Supporting Information for: Pre- and post-seismic deformation related to the 2015, $\mathcal{M}_w$7.8 Gorkha Earthquake, Nepal

S1. Variational bayesian Independent Component Analysis (vbICA)

Standard ICA techniques are based on a so-called mapping approach (e.g. Hyvärinen and Oja, 1997; Cardoso, 1999). The vbICA algorithm differs in that it uses a modelling approach. It is then necessary to evaluate the posterior pdf of the parameters (or 'weights' $\mathbf{W}$) of the model: $p(\mathbf{W})$. The evaluation of the posterior pdf of the parameters is computationally very costly if performed in a fully bayesian fashion. For this reason, Choudrey (2002) and Choudrey and Roberts (2003) introduced a variational approach to approximate the posterior pdf of the weights. The general concept can be stated as follows. An approximated pdf for the weights is introduced ($p'(\mathbf{W})$), and the problem is transformed into a maximization problem, where the cost function to be maximized is the so-called Negative Free Energy (NFE) of the model. Such a maximization automatically minimizes the Kullback-Liebler (KL) divergence between the true $p(\mathbf{W})$ and the approximated $p'(\mathbf{W})$ posterior pdf of the weights, and it is performed via an Expectation-Maximization algorithm. For this reason, the performance of the variational approximation depends on the particular factorization chosen for $p'(\mathbf{W})$. Here we adopt the same factorization of Choudrey (2002), and then used by Gualandi et al. (2015), where the original updating equations have been modified in order to take into account missing data, as proposed by Chan et al. (2003).

Following Choudrey (2002) and Gualandi et al. (2015), we first run a vbICA using a mix of four Gaussians to describe each IC, and starting the analysis with loose priors. Then we use a criterion based on an Automatic Relevance Determination (ARD) method to determine the most appropriate number of ICs to be retained to explain the data, and, among a pre-selected number of test set, we pick the set of priors that maximizes the NFE (see Gualandi et al., 2015).

S2. Uncertainties propagation

In this Section we use the theory of error propagation and the statistical properties of the covariance between random variables (rvs) to determine the final uncertainties on the quantities of interest such as displacement at the surface (Section S2.1), afterslip on the fault (Section S2.2), and moment released by afterslip (Section S2.3). Throughout the entire Section we perform an analysis only of the random uncertainties. Eventual systematic uncertainties are not quantified. We represent every rv through its mean and standard deviation. In showing the final value of a rv x we will write it as $x_{best} \pm \sigma_x$, where x_{best} is the expected value of the rv x and σ_x is its standard deviation. In the main text, when we refer to the 1σ uncertainties we imply one standard deviation. Following Taylor (1982), we present the results using the same number of decimal position as the associated uncertainty, and this is rounded to one significant figure. The only exception is in cases where the first significant figure of the uncertainty is a 1 or 2: in such a case we retain two significant figures for the uncertainty (for more details, see Taylor, 1982).

The aforementioned quantities for which we want to determine the final random uncertainty are further rvs, and in general they are the result of some function of several rvs: $z = f(x_1, \ldots, x_N)$. The calculation of the best value of the rv z obtained as a function of N rvs is performed using all the available digits, without considering the truncation due to the uncertainty estimation. The calculation of the standard deviation of the variable z is in general a complicated task, but we can resort to an approximation. The delta method approximates the variance of a function of a rv using the first-order approximation of a Taylor series. In the multivariate case it is still possible to

apply the delta method approximation, taking into account also the covariances between the rvs. For $N = 2$ the equation can be written as (see also formula (9.9) of Taylor, 1982):

$$var(z) \approx \left(\frac{\partial f}{\partial x_1}\Big|_{E[x_1]}\right)^2 var(x_1) + \left(\frac{\partial f}{\partial x_2}\Big|_{E[x_2]}\right)^2 var(x_2) + 2\frac{\partial f}{\partial x_1}\Big|_{E[x_1]}\frac{\partial f}{\partial x_2}\Big|_{E[x_2]} cov(x_1, x_2) \quad \text{(S1)}$$

where the notation $\frac{\partial f}{\partial x_*}\Big|_{E[x_*]}$ indicates the partial derivative of f w.r.t to x_* calculated in $E[x_*]$.

In general, we have:

$$var(z) \approx \sum_{j=1}^{N}\sum_{k=1}^{N} \frac{\partial f}{\partial x_j}\Big|_{E[x_j]}\frac{\partial f}{\partial x_k}\Big|_{E[x_k]} cov(x_j, x_k) \quad \text{(S2)}$$

In an even more general situation, we may want to calculate what is the covariance between the rvs obtained using two different functions, which depend on two different set of rvs. Let us imagine to have two scalar functions, f and g, whose inputs are a vector of rvs $\mathbf{x} = [x_1, \ldots, x_{N_f}]^T$ and a vector of rvs $\mathbf{y} = [y_1, \ldots, y_{N_g}]^T$, respectively. Their output will be the two scalar rvs $z_f = f(\mathbf{x})$ and $z_g = g(\mathbf{y})$. We can deduce the covariance between $z_{f(\mathbf{x})}$ and $z_{g(\mathbf{y})}$ using the same approximation method based on the Taylor expansions. The result is:

$$cov(z_{f(\mathbf{x})}, z_{g(\mathbf{y})}) \approx \sum_{j=1}^{N_f}\sum_{k=1}^{N_g} \frac{\partial f}{\partial x_j}\Big|_{E[x_j]}\frac{\partial g}{\partial y_k}\Big|_{E[y_k]} cov(x_j, y_k) = (\nabla_{\mathbf{x}} f\Big|_{E[\mathbf{x}]})^T C_{\mathbf{x},\mathbf{y}} \nabla_{\mathbf{y}} g\Big|_{E[\mathbf{y}]} \quad \text{(S3)}$$

When we have the possibility to calculate the exact value of the variance of the rv z we will use the exact value instead of the approximated one.

S2.1. Interpolated and reconstructed displacements

An advantage of the vbICA method w.r.t. the classical PCA and ICA analysis consists in having the full pdfs of both the mixing matrix and the sources. In our case, the mixing matrix corresponds to the spatial distributions U, and the sources to the temporal functions V. Using the same notation as in Gualandi et al. (2015), we introduce also a weighting diagonal matrix S, that is composed just by scalar errorless quantities. Here we calculate the random uncertainty related to the displacement at every single station (for all the three directions, East, North, and Vertical, independently) and at every single epoch. The total number of time series is M (equal to 3 times the number of stations), and the total number of epochs is T. We use the indexes j and t that run from 1 to M and from 1 to T, respectively, to indicate the generic time series or epoch. We consider the rvs u_{ji} and v_{it}, i.e. the elements of the vector representing a given IC spatial distribution or temporal function, to be normally distributed around their best value, and to be independent and identically distributed (iid) to the other elements of the same IC. The assumption that the rv v_{it} is normally distributed is not in conflict with the fact that the ICs are not normally distributed. Indeed, the $\{v_{it}\}_{t=1}^{T}$ constitutes a sample of the rv $\mathbf{v}_i$, and this rv is following a distribution different from the normal one. Our assumption reflects the idea that the nominal uncertainty in the position calculated at a certain epoch and a given site do not depend on the uncertainties calculated at a different time and the same site or at a different site and the same time. Furthermore, we neglect possible correlations between the position at a given site along different directions (East, North,

and Vertical). In other words, we are neglecting the correlations between the data. This is, in general, not true, since we know that coloured noise is present in daily position GPS time series (e.g. Mao et al., 1999; Williams, 2003), and correlations exist between the position along the three directions at a specified site, but it is a good first approximation.

Using the equation 1 to reconstruct the observed data X, we can write the element x_{jt} of the reconstructed surface displacement matrix as:

$$x_{jt}^{ICA} = \sum_{i=1}^{R} x_{jt}^{IC_i} = \sum_{i=1}^{R} x_{jti} = \sum_{i=1}^{R} u_{ji} s_{ii} v_{it} \tag{S4}$$

where $j = 1, \ldots, M$ is the number of time series, $t = 1, \ldots, T$ is the number of epochs, and R is the number of components used to approximate the original data matrix.

Because of the independence between the rvs $\{x_{jti}\}_{i=1}^{R}$, the variance of x_{jt}^{ICA} is given by:

$$var(x_{jt}^{ICA}) = var\Big(\sum_{i=1}^{R} u_{ji} s_{ii} v_{it}\Big) = \sum_{i=1}^{R} var(u_{ji} s_{ii} v_{it}) \tag{S5}$$

Thus, we have to calculate the variance of the contribution of every single IC. For the i-th IC we have:

$$var(x_{jti}) = var(u_{ji} s_{ii} v_{it}) = s_{ii}^2 \Big[var(u_{ji}) var(v_{it}) + var(u_{ji}) E[v_{jt}]^2 + E[u_{ji}]^2 var(v_{jt})\Big] \tag{S6}$$

The expected value and the variance for each component are a direct result of the vbICA analysis, and is obtained from the dispersion of the actual data. We can thus calculate the variance of the reconstructed displacement. The corresponding uncertainty will be the square root of the variance, i.e. the standard deviation. In the current version of the vbICA algorithm we are taking into account only the actual data, weighting 0 the missing ones. All the actual data are weighted equally, i.e. we are not considering the differences between the uncertainties on two different points belonging to the dataset.

Equation S6 is an example where the exact variance can be calculated. Indeed, the approximated variance calculated via equation S2 would neglect the contribution of the product of the variances. This approximation is reasonable if the variances are small compared to the expected values of the multiplied rvs.

In order to take into account the possible errors introduced in correcting the time series of the post-seismic phase we have to assess the uncertainty related to the quantity that we subtract from the original time series. It includes the co-seismic offsets, the seasonal, and the linear signals. The estimated displacement signal that we subtract from the raw time series is thus:

$$\begin{aligned} x_{jt}^{est} = x_{jt}^{est\ co} + x_{jt}^{est\ seasonal} + x_{jt}^{est\ linear} = \sum_{i=1}^{N_{co}} x_{jt}^{est\ co_i} + \sum_{i=1}^{\hat{R}} x_{jt}^{IC_i^{seasonal}} + x_{jt}^{est\ linear} = \\ = \sum_{i=1}^{N_{co}} c_{co_i} H(t - t_{co_i}) + \sum_{i=1}^{\hat{R}} u_{ji}^{krig} s_{ii} v_{it}^{stack} + u_j^{linear} s^{linear} v_t^{linear} \end{aligned} \tag{S7}$$

where N_{co} is the number of corrected co-seismic offsets, c_{co_i} and t_{co_i} are the offset relative to the i-th seismic event estimated from the data (see Section 2) and the relative epoch, H is the Heaviside function. In our case $\hat{R} = 3$ because the analysis on the pre-seismic data indicates 3 as the number of ICs to retain (see Section 2), and $s^{linear} = 1\ yr$ since we have interpolated the velocity field u^{linear} that is expressed in mm/yr while v^{linear} is non-dimensional.

We make the same assumption made for the estimation of the variance on the displacement reconstructed by the ICA: we consider every single element u^*_{ji} and v^*_{it} representing a given IC spatial distribution or temporal function as iid to the other elements of the same component. It follows that the variance of x^{est}_{jt}, that is the variance of the sum of independent rvs, can be obtained as the sum of the variances of every single component:

$$\begin{aligned} var(x^{est}_{jt}) &= var\Big(\sum_{i=1}^{N_{co}} x^{est\ co_i}_{jt} + \sum_{i=1}^{\hat{R}} x^{IC_i^{seasonal}}_{jt} + x^{est\ linear}_{jt}\Big) = \\ &= var\Big(\sum_{i=1}^{N_{co}} x^{est\ co_i}_{jt}\Big) + var\Big(\sum_{i=1}^{\hat{R}} x^{IC_i^{seasonal}}_{jt}\Big) + var(x^{est\ linear}_{jt}) = \\ &= \sum_{i=1}^{\hat{R}} var(x^{IC_i^{seasonal}}_{jt}) + var(x^{est\ linear}_{jt}) \end{aligned} \tag{S8}$$

where we have neglected the uncertainty associated to the determination of the co-seismic offsets in the calculation of the final uncertainty.

Thus we have to calculate the variance of every single seasonal component:

$$\begin{aligned} var(x^{IC_i^{seasonal}}_{jt}) &= var(u^{krig}_{ji} s_{ii} v^{stack}_{it}) = s^2_{ii} var(u^{krig}_{ji} v^{stack}_{it}) = \\ &= s^2_{ii}\Big[var(u^{krig}_{ji})var(v^{stack}_{it}) + var(u^{krig}_{ji})E[v^{stack}_{it}]^2 + E[u^{krig}_{ji}]^2 var(v^{stack}_{it})\Big] \end{aligned} \tag{S9}$$

and of the linear term:

$$var(x^{est\ linear}_{jt}) = \Big(s^{linear} E[v^{linear}_t]\Big)^2 var(u^{linear}_j) \tag{S10}$$

where we have assumed an errorless linear time function, i.e. $var(v^{linear}_t) = 0$.

From the kriging we have all the variances of the spatial distributions u^{krig}_{ji} and u^{linear}_j. From the stacking interpolation of the ICs we have the sample variance of the quantities v^{stack}_{it}. Thus, we have all the quantities appearing in the right hand side of equations S9 and S10. The final uncertainty on the interpolated displacement is calculated as the square root of equation S8. Also in this case we have used the exact expression for the variance of x^{est}_{jt} instead of its approximated value.

S2.2. Slip model

Let us now calculate the uncertainty on the slip model. In the following we will use the symbol $\widetilde{\delta}_{as\ pt}$ to indicate the afterslip on the patch p at time t as deduced from the inversion of the data. In

general, it is possible to substitute $\widetilde{\delta}_{as\ pt}$ with any other slip, $\widetilde{\delta}_{pt}$, deduced on the fault (co-seismic as well as transient slip). From the inversion method adopted we have calculated both the a posteriori model vector and its corresponding covariance matrix. The model vector for the slip deduced from the i-th IC is constructed as follows:

$$\mathbf{m}_i = \begin{bmatrix} m_{1i} \\ m_{2i} \\ \vdots \\ m_{Pi} \\ m_{P+1,i} \\ m_{P+2,i} \\ \vdots \\ m_{2P,i} \end{bmatrix} = \begin{bmatrix} l_{1i}^{strike} \\ l_{2i}^{strike} \\ \vdots \\ l_{Pi}^{strike} \\ l_{1i}^{dip} \\ l_{2i}^{dip} \\ \vdots \\ l_{Pi}^{dip} \end{bmatrix} = \mathbf{l}_i \tag{S11}$$

where $i = 1, \ldots, R$ indicates the ICs, and P is the number of patches on the fault for which we invert the data. We will refer to this vector as to the IC slip vector. We also will refer to the strike and dip components separately, thus l_{pi}^{strike} and l_{pi}^{dip} will be the strike and dip IC slip rvs relative to the patch p and the IC i.

If we want to invert R ICs, then we perform R different inversions. In our case, since $R = 1$, we have to perform only a single inversion. The a posteriori covariance matrix of $\mathbf{m}_i$ (or $\mathbf{l}_i$) is given by:

$$C_{\mathbf{m}_i} = \begin{bmatrix} cov(l_{1i}^{strike}, l_{1i}^{strike}) & \ldots & cov(l_{1i}^{strike}, l_{Pi}^{strike}) & cov(l_{1i}^{strike}, l_{1i}^{dip}) & \ldots & cov(l_{1i}^{strike}, l_{Pi}^{dip}) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ cov(l_{Pi}^{strike}, l_{1i}^{strike}) & \ldots & cov(l_{Pi}^{strike}, l_{Pi}^{strike}) & cov(l_{Pi}^{strike}, l_{1i}^{dip}) & \ldots & cov(l_{Pi}^{strike}, l_{Pi}^{dip}) \\ cov(l_{1i}^{dip}, l_{1i}^{strike}) & \ldots & cov(l_{1i}^{dip}, l_{Pi}^{strike}) & cov(l_{1i}^{dip}, l_{1i}^{dip}) & \ldots & cov(l_{1i}^{dip}, l_{Pi}^{dip}) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ cov(l_{Pi}^{dip}, l_{1i}^{strike}) & \ldots & cov(l_{Pi}^{dip}, l_{Pi}^{strike}) & cov(l_{Pi}^{dip}, l_{1i}^{dip}) & \ldots & cov(l_{Pi}^{dip}, l_{Pi}^{dip}) \end{bmatrix} \tag{S12}$$

This matrix can be seen as composed by four block matrices:

$$C_{\mathbf{m}_i} = \begin{bmatrix} C_{\mathbf{l}_i^{strike}} & C_{\mathbf{l}_i^{strike,\ dip}} \\ C_{\mathbf{l}_i^{strike,\ dip}} & C_{\mathbf{l}_i^{dip}} \end{bmatrix} \tag{S13}$$

where $C_{\mathbf{l}_i^{strike}}$, $C_{\mathbf{l}_i^{dip}}$, and $C_{\mathbf{l}_i^{strike,\ dip}}$ are the covariance matrices between the IC slip components along strike and dip.

We can calculate the slip on the patch p along the strike and dip directions as the sum of the contribution of the different ICs:

$$\begin{array}{rcl} \widetilde{\delta}_{as\ pt}^{strike} & = & \sum_{i=1}^{R} l_{pi}^{strike} s_{ii} v_{it} \\ \widetilde{\delta}_{as\ pt}^{dip} & = & \sum_{i=1}^{R} l_{pi}^{dip} s_{ii} v_{it} \end{array} \tag{S14}$$

and the total slip on the patch p will be:

$$\widetilde{\delta}_{as\ pt} = \sqrt{(\widetilde{\delta}^{strike}_{as\ pt})^2 + (\widetilde{\delta}^{dip}_{as\ pt})^2} = \sqrt{\Big(\sum_{i=1}^{R} l^{strike}_{pi} s_{ii} v_{it}\Big)^2 + \Big(\sum_{i=1}^{R} l^{dip}_{pi} s_{ii} v_{it}\Big)^2} \tag{S15}$$

This means that the rv $\widetilde{\delta}_{as\ pt}$ is a function of $3R$ rvs: $\{l^{strike}_{pi}\}_{i=1}^{R}$, $\{l^{dip}_{pi}\}_{i=1}^{R}$, and $\{v_{it}\}_{i=1}^{R}$, and $\{s_{ii}\}_{i=1}^{R}$ are just constant errorless quantities. We know that the covariance matrix of the IC slip vector is not diagonal, thus we want to take into account the correlations between the patches. With respect to the general equation S3 we have the following simplification: $f = g$. Indeed we want to calculate the covariance between the slip on two different patches, and the slip is calculated in the same exact way on both of them. Consequently, the set of rvs $\mathbf{x}$ and $\mathbf{y}$ will have the same size $N = N_x = N_y$, that in the general case is $N = 3R$. We can thus approximate the covariance between the slip on two different patches q and r as:

$$\begin{aligned} cov(\widetilde{\delta}_{as\ qt}, \widetilde{\delta}_{as\ rt}) \approx \sum_{j=1}^{3R}\sum_{k=1}^{3R} \frac{\partial \widetilde{\delta}_{as\ qt}}{\partial x_{qt\ j}}\Big|_{E[x_{qt\ j}]} \frac{\partial \widetilde{\delta}_{as\ rt}}{\partial y_{rt\ k}}\Big|_{E[y_{rt\ k}]} cov(x_{qt\ j}, y_{rt\ k}) = \\ = (\nabla_{\mathbf{x}_{qt}} \widetilde{\delta}_{as\ qt}\Big|_{E[\mathbf{x}_{qt}]})^T C_{\mathbf{x}_{qt},\mathbf{y}_{rt}} \nabla_{\mathbf{y}_{rt}} \widetilde{\delta}_{as\ rt}\Big|_{E[\mathbf{y}_{rt}]} \end{aligned} \tag{S16}$$

where the vectors $\mathbf{x}_{qt}$ and $\mathbf{y}_{rt}$ are given by:
$\mathbf{x}_{qt} = [l^{strike}_{q1}, \ldots, l^{strike}_{qR}, l^{dip}_{q1}, \ldots, l^{dip}_{qR}, v_{1t}, \ldots, v_{Rt}]^T$ and
$\mathbf{y}_{rt} = [l^{strike}_{r1}, \ldots, l^{strike}_{rR}, l^{dip}_{r1}, \ldots, l^{dip}_{rR}, v_{1t}, \ldots, v_{Rt}]^T$, and the subscripts indicate that these vectors depend on a given patch (q or r) and on a given epoch (t). This means that the covariance matrix $C_{\mathbf{x}_{qt},\mathbf{y}_{rt}}$ between $\mathbf{x}_{qt}$ and $\mathbf{y}_{rt}$ is a $3R \times 3R$ matrix given by:

$$C_{\mathbf{x}_{qt},\mathbf{y}_{rt}} = \begin{bmatrix} cov(l^{strike}_{q1},l^{strike}_{r1}) & \ldots & cov(l^{strike}_{q1},l^{strike}_{rR}) & cov(l^{strike}_{q1},l^{dip}_{r1}) & \ldots & cov(l^{strike}_{q1},l^{dip}_{rR}) & cov(l^{strike}_{q1},v_{1t}) & \ldots & cov(l^{strike}_{q1},v_{Rt}) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ cov(l^{strike}_{qR},l^{strike}_{r1}) & \ldots & cov(l^{strike}_{qR},l^{strike}_{rR}) & cov(l^{strike}_{qR},l^{dip}_{r1}) & \ldots & cov(l^{strike}_{qR},l^{dip}_{rR}) & cov(l^{strike}_{qR},v_{1t}) & \ldots & cov(l^{strike}_{qR},v_{Rt}) \\ cov(l^{dip}_{q1},l^{strike}_{r1}) & \ldots & cov(l^{dip}_{q1},l^{strike}_{rR}) & cov(l^{dip}_{q1},l^{dip}_{r1}) & \ldots & cov(l^{dip}_{q1},l^{dip}_{rR}) & cov(l^{dip}_{q1},v_{1t}) & \ldots & cov(l^{dip}_{q1},v_{Rt}) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ cov(l^{dip}_{qR},l^{strike}_{r1}) & \ldots & cov(l^{dip}_{qR},l^{strike}_{rR}) & cov(l^{dip}_{qR},l^{dip}_{r1}) & \ldots & cov(l^{dip}_{qR},l^{dip}_{rR}) & cov(l^{dip}_{qR},v_{1t}) & \ldots & cov(l^{dip}_{qR},v_{Rt}) \\ cov(v_{1t},l^{strike}_{r1}) & \ldots & cov(v_{1t},l^{strike}_{rR}) & cov(v_{1t},l^{dip}_{r1}) & \ldots & cov(v_{1t},l^{dip}_{rR}) & cov(v_{1t},v_{1t}) & \ldots & cov(v_{1t},v_{Rt}) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ cov(v_{Rt},l^{strike}_{r1}) & \ldots & cov(v_{Rt},l^{strike}_{rR}) & cov(v_{Rt},l^{dip}_{r1}) & \ldots & cov(v_{Rt},l^{dip}_{rR}) & cov(v_{Rt},v_{1t}) & \ldots & cov(v_{Rt},v_{Rt}) \end{bmatrix} \tag{S17}$$

such that $C_{\mathbf{x}_{qt},\mathbf{y}_{rt}} = C^T_{\mathbf{y}_{rt},\mathbf{x}_{qt}}$. Several elements of this matrix are zero because of the assumptions of independence between rvs. Indeed, $cov(l^*_{qj}, l^*_{rk}) = 0$ for every $j \neq k$, i.e. for every set of different IC slip. Moreover, we assume independence between the spatial and temporal vectors: $cov(l^*_{pi}, v_{it}) = 0\ \forall\ i, p, t$. Finally, also the covariance between the temporal functions of two different ICs is zero: $cov(v_{jt}, v_{kt}) = 0$ for every $j \neq k$. Having performed the inversion of just one IC, we have that this matrix is the simplest possible in our case:

$$C_{\mathbf{x}_{qt},\mathbf{y}_{rt}} = \begin{bmatrix} cov(l_{q1}^{strike}, l_{r1}^{strike}) & cov(l_{q1}^{strike}, l_{r1}^{dip}) & cov(l_{q1}^{strike}, v_{1t}) \\ cov(l_{q1}^{dip}, l_{r1}^{strike}) & cov(l_{q1}^{dip}, l_{r1}^{dip}) & cov(l_{q1}^{dip}, v_{1t}) \\ cov(v_{1t}, l_{r1}^{strike}) & cov(v_{1t}, l_{r1}^{dip}) & cov(v_{1t}, v_{1t}) \end{bmatrix} =$$

$$= \begin{bmatrix} cov(l_{q1}^{strike}, l_{r1}^{strike}) & cov(l_{q1}^{strike}, l_{r1}^{dip}) & 0 \\ cov(l_{q1}^{dip}, l_{r1}^{strike}) & cov(l_{q1}^{dip}, l_{r1}^{dip}) & 0 \\ 0 & 0 & var(v_{1t}) \end{bmatrix} \tag{S18}$$

For every epoch t we have $P \times P$ of these matrices, for a total of $T \times P \times P$. From each of these matrices we deduce the covariance between the slip on patch q at time t and the slip on patch r at time t (see equation S16). As we already said, we are neglecting temporal correlations. This means that we end up with a $P \times P$ slip covariance matrix $C_{\widetilde{\boldsymbol{\delta}}_{as\ t}}$ at time t, and we have T of these matrices.

When we refer to the uncertainty on the slip (see Section 4 and Figure 10) we are using the standard deviation, i.e. the square root of the variance associated to the rv describing the slip: $\sigma_{\widetilde{\delta}_{as\ pt}} = \sqrt{var(\widetilde{\delta}_{as\ pt})}$. This means that when we show these results we are ignoring the covariance with the other patches. Anyway, we have calculated these covariances, and we will see in the next Section that they play a role in the determination of the final uncertainty on the moment released by the slip distribution.

For the sake of clarity, let us write down explicitly the values of $\nabla\widetilde{\delta}_{as\ pt}$.

$$\nabla\widetilde{\delta}_{as\ pt} = \begin{bmatrix} \frac{\partial\widetilde{\delta}_{as\ pt}}{\partial l_{p1}^{strike}} \\ \vdots \\ \frac{\partial\widetilde{\delta}_{as\ pt}}{\partial l_{pR}^{strike}} \\ \frac{\partial\widetilde{\delta}_{as\ pt}}{\partial l_{p1}^{dip}} \\ \vdots \\ \frac{\partial\widetilde{\delta}_{as\ pt}}{\partial l_{pR}^{dip}} \\ \frac{\partial\widetilde{\delta}_{as\ pt}}{\partial v_{1t}} \\ \vdots \\ \frac{\partial\widetilde{\delta}_{as\ pt}}{\partial v_{Rt}} \end{bmatrix} = \begin{bmatrix} s_{11} v_{1T} \frac{\widetilde{\delta}_{as\ pt}^{strike}}{\widetilde{\delta}_{as\ pt}} \\ \vdots \\ s_{RR} v_{RT} \frac{\widetilde{\delta}_{as\ pt}^{strike}}{\widetilde{\delta}_{as\ pt}} \\ s_{11} v_{1T} \frac{\widetilde{\delta}_{as\ pt}^{dip}}{\widetilde{\delta}_{as\ pt}} \\ \vdots \\ s_{RR} v_{RT} \frac{\widetilde{\delta}_{as\ pt}^{dip}}{\widetilde{\delta}_{as\ pt}} \\ s_{11} \frac{l_{pt}^{strike}\widetilde{\delta}_{as\ pt}^{strike} + l_{pt}^{dip}\widetilde{\delta}_{as\ pt}^{dip}}{\widetilde{\delta}_{as\ pt}} \\ \vdots \\ s_{RR} \frac{l_{pt}^{strike}\widetilde{\delta}_{as\ pt}^{strike} + l_{pt}^{dip}\widetilde{\delta}_{as\ pt}^{dip}}{\widetilde{\delta}_{as\ pt}} \end{bmatrix} \tag{S19}$$

We have all the quantities needed to calculate the elements of the covariance matrix of the slip at time t (equation S16).

S2.3. Moment M_0 and moment magnitude $\mathcal{M}_w$

We now want to calculate the variance of the moment M_0. In particular, we calculate three different moments: 1) $\widetilde{M}_{0\ t}$, 2) $M_{0\ t}$, and 3) $M_{0\ t\to\infty}$ defined as:

$$\widetilde{M}_{0\ t} = \sum_{p=1}^{P} \widetilde{M}_{0\ pt} = \sum_{p=1}^{P} \mu_p A_p \widetilde{\delta}_{as\ pt} \tag{S20}$$

$$M_{0\ t} = \sum_{p=1}^{P} M_{0\ pt} = \sum_{p=1}^{P} \mu_p A_p \delta_{as\ pt} \tag{S21}$$

$$M_{0\ t\to\infty} = \sum_{p=1}^{P} M_{0\ pt\to\infty} = \sum_{p=1}^{P} \mu_p A_p \delta_{as\ pt\to\infty} \tag{S22}$$

where $\delta_{as\ pt}$ is the afterslip predicted by the spring-slider model and is obtained combining equations A1 and A2:

$$\delta_{as\ pt} = V_{0p} t_{rp} \ln\Big[1 + q_p\Big(\exp(t/t_{rp}) - 1\Big)\Big] - V_{0p} t \tag{S23}$$

Finally, $\delta_{as\ pt\to\infty}$ is the asymptotic value of $\delta_{as\ pt}$ for $t \to \infty$:

$$\delta_{as\ pt\to\infty} = V_{0p} t_{rp} \ln q_p \tag{S24}$$

In words, the difference between these three moments is the following. 1) $\widetilde{M}_{0\ t}$ corresponds to the moment calculated using the slip deduced from the inversion of the data, and corresponds to the moment released by afterslip between the first and last epoch available in the data set. 2) $M_{0\ t}$ is the afterslip moment produced combining P non interacting spring-sliders (equation S23). 3) $M_{0\ t\to\infty}$ is the asymptotic value of $M_{0\ t}$.

Once the moment and its variance are calculated, then it is possible to obtain the equivalent moment magnitude and an approximation of the corresponding variance. Let us take for example $M_{0\ t}$, but the same argumentation is valid also for $\widetilde{M}_{0\ t}$ and $M_{0\ t\to\infty}$. The moment magnitude is a rv obtained as a function of another rv, the moment (see equation B2). This means that equation S2 is in its simplest form, and the variance of the moment magnitude associated to $M_{0\ t}$ is approximated by:

$$var(\mathcal{M}_{w\ t}) \approx \Big(\frac{\partial \mathcal{M}_{w\ t}}{\partial M_{0\ t}}\Big)^2 var(M_{0\ t}) = \Big(\frac{2}{3\ln 10}\frac{1}{M_{0\ t}}\Big)^2 var(M_{0\ t}) \tag{S25}$$

S2.3.1. Variance of $\widetilde{M}_{0\ t}$

Let us start with the calculation of the variance of the quantity $\widetilde{M}_{0\ t}$ defined by equation S20. In general, the P rvs $\{\widetilde{M}_{0\ pt}\}_{p=1}^{P}$ are correlated. Indeed, as mentioned in the main text (see Section 3), we assume a non null covariance of the rigidity modulus between two patches belonging to the same layer of the stratified half space. Moreover, the slip on two different patches are correlated, as discussed in previous Section. The variance of the sum of P correlated rvs is:

$$var(\widetilde{M}_{0\ t}) = var\Big(\sum_{p=1}^{P} \widetilde{M}_{0\ pt}\Big) = \sum_{q=1}^{P}\sum_{r=1}^{P} cov(\widetilde{M}_{0\ qt}, \widetilde{M}_{0\ rt}) \tag{S26}$$

Since $\widetilde{M}_{0\ pt}$ is a function of three rvs (μ_p, A_p, and $\widetilde{\delta}_{as\ pt}$), we can apply the approximating equation S3 to determine the element $cov(\widetilde{M}_{0\ qt}, \widetilde{M}_{0\ rt})$, where now the vectors $\mathbf{x}$ and $\mathbf{y}$ are

$\mathbf{x}'_{qt} = [x'_{qt\ 1}, x'_{qt\ 2}, x'_{qt\ 3}]^T = [\mu_q, A_q, \widetilde{\delta}_{as\ qt}]^T$ and $\mathbf{y}'_{rt} = [y'_{rt\ 1}, y'_{rt\ 2}, y'_{rt\ 3}]^T = [\mu_r, A_r, \widetilde{\delta}_{as\ rt}]^T$. The covariance matrix between $\mathbf{x}'_{qt}$ and $\mathbf{y}'_{rt}$ can be written using the block matrices containing the covariances between the different rvs:

$$C_{\mathbf{x}'_{qt},\mathbf{y}'_{rt}} = \begin{bmatrix} C_{\boldsymbol{\mu}} & C_{\boldsymbol{\mu},\mathbf{A}} & C_{\boldsymbol{\mu},\widetilde{\boldsymbol{\delta}}_{as\ t}} \\ C_{\mathbf{A},\boldsymbol{\mu}} & C_{\mathbf{A}} & C_{\mathbf{A},\widetilde{\boldsymbol{\delta}}_{as\ t}} \\ C_{\widetilde{\boldsymbol{\delta}}_{as\ t}} & C_{\widetilde{\boldsymbol{\delta}}_{as\ t},\mathbf{A}} & C_{\widetilde{\boldsymbol{\delta}}_{as\ t}} \end{bmatrix} \tag{S27}$$

This is a $3P \times 3P$ matrix, and once we have specified the values of the block matrices that compose it we can derive the approximated variance of $\widetilde{M_{0\ t}}$.

Let us start from the diagonal blocks of $C_{\mathbf{x}'_{qt},\mathbf{y}'_{rt}}$.

The rigidity modulus $\{\mu_p\}_{i=1}^P$ is assumed to vary with depth. We use the same values adopted by Galetzka et al. (2015), which considered a layered half space. We make the hypothesis that the error on the determination of μ_p is increasing with the depth, so at every layer there is a different uncertainty associated. Moreover, we assume that the uncertainties associated to a given layer correlate between the patches belonging to the same layer. In other words, if the estimation of μ_p at a certain point is incorrect, then also the estimation at all the patches belonging to the same layer is incorrect of the same amount. This means that the covariance matrix of μ_p has the following structure:

$$C_{\boldsymbol{\mu}} = \begin{bmatrix} C_{\boldsymbol{\mu}^{layer\ 1}} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & C_{\boldsymbol{\mu}^{layer\ N}} \end{bmatrix} \tag{S28}$$

where we have assumed N layers. Since we are assuming that the rigidity modulus μ_q at a given layer n correlate with all the μ_r of patches belonging to the same layer, we have that $cov(\mu_q, \mu_r) = var(\mu_q)$ if $q, r \in n$-th layer, and 0 otherwise.

Let us move on considering the second random variable, the area of a given patch A_p. In this study we test only one simple planar geometry approximating the fault. Even if such approximation may be the cause of part of the mismodelling near the rupture (see Section 4) we think that the difference in the area of a more refined geometry and the one here adopted is negligible if compared to other sources of uncertainty (like, for example, the uncertainties estimated for the rigidity modulus, and for the long-term creeping velocity on each patch deduced using the coupling map). Thus, we neglect the uncertainty in the area of a single patch, and set $var(A_p) = 0$. We also neglect the correlations with adjacent patches, since the area of a patch is not influencing the area of the next one. This means that we assume a null covariance matrix for the variable A_p: $C_{\mathbf{A}} = \mathbf{0}$.

The last term of the diagonal of equation S27, $C_{\widetilde{\boldsymbol{\delta}}_{as\ t}}$, has been deduced in the previous Section.

Let us now consider the off-diagonal terms. There are no obvious reasons to consider correlations between the rigidity modulus μ_p and the area of a patch A_p. Instead, it is certainly true that the afterslip determined on a given patch depends on the particular geometry used to approximate the fault. Anyway, because of the assumption on the covariance of the area, $C_{\mathbf{A}} = 0$, we can treat the rv A_p as a constant. Taking the definition of covariance between two rvs we have:

$$cov(\mu_q, A_r) = E[\mu_q A_r] - E[\mu_q]E[A_r] = A_r E[\mu_q] - A_r E[\mu_q] = 0 \tag{S29}$$

$$cov(\widetilde{\delta}_{as\ qt}, A_r) = E[\widetilde{\delta}_{as\ qt} A_r] - E[\widetilde{\delta}_{as\ qt}]E[A_r] = A_r E[\widetilde{\delta}_{as\ qt}] - A_r E[\widetilde{\delta}_{as\ qt}] = 0 \tag{S30}$$

that means that $C_{\boldsymbol{\mu},\mathbf{A}} = C_{\mathbf{A},\boldsymbol{\mu}} = \mathbf{0}$ and $C_{\widetilde{\boldsymbol{\delta}}_{as\ t},\mathbf{A}} = C_{\mathbf{A},\widetilde{\boldsymbol{\delta}}_{as\ t}} = \mathbf{0}$.

Furthermore, the Green's functions G used to determine the slip ICs consider a homogeneous half-space, and depend only on the Poisson's modulus. Assuming a constant Poisson's modulus means that when the rigidity modulus changes than also the first Lamé parameter changes accordingly. It follows that there is a correlation between the two Lamé parameters, but this correlation does not affect the Poisson's modulus. Thus, we further assume that $cov(\mu_q, \widetilde{\delta}_{as\ rt}) = 0\ \forall\ q, r$, i.e. $C_{\boldsymbol{\mu},\widetilde{\boldsymbol{\delta}}_{as\ t}} = C_{\widetilde{\boldsymbol{\delta}}_{as\ t},\boldsymbol{\mu}} = \mathbf{0}$.

These assumptions lead to a very simple structure for the covariance matrix defined in equation S27:

$$C_{\mathbf{x}'_{qt},\mathbf{y}'_{rt}} = \begin{bmatrix} C_{\boldsymbol{\mu}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & C_{\widetilde{\boldsymbol{\delta}}_{as\ t}} \end{bmatrix} \tag{S31}$$

The gradient of the rvs $\widetilde{M}_{0\ qt}$ and $\widetilde{M}_{0\ rt}$ w.r.t. $\mathbf{x}'_{qt}$ and $\mathbf{y}'_{rt}$ is also easy to calculate, since the moment is just the product of the three rvs contained in the vectors. This simple relationship between the rvs allows us to even use an exact expression for the covariance of $\widetilde{M}_{0\ qt}$ and $\widetilde{M}_{0\ rt}$. Indeed, assuming the rvs to be normally distributed, the general equation of the covariance of the product of two rvs is (see equation (13) of Bohrnstedt and Goldberger, 1969):

$$\begin{aligned} cov(x_1x_2, y_1y_2) = E[x_1]E[y_1]cov(x_2, y_2) + E[x_1]E[y_2]cov(x_2, y_1) + \\ + E[x_2]E[y_1]cov(x_1, y_2) + E[x_2]E[y_2]cov(x_1, y_1) + \\ + cov(x_1, y_1)cov(x_2, y_2) + cov(x_1, y_2)cov(x_2, y_1) \end{aligned} \tag{S32}$$

We can develop this formula for the product of three rvs, that is the case we are interested in:

$$\begin{aligned} cov(x_1x_2x_3, y_1y_2y_3) = E[x_1x_2]E[y_1y_2]cov(x_3, y_3) + E[x_1x_2]E[y_3]cov(x_3, y_1y_2) + \\ + E[x_3]E[y_1y_2]cov(x_1x_2, y_3) + E[x_3]E[y_3]cov(x_1x_2, y_1y_2) + \\ cov(x_1x_2, y_1y_2)cov(x_3, y_3) + cov(x_1x_2, y_3)cov(x_3, y_1y_2) \end{aligned} \tag{S33}$$

Since all the off-diagonal values of matrix S31 are zero we have that the expectation values of the product of two rvs in equation S33 can be written as the product of the expectation values of the two rvs separately. Moreover, the covariances of crossed rvs are zero: $cov(x_i, y_jy_k) = 0$ for $j, k \neq i$. We can thus rewrite equation S33 as:

$$\begin{aligned} &cov(x_1x_2x_3, y_1y_2y_3) = \\ &= E[x_1]E[x_2]E[y_1]E[y_2]cov(x_3, y_3) + E[x_3]E[y_3]cov(x_1x_2, y_1y_2) + cov(x_1x_2, y_1y_2)cov(x_3, y_3) \end{aligned} \tag{S34}$$

We know from equation S32 what is the value of $cov(x_1x_2, y_1y_2)$. Using the simplifications introduced by the assumptions on the covariances between the rvs (see equation S31), we have:

$$cov(x_1x_2, y_1y_2) = E[x_1]E[y_1]cov(x_2, y_2) + E[x_2]E[y_2]cov(x_1, y_1) + cov(x_1, y_1)cov(x_2, y_2) \tag{S35}$$

and introducing this in equation S34 we get:

$$
\begin{aligned}
cov(x_1x_2x_3, y_1y_2y_3) = \\
&E[x_1]E[y_1]E[x_2]E[y_2]cov(x_3,y_3) + E[x_1]E[y_1]E[x_3]E[y_3]cov(x_2,y_2) + \\
&+ E[x_2]E[y_2]E[x_3]E[y_3]cov(x_1,y_1) + E[x_1]E[y_1]cov(x_2,y_2)cov(x_3,y_3) + \\
&+ E[x_2]E[y_2]cov(x_1,y_1)cov(x_3,y_3) + E[x_3]E[y_3]cov(x_1,y_1)cov(x_2,y_2) + \\
&+ cov(x_1,y_1)cov(x_2,y_2)cov(x_3,y_3)
\end{aligned} \tag{S36}
$$

The approximation given by equation S3 would basically neglect the last three terms of the right hand side. To make it clearer, we substitute the x_j and y_k with the actual name of the rvs used to calculate $\widetilde{M_{0\ t}}$:

$$
\begin{aligned}
cov(\mu_q A_q \widetilde{\delta}_{as\ qt}, \mu_r A_r \widetilde{\delta}_{as\ rt}) = \\
&E[\mu_q]E[\mu_r]E[A_q]E[A_r]cov(\widetilde{\delta}_{as\ qt}, \widetilde{\delta}_{as\ rt}) + E[\mu_q]E[\mu_r]E[\widetilde{\delta}_{as\ qt}]E[\widetilde{\delta}_{as\ rt}]cov(A_q, A_r) + \\
&+ E[A_q]E[A_r]E[\widetilde{\delta}_{as\ qt}]E[\widetilde{\delta}_{as\ rt}]cov(\mu_q, \mu_r) + E[\mu_q]E[\mu_r]cov(A_q, A_r)cov(\widetilde{\delta}_{as\ qt}, \widetilde{\delta}_{as\ rt}) + \\
&+ E[A_q]E[A_r]cov(\mu_q, \mu_r)cov(\widetilde{\delta}_{as\ qt}, \widetilde{\delta}_{as\ rt}) + E[\widetilde{\delta}_{as\ qt}]E[\widetilde{\delta}_{as\ rt}]cov(\mu_q, \mu_r)cov(A_q, A_r) + \\
&+ cov(\mu_q, \mu_r)cov(A_q, A_r)cov(\widetilde{\delta}_{as\ qt}, \widetilde{\delta}_{as\ rt})
\end{aligned} \tag{S37}
$$

Because of the assumption of errorless area, we are setting to zero the fourth, sixth, and seventh terms of the right hand side. This means that the approximation introduced by equation S3 differs from the exact form S36 only because of the term $E[A_q]E[A_r]cov(\mu_q, \mu_r)cov(\widetilde{\delta}_{as\ qt}, \widetilde{\delta}_{as\ rt})$. For all our calculations this difference is negligible. In any case, we have all the quantities to calculate the full covariance matrix of $\widetilde{M_{0\ pt}}$. We can then introduce these values in equation S26 to get the variance of $\widetilde{M_{0\ t}}$. Since we are neglecting possible correlations in time, the variance of $\widetilde{M_{0\ t}}$ is fully expressing the uncertainty on the moment.

S2.3.2. Variance of $M_{0\ t}$

In order to calculate the variance of $M_{0\ t}$ we can proceed in the same exact way we did for the calculation of the variance of $\widetilde{M_{0\ t}}$. The only difference will be in the rv related to the slip. Instead of using $\widetilde{\delta}_{as\ pt}$ we use $\delta_{as\ pt}$ and its covariance matrix:

$$
var(M_{0\ t}) = var\Big(\sum_{p=1}^{P} M_{0\ pt}\Big) = \sum_{q=1}^{P}\sum_{r=1}^{P} cov(M_{0\ qt}, M_{0\ rt}) \tag{S38}
$$

and

$$
\begin{aligned}
cov(\mu_q A_q \delta_{as\ qt}, \mu_r A_r \delta_{as\ rt}) = \\
&E[\mu_q]E[\mu_r]E[A_q]E[A_r]cov(\delta_{as\ qt}, \delta_{as\ rt}) + E[\mu_q]E[\mu_r]E[\delta_{as\ qt}]E[\delta_{as\ rt}]cov(A_q, A_r) + \\
&+ E[A_q]E[A_r]E[\delta_{as\ qt}]E[\delta_{as\ rt}]cov(\mu_q, \mu_r) + E[\mu_q]E[\mu_r]cov(A_q, A_r)cov(\delta_{as\ qt}, \delta_{as\ rt}) + \\
&+ E[A_q]E[A_r]cov(\mu_q, \mu_r)cov(\delta_{as\ qt}, \delta_{as\ rt}) + E[\delta_{as\ qt}]E[\delta_{as\ rt}]cov(\mu_q, \mu_r)cov(A_q, A_r) + \\
&+ cov(\mu_q, \mu_r)cov(A_q, A_r)cov(\delta_{as\ qt}, \delta_{as\ rt})
\end{aligned} \tag{S39}
$$

From equation S23 we deduce the expected value of $\delta_{as\ pt}$ once we know the expected values of the three rvs V_{0p}, t_{rp}, and q_p.

For the velocity V_{0p} we use the long-term creeping velocity $\hat{V}_0 = 20.2 \pm 1.1\ mm/yr$ obtained by Stevens and Avouac (2015), and we adjust it to each patch p multiplying it for $(1 - c_p)$, where c_p is the coupling coefficient varying from 0 to 1 obtained again by Stevens and Avouac (2015). We consider the coupling map to be exact, and we just rescale the uncertainty in the velocity $\hat{V}_0$ by the constant $(1 - c_p)$. For the fully locked patches ($c_p = 1$) we assign a coupling value equal to the highest c_p lower than 1. In this way we still admit afterslip on locked portions, as derived from our Model 1. If we do not operate in this way, and we completely neglect the afterslip on locked regions, we would end up with an underestimation of the total afterslip, and thus of the moment released. From the fitting of equation 7 we obtain for each patch the quantities $t_{rp} \pm \sigma_{t_{rp}}$ and $q_p \pm \sigma_{q_p}$. This means that for all these three rvs we do not have any information of the possible correlations. Anyway, we know for sure that V_{0p}, t_{rp}, and q_p are not independent. Indeed, from their definition we know that both t_{rp} and q_p depend on $1/V_{0p}$ (see Perfettini and Avouac, 2004). Even if we can not take into account all the correlations, we can give an estimation of the upper bound of $\delta_{as\ pt}$ final uncertainty. Indeed, using formula (3.48) of Taylor (1982) we have that:

$$\sigma_{\delta_{as\ pt}} \leq \left|\frac{\partial \delta_{as\ pt}}{\partial V_{0p}}\right| \sigma_{V_{0p}} + \left|\frac{\partial \delta_{as\ pt}}{\partial t_{rp}}\right| \sigma_{t_{rp}} + \left|\frac{\partial \delta_{as\ pt}}{\partial q_p}\right| \sigma_{q_p} \tag{S40}$$

The three partial derivatives are easy to calculate:

$$\begin{aligned}
\frac{\partial \delta_{as\ pt}}{\partial V_{0p}} &= t_{rp} \ln\left[1 + q_p\Big(\exp(t/t_{rp}) - 1\Big)\right] - t \\
\frac{\partial \delta_{as\ pt}}{\partial t_{rp}} &= V_{0p} \ln\left[1 + q_p\Big(\exp(t/t_{rp}) - 1\Big)\right] - V_{0p} q_p \frac{t}{t_{rp}} \frac{\exp(t/t_{rp})}{1 + q_p\Big(\exp(t/t_{rp}) - 1\Big)} \\
\frac{\partial \delta_{as\ pt}}{\partial q_p} &= V_{0p} t_{rp} \frac{\exp(t/t_{rp}) - 1}{1 + q_p\Big(\exp(t/t_{rp}) - 1\Big)}
\end{aligned} \tag{S41}$$

We now have all the quantities needed to calculate the upper bound of the uncertainty on the predicted moment $M_{0\ t}$. We will use this upper bound as the actual final uncertainty, and the variance will be just the square of such value. In our analog model we are considering independent spring-sliders. This means that the correlation between the slip on two patches q and r is zero, i.e. the covariance matrix of $\boldsymbol{\delta}_{as\ t}$ is diagonal, with null off-diagonal values.

S2.3.3. Variance of $M_{0t\to\infty}$

Using the same argumentation adopted for the calculation of the variance of M_{0t}, we can set the upper bound of $\delta_{as\ pt\to\infty}$ uncertainty to:

$$\sigma_{\delta_{as\ pt\to\infty}} \leq \left|\frac{\partial \delta_{as\ pt\to\infty}}{\partial V_{0p}}\right| \sigma_{V_{0p}} + \left|\frac{\partial \delta_{as\ pt\to\infty}}{\partial t_{rp}}\right| \sigma_{t_{rp}} + \left|\frac{\partial \delta_{as\ pt\to\infty}}{\partial q_p}\right| \sigma_{q_p} \tag{S42}$$

and we take for the variance $var(\delta_{as\ pt})$ the square of this upper bound.

It is easy to calculate the partial derivatives in equation S42, given equation S24:

$$\begin{aligned}
\frac{\partial \delta^{final}_{as\ pt}}{\partial V_{0p}} &= t_{rp} \ln q_p \\
\frac{\partial \delta^{final}_{as\ pt}}{\partial t_{rp}} &= V_{0p} \ln q_p \\
\frac{\partial \delta^{final}_{as\ pt}}{\partial q_p} &= \frac{V_{0p} t_{rp}}{q_p}
\end{aligned} \tag{S43}$$

Thus we have all the quantities we need to calculate the variance on the asymptotic value predicted for the moment released by the afterslip modelling every patch as a single 1-dimensional slider.

S3. Determination of Model 2 from the inversion of the pre- and post-seismic time series

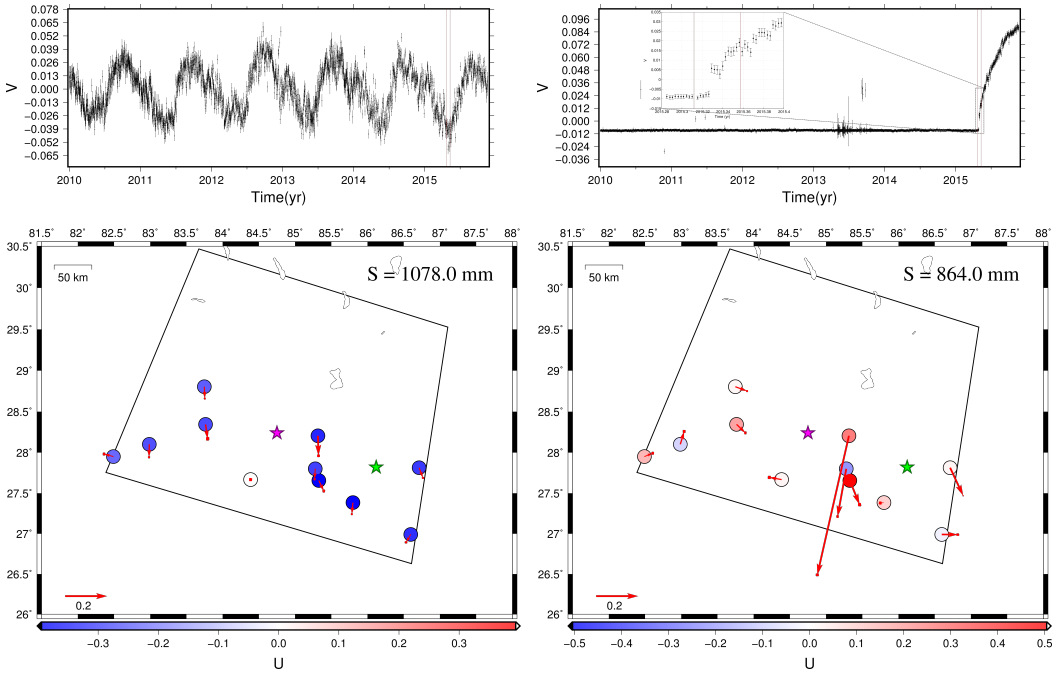
We present here the analysis of the near field GPS stations ($< 300\ km$ from the hypocenter) that recorded the post-seismic deformation and at least 2 years of pre-seismic deformation prior to the Gorkha earthquake, i.e. from 2010.0 to the last available epoch, 2015.8877. This continuity in time allows us to correct each time series for the long term linear trend. This selection limits the number of stations to 10. We have included also the station NAST (1.23 yr of data in the pre-seismic stage) because of its strategic location for the post-seismic analysis (see Figure 1). The total number of stations used in this analysis is thus 11 (see column 5, Table 1). As we can see from Figure 3, even if NAST has a shorter pre-seismic record, it is well reproduced by the ICA decomposition. The vbICA analysis applied to the corrected time series (i.e., detrended and with offsets corrected for) yields that four ICs are optimal to reconstruct the time series, based on the ARD criterion (Gualandi et al., 2015). The ICA does well at separating the post-seismic signal, which clearly shows up in the second IC, and the seasonal signals (see Figure S1). A delay of 5 days (see box in Figure S1b) in the starting of the post-seismic deformation is observed and is probably due to the fact that some post-seismic data are still explained by the Gaussian distributions used to model the pre-seismic stage. Indeed, even if we correctly separate the sources, the pdf of each source is modeled by a mix of Gaussian distributions. If we model the pdf of a source using a number of Gaussians that is not large enough to capture all the details of the actual source pdf, then we can have that some points are not optimally reconstructed. This may be the case when we are moving from the pre- to the post-seismic phases. In order to prevent this delay it might be necessary to increase the number of Gaussians in the mix, but such an increase in the number of parameters of the model is not justified by an F-test.

We do not observe any significant pre-seismic deformation, neither in the time function related to the post-seismic component nor in any other component (Figure S1). This is consistent with the fact that we did not observe any pre-seismic anomaly in the position time series, and indicates that there is no subtle signal coherent among the stations that the ICA could have picked up. Any pre-seismic signal would have to be hidden in the residuals obtained by subtracting the reconstructed times series from the observed ones. The Root Mean Square of these residuals is $\sim 2\ mm$ for the horizontal components, and $\sim 7\ mm$ for the vertical.

A frequency analysis of the temporal pattern associated to each IC shows that the first, third, and fourth ICs contain signals at 1 and 2 yr^{-1} frequencies (see Figure S2 of the Supplementary Material), proving that the seasonal contributions have been correctly separated from the post-seismic decay. As in the analysis reported in the main text, we observe that the CTWN station, which sits on alluvial sediments in the piedmont of the Himalaya, is clearly out-of-phase compared to all the other stations, probably due to poro-elastic deformation of the sediments. Therefore, we treated it independently (the poro-elastic signal at this station is represented by a single component determined from stacking the time series before the Gorkha earthquake). CTWN is the only station among the 10 considered that lies in the foreland sedimentary basin.

In order to invert the post-seismic signal, associated to the second IC, we follow the same iterative scheme described in the main text (see Section 3). The afterslip distribution that we

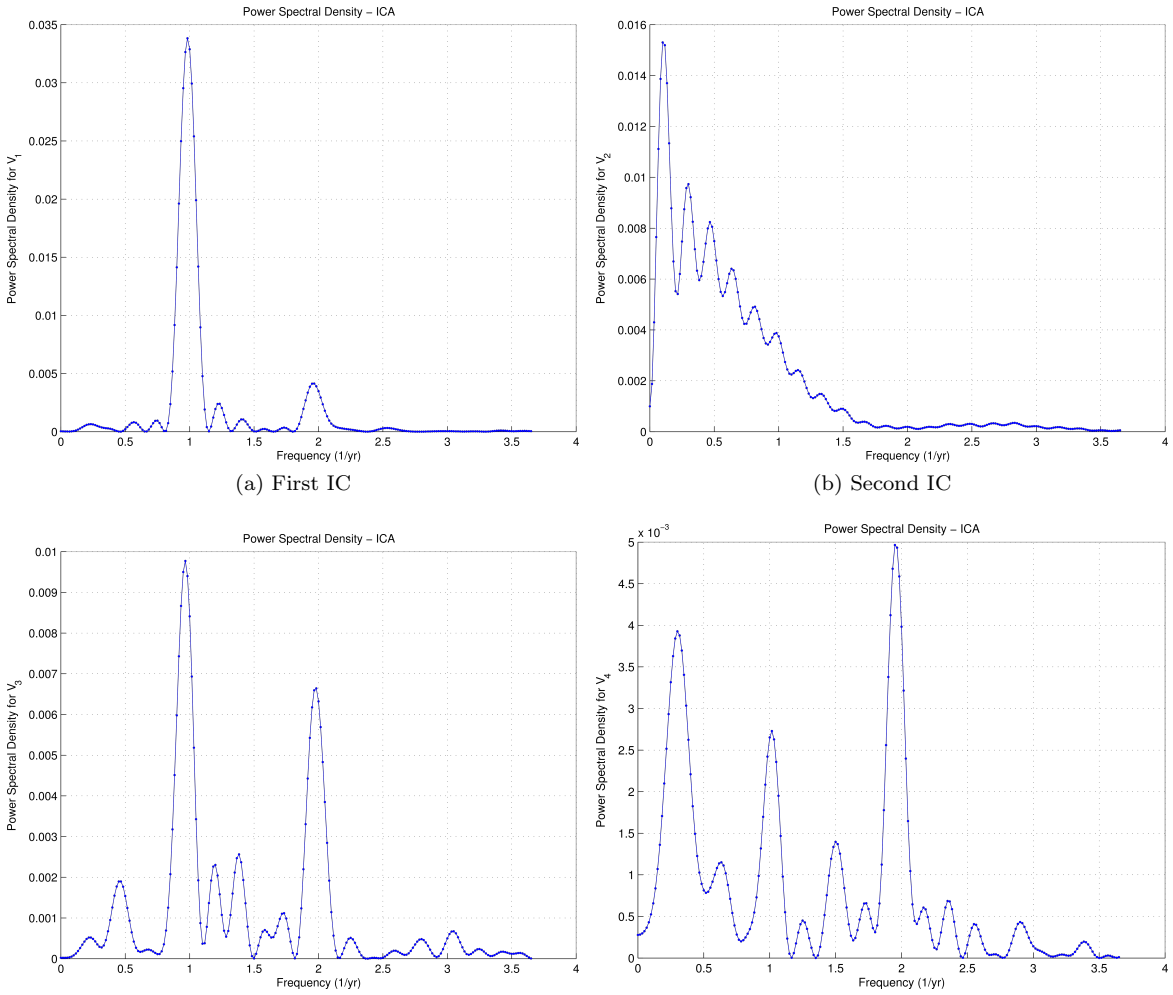
obtain is shown in Figure S17. The large afterslip patch downdip of the co-seismic rupture is a common feature between this model and Model 1 which is discussed in the main text (Figure 6), and was derived from the analysis of the complete dataset operating during the post-seismic phase. This patch does not overlap with the co-seismic ruptures of the mainshock and of the May 12^{th} aftershock. Our Model 2 is not resolved spatially as well as Model 1, resulting in the blurring of the patches belonging to regions A and B, and in the failure of detection of afterslip in region C. This is likely due to the less density of GPS stations adopted for the analysis.



(a) First IC (b) Second IC

(c) Third IC (d) Fourth IC

Figure S1: ICs of the time series, corrected for a linear trend and two co-seismic offsets, at 11 continuous GPS stations (column 5, Table 1) for the time span [2010.0, 2015.8877]. Top of each panel: temporal evolution of the IC. Map view: corresponding spatial response. Vectors: horizontal contributions. Colored dots: vertical response. The station CTWN has not been used to derive the ICs. Its spatial distribution to the second IC is calculated projecting the data onto such IC after correction of seasonal signals by ad hoc stacking. S: weight of each component. Stars as in Figure 1.



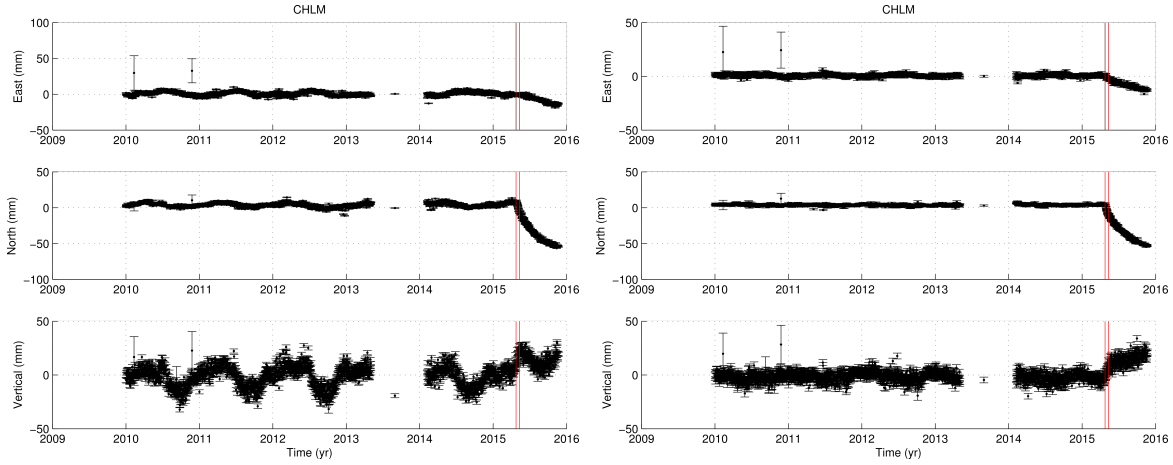
(a) First IC

(b) Second IC

(c) Third IC

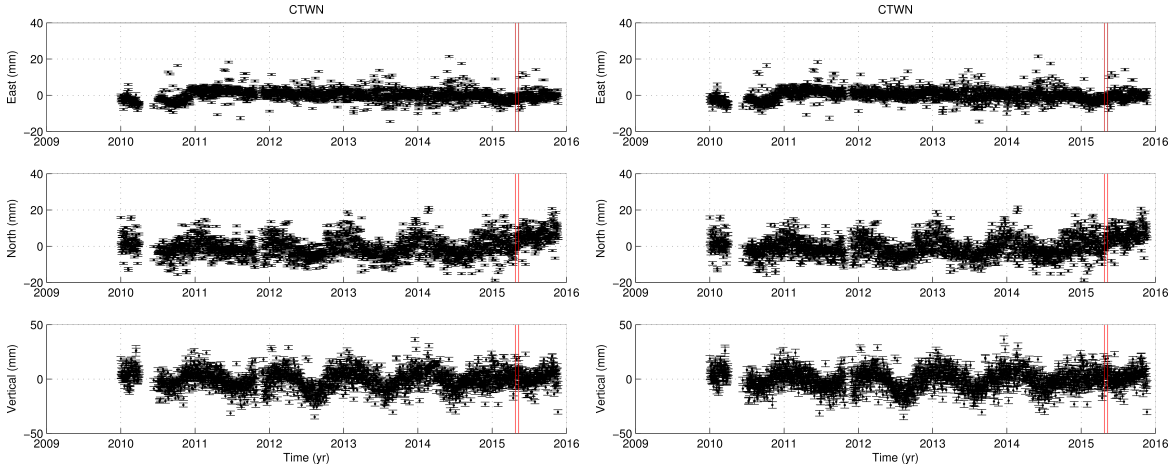
(d) Fourth IC

Figure S2: Power Spectral Density (PSD) showing the frequency content of the temporal signals of Figure S1. For the first, third, and fourth ICs we clearly identify two peaks corresponding to periods of 0.5 and 1 yr. The fourth IC shows also a peak at very low frequencies ($\sim 0.3\ yr^{-1}$), corresponding to a period of $\sim 3\ yr$. Since we are studying only 6 yr of data, such a multi-annual signal can not be properly assessed.



(a) CHLM

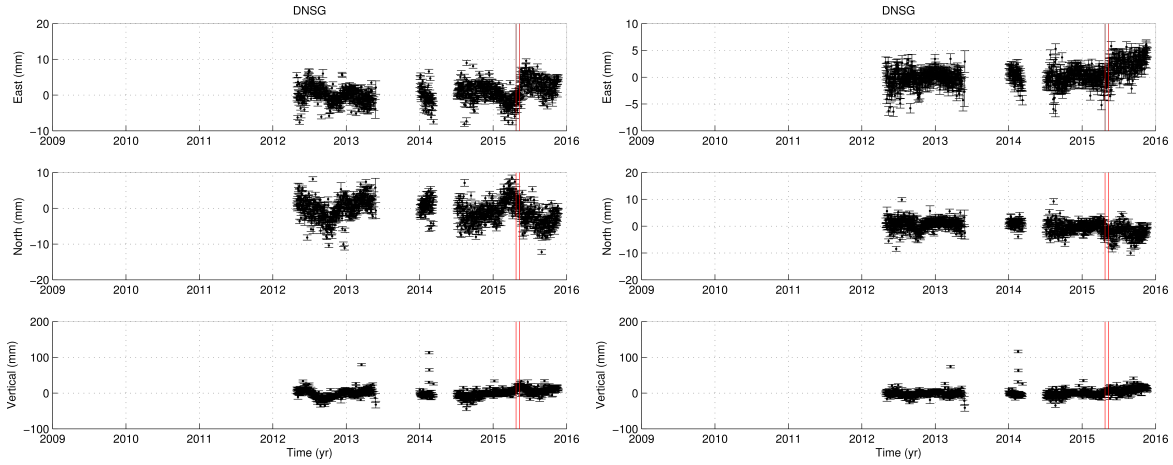
(b) CHLM



(c) CTWN

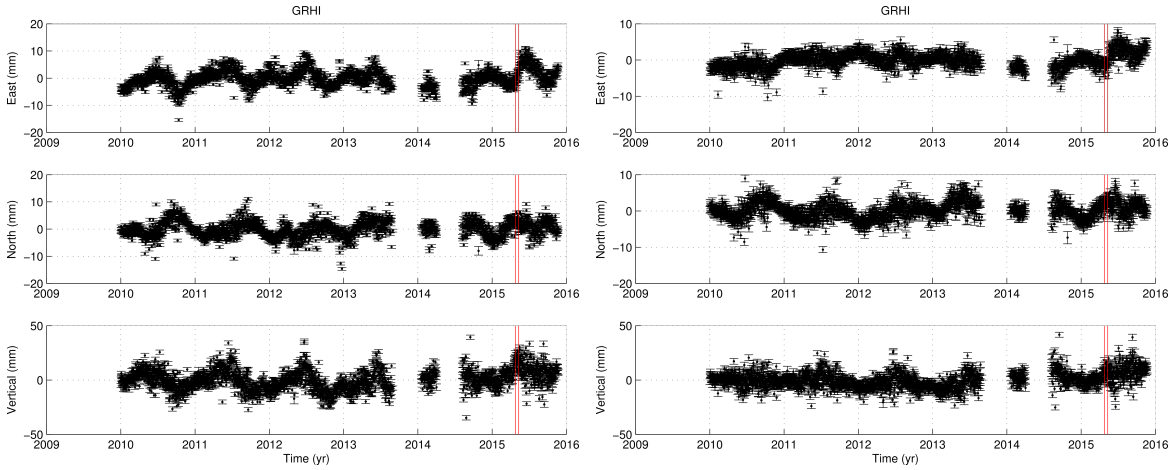
(d) CTWN

Figure S3: Left: Detrended and offset corrected position time series for the stations CHLM and CTWN. Right: Detrended, offset, and seasonal signals as retrieved from the vbICA corrected position time series for the same stations.



(a) DNSG

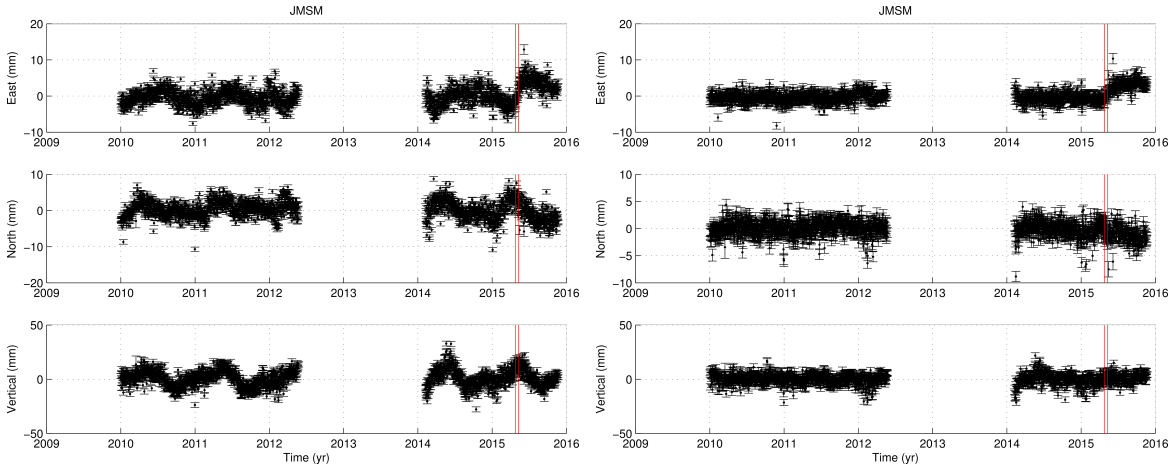
(b) DNSG



(c) GRHI

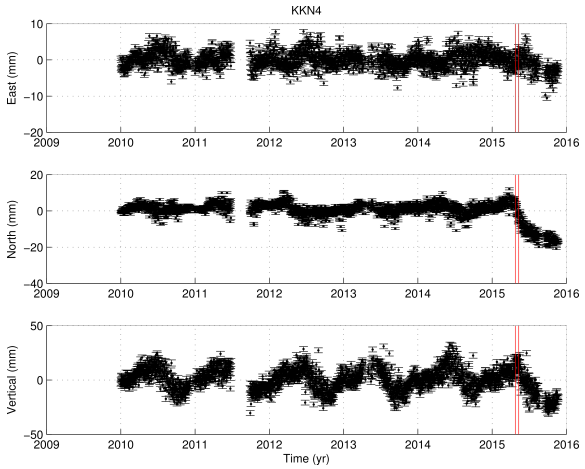
(d) GRHI

Figure S4: Left: Detrended and offset corrected position time series for the stations DNSG and GRHI. Right: Detrended, offset, and seasonal signals as retrieved from the vbICA corrected position time series for the same stations.

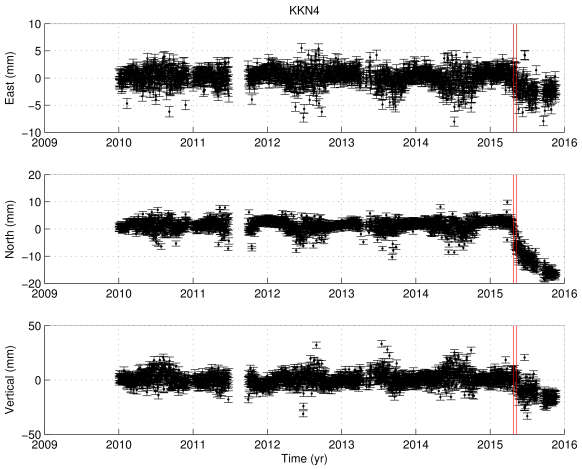


(a) JMSM

(b) JMSM

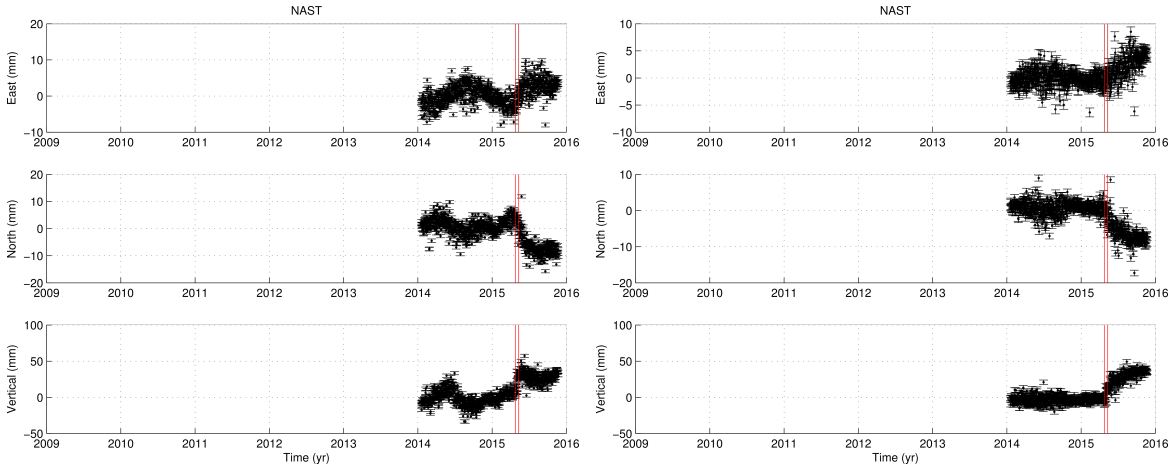


(c) KKN4



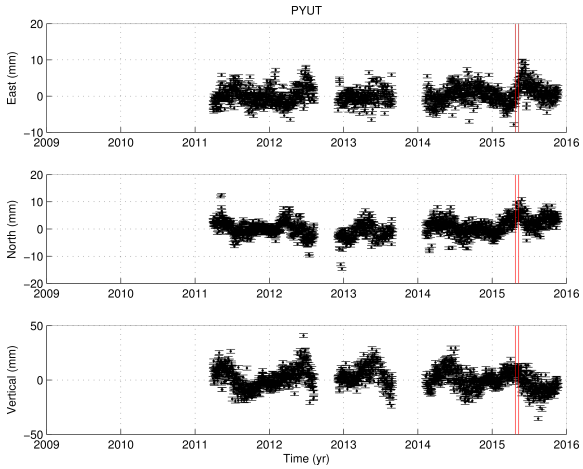
(d) KKN4

Figure S5: Left: Detrended and offset corrected position time series for the stations JSMS and KKN4. Right: Detrended, offset, and seasonal signals as retrieved from the vbICA corrected position time series for the same stations.

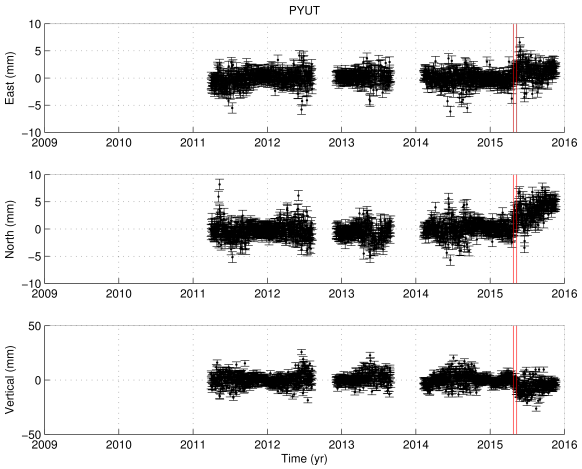


(a) NAST

(b) NAST

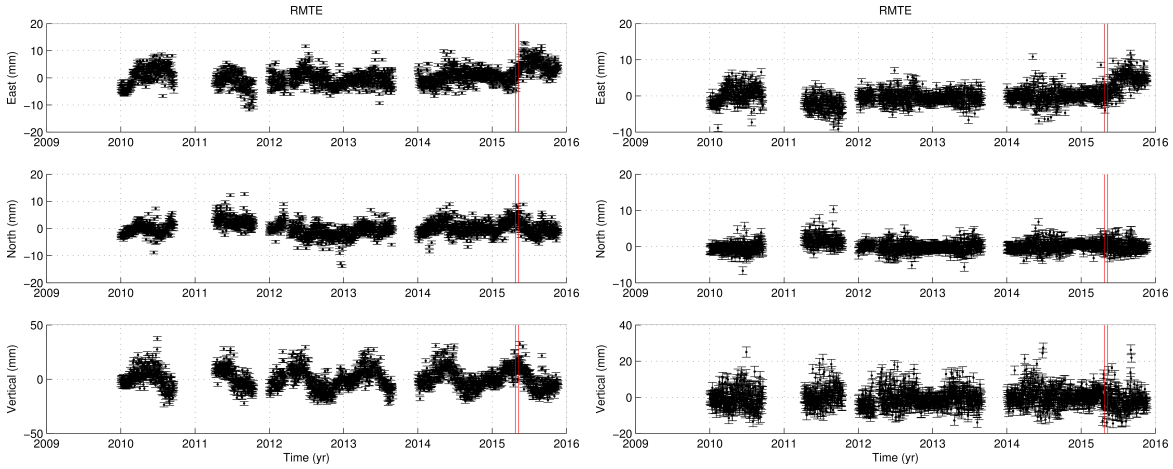


(c) PYUT



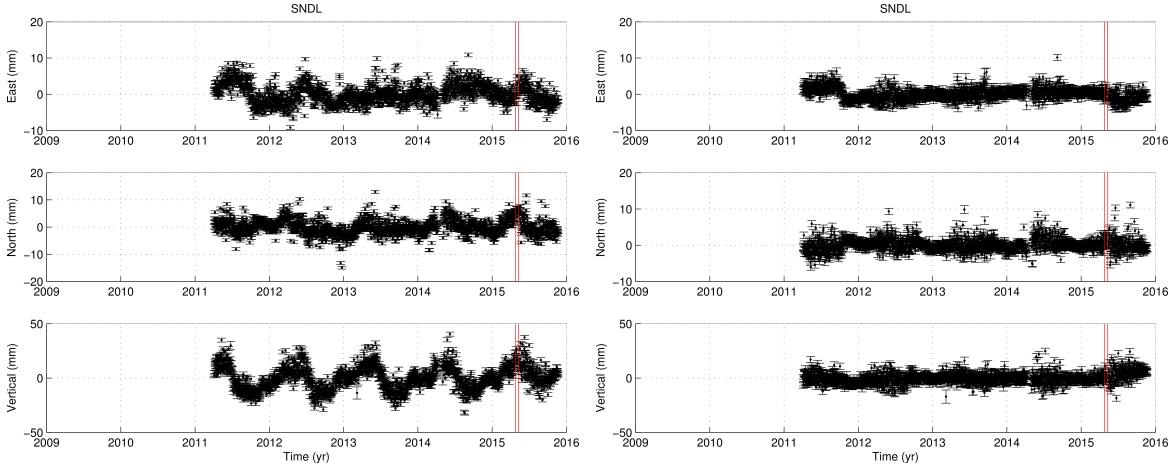
(d) PYUT

Figure S6: Left: Detrended and offset corrected position time series for the stations NAST and PYUT. Right: Detrended, offset, and seasonal signals as retrieved from the vbICA corrected position time series for the same stations.



(a) RMTE

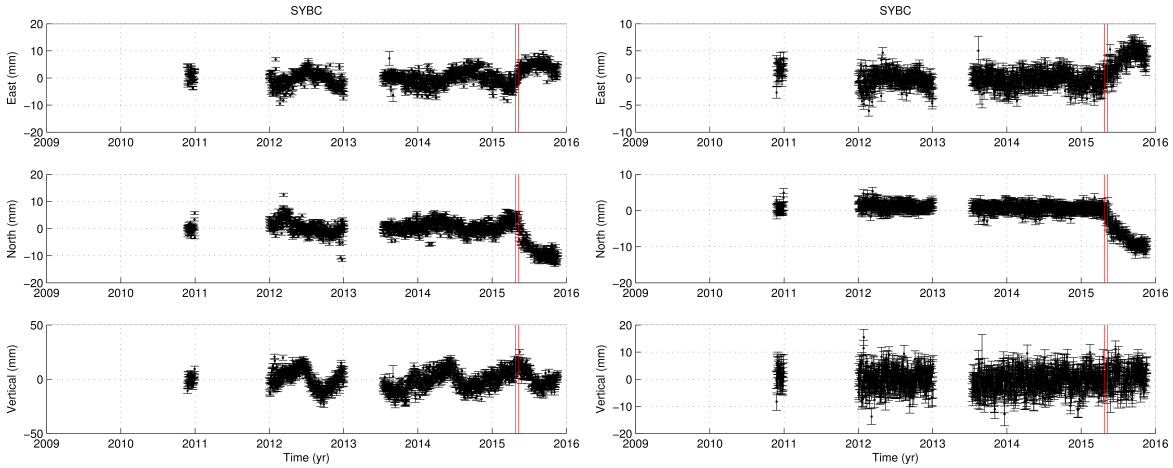
(b) RMTE



(c) SNDL

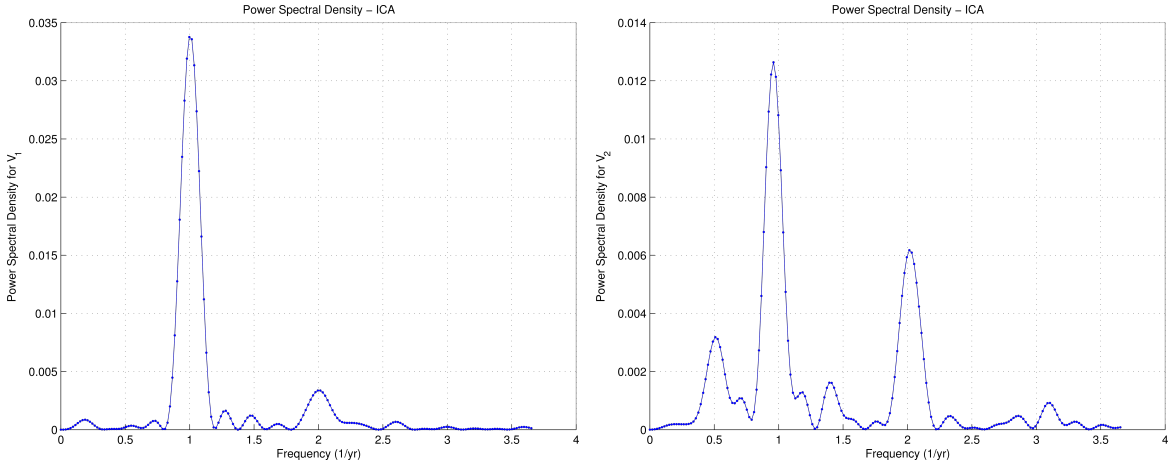
(d) SNDL

Figure S7: Left: Detrended and offset corrected position time series for the stations RMTE and SNDL. Right: Detrended, offset, and seasonal signals as retrieved from the vbICA corrected position time series for the same stations.



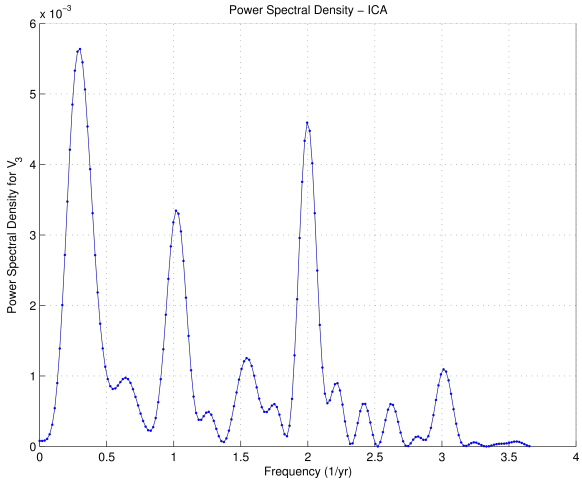
(a) SYBC (b) SYBC

Figure S8: Left: Detrended and offset corrected position time series for the station SYBC. Right: Detrended, offset, and seasonal signals as retrieved from the vbICA corrected position time series for the same station.



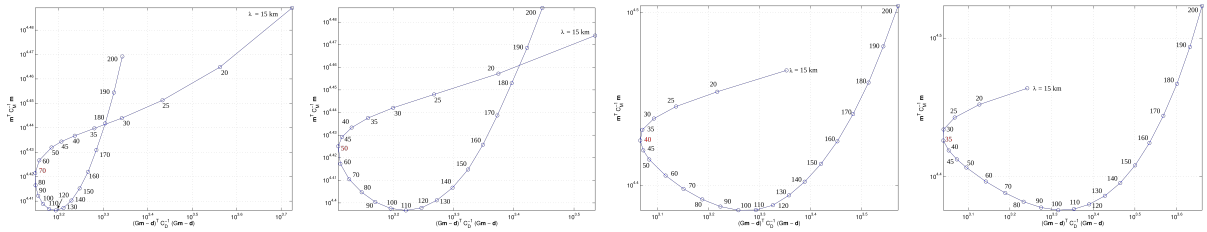
(a) First IC

(b) Second IC

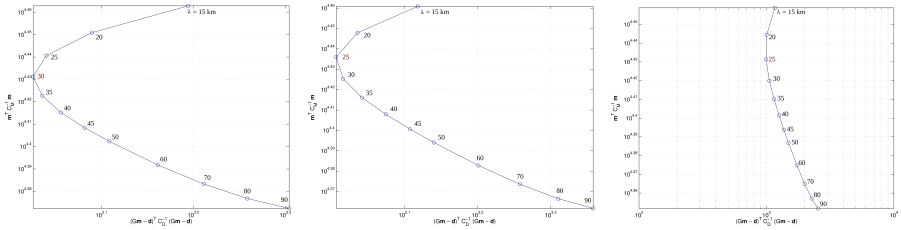


(c) Third IC

Figure S9: Power Spectral Density (PSD) showing the frequency content of the temporal signals of Figure 4. All the ICs clearly show two peaks corresponding to periods of 0.5 and 1 yr. The third IC shows also a peak at very low frequencies ($\sim 0.3\ yr^{-1}$), corresponding to a period of $\sim 3\ yr$. Since we are studying less than 6 yr of data, such a multi-annual signal can not be properly assessed.

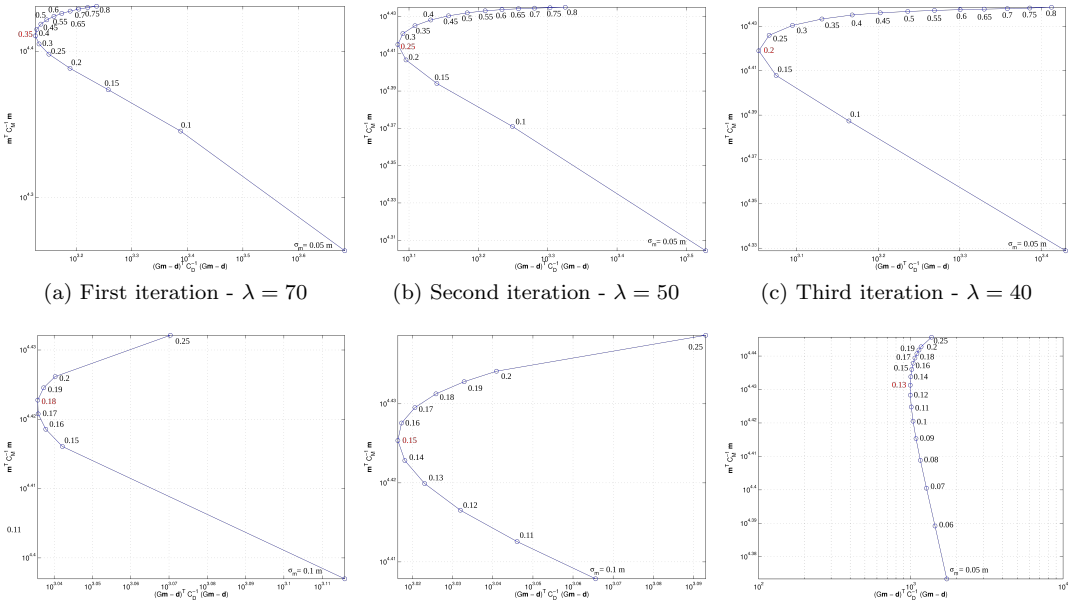


(a) First iteration - $\sigma_{\mathbf{m}} = 0.5$ (b) Second iteration - $\sigma_{\mathbf{m}} = 0.35$ (c) Third iteration - $\sigma_{\mathbf{m}} = 0.25$ (d) Fourth iteration - $\sigma_{\mathbf{m}} = 0.2$



(e) Fifth iteration - $\sigma_{\mathbf{m}} = 0.18$ (f) Sixth iteration - $\sigma_{\mathbf{m}} = 0.15$ (g) Seventh iteration - $\sigma_{\mathbf{m}} = 0.13$

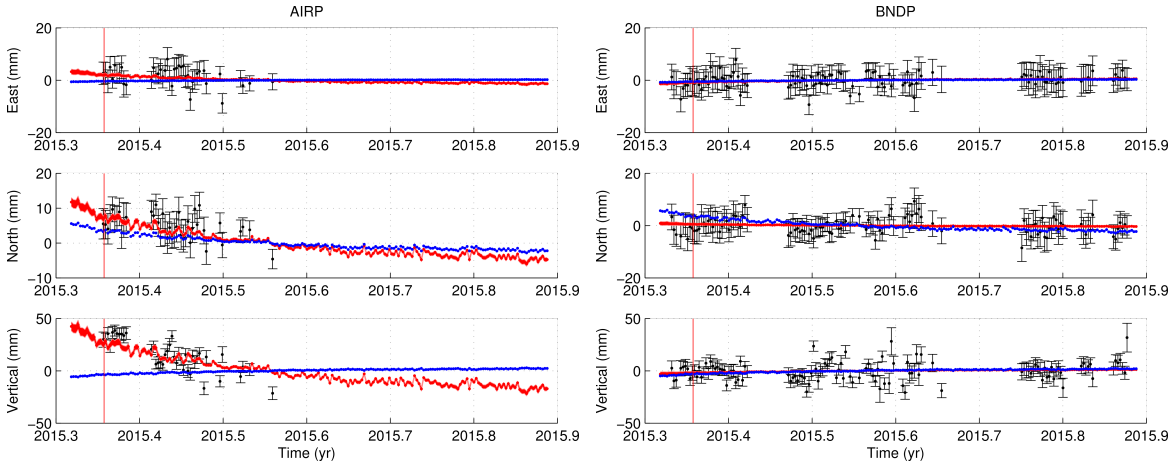
Figure S10: Norm of the model vs norm of the misfit space. We chose the λ value corresponding to the minimum of the misfit (value highlighted in red).



(a) First iteration - $\lambda = 70$ (b) Second iteration - $\lambda = 50$ (c) Third iteration - $\lambda = 40$

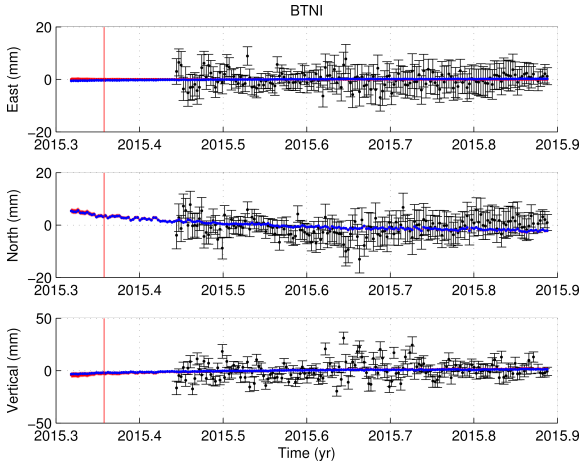
(d) Fourth iteration - $\lambda = 35$ (e) Fifth iteration - $\lambda = 30$ (f) Sixth iteration - $\lambda = 25$

Figure S11: Norm of the model vs norm of the misfit space. We chose the $\sigma_{\mathbf{m}}$ value corresponding to the minimum of the misfit (value highlighted in red).

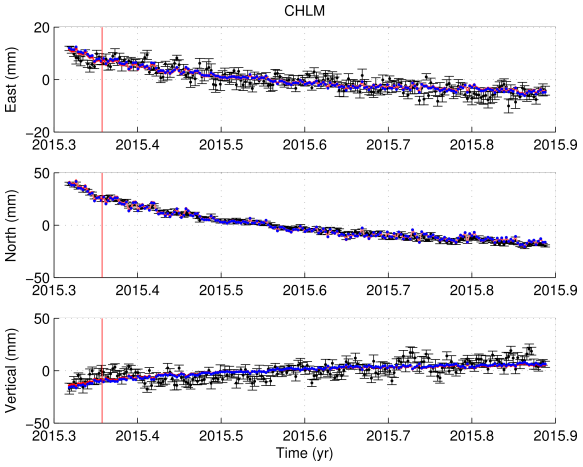


(a) AIRP

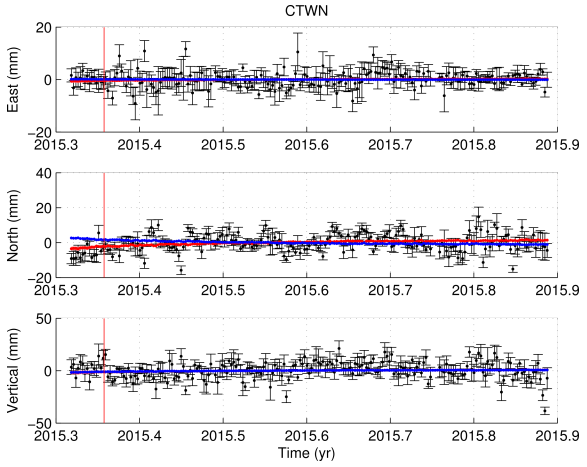
(b) BNDP



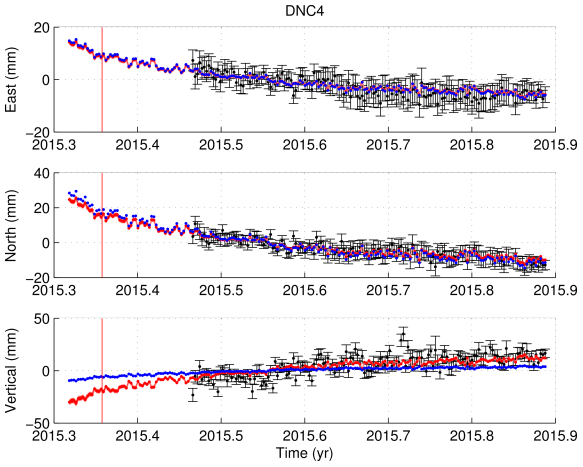
(c) BTNI



(d) CHLM

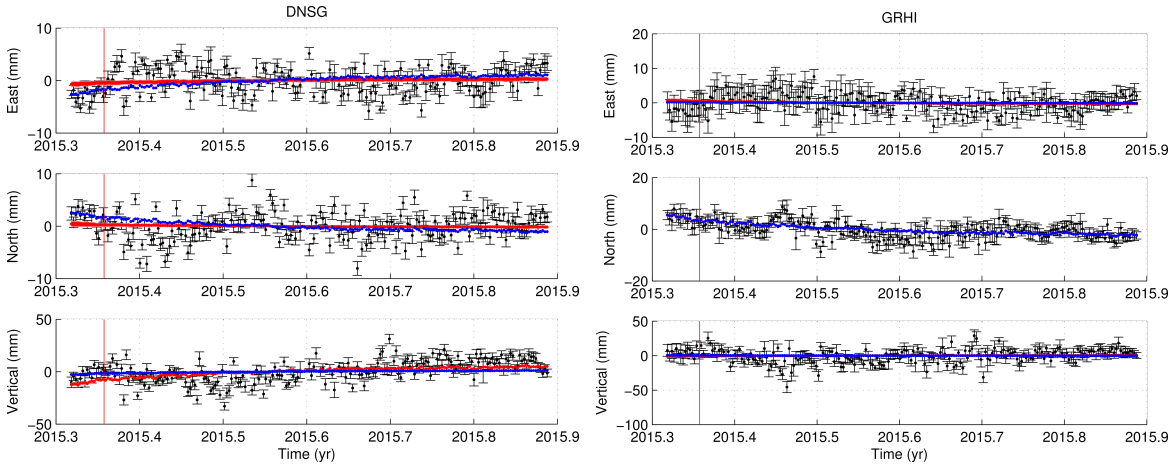


(e) CTWN



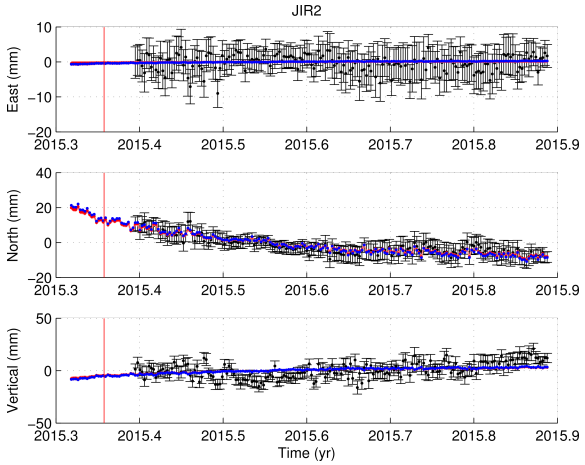
(f) DNC4

Figure S12: Detrended and offsets corrected position time series (black dots) and corresponding 1σ uncertainty (black errorbars) for the stations AIRP, BNDP, BTNI, CHLM, CTWN, and DNC4. Red dots: contribution of the first IC for analysis 1. Blue dots: modeled displacements related to the first IC. Red vertical line: main aftershock epoch.

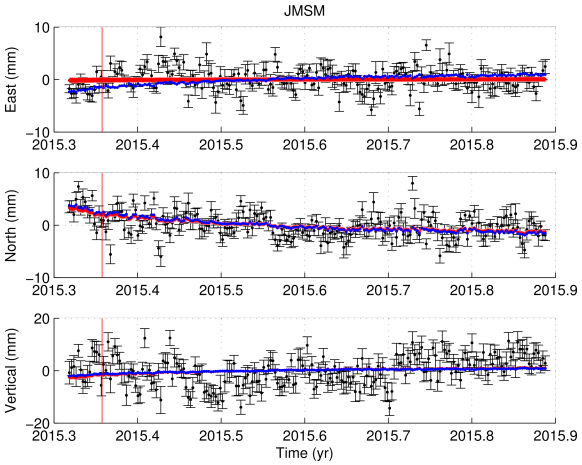


(a) DNSG

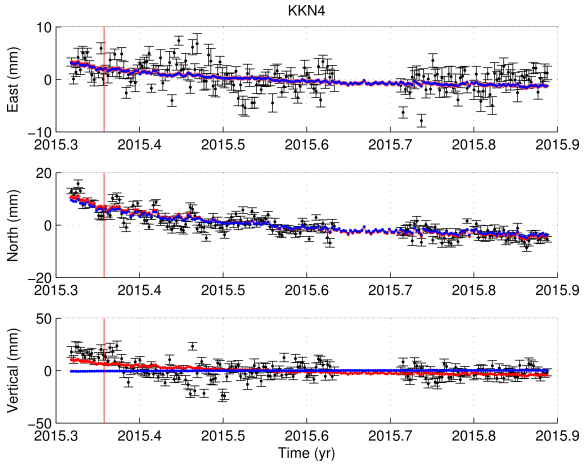
(b) GRHI



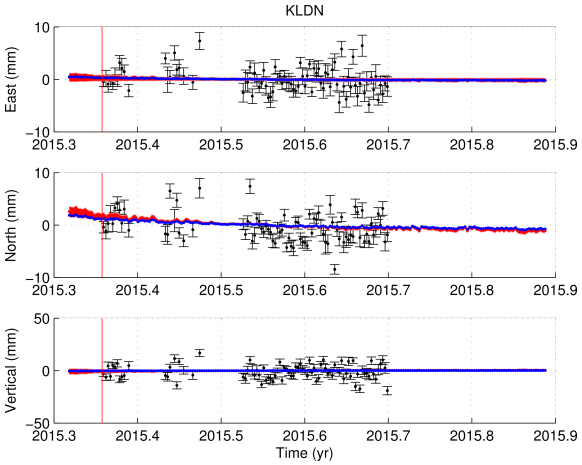
(c) JIR2



(d) JMSM

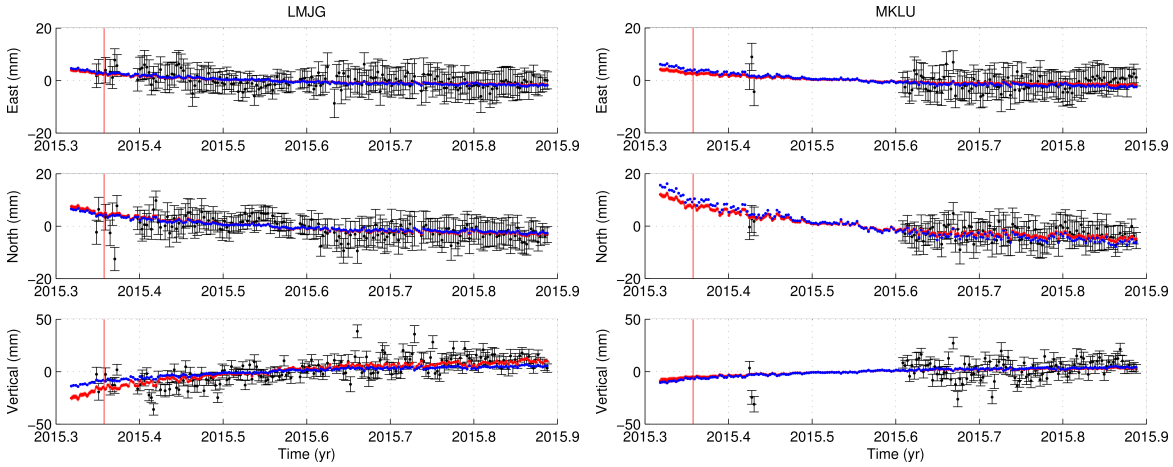


(e) KKN4



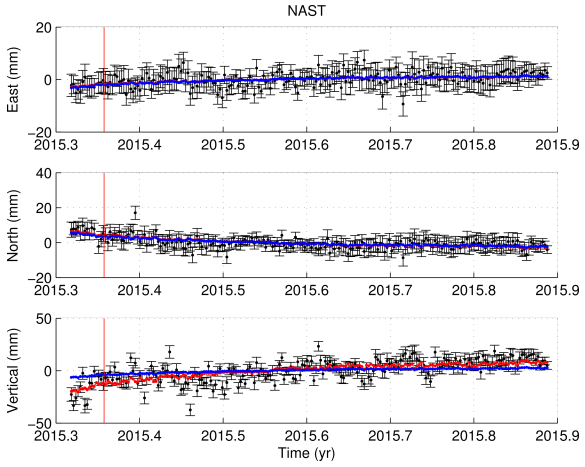
(f) KLDN

Figure S13: Detrended and offsets corrected position time series (black dots) and corresponding 1σ uncertainty (black errorbars) for the stations DNSG, GRHI, JIR2, JMSM, KKN4, and KLDN. Red dots: contribution of the first IC for analysis 1. Blue dots: modeled displacements related to the first IC. Red vertical line: main aftershock epoch.

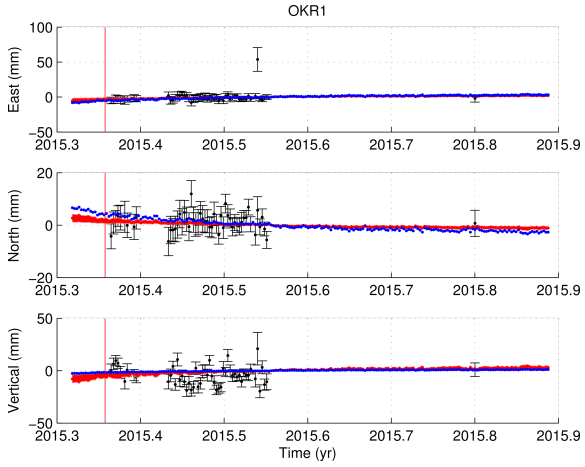


(a) LMJG

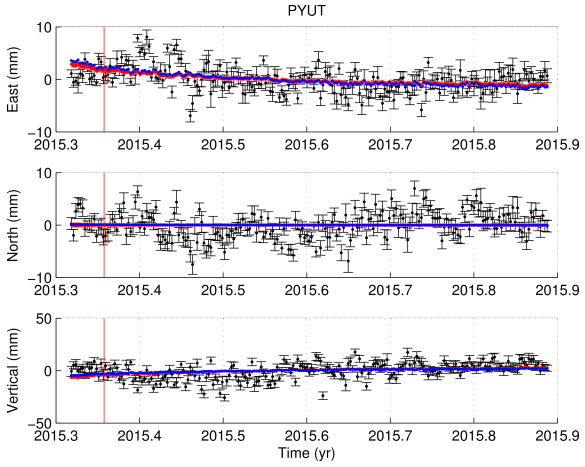
(b) MKLU



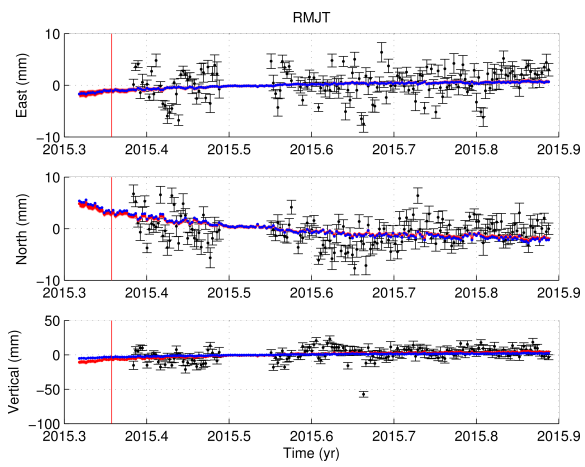
(c) NAST



(d) OKR1

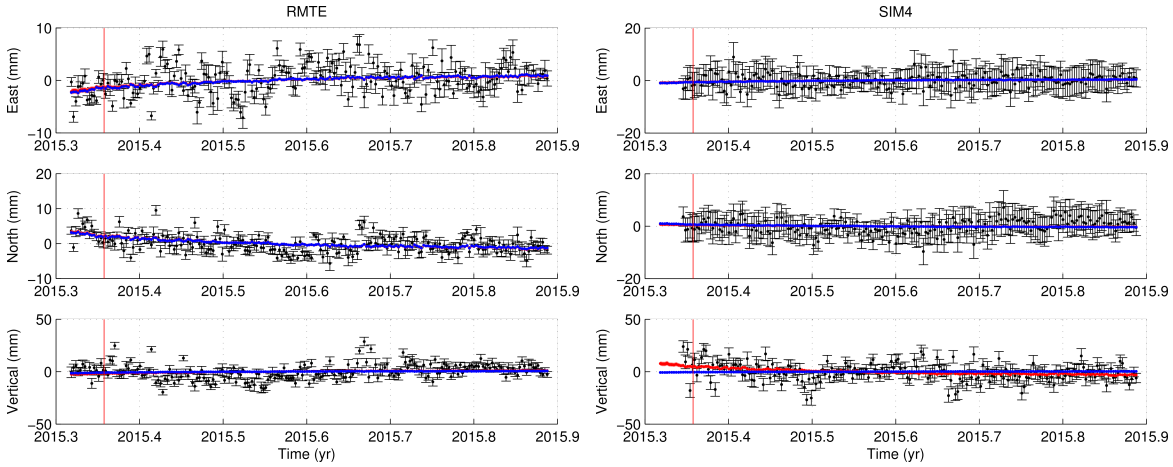


(e) PYUT



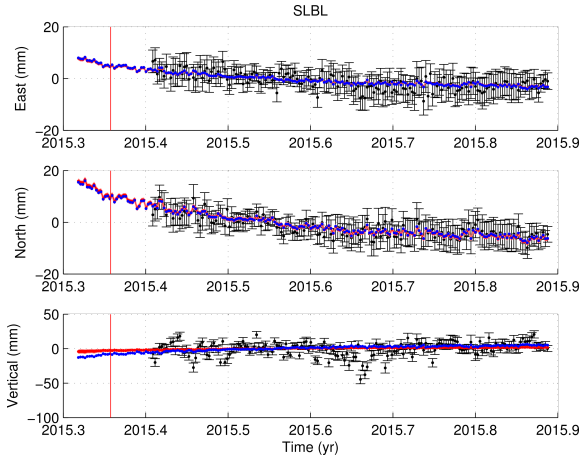
(f) RMJT

Figure S14: Detrended and offsets corrected position time series (black dots) and corresponding 1σ uncertainty (black errorbars) for the stations LMJG, MKLU, NAST, OKR1, PYUT, and RMJT. Red dots: contribution of the first IC for analysis 1. Blue dots: modeled displacements related to the first IC. Red vertical line: main aftershock epoch.

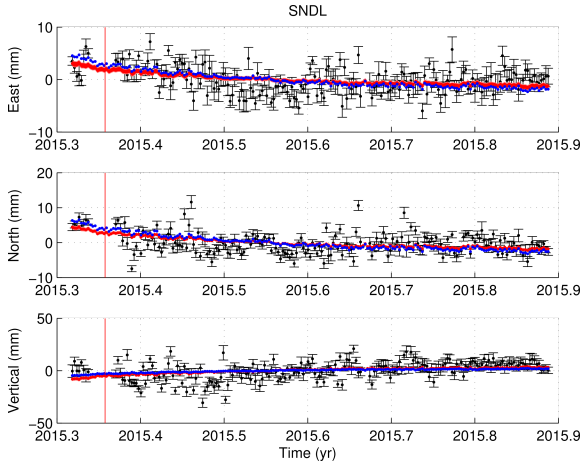


(a) RMTE

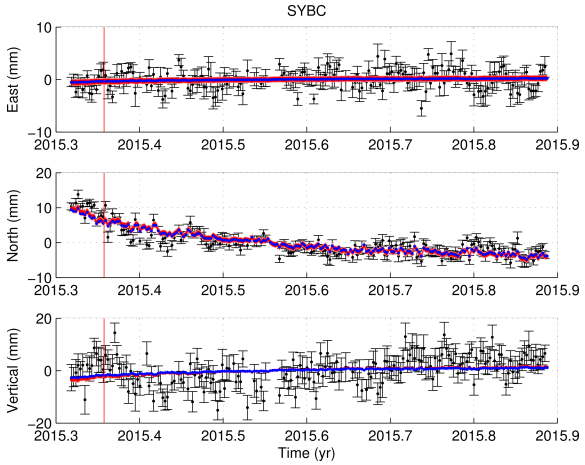
(b) SIM4



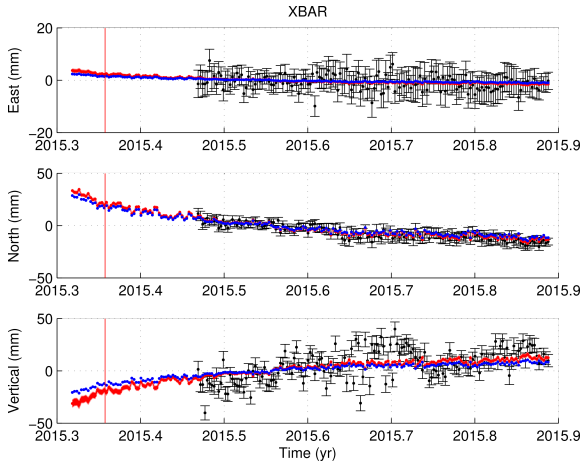
(c) SLBL



(d) SNDL

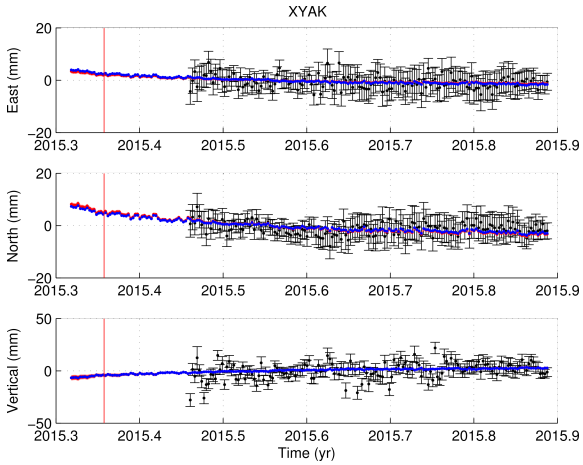


(e) SYBC



(f) XBAR

Figure S15: Detrended and offsets corrected position time series (black dots) and corresponding 1σ uncertainty (black errorbars) for the stations RMTE, SIM4, SLBL, SNDL, SYBC, and XBAR. Red dots: contribution of the first IC for analysis 1. Blue dots: modeled displacements related to the first IC. Red vertical line: main aftershock epoch.



(a) XYAK

Figure S16: Detrended and offsets corrected position time series (black dots) and corresponding 1σ uncertainty (black errorbars) for the station XYAK. Red dots: contribution of the first IC for analysis 1. Blue dots: modeled displacements related to the first IC. Red vertical line: main aftershock epoch.

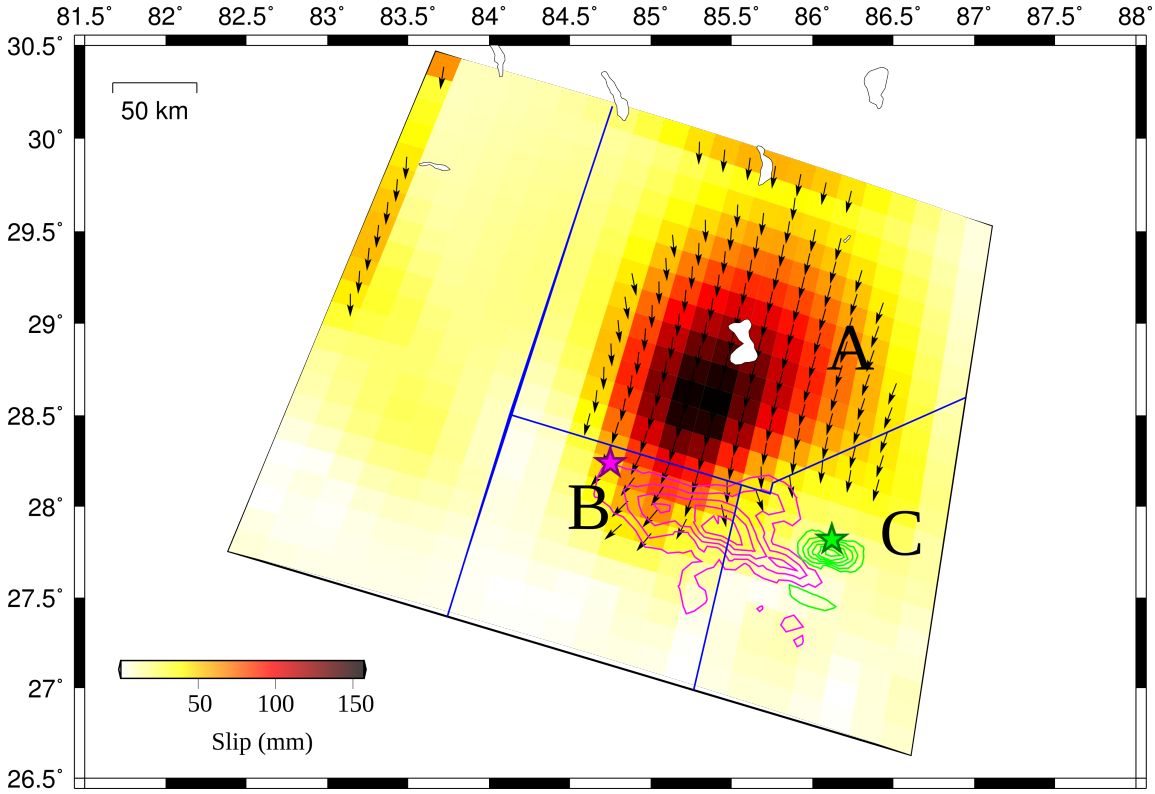
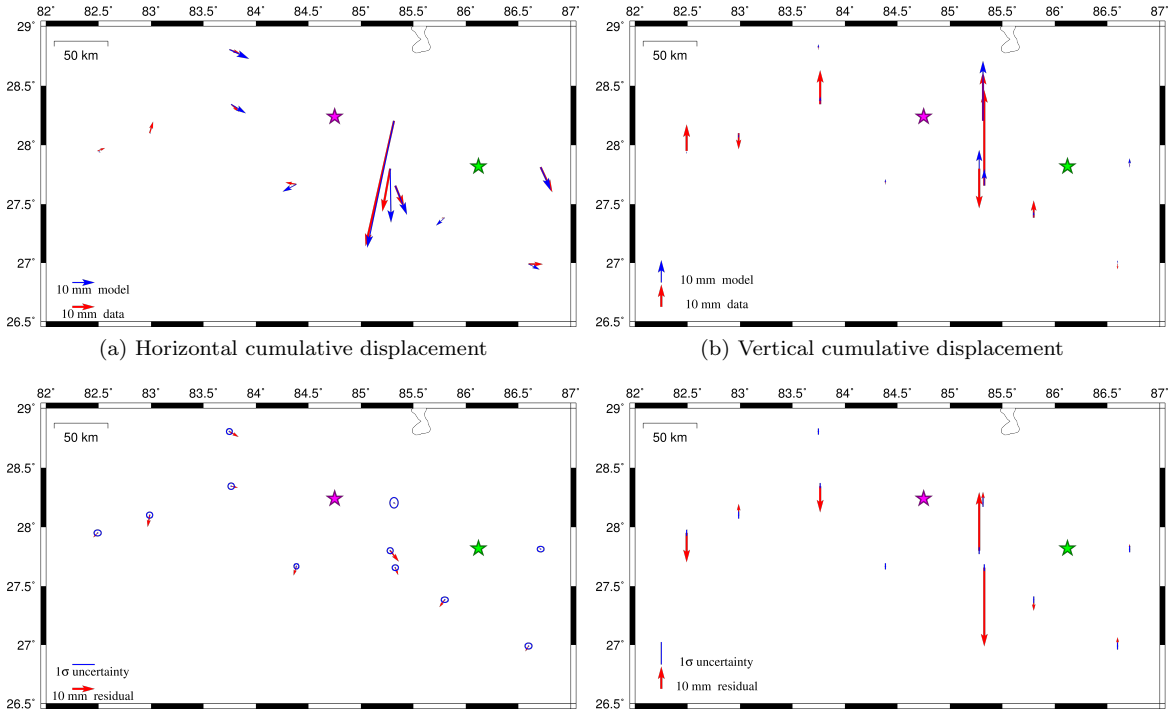


Figure S17: Afterslip distribution (color palette) 210 d after the mainshock from the inversion of IC number 2 (Figure S1b) deduced from the study of a subset of 11 stations (column 5 Table 1). Arrows: slip direction for patches with more than 25% of maximum slip. Stars, contours, and macro-region (A, B, and C) as in Figure 6.



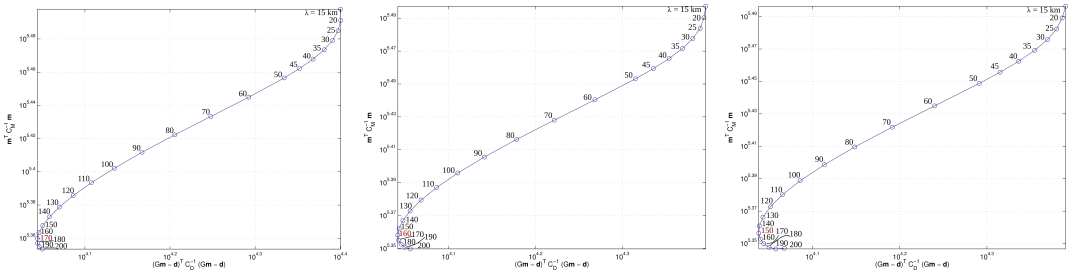
(a) Horizontal cumulative displacement

(b) Vertical cumulative displacement

(c) Horizontal residuals

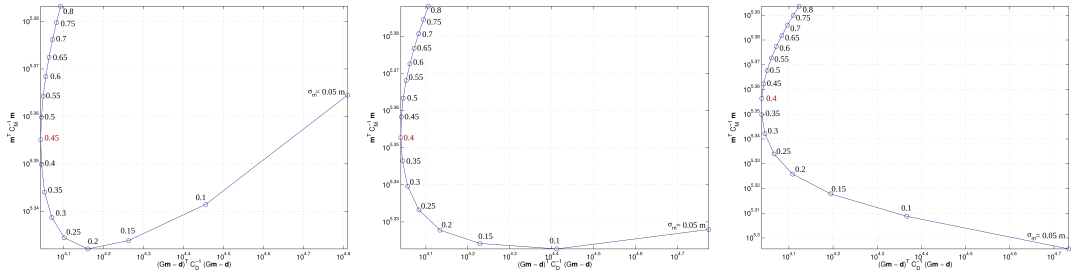
(d) Vertical residuals

Figure S18: Cumulative displacement vectors for horizontal (S18a) and vertical (S18b) components. (S18c) and (S18d): residuals for horizontal and vertical components (red arrows) and 1σ uncertainties (blue ellipses and bars). Stars as in Figure 1.



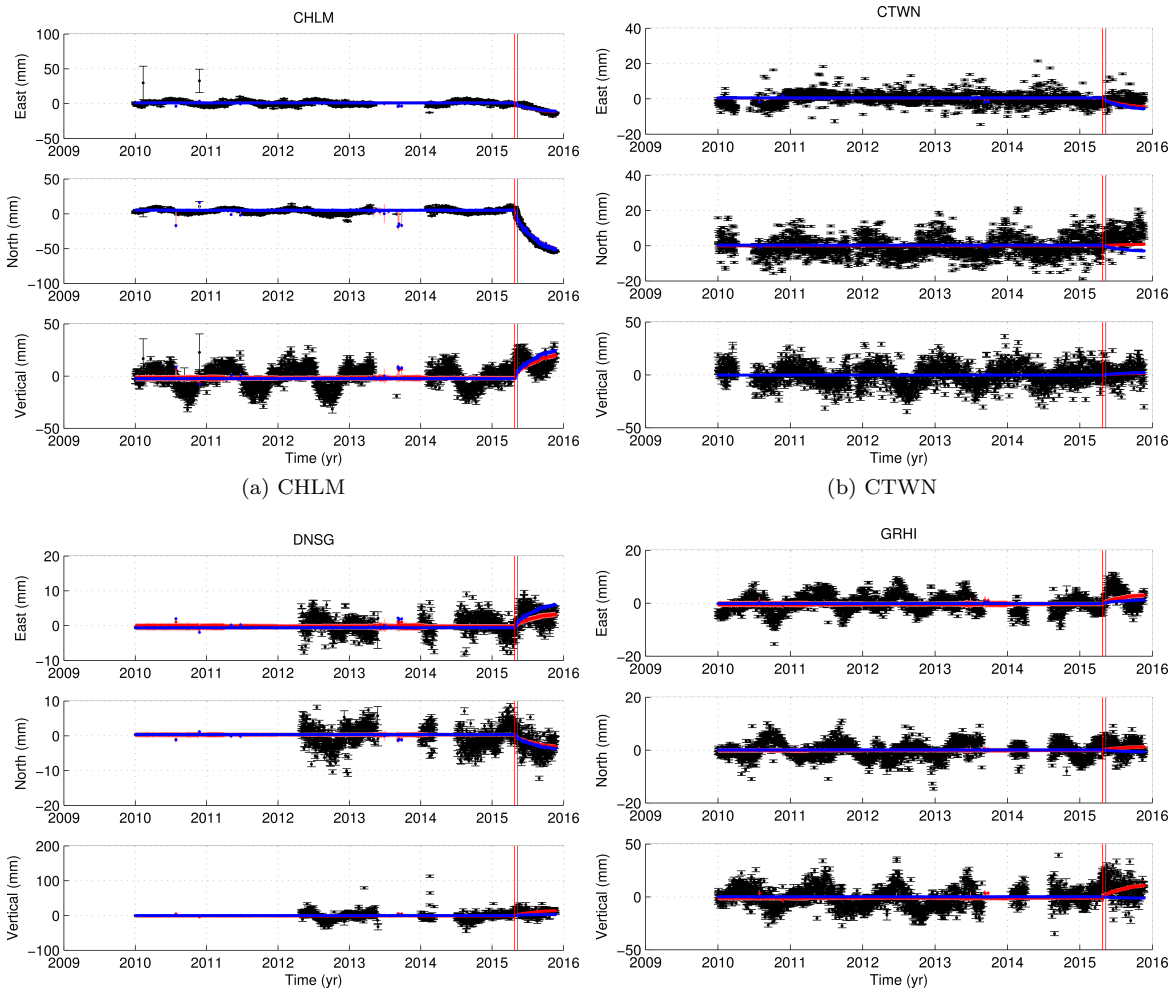
(a) First iteration - $\sigma_{\mathbf{m}} = 0.5\ m$ (b) Second iteration - $\sigma_{\mathbf{m}} = 0.45\ m$ (c) Third iteration - $\sigma_{\mathbf{m}} = 0.4\ m$

Figure S19: Norm of the model vs norm of the misfit space. We chose the λ value corresponding to the minimum of the misfit (value highlighted in red).



(a) First iteration - $\lambda = 170\ km$ (b) Second iteration - $\lambda = 160\ km$ (c) Third iteration - $\lambda = 150\ km$

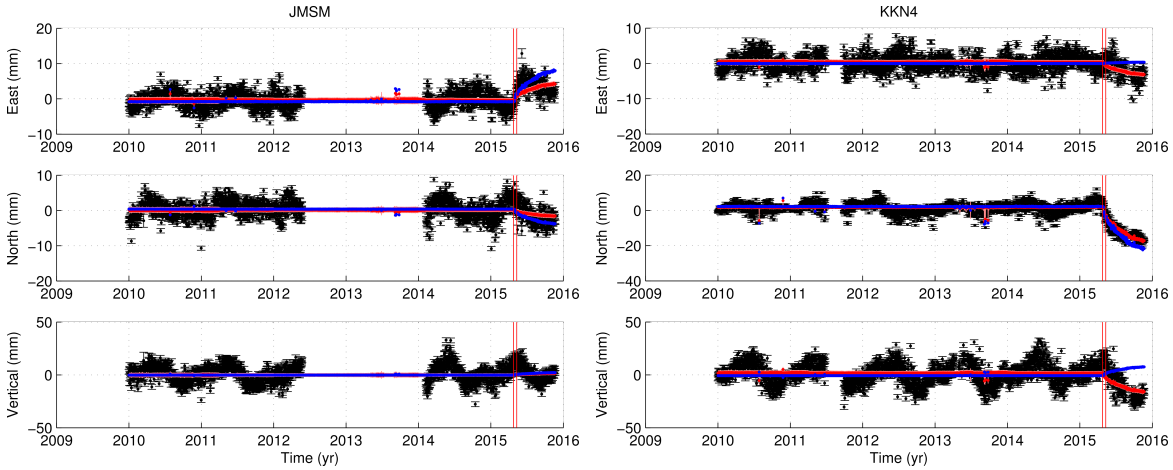
Figure S20: Norm of the model vs norm of the misfit space. We chose the $\sigma_{\mathbf{m}}$ value corresponding to the minimum of the misfit (value highlighted in red).



(a) CHLM (b) CTWN

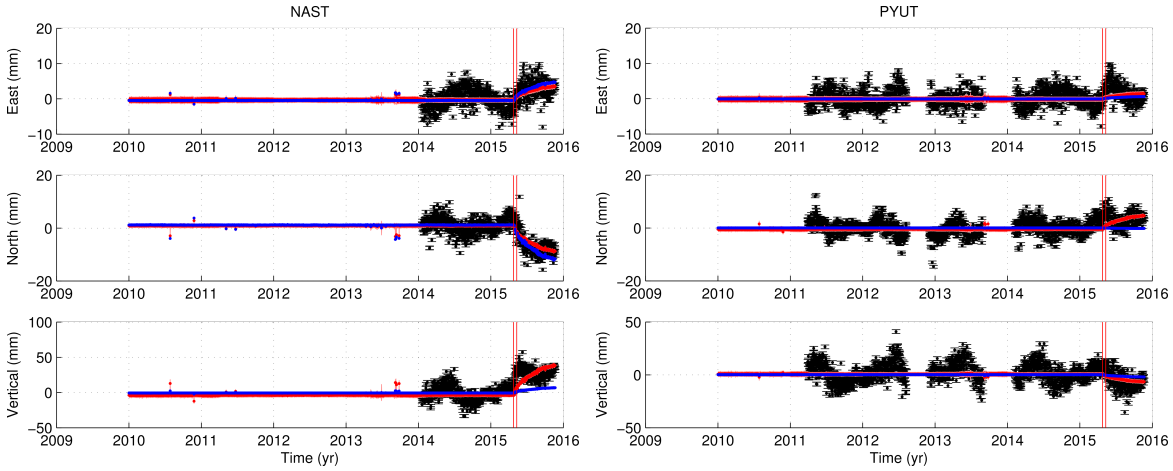
(c) DNSG (d) GRHI

Figure S21: Detrended and offsets corrected position time series (black dots) for the stations CHLM, CTWN, DNSG, and GRHI. Red dots: contribution of the second IC for analysis 2. Blue dots: modeled displacements related to the second IC. Red vertical lines: mainshock and main aftershocks epochs.



(a) JMSM

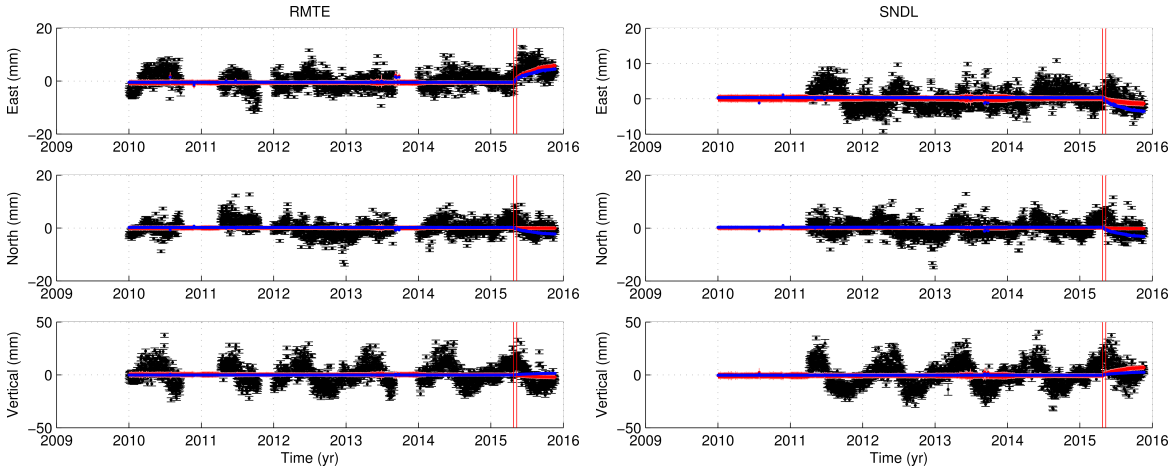
(b) KKN4



(c) NAST

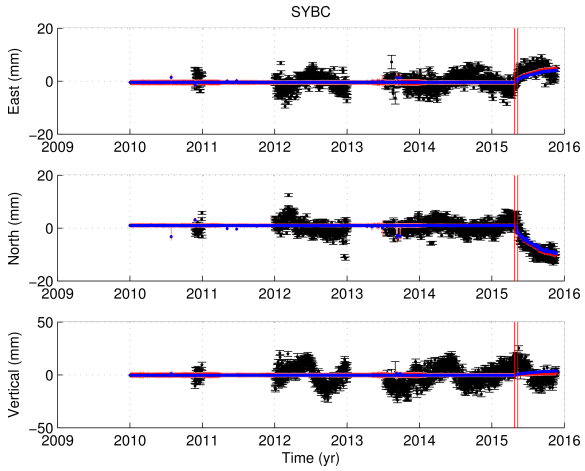
(d) PYUT

Figure S22: As Figure S21, but for stations JMSM, KKN4, NAST, and PYUT.



(a) RMTE

(b) SNDL



(c) SYBC

Figure S23: As Figure S21, but for stations RMTE, SNDL, and SYBC.

References

G.W. Bohrnstedt and A.S. Goldberger. On the Exact Covariance of Products of Random Variables. *Journal of the American Statistical Association*, 64(328):1439–1442, 1969.

J.F. Cardoso. High-order contrasts for Independent Component Analysis. *Neural Computation*, 11: 157–192, 1999.

K. Chan, T.-W. Lee, and T.J. Sejnowski. Variational Bayesian Learning of ICA with Missing Data. *Neural Computation*, 15(8):1991–2011, 2003.

R.A. Choudrey. *Variational Methods for Bayesian Independent Component Analysis*. Pattern analysis and machine learning - robotics research group, University of Oxford, 2002.

R.A. Choudrey and S.J. Roberts. Variational Mixture of Bayesian Independent Component Analyzers. *Neural Computation*, 15(1):213–252, 2003.

Galetzka et al. Slip pulse and resonance of Kathmandu basin during the 2015 Mw 7.8 Gorkha earthquake, Nepal imaged with geodesy. *Science*, 349(6252):1091–1095, 2015. doi: 10.1126/science.aac6383.

A. Gualandi, E. Serpelloni, and M.E. Belardinelli. Blind source separation problem in GPS time series. *J. Geodesy*, pages 1–19, 2015. doi:10.1007/s00190-015-0875-4.

A. Hyvärinen and E. Oja. A fast fixed-point algorithm for Independent Component Analysis. *Neural Computation*, 9(7):1483–1492, 1997.

A. Mao, C.G.A. Harrison, and T.H. Dixon. Noise in GPS coordinate time series. *Journal of Geophysical Research: Solid Earth*, 104(B2):2797–2816, 1999. doi: 10.1029/1998JB900033.

H. Perfettini and J.P. Avouac. Postseismic relaxation driven by brittle creep: a possible mechanism to reconcile geodetic measurements and the decay rate of aftershocks, application to the Chi-Chi earthquake. *Journal of Geophysical Research*, 109(B02304), 2004. doi:10.1029/2003JB002488.

V. Stevens and J.P. Avouac. Interseismic coupling on the main Himalayan thrust. *Geophysical Research Letters*, 42(14):5828–5837, 2015.

J.R. Taylor. *An Introduction to Error Analysis - The study of uncertainties in physical measurements*. University Science Books, Sausalito, California, 2nd edition, 1982.

S.D.P. Williams. Offsets in global positioning system time series. *Journal of Geophysical Research*, 108(B6), 2003. doi:10.1029/2002JB002156.