Supplementary Material

Global economic costs of aquatic invasive alien species

Ross N. Cuthbert[a,b,*], Zarah Pattison[c], Nigel G. Taylor[d], Laura Verbrugge[e,f], Christophe Diagne[g], Danish A. Ahmed[h], Boris Leroy[i], Elena Angulo[g], Elizabeta Briski[a], César Capinha[j], Jane A. Catford[k,l], Tatenda Dalu[m,b], Franz Essl[n], Rodolphe E. Gozlan[o], Phillip J. Haubrock[p,q], Melina Kourantidou[r,s,t], Andrew M. Kramer[u], David Renault[v,w], Ryan J. Wasserman[x,b], Franck Courchamp[g]

[a]GEOMAR, Helmholtz-Zentrum für Ozeanforschung Kiel, 24105 Kiel, Germany

[b]South African Institute for Aquatic Biodiversity, Makhanda, 6140, South Africa

[c]Modelling, Evidence and Policy Research Group, School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK

[d]Tour du Valat, Research Institute for the Conservation of Mediterranean Wetlands, Arles 13200, France

[e]University of Helsinki, Faculty of Agriculture and Forestry, Department of Forest Sciences, P.O. Box 27, 00014 Helsinki, Finland

[f]Aalto University, Department of Built Environment, Water & Development Research Group, Tietotie 1E, FI-00076 Aalto, Finland

[g]Université Paris-Saclay, CNRS, AgroParisTech, Ecologie Systématique Evolution, 91405, Orsay, France

[h]Center for Applied Mathematics and Bioinformatics (CAMB), Department of Mathematics and Natural Sciences, Gulf University for Science and Technology, P.O. Box 7207, Hawally 32093, Kuwait

[i]Biologie des Organismes et Ecosystèmes Aquatiques (BOREA), Muséum national d'Histoire naturelle, CNRS, IRD, Sorbonne Université, Université Caen-Normandie, Université des Antilles, 43 rue Cuvier, CP 26, 75005 Paris, France

[j]Centro de Estudos Geográficos, Instituto de Geografia e Ordenamento do Território – IGOT, Universidade de Lisboa, Lisboa, Portugal

[k]Department of Geography, King's College London, Strand, WC2B 4BG, UK;

[l]School of BioSciences, University of Melbourne, Vic 3010, Australia

[m]School of Biology and Environmental Sciences, University of Mpumalanga, Nelspruit 1200, South Africa

[n]BioInvasions. Global Change. Macroecology-group, Department of Botany and Biodiversity Research, University Vienna, Rennweg 14, 1030 Vienna, Austria

[o]ISEM UMR226, Université de Montpellier, CNRS, IRD, EPHE, 34090 Montpellier, France

[p]Senckenberg Research Institute and Natural History Museum Frankfurt, Department of River Ecology and Conservation, Gelnhausen, Germany.

[q]University of South Bohemia in České Budějovice, Faculty of Fisheries and Protection of Waters, South Bohemian Research Center of Aquaculture and Biodiversity of Hydrocenoses, Zátiší 728/II, 389 25 Vodňany, Czech Republic

[r]Woods Hole Oceanographic Institution, Marine Policy Center, Woods Hole, MA 02543, United States

[s]Institute of Marine Biological Resources and Inland Waters, Hellenic Center for Marine Research, Athens 164 52, Greece

[t]University of Southern Denmark, Department of Sociology, Environmental and Business Economics, Esbjerg, 6700, Denmark

[u]Department of Integrative Biology, University of South Florida, Tampa, Florida 33620, United States

[v]Univ Rennes, CNRS, ECOBIO [(Ecosystèmes, biodiversité, évolution)] - UMR 6553, F 35000 Rennes, France

[w]Institut Universitaire de France, 1 Rue Descartes, 75231 Paris cedex 05, France

[x]Department of Zoology and Entomology, Rhodes University, Makhanda, South Africa

*Correspondence: rossnoelcuthbert@gmail.com

Supplementary Material 1

1. Materials and methods

1.1. Original data

Costs of aquatic invasive alien species (IAS) in the InvaCost database were sourced using both a systematic range of standardised search strings in online repositories (i.e. Web of Science, Google Scholar and Google search engine) and targeted searches of other relevant materials; the original database compiled 2,419 reported costs of IAS. Additional non-English data were included and originated from a search focused on fifteen of the most spoken languages, either as a mother tongue or second language, including Chinese, Spanish, Portuguese, Russian and Japanese. The data search was similar to the InvaCost English version, however the majority of the data came from targeted searches, i.e. searching webpages and emailing experts or stakeholders to request documents or files containing cost information. Further, additional search results were checked and integrated into InvaCost version 3 which we used for analyses (as of December 2020; https://doi.org/10.6084/m9.figshare.12668570). Efforts were made to check and remove duplicates of costs considering taxonomic, temporal and spatial descriptors, however, for some studies the limited information provided negated this decision. Other errors in the data encountered during this process were also corrected. In order to allow for comparisons, we considered the cost estimates standardised to 2017 equivalent United States dollars (US$) using the market exchange rate (World Bank) and accounting for inflation (Consumer Price Index of the year the cost was estimated for in each study) (see Diagne et al., 2020 for full details).

For the purposes of capturing all species with a distinct and important relationship to aquatic systems, we used the Environment_IAS" column. To be clear, this is based on the following definitions:

(1) Fully aquatic: Species that completes its life cycle completely within water. Reproduction, development and foraging all occur within water. Species may briefly occur in or use terrestrial habitats (e.g. overland movements of crayfish).

(2) Semi-aquatic: (a) Species that habitually utilizes both aquatic and terrestrial habitats, for reproduction, development and/or foraging (e.g. *Aedes albopictus*, *Neovison vison*, *Branta canadensis*) or (b) emergent plant species that commonly occurs in wetlands (marshes, swamps, peatlands) (e.g. *Phragmites australis*, *Spartina alterniflora*).

(3) Terrestrial: Species that completes its life cycle completely on land. Reproduction, development and foraging all occur on land. Species may briefly occur in or use aquatic habitats (e.g. for bathing, or crossing water bodies during migrations).

Aquatic entries unassigned to either "fully aquatic" or "semi-aquatic" were included in overall analyses of aquatic species, unless specified otherwise. Entries unattributable between "terrestrial" and "fully aquatic" or "semi-aquatic" environments were deemed "diverse/unspecified"; those unattributable between "aquatic" and "semi-aquatic" were "diverse/unspecified aquatic".

To account for cost duration in our estimation, we performed analyses on the annual cost data obtained through expanding the cost data, so that each line of the dataset corresponds to a cost entry for a single year (*expandYearlyCosts* function, 'invacost' R package; Leroy et al., 2020) based on the difference between the starting ("Probable_starting_year_adjusted" column) and ending ("Probable_ending_year_adjusted" column) years of reported costs (initial aquatic cost entries: 2,317, expanded aquatic cost entries: 5,682; initial terrestrial cost

entries: 6,433, expanded terrestrial cost entries: 19,499). When no years were specified in these column(s), we removed the cost entry from our analyses so as not to induce temporal biases in our results. This was even though the cost might have been repeated over many years, even up to the present time.

1.2. Cost descriptors

Key descriptors included:

(1) broad taxonomic groupings ("Phylum" column): we further broadly categorised taxa into invertebrates (i.e. "Arthropoda", "Annelida", "Cnidaria", "Ctenophora", "Mollusca", "Mollusca/Arthropoda", "Nematoda" and "Platyhelminthes"), vertebrates (i.e. "Chordata") and plants (i.e. "Chlorophyta" and "Tracheophyta"). Note that the entries in the Class "Ascidiacea" within "Chordata" were reclassified to invertebrates here. Remaining phyla were classified together into a separate category "Other" (i.e. "Arthropoda/Chordata", "Arthropoda/Oomycota", "Chytridiomycota", "Diverse/Unspecified", "Haptophyta", "Ochrophyta", "Oomycota" and "Proteobacteria"); finally, we produced a list of "top 10 costly aquatic genera", based on the greatest total reported costs using all expanded database entries;

(2) perceived reliabilities of cost estimates based on the type of publication and method of estimation ("Method_reliability" column);

(3) implementation types, which refers to whether the cost estimate was, at the time of estimation, actually realised in the invaded habitat or whether it was a potential cost (i.e. an expected cost beyond its actual distribution area and/or predicted over time within or beyond its actual distribution area) ("Implementation" column);

(4) geographic areas ("Geographic_region" and "Official_country" columns), whereby costs attributed to multiple geographic regions, or for which no geographic region was specified, were categorised as "Diverse/Unspecified";

(5) cost types ("Type_of_cost_merged" column): types grouped under three key categories: (*i*) "Damage" included damages or losses incurred by the invasion (e.g. costs for damage repair, resource losses, medical care); (*ii*) "Management" comprised control-related expenditure (e.g. monitoring, prevention, eradication) and costs related to administration, education, research etc.; (*iii*) "Mixed" included mixed damage and management costs (cases where reported costs were not clearly distinguished). Full explanation of the above descriptors is provided at https://doi.org/10.6084/m9.figshare.12668570.

1.3. Spatial and taxonomic connectivity

This network was composed of 141 species and 84 country nodes as well as 323 links. We illustrated the network under Gephi 0.9.2 (Bastian et al., 2009) with the *ForceaAtlas2* algorithm. To detect groups of countries with similar combinations of IAS costs, we applied the *Map Equation* community-detection algorithm (version 0.19.12, www.mapequation.org; Rosvall and Bergstrom, 2008; Rosvall et al., 2009). This algorithm has been used previously, with success, on bipartite species-site networks (Leroy et al., 2019). We ran the *Map Equation* algorithm optimised for a two-level partition of the network with 1,000 trials to ensure stability of the identified clusters. Network analyses were performed with the 'biogeonetworks' R package (Leroy, 2019). The *ForceaAtlas2* algorithm groups nodes with strong links and separates groups of nodes that do not have links; in other words, nodes clustered together on the graph share costs (e.g. countries impacted by the same species), whereas separated nodes do not share costs. Large links indicate high costs and node sizes are proportional to the total cumulative costs with a log spline. Similarly, node size for species

indicates the total costs over all countries while for country nodes it represents the total costs of all species in that country.

The *Map Equation* clusters together nodes that have high intra-group connectivity, but low inter-group connectivity; in other words, countries which share costs from the same set of IAS are clustered together. However, some nodes were difficult to classify into a particular cluster because they have high overall connectivity (e.g. an IAS having costs spread over multiple clusters of countries). In this case, the node was identified as a distinct cluster, a particularity of the *Map Equation* algorithm (see Vilhena and Antonelli, 2015).

1.4. Prediction of annual costs for aquatic IAS

Quantiles deduced from the difference between "Impact_year" and "Publication_year" were used as an indication of cost completeness using the *expandYearlyCosts* function (Fig. S1). This analysis indicated that 2020 had below 25% completeness, three years preceding 2020 would have below 50% completeness, and seven years below 75% completeness. In other words, it would take three years to capture 50% of cost data and seven years for 75%. The years from 2013 to 2020 were thus excluded owing to cost completeness below 75%.

To select the best model, we followed two key criteria: (1) models with lowest root mean square error (RMSE); (2) simplicity (fewer parameters) was preferred in models with similar RMSE. Furthermore, a qualitative assessment was used to exclude models, which projected decreasing costs in recent years (i.e. previous seven years), given that numbers of IAS (and thus probably costs) are increasing exponentially globally (Seebens et al., 2017). We also considered that such a decrease in costs would be driven by model sensitivity to incomplete quantifications over recent years (see explanation of lag time above). As such, models that projected cost decreases since 2013 (i.e. incomplete years preceding 2020) were not considered robust.

1.5. Mathematical modelling for cumulative temporal cost of impacts

We used the general model proposed by Yokomizo (2009) to express the cumulative cost impact C in terms of the population density u, which reads:

$$C(u) = AC_{\max}\left[\frac{1}{\frac{1-B}{B}\exp\left(-\frac{u}{s_2K}\right)+1} - B\right], \quad 0 \le u < K \tag{1}$$

with parameters:

$$A = \frac{1+e^{-s}}{1-B(1+e^{-s})}, \qquad B = \frac{1}{1+\exp\left(\frac{1}{s_2}-s\right)}, \qquad s = \frac{1-s_1}{s_2}, \tag{2}$$

where $C_{\max}$ is the cumulative maximum cost of impact and K is the carrying capacity. The model describes four different types of cumulative density-impact curves, characterised by the shape parameters s_1 and s_2, i.e. (*i*) low threshold curve $s_1 = 0, s_2 = 0.1$, (*ii*) S-shaped curve $s_1 = 0.5, s_2 = 0.1$, (*iii*) linear curve $s_1 = 1, s_2 = 1$ and (*iv*) high threshold curve $s_1 = 1, s_2 = 0.1$. See Fig. 1 in Yokomizo et al., (2009) for an illustration of the differences between these cost curve shapes.

To express the cumulative cost purely as a function of time, it is also required to model the temporal dynamics of the 'population'. The simplest model that exhibits reasonable properties and takes into account population multiplication, natural mortality and the impact of intraspecific competition considers logistic growth, given by the equation:

$$\frac{du}{dt} = uf(u) = \alpha u\left(1-\frac{u}{K}\right), \quad 0 \le u < K, \quad u = u(0) = u_0, \tag{3}$$

where $f(u) = \alpha\left(1-\frac{u}{K}\right)$ is the *per capita* growth rate, which in this case is a monotonically decreasing function of density. Here, α is the intrinsic growth rate, K is the carrying capacity and u_0 is the initial population density. The solution can readily be obtained, which reads:

$$u(t) = \frac{K}{1 + \left(\frac{K}{u_0} - 1\right) e^{-\alpha t}}, \tag{4}$$

see, for example, Petrovskii (2006) or Lewis (2016). The model demonstrates that the population density is bounded, since we have that $u \to K$ as $t \to \infty$. Combining equation (1) with the defined parameters in (2) and the population density expressed as a function of time in equation (4), one can express the cost solely in terms of time.

For the particular case of the linear curve model, which is important for our purposes, the shape parameters take specific values $s_1 = 1, s_2 = 1$. The general model given by equations (1)-(2) then reduces to:

$$C(t) = \frac{2C_{\max}}{e - 1}\left[\frac{e + 1}{\exp\left(1 - \frac{u(t)}{K}\right) + 1} - 1\right] \tag{5}$$

where $u(t)$ is given by equation (4) on assuming a logistic population and $u_0 = 1$ is the initial density. This model can be used as a crude estimate of the marginal cost impact over some time interval $t_1 < t < t_2$, computed as $C(t_2) - C(t_1)$.

Note that, since the model is non-spatial, $C_{\max}$ represents a localised maximum cost – which is a crude estimate for the actual real cost. One can assume that this may correspond to an empirical situation in which the rate of spread of the IAS is stagnant (i.e. reached its bioclimatic niche limits), but this is rarely the case in a more realistic scenario.

We tested each of the four model types (low threshold, S-shaped, linear and high threshold curves) given by the general model against all the data and also separately by excluding four extreme cost values (2003, US$25.06 billion), (2005, US$21.07 billion), (2009, US$18.30 billion) and (2011, US$52.61 billion), corresponding to times $t = 43, 45, 49$ and 51, respectively. These upper end extreme values were excluded if any cost value was greater than $Q_3 + 1.5 \times \text{IQR} = \text{US\$14.66 billion}$, where Q_3 is the upper quartile and the IQR is the interquartile range of the data set.

1.6. Reporting of invasion costs from aquatic IAS compared to terrestrial IAS

A list of known alien species worldwide was sourced from various databases (i.e. Meyer, 2000; Fofonoff et al., 2003; Wonham and Carlton, 2005; Carlton and Eldredge, 2009; Briski et al., 2016; Casties et al., 2016; DAISIE, 2016; Galil et al., 2016; Fofonoff et al., 2017; US Geological Survey, 2017; Richardson et al., 2020; Schwindt et al., 2020; unpublished data provided by Prof. Chad Hewitt and Prof. Marnie Campbell) and species were grouped between aquatic and terrestrial habitat types based on the criteria aforementioned. Where possible, we removed duplicated species among data sources, however, note that synonyms in the combined dataset could remain due to taxonomic discrepancies.

2. Results

2.1. Geographic regions

Taxonomically, in North America, 54% of costs of aquatic IAS were caused by vertebrates (US$89 billion), followed by invertebrates (36%, US$60 billion) and fewer by plants (4%, US$6.3 billion) (Fig. 3a). Invertebrates caused most costs in Asia (88%, US$40 billion), South America (96%, US$19 billion) and Africa (55%, US$0.94 billion). In Europe, invasive alien plants held slightly the biggest share of costs (39%, US$8.4 billion), followed closely by vertebrates (35%, US$7.6 billion). In Oceania-Pacific Islands, aquatic vertebrates caused the biggest share of costs (44%, US$0.18 billion), followed by plants (27%, US$0.11 billion). All costs in the Antarctic-Subantarctic were caused by vertebrates (US$376 thousand).

In North America, Asia, South America and in diverse or unspecified regions, costs mostly comprised damages (≥ 70% per region; Fig. 3b), whereas the opposite was true in Oceania and Pacific Islands (91% of costs associated with management, US$0.38 billion) and Antarctica (100% of costs associated with management, US$376 thousand). In Europe 41% (US$9.0 billion) of costs were associated with management, but 57% (US$12 billion) to

damage, whereas 49% of costs in Africa were mixed (US$0.83 billion) and 47% damage (US$0.81 million). Note that costs in Antarctica are comprised of French Southern and Antarctic Lands.

2.2. Prediction of annual costs for aquatic IAS

The OLS linear model predicted annual costs of US$383 billion, and robust linear regression US$416 billion in 2020. Quadratic fits of those models predicted costs of US$75 billion and US$16 billion, respectively. Whilst most models demonstrated an increase in costs over time, the quadratic robust regression predicted a reduction in costs over recent years and was thus not considered further, owing to effects of data incompleteness, but also a high RMSE (Fig. S2). Considering quantile regressions, 0.1, 0.5 and 0.9 relationships were increasingly convergent in recent years, indicating greater average annual cost amplitude similarities (i.e. among lower boundary, median value and upper boundary; Fig. 6e). The best fitting GAM indicated an increasing rate of costs over the last several decades, covering several orders of magnitude.

References

Bastian, M., Heymann, S., Jacomy, M. 2009. Gephi: An Open Source Software for Exploring and Manipulating Networks. International AAAI Conference on Weblogs and Social Media.

Briski, E., Ghabooli, S., Bailey, S.A., MacIsaac, H.J. 2016. Are genetic databases sufficiently populated to detect non-indigenous species? *Biological Invasions*, 18, 1911–1922.

Carlton, J.T., Eldredge, L.G. 2009. Marine bioinvasions of Hawaii. The introduced and cryptogenic marine and estuarine animals and plants of the Hawaiian archipelago. *Bishop Museum Bulletins in Cultural and Environmental Studies*, 4, 202 pp.

Casties, I., Seebens, H., Briski, E. 2016. Importance of geographic origin for invasion success: a case study of the North and Baltic Seas versus the Great Lakes-St. Lawrence River region. *Ecology and Evolution*, 6, 8318–8329.

Delivering Alien Invasive Species Inventories for Europe (DAISIE). 2016. Accessed in February 2016; Available at: http://www.europe-aliens.org/.

Diagne, C., Leroy, B., Gozlan, R.E., Vaissiere, A.C., Assailly, C., et al. 2020. INVACOST: a public database of the economic costs of biological invasions worldwide. *Scientific Data*, 7, 277.

Fofonoff, P.W., Ruiz, G.M., Steves, B., Hines, A.H., Carlton, J.T. 2003. National exotic marine and estuarine species information system (NEMESIS): Chesapeake Bay introduced species database. Accessed in February 2017; Available at: http://invasions.si.edu/nemesis/chesapeake.html.

Fofonoff, P.W., Ruiz, G.M., Steves, B., Carlton, J.T. 2017. California Non-native estuarine and marine organisms (Cal-NEMO) System. Accessed in February 2017; Available at http://invasions.si.edu/nemesis/.

Galil, B.S., Marchini, A., Occhipinti-Ambrogi, A. 2016. East is east and West is west? Management of marine bioinvasions in the Mediterranean Sea. *Estuarine, Coastal and Shelf Science,* 201, 7–16.

Leroy, B. 2019. biogeonetworks: Biogeographical Network Manipulation and Analysis, R package version 0.1.1.

Leroy, B., Dias, M.S., Giraud, E., Hugueny, B., Jézéquel, C., et al. 2019. Global biogeographical regions of freshwater fish species. *Journal of Biogeography*, 46, 2407–2419.

Leroy, B., Kramer, A., Vaissière, A-C., Courchamp, F., Diagne, C. 2020. Analysing global economic costs of invasive alien species with the invacost R package. *bioRXiv*, https://doi.org/10.1101/2020.12.10.419432.

Lewis, M., Petrovskii, S.V., Potts, J. 2016.*The Mathematics Behind Biological Invasions*. Springer.

Meyer, J.-Y. 2000. Preliminary review of the invasive plants in the Pacific islands (SPREP Member Countries) by the South Pacific Regional Environment Programme. Accessed in February 2017; Available at: www.sprep.org.ws.

Petrovskii, S.V., Li, B.L. 2006. *Exactly solvable models of biological invasion*. Chapman Hall/CRC, Boca Raton, FL.

Richardson, D.M., Foxcroft, L.C., Latombe, G., Le Maitre, D.C., Rouget, M., et al. 2020. The biogeography of South African terrestrial plant invasions. In: Van Wilgen, B.W., Measey, J. Richardson, D.M., Wilson, J.R.U., G.J., Zengeya, T. (Eds.). *Biological invasions in South Africa*. Springer, Berlin.

Rosvall, M., Bergstrom, C.T. 2008. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, 105, 1118–1123.

Rosvall, M., Axelsson, D., Bergstrom, C.T. 2009. The map equation. *European Physical Journal: Special Topics*, 178, 13–23.

Schwindt, E., Carlton, J.T., Orensanz, J.M., Scarabino, F., Bortolus, A. 2020. Past and future of the marine bioinvasions along the Southwestern Atlantic. *Aquatic Invasions*, 15, 11–29.

US Geological Survey. 2017. Nonindigenous Aquatic Species Database. Gainesville, Florida. Accessed in March 2017; Available at: https://nas.er.usgs.gov/queries/SpeciesList.aspx?group=&state=WA&Sortby=1#.

Vilhena, D.A., Antonelli, A. 2015. A network approach for identifying and delimiting biogeographical regions. *Nature Communications*, 6, 6848.

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. *Biological Invasions*, 7: 369–392.

Yokomizo, H., Possingham, H., Thomas, M., Buckley, Y. 2009. Managing the impact of invasive species: the value of knowing the density-impact curve. *Ecological Applications*, 19, 376–386.

Table S1. Parameters of general cumulative cost model fit against both all the data and the adjusted data, assuming temporal logistic population growth. Each model is characterised by distinct shape parameters. Here, we report the computed values of the squared correlation coefficient (r^2) and the root mean square error (RMSE).

Model	Shape parameters	All data		Adjusted data	
		r^2	RMSE	r^2	RMSE
(*i*) Low threshold curve	$s_1 = 0, s_2 = 0.1$	0.371	88.60	0.331	56.00
(*ii*) S-shaped curve	$s_1 = 0.5, s_2 = 0.1$	0.278	94.10	0.173	62.20
(*iii*) Linear curve	$s_1 = 1, s_2 = 1$	0.996	6.73	0.999	2.26
(*iv*) High threshold curve	$s_1 = 1, s_2 = 0.1$	0.994	8.57	0.998	3.06

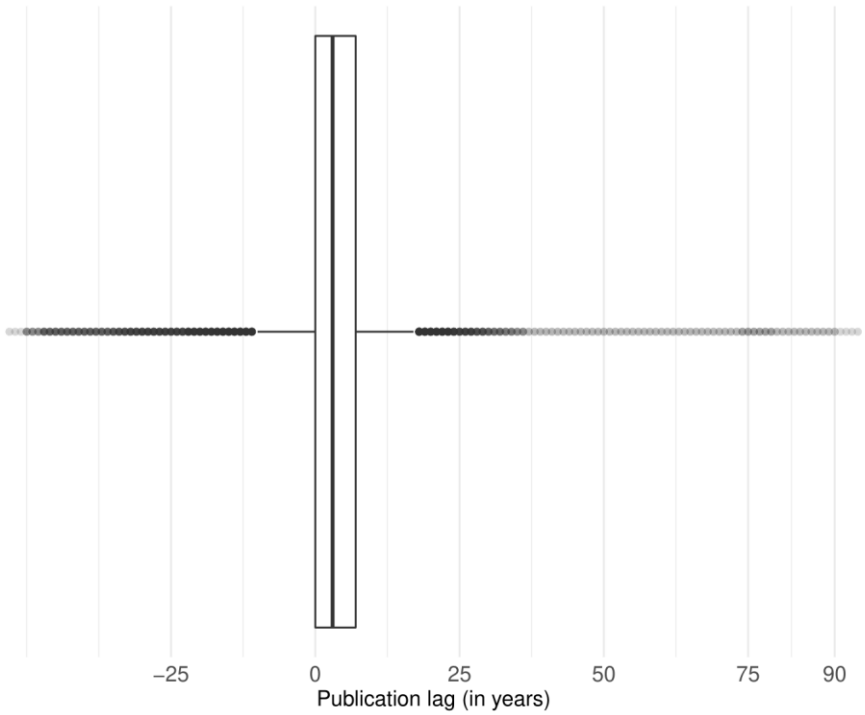

Fig. S1. Boxplot showing publication lag of reported costs for aquatic invasive alien species, calculated as the difference between year of publication (i.e. when cost was published) and year of impact (i.e. when cost was incurred). The box illustrates the median (50%) and interquartile ranges (25% and 75%), and horizontal lines represent minimum and maximum values. Raw points represent outliers ($\alpha = 0.2$). Negative values occur due to cost projections in future (i.e. impact years) beyond publication years. The median value is 3 years.

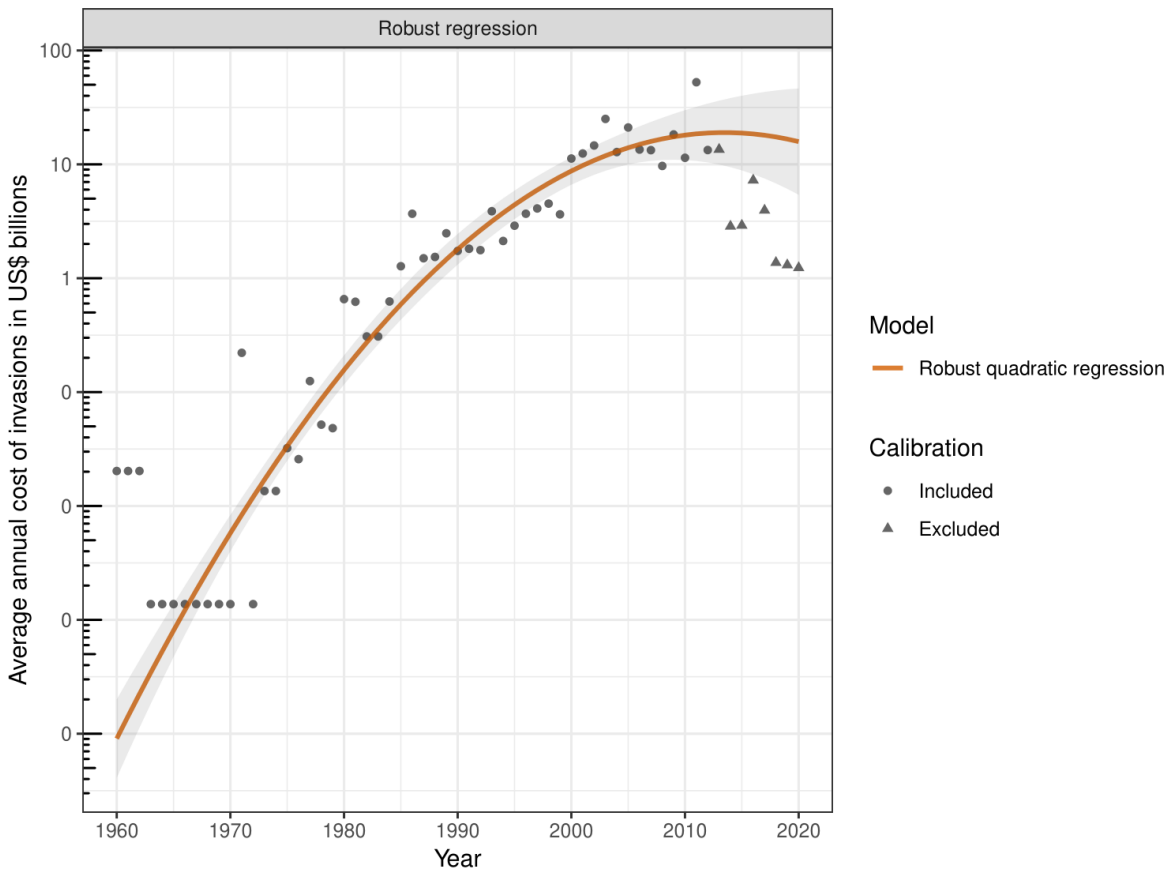

Fig. S2. Robust quadratic regression considering global aquatic invasion costs over time (RMSE = 0.63). Points are annual total costs. Note the *y*-axis is on a $\log_{10}$ scale. Shaded areas are 95% confidence intervals.