IDA

I N S T I T U T E   F O R   D E F E N S E   A N A L Y S E S

Bayesian Analysis in R/STAN

Kassandra Fronczyk

INSTITUTE FOR DEFENSE ANALYSES
4850 Mark Center Drive
Alexandria, Virginia 22311-1882

IDA

The Institute for Defense Analyses is a non-profit corporation that operates three federally funded research and development centers to provide objective analyses of national security issues, particularly those requiring scientific and technical expertise, and conduct related research on other national challenges.

About This Publication

In an era of reduced budgets and limited testing, verifying that requirements have been met in a single test period can be challenging, particularly using traditional analysis methods that ignore all available information. The Bayesian paradigm is tailor made for these situations, allowing for the combination of multiple sources of data and resulting in more robust inference and uncertainty quantification. Consequently, Bayesian analyses are becoming increasingly popular in T&E. This tutorial briefly introduces the basic concepts of Bayesian Statistics, with implementation details illustrated in R through two case studies: reliability for the Core Mission functional area of the Littoral Combat Ship (LCS) and performance curves for a chemical detector in the Bio-chemical Detection System (BDS) with different agents and matrices. Examples are also presented using STAN, a high-performance open-source software for Bayesian inference on multi-level models.

INSTITUTE FOR DEFENSE ANALYSES

IDA Document NS D-5831

Bayesian Analysis in R/STAN

Kassandra Fronczyk

Executive Summary

In an era of reduced budgets and limited testing, verifying that requirements have been met in a single test period can be challenging, particularly using traditional analysis methods that ignore all available information. The Bayesian paradigm is tailor made for these situations, allowing for the combination of multiple sources of data and resulting in more robust inference and uncertainty quantification. This tutorial is meant to introduce the basic concepts of Bayesian analysis. Further illustration of the flexibility and applicability of these methods is shown with examples from the T&E community, including implementation details. The course consists of four sections:

- Fundamentals of Bayesian Analysis: This section provides the basic concepts common to all Bayesian analyses, including the specifications of prior distributions, likelihood functions, and posterior distributions. A simple example is used for demonstrative purposes, including a short sensitivity study.

- Case study: Littoral Combat Ship (LCS). This section extends the simple component/subsystem level analysis shown in the first section and extends it to a system reliability calculation. Here, LCS has a functional area reliability requirement, but data were collected on various subsystems of different data types. Implementation details are given in code snippets.

- Case Study: Bio-chemical Detection System (BDS). This example takes the simple modeling case from the introduction and introduces factors of interest. A logistic regression model is fit to determine the probability of detection across different concentration levels of various chemical or biological agent/matrix concentrations. Implementation details are provided, along with a list of R packages that can aid in the fitting of these types of models.

i

Bayesian Data Analysis in R/STAN

Dr. Kassie Fronczyk

Institute for Defense Analyses

Dr. James Brownlow

US Air Force

IDA

Outline

IDA

- **Fundamentals of Bayesian Analysis**

- **Case study: Littoral Combat Ship (LCS)**

- **Case Study: Bio-chemical Detection System (BDS)**

- **STAN Examples**

Acceptance of Bayesian Techniques

Bayesian methods are commonly used and becoming more widely accepted

- **Applications**
  - FAA/USAF in estimating probability of success of launch vehicles
  - Delphi Automotive for new fuel injection systems
  - Science-based Stockpile Stewardship program at LANL for nuclear warheads
  - Army for estimating reliability of new aircraft systems
  - FDA for approving new medical devices

- **Recent High Profile Successes:**
  - During the search for Air France 447 (2009-2011), black box location
  - The Coast Guard in 2013 found the missing fisherman, John Aldridge

Classical versus Bayesian Statistics

- **Experiment 1:**
  - A fine musician, specializing in classical works, tells us that he is able to distinguish if Haydn or Mozart composed some classical song. Small excerpts of the compositions of both authors are selected at random and the experiment consists of playing them for identification by the musician. The musician makes 10 correct guesses in 10 trials.

- **Experiment 2:**
  - The guy next to you at the bar says he can correctly guess in a coin toss what face of the coin will fall down. Again, after 10 trials the man correctly guesses the outcomes of the 10 throws.

Classical versus Bayesian Statistics

- **Classical Statistics Analysis**
  - You have the same confidence in the musician's ability to identify composers as in the bar guy's ability to predict coin tosses. In both cases, there were 10 successes in 10 trials.

- **Bayesian Statistics Analysis**
  - Presumably, you are inclined to have more confidence in the musician's claim than the guy at the bar's claim. Post analysis, the credibility of both claims will have increased, though the musician will continue to have more credibility than bar guy.

Bayesian Statistics 101

Simple Example

- **We have a system comprised of 2 components: Component 1 and Component 2.**

- **For each of the two components, 10 pass/fail tests are administered and results are recorded. Component 1 fails twice and Component 2 fails zero times.**
  - We can calculate the reliability of each component, R_1 and R_2.
  - We also want an assessment of the system reliability, assuming the components work in series.

$$R_{system} = R_1 * R_2 = \left(1 - \frac{2}{10}\right) * \left(1 - \frac{0}{10}\right) = 0.8 * 1 = 0.8$$

*** For the purposes of the next few slides, focus on Component 1.*

Bayesian Statistics Basics

1. **Construct prior from prior information,** $f(\theta)$

2. **Construct likelihood from test data,**

$$L(\boldsymbol{x} \mid \theta) = \prod_{\{i=1\}}^{n} f(x_i \mid \theta)$$

3. **Estimate posterior distribution using Bayes Theorem**

$$f(\theta|x) = \frac{L(x \mid \theta)f(\theta)}{\int L(x \mid \theta)f(\theta)}$$

$$\propto L(x \mid \theta)f(\theta)$$

Each of these steps requires careful consideration of both the system and the statistics!

Priors

- **The prior distribution of the reliability, $f_{prior}(R)$, is constructed from previous data or expert knowledge. This is your first assessment of the system.**

- **Say Component 1 was previously tested and failed 3 out of 40 tests: use Beta distribution.**

$$f_{prior}(R_1) \propto R_1^{\,n_p p} (1 - R_1)^{n_p(1-p)}$$

- **p is the reliability estimate and $n_p \geq 0$ weights the relevance of the prior test data.**

Careful thought should always be put into the prior distribution!

IDA

Likelihood

- **Tests are performed and the resulting test data is used in the likelihood function, L(x|R).**

- **The binary test data of Component 1 follows a Binomial distribution with probability of a pass of R_1**

$$L(x|R_1) \propto R_1^{s1}(1 - R_1)^{f1}$$

s_1 is the number of successes and f_1 is the number of failures from Component 1.

Posterior Distribution

- **Bayes' theorem is used to find the posterior reliability distribution, f_posterior(R|data).**

- **The posterior distribution is proportional to the product of the prior distribution and the likelihood function.**

- **For our example, choosing the Beta distribution as a prior is ideal for a few reasons: it ensures that R is between (0, 1) and it is the "conjugate" prior for the Binomial distribution.**

$$f_{posterior}(R_1) \propto R_1^{s1}(1-R_1)^{f1}R_1^{n_p p}(1-R_1)^{n_p(1-p)}$$
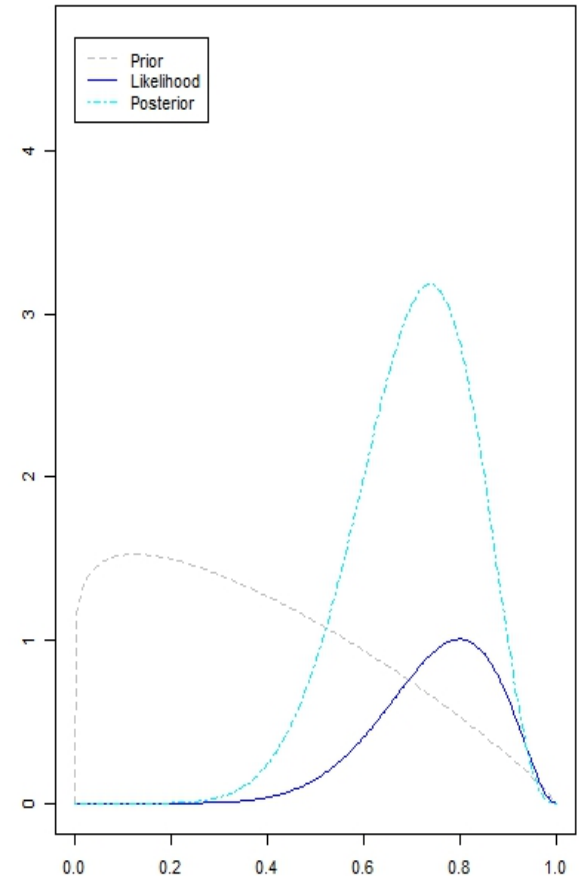$$\propto R_1^{s1+n_p p}(1-R_1)^{f1+n_p(1-p)}$$

$\longrightarrow$ *Also a Beta distribution with updated parameters: $s_1 + n_p p + 1$ and $f_1 + n_p(1-p) + 1$.*

Visualizing Bayesian Statistics
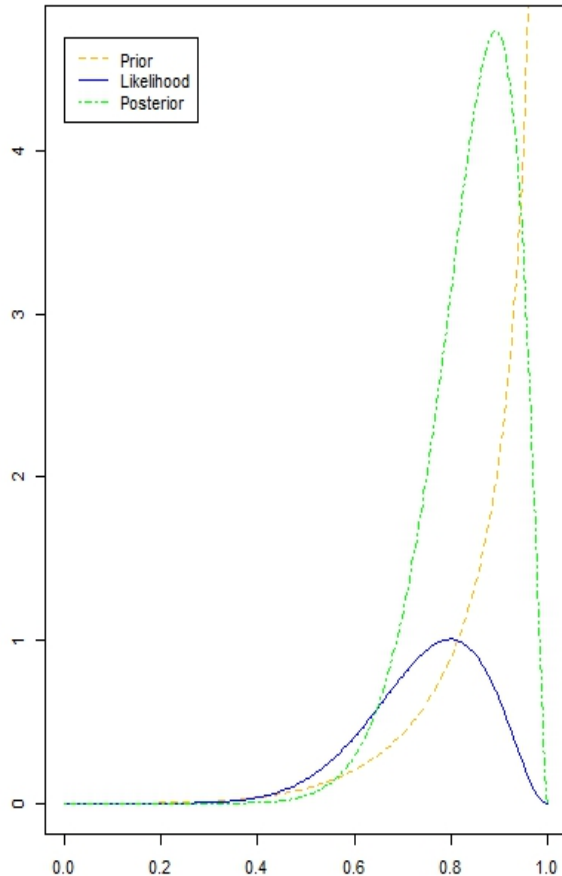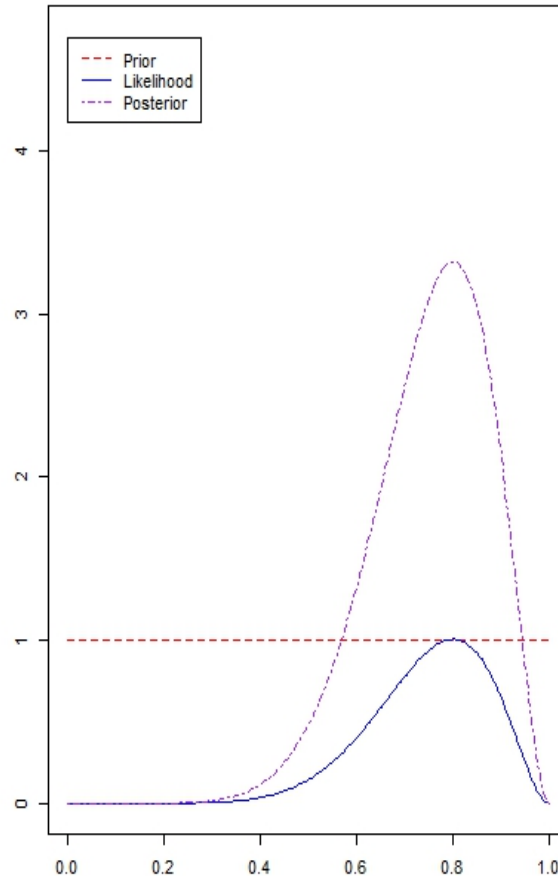
0.86 (0.76, 0.95)   *0.75 (0.58, 0.89)*   *0.70 (0.54, 0.85)*

Classical Estimate: 0.8 (0.55, 0.95)

When Should We Think About Using Bayesian Techniques

- **To obtain interval estimates (credible intervals) when there are zero failures**
  - Mean time between failure for short tests or for highly reliable systems
  - Interval estimates in kill-chain analysis where zero failures occur at any point along the kill-chain

- **If you are assessing a complex system mission reliability**
  - LCS Example - Confidence intervals are not straightforward to obtain using frequentist methods, impossible with zero failures in any sub-system

- **If there is relevant prior information to be incorporated in your analysis – this may include previous developmental (or operational) test data, engineering analyses, or information from modeling and simulation.**
  - BDS Example

Case Study: LCS

- **The Capability Development Document for LCS provides a reliability threshold for Core Mission functional area.**

Critical Subsystem	Total System Operating Time	Operational Mission Failures
Total Ship Computing Environment (full-time)	4500 hours	1
Sea Sensors and Controls (underway)	2000 hours	3
Communications (full-time)	4500 hours	0
Sea Engagement Weapons (on-demand)	11 missions	2

- **The target reliability for Core Mission is 0.80 in 720 hours.**

➢ Assume the functional area is a series system: system is up if all subsystems are up.

Data are notional, based on preliminary results.

IDA
Prior Assumptions: LCS

- **On-demand system**
  - Assume no belief in the relevance of prior knowledge, $n_p = 0$

- **Continuous systems**
  - The Gamma prior parameter a is set to 1, giving large variance. To ensure the 50th percentile is set at λ_{50}=1/MTBF$_{guess}$, choose b=log(2)×MTBF$_{guess}$
  - MTBF$_{guess}$ chosen by solving the reliability function at the requirement

*Guiding Principles in Prior Selection:*

- *Start with the properties of the parameter of interest*
- *Decide on what prior information to use*
- *Allow for the analysis to change freely based on the data observed*
- *Priors specified at the system level, as opposed to mission level – check impact on system prior*
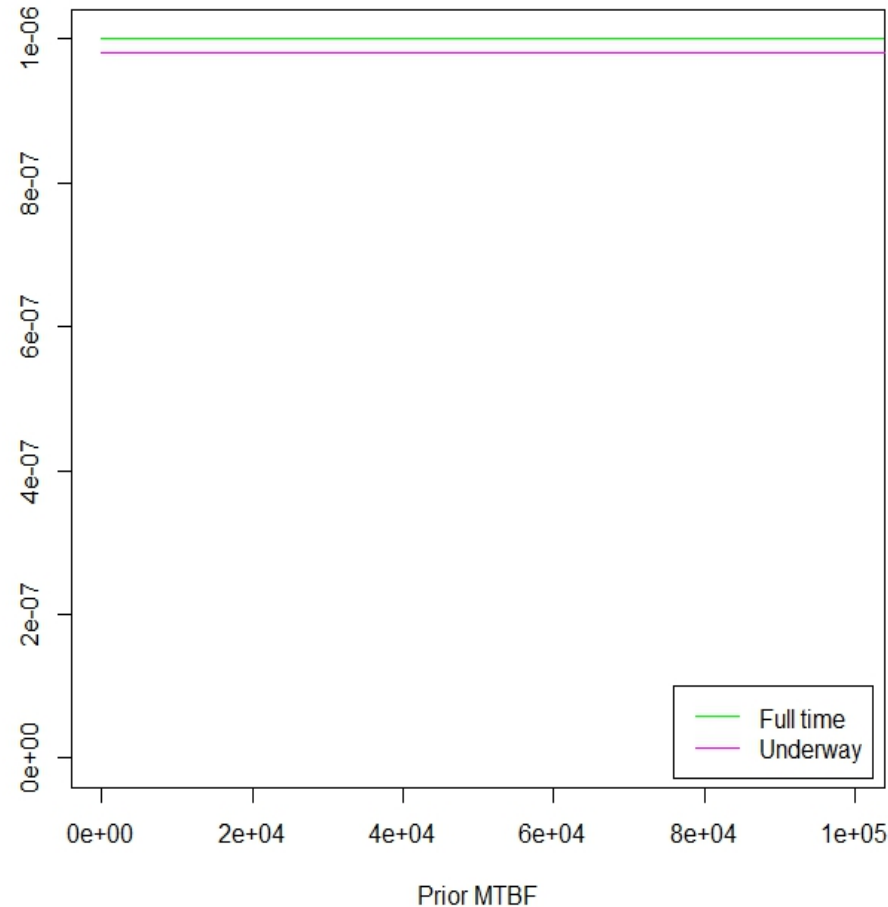
IDA   Prior Specification: On Demand System
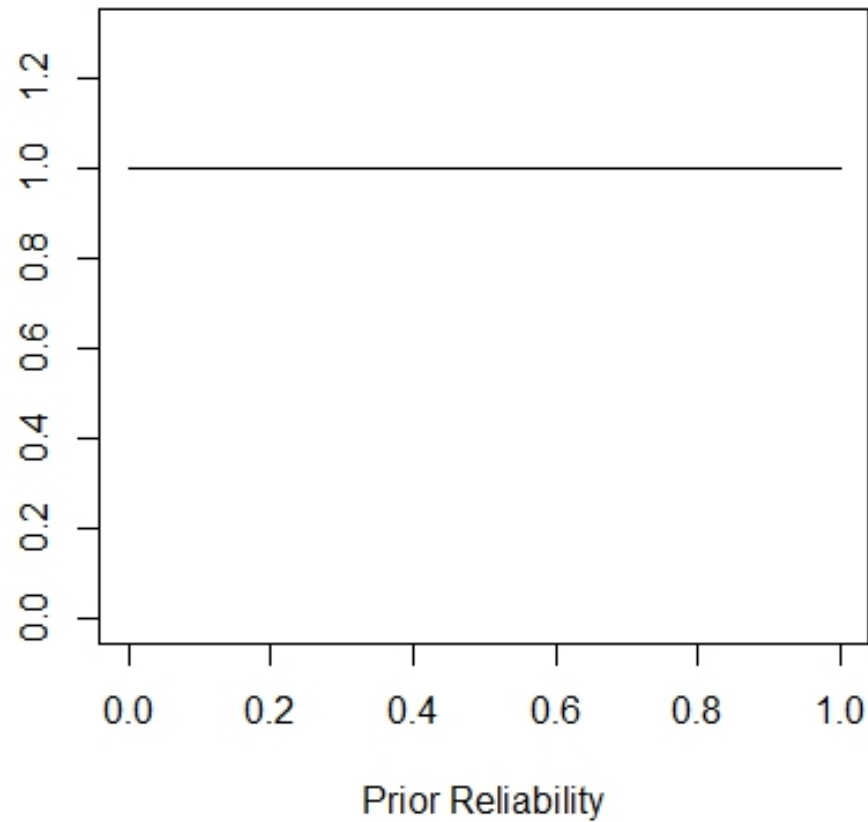
Visualizing the Prior Specification: LCS

Possible Prior 1: Too much information, if the prior probability of a value is 0, no amount of data can move the posterior MTBF there!
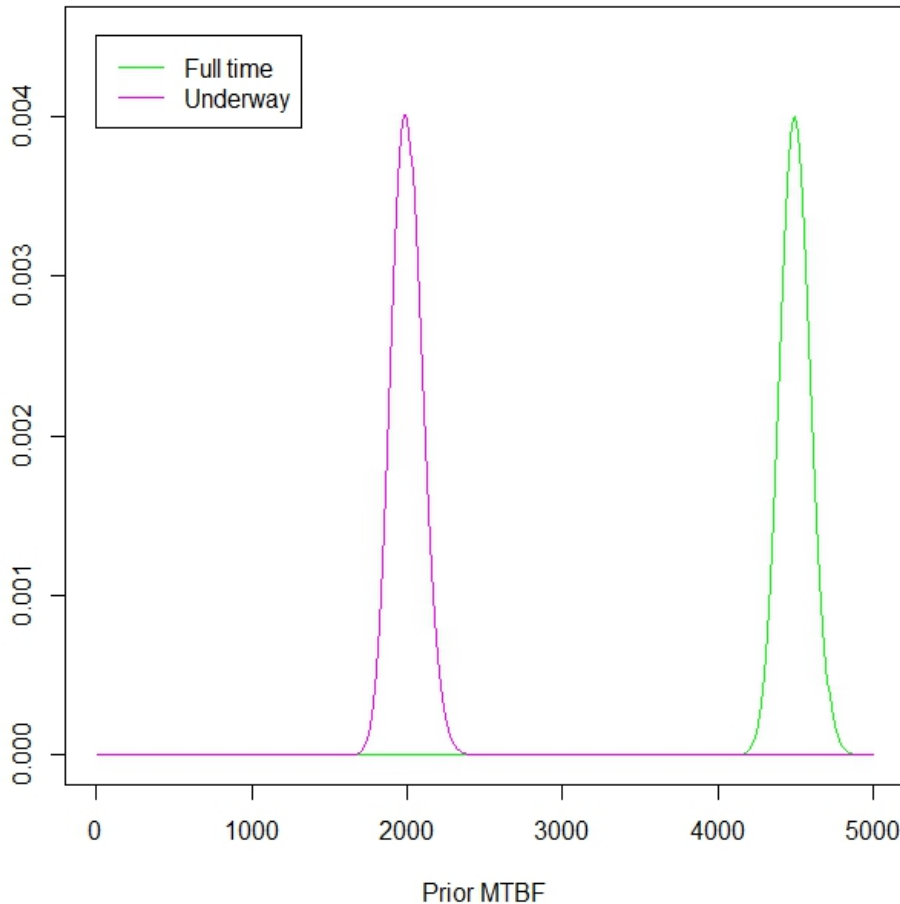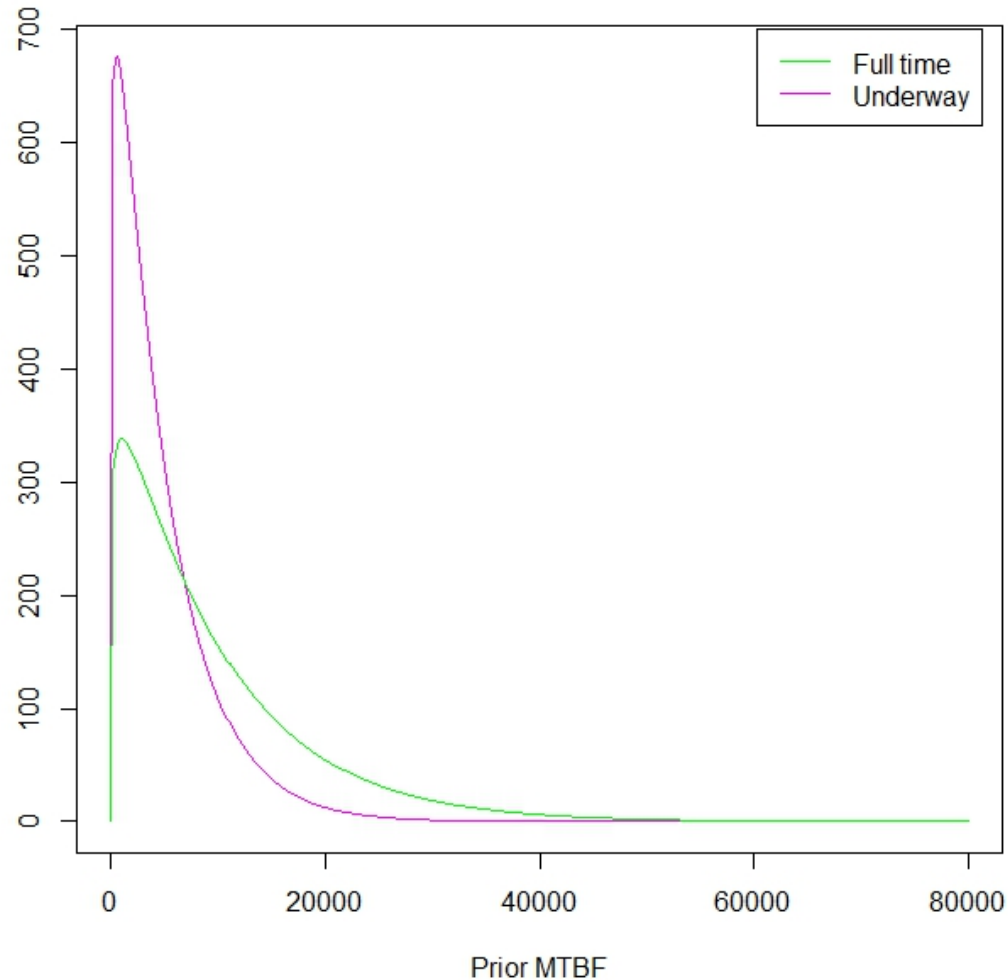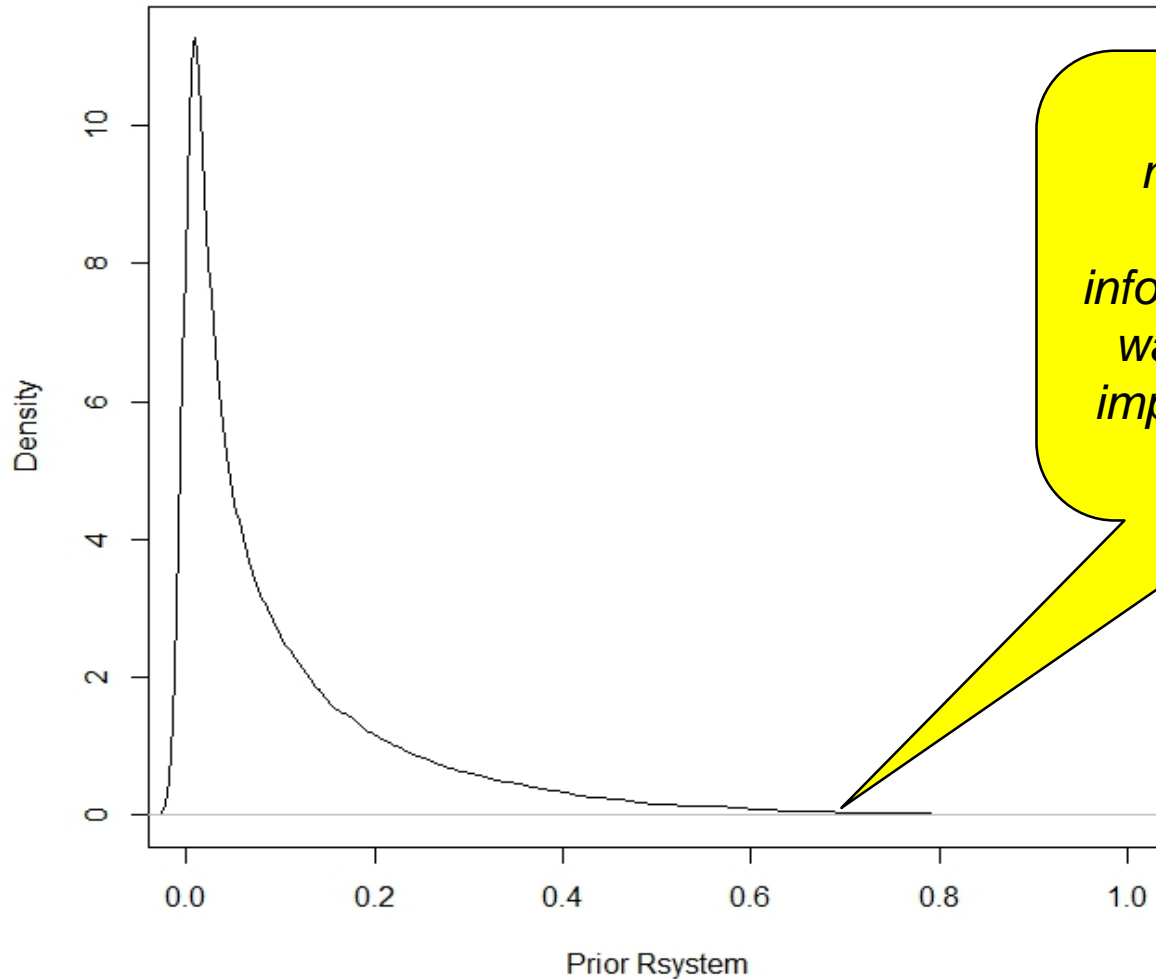
Possible Prior 2: Too little information, for continuous measures, flat priors can be problematic when there are few or zero failures.

Visualizing the Prior Specification: LCS

IDA

Possible Prior 3: Bounding the prior in case there are few failures, but ensuring enough flexibility for the data to speak for itself.

Core Mission Reliability Prior

R Code for implementation

```
> for(i in 1:B){
+
+        lambdaTC[i]=rgamma(1,TSCE+aT,TSCET+bT)
+
+        lambdaSSC[i]=rgamma(1,SensCont+aS,SensContT+bS)
+
+        lambdaCOMM[i]=rgamma(1,Comm+aC,CommT+bC)
+
+        pSEW[i]=rbeta(1,SEWs + pguess*pweight,SEWf+(1-pguess)*pweight)
+
+        Rsys[i]=exp(-720*lambdaTC[i])*exp(-720*lambdaSSC[i])*exp(-720*lambdaCOMM[i])*pSEW[i]
+ }
```
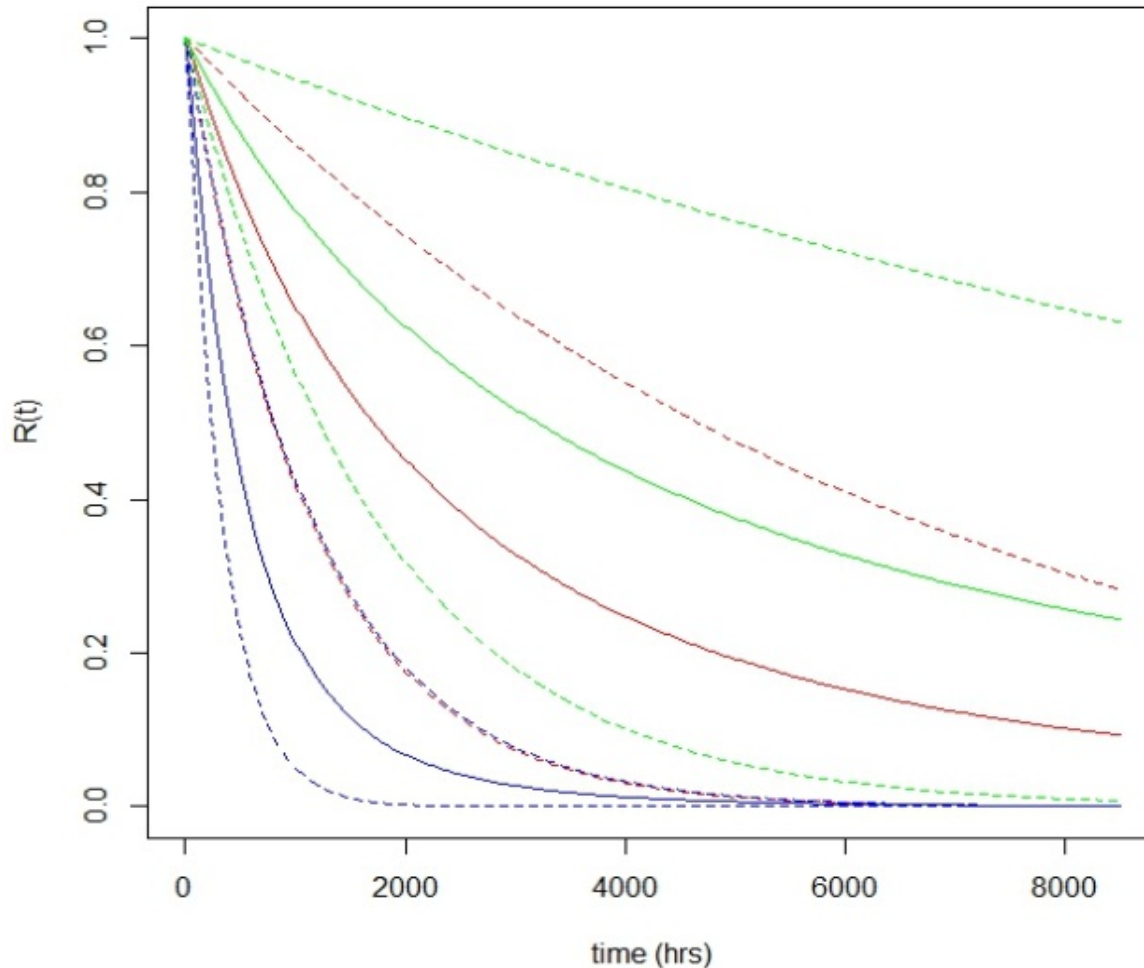
IDA
Comparison of Results

	Classical MTBOMF	Classical Reliability at 720hrs	Bayesian MTBOMF	Bayesian Reliability at 720hrs
TSCE	4500 hrs (1156 hrs, 42710 hrs)	0.85 (0.54,0.98)	3630 hrs (1179 hrs, 6753 hrs)	0.73 (0.54,0.90)
SSC	667 hrs (299 hrs, 1814 hrs)	0.33 (0.09,0.67)	697 hrs (332 hrs, 1172 hrs)	0.31 (0.11,0.54)
Comm	> 2796 hrs	> 0.77*	10320 hrs (1721 hrs, 18210 hrs)	0.83 (0.66,0.96)
SEW		0.82 (0.58,0.95)		0.77 (0.62,0.91)
Core Mission		?????		0.15 (0.05, 0.27)

Many ways to think about calculating this, none of which are particularly satisfactory

Note the impact of the prior is greater in the one failure system

Full Mission mean is comparable with the simple point estimate

TSCE: Total Ship Computing Environment
SSC: Sea Sensors and Controls
Comm: Communications
SEW: Sea Engagement Weapons

8/1/2016-20   *Zero failures occurred in the notional on-demand system data*

Core Mission Reliability Over Time

IDA

Posterior reliability as a function of time for TSCE (red), SSC (blue), and Communications (green)

Reporting of Results

IDA

Posterior mean and 80% intervals for each subsystem and the total system reliability over 15 days (light blue) and 30 days (dark blue) for the notional example.

Core Mission

Legend:
- t=360hrs
- t=720hrs

X-axis: TSCE, SeaSensors&Controls, Communications, SeaEngagementWeapons, Total

Y-axis: R(t)

8/1/2016-22

Value of Bayesian Statistics for LCS
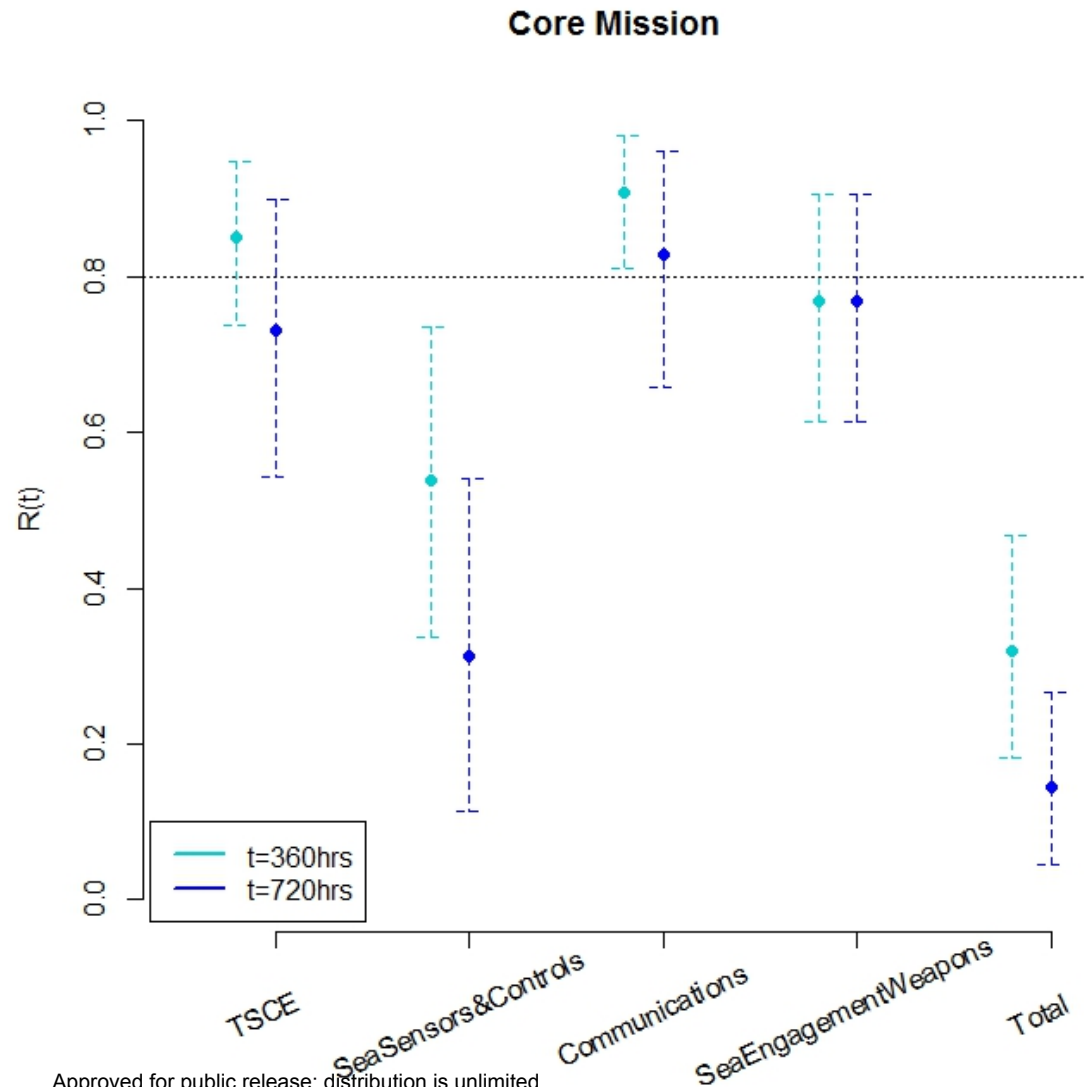
- **Avoids unrealistic reliability estimates when there are no observed failures.**

- **In our notional example (zero failures for the Communications system), the Bayesian approach helped us solve an otherwise intractable problem.**

- **Obtaining interval estimates is straightforward for system reliability**
  - Frequentist methods would have to employ the Delta method, Normal approximations, or bootstrapping.

- **Flexibility in developing system models**
  - We used a series system for the core mission reliability
  - Many other system models are possible and we can still get full system reliability estimates with intervals.

Case Study: Biochem Detection System

- **Bio-chemical Detection System analyzes environmental samples and identifies chemical, biological, radiological agents. Each subsystem is comprised of a collection of components of various sensitivity.**

- **KPP performance requirement for each subsystem: detect 85-90% of samples that come into the lab.**

- **Multiple Tiers of testing**
  - Tier 2: component level testing with agent in pristine matrix to each device (vendor testing)
    - » 5 components: total of almost 2000 runs
  - Tier 3: component level testing with agent in various matrices, such as soil, food, or swab (vendor testing)
    - » 8 components: total of about 3600 runs
  - DT/OT: subsystem level test with agent in matrices, triage procedures (government testing)
    - » 80-90 samples tested on multiple components, final call made by operator based on component output

Approved for public release; distribution is unlimited.

IDA   Case Study: Biochem Detection System

- **DT/OT: set concentration levels, comparatively small sample size**

- **Standard logistic regression on the Tier 3 data could be problematic**
  - **All detections or non-detections**
- **Bayesian approach with a dispersed prior:**

$$logit(P_D) = \beta_1 * conc + \beta_2^{matrix} + \beta_3^{agent}$$
$$(\beta_1, \boldsymbol{\beta_2}, \boldsymbol{\beta_3}) \sim Multivariate\ Normal(\boldsymbol{0}, \boldsymbol{W})$$

- **Explicitly forcing a dependence on concentration.**

- **Leverage all device runs to learn about each agent/matrix combination performance curve.**
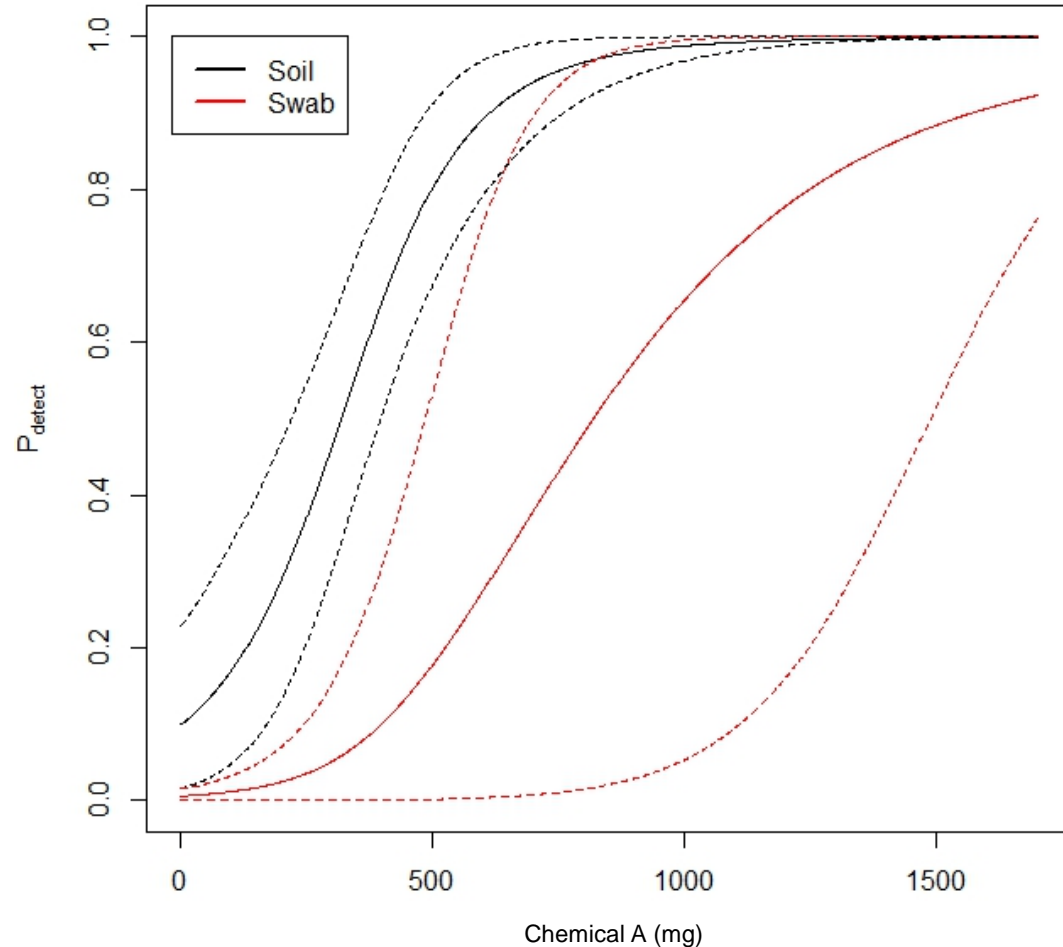
IDA

R code implementation

```
> post=function(beta1,beta2,beta3){
+            betavec=c(beta1,beta2,beta3)
+
+            val=dtnorm(betavec[1],0,10^3,lower=0,log=TRUE)+
+
+                dmvnorm(betavec[-1],rep(0,length(betavec[-1])),diag(rep(10^3,length(betavec)-1)),log=TRUE)+
+
+                sum(dbinom(y,1,mylogit(beta1*log(X[,1])+beta2[X[,2]]+beta3*X[,3]),log=TRUE))
+
+
+            return(val)
+       }
```

R code implementation

IDA
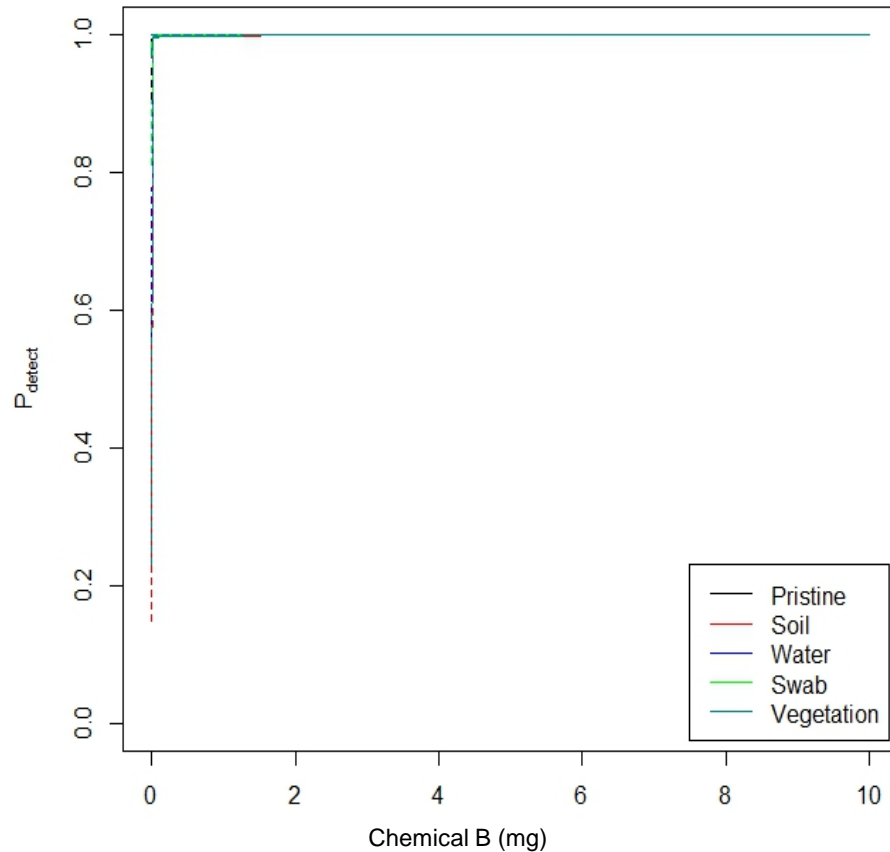
```
> for(i in 1:size){
+
+               #update beta1
+               cand.beta1=rnorm(1,beta1,0.1)
+               if(cand.beta1 > 0){
+               r = post(cand.beta1,beta2,beta3) - post(beta1,beta2,beta3)
+               u = runif(1) <= exp(r)
+               arate1 = arate1 + u
+               beta1 = cand.beta1*(u==1) + beta1*(u==0)}
+
+               #update beta2
+               for(j in 1:length(unique(X[,2]))){
+                       cand.beta2=beta2
+                       cand.beta2[j]=rnorm(1,beta2[j],0.7)
+                       r = post(beta1,cand.beta2,beta3) - post(beta1,beta2,beta3)
+                       u = runif(1) <= exp(r)
+                       arate2[j] = arate2[j] + u
+                       beta2[j] = cand.beta2[j]*(u==1) + beta2[j]*(u==0)}
+
+               #update beta3
+               for(j in 2:length(unique(X[,3]))){
+                       cand.beta3=beta3
+                       cand.beta3[j]=rnorm(1,beta3[j],0.7)
+                       r = post(beta1,beta2,cand.beta3) - post(beta1,beta2,beta3)
+                       u = runif(1) <= exp(r)
+                       arate3[j] = arate3[j] + u
+                       beta3[j] = cand.beta3[j]*(u==1) + beta3[j]*(u==0)}
+
+
+
+               par[i,] = c(beta1,beta2,beta3) #save current parameter values
+
+       }
```
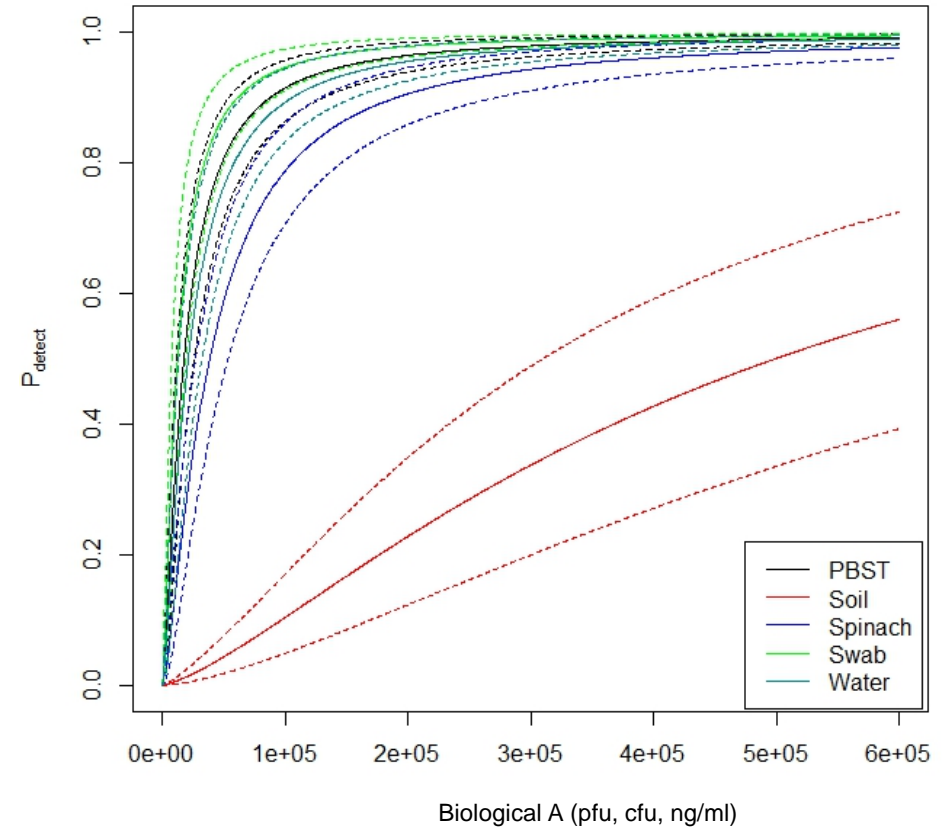
R code implementation

IDA

```
#update beta2
for(j in 1:length(unique(X[,2]))){
        cand.beta2[j]=rnorm(1,beta2[j],0.7)
        r = post(beta1,cand.beta2,beta3) - post(beta1,beta2,beta3)
        u = runif(1) <= exp(r)
        arate2[j] = arate2[j] + u
        beta2[j] = cand.beta2[j]*(u==1) + beta2[j]*(u==0)}
```

Case Study: Biochem Detection System

IDA

Component: Chemical Detector 1
Agent: Chemical A
Matrix: Soil, Swab

IDA

Case Study: Biochem Detection System

Chemical Detector 2

Biological Detector 1

Case Study: Biochem Detection System

Chemical C (mg)

Chemical Detector 3

Legend (left plot):
- Rainwater
- Swab
- Soil
- Veg

Concentration (ug/L)

Chemical Detector 4

Legend (right plot):
- Chemical A (ug)
- Chemical B (ug)
- Chemical C (ug)
- Chemical C (ug)
- Chemical D (ug)
- Chemical E (ug)
- Chemical F (ug)
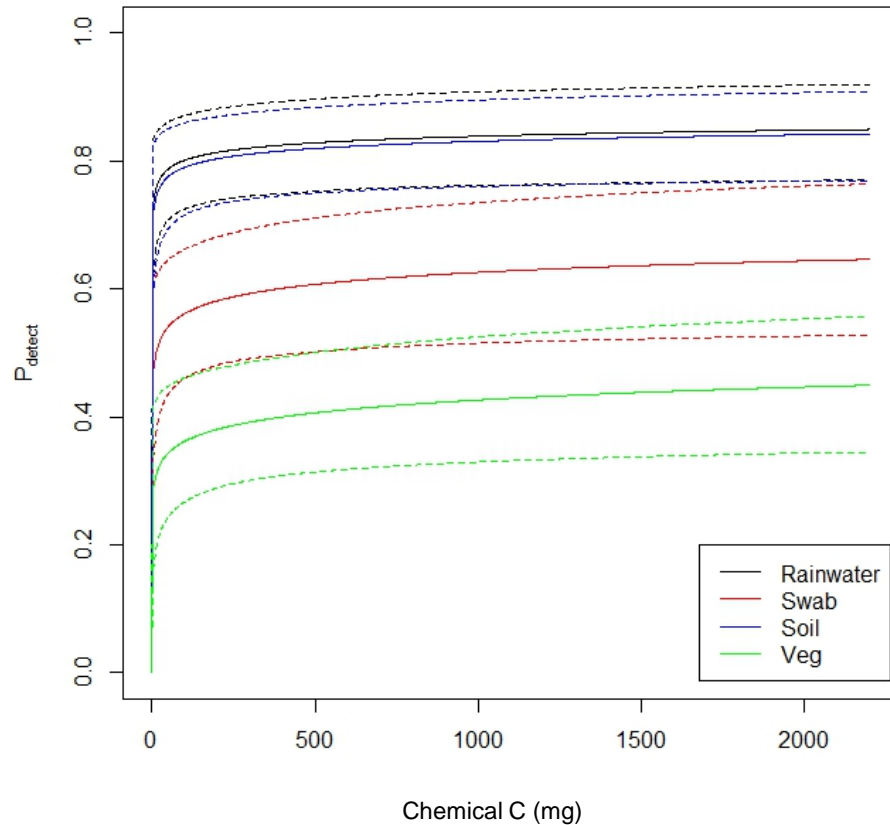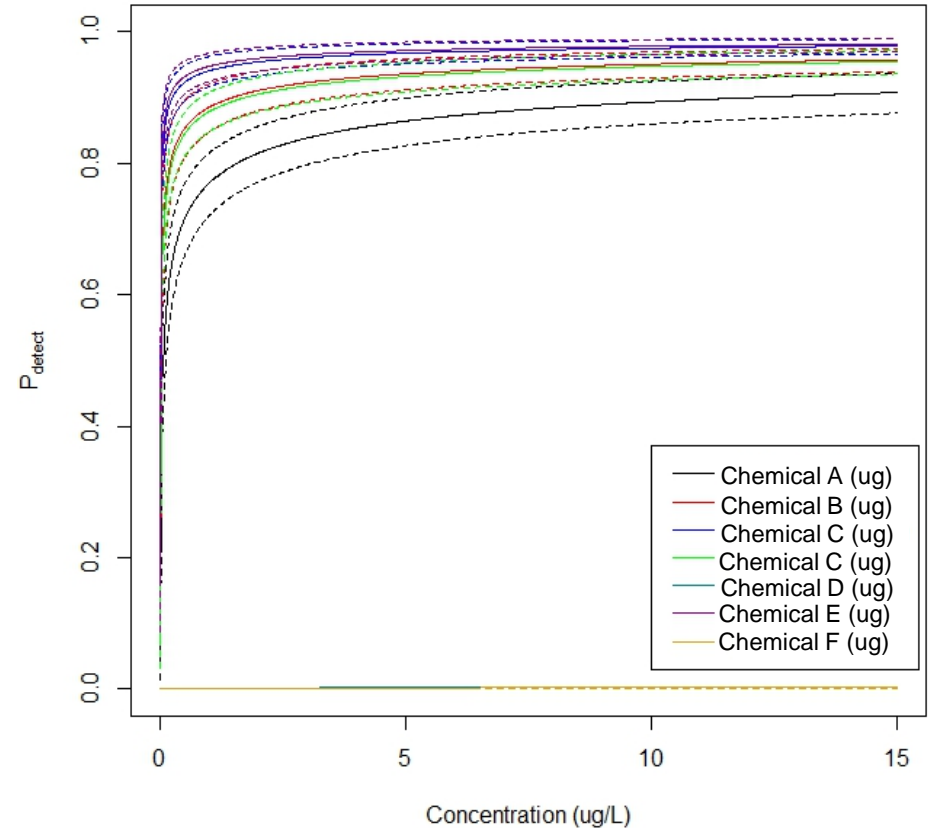
Value of Bayesian Statistics
for Biochem Detection System

- **Tier 2 and Tier 3 produced a lot of data which we can leverage to make informed decisions.**

- **By knowing the concentrations of agents within various matrices that each component can detect, we can determine concentrations that the system of devices might be easy or difficult for the operators to identify in DT/OT.**

- **This analysis can serve as the basis for the analysis of the DT/OT data.**

IDA   **Discussion: When Is it Worth the Effort?**

- **Inclusion of prior information from prior testing, modeling and simulation, or engineering analyses only when it is relevant to the current test. We do not want to bias the OT results.**

- **Even when including prior information, the prior must have enough variability to allow the estimates to move away from what was previously seen if the data support such values.**

- **We can use very flexible models for many types of test data (e.g. kill chains, complex system structures, linking EFFs to SA) and obtain estimates more readily than with the frequentist paradigm. The model and assumptions have to make sense for the test at hand.**

Discussion: Other Resources

IDA

- **For other R packages that provide easy to implement tools and short but informative how to guides with examples, see**

  https://cran.r-project.org/web/views/Bayesian.html

  ➢ arm, bayesm, and bayesSurv are good places to start

As with any new statistical method, it is important to have an expert review your work the first few times you apply these techniques.
There are many ways to accidentally do bad statistics!

Backup Slides

Common Conjugate Models

Likelihood	Parameter	Prior	Posterior
Binomial(s+f, R)	$0 \leq R \leq 1$	Beta(a,b) $a > 0, b > 0$	Beta(a',b') a'= a + s b'= b + f
Poisson(λ)	$\lambda > 0$	Gamma(a,b) $a > 0, b > 0$	Gamma(a',b') a' = a + n b' = b + Σ t
Exponential(λ)	$\lambda > 0$	Gamma(a,b) $a > 0, b > 0$	Gamma(a',b') a' = a + n b' = b + Σ t

** For more examples, see Conjugate Priors Wikipedia page or "Bayesian Reliability" pg 48