



**Fisheries New Zealand**

Tini a Tangaroa

# An investigation of the performance of CPUE modelling approaches – a simulation study

New Zealand Fisheries Assessment Report 2019/57.

A.D. Langley

ISSN 1179-5352 (online)

ISBN 978-1-99-000865-8 (online)

October 2019



Requests for further copies should be directed to:

Publications Logistics Officer  
Ministry for Primary Industries  
PO Box 2526  
WELLINGTON 6140

Email: [brand@mpi.govt.nz](mailto:brand@mpi.govt.nz)

Telephone: 0800 00 83 33

Facsimile: 04-894 0300

This publication is also available on the Ministry for Primary Industries websites at:

<http://www.mpi.govt.nz/news-and-resources/publications>

<http://fs.fish.govt.nz> go to Document library/Research reports

**© Crown Copyright – Fisheries New Zealand**

## TABLE OF CONTENTS

<b>EXECUTIVE SUMMARY</b>	<b>1</b>
<b>1. INTRODUCTION</b>	<b>3</b>
<b>2. METHODS</b>	<b>4</b>
2.1 Simulator configuration	4
<b>3. RESULTS</b>	<b>12</b>
3.1 Simulator evaluation	12
3.2 CPUE simulation modelling – <i>Trawl</i> based data sets	15
3.3 CPUE simulation modelling – <i>Day</i> based data sets	21
3.4 Simulating changes in species catch reporting	26
3.5 Model diagnostics	31
<b>4. DISCUSSION</b>	<b>35</b>
<b>5. RECOMMENDATIONS</b>	<b>36</b>
<b>6. ACKNOWLEDGMENTS</b>	<b>37</b>
<b>7. REFERENCES</b>	<b>37</b>



## EXECUTIVE SUMMARY

**Langley, A.D. (2019). An investigation of the performance of CPUE modelling approaches – a simulation study.**

*New Zealand Fisheries Assessment Report 2019/57. 50 p.*

This study was conducted using a simple simulation framework to evaluate a range of standard CPUE modelling approaches. The primary focus of the study was to investigate the influence of the catch reporting regime and CPUE data processing in the evaluation of the CPUE modelling approaches routinely applied for the derivation of CPUE indices.

The key conclusions of the simulation study are, as follows.

- Positive catch models (lognormal, Gamma and Weibull) yielded biased estimates of trends in relative abundance for all species, with the exception of those species that were ubiquitous and occurred in almost all (> 95%) of the CPUE records. In all other cases, the positive catch CPUE indices under-estimated the scale of the variation in species abundance (hyper-stability) and this was most pronounced for species of intermediate abundance (second tier).
- The inclusion of the binomial component in the combined delta-lognormal, delta-Gamma and delta-Weibull CPUE models generally compensates for the bias in the positive catch component of the CPUE models, particularly for the main species in the total catches. The combined CPUE indices closely approximated the trend in the species abundance for the five most abundant species; i.e., the species that were most regularly reported (either by trawl or day).
- For the lower tier (less abundant) species, the delta-Gamma and delta-Weibull models yielded CPUE indices that were less biased than the delta-lognormal model.
- The performance of the delta-lognormal models was more sensitive to the underlying distributional assumptions of the species (species occurrence and variability). The performance of the delta-lognormal models deteriorated with the increasing contrast in the distribution of a species, corresponding to an increasing proportion of zero-catch records being included in the simulated data set (i.e., higher variation in species occurrence and catch rate). In general, the delta-lognormal CPUE indices tended to over-estimate the extent of the trends in relative abundance for those species with at least 50–60% of the data records comprising zero catches (i.e. the CPUE indices exaggerate the magnitude of the increase or decrease in stock abundance).
- Overall, the delta-Weibull and delta-Gamma CPUE models were relatively insensitive to distributional assumptions of the simulated data and performed well even for data sets with a high proportion of zero-catch records.
- While these “two-stage” models performed well for the range of species, in terms of accurately estimating the abundance trends, an examination of the model diagnostics (QQplots, residual plots, etc.) indicated that the positive catch components of these models may under-estimate the extent of variation in catches. Therefore, while the model diagnostics may indicate a deficiency in the overall performance of the CPUE model, the diagnostics do not provide a strong basis for deciding whether or not to accept the resulting CPUE indices as indices of relative abundance.
- The deficiencies in the diagnostics from the negative binomial CPUE models were considerably greater than from the components of the “two-stage” models. The modelling results indicate that the structure of CPUE data do not conform to a negative binomial distribution and, consequentially, the negative binomial CPUE models were not considered appropriate for statistical inference.
- Changes in species catch reporting have the potential to introduce a significant bias in the positive catch component of the CPUE indices for the lower tier (less abundant) species. The influence of changes in reporting was considerably lower for the “two-stage” combined models (delta-lognormal, delta-Gamma and delta-Weibull), although the delta-Gamma and delta-Weibull CPUE models were considerably less sensitive to changes in reporting behaviour than the delta-lognormal models.
- The concentration of fishing effort in areas of highest species abundance is likely to result in a degree of hyper-stability of the CPUE indices derived for the key species of interest. The extent of

the hyper-stability of the CPUE indices will be related to the underlying species distribution and the movement rates of the species. There is potential to reduce these effects by incorporating spatially structured approaches in the derivation of the CPUE indices.

There is considerable scope to extend the current simulation approach to more realistically represent the spatial distributions of species assemblages, relative to habitat preferences, within specific fisheries areas as informed by spatially resolved data from trawl surveys and/or fisheries catch and effort data. Similarly, fisheries data could be applied to formulate alternative assumptions for the distribution of fishing effort relative to the abundance of individual species.

Recommendations from the current study are as follows.

- CPUE analyses should include a characterisation of the CPUE data set, including metrics that describe the overall prevalence of the species, the frequency of individual catch reporting, and the derivation of the positive catch records (via catch- or effort-based apportionment of total catches). The CPUE data set should be limited to the distribution of the species of interest; i.e., the data set should exclude effort records from outside the range of the species (defined by depth and/or area).
- Characterising the intensity of fishing effort relative to the abundance of the species (e.g. concentration indices) is likely to be informative regarding the degree of hyper-stability in the CPUE indices. This is likely to be more important for the main target species.
- Positive catch CPUE models (lognormal, Gamma or Weibull) are appropriate for ubiquitous species only; i.e., those species that occur in the reported estimated catches for most of the CPUE records (at least 95% of records in each year).
- In all other cases, combined “two stage” models should be applied to determine CPUE indices. The Weibull and Gamma (log link) positive catch models appear more robust than the lognormal model across the range of species, although the lognormal models were comparable for the more abundant species. The final model selection should be based on standard model selection criteria (i.e., the proportion of variance explained).
- Caution should be applied when evaluating CPUE indices that are based on positive catch records that include a considerable proportion (more than 20%) derived from the apportionment of landed catches amongst all associated fishing effort records (generating multiple, small notional catches).

## 1. INTRODUCTION

Trends in the catch rates of fish species derived from commercial fishing activities are routinely used to monitor the abundance of fish stocks. The species catch rates from fisheries are typically expressed as annual indices of Catch Per Unit Effort (CPUE) and the indices are generally assumed to be proportional to the abundance of the stock or, at least, the component of the stock that is vulnerable to a fishery (through selectivity).

A review of the stock assessments of New Zealand fish stocks revealed that 53 marine finfish stocks included CPUE indices as either the primary indicator of stock abundance or a key abundance index incorporated in the assessment modelling for the fish stock (Fisheries New Zealand 2018). A disproportionate number of the stocks assessed based primarily on CPUE indices are inshore finfish species, that are predominantly caught in trawl fisheries (such as red gurnard, John dory, flatfish, tarakihi, trevally, red cod, rig). In some regions, these CPUE indices augment trends in stock abundance derived from inshore trawl surveys, while in other areas the CPUE indices represent the primary index of stock abundance.

The prevalence of CPUE indices in stock assessments is due, in part, to the availability of the catch and effort data routinely collected from the fisheries via statutory reporting forms. However, fishing activity is a non-random process, relative to the main species of interest, and may introduce significant biases in the estimation of trends in relative abundance from the resultant catch and effort data. The potential sources of bias in the CPUE indices are generally quite well understood and are primarily related to the degree of targeting (or avoidance) of an individual species, corresponding to changes in the distribution of fishing activity or changes in the efficiency of the fishing fleet.

Generalised linear modelling (GLM) approaches are routinely used to “standardise” catch and effort data to derive annual CPUE indices for a species of interest, incorporating a range of potential explanatory variables that may influence the catchability of the species. Maunder & Punt (2004) provide a comprehensive review of CPUE modelling approaches.

In the past, standardised CPUE analyses have tended to focus on the main species caught in a fishery and modelled the CPUE indices based primarily on the magnitude of the catch of the species (positive catch CPUE indices). Increasingly, CPUE analyses have also been applied to monitor the less abundant species, necessitating the inclusion of CPUE models that also take into account the encounter rate of the species (Lo et al. 1992, Stefansson 1996). These models are termed “combined”, “two-stage” or “hurdle” models as the CPUE indices are derived from two models: a model that estimates the probability of the catch (usually a binomial model), and a model that estimates the magnitude of the (non-zero) catch. Other approaches, such as negative binomial models, zero-inflated negative binomial models (Hilbe 2007) and Tweedie models, are also used to simultaneously model the null and positive catch components (Brodziak & Walsh 2013).

For New Zealand inshore finfish fisheries, procedures have been developed for the processing of catch and effort data prior to the inclusion of these data in a standardised CPUE analysis (Starr 2007). These procedures primarily apportion the total landed catch of a species from the fishing trip amongst fishing event records, usually in proportion to the estimated catch. For the less important species, the reliability of the recording (reporting) of the catches may vary and CPUE indices based solely on positive catch records are likely to be biased (Langley 2015). Potential changes in the reliability of species catch reporting have contributed to the general adoption of two-stage CPUE models by the Inshore Fisheries Working Group. It is also recognised that for the less abundant species the changes in the encounter rate (species occurrence) may be an important indicator of a changes in overall species abundance and should be incorporated in the CPUE modelling approach.

Despite the reliance of CPUE indices for the monitoring of many stocks, little attention has been given to the robustness of the approaches used to derive CPUE indices. A review of the published literature identified a limited number of simulation studies that have investigated the reliability of alternative CPUE indices using different modelling approaches (Campbell et al. 2017, Forrestal et al. 2017, Goodyear 2006, Neubauer 2017, Okamura et al. 2017, Petrere et al. 2010). Those studies generally focused on the derivation of CPUE indices for a single species of interest. However, for trawl fisheries, the catch and effort data are generally reported for multiple species caught from a species assemblage.

Therefore, the recording of a catch of an individual species will be influenced by the reporting requirements for the fishery; for example, for some inshore trawl fisheries only the catches of the five most predominant species caught (top 5) were routinely recorded for each trawl. Therefore, the abundance of an individual species, relative to the abundance of the other species in the assemblage, will influence the frequency of the reporting of catches of the species of interest and, hence, the observed encounter rate of the species. Simulation modelling provides an opportunity to investigate the most appropriate approach to model CPUE indices given the constraints of multi-species reporting and the established procedures used to process catch and effort data. Further, the approach can also be applied to investigate the potential effects of changes in species reporting over time.

A proposal to conduct a CPUE simulation study was developed through the Stock Assessment Methods Working Group. The study was relatively modest in scope and was focussed on the development of a multi-species simulator to investigate the robustness of alternative CPUE modelling approaches and provide recommendations of the most appropriate approaches to determine standardised CPUE indices. The project was funded by Fisheries New Zealand under project code SEA2018-18 CPUE.

## **2. METHODS**

A simulation approach was used to investigate the utility of a range of modelling approaches for standardising catch and effort data to derive CPUE indices of relative abundance. The range of candidate CPUE models included alternative distributions of catch, the modelling of the species occurrence, and the application of two-stage models.

The accuracy of the reporting of catch data for an individual species is likely to be influenced by the abundance and prevalence of the species relative to other species in the composite catches. The simulator was configured to represent the catch reporting procedures typical of New Zealand inshore trawl fisheries, whereby multiple species (up to eight species) may be reported from each individual fishing event (trawl). Further, the simulator also attempts to mimic the approaches used to process the catch and effort prior to the CPUE modelling. The results of the simulations were used to evaluate the performance of each CPUE modelling approach for a range of species categorised according to their relative abundance and occurrence in the catch.

### **2.1 Simulator configuration**

The simulator comprised five elements: species assemblage module, fishery module, reporting module, data processing module and a CPUE analysis module.

#### **Species assemblage module**

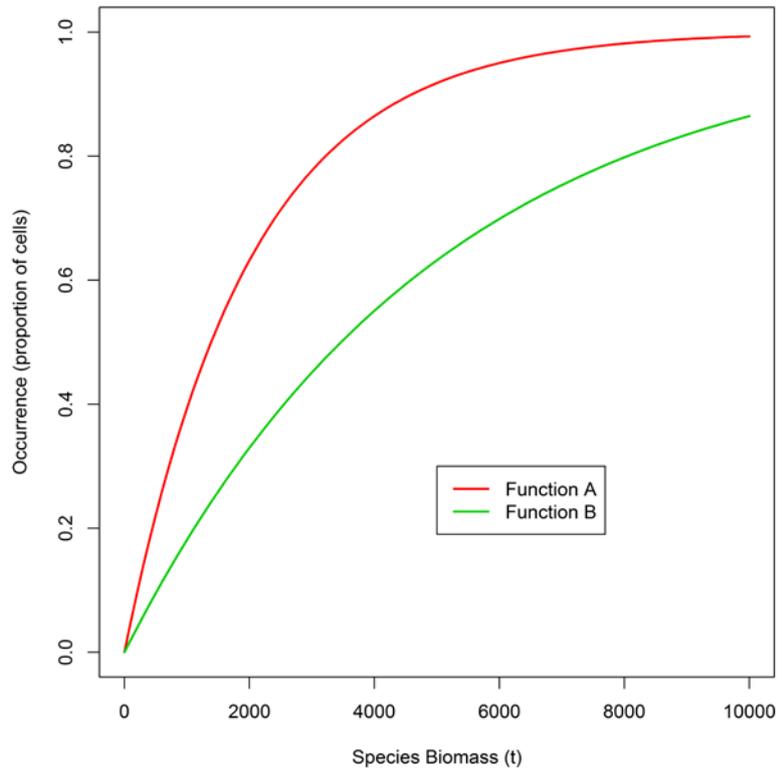
The simulator was configured to include nine (9) fish species that vary in the overall magnitude of abundance and independently vary in abundance over time. The overall scale of abundance amongst species was assumed to increase exponentially from the least abundant (species 9) to the most abundant species (species 1). The abundance of each species varied in accordance with random variation in annual recruitment and a constant rate of total mortality. The simulated catch from the fishery was not removed from the species population; i.e. the simulator was not a fully-integrated multispecies model, but rather was kept at a relatively simple level to facilitate interpretation and understanding. Levels of notional average species abundance ranged from about 1200 t (species 9) to 8000 t (species 1), with coefficients of variation of 8–20%.

The abundance of each species was simulated over a 45 year period with the initial 25 years representing a “burn in” period. The subsequent 20 year period of species abundance represented the simulated time series of species abundance. For each species, the biomass (at year  $yr$ ) was distributed across a spatial grid (matrix  $10 \times 10$  cells) based on a joint distribution of the presence and abundance of the species (at year  $yr$ ). The proportion of cells where the species occurred (at year  $yr$ ) was defined as a function of the abundance of the species (at year  $yr$ ). The occurrence of the species is considered to approach zero at very low levels of abundance and increase with increasing species abundance, although the nature of the relationship will vary depending on the degree of aggregation of the species.

Two alternative formulations of the relationship between species occurrence and species abundance were defined. The first (Function A) was informed by a summary of the catch and effort data from

trawls in the WCNI, Hauraki Gulf and Bay of Plenty inshore trawl fisheries. For the twelve main species reported from the fishery, annual average catch rates and the frequency of positive catches were compared (for 2007/08 to 2016/17). These data showed similar patterns for each of the species examined, with the probability of catch increasing disproportionately relative to the catch rate of the species and approaching a high probability of catch at moderate levels of abundance, as generalised by Function A (Figure 1).

An alternative formulation (Function B) was based on a simple simulation of species occurrence derived from a randomly distributed fish population (Figure 1). The probability of catch increases at a lower rate compared to Function A and reaches a lower maximum occurrence at high biomass levels.



**Figure 1: Assumed relationships between species abundance and the proportion of spatial cells where the species is present (occurrence).**

For each time interval, the proportion of cells where the species occurred (based on Function A or B) was applied to randomly select the individual cells (*rbinom* function) of the spatial grid for the allocation of biomass of the individual species.

The relative proportion of the biomass of the species (at each time interval) was allocated to each selected cell based on a lognormal distribution with a specified standard deviation (mean=0, sd of 0.5 or 1.0). This generated considerable variation in the distribution of species biomass amongst the individual cells. The extent of the variation was determined (in part) by the variation in the catches amongst the individual trawl records and, hence, the explanatory power of the CPUE models (so that  $R^2$  values were similar to those achieved from actual CPUE studies).

The biomass of each species was not allocated specifically to a specific set of cells (of the spatial grid) in a manner that would approximate the habitat distribution of the species. Instead, the spatial allocation of the biomass was simply used as a mechanism for randomly distributing the biomass based on the distributional assumptions.

Four main sets of simulated data (scenarios) were configured based on the combinations of the two functions for determining species occurrence and the two different levels of variation in abundance (Table 1). For these simulated datasets, the fishing effort was randomly allocated across the spatial grid. Two additional sets of simulated data were generated that allocated fishing effort in a non-random

manner either in proportion to the spatial distribution of the biomass of the most abundant species (Species1) or in proportion to the spatial distribution of the biomass of all nine species combined (Table 1).

**Table 1: Configuration of the sets of simulated data (scenarios) generated based on different assumptions regarding species distribution (occurrence and variability) and the distribution of fishing effort amongst cells of the spatial grid. Occurrence functions A and B are as depicted in Figure 1.**

Scenario	Species distribution		Fishing effort distribution
	Occurrence function	Variation	
<i>Scenario 1</i>	A	0.5	Random
<i>Scenario 2</i>	A	1.0	Random
<i>Scenario 3</i>	B	0.5	Random
<i>Scenario 4</i>	B	1.0	Random
<i>TrawlBiomass1</i>	A	0.5	Proportional to Species1 biomass
<i>TrawlBiomassAll</i>	A	0.5	Proportional to Sum(Species1...9) biomass

### Fishery module

The simulator included a fishery module to generate catch and effort data for a simple fisheries configuration. The simulated fishery operated throughout the 20 year period at a constant level of fishing activity with four vessels each conducting 27 trawls per month (three trips of three days duration conducting three trawls per day). Fishing activity was allocated across the spatial grid (100 cells  $x=1:10, y=1:10$ ) at each time interval (year, month). All simulated trawls were of a constant distance/duration.

For each simulation, a set of vessel specific fishing efficiency coefficients ( $Vesselq$ ,  $n=4$ ) was derived from a normal distribution (mean 0, std dev 0.30). The scale of the variation in the vessel coefficients is comparable to the variation in vessel coefficients from GLM standardisations of CPUE data in inshore trawl fisheries. Similarly, for each simulation, species availability varied seasonally, configured using a sine curve with variation in the timing and amplitude of the seasonal effect ( $MonthSpecies\ 1...9$ ).

Each trawl had an associated catch of each species (1 to 9). The magnitude of an individual *trawl* catch of a species by vessel  $v$  at a specific  $x,y$  location ( $loc$ ) in year  $yr$  and month  $m$  was determined by:

$$CatchSpeciesI_{trawl} = BiomassSpeciesI_{yr,loc} * exp(Vesselq_v) * exp(MonthSpeciesI_m)$$

Zero (null) catches were assigned to trawl records with no catch of the species (i.e. from trawls in cells with no biomass).

No additional sources of variation in trawl catches were incorporated in the derivation of the catch and effort data. The main source of variation was attributable to the variation in biomass amongst the individual cells of the spatial grid defined by the species assemblage module.

Fishing activity (trawls) was allocated across the spatial grid ( $x,y$ ) at each time interval (year, month). For the initial scenarios, individual trawls were allocated randomly amongst all the cells of the spatial grid (Table 1). Two additional scenarios were configured based on alternative assumptions for the spatial distribution of trawls at each time interval: in proportion to the biomass of a single species with the highest overall abundance (Species1) or in proportion to the aggregate biomass of all nine species (Table 1).

## Reporting Module

The simulated trawl records were configured in a format to approximate the statutory reporting forms used in the trawl fisheries over the last 10–20 years. Two main form types were considered: the CELR which essentially reports fishing effort and catch of the five main species for each day of fishing and the TCEPR which records individual trawl activity and the catch of the five main species per trawl.

For the TCEPR *trawl* data format, the reporting module derived a data record for each individual fishing event. The species catches from each trawl were ranked by catch weight and the catches of the five largest species catches were recorded for the individual trawl (representing the estimated catches). For the CELR *day* data format, the reporting module aggregated all trawls conducted by a vessel during a fishing day. The daily catches were aggregated by species and ranked by the total catch weight. The daily catches of the five species with the largest catches were recorded for the specific fishing day.

For each fishing trip, the aggregated catch of each species from the trip was also determined, representing the landed catch records for the trip. All simulated trawl catches, daily catches and landed catches were reported without error.

## Data processing module

The simulator included a data processing module to approximate the established protocols used to formulate CPUE data sets from the reported catch and effort data (Starr 2007). The data processing primarily involves the apportionment of the landed species catches from a fishing trip amongst the associated fishing records with reported catches of the species (catch based allocation). For those species with no corresponding reported catch records, the landed catch of the species is apportioned equally amongst the associated fishing effort records (effort based allocation). Typically, the effort based allocation generates effort records with numerous small catches of a species and no zero-catch records (for the fishing trip).

The data processing was conducted for the *trawl* and *day* catch and effort data sets separately. The resultant allocated catches represent the species catches that are included in the CPUE modelling. For each species, the proportion of the catch records generated by either allocation method was summarised for each simulated data set.

## CPUE analysis module

For each simulation, a Generalised Linear Modelling (GLM) approach, implemented in the *R* statistical software package, was applied to the catch and effort records for each species (1 to 9) to derive sets of CPUE indices (following Venables & Ripley 1999). The standardised CPUE models had the generic format:

$$\text{SpeciesCatch}_{\text{trawl}} \sim \text{FishingYear}_{\text{yr}} + \text{Vessel}_{\text{v}} + \text{Month}_{\text{m}} + \text{Error}_{\text{trawl}}$$

where the three explanatory variables were included in the models as categorical variables. All trawls were of a constant distance/duration and, hence, no additional effort variables were incorporated in the CPUE models.

A range of different model options were evaluated using various distributional assumptions (Table 2). For each model option, the CPUE indices (for each Species 1...9 and Simulation 1...100) were derived from individual *FishingYears* using the *predict* function and normalised to the average of the 20 year time series. CPUE indices were also derived from the two-stage models which combined the binomial indices with the indices from the Log-normal (delta-lognormal), Gamma (delta-Gamma) and Weibull (delta-Weibull) models (Lo et al. 1992, Stefansson 1996).

Initially, the range of candidate CPUE model options also included a Zero-inflated negative binomial (ZipNB) model. However, problems were encountered in attaining an adequate level of convergence for these models and the ZipNB model was excluded from the final suite of CPUE models evaluated.

**Table 2: The range of model options included in the set of CPUE models evaluated.**

Distribution	Link function	Dependent variable
Binomial	logit	Zero catch (0); positive catch (1)
Log-normal	log	Catch, excluding zeroes
Gamma	log	Catch, excluding zeroes
Weibull	log	Catch, excluding zeroes
Negative binomial	log, theta=2	Catch, incl. zero
Tweedie (compound Poisson distribution)	tweedie(var.power = 1.5)	Catch, incl. zero

### Summary Statistics

For each Scenario (see Table 1), a set of 100 simulations was conducted. Each set of simulations included the derivation of CPUE indices for the nine species from the nine CPUE model options.

For each CPUE model, the proportion of the total deviance (in species catch) explained by the CPUE model was determined. The individual sets of *CPUE* indices derived from each model were compared to the time series of biomass for the species, normalised for the 20 year period (*NormBiomass*).

The performance of each CPUE model was determined from the correlation coefficient and the slope of the linear relationship (*b*) between the predicted and “actual” simulated species abundance. The correlation coefficients reveal the extent of the deviation in the CPUE indices from the annual species abundance, while the slope parameter reveals the proportionality of the relationship between the CPUE indices and the species abundance. A slope parameter of less than 1.0 indicates that the CPUE indices vary less than the species abundance (hyper-stability), while a slope of greater than 1.0 indicates that the scale of the variation in CPUE indices is greater than the variation in species abundance (e.g. hyper-depletion).

For presentation purposes, the simulation results were aggregated by species and CPUE modelling approach.

### Simulated data sets

The simulation data sets were characterised using a range of metrics for each species (combined for all simulations). These metrics are similar to summary statistics used to characterise the catch and effort data included in actual CPUE analyses. The specific metrics are as follow.

- Average rank of species in the total catch (*TripRank*).
- Average rank of species in individual reported (estimated) catches (where catch > 0) (*CatchRank*).
- Proportion of positive catch records (proportion non-zero) (*PropCatch*).
- Proportion of landed catch allocated to estimated catches (allocation flag 1) (*PropCatch1*).
- Proportion of landed catch allocated based on effort (allocation flag 2) (*PropCatch2*).
- Proportion of positive catch records derived from estimated catches (allocation flag 1) (*PropRecord1*).
- Proportion of positive catch records derived from allocation based on effort (allocation flag 2) (*PropRecord2*).

For the *trawl* based data, the rank of the individual species in the total catches (*TripRank*) is consistent with the simulated level of biomass for each species (Table 3). However, the rankings of the catches of the species in the individual catches (*CatchRank*) tend to be lower than the overall magnitude of the abundance of the species, especially for the species below the top tier of species (Species 1–3). This relates to the lower proportion of positive reported catches of these species and hence the individual species are only reported when they occur amongst the five largest catches. Increasing the variation in species biomass (SD = 0.5 or 1.0) resulted in a lower proportion of positive reported catches for all species, although the effect was greater for the lower tier species (Species 7–9) (Table 3).

For all species, most of the landed catch is allocated to the individual trawl records based on trawl catches; i.e. catch based allocation (Table 3). However, for the lower tier species (Species 7–9) a

considerable proportion of the trawl records with a positive catch were derived via the effort based allocation (i.e. trips with no positive catches reported for the species) (Table 3).

The aggregation of trawl catches in the *day* format yields a higher proportion of positive catches compared to the *trawl* data set (Table 4). However, for the lower tier species (Species 7–9) a higher proportion of the species catches were derived from the effort based allocation of landed catch, related to the lower likelihood of reporting catches of the lower ranked species amongst the five main species caught during a day. Correspondingly, a substantial proportion of the positive catch records (number of records) for the lower tier species were derived based on the effort allocation of landed (trip) catch. Further, a considerable proportion of the intermediate tier species (Species 4–6) were also derived from the effort allocation of catch (Table 4).

**Table 3: Summary statistics for the *Trawl* based simulation data sets derived for each of the main data configuration scenarios. The data scenarios are defined in Table 1 and the summary metrics and described in Section 3.1.**

								Scenario 1
Species	TripRank	CatchRank	PropCatch	PropCatch1	PropCatch2	PropRecord1	PropRecord2	
1	1.5	0.9	0.846	0.999	0.001	0.984	0.016	
2	2.2	1.4	0.787	0.998	0.002	0.976	0.024	
3	3.2	1.8	0.721	0.997	0.003	0.970	0.030	
4	4.0	2.0	0.652	0.994	0.006	0.949	0.051	
5	4.9	2.2	0.580	0.988	0.012	0.905	0.095	
6	5.9	2.4	0.534	0.975	0.025	0.834	0.166	
7	7.1	2.6	0.481	0.963	0.037	0.771	0.229	
8	7.7	2.5	0.478	0.943	0.057	0.652	0.348	
9	8.5	2.6	0.468	0.918	0.082	0.535	0.465	
								Scenario 2
Species	TripRank	CatchRank	PropCatch	PropCatch1	PropCatch2	PropRecord1	PropRecord2	
1	1.5	1.0	0.729	1.000	0.000	1.000	0.000	
2	2.5	1.2	0.662	1.000	0.000	0.999	0.001	
3	3.2	1.6	0.587	1.000	0.000	0.997	0.003	
4	3.9	1.7	0.513	1.000	0.000	0.991	0.009	
5	5.1	1.8	0.448	0.999	0.001	0.980	0.020	
6	5.8	1.9	0.392	0.998	0.002	0.962	0.038	
7	6.8	1.9	0.344	0.996	0.004	0.922	0.078	
8	7.8	1.9	0.303	0.993	0.007	0.873	0.127	
9	8.5	2.0	0.269	0.989	0.011	0.810	0.190	
								Scenario 3
Species	TripRank	CatchRank	PropCatch	PropCatch1	PropCatch2	PropRecord1	PropRecord2	
1	1.4	1.4	0.792	1.000	0.000	0.995	0.005	
2	2.3	1.7	0.737	1.000	0.000	0.996	0.004	
3	3.0	1.8	0.679	0.999	0.001	0.989	0.011	
4	4.0	1.9	0.607	0.999	0.001	0.979	0.021	
5	5.1	2.0	0.545	0.998	0.002	0.964	0.036	
6	6.2	2.3	0.488	0.996	0.004	0.944	0.056	
7	6.8	2.3	0.452	0.991	0.009	0.858	0.142	
8	7.8	2.5	0.419	0.984	0.016	0.779	0.221	
9	8.4	2.6	0.400	0.979	0.021	0.708	0.292	
								Scenario 4
Species	TripRank	CatchRank	PropCatch	PropCatch1	PropCatch2	PropRecord1	PropRecord2	
1	1.5	1.0	0.724	1.000	0.000	1.000	0.000	
2	2.3	1.3	0.657	1.000	0.000	0.999	0.001	
3	3.0	1.7	0.586	1.000	0.000	0.998	0.002	
4	4.1	2.0	0.511	1.000	0.000	0.994	0.006	
5	4.9	2.0	0.447	1.000	0.000	0.985	0.015	
6	5.9	2.0	0.391	0.999	0.001	0.968	0.032	
7	6.8	2.0	0.340	0.998	0.002	0.936	0.064	
8	7.9	2.0	0.298	0.997	0.003	0.896	0.104	
9	8.6	2.0	0.266	0.995	0.005	0.853	0.147	

**Table 4: Summary statistics for the *Day* based simulation data sets derived for each of the main data configuration scenarios. The data scenarios are defined in Table 1 and the summary metrics and described in Section 3.1.**

Scenario 1							
Species	TripRank	CatchRank	PropCatch	PropCatch1	PropCatch2	PropRecord1	PropRecord2
1	1.5	0.8	0.935	0.992	0.008	0.932	0.068
2	2.2	1.2	0.898	0.984	0.016	0.908	0.092
3	3.2	1.7	0.859	0.970	0.030	0.872	0.128
4	4.0	2.1	0.819	0.946	0.054	0.809	0.191
5	4.9	2.3	0.787	0.903	0.097	0.687	0.313
6	5.9	2.4	0.785	0.841	0.159	0.558	0.442
7	7.1	2.8	0.772	0.761	0.239	0.434	0.566
8	7.7	2.9	0.799	0.700	0.300	0.329	0.671
9	8.5	3.1	0.825	0.600	0.400	0.235	0.765
Scenario 2							
Species	TripRank	CatchRank	PropCatch	PropCatch1	PropCatch2	PropRecord1	PropRecord2
1	1.5	1.1	0.879	0.994	0.006	0.950	0.050
2	2.5	1.6	0.837	0.991	0.009	0.939	0.061
3	3.2	1.7	0.793	0.980	0.020	0.885	0.115
4	3.9	1.8	0.759	0.966	0.034	0.804	0.196
5	5.1	2.2	0.722	0.943	0.057	0.726	0.274
6	5.8	2.4	0.701	0.919	0.081	0.641	0.359
7	6.8	2.4	0.688	0.884	0.116	0.534	0.466
8	7.8	2.6	0.675	0.838	0.162	0.440	0.560
9	8.5	2.6	0.659	0.797	0.203	0.360	0.640
Scenario 3							
Species	TripRank	CatchRank	PropCatch	PropCatch1	PropCatch2	PropRecord1	PropRecord2
1	1.4	1.1	0.876	0.995	0.005	0.951	0.049
2	2.3	1.5	0.834	0.990	0.010	0.932	0.068
3	3.0	1.7	0.803	0.981	0.019	0.877	0.123
4	4.0	1.9	0.759	0.966	0.034	0.814	0.186
5	5.1	2.3	0.727	0.944	0.056	0.738	0.262
6	6.2	2.6	0.703	0.909	0.091	0.641	0.359
7	6.8	2.5	0.731	0.873	0.127	0.505	0.495
8	7.8	2.6	0.745	0.822	0.178	0.396	0.604
9	8.4	2.7	0.754	0.775	0.225	0.321	0.679
Scenario 4							
Species	TripRank	CatchRank	PropCatch	PropCatch1	PropCatch2	PropRecord1	PropRecord2
1	1.5	1.2	0.851	0.997	0.003	0.965	0.035
2	2.3	1.7	0.801	0.994	0.006	0.952	0.048
3	3.0	1.7	0.759	0.990	0.010	0.902	0.098
4	4.1	2.0	0.710	0.981	0.019	0.849	0.151
5	4.9	2.1	0.684	0.968	0.032	0.760	0.240
6	5.9	2.3	0.662	0.951	0.049	0.680	0.320
7	6.8	2.5	0.648	0.931	0.069	0.576	0.424
8	7.9	2.6	0.634	0.906	0.094	0.488	0.512
9	8.6	2.8	0.618	0.876	0.124	0.421	0.579

### 3. RESULTS

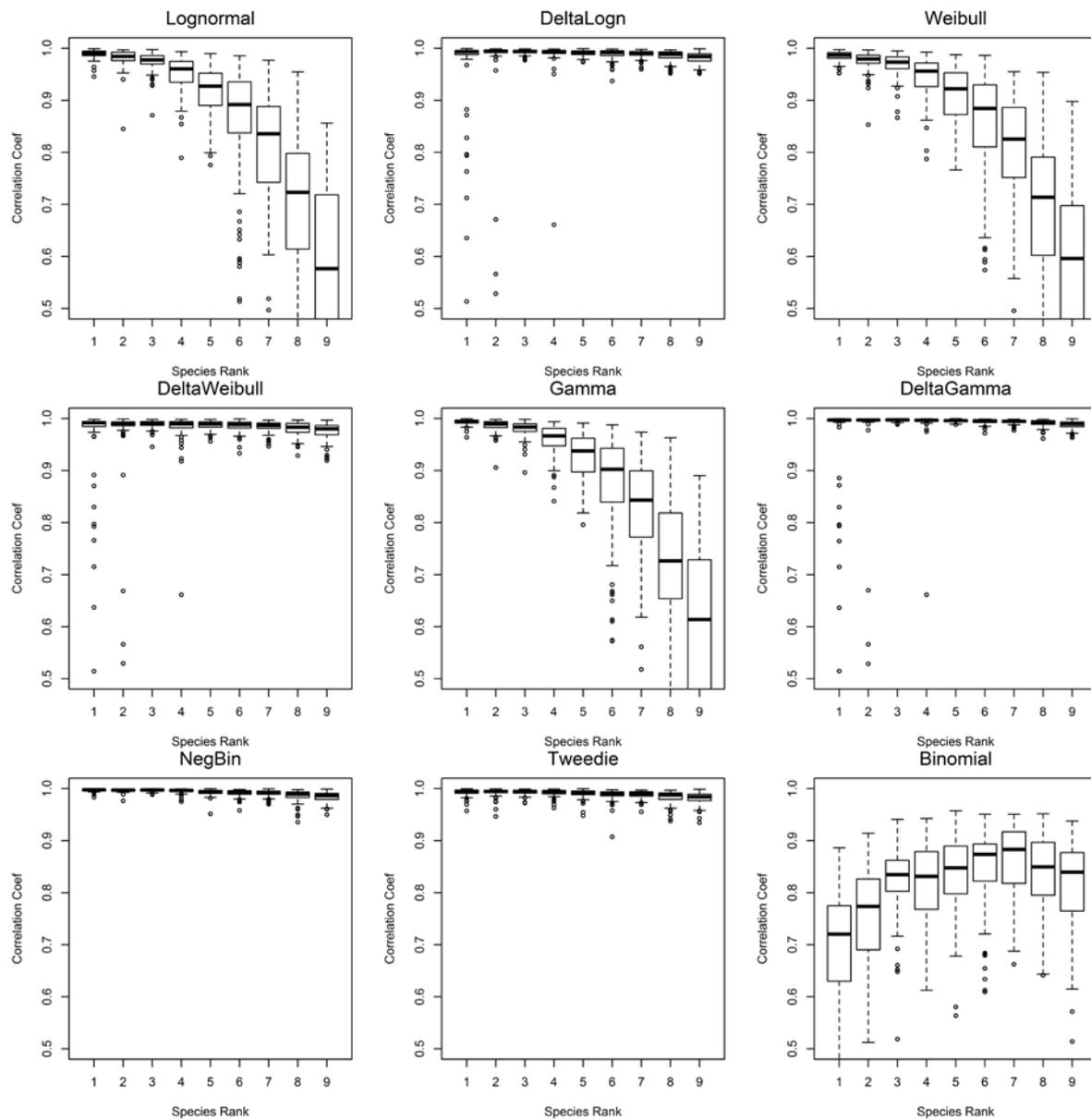
#### 3.1 Simulator evaluation

Initially, the performance of the simulator was evaluated by generating data sets without the inclusion of the reporting module. This generated trawl catches for all nine species that were not influenced by the constraints related to reporting catches of only the five most abundant species in the catch. The expectation was that the resulting species specific CPUE models would be consistent with the basic structure of the data set and, hence, generate CPUE indices that were consistent with the trends in species biomass.

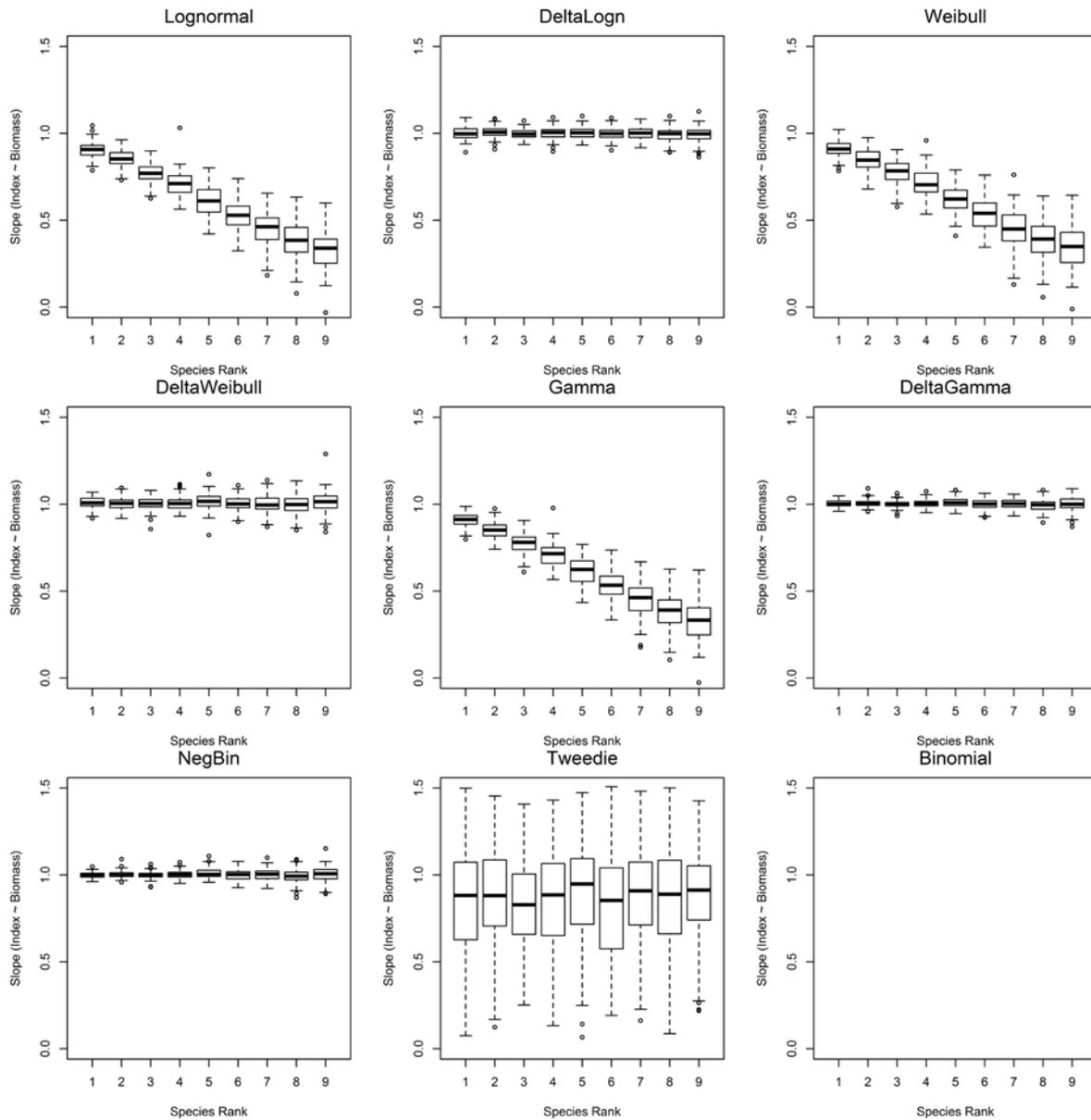
The positive catch models (lognormal, Weibull, Gamma) yielded CPUE indices that were accurate for the most abundant species, with correlation coefficients (Figure 2) and slope parameters (Figure 3) approaching 1.0 for Species Rank 1. However, the positive catch CPUE indices were increasingly biased for the less abundant species, as indicated by the decline in the correlation coefficients (Figure 2) and slope parameters (Figure 3) for the lower ranked species. This corresponds with the simulated spatial distribution of each species, with a lower proportion of cells assigned biomass for the less abundant species and hence a lower probability of catching those species. An index composed of only the positive catches of the species will not incorporate the change in the encounter rate of the species and will result in hyperstability of the derived indices (slope parameters less than 1.0) (Figure 3).

Conversely, the model options that incorporated the zero-catch component yielded CPUE indices that were consistent with the biomass trends for all species; i.e. correlation coefficients approaching 1.0 and slopes between the CPUE indices and biomass that approximated 1.0 (Figure 2 and Figure 3). Overall, the delta-lognormal, delta-Weibull, delta-Gamma and Negative binomial models yielded CPUE indices that closely represent the underlying biomass trends.

These preliminary results indicated that the simulator was performing in accordance with the underlying assumptions regarding the distribution of biomass and provided an appropriate platform for evaluating more complex data sets (including the reporting module).



**Figure 2: Boxplots of the correlation coefficients between the CPUE indices and simulated species abundance for each of the 9 species (ranked 1 to 9, most abundant to least abundant) for each CPUE modelling approach. The simulated data sets ( $n = 100$ ) were configured for initial evaluation of the simulator, as follows: data format trawl, species occurrence function A (Figure 1), variation in species distribution SD 0.5, trawl distribution random, reporting module excluded.**



**Figure 3: Boxplots of the slope ( $b$ ) of the linear relationship between the CPUE indices and simulated species abundance for each of the 9 species (ranked 1 to 9, most abundant to least abundant) for each CPUE modelling approach. The simulated data sets ( $n = 100$ ) were configured for initial evaluation of the simulator, as follows: data format trawl, species occurrence function A (Figure 1), variation in species distribution SD 0.5, trawl distribution random, reporting module excluded.**

### 3.2 CPUE simulation modelling – Trawl based data sets

The simulator was applied to evaluate the reliability of CPUE models derived from trawl-based data simulated, including the reporting module, for each Scenario prescribed in Table 1. For presentation purposes, detailed results are described for a single scenario (Scenario 1) and any key differences in the results from the other data scenarios are highlighted. Summary results from all (6) scenarios are also presented in Appendix 1 (Tables A1–A4).

The positive catch CPUE models (lognormal, Weibull and gamma) of the *trawl* based data (Scenario 1) accounted for approximately 30–50% of the total variation in the trawl catches of each species (Figure 4). These simulated results are broadly consistent with the proportion of the variation explained in the analysis of trawl based CPUE data from inshore trawl fisheries (for example, SNA 8  $R^2$  0.37, JDO 1 WCNI  $R^2$  0.372, JDO 1 HG-ENLD  $R^2$  0.441, JDO 1 BPLE  $R^2$  0.291). By comparison, the CPUE models that include the zero-catch component (binomial, negative binomial and Tweedie) explained a considerably lower proportion of the total variance in catch (Figure 4).

For the most abundant species (Species 1 to 3), there was a strong correlation between the CPUE indices for all model options (corr coef close to 1.0) indicating that the CPUE models reliably estimated the annual variation in species biomass (Figure 5). However, for the positive catch models (lognormal, Weibull and Gamma), the slope of the relationship between the CPUE indices and biomass of the most abundant species was considerably less than unity (slope = 1), indicating the CPUE models consistently under-estimated the overall scale of the trend in relative abundance (Figure 6).

For the least abundant species (Species 5 to 9), the positive catch models (lognormal, Weibull and Gamma) poorly estimated the variation (Figure 5) and trend (Figure 6) in species abundance.

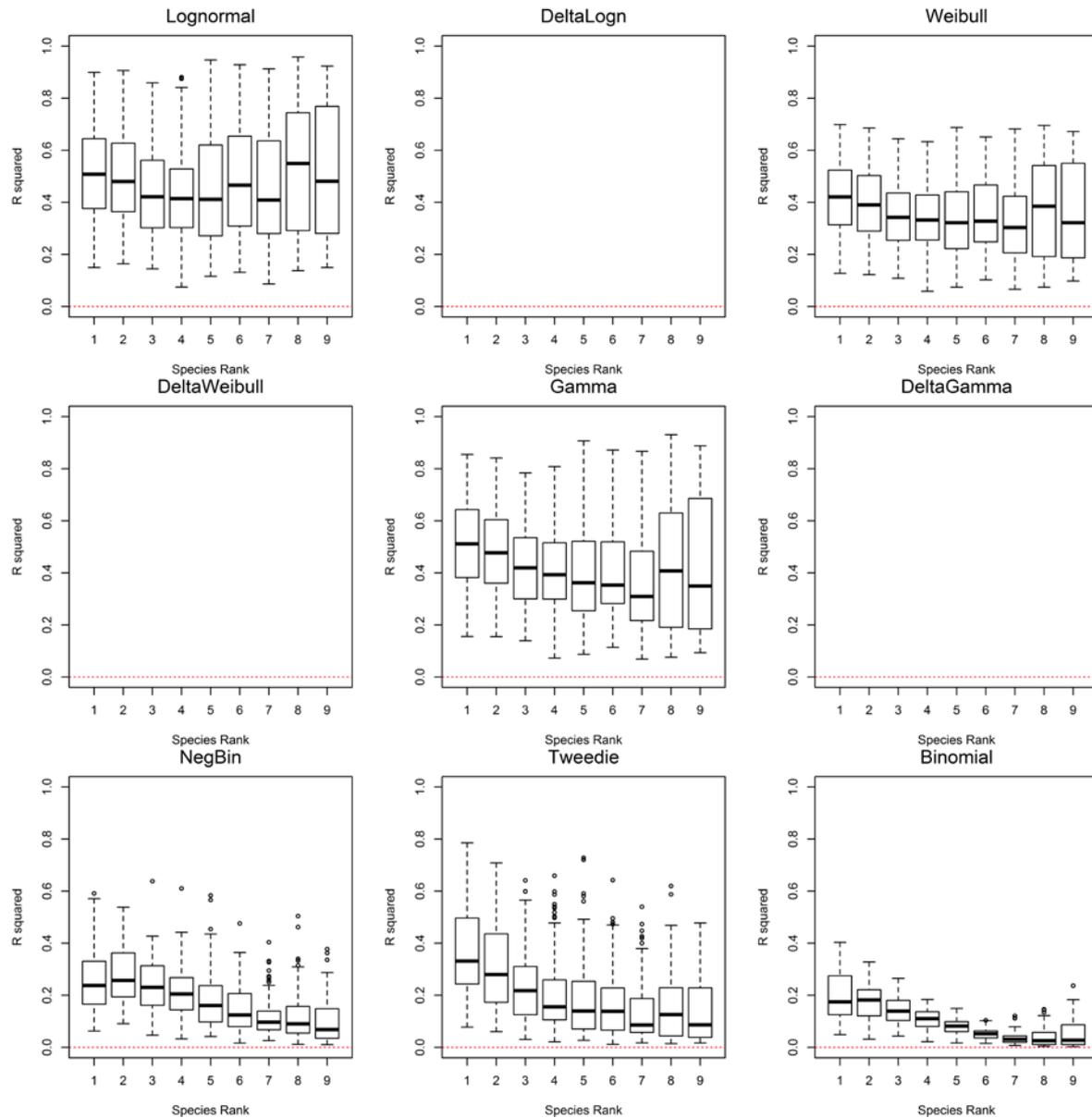
For the full range of species, the hurdle models (delta-models and negative binomial models) yielded the most reliable estimates of trends in species abundance (i.e. correlation coefficients and slope parameters approximating 1.0) (Figure 6). However, for the less abundant species (Species 5 to 9), the delta-lognormal models tended to over-estimate the scale of the variation in the abundance trend (slope parameters greater than 1.0) (Figure 6). The Tweedie models tended to perform relatively poorly across the range of species (median slope parameters approximately 0.85).

The differences in the performance of the delta-lognormal models amongst species tended to be more pronounced for the data sets with higher variation in the positive catches and with a higher proportion of zero catches (Occurrence function B; Figure 1) incorporated in the simulated data sets (Figure 7). For the four data scenarios, the delta-lognormal models performed well for the four most abundant species, but model performance deteriorated for the less abundant species, especially with a higher proportion of zero catches (Occurrence function B; Figure 1) incorporated in the simulated data sets. In general, the performance of the delta-lognormal models was lower for data sets that included zero-catch records for more than 50% of the records (see Table 3). These species data sets were also composed of at least 10% of positive catch records that were generated by the allocation of catches amongst the effort records (effort-based allocation) rather than amongst recorded catches (catch-based allocation).

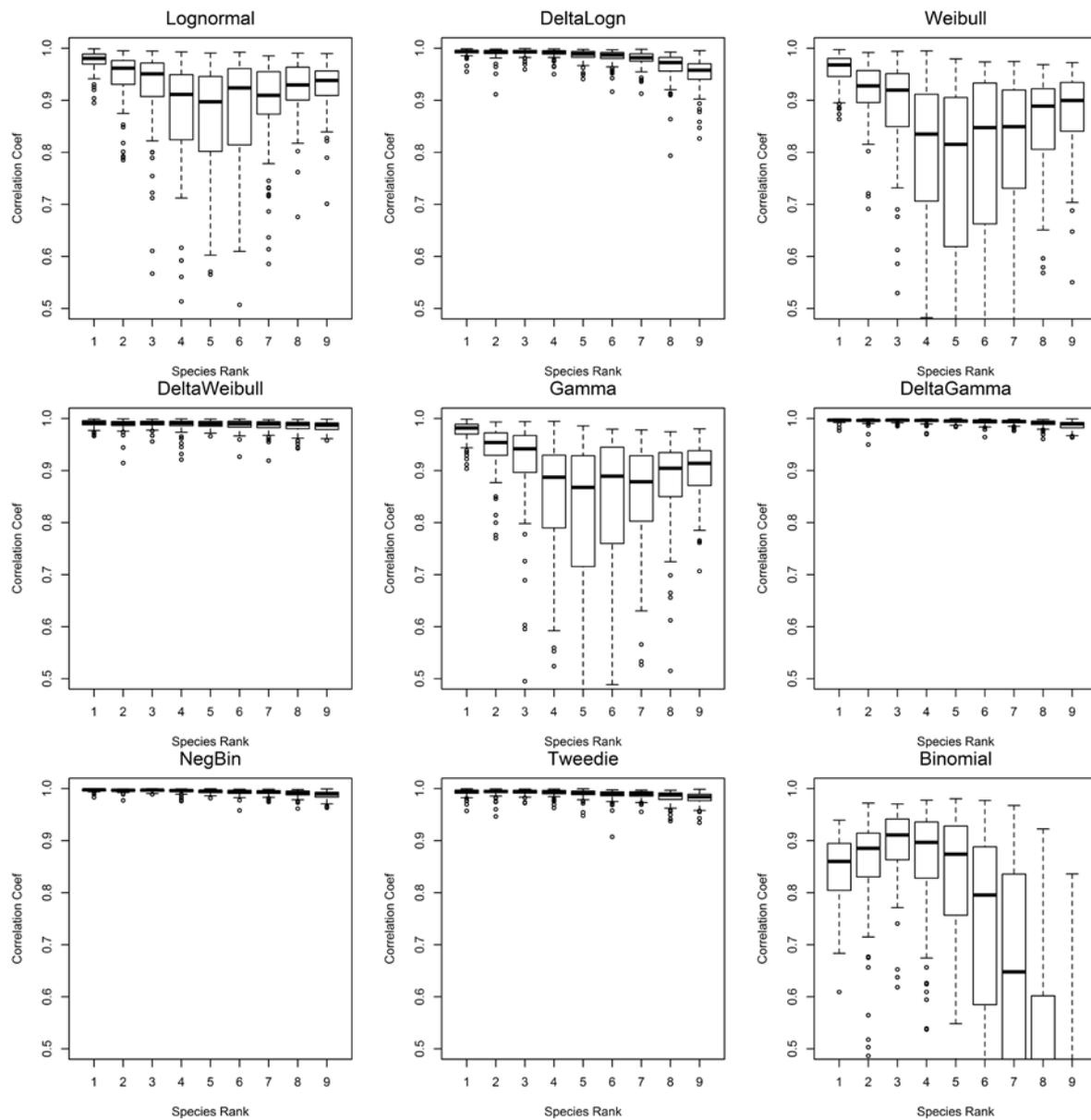
Overall, the Negative binomial, delta-Weibull, delta-Gamma models performed well for the range of species (1 to 9) and for the four different data configurations with correlation coefficients and slope metrics of approximately 1.0 (Figure 7). For these model options, there was only a small deterioration in model performance for the lowest ranked species (8 and 9) for the data configurations with the higher proportion of zero catches (correlation coefficients less than 0.95, slope approximately 1.0) (Appendix 1 Tables A1, A2).

Three alternative mechanisms for allocating trawl effort were considered: random allocation of effort (Scenario 1), the allocation of effort in proportion to the abundance of a single (target) species (Species 1) (Scenario *TrawlBiomass1*) and the allocation in proportion to the total abundance of all species (Species 1 to 9) (Scenario *TrawlBiomassAll*). The comparative models revealed that the allocation of effort in proportion to the abundance of a single (target) species introduced a degree of hyper-stability (slope parameters less than 1.0) in the relationship between the CPUE indices and the biomass of the species of interest (target species) for all model options (Figure 8, Appendix 1 Tables A3 and A4). An additional set of trials revealed that the biases in the CPUE indices were not ameliorated by reweighting

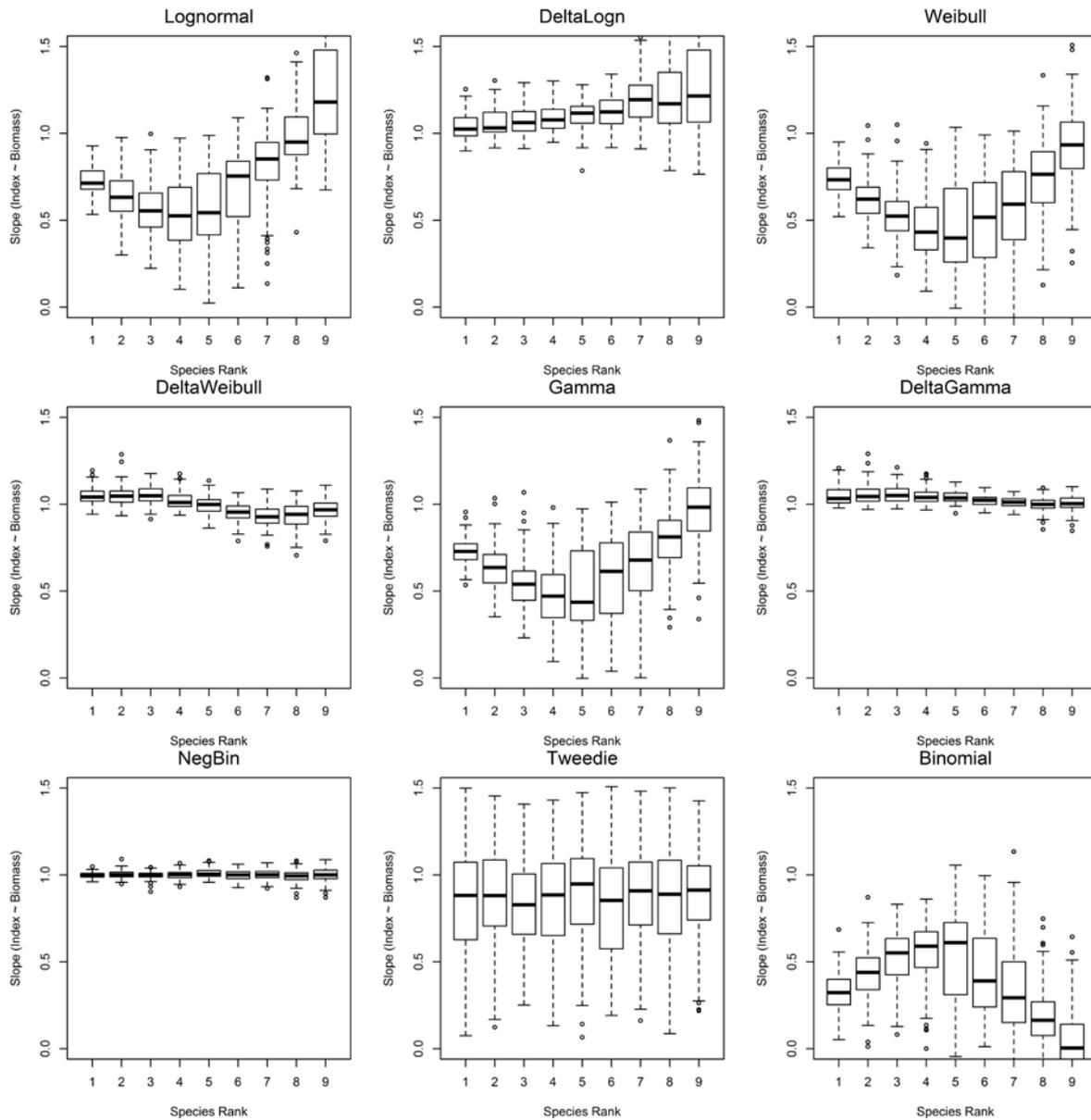
of the data records; i.e. weighting each record inversely proportional to the total number of records in each x, y, cell per year.



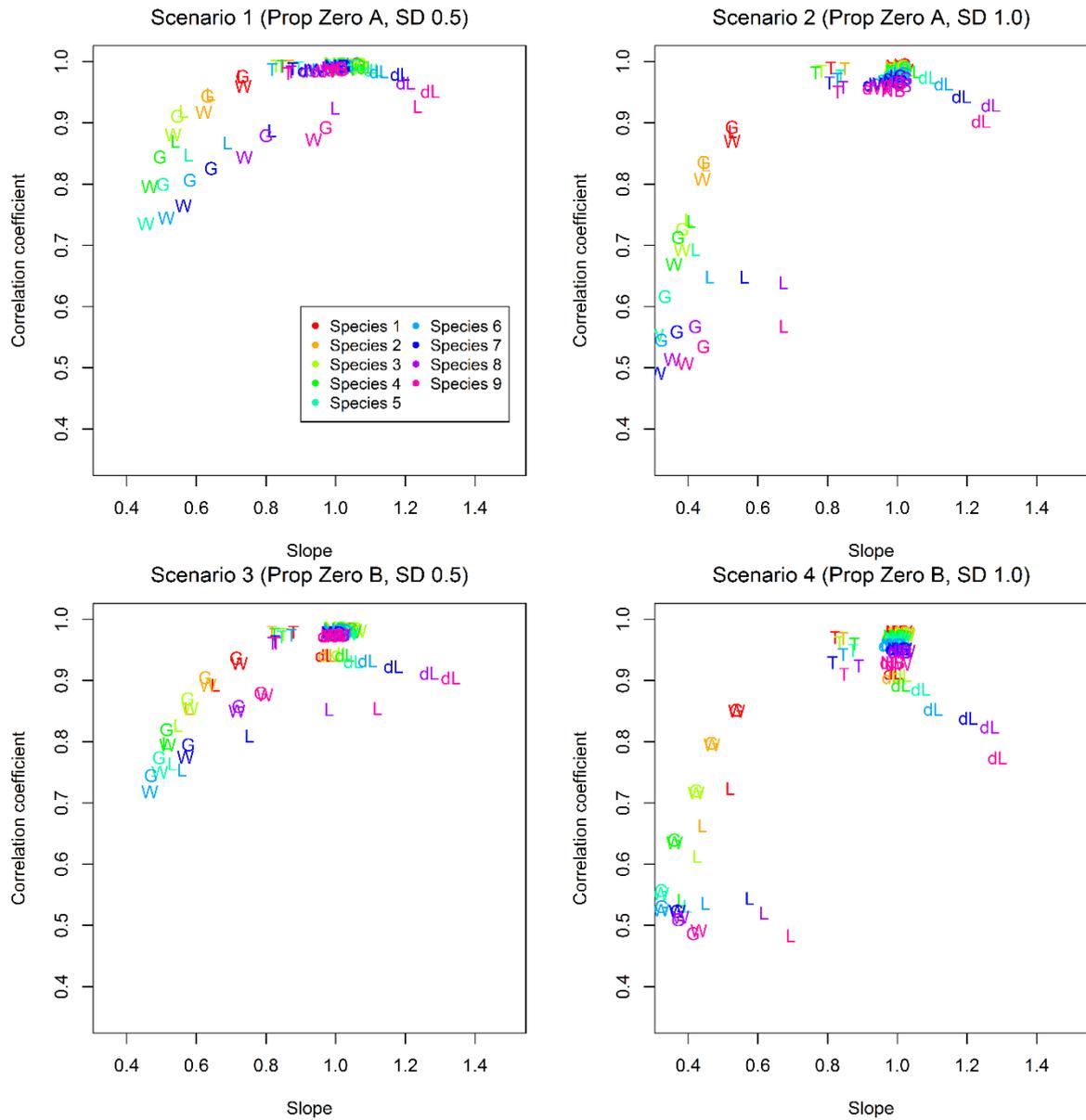
**Figure 4: Boxplots of the  $R^2$  statistics from each CPUE modelling approach for each of the 9 species (ranked 1 to 9, most abundant to least abundant) from the simulated *trawl* data sets ( $n = 100$ ) for Scenario 1; i.e., species occurrence function A (Figure 1), variation in species distribution SD 0.5, trawl distribution random, reporting module included.**



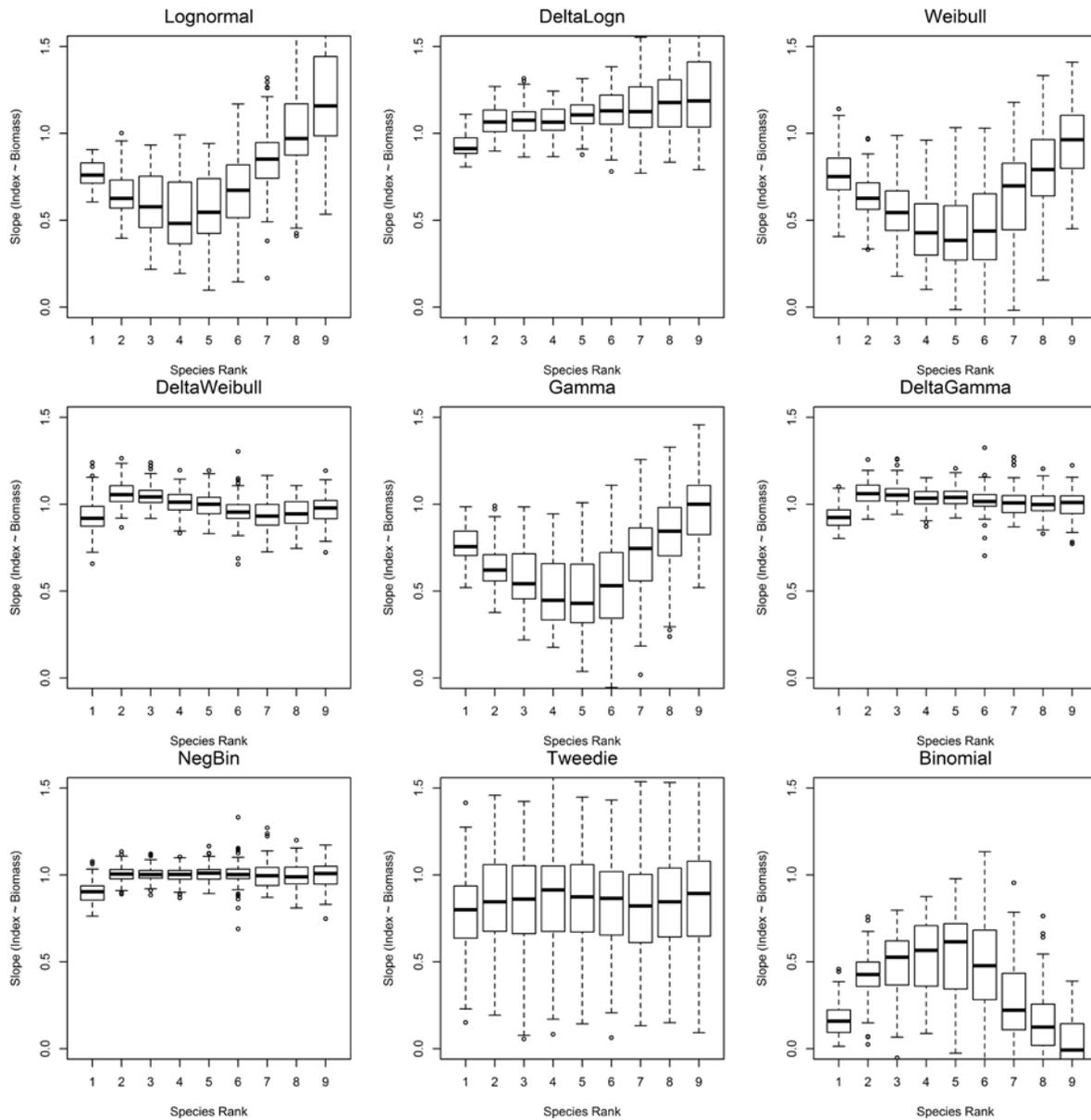
**Figure 5: Boxplots of the correlation coefficients between the CPUE indices and simulated species abundance for each of the 9 species (ranked 1 to 9, most abundant to least abundant) for each CPUE modelling approach. The simulated *trawl* data sets ( $n = 100$ ) for Scenario 1; i.e., species occurrence function A (Figure 1), variation in species distribution SD 0.5, trawl distribution random, reporting module included.**



**Figure 6: Boxplots of the slope ( $b$ ) of the linear relationship between the CPUE indices and simulated species abundance for each of the 9 species (ranked 1 to 9, most abundant to least abundant) for each CPUE modelling approach. The simulated *trawl* data sets ( $n = 100$ ) for Scenario 1; i.e., species occurrence function A (Figure 1), variation in species distribution SD 0.5, trawl distribution random, reporting module included.**



**Figure 7: Average slope and correlation coefficients (from 100 simulations) for each species (1 to 9, colour coded) from selected *trawl* based CPUE models (L, lognormal; dL, Delta-lognormal; W, Weibull; dW, delta-Weibull; G, Gamma; dG, Delta-Gamma; T, Tweedie; NB, negative binomial) and different data configurations (panels).**



**Figure 8: Boxplots of the slope ( $b$ ) of the linear relationship between the CPUE indices and simulated species abundance for each of the 9 species (ranked 1 to 9, most abundant to least abundant) for each CPUE modelling approach. The simulated *trawl* data sets ( $n = 100$ ) for Scenario 1; i.e., species occurrence function A (Figure 1), variation in species distribution SD 0.5, trawls distributed relative to the abundance of Species 1, reporting module included.**

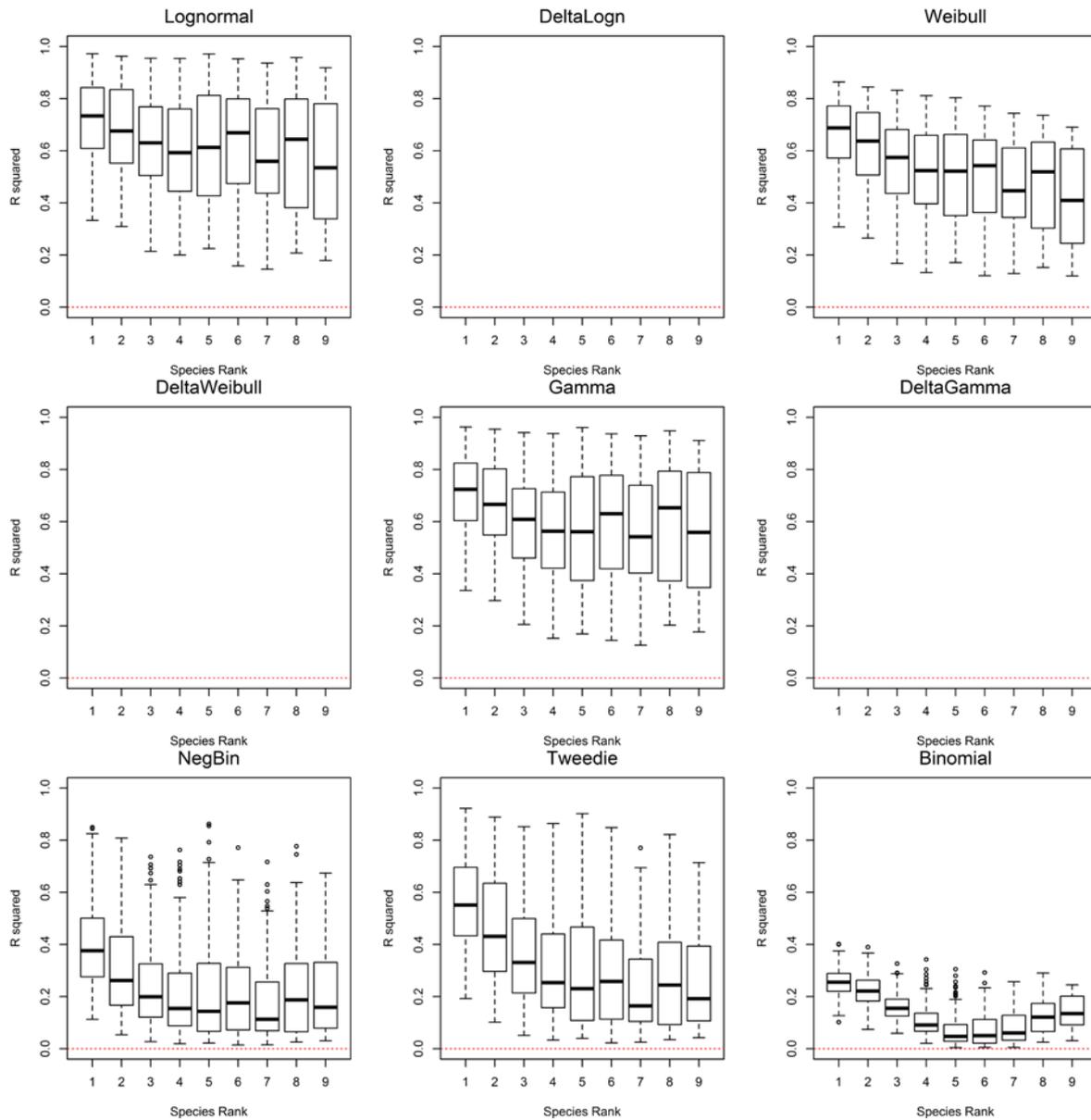
### 3.3 CPUE simulation modelling – Day based data sets

The CPUE modelling was repeated for the simulated data sets (specified in Table 1) with the data reformatted in the *day* summary format. For comparison with the *trawl* based simulations, detailed results are presented for data Scenario 1 and summarised results of the range of scenarios are presented in Appendix 1 (Tables A5–A8).

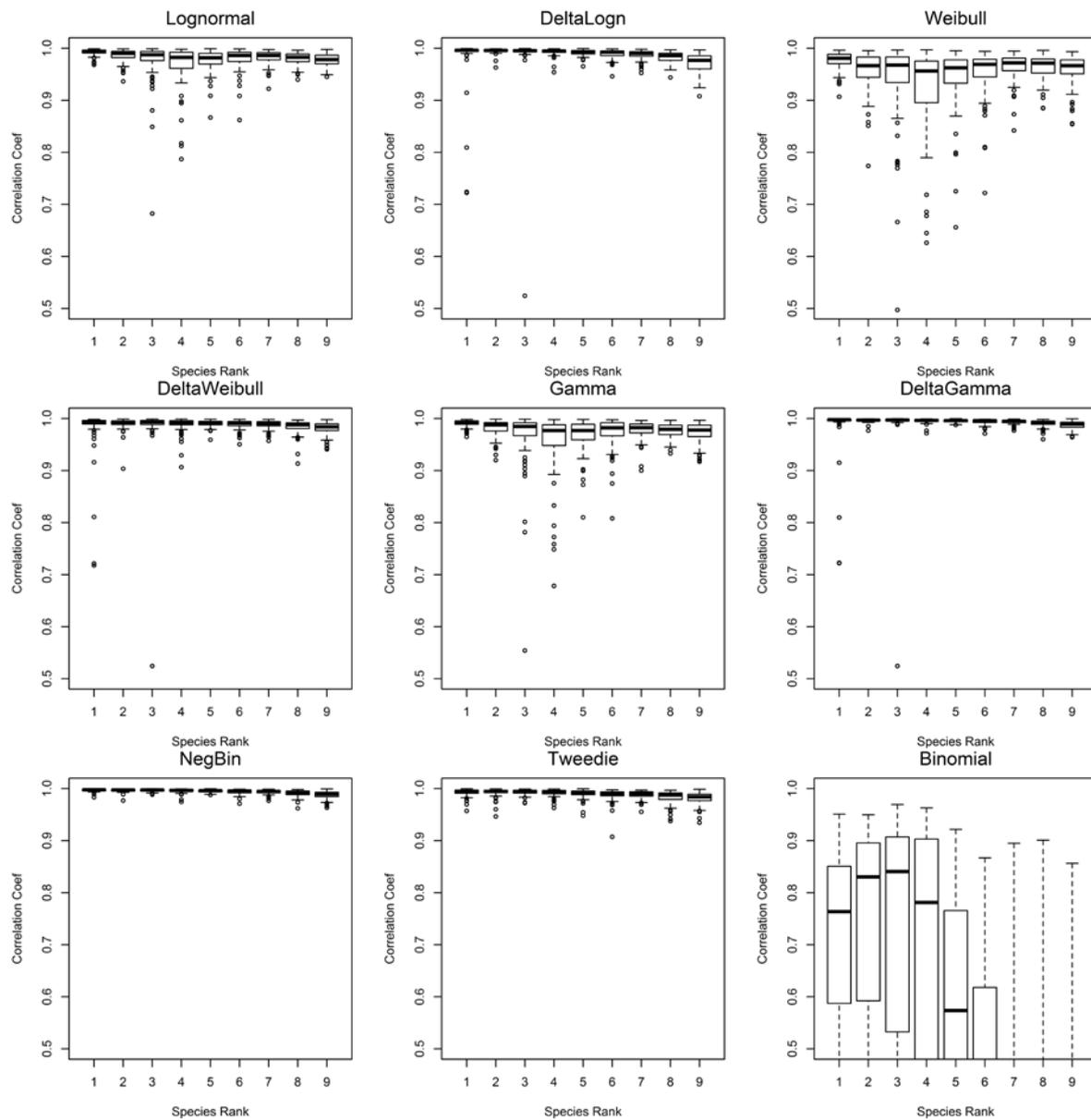
The positive catch CPUE models (lognormal, Weibull and gamma) of the *day* based data (Scenario 1) accounted for approximately 50–70% of the total variation in the daily catches of each species (Figure 9). The proportion of the variation explained is considerably higher than for the corresponding *trawl* based CPUE models due to the aggregation of the individual trawls records. However, these simulated results also explain a considerably higher proportion of the variation than that typically explained in the analysis of *day* based CPUE data from inshore trawl fisheries (for example, TAR 1 BPLE  $R^2$  0.401, TAR 2  $R^2$  0.551, TAR 3  $R^2$  0.373, GUR 2). This suggests that the simulated data set (Scenario 1) does not adequately represent the variation in the daily catches observed in the inshore commercial fisheries. However, the CPUE models derived from the other data scenarios (2–4) yielded  $R^2$  values comparable to the results of the published studies.

In general, the CPUE modelling of *day* aggregated catch and effort records yielded results that were comparable to the results from the modelling of the *trawl* based data. Most of the CPUE models yielded indices that were well correlated with the biomass trends for each species, with the exception of the binomial model (Figure 10). The lognormal, Weibull and Gamma positive catch models tended to under-estimate the scale of the trend in relative biomass for the main species (Species 1–5) (slope parameter less than 1.0) and over-estimate the scale of the trend in relative biomass for the lesser species (Species 7–9) (slope parameter greater than 1.0) (Figure 11).

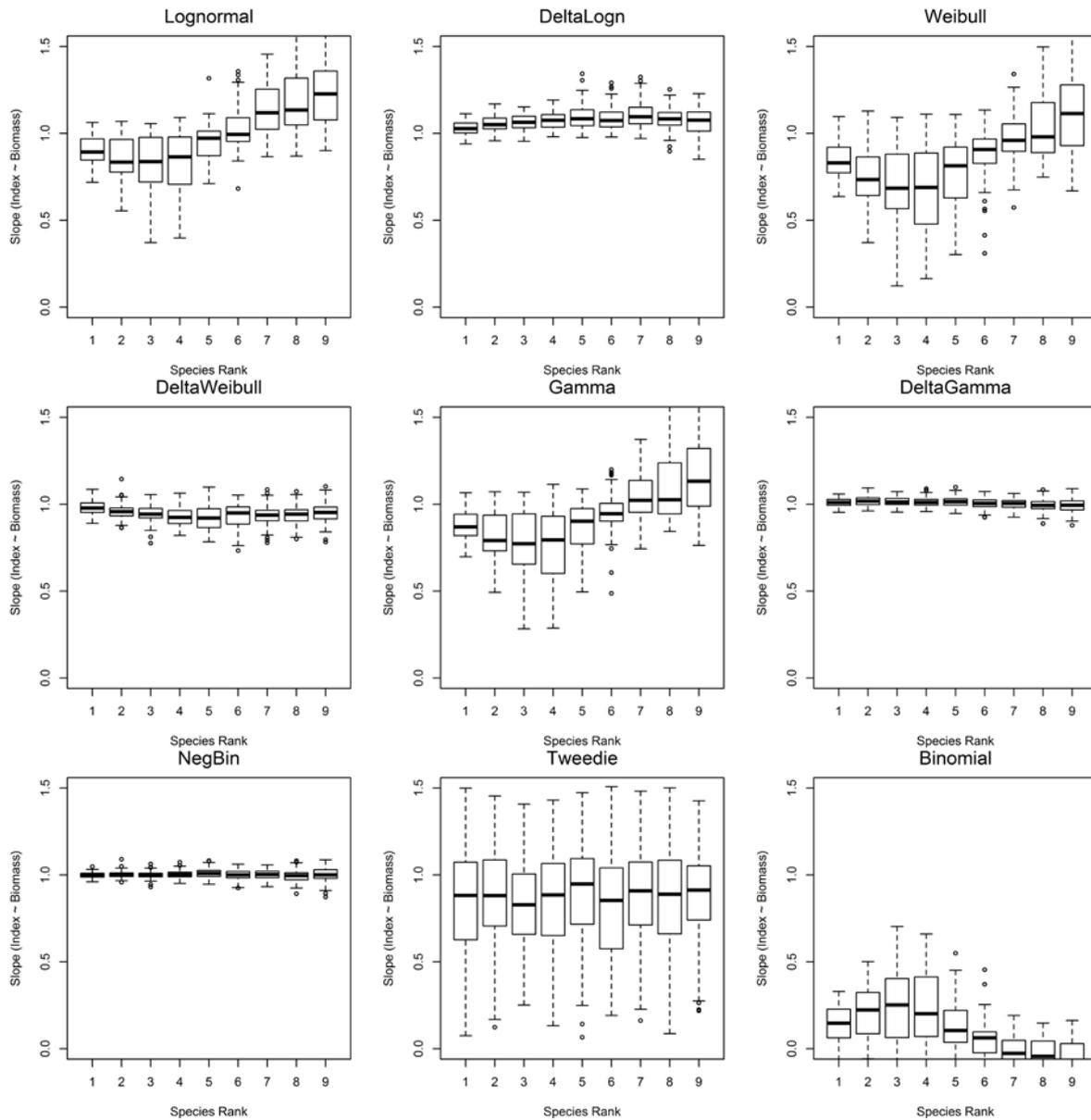
Incorporating the binomial CPUE indices in the two-stage models (delta-lognormal, delta-Weibull, delta-Gamma) CPUE indices moderated the trends in the positive catch indices and the three sets of combined indices yielded relatively unbiased CPUE indices for the full range of species (Species 1 to 9) (Figure 11). However, for most species, the performance of the delta-lognormal CPUE indices deteriorated when the variation in species distribution was increased (SD 1.0) and the proportion of zero catches was increased (under the Figure 1 Occurrence function B) (Figure 12). The delta-Weibull and delta-Gamma models were relatively robust to the changes in data formulation. Nonetheless, the Negative Binomial models appeared to perform marginally better than the two-stage CPUE models for all species, with correlation coefficients close to 1.0 and slope values approximating 1.0 (Figure 10 and Figure 11).



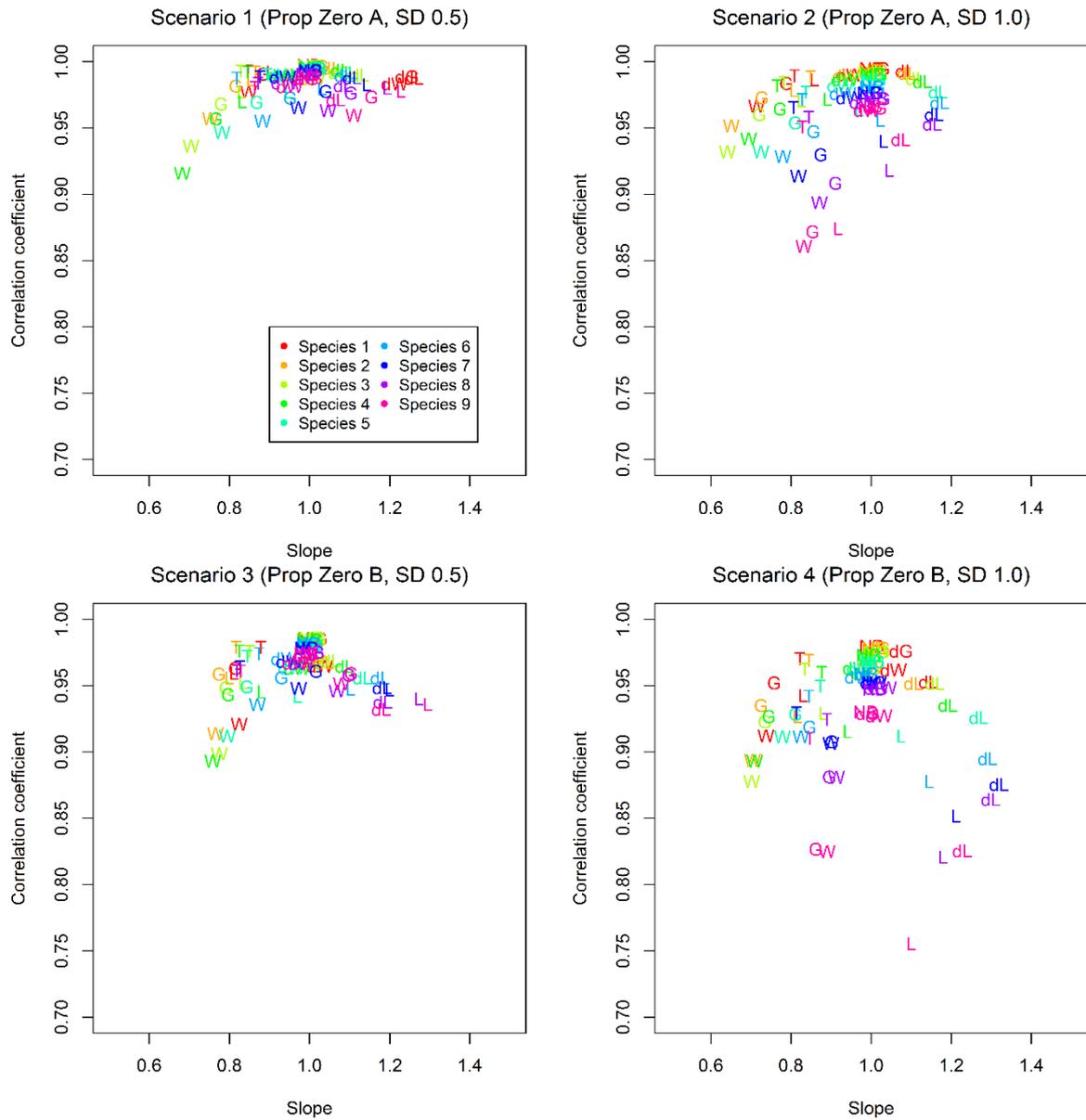
**Figure 9: Boxplots of the  $R^2$  statistics from each CPUE modelling approach for each of the 9 species (ranked 1 to 9, most abundant to least abundant) from the simulated *day* data sets ( $n = 100$ ) for Scenario 1; i.e., species occurrence function A (Figure 1), variation in species distribution SD 0.5, trawl distribution random, reporting module included.**



**Figure 10:** Boxplots of the correlation coefficients between the CPUE indices and simulated species abundance for each of the 9 species (ranked 1 to 9, most abundant to least abundant) for each CPUE modelling approach. The simulated *day* data sets ( $n = 100$ ) for Scenario 1; i.e., species occurrence function A (Figure 1), variation in species distribution SD 0.5, trawl distribution random, reporting module included.



**Figure 11: Boxplots of the slope ( $b$ ) of the linear relationship between the CPUE indices and simulated species abundance for each of the 9 species (ranked 1 to 9, most abundant to least abundant) for each CPUE modelling approach. The simulated *day* data sets ( $n = 100$ ) for Scenario 1; i.e., species occurrence function A (Figure 1), variation in species distribution SD 0.5, trawl distribution random, reporting module included.**



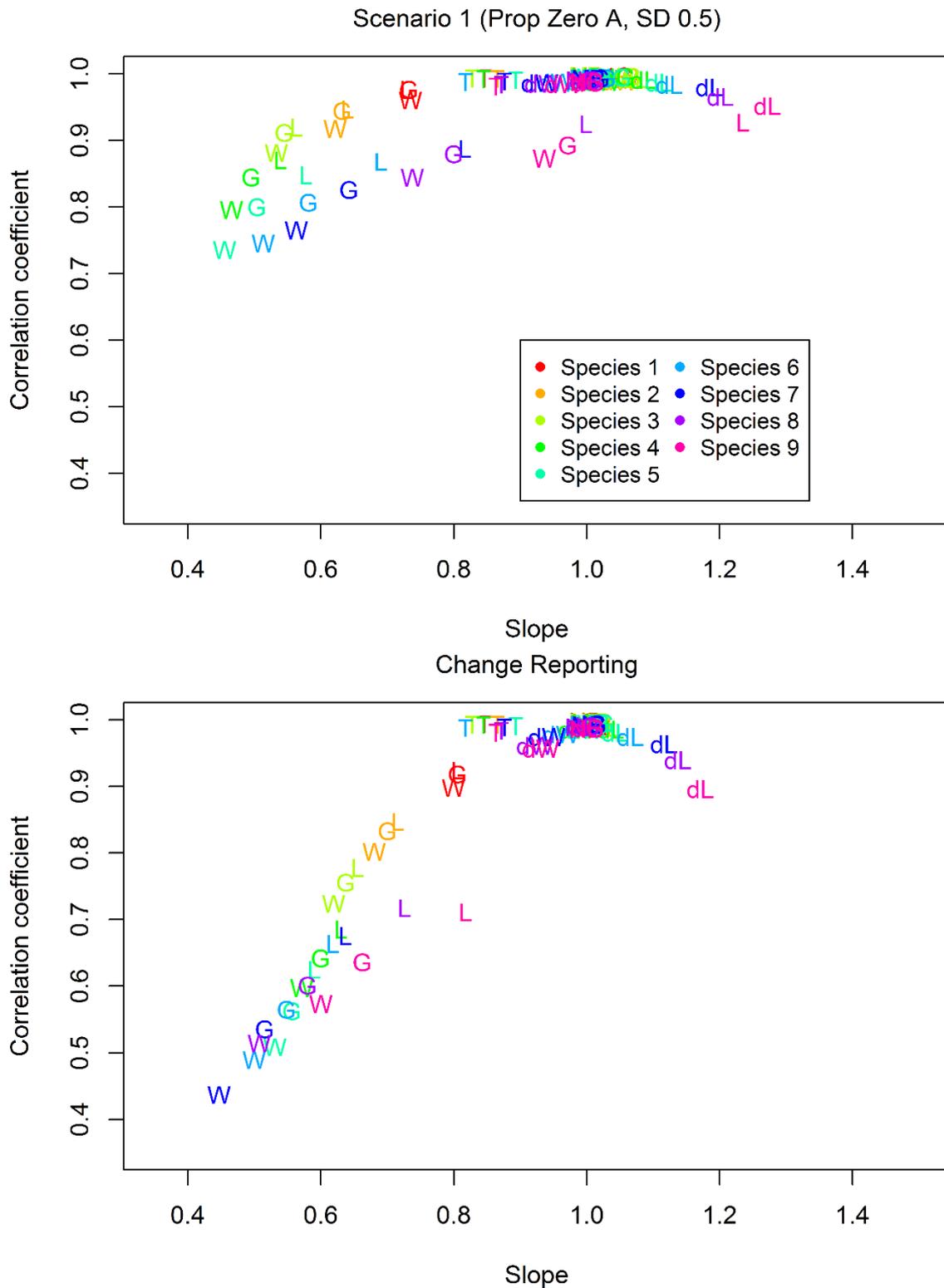
**Figure 12: Average slope and correlation coefficients (from 100 simulations) for each species (1 to 9, colour coded) from selected *day* based CPUE models (L, lognormal; dL, Delta-lognormal; W, Weibull; dW, delta-Weibull; G, Gamma; dG, Delta-Gamma; T, Tweedie; NB, negative binomial) and different data configurations (panels).**

### 3.4 Simulating changes in species catch reporting

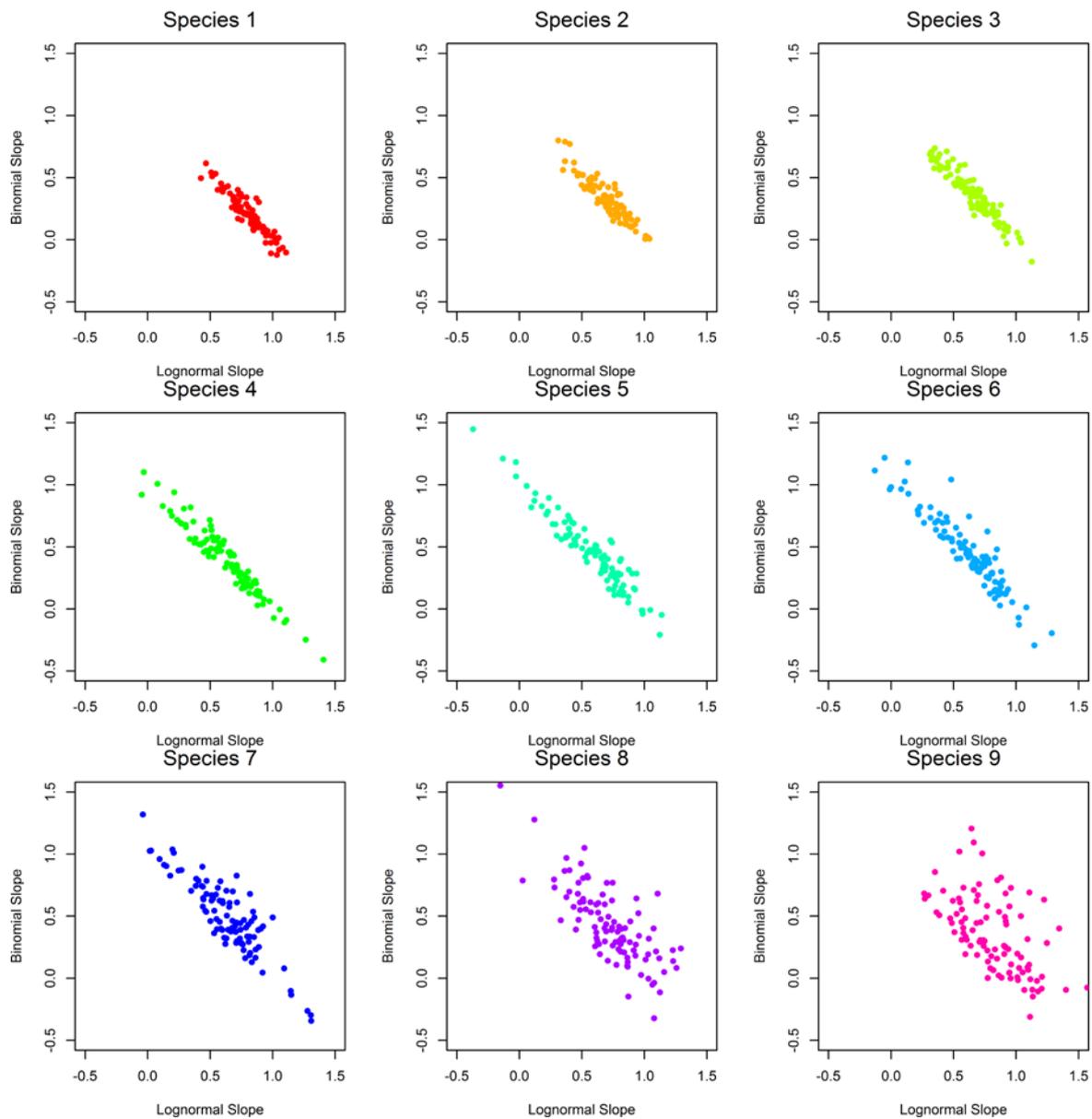
An additional set of simulations were conducted that changed the level of reporting of trawl catches during the time period; trawl based catches were reported for the five most abundant species for the first 10 years and then reporting was extended to encompass the eight most abundant species for the subsequent 10 years. The simulations were based on trawl data simulated using the equivalent configuration to Scenario 1 (SD 0.5, Occurrence function A; Figure 1).

These simulations revealed a marked deterioration in the accuracy of the positive catch models (lognormal, Weibull and Gamma) for almost all species, although the effect was most pronounced for the medium and lower tier species (Species 4 to 9) (Figure 13). There was also a considerable deterioration in the accuracy of the CPUE indices from the binomial model associated with the change in reporting. However, the corresponding trends in the binomial and lognormal CPUE indices are generally negatively correlated (relative to the trend in biomass) (Figure 14) and, hence, the combined model (delta-lognormal) has a tendency to moderate the influence of the change in reporting. For example, the increased level of reporting will result in a lower proportion of zero catches being reported during the latter period (positively biasing the binomial index) while the change in reporting will negatively bias the lognormal index as there are more, smaller catches during the latter period.

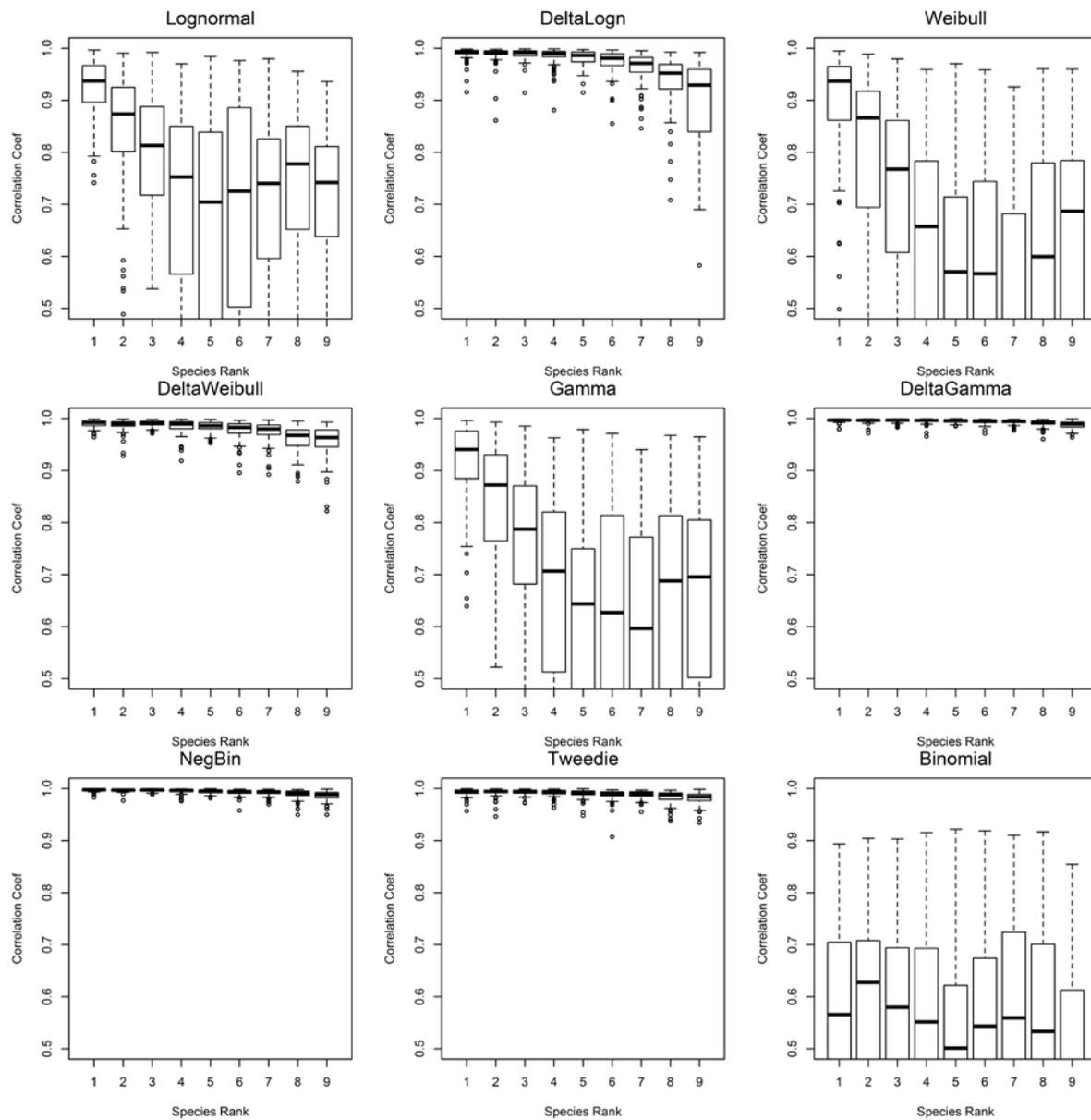
The simulated results revealed that the two-stage models (delta-lognormal, delta-Weibull, delta-Gamma) models yielded CPUE indices that compensated for the change in reporting for all but the lowest tier of the species (Species 8 and 9) while the negative binomial model yielded unbiased CPUE indices for the entire range of species (Figure 15 and Figure 16).



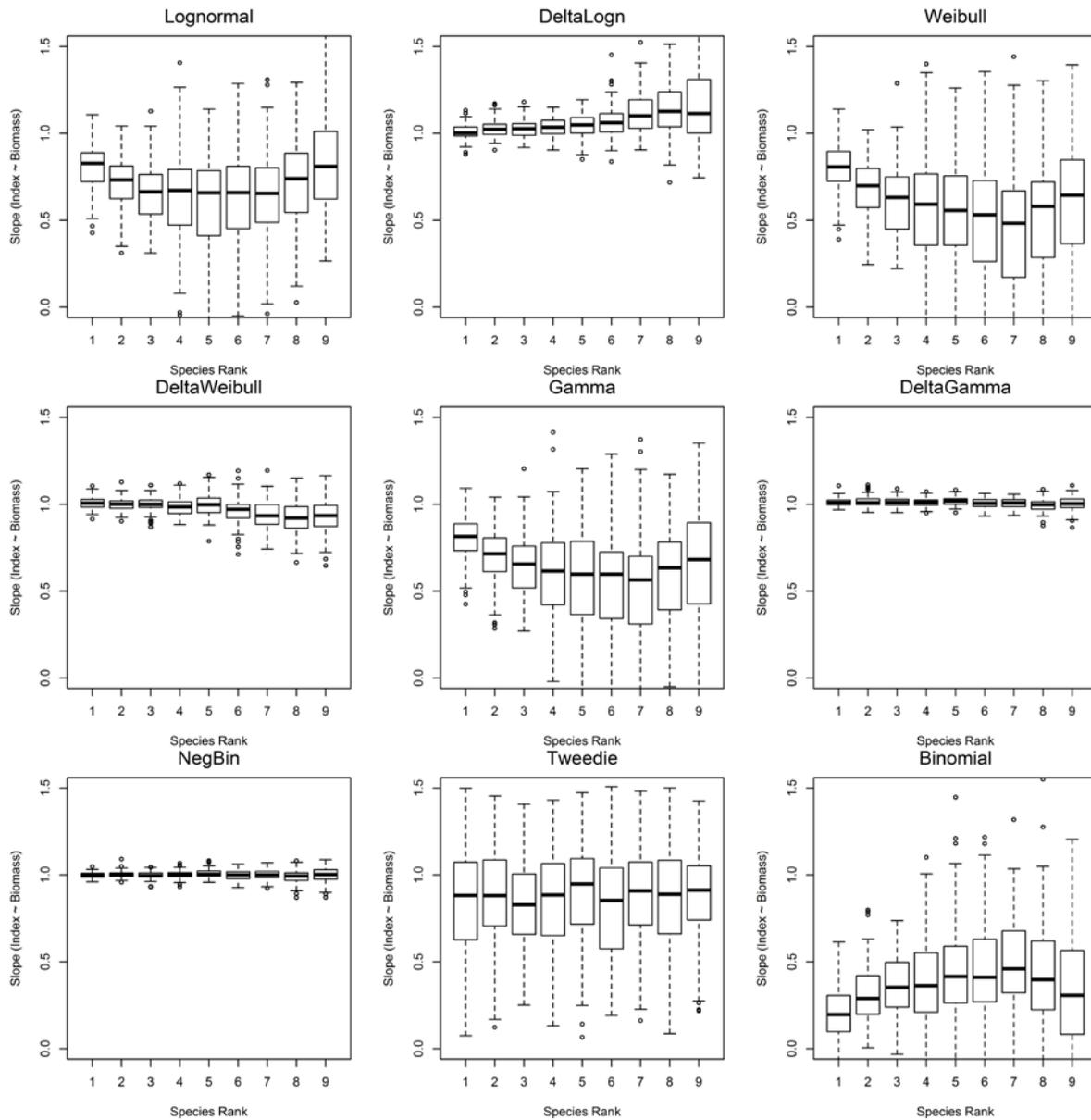
**Figure 13:** Average slope and correlation coefficients (from 100 simulations) for each species (1 to 9, colour coded) from selected *trawl* based CPUE models (L, lognormal; dL, Delta-lognormal; W, Weibull; dW, delta-Weibull; G, Gamma; dG, delta-Gamma; T, Tweedie; NB, negative binomial) for data configuration Scenario 1 (top panel) and a scenario with a change in the reporting of species catches (from five species to eight species).



**Figure 14: A comparison of the slope of the relationships between the binomial CPUE indices and simulated species abundance (y-axis) and the lognormal CPUE indices and simulated species abundance (x-axis) from 100 individual simulations for each of the 9 species (ranked 1 to 9, most abundant to least abundant). The simulated *trawl* data sets ( $n = 100$ ) is based on Scenario 1 and includes a change in species catch reporting.**



**Figure 15: Boxplots of the correlation coefficient between the CPUE indices and simulated species abundance for each of the 9 species (ranked 1 to 9, most abundant to least abundant) for each CPUE modelling approach. The simulated *trawl* data sets ( $n = 100$ ) is based on Scenario 1 and includes a change in species catch reporting.**



**Figure 16: Boxplots of the slope ( $b$ ) of the linear relationship between the CPUE indices and simulated species abundance for each of the 9 species (ranked 1 to 9, most abundant to least abundant) for each CPUE modelling approach. The simulated *trawl* data sets ( $n = 100$ ) is based on Scenario 1 and includes a change in species catch reporting.**

### 3.5 Model diagnostics

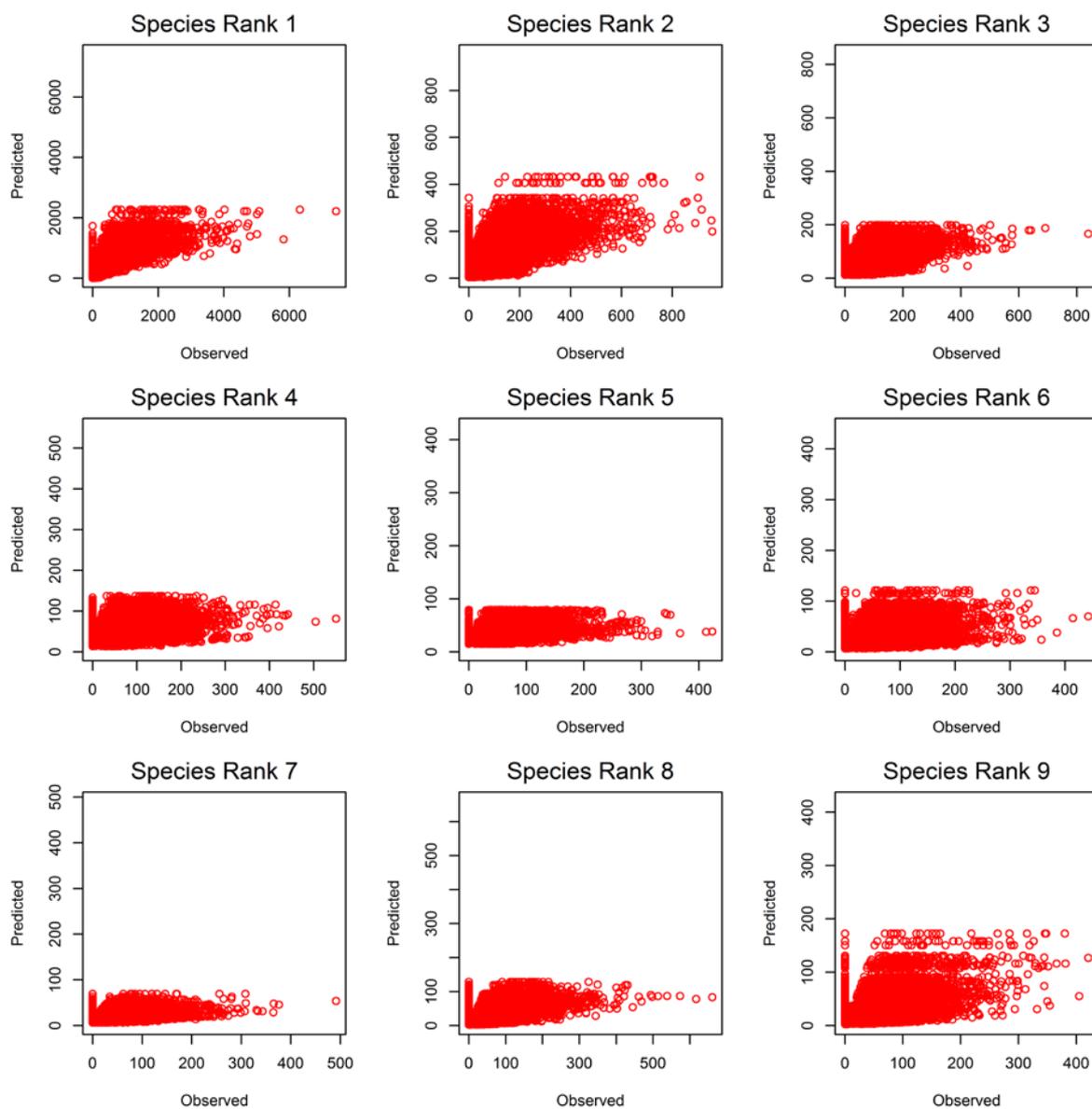
Standard model diagnostics are an important tool for the routine evaluation of positive catch CPUE models and, to a lesser extent, binomial models. The model diagnostics tend to be limited to distributional plots of model residuals, plots of observed vs expected values and quantile-quantile plots.

For a subset of the simulated results, model diagnostics were examined for the lognormal, Weibull and Gamma CPUE models. Similarly, comparative plots of the observed and predicted probability of catch were examined for binomial models. Overall, the binomial models accurately predicted the annual proportion of positive catch records in each species data set (Appendix 2, Figure A1). For the three positive catch models, the predicted distributions of catches were generally comparable to the observed catches for the most abundant species (Species 1–3) (Appendix 2, Figure A2–4). For the intermediate species (Species 4–7) the three models generally under-estimated the proportion of very small catches and also under-estimated the upper distribution of the observed catches (Appendix 2, Figure A2–4), although the average catch is well estimated. Qualitatively, these diagnostics might be interpreted as a poor performance of the positive catch CPUE models (from QQ plots, residuals, etc.).

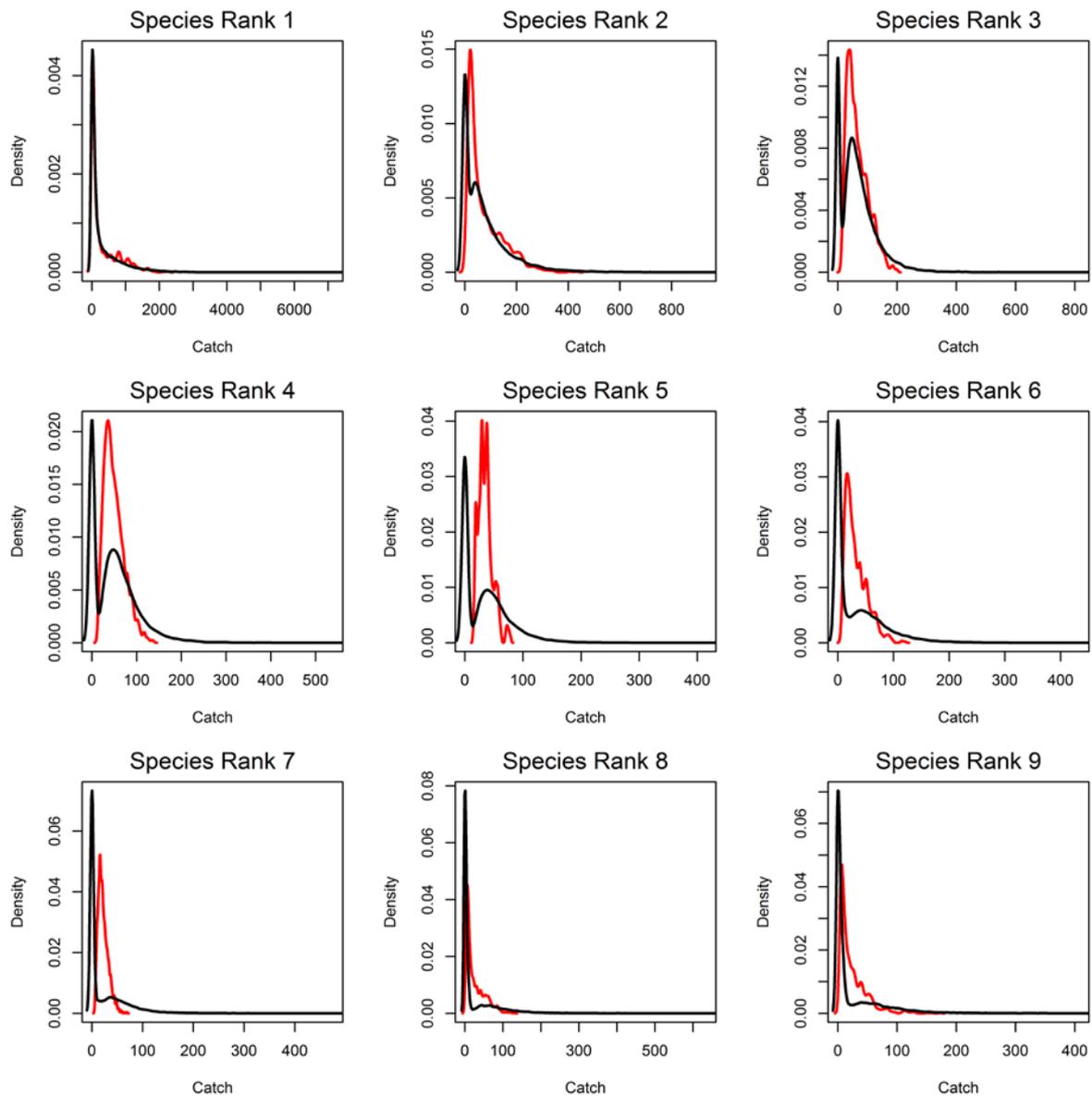
For the positive CPUE models for the two least abundant species (Species 8 and 9) the predicted distributions of catches were generally comparable to the observed catches (Appendix 2, Figure A2–4), although the Gamma model performed best for the smallest catches.

A more thorough evaluation of model performance was conducted for the negative binomial models. The CPUE indices derived from these models very closely approximated the trends in species abundance, seemingly exceeding the performance of the two-stage (delta-lognormal, delta-Gamma and delta-Weibull) CPUE models. However, an examination of the diagnostics reveals that the models represent a very poor fit to the individual observations (Figure 17 and Figure 18). Overall, the models estimate a very low proportion of zero catches, instead estimating a higher proportion of relatively small catches. The models also under estimate the extent of the upper range of observed catches, while, for each species, the negative binomial models almost exactly predicts the arithmetic mean of the annual catch rates (Figure 19) and, hence, the species relative abundance.

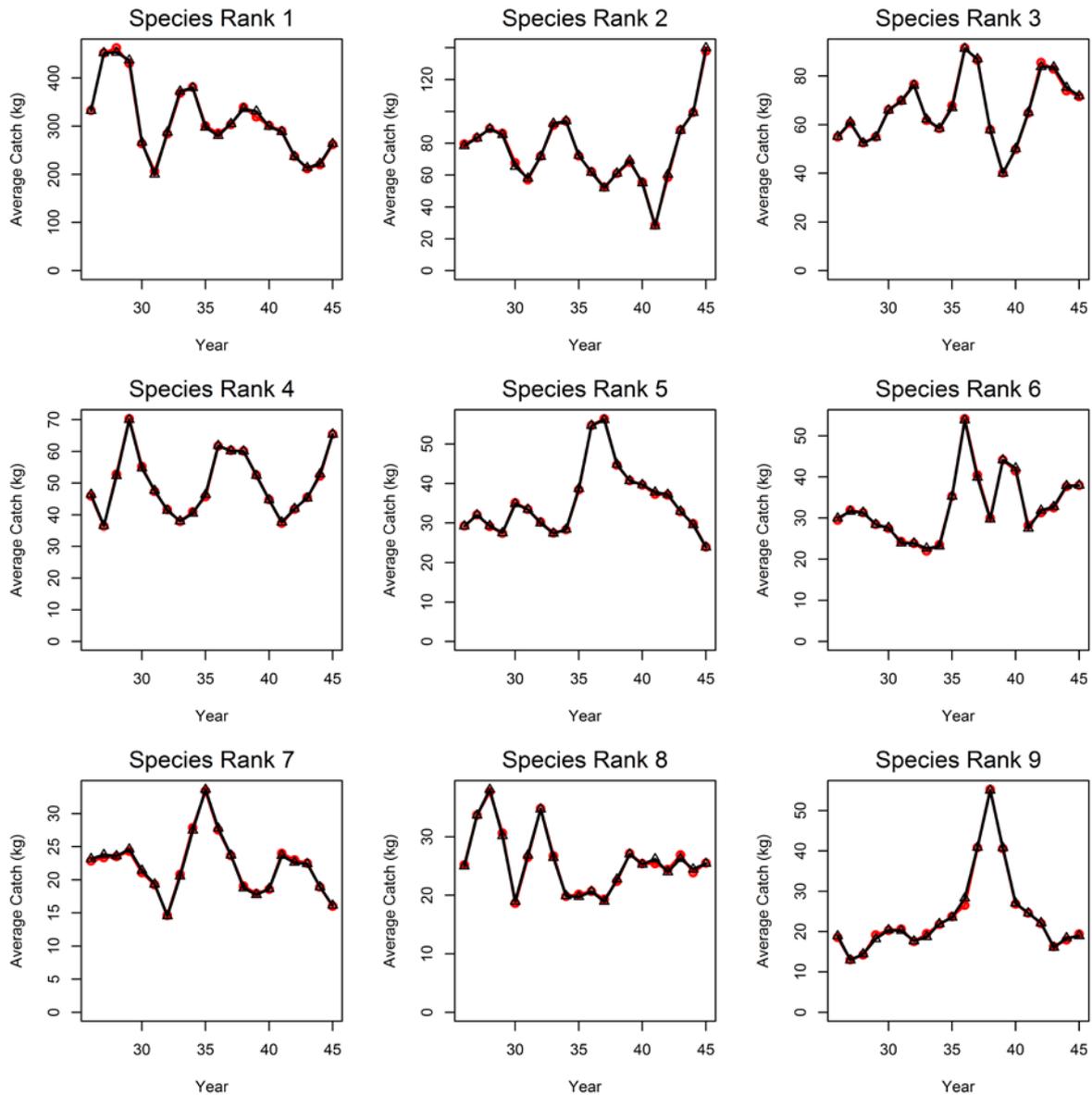
These diagnostics reveal that the distributional structure of the catch and effort data sets is not consistent with the negative binomial distribution. Consequently, the resultant models are not statistically robust for model inference, including the determination of confidence intervals for the key parameters of interest (especially those associated with the annual coefficients).



**Figure 17: An example of a comparison of individual observed catches and predicted catches from negative binomial CPUE models by species from one simulated data set (*trawl* data configured for Scenario 1; i.e., species occurrence function A (Figure 1), variation in species distribution SD 0.5, trawl distribution random, reporting module included).**



**Figure 18:** An example of a comparison of density plots of observed (black line) and predicted (red line) catches from negative binomial CPUE models by species from one simulated data set (*trawl* data configured for Scenario 1; i.e., species occurrence function A (Figure 1), variation in species distribution SD 0.5, trawl distribution random, reporting module included).



**Figure 19:** An example of a comparison of the observed arithmetic mean CPUE (catch per trawl) (black line and points) and the predicted (red line and points) annual CPUE from negative binomial CPUE models by species from one simulated data set (*trawl* data configured for Scenario 1; i.e., species occurrence function A (Figure 1), variation in species distribution SD 0.5, trawl distribution random, reporting module included).

#### 4. DISCUSSION

This study was conducted using a simple simulation framework to evaluate a range of standard CPUE modelling approaches. The primary focus of the study was to investigate the influence of the catch reporting regime and CPUE data processing in the evaluation of the CPUE modelling approaches routinely applied for the derivation of CPUE indices. The *trawl* and *day* based reporting regimes, approximated in the simulation process (reporting module), will tend to result in the under-reporting of the presence of an individual species in the total catches, especially for the less abundant species. Within the CPUE modelling framework, poor reporting of small catches has the potential to result in the over-estimation of the probability of a zero catch in a binomial CPUE model, thereby negatively biasing the binomial component of the CPUE index. The data processing applied to the CPUE data sets apportion all the catch of these species amongst the individual effort records from a fishing trip; the catch is allocated amongst the individual efforts records with estimated catches of the species or amongst all effort records, if there are no estimated catches reported. This may result in the magnitude of the observed catches being inflated through the catch-based allocation process or the generation of additional small catch records (creating false positive catch records). Thus, the data processing methodology may introduce additional biases in the catch observations included in the CPUE data sets. The simulation modelling approach provided the framework to evaluate the capability of various commonly used CPUE standardisation approaches to adjust for these potential sources of bias.

The simulated data sets do not fully accommodate the degree of complexity associated with the distributions of individual species and the overlaying complexity of the distribution of fishing effort and the associated variations in catchability. A limited range of assumptions were applied to the derivation of the individual species distributions to simplify interpretation and to ensure the conclusions of the study were robust to a reasonable range of assumptions.

The key conclusions of the simulation study are as follow.

- Positive catch only models (lognormal, Gamma and Weibull) yielded biased estimates of trends in relative abundance for all species, with the exception of those species that were ubiquitous and occurred in almost all (more than 95%) of the CPUE records. In all other cases, the positive catch CPUE indices under-estimated the scale of the variation in species abundance (hyper-stability) and this was most pronounced for species of intermediate abundance (second tier).
- The inclusion of the binomial component in the combined delta-lognormal, delta-Gamma and delta-Weibull CPUE models generally compensated for the bias in the positive catch component of the CPUE models, particularly for the main species in the total catches. The combined CPUE indices closely approximated the trend in the species abundance for the five most abundant species; i.e., the species that were most regularly reported (either by trawl or day).
- For the lower tier (less abundant) species, the delta-Gamma and delta-Weibull models yielded CPUE indices that were less biased than the delta-lognormal model.
- The performance of the delta-lognormal models was more sensitive to the underlying distributional assumptions of the species (species occurrence and variability). The performance of the delta-lognormal models deteriorated with the increasing contrast in the distribution of a species, corresponding to an increasing proportion of zero-catch records being included in the simulated data set. In general, the delta-lognormal CPUE indices tended to over-estimate the extent of the trends in relative abundance for those species with at least 50–60% of the data records comprising zero catches (i.e. the CPUE indices exaggerated the increasing or decreasing trends in species abundance).
- Overall, the delta-Weibull, delta-Gamma and negative binomial CPUE models were relatively insensitive to distributional assumptions of the simulated data and performed well even for data sets with a high proportion of zero-catch records. The improvement of the performance of these two models, relative to the delta-lognormal CPUE model, is presumably related to the greater flexibility of the Weibull and Gamma model components due to the estimation of the additional shape parameters for these distributions.
- While these “two-stage” models performed well for the range of species, in terms of accurately estimating the abundance trends, an examination of the model diagnostics (QQplots, residual plots, etc.) indicated that the positive catch components of these models may under-estimate the extent

of variation in catches. Therefore, while the model diagnostics may indicate a deficiency in the overall performance of the CPUE model, the diagnostics do not provide a strong basis for deciding whether or not to accept or reject the resulting CPUE indices.

- The performance of the negative binomial CPUE model was similar to the performance to the delta-Weibull and delta-Gamma CPUE models, in terms determining the point estimate of the annual indices. However, the negative binomial models did not reliably represent the distribution of the simulated data, particularly the proportion of zero catches in the data sets. Thus, the deficiencies in the diagnostics from the negative binomial CPUE models were considerably greater than from the components of the ‘two-stage’ models. The modelling results indicate that the structure of CPUE data do not conform to a negative binomial distribution and, consequentially, the negative binomial CPUE models were not considered appropriate for statistical inference and are unlikely to provide reliable estimates of uncertainty associated with the annual indices. Initial investigations of the application of Zero-Inflated negative binomial models did not reveal an appreciable improvement in model performance relative to the negative binomial models. Changes in species catch reporting have the potential to introduce a significant bias in the positive catch component of the CPUE indices for the lower tier (less abundant) species. The influence of changes in reporting was considerably lower for the ‘two-stage’ combined models (delta-lognormal, delta-Gamma and delta-Weibull), although the delta-Gamma and delta-Weibull CPUE models were considerably less sensitive to changes in reporting behaviour than the delta-lognormal models.
- The concentration of fishing effort in areas of highest species abundance is likely to result in a degree of hyper-stability of the CPUE indices derived for the key species of interest. The extent of the hyper-stability of the CPUE indices will be related to the underlying species distribution and the movement rates of the species. There is potential to reduce these effects by incorporating spatially-structured approaches in the derivation of the CPUE indices (Campbell 2004, 2015, Thorson 2015).
- The conclusions of this study are relevant to both individual trawl observations and aggregated daily catch and effort data (*trawl* or *day* format), although the variation in the results was considerably more pronounced for the *trawl* based data set. The aggregation of data by fishing day reduced the variability in the data set and, hence, reduced the magnitude of the bias introduced in the simulated CPUE indices.

There is considerable scope to extend the current simulation approach to more realistically represent the spatial distributions of species assemblages, relative to habitat preferences, within specific fishery areas as informed by spatially resolved data from trawl surveys and/or fishery catch and effort data. Similarly, fisheries data could be applied to formulate alternative assumptions for the distribution of fishing effort relative to the abundance of individual species.

Conceptually, the negative binomial and ZipNB modelling approaches are appealing as they have the potential to more directly model the proportion of zero catches. However, this study suggests that these methods may not be as capable of reflecting the distributional structure of the underlying data as well as the delta methods. This may relate, in part, to the structure of the data, specifically the generation of false zero catches through the species reporting process and/or the generation of false positive catches via the allocation of small landed catches amongst multiple effort records (catch allocation process).

Extension of the simulation studies should include further refinement of the consideration of alternative CPUE modelling approaches and identify the key criteria for determining the suitability of catch and effort data for the derivation of reliable CPUE indices. Such an analysis may also enable the quantification of the likely extent of the potential bias in CPUE indices, which may be informative in the treatment of these indices; for example, the inclusion of CPUE indices within an assessment model with associated priors related to the link between CPUE indices and abundance.

## 5. RECOMMENDATIONS

A key purpose of this study was to provide some guidance for the analysis of catch and effort data for the derivation of CPUE analyses. The study is most relevant to data sets collected from inshore trawl fisheries under the reporting framework in existence over the last few decades. However, the results of the study are also likely to be applicable in a wider context.

For trawl event-based catch and effort data, positive catch models (e.g. lognormal, Weibull, Gamma) yield biased (hyper-stable) CPUE indices and should be disregarded, except for species with catches recorded in almost all (more than 95%) relevant fishing effort records. In other cases, it is recommended that combined CPUE models are used. Two-stage delta-lognormal, delta-Gamma and delta-Weibull CPUE models are appropriate for those species that are caught with moderate frequency (i.e. 50–95% of records). However, for less frequently reported species (catches recorded for less than about 40–50% of data records) delta-Gamma and delta-Weibull models are more appropriate than delta-lognormal models.

For the most ubiquitous species, the performance of the combined CPUE models may be compromised by the poor performance of the binomial model component. It is therefore recommended that combined models are not applied to derive CPUE indices for species that are recorded in more than 95% of the records from each year of the time series. Instead, positive catch models (Gamma, Weibull or lognormal) should be applied.

For the less abundant species, the relatively poor performance of the delta-lognormal models was related to the proportion of positive catch records that were generated from the effort based allocation of catches. The effort based allocation approach generates multiple small, notional catches (including false positive catches). The performance of the delta-lognormal models deteriorated when these notional catches represented at least 20% of the positive catches in the CPUE data set. While the delta-Gamma and delta-Weibull models performed well for these lower tier species, it is recommended that CPUE indices should not be derived for data sets that incorporate a considerable proportion (more than 20–30%) of positive catch records generated from the effort based allocation.

For daily aggregated catch and effort data, combined delta-lognormal, delta-Weibull and delta-Gamma models are recommended with the latter model achieving a higher level of performance over the range of species (regardless of prevalence).

Further simulation-based analyses are recommended to investigate the influence of other potential sources of bias in the derivation of CPUE indices, specifically the spatial structure of catch and effort data, changes in fleet configuration (including criteria for defining a core fleet) and changes in fishing efficiency. The current study could also be extended to further investigate the implications of changes in the catch and effort regime, particularly the impending changes associated with the introduction of electronic reporting (ERS), including changes to the reporting of individual species catches.

Further consideration is required to formulate a set of model diagnostics and/or performance criteria that are likely to be indicative of relative performance of the CPUE indices. These performance criteria could be informed by simulation testing across a wide range of scenarios encompassing assumptions related to the distribution of the species of interest and the associated fishing activity.

## 6. ACKNOWLEDGMENTS

Dr Pamela Mace (Fisheries New Zealand), in her role as Chair of the Statistics, Assessments and Methods (SAM) Working Group, supported the development of the research proposal. Members of the SAM Working Group provided considerable guidance in the development of the simulations and interpretation of results. The study was funded by Fisheries New Zealand under Project Code SEA2018-18.

## 7. REFERENCES

- Brodziack, J.; Walsh, W.A. (2013). Model selection and multimodel inference for standardizing catch rates of bycatch species: a case study of oceanic whitetip shark in the Hawaii-based longline fishery. *Canadian Journal of Fisheries and Aquatic Science* 70: 1723–1740.
- Campbell, R., (2004). CPUE standardisation and the construction of indices of stock abundance in a spatially varying fishery using general linear models. *Fisheries Research* 70, 209–227.
- Campbell, R., (2015). Constructing stock abundance indices from catch and effort data: some nuts and bolts. *Fisheries Research* 161, 109–130.

- Campbell, R.; Zhou, S.; Hoyle, S.; Hillary, R.; Haddon, H.; Auld, S. (2017). Developing innovative approaches to improve CPUE standardisation for Australia's multispecies pelagic longline fisheries. Final report for project 2014-021 to the Fisheries Research Development Corporation, Canberra, Australia.
- Fisheries New Zealand (2018). Fisheries Assessment Plenary, May 2018: stock assessments and stock status. Compiled by the Fisheries Science and Information Group, Fisheries New Zealand, Wellington, New Zealand. 1674 p.
- Forrestal, F.C.; Goodyear, C.P.; Schirripa, M.; Babcock, E.; Laretta, M.; Sharma, R. (2017). Testing robustness of CPUE standardization using simulated data: findings of initial blind trials. *Collective Volume Scientific Papers, ICCAT, 74(2)*: 391–403.
- Goodyear, C.P. (2006). Performance diagnostics for the longline CPUE simulator. *Collective Volume Scientific Papers, ICCAT, 59(2)*: 615–626.
- Hilbe, J.M. (2007). Negative binomial regression. University Press, Cambridge. 553 p.
- Langley A.D. (2015). Fishery characterisation and Catch-Per-Unit-Effort indices for John dory in JDO 1. *New Zealand Fisheries Assessment Report 2015/47*. 76 p.
- Lo, N.; Jacobson, L.; Squires, D. (1992). Indices of Relative Abundance from Fish Spotter Data based on Delta-Lognormal Models. *Canadian Journal of Fisheries and Aquatic Science* 49: 2515–2526.
- Maunder, M.N.; Punt, A.E. (2004). Standardizing catch and effort data: a review of recent approaches. *Fisheries Research* 70 (2004): 141–159.
- Neubauer, P. (2017). Spatial bias in pāua *Haliotis iris* catch-per-unit-effort. *New Zealand Fisheries Assessment Report 2017/57*. 33 p.
- Okamura, H.; Morita, S.H.; Funamoto, T.; Ichinokawa, M.; Eguchi, S. (2017). Target-based catch-per-unit-effort standardization in multispecies fisheries. *Canadian Journal of Fisheries and Aquatic Science* 75: 452–463.
- Petrere, Jr.M.; Giacomini, H.C.; De Marco Jr.P. (2010). Catch-per-unit-effort: which estimator is best? *Brazilian Journal of Biology, 70 (3)*, 483–491.
- Starr, P.J. (2007). Procedure for merging Ministry of Fisheries landing and effort data, version 2.0. Report to the Adaptive Management Programme Fishery Assessment Working Group: Document 2007/04, 17 p. Unpublished document held by Fisheries New Zealand, Wellington.
- Stefansson, G. (1996). Analysis of groundfish survey abundance data: combining the GLM and delta approaches. *ICES Journal of Marine Science, 53*: 577–588.
- Thorson, J.T.; Shelton, A.O.; Ward, E.J.; Skaug, H.J. (2015). Geostatistical delta-generalized linear mixed models improve precision for estimated abundance indices for West Coast groundfishes. *ICES Journal of Marine Science* 72,1297–1310.
- Venables, W.N.; Ripley, B.D. (1999). Modern applied statistics with S-PLUS (third edition). Springer-Verlag, New York. 501 p.

## APPENDIX 1

**Table A1: Average correlation coefficients for the CPUE indices derived for each species and model option for the *trawl* based data sets comparing different species distributional assumptions (Scenarios).**

Species	Lognorm	Binom	DeltaLog n	Gamma	DeltaGa mma	Weibull	DeltaWei bull	Tweedie	NegBin
Scenario 1									
1	0.976	0.835	0.993	0.976	0.996	0.960	0.991	0.993	0.997
2	0.947	0.832	0.991	0.944	0.995	0.917	0.989	0.993	0.996
3	0.920	0.867	0.992	0.911	0.996	0.881	0.990	0.993	0.996
4	0.870	0.842	0.991	0.845	0.995	0.796	0.988	0.992	0.995
5	0.848	0.796	0.987	0.800	0.994	0.735	0.989	0.991	0.994
6	0.868	0.709	0.984	0.807	0.993	0.745	0.987	0.988	0.993
7	0.887	0.600	0.980	0.826	0.993	0.765	0.986	0.989	0.993
8	0.924	0.371	0.965	0.880	0.991	0.844	0.985	0.984	0.990
9	0.927	0.017	0.951	0.893	0.987	0.873	0.985	0.981	0.988
Scenario 2									
1	0.886	0.874	0.991	0.893	0.995	0.871	0.986	0.990	0.993
2	0.831	0.884	0.989	0.836	0.994	0.808	0.985	0.989	0.992
3	0.742	0.862	0.987	0.726	0.991	0.692	0.981	0.984	0.987
4	0.740	0.861	0.983	0.713	0.991	0.669	0.980	0.982	0.986
5	0.693	0.843	0.974	0.617	0.987	0.553	0.975	0.977	0.980
6	0.649	0.809	0.963	0.545	0.981	0.483	0.968	0.972	0.975
7	0.648	0.765	0.943	0.559	0.977	0.491	0.963	0.966	0.968
8	0.639	0.699	0.929	0.567	0.973	0.513	0.962	0.958	0.961
9	0.568	0.615	0.903	0.535	0.966	0.507	0.957	0.951	0.953
Scenario 3									
1	0.893	0.779	0.941	0.937	0.985	0.928	0.982	0.979	0.986
2	0.855	0.806	0.942	0.905	0.985	0.893	0.982	0.979	0.985
3	0.826	0.836	0.944	0.870	0.985	0.854	0.981	0.976	0.986
4	0.798	0.832	0.941	0.820	0.984	0.795	0.979	0.976	0.984
5	0.764	0.820	0.931	0.774	0.980	0.750	0.976	0.972	0.981
6	0.755	0.822	0.933	0.745	0.982	0.719	0.977	0.974	0.982
7	0.810	0.726	0.923	0.795	0.978	0.775	0.974	0.964	0.978
8	0.853	0.571	0.912	0.858	0.975	0.851	0.974	0.961	0.975
9	0.855	0.401	0.904	0.880	0.974	0.876	0.973	0.961	0.974
Scenario 4									
1	0.724	0.844	0.912	0.852	0.980	0.850	0.979	0.970	0.980
2	0.663	0.863	0.905	0.798	0.978	0.795	0.977	0.969	0.978
3	0.613	0.847	0.908	0.720	0.976	0.716	0.974	0.963	0.976
4	0.541	0.838	0.892	0.640	0.973	0.636	0.972	0.960	0.973
5	0.532	0.829	0.886	0.557	0.968	0.552	0.967	0.950	0.968
6	0.536	0.804	0.853	0.531	0.959	0.525	0.958	0.942	0.959
7	0.545	0.757	0.839	0.524	0.953	0.523	0.952	0.930	0.952
8	0.520	0.736	0.824	0.509	0.948	0.513	0.948	0.925	0.947
9	0.483	0.636	0.774	0.487	0.930	0.492	0.927	0.910	0.930

**Table A2: Average slope parameters for the CPUE indices derived for each species and model option for the *trawl* based data sets comparing different species distributional assumptions (Scenarios).**

Species	Lognorm	Binom	DeltaLog n	Gamma	DeltaGa mma	Weibull	DeltaWei bull	Tweedie	NegBin
Scenario 1									
1	0.729	0.323	1.040	0.732	1.046	0.735	1.050	0.846	0.999
2	0.641	0.435	1.062	0.632	1.056	0.622	1.048	0.866	1.002
3	0.564	0.521	1.072	0.545	1.057	0.533	1.049	0.828	0.999
4	0.539	0.550	1.086	0.495	1.048	0.466	1.023	0.845	1.002
5	0.578	0.538	1.109	0.504	1.042	0.455	0.997	0.894	1.009
6	0.690	0.430	1.124	0.581	1.019	0.513	0.953	0.818	0.999
7	0.817	0.352	1.184	0.642	1.010	0.563	0.929	0.876	1.003
8	0.999	0.178	1.201	0.799	0.998	0.738	0.934	0.872	0.994
9	1.235	-0.011	1.272	0.971	1.002	0.936	0.965	0.863	1.000
Scenario 2									
1	0.527	0.484	1.007	0.525	1.006	0.526	1.008	0.809	1.001
2	0.451	0.549	1.008	0.444	1.003	0.441	1.000	0.849	0.998
3	0.402	0.617	1.017	0.383	1.003	0.382	1.004	0.783	0.998
4	0.408	0.631	1.041	0.370	1.008	0.359	0.999	0.765	1.000
5	0.422	0.672	1.088	0.332	1.008	0.306	0.983	0.837	1.004
6	0.462	0.673	1.134	0.322	1.003	0.282	0.964	0.828	1.005
7	0.561	0.622	1.184	0.367	0.994	0.312	0.939	0.806	0.995
8	0.672	0.567	1.269	0.420	1.010	0.354	0.942	0.845	1.007
9	0.675	0.518	1.241	0.443	0.993	0.392	0.938	0.828	0.988
Scenario 3									
1	0.654	0.328	0.969	0.715	1.028	0.723	1.036	0.879	1.001
2	0.577	0.400	0.978	0.625	1.029	0.634	1.037	0.817	0.998
3	0.548	0.475	1.015	0.574	1.043	0.582	1.051	0.850	1.003
4	0.515	0.517	1.026	0.514	1.031	0.517	1.035	0.824	0.999
5	0.531	0.533	1.050	0.493	1.023	0.495	1.027	0.844	1.002
6	0.558	0.546	1.092	0.470	1.012	0.466	1.009	0.873	0.998
7	0.752	0.415	1.166	0.577	0.995	0.569	0.987	0.827	0.993
8	0.981	0.276	1.269	0.721	1.007	0.717	1.003	0.819	1.007
9	1.120	0.166	1.331	0.784	0.981	0.796	0.995	0.828	0.991
Scenario 4									
1	0.521	0.458	0.988	0.537	1.004	0.538	1.005	0.822	1.001
2	0.441	0.541	0.984	0.466	1.009	0.468	1.011	0.845	1.007
3	0.426	0.586	1.014	0.422	1.013	0.423	1.014	0.834	1.011
4	0.381	0.628	1.010	0.360	0.994	0.361	0.995	0.877	0.993
5	0.398	0.673	1.067	0.322	1.003	0.321	1.002	0.873	1.001
6	0.450	0.657	1.103	0.323	0.985	0.322	0.984	0.844	0.988
7	0.576	0.636	1.204	0.367	1.004	0.368	1.004	0.814	1.005
8	0.617	0.614	1.265	0.371	1.005	0.379	1.014	0.890	1.009
9	0.693	0.537	1.288	0.414	0.984	0.430	1.003	0.848	0.987

**Table A3: Average correlation coefficients for the CPUE indices derived for each species for comparative model options for the *trawl* based data sets comparing different distributions of fishing effort.**

Species	Lognorm	Binom	DeltaLog n	Gamma	DeltaGa mma	Weibull	DeltaWei bull	Tweedie	NegBin
	Scenario 1								
1	0.976	0.835	0.993	0.976	0.996	0.960	0.991	0.993	0.997
2	0.947	0.832	0.991	0.944	0.995	0.917	0.989	0.993	0.996
3	0.920	0.867	0.992	0.911	0.996	0.881	0.990	0.993	0.996
4	0.870	0.842	0.991	0.845	0.995	0.796	0.988	0.992	0.995
5	0.848	0.796	0.987	0.800	0.994	0.735	0.989	0.991	0.994
6	0.868	0.709	0.984	0.807	0.993	0.745	0.987	0.988	0.993
7	0.887	0.600	0.980	0.826	0.993	0.765	0.986	0.989	0.993
8	0.924	0.371	0.965	0.880	0.991	0.844	0.985	0.984	0.990
9	0.927	0.017	0.951	0.893	0.987	0.873	0.985	0.981	0.988
	Trawl Biomass Species1								
1	0.976	0.787	0.986	0.942	0.964	0.862	0.905	0.960	0.963
2	0.934	0.809	0.979	0.928	0.982	0.899	0.974	0.979	0.983
3	0.904	0.817	0.979	0.893	0.983	0.861	0.975	0.979	0.983
4	0.852	0.811	0.976	0.819	0.981	0.760	0.973	0.976	0.981
5	0.817	0.803	0.972	0.761	0.975	0.695	0.966	0.971	0.975
6	0.829	0.711	0.954	0.765	0.966	0.689	0.959	0.962	0.966
7	0.872	0.511	0.942	0.814	0.962	0.756	0.953	0.955	0.962
8	0.896	0.284	0.935	0.846	0.954	0.815	0.948	0.949	0.954
9	0.882	0.006	0.903	0.846	0.939	0.828	0.938	0.932	0.940
	Trawl Biomass All Species								
1	0.980	0.837	0.995	0.971	0.994	0.935	0.978	0.991	0.994
2	0.954	0.859	0.993	0.945	0.993	0.904	0.978	0.990	0.993
3	0.909	0.834	0.990	0.893	0.992	0.849	0.980	0.988	0.992
4	0.889	0.817	0.990	0.866	0.992	0.825	0.983	0.988	0.992
5	0.837	0.776	0.984	0.777	0.990	0.699	0.978	0.984	0.990
6	0.894	0.682	0.981	0.843	0.990	0.782	0.981	0.984	0.990
7	0.907	0.583	0.974	0.846	0.987	0.781	0.979	0.982	0.987
8	0.919	0.289	0.958	0.882	0.983	0.854	0.976	0.977	0.983
9	0.932	0.017	0.950	0.893	0.977	0.871	0.974	0.971	0.978

**Table A4: Average slope for the CPUE indices derived for each species and comparable model options for the *Trawl* based data sets comparing different distributions of fishing effort.**

Species	Lognorm	Binom	DeltaLog n	Gamma	DeltaGa mma	Weibull	DeltaWei bull	Tweedie	NegBin
	Scenario 1								
1	0.729	0.323	1.040	0.732	1.046	0.735	1.050	0.846	0.999
2	0.641	0.435	1.062	0.632	1.056	0.622	1.048	0.866	1.002
3	0.564	0.521	1.072	0.545	1.057	0.533	1.049	0.828	0.999
4	0.539	0.550	1.086	0.495	1.048	0.466	1.023	0.845	1.002
5	0.578	0.538	1.109	0.504	1.042	0.455	0.997	0.894	1.009
6	0.690	0.430	1.124	0.581	1.019	0.513	0.953	0.818	0.999
7	0.817	0.352	1.184	0.642	1.010	0.563	0.929	0.876	1.003
8	0.999	0.178	1.201	0.799	0.998	0.738	0.934	0.872	0.994
9	1.235	-0.011	1.272	0.971	1.002	0.936	0.965	0.863	1.000
	Trawl Biomass Species1								
1	0.767	0.169	0.931	0.767	0.931	0.763	0.927	0.781	0.901
2	0.653	0.421	1.072	0.643	1.064	0.634	1.058	0.836	1.005
3	0.595	0.490	1.074	0.576	1.060	0.565	1.051	0.825	1.005
4	0.539	0.532	1.072	0.496	1.034	0.467	1.009	0.874	0.998
5	0.562	0.557	1.113	0.484	1.041	0.435	0.996	0.851	1.006
6	0.654	0.479	1.131	0.541	1.022	0.474	0.957	0.821	1.004
7	0.849	0.293	1.158	0.699	1.007	0.634	0.940	0.817	1.000
8	1.012	0.145	1.193	0.827	1.002	0.775	0.948	0.834	0.997
9	1.221	-0.026	1.237	0.991	1.002	0.964	0.974	0.855	1.001
	Trawl Biomass All Species								
1	0.747	0.339	1.071	0.752	1.078	0.756	1.083	0.923	1.028
2	0.643	0.436	1.066	0.647	1.074	0.651	1.080	0.908	1.020
3	0.588	0.487	1.072	0.575	1.061	0.566	1.054	0.804	1.008
4	0.600	0.483	1.083	0.564	1.052	0.540	1.032	0.817	1.000
5	0.571	0.512	1.073	0.510	1.020	0.470	0.984	0.773	0.989
6	0.713	0.382	1.099	0.622	1.013	0.564	0.957	0.810	0.993
7	0.820	0.323	1.153	0.670	1.004	0.597	0.930	0.836	0.995
8	1.033	0.130	1.197	0.837	0.996	0.784	0.940	0.864	0.992
9	1.188	-0.011	1.216	0.972	0.995	0.944	0.966	0.837	0.992

**Table A5: Average correlation coefficients for the CPUE indices derived for each species and model option for the *Day* based data sets comparing different species distributional assumptions (Scenarios).**

Species	Lognorm	Binom	DeltaLog n	Gamma	DeltaGa mma	Weibull	DeltaWei bull	Tweedie	NegBin
Scenario 1									
1	0.992	0.584	0.988	0.991	0.989	0.977	0.983	0.993	0.997
2	0.986	0.628	0.995	0.982	0.997	0.957	0.990	0.993	0.997
3	0.977	0.658	0.990	0.968	0.992	0.937	0.987	0.993	0.997
4	0.970	0.655	0.993	0.957	0.996	0.916	0.989	0.992	0.996
5	0.977	0.478	0.992	0.970	0.995	0.947	0.990	0.991	0.995
6	0.980	0.252	0.990	0.973	0.994	0.955	0.989	0.988	0.994
7	0.983	-0.119	0.988	0.978	0.993	0.965	0.988	0.989	0.993
8	0.980	-0.190	0.982	0.976	0.991	0.963	0.984	0.984	0.991
9	0.978	-0.342	0.971	0.974	0.987	0.960	0.981	0.981	0.988
Scenario 2									
1	0.986	0.821	0.993	0.984	0.995	0.967	0.990	0.990	0.995
2	0.978	0.843	0.991	0.973	0.993	0.952	0.989	0.989	0.994
3	0.971	0.819	0.987	0.960	0.991	0.932	0.986	0.984	0.991
4	0.972	0.748	0.985	0.964	0.991	0.942	0.987	0.982	0.991
5	0.964	0.647	0.977	0.954	0.986	0.932	0.982	0.977	0.987
6	0.956	0.504	0.969	0.947	0.981	0.929	0.976	0.972	0.981
7	0.940	0.359	0.960	0.930	0.977	0.914	0.973	0.966	0.977
8	0.918	0.258	0.953	0.909	0.972	0.894	0.970	0.958	0.973
9	0.874	0.305	0.941	0.872	0.965	0.861	0.963	0.951	0.966
Scenario 3									
1	0.959	0.705	0.968	0.963	0.985	0.921	0.965	0.979	0.986
2	0.956	0.734	0.967	0.959	0.985	0.914	0.968	0.979	0.985
3	0.952	0.648	0.969	0.948	0.985	0.899	0.966	0.976	0.986
4	0.945	0.625	0.965	0.943	0.984	0.893	0.963	0.976	0.984
5	0.942	0.530	0.956	0.949	0.981	0.912	0.963	0.972	0.981
6	0.948	0.195	0.956	0.956	0.982	0.936	0.970	0.974	0.982
7	0.947	-0.077	0.948	0.961	0.978	0.948	0.968	0.964	0.978
8	0.940	-0.320	0.938	0.958	0.975	0.946	0.967	0.961	0.975
9	0.936	-0.386	0.932	0.959	0.973	0.951	0.967	0.961	0.974
Scenario 4									
1	0.942	0.805	0.953	0.952	0.976	0.912	0.962	0.970	0.980
2	0.927	0.828	0.952	0.935	0.978	0.894	0.964	0.969	0.978
3	0.929	0.786	0.952	0.923	0.976	0.878	0.959	0.963	0.976
4	0.916	0.750	0.935	0.927	0.973	0.893	0.963	0.960	0.973
5	0.912	0.649	0.926	0.929	0.968	0.912	0.963	0.950	0.968
6	0.878	0.524	0.895	0.919	0.959	0.912	0.956	0.942	0.959
7	0.852	0.356	0.875	0.908	0.952	0.907	0.952	0.930	0.952
8	0.821	0.312	0.864	0.881	0.948	0.881	0.948	0.925	0.948
9	0.756	0.259	0.825	0.827	0.929	0.825	0.927	0.910	0.930

**Table A6: Average slope parameters for the CPUE indices derived for each species and model option for the Day based data sets comparing different species distributional assumptions (Scenarios).**

Species	Lognorm	Binom	DeltaLog n	Gamma	DeltaGa mma	Weibull	DeltaWei bull	Tweedie	NegBin
Scenario 1									
1	0.903	0.394	1.260	0.883	1.243	0.848	1.212	0.846	0.999
2	0.854	0.211	1.054	0.815	1.018	0.753	0.959	0.866	1.003
3	0.830	0.312	1.124	0.779	1.076	0.704	1.006	0.828	1.000
4	0.832	0.243	1.077	0.766	1.012	0.682	0.929	0.845	1.003
5	0.948	0.135	1.096	0.867	1.014	0.782	0.927	0.894	1.010
6	1.031	0.034	1.086	0.951	1.004	0.883	0.933	0.818	1.001
7	1.142	-0.059	1.106	1.039	1.002	0.972	0.933	0.876	1.002
8	1.194	-0.118	1.084	1.103	0.994	1.046	0.937	0.872	0.995
9	1.228	-0.155	1.066	1.154	0.993	1.110	0.950	0.863	1.001
Scenario 2									
1	0.858	0.238	1.085	0.788	1.017	0.714	0.945	0.809	1.001
2	0.808	0.285	1.090	0.727	1.011	0.651	0.937	0.849	0.999
3	0.827	0.292	1.111	0.720	1.009	0.641	0.932	0.783	0.997
4	0.891	0.237	1.128	0.772	1.011	0.694	0.934	0.765	1.003
5	0.969	0.191	1.167	0.809	1.008	0.725	0.923	0.837	1.004
6	1.023	0.143	1.171	0.856	1.004	0.780	0.929	0.828	1.001
7	1.031	0.119	1.158	0.874	1.000	0.819	0.945	0.806	0.995
8	1.045	0.100	1.153	0.910	1.018	0.871	0.979	0.845	1.011
9	0.918	0.151	1.074	0.853	1.010	0.832	0.989	0.828	0.997
Scenario 3									
1	0.811	0.215	1.009	0.814	1.015	0.824	1.025	0.879	1.001
2	0.799	0.239	1.036	0.774	1.013	0.765	1.006	0.817	0.998
3	0.833	0.223	1.053	0.790	1.012	0.775	0.997	0.850	1.003
4	0.880	0.204	1.087	0.796	1.005	0.758	0.968	0.824	0.999
5	0.970	0.151	1.132	0.843	1.004	0.795	0.955	0.844	1.002
6	1.102	0.047	1.176	0.929	0.998	0.870	0.937	0.873	0.997
7	1.200	-0.039	1.181	1.015	0.992	0.973	0.949	0.827	0.992
8	1.273	-0.102	1.183	1.097	1.007	1.069	0.978	0.819	1.006
9	1.295	-0.121	1.179	1.103	0.988	1.079	0.964	0.828	0.991
Scenario 4									
1	0.830	0.325	1.143	0.758	1.074	0.737	1.055	0.822	1.001
2	0.816	0.296	1.106	0.725	1.018	0.704	0.998	0.845	1.007
3	0.881	0.285	1.159	0.734	1.018	0.702	0.986	0.834	1.011
4	0.941	0.249	1.191	0.744	0.996	0.709	0.962	0.877	0.993
5	1.075	0.190	1.268	0.810	1.005	0.778	0.972	0.873	1.002
6	1.145	0.137	1.291	0.845	0.989	0.825	0.967	0.844	0.987
7	1.212	0.103	1.319	0.901	1.008	0.899	1.006	0.814	1.006
8	1.179	0.112	1.298	0.895	1.012	0.913	1.031	0.890	1.008
9	1.100	0.121	1.227	0.861	0.991	0.890	1.019	0.848	0.986

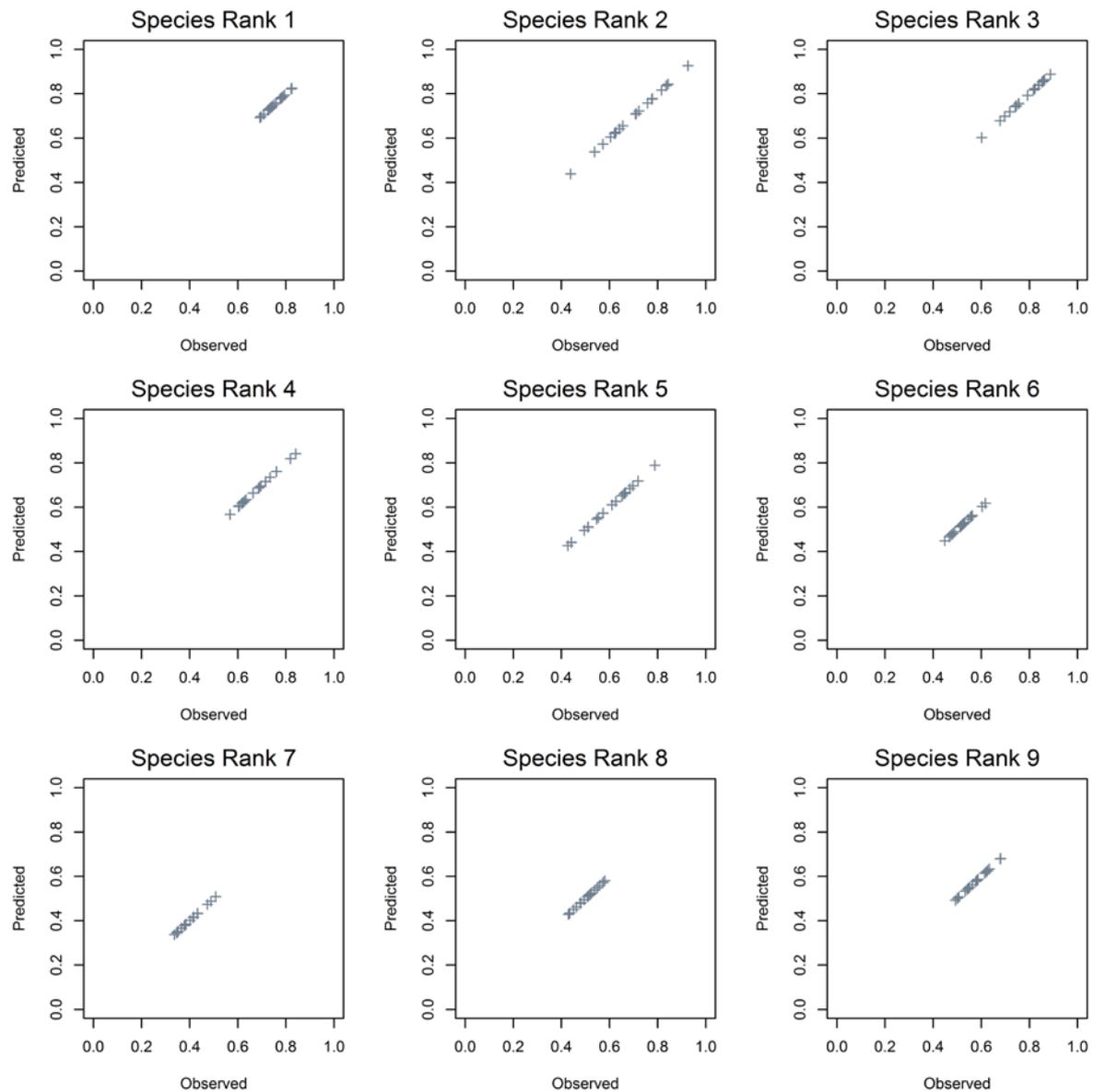
**Table A7: Average correlation coefficients for the CPUE indices derived for each species for comparative model options for the *Day* based data sets comparing different distributions of fishing effort.**

Species	Lognorm	Binom	DeltaLog n	Gamma	DeltaGa mma	Weibull	DeltaWei bull	Tweedie	NegBin
	Scenario 1								
1	0.992	0.584	0.988	0.991	0.989	0.977	0.983	0.993	0.997
2	0.986	0.628	0.995	0.982	0.997	0.957	0.990	0.993	0.997
3	0.977	0.658	0.990	0.968	0.992	0.937	0.987	0.993	0.997
4	0.970	0.655	0.993	0.957	0.996	0.916	0.989	0.992	0.996
5	0.977	0.478	0.992	0.970	0.995	0.947	0.990	0.991	0.995
6	0.980	0.252	0.990	0.973	0.994	0.955	0.989	0.988	0.994
7	0.983	-0.119	0.988	0.978	0.993	0.965	0.988	0.989	0.993
8	0.980	-0.190	0.982	0.976	0.991	0.963	0.984	0.984	0.991
9	0.978	-0.342	0.971	0.974	0.987	0.960	0.981	0.981	0.988
	Trawl Biomass Species1								
1	0.971	0.607	0.922	0.954	0.911	0.894	0.873	0.960	0.963
2	0.972	0.533	0.981	0.967	0.982	0.942	0.975	0.979	0.984
3	0.967	0.583	0.982	0.959	0.983	0.928	0.977	0.979	0.983
4	0.952	0.643	0.980	0.937	0.981	0.895	0.976	0.976	0.981
5	0.954	0.499	0.974	0.941	0.976	0.911	0.969	0.971	0.976
6	0.953	0.193	0.962	0.944	0.967	0.925	0.962	0.962	0.967
7	0.953	-0.054	0.957	0.948	0.962	0.933	0.955	0.955	0.963
8	0.947	-0.215	0.947	0.942	0.954	0.927	0.946	0.949	0.954
9	0.932	-0.286	0.929	0.926	0.940	0.911	0.933	0.932	0.940
	Trawl Biomass All Species								
1	0.992	0.613	0.974	0.987	0.973	0.960	0.958	0.991	0.994
2	0.984	0.692	0.985	0.976	0.984	0.943	0.971	0.990	0.993
3	0.977	0.600	0.993	0.965	0.992	0.924	0.982	0.988	0.992
4	0.977	0.526	0.993	0.966	0.993	0.932	0.985	0.988	0.993
5	0.969	0.488	0.990	0.953	0.990	0.916	0.980	0.984	0.990
6	0.979	0.260	0.988	0.970	0.990	0.949	0.981	0.984	0.990
7	0.979	-0.081	0.984	0.973	0.987	0.957	0.979	0.982	0.987
8	0.974	-0.249	0.977	0.967	0.983	0.951	0.973	0.977	0.983
9	0.972	-0.273	0.971	0.963	0.978	0.943	0.966	0.971	0.978

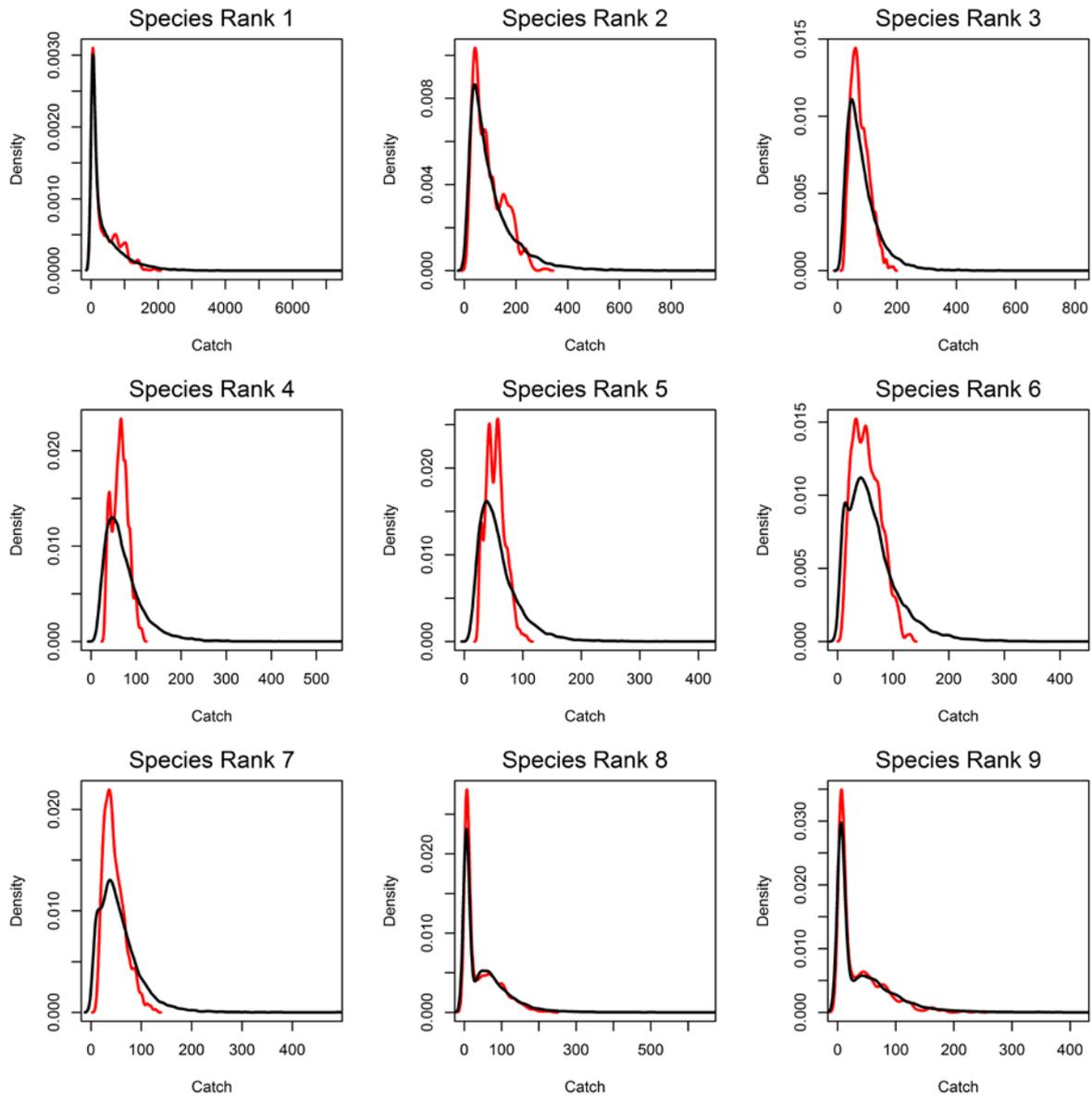
**Table A8: Average slope for the CPUE indices derived for each species and comparable model options for the *Day* based data sets comparing different distributions of fishing effort.**

Species	Lognorm	Binom	DeltaLog n	Gamma	DeltaGa mma	Weibull	DeltaWei bull	Tweedie	NegBin
Scenario 1									
1	0.903	0.394	1.260	0.883	1.243	0.848	1.212	0.846	0.999
2	0.854	0.211	1.054	0.815	1.018	0.753	0.959	0.866	1.003
3	0.830	0.312	1.124	0.779	1.076	0.704	1.006	0.828	1.000
4	0.832	0.243	1.077	0.766	1.012	0.682	0.929	0.845	1.003
5	0.948	0.135	1.096	0.867	1.014	0.782	0.927	0.894	1.010
6	1.031	0.034	1.086	0.951	1.004	0.883	0.933	0.818	1.001
7	1.142	-0.059	1.106	1.039	1.002	0.972	0.933	0.876	1.002
8	1.194	-0.118	1.084	1.103	0.994	1.046	0.937	0.872	0.995
9	1.228	-0.155	1.066	1.154	0.993	1.110	0.950	0.863	1.001
Trawl Biomass Species1									
1	0.848	0.838	1.620	0.838	1.612	0.814	1.590	0.781	0.901
2	0.877	0.243	1.108	0.844	1.076	0.790	1.025	0.836	1.004
3	0.855	0.215	1.061	0.807	1.015	0.739	0.952	0.825	1.004
4	0.840	0.232	1.077	0.769	1.006	0.684	0.922	0.874	0.999
5	0.941	0.145	1.095	0.856	1.009	0.764	0.917	0.851	1.005
6	1.043	0.032	1.099	0.953	1.007	0.877	0.928	0.821	1.005
7	1.137	-0.068	1.083	1.056	1.002	0.999	0.945	0.817	1.001
8	1.180	-0.110	1.085	1.089	0.995	1.034	0.939	0.834	0.998
9	1.205	-0.125	1.081	1.118	0.995	1.067	0.945	0.855	1.001
Trawl Biomass All Species									
1	0.922	0.515	1.395	0.902	1.376	0.863	1.340	0.923	1.028
2	0.850	0.317	1.146	0.818	1.117	0.765	1.068	0.908	1.020
3	0.840	0.229	1.065	0.794	1.021	0.730	0.960	0.804	1.008
4	0.886	0.167	1.056	0.839	1.009	0.777	0.948	0.817	1.001
5	0.913	0.144	1.069	0.838	0.994	0.756	0.910	0.773	0.990
6	1.010	0.037	1.064	0.945	0.998	0.893	0.945	0.810	0.994
7	1.118	-0.066	1.077	1.038	0.996	0.981	0.938	0.836	0.995
8	1.187	-0.125	1.073	1.100	0.987	1.049	0.936	0.864	0.992
9	1.181	-0.126	1.058	1.106	0.984	1.062	0.941	0.837	0.992

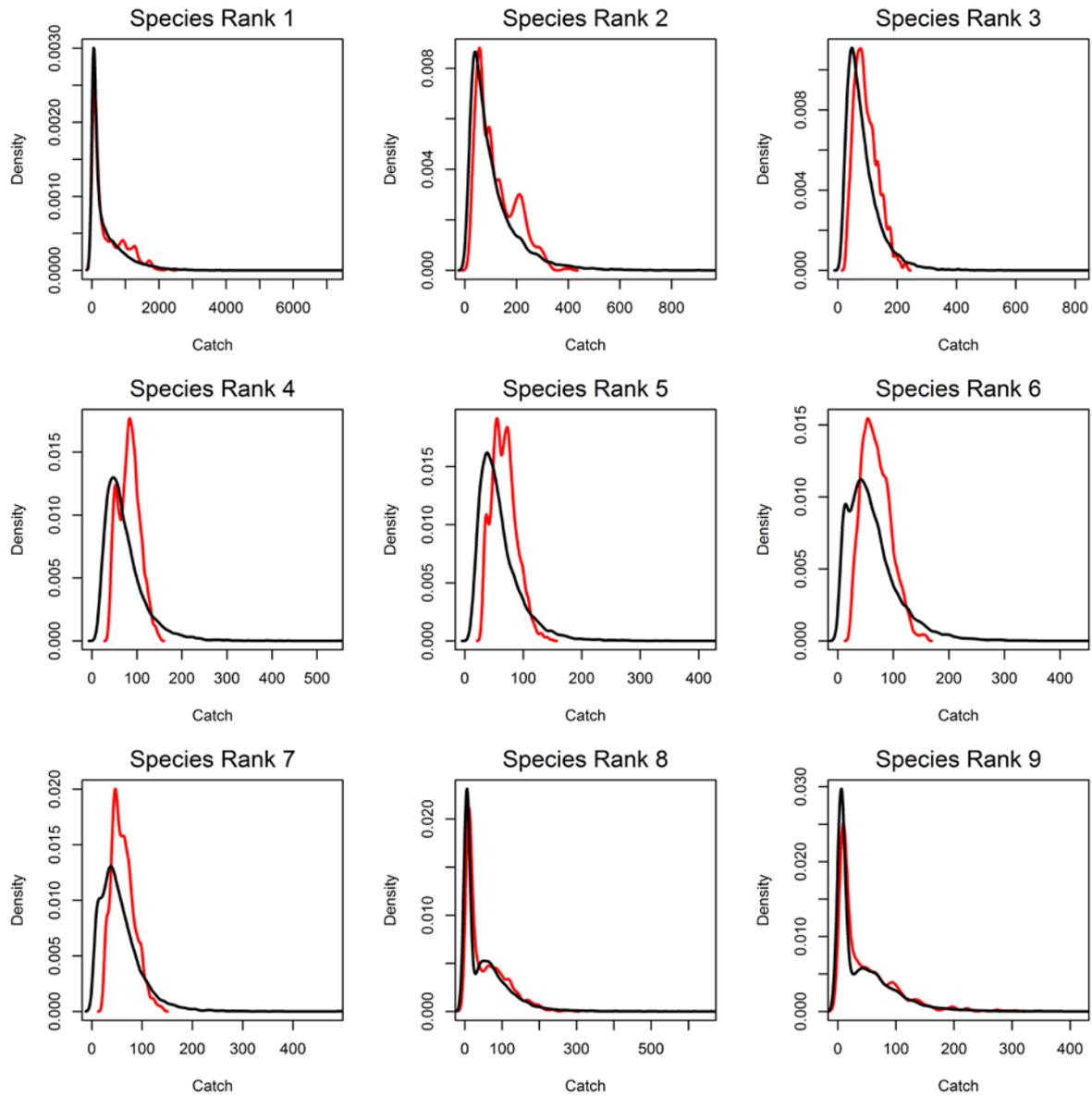
## APPENDIX 2



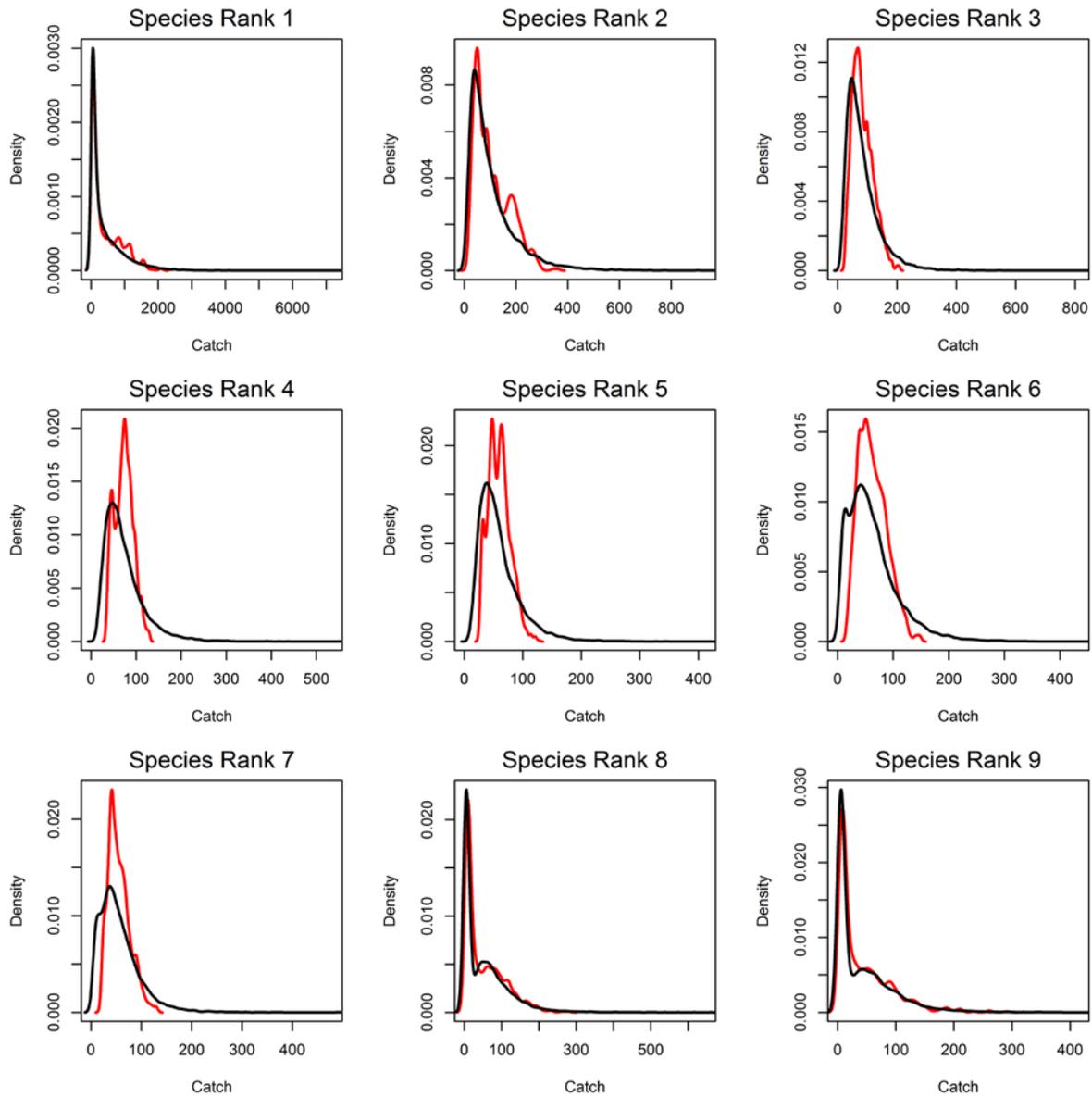
**Figure A1: An example of a comparison the annual observed and predicted proportion of positive catches from Binomial CPUE models by species from one simulated data set (*trawl* data configured for Scenario 1; i.e., species occurrence function A (Figure 1), variation in species distribution SD 0.5, trawl distribution random, reporting module included).**



**Figure A2:** An example of a comparison of density plots of observed (black line) and predicted (red line) positive (non-zero) catches from Lognormal CPUE models by species from one simulated data set (*trawl* data configured for Scenario 1; i.e., species occurrence function A (Figure 1), variation in species distribution SD 0.5, trawl distribution random, reporting module included).



**Figure A3:** An example of a comparison of density plots of observed (black line) and predicted (red line) positive (non-zero) catches from Weibull CPUE models by species from one simulated data set (*trawl* data configured for Scenario 1; i.e., species occurrence function A (Figure 1), variation in species distribution SD 0.5, trawl distribution random, reporting module included).



**Figure A4:** An example of a comparison of density plots of observed (black line) and predicted (red line) positive (non-zero) catches from Gamma CPUE models by species from one simulated data set (*trawl* data configured for Scenario 1; i.e., species occurrence function A (Figure 1), variation in species distribution SD 0.5, trawl distribution random, reporting module included).