Supplementary Data File:

“Evaluation of spatiotemporal imputations for fishing catch rate standardisation”

By Ross J. Marriott, Berwin Turlach, Kevin Murray, David V. Fairclough.

S.1. Simulation model: additional scenarios.

Alternative scenarios were simulated for the movement of Snapper among adjacent spatial sub-units \((a)\) within each time step \((y)\) and for a positively autocorrelated spatial distribution of the West Australian Dhufish pre-fishing sub-unit abundances \((N_{a,0})\).

S.1.1. Post-recruitment fish movements.

Snapper was selected for (post-recruit) movement scenarios because this species was suspected to have the greatest rates of movement, between spatial units that approximated the simulated sub-units in size (Lenanton et al. 2009; Wakefield et al. 2011; Fairclough et al. 2011, 2013). The movement scenarios were Diffusion and DDHS (or Density-Dependent Habitat Selection; MacCall 1990) movements. Both scenarios simulated the net number moving among adjacent sub-units in each time step (i.e., \([N_{a,y}^{\text{move}}]\) being, on average, ten percent of the pre-movement sub-unit abundances. This figure was within the range reported by Lenanton et al. (2009) for this stock. In the simulation, all post-recruit movements occurred at once, after catch removals by the simulated fishing fleet.

DDHS captured the prospect for possible spillover (Russ 2002) caused by post-recruit individuals migrating from sub-units with higher densities to others with more resources available per capita, as determined by that sub-unit's realised habitat suitability score in time step \(y\) \((S_{a,y})\). The size of each density-dependent movement was determined by the difference in \(S_{a,y}\) between adjacent sub-units before those simulated movements. Realised \(S_{a,y}\) were recalculated after fishing had occurred within each time step thus:

\[
S_{a,y} = rS_{a,0} - \beta N_{a,y},
\]

(S.1)

where \(\beta = r/N_{,0}\) and the inherent habitat suitability \(S_{a,0} = N_{a,0}/N_{,0}\), with \(N_{,0}\) representing the average starting abundance per sub-unit\(^1\). The number moving to a sub-unit \((a)\) from an adjacent one \((a')\) was then calculated by this difference in realised \(S_{a,y}\), scaled by a viscosity parameter \((V)\):

\(^1\) The average starting abundance per sub-unit was calculated from \(N_{,0} = 2N_{,15} / (1 + D/100)\), with \(N_{,15}\) = the average sub-unit abundance in Year 15, at the mid-point of the simulation; \(D\) = a fixed input for the relative depletion by the end year, Year 30; and \(N_{,15}\) was calculated by tuning the simulation model to result in this level of \(D\) (to within \(\pm\) 1%) for a given inputted value for the harvest ratio, \(H\).
\[ N_{a,y}^{\text{move}} = \sum_{a' \rightarrow a} V \left( S_{a',y} - S_{a,y} \right) \]

(S.2)

where all post-recruit fish movements were simulated to occur simultaneously among all adjacent sub-units. As for the Diffusion movements (and for simplicity), adjacent movements could occur only in North, East, South or West directions. Following MacCall (1990) the constraint that not more than 50 % of fish can move from any sub-unit in any time step was imposed. Values of \( V \) in Table 2 were calculated, for each combination of \( r \) and \( D \), as the value that produced 10 % average movements among sub-units for iteration 1 whilst fixing the inputted \( H \) at the value used to obtain 10 % diffusion.

S.1.2. Spatially autocorrelated starting fish abundances.

A spatially autocorrelated stochastic process was also used to generate starting abundances within spatial sub-units for West Australian Dhufish, because significant (positive) spatial autocorrelation was detected from a preliminary analysis for this species, but not for Snapper or Baldchin Groper (see Section S.2). This process was simulated using:

(S.3) \[ Y_i = \mu + \eta_i \]

\[ \eta_i = \lambda \left( \sum_{j=1}^{n} w_{ij} \eta_j \right) + \epsilon_i, \quad i = 1, 2, \ldots, n \]

following Plant (2012), where \( \mu \) represented the mean population abundance per sub-unit and \( \eta_i \) was the spatially autocorrelated random process. The \( \eta_i \) was comprised of: \( \lambda \), the autocorrelation coefficient measuring the overall strength of autocorrelation; \( w_{ij} \), a spatial weights matrix measuring connectivity between each \( Y_i \) and \( Y_j \) across \( n \) spatial units; and \( \epsilon_i \) was an uncorrelated random variable where \( \epsilon_i \sim N(0, \sigma^2) \). For this simulation, the \( Y_i \) were the population sub-unit starting abundances (\( N_{a,0} \)).

S.2. Simulation model: additional details on model inputs.

Inputs for simulation modelling were obtained from historical catch estimates provided from the former Catch And Effort System (CAES) reporting system, analyses of the first year (2008) of CPUE data collected from the current statutory fishing returns for the WCDSIMF, and available estimates of population parameters for the simulated stocks.

S.2.1. Fleet dynamics: multinomial distribution parameters.

The input value used for total simulated fleet size (\( n_f \)) was calculated from daily/trip logbook data as the number of vessels catching the study species in 2008. The mean number of sub-units fished per vessel was calculated by taking the mean of the
number of 10’ blocks for which each vessel reported catches of that species in 2008. The corresponding size parameter \( \theta \) of a multinomial distribution, which was the number of effort units fished by each vessel in a year \( (i.e., \theta_{v,y} = \sum_i \sum_{a \in m} E_{i,a,v,y}) \), was then calculated. The \( \theta_{v,y} \) was the integer value that produced a mean number of sub-units fished per vessel for a simulated dataset that most closely approximated the observed mean.

**S.2.2. Variance components: Sub-unit starting abundances and fleet dynamics.**

Rescaled estimated variance components from linear mixed models (LMMs) fitted to CPUE data were used to parameterise distributions for Monte Carlo resampling. Explanatory variables were selected and LMMs fitted following the methods of Fairclough et al. (2014). Random effects were estimated for 10’ block, \( \nu_1 \sim N(0, \hat{\sigma}_1^2) \), vessel, \( \nu_2 \sim N(0, \hat{\sigma}_2^2) \), and \( \epsilon \sim N(0, \hat{\sigma}_e^2) \) was the estimated residual or within-group variance. Estimated coefficients of variation (CV) were calculated by dividing the square root of estimated variance components by the mean of the normally-distributed responses (\( \hat{\mu} \)).

Means of EMMs (\( \hat{\mu} \)) and random effects as estimated from fitted LMMs were similar for Snapper and West Australian Dhufish, but a notably lower \( \hat{\mu} \) and block variance \( \hat{\sigma}_1^2 \) and a higher vessel variance \( \hat{\sigma}_2^2 \) and residual (that is, within-vessel and block) variance \( \hat{\sigma}_e^2 \) was estimated for Baldchin Groper (Table S1). These estimates resulted in greater differences in assumed inputs for modelling stochastic processes for the Baldchin Groper stock as compared to the other two species.

**Table S1.** Population estimates of mean and among-block variance of relative abundance from linear mixed models fitted to log-CPUE data recorded during the first year (2008) of implementation of fine resolution daily-trip logbooks. Block = 10’ grid block corresponding to simulated population sub-units; \( s_\alpha, s_\nu, s_\epsilon \) are CV estimates.

<table>
<thead>
<tr>
<th></th>
<th>Snapper</th>
<th>Baldchin Groper</th>
<th>West Australian Dhufish</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block ( \hat{\sigma}_1^2 )</td>
<td>0.037</td>
<td>0.188</td>
<td>0.045</td>
</tr>
<tr>
<td>( \hat{\mu} )</td>
<td>4.096</td>
<td>2.451</td>
<td>3.570</td>
</tr>
<tr>
<td>( s_\alpha = \hat{\sigma}_1 / \hat{\mu} )</td>
<td>0.047</td>
<td>0.177</td>
<td>0.059</td>
</tr>
<tr>
<td>Vessel ( \hat{\sigma}_2^2 )</td>
<td>0.208</td>
<td>1.081</td>
<td>0.254</td>
</tr>
<tr>
<td>( \hat{\mu} )</td>
<td>4.096</td>
<td>2.450</td>
<td>3.569</td>
</tr>
<tr>
<td>( s_\nu = \hat{\sigma}_2 / \hat{\mu} )</td>
<td>0.111</td>
<td>0.424</td>
<td>0.141</td>
</tr>
<tr>
<td>Error ( \hat{\sigma}_e^2 )</td>
<td>0.788</td>
<td>0.841</td>
<td>0.605</td>
</tr>
<tr>
<td>( \hat{\mu} )</td>
<td>4.096</td>
<td>2.450</td>
<td>3.569</td>
</tr>
<tr>
<td>( s_\epsilon = \hat{\sigma}_e / \hat{\mu} )</td>
<td>0.217</td>
<td>0.374</td>
<td>0.218</td>
</tr>
</tbody>
</table>

CV estimates for block (\( s_\alpha \)), vessel (\( s_\nu \)) and within block and vessel error (\( s_\epsilon \)) were then rescaled for use as simulation model input variances. The rescaled variances for simulation were: of log-normally distributed unfished population abundances among sub-units (Var[log \( (N_{i,a}) \)]) of average log CPUE among vessels (Var[log \( (C_{i,y}) \)]) to represent variation in fishing power within the fleet (Beverton and Holt 1957); and of log-normally distributed catches among vessels within each sub-unit and year (Var[\( \epsilon_{i,a,y} \)]) to represent variation in CPUE among fishing events for each vessel,
respectively. Rescaling involved multiplying the estimated CV by the respective mean of the log-transformed simulated quantity and then squaring this product.

S.2.3. Spatial autocorrelation parameters.

A spatial analysis of the standardised 2008 CPUE data was done to assess the significance of spatial autocorrelations. A GLM with the same set of explanatory variables as retained in the fitted LMMs was fitted to the log CPUE data and used to predict a matrix of estimated population marginal means (EMMs) for each stock (Searle et al. 1980). These matrices of EMMs were then converted into a spatial data objects (log $C_{x,y}$) by creating data vectors of latitude ($x$) and longitude ($y$) from the centroids of 10' blocks using the spdep and rgdal packages in R (R Development Team 2014). Detrending models were then tested for explaining significant ($\alpha < 0.05$) variation within each of these spatial data objects and, if significant, were applied to remove stationarity prior to spatial analysis following Plant (2012).

These data were then tested for spatial autocorrelation using the Moran's $I$ statistic. For stocks with significant spatial autocorrelation detected a spatially autocorrelated random process was generated:

$$\log C_{x,y} = T_{x,y} + \eta_{x,y} + \epsilon_{x,y},$$

(S.4)

where $x$ and $y$ represented data vectors of latitude and longitude, $T_{x,y}$ was a detrending model applied to achieve stationarity, $T_{x,y} = x + y + x^2 + y^2 + xy$, $\eta_{x,y}$ was the spatially autocorrelated random process and $\epsilon_{x,y}$ was an uncorrelated random variable, $\epsilon_{x,y} \sim N(0, \sigma_{\epsilon,x,y}^2)$ (Plant 2012). The spatially autocorrelated random process was generated by firstly constructing a spatial weights matrix using an inputted value for the parameter $\lambda$, which represented the strength of spatial autocorrelation occurring over the interval [-1,1] and assuming, for simplicity, rook's case with row normalized weights corresponding to the observed range of $x$ and $y$. Two boundary rows and columns were then removed to account for the initial transient effect, following Plant (2012).

As it was unclear which input value to use for $\lambda$, a fixed input of $\lambda = 0.75$ was assumed for cases where positive spatial autocorrelation was simulated. This value was arbitrarily selected to ensure a sufficiently strong spatial autocorrelation for contrasting with corresponding base case scenarios of no spatial autocorrelation ($\lambda = 0$). An input value for $\sigma_{\epsilon,x,y}$ was the value for simulating spatial autocorrelation, using Equation (S.3) and $\lambda = 0.75$, which resulted in an among sub-unit variance in relative abundance that matched the observed $\delta_t^2$.

The non-significant ($P > 0.05$) fit of detrending models (S.4) to the resulting standardised log(CPUE) spatial data objects indicated that detrending models were not necessary. Positive spatial autocorrelation was significant for West Australian Dhufish (Moran's $I = 0.59, P < 0.001$) but was non-significant for Snapper (Moran's $I = -0.04, P = 0.59$) and Baldchin Groper (Moran's $I = 0.08, P = 0.15$). This result indicated that it was appropriate to simulate an additional scenario with positive spatial autocorrelation for West Australian Dhufish but not for Snapper or
Baldchin Groper. Using $\sigma_{e,x,y} = 0.104$ and $\lambda = 0.75$ produced a simulated value for $\hat{\sigma}^2$ that matched the observed estimate so these values were used to generate a spatially autocorrelated distribution for West Australian Dhufish in subsequent simulations.

S.2.4. Number of bootstrap replications.

Fifty replications of bootstrapped CIs for imputed $I_{k,y}$ for Snapper (High Growth High Depletion scenario) showed that the standard deviation of bias-adjusted 80% CI widths decreased from bootstrap lengths of 20 to 2,000 for all imputation methods (Base, Linear, Geometric, Negative Exponential) and types (Before, Gap, After; Figs. S1, S2). This showed the benefit in increasing the length of bootstrap replications for lowering the standard deviation (and thus increasing precision) of this interval estimator, with smaller improvements apparent for higher bootstrap lengths (that is, from 1,000—2,000). Steeper declines were observed for later years of imputation periods except for the Base method, which imputed a constant $I_{k,y}$ with the same precision across years. On the basis of this evidence, and considering computational time constrains, it was decided that a bootstrap length of 1,000 was sufficient to achieve satisfactory precision for simulation evaluations.

References


Figure S1. Standard deviations of the widths of 80% bias adjusted bootstrapped confidence intervals (CIs) with number of bootstrap replications: Snapper, High Growth High Depletion scenario with no movement, by imputation method and type. Odd years shown.
Figure S2. Standard deviations of the widths of 80% bias adjusted bootstrapped confidence intervals (CIs) with number of bootstrap replications: Snapper, High Growth High Depletion scenario with no movement, by imputation method and type. Even years shown.
S.3. Results for alternative simulated scenarios.

**Figure S3.** Comparison of mean imputed values with population abundance for each stock and type of missing data period: Low Growth High Depletion scenario. Error bars are standard errors presented for population abundance and Geometric-imputed values to indicate relative precision of presented means. Grey shading covers observed CPUE data that were used to calculate imputed values.
**Figure S4.** Comparison of mean imputed values with population abundance for each stock and type of missing data period: High Growth, Moderate, Depletion scenario. Error bars are standard errors presented for population abundance and Geometric-imputed values to indicate relative precision of presented means. Grey shading covers observed CPUE data that were used to calculate imputed values.
Figure S5. Comparison of mean imputed values with population abundance for each stock and type of missing data period: Low Growth Moderate Depletion scenario. Error bars are standard errors presented for population abundance and Geometric-imputed values to indicate relative precision of presented means. Grey shading covers observed CPUE data that were used to calculate imputed values.
**Figure S6.** Comparison of mean imputed values with population abundance for each stock and type of missing data period: High Growth High Depletion scenarios with alternative simulated mechanisms for fish movement (Diffusion; DDHS = Density-Dependent Habitat Selection) and spatial distribution (positive spatial autocorrelation). Grey shading covers observed CPUE data for that analysis level block, which were used to calculate the imputed values.
Figure S7. Comparison of mean imputed values with population abundance for each stock and type of missing data period: Low Growth High Depletion scenarios with alternative simulated mechanisms for fish movement (Diffusion; DDHS = Density-Dependent Habitat Selection) and spatial distribution (positive spatial autocorrelation). Grey shading covers observed CPUE data for that analysis level block, which were used to calculate the imputed values.
Figure S8. Comparison of mean imputed values with population abundance for each stock and type of missing data period: High Growth Moderate Depletion scenarios with alternative simulated mechanisms for fish movement (Diffusion; DDHS = Density-Dependent Habitat Selection) and spatial distribution (positive spatial autocorrelation). Grey shading covers observed CPUE data for that analysis level block, which were used to calculate the imputed values.
Figure S9. Comparison of mean imputed values with population abundance for each stock and type of missing data period: Low Growth Moderate Depletion scenarios with alternative simulated mechanisms for fish movement (Diffusion; DDHS = Density-Dependent Habitat Selection) and spatial distribution (positive spatial autocorrelation). Grey shading covers observed CPUE data for that analysis level block, which were used to calculate the imputed values.
Figure S10. Box and whisker plots of average MSEs for stocks with alternative simulated scenarios for fish movement (Diffusion; DDHS= Density-Dependent Habitat Selection) and spatial distribution (A-C = positive spatial autocorrelation). Average MSE = MSE of imputed values averaged across years within each model iteration. Medians represented as horizontal white lines.
Figure S11. Mean normalised $I_y$ and mean $N_y$ (± standard error for population abundance and Geometric-imputed indices), Low Growth, High Depletion scenario: Stocks. Error bars are standard errors for mean $N_y$ representing stochastic variation across 200 iterations of the simulation model.
Figure S12. Mean normalised $I_y$ and mean $N_y$ ($\pm$ standard error for population abundance and Geometric-imputed indices), High Growth, Moderate Depletion scenario: Stocks. Error bars are standard errors for mean $N_y$ representing stochastic variation across 200 iterations of the simulation model.
Figure S13. Mean normalised $I_y$ and mean $N_y$ (± standard error for population abundance and Geometric-imputed indices), Low Growth, Moderate Depletion scenario: Stocks. Error bars are standard errors for mean $N_y$ representing stochastic variation across 200 iterations of the simulation model.
Figure S14. Mean normalised $I_y$ and mean $N_y$ (± standard error for population abundance and Geometric-imputed indices), High Growth, High Depletion scenario: Alternative simulated scenarios for fish movement (Diffusion; “Density-Dependent”= Density-Dependent Habitat Selection) and spatial distribution. Error bars are standard errors for mean $N_y$, representing stochastic variation across 200 iterations of the simulation model.
Figure S15. Mean normalised $I_y$ and mean $N_y$ (± standard error for population abundance and Geometric-imputed indices), Low Growth, High Depletion scenario: Alternative simulated scenarios for fish movement (Diffusion; “Density-Dependent”= Density-Dependent Habitat Selection) and spatial distribution. Error bars are standard errors for mean $N_y$, representing stochastic variation across 200 iterations of the simulation model.
Figure S16. Mean normalised $I_y$ and mean $N_y$ (± standard error for population abundance and Geometric-imputed indices), High Growth, Moderate Depletion scenario: Alternative simulated scenarios for fish movement (Diffusion; “Density-Dependent”= Density-Dependent Habitat Selection) and spatial distribution. Error bars are standard errors for mean $N_y$ representing stochastic variation across 200 iterations of the simulation model.
Figure S17. Mean normalised $I_y$ and mean $N_y$ (± standard error for population abundance and Geometric-imputed indices), Low Growth, Moderate Depletion scenario: Alternative simulated scenarios for fish movement (Diffusion; “Density-Dependent”= Density-Dependent Habitat Selection) and spatial distribution. Error bars are standard errors for mean $N_y$ representing stochastic variation across 200 iterations of the simulation model.
Figure S18. Mean relative error, Low Growth, High Depletion scenario: Stocks. Relative error, \( RE_y = \log(\text{normalised } I_y) - \log(\text{normalised } N_y) \).
**Figure S19.** Mean relative error, High Growth, Moderate Depletion scenario: Stocks. Relative error, $RE_y = \log(\text{normalised } I_y) - \log(\text{normalised } N_y)$. 

$\text{Method:}$
- Main Effects
- No Impute
- Base
- Linear
- Geometric
- Neg. Exp.
- Logistic
Figure S20. Mean relative error, Low Growth, Moderate Depletion scenario: Stocks. Relative error, $RE_y = \log(\text{normalised } I_y) - \log(\text{normalised } N_y)$. 

Method:
- Main Effects
- No Impute
- Base
- Linear
- Geometric
- Neg. Exp.
- Logistic
Figure S21. Mean relative error, High Growth, High Depletion scenario: Alternative simulated scenarios for fish movement (Diffusion; “Density-Dependent” = Density-Dependent Habitat Selection) and spatial distribution. Relative error, $RE_y = \log(\text{normalised } I_y) - \log(\text{normalised } N_y)$. 

Method:
- Main Effects
- No Impute
- Base
- Linear
- Geometric
- Neg. Exp.
- Logistic
Figure S22. Mean relative error, Low Growth, High Depletion scenario: Alternative simulated scenarios for fish movement (Diffusion; “Density-Dependent” = Density-Dependent Habitat Selection) and spatial distribution. Relative error, $RE_y = \log(\text{normalised } I_y) - \log(\text{normalised } N_y)$. 

Method:
- Main Effects
- No Impute
- Base
- Linear
- Geometric
- Neg. Exp.
- Logistic
Figure S23. Mean relative error, High Growth, Moderate Depletion scenario: Alternative simulated scenarios for fish movement (Diffusion; “Density-Dependent”= Density-Dependent Habitat Selection) and spatial distribution. Relative error, $RE_y = \log(\text{normalised } l_y) - \log(\text{normalised } N_y)$.
Figure S24. Mean relative error, Low Growth, Moderate Depletion scenario: Alternative simulated scenarios for fish movement (Diffusion; “Density-Dependent” = Density-Dependent Habitat Selection) and spatial distribution. Relative error, $RE_y = \log(\text{normalised } I_y) - \log(\text{normalised } N_y)$. 