Text S1.

Comparison of simultaneous and iterative decompositions

In principle it is necessary to solve for all components simultaneously to obtain the optimal rank-k approximation of the data matrix (note that this is not true for standard or weighted singular value decomposition). However, we ran into the complication of different components almost exclusively explaining separate datasets (Fig. **??**), annihilating the advantages of a joint inversion. The temporal functions for most components are discontinuous precisely at the epochs corresponding to the SAR images (Fig. **??**). To overcome that problem we solve for each component iteratively. This approach does not yield an optimal rank-k approximation of the data matrix, but it helps force each component to explain a significant fraction of all datasets. The procedure runs as follows:

1. Set X equal to the original data matrix.
2. Set $X_1 = X$.
3. Calculate the best rank-1 (1-component) model $\hat{X}_i$ of the data matrix X_i
4. Set $X_{i+1} = X_i - \hat{X}_i$
5. Return to step 3 unless termination condition has been reached.

The matrices $\hat{X}_i$ are the outer product of our spatial and temporal functions for the ith component. Due to the variation of optimal value of components with the total number of components (unlike in SVD, where the best ith component is independent of how what rank-k approximation is sought), the resultant principal components are close to but not exactly orthogonal. The PCAIM approach is valid for any linear combination of components regardless of orthogonality, even though mathematically it may not be optimal in terms of matrix approximation.

Comparison of InSAR-only, EDM-only and InSAR + EDM joint inversions

We demonstrate here the benefit of the joint inversion of InSAR and EDM data, as compared to InSAR-only and EDM-only inversions. After the PCA decomposition, both the joint and the

InSAR-only decomposition yield similar spatial functions for the 1st and 2nd components (Fig. **??**). By contrast the spatial functions of the 3rd and higher order components are quite different. Since the EDM data can be mostly reconstructed with only 2 components (notice how small the principal vectors are in the 3rd component of the joint decomposition), in the joint decomposition most of the InSAR signal that is not spatially coherent with the EDM data is taken account by the 3rd and higher order components. This leaves basically noise to higher components (in particular the tropospheric effects and the tectonic signals not visible in the EDM data). In the InSAR-only decomposition, without the guidance of EDM data, the spatial function can be *any* signal with coherent time history, and therefore what is extracted can be the coherent portion of atmospheric noises.

With regard to the time function, the joint and EDM-only decomposition yield very similar results for the first 3 components, whereas the joint and InSAR-only decomposition yield quite different results from the 2nd component and on. Again, since we assume that EDM data is better corrected for tropospheric noises and sampled in time, the incorporation of EDM data becomes a necessity because InSAR by itself cannot offer reliable time evolution history.

At the decomposition stage it is clear that with our iterative approach the time functions are mostly constrained by the EDM data while the spatial functions are mostly controlled by the InSAR data. When it comes to source modeling, the benefit of using InSAR data becomes explicit in that it provides much better spatial constraints than the EDM data. If we consider only the EDM data, the inversion problem is highly underdetermined given the chosen gridded source of magmatic inflation (more than 6000 point sources as compared with 8 observations from the spatial function of each component). The inversion works but the resulting model depends heavily on the regularization. In this relatively simple magmatic inflation example, one can certainly use a source model defined with fewer adjustable parameters, such as the traditional single point source of inflation (Mogi) or the prolate ellipsoid model of inflation (*Yang et al.*, 1988), so that including InSAR data does not make huge difference. In a more complicated example such as the slip model on a fault plane, involving InSAR data greatly helps improve the resolution of the slip pattern on different fault patches.

The joint and InSAR-only inversions give similar source model. The best-fit results yield reduced

Chi-square (χ_r^2) of 0.96 and 0.90 for the 1st and 1st+2nd component joint inversion respectively, and 1.07 and 0.89 for the 1st and 1st+2nd component InSAR-only inversion. However, as seen in the F-test (Table S1), with InSAR data only it is difficult to determine the cut-off component to be used in inversion, and we lose the temporal resolution inherent in the EDM data. The time evolution obtained from the InSAR-only inversion is coarser and partly biased by atmospheric effects which are present in all the components rather than being rejected in the higher order components as happens in the joint inversion.

References

Yang, X.-M., P. M. Davis, and J. H. Dieterich (1988), Deformation from inflation of a dipping finite prolate spheroid in an elastic half-space as a model for volcanic stressing, *J. Geophys. Res.*, *93*, B5, 4249–4257.

Table S1. F-test determination of the number of significant components needed for the InSAR and the EDM data based on our iterative decomposition and gridded source inversion

	Decomposition		Model Inversion	
	InSAR	EDM	InSAR	EDM
$\mathbf{P}(F \geq F_{1,2})^*$	$\sim 0^{**}$	$\sim 0^{**}$	$\sim 0^{**}$	$\sim 0^{**}$
$\mathbf{P}(F \geq F_{2,3})$	$\sim 0^{**}$	$5.2 \times 10^{-3**}$	$\sim 0^{**}$	0.06
$\mathbf{P}(F \geq F_{3,4})$	$\sim 0^{**}$	0.85	$\sim 0^{**}$	0.81
$\mathbf{P}(F \geq F_{4,5})$	$\sim 0^{**}$	2		
$\mathbf{P}(F \geq F_{5,6})$	$\sim 0^{**}$	2		
$\mathbf{P}(F \geq F_{6,7})$	$\sim 0^{**}$	2		

$^*F_{\mathrm{i,i+1}}$ refers to the test between using the first i and the first $i+1$ components. When probability $\mathbf{P}(F \geq F_{\mathrm{i,i+1}})$ is smaller than 0.05, we consider this test as statistically significant, which allows us to make the claim that the incorporation of the $(i+1)$th component does improve the fit significantly at the 95% confidence level (the probability that the improvement would be due to pure chance is less than 5%).

**significant; $\sim$0 indicates probability less than 10^{-323}