SAND2022-2078C

Sandia National Laboratories

Exceptional service in the national interest

Target Detection on Hyperspectral Images Using MCMC and VI Trained Bayesian Neural Networks

Daniel Ries, Jason Adams, Joshua Zollweg

3/10/2022

IEEE Aerospace 2022

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

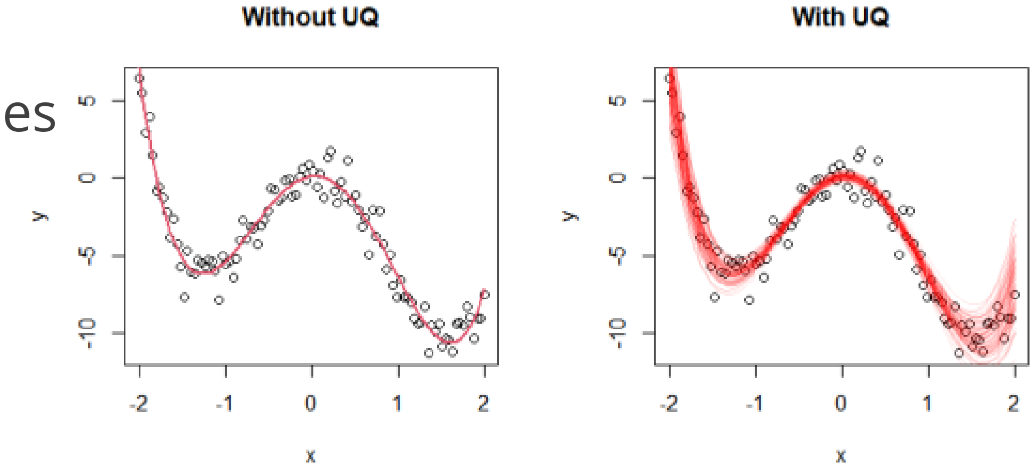
Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Motivation

- Deep learning (DL) has become popular tool for finding trends in hyperspectral imagery (HSI)

- Traditional DL does not quantify uncertainty of predictions
  - This is problematic for high **consequence** problems

- Bayesian neural networks (BNN) provides uncertainties for powerful DL predictions



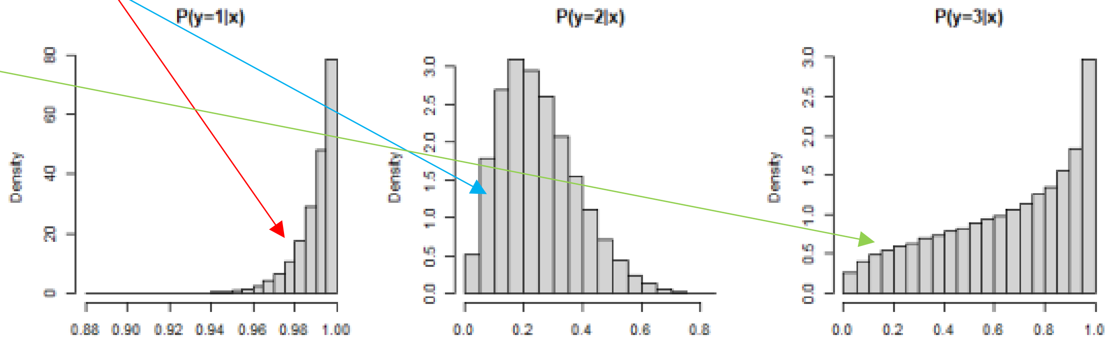
Target Detection Example

Do these pixels contain target? **($y_i = 1$ means target)**

- Pixel 1: $P(y_i \widehat{= 1} | x_1, \theta) = 0.99$
- Pixel 2: $P(y_i \widehat{= 1} | x_2, \theta) = 0.25$
- Pxiel 3: $P(y_i \widehat{= 1} | x_3, \theta) = 0.70$

How confident are we in *estimates* of $P(y_i = 1 | x_i, \theta)$?

- Pixel 1: We are very confident there's a target
- Pixel 2: We are pretty sure there's no target
- Pixel 3: Our "best" estimate is 0.7,
  but who knows?

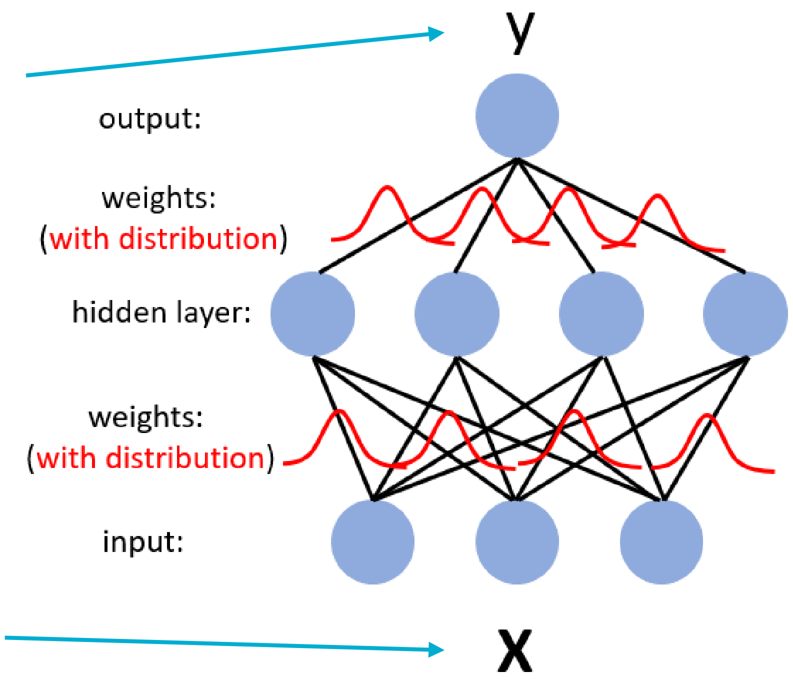


Bayesian Neural Network

Output model prediction for pixel *i*: $P(y_i = 1|x_i, \theta)$

Posterior distribution $\mathrm{p}(\theta|y,x)$

Input spectrum from pixel *i*



Goal: estimate posterior distribution of θ, $\mathrm{p}(\theta|y,x)$, so we can estimate $P(y_i = 1|x_i, \theta)$, with uncertainty

[1]Image credit: A Batched Scalable Multi-Objective Bayesian Optimization Algorithm https://arxiv.org/pdf/1811.01323.pdf

Operationalizing Decisions using Uncertainty

We can use **high confidence (HC) sets** to help make decisions about our target detection models' predictions

HC set contains all pixels such that:

- $P(P(y_i = 1|x_i, \theta), < L|y) > 1 - \alpha$ OR $P(P(y_i = 1|x_i, \theta), > U|y) > 1 - \alpha$
  - $(1 - \alpha)$ is the confidence you want in your prediction
  - L, U are your decision thresholds for not-target and target, respectively

Choosing α, L, U according to your application, we can:

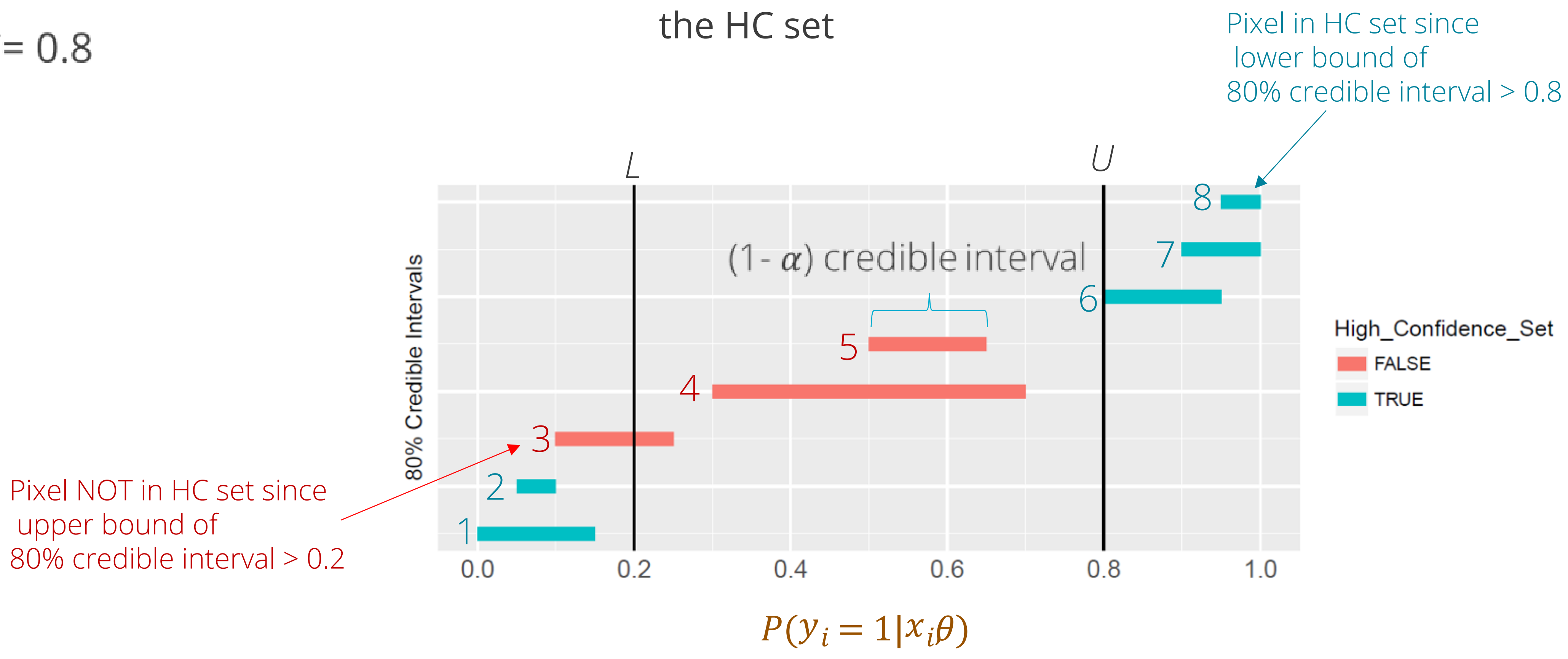
- Reduce false alarm rates
- Reduce burden on analysts
  - E.g. Automate predictions/decisions on HC pixels, send remaining to analysts for further review

HC Set Example

Example with:

- $\alpha = 0.2$ (80% confidence)
- $L = 0.2$
- $U = 0.8$

- Locations 1,2,6,7,8 are in the HC set
- Locations 3,4,5 are not in the HC set

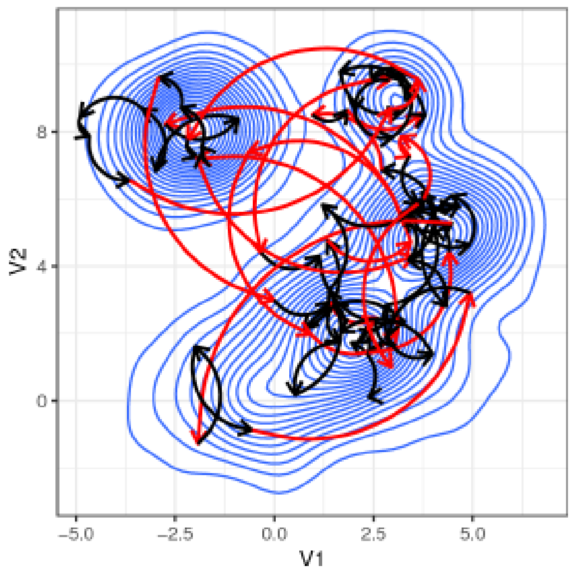


How do we train a BNN?

- Markov Chain Monte Carlo (MCMC)
  - Very accurate
  - Relatively slow
  - MCMC's approximation improves as the number of MC samples increases
- Variational Inference (VI)
  - Very fast
  - Mean-field assumption can affect accuracy of results
  - VI's approximation improves as the sample size *n* increases

VI is constrained in its ability to approximate the posterior by data size and MCMC is constrained by computation time.

MCMC Random Walk



VI Schematic

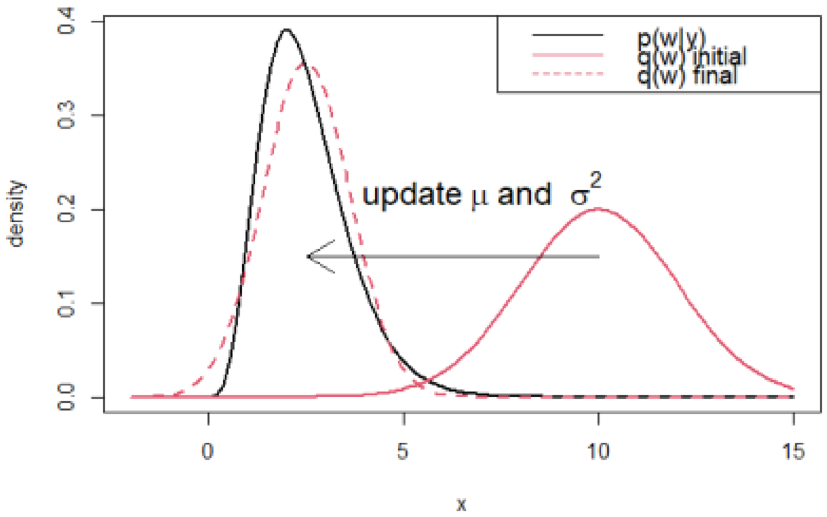


Image credit: Adaptive MCMC for multimodal distributions. The Annals of Statistics. 2020.

Megascene

Simulate **9** HSI scenes "Megascene" from DIRSIG

- **Three** MODTRAN-based atmospheres
  - Mid-latitude summer (MLS)
  - Sub-artic summer (SAS)
  - Tropical (TROP)
- **Three** times of day
  - 12:00
  - 14:30
  - 15:45

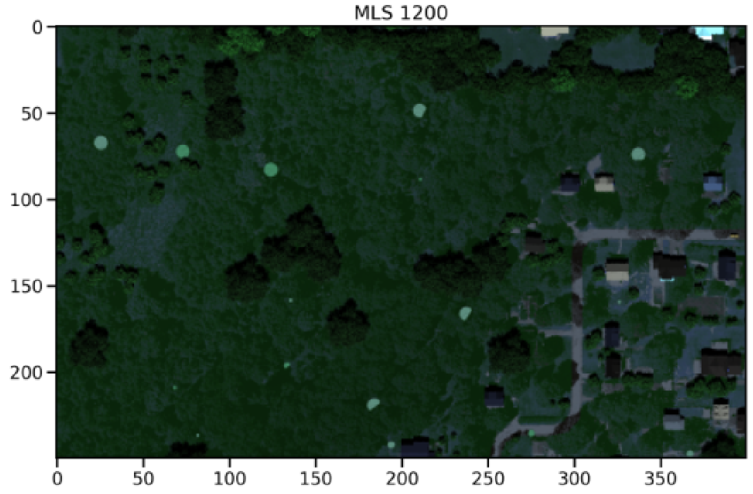
At every pixel, we have a full spectrum response across 211 spectral bands using AVIRIS-like sensor

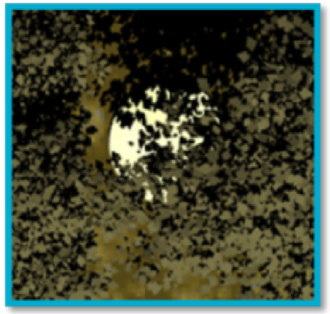
- 0.4 to 2.5 μm
- Elevation 4km
- Pixel size $1m^2$

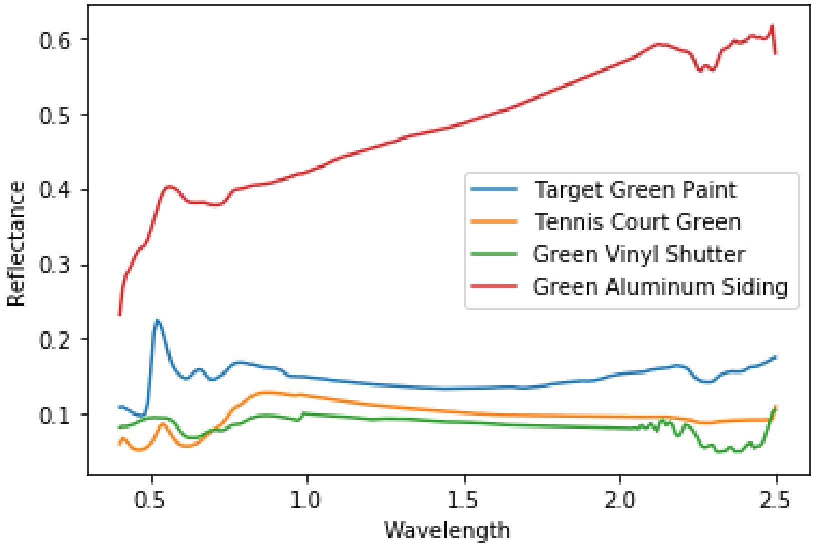
Megascene Targets

We manually add targets (green paint) to the scene

- 125 green discs in each of the **9** scenes, in different locations for each scene
- Radii of discs ranges from 0.1 to 4m
- Scene contains other green elements with similar spectral signatures





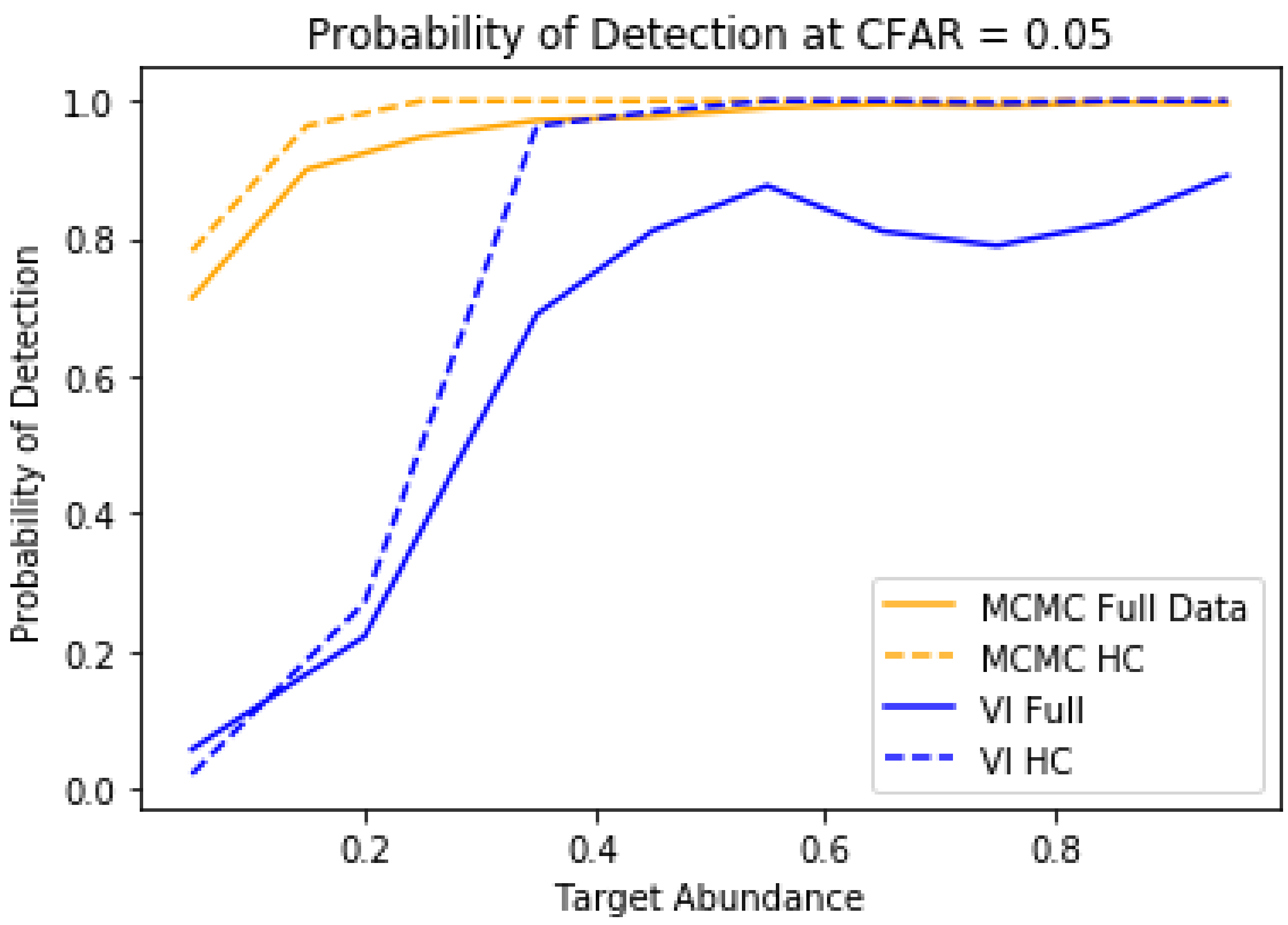


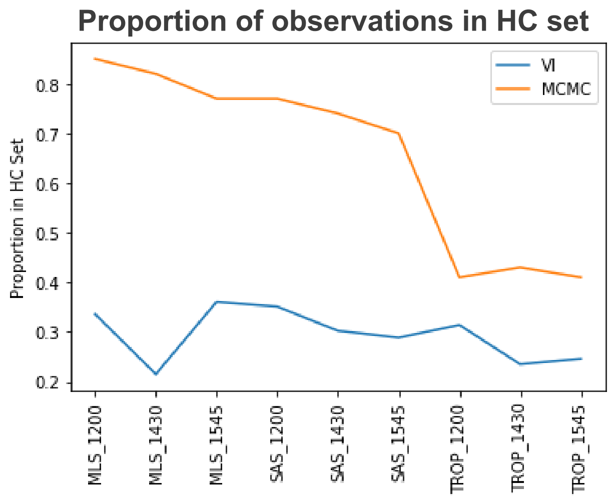
BNN Architecture Details

- 3 hidden layers
- 10 nodes per layer
- Sigmoid activation functions
- Priors on all model weights ~ N(0,10)
- Do fPCA on each spectrum, and use first 25 PCs as input features for that pixel

Comparison of MCMC vs VI and Full vs HC

- MCMC vastly outperforms VI at low target abundance levels
  - Methods converge at ~35% target abundance for HC data
  - MCMC outperforms VI for full data at all target abundances



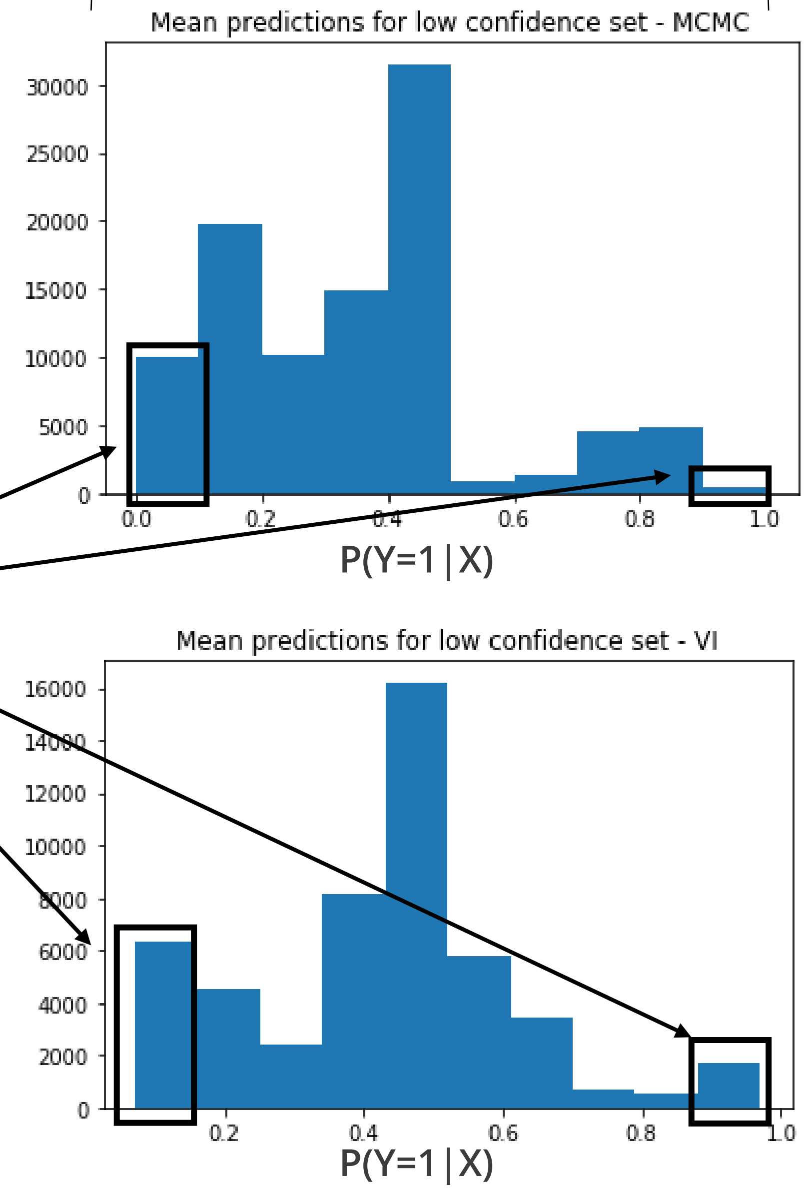


Doesn't P(Y=1|X) give my uncertainty?

- "If $P(y_i = 1|x_i, \theta)$ = 0.95, shouldn't I be 95% confident in my prediction?"
  - $P(y_i = 1|x_i, \theta)$ is an *estimate* itself

- What if your estimate of 0.95 had a confidence interval of (0.05,0.98)?
  - This isn't just a thought exercise
    - There were a significant number of pixels whose estimated P(Y=1|X) was close to 0 or 1, but whose CI width was >0.8 (a width of 1 is a CI of 0-1!)

Distribution of estimates with CI widths > 0.8



Conclusions

- Quantifying uncertainty in target detection has **significant** benefits
  - Create HC sets for quicker analyses
  - Reduce false alarms

- **Don't** treat an estimated probability ($P(y_i = 1|x_i, \theta)$) as your only *degree of confidence* for target/non-target
  - Particularly for high-consequence problems

- MCMC gave **better** results than VI
  - Although computationally faster, VI took a lot of work to train
  - Recommend caution with VI algorithms by non-Bayesian experts

Questions?

Thank you for listening!

dries@sandia.gov

Backup Slides

Bayesian Uncertainty for Target Detection

For the HSI target detection problem, we are interested in:

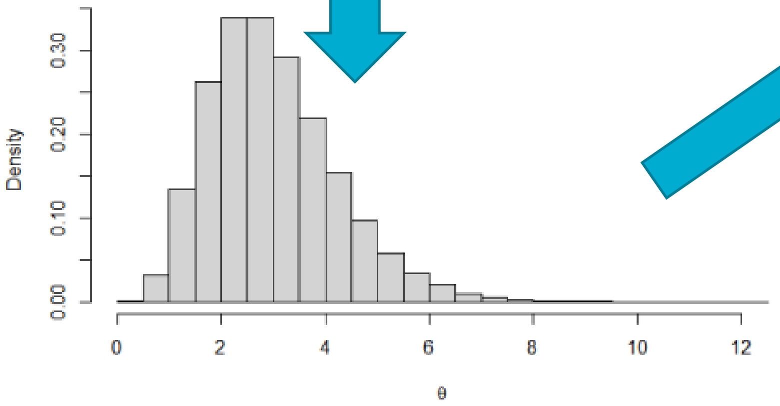
$$P(pixel\ i\ contains\ target|spectral\ measurements) \equiv P(y_i = 1|x_i, \theta)$$

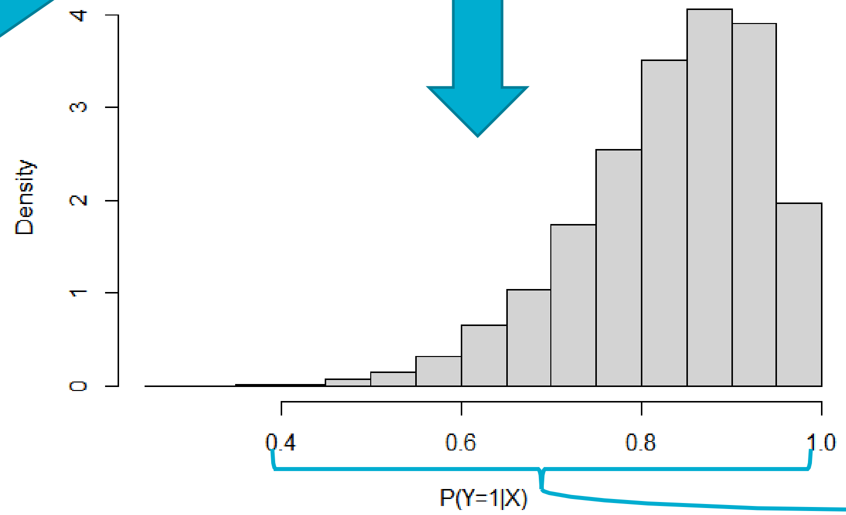
Uncertainty in this estimate can be captured by the **posterior distribution**, $p(\theta|y,x)$

$p(\theta|y,x)$

Apply posterior to quantity of interest

$P(y_i = 1|x_i\theta)$



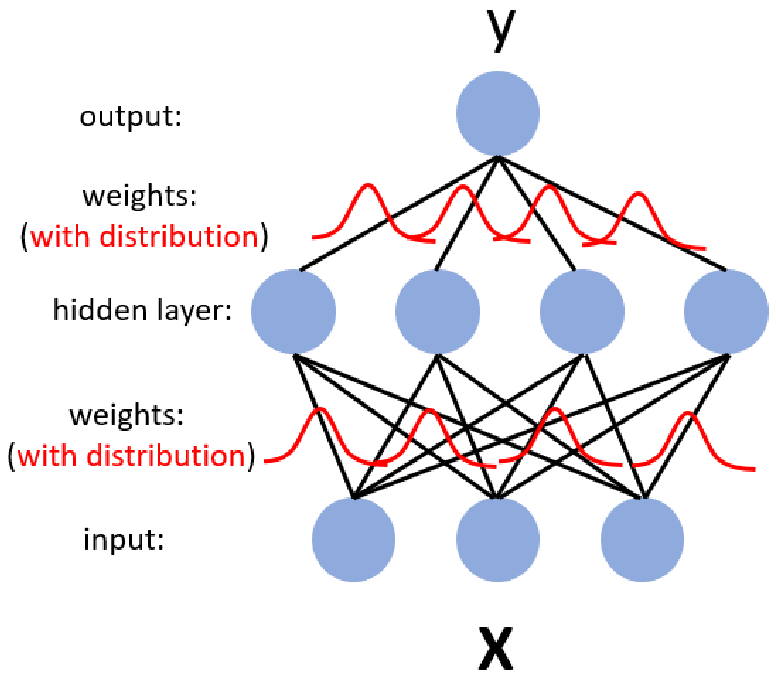


$P(y_i = 1|x_i, \theta)$ is *probably* about0.8, pretty sure it's >0.6 at least

Bayesian Neural Network

A Bayesian neural network (BNN) for target detection can be written mathematically by:

$$y_i|x_i,\theta \sim Bernoulli(\mu(x_i;\theta))$$
$$\mu(x_i;\theta) = f_{\theta_o}\left(f_{\theta_L}\left(f_{\theta_{L-1}}(\dots f_{\theta_1}(x_i)\dots)\right)\right)$$
$$\theta \sim D(\Psi)$$



y_i: binary indicator for whether pixel i contains target or not

x_i: spectral band (vector of length 211)

$\mu(x_i;\theta)$: $P(y_i = 1|x_i,\theta)$, or the probability that pixel i contains target

θ: model parameters/weights to be estimated ($\theta = (\theta_o, \theta_1, \theta_2, \dots, \theta_L)$)

Ψ: hyperpriors (assumed to be known)

f_{θ_o}: output function with parameters θ_o

f_{θ_l}: nonlinear activation function with parameters θ_l, $l = 1, \dots, L$

Goal: estimate posterior distribution of θ, $p(\theta|y,x)$, so we can estimate $P(y_i = 1|x_i,\theta)$, with uncertainty

[1]Image credit: A Batched Scalable Multi-Objective Bayesian Optimization Algorithm https://arxiv.org/pdf/1811.01323.pdf

Training/Test Splits

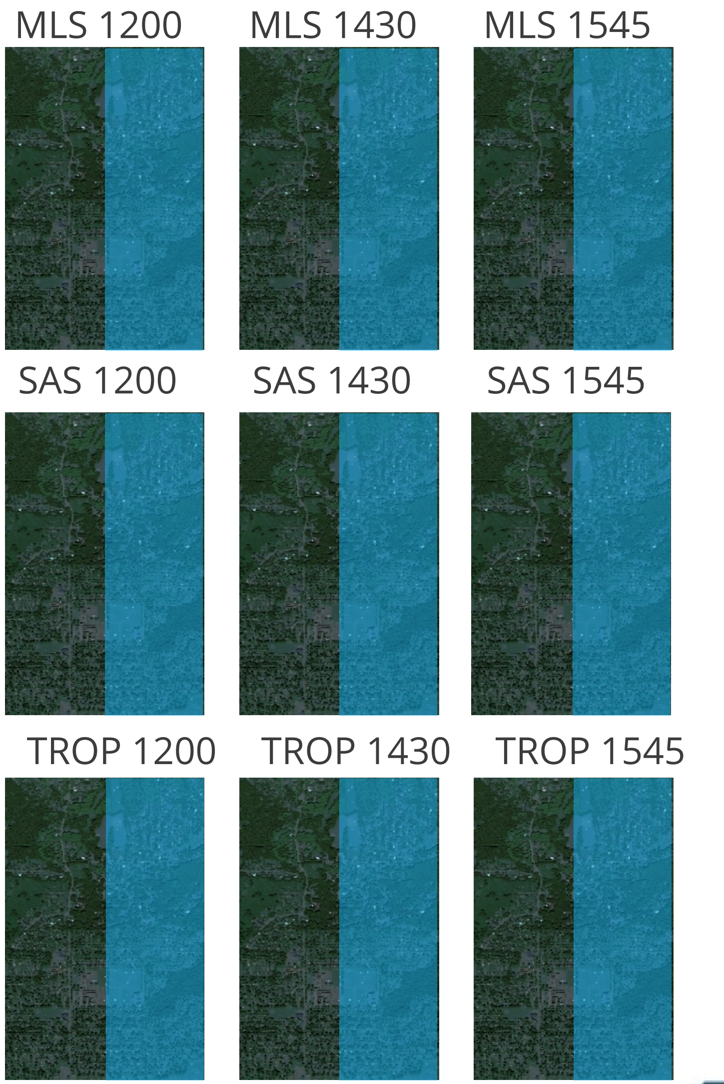
In order to understand the generalizability of the models across atmospheres, time, and space:

- Train on left hand side of **only** MLS1200
- Test on right hand side of all **9** scenes

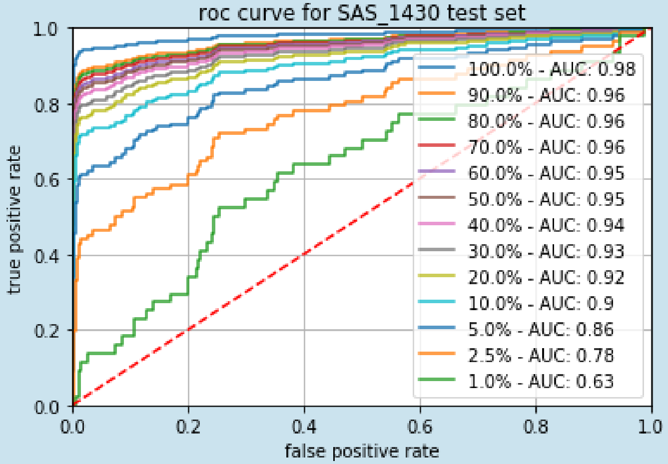
Train on pixels in shaded region

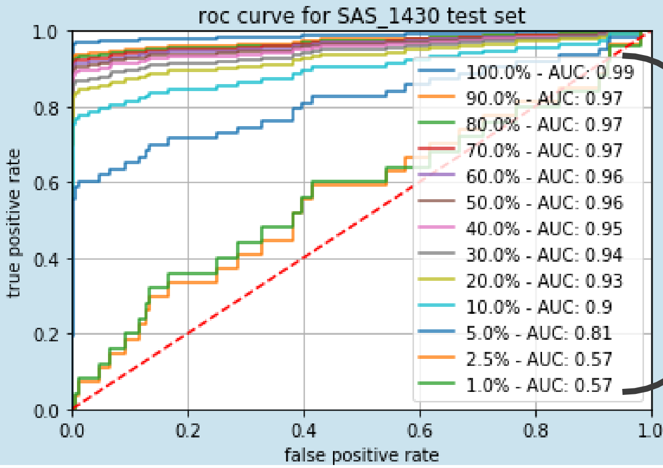
MLS 1200

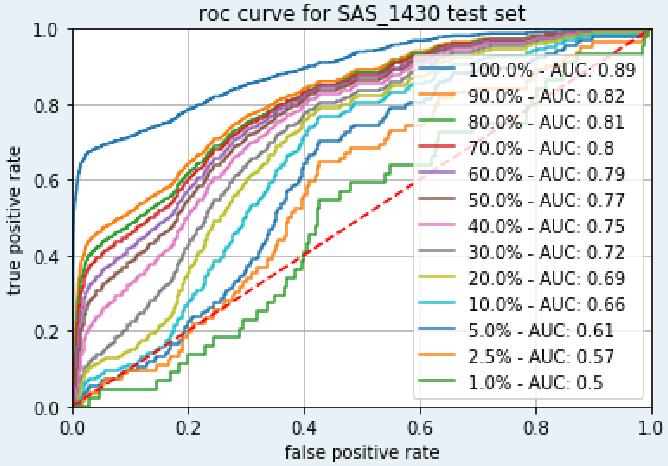
Test on pixels in shaded region

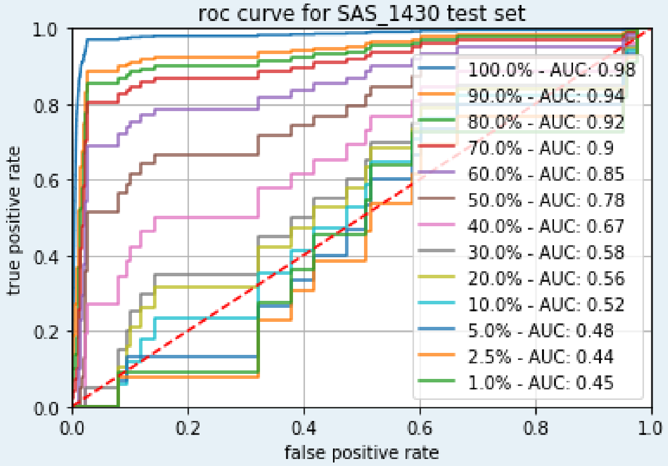


ROCs and AUCs

<table>
<tr><th></th><th>Full Data</th><th>HC Set</th></tr>
<tr><td>MCMC</td><td>


</td><td>


</td></tr>
<tr><td>VI</td><td>


</td><td>


</td></tr>
</table>

Each XX% line is ROC and AUC for pixels containing <XX% target abundance

E.g. 60% line is ROC for all pixels with <60% abundance