Towards improving precision in South African demersal trawl survey indices using geostatistical GLMMs

Henning Winker[1a], James Thorson[b], Tracy Fairweather[a], Rob Leslie[a], Deon Durholtz[a]

[a] *DAFF - Department for Agriculture, Forestry & Fisheries, Cape Town, South Africa*

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

Summary

This document seeks to introduce a recently commenced collaborative research project that aims to improve the precision of South African demersal trawl survey abundance indices, with the initial focus on priority hake trawl by-catch species as identified by a Demersal Scientific Working Group task team. The proposed objectives of this project are: (1) Develop a standardized demersal trawl database data extraction procedure to improve replicability of abundance estimation and facilitate regular updates and analysis speed; (2) Implement the current ACCESS-based abundance index calculation in R. (3) Explore statistical error distributions as likely more appropriate alternatives to the currently assumed normal distribution (e.g. delta-lognormal, delta-gamma); Apply the recently developed Geostatistical delta-GLMM framework (Thorson et al. 2015) to data for at least 10 selected species per region, including all prioritised by-catch species; and (5) Compare the performance of the Geostatistical delta-GLMM and current design-based estimator in terms of mean, precision and statistical properties. Preliminary results for eight species sampled during west coast demersal summer surveys suggest that geostatistical delta-GLMM has potential to increase the precision and reduce process noise compared to the currently employed design-based estimator. Further analysis and method testing is therefore warranted.

Introduction

Indices of abundance are among the most important inputs for stock assessment models and harvest control rules and are thus a key information source for effective fisheries management procedures. In South Africa, the collection of data for estimating abundance indices for ground

[1] Corresponding author
Email: henning.winker@gmail.com

fish stocks has been the main focus of regular fisheries-independent demersal trawl research surveys, conducted by DAFF since 1985. The abundance indices are derived using conventional design-based estimators (c.f. Smith 1990). The approach involves: (1) calculating the average catch rate (in weight per km^2) within several predefined depth strata; (2) raise the average catch rate to the total area of each stratum to obtain estimate of abundance for each area and (3) finally sum the abundance estimates of all strata to obtain total abundance. This procedure, also referred to a "swept-area method", has been suggested to produce fairly precise abundance estimates for several of the main target species of the demersal trawl industry, such as hakes *Merluccius sp.* and monk *Lophius vomerinus* (de Moor *et al.* 2015). However, abundance estimates for by-catch species, which may be less commonly encountered, exhibit a patchier distribution or only occur within a smaller portion of the surveyed area (e.g. only inshore) are typically considerable less precise and therefore associated with larger uncertainty. This may also have partially hampered the implementation of more formal management procedures for several priority by-catch (or joint product) species, some of which being important target species of other fisheries sectors.

Design-based approaches rely on the assumption that the species under assessment is homogeneously distributed within a given stratum (Petitgas 2001), but this assumption can be severely violated for species that exhibit a patchy distribution with distinct high and low density areas within a sampling stratum. Shelton et al. (2014) demonstrated that random allocation of sampling stations within a stratum can introduce large variation in catch rates depending on whether or not the sampling takes place in a suitable or unsuitable habitat. Recent research suggests that geostatistical models applied to survey data can produce less noisy and more precise abundance indices than conventional design-based approaches (Shelton *et al.* 2014; Thorson *et al.* 2015). In contrast to design-based methods, geostatistical models allow to explicitly accounting for spatial dependency of population densities, with densities between nearby sites assumed to be more similar (spatially correlated) than densities between sites that are farther apart. By explaining a substantial portion of the spatial variation in the survey catch rates, it is possible to increase precision of the temporal abundance trends, which may also reduce the occurrence of spurious spikes in abundance trends that are implausible from a population dynamics perspective (Thorson *et al.* 2015)

Thorson et al. (2015) developed a Geostatistical delta-Generalized Linear Mixed Model (Geostatistical delta-GLMM) specifically designed for trawl survey index standardization. When compared to design-based approaches, the Geostatistical delta-GLMM substantially increased the precision in abundance indices for 28 demersal fish species sampled during the US West Coast trawl surveys, while accompanying simulation-testing could demonstrate that the precision estimates (95% CI's) were well calibrated with no evidence for overestimated precision. The geostatistical delta-GLMM framework has also formed the foundation for recent research progress in estimating shifts in species distribution, area occupied and centre of gravity (Thorson *et al.* 2016b). The approach simultaneously estimates the effects spatial variation and fishing behaviour within a multi-species catch per unit effort standardization model (Thorson *et al.* 2016a). An analysis of trawl survey data for 120 species across six ecosystems, including the South African West Coast and South Coast survey data (Thorson *et al.* 2016c) suggests a positive abundance-area relationship in support of basin-model.

The purpose of this document is to introduce a recently commenced collaborative research project that aims to improve the precision of South African demersal trawl survey biomass indices, with the initial focus on priority hake trawl by-catch species as identified by a Demersal Scientific Working Group task team.

The objectives are:

1) Develop a standardized demersal trawl database data extraction procedure to improve replicability of abundance estimation and facilitate regular updates and analysis speed.
2) Implement the current ACCESS-based abundance index calculation in R.
3) Explore statistical error distributions as likely more appropriate alternatives to the currently assumed normal distribution (e.g. delta-lognormal, delta-gamma).
4) Apply the Geostatistical delta-GLMM to the most recently updated (2016) data (from point 1) for at least 10 selected species per region, including all prioritised by-catch species
5) Compare the performance of the Geostatistical delta-GLMM and current design-based estimator in terms of mean, precision and statistical properties.

In the following sections, we provide a brief description of the methodology and present preliminary results based on catch rates of eight selected species sampled during summer West Coast surveys (1985-2015).

Materials and Methods

Data sources

The data used in this analyses were obtained from the demersal research trawl summer surveys of the West Coast of South Africa (1986-2016), conducted by the Fisheries Branch of DAFF. The West Coast survey area is divided into the following six depth strata: (1) 0-100m, (2) 101-200m, (3) 201-300m, (4) 301-400m and (5) 401-500m. Stations at locations deeper than 500m were excluded from the analysis. As part of the data preparation for the Geostatistical delta-GLMM, we proceed by using a 2.5’ × 2.5’ grid that encompasses the entire spatial domain for the available survey data, comprising 3089 trawl stations. For this preliminary analysis, we calculated the catch rates (kg per nm^2 area swept) for the following eight species: shallow water hake (*Merluccius Capensis*, HKC), deepwater hake (*Merluccius paradoxus*, HKP), Monk (*Lophius vomerinus*, MONK), kingklip (*Genypterus capensis*, KKLP), Angelfish (*Brama brama*, APMF), cape dory (*Zeus capensis*, ZEUSCP), biscuit/thornback skate (Raja straeleni, TBSK) and Jacopever (*Helicolenus dactylopterus*, HELDAC). The latter four species have identified as priority by-catch species on the West Coast.

Design-based swept area method

The swept area method is based on the assumptions that the species under assessment is homogeneously distributed across a given sampling stratum and that all fish in the trawl path are caught, so that each trawl *i* within stratum *j* gives and an independent estimate of density $d_{i,j}$ within that stratum. If the latter assumption is violated, the abundance index should be treated as relative estimate

Species-specific catch rates were calculated as density, such that:

$$d_{i,j,y} = \frac{C_{i,j,y}}{a_{i,j,y}},$$

where $C_{i,j.y}$ is the catch of trawl i in stratum j in year y and $a_{i,j,y}$ is the corresponding swept area in nm^2, calculated as:

$$a_{i,j,y} = s_{i,j,y} \frac{t_{i,j,y}}{60} \frac{w_{i,j,y}}{1852}$$

where $s_{i,j,y}$ is the towing speed in knots (nm/h) of trawl i in stratum j and year y, $t_{i,j,y}$ the trawl duration in minutes and $w_{i,j,y}$ is the horizontal mouth width.

The mean density and standard error for each stratum in year y is then given by:

$$\bar{d}_{j,y} = \frac{1}{n_{i,j,y}} \sum_{i=1}^{n_{j,y}} d_{i,j,y}$$

and

$$SE_{j,y} = \frac{1}{\sqrt{n_{i,j,y}}} \sqrt{\frac{\sum_{i=1}^{n_{i,j,y}} (\bar{d}_{j,y} - d_{i,j,y})^2}{n_{j,y} - 1}}$$

where $n_{i,j,y}$ is the number of trawl stations i in stratum j and year y.

The total biomass (tons) in summed over all strata in year y is given by

$$B_y = \sum_{j=1}^{n_j} \frac{\bar{d}_{j,y} A_j}{1000}$$

where n_j denotes the number of strata (here 5). To approximate the associated standard error of the biomass estimate (excl. area weighting and assuming normality) for each year y was calculated as:

$$SE_y = \sqrt{\sum_{j=1}^{n_j} (SE_{j,y})^2} \ .$$

Geostatistical delta-GLMM

The geostatistical delta-GLMM (Thorson *et al.* 2015) seeks to estimate spatial variation in fish density using fishery-independent data, with aim to derive total abundance indices for stock assessment. As for any design-based approach, estimates of total abundance (here biomass) rely on the validity assumption that all fish in the trawl path are caught. If this assumption is violated, the abundance index should be treated as relative estimate. In contrast to design-based approaches, this geostatistical approach does not require that fish density is homogeneously distributed across a given sampling stratum. Instead, the geostatistical delta-GLMM estimates spatial and spatiotemporal variation in fish densities, which are approximated as Gaussian Markov random fields, assuming that variation in catch rates is spatially correlated and decreases as a function of distance.

A two-stage delta-model (or hurdle or zero-adjusted model) error model is implemented within the GLMM framework as is common practice for fisheries data that often contain many zeros and typically exhibit right-skewed distribution of positive catches (Lo *et al.* 1992; Steffanson 1996; Maunder and Punt 2004). Here, we assume a Binomial distribution to model the encounter probability and Gamma distribution for positive catch rates because of its higher flexibility compared to the log-normal. Each submodel (Binomial and Gamma GLMM) incorporates variation in density among years as a fixed effect. In addition variation among sampling vessels and additional predictor variables (e.g. depth) can be incorporated. However, a random vessel effect is only estimable if there is sufficient overlap between vessels within surveys, which is therefore not directly applicable to South African demersal trawl survey data.

The standard output produced by the geostatistical delta-GLMM, includes several model validation graphs and csv file comprising the assessment year, annual abundance estimates and the standard errors of the log of the expected values, which is designed facilitate routine input for the ADMB based stock assessment software Stock Synthesis 3 (Methot and Wetzel 2013). In addition, three types of maps are produced to visualize the spatial and temporal-spatial dynamics in estimated densities, encounter probabilities and positive catches. Density maps for the period 1986-2016 for kingklip are illustrated as an example in Fig. 1.

Model parameters are estimated using maximum marginal likelihood techniques using the Laplace approximation as implemented in Template Model Builder (Kristensen *et al.* 2014) called from within the R statistical platform. For specific details on geostatistical delta-GLMM refer to the work by Thorson *et al.* (2015) and references therein. Code for adapting this geostatistical delta-GLMM, including a worked example for jacopever based on the South African West Coast survey catch rates (summer; old gear only), is publicly available on GitHub (https://github.com/nwfsc-assess/geostatistical_delta-GLMM).

Comparing Precision

Direct comparison of the estimated standard errors (SEs) of annual abundance was possible as the currently employed design-based estimator approach assumes normality, whereas geostatistical delta-GLMM produces standard errors of the log of the annual abundance estimate. To first facilitate comparisons SEs were therefore transformed into observation error CVs.

The CVs for the designed-based estimator were calculated as:

$$CV_y^{DB} = \frac{SE_y^{DB}}{B_y^{DB}}$$

and the CVs for the geostatistical delta-GLMM were calculated as:

$$CV_y^{GEO} = \sqrt{\exp(\mathrm{log.se}_{\mathrm{y}}^2) - 1},$$

where *y* is the year, the superscripts *DB* and *GEO* denote design-based and geostatistical delta-GLMM, respectively, and log.se is the standard error of the log of the annual abundance as estimated by geostatistical delta-GLMM

In addition, we seek to obtain a quantity that allows comparing the process error ‘noise’ associated with each method and considered the loess smoother method described in Francis (Francis 2011) for this purpose. This method involves fitting a log-transformed abundance index using loess smoothers, and calculating the process error CV from the residuals of the fit of the smoother to the data.

Preliminary Results and Discussion

On broad comparisons of the estimated abundance indices, we found overall agreement between mean biomass values from design-based estimator and the geostatistical delta-GLMM across all eight species (see hakes' examples Figs. 2-3). However, in most cases the 95% interval of the design-based estimator is considerable wider than that for the geostatistical delta-GLMM index, which is in agreement with the consistently lower observation error CV means calculated for geostatistical index (Table 1, Fig. 4). Similarly, the process error CV (Francis 2011), which we used as an ad-hoc indicator of 'process noise', was consistently lower for geostatistical index (Table 1; Figs 5-6). However, this reduction in process noise appeared to be most pronounced for the two bycatch species cape dory and angelfish and somewhat surprisingly for shallow-water hake, being the most commonly encountered species in west coast summer surveys.

Several indices indicate the presence of a vessel and/or gear effect on the catchability. Seemingly strong signals of a vessel effect were most notably for deep-water hake (Fig. 2), monk (Fig. 3), cape dory (Fig. 7) and biscuit skate (Fig. 8), as indicated abrupt decline between the *MS Africana* and the *Andromeda* research survey biomass estimates. A potential gear-effect between the old and new trawl set up was generally less apparent upon visual inspection, with the exception of biscuit skate (Fig. 8) for which the new gear is evidently associated with a lower catchability then the old gear.

Our preliminary results for the eight species sampled during west coast demersal summer surveys suggest that geostatistical delta-GLMM has potential to increase the precision and reduce process noise compared to the currently employed design-based estimator, which may be substantial for some but less so for other species. Further analysis and method testing is therefore warranted. An important aspect will be to carefully validate and standardize the data sourcing process, with the aim to improve efficacy and reproducibility of abundance estimates as inputs for stock assessments and management procedures. The likely presence of species-specific gear- and vessel- effects of varying strength have be taken into consideration when the abundance index is used as input for stock assessments and ways of doing so should remain a high priority of future research.

Table 1. Species-specific means (sd) of estimated annual observation error CVs and approximated process error CVs using the loess smoother methods (Francis, 2011) from the current design-based estimator and the proposed geostatistical delta-GLMM.

Species	Code	Observation error CV		Process error CV	
		Design-based	Geo-GLMM	Design-based	Geo-GLMM
Merluccius paradoxus	HKP	0.179 (0.053)	0.148 (0.014)	0.281	0.260
Merluccius Capensis	HKC	0.184 (0.061)	0.122 (0.010)	0.533	0.457
Lophius vomerinus	MONK	0.126 (0.021)	0.106 (0.008)	0.188	0.185
Genypterus capensis	KKLP	0.219 (0.061)	0.171 (0.022)	0.327	0.305
Brama brama	APMF	0.326 (0.111)	0.315 (0.142)	0.579	0.467
Zeus capensis	ZEUSCP	0.387 (0.122)	0.274 (0.045)	0.491	0.354
Raja straeleni	TBSK	0.179 (0.035)	0.144 (0.009)	0.315	0.304
Jacopever	HELDAC	0.167 (0.045)	0.123 (0.009)	0.209	0.195

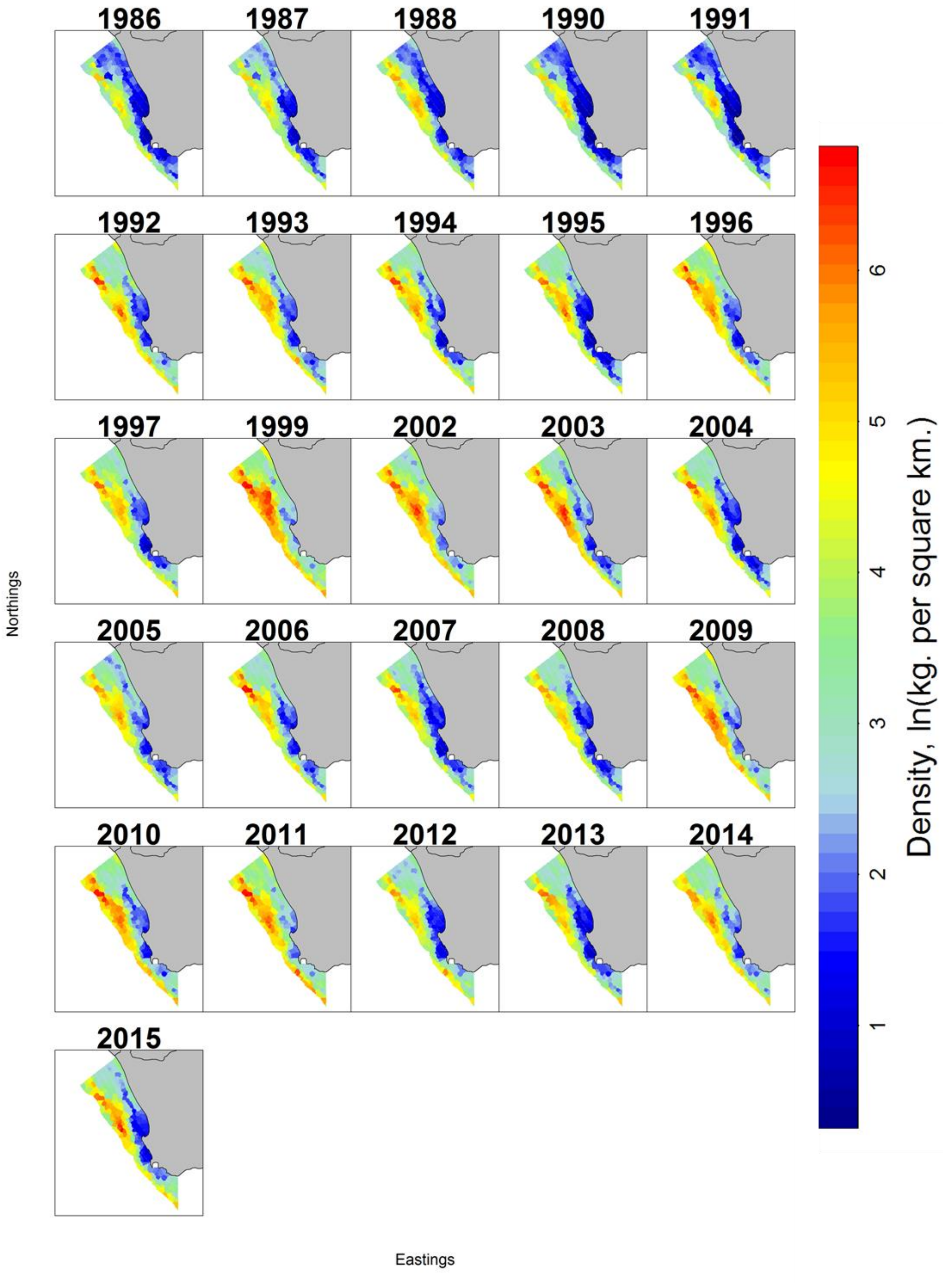

Fig. 1. Spatio-temporal density estimates for kingklip 1986–2015, estimated by the geostatistical delta-generalized linear mixed model

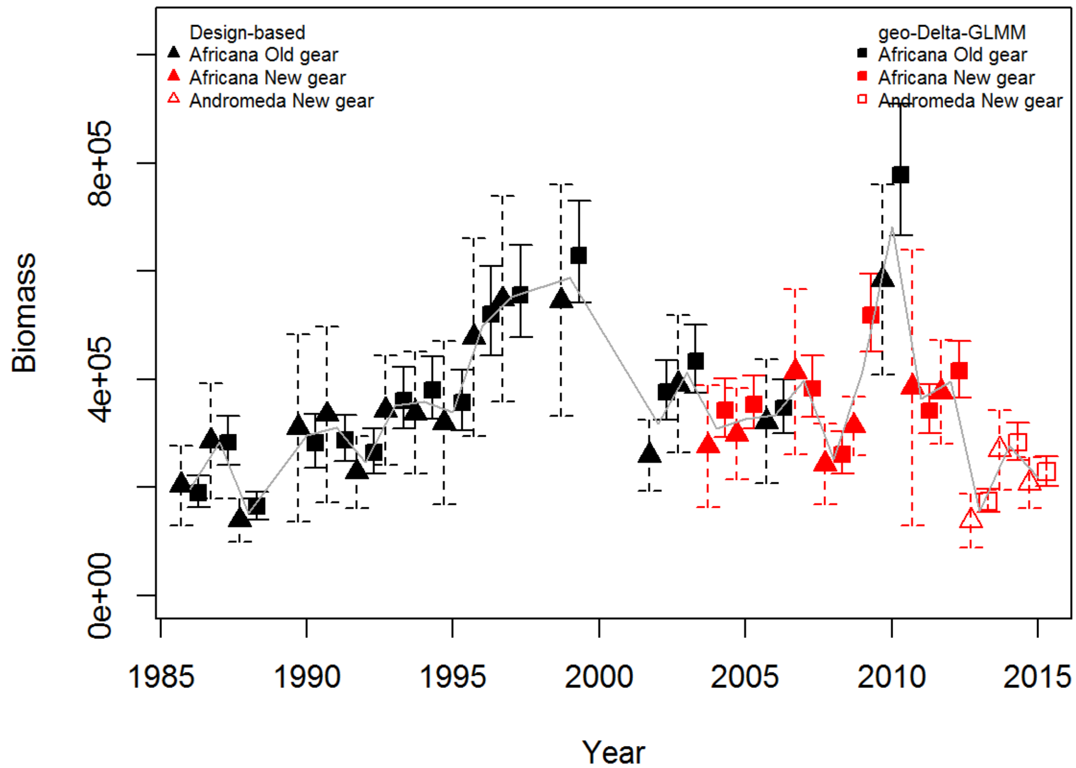

Fig. 2. Estimated indices of abundance for deep-water hake *Merluccius paradoxus* (Biomass in tons) using the design-based estimator and geostatistical delta-GLMM, with error bars denoting the approximated with 95% intervals, the grey line indicating the average abundance estimate between the methods and gear and vessel changes over time being highlighted by closed/open symbols and by a black/red color scheme, respectively.

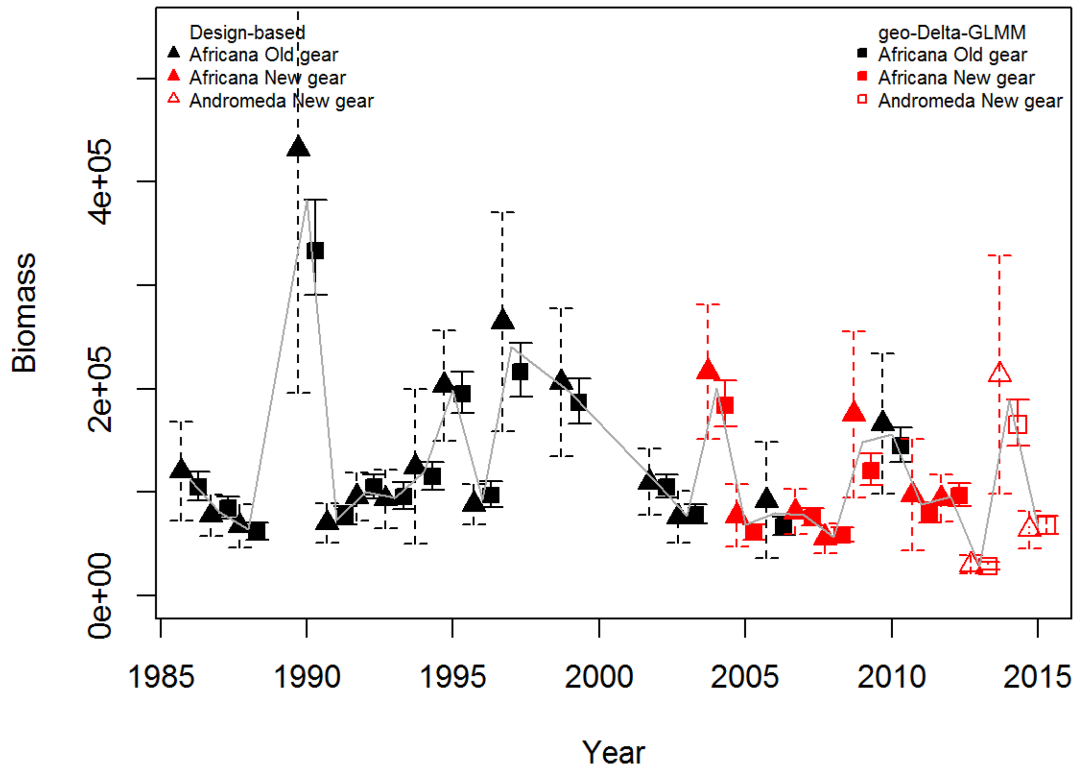

Fig. 3. Estimated indices of abundance for shallow-water hake *Merluccius capensis* (Biomass in tons) using the design-based estimator and geostatistical delta-GLMM, with error bars denoting the approximated with 95% intervals, the grey line indicating the average abundance estimate between the methods and gear and vessel changes over time being highlighted by closed/open symbols and by a black/red color scheme, respectively.

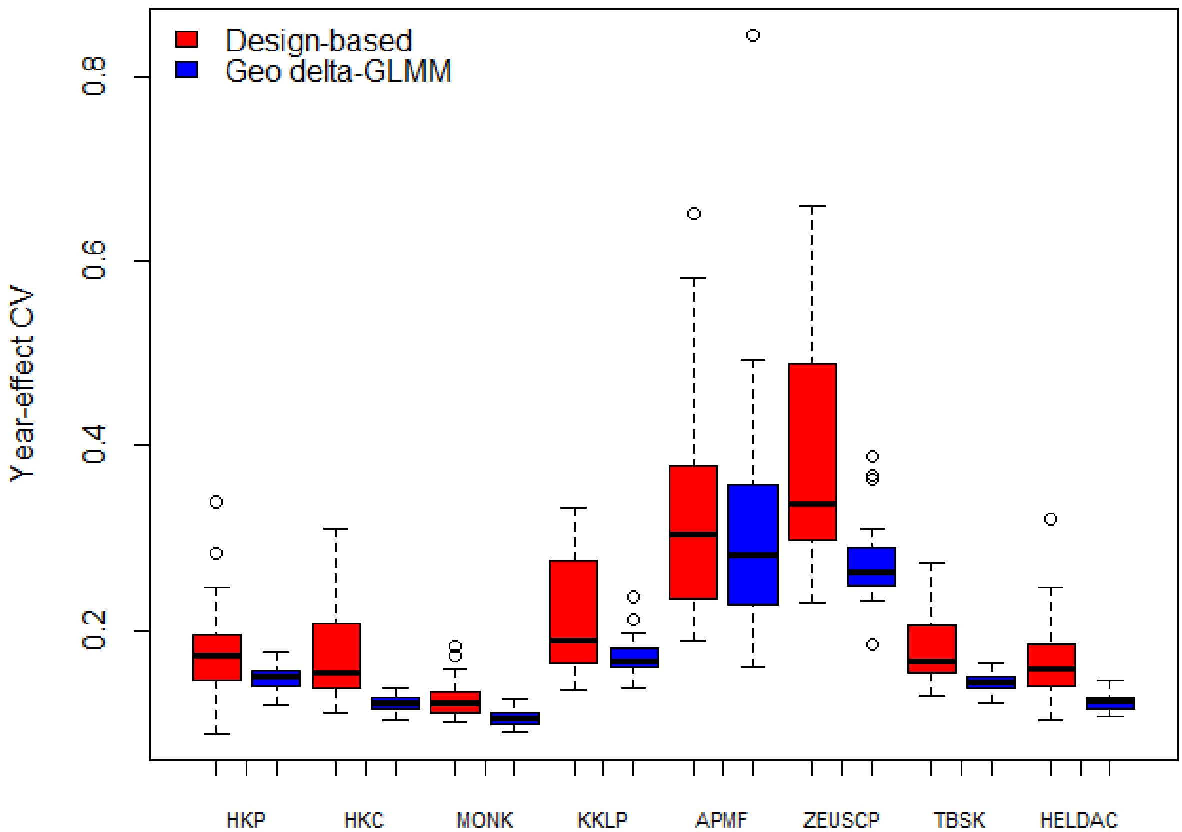

Fig. 4. Barplot showing the observation error CVs for annual abundance estimates by species based on the current design-based estimator and the proposed geostatistical delta-GLMM. HKP: *Merluccius paradoxus*, HKC: *Merluccius Capensis*, MONK *Lophius vomerinus*, KKLP: *Genypterus capensis,* APMF: *Brama brama*, ZEUSCP: *Zeus capensis*, HELDAC: *Helicolenus dactylopterus*

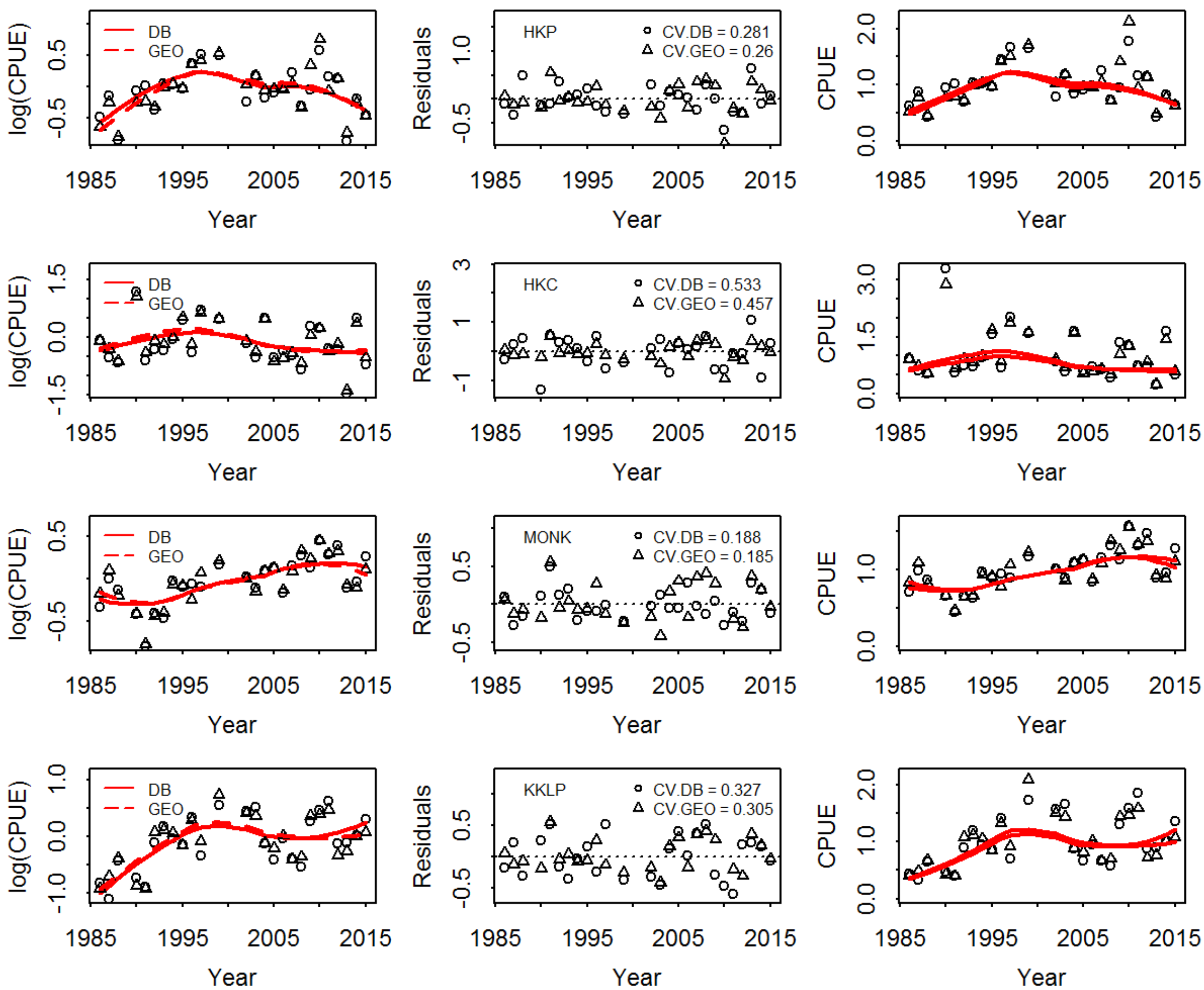

Fig. 5. Loess smoother fits to estimate process error CVs (c.f. Francis 2011) for abundance indices obtained from on the current design-based estimator and the proposed geostatistical delta-GLMM. Left panel: Smoother fits to log(CPUE) data; Middle panel: Residual plots and estimated CVs for each times series. Right panel: Loess smoother fits illustrated for CPUE indices. HKP: *Merluccius paradoxus*, HKC: *Merluccius Capensis*, MONK *Lophius vomerinus*, KKLP: *Genypterus capensis*

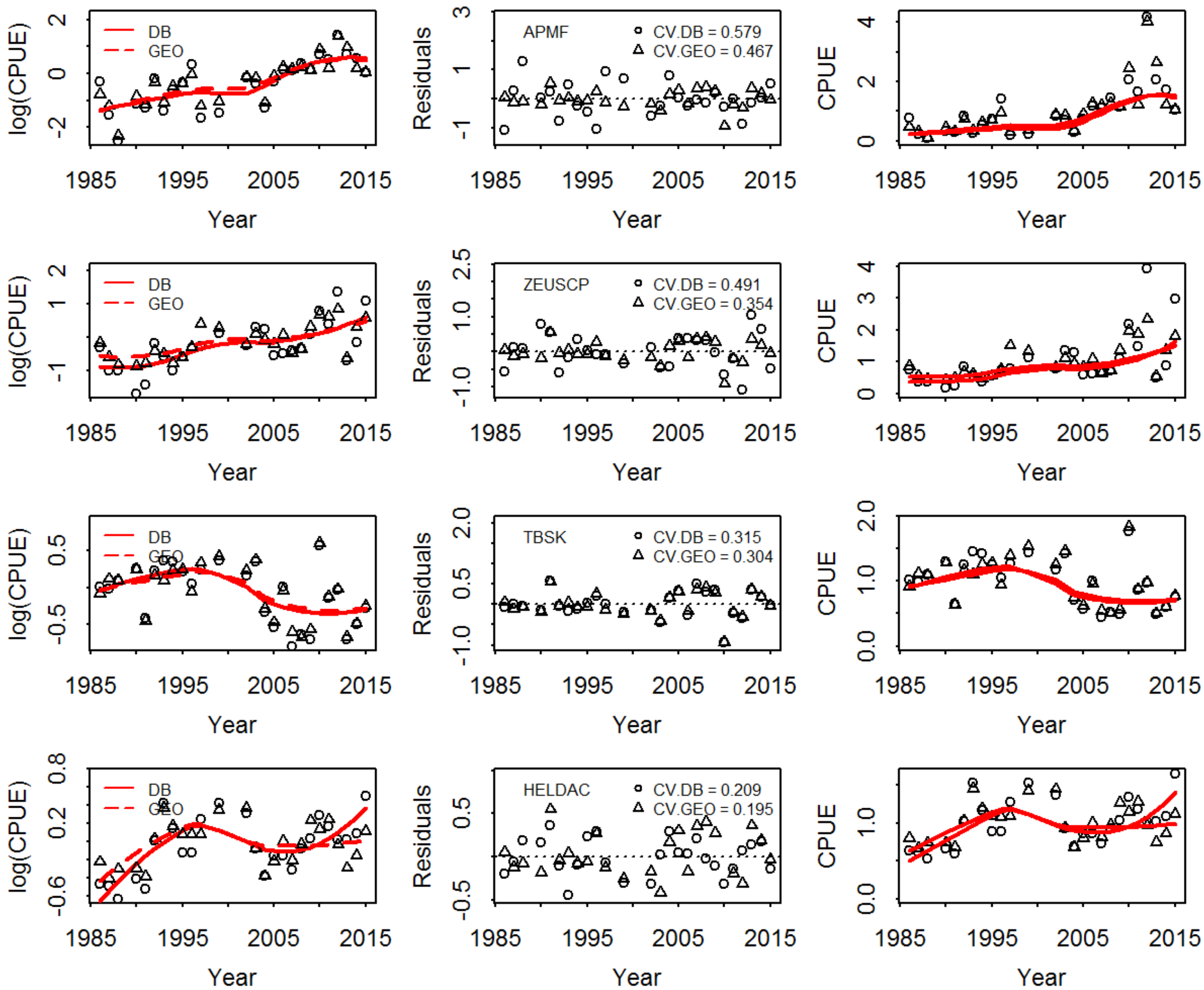

Fig. 6. Loess smoother fits to estimate process error CVs (c.f. Francis 2011) for abundance indices obtained from on the current design-based estimator and the proposed geostatistical delta-GLMM. Left panel: Smoother fits to log(CPUE) data; Middle panel: Residual plots and estimated CVs for each times series. Right panel: Loess smoother fits illustrated for CPUE indices. APMF: *Brama brama*, ZEUSCP: *Zeus capensis*, HELDAC: *Helicolenus dactylopterus*

Bibliography

Francis, R.I.C.C. (2011) Data weighting in statistical fisheries stock assessment models. *Canadian Journal of Fisheries and Aquatic Sciences* **68**, 1124–1138.

Kristensen, K., Thygesen, U.H., Andersen, K.H., Beyer, J.E. and Jech, J.M. (2014) Estimating spatio-temporal dynamics of size-structured populations. *Canadian Journal of Fisheries and Aquatic Sciences* **71**, 326–336.

Lo, N.C., Jacobson, L.D. and Squire, J.L. (1992) Indices of relative abundance for fish spotter data based on delta-lognormal models. *Canadian Journal of Fisheries and Aquatic Sciences* **49**, 2515–2526.

Maunder, M.N. and Punt, A.E. (2004) Standardizing catch and effort data: a review of recent approaches. *Fisheries Research* **70**, 171–195.

Methot, R.D. and Wetzel, C.R. (2013) Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management. *Fisheries Research* **142**, 86–99.

de Moor, C., Johnston, S., Brandão, A., Rademeyer, R., Glazer, J., Furman, L. and Butterworth, D. (2015) A review of the assessments of the major fisheries resources in South Africa. *African Journal of Marine Science* **37**, 285–311.

Petitgas, P. (2001) Geostatistics in fisheries survey design and stock assessment: Models, variances and applications. *Fish and Fisheries* **2**, 231–249.

Shelton, A.O., Thorson, J.T., Ward, E.J. and Feist, B.E. (2014) Spatial semiparametric models improve estimates of species abundance and distribution. *Canadian Journal of Fisheries and Aquatic Sciences* **71**, 1655–1666.

Smith, S.J. (1990) Use of statistical models for the estimation of abundance from groundfish trawl surveys. *Canadian Journal of Fisheries and Aquatic Sciences* **47**, 894–903.

Steffanson, 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., Fonner, R., Haltuch, M.A., Ono, K. and Winker, H. (2016a) Accounting for spatio-temporal variation and fisher targeting when estimating abundance from multispecies fishery data. *Canadian Journal of Fisheries & Aquatic Sciences* **14**, 1–38.

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 in Ecology and Evolution* **7**, 990–1002.

Thorson, J.T., Rindorf, A., Gao, J., Hanselman, D. and Winker, H. (2016c) Evaluating density-dependent changes in effective area occupied for bottom-associated marine fishes. *Philosophical Transactions of the Royal Society B: Biological Sciences* **283**, 20161853.

Thorson, J.T., Shelton, A.O., Ward, E.J. and 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.