Long Valley volcanic region long-term vent opening maps

Andrea Bevilacqua[(1)], Marcus Bursik[(1)], Abani Patra[(2)], Bruce Pitman[(3)]

(1) University at Buffalo, Dept. Geology
(2) University at Buffalo, Dept. Mechanical and Aerospace Engineering
(3) University at Buffalo, Dept. Materials Design and Innovation

2nd VIMESEA meeting
22-23 September, LMV - Clermont Ferrand

Doubly stochastic methods

The volcano will be presented as a **random system** that must be assessed with incomplete and **uncertain information**.

Adopting a **doubly stochastic approach,** some ill-constrained parts of the long-term probability models will be randomly changed.

Uncertainty quantification assumes a great importance, and we distinguished:

I. the **physical variability,** i.e. the intrinsic randomness of the system under study,

II. the **epistemic uncertainty** due to the imperfect knowledge/interpretation of the system.

As a consequence of this approach, some probability estimates will have their own confidence intervals.

Even the final probability maps will be affected by uncertainty: we calculated the **mean, 5th and 95th percentile** values for the vent opening probability density functions.

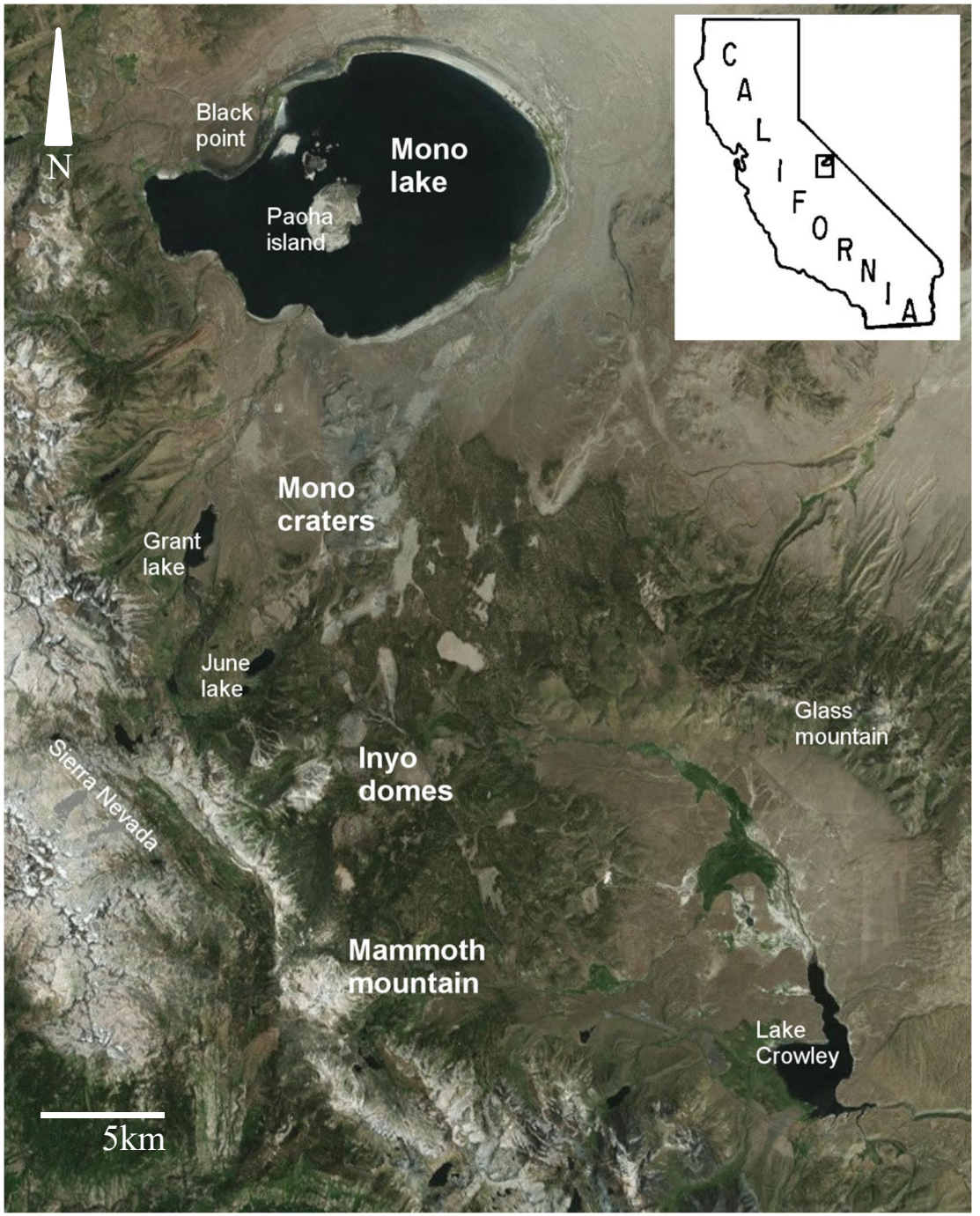
The Long Valley volcanic region

Long Valley volcanic region (LVVR) in eastern California includes the Long Valley caldera (LVC), created by the eruption of >200km^3 tephra ~ 760ka BP (Bishop tuff).

Over the last 180ka the eruptions have been localized at **Mammoth mountain,** on the western rim of LVC and along the **Mono-Inyo craters** volcanic chain, stretching ~45km North towards Mono lake.

The most recent period of **unrest started in 1978**, with several seismic swarms in LVC and below Mammoth mountain including one in 2014, and diffuse volcanic CO_2 emissions.

The study concerns LVVR **long-term vent opening maps**, primarily based on past eruption data and on the structural features of the volcanic system.



LVVR: volcanological datasets

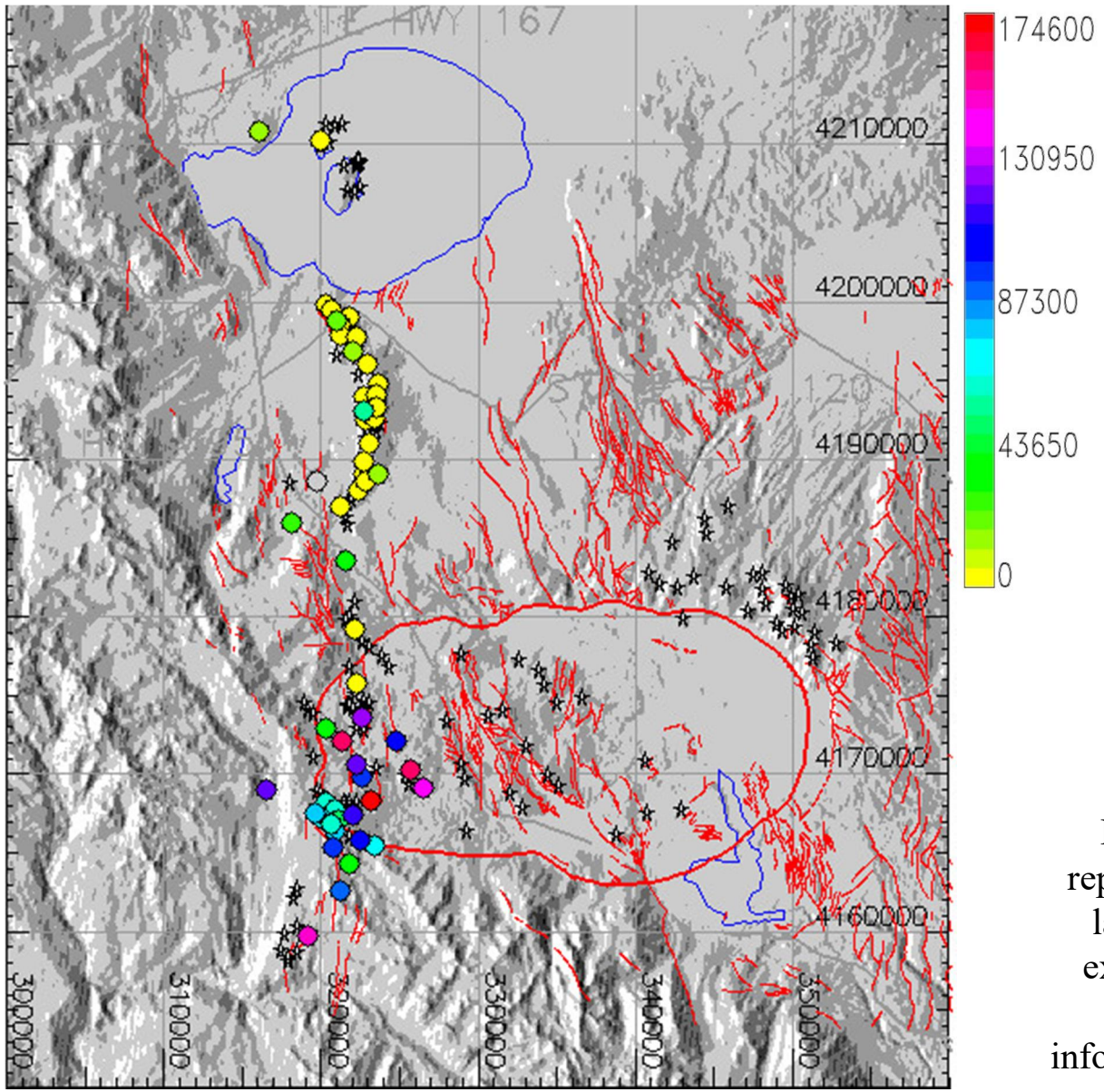


Figure 1 shows the spatio-temporal past vents dataset of the last 180ka. Colors refer to the ages BP.

The main dataset is composed of **uncertainty disks** (500m radius) defined on past vent locations of the last 180ka.

Temporal information about **sequence and age** is included with uncertainty.

Another dataset concerns the **fault extension** rates mm/yr of the last 130 ka.

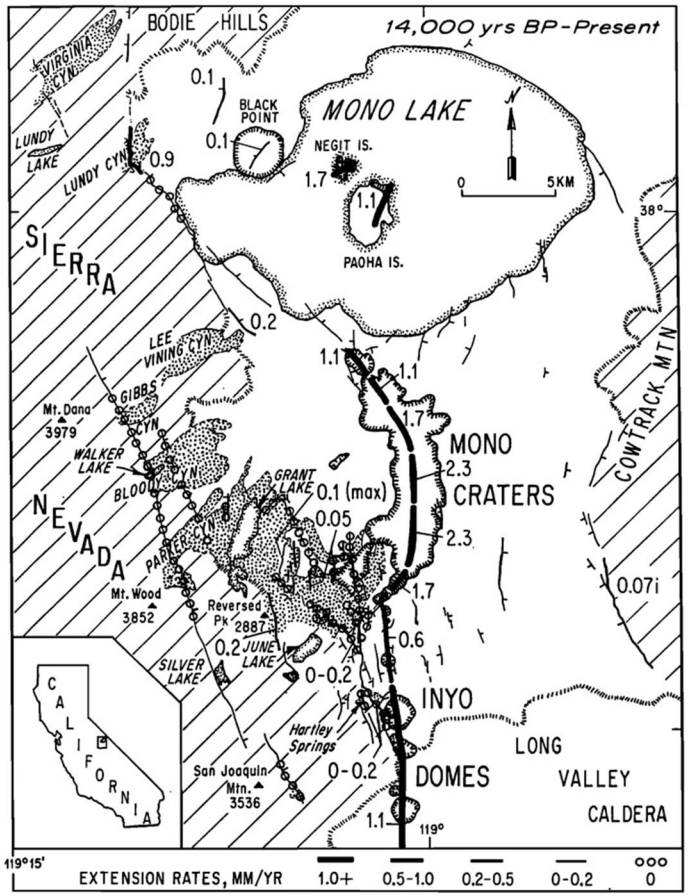


Figure 2 reports the last 14ka extension rate information on a map of the Mono basin [Bursik and Sieh 1989].

Time record description

A small number of events have been **removed** from the record to eliminate duplication.
Any pair of eruptions which were less distant than 150 years and 1km was considered as a single event.

A criterion based on additional volcanological information would be more robust.
This part is under development.

Past record was divided into five subgroups.

A

180/149 ka BP - initial sequence,
before a long period of quiescence

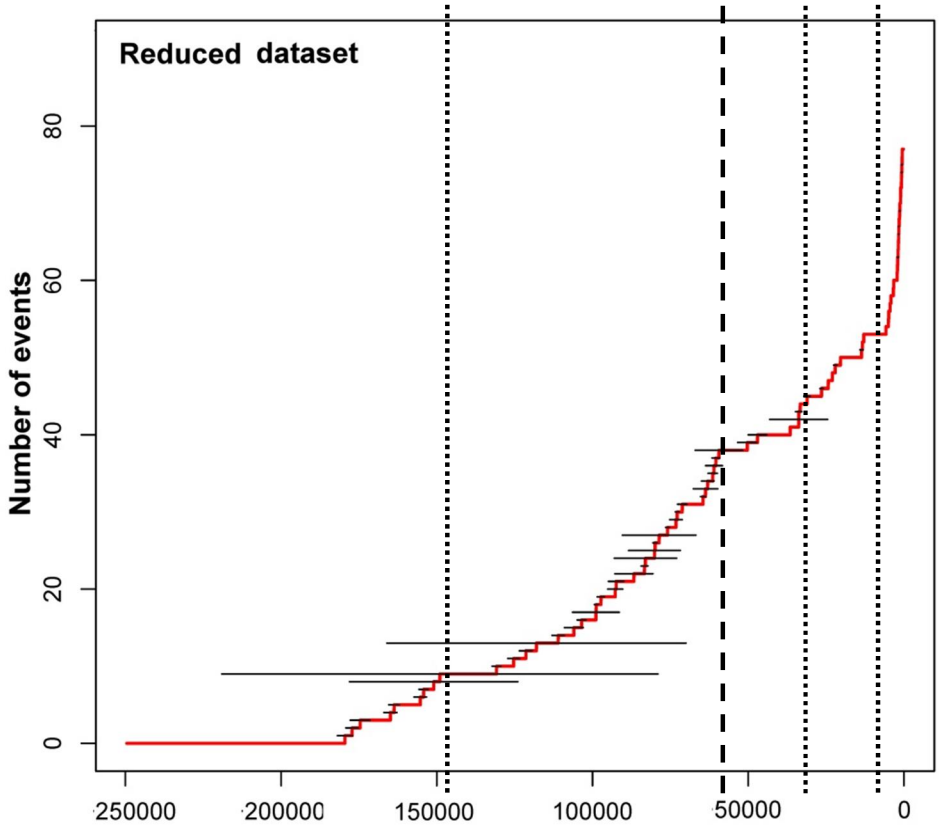
131/60 ka BP - activity centering on
Mammoth mountain area

B

60/31 ka BP - shift of the activity from
Mammoth region to Mono basin.

26/13 ka BP - activity centering on Mono basin

< 6 ka BP - most recent events,
after a period of quiescence.



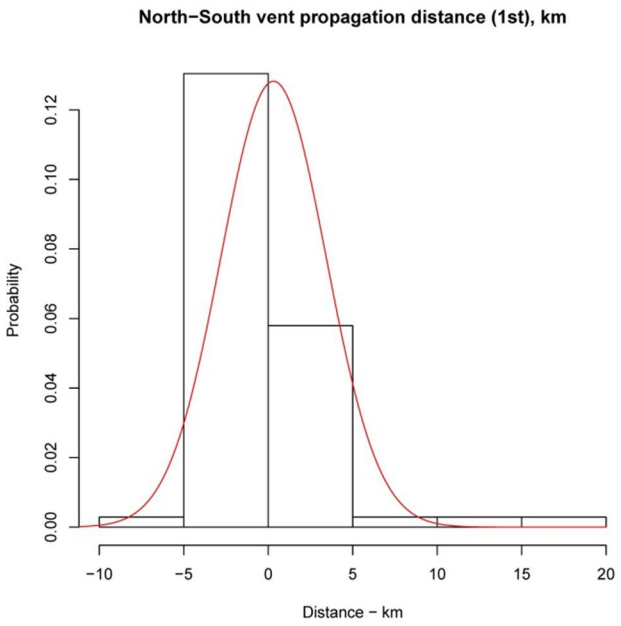
The figure reports the plots of the cumulative number of past event as a function of time. The vertical lines separate the subgroups.

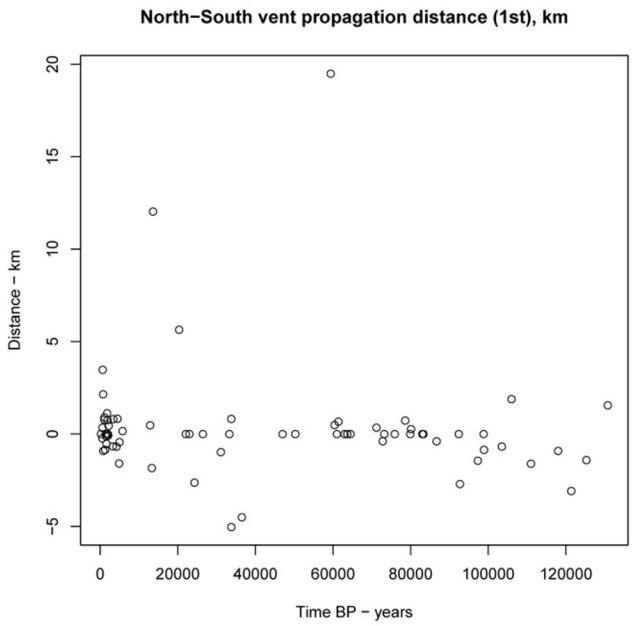
The separation around 60ka BP corresponded with the **activation of the northern part** of the region.

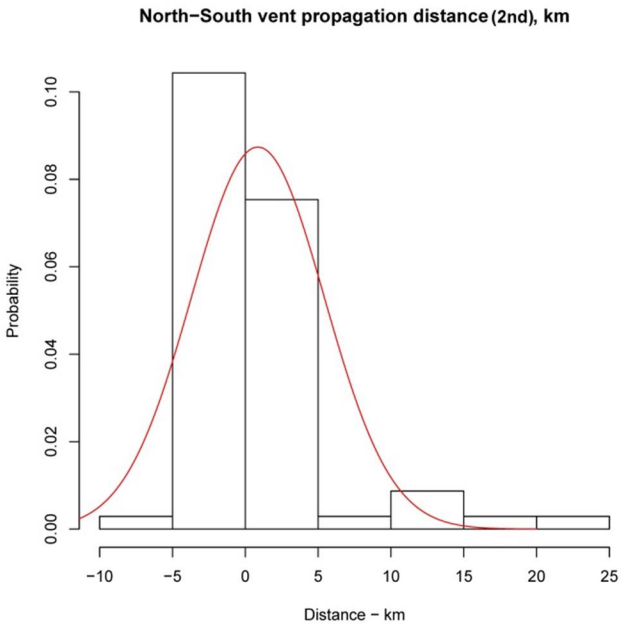
It is remarkable that at 5.8ka BP the rate of events per year rose **~13.6 times higher** than before, even if most of the removed events were in that interval.

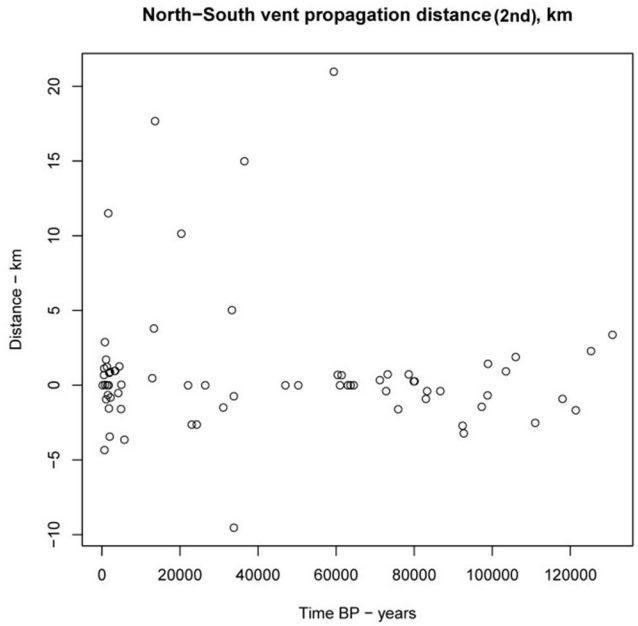
Past vents distance of propagation

A quantitative parameter of main importance, is the **distance of propagation** to each new vent from the set of the previous vent locations.









Alternative distances are easily defined: the closest previous vent, the second closest vent, etc.

A **20km jump** in the N/S distance of propagation is noticed around 60ka BP.

There is a significant increase of N/S propagation distances after such jump.

The figure reports the distribution of N/S distances of propagation from 1st and 2nd closest previous vent.

The 9 most ancient vents were assumed as a starting dataset and are not included.

On the right are the data samples; on the left are their histogram representations with Maximum likelihood Gaussian fits.

States A and B of the volcano

Two different states of volcanic activity named A and B modeled the past vent locations, with a probability *p* of changing state when a new event will occur.

The state A corresponded with the activity between 180ka and 60ka BP, the state B with the activity between 60ka and the present time.

Geographically, state A concerned Mammoth mountain area, whereas state B mostly the Mono basin.

The value of *p* is **very uncertain**: the Mammoth mountain system is likely crystallized, but being this volcanism related with tectonics, something like a **seismic gap** behavior could not be excluded for the future reactivation of that region.

For this reason *p* was uniformly sampled from 0% to 47.43%, the frequency of events in state A. In this way two **opposite paradigms** were interpolated in the epistemic uncertainty:

[the volcano is going to remain in state B // states A and B are equally probable in the future]

The tectonic data currently available mostly concerns the Mono basin, for this reason that type of data was only included in state B inferences. Additional faults data are under development for state A.

Probability mixture approach

The main sources of epistemic uncertainty affecting the problem are:

- the **uncertainty on location, age and number** of past eruptive vents;
- the **uncertainty on the relevance** of the different strands of data contributing to the map definition.

This talk will be mostly focused on the second type of uncertainty.

The physical variability is assessed as a **linear combination** of 5 spatial distributions $(\mu_i)_{i=1,\dots,d}$:

Past Vents A	Past Vents B	Faults A	Faults B	Uniform
180ka to 60ka	60ka to present time	Mono Region younger than14ka	Mono Region older than 14ka	

A uniform probability measure was also included, **homogeneous** over 20km distance from the considered past vents, for representing the possible lack of information.

The assessment of the vent opening map $\sum_{i=1}^{d} \alpha_i(e)\mu_i$, is reduced to find the distribution of the positive coefficients $\alpha = (\alpha_i)_{i=1,\dots,d}$, $\sum_i^d \alpha_i = 1$, depending on the epistemic uncertainty *e*.

Past vents data: kernel density estimates

The first spatial distributions were based exclusively on past vent locations data.

MAIN IDEA: given the locations $(x_i, y_i)_{i=1,\dots,N}$ of the past N events, a new event propagates from one event location **randomly chosen** from the previous, to a **random distance**:

$$\boldsymbol{X} = (x_{\boldsymbol{k}} + d_1, y_{\boldsymbol{k}} + d_2),$$

where $\boldsymbol{X}$ is the spatial location of the next vent,
$\boldsymbol{k}$ is a discrete random variable in $\{1,\dots,N\}$ sampling one of the previous vents, and
$\boldsymbol{d} = (d_1, d_2)$ is a two dimensional Gaussian random vector with mean μ and covariance matrix Σ.

The density f of $\boldsymbol{X}$ is obtained by convolving an anisotropic **probability kernel** describing $\boldsymbol{d}$ with the past vent location disks $(D_i)_{i=1,\dots,N}$, each weighted relying on the distribution of $\boldsymbol{k}$:

$$f(\mathbf{x}) = \sum_{i=1}^{N} \frac{k(i)}{2\pi\sqrt{\det(\Sigma)}} \int_{D_i} \exp\left(-\frac{1}{2}(\mathbf{x}-\xi-\mu)^T\Sigma^{-1}(\mathbf{x}-\xi-\mu)\right) d\xi.$$

The random variables $\boldsymbol{k}$, $\boldsymbol{d}$ change according with the state: defined as $\boldsymbol{k_a}$, $\boldsymbol{d_a}$ for A and $\boldsymbol{k_b}$, $\boldsymbol{d_b}$ for B.
For simplicity the variable $\boldsymbol{k}$ was equally distributed on the considered vents.

Bayesian model averaging for the propagation distance

Alternative estimates of $\boldsymbol{d_a}$ and $\boldsymbol{d_b}$ were produced considering the closest event (model M_1), the second closest (model M_2), the third closest (model M_3), etc.

The choice of the distance measurement was done with a **Bayesian Model Averaging** (BMA) approach.

The basic idea was to **weight** the different models in proportion to the **likelihood** of past data – a **"cross-validation"** procedure was accomplished on the most recent part of the datasets, 120-60ka BP for state A, the last 6ka for state B.

In detail, for each model M_j and for each x_i of the $m = 25$ most recent events the likelihood L_j^i was the integral of the density f_j^i of the random variable $\boldsymbol{X_j^i}$ inside its uncertainty disk:

$$L_j^i = \int_{D_i} f_j^i(\mathbf{x}) d\mathbf{x}$$

Such $\boldsymbol{X_j^i}$ was defined with $\boldsymbol{k}$ restricted to the events before ith, and with $\boldsymbol{d}$ calculated applying M_j on such reduced dataset.

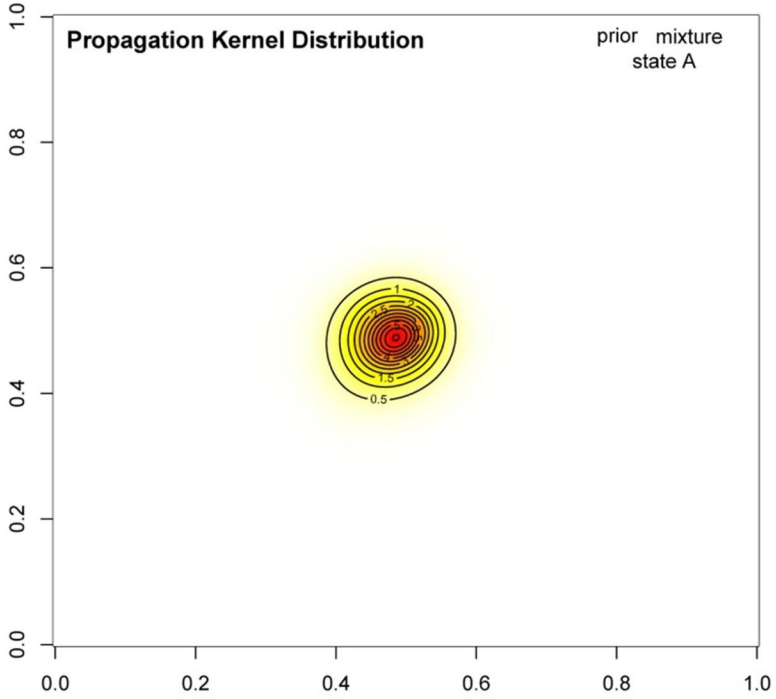
For each model M_j, the **total likelihood** L_j was the product of the likelihoods of the considered events:

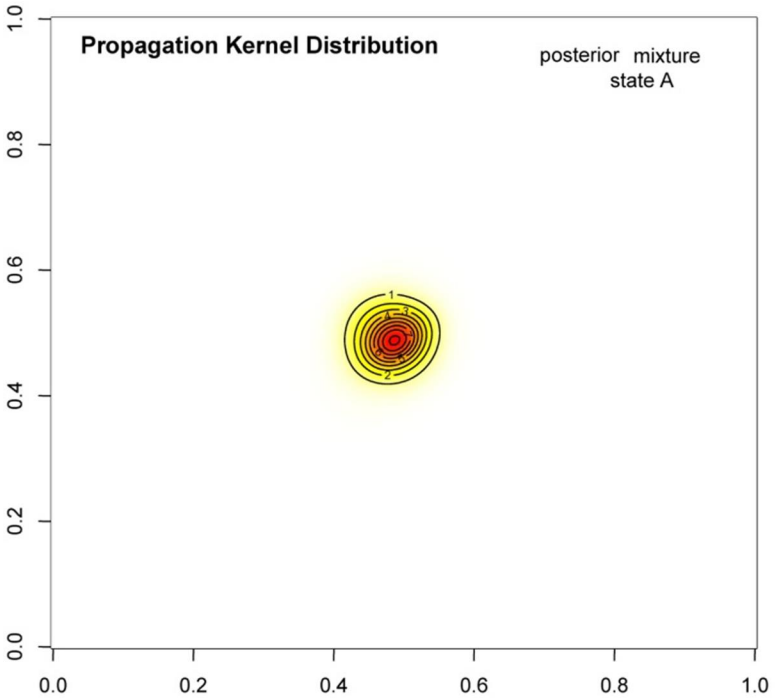
$$L_j = \prod_{i=1}^{m} L_j^i.$$

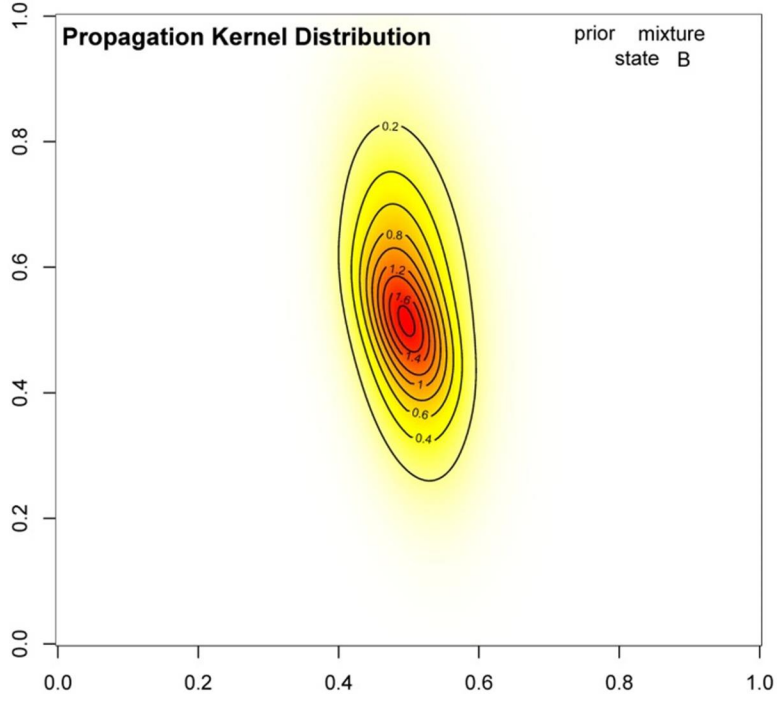
Applying the Bayes formula on equally assigned prior weights, for each j the posterior weight of M_j was:

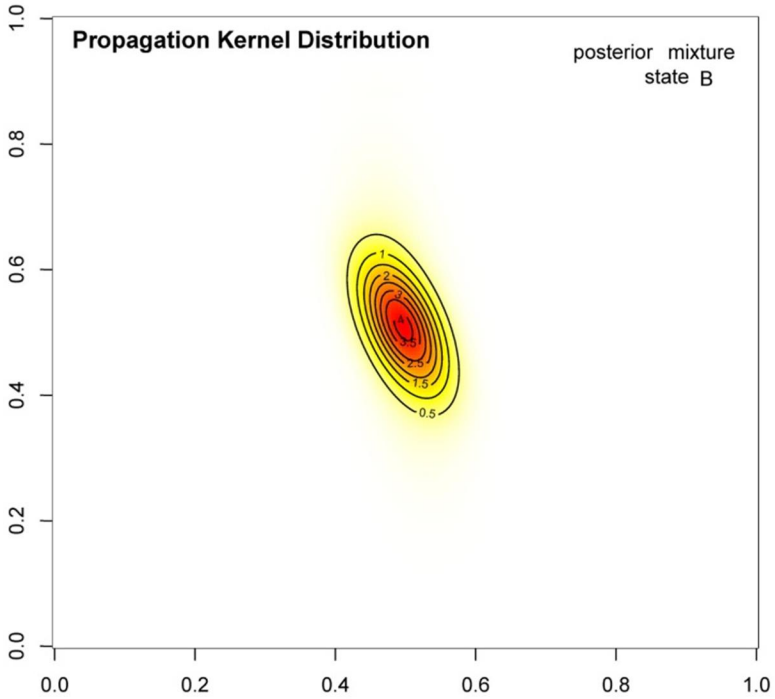
$$w_j = L_j / \sum_k L_k.$$

Combining the Gaussian kernels









In state A the weights are:
w_1=32.00%,
w_2=52.00%,
w_3=3.32%,
w_4=3.16%,
w_5=1.32%,
w_6=8.19%.

In state B the weights are:
w_1=76.92%,
w_2=10.10%,
w_3=6.18%,
w_4=6.80%.

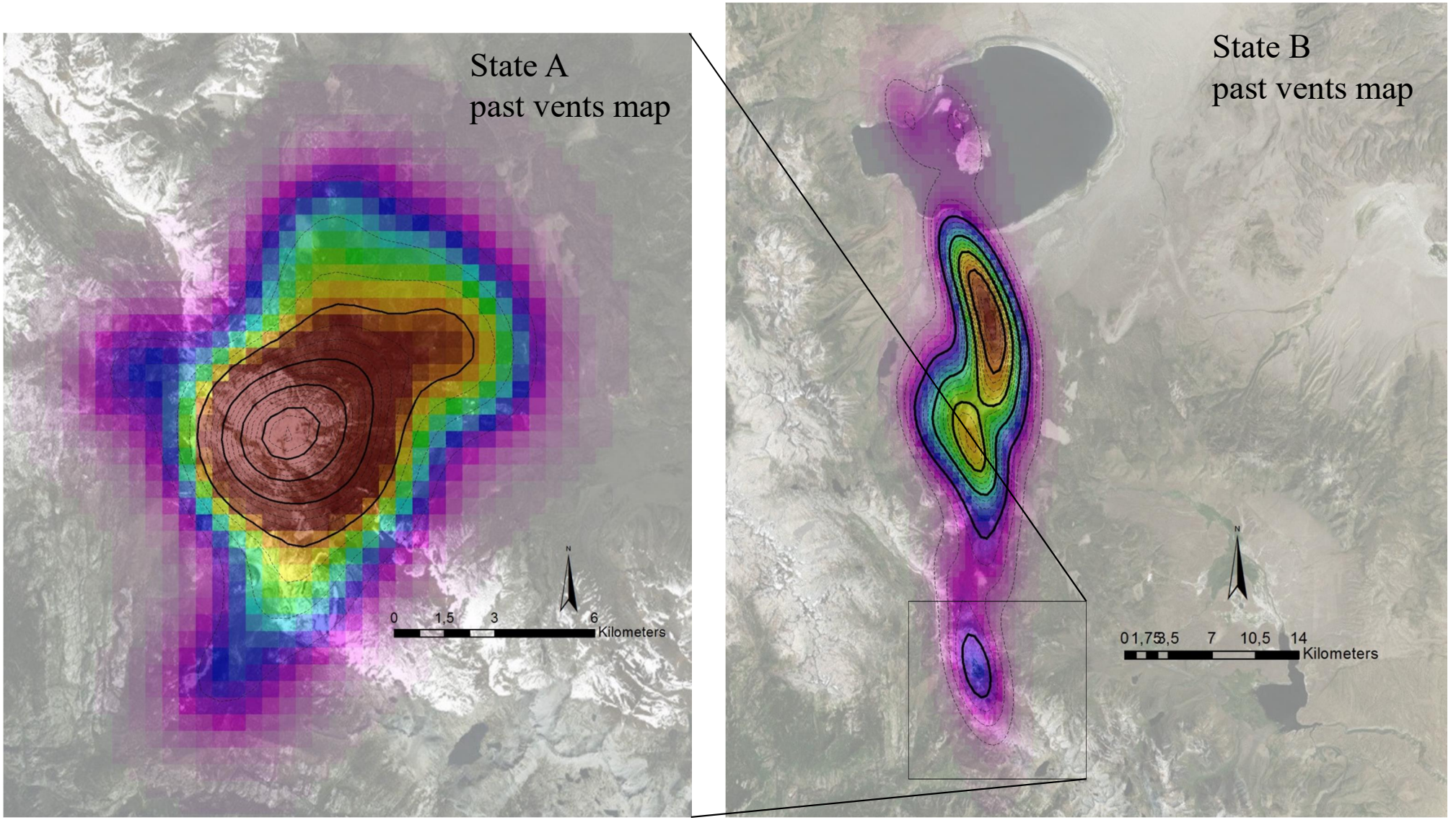
In the figure are the plots of the kernel distributions.

The prior frames concern mixtures assuming equal weights for $(M_i)_{i=1,\ldots,7}$.

The frame sides sizes are 40km.

Vent opening maps based on past vents data

The map of vent opening is the percentage estimate for the probability of vent opening per km^2 in each point the region of interest.



Contour lines report the vent opening probability percentage per km^2.
For state A the bold lines are 0.8% paced, the thin pointed lines are 0.2% paced.
For state B the bold lines are 0.2% paced, the thin pointed lines are 0.05% paced.

The tectonic maps: a source of epistemic uncertainty

Faults estension measurements depend on the **regional stresses** and are a useful marker for potential **dike intrusion/deviation**. Additional spatial distributions were based on the tectonic data.

Four different strands of data are available for the Mono Basin, covering the last 130ka.

As a first guess the younger fault extensions <14ka BP may be the only dataset to consider in the map.

However they are strongly correlated with the **dike accommodation** below the Mono domes, and the older faults extension values could be still valuable data for the long term tectonics of the region.

For this reason **two different maps** have been produced, one assuming the most recent strand of data, and the other picking the max value from the three older ones.

A **probability kernel** was applied on the faults locations. This represented the possibility of a new vent opening at a random distance from a random fault section.

A **Cauchy kernel** was preferred to a Gaussian kernel, due to the major level of uncertainty linking the faults location to the vents:

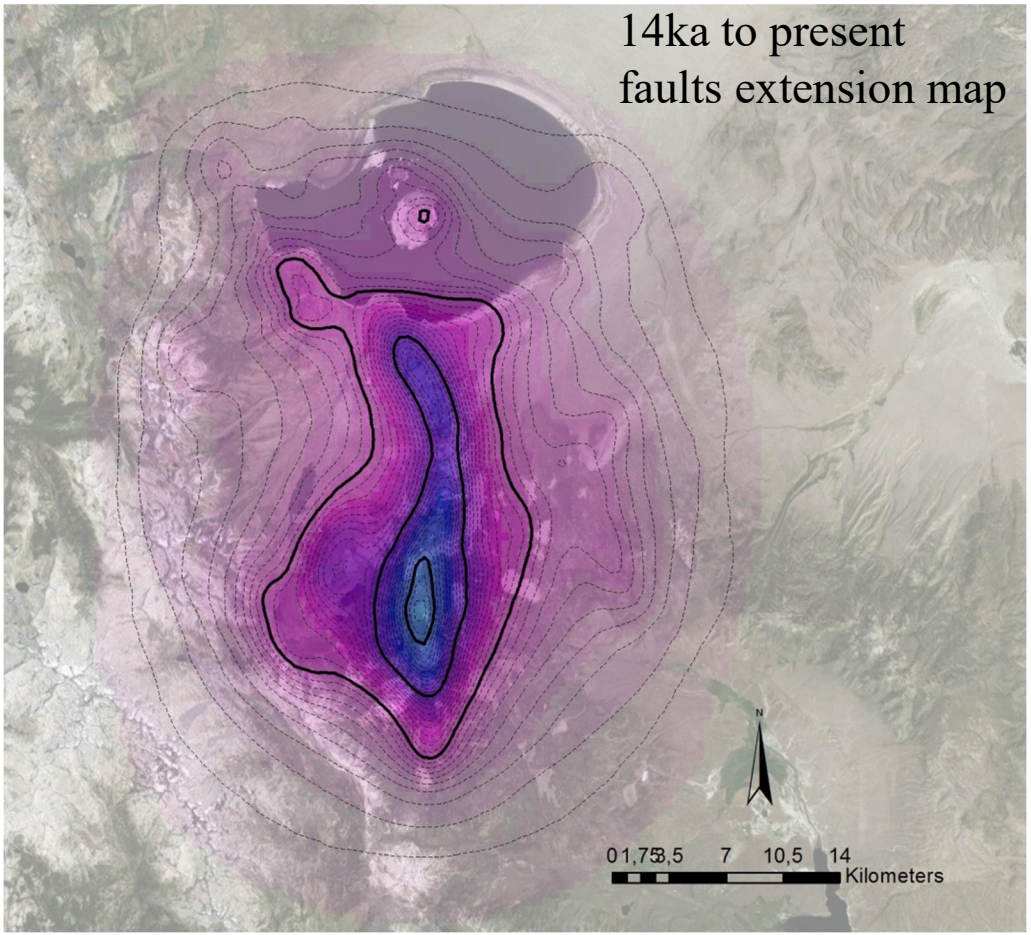
$$f(\mathbf{x}) = \frac{1}{\pi\gamma} \int_{M_f} \frac{\gamma^2}{|\mathbf{x} - \xi|^2 + \gamma^2} d\xi$$

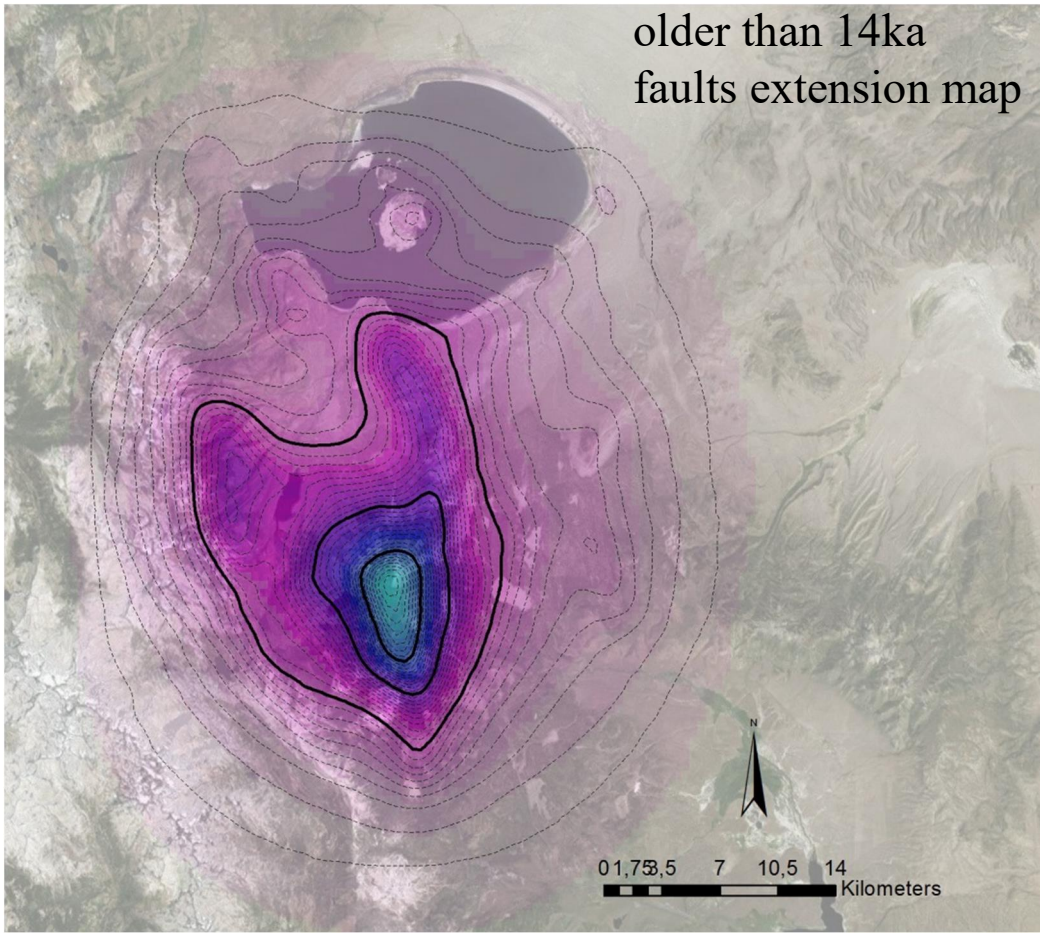
where γ is a scaling parameter and M_f is the faults data matrix.

The tectonic maps: kernel density estimation

A 100m **uncertainty buffer** was defined around the faults lines,
weighted in proportion to their **extension value in a log-scale**.

A one-dimensional Cauchy distribution was assumed on the distance (isotropic).
The 90% of the mass of the distribution was imposed to fell inside a **10km radius** - a proxy for the tectonic wavelength. A minimum value of 0.01 mm/yr was imposed.

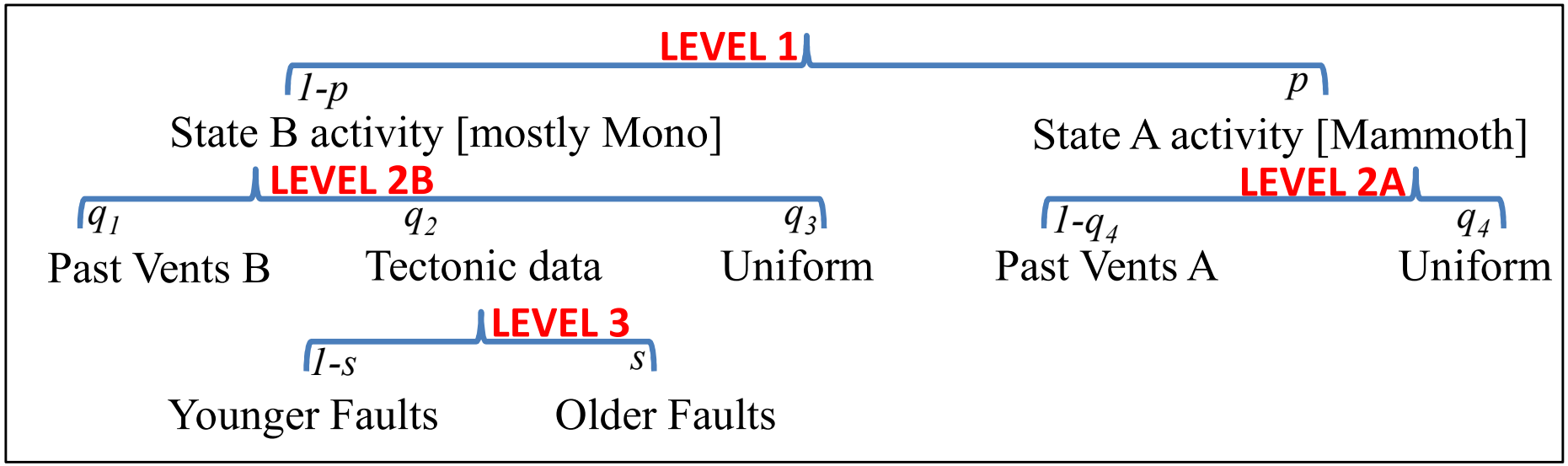




Contour lines report the vent opening probability percentage per km^2.
The bold lines are 0.1% paced, the thin pointed lines are 0.01% paced.

The logic tree for LVVR

To simplify/constrain the quantification of the weights of each spatial distribution to combine, a **simple hierarchical logic tree** was defined.



Each branch compared the **relative importance** of one variable or feature of the system versus others.

Level 1 relied on the unknown parameter *p*, ranging in [0 - 47.43%].

Level 2B on the parameter array (q_1,q_2,q_3), fixed through a "cross-validation" inference on the events younger than 40ka.

Level 2A was based on the parameter q_4, preliminarily fixed to 66.67% (this is under development).

Level 3 on the unknown parameter *s*, ranging in [0 - 100%].

The single estimates over each branch were multiplied in a **Monte Carlo simulation** to obtain the probability distributions of the weights.

Bayesian model averaging for the vent opening maps

A Bayesian model averaging approach was applied for defining the weights (q_1,q_2,q_3).

The basic idea was again to **weight** the different models in proportion to the **likelihood** that they give to the past data – a "**cross-validation**" procedure was accomplished on the < 40ka BP part of the dataset.

The algorithm was similar to what done for choosing the vent propagation distance model, except for an important detail: in this case was assumed a potentially **different model choice** for each of the events in the validation set.

For this reason the likelihood and the posterior weights of model *j* were calculated separately for each vent *i* in the 28 in the validation set.

Their average permitted to give an estimate for (q_1,q_2,q_3): $L_j = \sum_{i=1}^{M} L_j^i / M.$

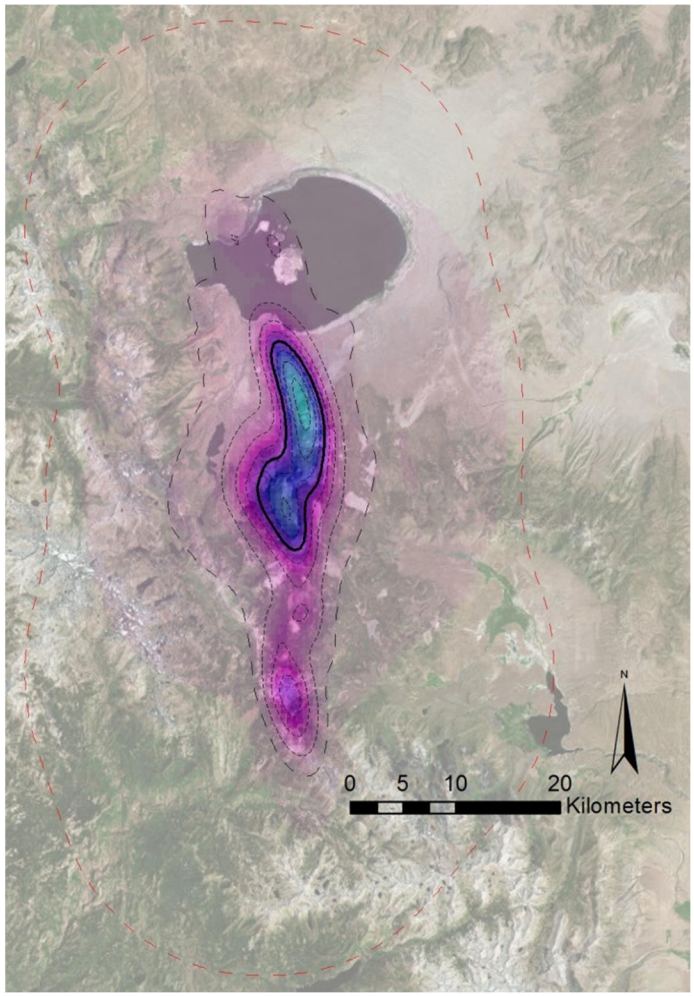
In particular:

	Past vents	Tectonics	Uniform
5%ile	58.71%	27.10%	6.15%
mean	61.56%	31.38%	7.06%
95%ile	64.67%	35.14%	8.23%

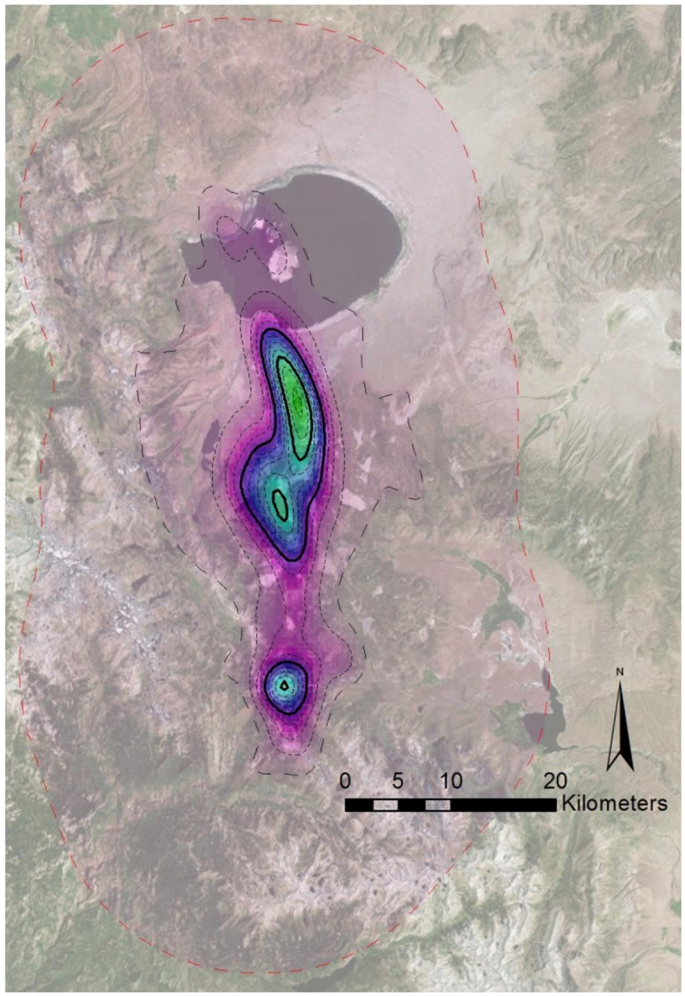
The epistemic uncertainty affecting the values is an effect of the random choice of the tectonic map.

Preliminary vent opening maps

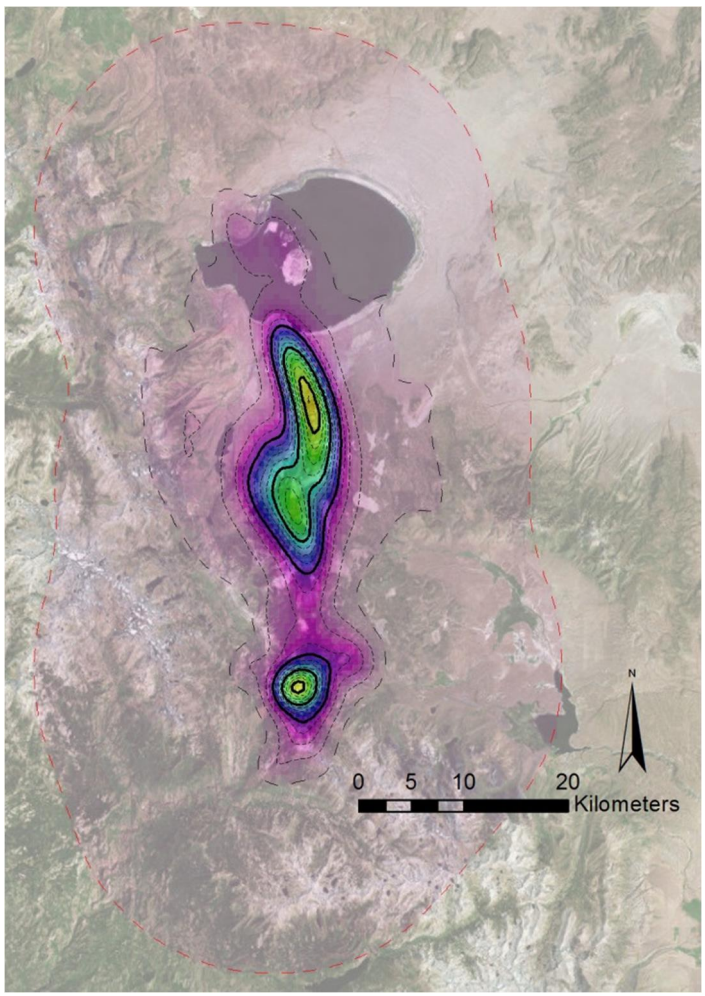
5%ile map



Mean map



95%ile map



Contour lines report the vent opening **probability percentage per km^2**.
The bold lines are 0.2% paced, the thin pointed lines are 0.05% paced.
The additional dashed line remarks the 0.02% level.
The thin red dashed line marks the current domain boundary, fixed at **20km** from the past activity (this could be modified according to the large scale morphology of the region).

The maps are defined on a matrix of 200x200, cell size 500m, lower left corner coordinates: 2710.501, 41342.49.

Concluding remarks

- **Doubly stochastic models** are a very general tool for assessing random systems that depend on uncertain information/interpretation, as in the case of volcanic processes.

- Preliminary vent opening maps for LVVR were produced as the **linear combination** of five spatial distributions, with a **main epistemic uncertainty** source concerning states A and B of past activity.

- The '**logic tree**' approach is an easy method for combining alternative vent opening maps based on key volcanologic features, inside a doubly stochastic model.

- The **Bayesian average** is a powerful tool for model selection, and the **probability kernels** for describing vent propagation distance from unknown regions of past data spatial distribution.

These are results still under development and there are some main open problems:
1) more detailed information/uncertainty quantification is needed on the **past events number** and concerning **Mammoth region tectonic data**.

2) **tectonic data** could be assumed as a **prior map to be updated** with past vents locations data, their likelihood function relying on a physical model for the fault/vent propagation.
This would lead to an alternative model.