Supplement of

A robust multi-indicator framework for landslide early warning using complementary statistical physics-based diagnostics

Qinghua Lei[1], Didier Sornette[2]

[1]Department of Earth Sciences, Uppsala University, Uppsala, 752 36, Sweden
[2]Institute of Risk Analysis, Prediction and Management, Academy for Advanced Interdisciplinary Studies, Southern University of Science and Technology, Shenzhen, 518055, China

Contents of this file

Introduction

This document provides supplementary materials to complement the methods, results, and discussions in the main text. Text S1 describes the method for estimating inverse gamma parameters. Text S2 details the LPPLS fitting procedure. Text S3 presents the derivation of LPPLS diagnostic metrics. Figure S1–S3, S4–S6, and S7–S9 show additional result analysis of the Preonzo, Veslemannen, and Stampa landslides, respectively.

Text S1. Parameter estimation for the inverse gamma distribution

For N measurements of slope velocities, $\mathbf{V} = \{v_1, v_2, \ldots, v_N\}$, they are first ranked in an ascending order as $\mathbf{V} = \{v_{(1)} \le v_{(2)} \le \ldots \le v_{(N)}\}$. To estimate a, b, and c from a subset $\mathbf{v} = \{v_{(1)} \le v_{(2)} \le \ldots \le v_{(n)}\} \subseteq \mathbf{V}$, $n \le N$, the following likelihood function is constructed:

$$L(a,b,c;\mathbf{v}) = \prod_{i=1}^{n} \frac{a^b}{\Gamma(b)} \left(\frac{1}{v_i - c}\right)^{b+1} \exp\left(-\frac{a}{v_i - c}\right), \tag{S1}$$

with the corresponding log-likelihood function given as:

$$l(a,b,c;\mathbf{v}) = \ln L(a,b,c;\mathbf{v}) = nb\ln a - n\ln\Gamma(b) - (b+1)\sum_{i=1}^{n}\ln(v_i - c) - a\sum_{i=1}^{n}\frac{1}{v_i - c}, \tag{S2}$$

The parameter values can be estimated by solving the following maximisation problem:

$$[\hat{a},\hat{b},\hat{c}] = \arg\max_{a,b,c} l(a,b,c;\mathbf{v}), \tag{S3}$$

which is associated with strong nonlinearity. To address this issue, the profile likelihood method is used (Barndorff-Nielsen and Cox, 1994) with a and b treated as nuisance parameters, such that the profile log-likelihood is given by:

$$l_{\mathrm{p}}(c;\mathbf{v}) = \max_{a,b|c} l(a,b,c;\mathbf{v}) = l(\hat{a},\hat{b},c;\mathbf{v}), \text{ with } [\hat{a},\hat{b}] = \arg\max_{a,b|c} l(a,b,c;\mathbf{v}). \tag{S4}$$

We scan a grid of fixed values of c within a plausible range $[-v_{\max}, v_{\min})$ with $v_{\max}$ and $v_{\min}$ being the maximum and minimum values in $\mathbf{v}$, respectively. To reduce computational cost, the search is first performed on a coarse grid and then adaptively refined in the vicinity of the maximum. For each fixed value of c, the parameters a and b are estimated by enforcing $\partial l/\partial a = \partial l/\partial b = 0$, which yields:

$$\frac{n\hat{b}}{\hat{a}} - \sum_{i=1}^{n}\frac{1}{v_i - c} = 0, \tag{S5}$$

$$n\ln\hat{a} - n\psi(\hat{b}) - \sum_{i=1}^{n}\ln(v_i - c) = 0, \tag{S6}$$

where $\psi(\cdot)$ is the digamma function. The preceding equations can be further reduced to the following form:

$$g(\hat{b}) = \ln n + \ln\hat{b} - \ln\sum_{i=1}^{n}\frac{1}{v_i - c} - \psi(\hat{b}) - \frac{1}{n}\sum_{i=1}^{n}\ln(v_i - c) = 0, \tag{S7}$$

and solved using the Newton-Raphson method (Süli and Mayers, 2003) with an initial guess for b obtained from the method of moments (Forbes et al., 2011) as $[\mu(\mathbf{v}-\mathrm{c})/\sigma(\mathbf{v}-\mathrm{c})]^2+2$, where $\mu(\cdot)$ and $\sigma(\cdot)$ denote the mean and standard deviation. Finally, the estimate of c is obtained from:

$$\hat{c} = \arg\max_{c} l_{\mathrm{p}}(c;\mathbf{v}). \tag{S8}$$

Text S2. Calibration scheme for the LPPLS model

We adopt a stable and robust scheme (Filimonov and Sornette, 2013) to calibrate the LPPLS model to time series data. Consider the time series data of observations $\Omega(t_i)$ recorded at times $t_i \in [\tau, t_{\mathrm{end}}]$, with $i = 1, \ldots, n$, where n is the number of time stamps, τ and t_{end} respectively denote the start and end times of the time window for fitting the LPPLS model. We rewrite the LPPLS formula as:

$$\Omega(t) = A + B(t_{\mathrm{c}} - t)^m + C_1(t_{\mathrm{c}} - t)^m \cos\left[\omega\ln(t_{\mathrm{c}} - t)\right] + C_2(t_{\mathrm{c}} - t)^m \sin\left[\omega\ln(t_{\mathrm{c}} - t)\right], \tag{S9}$$

where $C_1 = C\cos\phi$ and $C_2 = C\sin\phi$, such that the parameter set $\mathbf{\theta} = \{A, B, C, t_c, m, \omega, \phi\}$ (with three linear and four nonlinear parameters) of the original LPPLS formula given by equation (1) in the main text is transformed into a new parameter set $\mathbf{\theta} = \{A, B, C_1, C_2, t_c, m, \omega\}$ (with four linear and three nonlinear parameters). We define the sum of squared residuals as the cost function:

$$r^2(\mathbf{\theta};\mathbf{\Omega},\mathbf{t}) = \sum_{i=1}^{n}\left\{\Omega_i - A - B(t_c - t_i)^m - C_1(t_c - t_i)^m \cos\left[\omega\ln(t_c - t_i)\right] - C_2(t_c - t_i)^m \sin\left[\omega\ln(t_c - t_i)\right]\right\}^2, \quad \text{(S10)}$$

which is minimized using the ordinary least squares method to estimate the model parameters:

$$\hat{\mathbf{\theta}} = \arg\min_{\mathbf{\theta}} r^2(\mathbf{\theta};\mathbf{\Omega},\mathbf{t}). \quad \text{(S11)}$$

The four linear parameters $\{A, B, C_1, C_2\}$ are enslaved to the three nonlinear ones $\{t_c, m, \omega\}$, such that the minimization problem is reduced to:

$$\{\hat{t}_c, \hat{m}, \hat{\omega}\} = \arg\min_{t_c, m, \omega} r^2(t_c, m, \omega), \quad \text{(S12)}$$

where the profiled cost function is defined as:

$$r^2(t_c, m, \omega) = \min_{A,B,C_1,C_2} r^2(A, B, C_1, C_2, t_c, m, \omega) = r^2(t_c, m, \omega, \hat{A}, \hat{B}, \hat{C}_1, \hat{C}_2). \quad \text{(S13)}$$

The estimates for the four linear parameters can be obtained by solving the optimization problem for fixed values of the nonlinear ones:

$$\{\hat{A}, \hat{B}, \hat{C}_1, \hat{C}_2\} = \arg\min_{A,B,C_1,C_2} r^2(A, B, C_1, C_2, t_c, m, \omega), \quad \text{(S14)}$$

whose solution can be obtained by analytically solving the following system of linear equations:

$$\begin{bmatrix} n & \sum f_i & \sum g_i & \sum h_i \\ \sum f_i & \sum f_i^2 & \sum f_i g_i & \sum f_i h_i \\ \sum g_i & \sum f_i g_i & \sum g_i^2 & \sum g_i h_i \\ \sum h_i & \sum f_i h_i & \sum g_i h_i & \sum h_i^2 \end{bmatrix} \begin{bmatrix} \hat{A} \\ \hat{B} \\ \hat{C}_1 \\ \hat{C}_2 \end{bmatrix} = \begin{bmatrix} \sum \Omega_i \\ \sum \Omega_i f_i \\ \sum \Omega_i g_i \\ \sum \Omega_i h_i \end{bmatrix}, \quad \text{(S15)}$$

where $f_i = (t_c - t_i)^m$, $g_i = (t_c - t_i)^m \cos\left[\omega\ln(t_c - t_i)\right]$, and $h_i = (t_c - t_i)^m \sin\left[\omega\ln(t_c - t_i)\right]$.

Text S3. Derivation of LPPLS diagnostic metrics

Let t_2 denote the analysis time (i.e. the current time up to which data are available), and t_c denote the predicted failure time obtained from the LPPLS model. For each sensor and each analysis time t_2, a probability density function (PDF) of t_c, denoted as $\rho_i(t_c|t_2)$, is obtained from bootstrap realisations of the LPPLS fit. When multiple sensors are available, the individual PDFs are combined into a joint PDF using the Bayesian product rule:

$$\rho(t_c \mid t_2) \propto \prod_{i=1}^{N_s} \rho_i(t_c \mid t_2), \quad \text{(S16)}$$

where N_s is the number of sensors. The joint PDF is subsequently normalised such that

$$\int \rho(t_c \mid t_2) dt_c = 1, \quad \text{(S17)}$$

From the joint PDF, several diagnostic metrics are extracted. The lead time is defined as the difference between the median predicted failure time and the analysis time:

$$L(t_2) = \text{median}[t_c \mid t_2] - t_2, \quad \text{(S18)}$$

which quantifies the expected remaining time to failure at each t_2 and can be used to assess potential drifting behaviour of the forecast over time.
The uncertainty width is quantified using the interquantile range of the predicted t_c:

$$\Delta_{t_c}(t_2) = t_c^{(q_{75})}(t_2) - t_c^{(q_{25})}(t_2), \tag{S19}$$

where $t_c^{(q_{75})}$ and $t_c^{(q_{25})}$ denote the 75th and 25th percentiles of the joint PDF, respectively.
The normalised fluctuation is defined as the ratio of the uncertainty width to the lead time:

$$\eta_{t_c}(t_2) = \frac{\Delta_{t_c}(t_2)}{L(t_2)}, \tag{S20}$$

which provides a dimensionless measure of the relative uncertainty in the forecast.
In addition, the evolution of the critical exponent m is derived, with its median and interquartile range across sensors at each analysis time tracked. These metrics provide complementary information on the evolving dynamics of the system and the stability of the LPPLS fits.

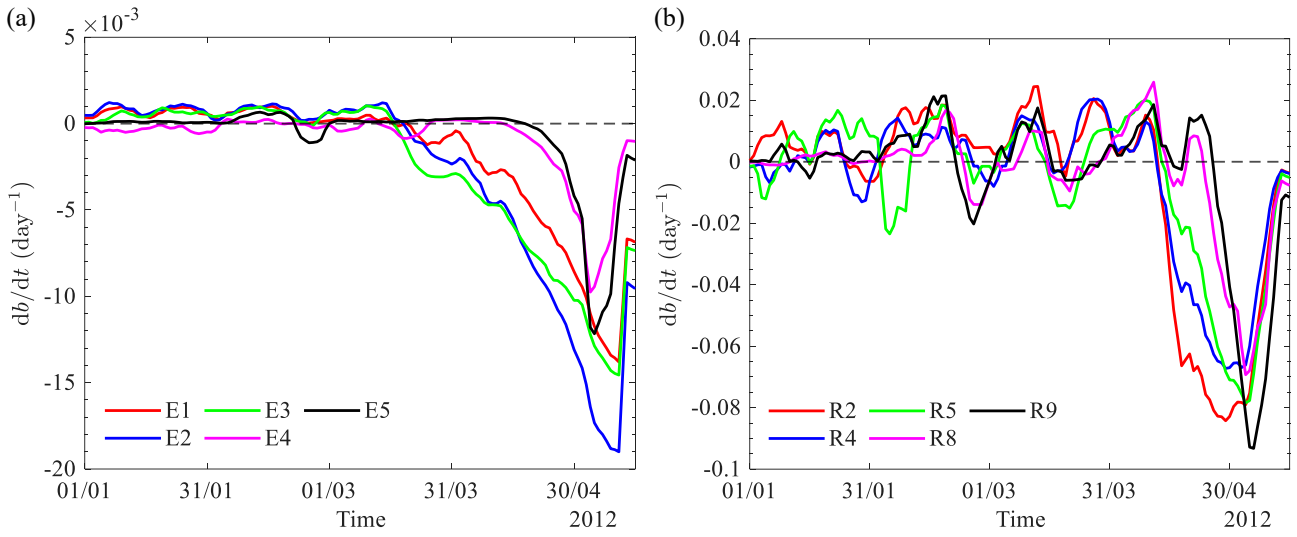

Figure S1. Temporal evolution of the time derivative of velocity *b*-value, derived from (a) extensometers and (b) reflectors installed at the Preonzo landslide.

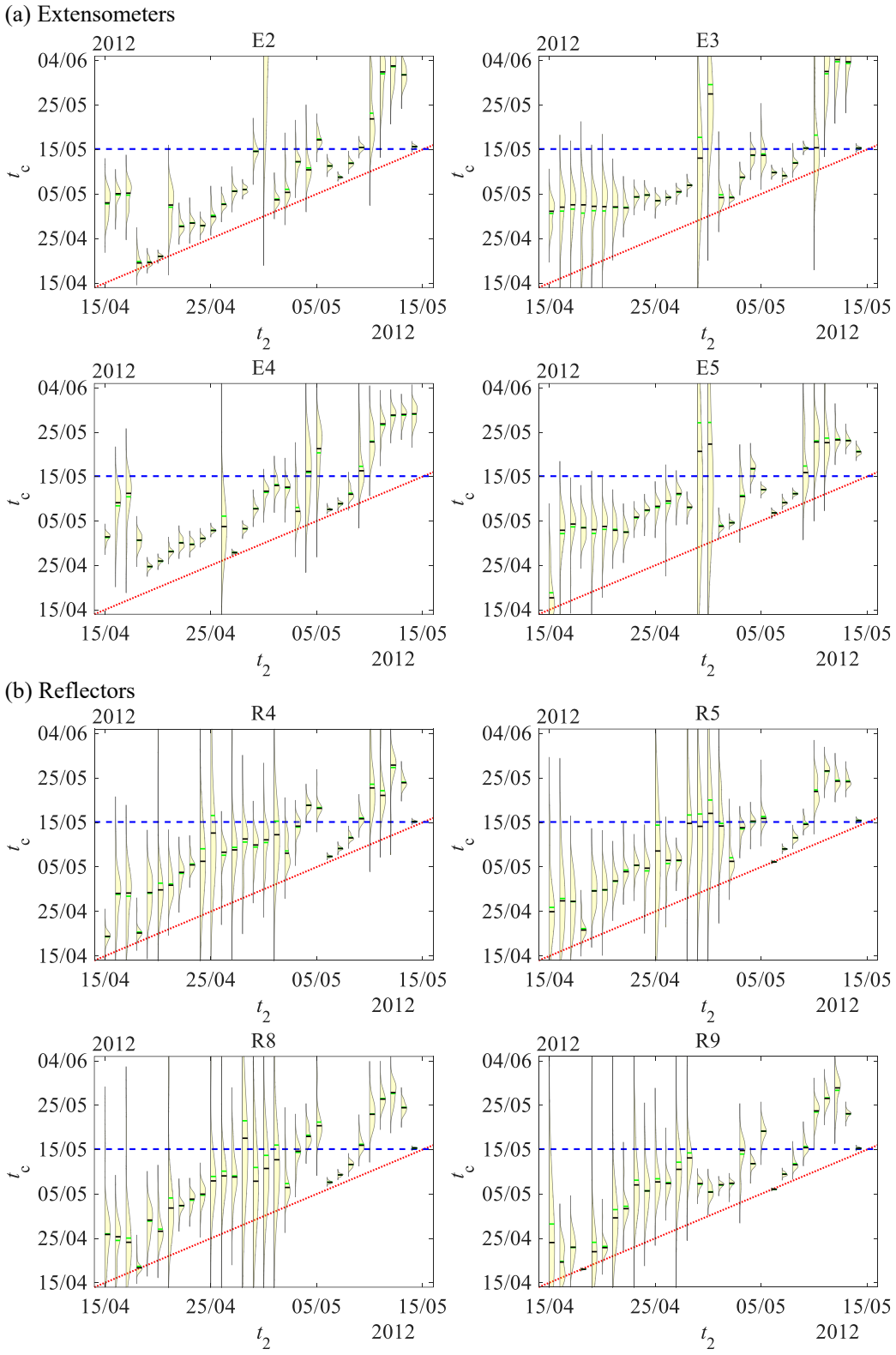

Figure S2. Evolution of the probability density function of anticipated failure time t_c as a function of analysis time t_2, analysed using extensometers E2–E5 and reflectors R4, R5, R8, and R9 at the Preonzo landslide. The green and black lines within each violin indicate the mean and median of the predicted t_c distribution. The red dotted line indicates the reference line $t_c = t_2$, while the blue dashed line marks the observed failure time.

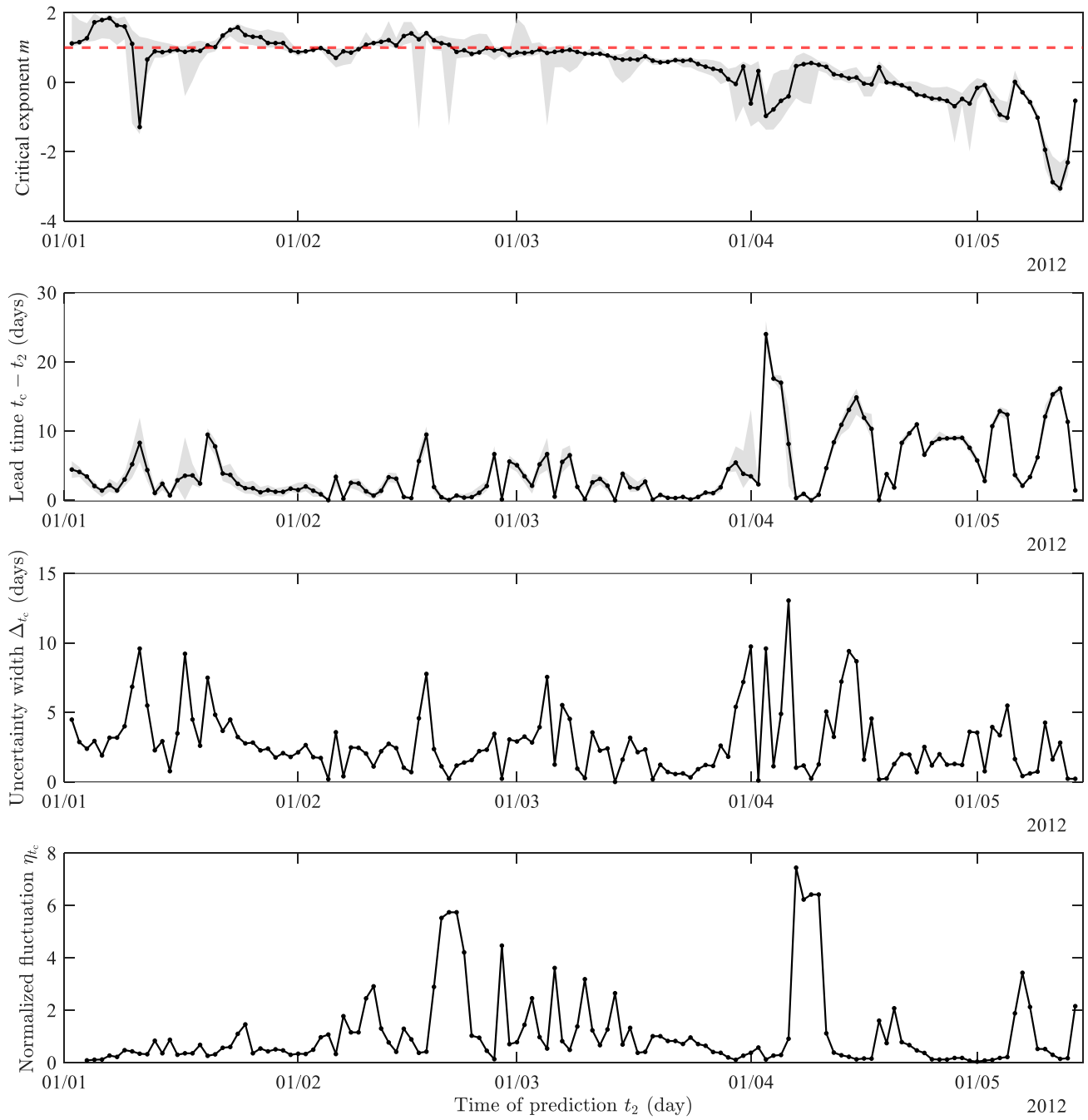

Figure S3. Diagnostics of the LPPLS-based time-to-failure prediction for the Preonzo landslide, including the evolution of the critical exponent, lead time, uncertainty width, and normalized fluctuation (see Text S3 for their definitions).

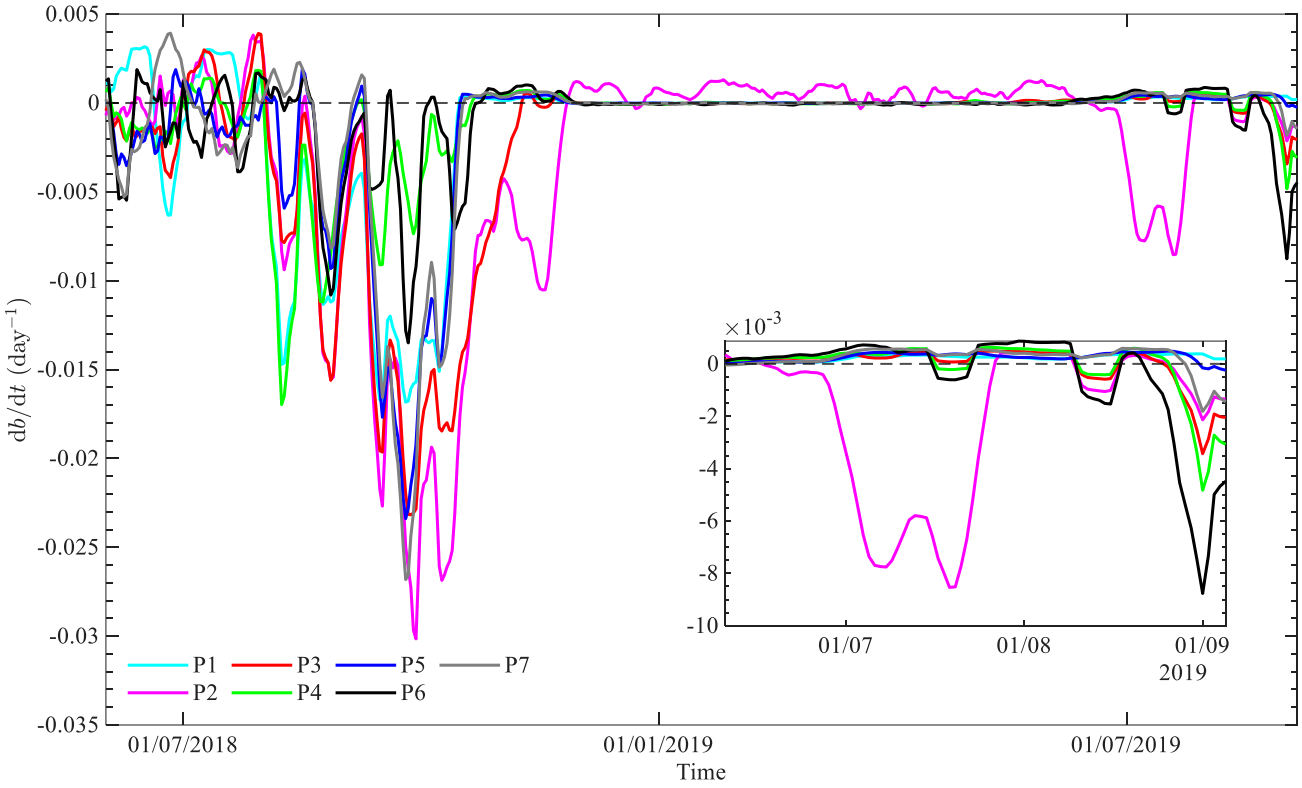

Figure S4. Temporal evolution of the time derivative of velocity *b*-value, derived from seven radar points for the Veslemannen landslide.

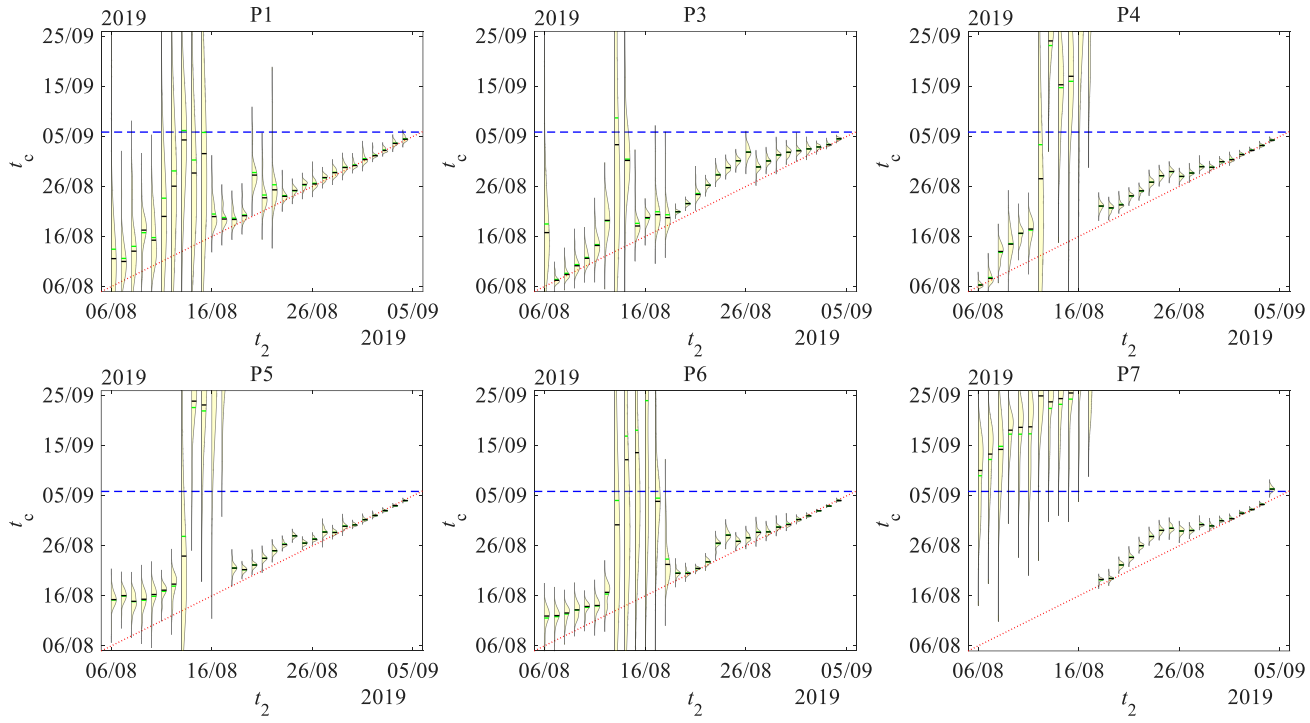

Figure S5. Evolution of the probability density function of anticipated failure time t_c as a function of analysis time t_2, analysed using radar points P1 and P3–P7 at the Veslemannen landslide. The green and black lines within each violin indicate the mean and median of the predicted t_c distribution. The red dotted line indicates the reference line $t_c = t_2$, while the blue dashed line marks the observed failure time.

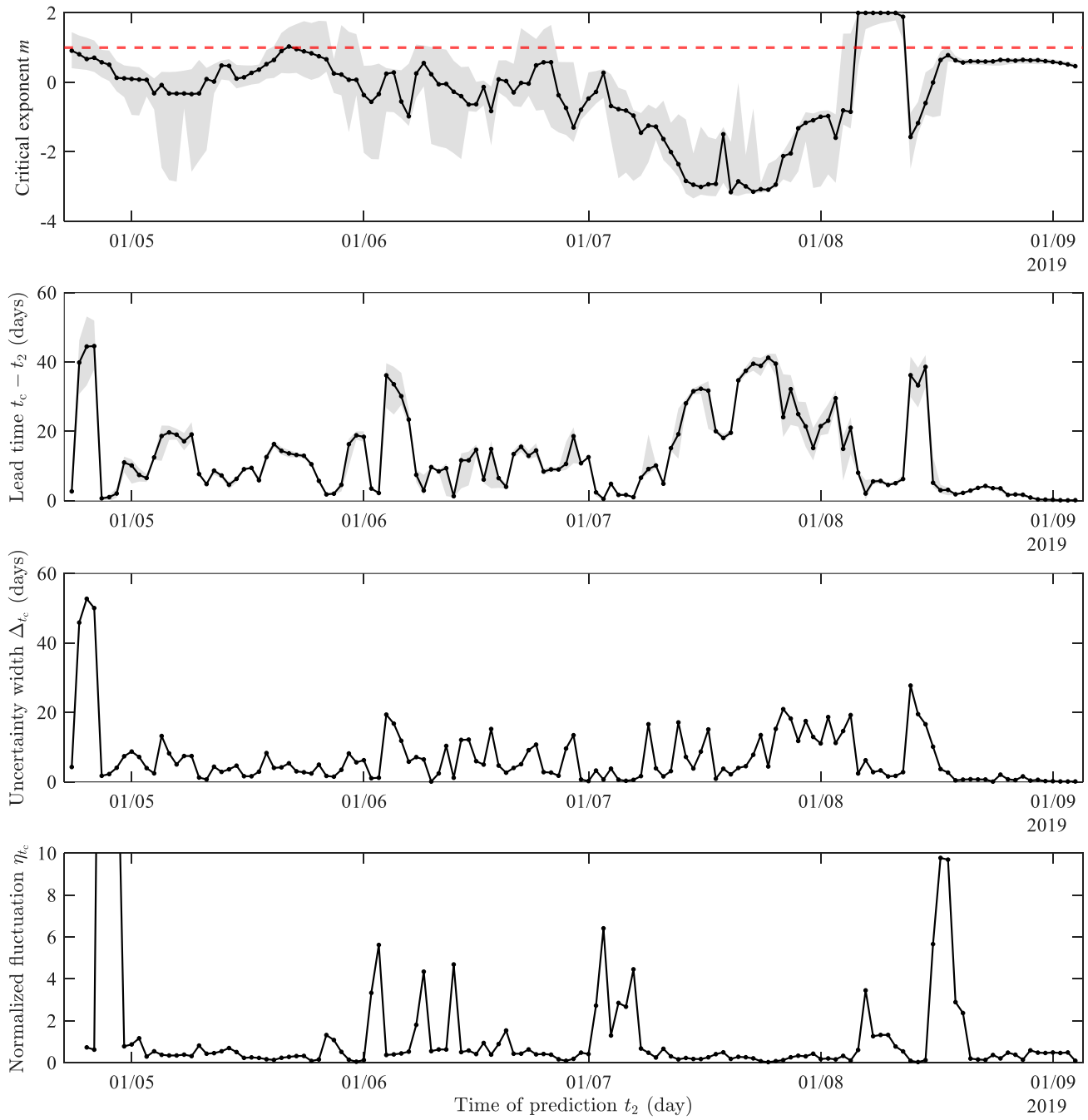

Figure S6. Diagnostics of the LPPLS-based time-to-failure prediction for the Veslemannen landslide, including the evolution of the critical exponent, lead time, uncertainty width, and normalized fluctuation (see Text S3 for their definitions).

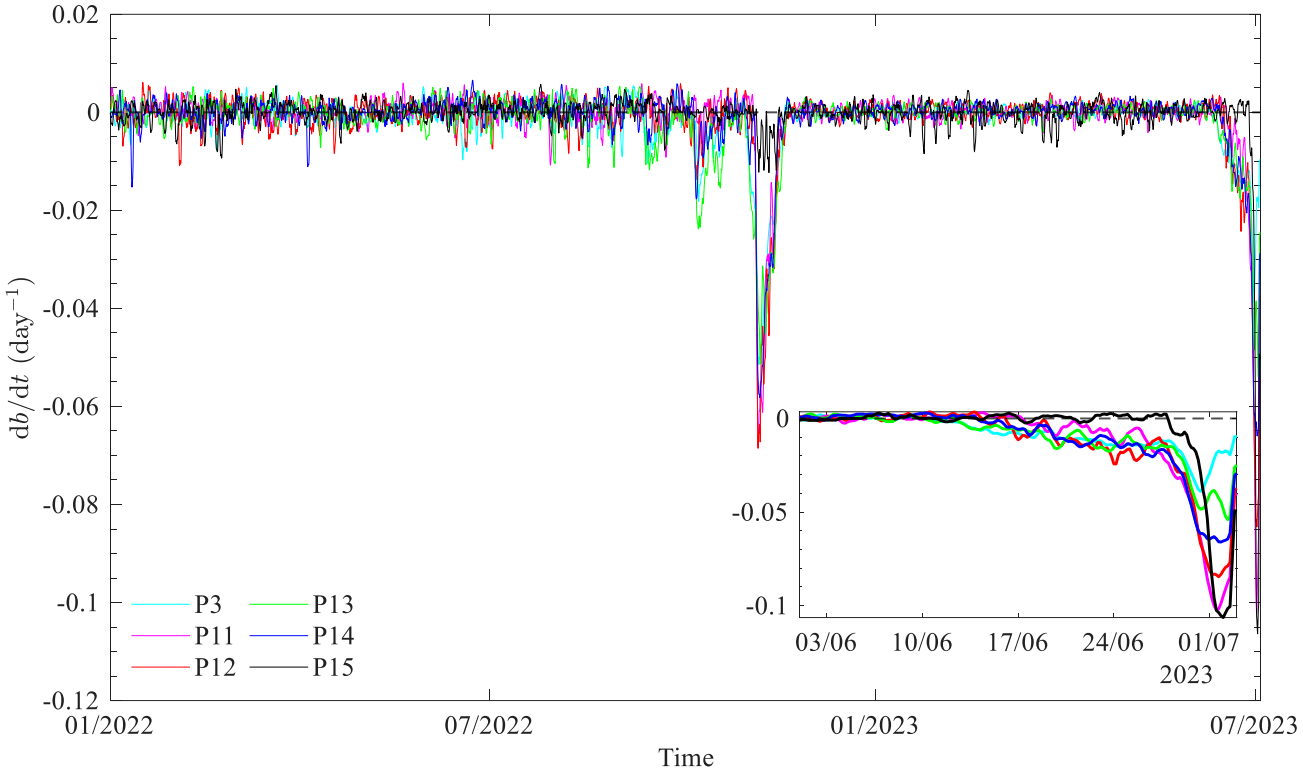

Figure S7. Temporal evolution of the time derivative of velocity *b*-value, derived from seven radar points for the Veslemannen landslide.

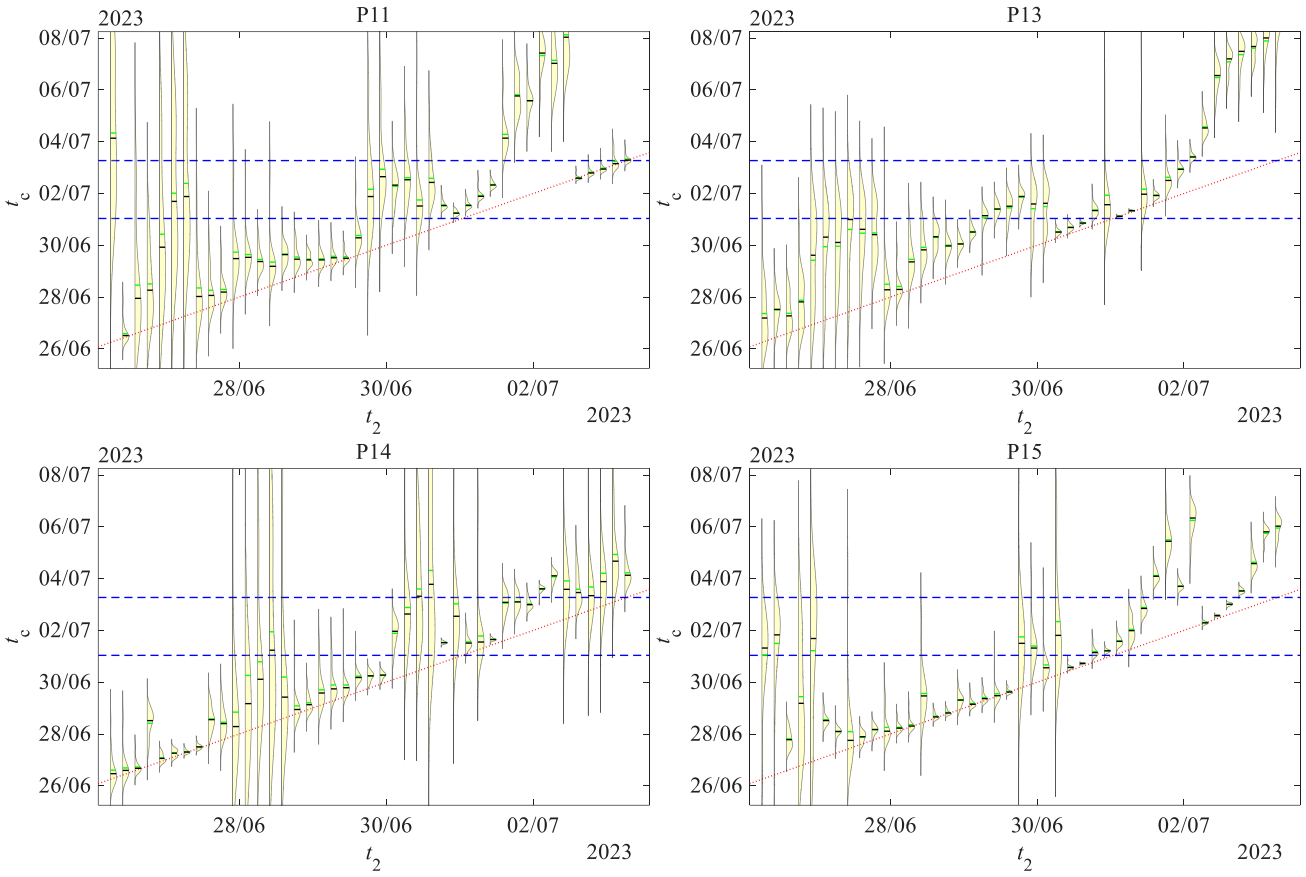

Figure S8. Evolution of the probability density function of anticipated failure time t_c as a function of analysis time t_2, analysed using radar points P11, P13, P14, and P15 at the Stampa landslide. The green and black lines within each violin indicate the mean and median of the predicted t_c distribution. The red dotted line indicates the reference line $t_c = t_2$, while the two blue dashed lines mark the observed two failure events (01 July 2023, 00:54 and 03 July 2023, 06:33).

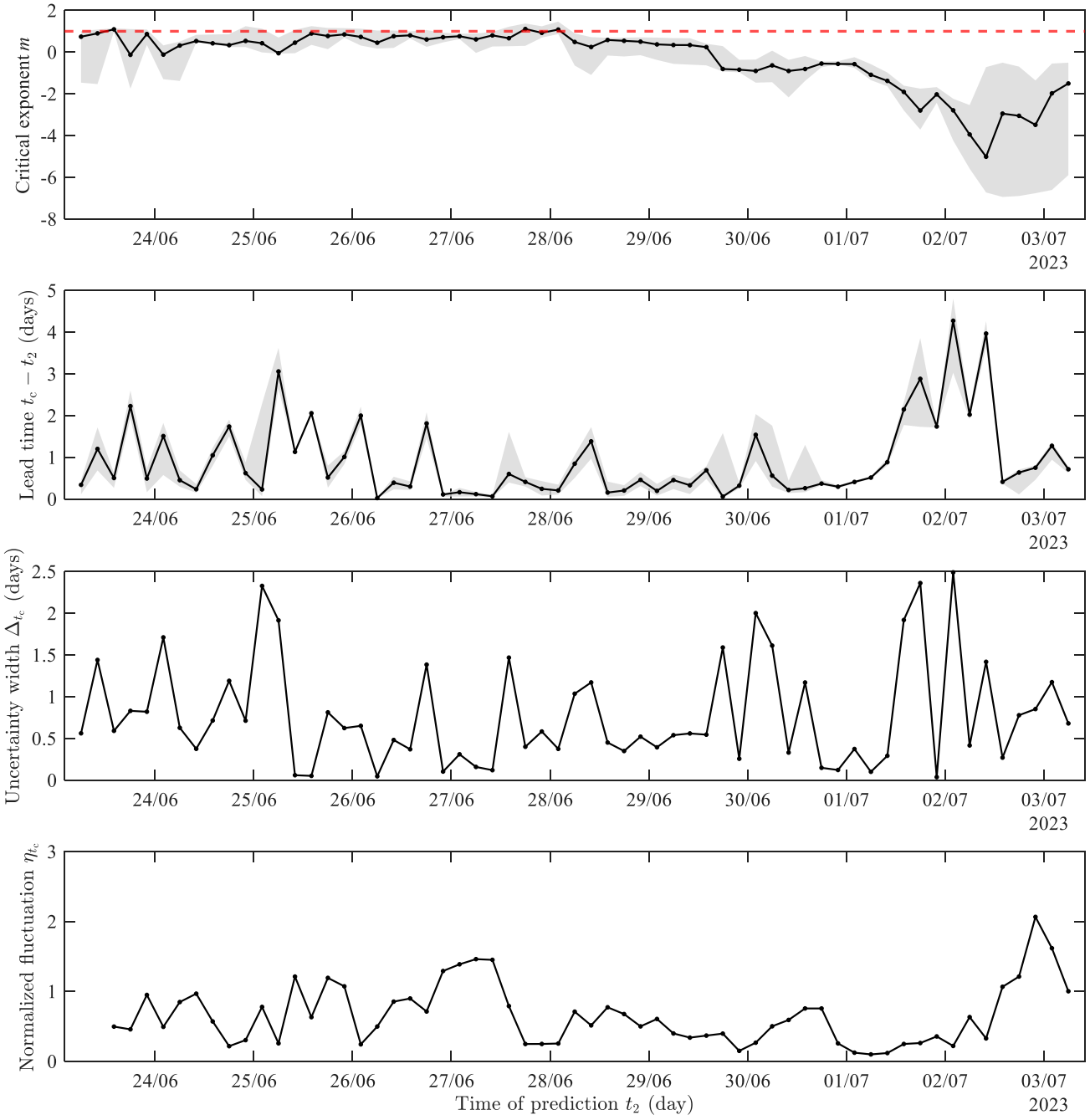

Figure S9. Diagnostics of the LPPLS-based time-to-failure prediction for the Stampa landslide, including the evolution of the critical exponent, lead time, uncertainty width, and normalized fluctuation (see Text S3 for their definitions).

References

Barndorff-Nielsen, O. E. and Cox, D. R.: Inference and Asymptotics, 1st ed., Chapman & Hall/CRC, Boca Raton, 1994.

Filimonov, V. and Sornette, D.: A stable and robust calibration scheme of the log-periodic power law model, Physica A Stat. Mech. Appl., 392, 3698–3707, https://doi.org/10.1016/j.physa.2013.04.012, 2013.

Forbes, C., Evans, M., Hastings, N., and Peacock, B.: Statistical Distributions, 4th ed., John Wiley & Sons, New Jersey, 2011.

Süli, E. and Mayers, D. F.: An Introduction to Numerical Analysis, Cambridge University Press, Cambridge, 2003.