A hierarchical Bayesian approach to estimate growth parameters from length data of narrow spread

Shijie Zhou, Sarah Martin, Dan Fu

Introduction

Neritic tuna species have substantial commercial and food security value to Indian Ocean coastal states, as well as the Distant Water Fishing Nations operating in the IOTC area of competence. However, compared with more economically valuable species, biology and population dynamics are less studied for neritic tunas. Data limitation has promoted adoption of data-poor stock assessment approaches for these species in the recent years (IOTC Secretariat, 2015a, 2016a).The goal of IOTC is to collect more data and move to full stock assessment in the near future. Both data-poor and full assessment require basic life history parameters, including growth information.

Estimates of growth for neritic tuna species in the Indian Ocean are highly variable, based on a number of independent studies that have taken place in particular regions for particular time periods (IOTC Secretariat, 2015b, 2015c, 2016b, 2016c). This may be due to the presence of subpopulations of the stocks with different growth rates and maximum sizes or may be due to differences based on the sampling methods such as the size selectivity of different gear types. The large differences may have also resulted from different analytical methods. The majority of studies have used ELEFAN (IOTC Secretariat, 2015b, 2015c, 2016b, 2016c) which was developed for closed populations (Pauly, 1987), where modal progression can provide better estimates of growth, and so may not provide good estimates of growth for migratory species. There is likely to be migration of fish across the entire area and so isolated studies using these techniques may not be appropriate for coastal tuna populations. Therefore, the IOTC Working Party on Neritic Tuna requested a meta-analysis take place which collates the local area studies to get an overview of parameters related to growth in the Indian Ocean.

Classical meta-analyse involves analysing the results from many studies. An important limitation of this type of meta-analysis is that its summary results can only be as good as the data from the original studies, and so is dependent on the validity of these studies. As discussed above, there is a concern regarding the methodologies and assumptions in the original studies. Consequently, a completely new approach is used here to tackle the issue by analysing all of the available data that has been provided to the IOTC under a hierarchical Bayesian framework. Although this requires much more effort and time, such a consistent meta-analysis should yield more reliable and credible results than summarizing the existing studies that use limited data in isolation.

The IOTC maintains a length-frequency database for six neritic tuna species. The database covers samples taken between 1983 and 2014 from 10 countries (fleets) using a range of fishing gears. The sample size at each {species, fleet, year, month, grid, gear} stratum ranged from 1 to over 56,000 fish. This large database is valuable for deriving growth parameters. However, it also presents several major challenges:

1. Samples were taken sporadically over time and space and the sample size varied enormously between strata;

2. The range of fish size is narrow in each stratum, resulting in few identifiable age classes;

3. There is a lack of knowledge about the number of age classes;

4. There is a lack of clear length-modal progression within a region, perhaps due to fish migration;

5. There is limited information from other methods such as tagging and aging studies.

To tackle these difficulties, a Bayesian meta-analysis approach is developed to simultaneously model multiple datasets from a wide range of space over a long time period. It is assumed the as an identical species, fish at different strata share some similar life history traits but may vary subtly across spatial and temporal dimensions. Two growth models are examined: the classic von Bertalanffy model and the re-parameterized Schnute-Fournier model (Schnute and Fournier, 1980). The analysis is carried out in R language (R Core Team, 2017) and WinBUGS/OpenBUGS/JAGS (Lunn *et al.*, 2000; Plummer, 2003).

Material methods

Data

The IOTC maintains a length-frequency database for six neritic tuna species. The database includes fleet (country), year, month, grid (area fished), gear, total number of fish measured, and length-frequency at 1-cm interval. It covers samples taken between 1983 and 2014 from 10 regions (fleets) using various (total of 15) gear-types (Tables 1-4). The sample size at each {species, fleet, year, month, grid, gear} stratum ranged from 1 to over 56,000 fish. For a single species, it is likely that fish within different strata share some similar life history traits but vary subtly across wide spatial and temporal dimensions. To model their growth at varying scales, we define fish captured in each stratum as a unique population for the purpose of some parameters of the growth model (e.g., age and size).

Growth model

Several alternative growth models have been used for tuna studies. However, the von Bertalanffy (vB) growth model is most widely adopted for tunas (Kolody *et al.*, 2016) and is used in this study,

$$L_t = L_\infty\left[1 - \exp(-K(t - t_0))\right] \qquad (1)$$

where L_∞ is the infinite length, L_t is the length at age t, K is the vB growth parameter, and t_0 is estimated theoretical age when size is zero. This model requires aging information, which is often difficult and costly to obtain. Alternatively, growth is estimated using length frequency data, where the first length mode is typically assumed to be one year-old fish. Schnute and Fournier (1980) re-parameterize this model as

$$\mu_i = L_{\min} + (L_{\max} - L_{\min})\frac{1-k^{i-1}}{1-k^{A-1}}; \text{ i} = 1, \ldots, A. \qquad (2)$$

This equation (referred to as Schnute and Fournier or SF model here) considers the length data at hand, where $L_{\min}$ and $L_{\max}$ are the mean length of the first and last age-classes in the data, k = exp(-K) Ford growth parameter, A is the number of age-classes in the length frequency dataset, and μ_i is the mean length of fish at ith age-class. Note that the first age-class is not necessary one year-old.

Modelling approaches 1: a 2-step method

The model

The classic methods of estimating growth from size frequency data involve two independent steps (Gulland and Rosenberg, 1992; Pauly, 1987). First, the mode of each annual cohort is identified from length or weight frequency distribution. Second, growth parameters are estimated from the cohort modes.

The challenge in modelling neritic tuna length data is that very few modes (typically 1 to 3, see below) can be identified from each dataset (population). Estimating three parameters, either K, L_∞, t_0 in vB model or k, $L_{\min}$ and $L_{\max}$ in SF model from essentially 1 to 3 data points is extremely difficult. We overcome this challenge through a meta-analysis that uses multiple datasets and a hierarchical model structure. We assume that for species with a wide geographic distribution, biological characteristics, including growth, may vary somewhat from region to region. However, while there may be some regional variation in biological characteristics, this will be limited for any one given species which will also share a number of attributes so it is assumed that all fish share a common underlying distribution across regions within the whole ocean.

Identifying modes

Growth models based on length frequency data typically require knowledge of the number of age groups in the length data (Fournier *et al.*, 1990; Schnute and Fournier, 1980). To determine the number of age-classes we assume: (1) length frequency of each age-class follows a normal distribution; (2) length increment from one age to the next is greater than a minimum amount defined by the vB model. We apply function normalmixEM in the R package mixtools (R Core Team, 2017) to the length frequency data. Because including more age-classes tends to improve model fitting, we accept a model with one more age-class only when it is significantly better than the model with one age-class less. This Step-1 provides the best number of age-classes in each population which is used as part of the 2-step method as well as in the following integrated method. It also produces mean lengths and variance at modes to be used in Step-2 but does not infer the actual age or size increment.

Bayesian hierarchical modelling

Natural size variability between individuals of the same age-class is an important feature and generally age-specific, i.e., variance is smaller in younger year-classes than in older year-classes. To capture this variability and transfer it from step-1 to step-2 we introduce a latent length:

$$lL_{f,y,m,g,t} \sim norm(L^M_{f,y,m,g,t}, \sigma^M_{f,y,m,g,t})$$

where $L^{M}_{f,y,m,g,t}$ is the mean length of an age-class and subscript *f* is fleet (or region), *y* is year, *m* is month, *g* is grid, and *t* is age-class (mode in the length frequency data). $\sigma^{M}_{f,y,m,g,t}$ is the standard deviation associated with the mean length.

There is an additional difficulty that the actual age associated with each mode is unknown. For example, fish belong to the first mode may not be exactly age 1 (e.g., 0.7 or 1.3 years old) or can even be much older than age 1 (e.g., 2.3 or 3.4 year-old). We overcome these difficulties by using a modified vB model (1)

$$L^{M}_{f,y,m,g,t} = L_{\infty,f}\left[1-\exp(-K_f\,(tt_{f,y,m,g,t}-t_{0,f}))\right],\ t = 1, 2, \ldots \text{nage} \qquad (3)$$

Because the first mode identified in each dataset is not necessary one year-olds and the age of the modes may differ between populations, the model attempts to estimate a “true age” at finer scale than an integer age:

$$tt_{f,y,m,g,t} = a_{f,y,m,g} + t - 1 + \delta t.$$

Here an additional parameter $a_{f,y,m,g}$ is the age of fish belonging to the first mode in that unique population and δt is a small random number that allows each age to diverge from exactly integer age. This model assumes that fish capture by the same fleet (country) share a same growth model described by three parameters $L_{\infty,f}$, K_f, $t_{0,f}$. For each population the length-at-age may be located at varying points along the vB curve. Hence, multiple populations, even with only 2 or 3 length-age data points each, facilitate the estimation of common parameters.

Model (3) has a hierarchical structure where the first age is estimated at population (stratum) level, the growth parameters are estimated at fleet level, and the underlying hyper-parameters are at species level in the whole Indian Ocean (Figure 1).

In the second step, the mean length-at-age-class from known fleet, year, month, and grid is modelled as normal distribution along the mean of the vB model:

$$L^{M}_{f,y,m,g,t} \sim norm\left(L^{vB}_{f,y,m,g,t}, \sigma_{f,y,m,g,t}\right).$$

In this model, $L^{m}_{f,y,m,g,t}$ is the maximum likelihood estimate of length mode from Step-1 and

$L^{vB}_{f,y,m,g,t}$ is the vB predicted mean length. To ensure the standard deviation is positive and realistic, we assume

$$\tau_{f,y,m,g,t} \sim gamma(0.01, 0.01)$$
$$\sigma_{f,y,m,g,t} = 1/\sqrt{\tau_{f,y,m,g,t}}$$

The first age parameter (actual age of the first age-class) is assumed to follow a normal distribution:

$$a_{f,y,m,g} \sim norm(1,1).$$

We are particularly interested in the growth parameters at regional level and assume the following distributions:

$$K_f \sim \log norm(\mu_K, \sigma_K)$$

$$L_{\infty,f} \sim norm(\mu_{L\infty}, \sigma_{L\infty})$$

$$t_{0,f} \sim norm(\mu_{t0}, \sigma_{t0})$$

Weak hyper-priors are used for K_f and $L_{\infty,f}$:

$$\mu_K \sim norm(-1, 0.01)$$

$$\tau_K \sim gamma\,(0.01, 0.01) \text{ and } \sigma_K = 1/\sqrt{\tau_K}$$

$$\mu_{L\infty} \sim norm(100, 0.000001)$$

$$\tau_{L\infty} \sim gamma\,(0.01, 0.01) \text{ and } \sigma_{L\infty} = 1/\sqrt{\tau_{L\infty}}$$

Modelling approaches 2: Integrated method

The model

The 2-step classic approach assumes that uncertainty in identifying age and uncertainty in fitting the growth model are independent, even though all information is based on frequency distribution. When the size frequency data lack clearly distinguishable modes, it can difficult to separate these two sources of uncertainties and the estimated growth parameters may be unreliable. We develop a Bayesian hierarchical approach to analyse length data in one integrated step. Unlike the 2-step method and other approaches that model length-frequency data at bin level, this integrated method models length of each individual fish *i* instead of mode:

$$L_{f,y,m,g,i} \sim norm(L^{vB}_{f,y,m,g,t}, \sigma_{f,t}).$$

This formulation attempts to model the probability of each fish *i* to age-class *t* where the mean length-at-age and variance-at-age are population specific and is restrained by the vB model (3). In other words, the method does not try to identify modes first and then fit the growth model to the modes. Instead, this approach integrates vB model into the process of estimating the probability of each fish belonging to one of the several possible age classes.

A key question is to determine the age-class that each fish belongs to. We use Dirichlet distribution and categorical distribution to assign the probability that a fish is in the age-class *t*:

$$P_{f,y,m,g,t} \sim dirichlet(\alpha_{1:nage})$$

$$t_{f,y,m,g,i} \sim cat(P_{f,y,m,g,t})$$

Preliminary analysis indicated that parameter estimates are unreliable with too few samples while Bayesian MCMC process takes too much computing power (time and computer memory) with large sample size. Therefore, we excluded populations with sample size < 30 and reduced sample size to 300 fish for strata of large measurements by $N_i \times 300/\sum N_i$

where N_i is the number of fish in length interval i. Other parameters use the same priors as in the 2-step method above.

Validating the integrated method by simulation

Since the integrated method is a new approach, a simulation exercise was carried out to examine its reliability. As the major challenge in the neritic tuna length data is the limited number of age-classes captured in each population, the simulation focuses on this aspect.

Known mean length-at-age data L^m are generated using equation (1). The following values are used: K=0.4, L_∞ = 100, t_0 =-0.2, and age from 0 to 8 by an increment of 0.2 year. For each of the L_t^M 100 random lengths are produced by $norm(L_t^M, \sigma)$ where σ = 3 is used. These 41*100 fish are distributed and harvested in 10 strata, each with only two age-classes with varying ages. This assumption mimics fish migration from one area to another, resulting in age variation of several months in different regions. Only fish from age 1 to age 4 are "captured".

Two model structures were explored: (1) assuming that the ten populations are independent of each other so all the growth parameters (K, L_∞, t_0, a and their variances) are estimated for each of the populations. (2) assuming that the ten populations have an identical underlying growth pattern (i.e., one set of K, L_∞, t_0 and their variances) but the age of the first age-class differs among the populations. Same priors as described above are adopted for the simulated data.

Model implementation

The two forms of hierarchical Bayesian growth models are implemented in OpenBUGS and JAGS (also WinBUGS). Preparation of the model input data and initial values is quite complicated because the data are unbalanced and have multiple levels and varying number of age-class. After properly assembling the input and initials, we set up three MCMC chains and keep samples at every 20 iterations. Burn-in periods vary between models and species. Chain convergence is verified by visual inspection of the MCMC history as well as Gelman-Robin diagnostic. Generally, this requires the first several hundred thousand samples to be discarded. After convergence the model is run for additional 10,000 iterations to be saved for parameters inference. The 2-step method involves fewer data because it uses fish at stratum-age level so the speed is fast. However, the integrated method has many thousands of fish so the MCMC process is very slow, typically taking many days.

Results

Simulation test of the integrated method

We first tested whether the model structure that attempts to estimate growth parameters for each population is feasible. We found that MCMC chains were unable to converge after 50,000 iterations (due to few data points in each stratum but many parameters). From these un-converged posterior distribution, mean K of the vB model varies between 0.17 and 1.88, and the mean L_∞ varies between 73 and 184. The Schnute-Fourier model performs even more poorly, with all K > 1.1 and L_∞ > 189.

For the second model structure that assumes a common growth pattern, the SF model still produces unrealistic results (too large K and L_∞). However, the vB model yields a nice fit of length frequency (Figure 2) for most of the populations. The match between observed length frequency and the posterior length frequency indicate that accuracy of three parameters: mean length at age $L^M_{\bullet,t}$, proportion of fish at age $P_{\bullet,t}$, and standard error of length at age $\sigma_{\bullet,t}$. The posterior mean length-at-age and the mean age are comparable with the true values (Figure 3). The posterior mean $K = 0.406$ (sd = 0.014) and $L_\infty = 99.60$ (sd = 1.37) are very close to the true values of 0.4 and 100. However, the posterior mean $t_0 = 0.004$ is larger than the true value of -0.2. There is a high correlation between K and L_∞ ($r = -0.95$).

Spanish Mackerel

Identifying age-class by length modes

The length frequency database contains a total of 657 populations (i.e., {fleet, year, month, and grid} strata) for Spanish mackerel. We assume that there are possibly 1 to 6 age-classes in each population. One normal distribution model and five multi-normal mixture models are fitted to each of the populations. In many populations, fitting multi-normal mixture models do not significantly reduce AIC from the normal distribution model, indication that there is likely only one age-class. For some populations, multi-normal mixture models can significantly reduce AIC but the length increase is too small to be biologically realistic. Consequentially, the screen criteria result in a total of 152 populations (Table 5) where two or more age-classes are identified (Figure 4 and Appendix 2).

2-step method

The multi-normal mixture models identify that the most likely number of age-class ranges from 2 to 4 in the 152 populations. Since the location of modes varies between populations even when the number of age-class is the same, these 152 populations comprise many age-classes at finer time step than 1 year. Meta-analysis of these populations together enables estimating the true age. Note the fine scale of the posterior age on the x-axis of Figures 5 and 6. The length modes from Step 1 scatter around the vB curve. The Bayesian model (Step 2) tends to overestimate the length for smaller fish but underestimate it for larger fish (Figure 7).

The posterior mean K varies from 0.25 for the LKA region to 0.39 for the OMN region (Table 6). The ocean-wide hyper-parameter *muK* is 0.32 (95%CI between 0.18 and 0.48). L_∞ is similar across regions. One possible reason for this less variability is the shrinkage effect of hierarchical Bayesian model. The ocean-wide hyper-parameter muL_∞ is 157.0 cm (95%CI between 136.5 and 176.6). Compared to the hyper-priors, the posteriors exhibit a fairly narrow distribution (Figure 8).

Integrated method

The major drawback of the Bayesian approach using Markov chain Monte Carlo technique is the slow computing speed. To reduce lengthy sampling time, we proportionally scale large sample size to a maximum of 300 in each population. This reduction process results in about 40,000 length measurements for Spanish mackerel.

Waiting for the results…

Discussion

This draft paper focuses on the development of a meta-analysis method for analysing length measurement data that have very few identifiable age-classes. We fails to achieve MCMC chain convergence when applying a growth model to each strata separately. However, reliable results are obtained by Bayesian hierarchical models applied to multiple strata simultaneously. There are two possible approach under the Bayesian framework: one uses length-frequency modes as data and the other models individual fish length measurement. We discuss the performance of the approaches and some of the features.

Comparing with other studies

A range of studies have been carried out in separate locations around the Indian Ocean. We compare our results with 32 studies that report growth parameters (IOTC Secretariat, 2015c). Using various methods and data from various area and time, the estimated L_∞ ranges from 110 cm to 230 cm with a mean of 150.9 cm. The estimated K ranges from 0.12 y^{-1} to 0.78^{-1} with a mean of 0.35 (Figure 9). Interestingly, our results are very close to these values.

Assumption about the same growth model at region level

References

Fournier, D. A., Sibert, J. R., Majkowski, J., and Hampton, J. 1990. MULTIFAN a likelihood-based method for estimating growth parameters and age composition from multiple length frequency data sets illustrated using data for Southern Bluefin Tuna (Thunnus maccoyii). Canadian Journal of Fisheries and Aquatic Sciences, 47: 301–317.

Gulland, J. A., and Rosenberg, A. A. 1992. A review of length-based approaches to assessing fish stocks. FAO Fisheries Technical Paper No. 323. Rome, FAO. 100 pp.

IOTC Secretariat. 2015a. Report of the 5th Session of the IOTC Working Party on Neritic Tunas. Zanzibar, Tanzania, 26-29 May 2015. 105 pp.

IOTC Secretariat. 2015b. Population parameters: Kawakawa (Euthynnus affinis). IOTC–2015–WPNT05–DATA12. 28 Appril 2015. 7 pp.

IOTC Secretariat. 2015c. Population parameters: Narrow-barred Spanish mackerel (Scomberomorus commerson). IOTC–2015–WPNT05–DATA14 Rev_1. 28 April 2015. 9 pp.

IOTC Secretariat. 2016a. Report of the 6th Session of the IOTC Working Party on Neritic Tunas. Mahe, Seychelles. Indian Ocean Tuna Commission, 21-24 June 2016. 89 pp.

IOTC Secretariat. 2016b. Population parameters: Bullet tuna (Auxis rochei). IOTC–2016–WPNT06–DATA14. 30 May 2016. 6 pp.

IOTC Secretariat. 2016c. Population parameters: Indo-Pacific King mackerel (Scomberomorus guttatus). IOTC–2016–WPNT06–DATA12. 30 May 2016. 9 pp.

Kolody, D. S., Eveson, J. P., and Hillary, R. M. 2016. Modelling growth in tuna RFMO stock assessments : Current approaches and challenges. Fisheries Research, 180: 177–193. Elsevier B.V. http://dx.doi.org/10.1016/j.fishres.2015.06.016.

Lunn, D. J., Thomas, A., Best, N., and Spiegelhalter, D. 2000. A Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing, 10: 325–337.

Pauly, D. 1987. A review of the ELEFAN system for analysis of length-frequency data in fish and aquatic invertebrates. *In* Length-based Methods in Fishereis Research. ICLARM Conference Proceedings 13, 468p., pp. 7–34. Ed. by D. Pauly and G. R. Morgan. International Center for Living Aquatic Resources Management, Manila, Philippines.

Plummer, M. 2003. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *In* Proceedings of the 3rd International Workshop on Distributed Statistical Computing.

R Core Team. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. URL.

Schnute, J., and Fournier, D. 1980. A new approach to length-frequency analysis: growth structure. Canadian Journal of Fisheries and Aquatic Sciences, 37: 1337–1351.

Table 1. Length measurements for the six neritic tuna species from 1983 to 2014.

Year	BLT	COM	GUT	KAW	LOT	FRI	Sum
1983				5,761	5,151	692	-
1984	30			606		503	30
1985	265			4,173	1,310	3,057	265
1986	90			9,288	564	5,227	90
1987		3,214		9,923	9,379	3,470	3,214
1988	4,055	13,605	372	143,488	33,991	64,517	18,032
1989	4,147	15,194	14	95,182	5,047	162,046	19,355
1990	10,403	16,761	1	104,657	378	118,188	27,165
1991	7,454	4,113		68,821	187	83,468	11,567
1992	6,612	5,471		108,291	8,721	60,968	12,083
1993	4,694	9,876	1	14,359	7,522	101,489	14,571
1994	2,308	1,739		11,106	4,013	34,493	4,047
1995	5,140	499		13,175	275	31,856	5,639
1996	3,274	568		10,642	14,035	26,965	3,842
1997	1,500	511		5,920	14,591	19,732	2,011
1998	3,020	396		6,812	12,726	13,200	3,416
1999	2,349	162		7,988		9,218	2,511
2000	1,792	315		3,744	1,501	5,667	2,107
2001	2,230	437		2,526	148	3,093	2,667
2002	2,016	507		3,769	1,241	4,524	2,523
2003	1,927	261		3,348	1,211	7,792	2,188
2004	712	146		1,340	676	10,410	858
2005	416	250	10	2,102	201	7,754	676
2006	8,810	377	156	17,815	7,657	13,475	9,343
2007	556	109		1,030		1,135	665
2008				4,894	13,344		-
2009		22,370		10,136	35,297		22,370
2010		12,891		9,843	13,898	556	12,891
2011	9	29,172		15,786	14,716	355	29,181
2012		21,728		20,105	28,303		21,728
2013	1	16,842		15,001	25,113	62	16,843
2014		21,091		12,311	12,463	32,910	21,091
Sum	73,810	198,605	554	743,942	273,659	826,822	272,969

Table 2. Length measurements for neritic tuna species by countries from 1983 to 2014.

Year	EUESP	IDN	IRN	KOR	LKA	MDV	MYS	OMN	PAK	THA	Sum
1983						827	10,357				11,184
1984						659					659
1985						2,338	2,170				4,508
1986						3,136	807				3,943
1987						3,147	11,632	6,307	190		21,276
1988					189,398	3,417	52,624	12,868	1,037		259,344
1989					265,931	1,664		10,632	2,596		280,823
1990					247,139	1,957		744	223		250,063
1991					161,840	1,627			576		164,043
1992			6,436		176,010	3,552			4,065		190,063
1993			5,843		120,986	4,482			6,630		137,941
1994			2,598		10,184	38,230			2,647		53,659
1995			275		23,827	26,843					50,945
1996			14,035		21,655	19,794					55,484
1997			14,591		12,567	15,096					42,254
1998			12,726		15,069	8,359					36,154
1999					17,349	2,368					19,717
2000			1,501		8,637	2,881					13,019
2001			148		7,523	763					8,434
2002			1,241		9,381	1,435					12,057
2003			1,211		7,720	5,608					14,539
2004			676		3,040	9,568					13,284
2005					2,429	6,846				1,458	10,733
2006			925		1,313	3,079				42,973	48,290
2007					2,830						2,830
2008			18,238								18,238
2009			45,086					22,717			67,803
2010			34,569			2,619					37,188
2011			58,890			773			375		60,038
2012			69,937			199					70,136
2013			56,254	2		296	467				57,019
2014	32,335		39,357	46		1,373	5,664				46,440
Sum	32,335	-	384,537	48	1,304,828	172,936	83,721	53,268	18,339	44,431	2,062,108

Table 3. Length measurements for neritic tuna species by gear type and countries.

Gear	EUESP	IDN	IRN	KOR	LKA	MDV	MYS	OMN	PAK	THA	Sum
BB					61	115,831					115,892
BS		1,067									1,067
G/L					5,602						5,602
GILL		4,103	364,096		1,256,977		230	43,786	18,339		1,687,531
HAND					3,817	295		9,482			13,594
HARP					20						20
HATR					406						406
HOOK			6,737								6,737
LLCO					630						630
PS	32,335		12,608								44,943
PSOB				48							48
PSS		14,306					13,888			44,431	72,625
RIN					729						729
TROL		3,473	1,096		35,508	807	69,603				110,487
UNCL					1,078	56,003					57,081
Sum	32,335	22,949	384,537	48	1,304,828	172,936	83,721	53,268	18,339	44,431	2,117,392

Table 4. Length measurements for six neritic tuna species by countries.

Species	EUESP	IDN	IRN	KOR	LKA	MDV	MYS	OMN	PAK	THA	Sum
BLT		385		1	64,189				9	9,226	73,810
COM			119,784		52,732			17,323	8,486		198,325
FRI	32,335	6,842		44	637,956	132,919	6,170		419	10,137	826,822
GUT					388					166	554
KAW		15,408	77,962	3	549,324	40,017	36,240	5,205	2,094	17,689	743,942
LOT		314	186,791		239		41,311	30,740	7,331	6,933	273,659
Sum	32,335	22,949	384,537	48	1,304,828	172,936	83,721	53,268	18,339	44,151	2,117,112

Table 5. Number of unique {Fleet, Year, Month, Grid} strata by fleet and year for Spanish mackerel used for the growth modelling.

Fleet	Year	N strata
IRN	2009	7
IRN	2010	6
IRN	2011	15
IRN	2012	7
IRN	2013	10
IRN	2014	13
LKA	1988	7
LKA	1989	11
LKA	1990	11
LKA	1991	8
LKA	1992	3
LKA	1993	3
LKA	2002	1
OMN	1987	10
OMN	1988	11
OMN	1989	11
OMN	1990	1
OMN	2009	5
PAK	1988	2
PAK	1989	2
PAK	1991	1
PAK	1992	1
PAK	1993	6
		152

Table 6. Posterior key vB model parameters for Spanish mackerel based on the 2-step method.

Param	mean	sd	val2.5pc	val97.5pc
K[1]	0.30	0.06	0.22	0.45
K[2]	0.25	0.05	0.18	0.37
K[3]	0.39	0.04	0.31	0.48
K[4]	0.31	0.07	0.21	0.46
K[5]	0.37	0.08	0.25	0.54
Linf[1]	156	12.21	131.1	176.8
Linf[2]	156.8	12.16	134	177.2
Linf[3]	158.6	9.645	140.4	177.7
Linf[4]	155.7	11.93	130.9	175.3
Linf[5]	157.5	12.98	131.8	181.6
t0[1]	-1.04	0.21	-1.49	-0.66
t0[2]	-1.59	0.31	-2.25	-1.03
t0[3]	-0.91	0.06	-1.03	-0.81
t0[4]	-1.49	0.37	-2.17	-0.93
t0[5]	0.37	0.47	-0.27	1.49
Kmu	0.32	0.08	0.18	0.48
Linfmu	157	11.25	136.5	176.6
t0mu	0.02	3.16	-6.19	6.22

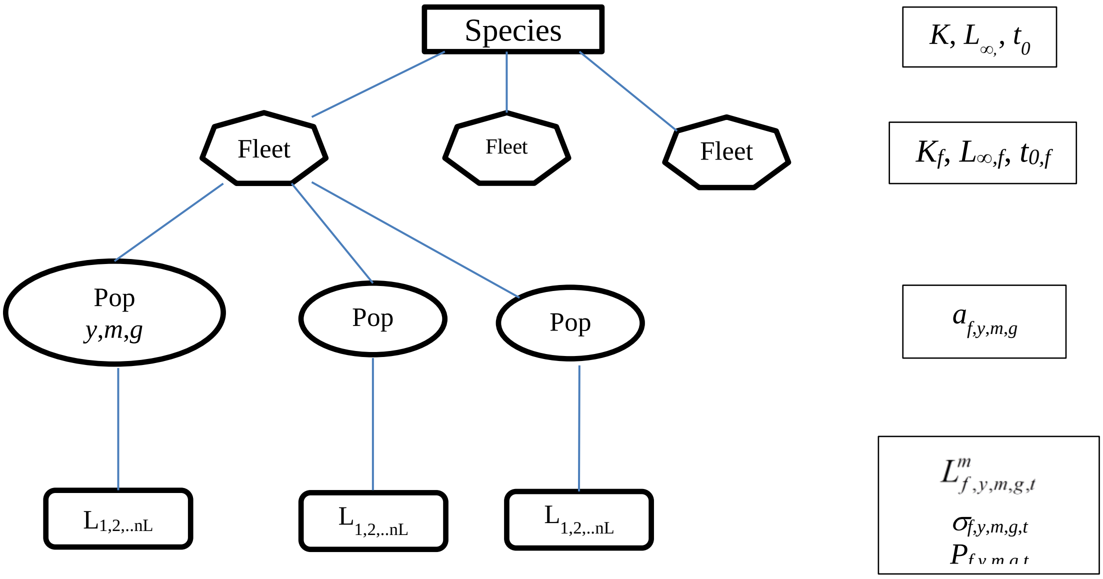


Figure 1. Schematic design of the integrated hierarchical model structure and model parameters.

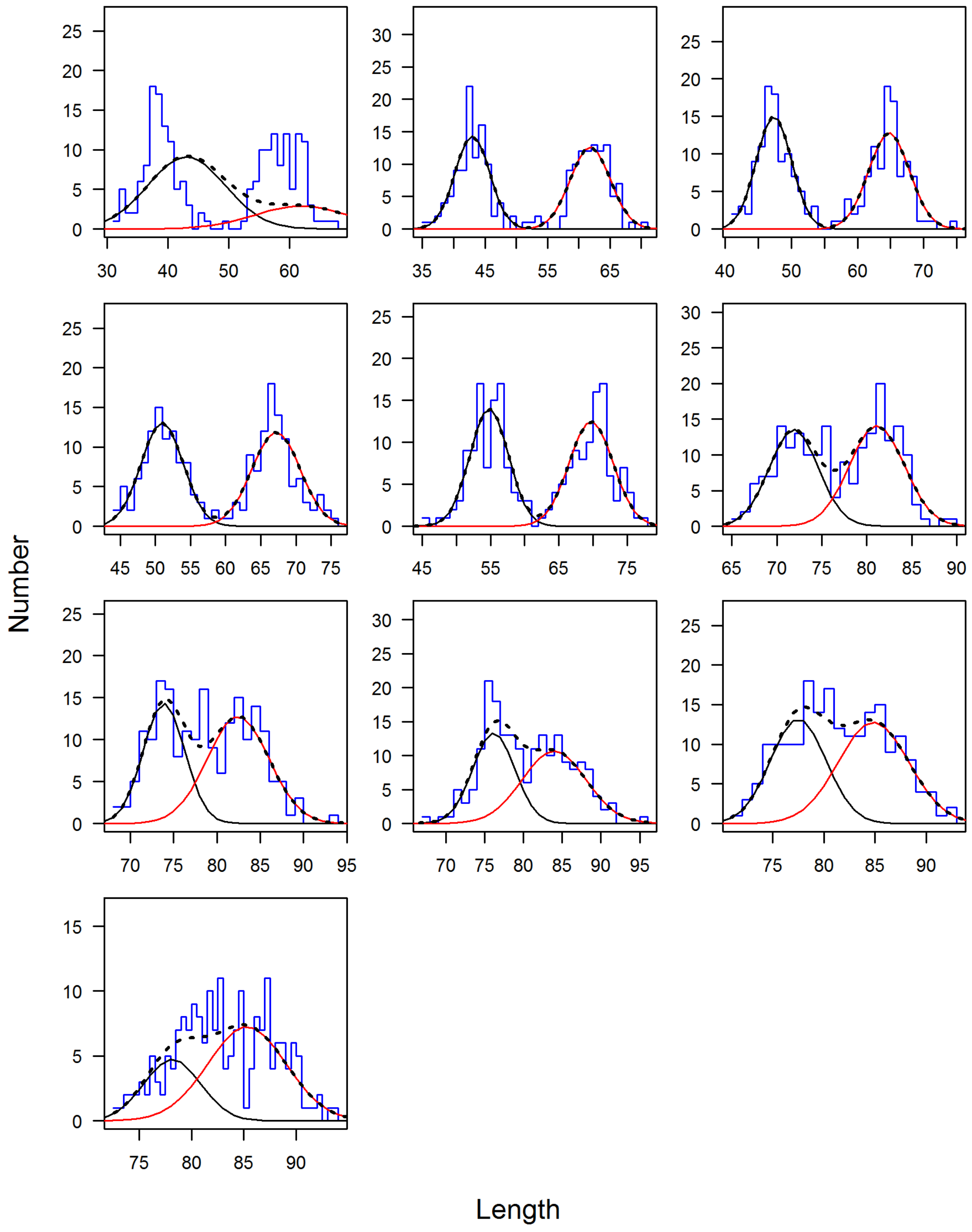


Figure 2. The classic vB growth model fitted to the simulated length data for 10 populations. The blue histogram is the input length-frequency, the solid lines are posterior distributions of distinctive age groups, and the dashed lines are the cumulative distribution.

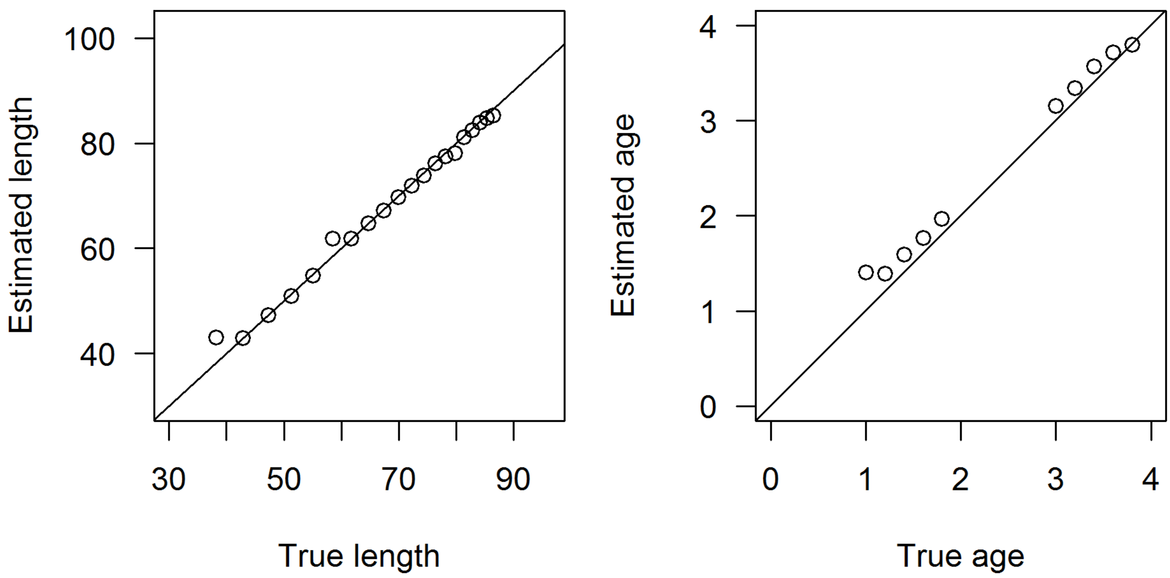


Figure 3. Comparison between the posterior mean length and age with the true simulated length and age.

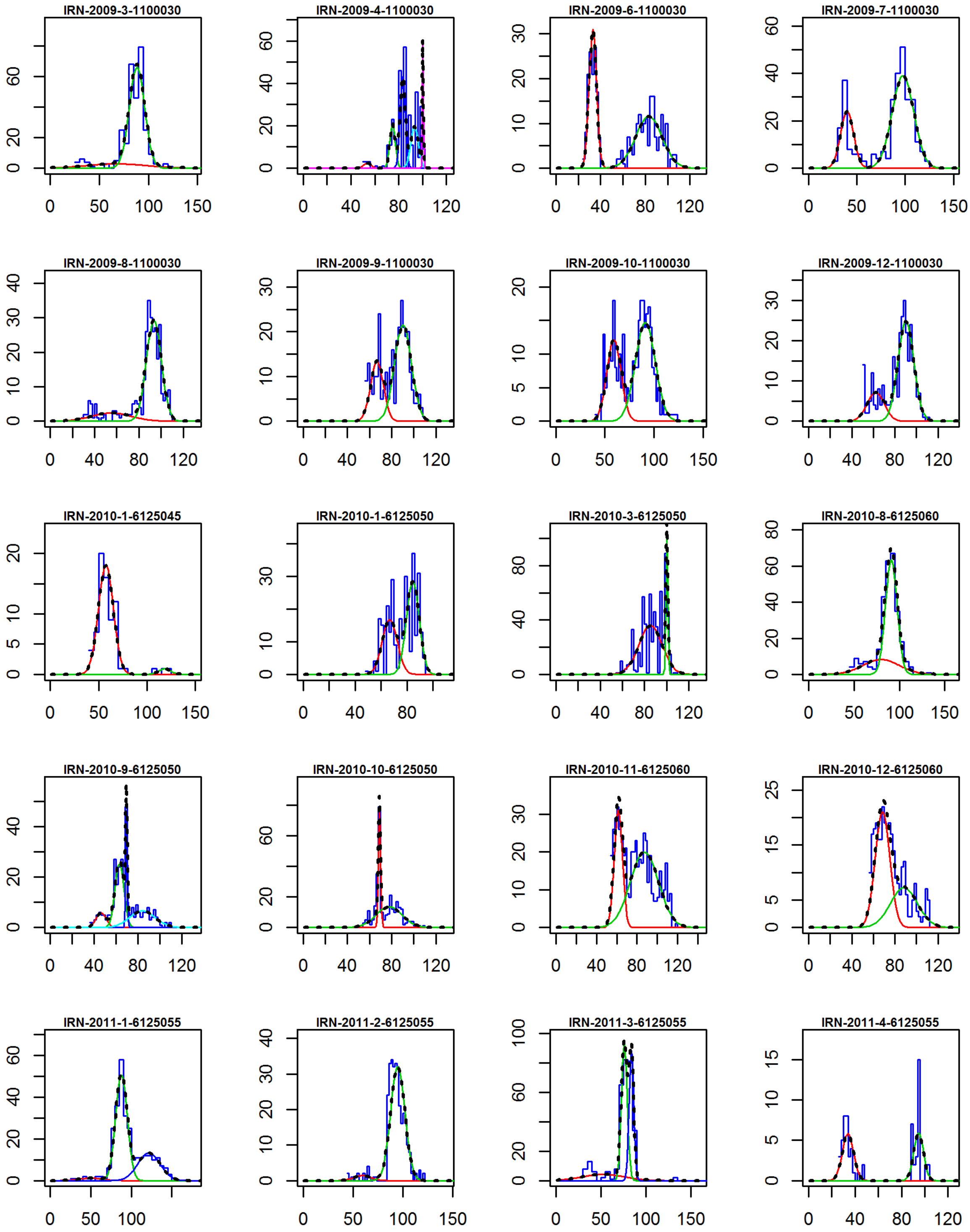


Figure 4. Fitting multi-normal mixture model to length-frequency of Spanish mackerel. The red line is the first mode, green the second, blue the third, and light-blue the fourth. The dashed lines are the cumulative probability curves.

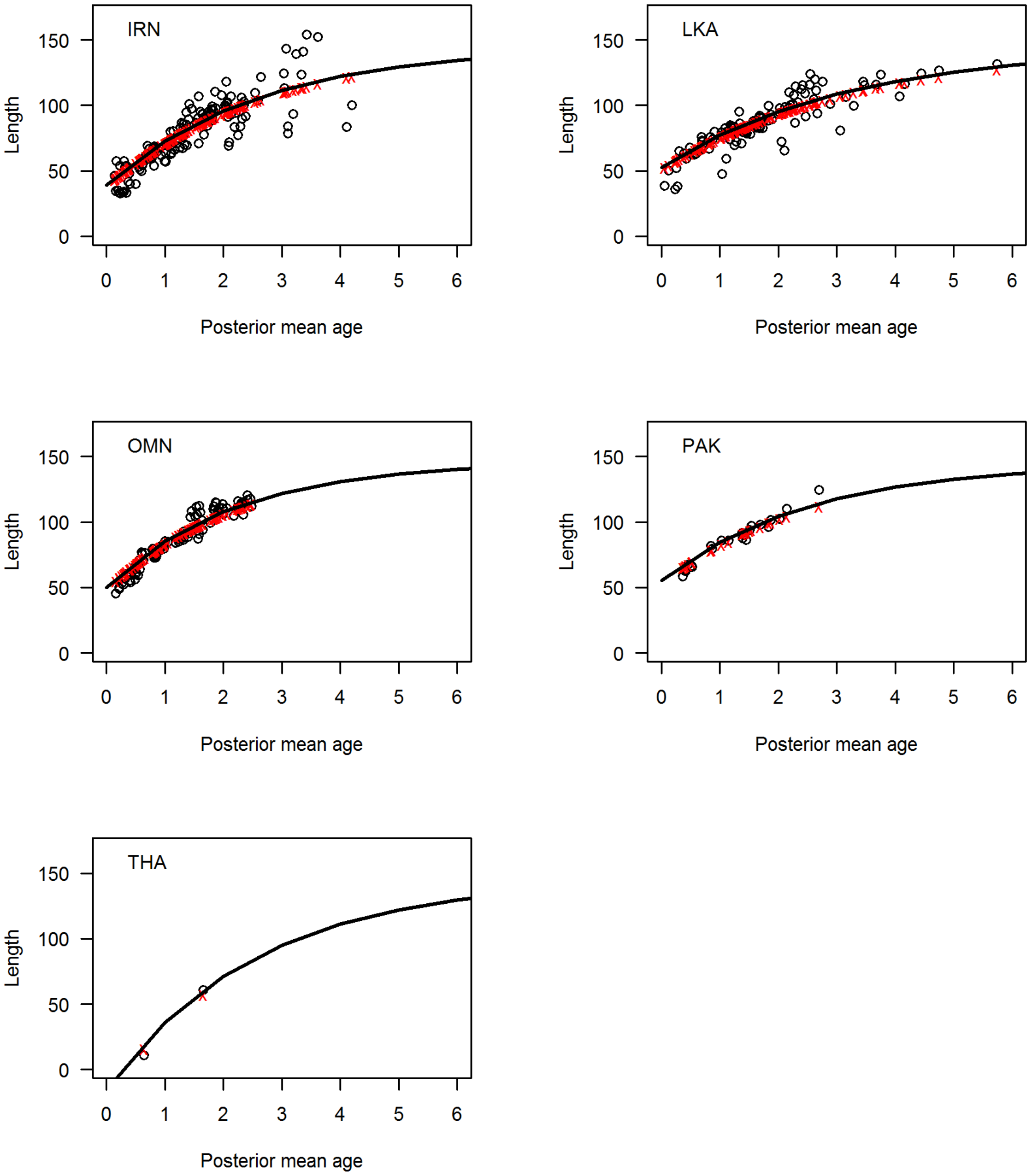


Figure 5. Posterior mean length and mean age for Spanish mackerel in four regions from the 2-step method. The circles are ML estimates and the red xs are corresponding posterior mean. The curves are based on fleet level K, Linf, and t_0.

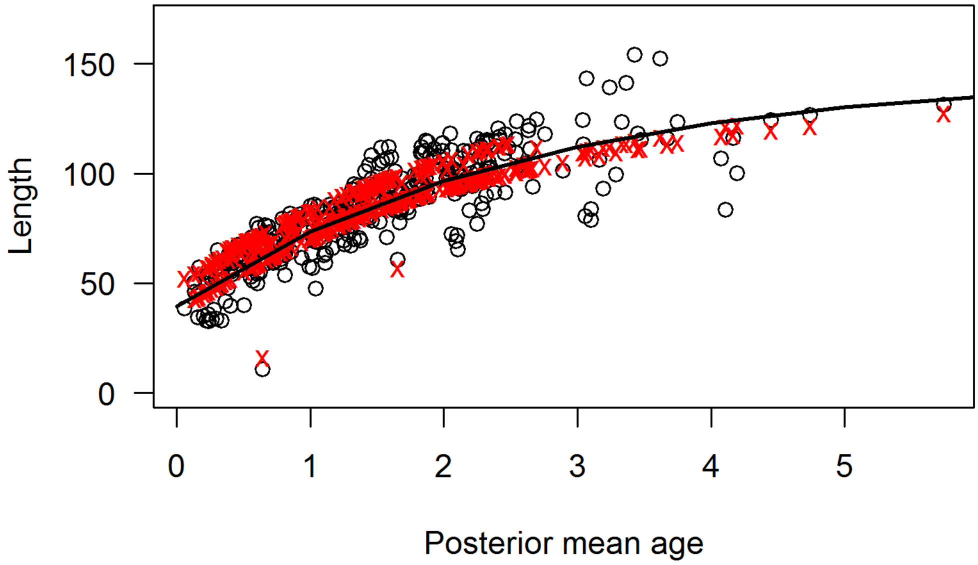


Figure 6. Posterior mean length and mean age for Spanish mackerel in combined regions from the 2-step method. The circles are ML estimates and the red xs are corresponding posterior mean. The curves are based on mean hyper-parameter mu[K], mu[Linf], and mean t_0.

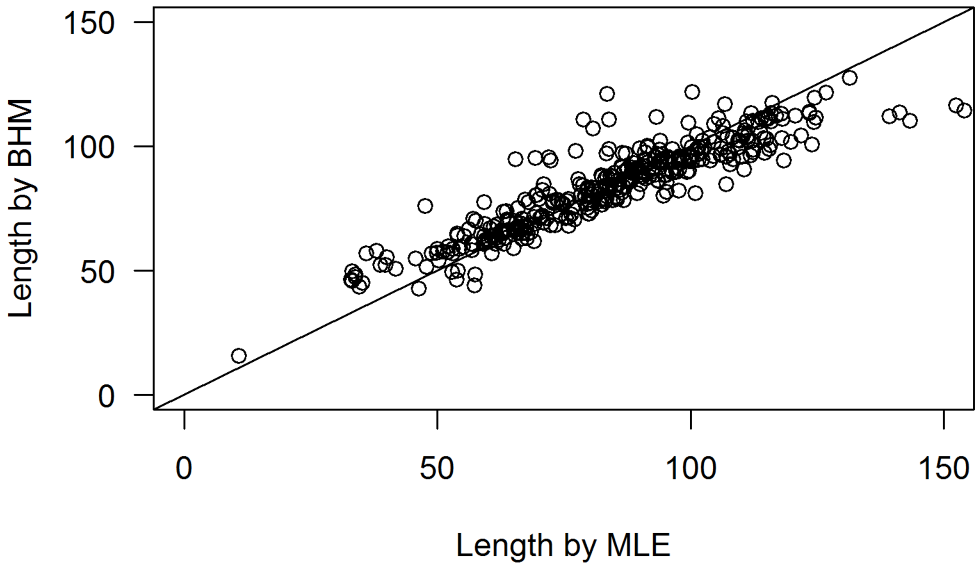


Figure 7. Comparison of length modes between multi-normal mixture model (MLE) and Bayesian hierarchical vB model for Spanish mackerel.

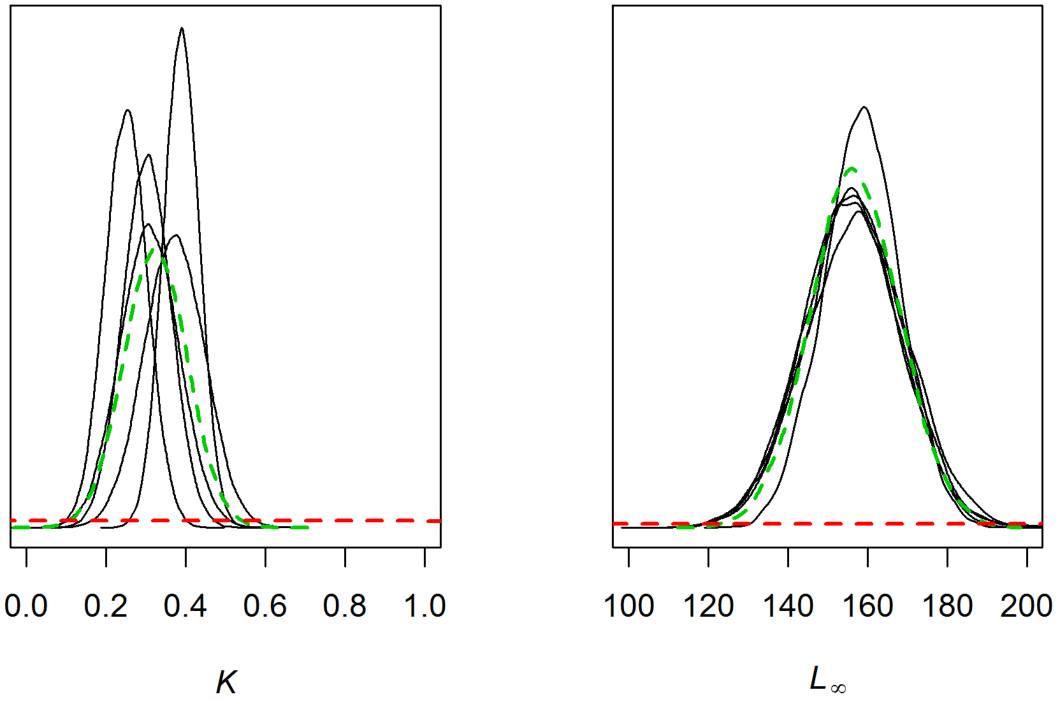
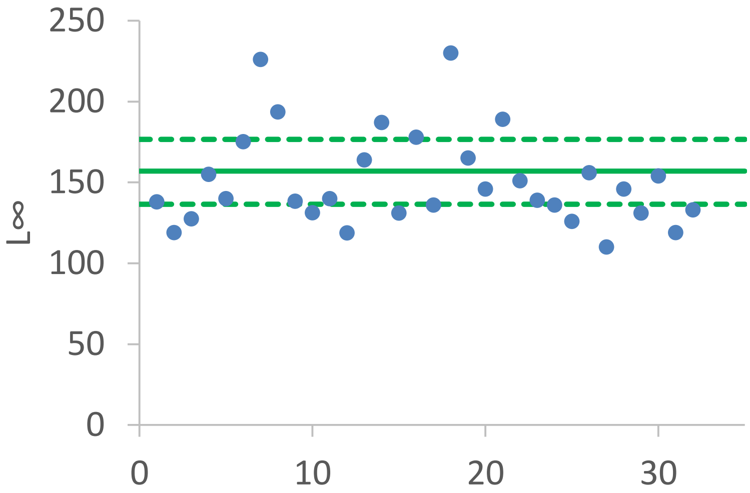


Figure 8. Density of hyper-prior and posterior for Spanish mackerel. The thin lines are for five fleet. The red lines are the hyper-prior and the green lines are posterior of the hyper-parameters.



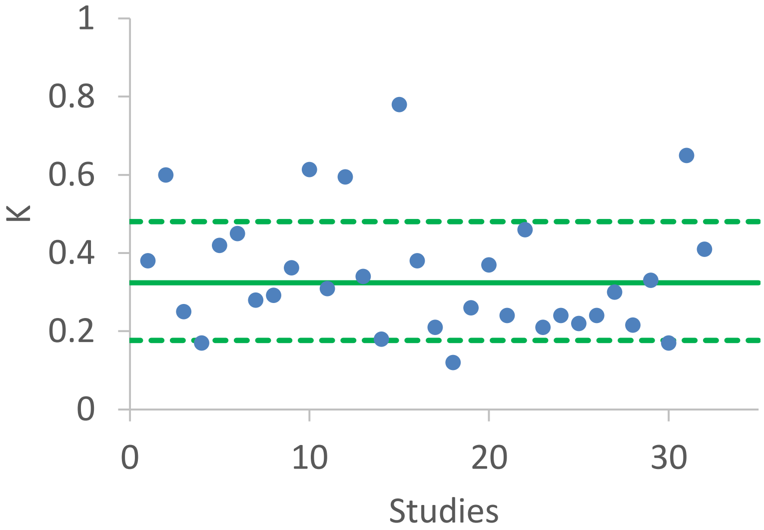


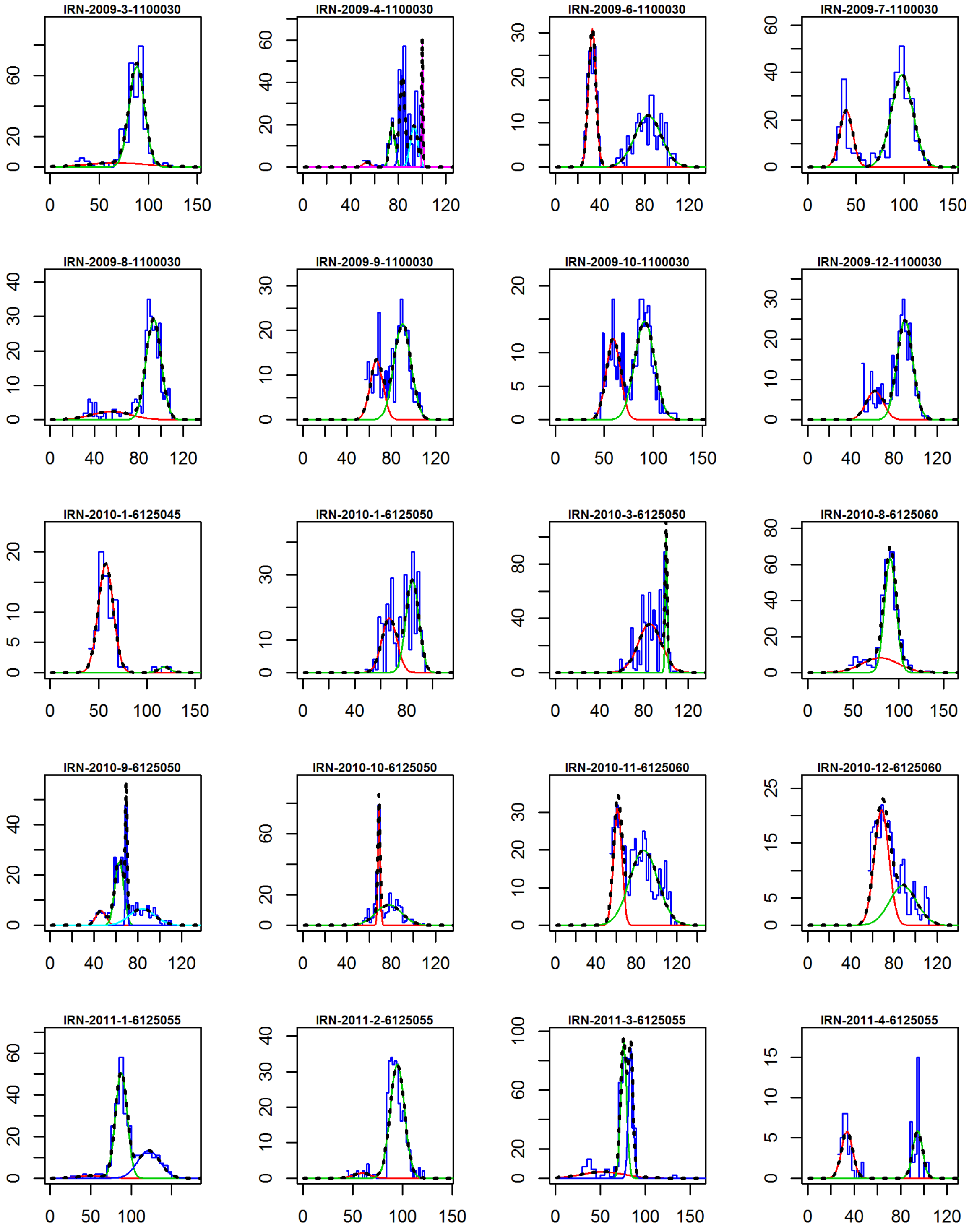
Figure 9. Comparing posterior hyper-parameters L_∞ and K from 2-step method for Spanish mackerel with results from other 32 studies in the Indian Ocean. The solid lines are the mean and the dashed lines are 95% credible interval from current meta-analysis.

Appendix: Bugs code for hierarchical Bayesian model

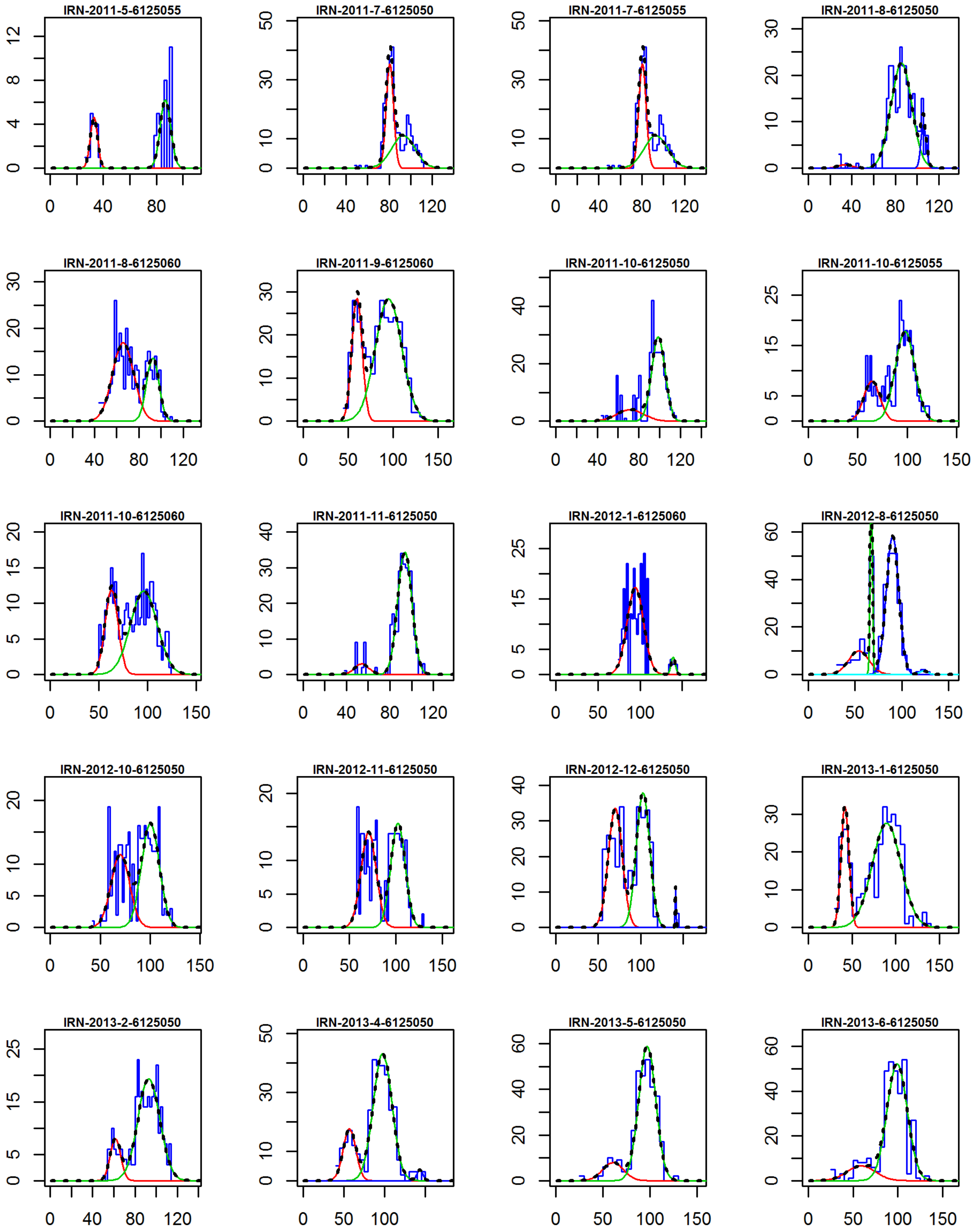
```
model{
# model each length to age-class
for (i in 1:nL){
        L[i]~ dnorm(muL[fleet[i], year[i], mon[i], grid[i], age[i]], tauL[fleet[i], year[i],
mon[i], grid[i], age[i]])
        age_opt[i,1]~dcat(Page2[fleet[i], year[i],mon[i],grid[i],])
         age_opt[i,2]~dcat(Page3[fleet[i], year[i],mon[i],grid[i],])
         age_opt[i,3]~dcat(Page4[fleet[i], year[i],mon[i],grid[i],])
        age_opt[i,4]~dcat(Page5[fleet[i], year[i],mon[i],grid[i],])
        age[i] <- age_opt[i, (Nage[i]-1)]
}

# hyper-priors
muK ~ dnorm(-1, 0.1)
tauK ~ dgamma(0.1,0.1)
muLinf~dnorm(100,0.000001)
tauLinf~dgamma(0.01, 0.01) # dunif(0, 10) #

# fitting vB model
for (f in 1:nf) {
# priors
        K[f]~ dlnorm(muK, tauK)I(0,3)
        Linf[f]~dnorm(muLinf, tauLinf) I(0,)
        t0[f] ~ dnorm(0, 0.1)I(-1,1)

        for (y in 1:ny[f]){
          for (m in 1:nm[f,y]){
            for (g in 1:ng[f,y,m]){
                Page2[f,y,m,g, 1:2] ~ ddirch(alpha2[])
                Page3[f,y,m,g, 1:3] ~ ddirch(alpha3[])
                Page4[f,y,m,g, 1:4] ~ ddirch(alpha4[])
                Page5[f,y,m,g, 1:5] ~ ddirch(alpha5[])
                 age0[f,y,m,g] ~ dnorm(0,0.1)I(-0.5,3)

                for ( a in 1:maxNa ) {
                   sdL[f,y,m,g,a] ~ dnorm(0, 0.1)I(0.01,100)
                    muL[f,y,m,g,a] <- Linf[f]*(1-exp(-K[f]*(age0[f,y,m,g] +a-t0[f])))
                                                }
                                            }
                                        }
                                 }
                            }
```

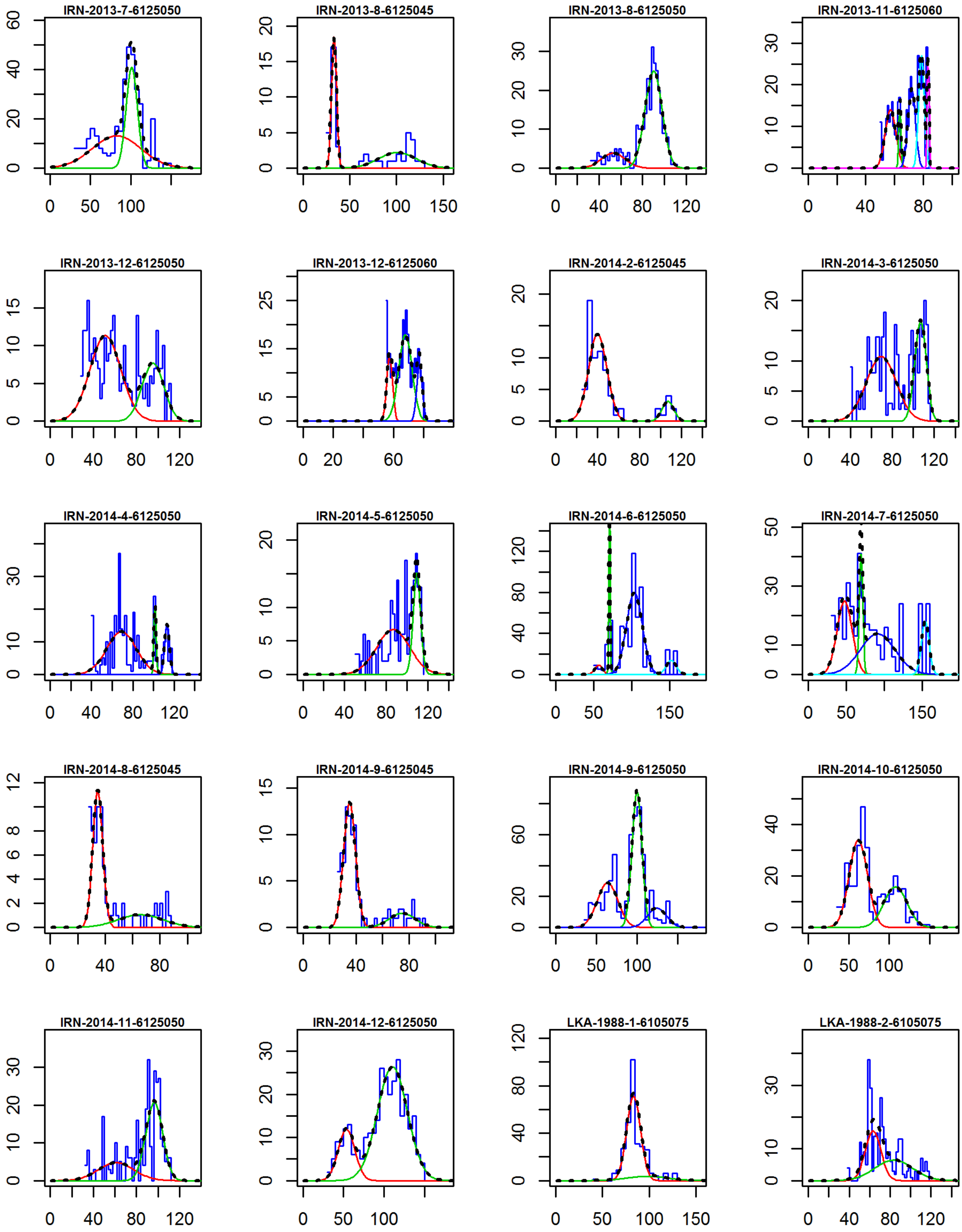
```
}
```

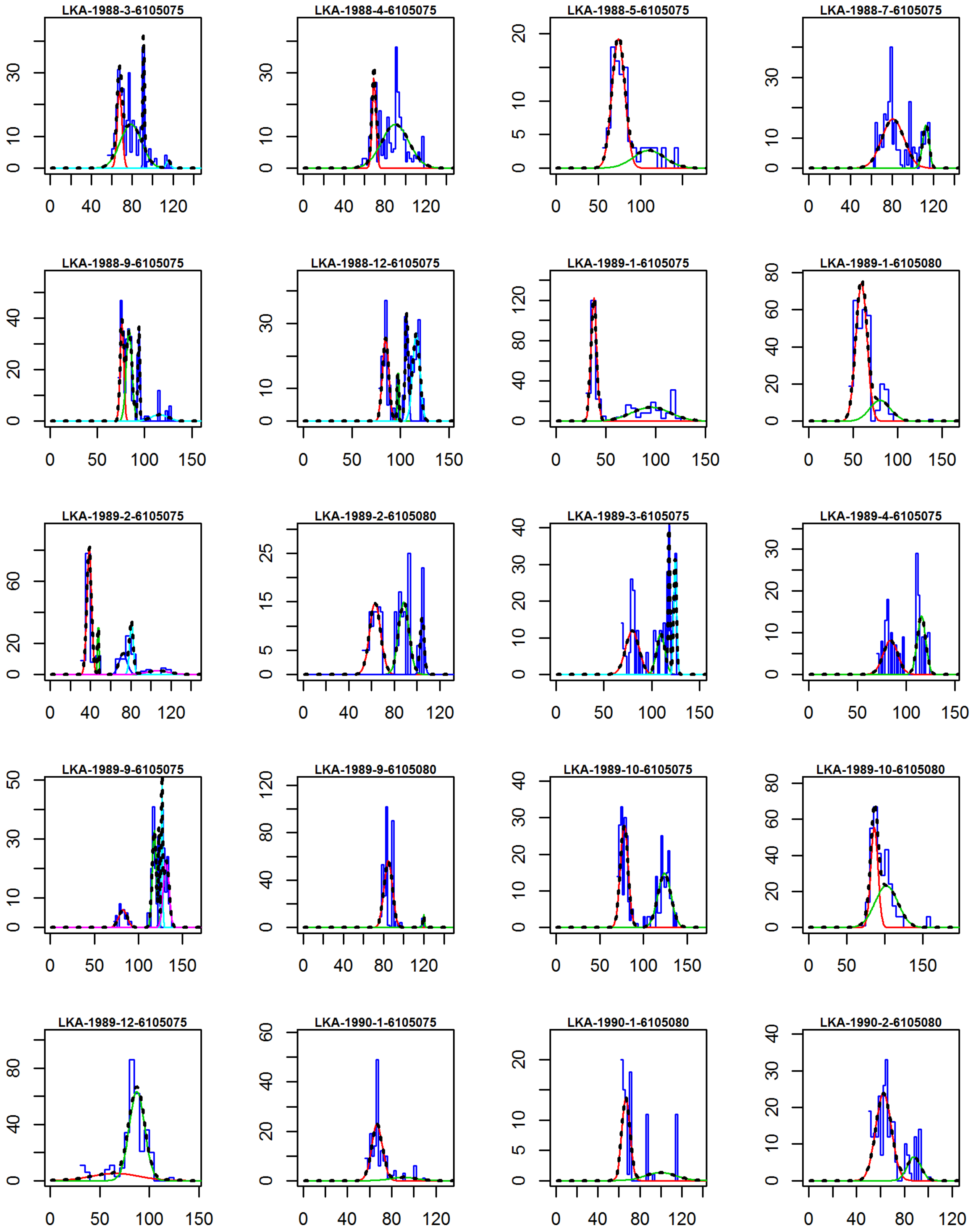
Appendix 2. Fitting multi-normal mixture model to length-frequency of Spanish mackerel. The red line is the first mode, green the second, blue the third, and light-blue the fourth. The dashed lines are the cumulative probability curves.

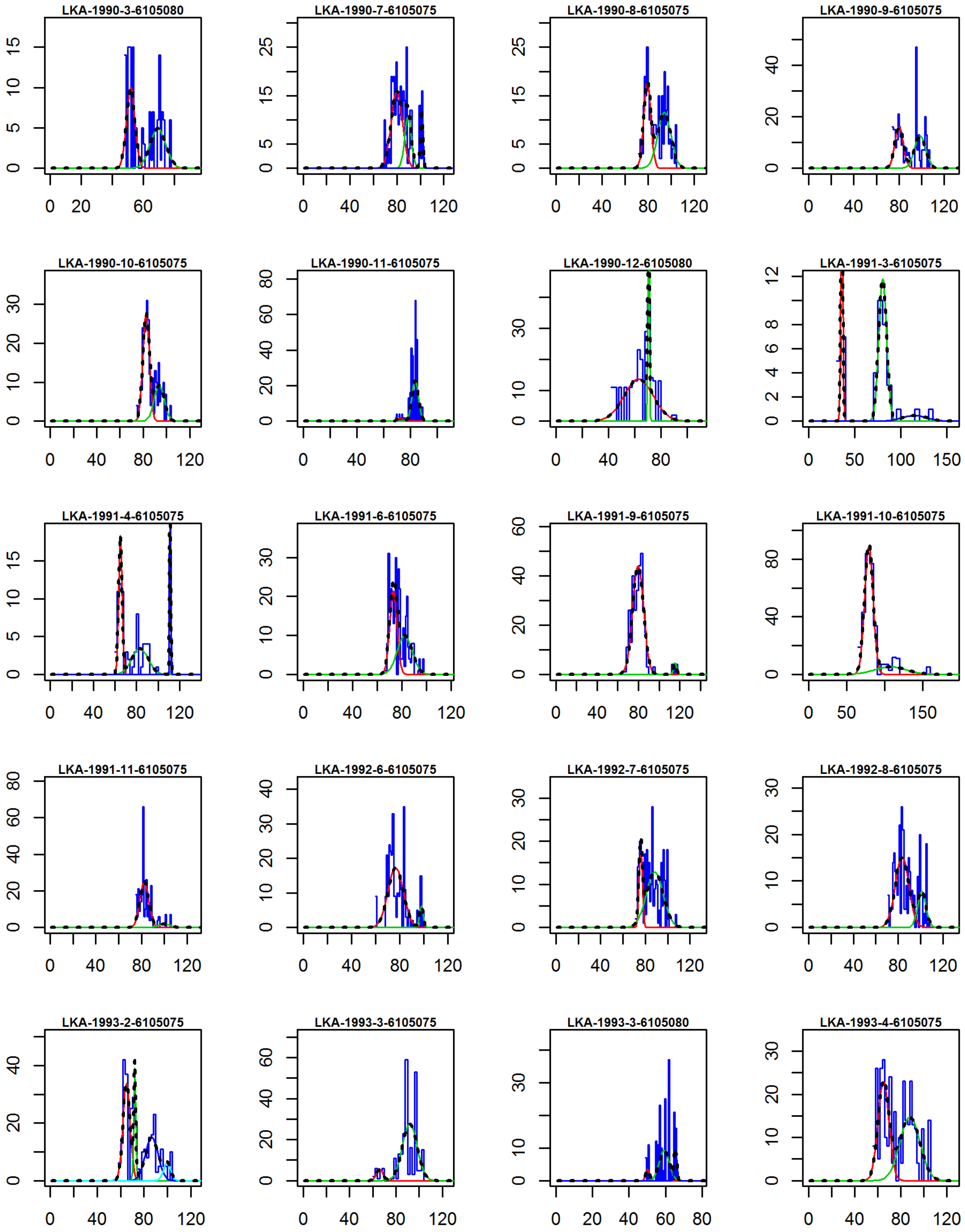


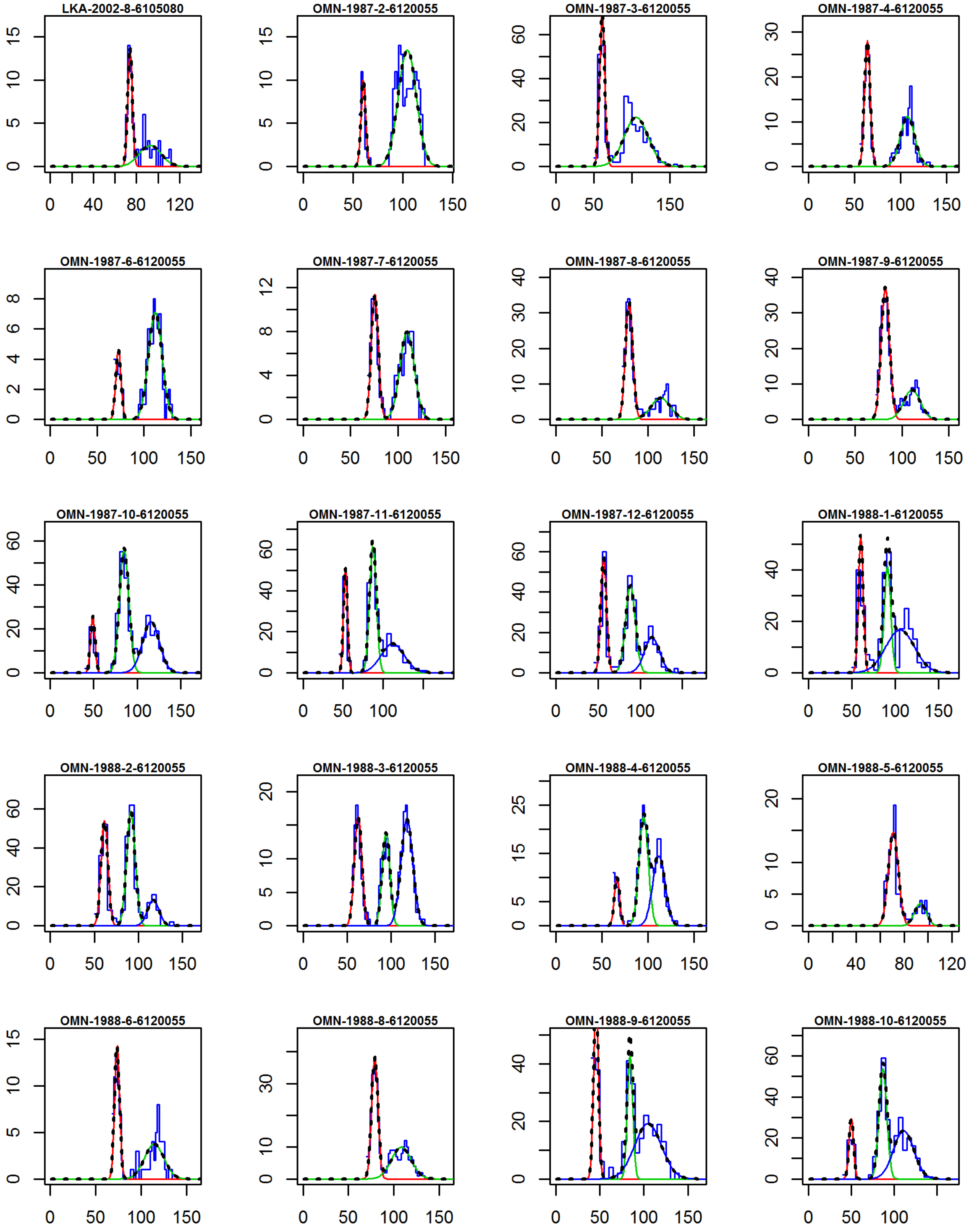
IRN-2009-3-1100030
IRN-2009-4-1100030
IRN-2009-6-1100030
IRN-2009-7-1100030
IRN-2009-8-1100030
IRN-2009-9-1100030
IRN-2009-10-1100030
IRN-2009-12-1100030
IRN-2010-1-6125045
IRN-2010-1-6125050
IRN-2010-3-6125050
IRN-2010-8-6125060
IRN-2010-9-6125050
IRN-2010-10-6125050
IRN-2010-11-6125060
IRN-2010-12-6125060
IRN-2011-1-6125055
IRN-2011-2-6125055
IRN-2011-3-6125055
IRN-2011-4-6125055

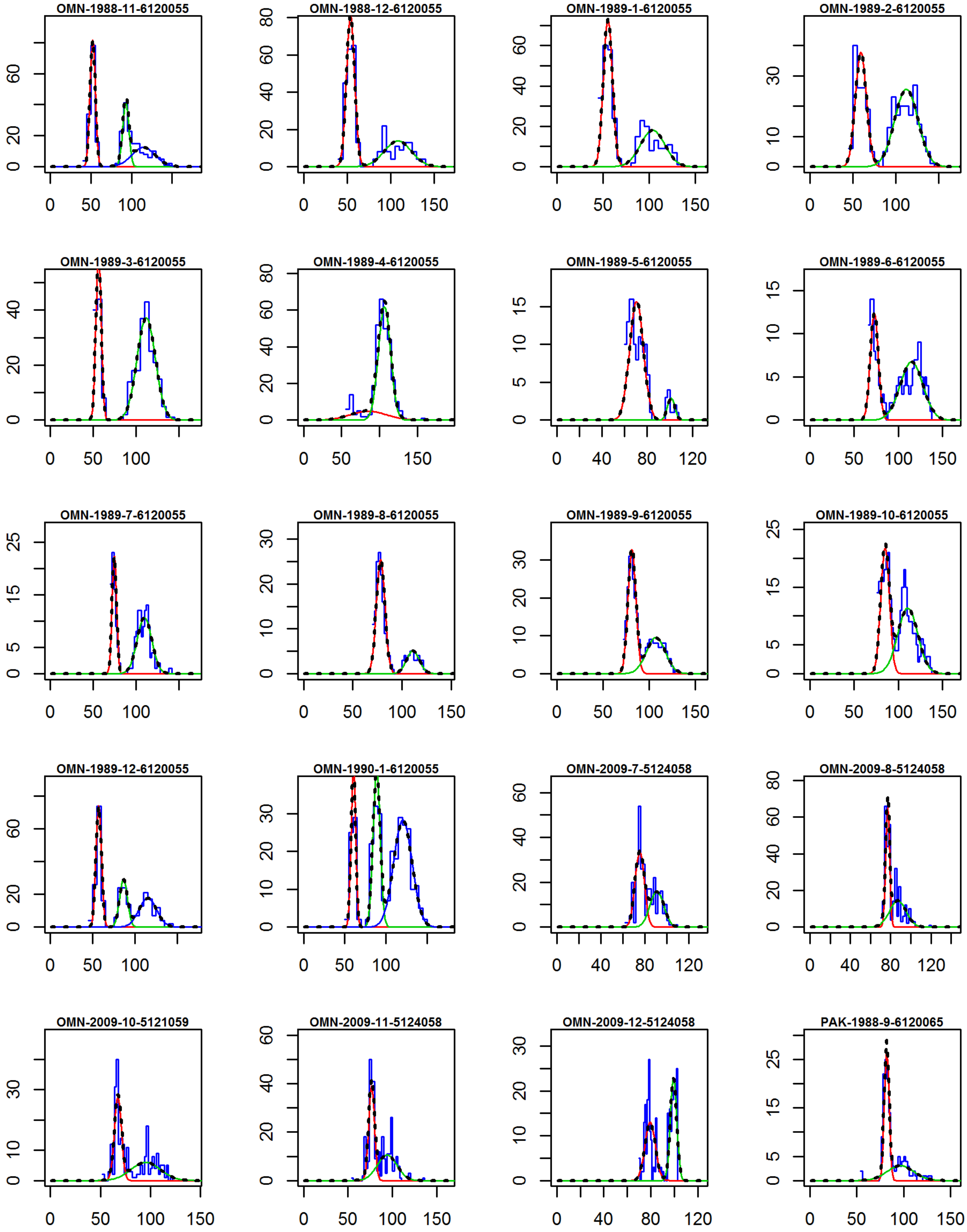

IRN-2011-5-6125055
IRN-2011-7-6125050
IRN-2011-7-6125055
IRN-2011-8-6125050
IRN-2011-8-6125060
IRN-2011-9-6125060
IRN-2011-10-6125050
IRN-2011-10-6125055
IRN-2011-10-6125060
IRN-2011-11-6125050
IRN-2012-1-6125060
IRN-2012-8-6125050
IRN-2012-10-6125050
IRN-2012-11-6125050
IRN-2012-12-6125050
IRN-2013-1-6125050
IRN-2013-2-6125050
IRN-2013-4-6125050
IRN-2013-5-6125050
IRN-2013-6-6125050

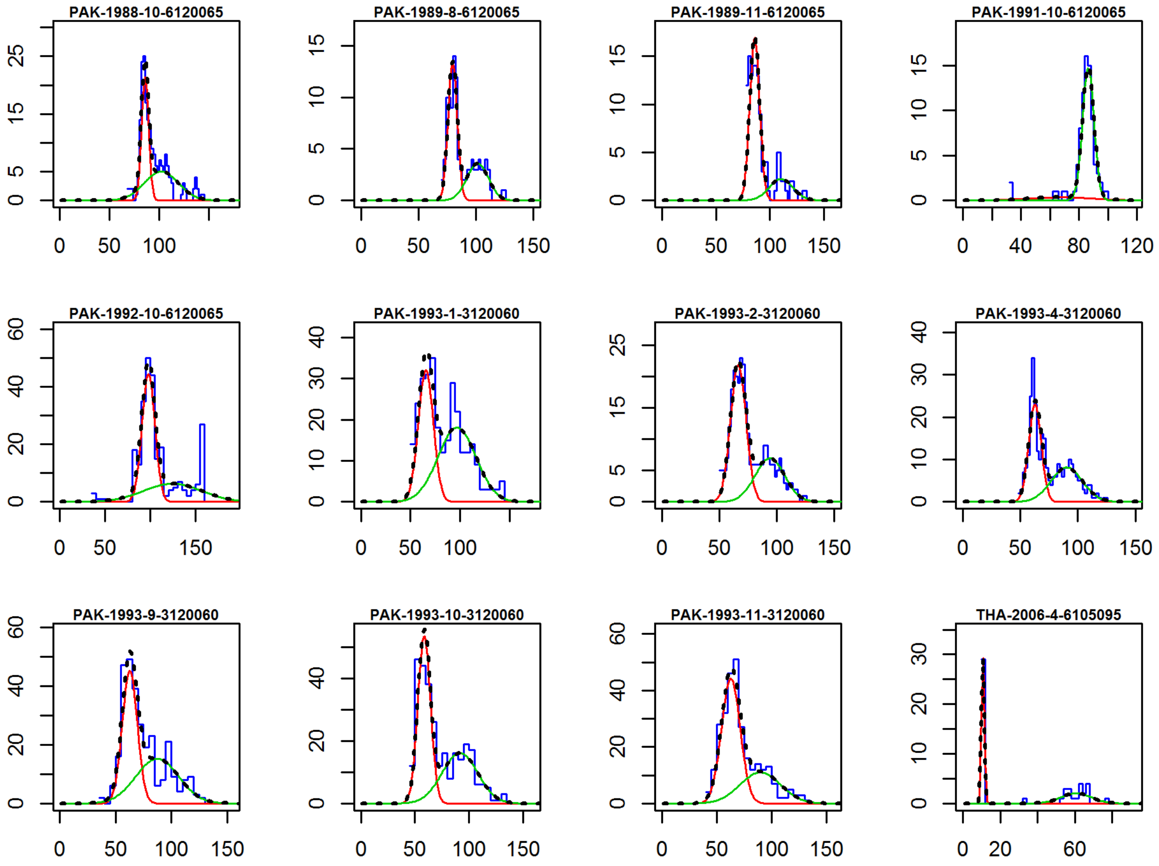

IRN-2013-7-6125050
0 20 40 60
0 50 100
IRN-2013-8-6125045
0 5 10 15 20
0 50 100 150
IRN-2013-8-6125050
0 10 20 30
0 40 80 120
IRN-2013-11-6125060
0 10 20 30
0 40 80
IRN-2013-12-6125050
0 5 10 15
0 40 80 120
IRN-2013-12-6125060
0 5 15 25
0 20 60
IRN-2014-2-6125045
0 5 10 20
0 40 80 120
IRN-2014-3-6125050
0 5 10 20
0 40 80 120
IRN-2014-4-6125050
0 10 30
0 40 80 120
IRN-2014-5-6125050
0 5 10 20
0 40 80 120
IRN-2014-6-6125050
0 40 80 120
0 50 150
IRN-2014-7-6125050
0 10 30 50
0 50 150
IRN-2014-8-6125045
0 2 4 6 8 12
0 40 80
IRN-2014-9-6125045
0 5 10 15
0 40 80
IRN-2014-9-6125050
0 20 60
0 50 100
IRN-2014-10-6125050
0 20 40
0 50 100
IRN-2014-11-6125050
0 10 20 30
0 40 80 120
IRN-2014-12-6125050
0 10 20 30
0 50 100
LKA-1988-1-6105075
0 40 80 120
0 50 100 150
LKA-1988-2-6105075
0 10 30
0 40 80 120


LKA-1988-3-6105075
LKA-1988-4-6105075
LKA-1988-5-6105075
LKA-1988-7-6105075
LKA-1988-9-6105075
LKA-1988-12-6105075
LKA-1989-1-6105075
LKA-1989-1-6105080
LKA-1989-2-6105075
LKA-1989-2-6105080
LKA-1989-3-6105075
LKA-1989-4-6105075
LKA-1989-9-6105075
LKA-1989-9-6105080
LKA-1989-10-6105075
LKA-1989-10-6105080
LKA-1989-12-6105075
LKA-1990-1-6105075
LKA-1990-1-6105080
LKA-1990-2-6105080


LKA-1990-3-6105080
0 5 10 15
0 20 60
LKA-1990-7-6105075
0 5 15 25
0 40 80 120
LKA-1990-8-6105075
0 5 15 25
0 40 80 120
LKA-1990-9-6105075
0 20 40
0 40 80 120
LKA-1990-10-6105075
0 10 20 30
0 40 80 120
LKA-1990-11-6105075
0 20 40 60 80
0 40 80
LKA-1990-12-6105080
0 10 30
0 40 80
LKA-1991-3-6105075
0 2 4 6 8 12
0 50 100 150
LKA-1991-4-6105075
0 5 10 15
0 40 80 120
LKA-1991-6-6105075
0 10 20 30
0 40 80 120
LKA-1991-9-6105075
0 20 40 60
0 40 80 120
LKA-1991-10-6105075
0 40 80
0 50 150
LKA-1991-11-6105075
0 20 40 60 80
0 40 80 120
LKA-1992-6-6105075
0 10 20 30 40
0 40 80 120
LKA-1992-7-6105075
0 10 20 30
0 40 80 120
LKA-1992-8-6105075
0 10 20 30
0 40 80 120
LKA-1993-2-6105075
0 20 40
0 40 80 120
LKA-1993-3-6105075
0 20 40 60
0 40 80 120
LKA-1993-3-6105080
0 10 30
0 20 40 60 80
LKA-1993-4-6105075
0 10 20 30
0 40 80 120


LKA-2002-8-6105080
OMN-1987-2-6120055
OMN-1987-3-6120055
OMN-1987-4-6120055
OMN-1987-6-6120055
OMN-1987-7-6120055
OMN-1987-8-6120055
OMN-1987-9-6120055
OMN-1987-10-6120055
OMN-1987-11-6120055
OMN-1987-12-6120055
OMN-1988-1-6120055
OMN-1988-2-6120055
OMN-1988-3-6120055
OMN-1988-4-6120055
OMN-1988-5-6120055
OMN-1988-6-6120055
OMN-1988-8-6120055
OMN-1988-9-6120055
OMN-1988-10-6120055
0
40
80
120
50
100
150


OMN-1988-11-6120055
OMN-1988-12-6120055
OMN-1989-1-6120055
OMN-1989-2-6120055
OMN-1989-3-6120055
OMN-1989-4-6120055
OMN-1989-5-6120055
OMN-1989-6-6120055
OMN-1989-7-6120055
OMN-1989-8-6120055
OMN-1989-9-6120055
OMN-1989-10-6120055
OMN-1989-12-6120055
OMN-1990-1-6120055
OMN-2009-7-5124058
OMN-2009-8-5124058
OMN-2009-10-5121059
OMN-2009-11-5124058
OMN-2009-12-5124058
PAK-1988-9-6120065
0
50
100
150
40
80
120


PAK-1988-10-6120065
PAK-1989-8-6120065
PAK-1989-11-6120065
PAK-1991-10-6120065
PAK-1992-10-6120065
PAK-1993-1-3120060
PAK-1993-2-3120060
PAK-1993-4-3120060
PAK-1993-9-3120060
PAK-1993-10-3120060
PAK-1993-11-3120060
THA-2006-4-6105095
0
5
15
25
50
100
150
10
40
80
120
20
30
60