Enhancing the Failure Forecast Method using a noisy mean-reverting process

Andrea Bevilacqua[1,2], E. Bruce Pitman[3,1], Abani K. Patra[4,2,1], Augusto Neri[5]

(1) Center for Geohazard Studies, Department of Earth Sciences, University at Buffalo, SUNY
(2) Computational Data Science and Engineering Program, University at Buffalo, SUNY
(3) Department of Materials Design and Innovation, University at Buffalo, SUNY
(4) Department of Mechanical and Aerospace Engineering, University at Buffalo, SUNY
(5) Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Pisa.

ID 279, S02.13
Strategies for eruption forecasting and early-warning to support operational hazard evaluation during volcanic crises

3 Settembre 2018, Napoli

Cities on Volcanoes 10
NAPOLI | ITALY SEPTEMBER 2-7 2018

The Failure Forecast Method (FFM)

The FFM is a well-established tool in the interpretation of monitoring data as possible **precursors**, providing quantitative predictions of the eruption onset (Voight, 1988).

The model represents the potential **cascade** of precursory signals leading to a significant rupture of materials, with t_f a proxy for the eruption onset t_e.

The FFM has been retrospectively applied to several volcanic systems, including dome forming and **explosive** eruptions.

Seismic data are the type of signals most extensively studied with the method.

EXAMPLE

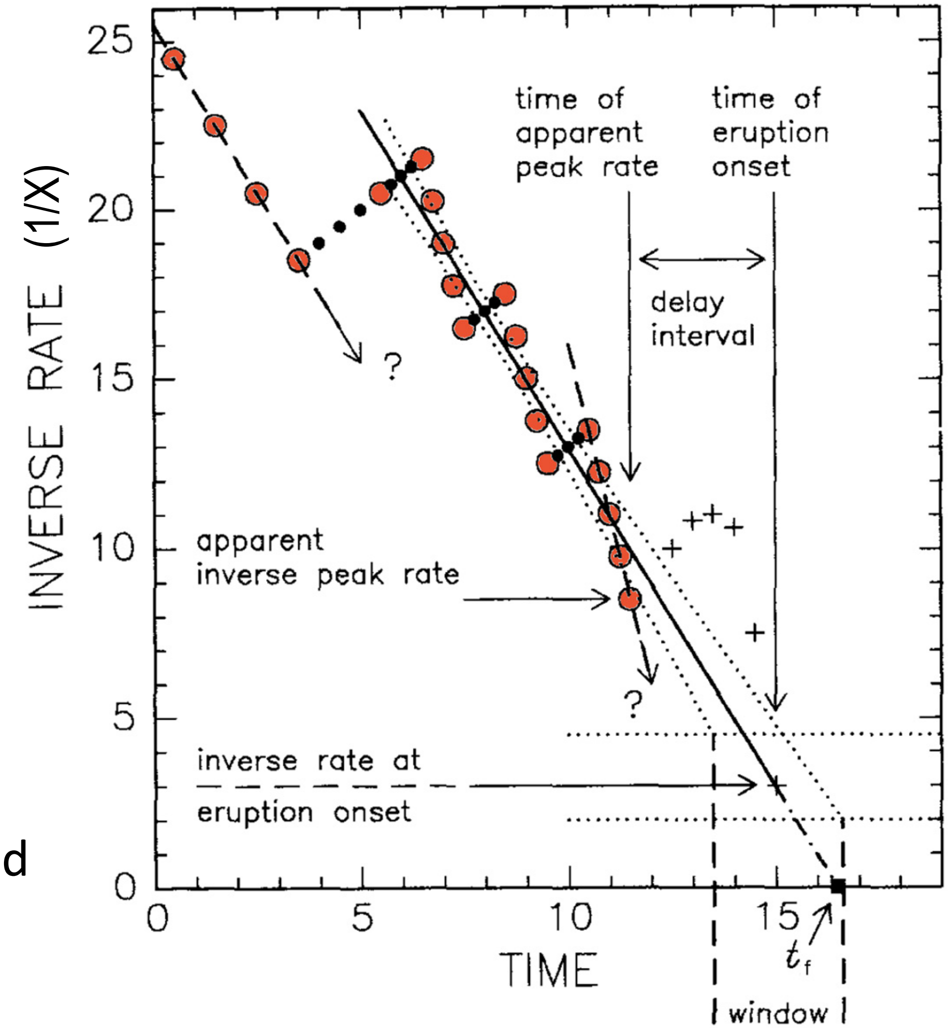


Figure. Example scheme of linear regression of inverse rate of cascading signals (modified from Cornelius & Voight, 1995)

Motivations and Outline

FFM is known to be affected by sources of **uncertainty**, like:

- the occurrence of multiple phases of acceleration in the signals
- the superposition of signals originating from different causes
- heterogeneity in the breaking material, producing changes in the signals.

In addition, the statistical fitting of model parameters can be poorly constrained.

A **full probability assessment** of FFM, with uncertainty quantification, is the motivation of this study.

In particular, we enhance the classical FFM by:

- systematically characterizing the **uncertainty**, including both aleatoric and epistemic sources;
- incorporating a stochastic **noise** in the equations, and a mean-reversion property to constrain it.

We retrospectively test the enhanced FFM over four datasets from Voight, 1988.

These refer to:
St. Helens, 1981-82,
Bezymianny, 1960,
Mt. Toc (Vajont), 1963.

Our aim is to produce **probability forecasts** with the FFM, instead of deterministic predictions.

The FFM differential equations (ODE)

α - convexity parameter
A - slope parameter
t_0 - initial time

$$\boxed{dX/dt = AX^{\alpha}}$$

where X is the time rate of signals

$$\boxed{X(t) = \left[(1-\alpha)A(t-t_0) + X(t_0)^{1-\alpha}\right]^{\frac{1}{1-\alpha}}}$$

SOLUTION

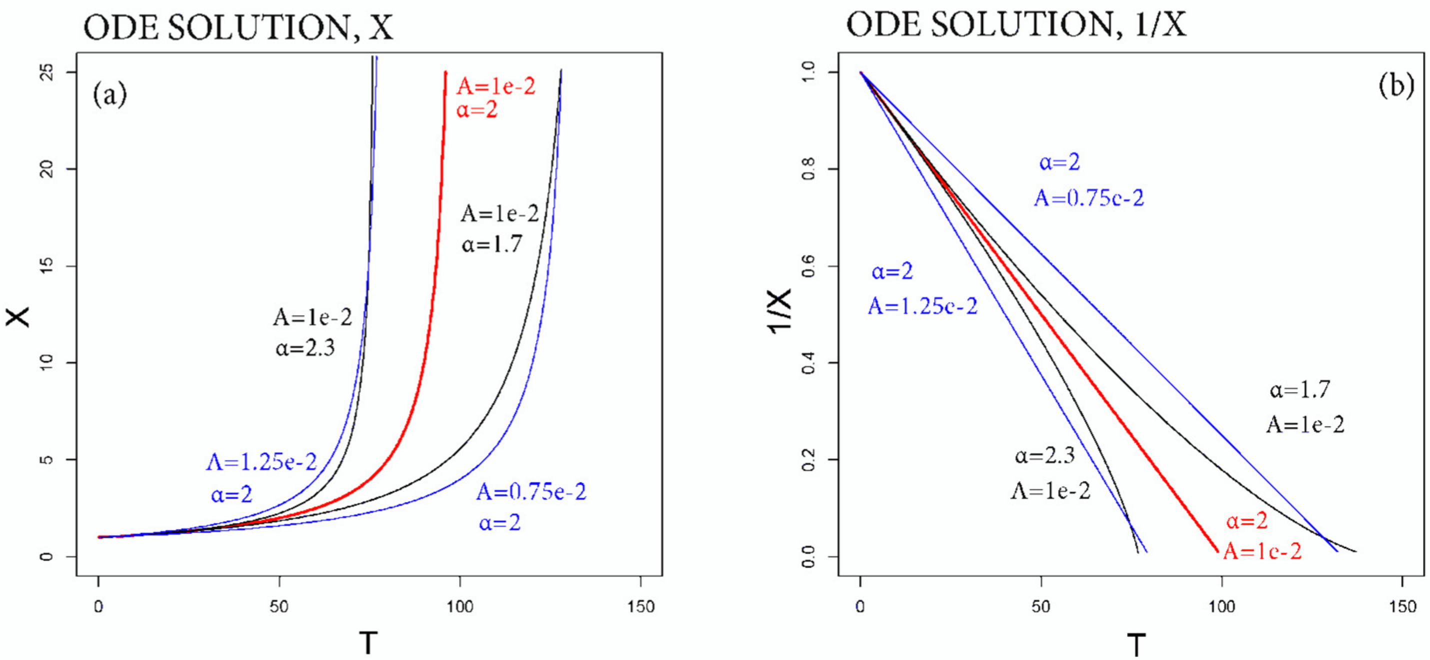


Figure.
ODE solution,
(a) X, and (b) 1/X.

Note the effect of varying slope and convexity parameters.
α, A learnt from data

$$\eta := X^{1-\alpha}$$

change of variables $\Longrightarrow$ *linearization*

$$\frac{d\eta}{dt} = (1-\alpha)A$$

$$\eta(t) = (1-\alpha)A(t-t_0) + \eta(t_0)$$

t_f - failure time

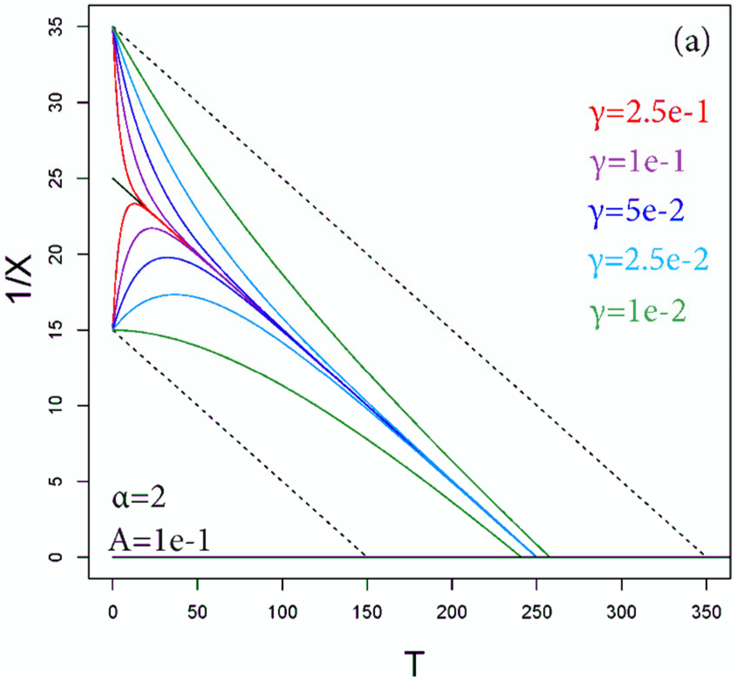
$$\boxed{t_f = \inf\{t : \eta(t) = 0\}}$$

The new terms: mean-reversion and stochastic noise

$$\eta(t) = (1-\alpha)A(t-t_0) + \underbrace{\beta\exp(-\gamma t)}_{\text{mean-reversion term}} + \eta(t_0)$$

It makes every perturbation decay with time

MEAN-REVERTING EFFECT



STOCHASTIC NOISE EFFECT

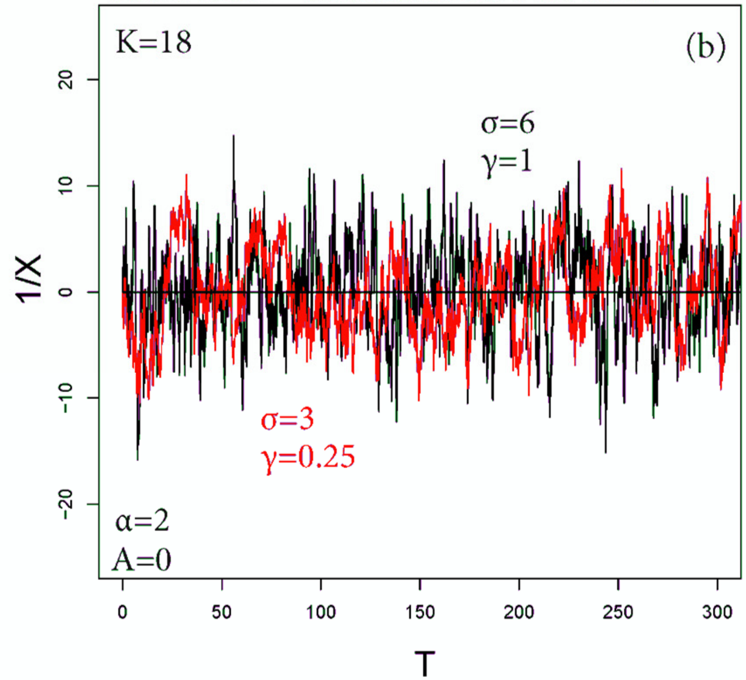


Figure.
(a)
1/X with $\alpha=2$, $A=0.1$, $\beta=\pm 10$. The colors show different γ.

(b)
$(\eta_t)_{t>0}$ with $A=0$. The colors show different (γ, σ), with equal $K=\sigma^2/\gamma$.

$$d\eta_t = -\gamma\eta_t dt + \underbrace{\sigma dW_t}_{\text{noise term}}$$

(A=0)

γ - mean-reversion parameter
σ - noise parameter
β - initial perturbation

SOLUTION

$$\eta_t \sim \mathcal{N}\left(0, \frac{\sigma^2}{2\gamma}\left[1-\exp(-2\gamma t)\right]\right) \simeq \mathcal{N}\left(0, \frac{\sigma^2}{2\gamma}\right)$$

Parameters are based on the residuals in the linearized problem.

The FFM stochastic differential equations (SDE)

$$d\eta_t = \{\underbrace{\gamma\,[(1-\alpha)A(t-t_0)+\eta_{t_0}-\eta_t]}_{\text{mean-reversion terms}} + \underbrace{(1-\alpha)A\}\,dt}_{\text{classical FFM}} + \underbrace{\sigma dW_t}_{\text{noise term}}$$

$$X_t = \left\{X_{t_0}^{1-\alpha} + \int_{t_0}^{t}\{\gamma\,[(1-\alpha)A(s-t_0)+X_{t_0}^{1-\alpha}-X_s^{1-\alpha}]+(1-\alpha)A\}\,dt + \int_{t_0}^{t}\sigma dW_s\right\}^{\frac{1}{1-\alpha}}$$

NONLINEAR FORMULATION

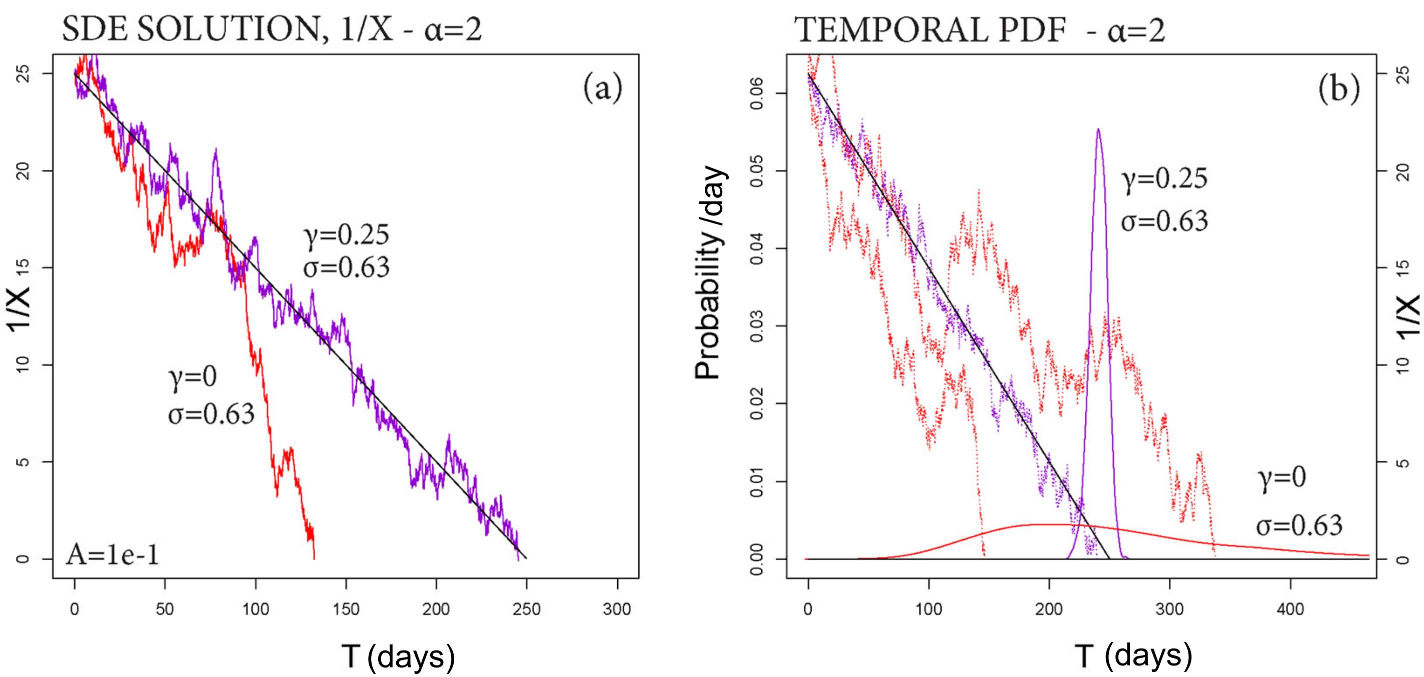


Figure. SDE solutions, with α=2, A=0.1.

The black line is the mean solution.

(a) colored lines are random paths, γ=0 or γ=0.25.

(b) also shows g_{tf}. The solution 1/X is reported again.

$$t_f(\omega) = \inf\{t : X^{-1}(\omega,t)=0\}$$

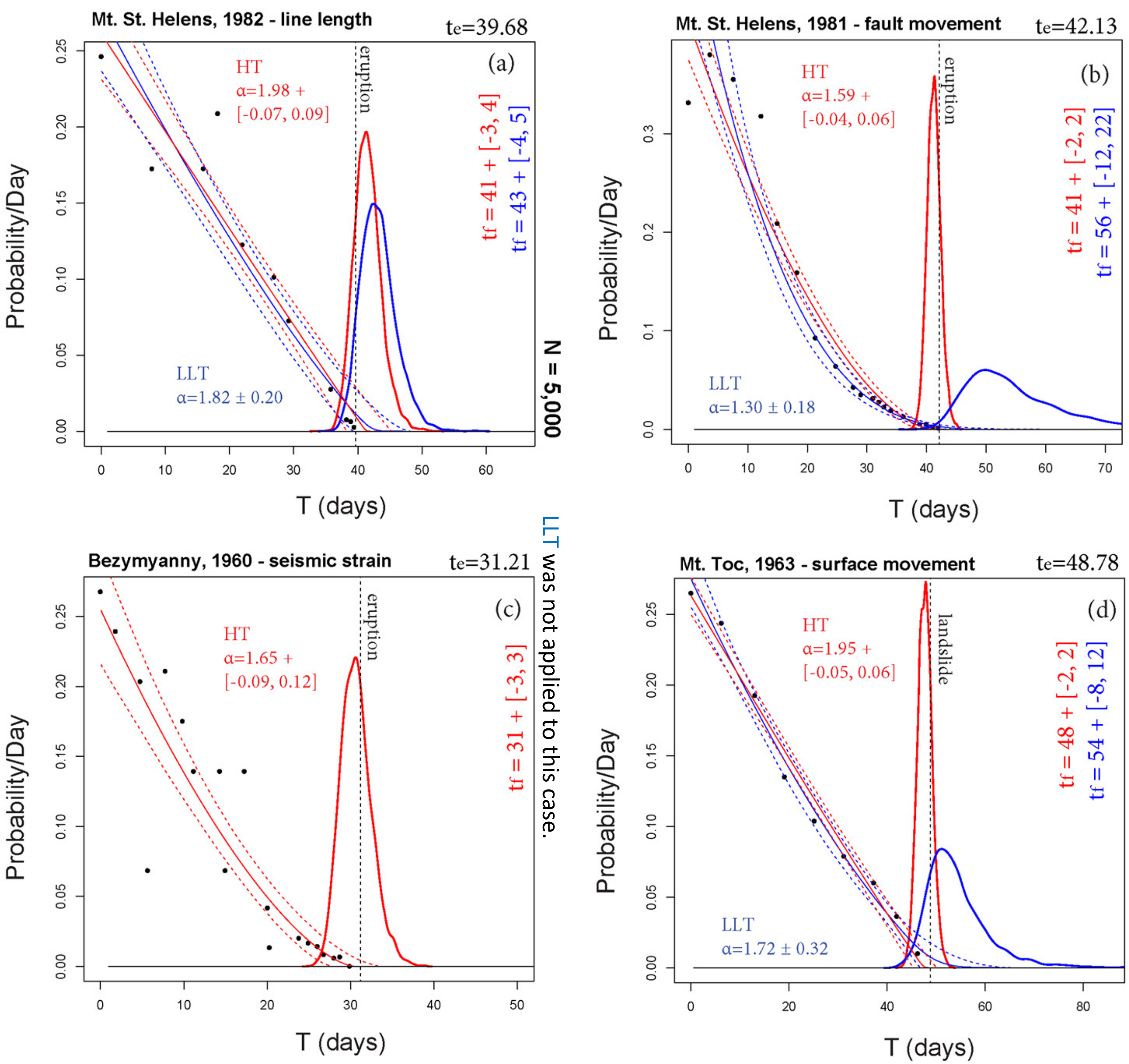
random variable

$$g_{t_f} : \mathbb{R} \rightarrow \mathbb{R}+, \quad \int_0^\infty g_{t_f}(x)dx = 1$$

t_f probability density function

Method 1, ODE estimators of t_f

Method 1 characterizes the **epistemic uncertainty** related to the parameter fitting in the classical FFM.



The Log-rate vs Log-acceleration Technique (LLT), and the Hindsight Technique (HT) are estimators of α (Cornelius & Voight, 1995).

HT requires that we know t_e and hence can only be used in retrospective analysis.

LLT is less accurate than HT and needs a second order derivative of data.

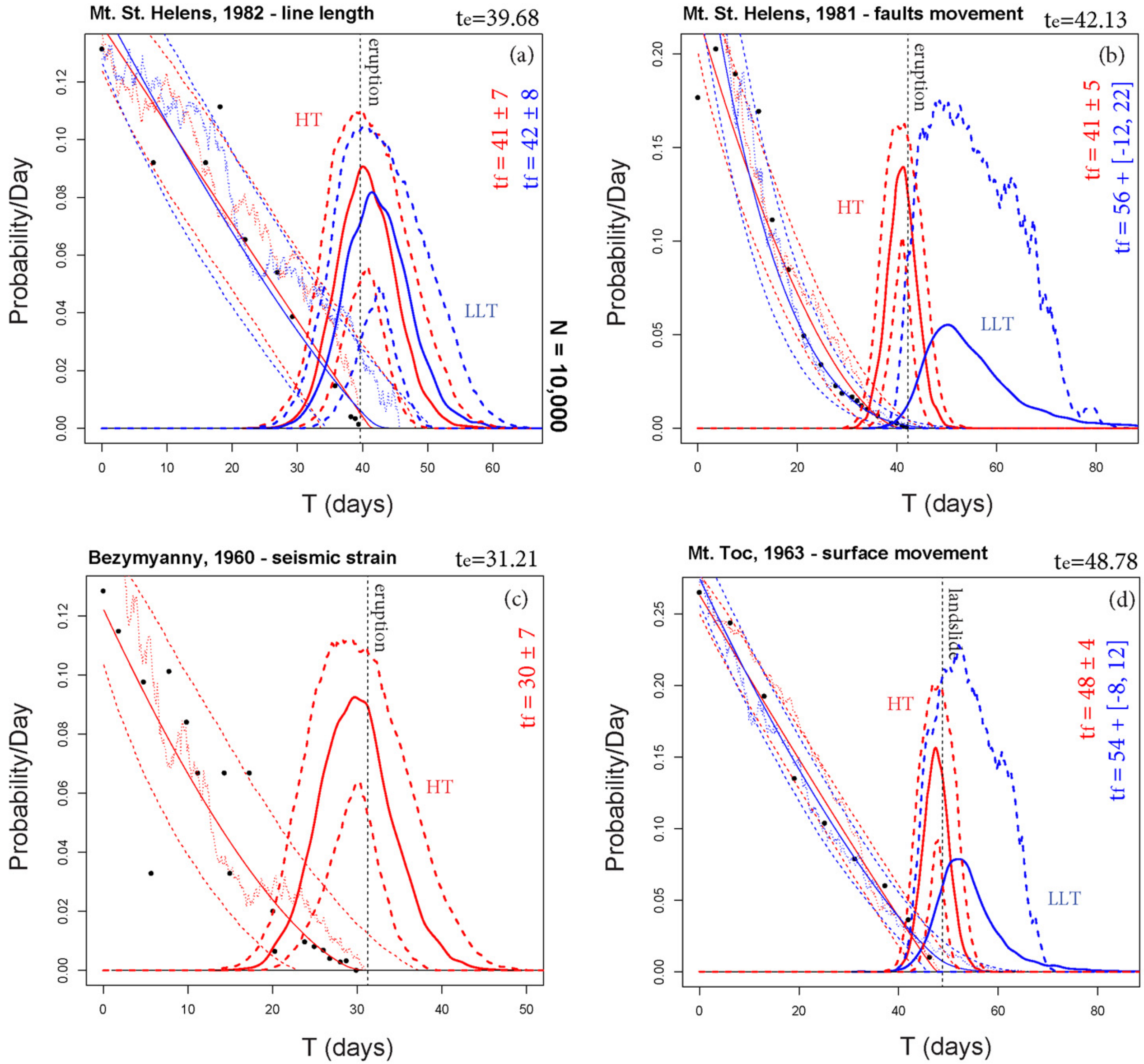
Figure.
Colored lines assume α as from LLT or HT.

The bold line is g_{tf}. Dashed lines bound a 90% confidence of solutions.

A dashed vertical line is t_e, the black dots are data.

Method 2, doubly stochastic estimators of t_f

Method 2 allows **excursions** from the classical FFM solutions, modeling **aleatoric uncertainty** sources.
It also models the epistemic uncertainty related to parameter fitting, like in Method 1.



Our doubly stochastic formulation allows users to determine a "**worst case scenario"** with a specified level of confidence (Bevilacqua, 2016).

Figure.
Colored lines assume α as from LLT or HT.

The bold line is g_{tf}.
Bold dashed lines are its 5th and 95th percentile values.

Dashed lines bound a 90% confidence of solutions.
Dotted lines show examples of random paths.

A dashed vertical line is t_e, the black dots are data.

Material failure likelihood - $g_{tf}(t_e)$

The reported value is the pdf in the day t_e, as displayed in the previous figures.

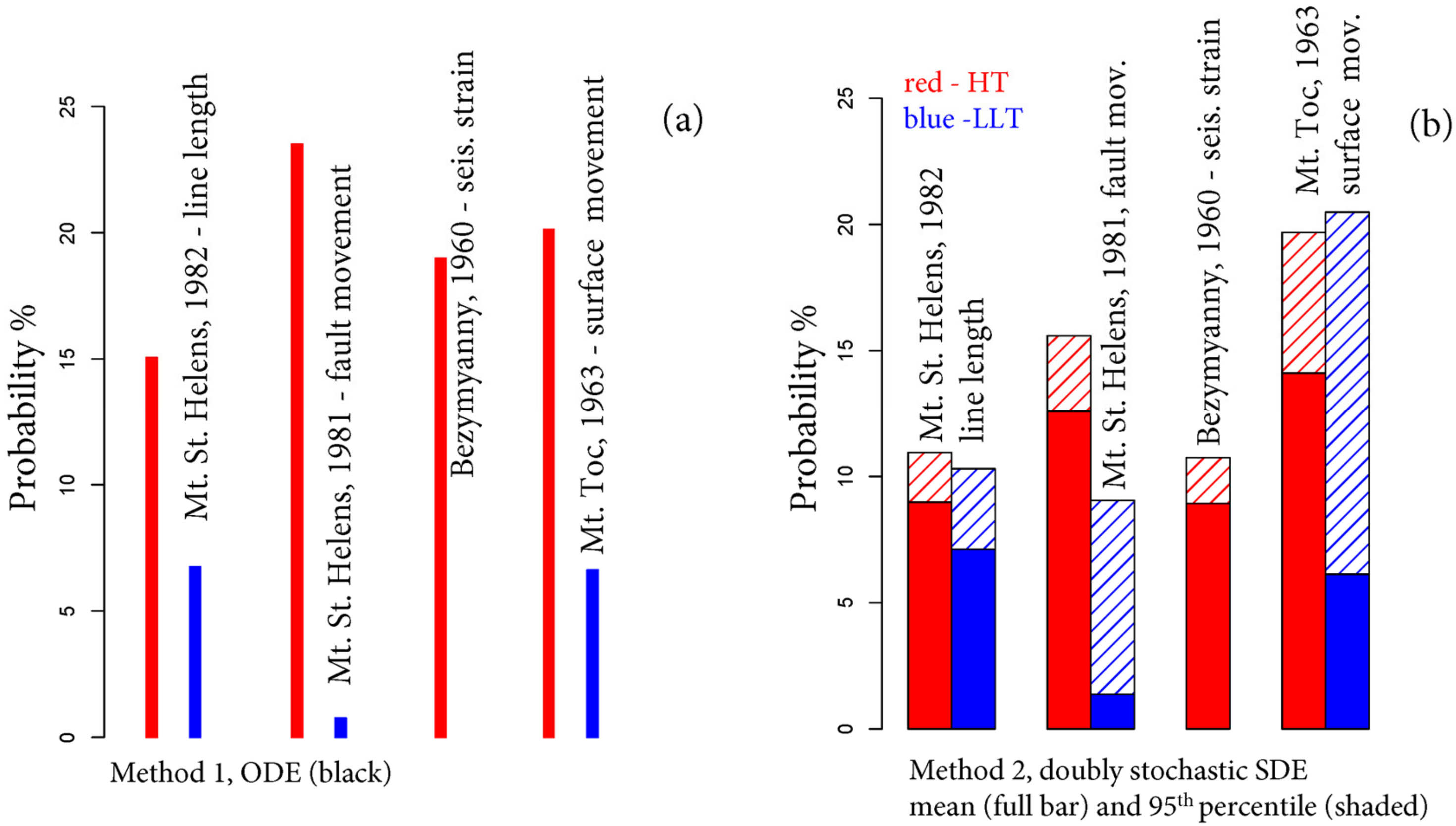


Figure. Barplots of the likelihood $g_{tf}(t_e)$.

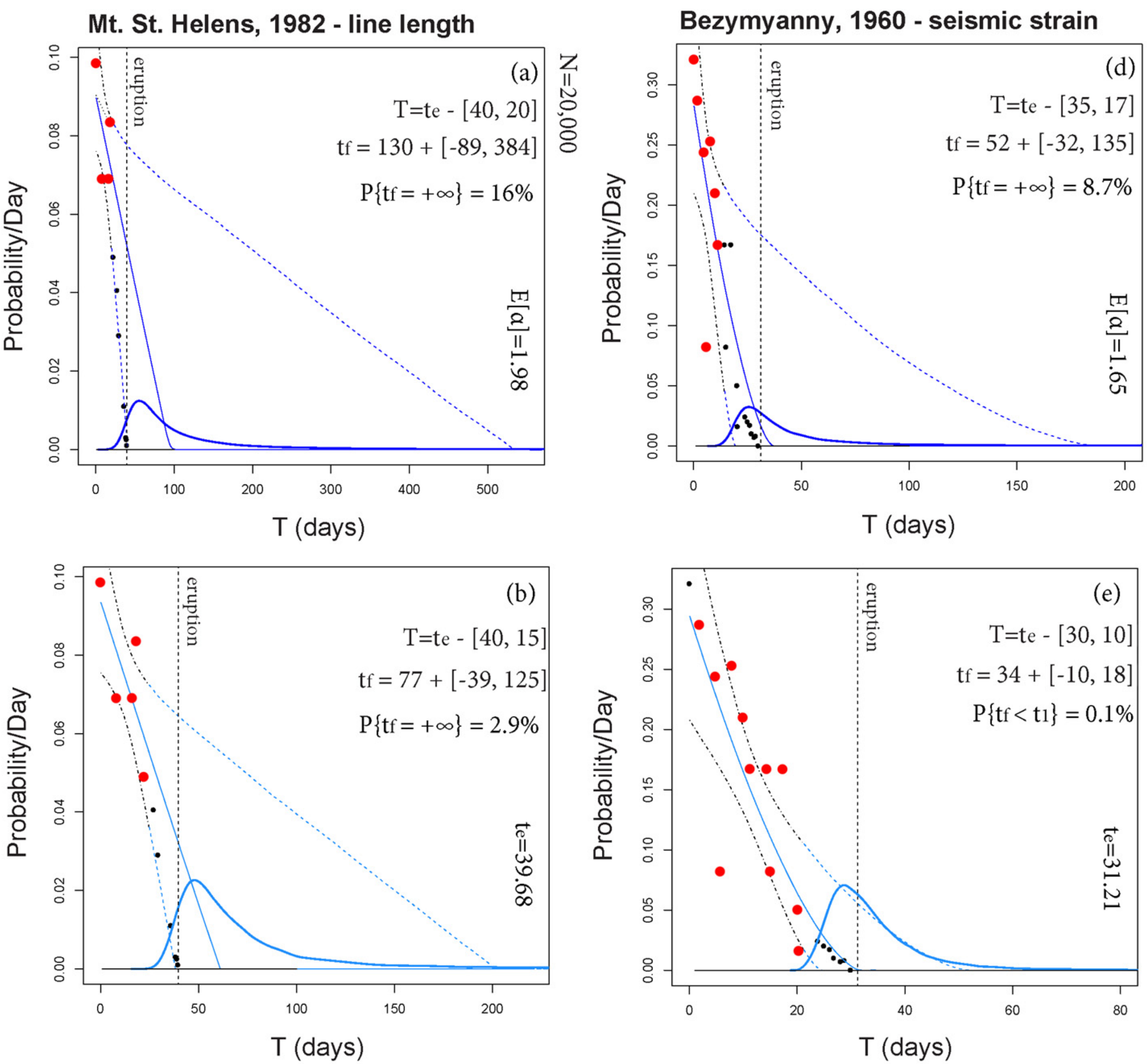
(a) the colored bars assume Method 1.

(b) assumes Method 2. Full bars are the mean values, shaded bars are the 95th percentile values of the likelihood.

If α is based on LLT Method 1 provides low likelihoods, below 1% in some case. The 95th percentile values of Method 2 clearly **outperforms** Method 1.

We remark that the HT method cannot be used in forward forecasting, but only retrospectively.

Method 1, ODE forecasts of t_f



We compare two **time windows** with extremes reported in figure. They include different subsequences of data.

Estimators based on the whole sequence of signals are not **forecasts** (Boué et al. 2015).

We assume that data are available only up to time $t<t_e$.

Forecasts can be significantly **uncertain**, because based on fewer data.

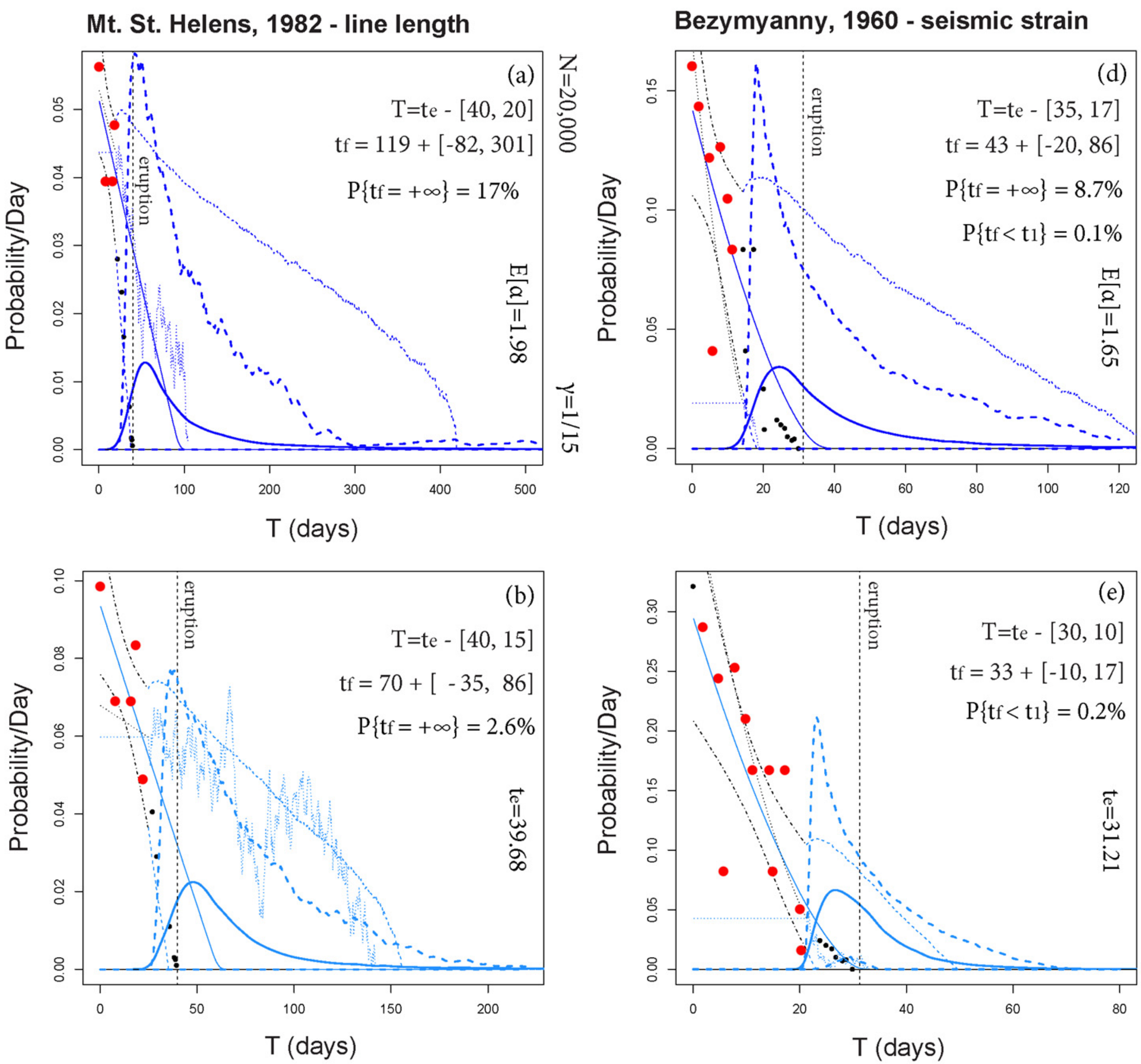
Figure.
Forecasts of t_f based on different time windows T.

The bold line is g_{tf}.
Dashed lines bound a 90% confidence of solutions.

A dashed vertical line is t_e.
The dots are data, in red if belonging to T.

Method 2, doubly stochastic forecasts of t_f

If the forecast is poorly constrained, Method 2 typically **reduces the uncertainty** affecting t_f, compared to Method 1. Indeed the noise can push $1/X$ to hit zero, when $1/X$ is small enough.



The **doubly stochastic** formulation of Method 2 appears to have an impact.

Figure. Forecasts of t_f based on different time windows T.

The bold line is g_{tf}. Bold dashed lines are its 5^{th} and 95^{th} percentile values.

Dashed lines bound a 90% confidence of solutions. Dotted lines show examples of random paths.

A dashed vertical line is t_e. The dots are data, in red if belonging to T.

Material failure likelihood - $g_{tf}(t_e)$

Method 2 mean forecasts provide consistent likelihoods with Method 1.

The 95th percentile values are **significantly higher** than other forecasts, from 5% to 10% in the first and second time windows, and above 15% in the third.

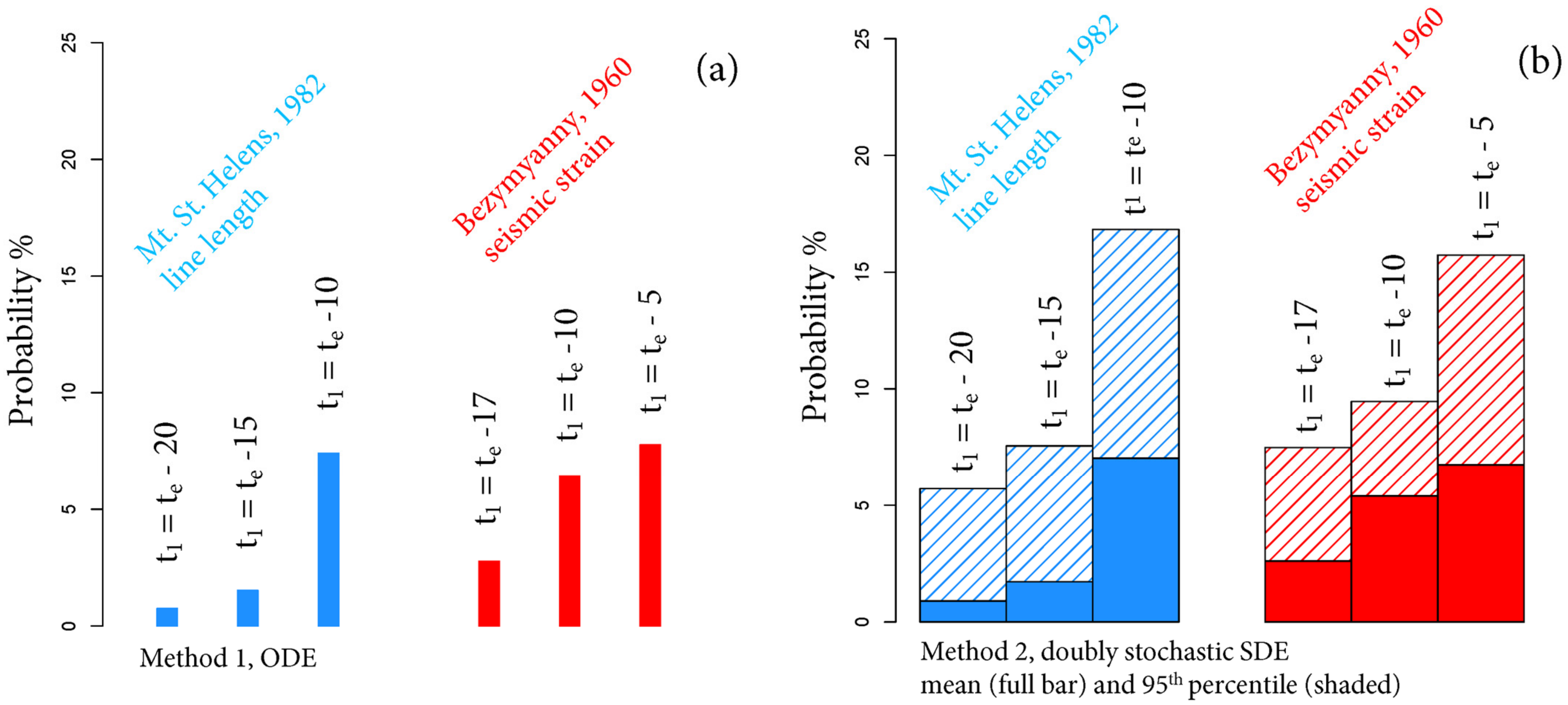


Figure. Barplots of the likelihood $g_{tf}(t_e)$ on three time windows. In (a) the colored bars assume Method 1. (b) assumes Method 2. Full bars are the mean values, shaded bars are the 95th percentile values.

- Summary and conclusions -

We have introduced a new method for performing short-term eruption timing forecasts, when the eruption onset is related to a significant rupture of materials.

- The method enhances the well known FFM equation. We allow random excursions from the classical solutions. This provides **probability forecasts** instead of deterministic predictions.

- Our doubly stochastic formulation can consider the “**worst case scenario”** with a probability of occurrence of at least 5%. This was not possible in the classical formulation.

- We compared two formulations of the method on historical datasets of precursory signals. The data show the **increased forecasting skill** of the doubly stochastic formulation, expressed as the likelihood in the day of the actual eruption.

This approach is the subject of ongoing and future work, with the purpose to further test its forecasting robustness over more complex sequences of signals.

Doubly stochastic enhancement of the Failure Forecast Method using a noisy mean-reverting process and application to volcanic eruption forecasts, https://arxiv.org/abs/1805.11654.

This research is supported by:

- *Collaborative research: Using precursor information to update probabilistic hazard maps*, National Science Foundation (NSF).

- *Sale Operative Integrate e Reti di Monitoraggio del Futuro: l’INGV 2.0*, Fondo Integrativo Speciale per la Ricerca (FISR).