re

Final Report

SCEC Project 09095: Model-Based Geodetic Transient Detection

Clifford Thurber, Principal Investigator
thurber@geology.wisc.edu

Kurt L. Feigl, Co-Principal Investigator
feigl@wisc.edu

Department of Geoscience
University of Wisconsin-Madison
1215 W. Dayton St.
Madison, WI 53706

Date of Report:	2010-Mar-31/ 12:13
Amount of Award:	$20,000 (1 year)
Proposal Category:	B. Integration and Theory
SCEC3 Science Priority Objectives:	A5
Disciplinary Activities:	Tectonic Geodesy
Interdisciplinary Focus Areas:	Crustal Deformation Modeling

Model-Based Geodetic Transient Detection

Project Summary

A research project has been completed at the University of Wisconsin-Madison by graduate student Summer Ohlendorf (a former SCEC intern) under the joint supervision of Professors Clifford Thurber and Kurt Feigl. The project evaluated a model-based approach to the detection of transient signals in geodetic time series. The basic strategy estimates the distribution of slip on the fault surface from geodetic data. Our approach provides two measures for detecting possible transients: changes in the slip model parameters and changes in the data fit. To validate this approach, we consider the transient postseismic signal observed at Parkfield, California, as recorded by continuous GPS in the years following the 2004 earthquake. We analyze how the two statistical measures vary with the duration of the time interval spanned by the GPS measurements. The signal-to-noise ratio of the total estimated slip to its uncertainty could, in principle, be used as a criterion to assess the significance of a transient signal.

Data and Method

The postseismic transient slip behavior along the San Andreas fault (SAF) following the Parkfield main shock on 28 September 2004 has been well documented (Johanson et al., 2006; Murray and Langbein, 2006; Langbein et al., 2006; Lienkaemper et al., 2006). Since most of the postseismic transient occurs during the first 230 days following the main shock, we restrict our analysis to the time interval beginning 29 September 2004 and ending 17 May 2005. We are particularly interested in determining in the length of the measurement interval required to yield a reliable detection of the transient. Since the postseismic transient decays with time, we adopt the latter date as our "stop time" and then analyze data spanning progressively longer intervals extending earlier (nearer to the date of the main shock). Accordingly, we denote the time interval in days as

$$\Delta t_i = 230 - t_i \tag{1}$$

Since shorter time intervals will contain less deformation than longer time intervals, we interpret our results in terms of the algorithm's ability to detect a small transient. In other words, the "detection threshold" will depend on the length of the shortest time interval that yields a change that is significantly different from zero by at least one of the following two criteria: (1) the misfit of model to data, as measured by the chi-squared statistic; and/or (2) the confidence in the estimated fault slip, as measured by the ratio of the estimated total slip divided by its uncertainty.

Following this strategy, we have modified the existing geodetic inversion algorithm of Murray and Langbein (2006). For a unit slip on each discrete element of the fault, the forward problem calculates the resulting displacement vector at each GPS station on the Earth's surface. This information is assembled in a matrix of Green's functions dimensioned to account for 489 fault elements and 13 GPS stations. The free parameters include one component (horizontal strike-slip) of slip for each fault element. The data include the three components of displacement (with respect to reference station CRBT) and their scaled uncertainties estimated from the GPS position time series over the relevant time interval (Figure 1). In addition, the data from the USGS alinement [sic] arrays (Lienkaemper et al., 2006) are also included, as in Murray and

Langbein (2006). The fault is parameterized using a triangular mesh (Figure 2). The elastic properties are assumed to be uniform throughout the half space with a Poisson's ratio equal to 0.25. To regularize the problem, a smoothing operator is imposed, as discussed in Maerten et al. (2005). The optimal weighting value of this smoothing operator is selected by cross-validation, as shown in Figure 3. In addition, the estimated slip values must all be right-lateral. This constraint is imposed by the non-negative least squares algorithms implemented in the Matlab routines NNLS, LSQLIN, or LSEI_ML (Lawson and Hanson, 1974; Hanson and Haskell, 1982). To describe long-term interseismic motion, three of the fault elements are constrained to slip at constant rates of 27 ± 1.5 mm/yr for the creeping section of the SAF northwest of Parkfield, 1 ± 1 mm/yr for the locked section of the SAF southeast of Parkfield, and 32.6 ± 1 mm/yr below the seismogenic depth of 14 km.

To evaluate the misfit of the model to the data, we calculate, for each time interval, a chi-squared statistic

$$\chi^2\left(\Delta t_i\right) = \sum_{j=1}^{3N(\text{GPSstations})} \left(\frac{u_j^{(\text{obs})} - u_j^{(\text{model})}}{\sigma_j} \right)^2_{\Delta t_i = 230 - t_i} \tag{2}$$

where the "observed" displacement $u^{(obs)}$, its uncertainty σ, and the modeled displacement $u^{(model)}$ have each been estimated from the GPS time series over the time interval Δt_i. As shown in Figure 4, the misfit does not exhibit any simple relationship with the time interval. Accordingly, a detection criterion based on misfit alone would not be sensitive to transient signals.

As an alternative to a criterion based on the misfit of the model to the data, we consider a criterion based instead on the estimated model parameters and their uncertainties. To find the uncertainty of the estimated slip parameters, we have explored three options: jack-knife (Tichelaar and Ruff, 1989), boot-strap (Efron and Tibshirani, 1986) and a constrained least-squares algorithm called LSEI (Hanson and Haskell, 1982). The three approaches yield different results for estimated total slip $U^{(est)}$ (summed over all elements on the fault) and its uncertainty σ_U, as shown by the dashed lines in Figure 5.

The jackknife approach finds large values (± ~ 5 m) for the uncertainty σ_U of the total slip. In calculating these values, we applied equation (5) in Tichelaar and Ruff (1989)

$$\hat{\sigma}_{JACK} = \left[\frac{k-p+1}{n-k} \sum_i w_i^* \left(\hat{\theta}_i^* - \tilde{\theta}\right)^T \left(\hat{\theta}_i^* - \tilde{\theta}\right) \right]^{1/2} \tag{3}$$

where θ is the model parameter estimated from various jackknife subsets of the data, $p = 489$ is the number of parameters, n is the number of data, and k is the size of each jackknife subset. To evaluate this expression, we first remove the three displacements for each GPS station in turn, thus setting the subset size to $k = n - 3$. Second, we set the "weights" w^*_i to unity. Third, we set n to the total number of equations (582) to account for the constraint equations. As a result, the variance factor $(k - p + 1)/(n - k)$ is approximately 20. This value may be too large because it treats the constraint equations as independent. On the other hand, this value is of the same order of magnitude as the chi-squared χ^2 statistic (20 to 40). The latter value can also be used to account for the number of degrees of freedom in constrained solutions (Aster et al., 2005). The

bootstrap algorithm yields small uncertainties (black crosses labelled "boot" in Figures 4-6), presumably because it under-estimates the uncertainties of the observed displacements. Although the diagonal elements of the data covariance matrix have been scaled to reflect scatter in the residuals, the off-diagonal elements are assumed to be null, neglecting correlations between the displacements at neighboring stations.

The LSEI algorithm (blue circles labeled "lsei" in Figures 4-6), yields uncertainties for the total slip that decrease with increasing duration of the time interval Δt_i, as expected. The LSEI is the only one of the three approaches to exhibit this (useful) attribute. In other words, the LSEI algorithm returns a statistic that is a sensitive measure of uncertainty.

To assess the confidence of the estimated model parameters, we calculate, for each time interval, the ratio of the "signal" (total slip $U^{(\mathrm{est})}$) to the "noise" (uncertainty σ_U) as

$$R = \frac{U^{(\mathrm{est})}}{\sigma_U} = \frac{\sum_{i=1}^{p} u_i^{(\mathrm{est})}}{\sqrt{\sum_{i=1}^{p} \sigma_i^2}} \tag{4}$$

The value of the signal-to-noise ratio R of the model estimate of the total slip $U^{(\mathrm{est})}$ to its uncertainty σ_U from equation (4) as a function of time interval Δt_i increases with increasing duration of the time interval, as shown in . The LSEI estimates (blue circles) increase from $R \sim 10$ for a short 30-day interval to $R \sim 20$ for a longer 130-day interval. Accordingly, this measure could be used as a criterion for detection.

Conclusions

The signal-to-noise ratio R of the total estimated slip to its uncertainty increases with the duration of the time interval. Although the three different approaches for calculating the uncertainty yield results that vary widely, each could be used individually to compare different time intervals. This ratio could, in principle, be used as a criterion to assess the significance of a transient signal such as the postseismic deformation following the 2004 Parkfield earthquake. Detecting transient phenomena on earthquake-capable faults is a high-priority goal for SCEC, appearing as objective A5 to "develop a geodetic network processing system that will detect anomalous strain transients." Our work has contributed directly towards this high priority objective by demonstrating the feasibility of a novel approach for transient detection. As a next step, this approach could be applied to the SCEC Community Fault Model (CFM) using GPS data from southern California. Toward this end, our revised Matlab source code, including the uncertainty analyses, has been returned to its original author, Jessica Murray-Moraleda at the USGS in Menlo Park. She deserves our thanks for generously sharing the code and taking the time to help us with it.

References

Aster, R., B. Borchers, and C. Thurber, *Parameter Estimation and Inverse Problems*, Elsevier/Academic Press, 296 pp., 2005.

Efron, B., and R. Tibshirani (1986), Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy, *Statistical Science,* 1, 54-77.

Hanson, R., and K. Haskell (1982), Algorithm 587: two algorithms for the linearly constrained least squares problem, *ACM Transactions on Mathematical Software* (TOMS), 8, 333; http://doi.acm.org/10.1145/356004.356010.

Johanson, I. A., E. J. Fielding, F. Rolandone, and R. Burgmann (2006), Coseismic and postseismic slip of the 2004 Parkfield earthquake from space-geodetic data, *Bull. Seism. Soc. Am.,* 96, S269-S282.

Langbein, J., J. R. Murray, and H. A. Snyder (2006), Coseismic and initial postseismic deformation from the 2004 Parkfield, California, earthquake, observed by Global Positioning System, electronic distance meter, creepmeters, and borehole strainmeters, *Bull. Seism. Soc. Am., 96,* S304-320.

Lawson, C. L., and R. J. Hanson (1974), *Solving least squares problems*, Prentice-Hall.

Lienkaemper, J. J., B. Baker, and F. S. McFarland (2006), Surface Slip Associated with the 2004 Parkfield, California, Earthquake Measured on Alinement Arrays, *Bull. Seism. Soc. Am., 96*, S239-249.

Maerten, F., P. Resor, D. D. Pollard, and L. Maerten (2005), Inverting for slip on three-dimensional fault surfaces using angular dislocations, *Bull. Seism. Soc. Am.* 95, 1654-1665, doi:10.1785/0120030181.

Murray, J. R., and J. Langbein, Slip on the San Andreas fault at Parkfield, California over two earthquake cycles and the implications for seismic hazard, *Bull. Seism. Soc. Am*. 96, doi:10.1785/0120050820, S283-S303, 2006.

Tichelaar, B., and L. Ruff (1989), How good are our best models?; jackknifing, bootstrapping, and earthquake depth, *EOS Transactions AGU*, 70, 593.

Figures

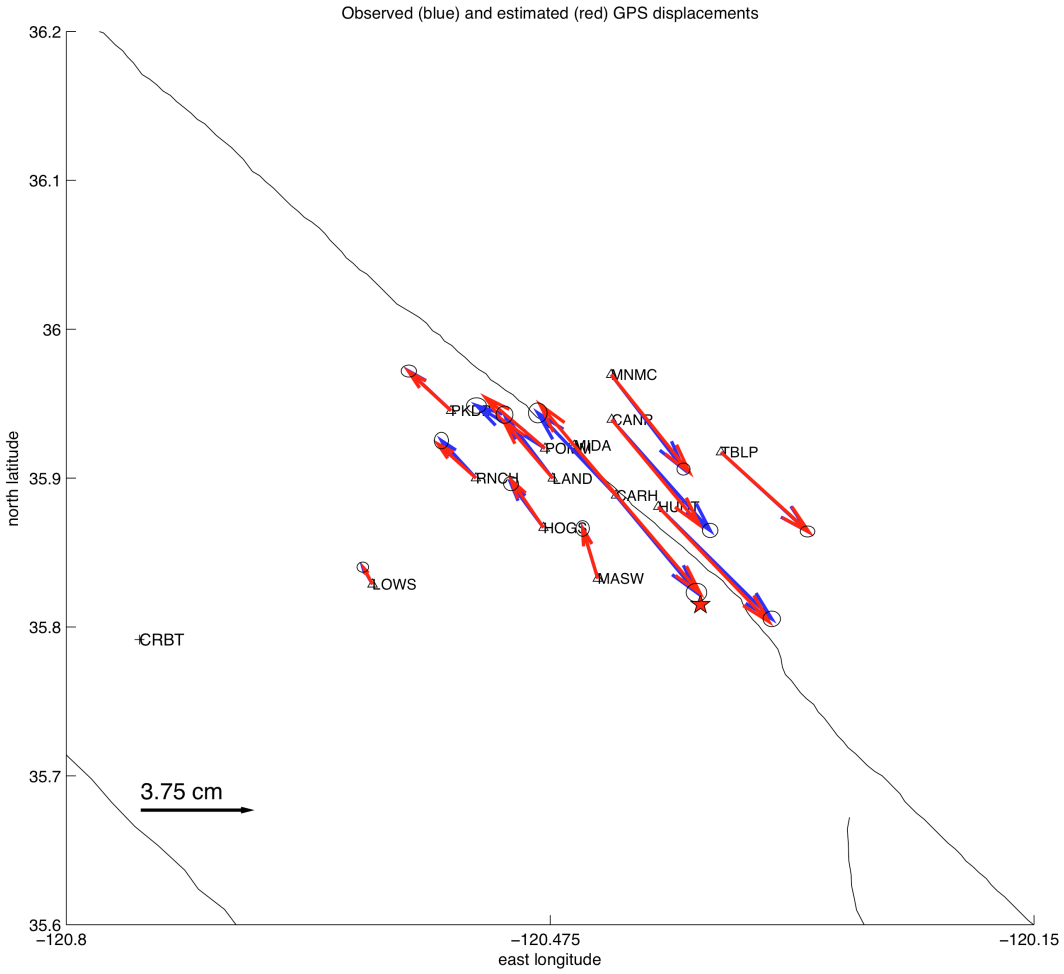
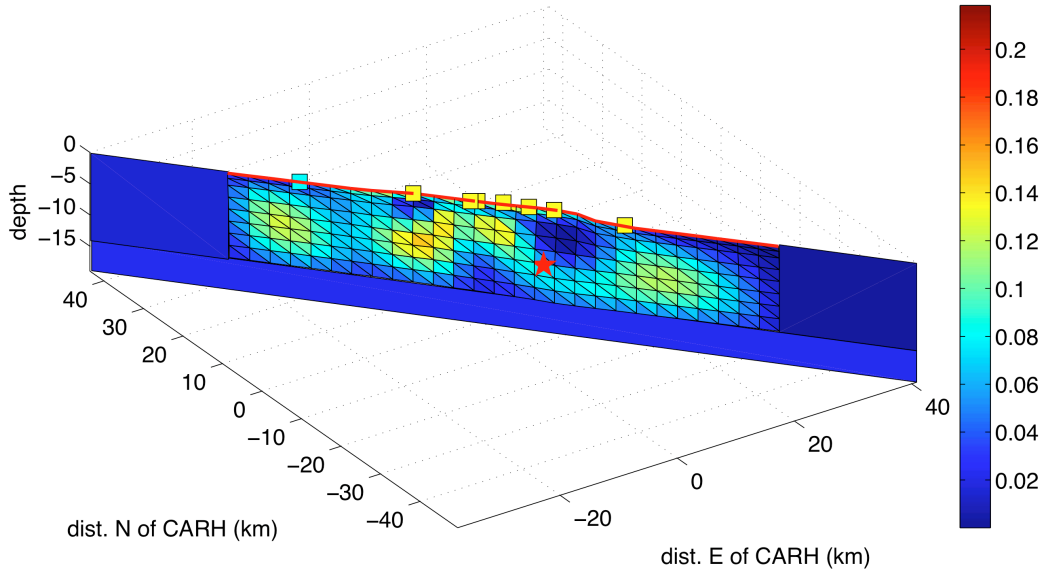


Figure 1. Example of displacement vectors for the full 230-day long postseismic time interval beginning 29 September 2004 and ending 17 May 2005, showing observed displacement (blue) and modeled displacements (red arrows). Ellipses include the area of 95 percent confidence, after scaling by the estimated variance factor.



slip full post allstn.pdf /mnt/s22/summer/Parkfield GPS/Interseismic 2010-03-30 11:13:58 summer

Figure 2. Slip distribution estimated from the GPS data for the full 230-day long postseismic time interval beginning 29 September 2004 and ending 17 May 2005. This 3-dimensional representation shows the flexible triangular mesh for representing a complex fault surface, as developed by Murray and Langbein (2006). The boxes along the earth's surface denote the locations of the USGS alinement [sic] arrays (Lienkaemper et al., 2006).

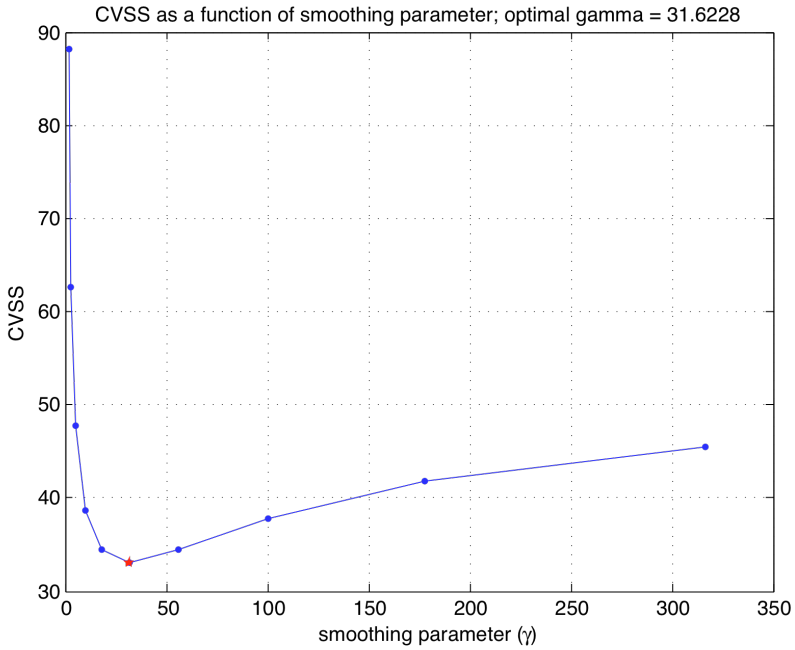
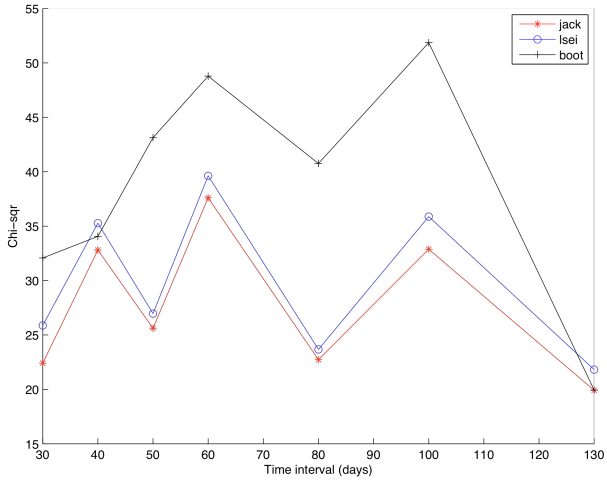
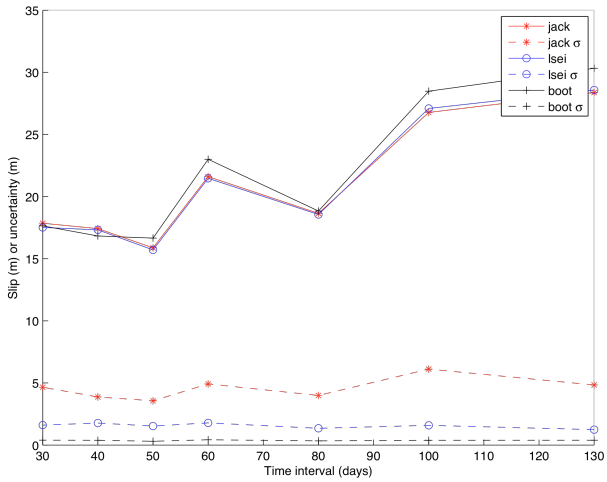


Figure 3. Optimal value of smoothing parameter (horizontal axis) is identified by the smallest value of the cross-validation sum of squares CVSS (vertical axis). The star denotes the optimal value of the smoothing parameter $\gamma = 31.62$ used in all subsequent runs.



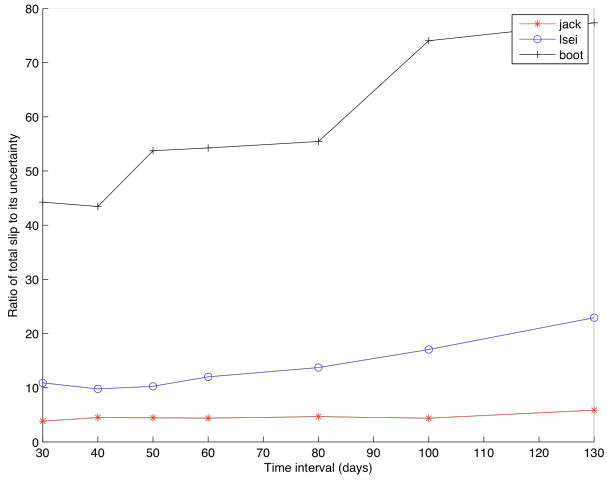
chiVdt.pdf /usr1/feigl/parkfield/Interseismic 2010-03-30 16:32:16 feigl

Figure 4. Misfit of modeled to observed displacement, as measured by chi-squared statistic χ^2, as a function of time interval, as calculated in equation (2). The three different curves reflect different schemes for estimating the model parameters and their uncertainties: jackknife (red curve with asterisks labeled "jack"), constrained least squares (blue curve with circles labeled "lsei"), and bootstrap (black curve with crosses labeled "boot").



dt.pdf /usr1/feigl/parkfield/Interseismic 2010-03-30 16:32:18 feigl

Figure 5. Estimated total slip $U^{(est)}$ (solid lines) and its uncertainty σ_U (dashed lines) as a function of time interval.



snrVdt.pdf /usr1/feigl/parkfield/Interseismic 2010-03-30 16:32:19 feigl

Figure 6. Model uncertainty, as measured by the "signal to noise" ratio R of the total slip $U^{(est)}$ to its uncertainty σ_U from equation (4) as a function of time interval.