PAME II-2020: Agenda 4.1
Modelling Arctic oceanographic connectivity to further develop PAME's MPA toolbox
Draft report

Lead country: Sweden (Swedish Agency for Marine and Water Management, SwAM)
Dr Jessica Nilsson, SwAM
Prof. Göran Broström and Prof. Per Jonsson, Department of Marine Sciences, University of Gothenburg

Context

This project aims to further develop PAME's toolbox for Marine Protected Areas (MPA) by mapping oceanographic connectivity in the Arctic region using biophysical modelling, and to use this data resource to identify major barriers to dispersal, gene flow and efficient design of MPA networks. The research may assist PAME to promote an 'ecologically connected, representative and effectively managed network of protected and specially managed areas'. The project could further inform the scoping process that is planned in the Conservation of Arctic Flora and Fauna's (CAFF) Circumpolar Biodiversity Monitoring Program (CBMP) and other users. This project communicates with the vast scientific community connected to PAME, as well as with CAFF, WWF and others.

Due to Covid 19 the project has been slightly delayed and a complete research report will be presented at PAME I-2021. The project team will continue reaching out to the PAME community and other parties during fall 2020.

Time plan

- The project started in November 2018
- Endorsed at PAME I-2019
- Preliminary test for particle tracking model (July 2019)
- Literature study on key species, their distribution, spawning season, and larvae behavior (September 2019)
- Main particle simulation result (Jan 2020)
- MPA analysis (April 2020)
- Trend analysis (October 2020)
- Delivery of report (December 2020)
- Presentation at PAME I-2021

Modelling Arctic oceanographic connectivity to further develop PAME's MPA toolbox

Table of Contents

I. Executive Summary

To be added

II. Objectives

This report (to be completed with further research, and presented at PAME I-2021) is aimed as a working document and is a contribution to the protection of the Arctic marine and coastal environments from the Swedish delegation within the Arctic Council working group 'Protection of the Arctic Marine Environment (PAME)'. The overall aims are to contribute to the PAME's MPA toolbox by:

- mapping **connectivity** among populations of key marine species in the Arctic region using biophysical modelling;
- describing dispersal distance within the Arctic Sea which may be used to infer the **minimum size** of MPAs for sufficient self-recruitment;
- identifying **sources and sinks** for MPAs and other valuable areas, e.g. fish spawning grounds; and
- identifying major **barriers** to dispersal and gene flow.

III. Background

Protection of the Arctic Marine Environment

The Arctic Council working group *Protection of the Arctic Marine Environment* (PAME) aims to suggest and support science and actions for the protection of the Arctic marine environment. Ongoing climate change is already facilitating increased access to the Arctic region leading to greater use and new economic opportunities, but also bringing potential threats to the Arctic marine and coastal environments. These changes require more integrated approaches to Arctic marine protection, including spatial planning of protection measures to support sustainable use of the Arctic environment.

Marine protected areas – an important tool for ecosystem-based management

Networks of Marine Protected Areas (MPAs) are considered to be an effective instrument to mitigate extractive and local disturbance effects on harvested stocks, general biodiversity, and ecosystem services (Lester & Halpern 2008). MPAs are also believed to improve local resilience to large-scale pressures like effects from climate change (Micheli et al. 2012). When selecting areas as MPAs it is important to ensure that they are designed to have the capacity to protect target populations. As an example, one design criteria can be to create MPAs that provide 'ecologically coherent' networks that ensure dispersal, migration and genetic exchange of individuals between relevant sites improving resilience to disturbance or damage caused by natural and anthropogenic factors (OSPAR 2013, HELCOM 2016, Jonsson et al. 2020). Key aspects of this design are the size of individual MPAs and how they are connected through dispersal to the ambient environment. Organisms with long-distance dispersal may require very large MPAs, or a network of smaller MPAs that can exchange dispersal stages within the network, or with surrounding areas.

Biophysical modelling of dispersal

More than 70% of marine invertebrates and fish disperse with large numbers of microscopic eggs and larvae that may drift for days to months with the ocean circulation. Although larvae disperse passively, they may nonetheless influence their transport through vertical migration to different depths with different current patterns. Thus, it is very difficult to make direct observations of dispersal in the field, although genetic methods can be used to coarsely infer dispersal (e.g. Jorde et al. 2015, Jahnke et al. 2018). *Biophysical modelling* is increasingly used to estimate dispersal in the seascape (Cowen & Sponaugle 2009, van Sebille et al. 2018). Here a physical circulation model first predicts how the physical water transport, as well as temperature, salinity and ice, vary in space and time. The oceanographic model of sea currents is then combined with a biological model that defines the traits for a particular species (or dispersal strategy), such as spawning time, drift duration of the larvae (pelagic larval duration, PLD), and any larval behaviour, e.g. vertical migration or ontogenetic shifts in drift depth. Such a biophysical model can simulate the transport paths of virtual larvae *in silico* allowing the "release" of many millions of virtual larvae at many spawning sites (sources), and includes temporal variability in currents on scales from days to years.

The model results of each larval dispersal simulation are summarized in a connectivity matrix where each element gives the probability of dispersal from site A to site B (Jonsson et al. 2020). For the area included in the model, the connectivity matrix fully describes connectivity for the target species (or the dispersal strategy) in the seascape. A major advantage with a database of connectivity matrices based on combinations of trait values as spawning time, drift depth and PLD, is that these matrices can be combined and averaged to represent many different species (Jonsson et al. 2020).

IV. Methods

Review of published information on larval dispersal traits in the Arctic Ocean

A starting point of this research was to review studies from the Arctic Sea that report data on traits for larvae of marine invertebrates and fish that may affect dispersal patterns. The focus is on dispersal of planktonic larvae that largely drift with the ocean circulation. Traits assumed to be most influential on dispersal are spawning season, the amount of time larvae drift in the water column (Pelagic Larval Duration, PLD), and the depth interval where larvae are drifting (Corell et al. 2012). There is also the possible occurrence of more complex, behavioural vertical migration schemes. However, considering the sparse literature, it is not surprising that this information is lacking at present. Thus, we will only consider relatively simple traits in this pre-study.

A number of search strings were used to query the Web of Science database (Clarivate Analytics), including 'marine', 'larva*', 'arctic', 'polar', 'dispersal', and 'connectivity'. We also received some input from Dr. Tom Christensen at the Arctic Council working group CAFF and Dr. Jørgen Hansen at the University of Aarhus. This search resulted in only few publications with data on larval traits, and only for few species. Only few relevant publications were found, but some data could guide the selection of major spawning time and dominant depth intervals for larval dispersal as simulated with the biophysical model.

Oceanographic circulation model

We have used two existing oceanographic models, the Arctic4 and the TOPAZ, that include the Arctic Sea and North Atlantic. *TOPAZ* (e.g., https://os.copernicus.org/articles/8/633/2012/) was developed at the Nansen Centre in Bergen, and the Norwegian Meteorological Institute (MET) makes an operational forecast once a week. It is also run in a re-analysis mode by the Nansen centre. It is the main operational, and re-analysis model for the Arctic Sea in the marine Copernicus data portal (https://marine.copernicus.eu). This model has stereographic projection with 12.5 km horizontal resolution and has 28 vertical (isopycnal) layers. Results for the 5 m level has been used for the present study (the TOPAZ products in the Copernicus data portal are interpolated to 12 unevenly (Levitus) spaced vertical levels: 5, 30, 50, 100, 200, 400, 700, 1000, 1500, 2000, 2500 and 3000 m), although a new re-analysis with a higher vertical resolution output at Copernicus data portal will be available in the near future. We here used daily averaged velocity fields from 1991 to 2014. The TOPAZ reanalysis do not contain tides.

The *Arctic4* model has a horizontal resolution of 4 km and 32 vertical sigma-coordinate layer (i.e. the layers follow the bottom topography and vertical resolution is thus higher at shallower regions) and was developed by MET and AkvaPlan-Niva with the aim to become an operational model, but this plan was abandoned with the introduction of TOPAZ as the main operational model for the Arctic Ocean. Arctic4 has been run for more than 10 years by researchers at MET and AkvaPlan-Niva but has not gone through the same extensive validation as is typical for operational models. In this study, daily averaged velocity fields between 2007-2016 were used. The Arctic 4 model contain tides.

All available velocity fields have been downloaded and pre-processed for use in a Lagrangian particle tracking model that simulates dispersal trajectories. The pre-processing includes extracting data north of 65°N and saving as single precision in Matlab native format for rapid reading; Arctic 4 data were interpolated to prescribed depth levels. Mean drift patterns as well as interannual variability in dispersal distance and connectivity have been targeted. It can be noted that there exist oceanographic model setups with higher resolution but results are not easily available for long-term analyses (i.e. may require new production of velocity fields from the hydrodynamical model). Furthermore, not all models use data assimilation as e.g. TOPAZ include.

Particle tracking model

A Lagrangian particle-tracking model uses available model prediction of ocean currents to move particles in the ocean (van Sebille et al. 2018). An in-house particle-tracking model was used and developed in MATLAB® (MathWorks Inc.), but there exist a number of similar models in the literature and downloadable on the web using different programming languages. For the present study, we used a time step of 1 hour since only daily averaged velocity fields were available. In the present analysis we only consider horizontal drift at a specified depth (although we use a number of specified depths, see below). Furthermore, we do not use any horizontal dispersion (or random walk) of particles as is frequently used to mimic unresolved turbulent processes. One reason is that the long-term mean of the particle fields (or velocity fields) are very dispersive in themselves (see e.g. Fig. 6) and a small “extra” dispersion does not change results significantly (as verified by preliminary tests).

A critical decision for the particle tracking simulations is how many release sites to consider, the number of release time points, and how many particles to release on each occasion. To provide non-trivial results for the last case would require numerical dispersion to be included. The ideal situation is if all model grid cells that overlap with the target habitat are included as sources of particles in the Lagrangian tracking model. As a compromise between conservation interests and practical limits of numerical computation it was decided to include all areas shallower than 500 m in the TOPAZ model bathymetry above the arctic circle. With a 12km grid, this area resulted in 40893 release points and this model domain is shown in Fig. 1. From each release point, one particle ('virtual larva') was released every day and tracked through its PLD. Daily release was repeated all-year-round for at least 10 years. For the Arctic4 model particles were released between 2007-2016, and for the TOPAZ model particles were released between 1991-2014. Based on velocity fields from the Arctic4 model particles were released and locked at the following depths: 0, 5, 10, 15, 20, 30, 50, 70, 100 and 150 m depth. Velocity fields from the TOPAZ model was only used to track particles in the surface layer (0-5 m). The positions of each particle were saved daily, but in subsequent analyses we only consider drift after 5, 10, 15, 30, 45, 60 and 90 days.

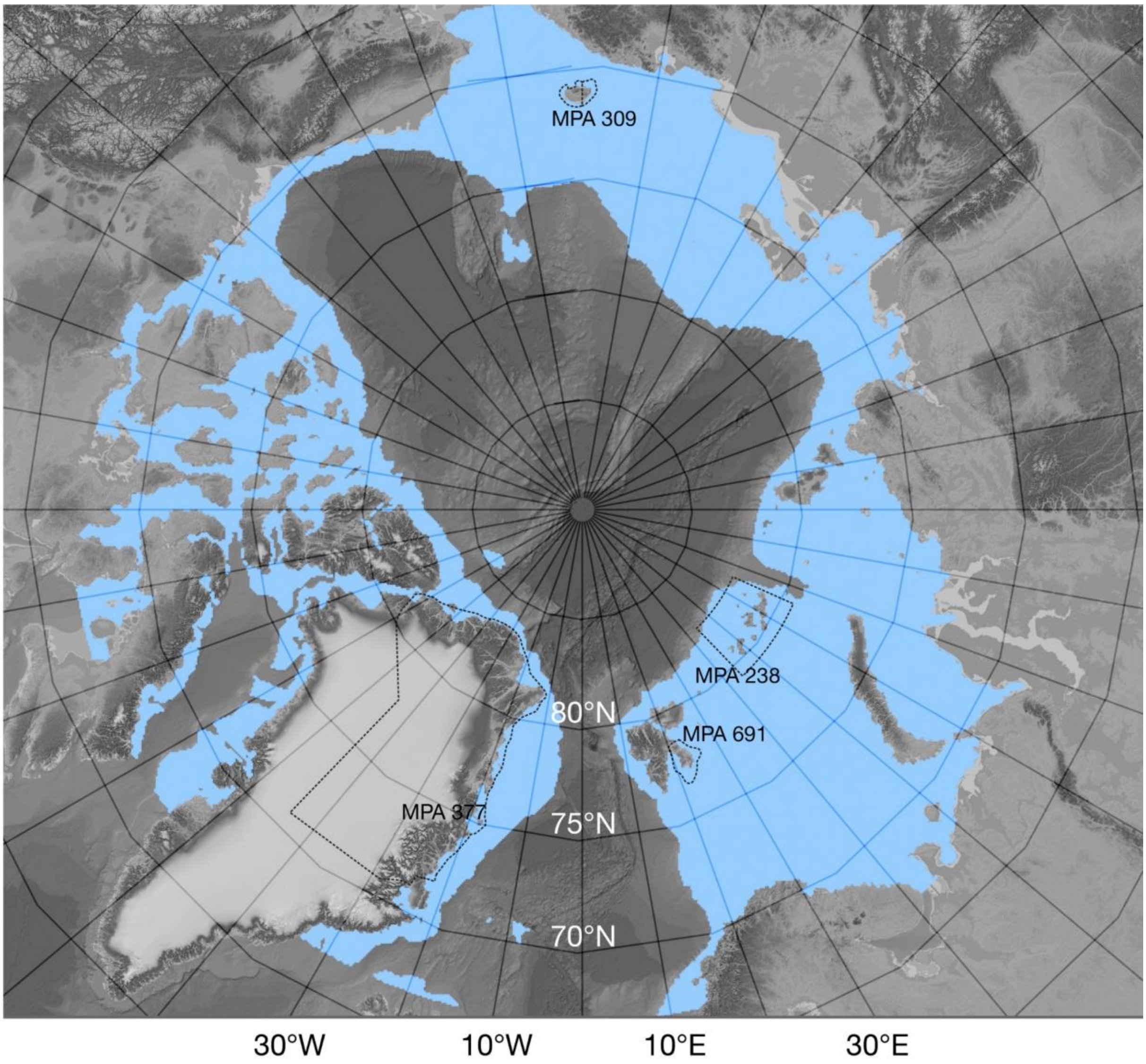
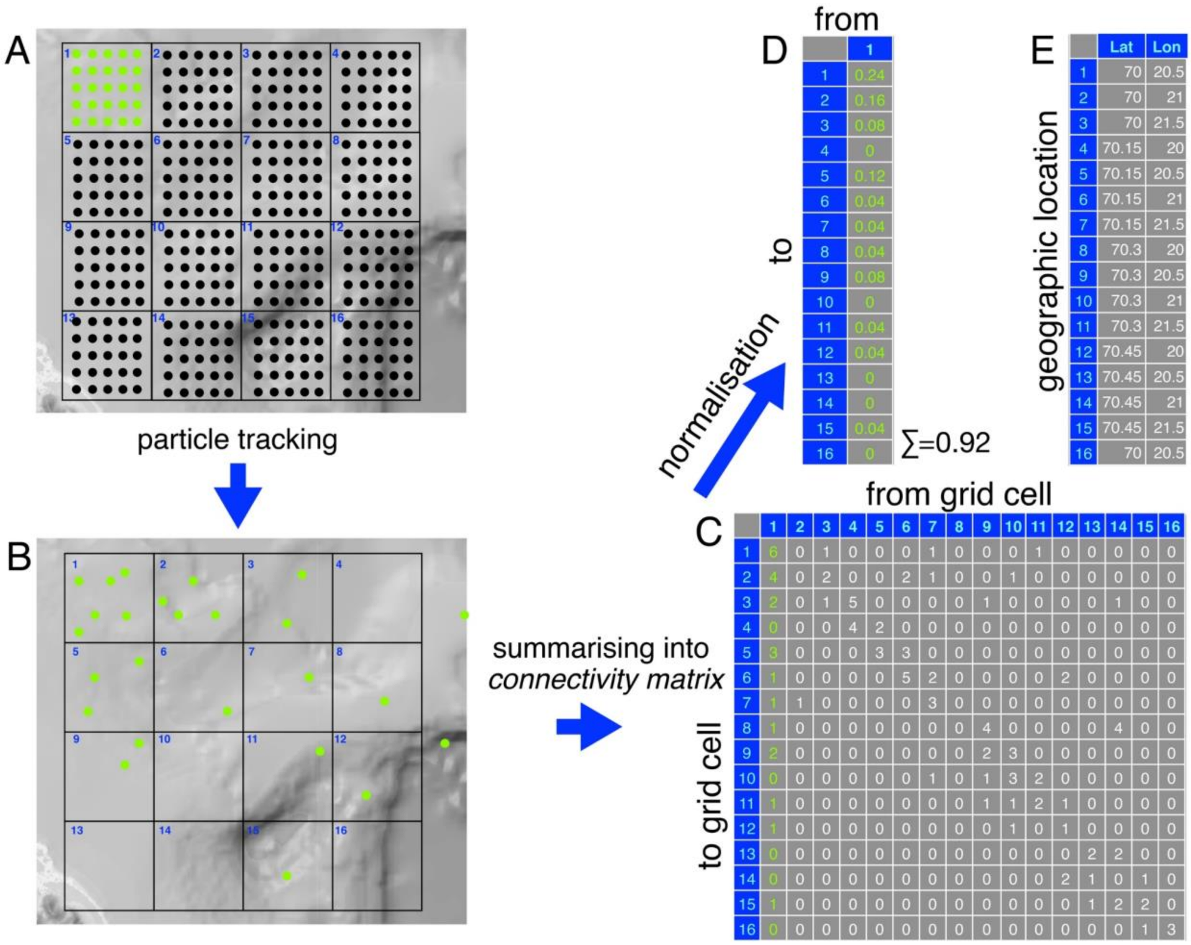


Fig. 1. The Arctic Sea with the biophysical model domain (depth ≤ 500 m) shown in blue. Also shown are the selected Marine Protected Areas: MPA 238 – Franz Josef Land, MPA 309 - Wrangel Island, MPA 377 – East Greenland, MPA 691 – Svalbard.

The dispersal trajectory data produced by the Lagrangian particle tracking model are extensive (ca. 10^{10} positions) and are summarized into mean connectivity matrices specifying dispersal probabilities between all 40893 model grid cells. Each connectivity matrix consists of 40893 rows and columns with each element specifying the probability to disperse from grid cell *j* (column *j*) to grid cell *i* (row *i*). Figure 2 schematically shows the construction of the connectivity matrix. For the connectivity matrices we averaged dispersal probability for each month and finally for a 10-year period. We further constructed separate connectivity matrices



C

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	6	0	1	0	0	0	1	0	0	0	1	0	0	0	0	0
2	4	0	2	0	0	2	1	0	0	1	0	0	0	0	0	0
3	2	0	1	5	0	0	0	0	1	0	0	0	0	1	0	0
4	0	0	0	4	2	0	0	0	0	0	0	0	0	0	0	0
5	3	0	0	0	3	3	0	0	0	0	0	0	0	0	0	0
6	1	0	0	0	0	5	2	0	0	0	0	2	0	0	0	0
7	1	1	0	0	0	0	3	0	0	0	0	0	0	0	0	0
8	1	0	0	0	0	0	0	0	4	0	0	0	0	4	0	0
9	2	0	0	0	0	0	0	0	2	3	0	0	0	0	0	0
10	0	0	0	0	0	0	1	0	1	3	2	0	0	0	0	0
11	1	0	0	0	0	0	0	0	1	1	2	1	0	0	0	0
12	1	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	2	2	0	0
14	0	0	0	0	0	0	0	0	0	0	0	2	1	0	1	0
15	1	0	0	0	0	0	0	0	0	0	0	0	1	2	2	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	3

D

	1
1	0.24
2	0.16
3	0.08
4	0
5	0.12
6	0.04
7	0.04
8	0.04
9	0.08
10	0
11	0.04
12	0.04
13	0
14	0
15	0.04
16	0

E

	Lat	Lon
1	70	20.5
2	70	21
3	70	21.5
4	70.15	20
5	70.15	20.5
6	70.15	21
7	70.15	21.5
8	70.3	20
9	70.3	20.5
10	70.3	21
11	70.3	21.5
12	70.45	20
13	70.45	20.5
14	70.45	21
15	70.45	21.5
16	70	20.5

Fig. 2. Schematic drawing showing the particle tracking procedure and the construction of the connectivity matrix. (A) Each grid cell in the model domain is seeded within some season and depth with a number of particles, in this example 25 particles. Note that each grid cell has an identification number, here from 1 to 16. (B) When the seeded particles have been transported by the ocean circulation for a specified PLD the positions are recorded. This panel shows that positions of the particles seeded in grid cell No 1 (the green circles). (C) The result is summarised into a connectivity matrix where each element indicates how many particles seeded in column j ended up in row i, where j and i are the identities of the source and sink grid cells, respectively. For the green particles seeded in grid cell 1 the results are filled into the rows of column 1, e.g. grid cell 9 received 2 particles from grid cell 1. (D) The numbers in the primary connectivity matrix are conveniently converted or normalised to probability by dividing by the number of particles seeded into each grid cell, in this case 25. Thus, the table in panel D shows the dispersal probability of dispersing from grid cell 1 to grid cells 1 to 16. (E) Also needed is the geographic location of each grid cell as latitude and longitude where each row number corresponds to the grid cell number.

for the warm season (March-October) and the cold season (November-February). This division into two seasons was based on the literature review of larval abundance (Fig. 5) and the ice cover (National Snow and Ice Data Centre 2019). Separate connectivity matrices were constructed for the drift depths 0, 10, 15, 30, 50, 70, 100 and 150 m and for the PLD 5, 10, 15, 30, 45, 60 and 90 days resulting in a total of 56 matrices for each season. In addition, 7 connectivity matrices per season were constructed for the surface drift based on the TOPAZ4 model. Each connectivity matrix may be viewed as the dispersal pattern for a specific combination of dispersal traits in terms of spawning season, drift depth and PLD. It is also possible to generate more complex dispersal strategies by combining several connectivity matrices (e.g. Jonsson et al. 2016). An example of a connectivity matrix produced within the project is shown in Fig. 3 where all elements greater than zero, i.e. connectivity between grid cells, is indicated in blue.

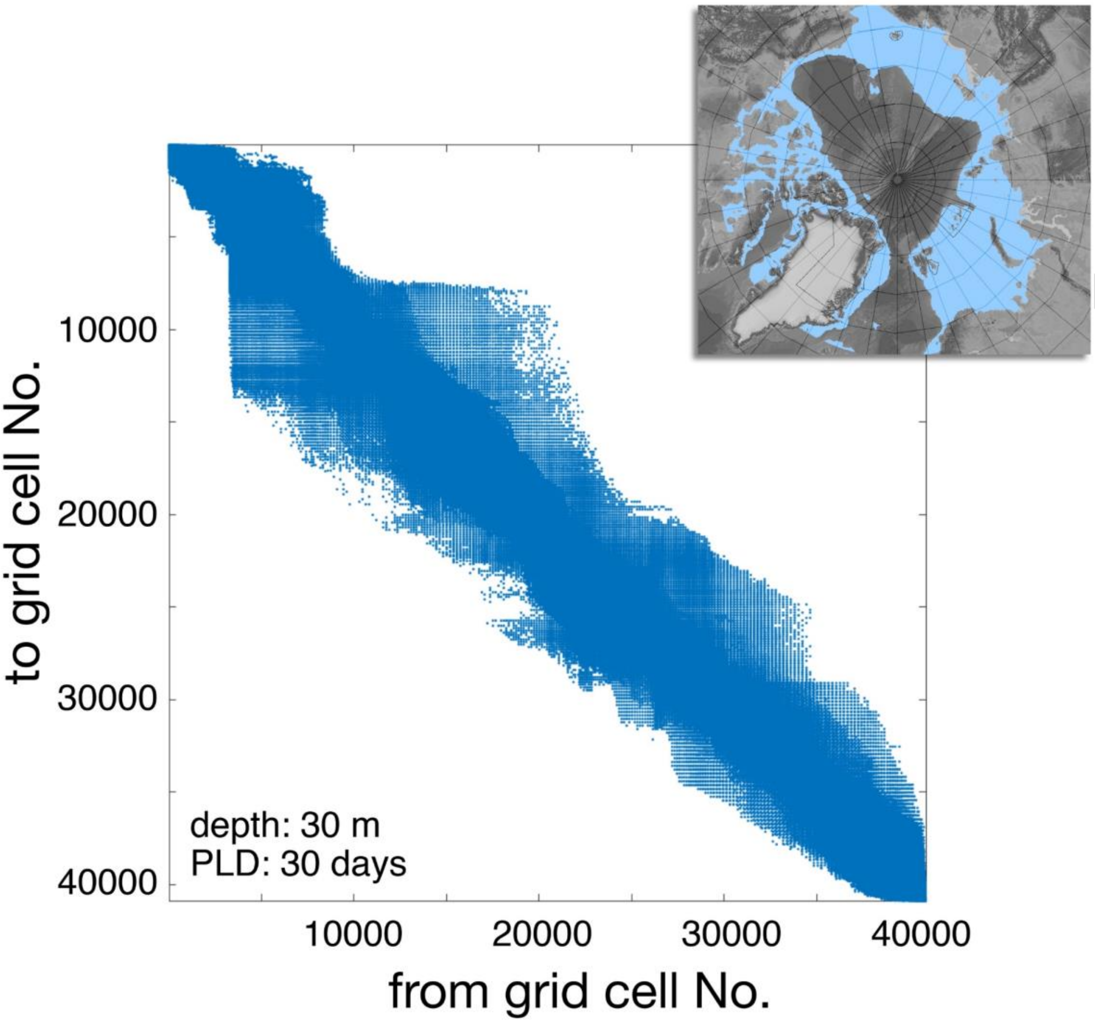


Fig. 3. A visualisation of a connectivity matrix for the whole model domain of 40893 grid cells (inset). The blue elements indicate all non-zero dispersal probabilities between grid cells. Note that the connectivity matrix is read as probability of dispersal from column j to row i. As seen the matrix is sparse and we only save non-zero data on disk to save space.

Calculation of dispersal distance

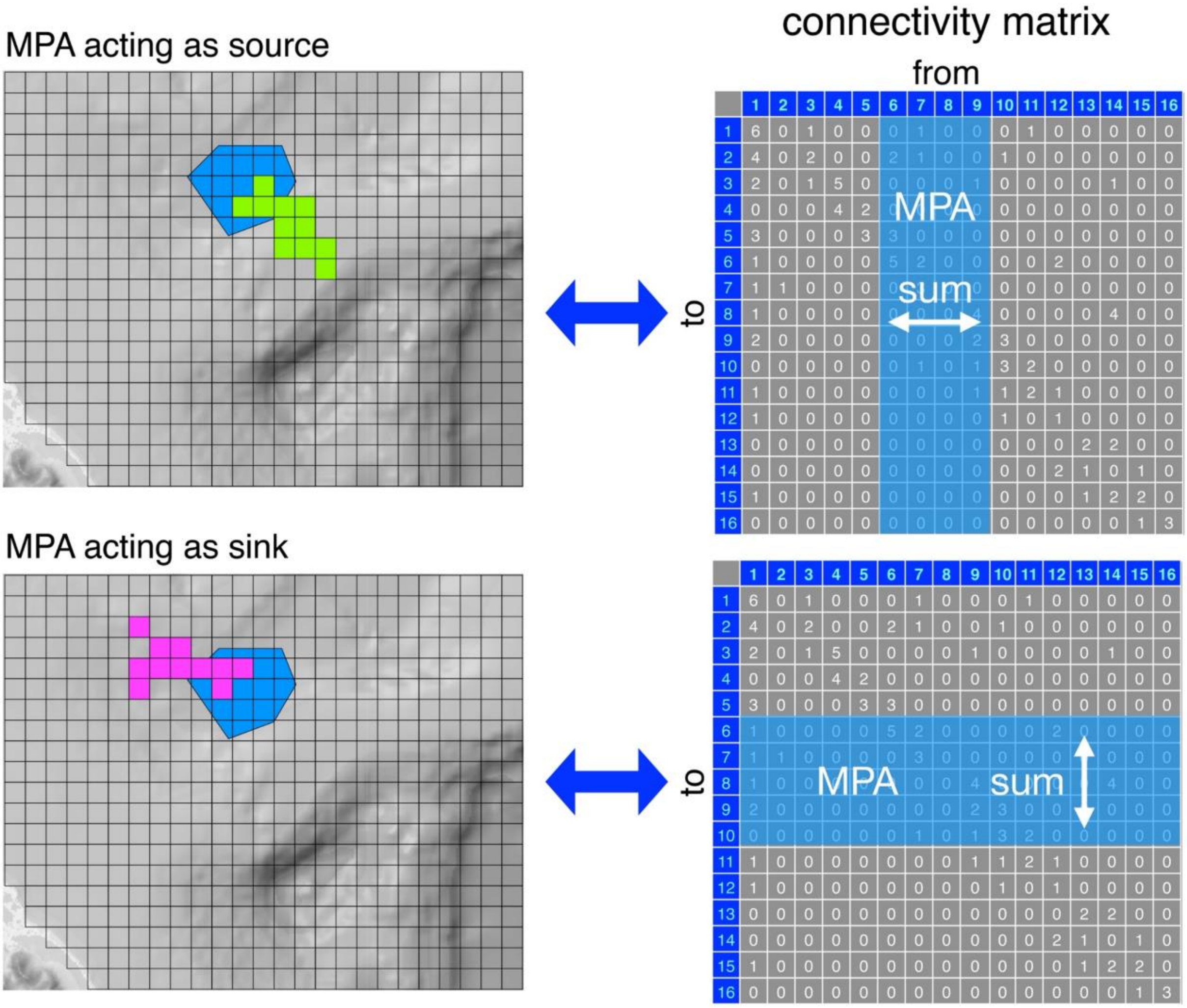
For each model grid cell that is included in the Lagrangian trajectory model (Fig. 1), the weighted mean of local dispersal distance was calculated based on the dispersal probability and the Euclidean distance between grid cells specified by the connectivity matrices. The weighted mean dispersal distance ($\bar{l}$) from each source grid cell *i* was estimated as:

$$\bar{l}_i = \sum_i^N C_{ij} \cdot D_{ij} \qquad \text{eq. 1}$$

where C_{ij} is a vector with connectivity from grid cell *i* to all other *N* grid cells, and D_{ij} is a vector of geographic distance from grid cell *i* to all other grid cells. This mean dispersal distance was calculated for a selection of combinations between drift depth and PLD (see results). Local dispersal distance can be important information when designing MPAs and considering the adequate size that can allow sufficient self-recruitment and thus ensure population persistence (Jonsson et al. 2020).

Identification of sources and sinks

From the connectivity matrix, it is also possible to identify the sources and sinks to and from a particular area, e.g. an MPA, a spawning ground or a nursery area. Sources may include a tracer of some pressure (e.g. contaminants or suspended matter) or biological propagules like seeds, eggs or larvae. The areas receiving propagules from an MPA acting as a source are found by summing columns representing locations where the MPA overlaps with the model grid cells (Fig. 4). Areas acting as sources for particular MPA are instead found by summing the rows representing locations where the MPA overlaps with the model grid cells. As an example, source and sink areas were identified for four selected areas (for illustration), with some marine extension, included in CAFF protected areas (CAFF 2013). The selected areas were ID258-Franz Josef Land, ID309-Wrangel Island, ID377-North-East Greenland, ID691-Svalbard. Source and sink analyses were carried out for combinations of drift depth and PLD. This analysis assumes that an abiotic tracer or biological propagules maintain their position within specific depth intervals during transport.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	6	0	1	0	0	0	1	0	0	0	1	0	0	0	0	0
2	4	0	2	0	0	2	1	0	0	1	0	0	0	0	0	0
3	2	0	1	5	0	0	0	0	1	0	0	0	0	1	0	0
4	0	0	0	4	2	0			0	0	0	0	0	0	0	0
5	3	0	0	0	3	3	0	0	0	0	0	0	0	0	0	0
6	1	0	0	0	0	5	2	0	0	0	0	2	0	0	0	0
7	1	1	0	0	0	0			0	0	0	0	0	0	0	0
8	1	0	0	0	0	0	0	0	4	0	0	0	0	4	0	0
9	2	0	0	0	0	0	0	0	2	3	0	0	0	0	0	0
10	0	0	0	0	0	0	1	0	1	3	2	0	0	0	0	0
11	1	0	0	0	0	0	0	0	1	1	2	1	0	0	0	0
12	1	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	2	2	0	0
14	0	0	0	0	0	0	0	0	0	0	0	2	1	0	1	0
15	1	0	0	0	0	0	0	0	0	0	0	0	1	2	2	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	3

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	6	0	1	0	0	0	1	0	0	0	1	0	0	0	0	0
2	4	0	2	0	0	2	1	0	0	1	0	0	0	0	0	0
3	2	0	1	5	0	0	0	0	1	0	0	0	0	1	0	0
4	0	0	0	4	2	0	0	0	0	0	0	0	0	0	0	0
5	3	0	0	0	3	3	0	0	0	0	0	0	0	0	0	0
6	1	0	0	0	0	5	2	0	0	0	0	2	0	0	0	0
7	1	1	0	0	0	0	3	0	0	0	0	0		0	0	0
8	1	0	0				0	0	4					4	0	0
9	2	0	0	0	0	0	0	0	2	3	0	0		0	0	0
10	0	0	0	0	0	0	1	0	1	3	2	0	0	0	0	0
11	1	0	0	0	0	0	0	0	1	1	2	1	0	0	0	0
12	1	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	2	2	0	0
14	0	0	0	0	0	0	0	0	0	0	0	2	1	0	1	0
15	1	0	0	0	0	0	0	0	0	0	0	0	1	2	2	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	3

Fig. 4. Schematic drawing illustrating the identification of areas receiving propagules from an MPA (blue polygon), and areas providing propagules to the MPA. In the top left panel the MPA acts as a source to ambient areas (green grid cells). The green areas are found by first identifying the grid cell numbers matching the location of the MPA. The columns representing these grid cells are summed row-wise (blue columns) and normalised to the number of columns. The resulting non-zero elements represent the areas (green grid cells) receiving propagules from the MPA. Symmetrically, the pink areas providing propagules to the MPA in the bottom-left panel can be found by identifying the row numbers corresponding to the grid cells matching the MPA location (blue rows). Summing these rows column-wise produces a vector where the non-zero elements represent the pink areas acting as a source to the MPA.

Calculation of local retention for MPAs

For existing or planned MPAs, the connectivity matrix can be directly used to calculate the predicted local retention of larvae, i.e. the probability that larvae released within the MPA also end their trajectory within the same MPA. The elements in the connectivity matrix overlapping with each MPA were extracted and summed to obtain the estimated mean local retention ($\bar{r}$) for each MPA as:

$$\bar{r} = \sum_{1}^{n \cdot n} C_{MPA,MPA} \cdot 1/n \qquad \text{eq. 2}$$

where $C_{\mathrm{MPA,\ MPA}}$ is a sub-matrix with connectivity for the *n* grid cells located within the MPA. Local retention is a necessary condition for self-recruitment, although not the only.

Identification of dispersal barriers

Bathymetric features, habitat distribution and consistent circulation patterns may lead to dispersal barriers in the seascape with consequences for exchange of individuals and genes between sub-populations. A previously developed clustering method was employed to identify partial dispersal barriers from the constructed connectivity matrices (Nilsson Jacobi et al. 2012, Jonsson et al. 2020). This theoretical framework finds clusters as a signature of partially isolated sub-populations. Identification of subpopulations is formulated as a minimization problem with a tuneable penalty term that makes it possible to generate population subdivisions with varying degree of dispersal restrictions. Areas that have an internal connectivity above the dispersal restriction are colour-coded, and the transitions of colours thus indicate partial dispersal barriers. Barriers may differ among dispersal strategies and habitats. Strong barriers may indicate the presence of locally adapted sub-populations with unique genetic compositions. Also relatively weak barriers may indicate limited exchange of individuals, which may call for separate management plans for harvested populations.

V. Results

Review of published information on larval dispersal traits in the Arctic Ocean

The most commonly reported data is the occurrence of larvae over time in the water column, often with only broad taxonomic identification. It was also possible in many cases to infer the depth interval where larvae were caught, although only a few studies presented depth profiles. Few studies reported data on PLD, although PLD could in some cases be coarsely inferred from the occurrence over time in the water column.

Table A1 (see Appendix) shows extracted data from 10 publications. A coarse overview of the temporal occurrence of larvae in the water column is shown in Fig. 5 expressed as number of taxa found at different time periods. The blue line indicates the time of the year when larvae are first observed and the orange line represents when they disappear. It is clear that the majority of larvae disperse during the warmer half of the year with peak abundances in May to August. This graph guided the subdivision of connectivity matrices into the two seasons "warm" (March-October) and "cold" (November-February).

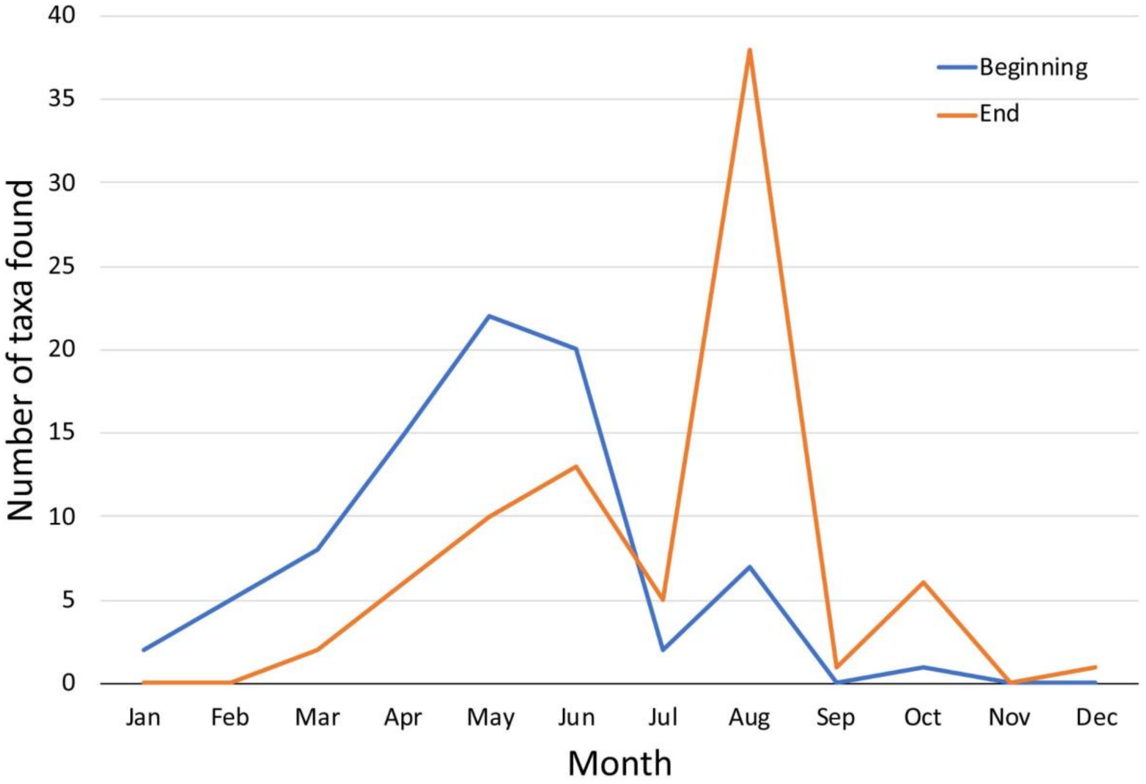


Fig. 5. Data on seasonal occurrence of taxa with planktonic larvae in the Arctic Sea extracted from published papers (see Table A1). The two curves indicate the beginning (blue) and end (orange) of occurrence.

The differences in mean time between the beginning and the end of occurrence give a very approximate estimate of the mean PLD, which is calculated to 63 day in this report. Vestfal et al. (2019) reported a PLD for polar and saffron cods of 60-90 days but that can be preceded by an egg stage of 35-80 days. For the bivalves *Hiatella arctica* and *Mya truncata* PLDs of 60-100 and 90 days are reported, respectively (Brandner et al. 2017).

Information about the sample depth was mainly reported rather than vertical profiles of larval abundance. For the few studies that sampled a depth profile, larvae were concentrated in the upper 50 m (e.g. Kuklinski et al. 2013). No study sampled deeper than 200 m.

Based on the very limited input from the literature, it is only possible to use reported data on larval traits as a rough guideline of the bounds of included parameters in the biophysical model. It is also likely that PLD shows interannual variation because of water temperature. The strategy in this work was thus to include a broad range of parameter values for season, drift depth and PLD. This allows the construction of an extensive database of modelled connectivity for many possible parameter combinations, giving flexibility when more empirical data on larval traits become available.

Validation of the oceanographic circulation models

We do not provide any validation of the ocean models within this study. The TOPAZ system is an operational model run at the Norwegian Meteorological Institute, and it is an official part of the European Union (marine) Copernicus web portal. Notably, it also contains a data assimilation scheme taking advantage of available observations. The model is used in both predictive mode (using real time observations) and in a re-analysis mode (using all available data). The model undergoes a continuous validation, for this report the validation of drift buoys is perhaps the most relevant (https://cmems.met.no/ARC-MFC/V2Validation/buoyDrift/index.html). Temperature and salinity are validated here (https://cmems.met.no/ARC-MFC/V2Validation/TSprofile/index.html). Thus, we consider TOPAZ to be well validated and it would require a significant effort to “improve” validation relevant for this study. The Arctic4 model does not undergo a similar validation process. The main validation we consider in this study is comparing the Arctic4 toward the TOPAZ, which is well validated. We do get similar results for mean currents and its variability, also the results from connectivity matrix analysis does not show significant deviations in strength and direction of dispersion of particles (Fig. 6). Thus, the similarity in the outputs from both models provide a validation for model results.

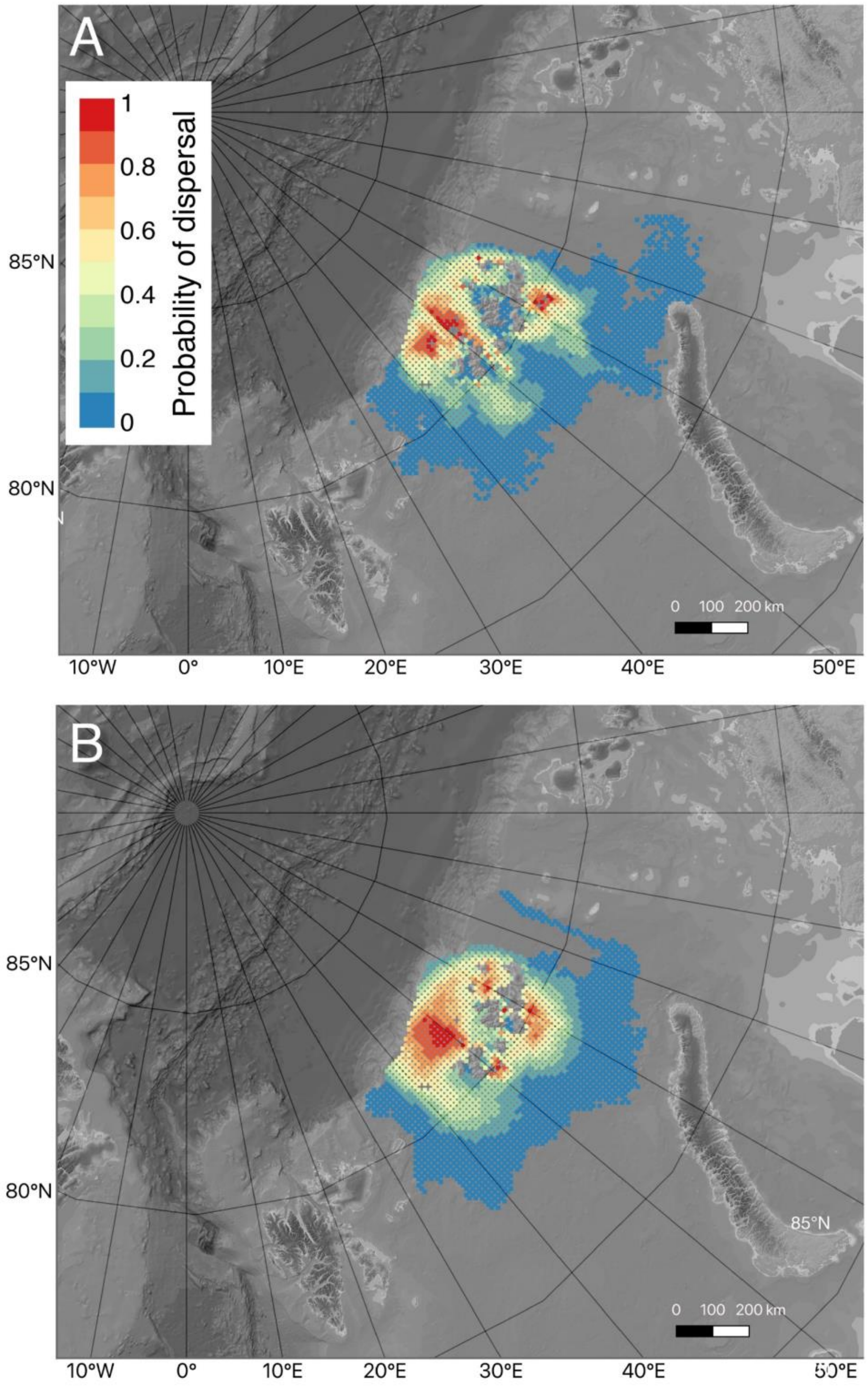


Fig. 6. A comparison of connectivity matrices produced by the two models (A) Arctic4 and (B) TOPAZ. Here particles were seeded within the MPA 238 (Franz Josef Land) shown by the dotted polygon. Probability of dispersal in the surface water with a PLD of 45 days is shown as the colour-coded fields.

Local dispersal distance as a function of biological traits

Modelled dispersal and effects of biological traits is best visualised as dispersal distance from each release point in the seascape. Figure 7 shows dispersal distance for each of the 40893 grid cells for a drift depth of 10 m and for increasing days of PLD. As expected, dispersal distance increases with PLD but the distance is also highly dependent on the location in the seascape. In many areas, dispersal distance is less than 100 km but extends to several 100 km along the east Greenland slopes, north of Novaya Zemlya and north of Alaska. Dispersal distance tends to decrease with increasing depth (Fig. 8), e.g. in the Kara Sea. However, in most areas the depth interval explored in the present study (0-150 m) had a modest effect of modelled dispersal distance. This report only presents maps of dispersal distance for a subset of the available connectivity matrices but these calculations can easily be extended.

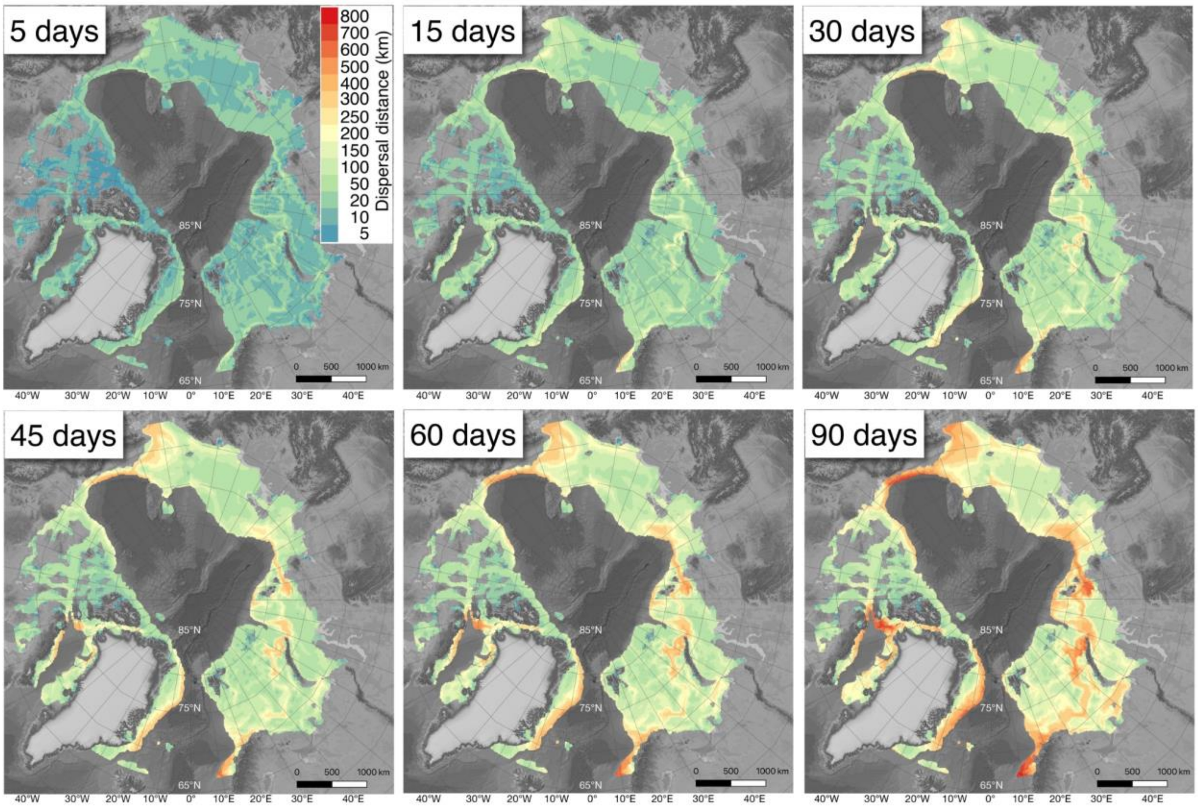


Fig. 7. Maps showing the mean dispersal distance for each of 40893 grid cells during the warm season (March-October) for modelled larvae drifting at 10 m depth for PLDs of 5, 15, 30, 45, 60 and 90 days. Results are means over 10 years (2007 – 2016). Results are from Arctic4 model.

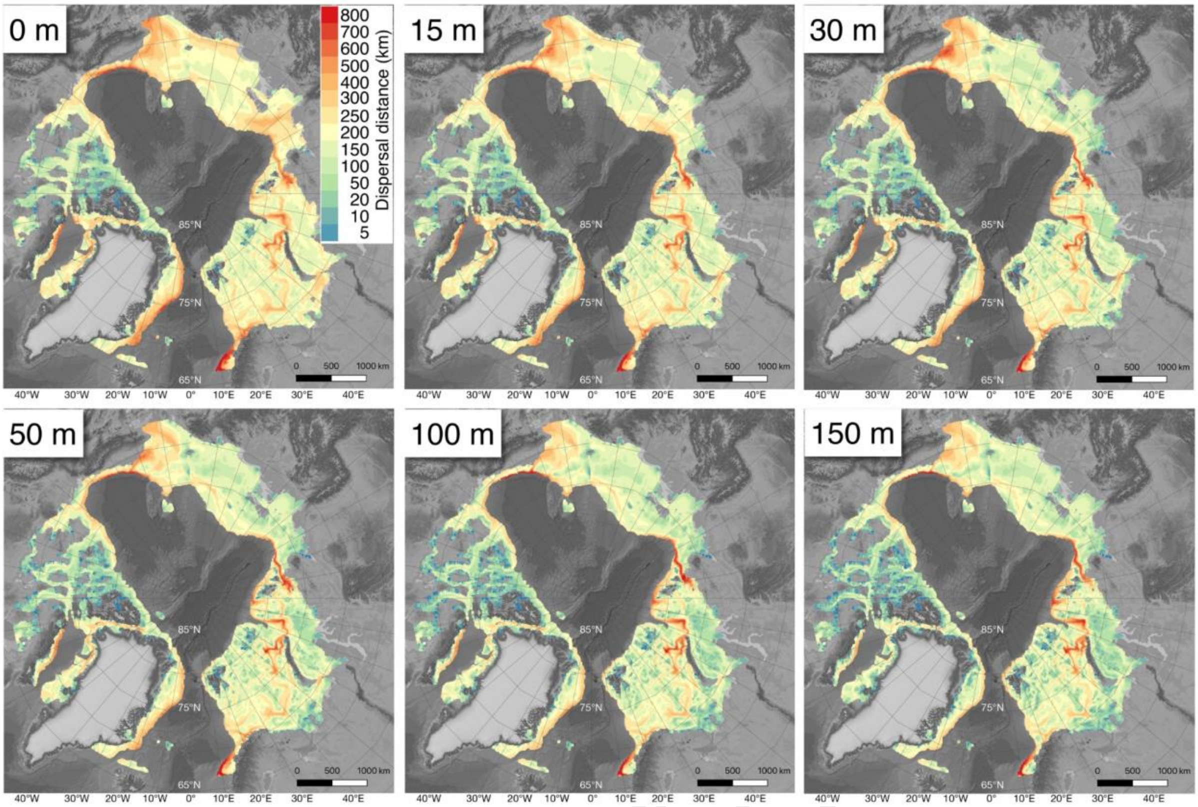


Fig. 8 Maps showing the mean dispersal distance for each of 40893 grid cells during the warm season (March-October) for modelled larvae drifting at 6 different depths (0, 15, 30, 50, 100 and 150 m) for a PLD of 30 days. Results are means over 10 years (2007 – 2016). Results are from Arctic4 model.

Identification of sources and sinks – a brief demonstration

In many cases mores specific information about seascape connectivity than dispersal distance is needed. One example is to estimate source-sink relationships between areas. This can apply, e.g. to how MPAs may function as sources of recruits to ambient areas, the risk of MPAs acting as sinks for contaminants from external sources, or how fish spawning grounds connect via larval dispersal to possible nursery areas. Figures 9 and 10 show a few examples of source-sink analyses for a protected area around Franz Josef Land (CAFF 2013), mainly as an example of how connectivity matrices can be used to identify areas around MPAs that either receive larvae or other propagules from organisms spawning with an MPA, or to identify ambient areas that provide an MPA with recruits. Different MPAs may show very different source patterns as illustrated in Fig. 11 for 4 selected areas.

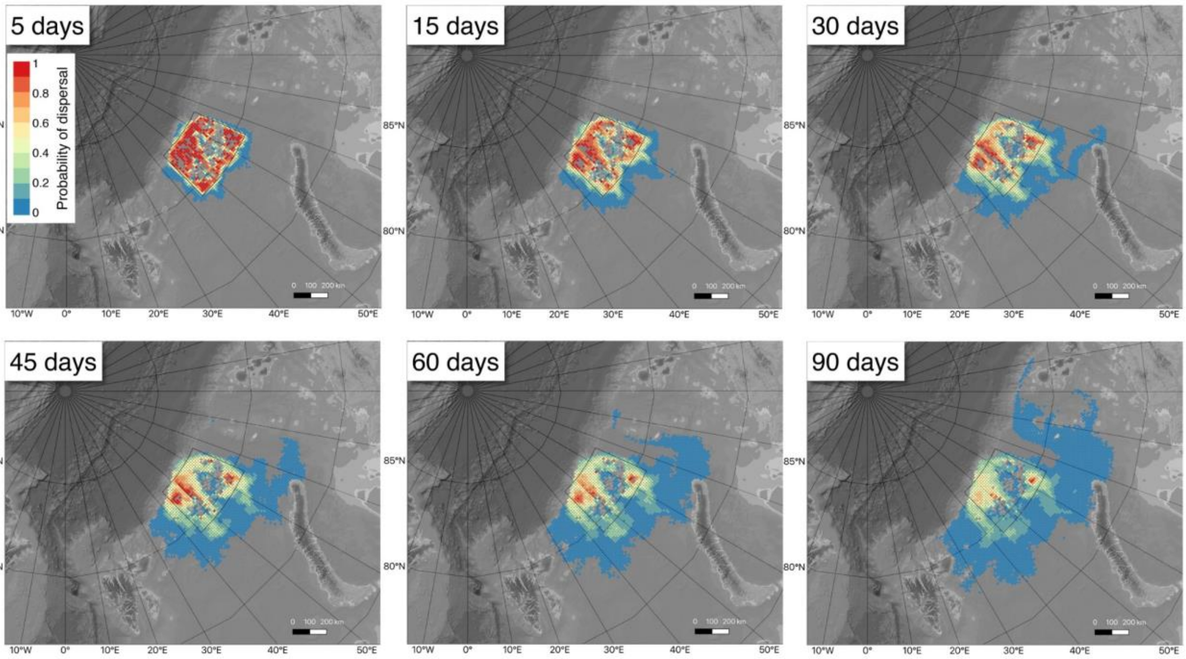


Fig. 9. Areas receiving propagules from the MPA 238 – Frans Josef Land (dotted area), i.e. the MPA acting as a source. The six panels show the effect of increasing PLD. The colour scale (from blue to red) indicates relative probability (0-1) of propagules originating from the MPA.

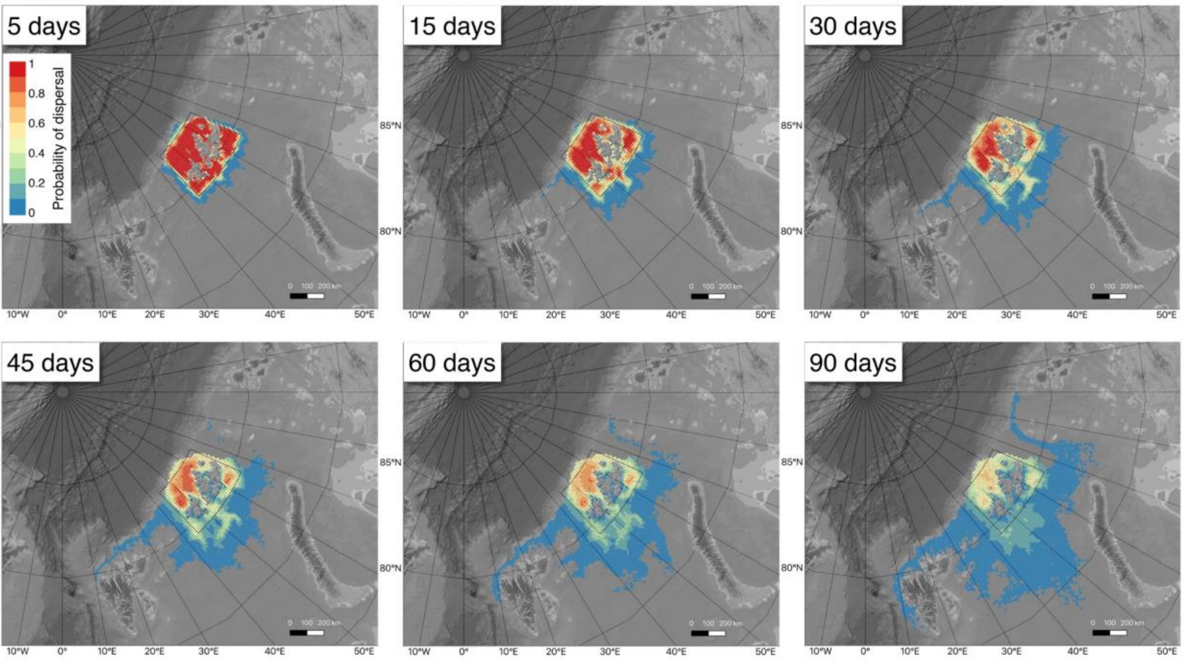


Fig. 10. Areas providing propagules to the MPA 238 – Frans Josef Land (dotted area), i.e. the MPA acting as a sink. The six panels show the effect of increasing PLD. The colour scale (from blue to red) indicates relative probability (0 to 1) of propagules ending up in the MPA.

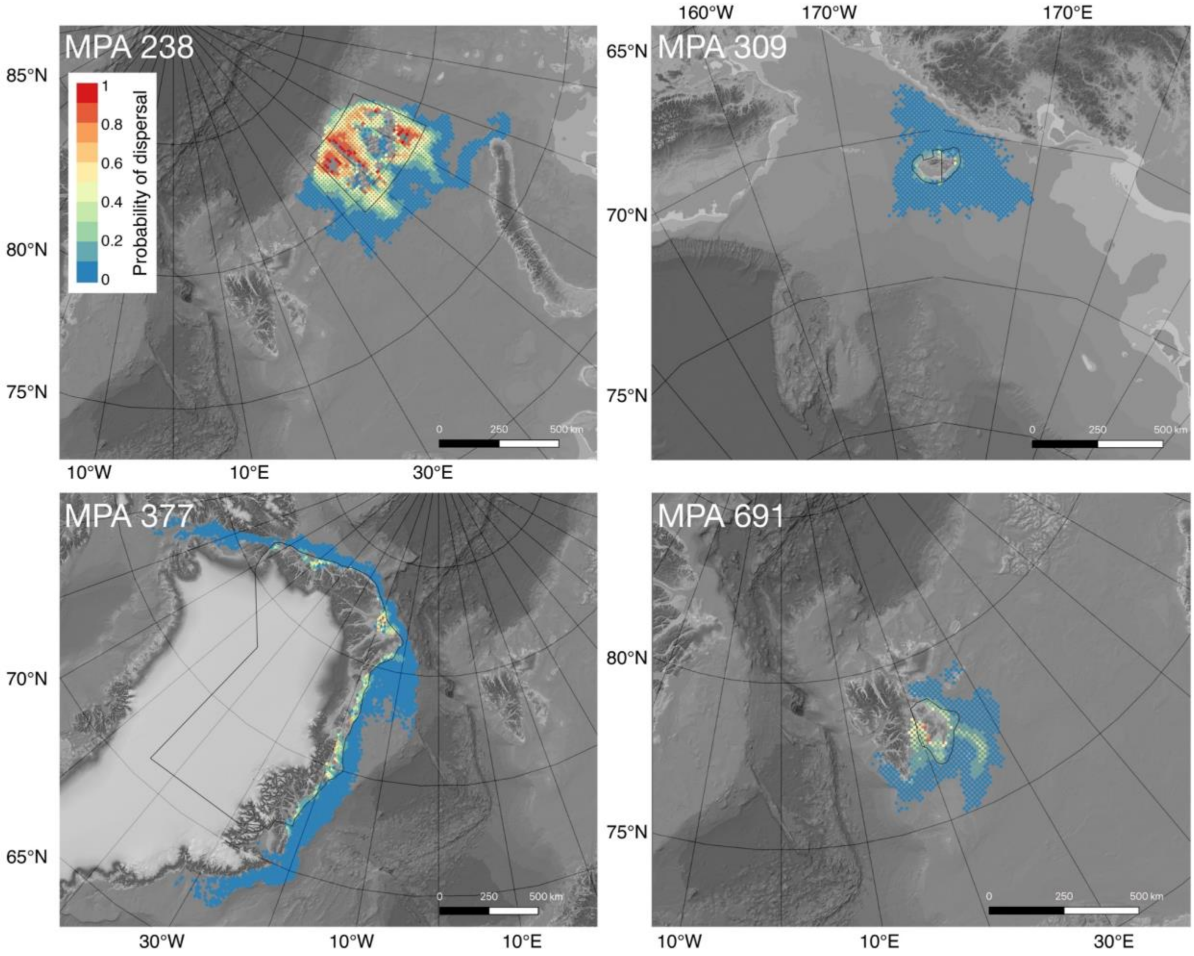


Fig. 11. Areas receiving propagules from 4 selected MPAs (dotted areas, see Fig. 1 for overview). The drift depth was 10 m and the PLD 30 days. The colour scale (from blue to red) indicates relative probability (0 to 1) of propagules originating from the MPA.

Examples of local retention with implications for self-recruitment and persistence

A critical aspect in the design of MPAs is the potential for self-recruitment within the designated MPA. For sufficient self-recruitment, a significant portion of released larvae need to be locally retained within the MPA or return to the MPA at the time for settlement. The local retention of released larvae may easily be calculated from the connectivity matrix as the diagonal elements for those grid cells overlapping with the MPA (eq. 2). It is not well understood what level of self-recruitment that is necessary to maintain persistent populations within an MPA but for many fish species, a level of 40% has been suggested (Kaplan et al. 2006) although this could differ between species and areas. Figure 12 shows local retention for the 4 selected MPAs and for combinations of drift depth and PLD. Reflecting the variation in dispersal distance (Figs. 7-8) local retention within MPAs decreases with PLD but is marginally affected by drift depth. There is also an effect of MPA size and shape where the relatively small MPA309 shows low levels of local retention.

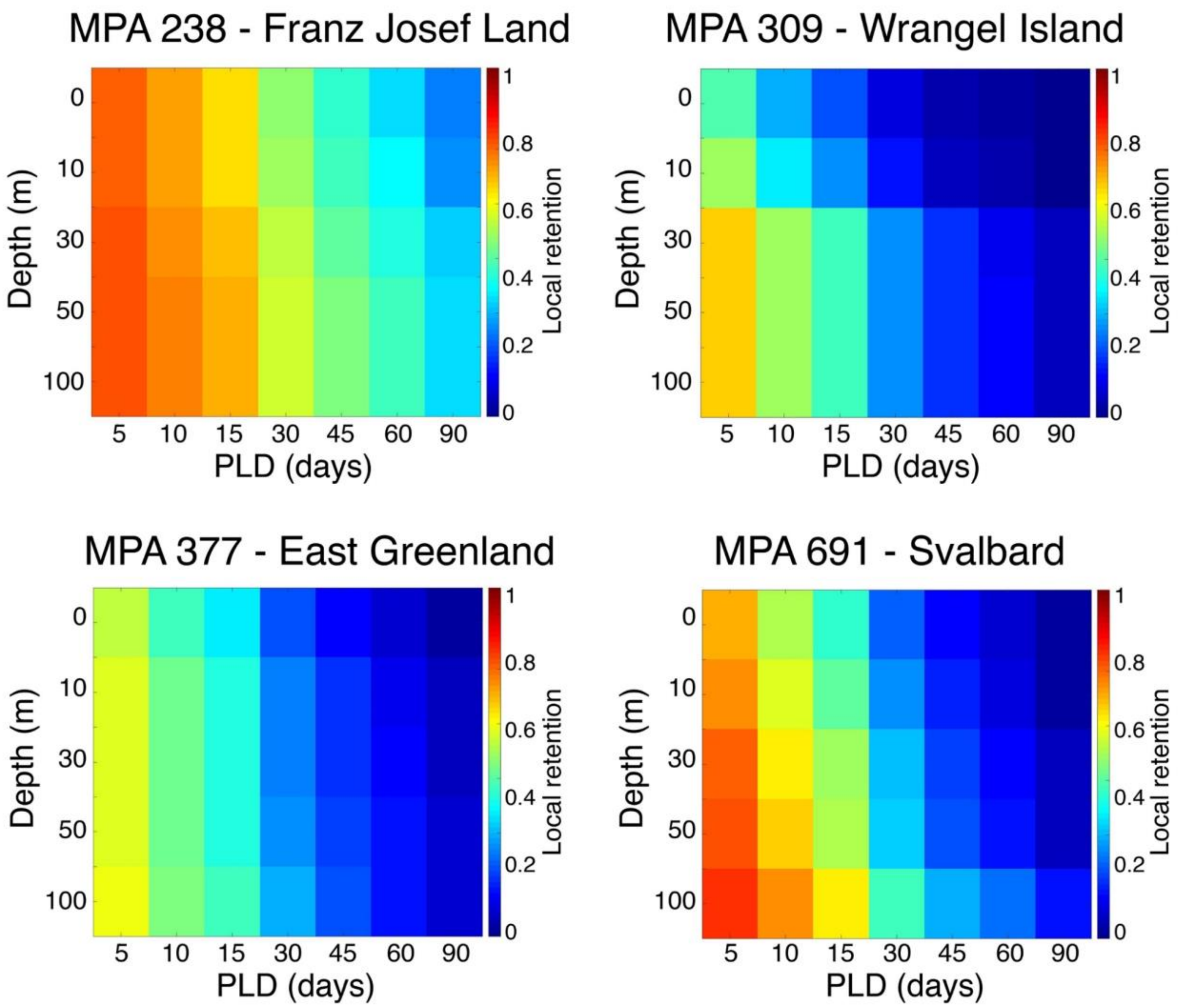


Fig. 12. Graphs showing the local retention (1 is complete retention) in 4 selected MPAs for larval dispersal at 5 depths and for 7 PLDs.

Identification of dispersal barriers

We used a cluster method (Nilsson Jacobi et al. 2012) applied to the connectivity matrices to identify potential dispersal barriers in the domain modelled in the Arctic Sea. This method requires that the user specifies a threshold of minimum "leakage" of propagules between partially isolated areas that define a dispersal barrier. Typically, the connectivity within areas is set to more than 100 times greater than connectivity across barriers. However, this level may be varied depending on the biological process of interest. The potential of genetic differentiation may require strong barriers to gene flow, while demographic independence, e.g. of fish stocks, may develop even when substantial dispersal occurs across barriers. In Fig. 13, four examples are shown of the mapping of connectivity clusters separated by partial dispersal barriers. With increasing threshold of allowed dispersal across barriers, the domain is partitioned to more sub-populations. Not surprisingly, by increasing the PLD from 30 to 90 days, the areas of high internal connectivity increase in size. The open shelf areas show few barriers with large areas that have high internal connectivity, e.g. the Kara Sea. The more complex geomorphology in the Arctic Archipelago also leads to more dispersal barriers.

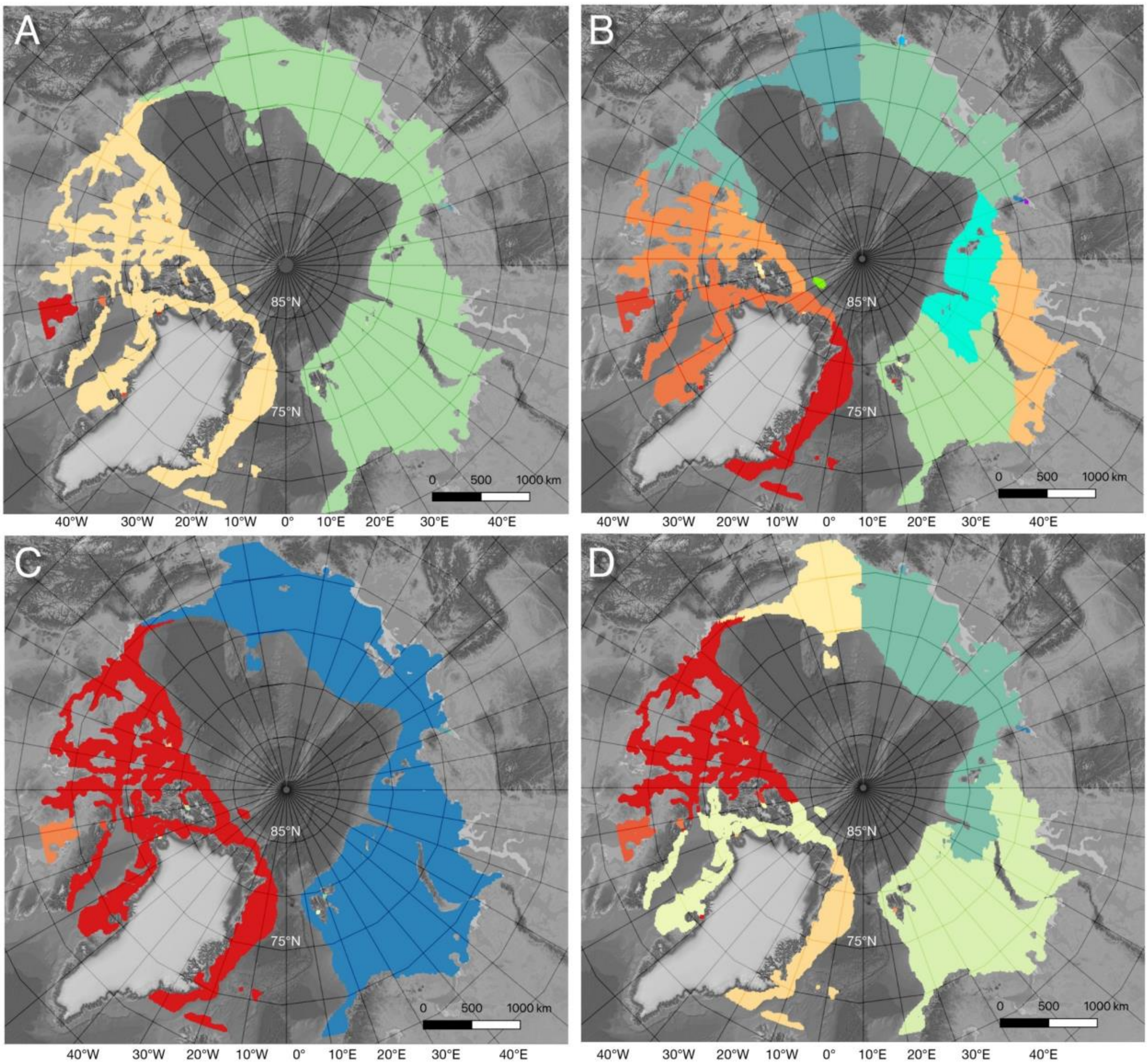


Fig. 13. Identification of dispersal barriers on the Arctic Sea shelf based on the biophysical model of larval dispersal for some dispersal trait combinations. The top panels show barriers for dispersal at 10 m depth with a PLD of 30 days and with a threshold of dispersal across barriers of (A) 0.0007 and (B) 0.01. The bottom panels show barriers for dispersal at 10 m depth with a PLD of 90 days and with a threshold of dispersal across barriers of (C) 0.0009 and (D) 0.01 Colour transitions between areas indicate partial dispersal barriers. Colours are arbitrarily chosen.

VI. Conclusions and future perspectives

In this project we used a biophysical model to estimate seascape connectivity within the continental shelf of the Arctic Sea with a focus on organisms dispersing with free-drifting larvae. The set of tools described here to assess seascape connectivity are based on biophysical modelling of larval dispersal and is mainly relevant for organisms with sedentary adults where connectivity largely depends on physical water transport of larvae or other propagules. A large number of dispersal trajectories were simulated including many combinations of spawning season, drift depth and Pelagic Larval duration (PLD). We used two physical models (Arctic4 and TOPAZ), which give similar results for the connectivity between grid cells in the model domain. The modelling results are primarily stored as connectivity matrices, which comprise a database of dispersal probability between 40893 selected grid cells in the model. In this report we have demonstrated how these connectivity matrices can be applied into tools for analyses and design of MPAs. Increasingly, MPAs are designed as ecologically coherent networks (HELCOM 2016) and the toolbox demonstrated here can be used to assess relevant size and shape of individual MPAs (adequacy) and how they act as a network based on connectivity, both within the network and with non-protected habitat (Jonsson et al. 2020).

Dispersal distance is more sensitive to increasing PLD and less influenced of the drift depth. The preliminary analysis indicates that expected dispersal distance in the Arctic Sea may be greater for a given PLD than in some previously modelled coastal areas in the North Sea (Jonsson et al. 2016) and in the Baltic Sea (Jonsson et al. 2020). Generally, the many open coastal areas may explain this more extensive dispersal. In some areas where the coast is complex, e.g. the Arctic Archipelago, dispersal distance is more modest. The generally large dispersal distance per day and also the expected long PLDs in cold waters (e.g. Vestfal et al. 2019) imply that MPAs need to be relatively large to ensure a high level of local retention and self-recruitment.

The connectivity matrix further gives a flexible possibility to identify sources and sinks to a specified area, e.g. MPAs. Sources may be areas that supply an MPA with recruits, but also areas causing environmental impact on protected areas, e.g. non-indigenous species, discharge of contaminants or oil spill. The latter can be important as less ice can open the Northwest Passage for shipping, also new gas and oil reserves can be found and exploited. The main message is that also activities in a considerable area outside an MPA may affect the conditions within the boundaries of MPAs. However, the shape of this external area will depend on the local circulation pattern. The source area will be a function of residence time in the water, the depth of dispersal and seasonal variation in circulation. In a similar way, it is easy to extract from the connectivity matrix the probability that fertilized eggs, larvae or other propagules released within an MPA also settle within that MPA. High local retention may lead to a largely self-recruited, closed local population, which may persist without immigration from other protected or unprotected local populations.

Connectivity patterns in the seascape, e.g. specified by a connectivity matrix, may reveal areas with high internal connectivity with partial dispersal barriers to other such areas. Barriers may indicate demographically independent local populations (stocks) or genetically differentiated populations with local adaptations if barriers are sufficiently strong (Allendorf et al. 2013). Based on the connectivity matrix, well-connected clusters can be identified in a way to minimize the total dispersal (leakage) among such groups, subjected to some penalty of aggregating groups (Nilsson Jacobi et al. 2012). This tool aids to visualize the structure of the connectivity matrix projected onto a geographic map. Colour-coded areas indicate management units (Palsbøll et al. 2007) separated by dispersal barriers. Dispersal barriers are generally partial and the number of dispersal barriers decrease as less dispersal is allowed across barriers, which is specified by the user. Genetically differentiated local populations are expected to be associated with fewer but stronger barriers (Jahnke et al. 2018), while a larger number of more "leaky" barriers may represent the distribution of demographically independent stocks. Such local populations or stocks may require separate conservation and management actions, and siting of MPAs may be stratified across such management units.

As a future perspective, it will now be possible to apply the modelled connectivity matrices in network analyses of complete MPA networks using a metapopulation perspective (Jonsson et al. 2020). We have previously developed a framework based on Eigenvalue Perturbation Theory (EPT, Ovaskainen & Hanski 2003, Nilsson Jacobi & Jonsson 2011). Connectivity is here directly linked to metapopulation dynamics to identify optimal MPA networks. By applying EPT to the connectivity matrix it is possible to find the best network of MPAs that maximizes the growth rate of the global metapopulation (protected and unprotected areas) when the metapopulation is small, which is typical of threatened species. One advantage of this approach is that there is a unique network of MPAs for each connectivity matrix and the total protected area, and that this network is directly linked to persistence of the whole metapopulation. If habitat information is available, e.g. presence-absence or habitat quality (Berglund et al. 2012, Virtanen et al. 2018), and it is also possible to identify consensus MPA networks targeting multiple species with different dispersal strategies (Jonsson et al. 2016).

The database of connectivity matrices produced in this project together with the demonstration of suitable analytical methods will be part of PAME's MPA toolbox.

VII. Data availability

The primary data produced within this project consist of a set of connectivity matrices specifying the probability of dispersal within the studied seascape for combinations of year, month, drift depth and pelagic larval duration (drift duration in days). Each connectivity matrix has a data structure of 40893 rows and columns and is saved as sparse matrices in Matlab .mat binary format. In total there are 9960 connectivity matrices based on the Arctic4 model and 2016 matrices based on the TOPAZ model. Each connectivity matrix requires around 1 MB of digital memory. There is also a set of 126 aggregated connectivity matrices which are averaged over all years and divided into two seasons (March-October and November-February). Aggregated connectivity matrices based on Arctic4 are available for the depths 0, 10, 15, 30, 50, 70, 100 and 150 m, and for the PLDs 5, 10, 15, 30, 45, 60 and 90 days. For TOPAZ there are aggregated connectivity matrices for 0 m depth, and for the PLDs 5, 10, 15, 30, 45, 60 and 90 days. The aggregated connectivity matrices each requires between 10-50 MB of digital memory.

In addition to the primary data in the form of connectivity matrices there are derived datasets based on the connectivity matrices. First there are results from the barrier analysis assigning each of the 40893 locations included in the model domain to a cluster of high within-cluster connectivity. This analysis was performed for a large set of aggregated connectivity matrices and for a range of parameters representing the level of allowed dispersal between clusters. In total there are 2149 files (.txt) representing combinations of drift depth, PLD and allowed dispersal between clusters. Second, there are visualisations of data in the form of maps, many of them included in this report. The maps are available as shape files.

The intention is that all or selected data could be open access given an infrastructure for storage and downloading can be arranged.

VIII. Meetings and contacts with reference group

To be added

IX. References

Allendorf, F. W., G. H. Luikart, and S. N. Aitken. 2013. Conservation and the genetics of populations. John Wiley & Sons.

Berglund, M., M. Nilsson Jacobi, and P. R. Jonsson. 2012. Optimal selection of marine protected areas based on connectivity and habitat quality. Ecological Modelling 240:105-112.

Brandner, M. M., E. Stübner, A. J. Reed, T. M. Gabrielsen, and S. Thatje. 2017. Seasonality of bivalve larvae within a high Arctic fjord. Polar Biology 40:263-276.

CAFF. 2013. Arctic Biodiversity Assessment. Status and trends in Arctic biodiversity. Akureyri. 557 pp.

Corell, H., P. O. Moksnes, A. Engqvist, K. Döös, and P. R. Jonsson. 2012. Depth distribution of larvae critically affects their dispersal and the efficiency of marine protected areas. Marine Ecology Progress Series 467:29-46.

Cowen, R. K., and S. Sponaugle. 2009. Larval dispersal and marine connectivity. Annual Review of Marine Science 1:443-466.

HELCOM. 2016. Ecological coherence assessment of the Marine Protected Area network in the Baltic Sea. Baltic Sea Environment Proceedings 148:74 pp.

Jahnke, M., Jonsson, P.R., Moksnes, P.O., Loo, L.O., Nilsson Jacobi, M. & Olsen, J.L. 2018. Seascape genetics and biophysical connectivity modelling support conservation of the seagrass *Zostera marina* in the eastern North Sea. Evolutionary Applications*,* 11**:** 645-661.

Jonsson, P. R., P. O. Moksnes, H. Corell, E. Bonsdorff, and M. Nilsson Jacobi. 2020. Ecological coherence of Marine Protected Areas: New tools applied to the Baltic Sea network. Aquatic Conservation: Marine and Freshwater Ecosystems 30:743-760.

Jonsson, P. R., M. Nilsson Jacobi, and P. O. Moksnes. 2016. How to select networks of marine protected areas for multiple species with different dispersal strategies. Diversity and Distributions 22:161-173.

Jorde, P. E., G. Sovik, J. I. Westgaard, J. Albretsen, C. Andre, C. Hvingel, T. Johansen, A. D. Sandvik, M. Kingsley, and K. E. Jorstad. 2015. Genetically distinct populations of northern shrimp, Pandalus borealis, in the North Atlantic: adaptation to different temperatures as an isolation factor. Molecular Ecology 24:1742-1757.

Kaplan, D. M., and J. Largier. 2006. HF radar-derived origin and destination of surface waters off Bodega Bay, California. Deep-Sea Research II 53:2906-2930.

Kuklinski , P., J. Berge, L. McFadden, K. Dmoch, M. Zajaczkowski, H. Nygåd, K. Piwosz, and A. Tatarek. 2013. Seasonality of occurrence and recruitment of Arctic marine benthic invertebrate larvae in relation to environmental variables. Polar Biology 36:549-560.

Lester, S. E., and B. S. Halpern. 2008. Biological responses in marine no-take reserves versus partially protected areas. Marine Ecology Progress Series 367:49-56.

Micheli, F., A. Saenz-Arroyo, A. Greenley, L. Vazquez, J. A. Espinoza Montes, M. Rossetto, and G. A. De Leo. 2012. Evidence that marine reserves enhance resilience to climatic impacts. PLoS ONE 7:e40832.

National Snow and Ice Data Center. 2019. Arctic sea and ice news & analysis. https://nsidc.org/arcticseaicenews/charctic-interactive-sea-ice-graph/, accessed August 21, 2020.

Nilsson Jacobi, M., C. André, K. Döös, and P. R. Jonsson. 2012. Identification of subpopulations from connectivity matrices. Ecography 35:1004-1016.

Nilsson Jacobi, M., and P. R. Jonsson. 2011. Optimal networks of nature reserves can be found through eigenvalue perturbation theory of the connectivity matrix. Ecological Applications 21:1861-1870.

OSPAR. 2013. An assessment of the ecological coherence of the OSPAR Network of Marine Protected Areas in 2012. OSPAR Commission, London.

Ovaskainen, O., and I. Hanski. 2003. How much does an individual habitat fragment contribute to metapopulation dynamics and persistence? Theoretical Population Biology 64:481-495.

Palsbøll, P. J., M. Bérubé, and F. W. Allendorf. 2007. Identification of management units using population genetic data. Trends in Ecology and Evolution 22:11-16.

van Sebille, E., S. M. Griffies, R. Abernathey, T. P. Adams, P. Berloff, A. Biastoch, B. Blanke, E. P. Chassignet, Y. Cheng, C. J. Cotter, E. Deleersnijder, K. Döös, H. F. Drake, S. Drijfhout, S. F. Gary, A. W. Heemink, J. Kjellsson, I. M. Koszalka, M. Lange, C. Lique, G. A. MacGilchrist, R. Marsh, C. G. M. Adame, R. McAdam, F. Nencioli, C. B. Paris, M. D. Piggott, J. A. Polton, S. Ruhs, S. H. A. M. Shah, M. D. Thomas, J. B. Wang, P. J. Wolfram, L. Zanna, and J. D. Zika. 2018. Lagrangian ocean analysis: Fundamentals and practices. Ocean Modelling 121:49-75.

Vestfals, C. D., F. J. Mueter, J. T. Duffy-Anderson, M. S. Busby, and A. De Robertis. 2019. Spatio-temporal distribution of polar cod (Boreogadus saida) and saffron cod (Eleginus gracilis) early life stages in the Pacific Arctic. Polar Biology 42:969-990.

Virtanen, E. A., M. Viitasalo, J. Lappalainen, and A. Moilanen. 2018. Evaluation, gap analysis, and potential expansion of the Finnish marine protected area network. Frontiers in Marine Science 5:402.

X. Appendices (to be included)

Table A1. Extracted data from a literature review of traits relevant for dispersal of invertebrate and fish larvae. Excel file ‘’Table A1.xlsx’

List of connectivity matrices. Excel file ‘List of Connectivity Matrices.xlsx'