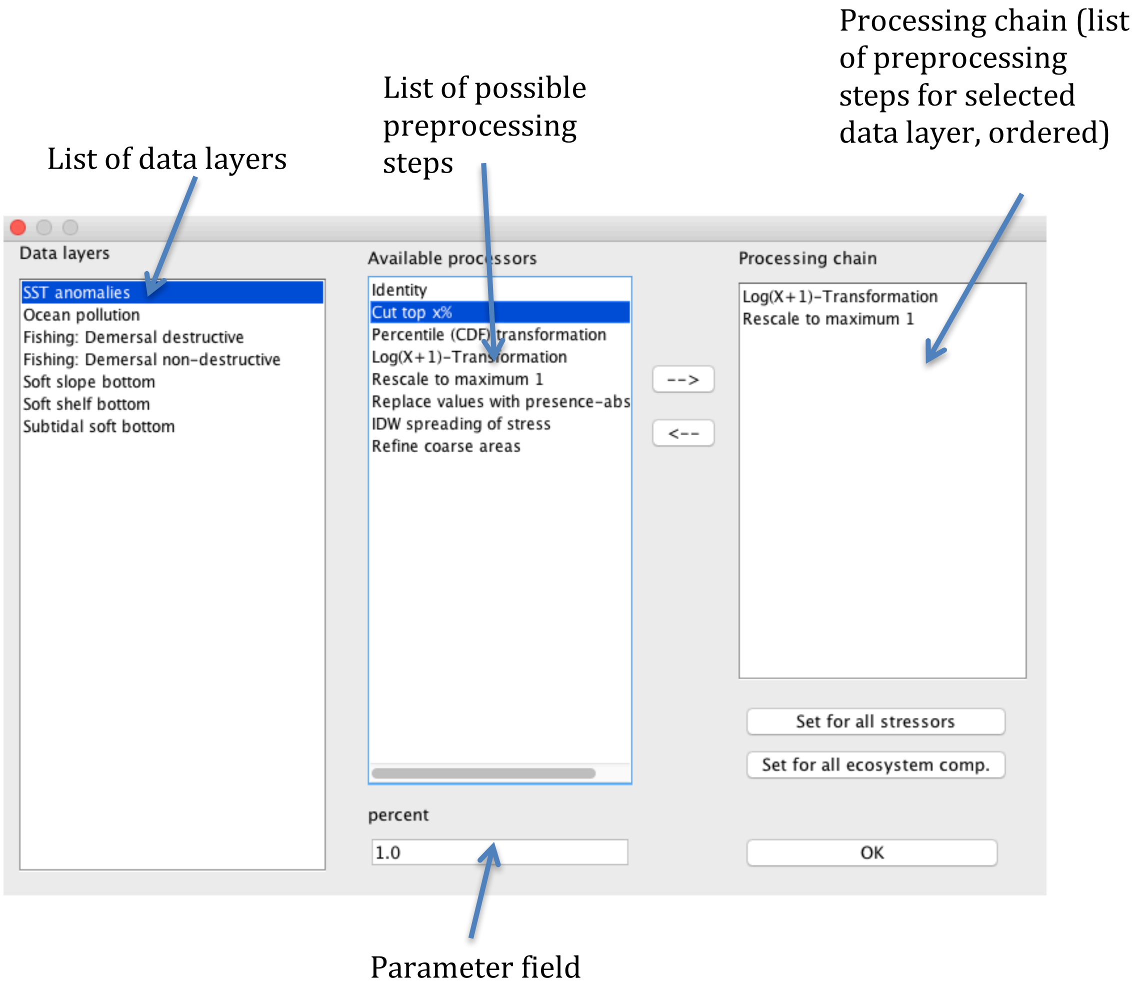
- The 1st column contains the % of the study area with a diversity index equal to or higher than the value in the 2nd column.
- The 2nd column contains all values that the ecological diversity index takes on, in decreasing order.
- The following columns contain the % of **unprocessed** stressor intensities occurring within that part of the study area. For example, if you include a fishing stressor where intensity is measured as effort, the respective column will include the % of effort occurring within the respective part of the study area. This would allow results like "60% of fishing effort occur in the 20% of the study area with the highest diversity index".

Note that for the example data, because the diversity index takes only one value (1), this function is not useful.

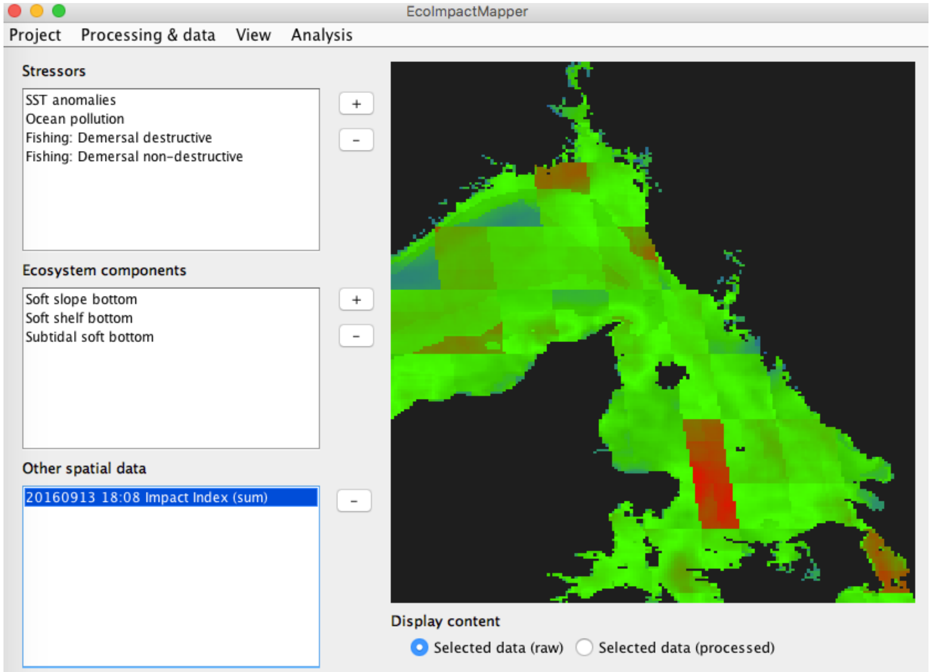
Spatial data preprocessing

In Halpern et al.'s original model, all stressor data layers are log(X+1)-transformed and rescaled so that the maximum of each layer is 1. Per default, the EcoImpactMapper does the same, but it is possible to change the pre-processing. For example, let's calculate an impact index from the example data without Log(X+1)-transformation.

- ***Load the example project.***
- ***Select Processing & data/Preprocessing from the menu on top of the main window.***



- ***Click on an ecosystem component data layer in the data layer list. The processing chain only contains "Identity"; that means that this data layer is not transformed in any way.***
- ***Click on a stressor data layer in the data layer list. The processing chain now contains two entries: "Log(X+1)-transformation" and "Rescale to maximum 1". The entries in the processing chain are ordered; i.e. the layer is <u>first</u> Log(X+1)-transformed and <u>then</u> rescaled.***
- ***Select "Log(X+1)-Transformation" in the processing chain and click the "←" button next to it. This processing step is now removed. Repeat for the other stressor layers or click on "Set for all stressors".***
- ***Click "Ok" and calculate an impact index as described above. This time, no log-transformation of stressor data will take place.***
- ***Some pre-processors have a parameter that can be entered in the parameter field. Otherwise, the parameter field is not active.***



In addition to the log[X+1]-transformation and rescaling, the EcoImpactMapper includes the following alternative transformations and other pre-processing steps:

- *Identity*: No pre-processing
- *Cut top x%*: This preprocessing step sets the highest x% of cell values to the 100-x percentile, where x is given as a parameter. For example, if the parameter is 1.0, then this preprocessing step will set all values greater than the 99-percentile to the 99-percentile. In this way, it removes extremely high values from the data, but in contrast to the log[x+1]-transformation, leaves smaller values untransformed.
- *Percentile (CDF) transformation*: This preprocessing step sets all values in a data layer to the quantile that they correspond to.
- *Replace values with presence-absence*. This preprocessing step replaces all values in a data layer with 1 if they are greater than 0. It can for example be used to map how many stressors there are in each grid cell.

The following two pre-processing functions work well only for some data layers. Their main use is to investigate how they affect your input data before setting ranges or levels for the uncertainty and sensitivity analyses.

- *IDW spreading of stress*: Linear inverse distance weighted decay of stress from e.g. point sources like fish farms or industrial ports. This pre-processing step has a parameter, the distance up to which the stressor spreads from its sources, given in coordinate system units.
- *Refine coarse areas*: This pre-processing step replaces coarse areas (neighboring grid cells with the same value) with a randomly generated

finer-resolution version. This is achieved by generating a random noise grid, locally rescaling it so that the spatial patterns of the original coarse areas are retained, and then applying a low-pass filter. This pre-processing step has a parameter, the distance over which the low-pass filter is applied.

Saving the mapping project, project file format and sharing projects

- ***Save the mapping project by selecting Project/Save from the menu on top of the main window, then providing a filename for saving, e.g. "myproject.csv".***

- ***Like all inputs and outputs of the EcoImpactMapper, the project file is a CSV file. Open it, e.g. in Excel or a text editor, and take a look at its contents.***

Note: The EcoImpactMapper does automatically save all outputs, but not the project as such. There is no warning when you close the EcoImpactMapper without saving changes to the project first. Be sure to save the project often, and to make backups of your project before major changes or manually editing files.

The project file contains the following information:

- a list of stressors with information about the source files
- a list of ecosystem components with information about the source files
- a list of results (e.g. impact indices) with information about the source files
- the path to the file containing sensitivity weights (if loaded)
- the path to the file containing processing chains
- the paths to three files containing settings for uncertainty and sensitivity analysis (see next section)
- the path to the file containing the regions (if loaded)
- information about the color scale for data visualization

You can make changes to the project, e.g. add data, directly by editing the project file; <u>but be sure to make a backup copy in case the file becomes invalid due to manual editing.</u>

The processing chains for all data layers are stored in the same folder as your project file, in a file ending with "pchains.csv". For example, if you saved your project as "myproject.csv", the processing chains will be stored in "myproject_pchains.csv". The processing chain file lists the preprocessing steps for each data layer in the order in which they will be applied. The settings related to uncertainty and sensitivity analysis are stored in the files starting with the project name and ending with "_SF.csv" (selective factors), "_mcsettings.csv" (settings for Monte Carlo simulations) and "_factors.csv" (factor levels for elementary effects analysis).

- ***To share your project, share your project file and all files to which it refers.***
- ***The project file refers to the input- and output data files using relative paths. The data are not stored inside the project file. They must thus be shared, too.***
- ***It is thus strongly recommended that you establish a project folder for each new project, and keep all input data, outputs, the project file and related files in that folder or its subfolders. In this way, you can simply share your whole project folder without worrying about compiling the required files.***

Uncertainty and sensitivity analysis

Since version 1.2 (September 2016), the EcoImpactMapper contains functions for uncertainty analysis (UA) and sensitivity analysis (SA):

- Monte Carlo simulations for UA.
- The elementary effects method for SA.

Both methods are global and thus can provide a reliable picture of the uncertainty related to model assumptions and errors in the input data. They work by making hundreds or even thousands of human impact maps with different but ecologically plausible model assumptions and random errors in the input data. In the following, I will call these assumptions and errors <u>factors.</u>

The EcoImpactMapper currently supports nine factors:

- Missing stressor data
- Errors in sensitivity weights
- *The distance over which stress, e.g. from point sources, decays
- The shape of ecosystem components' responses to different types of stress (linear vs. threshold responses)
- Spatial resolution of the analysis
- *Refined resolution for stressors represented by large areas.
- Aggregation of impacts on ecosystem components as sum or mean
- Type of stressor data transformation (none, log(X+1), cut highest 1%, CDF)
- Multiple stressor effects model (additive, dominant stressor, antagonistic)

The two factors marked with * are only applied to stressor data layers that you assign them to. In the EcoImpactMapper, such factors are called "selective", because they are only applied to layers selected by the user. For example, it may be reasonable to include spatial decay of stress from fish farms represented by presence-absence, but not from a satellite-derived map of sea surface temperatures (commonly used as a proxy for ocean warming).

These methods and factors are described and justified in detail by Stock & Micheli (2016, see footnote 2, page 2). Great in-depth yet readable explanations of the elementary effects method are found in the books by Saltelli et al.[5]

The following sections explain the factors and methods using the example data.

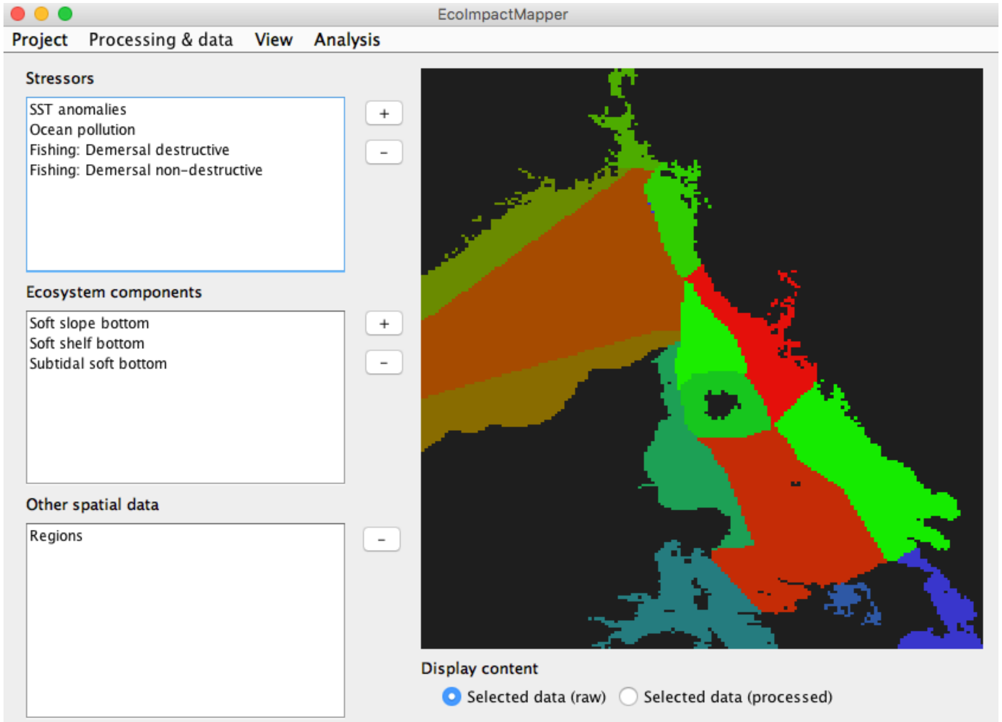
Selective factors and regions

There are two additional kinds of information required for uncertainty and sensitivity analyses with the EcoImpactMapper:

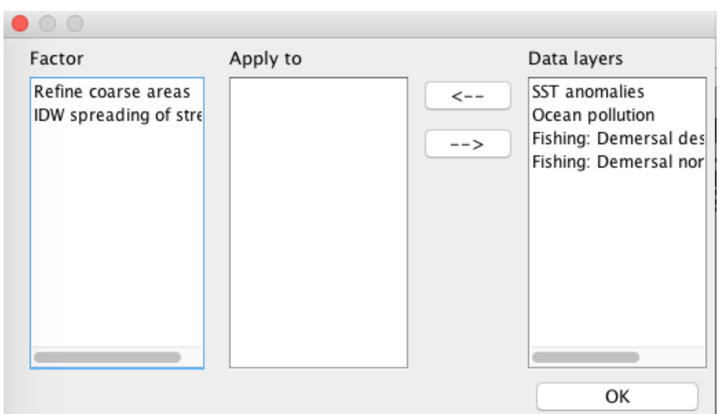
- A sub-division of the study area into regions (the mean value of the impact index in each region will be used to rank them as a proxy for broad-scale spatial patterns)
- Which stressor data layers, if any, the two selective factors should be applied to.

You can load regions from the "Processing and data/Load regions" menu. Like any other spatial data layer in the EcoImpactMapper, they are loaded from a CSV file with X, Y and value fields. The value should be an integer "region code", i.e. all grid cells with the same value will be considered a region. This tutorial includes an example region file called "example_regions.csv". Load the example mapping project and add the regions before proceeding to the next sections. Like for stressors and ecosystem components, you will have to select an x field, a y field, and a value field (in this case, "region"). You cannot change this data layer's name.

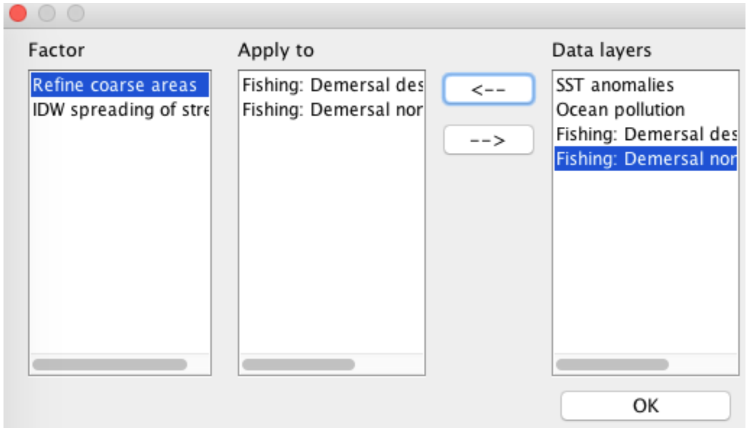
[5] *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models* (2004) and *Global Sensitivity Analysis: The Primer* (2008).



After loading the regions, assign the two selective factors to stressor data layers via "Analysis/Uncertainty/Assign selective factors":

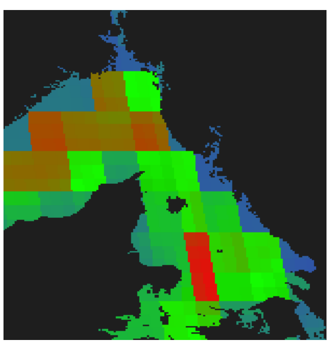


You can select one of the factors on the left, one of the data layers on the right, and use the "<--" button to apply the factor to the data layer. Assuming spreading of stress from its source (i.e. linear decay) only makes sense for stressor data layers that do not represent the spatial extent of stressors, like presence-absence of fish farms. This factor should thus not be assigned to any data layer in the example project. The two fishing data layers, however, represent fishing effort as coarse areas. To include the effects of refined resolution of these data, assign the "refine coarse areas" factor here:

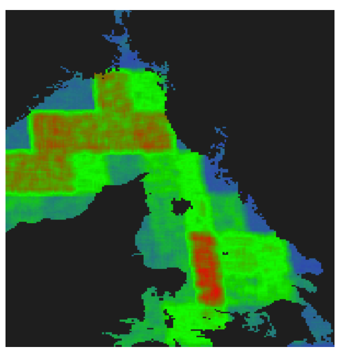


Click "OK" to close the selective factors window and save the project.

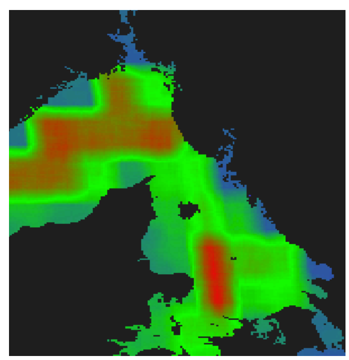
Resolution refinement is based on a random noise surface that is locally rescaled to mimic the spatial patterns of the original data, and then low-pass filtered. For example, if turned on, it replaces the original demersal destructive fishing data set with randomly generated ones like this:



Original (log-transformed)



Refined (10km)



Refined (20km)

The numbers in brackets are the low-pass filter distance, which can be set in the sensitivity analysis or uncertainty analysis window (see below). You can investigate the effects of different low-pass filter distances by applying the "Refine coarse areas" preprocessor to your data layers (see the "Spatial Data Prpeprocessing" section):.

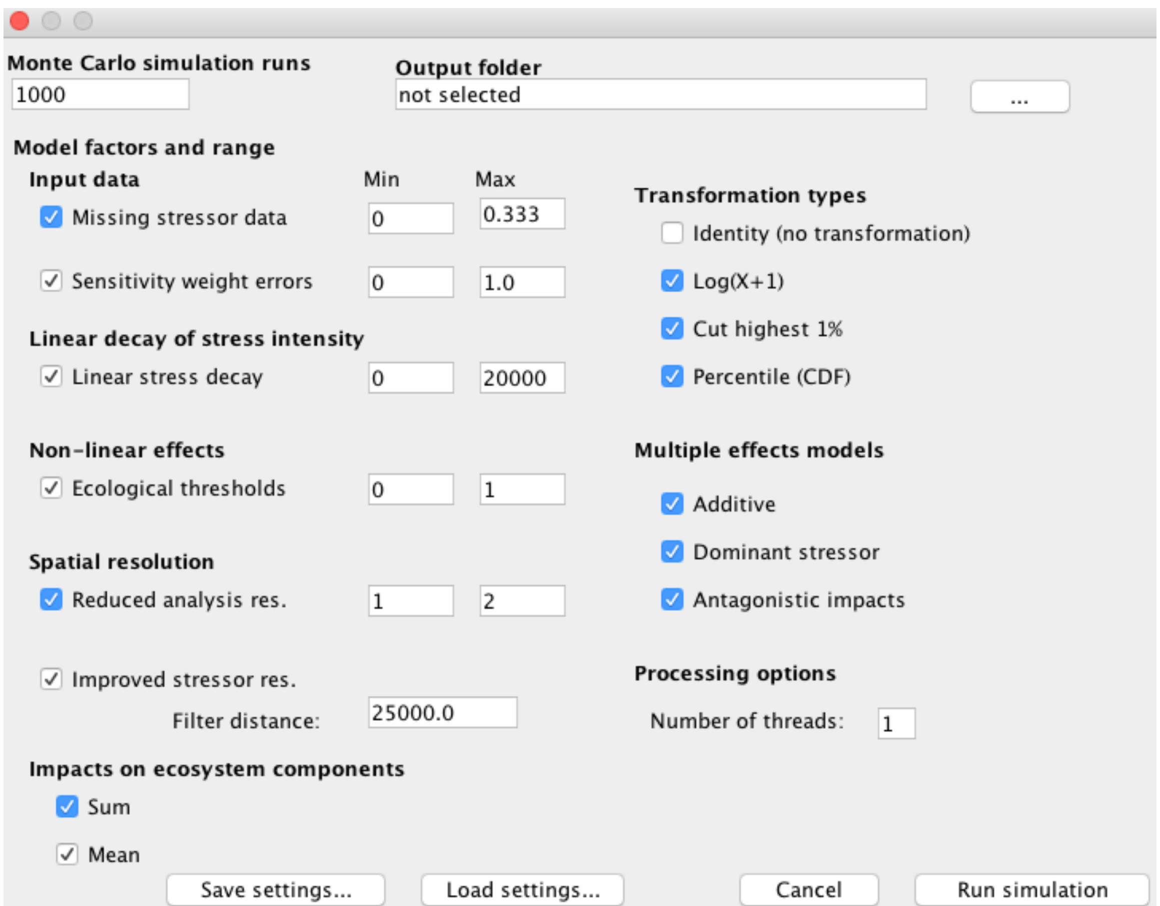
Uncertainty analysis with Monte Carlo simulations

This function allows you to distinguish robust patterns in your human impact map from patterns that are artifacts of specific modeling assumptions, or sensitive to data quality problems. This is achieved by making hundreds or thousands of maps with different ecologically plausible assumptions and random errors in the input data, and counting how often different results are obtained.

The EcoImpactMapper currently supports nine model assumptions and data quality problems (factors). They are described in detail by Stock & Micheli (2016, see footnote 2 on page 2). A major shortcoming of this approach is that there are other assumptions and types of errors that are not considered.

Note that the preprocessing steps are determined by the uncertainty analysis settings described below, not by the steps selected in the Pre-processing window.

- ***Load the example project.***
- ***Open the uncertainty analysis window via "Analysis/Uncertainty/Monte Carlo with random sampling":***



The following list describes all options. You can also read a description by hovering your mouse cursor over most labels and fields in the uncertainty analysis window.

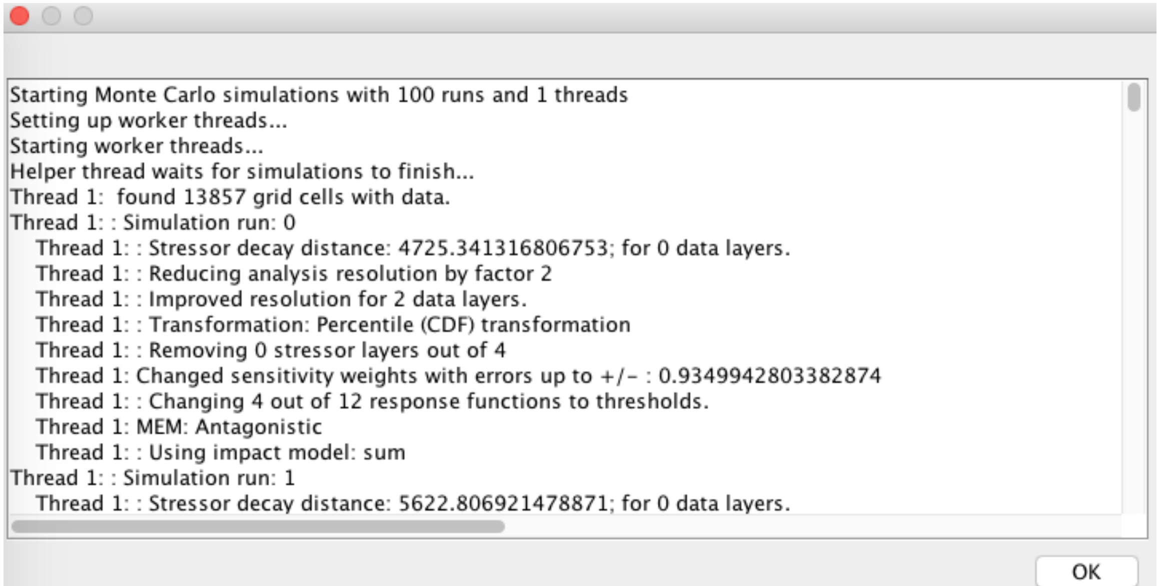
- ***Monte Carlo simulation runs***: The number of simulation runs, i.e. the number of maps with randomly assembled assumptions and random errors in the input data that is generated.
- ***Output folder:*** Where the output files will be stored.
- ***Missing stressor data:*** Proportion of input stressor data sets that will be excluded in each simulation run. Takes values from 0 to 1. For example, the default minimum of 0 and maximum of 0.333 mean that in each simulation run, up to 1/3 of stressor data sets (but possibly fewer) will be excluded. How many and which is randomly decided in each simulation run. As another example, if you were to set both minimum and maximum to 0.5, half of the stressor data sets would be excluded in each simulation run. Each stressor has the same probability of being excluded.
- ***Sensitivity weight errors***. Magnitude of random errors added to sensitivity weights, ranging from 0 to 1. If *M* is the range of the sensitivity weights, the errors are drawn from a uniform distribution from *–r*M* to *r*M*, where *r* is between *min* and *max*. For example, if the largest sensitivity weight is 4 and the smallest sensitivity weight is 1 (i.e. the range is 3) like in the example project, and *min=0* and *max=1*, each simulation run will have sensitivity weight errors between *-3*r* and *3*r*. In each run, *r* will be a new random number between 0 and 1. Similarly, if both *min* and *max* were *0.5*, then each simulation run would have sensitivity weight errors drawn from a uniform distribution *U(-1.5,1.5)*. If an error is so large that adding it to a sensitivity weight would make it fall outside the original weights' range, the minimum or respectively, maximum of the range will be used.
- ***Linear stress decay.*** This is a selective factor that can e.g. be assigned to stressor data layers representing point sources of stress. Stress will be assumed to spread and decay linearly up to a randomly chosen distance between *min* and *max* in each simulation run. *Min* and *max* must be given in coordinate system units. The decay distance is the same for all stressors to which this factor is assigned. The example project's coordinate system uses meters as basic units. Thus, *min=0* and *max=20000* means that in each simulation run, the decay distance will be between 0 and 20km. *Note that you can experiment with different decay distances and see how they work out for your data by using the "IDW spreading of stress" preprocessor.*
- ***Ecological thresholds.*** Proportion of stressor-ecosystem combinations that are assumed to have threshold instead of linear responses to increasing stress. In each simulation run, a proportion between *min* and *max* will be chosen, and a respective number of randomly selected stressor-ecosystem component combinations will be assigned threshold responses. All others will keep assuming linear responses. For example, with *min=0* and *max=1*, some simulation runs will have almost all linear responses, some will have almost all threshold responses, and most will have a mix of both in varying

proportions. If *min=max=0.5*, half of the stressor-ecosystem component combinations will have linear, and the other half will have threshold responses. Which responses are thresholds will be chosen randomly in each simulation run.

- ***Reduced analysis resolution.*** This factor reduces the analysis resolution by a given factor. The example project's spatial resolution is 2km. A value of 1 means using the original resolution, 2 means that the spatial resolution is reduced by factor 2 (i.e. to 4km cells in the example data), and so on. The factor is chosen randomly between *min* and *max* in each simulation run. For example, with *min=1* and *max=2*, about half the simulation runs will use the original resolution, whereas the rest will convert all spatial input data to half the original resolution.
- ***Improved stressor resolution.*** This is a selective factor. If enabled, in each simulation run, all stressor data sets to which it is assigned will be replaced with randomly generated finer-resolution versions that retain the original data's broad-scale patterns. The larger the filter distance, the smoother (less noisy) these finer-resolution data sets are. The filter distance is given in coordinate system units. *Note that you can experiment with different low-pass filter distances and see how they work for your data by using the "Refine coarse areas" preprocessor.*
- ***Impacts on ecosystem components.*** This factor determines if impacts are aggregated as sum or mean for all ecosystem components in a given grid cell. If you enable only one of these options, it will be used in all simulation runs; if both are enabled, a coin is flipped in each simulation run.
- ***Transformation types.*** This factor determines which transformation type is used (log[x+1], cut the highest 1%, CDF-transformation, or no transformation at all). At least one type must be enabled. If several types are enabled, one will be randomly chosen and applied to all stressors in each simulation run.
- ***Multiple stressor effects models.*** At least one must be enabled. In each simulation run, one of the enabled multiple stressor effects models (additive, dominant stressor or antagonistic) is randomly selected.
- ***Number of threads.*** This option allows parallel processing, which speeds up calculations on most modern computers but increases memory requirements. See the section on parallel processing below.

You can save and load your settings with the respective buttons.

- ***For now, only select an output folder and keep the default settings for everything else. Click on "Run simulation".***
- ***The status window provides information about the factor values in each simulation run:***



In the first simulation run, the decay distance was about 4525m but applied to no data layers (because we did not assign any data layers to this selective factor); the analysis resolution was reduced by factor 2; the spatial resolution was refined for the two data layers to which this factor was assigned; the CDF-transformation was used; no stressor data layer was excluded; errors up to +/- 0.93 were added to the sensitivity weights; 4 response functions were set to thresholds; and the antagonistic multiple stressor effects model was used. Note that your results will differ, as the factors are set randomly based on your options. This information is also stored in your session log (see "Logging" section below).

Once the calculations are completed, the "OK" button will be enabled. Your selected output folder will contain the following files.

simulationsettings.csv

This file stores the Monte Carlo simulation settings for your record. You can also load it in the uncertainty analysis window ("Load settings" button) if you want to repeat an earlier analysis.

ecocompranks.csv:

Ecocomp	Cellsum	HighestRank	LowestRank	RankRange	InTop25p	InBottom25p
Soft slope bo	1948	3	1	2	63	19
Soft shelf bo	2109	3	1	2	17	41
Subtidal soft	9800	3	1	2	20	40

Contains the number of cells in which each ecosystem component is present, its highest and lowest rank and range of ranks during all simulation runs, and the % of simulation runs in which it was among the most and least impacted 25% of ecosystem components (because there are only 3 ecosystem components in the example project, this corresponds to being the most or the least impacted). Each ecosystem component was the most and the least impacted in at least one simulation run. Soft slope bottoms were the most impacted ecosystem component in 63% of simulation runs, and percentages are lower for all other results. Thus, in the example project, which ecosystem components are the most and the least impacted depends much on model assumptions and data quality.

stressorranks.csv:

Stressor	Times includ	HighestRank	LowestRank	RankRange	InTop25p	InBottom25p
Ocean pollut	85	1	0.33333334	0.6666666	52.9411765	1.17647059
Fishing: Dem	85	1	0.33333334	0.6666666	54.1176471	3.52941176
SST anomalie	87	1	0.25	0.75	4.59770115	45.9770115
Fishing: Dem	86	1	0.25	0.75	5.81395349	65.1162791

This file shows how often each stressor was included in the simulations, its highest and lowest rank (note that ranks are normalized for stressors as the number of included stressors can change between simulation runs), their difference (range) as well as the proportion of simulation runs in which it was among the most and least important 25% of stressors (counting only runs were the stressor was included). Because the example project contains only 4 stressors, this means being the single most or least important stressor here. One robust result is that SST anomalies and demersal non-destructive fishing were the most important stressor only under quite rare circumstances, whereas ocean pollution and demersal destructive fishing were rarely the least important stressor. Which stressor was more important within these pairs, however, depended much on model assumptions and data quality.

regionranks.csv:

Region	HighestRank	LowestRank	RankRange	InTop25p	InBottom25p
12	14	1	13	72	10
1	13	1	12	80	4
7	14	1	13	36	13
9	14	2	12	20	62
10	14	2	12	17	23
13	14	1	13	70	4
3	12	1	11	33	5
8	12	1	11	47	3
11	13	2	11	5	13
14	12	5	7	0	29
2	14	4	10	1	88
6	12	3	9	11	19
4	14	3	11	1	78
5	14	2	12	7	49

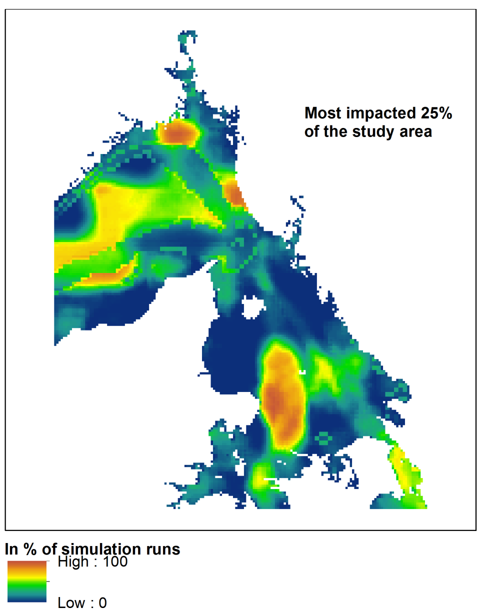
This file contains each region's highest rank, lowest rank, the range of ranks, and the proportion of simulation runs in which it was among the most and least impacted 25% of regions. As there are 14 regions in the example project, this means being among the most or least impacted 4 regions here. Region 1 was among the 4 most impacted regions in 80% of simulation runs. Region 2 was among the least impacted regions in 88% of simulation runs, and several regions were in one of these categories in 70% or more of simulation runs. Most regions, however, were not robustly in either of these categories.

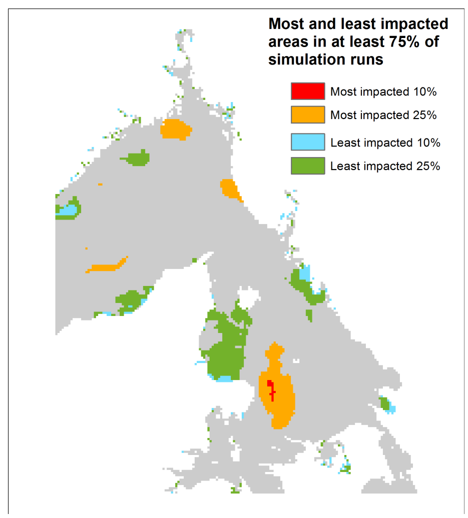
highest25p.csv:

x	y	value
685775.75	6772416	95
685775.75	6774416	95
773775.75	6520416	94
773775.75	6522416	94
765775.75	6526416	94
775775.75	6528416	94
773775.75	6530416	94
777775.75	6530416	94

...

This file contains how often each grid cell was among the 25% with the highest impact score. Like all spatial outputs, it has an x field, a y field, and a value field. The remaining output files contain the same information for the lowest 25 percent of simulation runs, and for other percentages (e.g. highest05p.csv and lowest05p.csv, the proportion of simulation runs in which each grid cell was among the most and least impacted 5%). Loaded into GIS software, the example results look as follows:



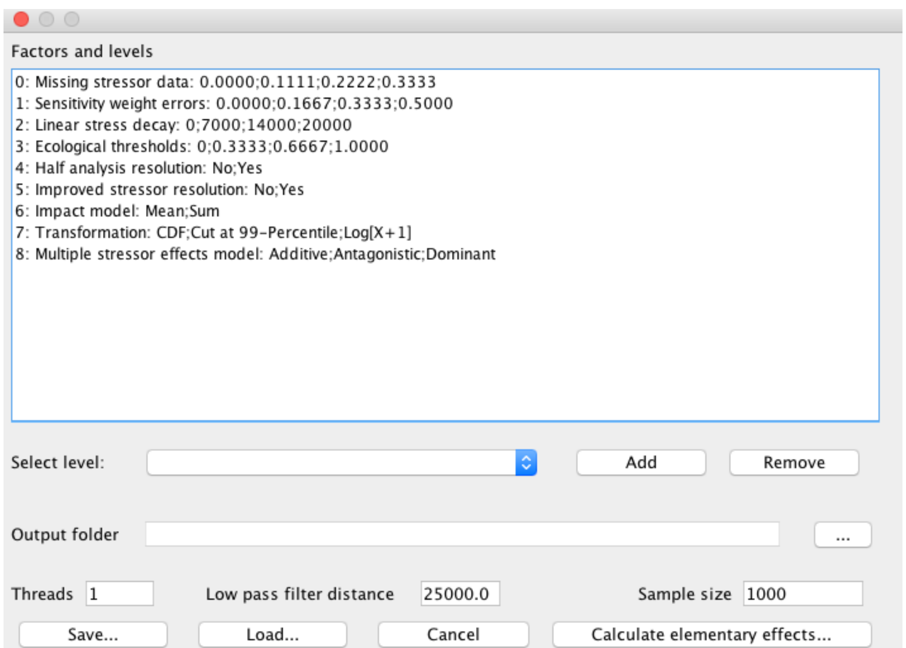


As these maps show, there were some relatively small areas that were robustly among the most and least impacted parts of the study area. Note again that this is just an example – while these maps are based on real data for the Kattegat and Skagerrak, in reality, there are many more stressors and ecosystem components to consider. The example project also uses a subset of global data that may not get local patterns right. The examples should thus not be seen as telling anything real about that region.

Sensitivity analysis with the elementary effects method

The elementary effects method[6] allows ranking of factors by their influence on model outputs. In the EcoImpactMapper, it has been adjusted to work with spatial human impact models (see Stock & Micheli 2016, footnote 2 on page 2). The factors are the same as for the Monte Carlo simulations (with some limitations; for example, the analysis resolution can only be reduced by factor 2, not any factor). Furthermore, quantitative factors cannot take on any value from their range, but a usually small number of levels. There are great books explaining the method and setting factor levels in detail (Saltelli et al. 2004, 2008; see footnote 5, page 17).

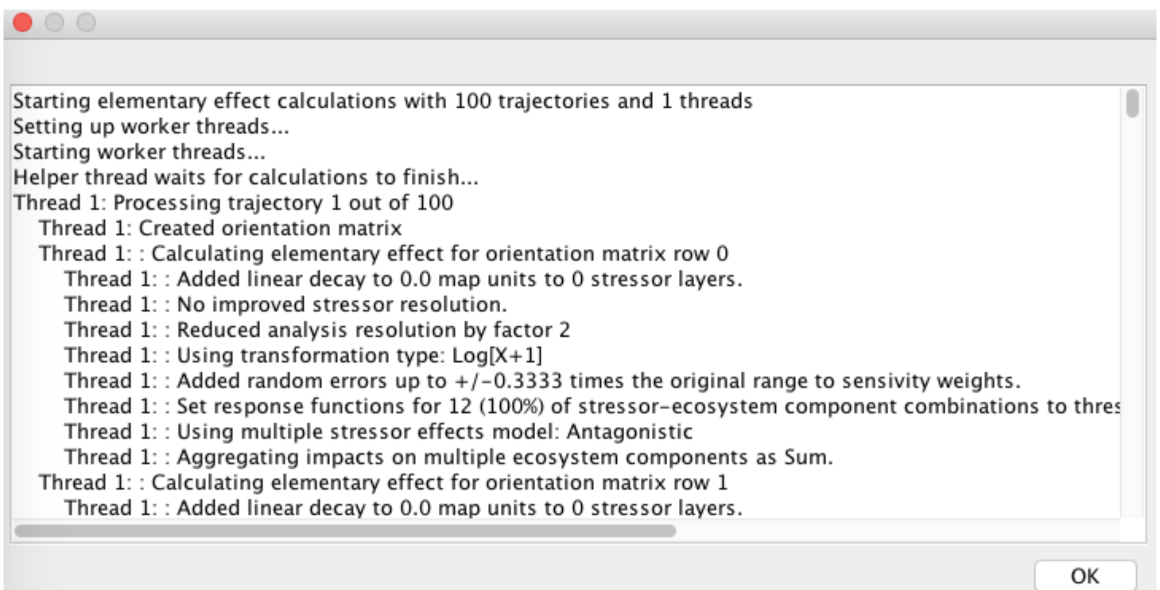
- ***Load the example project and regions.***
- ***Assign the selective factors as above, i.e. IDW spreading of stress should not be assigned to any data layers, and refining coarse areas should be assigned to the two fishing layers.***
- ***Open the sensitivity analysis window from "Analysis/Uncertainty/Elementary effects with Morris design":***



---

[6] Morris, M. D. (1991). Factorial sampling plans for preliminary computational experiments. *Technometrics*, *33*(2), 161-174.

- ***The list in the top half of the window shows the factors and their levels. In this tutorial, we will use the default levels. You could however select a factor in the list, then use the "Select level" field with the "Add" and "Remove" buttons to select your own levels.***
- ***You can disable a factor by using exactly one level.***
- ***Choose an output folder, let the number of threads remain 1, and the low pass filter distance (used for refining the resolution of coarse areas as described above) remain 25,000.***
- ***The elementary effects method evaluates the model 10 times to calculate a single elementary effect for each of the nine factors. While for real projects, you may want to use a larger sample size, reduce it to 100 here. Also note that the elementary effects method tends to reliably rank factors by their importance even at very small sample sizes.***
- ***You can save or load your settings with the respective buttons on the bottom of the window.***
- ***Click on "Calculate elementary effects…". The status window will show the calculations' progress. This information will also be stored in a log file (see "Logging" section below). Once finished, the "OK" button will become enabled.***



The output folder now contains the results of the elementary effects calculation.

levels.csv
Lists the factor levels used. This file can also be loaded in the elementary effects window if you need to repeat calculations.

ecocomps_mustar.csv, ecocomps_sigmastar.csv, stressors_mustar.csv, stressors_sigmastar.csv, regions_mustar.csv, regions_sigmastar.csv

The adjusted elementary effects method as implemented in the EcoImpactMapper produces 2 measures: μ*, a measure of a factor's overall effect on a model output, and σ*, a measure of how much this factor's effect depends on interactions with other factors and random components in the model (e.g. *which* response functions are assumed to be linear and *which* are assumed to be thresholds). These measures are calculated for three types of model outputs – the ranks of stressors, the ranks of ecosystem components, and the ranks of regions impacted (as a proxy of broad-scale spatial patterns) from most to least impacted.

For example, regions_mustar.csv looks as follows:

Region	0: Missing st	1: Sensitivity	2: Linear stre	3: Ecological	4: Half analy	5: Improved	6: Impact mc	7: Transform	8: Multiple stressor effects model
1	1.951	1.3310382	0	2.4309447	0.491	0.436	0.382	1.561	1.232
2	1.129	0.8560352	0	1.6509691	0.556	0.253	0.279	1.035	1.404
3	1.798	1.4875469	0	3.0429504	0.914	0.637	0.823	1.493	1.228
4	1.393	0.9345284	0	2.319951	0.461	0.429	0.321	2.04	1.007
5	2.0765	1.2090349	0	3.7404506	0.7	0.711	0.57	2.454	1.59
6	1.537	1.3515128	0	2.796476	0.588	0.676	0.436	2.429	1.329
7	1.4515	2.0200481	0	4.1749096	1.037	0.584	0.929	2.699	2.001
8	1.5485	1.3605105	0	3.047457	0.526	0.847	0.382	1.741	1.403
9	2.4835	1.1820081	0	3.129961	1.519	0.377	1.016	1.793	3.614
10	1.639	1.9685452	0	3.8734317	1.194	0.973	1.04	2.162	1.825
11	1.9295	1.3880428	0	2.37796	0.662	0.635	0.464	1.961	1.233
12	3.1465	1.928084	0	3.1674547	0.431	0.535	0.379	2.047	1.224
13	2.1915	1.4275389	0	2.8529441	0.684	0.651	0.575	1.684	1.235
14	1.2715	1.0430338	0	2.4104514	0.535	0.65	0.364	1.363	0.975

Rows are regions and columns are factors. The number in each cell is μ* for that region and factor, i.e. a measure of how much the factor influences the region's rank. In order to get a measure of how much each factor influences region ranks overall, let's calculate the mean of each factor's μ* over all regions:

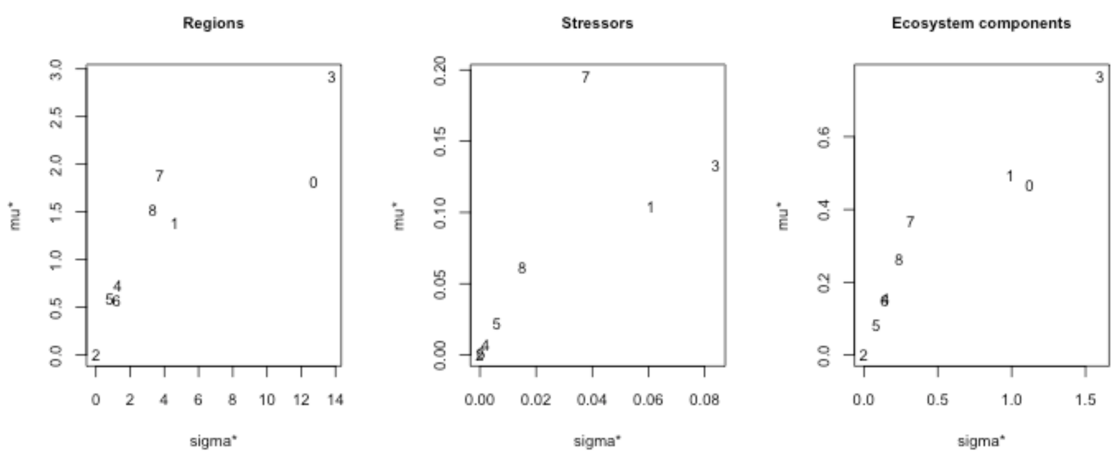
Region	0: Missing st	1: Sensitivity	2: Linear stre	3: Ecological	4: Half analy	5: Improved	6: Impact mc	7: Transform	8: Multiple stressor effects model
1	1.951	1.3310382	0	2.4309447	0.491	0.436	0.382	1.561	1.232
2	1.129	0.8560352	0	1.6509691	0.556	0.253	0.279	1.035	1.404
3	1.798	1.4875469	0	3.0429504	0.914	0.637	0.823	1.493	1.228
4	1.393	0.9345284	0	2.319951	0.461	0.429	0.321	2.04	1.007
5	2.0765	1.2090349	0	3.7404506	0.7	0.711	0.57	2.454	1.59
6	1.537	1.3515128	0	2.796476	0.588	0.676	0.436	2.429	1.329
7	1.4515	2.0200481	0	4.1749096	1.037	0.584	0.929	2.699	2.001
8	1.5485	1.3605105	0	3.047457	0.526	0.847	0.382	1.741	1.403
9	2.4835	1.1820081	0	3.129961	1.519	0.377	1.016	1.793	3.614
10	1.639	1.9685452	0	3.8734317	1.194	0.973	1.04	2.162	1.825
11	1.9295	1.3880428	0	2.37796	0.662	0.635	0.464	1.961	1.233
12	3.1465	1.928084	0	3.1674547	0.431	0.535	0.379	2.047	1.224
13	2.1915	1.4275389	0	2.8529441	0.684	0.651	0.575	1.684	1.235
14	1.2715	1.0430338	0	2.4104514	0.535	0.65	0.364	1.363	0.975
Mean	**1.82471429**	**1.39196484**	**0**	**2.92973652**	**0.73557143**	**0.59957143**	**0.56857143**	**1.89014286**	**1.52142857**

We see that factor 3 – assuming ecological thresholds – has the largest μ*, i.e. the largest effect on region ranks in the example project. Also note that factor 2 – Linear stress decay – has no effect at all, because we did not apply this selective factor to any stressors data layers.

There is only one ecosystem component in each cell. You may wonder why factor 6 – calculating overall impacts in a cell as sum or as mean of impacts on the present ecosystem components – seems to have a small effect. This happens because when the spatial resolution of the input data is reduced automatically in some simulation runs, adjacent ecosystem components may overlap. Thus, the resolution reduction results in more than one ecosystem component per cell in some places. Indeed, if we had disabled the resolution reduction, the μ*'s for factor 6 would all have been 0. This somewhat artificial interaction is in practice no big problem because most real-world impact mapping projects have several ecosystem components in each cell anyway, and cover large areas compared to the cell size, so that boundaries between adjacent ecosystem components make up only a small fraction of the area of each region.

Similarly, the σ* for individual regions, stressors and ecosystem components can be summarized into an overall measure by calculating the mean.

It is common to plot μ* and σ*together as follows:



The numbers correspond to the factor numbers in the software. The most influential factor for the ranks of regions and ecosystem components is factor 3: ecological thresholds. In contrast, the ranking of stressors is most influenced by factor 7, the choice of transformation. Note that σ*should be interpreted relative to μ*, and is a variance. To be comparable to μ*, the square root must thus be taken.

Parallel processing

Uncertainty and sensitivity analysis can require hundreds or thousands of model evaluations. One way to speed up calculations is parallel processing. Most computer processors have several cores, which means that they can physically perform several calculations at the same time. The number of cores ranges from 2 in many standard laptops to more than 20 in some desktop workstations. These computers could physically perform 2 or respectively 20, operations at the same time.

If you use k threads for uncertainty and sensitivity analysis, the EcoImpactMapper will split the problem into k chunks and process each of them in parallel. If your computer has k or more cores, the processing will be respectively faster. For example, with 2 threads, the calculations could be 2 times faster – except that in practice, there is some overhead because the threads need to be coordinated, one thread may take longer than others if it happens to use more computationally expensive preprocessors, and the threads' results must be merged in the end. Still, depending on your computer, using an appropriate number of threads will much speed up calculations. The downside is that because each thread works independently and changes the input data (e.g. by applying different pre-processing steps), the memory requirements for processing also increase by factor k. Thus, if you get error messages saying that there is not enough memory, try using fewer threads.

All common operating systems include system monitor programs that allow you to check the CPU load and memory use while running the EcoImpactMapper. They can help you choose the best number of threads.

Logging

The EcoImpactMapper automatically sets up a log file each time it is started. These files are found in the "Logs" folder in your executable's directory. They contain the information shown in the status window, as well as additional information about processes occurring in the background. The log files' primary purpose is to help debugging the EcoImpactMapper. But you may also use them to find information about previous sessions, for example, about where the results from yesterdays' work were saved. They also contain detailed information about each model run in the uncertainty and sensitivity analyses.

The files are created, but not deleted, automatically. If you think you don't need them, it is fine to delete them periodically.

Contact, reporting bugs & support

You can contact me by writing to astock{at}stanford{dot}edu. I'd be grateful to receive reports of any bugs or other troubles – including unclear explanations in this tutorial – associated with the EcoImpactMapper, and will do my best to answer questions and fix any issues quickly.

Acknowledgments

While developed independently, the EcoImpactMapper draws on my experience in the HARMONY project. When first developing the EcoImpactMapper, I was funded by a Davis Family E-IPER Fellowship. I'm currently maintaining and improving it under a Goldman Graduate Fellowship from Stanford's Woods Institute for the Environment.