**Four problems with the conventional delta-model for biomass sampling data, and one suggested alternative**

Authors: James Thorson

Potential authors TBD.

Fisheries Resource Assessment and Monitoring Division, Northwest Fisheries Science Center, National Marine Fisheries Service, NOAA, Seattle, WA, USA

[James.Thorson@noaa.gov](mailto:James.Thorson@noaa.gov)

Keywords: delta-generalized linear model; hurdle model; Tweedie distribution; spatio-temporal model; encounter probability;

**Abstract**

Ecologists often collect and analyse survey data measured as biomass. Examples include weighing fishes from bottom trawls, plant reproductive output from seed traps, or insect density from light traps. In these and other cases, sampling will often result in many zeros (samples where a given taxon is not encountered). These data are often analysed using two generalized linear models (one for encounter probability, and the other for sampled biomass given an encounter), and the combination is called a “delta” (or hurdle) model. I first discuss three theoretical problems with delta-models, including: (1) the assumed independence of encounter probability and biomass given an encounter; (2) difficulties in interpreting the impact of coefficients for the encounter-probability component on predicted population density; and (3) difficulties in restricting parameters for either model component (e.g., to potentially increase model parsimony when data are sparse). I then derive an alternative model from an assumed Poisson-process for individuals in the vicinity of sampling, and discuss how this model responds to these theoretical problems. Finally, I obtain biomass sampling data for 125 fishes in eight marine ecosystems, and compare relative performance of conventional and Poisson-process delta-models. This comparison shows that the Poisson-process model improves fit for >80% of populations in 7 of 8 ecosystems. It also informs estimates of sampled biomass using information about predicted encounter probability, thereby decreasing the average magnitude of residual spatial and spatio-temporal variation. However, both models have similar confidence-interval width when predicting total population biomass. I conclude that the Poisson-process model is a useful alternative to the conventional delta-model both theoretically and empirically, and recommend ongoing research regarding efficient approximations to otherwise intractable processes arising in ecology.

**Introduction**

Ecologists often estimate biological rates (e.g., survival, stage-transition probabilities, per-capita productivity) by fitting ecological models to available data. Data often arises from biological surveys, where observers visit a defined site and either record the species they encounter (occupancy data) or measure the quantity of each species (counts or weights). In some cases, sampling can be replicated on a “demographically closed” segment of a population (i.e., a local scale with no births, deaths, or movement), and this replicated sampling allows estimation of detection probabilities (MacKenzie et al. 2003, Royle and Dorazio 2008). However, the assumption of demographic closure is not justified for many sampling programs, including the majority of marine-fish sampling programs, and in these cases it is common to fit models directly to sampling data while assuming detection probabilities are constant across space and time (Kristensen et al. 2014, Thorson et al. 2015a).

Some of the most common ecological analyses include fitting a regression model to sampling data. Examples include species distribution models, climate envelope analysis, and habitat utilization models. Survey data for ecological communities often include a few species that are frequently encountered, and a long list of species that are encountered rarely (McCune et al. 2002). Species may be rarely encountered either because they occur at low densities (and hence require frequent sampling to have a high probability of detection), or because the majority of sampled habitats occur outside their occupied habitats (Harte 2011). Thus, a sample with a zero-count for a given species could arise because it is outside the species’ habitat (a “true” zero), or because it was not detected despite being present (a “false” zero, Martin et al. (2005)). Therefore, ecologists frequently analyse count data as arising from a combination of two process: the expected probability distribution for counts in occupied habitats, and an additional probability of zero-counts arising from sampling in unoccupied habitats. This “hurdle” model has two components: the probability distribution for counts, and the probability of “excess” zeros. Each probability can be usefully interpreted, e.g., to infer the proportion of unoccupied habitats from the probability of excess zeros.

Alternatively, ecologists will sometimes analyse survey data obtained as the weight of species encountered. For example, this is common in marine fish sampling, where thousands of fishes will be captured simultaneously by a trawl gear. In this case, it is easiest to sort catch to species, weigh the catch for each species, and potentially subsample to determine weight, sex and age (which can then be used to estimate catch in numbers). Other examples of sampling species weight include insect traps and leaf-litter traps (Clark 2016). In each case, sampling data yields some proportion of zeros (e.g., where no fish of a given species were encountered), and also a continuous-valued measure of density (e.g., catch-weights where a fish was encountered). These data are often analysed using a “delta-model” (Lo et al. 1992, Stefansson 1996) that again includes two components: the probability of encountering the species, and the probability distribution for weights given that the species is encountered.

This “delta-model” remains one of the most common types of regression used by ecologists and fisheries scientists (Maunder and Punt 2004, Zuur et al. 2009). However, it has several theoretical and practical draw-backs. In the following, we first define the delta-model in detail, and outline three theoretical problems with its use. We then define an alternative “Poisson-process” model for analysing similar data, and describe how this “Poisson-process” model rectifies these theoretical problems, while also having additional potential for model parsimony. Finally, we use data from 120 fishes from six marine ecosystems in North America and Europe to show that the proposed “Poisson-process” model often has better fit and reduces unexplained variation relative the conventional delta-model.

**Methods**

**The conventional delta-model**

Fisheries scientists have analysed catch-rate data using “delta-models” for over thirty years (Lo et al. 1992, Stefansson 1996). Historically, these delta-models have been fitted to data by estimating parameters for two separate and independent generalized linear models (GLMs):

1. *Encounter probability*: the probability of encountering the species at a given place and time;
2. *Positive catch rates*: the catch in weight given that the species is encountered.

In particular, the “encounter probability” component of the delta-model defines the probability that catch at location and time is non-zero, . Historically, is typically modelled via a link function, where is a linear function of fixed and random effects. In the following, I use a logit link function, . The “positive catch rate” component then defines the probability density function for positive catch rates:

|  |  |
| --- | --- |
|  | (1) |

where is a probability density function, and in the following we use a lognormal distribution with a log-link, i.e., , and where is modelled via a linear predictor. Given these two model components, population density at location and time can be predicted, , and useful population metrics (total abundance, center of distribution, or effective population area) can be easily calculated (Thorson et al. 2016b).

**Three “theoretical” problems with conventional delta-models**

I see three major draw-backs to using the conventional delta-model:

1. Independence between model components;
2. Difficulties in interpreting covariates;
3. Difficulties in restricting parameters when data are sparse.

I first discuss these three reasons in detail, before also presenting an empirical argument against using the conventional delta-model for marine fish sampling data.

*Independence among components:* Using the conventional delta-model, the “encounter probability” and “positive catch rate” components are statistically independent, i.e., knowledge about (or even that the species was encountered, ) gives no information about the distribution for positive catches . This assumption is contrary to a large body of evidence suggesting (1) that abundant species have wide ranges, such that frequently encountered species also have higher density throughout their range (Gaston 1994), and also (2) that an increase in local density will directly cause an increased probability of encounter at that site (Royle and Nichols 2003). Both phenomena suggest that a location with increased probability of encounter (higher ) will tend to have greater catch rates given an encounter (higher ).

*Difficulties in interpreting coefficients*: Using the conventional delta-model, an ecologist can include covariates affecting the probability of detection, and/or the positive catch rate. Specifically, the logit- (or probit-)transformed encounter probability is often specified using a linear predictor:

|  |  |
| --- | --- |
|  | (2) |

where and are predictors for fixed-effects and random effects associated with the th observation. Fixed and random-effects then affect the “odds-ratio” for encounter probability. However, it is easy (or even feasible) to summarize the effect of covariates on population density . Furthermore, a random effect may have a variance of , but there is no closed-form equation for calculating the resulting variance in population density . We suspect that many ecologists would prefer to estimate the impact of a covariate (e.g., bottom temperature) on expected fish densities, rather than the “odds ratio” (the ratio of encounter probability and non-encounter probability).

*Difficulties in restricting parameters*: Ecologists frequently have relatively little data with which to estimate a multitude of potential ecological processes. “Tapering effects” (i.e., the presence of many ecological processes with gradually declining effect sizes for any given system) has driven interest in identifying “parsimonious” ecological models (Burnham and Anderson 2002). Parsimony in this case is defined as an appropriate number of parameters that minimizes total predictive error (simultaneously low bias and imprecision). In many cases, parsimony is achieved by identifying a flexible family of models, where analysts can allow data to discriminate the appropriate degree of model complexity. However, a delta-model will generally require estimating covariates (fixed effects) and unexplained variation (random effects) for both encounter probability and positive catch rates. For example, recent spatio-temporal models have generally estimated spatial variation in both model components (Shelton et al. 2014, Thorson et al. 2015b). Eliminating spatial for either component requires assuming that either encounter probabilities or positive catch rates are perfectly explained by measured covariates.

**Solutions from using an alternative “Poisson-process” model**

To address these theoretical problems, I propose an alternative “Poisson process” version of this delta-model. This alternative is derived by defining as the average population density (in numbers) over a time interval in the neighbourhood to , where the precise number of individuals at that location in time follows a Poisson process. The encounter probability from this Poisson process is then:

|  |  |
| --- | --- |
|  | (3) |

where is the proportion of habitat in the vicinity to that is never occupied regardless of population density (i.e., the probability of a “true zero”, bounded from zero to one), such that as . is then modelled via a log-linked linear predictor:

|  |  |
| --- | --- |
|  | (4) |

These equations reduce to a complementary log-log link for encounter probability as a function of log-density (in numbers), , whenever . We then derive the expectation of positive catch rates from the definition of density:

|  |  |
| --- | --- |
|  | (5) |

where is average weight per group of individuals. can optionally be specified via log-linked linear predictor:

|  |  |
| --- | --- |
|  | (6) |

although the model is also interpretable if is constant over space and time (i.e., and ). This model implies that locations with above-average encounter probabilities will also have above-average positive catch rates , where this correlation is stronger for frequently-encountered species and approaches zero when encounter probability is low. Population density (in biomass) is predicted as the product of density (in numbers) and average weight, . This derivation has many similarities to a compound Poisson-gamma distribution (a member of the Tweedie family when the power parameter ), although it has a different mean-variance relationship (Appendix S1).

This model ameliorates all three theoretical problems with the conventional delta-model:

1. *Independence among components:* The “poisson-process” model induces a correlation between encounter probability and positive catch rates that is interpretable biologically. When densities are low (), an increase in density results in an increase in encounter probability (). In this case, increasing density results in a greater proportion of encounters, which each are likely to encounter a single individual. As density becomes large (), encounter probability will plateau (), and further increases in density are accompanied by increasing positive catch rates (). The dependency between density , encounter probability , and positive catch rate is therefore interpretable biologically.
2. *Difficulties in interpreting coefficients:* The Poisson-process model also simplifies interpretation of covariates. In particular, covariates and both predict changes in log-density, so e.g., a 0.01 increase in is associated with approximately a 1% increase in density. Similarly, a random effect with a standard deviation of 0.01 explains approximately a 1% standard deviation in density. Both fixed- and random-effects therefore have a similar interpretation in predicting variation in population density.
3. *Difficulties in restricting parameters:* Finally, the Poisson-process model allows for clear restrictions on parameters to achieve parsimony. For example, an ecologist may estimate covariates while restricting , thus eliminating many potential parameters. In this case, covariates are still affecting both encounter probability and positive catch rates . More importantly, an ecologist may want to specify a biological model for changes in local population density over time (Thorson et al. In press, Kristensen et al. 2014). In this case, the analyst can specify and such that population density (in units biomass) is explained by a single linear predictor (e.g., derived from last year’s density, surplus production, and removals in Thorson et al. (In press)). In this case, a single variable (population density) can approximate changes in both encounter probability and positive catch rates (Fig. 1).

**Spatio-temporal index standardization**

I compare these two parameterizations using a spatio-temporal modelling framework that includes both spatial and spatio-temporal variation, and which estimates a fixed effect for each year. For the conventional delta-model, I specify:

|  |  |
| --- | --- |
|  | (7a) |
|  | (7b) |

where:

|  |  |
| --- | --- |
|  | (8a) |
|  | (8b) |

and where is the spatial correlation given estimated decorrelation distance , is the estimated pointwise variance of spatial variation in , is the estimated pointwise variance of spatio-temporal variation in , and and are defined identically but with separate estimates of spatial variance and spatio-temporal variance (Thorson et al. 2015b). For the alternative Poisson-process model, I specify:

|  |  |
| --- | --- |
|  | (9a) |
|  | (9b) |

where spatial and spatio-temporal terms (e.g., Eq. 8a-8b) are defined analogously to the conventional delta-model.

Parameters for both models are estimated using maximum marginal likelihood, using the Laplace approximation to approximate the integral across the joint probability of fixed and random effects. Parameter estimation is performed using package *VAST* ([www.github.com/nwfsc-assess/VAST](http://www.github.com/nwfsc-assess/VAST)), which is implemented using Template Model Builder (Kristensen et al. 2016) within the R statistical platform (R Core Team 2015). Model selection is conducted using the Akaike information criterion (Akaike 1974), based on the marginal likelihood and the number of fixed effects.

After estimating parameters, I then evaluate model performance by comparing the estimated standard deviation of spatial and spatio-temporal for “encounter probability” in the conventional model with the same parameters for “average weight” in the alternative model. If the alternative model is parsimonious, I expect that the correlation between encounter probability and positive catch rates in the “Poisson-process” model will result in a lower variance for spatial and spatio-temporal terms of average weight than the comparable variance parameters for positive catch rates in the conventional version.

**Case study data: Bottom trawl survey database**

I seek to evaluate the relative goodness-of-fit of the conventional and alternative Poisson-process models to fisheries data. To do so, I obtain bottom-trawl survey data from eight marine ecosystems:

1. *Eastern Bering Sea* – Survey operated by the Alaska Fisheries Science Center (AFSC) obtained from a fixed-station design;
2. *Gulf of Alaska* – Survey operated by the AFSC obtained from a randomized design;
3. *Aleutian Islands* – Survey operated by the AFSC obtained from a randomized design;
4. *US West Coast* –The West Coast groundfish bottom trawl survey (Bradburn et al. 2011) operated by the Northwest Fisheries Science Center (NWFSC), obtained from a stratified-random design;
5. *North* *Sea –* The North Sea international bottom trawl survey (NS*-*IBTS), restricting data to 1991-2015 obtained using a “Gov” gear in quarter 1 (winter);
6. *Baltic Sea –* The Baltic international trawl survey (BITS), restricting data to 1991-2015 obtained using a “Gov” gear in quarter 1;
7. *Scottish West Coast* – The Scottish West Coast international bottom trawl survey (SWC-IBTS), restricting data to 1991-2015 obtained using a “Gov” gear in quarter 1;
8. *Celtic Sea* –The French demersal survey (EVHOE) of the Celtic Sea, operating from 1997-2015 in quarter 4 (fall).

Surveys 1-5 included data for many species, and we restricted data to the twenty most frequently encountered fishes (see Fig. 2 for annual sample sizes). Other surveys had publicly available data for fewer than twenty species, so we used as many as were available. All surveys are publicly available and can be accessed using R package *FishData* (<https://github.com/james-thorson/FishData>), which in turn uses R package *icesDatras* (<https://github.com/ices-tools-prod/icesDatras>) to download data for surveys 5-8.

I fitted both conventional and Poisson-process models to data for each species in each survey, and summarize results for models all converged models (i.e., with a maximum absolute-gradient < 0.01 and a positive-definite Hessian matrix). In particular, I record (1) the proportion of species for each region where the conventional or alternative model was selected as parsimonious using the Akaike Information Criterion (AIC); (2) the pointwise (a.k.a. marginal) standard deviation of spatial and spatio-temporal variance for both model components; (3) the predictive standard deviation of an abundance index derived from each model (indices are area-weighted following Thorson et al. (2015b)). I hypothesize that the Poisson-process model will be more parsimonious than the conventional delta-model for the majority of species. The pointwise variances and from the conventional model and and from the alternative model are directly comparable (both affect the distribution of catches given an encounter, , via a log-link function), and I hypothesize that spatial and spatio-temporal variances for the alternative model will be lower because the encounter probability (estimated from proportion of nearby samples that encounter the species) is informative about local positive catch rates .

**Results**

Data from eight regions results in converged estimates of parameters for both conventional and alternative models for 125 populations total (Fig. 3). In general, the alternative model results in better fit (a high log-likelihood of available data) for the vast majority of populations in 7 out of 8 regions. Both models have an identical number of estimated parameters, so a high log-likelihood also indicates greater parsimony (e.g., using the Akaike Information Criterion). The average AIC weight for the two models is >80% for the same 7 regions. The exception is for the California Current, where the two models are each selected 10 out of 20 times. In this region, the implied correlation between encounter probability and positive catch rates apparently does not improve model fit. However, the implied correlation does improve fit for the majority of populations in other regions.

The conventional and alternative models have essentially identical estimates of residual variation (), indicating that both models attribute a roughly identical portion of variation to the combination of spatial and spatio-temporal variation (Fig. 4a). As hypothesized, however, the Poisson-process model results in a lower standard deviation for spatial and spatio-temporal variation (Fig. 4b). The standard deviation is not directly comparable for the first-model component between models, because and (from the conventional model; top row of Fig. 4b) affect via a logit-link function while and (from the alternative model; middle row of Fig. 4b) affect via a complementary log-log link function. However, the standard deviations for residual variation and for the second-component are comparable (both , and , affect positive catch rates via log-link function). For this second component (Fig. 4b, bottom row), the delta-model has a pointwise standard deviation of 1.47, whereas the Poisson-process model has 1.10. Therefore, including local densities and encounter probabilities ( as a predictor of shrinks the magnitude of unexplained spatial variation by . Similarly, the Poisson-process model shrinks the magnitude of unexplained spatio-temporal variation by on average across populations.

Despite resulting in better fit and also shrinking the magnitude of explained variation in positive catch rates, the Poisson-process model does not consistently decrease the log-standard deviation of confidence intervals for estimated abundance indices. Across all eight regions, the Poisson process model has similar or slightly wider confidence intervals (0-4% wider) for six regions, and results in a substantially smaller interval width (10-18% narrower) for two regions (Gulf of Alaska, and Aleutian Islands). Even for these two regions, the confidence intervals widths are very similar between models, and the difference arises primarily due smaller intervals for 1-2 populations in each region.

**Discussion**

I have presented three theoretical arguments for why the conventional delta-model is unsatisfactory, including the lack of dependence between encounter probability and positive catch rates, difficulties in interpreting how parameters in the encounter-probability component affect population density, and difficulties in restricting parameters for either model component. I have then shown how these three difficulties are addressed using a new “Poisson-process” model, derived from the biological assumption that groups of individuals follow a Poisson-process in the vicinity of sampling. Application to 125 populations in eight marine regions shows that the Poisson-process model substantially improves fit for, uses knowledge of encounter probabilities to decrease otherwise-unexplained variation in positive catch rates, but doesn’t consistently decrease confidence-interval width for abundance indices. I therefore conclude that the Poisson-process model is not likely to substantially increase the information available to stock assessments when used to estimate abundance indices. However, improvements in fit, interpretability, and parsimony are still likely to be useful when estimating habitat maps, estimating habitat associations, and fitting ecological models to samples of fish biomass.

Restricted version of the Poisson-process model have already proved useful when fitting ecological models directly to catch-rate data sampled in species biomass. For example, Thorson et al. (2016a) specified for all samples (i.e., and ) to jointly model survey data for 10 groundfishes in the Eastern Bering Sea, and Thorson et al. (In press) used to same restriction to approximate individual movement within a spatial population-growth model for a big skate in the California Current. In these cases, the ability to restrict variation in precluded estimating nearly half of all random effects, and therefore allowed afforded greater computational complexity in other model components. However, the present study suggests that substantial variation remains in even when using the Poisson-process model. When possible, I therefore recommend testing for evidence of variable “average weight” in excess of predictions from a Poisson-process model for local densities.

The Tweedie (a.k.a. “compound Poisson-gamma”) distribution is a noteworthy alternative for modelling species biomass data (given restrictions on the index parameter: ) The Tweedie distribution is derived from the assumption that catch rates arise from a Poisson distribution for the number of captured individuals, and a Gamma distribution for the average weight of individuals captured in that sample. At its most restricted, the Tweedie distribution has three parameters: the expected number of individuals, the average weight of individuals, and the variation in average individual weight among samples. Using these three parameters, the Tweedie distribution generates a “power-law” (2-parameter) relationship between sample mean and variance, (where is the index parameter). By contrast, our Poisson-process model implies that , and therefore is less flexible about the mean-variance relationship than the Tweedie model (see Appendix S1 for more details). However, the Tweedie distribution has no closed-form equation for calculating the log-likelihood (Dunn and Smyth 2005), and parameters are typically estimated either by approximating the sum of an infinite series (i.e., by summing from zero to an arbitrary upper bound; Foster and Bravington (2012)) or using Markov-chain sampling (Lecomte et al. 2013). Both approaches are still computationally challenging, i.e., where the computational order scales with the upper bound of the approximation. By contrast, the “Poisson-process” model allows efficient and closed-form calculation of the model likelihood, and is intended to ease the computational burden to afford greater computational detail for other model components.

I envision several useful avenues for future research. Most obviously, the Poisson-process model could be compared with the Tweedie model in cases where it is computationally feasible to do so. This comparison is unlikely to be feasible given the spatio-temporal models explored here, e.g., because previously published applications of the Tweedie distribution have not included either spatial or spatio-temporal variation (Foster and Bravington 2012, Lecomte et al. 2013). I also recommend future research exploring the potential consequences of restricting (as used in previous spatio-temporal models for biomass data). Application to 125 populations worldwide shows that there is substantial variation in even after accounting for the effect of encounter probabilities, but determining the impact of this restriction on model performance will require a simulation experiment. This simulation experiment could easily be conditioned on the range of spatial and spatio-temporal variances estimated in this study.

Finally, the past decade has seen rapid growth in a variety of useful approximations for otherwise slow or intractable models in ecology. Examples include approximating individual demographics using Markov chains (Hubbell 2011), estimating mixed-effects models by maximizing the Laplace approximation to the marginal likelihood (Skaug and Fournier 2006, Kristensen et al. 2016), or approximating spatial variation using finite-element analysis methods (Lindgren et al. 2011). Collectively, these approximations are useful when they permit the development of models with increased realism regarding otherwise-neglected components of ecological systems (e.g., linkages between regional and local species pools; Hubbell (2011)). In this light, the Poisson-process model can be viewed as a computationally-efficient approximation to the Tweedie model. I recommend ongoing development and testing of efficient approximations to ecological processes, and hope that these approximations will collectively allow biological rates (births, deaths, and movement) to be simultaneously estimated for entire communities occurring on heterogenous landscapes using available data worldwide (e.g., filling in white spaces in Fig. 1).

**Acknowledgements**

**Bibliography**

Akaike, H. 1974. New look at statistical-model identification. IEEE Trans. Autom. Control **AC19**(6): 716–723.

Bradburn, M.J., Keller, A.A., and Horness, B.H. 2011. The 2003 to 2008 US West Coast bottom trawl surveys of groundfish resources off Washington, Oregon, and California: estimates of distribution, abundance, length, and age composition. NOAA Tech. Memo., US Department of Commerce, National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Northwest Fisheries Science Center, Seattle, WA.

Burnham, K.P., and Anderson, D. 2002. Model Selection and Multi-Model Inference. *In* 2nd edition. Springer, New York.

Clark, J.S. 2016. Why species tell more about traits than traits about species: Predictive analysis. Ecology.

Dunn, P.K., and Smyth, G.K. 2005. Series evaluation of Tweedie exponential dispersion model densities. Stat. Comput. **15**(4): 267–280.

Foster, S.D., and Bravington, M.V. 2012. A Poisson–Gamma model for analysis of ecological non-negative continuous data. Environ. Ecol. Stat. **20**(4): 533–552. doi:10.1007/s10651-012-0233-0.

Gaston, K.J. 1994. Measuring geographic range sizes. Ecography **17**(2): 198–205. doi:10.1111/j.1600-0587.1994.tb00094.x.

Harte, J. 2011. Maximum entropy and ecology: a theory of abundance, distribution, and energetics. Oxford University Press, Oxford, UK. [accessed 1 March 2016].

Hubbell, S.P. 2011. The Unified Neutral Theory of Biodiversity and Biogeography (MPB-32). Princeton University Press.

Kristensen, K., Nielsen, A., Berg, C.W., Skaug, H., and Bell, B.M. 2016. TMB: Automatic Differentiation and Laplace Approximation. J. Stat. Softw. **70**(5): 1–21. doi:doi: 10.18637/jss.v070.i05.

Kristensen, K., Thygesen, U.H., Andersen, K.H., and Beyer, J.E. 2014. Estimating spatio-temporal dynamics of size-structured populations. Can. J. Fish. Aquat. Sci. **71**(2): 326–336. doi:10.1139/cjfas-2013-0151.

Lecomte, J.-B., Benoît, H.P., Ancelet, S., Etienne, M.-P., Bel, L., and Parent, E. 2013. Compound Poisson-gamma vs. delta-gamma to handle zero-inflated continuous data under a variable sampling volume. Methods Ecol. Evol. **4**(12): 1159–1166. doi:10.1111/2041-210X.12122.

Lindgren, F., Rue, H., and Lindström, J. 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. J. R. Stat. Soc. Ser. B Stat. Methodol. **73**(4): 423–498. doi:10.1111/j.1467-9868.2011.00777.x.

Lo, N.C., Jacobson, L.D., and Squire, J.L. 1992. Indices of Relative Abundance from Fish Spotter Data based on Delta-Lognormal Models. Can. J. Fish. Aquat. Sci. **49**(12): 2515–2526.

MacKenzie, D.I., Nichols, J.D., Hines, J.E., Knutson, M.G., and Franklin, A.B. 2003. Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. Ecology **84**(8): 2200–2207.

Martin, T.G., Wintle, B.A., Rhodes, J.R., Kuhnert, P.M., Field, S.A., Low-Choy, S.J., Tyre, A.J., and Possingham, H.P. 2005. Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. Ecol. Lett. **8**(11): 1235–1246.

Maunder, M.N., and Punt, A.E. 2004. Standardizing catch and effort data: a review of recent approaches. Fish. Res. **70**(2-3): 141–159. doi:10.1016/j.fishres.2004.08.002.

McCune, B., Grace, J.B., and Urban, D.L. 2002. Analysis of ecological communities. MjM software design Gleneden Beach, OR. Available from http://tocs.ulb.tu-darmstadt.de/106964135.pdf.

R Core Team. 2015. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Available from https://www.R-project.org/.

Royle, J.A., and Dorazio, R.M. 2008. Hierarchical modeling and inference in ecology: the analysis of data from populations, metapopulations and communities. *In* 1st edition. Academic Press, London.

Royle, J.A., and Nichols, J.D. 2003. Estimating abundance from repeated presence-absence data or point counts. Ecology **84**(3): 777–790.

Shelton, A.O., Thorson, J.T., Ward, E.J., and Feist, B.E. 2014. Spatial semiparametric models improve estimates of species abundance and distribution. Can. J. Fish. Aquat. Sci. **71**(11): 1655–1666. doi:10.1139/cjfas-2013-0508.

Skaug, H., and Fournier, D. 2006. Automatic approximation of the marginal likelihood in non-Gaussian hierarchical models. Comput. Stat. Data Anal. **51**(2): 699–709.

Stefansson, G. 1996. Analysis of groundfish survey abundance data: combining the GLM and delta approaches. ICES J Mar Sci **53**(3): 577–588.

Thorson, J.T., Ianelli, J.N., Larsen, E.A., Ries, L., Scheuerell, M.D., Szuwalski, C., and Zipkin, E.F. 2016a. Joint dynamic species distribution models: a tool for community ordination and spatio-temporal monitoring. Glob. Ecol. Biogeogr. **25**(9): 1144–1158. doi:10.1111/geb.12464.

Thorson, J.T., Ianelli, J.N., Munch, S.B., Ono, K., and Spencer, P.D. 2015a. Spatial delay-difference models for estimating spatiotemporal variation in juvenile production and population abundance. Can. J. Fish. Aquat. Sci. **72**(12): 1897–1915. doi:10.1139/cjfas-2014-0543.

Thorson, J.T., Jannot, J.E., and Somers, K. In press. Using spatio-temporal models of population growth and movement to monitor overlap between human impacts and population density. J. Appl. Ecol.

Thorson, J.T., Pinsky, M.L., and Ward, E.J. 2016b. Model-based inference for estimating shifts in species distribution, area occupied and centre of gravity. Methods Ecol. Evol. **7**(8): 990–1002. doi:10.1111/2041-210X.12567.

Thorson, J.T., Shelton, A.O., Ward, E.J., and Skaug, H.J. 2015b. Geostatistical delta-generalized linear mixed models improve precision for estimated abundance indices for West Coast groundfishes. ICES J. Mar. Sci. J. Cons. **72**(5): 1297–1310. doi:10.1093/icesjms/fsu243.

Zuur, A.F., Ieno, E.N., Walker, N., Saveliev, A.A., and Smith, G.M. 2009. Mixed Effects Models and Extensions in Ecology with R. *In* 1st edition. Springer, New York.

Fig. 1 –

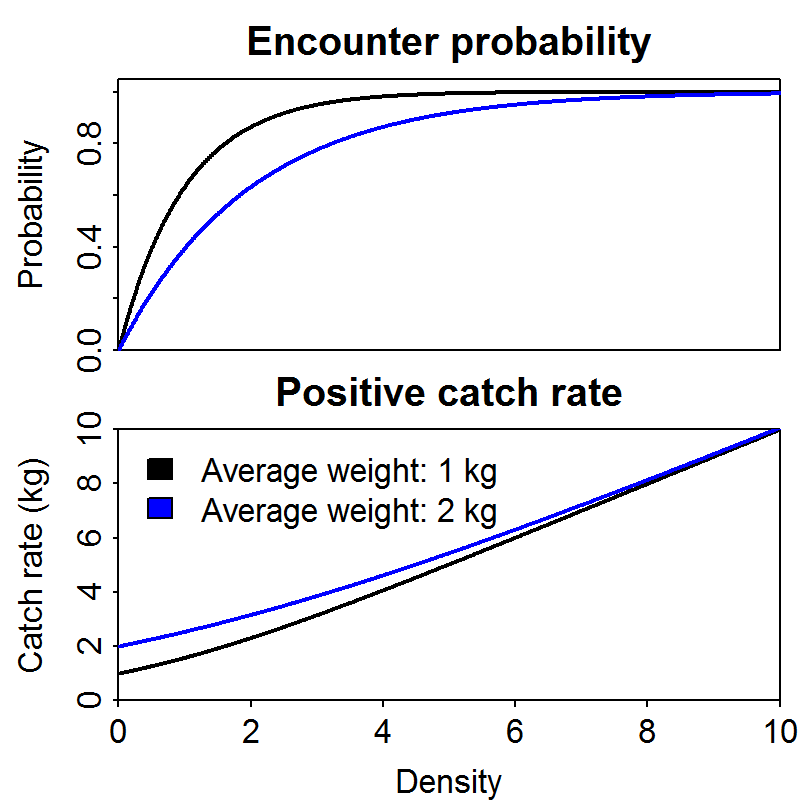


Fig. 2 – Spatial coverage (top panel) and availability (bottom panel) for eight bottom trawl surveys with publicly available Application Programming Interfaces, used for comparing performance of conventional and “Poisson-process” delta-models (colors are identical between panels, and can be used to match data availability to the spatial coverage for each survey).

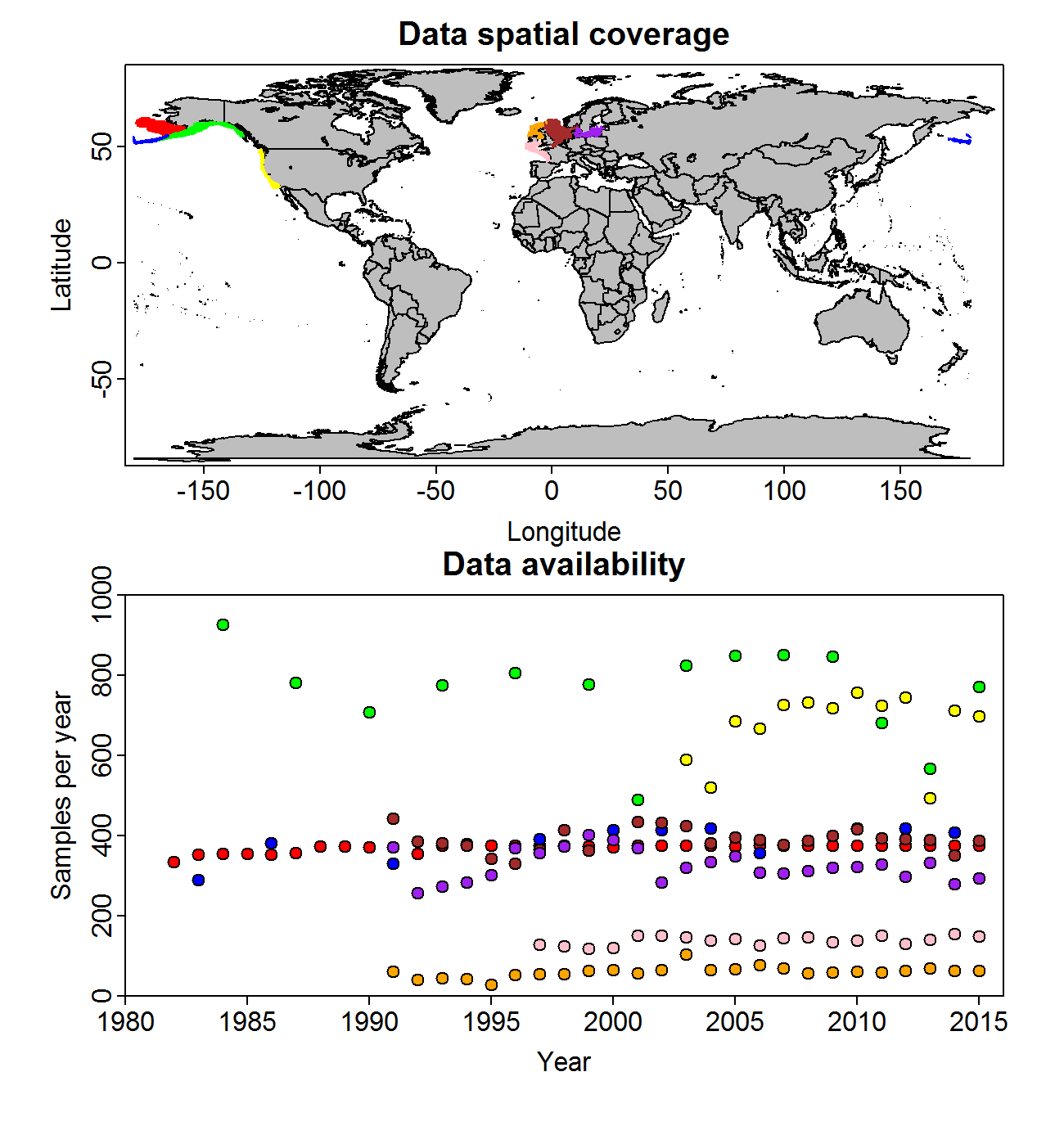


Fig. 3 – Selection of the conventional delta-lognormal or an alternative “Poisson-process” model for survey catch rate data in four bottom trawl surveys

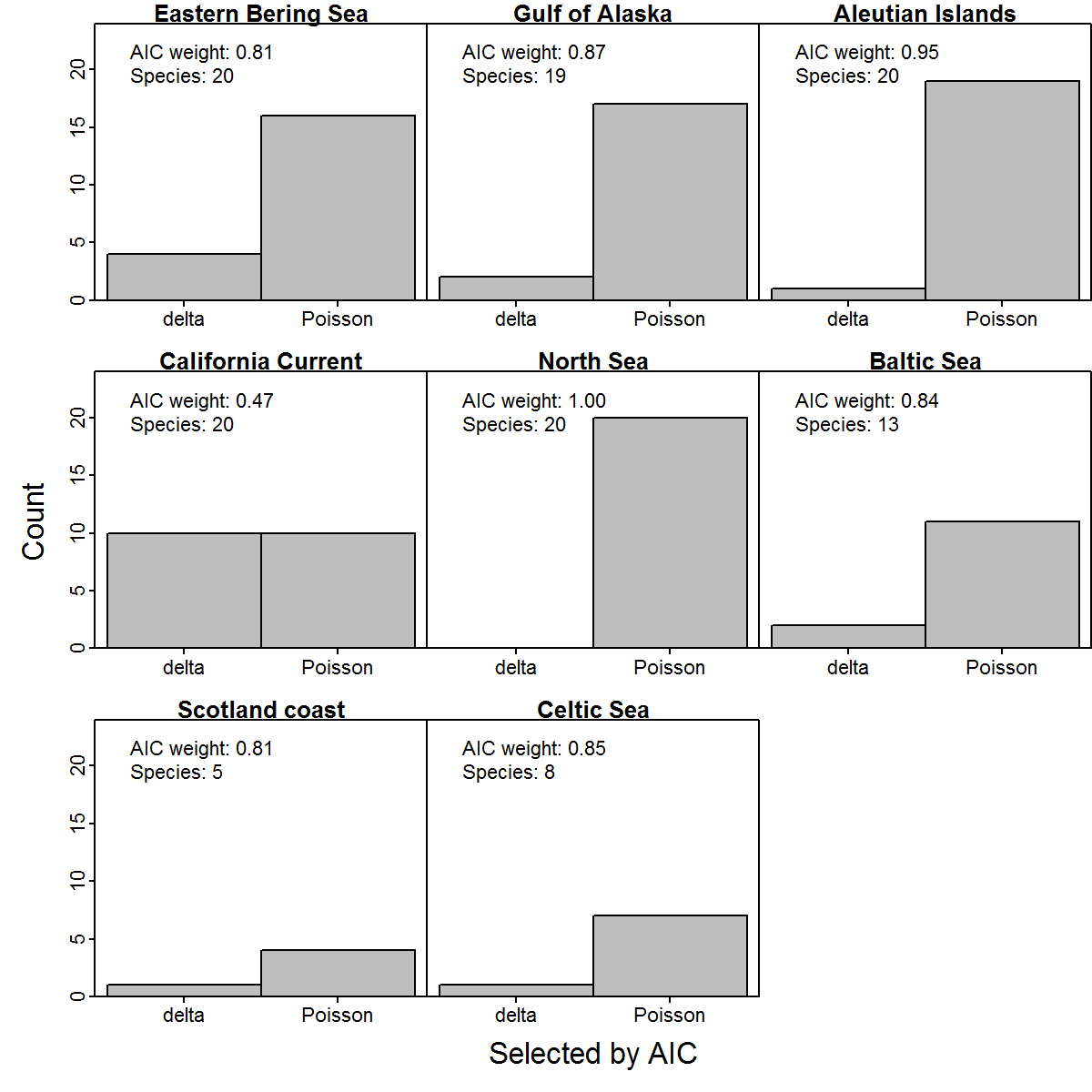


Fig. 4a – Standard deviation estimates across all 125 stocks in eight surveys for residual variation in positive catch rates (see Fig. 4b caption for details).

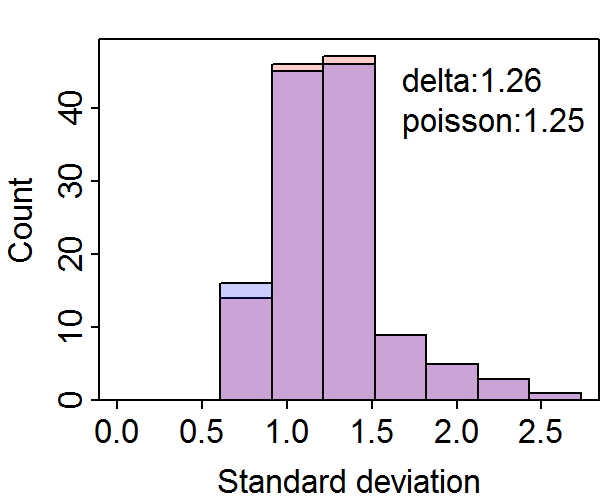


Fig. 4b – Standard deviation estimates across all 125 stocks in eight surveys for spatial (left column) and spatio-temporal variation (right column) using the conventional delta-model (red) or alternative Poisson-process version (blue), where standard deviations for encounter probability (top row) and density in numbers (middle row) are not directly comparable (because encounter probability uses a logit-link, while density uses a log-link), but where standard deviations for positive catch rates and average weight (bottom row) are directly comparable (because both use a log-link). In each panel, we display the average standard deviation in the top-right corner.

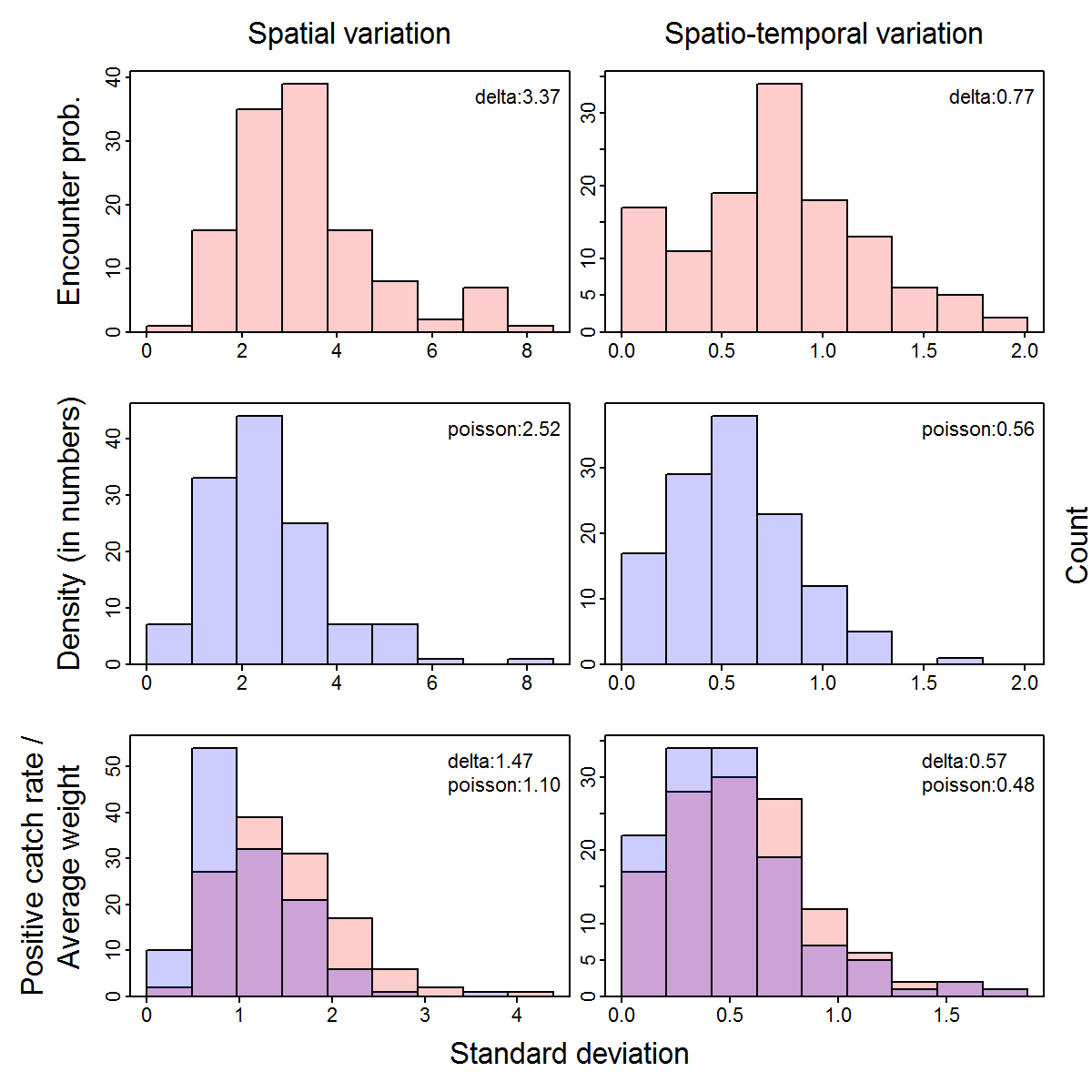


Fig. 5

