MIKE analysis for Africa - Technical Report (updated 24-Aug-2024)

2024-09-03

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. The method used here was published in 2020 on a GitHub repository CITESmike2020/MIKE-GLMM with the computer code (written in R) and several accompanying technical reports. The original technical report, which this document is based on, can be viewed by clicking here.

The computer code from the GIT hub repository has been modified to estimate yearly PIKE trends from 2003-2021 using only the unweighted marginal mean PIKE model (MM.p.uw), which does not require elephant population abundance data at the site-year level. The pro and cons of this approach are discussed in more detail in the original technical document.

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:

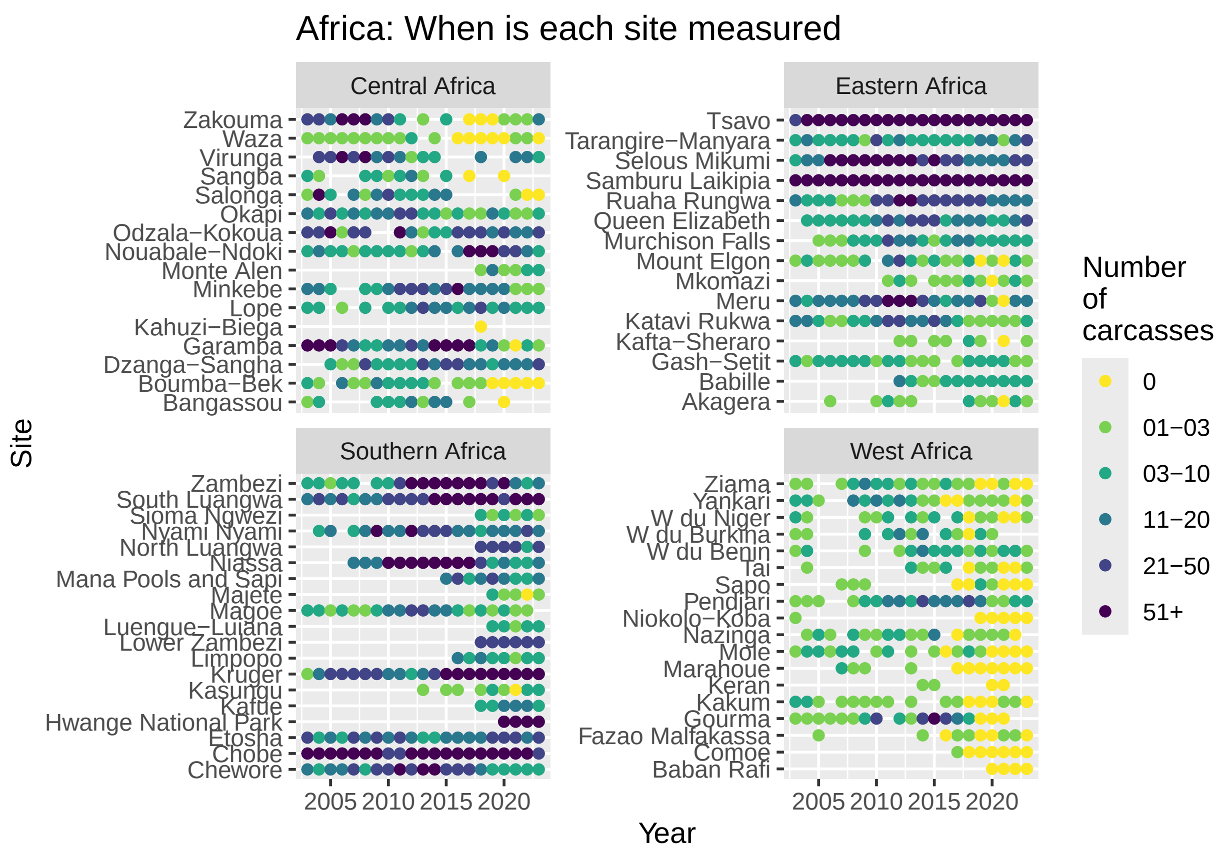
- Detection probabilities of elephant carcasses in various habitats differ.
- Levels of search effort and site coverage differ between sites.
- Not all sites report in all years.
- Number of carcasses in both categories varies considerably across space and time.

2 Exploration of PIKE data

2.1 MIKE sites with *PIKE* data

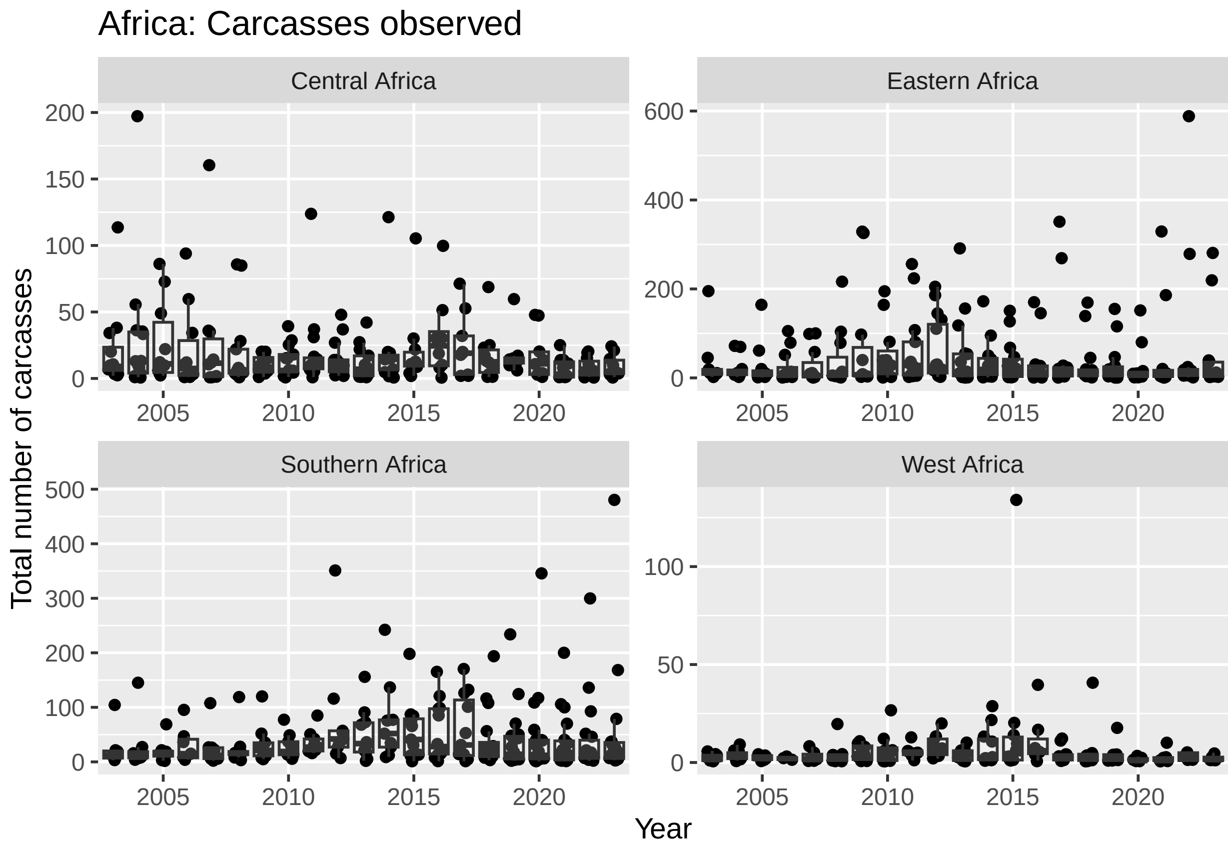
There are 69 MIKE sites that have reported data on the number of carcasses found and the number of illegally killed carcasses among these. This includes 90 site-years where sites have reported 0 carcasses and 909 site-years where sites have reported one or more elephant carcasses.

The current analysis treats a site that did not detect and report any carcasses in a year due to no patrol effort and a site that did conduct patrols but did not detect any carcasses in a year in the same manner. This is because information on patrol effort is not currently used in the analysis and only the number of carcasses detected and 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.



In total, there are 909 unique site-years in Africa since 2003 where data has been reported (and the number of reported carcasses > 0).

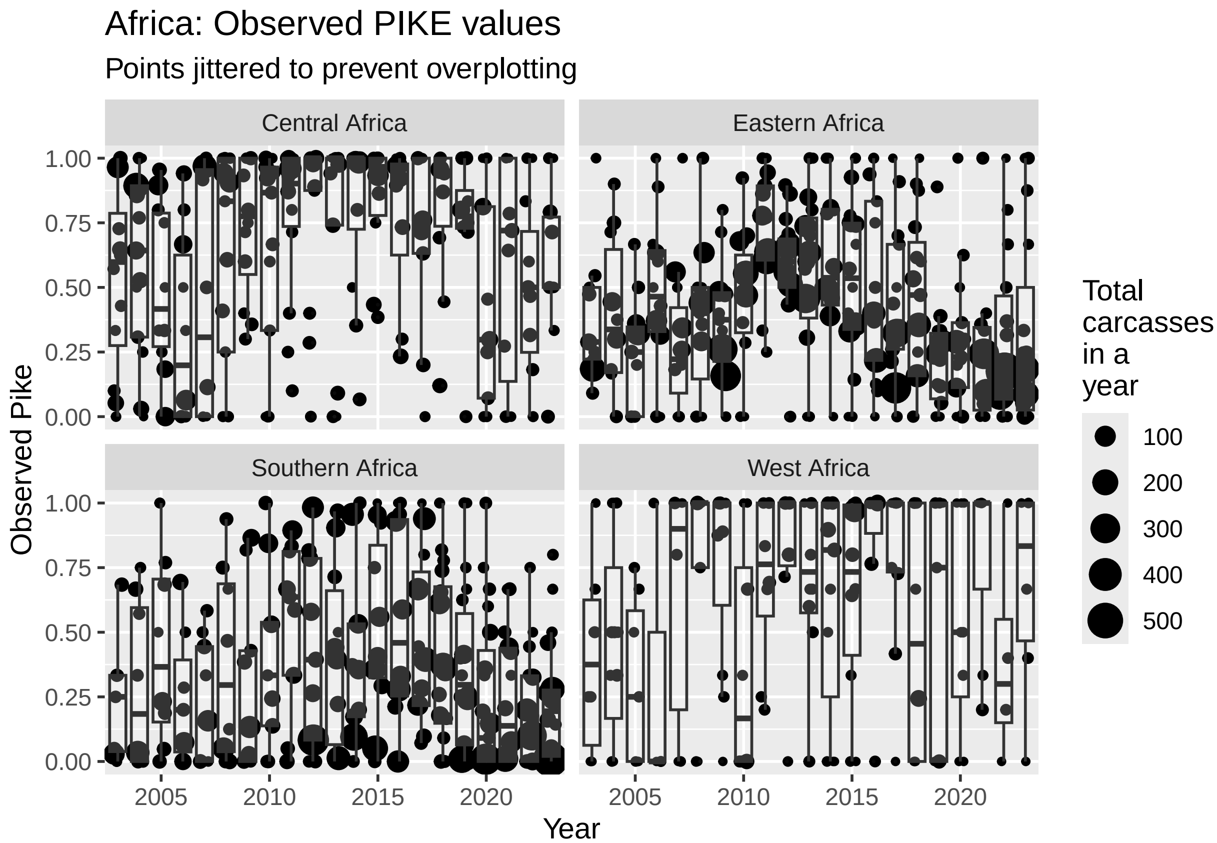
The number of carcasses reported in each site-year since 2003 varies enormously from 1 to 588 carcasses.



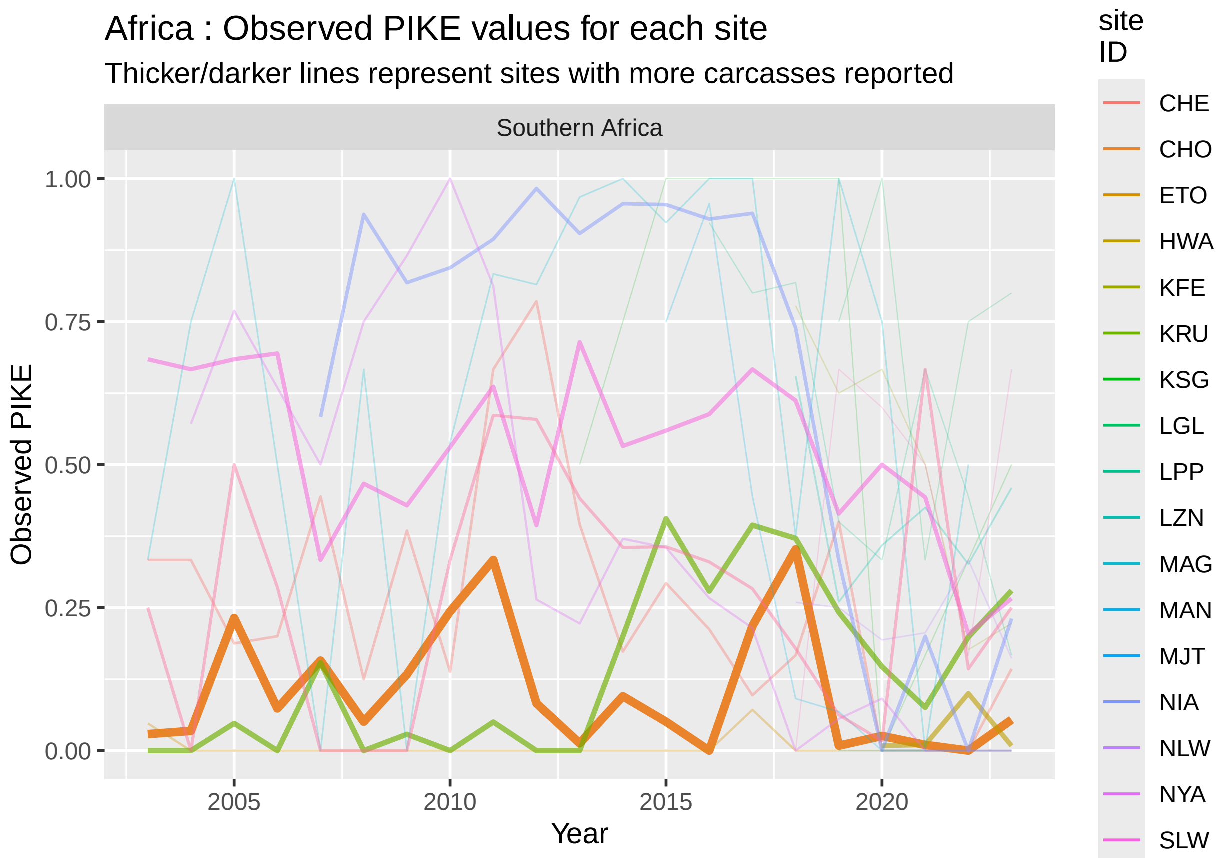
The unusual data point for West Africa where one site reported a large number of carcasses in one year is correct and corresponds to the MIKE site Gourma (GOU) with total number of carcasses equal to 134, of which 130 were poached by armed groups who entered the area.

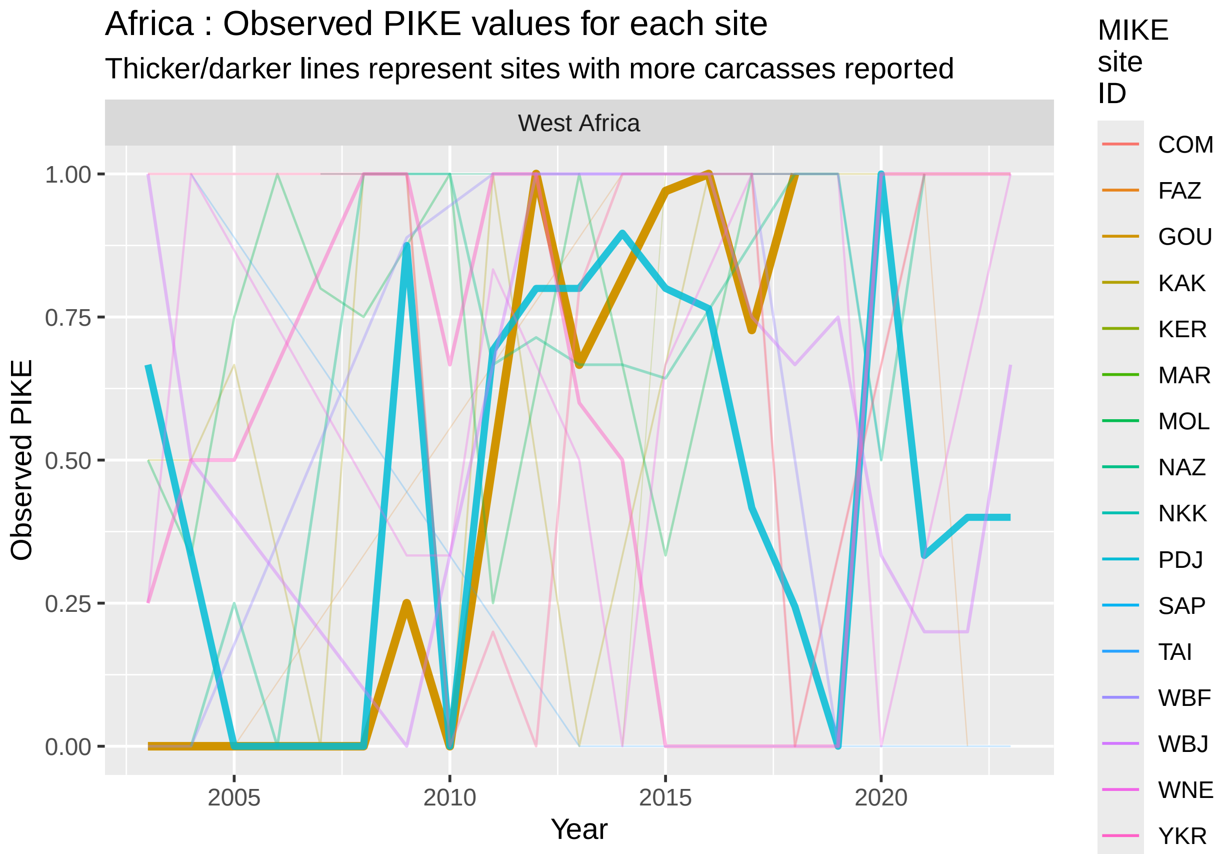
2.2 Observed *PIKE*

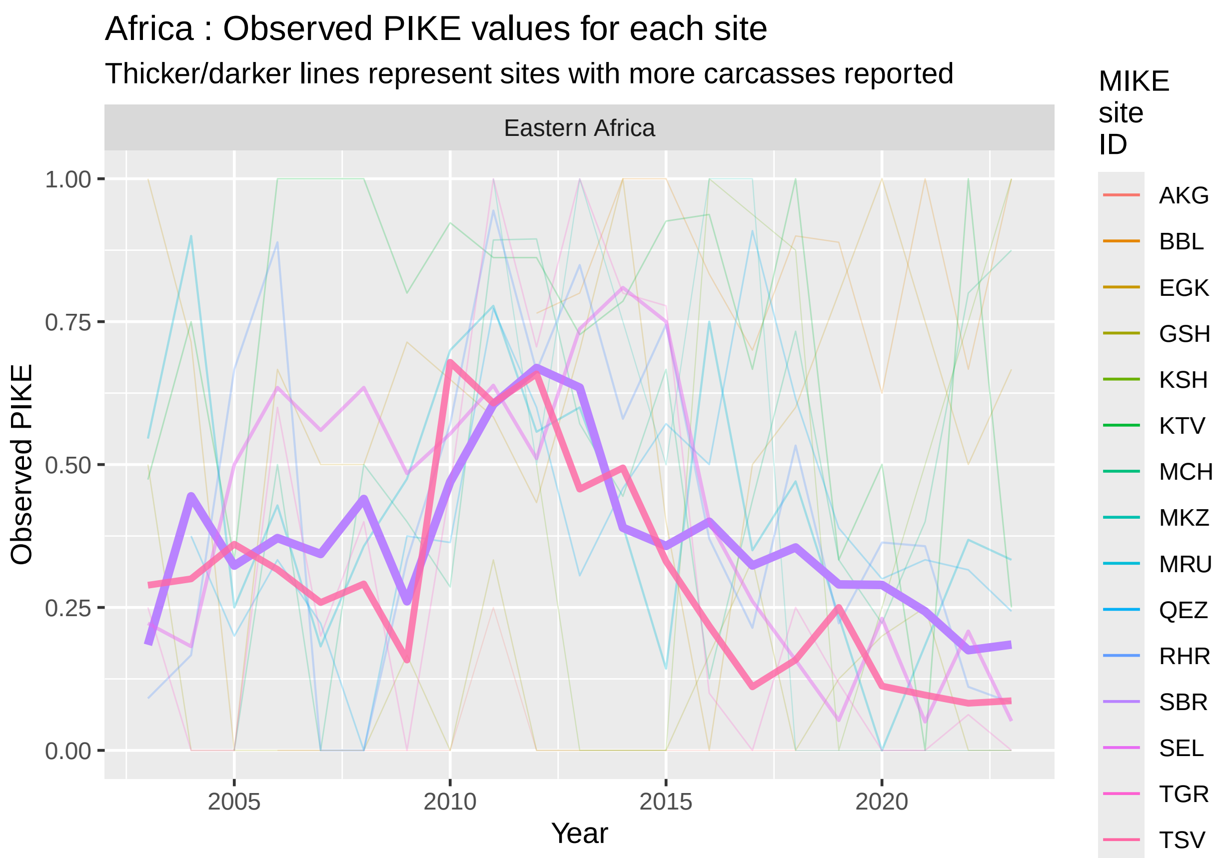
The observed *PIKE* is the value computed from the examined carcasses in a year which we hope reflects the actual *PIKE* for all elephants at the MIKE site. A plot of the observed *PIKE* values from each site-year shows a wide range in the observed *PIKE* values, but many of the observed *PIKE* values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

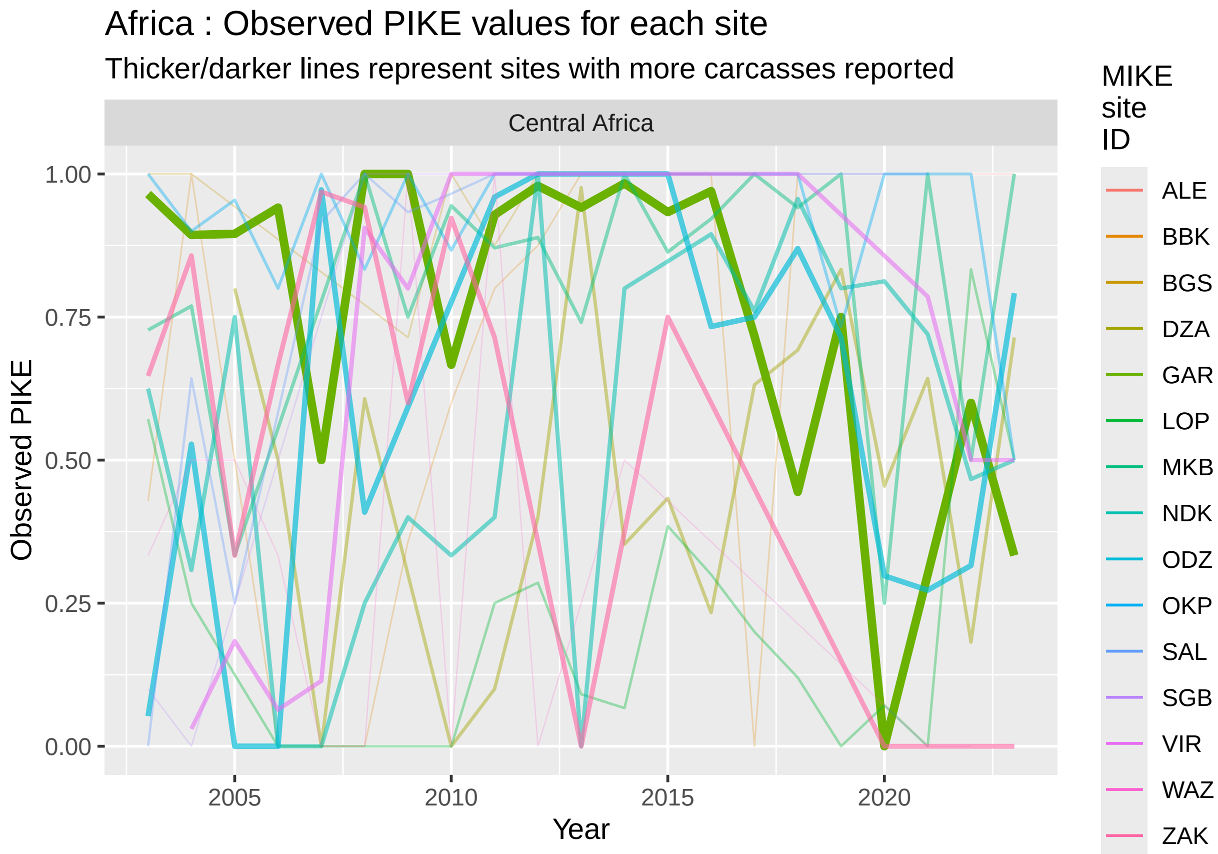


The trend in observed *PIKE* values for each site is:


Africa : Observed PIKE values for each site
Thicker/darker lines represent sites with more carcasses reported
Southern Africa
Observed PIKE
1.00
0.75
0.50
0.25
0.00
2005
2010
2015
2020
Year
site ID
CHE
CHO
ETO
HWA
KFE
KRU
KSG
LGL
LPP
LZN
MAG
MAN
MJT
NIA
NLW
NYA
SLW


Africa : Observed PIKE values for each site
Thicker/darker lines represent sites with more carcasses reported
West Africa
Observed PIKE
1.00
0.75
0.50
0.25
0.00
2005
2010
2015
2020
Year
MIKE site ID
COM
FAZ
GOU
KAK
KER
MAR
MOL
NAZ
NKK
PDJ
SAP
TAI
WBF
WBJ
WNE
YKR


Africa : Observed PIKE values for each site
Thicker/darker lines represent sites with more carcasses reported
Eastern Africa
Observed PIKE
1.00
0.75
0.50
0.25
0.00
2005
2010
2015
2020
Year
MIKE site ID
AKG
BBL
EGK
GSH
KSH
KTV
MCH
MKZ
MRU
QEZ
RHR
SBR
SEL
TGR
TSV



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.

2.3 Final dataset used

The final data set consists of 66 MIKE sites from 2003 to 2023 over the subregions as shown below:

Table 1: Summary of MIKE sites used in analysis

Subregion Name	Number of sites	# Site-Years	Mean # carcasses reported per year	Site IDs
Central Africa	15	231	18.5	ALE, BBK, BGS, DZA, GAR, LOP, MKB, NDK, ODZ, OKP, SAL, SGB, VIR, WAZ, ZAK
Eastern Africa	15	263	43.0	AKG, BBL, EGK, GSH, KSH, KTV, MCH, MKZ, MRU, QEZ, RHR, SBR, SEL, TGR, TSV
Southern Africa	19	243	43.0	CHE, CHO, ETO, HWA, KFE, KRU, KSG, LGL, LPP, LZN, MAG, MAN, MJT, NIA, NLW, NYA, SLW, SMN, ZBZ

West Africa	17	172	5.6	COM, FAZ, GOU, KAK, KER, MAR, MOL, NAZ, NKK, PDJ, SAP, TAI, WBF, WBJ, WNE, YKR, ZIA

3 The Bayesian model

3.1 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.

3.2 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).

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.

3.3 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 two 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.

3.4 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 in the original technical document.

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.

4 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.

4.1 Estimates of site and year.site variance components

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

Table 2: Estimated standard dev of Site and Year.Site effects

			95% CI			
	Mean	SD	Lower	Upper	Rhat	Eff n
sd.site.eff	1.74	0.19	1.41	2.16	1.002	1800
sd.year.site.eff	1.27	0.06	1.16	1.38	1.004	790

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)

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

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

Table 3: Estimated year effects on the logit scale

		95% CI

Year index	Year	Mean	SD	Lower	Upper
1	2003	-0.36	0.35	-1.06	0.32
2	2004	-0.32	0.34	-0.99	0.33
3	2005	-0.52	0.35	-1.21	0.16
4	2006	-0.35	0.38	-1.12	0.37
5	2007	-0.20	0.37	-0.95	0.51
6	2008	0.36	0.35	-0.34	1.07
7	2009	0.25	0.34	-0.41	0.90
8	2010	0.18	0.34	-0.49	0.84
9	2011	1.26	0.33	0.60	1.89
10	2012	1.08	0.33	0.43	1.71
11	2013	0.67	0.32	0.04	1.32
12	2014	1.08	0.33	0.44	1.72
13	2015	0.95	0.33	0.32	1.60
14	2016	0.86	0.34	0.20	1.51
15	2017	0.50	0.33	-0.14	1.15
16	2018	0.27	0.31	-0.35	0.87
17	2019	-0.51	0.32	-1.14	0.13
18	2020	-0.97	0.33	-1.62	-0.32
19	2021	-0.73	0.34	-1.40	-0.07
20	2022	-1.03	0.33	-1.68	-0.38
21	2023	-0.81	0.33	-1.44	-0.16

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.

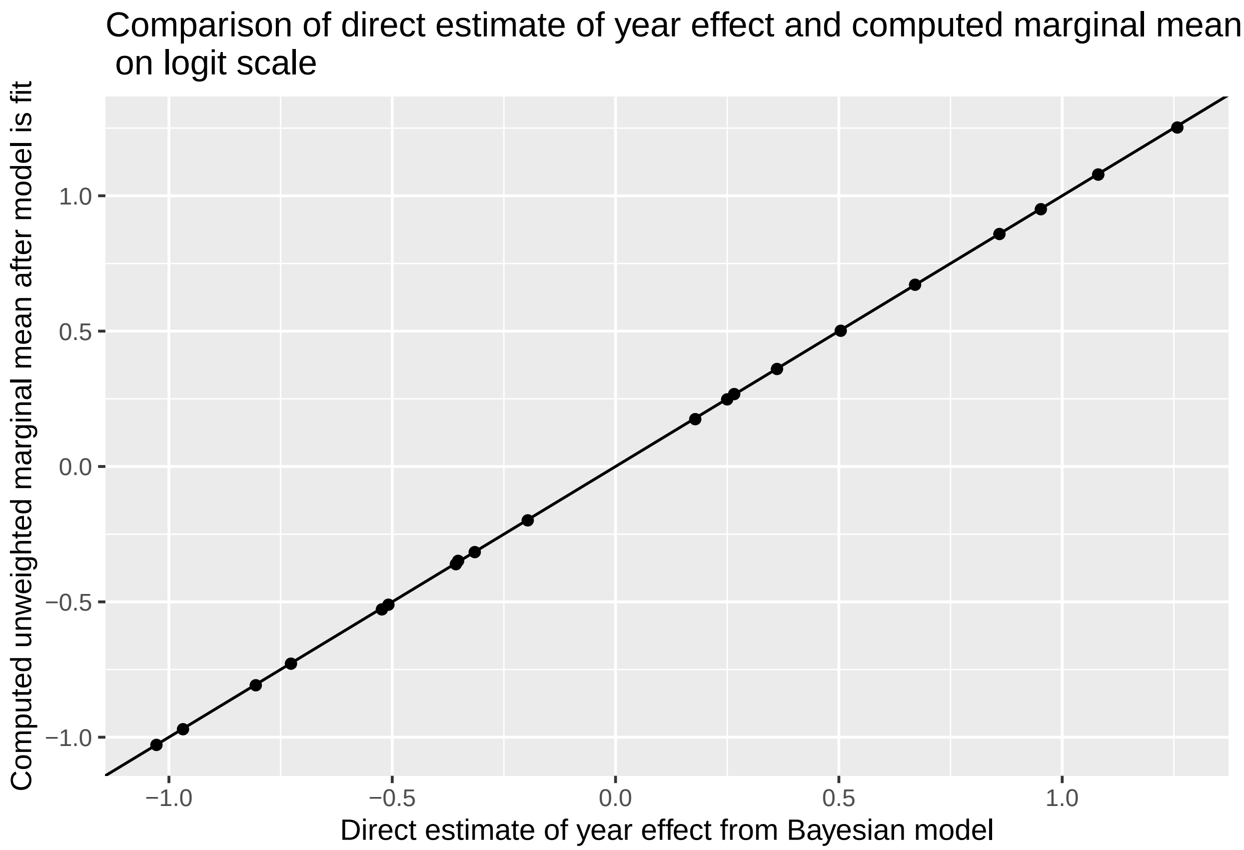
This is contrasted to the marginal means on the *logit* scale, i.e. the marginal mean $logit(PIKE)$ is computed in each year over sites that have data or sites with imputed site.years:

Table 4: Estimated marginal means on the logit scale

				95% CI	
Year index	Year	Mean	SD	Lower	Upper
1	2003	-0.36	0.23	-0.81	0.09
2	2004	-0.32	0.21	-0.73	0.09
3	2005	-0.53	0.23	-0.99	-0.09
4	2006	-0.35	0.27	-0.86	0.17
5	2007	-0.20	0.25	-0.69	0.29
6	2008	0.36	0.23	-0.09	0.82
7	2009	0.25	0.21	-0.16	0.65
8	2010	0.17	0.21	-0.25	0.59
9	2011	1.25	0.19	0.89	1.63
10	2012	1.08	0.19	0.70	1.46
11	2013	0.67	0.19	0.30	1.05
12	2014	1.08	0.20	0.68	1.48
13	2015	0.95	0.20	0.57	1.34

14	2016	0.86	0.21	0.44	1.26
15	2017	0.50	0.20	0.10	0.89
16	2018	0.27	0.18	-0.08	0.62
17	2019	-0.51	0.18	-0.86	-0.15
18	2020	-0.97	0.20	-1.38	-0.57
19	2021	-0.73	0.22	-1.15	-0.30
20	2022	-1.03	0.20	-1.41	-0.63
21	2023	-0.81	0.20	-1.20	-0.43

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



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)

4.3 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 5: Comparison of marginal mean PIKE on [0,1] scale

	anti-logit Year effect		mean antilogit Bootstrap		mean antilogit year.site	
Year	Mean	SD	Mean	SD	Mean	SD
2003	0.41	0.083	0.44	0.051	0.44	0.035
2004	0.42	0.080	0.46	0.050	0.46	0.032
2005	0.38	0.079	0.43	0.050	0.43	0.034
2006	0.42	0.089	0.46	0.054	0.46	0.039
2007	0.45	0.089	0.47	0.053	0.47	0.037
2008	0.59	0.084	0.56	0.050	0.56	0.032
2009	0.56	0.081	0.54	0.049	0.54	0.030
2010	0.54	0.083	0.53	0.049	0.53	0.030
2011	0.77	0.057	0.69	0.041	0.69	0.024
2012	0.74	0.062	0.66	0.044	0.66	0.024
2013	0.66	0.071	0.61	0.046	0.61	0.026
2014	0.74	0.063	0.65	0.045	0.65	0.025
2015	0.72	0.065	0.64	0.045	0.64	0.026
2016	0.70	0.070	0.63	0.047	0.63	0.027
2017	0.62	0.076	0.57	0.046	0.57	0.027
2018	0.56	0.075	0.54	0.046	0.54	0.024
2019	0.38	0.073	0.43	0.047	0.43	0.027
2020	0.28	0.066	0.36	0.046	0.36	0.029
2021	0.33	0.073	0.40	0.048	0.40	0.030
2022	0.27	0.064	0.35	0.044	0.35	0.028
2023	0.31	0.070	0.39	0.046	0.39	0.028

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 $logit(PIKE) = 0.84$.

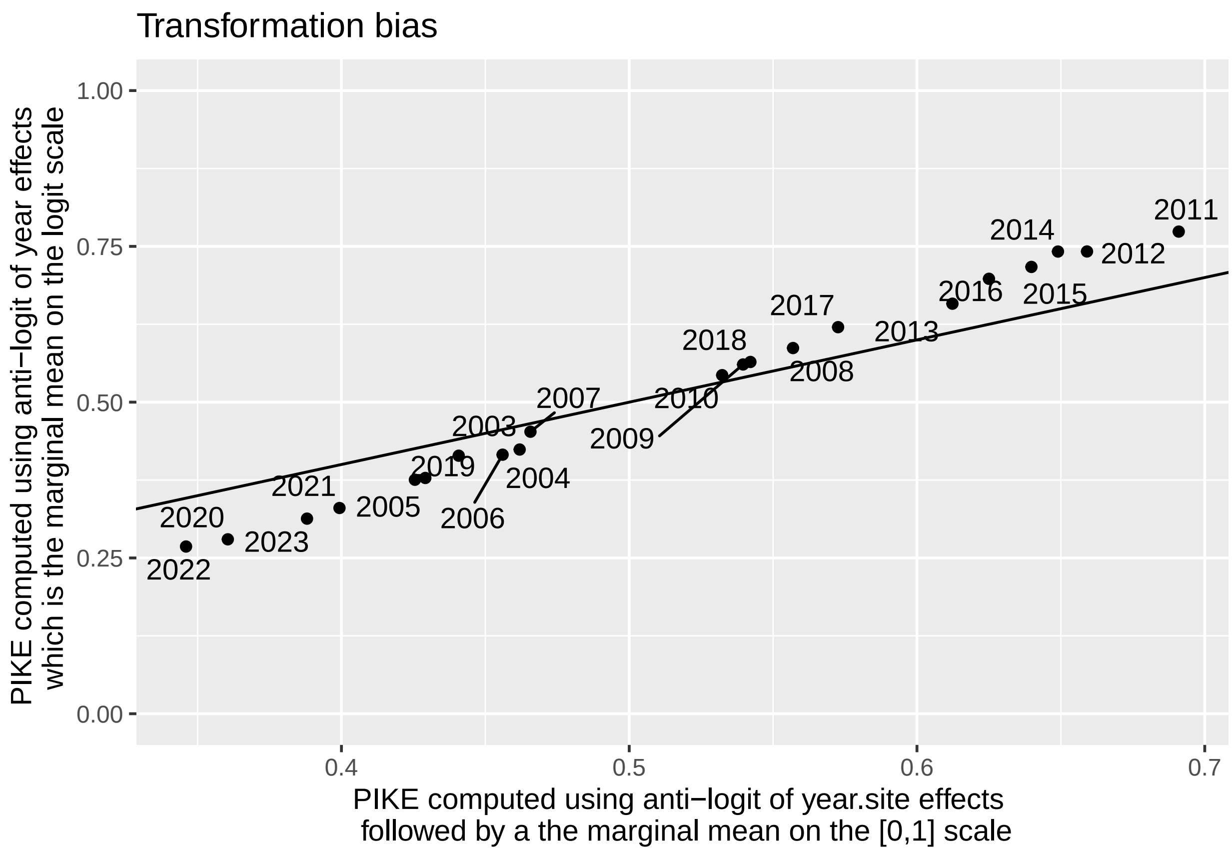
The year effect is estimated as the mean of the logit values

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

and $anit\text{-}logit(1.48) = 0.82$ which is larger than the mean *PIKE* of 0.8.

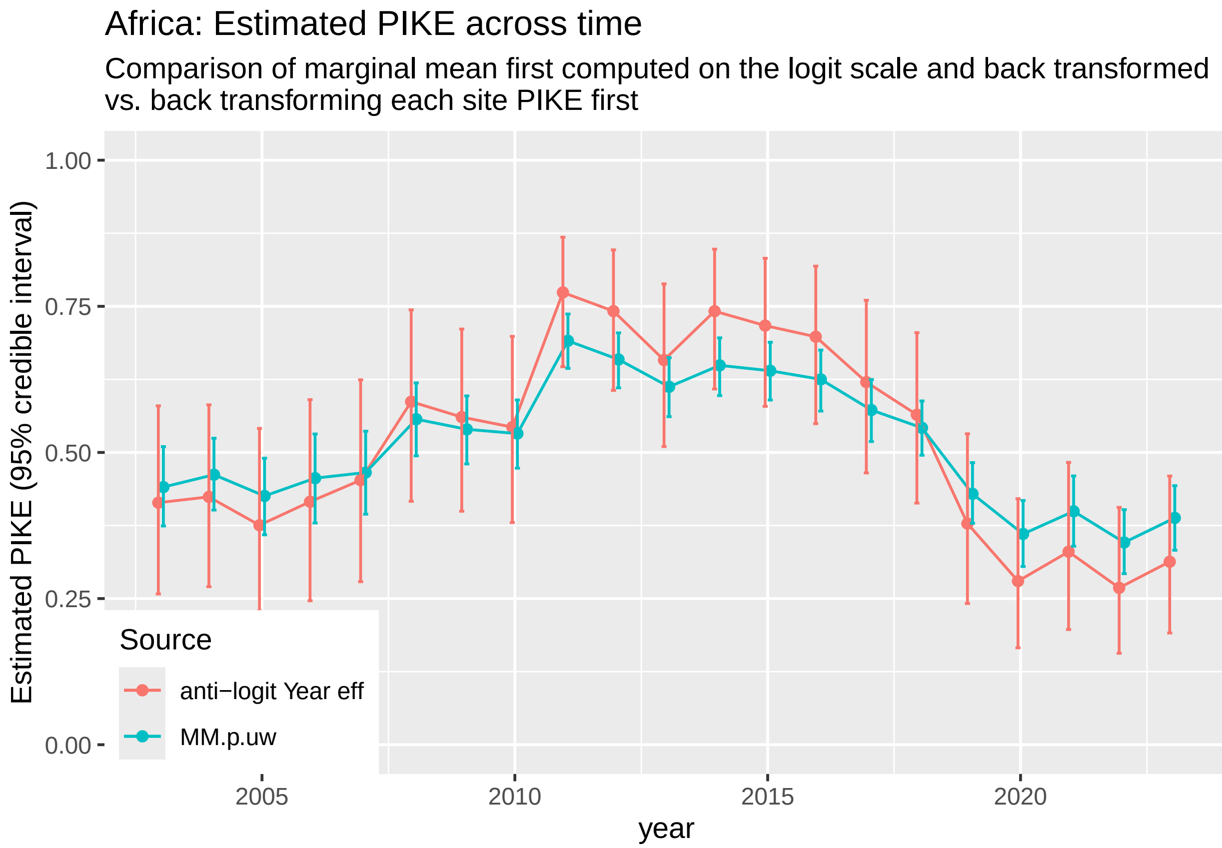
As noted previously, the transformation from the *logit* scale to the probability scale is not linear, and so back transformation of the mean *PIKE* over sites in a year on the *logit* scale is not equal to the mean of the back transformed *PIKE* for a site in a year to the [0,1] scale. The transformation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

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:



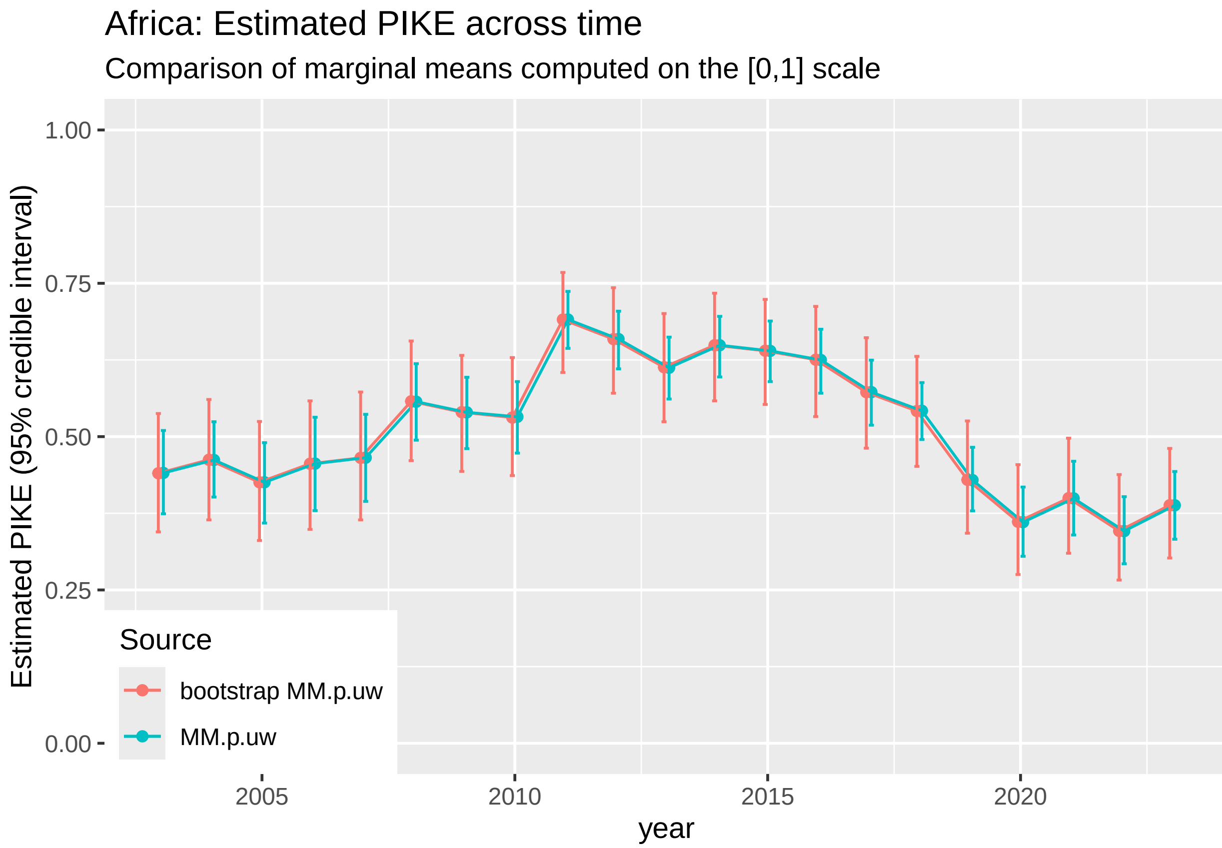
As expected (see earlier sections), a negative bias exists when the marginal mean on the logit-scale is back-transformed to the [0,1] scale when the *PIKE* is < 0.5 and a positive bias when the *PIKE* is > 0.5. This is why we first back transform to the [0,] scale before finding the marginal mean.

If we plot the trends over time:



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:



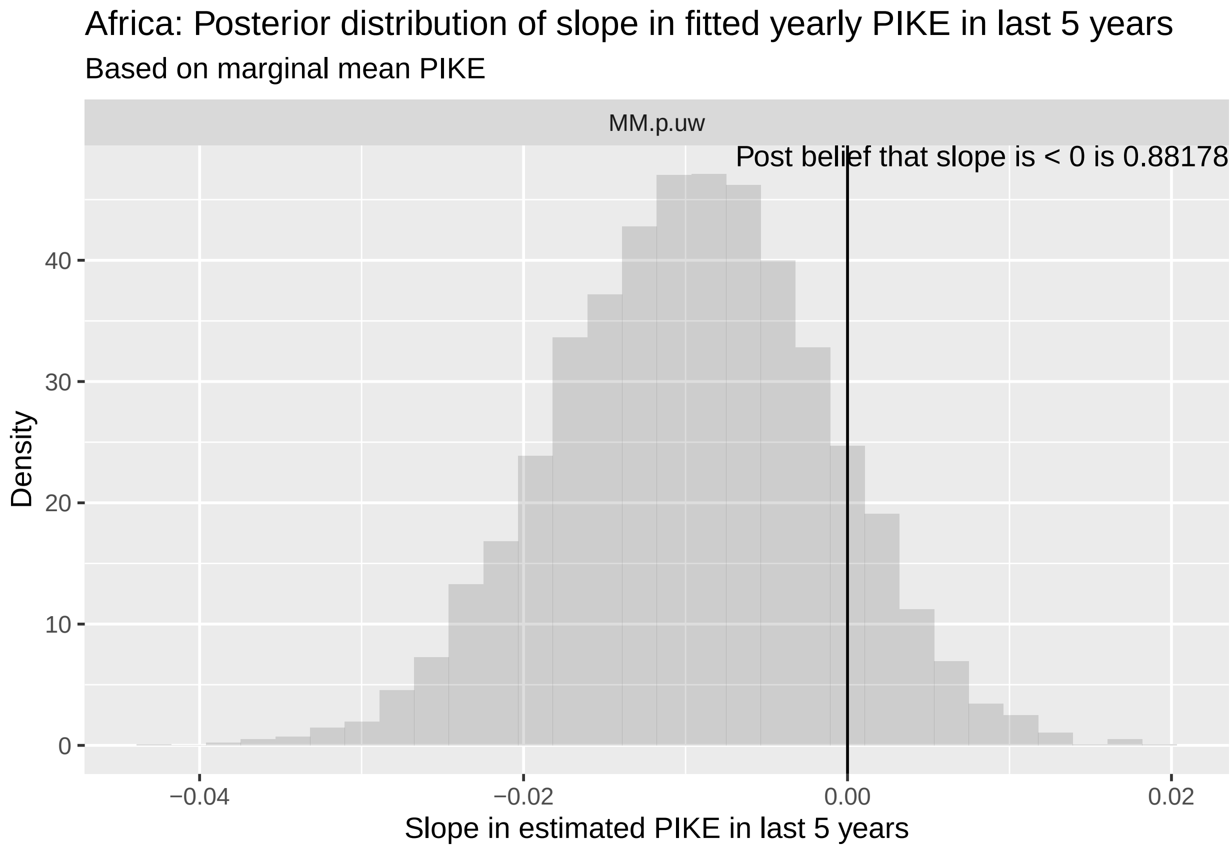
We see that they are identical (as they must be) but the uncertainty is larger in the bootstrapped marginal mean. This is because the uncertainty relates to how we interpret the marginal mean *PIKE*.

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.

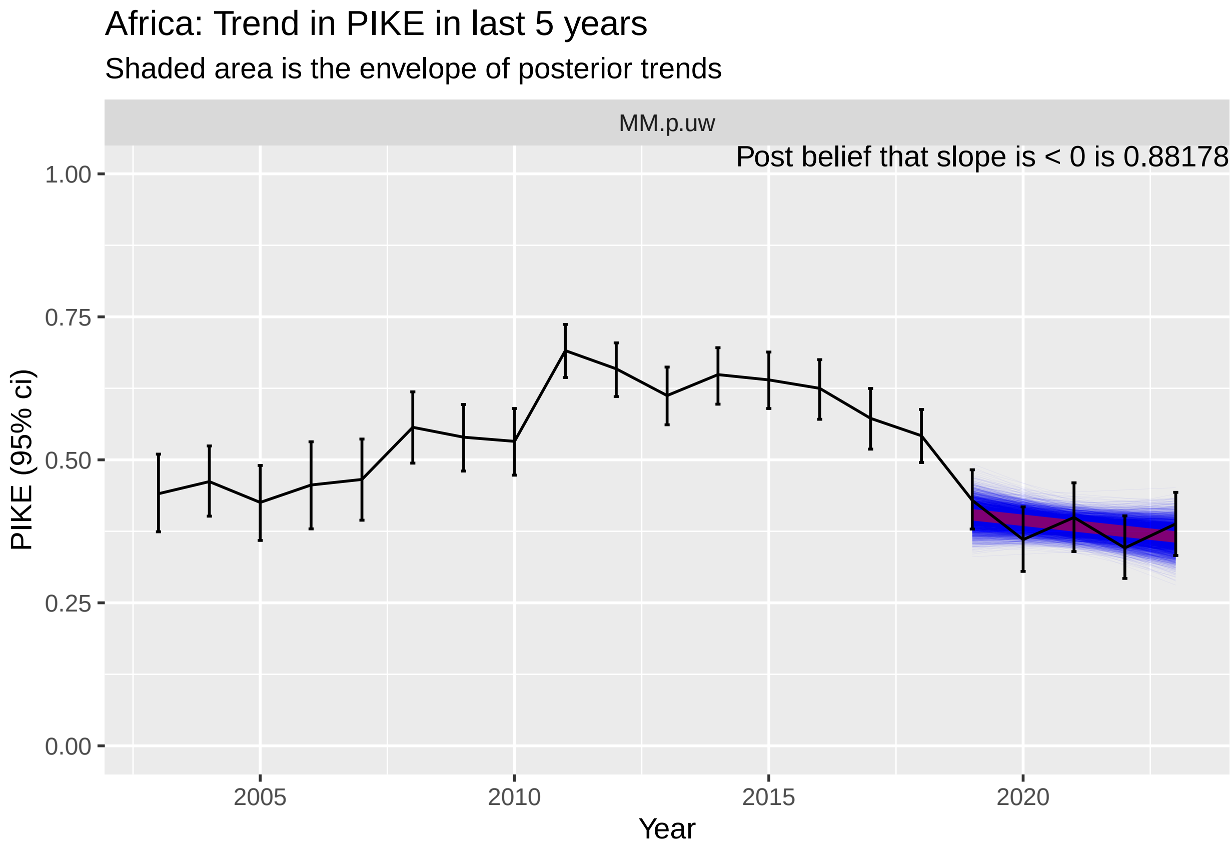
4.4 Posterior belief of trend in *PIKE* in last few years

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 shown in the graph below.



In this case, the posterior belief that the slope in PIKE is negative in the last 5 is given in top right corner of the graph.

The trend for each subregion is shown below, where the shaded area is the envelope of posterior trends.

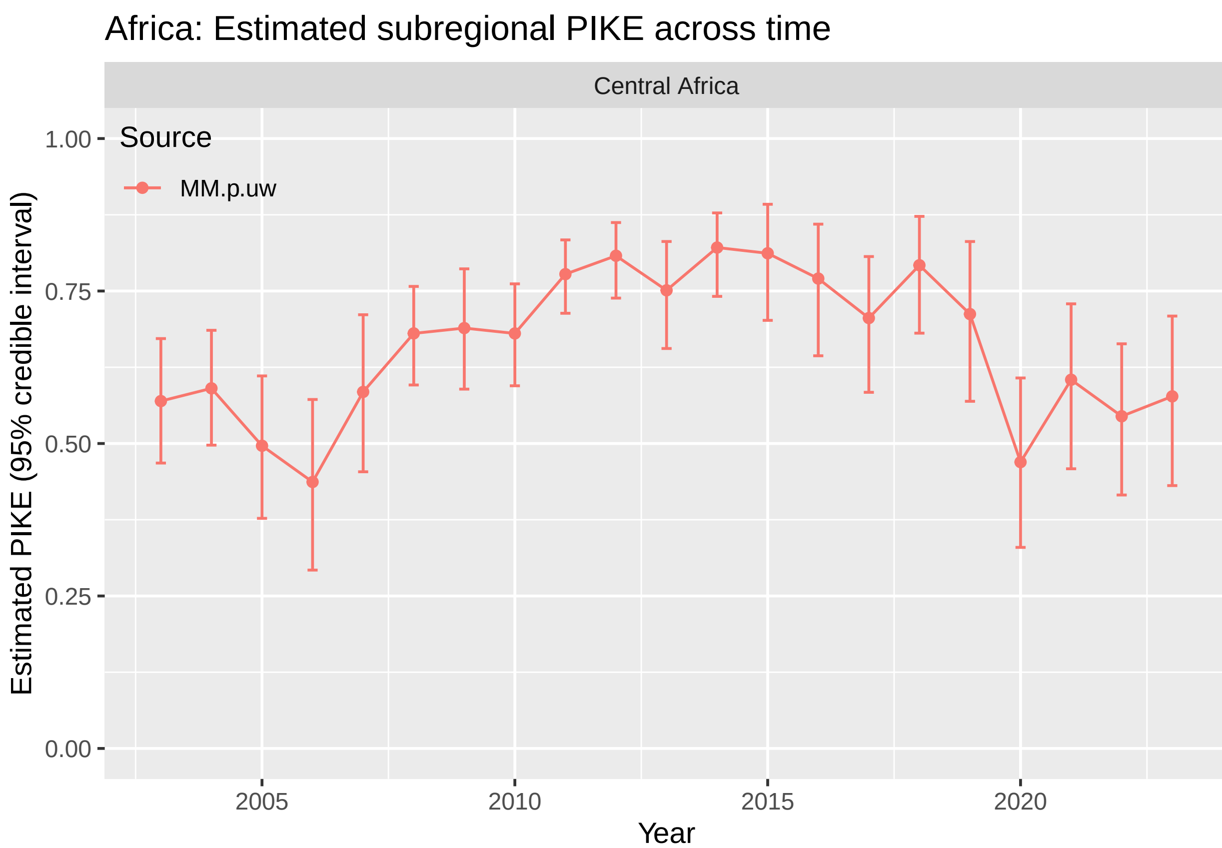


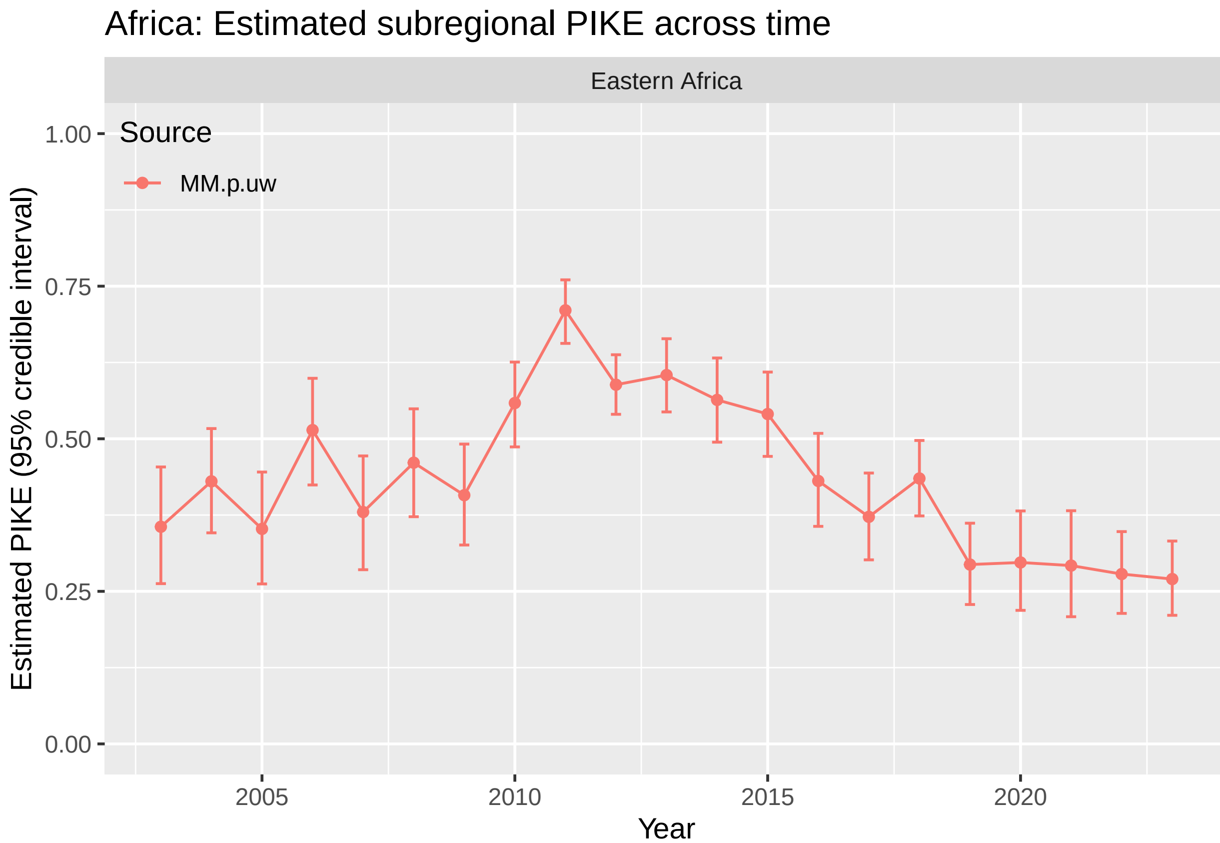
There is a strong posterior belief that the trend in the last 5 years is negative, i.e. the *PIKE* is declining in the last 5 years.

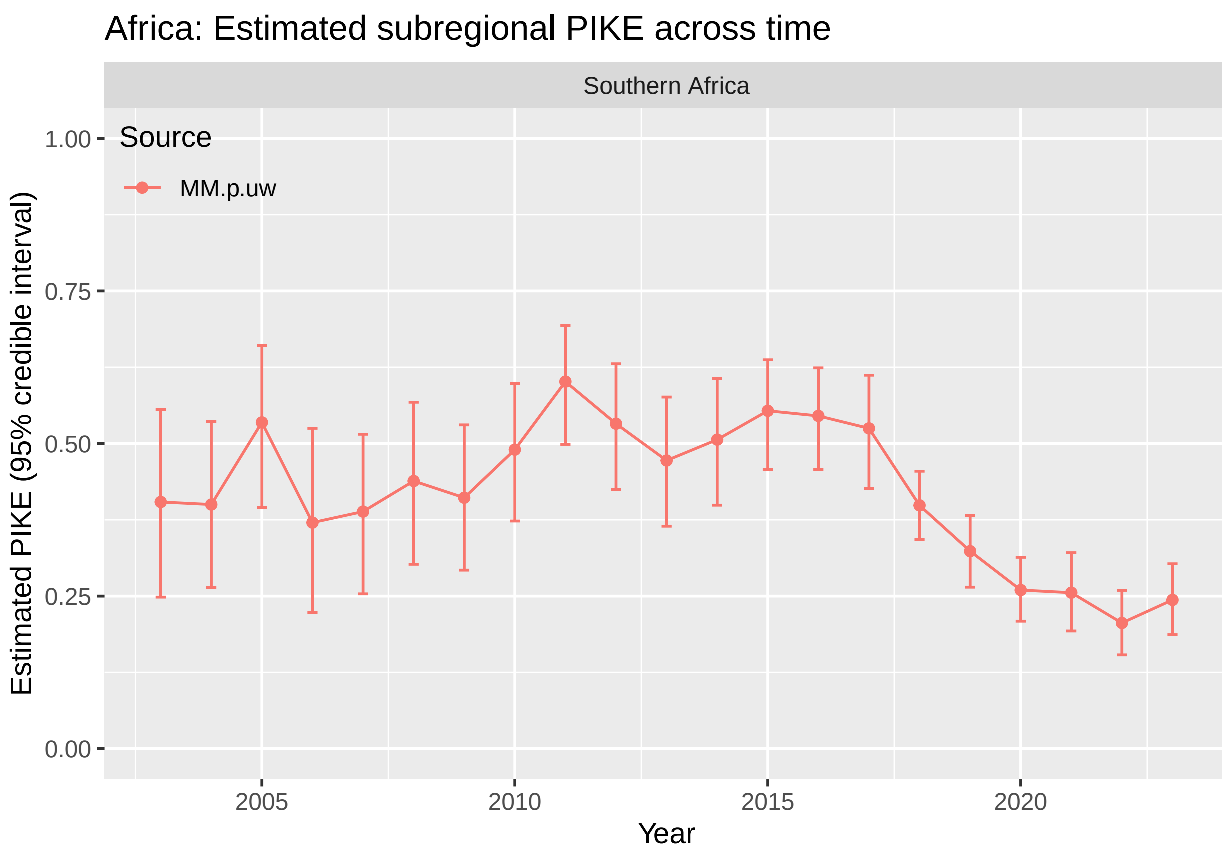
5 Subregional trends in *PIKE*

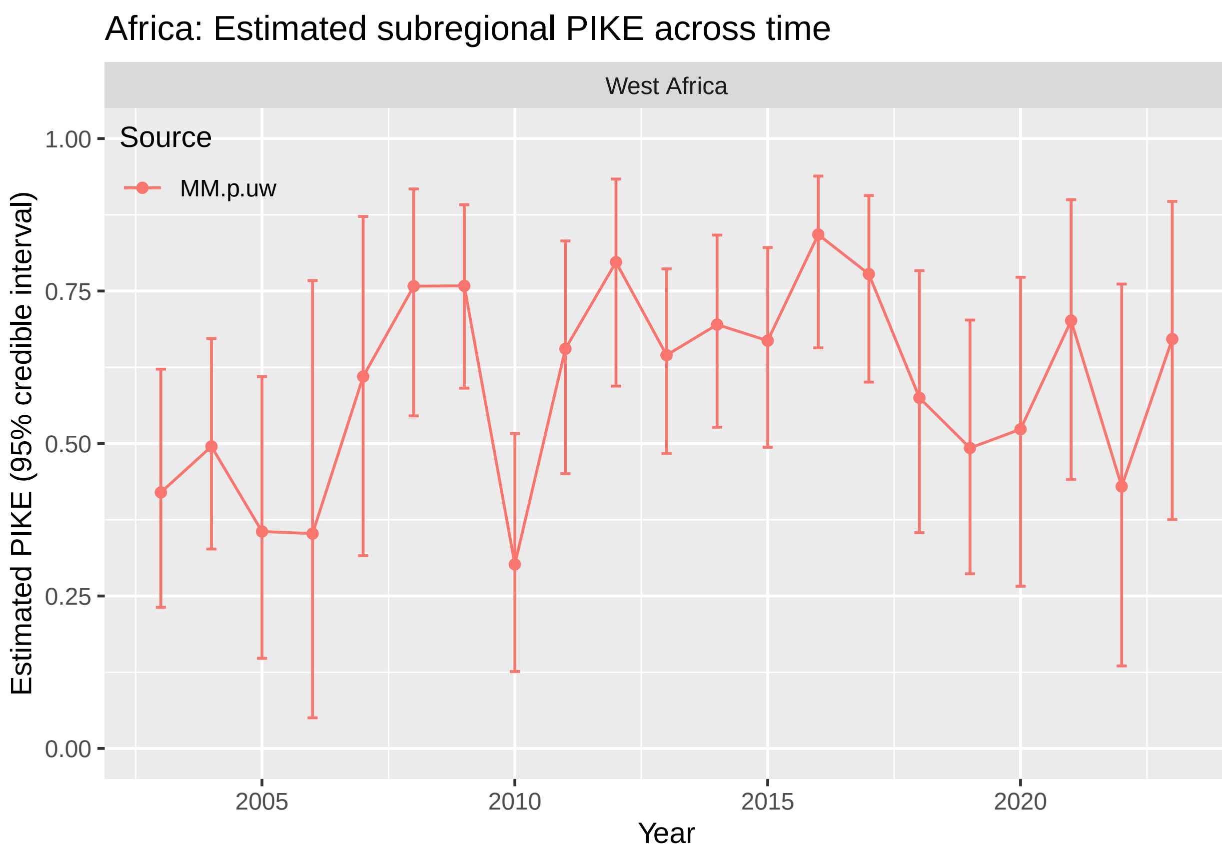
The above analyses were repeated at the sub-regional level. Only the data from each sub-region was used in each analysis, i.e., completely separate analyses were performed for each sub-region.

The following plots show the unweighted marginal *PIKE* values for the four sub-regions:

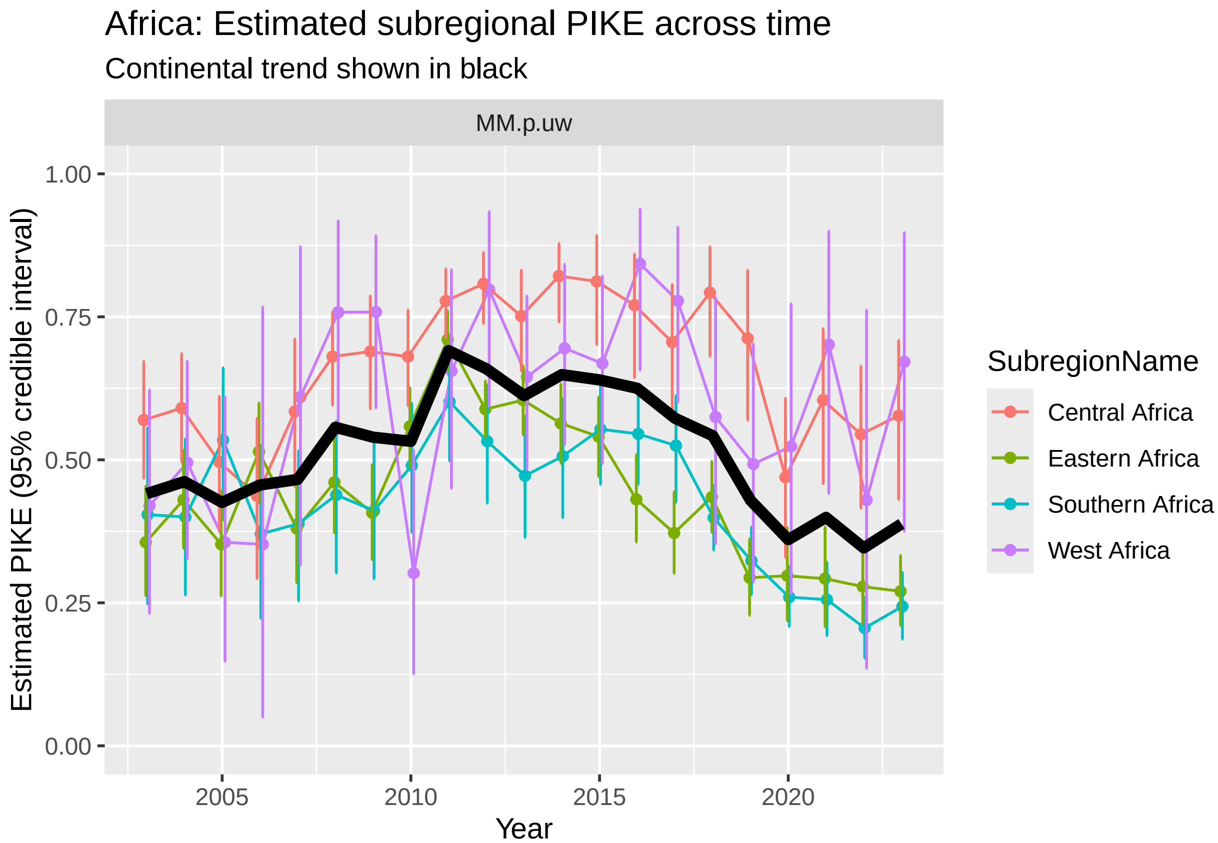

Africa: Estimated subregional PIKE across time
Central Africa
Source
MM.p.uw
Estimated PIKE (95% credible interval)
1.00
0.75
0.50
0.25
0.00
2005
2010
2015
2020
Year


Africa: Estimated subregional PIKE across time
Eastern Africa
Source
MM.p.uw
Estimated PIKE (95% credible interval)
1.00
0.75
0.50
0.25
0.00
2005
2010
2015
2020
Year


Africa: Estimated subregional PIKE across time
Southern Africa
Source
MM.p.uw
Estimated PIKE (95% credible interval)
1.00
0.75
0.50
0.25
0.00
2005
2010
2015
2020
Year



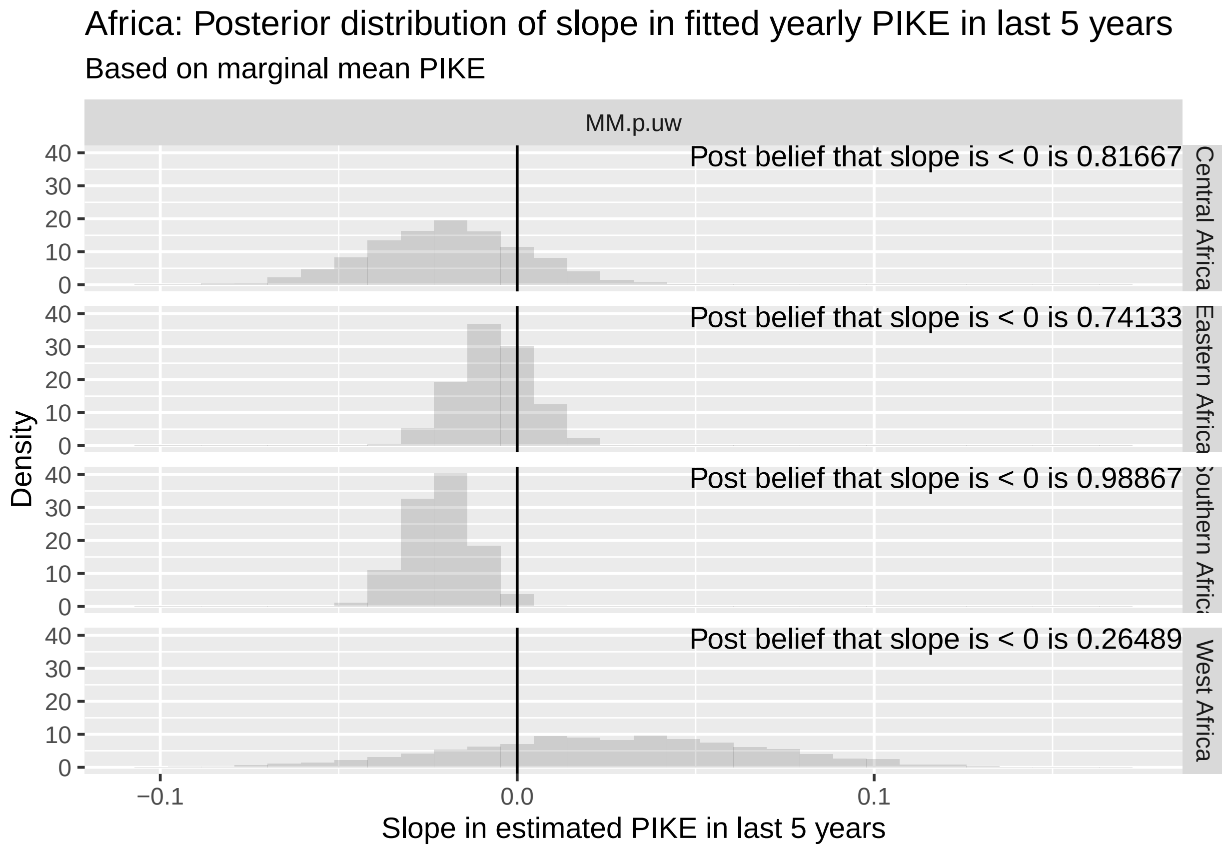
It is interesting to compare the regional trends with the continental trends (shown in black):



In most years, we see that PIKE in Southern and Eastern Africa is lower than the continental PIKE, while PIKE in Central and West Africa is higher than the continental PIKE.

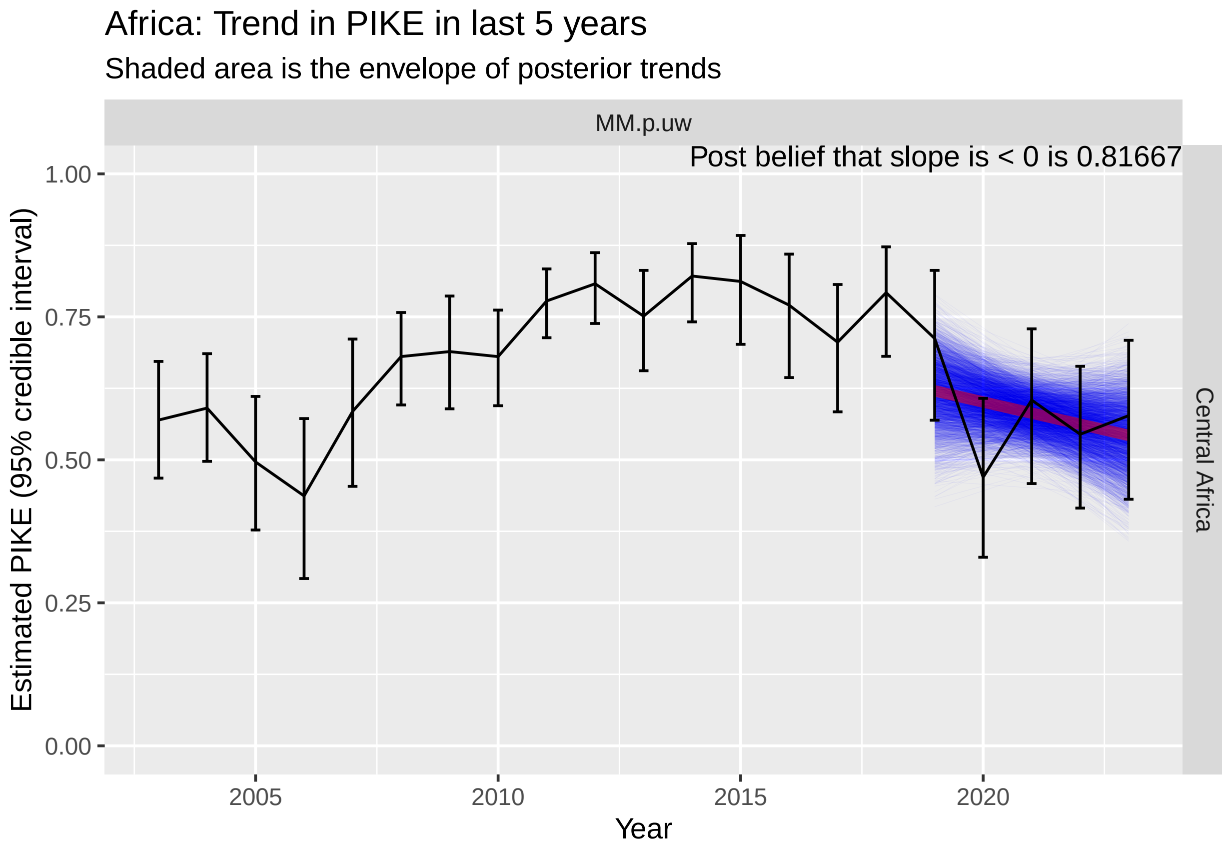
5.1 Posterior belief of trend in $PIKE$ in last few years

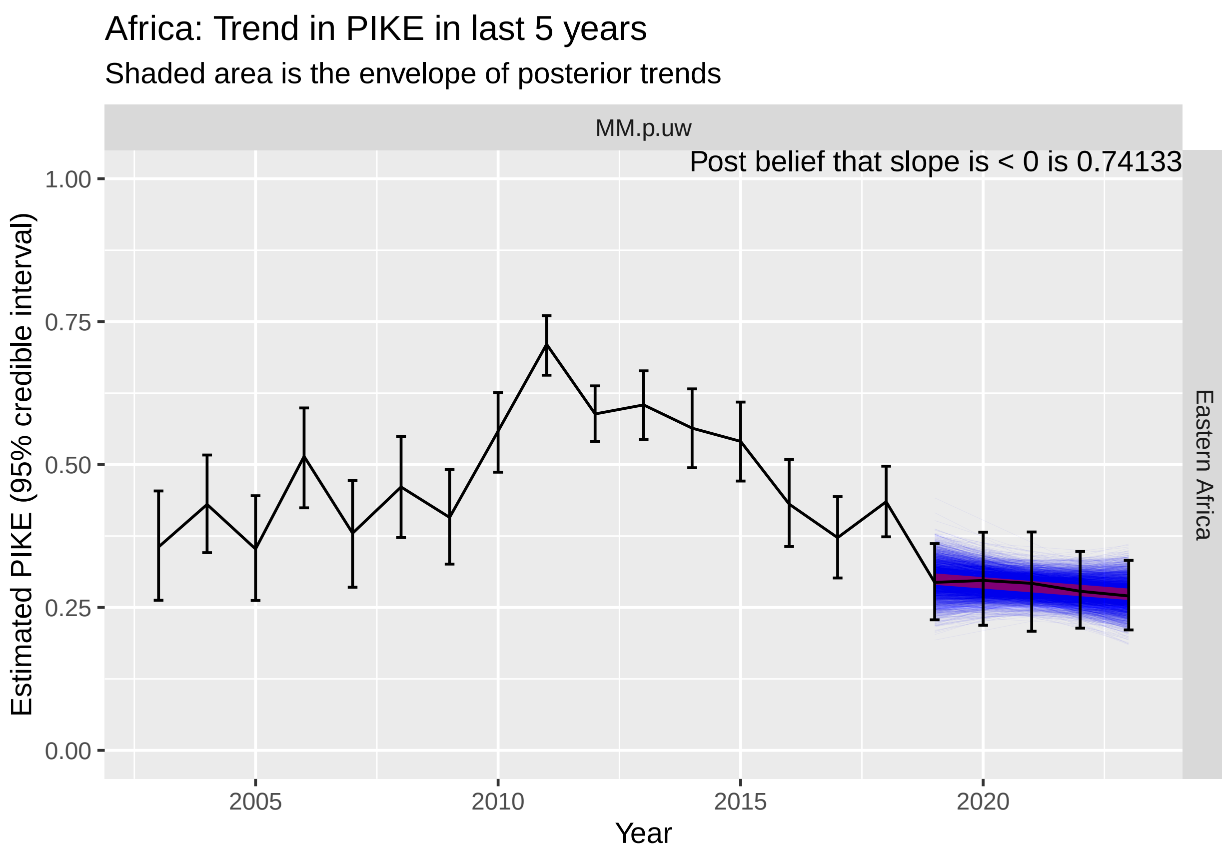
Once the sample from the posterior is available, it is relatively easy to estimate the posterior belief that the trend is negative in the last 5 years in each subregion. This is done by estimating the slope in the last 5 years for each sample from the posterior, and then the posterior belief that the 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 shown below.

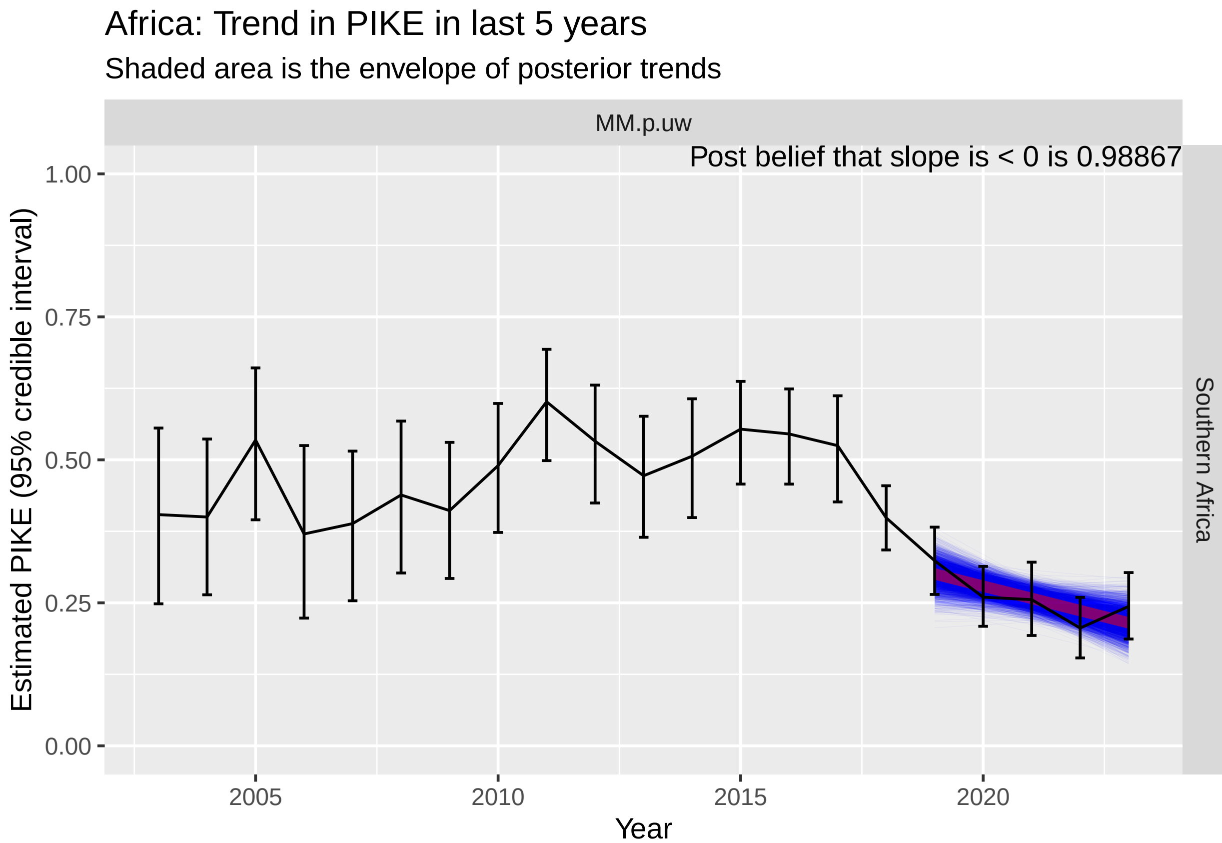


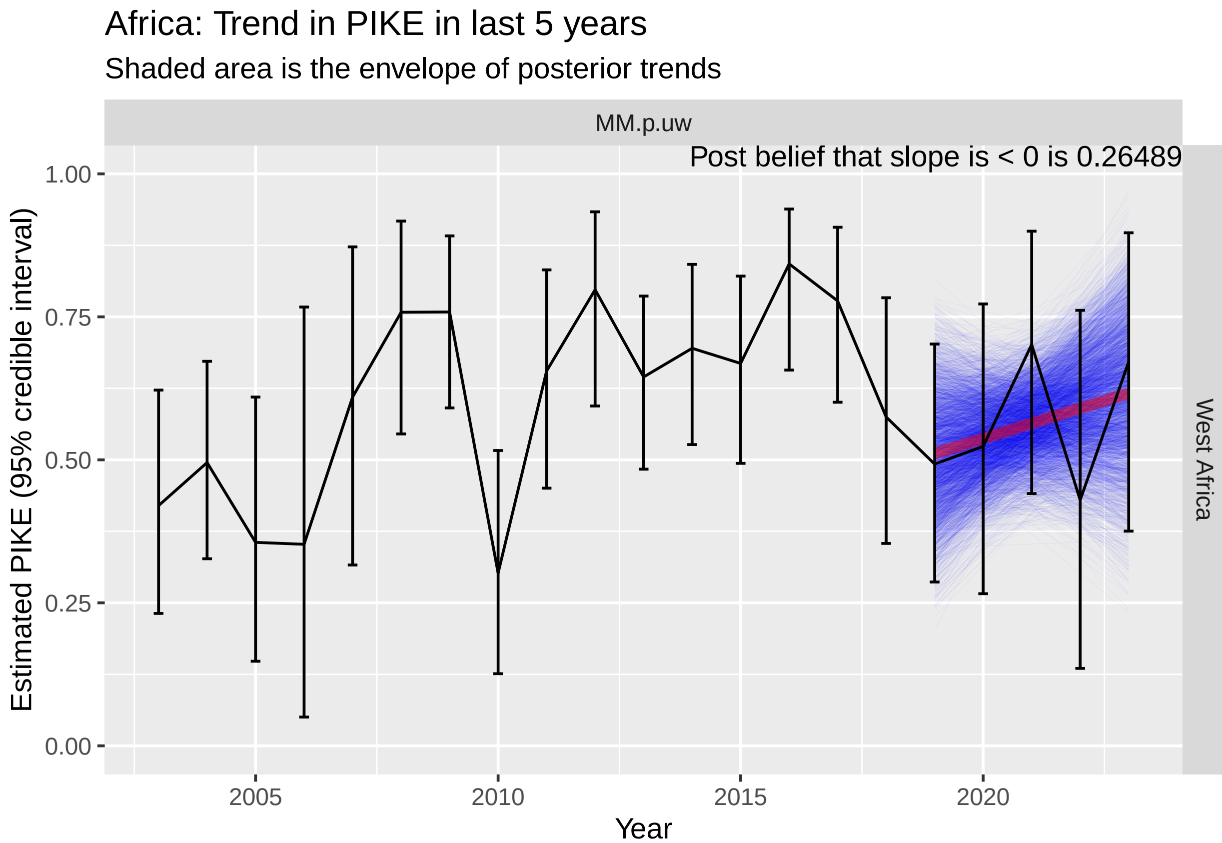
In this case, the posterior belief that the slope in PIKE is negative in the last 5 years is very high for the unweighted *PIKE*‘. The posterior belief that *PIKE* in the last 5 years is declining is displayed on the top-right corner of the graph.

The linear trend for the last 5 years is shown below. The shaded area (in blue) is the envelope of posterior trends.


Africa: Trend in PIKE in last 5 years
Shaded area is the envelope of posterior trends
MM.p.uw
Post belief that slope is < 0 is 0.81667
Central Africa
Estimated PIKE (95% credible interval)
1.00
0.75
0.50
0.25
0.00
2005
2010
2015
2020
Year


Africa: Trend in PIKE in last 5 years
Shaded area is the envelope of posterior trends
MM.p.uw
Post belief that slope is < 0 is 0.74133
Eastern Africa
Estimated PIKE (95% credible interval)
1.00
0.75
0.50
0.25
0.00
2005
2010
2015
2020
Year


Africa: Trend in PIKE in last 5 years
Shaded area is the envelope of posterior trends
MM.p.uw
Post belief that slope is < 0 is 0.98867
Southern Africa
Estimated PIKE (95% credible interval)
1.00
0.75
0.50
0.25
0.00
2005
2010
2015
2020
Year



The posterior belief that the trend in the last 5 is negative is displayed on the top-right corner of the graph.

6 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.

6.1 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 6: Rhat and Effective sample size for several parameter sets

Effect	Max Rhat	Max N.eff	Min N.eff

SD Site effects	1.002	1800	1800
SD Year Site effects	1.004	790	790
Site Effects	1.013	4500	160
Year Effects	1.004	4500	580

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

The effective sample size is small (<500) for 1 sites. The sites with small effective sample sizes are:

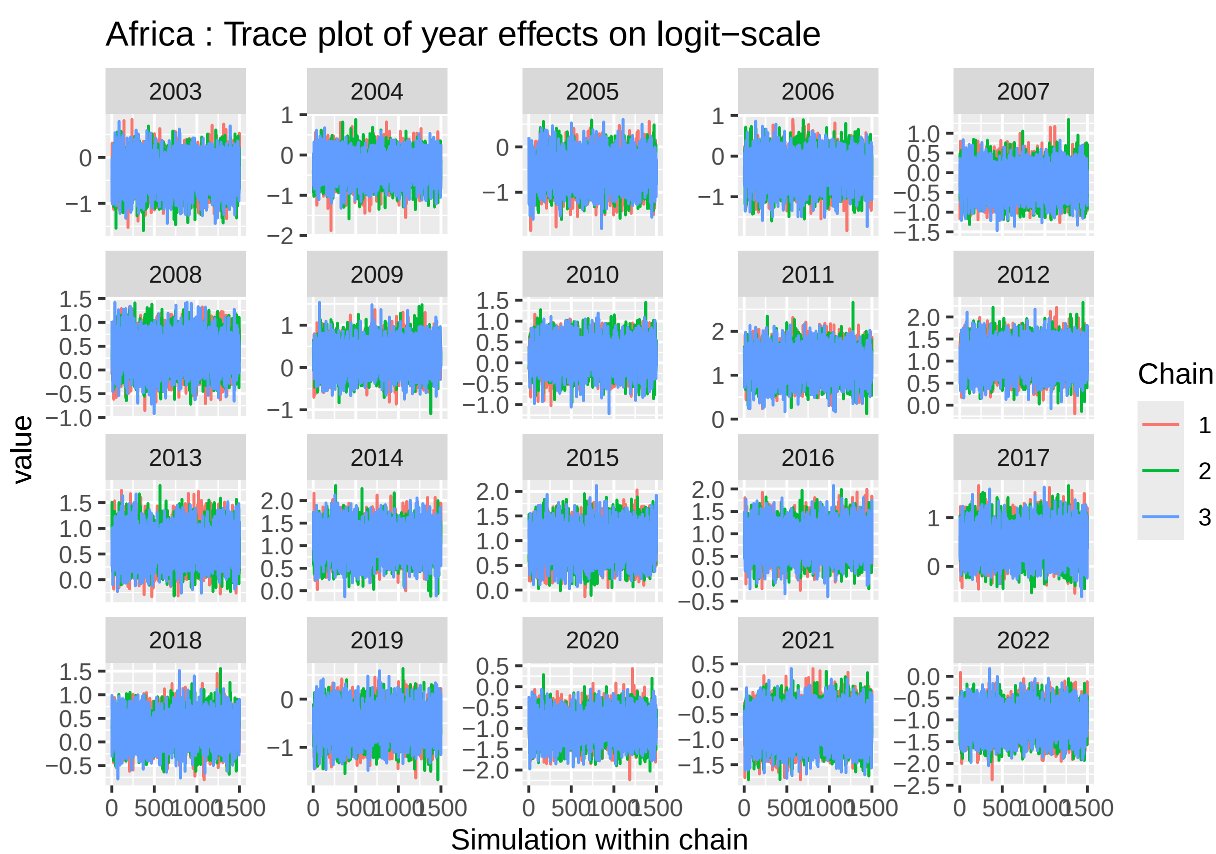
Table 7: Sites with small effective sample sizes

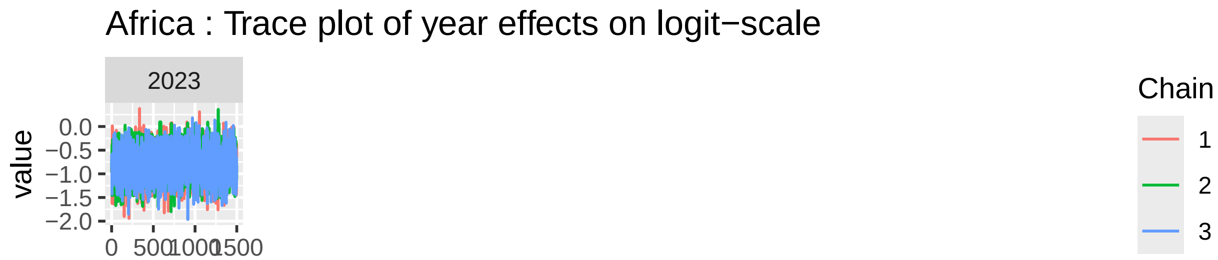
MIKE site	Avg PIKE	Site effect	Rhat	N eff
BGS	0.96	2.77	1.013	160

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.

6.2 Examination of trace plots

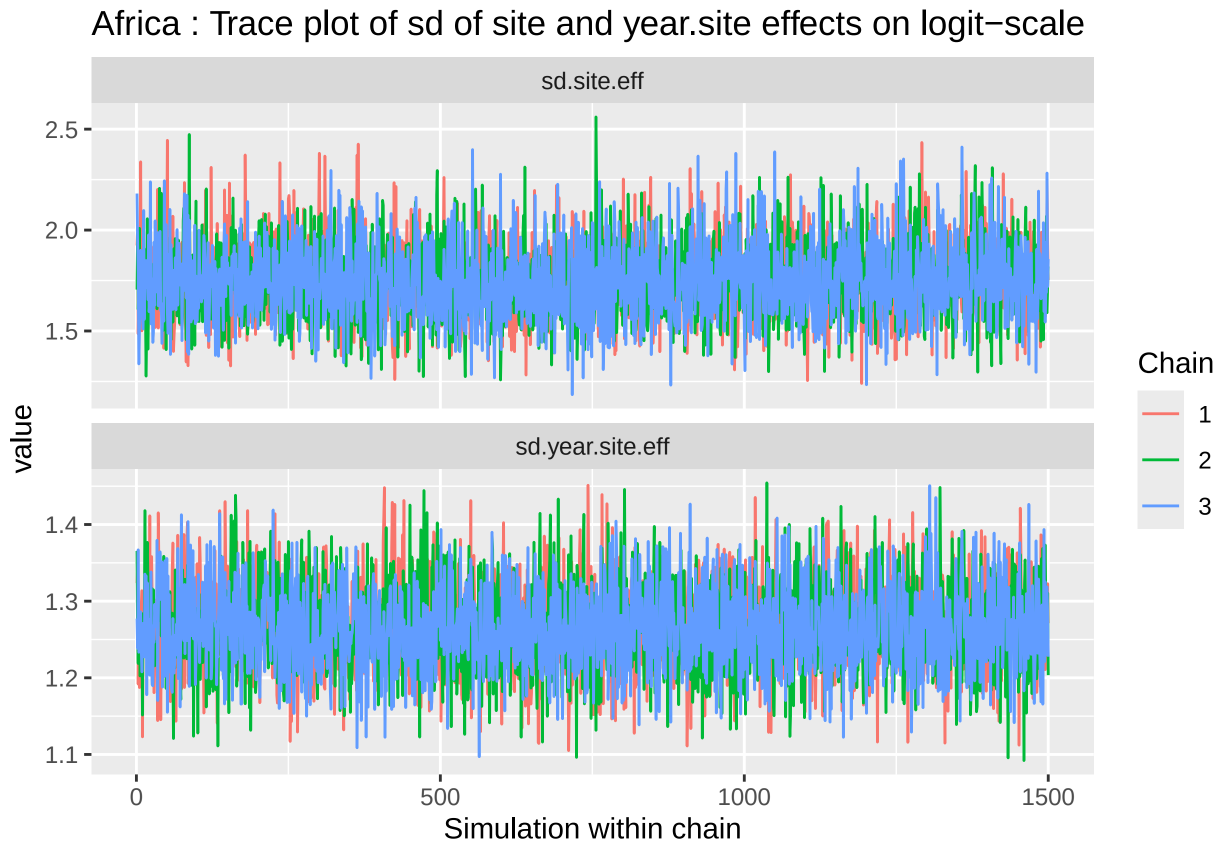
Trace plots were constructed for the yearly estimates of *PIKE* on the *logit* scale:





Simulation within chain

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



All plots show good evidence of mixing of the three chains sampled from the posterior.

6.3 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:

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

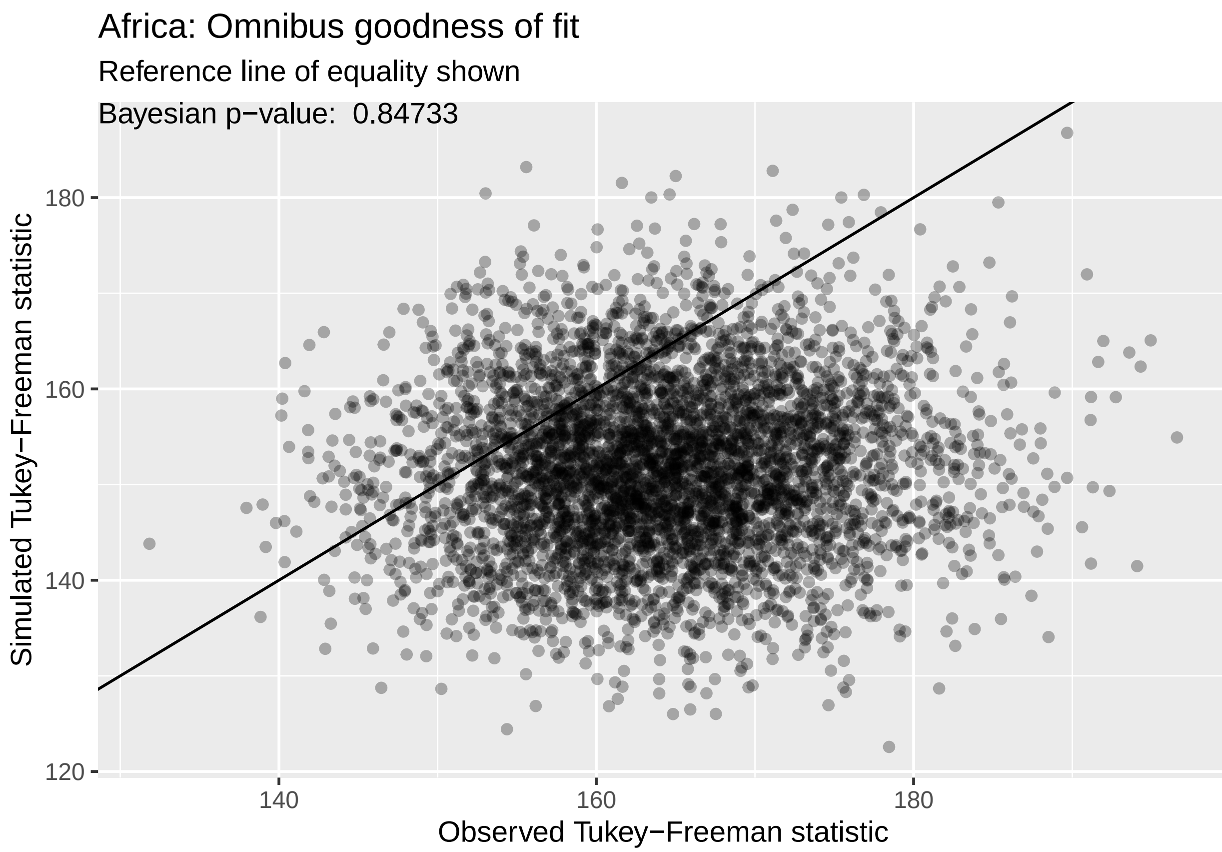
The simulated Tukey-Freeman statistic is found as the difference between a simulated number of illegally killed elephants and the expected number of illegally killed elephants:

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

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

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:



Because the Bayesian p-value is not extreme, the fit is deemed acceptable.

6.4 Over dispersion

6.4.1 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 909 site-year data points in the sum above.

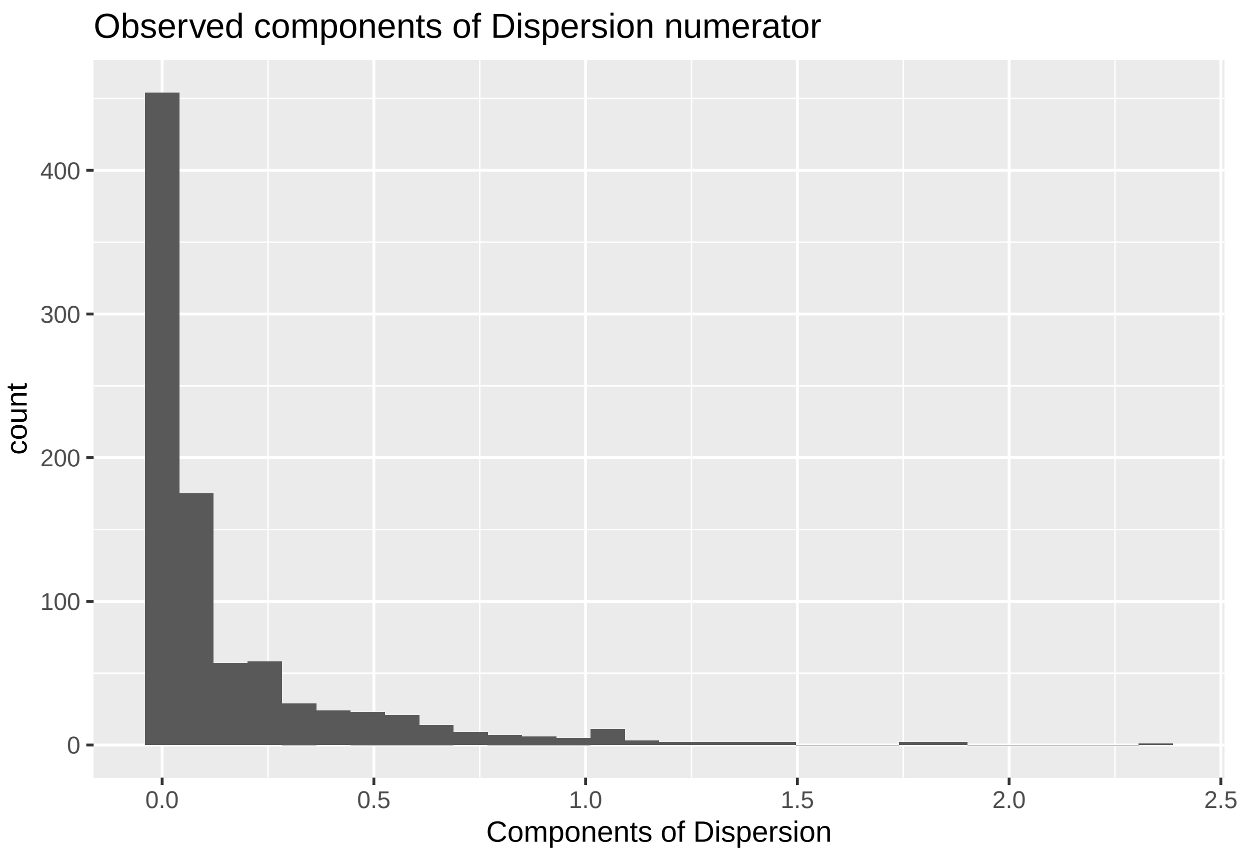
This is traditionally divided by the ($number\ of\ data\ points - the\ number\ of\ estimated\ parameters$). However, in Bayesian hierarchical models (such as this), the number of parameters is ill-defined. For example, we model site-effects as random variables from a common distribution. Is the number 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 996 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=867.9 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 $OD = 3.6$. 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:



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 8: 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
WAZ	2012	5	0	0.00	0.20	1.00	1.00
TGR	2023	34	0	0.00	0.03	1.01	1.01
KRU	2013	17	0	0.00	0.06	1.01	1.01
NDK	2006	5	0	0.00	0.20	1.01	1.01
GAR	2020	3	0	0.00	0.34	1.01	1.01
SAL	2003	2	0	0.00	0.51	1.02	1.02

LOP	2022	6	5	0.83	0.53	3.19	1.03
QEZ	2008	6	0	0.00	0.17	1.05	1.05
YKR	2015	2	0	0.00	0.52	1.05	1.05
GSH	2003	4	2	0.50	0.25	0.98	1.06
NYA	2018	10	0	0.00	0.11	1.06	1.06
ZIA	2012	2	0	0.00	0.54	1.07	1.07
PDJ	2010	6	0	0.00	0.18	1.08	1.08
KRU	2008	28	0	0.00	0.04	1.11	1.11
VIR	2004	33	1	0.03	0.08	2.76	1.12
ETO	2003	21	1	0.05	0.02	0.36	1.12
BBK	2006	12	0	0.00	0.10	1.21	1.21
WBF	2019	4	0	0.00	0.30	1.22	1.22
PDJ	2019	18	0	0.00	0.07	1.30	1.30
SMN	2018	8	0	0.00	0.16	1.31	1.31
MAG	2009	5	0	0.00	0.27	1.34	1.34
NIA	2020	11	0	0.00	0.13	1.39	1.39
ZAK	2023	15	0	0.00	0.10	1.43	1.43
ZIA	2010	6	0	0.00	0.24	1.44	1.44
CHO	2016	121	0	0.00	0.01	1.75	1.75
NDK	2013	10	0	0.00	0.18	1.76	1.76
GOU	2010	27	0	0.00	0.07	1.83	1.83
TAI	2004	2	2	1.00	0.39	0.79	1.87
ODZ	2005	73	0	0.00	0.03	2.35	2.35

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.

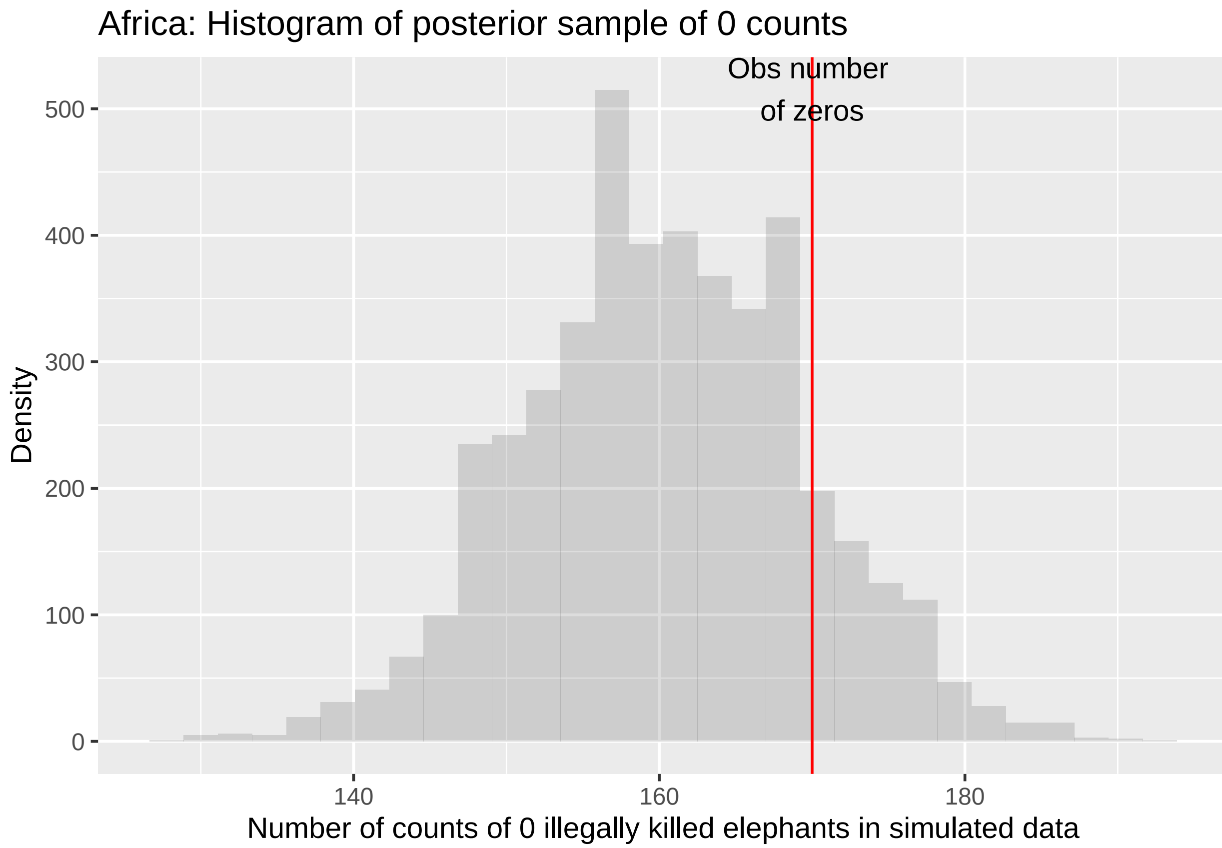
6.4.2 Overdispersion due to 0 counts

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 170 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:

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

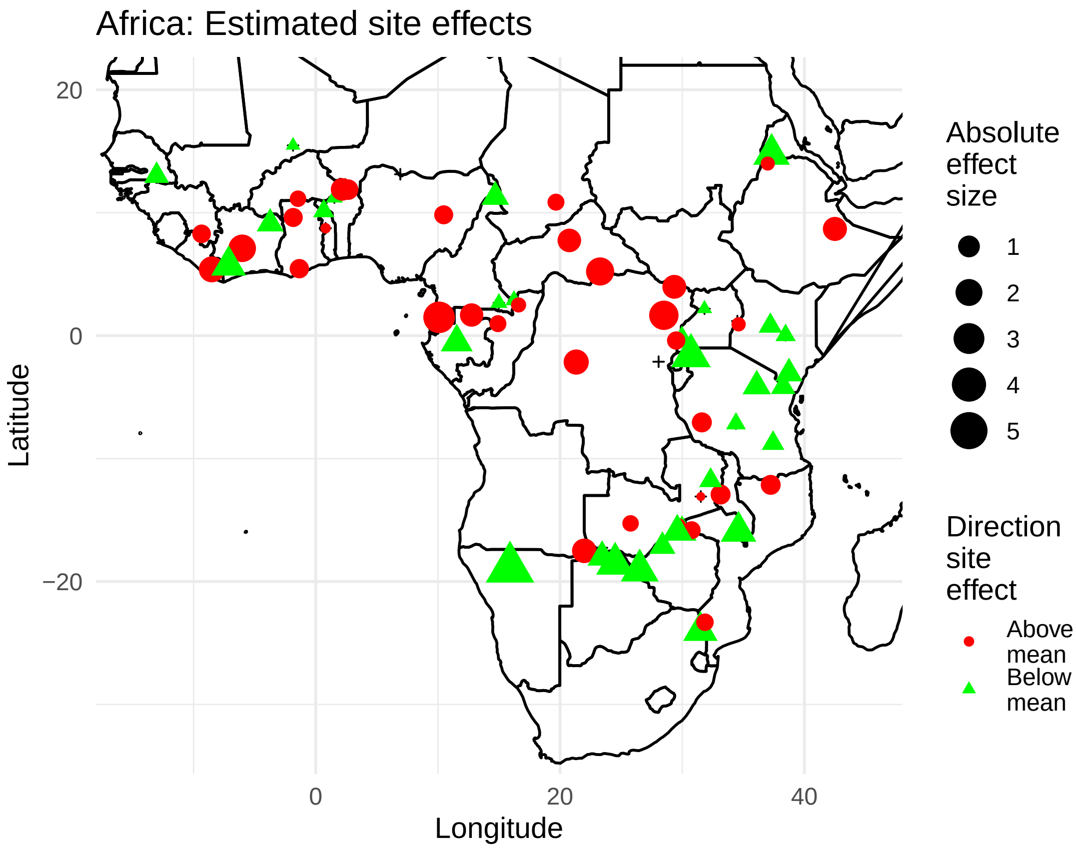
and count the number of times a count of 0 is obtained. This is compared to the observed number of times a 0 is obtained.



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

6.5 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.



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.

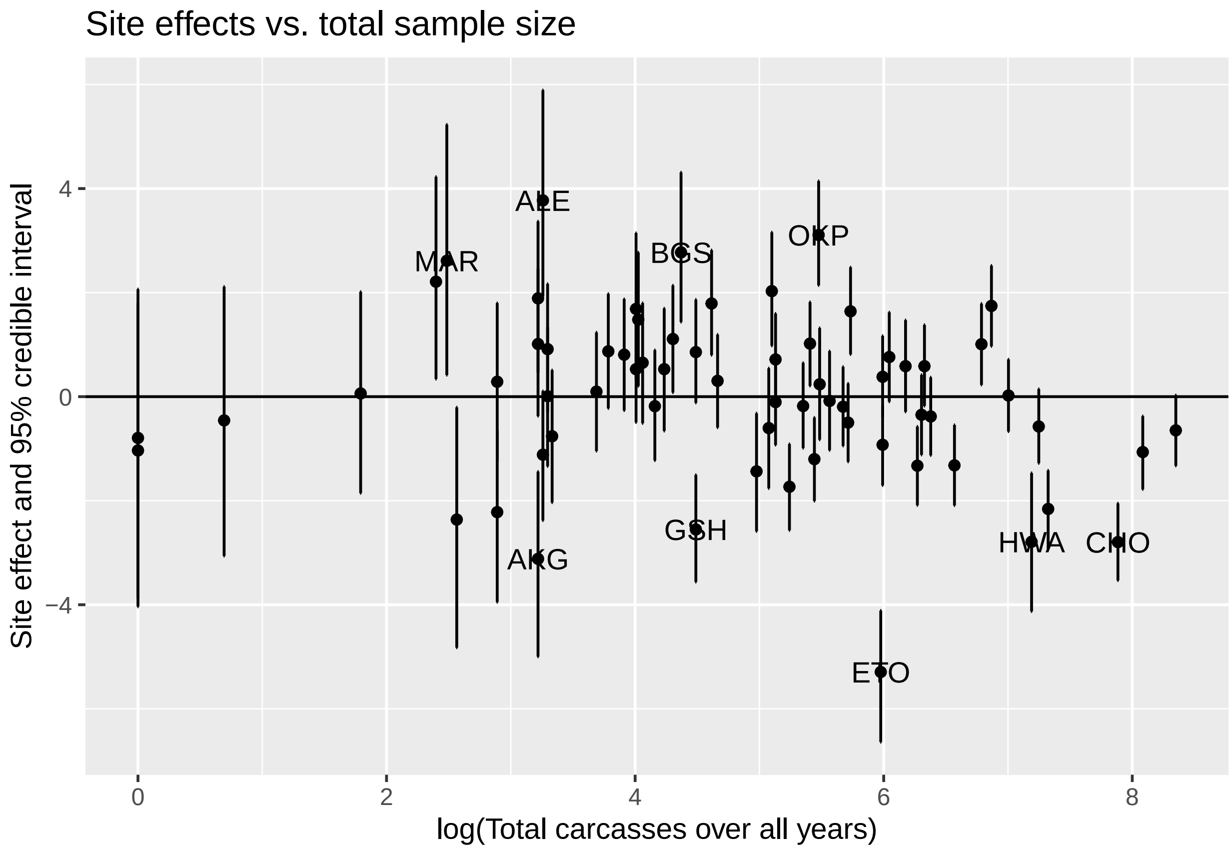
The current analysis is still valid, but inefficient because it has not used the spatial correlation to improve inference. If spatial autocorrelation is explicitly modelled, then information is shared among sites that are geographically close, i.e., if the *PIKE* increases in one site, then spatial autocorrelation 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.

6.6 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:

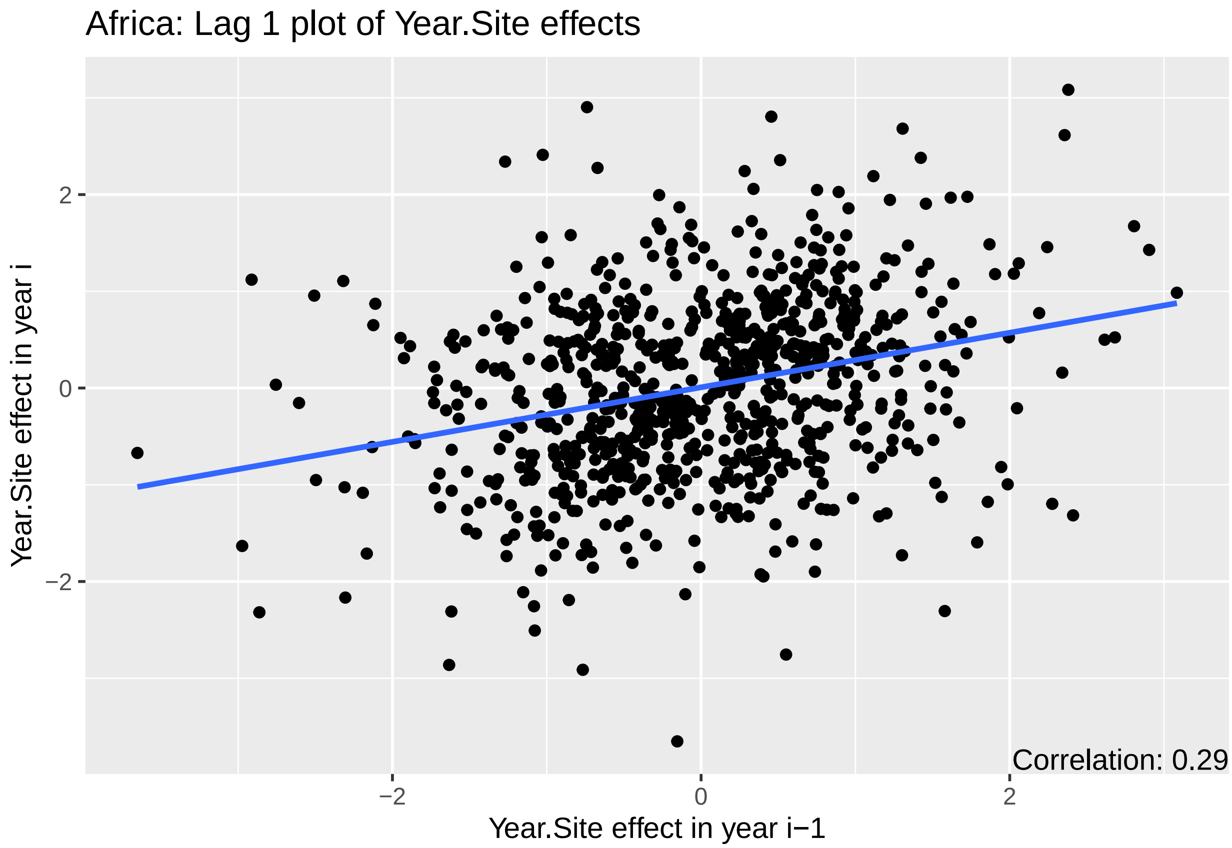


This plot shows that the uncertainty in the site effects declines with the total number of carcasses observed (as expected), and a random scatter about 0 (also as expected). There are a few MIKE sites with extreme site effects as labelled in the plot.

6.7 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.

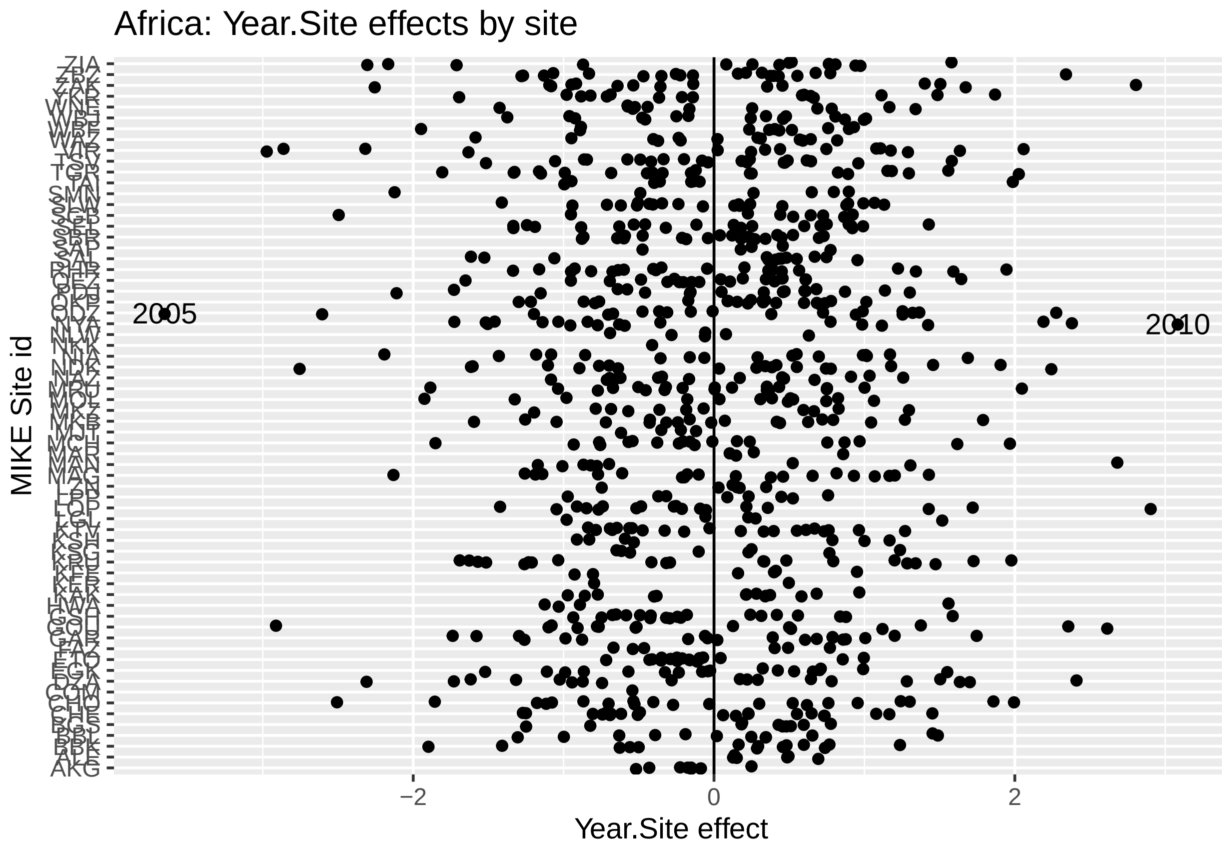
A plot of the $Year.Site_i$ vs. the $Year.Site_{i-1}$ (i.e. a lag 1 plot) is:



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.

6.8 Year.Site effect for each site

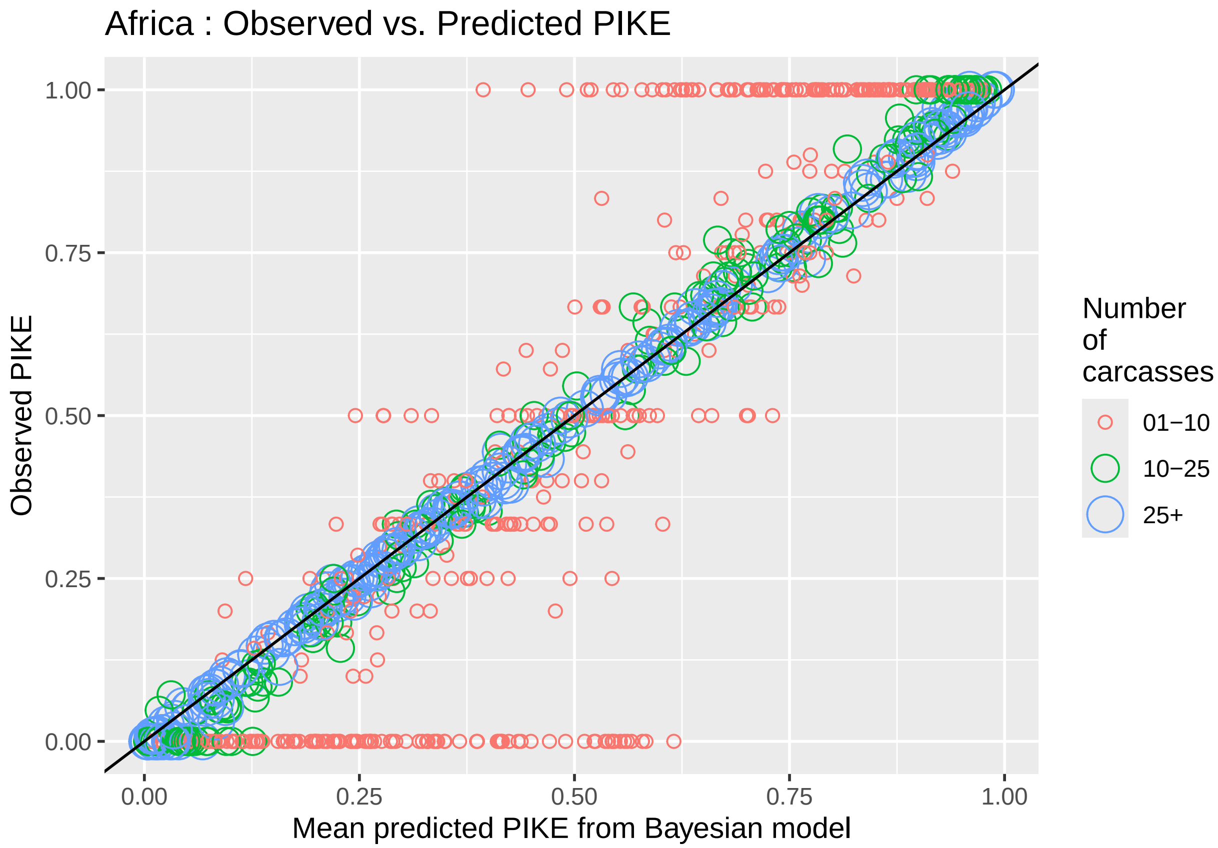
A plot of the $Year.Site$ effect for each site:



shows that only a few years had *PIKE* values within a site that could be considered unusual for that site.

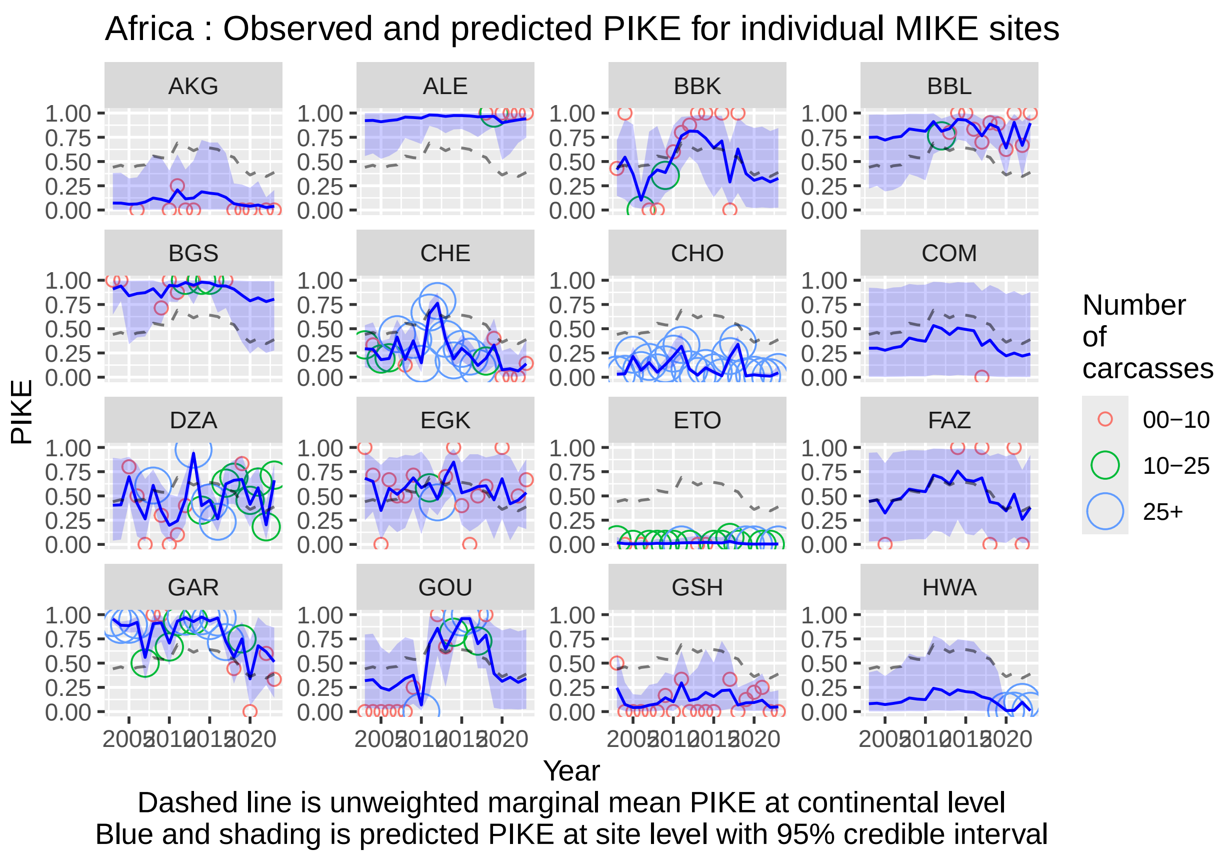
6.9 Observed vs. predicted *PIKE*

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



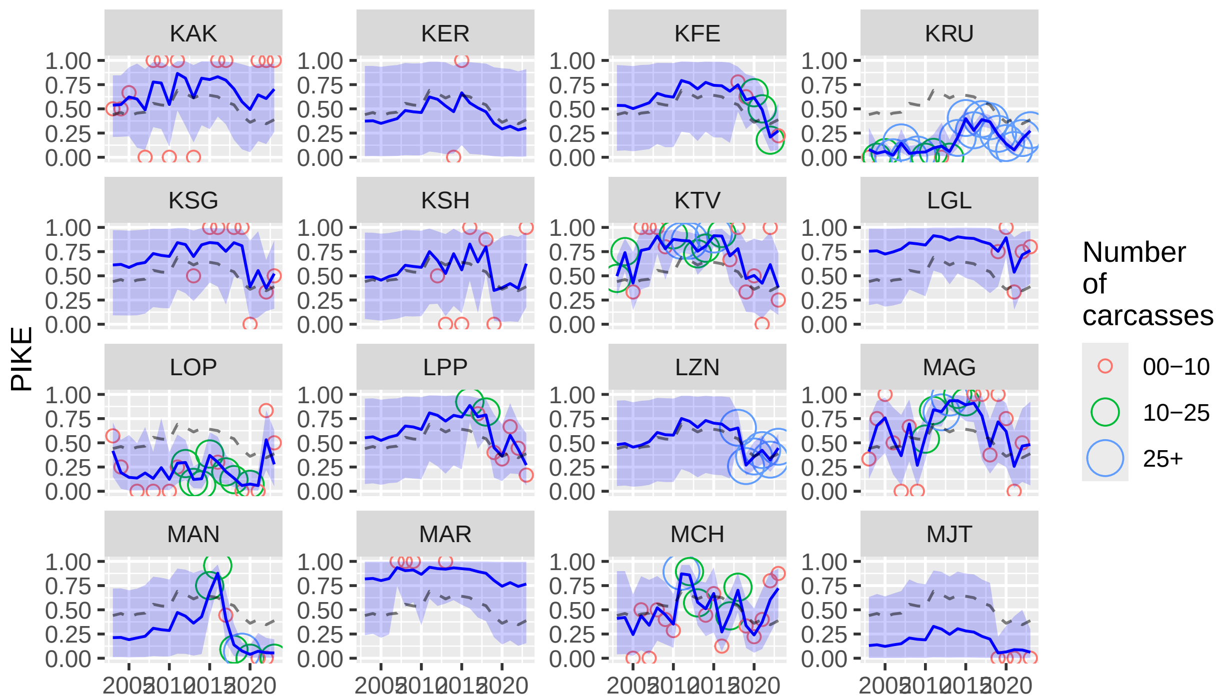
The fit is generally very good. For site-years where the number of carcasses was very small (< 10) and the observed *PIKE* was 0 or 1, the estimated *PIKE* is pulled towards the yearly average for that year. For site-years with large number of carcasses (> 25) the estimated *PIKE* matches closely with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



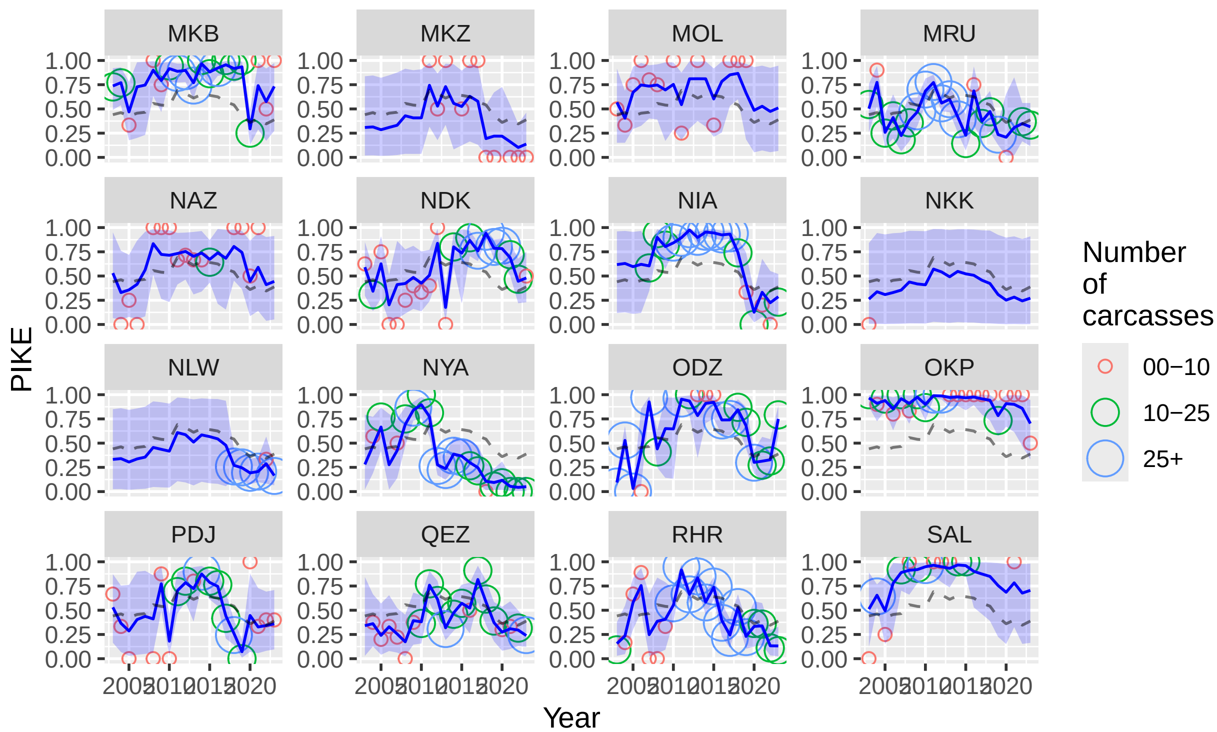
Africa : Observed and predicted PIKE for individual MIKE sites
AKG
ALE
BBK
BBL
BGS
CHE
CHO
COM
DZA
EGK
ETO
FAZ
GAR
GOU
GSH
HWA
1.00
0.75
0.50
0.25
0.00
PIKE
2005
2010
2015
2020
Year
Number of carcasses
00−10
10−25
25+
Dashed line is unweighted marginal mean PIKE at continental level
Blue and shading is predicted PIKE at site level with 95% credible interval

Africa : Observed and predicted PIKE for individual MIKE sites

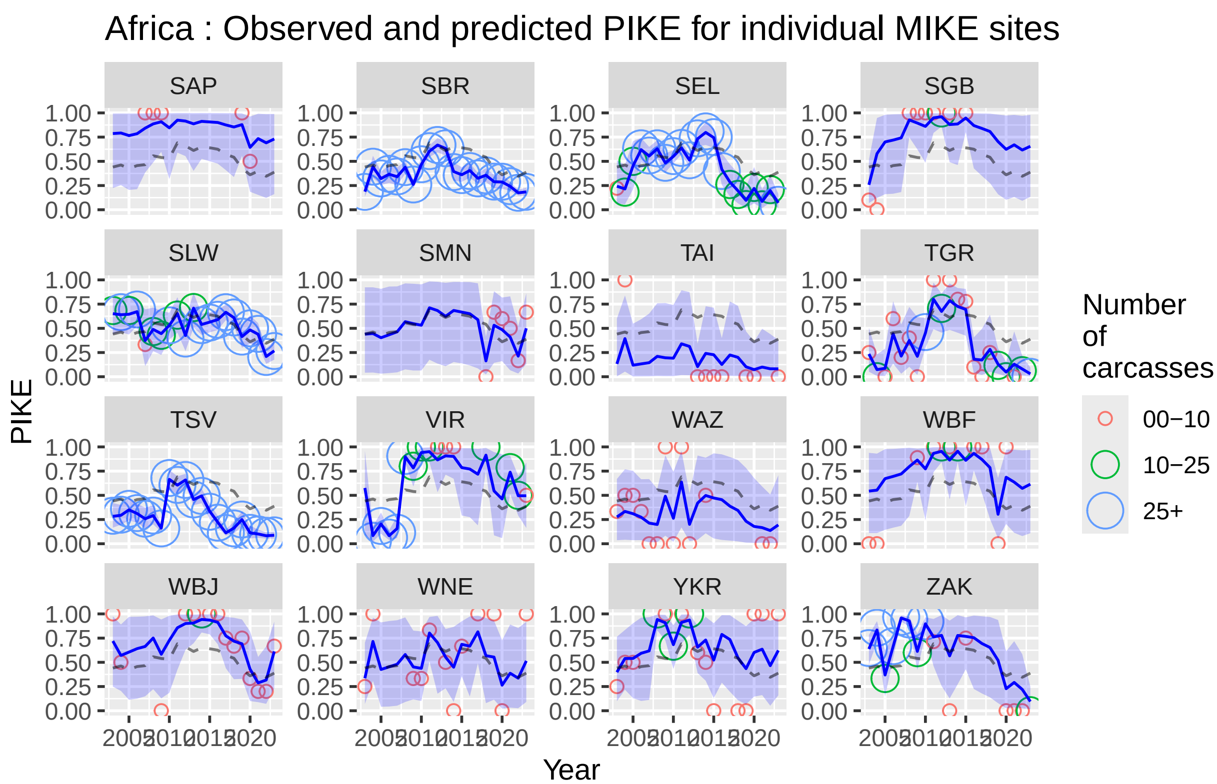


Dashed line is unweighted marginal mean PIKE at continental level
Blue and shading is predicted PIKE at site level with 95% credible interval

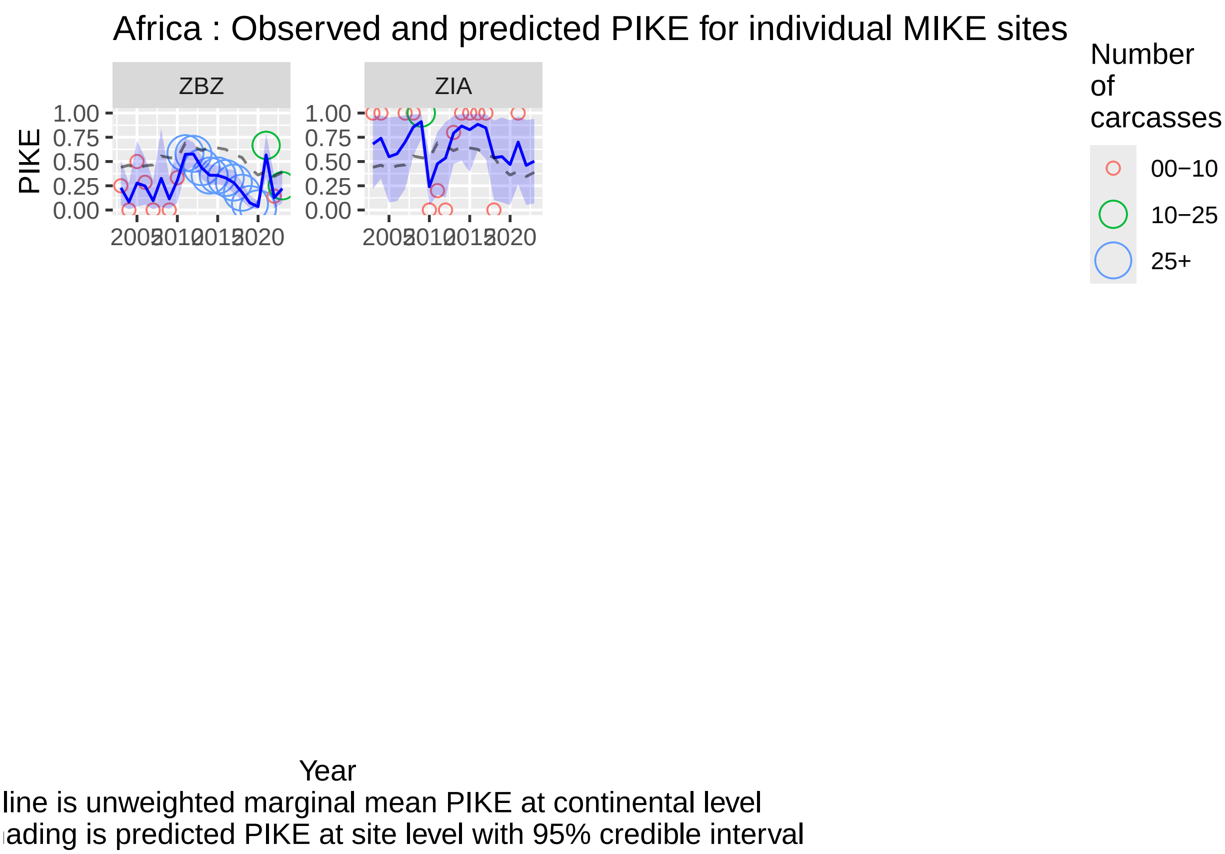
Africa : Observed and predicted PIKE for individual MIKE sites



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



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. 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.

7 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.

lsmeans (2019). pike TREND ANALYSIS USING THE LEAST-SQUARES MEANS APPROACH in R. https://github.com/CITES-MIKE/MIKE-LSMEANS

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.