

# **Modeling marine turtle somatic growth dynamics: Exploring more than just mean growth**

George Balazs<sup>1\*</sup>, Stacy Hargrove<sup>2</sup>, Milani Chaloupka<sup>3\*</sup>

<sup>1</sup> NOAA Pacific Islands Fisheries Science Center, Marine Turtle Biology and Assessment Program, Honolulu, Hawaii.

<sup>2</sup> NOAA Southwest Fisheries Science Center, Miami Florida

<sup>3</sup> Ecological Modelling Services, St. Lucia, Queensland, Australia

\*Corresponding Authors

## **Abstract**

*We found some benefit of explicitly accounting for the growth rate variability but it was limited for these small data sets. Nonetheless, the GAMLSS approach holds promise for modelling marine turtle somatic growth dynamics when larger data sets are available with a broader range of potentially informative covariates. Importantly, the random effects GAMLSS models fit using Bayesian inference suggested fewer significant covariates than the frequentist inferential procedure. This might be due to the Bayesian procedure being better suited to accounting for the uncertainty in model parameter estimates when applied to small data sets.*

## **Introduction**

Green sea turtles (*Chelonia mydas*) are globally distributed in tropical and subtropical waters and are listed as Threatened or Endangered under the US Endangered Species Act (ESA). The Hawaiian green turtle population is genetically distinct from other global green turtle stocks (Dutton et al. 2008) and belongs to the Central North Pacific

distinct population segment (DPS), which is currently listed as Threatened under the ESA but Least Concern Red List category by the International Union for Conservation of Nature [www.iucnredlist.org/details/16285718/0](http://www.iucnredlist.org/details/16285718/0). Hawaiian green turtles were at the brink of extinction in the 1960's, but state and federal protection resulted in four decades of population increases. The population is now among the best studied and best protected worldwide of any marine turtle population (Balazs et al. 2015). The strong dichotomy in conservation status between green turtle populations- some increasing or stable, and others declining- is one of the great challenges of our time in balancing protective restrictions versus human use, where such utilization might be advantageous to both turtles and people (see Broderick 2015 and Kondo et al. 2017 for relevant discussions).

Hawaiian green turtles live in coastal habitats and exhibit a high degree of site fidelity to their specific feeding locations. At some foraging locations, green turtles appear to be at or near carrying capacity (Wabnitz et al 2010). Decades of population growth have resulted in slow and declining somatic growth rates across the Hawaiian Archipelago (Balazs & Chaloupka 2004). Yet, some portions of the population behave differently, growing and maturing faster than expected (Van Houtan et al 2014). These individual differences in growth are the focus of this study.

Marine turtle somatic growth studies generally focus on the mean or expected growth behavior and do not address the variability in that growth (Chaloupka et al 2004, Balazs & Chaloupka 2004, Kubis et al 2009, Bjorndal et al 2016). In fact the mean response is the most commonly modeled summary statistic for most purposes across many disciplines (Kneib 2013). Marine turtle somatic growth rates can be highly

variable (Chaloupka & Balazs 2005, Bjorndal et al 2016,2017), and that heterogeneity might be a function of important predictors or covariates leading to further insight into marine turtle growth dynamics. Yet that variability has not been accounted for in any marine turtle growth rate study. So the main purpose of our study was to explore green turtle somatic growth dynamics in terms of not only the mean growth dynamics but also in terms of the variability in that growth behavior.

We address this issue of modeling beyond the mean response by using the GAMLSS regression modeling approach (Stasinopoulos & Rigby 2007). Potentially informative covariates can then be specified not only for the mean of the growth rate response but also for a range of other summary parameters of the response distribution such as the variance, skewness and kurtosis (Stasinopoulos & Rigby 2007, Kneib 2013).

## **Materials and Methods**

### Data set and sampling design

The data set comprised 267 growth records for carapace size of 108 individual immature green turtles captured in two Hawaiian Archipelago foraging grounds in an ongoing long-term and spatially extensive capture–mark–recapture program (Balazs & Chaloupka 2004, Chaloupka & Balazs 2005, Franke et al 2013). The foraging-ground samples were from sampling areas at Kawainui Marsh Estuary and Hanauma Bay, both on the Hawaiian Island of Oahu. Sampling occurred over a 16-year period from 1999-2014. There were a small number of growth rate measurements recorded

from earlier years well prior to 1999, but those measurements were not included here to maintain a continuous sampling period from 1999-2014.

The Kawainui sampling area is located in Kailua Bay on the southeastern coast of Oahu. The shallow benthic habit type at Kawainui is composed mainly of pavement-type coral reef and macroalgae (Franke et al 2013). The turtles here were all immature turtles with carapace length ranging from ca. 37-76 cm SCL. Hanauma Bay is a marine **embayment** formed within an extinct volcanic tuff ring located also along the southeast coast of Oahu. It is one of the most popular tourist destinations in Hawaii with more than 1 million visitors a year. It is also a Nature Preserve and a Marine Life Conservation District and the shallow benthic habit is pavement-type coral reef and macroalgae. These turtles were also immature with carapace length ranging from 38-64 cm SCL. Turtles at both study sites were captured, marked, and released using a variety of methods such as scoop nets or by hand (Franke et al 2013) and double-tagged using Passive Integrated Transponders (Balazs & Chaloupka 2004).

The capture–mark-recapture profiles recorded for each turtle included: (1) carapace size recorded to the nearest millimeter as straight carapace length (cm SCL) at first capture and any subsequent recaptures, (2) year of first capture and (3) the years at large since first capture or previous recapture. More details on measurement methods can be found in Balazs & Chaloupka (2004). The fibropapilloma disease status was also recorded for each turtle (Chaloupka & Balazs 2005). Absolute growth rates were derived from the capture–mark-recapture profiles for each foraging-ground sample, with negative or zero growth rates included, since these are part of the measurement error (Balazs & Chaloupka 2004).

## Statistical modeling approach

We used the GAMLSS semiparametric or generalized additive regression modeling approach for location, scale, and shape (Stasinopoulos & Rigby 2007), which is a form of structured additive distributional regression (Klein et al 2015). The GAMLSS approach supports the simultaneous modelling of (1) the mean (or expected) somatic growth rate response as a function of potentially informative covariates as well as (2) the variance or the dispersion of that same response also as a function of informative covariates. This approach was used recently to explore appropriate regression model structure for a study reviewing risk factors for seabird bycatch in the Hawaii-based pelagic longline tuna fishery (Gilman et al 2016). The covariates used here included mean carapace size, mean sampling year, growth increment duration, disease status, and the identity of each turtle (Balazs & Chaloupka 2004, Chaloupka & Balazs 2005, see also Bjorndal et al 2016 for modelling marine turtle growth rates within a spatio-temporal context).

The data sets in our study were limited in terms of negative growth estimates due to measurement error and small sample size, especially for each sampling year, so we used 2 different inferential procedures to fit the GAMLSS model structure given those data limitations: (1) direct optimization of a penalized maximum likelihood using the `gamlss` package for R (Stasinopoulos & Rigby 2007) and (2) Bayesian inference implemented with the Stan computation engine (Stan Development Team 2016) with NUTS sampling (Carpenter et al 2017) via the `brms` package for R (Bürkner in press).

The negative growth rate estimates in this study limited the range of model likelihood that could be used (Voudouris et al 2012). So we used a robust Student-*t* likelihood, as suggested elsewhere by Gilman et al (2016), and then we compared the Bayesian GAMLSS models to similar models fitted with Gaussian likelihood using leave-one-out cross-validation (LOO-cv) via the *loo* package for R (Vehtari et al 2017).

The GAMLSS regression models used in either framework comprised: (1) an identity link, (2) Student-*t* model likelihood to minimize outlier effects on parameter estimates and (3) either penalized spline or thin plate regression spline smooths (Wood 2006) to model any nonlinear functional form between mean turtle somatic growth rates and potentially informative covariates such as sampling year, turtle carapace size and the interval between subsequent recaptures. The somatic growth rate variance was also jointly modeled as a function of sampling year and turtle carapace size. We also included random effects (random intercepts) to account for any turtle-specific heterogeneity due to the variable number of growth rate measurements recorded for each turtle (Chaloupka & Balazs 2005, see also Bjorndal et al 2016). These are now referred to as random-effects GAMLSS or GAMMLSS models.

GAMMLSS regression models implemented using a frequentist inferential approach were then evaluated for goodness-of-fit using quantile residual diagnostics (Dunn & Smyth 1996) and worm plots (detrended Q-Q plots) that are useful in checking the assumed distribution of the response variable and also for checking that the response distribution was fitted adequately for the range of a specific explanatory variable (van Buuren & Fredriks 2001). The GAMMLSS models implemented in the Bayesian framework used weakly informative regularizing priors (Gelman et al 2008, Park &

Casella 2008) with posterior samples sourced from 5 chains and 50k iterations after a warmup of 2000 iterations. Bayesian GAMMLSS regression model fit was displayed using the ggplot2 package for R (Wickham 2016) and then evaluated using graphical posterior predictive checking procedures (Gelman & Hill 2007, Gelman et al 2014) via the bayesplot package for R (Gabry 2016). We also then used LOO-cv to compare the GAMMLSS fits to simpler GAMM models to determine if accounting explicitly for variability improved model inference about the mean growth rate dynamics.

## **Results and Discussion**

The 4 random-effects GAMMLSS or GAMMLSS model fits for the 2 foraging-ground population samples are summarized in Table 1. Using leave-one-out cross validation and the LOO Information Criterion metric or LOOIC (Vehtari et al 2017), we found that there was no significant difference between the Bayesian GAMMLSS models fitted with either robust Student-*t* or Gaussian likelihood. So we report only the model summaries in Table 1 for those GAMMLSS models fitted with Student-*t* likelihood. Convergence diagnostics such as the effective posterior sample size and the Gelman-Rubin statistic ( $R_{hat} < 1.01$ ) reflected convergence of all Bayesian models (Gelman & Hill 2007).

The models fitted by either inferential procedure found similar results for the Kawainui population except for the predicted variance of the mean year effect (Table 1). The frequentist GAMMLSS model with Student-*t* likelihood found that the year effect variability decreased from around the mid-2000s onward (Fig 1e) whereas the

Bayesian model found no significant variance effect for sampling year (Fig 2d, Table 1). Model fit diagnostics based on visual inspection of residual plots (Fig 3) and worm plots conditioned for example on the mean carapace size predictor (Fig 4) suggest that the frequentist model was an adequate fit to the Kawainui growth rate data. Similarly, posterior predictive check tests such as key parameter summaries (Fig 5) suggest that the Bayesian model was also an adequate fit to the Kawainui growth rate data. The trends for all parameters in either model were similar but the Bayesian model estimates mean and variance components for the GAMMLSS model with much greater uncertainty than the frequentist model (compare expected curves in the panels of Figs 1 and 2).

It is apparent using either inferential procedure that there was a significant decrease in growth rates for the Kawainui population attributable to severe fibropapillomatosis disease affliction (Figs 1a, 2a) and that both the mean (Figs 1c, 2e) and variance of the growth rates (Figs 1f, 2f) were linear functions of the growth increment duration (or the recapture duration) — with somatic growth rate variance decreasing with increasing increment duration due to many of the increments being for quite short intervals. Interestingly, while there appeared to be declining mean growth rates during the late 2000s (Figs 1b, 2c), this effect was not found to be a significant growth rate predictor using either procedure (Table 1). Similarly for the expected or mean size-specific effect for the Kawainui population, which appeared to suggest a declining functional form (Figs 1d, 2b) but this effect was not significant using either procedure (Table 1).



A Bayesian GAMM model that does not account explicitly for variance in growth rate was also fit to the Kawainui data (Table 1). But the GAMMLSS that accounts for growth variability was an improved fit compared to this mean-response-only model (GAMMLSS LOOIC =  $445.9 \pm 22.8$ ; GAMM LOOIC =  $474.5 \pm 23.1$ ). None of the covariates were found to be significant predictors based on the GAMM (Table 1), even though both GAMLSSs found that affliction with severe fibropapilloma disease was a significant predictor of mean growth and that recapture duration (or duration of growth increment) was a significant predictor of both growth rate mean and variance (Table 1).

On the other hand, the 2 inferential procedures resulted in quite different conclusions about significant predictors of somatic growth rates at the Hanauma Bay study site (Table 1). The Bayesian GAMMLSS found that growth rate variability was not a significant function of mean sampling year (Fig 6d) or recapture interval (Fig 6f). In fact it was questionable whether either mean growth or growth variability was a significant function of any predictors included in this model (Fig 6, Table 1). Graphical posterior predictive check tests such as a density overlay plot (Fig 7) of many simulations (replications) of the fitted model compared to the density function of the observed growth rates suggest that the Bayesian GAMMLSS was an adequate fit to the Hanauma Bay growth rate data. A graphical posterior predictive check test of the predicted maximum growth rate also provided strong evidence that this model was an adequate fit to the Hanauma Bay data (Fig 8).

It is apparent then that the intercepts for the mean and variance components account for the modelled data with limited marginal effect due to any predictor, which is due

mainly to the small sample size for the Hanauma Bay population (72 growth increments for 31 individual immature green turtles). Given the limited sample and high parameter uncertainty (Fig 6), it was not considered worthwhile assessing the frequentist GAMMLSS fit. Moreover, a GAMM that does not account for variability was an improved fit to the Hanauma Bay data compared to the GAMMLSS model (GAMM LOOIC =  $178.3 \pm 12.1$ ; GAMMLSS LOOIC =  $199.5 \pm 20.5$ ). So while the frequentist GAMMLSS found 4 significant predictors (Table 1), this finding was not supported by the models fitted using Bayesian procedures — except perhaps an apparent decline in mean growth rates over the past 14-15 years (Fig 6c, Table 1) but no such trend in the variance (Fig 6d, Table 1).

### **Literature Cited**

Balazs G, Chaloupka M (2004) Spatial and temporal variability in somatic growth of green sea turtles resident within the Hawaiian Archipelago. *Marine Biology* 145: 1043-1059

Balazs G, Van Houtan, K, Hargrove S, Brunson S, Murakawa, S. (2015) A review of the demographic features of Hawaiian green turtles (*Chelonia mydas*). *Chelonian Conservation and Biology* 14(2) 119-129.

Bjorndal K, Chaloupka M, Saba V, Diez C, van Dam R, + 69 Coauthors (2016) Somatic growth dynamics of West Atlantic hawksbill sea turtles: a spatio-temporal perspective. *Ecosphere* 7(5): e01279. [10.1002/ecs2.1279](https://doi.org/10.1002/ecs2.1279)

Bjorndal K, Bolten A, Chaloupka M, Saba V, Bellini C, +69 Coauthors (2017) Ecological regime shift drives declining growth rates of sea turtles throughout the West Atlantic. *Global Change Biology* DOI 10.1111/gcb.13712.

Broderick (2015) Grand challenges in marine conservation and sustainable use. *Frontiers in Marine Science* 2:1-3.

Bürkner P (in press) brms: An R Package for Bayesian multilevel models using Stan. *Journal of Statistical Software*

van Buuren S, Fredriks M (2001) Worm plot: a simple diagnostic device for modelling growth reference curves. *Statistics in Medicine* 20: 1259–1277

Carpenter B, Gelman A, Hoffman M, Lee D, Goodrich B, Betancourt M, Brubaker M, Guo J, Li P, Riddell A (2017) Stan: A Probabilistic Programming Language. *Journal of Statistical Software* 76(1): 1-32

Chaloupka M, Balazs G (2005) Modelling the effect of fibropapilloma disease on the somatic growth dynamics of Hawaiian green sea turtles. *Marine Biology* 147: 1251-1260

Chaloupka M, Limpus C, Miller J (2004) Green turtle somatic growth dynamics in a spatially disjunct Great Barrier Reef metapopulation. *Coral Reefs* 23: 325-335

DeCarlo T, Cohen A, Barkley H, Quinn Cobban Q, Young C, Shamberger K, Brainard R, Golbuu Y (2015) Coral macrobioerosion is accelerated by ocean acidification and nutrients. *Geology* 43: 7–10

Dmitriew C (2011) The evolution of growth trajectories: what limits growth rate? *Biological Reviews* 86: 97–116

Dunn P, Smyth G (1996) Randomised quantile residuals. *Journal of Computation and Graphical Statistics* 5: 236–244

Dutton P, Balazs G, LeRoux R, Murakawa S, Zarate P, Martinez L (2008) Composition of Hawaiian green turtle foraging aggregations: mtDNA evidence for a distinct regional population. *Endangered Species Research* 5:37-44

Francke D, Hargrove S, Vetter E, Winn C, Balazs G, Hyrenbach D (2013) Behavior of juvenile green turtles in a coastal neritic habitat: Validating time–depth–temperature records using visual observations. *Journal of Experimental Marine Biology and Ecology* 444: 55–65

Gabry J (2016) bayesplot: Plotting for Bayesian Models. R package version 1.1.0. <https://CRAN.R-project.org/package=bayesplot>

Gelman A, Hill J (2007) Data analysis using regression and multilevel/hierarchical models. Cambridge University Press, New York, USA

Gelman A, Hwang J, Vehtari A (2014) Understanding predictive information criteria for Bayesian models. *Statistics and Computing* 24: 997–1016

Gelman A, Jakulin A, Pittau M, Su Y (2008) A weakly informative default prior distribution for logistic and other regression models. *Annals of Applied Statistics* 2: 1360–1383

Gilman E, Chaloupka M, Peschon J, Ellgen S (2016) Risk factors for seabird bycatch in a pelagic longline tuna fishery. *PLoS One* 11(5): e0155477. doi:10.1371/journal.pone.0155477

Hudson I, Susan W. Kim S, Marie R. Keatley M (2010) Climatic influences on the flowering phenology of four Eucalypts: a GAMLSS approach. In: Hudson I, Keatley M (eds) *Phenological Research: methods for environmental and climate change analysis*. Springer Netherlands, p 209-228

Kiffner C, Lödige C, Alings M, Vor T, Rühle F (2011) Body-mass or sex-biased tick parasitism in roe deer (*Capreolus capreolus*)? A GAMLSS approach. *Medical and Veterinary Entomology* 25: 39-45

Klein N, Kneib T, Lang S, Sohn A (2015) Bayesian structured additive distributional regression with an application to regional income inequality in Germany. *Annals of Applied Statistics* 9: 1024-1052

Kondo S, Morimoto Y, Sato T, Suganuma H (2017) Factors affecting the long-term population dynamics of green turtles (*Chelonia mydas*) in Ogasawara, Japan. *Chelonian Conservation and Biology* 16(1): 83-92.

Kneib T (2013) Beyond mean regression. *Statistical Modelling* 13(4): 275–303

Krueger B, Chaloupka M, Leighton P, Dunn J, Horrocks J (2011) Somatic growth rates for a hawksbill turtle population in coral reef habitat around Barbados. *Marine Ecology Progress Series* 432: 269-276

Kubis S, Chaloupka M, Ehrhart L, Bresette M (2009) Growth rates of juvenile green turtles *Chelonia mydas* from three ecologically distinct foraging habitats along the east central coast of Florida. *Marine Ecology Progress Series* 389: 257-269

Park T, Casella G (2008) The Bayesian lasso. *Journal of the American Statistical Association* 103: 681-686

Stan Development Team (2016) The Stan C++ Library, version 2.10.0. <http://mc-stan.org/>

Stasinopoulos D, Rigby R (2007) Generalized additive models for location scale and shape (GAMLSS) in R. *Journal of Statistical Software* 23(7): 1–46

Vehtari A, Gelman A, Gabry J (2017) Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing* doi: 10.1007/s11222-016-9696-4

Voudouris V, Gilchrist R, Rigby R, Sedgwick J, Dimitrios S (2012) Modelling skewness and kurtosis with the BCPE density in GAMLSS. *Journal of Applied Statistics* 39: 1279-1293

Wabnitz CCC, Balazs G, Beavers S, Bjorndal KA and others (2010) Ecosystem structure and processes at Kaloko Honokōhau, focusing on the role of herbivores, including the green sea turtle *Chelonia mydas*, in reef resilience. *Mar Ecol Prog Ser* 420:27-44. <https://doi.org/10.3354/meps08846>

Wickham H (2016) *ggplot2: Elegant Graphics for Data Analysis*. 2nd Edition. Springer-Verlag, New York, USA

Wood S (2006) *Generalized additive models: an introduction with R*. Chapman and Hall/CRC Boca Raton, Florida, USA

**Table 1:** Model-specific summaries of parameter significance determined by either frequentist  $P < 0.05$  or where the Bayesian 95% credible interval doesn't include 0.

Study site	GAMMLSS model:		GAMM model:
	frequentist	Bayesian	Bayesian

Kawainui

*Mean terms:*

FP moderate	no	no	no
FP severe	yes	marginal	no
mean year	no	no	no
mean size	no	no	no
recapture duration	yes	marginal	no

*variance terms:*

mean year	yes	no
recapture duration	yes	yes

Hanauma Bay

*Mean terms:*

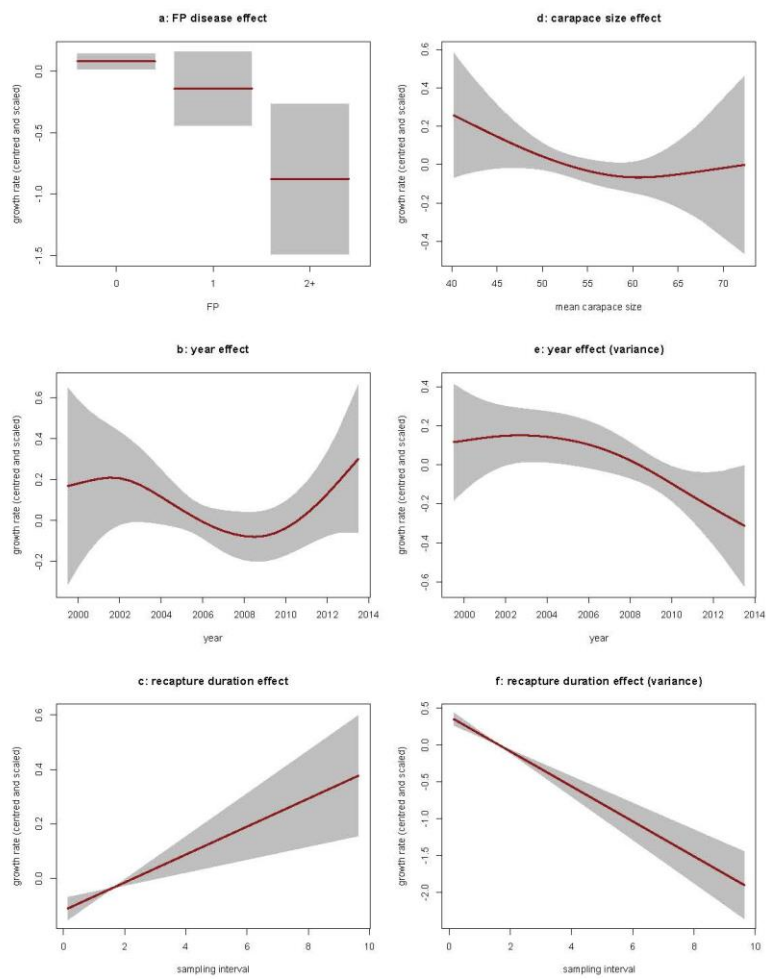
FP afflicted	no	no	no
mean year	yes	marginal	marginal
mean size	yes	no	no
recapture duration	yes	no	no

*variance terms:*

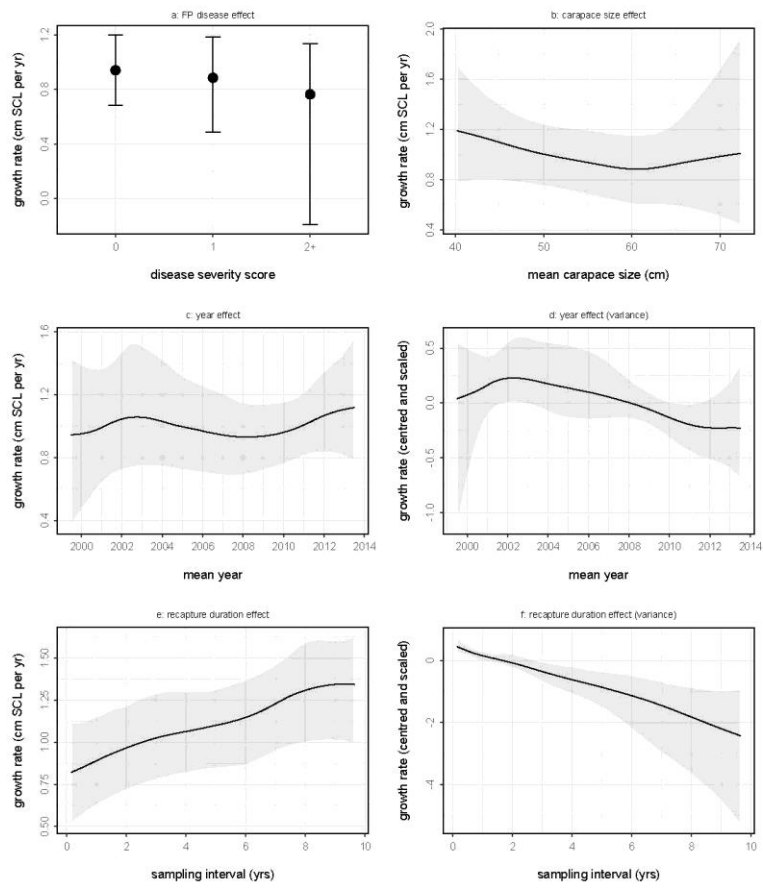
mean year	no	no
recapture duration	yes	no

---

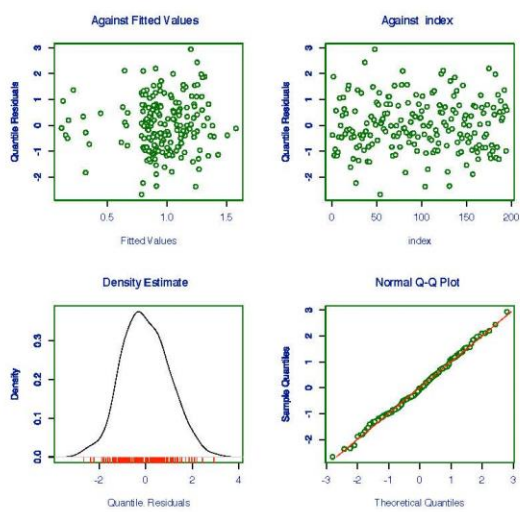




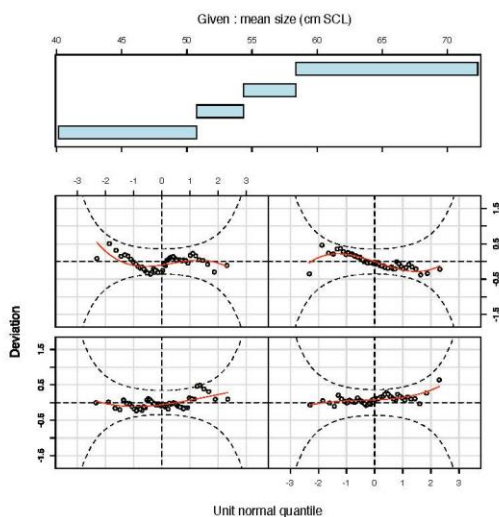
**Figure 1** Kawainui. Summary of a frequentist random-effects GAMLSS regression model with Student-*t* likelihood fitted to somatic growth rates with 4 mean predictors and 2 variance predictors. Expected response shown by solid curve, 95% confidence intervals around curves or factor levels shown by gray polygon. Panels (e and f) show the estimated variance effects, the other panels show the estimated mean effects.



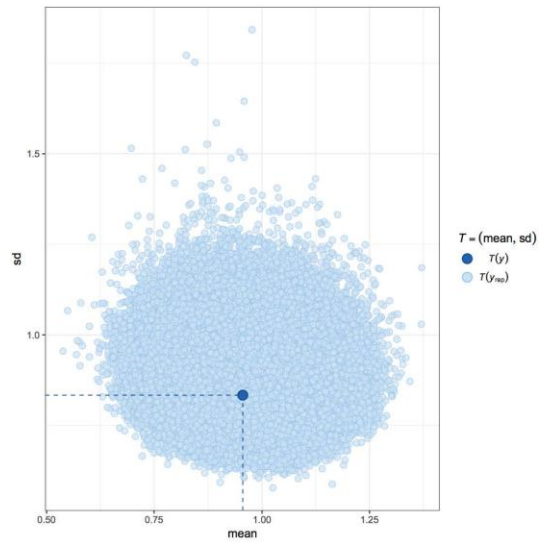
**Figure 2** Kawainui. Summary of a Bayesian random-effects GAMLSS model with Student-*t* likelihood fitted to somatic growth rates with 4 mean predictors and 2 variance predictors. Expected response shown by solid curve or dot, 95% credible or uncertainty intervals around the curves or factor levels shown by gray polygon or bar. Panels (d and f) show the estimated variance effects, the other panels show the estimated mean effects.



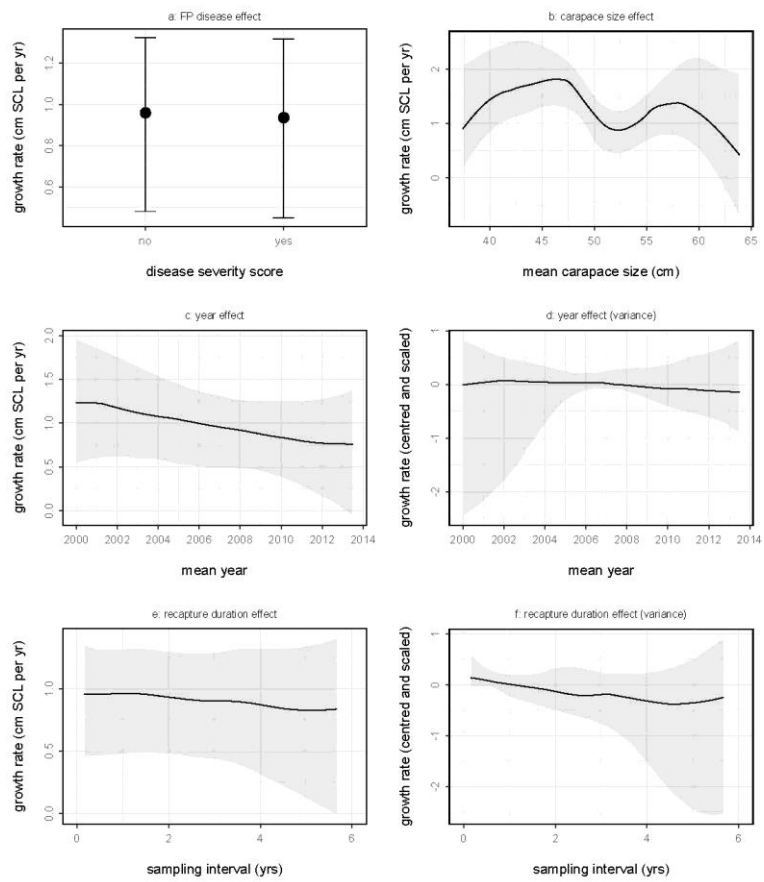
**Figure 3** Kawainui. Residuals summary plots for the model fit summarized in Figure 1.



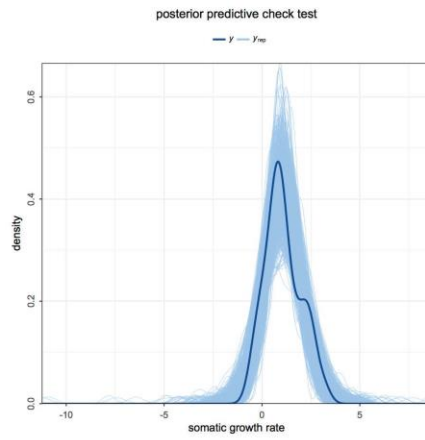
**Figure 4** Kawainui. Worm plot for model fit summarized in Figure 1 and conditioned on 1 of the predictors (mean carapace size: Fig 1d). All residuals fit within the 95% confidence ellipses and generally fluctuate around the zero horizontal centre line for each of the 4 segments of the predictor (so residual behavior is not a function of any level of the predictor)



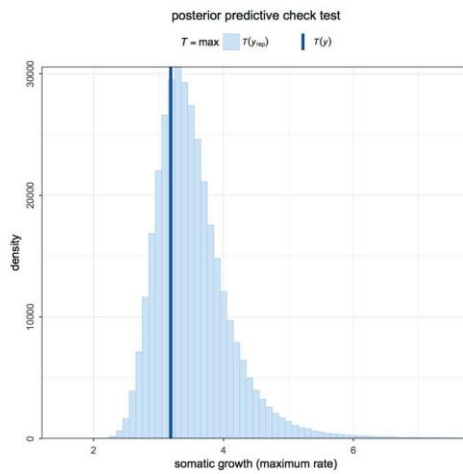
**Figure 5** Kawainui. Posterior predictive check of the 2 key summary parameters (the mean and standard deviation) for the model fit summarized in Figure 4. Solid blue dot shows the summary parameters for the observed data and the mass of blue dots shows the summary estimates for each of the 50,000 simulations of the model fit shown in Figure 4. So the fitted model makes realistic predictions of these summary parameters.



**Figure 6** Hanauma Bay. Summary of a Bayesian random-effects GAMLSS with Student-*t* likelihood fitted to somatic growth rates with 4 mean predictors and 2 variance predictors. Expected response shown by solid curve or dot, 95% credible or uncertainty intervals around curves or factor levels shown by gray polygon or bar. Panels (d and f) show the estimated variance effects, the other panels show the estimated mean effects.



**Figure 7** Hanauma Bay. Posterior predictive check of the response variable for the model fit summarized in Figure 6. Solid blue line shows the density curve summarizing the observed data and the mass of blue lines shows the density curve overlay for each of the 50,000 simulations of the fitted GAMLSS model shown in Figure 6. So the fitted model makes realistic predictions of this parameter.



**Figure 8** Hanauma Bay. Posterior predictive check of the maximum response variable for the model fit summarized in Figure 6. Solid blue line shows the maximum growth rate of the observed data and the light blue histogram summarizes the 50,000 simulations of the expected maximum for the fitted GAMLSS model shown in Figure 6. So the fitted model makes realistic predictions of this parameter.