Proposal for Monitoring Design and Analysis using the

Fish Sustainability Index

2017-03-08

Carl James Schwarz

StatMathComp Consulting by Schwarz

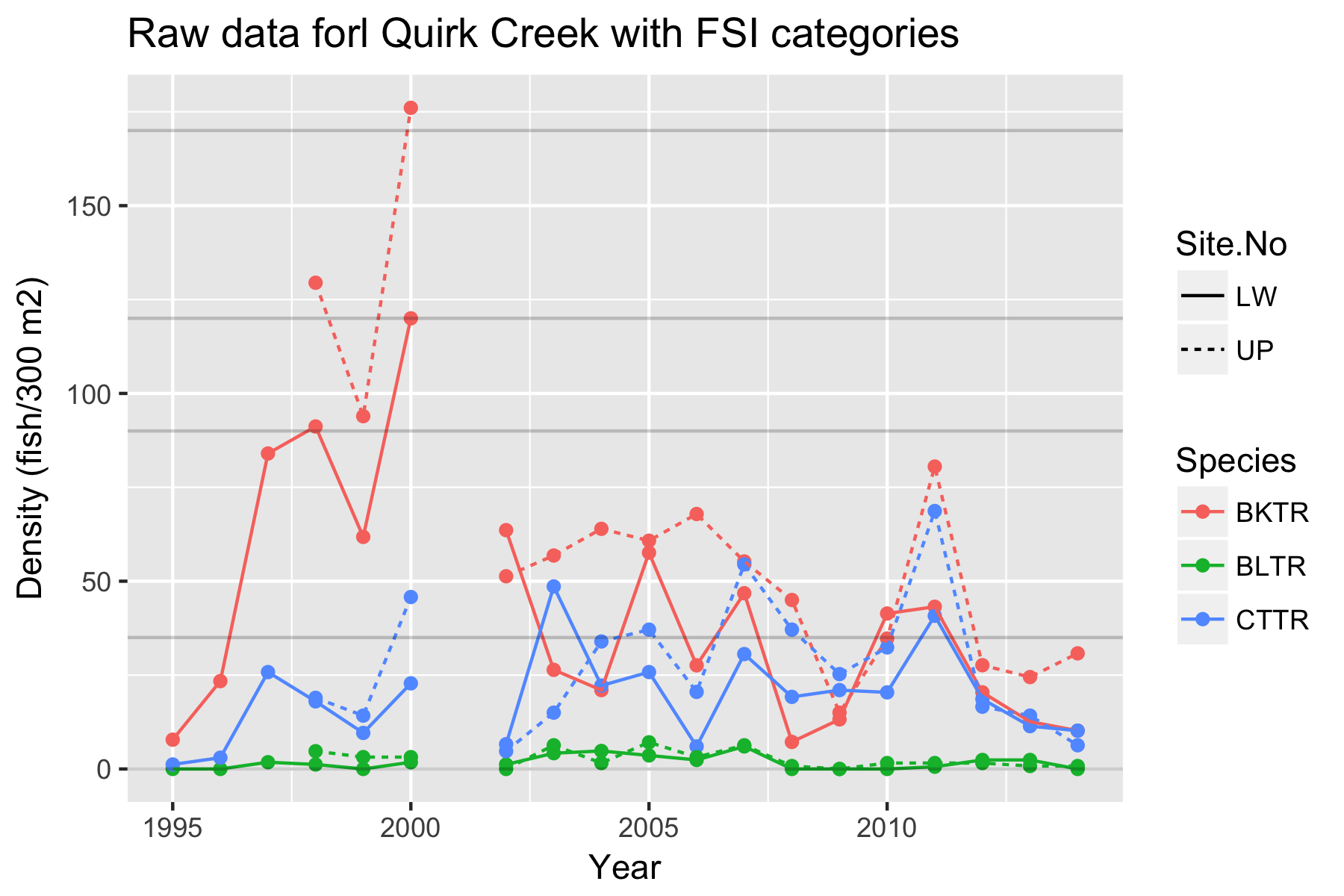
cschwarz.stat.sfu.ca@gmail.com

# 1. Introduction

The Fish Sustainability Index (FSI) is Alberta Fish and Wildlife’s method of assessing fish stocks on a provincial scale. One of the components of this index is Population Integrity which is partially assessed using population density. Density comparisons are referenced to what the watershed question could produce if it had no human impacts and consisted of the most ideal habitat in Alberta and is ranked on six point scale (0 to 5) from Functionally Extirpated to Low Risk. For example, draft guidelines for BLTR result in the following ranking based on observed catch per unit effort from one pass electrofishing.

|  |  |  |  |
| --- | --- | --- | --- |
| Table 1. Draft guidelines for FSI categorization for BLTR based on CPUE. Source “Draft FSI Thresholds for BLTR and MNHW.docx”. | | | |
| Theoretical FSI Category | Reduction Relative to Reference Condition (%) | CPUE (BLTR / 100m) | CPUE Range |
| 5 – Very Low Risk | 0 | 33.09 | > 33.1 |
| 4 – Low Risk | 30 | 23.163 | 23.2 - 33.1 |
| 3 – Moderate Risk | 50 | 16.545 | 16.5 - 23.2 |
| 2 – High Risk | 80 | 6.618 | 6.6 - 16.5 |
| 1 – Very High Risk | 100 | 0 | 0 - 6.6 |

Quantitative classification of a given population occurs by sampling a watershed at numerous sites (reaches) and observing the CPUE. However data can be very noisy. Consider, for example, a plot of CPUE for three species at Quirk Creek in Figure 1.

Figure 1a. Observed density in Quirk watershed measured at two sites (LW and UP). The horizontal lines correspond to the FSI category boundaries as extracted from a power point presentation. For simplicity it will be assumed that the categories were the same for all species.

In some situations, classification is straightforward. For example, if the (very) simplistic assumption is made that the FSI categories are identical for all species, then most readers would clearly the BLTR (green lines) in the VHR category in all years. However, what should be done with CTTR. In some years, some sites are above the threshold between VHR and HR, while in other years, both values are below the threshold. A single classification ignores the uncertainty in the data.

Similarly, consider Figure 2 extracted from results in the Macleod and Berland Watershed. While the mean CPUE all fall within the HR FSI category, the uncertainty in the mean CPUE is large enough that it may actually fall in any of the three categories.

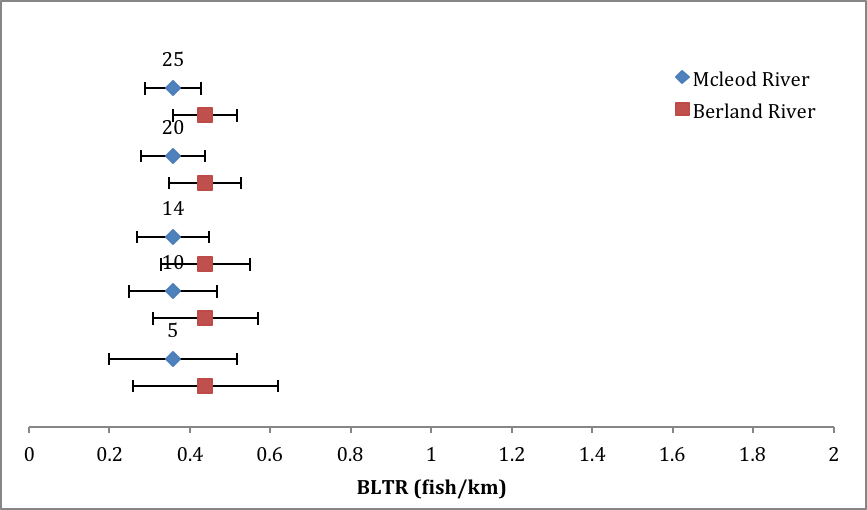


Figure 1b. Mean density from the Macleod and Berland Watershed with error bars representing ±1 SE of the mean and vertical dashes assumed to be FSI boundaries. Extracted from the provided “McLeod 2006 and BLTR data.xlsx” workbook.

Rather than trying to arbitrarily choose a single category and ignore uncertainty, it would be useful to view the assignment as a probabilistic task. For example, we would like to make statements such as “We are 80% certain that this watershed is in the HR category”.

This gives rise to a number of issues that need to be resolved:

1. How can such an assessment be done?
2. How many sites must be measured so that you are xx% certain that each watershed is in its appropriate risk category.
3. How can these assessments be tracked over time either through long-term trends or running averages
4. How many years of sampling and how many sites within each watershed are needed to be xx% certain of detecting a trend

I will used the Quirk Creek watershed as an example of how such questions may be approached.

# 2. Illustrative Example.

This section will serve as an illustrative example. It is not to be construed as a definitive analysis.

Data was collected on CPUE (fish/300m2)[[1]](#footnote-1) from 1995 to 2014 (except for a break in 2001) at two sites in the watershed (LW and UW) for three species (BKTR, CTTR, and BLTR).

FSI thresholds were read off a graph provided in a power point presentation and are:

FSI.cat,lower, upper

1, VHR, 0, 35

2, HR, 35, 90

3, MR, 90, 120

4, LR, 120, 170

5, VLR, 170, 3000

The same boundaries will be used for ALL species (even though this is unrealistic). Nothing in the example below precludes a separate set of FSI thresholds for each species.

A plot of the observed densities was shown in Figure 1 (in the introduction). There is a variation between the sites, and some consistency in year effects. For example, in 2011 most densities appear to be elevated.

A plot of the standard deviation (between the sites in a year) vs. the mean of the two sites in a year shows a pattern where the standard deviation increases with the mean. This is quite common when measuring abundance and indicates that a log-normal distribution will be good description of the distribution of measured densities among sites within a year. One of the properties of the log-normal distribution is that large outliers are quite common. To account for the impact of these large outliers on the mean across the sites in a year, the MEDIAN is preferred measure of overall trend and of classification to the FSI category. The median is the point where 50% of the values (measurements at sites) are predicted to be above and 50% of the values are predicted to be below. [With only two sites measured per year, the median is not well estimated!]. If the underlying distribution of readings follows a log-normal distribution, then the median can be estimated by the geometric mean of the raw observations.

## 2.1 Estimating the probability that the watershed is in each FSI category based on individual years.

From Figure 1, we see that the individual density measurements for BLTR are all close to zero and always within the first FSI category. The measurements for CTTR are more variable with the measurements at UP tending to be higher than the measurements at LW, but the median density between the two sites is mostly in the bottom FSI category except perhaps in 2011.

It seems natural to ask – what is the “probability that the median density is in each category”. This cannot be answered with classical statistic (mean and confidence intervals) because the actual category membership is fixed (non probabilistic) in any one year[[2]](#footnote-2). A Bayesian analysis comes to the rescue. A Bayesian would change the question slightly to “what is the belief that the median density is in each category”. The change is crucial, because belief can be expressed in a probabilistic framework.

An intuitive explanation for the process is as follows. For each year, estimate the parameters that describe that year’s distribution of density across the sites. If a log-normal distribution is assumed, the parameters are the log(median density) and the standard deviation on the log-scale across ALL possible sites. These values are estimated based (on this case) the two observed sites. Estimate the sampling distribution for the log(median) for ALL sites based on these values. A sampling distribution gives the distribution of plausible values for the log(median) for ALL sites. [A confidence interval can be computed from this sampling distribution but is not used.]. A Bayesian then says that the sampling distribution is interpreted as your belief in the distribution of the log(median) for all sites. Consequently, compute what fraction of the sampling distribution lies between the FSI boundaries and that is your belief (probability) that the median density is in each category.

The actually fitting process cannot be done by hand and requires a method called MCMC to do the Bayesian analysis. As part of the output from an MCMC analysis, are quantities such as the probability (belief) that the median is in each category.

A program was written in BUGS – a standard language for Bayesian analysis and run using JAGS (a program to do Bayesian analysis) from R (an open source statistical software package). All of these programs are available free of charge and can run on a standard desktop computer in a few seconds.

The actual analysis needs to account for several sources of variation

(a) Site-specific effects. Notice in Figure 1, that the counts for site UP tend to larger than the counts for LW. This is due to site specific factors that are currently unknown and are called random effects. If the site-specific factors are thought to be related to covariates of the site (e.g. water temperature), these can be included in the analysis.

(b) Year-specific factors Notice that in 2011, most counts took an upward tick. Again, the cause of these year-specific factors is unknown but if covariates are available (e.g. total precipitation), these can be included. We again assume that year-specific factors are random effects.

(c) Noise (or residual variation). Notice that while the site UP tends to have higher densities than the site LW across years, that the difference is not consistent. We also assume that residual variation is random.

A model was fit allowing for all of these sources of variation and the results are shown in Figure 2.

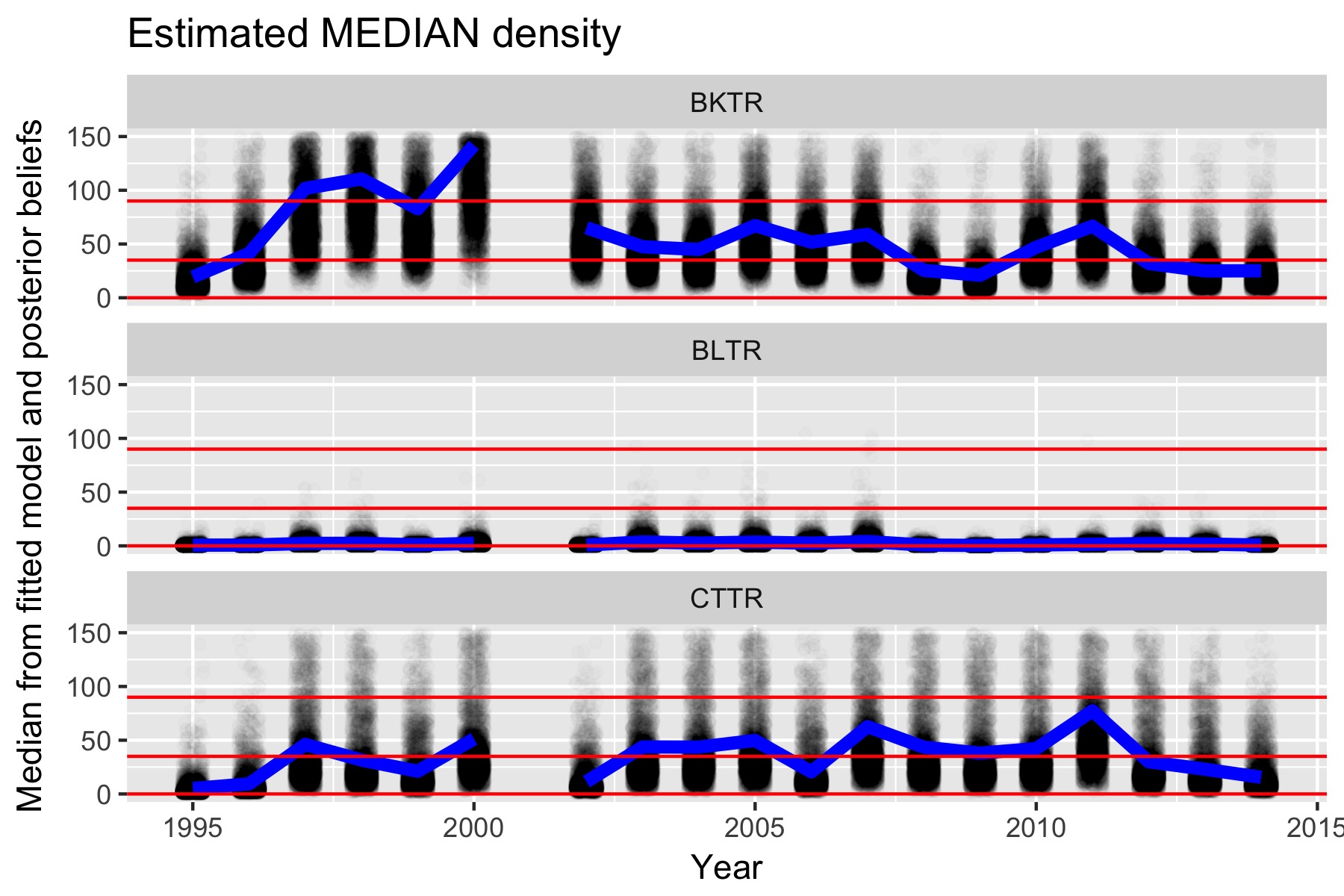


Figure 2. Estimated posterior (after the Bayesian analysis) plot of the fitted median (blue line) and belief about possible values (black dots) for each species in each year.

Consider species BLTR. The fitted line (blue) never ventures outside of the lowest FSI category, but given the small sample sizes, it is very slightly possible that the median may fall in the next category (shown by a faint smudge in 2007).

Consider species CTTR. The fitted median is mostly above the cutoff for the first category, but the black smudges indicate a high belief that it could still b in the lowest category. If you look at Figure 1, you see that the two sites straddle the boundary between the first two FSI categories. In particular, in 2009 both sites had readings below the first threshold, but the fitted line (Figure 2) is above the threshold. What happens is that the model looks at the relationship between the two sites over all of the years. The gap between the two sites is often larger than observed in 2009. Consequently, the model believe that the closeness of the two sites in 2009 is due to random change and takes that into account when fitting the model.

This sharing of information is a prime reason why Bayesian models work well in these context when there are small sample sizes in all years.

The probability that the median is in each FSI category is shown in Figure 3.

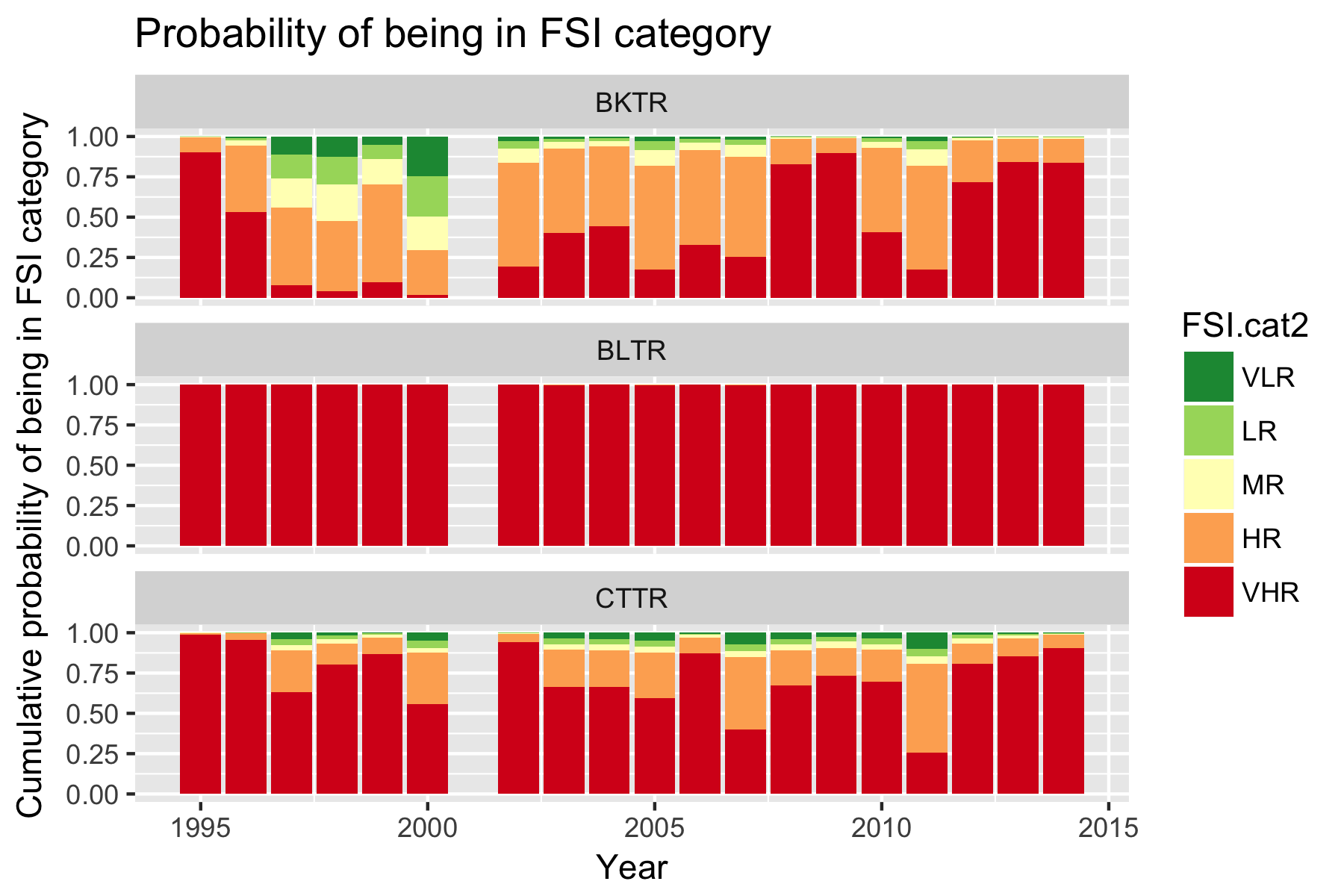


Figure 3. Posterior beliefs about belonging to each FSI category.

For example, there is virtually a 100% belief that Species BLTR belongs to the VHR category in all years. For species CTTR, there is a high probability in most years, but in 2011, the counts were high enough that there was shift towards the second category. This a small belief that the median could be in the VLR category as a consequence of the very small sample sizes in each year that allow for the possibility that the observed data were so low in each year just by chance.

## 2.2 How many sites are needed to be xx% certain that the correct category is chosen?

The required sample size to be xx% certain of assigning a watershed to the correct category in a particular year depends on three sources of variations:

1. the site-to-site variation. This is reflected by the fact that the densities in any one year are not identical at the two sites. As seen from Figure 1, some sites tend to have higher densities more or less consistently due to site-specific factors. By taking more sites, you can average over these site-specific factors.
2. Residual variation. When you take measurements at a particular site, that measurement is subject to error. You can reduce this error by sampling larger areas of the site.
3. Year-to-year variation (process error). Again referring back to Figure 1. If you are willing to assume that the system is in a “steady state”, the values seem to fluctuate up and down in a consistent pattern among sites within a year. For example, in 2011, the densities seem to be increased consistently across all sites due to factors that are year-specific. For example in that year, environmental conditions were such that the CPUE was elevated consistently.

The probability also depends on how far the underlying mean is from the boundary. For example, from Figure 3, you are very certain that BLTR is in the VHR category even with only two sites sampled. However, the uncertainty for BKTR is much higher because measurements for this site are bouncing around the boundary with site-to-site, year-to-year and residual variation all large.

If the year-to-year variation (process error) is large, it will very difficult (and perhaps impossible) to achieve the target certainty of assignment, regardless of how many sites you sample. Even with perfect information (measuring an infinite number of sites), the temporal variation cannot be predicted in advance its distribution may be so wide that you cannot achieve your goal.

For example, suppose, for simplicity, that we assume that all of the species are in steady state. We obtain the following information about the median density and variance components (on the log-scale):

Species med.log med.regular SDresid SDsite SDyear

1 BKTR 3.71 40.80 0.43 0.80 0.70

2 BLTR -0.16 0.85 1.41 0.88 1.05

3 CTTR 3.00 20.06 0.45 0.73 0.86

The boundary between the VHR and HR categories is around 30, so let us assume that species BLTR and CTTR belong to HVR category, and BKTR to the HR category with their respective median densities.

We see that the year-to-year variation (on the log-scale) is comparable to larger than the site-to-site variation and so for points near the boundary of a category, it may be difficult to reach the goal.

Figure 4 shows how the probability of correct classification varies by the number of sites measured each year. There is virtually no impact of measuring an excessive number of sites because year-to-year variation is so large (and not controllable) that you cannot be certain that a year-specific effect will push the median into the next category. Refer to Figure 3. We see that for CTTR, that the highest probability that it is in the VHR level occurred on 17/19 years.

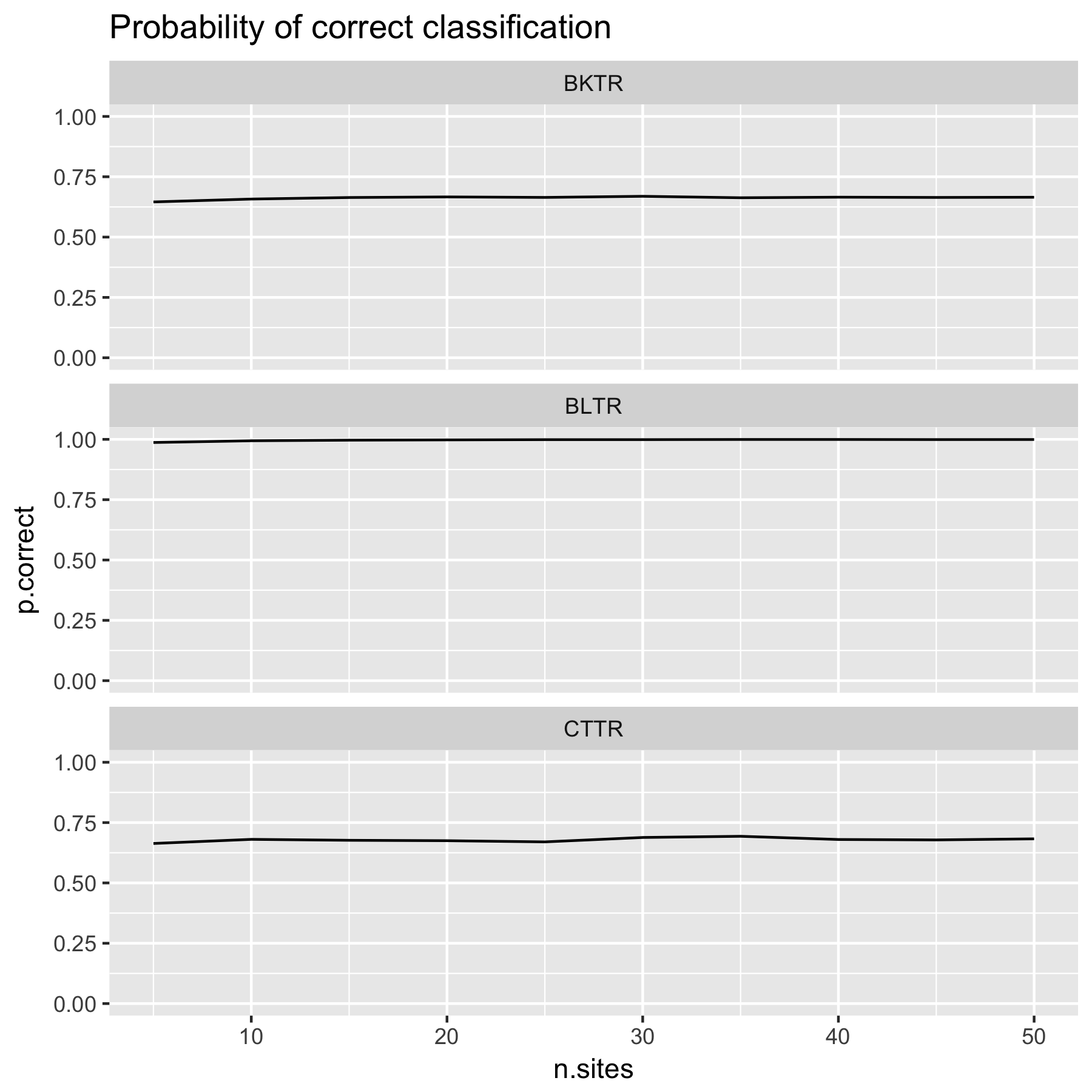


Figure 4. Probability of a correct classification (assuming that the system is currently in a steady state) as a function of the number of sites measured each year.

# 2.3 Smoothing the above results.

With very small sample sizes, the results can be quite variable from year to year as seen above. It is possible to “smooth” the data. I will demonstrate how to fit a linear trend line, but you could also easily fit a five-year running average with similar results.

When every fitting trends over time, you need to be concerned about the year-specific effects (also known as process error). Year-specific effect force the point in year above or below the trend line en masse, which is a violation of the key assumption of a regression analysis. For example, regression analysis assumes that the data are always centered about the regression line as shown in Figure 5.

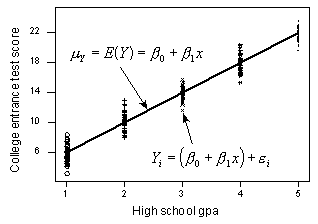


Figure 5. The standard assumption of regression analysis where at each X value, the points are randomly scattered about the trend line.

However, year specific-effects can push the set of points in a year higher or lower around the trend line as shown in Figure 6:

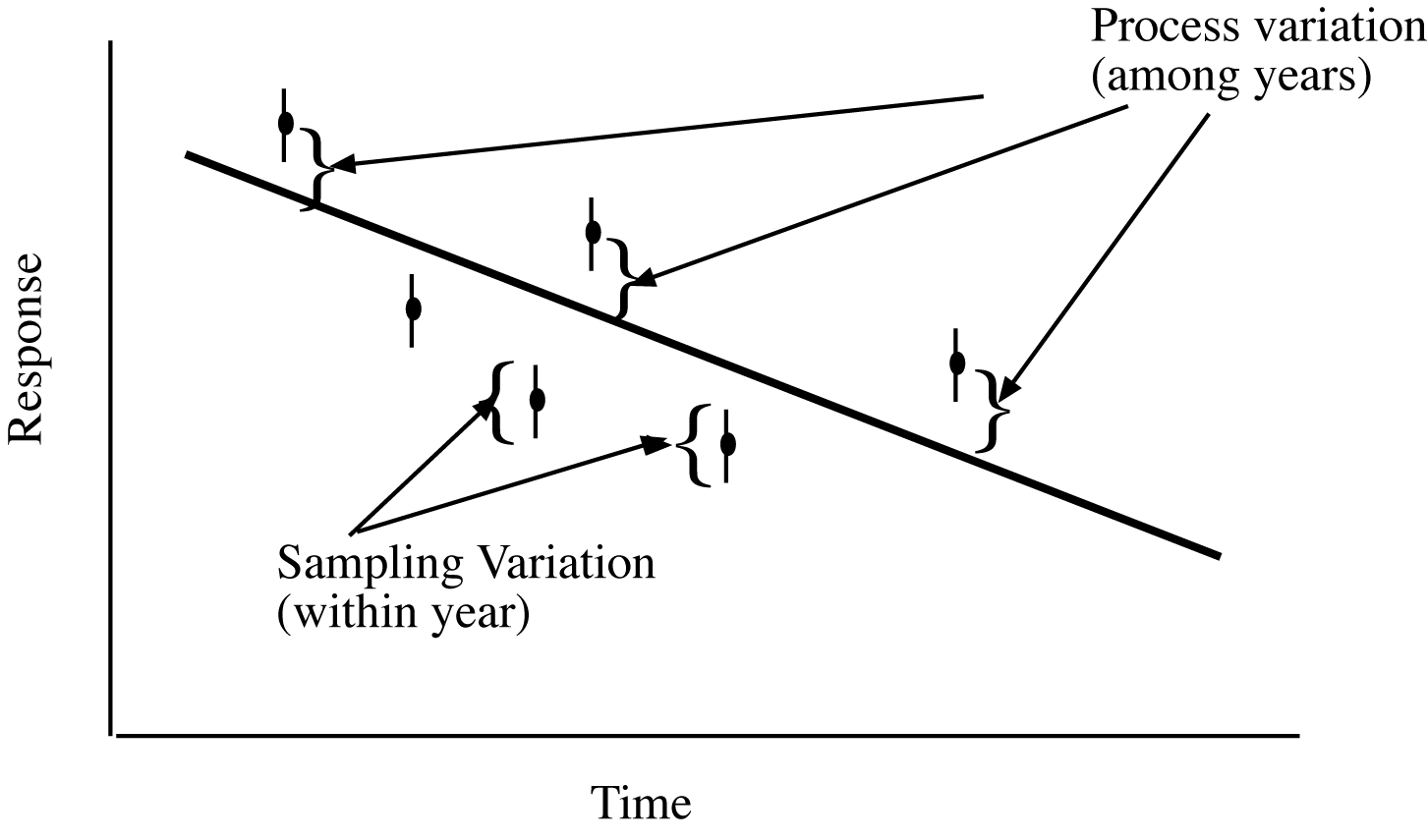


Figure 6. Process error will tend to push all points above or below the underlying trend line.

There is evidence of year-specific effects in this dataset. For example in Figure 1, all data points tended to be pushed up in 2011. The impact of year-specific effects (process error) are two fold. First, the individual data points become less and less important in determining the trend – only the variation in the AVERGE about the trend line is important. More importantly, to detect trends, the number of years of sampling then becomes the limiting factor as illustrated in Figure 7.

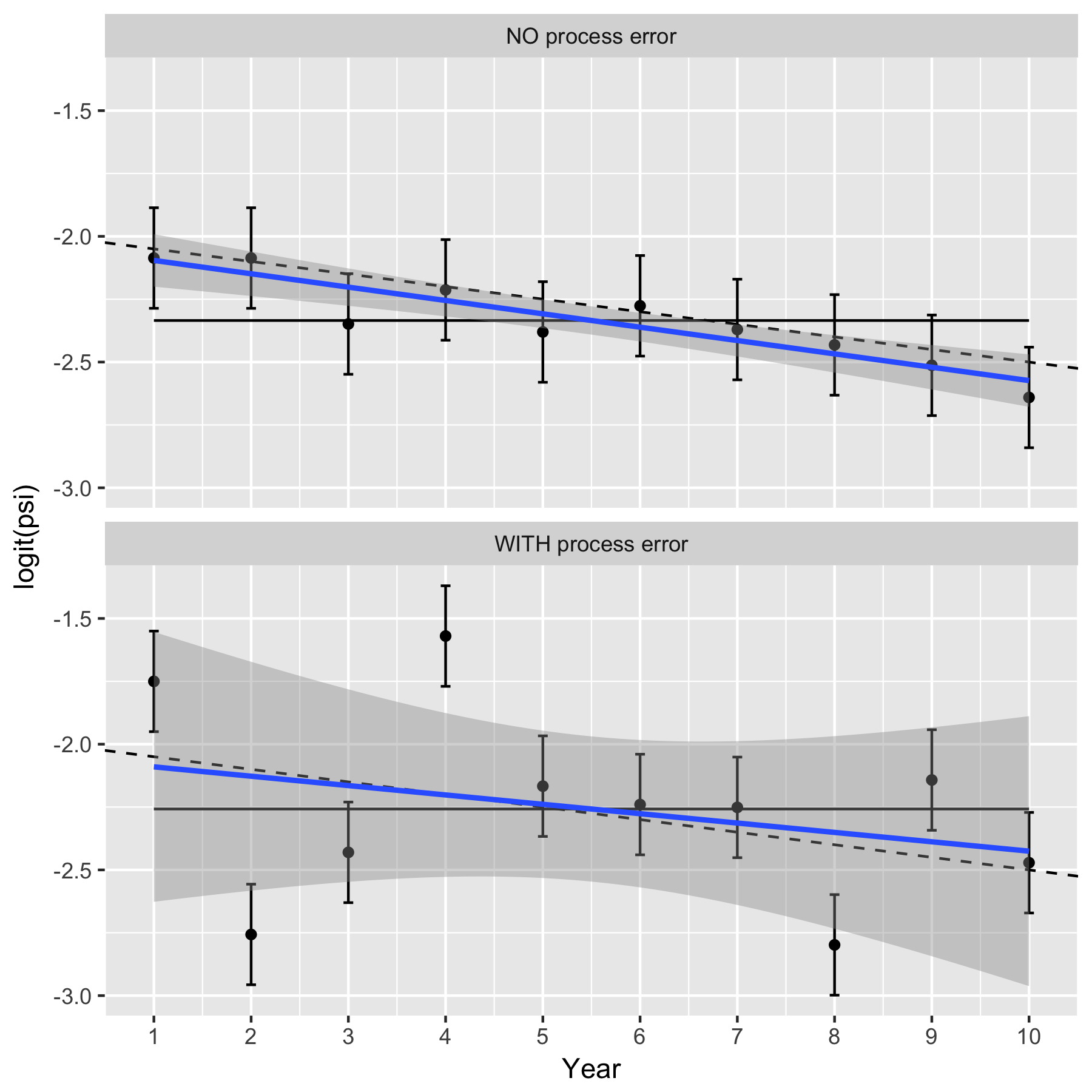


Figure 7. Illustrating the impact of year-specific factors on estimating a trend.

In the top panel, there are no year-specific effects (no process error) and the response changes over time based on the underlying trend line shown as a dashed line. Only sampling variation is present so 95% of the confidence intervals about the observed mean response each year will overlap with the underlying trend line. The fitted line will be close to the true underlying trend line (solid line). The uncertainty is small about the overall trend, and there is clear evidence that the trend is different from a 0 trend (which is shown by the horizontal line). In the bottom panel, the year-specific effects (process error) add extra variation to the underlying response each year due to effects such as weather, food etc. Now, the 95% confidence intervals for mean response still provide valid estimates for the yearly mean response values, but now may not overlap the true underling trend (in dashed). The fitted line will still be unbiased for the true trend (solid line), but the extra variation makes the uncertainty in the fitted line much larger and now there is no evidence that the trend line differs from 0.

If there is substantial year-specific effects (process error), then sampling more sites in a year will NOT be helpful. Sampling more sites in a year will shrink the size of the confidence intervals in the bottom panel above, but has NO impact on the process error and so the variation around the fitted line will only be reduced slightly. In cases of substantial process error, the limiting factor for detecting trends is likely to be the total years of sampling and not the number of sites/year that are sampled.

A second model was fit that allowed for a linear trend over time to smooth the results from the previous section. Process error was allowed for. The results are shown in Figure 8.

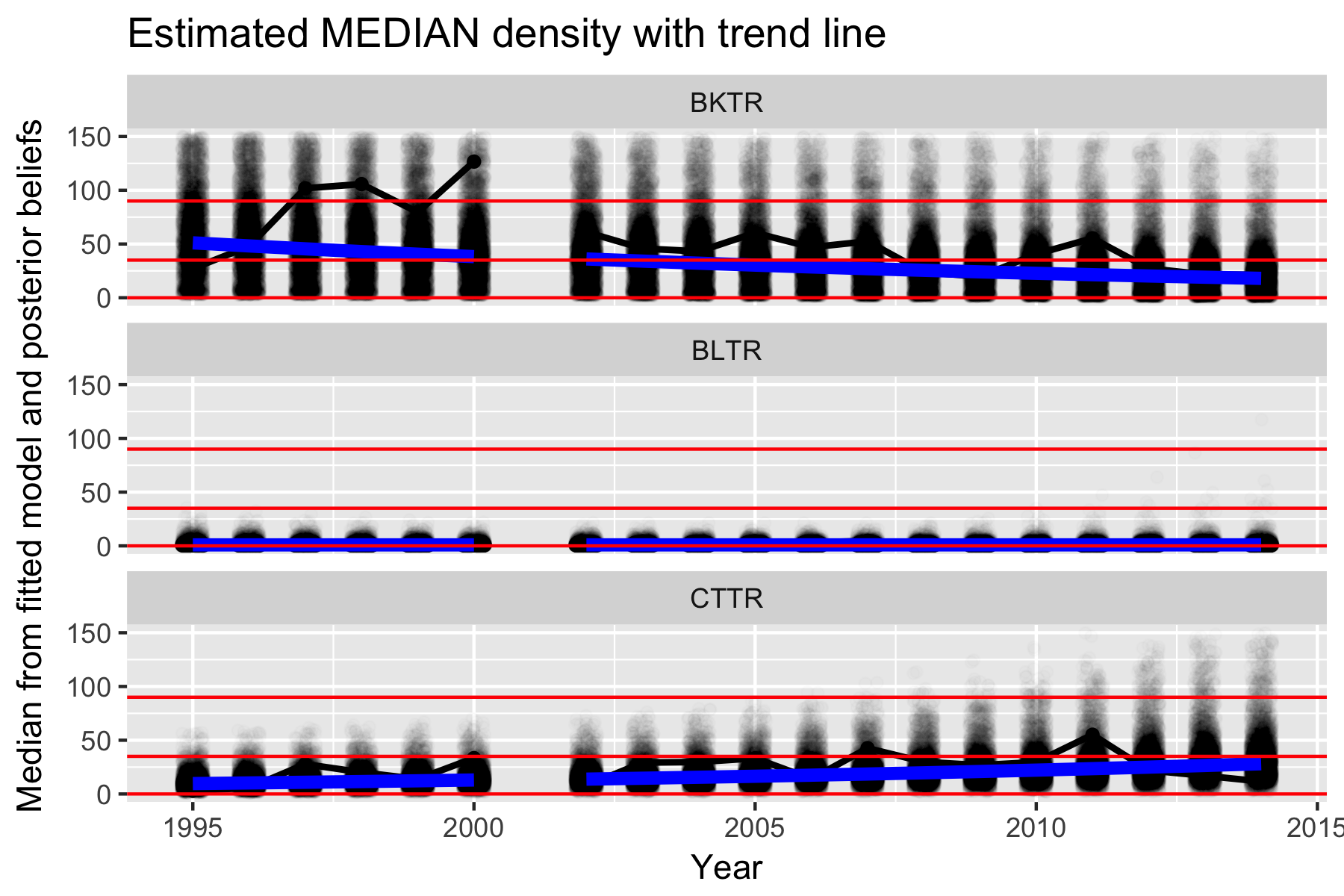


Figure 8. Posterior plot after model fitting with linear trend. The blue line is the fitted linear trend. The back line is the observed median with process error. The black dots ae the beliefs that the median on the underlying trend lies in the various FSI categories.

Here the trend line “smooths” the individual medians (the black line) and so now the beliefs in the various FSI categories (see next plot) is much smoother and less subject to random year-specific effects. We now see that the population status may be changing slowly. Notice that the log-normal distribution treats the large values for species BKTR in the late 1990s as mostly outliers. While I fit a linear trend here, you could also fit a running 5-year average or similar non-parameteric curves.

The beliefs that the system is in each category is shown in Figure 9.

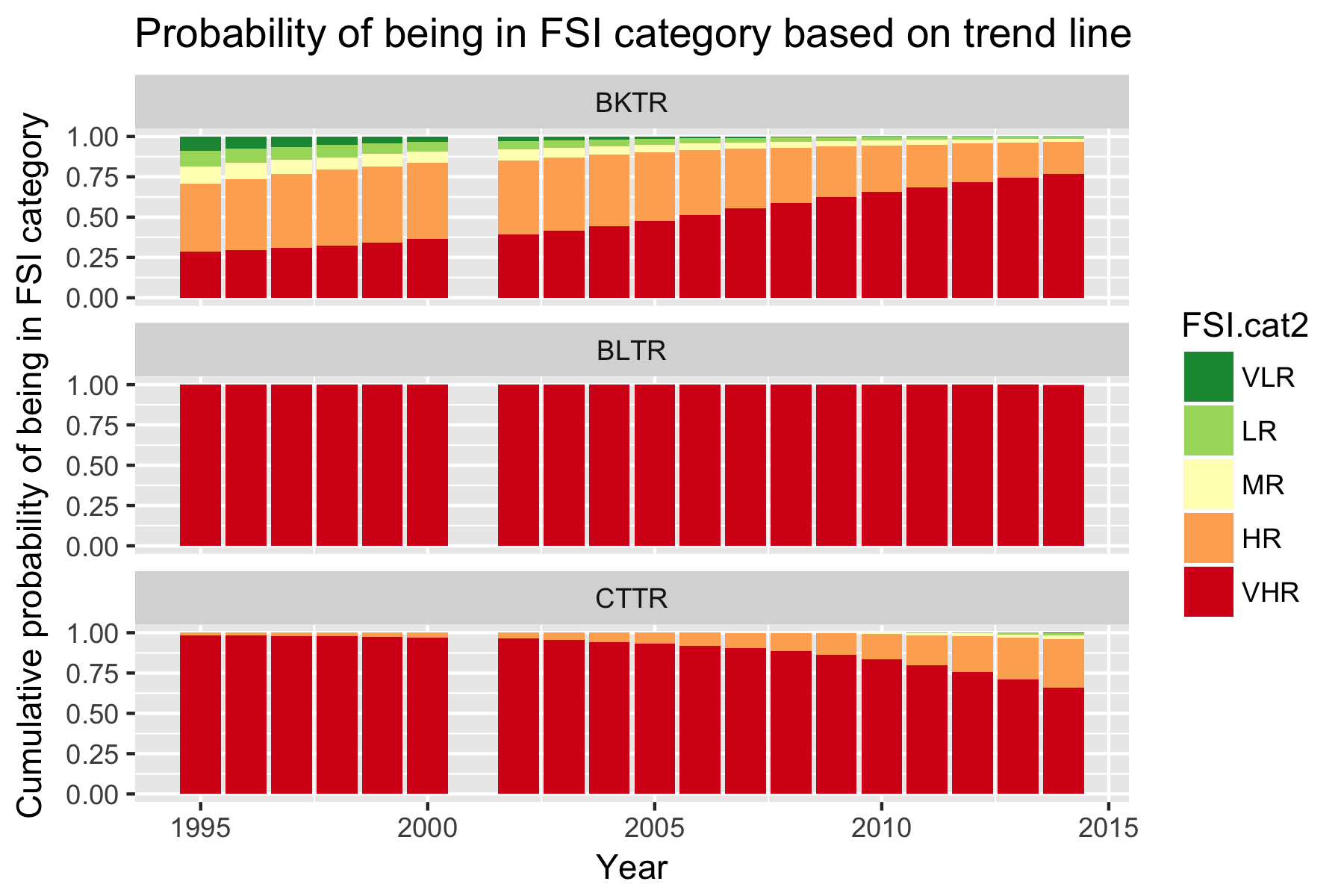


Figure 9. Posterior probability of belonging to each FSI category after a trendline is fit.

From Figure 9, you can see that the posterior probability of belonging to VHR is increasing for species BKTR and decreasing for species CTTR as expected given the trend line in Figure 7.

The estimated trend line (on the logarithmic scale) is shown in the following table.

Species.num Species slope sd p.slope.lt.0

1 1 BKTR -0.06 0.03 0.96

2 2 BLTR 0.00 0.07 0.49

3 3 CTTR 0.06 0.04 0.05

The interpretation of the slope and standard deviation are similar to classical regression. For example for species BKTR, the estimated slope (on the log-scale) is (negative) 0.06.(SD .04). Because you analyzed data on the logarithmic scale, this corresponds to an estimated 6% decline/year. The interpretation of the last column is different from the traditional p-value. In a Bayesian context, it is your belief that the trend line is negative. In this case, you have 96% belief that the trend line is negative.

Similarly for species CTTR, the estimated change is an increase of 6%/year and you have a 5% (95%) belief that the slope is negative (positive), respectively.

For species BLTR, the estimated slope is close of 0, and the belief that the slope if negative (positive) is 49% (51%) respectively. There is no evidence of a change in median density over time. Figure 9 shows that this population has a high posterior probability of being in the VHR category.

## 2.4 Sample size requirements to detect trends

The trend model provides estimates of the site effects standard deviation (the site-to-site variation), the year effects standard deviation (the year-to-year specific factor variation, a.k.a. the process error), and the residual standard deviation (noise). These are needed to determine how many sites and years are needed to be sampled to detect a specific trend. This in turn can be used to estimate how many years are needed to increase the probability of moving from one category to another category (not demonstrated here).

The estimated standard deviations are:

Species.num SDresid SDsite SDyear Species

1 1 0.4422217 0.9831795 0.5958415 BKTR

2 2 1.4248202 0.9117626 1.1082715 BLTR

3 3 0.4456816 0.6423131 0.8153628 CTTR

We notice that for all species, the year-effect standard deviation (SDyear column) is the same order of magnitude as the residual standard deviation column (SDresid). This implies that we are in the situation of Figure 6 where year-to-year specific effects push the observed values up or down from the trend line en masse and the limiting factor is the number of years of sampling and not the number of sites sampled in each year.

Figure 10 illustrates power computations to detect a 5% (increase/decrease) per year using the variance components from the above table with different number of sites sampled per year.

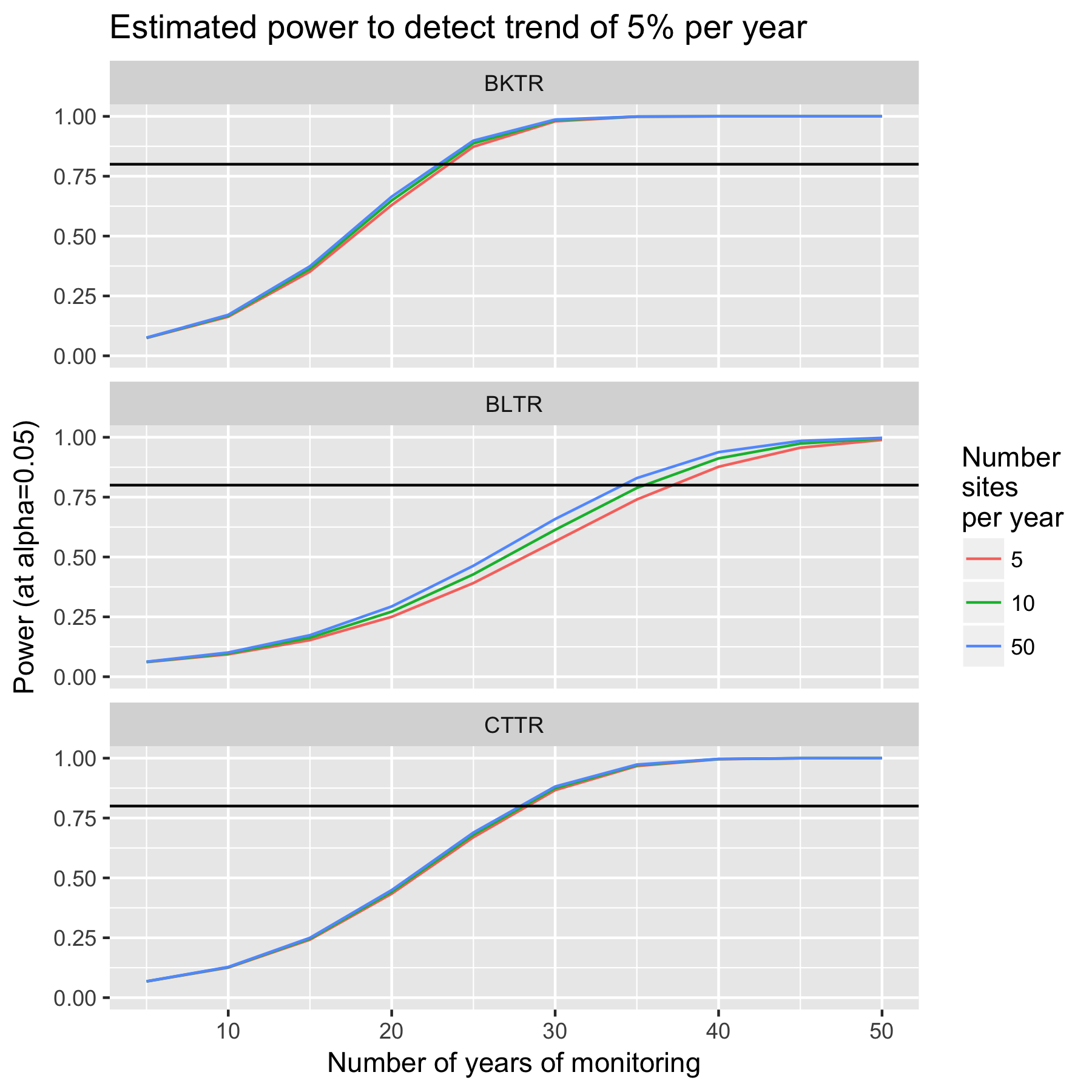


Figure 10. Estimated power to detect a 5% trend with different number of years and sites sampled per year.

The power essentially depends only on the number of years sampling and not the number of sites/years. The key issue is again the size of the process error (year-specific effects). This variation CANNOT be controlled by experimenter as it typically results from external forces such as weather. It would be possible to adjust for these effects if suitable covariates were available, but some caution is needed. For example, suppose that process error is related to climate change and closely related to mean winter temperature. If you include mean winter temperature in the model, you essentially have removed the effect of climate change (!) which likely is not what you want.

Given that the number of years of monitoring is the prime drive for power, it is possible to adjust the sampling plan by monitoring sites every 2nd year. For example, one set of sites is measured on years 1, 3, 5, 7 and the other set of sites is measured on years 2, 4, 6, 8. The power is not greatly impacted.

This power analysis can then be embedded into the FSI framework by projecting forward in time and computing the posterior probabilities of belonging to the category. This has not been illustrated in this brief report.

Similarly, the trend lines can be projected forward to predict the probability of being in each FSI category in future years, e.g. 2021. This has not been illustrated in this brief report.

In this particular example, the very small sample sizes (i.e. only 2 sites measured each year) has a great impact on the analysis, but Bayesian methods are able to compensate by “sharing information” across years. There appears to be very large year-specific effects which limit the ability to detect trends over short time horizons.

## 3.0 Proposal

This proposal is for the development and implementation of the above methodology for fish density information on watershed in Alberta.

Deliverables will be:

1. R code and documentation on the methodology. It will made available on an open source platform (e.g. GitHub) so that others can use, critique, and update as needed.
2. Analysis of data from watersheds in Alberta (assuming that the data is available in electronic format). An automated system (RMarkdown) will be used so that reports are generated automatically for each watershed and can be updated each year by simply updating the database and re-running the code. This will include estimates of the probability of being in each FSI category for each year (comparable to Section 2.1 of the report); sample size analysis (comparable to Section 2.2 of the report); trend analysis using regression methods and spline (similar to running averages) with projects from the fitted trend to 5, and 10 years in the future (Similar to section 2.3 of this report); sample size determination for the number of years of monitoring, frequency of monitoring, and number of sites to monitor to detect trends of specified magnitude (Similar to Section 2.4 of this report).  
     
   These documents can made web ready for easy posting.

Timeline: Project completion 30 June 2017 (or sooner)

Costing:

$15,000 – fixed cost price + GST.

If travel is required, will be charged at cost.

Data will be provided in electronic format. A number of formats is acceptable (e.g. Excel, Access, Other database systems).

Price valid until 2017-04-30.

1. Actual data as per 100m2, but FSI thresholds were in per 300m2 so all the values in the dataset were multiplied by 3. [↑](#footnote-ref-1)
2. This is analogous to asking “What is the probability that a woman is pregnant.” The pregnancy status is a fixed, non probabilistic value. However, what is your belief that a woman is pregnant is a sensible statement as a probability. [↑](#footnote-ref-2)