Forecasting Volcanic Activity Using An Event Tree Analysis System and Logistic Regression

William N. Junek
Central Florida Remote Sensing Laboratory, University of Central Florida

Doctoral Dissertation Defense
March 19, 2012

Presentation Outline

- Research Objectives
- Motivation
- Volcanic Processes
- Event Tree System
- Results
- Conclusions
- Future Work

Primary Research Objectives

- Develop a process for forecasting short term volcanic activity using monitoring data, source modeling results, and historic observations , which is easy to use, transportable, and can be updated as new information becomes available.

- Derive empirical statistical models via logistic regression using a dataset that is comprised of monitoring data, source modeling results, and historic eruption information acquired from collection of analog volcanoes.

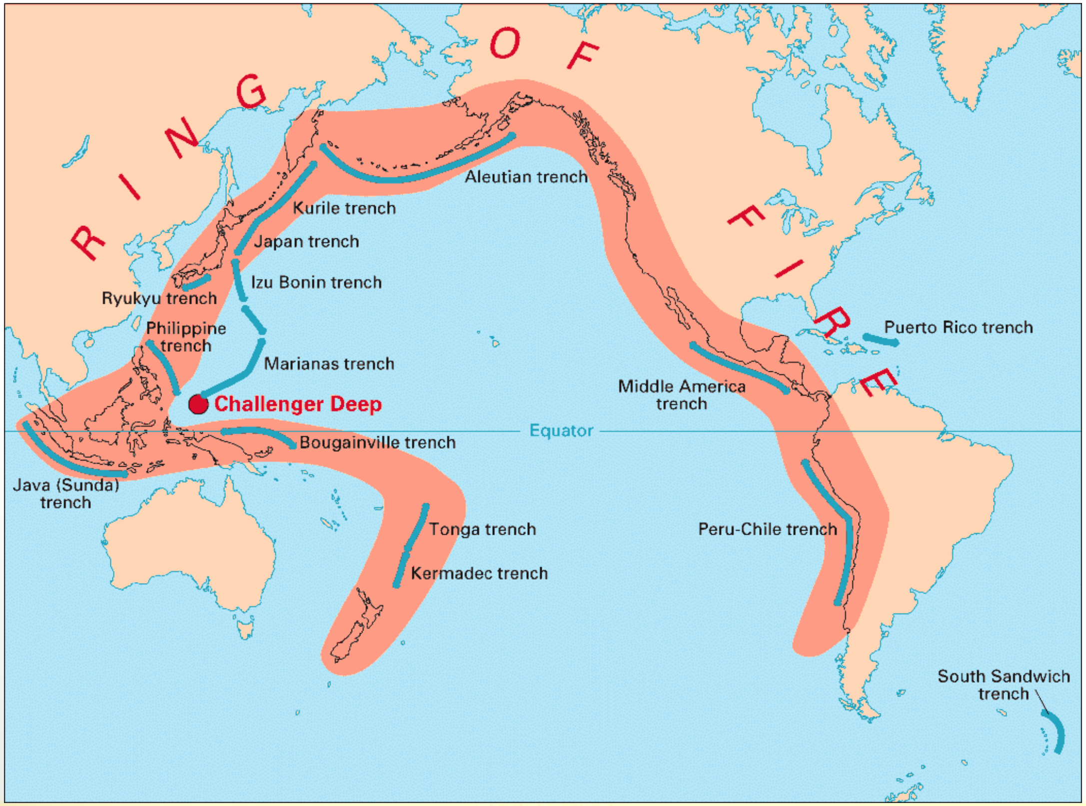
- Estimate probable volcanic vent locations using a two dimensional spatial probability density function derived from a combination of source modeling results and monitoring data.

- Produce hazard assessments in terms of the USGS ground-based color code system.

Motivation

- There are 169 geologically active volcanic centers in the United States.

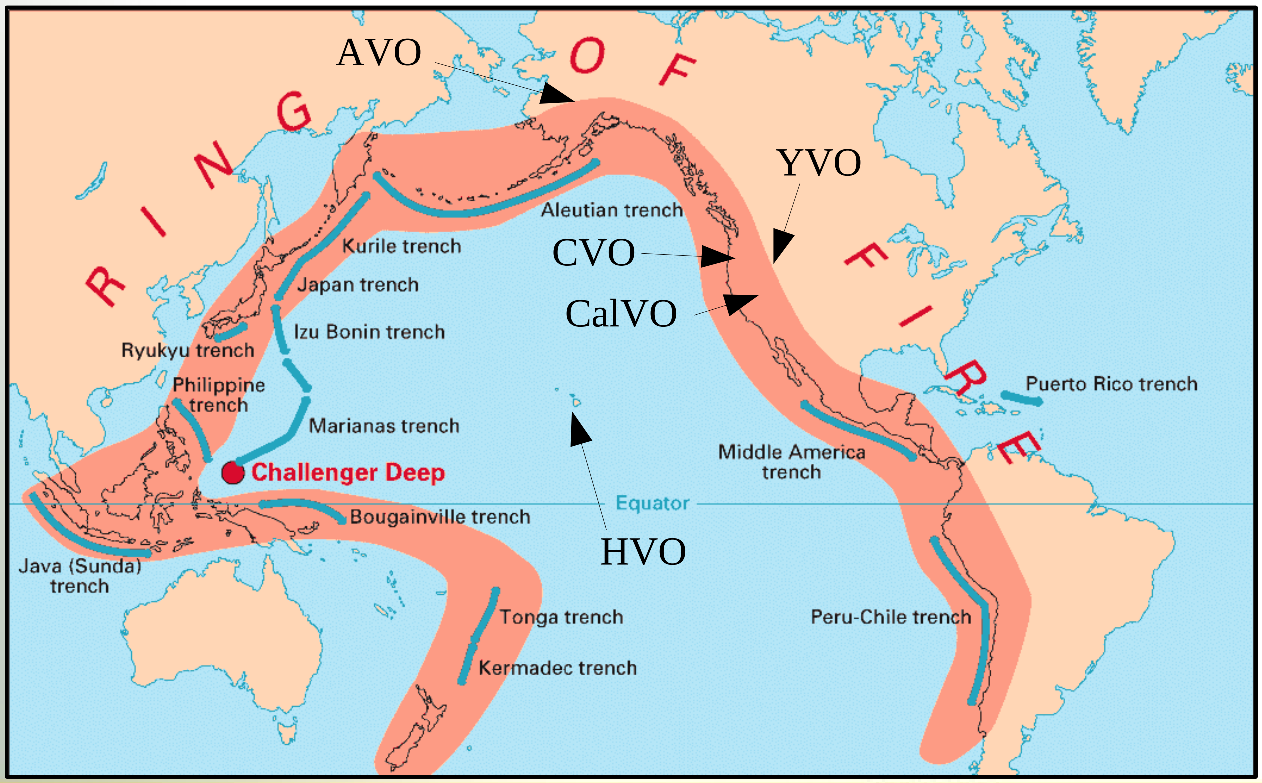
- There is a significant need to monitor volcanic activity within the United States to ensure major population centers can be evacuated and air traffic diverted in the event of an eruption.

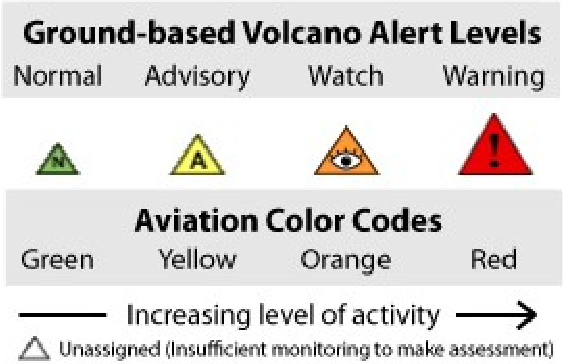


Motivation

- According to the Stafford Act (Public Law 93-288), the United States Geological Survey (USGS) is responsible for issuing timely warnings of volcanic eruptions to federal emergency management agencies.

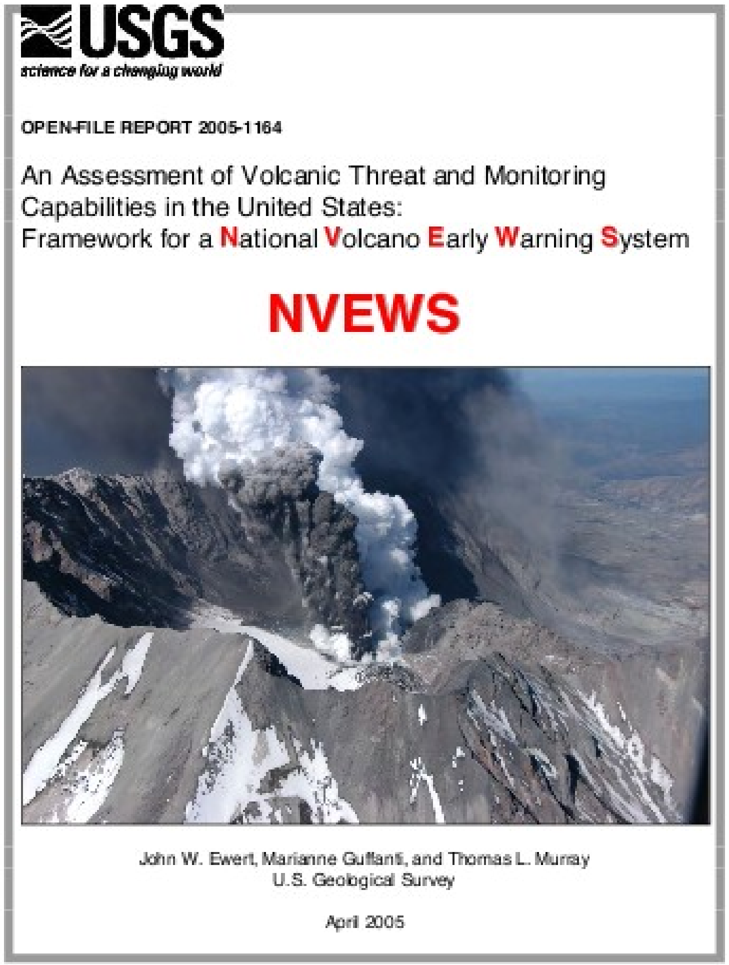
- This responsibility is carried out by a series of volcano observatories that are tasked to monitor their distinct volcano-tectonic region.





Images obtained from the USGS Volcanic Hazard Program Web page: http://volcanoes.usgs.gov/

Motivation


USGS
science for a changing world

OPEN-FILE REPORT 2005-1164

An Assessment of Volcanic Threat and Monitoring Capabilities in the United States:
Framework for a National Volcano Early Warning System

NVEWS

John W. Ewert, Marianne Guffanti, and Thomas L. Murray
U.S. Geological Survey

April 2005

- Monitoring techniques developed since the eruption of Mount Saint Helens are currently being applied on an ad hoc basis to volcanoes exhibiting heightened activity.

- The Consortium of U.S. Volcano Observatories (CUSVO) is working to solve this problem.

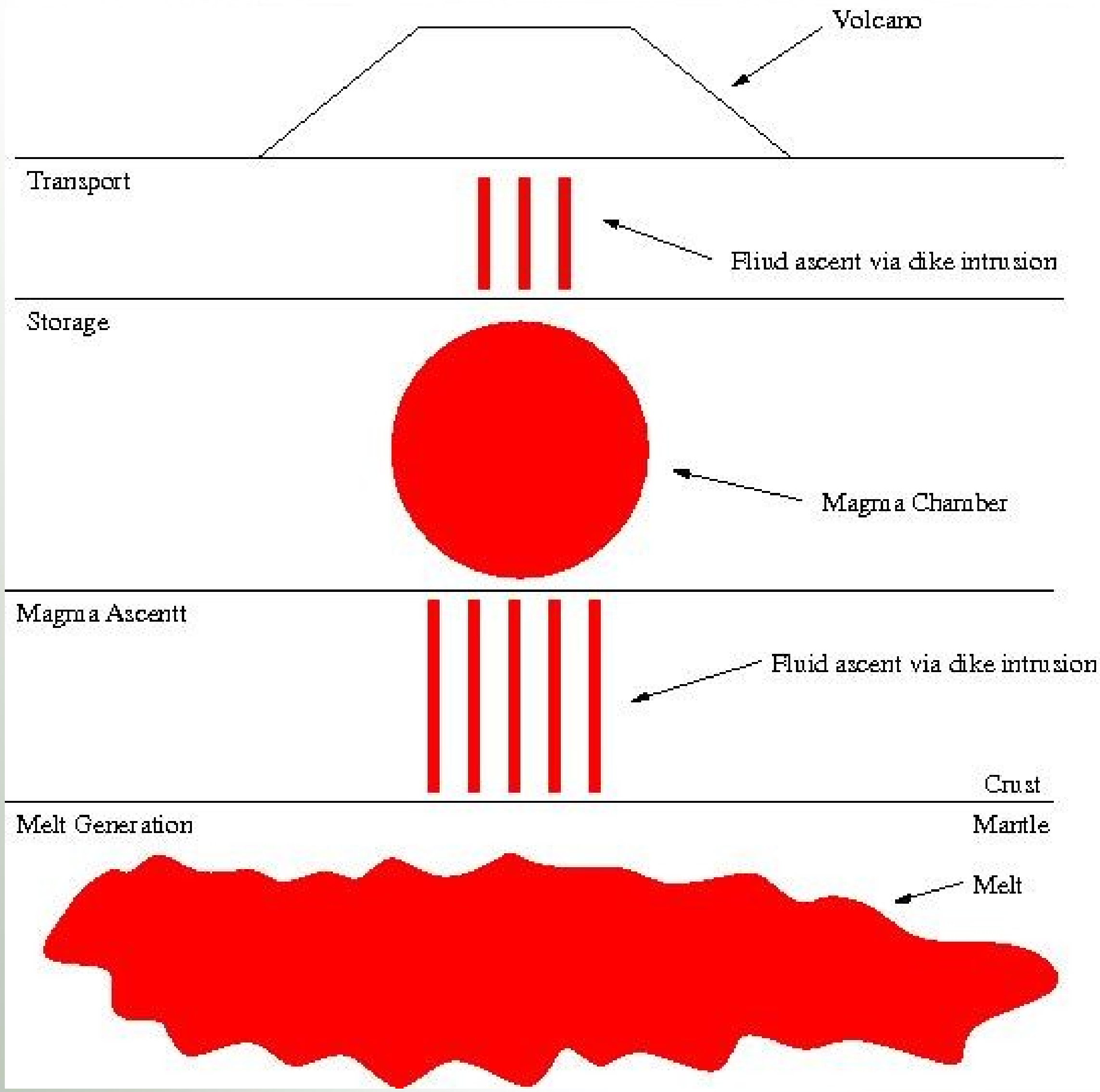
- The National Volcano Early Warning System (NVEWS) was announced in 2005.

- The NVEWS will focus on monitoring all high and moderate risk volcanoes in the U.S.

- The System will include a centralized “Watch Office” that will collect and analyze monitoring data.

- This report states: “Monitoring without research into the driving physico-chemical processes becomes mechanistic pattern recognition, an inadequate approach to phenomena as complex as volcanoes."

Volcanic Processes



Eruption Mechanics:

- Transport - The process that delivers magma from the storage area to the surface.

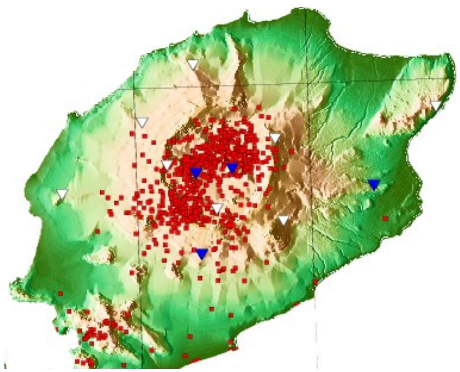
- Storage - A shallow chamber that stores magma transported from underlying melts.

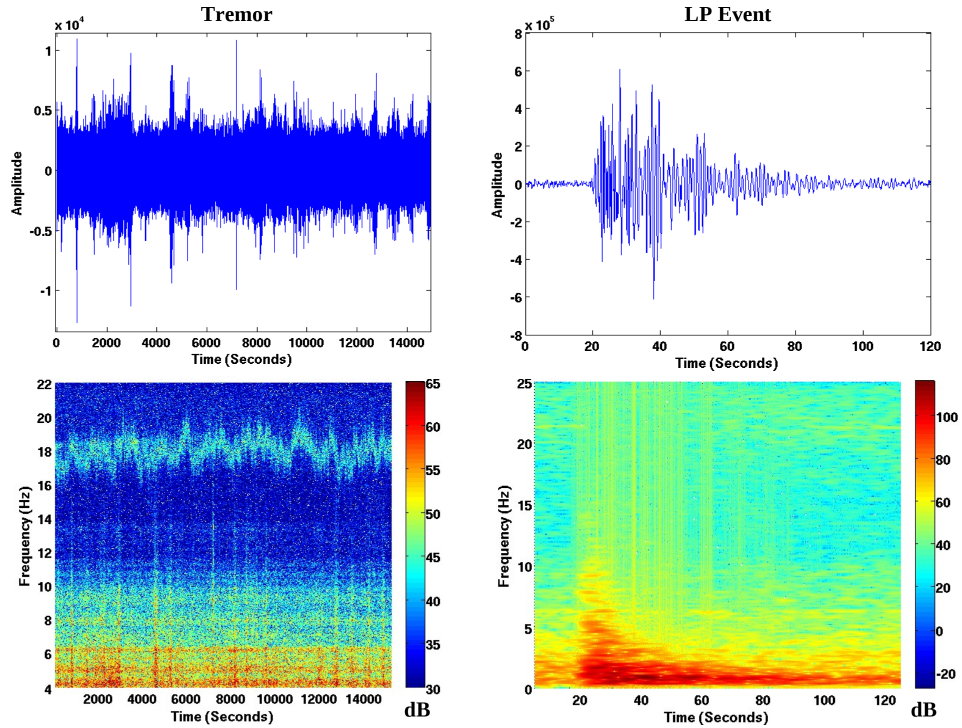
- Magma Ascent - The movement of magma through a series of dike intrusions that exist between the melt layer and an intermediate storage area or the surface.

- Melt Generation - The process of magma production occurs deep beneath the Earth's crust.

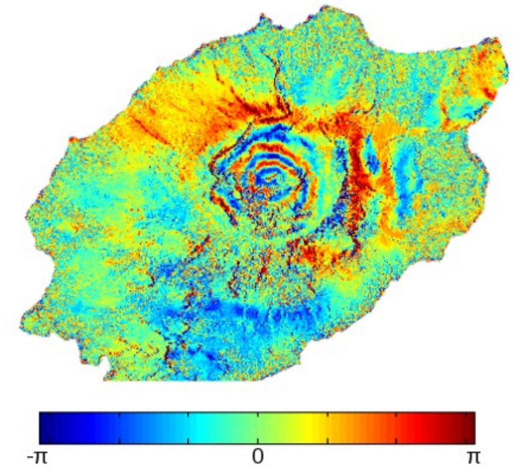
Monitoring Volcanic Processes

Seismic

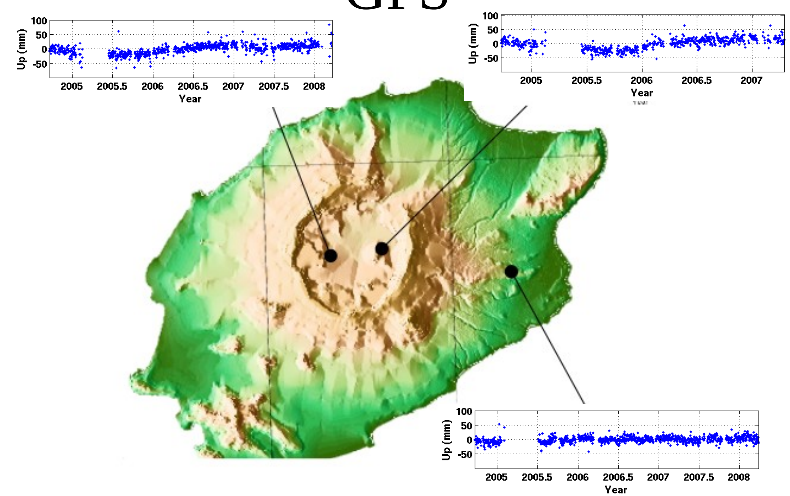




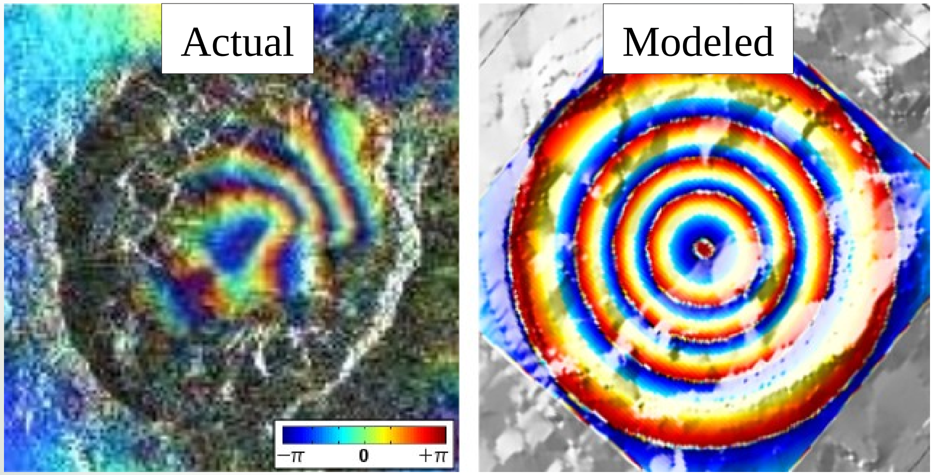
InSAR

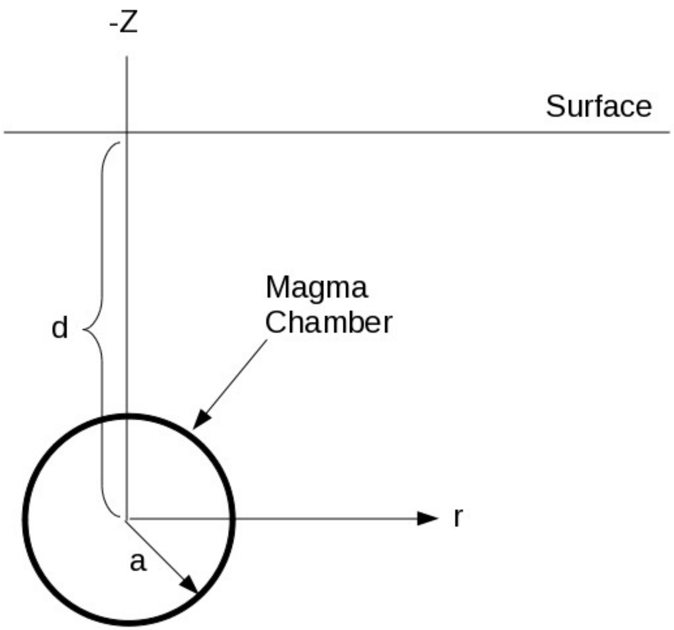


GPS



Modeling Volcanic Processes





Mogi Source

$$\Delta h(r) = C\frac{d}{(r^2+d^2)^{3/2}} \qquad \Delta r(r) = C\frac{r}{(r^2+d^2)^{3/2}} \qquad C = \frac{3a^3 P}{4\mu}$$

Where: P=pressure change and μ=Lame's constant

Combining Multidisciplinary Data

- By combining empirical and synthetic data an integrated view of the magma ascent process can be created.

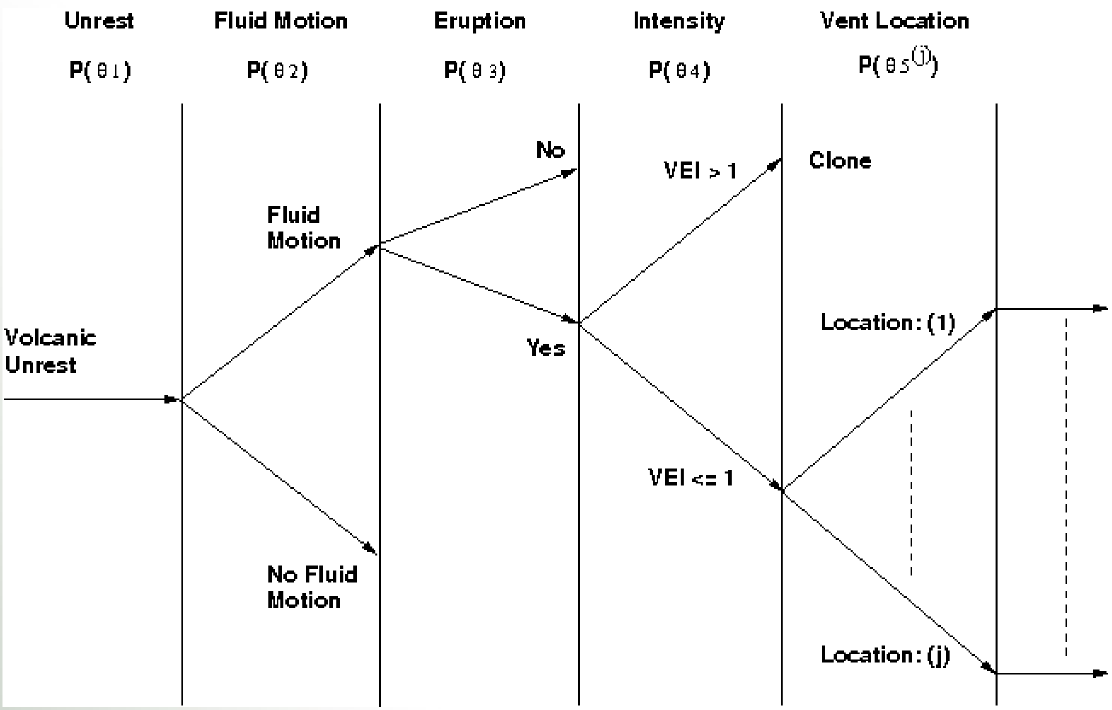
- This allows for the development of a physical modeled based pattern recognition system that uses monitoring data, modeling results, and historic information to forecast various types of volcanic activity.

  - **Prediction**: A statement that a particular event of a certain size will occur at a certain location and time.

  - **Forecast**: A statement of the probability that a particular event of a certain size may occur in a certain area and time frame.

Volcanic activity is comprised of a complex combination of geophysical that makes predicting the onset of an eruption impossible. Probabilistic forecasting techniques can be used to assist in the assessment of volcanic hazards and aid civil authorities in planning a response to a developing volcanic crisis that may require immediate action (e.g., evacuation)

Event Tree System



$$P(\theta_n) = P(n|n-1) = \frac{P(n-1|n)\,P(n)}{P(n-1|n)\,P(n) + P(n-1|n')\,P(n')}$$

Bayes Theorem

Decision Nodes:

- Unrest [$P(1)$] - Does the geophysical activity at the selected volcano exceed a predetermined threshold within the specified sampling window?
- Fluid Motion [$P(2|1)$] - Is the unrest the result of magma motion?
- Eruption [$P(3|2)$] - Does the detected fluid motion have the potential to reach the surface and cause an eruption?
- Intensity [$P(4|3)$] - What is the likely eruption intensity?
- Vent Location [$P(5|4)$] - Where is the eruption likely to occur?

Event Tree System

Probability [P(θ_1)]:

- Unrest: Set P(1) = 1.00 when the summation of a set of explanatory variables (**Xn**) exceeds 0.00.

Prior Probabilities [P(n)]:

- Fluid Motion: P(2) = P(*Fluid Motion=1*|***Xn***)
- Eruption: P(3) = P(*Eruption=1*|***Xn***)
- Intensity: P(4) = P(*Intensity>1*|***Xn***)

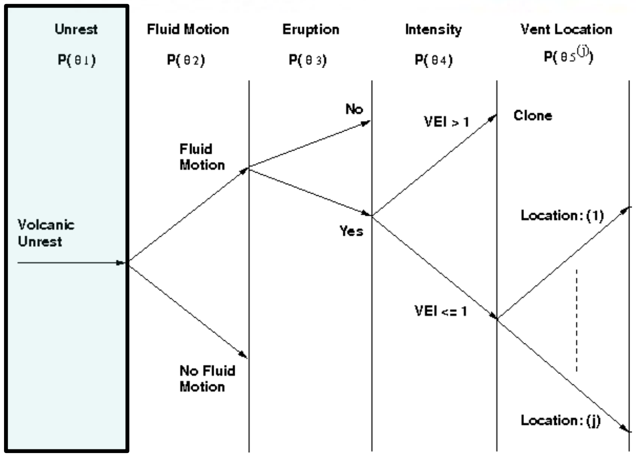
The probability the event in question will occur, given the values of a collection explanatory variables (i.e., empirical, modeled, and historic data), **Xn**.

- Vent Location: P(5^j) = P(*Vent Formation* ^{j}*=1*|*Xn* ^{j}), The probability of vent formation in the jth location given the values of a collection explanatory variables (empirical, modeled), **Xn**.

$$P(\theta_n)=P(n|n-1)=\frac{P(n-1|n)P(n)}{P(n-1|n)P(n)+P(n-1|n')P(n')}$$

Bayes Theorem

Node 1: Detecting Volcanic Unrest



- The value of ν (unrest severity) is the summation of a collection of explanatory variables.

$$\nu = \sum_{n=1}^{N} \beta_n X_n \quad \Longrightarrow \quad \nu = \beta(X_{sr} + X_{df} + X_{lm} + X_{md})$$

- Detection of unrest is declared (X_n=1) when monitored activity exceeds the outlier threshold.

- All variables carry identical weights (β_n = 0.25) and sum to 1.

$$P(\theta_1) = P(1) = \begin{cases} 1, \text{ if } \nu > 0 \\ 0, \text{ if } \nu = 0 \end{cases}$$

Explanatory Variables

Variables	Description	Threshold	Value
X_{sr}	Seismicity Rate	Outlier	0/1
X_{df}	Surface Deformation	Outlier	0/1
X_{lm}	Large Magnitude	Outlier	0/1
X_{df}	Modeling	1	0/1

Nodes 2 - 4

Empirical CDF via Logistic Regression:

- Logistic function output is bound between 0 and 1, while its input (z) can range between +/- infinity.

$$f(z) = \frac{1}{1+e^{-z}}$$

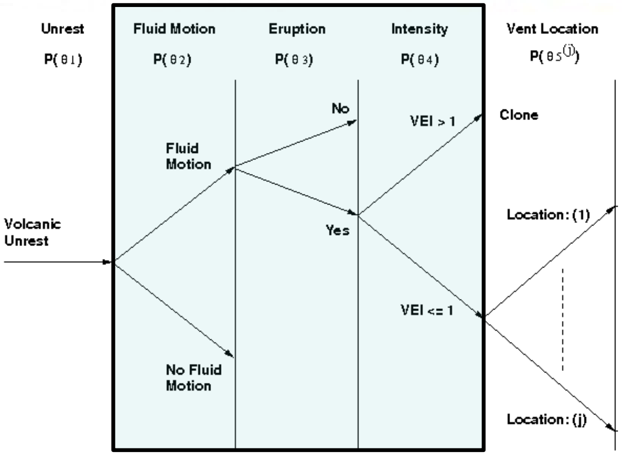
- z is the linear sum of a set of independent explanatory variables (Logistic Model).

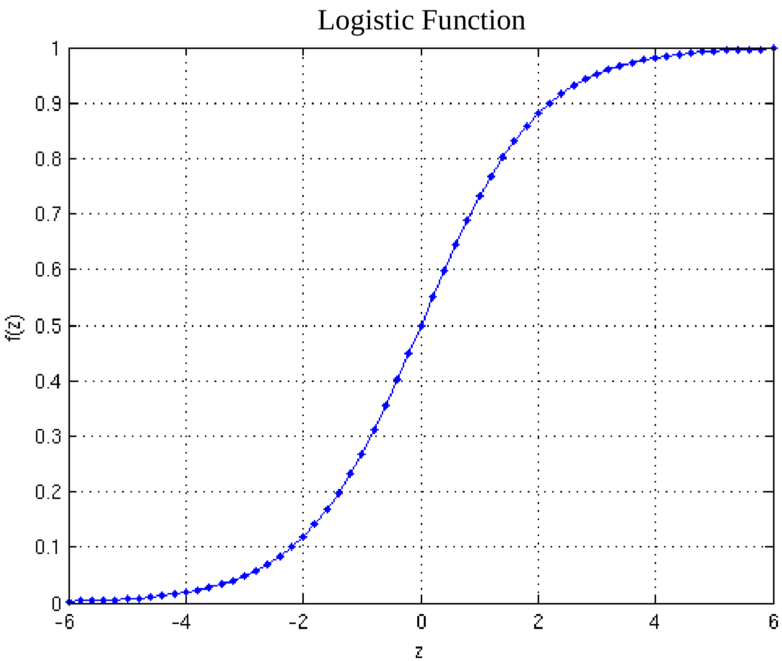
$$z = \beta_0 + \sum_{n=1}^{N} \beta_n X_n$$

- Logistic coefficients are computed using a generalized linear model and logit linking function (assumes binominal response variable)

- The conditional probability is defined as:

$$P(\theta = 1 | X_{1} .. X_{N}) = \frac{1}{1+e^{-z}}$$

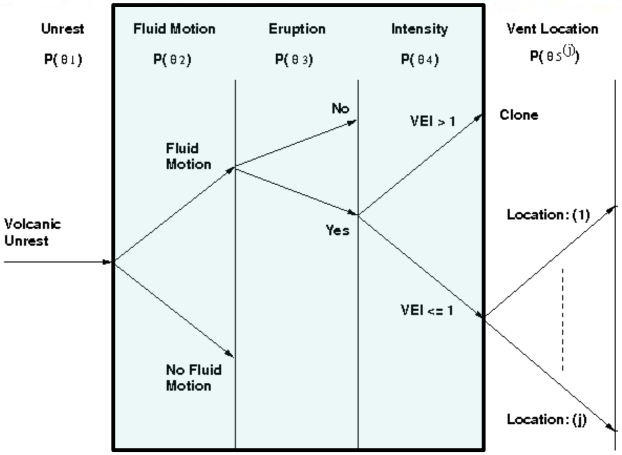




Nodes 2 - 4

Explanatory Variables:

- Independent variables relating a set of observables to an outcome.
- Decouple explanatory variables from modeling technique.
- Model based variables are not dependent on a specific source model.



Explanatory Variables

Variable	Variable Description	Values
X_{MM}	Unrest consistent with magmatic intrusion model	0 or 1
X_{NE}	Average Number of Earthquakes Per Day	0 - ∞
X_{CSM}	Average Normalized Cumulative Seismic Moment Per Day	0 - ∞
X_{DAYS}	Episode Duration in Days	0 - ∞
X_{ERH}	Average Eruption History	0 - ∞

Nodes 2 - 4: Training Data

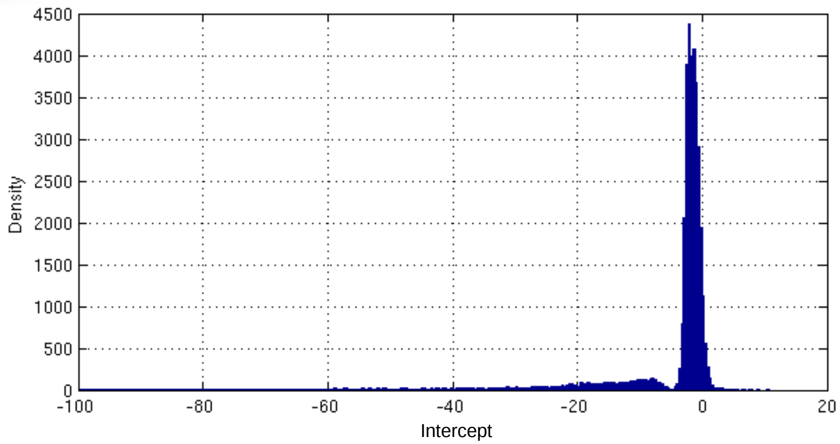
- Historic data was acquired from a combination of published reports and publicly available databases.
- It is assumed that this collection of events is a random sample of volcanic activity in the northern hemisphere and is representative of this population.
- Current data set contains 41 samples.

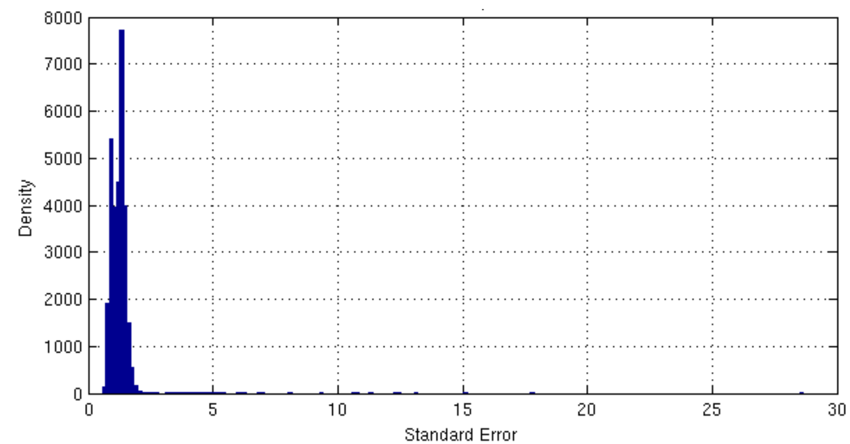
Samples		Response Variables			Independent Variables				
Volcano	Year	In	Er	VEI	MM	Cumulative Number of Earthquakes	Cumulative Seismic Moment	Days	Eruption* History
Medicine Lake	1993	0	0	0	0	115	6.90e+20	2492	1
Hengill	1994	1	0	0	1	63450	7.7e+23	1607	12
Iliamna	1996	1	0	0	1	1477	2.1e21	382	2
Shishaldin	1999	1	1	3	0	688	9.0e+22	42	34
Spurr	2004	0	0	0	0	2743	5.1e+20	239	2
Augustine	2005	1	1	3	1	2007	3.1e+20	80	9
Yellowstone	2008	1	0	0	1	2592	6.9e+22	49	0

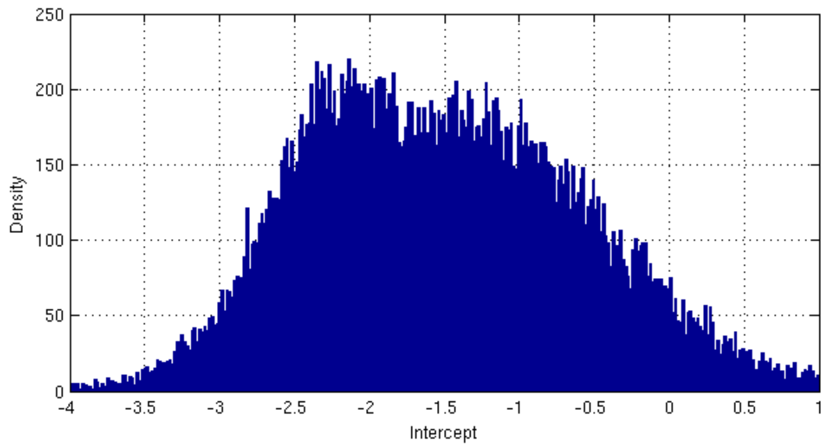
* Eruption history used for eruption node only.

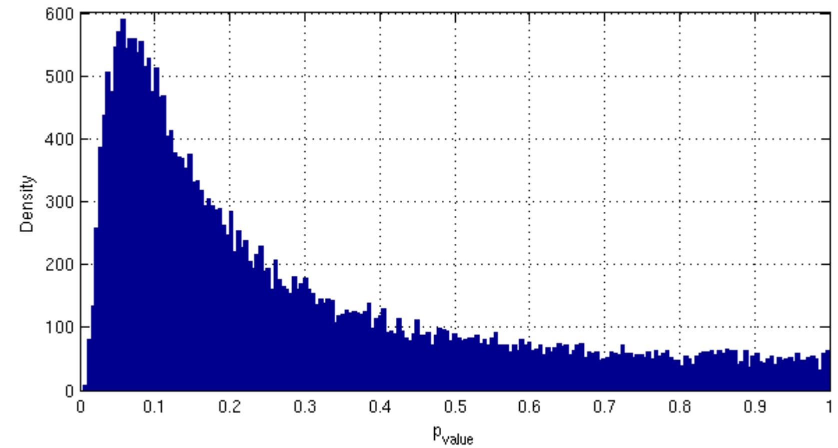
Bootstrapping Analysis Example: Node 2

- Node 2 (Fluid Motion): Intercept, standard error, and p-value distributions.



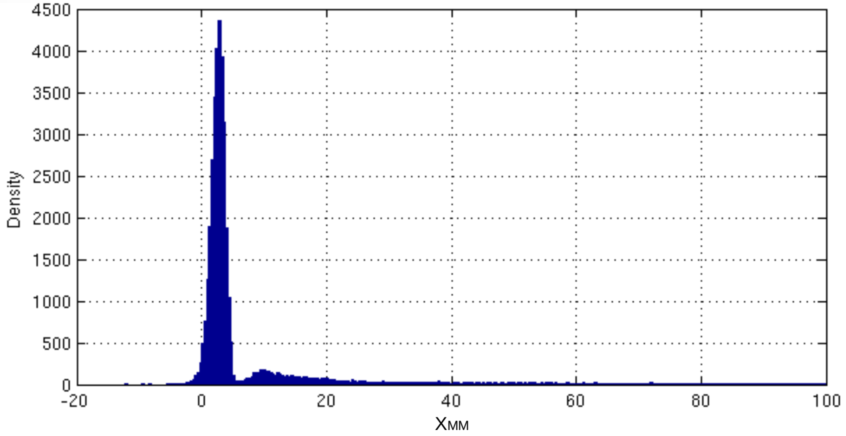


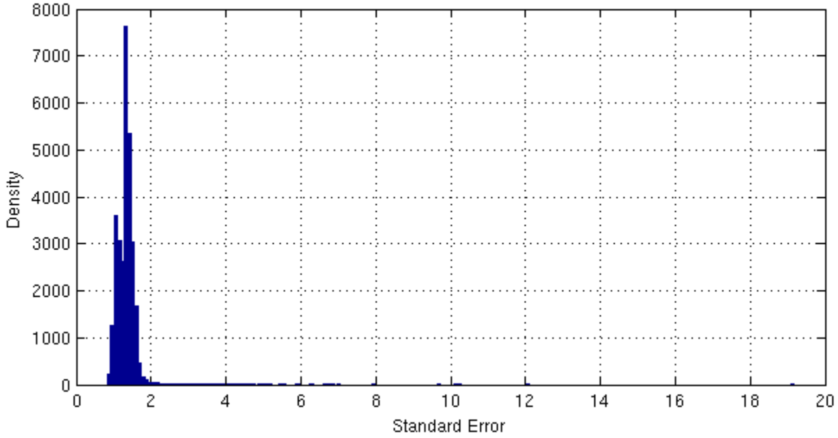


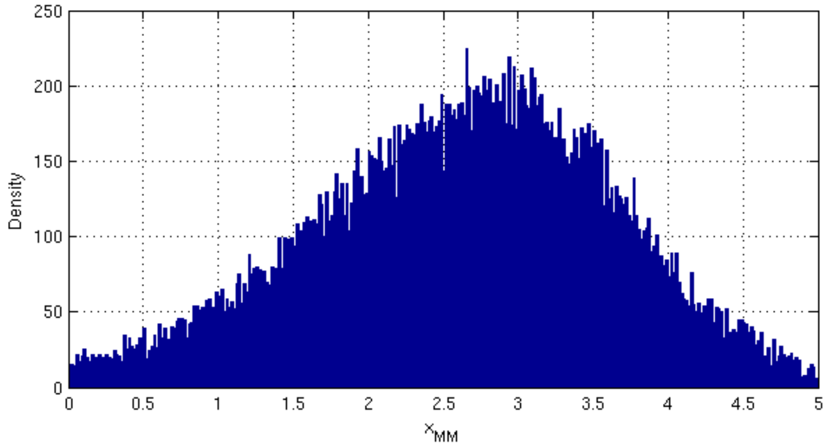


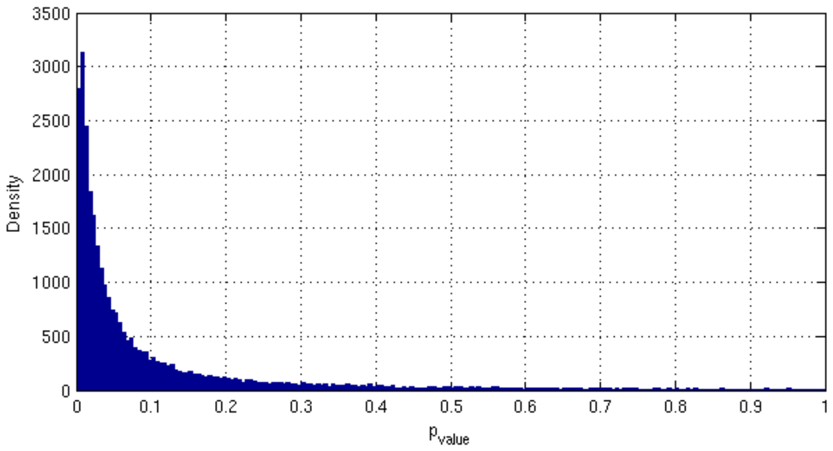
Bootstrapping Analysis: Node 2

- Node 2 (Fluid Motion): X_{MM}, standard error, and p-value distributions.

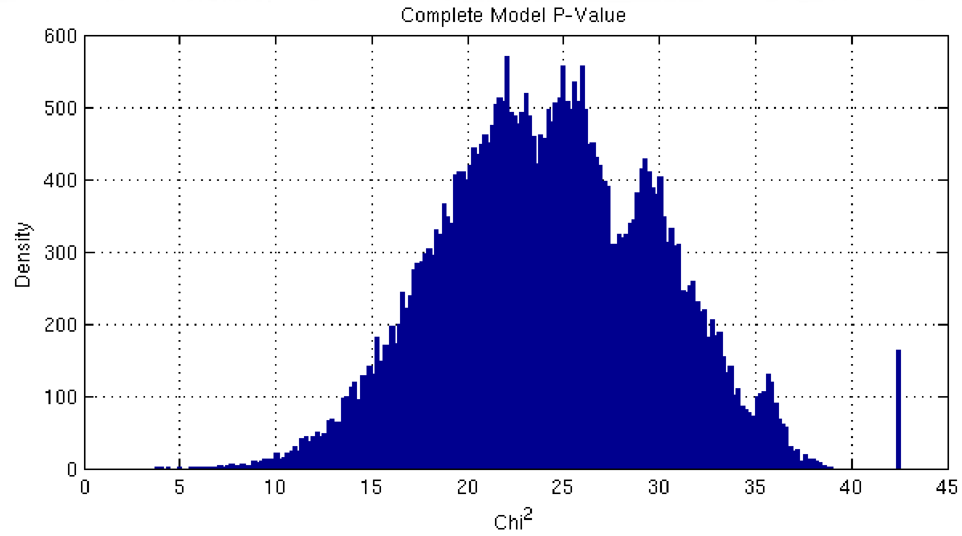


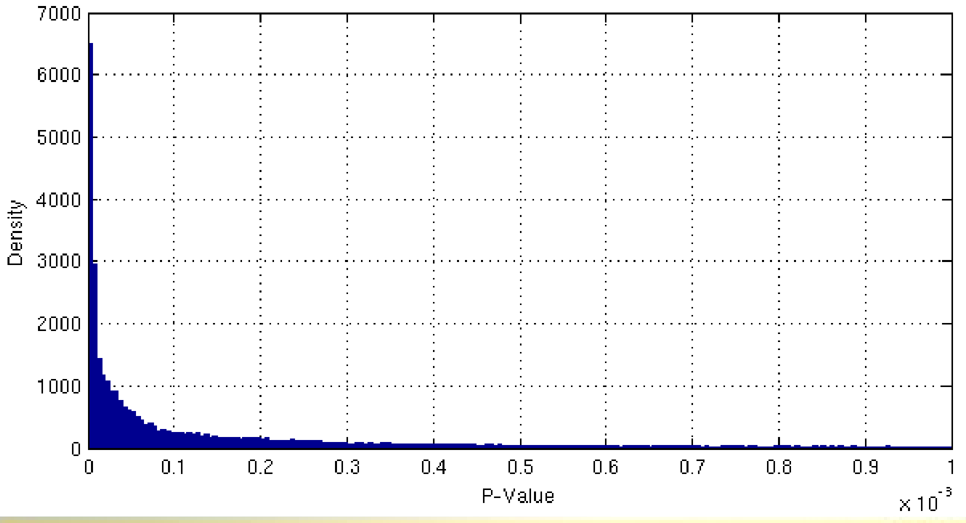






Bootstrapping Analysis Example: Node 2





- The goodness-of-fit (G) of a logistic model is often assessed using a likelihood ratio test.

- This test statistic is computed from the differences between the deviance ($-2\ln(L_n)$) of the null (intercept only) and full logistic models.

- The test statistic is computed from the log ratio

$$G = \chi^2 = -2\ln\left(\frac{L_{null}}{L_{full}}\right)$$

where L_{null} and L_{full} are the likelihoods of the null and full models.

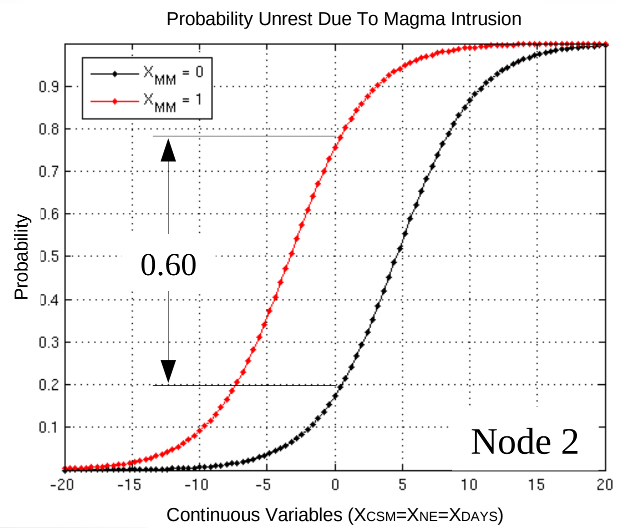
- The test statistic is Chi squared distributed, where its degrees of freedom are equivalent to the number of constrained predictors (explanatory variables) in the logistic model.

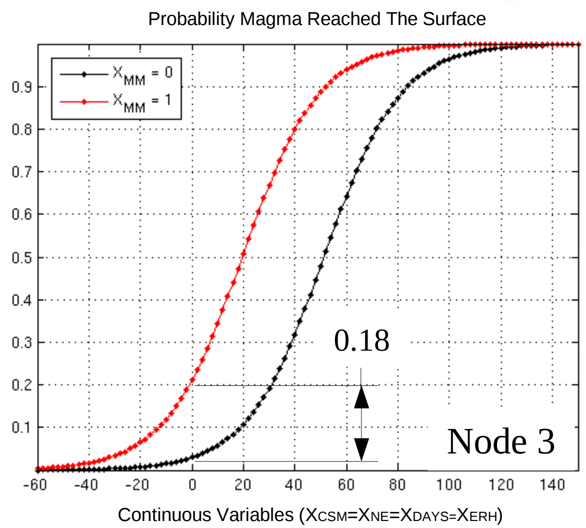
- H_o: Null Model fits data better

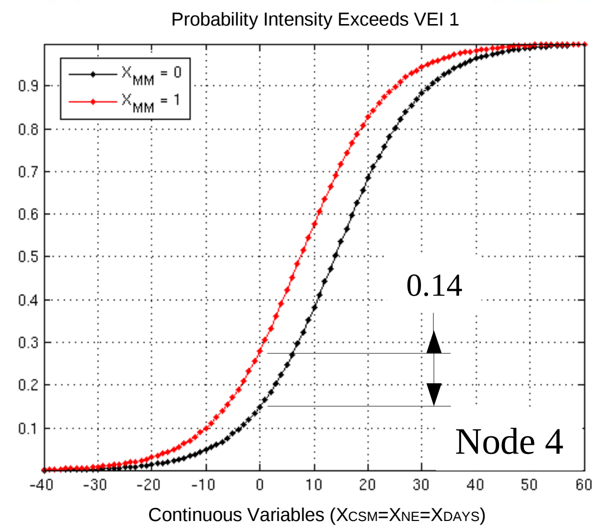
- H_a: Full Model fits data better

- If the difference is statistically significant (e.g., $p < 0.05$), then the full model fits the data better than the null model.

Empirical CDF Comparison







p value = $[P(\chi^2 > G)] = 0.001$

p value = $[P(\chi^2 > G)] = 0.014$

p value = $[P(\chi^2 > G)] = 0.007$

Node 2

$$z = 2.6869(X_{MM}) + 0.3465(X_{NE}) + 0.0041(X_{CSM}) - 0.0001(X_{DAYS}) - 1.5618$$

Node 3

$$z = 2.1401(X_{MM}) - 0.0056(X_{NE}) + 0.0023(X_{CSM}) - 0.0014(X_{DAYS}) + 12.8714(X_{ERH}) - 3.4589$$

Node 4

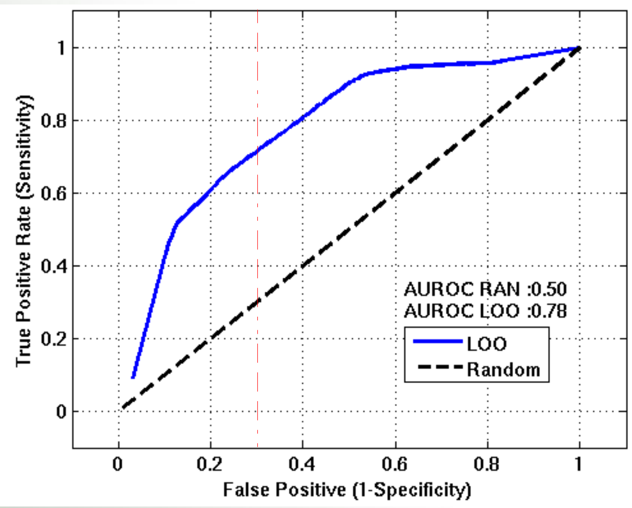
$$z = 0.7924(X_{MM}) + 0.1293(X_{NE}) + 0.0006(X_{CSM}) - 0.0033(X_{DAYS}) - 1.7369$$

Cross Validation

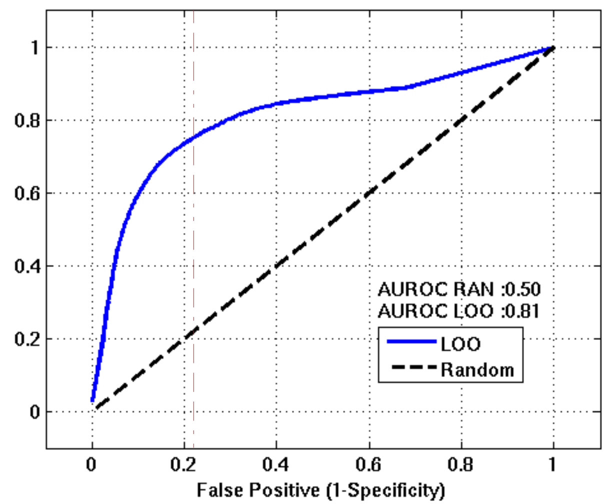
• Leave One Out: Exclude one sample, compute model without excluded sample, predict the excluded sample's outcome, and compare to actual result over a collection of detection thresholds.

• Random: Results expected by randomly selecting the event outcome.

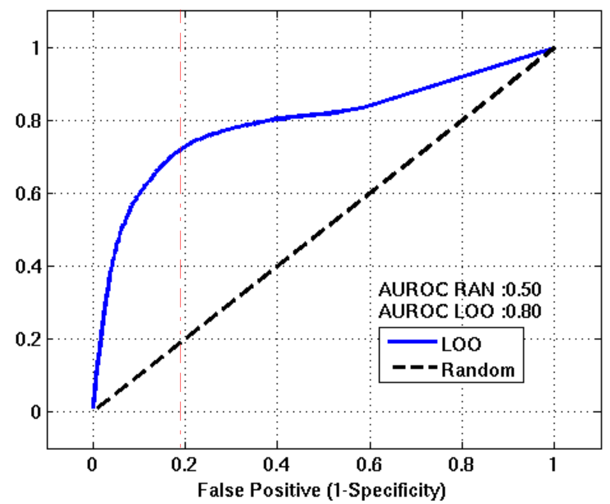
Node 2: Fluid Motion



Node 3: Eruption

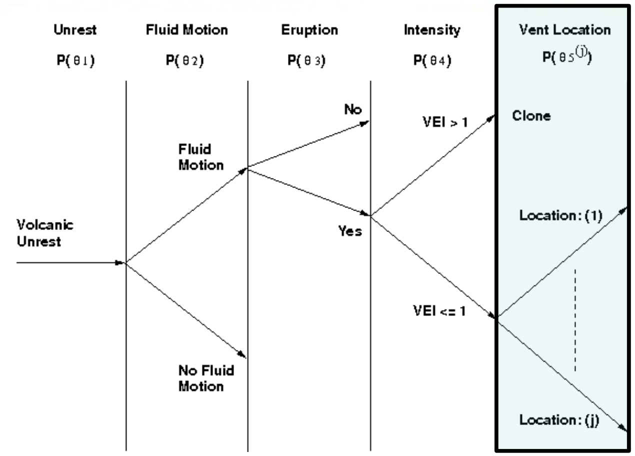


Node 4: Intensity



Node	Threshold	FPR	TPR
Fluid Motion	0.91	29%	71%
Eruption	0.47	21%	75%
Intensity	0.21	19%	73%

Node 5: Vent Location

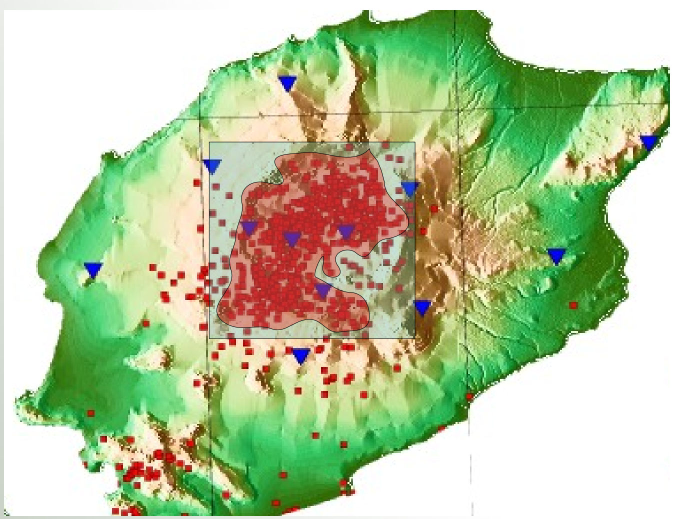


- The spatial PDF for estimating the probability of vent formation (VF) at the jth location is a function of the data used for monitoring a particular volcanic center.

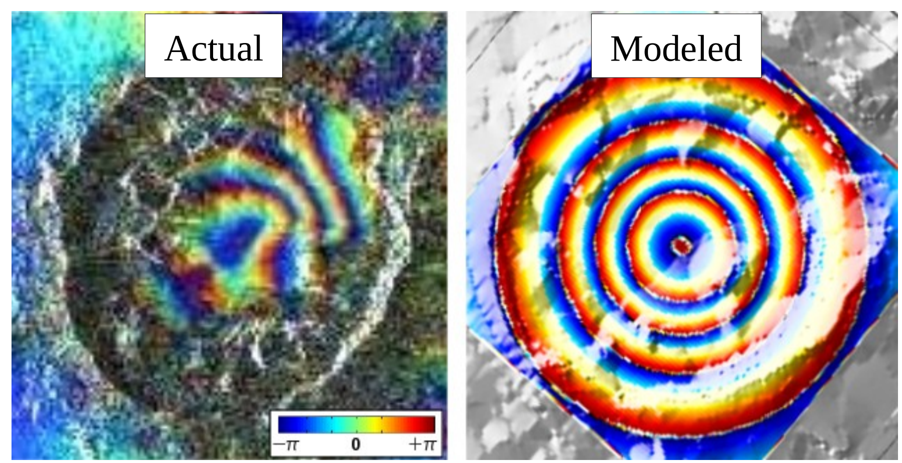
$$P(VF^j|X_n) = \frac{X_n^j}{\sum_{i=1}^{J} X_n^i}$$

$$\Rightarrow$$

$$P(VF^j|X_s, X_d) = \frac{X_s^j + X_d^j}{\sum_{i=1}^{J} (X_s^i + X_d^i)}$$

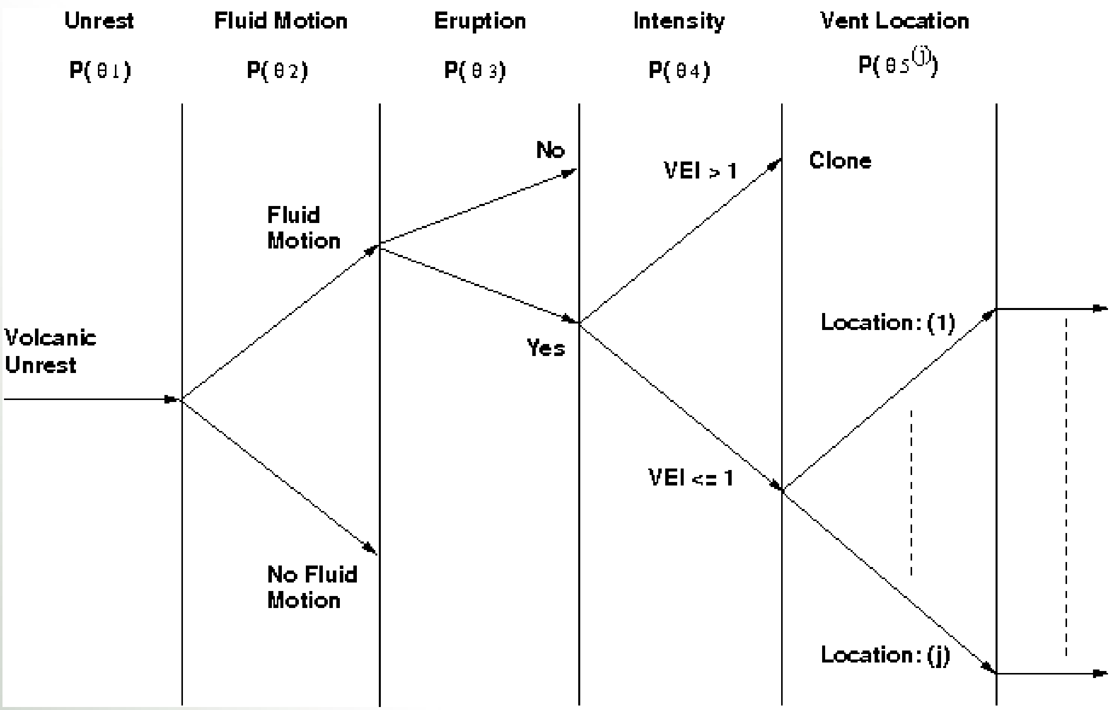


Seismic (X_s)



Deformation (X_d)

Event Tree System



$$P(\theta_1) = P(1)$$

$$P(\theta_2) = P(1)\,P(2|1)$$

$$P(\theta_3) = P(1)\,P(2|1)\,P(3|2)$$

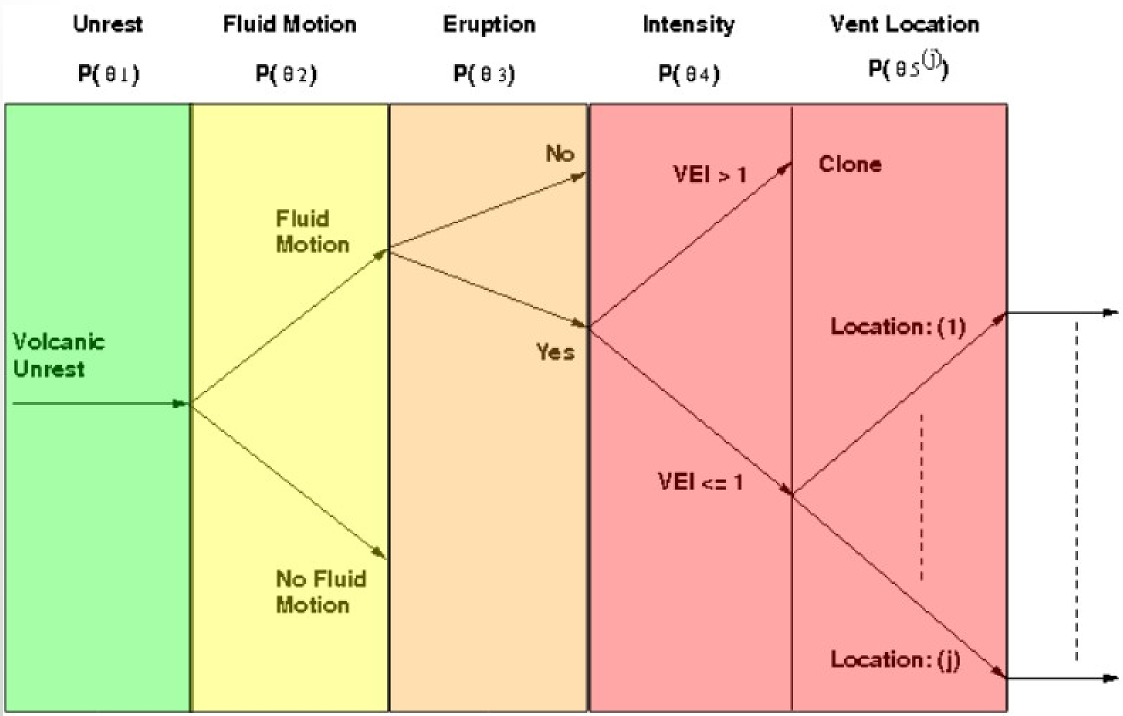
$$P(\theta_4) = P(1)\,P(2|1)\,P(3|2)\,P(4^j|3)$$

$$P(\theta_5) = P(1)\,P(2|1)\,P(3|2)\,P(4^j|3)\,P(5^{VEI}|4)$$

Event Probabilities:

- The probability of a particular event occurring is estimated from the product of conditional probabilities.

Event Tree System

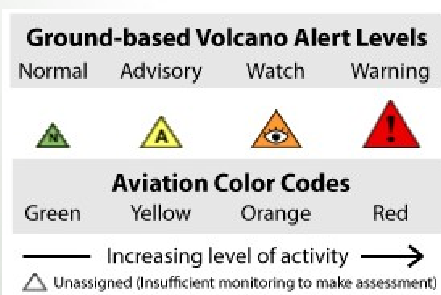


Quantification of USGS Color Code:

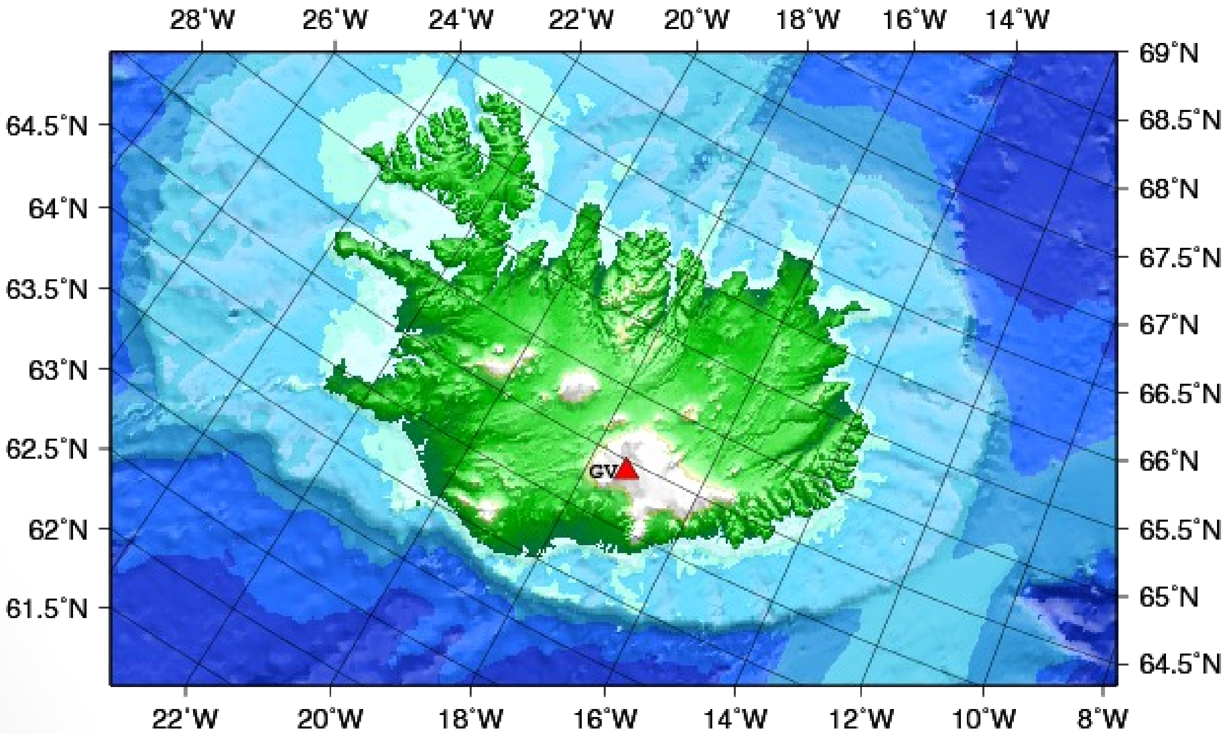
- Green (Normal): Non-eruptive conditions
- Yellow (Advisory): Heighten Unrest
- Orange (Watch): Escalating Unrest
- Red (Warning): Eruption Imminent

Hazard Declaration:

- If the probability of an event occurring at a specific node exceeds a predefined threshold, the process continues to the next node, otherwise it uses the last threshold crossing probability as the hazard declaration.



Results: Grimsvötn

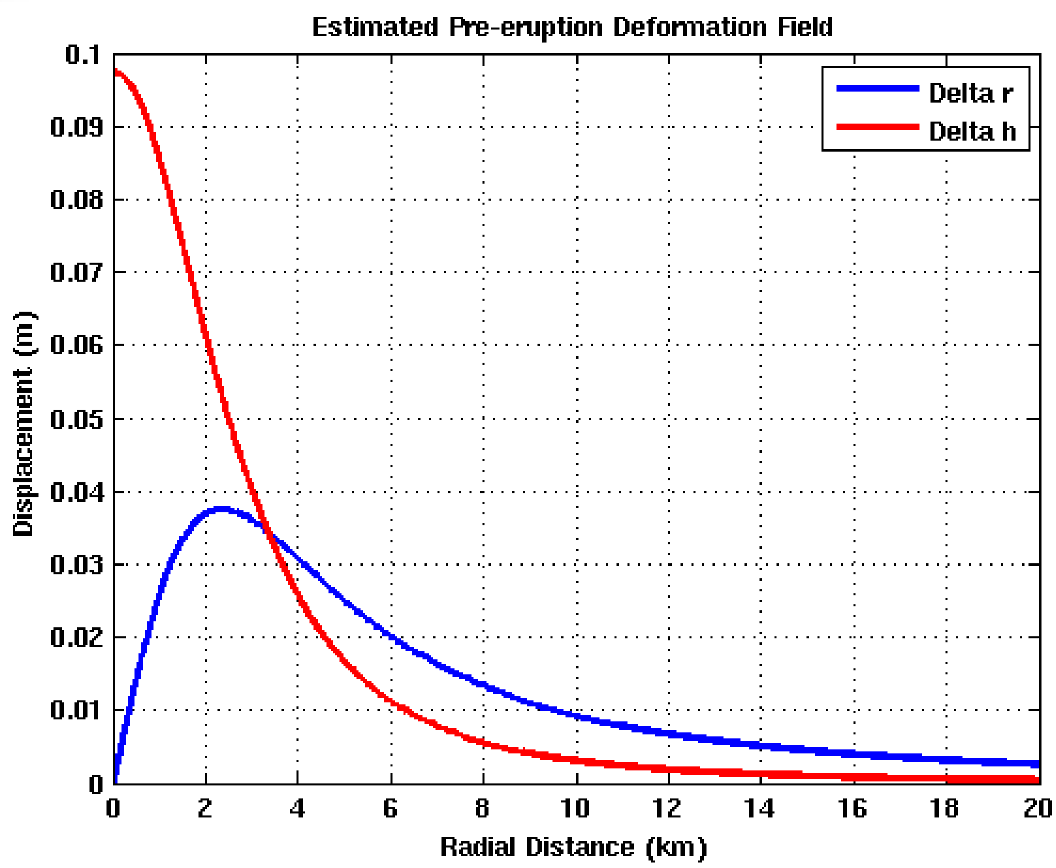


- Grimsvotn is located approximately 200m below the northwestern portion of the Vatnajokull icecap in southeastern Iceland.

- This volcano is among the most active Iceland, where approximately 29 eruptions occurred over the last 211 years.

- Its last eruption began on 21 May, 2011 and produced large ash clouds that extended approximately 20km into the atmosphere which disrupted European air traffic for several days .

Results: Grimsvötn

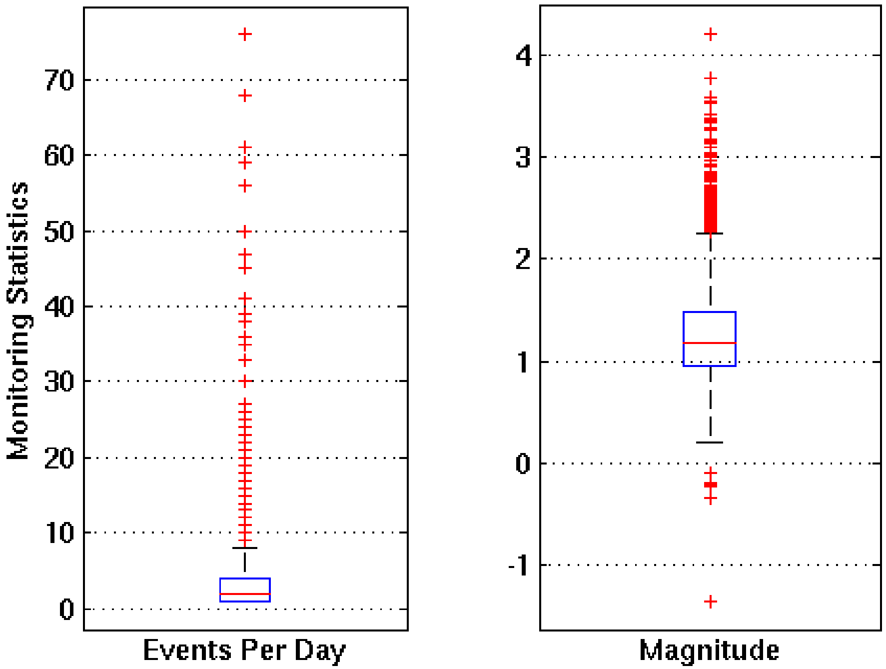


Source Modeling Results:

- Mogi source model solution based on a data acquired by a single GPS sensor (GFUM) approximately 3.0 km from the caldera center.
- GPS displacement observations are similar to those acquired from the last two eruptions.
- Modeling results indicate the magma chamber is between 1.6 – 3.3 km deep.
- Results are consistent with previously published results (Strukell et al. 2003).

Sample	Δh	Δr	d	C	Δv
Pre-Eruption	40 mm	36 mm	3.00 km	0.0011 km^3	0.0068 km^3
Post Eruption	250 mm	468 mm	1.60 km	-0.0061 km^3	-0.0383 km^3

Results: Grimsvötn

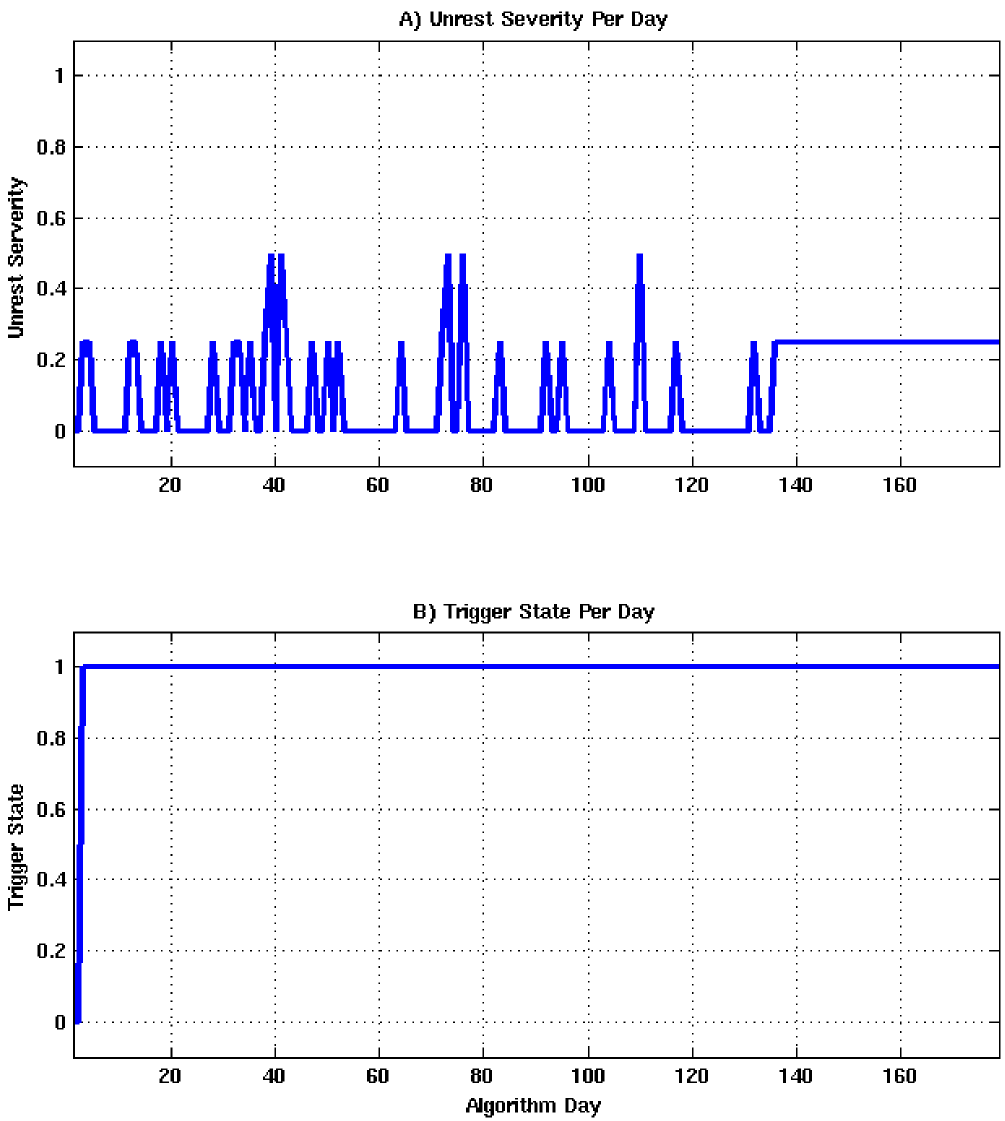


•Boxplots highlighting the distribution of seismicity and deformation beneath the Grimsvotn caldera between 2005 and 2011.

•Events per day and magnitude whiskers are set to 1.5 time the inter quartile range.

•Monitoring thresholds are 8.0 events per day and a ml of 2.27.

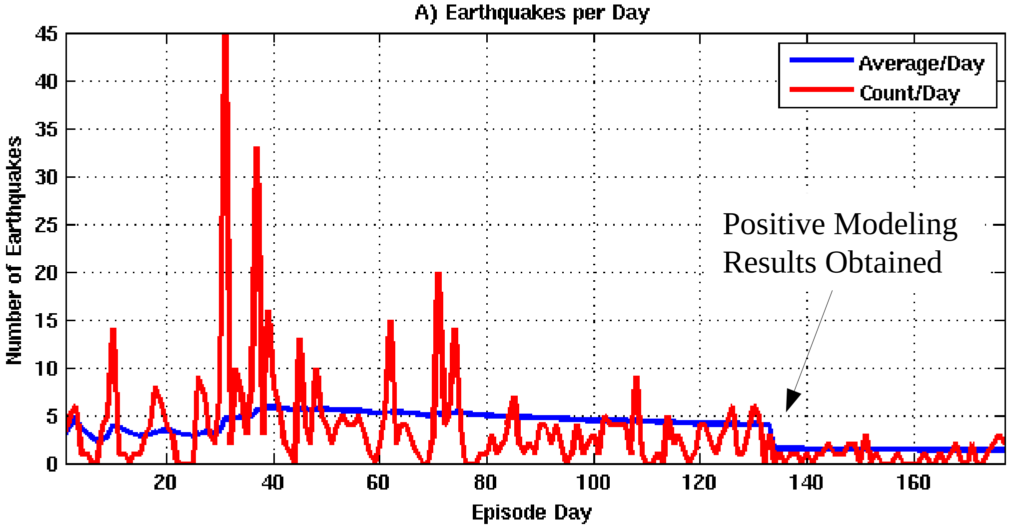
Results: Grimsvötn

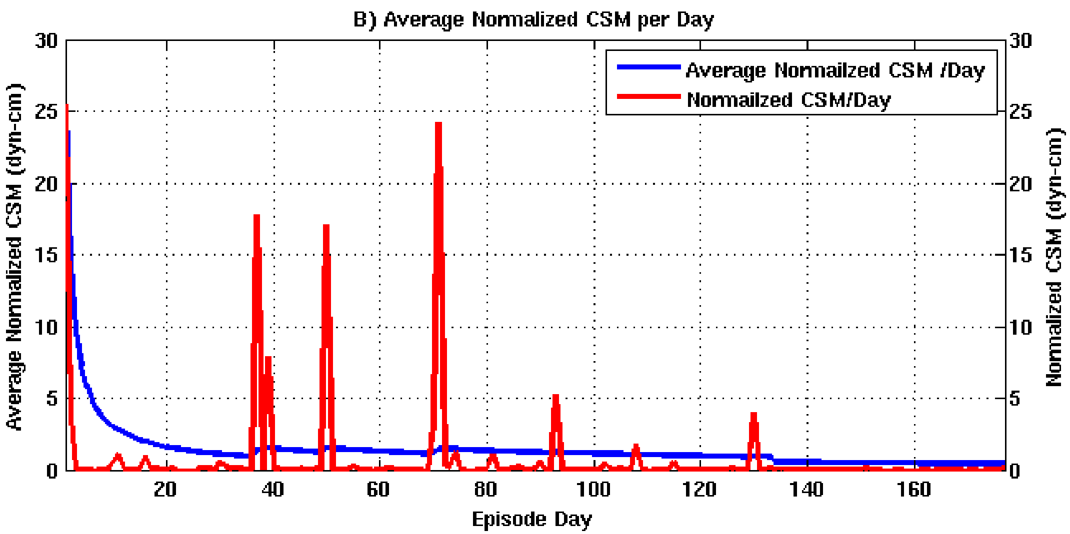


Unrest and Trigger State Vectors:

- Algorithm state as a function of processing day.
- Unrest severity state fluctuates between 0.25 (low) and 0.50 (moderate) over the course of the episode.
- Trigger state = 1, means the algorithm has detected unrest in is actively generating forecasts.
- Trigger state transitions from 0 to 1 on algorithm day 1.

Results: Grimsvötn

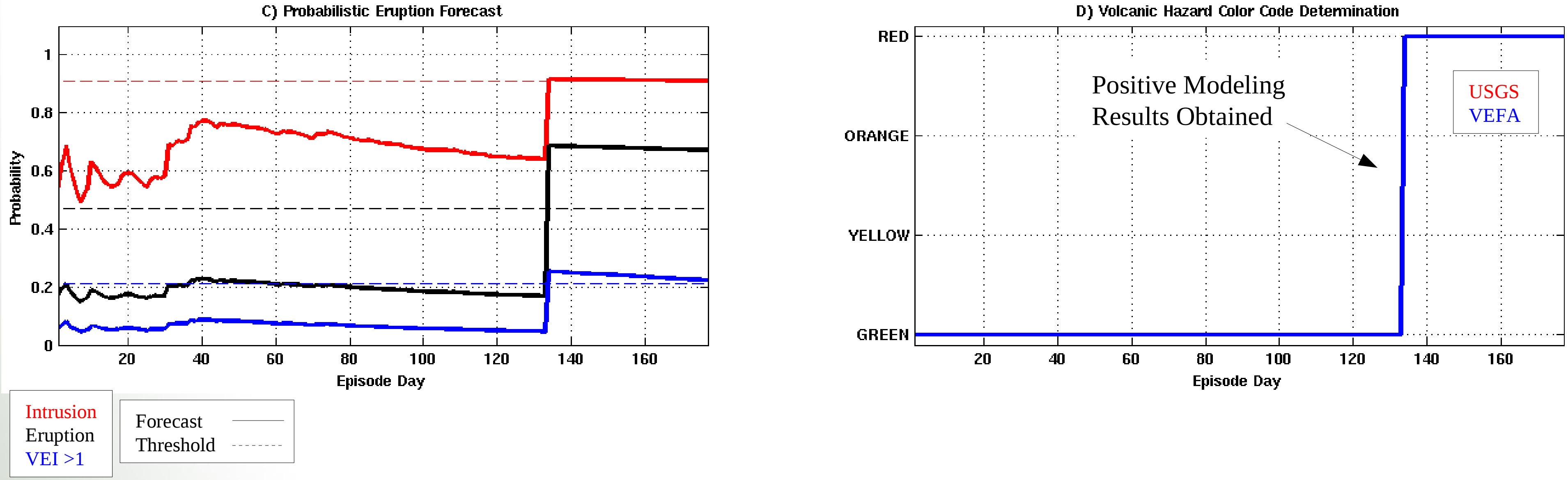




Input Data:

- Volcano monitoring data preceding Grimsvotn's 2011 eruption.
- Count and average (X_{NE}) number of earthquakes per episode day (X_{DAYS}), shown in blue and red.
- Count and average (X_{CSM}) seismic moment per episode day (X_{DAYS}), shown in blue and red.
- Historic eruption frequency index (X_{ERH}) value = 29

Results: Grimsvötn



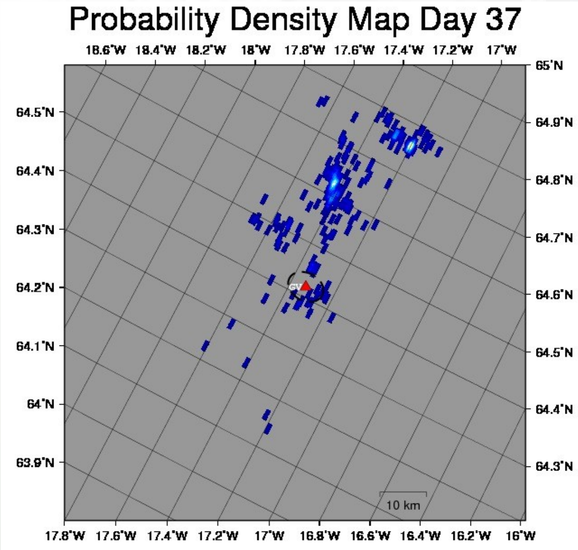
Forecasts:

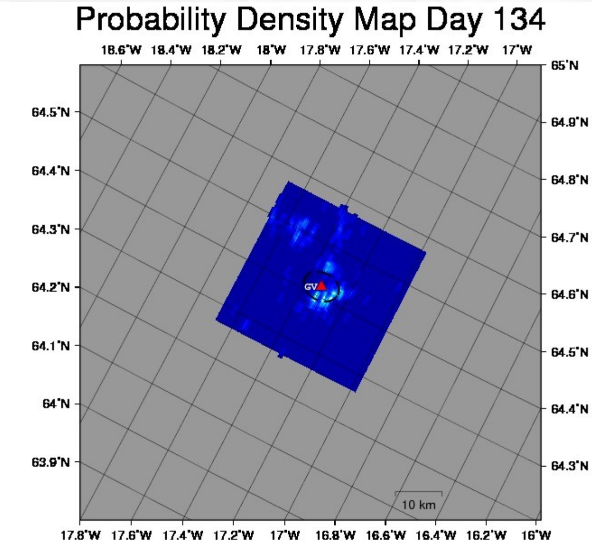
• Initial forecasts fluctuate between 0.60 - 0.80, 0.18 - 0.20, and 0.04 - 0.09 for the intrusion, eruption, and intensity estimates.

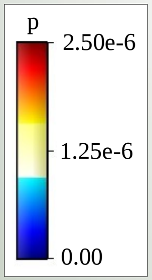
• Revised forecasts suggest this event has a 0.91 probability of being caused by a magmatic intrusion, a 0.68 probability of culminating into an eruption, and a 0.24 probability of exceeding an intensity of 1.0.

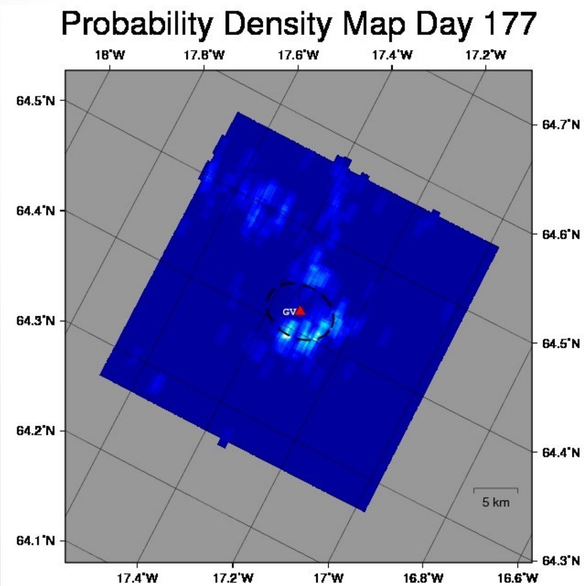
• Upon the determination that the event is the result of a magmatic intrusion, the hazard level immediately jumps to red, which an eruption of greater than VEI 1.0 may be imminent.

Results: Grimsvötn









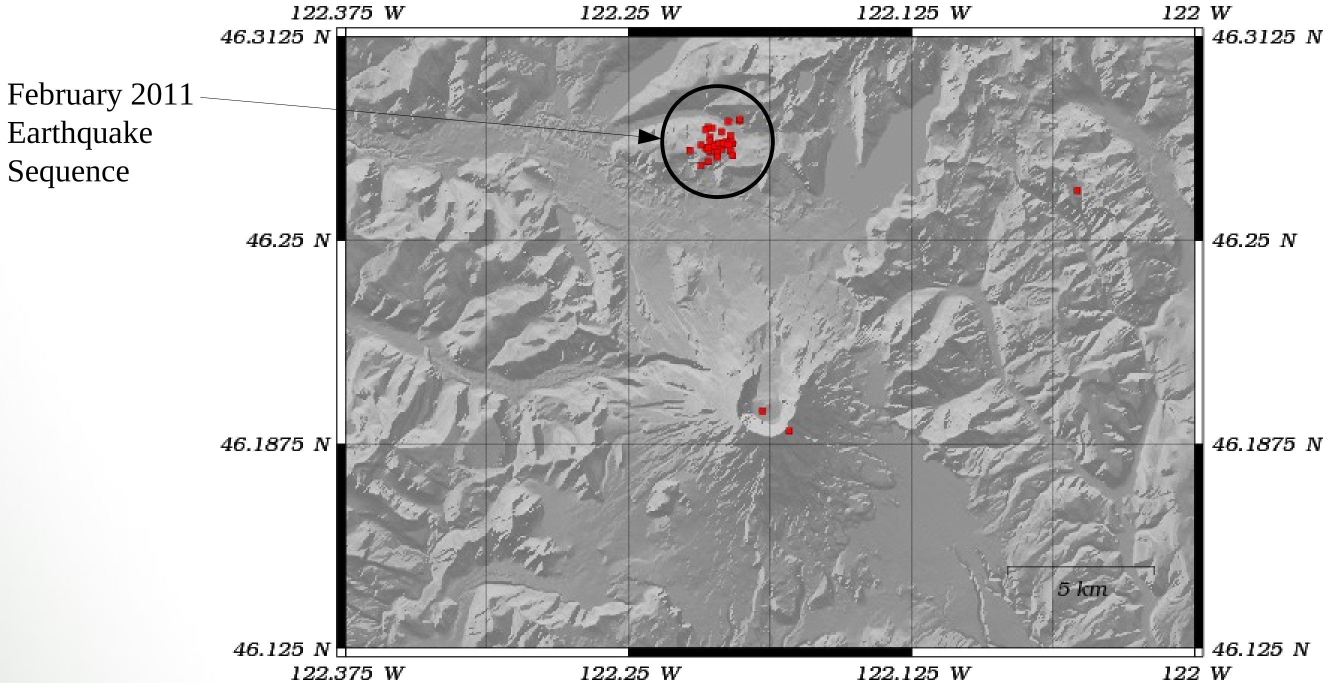
Probable Vent Location:

- PDF is fragmented and scattered over a large area (2979 km^2, J=7.1e6 probable locations).

- Modeling information provided a quantitative constraint on the forecast area (1156 km^2, J=2.8e6 probable locations)

- Eruption occurred in Southeastern section of the caldera.

Results: Mount Saint Helens



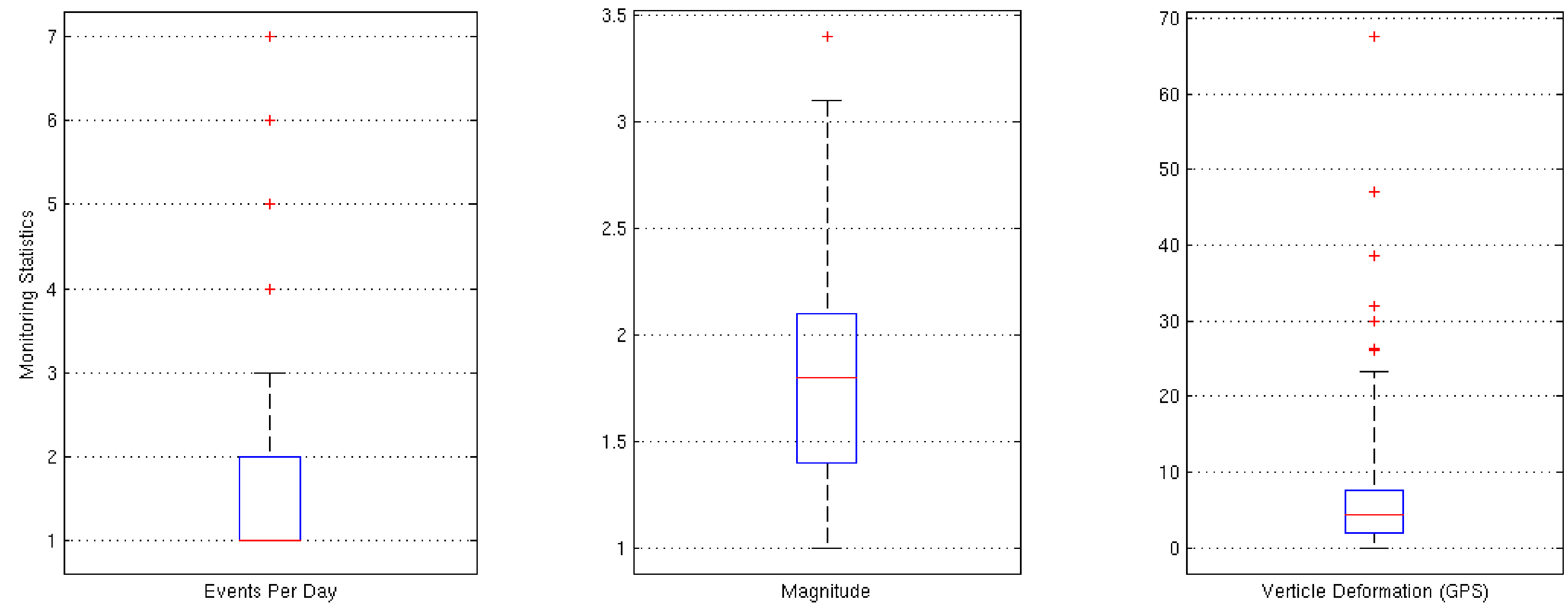
- Mount Saint Helens is an active stratovolcano located in the Cascade Mountain Range in the southwestern region of Washington state.

- It has erupted approximately fourteen times since 1800, which includes four in the last 30 years.

- The most recent eruption began in 2004 and subsided in 2008.

- The 1980 VEI 5 eruption destroyed the top section of the mountain (reducing its elevation by approximately 1000 feet), inflicted massive damage on the surrounding area, and caused some loss of life.

Results: Mount Saint Helens

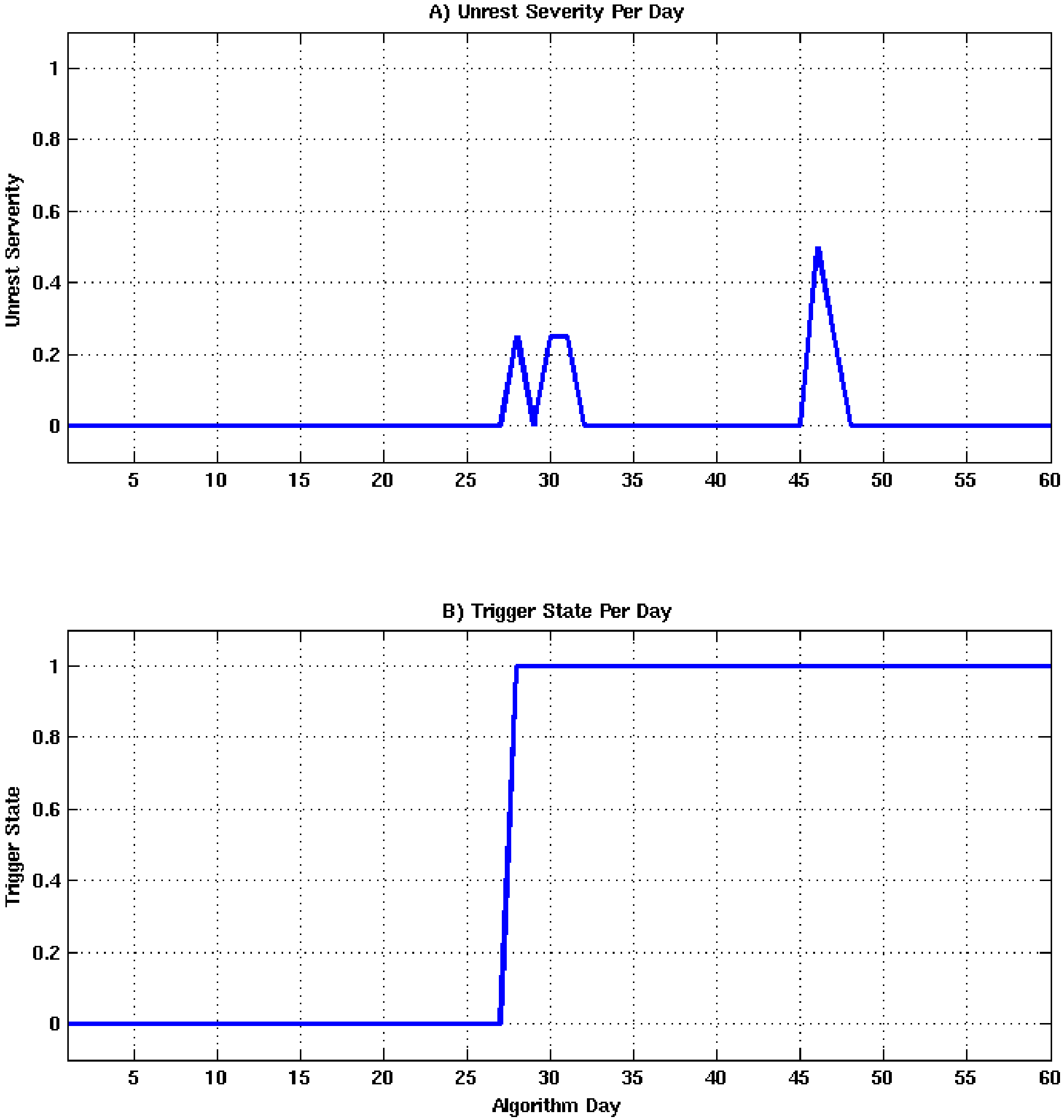


- Boxplots highlighting the distribution of at Mount Saint Helens between 2007 and 2011

- Events per day and magnitude whiskers are set to 1.5 and the vertical GPS deformation measurements are 3 time the interquartile range.

- Monitoring thresholds are 3.0 events per day, ml of 3.1, and 23.3 mm of deformation.

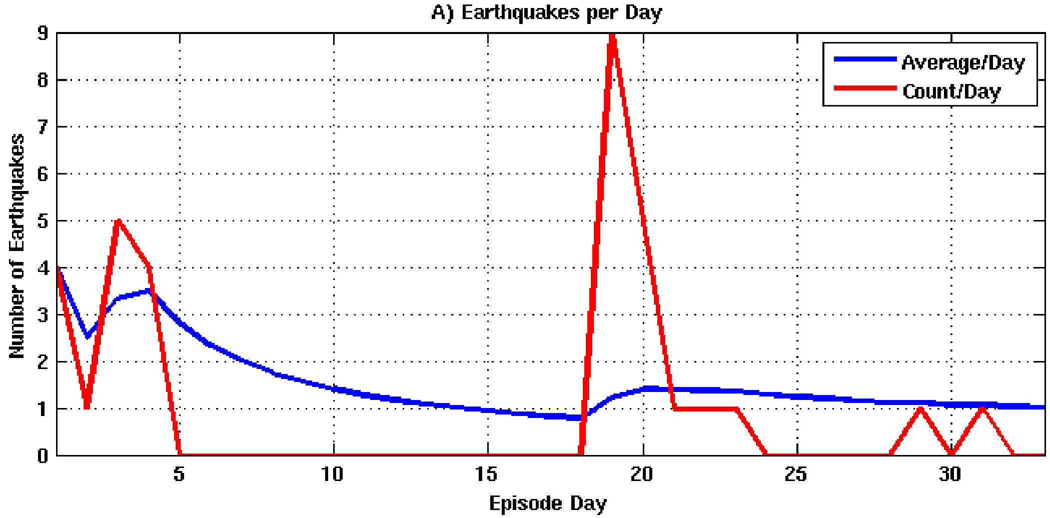
Results: Mount Saint Helens

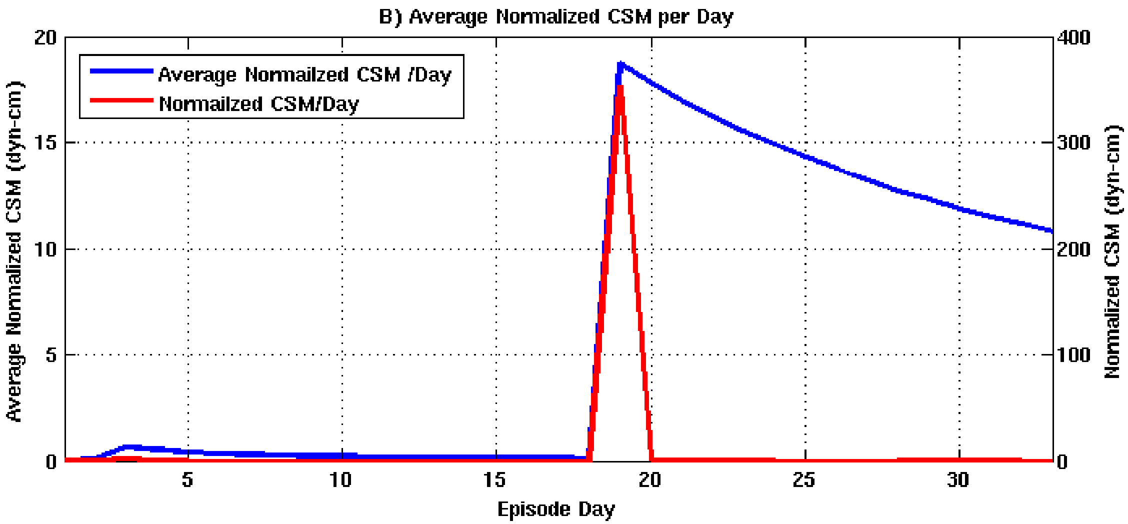


Unrest and Trigger State Vectors:

- Algorithm state as a function of processing day.
- Infrequent detections of 0.25 (low) and 0.50 (moderate) levels of unrest over the course of the episode.
- Trigger state = 1, means the algorithm has detected unrest and is actively generating forecasts.
- Trigger state transitions from 0 to 1 on algorithm day 28.

Results: Mount Saint Helens

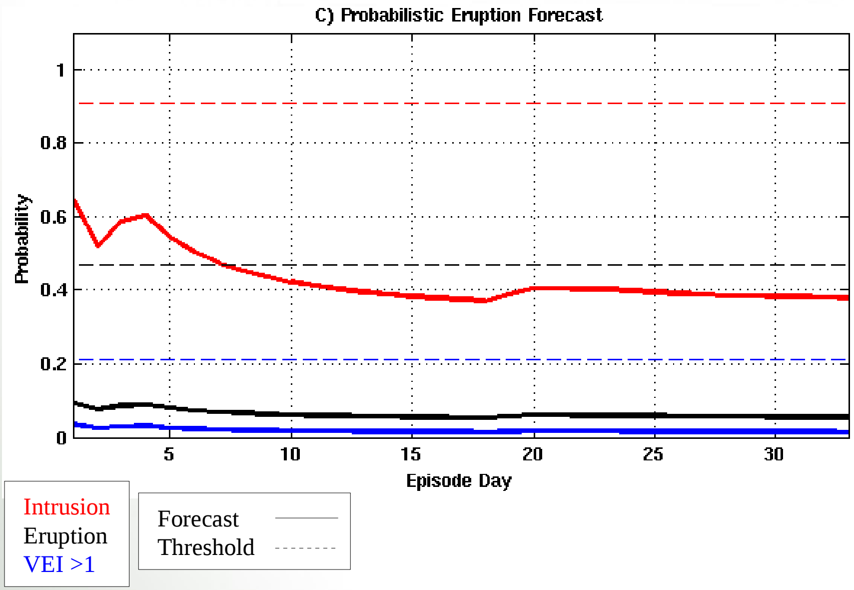


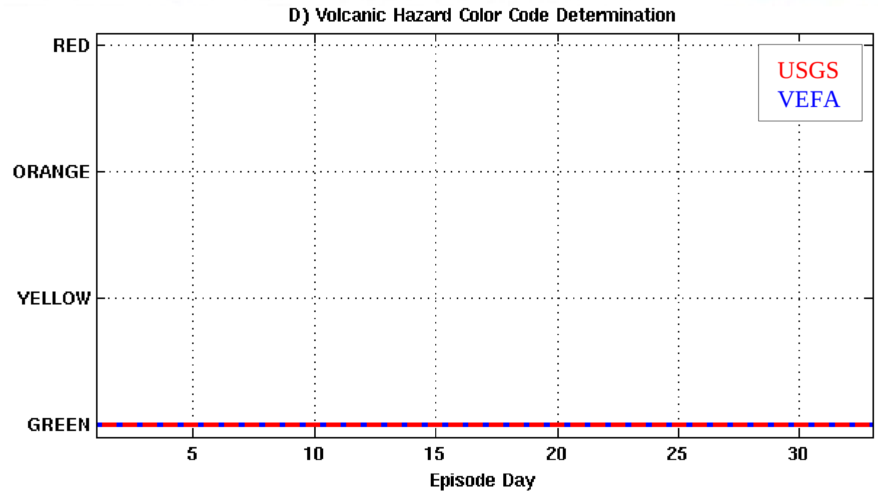


Input Data:

- Volcano monitoring data preceding the 2011 Mount Saint Helens earthquake sequence.
- Count and average (X_{NE}) number of earthquakes per episode day (X_{DAYS}), shown in blue and red.
- Count and average (X_{CSM}) seismic moment per episode day (X_{DAYS}), shown in blue and red.
- Historic eruption frequency index (X_{ERH}) value = 14

Results: Mount Saint Helens





Forecasts:

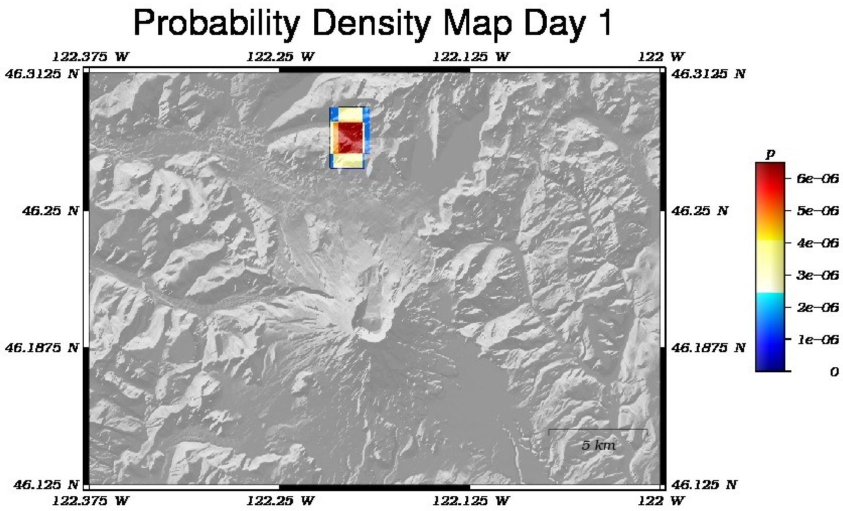
- All event tree forecasts fall below their respective detection thresholds.

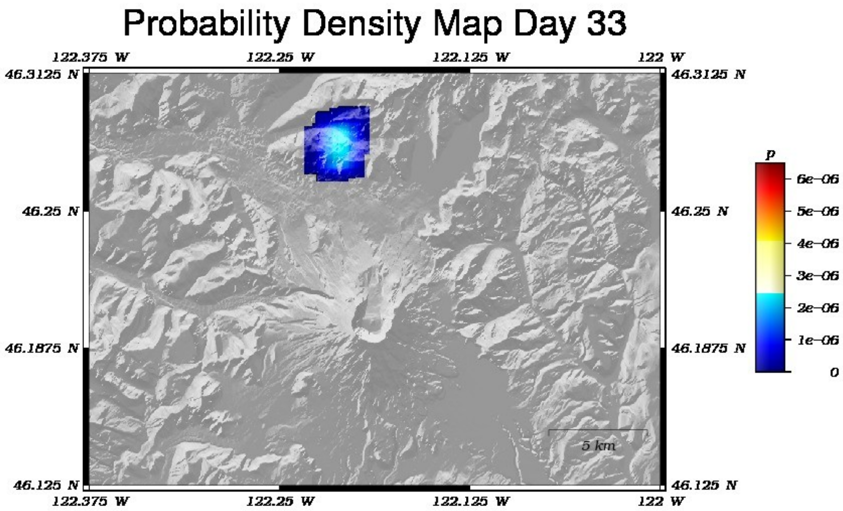
- Intrusion probabilities are initially in the 0.60 range and drop to the 0.37 range as the level of average seismicity per day decreases.

- Final activity forecasts for this episode range between 0.37 - 0.60, 0.05 - 0.09, and 0.01 - 0.03 that a magmatic intrusion is occurring and will result in an eruption that will exceed a VEI of 1.0.

- The color code declaration remains at green for the entire episode, which is consistent USGS hazard level.

Results: Mount Saint Helens

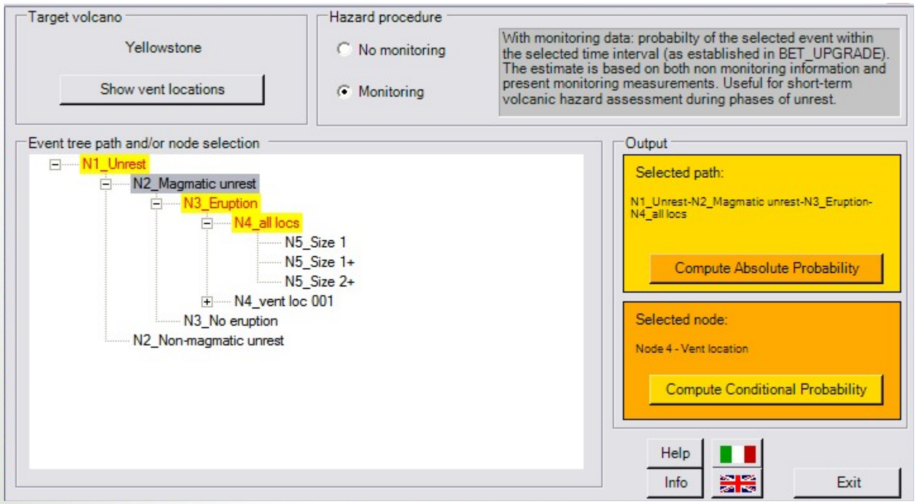


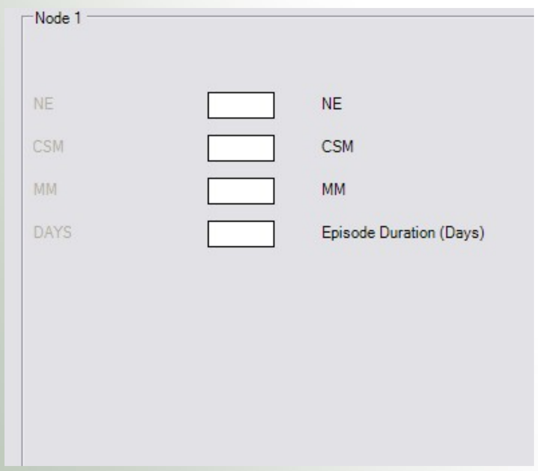


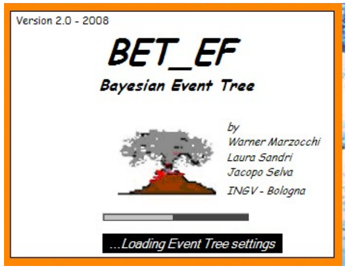
Probable Vent Location:

- PDF encompass a large area (400 km^2, J=1.4e6 probable locations).
- Lack of positive modeling information forces the entire area to be considered for vent formation.
- Modeling results (published by PNSN) suggest the episode is tectonic in nature.
- Seismicity is concentrated to the north of the 1980 eruption crater.
- Eruption at this location is extremely improbable.

Comparison of Results



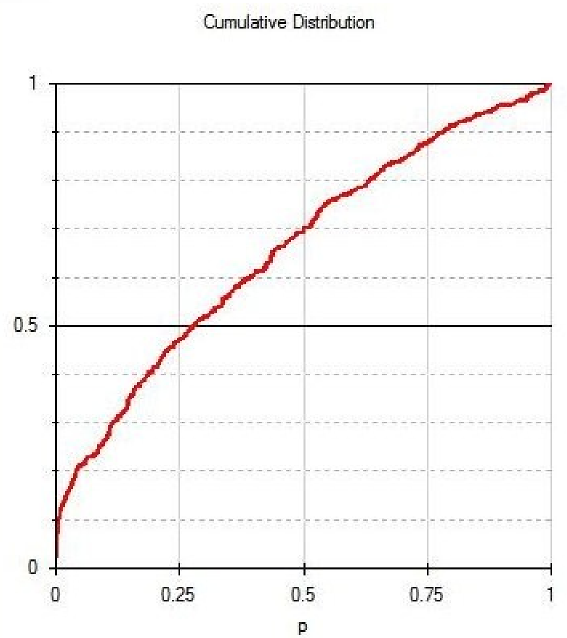


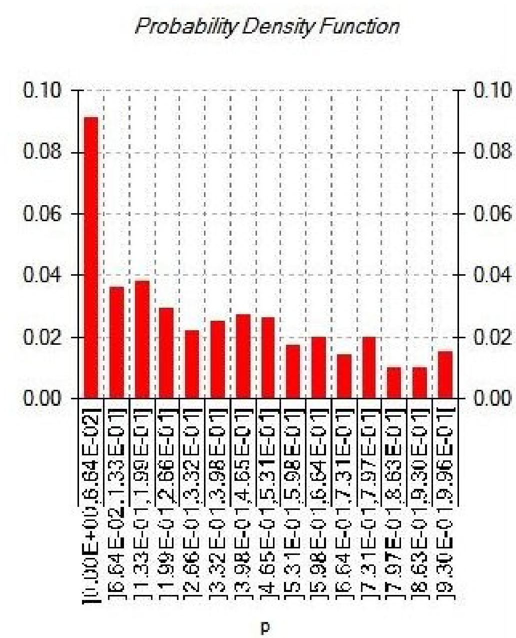


BETEF v2.0 Overview

- BETEF was developed by Warner Marzocchi, Laura Sandri, and Jacopo Selva and is available via the internet.
- Utilizes the fuzzy approach and beta distributions assembled from historic and monitoring data to produce probabilistic forecasts of various types of volcanic activity.
- Unique statistical models must be developed for each volcano being monitored.
- Employs assumptions and statistical models that are only valid for a specific volcano.
- Models are non-transportable and have no defined false positive rate or optimized detection thresholds.
- All model weighting coefficients and detection thresholds used by this application are selected subjectively by the user at the time of their development.

Comparison of Results





Junek_3+N1_Unrest-N2_Magmatic unrest_WM

Values of probability per unit

Average	6.58E-01
10th percentile	6.58E-03
Median	2.79E-01
90th percentile	7.83E-01

WARNING: Tabled values are approximated with 3 digits

Save

Help

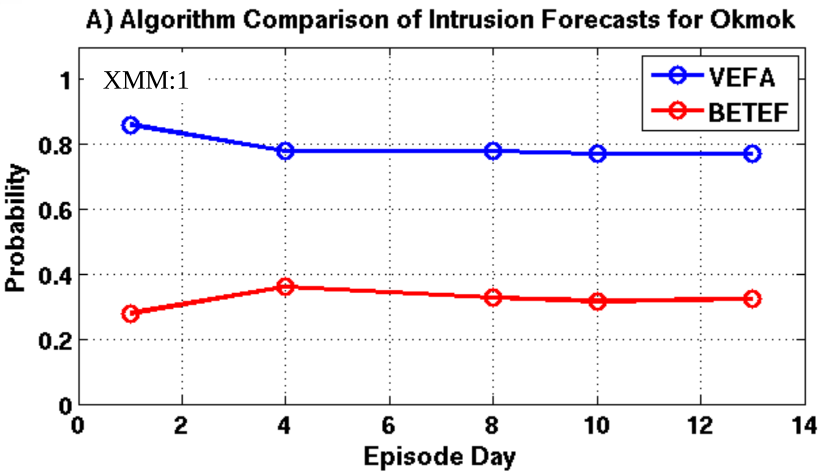
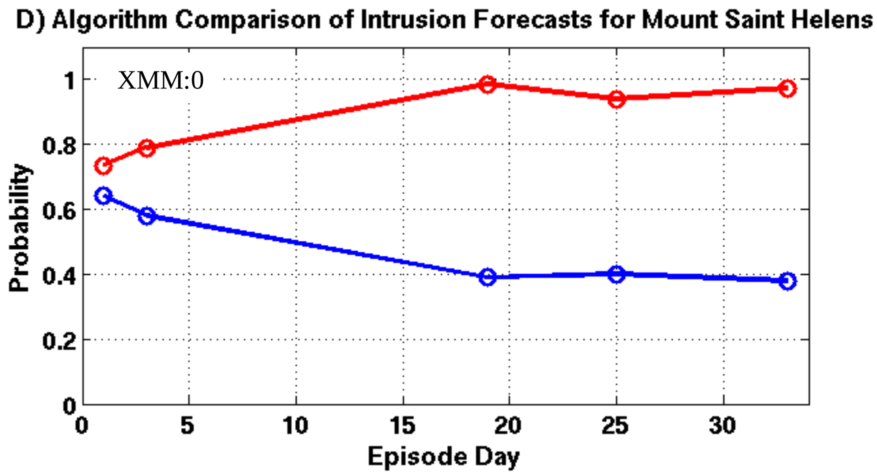
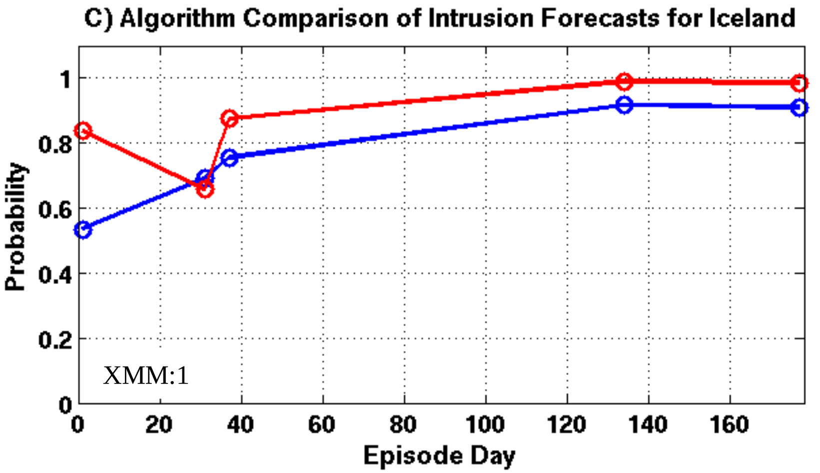
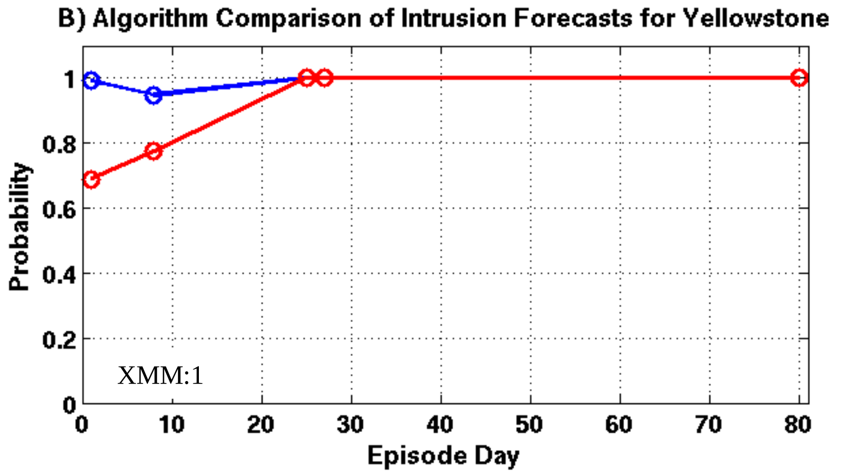
Back

Exit

Comparison Setup:

- Models used for each example were trained using monitoring and modeling data acquired from the most recent unrest episode preceding the event under test.
- Since eruption history is built directly into the BETEF model during the design process, the ERH parameter is not required as a BETEF input.
- Weighting coefficients for each of the BETEF input parameters (XNE, XCSM, and XMM) were set to 1.
- All VEFA and BETEF forecasts are based on the same input parameters to ensure comparability of results.
- The median value of the BETEF prediction is used for comparison purposes for all the examples shown below.

Comparison of Results

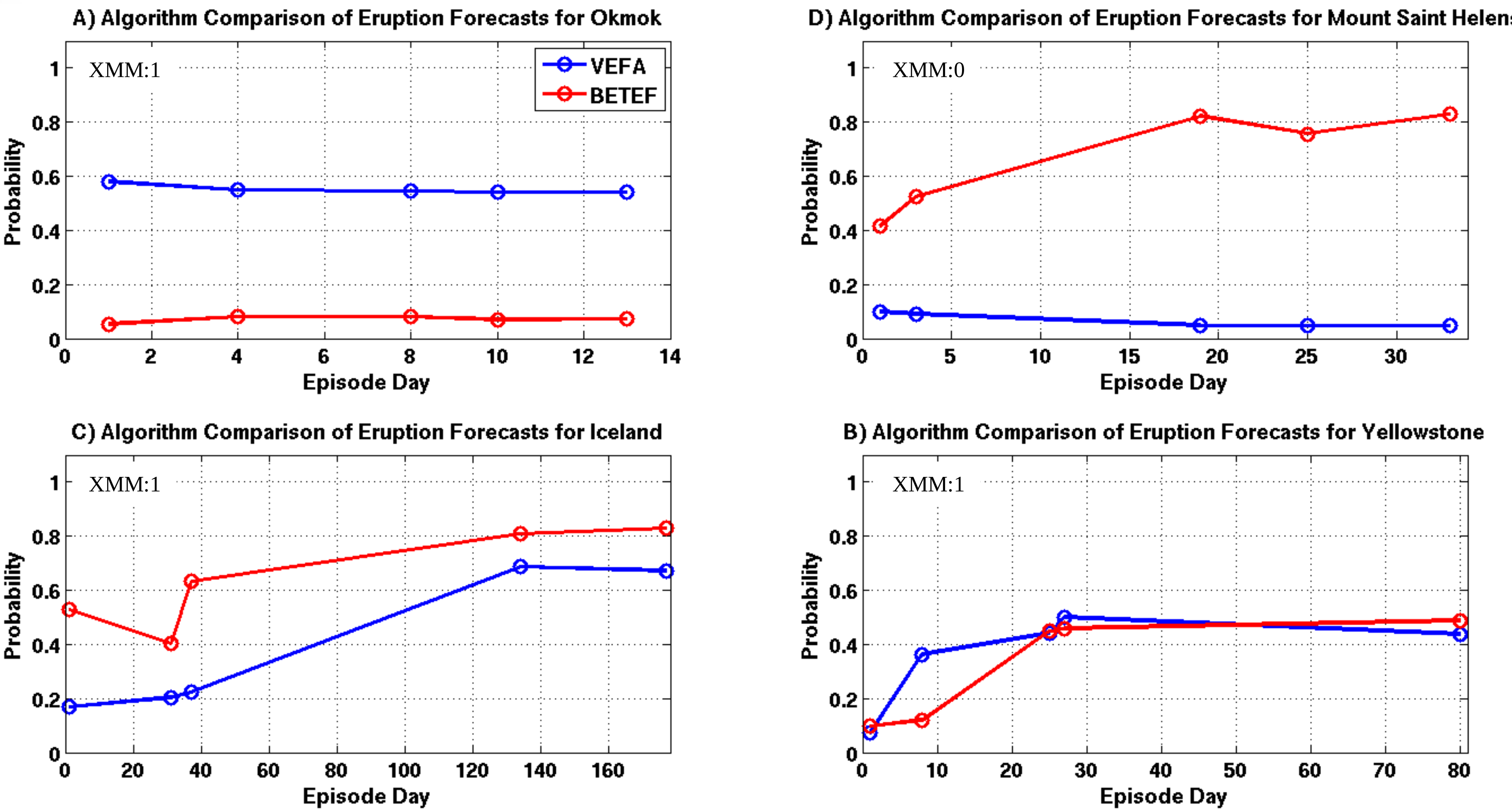
	Eruption	No Eruption
LS		
HS		

Comparison of Results

Eruption | No Eruption

LS

HS

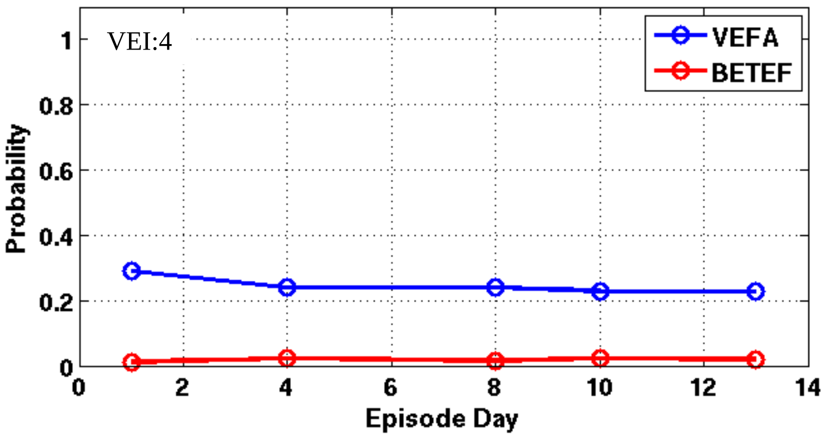


Comparison of Results

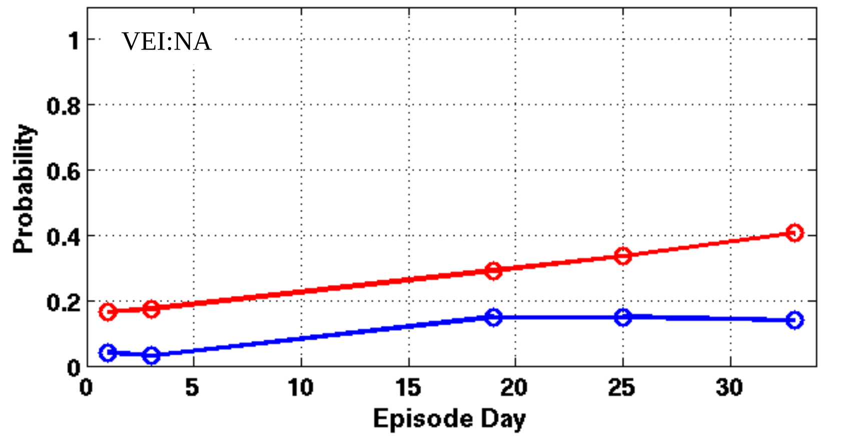
Eruption | **No Eruption**

LS

A) Algorithm Comparison of Intensity Forecasts for Okmok

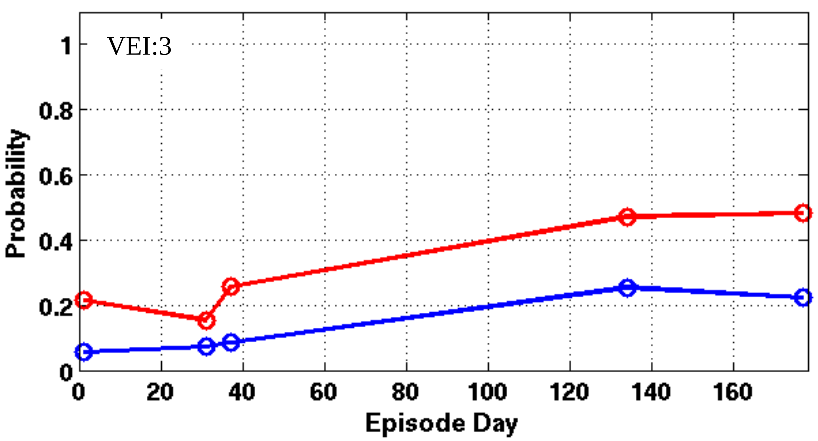


D) Algorithm Comparison of Intensity Forecasts for Mount Saint Helens

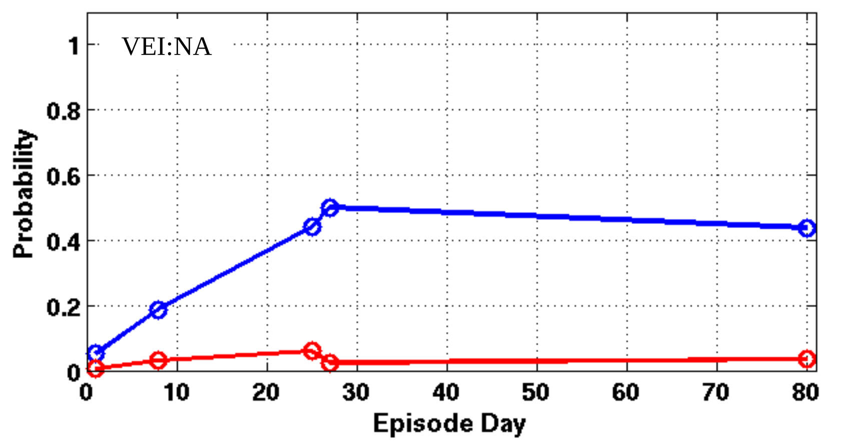


HS

C) Algorithm Comparison of Intensity Forecasts for Iceland



B) Algorithm Comparison of Intensity Forecasts for Yellowstone



Conclusions

- The relative weighting of the logistic coefficients showed modeling information (XMM) has a significant influence on the probability estimate.

- Source modeling information provided a quantitative constraint on the forecast area, and provides the justification for focusing attention on smaller areas that are directly experiencing the effects of magma ascent.

- A comparison of results generated by this method and a published approach illustrates the power of combining modeling and monitoring information with historic data to forecast short-term volcanic activity.

- Logistic regression approach performed comparable to, and in some cases, outperformed, no-transportable empirical models built from site specific information.

- Preliminary results show the logistic models are transportable and can be applied to volcanoes where modeling and monitoring information are available.

Future Work

- Expand the training dataset and re-evaluate the model coefficients using data provided by the upcoming WOVO volcanic unrest database.

- Identify new independent variables (e.g., thermal, geochemical) to add to the logistic models.

- Applying the algorithm to a variety of volcanic centers to identify existing and future volcanic hazards

Publications

Conference Proceedings:

[1] "Temporal Analysis of the Magma Supply System Beneath the Okmok Caldera by Interferometric Synthetic Aperture Radar and Statistical Seismology", **William N. Junek**, W. Linwood Jones, and Mark T. Woods, *IEEE International Geoscience and Remote Sensing Symposium*, 2010, pp. 1545-1548

[2] "Short Term Forecasts of Volcanic Activity Using an Event Tree Analysis System and Logistic Regression", **William N. Junek**, W. Linwood Jones, and Mark T. Woods, *2011 American Geophysical Union Conference*

[3] "Detecting Developing Volcanic Unrest", **William N. Junek**, W. Linwood Jones, and Mark T. Woods, *2012 IEEE South East Conference*

Journals:

[1] "Locating Incipient Volcanic Vents Using Multidisciplinary Remote Sensing Data and Source Modeling Information", **William N. Junek**, W. Linwood Jones, and Mark T. Woods, *IEEE Geoscience and Remote Sensing Letters*, Currently Under Review

[2] "Utilization of Logistic Regression for Forecasting Short-Term Volcanic Activity", **William N. Junek**, W. Linwood Jones, and Mark T. Woods, *Algorithms*, Planned Submission April 2012

Questions?

Backup

Terminology

Volcanic Explosivity Index (VEI):

- Derived from a combination of the volume of material expelled during the eruption, the column height, and a qualitative description of the event.

- Scale ranges between 0 and 8.

- Eruption intensity values increase one VEI unit when the volume of expelled material (tephra) increases one order of magnitude.

- For example, an eruption that expels 1e6 m^3 of tephra has a VEI of approximately 2. However, if the eruption produces 1e12 m^3 of tephra its VEI is approximately 8.

Seismic Moment:

- Seismic moment (M_0) relates earthquake size to a set of fundamental source parameters (Shear modules, fault displacement (slip), and fault area).

- Expressed in terms of dyne-cm (10^-7 Nm).

$$M_o = 10^{1.5(M_w + 10.73)}$$

Eruption Products

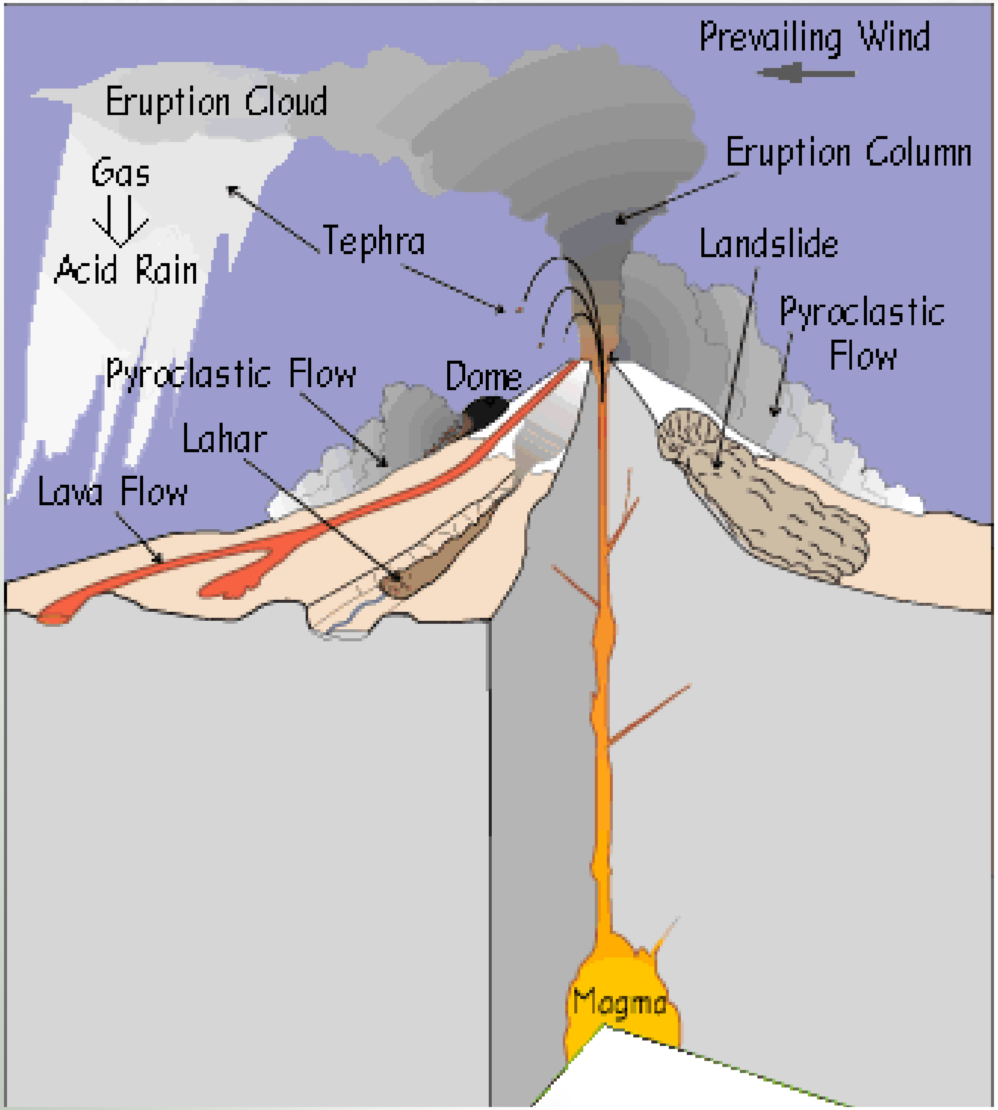


Image obtained from the USGS Volcanic Hazard Program Web page: http://volcanoes.usgs.gov/

Volcanic Hazards:

- Tephra
- Lahar
- Landslide
- Lava
- Pyroclasitc Flow
- Eruption (Ash) Cloud
- Acid Rain

Modeling Volcanic Processes

Estimating Source Behavior:

- Interferometric Synthetic Aperture Radar (InSAR) techniques can be used to measure the surface deformation around a volcano.

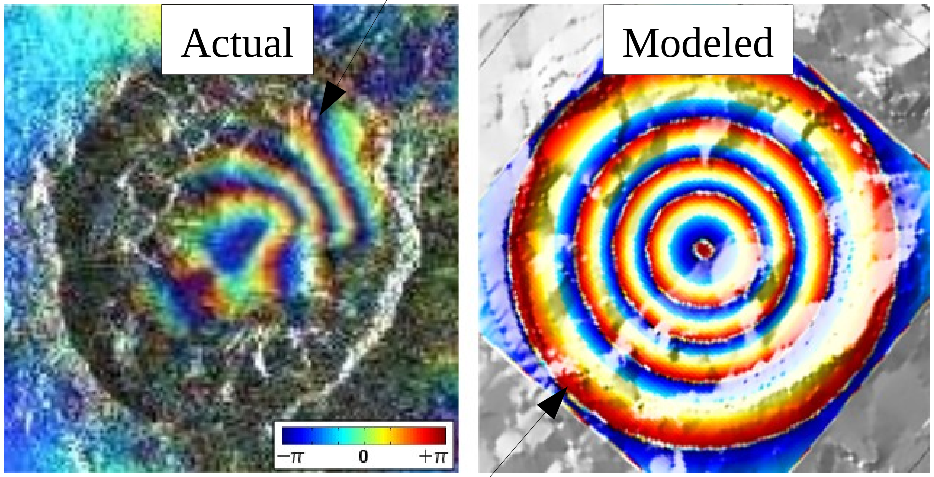
- Source modeling is performed by empirically matching a synthetic interferogram, generated by a Mogi source model, to the actual image.

- Synthetic interferograms are created by substituting selected values of C, d into the Mogi source equations for a particular location.

- The resulting ground deformation is wrapped at a rate equal to half the electromagnetic wavelength of the radar (λ = 2.83 cm)

- This creates a synthetic interferogram that can be compared to the actual image.

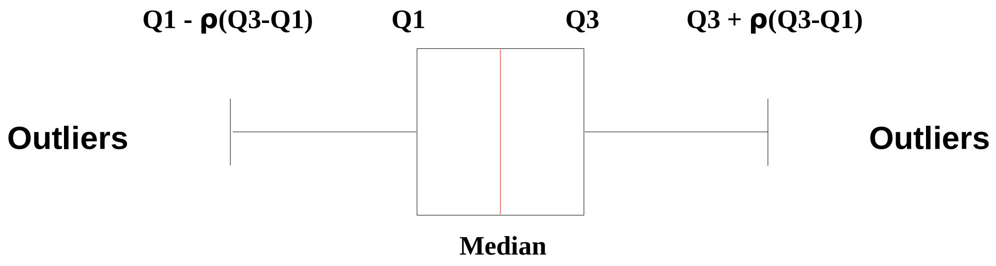
Each interferometric fringe is equal to a phase rotation of 2 π radians.



Chamber depth (4.0 km) obtained from (Junek, Jones, and Woods IGARSS, 2010)

Statistics

Box Plot



Outliers:

- Outliers are observations that are significantly different from other members of a sample distribution
- Here they are defined as measurements greater than a value given by

$$Threshold = Q_3 + \rho\left(Q_3 - Q_1\right)$$

- The constant ρ is determined empirically on a case by case basis.

Statistics

p-value:

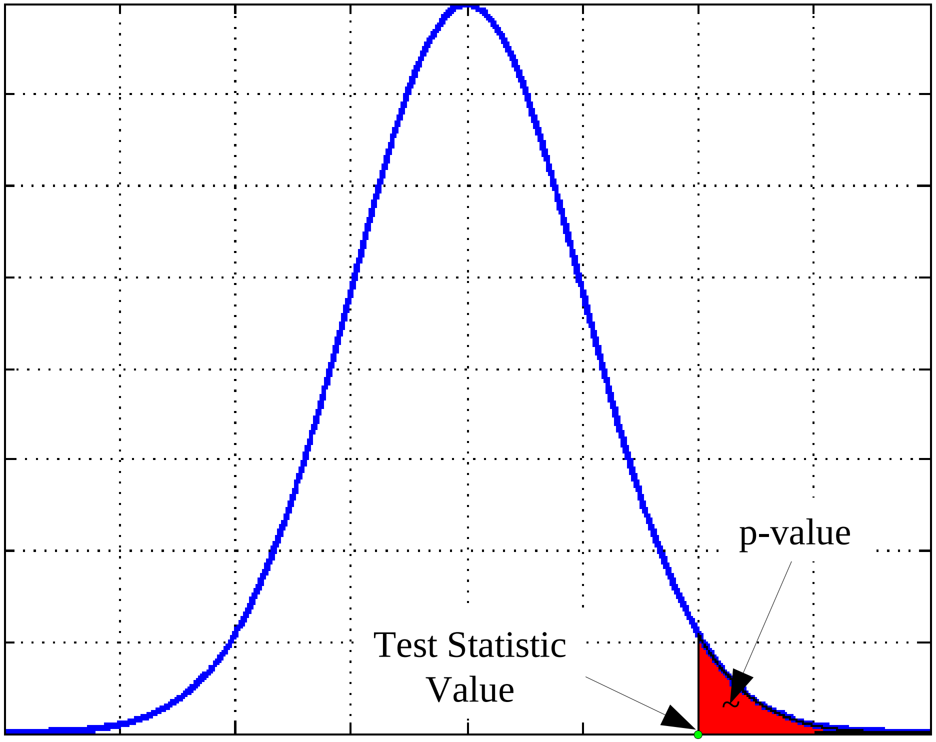
• Probability of obtaining an observation more extreme than the test statistic.

• Null hypothesis is rejected when the p-value is less than a user defined significance level (generally α=0.05).

• Smallest significance level for rejection of the null hypothesis.

• Result is declared “Statistically Significant” upon the rejection of the null hypothesis.

Normal Distribution



Range of Possible Observations

Generalized Linear Model

- Generalized linear model is an algorithm that employs least squares regression to fit data to a distribution belonging to exponential family (e.g. Binominal, Poisson, Bernoulli, Exponential, etc.)

$$f_Y(y|\theta,\tau) = h(y,\tau)\exp\frac{b(\theta)T(y)-A(\theta)}{d(\tau)}$$

- Information from a set of independent variables is incorporated into the model via a linear predictor.

$$\eta = \boldsymbol{X}\boldsymbol{\beta}$$

- A linking function, g, establishes the connection between the mean of the response variables, Y , and the linear predictor.

$$E[Y] = \mu = g^{-1}(\eta)$$

Generalized Linear Model

Estimation of Logit model coefficients:

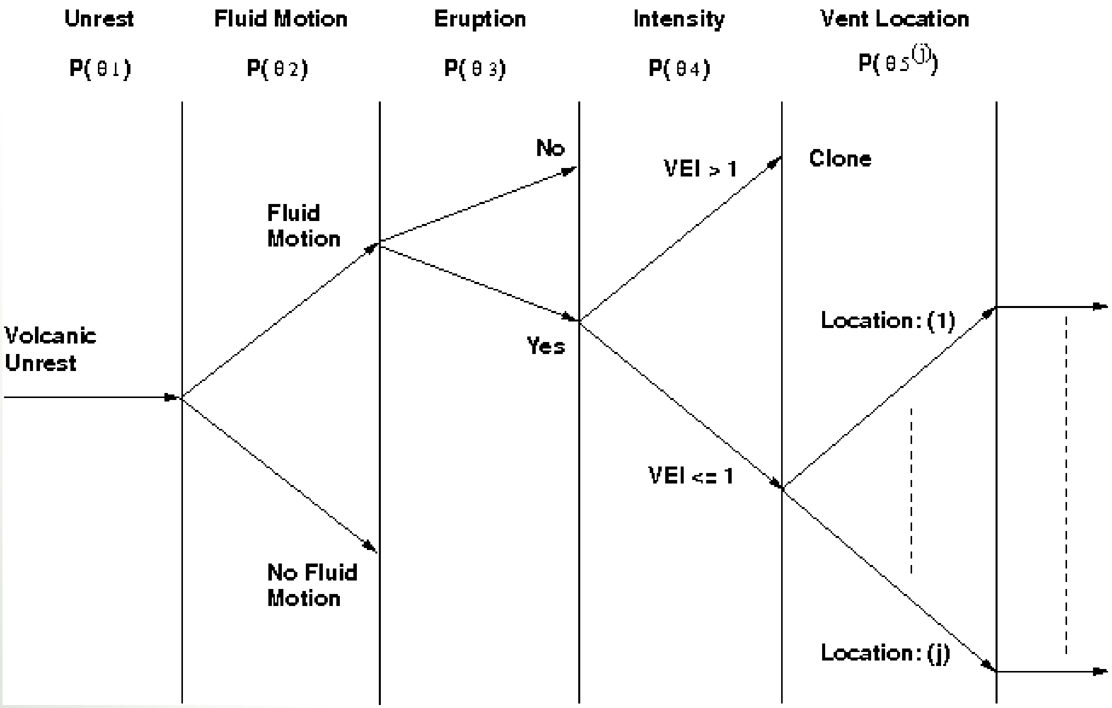
• Calculates a linear model that relates explanatory variables and observed outcomes that follow a binominal distribution via a logit linking function.

Linking Function	Mean
$\boldsymbol{X\beta} = \ln\frac{\mu}{1-\mu}$	$\boldsymbol{\mu} = \frac{1}{1+e^{-X\beta}}$

• Canonical linking function (e.g., Binominal, Poisson) implies $b(\mu)=\theta=\boldsymbol{X\beta}$ allowing X'Y to serve as a sufficient statistic for $\boldsymbol{\beta}$.

• The "*glmfit*" function in Matlab and R were used to compute the logit model coefficients.

Event Tree System



Likelihood Terms:

P(1|2) = 28/41

P(1|2')=13/41

P(2|3)=0.95

P(2|3')=17/41

P(3|4)=27/41

P(3|4')=14/41

$$P(\theta_n) = P(n|n-1) = \frac{P(n-1|n)\,P(n)}{P(n-1|n)\,P(n) + P(n-1|n')\,P(n')}$$

Bayes Theorem

Bootstrapping Analysis

- Ideally, the regression process is performed using many combinations of random samples from the population we are attempting to model.

- After many iterations, a distribution of parameter estimates, such as the sample mean or regression coefficients, can be produced and their true value estimated.

- In this case, however, there is no additional data available. Therefore, a bootstrapping approach is invoked to estimate the distribution of logistic model coefficients for nodes 2, 3, and 4.

- Bootstrapping is a resampling technique that produces M datasets from a single set of random samples taken from a specific population.

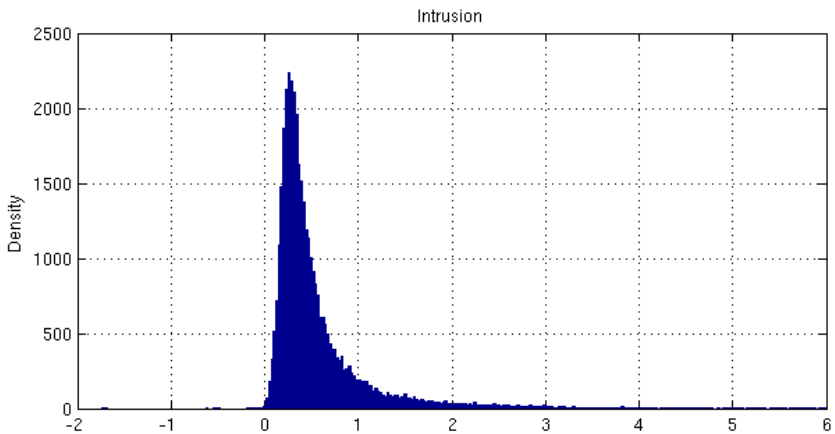
- Each of the newly constructed datasets contain a random combination of samples that were drawn from the original dataset.

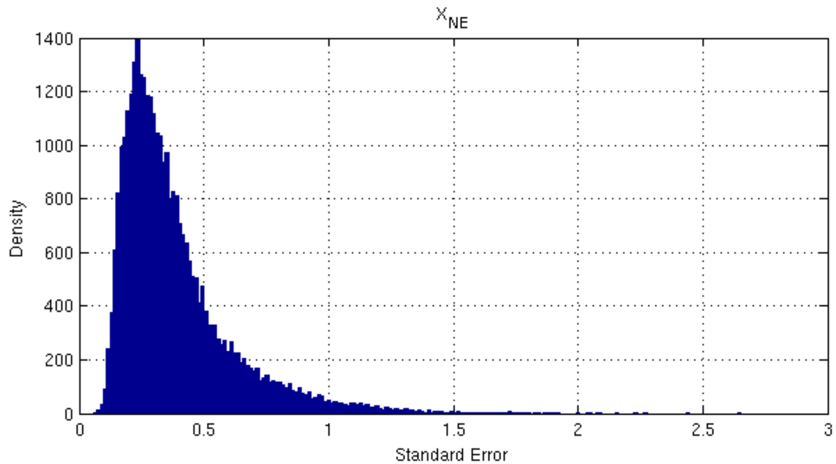
- Since a sample with replacement (Monte Carlo sampling) process is used, it is possible for each new dataset to contain multiple or no copies of any particular sample.

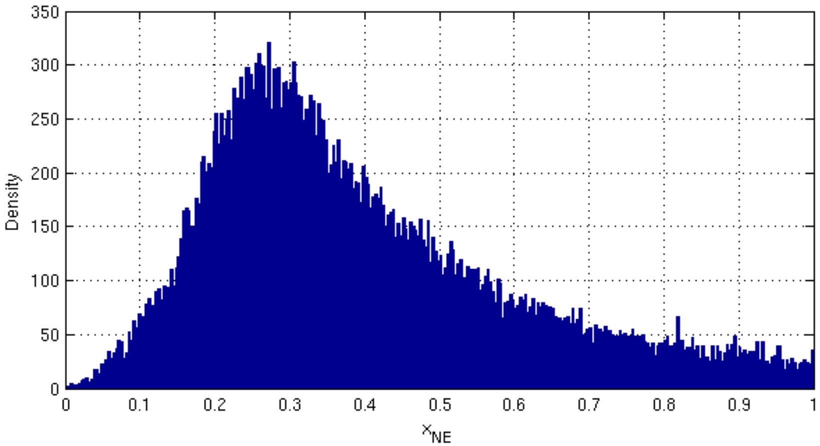
- Thus, the probability that a particular sample is selected for inclusion in a bootstrapped dataset each time a drawing is made is 1/N, where N is the number of samples in the original dataset.

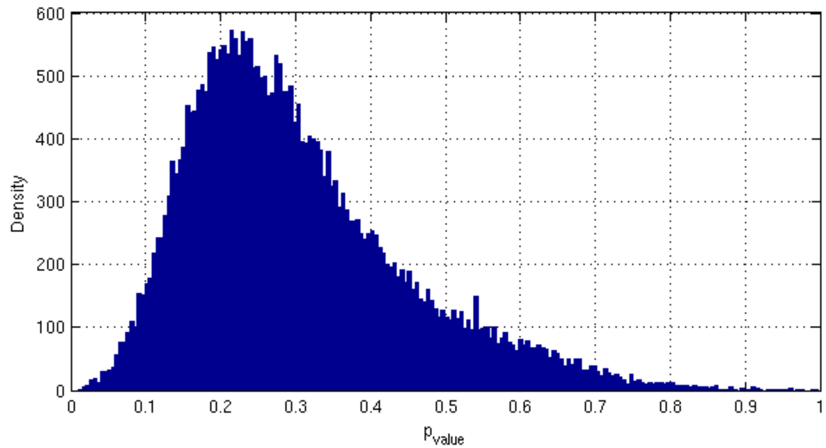
Bootstrapping Analysis: Node 2

- Node 2 (Fluid Motion): X_{NE}, standard error, and p-value distributions.



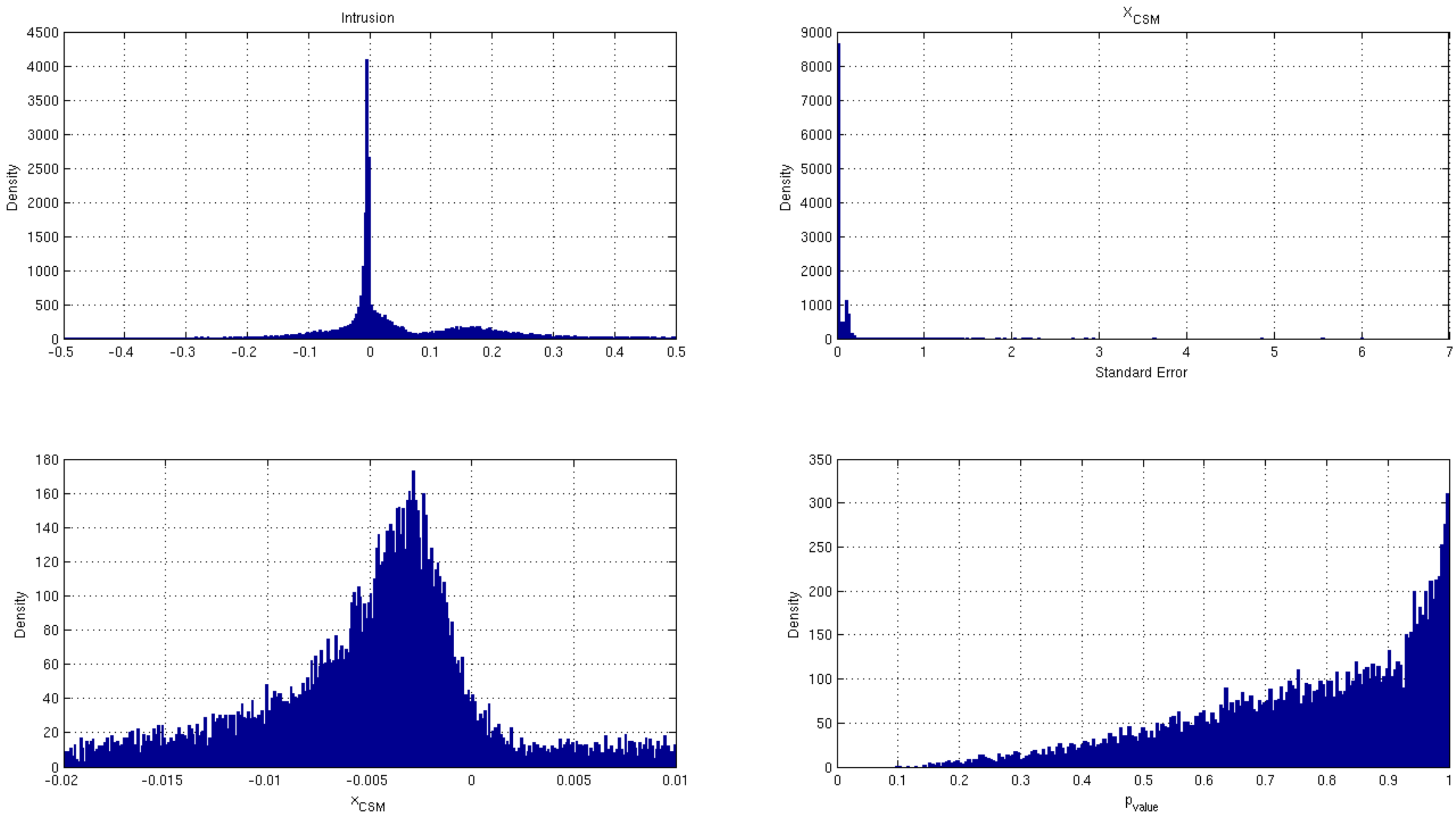






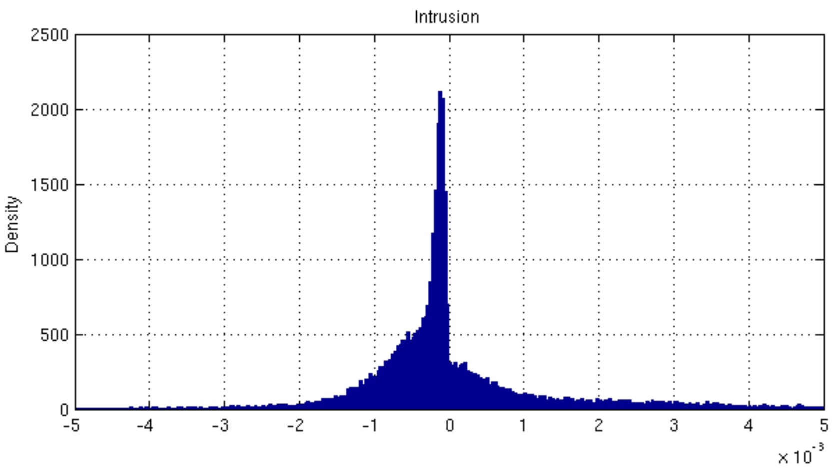
Bootstrapping Analysis: Node 2

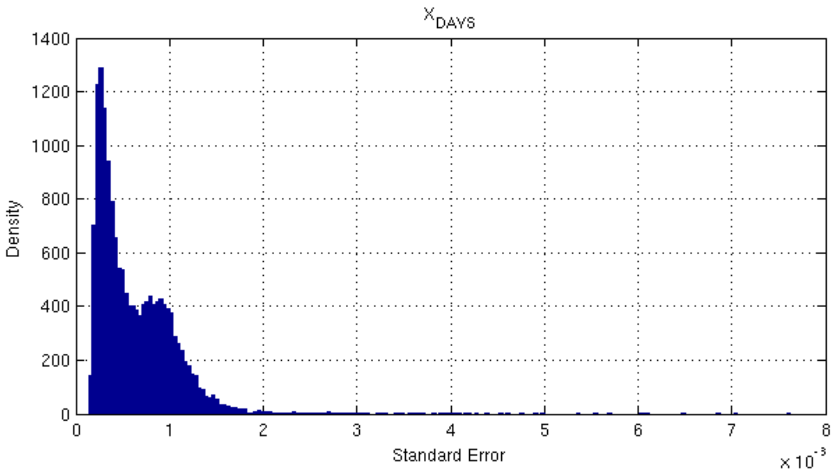
- Node 2 (Fluid Motion): X_{CSM}, standard error, and p-value distributions.

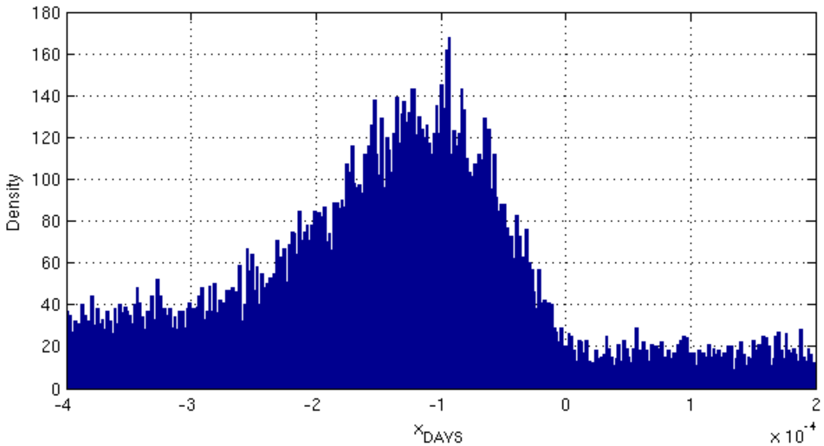


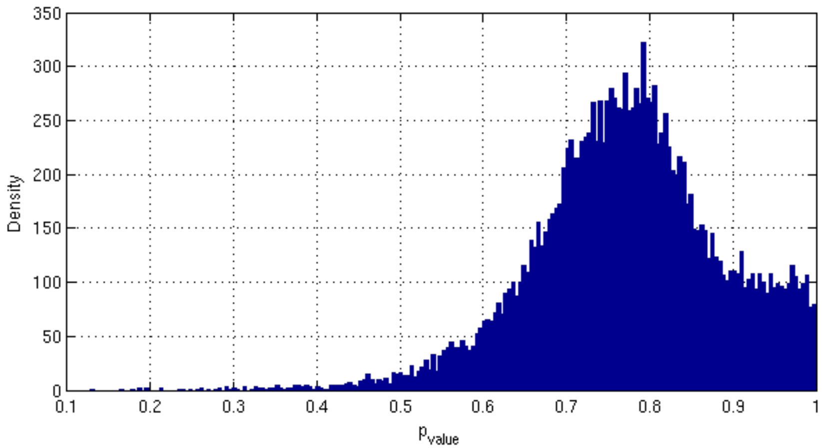
Bootstrapping Analysis: Node 2

- Node 2 (Fluid Motion): X_{DAYS}, standard error, and p-value distributions.

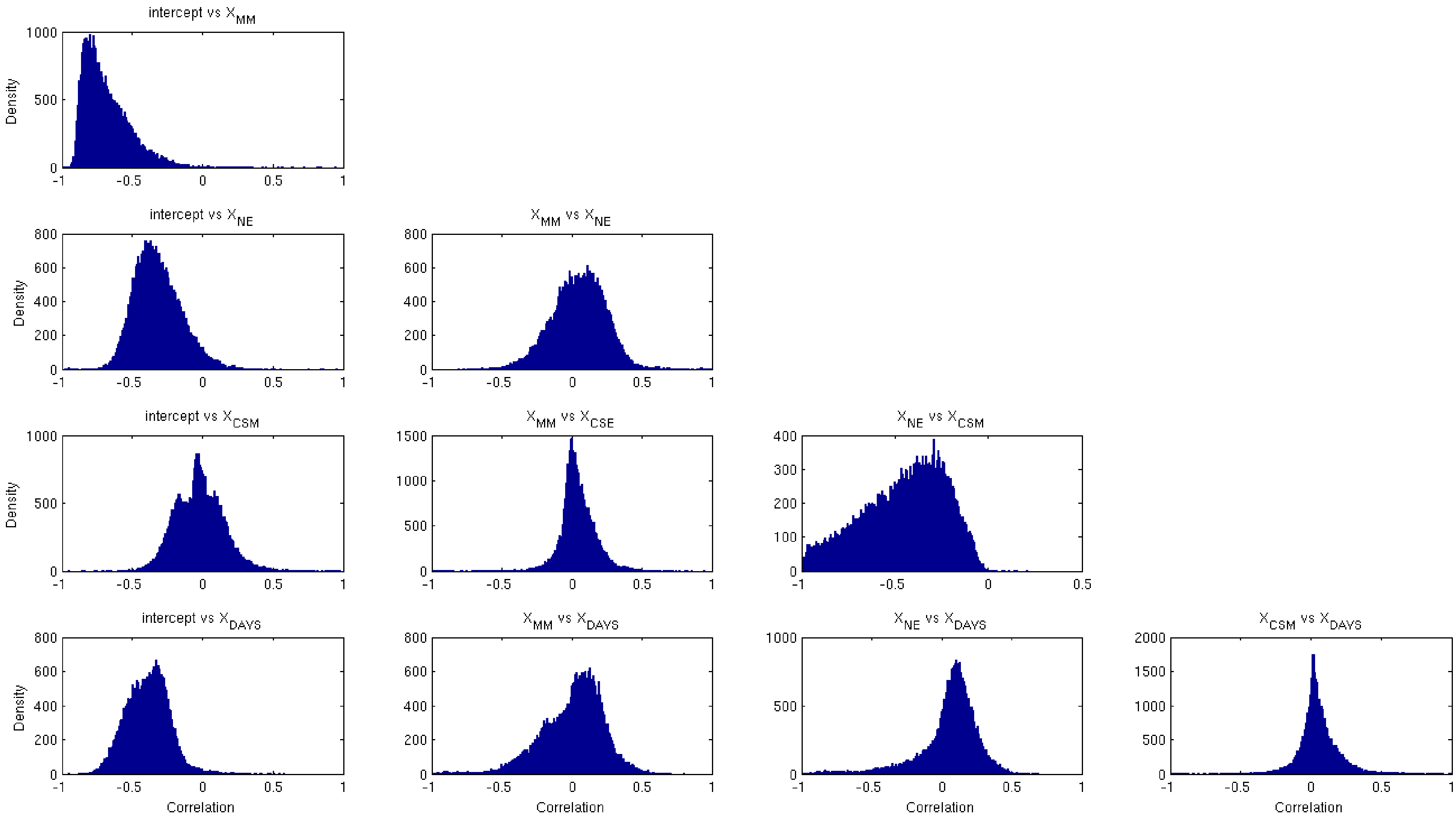








Bootstrapping Analysis: Node 2



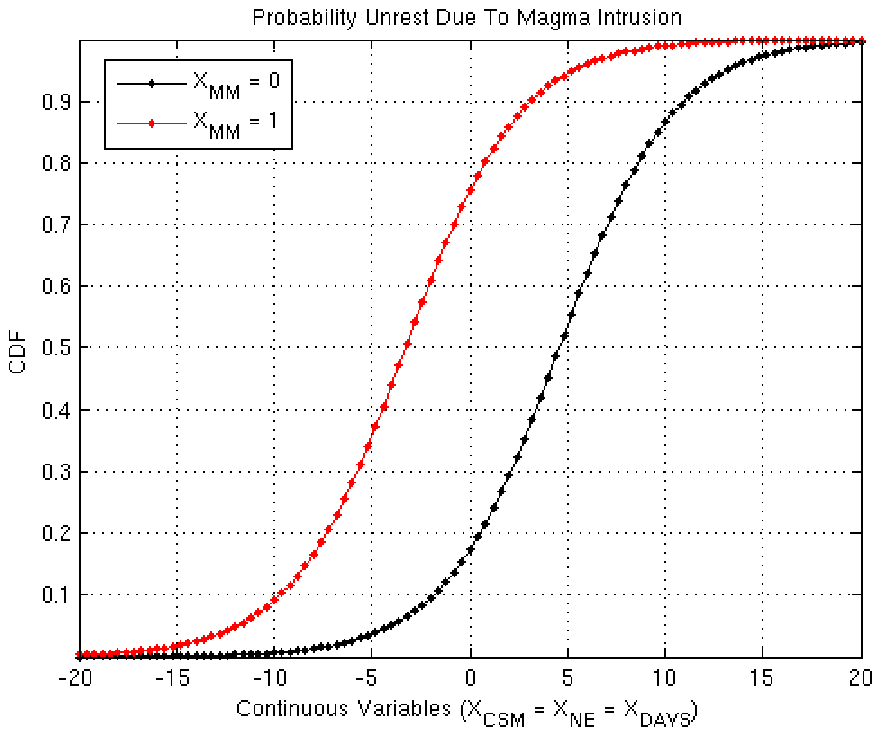
Node 2: Bootstrapping Results

Statistical Summary

Variable	Std Error	p Value* (Median)	p Value* (Mode)
Intercept	1.2611	0.1920	0.0573
X_{MM}	1.3350	0.0379	0.0082
X_{NE}	0.3226	0.2733	0.2149
X_{CSM}	0.0151	0.8061	0.9977
X_{DAYS}	0.005	0.7754	0.7929

Correlation Coefficients

Variable	Intercept	X_{MM}	X_{NE}	X_{CSM}	X_{DAYS}
Intercept	1.00	-0.77	-0.36	-0.04	-0.40
X_{MM}	-0.77	1.00	0.09	0.04	0.07
X_{NE}	-0.36	0.09	1.00	-0.37	-0.12
X_{CSM}	-0.04	0.04	-0.37	1.00	0.05
X_{DAYS}	-0.40	0.07	-0.12	0.05	1.00



p value = $[P(\chi^2 > G)] = 0.001$

Logistic Model

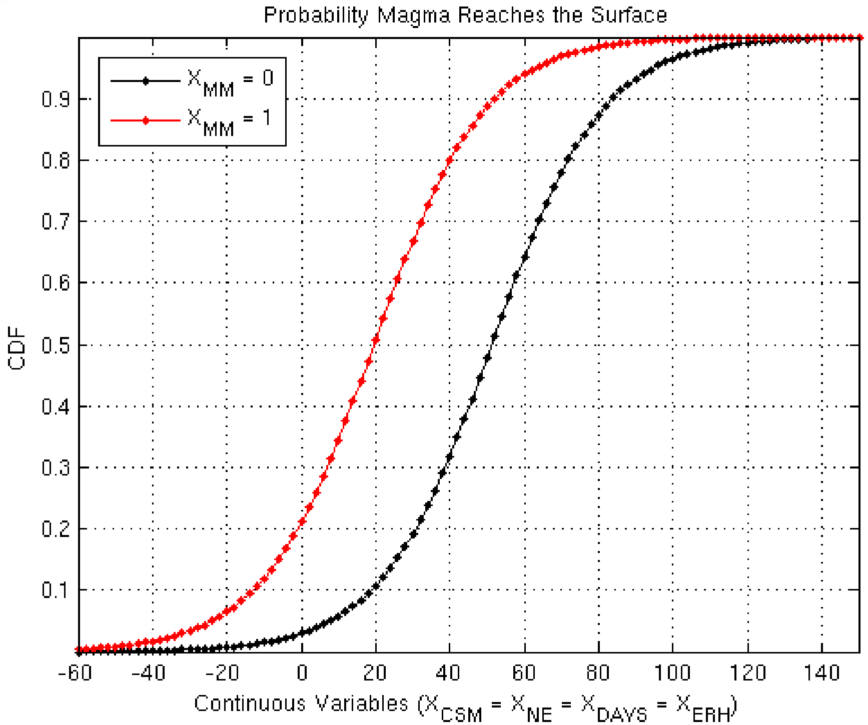
$$z = 2.6869(X_{MM}) + 0.3465(X_{NE}) + 0.0041(X_{CSM}) - 0.0001(X_{DAYS}) - 1.5618$$

*Ho: $\beta_n = 0$, Ha: $\beta_n \neq 0$

Node 3: Bootstrapping Results

Statistical Summary

Variable	Std Error	p Value* (Median)	p Value* (Mode)
Intercept	2.2786	0.1756	0.0657
XMM	1.9030	0.6525	0.1699
XNE	0.0332	0.7675	0.9378
XCSM	0.0079	0.7723	0.7181
XDAYS	0.0017	0.4478	0.4046
XERH	9.1913	0.1533	0.1151



p value = $[P(\chi^2 > G)] = 0.014$

Correlation Coefficients

Variable	Intercept	XMM	XNE	XCSM	XDAYS	XERH
Intercept	1.00	-0.89	-0.26	-0.13	-0.29	-0.71
XMM	-0.89	1.00	0.04	0.17	0.06	0.51
XNE	-0.26	-0.04	1.00	-0.33	0.11	0.26
XCSM	-0.13	0.17	-0.33	1.00	0.04	0.01
XDAYS	-0.29	0.06	-0.11	0.04	1.00	0.21
XERH	-0.71	0.51	0.26	0.01	0.21	1.00

Logistic Model

$$z = 2.1401(X_{MM}) - 0.0056(X_{NE}) + 0.0023(X_{CSM}) - 0.0014(X_{DAYS}) + 12.8714(X_{ERH}) - 3.4589$$

*Ho: $\beta_n = 0$, Ha: $\beta_n \neq 0$

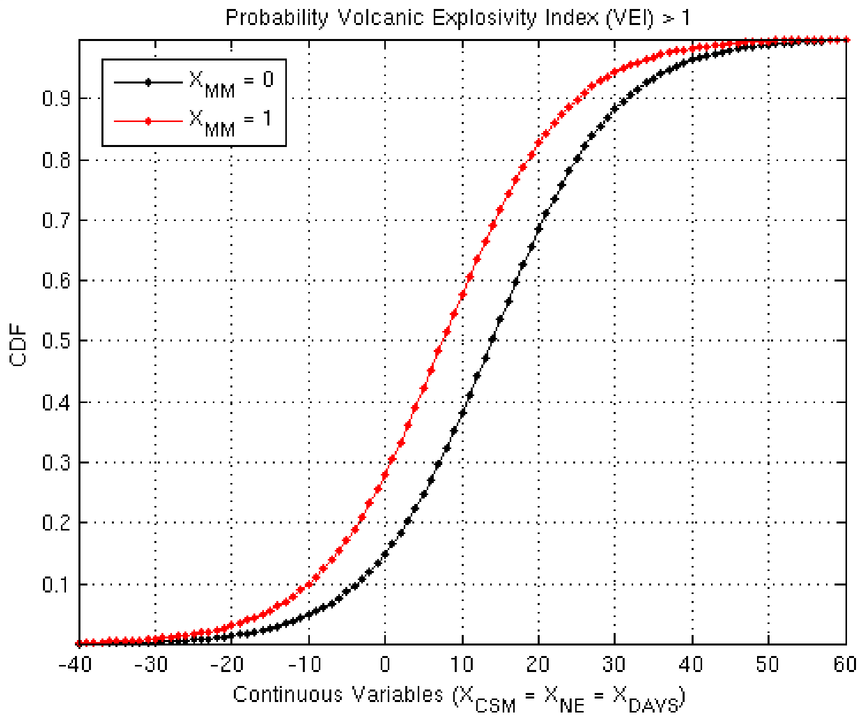
Node 4: Bootstrapping Results

Statistical Summary

Variable	Std Error	p Value* (Median)	p Value* (Mode)
Intercept	1.3550	0.1977	0.0605
X_{MM}	1.3594	0.3886	0.1118
X_{NE}	0.0761	0.0835	0.0438
X_{CSM}	0.0050	0.8719	0.9803
X_{DAYS}	0.0022	0.1296	0.0352

Correlation Coefficients

Variable	Intercept	X_{MM}	X_{NE}	X_{CSM}	X_{DAYS}
Intercept	1.00	-0.78	-0.51	-0.04	-0.21
X_{MM}	-0.78	1.00	0.29	0.06	0.10
X_{NE}	-0.51	0.29	1.00	-0.27	0.23
X_{CSM}	-0.04	0.06	-0.27	1.00	0.20
X_{DAYS}	-0.21	0.10	-0.23	0.20	1.00



p value = [P(χ^2 > G)] = 0.007

Logistic Model

$$z = 0.7924(X_{MM}) + 0.1293(X_{NE}) + 0.0006(X_{CSM}) - 0.0033(X_{DAYS}) - 1.7369$$

*Ho: $\beta_n = 0$, Ha: $\beta_n \neq 0$

Cross Validation

- Binary classification problems attempt to categorize the outcome of an event into one of two categories (true or false).

- This process can result in one of four possible outcomes:

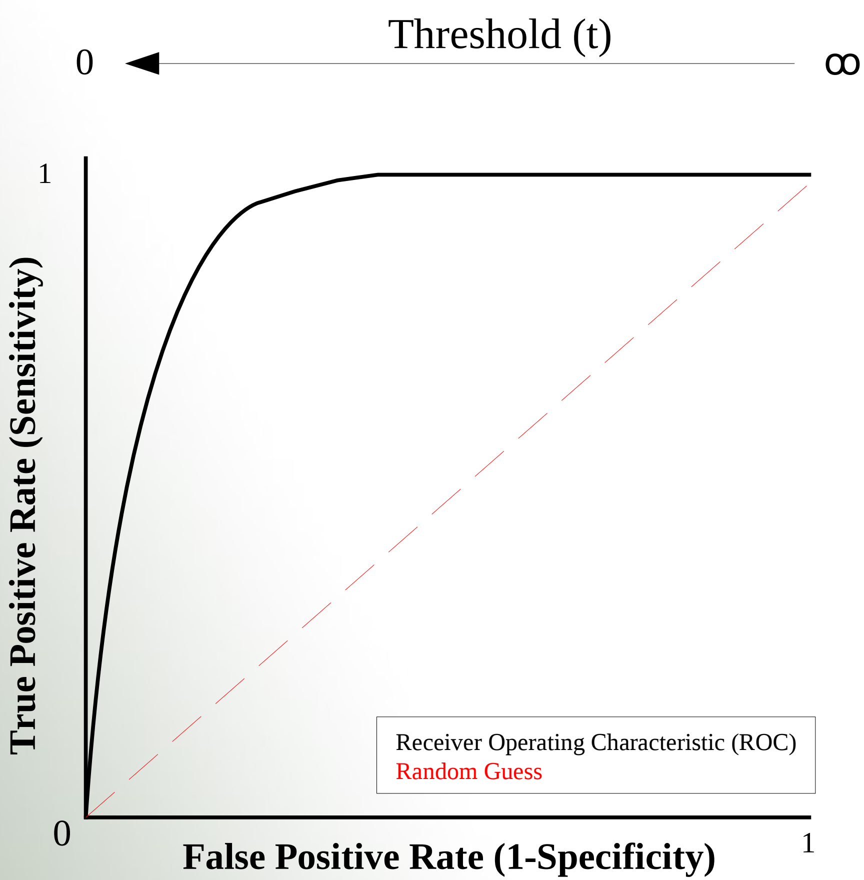
  - True Positive (*TP*): Predicted outcome matches the actual outcome.

  - False Positive (*FP*): Predicted result is true, but the actual result is false (i.e., type I error: incorrectly reject the null hypothesis).

  - False Negative (*FN*): Predicted result is false, but the actual result is true (i.e., type II error: incorrectly accept the null hypothesis).

  - True Negative (*TN*): Predicted and actual results are false.

Cross Validation

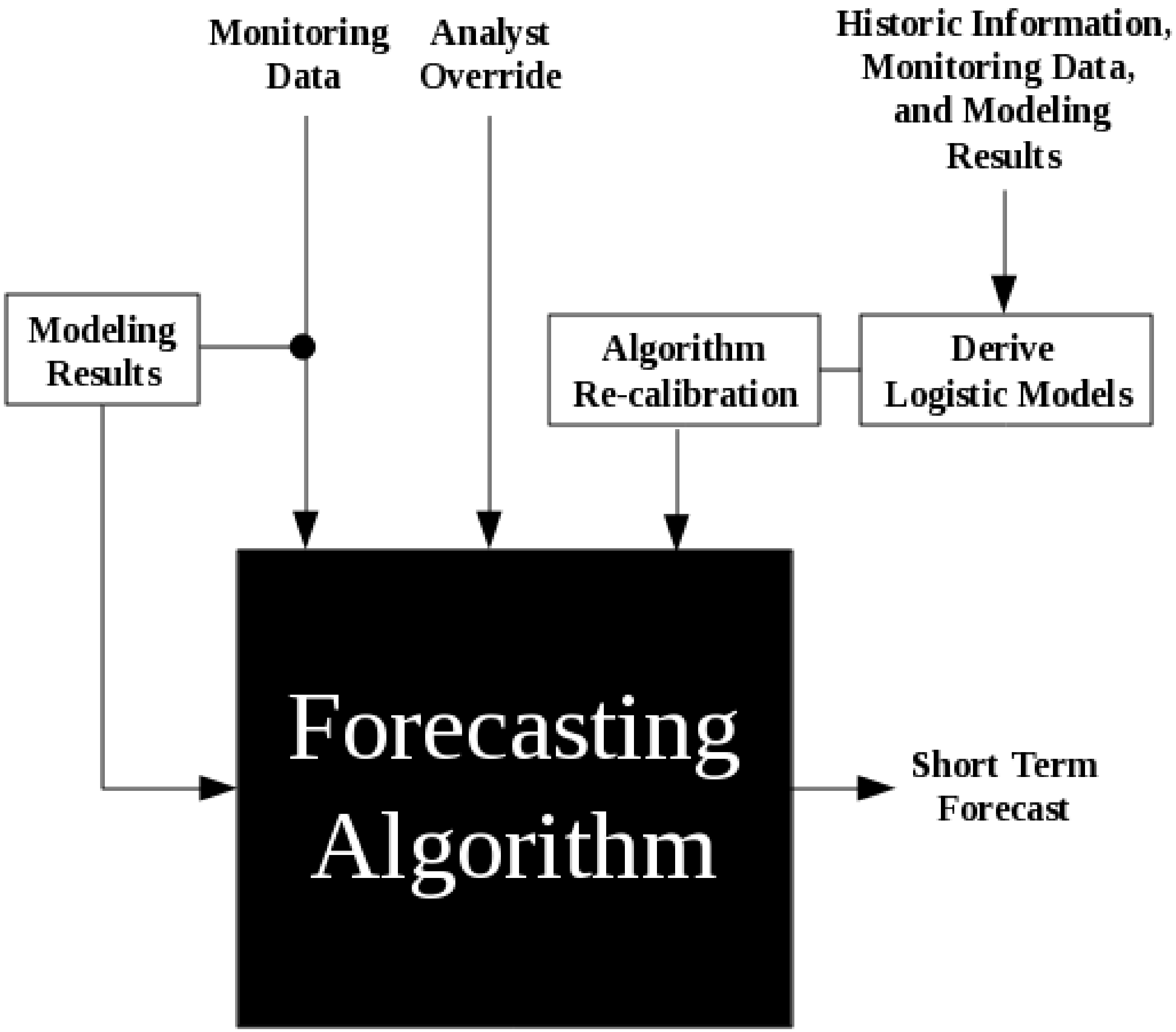


- The quality of a binary classifier is assessed through a receiver operating characteristic (ROC) analysis.
- A ROC curve is generated by plotting the prediction algorithm's true positive rate (TPR or sensitivity) versus its false positive rate (FPR or 1-specificity).
- The algorithms predictive capability is quantified by the area under the ROC (AUROC) curve.
- Its predictive power increases as the AUROC approaches 1.0 and decreases as its approaches 0.0.
- An AUROC of 0.5 is equivalent to randomly selecting an outcome.

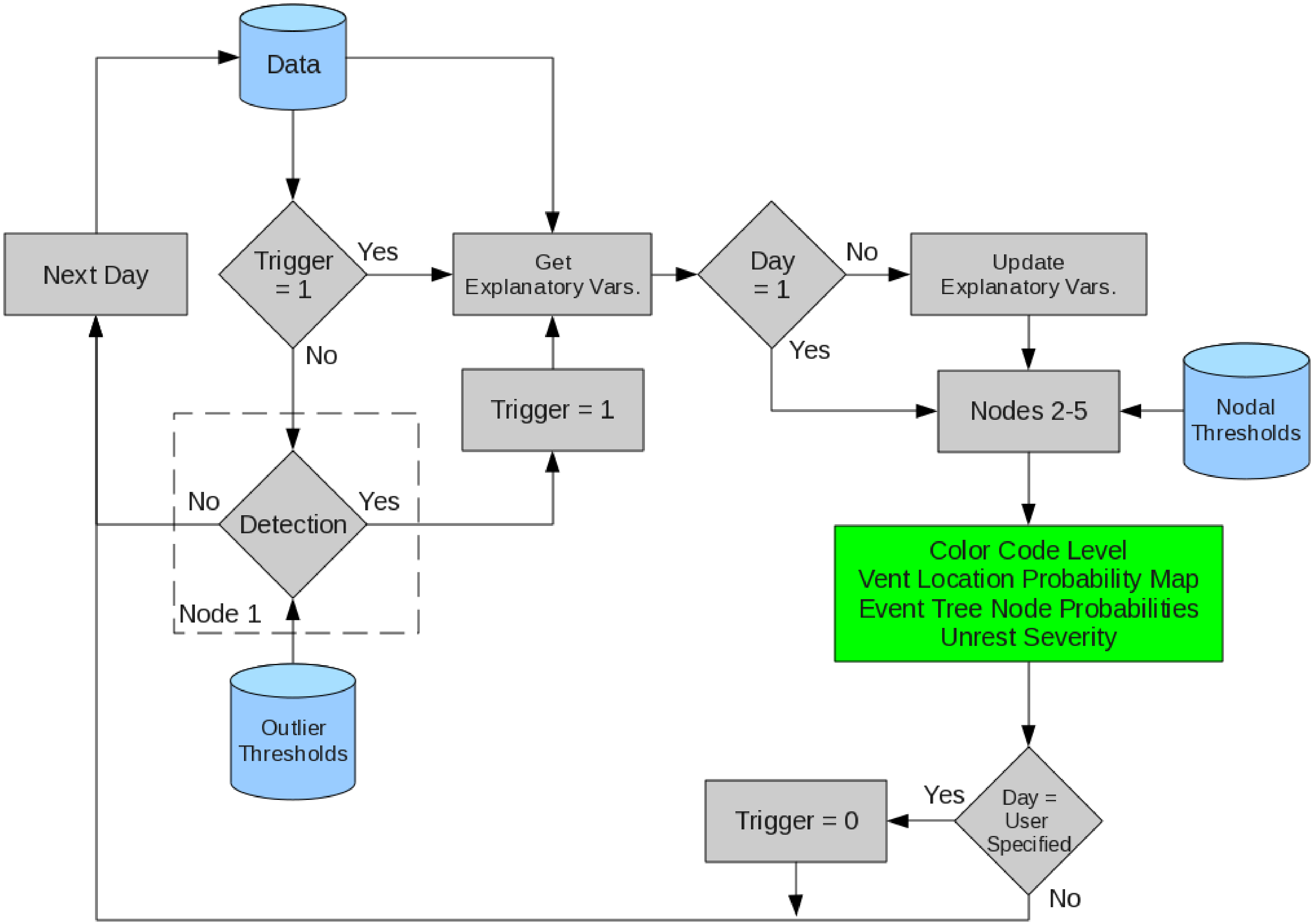
$$Sensitivity(t) = \frac{TP(t)}{TP(t)+FN(t)}$$

$$Specificity = \frac{TN(t)}{TN(t)+FP(t)}$$

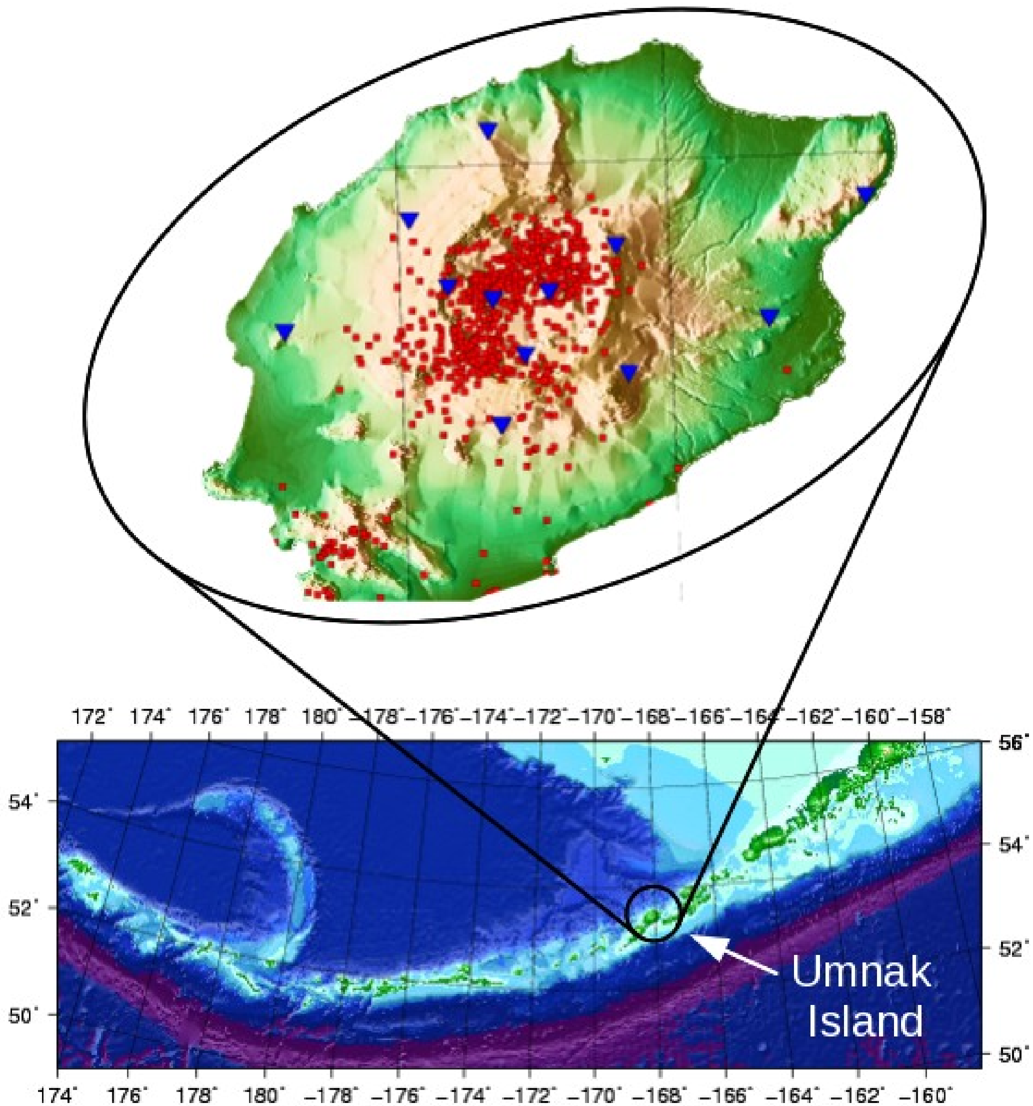
Implementation



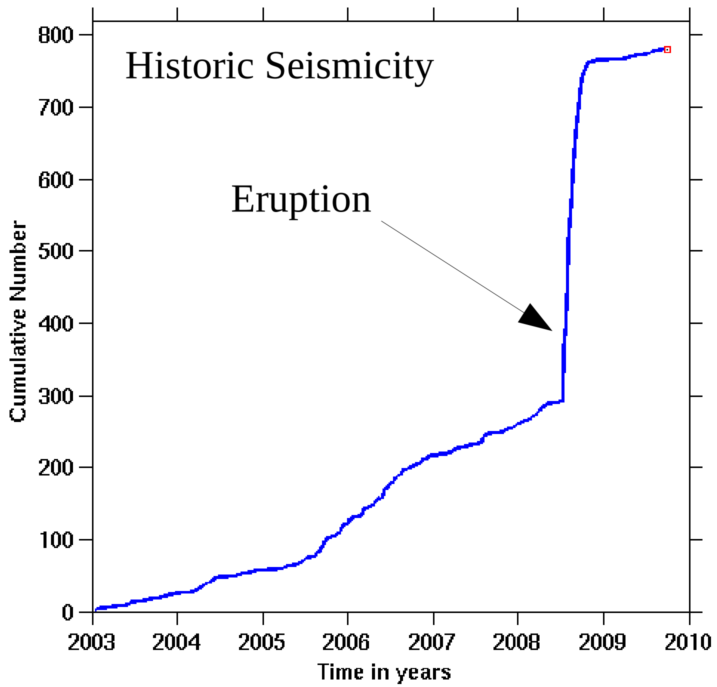
Implementation



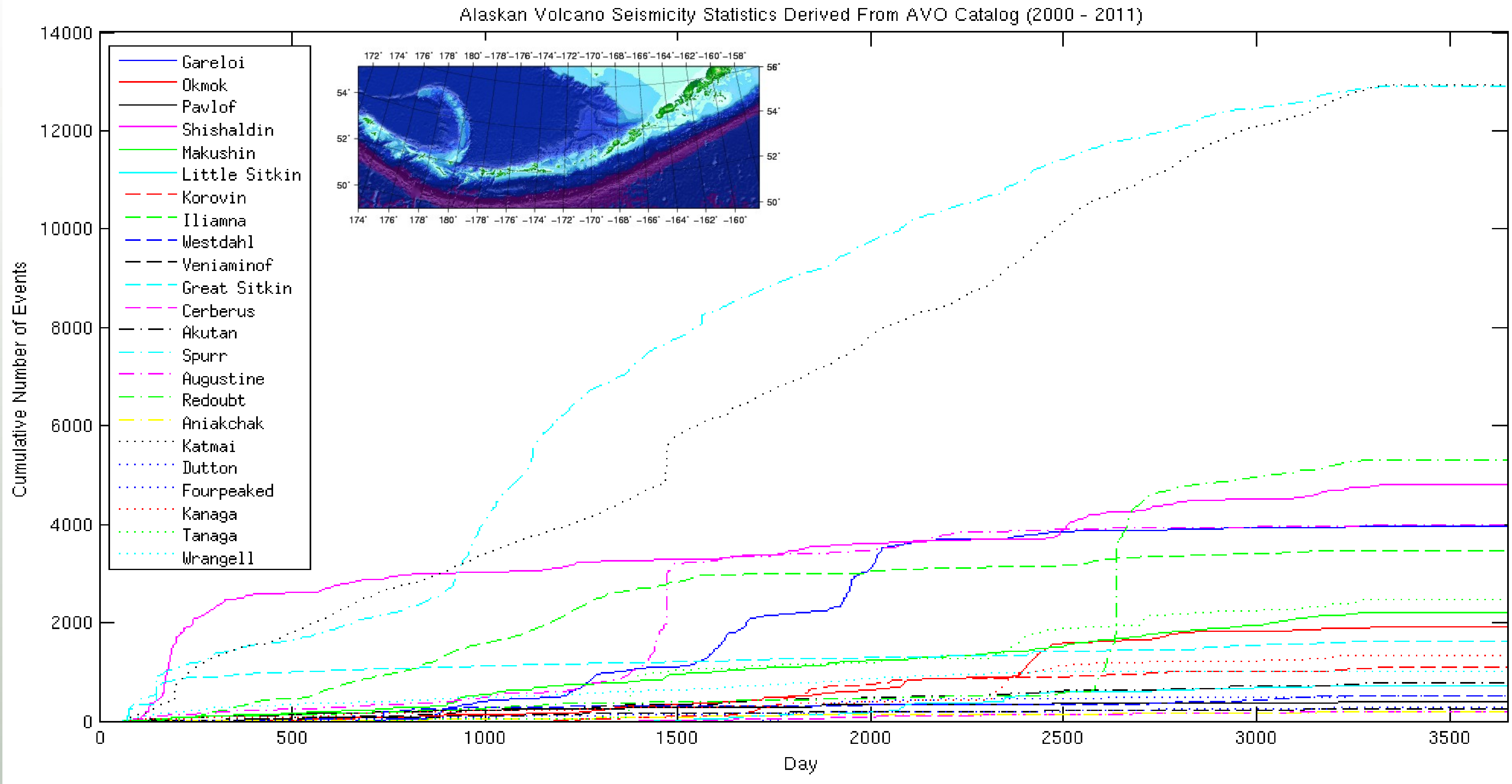
Results: Okmok



- Active shield volcano located near the center of the Aleutian arc.
- Most recent eruptions: May 1997 & July 2008.
- Seismic activity is monitored continuously by the University of Alaska Fairbanks/USGS Alaska Volcano Observatory (AVO).

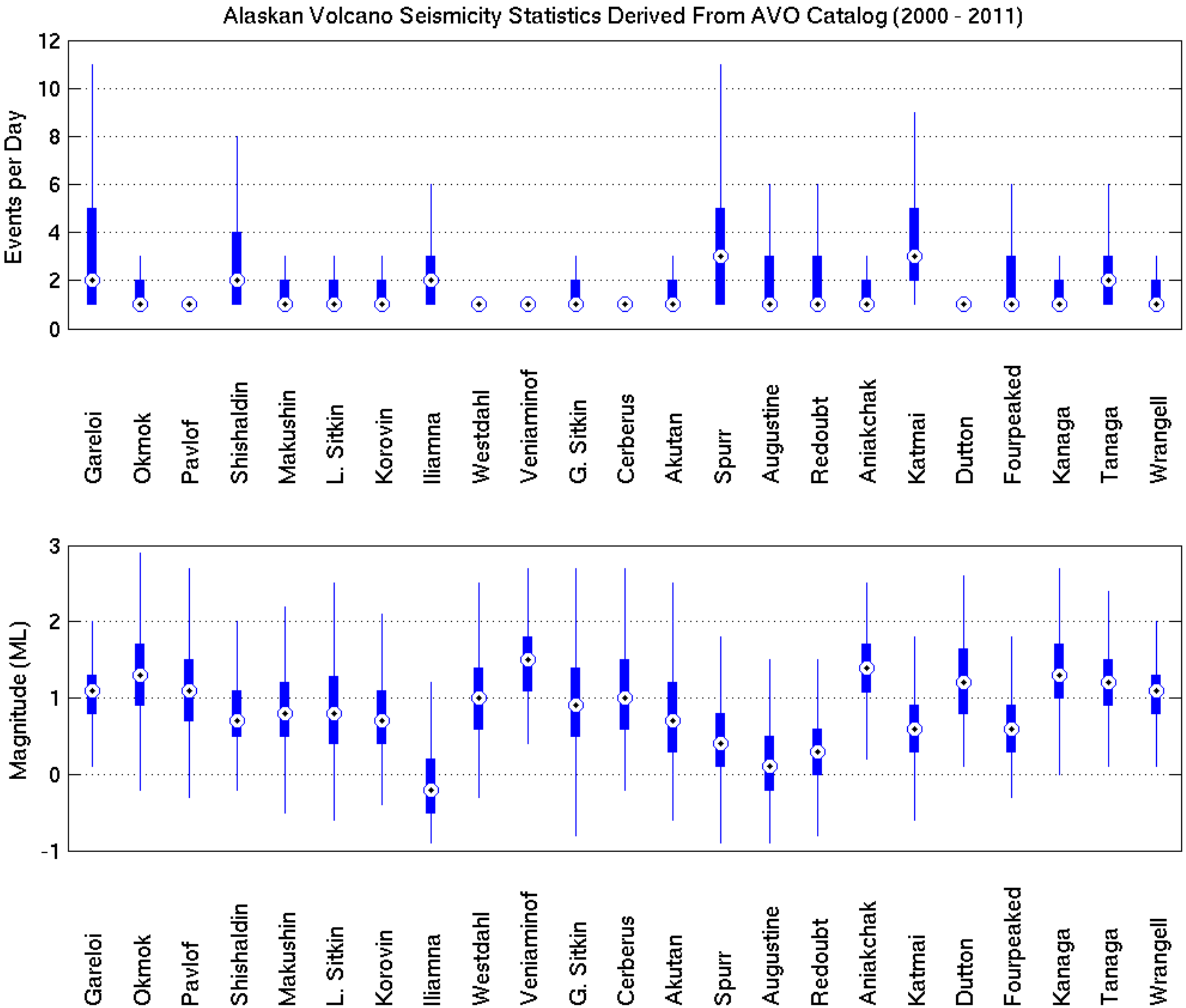


Results: Okmok



- Seismicity varies with time
- Each unrest episode is unique

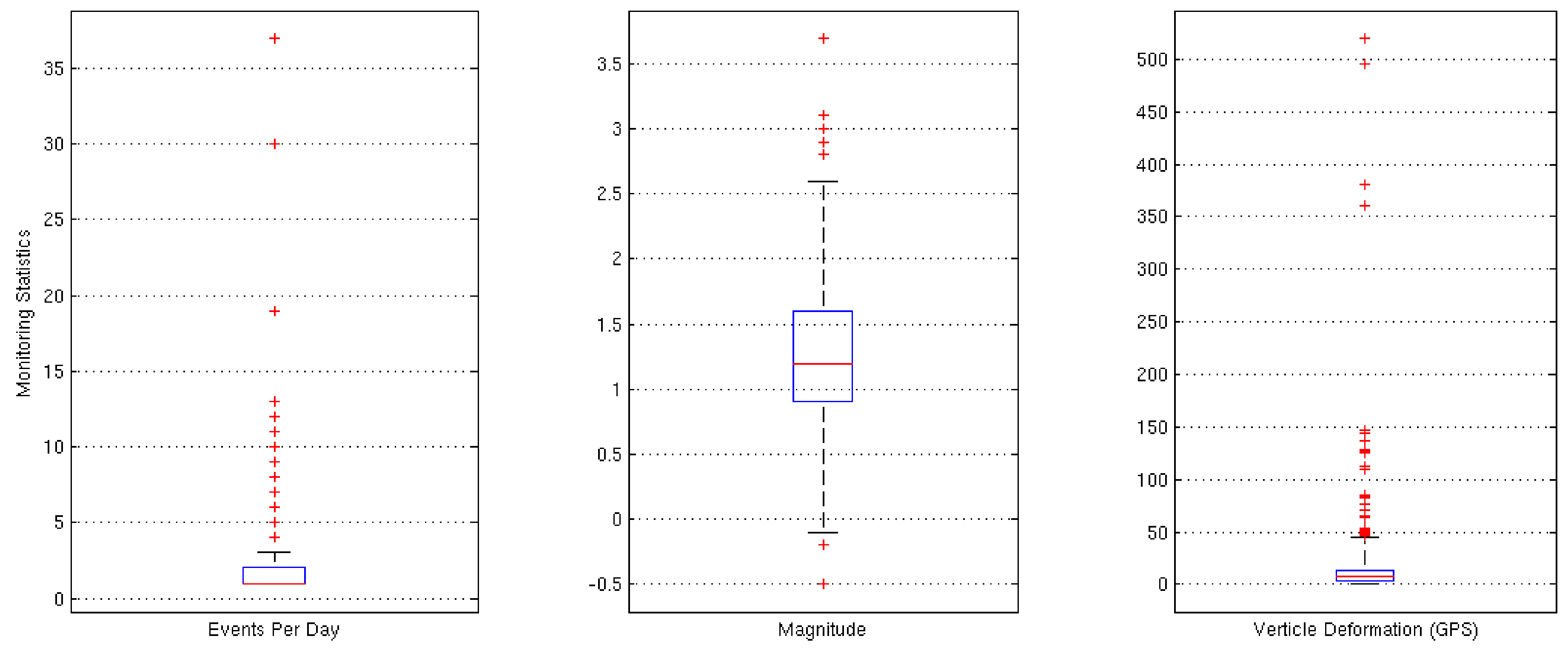
Results: Okmok



Volcano Seismicity:

- Seismicity varies from volcano to volcano
- No two volcanoes behave exactly the same
- Unrest thresholds must be determined on a volcano by volcano basis
- Historic observations must be leveraged for this purpose.
- Outlier analysis allows the the determination of anomalous values that could serves a an unrest detection threshold.

Okmok: Results

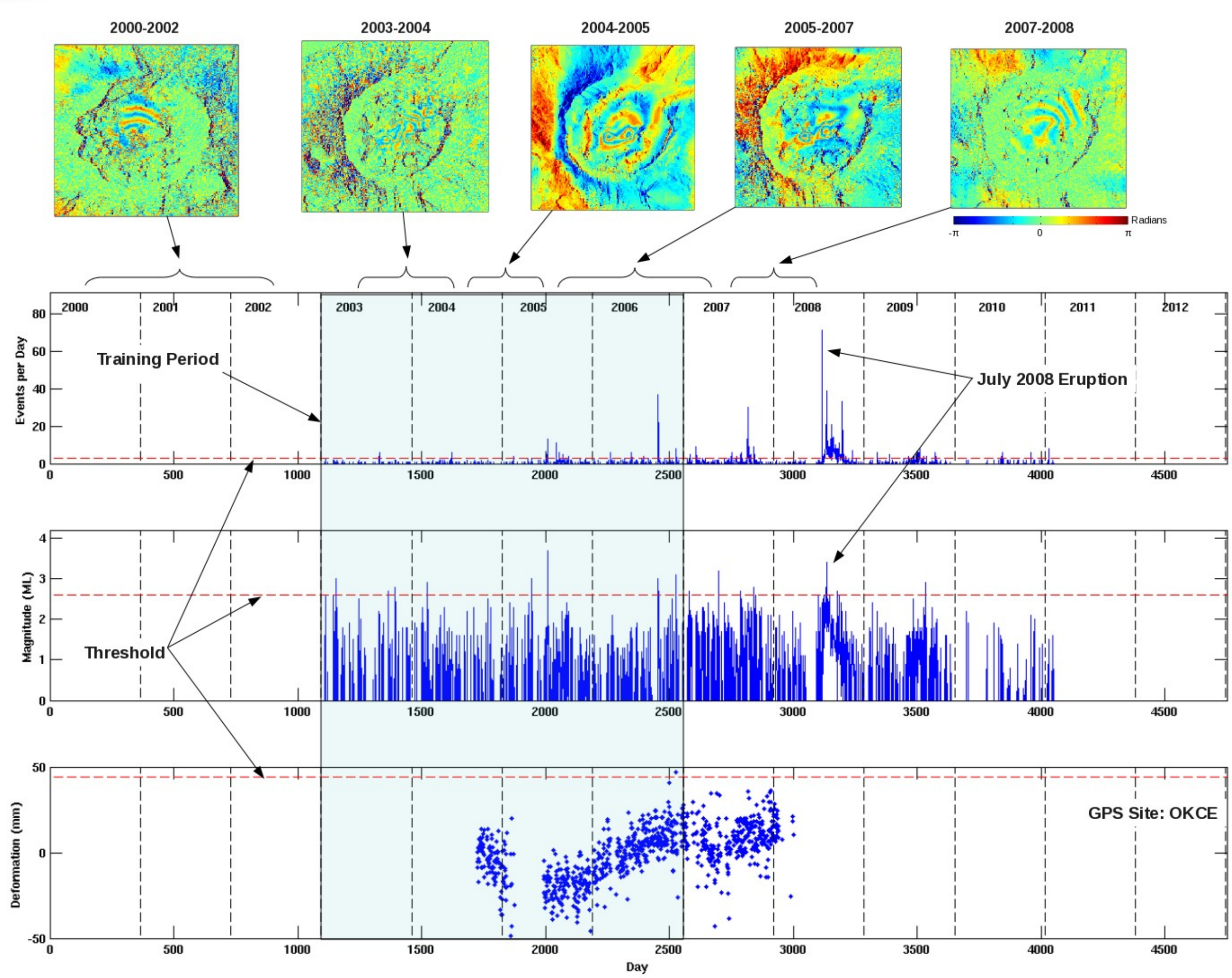


- Boxplots highlighting the distribution of seismicity and deformation on Umnak island between 2003 and 2008.

- Outlier whiskers for events per day and magnitude whiskers are set to 1.5 and the vertical GPS deformation measurements are 3 times the inter quartile range.

- Unrest thresholds are 3.0 events per day, ml of 2.6, and 44.3 mm of deformation.

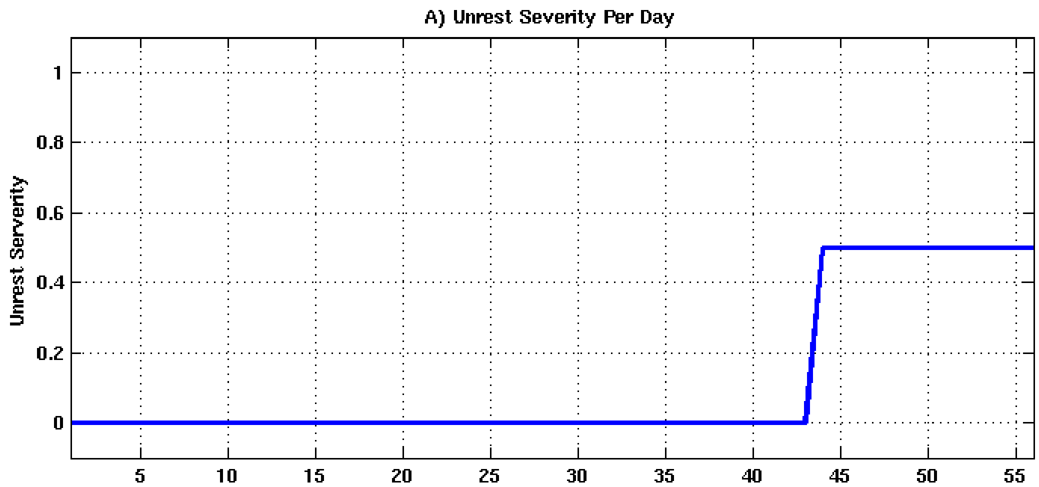
Results: Okmok

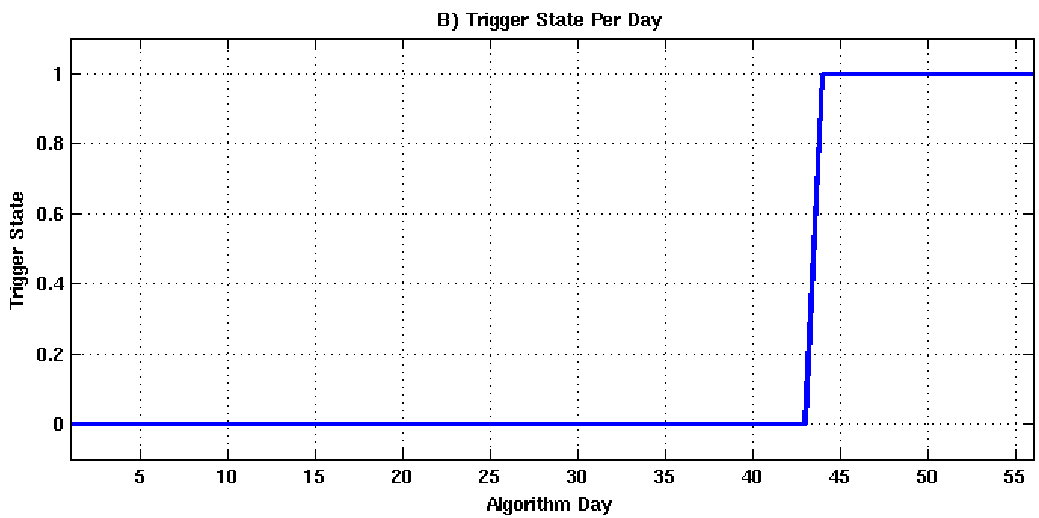


Monitoring Data:

- Time series of monitoring data acquired from the Okmok volcano between 2000 and 2011
- Red lines are the outlier thresholds derived from data in the blue window.
- GPS data shows vertical displacement between late 2004 and early 2008.

Results: Okmok



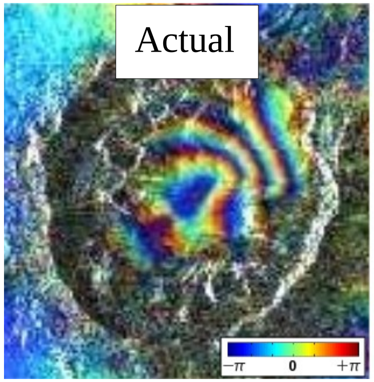


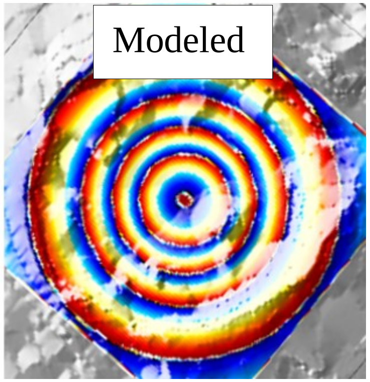
Trigger State Vector:

• Trigger state =1, means the algorithm has detected unrest and is actively generating forecasts.

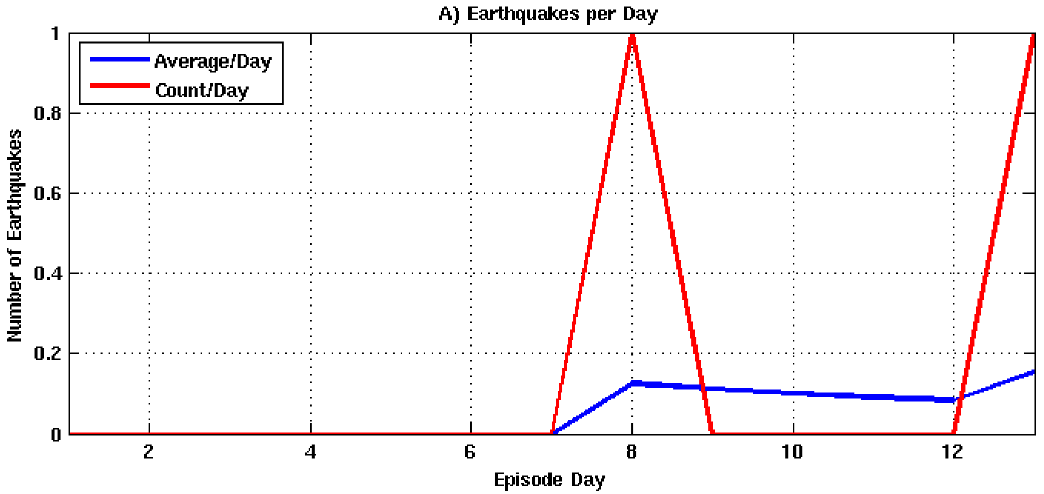
• Trigger state transitions from 0 to 1 on algorithm day 43.

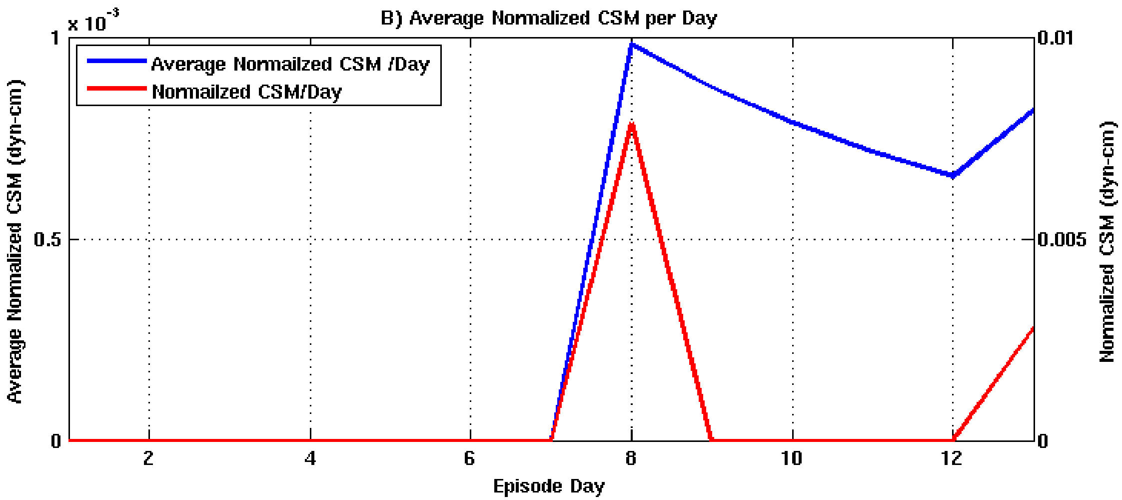
• Surface deformation and positive source modeling results (consistent with magmatic intrusion source model) trigger algorithm.





Results: Okmok

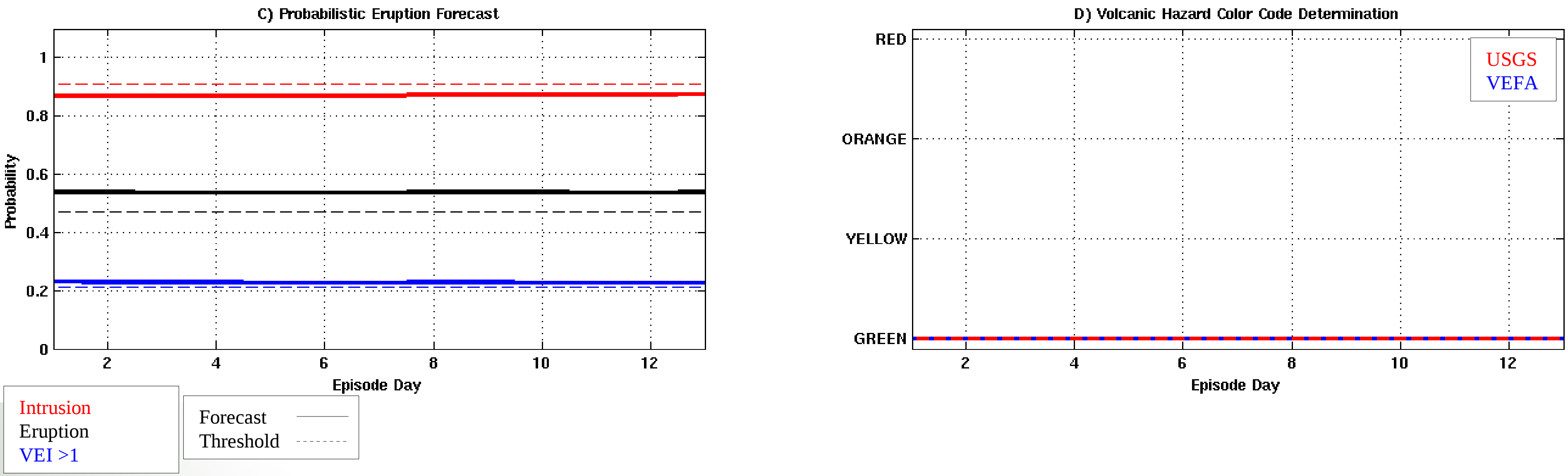




Input Data:

- Volcano monitoring data preceding Okmok's 2008 eruption.
- Count and average (X_{NE}) number of earthquakes per episode day (X_{DAYS}), shown in blue and red.
- Count and average (X_{CSM}) seismic moment per episode day (X_{DAYS}), shown in blue and red.
- Historic eruption frequency index (X_{ERH}) value = 16
- No anomalous precursory seismicity is detected.

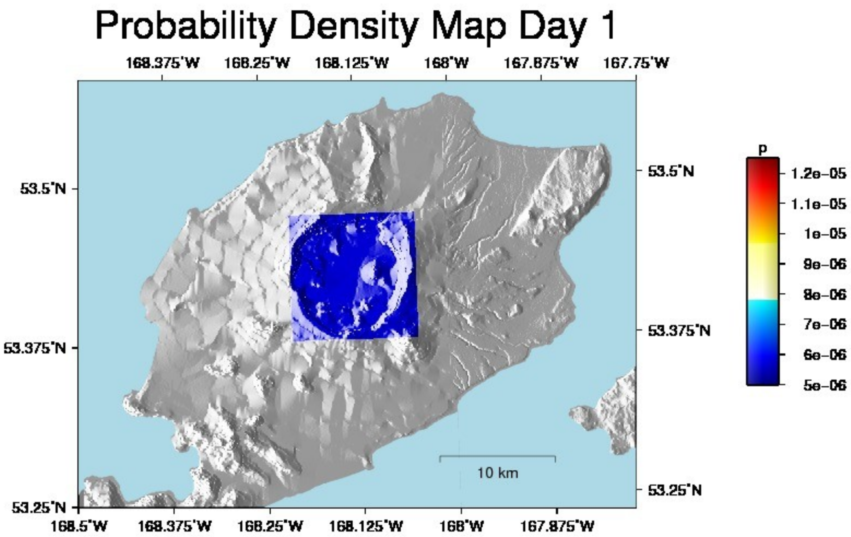
Results: Okmok

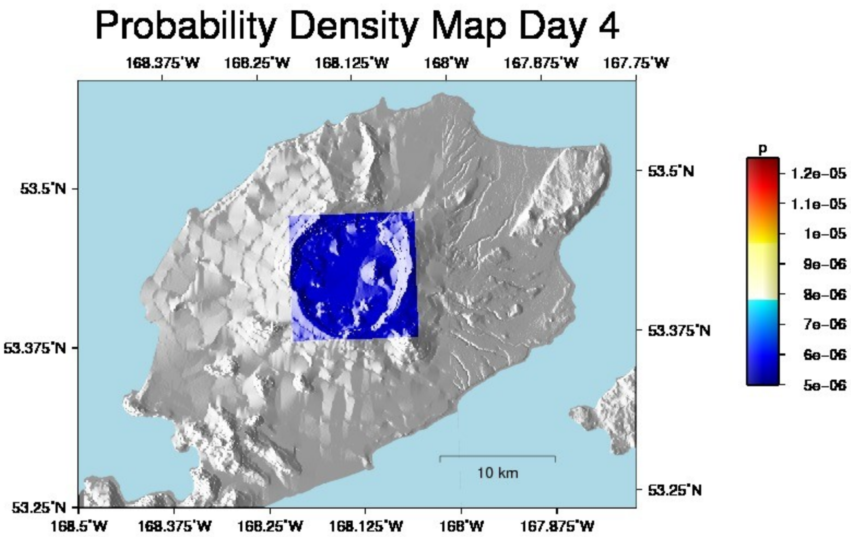


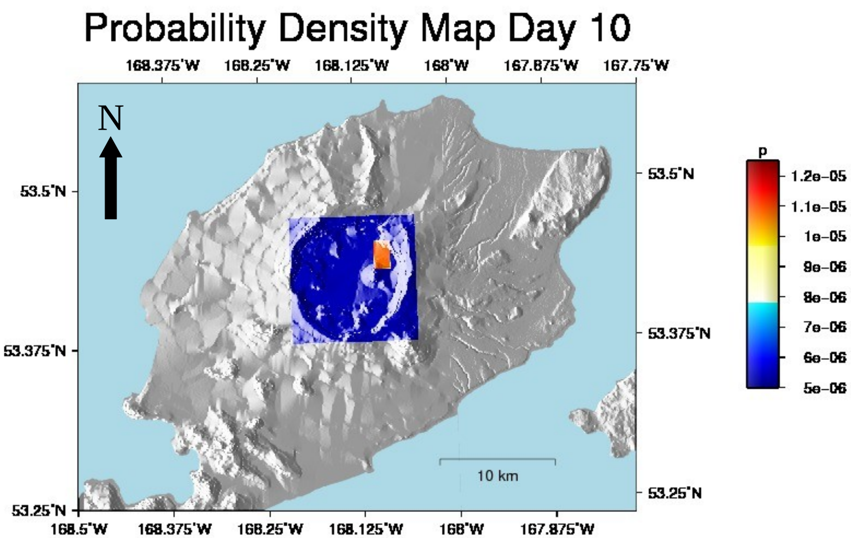
Forecasts:

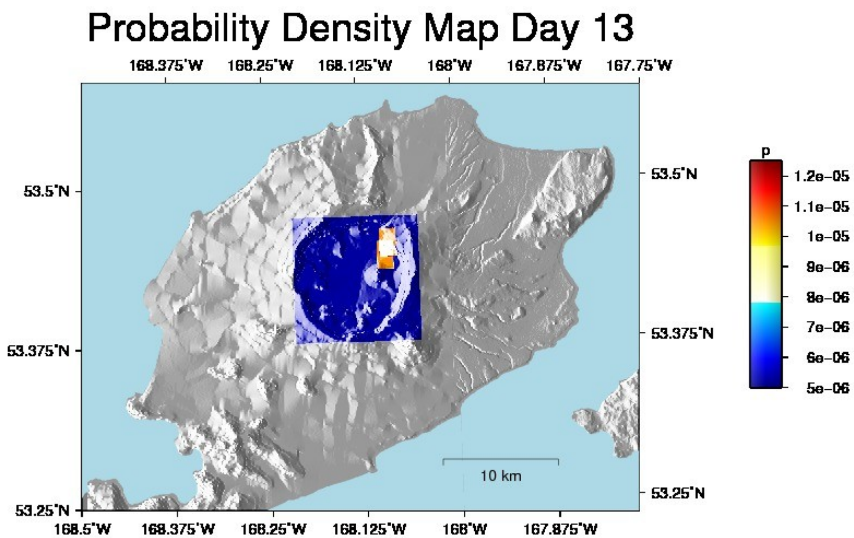
- Eruption probabilities remain relatively constant at 0.88, which is just below the detection threshold.
- The probability Okmok will erupt with an intensity greater than VEI 1.0 is approximately 0.23.
- Hazard declaration remains green throughout the episode.
- Eruption forecast is complicated by the absence of precursory seismicity.

Okmok: Results

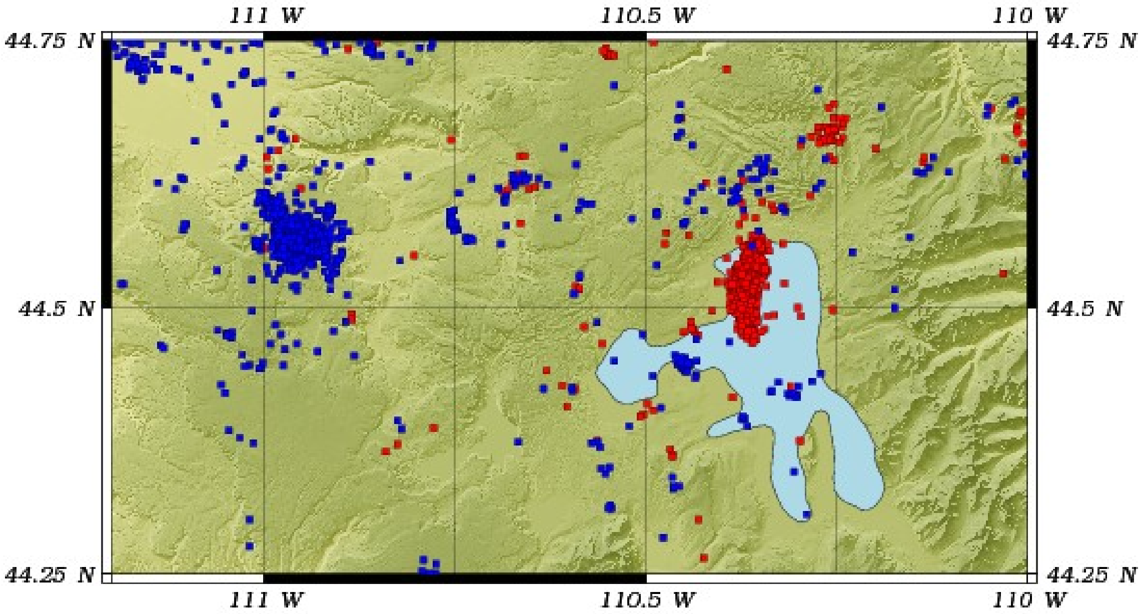








Results: Yellowstone



Yellowstone:

- The Yellowstone Caldera and Snake River valley are well known for displaying elevated signs of magmatic and tectonic activity.

- Between 2005 and 2010 an unprecedented episode of magmatic unrest took place within the Yellowstone caldera.

- Through extensive source modeling, this episode was determined to be the result of a complex magmatic intrusion occurring beneath the park (Chang et al. 2010).

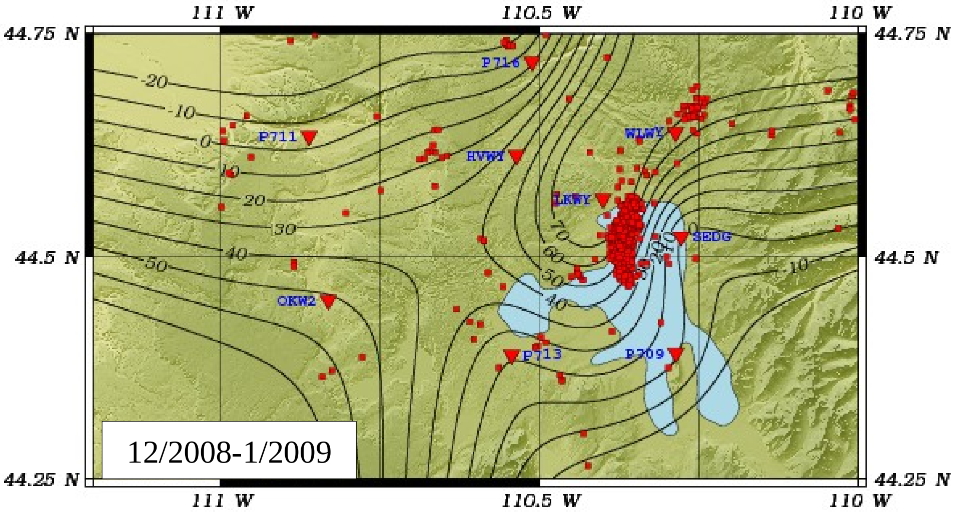
- Figure shows the location of two earthquake sequences that occurred within the caldera in 2008 (red) and 2010 (blue) as a result of the intrusion episode.

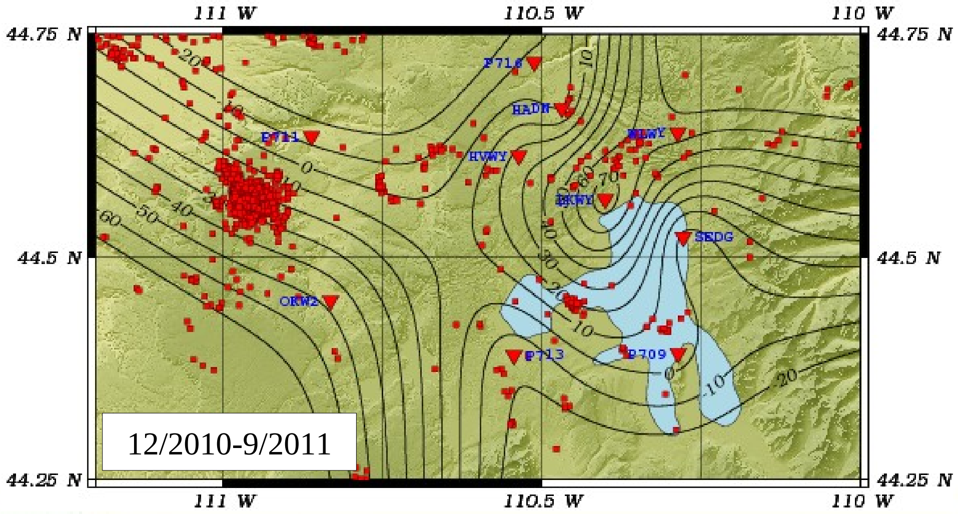
Results: Yellowstone

Collocated Unrest:

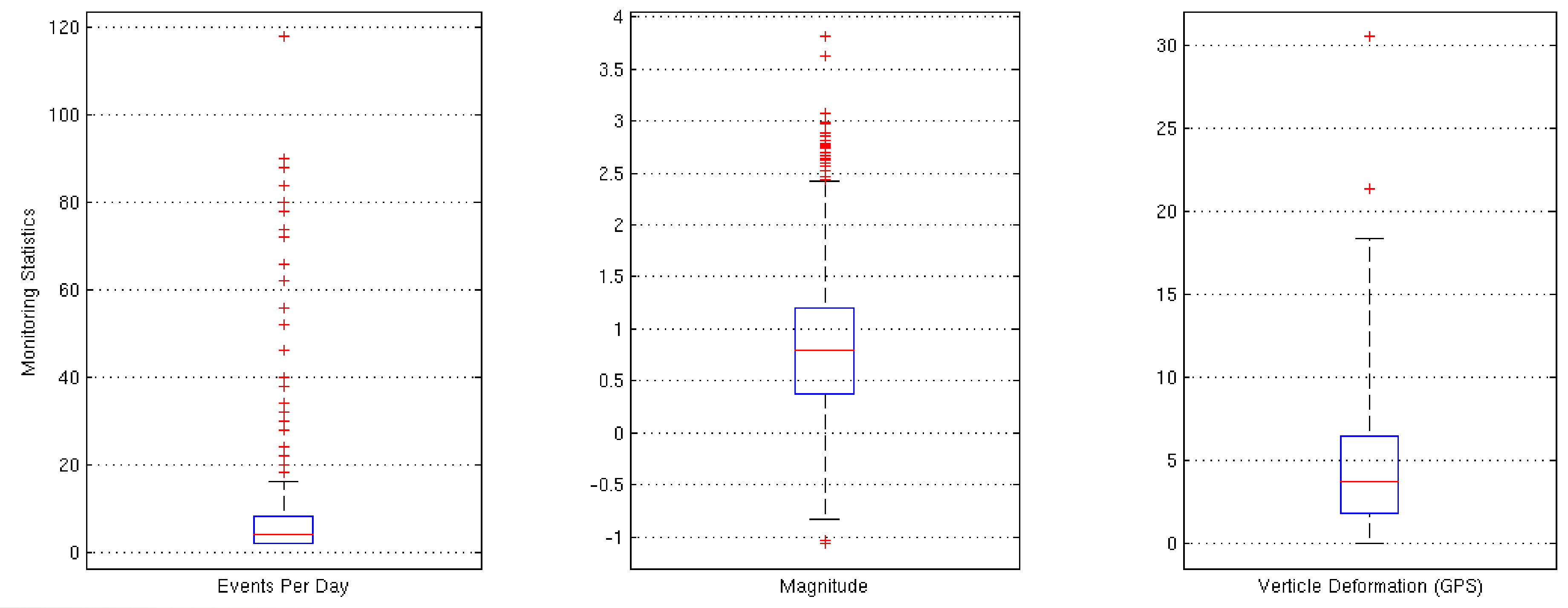
• Figures highlight the collocation of each earthquake sequence with regions having large surface deformation gradients.

• In each case, deformation surfaces, shown as contours, are computed by interpolation of displacement measurements acquired by a suite of GPS sensor located throughout Yellowstone National Park.





Results: Yellowstone

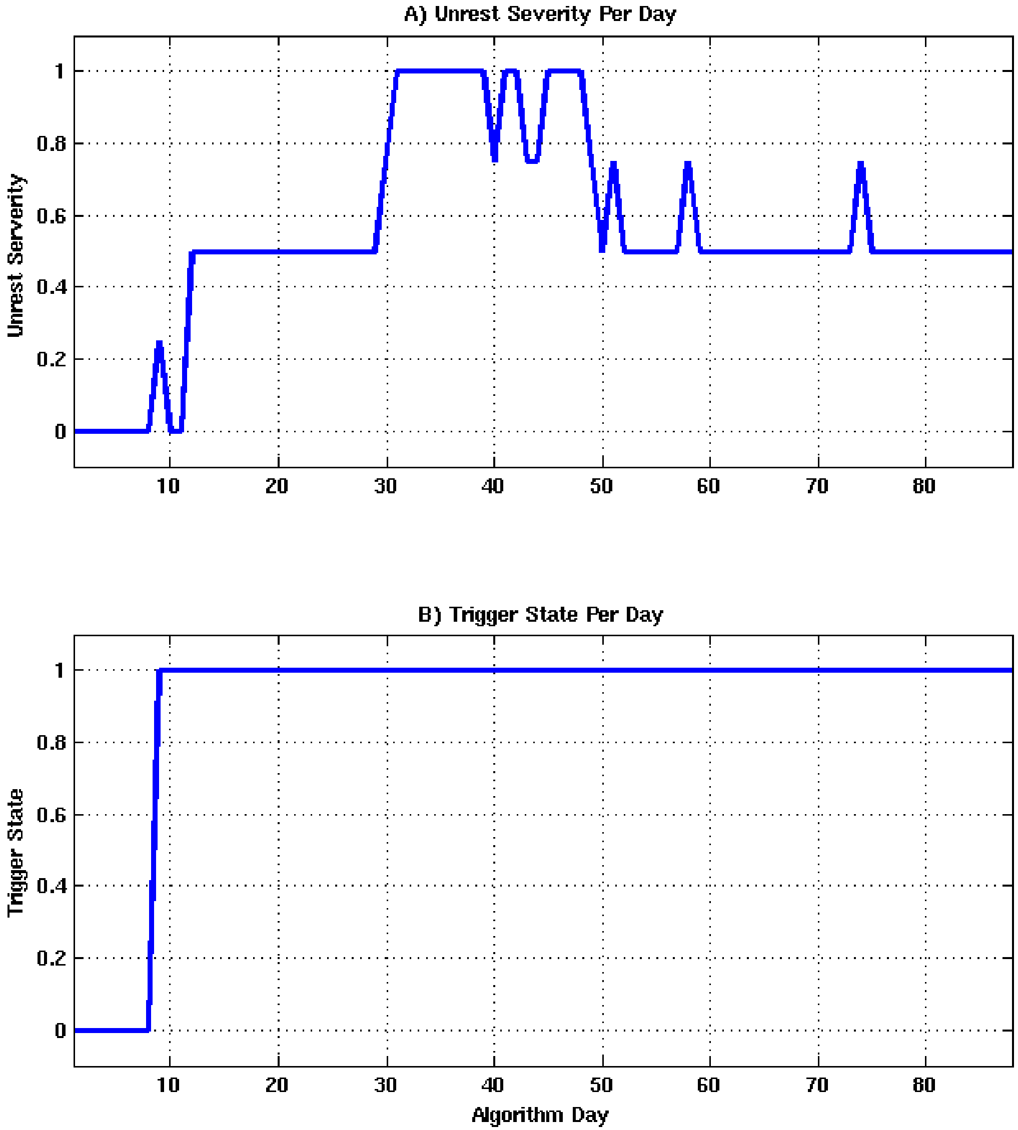


- Boxplots highlighting the distribution of seismicity and deformation within the Yellowstone caldera between2005 and 2008

- Events per day and magnitude whiskers are set to 1.5 and the vertical GPS deformation measurements are 3 times the inter quartile range.

- Monitoring thresholds are 16.0 events per day, ml of 2.42, and 18.3 mm of deformation.

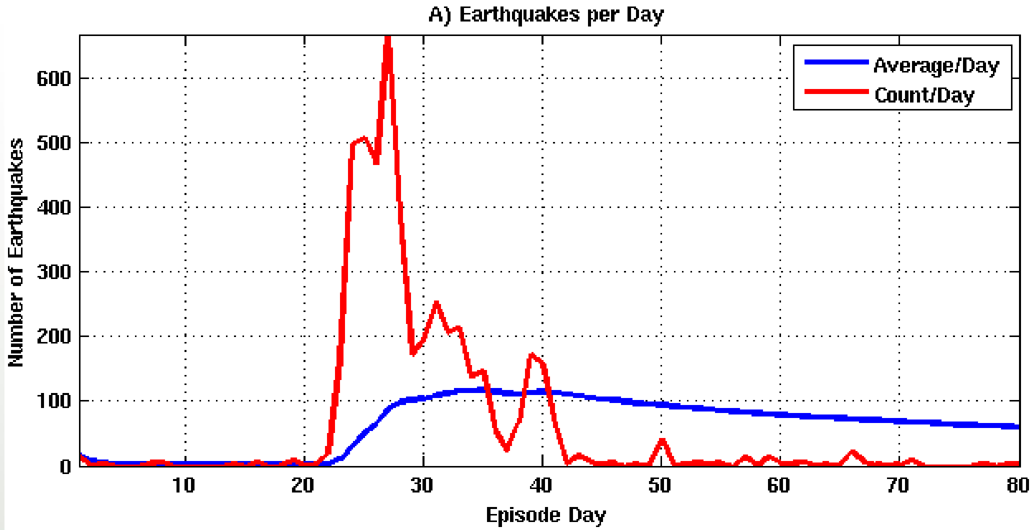
Results: Yellowstone

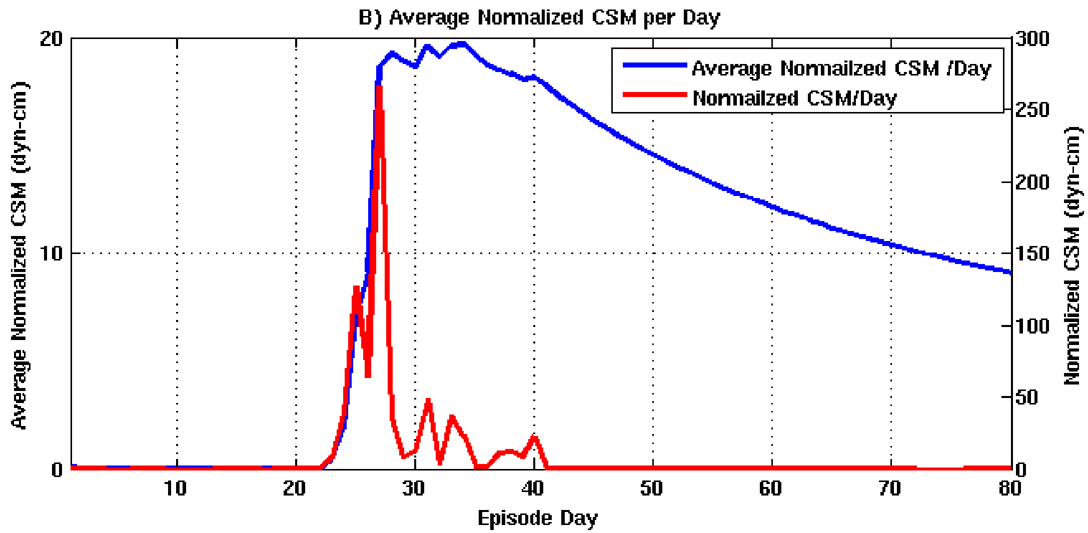


Unrest and Trigger State Vectors:

- Algorithm state as a function of processing day.
- Trigger state = 1, means the algorithm has detected unrest in is actively generating forecasts.
- Trigger state transitions from 0 to 1 on algorithm day 8.
- Unrest severity state is equal to one for approximately 15 days, indicating extreme unrest is detected by all monitoring techniques.

Results: Yellowstone

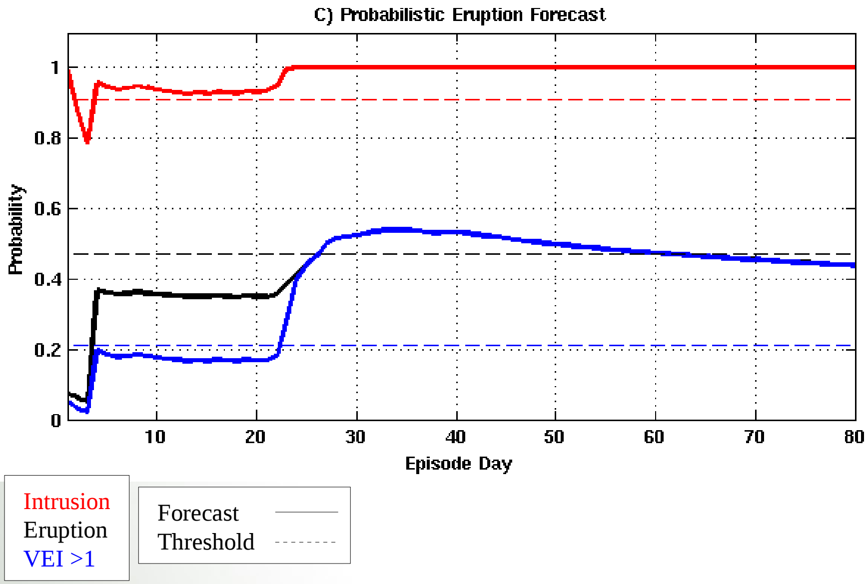


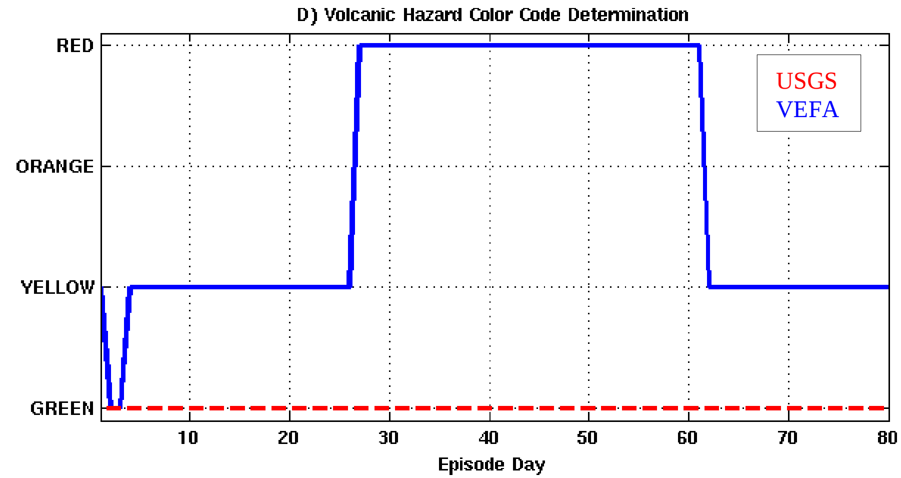


Input Data:

- Volcano monitoring data preceding the 2010 Yellowstone earthquake sequence.
- Count and average (X_{NE}) number of earthquakes per episode day (X_{DAYS}), shown in blue and red.
- Count and average (X_{CSM}) seismic moment per episode day (X_{DAYS}), shown in blue and red.
- Historic eruption frequency index (X_{ERH}) value = 0

Results: Yellowstone





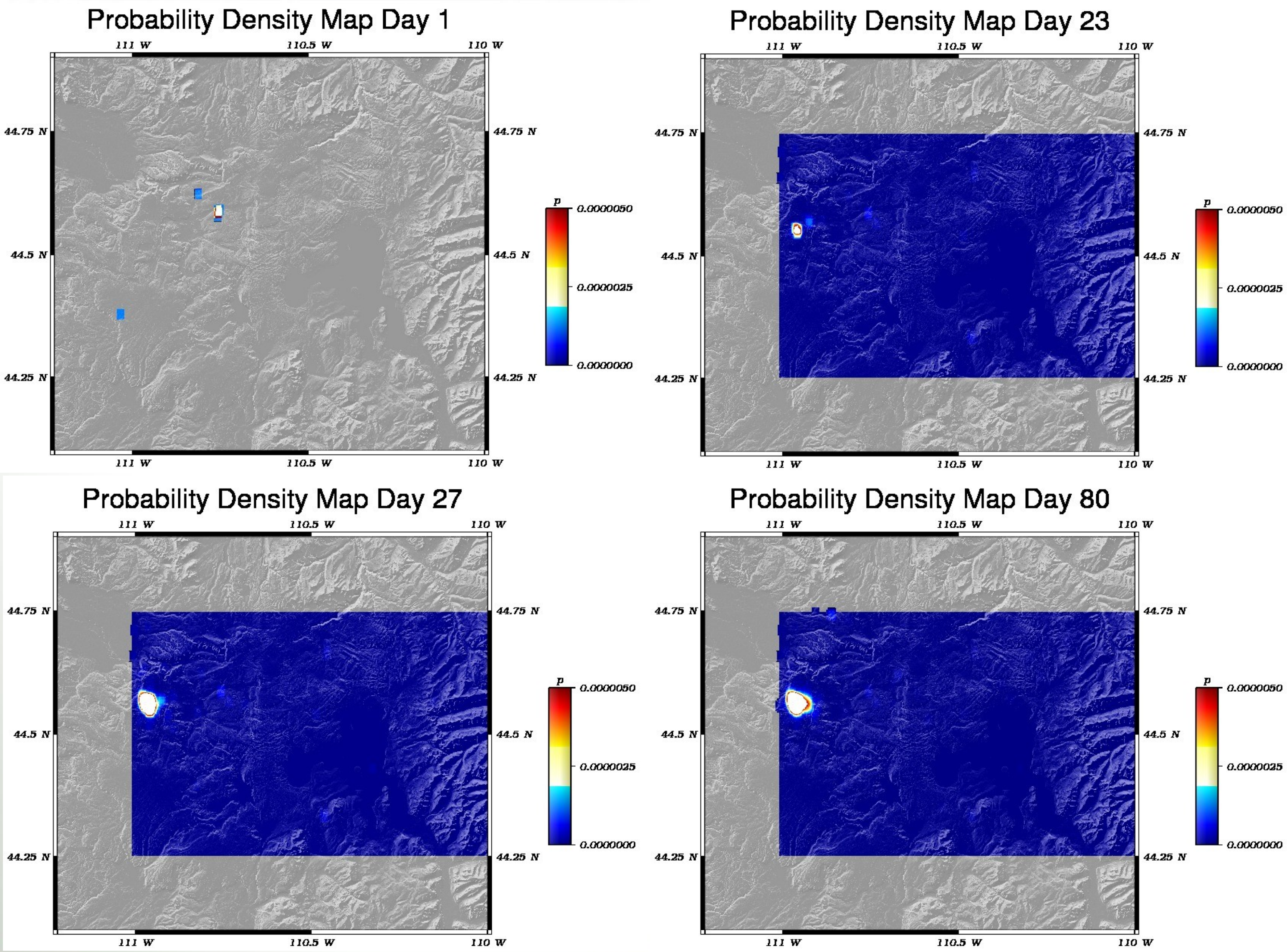
Forecasts:

- Eruption probability ranges between ~0.05 on day 1, peaks at 0.54 on day 35, and settles to approximately 0.44 on day 88.

- The probability of a catastrophic eruption at Yellowstone from this unrest episode is, on average, approximately 0.45 over the course of the episode.

- Initial USGS hazard declaration is yellow, drops to green for one day, increases to yellow for 22 days, elevates to red for 34 days, and eventually falls back to yellow.

Results: Yellowstone



Synthetic Aperture Radar

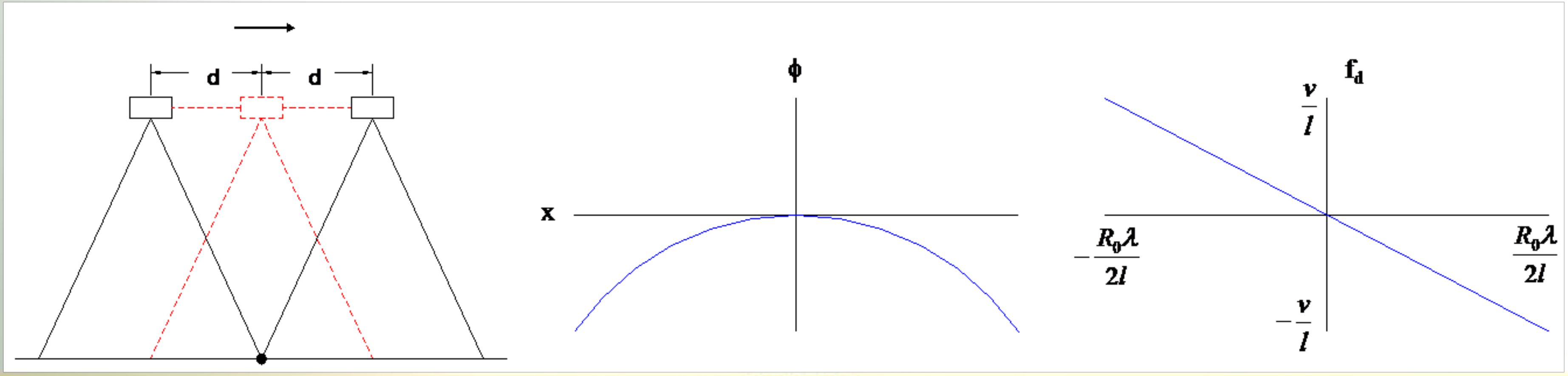
Synthetic Aperture Radar (SAR):

- A SAR uses the forward motion of its platform to produce the equivalent of a large aperture array from a relatively small antenna.

- By directing the antenna beam perpendicular to the platform motion and summing the returns from successive pulses, a synthesized along track array can be constructed

- Platform motion results produces a path length difference between the returns collected by the synthesized array elements.

- This difference produces a phase variation across the length of the array that defocuses the final SAR image.



SAR Concepts:

- Fine azimuth resolution can be achieved by applying a phase correction across the array to focus the image.

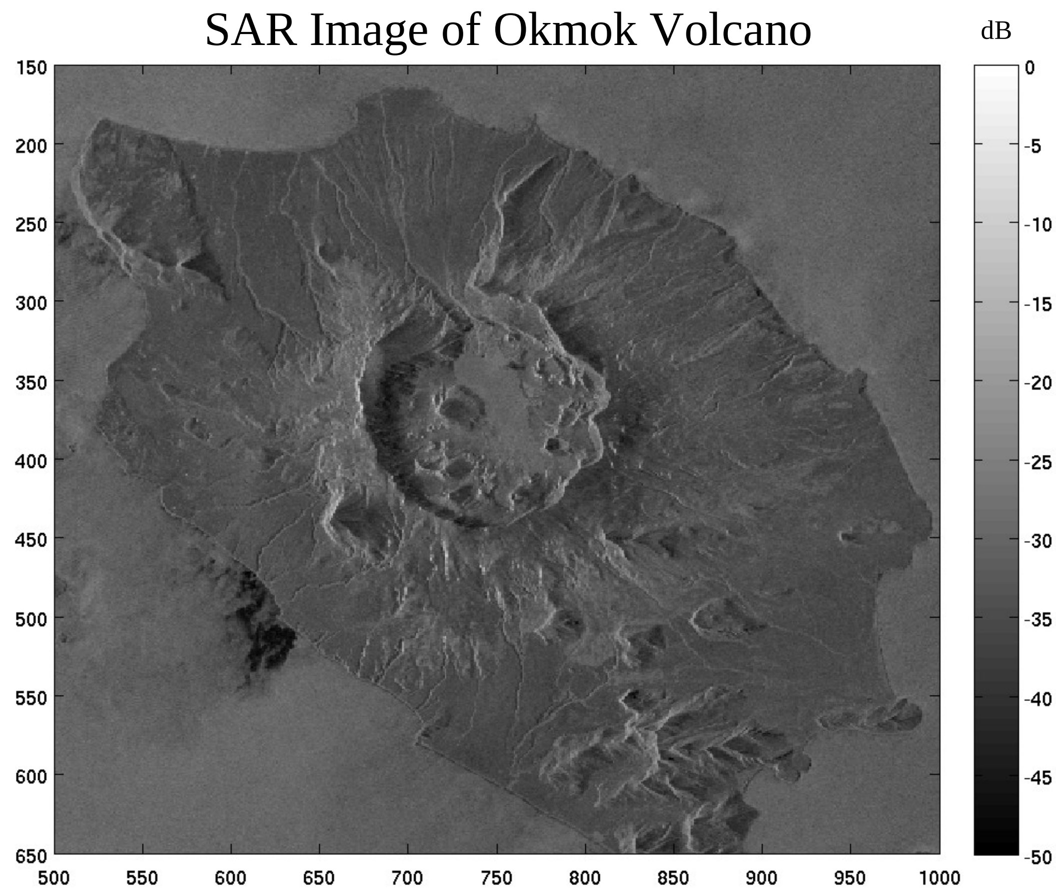
- A focused SAR is capable of an azimuth resolution as fine as half the length of the physical aperture of its antenna

- Range resolution is a function of the transmit pulse bandwitdh

- Image pixel size:

  - Azimuth: 90 m

  - Range: 30 m

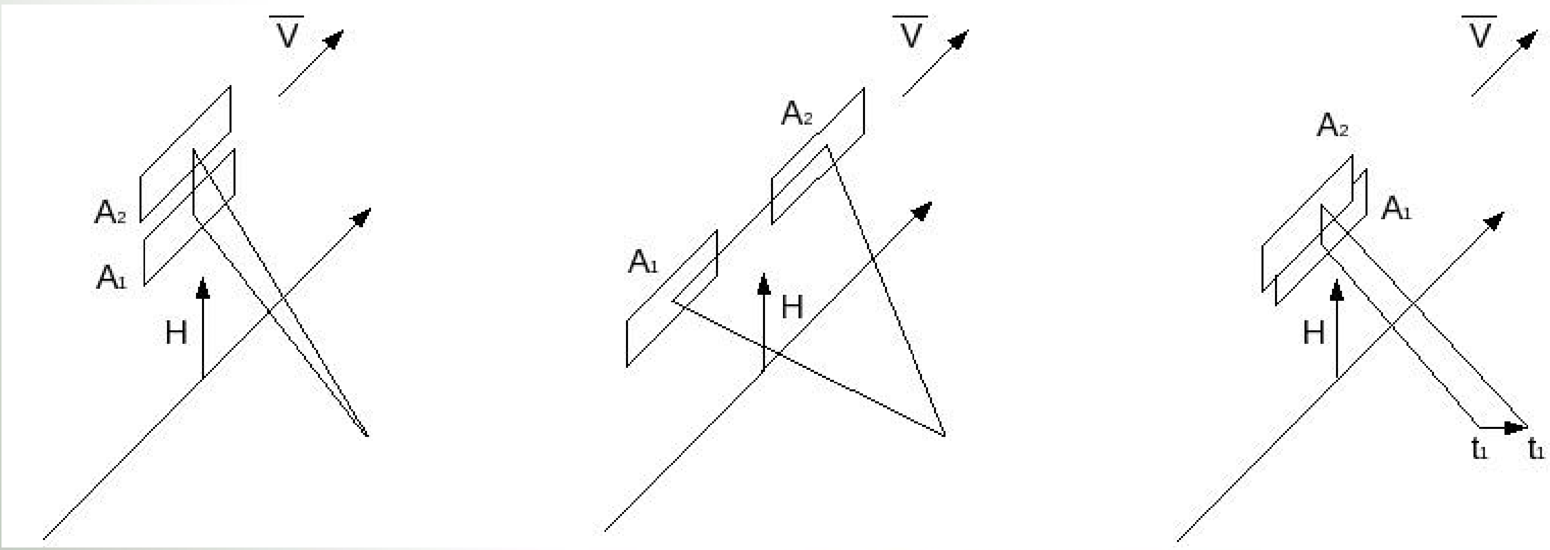


Interferometric Synthetic Aperture Radar (InSAR):

• InSAR exploits the phase difference between two complex SAR images of the same scene that are displaced in either space or time.

• Cross track interferometer (CTI) generates a pair of complex SAR images using two coherent radar systems separated by a vertical distance referred to as the baseline

• Along track interferometer (ATI) consists of two coherent radars that are separated by a horizontal baseline extending in the along track direction, where the image pairs are displaced in time.



InSAR Concepts:

- The phase difference between complex SAR images is referred to as the interferometric phase.

- Caused by path length difference between the backscattered signals from corresponding pixels in each complex SAR image and is a function of the interferometer geometry

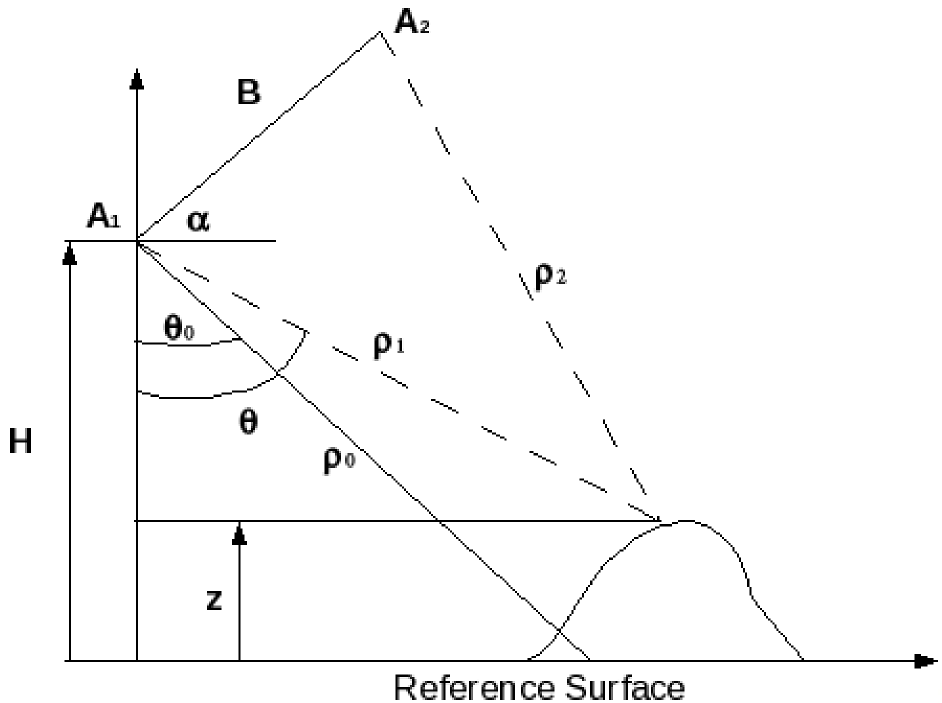
Interferometric Phase

$$\phi = \frac{-4\pi B \sin(\theta - \alpha)}{\lambda}$$

Interferometric Phase

$$\phi_{total} = \frac{-4\pi B \sin(\theta_0 - \alpha)}{\lambda} - \frac{4\pi z B \sin(\theta_0 - \alpha)}{\lambda \rho_0 \sin(\theta_0)}$$

Total   Flat Earth   Topographic



Synthetic Aperture Radar

InSAR Concepts:

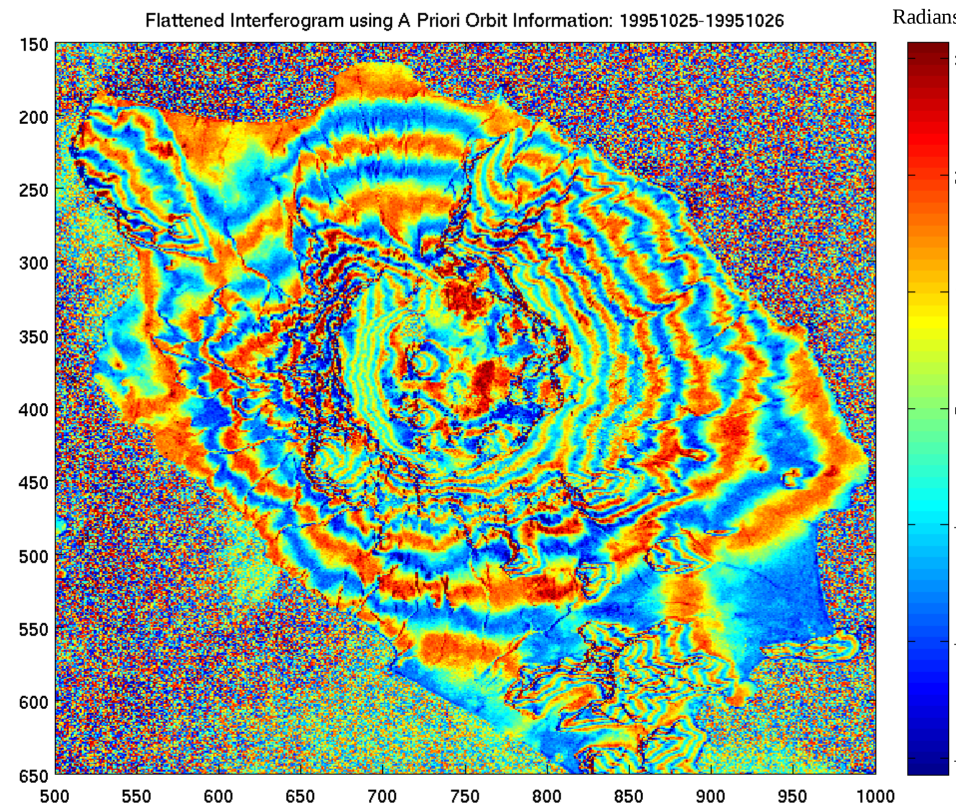
- An image illustrating the interferometric phase pattern over a geographic area is known as an interferogram.

- Each color cycle, or fringe, is equivalent to a complete phase revolution of 2 π radians.

- Vertical displacement in the line of sight direction is represented by a fringe pattern that increase or decrease as a function of the changing elevation.

- Each complete phase revolution is equivalent to an elevation change of h_a meters, which is referred to as the ambiguity height

- Ambiguity Height: 123 m



Ambiguity Height

$$h_a = \frac{\lambda \rho_0 \sin(\theta_0)}{2B\cos(\theta_o - \alpha)}$$

Synthetic Aperture Radar

InSAR Concepts:

- Elevation changes (surface deformation) occurring over the time separation between SAR images is easily detected and measured using InSAR techniques

- Each fringe in the displacement interferogram is equivalent to a distance equal to half the electromagnetic wavelength of the radar

Interferometric Phase

$$\phi_{total} = \frac{-4\pi B \sin(\theta_0 - \alpha)}{\lambda} - \frac{4\pi z B \sin(\theta_0 - \alpha)}{\lambda \rho_0 \sin(\theta_0)} + \frac{4\pi}{\lambda}\delta\rho_{disp}$$

Total — Flat Earth — Topographic — Displacement

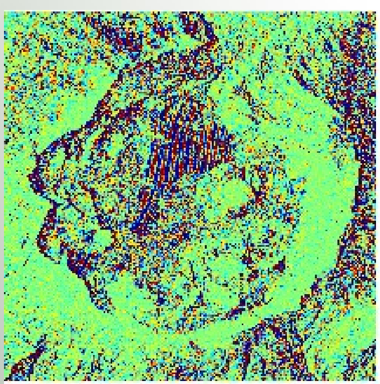
Interferometric Phase (displacement)

$$\phi_{disp} = \frac{4\pi}{\lambda}\delta\rho_{disp} = \frac{4\pi}{\lambda}\frac{\lambda}{2} = 2\pi$$

Synthetic Aperture Radar

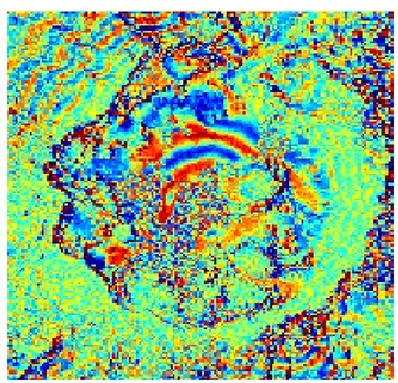
InSAR Concepts:

- The flat earth and topographic phase contributions must be simulated and subtracted from the total interferometric phase
- To remove the topographic phase contributions, a digital elevation model (DEM) of the area is required
- The DEM is used to estimate the topography in the scene that must be simulated and removed.
- The residual interferometric phase is the result of surface changes that have taken place over the elapsed time between images.



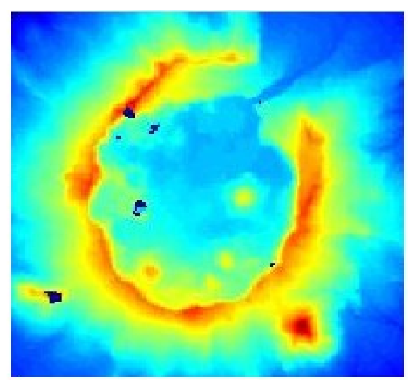
Raw Interferogram

-



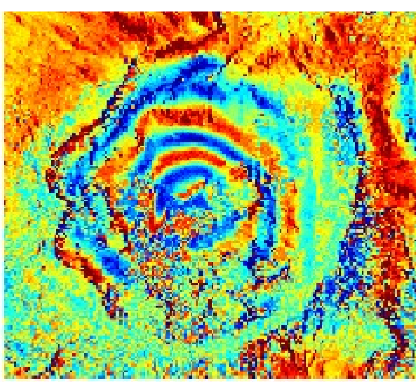
Flattened Interferogram

-



Digital Elevation Model

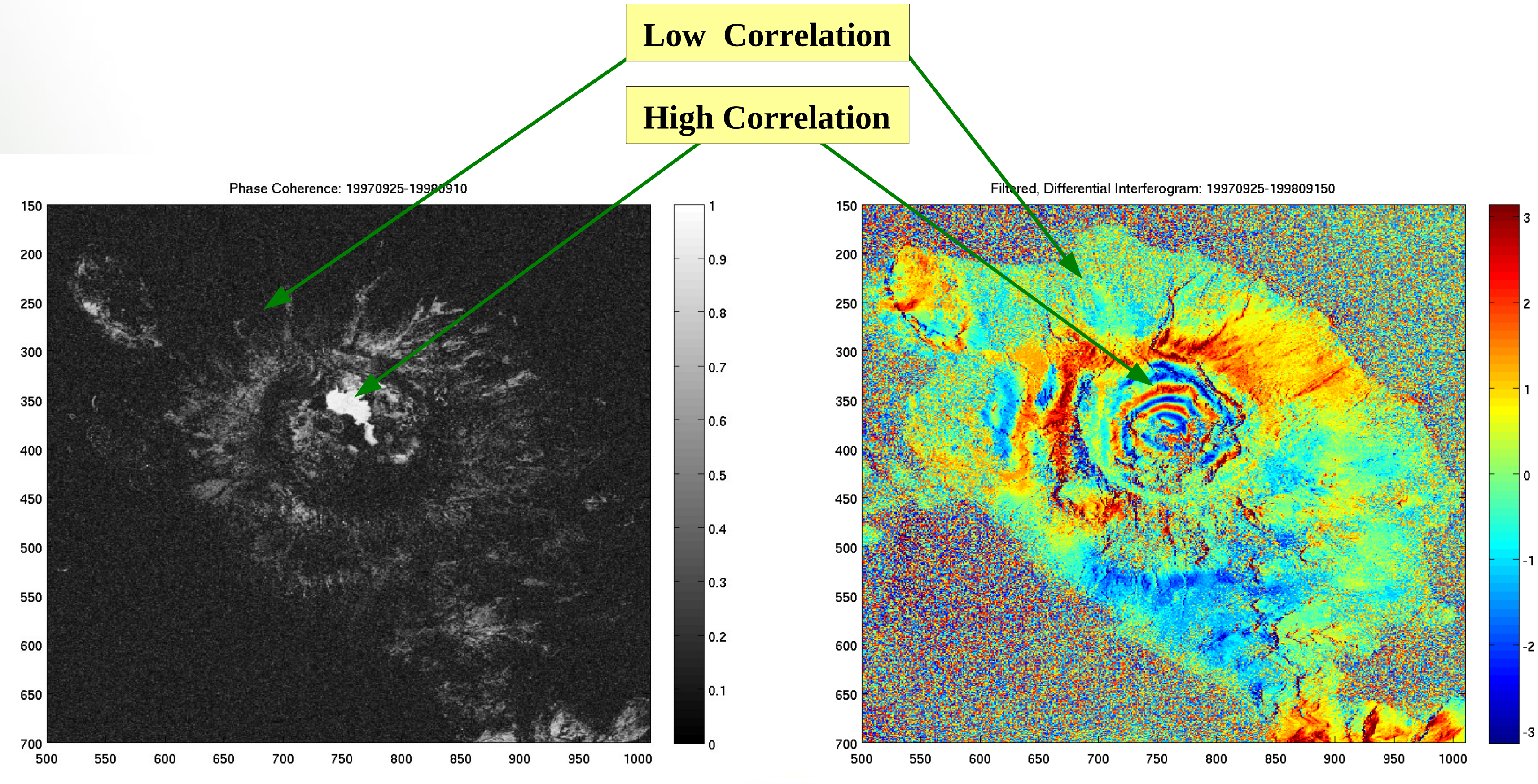
=



Deformation Induced Interferogram

InSAR Limitations:

- InSAR surface deformation measurements are limited to areas having high correlation between images.

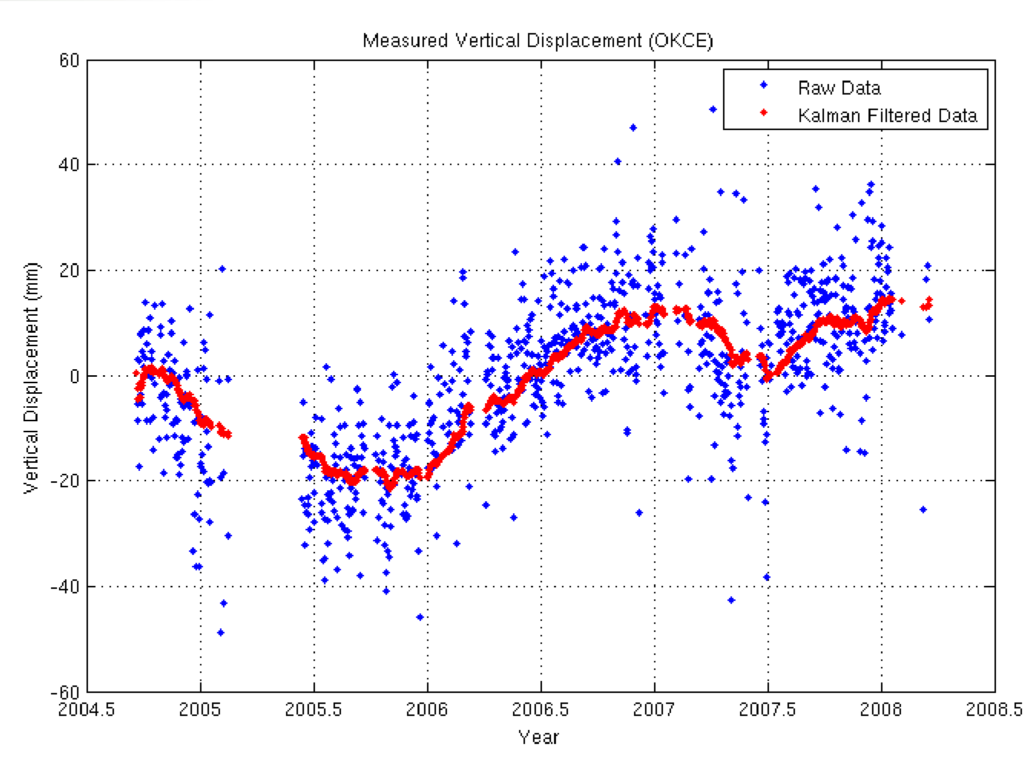


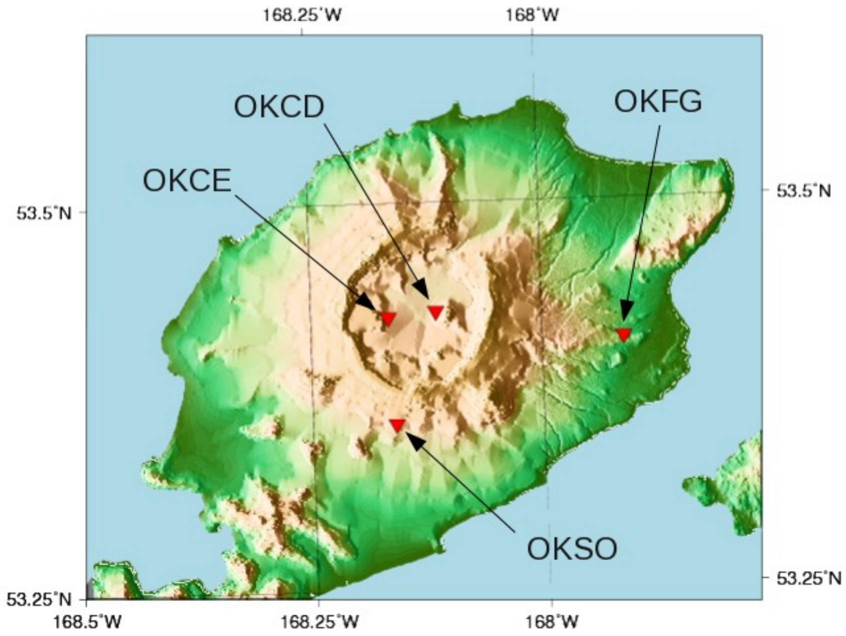
Global Positioning System

GPS Network:

- Okmok's GPS network can be used to fill the InSAR coverage gaps.

- A Kalman filter is used to reduce the variance in vertical displacement measurements.

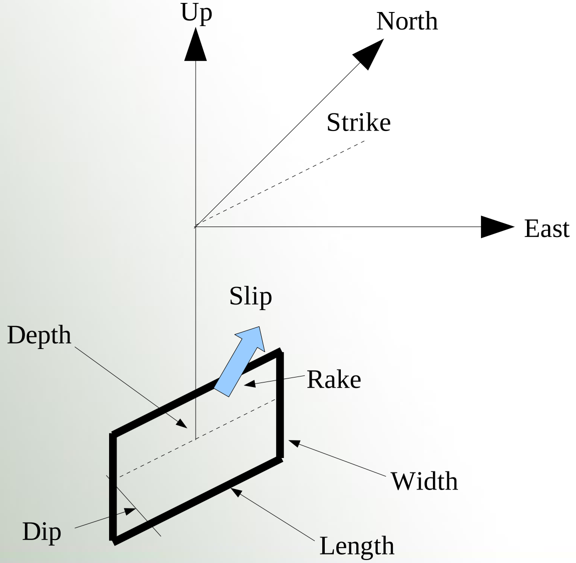


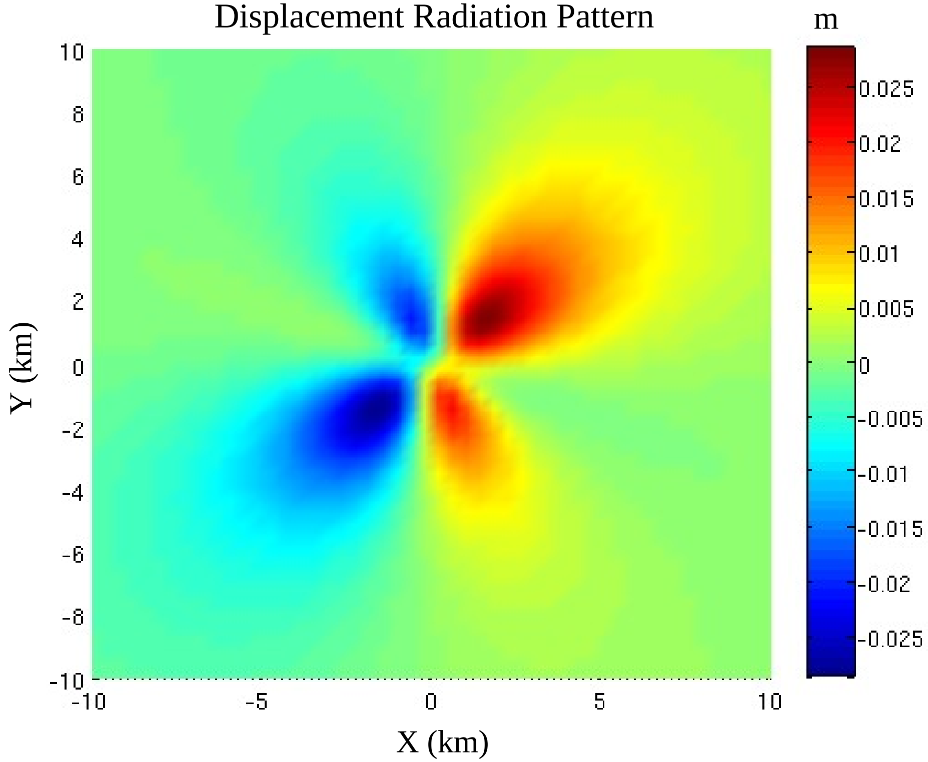


Surface Deformation Models

Rectangular Source

- Deformation pattern produced by a rectangular source (e.g., sill, dike).
- Can be used to simulate displacement produced by a fluid filled crack.
- Based on Okada 1985.





Model Parameters:

- Left Lateral, Strike-Slip, Model
- Depth = 2
- Length = Width = 2
- Strike = 0, Dip = 90, Rake = 0