Blue Growth & Marine
Spatial Planning Scheme

Modelling the Impact and Management of Marine Invasive species

Technical report

Lead Partner: University College of Dublin
Partners: Marine Institute
Authors: Clavel-Henry Morane, Crowe Tasman P, Giesler Rebecca J, Yearsley Jonathan M

Ireland's EU Structural and Investment Funds Programmes 2014 - 2020
Co-funded by the Irish Government and the European Union
An Roinn Talmhaíochta, Bia agus Mara
Department of Agriculture, Food and the Marine
EUROPEAN UNION
This measure is part-financed by the European Maritime and Fisheries Fund
Marine Institute
Foras na Mara

Operational Programme	European Maritime and Fisheries Fund (EMFF) Operational Programme 2014-2020
Priority	Union Priority 6 – Fostering the Implementation of the Integrated Maritime Policy
Thematic Objective	TO 6 – Preserving and protecting the environment and promoting resource efficiency
Specific Objective	SO1 - Development and implementation of the Integrated Maritime Policy
Measure	Blue Growth & Marine Spatial Planning Scheme
EMFF Certifying Body	Finance Division, Department of Agriculture, Food and the Marine
Managing Authority	Marine Agencies & Programmes Division, Department of Agriculture, Food and Marine
Specified Public Beneficiary Body	Marine Institute
Grant Rate	100%
EU Co-Financing Rate	50%
Legal Basis	Article 79 and 80 EMFF

This project or operation is part supported by the Irish government and the European Maritime & Fisheries Fund as part of the EMFF Operational Programme for 2014-2020

Ireland's EU Structural and Investment Funds Programmes 2014 - 2020
Co-funded by the Irish Government and the European Union

An Roinn Talmhaíochta, Bia agus Mara
Department of Agriculture, Food and the Marine

EUROPEAN UNION
This measure is part-financed by the European Maritime and Fisheries Fund

Marine Institute
Foras na Mara

Modelling the Impact and Management of Marine Invasive Species

Technical report

LIST OF TABLES

LIST OF MAPS

LIST OF FIGURES

1. General Introduction

Marine invasive species are a global concern in the world for their threats on biodiversity and socio-economical activities. Their introduction in new environments has been magnified by changes in coastal structure, the intensification of global maritime shipping movements, aquaculture, and climate change (Dukes and Mooney, 1999; Katsanevakis et al., 2013). The success of any many actions against invasive species also depends upon the invasion stage. It is admitted that they can be effective at preventing introduction and control establishment, but actions are generally ineffective against expansion of the species. Yet, the need to understand the processes leading to invasion and the monitoring of invasive species is fully acknowledged and encouraged by national and international infrastructures. In Europe, the EU Marine Strategy Framework Directive (MFSD), in the first MFSD version 2008[1] and then in the revised version in 2017[2], presented the non-indigenous species as one of the qualitative descriptors to measure the Good Environmental Status (GES) of the European seas. The GES of a region is gauged by surveying and estimating the species presence (see Descriptor 2 in Official Journal of the European Union 2010/477/EU), which for the latter relies on the fact that the species could be soon introduced, newly introduced, or established in the region. To fulfil the demand from the European infrastructure, the assessment of the invasion risk in a region can be investigated using several methods that would optimise monitoring and management strategies.

In the project "Modelling the Impact and Management of Marine Invasive Species" (MIMIS), the assessment of invasion risk in a region relied on combining two modelling methods (hereby called modelling tools) which estimated 1/ the transport of pelagic larvae away from known sites of an invasive species to the surrounding region based on numerical and stochastic computations from Lagrangian transport model and 2/ the habitat suitability for the establishment of new invasive species populations in the surrounding region based on a species distribution model. Modelling generally allows the complete picture of the invasive species spread and settlement in a region to be estimated where observation and sampling can only detect the species presence in a few places and with limited time. Models have a great plasticity which permits them to account for a species' known ecology, the observed occurrences of the species, and the local environmental conditions. The results are spatially resolved and can therefore have many levels of interpretations if combined with other spatial data (e.g. GIS layers on biodiversity, cultural sites, and socio-economic activities). Illustrated in the MIMIS Management report, the outputs from modelling tools were used to understand a species' invasion milestones and helped to frame the biological processes and physical mechanisms which contributed to a species' spread. Furthermore, as implied by the terminology, "modelling tools" is a toolbox of methods that can be adapted to the user objectives (e.g., adapted to a new invasive species). For instance, with a management purpose, a Lagrangian transport model can simulate a management action. Then, with the newly modelled simulations, we can assess the spread of invasive species and estimate the efficiency of the management action.

Overall, these modelling tools have already been developed in many cases implying management near Ireland coastal waters (Robins et al., 2017; Wood et al., 2021). The MIMIS project aimed to investigate the processes of spread and settlement for the management of marine invasive species specifically. In the present report, technical information is documented on the detailed modelling effort that underpins the project's objectives.

The specific objectives of the report are the following:

- Providing guidance for modelling marine invasive species spread and settlement.
- Giving explanations and rationales on modelling choices
- Representing the limits of the modelling
- Setting the parametrisation for modelling two targeted invasive species: *Magallana gigas* and *Hemigrapsus sanguineus.*

[1] https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32008L0056&from=EN
[2] https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32017D0848&from=EN

2. Modelling Framework

In the MIMIS project, the modelling analysis needed to follow several steps and to carry out sensitivity analyses (**Figure 1**). The main modelling effort involves running simulations of larval dispersal by coupling hydrodynamic and Lagrangian transport models with ecological information on the targeted marine invasive species (e.g. planktonic larval duration, vertical migration behaviour in the water column). This Lagrangian transport model produces thousands of simulated tracks of larval dispersal. This raw simulated spread output can be processed by discarding larvae that do not satisfy selected criteria (e.g. tracks that do not overlap with habitat suitable for larval settlement). One of these criteria is a habitat suitability map estimated from a species distribution model showing the favourable places for the species establishment. This criterion was the secondary modelling effort of the MIMIS project and was independent of the larval dispersal model. To sum up, the modelling framework is divided into four chronological sections (**Figure 1**) and are developed as the "Initialisation", the "transport scheme", the "settlement scheme", and the "Output analyses".

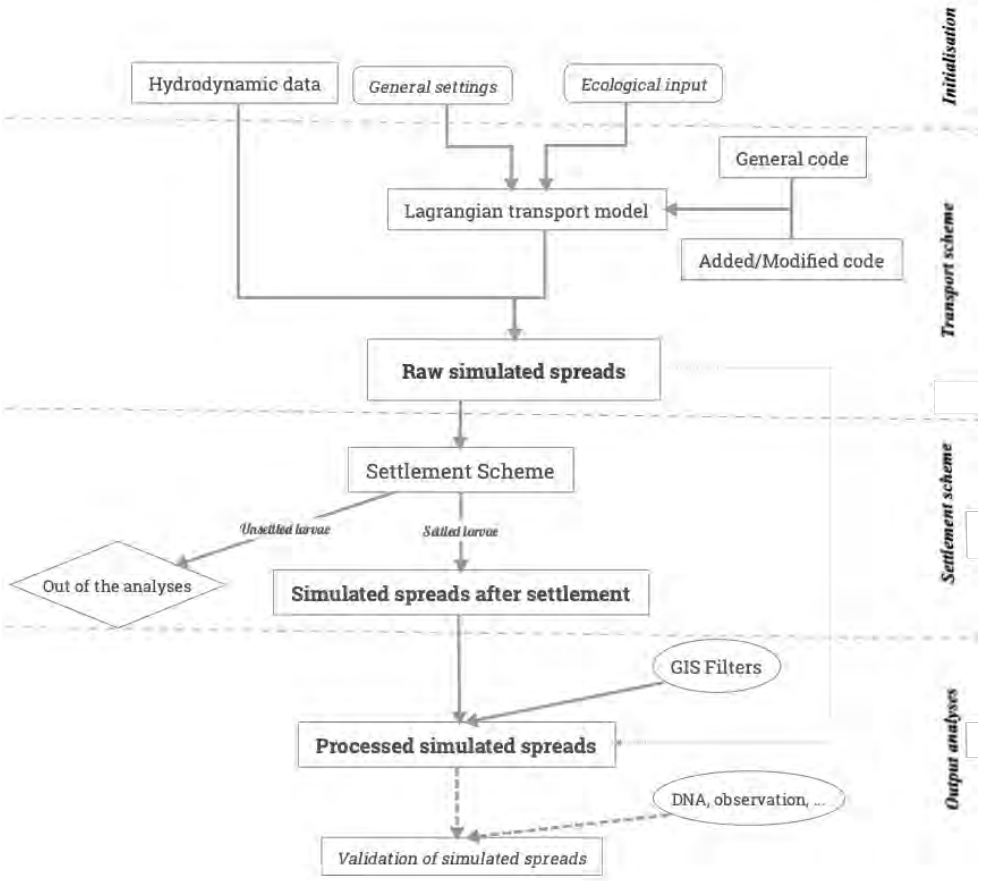


Figure 1. Chart of the modelling effort to map the invasive alien species in the MIMIS project

2.1. Initialisation

Larval dispersal simulations are run *offline*, which means the runs are done separately to the production of hydrodynamic fields. Tracking particles require mandatory inputs from hydrodynamic models and from the parametrisation of the Lagrangian tracking model. To specifically track larvae, seen as an individual, the parametrisation can be adjusted with ecological knowledge on the targeted species.

2.1.1. Mandatory inputs and parametrisation

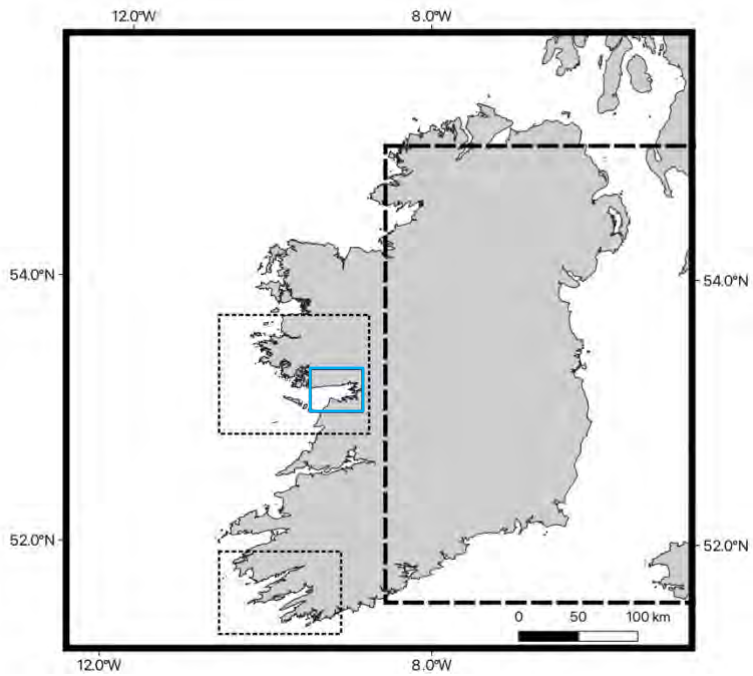
Hydrodynamic field

Hydrodynamic fields are gridded data over a spatial extent and provide temporal information of the water properties (salinity, temperature) and dynamics (velocity currents) in the three-dimension (latitudinal, longitudinal, depth) by resolving the Stokes equation for the movement of uncompressed fluids.

For running the Lagrangian transport model, the modeller often has limited choices on the available hydrodynamic fields. If not by default, the best choice of hydrodynamic fields is the one with relevant space and time resolutions regarding the modelled area and the studied species. Considering a coastal marine species highly dependent on the tidal cycle, a relatively fine resolution in time and space is required (i.e., a temporal resolution less than 6 hours).

We are aware of various hydrodynamics datasets for Irish waters that are available as inputs to Lagrangian transport models (**Map 1**):

- ***National scale:*** the whole Ireland with a Regional Ocean Modelling System (ROMS, Shchepetkin and McWilliams; 2005) model of 1.1-1.6 km spatial resolution and a daily temporal resolution (Nagy et al., 2020a).
- ***Regional scale***: two regions on the west of Ireland with ROMS models of ~200 m spatial resolution and 1h to 3h time resolution (Dabrowski et al, 2016; Nagy et al., 2020a; Nagy et al., 2020b).
- ***Regional scale***: a region of the east of Ireland and the west of the UK with a Stony Brook Parallel Ocean Modelling (sbPOM, Jordi and Wang, 2012) model of ~1.85 km spatial resolution and 30 minutes time resolution (Robins et al., 2017).
- ***Local scale***: a local area in the west of Ireland with ROMS model of 70 m spatial resolution and 30 minutes time resolution



Map 1. Available hydrodynamics models for Irish waters. Continuous thick black line: a portion of Ireland's ROMS model around the coastline, dashed black line: a portion of the sbPOM model; dotted black line: Extent of the regional ROMS models (Connemara model and Bantry Bay model; blue continuous line; extent of the local ROMS model.

Considering that the MIMIS project aims to study coastal species that are strongly associated with the intertidal area, the coastal water dynamics and interactions should be well represented. This narrowed the choice to the model with a high spatial and temporal resolution. Finally, hydrodynamic fields at a regional scale from the ROMS models with 200 m spatial resolution were selected for the years between 2013 and 2020. The fields have a temporal resolution of 3 hours between the years 2013 and 2016 and 1 hour between the years 2017 and 2020. These regional scales hydrodynamic models are run and stored by the Marine Institute (Oranmore, Ireland), co-partner of the MIMIS project.

General settings of the Lagrangian model

A Lagrangian transport model uses a scheme to compute the displacement of neutral small-sized mass-conservative particles at a time *t* based on an advection algorithm (e.g., fourth order Runge-Kutta). A set of parametrisations is absolutely needed for initialising and running the model. It consists of initial settings that are independent to what the particle symbolises (i.e., a larva in the MIMIS project) and optional settings that can be implemented with data species ecology (see 4.1.2). Lastly, some settings can have temporal variations through the dispersal simulation (i.e., larval swimming velocity).

Overall, the access to the setting modification is specific to the Lagrangian transport software and explained in guides and tutorials. The software choice is up to the modeller, the study interest, and the compatibility with the hydrodynamic model. For the MIMIS project, LTRANS software[3] (version 2b) was selected for its compatibility with the ROMS outputs and its already-existing application for tracking coastal oyster and crab species - the two selected target taxa for the MIMIS project. LTRANS detailed parametrisation and explanation of the transport schemes are provided in the User Guide (Schlag and North, 2012). The simulation runs rely on instructions and values given in the LTRANS.data file, which are later used/called in code chunks written

[3] https://northweb.hpl.umces.edu/LTRANS.htm and https://github.com/LTRANS/LTRANSv.2b (Accessed 29th Aug 2022)

in different files. In other words, LTRANS.data is a directive file and is often modified for initializing successful runs of larval transport simulation.

The settings that are fully independent to ecological knowledge are for solving the displacement schemes during the particle tracking simulation:

- *The external time step* (*dt* in LTRANS.data*)*: time step between hydrodynamic outputs.
- *The internal time step (idt* in LTRANS.data*)*: time step for tracking particles.
- *The diffusive displacement scheme* can 1) be enabled by a logical (TRUE/FALSE with *HTurbOn* and *VTurbOn* in LTRANS.data) and 2) parametrised with constant diffusivity coefficients (*ConstantHTurb* and *constAKs* in LTRANS.data) if they are not provided in the hydrodynamic data.

Settings are also needed to link the ROMS outputs to the Lagrangian transport model: specifying the variable names (e.g., current vectors), the size of the ROMS grid, and the path for accessing the ROMS output.

Another type of setting is related to the particle parametrisation (**Table 1**). When a particle represents a living organism (e.g., larvae), those settings can be associated with attributes depending on the species' ecology. The settings can be given as a unique constant for all particles, a constant drawn randomly from a specific distribution (e.g., Gaussian, Binomial) for each particle, or a value from a function that depends on external conditions and time. After parameterisation each particle becomes an individual that is independent from the others.

Table 1. General Lagrangian particle tracking parameters and their equivalence to marine species ecology. *: optional parameter; #: parameters that can be constant or that can change through time

Model parameterisation	***Species ecology***
Particle tracking duration	Pelagic life duration
Release coordinates and depth	Known or assumed spawning places
Release time	Gonad maturity time of the spawners
*Release frequency**	Spawning frequency
Boundary collision action	Mortality, Settlement
*Particle number**	Eggs/individuals per spawning event
Particle vertical displacement#*	Buoyancy, vertical swimming
Particle horizontal displacement#*	Horizontal swimming
*Particle characteristics changes**	Size, shape, conditions

Optional parameters are useful for adding ecologically relevant detail to the Lagrangian transport modelling but are unnecessary for a basic dispersal simulation (**Table 1**). Nonetheless, including these optional parameters should improve the accuracy in representing the dispersal of the targeted species.

In the MIMIS project, particles represent the pelagic stages of marine species, which could involve eggs, larval and post-larval phases. Ecological knowledge on the pelagic stages of marine species, is used to set up the Lagrangian model. When knowledge is lacking, we use information from species which are taxonomically close to the studied species.

2.1.2. Ecological mandatory inputs to the Lagrangian model

Two mandatory inputs to the Lagrangian transport modelling are release information about particles (i.e., coordinates, depth, and time/date of release) and the tracking duration.

Out of all particles settings that can be adapted to the ecology of a pelagic early-life individual, the *tracking duration* can have a major impact on the particle trajectory. *Tracking duration* is symbolised by the Pelagic Larval Duration (PLD). That is the time spent in pelagic waters from the release of a larval individual to the age when the individual becomes fixed to a substrate (if the species is a benthic species), or when advection stops being the limiting factor for dispersal (if the species is a pelagic species). The PLD parameter can be a constant or a relationship based on external and intrinsic factors (e.g., water temperature; O'Connor et al., 2007). In the MIMIS project, the PLD was constant for the two target species.

The release event is also important because of the spatiotemporal variability in the hydrodynamics. Different release times from the same location can lead to diverging larval trajectories. Similarly, larval trajectories from larvae released at the same time but at two close coordinates can also diverge. Knowledge on the reproductive ecology of the spawner helps in defining the location and time of larval release. Coordinates (Longitude and Latitude), depth, date and time of release are stored in a *csv* file whose name and file directory path are given in the LTRANS.data file. Each line in the csv file represents one particle. The total number of lines in the CSV file should be *numpar* in LTRANS.data.

2.1.3. Ecological optional inputs to the Lagrangian model

Additional, optional inputs can be enabled in the Lagrangian transport model and involve a deeper knowledge of the larval ecology. In LTRANS source codes, additions/modification related to the larval ecology and active larval displacement are programmed in the file called behavior_module.f90.

The optional inputs account for the vertical and horizontal positionings of the larvae and "larval attributes" during the simulations. The *vertical positioning* of a larva in the water column is one of the most used optional inputs because it has a major impact on the larval trajectory. In the MIMIS project, the larvae are defined so that they can change their vertical positioning pattern with age and their swimming capacity can increase with time. The *horizontal positioning* describes a larval swimming/crawling behaviour, but it was not activated in the MIMIS project. The *larval attributes* describe several independent modules for the larval status through time (e.g. growth, mortality or survival, larval development stage changes, and settlement competency). These types of modules are used to better describe the effect of the environment on larvae and to better represent the larval conditions at the end of the simulations (Mao et al., 2019; Vikebo et al., 2007). In the MIMIS project, among the larval attributes, only the larval development stages changed because it was the criterion for modifying the vertical positioning pattern (O'Connor et al, 2007).

2.1.4. Sensitivity analysis to model choices

Sensitivity of the particle tracking output to the model's inputs should be investigated to explain their selection and to detect the modelling limitations. In the MIMIS project, sensitivity analyses are carried out on three ecological and two physical input settings. Full details on the methodology and results are described in the annexes I to IV.

Ecological input settings

In the MIMIS project, an unbiased sensitivity analysis is done to determine the number of particles that should be released and the release frequency of particles (Annex I). Another sensitivity analysis looks at the effect of varying larval behaviour and release date upon the larval trajectories (MIMIS project Task 2.1: Model larval transport scenarios of *M. gigas* and *H. sanguineus* from specific release sites for a range of behaviours and

dates). The effect of a range of larval behaviours (i.e., particle vertical positioning) on the species spread has been assessed (Annex II).

Physical input settings

In the MIMIS project, the diffusivity coefficients are not recorded in the ROMS outputs. Diffusivity knowledge is hardly available from literature, but it has been possible to estimate a diffusivity coefficient adapted to our case study through an unbiased method (Annex III).

Though several years of modelled hydrodynamics are available for simulation of the larval tracking, the time resolutions of the ROMS model were 3 hours from 2013 to 2017 and 1 hour from 2018 to 2020. For coastal species in an environment where water currents are driven by tides and winds, the time resolution of 3 hours might be too low to accurately simulate their dispersal. We therefore tested the effect of the time step resolution from the hydrodynamic model on the particle dispersals (see Annex IV)

2.2. Transport scheme

In the MIMIS project, LTRANS source code (Version 2b) is adapted to the ROMS hydrodynamic model outputs and to the species of interest. To carry out these modifications, we follow two strategies (S1 and S2). If a scheme (e.g., accounting for a specific larval behavior) is already present in LTRANS.data (Strategy S1, **Figure 2**), the existing source code must be modified to the specific need of the modeller. If the scheme is new (Strategy S2), new parameters have to be 1) created in *LTRANS.data*, 2) defined in *LTRANS.h* and *parameter_module.f90*, 3) then set up in the functions of interest in the source codes, and finally, 4) the execution of the modified source codes has to be checked up (S2, **Figure 2**).

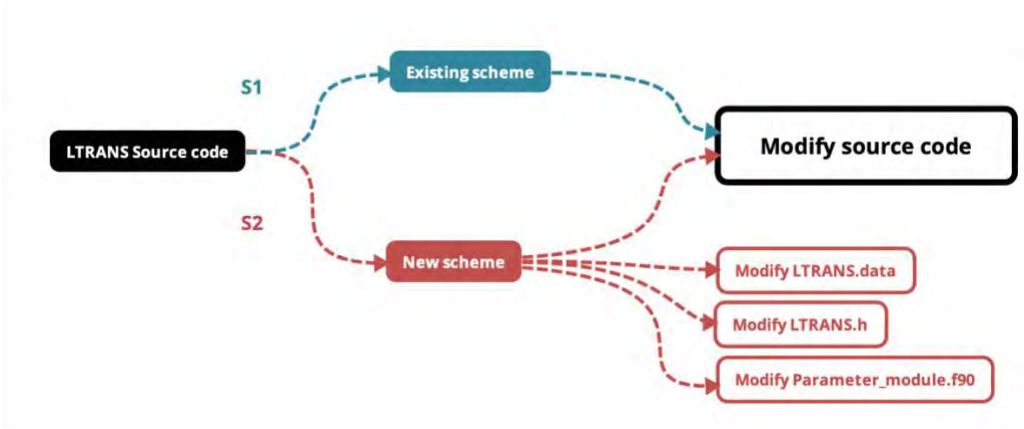


Figure 2. Charts of the strategies for modification of LTRANS codes.

2.2.1. Modifying LTRANS source code to read hydrodynamic data

A priori, LTRANS could not run with the regional scale ROMS (over the Connemara and Bantry regions in Figure 1) because the hydrodynamic fields lacked the vertical velocity variable (*w*) and associated variables (i.e., the *s-coordinate* and *Cs* values on the w-grid). Although the modeller can specify the use of a constant vertical velocity instead of the values from the hydrodynamic model, LTRANS still tries to read the variables from the hydrodynamic fields. To prevent that error, the scheme in *hydrodynamic_module.f90* was modified in order to not read these variables. The commands are commented out of the source code so the commands could be reinstated if the reading of the vertical velocity variables is required in the future.

2.2.2. *Modifying LTRANS source code to include specific larval ecology*

We added two new schemes of larval behaviour to the seven pre-set existing schemes in *behavior_module.f90*: one adapted to the larval behaviour of the Pacific oyster (called "Ebb/Flood transport") and one to the larval behaviour of the Asian crab (called the "Daily DVM").

For the Pacific oyster, between specific ages of the particle, the behaviour was dependent on the tide: at ebb tide, particles swim towards the bottom, and at flood tide, the particle swim towards the surface. This scheme is set up in the subroutine *behave* of the *behaviour_module.f90* file. The scheme assumes a flood tide occurs when the free-surface depth (*zeta*) at time *t-1* is lower than the free-surface depth at time *t*. If so, a Boolean value is set to 1. Otherwise, it assumes an ebb tide, and a Boolean value is set to 0. *Zeta* is a hydrodynamic variable which represents the sea surface variation and was a good proxy of the tidal cycle (**Figure 3**). According to the Boolean value, the scheme will call existing scheme: near-surface displacement if the Boolean value is 1, near-bottom displacement otherwise. This scheme is to make use of the existing scheme as done for the Salinity-oriented behaviours of two other oyster species set up in LTRANS source code (the 4th and 5th in the subroutine).

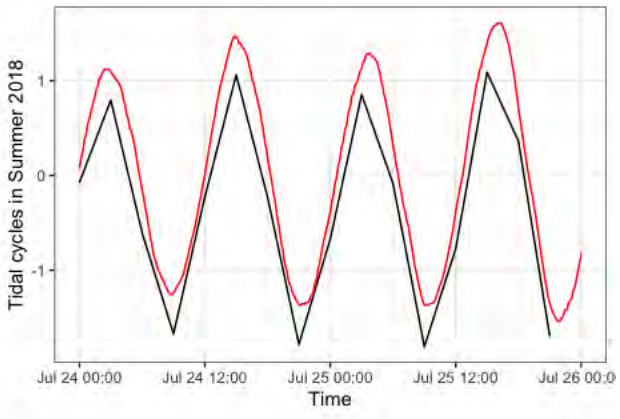


Figure 3. Tidal cycle in Galway Bay. In red, water level (meters) based on Ordnance Datum level from a gauge in Galway Port[4]. In black, free-surface depth (zeta, in meters) from the ROMS model in a location within Galway Bay

For the Asian shore crab, the behaviour was driven by the light intensity at specific larval ages. However, instead of using the existing Diel Vertical Migration (DVM) scheme in LTRANS source code, we added a new scheme very similar to the original except that the day length and the irradiance at solar noon had variations across days. In the *LTRANS.data* file and its *$dvmparam* section, the parameter *twistart* for the twilight start time, *twiend* for the twilight end time, *daylength* for the day length, and *Em* for the irradiance at solar noon are kept but in the new scheme, the code ignores the given values. Instead, we added a parameter called *dirDVM_table* which gives the directory path to a four-column *csv* file containing the day of the year, its length, and when the twilight starts and ends for one selected year (i.e., 2018). A parameter called *JDay* was also added and is the day of the year corresponding to the first hydrodynamic field used by LTRANS. For instance, if a particle simulation starts on July 1st of 2018, JDay value is 182. This value and the simulation timer help to find the day of the year for the particle at time *t*. Then, with the day of the year at time *t*, a subroutine called *Find_daylength* and added in the *behavior_module.f90* file looks for the day length, twilight start and end times in the *csv* file provided with *dirDVM_table* (**Figure 4**). Finally, the Irradiance at solar noon for that day of the year is then computed with a subroutine added in the *behavior_module.f90*. That subroutine is initially written in *light_v2_4BlueCrab.f90* kindly provided by Elizabeth North[5].

[4] Real tidal data from: www.digitalocean.ie/Data/DownloadTideData/Galway%20Port

[5] University of Maryland Center for Environmental Science, Horn Point Laboratory, Cambridge (MA, USA)

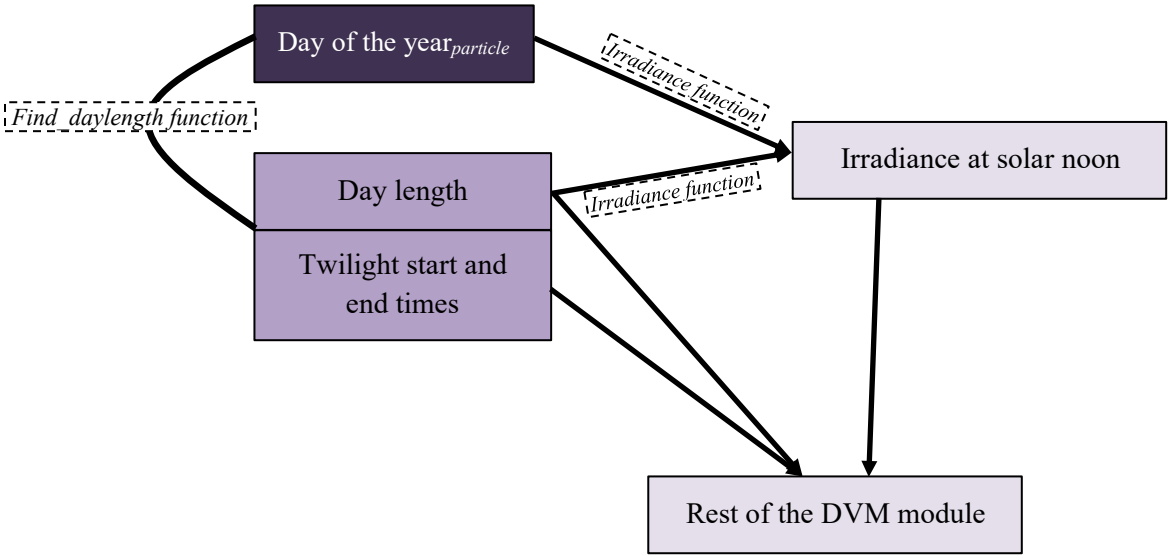


Figure 4. Description of the modification in the new DVM scheme in *behavior_module.f90*

The run of LTRANS tracking model produces a "raw" dataset that records the position (i.e., latitude, longitude, and depth) of the particle at every *advection time step*. They can be provided in the practical *netcdf* format or in the easily readable *csv* format. Additional variables are stored such as the particle age, the water properties at the particle position, the particle condition, etc… The dataset was then post-processed in R software to clean and shape the data for further analyses with GIS layers (**Figure 1**).

2.3. Settlement Scheme

A first post-process analysis effort was to define if the simulated larvae settled over suitable habitats. First, we estimated areas that were suitable for settlement and then, we identified the larvae that encountered these suitable areas during their competent period for settlement.

The suitable habitats can be based on adult population observations or its depth range, but in the MIMIS project, it is estimated by a species distribution model. The species distribution model is built on the relationship between spatial observation of the species and environmental variables. With such model, knowledge on occurrence distributions can be estimated at remote places and/or difficult to survey. The final model outcome is a GIS layer with predicted occurrences between 0 and 1. During the competent period for settlement, a larva is flagged as having arrived over a suitable habitat for settlement if the predicted habitat suitability exceeded some pre-defined threshold. In LTRANS, a settlement module can be used while the larval tracking simulations run (North et al., 2006; Schlag and North, 2012). However, in the MIMIS project, we flagged the particles as settled in a post-process analysis of the simulations. Coupling Lagrangian larval tracking and species distribution models is developed in many other cases, as, for example, in Storlazzi et al. (2017) and Cetina-Heredia et al. (2019).

The success of settlement also depends on the larval settlement competency, which spans over a species-dependent duration. In the MIMIS project, we checked up at an age t varying from the start to the end of the competent larval development stage if the larval location overlapped with suitable habitats. As soon as the location was favourable for settlement, the position and age of the larvae are stored, and all later larval positions were discarded.

Although possible in LTRANS, we ignored the fact that settlement ability of a larva can change during the competent period for settlement (see for example, Connolly and Baird, 2010; Cecino and Treml, 2021) due to

lack of knowledge. For the same reason, the fact that larvae can actively seek out a settlement location was not considered in the MIMIS project.

2.4. Outputs analysis

A second post-process analysis was to assess the invasion risk and impact from the larval spread to selected sites (e.g., Marine Community Types). We used different environmental and socio-economical GIS layers as informative layers about what damages and pressures the arrival of larvae growing into population can cause (Table 2). The methodology to extract information from gridded or polygon-shaped GIS layers is to extract the attributes of the GIS layer at a particle's position and at specific times. When the GIS layers are spatial point data, a circular buffer with a specific radius is set up around the point and if the particle falls within the buffer, the attributes of the GIS layers are extracted. The value of the radius is of 300 m, or approximatively the diagonal length for a grid cell of the regional scale hydrodynamic model. Analyses of the larval spread in these layers are provided in the MIMIS Management Report.

Table 2. GIS layers for management assessments in the MIMIS project

Name	Type	Age of particle at extraction	Origin
Empty Grid	Gridded (2D)	Age at settlement period	MIMIS Project
Daily Bottom Water temperature	Gridded (3D)	All ages during the settlement period	ROMS hydrodynamic fields
Galway Bay subdivisions	Shapefile type polygon	Age at settlement	MIMIS Project
Seafloor substrate	Shapefile type polygon	Age at settlement	EMODNET
Marine Community Type	Shapefile type polygon	Age at settlement	National Parks and Wildlife Service
Piers Boulder Marinas	Shapefile type points	Age at settlement	Clare Coastal Architectural Heritage Survey 2008 Galway County Council
Sites with Farmed and Wild Oysters	Shapefile type points	Age at settlement	Irish Department of Agriculture Food and the Marine Gov.ie

Seven GIS layers are used in the project (Table 2):

- The "Empty Grid" layer is used to compute the larval density maps and extends from 10.79 to 8.89 ºW and from 52.95 to 43.73 ºN with a horizontal resolution of ~ 250 m. Each grid cell of the layer contains an ID number which is a reference for aggregating all larvae falling in the same grid cell.
- The "Daily bottom water temperature" layer allows us to visualise the water temperature at settlement time and assess the quantity of larvae that may survive. It also permits to visualize the effect of global warming of the ocean on the survival of invasive species.
- The "Galway Bay subdivisions" gathers the local names of Galway's sub-bays and inlets. Their size and delimitation are subjective, but they are based on shellfish license site denominations. It helps to understand the connectivity between places in Galway Bay.

The remaining four GIS layers are to assess the conditions of the environment ("Seafloor substrate"), ecology ("Marine Community Types"), and socio-economy (i.e., "Piers, Boulder, Marinas" and "Oyster farms") at simulated larval settlement.

Future analyses could use many other layers, such as recreational areas (e.g., beaches) or saltmarsh habitats.

2.5. Validation and Limitations of the Modelling

2.5.1. Validation of the larval spread

The validation of the Lagrangian transport model outputs can be achieved using genetics, survey observations, or comparing the outputs with buoy trajectories.

In the MIMIS project, the validation of the outputs is limited and relies on the connectivity between the release sites of larvae and observed occurrences of the species. The possibility of using eDNA in the future for validating larval tracking simulations was discussed with the GMIT research group (SERV-19-MEFS-004).

2.5.2. Limitations of the Lagrangian Transport Modelling Approach

The presented modelling efforts were the best approach with the available tools and knowledge to represent the spread of invasive species in Ireland. Nonetheless, the limitations of the modelling approach must be acknowledged when using the results (e.g., deciding on management actions).

The uncertainties in the modelled species spread are primarily due to uncertainties in the hydrodynamic fields, the displacement schemes within the Lagrangian Transport model, and the ecological settings of the species. The major uncertainty of our results comes from the ecological settings. Indeed, the larval ecology and their settlement preferences are more complex than what we know and can parametrise in the settings. Uncertainties introduced by Lagrangian Transport model are mathematical. For instance, for displacements by advection only, the vertical and horizontal interpolation scheme of the current velocities to the particles and the advection model from a Lagrangian Transport model to another are different and substantiates variations in the modelled larval dispersal (see guidebooks of LTRANS; Schlag and North, 2012 and OpenDrift; Dagestad et al., 2018). Uncertainties issued from hydrodynamic fields are minor. Besides, the hydrodynamic models have passed several validation analyses. More details of these limitations and consequences for the management are described in the Section 4 of the MIMIS Final Report ("Applying modelling approaches to assess the spread and management of marine non-indigenous species").

3. Case Studies

Two targeted species, *Magallana gigas* and *Hemigrapsus sanguineus* were selected in the MIMIS project with a rationale explained in the MIMIS Management report.

We focused the mapping of the invasive species spread on Galway Bay because this region satisfies three important prerequisites:

- At least one of the two invasive species is reported in the region.
- It exists hydrodynamic fields data with a spatiotemporal resolution good enough to capture water circulation dynamics in tidal and coastal environments,
- The impact of the invasive species spread can be measured and assessed with multiple and diverse available GIS layers (see Table 2).

Only the region of Galway Bay can comply with the three conditions. The below sections gather detailed technical information on the spread modelling and post-process analyses of the two targeted species.

3.1. Hydrodynamics fields

The hydrodynamics of Galway Bay were modelled by a Regional Ocean Modelling System. The configuration and validation of this model are described in Nagy et al. (2020a) and Nagy et al. (2020b).

The model covers an extent from 10.8°W to 8.9°W and from 52.95°N to 53.73°N (Map 1). It is of ~200 m horizontal resolution and has 20 vertical sigma levels. The time resolution of the hydrodynamic fields is 3 hours from 2013 to end of 2016 and 1 hour from 2017 to end of 2020. To simulate the coastal marine species spread, we only select the last three years with an hourly time resolution (i.e., 2018-2020). The years 2019 and 2020 are characterised by water temperatures being cold and warm, respectively, while the 2018 year was a mild year (**Figure 5**). These variations are also reported with the North Atlantic Oscillation that represents anomalies in the surface sea-level pressure in latitudes including Ireland[6].

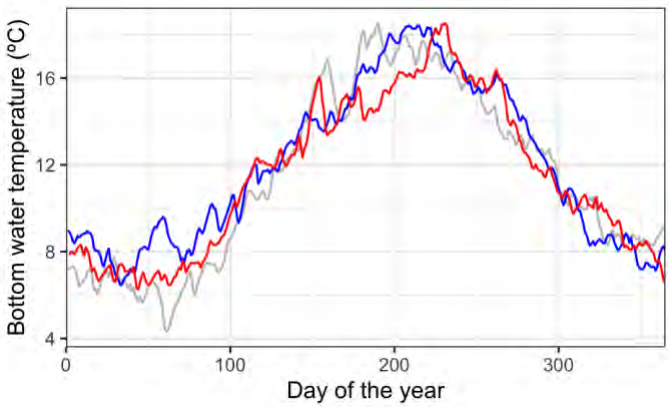


Figure 5. Daily variations of the bottom water temperature over seafloor depth between 0 and 5 m. The year 2018 is in grey line, 2019 in blue line, and 2020 in red line.

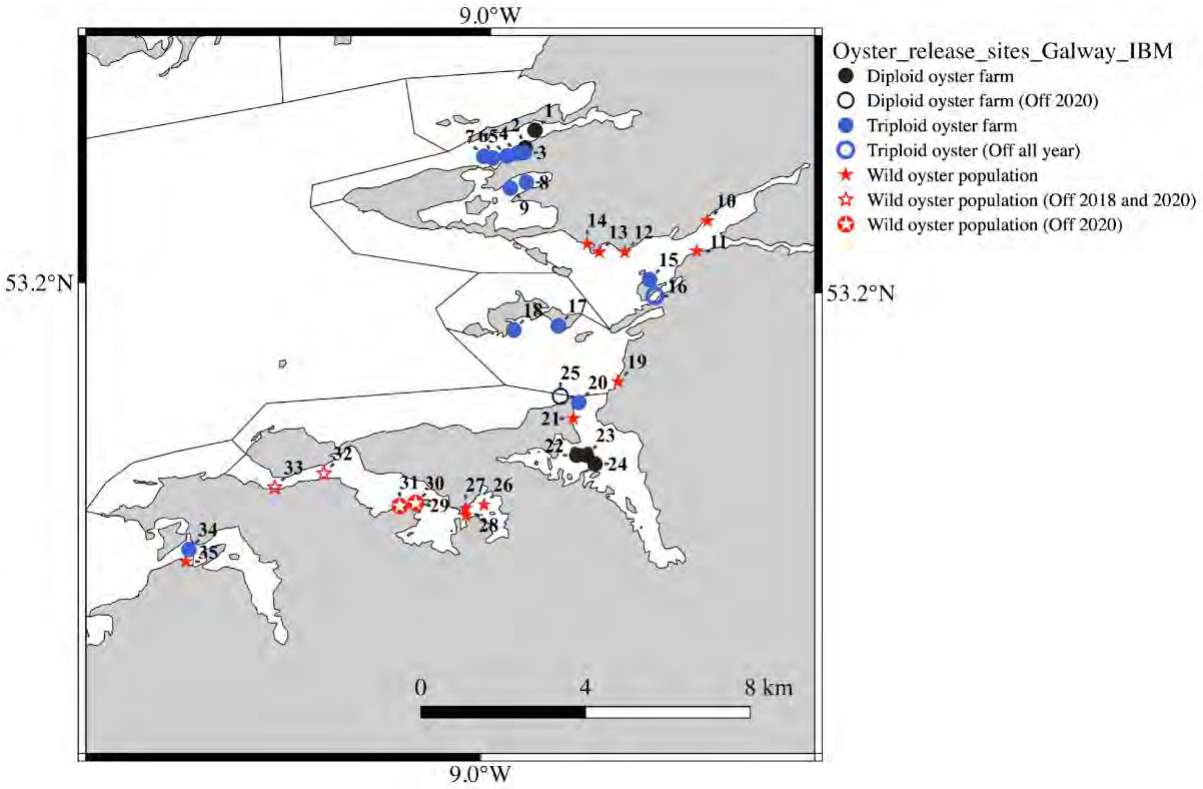
[6] https://www.cpc.ncep.noaa.gov/products/precip/CWlink/pna/nao.shtml

3.2. Case Study 1: Magallana (Crassostrea) gigas

The knowledge on this commercial species is well detailed for the adults, but not well developed for the pelagic phases. Some parameters were extracted from *Crassostrea virginica*, a species taxonomically closed to *M. gigas* and having several larval analogies with the Pacific oyster (Pauley, 1988). All GIS layers listed in **Table 2** are used for assessing the spread impact from the Pacific oyster.

Inputs in the Lagrangian tracking model

Larval transport simulations are initialized from licensed oysters farm sites (stipulated as diploid, triploid or und. breed in trestles) and sites of recorded wild oysters (see **Map 2**). When spatially projected against a hydrodynamic field (e.g., water temperature), four oyster sites are located over the land. The four sites are moved to the closest water sites (e.g., sites 8 and 9 in **Map 2**) or moved at the entrance of a sub-bay (e.g., sites 1 and 35 in **Map 2**) within the hydrodynamic fields. It is therefore assumed that larval dispersal from the modified locations of the sites would be similar to the dispersal from their original location.



Map 2. Source sites with adult oyster that are potentially capable of spawning larvae in Galway Bay.

Times for the spawning event are estimated per site according to a day degree threshold and the water temperature during 6-7 days after reaching the day degree threshold. More concretely, sites for which the day degree exceeds 592 and water temperature is warmer than 16 ºC for 6-7 days consecutively are suitable for larval releases. The day degree is the cumulative sum of the differences between the daily water temperature and a 10.55 ºC threshold (Mann, 1979; Mills, 2016). Day degree is an indicator of when the oyster gonad could reach maturity and the oyster could start spawning. The 16 ºC water temperature represents the minimum temperature observed for spawning event (Castaños *et al.*, 2009; Dankers *et al.*, 2004; Ruiz *et al.*, 1992), although the optimum temperature for a spawning event is 18 ºC. The rest of the parametrisation of the LTRANS.data file is provided in **Table 3**.

Table 3. Settings for the tracking of *M. gigas* larvae

Settings	Value/Fact	Criteria in LTRANS.data	Reference
Release event	At night. Between July 22 and September 6, 2018 Between July 21 and August 29, 2019 Between July 29 and September 17, 2020	parfile	592 days Mann, 1979
Release frequency	One day per sites	parfile	
Release location	34 sites	parfile	Farms and Wild oyster populations
Release depth	Near bottom of the release sites	partfile	
Particle number	400 per sites	numpar	See Annex I
Duration of tracking	31 days Larval duration of three weeks +7 days for which larvae were competent to settle.	timeparam	Coon et al. 1990 Collet et al. 1999 Troost, 2010
Behavior	Trochophore: Passive Veliger: Tidal Pediveliger: Near- Bottom	Behavior	Arakawa, 1990
Vertical swimming velocities	Veliger: 0.001 m/s Pediveliger: 0.01 m/s	Swimslow, swimfast	Suquet et al., 2012
Metamorphose age	Trochophore to Veliger: 2 days Veliger to Pediveliger: 21 days		Pauley, 1988

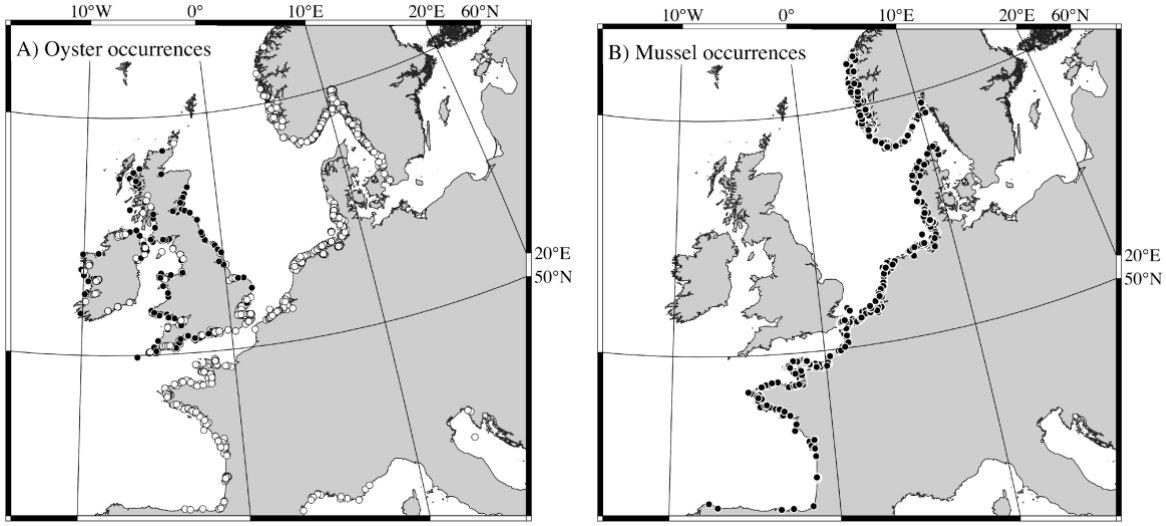
Settlement of the tracked larvae

Crassostrea gigas' suitable habitats for settlement in Ireland were estimated by species distribution models (SDMs) which combined two machine learning models: Maxent and Boosted Regression Trees with the *sdm* R package (Friedman, 2001; Naimi and Araújo, 2016; Phillips *et al.*, 2006). The SDMs rely on independent covariates (here environmental predictors: vertically averaged Temperature (Txx, ºC), Sea surface temperature (SST, ºC), Salinity (Sxx, g kg^{-1}), Sea Surface Salinity (SSS, g kg^{-1}), Dissolved Oxygen (DOx, mmol m^{-3}), pH (pHx) , Chlorophyll A concentration (CHL, mg m^{-3}), Particulate Inorganic Carbon (PIC, mol m^{-3}), Particulate Organic Carbon (POC, mg m^{-3}), Wave Exposure (WaveE), Current velocities (Vel, m s^{-1}), and Bathymetry (Bot2, m), **Box 1**) that are associated with the dependent variable (here occurrence observations).

Box 1. Information on the downloaded environmental predictors open-source data from the below platforms.

Source 1: EMODNET data - Bathymetry *Spatial coverage:* -12 to 13º E, 42.9 to 66º N *Horizontal resolution:* 1/16 arcminutes (59-72 m depending to the latitude)	**Source 2:** MODIS data (NASA) - Chlorophyll A, PIC (Particle inorganic Carbon), POC (Particle Organic Carbon) *Spatial coverage*: -12 to 13ºE, 42.9 to 66 ºN *Temporal coverage*: 2002, July 4 - 2020, July 4. Monthly average *Depth coverage:* 0-75 m *Spatial resolution:* 4 km
Source 3: Marine Copernicus Atlantic-European northwest shelf-ocean physics reanalyses. 3. Temperature, Salinity, Currents *Spatial coverage*: -12 to 13º E, 42.9 to 66º N *Temporal coverage*: 1992, Jan 16 - 2018, Nov 16. *Time resolution:* Monthly average *Depth Coverage*: 0-75 m *Spatial resolution:* 0.111º x 0.067º	**Source 4:** Marine Copernicus Atlantic-European northwest shelf-ocean biology reanalyses. 4. DO, pH *Spatial coverage:* -12 to 13º E, 42.9 to 66º N *Temporal coverage:* 1998, Jan 16 - 2018, Nov 16. *Time resolution:* Monthly average *Depth coverage:* 0-75 m *Spatial resolution:* 0.111º x 0.067º

The presence/absence dataset contains Pacific oyster occurrences (86 absences and 1464 presences) from European waters and mussel *Mytilus edulis* occurrences (1388 presence) (**Map 3**). Mussels are bivalve species with ecological patterns like *M. gigas* and they share similar habitats. Mussel presences were used as pseudo-absences (Barbet-Massin et al, 2012), which means they indicated an *'Absence'* of oyster observations. All extremely unrealistic occurrences (i.e., in the middle of a country) are removed. Mildly unrealistic occurrences (i.e., on land but close to the sea) are moved to the closest water location. Lastly, all records along the Irish coastlines are stored in a separate file for the validation of the species distribution. There were 25 absences and 133 presences of the oyster along the Irish coastline.



Map 3. Distribution of A) the Pacific oyster and B) the common Mussel occurrences in the European Waters. White dot indicates presence and black dot indicates absence/pseudo-absence of the oyster.

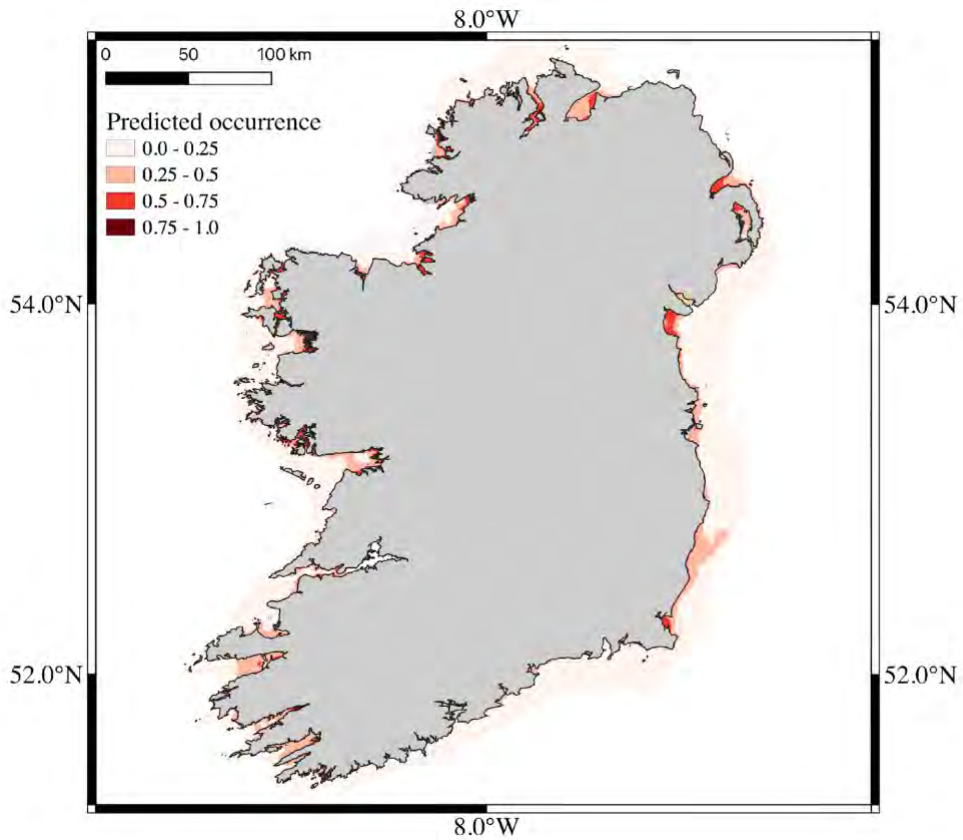
All independent covariates at the location of oyster presence and absence are extracted from gridded data by bilinear interpolation. Covariates are then centred and scaled. Averages and standard deviations from the scaling transformation are saved for the prediction analysis. The SDMs are trained on randomly selected 80% of the dataset. The remaining 20% was used to quantify independently the model performance through *AUC*, *kappa,* and *Specificity* metrics.

The performances of the BRT and Maxent models are adequate and similar to each other (**Table 4**). The chlorophyll A, current velocity, wave exposure, sea surface temperature, and the Particle Organic Components are, in a descending order, the most important covariates of the model.

Table 4. Information about the species distribution model parametrisation and its performance. The sorting of variable importance is based on AUC decreases when the variable is removed from the model. Performance is evaluated with independent data (the testing dataset).

Dataset characteristics	**Model parameterisation**	**Variable Importance**	**Performance**		
Size : 2751 points, Training size : 2200 points	**BRT**: 1300 trees Shrinkage: 0.01 Depth: 6 **Maxent**: Betamultiplier: 2	CHL, Vel, WaveE, SST, POC		**BRT**	**Maxent**
			AUC	0.90	0.89
			Kappa	0.63	0.61
			Spec.	0.72	0.84

Oyster occurrences in the Irish coastal waters are predicted with the built SDMs using gridded covariates. First, a dataset gathers the gridded independent covariates extending from 10.66449º to 5.20062º W and from 51.35768º to 55.45189º N with a high spatial resolution (grid spacing along x = 0.005463871º, grid spacing along y = 0.002408357º). Second, each covariate is centred and scaled using the average and standard deviation of the covariates from the dataset with observed occurrences. Last, an ensemble method for model prediction (Araújo and New, 2007) is selected because both BRT and Maxent models performed well. Therefore, the predictions from BRT and Maxent models are averaged (i.e., ensemble modelling). Maps of predicted occurrence probabilities show spatial variation along the coastline (**Map 4**). Predicted occurrences are high (> 0.75) in semi-enclosed areas like already observed in Lough Swilly (Northern Ireland; Tully and Clark, 2012) and low (> 0.25) along open coastlines.



Map 4. Predicted occurrence of the Pacific oyster in the Irish coastal waters

We conduct an additional statistical validation of the models applying the block cross-validation method (Valavi *et al.*, 2019) that has the following structure:

1) The dataset spatial domain (the European Union water excluding Irish coastal waters) is divided in blocks of 200 km resolution. Blocks without recorded occurrences are discarded from the rest of the analysis.
2) Then, each block is randomly assigned an ID number between 1 and 5.
3) The models are trained with data from four IDs and tested with data from the remaining ID number.
4) The step 3 is repeated four times with all the rest of ID combinations.
5) The Receiver Operating Characteristic (ROC) graph is calculated based on the results of the five testing datasets and on the model type (i.e., *maxent*, *GBM*, and *ensemble*).

Figure 5 shows that the three model types are performing well and that the *ensemble* model adequately retained the best characteristics of the *GBM* and *maxent* model types.

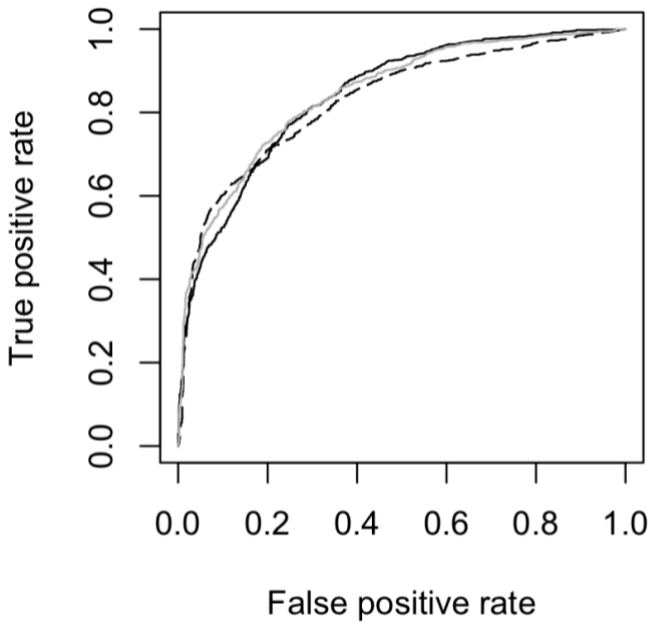


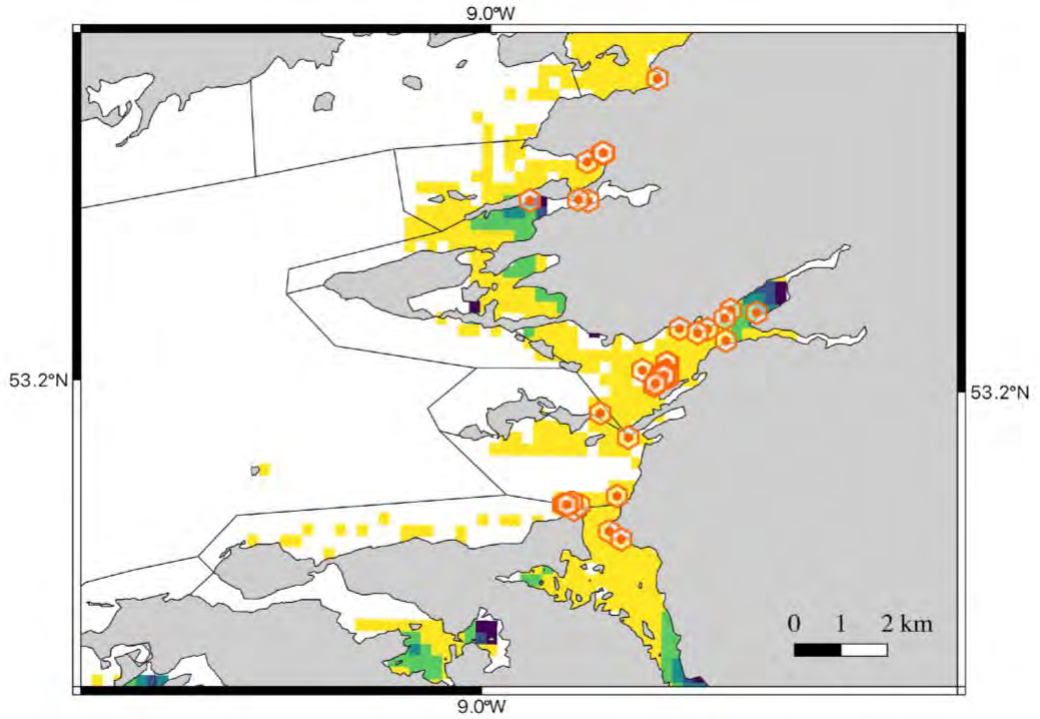
Figure 6. ROC from the block cross-validation with the maxent (continuous line), the GBM (dashed line) and the ensemble (grey lines).

With the predicted occurrence from the ensemble model, the larval settlement could be assessed. We focus on the larval location between the 21st day and the 31st days (10 days long) in the larval dispersal outputs. This time window represents the period during which oyster larvae have settlement competency. If a predicted occurrence from the species distribution model underneath the larval position is higher than a threshold of 0.5, the place is considered suitable for settlement. We record the position of the larvae and the time when the occurrence probability is for the first time higher than the threshold. Larvae that do not reach a suitable place during the competence time frame are flagged 'Unsettled' and could be disregarded from further analyses (i.e., the assessment of the invasive spread impact).

Validation of the models

Validation of the models (SDMs and larval transport model) is succinctly carried out with available independent data indicating the oyster presence. Records are obtained from an independent survey that sampled the eastern side of Galway Bay (**Map 5**). All record locations coincide with SDM predictions of highly suitable areas for settlements, and with surface areas expected to receive at least 1 simulated larval and maximum 1060

larvae. Three newly colonised sites from the independent survey are at locations where an important density of simulated larvae has been predicted (see Tawin, Kilcolgan). However, the remaining of the newly colonized sites are located where density of simulated larvae is low (1 to 44), which can be an indicator that the Lagrangian modelling parametrisation needs some improvements. As for the other places along the southern coastline of Galway Bay where high density of simulated larvae is estimated, the independent sampling survey was not carried out in these zones.



Map 5. Distribution of the wild population oyster (hexagon dots) from the 2018 survey in the north-eastern Galway Bay over the average density of settled larvae.

We also approach the model validation by analysing the connectivity between a recorded source of larvae and the wild oyster population. This approach also helps to understand the pattern of colonization in Galway Bay. The independent data are records of oyster wild populations observed before and after 2018. Pre-2018 records are data from the GBIF atlas and Kochman's study (2012). They were used in the larval dispersal modelling. Post-2018 records are obtained from the independent survey that sampled the eastern side of Galway Bay in 2018. For all records, we count the number of sources that arrive in their vicinity (i.e., within a buffer of ~ 300 m) and identify if they come 1) from wild population sources or from farms with diploid oysters and 2) from the same subdivision or another (**Table 5** and **Table 6**).

For the pre-2018 data locations, Kilcolgan, Ballinderreen, and Kinvarra subdivisions counted wild population sites supplied by diploid farms within or outside the subdivision (**Table 5**). A hypothesis would be that these are subdivisions with a primary introduction of the wild population in Galway Bay from aquaculture. On the other side, Ballyvelaghan and Ballyvaughan wild populations are supplied by wild oysters only. The hypothesis behind these links would be a secondary introduction from previously established wild populations. Besides, the result for Ballyvaughan shows a supply of oysters from the single wild population recorded within the subdivision. This statement reveals that the dispersal model could be underestimating larval transport as we could not detect a link from outside the division to the recorded population. This underestimation also applies to the links detected for the other subdivisions. Another justification could be given by the existence of diploid farms in that area that we were not aware of.

Table 5. Number of larval sources linked to wild oyster populations detected before an independent survey in 2018.

		From the same subdivision		**From outside the subdivision**		
Subdivision	Wild sites (N)	Diploid farm origin	Wild origin	Diploid farm origin	Wild origin	Total sources (N)
Ballyvelaghan	8		8		1	9
Kilcolgan	5		5	4	4	13
Ballinderreen	1		1	3	3	7
Kinvarra	1	2	1		3	6
Ballyvaughan	1		1			1

For the post-2018 data locations, 51 new wild populations are observed in eight subdivisions of Galway Bay. The origins of six new wild population in the subdivisions of Aille, Oranmore, and North of Tawin (i.e., an area without wild oyster population until the recent survey) would only be because of existing diploid farms (**Table 6)** and one new wild population in the North of Eddy Islands subdivision would only be because of an existing wild oyster population. Thirty-two new wild population sites are counted in the Kilcolgan subdivision, either geographically close to the existing oyster population or at more distant places within the subdivision (**Map 5**). In both cases, the simulated larval dispersal could explain the appearance of these new sites which could have been supplied by oyster population within or outside the subdivision.

Table 6. Number of larval sources linked to wild oyster populations detected after an independent survey in 2018.

		From the same subdividision		**From outside the subdivision**		
Subdivision	**New wild sites (N)**	Diploid farm origin	Wild origin	Diploid farm origin	Wild origin	**Total sources (N)**
Kilcolgan	32		4	4	5	13
Aille	5			2		2
Ballinderreen	5		1	4	4	9
Rosshill	4			2	1	3
Kinvarra	2	3	1	1	6	11
Oranmore	1			2		2
North of Tawin	1	2				2
North of Eddy Island	1				1	1

3.3. Case Study 2: Hemigrapsus sanguineus

The knowledge about the species is well described and its ecology is very similar to *Hemigrapsus takanoi*. In fact, the modelled spread can represent the two species. All GIS layers of **Table 2**, except "Sites with Farmed and Wild Oysters", are used for assessing the impact of crab spread.

Inputs in the Lagrangian tracking model

Larval transport is initialized from Galway harbour (53.268 ºN, 9.048 ºW). This simulates a hypothetical introduction of the crab by ballast water. Indeed, an introduction of the crab in Galway harbour can be expected through marine traffic, because Galway harbour has connections to Cork harbour, and Cork harbour is connected to harbours outside Ireland where the crab is present. Time of hatching is an ecological trait that was lacking in the literature. To estimate it, knowledge on the time for gonadal development, egg brooding and hatching was needed. According to Epifanio et al. (2013), the length of the reproductive season can vary with latitude and winter water temperature. The latitude of Galway harbour (53.268ºN) is higher than the sites in Epifanio's review and the average bottom water temperature for the first three months of 2018 was 7.5ºC. In European waters, *H. sanguineus* reproductive season could begin in April/ May and last for 5-6 months (Gothland et al., (2014), Geburzi et al., 2018) with a peak of ovigerous females in summer (Dauvin, 2009). In Galway Bay, we based the hatching time of *H. sanguineus* on *Hemigrapsus takanoi*: it needs a minimum water temperature of 15 ºC for ~20 days to have the brood hatching (Van Den Brink et al., 2013). Using this condition and knowing the water temperature at the release sites, we made larval releases between July 11th and September 12th (**Figure 7**).

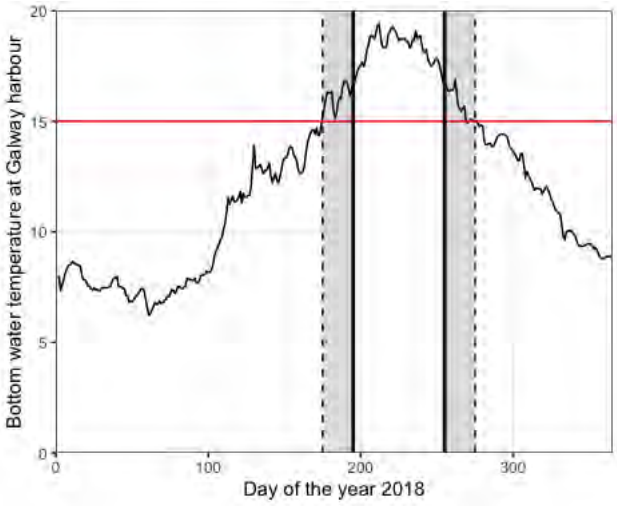


Figure 7. Bottom water temperature at Galway harbour and hatching time for *H. sanguineus*. The grey ribbon shows a duration of 20 days in which the water temperature was above 15ºC for the first time and the last time. Discontinuous lines show the day of the year where bottom water temperature was higher than 15ºC. Continuous lines frame the period of larval releases.

The rest of the parametrisation of the LTRANS.data file is provided in **Table 7**.

Table 7. LTRANS parametrisation settings for the tracking of *H. sanguineus* larvae

Settings	Value/Fact	Criteria in LTRANS.data	Reference
Release event	At night and end of flood tide between July 11th and September 12th.	parfile	Park et al., 2009 Epifanio et al., 2013 Van den Brink et al, 2013
Release frequency	Every night suitable for releases	parfile	See Annex I
Release location	Galway harbour	parfile	Likeliest location of introduction due to marine traffic
Release depth	Near bottom of the release sites	partfile	
Particle number	1000 per sites	numpar	See Annex I
Duration of tracking	35 days at water temperature around 18 ºC	timeparam	Gimenez et al., 2020
Behavior	First two zoeas: Near-surface Last three zoeas: Diel vertical migration Megalopae: Near- Bottom	Behavior	Park et al., 2004 Cohen et al. 2015
Vertical swimming velocities	Zoea I to Zoea II: 0.0233 m/s Zoea III to Megalopa: 0.035 m/s	Swimslow, swimfast	Landeira et al., 2020 Cohen et al., 2015
Metamorphose age	Zoea I to Zoea III: 10 days Zoea III to Megalopa: 15 days		Landeira et al., 2019 Cohen et al., 2015

Settlement of the tracked larvae

Hemigrapsus sanguineus' suitable habitats for settlement in Ireland are estimated using species distribution models. The species distribution modelling was carried out by Ultan O'Donnell (a MSc student at UCD). Details of the data, modelling training and modelling predictions are provided in Ultan's thesis.

Validation

As *Hemigrapsus sanguineus* has not arrived yet in Ireland, validation for the dispersal model cannot be done until detection of the crab in the bay. Validation for the SDM is discussed in the MSc report.

4. Conclusions

In the MIMIS project, the spread and settlement of marine invasive species was investigated with a modelling kit made of Lagrangian transport models and species distribution models. Guidance and explanation in this MIMIS Technical report is intended to help users develop their own studies on invasive marine species spread and/or adjusting existing models to their own needs. The use of two focal species illustrates the multiple steps needed to model the spread of a species and the modelling limits when the ecological knowledge is scarce (e.g., when a species is yet to be detected in a region).

5. References

Araújo, M. B., & New, M. (2007). Ensemble forecasting of species distributions. Trends in Ecology & Evolution, 22(1), 42-47. doi: 10.1016/j.tree.2006.09.010

Barbet-Massin, M., Jiguet, F., Albert, C. H., & Thuiller, W. (2012). Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution, 3*(2), 327-338. doi:10.1111/j.2041-210X.2011.00172.x

Castaños, C., Pascual, M., & Camacho, A. P. (2009). Reproductive biology of the nonnative oyster, *Crassostrea gigas* (Thunberg, 1793), as a key factor for its successful spread along the rocky shores of northern Patagonia, Argentina. *Journal of Shellfish Research*, 28(4), 837-847. doi:10.2983/035.028.0413

Cecino, G., & Treml, E. A. (2021). Local connections and the larval competency strongly influence marine metapopulation persistence. *Ecol Appl*, 31(4), e02302-e02302. doi:10.1002/eap.2302

Cetina-Heredia, P., Roughan, M., Liggins, G., Coleman, M. A., & Jeffs, A. (2019). Mesoscale circulation determines broad spatio-temporal settlement patterns of lobster. *PLoS One*, 14(2), e0211722. doi:10.1371/journal.pone.0211722

Cohen, J. H., Hanson, C. K., Dittel, A. I., Miller, D. C., & Tilburg, C. E. (2015). The ontogeny of larval swimming behavior in the *crab Hemigrapsus sanguineus*: Implications for larval transport. *Journal of Experimental Marine Biology and Ecology*, 462, 20-28. doi:10.1016/j.jembe.2014.10.003

Collet, B., Boudry, P., Thebault, A., Heurtebise, S., Morand, B., & Gérard, A. (1999). Relationship between pre- and post-metamorphic growth in the Pacific oyster *Crassostrea gigas* (Thunberg). *Aquaculture,* 175(3-4), 215-226. doi:10.1016/s0044-8486(99)00042-3

Connolly, S. R., & Baird, A. H. (2010). Estimating dispersal potential for marine larvae: dynamic models applied to scleractinian corals. *Ecology*, 91(12), 3572-3583. doi:10.1890/10-0143.1

Coon, S. L., Fitt, W. K., & Bonar, D. B. (1990). Competence and delay of metamorphosis in the Pacific oyster *Crassostrea gigas. Marine Biology*, 106(3), 379-387. doi:10.1007/BF01344316

Dabrowski, T., Lyons, K., Nolan, G., Berry, A., Cusack, C., & Silke, J. (2016). Harmful Algal Bloom warning system for SW Ireland. Part I: description and validation of an operational forecasting model. *Harmful Algae*, 53, 64-76.

Dagestad, K. F., Rohrs, J., Breivik, O., & Adlandsvik, B. (2018). OpenDrift v1.0: a generic framework for trajectory modelling. *Geosci Model Dev, 11*(4), 1405-1420. doi:10.5194/gmd-11-1405-2018

Dankers, N., Dijkman, E., de Jong, M., de Kort, G., & Meijboom, A. (2004). De verspreiding en uitbreiding van de Japanse Oester in de Nederlandse Waddenzee. *Alterra Report No 909* (Alterra, Wageningen). https://edepot.wur.nl/37096 (Accessed 24th Aug. 2022)

Dauvin, J.-C., Tous Rius, A., & Ruellet, T. (2009). Recent expansion of two invasive crabs species *Hemigrapsus sanguineus* (de Haan, 1835) and *H. takanoi* Asakura and Watanabe 2005 along the Opal Coast, France. *Aquatic Invasions*, 4(3), 451-465. doi:10.3391/ai.2009.4.3.3

Dukes, J. S., & Mooney, H. A. (1999). Does global change increase the success of biological invaders? *Trends in Ecology & Evolution, 14*(4), 135-139. doi:10.1016/S0169-5347(98)01554-7

EMODnet Bathymetry Consortium (2018) EMODnet Digital Bathymetry (DTM 2018). doi:10.12770/18ff0d48-b203-4a65-94a9-5fd8b0ec35f6 (Accessed 24th Aug. 2022)

EMODnet Geology, Seabed Substrate 1:100k. (Geological Survey of Finland GTK) Retrieved from: https://egdi.geology.cz/record/basic/a434a2bb3e860a1c5485124055c4458e7fdfe08b (Accessed 24th Aug. 2022)

Epifanio, C. E. (2013). Invasion biology of the Asian shore crab *Hemigrapsus sanguineus*: A review. *Journal of Experimental Marine Biology and Ecology*, 441, 33-49. doi: 10.1016/j.jembe.2013.01.010

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Ann. Statist.,* 29(5), 1189-1232. doi:10.1214/aos/1013203451

Geburzi, J. C. (2018). Full larval cycle and small-scale migration patterns of *H. takanoi* larvae in the recently invaded southwestern Baltic Sea. In New species from the Pacific establishment and dispersal of two invasive crabs (genus Hemigrapsus) in German coastal waters (pp. 67-94). Kiel University.

Giménez, L., Exton, M., Spitzner, F., Meth, R., Ecker, U., Jungblut, S., et al. (2020). Exploring larval phenology as predictor for range expansion in an invasive species. *Ecography*, 43(10), 1423-1434. doi: 10.1111/ecog.04725

Gothland, M., Dauvin, J. C., Denis, L., Dufossé, F., Jobert, S., Ovaert, J., et al. (2014). Biological traits explain the distribution and colonisation ability of the invasive shore crab *Hemigrapsus takanoi*. *Estuarine, Coastal and Shelf Science*, 142, 41-49. doi:10.1016/j.ecss.2014.03.012

Hartnett, M., Dabrowski, T., & Olbert, A. I. (2011). A new formula to calculate residence times of tidal waterbodies. *Proceedings of the Institution of Civil Engineers - Water Management*, 164(5), 243-256. doi:10.1680/wama.2011.164.5.243

Jordi A., Wang, & D.P. (2012) sbPOM: A parallel implementation of Princenton Ocean Model. *Environmental Modelling & Software* 38: 59–61. doi: 10.1016/j.envsoft.2012.05.013

Katsanevakis, S., Zenetos, A., Belchior, C., & Cardoso, A. C. (2013). Invading European Seas: Assessing pathways of introduction of marine aliens. *Ocean & Coastal Management, 76*, 64-74. doi:10.1016/j.ocecoaman.2013.02.024

Kochmann, J. (2012b). Documenting and predicting the spread of Pacific Oyster (*Crassostrea gigas*) in Ireland. (MSc). University College of Dublin

Landeira, J. M., Cuesta, J. A., & Tanaka, Y. (2019). Larval development of the brush-clawed shore crab *Hemigrapsus takanoi* Asakura & Watanabe, 2005 (Decapoda, Brachyura, Varunidae). *Journal of the Marine Biological Association of the United Kingdom*, 99(5), 1153-1164. doi:10.1017/s002531541900002x

Landeira, J. M., Liu, B., Omura, T., Akiba, T., & Tanaka, Y. (2020). Salinity effects on the first larval stage of the invasive crab *Hemigrapsus takanoi*: Survival and swimming patterns. *Estuarine, Coastal and Shelf Science*, 245, 106976. doi: 10.1016/j.ecss.2020.106976

Mann, R. (1979). Some biochemical and physiological aspects of growth and gametogenesis in *Crassostrea gigas* and *Ostrea edulis* grown at sustained elevated temperatures. *Journal of the Marine Biological Association of the United Kingdom*, 59(1), 95-110. doi:10.1017/s0025315400046208

Mao, X., Guo, X., Wang, Y., & Takayama, K. (2019). Influences of Global Warming on the Larval Survival and Transport of Snow Crab (*Chionoecetes opilio*) in the Sea of Japan. *Sustainability*, 11(8). doi:10.3390/su11082198

Marine Copernicus (2021), Ocean Physical and Biogeochemical Reanalysis (NWSHELF_MULTIYEAR_PHY_004_009, NWSHELF_MULTIYEAR_BGC_004_011), E.U. Copernicus Marine Service Information. doi: 10.48670/moi-00059 (Accessed 1st Oct. 2021).

Mills, S. R. A. (2016). Population structure and ecology of wild *Crassostrea gigas* (Thunberg, 1793) on the south coast of England. (PhD). Southampton,

Nagy, H., Lyons, K., Nolan, G., Cure, M., & Dabrowski, T. A. (2020a). Regional Operational Model for the North East Atlantic: Model Configuration and Validation. *J. Mar. Sci. Eng.,* 8, 673. doi :10.3390/jmse8090673

Nagy, H., Lyons, K., & Dabrowski, T. (2020b). A Regional Operational and Storm Surge Model for the Galway Bay: Model Configuration and Validation. *Earth and Space Science Open Archive*. doi: 10.1002/essoar.10502263.1

Naimi, B., & Araújo, M. B. (2016). sdm: a reproducible and extensible R platform for species distribution modelling. *Ecography*, 39(4), 368-375. doi: 10.1111/ecog.01881

NASA Goddard Space Flight Center, Ocean Ecology Laboratory, Ocean Biology Processing Group; (2014): MODIS-Aqua Ocean Color Data; NASA Goddard Space Flight Center, Ocean Ecology Laboratory, Ocean Biology Processing Group. doi: 10.5067/AQUA/MODIS_OC.2014.0 Accessed on 07/28/2015.

North, E. W., Vølstad, J. H., Christman, M. C., Hood, R. R., Zhong, L., Schlag, Z., et al. (2006). Linking larval transport and fisheries demographic models to study the influence of environmental variability and larval behavior on juvenile recruitment to oyster populations.

O'Connor, M. I., Bruno, J. F., Gaines, S. D., Halpern, B. S., Lester, S. E., Kinlan, B. P., & Weiss, J. M. (2007). Temperature control of larval dispersal and the implications for marine ecology, evolution, and conservation. *Proceedings of the National Academy of Sciences, 104*(4), 1266-1271. doi:10.1073/pnas.0603422104

Okubo, A. (1971). Oceanic diffusion diagrams. *Deep Sea Research and Oceanographic Abstracts*, 18(8), 789-802. doi: 10.1016/0011-7471(71)90046-5

Park, S., Epifanio, C. E., & Grey, E. K. (2004). Behavior of larval *Hemigrapsus sanguineus* (de Haan) in response to gravity and pressure. *Journal of Experimental Marine Biology and Ecology*, 307(2), 197-206. doi: doi.org/10.1016/j.jembe.2004.02.007

Park, S., Epifanio, C. E., & Iglay, R. B. (2005). Patterns of larval release by the asian shore crab *Hemigrapsus sanguineus* (de Haan): periodicity at diel and tidal frequencies. *Journal of Shellfish Research*, 24(2), 591-595. doi:10.2983/0730-8000(2005)24[591:POLRBT]2.0.CO;2

Pauley, G. B., Van Der Raay, B., & Troutt, D. (1988). Species profiles: life histories and environmental requirements of coastal fishes and invertebrates (Pacific Northwest): Pacific oyster. Biological Report - *US Fish & Wildlife Service*, 82(11), 28 pp.

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190(3), 231-259. doi: 10.1016/j.ecolmodel.2005.03.026

Robins, P. E., Tita, A., King, J. W., & Jenkins, S. R. (2017). Predicting the dispersal of wild Pacific oysters *Crassostrea gigas* (Thunberg, 1793) from an existing frontier population - a numerical study. *Aquatic Invasions*, 12(2), 117-131. doi:10.3391/ai.2017.12.2.01

Ruiz, C., Abad, M., Sedano, F., Garcia-Martin, L. O., & López, J. L. S. (1992). Influence of seasonal environmental changes on the gamete production and biochemical composition of *Crassostrea gigas* (Thunberg) in suspended culture in El Grove, Galicia, Spain. *Journal of Experimental Marine Biology and Ecology*, 155(2), 249-262. doi:10.1016/0022-0981(92)90066-j

Schlag, Z. R., & North, E. W. (2012). Lagrangian TRANSport model (LTRANS v.2) User's Guide. Cambridge, MD. https://northweb.hpl.umces.edu/LTRANS/LTRANS-v2/LTRANSv2_UsersGuide_6Jan12.pdf (Accessed on 24th Aug. 2022).

Simons, R. D., Siegel, D. A., & Brown, K. S. (2013). Model sensitivity and robustness in the estimation of larval transport: A study of particle tracking parameters. *Journal of Marine Systems*, 119–120, 19-29. doi: 10.1016/j.jmarsys.2013.03.004

Storlazzi, C. D., van Ormondt, M., Chen, Y.-L., & Elias, E. P. L. (2017). Modeling Fine-Scale Coral Larval Dispersal and Interisland Connectivity to Help Designate Mutually-Supporting Coral Reef Marine Protected Areas: Insights from Maui Nui, Hawaii. *Frontiers in Marine Science*, 4. doi:10.3389/fmars.2017.00381

Suquet, M., Le Mercier, A., Rimond, F., Mingant, C., Haffray, P., & Labbe, C. (2012). Setting tools for the early assessment of the quality of thawed Pacific oyster (*Crassostrea gigas*) D-larvae. *Theriogenology*, 78(2), 462-467. doi:10.1016/j.theriogenology.2012.02.014

Troost, K. (2010). Causes and effects of a highly successful marine invasion: Case-study of the introduced Pacific oyster *Crassostrea gigas* in continental NW European estuaries. *Journal of Sea Research*, 64(3), 145-165. doi:10.1016/j.seares.2010.02.004

Tully, O., & Clarke, S. (2012). The status and management of oyster (*Ostrea edulis*) in Ireland. *Irish Fisheries Investigations,* 24, Marine Institute. http://hdl.handle.net/10793/828 (Accessed on 24 Aug 2022).

Valavi, R., Elith, J., Lahoz-Monfort, J. J., & Guillera-Arroita, G. (2019). blockCV: An r package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods in Ecology and Evolution*, 10(2), 225-232. doi: 10.1111/2041-210X.13107

van den Brink, A., Godschalk, M., Smaal, A., Lindeboom, H., & McLay, C. (2012). Some like it hot: the effect of temperature on brood development in the invasive crab *Hemigrapsus takanoi* (Decapoda: Brachyura: Varunidae). *Journal of the Marine Biological Association of the United Kingdom*, 93(1), 189-196. doi:10.1017/s0025315412000446

Vikebø, F., Jørgensen, C., Kristiansen, T., & Fiksen, Ø. (2007). Drift, growth, and survival of larval Northeast Arctic cod with simple rules of behaviour. *Marine Ecology Progress Series*, 347, 207-219. doi:10.3354/meps06979

Wood, L. E., Silva, T. A. M., Heal, R., Kennerley, A., Stebbing, P., Fernand, L., & Tidbury, H. J. (2021). Unaided dispersal risk of *Magallana gigas* into and around the UK: combining particle tracking

modelling and environmental suitability scoring. *Biological Invasions*, 23(6), 1719-1738. doi:10.1007/s10530-021-02467-x

Annex A. Particle number and release frequency

Concept

Choosing an adequate number of particles for the particle tracking modelling is an important decision. Using the real number of larvae is often impractical or a waste of computational time. The alternative for modellers is to select a smaller number of particles that represents the distribution of larval dispersal scenarios to within some required tolerance. The approach is the same for selecting the particle release frequency.

Methodology

Based on Simons et al. (2013) methodology, we estimated a particle number and a release frequency for which the Fraction of Unexplained Variance (FUV) was smaller than 0.05. The FUV is a metrics produced by comparing the spatial larval dispersals and can inform about how similar or different two larval dispersals are.

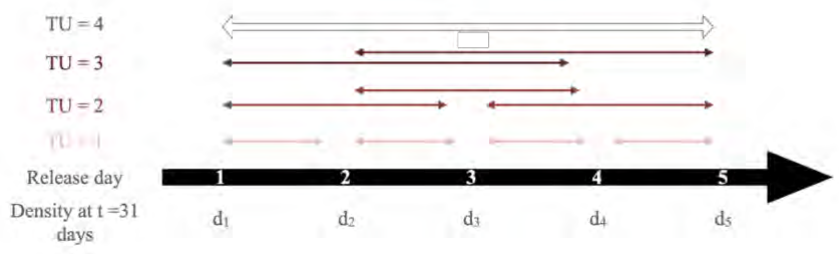
The FUV methodology for selecting a number n_i of particles has the following steps:

i) releasing and tracking a maximum number of particles N
ii) at a tracking time t, computing the spatial particle density D in a grid G
iii) at a tracking time t, randomly sampling the position of n_i particles out of the N particles and computing their spatial density d_i in the grid G
iv) Computing the Pearson coefficient of correlation r between D and d_i, and then the FUV with FUV = $1 - r^2$.
v) Repeating iii) and iv) 100 times and computing the average and standard deviation values of the 100 FUV for n_i particles.

The FUV methodology for selecting the final release frequency f_x of particles consists of getting the number of time units (TU) between releases that gives the lowest FUV. It is carried out with the following steps:

i) Consecutively releasing particles to the lowest frequency X_{min} over a period of time unit (P) and tracking the particles.
ii) at a tracking time t, computing the spatial particle density d_x in a grid G, x represents an index ranging between 1 and P.
iii) Computing the coefficient of correlation r_{TU} between d_x and d_{x+a}, and then the $FUV_{TU} = 1 - r_{TU}^2$, a is a value between 1 and $P-1$. This means that when $x+a$ increases, the frequency between releases lowers and TU between releases increases.
iv) Computing the average and standard deviation of the FUV for the same TU between releases (e.g., **Figure I.1**).

Figure I.1. Example of FUV computation between spatial particle densities d_x for the case where P = 5 days and X*min* = 1 day. The graph shows how the analysis is defined and the table shows the output of the four steps described in the text. The shading in the table gets darker when time units (TU) between releases increase



	d_1	d_2	d_3	d_4	d_5
d_1		FUV_1,	FUV_2	FUV_3	FUV_4
d_2			FUV_1	FUV_2	FUV_3
d_3				FUV_1	FUV_2
d_4					FUV_1
d_5					

Application. Number of crab larvae and release frequency

A total N of 19000 simulated passive larvae were released at one location in Galway Bay between July 11th and 30th , so 1000 larvae per night with high tide. Larvae were tracked for 31 days. Depth of release was near the bottom. The diffusivity coefficients were K_h = 0.1 m^2/s and K_v = 0.0001 m^2/s (see Annex III). The spatial extent of the grid G was -10.8 to -8.9º E and 52.95 to 53.73º N with a resolution of 0.003º along the longitude and 0.002º along the latitude. At a tracking time *t =31 days* and for each release between the 11th and 30th July (i.e., a period P of 19 days), we computed the FUV for 8 values of n_i. We also computed the FUV for 19 frequencies (*i.e., time units (TU) between releases are* 1 day to 18 days).

The results indicate that a number of particles higher than 999 would be enough to have the FUV from all the releases lower than 0.05 (**Figure I.2**). When studying the release frequency (**Figure I.3**), no FUV was lower than 0.05, which means than releasing larvae everyday (the highest possible release frequency) was needed to have a good representation of the dispersal variance

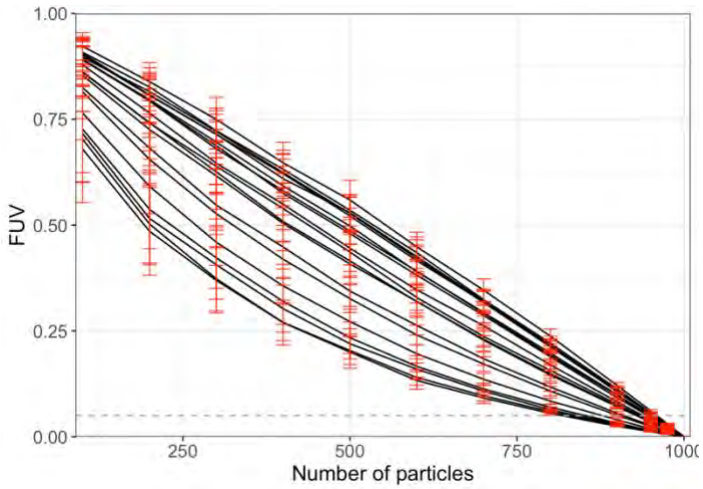


Figure I.2. Fraction of the Unexplained Variance with number of released particles. All continuous lines represent the FUV variation of one release day. The grey dashed line represents the threshold of 0.05.

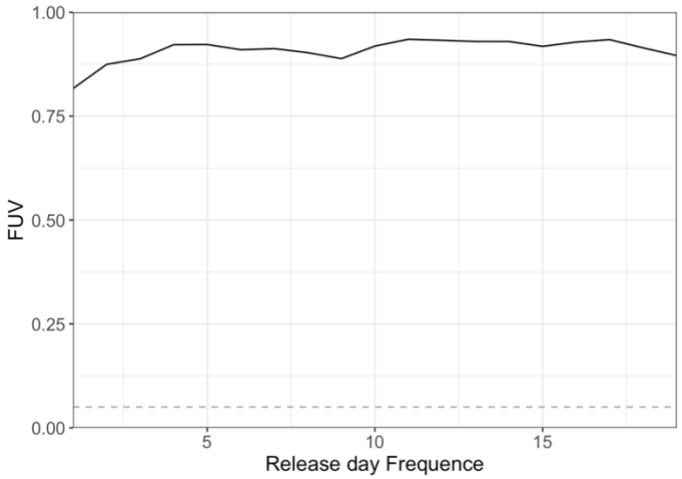


Figure I.3. Fraction of the Unexplained Variance with release frequency of particles. The grey dashed line represents the threshold of 0.05.

Annex B. Larval behaviour

Concept

Implementing a modification in the vertical position of a particle has an important impact on the dispersal. A general overview is that near-surface currents are enhanced by atmospheric forcings and the near-bottom currents are slowed due to contact and friction with the seafloor bottom. Between the surface and the bottom, the vertical profile of the currents can heterogeneous or homogenous. For example, in strongly mixed water, the water circulation is the same at all depths. Both the circulation mechanisms of a studied area and an active displacement of the particle in the water column have substantial impacts on the dispersal. For larvae, their active displacement in the vertical dimension is dependent on larval behaviour.

Methodology

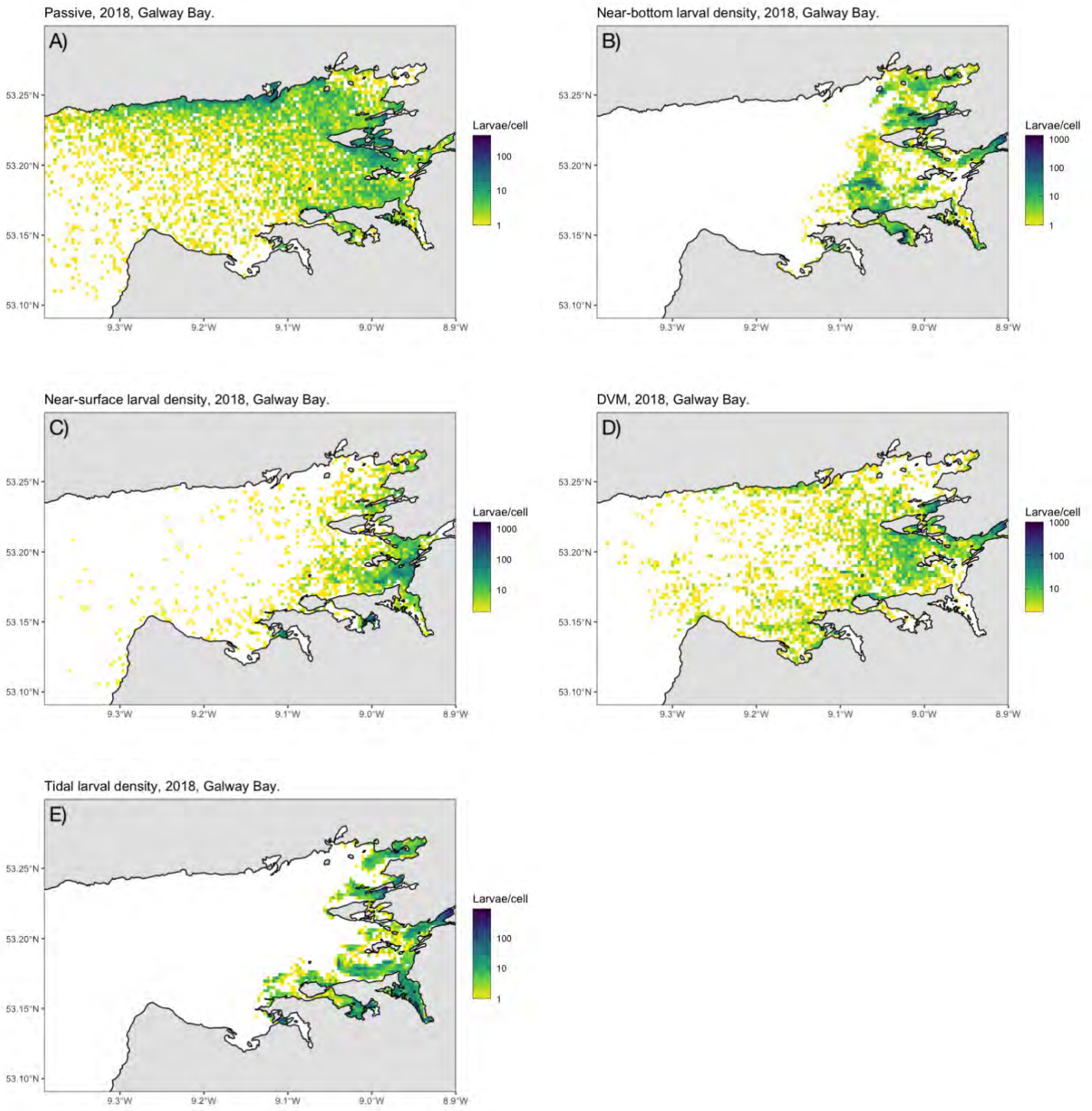
Simulated oyster larvae are released at 32 locations in Galway Bay and are tracked from the day that corresponded to oyster gonad maturation and spawning for each particular location. Larvae are tracked for 31 days. Five scenarios of larval behaviours are implemented:

- S1: Inactive simulated larvae (passive advection)
- S2: Near-bottom particles. Eighty percent of the particles are repositioned near the bottom if their depth is more than 1 m above the bottom. The remaining 20% stays at the same depth.
- S3: Near-surface particles. Eight percent of the particles are repositioned near the surface if their depth is less than 1 m deep. The remaining 20% stays at the same depth.
- S4: Particles with phototaxis (e.g., diel vertical migration, DVM). Particles are repositioned near the surface during nighttime and near the bottom during the daytime.
- S5: Tidally cued particles. Particles are repositioned near the surface during the flood tide and near the bottom during the ebb tide.

In addition to mapping the density distribution of the larvae, which provide a spatial overview of the impact of behaviour on the dispersal, the distance of the particle from the release sites and the mean squared displacement, a surface area, of particles from the same release site are computed.

Results

The use of different behavioural scenarios reveals distinct patterns of larval dispersal. On the one hand, simulated passive larvae and larvae with phototaxis (e.g., DVM) disperse across the whole of Galway Bay (**Map II.1**). On the other hand, tidal behaviour is the scenario with the most retentive larval transport, which consequently results in aggregated larvae within different inlets of Galway Bay. Surprisingly, larvae being kept at the surface are less dispersed in Galway Bay than expected and could accumulate in the south-eastern section of Galway Bay. On the opposite, larvae being kept near the bottom could be transported at distant places from the release sites such as the center of Galway Bay and in the southwestern part of the bay.



Map II.1. Dispersal of larvae with different behaviour classified from the most dispersive (top panel) to the less dispersive (bottom panel). One cell represents 0.06 km^2.

Simulated larvae travelled the furthest (10 km on average from their source sites) and dispersed over the largest surface area (~93 km^2) when they had no behaviour (i.e., passive, **Figure II.1, Table II.1**). Simulated larvae travelled the least (2.8 km) and dispersed over a smallest surface area (7.5 km^2) when they interacted with the tide. Even if the spatial distribution of simulated larvae with a diel vertical migration seems to cover Galway Bay, the larvae under this scenario did not spread (5.3 km and 18.4 km^2) as much as passive simulated larvae.

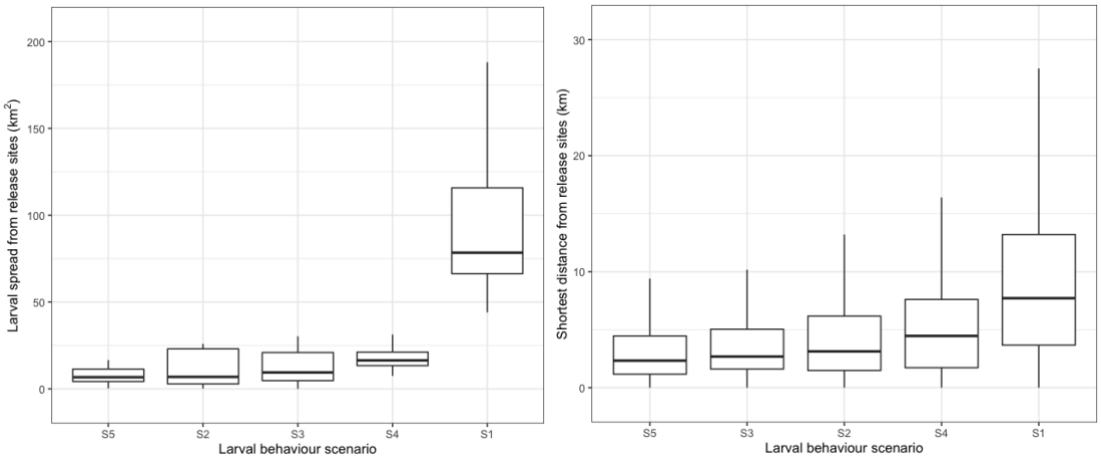


Figure II.1. Larval transport metrics distribution per scenarios using distance from source (in km, left panel) and mean spread dispersal (km^2, right panel).

Table I.1. Statistics on the distance and mean displacement surface area of simulated larval transport from four simulated larval behaviour. S1: Passive simulated larvae, S2: Near surface simulated larvae, S3: Near bottom simulated larvae, S4: Simulated larvae with phototaxis (DVM), S5: Tidally cued simulated larvae. Avg.: Average, SD: Standard deviation.

Scenario		**S1**	**S2**	**S3**	**S4**	**S5**
Distance (km)	Avg.	10.2	4.11	3.90	5.40	3.04
	SD	8.85	3.41	4.13	4.49	2.82
	Max	74.8	79.3	62.6	78.0	68.5
Surface area (km^2)	Avg.	95.8	12.5	15.7	19.0	7.79
	SD	47.7	10.2	17.1	8.95	5.37
	Max	264	25.9	77.1	37.3	24.9

Annex C. Turbulence diffusion

Concept

Turbulent diffusion is a chaotic and unstable displacement of the water body which can be accounted for in the hydrodynamic fields thanks to turbulent schemes. Sometimes, for lack of storage space, coefficients of diffusivity are not stored in the modelled hydrodynamic fields. In Lagrangian tracking models, the diffusivity coefficient can be given as a constant. To estimate the most fitted values for the studied area, a comparative study can be used to analyse the correlation between the Okubo's diffusive equation (Eq. III.1) and the estimated relationship of particle spread with time.

Eq. III.1. $$MSD(t) = 0.0108 \times t^{2.34}$$

where *MSD(t)* is the mean squared displacement of a particle after t simulated hours relative to the average position of all simulated particles.

Methodology

The relationship of particle spread with time is a relation estimated after several particle transport simulations were run. Each of the simulations uses a unique combination of horizontal and vertical diffusivity coefficient values. The diffusive transport of the particles in the horizontal (Δx) and vertical (Δz) dimensions during a time step Δt is computed by two random walk schemes Eq. III.2 and Eq. III.3, respectively:

Eq. III.2. $$\Delta x = R * \sqrt{\frac{2}{r}.K_h.\Delta t}$$

Eq. III.3. $$\Delta z = K'(z).\Delta t + R * \sqrt{\frac{2}{r}.K_v.(z + 0.5.K'(z)).\Delta t},$$ [7]

Where *R* is a random value from a Gaussian distribution *G(0,r)* with mean of zero and standard coefficient *r* (a constant of 1); K_h and K_v, the horizontal and vertical coefficients of diffusion; z, the vertical particle position at the previous time and $K'(z) = \Delta K/\Delta z$, represents the gradient of diffusivity over a small step in the vertical dimension.

We initialise the dispersal of 1000 particles per simulation and track them for 30 days. We run scenarios of particle dispersals using 4 release sites and two release depths: near-bottom or near-surface. The diffusivity coefficients are combined such as K_h is between 0 and 5 m^2/s and K_v being between 0 and 6 x 10^{-4} m^2/s. A total of 65 simulations are obtained.

Then, the mean squared displacement *MSD* is hourly computed for each simulation:

Eq. III.4. $$MSD(t) = \frac{1}{N} \times \sum_i^N d\{(\underline{X_t}, \underline{Y_t}), (Xi_t, Yi_t)\},$$

With *d{A,B}*, the distance between two coordinates using the spherical law of cosines; *N*, the number of particles released, $\underline{X_t}, \underline{Y_t}$, the center of the mass of *N* particles at time *t*, and Xi_t, Yi_t the *i-th* particle position at time *t*.

To establish the relationship between *MSD* and time *t*, we log-transform (base 10) the time and the MSD and trained a linear model such as:

Eq. III.5. $\log(MSD) \sim A + B * \log(t)$.

The fitted coefficients *A* and *B from Eq.III.5* would be equivalent of the coefficients 0.0108 and 2.34, respectively, from Okubo's diffusive equation (Eq. III.1).

[7] Formula provided in the LTRANS file "vert_turb_module.f90", adapted from Visser et al., 1997.

Then, we analyse the effect of the diffusivity coefficient combinations, the release location, and the release depth on the MSD.

Results

The horizontal and vertical diffusivity coefficients of 0.1 m^2/s and 4 x 10^{-4} m^2/s, respectively, seems to be graphically the closer to the Okubo's equation (**Figure III.1.C**). Indeed, the mean squared displacement seemed to be less divergent through the duration of simulations than the other combinations of diffusivity coefficients. The estimated intercept and coefficient (A and B in **Table III.1**) are close to Okubo's values.

We note that the MSD is physically close to Okubo's equation (i.e., blue lines in Figure III) when only vertical diffusivity (VV in **Figure III.I.D**) is accounted for than when using the combination of the two diffusivities (HV in **Figure III.I.D**), which is also perceived with the values of the coefficient and exponent B closer to 2.34 (eq.III.1, **Table III.1)**.

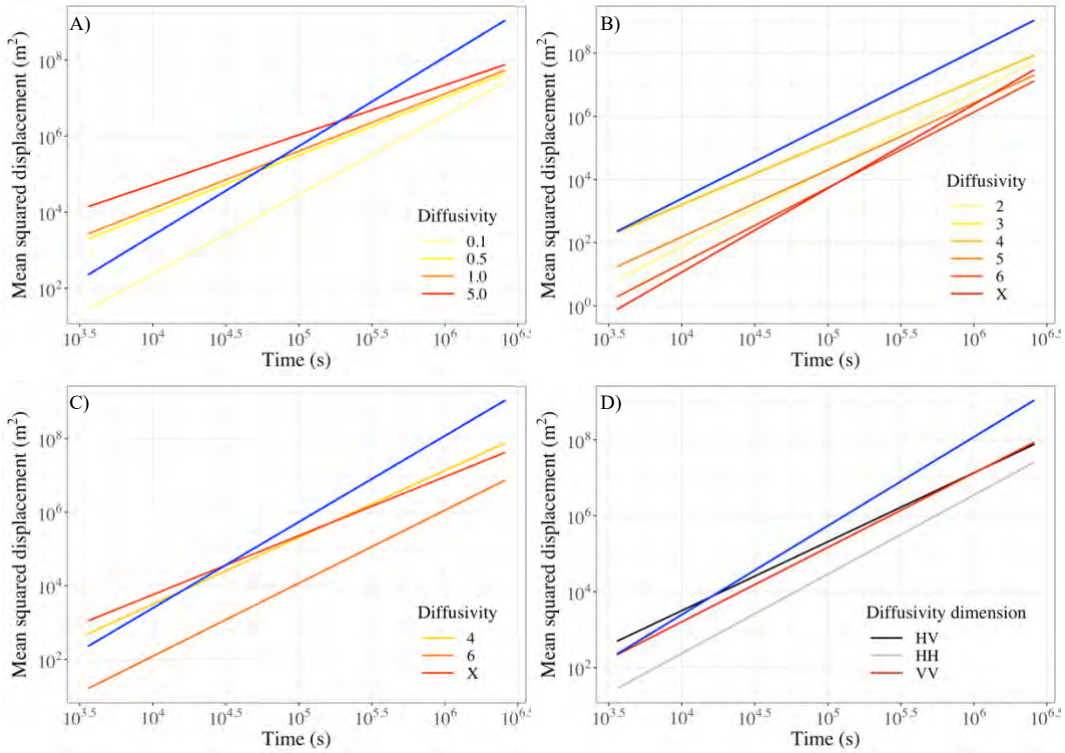


Figure III.1. Spread of the particles according to Okubo's equation (blue lines) and according to the particle dispersal simulations with various combinations of diffusivity coefficients: A) Only horizontal diffusivity, B) only vertical diffusivity, C) horizontal diffusivity of 0.1 m^2/s and vertical diffusivity, and D) three combinations with and without vertical diffusivity at 4 x 10^{-4} m^2/s and horizontal diffusivity coefficient at 0.1 m^2/s. X is a vertical diffusivity coefficient extracted from the ROMS model.

Table III.1. Values of the diffusivity coefficient (vertical K_h and horizontal K_v in m^2/s) used in the Lagrangian tracking models and estimated coefficient (B) and intercept (A) of Okubo's equation. Red shade indicates the coefficients kept in the dispersal modelling.

Coefficients	K_h	K_v (*10^{-4})	A	B	R^2
H	0.5	0	0.0079	1.51	98
	0.1	0	$1.07*10^{-6}$	2.08	94
	1	0	0.01	1.51	98
	5	0	0.31	1.31	98
V	0	2	$1.58*10^{-8}$	2.42	98
	0	3	$2.23*10^{-5}$	1.96	98
	0	4	$2.56*10^{-5}$	1.95	95
	0	5	$4.90*10^{-7}$	2.12	92
	0	6	$5.90*10^{-9}$	2.39	93
	0	X	$2.76*10^{-10}$	2.65	90
HV	0.1	4	0.00017	1.81	99
	0.1	X	0.0022	1.6	97
	0.1	6	$1.45*10^{-6}$	1.98	92
	0.5	4	$2.48*10^{-5}$	1.91	96

The sites of release may potentially influence the MSD, but over time, the differences decreased (**Figure III.2**). Only the mean squared displacement from releases at site 1 has a graphically important difference, which can be associated with the hydrodynamics of the area around site 1.

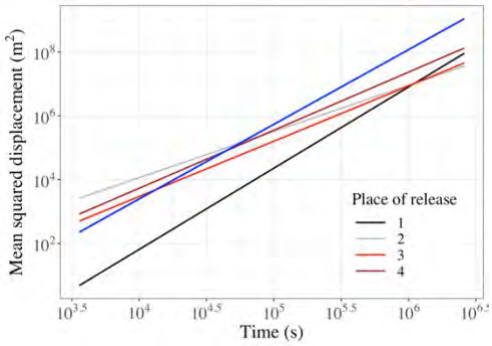


Figure III.2. Spread of the particles according to Okubo's equation (blue line) and according to the particle dispersal simulations with horizontal and vertical coefficient of diffusivity of 0.1 m^2/s and 4 x 10^{-4} m^2/s, respectively from four release sites: 1 at 53.12364 ºN and 9.15362 ºW; 2 at 53.16414 ºN and 8.966936 ºW; 3 at 53.20242 ºN and 8.941351 ºW; and 4 at 53.22808 ºN and 9.008547 ºW.

The depth of a particle's release has no influence the mean squared displacement, probably due to the shallow location of the release sites. (**Figure III.3**).

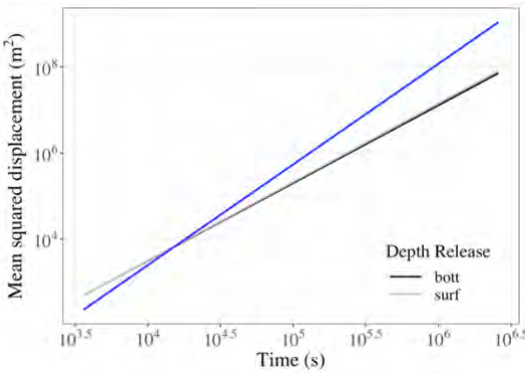


Figure III.3. Spread of the particles according to Okubo's equation (blue line) and according to the particle dispersal simulations with horizontal and vertical coefficient of diffusivity of 0.1 m^2/s and 4 x 10^{-4} m^2/s, respectively from two release depths.

Conclusion

Both vertical and horizontal diffusivity coefficients are important in Lagrangian transport modelling although the correlation between the estimated mean squared displacement and Okubo's results is more relevant with simulations done with only vertical diffusivity. This analysis shows the difficulty in selecting values relevant for the whole study area. Finally, we select the values of 0.1 m^2/s and 4 x 10^{-4} m^2/s for vertical and horizontal diffusivity coefficients, respectively.

Reference

Okubo, A. (1971). Oceanic diffusion diagrams. *Deep Sea Research and Oceanographic Abstracts, 18*(8), 789-802. Doi: 10.1016/0011-7471(71)90046-5

Visser, A. (1997). Using random walk models to simulate the vertical distribution of particles in a turbulent water column. *Marine Ecology – Progress Series, 158*, 275-281. Doi:10.3354/meps158275

Annex D. Hydrodynamic time step resolution

Concept

Hydrodynamic fields are recorded on a selected temporal scale that fits with the physics of the studied area. The temporal scale may be of importance for Lagrangian transport simulations. The ROMS in the CONNEMARA region (middle western side of Ireland, **Map 1**) has a 3-hour time-resolution between 2014 and 2016 and an hourly time-resolution between 2017 and 2020. The question was about whether a 3-hour time resolution is sufficient for the Lagrangian transport simulations, or whether the hourly time resolution is required.

Methodology

We compare the differences in the particle dispersals that are simulated in hydrodynamic fields with two different time resolutions. Subsetting the hourly hydrodynamic fields from year 2018, we get a 3-hour time resolution version of a hydrodynamic model. Then, we run the Lagrangian transport model by releasing 12800 particles from 32 locations and they are hourly advected for 31 days. We keep hourly tracks of their position during that period.

Three simulated larval behaviours of vertical movement are implemented:

- S1: the particles stay passive.
- S2: the particles are displaced near the bottom during ebb tide and near the surface during the flood tide.
- S3: the particles displace near the surface during night-time and near the floor during daytime.

In total, we run 6 simulations of larval dispersal.

Then, we compute the centre of the mass (CM) of particles positions (i.e., the average coordinates in a particle cloud), for each release location at each recorded time, and for the six simulations. For each scenario, we calculate the distance between CMs from simulations done whith hourly and 3-hour hydrodynamic fields. Similarly, we calculate the mean squared displacement of the particles (i.e., the sum of the average distance between the particle coordinates and the CM) with time and calculate the ratio of the mean squared displacement (MSD) at 1 hour to the mean spread at 3 hours.

Results

Overall, the time resolution of the hydrodynamic fields influenced the particle dispersals (**Figure IV.1**) and using a hourly hydrodynamic model lowered the errors. The time resolution impacted the average position of the particles and, additionally, it increased the discrepancy between the average positions through the duration of particle tracking (to 2 km with S1, to ~ 3.5 km with S2 and S3). The MSD of the particle also fluctuated with the duration of the particle tracking. The MSD stayed constant with S1 (~1.4), gradually became 5 times bigger with S2, and it peaked to 5 and 15 with S3.

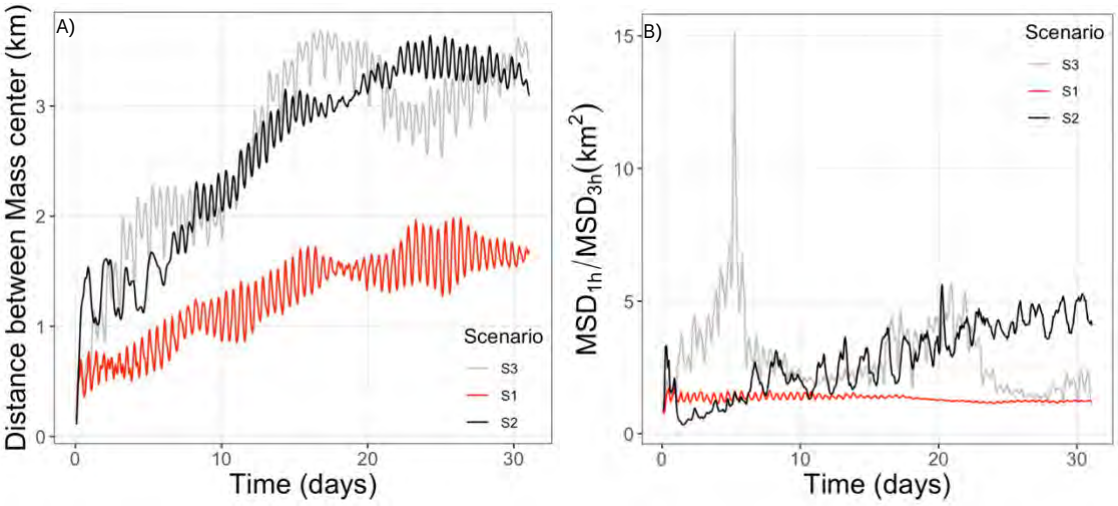


Figure IV.1. Temporal evolution of the distance between mass centre (A) and the ratio of the larval spread (B).

The two metrics for the simulation under S1 had the smallest differences compared to tidally impacted particles (S2) and light-driven particles (S3).

Conclusion

At Galway Bay scale (30 km x 10 km), the changes in the position and spread of the larvae because of the hydrodynamic field time resolution were important (up to 3.5 km and 5 times bigger, respectively). The behaviour of particles increased these changes. To simulate the larval dispersal of coastal species, we selected hydrodynamic fields with the highest temporal scale, here 1 hour.

Annex E. Presence/Absence coverage

Concept

Species distribution models were built with datasets made of four different presence/absence setups to test whether the size of the dataset, the spatial extent, and the balance presence/absence ratio could influence the results.

Methods

Two datasets gather oyster occurrences from UK and Irish waters ("UK&I") and European waters ("EU"). To the "EU" dataset, we added mussel occurrences from the mainland European waters to get a third dataset ("EU_plus"). A fourth dataset was obtained by removing occurrences recorded in the Irish water zone from EU_Plus ("EU_minus"). Presence of mussels with no oyster observations was used as a proxy for *'Absence'* of oyster observations.

A priori, the datasets "EU_plus" and "EU_minus" have variables with similar ranges (**Figure V.1**). Datasets "EU" and "UK&I" have variables with smaller (e.g., pH) or similar (e.g., PIC with "UK&I") ranges than/as "EU_plus" and "EU_minus".

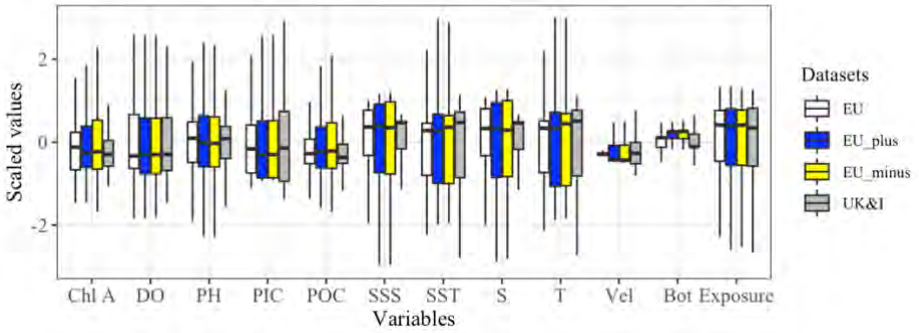


Figure V.1. Variable ranges in the four datasets. Boxplots mostly inform about the median (central bar) and the first and third quartiles (hinges), and the 95th quantiles.

A statistical approach using a mean comparison test and an analysis of variances corroborates the fact that variables and their range are different (Table 9). Only the differences in the mean and variance of the Wave Exposure (WaveE) were no significant, which can be interpreted in the fact that oyster and mussels are species living in areas protected from strong wave exposure.

Table V.1. A summary of the mean comparison test among the four datasets. *; p-values > 0.05, -; p-values < 0.05.

	CHL	DOx	PHx	Vel	PIC	POC	SSS	SST	Txx	Sxx-	WaveE	Bot2
EU	-	-	-	-	-	-	-	-	-	-	-	*
EU_plus	-	-	-	-	-	*	-	-	-	-	*	-
EU_minus	-	-	-	-	-	*	-	-	-	-	*	-
UK&I	*	-	-	-	*	*	-	*	*	-	*	*

All independent variables are centered and scaled after removals of unrealistic spatial occurrences (i.e., land locations). Averages and standard deviations from the scaling are recorded in a separate file for later uses when preparing the predictive dataset of independent variables. Eighty percent of the dataset is used to train the model. The remaining 20% is used to evaluate the model performance and to produce performance statistics (*AUC*, *kappa,* and *Specificity* metrics).

We build the SDM by combining two types of statistical model: Maxent and Boosted Regression Trees with the *sdm* R package (Friedman, 2001; Naimi and Araújo, 2016; Phillips *et al.*, 2006). Once the SDMs are built, prediction of oyster occurrences in the Irish coastal waters is carried out. A dataset gathers the gridded

independent variables extending from 10.66 to 5.20º W and from 51.36 to 55.45º N with a high spatial resolution (x = 0.005463871º, y = 0.002408357º). The independent variables dataset used for prediction are scaled by the average and standard deviation extracted for each dataset. The last step consists of combining the prediction from BRT and Maxent models (i.e., ensemble modelling; Araújo and New, 2007) by averaging their estimations.

Results

The performances of the eight trained models are all acceptable (**Table V.1**). No model stood out based on the performance metrics, and this was underlined as well with the ROC curves (**Figure V.2**).

Table V.2. Information about species distribution model parametrisations and performances. The sorting of variable importance is based on AUC decreases when the variable is removed from the model. Performance is evaluated with independent data (the testing dataset) and selected metrics (kappa, AUC, Specificity).

Datasets	**Model parameterisation**	**Variable Importance**	**Performance**		
"UK&I" Total occurrences: 736, Occurrences for model training: 588 Ratio Presence: 66%	BRT: 500 trees Shrinkage: 0.01 Depth: 6 **Maxent**: Betamultiplier: 2	DOx, PIC, Txx, Sxx and WaveE		**BRT**	**Maxent**
			AUC	0.93	0.90
			Kappa	0.69	0.63
			Spec.	0.91	0.89
"EU" Total occurrences: 1716, Occurrences for model training: 1372 Ratio Presence: 85%	**BRT**: 1300 trees Shrinkage: 0.01 Depth: 6 **Maxent**: Betamultiplier: 2	PHx, CHL, Txx, WaveE		**BRT**	**Maxent**
			AUC	0.97	0.86
			Kappa	0.72	0.47
			Spec.	0.95	0.78
"EU_plus" Total occurrences: 2938, Occurrences for model training : 2350 Ratio Presence: 50%	**BRT:** 1300 trees Shrinkage: 0.01 Depth: 6 **Maxent:** Betamultiplier: 2	WaveE, Vel, SST, CHL		**BRT**	**Maxent**
			AUC	0.90	0.87
			Kappa	0.63	0.55
			Spec.	0.76	0.80
"EU_minus" Total occurrences: 2751, Occurrences for model training: 2200 Ratio Presence: 48%	**BRT**: 1300 trees Shrinkage: 0.01 Depth: 6 **Maxent**: Betamultiplier: 2	CHL, WaveE, Vel, POC, DOx		**BRT**	**Maxent**
			AUC	0.91	0.87
			Kappa	0.65	0.57
			Spec.	0.75	0.83

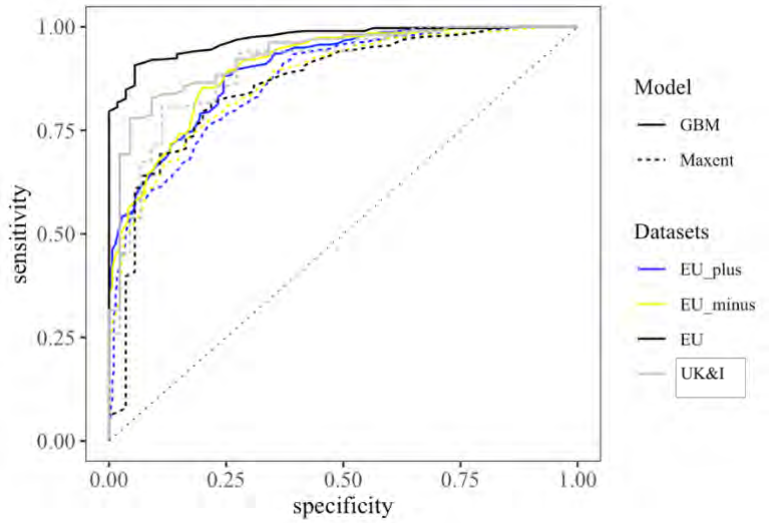
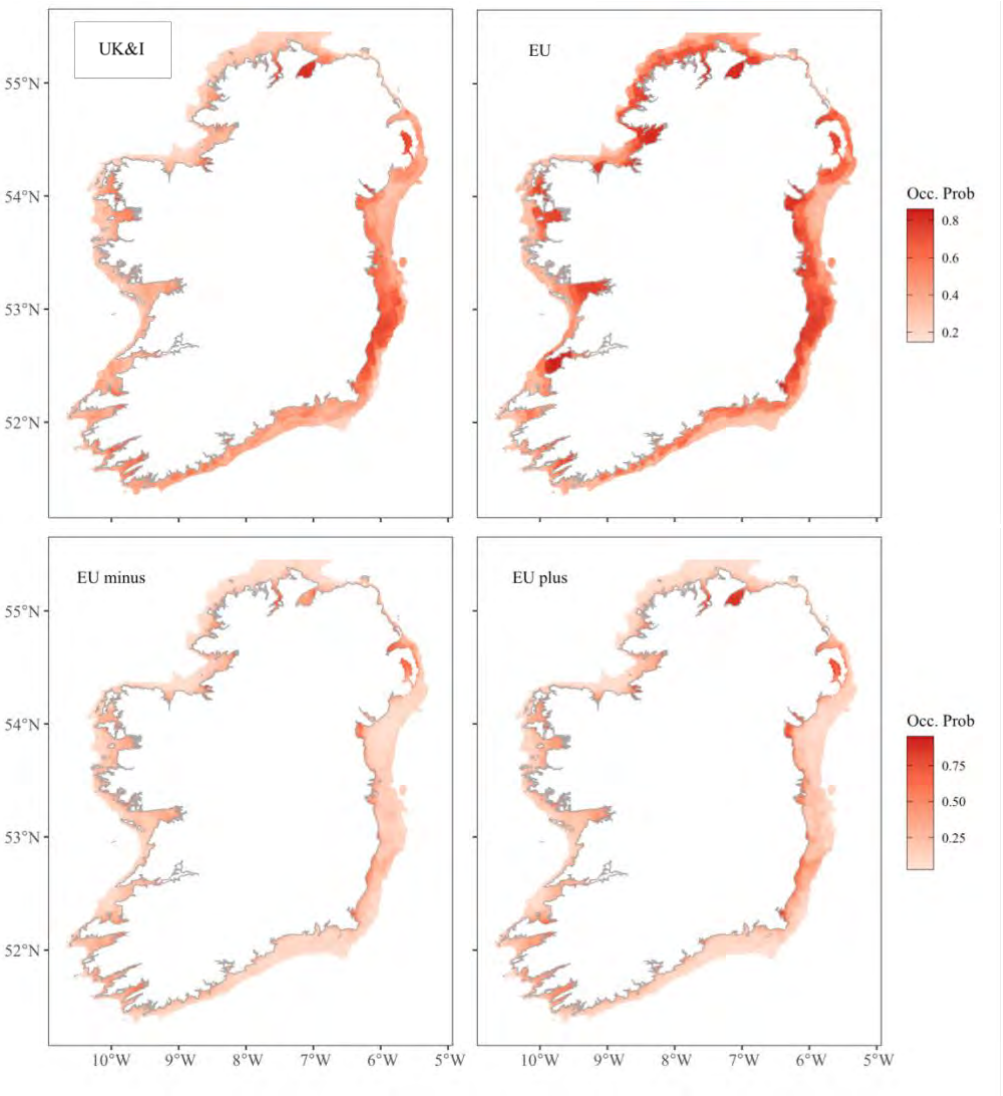


Figure V.2. Receiver operating characteristic (ROC) curves estimated from models built on different datasets and model type.

Maps showed small spatial variations in the predictions of occurrence probability related to the model and the dataset (**Figure V.3**). Estimated distribution of the occurrence probability was similar if the models were built on i) the "UK&I" and "EU" datasets, and ii) "EU_minus" and "EU_plus". In all maps, occurrence probability was high (> 0.7) in semi-enclosed area like Lough Silly (Northern Ireland). High values of occurrence probability, which were predicted with the "UK&I" and "EU" datasets, were distributed over a wider bottom range as seen over the eastern Irish continental shelf. This pattern was also observed in the northern Irish continental shelf when building a model with the "EU" dataset.



FigureV.3. Prediction of Oyster occurrence probability with models based on the "UK&I" (top left panel), the "EU" (top right panel), the "EU_minus" (bottom left panel), and the "EU_plus" (bottom right panel) datasets.

Conclusion

The prediction of occurrence suggested that with ("EU_plus") or without ("EU_minus") including observations of occurrences from the Irish coastal water, the model had acceptable performance. The analysis presented here briefly shows the importance of having a large enough spatial extent of the observations and a balanced ratio of presences and pseudo-absences. The study on *H. sanguineus* distribution carried out a more detailed sensitivity analysis over the dataset size and the ratio of presences (see Ultan's MSc thesis). During the analyses of the larval dispersal simulations, the model trained with the "EU_minus" dataset was used as a GIS layer indicating the suitable places for larval settlement and potential adult distribution.

Reference

Araújo, M. B., & New, M. (2007). Ensemble forecasting of species distributions. Trends in Ecology & Evolution, 22(1), 42-47. doi: 10.1016/j.tree.2006.09.010

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Ann. Statist.,* 29(5), 1189-1232. doi:10.1214/aos/1013203451

Naimi, B., & Araújo, M. B. (2016). sdm: a reproducible and extensible R platform for species distribution modelling. *Ecography*, 39(4), 368-375. doi: 10.1111/ecog.01881

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190(3), 231-259. doi: 10.1016/j.ecolmodel.2005.03.026

Further details available on www.emff.marine.ie

Managing Authority EMFF 2014-2020	Specified Public Beneficiary Body
Department of Agriculture Food & the Marine Clogheen, Clonakilty, Co. Cork. P85 TX47 Tel: (+)353 (0)23 885 9500 www.agriculture.gov.ie/emff	Marine Institute Rinville, Oranmore, Co. Galway, H91 R673 Phone: (+)353 (0)91 38 7200 www.marine.ie

This project or operation is part supported by the Irish government and the European Maritime & Fisheries Fund
as part of the EMFF Operational Programme for 2014-2020

Ireland's EU Structural and Investment Funds Programmes 2014 - 2020
Co-funded by the Irish Government and the European Union

An Roinn Talmhaíochta, Bia agus Mara
Department of Agriculture, Food and the Marine

EUROPEAN UNION
This measure is part-financed by the European Maritime and Fisheries Fund

Marine Institute
Foras na Mara