MIKE analysis for Africa - Technical

Carl Schwarz

2020-10-29

Contents

1 Introduction

This document estimates yearly-trends in the Proportion of Illegally Killed Elephants (PIKE) from MIKE (Monitoring Illegally Killed Elephants) monitoring sites in Africa since 2003.

Briefly, MIKE data is collected on an annual basis in designated MIKE sites by law enforcement and ranger patrols in the field and through other means. When an elephant carcass is found, site personnel try to establish the cause of death and other details, such as sex and age of the animal, status of ivory, and stage of decomposition of the carcass. This information is recorded in standardized carcass forms, details of which are then submitted to the MIKE Programme. As expected, different sites report widely different numbers of carcasses, as encountered carcass numbers are a function of: population abundance; natural mortality rates; the detection probabilities of elephant carcasses in different habitats; differential carcass decay rates; levels

of illegal killing; and levels of search effort and site coverage. Because of these features of the survey data, the number of carcasses found is unlikely to be proportional to the total mortality and trends in observed numbers of illegally killed elephants may not be informative of the underlying trends.

Consequently, the observed proportion of illegally killed elephants (PIKE) as an index of poaching levels has been used in the MIKE analysis in an attempt to account for differences in patrol effort between sites and over time:

$$PIKE_{sy} = \frac{Number\ of\ illegally\ killed\ elephants_{sy}}{Total\ Carcasses\ Examined_{sy}}$$

where the subscripts *sy* refer to site and year respectively.

Computing a continent-wide PIKE is challenging for several reasons, including as mentioned above:

- Detection probabilities of elephant carcasses in various habitats differ.
- Levels of search effort and site coverage differ between sites.
- Not all sites report in all years.
- Number of carcasses in both categories varies considerably across space and time.

1.1 The LSMeans approach

In past years, a simple linear model on the PIKE values was computed as described in https://cites.org/sites/default/files/notif/E-Notif-2019-046.pdf. This is denoted the *LSMeans* approach.

The PIKE trend is calculated using estimated marginal means of a linear model weighted by the total number of observed carcasses. The continental PIKE trend is estimated based on a model with subregion and year as factors, while the subregional trends are estimated from a model using country and year as factors.

The statistical model at the continental level in an *R* code syntax is:

$$lm(pike \sim SubRegion + YearC, data = ..., weight = TotalCarcasses)$$

and at the subregional level the model, for each subregion, is

$$lm(pike \sim Country + YearC, data = ..., weight = TotalCarcasses)$$

where, in both cases, *YearC* is a categorical year effect.

The *weight=TotalCarcasses* gives each examined carcass equal weight and essentially pools data to the sub-region/country-year combination. For example, suppose at the subregional level, a particular country in a year had 3 sites with the following *PIKE* and *TotalCarcasses* examined:

Site	PIKE	TotalCarcasses
A	.1	10
B	.2	100
C	.3	50

Then the *PIKE* computed for this country in this year is found as

$$PIKE_{country\ year} = \frac{10(.1) + 100(.2) + 50(.3)}{10 + 100 + 50} = \frac{36}{160} = 0.24$$

which is not the simple arithmetic mean of the site *PIKE* values but is found as total illegally killed carcasses / total carcasses examined in that country-year combination. Similarly at the continental level, all sites in a sub-region are pooled together for subregion-year combination (and country is ignored).

At the subregional level, the model estimates *PIKE* for each country and year, imputing missing values for particular country-year based on the relationship in *PIKE* among countries in other years. For example, if

country 1 tends to have a similar *PIKE* to another country, then the *PIKE* is imputed in missing years based on the *PIKE* in the other countries. At the continental level, any missing value at the sub-regional level is also imputed.

This imputation is done statistically and has a sound mathematical basis. Once the *PIKE* is estimated for every country-year combination, the mean *PIKE* over all countries in a year is found as the mean of the estimated *PIKE* for all countries in year (the *least square mean* or *expected marginal mean*) for each sub-region. Each country is given the same weight in the computation of the marginal mean *PIKE* in a year at the sub-regional level. Similarly, once the *PIKE* is estimated for each sub-region year combination, the continental-level *PIKE* is estimated as the simple average of the sub-regional *PIKE*. However, because different models (and different weightings) are used for the two scales, the results may not be consistent, i.e. the mean of the individual sub-regional *PIKE* from the sub-regional models may not match the continental *PIKE* estimated from the continental model.

As indicated in the MIKE report to the 18th meeting of the Conference of Parties to CITES available at https://cites.org/sites/default/files/eng/cop/18/doc/E-CoP18-069-02.pdf, the CITES Secretariat, in collaboration with the MIKE-ETIS TAG statisticians and an independent statistician, initiated a process to review the MIKE analytical methodology to determine whether it could be refined or its scientific robustness improved, and to further enhance the analytical basis for MIKE. The approach included a review of the current methodology, and consideration of new statistical developments and, therefore, alternative methods or models for *PIKE* trend analysis, while taking into consideration the imbalances and inconsistencies inherent in the data. Burn, Underwood and Blanc (2011) used a Bayesian hierarchical model based on a generalized linear mixed model (logistic regression with random effects) to correct for many of these issues.

1.2   Why change from the LSMeans approach

The *LSMeans* approach has been for the *PIKE* trend analysis in the reports to the two previous meetings of the Conference of the Parties (CoP16, Bangkok, 2013 in document CoP16 Doc. 53.1 and CoP17; Johannesburg, 2016, in document CoP17 Doc. 57.5), and to meetings of the Standing Committee (SC62, Geneva, July 2012, in document SC62 Doc. 46.1 (Rev. 1); SC65, Geneva, July 2014, in document SC65 Doc. 42.1; SC66, Geneva, January 2016, in document SC66 Doc. 47.1; SC69, Geneva, November 2017, in document SC69 Doc. 51.1; and SC70, Sochi, October 2018, in document SC70 Doc. 49.1). However, there are a number of issues with this approach

- Predicted *PIKE* values could be less than 0 or greater than 1
- Each country is given equal weight regardless of the number of *MIKE* sites or the abundance of elephants in each site at the sub-regional level. For example, in Eastern Africa, both Kenya (4 *MIKE* sites) and Eritrea (1 *MIKE* site) are given equal weight in computing the subregional PIKE. But at the continental level, countries are ignored and data pooled over all sites in a sub-region.
- The *LSMeans* is not consistent in how to aggregate to larger levels. The mean of the *LSMeans* estimates of *PIKE* at the sub-regional level from the individual sub-regional models, will not match the estimated *PIKE* at the continental level estimated using the continental model because country is or is not included in the two models.
- It is difficult to apply different weighting, e.g. by elephant populations at the *MIKE* or country level.
- The *LSMeans* approach imputes *PIKE* for missing country-year combinations at the sub-regional level, but the individual *MIKE* sites could differ in the patterns of missingness across years. Consequently, the country may appear to have data for every year, but the country-wide aggregate data is based on a different combination of *MIKE* sites across the years. A similar problem occurs in the analysis at the continental level.
- Not all sources of variation are included in estimates of uncertainty. For example, binomial variation (Section 5.2) at the site-year level is not included, i.e. if the actual *PIKE* of all elephants at the *MIKE* site is .20, then the observed *PIKE* in the sample of carcasses examined will vary around 0.20.

The impact of giving equal weight at the country level or at the *MIKE* site level is explored in more detail in Appendix 3 where this has a noticeable impact on the sub-regional *PIKE* for East Africa.

At the recent MIKE-ETIS TAG meeting (September 2019, Nairobi), the use of the generalized linear mixed

model (GLMM) was recommended going forward. This document provides an analysis of the *PIKE* data using a GLMM, compares the results to those from the previous analyses, and explores the impacts of various assumptions on the estimated *PIKE*.

The advantages of the GLMM approach are:

- The new model fully accounts for the binomial structure at the site-year level, i.e. of n carcasses observed, x are illegally killed.
- The new model fully accounts for different sample sizes, i.e. a *PIKE* based on observing 1 illegally killed elephants out of 2 elephant carcasses is given different weight than a *PIKE* based on observing 20 illegally killed elephants out of 40 elephant carcasses.
- Imputation of missing data takes place at the MIKE site level based on the relationship of *PIKE* in this site to the *PIKE* at other sites across time.
- Each *MIKE* site is given equal weight when computing the continental or sub-regional *PIKE*. Consequently, countries with more *MIKE* sites will automatically be given more weight in the aggregate *PIKE* estimates.
- It is easy to apply other weightings, e.g. by the elephant population abundances at each *MIKE* site when computing an aggregate *PIKE*
- Multiple sources of variation are automatically included, e.g. binomial variation at the site-year level, site-year interactions representing the changes in *PIKE* over time at a particular *MIKE* site, and site-to-site variation. This variation is automatically included in the uncertainty of the aggregate *PIKE* estimates.
- Uncertainty of the estimated *PIKE* at the continental or sub-regional level can be computed assuming that the current *MIKE* sites are index sites (and fixed) or random sample of potential *MIKE* sites. The current design is somewhat between these two extremes, and so the uncertainty reported under these two ways of viewing the current set of *MIKE* sites represents a lower and upper bound of uncertainty.
- It is easy to include estimates of uncertainty in elephant abundances when weighting *MIKE* sites by elephant abundances. This is currently not yet done because of issues in determining the uncertainty in these abundance estimates (see Section 8.7).
- The current model implicitly accounts for spatial autocorrelation in the *PIKE* among *MIKE* sites that are geographically close. The estimated site-level effects are similar for sites that are geographically close and have similar levels of governance and poaching.
- It will be possible to extend this model to account explicitly for spatial and spatial-temporal autocorrelation (Zuur, 2019)

1.3   *PIKE* as an index of poaching pressure

The value of *PIKE* computed by *LSMeans* and the *GLMM* approach should be considered as an INDEX of poaching pressure. We hope that trends in the index reflect trends in the actual levels of poaching. Converting the value of *PIKE* into a measure of actual poaching mortality is complicated due to the following:

- Effects of changing natural mortality over time are confounded with changes in *PIKE*.
- MIKE sites are not selected at random and so the marginal mean *PIKE* may not be representative of the actual marginal mean *PIKE*.
- Poaching at MIKE sites may not be representative of poaching in areas outside of the MIKE site.
- The unweighted marginal mean *PIKE* gives equal weight to every MIKE site. We have investigated in this report the impact of weighting sites by the population abundance of elephants associated with that site, but have not included estimates of uncertainty in the population abundance estimates.
- There are a number of issues with the collection of the data such as:
  - Management related deaths are included in the number of carcasses examined (the denominator of *PIKE*) at this point in time;
  - search patterns are not random and often directed to illegally killed elephants;
  - misclassification of animals as unknown causes of mortality rather than illegally or natural mortality based on the preponderance of evidence;
  - errors in the data

The CITES MIKE CCU is preparing a discussion paper that explores data issues in more detail.

For these reasons, great care should to be taken and assumptions should be well documented when converting the estimated *PIKE* to actual levels of poaching mortality.

2 Exploration of PIKE data

2.1 MIKE sites with GIS data

There are 68 MIKE sites in Africa broken into 4 regions for which GIS data (shape files) of their boundary are available:

Africa: Location of MIKE sites
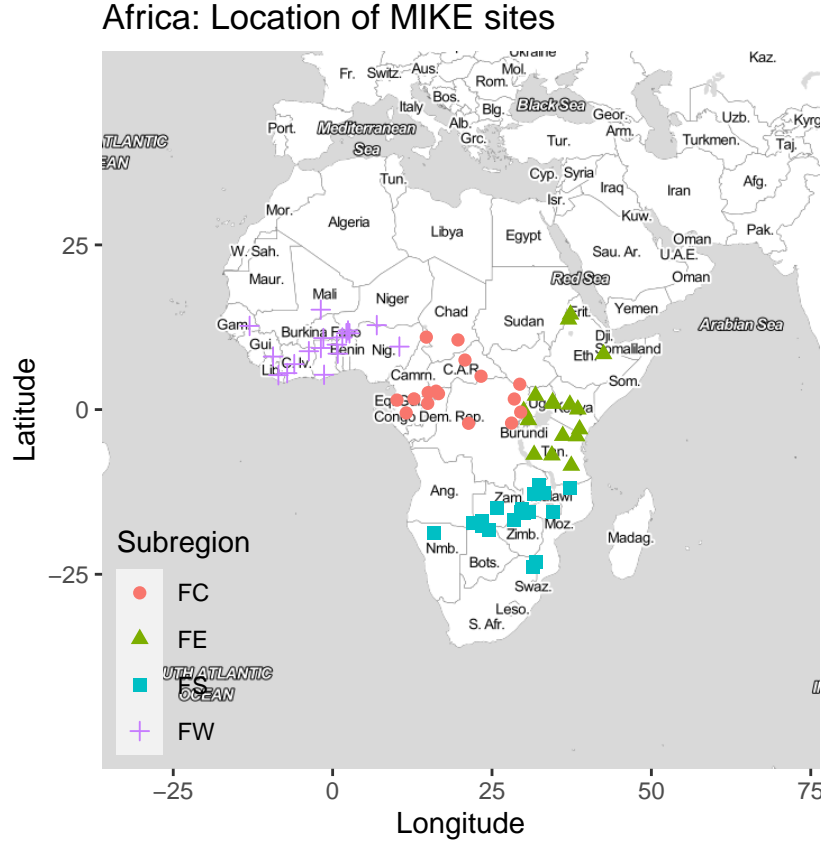
2.2 MIKE sites with *PIKE* data

There are 61 MIKE sites that have reported data on the number of carcasses found and the number of illegally killed carcasses among these. This includes 33 site-years where sites have reported 0 carcasses examined in a year.

The current analysis treats site that did not report on any carcasses in a year (no patrol effort) and a site that reports 0 carcasses examined in year (patrol effort but no carcasses found) in the same way. This is because information on patrol effort is not currently used in the analysis and only the number of carcasses examined and the number of illegally killed elephants in the sample of carcasses is used. In the latter case, 0 illegal carcasses out of 0 carcasses examined gives a *PIKE* for that site-year of 0/0 which is indeterminate and cannot be used in any mathematical analysis of *PIKE*.

The following plot shows that there are some sites that have reported data for at least one carcass in as little as one year, but other sites have reported data for at least one carcass in almost every year.

Africa: When is each site measured
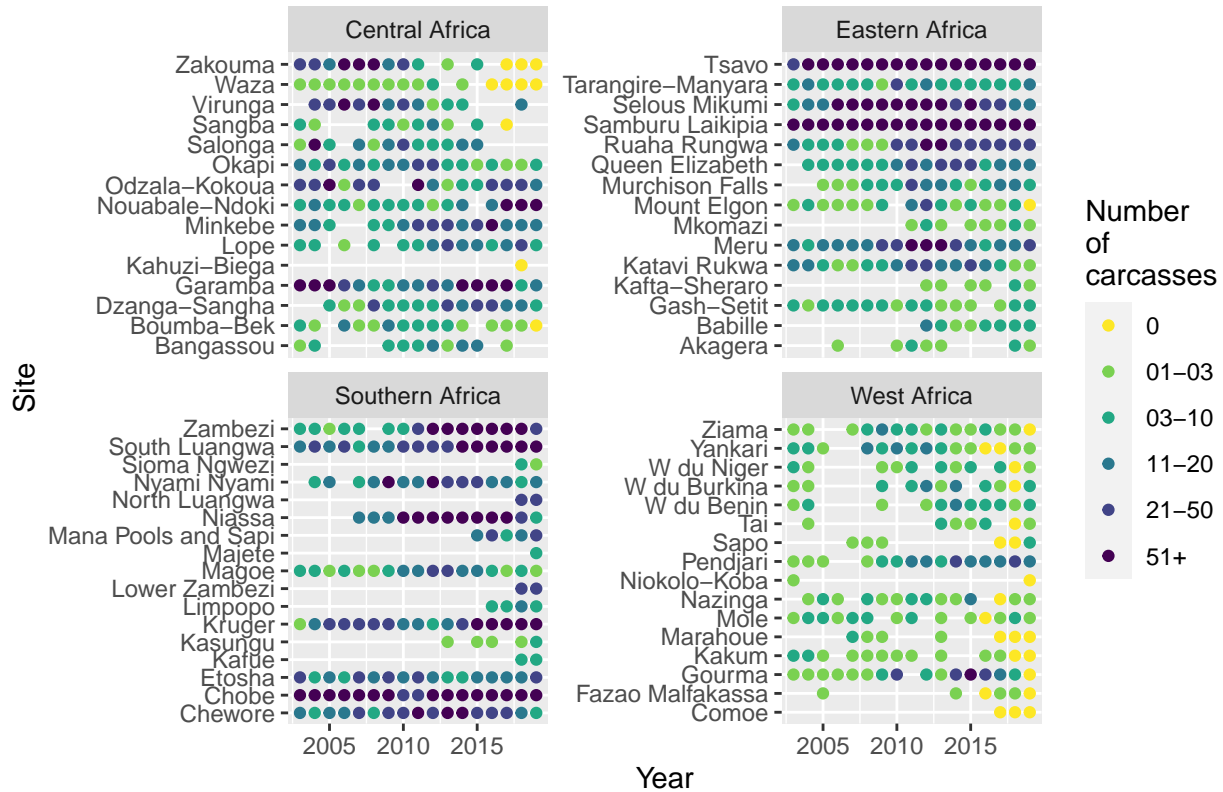
In total, there are 711 unique site-years in Africa since 2003 where data has been reported (and the number of reported carcasses > 0).

The number of carcasses reported in each site-year since 2003 varies enormously from 1 to 351 carcasses.

Africa: Carcasses observed
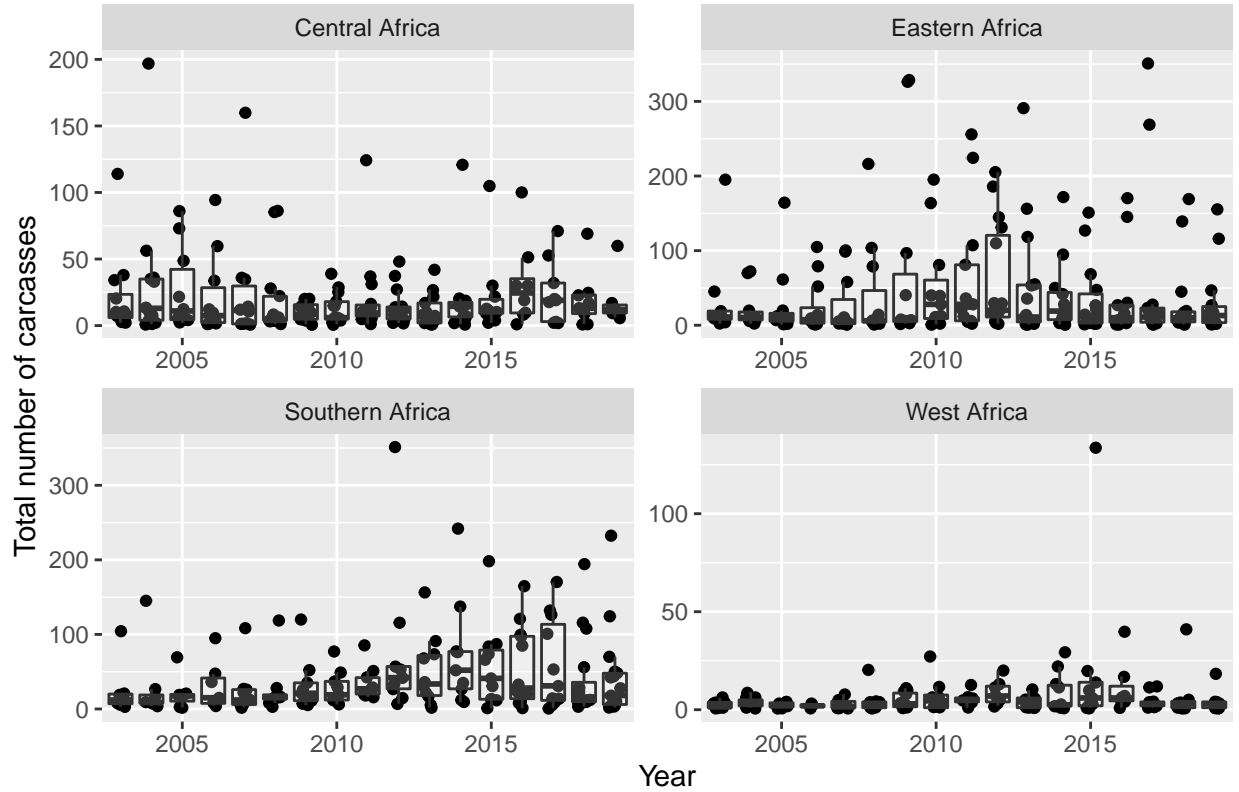
The unusual data point for West Africa where one site reported a large number of carcasses in one year is correct and corresponds to the MIKE site Gourma (GOU) with total number of carcasses equal to 134, of which 130 were poached by armed groups who entered the area.

2.3 Observed *PIKE*

The observed *PIKE* is the value computed from the examined carcasses in a year which we hope reflects the actual *PIKE* for all elephants at the MIKE site. A plot of the observed *PIKE* values from each site-year shows a wide range in the observed *PIKE* values, but many of the observed *PIKE* values close to 0 or 1 occur in sites with only a small number of carcasses examined in a year:

Africa: Observed PIKE values

Points jittered to prevent overplotting
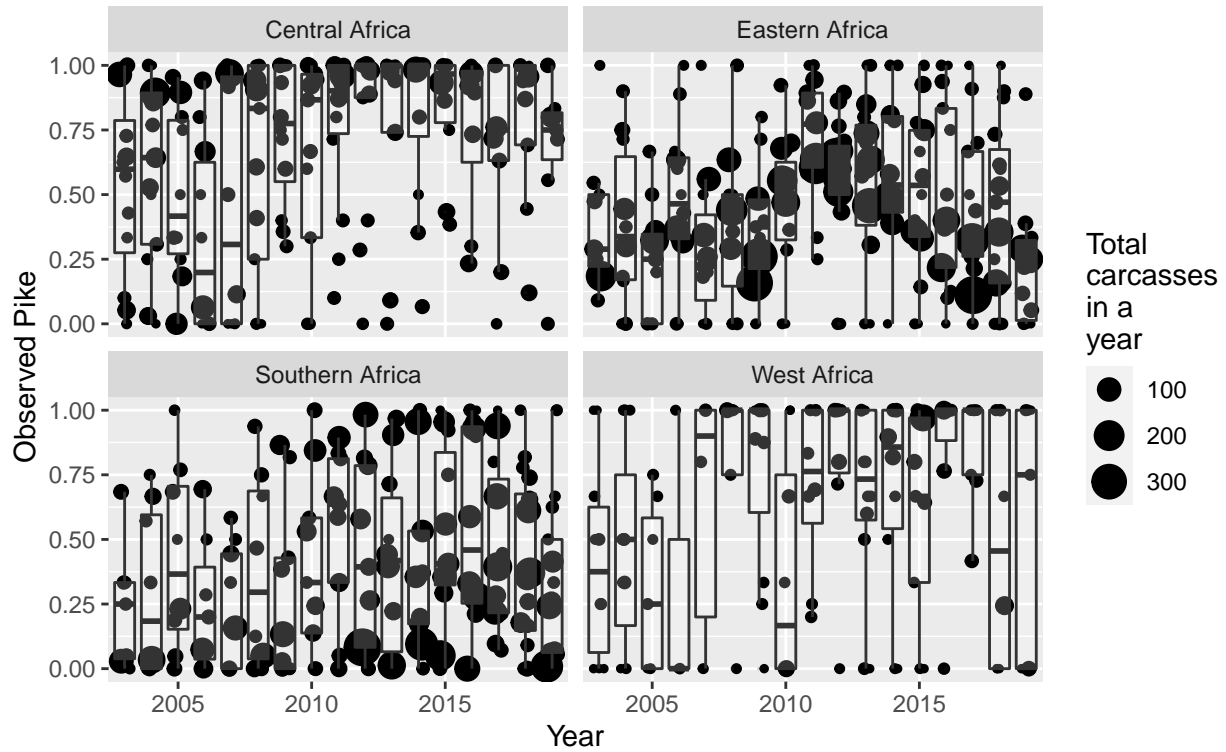
The trend in observed *PIKE* values for each site is:

Africa : Observed PIKE values for each site

Thicker/darker lines represent sites with more carcasses reported
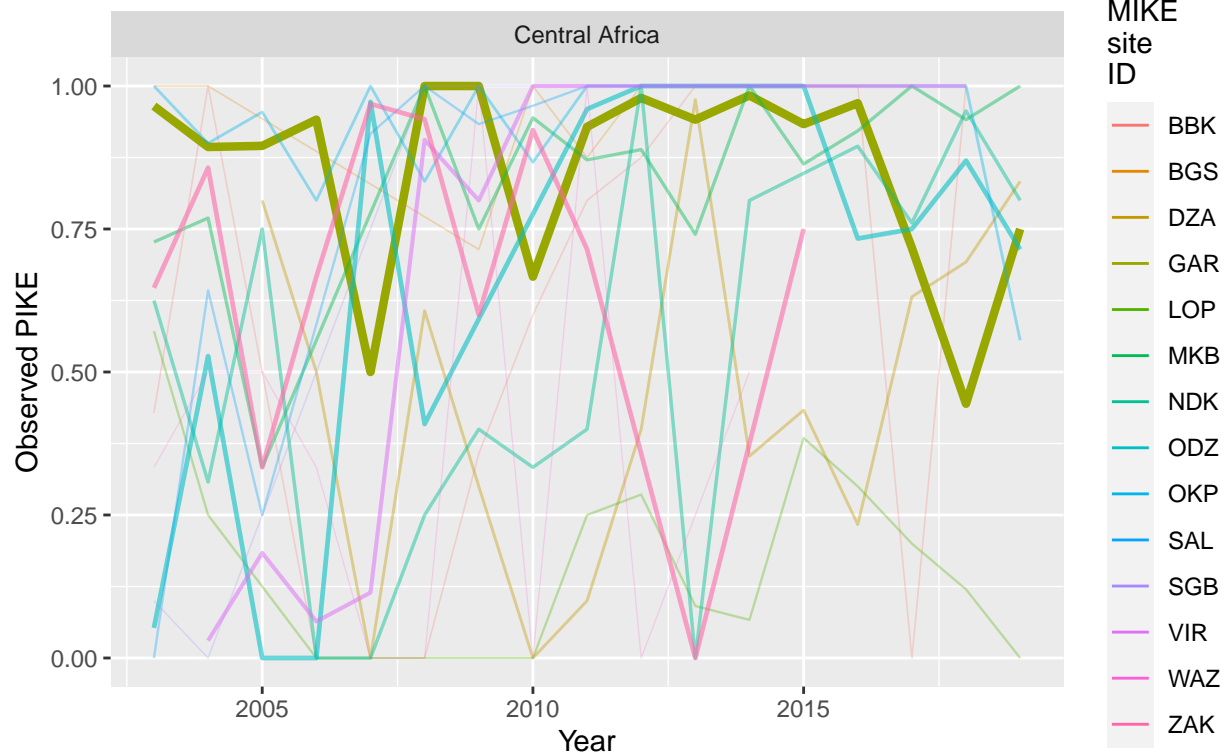
Africa : Observed PIKE values for each site

Thicker/darker lines represent sites with more carcasses reported
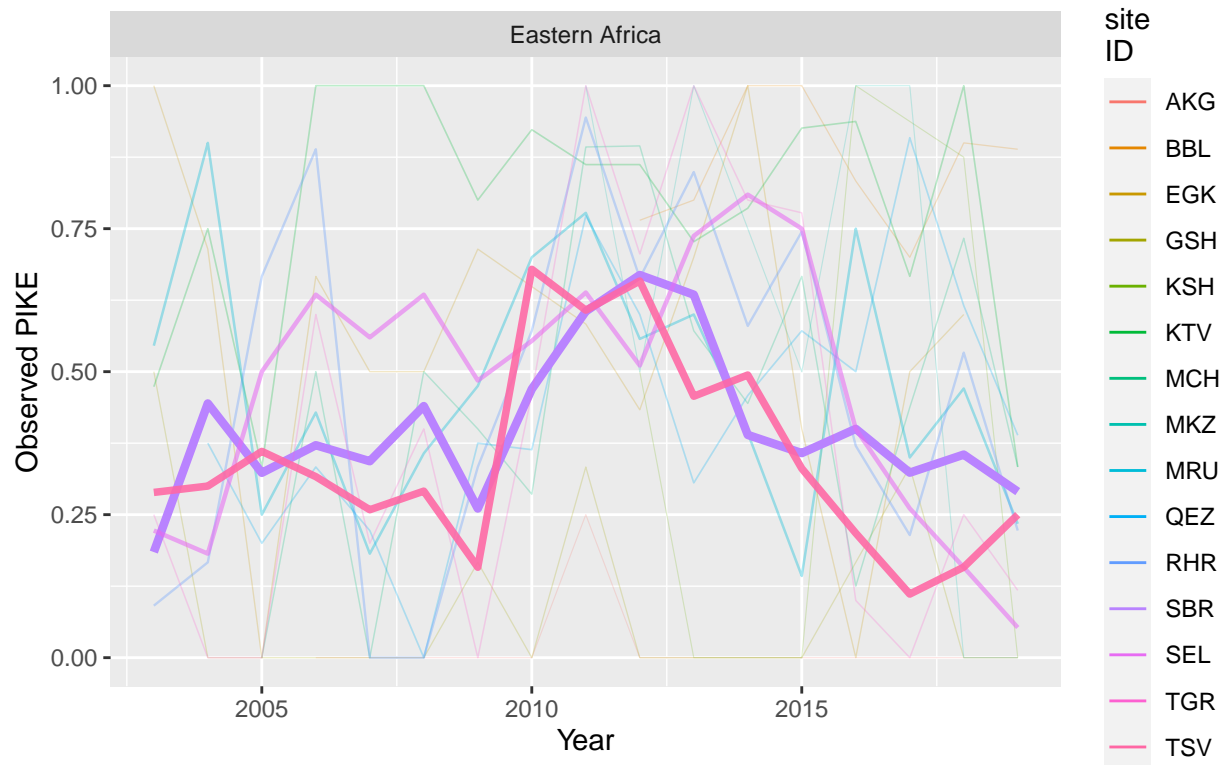
Africa : Observed PIKE values for each site
Thicker/darker lines represent sites with more carcasses reported
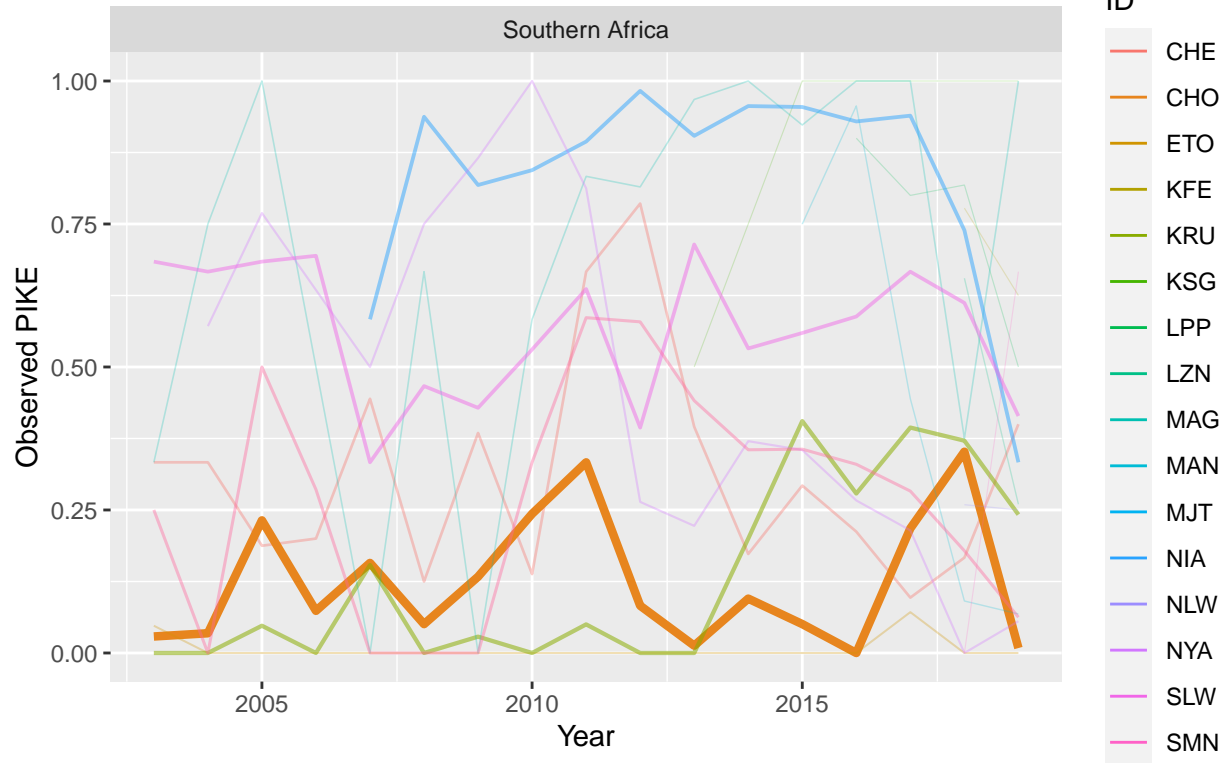
Africa : Observed PIKE values for each site
Thicker/darker lines represent sites with more carcasses reported
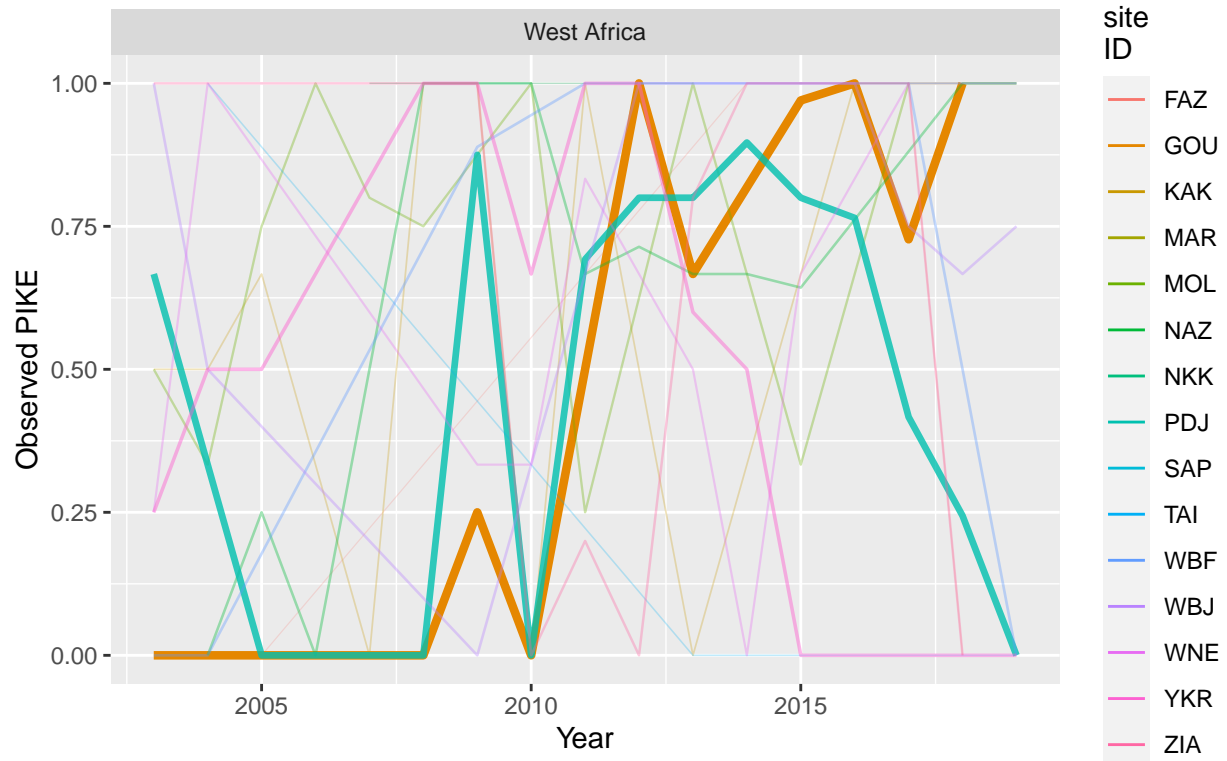
Note that with a small number of carcasses reported (e.g. 0 or 1) it is quite common for the reported *PIKE* to be 0 or 1 because either none or all of the carcasses have been illegally killed. Consequently, the trends are difficult to interpret for many sites with only a few carcasses reported.

2.4   MIKE sites with population data

Population data is available for 67 sites and has been extracted from Thouless et al (2016). Population surveys are not done each year and population numbers are not updated between population surveys. In these cases, for purposes of illustration, these missing values were imputed using the most recent abundance value. For example, if the estimated population abundances from a survey conducted in 2010 was 500 elephants and from a survey conducted in 2015 was 400 elephants, then the imputed population abundances for 2011, 2012, 2013, and 2014 is also 500 elephants.

For some sites, no population surveys have been done, and population estimates are "best guesses".

These population values are NOT the number of elephants in the MIKE site. Rather, the MIKE sites were chosen to be representative of a broader population and the population values here refer to this broader population. **The use of population values to compute a population-weighted PIKE is currently experimental and being evaluated and the actual implementation may change in the future.**

A plot of when an estimate of the population abundance is available by MIKE site is:
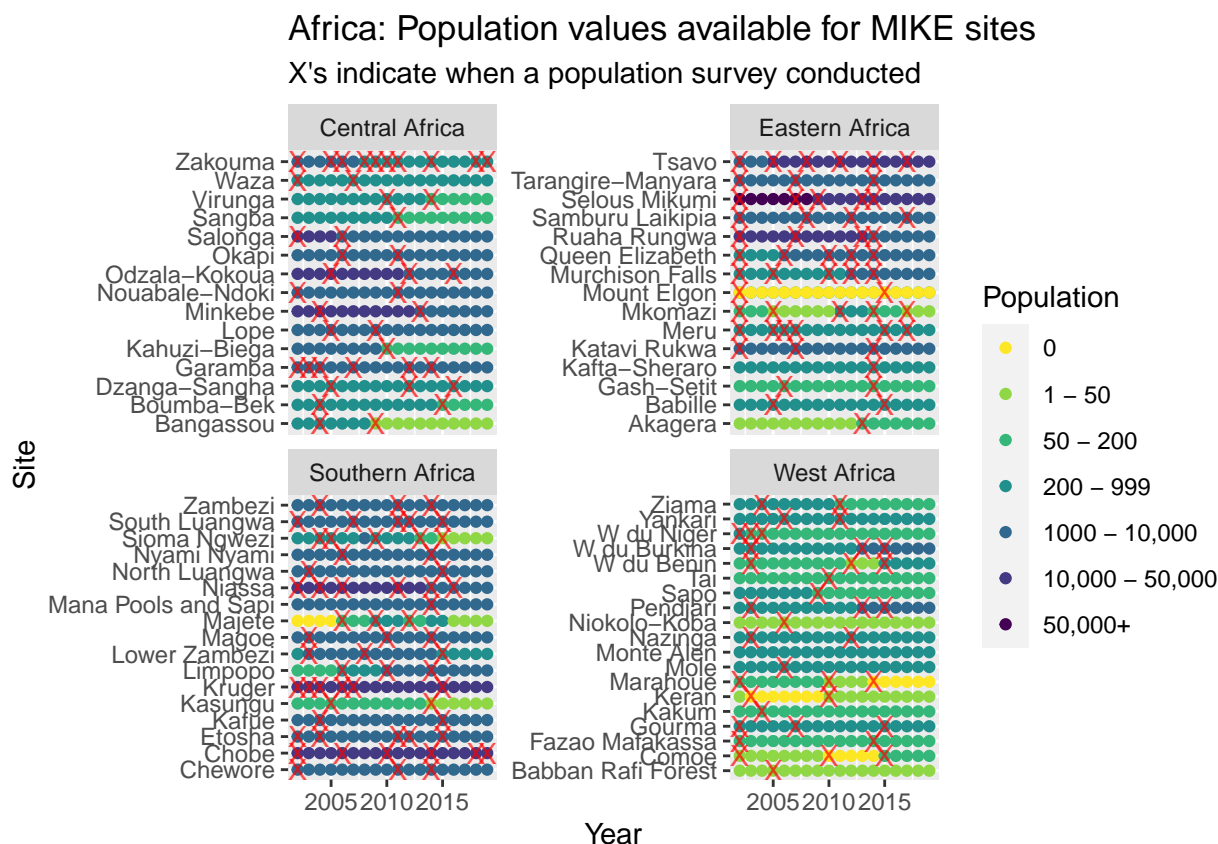
As a reminder, abundance surveys are NOT done in each year for each site. The sudden jumps between categories of abundance for a site may be an artefact of the imputation rule used between population surveys.

Patrol effort varies considerably among sites, so the number of examined carcasses only bears a weak relationship to estimated population abundances at the MIKE sites:

Africa : Relationship between number of reported carcasses and estimated population abundance
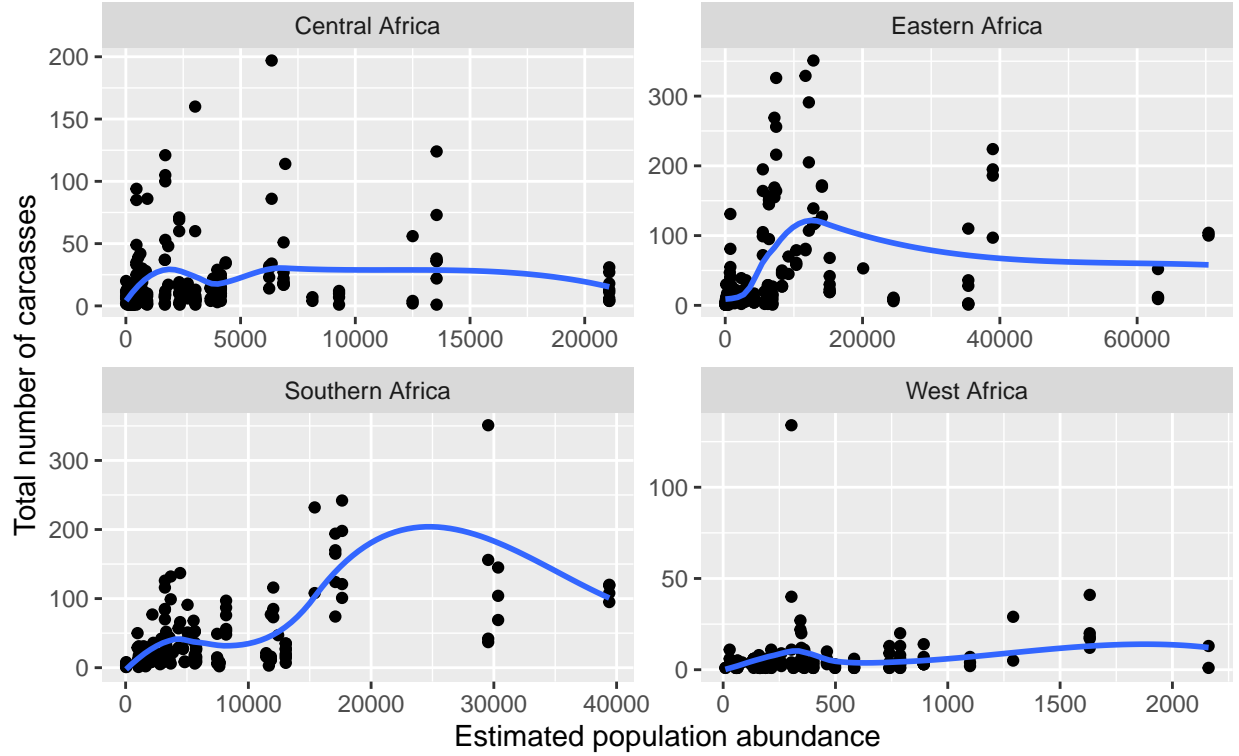
Consequently, the observed data on total number of carcasses found is unlikely to be a consistent fraction of the population abundance and aggregating the observed data on the number of illegally killed elephant carcasses may not properly reflect total poaching pressure Because the *PIKE* index does not depend on effort, it is thought to be a better indicator of poaching pressure than the observed number of carcasses of illegally killed elephants.

2.5   Final dataset used

The final dataset is an amalgamation of the MIKE population data and observed *PIKE* data. Only those sites for which population data is available and for which at least one carcass has been reported over the period of interest are used.

The following table identifies the MIKE sites not included in the analysis based on data from one of the data sources being absent:

Table 2: Summary of MIKE sites where data not present in all sources

MIKEsiteID	Population values available	PIKE reported	MIKE Center Available
ALE	TRUE	FALSE	TRUE
BBR	TRUE	FALSE	TRUE
COM	TRUE	FALSE	TRUE
EGU	TRUE	FALSE	TRUE
KER	TRUE	FALSE	TRUE
KHB	TRUE	FALSE	TRUE
LGL	FALSE	FALSE	TRUE

The final data set consists of 61 MIKE sites from 2003 to 2019 over the subregions as shown below:

Table 3: Summary of MIKE sites used in analysis

Subregion Name	Number of sites	# Site-Years	Mean # carcasses reported per year	Site IDs
Central Africa	14	188	20.4	BBK, BGS, DZA, GAR, LOP, MKB, NDK, ODZ, OKP, SAL, SGB, VIR, WAZ, ZAK
Eastern Africa	15	210	41.7	AKG, BBL, EGK, GSH, KSH, KTV, MCH, MKZ, MRU, QEZ, RHR, SBR, SEL, TGR, TSV
Southern Africa	17	169	42.8	CHE, CHO, ETO, KFE, KRU, KSG, LPP, LZN, MAG, MAN, MJT, NIA, NLW, NYA, SLW, SMN, ZBZ
West Africa	15	144	6.2	FAZ, GOU, KAK, MAR, MOL, NAZ, NKK, PDJ, SAP, TAI, WBF, WBJ, WNE, YKR, ZIA

3   Site, year, and site-year effects (a variance decomposition)

PIKE values vary spatially and temporally. A variance decomposition of the observed *PIKE* is helpful to understand the basis of the GLMM. Here is a (hypothetical) plot of *PIKE* for two sites in three years along with the yearly mean *PIKE* over both sites.

Hypothetical PIKE to illustrate variance decomposition
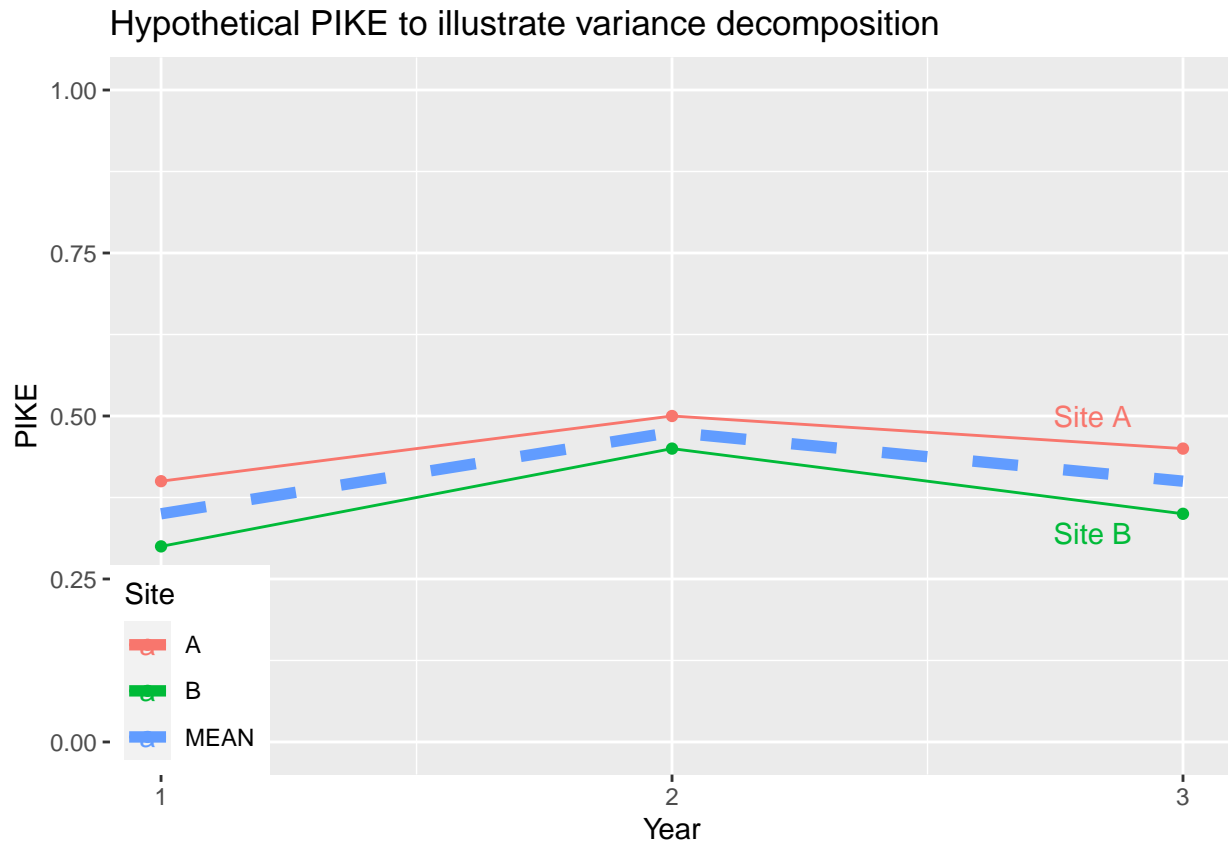
There are several sources of variation:

- temporal effects. The mean *PIKE* varies over time increasing in year 2 and then decreasing in year 3.
- site effects. Site *A* tends to consistently have a higher *PIKE* than site *B*.
- site-year effects. Even though Site *A* tends to have a larger *PIKE* than Site *B*, the difference is not consistent, i.e., the trend lines for Site *A* are not strictly parallel to the trend line for Site *B*.

The latter two effects can be more clearly seen if the yearly trend is subtracted from each value:

Hypothetical PIKE to illustrate variance decomposition

After removing yearly trend
with mean for each site over time also shown
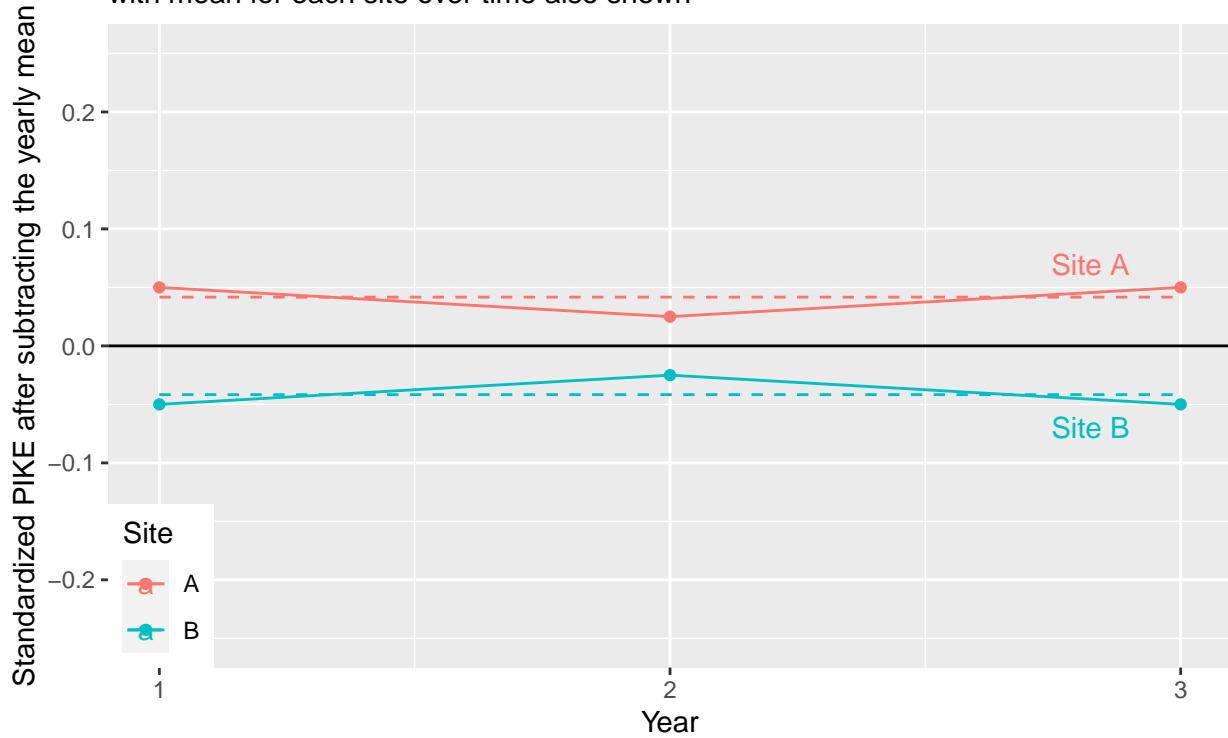
The *site effects* are the difference between the mean standardized PIKE and 0; the *site-year effects* are the differences between the *PIKE* for a particular site-year and the mean value for the site, i.e. the difference from each point to the corresponding mean (dashed line) for each site.

In this particular example, the *site effects* tend to be larger than the *site-year effects*, and so the variation in site effect will tend to be larger than the variation in site-year effects.

4 Marginal means

Given the observed *PIKE* at two or more MIKE sites, how are these values rolled up to, for example, the sub-regional or continental level. The aggregated *PIKE* is known as a marginal mean because it is formed as unweighted mean, weighted mean by carcass numbers, or a weighted mean by population abundance of the individual *PIKE* values. The averaging can also take place on the probability or on the logit (a transformation of the probability) scale.

4.1 A hypothetical example

The current estimate of the yearly mean *PIKE* values gives equal weight to every site regardless of the population abundance that the MIKE site represents. For example, suppose that the observed data, estimated *PIKE* values, and population abundances for two sites in a particular year were.

Site	Carcasses Found	Illegally Killed	*PIKE*	*logit(PIKE)*	Population abundance
A	100	30	0.3	-0.85	1000
B	50	20	0.4	-0.41	10,000

4.2 The overall observed *PIKE*

The overall observed *PIKE* value is found as the ratio of the totals of the observed carcass counts

$$PIKE_{overall\ observed} = \frac{30 + 20}{100 + 50} = .33$$

The observed pike value is heavily influenced by sites that report more carcasses.

4.3 The marginal mean *PIKE* computed on the *logit()* scale and back-transformed

The *logit* function ($logit(p) = \log(p/(1-p))$ where p is a probability and $\log()$ is the natural logarithmic function), transforms the probability scale ([0,1]) to an unbounded scale ($-\infty$ to $+\infty$) and is the basis for logistic regression. Here $logit(0.3)=\log(0.3/(1-0.3)) = -0.85$ and $logit(0.4)=\log(0.4/(1-0.4)) = -0.41$.

The unweighted estimated marginal mean $logit(PIKE)$ gives each MIKE site equal weight:

$$logit(PIKE)_{marginal} = \frac{-0.85 + -0.41}{2} = -0.63$$

and the back transformation to the [0, 1] scale gives a value of 0.348.

4.4 The unweighted marginal mean *PIKE* computed on back-transformed values

As noted below, the analysis of *PIKE* takes place on the *logit*() scale and the model is used to predict the *PIKE* for each site-year combination.

The unweighted estimated marginal mean *PIKE* gives each MIKE site equal weight and simply averages the estimated *PIKE* across sites:

$$PIKE_{marginal} = \frac{0.3 + 0.4}{2} = 0.35$$

The unweighted marginal means from the *logit()* and [0, 1] scale are similar here (0.348 vs 0.35) because at these value of *PIKE* the transformation is approximately linear. If the *PIKE* values were closer to 0 and 1, there would a larger discrepancy between these two values.

4.5 The weighted marginal mean *PIKE* computed on back-transformed values

The use of population values to compute a population-weighted PIKE is currently experimental and being evaluated and the actual implementation may change in the future.

The weighted estimated marginal mean *PIKE* weights the *PIKE* from each according to the population abundance

$$PIKE_{weighted} = \frac{1000 \times 0.3 + 10000 \times 0.4}{1000 + 10000} = 0.39$$

The unweighted and weighted marginal means would be similar if the number of carcasses examined were a consistent proportion (among sites) of the total carcasses available in a year. This is unlikely to be true.

If there is no data for a site in year, than the observed *PIKE* is also influenced by this omission. The unweighted and weighted marginal mean *PIKE* is derived using imputed values for the missing site-years as explained below.

4.6 Choice among estimates of marginal *PIKE*

The observed overall *PIKE* may be a suitable estimator for the marginal *PIKE* if carcasses reported were proportional to the population abundances (if you are interested in a weighted *PIKE*), equal across sites (if you are interested in an unweighted observed *PIKE*), and all sites reported in all years. The number of

carcasses reported is not equal across sites, only bears a weak relationship to the population abundances, and there is much missing data so the observed overall *PIKE* is difficult to interpret and seldom of interest.

The marginal *PIKE* computed on the *logit*() scale and then back-transformed suffers from a non-linear transformation effect. For example, if you wanted to compute the average temperature across sites in °C from measurements recorded in °F, it doesn't matter if you first convert each measurement to °C and then average, or if you average the measurements in °F and then transform to °C – you will get the same result. This is because the conversion between °C and °F is a linear transformation.

However, consider finding the average area when individual measurements are the length of each side of square (e.g., 3 or 5 m). Now averaging the length of the sides first (e.g. $\frac{3+5}{2} = 4$) and then squaring the average $4^2 = 16$ does NOT give the correct average area found by squaring first and then averaging ($\frac{3^2+5^2}{2} = 17$) because finding area from the length of a side is a non-linear transform.

Similarly, finding the mean of the *logit*() values and then back-transforming to the [0,1] scale does not give the same value as back-transforming the individual values to the [0,1] scale and then finding the average because the *logit*() transformation is non-linear. It can be shown that if the individual *PIKE* values are above 0.5, then the back-transformed unweighted mean $logit(PIKE)$ tends to overestimate the mean *PIKE* and conversely, if the individual *PIKE* values are below 0.5, then the back-transformed unweighted mean $logit(PIKE)$ tends to underestimate the mean *PIKE*. Consequently, it is *not recommended* to use the marginal mean $logit(PIKE)$ and then back-transform.

The choice between the weighted and unweighted marginal mean *PIKE* is more complex. The unweighted marginal mean *PIKE* should be thought of as an index of poaching pressure that gives each site the same weight regardless of the underlying population abundance. We hope that trends in the index are broadly informative of changes in poaching pressure at the continental scale. The weighted marginal mean *PIKE* may reflect actual trends in poaching at the population level under several assumptions (e.g. that the observed *PIKE* in a MIKE site is reflective of the poaching in the larger population; that the population abundance estimate for a MIKE site is reflective of the actual population abundance; that all populations are monitored (i.e. every population has a MIKE site), etc.). Consequently, the weighted marginal mean *PIKE* looks appealing but makes many assumptions in order to reflect the poaching pressure on a population level.

The use of population values to compute a population-weighted PIKE is currently experimental and being evaluated and the actual implementation may change in the future.

Given the very weak relationship between the number of carcasses examined and the population abundance, the observed overall *PIKE* may tend to track the weighted by population abundance marginal mean *PIKE* marginally better than the unweighted marginal mean *PIKE* values.

5 The Bayesian model

5.1 Why a Bayesian model?

The analysis of the *PIKE* data will follow a generalized linear mixed model. The model could be fit using maximum likelihood methods but often models with multiple levels of random effects (as in our model) are difficult to fit using likelihood methods (e.g. failure to converge), and it is difficult to create custom summaries (such as marginal means) using likelihood methods.

We have implemented our model in a Bayesian context using a method called MCMC (Markov chain Monte Carlo) sampling for the following reasons:

- It avoids many of the technical difficulties of a likelihood analysis.
- It easily deals with missing site-years rather than the integration needed in a likelihood approach.
- The hierarchical structure allows for easy sharing of information across sites. For example, the *PIKE* at a site relative to other sites provides information for missing site-years or site-years with little data.
- It easily allows computation of weighted *PIKE* in various ways at various scales (e.g. continental, subregional, etc).

18

- It deals with the binomial variation at the site-year level which was ignored in the *LSMeans* approach.
- It is easy to add uncertainty in the population abundance values and propagate this uncertainty to the marginal means.
- It will be easy to add uncertainty about the actual carcass data (e.g. certainty about a kill being illegal) if this information is collected in the future.
- Posterior inference is available on selected trends (e.g. last five years) which is not readily done using a likelihood approach.
- It implicitly accounts for spatial autocorrelation among sites that are geographically close.

A Bayesian model combines information from the prior beliefs about the values of certain parameters and the information about these parameters from the data (through a likelihood function). So a Bayesian model extends the likelihood model to allow incorporation of prior beliefs about parameters. If there is a large amount of data, the information about the parameters would usually overwhelm the information from the prior beliefs, but in cases of sparse data, the prior beliefs may be more important. The end product of a Bayesian analysis is the posterior distribution of belief about a parameter. In the context of this *PIKE* analysis, we have used vague priors with little information about the parameters so the final result is driven almost entirely by the data from the MIKE program.

The concept of posterior belief has an intuitive understanding that many people practice. For example, suppose you are waiting to be picked up by a friend. But the friend is late. A priori (prior belief), you could place different weights on two hypotheses of why your friend is late - traffic is bad and so your friend is delayed in traffic, or your friend forgot. So if your friend is generally reliable, you may play a higher prior weight on the hypothesis that traffic was bad. While waiting, you overhear from someone else about an accident on the road that has caused a large traffic jam. In light of this new information (data), you update your prior beliefs about the two hypotheses and would now place even more weight on the traffic jam hypothesis than the "forgot" hypothesis. The updated beliefs are your posterior beliefs about the hypotheses which is a combination of prior beliefs in the two hypotheses (prior distribution) and information (data, or likelihood). It would be sensible to make statements such as "I have a 75% belief that my friend is stuck in a traffic jam". Note that your friend either is or is not stuck in traffic – there is no probability associated with the actual state. As more and more data are received about the size of the traffic jam, your posterior beliefs about the two hypotheses will shift.

Similarly, you may have a prior belief about a persons age who you have never met. In this case, the prior belief is not two discrete hypotheses, but a continuum, i.e. you can picture a "normal"-like distribution of prior belief with a peak, say at age 50, but ranging from 40 to 60. As you get more information (data) such as the fact that person witnessed the fall of the Berlin Wall as a young adult, you would update your prior belief by shifting the peak of the prior age distribution and/or shrinking the range of possible ages. This is now your posterior belief summarized by a posterior distribution. It would sensible to make statements such as "I have a 40% posterior belief that this person is more than 55 years of age". Note that the actual age of the person does NOT have a distribution – the distribution refers to your knowledge of the age based on a combination of prior information and actual data. If you find the year of birth of the person, the posterior distribution becomes extremely concentrated because you now know the age of the person to within a year.

In an analogous fashion, a Bayesian analysis summarizes the posterior belief about a parameter using the posterior distribution. This distribution could be discrete (e.g. two competing hypotheses) or could be continuous (e.g. a range of values summarized by a distribution with a peak and spread.)

There are several summaries of the posterior distribution that are commonly reported:

- Mode of the posterior distribution representing the value of the parameter that is believed to be most likely.
- Mean of the posterior distribution the (mathematical) expectation of the belief distribution. If the posterior distribution is symmetrical, then the mean, median, and mode are often very close.
- Standard deviation of the posterior distribution representing a measure of uncertainty about our belief.
- A credible interval representing uncertainty in the belief. A common credible interval is found as the interval bounded by the 2.5^{th} and 97.5^{th} percentiles of the posterior distribution forming a 95% credible interval.

- Total belief that the parameter is above/below a certain value found as the portion of the posterior distribution that is above/below the specified value. As seen below, we are interested in our posterior belief that the slope in marginal *PIKE* was less than zero in the last five years representing our posterior belief that the marginal *PIKE* has declined in the last five years.

A rough correspondence exists between the results of a likelihood model and a Bayesian model. The maximum likelihood estimate roughly corresponds to the mode of the posterior distribution. If the posterior is symmetric, then this is also its mean; the standard error for a parameter from a likelihood fit roughly corresponds to the standard deviation of the posterior distribution; a confidence interval from a likelihood fit roughly corresponds to a credible interval from a Bayesian model. However, the correspondence is not "exact" and there are fundamental differences in the technical definitions of each measure that make them not commensurable. In particular, the total belief that a parameter is above/below a specific value has no correspondence in the likelihood context.

There are many other summary statistics of the posterior distribution that could be found, but many statistical packages that implement a Bayesian analysis (such as *JAGS*) typically report the mean, the standard deviation, and selected percentiles of the posterior distribution as summary measures. The packages also generate samples from the posterior (e.g. a sample of 10,000 values from the posterior) that can be used by the analyst to compute custom summary measures, or to compute posterior beliefs that the parameter value is above/below a certain value by finding what fraction of the sample of 10,000 values from the posterior are above/below the specified value.

5.2 Binomial variation within each site-year

In each site-year, the number of illegally killed elephant carcasses is a fraction of the total elephant carcasses examined. Consequently, we use a binomial distribution to model this part of the data:

$$IC_{sy} \sim Binomial(TC_{sy}, \pi_{sy})$$

where IC_{sy} is the number of illegally killed carcasses reported from site s in year y; TC_{sy} is the total number of carcasses located as reported from site s in year y; and π_{sy} is the probability that a reported carcass was defined as illegally killed in site s in year y.

5.3 Temporal and site effects

The value of π_{sy} (the *PIKE* in site s and year y) varies by time (temporal trends), by site (site effects) and over time within each site (site-year effects). Here is a key difference from the previous *LSMeans* models. The *LSMeans* model used sub-region as the unit of analysis for the continental estimates and country as the unit of analysis for the sub-regional estimates and gave each unit of analysis an equal weight when computing the aggregate estimate. In the Bayesian model, the site is the unit of analysis and the Bayesian model gives each site equal weight when aggregating to larger units (continental or subregional). This issue is discussed in detail in Appendix 3.

Because, the *PIKE* must be between 0 and 1, it is modelled on the logistic scale. Similar to (but not exactly the same as) Burn, Underwood and Blanc (2011), a Bayesian hierarchical model is adopted of the form:

$$logit(\pi_{sy}) = Year_y + Site_s(R) + SiteYear_{sy}(R)$$

where $Year_y$ is the effect of year y on the $logit(PIKE)$; $Site_s(R)$ is the (random) effect of site s on the $logit(PIKE)$; and $SiteYear_{sy}(R)$ is the (random) effect of site s in year y on the $logit(PIKE)$.

Here *year* is not modelled in a hierarchical fashion because we are interested in these particular years and do not believe that these years represent a (theoretical) sample from all possible years.

The random effects of site and site-year are modelled using a hierarchical model, i.e.

$$Site_s \sim Normal(0, \sigma_{site})$$

and
$$SiteYear_{sy} \sim Normal(0, \sigma_{site.year})$$

Here the $Year_y$ effects represents the average $logit(PIKE)$ over all sites giving each site an equal weight, analogous to the least-square means reported in previous analyses.

5.4   Marginal mean *PIKE*

Once the model is fit, the estimated $logit(PIKE)$ for all sites and years where no data are collected is found as:
$$\widehat{logit(\pi_{sy})} = \widehat{Year}_y + \widehat{Site}_s + \widehat{SiteYear}_{sy}$$

Note that if no data are collected in a particular site-year, the estimated *PIKE* is based purely on the estimated value from other years. Because all *Site.Year* effects are assumed to be independent among and within sites, so their values must be simulated from the posterior distribution.

Once the estimated site-year values are obtained, the marginal means are found in three ways:

1. The marginal mean on the *logit* scale

$$MM_y^{logit} = \frac{\sum_s \widehat{logit(\pi_{sy})}}{s}$$

   where s is the number of sites.

This marginal mean can also be interpreted as the $logit(PIKE)$ when the *Site* and *Site.Year* effects are zero, i.e. for an "average site".

This marginal mean can be back transformed to the [0,1] scale. Because the $logit()$ scale is a non-linear transformation of the [0,1] scale, this (default) method of computing a marginal mean is greatly influenced by $logit()$ values from *PIKE* that are close to 0 or 1, i.e., $logit(0) = -\infty$ and $logit(1) = +\infty$. Consequently, this marginal mean is **not recommended** for use.

2. The marginal mean on the probability (i.e. the 0-1) scale

$$MM_y^{unweighted} = \frac{\sum_s \widehat{\pi}_{sy}}{s}$$

   This is closest to the marginal means computed in the prior analysis (the *LSMeans* approach) and is the recommended approach for computing the unweighted marginal mean.

3. The weighted marginal mean by population abundance

$$MM_y^{weighted} = \frac{\sum_s \widehat{N}_{sy} \times \widehat{\pi}_{sy}}{\sum_s \widehat{N}_{sy}}$$

   Because the number of carcasses examined in a year are unlikely to be a consistent fraction of the carcasses available, this estimate attempts to correct for this source of bias. Uncertainty in the population estimates has NOT been taken into account in computing the marginal mean so the uncertainty in these estimates is expected to be understated.

The use of population values to compute a population-weighted PIKE is currently experimental and being evaluated and the actual implementation may change in the future.

5.5   Uncertainty about marginal mean *PIKE*

There are three sources of uncertainty that need to be considered when estimating the uncertainty about the marginal mean *PIKE*:

- choice of MIKE sites

- imputation of missing *PIKE* in year.sites where no data is collected
- estimation of *PIKE* in a year.site when only a small number of carcasses is measured.

If you believe that MIKE sites were chosen at random from a larger population of MIKE sites and you need to account for this initial selection of sites, then all three sources of uncertainty need to be incorporated into the estimates.

However, MIKE sites were selected to be representative of most major populations of elephants and the notion of a new sample of MIKE sites may not be realistic. In this case, the MIKE sites are "fixed" and only the last two elements of uncertainty need to be incorporated.

The differences between these two interpretations can be made clearer if asked what uncertainty should be reported if all MIKES reported in all years and had perfect information, i.e. the mortality of every single mortality in the associated population is known. If you believe that the current MIKE sites are a random sample from many potential MIKE sites, then there is sampling uncertainty associated with the marginal mean. If you believe that the current set of MIKE sites is fixed and representative, then marginal mean *PIKE* would then have an uncertainty of 0.

This issue is explored in more detail in Appendix 2.

It turns out that finding the uncertainty when MIKE sites are treated as "fixed" is automatically provided by the Bayesian analysis and no further computations are needed.

If the MIKE sites are to be treated as a random sample of sites taken from a larger population of MIKE sites, then the Bayesian uncertainty associated with the *Year.eff* term on the logit scale automatically incorporates all three sources of uncertainty. However, as noted previously and later in the document, you cannot simple take the anti-logit of the *Year.eff* to get the marginal mean *PIKE* on the [0,1] scale with the proper accounting of uncertainty because of the transformation bias induced by the anti-logit transform.

We derived the uncertainty of the marginal mean *PIKE* on the [0,1] scale accounting for a random sample of sites and correcting for the transformation bias, by using Bayesian Bootstrapping (Rubin, 1981; https://stats.stackexchange.com/questions/181350/bootstrapping-vs-bayesian-bootstrapping-conceptually). For each sample from the posterior, the year.site values for *PIKE* on the logit scale (accounting for uncertainty from a sample of carcasses and imputation for missing year.site values), are converted to the [0,1] scale. A sample of weights is generated from a Dirichlet distribution with prior weights all set to 1. The sample of weights are then used to compute a weighted average of the year.site values on the [0,1] scale.

More formally,

$$\mathbf{w} \sim Dirichlet(1, 1, 1, ...1_{Nsites})$$

$$MM_y^{BB,unweighted} = \sum_s w_i \times \widehat{\pi}_{sy}$$

The posterior distribution of the Bayesian bootstrap estimator will then account for all sources of uncertainty.

6 Continential trends in *PIKE*

The above model was coded using *BUGS* (Lunn et al, 2012), a common way to specify Bayesian models and run using *JAGS* (Plummer, 2003) within *R* (R Core Team, 2020).

Vague priors were specified for the year effects, and conjugate prior specified for the variance components of the *site* and *site.year* effects.

The model was run for 5000 iterations with the first 2000 iterations discarded as burnin and the MCMC samples thinned by a factor of 2. Multiple independent chains (3) were run and 1500 samples from the posterior samples were retained from each chain. A total of 4500 samples from the posterior from all chains were retained.

6.1 Estimates of site and year.site variance components

The estimated variance components (on the *logit* scale are):

Table 5: Estimated standard dev of Site and Year.Site effects

	Mean	SD	95% CI		Rhat	Eff n
			Lower	Upper		
sd.site.eff	1.66	0.19	1.33	2.06	1.006	360
sd.year.site.eff	1.27	0.06	1.15	1.39	1.026	87

The variation in *PIKE* across sites is larger than within site-years (as expected). This indicates that the *PIKE* varies more across sites, than the *PIKE* varies within a site (across years)

6.2 Estimates of year effects and marginal means on the (*logit* scale)

The estimated year effects (on the *logit* scale) are:

Table 6: Estimated year effects on the logit scale

Year index	Year	Mean	SD	95% CI	
				Lower	Upper
1	2003	-0.36	0.36	-1.07	0.31
2	2004	-0.33	0.33	-0.99	0.32
3	2005	-0.54	0.36	-1.24	0.17
4	2006	-0.37	0.38	-1.10	0.37
5	2007	-0.25	0.37	-0.95	0.46
6	2008	0.33	0.36	-0.38	1.04
7	2009	0.25	0.34	-0.40	0.92
8	2010	0.16	0.34	-0.52	0.84
9	2011	1.26	0.33	0.62	1.92
10	2012	1.08	0.33	0.44	1.70
11	2013	0.68	0.33	0.03	1.33
12	2014	1.12	0.34	0.46	1.81
13	2015	0.90	0.33	0.25	1.54
14	2016	0.83	0.34	0.16	1.51
15	2017	0.52	0.34	-0.13	1.17
16	2018	0.26	0.32	-0.37	0.91
17	2019	-0.62	0.33	-1.25	0.06

The year effects are the $logit(PIKE)$ for an "average site" in each year or for the average $logit(PIKE)$ over a random sample of sites (refer to the appendices for more details). The SD for this term depends on the variance components seen earlier and the number of sites and is only weakly dependent on the number of carcasses measured each year and the number of imputed values in a year.

This is contrasted to the marginal means on the *logit* scale, i.e. the marginal mean $logit(PIKE)$ is computed in each year over sites that have data or sites with imputed site.years:

Table 7: Estimated marginal means on the logit scale

Year index	Year	Mean	SD	95% CI Lower	95% CI Upper
1	2003	-0.37	0.23	-0.84	0.07
2	2004	-0.33	0.20	-0.73	0.07
3	2005	-0.54	0.24	-1.02	-0.09
4	2006	-0.37	0.26	-0.88	0.16
5	2007	-0.25	0.25	-0.74	0.24
6	2008	0.34	0.23	-0.12	0.79
7	2009	0.25	0.21	-0.17	0.65
8	2010	0.15	0.21	-0.26	0.56
9	2011	1.26	0.19	0.90	1.64
10	2012	1.08	0.19	0.71	1.44
11	2013	0.68	0.19	0.31	1.04
12	2014	1.12	0.20	0.72	1.51
13	2015	0.89	0.19	0.50	1.27
14	2016	0.83	0.21	0.43	1.24
15	2017	0.51	0.20	0.12	0.91
16	2018	0.26	0.17	-0.08	0.60
17	2019	-0.61	0.18	-0.98	-0.26

If these two values are plotted against each other for each year, they are very close (as expected and explained in the appendices):

Comparison of direct estimate of year effect and computed marginal mean on logit scale
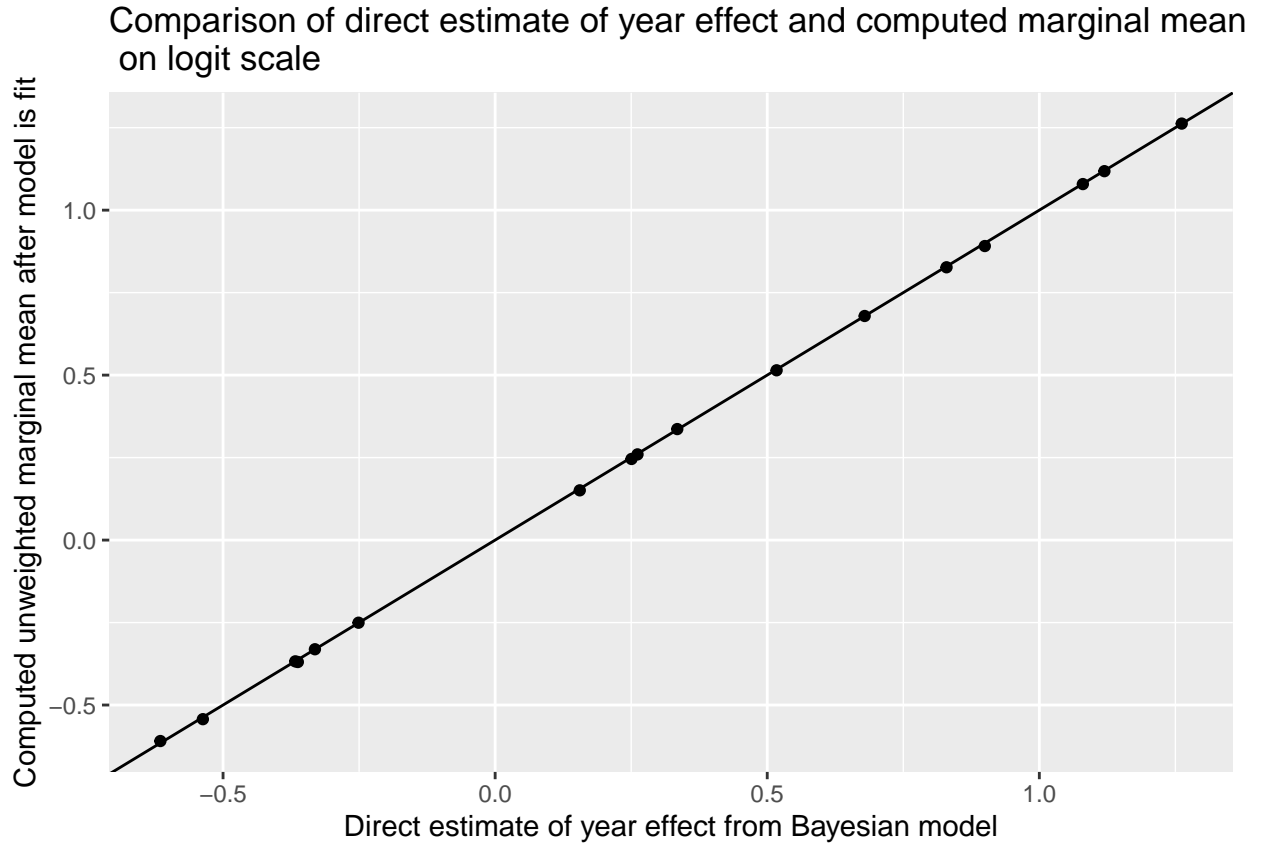
The standard deviation for the *Year.eff* can be interpreted as closest to the standard error of a mean, i.e. how uncertain are you about the mean $logit(PIKE)$ if you are willing to assume that the sites are a random sample from all possible sites etc. The standard deviation for the marginal mean $logit(PIKE)$ treats the sites chosen as a fixed index to all sites and so the concept of a random sample of sites has no meaning. The mean $logit(PIKE)$ is also an index to the overall $PIKE$ and uncertainty in this index is driven by the uncertainty in the individual site-year observed $PIKE$, i.e. by the number of carcasses monitored and the uncertainty in the imputation for site.years that are missing (see appendices for details)

6.3 Estimates of year effects and marginal means on the [0,1] scale

However, interest lies on the marginal mean *PIKE* on the [0,1] scale rather than the logit scale.

There are three possible estimates of these marginal mean *PIKE*:

- The anti-logit of the *Year* effect which was originally estimated on the logit scale.
- The marginal mean of the anti-logit of the boot-strapped site-year estimates
- The marginal mean of the anti-logit of the estimated site-year *logit(PIKE)* for these sites.

Table 8: Comparison of marginal mean PIKE on [0,1] scale

Year	anti-logit Year effect		mean antilogit Bootstrap		mean antilogit year.site	
	Mean	SD	Mean	SD	Mean	SD
2003	0.41	0.084	0.44	0.053	0.44	0.035
2004	0.42	0.079	0.46	0.050	0.46	0.031
2005	0.37	0.082	0.42	0.052	0.42	0.035
2006	0.41	0.088	0.45	0.056	0.45	0.039
2007	0.44	0.087	0.46	0.054	0.46	0.037
2008	0.58	0.086	0.56	0.052	0.56	0.032
2009	0.56	0.082	0.54	0.050	0.54	0.030
2010	0.54	0.082	0.53	0.050	0.53	0.030
2011	0.77	0.057	0.70	0.042	0.70	0.024
2012	0.74	0.061	0.66	0.044	0.66	0.023
2013	0.66	0.072	0.62	0.046	0.62	0.025
2014	0.75	0.063	0.66	0.047	0.66	0.026
2015	0.71	0.068	0.63	0.046	0.63	0.026
2016	0.69	0.071	0.62	0.048	0.62	0.026
2017	0.62	0.078	0.58	0.048	0.58	0.028
2018	0.56	0.077	0.54	0.045	0.54	0.024
2019	0.35	0.074	0.42	0.047	0.41	0.027

Notice that the estimated marginal mean *PIKE* of the last two methods are the same but the standard deviations differ.

The first estimate computed from the anti-logit of the year effect from the model is unsatisfactory because of the back-transformation bias. For example, consider three sites in one particular year: * Site A. $PIKE = 0.9$ or $logit(PIKE) = 2.20$. * Site B. $PIKE = 0.8$ or $logit(PIKE) = 1.38$. * Site C. $PIKE = 0.7$ or $logit(PIKE) = 0.84$.

The year effect is estimated as the mean of the logit values

$$Year\ effect = \frac{2.20 + 1.38 + 0.84}{3} = 1.48$$

and $anti-logit(1.48) = 0.82$ which is larger than the mean *PIKE* of 0.8.

As noted previously, the transformation from the *logit* scale to the probability scale is not linear, and so back transformation of the mean *PIKE* over sites in a year on the *logit* scale is not equal to the mean of the back transformed *PIKE* for a site in a year to the [0,1] scale. The transformation bias is positive if the mean *PIKE* is more than 0.5 and negative if the mean *PIKE* is less than 0.5.

As noted earlier, this marginal mean is closer in spirit to the marginal mean computed in the previous analysis (the *LSMeans* approach). These values differ from the year effects on the [0,1] scale because the range of *PIKE* values is very wide and so the *logit* scale is not longer linear.

The transformation bias (i.e., the anti-logit of the mean of the year-site estimates on the logit scale, vs. the mean of the anti-logit of the year-site estimates in a year) is shown in the following plot:
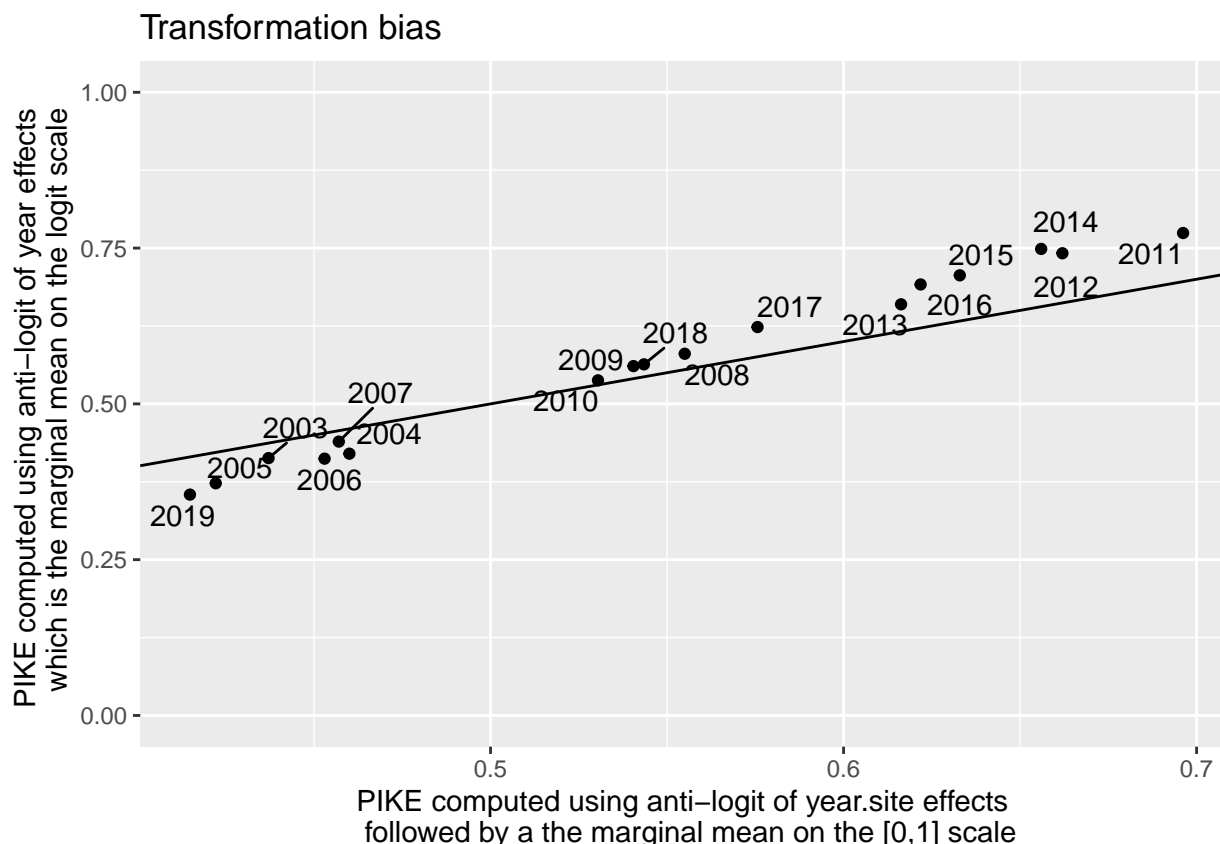

As expected (see earlier sections), a negative bias exists when the marginal mean on the logit-scale is back-transformed to the [0,1] scale when the *PIKE* is < 0.5 and a positive bias when the *PIKE* is > 0.5. This is why we first back transform to the [0,] scale before finding the marginal mean.

If we plot the trends over time:

Africa: Estimated PIKE across time

Comparison of marginal mean first computed on the logit scale and back transformed vs. back transforming each site PIKE first
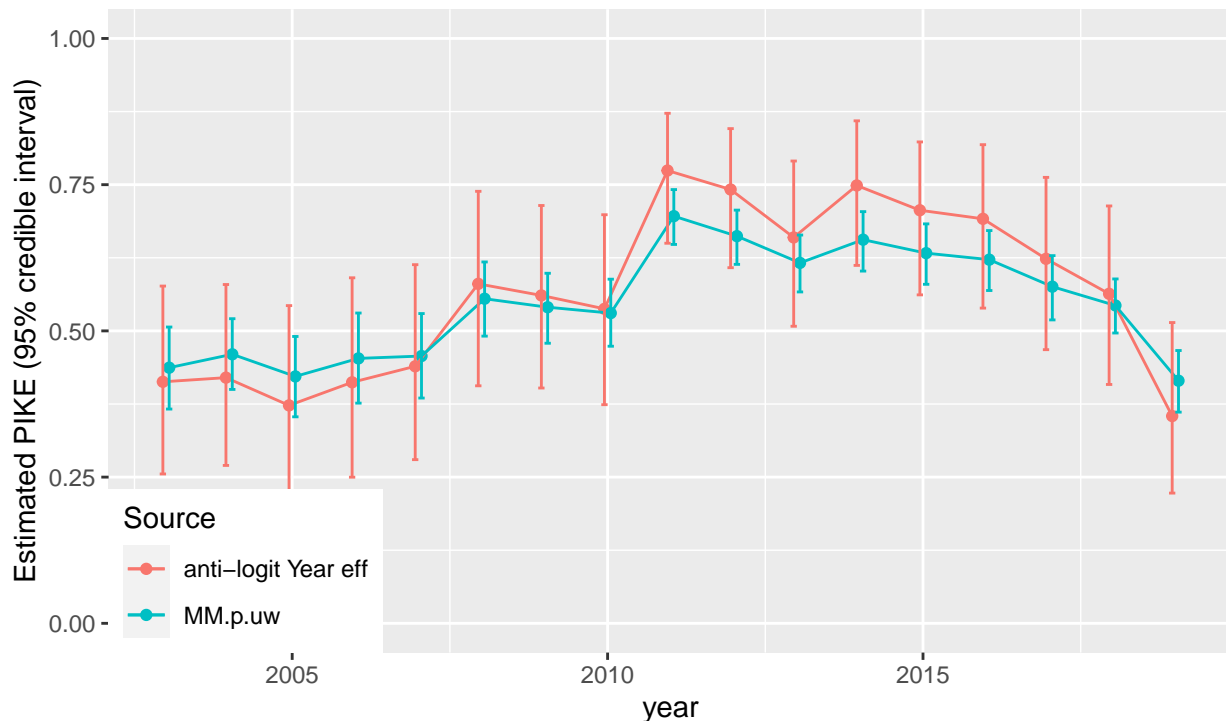
we see that when *PIKE* is > 0.5, the marginal mean computed on the *logit* scale and then back transformed ($MM.logit$) is consistently larger than the marginal means first computed by back-transforming the *PIKE* value for each year.site and then finding the marginal mean ($MM.p.uw$) and vice versa when *PIKE* is < 0.5. This is an artefact of the non-linear transformation from the *logit* scale to the $[0, 1]$ scale. **Consequently, it is recommended that the estimated *PIKE* for each year.site be first back-transformed before computing marginal means.**

We corrected for this transformation bias by first converting the site.year estimates of *logit(PIKE)* to the *PIKE* for each year, and then taking the average (last two sets of columns in the first table of this section). These last two estimates are plotted over time:

Africa: Estimated PIKE across time

Comparison of marginal means computed on the [0,1] scale
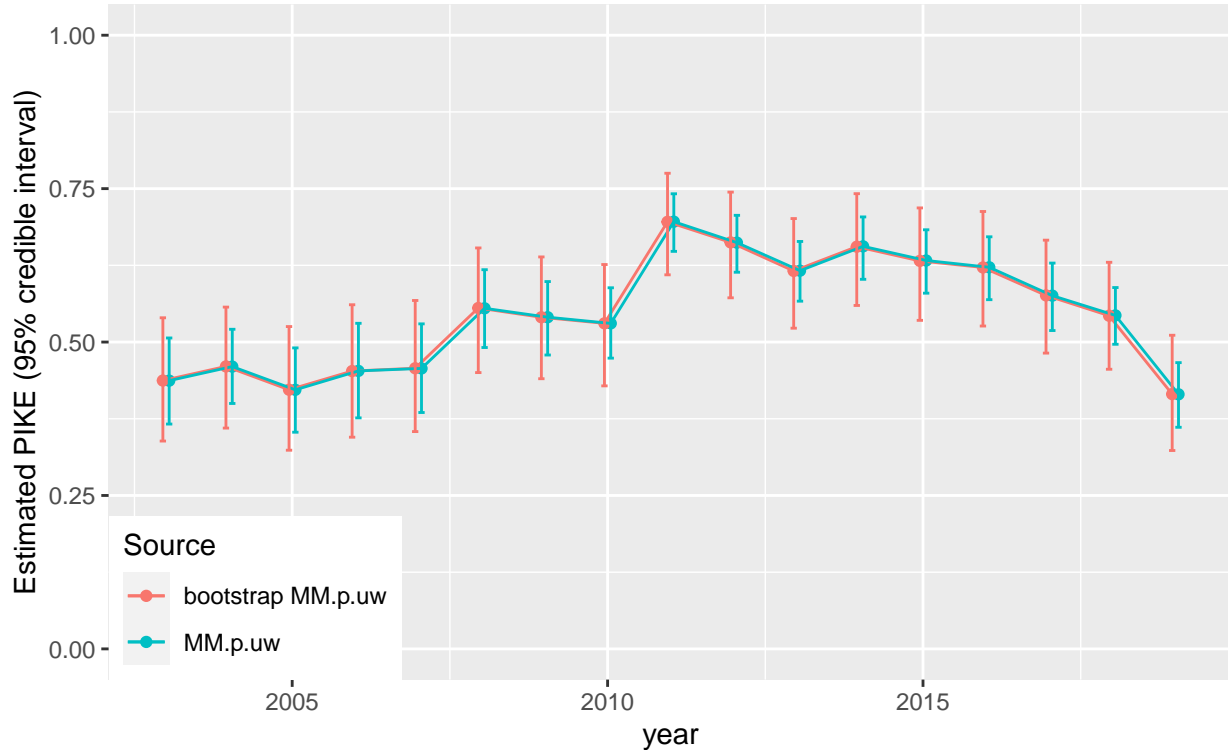
We see that they are identical (as they must be) but the uncertainty is larger in the bootstrapped marginal mean. This is because the uncertainty relates to how we interpret the marginal mean *PIKE*.

If we believe that MIKE sites are a true random sample from all sites with elephant populations and want to account for uncertainty in the continental mean due to the random sampling of sites, the uncertainty in *PIKE* in individual site.year, and the imputation process, then the uncertainty attached to the bootstrap marginal mean *PIKE* should be used. Even if every MIKE site had perfect information (e.g. every elephant mortality found and carcass status known with no missing values), there would still be uncertainty associated with the random sample of MIKE sites. This uncertainty is closest in spirit to the uncertainty reported from a random sample of numbers, i.e. the mean and standard error of the mean.

However, MIKE sites are not randomly selected but were purposely selected to be "representative" of the various elephant populations, then other MIKE sites that could have been selected are not relevant. Sites are treated as being fixed, and the only uncertainty of interest is due to a small sample of carcasses being monitored in each site-year and missing site-years. If every site has perfect information, the uncertainty of the *MM.p.uw* would be zero.

These (subtle) differences are explored more in Appendix 2 of this report.

6.4 Weighted marginal means on the [0.1] scale

Finally, the weighted (by the population abundances) marginal mean of the *PIKE* computed on the [0,1] scale is:

Table 9: Comparison of marginal weighted mean PIKE on [0,1] scale

Year	Bootstrapped marginal weighted PIKE		Marginal weighted PIKE	
	Mean	SD	Mean	SD
2003	0.34	0.074	0.34	0.037
2004	0.37	0.075	0.36	0.032
2005	0.41	0.064	0.41	0.041
2006	0.47	0.082	0.47	0.035
2007	0.44	0.072	0.44	0.033
2008	0.49	0.082	0.49	0.033
2009	0.45	0.075	0.45	0.035
2010	0.54	0.065	0.54	0.024
2011	0.68	0.065	0.68	0.017
2012	0.59	0.074	0.58	0.015
2013	0.53	0.088	0.52	0.018
2014	0.57	0.071	0.56	0.018
2015	0.56	0.069	0.56	0.017
2016	0.47	0.068	0.46	0.017
2017	0.44	0.058	0.44	0.019
2018	0.45	0.056	0.44	0.017
2019	0.31	0.056	0.30	0.017

As in the previous section, these two methods of finding the marginal weighted mean are identical but with different standard deviations.

If we believe that MIKE sites are a true random sample from all sites with elephant populations and want to account for uncertainty in the continental mean due to the random sampling of sites, the uncertainty in *PIKE* in individual site.year, and the imputation process, then the uncertainty attached to the bootstrap marginal weighted mean *PIKE* should be used. Even if every MIKE site had perfect information (e.g. every elephant mortality found and carcass status known with no missing values), there would still be uncertainty associated with the random sample of MIKE sites. This uncertainty is closest in spirit to the uncertainty reported from a random sample of numbers, i.e. the mean and standard error of the mean.

However, MIKE sites are not randomly selected but were purposely selected to be "representative" of the various elephant populations, then other MIKE sites that could have been selected are not relevant. Sites are treated as being fixed, and the only uncertainty of interest is due to a small sample of carcasses being monitored in each site-year and missing site-years. If every site has perfect information, the uncertainty of the *MM.p.uw* would be zero.

These (subtle) differences are explored more in Appendix 2 of this report.

These two estimates are plotted over time:

Africa: Estimated PIKE across time

Comparison of marginal weighted means computed on the [0,1] scale

Notice that the standard deviation is much wider from the bootstrap estimates because now it is possible that a different sample of sites with their associated weights could give quite a different estimate. For example, rather than only having one MIKE site representing 100,000 elephants, the bootstrap estimator allows for multiple MIKE sites each representing 100,000 elephants. If there are few (or no) other large populations not already represented in the current set of MIKE sites, then the bootstrap estimate of uncertainty will overstate the actual uncertainty.

6.5 Comparing the observed *PIKE*, the *LSMeans* of *PIKE*, and the weighted and unweighted marginal mean PIKE

We first compare the *LSMeans* computed using the previous method, the unweighted marginal mean computed using the Bayesian model ($MM.p.u$) and the observed *PIKE* value:

Africa: Estimated PIKE across time
Comparison of LSMeans, weighted and unweighted marginal means, and observed PIK
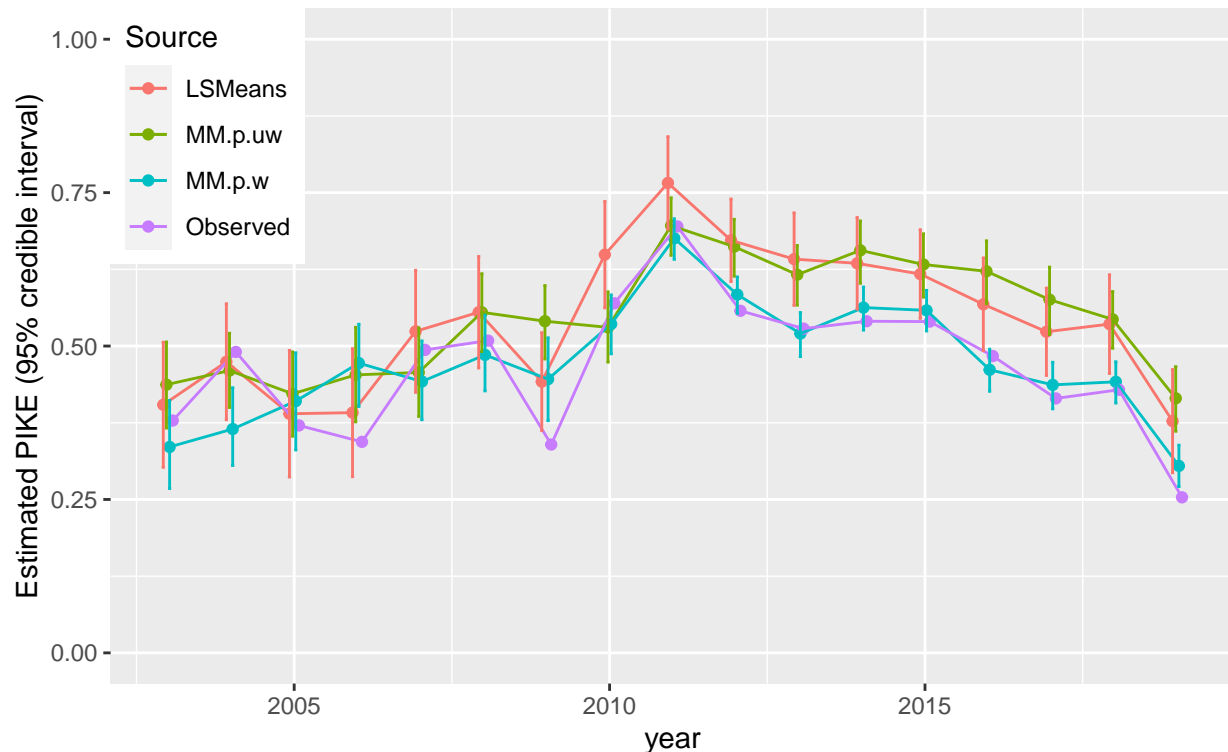
The new proposed unweighted marginal mean ($MM.p.uw$) tracks the previously computed $LSMeans$ fairly well except for 2009-2011. The fitted trend lines are consistently higher than the observed $PIKE$ values because the latter gives more weight to sites with more carcasses whereas the $MM.p.uw$ give equal weight to each site regardless of number of carcasses (or underlying population abundance). The weighted marginal mean ($MM.p.w$) tracks the observed $PIKE$ closely after 2010 because sites with large number of carcasses (and larger elephant populations) will dominate both in a similar way.

We can also plot the individual estimates against each other:

Africa : Estimated PIKE across time

Comparison of LSMeans, weighted and unweighted marginal means
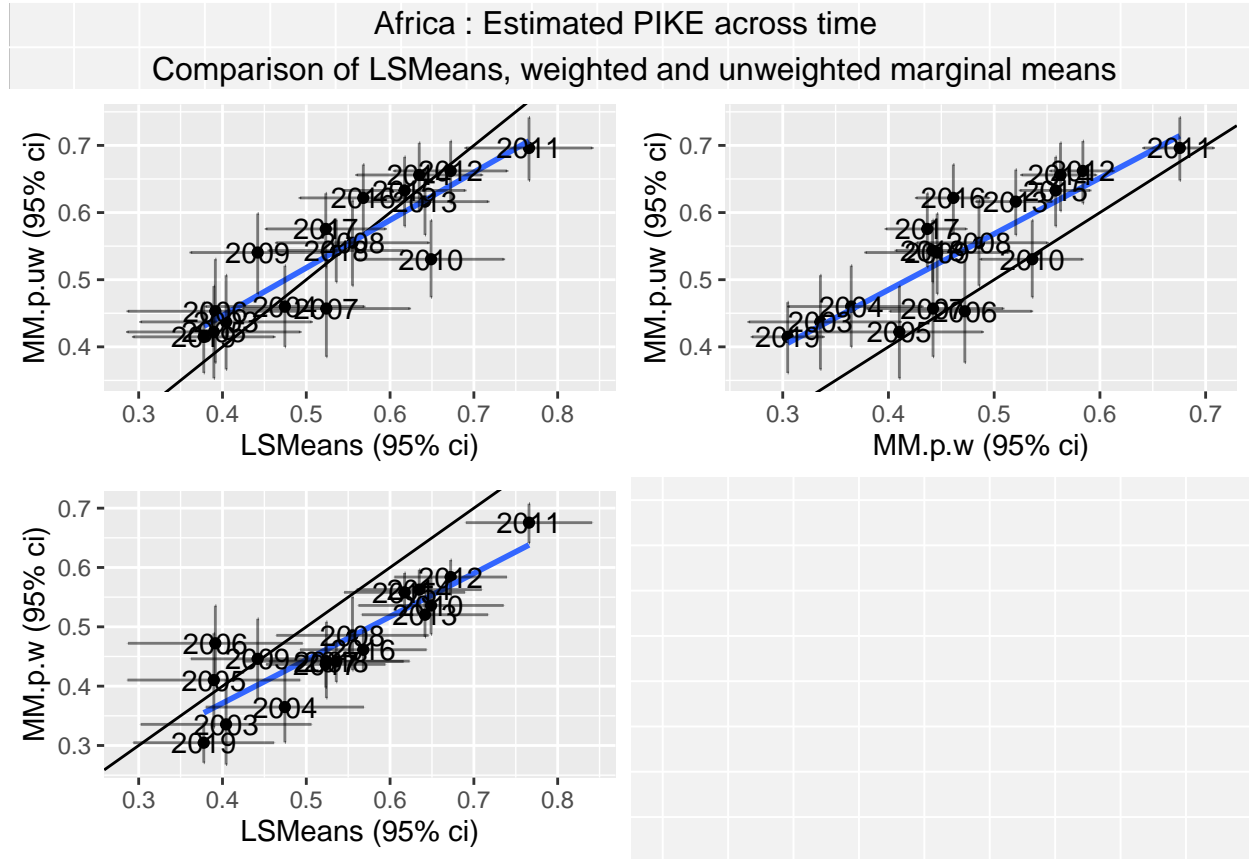
In general, the weighted (by population abundance) estimates tend to be lower than the unweighted estimates (top right plot). This is an artefact of the data where sites with larger abundances tend to have lower *PIKE*. The *LSMeans* estimates also tend to be lower than the weighted estimates (bottom left plot) for the same reason.

The *LSMeans* and unweighted estimates are similar, however, in years when the *PIKE* are smaller (i.e. closer to zero), the unweighted estimates are shrunk towards the grand mean and vice versa in years when the *PIKE* are larger (i.e. closer to one). This occurs because an observed *PIKE* close to 0 or 1 is likely an artefact of a small sample size (e.g. 1 illegally killed out of 6 carcasses examined; 6 illegally killed out of 6 carcasses examined) and the GLMM model will shrink these estimates to the overall mean because with a small sample size, observed *PIKE* are very variable.

6.6 Comparing the unweighted marginal mean and weighted marginal mean *PIKE*

We present here a more focused comparison of the unweighted marginal mean *PIKE* ($MM.p.uw$) vs. the weighted marginal mean *PIKE* ($MM.p.w$):

Africa: Estimated PIKE across time
### Comparison of the weighted and unweighted marginal mean PIKE
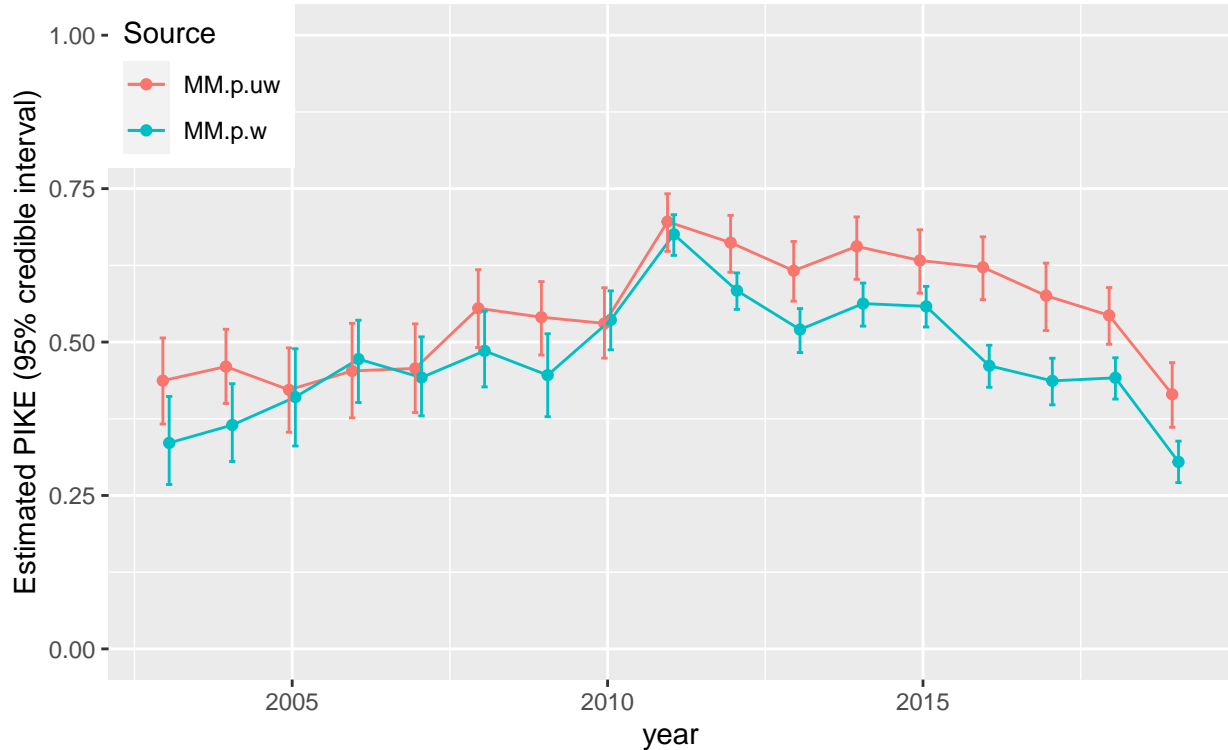
The weighted marginal mean (*MM.p.w*) weights each estimated *PIKE* by the estimated population abundance at the MIKE site while the unweighted marginal mean (*MM.p.uw*) gives equal weight to every site regardless of population abundance. In the MIKE dataset, sites that tend to have fewer elephants also tend to have higher reported *PIKE* and these sites will influence the unweighted marginal mean more than the weighted marginal mean.

The use of population values to compute a population-weighted PIKE is currently experimental and being evaluated and the actual implementation may change in the future.

6.7 Posterior belief of trend in *PIKE* in last few years

Once the sample from the posterior is available, it is relatively easy to estimate the posterior belief that the trend is negative in the last 5 years. This is done by estimating the slope in the last 5 years for each sample from the posterior, and then the posterior belief that the trend is negative is the proportion of fitted slopes that are less than zero. The posterior distribution of the slope in the last 5 years is

Africa: Posterior distribution of slope in fitted yearly PIKE in last 5 years
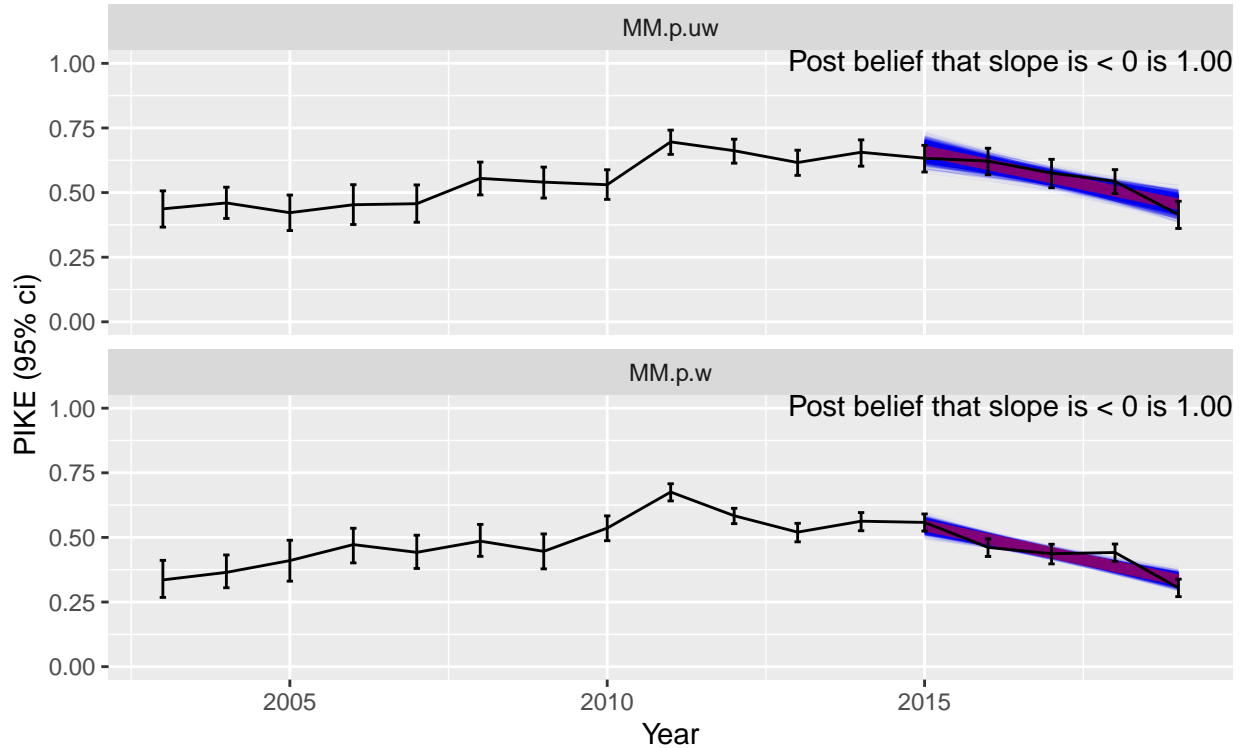Based on marginal mean PIKE
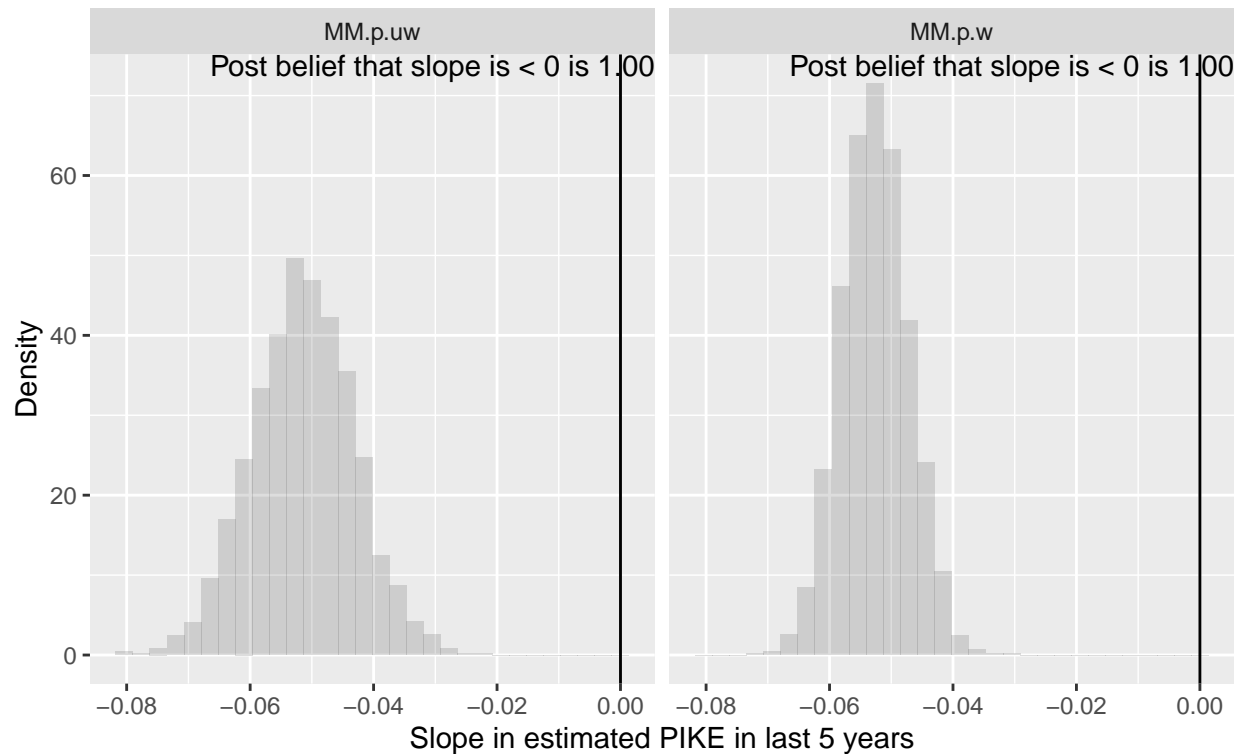
In this case, the posterior belief that the slope in PIKE is negative in the last 5 years is very high for both the unweighted and weighted *PIKE*', i.e. we have a high posterior belief that the *PIKE* in the last 5 years is declining.

This can be visualized:

Africa: Trend in PIKE in last 5 years

Shaded area is the envelope of posterior trends

There is a strong posterior belief that the trend in the last 5 years is negative, i.e. the *PIKE* is declining in the last 5 years.

7 Subregional trends in *PIKE*

The above analyses were repeated at the sub-regional level. Only the data from each sub-region was used in each analysis, i.e., completely separate analyses were performed for each sub-region.

A more complex model (Section 8.8) where the four sub-regions were modelled as part of a larger model could be attempted. This larger model would be useful in cases where a sub-region has sparse data but has a consistent relationship in *PIKE* trends with other sub-regions. This is similar to the case where sparse data for a particular MIKE site is improved if there is a consistent relationship between the *PIKE* in several MIKE sites. In these cases, the consistency in the relationship allows information from other sub-regions to improve the estimates of trend in the sub-region with sparse data.

This more complex model is under investigation, but there does not appear to be a great deal of similarity in trends among sub-regions, so we do not expect this more complex model to provide much of an improvement over individual models for each sub-region.
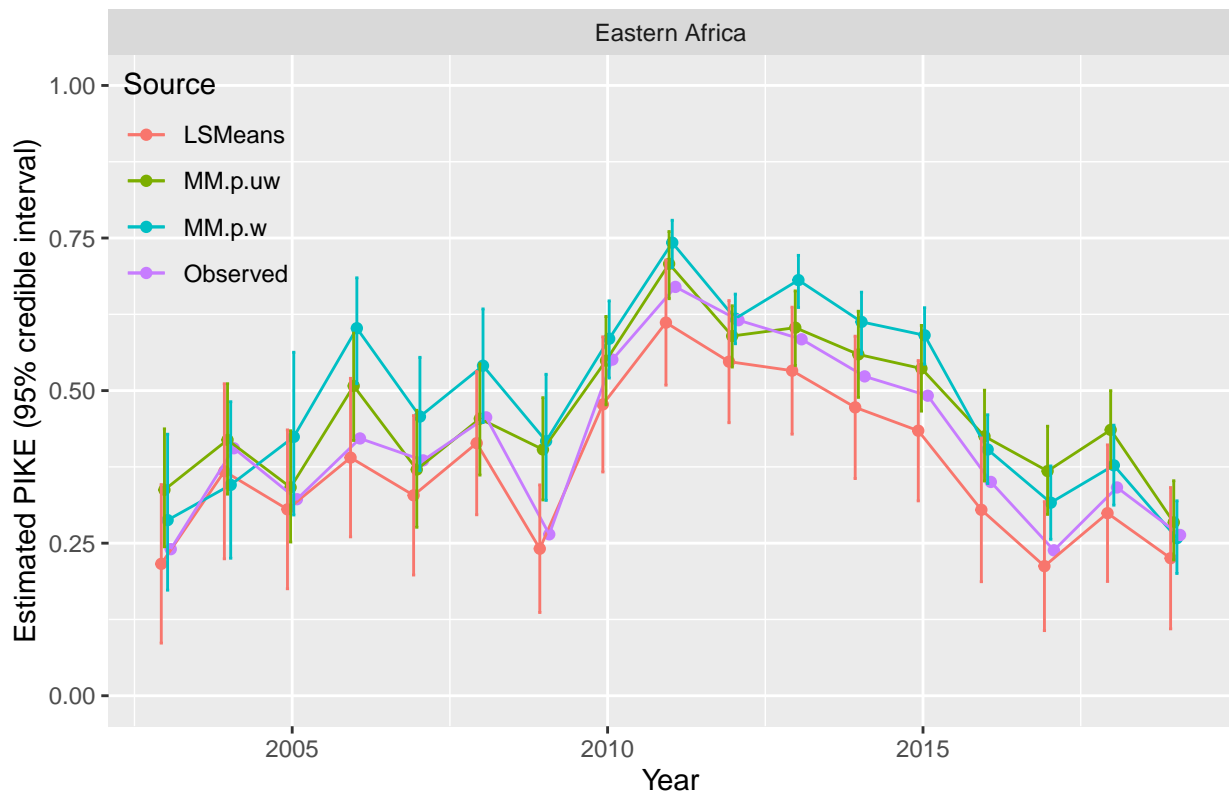
7.1 Comparison of *LSMeans*, observed, and unweighted marginal *PIKE*

The following plots compare the previously computed *PIKE* computed using the *LSMeans* method, the observed *PIKE* and the unweighted marginal *PIKE* values.

Africa: Estimated subregional PIKE across time
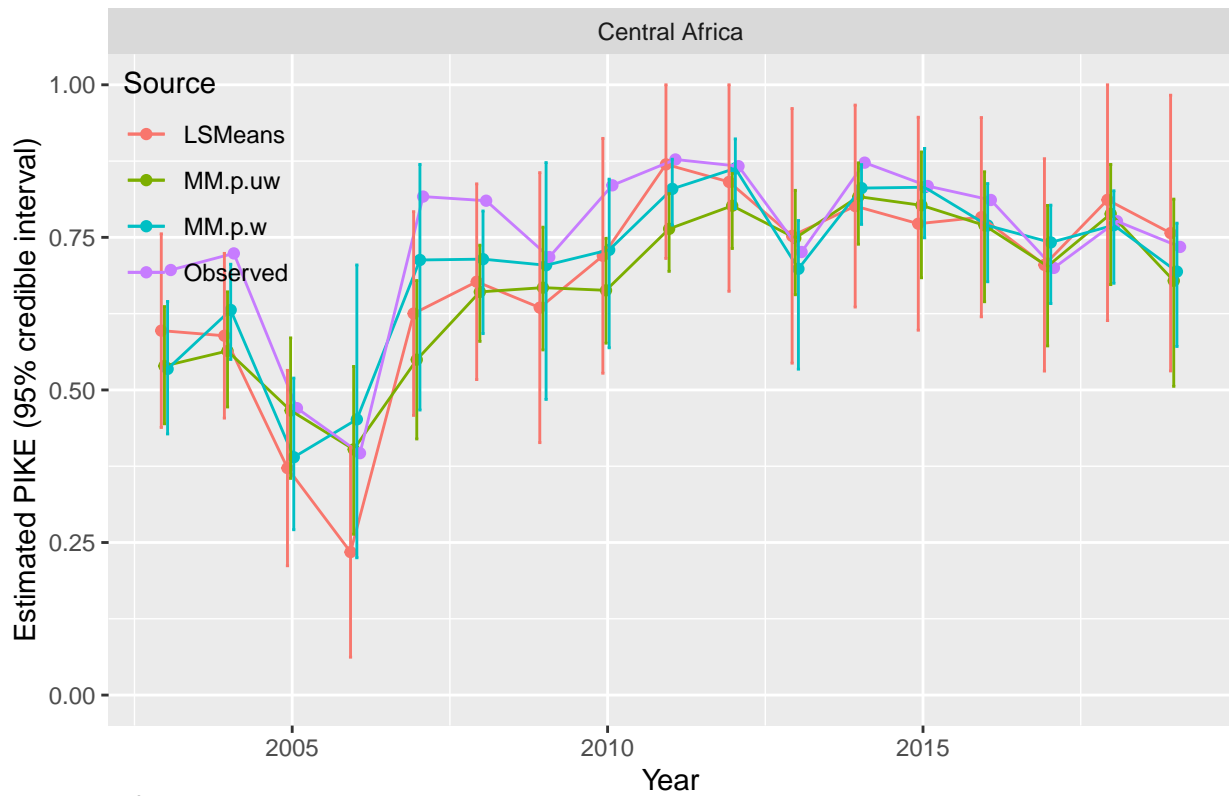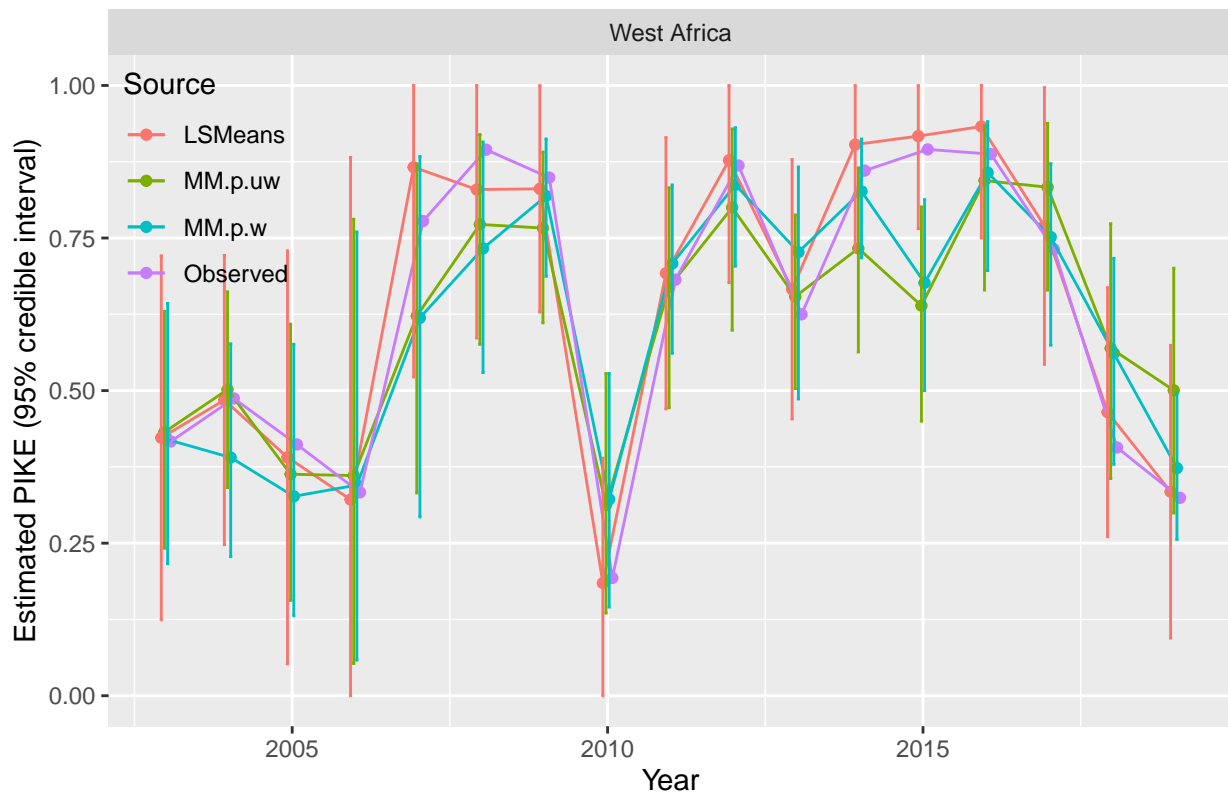

Central Africa

Africa: Estimated subregional PIKE across time

Eastern Africa
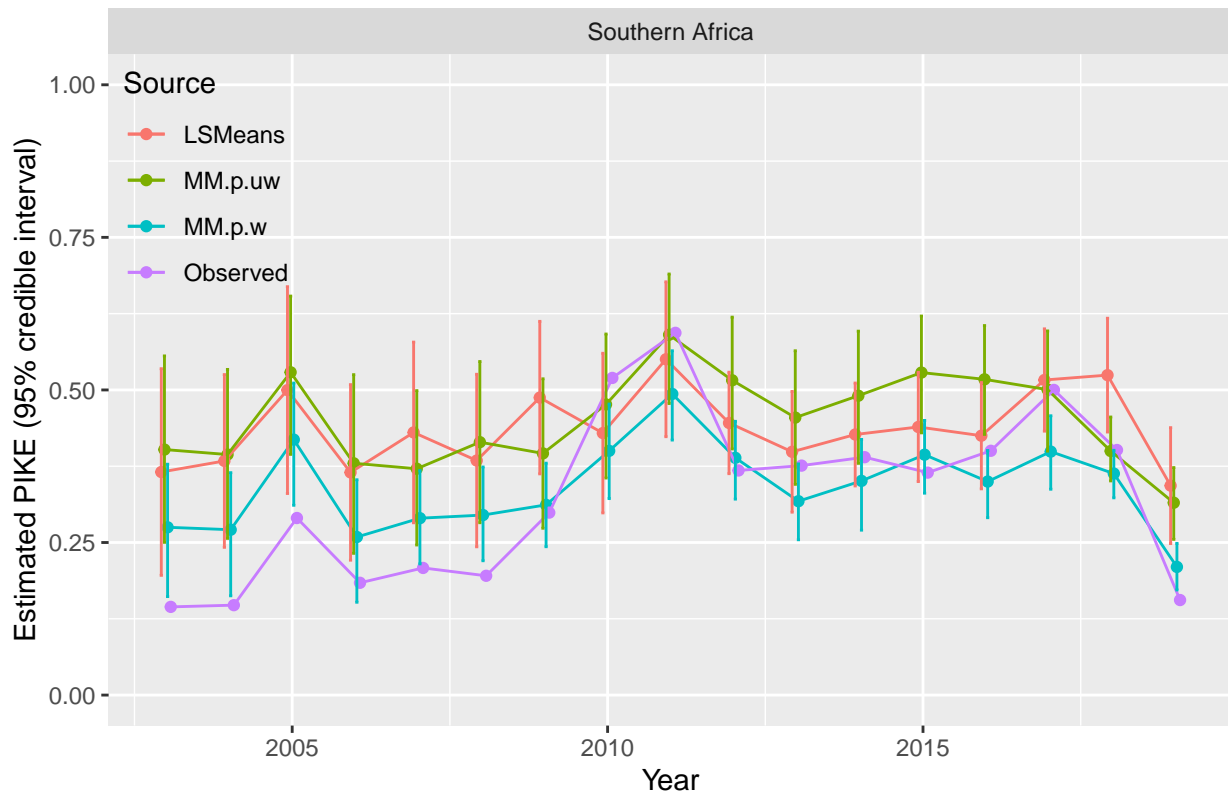
Africa: Estimated subregional PIKE across time

The estimates at each sub-region have wider confidence intervals because the amount of data is smaller in

each sub-region compared to the continental results.

It is hard to make a general statement comparing the estimates computed using $LSMeans$ and the unweighted marginal means ($MM.p.uw$) because of the wide confidence intervals, but the respective trend lines mostly track each other. The apparent consistent difference seen for Eastern Africa and Southern Africa are artefacts of how data are aggregated to the sub-regional level in the $LSMeans$ and Bayesian GLMM models (see Appendix 3 for details).

We can also plot each estimate against each other for the regions:
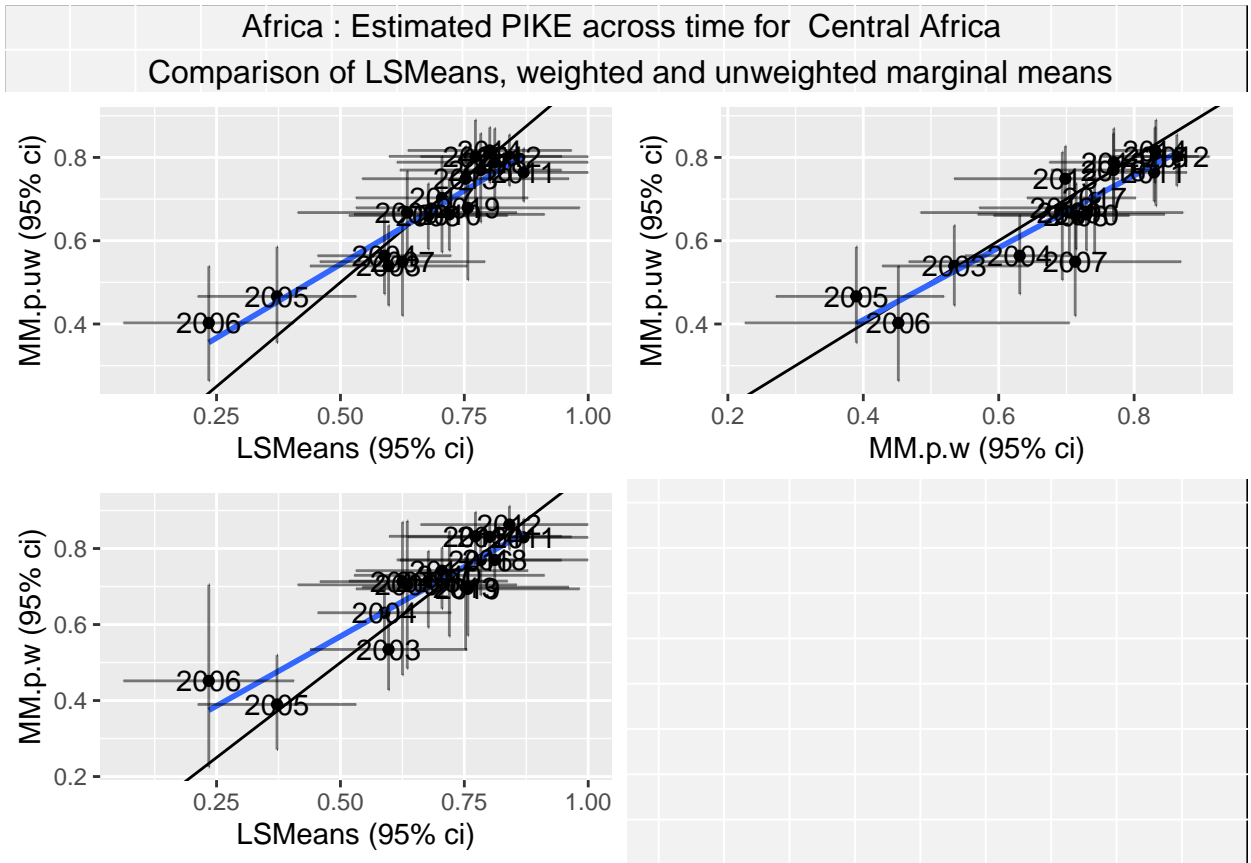
Africa : Estimated PIKE across time for Eastern Africa
Comparison of LSMeans, weighted and unweighted marginal means
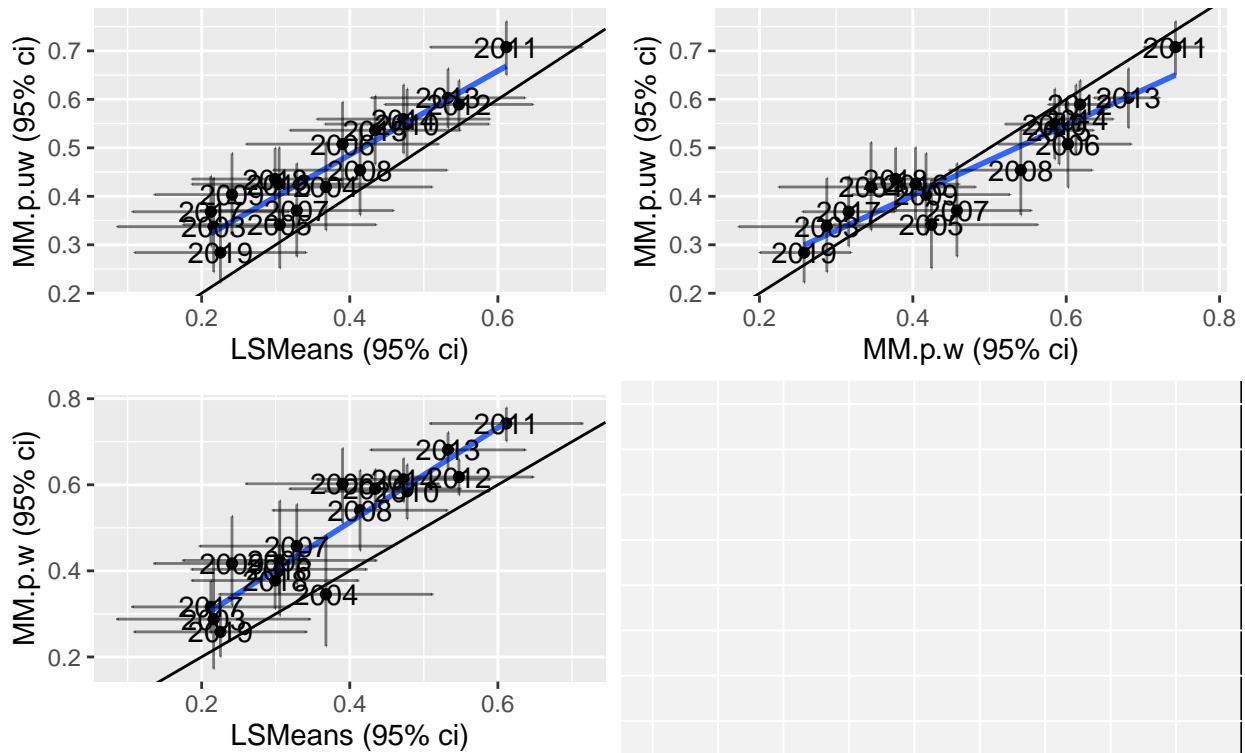
Africa : Estimated PIKE across time for Southern Africa
Comparison of LSMeans, weighted and unweighted marginal means

Africa : Estimated PIKE across time for West Africa
Comparison of LSMeans, weighted and unweighted marginal means

Here the weighted (by population abundance) estimates tend to be lower than the unweighted estimates (top right plot in each set) only for the 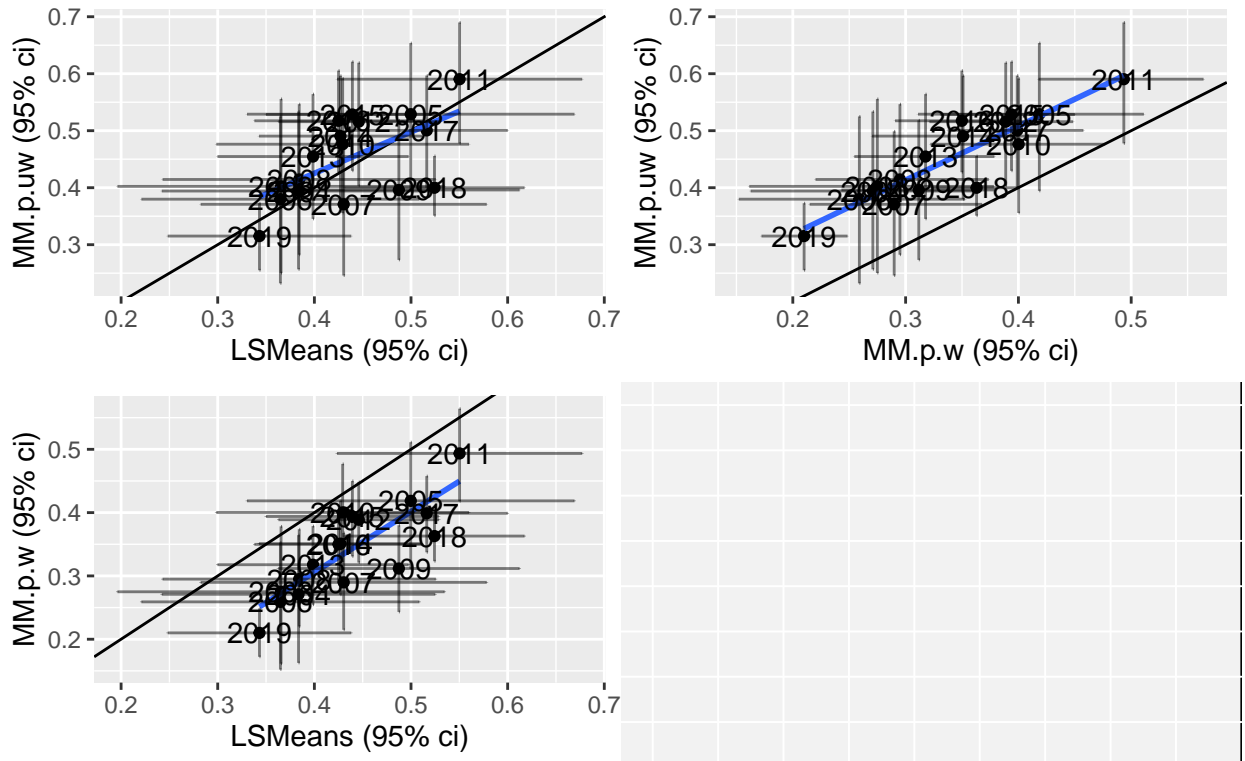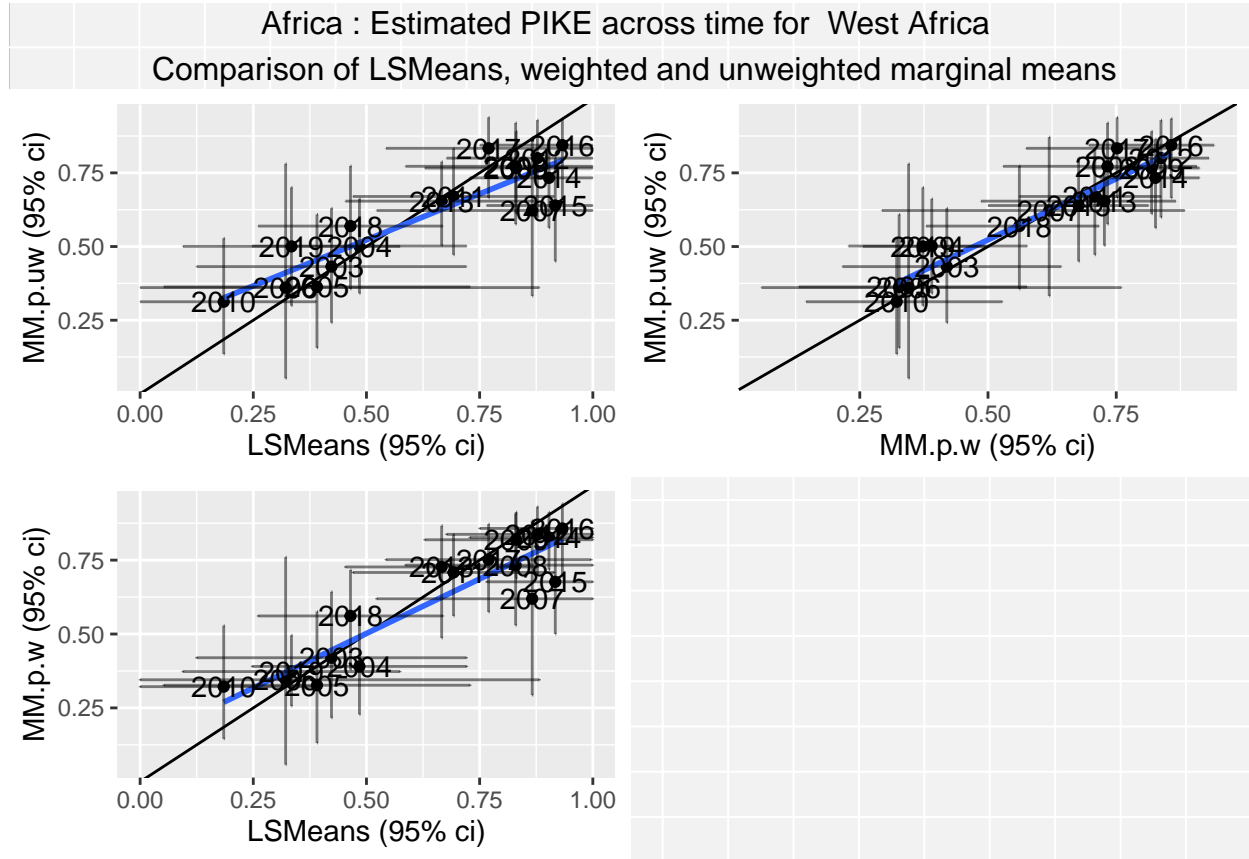South Africa sub-region. This is because in the other regions, the population sizes as the MIKE sites all tend to be similar.

The patterns when comparing the *LSMeans* estimates to the weighted estimates (top left plot in each set) is much more variable. In Central and West Africa they are similar; in Eastern Africa the *LSMeans* tend to be lower than the weighted estimates and in South Africa the *LSMeans* tend to be larger. These are artefacts of the relationship between the population abundance and the *PIKE* which is not consistent among regions.

Finally, the relationship between the *LSMeans* and the unweighted estimates (top right plot in each set) is also variable across the sub-regions and is likely an artefact of shrinkage with the mean number of carcasses examined varying considerably across the sub-regions.

7.2 Comparing the weighted and unweighted marginal mean *PIKE*

The following plots focus on comparing the weighted and unweighted marginal mean *PIKE* values:

Africa: Estimated subregional PIKE across time
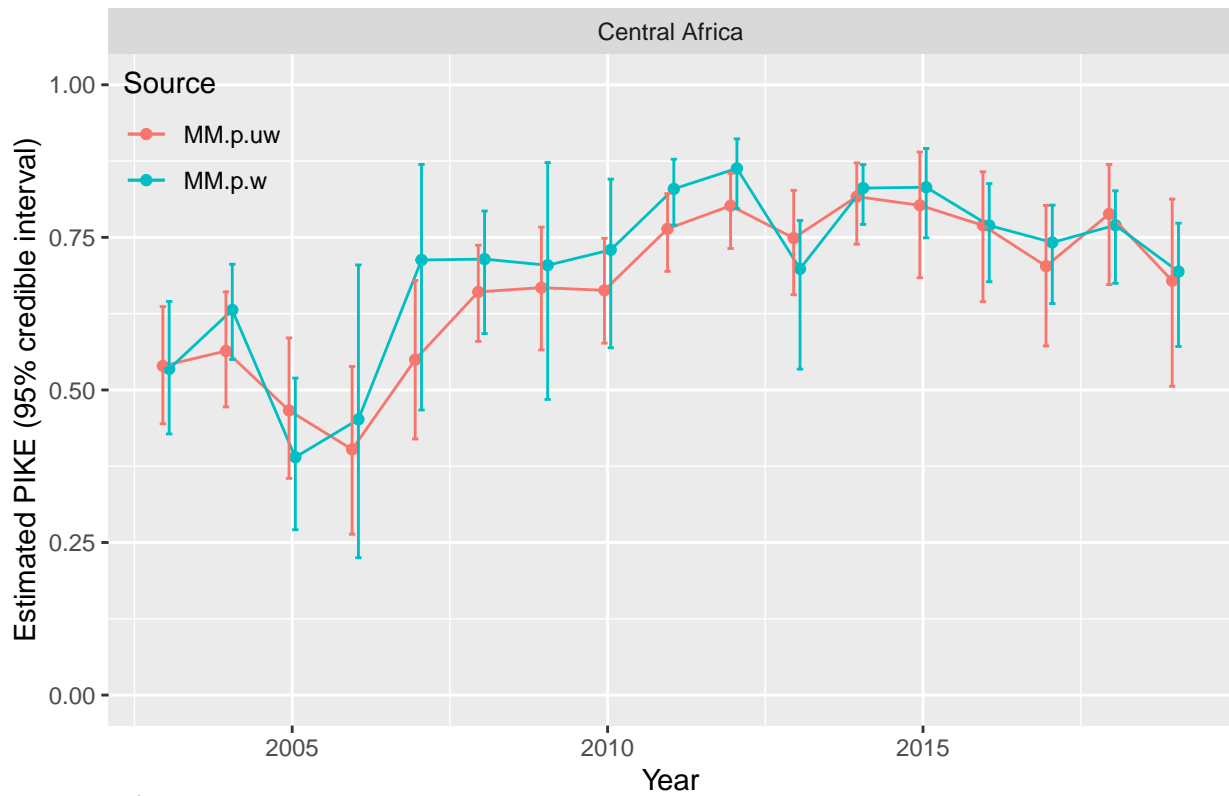
Africa: Estimated subregional PIKE across time
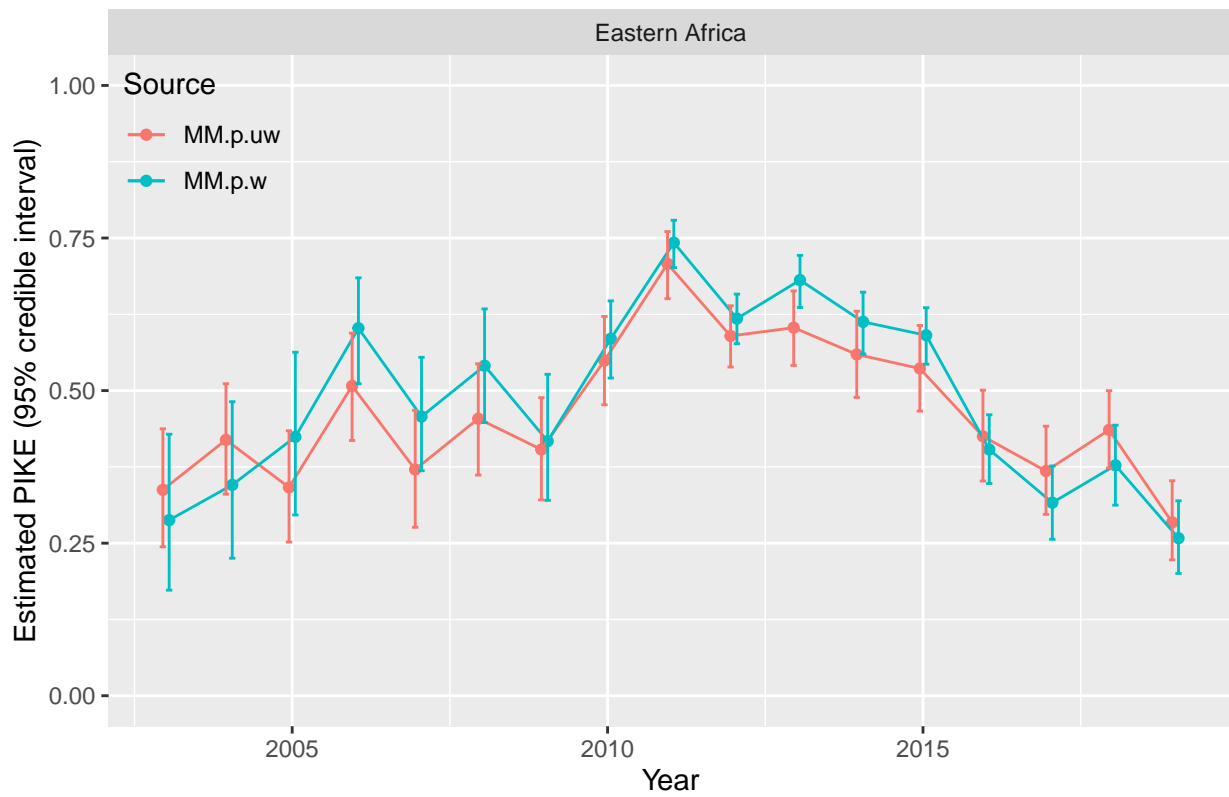
Africa: Estimated subregional PIKE across time
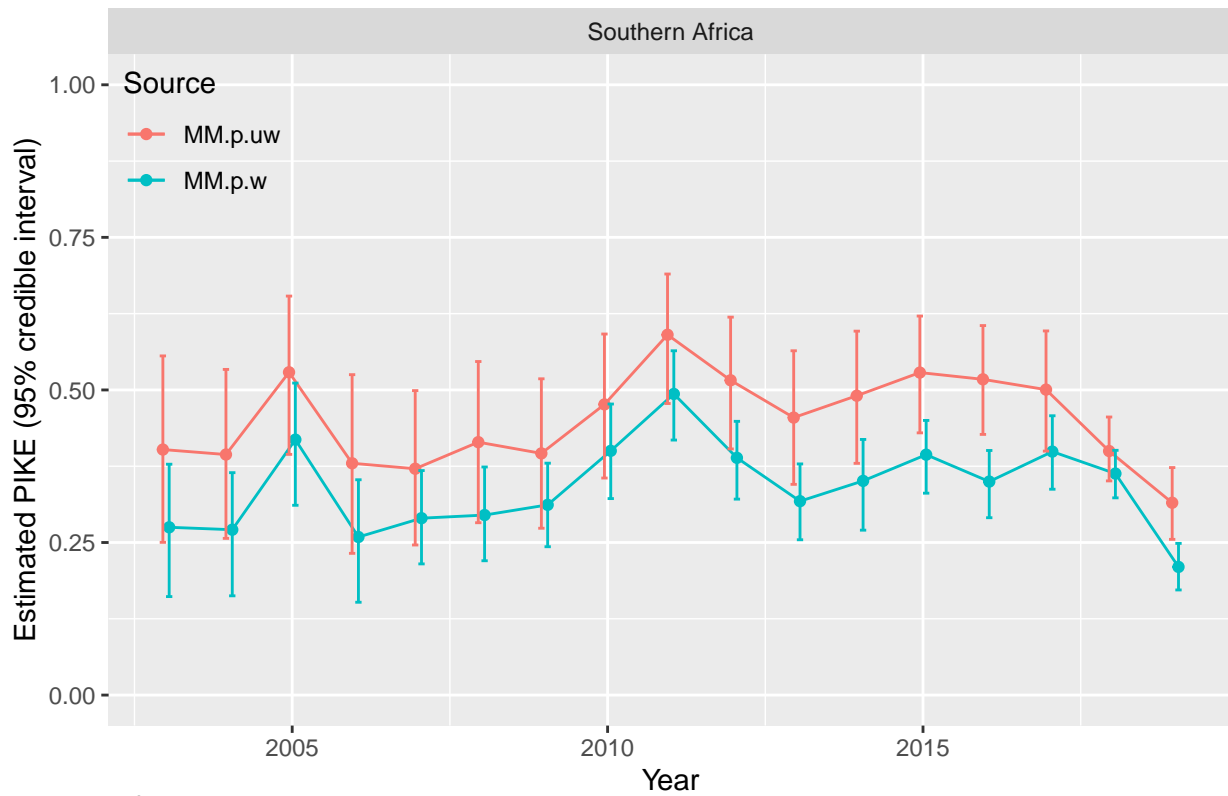
Africa: Estimated subregional PIKE across time

In Central Africa, the weighted marginal mean *PIKE* is consistently higher than the unweighted marginal

mean *PIKE* but the credible intervals overlap); in Eastern and West Africa, they track each other closely; while in Southern Africa, the unweighted marginal mean *PIKE* is consistently higher than the weighted mean *PIKE*. This "inconsistency" is a function of the relationship between the number of carcasses examined and the population abundance examined earlier in the document.

But regardless if the weighted or unweighted marginal mean *PIKE* are examined, trends are similar.

It is interesting to compare the regional trends with the continental trends
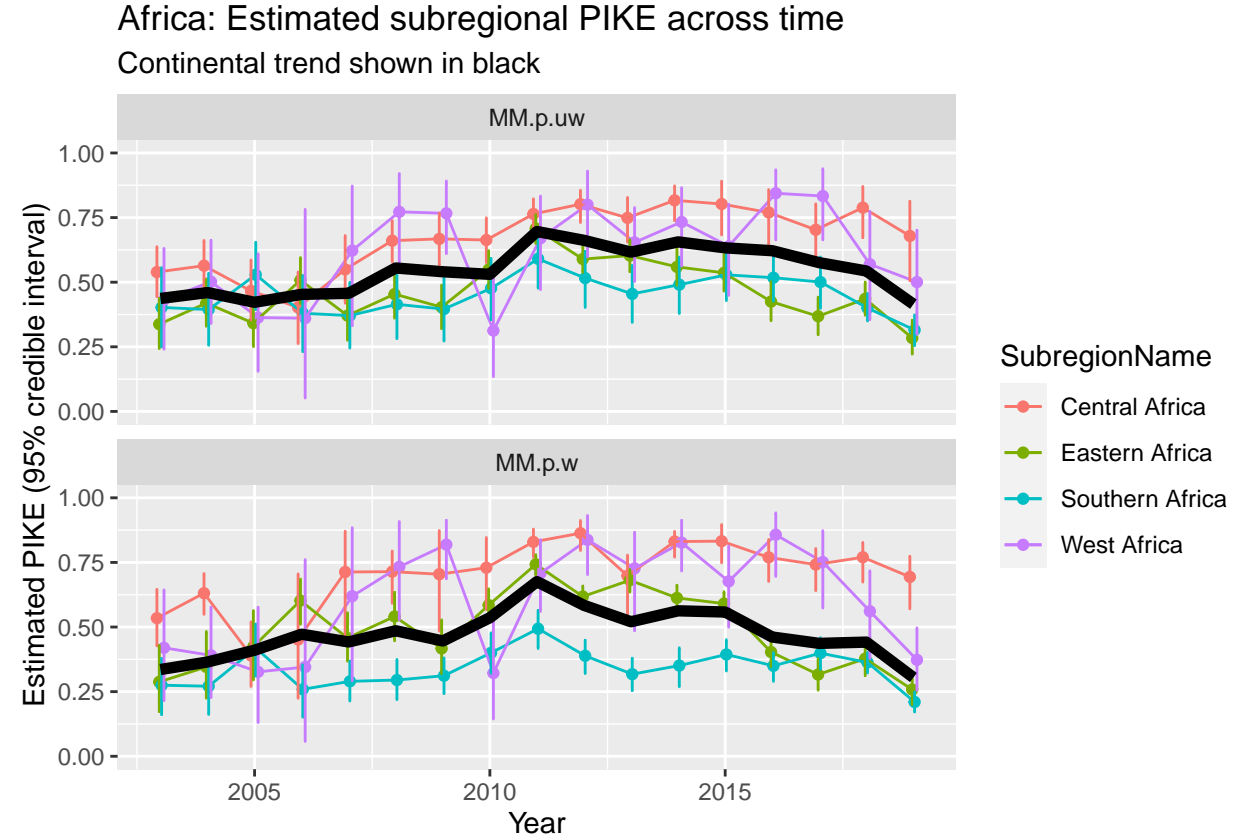
We see that *PIKE* in Southern Africa is consistently lower than the continental *PIKE*, while *PIKE* in Central Africa is consistently higher.

7.3 Posterior belief of trend in *PIKE* in last few years

Once the sample from the posterior is available, it is relatively easy to estimate the posterior belief that the trend is negative in the last 5 years in each subregion. This is done by estimating the slope in the last 5 years for each sample from the posterior, and then the posterior belief that the trend is negative is the proportion of fitted slopes that are less than zero. The posterior distribution of the slope in the last 5 years is

Africa: Posterior distribution of slope in fitted yearly PIKE in last 5 years
Based on marginal mean PIKE
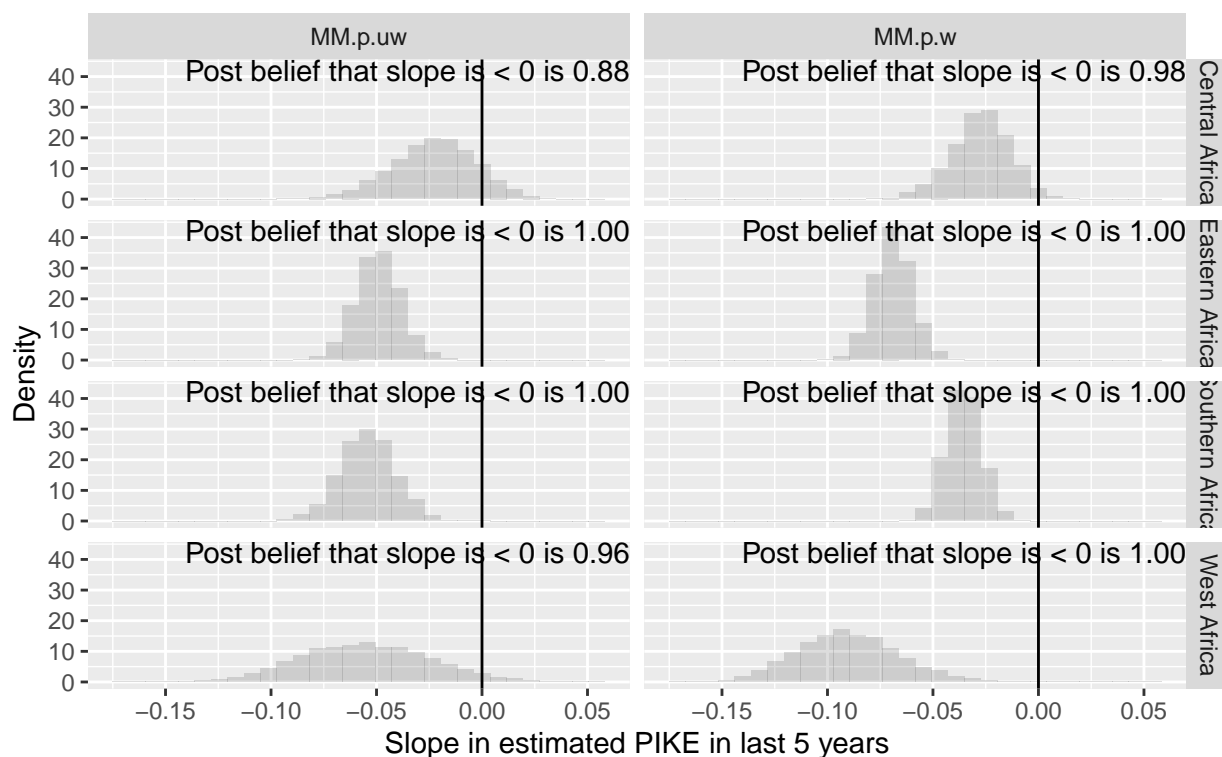
In this case, the posterior belief that the slope in PIKE is negative in the last 5 years is very high for both the unweighted and weighted *PIKE*', i.e. we have a high posterior belief that the *PIKE* in the last 5 years is declining in all subregions. However, the size of the decline varies considerably among the sub-regions with the decline in the Central Africa subregion closest to 0.

This can be visualized:

Africa: Trend in PIKE in last 5 years
Shaded area is the envelope of posterior trends
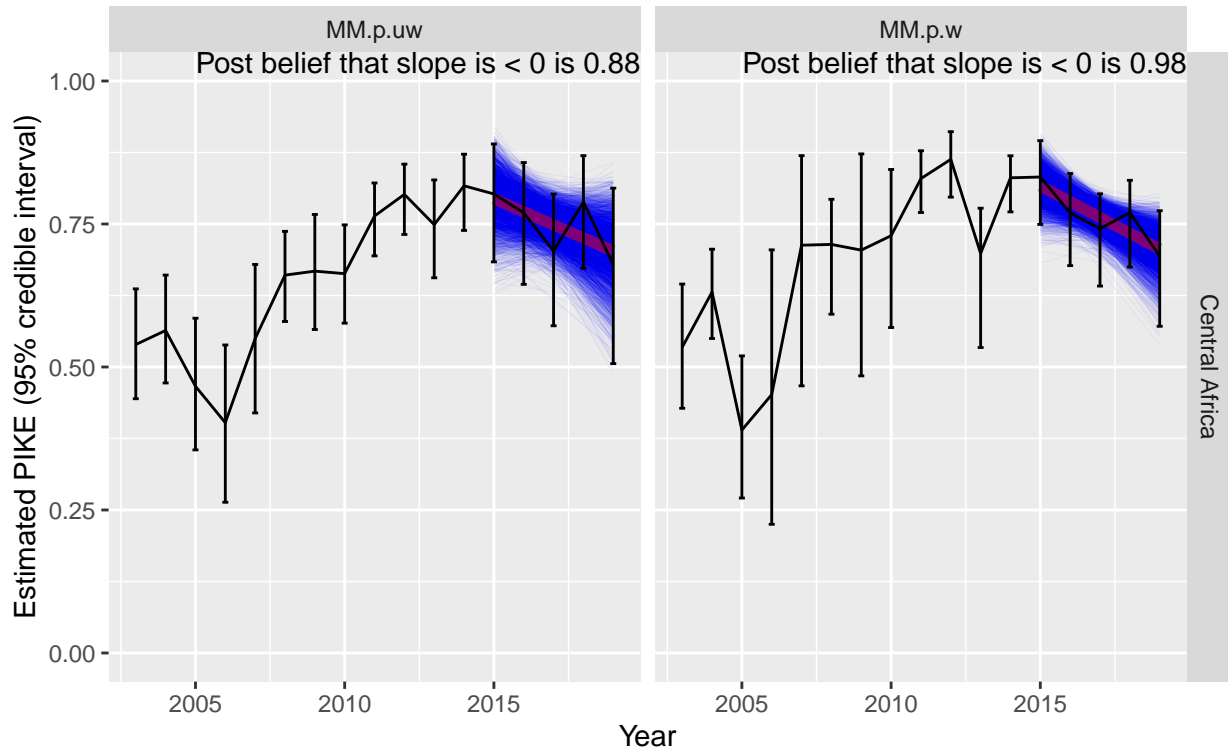
Africa: Trend in PIKE in last 5 years
Shaded area is the envelope of posterior trends
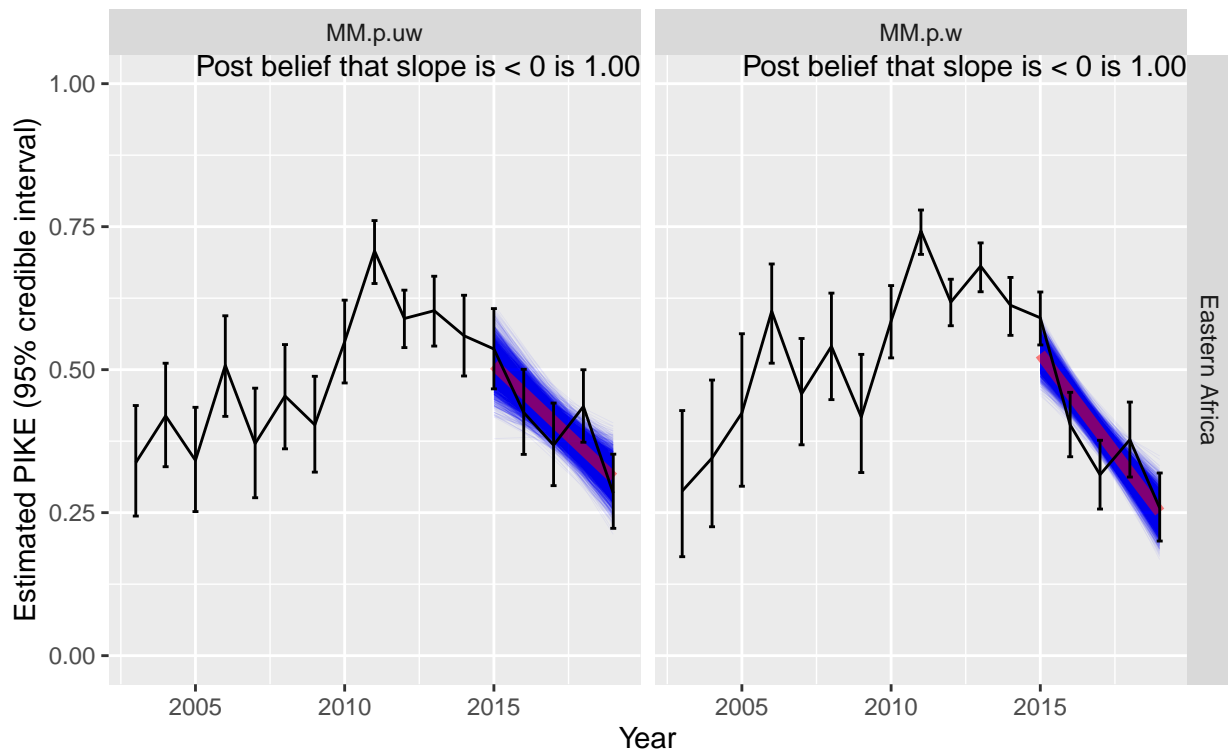
Africa: Trend in PIKE in last 5 years

Shaded area is the envelope of posterior trends
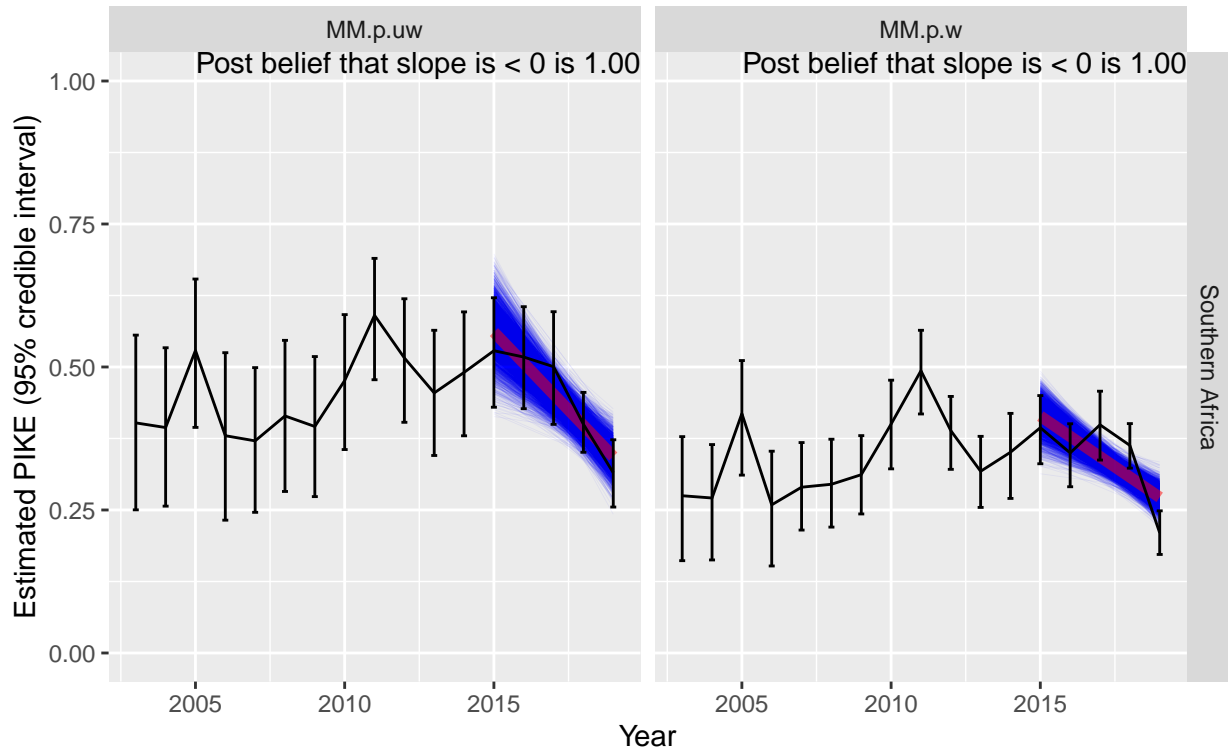
Africa: Trend in PIKE in last 5 years

Shaded area is the envelope of posterior trends
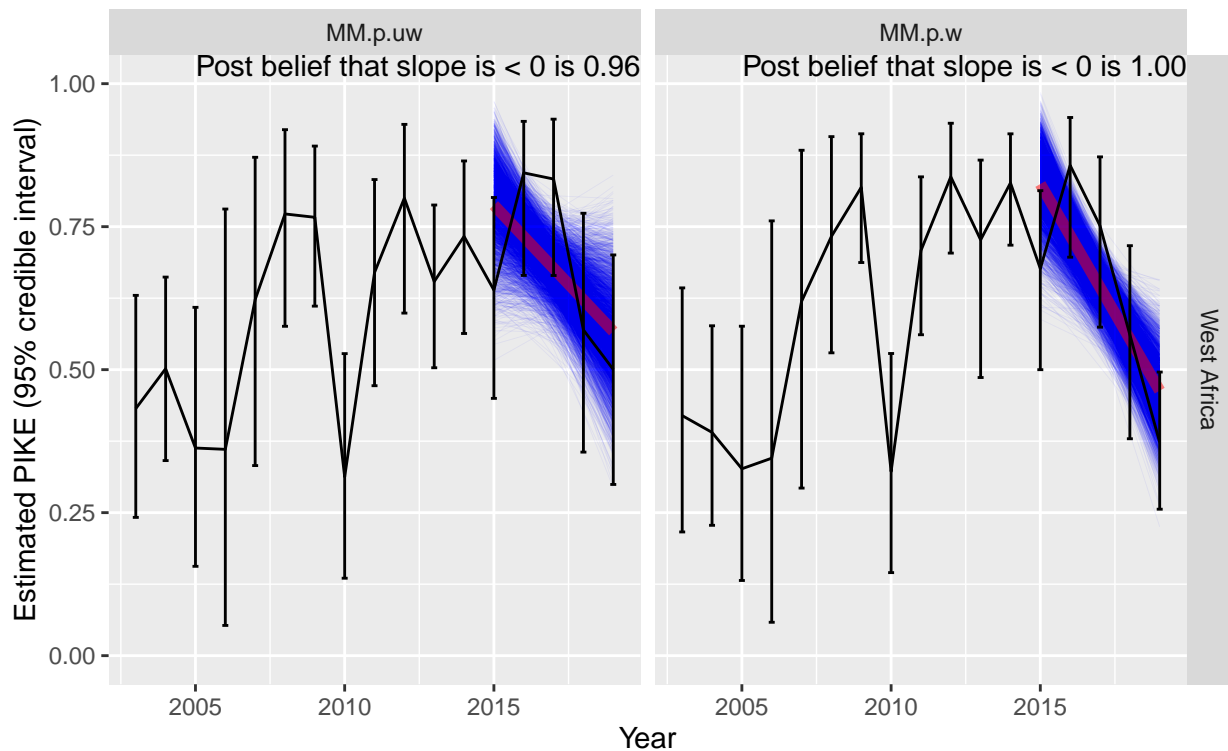
There is a strong posterior belief that the trend in the last 5 is negative, i.e. the *PIKE* is declining in the last

5 years in all subregions.

8 Model assessment

We performed model assessments of the model at the continental level and expect that similar findings will occur at the sub-regional levels.

8.1 Mixing of chains

The Gelman and Rubin's potential scale reduction factor statistic ($\widehat{R}$; Gelman et al, 2013) measures the relative variation in an estimated parameter among the multiple chains and the variation within a chain. Models should have value of $\widehat{R}$ close to 1 indicating that the posterior space covered by each chain is very similar. The effective sample size is an adjustment to the number of samples in the posterior for autocorrelation. If successive samples from the posterior have a high autocorrelation, then 10 samples from the posterior provide only incremental information over a single sample from the posterior. The effective sample should be reasonably large for all posterior samples to ensure that the posterior mean, standard deviation, and credible intervals are well estimated.

We examined $\widehat{R}$ and the effective sample size for several parameter sets:

Table 10: Rhat and Effective sample size for several parameter sets

| Effect	Max Rhat	Max N.eff	Min N.eff
SD Site effects	1.006	360	360
SD Year Site effects	1.026	87	87
Site Effects	1.065	4500	36
Year Effects	1.003	4500	800

Mixing appears to be adequate with small values of $\widehat{R}$ in all parameter sets.

The effective sample size is small (<500) for 4 sites. The sites with small effective sample sizes are:

Table 11: Sites with small effective sample sizes

MIKE site	Avg PIKE	Site effect	Rhat	N eff
BGS	0.96	2.72	1.007	340
ETO	0.01	-5.07	1.065	36
OKP	0.95	3.05	1.012	180
SGB	0.82	1.68	1.005	430

Sites with small effective sample sizes, tend to have *PIKE* that are very much larger or very much smaller than the average *PIKE* as estimated by their site effect. In particular, a site with a *PIKE* close to 0 or 1 will have a site effect with very small uncertainty and so repeated samples from the posterior will all be very similar. Mixing was adequate (as measured by $\widehat{R}$) and so these low effective sample sizes are acceptable.

8.2 Examination of trace plots

Trace plots were constructed for the yearly estimates of *PIKE* on the *logit* scale:
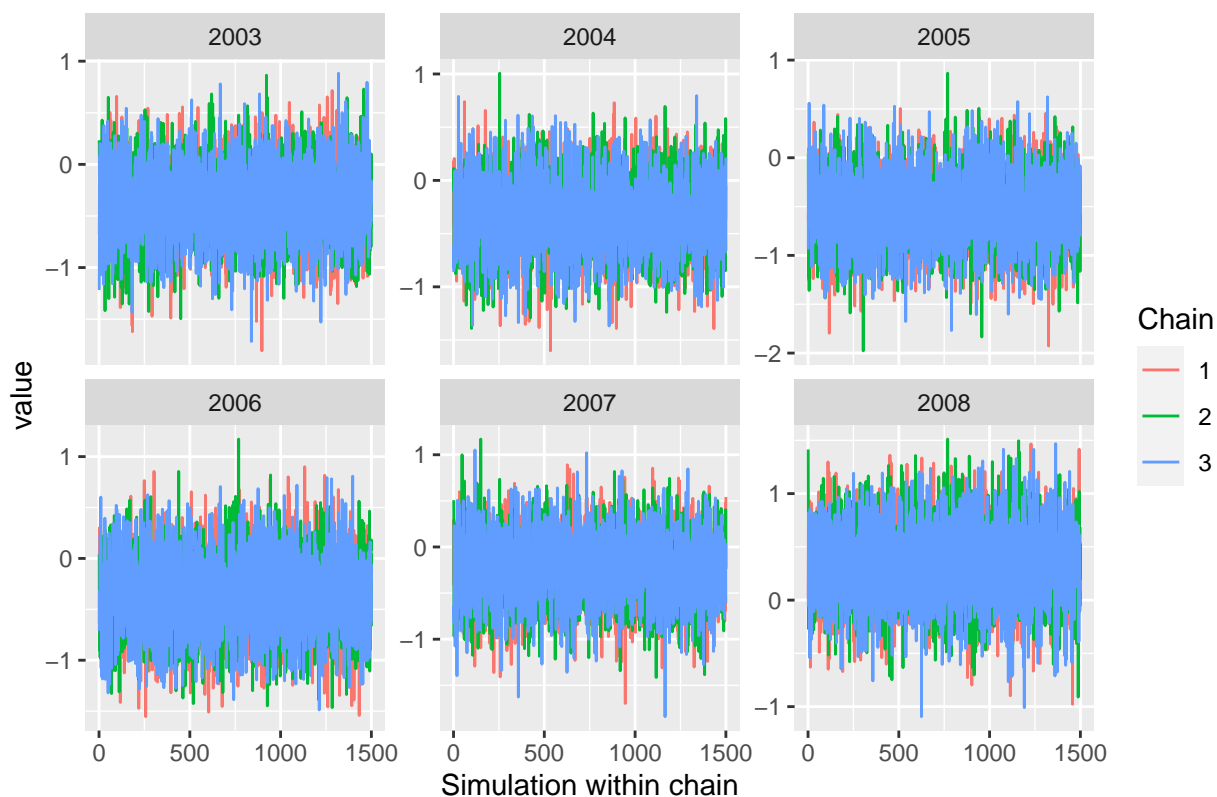
Africa : Trace plot of year effects on logit–scale

Africa : Trace plot of year effects on logit–scale
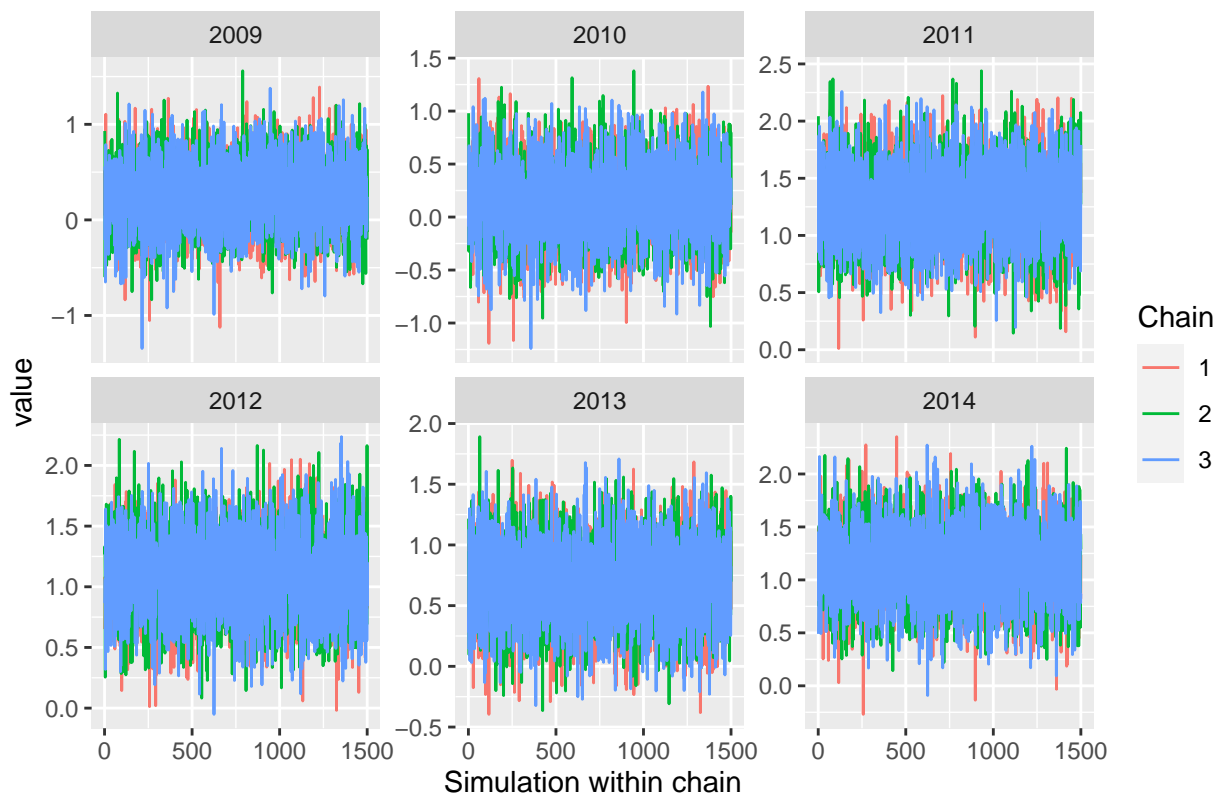
Africa : Trace plot of year effects on logit–scale
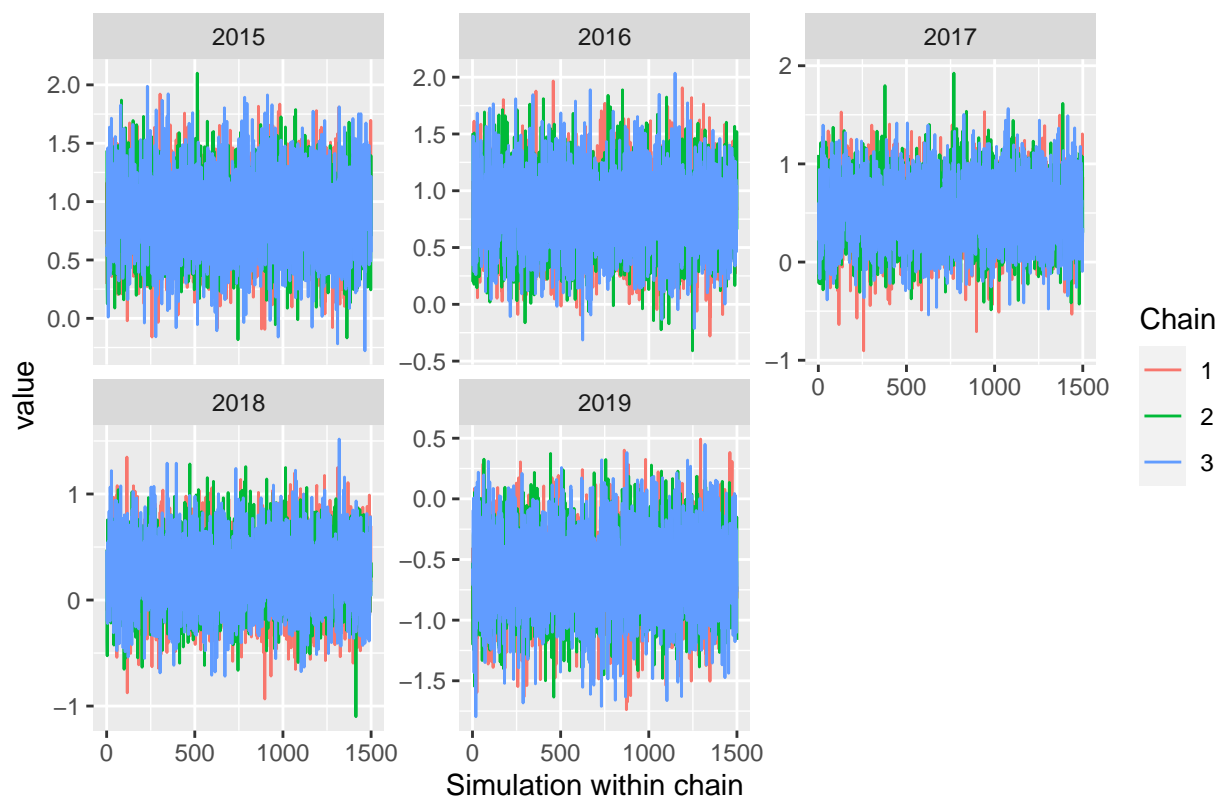
Similarly, trace plots were constructed for the estimated standard deviation of the *site* and *site.year* effects on the *logit* scale:

Africa : Trace plot of sd of site and year.site effects on logit–scale
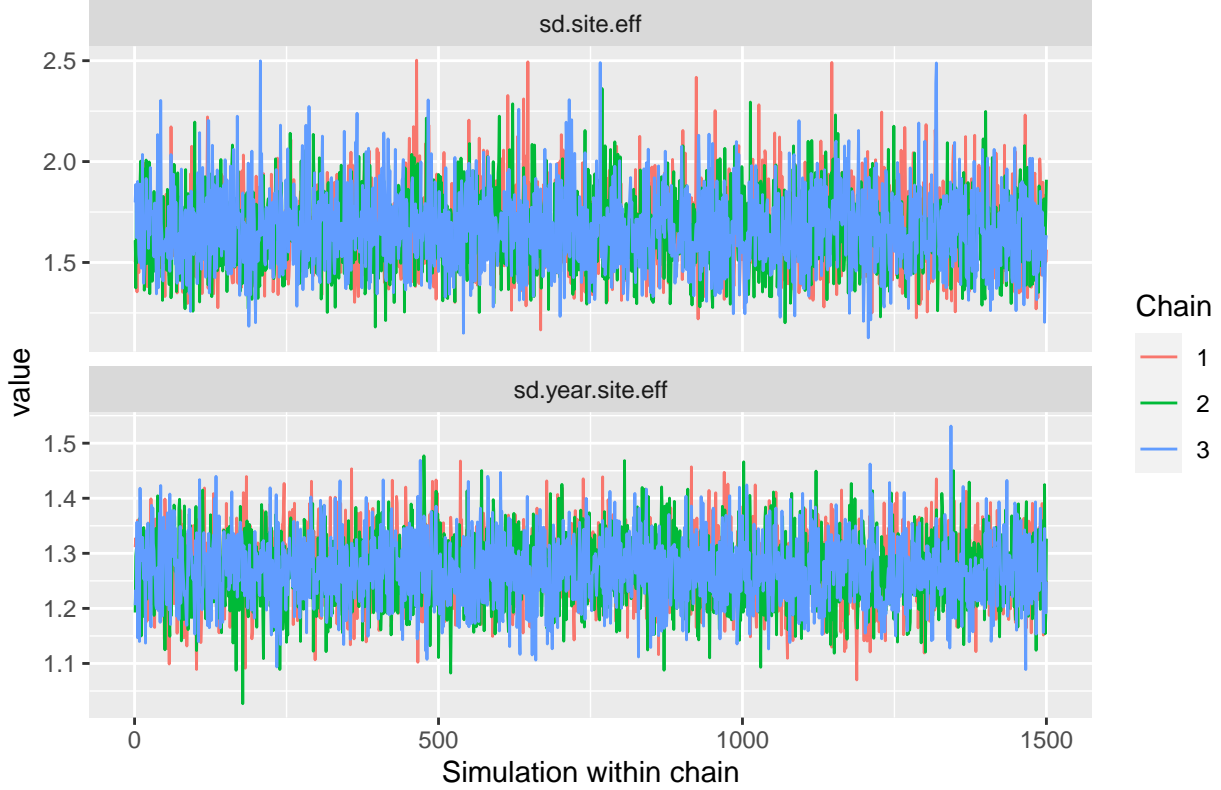
All plots show good evidence of mixing of the three chains sampled from the posterior.

8.3 Omnibus goodness of fit

An omnibus goodness-of-fit test can be constructed using Bayesian Predictive Plot (Gellman et al, 2013). For each sample from the posterior, the Tukey-Freeman statistic (Freeman and Tukey, 1950) is computed using the observed data and a simulated data based on the posterior sample. The Tukey-Freeman statistic is less sensitive to small observed and expected values than the usual chi-square goodness-of-fit test.

For example, for a particular value of the posterior sample, the observed Tukey-Freeman statistic is found as the difference between the observed number of illegally killed elephants and the expected number of illegally killed elephants:

$$TF.obs = \sum_{site.years} (\sqrt{IC_{site.year}} - \sqrt{TC_{site.year} \times \pi_{site.year}})^2$$

The simulated Tukey-Freeman statistic is found as the difference between a simulated number of illegally killed elephants and the expected number of illegally killed elephants:

$$IC.sim_{site.year} \sim Binomial(TC_{site.year} \times \pi_{site.year})$$

$$TF.sim = \sum_{site.years} (\sqrt{IC.sim_{site.year}} - \sqrt{TC_{site.year} \times \pi_{site.year}})^2$$

The value of the $TF.obs$ is plotted against the corresponding $TF.sim$ and the proportion of times that the observer Tukey-Freeman statistic exceeds the simulated Tukey-Freeman statistic is known as the Bayesian p-value. If the model fits well, then these two measures should be similar and the Bayesian p-value will be close to 0.5. If there is lack of fit, then the two measures will be discordant, and the Bayesian p-value will be close to 0 or 1.

The Bayesian Posterior Predictive plot for the omnibus goodness of fit is:

Africa: Omnibus goodness of fit

Reference line of equality shown

Bayesian p–value: 0.86
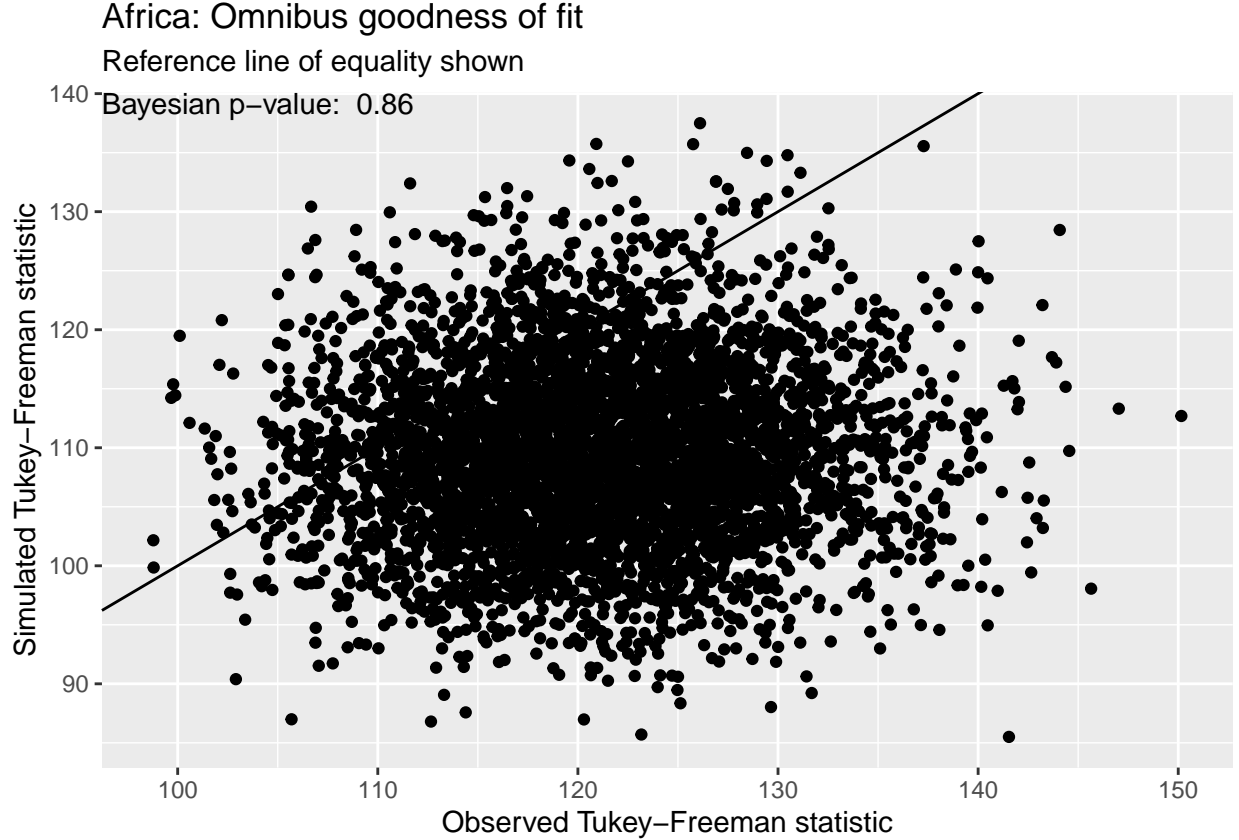
Because the Bayesian p-value is not extreme, the fit is deemed acceptable;

8.4 Over dispersion

8.4.1 General over dispersion

A general measure of over dispersion is to compute a statistic that compares the expected number of illegally killed elephants based on the fitted site-year *PIKE* with the observed number of illegally killed elephants.

$$Disperson = \sum_{sy} \frac{(TC_{sy} \times \widehat{\pi}_{sy} - IC_{sy})^2}{TC_{sy} \times \widehat{\pi}_{sy}}$$

There are 711 site-year data points in the sum above.

This is traditionally divided by the (*number of data points − the number of estimated parameters*). However, in Bayesian hierarchical models (such as this), the number of parameters is ill-defined. For example, we model site-effects as random variables from a common distribution. Is the number of parameters 2 (the mean and variance of the common distribution) or is it the number of sites (we need to estimate the individual site effects). Furthermore shrinkage in Bayesian models implies that the effective number of site estimates is smaller than the number of sites. A similar problem occurs with the site-year effects.

If you count the individual year effects, the individual site effects, and the individual site-year effects as separate parameters, this gives a total parameter count of 789 which is more than the number of data points.

The Bayesian output includes a measure pD defined as the effective number of parameters, i.e. after accounting for shrinkage. We obtain pD=672.9 which is considerably less and accounts for shrinkage (Spiegelhalter et al. 2002). This gives an over dispersion value of

$$OD = \frac{Dispersion}{\# \ data \ points - pD}$$

which gives $OD = 2.9$. This value is slightly above 2 indicating some evidence of over dispersion, but generally speaking is acceptable.

Some of the expected number of illegally killed elephants are very small which can inflate the numerator. A histogram of the individual components of the *Dispersion* numerator:

Observed components of Dispersion numerator
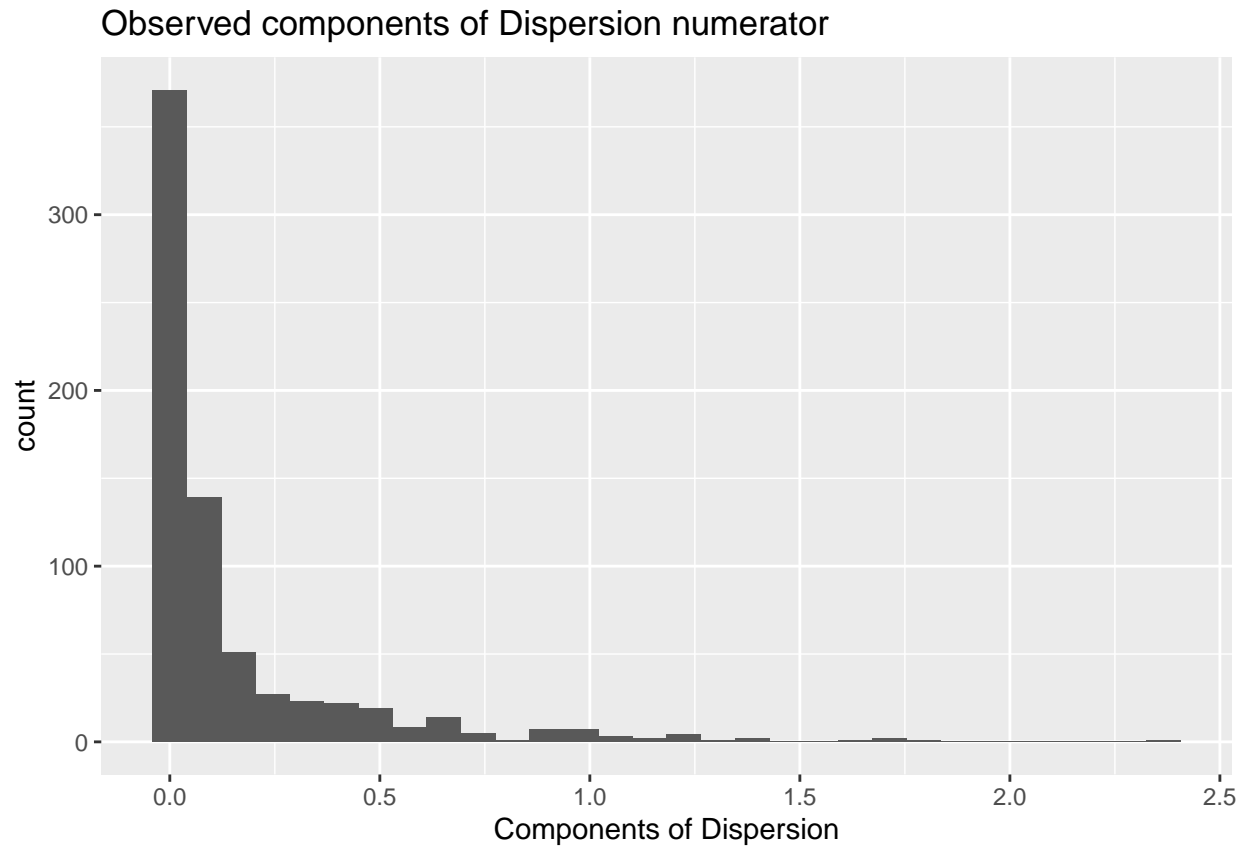
shows that the fit is generally good, with only a few site years where the contribution is large. The (few) site-years where the observed dispersion component is > 1 are shown below and are acceptable in terms of goodness of fit.

Table 12: Site-years with largest discrepancy in fit

Site ID	Year	Total number of carcasses	Number of Illegal Carcasses	Observed PIKE	Estimated PIKE	Estimated Number of Illegal Carcasses	Contribution to dispersion
PDJ	2010	6	0	0.00	0.17	1.02	1.02
ZIA	2012	2	0	0.00	0.51	1.02	1.02
SMN	2018	8	0	0.00	0.13	1.04	1.04
WAZ	2012	5	0	0.00	0.21	1.05	1.05
VIR	2004	33	1	0.03	0.08	2.75	1.11
WBF	2019	4	0	0.00	0.28	1.14	1.14
NYA	2018	10	0	0.00	0.12	1.20	1.20
BBK	2006	12	0	0.00	0.10	1.20	1.20
PDJ	2019	18	0	0.00	0.07	1.22	1.22
ETO	2003	21	1	0.05	0.02	0.35	1.24
MAG	2009	5	0	0.00	0.27	1.34	1.34
GSH	2003	4	2	0.50	0.22	0.90	1.36
ZIA	2010	6	0	0.00	0.23	1.38	1.38
NDK	2013	10	0	0.00	0.16	1.61	1.61
TAI	2004	2	2	1.00	0.41	0.82	1.72
CHO	2016	121	0	0.00	0.01	1.74	1.74
GOU	2010	27	0	0.00	0.07	1.81	1.81
ODZ	2005	73	0	0.00	0.03	2.37	2.37

These generally occur when no illegally killed elephants are reported with an intermediate number of total carcasses reported where the model predicts a non-zero *PIKE*. Refer to the earlier sections to look at the individual sites reported here.

8.4.2 Overdispersion due to 0 counts

The omnibus test is a general goodness-of-fit measure. The same logic can be used to investigate specific aspects of the fit. In particular, the number of times that the number of illegally killed elephants is reported as 0 is examined.

There were 122 cases over all sites and all years where the number of illegally killed elephant carcasses was reported as zero. After fitting the model, for each sample from the posterior, we simulate the number of illegally killed elephants in the same way as in the omnibus goodness of fit:

$$IC.sim_{site.year} \sim Binomial(TC_{site.year} \times \pi_{site.year})$$

and count the number of times a count of 0 is obtained. This is compared to the observed number of times a 0 is obtained.

Africa: Histogram of posterior sample of 0 counts
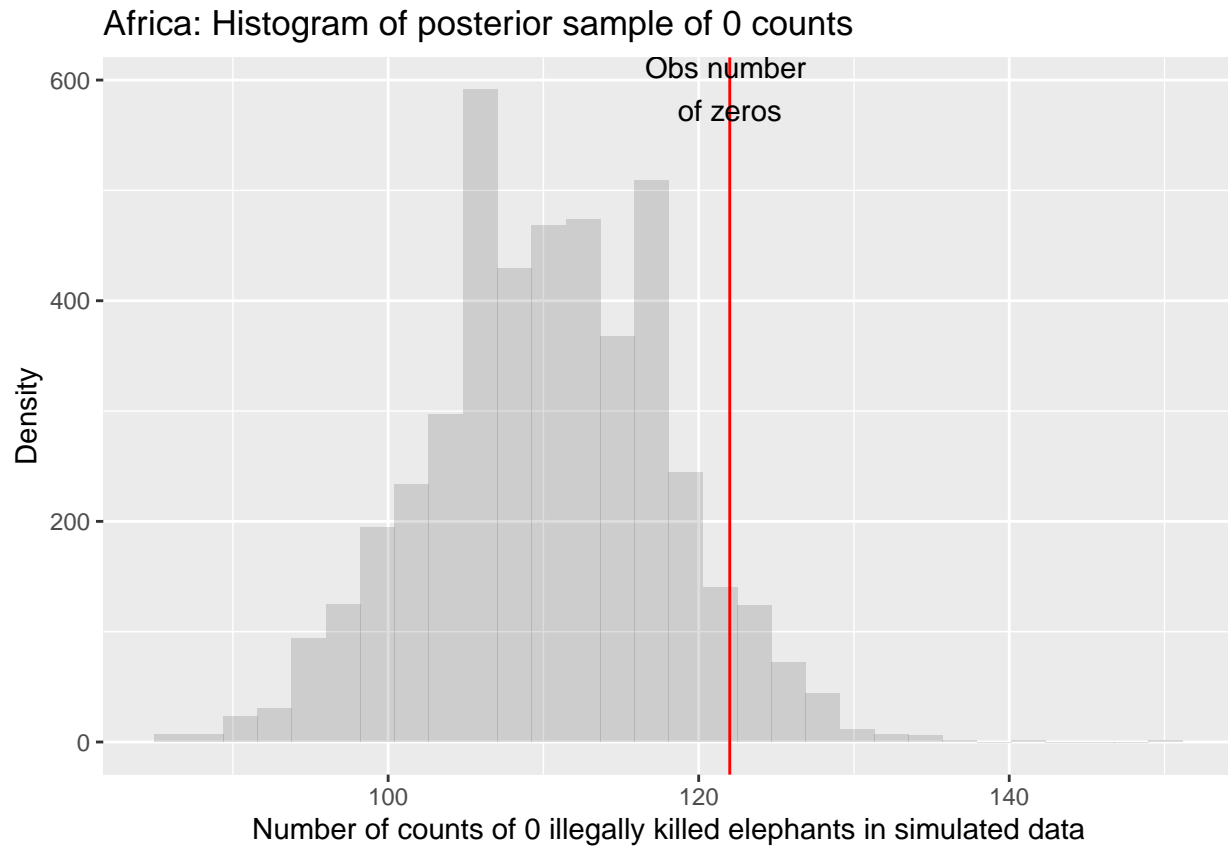
The number of 0 counts is on the higher side, but not unusual relative to that seen from simulated data.

8.5 Spatial correlation among site effects

The (random) site effects have been modelled as independent random effects without explicitly accounting for the spatial structure of the data. However, we find that sites that are close geographically have similar site effects.

Africa: Estimated site effects
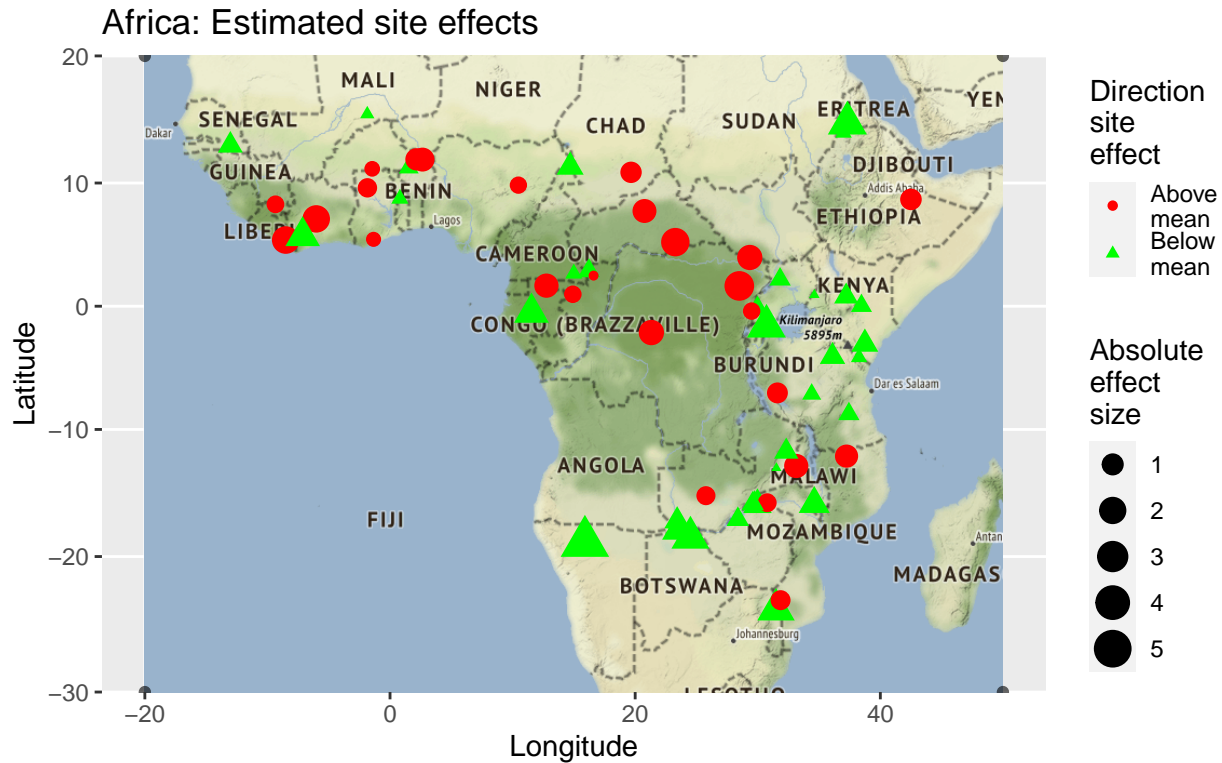
Sites that have *PIKE* consistently above the continental average are labelled as *Above the mean*; sites that have *PIKE* consistently below the continental average are labelled as *Below the mean.*

We notice that sites that are close geographically tend to have similar site effects (size of dot) and in the same direction (above or below the mean, color of dots). This implies there is a spatial correlation among the site effects that has not been directly accounted for in the analysis.

The current analysis is still valid, but inefficient because it has not used the spatial correlation to improve inference. If spatial autocorrelation is explicitly modelled, then information is shared among sites that are geographically close, i.e., if the *PIKE* increases in one site, then spatial autocorrelation would imply that it would tend to also increase in a nearby site. Of course, if the sites are in different countries with different levels of enforcement or other covariates that impact *PIKE*, an explicit spatial autocorrelation could introduce a spurious relationship between the *PIKE* in the two sites unless these other factors (law enforcement etc.) are also modelled. The explicit spatial autocorrelation models rapidly become more complex to account for these features.

Because the current analysis treats all sites as independent (rather than spatially correlated), the uncertainty in the overall yearly *PIKE* is slightly smaller than from a model with explicity spatial autocorrelation because the effective number of sites used in computing the overall yearly *PIKE* is smaller when autocorrelation is explicitly modelled. This in turn, implies that the uncertainty of a trend (e.g. trend in the last five years) in the currently model may be slightly understated as well and the posterior belief in a trend will be higher in the current model compared to the model with an explicity spatial autocorrelation. We believe such effects are minor given the spase data at many sites, the large amount of missing site.years and the potential breaking of spatial autocorrelation across country borders.

A potential improvement to the current analysis may be to add another level of random effects (country effects) so that points from the same country that have related site effects then experience a common country effect. This model is currently under investigation.

8.6 Site effects vs number of carcasses observed

A plot of the estimated site effects vs. the total number of carcasses observed over the year is:

Site effects vs. total sample size
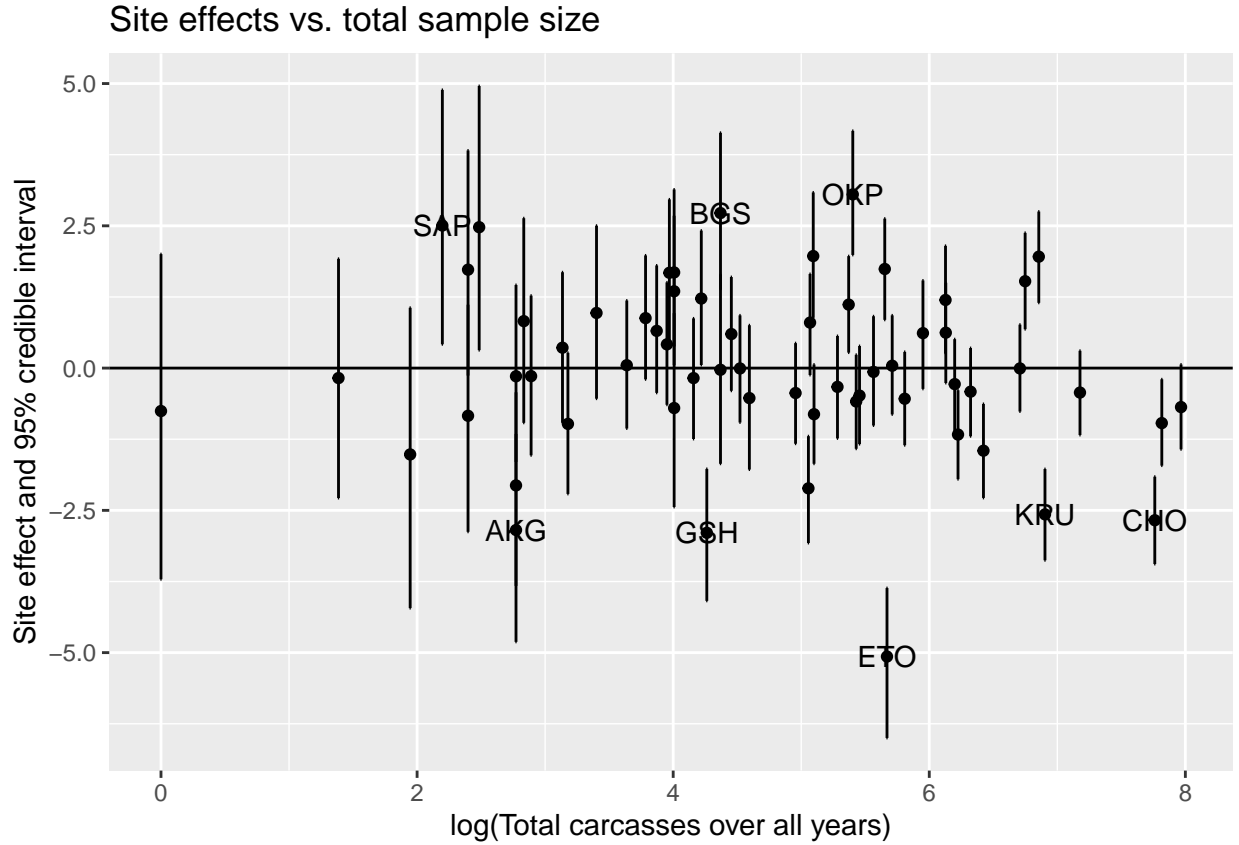
This plot shows that the uncertainty in the site effects declines with the total number of carcasses observed (as expected), and a random scatter about 0 (also as expected). There are a few MIKE sites with extreme site effects as labelled in the plot.

8.7 Autocorrelation in year.site effects

This model assumes that $Year.Site$ effects are independent from year-to-year. However, local effects may last for several years, and so there may be autocorrelation present in the $Year.Site$ effects.

A plot of the $Year.Site_i$ vs. the $Year.Site_{i-1}$ (i.e. a lag 1 plot) is:

Africa: Lag 1 plot of Year.Site effects

Correlation: 0.26

Year.Site effect in year i

Year.Site effect in year i−1
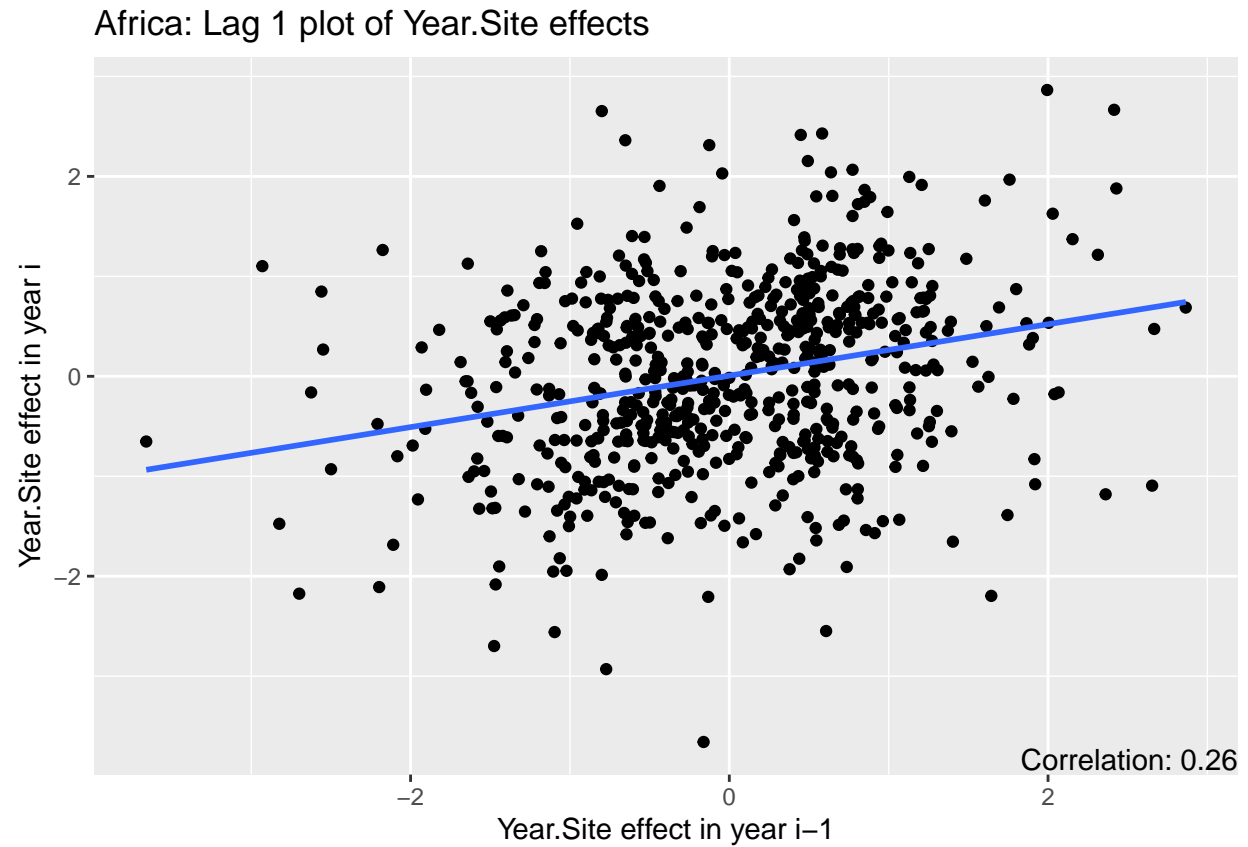
shows a very model correlation over time which is sufficiently small that is not a problem. Note that only those site-years where data are collected are used in the above plot.

8.8 Year.Site effect for each site

A plot of the $Year.Site$ effect for each site:

Africa: Year.Site effects by site
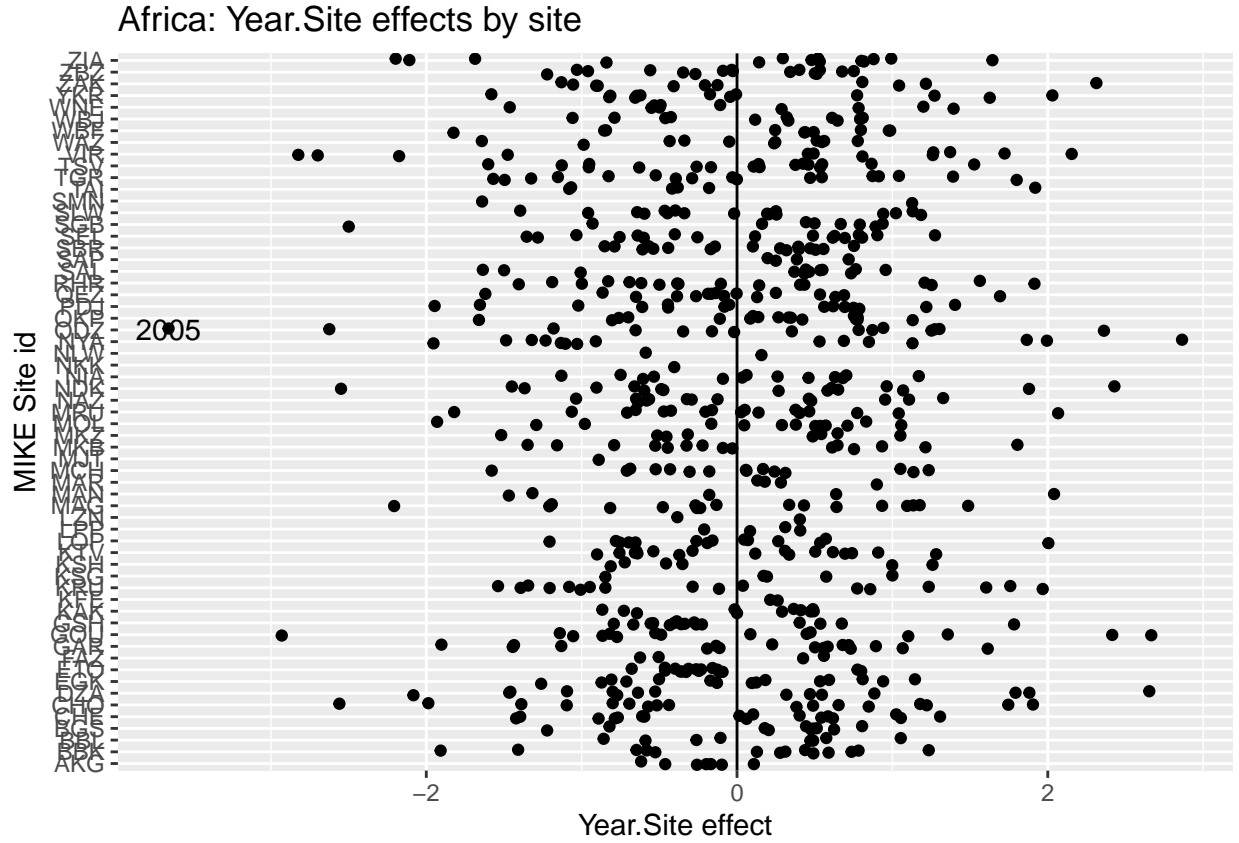
shows that only a few years had *PIKE* values within a site that could be considered unusual for that site.

8.9 Fit of a simpler model

A simpler model was also fit ignoring the *Year.Site* random effects, i.e.

$$logit(\pi_{sy}) = Year_y + Site_s(R)$$

This model is closer in form to the model of Burn et al (2011).

The Bayesian Posterior Predictive plot for the omnibus goodness of fit for this simpler model is:

Africa: Omnibus goodness of fit for model with no Site Year effect
Reference line of equality shown
Bayesian p–value: 1.00
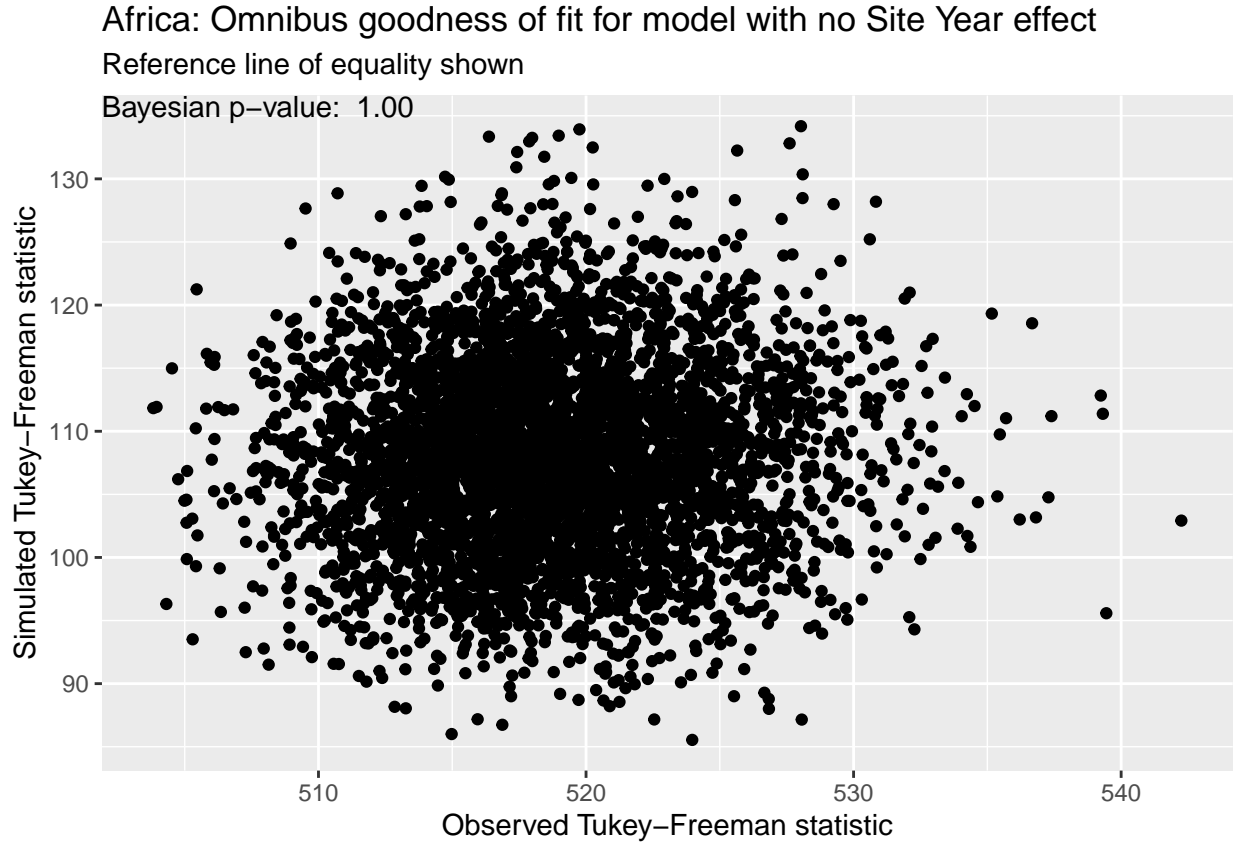
which indicates that this model has very evident lack of fit (the Bayesian p-value is very extreme).

A comparison of the DIC from the two models confirms this:

```
DIC.Full.Model DIC.No.Site.Year
      2788.376          4605.831
```

The full model has overwhelming support with a much smaller DIC compared to the simpler model.

However, Millar (2009) showed that DIC computed for hierarchical models conditions on the latent random effects (*site* and *site-year* in our model) and needs to be integrated over these random effects. If the integration is not carried out, then comparisons using DIC could be misleading. Alternative measures of model comparison, e.g. Bayes-factors or WAIC are being investigated, but given the very poor fit of the simpler model, we do not believe that our conclusions that the simpler model is not adequate will change.

While this simple model is closer in form to that of Burn et al (2011), this goodness-of-fit comparison is not a comparison with the Burn et al (2011) model. The Burn et al (2011) model was for a shorter time period and included country effects and covariates which may capture some of the differences between countries and sites. Therefore the poor fit of the simpler model does not indicate that the Burn et al (2011) model was a poor fit.

8.10   Observed vs. predicted *PIKE*

A plot of observed *PIKE* in each year.site vs. the predicted *PIKE* is:

Africa : Observed vs. Predicted PIKE
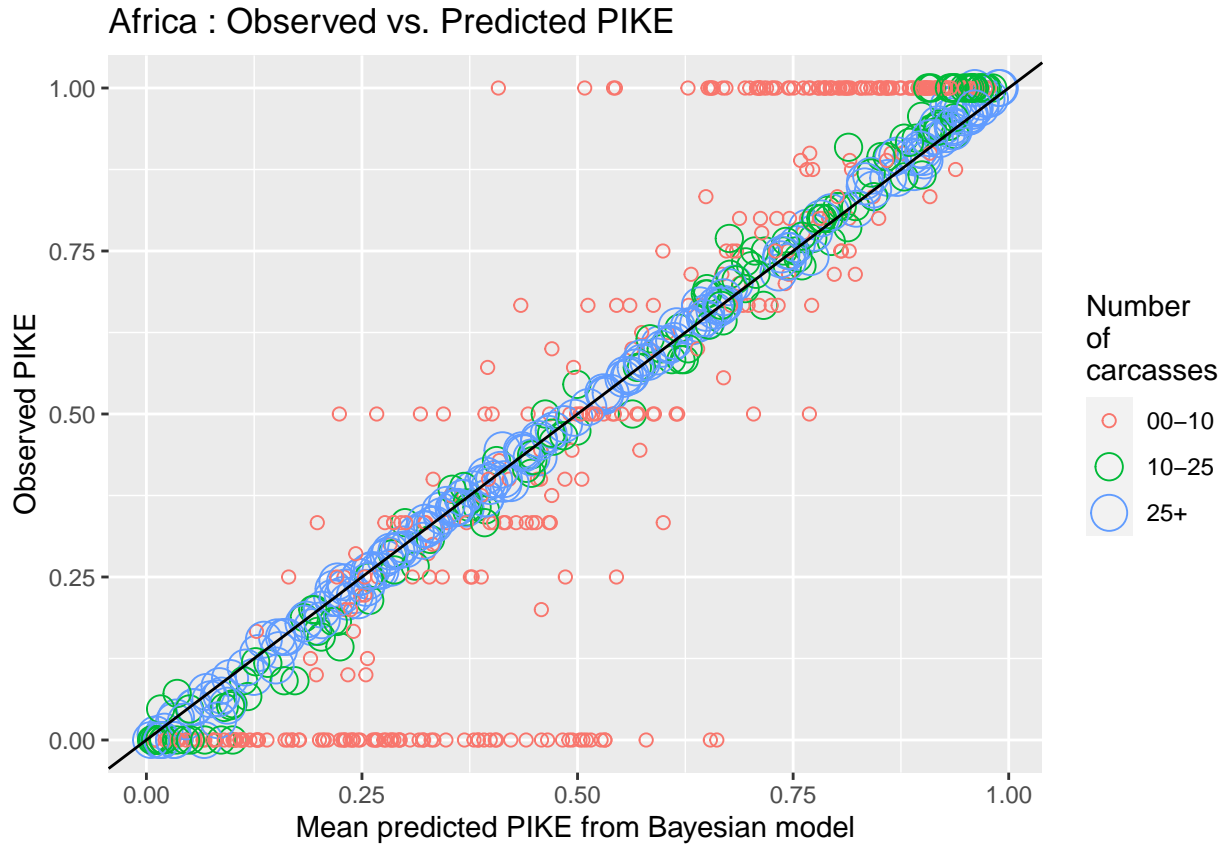
The fit is generally very good. For site-years where the number of carcasses was very small (< 10) and the observed *PIKE* was 0 or 1, the estimated *PIKE* is pulled towards the yearly average for that year. For site-years with large number of carcasses (> 25) the estimated *PIKE* matches closely with the observed *PIKE*. For site-years with intermediate number of carcasses, the estimates are shrunk slightly towards the mean for that year.

This can also be seen in the plots of observed and fitted *PIKE* for the individual MIKE sites:

Africa : Observed and predicted PIKE for individual MIKE sites

Dashed line is unweighted marginal mean PIKE at continental level
Blue and shading is predicted PIKE at site level with 95% credible interval

Africa : Observed and predicted PIKE for individual MIKE sites

Dashed line is unweighted marginal mean PIKE at continental level
Blue and shading is predicted PIKE at site level with 95% credible interval

Africa : Observed and predicted PIKE for individual MIKE sites

Dashed line is unweighted marginal mean PIKE at continental level
Blue and shading is predicted PIKE at site level with 95% credible interval

Africa : Observed and predicted PIKE for individual MIKE sites

Dashed line is unweighted marginal mean PIKE at continental level
Blue and shading is predicted PIKE at site level with 95% credible interval

There are several interesting patterns that illustrate the features of the model.

- Site *CHO*. This site reports nearly every year with a large number of carcasses (large blue circles). The estimated yearly site-level *PIKE* closely follows the observed data (as expected).

- Site *BBK*. In some years, this site reports a large number of carcasses and the estimated yearly site-level *PIKE* for these years matches the observed *PIKE*. In some years, the observed *PIKE* based on large number of carcasses is above the continental marginal mean (e.g. 2003) and in some some years it is below the continental marginal mean (e.g. 2006). On average, this site tracks the continental trend. So in years where this site only reports on a few carcasses (small red circles) and the observed *PIKE* is mostly 0 or 1, the estimated *PIKE* is close to the continental trend. For example, with a small number of carcasses examined, a value of 2 illegally killed elephants from 2 carcasses examined (observed *PIKE* of 1) is consistent with an estimated *PIKE* closer to the continental marginal *PIKE*. Notice that in years with a small number of carcasses reported, the credible interval for the estimated site-level *PIKE* is very wide.

- Site *AKG*. This site mostly has smallish sample sizes, but the observed *PIKE* is consistently close to 0. The estimated site-level *PIKE* is then also close to 0 in years with no reports, but notice the wide credible intervals.

- Site *KFE*. This is a new MIKE site with data only starting in 2018. In this year, the observed *PIKE* was slightly higher than the continental marginal mean *PIKE*, and so the estimated site-level *PIKE* for earlier years is slightly above the continental trend, but notice the wide credible intervals.
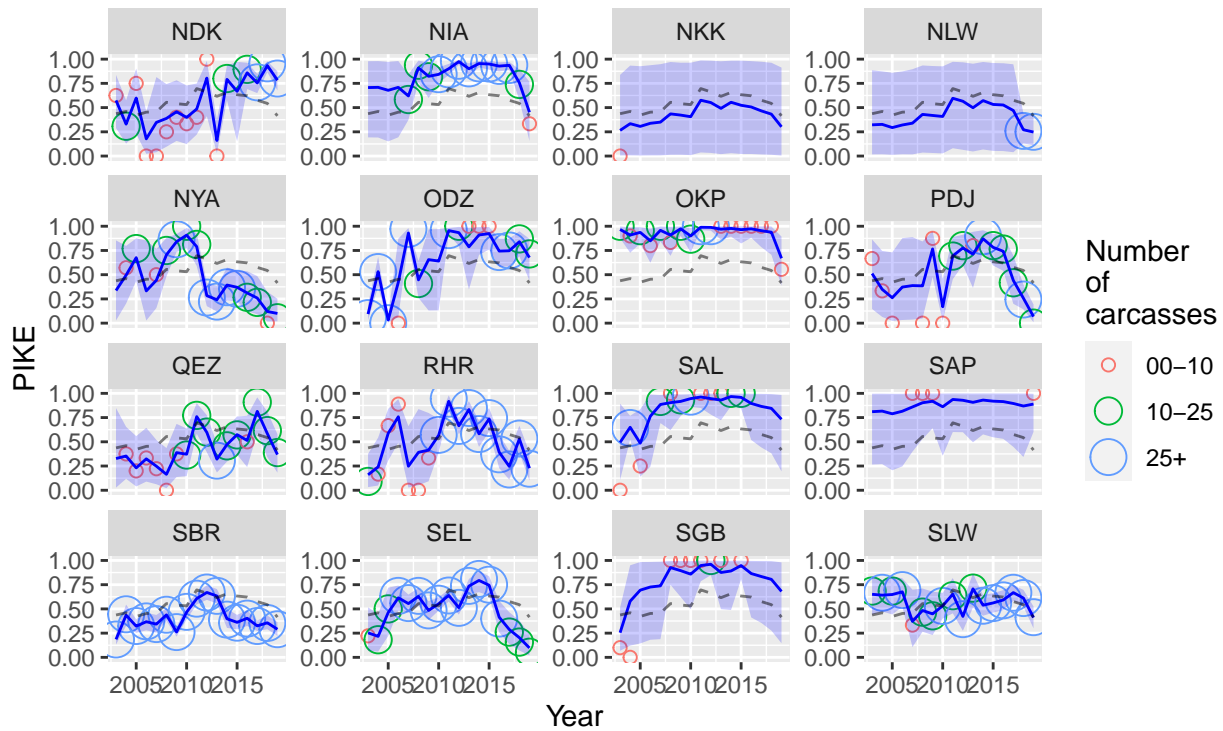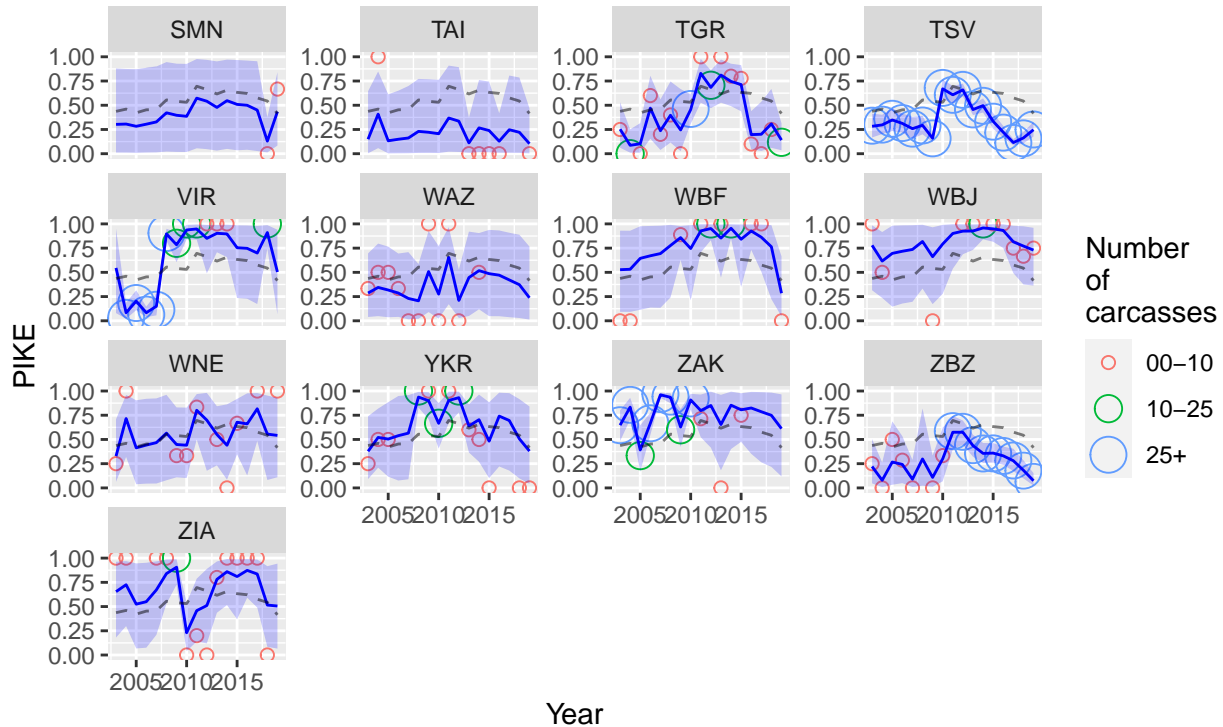
In summary, in years with many carcasses reported, the estimated site-year *PIKE* will closely match the observed site-year *PIKE*. In years with few carcasses reported, the estimated site-year *PIKE* will be pulled towards the continental trend after accounting for the observed relationship between this sites *PIKE* and the continental trend.

9 Sensitivity Analysis

In this section, we examine the sensitivity of the various PIKE measures to the data.

9.1 Effect of dropping a site

The yearly mean *PIKE* is computed based on unweighted or weighted averages of the individual site-year values. How sensitive are the yearly mean *PIKE* value to individual sites?

The yearly mean *PIKE* values are computed by dropping each site individually. The trend in yearly mean *PIKE* values from these fits are displayed over the yearly mean *PIKE* values using all sites below.

Effect of dropping individual sites on PIKE

Each black line corresponds to PIKE when a site is dropped
Red line is PIKE using all data

A separate line is drawn for the year *PIKE* values when dropping each site. In most cases, the effect of dropping individual sites is slight and so the individual lines "blend" together and cannot be separated except for a few cases when *PIKE* is computed by weighting by population abundance. In these cases, the yearly mean *PIKE* values can change substantially when a site with a large underlying population is dropped. This influence could push the year mean *PIKE* up or down depending if the particular site has a site-specific *PIKE* that was larger or smaller than the overall yearly mean *PIKE*.

The average change in mean *PIKE* when site s is removed is computed as:

$$avg.change_s = \sqrt{\frac{\sum_{years}(\widehat{PIKE}_{drop\ site_s} - \widehat{PIKE}_{all\ data})^2}{\#years}}$$

This is plotted against the average population abundance at the MIKE site:

Mean changes in yearly PIKE estimates when dropping a site
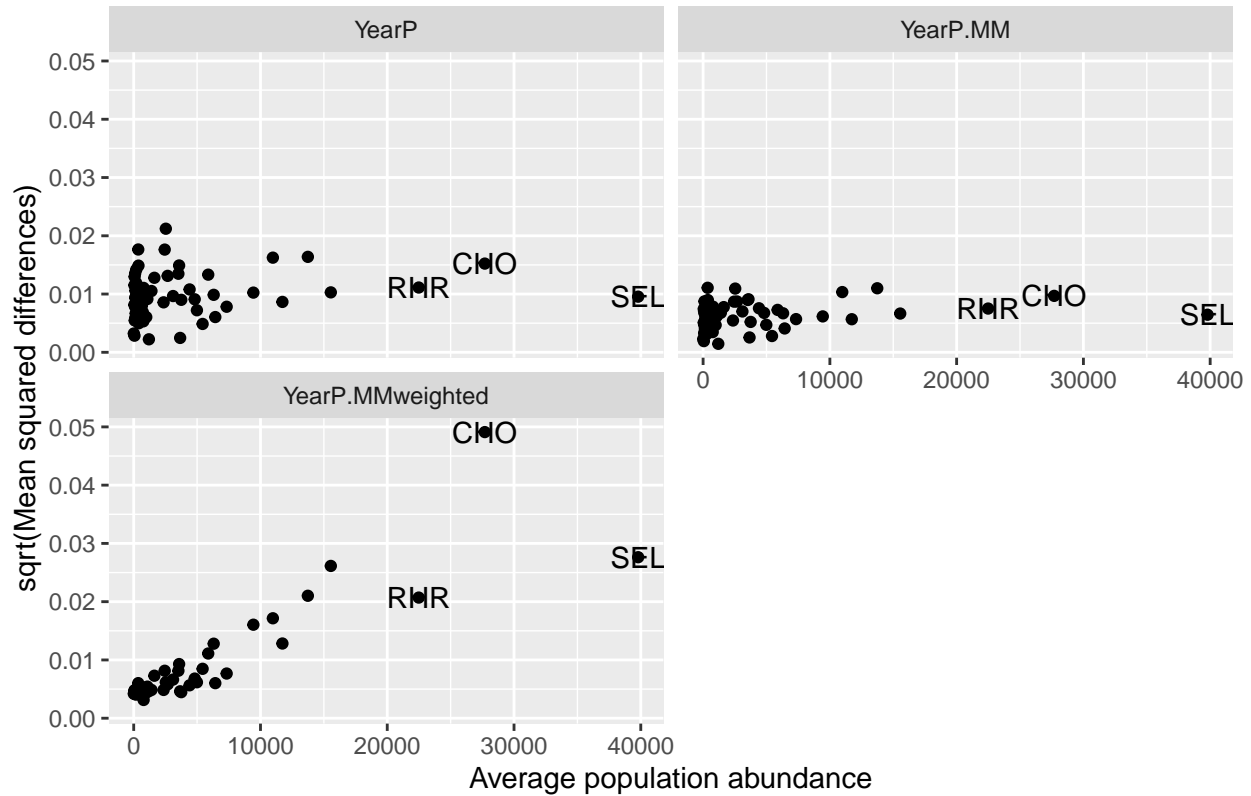
As expected, dropping a site with a large population abundance has a large influence on the weighted yearly mean *PIKE* but a relatively small impact on the unweighted yearly mean *PIKE*. Changes tend to be larger overall in the weighted yearly mean *PIKE*.

9.2 Adding new sites

New MIKE sites can be added in unpredictable ways. For example, in 2018, 7 sites were added (6 of these submitted data). A new site will have NO carcass data until year in which it was added.

Adding a new MIKE site may have impacts on previous estimated *PIKE* values. Presumably the MIKE site had elephants and potential illegal killings prior to being added to the programme. The current model will impute the *PIKE* for this site in the years before the site was added based on the yearly trend in *PIKE* in other sites and the relationship between the new site's *PIKE* and other sites' *PIKE*. For example, if the new MIKE site had a *PIKE* value that was 10 percentage points above the mean *PIKE*, then a *PIKE* will be imputed for this site for all past years that is also 10 percentage points above the yearly mean.

The imputation will be very coarse in the first few years after a MIKE site is added until enough yearly *PIKE* data has been collected to estimate reasonably well the relationship between this site's *PIKE* and the overall mean (i.e., several years will be needed to estimate the (random) *site* effect well).

Fortunately, the (unweighted) marginal mean *PIKE* is currently computed using over 50 MIKE sites and so adding a new MIKE site is expected to only have minimal impact on the yearly mean *PIKE* values unless the new site has an extreme *PIKE*. The weighted marginal mean *PIKE* could be influenced more if the new MIKE site represented a large elephant population that was not previously monitored. It is also possible, but unlikely, for a new MIKE site to represent an introduced population that did not exist in previous years (the inverse of an extirpation). The current unweighted mean *PIKE* currently does not have a mechanism to deal with introduced populations, but could be modified by creating an indicator variable that provides information if an elephant population existed in a particular site-year. Then a weighted mean using the indicator variable could account for newly introduced elephant populations. The weighted mean *PIKE* using

population abundance as the weighting variable would deal with introduced elephant populations in a similar fashion using a value of 0 for site-years when the new MIKE site did not have a population.

A small simulation was constructed to illustrate the impacts of a new MIKE site. The simulation considered adding a new MIKE site whose *PIKE* was 30 percentage points higher than, or -30 percentage points lower than the current mean *PIKE* of around 0.35. For each scenario, the population abundance was considered to be small (58) or large (10055) representing the 10th and 90th percentiles of the population abundances at the current MIKE sites. The simulated new MIKE site will report on 29 carcasses representing the mean number of carcasses currently reported across MIKE sites. The number of illegally killed carcasses will be set to the expected value give the *PIKE* and the number of monitored carcasses.

A total of 6 scenarios will be considered:

Table 13: Scenarios considered when adding a single new MIKE site

New Pike	New pop size	Total number of carcasses	Number of Illegal Carcasses
0.7	58	29	19
0.3	58	29	10
0.0	58	29	1
0.7	10055	29	19
0.3	10055	29	10
0.0	10055	29	1

A plot of the estimated year marginal *PIKE* when the simulated new MIKE site was added is:

Africa : Effects of adding a new MIKE site
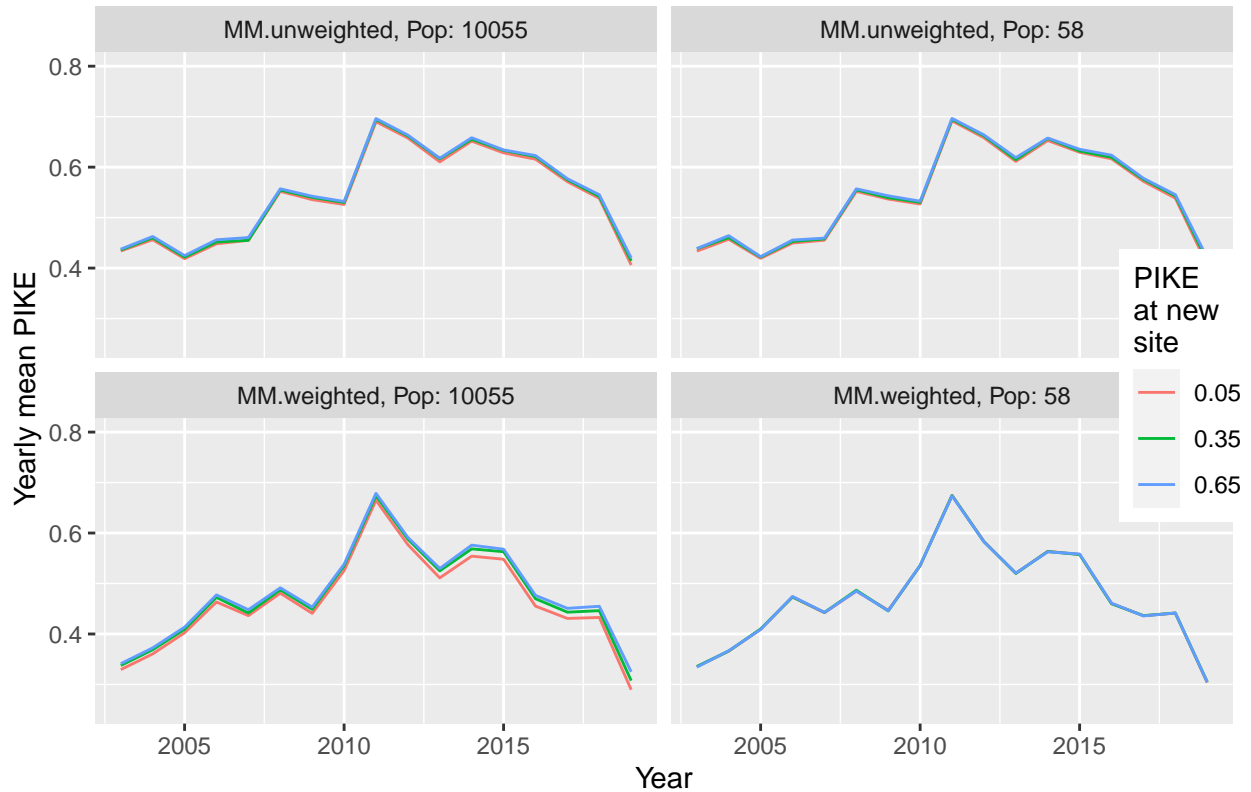
The plots above show minimal changes in the yearly mean *PIKE* when a single site is added except for the weighted (by population abundance) *PIKE* where adding a new MIKE site representing a large population has more effect in the later years. Even then, the effect is small relative to the uncertainty seen in the yearly estimated *PIKE* shown in earlier plots.

A similar study could be done to investigate the addition of multiple MIKE sites. It is expected that if the new multiple MIKE sites had *PIKE* values both above and below the current mean *PIKE* that the effects will also be minimal.

Because regional estimates of yearly mean *PIKE* are based on fewer MIKE sites, the impact of adding a new MIKE site could be larger. This has not been examined in in this report.

9.3   Extirpation at a MIKE site

It may happen that the population represented by a MIKE site is extirpated. In this, no carcasses are reported.

The models cannot distinguish between a MIKE site where the population is extirpated and a MIKE site that did not report in a year. In both cases, the model will extrapolate a *PIKE* for those site-years based on the relationship between the *PIKE* at that MIKE site in past years and the yearly mean *PIKE*.

The current unweighted mean *PIKE* currently does not have a mechanism to deal with extirpation, but could be modified by creating an indicator variable that is used to indicate if an elephant population existed in a particular site-year. Then a weighted mean using the indicator variable could account for an extirpated elephant population. The weighted mean *PIKE* using population abundance as the weighting variable would deal with an extirpated elephant population in a similar fashion using a value of 0 for site-years when the MIKE site did not have a population. This mechanism is similar to that used with introduced populations considered in the previous section.

9.4 Effects of the number of monitored carcasses

The sample size in each site-year (the number of carcasses monitored) also has an impact. For example, if a MIKE site reports 3 illegally killed carcasses out of 5 monitored carcasses, the observed number of illegally killed carcasses is consistent with underlying *PIKE* values ranging from about .20 to .90. But if the same site reported 30 illegally killed carcasses out of 50 monitored carcasses, the underlying *PIKE* values that are consistent with this data range from 0.45 to 0.72, a much narrower range.

With small sample sizes, the *PIKE* values for that MIKE site are 'unbiased' but have high uncertainty. When the yearly mean *PIKE* is computed over all MIKE sites, the uncertainty in the individual *PIKE* estimates is propagated into the uncertainty of the yearly mean *PIKE* values.

We simulated the effect of changing sample sizes by using the current *PIKE* data and multiplying the number of carcasses and number of illegally killed elephants by a constant multiplier. A total of 4 scenarios will be considered:

Table 14: Scenarios considered when sample size are increased by a given multiplier

Multiplier
1
4
10
25

The following plot shows the impact of the changing sample sizes on the estimated marginal yearly mean *PIKE*:

Africa : Effects of sample size on estimated unweighted marginal mean PIKE
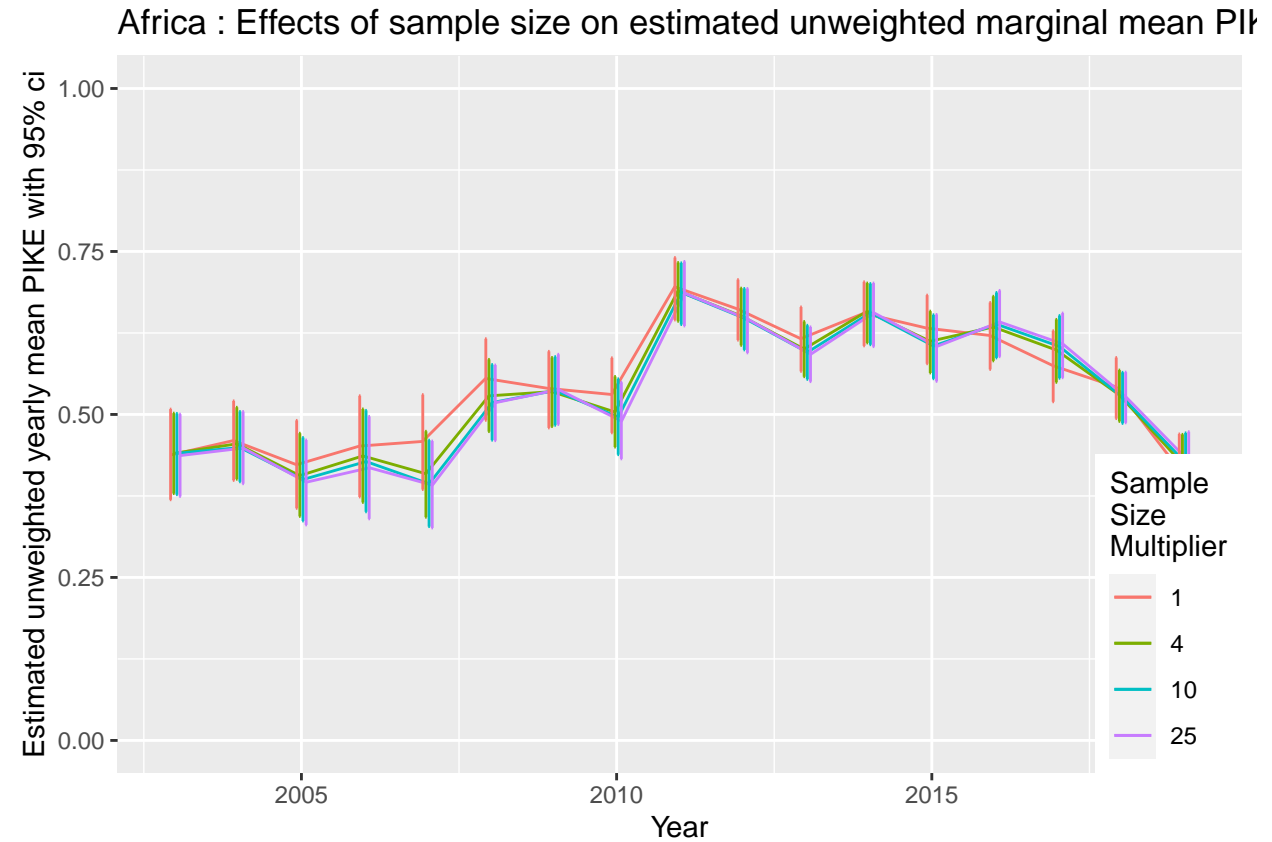
There is a minor effect of sample size showing that the precision (the 95% ci) is slightly narrower as the

sample size increases, but after an initial improvement, there does not appear to be a further improvement in precision.

Table 15: Average SD over all year.sites with multiplier sample size

Minimum sample size	Mean SD	Median SD	Max SD
1	0.0294	0.0279	0.0396
4	0.0266	0.0252	0.0368
10	0.0274	0.0256	0.0396
25	0.0278	0.0261	0.0401

At first glance, the LACK of a large effect of sample size on the precision of the estimated yearly marginal mean *PIKE* is somewhat surprising. However, a key determinant to the precision for a particular year is the number of imputed site-years which does not change if sample sizes in each MIKE site are increased. These missing values must still be imputed and the precision of the imputation does not really depend on the sample size of observed site-years. For example, in 2013, nearly all MIKE sites reported and in this year, the uncertainty does appear to decline with sample size.

9.5 Effects of a minimum number of monitored carcasses

The previous sensitivity analysis looked the impact of increasing the sample size at ALL sites. However, suppose that it were possible to insist that a minimum number of monitored carcasses be observed. Note that enforcing this restriction may not be logistically feasible if the underlying population for a MIKE site is very small. For example, if the population at a MIKE site was around 100 elephants, then with a 2% annual mortality rate, only about 2 carcasses would be observed in all years if the entire population was monitored.

As noted above, with small sample sizes, the *PIKE* values for that MIKE site are still 'unbiased' but have high uncertainty. When the yearly mean *PIKE* is computed over all MIKE sites, the uncertainty in the individual *PIKE* estimates is propagated into the uncertainty of the yearly mean *PIKE* values.

In this sensitivity analysis, we must use a simulated population unlike in the previous section. If the observed sample size was simply increased for small populations but the same observed *PIKE* of 0 or 1 was used, you would have better information for these sites that the actual *PIKE* was close to 0 or 1 and this would then pull down or up the marginal mean *PIKE* as well. But the actual *PIKE* could have been 0.5 and you were just unlucky to observe 0 or 3 illegal kills in a sample of size 3.

We retained the same structure for when each site was visited over time, but assigned each site a true underlying *PIKE* value selected from a Beta(3,3) distribution that is constant over time. This implies that the mean *PIKE* per year is 0.5.

We simulated the effect of a minimum sample size in each year by using the current *PIKE* data and multiplying the number of carcasses and number of illegally killed elephants by a multiplier to bring the total number of carcasses monitored to the minimum. A total of 4 scenarios will be considered:

Table 16: Scenarios considered with minimum sample size

Minimum sample size
3
5
10
20

The following plot shows the impact of different minimum sample sizes on the estimated marginal yearly mean *PIKE*:

Africa : Effects of minimum sample size on estimated unweighted marginal
Simulated population with constant marginal PIKE over time
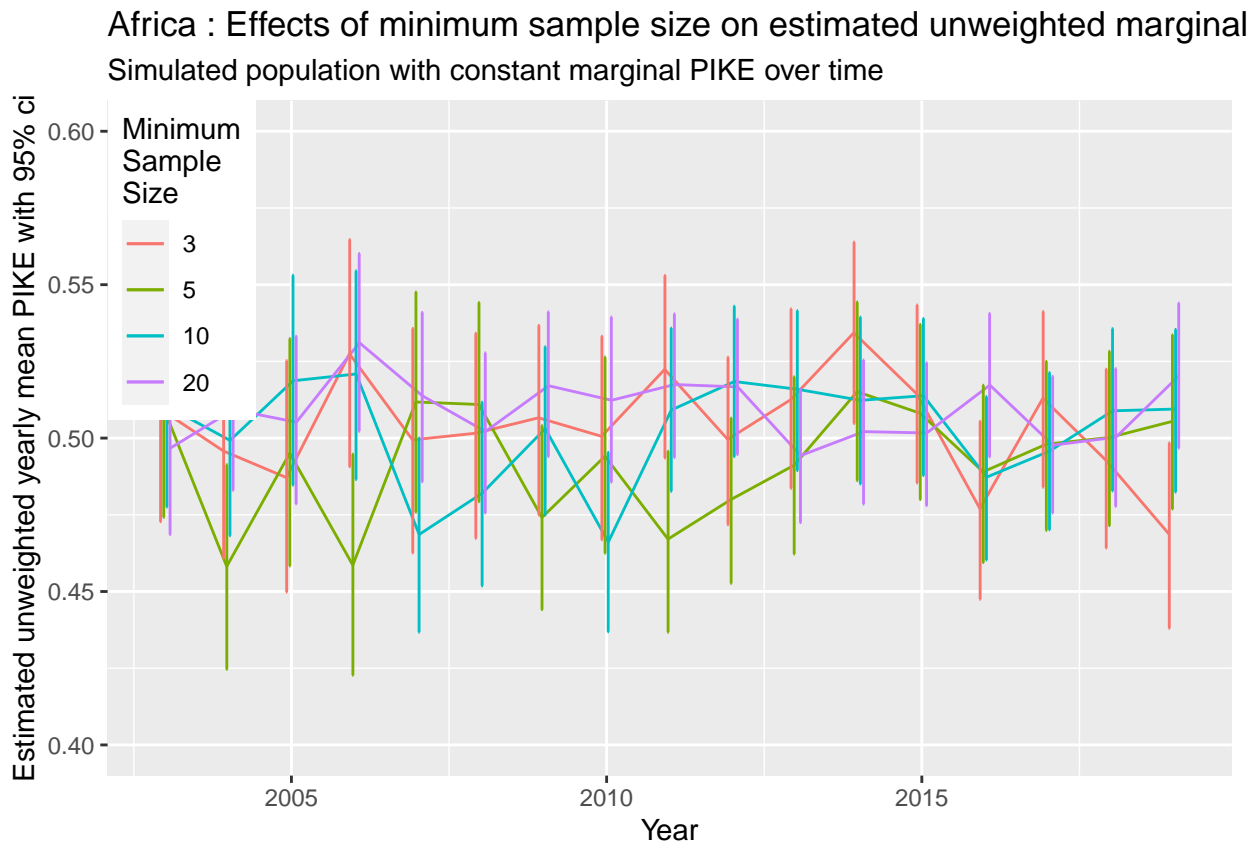
There is a minor effect of minimum sample size showing that the precision (the 95% ci) is slightly narrower as the sample size increases.

Table 17: Average SD over all year.sites with minimum sample size

Minimum sample size	Mean SD	Median SD	Max SD
3	0.0163	0.0153	0.0194
5	0.0158	0.0149	0.0191
10	0.0145	0.0138	0.0173
20	0.0126	0.0121	0.0147

As in Section 8.4, the LACK of a large effect of minimum sample size on the precision of the estimated yearly marginal mean *PIKE* is somewhat surprising. However, a key determinant to the precision for a particular year is the number of imputed site-years which does not change if minimum sample sizes in each MIKE site are increased. These missing values must still be imputed and the precision of the imputation does not really depend on the sample size of observed site-years. For example, in 2013, nearly all MIKE sites reported and in this year, the uncertainty does appear to decline with sample size.

9.6 Influence of priors

We assessed the sensitivity of the result to the prior by examining the overlap between the prior distribution and the posterior distribution (Gimenez et al, 2009).

Table 18: Overlap of Posterior with the Prior

Parameter	Overlap
Year.eff[1]	0.029
Year.eff[2]	0.027
Year.eff[3]	0.028
Year.eff[4]	0.032
Year.eff[5]	0.030
Year.eff[6]	0.029
Year.eff[7]	0.027
Year.eff[8]	0.030
Year.eff[9]	0.029
Year.eff[10]	0.028
Year.eff[11]	0.028
Year.eff[12]	0.026
Year.eff[13]	0.031
Year.eff[14]	0.026
Year.eff[15]	0.027
Year.eff[16]	0.029
Year.eff[17]	0.026
tau.site.eff	0.002
tau.year.site.eff	0.001

The last two parameters (*tau.xxx.eff*) represent parameters that control the variation of the *Site* and *Year.Site* effects.

A value less than 0.30 for the overlap is generally considered to be good evidence that the prior distribution has a small effect on the posterior (Gimenez et al. 2009). All of the overlap values are well below this threshold.

9.7   Effect of uncertainty in the abundance estimates

The weighted *PIKE* does not account for the uncertainty in the estimated elephant abundances. The uncertainty in the estimated elephant abundances is not readily available. We simulated the impact of uncertainty in the estimated elephant abundances by assuming that each estimate has an uncertainty of 25% of the estimate (i.e., the relative standard error, RSE, is 25% of the estimate). This implies that the 95% confidence interval for the elephant abundance for each site-year is approximately ± 50%.

A comparison of the estimates excluding and including estimates of uncertainty in the estimates of elephant abundance is:

Table 19: Impact of uncertainty in population weights. The RSE is assumed to be 0.25.

Year (1)	Bootstrap marginal weighted PIKE without/with uncertainty				Marginal weighted PIKE without/with uncertainty			
	Mean (2)	SD (3)	Mean (4)	SD (5)	Mean (6)	SD (7)	Mean (8)	SD (9)
2003	0.34	0.074	0.34	0.075	0.34	0.037	0.34	0.039
2004	0.37	0.075	0.37	0.077	0.36	0.032	0.37	0.038
2005	0.41	0.064	0.41	0.067	0.41	0.041	0.41	0.043
2006	0.47	0.082	0.47	0.086	0.47	0.035	0.47	0.041
2007	0.44	0.072	0.44	0.076	0.44	0.033	0.44	0.038
2008	0.49	0.082	0.48	0.085	0.49	0.033	0.49	0.040
2009	0.45	0.075	0.45	0.077	0.45	0.035	0.45	0.040
2010	0.54	0.065	0.54	0.066	0.54	0.024	0.54	0.029
2011	0.68	0.065	0.67	0.069	0.68	0.017	0.67	0.024
2012	0.59	0.074	0.59	0.076	0.58	0.015	0.58	0.025
2013	0.53	0.088	0.52	0.091	0.52	0.018	0.52	0.030
2014	0.57	0.071	0.56	0.074	0.56	0.018	0.56	0.025
2015	0.56	0.069	0.56	0.072	0.56	0.017	0.56	0.025
2016	0.47	0.068	0.47	0.070	0.46	0.017	0.46	0.025
2017	0.44	0.058	0.44	0.060	0.44	0.019	0.44	0.024
2018	0.45	0.056	0.45	0.058	0.44	0.017	0.44	0.022
2019	0.31	0.056	0.31	0.057	0.30	0.017	0.30	0.022

The uncertainty in the population weights only adds a small amount of uncertainty to the final *PIKE* estimate (compare columns 3 vs. 5 and columns 7 vs. 9). This is not unexpected because a positive error in the estimated abundance in one site will tend to cancel a negative error in abundance from another site if estimation errors are independent across sites.

This analysis ignores several issues

- Estimates are not obtained every year and intermediate values are set to the last survey until updated.
- Estimation errors in neighbouring sites may not be independent if both sites are surveyed at the same time in a combined survey.
- BIAS in the estimated abundances is assumed to be zero.

Consequently, the impact of uncertainty in the population weights reported above may be understated. The impact of the uncertainty in the population weights will be larger for smaller aggregates, i.e. the impact at the sub-regional level will be larger because the estimated sub-regional *PIKE* is based on fewer sites and so there is less certainty that the estimation errors will cancel.

9.8   Common model at continental and subregional level

At the moment, two similar models with the same structure are fit at the continental or subregional levels, but the fitting takes place independently. For example, the entire data set is used to fit the model at the continental level, and four separate fits using subsets of the data are used to fit the model at the subregional level.

Because the subregional estimates are based on non-overlapping subsets of *MIKE* sites, a single model could be fit at the continental level (as is currently done), and subregional estimates obtained by explicitly choosing the appropriate subset of the *MIKE* sites to compute the subregional *PIKE*.

The advantages of this approach are:

- All data are used to help impute missing site-year values for the sub-regional level rather than just the *MIKE* sites in the subregion. For example, a *MIKE* site just outside the subregion boundary could
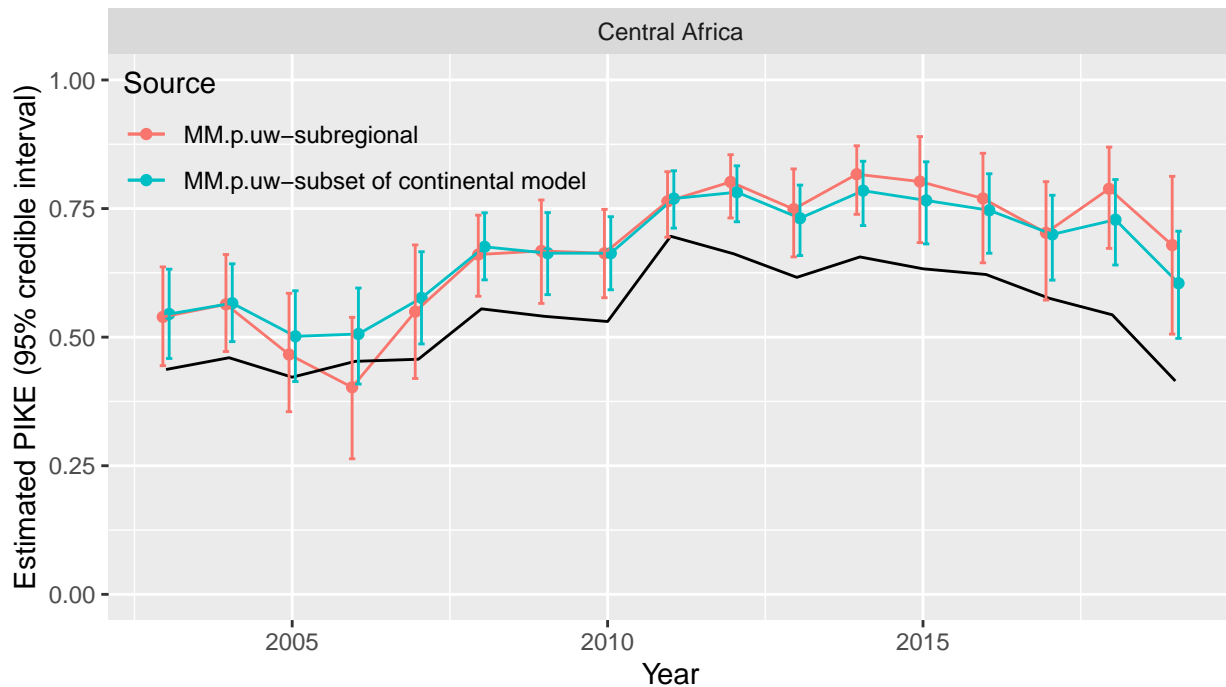
have a high correlation in *PIKE* with a *MIKE* site in a subregion. In the current approach, this high correlation is used to impute missing site-year combinations at the continental level, but not at the sub-regional level because the two sites are in two different sub-regions.

- Subregions with poor data will "borrow" information from other sub-regions. For example, if subregions with high amounts of data have *PIKE* trending downwards, this information is used to inform trends in the subregion with poor data.
- Estimates from the subregional level are consistent with those at the continental level, i.e. a suitably weighted average of the subregional estimates will equal exactly the estimate at the continental level. In the current approach where separate models are fit for each subregion and at the continental level, this is not always true.
- Estimates will be slightly more precise because all of the data from all regions is used to to impute missing site-year *PIKE*.
- Only a single model needs to be fit rather than the current 5 (one for the continental level, and 4 for the individual subregions).

A comparison of the unweighted (i.e. every site has equal weight) *PIKE* at the subregional level estimated from the continental level model and the individual sub-regional models is:

Africa: Estimated subregional PIKE across time

Comparing subregional estimates from subsetting continental model
and separate subregional models
Overall continental trend from global model in black

Africa: Estimated subregional PIKE across time

Comparing subregional estimates from subsetting continental model
and separate subregional models
Overall continental trend from global model in black

Africa: Estimated subregional PIKE across time

Comparing subregional estimates from subsetting continental model
and separate subregional models
Overall continental trend from global model in black

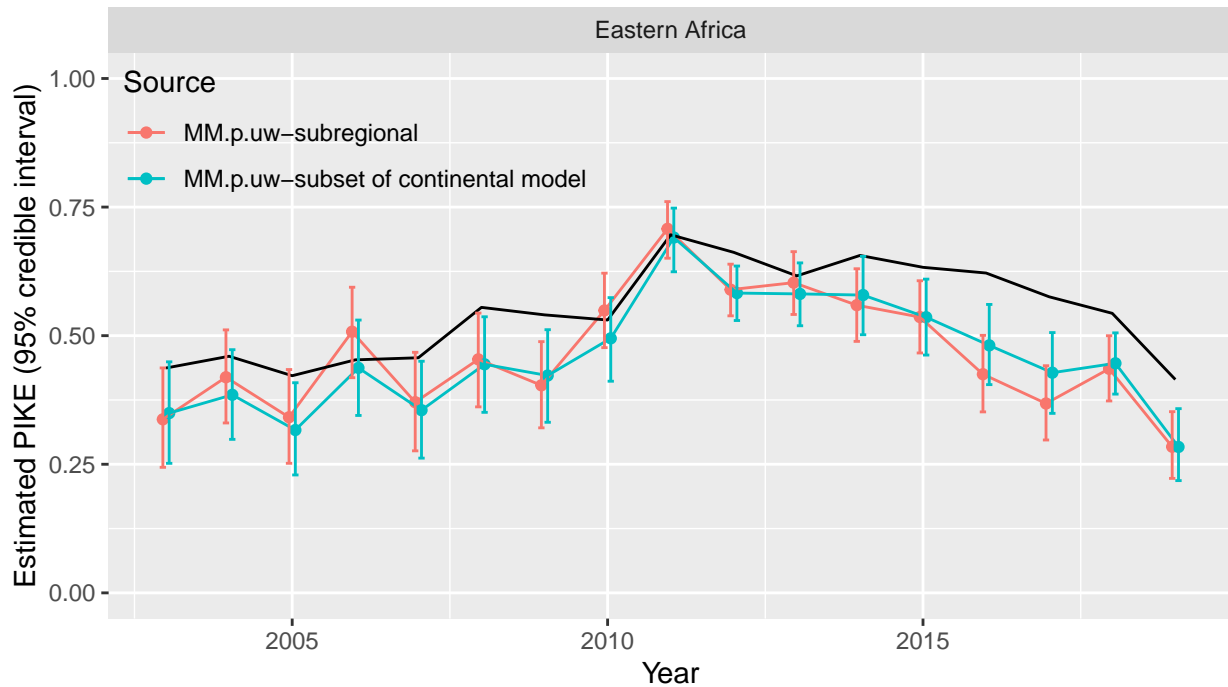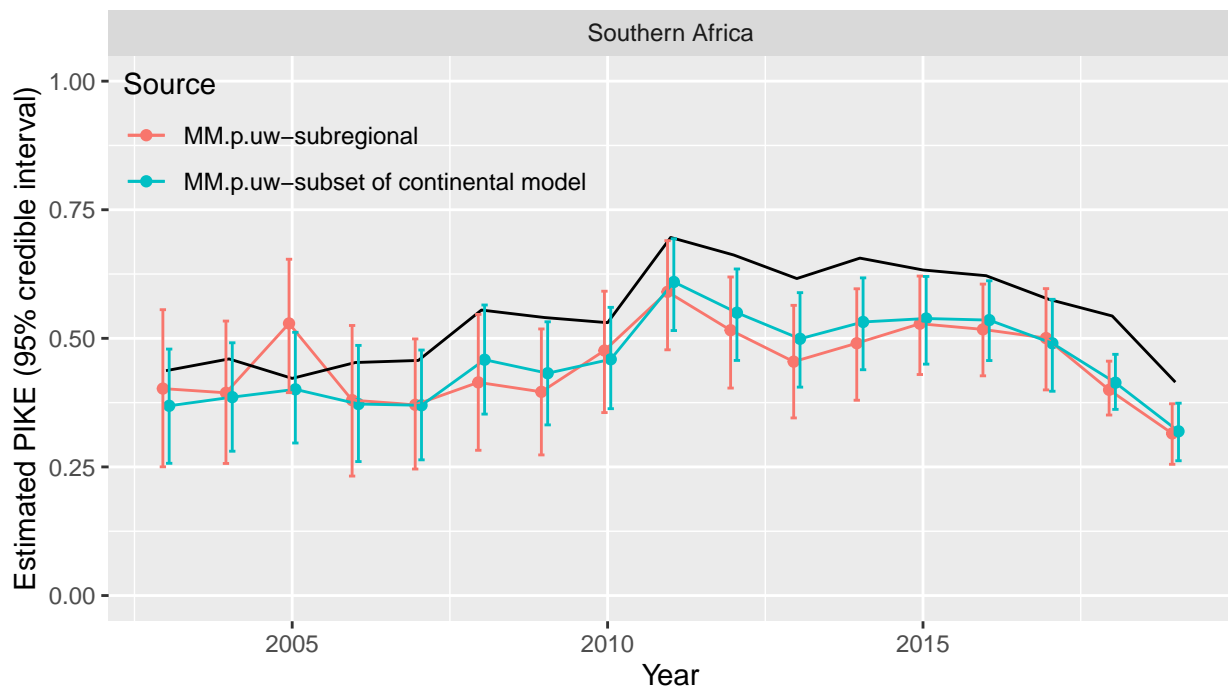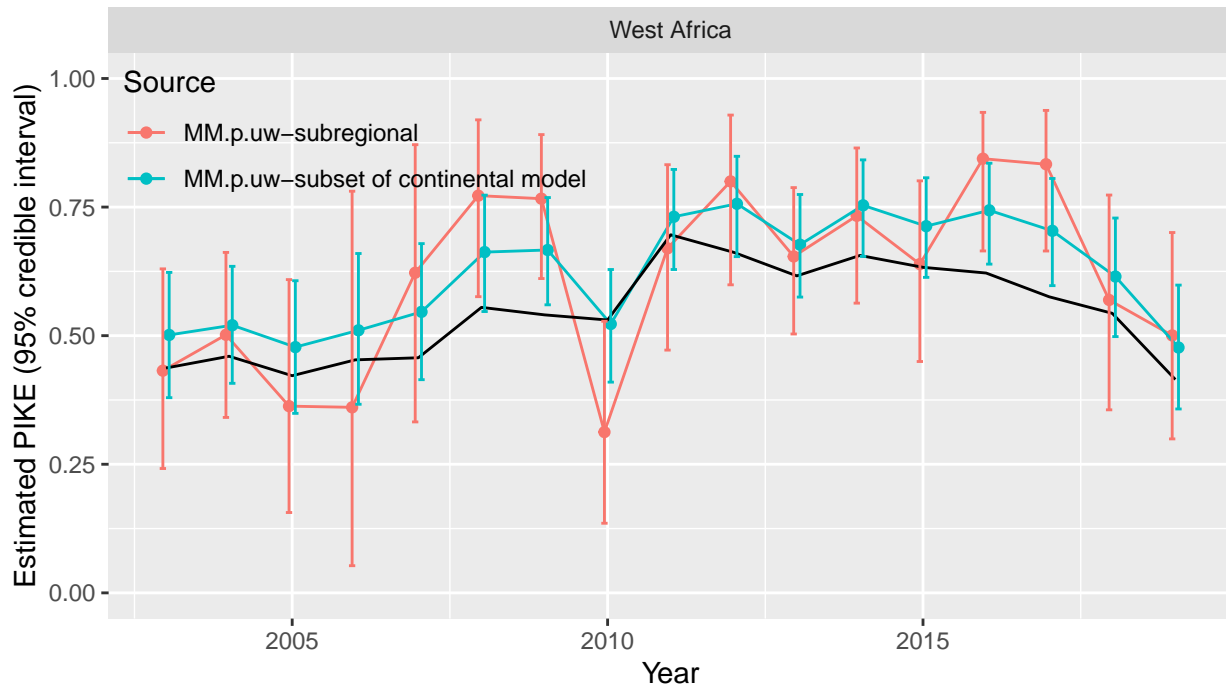**Africa: Estimated subregional PIKE across time**

Comparing subregional estimates from subsetting continental model
and separate subregional models
Overall continental trend from global model in black

In subregions with larger numbers of carcasses examined, the *PIKE* estimated from separate subregional models and estimated from the global model are very similar. Standard errors for the subregional *PIKE* estimated from the global model are slightly smaller than the standard errors from the individual subregional models.

The largest differences occur for the West Africa subregion, especially in the early 2000s. Here the data for the West Africa subregion is very sparse (especially for 2006 when only 2 sites reported each with a very small number of carcasses examined). Based on a small number of sites reporting and with smallish number of carcasses examined in the early 2000s, the individual subregional model estimates a very low *PIKE* (close to 0, but with very wide confidence intervals), but based on the global model, the sparse data could also be indicative of *PIKE* closer to the continental trends. The estimated subregional trend line from the global model is above the continental trend in the early 2000s because the trend in later years (e.g. 2010+) where there is more data indicates that the *PIKE* are above the continental trends.

Similar effects are expected with the weighted (by population abundances) estimates, or the bootstrap estimates and are not shown here.

10 Advanced models - future work

10.1 Spatial effect models

Zuur (2019) considered a model for *PIKE* that accounted for spatial autocorrelation among the site effects. In this way, MIKE sites that are spatially close together and highly correlated in their *PIKE* over time, are not treated as independent observations. This model may be useful if several MIKE sites monitor the same transitory elephant population and poaching pressure is similar in the site, e.g. within the same country.

This spatial model requires specialized software (the *INLA* package in *R*) and is computationally very demanding and requires a high level of expertise to use.

This new model is current under review by the MIKE-ETIS TAG.

10.2 Spatial-Temporal Effects

Zuur (2019) also considered a model for *PIKE* that accounted for spatial autocorrelation AND temporal autocorrelation. Temporal autocorrelation would be induced if a underlying temporal trend in yearly mean *PIKE* was estimated, e.g., a smoothing spline. In our case, we consider each year's *PIKE* to be of interest and the underlying smooth trend is not of interest. Such a model may be more applicable if covariates that vary on a yearly scale (e.g. ivory prices) were used as an explanatory variable for the temporal trends seen.

This spatial-temporal model again requires specialized software (the *INLA* package in *R*) and is extremely computationally demanding, and requires a high level of expertise to use.

This new model is current under review by the MIKE-ETIS TAG, but this model is unlikely to be implemented in the near future.

11 References

Burn, R.W., Underwood, F.M., Blanc J. (2011). Global Trends and Factors Associated with the Illegal Killing of Elephants: A Hierarchical Bayesian Analysis of Carcass Encounter Data. PLoS ONE 6(9): e24165. https://doi.org/10.1371/journal.pone.0024165

Chen, Ming-Hui, and Qi-Man Shao. (1999). Monte Carlo Estimation of Bayesian Credible and HPD Intervals. Journal of Computational and Graphical Statistics 8, 69-92. doi:10.2307/1390921.

Freeman, M.F. & Tukey, J.W. (1950). Transformations related to the angular and square root. Annals of Mathematical Statistics, 221, 607–611.

Gelman, A, Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A. and Rubin, D.R. (2013). Bayesian Data Analysis, 3rd Edition. Chapman and Hall/CRC.

Gimenez, O., Morgan, B.J., and Brooks, S. (2009). Weak identifiability in models for mark-recapture-recovery data. pp.1055-1068 in Thomson, Cooch and Conroy (eds) Modeling demographic processes in marked populations. Springer.

Lunn, D., Jackson, C., Best, N., Thomas, A. and Spiegelhalter, D. (2012). The BUGS Book – A practical introduction to Bayesian Analysis. Chapman and Hall/CRC Press.

Millar, Russell B. (2009). Comparison of Hierarchical Bayesian Models for Overdispersed Count Data Using DIC and Bayes' Factors. Biometrics, 65, 962-69.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003), March 20–22, Vienna, Austria. ISSN 1609-395X.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rubin,D. B. (1981) The Bayesian Bootstrap. The Annals of Statistics 9, 130-134. http://www.jstor.org/stable/2240875

Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64, 583-639. doi:10.1111/1467-9868.00353

Thouless, C.R., H.T. Dublin, J.J. Blanc, D.P. Skinner, T.E. Daniel, R.D. Taylor, F. Maisels, H. L. Frederick and P. Bouché (2016). African Elephant Status Report 2016: an update from the African Elephant Database. Occasional Paper Series of the IUCN Species Survival Commission, No. 60 IUCN / SSC Africa Elephant Specialist Group. IUCN, Gland, Switzerland. vi + 309pp

Zuur, A. F. (2019). Statistical analysis of spatial-temporal elephant poaching data using R-INLA. Prepared for CITES.

12   Appendix 1 - A note on the new MIKE sites added in 2019

An additional seven MIKE sites were added to the set of sites in 2019. Population values were obtained back to 2002.

Because few of these sites will have any carcass data prior to 2019, the observed *PIKE* in 2019 onward and the relationship of the observed *PIKE* in this year to the overall continental trend *PIKE* values in other sites will be used to impute values for *PIKE* for the new sites for all years prior to 2019.

There will be slight affect on estimates of the continental *PIKE* for all years prior to 2019, however, as noted earlier in the document, this is expected to be small unless the observed *PIKE* are extreme.

Of more interest is the Majeta site in Malawi. The estimated population represented by this site is 0 in 2002-2005. In 2006, 70 elephants were introduced from Liwonde National park and in 2008 another 64; and in 2009 another 83 were introduced; and then jumps to 70 animals, corresponding to an introduction of animals at this site.

As noted earlier, the *PIKE* will still be extrapolated to years prior to the introduction in 2006. The weighted by population abundance marginal *PIKE* estimates will automatically account for no animals present in 2002 to 2005, but the unweighted marginal *PIKE* estimate will include this sites for these years when no elephants are present.

Given that this problem occurred very long ago in the time series, no adjustment has been made to the unweighted marginal *PIKE* estimates.

13   Appendix 2 - Simulation to illustrate how to interpret estimates

Interpretation of the estimates obtained from the real data is difficult because of the differing number of carcasses monitored in each site-year, missing site-years follow no obvious pattern, and no year has complete data.

Consequently, this appendix will use a small simulated population to illustrate the effects of different sample sizes in each site-year and the effect of missing site years on the estimates and how to interpret each estimate.

Simulated data were generated with 10 years of data for 10 sites. In the first half of the study, the total number of monitored carcasses was 50. In the second half of the study, the total number of monitored carcasses was 50000. So in the second half of the study, there is essentially NO uncertainty about the *PIKE* in those site years.

We examine years with complete data and varying number of site with missing data. A plot of when each site was measured is:

Years in which data collected and number of carcasses examined
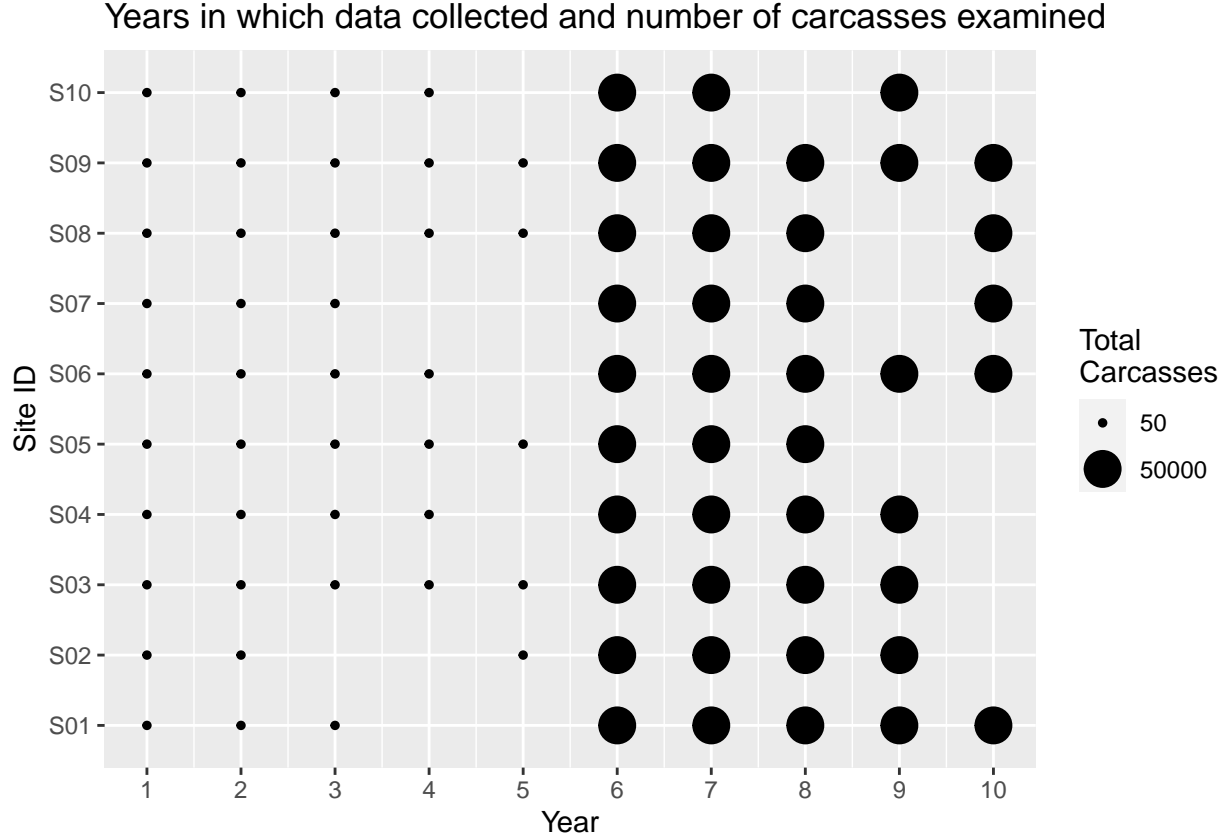
We set the actual *PIKE* to slowly increase over the years of the study, and then simulated (random) site and (random) year.site effects with standard deviation of 2 and 1 respectively (on the logit scale). The year, site, and year.site effects were used to generate the underlying *PIKE* for a year.site. The number of illegally killed carcasses was then generated using a binomial distribution.

The Bayesian GLMM was then fit to this simulated data

Estimates of the variance components are:

Table 20: Estimated standard dev of site and year.site effects

	Mean	SD	95% CI		Rhat	Eff n
			Lower	Upper		
sd.site.eff	2.50	0.69	1.54	4.18	1.001	4500
sd.year.site.eff	0.97	0.10	0.79	1.20	1.004	570

The estimates of the variance of the random effects are close to the true parameter values (as expected).

A plot of the estimates for each site over time are:

Simulated data: Observed and Predicted PIKE for individual sites
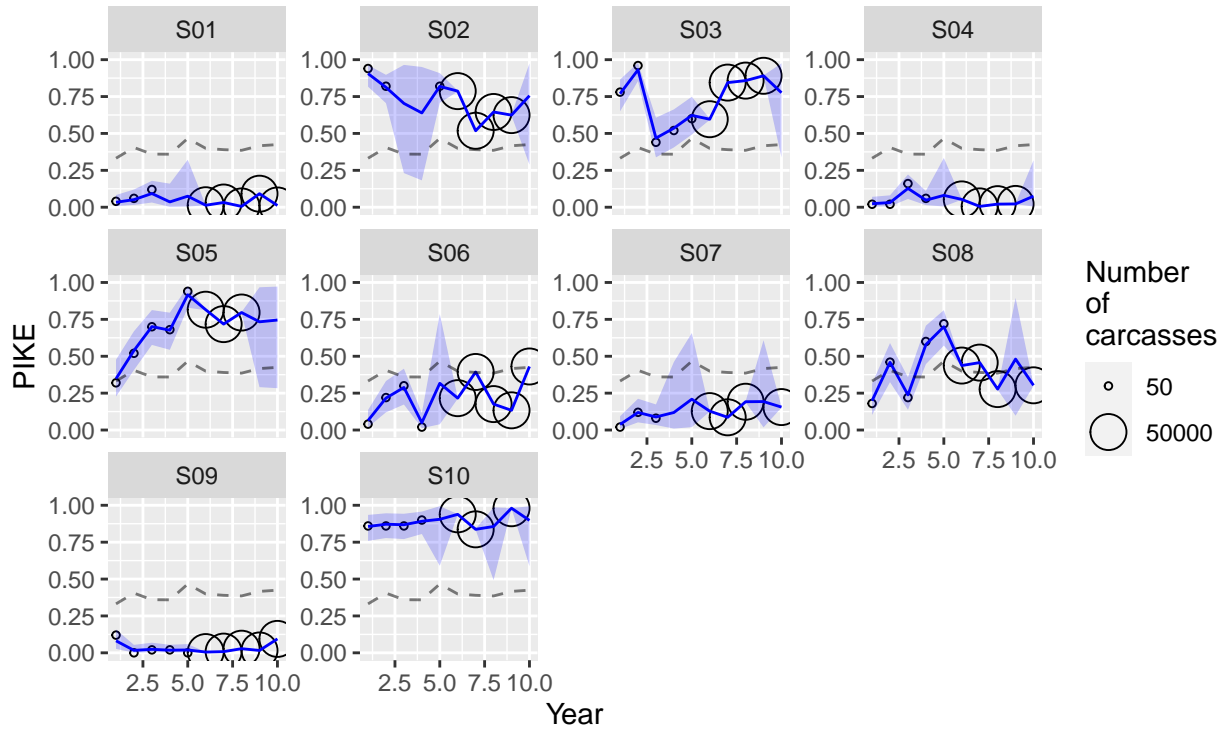
Dashed line is unweighted marginal mean PIKE at continental level
Blue and shading is predicted PIKE at site level with 95% credible interval

Notice that there is essentially no uncertainty about the *PIKE* in a year.site when a large number of carcasses are observed (large circles). When a site is missing in year, the imputed *PIKE* for that year.site has a large uncertainty.

We then extracted the year effect term (*Year.eff* on the logit scale) and the marginal mean logit(*PIKE*) (*Year.est.MM*).

Table 21: Specifications for simulation and estimated yearly effects (on logit scale)

Year	Proportion Sites Missing	Total Carcasses	logit(PIKE) at average site - Year.eff		Marginal mean logit(PIKE) -Year.e	
			mean	sd	mean	sd
1	0.0	50	-1.27	0.89	-1.29	0.18
2	0.0	50	-0.82	0.88	-0.83	0.19
3	0.1	50	-0.89	0.88	-0.91	0.19
4	0.3	50	-1.24	0.92	-1.27	0.29
5	0.5	50	-0.45	0.96	-0.47	0.41
6	0.0	50000	-0.99	0.87	-1.00	0.01
7	0.0	50000	-1.19	0.86	-1.20	0.01
8	0.1	50000	-1.04	0.87	-1.06	0.11
9	0.3	50000	-0.59	0.90	-0.62	0.22
10	0.5	50000	-0.55	0.93	-0.56	0.33

As noted earlier in the report, the *Year.eff* and *Year.est.MM* are similar. The *Year.eff* term in the model estimates the (logit) year effect for a site whose (random) site and year.site effects are 0, i.e. for an "average site". The *Year.est.MM* parameter estimates the mean of the individual logit(*PIKE*) over the sites within

a year. Because both the (random) site and (random) year.site effect have a mean of 0, the total of the observed site and year.site effects will be close to 0 and so it is not surprising that the estimates match those for an "average" site whose site and year.site effects are both zero.

While the estimates are similar, the uncertainty (SD) of these estimates is quite different and responds to sample size in each year.site differently.

The SD for the *Year.eff* is fairly constant regardless of the sample size and increases weakly with the proportion of missing sites in a year, and measures the uncertainty of the logit($PIKE$) in a specific year at an "average" site, i.e. with the (random) site and year.site effects of 0. Because this is based on real and imputed data for a year, the SD is found approximately like the standard error of a mean in a likelihood setting as

$$SD_{Year.eff} \approx \sqrt{\frac{Site\ effect\ variation + Year.Site\ effect\ variation}{\#\ sites}} = \sqrt{\frac{2.5^2 + 1^2}{10}} = 0.85$$

The sample size within a year.site (which affects the uncertainty of the estimated *PIKE* in a year.site) plays a small role as long as the total of the *Site* effect and *Year.Site* effect variation is large. To reduce this standard deviation, more sites are needed or the variation among sites must be reduced. This behaviour is similar to the standard error of a mean of a random sample from some distribution.

On the other hand, the SD for the marginal mean logit($PIKE$) when there are a large number of carcasses measured and no missing sites in a year (e.g. Years 6 and 7) is very small. With a large number of carcasses measured there is very little uncertainty in the logit($PIKE$) for a particular year.site, and so the average over the sites will also have little uncertainty. If the *PIKE* were known with certainty in all sites for a year, the SD of the mean logit($PIKE$) would be zero. This may seem surprising, but is correct – there is then no uncertainty in the mean *PIKE* for THESE PARTICULAR sites.

Conversely, the SD for the marginal mean logit($PIKE$) when there are a small number of carcasses measured and no missing sites in a year (e.g. Years 1 and 2) is larger. With a smaller number of carcasses measured there is larger uncertainty in the logit($PIKE$) for a particular year.site, and so the average over the sites will also have a larger uncertainty.

The impact of missing sites in year is dramatic. As seen in the plots of the *PIKE* for each site earlier, the imputed year.site *PIKE* value have large uncertainty. Consequently, the average over the sites must have a larger uncertainty. As the number of sites with missing values increase in a year, the uncertainty becomes larger as well.

Similar findings hold for the marginal unweighted mean unweighted *PIKE*:

Table 22: Specifications for simulation and estimated unweighted marginal mean PIKE

Year	Proportion Sites Missing	Total Carcasses	Unweighted marginal mean PIKE		Bootstrap unweighted marginal mean	
			mean	sd	mean	sd
1	0.0	50	0.33	0.014	0.33	0.107
2	0.0	50	0.40	0.015	0.40	0.107
3	0.1	50	0.36	0.027	0.36	0.095
4	0.3	50	0.36	0.032	0.36	0.106
5	0.5	50	0.47	0.043	0.47	0.117
6	0.0	50000	0.40	0.000	0.40	0.105
7	0.0	50000	0.39	0.001	0.39	0.098
8	0.1	50000	0.39	0.013	0.39	0.105
9	0.3	50000	0.42	0.037	0.42	0.117
10	0.5	50000	0.42	0.042	0.43	0.112

When every site occurs in year with a large number of carcasses (years 6 and 7), there is essentially no

uncertainty about the *PIKE* for each year.site, and hence no uncertainty about the average *PIKE* FOR THESE SITES. Again, notice the impact of missing sites within a year on the uncertainty of the marginal mean *PIKE* – the imputed *PIKE* for the missing sites increases the uncertainty of the average *PIKE* FOR THESE SITES. Columns 4 and 5 correspond to "fixed" MIKE sites and so no uncertainty from random sampling of sites has been included.

In contrast, columns 6 and 7 include uncertainty due to random sampling of sites. This would be appropriate if MIKE sites are considered to be a random sample of sites. Even with perfect information in every MIKE site (years 6 and 7) there is still uncertainty in this marginal mean.

In summary, there are two ways to measure the mean logit *PIKE* over many sites. The *Year.eff* term from the model behaves the closest to the mean of a random sample of logit *PIKE* in classical sampling theory. Here the implicit assumption is that the sites chosen are a random sample taken from all possible sites and interest lies in the overall mean. The SD for this parameter is closest in spirit to the traditional standard error of a mean. The SD depends mainly on the number of sites and the site-to-site and year.site variation in the logit *PIKE*. Sample size within a site has only a weak effect on the uncertainty for this estimate.

The marginal mean logit *PIKE* treats the sites as "fixed" and as an "index" to the population. Consequently, the mean logit *PIKE* is also an index. The SD for the marginal mean logit *PIKE* only reflects the uncertainty of the estimated *PIKE* in each year.site and represents the uncertainty in the index. No attempt is made to measure the uncertainty in the mean *PIKE* induced by a random sample of sites because, in this study, sites were not selected at random.

When you are interested in the marginal mean *PIKE*, you must also consider which measure of uncertainty is of interest.

For these reasons, and because of the issues with non-random sampling of MIKE sites, the non-representation of the *PIKE* at the sites compared to the surrounding areas, etc (as summarized in the introduction), that we again reiterate that the marginal mean unweighted or weighted *PIKE* should be thought of as an INDEX to the underlying poaching pressure.

14   Appendix 3 - Key difference between the previous analysis (LSMeans) and the Bayesian GLMM

A key difference between the previous (*LSMeans*) and the proposed Bayesian GLMM approach (aside from the frequentist vs. Bayesian approaches) lies in how the data from sites are aggregated to the sub-regional or continental level.

In many cases, the two methods give similar results, but this is not always true. Consider, for example, the sub-regional estimate of marginal mean *PIKE* for Eastern Africa. Plots of the results from the two methods were previously shown, but are reproduced here for ease of reference.

Estimated sub−regional PIKE estimated from the previous and proposed m
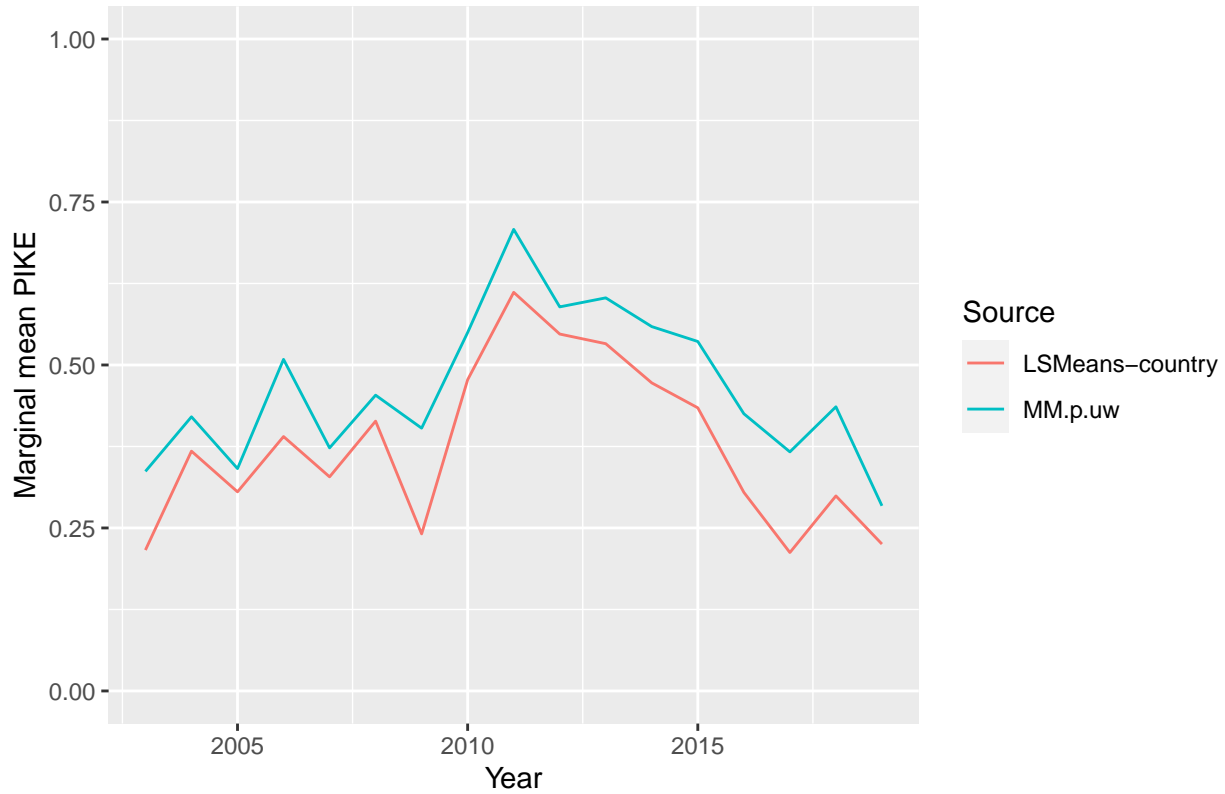
There is an approximately consistent difference between the two curves with the marginal mean *PIKE* higher when estimated using the Bayesian method.

The reason is related to how *PIKE* from sites are aggregated to the sub-regional level.

In the previous approach (the *LSMeans* approach), the model used to fit the observed pike at the subregional level is

$$lm(pike \sim Country + YearF, weight = TotalNumberOfCarcasses)$$

where *Country* is a country effect and *Y earF* is the categorical year effect. This model effectively totals the carcasses monitored and the number of illegally killed elephants to the country level and then gives each country equal weight in computing the sub-regional *PIKE*.

In the Bayesian GLMM approach, the model is similar to:

$$lm(pike \sim Site + YearF, weight = TotalNumberOfCarcasses)$$

where *Site* are the individual sites in Eastern Africa. Now each site is given equal weight in computing the sub-regional *PIKE*. The results from this "approximation" model is also displayed on the previous plot

Estimated sub–regional PIKE
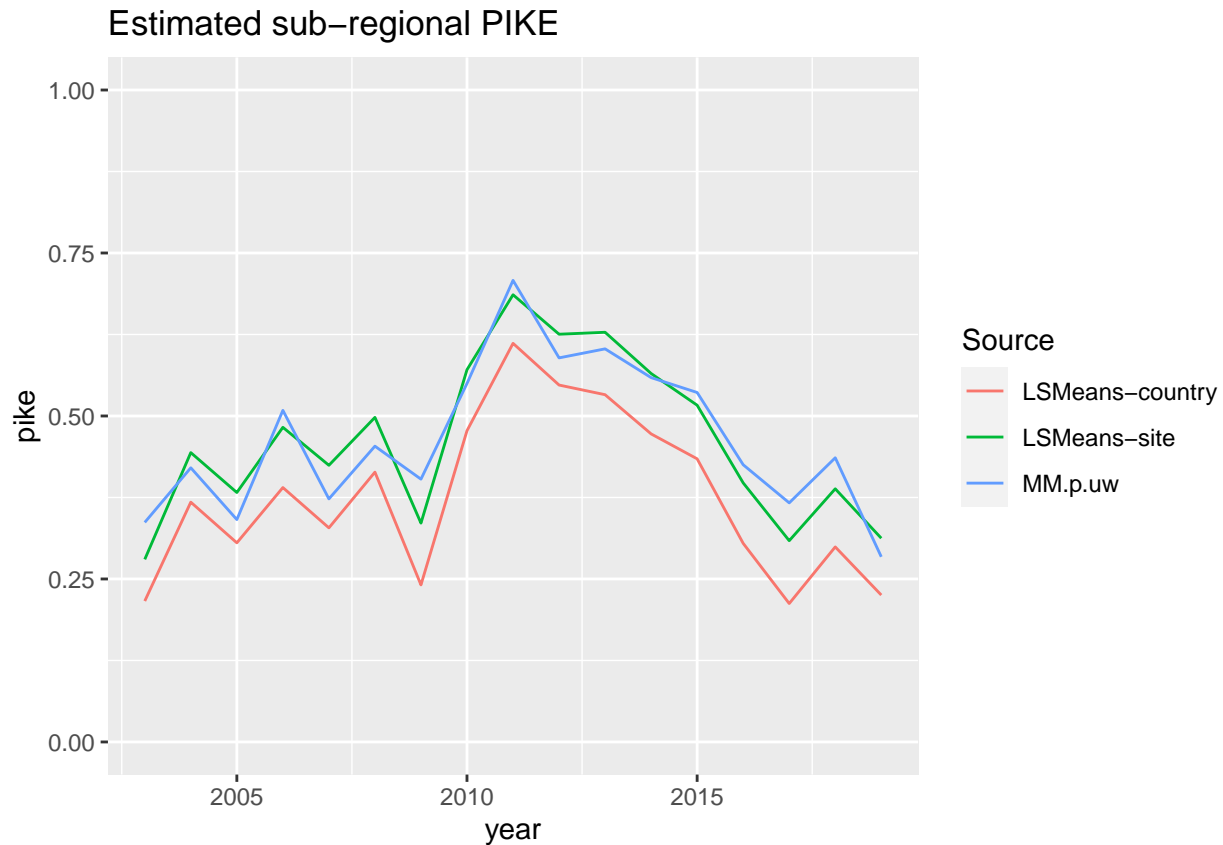
Notice the results from the *LSMeans* method but analyzed at the site level is very close to the Bayesian approach.

Why does equal weighting by country or equal weighting by site make such a consistent difference in this case. If we look at the number of sites for each country, and the observed *PIKE* in each country, we obtain:

Table 23: Summary of number of sites and carcasses by country in Eastern Africa

Country	# sites	Total Carcasses	Total Illegally Killed	Observed PIKE
Eritrea	1	71	6	0.08
Ethiopia	2	84	67	0.80
Kenya	4	6011	2299	0.38
Rwanda	1	16	1	0.06
Tanzania, United Republic of	5	2198	1297	0.59
Uganda	2	370	193	0.52
Weighted by country				**0.41**
Weighted by site				**0.48**

Notice that the difference in the weighted *PIKE* is about .07, approximately the same size as in the previous plots. When equal weight is given to each country (the previous *LSMeans* approach), equal weight is given to

the United Republic of Tanzania with a higher *PIKE* as to Eritrea with a lower *PIKE* despite the former having 5x the sites as the latter. In the Bayesian GLMM approach, sites are given equal weight, so countries with more sites are given more weight in the sub-regional *PIKE*. In this case, more weight is given to the United Republic of Tanzania (5 sites) with a higher *PIKE* than Eritrea (1 site) with a lower *PIKE*. Consequently, the Bayesian GLMM model will tend to give a higher value for the subregional trend for Eastern Africa.

If we examine the other subregions in a similar way:

Table 24: Summary of comparison of PIKE weighted by country or by site

Subregion	Average PIKE weighted by country	Average PIKE weighted by site
Central Africa	0.68	0.68
Eastern Africa	0.41	0.48
Southern Africa	0.42	0.50
West Africa	0.63	0.63

we see that the biggest difference occurs in Eastern Africa as seen in the results at the sub-regional level.

In summary, difference between the marginal mean *PIKE* estimated using the previous approach (*LSMeans*) and the new approach (Bayesian GLMM) is due primarily to the different weighting, previously by country and now by site. There are also subtle effects in each year when sites are missing, but the basic picture does not change.

It should be noted that these differences will not occur if weighting is done by population abundance. Presumably the population values at the regional level are the sum of the population values at the site level, so data will be equally weighted under either approach.

15 Appendix 4 - Frequently Asked Questions (FAQ)

15.1 Why are the weighted and unweighted marginal mean *PIKE* so different?

The unweighted marginal mean *PIKE* gives equal weight to each site regardless of the underlying population abundance the MIKE site represents. So a *PIKE* of 0.90 from a population of 100 elephants is treated the same as as *PIKE* of 0.90 from a population of 10,000 elephants, even though the latter actually implies a larger number of illegally killed elephants – all else being equal.

The weighted marginal mean *PIKE* weights each site's yearly *PIKE* value by the underlying population abundance of elephants. So a *PIKE* of .90 from a population of 10,000 elephants, is given 100 times the weight than a *PIKE* of 0.90 from a population of 100 elephants. Note that uncertainty in the population estimates has not been included when finding credible intervals for the weighted marginal mean *PIKE* at this time.

IN THEORY, weighting by population abundance and using the weighted marginal mean *PIKE* would make it possible to make statements about rates of death per head of population – that is the weighted marginal mean *PIKE* value would be more closely expressing poaching mortality (ignoring other issues such as varying natural mortality rates) because rates for sites with more elephants would contribute more to the average

than sites with fewer elephants. The unweighted marginal mean *PIKE* uses rates per site rather than per head of the population and so is a site-based average.

Weighted and unweighted mean marginal *PIKE* should be used only as an index to poaching pressure. As outlined in the introduction, a naive use of the marginal mean *PIKE* is not recommended to estimate actual poaching rates – the marginal mean *PIKE* should be viewed as an index to poaching pressure and would require a very careful analysis and far more data than is available in this report to convert to an actual poaching rate.

15.2 What would be the impact of including uncertainty in the population estimates

At the moment, uncertainty in the population estimates used in the weighted marginal mean *PIKE* is not incorporated into the credible intervals. It would be straight forward to include this uncertainty if the uncertainty for each estimate were available.

It is expected that including such uncertainty will make the final credible intervals slight wider. This issue was discussed in the Sensitivity Analysis section.

15.3 How are credible intervals found for derived parameters such as the marginal mean *PIKE*?

A careful reader will note that the marginal mean *PIKE* is a derived parameter, i.e. it does not appear in the likelihood or prior specification of the model but is a function of the actual parameters estimated by the model which are the year, site, and site-year effects (on the logit scale) along the the standard deviations of the distribution for the hierarchical parameters.

The MCMC algorithm used in this analysis will generate samples from the posterior distribution of the year, site, ad site-year effects. The 2.5^{th} and 97.5^{th} percentiles of the sample from the posterior will be an estimate of the 95% credible intervals for these parameters.

In likelihood analyses, standard errors and confidence intervals for derived parameters are found using the Delta method (https://cran.r-project.org/web/packages/modmarg/vignettes/delta-method.html). With Bayesian methods this is unnecessary and standard deviations and credible intervals for derived parameters are found in a straight forward manner.

We begin with a matrix of samples from the posterior with rows of the matrix representing different samples from the posterior and columns representing values from the posterior for each parameter. This matrix automatically incorporates any cross parameter correlations, or skewness in the posterior distributions. Each column represents the marginal posterior distribution for the parameter represented by that column. In our case, we retained 4500 samples from the posterior (number of rows) for the year, site, and site-year effects.

For each row, we then estimate the (logit) *PIKE* for each site-year combination and create additional columns in the matrix for these derived parameters. Each new column now represents a sample from the posterior distribution for that derived parameter and no further mathematical details are needed (i.e. the delta-method is implicitly done). We can then use these derived parameters to generate site-year (logit) *PIKE* values and find the site-year *PIKE* on the [0,1] scale using the anti-logit transformation and add them to the columns of the matrix. Again, each new column represents the marginal posterior distribution for the new derived parameter. Finally, we find the yearly marginal mean unweighted or weighted *PIKE* and add these new columns to the matrix. The final posterior distribution is found using the column in the same way, i.e. the the 2.5^{th} and 97.5^{th} percentiles of the final sample from the posterior will be an estimate of the 95% credible intervals for these final derived parameters.

Full theoretical details are available in Chen and Shao (1999).