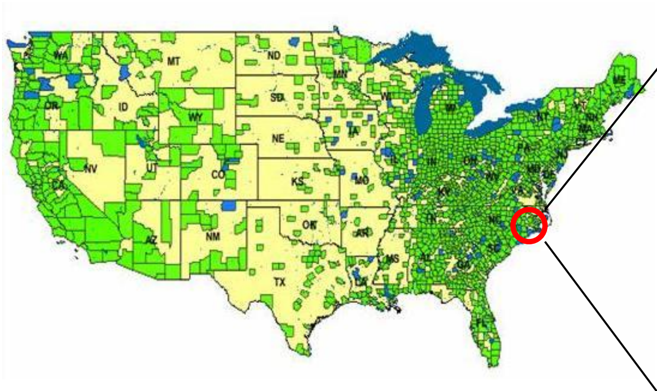
A Generalized Fast Subset Sums Framework for Bayesian Event Detection

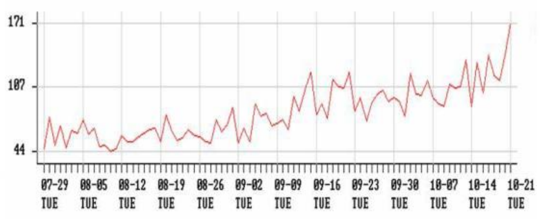
Daniel B. Neill
Event and Pattern Detection Laboratory
Carnegie Mellon University
neill @ cs.cmu.edu

This is joint work with **Kan Shao** and **Yandong Liu**, and was partially supported by NSF grants IIS-0916345, IIS-0911032, and IIS-0953330.

Multivariate event detection



Daily health data from thousands of hospitals and pharmacies nationwide.



Time series of counts $c_{i,m}^t$ for each zip code s_i for each data stream d_m.

d_1 = respiratory ED

d_2 = constitutional ED

d_3 = OTC cough/cold

d_4 = OTC anti-fever

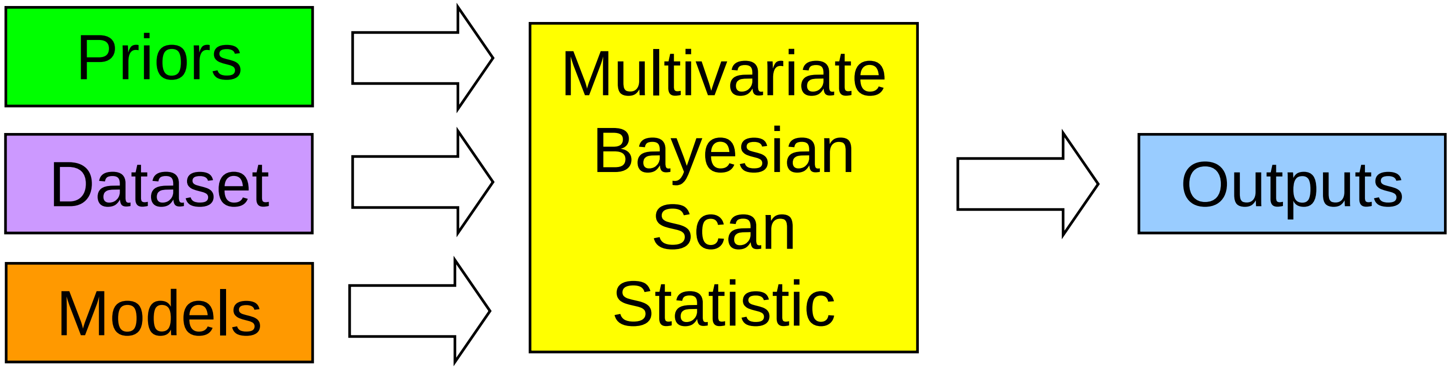
etc.

Given all of this nationwide health data on a daily basis, we want to obtain a complete situational awareness by integrating information from the multiple data streams.

More precisely, we have three main goals: to detect any emerging events (i.e. outbreaks of disease), characterize the type of event, and pinpoint the affected areas.

Overview of the MBSS method

(Neill and Cooper, *Machine Learning*, 2010)

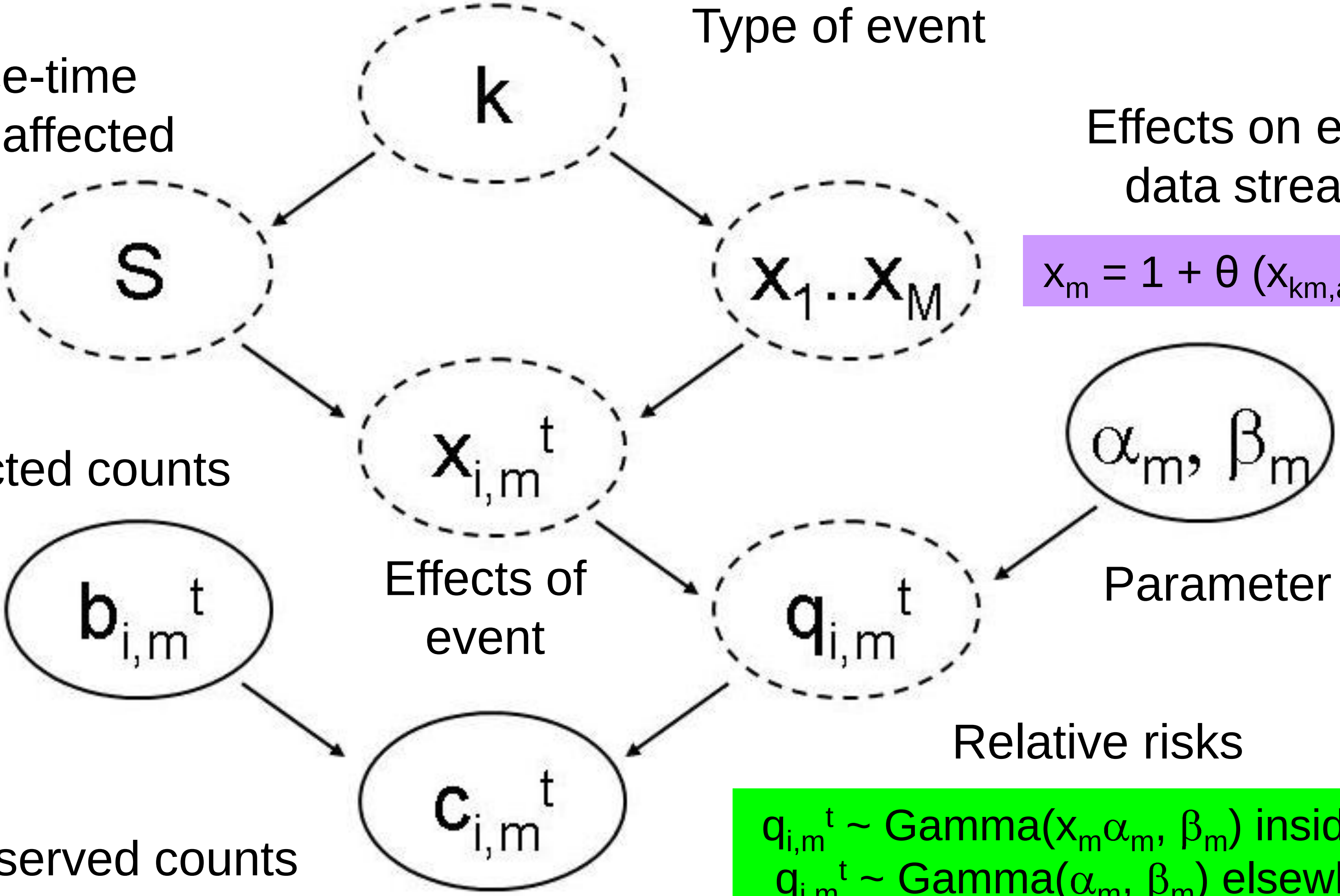


Given a set of event types E_k, a set of space-time regions S, and the multivariate dataset D, MBSS outputs the posterior probability $Pr(H_1(S, E_k) \mid D)$ of each type of event in each region, as well as the probability of no event, $Pr(H_0 \mid D)$.

We must provide the prior probability $Pr(H_1(S, E_k))$ of each event type E_k in each region S, as well as the prior probability of no event, $Pr(H_0)$.

MBSS uses Bayes' Theorem to combine the data likelihood given each hypothesis with the prior probability of that hypothesis: $Pr(H \mid D) = Pr(D \mid H) Pr(H) / Pr(D)$.

The Bayesian hierarchical model



Type of event

Space-time region affected

Effects on each data stream

$x_m = 1 + \theta\,(x_{km,avg} - 1)$

Expected counts

Effects of event

Parameter priors

Relative risks

$q_{i,m}^t \sim \mathrm{Gamma}(x_m \alpha_m, \beta_m)$ inside S,
$q_{i,m}^t \sim \mathrm{Gamma}(\alpha_m, \beta_m)$ elsewhere

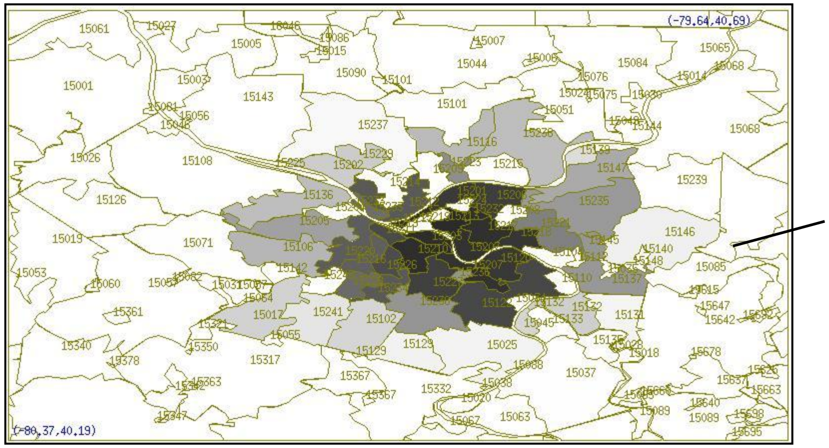
Observed counts

$c_{i,m}^t \sim \mathrm{Poisson}(q_{i,m}^t b_{i,m}^t)$

Interpretation and visualization

MBSS gives the total posterior probability of each event type E_k, and the distribution of this probability over space-time regions S.

Visualization: $\Pr(H_1(s_i, E_k)) = \sum \Pr(H_1(S, E_k))$ for all regions S containing location s_i.



Posterior probability map

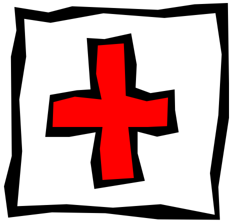
Total posterior probability of a respiratory outbreak in each Allegheny County zip code.

Darker shading = higher probability.

MBSS: advantages and limitations

MBSS can detect faster and more accurately by integrating multiple data streams.

MBSS can model and differentiate between multiple potential causes of an event.



MBSS assumes a uniform prior for circular regions and zero prior for non-circular regions, resulting in low power for **elongated** or **irregular** clusters.

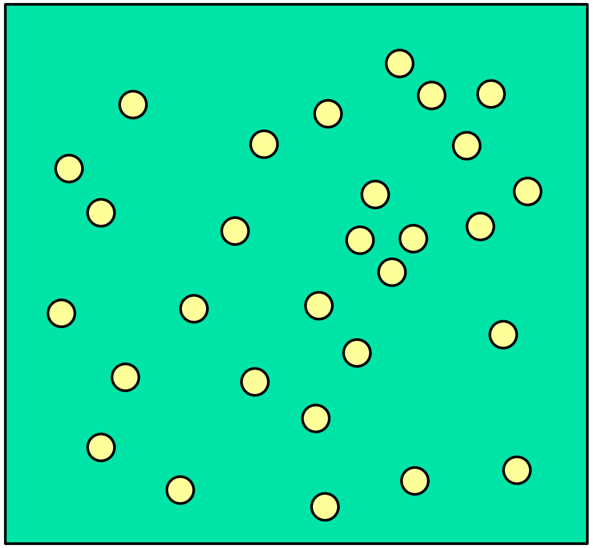
There are too many subsets of the data (2^N) to compute likelihoods for all of them!

How can we extend MBSS to **efficiently** detect irregular clusters?

Generalized Fast Subset Sums

We define a non-uniform prior $Pr(H_1(S, E_k))$ over all 2^N subsets of the data.

This prior has hierarchical structure:

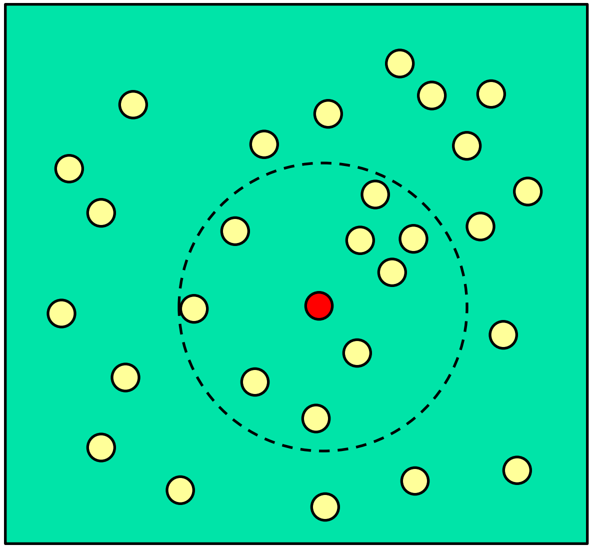


Generalized Fast Subset Sums

We define a non-uniform prior $\Pr(H_1(S, E_k))$ over all 2^N subsets of the data.

This prior has hierarchical structure:

1. Choose **center location s_c** from $\{s_1 \ldots s_N\}$, given multinomial $\Pr(s_i)$.
2. Choose **neighborhood size n** from $\{1 \ldots n_{max}\}$, given multinomial $\Pr(n)$.

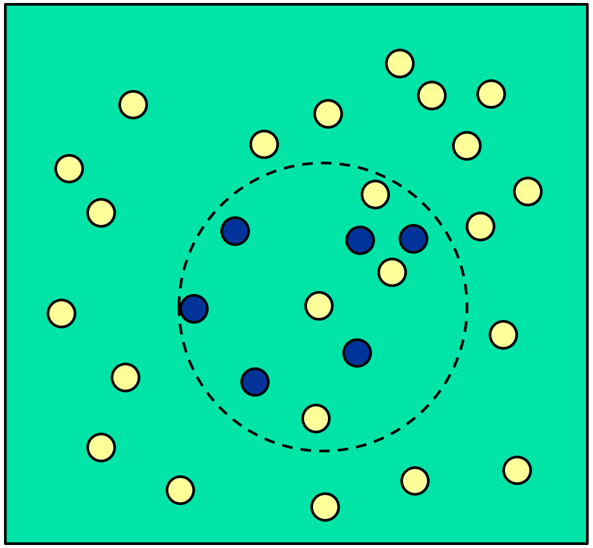


Generalized Fast Subset Sums

We define a non-uniform prior $Pr(H_1(S, E_k))$ over all 2^N subsets of the data.

This prior has hierarchical structure:

1. Choose **center location s_c** from $\{s_1 \ldots s_N\}$, given multinomial $Pr(s_i)$.
2. Choose **neighborhood size n** from $\{1 \ldots n_{max}\}$, given multinomial $Pr(n)$.
3. For each $s_i \in S_{cn}$, include s_i in S with probability p, for a fixed $0 < p \leq 1$.



This prior distribution has non-zero prior probabilities for any given subset S, but more compact clusters have larger priors.

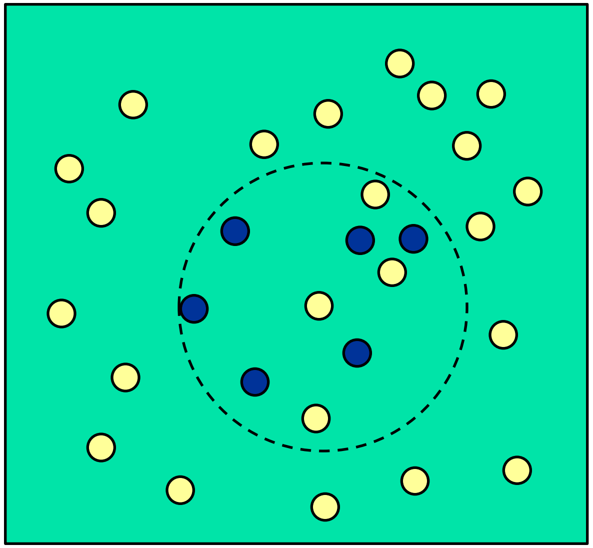
Parameter p controls the sparsity of detected clusters. Large p = compact clusters. Small p = dispersed clusters.

Generalized Fast Subset Sums

We define a non-uniform prior $\Pr(H_1(S, E_k))$ over all 2^N subsets of the data.

This prior has hierarchical structure:

1. Choose **center location s_c** from $\{s_1 \ldots s_N\}$, given multinomial $\Pr(s_i)$.
2. Choose **neighborhood size n** from $\{1 \ldots n_{max}\}$, given multinomial $\Pr(n)$.
3. For each $s_i \in S_{cn}$, include s_i in S with probability p, for a fixed $0 < p \leq 1$.



p = 0.5 corresponds to the original Fast Subset Sums approach described in (Neill, *Stat. Med.*, 2011), assuming that all subsets are equally likely given the neighborhood.

p = 1 corresponds to MBSS, searching circular regions only.

Generalized **Fast** Subset Sums

Naïve computation of posterior probabilities using this prior requires summing over an exponential number of regions, which is infeasible.

However, the total posterior probability of an outbreak, $\Pr(H_1(E_k) \mid D)$, and the posterior probability map, $\Pr(H_1(s_i, E_k) \mid D)$, can be calculated efficiently **without** computing the probability of each region S.

In the original MBSS method, the **likelihood ratio** of spatial region S for a given event type E_k and event severity θ can be found by multiplying the likelihood ratios $LR(s_i \mid E_k, \theta)$ for all locations s_i in S.

In GFSS, the **average likelihood ratio** of the 2^n subsets for a given center s_c and neighborhood size n can be found by multiplying the quantities $(p \times LR(s_i \mid E_k, \theta) + (1-p))$ for all locations s_i in S.

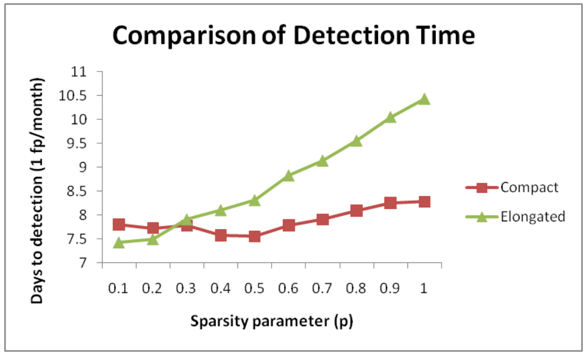
Since the prior is uniform for a given center and neighborhood, we can compute the posteriors for each s_c and n, and marginalize over them.

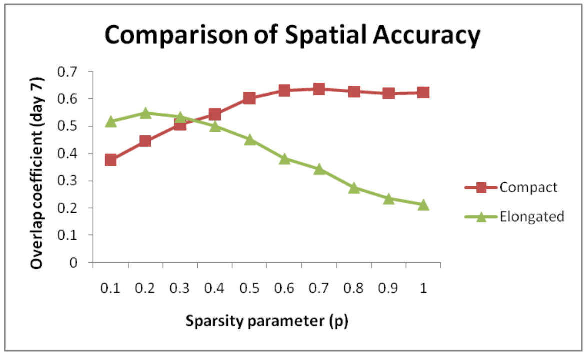
Preliminary results

We injected simulated disease outbreaks into two streams of Emergency Department data from 97 Allegheny County zip codes.

Results were computed for ten different outbreak shapes, including compact, elongated, and irregularly-shaped (200 injects of each type).

Runtime of GFSS was extremely fast, computing the posterior probability map for each day of data in less than nine seconds.





Smaller values of the sparsity parameter p achieve higher detection performance for elongated clusters, and larger p for compact clusters.

Learning the sparsity parameter

We demonstrate that the sparsity parameter can be **learned** from a set of labeled training examples $S_1 \ldots S_J$. For each S_j, we are given the affected region S, but not the values of the latent parameters (center location s_c, neighborhood size n, and sparsity parameter p).

To learn the distribution of p, we must marginalize over s_c and n:

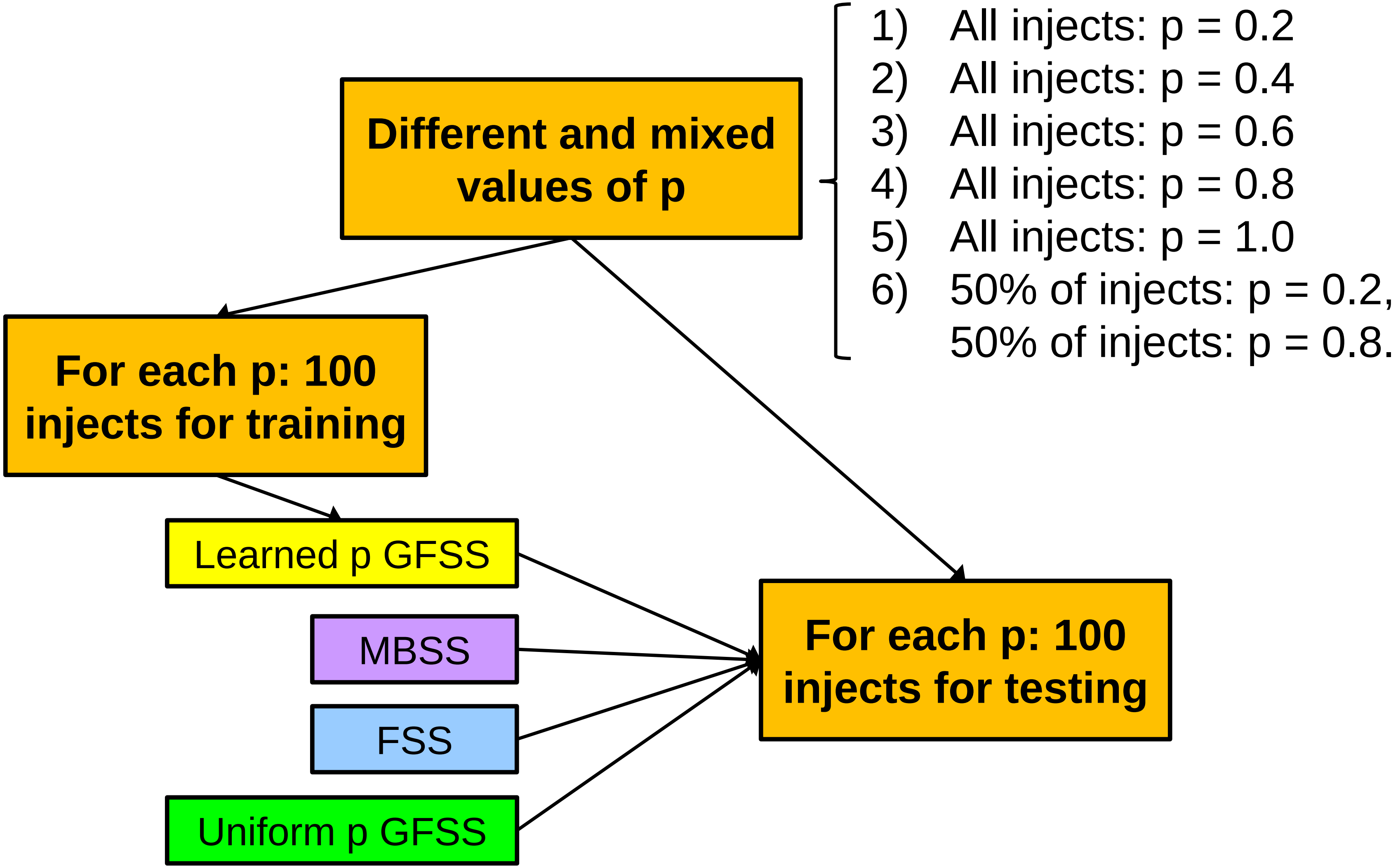
$$\Pr(S_j \mid p) = \sum_{s_c} \sum_n \Pr(S_j \mid p, s_c, n) \Pr(s_c) \Pr(n)$$

$$\Pr(S_j \mid p, s_c, n) = p^{|S_j|} (1 - p)^{n - |S_j|} 1\{S_j \subseteq S_{cn}\}$$

We assume that the p_j for each S_j is drawn from a multinomial distribution θ over {0.1, 0.2, …, 1.0}, assuming a Dirichlet prior on θ.

$$\theta_k = \Pr(p = 0.1 \times k) = \frac{0.1 + \sum_{j=1..J} \frac{\Pr(S_j \mid p = 0.1 \times k) \Pr(p = 0.1 \times k)}{\sum_{k=1..10} \Pr(S_j \mid p = 0.1 \times k) \Pr(p = 0.1 \times k)}}{1 + J}$$

Evaluation framework



Different and mixed values of p

1) All injects: p = 0.2
2) All injects: p = 0.4
3) All injects: p = 0.6
4) All injects: p = 0.8
5) All injects: p = 1.0
6) 50% of injects: p = 0.2,
   50% of injects: p = 0.8.

For each p: 100 injects for training

Learned p GFSS

MBSS

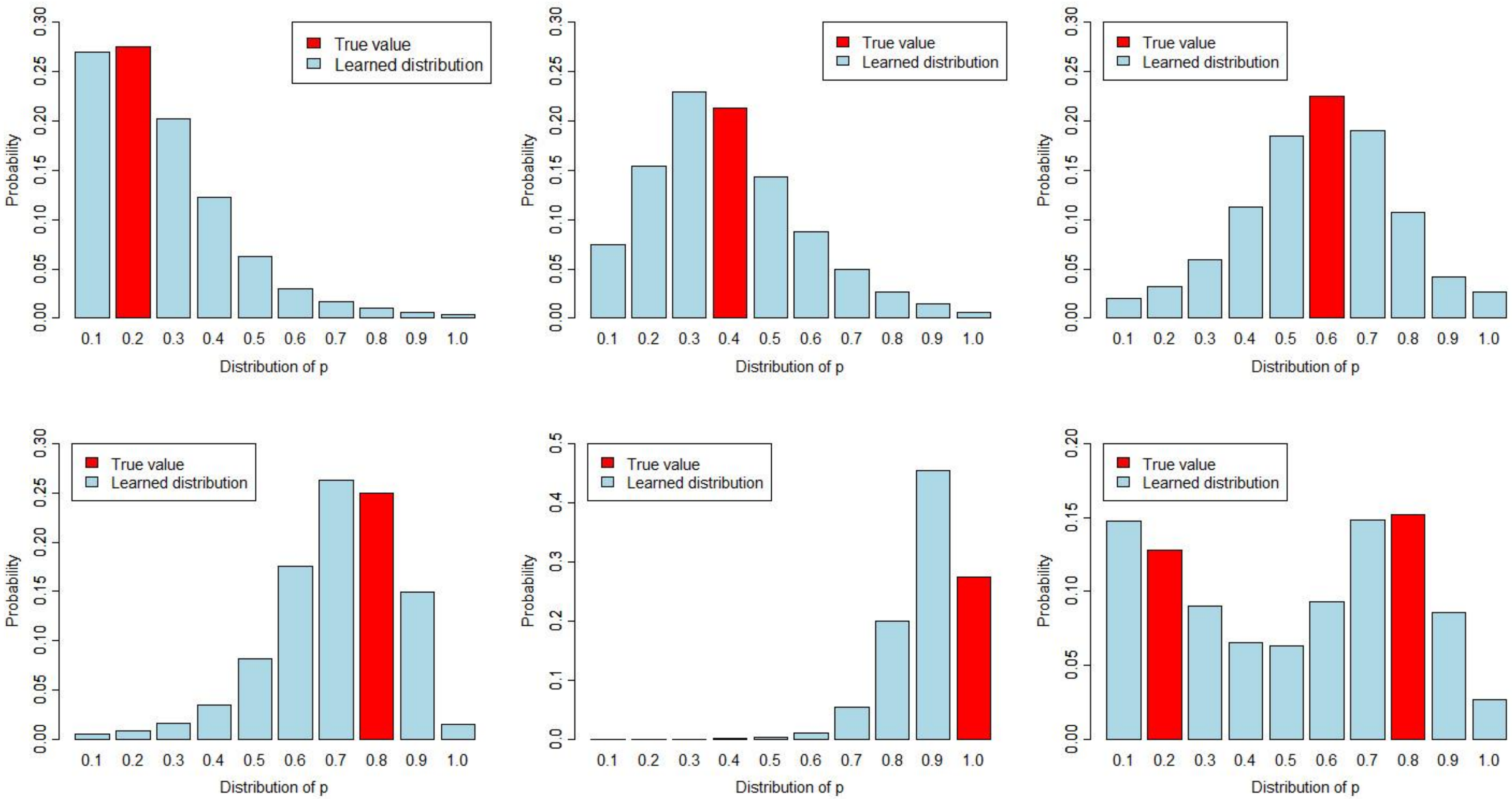
FSS

Uniform p GFSS

For each p: 100 injects for testing

Results: accuracy of learned model

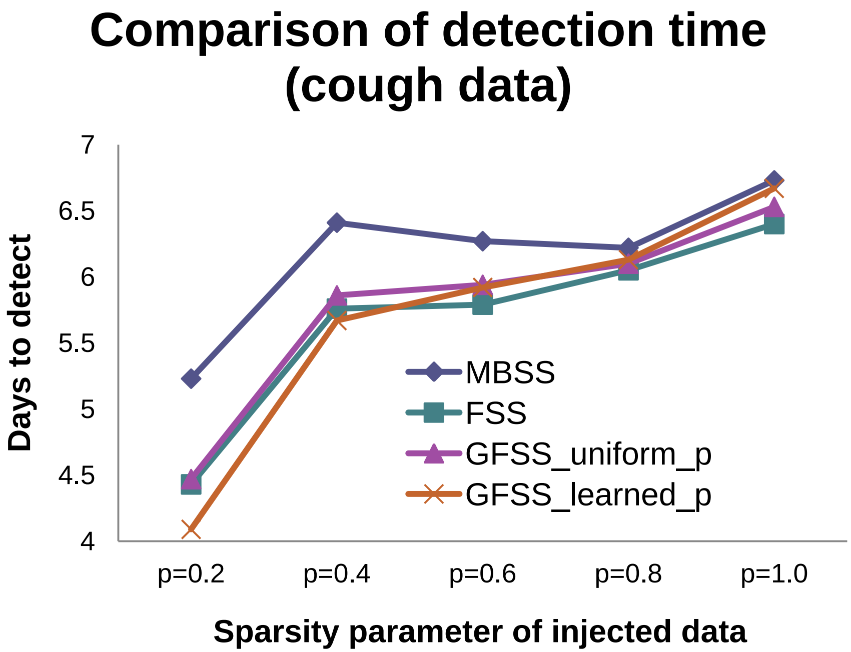
(100 injects, cough data)

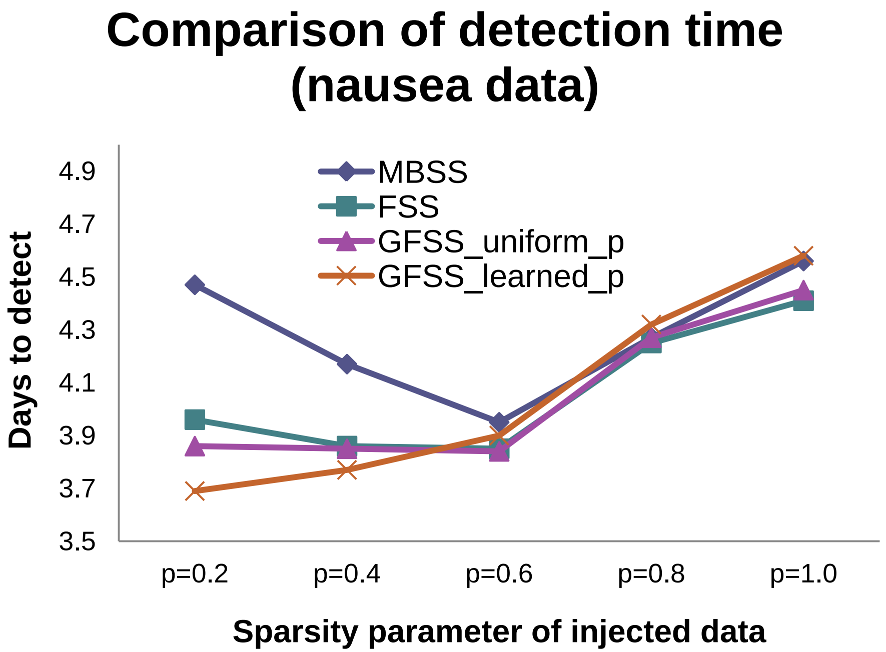


GFSS was able to estimate the true value(s) of p. Results were very similar for nausea outbreaks and for as few as 25 injected outbreaks.

Results: detection power

We compared the average time to outbreak detection for the Learned-p GFSS, Uniform-p GFSS, MBSS, and FSS methods, at a fixed false positive rate of 1/month.


Comparison of detection time (cough data)

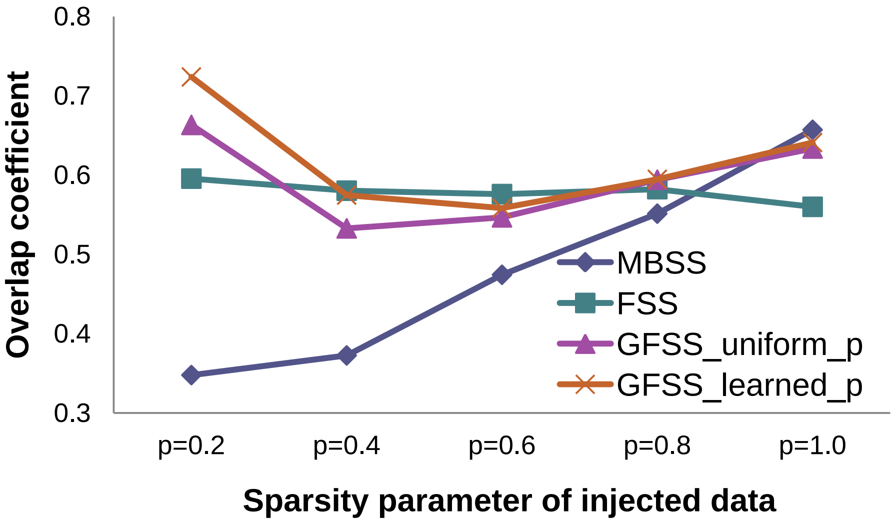

Comparison of detection time (nausea data)

When the value of p is small, corresponding to an elongated or irregular outbreak region, GFSS with learned p is able to detect substantially earlier than the other methods.

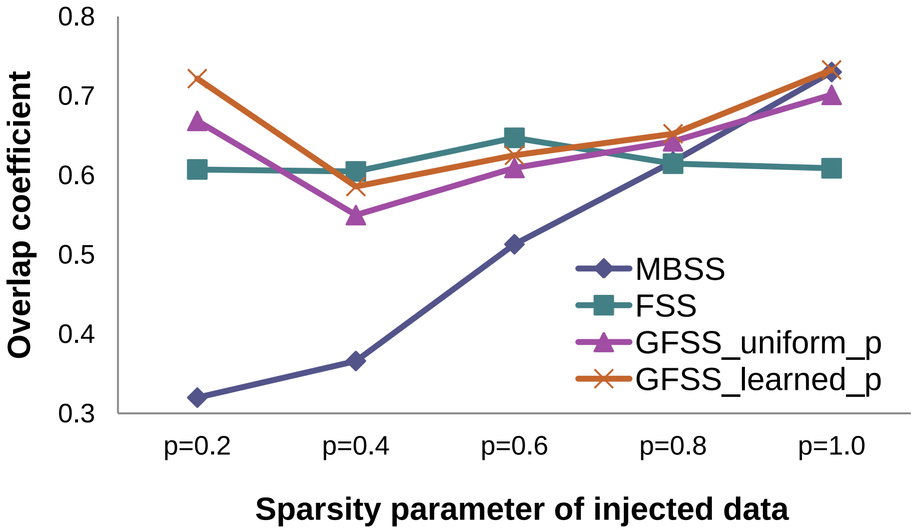
Results: spatial accuracy

We compared the spatial accuracy (average overlap coefficient between true and detected clusters at day 7 of the outbreak) for Learned-p GFSS, Uniform-p GFSS, MBSS, and FSS.

Comparison of spatial accuracy (cough data)



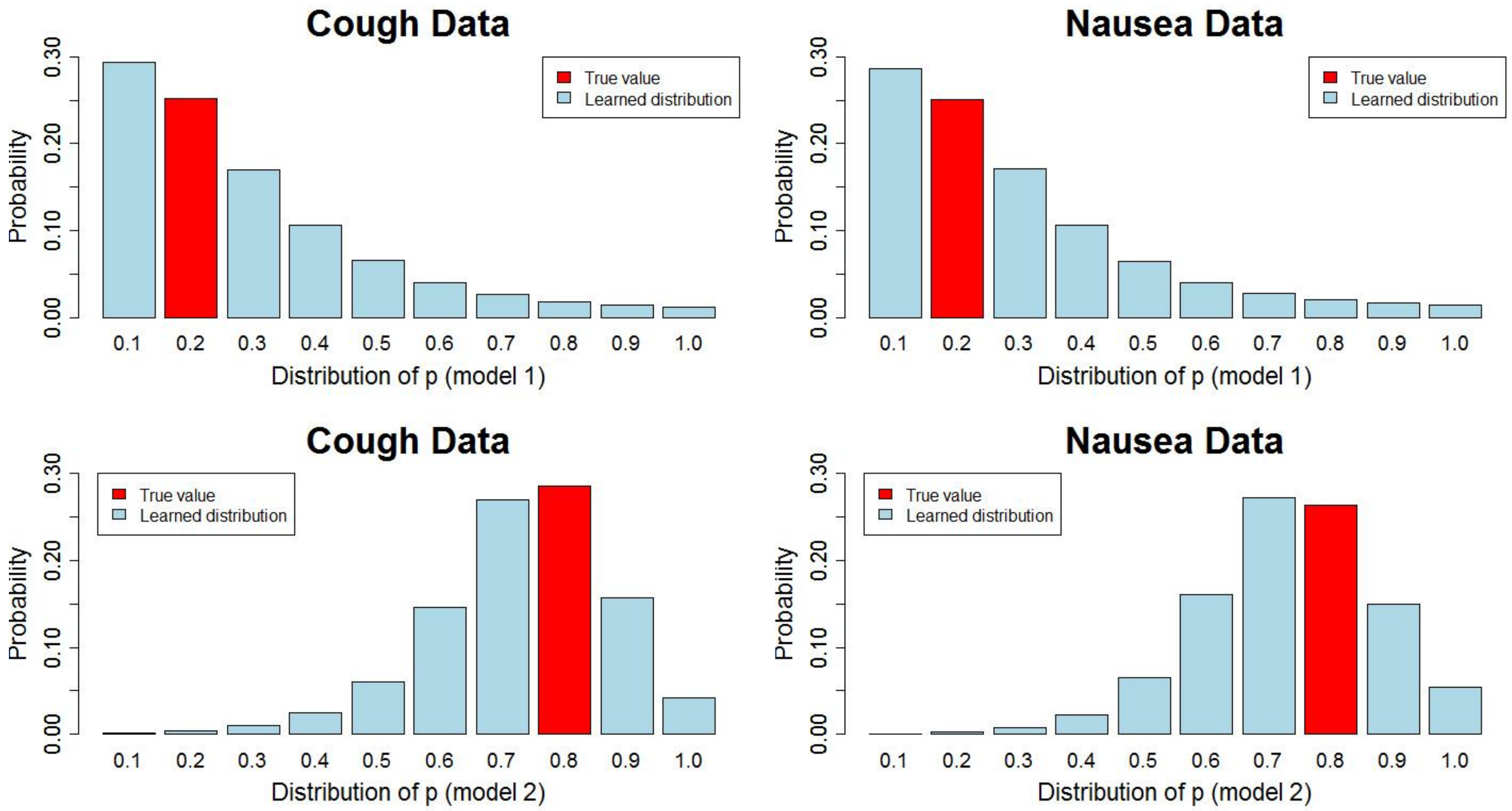
Comparison of spatial accuracy (nausea data)



GFSS with learned p achieves high spatial accuracy across the entire range of p values.

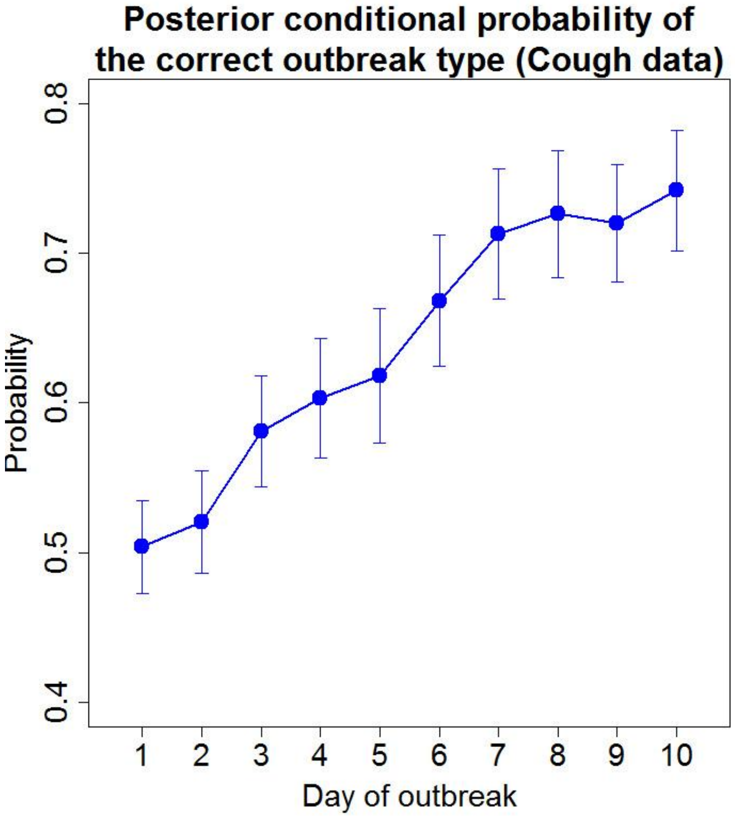
Results: distinguishing outbreak types

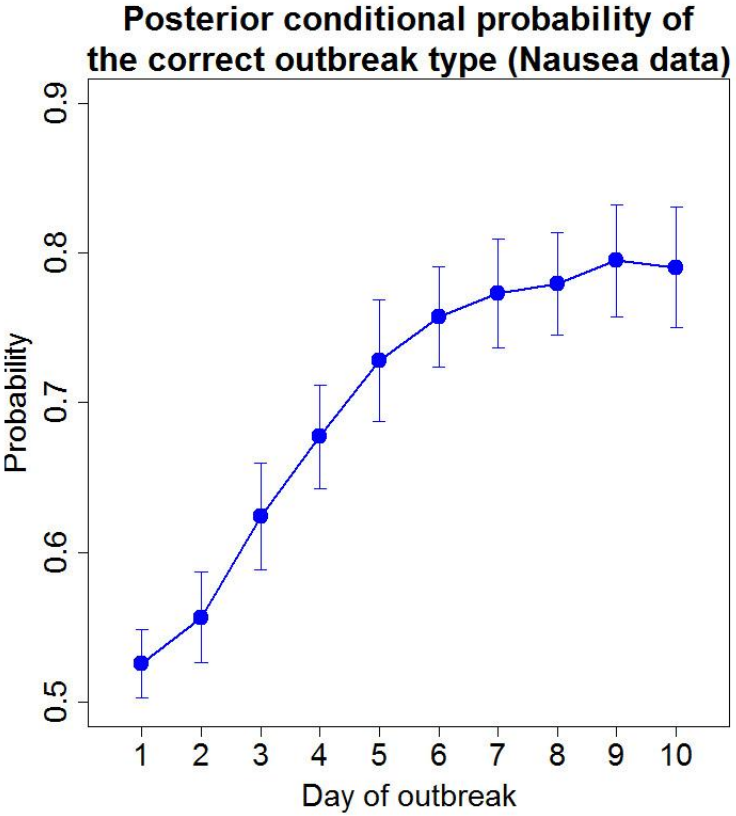
We used the mixture outbreak data to evaluate the ability of GFSS to learn and distinguish between two outbreak types with different values of the sparsity parameter (p = 0.2 and p = 0.8).



Results: distinguishing outbreak types

We used the mixture outbreak data to evaluate the ability of GFSS to learn and distinguish between two outbreak types with different values of the sparsity parameter (p = 0.2 and p = 0.8).

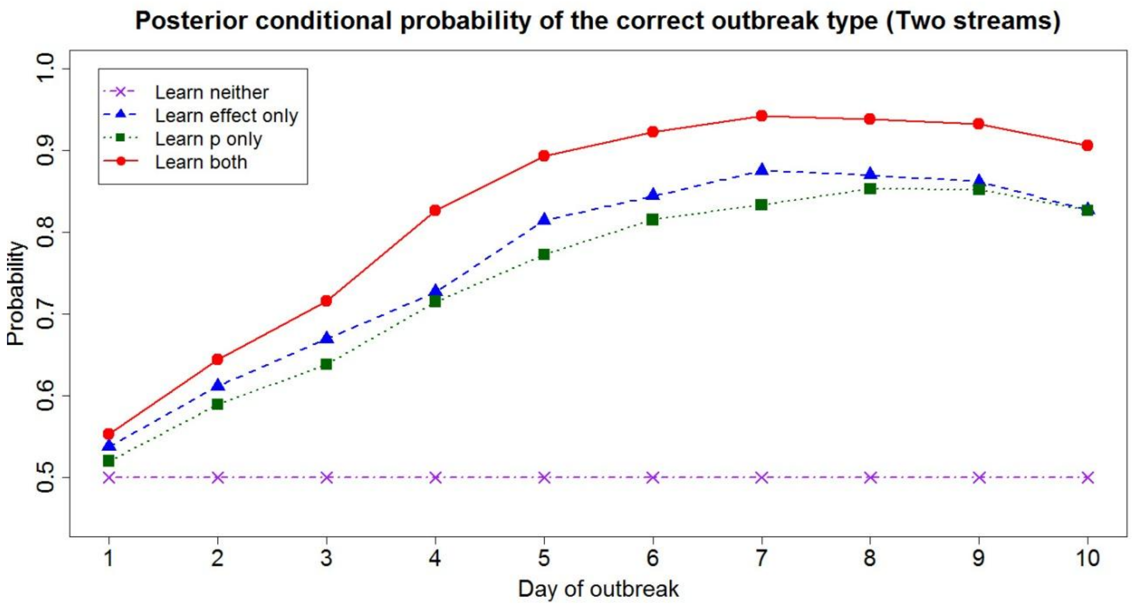




Results: distinguishing outbreak types

We used the mixture outbreak data to evaluate the ability of GFSS to learn and distinguish between two outbreak types with different values of the sparsity parameter (p = 0.2 and p = 0.8).

When the two outbreaks also had different effects on the two monitored data streams (cough and nausea), learning both the relative effects and the sparsity further improved detection.



Conclusions

GFSS shares the essential advantages of MBSS: it can integrate information from **multiple data streams**, and can accurately distinguish between **multiple outbreak types**.

As compared to the MBSS method, GFSS substantially improves **accuracy** and **timeliness** of detection for elongated or irregular clusters, with similar performance for compact clusters.

While a naïve computation over the exponentially many subsets of the data is computationally infeasible, GFSS can **efficiently** and **exactly** compute the posterior probability map.

We can **learn** the distribution of the sparsity parameter p for multiple event types using a small amount of labeled training data.

Conclusions

Learning the distribution of the sparsity parameter not only improves detection power, but enables us to accurately differentiate between multiple, similar types of outbreak.

We are currently extending GFSS to simultaneously learn the distributions over the center location, neighborhood size, and the sparsity parameter p, using an EM-based approach.

In future work, we will also extend GFSS to the case of partially labeled training data, when only a small subset of affected locations are identified for each labeled event.

Thanks for listening!
Any questions?