

# Marine Connectivity Explorer

**Risk Assessment Tool for assessment of the risk of natural dispersal of marine non-indigenous species**

## Documentation

December 2025

## Contents

1	Introduction	3
2	Biophysical modelling	4
2.1	Hydrodynamic data	4
2.2	Agent Based Model (ABM)	5
2.3	Scenarios	6
3	Marine connectivity database	7
3.1	Connectivity grid	7
3.2	Connectivity matrices	8
3.3	Total release	8
3.4	Salinity statistics matrices	8
3.5	Data formats	8
4	User Interface	8
4.1	Overview	9
4.2	Query input (left side panel)	9
4.2.1	Step 1 – Select Marine Connectivity Database	10
4.2.2	Step 2 – Select Dispersal Traits	10
4.2.3	Step 3 – Select Habitat Preferences	12
4.2.4	Step 4 – Run Query	13
4.2.5	Query Summary (Main panel)	14
4.3	Connectivity Analyses (Main panel)	16
4.3.1	Dispersal Probability Map tab	16
4.3.2	Port Connectivity Matrix tab	19
4.3.3	Clustering Analysis	19
4.3.4	Sink Source maps	21
5	Notes	22
5.1	Model boundaries	22
5.2	Mortality	23
5.3	Interpretation of port dispersal probability	23
5.4	Interpretation of cluster maps	23
5.5	Risk assessment	24
6	References	24
	Appendix	26

# 1 Introduction

This is the documentation of the Marine Connectivity Explorer, - an r-shiny app, which in this current version, is specifically designed to support the work process of risk assessment of the natural dispersal of marine non-indigenous species and their pelagic life-stages, as proposed by Hansen et al. (2024).

Marine connectivity can be defined as a network of links and exchanges—of organisms, genes, nutrients, energy, and even human activities—between different parts of the ocean, crucial for maintaining biodiversity, ecosystem health, and effective management,

Analyses of Marine connectivity of pelagic life stages of sessile benthic organisms typically involve the development, setup and execution of biophysical models for predicting the dispersal of pelagic life-stages. A biophysical model is a combination of hydrodynamic data originating from a hydrodynamic model in terms of predicted current velocities, and a dispersal model, predicting the dispersal of pelagic life stages as a function of simulated current velocities. Based on the results from the biophysical model, connectivity matrices are calculated by aggregating all the connectivity information into a dense data format for further analysis.

The biophysical modelling is often time consuming both in terms of man hours and computational time. The present concept described in this document demonstrates an alternative to this more traditional work process of marine connectivity analyses. This alternative concept (figure 1) relies on a large number of general biophysical model simulations and pre-calculated connectivity matrices stored in a data storage or database. The way the simulations are designed and the way the calculated matrices are structured, the connectivity analysis can be done on-the-fly producing a variety of maps and other connectivity related outputs within a few minutes. Prior to connectivity analyses, the user will have to select several input parameters including dispersal traits, habitat preferences and environmental tolerances. Based on the user selection, connectivity matrices are extracted from the connectivity database and processed and compiled into a single matrix representing the user inputs.

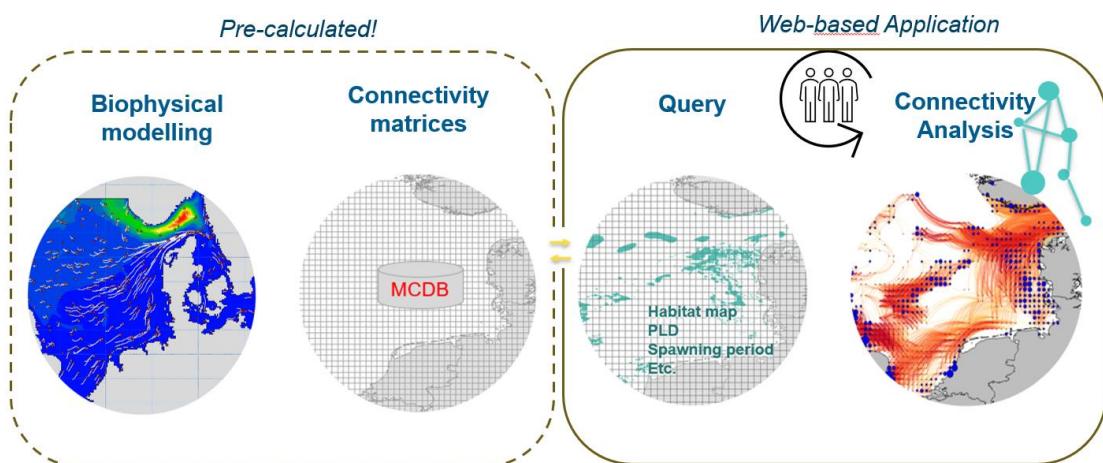


Figure 1 The conceptual approach of the Marine Connectivity Explorer

The development of the tool has been funded by the Danish Maritime Fund and the Danish Ministry of Higher Education and Science.

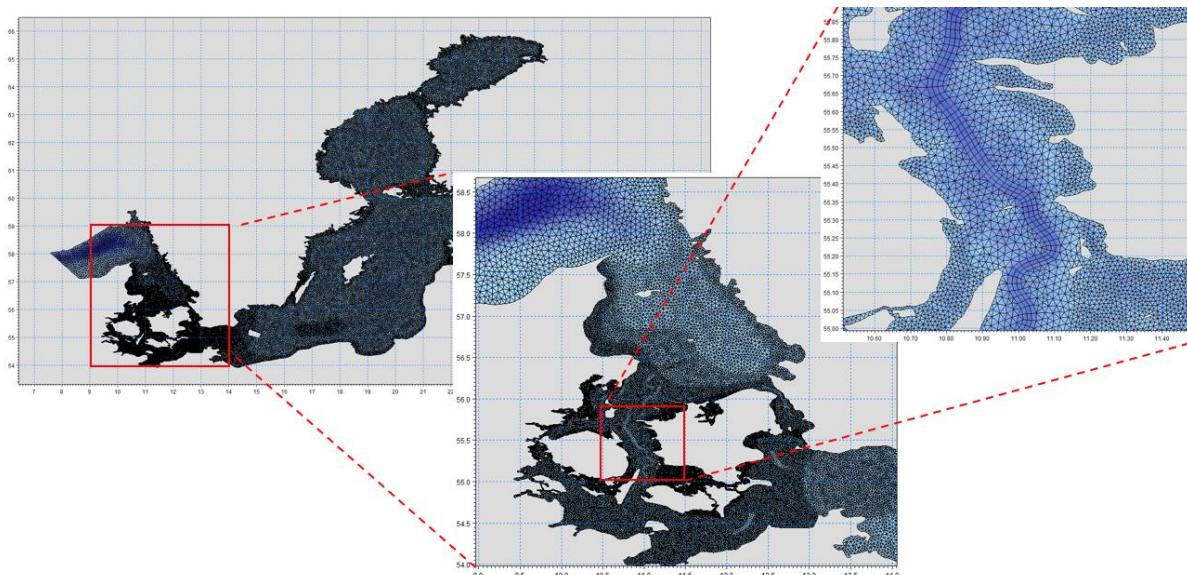
## 2 Biophysical modelling

### 2.1 Hydrodynamic data

The hydrodynamic data originates from a MIKE 3 FM hydrodynamic model of the Baltic Sea, the Inner Danish Straits, Kattegat and Skagerrak (DHI 2018). The modelling system (DHI 2025a) is based on the numerical solution of the three-dimensional (3D) incompressible Reynolds averaged Navier-Stokes equations invoking the assumptions of Boussinesq and of hydrostatic pressure.

The computational grid (Figure 2) is based on finite element triangular and squared mesh with a horizontal spatial resolution varying between  $\leq 5$  km in the open and central parts of the system down to ca. 0.5 km in the narrow straits and parts of the coastal areas (measured as the distance between centre points of neighbouring elements that share a common element outline). The vertical resolution of up to 243 layers, and consists of 10 sigma layers of a maximum thickness of 1 m of the upper part of the water column, up to 210 z-layers of 1 m thickness of the intermediate depth ranges and up to 23 z-layers of increasing thicknesses from 5 m – 20 m within depth ranges of the bottom of the deepest parts of the area in Skagerrak. The model grid is based on geographical coordinates using datum WGS84.

The model was originally run using a 5 min timestep. Result files were saved with 1 hour storage frequency including horizontal (u and v) and vertical (w) velocity components, salinity and temperature.



**Figure 2** The computational mesh used in the MIKE 3 FM hydrodynamic model for the Danish Straits and the Baltic Sea (DHI 2018). Grid resolution ranges from approx. 0.5 to 5 km

The depth within the study area ranges between ca. 0 - 700 m (Figure 3, left). Water levels in the inner Danish straits and western Baltic Sea are primarily determined by meteorological driven water movements, and only secondarily by tides. Salinity conditions are highly variable in the central parts of the system comprising in particular the

southern parts of Kattegat and the Inner Danish Straits, including both spatial (horizontal and vertical) and temporal variations (Figure 3, right).

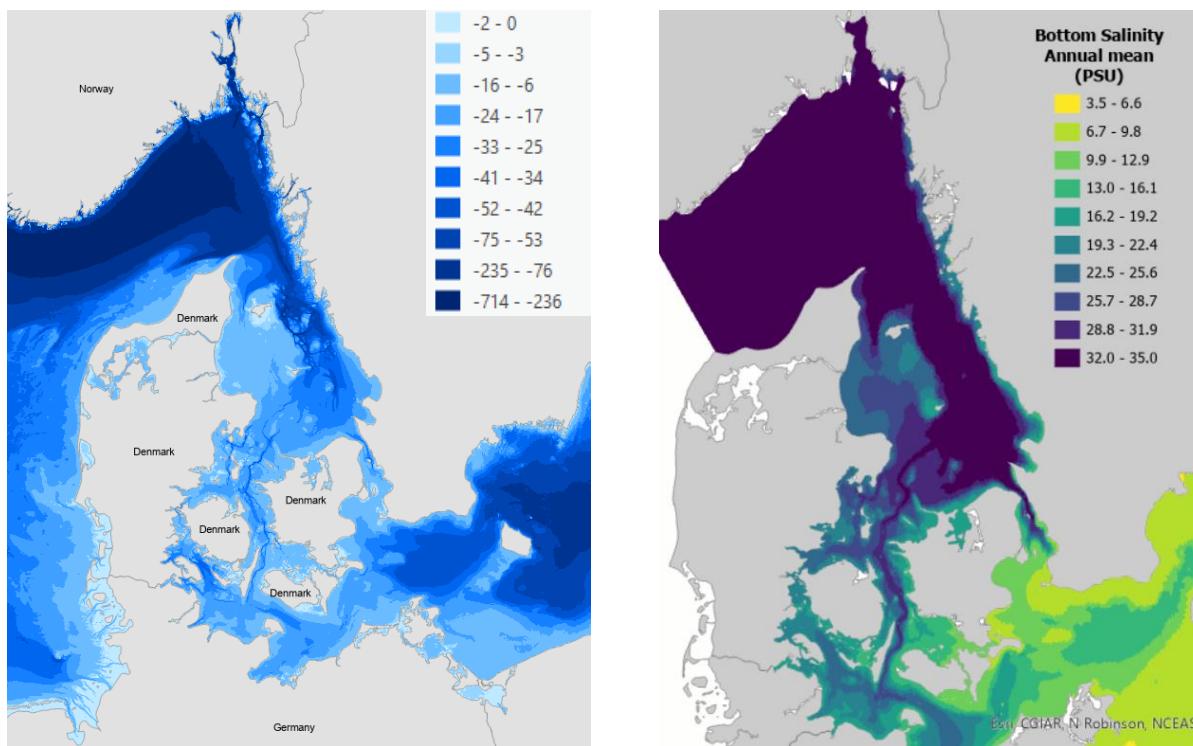


Figure 3 Left: Bathymetry data (m) based on the GEBCO bathymetry dataset: IOC, IHO and BODC (2003). Legend using quantiles. Right: Bottom salinity, annual means, DHI (2012). .

## 2.2 Agent Based Model (ABM)

The dispersal of pelagic life-stages were simulated using the agent-based modelling framework, ABM Lab, which is an integrated modelling component of the open equation solver, ECO Lab. ECO Lab is an add-on module to MIKE hydrodynamic simulation modelling software designed for building and executing ecological models, combining and integrating Eulerian (concentration based) and Lagrangian (agent-based) models (DHI 2025b).

Simulations were setup for each day of the period 1 Jan – 10 Nov, for each year. In each simulation 100 000 agents were released (spawned) at the seabed randomly distributed in space within the study area and randomly distributed in time within the first 24 hours of the simulation. Each simulation was run for 50 days, saving results for each day (1-50). To optimize the simulation time for each simulation a sub-mesh was extracted of the original computational mesh covering the entire Baltic Sea to include only the most western parts of the Baltic Sea starting from east of Bornholm with an open boundary condition allowing agents to disperse across the boundary.

Agent state-variables saved for each agent include:

- age
- X, Y position start
- X, Y position end

- Z-position at the time of “settling”
- Minimum salinities experienced since release/spawning
- Maximum salinities experienced since release/spawning
- Trapped – no. of timesteps with no horizontal movement

A simulation timestep of 15 minutes was used in the ABM simulations. Result files were saved every 24 hours of the simulation duration for 50 days. Each result file represents an age distribution approximately uniformly distributed between the time stamp of the result file and the preceding 24 hours (mean  $\sim (a - a_{t-1})/2$ , st.dev.  $\sim 0.28$ ; a =age, t: day number).

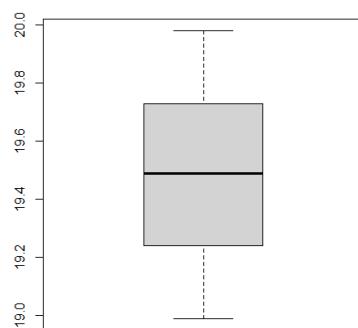


Figure 4 Boxplot of an example of the age distribution of 98436 agents in a result file representing day 20. Mean: 0.489. St.dev.: 0.285.

Agents may get stuck during the simulation when hitting the land-sea-interface due to limitations in the spatial resolution of the model, due to actual stranding processes, or due to agents getting trapped in local depressions in the computational grid. The horizontal movement is not adjusted when hitting the land-sea interface. The agent will remain at the position until currents will turn away from the shore. To adjust for agents getting trapped at the seabed due to local depressions in, or scale dependent simplifications of, the computational grid, vertical adjustments were included to move agents upwards 1 m per timestep until horizontal movement continues. Furthermore, to minimize the number of agents getting trapped, agents were forced to a position of 1 m above seabed, i.e. if the z-position of an agent was less than 1 meter above seabed, it was moved 1 m upwards the following time step. For scenarios (see later) considering drift confined to a certain depth or depth interval, a positioning adjustment was included in terms of a vertical speed corresponding to the vertical distance from the depth threshold divided by the time step. For postprocessing purposes a state variable keeps track of how many timesteps each agent hits the land-sea-interface.

To account for hydrodynamic currents not resolved by the hydrodynamic model horizontal (1 m<sup>2</sup>/s) and vertical (0.0001 m<sup>2</sup>/s) dispersion factors were included.

## 2.3 Scenarios

Two scenarios were run for each year and for each release day representing passive drift in 3D assuming neutral buoyancy (referred to as “3D”) and surface drift constrained to the upper 0 – 1 m depth (referred to as “SF”).

Each scenario was run for 3 years including 2005, 2010 and 2012. These three years were selected based on the North Atlantic Oscillation index representing a neutral, a negative and a positive NAO index respectively. For details see: Hansen et al. (2018).

## 3 Marine connectivity database

The connectivity database is based on a subdivision of the study area (a connectivity grid) and comprises pre-calculated connectivity matrices, matrices representing statistics on salinity experienced during dispersal of each agent, and data files keeping track of the number of agents released. These are calculated for each combination of year (2005, 2010 and 2012), scenario (3D or SF), release day of the year (1 Jan – 1 Nov) and drift duration (1-50 days). A total of 364 800 data files.

### 3.1 Connectivity grid

Connectivity matrices were computed based on a subdivision of the study area divided into a quadratic grid of 10x10 km using Lambert Equal Area Projection (LAEA Europe ETRS89), covering an area between 3423856 - 4063856 N, and 3423856 - 4063856 E, comprising in total  $44*64=2816$  grid elements of which 1291 elements include sea water. For coding simplicity and for generic purposes a matrix dimension of 2816x2816 rows/columns were used and sparse matrix formats were applied in data processing when required for optimization.

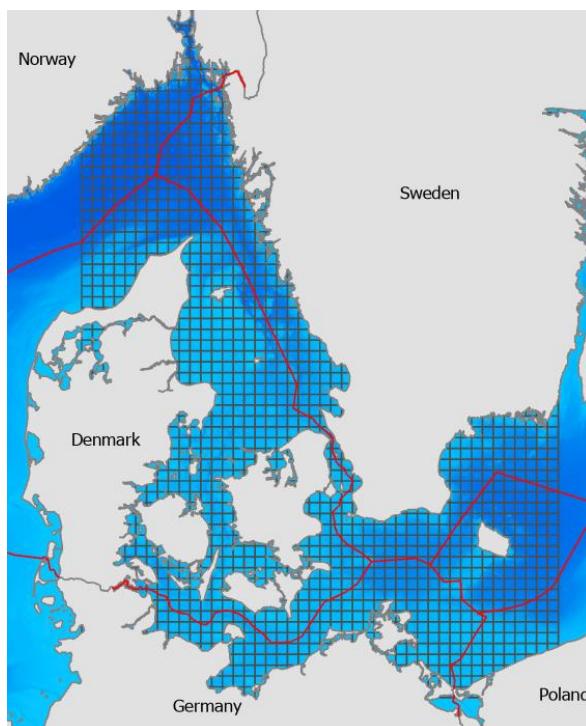


Figure 5 The extent of the subdivision of the study area into a 10x10 km connectivity grid. Red lines indicate the EEZ boundaries between Denmark, Sweden, Norway, Germany and Poland.

## 3.2 Connectivity matrices

For each result file (year, release day, and day number since release), start and end positions of all agents were used to populate a connectivity matrix where the number in each matrix element represents the number of agents with a start position in the specific element (= row number) that has an end position in another (or the same) specific element (column number). Diagonals represent local retention, i.e. the number of agents in each element with a start and end position in the same element.

Prior to calculation of connectivity matrices, to reduce the bias of entrapment of agents hitting the land-sea interface, all agents hitting land more than 5 % of the time since release were excluded. Most of all sea-land interactions are assumed to be due to limitations in the spatial resolution of the hydrodynamic model and/or the time step used.

## 3.3 Total release

The total number of agents released in each connectivity grid element on each release day, was stored in separate files. Total release is used for calculating connectivity probability matrices (see later). Total number of agents released include agents that may have been transported out of the model domain but exclude agents exceeding the 5% criteria for entrapment.

## 3.4 Salinity statistics matrices

Salinity statistics matrices include salinity values representing the mean of the minimum and the mean of the maximum salinity of all agents in each connection (=connectivity matrix element).

## 3.5 Data formats

All data files are stored in fst-file format (<http://www.fstpackage.org>) which supports a multithreaded serialization of compressed data frames. Each data file includes strings referring to release date and no. of days since release. Data file prefixes include "cmn", "rel", "csmn" and "csmx" representing the connectivity matrix, total release per connectivity grid element, matrix including mean of minimum salinity experience by agents and matrix including mean of maximum of salinity experienced by agents, respectively:

cmn\_"yyyy"-“mm”-“dd”\_”days-since-release”.fst (e.g. cmn\_2012-05-22\_14fst)

## 4 User Interface

The user interface, IU, is developed using the R-shiny package in R and all data processing was done using R version 4.5.1 (R core team 2024). All R-packages used are listed in appendix.

## 4.1 Overview

The user interface, UI, is divided into a left panel and several tabs in the main panel.

The “About” tab in the main panel include information about the tool including a map of the areal extent and a subdivision of the area represented by a connectivity grid of which connectivity matrices have been calculated, countries outline and EEZ boundaries between countries.

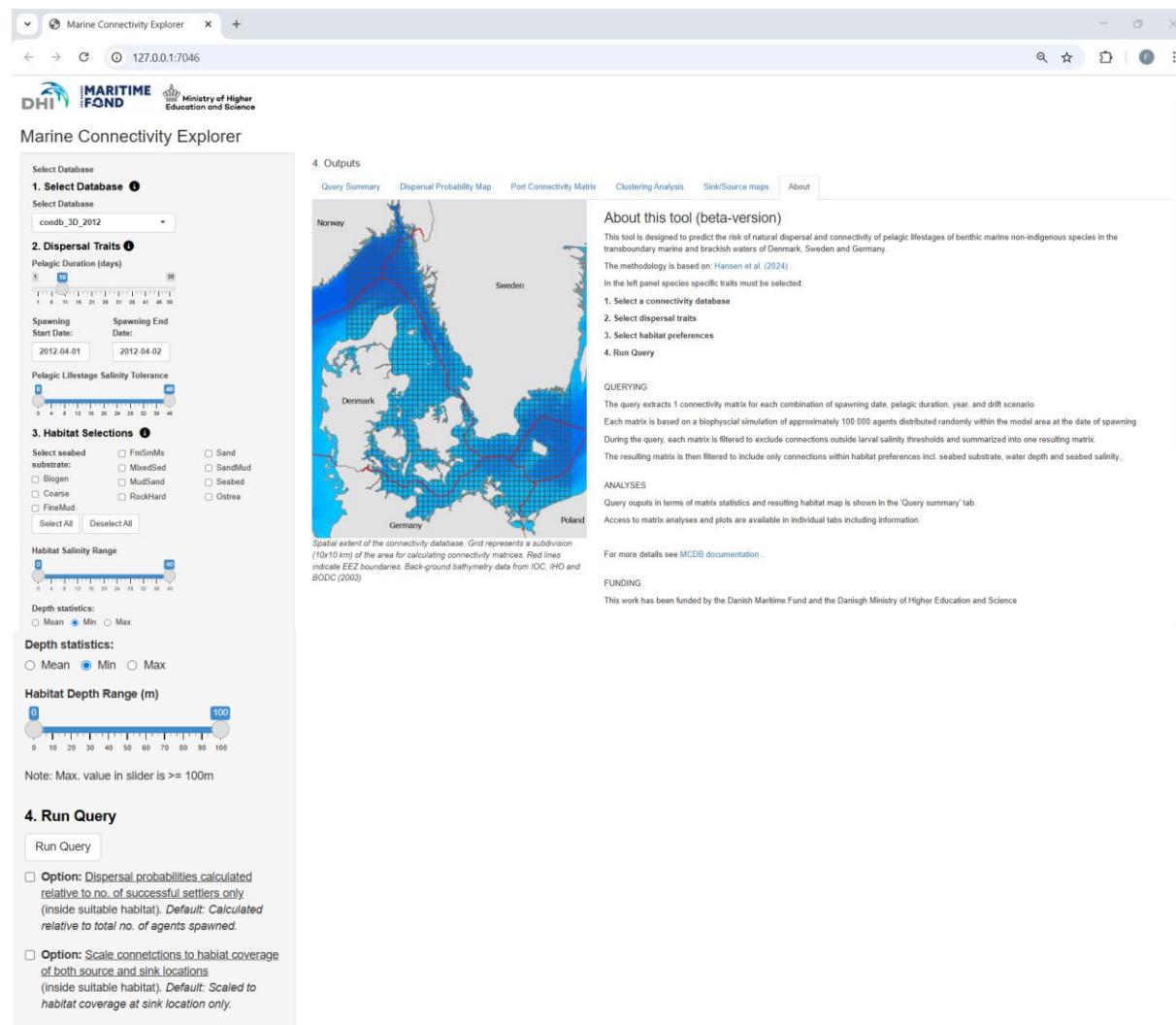


Figure 6 Marine Connectivity Explorer User Interface

## 4.2 Query input (left side panel)

The query input in the left side panel is divided into 4 sections:

- Select connectivity database
- Select Dispersal Traits
- Select Habitat preferences
- Run query

A query is a look-up in the connectivity database or data storage to extract matrices and data that match the user input in the left panel. After specifying the user input, the “run” button can be activated to start data extraction. At the end of the data extraction, the main panel will change to the “Query summary” tab. Once a query has been run, different types of analysis can be selected to create multiple connectivity analysis outputs, see section 4.3.

#### 4.2.1 Step 1 – Select Marine Connectivity Database

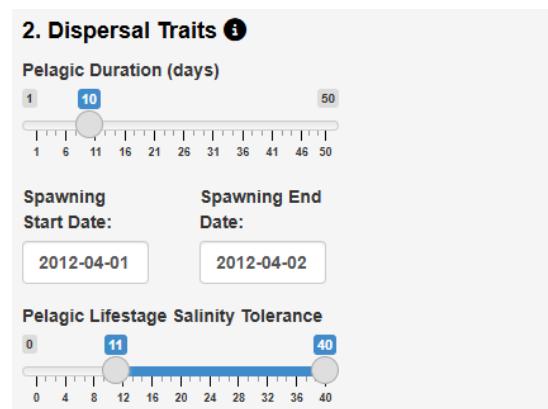
The datafiles in the connectivity database are grouped by ‘year’ and “drift scenario” referring to the year of the biophysical modelling simulation (2005, 2010, or 2012) and the drift scenario (3D passive drift or Surface drift 0-1 m). The database is selected from a drop-down menu “Select Database”:



#### 4.2.2 Step 2 – Select Dispersal Traits

The dispersal traits available for selection include:

- Pelagic life-stage duration (PLD)
- Spawning start and end dates
- Pelagic life-stage salinity tolerance



Pelagic life-stage duration (PLD) represents the drift/dispersal duration from the time of spawning and until the specified PLD value and ranges between 1 and 50 days. Note that for the highest PLD values and for locations of high current velocities, a large number of agents may be conveyed out of the model domain, and this can affect analysis outputs (section 5.3). In general, the larger the PLD values the more uncertain the analysis outputs become close to the model boundaries.

Spawning start date and end dates are selected from the calendar drop down menus. Selected dates cannot exceed the date range from 1 January to 1 November within selected year. Only dates within the chosen year (of selected database) can be selected.

Pelagic Life Stage Salinity Tolerance is the salinity range tolerated by the pelagic life stage during dispersal. Connections where the mean of minimum and/or mean of maximum salinities experienced by the individual agents during dispersal are exceeding the salinity tolerance thresholds specified, are discarded from the connectivity matrices, i.e. elements in the connectivity matrix are set to "0". In addition, the salinity conditions at the seabed corresponding to the horizontal location where each agent is located at the time of the end for the PLD (= corresponding to the time of settling) is included to exclude connections where seabed salinity exceeds the pelagic life stage salinity tolerance. The salinity conditions at the seabed are extracted from the 3D hydrodynamic salinity data as daily values (extracted at 12.00 am) at the centroid's positions of each element in the connectivity grid. The salinity conditions at the seabed are necessary to include since the vertical positions of the agents at the time of the end of the chosen PLD will be distributed throughout the water column (in the 3D scenario) or at the surface (in the SF scenario). Hence, the evaluation of the bottom salinity conditions at the current horizontal position of each agent assumes agents settle instantaneous within a single time step.

Note that while each connectivity matrix in the database represents a combination of spawning day and dispersal duration (PLD) the salinity tolerance thresholds are evaluated matrix by matrix. Thus, some connections may be included for one or some of the daily connectivity matrices while the same connections may be excluded in other daily matrices where dispersal conditions are different and the salinity conditions are exceeded.

The resulting connectivity matrix,  $S$ , represents the sum of connectivity matrices extracted from the database within the spawning period excluding the connections exceeding pelagic life stage salinity tolerance thresholds:

$$(1) \quad M1_{ij}^{(k)} = \begin{cases} 1, & \text{if } T_{min} \leq L_{ij}^{(k)} \leq T_{max} \text{ and if } T_{min} \leq U_{ij}^{(k)} \leq T_{max} \\ 0, & \text{otherwise} \end{cases}$$

$$(2) \quad M2_{ij}^{(k)} = \begin{cases} 1, & \text{if } T_{min} \leq b_j^{(n)} \leq T_{max} \text{ and if } T_{min} \leq b_j^{(n)} \leq T_{max} \\ 0, & \text{otherwise} \end{cases}$$

$$(3) \quad S = \sum_{k=1}^n (A^{(k)} \circ M1^{(k)} \circ M2^{(k)})$$

- $k = 1, \dots, n$ : Release day number in the release/spawning period specified.
- $A^{(k)} = [a_{ij}] \in \mathbb{R}^{n \times n}$  : Connectivity matrix for release day  $k$  (number of agents moving from  $i$  to  $j$  from release day  $k$  to the end of PLD)
- $L^{(k)} = [l_{ij}] \in \mathbb{R}^{n \times n}$  : Mean of minimum salinity experienced by agents during dispersal from connectivity grid element  $i$  to element  $j$  (during dispersal beginning at release day  $k$  and to the end of PLD)
- $U^{(k)} = [u_{ij}] \in \mathbb{R}^{n \times n}$  : Mean of maximum salinity experienced by agents during dispersal from connectivity grid element  $i$  to element  $j$  (during dispersal beginning at release day  $k$  and to the end of PLD)

- $M1^{(k)} = [m_{ij}] \in \mathbb{R}^{n \times n}$ : Mask matrix for masking the connections from connectivity grid element i to j exceeding the pelagic life stage salinity tolerance thresholds (during dispersal beginning at release day k and to the end of PLD)
- $M2^{(k)} = [m_{ij}] \in \mathbb{R}^{n \times n}$ : Mask matrix for masking the connections from connectivity grid element i to j where the bottom salinity at the time of settling (end of PLD) exceed the pelagic life stage salinity tolerance thresholds.
- $T_{min}$ : Pelagic Life-stage salinity tolerance threshold constant, minimum
- $T_{max}$ : Pelagic Life-stage salinity tolerance threshold constant, maximum

#### 4.2.3 Step 3 – Select Habitat Preferences

The habitat preferences available for selection include:

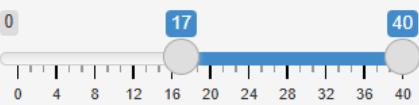
- Seabed Substrate
- Habitat Salinity Range
- Habitat Depth Range

**3. Habitat Selections** 

**Select seabed substrate:**

<input type="checkbox"/> Biogen	<input checked="" type="checkbox"/> Coarse	<input type="checkbox"/> FineMud	<input type="checkbox"/> FmSmMs	<input checked="" type="checkbox"/> MixedSed	<input type="checkbox"/> MudSand	<input type="checkbox"/> Sand	<input type="checkbox"/> SandMud	<input type="checkbox"/> Seabed	<input type="checkbox"/> Ostrea
---------------------------------	--	----------------------------------	---------------------------------	--	----------------------------------	-------------------------------	----------------------------------	---------------------------------	---------------------------------

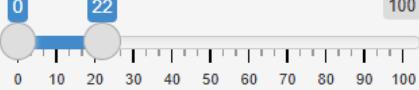
**Habitat Salinity Range**



**Depth statistics:**

Mean  Min  Max

**Habitat Depth Range (m)**



Note: Max. value in slider is >= 100m

Seabed substrate classes are derived from EU EMODNET (2018) where seabed substrates are divided into 11 classes. The seabed substrate data is available as GIS layers as polygons. For this tool the coverage of each substrate class within each 10x10 km element in the connectivity grid were calculated. The sum of all substrate classes in each element equals 1. The substrate coverages are used to weight the connectivity matrix representing the sum of the extracted connectivity matrices from the data base, step 2. This is done by multiplying the connectivity matrix rows and/or columns by the substrate coverage of the corresponding element in connectivity grid.

Habitat salinity range refers to the salinity tolerance range of the post-settlement sessile juvenile and adult life stages. For some species the upper and lower salinity tolerance thresholds are not always identical to the salinity tolerance threshold of the pelagic life stages. The habitat salinity range is used to exclude all habitats where salinity tolerance for juveniles and adults is exceeded. This is done by discarding connections in the connectivity matrix, step 2, i.e. by setting all values in the rows and columns of the connectivity matrix that correspond to habitat elements outside the salinity tolerance range to "0". Habitat salinity range is evaluated based on annual minimum and maximum bottom salinities extracted from the hydrodynamic model for each centroid of the 10x10 km elements in the connectivity grid. A gross extent of the habitat is used, where annual minimum and/or maximum salinities fall inside the habitat salinity range.

Similarly, the Habitat Depth Range refers to the depth interval of suitable seabed habitat. The depth range is used to exclude habitats outside the selected depth range. This is done by discarding connections in the connectivity matrix, step 2, i.e. by setting all values in the rows and columns of the connectivity matrix that correspond to habitat elements outside the selected habitat depth range to "0". Depth data are based on GEBCO bathymetry raster data (IOC, IHO and BODC 2003). For this tool depth statistics of each 10x10 km element in the connectivity grid were calculated including mean, minimum and maximum depth. The choice of depth statistic to use defining the extent of the suitable habitat in relation to water depth, is chosen using the available radio buttons. Default choice is set to "minimum" depth representing the maximum extent of habitats with increasing depth.

The resulting connectivity matrix,  $W$ , represents the filtered (depth and salinity range) and weighted (seabed substrate) connectivity matrix from step 2:

$$(4) \quad W = S \circ [(h^\alpha \circ m)(h^\beta \circ m)^T]$$

Where:

- $S = [s_{ij}] \in \mathbb{R}^{n \times n}$ : The connectivity matrix as the sum of extracted connectivity matrices (equation 3) filtered to exclude connections outside the pelagic life stage salinity thresholds (equations 1 and 2). Each matrix element represents the dispersal from connectivity grid element  $i$  to element  $j$ .
- $h = (h_1, \dots, h_n)^T \in [0,1]^n$ : Habitat Seabed Substrate habitat coverage vector
- $m = (m_1, \dots, m_n)^T \in \{0,1\}^n$ : Mask vector for habitat depth and salinity range
- exponents  $\alpha, \beta \in \{0,1\}^n$  controlling whether habitat weighting is applied to origin and/or the destination (rows and columns respectively) is applied (see: "options", step 4)

#### 4.2.4 Step 4 – Run Query

The Run Query section includes a button to start data extraction and two options that can be selected prior to initiating the data extraction.

The button "Run Query" executes a script code based on the selection inputs in steps 1-3, and the corresponding equations 1-4.

#### 4. Run Query

- Option:** Dispersal probabilities calculated relative to no. of successful settlers only  
(inside suitable habitat). **Default:** Calculated relative to total no. of agents spawned.
- Option:** Scale connections to habitat coverage of both source and sink locations  
(inside suitable habitat). **Default:** Scaled to habitat coverage at sink location only.

Prior to running a query, the two options include:

- Option 1: Dispersal probabilities calculated relative to no. of successful settlers only (inside suitable habitat). **Default:** Calculated relative to total no. of agents spawned.
- Option 2: Weight connections by habitat coverage of both source and sink locations (inside suitable habitat). **Default:** Weight by habitat coverage at the sink location only.

Options 1 relates to how to calculate dispersal probabilities between each pair of elements in the connectivity grid. This option is used for the available connectivity analyses accessible via individual tabs in the main panel (see section 5.3).

By default, the dispersal probabilities (from origin i to destination j) are calculated relative to the total no. of released agents inside the origin element, i.e. the sum of released agents per day during the selected spawning period. The total release includes both agents that settle within suitable habitats as well as agents settling outside suitable habitats, including agents that may be conveyed across open boundaries of the biophysical model, and agents settled outside the connectivity grid, but excluding agents that may be discarded due to entrapment of agents hitting the land-sea interface more than 5 % of the time (see: section 4).

When selecting option 1, dispersal probabilities (from origin to destination) are calculated relative to no. of successful settlers only, i.e. the of number agents that originates from element “a” that settle successfully inside suitable habitat in element “b”, excluding agents released that settle outside suitable habitats.

Option 2 relates to how to weight connections by habitat coverage of source and/or sink locations. By default, connections are weighed by habitat coverage of sink locations only. When selecting this option connections are weighted by habitat coverages of both source and sink locations.

#### 4.2.5 Query Summary (Main panel)

Proceeding the finalisation of data extraction from the database (step 4, section 5.2.4), the query summary will be displayed in the “Query Summary” tab in the main panel:

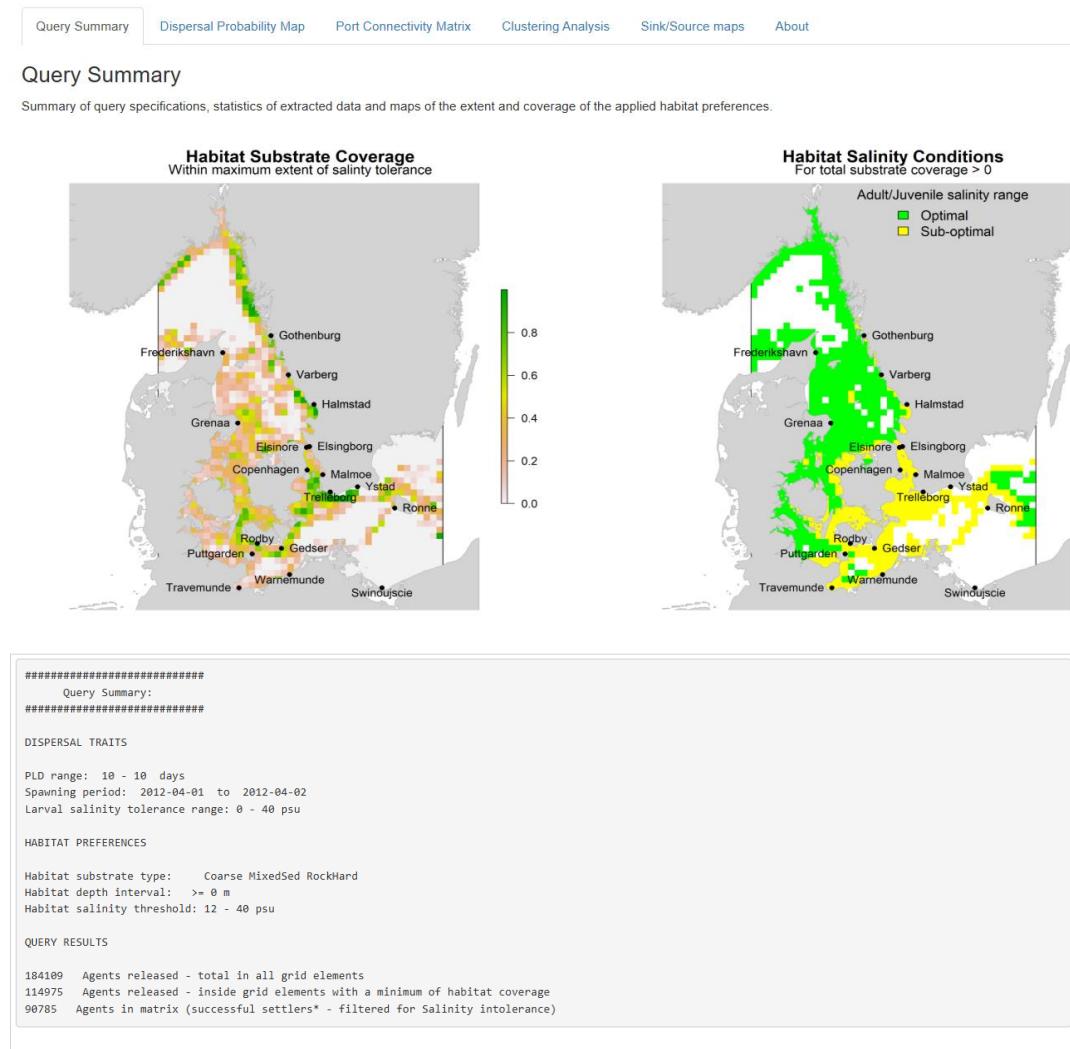


Figure 7. Top left: Habitat coverage (0-1) of selected seabed substrates within salinity threshold tolerance range for juvenile/adult life-stages. Top right: Habitat Salinity Conditions above seabed for juvenile/adult life-stages. Bottom: Summary of query selections and output matrix statistics. See text for more information.

The Query Summary include:

- Habitat Coverage Map
- Seabed Salinity Conditions Map
- Query Summary Results

The habitat coverage map shows the extent and the coverage of the habitat preferences selected in the left panel based on seabed substrate, water depth and bottom salinity conditions. The habitat coverage value between 0 and 1 represent the proportion of the connectivity grid element covered by the sum of the coverages of the selected seabed substrates.

Seabed salinity condition map classifies habitat quality as "optimal" or "sub-optimal" based on consistent adherence to adult salinity tolerance. "Optimal" conditions indicate connectivity grid elements where bottom salinity consistently aligns with the habitat salinity range, while "sub-optimal" conditions involve periods where salinity exceeds (higher or lower) habitat salinity range. The larger the distance to the optimal salinity

conditions (green), the more frequent and/or the longer periods juvenile and adult life stages may be exposed to salinity conditions outside the tolerated habitat salinity range.

The Query Summary Results summarize the selections made in the left panel for extracting, filtering and weighting matrices from the connectivity database, and a summary of the number of agents represented by the extracted matrices:

- Agents released - total in all grid elements

This is the sum of the total number of agents released each day during the selected spawning period and within the entire connectivity grid but prior to discarding connections outside pelagic life stage salinity tolerance thresholds and connections outside preferred habitat conditions. The numbers include agents that may be transported out of the model domain across open boundaries and agents settled outside the connectivity grid, but exclude agents discarded due to land entrapment (section 4.2)

- Agents released - inside grid elements with a minimum of habitat coverage

This is the sum of agents released each day during the spawning period inside grid elements in the connectivity grid with a minimum of habitat coverage only, and prior to matrix weighting by seabed substrate coverage.

- Agents in matrix (successful settlers- filtered for Salinity intolerance)

This is the sum of the total number of agents released each day during the spawning period inside the grid elements of the connectivity grid with a minimum of habitat coverage that connects to other elements with a minimum of habitat coverage, i.e. equal to the sum of agent numbers in the resulting matrix produced by the query prior to habitat coverage weighting.

## 4.3 Connectivity Analyses (Main panel)

Apart from the “About” and “Summary” tabs, four additional tabs are used for connectivity analyses to be conducted after running a query:

- Dispersal Probability Map
- Port Connectivity Matrix
- Cluster analysis
- Sink/Source Maps

### 4.3.1 Dispersal Probability Map tab

In the dispersal probability map tab, dispersal probabilities from and to selected ports, can be extracted and mapped. The user inputs include:

- Port selection
- Dispersal direction:
- Downstream (export)

- Upstream (import)
- No. of generations (1 or 2)
- Recruitment rate (0 – 1)

Ports are selected from the “Select port” drop-down list. The corresponding coordinates of the port location are shown as greyed out. Coordinates are in the LAEA ETRS89 projection. Note that some ports are duplicated with the extensions ..1 and ..2. These duplicates are included for ports where the port location is close to a boundary between 2 neighbouring elements in the connectivity grid, or if the element of the connectivity grid where the port is located includes a large proportion of land. In such cases, dispersal probabilities to and from the port can be extracted from an alternative neighbouring element. Port names with extension ..1 will refer to the element in the connectivity grid with the actual location of the port.

“Dispersal direction” indicates which dispersal probability map to display:

- “Downstream (export)” refers to the dispersal probability that an agent with a start position in the element of the selected port will end up in each of the other connectivity grid elements. Probabilities are calculated relative to the total number of agents released (default) or relative to the total number of successful setters (optional), depending on the selection in the left panel prior to database querying, see section 5.2.4.
- “Upstream (import)” refers to the probability of an agent with an end position in the element of the selected port that will have a start position in each of any of the other connectivity grid elements. Probabilities are calculated relative to the total number of agents with an end position inside the element of the selected port.

“No. of generations” refers to the dispersal via 1 generation (direct dispersal) or via 2 generations (stepping-stone dispersal). Dispersal via 2 generations is calculated following Legrand et al. (2021):

$$(5) \quad P^{G=2} = P + rP(L \circ P)$$

Where:

- $P = [p_{ij}] \in \mathbb{R}^{n \times n}$  : The downstream (export) or upstream (import) connectivity probability matrix with values representing the dispersal probability from element i to element j in the connectivity grid, or to element i from element j, respectively.
- $G$  : Generation number.
- $L = [l_{ij}] \in \mathbb{R}^{n \times n}$  : All-ones matrix minus the identity matrix, i.e. diagonal values set to “0”.
- $r$ : Between generation recruitment rate (between 0 and 1)

A “recruitment rate” can be set between 0 and 1 to reflect the recruitment probability in-between the 2 generations. The recruitment rate represents the proportion of “settled” life stages that survive to the adult stage to produce a new generation.

Probability values between 0.001 and  $10^{-6}$  are shown with hatched symbolic. Values less than  $10^{-6}$  are not shown.

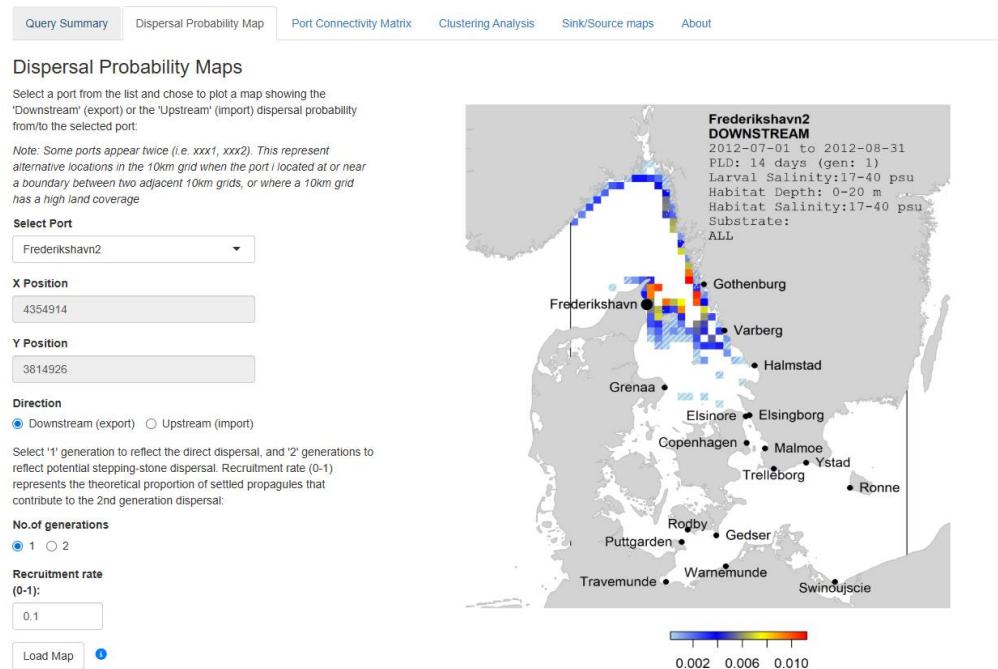


Figure 8 Example of a Downstream (export) dispersal probability map.

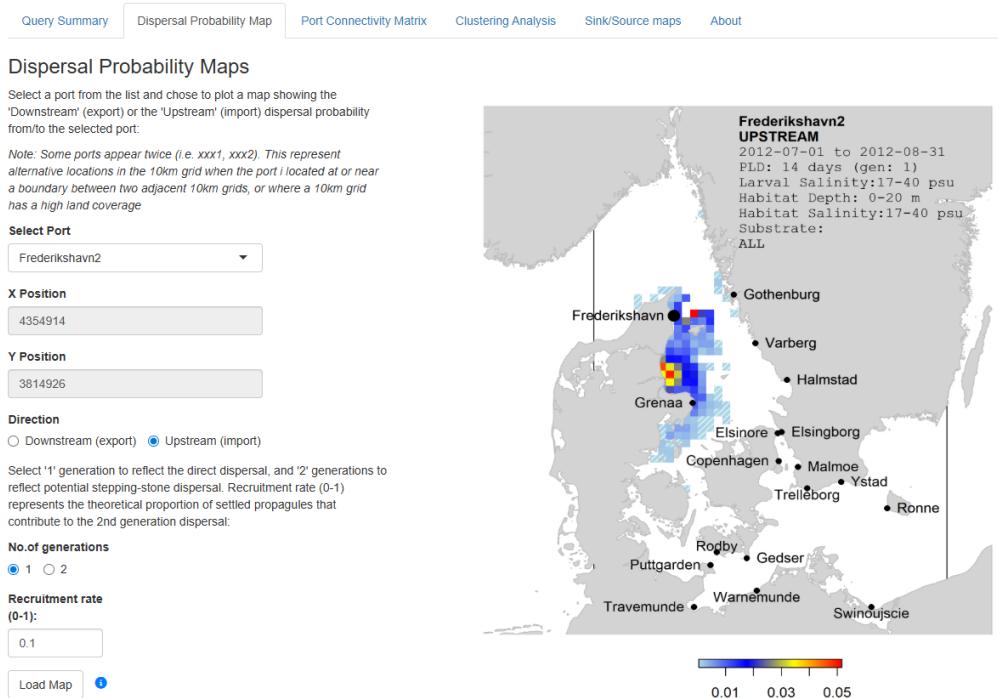


Figure 9. Example of an upstream (import) dispersal probability map.

### 4.3.2 Port Connectivity Matrix tab

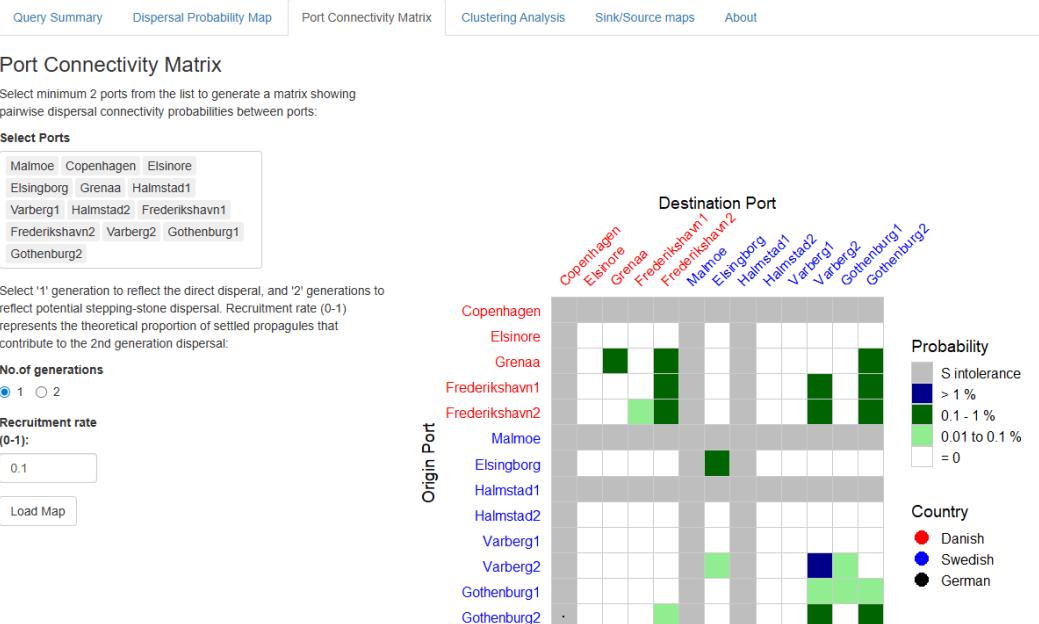
In the port connectivity matrix tab, a port connectivity matrix can be generated representing a summary of pairwise connectivity probabilities (downstream) between selected ports. The user inputs include:

- Port selection
- No. of generations (1 or 2)
- Recruitment rates (0 - 1)

Ports are selected in the “Select Ports” input box. A minimum of 2 ports must be selected.

“No of generations” and “Recruitment rate (0-1)” are explained in previous section, section 5.3.1.

The “Load Map” button plots the port connectivity matrix including the selected ports. Matrix rows represent the Origin Port (from), and the matrix columns represent the Destination Port (to). Port names are grouped by country: Red = Denmark; Blue = Sweden; Black = Germany and/or Poland. Matrix element colours represent dispersal probability intervals. White elements indicate no connectivity within the threshold set as input parameters. Grey elements indicate port connections outside the salinity tolerance threshold for pelagic life-stages.



### 4.3.3 Clustering Analysis

In the cluster analysis tab, cluster analysis (also referred to as community detection) identifies clusters of connectivity grid element with relatively strong connections between elements within the same cluster and with relatively weaker connections to elements in other clusters.

The user inputs include:

- Select cluster algorithm
  - Louvain
  - Infomap
- Gamma (Louvain only)
- Min. Cluster size

The output from the cluster analysis can be used to identify potential dispersal barriers assuming the species are distributed throughout all potential habitats evenly or scaled to the habitat coverage depending on the substrate coverage selected, section 5.2.5. Prior to clustering the connectivity matrix is translated into a graph, i.e. a data structure that accommodate network analysis techniques referred to as graph theory. Graphs consist of nodes and edges where nodes in marine connectivity context refer to the individual elements in a connectivity grid, and edges refer to the connections between nodes in terms of total number of agents or probabilities of individual connections.

The cluster algorithms available include the Louvain and the Infomap algorithms.

The Louvain algorithm (Blondel et al. 2008) iteratively improves community assignments for maximizing modularity (e.g. Pastor et al. 2015). The Louvain algorithm assumes undirected graphs, and while graphs based on connectivity matrices are directed (edge weights differ in opposite directions), these edge weights are collapsed:

$$(6) \quad w_{undirected}(A, B) = w(A, B) + w(B, A)$$

Where:

$w_{undirected}(A, B)$ : Weights of the undirected edges between nodes A and B.

$w(A, B)$ : Weights between nodes A and B

$w(B, A)$ : Weights between nodes B and A

The Infomap algorithm (Rosvall et al. 2009) minimizes the description (information) length of a random walker's movements, accommodating directionality of the graph and producing variable community sizes (e.g. Pastor et al. 2025).

For the Louvain algorithm a  $\lambda$  (gamma) parameter can be set for producing fewer and coarser clusters ( $\lambda < 1$ ) or more and smaller clusters ( $\lambda > 1$ ). Default value is "1",

For both algorithms a minimum size of a cluster can be set representing a minimum number of connectivity grid elements.

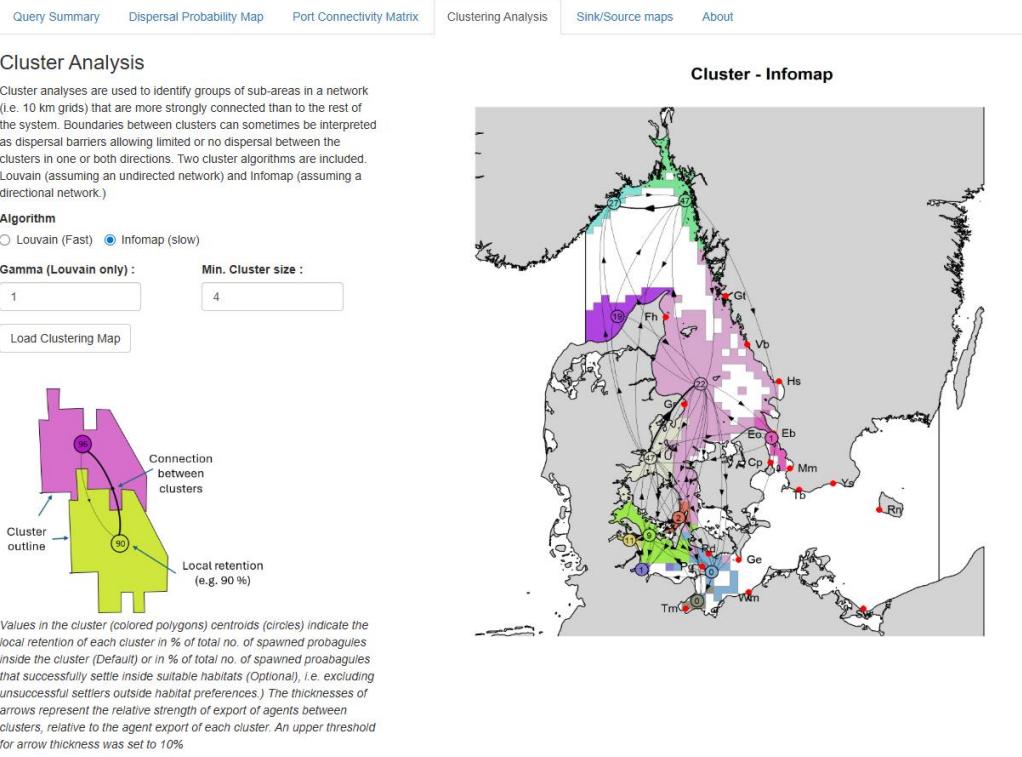


Figure 11. Example of a cluster map based on the Infomap cluster algorithm:

#### 4.3.4 Sink Source maps

In the sink/source maps tab, selected graph metrics can be shown in a map. Graph metrics refers to several metrics that can be calculated for any graph representing specific graph properties. Some of these properties are relevant for ecological networks like the marine connectivity networks. For more details: Treml et al. (2008), Ospinaz-Alvarez et al. (2020). Here we include 2 sets of metrics that relate to the identification of sinks and sources:

- In- and out-degree
- In- and out-strength

*Example of an output showing graph metric representing sinks:: “in-strength”*

Degree refers to the number of nodes (connectivity grid elements) each node is connected to, i.e. the number of edges (connections) between a node to and from other nodes without considering how many agents each edge is represented by. E.g. a connection counts as “1” for both an edge represented by only one agent and an edge represented by 100 agents. In-degree refers to the number of in-going edges to each node (=import). Out-degree refers to the number of out-going edges from each node (=export).

Strength, also known as “weighted degree”, is calculated as the sum of weights, for in-strength,  $s_i^{out}$  (=import) and out-strength,  $s_i^{in}$  (=export), respectively.

Out-strength:

$$(7) \quad s_i^{out} = \sum_{j \in V} w_{ij}$$

In-strength:

$$(8) \quad s_i^{in} = \sum_{j \in V} w_{ji}$$

Where:

w<sub>ij</sub> : number of agents (or probability of dispersal) from node i to node j

w<sub>ji</sub> : number of agents (or probability of dispersal) from node j to node i

V : nodes in the network

In-strength values are normalised to represent the fraction (0-1) of the sum of imports to all nodes in the network ("no. of agents" or "no. of agents weighted relative to habitat coverage"), where e.g. 0.015 corresponds to 1.5 % of the sum of all node-imports.

Out-strength values are normalized to represent the fraction (0-1) of the sum of exports from all nodes in the network ("no. of agents" or "no. of agents weighted relative to the habitat coverage"), i.e. 0.015 corresponds to 1.5 % of the sum of all node-exports.

The identification of major sinks can provide guidance for the most likely locations for observing one or multiple species if the species has been observed but the distribution is unknown. As an example, see Lindegren et al. (2022) where predicted larval settlements are equivalent to in-strength in the current tool.

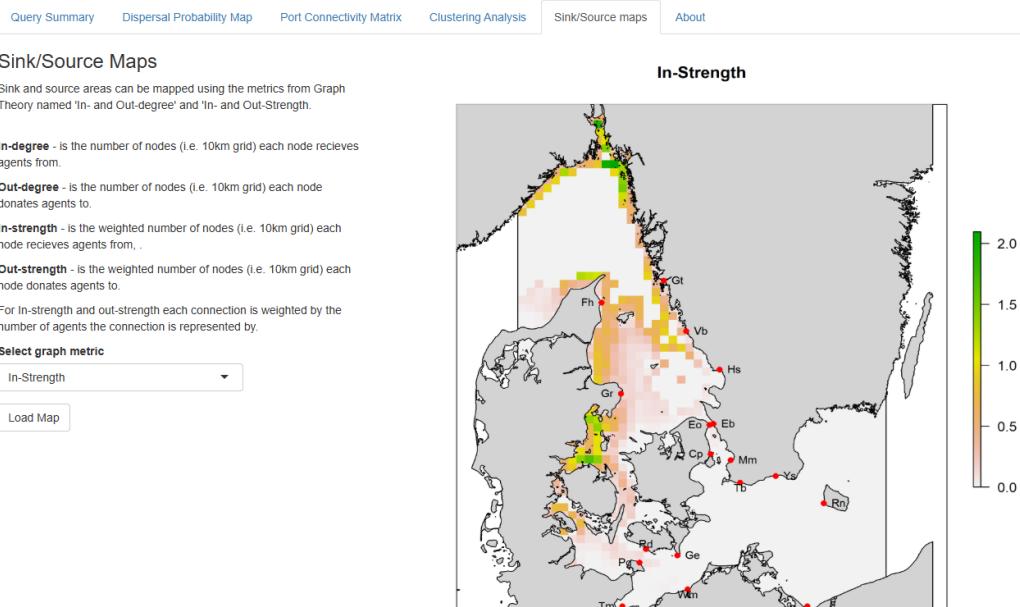


Figure 12. Example of an output showing graph metric representing sinks: "in-strength".

## 5 Notes

### 5.1 Model boundaries

In the biophysical model simulated agents may migrate out of the extent of the connectivity grid, and further across the open boundaries of the model domain. Thus, the connectivity metrics may become more uncertain close to the boundaries toward the

north-northwestern part of the connectivity grid (in Skagerrak) and to the southeastern part of the connectivity grid (in the Baltic Sea east of Bornholm). While this uncertainty is small and limited to an area close to the boundary for short dispersal durations (PLD), the extent of the area affected will increase with increasing PLD. A quick check on how boundaries may affect the dispersal at and around a selected port, a plot of the dispersal probabilities to and from the port can give an indication if the dispersal probabilities extend within or beyond the outline of the connectivity grid.

## 5.2 Mortality

Mortality is not included explicitly in the biophysical modelling. Instead, mortality can be included in the postprocessing in two ways:

- By setting pelagic life-stage salinity tolerance thresholds prior to running a query. Connections outside the salinity threshold range are considered subject to mortality.
- By manually adjusting the predicted connectivity probabilities calculated, i.e. if the dispersal probability range in a port connectivity matrix is assigned a probability range of e.g. 0.01-0.001, an assumption of a “background” mortality of pelagic life stages of e.g. 90 %, the probability range in the port connectivity matrix can be adjusted manually by dividing by 10 to 0.001-0.0001. This assumes that the background mortality of the duration of the pelagic life stage is constant in time and space,

## 5.3 Interpretation of port dispersal probability

The dispersal probability to and from port locations are summarized in the port connectivity matrix. The absence of port connections in the summary matrix can be double-checked by plotting the import and/or export dispersal probability for the ports of interest. If the dispersal probability maps indicate dispersal probabilities from neighboring ports reaching connectivity grid elements close to the port location of interest, a port connection may be considered. This is especially the case if selected habitat substrates are included in the query. Habitat substrate mapping based on EU EMODNET dataset is relatively coarse, and local habitat substrate may not be sufficiently represented, e.g. hard substrates of man-made structures including port structures, coastal protection, etc.

When selecting a specific habitat substrate type where local knowledge suggests that the habitat data is not adequate, a query including all habitats substrates selected can give an indication if ports may be connected or not.

In cases where larval salinity tolerance thresholds are included, local salinity conditions in a port location may be affected by large freshwater run-off from rivers nearby (e.g. Port of Halmstad). River outflow is included in the hydrodynamic model data used.

## 5.4 Interpretation of cluster maps

Clustering is based on iterative computations and optimizations and uses different types of algorithms. Cluster algorithms are all approximations of a highly complex, high-dimensional system involving a very large number of potential connection pathways among sites. Thus, applying different clustering algorithms and different clustering

parameters will improve the understanding of the connectivity network, - i.e. in some parts of the marine connectivity network the location and outline of clusters may be consistent between methods, while in other parts of the network cluster locations and outline may be more “fuzzy”. For more details on this topic, see Pastor Rollan et al. (2025). The interpretation of the clustering results is also recommended to include an evaluation of the between-cluster connectivity as well as the local retention.

When selecting option 1 (section 4.2.4: "Dispersal probabilities calculated relative to no. of successful settlers only (inside suitable habitat)", the local retention in each cluster will indicate the proportion of agents with a start position inside each cluster that settle in the same cluster. If this value is close to 100 %, this will indicate a very low export to habitats in other clusters, while a small value, in contrast, will indicate a high export to habitats in other clusters.

To assess the degree to which each port belongs to its assigned cluster, the export and import probability maps for each port can be compared with the outlines of the individual clusters.

## 5.5 Risk assessment

When using the tool in a risk assessment context, the connectivity metrics, should not be the sole source of information to consider. The need for supplementary risk assessment indicators is discussed in Hansen et al. (2024):

1. Dispersion Potential Indicator: Provides insight into the extent to which the NIS may utilize the identified potential connectivity.
2. Impact Potential Indicator: Evaluates the potential impact of the NIS.
3. Invasion Status Indicator: Indicates the presence, absence, or distribution of the NIS in the SRA candidate.
4. Recognition by Authorities Indicator: Notes whether the NIS is recognized by local/regional authorities as a target species.
5. Anthropogenic Dispersal Indicator: Considers whether the NIS may be subject to additional anthropogenic dispersal vectors, such as hull fouling, or via networks of aquaculture sites.

## 6 References

Blondel V D, Guillaume J L, Lambiotte R & Lefebvre E (2008) Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.*

DHI (2018) DKBS2 Hydrodynamic Model Setup and Validation. Technical Note, August 2018

DHI (2025a) MIKE 3 Flow Model FM. Hydrodynamic and transport module. Scientific documentation.

[https://manuals.mikepoweredbydhi.help/latest/Coast\\_and\\_Sea/MIKE\\_3\\_Flow\\_FM\\_Scientific\\_Doc.pdf](https://manuals.mikepoweredbydhi.help/latest/Coast_and_Sea/MIKE_3_Flow_FM_Scientific_Doc.pdf)

DHI (2025b) MIKE ECO Lab. Numerical Lab for Ecological and Agent Based Modelling. Short scientific documentation.

[https://manuals.mikepoweredbydhi.help/latest/General/MIKE\\_ECO\\_Lab\\_ShortScientific.pdf](https://manuals.mikepoweredbydhi.help/latest/General/MIKE_ECO_Lab_ShortScientific.pdf)

EMODNET (2018) European Marine Observation and Data Network. European Marine Observation Data Network (EMODnet) seabed habitats project. [www.emodnetseabedhabitats.eu/](http://www.emodnetseabedhabitats.eu/) Accessed 2018

Hansen F T, Pastor A, Christensen A, Stuer-Lauridsen F (2024) Using Biophysical modelling and marine connectivity to assess the risk of natural dispersal of non-indigenous species to comply with the Ballast Water Management Convention. *Biol Invasions*. Published online: 09 May 2024. <https://DOI:10.1007/s10530-024-03327-0>

IOC, IHO and BODC (2003) Centenary Edition of the GEBCO Digital Atlas", published on CD-ROM on behalf of the Intergovernmental Oceanographic Commission and the International Hydrographic Organization as part of the General Bathymetric Chart of the Oceans. British Oceanographic Data Centre, Liverpool

Legrand T, Chenuil A, Ser-Giacomi E et al. (2022) Spatial coalescent connectivity through multi-generation dispersal modelling predicts gene flow across marine phyla. *Nat Commun* 13, 5861 <https://doi.org/10.1038/s41467-022-33499-z>

Lindgren M, Gabellini AP, Munk P et al (2022) Identifying key processes and drivers affecting the presence of non-indigenous marine species in coastal waters. *Biol Invasions* 24:2835–2850. <https://doi.org/10.1007/s10530-022-02802-w>

Ospina-Alvarez A, de Juan S, Alós J, Basterretxea G et al. (2020) MPA network design based on graph theory and emergent properties of larval dispersal. *Mar Ecol Prog Ser* 650:309–326 <https://doi.org/10.3354/meps13399>

Pastor Rollan A, Berlow E L, Williams R et al. (2025) Managing the fuzzy boundaries and partitions of marine ecological systems using network theory. *Sci Rep* 15, 25803 <https://doi.org/10.1038/s41598-025-09601-y>

R Core Team (2024) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. <https://www.R-project.org/>

Rosvall M, Axelsson D & Bergstrom C T (2009) The map equation. *Eur. Phys. J. Special Top.* 178, 13–23

Trembl EA, Halpin Urban DL, Pratson LF (2008) Modelling population connectivity by ocean currents, a graph-theoretic approach for marine conservation. *Landscape Ecol* 23:19–36. <https://doi.org/10.1007/s10980-007-9138-y>

## Appendix

R packages used:

Package	Version	Maintainer
calibrate	1.7.7	Jan Graffelman <jan.graffelman@upc.edu>
ConnMatTools	0.3.5	David M. Kaplan <dmkaplan2000@gmail.com>
data.table	1.17.4	Tyson Barrett <t.barrett88@gmail.com>
diagram	1.6.5	Karline Soetaert <karline.soetaert@nioz.nl>
dplyr	1.1.4	Hadley Wickham <hadley@posit.co>
fields	16.3.1	Douglas Nychka <douglasnychka@gmail.com>
fs	1.6.6	Gábor Csárdi <csardi.gabor@gmail.com>
fst	0.9.8	Mark Klik <markklik@gmail.com>
ggplot2	3.5.2	Thomas Lin Pedersen <thomas.pedersen@posit.co>
gttext	0.1.2	Brenton M. Wiernik <brenton@wiernik.org>
grid	4.5.1	R Core Team <do-use-Contact-address@r-project.org>
gttable	0.3.6	Thomas Lin Pedersen <thomas.pedersen@posit.co>
igraph	2.1.4	Kirill Mäller <kirill@cynkra.com>
jsonlite	2.0.0	Jeroen Ooms <jeroenooms@gmail.com>
lubridate	1.9.4	Vitalie Spinu <spinuvit@gmail.com>
mapplots	1.5.2	Hans Gerritsen <hans.gerritsen@marine.ie>
mapproj	1.2.12	Alex Deckmyn <alex.deckmyn@meteo.be>
maps	3.4.3	Alex Deckmyn <alex.deckmyn@meteo.be>
Matrix	1.7-3	Martin Maechler <mmaechler+Matrix@gmail.com>
randomcoloR	NA	NA
raster	3.6-32	Robert J. Hijmans <r.hijmans@gmail.com>
RColorBrewer	1.1-3	Erich Neuwirth <erich.neuwirth@univie.ac.at>
reshape2	1.4.4	Hadley Wickham <h.wickham@gmail.com>
rlist	0.4.6.2	Kun Ren <ken@renkun.me>
sf	1.0-21	Edzer Pebesma <edzer.pebesma@uni-muenster.de>
shapefiles	0.7.2	Ben Stabler <benstabler@yahoo.com>
shiny	1.11.1	Winston Chang <winston@posit.co>
shinyBS	0.61.1	Eric Bailey <ebailey@idem.in.gov>
shinybusy	0.3.3	Victor Perrier <victor.perrier@dreamrs.fr>
shinyFiles	0.9.3	Thomas Lin Pedersen <thomas.pedersen@posit.co>
shinyjs	2.1.0	Dean Attali <daattali@gmail.com>
sp	2.2-0	Edzer Pebesma <edzer.pebesma@uni-muenster.de>
svDialogs	1.1.0	Philippe Grosjean <phgrosjean@sciviews.org>
stringr	1.5.1	Hadley Wickham <hadley@posit.co>
this.path	2.7.0	Iris Simmons <ikwsimmo@gmail.com>
tidyR	1.3.1	Hadley Wickham <hadley@posit.co>
tools	4.5.1	R Core Team <do-use-Contact-address@r-project.org>
utils	4.5.1	R Core Team <do-use-Contact-address@r-project.org>
viridis	0.6.5	Simon Garnier <garnier@njit.edu>
XML	3.99-0.18	CRAN Team <CRAN@r-project.org>