Report Cover Page

ACERA Project
1104E

Title
Ballast Water Risk Table Reconstruction

Author(s) / Address (es)
Sam Zhao, ABARES; DAFF
Peter Caley, Mathematics, Informatics and Statistics; CSIRO
Tony Arthur, ABARES, DAFF
Andrew Robinson, ACERA; The University of Melbourne
Matthew Gregg, Invasive Marine Species Program; DAFF.

Material Type and Status (Internal draft, Final Technical or Project report, Manuscript, Manual, Software)
Final Report

Summary

Background: Estimates of risk for the uptake, transportation and establishment of invasive marine pest species between marine ports provide key evidence that underpins ballast water management systems. Previously, CSIRO developed methods to generate tables that report estimates of the risk of the transfer of marine pests by ballast water between Australian ports (ballast water risk tables).

Overview: We reviewed the existing workflow for risk table construction and reconstituted the entire construction process within the readily available statistical software environment R. The table generation process was further modularised for ease of use and to allow risk tables to be updated without necessarily running the entire process.

Outcomes: We identify under what circumstances risk table updates should occur and how they should be carried out. The new R code was tested by reproducing CSIRO's previous results for the ballast water risk tables, using the same past input datasets. The new code has been used to produce a revised set of ballast water risk tables using preliminary updated port survey datasets. A key outcome of this project is that ABARES now has a good understanding of the risk table methodology and the code used to generate the risk tables.

Recommendations: That ABARES undertake the regular analytical maintenance of the risk tables on behalf of DAFF, with appropriate remuneration from the client Program.

ACERA Use only	Received By:		Date:
	ACERA / AMSI SAC Approval:		Date:
	DAFF Endorsement: ( ) Yes ( ) No		Date:

Ballast Water Risk Table Reconstruction

A report arising from ACERA project 1104E

ABARES project No. 43325

Sam Zhao[1], Peter Caley[2], Tony Arthur[1], Andrew Robinson[3] & Matthew Gregg[4]

[1] Australian Bureau of Agricultural and Resource Economics and Sciences (ABARES), GPO Box 1563 Canberra ACT 2601

[2] CSIRO Mathematics, Informatics and Statistics, GPO Box 664, Canberra, ACT 2601

[3] ACERA & Department of Mathematics and Statistics, University of Melbourne, VIC 3010

[4] Department of Agriculture, Fisheries and Forestry, GPO Box 858, Canberra ACT 2601

November 2012

Acknowledgements

Table of Contents

List of Tables

List of Figures

Executive Summary

Estimates of risk for the uptake, transportation and establishment of invasive marine pest species between marine ports provide key evidence underpinning ballast water management systems.

In this collaborative project between CSIRO, DAFF (ABARES and Biosecurity Animal Division) and ACERA, we updated and streamlined the methods developed by the CSIRO to generate tables that report estimates of the risk of the transfer of marine pests by ballast water between Australian ports, namely, ballast water risk tables (Zhao et al., 2012). We reviewed the existing workflow for risk table construction and reconstituted the entire construction process within the readily available statistical software environment R, after reproducing calculations previously undertaken using Visual Basic. Microsoft Excel is now used only for the storage of input data and outputted results. The code was modified to allow temperature tolerance data for the species of concern to be entered into an Excel worksheet, instead of being "hard-wired" within the R code and hence relatively inaccessible. The table generation process was further modularised for ease of use and to allow risk tables to be updated without necessarily running the entire process.

We identify under what circumstances risk table updates should occur and how they should be carried out. The new R code was tested by reproducing CSIRO's previous results for the ballast water risk tables, using the same past input datasets. The new code has been used to produce a revised set of ballast water risk tables using preliminary updated port survey datasets.

A key outcome of this project is that ABARES now has a good understanding of the risk table methodology and the code used to generate the risk tables. We recommend that ABARES undertake the regular analytical maintenance of the risk tables on behalf of DAFF, with appropriate remuneration from the client Program.

Introduction

Exotic species carried by ballast water could potentially threaten the marine environment of recipient ports. To estimate and manage the risks of these species within Australia, CSIRO and DAFF developed the Australian ballast water risk assessment (hereafter called the BWRA) (Hayes *et al.* 2007; Hayes *et al.* 2008; Hayes *et al.* 2009; Paini *et al.* 2011). The BWRA is a modular, species-specific system that estimates the risk that a species could be taken up from one Australian port (a donor port), transported to another Australian port (a recipient port) and successfully establish there, for any given month in the year. The risk is expressed in binary form as either a "1" for high risk or a "0" for low risk. If the risk is '1', then the vessel must manage its ballast water before arriving in the recipient port. Currently 129 ports are considered.

The original version of the BWRA was constructed using a combination of code written for the open-source statistical environment R (R Development Core Team 2011) and Visual Basic in Microsoft Excel. The process was unnecessarily complicated by the need to move backwards and forwards between these two programs, and the lack of in-house skills in Visual Basic programming within DAFF made code maintenance potentially difficult or expensive. In addition, the temperature tolerances for the species life-stages were embedded within the R code, which made updating difficult.

These complications reduced the utility of this system within DAFF. The current project was established to:

1. Determine whether the BWRA could be updated to carry out all of the analysis in R, with Excel being used only to hold input data and results,
2. Determine whether the system could be compartmentalised so that certain types of updates of the final tables would not require carrying out the full model refitting process (as documented herein),
3. Modify the process to allow for easy alteration of temperature tolerance data for species life stages as new data become available,
4. Develop the skills within ABARES to construct the risk tables to facilitate their use within DAFF,
5. Identify whether material could be published in the peer-reviewed literature.

In this report we:

1. Review the workflow for risk table construction and describe the mechanics of how we carry out the process now, with reference to the earlier version of the BWRA,
2. Compare the results obtained from the new version with those obtained under the old version to confirm that the new version is functioning correctly,
3. Describe the varying circumstances under which the risk tables would need to be updated, relate this to the mechanics of the new version of the BWRA and provide estimates of the time required for the various updates,
4. Update the risk tables with the latest port survey data.

Methodology

Risk table construction overview

A detailed description of the theory behind the construction of the risk tables is in Hayes *et al.* (2009) and references therein. Here we provide an overview of the process so that it can be clearly related to the mechanics of the new version of the BWRA. The risk analysis focuses on marine species that were selected by a hazard analysis process that was ratified by the National Introduced Marine Pests Coordination Group (NIMPCG), which was chaired by DAFF and had representatives from the States and the Northern Territory, industry, scientific and environmental organisations. The hazard analysis criteria are as follows.

By definition, the potential risk species are those that have ballast water mediated invasion history, have yet to reach the full extent of their potential range in Australia, and can be managed by ballast water exchange or treatment. Currently 7 species are included in the BWRA, namely the sub-tidal invertebrates *Asterias amurensis, Carcinus maenas, Varicorbula gibba, Musculista senhousia* and *Sabella spallanzani*, the sub-tidal plant *Undaria pinnatifida,* and the inter-tidal invertebrate *Crassostrea gigas*. The list will be updated as other potentially invasive species establish in Australian waters.

For any particular pest, the BWRA considers that there is a risk to a recipient port if, under simulated conditions:

1. The pest is present in the donor port. If the port has not been surveyed then it is assumed that the pest is present if it could complete at least 80% of its lifecycle if it were to invade in any month of the year (more detail about this is provided below), and,

2. Larva of the pest that could be taken up in ballast water are present in the water column in the month of departure from the donor port assuming the donor port has an existing population of the pest, and,

3. The pest survives in the ballast water during transit. At this stage the probability of survival is considered to be 1 for all transits due to a lack of information on species-specific survival in ballast tanks, and,

4. The pest arrives in the recipient port where surveys have shown that it is not present and in a month where, if it were released from ballast water, it would be able to complete at least 80% of its lifecycle. If the port has not been surveyed then it is assumed that the pest is not present.

These four points are equivalent to the modules described in earlier documentation for the BWRA (Hayes *et al.* 2009 and reference therein). Note that unsurveyed ports are treated differently depending on whether they are considered a donor or recipient port for the particular calculation. In any case, the most risk-averse approach is adopted.

As can be seen from the four points, construction of the risk tables requires a survey from the ports to determine whether the species are present or not (or the above assumption), an estimate of when the pest may be taken up in ballast water, an estimate of the probability the pest will survive transit (currently assumed to be 1), and an estimate of how much of the lifecycle of the pest will be completed should the pest arrive in a port in a particular month. Estimating the latter is the most detailed step, and has involved the most significant changes to the mechanics between the old BWRA and the new BWRA.

Construction of the risk tables follows the steps and draws on the datasets shown in Figure 1. Here we describe the steps in detail and how we have changed their implementation in the new version of the BWRA.
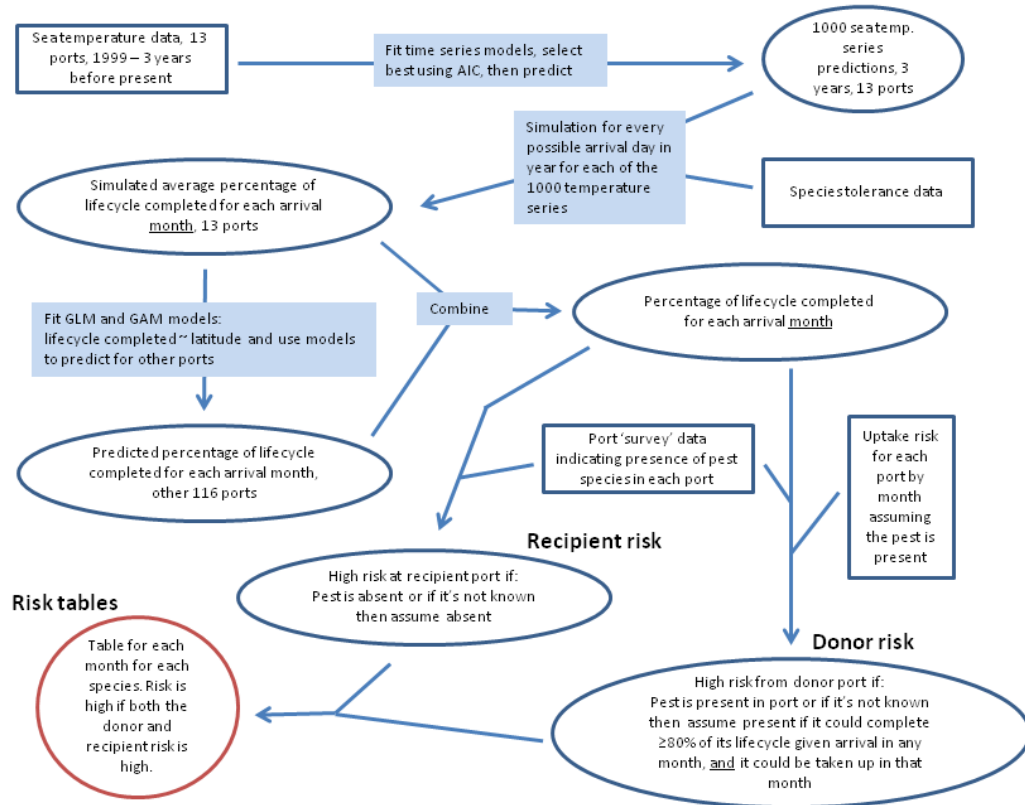
Figure 1. Flowchart for the process underlying the generation of Ballast Water Risk Tables. Open rectangles show data files. Ovals show derived files/information. Shaded rectangles describe manipulations.

Port survey status

Pest presence for each port has been derived from Monitoring Design Assessment Panel (MDAP) approved port monitoring surveys or the Consultative Committee on Introduced Marine Pest Emergencies (CCIMPE) range extension reports. Originally all of the port survey data in the database was based on the port surveys undertaken by CSIRO's Centre for Research on Introduced Marine Pests (CRIMP) and other agencies according to the protocols laid out in Hewitt and Martin (1996, 2001). This information was sometimes supplemented by other sources such as published literature and records of new discoveries, subject to CCIMPE and NIMPCG approval. The original surveys were later endorsed by MDAP (Hayes, pers. comm. 2013).

The data are held in an Excel spreadsheet with 129 ports in rows and 7 target species in the 2012 table. (Note there were 11 target species in the 2008 table) in columns. In the spreadsheet a "1" indicates presence and a "0" indicates absence. If a port has not been surveyed for that species then a blank is recorded. When the data are used in a risk table calculation they are read into program R. If an unsurveyed port is being considered as a donor port then presence is signified by a "1"; if the port is being considered as a recipient port then presence is signified by a "0". We do this because the final risk is determined by multiplying all the risks together and the only combination that leads to a final

risk of "1" is: present in donor (1) x taken up in ballast (1) x survives transit (1) x survives port arrival (1) x <u>absent</u> in recipient (1).

As indicated above, the assumption about whether a pest is present in an <u>unsurveyed</u> port is different depending on whether the port is being considered a donor or recipient port for the particular calculation. If the port in question is treated as a donor port then the pest is considered 'present' if it is forecast to complete 80% or more of its life-cycle if it were to arrive at the port in any month of a year, otherwise it is considered absent. If treated as a recipient port then the pest is considered absent (Hayes *et al.* 2007; Hayes *et al.* 2008; Hayes *et al.* 2009; Paini *et al.* 2011). This risk-averse protocol was approved by NIMPCG (Hayes, pers. comm., 2013).

Uptake in ballast water – vessel infection status

The data for whether a pest's life-stage is likely to be present in the water column and therefore available for uptake in ballast water ("1" = yes and "0" = no) in a particular month is held in an Excel spreadsheet "Vessel infection status CSIRO Oct 08.xlsx". This file comprises a sheet for each species, and each sheet the ports are in rows and months of the year are in columns. The data are based on the lifecycle of the species and the conditions at the local port. They do not depend on whether the pest has been detected at the port, i.e. they assume that the pest is present. At present they would be updated manually if additional information about the timing of the reproductive cycle of the species were obtained. The data are read into program R when required (described below). We refer to this as the 'vessel infection status' module to maintain consistency with earlier reports on the BWRA. An example is provided in Table 1.

Table 1. Example showing the risk that an *A. amurensis* life-stage that could be taken up in ballast water is present in a given month assuming the pest is extant in that port. Risk = "1" and "0" = no risk. Note this is for a subset of ports.

PortName	January	February	March	April	May	June	July	August	September	October	November	December
Abbot_Point	0	0	0	0	1	1	1	1	1	1	1	1
Adelaide	0	0	0	0	1	1	1	1	1	1	1	1
Airlie_Terminal	0	0	0	0	1	1	1	1	1	1	1	1
Albany	0	0	0	0	1	1	1	1	1	1	1	1
Ardrossan	0	0	0	0	1	1	1	1	1	1	1	1
Ballast_Head	0	0	0	0	1	1	1	1	1	1	1	1
Bamaga	0	0	0	0	1	1	1	1	1	1	1	1
Barrow_Island	0	0	0	0	1	1	1	1	1	1	1	1
Barry_Beach	0	0	0	0	1	1	1	1	1	1	1	1
Beadon	0	0	0	0	1	1	1	1	1	1	1	1
Beauty_Point	0	0	0	0	1	1	1	1	1	1	1	1
Bell_Bay	0	0	0	0	1	1	1	1	1	1	1	1
Bing_Bong	0	0	0	0	1	1	1	1	1	1	1	1
Botany_Bay	0	0	0	0	1	1	1	1	1	1	1	1
Bowen	0	0	0	0	1	1	1	1	1	1	1	1
Brisbane	0	0	0	0	1	1	1	1	1	1	1	1
Broome	0	0	0	0	1	1	1	1	1	1	1	1
Buffalo_Venture	0	0	0	0	1	1	1	1	1	1	1	1

Calculation of species lifecycle completion

Marine pest species require appropriate conditions at every stage of their lifecycle to progress to the next stage and eventually establish a new population. A critical component of the BWRA is the calculation of how much of the lifecycle is likely to be completed given propagules are delivered via ballast water into a new port in a particular month. In the BWRA it is assumed that if a species can complete at least 80% of its lifecycle then there is a risk it will establish (based on NIMPCG

agreement, Hayes, pers. comm., 2013). The BWRA calculates the percentage of the lifecycle completed in two different ways. First, for a subset of ports at which there are data available from Bureau of Meteorology's (BOM) SEAFRAME tidal gauges (*n*=15 excluding Cocos Island; http://www.bom.gov.au/oceanography/projects/abslmp/data/index.shtml), the lifecycle is simulated directly based on life history parameters of the species and daily maximum and minimum sea surface temperatures. These life history temperature tolerance data are now held in an Excel spreadsheet. The results from these simulations are then used to develop statistical models that relate the percentage of the lifecycle completed to latitude (a proxy for sea surface temperatures) for each month of the year using either generalised linear models (GLMs) if the relationship between survival and latitude is monotonic (e.g. either increasing or decreasing), or generalised additive models (GAMs) where the relationship is non-monotonic (e.g. low survival in low and high latitudes with high survival in middle latitudes). These models are then used to predict the percentage of the lifecycle completed for the remaining ports for each month based on their latitude. Hence, we end up with a prediction of the percentage of lifecycle completed for each species in each of the 129 ports given arrival in each month of the year, amounting to 129 x 7 x 12 = 10836 predictions.

Direct simulation of lifecycle completion for a subset of ports

The details of how these calculations are carried out can be found in Hayes *et al*. (2008). We have not changed the approach, just the mechanics of implementation. Briefly, for each species there is a stochastic model written in R that progresses the species through its lifecycle stages based on the daily synthetic sea water temperature data, which is stochastically generated from time series models that have been developed for that port (see below). For intertidal species (e.g. *C. gigas*), air temperature is also incorporated. The models account for the fact that there are critical temperature thresholds outside which lifecycle stages cannot survive. Temperature can also influence how quickly a species progresses through a lifecycle stage.

In the old implementation of the model, different parameter values for temperature tolerances were "hard coded" into program R. Except for *C. gigas,* which has been subject to considerable additional modelling – see (Paini *et al.* 2011), we have extracted the values into an Excel file with life-stages of each species in rows and duration and temperature tolerances in columns. This will allow for easier updating should these values change, and allows users to easily check what these values are. An example of how the data are stored is in Table 2. We note that there is a heat budget model developed for *C. gigas*, but did not consider it to be operational at this stage (Paini *et al.* 2011).

Table 2. Life-cycle and temperature tolerance for *Asterias amurensis*. The first column lists life stage. The "aalarv" within the Larval duration implies that at the stage of larva, we need to use a special function to simulate the duration. "rpert" indicates that we need to calculate the maximal temperature tolerance from the Pert Distribution, with parameters *a*, *b* and *c* given in the last three columns. L = lower and U = upper.

Life-stage	Larval duration		Min. spawning tolerance		Max. spawning tolerance		Min. temp. tolerances		Max. temp. tolerances		Rpert parameters		
	Min	Max	L	U	L	U	L	U	L	U	*a*	*b*	*c*
Gamete	1	3	NA	NA	NA	NA	0	10	24	30	NA	NA	NA
Larva	aalarv	aalarv	NA	NA	NA	NA	0	8	rpert	rpert	20	22	26
Juvenile	270	730	NA	NA	NA	NA	1	7	rpert	rpert	22	25	27
Adult	NA	NA	5	5	23	23	1	1	rpert	rpert	22	25	27

First, autoregressive integrated moving average models are developed for the daily maximum and minimum sea and air temperature for 13 ports around Australia where detailed temperature records are available from the SEAFRAME tidal gauges (Hayes *et al.* 2008). These ports are Broome, Burnie, Cape Ferguson, Darwin, Esperance, Hilarys, Groote Eylandt, Rosslyn Bay, Port Kembla, Portland, Port Stanvac, Spring Bay and Thevenard. The current analysis includes data over 11 years from 1999 to 2009. (NB CSIRO's previous analysis used data spanning the 9 years from 1999 to 2007). These

time series models are then used to stochastically generate synthetic runs of daily maximum and minimum temperature data. This step is quite time consuming, so we have introduced the ability to save the simulations as R-image files (AirTempSim.RData and SeaSurSim.RData) to avoid having to repeat the process. These simulated data runs are then used to estimate the proportion of lifecycle completion for each day of the year for each species in each of the 13 ports. For each combination of month, species and port, the estimated life-cycle completion is averaged over the days of the month. As an example, the results for *A. amurensis* are given in Table 3.

The original code was implemented in R and we have not changed the underlying simulation approach.

Table 3. Example showing the excel table with the average proportion of lifecycle completed for *A. amurensis* from simulation of the lifecycle model. Note this is for a subset of ports.

Port_name	January	February	March	April	May	June	July	August	September	October	November	December
Broome	0.00	0.00	0.00	0.00	0.01	0.12	0.13	0.06	0.01	0.00	0.00	0.00
Burnie	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Cape_Ferguson	0.00	0.00	0.00	0.00	0.05	0.17	0.15	0.08	0.02	0.00	0.00	0.00
Darwin	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Esperance	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.96
Hilarys	0.04	0.20	0.60	0.80	0.79	0.76	0.74	0.71	0.65	0.42	0.16	0.10
Groote_Eylandt	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.01	0.00	0.00	0.00	0.00
RosslynBay	0.00	0.00	0.00	0.04	0.20	0.27	0.21	0.15	0.06	0.01	0.00	0.00
Port_Kembla	0.10	0.11	0.21	0.47	0.88	0.96	0.95	0.95	0.94	0.93	0.85	0.59
Portland	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Port_Stanvac	0.27	0.28	0.65	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.79
Spring_Bay	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Thevenard	0.31	0.40	0.78	0.91	0.90	0.88	0.87	0.85	0.84	0.82	0.61	0.31

Estimating the percentage of lifecycle completed using statistical models

For each possible arrival month and species, the relationship between the proportion of a species' lifecycle completed and latitude is modelled either using a generalised linear model (GLM) (e.g. Figure 2) or a generalised additive model (GAM) (e.g. Figure 3). The input data is based on previous simulation results from the 13 ports.

These statistical models are then used to predict the proportions in all other ports. In the original BWRA the models were fitted in R and then the parameter estimates from the models were copied into Excel where a Visual Basic macro was used to make the predictions. We have modified the code so that the predictions are made in R and then written to two Excel files, one reporting the GLM fitting and another reporting the GAM fitting, without resorting to Visual Basic code. Each sheet stores particular species' estimated lifecycle completion by month of introduction for all 129 ports.
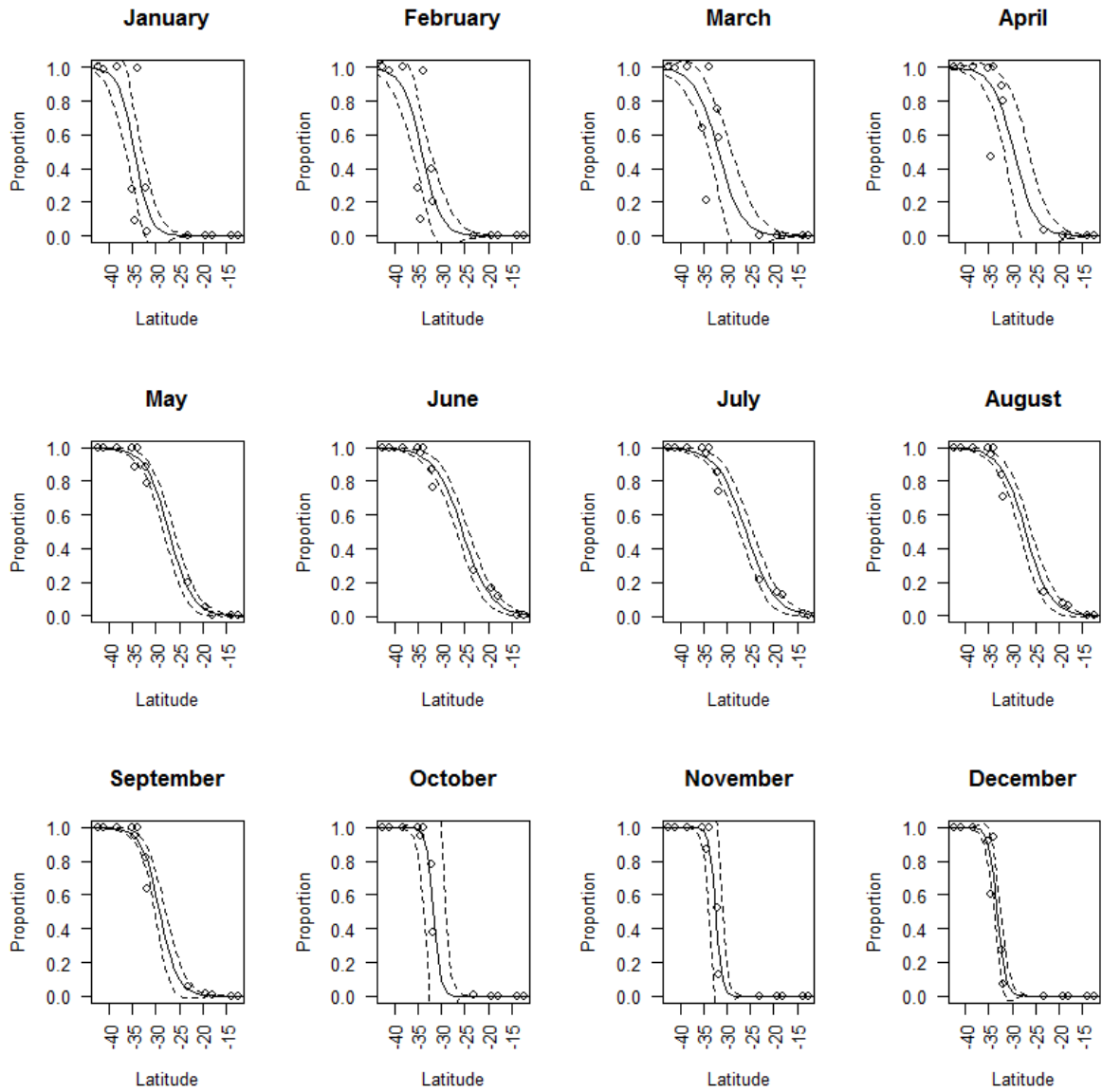
Figure 2. Relationship between the proportion of lifecycle completed upon introduction and latitude for the species *Asterias amurensis*. Circles are datapoints derived from simulation model, and lines are fitted GLM models with 95% confidence intervals.
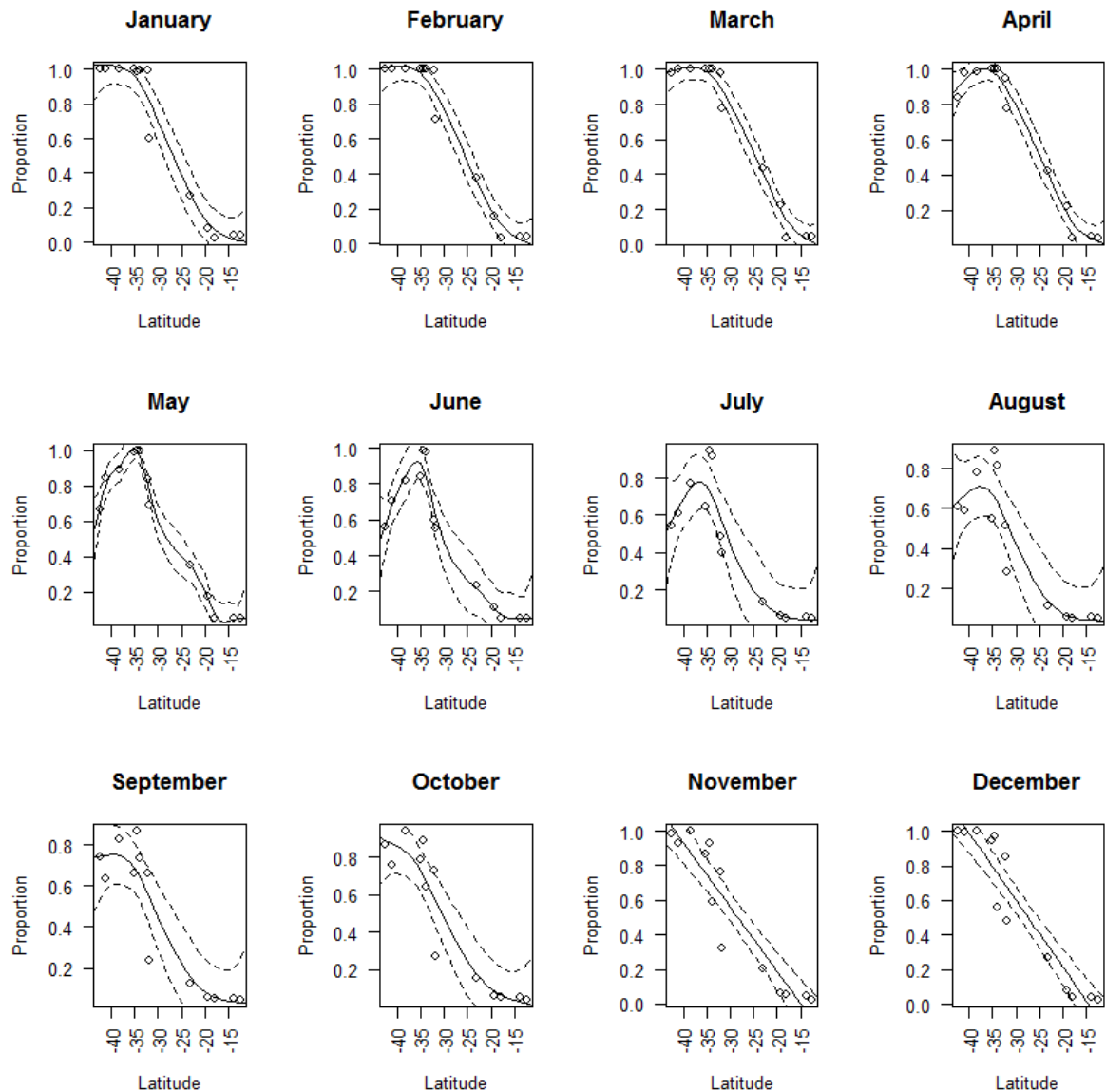
Figure 3. GAM fitted proportions of lifecycle completed against latitude for the species *C. gigas*. Circles are datapoints derived from simulation model, and lines are fitted GAM models with 95% confidence intervals.

Constructing risk tables between donor and recipient ports by month for all species

Using the predictions for the percentage of lifecycle completed for each combination of species, month and port, we assign a "1" to ports where ≥ 80% percentage of the lifecycle is completed. Note that for the 13 ports used in direct simulation the 80% comparison in the original BWRA was based on the <u>simulated</u> value, not the statistical model value. We have maintained this approach with the new BWRA, but point out that the result from the simulation can differ from the result from the statistical model Figure 2.

The risk tables are *n* x *n* (*n* refers to the total number of ports) matrices recording the risk of transferring and establishing ("1" for high risk and "0" for low risk) the pest species between donor ports and recipient ports. Donor ports are in rows and recipient ports are in columns and there is a new table for each species for each month. To construct the final risk tables we combine the information

13

from the four 'modules' for every combination of port, month and species. Note that at present the old BWRA and current version assume that if a vessel leaves a donor port in a particular month it arrives in the recipient port in the <u>same</u> month. We have maintained this assumption in the current version of the BWRA because the current operational software can only use the information in this form. We note that the majority of domestic journeys are of a relatively short (< 5 days) duration. Keith Hayes did a quick check of the Lloyds' data for 2002 and it suggested that about 60% of domestic journeys are five days or less, and 78% were 10 days or less. Note this high proportion of short journeys was also partly the reason for not investing much effort into the journey survival modelling. Initial modelling documented in Hayes et al (2008) suggests that relatively little risk reduction would be achieved for journeys of less than 10 days (Hayes, pers. comm., 2013).

With the information in each 'module' structured as either 1's or 0's, the final risk is determined by multiplication. Four 1's are required to give a high risk = 1. If there is a zero in any of the modules then the final risk = 0. We carry this out as follows.

1. First we construct column vectors for donor ports by combining the 'presence' of a species in <u>donor</u> ports with the risk that, assuming a pest is extant, propagules are present at the time of ballast water uptake:

$$vector.donor = \begin{pmatrix} survey\ status\ in\ donor\ port\ 001 \\ . \\ . \\ . \\ survey\ status\ in\ donor\ port\ 129 \end{pmatrix} \times \begin{pmatrix} vessel\ infection\ in\ port\ 001 \\ . \\ . \\ . \\ vessel\ infection\ in\ port\ 129 \end{pmatrix}$$

$$= \begin{pmatrix} 0 \\ 1 \\ . \\ . \\ 1 \\ 1 \end{pmatrix} \times \begin{pmatrix} 1 \\ 1 \\ . \\ . \\ 0 \\ 1 \end{pmatrix}$$

$$= \begin{pmatrix} 0 \\ 1 \\ . \\ . \\ 0 \\ 1 \end{pmatrix}$$

Here the first column vector, i.e. the survey status vector, is derived from port survey tables for the given species, e.g. *A. amurensis*, and the second column vector, i.e. the vessel infection status vector, is taken from Table 1 for the given species and given month.

2. Then we construct row vectors for recipient ports by combining the 'presence' of a species in recipient ports with the risk that it can complete $\geq 80\%$ of its lifecycle given that it arrived in that month (survival status). Remember 'presence' for unsurveyed ports is defined differently for donor and recipient ports as described above:

$vector.recipient$
$= (survey\ status\ in\ recipient\ port\ 001, \ldots, survey\ status\ in\ recipient\ port\ 129)$
$\times (survival\ status\ in\ recipient\ port\ 001, \ldots, survival\ status\ in\ recipient\ port\ 129)$
$= (0,1,\ldots,1,0) \times (1,0,\ldots,1,0)$
$= (0,0,\ldots,1,0)$.

3. For a given species and a given month the risk tables can be constructed by multiplying the above column and row vectors:

$$risk.table = vector.donor \times vector.recipient$$

These operations are carried out in R, reading the relevant Excel files for species and month and then writing the results back out to be stored in Excel files. Tables are generated for each species using both GLM and GAM models, but in final summaries the most appropriate model is chosen (see below). Each file contains sheets with risk tables for each month, and a summary sheet which shows the number of months in which there is a high risk for that species. We show an example in Table 4. In the example there is a high risk of transferring *A. amurensis* from Melbourne to Newcastle in 6 months of the year.

Table 4. Yearly risk table of *A. amurensis* **between donor ports (rows) and recipient ports (columns). The risks in this table are sums of monthly risks. We show the yearly number of risks between 17 donor ports and 14 recipient ports. Here the ports are listed according to the order of "Domestic Ballast Export Rank".**

Donor.port.pame	Domestic.ballast.export.rank	Domestic.ballast.Import.rank	State	Newcastle	Melbourne	Port_Kembla	Gladstone	Brisbane	Thursday_Island	Fremantle	Botany_Bay	Geelong	Port_Dalrymple	Port_Hedland	Sydney	Weipa	Townsville
Newcastle	1	1	New South Wales	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Melbourne	2	5	Victoria	6	0	7	0	0	0	5	7	0	0	0	7	0	0
Port_Kembla	3	4	New South Wales	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Gladstone	4	2	Queensland	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Brisbane	5	7	Queensland	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Thursday_Island	6	35	Northern Territory	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Fremantle	7	6	Western Australia	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Botany_Bay	8	8	New South Wales	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Geelong	9	9	Victoria	6	0	7	0	0	0	5	7	0	0	0	7	0	0
Port_Dalrymple	10	3	Queensland	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Port_Hedland	11	13	Western Australia	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sydney	12	15	New South Wales	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Weipa	13	11	Northern Territory	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Townsville	14	17	Queensland	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Adelaide	15	12	South Australia	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Devonport	16	14	Tasmania	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Dampier	17	10	Western Australia	0	0	0	0	0	0	0	0	0	0	0	0	0	0

4. If there is a high risk of transferring any pest species between ports then ballast water management should occur by conduct of a ballast water exchange outside of the territorial seas. In the final step we generate an excel file using R code that shows the number of species that could be transferred between ports in a given month. To construct that summary file we use GAM fitted proportions of lifecycle completed for *C. gigas and M. senhousia* and GLM fitted proportions for all other species. We also produce a summary sheet which shows the number of months in which there is a high risk of transferring at least one species between ports. The risk tables for all species are stored in an Excel file "RiskTable_AllSpecies.xlsx", which has a similar structure to the individual tables.

Comparison with original results

To check that the new mechanics are reproducing the results of the original BWRA mechanics, we used the same sea surface and air temperature datasets used by CSIRO in 2008. The final risk tables were very similar with only the occasional combination of species, month and donor/recipient port

differing. The differences arise because of the stochastic nature of the simulations (directly) and slight differences in the fitted statistical models from these simulations, combined with the 80% cut-off for risk of establishment. Our current simulated results for the proportion of a species' lifecycle completed were very close to those obtained by CSIRO's for nearly every combination of species and month for all the simulation ports. Table 5 lists a few notable differences.

Table 5. Comparison of a few notably different results obtained from our current and previous CSIRO's projects for simulated proportions of lifecycle completed. The temperature data, which were used to predict the proportions, spanned 9 years from 1999 to 2007.

Genus	Port	Month	Current result	Previous result
Asterias	Port Stanvac	December	0.78	0.87
Musculista	Burnie	January	0.80	0.78
Musculista	Burnie	December	0.19	0.21
Undaria	Port Stanvac	February	0.39	0.33
Undaria	Port Stanvac	March	0.80	0.79

The differences result from the stochastic simulations and are consistent with observations of the earlier CSIRO work. For example, for *A. amurensis* in Port Stanvac during December, one of CSIRO's simulations produced 0.87 (shown in Table 5) while in another simulation they obtained 0.75. This results from high variability in the lifecycle completion of *A. amurensis* in Port Stanvac in December arising from conditions becoming unsuitable for larval survival during this month, with life-cycle completion dropping from near 100% to near zero. Yearly variation in the timing of water temperatures exceeding the tolerance thresholds drives this variation.

For the statistical models, on the odd occasion the predicted lifecycle completed went from just above the 80% cut-off to just below and vice versa. For example, we obtained a value for *A. amurensis* in Clarence River during June of 0.79, whereas the previous analysis gave 0.80. We provide a further comparison of the determination of the risk outcomes in Table 6, which reports the outcomes of risk identification across all ports, species, and months. The simulation-based results are very similar, and the statistical results are a little less so. For the statistical results, the major differences are in the species *C. gigas* and *M. senhousia*, which were fitted using a GAM. A large number of the predicted lifecycle completion proportions for these species were close to 0.8, hence random fluctuations created the illusion of considerable differences. A reasonable explanation for the differences is that the original tables were constructed using temperature data up to 2007, whereas the reconstruction of the tables included data up to 2009. We repeated the analysis using only the data up to 2007 and the differences were much less substantial, although some persisted, which we believe to be due to random fluctuations in the simulation outcomes.

Table 6: Comparison of simulation results between previous and current tables. Note that this simple tabular summary ignores correlations between and within months, between and within ports, and within species.

12 Simulation Ports		Previous Risk			Non-simulation Ports		Previous Risk	
		Yes	No				Yes	No
New Risk	Yes	538	4		New Risk	Yes	4527	106
	No	6	460			No	145	5218

Update of the risk tables with the new survey data

We have used preliminary new survey data to produce updated risk tables with the new code. Compared with the old data, data from six ports has changed, i.e. Adelaide, Dampier, Darwin, Port Hedland, Portland and Westernport. *A. amurensis* and *S. spallanzani* are now shown as present in the ports of Westernport and Portland, respectively. *C. maenas* and *M. senhousia* which were previously shown as present in the Port of Adelaide are now shown as absent after consecutive MDAP approved

surveys have shown these species not to be present. Dampier has changed from 'unsurveyed' to no target species present after an MDAP approved port survey. The Ports of Darwin and Port Hedland were marked as "Newly updated data" but the status of the target species was the same as in the previous dataset.

When do the tables need to be updated?

Carrying out every process in the construction of the risk tables with the new mechanics takes about 12 hours of computer time using a HP Z600 workstation. The most time-consuming step for the computer is the simulation of the lifecycle based on time series data. However, the process is modularised in such a way that it is not necessary for the entire process to be carried out each time. The time consuming component of any update is ensuring that the input files are in the correct format and checking that the results are generally consistent with past results and organising files. As a guide, the process of fully updating the tables with new survey data, including associated meetings and reporting back, would take around 1 week. Here we summarise when tables should be updated and highlight any issues.

Updated survey data

The most obvious update to the tables would occur when the presence status of a target species in any port changes, either because of port survey or by other means. When this occurs the port survey spreadsheets would be updated and the risk tables could be updated by running the last step of the process. The last step includes generating the donor port survey and recipient port survey vectors (accounting for the unsurveyed ports) and combining all the modules.

Updated temperature data

Ideally, the data underlying the sea surface temperature model simulations would be updated yearly as recommended in Hayes *et al.* (2008), and consideration given to using a moving "window" of these data of the order of 10 years, depending on the annual variability. The rationale would be that we expect long term trends in climate and we should be updating our analyses to reflect this. This would require the whole process to be rerun. The results in this project of comparing models fitted to 1999–2007 data versus 1999–2009 data suggest that the differences in the risk tables arising from the addition of these data are small, but this may be due to the way this has been coded in the original methodology.

Changes to species tolerance information

If the species tolerance information were to change the entire process would need to be rerun after entering the new information into the AllSpeciesData.xlsx worksheet. This will probably take 1 week.

Addition of new species

The addition of a new species would require new code to be written to cover the lifecycle of that species and then the entire process would need to be rerun. This would be a much more involved process compared with the other changes, as it would require the development and parameterisation of the lifecycle model.

Addition of new port

The addition of a new port would require the relevant details to be added to the blank sheet for entering estimated lifecycle completion (currently named "BlankSurvivalSheet.csv"), the VESSEL.STATUS.FILE (currently named "Vessel infection status CSIRO Oct 08.xlsx" – note there are multiple sheets to be updated) and the SURVEY.STATUS.FILE (currently named "Port survey status June 2012.xlsx"). It is important that the ports are in the exact same order in each of these files and are given exactly the same name.

References

Hayes, K. R., Leriche, A., McEnnulty, F., Patil, J., Barne, M., and Cooper, S. (2009). Ballast Water Service Level Agreement (SLA) - Part IV. Final report to the Australian Government Department of Agriculture, Fisheries and Forestry. (CSIRO Mathematical and Information Sciences: Hobart).

Hayes, K. R., McEnnulty, F., Sutton, C., and Cooper, S. (2008). Ballast Water Decision Support System (DSS) Service Level Agreement (SLA) - Part III. Final report (08/83) to the Australian Government Department of Agriculture, Fisheries and Forestry. (CSIRO Mathematical and Information Sciences: Hobart).

Hayes, K. R., McEnnulty, F. R., R.M., G., Patil, J. G., Green, M., Lawrence, W., Barry, S., Sliwa, C., Migus, S., and Sutton, C. (2007). Ballast Water Decision Support System (DSS) Service Level Agreement (SLA) – Part II. Final report for the Australian Government Department of Agriculture Fisheries and Forestry. (CSIRO Marine and Atmospheric Research: Hobart).

Hewitt, CL & Martin, RB 1996, *Port Surveys for Introduced Marine Species – Background considerations and sampling protocols*, CRIMP Technical Report 4, CSIRO Division of Fisheries, Hobart, p. 40.

Hewitt, C.L. and Martin, R.B. 2001, *Revised protocols for baseline port surveys for introduced marine species: survey design, sampling protocols and specimen handling*, CRIMP Technical Report Number 22, Centre for Research on Introduced Marine Pests - CSIRO Marine Research, Hobart.

Paini, D., Caley, P., Adams, M., Hayes, K., Lin, X., and Murphy, B. (2011). Ballast Water Service Level Agreement (SLA) - Part V FINAL REPORT (CSIRO Ecosystem Sciences: Canberra).

R Development Core Team (2011). R: A language and environment for statistical computing. http://www.R-project.org

Zhao, S., Caley, P., Arthur, T, Robinson, A, and Gregg, M. (2012). Ballast Water Risk Table Reconstruction Internal Advice Note Arising from ACERA Project 1104E/ABARES project No. 43325. Internal DAFF Report.