

Supplementary Information of:
The risk of marine bioinvasion caused by global
shipping

Hanno Seebens¹, Michael T. Gastner^{2,3}, Bernd Blasius^{1*}

¹Institute for Chemistry and Biology of the Marine Environment,
Carl-von-Ossietzky University, Oldenburg, Germany

²Department of Mathematics, Imperial College London, UK

³Department of Engineering Mathematics, University of Bristol, UK

*To whom correspondence should be addressed; E-mail: blasius@icbm.de.

Contents

1 Model formulation	2
2 Model assumptions	4
3 Model variants and model selection	7
4 References	10
5 Figures and tables	13

1 Model formulation

The invasion model estimates the probability of new primary invasions for every port call of a vessel from a large pool of non-native species from all previous stopover sites. Consider a ship that sails from donor port i to a recipient port j along a route r of successive stopover sites (Fig. S3). The model accounts for the fact that biological invasions constitute a dynamic multi-stage process [1]. A species must pass three distinct invasion transitions, each describing a different, important aspect of the invasion process, before it can successfully invade the recipient port along route r : (i) the species must be non-native in the destination port, (ii) it must be introduced alive into the new habitat, and (iii) it must establish a self-sustaining population.

(i) The invasion process begins with propagules being loaded by ballast water in their native range. Not all of those organisms play a role for invasion as ships can introduce species which are already present in the new locations. To incorporate this in our model, we introduce the probability to be non-native $P_{ij}(\text{Alien})$ which describes the likelihood that a native species at donor port i is non-native at recipient port j . We estimate this probability by the biogeographical dissimilarity between the ecological communities at donor and recipient ports. Adopting well-established functional relationships about the distance dependence of ecological similarity [2, 3, 4, 5], we assume that this dissimilarity increases with the geographic distance d_{ij} (obtained by calculating the great circle distance between native and invaded ports). However, ecological dissimilarity does not increase with distance at a constant rate. Instead, following common concepts of biogeography, the species composition within a defined marine ecoregion can be expected to be rather homogeneous [6, 7]. Thus, the amount of non-native species should be disproportionately low in the direct vicinity of a port. This gives rise to a sigmoidal dependence of biogeographic dissimilarity with distance which effectively suppresses the invasion risk at short distances. We found that the model results are largely independent of the specific analytic form which is used to model this relationship. In this study, we use the following formula

$$P_{ij}(\text{Alien}) = \left(1 + \frac{\gamma}{d_{ij}}\right)^{-\beta}. \quad (\text{S1})$$

Here, β is a shape parameter (increasing β makes the functional dependence on the distance d_{ij} more sigmoidal) and γ describes the characteristic geographic scale over which species composition changes at a disproportional low rate. However, very similar results are achieved by using, for instance,

$$P_{ij}(\text{Alien}) = \begin{cases} 0 & , d_{ij} < \gamma \\ [1 - e^{-\beta(d_{ij}-\gamma)}] & , d_{ij} \geq \gamma. \end{cases} \quad (\text{S2})$$

Note that in our model $P_{ij}(\text{Alien})$ depends only on the geographic distance d_{ij} between ports, but not on the route r itself.

(ii) The second invasion stage is described by the probability of introduction $P_r(\text{Intro})$. Non-native species must be introduced alive into the new habitat, whereas most organisms perish during the transport to a new locale. For most species, ballast tanks are extreme habitats owing to the lack of light, varying temperature, salinity, and oxygen concentrations, but also due to the ship's rolling that imposes mechanical stress on inoculated species. Field studies have shown that population densities in ballast water tanks decline approximately exponentially with time [8, 9, 10, 11]. Thus, the probability of introduction most notably accounts for the survival of species entrained in ballast tanks and is modeled as an exponentially decaying function of travel time Δt_r on route r with mortality rate μ . Further, the probability of introduction increases with the amount of released ballast water B_r originating from port i . B_r is calculated as $B_r = zW_r(1 - zW_r/V_r)^{\delta_r}$, where W_r denotes the amount of released ballast water in m^3 calculated from regression fits shown in Fig. S1, V_r is the mean volume of ballast tanks of a ship in m^3 depending on the size of the ship [12], δ_r represents the number of intermediate stopovers on route r , and z is the fraction of zero releases depending on the type and the size of a ship (Fig. S1). $1 - zW_r/V_r$ is therefore the fraction of ballast water remaining in the ballast tanks. Finally, the probability of introduction is reduced by the factor ρ_r to incorporate the potential suppression of invasion risks by ballast water treatment. This factor can be set specific to different ship movements, ports, or ship types and thus allows modeling different ballast water treatment scenarios. If, for example, ballast water is treated equally at every port call of the ship and ρ describes the fraction of species left in the ballast water after a single treatment, the risk reduction over the full trajectory is given by $\rho_r = \rho^{\delta_r+1}$. Putting everything together, the probability of introduction is described as

$$P_r(\text{Intro}) = \rho_r (1 - e^{-\lambda B_r}) e^{-\mu \Delta t_r} \quad (\text{S3})$$

with λ being a characteristic constant. In contrast to $P_{ij}(\text{Alien})$, the probability of introduction $P_r(\text{Intro})$ depends on the full history of the ship route r .

(iii) Once released in the recipient port, the introduced organisms must establish a self-sustaining population which is described by the probability of establishment $P_{ij}(\text{Estab})$. The chances for a non-native species to survive and establish a population depend on multiple abiotic and biotic factors at the recipient region. The abiotic conditions included in the model are the most dominant environmental factors for marine organisms: water temperature and salinity. We assume that $P_{ij}(\text{Estab})$ increases with the environmental similarity between recipient and donor ports as a Gaussian function of the differences of water temperature ΔT_{ij} and salinity ΔS_{ij}

$$P_{ij}(\text{Estab}) = \alpha e^{-\frac{1}{2} \left[\left(\frac{\Delta T_{ij}}{\sigma_T} \right)^2 + \left(\frac{\Delta S_{ij}}{\sigma_S} \right)^2 \right]}. \quad (\text{S4})$$

Here, σ_T and σ_S represent the width of the species' ecological niches. Note that the probability of establishment depends only on the environmental conditions at source and destination ports, but not on the route r .

The full invasion probability along the route r is obtained by combining the three probabilities for the specific invasion stages. Our numerical calculations show that these three probabilities are only weakly correlated (Fig. S4). Assuming statistical independence, we can calculate the probability of new invasions originating from one particular ship movement r by taking the product $P_{ij}(\text{Alien})P_r(\text{Intro})P_{ij}(\text{Estab})$. However, as ships are calling many ports, species released from ballast water exchange may originate not only from the last port of call (direct link), but also from all ports previously called by the ship (indirect links). Therefore, the total invasion risk between a pair of ports can only be calculated by accounting for the full history of the shipping trajectory (Fig. S3). The total invasion risk $P_{ij}(\text{Inv})$ from port i to j (i.e., the probability of observing at least one invasion during the study period of 2007-2008) is then obtained by merging the invasion probability over every possible shipping trajectory r of every ship in the network that is connecting donor port i to recipient port j either directly or via intermediate ports. In practice, this is calculated as the complement of species failing to invade on all shipping routes r in the set $R(i, j)$ of all (possibly indirect) routes from i to j :

$$P_{ij}(\text{Inv}) = 1 - \Pi_{r \in R(i,j)} [1 - P_{ij}(\text{Alien})P_r(\text{Intro})P_{ij}(\text{Estab})]. \quad (\text{S5})$$

The risk for a new invasion at port j is obtained by aggregating invasion risks in a similar way over all shipping connections passing through j

$$P_j(\text{Inv}) = 1 - \prod_i [1 - P_{ij}(\text{Inv})]. \quad (\text{S6})$$

Aggregated invasion risks can be obtained by merging invasion probabilities in a similar way over different ships, ship types (Fig. S5), ports (Fig. S6), geographic regions, or ecosystems (Fig. 3). The expected number of new establishments in port j in a time period τ can be calculated as $N_j = -\log(1 - P_j(\text{Inv})) \cdot \tau/\tau_0$, with the duration of the study period $\tau_0 = 2$ years. The setup also allows us to resolve the temporal dynamics of invasion risks such as risk development during the voyage of a ship (Fig. S7) or the seasonal dynamics of ecoregions (Fig. S8).

2 Model assumptions

In the following list, we present and briefly discuss the main model assumptions connected to each of the three probabilities (Eq. S1, Eq. S3, and Eq. S4).

P(Alien):

1. *The dissimilarity of ecological communities increases with geographic distance.*

It seems safe to assume that two communities located nearby are on average more similar in their species composition than two communities which are separated by greater distance. This pattern is also supported by various studies [2, 3, 4, 5].

2. *The dissimilarity does not increase at a constant rate with geographic distance.*

At small distances, say within the range of an ecoregion, the species composition can be assumed to be more or less homogeneous [6, 7] and thus the dissimilarity of two communities changes at a disproportionately low rate compared to larger distances. At large distances, the dissimilarity should saturate at distances when on average two communities do not share any species. In this case, a further increase in distance does not increase the dissimilarity. This gives rise to a sigmoid function as described by Eq. S1.

P(Intro):

1. *Population densities decrease exponentially during transportation in ballast tanks.*

Many field studies have shown that in general this assumption holds [8, 9, 10, 11] although some of these studies also highlight that it is possible that populations remain stable or even increase during the travel.

2. *A ship that enters a port is assumed to exchange ballast water.*

This assumption has to be made as no data about the actual ballast water management of ships are currently available. We acknowledge that many ships do not exchange ballast water. As shown in Fig. S1 we consider this in the model by reducing the amount of exchanged ballast water by a characteristic fraction of zero-releases taken from field data (see “Model formulation”). For a single ship this averaging may lead to substantial errors in the estimated amount of exchanged ballast water. However, the invasion risk between a pair of ports is calculated from the average over many ship movements (the top 100 000 high risk routes are connected on average by 432 ships movements) which seems to be adequate for the assumption of a mean ballast water flow.

3. *Ballast water is (dis)charged only in, or within the vicinity of, a port.*

This and the following assumptions 4-5 have to be made due to the lack of more detailed data. A ship can charge or discharge ballast water at almost any time and, for example, may adjust the total ballast water weight depending on weather conditions. However no data about these exchanges is available yet. Therefore, the safest assumption is to assume that ballast water is charged/discharged when cargo is exchanged or when a ship enters calmer waters within the vicinity of a port. We would like to note that this is a limitation of current data availability but not of the model itself as more detailed data, if available, can easily be incorporated in the model.

4. *The same ship releases the same amount of ballast water at each port call.*

Field data show that this assumption is violated. This is also indicated by the large scatter of released ballast water in Fig. S1. The effect of this variability is, however,

reduced the more ships travel the same link (see comment 2 of $P(\text{Intro})$). This assumption also implies that the transport of ballast water is not directed, which in reality is wrong for many ships. For example tankers and bulkers often transport cargo in one direction and ballast water in the other (when empty). This assumption might give systematic errors in risk estimates, however due to the lack of data, this limitation can not be resolved yet.

5. *Ships of the same size and the same type release the same amount of ballast water.*
See comment to the previous assumption.
6. *Ballast water exchange increases as a power law with ship size.*
Although there is clearly a large variation of the amount of released ballast water (Fig. S1), this assumption seems to hold on average. See also comment 4 of $P(\text{Intro})$.
7. *Ballast water is homogeneously mixed throughout all ballast tanks of the same ship.*
This assumption is clearly violated, but a necessary simplification due to the lack of detailed data. See also comment 4 of $P(\text{Intro})$.
8. *Species are introduced only by ballast water exchange.*
Ballast water exchange may be the most important vector for marine invasions, but it is definitely not the only invasion vector. Other invasion vectors, such as hull-fouling and the release from aquaculture, also play an important role. We have explored the role of hull-fouling as an additional invasion vector in our model variant 5 (see below).

P(Estab):

1. *Environmental conditions are mainly characterized by temperature and salinity.*
There are definitely many more environmental conditions that play a role for the establishment of a species, but for marine invasions in general surface water temperature and salinity are the most important ones, relevant for almost all species.
2. *The chance to establish depends only on the match of environmental conditions.*
This assumption implies that all species whose requirements match those of the port of their introduction are able to establish a population, meaning that biotic interactions do not play any role at the new locale. This assumption is clearly a strong simplification. Biotic interactions are, however, difficult to model in a non-species-specific approach. One possibility, as done in model version 2 (see “Model variants and model selection” below), is to use the species richness in the recipient port as a surrogate for the biotic pressure imposed on the introduced species. This biotic resistance hypothesis [13] is, however, controversially discussed and empirical evidence that supports this hypothesis is weak.

3 Model variants and model selection

The model presented in this study was selected out of a set of different model versions and represents the model which fitted field data of marine invasions best. In the following, we present these model versions in the form of modifications of Eqs. S1, S3, and S4, as well as the fitting procedure.

Model version 1:

Ship-specific ballast water, no species richness (standard model)

In this version, each ship is assigned a ship-size specific and ship-type specific release of ballast water calculated from ballast water release protocols obtained from NBIC [14] as described in chapter “Model formulation”. In all model variants except model version 6 ballast water releases are considered in this way. This model version represents the standard model used in this study and is described in detail in chapter “Model formulation” of this supplement and in the main text. Species richness is not included.

Model version 2:

Ship-specific ballast water, species richness in recipient port

Species richness at the recipient port can be considered as a surrogate for the biotic pressure imposed on newly introduced species in the new habitat as stated by the biotic resistance hypothesis [13]. According to the hypothesis, the invasion risk should decrease with increasing species richness in the recipient port, which is included in this model version in the form of an additional factor $e^{-\eta R_j}$ in Eq. S4

$$P_{ij}(Estab) = \alpha e^{-\eta R_j} e^{-\frac{1}{2} \left[\left(\frac{\Delta T_{ij}}{\sigma_T} \right)^2 + \left(\frac{\Delta S_{ij}}{\sigma_S} \right)^2 \right]}.$$

Here, R_j is the species richness in the recipient port j and η is a characteristic constant. Species richness data of marine ecoregions were taken from Tittensor et al. [15]. The authors provided numbers of marine species across taxonomic levels for each 880-km equal-area grid at a global resolution. From these data, we selected taxonomic groups likely to be transported by ballast water (i.e., planktonic species, coastal fishes, corals, and sea grasses) and assigned each port the species richness R_j of the grid it is located in. We also tested other combinations of taxonomic groups but did not find a significant change of results.

Model version 3:

Ship-specific ballast water, species richness in donor port

Species richness can also be considered as a measure for the number of non-native species with the potential to establish a population in the donor habitat. In this model variant

we assume that the probability of invasion increases with the species richness in the donor region. This is modelled by including the term $(1 - e^{-\chi R_i})$ in Eq. S3

$$P_r(\text{Intro}) = \rho_r (1 - e^{-\chi R_i}) (1 - e^{-\lambda B_r}) e^{-\mu \Delta t_r}.$$

Here R_i is the species richness in the donor port and χ is a characteristic constant. Data of species richness are described in paragraph “Model version 2”.

Model version 4:

Ship-specific ballast water, species richness in donor and recipient port

In this model version we combine model variants 2 and 3 and include species richness twice, in the donor and in the recipient region, as described above.

Model version 5:

Ship-specific ballast water, hull-fouling

In this model variant we consider biofouling as an additional vector of ship associated invasions. A recent study emphasizes the role of the sailing speed and the duration the ship stays in a port for a successful introduction of species attached to ship hulls [16]. The sailing speed is not known for our data, however, the time the ship spent in a port can be included in the following way

$$\hat{P}_r(\text{Intro}) = \nu P_r(\text{Intro}) + (1 - \nu)(1 - e^{-\omega D_r}),$$

where D_r denotes the duration a ship stayed in the last ports on route r and ω is a constant. In this equation the first term describes the probability of introduction by ballast water and is calculated by Eq. S3, while the second term describes the probability of introduction via hull fouling. Note that we do not consider a mortality loss with travel time for organisms attached to the hull of a ship. The constant ν describes the mixing of both invasion vectors. This is a simplistic approach to incorporate biofouling and it may be possible to establish other models additionally considering e.g. the environmental conditions during the journey. However, the lack of information about the exact travel routes and sailing speeds hinders a more sophisticated approach. Species richness is not considered.

Model version 6:

Simplified ballast water release model

This is the simplest model variant. We assume that each vessel has the same ballast water capacity, independent of ship size and ship type, and that a constant fraction ϵ of ballast water is discharged at each port of call. To include this into the model we assume that the probability of introduction scales with the fraction ϵ of released ballast water at the

recipient port, and is reduced by the dilution factor $(1 - \epsilon)$ to account for ballast water exchange at each intermediate stop at the δ_r ports visited on the route r . This give rise to the following equation (compare to Eq. S3)

$$P_r(\text{Intro}) = \rho_r \epsilon (1 - \epsilon)^{\delta_r} e^{-\mu \Delta t_r}.$$

Species richness is not included.

Model selection

All six model versions were fitted to data of reported invasions in marine habitats. We selected studies which provide information about the source regions from which the species were likely to be introduced and preferably studies which distinguished the vector of introduction. We included four case studies of highly invaded sites (North Sea, Europe [17], Pearl Harbor, Hawaii [18], Port Phillip Bay, Australia [19], and San Francisco Bay, USA [20]) and one study providing numbers of exotic species for all ecoregions of the world [21]. The simplest model version which fitted the data best was selected as the standard model presented in this study.

Invasion risks calculated by the model cannot be directly compared to the number of reported invasions because the model describes the risk of invasion during the study period of two years while the reported invasion events result from a decades long history of introductions. We therefore calculated the expected number of new establishments in port j in a time period τ as $N_{inv} = -\log(1 - P_{inv}) \cdot \tau / \tau_0$, with the duration of the study period $\tau_0 = 2$ years. N_{inv} was then used to fit the model to data.

The fit between model data x_i and field data y_i was calculated by the root mean square error, $\text{RMSE} = \sqrt{\sum_{i=1}^N (x_i - y_i)^2 / N}$, with N being the number of values to be fitted. The RMSE was calculated for data of each of the five studies separately and then averaged. To provide a meaningful scale of N_{inv} , we assumed that on average one new species is able to establish in an ecoregion per year as reported for some sites [17, 18, 19, 20, 22].

We applied an optimization algorithm (simulated annealing [23]) to find the combination of parameter values which yield the best fit of the reported data (i.e., the lowest RMSE). The parameter optimization was run for each model version independently. The ranking of the best-fitting models revealed that the RMSE is lowest for model version (2) incorporating species richness in the recipient port (Table S4). The main effect of including species richness in the recipient region is that invasion risks at ports located in areas of high species richness, i.e., the tropics, are slightly reduced. This yields a marginal better overall fit of the reported invasion risks world-wide. However, the differences of the RMSE between the model versions was very small, i.e., the RMSE of the fourth best model versions differed by less than 2.5% from the top fitting model, and thus these models fit the data almost equally well. We therefore decided to use model version (1) which is the simplest of the best-fitting models and ignores species richness.

The set of model parameters used in this study is informed by the results obtained from the parameter optimization, except for σ_S and μ which differed distinctly from values reasonable for the invasion process. The species tolerance level for salinity differences, σ_S , reached very large values during the optimization process, rendering the influence of salinity on the invasion risk very low. This may be explained by the fact that our risk estimations are based on data sets of surface water salinities that characterize the conditions in ports (that may contain almost fresh water), whereas the field studies used for model fitting provide numbers of reported non-native species in the vicinity of a port which is usually marine. This does not imply that salinity conditions are redundant in the model, but highlights that samplings of non-native species in ports are required for a more reliable model validation. The optimal value for the mortality rate in the ballast water tank during transport, μ , selected by the simulated annealing was more than a magnitude lower than those reported in the literature [11, 10]. We therefore set $\mu = 0.02$ representing the measured median of mortality rates for various taxa in ballast tanks without any ballast water treatment [11].

4 References

- [1] Mack, R.N. *et al.* (2000) Biotic invasions: Causes, epidemiology, global consequences, and control. *Ecol. Appl.* 10:689–710.
- [2] Nekola, J.C. & White, P.S. (1999) The distance decay of similarity in biogeography and ecology. *J. Biogeogr.* 26:867–878.
- [3] Tuomisto, H., Ruokolainen, K. & Yli-Halla, M. (2003) Dispersal, environment, and floristic variation of western Amazonian forests. *Science* 299:241–244.
- [4] Soininen, J., McDonald, R. & Hillebrand, H. (2007) The distance decay of similarity in ecological communities. *Ecography* 30:3–12.
- [5] Thielges, D.W. *et al.* (2009) Distance decay of similarity among parasite communities of three marine invertebrate hosts. *Oecologia* 160:163–173.
- [6] Hempel, G. & Sherman, K. (eds) (2003) *Large Marine Ecosystems of the World: Trends in Exploitation, Protection, and Research* (Elsevier Science).
- [7] Spalding, M.D. *et al.* (2007) Marine Ecoregions of the World : A Bioregionalization of Coastal and Shelf Areas. *BioScience* 57:573–583.

- [8] Gollasch, S., Lenz, J., Dammer, M. & Andres, H.G. (2000) Survival of tropical ballast water organisms during a cruise from the Indian Ocean to the North Sea. *J. Plankton Res.* 22:923–937.
- [9] Olenin, S., Gollasch, S., Jonušas, S. & Rimkut, I. (2000) En-Route Investigations of Plankton in Ballast Water on a Ship's Voyage from the Baltic Sea to the Open Atlantic Coast of Europe. *Int. Rev. Hydrobiol.* 85:577–596.
- [10] Wonham, M.J., Walton, W.C., Ruiz, G.M., Frese, A.M. & Galil, B.S. (2001) Going to the source: role of the invasion pathway in determining potential invaders. *Mar. Ecol. Prog. Ser.* 215:1–12.
- [11] Wonham, M.J. & Carlton, J.T. (2005) Trends in marine biological invasions at local and regional scales : the Northeast Pacific Ocean as a model system. *Biol. Invasions* 7:369–392.
- [12] ABS (American Bureau of Shipping, 2011). Ballast Water Treatment Advisory. Available from <http://www.eagle.org>.
- [13] Kennedy, T.A. et al. (2002) Biodiversity as a barrier to ecological invasion. *Nature* 417:636–8.
- [14] National Ballast Information Clearinghouse (NBIC) Online Database. Electronic publication, Smithsonian Environmental Research Center & United States Coast Guard, 2008. Available from <http://invasions.si.edu/nbic/search.html>; searched at 2012-10-16.
- [15] Tittensor, D.P., Mora, C., Jetz, W., Lotze, H.K., Ricard, D., Berghe, E.V., et al. (2010). Global patterns and predictors of marine biodiversity across taxa. *Nature* 466:1098–101. Data available at <http://www.mathstat.dal.ca/~derekt/publications>.
- [16] Sylvester, F. & MacIsaac, H.J. (2010) Is vessel hull fouling an invasion threat to the Great Lakes? *Div. Distr.* 16:132–143.
- [17] Reise, K., Gollasch, S. & Wolff, W.J. (1998). Introduced marine species of the North Sea coasts. *Helgoland Mar. Res.* 52:219–234.
- [18] Coles, S.L., DeFelice, R.C., Eldredge, L.G. & Carlton, J.T. (1999). Historical and recent introductions of non-indigenous marine species into Pearl Harbor, Oahu, Hawaiian Islands. *Mar. Biol.* 135:147–158.
- [19] Hewitt, C.L., Campbell, M.L., Thresher, R.E., Martin, R.B., Boyd, S., Cohen, B.F., et al. (2004). Introduced and cryptogenic species in Port Phillip Bay, Victoria, Australia. *Mar. Biol.* 144:183–202.

- [20] Cohen, A.N. & Carlton, J.T. (1998). Accelerating invasion rate in a highly invaded estuary. *Science* 279:555–558.
- [21] Molnar, J.L., Gamboa, R.L., Revenga, C. & Spalding, M.D. (2008). Assessing the global threat of invasive species to marine biodiversity. *Front. Ecol. Environ.* 6:485–492.
- [22] Strayer, D.L. (2012). Eight questions about invasions and ecosystem functioning. *Ecol. Lett.* 15:1199–1210.
- [23] Kirkpatrick, S., Gelatt, C.D. Jr., Vecchi, M.P. (1983). Optimization by simulated annealing. *Science* 220:671–680.
- [24] Keller R.P., Drake J.M., Drew M.B. & Lodge D.M. (2010) Linking environmental conditions and ship movements to estimate invasive species transport across the global shipping network. *Divers Distrib* 17:93–102.
- [25] Kaluza P., Kölzsch A., Gastner M.T. & Blasius B. (2010) The complex network of global cargo ship movements. *J Royal Soc Interface* 7:1093–1103.

5 Figures and tables

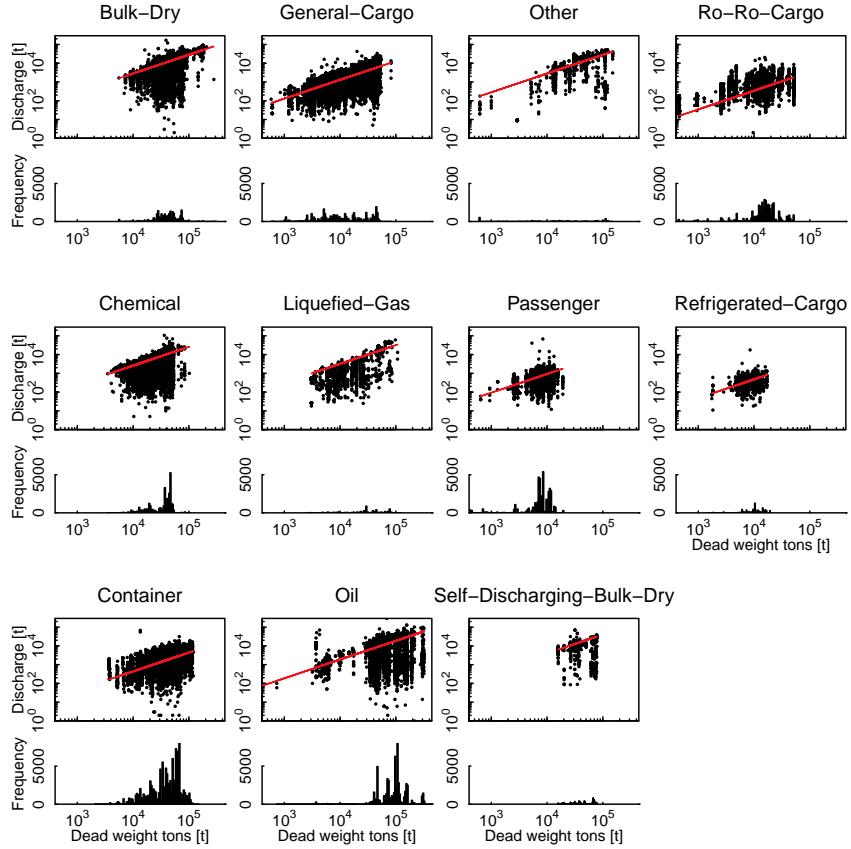


Fig. S1. Comparison of ballast water discharge and ship size. The plot shows regression fits of ship sizes and ballast water discharge volumes for various ship types (scatter plots) and histograms of port calls without discharges (histograms). Data of ballast water discharge volumes were taken from the National Ballast Information Clearinghouse [14] for those ships common to our data base (42%). Altogether 717 250 releases during 2004-01-01 to 2012-07-12 were used with an average of 65 205 releases for each ship type. The scatter plots show the non-zero discharges ($> 0 \text{ m}^3$) which were used to calculate linear regression fits between ship sizes and discharge volumes on a double-logarithmic scale. With the slopes of the regression fits we calculated for each ship in our data base a type-specific and size-specific ballast water release (see also “Methods” in the main text). In 42-88 % of all port calls, depending on the ship type, no ballast water is released at all, which is shown in the histograms. The fraction of these zero-discharges were included separately into the model (see section “Model formulation” in this supplement).

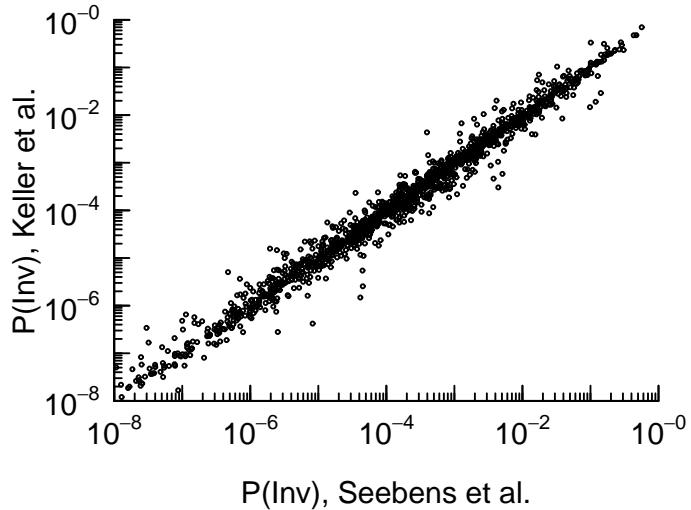


Fig. S2. Comparison of calculated port invasion risks based on two different data sets for port salinities. The characterization of port environmental conditions by means of salinity is hampered by the strong natural variability in many ports. To test the robustness of our results, we compared the salinities used in this study with salinity data provided by another study [24] for a total of 1321 common ports (90%) and found a correlation of both data sets of Spearman's $\rho = 0.71$. This deviation might be explained by the coarse resolution of the WOA, false predictions of a model used to determine salinities, and unexplained large values provided by Lloyd's Register Fairplay. However, when we run our invasion model with the salinity data of [24], we found only small differences to our original results as shown in the figure, i.e., the estimated port invasion risks based on either of the two salinity data sets show a strong correlation (Spearman's $\rho = 0.99$). This indicates that our results are robust to variations in the environmental conditions.

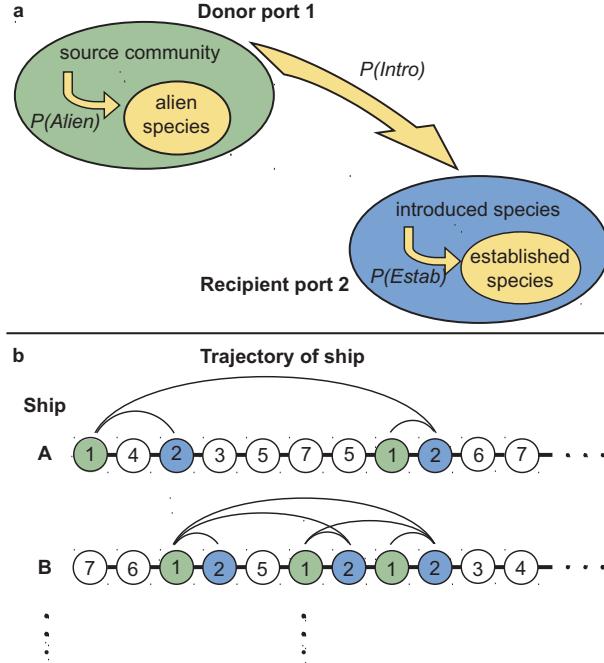


Fig. S3. Schematic illustration of the risk model. (a) To describe the invasion risk from a source community at donor port i to a recipient community at port j along a ship route r we calculate three probabilities: (1) the probability of a species in the source community to be non-native in the recipient community, $P_{ij}(\text{Alien})$, (2) the probability to be introduced alive in the new locale, $P_r(\text{Intro})$, and (3) the probability to establish a population, $P_{ij}(\text{Estab})$. The product of the three probabilities, $P_{ij}(\text{Alien})P_r(\text{Intro})P_{ij}(\text{Estab})$, measures the invasion probability resulting from that particular ship movement r . (b) This procedure is repeated for every realized ship movement r (indicated as arcs) of every ship (A, B, \dots) in the network connecting i and j . The figure, for example, indicates all realized connections between donor port 1 (green circles) to recipient port 2 (blue circles) either directly or via intermediate ports (3, 4, ..., white circles) along the route. The full invasion probability $P_{ij}(\text{Inv})$ is obtained by combining the invasion probabilities of all realized ship movements (Eq. S5).

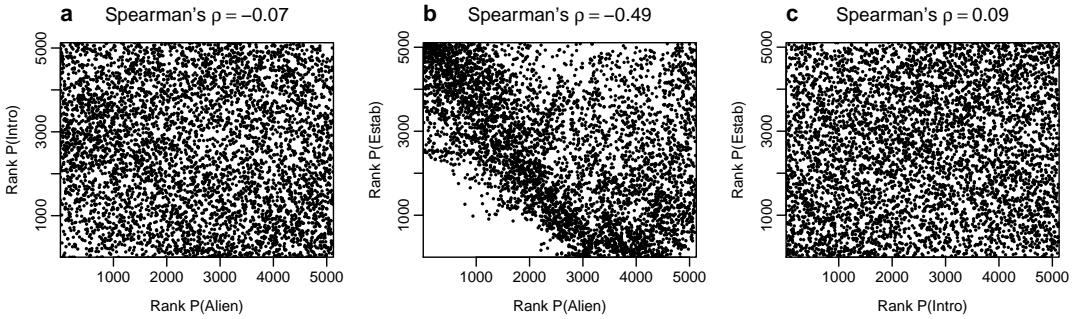


Fig. S4. Correlations of the probabilities $P(\text{Alien})$, $P(\text{Intro})$, and $P(\text{Estab})$. The figure shows the rank of (a) $P(\text{Alien})$ vs. $P(\text{Intro})$, (b) $P(\text{Alien})$ vs. $P(\text{Estab})$, and (c) $P(\text{Intro})$ vs. $P(\text{Estab})$ calculated for a randomly chosen subsample (1% of the data) of trajectories in the network. Spearman's ρ was calculated using all port connections ($n = 512\,603$). The figure shows that the probabilities on average can be regarded to be uncorrelated, except for the combination of small $P(\text{Alien})$ and $P(\text{Estab})$ (lower left corner in b). This can be explained as follows: pairs of ports with small values of $P(\text{Alien})$ have small geographic distance and thus usually a high environmental similarity, resulting in large values of $P(\text{Estab})$. Vice-versa, with increasing geographic distance (i.e., increasing $P(\text{Alien})$) in general the similarity of their environmental conditions is reduced (i.e., decreasing $P(\text{Estab})$), which resulted in the observed pattern in (b). This correlation is the source of a systematic error in the risk calculation of closely located ports because then $P(\text{Alien})$ and $P(\text{Estab})$ are not independent. However, this effect plays no role for the invasion dynamics because due to the small $P(\text{Alien})$ invasion risks between those ports can be neglected. We found that the correlation vanishes for the top 100 000 high risk links (Spearman's $\rho = -0.17$).

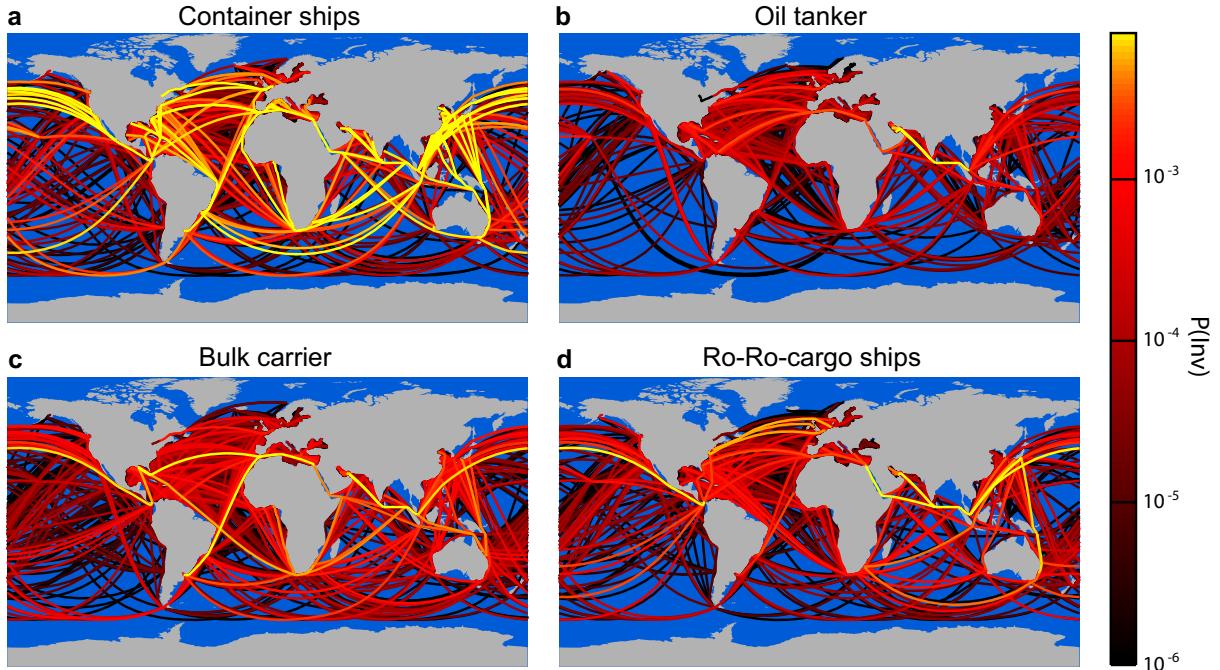


Fig. S5. Invasion risk of ship routes specific to ship types. The global shipping network consists of multiple layers each representing the movements of a different ship type with specific mobility patterns [25]. These translate into specific global patterns of invasion risks. The figure shows the invasion risks (similar to Fig. 2A) restricted to four specific ship types: (a) Container ships, (b) oil tankers, (c) bulk carriers, and (d) ro-ro-cargo ships. The high-risk routes clearly vary among ship types and mainly reflect the trading regimes between which the ships operate. The network of container ships, for example, has a high diversity of high-risk routes connecting ports all over the world. The invasion highways of oil tankers, on the other hand, are restricted to a few high-risk routes. These routes are mostly connecting the main oil exporting regions in the world to sites of similar climatic regimes. The highest invasion risk from bulk carriers emerges between countries of high raw material deposits and industrial nations while the high-risk routes of ro-ro-ships connect industrial nations reflecting the exchange of produced products. Shipping routes are calculated as the shortest distance between two ports. Color-coding indicates the risk level.

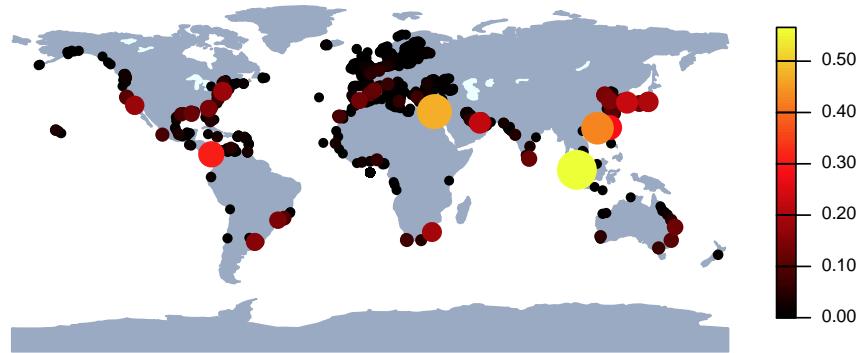


Fig. S6. Hot spots of bioinvasion. Total invasion risk $P_j(\text{Inv})$ of every port in the data base indicated by the color and the size of the dots. A ranking of the top 20 invasion hot spots is provided in Table S1.

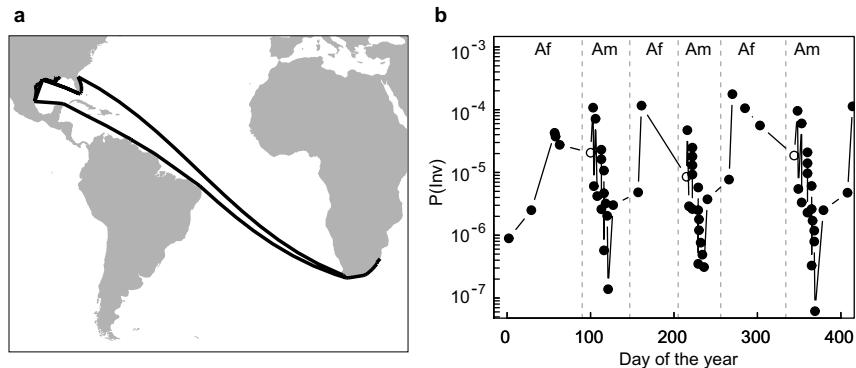


Fig. S7. Temporal development of the invasion risk during the journey of a single ship. The plot shows an example shipping trajectory hopping several times between ports in America (Am) and Africa (Af) (a), and the associated invasion probabilities assigned to the ship at each port call (dots) on that trajectory (b). Dashed lines indicate ocean crossings of the ship. The probability to introduce new species fluctuates tremendously during the voyage of the vessel. Every time the ship crossed the Atlantic, $P(\text{Inv})$ increased owing to the high probability to entrain non-native species (high $P(\text{Alien})$). However, the invasion risk was not necessarily increasing in the first port after an ocean crossing if environmental conditions did not match. In the example, the first American port was Altamira, (Mexico, open circles) which experiences a low salinity while all other ports are marine. Hence, after crossing the ocean and calling Altamira, the small value of $P(\text{Estab})$, i.e., the low habitat similarity, more than compensated the high value of $P(\text{Alien})$ resulting in a low $P(\text{Inv})$.

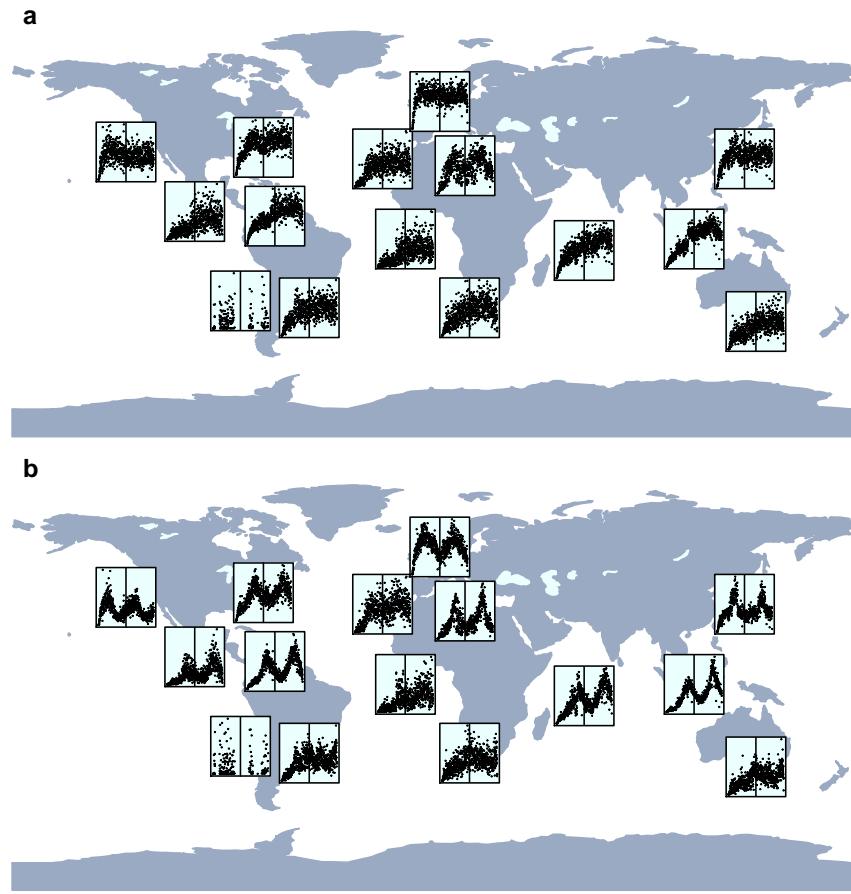


Fig. S8. Seasonality of invasion risks for ecoregions. Invasion risk as a function of season owing to (a) intraannual variation in shipping intensity (environmental conditions are kept constant) and (b) intraannual variation of both shipping intensity and environmental conditions. The scatter plots show the development of estimated invasion probabilities over the study period of two years aggregated for each of the 15 marine ecoregions (y-axes vary in scale). The two years are separated by a vertical line. Most ecosystems, including tropical regions, exhibit pronounced seasonal changes of invasion risks which are caused by intra-annual variations in shipping intensity and water temperatures. This temporal pattern varies strongly within the same climatic zone, e.g. in North America the highest invasion risk occurs at the west coast in winter, but at the east coast during summer. This pattern results from the seasonal variation of the environmental match of the donor and the recipient regions: For example, water temperatures in East Asia are higher and fluctuate at a larger amplitude compared to water temperatures at northern American ports. The highest environmental match is achieved during winter when temperatures in East Asia are lowest. The steep increase of invasion probability at the beginning of the year in some regions is an artefact of the initialization period of the model (i.e., the missing history of earlier shipping trajectories at the begin of the study period).

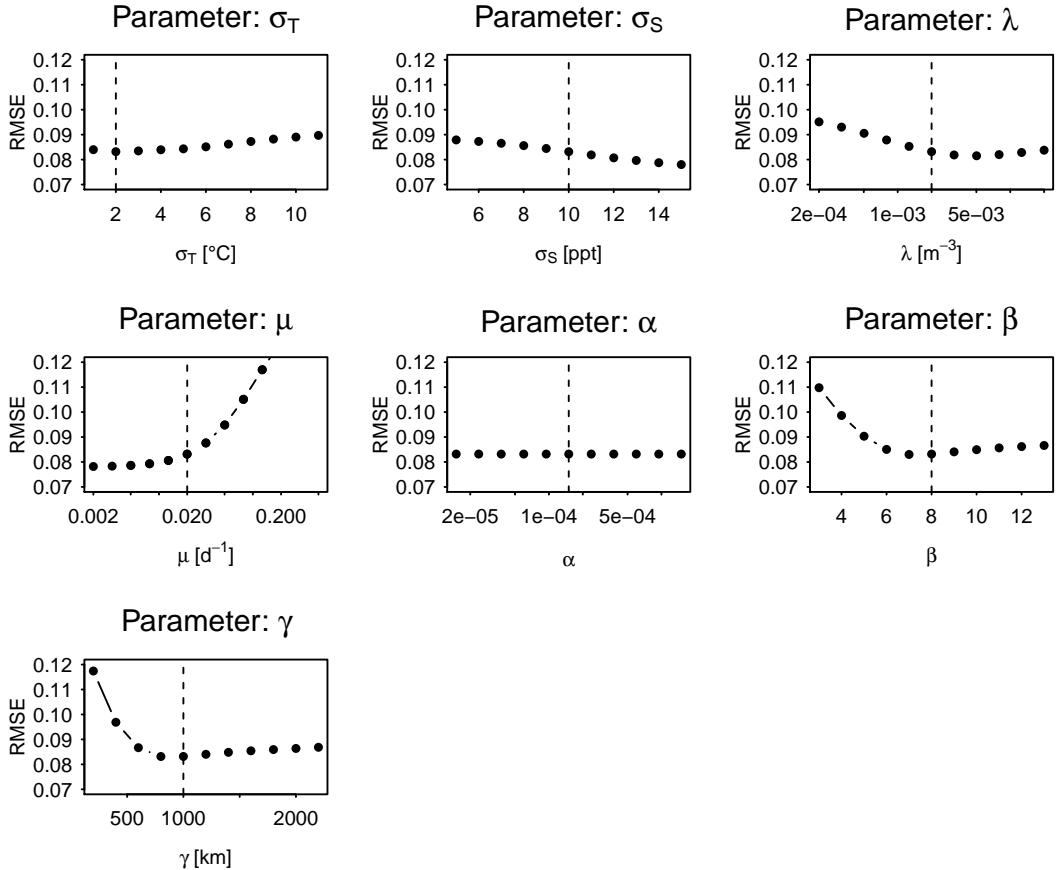


Fig. S9. Sensitivity analysis. The figure shows the influence of parameter variations on the goodness-of-fit for each parameter of the model. The goodness-of-fit is measured as the root mean square error (RMSE) of the deviations between field data shown in Fig. 4 in the main text and the model outcome (see “Model variants and model selection”). The vertical dashed lines indicate the parameter values used in this study. In most cases a change in a parameter setting has only minor effects on the RMSE. RMSE considerably increases if $\beta < 6$, $\gamma < 600$ km or $\mu > 0.02 d^{-1}$.

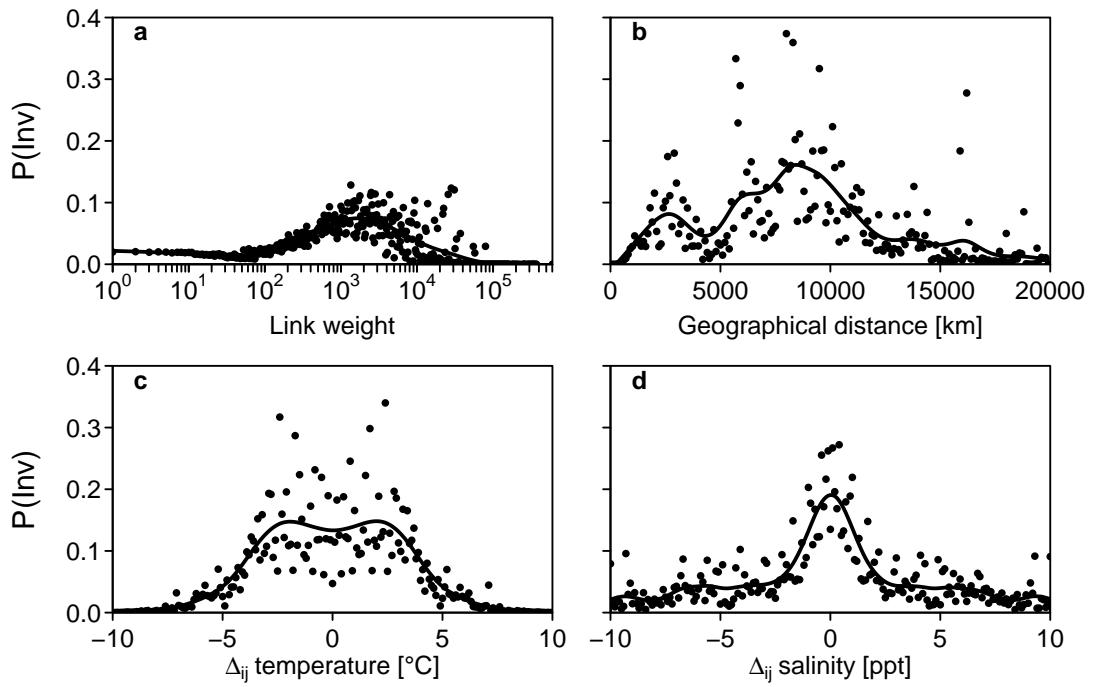


Fig. S10. Influence of risk components. Probability of invasion for each pair of ports connected by a ship as a function of (a) the number of ship movements between recipient and donor ports, (b) their geographical distance d_{ij} , (c) temperature difference ΔT_{ij} , (d) and salinity difference ΔS_{ij} . The figure reveals a maximal invasion risks for intermediate geographical distances ($d_{ij} \approx 9\,000$ km) and intermediate temperature difference ($|\Delta T_{ij}| \approx 3^{\circ}\text{ C}$) but vanishing salinity difference $\Delta S_{ij} = 0$ ppt.

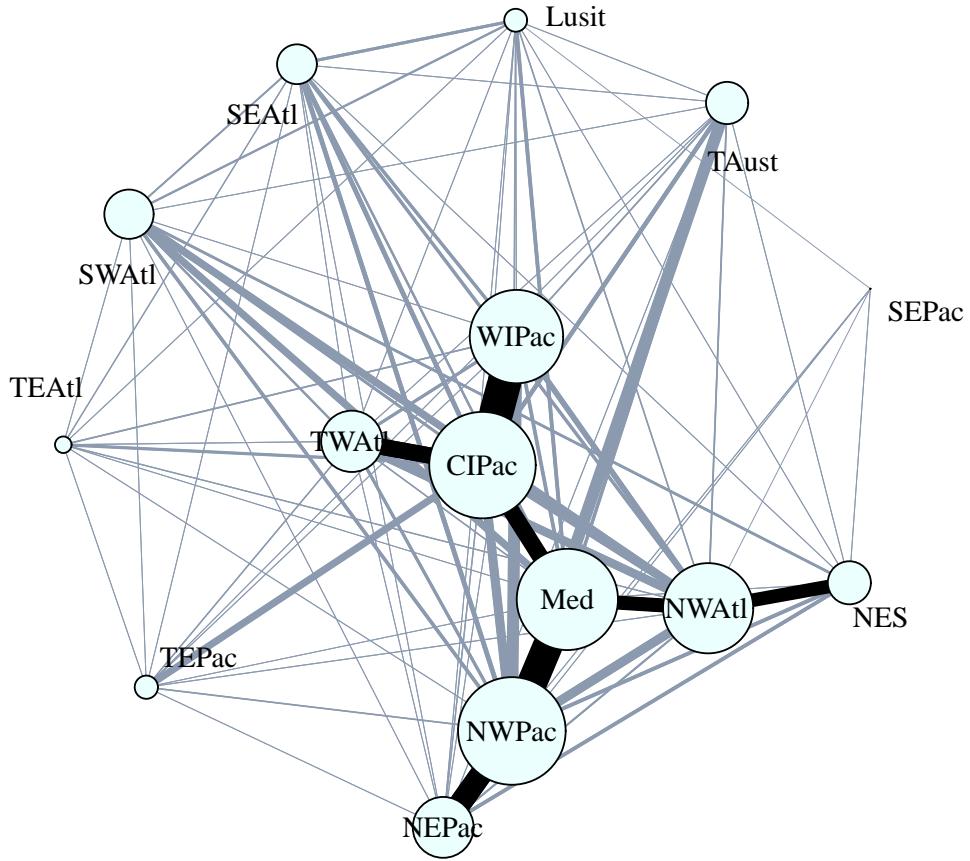


Fig. S11. Network of mutual invasion risks between ecoregions. The network is a visualization of the invasion risks shown in Table S3 (see also Fig. 3). Ecoregions are represented as circles (i.e., the nodes of the network). A link between two nodes is drawn if the invasion risk between two ecoregions exceeds a certain threshold, $P_{ij}(Inv) > 0.001$. The size of the nodes scales with the total invasion risk $P_j(Inv)$ of that ecoregion. The width of the links scales with the risk of invasion between two ecoregions $P_{ij}(Inv)$. The links of highest invasion risks ($P_{ij}(Inv) > 0.215$) are marked in black and highlight the core network of high-risk links and ecoregions. Pac: Pacific, IPac: Indo-Pacific, Atl: Atlantic, Med: Mediterranean Sea, Lusit: Lusitanian, Aust: Australasia, N: North, E: East, S: South, W: West, T: Tropical, C: Central.

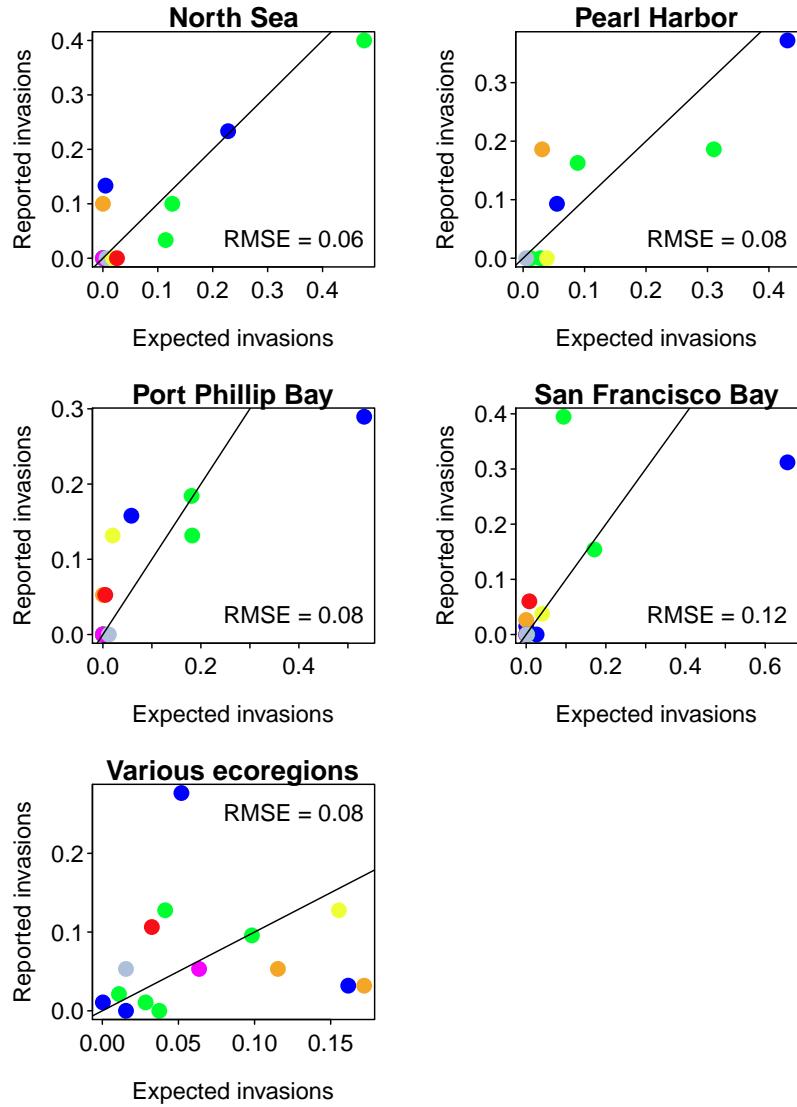


Fig. S12. Comparison of estimated and reported invasion risks. The figures show the normalized number of reported invasions as a function of the normalized number of expected invasions for five sample studies as presented in Fig. 4 in the main text. The colors represent different ocean (same color coding as in Fig. 4). The goodness-of-fit is calculated as the root mean square error (RMSE) which determines the mean deviation from the bisecting line (solid line).

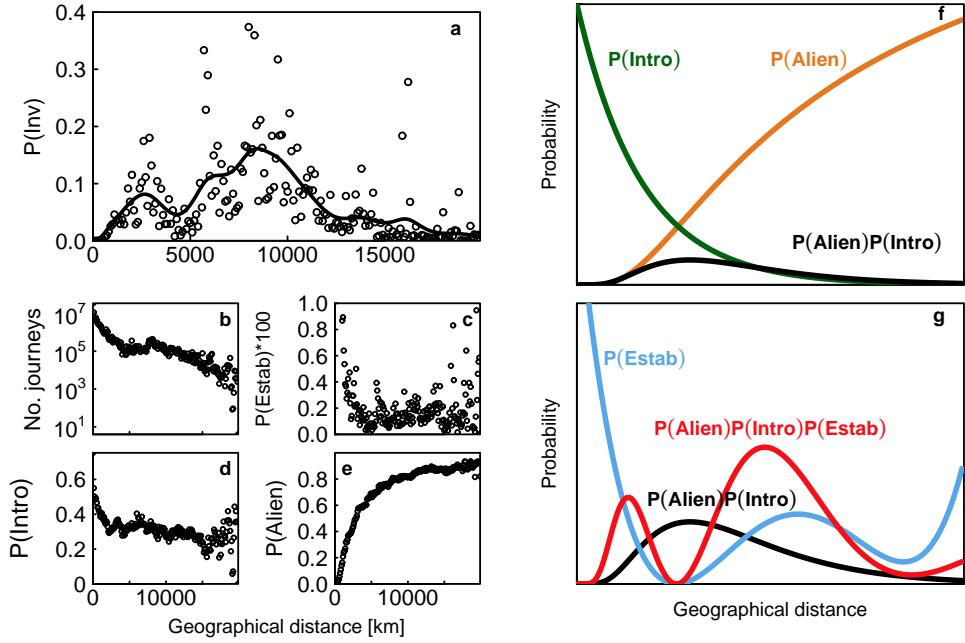


Fig. S13. Intermediate distance hypothesis. Histogram of invasion probabilities $P_{ij}(Inv)$ (a) and of selected model components (b–e) from numerical simulation, plotted as a function of the geographical distance d_{ij} between source and destination ports (bin size 100 km). (a) The model identifies highest invasion risks at intermediate distances of around 9 000 km. This pattern emerges as a result of the combination of (b) traffic volume, i.e., the number of ship movements in the data base connecting ports with a certain geographic distance, (c) the risk of establishment $P(Estab)$, (d) the risk of introduction $P(Intro)$, and (e) the risk of transporting non-native species $P(Alien)$. (c) The risk of establishment $P(Estab)$ declines sharply over short distances, because the environmental similarity between ports is reduced with increasing distance. Over larger distances, beyond a minimum of around 4 000 km, $P(Estab)$ increases again because of an increasing environmental similarity between ports which may be located in similar climatic regimes but on different hemispheres. (d) The risk of introduction $P(Intro)$ decays with geographic distance, because the travel time and the number of stopover sites increases on average with the geographic distance between ports. (f, g) Theoretical explanation for the pattern shown in (a). (f) Since $P(Alien)$ (orange) is increasing and $P(Intro)$ (green) is decreasing with distance, their product gives a hump-shaped function (black). This is plausible as individuals transported over short distances are less likely non-native, while individuals transported over long distances are less likely to survive the journey. (g) Probability of establishment $P(Estab)$ (blue) from (c) fitted as a function of distance by a fourth-order polynomial. Multiplying all three probabilities $P(Alien)P(Intro)P(Estab)$ (red) gives a good representation of $P(Inv)$ (see (a)), but also of the field data shown in Fig. 5. Note that for visualization the curves are scaled differently.

Rank	Port	Country	P(Inv)
1	Singapore	Singapore	0.566
2	Suez Canal	Egypt	0.472
3	Hong Kong	China	0.431
4	Panama Canal	Panama	0.307
5	Kaohsiung	Taiwan	0.288
6	Suez	Egypt	0.279
7	Port Said	Egypt	0.267
8	Busan	Korea (South)	0.23
9	Jebel Ali	United Arab Emirates	0.226
10	Kawasaki	Japan	0.2
11	Durban	South Africa	0.191
12	Yokohama	Japan	0.181
13	New York New Jersey	United States of America	0.178
14	Long Beach	United States of America	0.177
15	Xiamen	China	0.163
16	Fujairah	United Arab Emirates	0.158
17	Los Angeles	United States of America	0.158
18	La Plata	Argentina	0.157
19	Qingdao	China	0.147
20	Santos	Brazil	0.143

Table S1. Ranking of the 20 ports of highest invasion risk.

Rank	Donor port		Recipient port		P(Inv)
	Port	Country	Port	Country	
1	Jebel Ali	United Arab Emirates	Singapore	Singapore	0.121
2	Singapore	Singapore	Jebel Ali	United Arab Emirates	0.121
3	Singapore	Singapore	Fujairah	United Arab Emirates	0.111
4	Hong Kong	China	Panama Canal	Panama	0.093
5	Panama Canal	Panama	Hong Kong	China	0.089
6	Fujairah	United Arab Emirates	Singapore	Singapore	0.082
7	Hong Kong	China	Suez Canal	Egypt	0.064
8	Kaohsiung	Taiwan	Panama Canal	Panama	0.059
9	Suez Canal	Egypt	Hong Kong	China	0.055
10	Panama Canal	Panama	Kaohsiung	Taiwan	0.054
11	Xiamen	China	Suez Canal	Egypt	0.046
12	Singapore	Singapore	Khor Fakkan	United Arab Emirates	0.043
13	Busan	Korea (South)	Long Beach	United States of America	0.039
14	Long Beach	United States of America	Busan	Korea (South)	0.038
15	Kaohsiung	Taiwan	Singapore	Singapore	0.038
16	Singapore	Singapore	Kaohsiung	Taiwan	0.037
17	Singapore	Singapore	Colombo	Sri Lanka	0.037
18	Khor Fakkan	United Arab Emirates	Singapore	Singapore	0.036
19	Colombo	Sri Lanka	Singapore	Singapore	0.035
20	Busan	Korea (South)	Suez Canal	Egypt	0.033

Table S2. Ranking of the 20 links of highest invasion risk.

	Central Indo-Pacific	Lusitanian	Mediterranean Sea	North East Pacific	Northern European Seas	North West Atlantic	North West Pacific	South East Atlantic	South East Pacific	South West Atlantic	Temperate Australasia	Tropical East Atlantic	Tropical East Pacific	Tropical West Atlantic	Western Indo-Pacific
Central Indo-Pacific	0.254	0.002	0.296	0.018	0	0.107	0.147	0.072	0	0.044	0.072	0.051	0.124	0.289	0.609
Lusitanian	0.001	0.008	0.055	0.002	0.004	0.021	0.012	0.047	0	0.031	0.002	0.001	0	0.001	0.002
Mediterranean Sea	0.27	0.063	0.306	0.05	0.011	0.227	0.491	0.033	0	0.126	0.056	0.009	0.003	0.016	0.045
North East Pacific	0.017	0.003	0.036	0.031	0.055	0.023	0.399	0.002	0.001	0.006	0.012	0	0.007	0.007	0.004
Northern European Seas	0	0.004	0.011	0.062	0.13	0.241	0.059	0.008	0.001	0.043	0.013	0	0	0	0
North West Atlantic	0.099	0.017	0.248	0.033	0.216	0.008	0.154	0.043	0.001	0.125	0.026	0.019	0.013	0.143	0.085
North West Pacific	0.159	0.024	0.537	0.39	0.061	0.16	0.064	0.073	0.001	0.056	0.18	0.002	0.024	0.052	0.212
South East Atlantic	0.079	0.049	0.036	0.002	0.009	0.055	0.08	0	0	0.023	0.012	0.017	0.002	0.006	0.045
South East Pacific	0	0.001	0	0.001	0.002	0	0.004	0	0	0	0	0	0	0	0
South West Atlantic	0.042	0.032	0.113	0.003	0.04	0.126	0.056	0.028	0	0.001	0.002	0.007	0.002	0.036	0.008
Temperate Australasia	0.073	0.002	0.046	0.012	0.01	0.021	0.21	0.004	0	0.001	0.025	0	0.002	0.008	0.024
Tropical East Atlantic	0.036	0.001	0.007	0	0	0.014	0.001	0.015	0	0.008	0	0.006	0.001	0.014	0.021
Tropical East Pacific	0.094	0	0.004	0.006	0	0.013	0.017	0.001	0	0.002	0.002	0.002	0.005	0.019	0.008
Tropical West Atlantic	0.289	0.001	0.021	0.004	0	0.147	0.045	0.007	0	0.029	0.011	0.013	0.017	0.167	0.049
Western Indo-Pacific	0.563	0.006	0.062	0.007	0	0.098	0.154	0.055	0	0.007	0.017	0.024	0.009	0.048	0.129
Total risk	0.91	0.195	0.887	0.517	0.439	0.747	0.896	0.328	0.005	0.408	0.366	0.142	0.195	0.589	0.801
Mean time [years]	0.415	4.602	0.46	1.374	1.729	0.727	0.442	2.513	199.061	1.909	2.196	6.543	4.598	1.124	0.619
Evenness	0.784	0.719	0.742	0.516	0.613	0.833	0.801	0.802	0.717	0.766	0.694	0.731	0.557	0.685	0.621

Table S3. Matrix of mutual invasion risks between ecoregions (see Fig. 3). Matrix entries P_{ij} indicate the probability of an invasion from an ecoregion i (rows) to ecoregion j (columns). The total risk for an invasion in ecoregion j is given by $P(Inv)_j = 1 - \prod_i (1 - P_{ij}(Inv))$. The mean time to the first invasion in an ecoregion is estimated as $T_j = -1 / \ln(1 - P_j)$. The evenness of risk distribution is calculated with the Pielou index $E_j = -\sum_i p_{ij} \log P_{ij} / \log(n)$ where $n = 15$ is the number of ecoregions. Here, $p_{ij} = P_{ij}(Inv) / \sum_i P_{ij}(Inv)$ is the proportional risk for invasion from other ecosystems, i.e., given a specific invasion event in ecoregion j , p_{ij} estimates the likelihood that this invasion originated in ecoregion i .

Model	RMSE	\bar{r}
Version 2 (specBW, SR in recipient port)	0.066	0.77
Version 4 (specBW, SR in donor+recipient port)	0.066	0.77
Version 3 (specBW, SR in donor port)	0.067	0.71
Version 1 (specBW)	0.068	0.71
Version 5 (specBW, biofouling)	0.069	0.69
Version 6 (constBW)	0.072	0.68

Table S4. Ranking of different model versions. To each model version (1-6) optimal parameter values were obtained by applying an optimization algorithm (simulated annealing) which maximizes the fit between field data and model results. The goodness-of-fit was measured by the root mean square error (RMSE). In addition, the mean correlation coefficient (\bar{r}) between field data and model predictions is shown. The model variants, the data, and the fitting procedure are described in more detail in chapter “Model variants and model selection” of this supplement. The RMSE of the four best-fitting models vary by 2.5% or less, indicating that all four versions fitted the field data almost equally well. Likewise, for the four best-fitting model variants the mean correlation coefficient (\bar{r}) between field data and model results was always larger than 0.7.