

# Automatic Decomposition of Geodetic Time Series for Studies of Ground Deformation

Thesis by  
Bryan Riel

In Partial Fulfillment of the Requirements for the  
degree of  
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY  
Pasadena, California

2017  
Defended September 6, 2016

© 2017

Bryan Riel

ORCID: 0000-0003-1940-3910

All rights reserved

## ACKNOWLEDGEMENTS

It is hard to put into words the gratitude I feel towards so many people who have helped me along the way to reach the end of this degree. For me, the people that I have met and the experiences we shared hold far more significance than the actual research I have done. What I will say here will not even come close to expressing my gratitude, but I will give it a shot anyways. Before I start, I would like to acknowledge financial support from the NASA Earth and Space Science Fellowship with grant number NNX12AN65H.

Firstly, I thank my research advisor, Mark Simons. From day one, he understood my strengths and suggested a research topic tailored to my interests. I came into Seismolab at Caltech with zero knowledge of geophysics, and his support and guidance has been invaluable for keeping my head above water. It was once said that Mark is “practical”. When I first heard that, the full significance of that statement eluded me. Now, I have come to understand how important practicality is, as well as how limited that word is to describe the ideas that Mark has on a day-to-day basis.

I also thank the members of my thesis advisory committee: Jennifer Jackson, Pablo Ampuero, Victor Tsai, and Rob Clayton. Your guidance and advice, ranging from formal TAC meetings to “aha” moments in the hallway, have been very useful. Every faculty member in Seismolab has been inspirational in one way or another, and I am very lucky to have shared the hallways of South Mudd with them.

Before I came to Caltech, I pursued a Masters degree in aerospace engineering at the University of Texas at Austin. I would not be where I am without the help and guidance of my advisor there, Sean Buckley. Sean was the first person to introduce me to remote sensing, and he was instrumental in setting me along the career path I am now on. His tireless work in finding the best opportunities for his students cannot be overstated. While at UT, I was introduced to the airborne lidar group at the Center for Space Research where I worked with Roberto Gutierrez and Amy Neuenschwander. Roberto was a great supervisor and a great friend. The afternoons we spent discussing life and its intricacies will never be forgotten, nor will the hours spent in a small Cessna flying over sand dunes and coastlines.

Through Sean, I was also introduced to the radar group at the Jet Propulsion Laboratory (JPL). The enthusiasm, kindness, and talent of the people there made the summer of 2009 one of my favorite summers ever. Thank you Marc, Naiara, Yunling,

Brian, Sarah, Ian, Kean, Yang, and many others for your friendship and support. I look forward to working with you all again.

At Caltech, the Mark Simons research group (both old and new members) has been instrumental in providing critical guidance and technical knowledge throughout the years. It was a joy to collaborate with Romain Jolivet and Zacharie Duputel, who are both well on their way to standout careers in France. Piyush Agram was, and still is, the all-things-technical guru who has saved me on many occasions and provided me with countless coding and mathematical tips. I also owe thanks to Francisco, Junle, Hilary, Sarah, Heresh, Nina, and others for their help along the way. I would also like to extend my gratitude to the administrative staff, whom I consider to be my anchor to the real world and are responsible for keeping Seismolab running and instilling a good mood to anyone lucky enough to ask them for help. Thank you Donna, Rosemary, Kim, Sarah, and Priscilla.

Of course, my friends at Caltech have been a very important fixture these past few years. My officemates Daniel Bowden, Asaf Inbal, and Brent Minchew were unfailingly good sources of laughter and syntax knowledge. In particular, Brent Minchew has been a close friend since my time at UT and has provided support, life lessons, and a healthy dose of sarcasm like no one else can. Stephen Perry and Semechah Lui have been close friends, study partners, and traveling companions since we first started at Caltech, and I look forward to extending our friendship far into the future. I would also like to thank my roommates Chris Rollins and Tobias Bischoff. Our travels, discussions in cars, and discussions in the living room speak for themselves and set an example for what roommates should be.

Finally, I thank my family in Texas for their neverending support and love. They fill my life with laughter, and they make each step forward in life ten times easier. They say everything is bigger in Texas. When it comes to the love they send my way every day, I wholeheartedly agree.

## ABSTRACT

Geodetic measurements of surface deformation have been used for several decades to study how the Earth's surface responds to a wide range of geophysical processes. Geodetic time series acquired over a finite spatial extent can be used to quantify the time dependence of surface strain for a wide range of spatial and temporal scales. In this thesis, we present a new method for automatically decomposing geodetic time series into temporal components corresponding to different geophysical processes. This method relies on constructing an overcomplete temporal dictionary of reference functions such that any geodetic signal can be described by a linear combination of the functions in the dictionary. By solving a linear least squares problem with sparsity-inducing regularization, we can limit the total *number* of dictionary elements needed to reconstruct a signal. In Chapter 2, we present the development of this method in the context of transient detection, where we define transient deformation as nonperiodic, nonsecular accumulation of strain in the crust. The sparsity regularization term automatically localizes the dominant timescales and onset times of any transient signals. We apply this method to Global Positioning System (GPS) data for a slow slip event in the Cascadia subduction zone while incorporating a spatial weighting scheme that filters for spatially coherent signals. In Chapter 3, we use a combination of unique space geodetic measurements and seismic observations to study the 2014 collapse of Bárðarbunga Caldera in Iceland associated with a major eruption event. The eruption sequence, which involved deflation of a magma chamber underneath the caldera and emplacement of a dike leading to lava flow, resulted in rapid subsidence of the glacier surface overlying the caldera and wide-scale ground deformation encompassing the rift zone associated with the dike emplacement. We present a model of the collapse that suggests that the majority of the observed subsidence occurs aseismically via a deflating sill-like magma chamber. In Chapter 4, we extend upon the transient detection framework presented in Chapter 2 to study complex surface deformation over groundwater basins near Los Angeles, California. We develop a distributed time series analysis framework based on the sparse estimation techniques of Chapter 2 and apply it to an 18-year interferometric synthetic aperture radar (InSAR) time series covering the Los Angeles area. We compare long- and short-term ground deformation signals to hydraulic head data from monitoring wells to understand the mechanical link between pressure variations in subsurface aquifers and observed ground deformation.

## PUBLISHED CONTENT AND CONTRIBUTIONS

1. Riel, B., Simons, M., Agram, P. & Zhan, Z. Detecting transient signals in geodetic time series using sparse estimation techniques. *Journal of Geophysical Research: Solid Earth* **119**. doi:10.1002/2014JB011077 (2014).  
B.R. developed the algorithm for the transient detection, performed synthetic experiments, analyzed the algorithm performance on GPS data, and wrote the manuscript.
2. Riel, B., Milillo, P., Simons, M., Lundgren, P., Kanamori, H. & Samsonov, S. The Collapse of Bárðarbunga Caldera, Iceland. *Geophysical Journal International* **202**. doi:10.1093/gji/ggv157 (2015).  
B.R. gathered and analyzed the seismic and GPS data, performed the geophysical model inversions, aided in the development of the volcano source model, and wrote the manuscript.

## TABLE OF CONTENTS

Acknowledgements . . . . .	iii
Abstract . . . . .	v
Published Content and Contributions . . . . .	vi
Table of Contents . . . . .	vii
List of Illustrations . . . . .	viii
List of Tables . . . . .	xxii
Chapter I: Introduction . . . . .	1
Chapter II: Detecting Transient Signals in Geodetic Time Series Using Sparse	
Estimation Techniques . . . . .	5
2.1 Introduction . . . . .	5
2.2 Transient Detection . . . . .	7
2.3 Synthetic Example . . . . .	13
2.4 Spatial Sparsity Weighting . . . . .	21
2.5 Slow Slip Events in Cascadia . . . . .	23
2.6 Discussion . . . . .	32
2.7 Conclusions . . . . .	35
Chapter III: The collapse of Bárðarbunga caldera, Iceland . . . . .	40
3.1 Introduction . . . . .	40
3.2 Geodetic and seismic data . . . . .	42
3.3 Source models for rift zone and magma chamber . . . . .	44
3.4 Discussion . . . . .	46
3.5 Conclusions . . . . .	50
3.6 Supplementary: Materials and Methods . . . . .	52
Chapter IV: Quantifying Ground Deformation in the Los Angeles and Santa	
Ana Coastal Basins due to Groundwater Pumping . . . . .	64
4.1 Introduction . . . . .	64
4.2 Background . . . . .	67
4.3 Groundwater Level and Ground Deformation Time History . . . . .	70
4.4 Central and Santa Ana Basin InSAR Time Series . . . . .	80
4.5 Simultaneous Time Varying Seasonal and Long-Term Ground De-	
formation Using InSAR Time Series Analysis . . . . .	84
4.6 Discussion . . . . .	93
4.7 Conclusions . . . . .	100
4.A Appendix A: Spatiotemporal InSAR Time Series Analysis . . . . .	101
4.B Appendix B: Supplemental Figures . . . . .	109
Chapter V: Conclusions . . . . .	114

## LIST OF ILLUSTRATIONS

<i>Number</i>	<i>Page</i>
2.1 Uniform integrated B-splines ( $B^i$ -splines) of various timescales used as candidate temporal displacement functions. The characteristic timescales for the $B^i$ -splines are determined by dividing the time span into uniformly spaced intervals. The filled circles denote the interval spacings for each $B^i$ -spline and differ by factors of two. . . . .	7
2.2 Individual model components input into the synthetic GPS time series. The seasonal signals are constructed using annual and semi-annual periods while the transients are formed using arctangent functions of various amplitudes and durations. Additionally, white and colored noise are added to the model to simulate measurement noise typically found in geodetic time series. . . . .	14
2.3 Convergence test of the Gibbs sampler for coefficients corresponding to several $B^i$ -splines. The mean of the samples drawn from the posterior distribution for each $B^i$ -spline is monitored for increasing number of samples. The means are normalized by the mean computed using $2e^5$ samples. After $O(7e^4)$ samples, the means are within 0.2% of their final values. . . . .	15
2.4 $K$ -fold cross-validation results for selection of the penalty parameter, $\lambda$ , for the synthetic time series. Fourteen separate cross-validation experiments were performed for fourteen independent subsets of the time series. The solid blue line shows the mean testing error for all experiments while the shaded region denotes the standard deviation. The vertical black dashed line marks the mean $\lambda$ as determined by the Gibbs sampler which shows relatively good agreement with the optimal $\lambda$ obtained from the cross-validation (red circle). . . . .	15

- 2.5 Reconstructed time series for the synthetic data with the simultaneously estimated seasonal and secular signal removed. (a) Full transient signal for models corresponding to different values of the penalty parameter  $\lambda$ . The black circles show the input data after removing the true secular and seasonal signals. Lower values of  $\lambda$  correspond to rougher models. The model for  $\lambda = 10$  and the point estimate derived from the Gibbs samples for the same  $\lambda$  are able to nearly exactly reproduce the input signal. (b) The high frequency component of the reconstructed transient signal compared with the input colored noise. The  $\lambda = 0.66$  model (chosen through cross-validation) is able to capture nearly all of the temporally correlated noise structure while higher penalties result in smoothing over the higher frequency variations. . . . . 16
- 2.6 Synthetic scalograms showing the amplitudes of the  $B^i$ -splines estimated using three different methods: (a) Tikhonov ( $\ell_2$ -norm) regularization; (b) Sparse ( $\ell_1$ -norm) regularization with  $\lambda = 10$ ; and (c) posterior distribution means derived from the Gibbs samples. Each row of the scalograms correspond to  $B^i$ -splines of a given timescale. Tikhonov regularization results in non-zero amplitudes for nearly all  $B^i$ -splines and tends to spread energy across timescales. The sparsity-promoting regularization methods zero-out nearly all of the  $B^i$ -splines, leaving only those that describe transients in the data. The estimated secular rates for the Tikhonov, sparse regularization, and Gibbs solutions are 12.0, 9.7, and 8.6 mm/yr, respectively (for an input secular rate of 10 mm/yr). . . . . 18

- 2.7 Two-dimensional histograms of samples drawn from posterior distributions for different pairs of  $B^i$ -splines. (a) When two  $B^i$ -splines have nearly coincident time centroids and are both able to reconstruct a transient signal, their joint posterior exhibits a strong negative covariance. (b) For adjacent  $B^i$ -splines of the same timescale where the centroid of the true signal lies between the centroids of the  $B^i$ -splines, their amplitudes will co-vary in a positive manner. (c) Long-timescale  $B^i$ -splines and the secular rate will exhibit covariances similar to case (a). Note the sharp corner in the joint distribution due to the  $\ell_1$ -norm penalty. (d) Two  $B^i$ -splines located in a time window where no transient signals occur will have samples located very close to the origin with a structure resembling the prior (bi-variate Laplace distribution). 19
- 2.8 Transient signal recovery using subsampled data with increasing subsampling factors. The limited data resolution of the smallest timescale  $B^i$ -splines requires 3-4 observations to resolve a transient signal. In the case of the smallest synthetic transient around year 10, the temporal spacing of  $\Delta t = 256$  days is too large to resolve the short-timescale duration, and the reconstructed signal is smoothed (bottom panel). However, this result still demonstrates that a positive detection is still possible with temporally coarse time series. . . . . 21
- 2.9 Computed scale length (correlation length) of the GPS network used in the SCEC Phase III validation exercise. Scale length is computed using the average distance from each station to the nearest three stations. 24
- 2.10 Reconstructed transient signal (red arrows) corresponding to simulated thrust event (blue arrows). Most of the signal is isolated close to the fault patches, although a few stations outside of the basin region show extraneous transients. The large east error at station DSHS is due to a large time-correlated noise signal. . . . . 24
- 2.11 Reconstructed north component transient time series for SCEC Phase III Set D stations near the Santa Monica fault. The data (dots) have the estimated secular and seasonal signals removed to match the reconstructed signal (bold line). Dashed black lines show the true ground signal for each station. Without spatial weighting (left), the transient signal is corrupted by time-correlated colored noise. With spatial weighting (right), the signal from the thrust event is isolated. . 25

- 2.12  $K$ -fold cross-validation results for selection of the penalty parameter  $\lambda$  for the east-component ALBH time series. Cross-validation was performed on three independent, equally-sized subsets. The blue line shows the mean testing error and the shaded region denotes the standard deviation. . . . . 26
- 2.13 Transient detection results for the east-component ALBH time series. (a) GPS data with estimated seasonal and secular signals removed (black circles) and estimated transient signal from a reduced dictionary corresponding to a 99% variance reduction (red line). Shaded area denotes  $5\text{-}\sigma$  uncertainties on predicted displacement using the method in Section 2.2. (b) Posterior data covariance matrix of the time series fit scaled to unity. (c) Scalogram for all  $B^i$ -spline coefficients. Comparison with the time series shows the direct correspondence between the slow slip events and the non-zero 3- and 7-week  $B^i$ -splines. No strong long-term transient signals are present. . . . . 27
- 2.14 Evolution of the posterior covariance with penalty  $\lambda$  between two centroid-coincident  $B^i$ -splines corresponding to the January 2007 slow slip event. Solid and dashed lines mark lines of constant posterior density in the following manner: black dashed  $\rightarrow$  0.03, white dashed  $\rightarrow$  0.2, black solid  $\rightarrow$  0.2, white solid  $\rightarrow$  0.95. For small  $\lambda$ , the joint distribution is diffuse with larger amplitude samples drawn for the shorter timescale  $B^i$ -spline. As  $\lambda$  increases, the distribution shrinks and moves along the vertical axis. Eventually, the distribution transitions to larger values for the longer timescale  $B^i$ -spline and small values for the shorter timescale  $B^i$ -spline. For the largest  $\lambda$ , the samples are tightly clustered around the origin. . . . . 28
- 2.15 Map of select GPS stations from the PANGA network used in the spatial sparsity weighting. . . . . 29
- 2.16 Amplitudes corresponding to 3-month  $B^i$ -splines used to fit east component time series data from Cascadia GPS stations. Stations are arranged by increasing latitude, and spatial sparsity weighting was applied to isolate spatially coherent  $B^i$ -splines. The episodic slow slip events are easily visualized, and longer time scale propagation effects can be seen in two of the events (mid-2009 and late-2010). . . 30

- 2.17 Transient detection with spatial sparsity weighting for 33 GPS stations within the Cascadia region. Three separate slow slip events are shown: (*left*) January 2007; (*middle*) mid-2008; and (*right*) mid-2009. The top panels show the normalized east-component displacements corresponding to the slow slip events. The displacement time series are ordered in distance along a 40 km depth contour. The bottom panels show the transient ground motion during each slow slip event where the marker color indicates days from the start of the event. . . . . 31
- 3.1 Map of northwestern Vatnajökull, Iceland with the locations of known calderas/volcanos indicated by barbed lines: As - Askja, Bá - Bárðarbunga, Gr - Grímsvötn, H - Herðubreið, Kv - Kverkfjöll, Tu - Tungnafellsjökull. Colored fringes represent contours of the line-of-sight (LOS) component of ground motion at 15 cm per color cycle. Areas of high interferometric phase noise have been masked out. The interferogram over the ice was formed from COSMO-SkyMED (CSK) images acquired from Sep. 12-13, 2014 while the 24-day interval interferogram over the ground was formed from RADARSAT-2 (RS2) images acquired on Aug. 8 and Sep. 1, 2014. A clear bullseye pattern over the Bárðarbunga caldera indicates subsidence of the glacier surface. The RS2 interferogram shows the effects of the rifting associated with spreading of the ground away from the active dike. Black arrow indicates the satellite-to-ground direction for both the CSK and RS2 interferograms. White dots indicate earthquakes that occurred between August 15, 2014 and February 1, 2015, as recorded by the SIL network in Iceland. Lower-left inset shows the location of the study area in Iceland with respect to the glaciers and active volcanic zones (dark grey shading). Upper-left inset shows surface rupture near the Herðubreið volcano (denoted by white box) observed in the RS2 interferogram. . . . . 41

- 3.2 (A)-(E) Evolution of subsidence within the caldera observed from five one-day interval CSK interferograms. Note the similarity in LOS directions for the interferograms, with the exception of the Sep. 17-18 interferogram. Colored circles indicate earthquakes that occurred during the time span of the interferogram and are colored by hours elapsed from the acquisition time of the first image of a given pair. Earthquakes with  $M_w > 4.9$  are shown with focal mechanisms derived from moment tensors obtained from the GFZ Potsdam catalog. The location of the transects in Fig. 3.S2 are indicated in the first panel for the August 27-28 interferogram. (F) Distribution of all seismicity near the caldera between August 16 to October 21, 2014. The colored focal mechanisms are colored by days elapsed since August 23, 2014. Yellow dots denote the locations of the earthquakes presented in Fig. 3.1. The black focal mechanisms with vertical tensional axes correspond to selected CLVD events at Bárðarbunga between 1976-1996 [18]. . . . . 43
- 3.3 (A) Difference between Sep. 12-13 and Sep. 13-14 interferograms for isolating the ground deformation associated with a  $M_w$  4.9 earthquake that occurred on Sep. 14. Also shown are synthetic LOS displacements due to slip along a vertical ring fault (B) and a closing crack (C). For the ring fault (thick black line), the fault is placed at a depth of 2 km with a width of 1 km and approximately 3 meters of normal motion. The closing crack is simulated using the model of Fialko, et al. (2001) at a depth of 2 km, a radius of 1 km, and a volume change of  $0.003 \text{ km}^3$ . (D) Seismic potency for the northern and southern halves of the caldera compared with estimated volume change rate ( $dV/dt$ ) within the caldera inferred from the one-day interferograms. The thick blue and red lines correspond to the cumulative potency of the northern and southern regions, respectively, of the caldera. To calculate seismic potency, we divide the seismic moment by a shear modulus of 24 GPa. Thin vertical lines mark the occurrence of earthquakes greater than  $M_w$  4.9. The shaded brown boxes show the time spans of each one-day interferogram and the estimated volume change during that time span (right ordinate). The volume change for each interferogram was computed by integrating the projected vertical displacements within the caldera boundary. . . . . 45

- 3.4 A conceptual model for the mechanics of the collapsing caldera. Magma migration out of the chamber into the dike system causes depressurization of the chamber. The deflating chamber results in subsidence of a coherent block above the chamber, as well as subsidence of the overlying ice. Sustained depressurization of the chamber leads to failure of internal support structures (causing seismic events marked by inward vertical motion and outward horizontal expansion) or rupturing of curved, inward-dipping ring faults. Cross-sections of focal mechanisms representative of these events are shown on the edge of the chamber. . . . . 48
- 3.S1 Ascending and descending one-day CSK interferograms with time spans of Sep. 13-14 and Sep. 17-18, respectively. The two viewing geometries allow for separation of LOS displacements into vertical and horizontal components using the method of Yun et al. (2006), assuming that the two interferograms are measuring roughly the same ground deformation. While the two interferograms appear to measure similar levels of ice subsidence, the earthquake perturbations are more pronounced for the Sep. 13-14 interferogram (see Fig. 3.2). Nevertheless, the ratio of horizontal to vertical displacements does not exceed  $\approx 0.3$  within the caldera. . . . . 53
- 3.S2 Profiles of projected vertical subsidence observed in the five one-day CSK interferograms. Profile locations are shown in Fig. 3.2. Vertical displacements were computed by dividing the line-of-sight range changes by the cosine of the incidence angle for each interferogram. Solid lines are constructed from interferogram data with good coherence while the dashed lines interpolate through areas of low coherence. The gold line shows transects extracted from a 2013 one-day interferogram (Fig. 3.S3), confirming that the current subsidence is directly related to the ongoing magmatic activity. . . . . 54
- 3.S3 One day interferogram of Bárðarbunga caldera and surrounding ice-covered regions formed from CSK images acquired on June 26, 2013 and June 27, 2013. No significant subsidence over the caldera is observed, although LOS displacements due to ice flow out of the east side of the caldera are consistent with the 2014 interferograms. . . . . 54

3.S4 Time spans of interferograms used for estimating the distribution of tensile opening along the dike interface. All interferogram time spans are referenced to Aug. 16, 2014, indicated by the vertical dashed line. The yellow lines correspond to RS2 interferograms while the red lines correspond to CSK interferograms. The orbital directions of the images used for the interferograms are indicated as either ascending (ASC) or descending (DESC). The alternating gray and white columns correspond to the temporal subdomains used in the inversion. . . . . 55

3.S5 Estimated distribution of opening along the dike interface for the time span corresponding to the RS2 interferogram in Fig. 3.1. The synthetic interferogram (see Fig. 3.S6) is shown above the fault, and the white dots correspond to earthquakes located along the dike using the SIL network. . . . . 55

3.S6 Synthetic ground deformation for the elastic model of tensile opening along the dike interface. (A) RS2 interferogram from Fig. 3.1 and GPS displacements with 90% confidence error ellipses used in inversion. (B) Synthetic ground displacements from the estimated model. The modeled LOS displacements include a bilinear ramp to account for orbital errors and long-wavelength deformation. Ellipses represent 90% confidence posterior uncertainties in the modeled GPS displacements. (C) Residual interferogram and GPS displacements after subtracting the synthetic ground displacements. Larger residuals near the surface trace north of Vatnajökull may indicate potential asymmetries due to dike dip not captured in our vertical dike model. (D) Cumulative geodetic potency from the dike model in Fig. 3.S5 for Aug. 16 - Sep. 20 vs. a histogram of the depth distribution of earthquakes located along the dike. . . . . 56

3.S7 (A) Samples for magma chamber depth and radius obtained from an adaptive Metropolis algorithm. Red and blue circles correspond to samples where the prior distribution for the chamber depth has a maximum at 3 and 6 km, respectively. The black dashed lines indicate lines of constant depth-to-radius ratio,  $h$ . (B) Samples for magma chamber pressure (in units of the shear modulus  $\mu$ ) and radius. 57

- 3.S8 Modeling of the source magma chamber using geodetic measurements on ice-free areas. The chamber is modeled as a horizontal circular sill at depth. (A) Unwrapped RS2 interferogram formed from images acquired in an ascending orbit on Aug. 1, 2014 and Sep. 18, 2014, and GPS displacements with 90% confidence error ellipses for VONC and HNIF after removing the estimated signal from the rift zone; (B) unwrapped RS2 interferogram from images acquired in a descending orbit on Aug. 27, 2014 and Sep. 20, 2014; (C) modeled GPS displacements and LOS displacements for the ascending interferogram for a chamber at a depth of 8 km and a radius of 2.3 km (location and size indicated by red dashed circle) or a depth of 4 km and radius of 1.1 km (blue dashed circle); (D) modeled LOS displacements for the descending interferogram; (E) residual between (A) and (C); and (F) residual between (B) and (D). The estimated volume changes were  $0.43 \text{ km}^3$  and  $0.14 \text{ km}^3$  for the ascending and descending interferograms, respectively. . . . . 58
- 3.S9 Seismicity rates for earthquakes occurring within the dike (top) and caldera (bottom). The number of earthquakes are binned in latitude and time where 1-day intervals are used for the time bins. The count is saturated at 15 earthquakes/day for visualization clarity. Bins with less than two earthquakes are colored white. High seismicity rates within the dike for the first 10 days are associated with the initial dike emplacement. While seismicity along the dike has decreased with time, seismicity along the northern rim of the caldera has increased. . . . . 59
- 4.1 Location and tectonic setting of the Los Angeles Central and West Coast Basins and the Santa Ana Coastal Basin. The thick gold lines represent major faults in the area, including the Newport-Inglewood Fault (NIF), the Whittier Fault (WF), the Palos Verdes Fault (PVF), and the Hollywood Fault (HF). The thin black lines indicate county boundaries, and the thick blue line corresponds to the Santa Ana River. The dashed red line indicates the barrier between the forebay and confined areas of the groundwater system modified from estimates by the WRD and OCWD. The inset shows the location of the study area along the California coast. . . . . 67

- 4.2 Distribution of GPS and well data used in this study. The white circles correspond to WRD wells used in this study while the white diamonds correspond to OCWD wells. The black triangles show the GPS coverage in this area. . . . . 71
- 4.3 Hydraulic head time series for OCWD well SAR-9 for selected ports in the shallow, principal, and deep aquifers. The dots represent the raw head data while the solid lines are interpolated data using smoothing splines. The green bars represent the average rainfall in the Orange County area in 3-month intervals from 2000 to 2013. The light gray area indicates the time span for available rainfall data. . . . 71
- 4.4 Depth profile of amplitude and phase delay of SAR-9 hydraulic head data. The blue lines correspond to the average seasonal oscillations while the red lines correspond to the uplift signal initiating in 2004/2005 due to a period of heavy rainfall. The horizontal shaded areas centered on the blue and red lines indicate the uncertainties associated with those values. The vertical shaded areas represent the depths corresponding to the three aquifer layers as estimated by the OCWD three-layer model, and the brown hatched regions indicate the approximate range of depths of the aquitards separating the aquifer layers. . . . . 75
- 4.5 Time to peak seasonal signal for hydraulic head data from well ports located in the principal aquifer. Wells in the center of the basin reach their peak signals earlier in the year compared to wells in the margins of the basin. Inset shows timing of wells along the transect indicated by the A-A' line. . . . . 76

- 4.6 Cross correlation analysis between SAR-9 head data and SACY GPS data. The upper plots correspond to short-term, seasonal signals for both data sets while the lower plots correspond to long-term signals. We iterate over the time series at each port depth, compute the Pearson correlation coefficient between well and GPS data, and estimate the storage coefficient (scaling parameter) to best match the well and GPS data. The right plots show the Pearson correlation coefficient between the time series at each port depth and the SACY GPS time series. The dashed black line corresponds to a third-order polynomial fit to the correlation coefficients, and the red star indicated the location of maximum correlation. For the left plots, SACY data are shown with blue dots and the scaled well data for the depth of maximum correlation are shown solid red lines. The red shaded area represents the uncertainty in the scaling of the well data. . . . . 79
- 4.7 Maps of seasonal peak-to-peak amplitude and phase delay. The dashed black lines show the location of the transects used for the bottom plots. For the lower transect plots, blue lines correspond to the phase delay while the dashed red lines correspond to the peak-to-peak amplitude. The red circle in the map of seasonal phase delay shows the location of a discontinuity in amplitude and phase within the basin. The arrow indicates the satellite-to-ground line-of-sight (LOS) direction. . . . . 81
- 4.8 Map of seasonal phase delay (left) with contour lines corresponding to the depth of the bottom of the principal aquifer. The dashed blue line corresponds to the approximate boundary between the forebay and confined areas as defined by OCWD. The right plot shows the transects (C - C') for the seasonal phase (solid blue) and aquifer depth (dashed black). . . . . 83
- 4.9 Maps of seasonal peak-to-peak LOS amplitude (left) and seasonal groundwater pumping (right). The locations with the highest seasonal pumping correspond to the locations of the highest seasonal ground deformation within the pressure area. The red outline encompasses the seasonal ground deformation as observed in the InSAR map of seasonal phase delay. The dashed white line indicates the approximate boundary between the forebay and confined areas. . . 83

- 4.10 InSAR time series reconstruction compared with three GPS stations (LBC1, BLSA, SACY) located in the basin. The GPS data are projected onto the radar LOS direction. Both the InSAR and GPS time series have been decomposed into short-term, seasonal signals (A) and long-term, transient signals (C). GPS data are shown with blue dots, and InSAR data are shown with solid red lines. The shaded red regions in (A) and (B) represent the uncertainties in the reconstructed InSAR time series. Note the increase in uncertainty during time periods with no SAR acquisitions. Thin vertical black lines correspond to the SAR acquisition times. (B) and (D) show the residuals between the GPS and InSAR data in (A) and (C), respectively. The locations of the GPS stations are shown in (E). . . . . 86
- 4.11 Maps of March to September LOS deformation for three different years: (A) 2004; (B) 2008; and (C) 2009. The 2004 ground deformation is consistent with the maps of seasonal amplitudes, whereas the 2008 and 2009 maps show starkly different spatial patterns and amplitudes due to cessation of the STSS program. . . . . 87
- 4.12 (A)-(C) show maps of subsidence and uplift rates for the time periods 1995 - 2004 (A), 2007 - 2010 (B), and 2004 - 2007 (C). The labels in (A) indicate the subsidence signature in Long Beach (LB) due to the Wilmington oil Field, uplift in the Sante Fe Springs Oil Field (SF), and subsidence in Pomona Basin (PM) due to groundwater pumping. The transect profile is shown by the black dashed line, M - M'. (D) The solid green line corresponds to the transect for the 1995 - 2004 time period, and dashed blue line corresponds to 2007 - 2010, and the red dashed line corresponds to 2004 - 2007. The transect data show very consistent deformation rate profiles for the two time periods associated with subsidence, whereas the uplift profile has a distinctly different profile. . . . . 88
- 4.13 (A) Close-up maps of long-term LOS displacement from 1995 - 2004, seasonal phase delay, and seasonal peak-to-peak amplitude. The white diamonds in the subsidence map correspond to large-scale OCWD production wells. The yellow squares show the locations of the time series in (B) and (C). (B) shows the time series for the long-term displacements while (C) shows the time series for the short-term, seasonal displacements. . . . . 90

- 4.14 Comparison between long-term LOS subsidence from 1995 - 2004 (A) and residual subsidence after removing a synthetic subsidence signal corresponding to the scaled seasonal displacement (B). . . . . 91
- 4.15 Depth of maximum correlation between hydraulic head data and InSAR displacement are denoted with colored circles for OCWD wells for seasonal signals (A) and long-term signals (B). The background color in the basin corresponds to the depth of the bottom of the principal aquifer. (C)-(F) show the InSAR time series (blue) and scaled head for the well port with maximum correlation (red) for select wells in the basin. . . . . 92
- 4.16 Time series of total volumetric basin deformation vs. total OCWD groundwater production for seasonal signals (A) and long-term signals (B). The volumetric basin deformation was computed by summing the product of pixel area and displacement at each pixel. The dashed red lines correspond to a model of the production time series consisting of annual and semiannual sinusoids for (A) and  $B^i$ -splines for (B). In (B), the data are not shown for visual clarity, and the red shaded area indicates the 1-sigma scatter of the data. . . . . 94
- 4.17 Estimation of storage coefficients for each aquifer and aquitard layer using vertical ground displacements from GPS station SACY and hydraulic head from OCWD well SAR-9. GPS data are shown in blue dots while the model (summation of hydraulic head in each layer scaled by a storage coefficient) is shown in red for short-term (A) and long-term (B) variations. (C) and (D) show the estimated storage coefficients in green and specific storage in magenta for short- and long-term signals, respectively. . . . . 96

- 4.18 (A)-(C): model setup for one-dimensional diffusion simulation with a periodic pressure source placed in the center of the domain. The three different models are: (A) constant hydraulic conductivity and aquifer thickness; (B) constant hydraulic conductivity and a step decrease in aquifer thickness; and (C) narrow fault zone of low hydraulic conductivity and constant aquifer thickness. (F) shows the specific values for thickness and hydraulic conductivity for the three different scenarios. The amplitude and phase of the simulated hydraulic head in response to the periodic pressure source are shown in (D) and (E), respectively. The colors corresponding to the different models are: (A) blue, (B) dashed blue, and (C) red. . . . . 99
- 4.B1 Cross-section of the Santa Ana Coastal Basin aquifers and partitioning into three-layer model (modified from Woodside & Westropp [13]). The distribution of aquifers and interleaving aquitards can be aggregated into a simplified three-layer model where the aquitards separating the three layers are laterally extensive. . . . . 109
- 4.B2 Spatial distribution of largest  $B^i$ -spline coefficients determined from the recursive ADMM InSAR time series analysis. For each image, the color at each pixel corresponds to the amplitude of the specified  $B^i$ -spline. The largest estimated coefficient is associated with a rapid subsidence signal with a time scale of 2.4 years centered in the beginning of 2007 (A). We can also observe high values for the uplift event due to heavy rainfall in (D) and (E). In (H), we see that the uplift in the Santa Fe Spring oil field and subsidence in the Pomona basin are well described by a long-term  $B^i$ -spline centered on 1999. . 110

## LIST OF TABLES

<i>Number</i>	<i>Page</i>
3.S1 Comparison of select earthquakes from the GFZ Potsdam and Global CMT (GCMT) catalogs. Focal mechanisms and moment magnitudes are shown for the earthquakes from each catalog. We compute seismic moment as $M_0 = 1/\sqrt{2} \left( \sum_{ij} M_{ij} \right)^{1/2}$ . Both the mechanisms and moment magnitudes are consistent between the catalogs. . . . .	57
3.S2 Focal mechanisms and moment magnitudes for CLVD events occurring near Bárðarbunga from 1976-1996. The focal mechanisms were formed using the moment tensor elements estimated by Nettles & Ekström (1998). These prior events exhibit opposite polarities from the current events (vertical tensional vs. compressional axes). . . . .	60

*Chapter 1*

## INTRODUCTION

Measurements of surface deformation on the Earth have long been used to study the behavior of various geophysical processes, such as motion of tectonic plates, creeping on major fault systems, magmatic activity, and hydrological activity. Initially, these measurements were limited to measurements of changing surface elevations through leveling and tiltmeters at discrete points, which have been successfully used to assess fault slip over major subduction zones and measure uplift over inflating magma chambers. With the advent of the Global Positioning System (GPS), which allows GPS receivers to measure their position in three components to high precision, scientists have been able to study the behavior of geophysical processes in much greater detail. Extensive GPS networks consisting of hundreds of stations have been deployed all over the globe. Most GPS stations are operated in continuous mode, i.e. position measurements are recorded on a daily or subdaily basis. This sampling mode provides a way to investigate the spatial and temporal evolution of surface deformation, which can be used to constrain time-varying mechanical models and monitor changes in surface strain for natural hazard assessment. In the early 1990s, repeat pass satellite interferometric synthetic aperture radar (InSAR) was introduced. SAR instruments transmit microwaves towards the Earth's surface, measuring the distance between the satellite and the surface. SAR images acquired at different times from the same observation point can be combined using interferometry to obtain high-resolution maps of surface displacement called interferograms. Combining the temporal continuity of GPS observations with the spatial continuity of InSAR observations has greatly enhanced our ability to monitor the Earth's surface in great detail.

One class of deformation signals that have been of interest to the geodetic community are transient signals, which are signals that can be defined as non-periodic, non-secular accumulation of strain in the crust. These signals can be caused by a wide range of geophysical processes, including slow slip events, migration of magma, groundwater flow, etc. Consequently, transients can have a wide range of timescales and spatial extents that are a priori unknown. Such signals can often be quite subtle and embedded in geodetic data with other confounding signals, such as secular plate motion or periodic variations caused by solid earth tides. With the large volume

of geodetic data available from GPS networks and interferograms collected over long time spans, it is imperative that detection of transient signals be automated and reliable. Automatic detection of transients can be useful for detecting the onset of a slow slip earthquake in subduction zones, ground deformation due to pre-eruptive magma flow in volcanoes, or localized creep events on strike slip faults. Such detection procedures can also be used in a retrospective manner to study the behavior of previous transient events by acting as a spatiotemporal filter for geodetic data to reduce the effect of observation noise and non-transient signals.

In addition to transient signals, geodetic data frequently record phenomena that are often assumed to be steady-state, such as secular plate motion or seasonal effects, but these signals can often be time varying due to changes in mechanical properties or their driving forces. Thus, studies that aim to quantify the magnitudes and spatial characteristics of these processes can be corrupted by inaccurate assumptions about the steadiness/stationarity of the signals. Importantly, detecting transient signals and quantifying the behavior of non-steady signals are not mutually exclusive. A time series modeling framework that has the flexibility to detect and model signals of any duration and temporal pattern while providing a level of robustness to limit the effect of observation noise will be invaluable to the geodetic and natural hazard community.

In Chapter 2, we present a new method for automatically detecting transient deformation signals in geodetic time series. The method is based on compressed sensing, a recent technique in signal processing that allows for the reconstruction of *sparse* signals using potentially underdetermined linear systems. By constructing an overcomplete, non-orthogonal design matrix (dictionary) of displacement functions in time that resemble transient signals of various timescales, we can apply a sparsity-promoting regularization to limit the total number of dictionary elements needed to reconstruct transients. We also present an equivalent Bayesian inference approach to sample the full model space for each dictionary element while quantifying uncertainties for the transient signal reconstruction. The detection method localizes the dominant timescales and onset times and enhances the interpretability of transients. We incorporate a spatial weighting scheme that self-adjusts to the local network density and filters for spatially coherent signals. To demonstrate this approach, we apply the spatial weighting to slow slip events in the Cascadia region to automatically characterize their propagation and displacement characteristics.

In Chapter 3, we demonstrate how a combination of unique space geodetic mea-

measurements and seismic observations can be used to study and constrain physical models of an active volcanic event. We describe a model for the caldera collapse at Bárðarbunga stratovolcano which lies beneath the Vatnajökull ice cap in Iceland. The collapsing caldera was characterized by rapid 50 cm/day subsidence of the overlying glacier surface. Associated with this collapse was the initiation of a plate boundary rifting episode extending tens of kilometers northwards of the caldera. The close spatiotemporal association of the collapsing caldera, anomalous seismicity along the caldera rim, and the active rifting provided us with an unprecedented opportunity to understand the mechanics of a caldera collapse in a basaltic system. We present a model of the collapse process consistent with the available geodetic and seismic observations that suggests that the bulk of the observed subsidence occurs aseismically via a deflating sill-like magma chamber. Deflation of the chamber results in anomalous seismic events generated by a rapidly closing crack or rupture on curved, inward-dipping ring faults at depths shallower than the magma chamber. This model differs from prior interpretations of the seismic events that suggested that deformation of a very shallow magma chamber generates seismic events on deeper, outward-dipping ring faults. The model we present is more consistent with the modern geodetic observations and provides a simplified framework to explain the past and present anomalous seismic events.

In Chapter 4, we investigate complex surface deformation within the Los Angeles and Santa Ana coastal basins due to groundwater pumping/recharge and subsequent aquifer compaction/expansion. We extend upon the technique presented in Chapter 2 and apply it to an 18-year InSAR time series consisting of 881 interferograms. This large dataset required the development of a distributed time series analysis framework able to utilize the power of high performance cluster computing. With this framework, we are able to decompose both InSAR and GPS time series into long- and short-term signals, allowing us to study distinct processes occurring within the groundwater system underneath the basins. We find that short-term, seasonal oscillations of ground elevations due to the annual cycle of groundwater pumping and recharge are unsteady and time varying due to changes in groundwater pumping practices by major water districts. We also quantify long-term signals over broad areas within the basin, including subsidence due to slowly declining groundwater levels and a transient uplift event due to a short period of heavy rainfall. Comparison of the decomposed geodetic data with hydraulic head data from major water districts allows for a detailed investigation of the relationship between aquifer pressures and ground deformation. Specifically, we find evidence to suggest that different segments

of the groundwater system are responsible for different temporal components in the observed ground deformation. These results demonstrate the potential for geodetic analysis to be an important tool for groundwater management to maintain sustainable pumping practices.

*Chapter 2*DETECTING TRANSIENT SIGNALS IN GEODETIC TIME  
SERIES USING SPARSE ESTIMATION TECHNIQUES

1. Riel, B., Simons, M., Agram, P. & Zhan, Z. Detecting transient signals in geodetic time series using sparse estimation techniques. *Journal of Geophysical Research: Solid Earth* **119**. doi:10.1002/2014JB011077 (2014).

**2.1 Introduction**

We define transient deformation signals as non-periodic, non-secular accumulation of strain in the crust. Over seismically active regions, transients are often the surface manifestations of slow slip events that are difficult to measure directly with traditional seismological instruments [e.g., 1–5]. In volcanically active regions, transients frequently correspond to periods of ground deformation caused by underlying magmatic activity [e.g., 6–9]. Previously studied transient events vary widely in magnitude, ranging from spatially coherent surface motions of several centimeters to more subtle motions of only a few millimeters. These signals have also varied widely in duration from year-long signals in subduction zone areas [e.g., 10] to very short episodes lasting only a few days [e.g., 1]. Despite the highly non-uniform properties associated with transients, their detection has relied on their combined temporal and spatial coherency, i.e. they are defined by a measurable temporal evolution and systematic spatial structures [e.g., 11].

Detection of transient events with unknown magnitudes and durations requires precise measurements of surface displacements over sufficiently large regions. Over the past two decades, the availability of such measurements for monitoring crustal deformation has rapidly increased. Large-scale continuously-operating GPS networks, such as the Plate Boundary Observatory (PBO) network in the western United States (<http://pboweb.unavco.org>), are used to derive station positions with typical daily repeatabilities of 2-3 mm for horizontal positions and 7-8 mm for vertical positions [12]. In addition to regional coverage, many GPS networks are also very dense, with 1100 permanent GPS stations for PBO, 1200 for Japan's GEONET network ([13]), over 200 for Taiwan ([14, 15]), over 200 for New Zealand ([16]), etc. Geodesy based on repeat imagery (e.g., InSAR) naturally provides spa-

tially dense observations of surface motion but typically suffers from poor temporal sampling. However, the advent of long-timespan InSAR time series, new methods for analyzing the temporal evolution of signals contained in interferograms, and future InSAR missions with approximately weekly repeat times will enable large-scale, high-resolution studies of crustal deformation with sufficiently high temporal resolution to capture many transient processes [17–19].

Each geodetic data type has its own unique set of error characteristics which complicate transient detection. GPS time series typically contain Gaussian white noise plus time-correlated random walk components which can resemble transient signals [20, 21]. GPS networks exhibit spatially-correlated common mode errors which must be estimated and removed as part of any analysis [22]. Many of these errors can be mitigated by analyzing an ensemble of datasets. For a given GPS network density, time-correlated signals in GPS data that are evident only at individual stations can be classified as either colored noise or more local processes. Similarly for InSAR time series, phase delays induced by heterogeneous propagation velocities in the atmosphere can be modeled out or mitigated by ensemble averaging of line-of-sight velocities or other time series techniques [e.g., 18, 23–25]. In addition to noise characteristics, the presence of other confounding signals, such as secular and seasonal effects, can complicate detection of transients. When data volume is large, properly handling sources of errors and non-transient signals makes manual inspection of the data infeasible and requires a sufficiently automated detection algorithm.

We propose a new method for estimating the time and duration of anomalous transient signals in geodetic time series by employing sparse estimation techniques. This method makes use of a dictionary of non-orthogonal time evolution functions that resemble temporally correlated transient events. Estimating the coefficients of the dictionary allows for the reconstruction of transient signals of varying durations and start times. The flexibility of the dictionary also allows for inclusion of known signals, such as seasonal, secular, co-seismic/instruments offsets, post-seismic deformation, etc. The non-orthogonality of the dictionary requires regularization during least squares estimation of the dictionary coefficients. We employ a sparsity-promoting regularization approach to compactly reconstruct the underlying transient signal. Additionally, we introduce a Bayesian sampling scheme for the estimation problem to rigorously assess the uncertainties associated with coupling a non-orthogonal, overcomplete dictionary with sparsity-promoting techniques. The final reconstruction inherently includes information about the dominant timescales

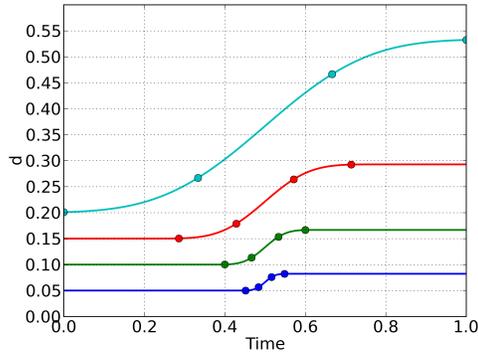


Figure 2.1: Uniform integrated B-splines ( $B^i$ -splines) of various timescales used as candidate temporal displacement functions. The characteristic timescales for the  $B^i$ -splines are determined by dividing the time span into uniformly spaced intervals. The filled circles denote the interval spacings for each  $B^i$ -spline and differ by factors of two.

and likely start times of any transient signals. When time series are available from multiple locations within a certain region, we can exploit the expected spatial coherency of transient signals with a straightforward spatial weighting scheme that encourages selection of dictionary elements that are common to stations within a given length-scale. Application of this method to both synthetic and real GPS time series from the Cascadia region demonstrates the successful recovery of signals of different timescales and magnitudes while providing a direct estimate of the long-term tectonic signal.

## 2.2 Transient Detection

Here, we assume no *a priori* information about the underlying physical mechanisms responsible for a given transient signal. This assumption prevents us from imposing time functions corresponding to a specific physical description. Instead, we use a flexible approach that parameterizes time-dependent deformation with an *over-complete* set, or dictionary, of functions that describe the full suite of behaviors we would expect to be present in a given time series. This dictionary can in principle include sinusoidal functions to model seasonal signals, linear terms for secular velocities, heaviside functions for co-seismic offsets, etc. For transient signals of unknown initiation times and durations, we populate the dictionary with 3rd-order time-integrated B-splines, hereafter referred to as  $B^i$ -splines, which exhibit one-sided behavior of a particular timescale (Figure 2.1) [18]. By dividing the time span of a time series into uniformly spaced intervals, we can generate a series of  $B^i$ -splines centered at the endpoints of the intervals with durations proportional

to the interval durations. In this work, a dyadic spacing scheme (i.e.,  $B^i$ -splines of duration  $T/4, T/8, T/16$ , etc., for a given timespan  $T$ ) is used to populate the dictionary. A similar approach is used in wavelet analysis to efficiently cover the frequency spectrum of a signal [e.g., 26]. At this stage, modeling time-dependent deformation for time series data,  $\mathbf{d}$ , is reduced to estimating the coefficients,  $\mathbf{m}$ , of the dictionary elements in  $\mathbf{G}$ , while imposing a linear relation between the model parameters and data,  $\mathbf{G}\mathbf{m} = \mathbf{d}$  (here, boldface indicates matrix or vector quantities).

### Regularized Least Squares

Due to the non-orthogonality of the  $B^i$ -splines in  $\mathbf{G}$ , any estimate of  $\mathbf{m}$  derived using ordinary least squares methods will be particularly sensitive to the data noise and will exhibit large variances for the estimated parameters. Regularization techniques aim to reduce this sensitivity by jointly minimizing a measure of the residual  $\|\mathbf{G}\mathbf{m} - \mathbf{d}\|_2^2$ , where  $\|\cdot\|_2$  denotes the Euclidean or  $\ell_2$ -norm, and a regularizing function that incorporates *a priori* information about the solution. Typically, regularized least squares optimization minimizes the unconstrained cost function,  $\varphi(\mathbf{m})$ :

$$\varphi(\mathbf{m}) = \|\mathbf{G}\mathbf{m} - \mathbf{d}\|_2^2 + \lambda F(\mathbf{m}), \quad (2.1)$$

where  $\lambda > 0$  controls the degree of regularization and  $F(\mathbf{m})$  is the regularizing function. The above formulation can be modified to incorporate uncertainties on the observed data in  $\mathbf{d}$ . Traditional zeroth-order Tikhonov regularization, where  $F(\mathbf{m}) = \|\mathbf{m}\|_2^2$ , minimizes the size or energy of the solution  $\mathbf{m}$ . For a mixed dictionary of  $B^i$ -splines and steady-state functions, such as seasonal and secular terms, we generally only penalize the  $B^i$ -splines and allow the steady signals to compensate for the rest of the displacement provided by the data. Thus, in Equation 2.1,  $F(\mathbf{m}) \rightarrow F(\mathbf{m}^{B^i})$ , where  $\mathbf{m}^{B^i} \subseteq \mathbf{m}$  denotes the  $B^i$ -spline coefficients. For brevity in the following discussion, we assume  $\mathbf{m} = \mathbf{m}^{B^i}$ . In a Bayesian framework, Tikhonov regularization implies an uncorrelated zero-mean Gaussian prior for the coefficients of the  $B^i$ -splines and uniform priors for everything else.

However, transient events are not well described by Gaussian statistics; rather, transients are sparse and irregular in nature. To promote sparsity, we would like to use  $F(\mathbf{m}) = \|\mathbf{m}\|_0$ , where  $\|\cdot\|_0$  denotes the counting pseudo-norm and measures the number of non-zero elements in  $\mathbf{m}$  [27, 28]. Since this formulation results in an intractable combinatorial problem, it is common to use an approximation with  $F(\mathbf{m}) = \|\mathbf{m}\|_1$ , where  $\|\cdot\|_1$  is the  $\ell_1$ -norm or the sum of the absolute values of  $\mathbf{m}$ ,

leading to a *convex* cost function ([28–30]):

$$\varphi(\mathbf{m})_{\ell_1} = \|\mathbf{G}\mathbf{m} - \mathbf{d}\|_2^2 + \lambda\|\mathbf{m}\|_1. \quad (2.2)$$

Using  $\ell_1$ -norm regularization, the solution still maintains sparsity, i.e. many components of  $\mathbf{m}$  are very close to zero and the remaining components can still effectively describe the data. Furthermore, a convex cost function results in a solution that is guaranteed to be globally optimal [31]. We cast the sparse regularization problem as a quadratic program and solve for  $\mathbf{m}$  using the CVXOPT software (<http://abel.ee.ucla.edu/cvxopt/index.html>).

Sparsity-promoting regularization with the  $\ell_1$ -norm has been effectively used for recovering isolated spikes in seismic data ([32]), detecting sharp discontinuities in tomography studies ([33]), and estimating compact distributions of fault slip for large earthquakes ([34]). In the context of transient detection, this method automatically determines the  $B^i$ -splines which best model the deformation while zeroing out the others. The advantage of enforcing a sparse set of  $B^i$ -splines is that we automatically place higher importance to  $B^i$ -splines that have nearly the same timescales and onset times as any transient signals present in the data and heavily penalize those that do not significantly improve our data fit. Steady signals, such as those from seasonal and secular processes, are also required to be consistent with a sparse set of  $B^i$ -splines. This requirement is beneficial for ensuring that any estimated steady signals do not overly accommodate the observed displacement which can lead to false positives in transient signal detection (Section 2.5).

The effectiveness of the  $\ell_1$ -norm for recovering sparse solutions can be enhanced by adaptive re-weighting techniques. In Candés *et al.* [35], an iterative re-weighting algorithm was introduced where each coefficient,  $m_i$ , is assigned a different penalty parameter,  $\lambda_i$ , which is inversely proportional to  $|m_i|$  at the current iteration. By initializing the algorithm with uniform values for  $\lambda_i$ , each successive iteration causes larger coefficients to be penalized less heavily than smaller coefficients, leading to a solution that enhances the most dominant  $B^i$ -splines. Candés *et al.* [35] demonstrated that re-weighting brings the  $\ell_1$ -norm closer to the  $\ell_0$ -pseudo-norm by increasing the strength of the regularizing function  $F(\mathbf{m})$  near the origin. Theoretically, one could choose from a multitude of functional forms relating  $\lambda_i$  to  $|m_i|$ , e.g.,  $\lambda_i \propto |m_i|^{-1}$ ,  $\lambda_i \propto m_i^{-2}$ ,  $\lambda_i \propto \log(|m_i|^{-1})$ , etc. Larger negative powers for  $m_i$  will increase the strength of the sparsity constraint and bring the  $\ell_1$ -norm very close to the  $\ell_0$ -norm. In practice, we have found that using the stronger re-weighting

functions favors selection of short-timescale  $B^i$ -splines while the logarithmic functions favor longer-timescales. Thus, selecting the appropriate re-weighting function can depend on the expected timescales of the transient signals in a data set. Typically, for all re-weighting functions, 5-10 re-weighting iterations are required for convergence.

### Posterior Uncertainties

In ordinary least squares problems where all variables are assumed to be Gaussian (unregularized or Tikhonov regularization), analytic relations exist to estimate model and predicted data uncertainties [36]. While the  $\ell_1$ -norm regularization prevents us from directly using those relations, we can interpret the minimization of the cost function in Equation 2.2 as choosing the optimum subset of the elements of  $\mathbf{G}$  that minimize the data misfit as well as the number of elements used for the solution and determining the coefficients of those elements. Thus, for a given solution vector  $\mathbf{m}$ , we can construct a compact dictionary  $\tilde{\mathbf{G}}$  populated with the elements corresponding to the largest coefficient absolute values in  $\mathbf{m}$ . We can construct a diagonal prior covariance matrix,  $\mathbf{C}_m$ , where the values along the diagonal correspond to the squared coefficients in  $\mathbf{m}$ . Then, for a given data covariance matrix  $\mathbf{C}_d$ , we can apply the standard least squares formulation to obtain a solution  $\tilde{\mathbf{m}}$ :

$$\tilde{\mathbf{m}} = \left( \tilde{\mathbf{G}}^\top \mathbf{C}_d^{-1} \tilde{\mathbf{G}} + \mathbf{C}_m^{-1} \right)^{-1} \tilde{\mathbf{G}}^\top \mathbf{C}_d^{-1} \mathbf{d}. \quad (2.3)$$

Various stopping criteria can be applied for determining the number of elements to include in  $\tilde{\mathbf{G}}$ . Here, we apply a variance reduction criterion where we first remove the estimated steady-state signals from the data and iteratively remove modeled transient displacements corresponding to the largest values of  $\mathbf{m}$  from the data until the variance reduction reaches a prescribed threshold. This approach is similar to matching pursuit methods that iteratively search through non-orthogonal bases and add those to the dictionary that are most correlated with the data residual at each iteration [e.g., 26]. In our case, the ‘‘best basis’’ is determined in one step through the convex optimization and is less susceptible to high data noise or initial errors in basis selection [30].

Using knowledge about uncertainties for the observed data in  $\mathbf{C}_d$ , the posterior model covariance matrix can be computed as:

$$\tilde{\mathbf{C}}_m = \left( \tilde{\mathbf{G}}^\top \mathbf{C}_d^{-1} \tilde{\mathbf{G}} + \mathbf{C}_m^{-1} \right)^{-1}. \quad (2.4)$$

Since the resultant compact dictionary  $\tilde{\mathbf{G}}$  is in most cases composed of non-orthogonal elements and will be ill-posed, we can expect some large off-diagonal

components in  $\tilde{\mathbf{C}}_m$ . The probability density function of the predicted data is then a multivariate Gaussian distribution with a mean of  $\tilde{\mathbf{G}}\tilde{\mathbf{m}}$  and a covariance matrix,  $\tilde{\mathbf{C}}_d$ , given by:

$$\tilde{\mathbf{C}}_d = \tilde{\mathbf{G}}\tilde{\mathbf{C}}_m\tilde{\mathbf{G}}^\top. \quad (2.5)$$

We note that while this approach is useful for assessing the uncertainties of the coefficients of  $\tilde{\mathbf{G}}$  and the predicted data, it does not address the uncertainties associated with the subsetting of  $\mathbf{G}$  to form  $\tilde{\mathbf{G}}$ . Furthermore, this approach relies on the assumption that the model parameters are normally distributed, which is contrary to our assumption of transients as temporally sparse.

### Bayesian Sampling

As was previously mentioned, the commonly implemented form of Tikhonov regularization is equivalent to enforcing a Gaussian prior on the elements of  $\mathbf{m}$ . Analogously, sparsity-promoting regularization in its most basic form can be achieved by enforcing a Laplace prior for  $\mathbf{m}$ , which has the form  $p(m_i) \propto \exp\{-\lambda|m_i|\}$  [29]. We can see this result by considering Bayes' theorem,  $P(\mathbf{m}|\mathbf{d}) \propto P(\mathbf{d}|\mathbf{m})P(\mathbf{m})$ , where  $P(\mathbf{m}|\mathbf{d})$  is the posterior distribution of our model coefficients (i.e., the distribution of values for  $\mathbf{m}$  that explain the data),  $P(\mathbf{d}|\mathbf{m})$  is the data likelihood, and  $P(\mathbf{m})$  is the prior distribution of the coefficients. By maximizing  $\exp\{-\varphi(\mathbf{m})_{\ell_1}\}$ , it can be shown that the regularized least squares solution is equivalent to maximizing the posterior distribution with a Gaussian misfit between the data and model prediction and a Laplace prior on the model coefficients.

Laplace priors are characterized by high probabilities near the origin with long-tails to allow for an increased likelihood of arbitrarily large values relative to a Gaussian prior. The penalty term  $\lambda$  acts as a scale factor that controls the width of the distribution and the probability that the elements of  $\mathbf{m}$  will be sparse. Since there is no convenient conjugate relation between a Gaussian likelihood and Laplace prior, we cannot derive a closed-form solution for the posterior distribution  $p(\mathbf{m})$ . Instead, we employ a Gibbs sampler to draw samples from the posterior distribution [37]. The Gibbs sampler explores the posterior distribution of each variable in the model using distributions conditional on the current values for all other variables. Following the approach of Park & Casella [38], we group the coefficient amplitudes,  $m_i$ , as a single variable and the coefficient *precisions*,  $\tau_i$ , as another group of variables. The conditional distributions relating the coefficient amplitudes and precisions are obtained by expressing the Laplace prior as a scale mixture of normals with an

exponential mixing density:

$$\frac{\lambda}{2} e^{-\lambda m_i} = \int_0^\infty \frac{1}{\sqrt{2\pi\tau_i}} e^{-m_i^2/(2\tau_i)} \frac{\lambda^2}{2} e^{-\lambda^2\tau_i/2} d\tau_i. \quad (2.6)$$

The posterior distribution can now be expressed as a product of a Gaussian data likelihood, Gaussian priors for the dictionary coefficients, and exponential hyperpriors for the coefficient precisions. This hierarchical representation where each  $\tau_i$  is treated as a hyperparameter is analogous to the re-weighting scheme described in Section 2.2. The conditional distributions in this hierarchy are straightforward to sample from with block updates of  $\mathbf{m}$  and  $(\tau_1^2, \dots, \tau_p^2)$ .

There are several advantages that favor a Bayesian sampling method over regularized least-squares. For one, optimization-based approaches for non-Gaussian priors do not permit us to readily compute uncertainties associated with our estimate for  $\mathbf{m}$ . While we could apply least squares theory using a sparse subset of  $\mathbf{G}$  that best explain the data (see Section 2.2), we would still require the assumption that the model parameters are normally distributed. We also do not obtain much information about the full solution space, which is necessary if different families of solutions exist with nearly the same predictive power as the optimal solution. For example, consider the inherent trade-off between a single  $B^i$ -spline to model a transient signal and two shorter-timescale  $B^i$ -splines located at the same time. While selecting the single  $B^i$ -spline would be the sparser solution, we can imagine a situation where the combination of the two shorter  $B^i$ -splines provided a better fit to the data. In this case, the single, longer timescale  $B^i$ -spline would trade off with the shorter ones where the strength of the trade-off would be dependent on the value of the penalty  $\lambda$ . Bayesian sampling allows for sampling from the full solution space where models are produced in numbers proportional to their probability given the data [e.g., 36, 37].

### Selecting the Penalty Parameter

The parameter  $\lambda$  in the Laplace prior controls the relative strengths of the steady-state terms and the  $B^i$ -spline coefficients. Larger values of  $\lambda$  will minimize the contributions of the  $B^i$ -splines, leading to a smoother solution. Smaller values of  $\lambda$  will distribute more weight across the  $B^i$ -splines, leading to a rougher solution. Thus, the amplitude of the steady-state signals is also sensitive to the value of  $\lambda$  and may vary as  $\lambda$  changes. For sparse regularization optimization problems, we use  $K$ -fold cross-validation to select the optimal values for  $\lambda$  where  $K$  depends on the number of data points available for partitioning into training and testing subsets.

Due to the sparsity-enforcing regularization, we can directly quantify the effective data resolution of  $\mathbf{G}$  (i.e., the shortest-duration resolvable signal) as the duration of the shortest timescale  $B^i$ -spline included in  $\mathbf{G}$ . A single  $B^i$ -spline is only able to predict four independent observations spaced  $T_k/2$  time units apart, where  $T_k$  is the effective timescale of the  $k$ -th  $B^i$ -spline. For daily observations, we first partition the data into  $S$  random subsets where  $S$  is the number of days spanned by the time series divided by the number of observations predictable by  $\mathbf{G}$ . Within each subset, we further divide the data into  $K$  random partitions where one of the partitions is used to compute the data misfit for the proposed model  $\mathbf{m}$  trained by the other  $K - 1$  partitions. We then average over  $S \cdot K$  cross-validation experiments to obtain the average data misfit for the current value of the penalty parameter,  $\lambda$ .

Since cross-validation would be computationally expensive for a high number of Gibbs sampling runs, alternative methods are required. A variety of model class selection methods are available that allow for estimation of the *evidence* of a model class, where a model class is defined by the value of  $\lambda$  [39, 40]. The evidence measures the average data fit for a model class and the amount of information the model class extracts from the data, i.e. some metric of distance between the posterior and prior distributions [39]. Unfortunately, model class selection is highly influenced by the choice of the prior distribution, which could lead to significant biases for values of  $\lambda$  that maximize the evidence [40]. Here, we sample for  $\lambda$  by assigning it a diffuse hyperprior, allowing for a wide range of possible widths for the corresponding Laplace priors [38]. We consider a gamma hyperprior on  $\lambda^2$  such that the prior density is relatively flat up to  $\lambda \approx 10^3$  and then decreases steeply to penalize very large values.

### 2.3 Synthetic Example

To test the temporal transient detection capabilities of the proposed method, we generated a 20-year synthetic daily GPS time series consisting of seasonal, secular, and transient deformation. The seasonal signals are a linear combination of annual and semi-annual sinusoids. A transient signal is constructed using arctangent functions of three different amplitudes, timescales, and centroids to simulate slow deformation events with various properties (Figure 2.2). We add white noise plus colored noise using a power law model to mimic errors commonly found in geodetic data [21]. Robust detection of transients is traditionally difficult when the amplitude of the temporally correlated colored noise is on the same order as the signal of interest (as is the case for the weakest synthetic transient). In practice, this problem

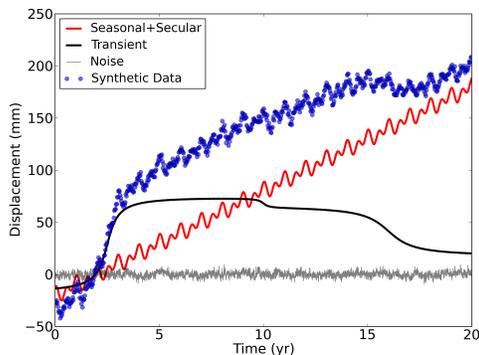


Figure 2.2: Individual model components input into the synthetic GPS time series. The seasonal signals are constructed using annual and semi-annual periods while the transients are formed using arctangent functions of various amplitudes and durations. Additionally, white and colored noise are added to the model to simulate measurement noise typically found in geodetic time series.

can be mitigated by exploiting coherency within a geodetic network, but we will demonstrate that successful detection is still possible with a single time-series.

We construct  $\mathbf{G}$  using a dictionary of reference functions that include simple sinusoidal and linear functions to capture seasonal and secular terms, respectively. Additionally, the dictionary includes  $B^i$ -splines with effective timescales of 0.3, 0.6, 1.3, 2.7, 5.7, and 13.3 years (chosen by dividing the 20-year time period into 128, 64, 32, 16, 8, and 4 uniformly spaced intervals, respectively). The coefficients of the dictionary terms are estimated using two different methods: least squares optimization with sparsity-inducing regularization on the  $B^i$ -spline coefficients and Gibbs sampling of the posterior distribution. As described earlier, we prescribe Gaussian priors on the seasonal and secular terms and Laplace priors on the  $B^i$ -splines, although the sampler is initialized with random variates from a wide Gaussian distribution for all coefficients. For this example, we run the Gibbs sampler for  $10^5$  samples, which is far more than required for convergence but allows the posterior means to evolve to within  $< 1\%$  of their final values (Figure 2.3).

After cross-validation of 14 independent data subsets, the average optimal penalty parameter was  $\lambda = 0.66$ , which agreed fairly well with the mean of the posterior distribution of  $\lambda$  constructed with the Gibbs sampler (Figure 2.4). Prescribing  $\lambda = 0.66$  results in a reconstructed transient signal that is rougher than the input transient (Figure 2.5a). This behavior is primarily due to the temporally-correlated colored noise causing several false detections throughout the time series. The random walk characteristics of the noise resemble small transient events that are

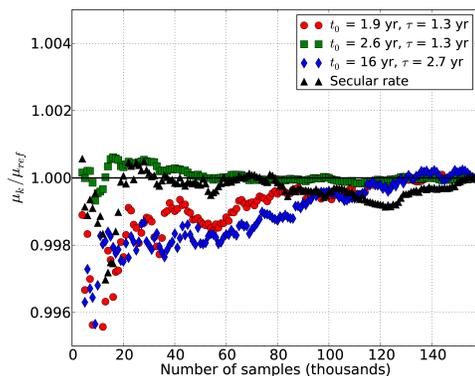


Figure 2.3: Convergence test of the Gibbs sampler for coefficients corresponding to several  $B^i$ -splines. The mean of the samples drawn from the posterior distribution for each  $B^i$ -spline is monitored for increasing number of samples. The means are normalized by the mean computed using  $2e^5$  samples. After  $O(7e^4)$  samples, the means are within 0.2% of their final values.

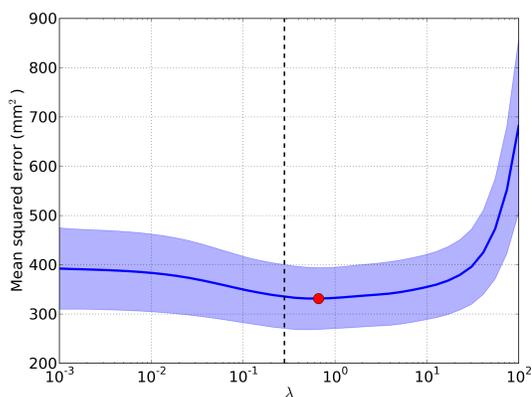


Figure 2.4:  $K$ -fold cross-validation results for selection of the penalty parameter,  $\lambda$ , for the synthetic time series. Fourteen separate cross-validation experiments were performed for fourteen independent subsets of the time series. The solid blue line shows the mean testing error for all experiments while the shaded region denotes the standard deviation. The vertical black dashed line marks the mean  $\lambda$  as determined by the Gibbs sampler which shows relatively good agreement with the optimal  $\lambda$  obtained from the cross-validation (red circle).

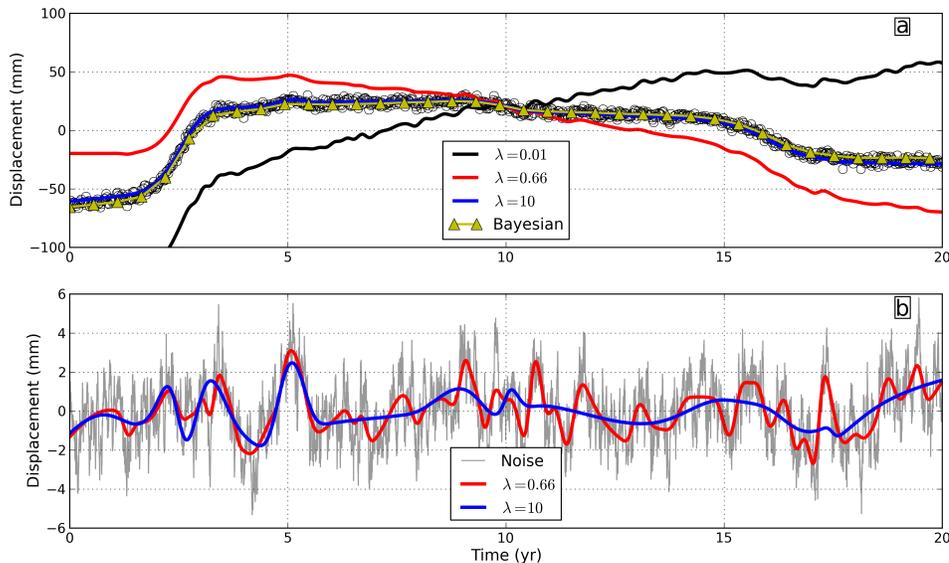


Figure 2.5: Reconstructed time series for the synthetic data with the simultaneously estimated seasonal and secular signal removed. (a) Full transient signal for models corresponding to different values of the penalty parameter  $\lambda$ . The black circles show the input data after removing the true secular and seasonal signals. Lower values of  $\lambda$  correspond to rougher models. The model for  $\lambda = 10$  and the point estimate derived from the Gibbs samples for the same  $\lambda$  are able to nearly exactly reproduce the input signal. (b) The high frequency component of the reconstructed transient signal compared with the input colored noise. The  $\lambda = 0.66$  model (chosen through cross-validation) is able to capture nearly all of the temporally correlated noise structure while higher penalties result in smoothing over the higher frequency variations.

indistinguishable from true transients (for data from a single station). Colored noise also has the effect of biasing the estimate of the long-term secular rate as demonstrated by the under- and over-estimation of the secular rate for the models constructed with  $\lambda = 0.01$  and  $\lambda = 0.66$ , respectively. Low values of  $\lambda$  allow selection of more  $B^i$ -splines from the dictionary to fit the the smaller signals. By plotting the reconstructed high-frequency signal against the input colored noise (Figure 2.5b), we observe that enforcing  $\lambda = 0.66$  allows us to reconstruct a majority of the structure of the input noise. The combined dictionary plus sparsity-inducing regularization approach thus acts as a smoothing method where  $\lambda$  controls the degree of smoothness of the reconstructed signal. Selecting a much higher value of  $\lambda = 10$  still allows us to reconstruct the largest signals of the input noise while smoothing over the higher-frequency variations.

Without spatial information, it is impossible to distinguish between colored noise

and true transient signals for tuning  $\lambda$  to recover the correct secular rate. Data from multiple stations must be used to determine whether a temporally coherent signal persists over a finite region (Section 2.4). In a separate cross-validation experiment performed on synthetic data with white noise only resulted in optimal values of  $\lambda \approx 10$ , which successfully isolates the transient signals. Since the focus of this paper is on transient detection, we proceed with the model resulting from  $\lambda = 10$  to smooth over the noise signals, but we emphasize that in general applications, the issue of colored noise must not be neglected.

The reconstructed time series with  $\lambda = 10$  successfully models all three input transient signals (Figure 2.5a). A nearly identical model is achieved by deriving a point estimate from the means of the posterior distribution constructed from the Gibbs sampler with the same value for  $\lambda$ . Even with this higher value of  $\lambda$ , both the optimization and Bayesian solutions are corrupted by a relatively long-duration random walk process that started around year 4.5. The duration of this colored noise was of sufficient length to be modeled by one of the  $B^i$ -splines in the dictionary. Increasing the penalty parameter further would smooth over this noise signal at the cost of losing recovery of the smallest transient signal at year 10.

### **Coefficient Scalograms**

The limited data resolution (as opposed to model resolution) of the dictionary  $\mathbf{G}$  will limit the precision of our estimates for transient durations. One can also expect significant covariances between dictionary elements that may have different timescales but share common centroid times. Viewing the estimated  $B^i$ -spline coefficients in a scalogram-type fashion reveals the effectiveness of the different estimators in limiting the intra-dictionary covariances (Figure 2.6). For comparative purposes, we also show the scalogram for a model using a traditional zeroth-order Tikhonov regularization scheme. The Tikhonov estimator locates the onset times of the largest transient signal relatively well but tends to spread the energy across the timescales, much the same way that a wavelet transform would spread the energy across wavelet scales. In fact, previous methods using wavelet transforms to pick onset times of transients in GPS data have presented scalograms that suggest permanent deformation across all temporal scales [e.g., 5]. The advantage of the sparse estimation techniques is a much stronger localization of energy to very few scales and times, as seen on the bottom two plots in Figure 2.6. The majority of each input transient signal can be recovered by just two distinct  $B^i$ -splines, allowing for more precise estimates of the transient start times and durations. We reiterate

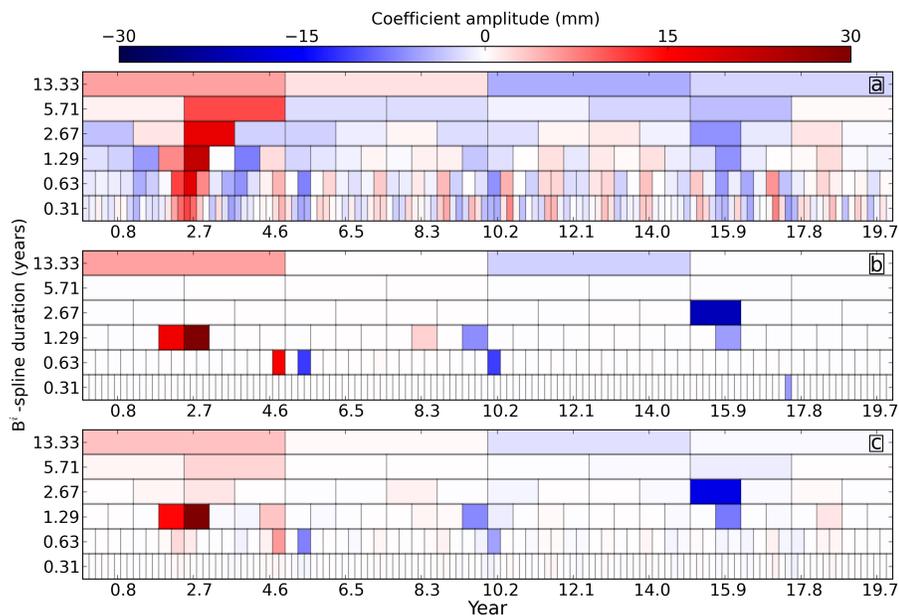


Figure 2.6: Synthetic scalograms showing the amplitudes of the  $B^i$ -splines estimated using three different methods: (a) Tikhonov ( $\ell_2$ -norm) regularization; (b) Sparse ( $\ell_1$ -norm) regularization with  $\lambda = 10$ ; and (c) posterior distribution means derived from the Gibbs samples. Each row of the scalograms correspond to  $B^i$ -splines of a given timescale. Tikhonov regularization results in non-zero amplitudes for nearly all  $B^i$ -splines and tends to spread energy across timescales. The sparsity-promoting regularization methods zero-out nearly all of the  $B^i$ -splines, leaving only those that describe transients in the data. The estimated secular rates for the Tikhonov, sparse regularization, and Gibbs solutions are 12.0, 9.7, and 8.6 mm/yr, respectively (for an input secular rate of 10 mm/yr).

that these estimates have been obtained almost completely automatically, with some minor supervision for selection of the penalty parameter (i.e., choosing the correct data subset size for cross-validation such that the optimal estimate of  $\lambda$  does not change significantly for slightly different subset sizes).

### Covariances Between Dictionary Elements

Estimating the uncertainties of the  $B^i$ -spline coefficients is straightforward since samples have been drawn from the posterior distribution via the Gibbs sampler. Furthermore, we can directly investigate trade-offs between different parameters and gain insight into the nature of sparse transient detection with a non-orthogonal dictionary. For example, the longer duration transient centered around the 16 year mark can be described well by both a 2.67-year and a 1.29-year  $B^i$ -spline, leading to large standard deviations for both coefficients accompanied by a strong trade-off

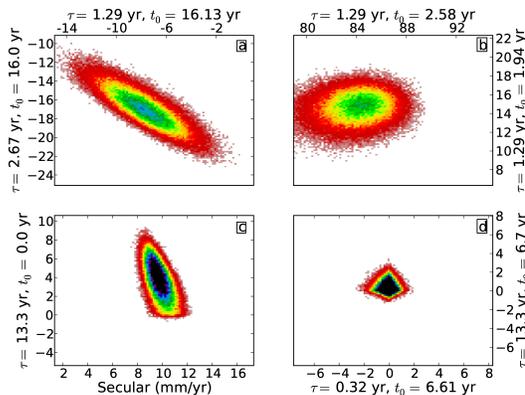


Figure 2.7: Two-dimensional histograms of samples drawn from posterior distributions for different pairs of  $B^i$ -splines. (a) When two  $B^i$ -splines have nearly coincident time centroids and are both able to reconstruct a transient signal, their joint posterior exhibits a strong negative covariance. (b) For adjacent  $B^i$ -splines of the same timescale where the centroid of the true signal lies between the centroids of the  $B^i$ -splines, their amplitudes will co-vary in a positive manner. (c) Long-timescale  $B^i$ -splines and the secular rate will exhibit covariances similar to case (a). Note the sharp corner in the joint distribution due to the  $\ell_1$ -norm penalty. (d) Two  $B^i$ -splines located in a time window where no transient signals occur will have samples located very close to the origin with a structure resembling the prior (bi-variate Laplace distribution).

(Figure 2.7a). The longer duration  $B^i$ -spline more closely matches the duration of the input transient but is slightly time-shifted from the true centroid time, whereas the 1.29-year  $B^i$ -spline is more closely aligned with the centroid time but under-represents the signal duration. The negative slope in the covariance plot is a direct result of the Laplace prior placed on  $m_i$ , which tries to drive the coefficients closer to zero. In the case when two coefficients have nearly equal probability in matching the data, many samples will be drawn from the model space where both coefficients are non-zero. However, this behavior is entirely dependent on the value of the penalty parameter  $\lambda$ . Higher values of  $\lambda$  would draw more samples for the more probable coefficient and less for the less probable coefficient (see Section 2.5).

We can observe a similar trade-off between a long-timescale  $B^i$ -spline with a time centroid of  $t = 0$  and the secular rate, implying that long-timescale  $B^i$ -splines are nearly as effective in modeling long-term, steady displacement signals (Figure 2.7c). Trade-offs between  $B^i$ -splines adjacent in time with identical durations also show large standard deviations but with slightly weaker trade-offs (Figure 2.7b). In this case, we can infer that the centroid of the true signal is most likely between the two and has a time duration of  $\approx (3/2)T_k$ . At the other extreme, when two  $B^i$ -splines are

centered in a period when no transient signal occurs, nearly all of the samples lie very close to the origin (Figure 2.7d), replicating the expected probability distributions of two sparse parameters [41]. Another feature evident in many coefficient posterior distributions is the sharp corner seen in Figure 2.7c which is a consequence of the  $\ell_1$ -norm strongly penalizing coefficients of the wrong sign [38]. Thus, a simple viewing of the posterior samples can provide a strong indication of the most likely sign of displacement of a detected transient signal.

### Data Subsampling

As discussed previously, when geodetic time series have very high temporal sampling rates, the limited data resolution of  $\mathbf{G}$  requires a certain level of data subsampling to obtain reliable estimates of the penalty  $\lambda$  during cross-validation. On the other hand, for studies where time series are expected to have poorer temporal sampling, we can estimate the minimum amount of data required to detect a transient signal of a given duration since a  $B^i$ -spline of duration  $T_k$  is expected to predict four observations spaced  $T_k/2$  time units apart. While daily GPS solutions provide adequate sampling rates for capturing many transient processes, other geodetic time series, such as InSAR, provide observations that are typically sparser in time. If the duration of a transient process is appreciably shorter than the time interval between observations, there is a risk of severely mis-estimating the timescale of the signal or not detecting the signal at all. To investigate the effect of sampling rate on transient detection, we repeated the least squares optimization with sparsity-promoting regularization on the synthetic time series with increasing data decimation factors. We varied  $\lambda$  for each decimation factor in order to keep the results consistent. Remarkably, the three input transient events were successfully recovered up to a decimation factor of 256 (Figure 2.8a). For this highly ill-posed case where the number of candidate features (252  $B^i$ -splines) is significantly greater than the number of data points (29 points), sparse regularization was able to recover a stable solution. Reconstruction of the input transients becomes less accurate with increasing time intervals between observations, but even for the smallest signal, only 2-3 data points are required to register a positive detection, in agreement with the expected data resolution of the  $B^i$ -splines (Figure 2.8b).

Decimation also reduces the influence of higher-frequency time correlated noise by effectively low-pass filtering the data before estimation of  $\mathbf{m}$ . We can observe this effect from the increasing smoothness of the reconstructed signal for larger decimation factors. This result suggests that a data cascading approach could be

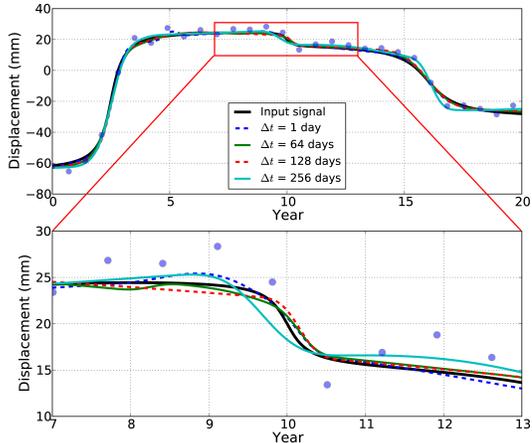


Figure 2.8: Transient signal recovery using subsampled data with increasing subsampling factors. The limited data resolution of the smallest timescale  $B^i$ -splines requires 3-4 observations to resolve a transient signal. In the case of the smallest synthetic transient around year 10, the temporal spacing of  $\Delta t = 256$  days is too large to resolve the short-timescale duration, and the reconstructed signal is smoothed (bottom panel). However, this result still demonstrates that a positive detection is still possible with temporally coarse time series.

beneficial in recovering the strongest signals. For cascading, we would estimate the  $B^i$ -spline coefficients using only a subset of the data and use those results as an *a priori* estimate of the coefficients for use with a larger subset of data [42]. The *a priori* estimate could then be integrated with the re-weighting approach in Section 2.2 to impose smaller penalties on  $B^i$ -splines that have larger values in the initial estimate. Subsequent estimates of  $\mathbf{m}$  using more data would enhance the stronger signals and reduce the overall effect of high-frequency colored noise. This approach would be useful for combining coincident geodetic time series with different temporal sampling rates and noise characteristics, e.g., InSAR and GPS. Transient detection would first be performed with an InSAR time series. Feeding this initial result into a detection procedure with the GPS data would then enhance the transient signals that are consistent between the two data types.

## 2.4 Spatial Sparsity Weighting

Much of the ambiguity over the correct choice of  $\lambda$  and the influence of local noise can be mitigated by using data from multiple surrounding stations. By the adopted definition of what constitutes a transient signal, displacements should be coherent over a finite region and would lead to common non-zero  $B^i$ -splines over multiple stations. Thus, the re-weighting scheme discussed in Section 2.2 can be

performed in a spatial sense where the coefficient-dependent penalties are enforced to be consistent over stations within a certain length-scale. Using this approach, the spatial re-weighting is initialized by performing a single iteration of the minimization of Equation 2.2 independently for each GPS station. After this iteration, candidate coefficient-dependent penalty parameters  $\lambda_i$  are computed for each station based on that station's current initial estimate for  $\mathbf{m}$ . To enforce spatial consistency for  $n$  stations, we select  $\lambda_i = f(\lambda_i^1, \dots, \lambda_i^n, w^1, \dots, w^n)$ , where  $f()$  is a weighted median and  $w^j$  are the weights assigned to each station. The weights are re-computed at every  $k$ -th station using the spatial weighting function:

$$w^j = \exp\left(-\frac{d(j, k)}{L_0^j}\right), \quad (2.7)$$

where  $d(j, k)$  is the distance between station  $j$  and  $k$  and  $L_0^j$  is a prescribed correlation length for the  $j$ -th station. To account for variable station densities within GPS networks, we choose  $L_0^j$  to be the average distance from station  $j$  to the nearest three or four stations, allowing us to detect spatially correlated transients with length scales proportional to the resolving power of a given network.

Since the spatial sparsity weighting method relies on spatial consistencies of scalograms for neighboring stations, we must ensure that the  $B^i$ -spline decomposition of transient signals is translation-invariant. Translation-invariance states that a time shift of the input signal (the data) will only result in an equivalent time shift of the selected  $B^i$ -splines without modification of the amplitudes [26]. In wavelet analysis, wavelet transforms that are not translation-invariant can result in vastly different decompositions for small time shifts in the data. For  $\ell_1$ -regularized least squares problems, we can achieve translation-invariance by constructing a translation-invariant dictionary  $\mathbf{G}$ . A given dictionary  $\mathbf{G}$  is translation-invariant if for any temporal function  $\mathbf{g}_i(t) \in \mathbf{G}$  and  $t_0 \in \Delta t * [0, N - 1]$ , where  $N$  is the number of data points and  $\Delta t$  is the time duration between observations, then  $\mathbf{g}_i[t - t_0] \in \mathbf{G}$  [26]. In other words, we construct  $\mathbf{G}$  such that every observation epoch in the time series is associated with a  $B^i$ -spline of all valid temporal scales, which would result in a  $\mathbf{G}$  matrix with  $N$  rows and  $N \log_2 N$  columns if a dyadic scale approach is used for the  $B^i$ -splines. Due to the large number of parameters associated with translation-invariant dictionaries, we generally only enforce translation-invariance when applying the spatial sparsity weighting to time series that potentially contain rapidly propagating transients, such as in Cascadia. In practice, we have found that longer-duration transients

can be effectively isolated using the spatial sparsity weighting with the standard  $\mathbf{G}$  construction, i.e. 4  $B^i$ -splines of duration  $T/4$ , 8  $B^i$ -splines of duration  $T/8$ , etc.

### **Example: SCEC Validation Exercises**

Since 2009, the Southern California Earthquake Center (SCEC) community has coordinated transient detection validation workshops where participants are able to test their detection methods on several synthetic time series resembling data from southern California GPS stations [43]. The data are generated by the Fakenet package which simulates transient processes of varying complexity while including additional signals from seasonal and secular processes, random and common mode noise, and data gaps [44]. Four phases of testing from 2009 to 2012 were performed with transient sources ranging from slow slip events with strike-slip and thrust motions to small- and large-scale aquifer inflations. We apply the spatial sparsity weighting approach to a 10-year synthetic dataset from Phase 3 (set D) which contains signals from a simulated thrust event on the Santa Monica fault. As before, we populate a global temporal dictionary with  $B^i$ -splines of timescales of  $\approx 0.16, 0.32, 0.65, 1.33, 2.86,$  and  $6.67$  years, as well as functions for seasonal and secular processes. The correlation lengths  $L_0^j$  are computed for each station using the average distance to the nearest three stations, resulting in strong resolution power over the Los Angeles basin and weaker resolution near the California-Mexico border and islands (Figure 2.9).

After about twenty iterations of the spatial sparsity weighting, the thrust event is strongly localized to the stations nearest to the Santa Monica fault with good agreement with the true signal (Figure 2.10). While there are a few errant non-zero signals for stations outside of the deforming zone, their spatial characteristics suggest these signals are primarily from local noise processes that are not removed in the weighting. Overall, the weighting greatly improves the spatial consistency of the modeled transient signal. The reconstructed transient time series corresponding to time scales of approximately four months show that most time-correlated signals that are not persistent over multiple stations are smoothed over in the spatially-weighted solution, thus isolating the signal due to the thrust event (Figure 2.11).

## **2.5 Slow Slip Events in Cascadia**

Continuous GPS measurements above the Cascadia subduction zone have revealed episodic slow slip events located deep on the plate interface that are accompanied by subduction-related tremor signals [1, 5]. These slow slip events exhibit a fairly per-

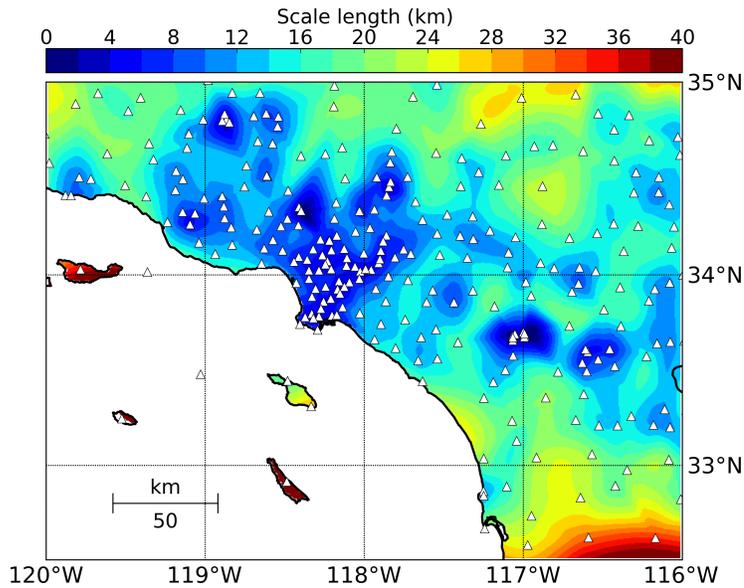


Figure 2.9: Computed scale length (correlation length) of the GPS network used in the SCEC Phase III validation exercise. Scale length is computed using the average distance from each station to the nearest three stations.

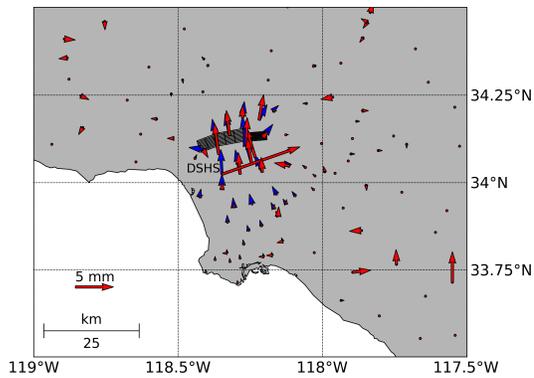


Figure 2.10: Reconstructed transient signal (red arrows) corresponding to simulated thrust event (blue arrows). Most of the signal is isolated close to the fault patches, although a few stations outside of the basin region show extraneous transients. The large east error at station DSHS is due to a large time-correlated noise signal.

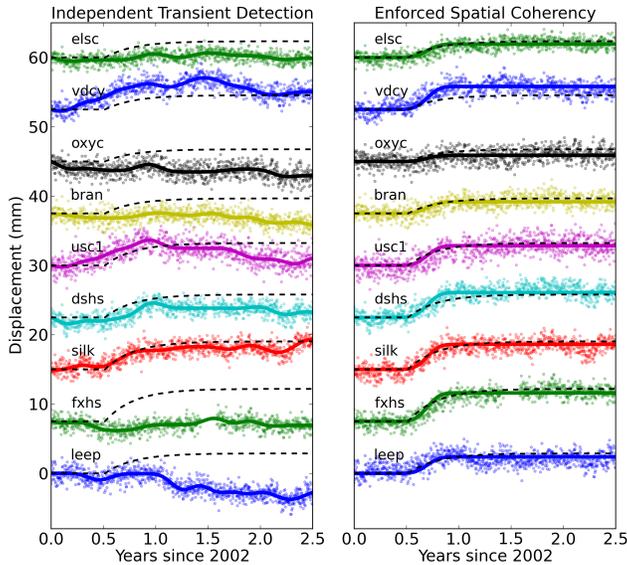


Figure 2.11: Reconstructed north component transient time series for SCEC Phase III Set D stations near the Santa Monica fault. The data (dots) have the estimated secular and seasonal signals removed to match the reconstructed signal (bold line). Dashed black lines show the true ground signal for each station. Without spatial weighting (left), the transient signal is corrupted by time-correlated colored noise. With spatial weighting (right), the signal from the thrust event is isolated.

sistent quasi-periodicity in this region ( $\approx 14$  months) and inform our understanding of the fault physics and frictional properties through the slow slip location, amplitude, and timing. However, the periodicity and amount of slip for each transient event are both spatially and temporally variable and *a priori* unknown. Inference of these slow slip properties can be obtained with precise measurements of the surface deformation field over time.

To test the temporal transient detection capabilities of our proposed method, we use daily GPS solutions for the east component of station ALBH located within the Pacific Northwest Geodetic Array (PANGA). The data cover the timespan from 2005 to mid-2012 and were processed by the Scripps Orbit and Permanent Array Center (SOPAC) with regional filtering applied to remove common mode errors [12]. Known offsets due to hardware changes were removed before analysis. For this work, we examine the east component of the data since the surface deformation in this area occurs primarily in an east-west fashion. Manual inspection of the time series reveals at least six distinct slow slip events. As in the synthetic example, the dictionary  $\mathbf{G}$  consists of seasonal, secular, and transient displacement functions. We uniformly subdivide the 7.5-year time series into 256, 128, 64, 32, 16, 8, and

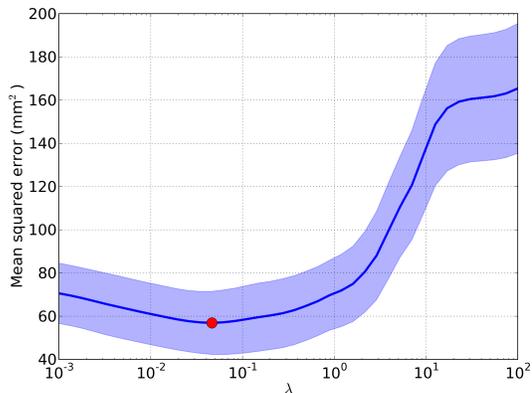


Figure 2.12:  $K$ -fold cross-validation results for selection of the penalty parameter  $\lambda$  for the east-component ALBH time series. Cross-validation was performed on three independent, equally-sized subsets. The blue line shows the mean testing error and the shaded region denotes the standard deviation.

4 intervals to construct  $B^i$ -splines of timescales of  $\approx 3, 6, 12, 25, 52, 113,$  and  $263$  weeks, respectively. Slow slip events typically have recorded durations of 3 weeks [1]. Both the regularized least squares approach and the Gibbs sampler are used to estimate the dictionary coefficients, where the latter is used to construct the full posterior distribution. Cross-validation was performed to select the optimal penalty parameter  $\lambda$  (Figure 2.12). The increased number of transient events recorded in the ALBH time series favored a smaller penalty parameter than the synthetic data.

As was seen with the synthetic time series case, a challenge for transient detection is correctly estimating the contribution from seasonal and secular processes. Our estimates of these steady processes ultimately govern the magnitude of the detected transient events and their overall interpretation in terms of onset times and durations. Using our approach, removal of the estimated seasonal and secular displacements results in a modeled transient evolution with distinct, step-like motions corresponding to the slow-slip events (Figure 2.13a). By comparing the structure of these transients with those presented in previous studies ([e.g., 5]), we can observe a distinctly different interpretation of the relative contributions from the secular rate and transient events. Here, removal of the estimated secular rate results in slow slip events that are modeled as displacement phenomena that occur in an otherwise quiescent field. Other studies have presented the transient time series in a more sawtooth-like fashion in which the station moves slowly eastward in the inter-event period (superposed on the long-term plate rate) before moving rapidly westward. In our framework, such an interpretation would require a positive, longer duration  $B^i$ -

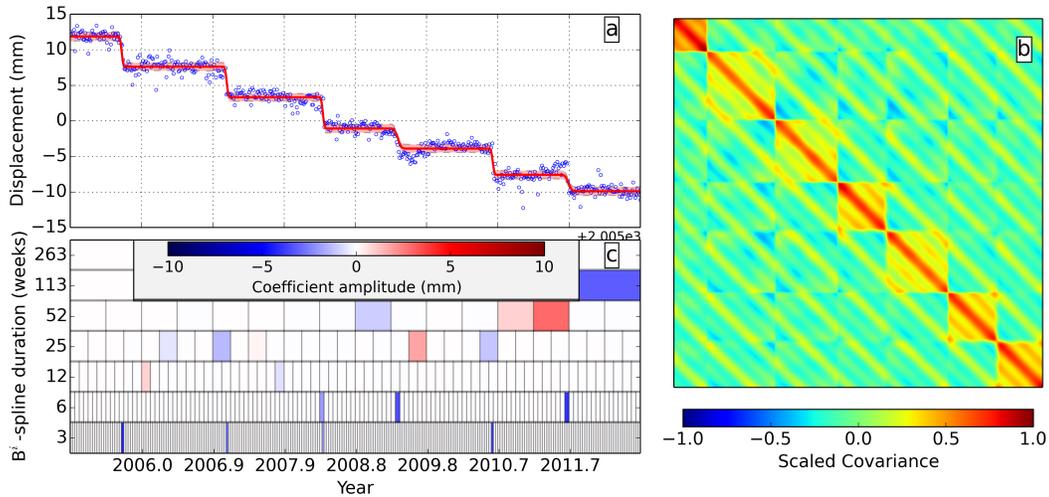


Figure 2.13: Transient detection results for the east-component ALBH time series. (a) GPS data with estimated seasonal and secular signals removed (black circles) and estimated transient signal from a reduced dictionary corresponding to a 99% variance reduction (red line). Shaded area denotes  $5\text{-}\sigma$  uncertainties on predicted displacement using the method in Section 2.2. (b) Posterior data covariance matrix of the time series fit scaled to unity. (c) Scalogram for all  $B^i$ -spline coefficients. Comparison with the time series shows the direct correspondence between the slow slip events and the non-zero 3- and 7-week  $B^i$ -splines. No strong long-term transient signals are present.

spline before each event, followed by a short duration, negative  $B^i$ -spline to model the actual event. This model would be unfavorable due to the sparse regularization which penalizes the total number of non-zero  $B^i$ -splines. The optimal model is the one with the most compact representation, which we can confirm from the scalogram of the  $B^i$ -spline coefficients (Figure 2.13c). In other words, the sparsity-promoting regularization allows us to automatically estimate the inter-event secular rate rather than an average rate that mixes the contributions from the secular rate and transient processes.

Here, all of the slow slip events are modeled either by 3- or 7-week  $B^i$ -splines, although increasing the penalty  $\lambda$  would start to over-smooth the data by selecting longer 14-week  $B^i$ -splines. The posterior data covariance, obtained using the procedure outlined in Section 2.2, shows stronger covariances between modeled displacements in the inter-event period, which is a consequence of the finite support of the  $B^i$ -splines (Figure 2.13b). Similarly, the striping in the off-diagonal terms during the inter-event period is also due to the construction of the temporal dictionary since the modeled value at one observation epoch will covary with the other

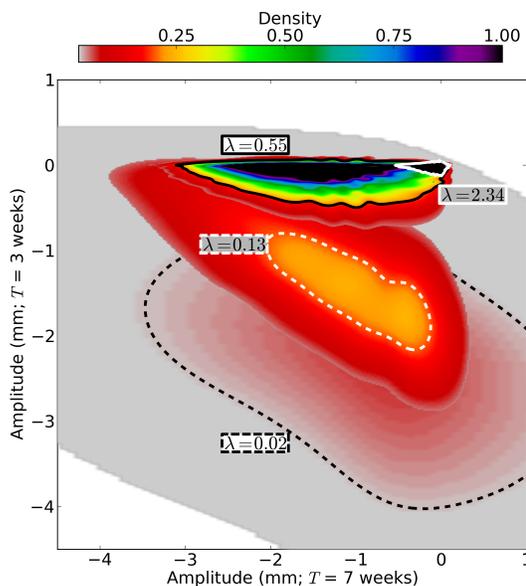


Figure 2.14: Evolution of the posterior covariance with penalty  $\lambda$  between two centroid-coincident  $B^i$ -splines corresponding to the January 2007 slow slip event. Solid and dashed lines mark lines of constant posterior density in the following manner: black dashed  $\rightarrow$  0.03, white dashed  $\rightarrow$  0.2, black solid  $\rightarrow$  0.2, white solid  $\rightarrow$  0.95. For small  $\lambda$ , the joint distribution is diffuse with larger amplitude samples drawn for the shorter timescale  $B^i$ -spline. As  $\lambda$  increases, the distribution shrinks and moves along the vertical axis. Eventually, the distribution transitions to larger values for the longer timescale  $B^i$ -spline and small values for the shorter timescale  $B^i$ -spline. For the largest  $\lambda$ , the samples are tightly clustered around the origin.

observations depending on the  $B^i$ -spline coefficients.

### Dictionary Covariances

For all slow-slip events, there are strong covariances between  $B^i$ -splines that have time-coincident centroids, and the strongest covariances are between 3-week and 7-week  $B^i$ -splines. As with the synthetic data, the strength of the covariance is primarily determined by the value of  $\lambda$  which moves the areas of high posterior probability along a fairly defined trajectory. The covariance behavior between 3-week and 7-week  $B^i$ -splines with time centroids corresponding to a slow-slip event in January 2007 shows that for low values of  $\lambda$ , the joint posterior distribution is fairly diffuse and exhibits large variances with the probability peaking over the shorter-timescale  $B^i$ -spline (Figure 2.14). However, we can still observe a tail in the joint distribution that points towards a non-zero value for the longer-timescale  $B^i$ -spline. Increasing  $\lambda$  is equivalent to shrinking the prior densities for all  $B^i$ -splines, causing the joint distribution to move along the vertical axis defined by



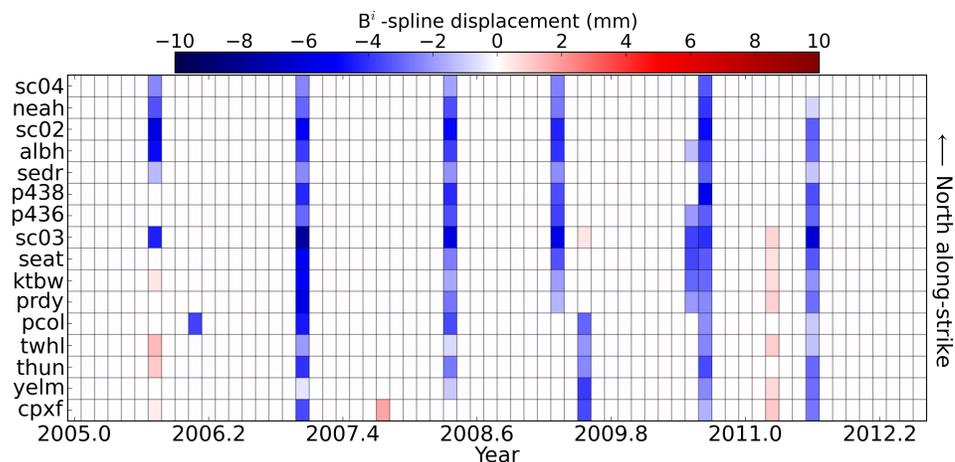


Figure 2.16: Amplitudes corresponding to 3-month  $B^i$ -splines used to fit east component time series data from Cascadia GPS stations. Stations are arranged by increasing latitude, and spatial sparsity weighting was applied to isolate spatially coherent  $B^i$ -splines. The episodic slow slip events are easily visualized, and longer time scale propagation effects can be seen in two of the events (mid-2009 and late-2010).

and seasonal and secular processes; 2) an analysis phase where we select only a year of data encompassing the detected slow slip events and the dictionary is populated exclusively with  $B^i$ -splines ( $\tau_k > 5$  days). Prior to the analysis phase, we remove the estimated signals from seasonal and secular processes, as well as detected transients with timescales longer than one year. While the  $\approx 1 - 3$  month  $B^i$ -splines used in the detection phase will over-smooth many slow slip events, we still gain information about the dominant event durations and onset times (Figure 2.16). The detailed analysis phase can then reveal any subtle propagation behaviors of detected slow slip events and constrain the spatial extent of the transient surface strain. In cases where the data size of the time series are relatively small, the detection phase may be skipped. We utilize it for the Cascadia data because the time scales of the slow slip events are considerably shorter than the time extent of the data.

For the analysis phase, we focus on three slow slip events: 1) January 2007; 2) mid-2008; and 3) mid-2009 (Figure 2.17). For all three events, we can observe several characteristics common to all events. First, slow slip tends to nucleate on the eastern Olympic peninsula near the Seattle area. The slow slip front (determined by the GPS stations' peak ground velocities) propagates bi-laterally with total ground motion larger at stations higher than  $\approx 47.5^\circ$  N. The southern stations tend to show transient motion near the cessations of the slow slip events, which may suggest a change in

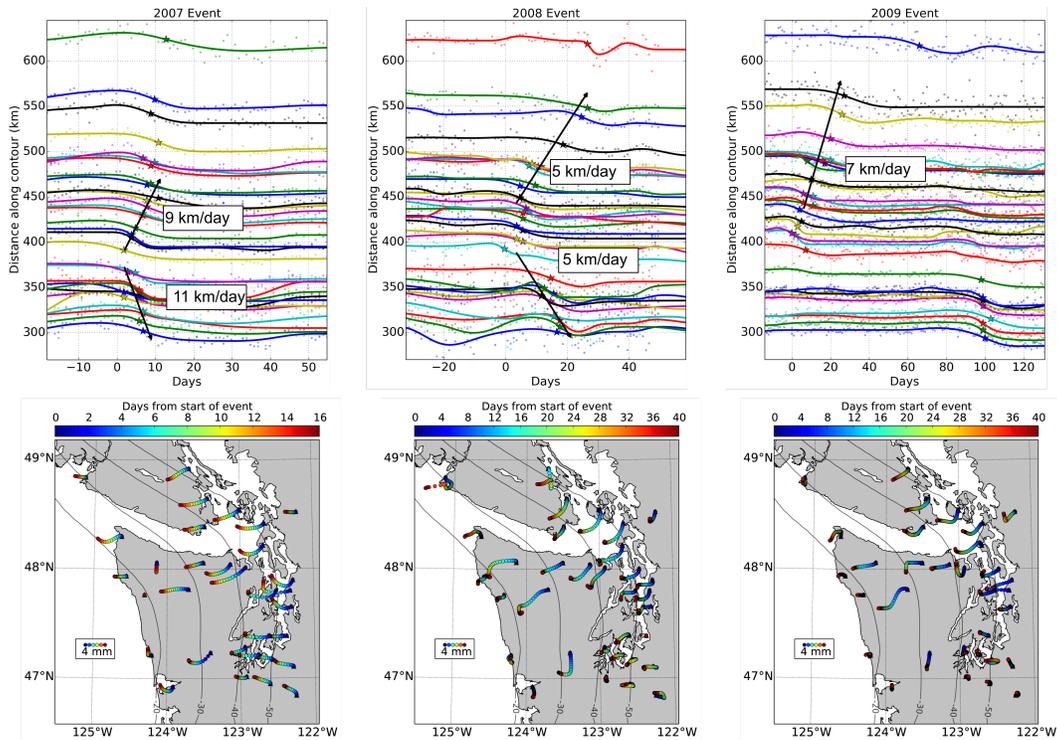


Figure 2.17: Transient detection with spatial sparsity weighting for 33 GPS stations within the Cascadia region. Three separate slow slip events are shown: (*left*) January 2007; (*middle*) mid-2008; and (*right*) mid-2009. The top panels show the normalized east-component displacements corresponding to the slow slip events. The displacement time series are ordered in distance along a 40 km depth contour. The bottom panels show the transient ground motion during each slow slip event where the marker color indicates days from the start of the event.

frictional properties or fault geometry around  $47^\circ$  N. For the 2007 event, the GPS stations show distinctly different displacement azimuths with larger overall velocities (larger displacements in a shorter time period), which may indicate a different slipping area in the underlying thrust fault [45]. Additionally, the propagation speed is significantly faster than either the 2008 or 2009 events. Ground motions for the 2008 and 2009 events initiate at a higher latitude, and the northern stations exhibit stronger southward motion in the first few days of each event. Similarly, station P418 in the southern section shows strong southward motion in the first 10 days of each event. Comparison with previous studies estimating static slip for the 2007 and 2008 events (e.g., [45, 46]) reveals that the areas of peak static slip on the underlying fault are closely located to GPS stations with the first ground motions for each event. This relationship suggests that the nucleation zones on the fault experience the highest cumulative slip. For all three events, we can also observe that many of the stations

cease their north-south motions in the second half of the event and move primarily in an east-west fashion.

## 2.6 Discussion

Generally, transient signal reconstructions are remarkably consistent between the least squares optimization with sparsity-inducing regularization and the Bayesian sampling approach with a Laplace prior. Moreover, the Bayesian approach does not explicitly implement re-weighting for enhancing sparsity as mentioned in Section 2.2 and in Candés *et al.* [35] and can be viewed as solving a single iteration of the sparsity-regularized least squares problem. Even so, the scalogram comparison between the two different approaches (Figure 2.6) confirms that Gibbs sampling can recover the same  $B^i$ -splines corresponding to transient signals while zeroing out the contributions from other  $B^i$ -splines. The fact that we use the mean of the posterior distribution rather than the mode as a point estimate is because we have assigned a squared-error “loss-function” to the data misfit [36]. In a study by Hans [47], it was shown that the posterior mode for a Laplace prior can be interpreted as a limiting case corresponding to a zero-one loss function, which resembles the  $\ell_0$ -norm penalty. However, using the mode does not capture the best point estimate when the marginal posteriors are skewed, as was observed in Figure 2.7. While the penalty  $\lambda$  controls the degree of skewness of the marginal posteriors, it is always optimal to use the mean as a point estimate when a Gaussian data misfit is used.

In Section 2.3, we demonstrated the successful detection of multiple transient signals with a relatively sparse data set. In practical applications, these results suggest that a minimum of  $\approx 3 - 4$  data points spanning the transient process would be required to recover its signal. For daily GPS solutions, we could therefore potentially recover very rapid processes, provided that the signal-to-noise ratio was high enough and the dictionary contained  $B^i$ -splines with equally short timescales. On the opposite end of the spectrum, time series with much coarser temporal resolution, such as an InSAR time series, could still detect transient signals of durations comparable to the time spacing between data acquisitions. For both of these cases, the penalty parameter could be chosen using  $K$ -fold cross-validation, which was demonstrated to favor lower penalties to allow for robust reconstruction of very small signals.

When two  $B^i$ -splines are centered over the same time epoch and can both fit the data reasonably well (as in Section 2.5), lower values of the penalty parameter  $\lambda$  will tend to allow larger amplitudes for the shorter-timescale  $B^i$ -spline, allowing the

reconstructed transient signal to contain higher frequency components. Increasing  $\lambda$  is equivalent to shrinking the prior probabilities of the  $B^i$ -splines, forcing the solution to favor the longer-timescale  $B^i$ -spline and smoothing the reconstructed signal. For real geodetic time series, this behavior implies that successful detection of very short-duration transient signals would most likely require a fairly low value for  $\lambda$  which could be chosen through  $K$ -fold cross-validation. The side-effect of using a small  $\lambda$  is that the reconstructed signal would have a higher probability of being corrupted by colored noise processes in the data. If the signal of interest is known to have a longer duration than typical colored noise, then one could safely choose a higher value of  $\lambda$  to favor longer-timescale  $B^i$ -splines. An alternative approach would be to construct the dictionary such that it included only longer-timescale  $B^i$ -splines, limiting the effective data resolution of the model.

The spatial sparsity weighting discussed in Section 2.4 was effective for minimizing the effects of local ground motion and colored noise. However, the performance of the spatial weighting is inherently dependent on the density of the network and the value of the correlation length used in the distance weighting scheme. Larger correlation lengths will tend to reconstruct long-wavelength deformation fields while smaller correlation lengths will reconstruct more local ground motions. The variable correlation length approach used in this work has the advantage of only reconstructing signals that are resolvable by the geodetic network and minimizing false detections of spurious transients due to data noise. Our approach of prescribing the correlation length,  $L_0^j$ , to be the average distance from station  $j$  to the nearest three or four stations will be affected by the addition of a new station in the vicinity of station  $j$ . However, the addition of stations can only decrease the correlation length. From Equation 2.7, the weighting function would decrease in strength at a given distance, corresponding to an increase in the effective spatial resolution at the current location. Likewise, station removals would increase the correlation length and decrease the effective spatial resolution. For both cases, changes in the station distribution would only affect the spatial reconstructions in the vicinity of the station addition/removal, and the solutions for the remaining stations in the network will be unaffected. In terms of solution stability, the weighting scheme is thus stable and adaptive to changes in network geometry.

The spatial weighting can also be formulated in a Bayesian sense by using the cascading approach discussed earlier and in Minson *et al.* [42]. We can write the posterior distribution of the  $B^i$ -spline coefficients for the  $k$ -th GPS station as

$p(\mathbf{m}_k|\mathbf{d}_k) \propto p(\mathbf{d}_k|\mathbf{m}_k)p(\mathbf{m}_k)$ . By assuming that the terms in  $\mathbf{m}_k$  are correlated with the same terms in  $N$  surrounding stations, we can setup a joint estimation problem where  $\mathbf{M} = [\mathbf{m}_1, \dots, \mathbf{m}_N]$  and  $\mathbf{D} = [\mathbf{d}_1, \dots, \mathbf{d}_N]$ . Then, for the  $k$ -th station, where  $k \leq N$ , the joint posterior distribution would be:

$$\begin{aligned} p(\mathbf{m}_k|\mathbf{D}) &\propto \left[ \prod_{j=1, j \neq k}^N p(\mathbf{d}_j|\mathbf{m}_j)p(\mathbf{m}_j) \right] p(\mathbf{d}_k|\mathbf{M})p(\mathbf{M}) \\ &\propto \left[ \prod_{j=1, j \neq k}^N p(\mathbf{m}_j|\mathbf{d}_j) \right] p(\mathbf{d}_k|\mathbf{M})p(\mathbf{M}). \end{aligned} \quad (2.8)$$

The prior distribution  $p(\mathbf{M})$  would account for spatial coherency between  $B^i$ -spline amplitudes by incorporating a prior covariance matrix with non-zero off-diagonal components for the elements in  $\mathbf{M}$ . As before, the structure of the prior covariance matrix would depend on some form of distance weighting between stations and would be re-computed for each  $k$ -th station. Then, independent Bayesian sampling runs performed for each station would be combined to form the product in brackets in Equation 2.8, and the final posterior distribution  $p(\mathbf{m}_k|\mathbf{D})$  would be sampled by constructing an appropriate likelihood function  $p(\mathbf{d}_k|\mathbf{M})$ .

The reconstructed transient motions and propagation characteristics of the Cascadia slow slip events agree well with independent studies of tremor space-time propagation [e.g., 48]. The geodetic propagation speeds estimated here fall within the range of the tremor propagation speeds, and the nucleation zones for tremor activity correspond well to the first ground motions for each event. Inspection of the reconstructed transient time series for each station (Figures 2.13 and 2.17) indicate that the reconstructions do not suffer from significant oversmoothing. The inclusion of short-timescale  $B^i$ -splines in the temporal dictionary  $\mathbf{G}$  permit us to detect and model very subtle ground motions due to slow slip while the spatial sparsity weighting prevents us from overfitting local ground motions.

Finally, the construction of the transient detection method as a linear model results in very efficient time series processing which can be easily scaled for large geodetic networks. As with any linear model, the efficiency of the linear algebra routines used for performing matrix-vector operations will greatly impact the detection speed and can be enhanced with straightforward parallelization. An alternative to batch estimation of the coefficients is a recursive approach where estimation is performed for each data point in a sequential fashion. The  $\ell_1$ -norm penalty would be enforced as a pseudo-measurement with an associated Kalman gain that tracks the  $\ell_1$ -norm

of the current state of coefficients [49]. A recursive approach would require storing only a single row of the dictionary  $\mathbf{G}$  into computer memory, limiting the computational cost of the estimation and allowing an arbitrarily large number of dictionary elements. Furthermore, the update would be very fast for a single observation and would provide a real-time transient detection capability.

## 2.7 Conclusions

We demonstrated successful transient detection for a single geodetic time series by estimating the coefficients corresponding to a highly overcomplete dictionary (design matrix) of integral B-splines that resemble transient events of various timescales and start times. Regularizing the estimation procedure with an  $\ell_1$ -norm on the coefficients favors *sparse* solutions, limiting the number of B<sup>*i*</sup>-splines needed to describe transient events while still providing a good fit to the data. For GPS networks with sufficient station density, we can perform the regularization simultaneously across the whole network with a distance weighting procedure to enhance signals that are spatially coherent over a given length-scale. The reconstructed time series essentially resemble smoothed versions of the input data but with additional critical information regarding transient event start times and durations. The temporal resolution of the detection method is only limited by the signal-to-noise ratio of the data and the smallest timescale B<sup>*i*</sup>-splines included in the dictionary. The spatial resolution is then limited by the density of the geodetic network. Detection sensitivity is controlled by the penalty parameter on the  $\ell_1$ -norm, which can be robustly and automatically chosen with cross-validation. We also presented a Gibbs sampling approach to construct the full posterior distribution of each element in the dictionary which, in addition to stand-alone point estimates of the coefficients via the posterior means, provides quantifiable uncertainties on the coefficients and valuable insight into trade-offs between dictionary elements. In the absence of *a priori* knowledge about transient event start times, durations, and physical sources, this method automatically and efficiently determines the most dominant signals in a time series in a compact and interpretable manner.

## References

1. Rogers, G. & Dragert, H. Episodic Tremor and Slip on the Cascadia Subduction Zone: The Chatter of Silent Slip. *Science* **300**, 1942–1943 (June 2003).

2. McGuire, J. J. & Segall, P. Imaging of Aseismic Fault Slip Transients Recorded by Dense Geodetic Networks. *Geophysical Journal International* **155**, 778–788 (2003).
3. Gombert, J. *et al.* Slow-slip phenomena in Cascadia from 2007 and Beyond: A Review. *GSA Bulletin* **122**, 963–978 (July 2010).
4. Kato, A., Obara, K., Igarashi, T., Tsuruoka, H., Nakagawa, S. & Hirata, N. Propagation of Slow Slip Leading Up to the 2011  $M_w$  9.0 Tohoku-Oki Earthquake. *Science* **335**, 705–708 (2012).
5. Szeliga, W., Melbourne, T., Santillan, M. & Miller, M. GPS Constraints on 34 Slow Slip Events Within the Cascadia Subduction Zone, 1997-2005. *Journal of Geophysical Research* **113** (2008).
6. Ji, K. H. & Herring, T. A. Transient Signal Detection Using GPS Measurements: Transient Inflation at Akutan Volcano, Alaska, during early 2008. *Geophysical Research Letters* **38** (2011).
7. Langbein, J. O. Deformation of the Long Valley Caldera, California: Inferences from Measurements from 1988 to 2001. *Journal of Volcanology and Geothermal Research* **127**, 247–267 (2003).
8. Masterlark, T. & Lu, Z. Transient Volcano Deformation Sources Imaged with Interferometric Synthetic Aperture Radar: Application to Seguam Island, Alaska. *Journal of Geophysical Research* **109** (2004).
9. Pritchard, M. & Simons, M. An InSAR-based survey of volcanic deformation in the central Andes. *Geochemistry, Geophysics, Geosystems* **5** (2004).
10. Miyazaki, S., McGuire, J. J. & Segall, P. A Transient Subduction Zone Slip Episode in Southwest Japan Observed by the Nationwide GPS Array. *Journal of Geophysical Research* **108** (2003).
11. Ji, K. H. & Herring, T. A. A Method for Detecting Transient Signals in GPS Position Time-Series: Smoothing and Principal Component Analysis. *Geophysics Journal International* **193** (2013).
12. Williams, S. D. *et al.* Error analysis of continuous GPS position time series. *Journal of Geophysical Research* **109**. doi:10.1029/2003JB002741 (2004).
13. Sagiya, T. A Decade of GEONET: 1994-2003-The Continuous GPS observation in Japan and its impact on earthquake studies. *Earth Planets and Space* **56** (2004).
14. Hsu, Y.-J., Yu, S.-B., Simons, M., Kuo, L.-C. & Chen, H.-Y. Interseismic crustal deformation in the Taiwan plate boundary zone revealed by GPS observations, seismicity, and earthquake focal mechanisms. *Tectonophysics* **479**, 4–18 (2009).

15. Hung, H.-K. & Rau, R.-J. Surface waves of the 2011 Tohoku earthquake: Observations of Taiwan's dense high-rate GPS network. *Journal of Geophysical Research: Solid Earth* (2013).
16. Wallace, L. M. & Beavan, J. Diverse slow slip behavior at the Hikurangi subduction margin, New Zealand. *Journal of Geophysical Research* **115** (2010).
17. Covello, F. *et al.* COSMO-SkyMed an existing opportunity for observing the Earth. *Journal of Geodynamics* **49**, 171–180 (2010).
18. Hetland, E., Musé, P., Simons, M., Lin, Y., Agram, P. & DiCaprio, C. Multi-scale InSAR Time Series (MInTS) Analysis of Surface Deformation. *Journal of Geophysical Research* **117** (2012).
19. Agram, P. *et al.* New Radar Interferometric Time Series Analysis Toolbox Released. *Eos, Transactions American Geophysical Union* **94**, 69–70 (2013).
20. Zhang, J. *et al.* Southern California Permanent GPS Geodetic Array: Error Analysis of Daily Position Estimates and Site Velocities. *Journal of Geophysical Research* **102**, 18035–18055 (Aug. 1997).
21. Langbein, J. Noise in Two-Color Electronic Distance Meter Measurements Revisited. *Journal of Geophysical Research* **109** (2004).
22. Dong, D. *et al.* Spatiotemporal Filtering Using Principle Component Analysis and Karhunen-Loeve Expansion Approaches for Regional GPS Network Analysis. *Journal of Geophysical Research* **111** (2006).
23. Williams, S., Bock, Y. & Fang, P. Integrated Satellite Interferometry: Tropospheric Noise, GPS Estimates and Implications for Interferometric Synthetic Aperture Radar Products. *Journal of Geophysical Research* **103**, 27051–27067 (1998).
24. Berardino, P., Fornaro, G., Lanari, R. & Sansosti, E. A new algorithm for surface deformation monitoring based on small baseline differential SAR interferograms. *Geoscience and Remote Sensing, IEEE Transactions on* **40**, 2375–2383 (2002).
25. Jolivet, R., Grandin, R., Lasserre, C., Doin, M.-P. & Peltzer, G. Systematic InSAR tropospheric phase delay corrections from global meteorological reanalysis data. *Geophysical Research Letters* **38** (2011).
26. Mallat, S. A Theory for Multiresolution Signal Decomposition: The Wavelet Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11**, 674–693 (1989).
27. Candés, E. J. & Wakin, M. B. An Introduction to Compressive Sampling. *IEEE Signal Processing Magazine* **21** (Mar. 2008).
28. Donoho, D. L. For Most Large Underdetermined Systems of Linear Equations the Minimal  $\ell^1$ -norm Solution is also the Sparsest Solution. *Communications on Pure and Applied Mathematics* **59**, 797–829 (2006).

29. Tibshirani, R. Regression Shrinkage and Selection Via the Lasso. *J. R. Stat. Soc. B* **58**, 267–288 (1996).
30. Chen, S. S., Donoho, D. L. & Saunders, M. A. Atomic Decomposition by Basis Pursuit. *SIAM Journal on Scientific Computing* **20**, 33–61 (1998).
31. Boyd, S. & Vandenberghe, L. *Convex Optimization* (Cambridge University Press, 2004).
32. Taylor, H., Banks, S. & McCoy, J. Deconvolution with the  $\ell_1$ -norm. *GEO-PHYSICS* **44**, 39–52 (1979).
33. Gholami, A. & Siahkoohi, H. Regularization of Linear and Non-linear Geophysical Ill-posed Problems With Joint Sparsity Constraints. *Geophysical Journal International* **180**, 871–882 (2010).
34. Evans, E. L. & Meade, B. J. Geodetic imaging of coseismic slip and postseismic afterslip: Sparsity promoting methods applied to the great Tohoku earthquake. *Geophysical Research Letters* **39** (2012).
35. Candès, E. J., Wakin, M. B. & Boyd, S. P. Enhancing Sparsity by Reweighted  $\ell_1$  Minimization. *Journal of Fourier Analytical Applications* **14**, 877–905 (2008).
36. Tarantola, A. *Inverse Problem Theory and Methods for Model Parameter Estimation* (Society for Industrial and Applied Mathematics, 2005).
37. Gelman, A., Carlin, J. B., Stern, H. S. & Rubin, D. B. *Bayesian Data Analysis* chap. 11 (Chapman & Hall/CRC, 2004).
38. Park, T. & Casella, G. The Bayesian Lasso. *Journal of the American Statistical Association* **103:482**, 681–686 (2008).
39. Beck, J. L. & Yuen, K.-V. Model Selection Using Response Measurements: Bayesian Probabilistic Approach. *Journal of Engineering Mechanics* **130** (Feb. 2004).
40. Ching, J., Muto, M. & Beck, J. Bayesian Linear Structural Model Updating Using Gibbs Sampler with Modal Data. *Computer-Aided Civil and Infrastructural Engineering* **21**, 242–257 (May 2006).
41. Tipping, M. E. Bayesian Inference: An Introduction to Principles and Practice in Machine Learning. *Advanced Lectures on Machine Learning* (2004).
42. Minson, S., Simons, M. & Beck, J. Bayesian inversion for finite fault earthquake source models I — theory and algorithm. *Geophysical Journal International* (2013).
43. Lohman, R. B. & Murray, J. R. The SCEC Geodetic Transient-Detection Validation Exercise. *Seismological Research Letters* **84**, 419–425 (2013).
44. Agnew, D. C. Realistic simulations of geodetic network data: The Fakenet package. *Seismological Research Letters* **84**, 426–432 (2013).

45. Wech, A. G., Creager, K. C. & Melbourne, T. I. Seismic and geodetic constraints on Cascadia slow slip. *Journal of Geophysical Research: Solid Earth (1978–2012)* **114** (2009).
46. Aguiar, A. C., Melbourne, T. I. & Scrivner, C. W. Moment Release Rate of Cascadia Tremor Constrained by GPS. *Journal of Geophysical Research: Solid Earth (1978–2012)* **114** (2009).
47. Hans, C. Bayesian Lasso Regression. *Biometrika* **96**, 835–845 (2009).
48. Houston, H., Delbridge, B. G., Wech, A. G. & Creager, K. C. Rapid tremor reversals in Cascadia generated by a weakened plate interface. *Nature Geoscience* **4**, 404–409 (2011).
49. Carmi, A., Gurfil, P. & Kanevsky, D. Methods for Sparse Signal Recovery Using Kalman Filtering With Embedded Pseudo-Measurement Norms and Quasi-Norms. *IEEE Transactions On Signal Processing* **58** (Apr. 2010).

## THE COLLAPSE OF BÁRÐARBUNGA CALDERA, ICELAND

1. Riel, B., Milillo, P., Simons, M., Lundgren, P., Kanamori, H. & Samsonov, S. The Collapse of Bárðarbunga Caldera, Iceland. *Geophysical Journal International* **202**. doi:10.1093/gji/ggv157 (2015).

### 3.1 Introduction

On August 16, 2014, a swarm of earthquakes was detected underneath Bárðarbunga caldera in Iceland, a stratovolcano located within the Eastern Volcanic Zone (EVZ) and completely covered by the Vatnajökull ice cap [1]. These earthquakes signaled the onset of subsurface magma movement. As magma propagated out of the confines of the caldera, earthquake activity tracked its motions, revealing the emplacement of a large dike along a northeast oriented fissure swarm of the Bárðarbunga volcanic system, consistent with a plate boundary rifting event [2, 3]. The dike intrusion triggered an effusive eruption 40 km away from the caldera at the surface of the Holuhraun lava field north of Vatnajökull beginning on Aug. 29, 2014. Prior to the surface eruption, geodetic observations revealed that the ice over the caldera was subsiding rapidly, with measured rates of approximately 50 cm/day (Fig. 3.1). The rapid subsidence was accompanied by moderate earthquakes ( $M_w > 5$ ) with epicenters concentrated along the caldera rim (Fig. 3.2). Nearly all of these earthquakes exhibited anomalous behavior with large deviations from traditional double-couple sources (i.e., motion confined to a shear fault plane) [4]. These earthquakes appear to be the manifestations of simultaneous vertical compression and outward horizontal expansion. Such motion is commonly interpreted as a compensated linear vector dipole (CLVD) [5]. The close spatiotemporal association of the caldera collapse, anomalous seismicity, and large-scale rifting provides a unique opportunity to study the mechanics of a caldera collapse in a basaltic system. The large subsidence within the caldera rim, which has never been previously observed at Bárðarbunga, provides critical constraints on the collapse sequence within the caldera.

Since the start of the eruption, a suite of synthetic aperture radar (SAR) images over northwest Vatnajökull and adjacent regions has been acquired by the international constellation of radar satellites. With these images, we can use interferometric

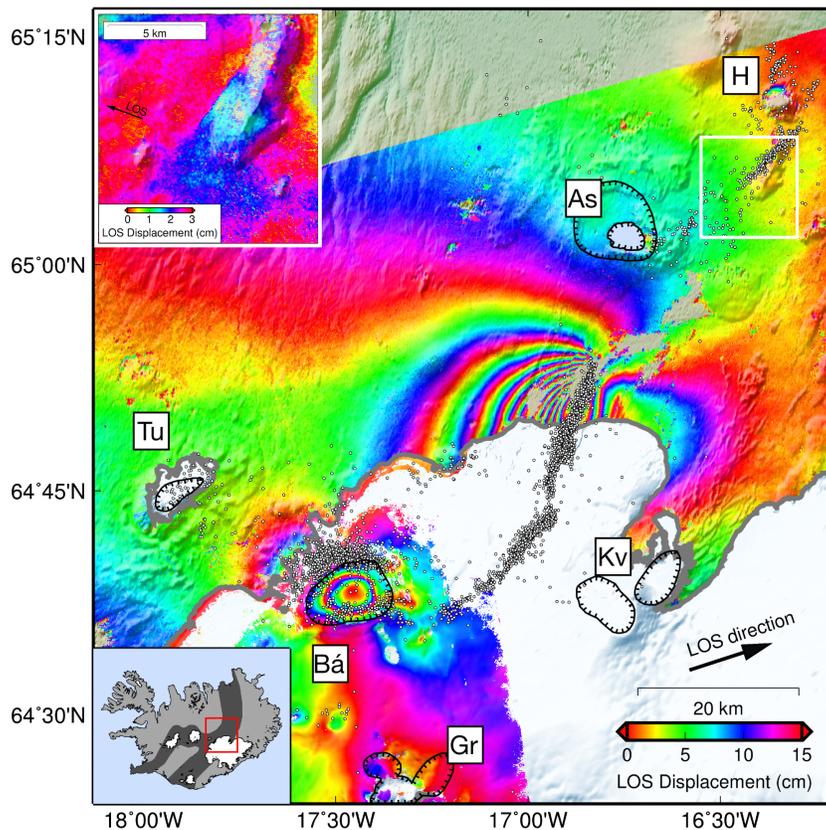


Figure 3.1: Map of northwestern Vatnajökull, Iceland with the locations of known calderas/volcanos indicated by barbed lines: As - Askja, Bá - Bárðarbunga, Gr - Grímsvötn, H - Herðubreið, Kv - Kverkfjöll, Tu - Tungnafellsjökull. Colored fringes represent contours of the line-of-sight (LOS) component of ground motion at 15 cm per color cycle. Areas of high interferometric phase noise have been masked out. The interferogram over the ice was formed from COSMO-SkyMED (CSK) images acquired from Sep. 12-13, 2014 while the 24-day interval interferogram over the ground was formed from RADARSAT-2 (RS2) images acquired on Aug. 8 and Sep. 1, 2014. A clear bullseye pattern over the Bárðarbunga caldera indicates subsidence of the glacier surface. The RS2 interferogram shows the effects of the rifting associated with spreading of the ground away from the active dike. Black arrow indicates the satellite-to-ground direction for both the CSK and RS2 interferograms. White dots indicate earthquakes that occurred between August 15, 2014 and February 1, 2015, as recorded by the SIL network in Iceland. Lower-left inset shows the location of the study area in Iceland with respect to the glaciers and active volcanic zones (dark grey shading). Upper-left inset shows surface rupture near the Herðubreið volcano (denoted by white box) observed in the RS2 interferogram.

SAR (InSAR) to measure surface deformation between two successive SAR images along a line-of-sight (LOS) direction (e.g., Simons & Rosen 2007). We use images acquired by the COSMO-SkyMed (CSK) constellation, which consists of four X-band radar satellites operated by the Italian Space Agency (ASI), to image ground deformation within the vicinity of the Bárðarbunga caldera (Fig. 3.1). One-day separation between CSK images over the collapsing ice-covered caldera permits the formation of high-resolution interferograms with good coherence, providing snapshots of daily subsidence of the overlying ice. We complement the CSK data with 24- and 48-day-interval InSAR observations from RADARSAT-2 (RS2), a C-band satellite operated by the Canadian Space Agency, to measure ground deformation in ice-free regions north of Vatnajökull. This deformation primarily results from emplacement of an intrusive dike, producing 1.5 m of surface ground motion very close to the dike and measurable deformation as far as 60 km away from the surface trace. Near Herðubreið volcano northeast of Askja, a cluster of earthquakes is associated with left-lateral fault motion where discrete centimeter-scale surface rupture can be observed. The left-lateral motion agrees with the previously inferred bookshelf faulting for that area and implies that dike emplacement enhanced the background stress field in the Askja rift segment [2, 6]. In the ice-free areas adjacent to and west of Bárðarbunga caldera, inspection of the LOS displacements in the RS2 data reveals several centimeters of motion consistent with a deflating magma chamber beneath the caldera.

### 3.2 Geodetic and seismic data

We formed five one-day interferograms from CSK images acquired on August 27-28, September 12-13, September 13-14, September 17-18, and October 19-20 (Fig. 3.2). All interferograms show strong subsidence signals within the caldera boundary, presumably due to subsidence of the caldera floor. While melting of the overlying ice at its base could also result in the observed subsidence, there has been no evidence of anomalous glacial outwash or changes in the ice flow rates for the central Bárðarbunga caldera [3]. The first two and the last interferogram show similar, axisymmetric bullseye patterns due to the subsidence of the ice over the caldera. However, only a day after the Sep. 12-13 interferogram, the Sep. 13-14 interferogram shows a distinctly different deformation pattern with greater displacement near the northern rim of the caldera. Because ground displacement in one satellite line-of-sight direction is generally a combination of horizontal and vertical motion, we use the method of Yun et al. (2006) to estimate the ratio of

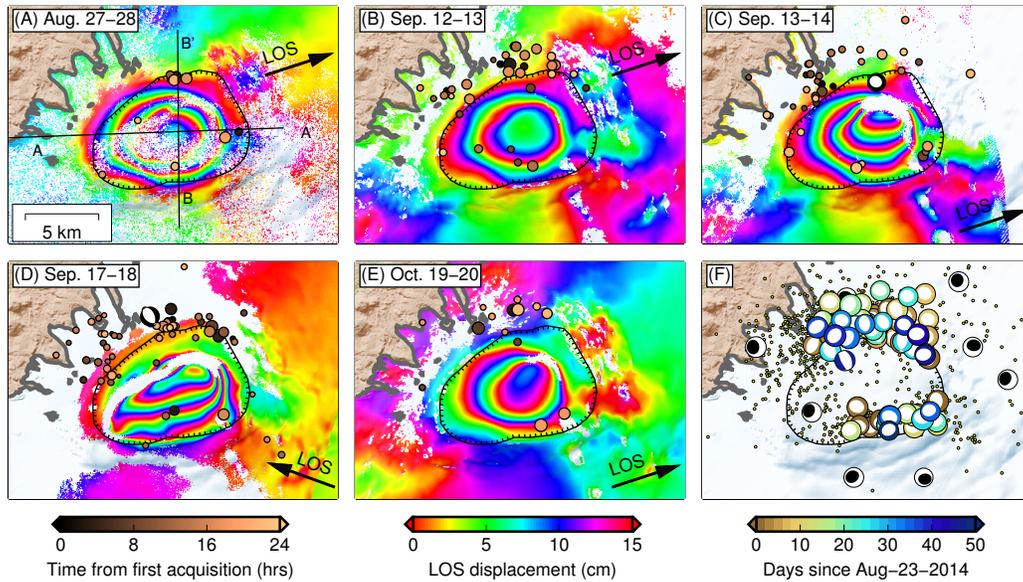


Figure 3.2: (A)-(E) Evolution of subsidence within the caldera observed from five one-day interval CSK interferograms. Note the similarity in LOS directions for the interferograms, with the exception of the Sep. 17-18 interferogram. Colored circles indicate earthquakes that occurred during the time span of the interferogram and are colored by hours elapsed from the acquisition time of the first image of a given pair. Earthquakes with  $M_w > 4.9$  are shown with focal mechanisms derived from moment tensors obtained from the GFZ Potsdam catalog. The location of the transects in Fig. 3.S2 are indicated in the first panel for the August 27-28 interferogram. (F) Distribution of all seismicity near the caldera between August 16 to October 21, 2014. The colored focal mechanisms are colored by days elapsed since August 23, 2014. Yellow dots denote the locations of the earthquakes presented in Fig. 3.1. The black focal mechanisms with vertical tensional axes correspond to selected CLVD events at Bárðarbunga between 1976-1996 [18].

horizontal to vertical motion on the ice by comparing interferograms with different viewing geometries (Fig. 3.S1). We estimate that the ratio of horizontal to vertical motion does not exceed  $\approx 0.3$  within the caldera, and any horizontal motion is diminished close to the center of the subsidence signal. Thus, assuming purely vertical motion, we can extract profiles across the caldera to estimate instantaneous subsidence rate for each of the interferograms (Fig. 3.S2). The northward trending subsidence pattern in the Sep. 13-14 interferogram is clearly associated with a sharp increase in apparent instantaneous subsidence rate over the background rate: 20 cm increase in the center of the caldera and 25 cm on the northern edge of the caldera.

Two earthquakes with  $M_w$  4.9 and  $M_w$  5.3 occurred on the north rim of the caldera

during the time spans of the Sep. 13-14 and Sep. 17-18 interferograms, respectively (Fig. 3.2). These two earthquakes were part of a persistent sequence of  $M_w > 5$  earthquakes along the rim of the caldera with highly non-double-couple characteristics, exhibiting dominant vertical compressional axes and horizontal expansions consistent with CLVDs (Fig. 3.2) [7]. The focal mechanisms for this sequence of events are consistent between different earthquake catalogs (Table 3.S1). While previous studies of global CLVD earthquakes in the vicinity of active volcanoes have shown that such earthquakes tend to have longer-than-average source durations for their magnitudes, estimates of the magnitudes themselves are weakly dependent on the complexity of the source time functions [8, 9]. Therefore, the small differences in moment magnitudes between different catalogs for the current event (Table 3.S1) suggest that the magnitudes are correct to within approximately 0.1 magnitude unit.

The  $M_w$  4.9 event has a clear CLVD component in the moment tensor while the  $M_w$  5.3 event appears to show more normal fault motion. The location, size, and focal mechanisms of these events, and the absence of similar events during the time spans of the Aug. 27-28, Sep. 12-13, and Oct. 19-20 interferograms, suggests that the asymmetry of the Sep. 13-14 and Sep. 17-18 interferograms is linked to the occurrence of the larger earthquakes on the northern rim. A simple relationship between earthquake size and ground displacement, such as assuming seismic potency is proportional to areally integrated displacements of the ice surface, does not appear to apply to these events since the potency associated with the earthquake that occurred during the Sep. 13-14 interval was nearly five times less than the earthquake that occurred during the Sep. 17-18 interval, yet the estimated volume change within the caldera as estimated from the integrated displacements is larger for the former (Fig. 3.3D). Therefore, most of the subsidence occurs aseismically while the larger  $M_w > 5$  events may produce localized additional displacements of up to 25 cm along the caldera rim as measured at the ice surface.

### 3.3 Source models for rift zone and magma chamber

Experimental and numerical studies of caldera collapse consistently show that the size, shape, and depth of subsurface magma chambers strongly affect the final geometry of the collapsed caldera, as well as the rate at which it will form [10–13]. In order to estimate the depth of a subsurface magma chamber while avoiding modeling errors due to uncertain ice-rock coupling, we consider LOS displacements of the subsidence signal from the ice-free areas in interferograms formed from longer time interval RS2 images (24 and 48 days, Fig. 3.S8). We also include data from

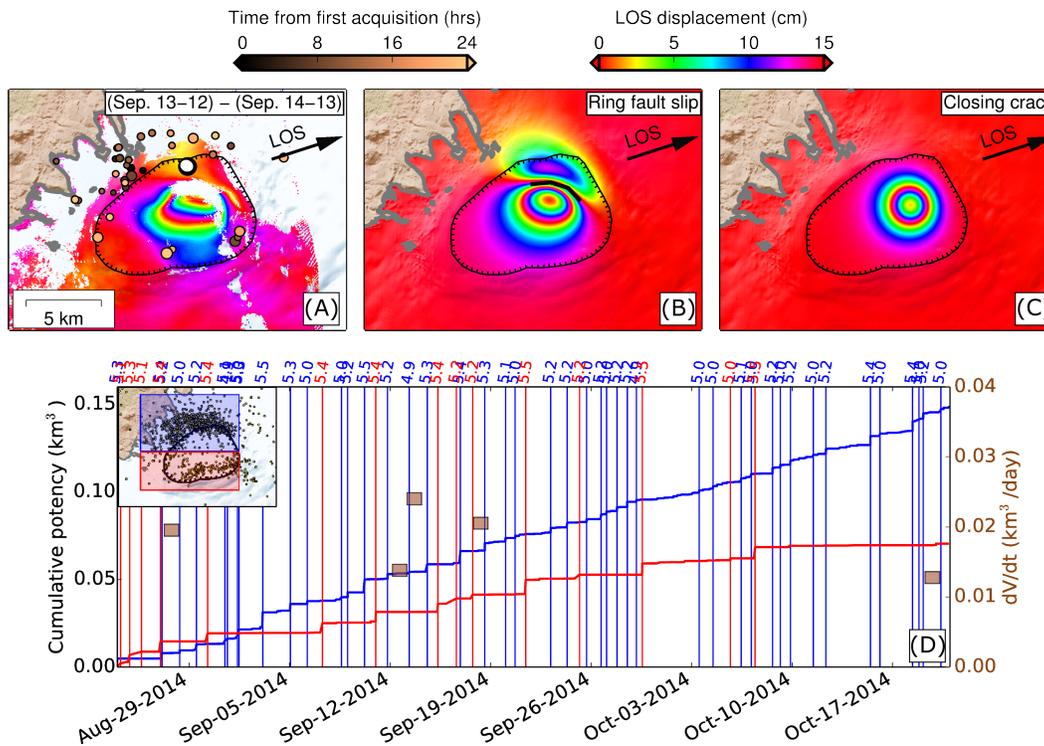


Figure 3.3: (A) Difference between Sep. 12-13 and Sep. 13-14 interferograms for isolating the ground deformation associated with a  $M_w$  4.9 earthquake that occurred on Sep. 14. Also shown are synthetic LOS displacements due to slip along a vertical ring fault (B) and a closing crack (C). For the ring fault (thick black line), the fault is placed at a depth of 2 km with a width of 1 km and approximately 3 meters of normal motion. The closing crack is simulated using the model of Fialko, et al. (2001) at a depth of 2 km, a radius of 1 km, and a volume change of  $0.003 \text{ km}^3$ . (D) Seismic potency for the northern and southern halves of the caldera compared with estimated volume change rate ( $dV/dt$ ) within the caldera inferred from the one-day interferograms. The thick blue and red lines correspond to the cumulative potency of the northern and southern regions, respectively, of the caldera. To calculate seismic potency, we divide the seismic moment by a shear modulus of 24 GPa. Thin vertical lines mark the occurrence of earthquakes greater than  $M_w$  4.9. The shaded brown boxes show the time spans of each one-day interferogram and the estimated volume change during that time span (right ordinate). The volume change for each interferogram was computed by integrating the projected vertical displacements within the caldera boundary.

continuous Global Positioning System (GPS) stations located north of Vatnajökull [3]. As part of the analysis, we remove ground displacements due to the rift zone by estimating an elastic model for tensile opening along the dike interface using a collection of RS2 and CSK interferograms and the GPS data (see Materials and Methods, Fig. 3.S4). While the geodetic data alone cannot resolve opening of the dike under the ice, the robust components of the model include approximately 5 meters of opening in the upper 5 km of the resolvable dike segment (equivalent to 200-300 years of cumulative plate motion) and peak geodetic potency occurring at a shallower depth ( $\approx 2$  km) than the peak earthquake density, which occurs at around 6 km depth (Sigmundsson et al. 2014; Fig. 3.S6). These results agree with estimates of the dike geometry from other studies (e.g., Gudmundsson et al. 2014; Sigmundsson et al. 2014).

After removal of the rift zone signal, we model the chamber as a collapsing horizontal circular crack in an elastic half-space [14](Materials and Methods). The parameterization of the chamber parameters is such that there are strong trade-offs between the chamber depth, radius, and excess pressure (difference between the magma chamber and lithostatic pressures). However, we are able to resolve a consistent depth-to-radius ratio of approximately 3.6 (Fig. 3.S7). By adjusting the excess pressure, we can explain the geodetic observations equally well with a shallow, small chamber or a deep, larger chamber. Therefore, determining the “true” depth of the magma chamber would require other independent observations, i.e. re-located seismicity of earthquakes occurring within and around the caldera or an upper bound on the allowable values for excess pressure. Nevertheless, the ability of our model to fit the displacements at distances greater than three times the chamber radius validates the assumption of a symmetric source with a uniform pressure difference since no obvious asymmetries appear in the residuals. Steady deflation of the chamber is thus the primary contributor to subsidence observations both on and off the ice. While we cannot rule out the possibility of additional deeper magma chambers (depth  $> 10$  km), our lack of reliable ground measurements over the caldera limits our ability to resolve multiple chambers.

### 3.4 Discussion

One of the most interesting aspects of the collapse sequence has been the occurrence of the moderate earthquakes along the caldera rim with large CLVD components in their focal mechanisms. Assuming that the background subsidence rate is nearly constant between Sep. 12-14, we can isolate the ground deformation associated with

one of the earthquakes ( $M_w$  4.9 on Sep. 13) by forming a residual interferogram of the difference between the Sep. 12-13 and Sep. 13-14 CSK interferograms since only the Sep. 13-14 image spanned the time of the earthquake (Fig. 3.3A). The highly localized deformation due to the earthquake indicates that the source is most likely shallow. For earthquakes with vertical CLVD focal mechanisms in the vicinity of volcanoes, the two most likely physical processes are rupture on curved ring faults or opening/closing of a crack under tension/compression [9, 15]. Static simulations for both processes require source depths of approximately 2 km to roughly match the InSAR observations (Fig. 3.3B-C). For reference, a  $M_w$  4.9 event with an effective shear modulus of 5 GPa (typical for shallow depths in volcanic systems within an extensional regime [16]) has a seismic potency of approximately  $0.004 \text{ km}^3$ . However, the simulated ring fault rupture requires a much larger potency ( $\approx 0.01 \text{ km}^3$ ), which implies a very low effective shear modulus ( $\approx 1 \text{ GPa}$ ) consistent with very weak materials such as water-saturated basaltic tuff. While the closing crack provides a more realistic measure of potency (equivalent volume change of about  $0.003 \text{ km}^3$ ), the synthetic ground deformation lacks the asymmetry observed in the residual interferogram. Therefore, for both cases, it is likely that the ground deformation is a combination of seismic deformation (due to ring faulting or a closing crack) and aseismic slip on the ring fault. The inconsistency between earthquake size and ground deformation (Fig. 3.3D) suggests that the amount of aseismic slip per event is highly variable.

Based on the seismic evidence and the inversion results for the chamber geometry, we propose a model for the sequence of the caldera collapse and dike emplacement (Fig. 3.4). The initial seismic activity on the southern edge of the caldera and subsequent propagation of an oblique dike caused depressurization of a magma chamber that can be approximated as a horizontal circular sill. Magma migrated out of the chamber to the short, oblique dike and eventually migrated to the larger, regional-scale dike. The underpressure in the magma chamber resulted in subsidence of the caldera surface and overlying ice. Stress concentrations in the vicinity of the deflating magma chamber led to the initiation of  $M_w > 5$  seismic events located along the caldera rim. The large CLVD components in the earthquake focal mechanisms indicate a seismic process characterized either by downward vertical motions and horizontal expansions or rupture on a curved fault. The former process has been observed in mine collapses where rapid closing of a horizontally-oriented underground cavity (the mine) leads to CLVD components in the seismic moment tensor [17]. A mine collapse mechanism could imply failure of brittle support structures within the

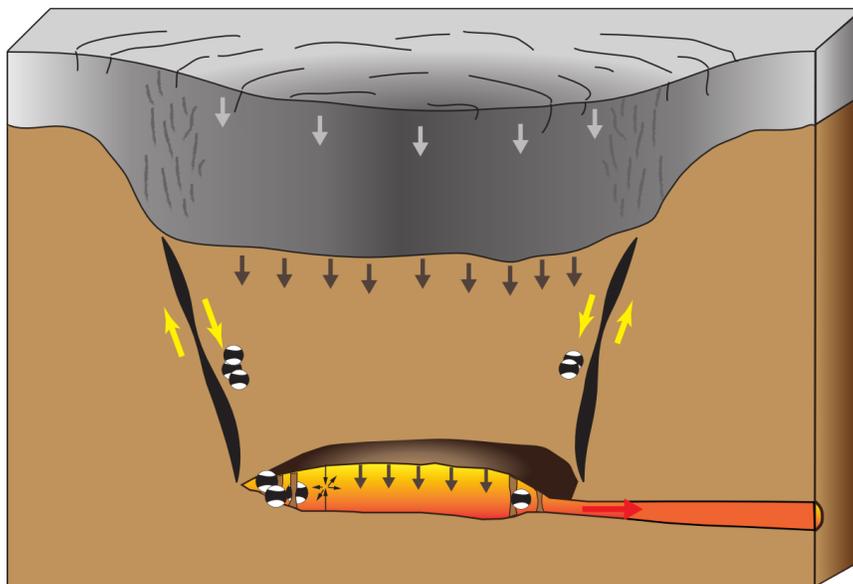


Figure 3.4: A conceptual model for the mechanics of the collapsing caldera. Magma migration out of the chamber into the dike system causes depressurization of the chamber. The deflating chamber results in subsidence of a coherent block above the chamber, as well as subsidence of the overlying ice. Sustained depressurization of the chamber leads to failure of internal support structures (causing seismic events marked by inward vertical motion and outward horizontal expansion) or rupturing of curved, inward-dipping ring faults. Cross-sections of focal mechanisms representative of these events are shown on the edge of the chamber.

partially molten magma chamber due to large compressive stresses, which would require a larger (and thus deeper) magma chamber since the earthquake clusters on the northern and southern rim of the caldera are separated by approximately 8-10 km. Alternatively, the deflating chamber could impart large shear stresses on existing ring faults, leading to seismogenic slip on limited portions of a ring fault. Static stress changes associated with both mechanisms could then trigger larger aseismic slip on the ring fault with slip most likely confined to shallower depths.

Earthquakes with significant non-double-couple components in their focal mechanisms have occurred at Bárðarbunga in the decades prior to the current event (Table 3.S2) [18–20]. However, the polarities of the focal mechanisms for the current Bárðarbunga eruption are opposite to those observed for the earlier earthquakes, i.e. vertical compressional axes for the former and tensional axes for the latter (Fig. 3.2F). One interpretation of these earlier events suggested that their faulting mechanisms are primarily due to rupture on outward dipping ring faults which are activated by inflation of a very shallow magma chamber [18]. However, this interpretation

is highly non-unique, since a deflating magma chamber below the ring faults could also produce similar focal mechanisms [21]. Our inversion results for the magma chamber predict a chamber radius smaller than the caldera radius for chamber depths less than 15 km (Figs. 3.S7,3.S8), which implies inward dipping ring faults. The dip of the ring faults controls the rupture arc-length necessary to create the CLVD focal mechanisms. As an example, ruptures with arc-lengths of  $180^\circ$  would require faults with dips shallower than  $70^\circ$  [21]. For this geometry, the earlier events can be explained by inflation of the magma chamber, leading to reverse motion on those ring faults.

In addition to mine collapses and ring fault rupture, earthquakes with large CLVD components like the ones observed on the rim of the Bárðarbunga caldera have been observed for tensile failure due to high fluid pressure [22] and magma injection in water-saturated environments [23]. An earthquake with a strong CLVD component was observed near Tori Shima, Japan in 1984. Since the earthquake was tsunamigenic, the preferred explanation was horizontal injection of magma into water-saturated sediment with a volumetric component resulting from the explosive magma-water interaction [23]. A global search for earthquakes prior to 2013 with vertical-CLVD mechanisms revealed  $\approx 100$  vertical-CLVD earthquakes located near active volcanoes over the past century [8]. Thus, the close temporal and spatial association of the current Bárðarbunga anomalous events with the active fissuring lends credibility to a collapse mechanism driven by a drop in magma pressure. A sudden vertical collapse with no volume change would be characterized by downward vertical and outward horizontal motions.

Our model of caldera collapse due to magma withdrawal has been proposed for volcanic systems in Iceland neighboring Bárðarbunga. North of the EVZ is the Northern Volcanic Zone (NVZ), consisting of five *en echelon* volcanic systems aligned with the boundary of the North American and European plates [24]. Within the NVZ, the Krafla caldera system was the site of a major rifting event between 1975-1984. Eruptive activity initiated with inflation of the central caldera, which then underwent rapid deflation, leading to lateral basaltic magma injection into a northward trending fissure swarm. A shallow magma chamber at approximately 3 km depth has been inferred at Krafla via inversions of geodetic observations [25]. Seismic tremor amplitude associated with dike emplacement was correlated with subsidence at the caldera [24], suggesting a mechanistic link between chamber pressure and stress on the dike. Numerical models support the hypothesis that magma

pressure, dike overpressure, and background tectonic stresses are the primary factors controlling dike propagation [26]. Recent studies of the current dike emplacement at Bárðarbunga also predict a feedback mechanism where dike-induced stresses trigger seismicity at the caldera itself [2]. Kinematically, however, the two volcanic systems are quite different since the maximum subsidence for Krafla (2 m) was substantially smaller than that of Bárðarbunga (about 60 m). Additionally, the Krafla episode experienced periodic inflation/deflation over 10 years whereas Bárðarbunga has thus far only experienced rapid deflation and over a much shorter time period.

On a global scale, there have only been a few observed eruptive events and subsequent caldera collapses in basaltic systems, although none with volume changes as large as Bárðarbunga. The 1968 caldera collapse at Isla Fernandina in the Galápagos Islands occurred on a single shield volcano, and aerial observations indicated that the caldera floor collapsed 300 m as a 3 km wide coherent block with motion confined to an elliptical boundary fault formed from a prior collapse event [27–29]. In 2000, the Miyakejima stratovolcano in the Izu-Bonin volcanic chain experienced 12 days of increased seismicity due to magma intrusion at its northwest flanks, which was then followed by a minor phreatic eruption and formation of a collapsed caldera 1.6 km in diameter [30, 31]. Tiltmeters stationed around the summit indicated intermittent abrupt uplift events superposed on the longer-term subsidence. The uplift events were accompanied by very long period (VLP) seismic events detectable over a wide area [30, 32]. Similar observations were collected for the 2007 Piton de la Fournaise caldera collapse during its largest historical eruption [33]. For Miyakejima, all of the VLP events exhibited large CLVD components for their focal mechanisms, which was explained to be slip on both inward- and outward-dipping faults [8, 31]. 60 km south of the Isla Fernandina caldera, an inflation event occurred at the Sierra Negra caldera which was followed a few years later by a  $M_w$  5.5 CLVD earthquake with a vertical tensional axis [8, 34], consistent with the interpretation of the older Bárðarbunga events being caused by inflation of a central magma chamber.

### 3.5 Conclusions

While the unique geodetic observations of ground deformation within and around the Bárðarbunga caldera during its collapse were the first of its kind for the caldera, there are still large uncertainties regarding the mechanics of the collapse process. The difficulties associated with the unknown interaction between the overlying glacier and the bedrock limits the spatial extent of usable data for estimating the geometry of the underlying magma chamber and active ring fault systems. We have shown that a

majority of the ice-free ground deformation can be attributed to steady deflation of a horizontal circular sill. However, the physical process driving the anomalous  $M_w > 5$  earthquakes along the caldera rim is still uncertain. The geodetic signature of one of these events, as measured by the differences in successive one-day interferograms over the ice, suggests a shallow seismic source caused by rupture on a ring fault or a rapidly closing crack. Since the amount of ground deformation expected for either seismic mechanism is significantly less than the observed deformation, we believe that the total ground deformation is caused by a combination of seismic processes and aseismic slip on ring faults. However, if the shear modulus in the vicinity of the caldera were much lower than expected (perhaps due to the presence of water-saturated basaltic tuff), then the predicted deformation for the seismic component would be much larger. In that case, the earthquakes can be explained entirely by rupture on dipping ring faults.

### 3.6 Supplementary: Materials and Methods

#### Elastic model for dike opening in the active rift zone

We estimate a model of tensile opening for a dike embedded in an elastic halfspace to account for ground displacements due to the active rift zone. The dike trace is defined by the earthquake locations in the SIL catalog (Fig. 3.1). We use six longer-term RS2 and CSK interferograms with time spans ranging from 24 to 48 days (Fig. 3.S4). In order to account for the different timespans of the interferograms, we divide the InSAR observations into four temporal subdomains spanning Aug. 16 to Sep. 20, 2014 (since we expect minimal ground deformation prior to Aug. 16) and solve for the distribution of tensile opening for each subdomain. The predicted LOS displacements in the interferograms are then linear combinations of the predicted displacements from one or more subdomains [35]. For interferograms with acquisition times falling within the span of a subdomain, we linearly extrapolate the predicted displacements associated with the nearest subdomain. Additionally, we use horizontal GPS displacements from Sigmundsson et al. (2014) as additional data for the first three temporal subdomains. We use the predicted ground displacements to form residual interferograms and GPS displacements for estimating the magma chamber geometry (Fig. 3.S8).

#### Source modeling for magma chamber beneath the Bárðarbunga caldera

To estimate the geometry of a magma chamber beneath the Bárðarbunga caldera while avoiding modeling errors due to uncertain ice-rock coupling, we consider LOS displacements from the ice-free areas adjacent to the caldera from longer-term RS2 interferograms. We use two RS2 interferograms (after removal of any signal due to the rift zone) with two different viewing angles (ascending and descending) to improve resolution of the chamber location. We also include horizontal displacements from GPS stations VONC and HNIF (also after removing the estimated signal from the rift zone). We model the chamber as a collapsing horizontal circular crack in an elastic half-space [14] while employing an adaptive Metropolis algorithm to sample the posterior probability density function for the chamber parameters [36]. We use the centroid of the subsidence pattern in the Aug. 27-28 and Sep. 12-13 CSK interferograms to define strong priors for the horizontal location of the chamber and sample for depth, radius, and difference between the magma pressure and lithostatic pressure. Since the amplitudes of the subsidence signals in the RS2 interferograms are substantially different, we assign independent pressure differences for each interferogram. We also sample for the coefficients of a bilinear ramp for

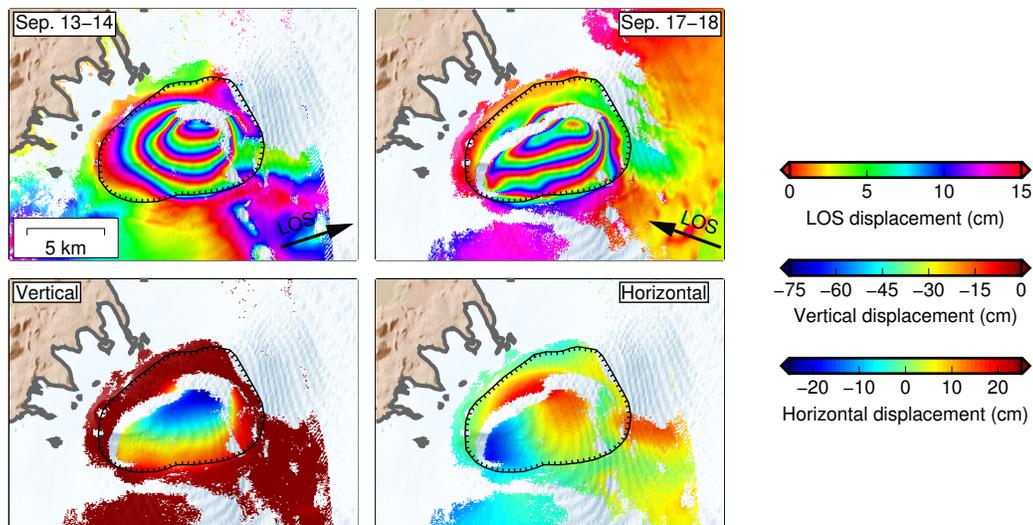


Figure 3.S1: Ascending and descending one-day CSK interferograms with time spans of Sep. 13-14 and Sep. 17-18, respectively. The two viewing geometries allow for separation of LOS displacements into vertical and horizontal components using the method of Yun et al. (2006), assuming that the two interferograms are measuring roughly the same ground deformation. While the two interferograms appear to measure similar levels of ice subsidence, the earthquake perturbations are more pronounced for the Sep. 13-14 interferogram (see Fig. 3.2). Nevertheless, the ratio of horizontal to vertical displacements does not exceed  $\approx 0.3$  within the caldera.

each interferogram to account for long-wavelength deformation. To maintain consistency between the ice-free interferograms and the CSK interferograms over the ice, we include one-day LOS displacements over the caldera from the Sep. 12-13 CSK interferogram to compare against the modeled vertical displacements over the caldera. While we assign those data low weights during the inversion due to the uncertain ice-rock coupling, their inclusion helps to constrain the pressure differences, which in turn provides better constraints for the chamber geometry. Finally, we acknowledge that there is likely some contribution to the observed ground displacement due to slip on caldera ring faults. However, the lack of reliable observations over the ice, as well as the rapid decay of deformation associated with fault slip with distance away from the caldera, does not allow for reliable estimations of fault geometry and/or slip. Therefore, we assume that most of the ground deformation on the ice-free areas is due to subsidence caused by the deflating magma chamber.

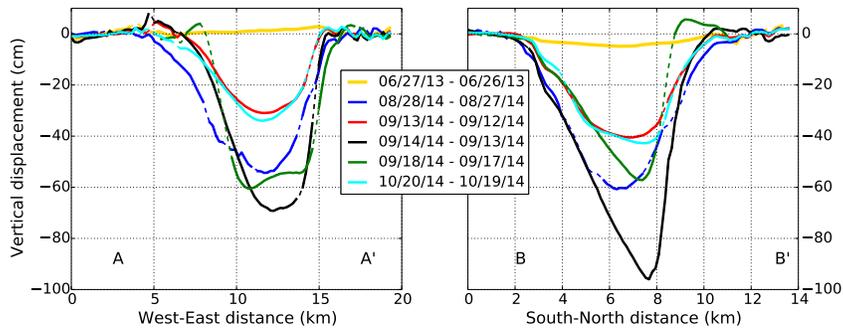


Figure 3.S2: Profiles of projected vertical subsidence observed in the five one-day CSK interferograms. Profile locations are shown in Fig. 3.2. Vertical displacements were computed by dividing the line-of-sight range changes by the cosine of the incidence angle for each interferogram. Solid lines are constructed from interferogram data with good coherence while the dashed lines interpolate through areas of low coherence. The gold line shows transects extracted from a 2013 one-day interferogram (Fig. 3.S3), confirming that the current subsidence is directly related to the ongoing magmatic activity.

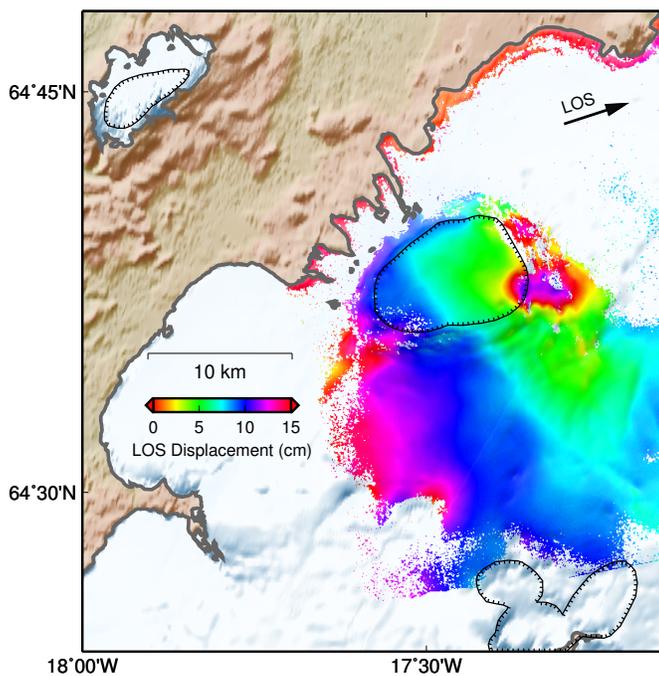


Figure 3.S3: One day interferogram of Bárðarbunga caldera and surrounding ice-covered regions formed from CSK images acquired on June 26, 2013 and June 27, 2013. No significant subsidence over the caldera is observed, although LOS displacements due to ice flow out of the east side of the caldera are consistent with the 2014 interferograms.

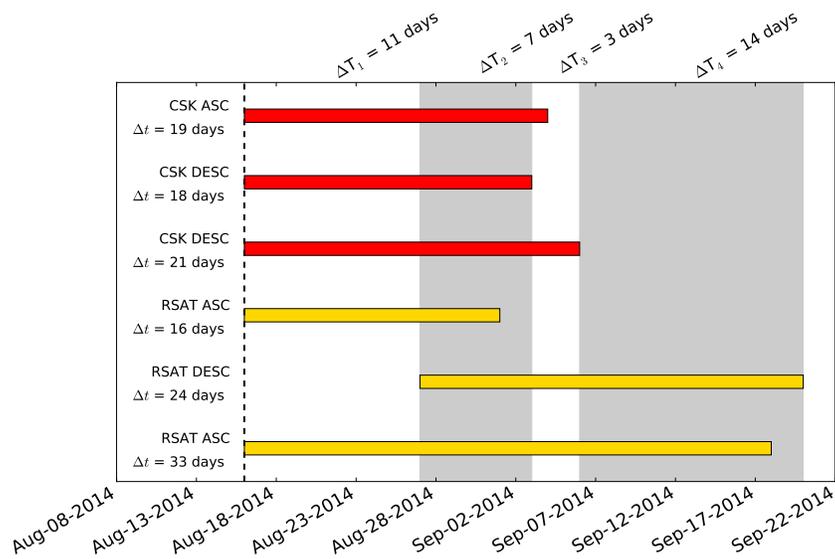


Figure 3.S4: Time spans of interferograms used for estimating the distribution of tensile opening along the dike interface. All interferogram time spans are referenced to Aug. 16, 2014, indicated by the vertical dashed line. The yellow lines correspond to RS2 interferograms while the red lines correspond to CSK interferograms. The orbital directions of the images used for the interferograms are indicated as either ascending (ASC) or descending (DESC). The alternating gray and white columns correspond to the temporal subdomains used in the inversion.

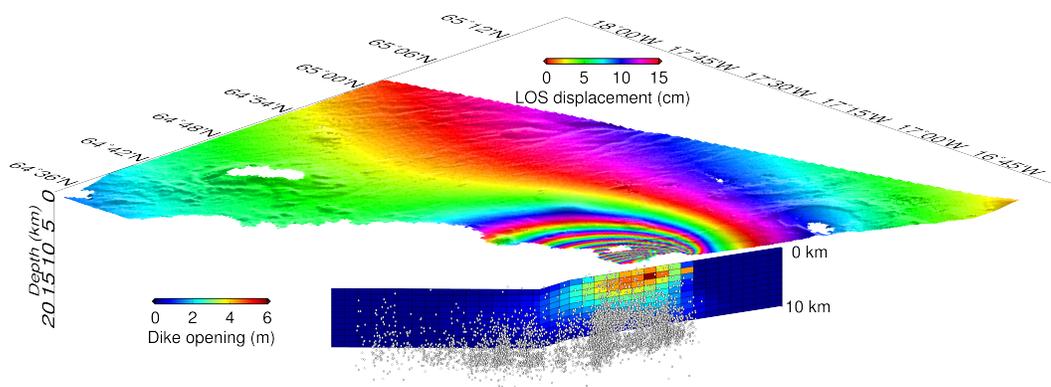


Figure 3.S5: Estimated distribution of opening along the dike interface for the time span corresponding to the RS2 interferogram in Fig. 3.1. The synthetic interferogram (see Fig. 3.S6) is shown above the fault, and the white dots correspond to earthquakes located along the dike using the SIL network.

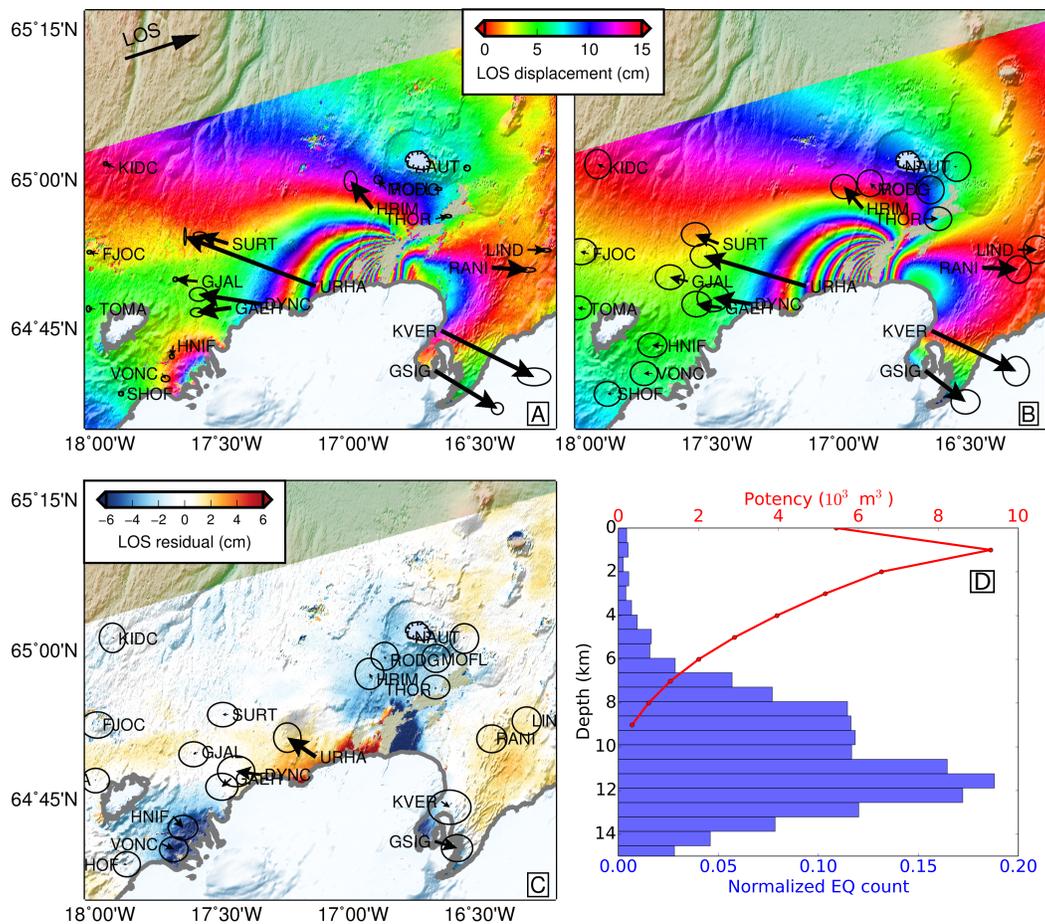


Figure 3.S6: Synthetic ground deformation for the elastic model of tensile opening along the dike interface. (A) RS2 interferogram from Fig. 3.1 and GPS displacements with 90% confidence error ellipses used in inversion. (B) Synthetic ground displacements from the estimated model. The modeled LOS displacements include a bilinear ramp to account for orbital errors and long-wavelength deformation. Ellipses represent 90% confidence posterior uncertainties in the modeled GPS displacements. (C) Residual interferogram and GPS displacements after subtracting the synthetic ground displacements. Larger residuals near the surface trace north of Vatnajökull may indicate potential asymmetries due to dike dip not captured in our vertical dike model. (D) Cumulative geodetic potency from the dike model in Fig. 3.S5 for Aug. 16 - Sep. 20 vs. a histogram of the depth distribution of earthquakes located along the dike.

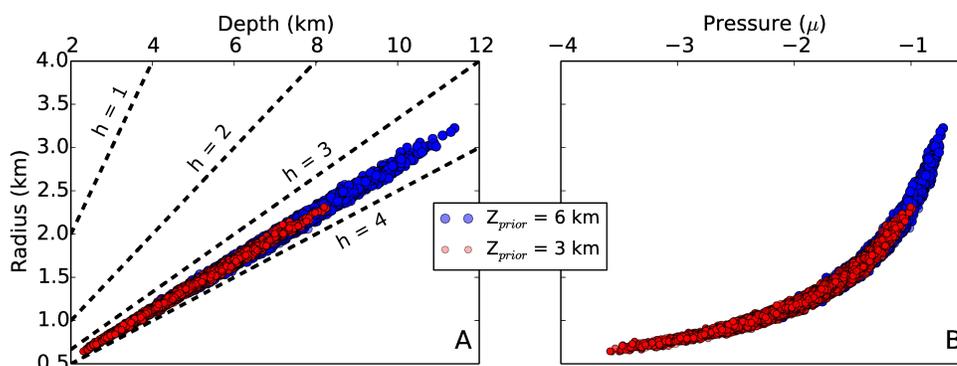


Figure 3.S7: (A) Samples for magma chamber depth and radius obtained from an adaptive Metropolis algorithm. Red and blue circles correspond to samples where the prior distribution for the chamber depth has a maximum at 3 and 6 km, respectively. The black dashed lines indicate lines of constant depth-to-radius ratio,  $h$ . (B) Samples for magma chamber pressure (in units of the shear modulus  $\mu$ ) and radius.

Table 3.S1: Comparison of select earthquakes from the GFZ Potsdam and Global CMT (GCMT) catalogs. Focal mechanisms and moment magnitudes are shown for the earthquakes from each catalog. We compute seismic moment as  $M_0 = 1/\sqrt{2} (\sum_{ij} M_{ij})^{1/2}$ . Both the mechanisms and moment magnitudes are consistent between the catalogs.

Date of event	GFZ		GCMT	
	Focal mechanism	$M_w$	Focal mechanism	$M_w$
08/26/2014 01:26	⊙	5.43	⊙	5.48
08/27/2014 02:50	⊙	5.37	⊙	5.45
08/28/2014 08:13	⊙	5.51	⊙	5.56
09/01/2014 11:41	⊙	5.53	⊙	5.60
09/03/2014 03:09	⊙	5.47	⊙	5.56
09/07/2014 07:07	⊙	5.57	⊙	5.65
09/15/2014 08:05	⊙	5.54	⊙	5.58
09/25/2014 05:00	⊙	5.18	⊙	5.25
09/29/2014 13:43	⊙	5.63	⊙	5.64
10/07/2014 10:22	⊙	5.57	⊙	5.63

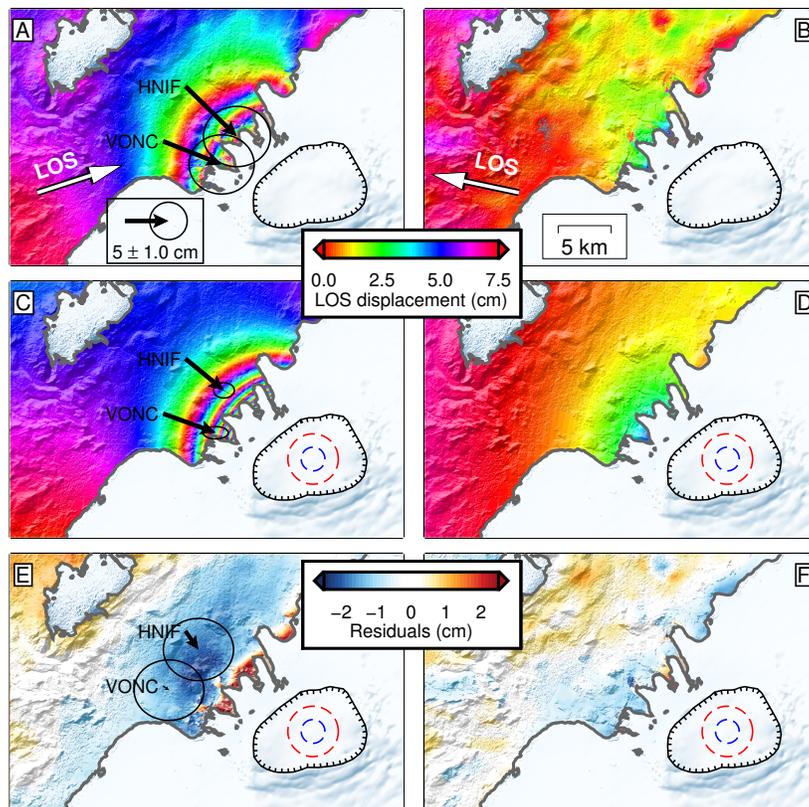


Figure 3.S8: Modeling of the source magma chamber using geodetic measurements on ice-free areas. The chamber is modeled as a horizontal circular sill at depth. (A) Unwrapped RS2 interferogram formed from images acquired in an ascending orbit on Aug. 1, 2014 and Sep. 18, 2014, and GPS displacements with 90% confidence error ellipses for VONC and HNIF after removing the estimated signal from the rift zone; (B) unwrapped RS2 interferogram from images acquired in a descending orbit on Aug. 27, 2014 and Sep. 20, 2014; (C) modeled GPS displacements and LOS displacements for the ascending interferogram for a chamber at a depth of 8 km and a radius of 2.3 km (location and size indicated by red dashed circle) or a depth of 4 km and radius of 1.1 km (blue dashed circle); (D) modeled LOS displacements for the descending interferogram; (E) residual between (A) and (C); and (F) residual between (B) and (D). The estimated volume changes were  $0.43 \text{ km}^3$  and  $0.14 \text{ km}^3$  for the ascending and descending interferograms, respectively.

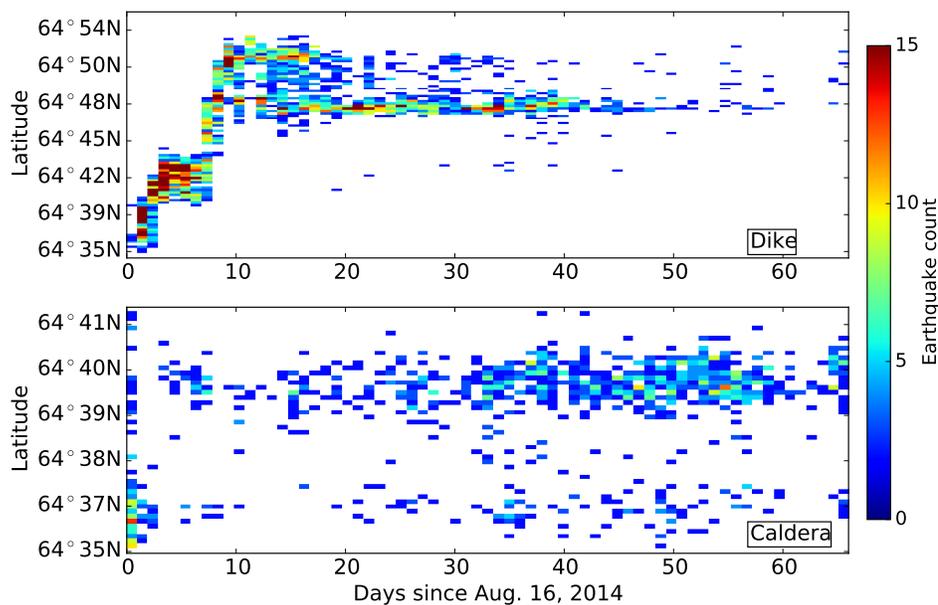


Figure 3.S9: Seismicity rates for earthquakes occurring within the dike (top) and caldera (bottom). The number of earthquakes are binned in latitude and time where 1-day intervals are used for the time bins. The count is saturated at 15 earthquakes/day for visualization clarity. Bins with less than two earthquakes are colored white. High seismicity rates within the dike for the first 10 days are associated with the initial dike emplacement. While seismicity along the dike has decreased with time, seismicity along the northern rim of the caldera has increased.

Table 3.S2: Focal mechanisms and moment magnitudes for CLVD events occurring near Bárðarbunga from 1976-1996. The focal mechanisms were formed using the moment tensor elements estimated by Nettles & Ekström (1998). These prior events exhibit opposite polarities from the current events (vertical tensional vs. compressional axes).

Date of event	Focal mechanism	$M_w$
07/27/1976		5.4
12/28/1977		5.4
06/22/1979		5.2
08/12/1980		5.4
02/03/1989		5.1
09/15/1990		5.6
09/26/1992		5.6
06/22/1993		5.4
05/05/1994		5.4
09/29/1996		5.6

## References

1. National Earthquake Information Center 2014.
2. Gudmundsson, A., Lecoeur, N., Mohajeri, N. & Thordarson, T. Dike emplacement at Bardarbunga, Iceland, induces unusual stress changes, caldera deformation, and earthquakes. *Bulletin of Volcanology* **76**, 1–7 (2014).
3. Sigmundsson, F. *et al.* Segmented lateral dyke growth in a rifting event at Bardarbunga volcanic system, Iceland. *Nature* (2014).
4. Aki, K. & Richards, P. G. *Quantitative seismology* (2002).
5. Knopoff, L. & Randall, M. J. The compensated linear-vector dipole: A possible mechanism for deep earthquakes. *Journal of Geophysical Research* **75**, 4957–4963 (1970).
6. Green, R. G., White, R. S. & Greenfield, T. Motion in the north Iceland volcanic rift zone accommodated by bookshelf faulting. *Nature Geoscience* **7**, 29–33 (2014).
7. Centre, G. D. *GEOFON Seismic Network. Deutsches GeoForschungsZentrum GFZ. Other/Seismic Network* Last accessed: 2014-10-13. 1993.
8. Shuler, A., Nettles, M. & Ekström, G. Global observation of vertical-CLVD earthquakes at active volcanoes. *Journal of Geophysical Research: Solid Earth* **118**, 138–164 (2013).
9. Shuler, A., Ekström, G. & Nettles, M. Physical mechanisms for vertical-CLVD earthquakes at active volcanoes. *Journal of Geophysical Research: Solid Earth* **118**, 1569–1586 (2013).
10. Roche, O., Druitt, T. & Merle, O. Experimental study of caldera formation. *Journal of Geophysical Research: Solid Earth (1978–2012)* **105**, 395–416 (2000).
11. Geyer, A., Folch, A. & Marti, J. Relationship between caldera collapse and magma chamber withdrawal: an experimental approach. *Journal of Volcanology and Geothermal Research* **157**, 375–386 (2006).
12. Gudmundsson, A. Conceptual and Numerical Models of Ring-Fault Formation. *Journal of Volcanology and Geothermal Research* **164**, 142–160 (2007).
13. Gudmundsson, A. Magma-chamber geometry, fluid transport, local stresses and rock behaviour during collapse caldera formation. *Developments in Volcanology* **10**, 313–349 (2008).
14. Fialko, Y., Khazan, Y. & Simons, M. Deformation due to a pressurized horizontal circular crack in an elastic half-space, with applications to volcano geodesy. *Geophysical Journal International* **146**, 181–190 (2001).

15. Acocella, V. Understanding caldera structure and development: An overview of analogue models compared to natural calderas. *Earth-Science Reviews* **85**, 125–160 (2007).
16. Smith, R. P., Jackson, S. M. & Hackett, W. R. Paleoseismology and seismic hazards evaluations in extensional volcanic terrains. *Journal of Geophysical Research: Solid Earth (1978–2012)* **101**, 6277–6292 (1996).
17. Dreger, D. S., Ford, S. R. & Walter, W. R. Source analysis of the Crandall Canyon, Utah, mine collapse. *Science* **321**, 217–217 (2008).
18. Nettles, M. & Ekström, G. Faulting mechanism of anomalous earthquakes near Bardarbunga Volcano, Iceland. *Journal of Geophysical Research: Solid Earth (1978–2012)* **103**, 17973–17983 (1998).
19. Konstantinou, K. I., Kao, H., Lin, C.-H. & Liang, W.-T. Analysis of broadband regional waveforms of the 1996 September 29 earthquake at Bárðarbunga volcano, central Iceland: investigation of the magma injection hypothesis. *Geophysical Journal International* **154**, 134–145 (2003).
20. Tkalčić, H., Dreger, D. S., Foulger, G. R. & Julian, B. R. The puzzle of the 1996 Bárðarbunga, Iceland, earthquake: No volumetric component in the source mechanism. *Bulletin of the Seismological Society of America* **99**, 3077–3085 (2009).
21. Ekström, G. Anomalous earthquakes on volcano ring-fault structures. *Earth and Planetary Science Letters* **128**, 707–712 (1994).
22. Sipkin, S. A. Interpretation of non-double-couple earthquake mechanisms derived from moment tensor inversion. *Journal of Geophysical Research: Solid Earth (1978–2012)* **91**, 531–547 (1986).
23. Kanamori, H., Ekström, G., Dziewonski, A., Barker, J. S. & Sipkin, S. A. Seismic radiation by magma injection: An anomalous seismic event near Tori Shima, Japan. *Journal of Geophysical Research: Solid Earth (1978–2012)* **98**, 6511–6522 (1993).
24. Einarsson, P. Earthquakes and present-day tectonism in Iceland. *Tectonophysics* **189**, 261–279 (1991).
25. Ewart, J. A., Voight, B. & Björnsson, A. Elastic deformation models of Krafla Volcano, Iceland, for the decade 1975 through 1985. *Bulletin of volcanology* **53**, 436–459 (1991).
26. Buck, W. R., Einarsson, P. & Brandsdóttir, B. Tectonic stress and magma chamber size as controls on dike propagation: Constraints from the 1975–1984 Krafla rifting episode. *Journal of Geophysical Research: Solid Earth (1978–2012)* **111** (2006).
27. Simkin, T. & Howard, K. A. Caldera Collapse in the Galápagos Islands, 1968 The largest known collapse since 1912 followed a flank eruption and explosive volcanism within the caldera. *Science* **169**, 429–437 (1970).

28. Filson, J., Simkin, T. & Leu, L.-k. Seismicity of a caldera collapse: Galapagos Islands 1968. *Journal of Geophysical Research* **78**, 8591–8622 (1973).
29. Francis, T. A new interpretation of the 1968 Fernandina Caldera collapse and its implications for the mid-oceanic ridges. *Geophysical Journal International* **39**, 301–318 (1974).
30. Ukawa, M., Fujita, E., Yamamoto, E., Okada, Y. & Kikuchi, M. The 2000 Miyakejima eruption: Crustal deformation and earthquakes observed by the NIED Miyakejima observation network. *Earth Planets and Space* **52**, xix–xxvi (2000).
31. Geshi, N., Shimano, T., Chiba, T. & Nakada, S. Caldera collapse during the 2000 eruption of Miyakejima Volcano, Japan. *Bulletin of Volcanology* **64**, 55–68 (2002).
32. Kumagai, H. *et al.* Very-long-period seismic signals and caldera formation at Miyake Island, Japan. *Science* **293**, 687–690 (2001).
33. Michon, L., Villeneuve, N., Catry, T. & Merle, O. How summit calderas collapse on basaltic volcanoes: New insights from the April 2007 caldera collapse of Piton de la Fournaise volcano. *Journal of Volcanology and Geothermal Research* **184**, 138–151 (2009).
34. Yun, S., Segall, P. & Zebker, H. Constraints on magma chamber geometry at Sierra Negra Volcano, Galápagos Islands, based on InSAR observations. *Journal of Volcanology and Geothermal Research* **150**, 232–243 (2006).
35. Pritchard, M. & Simons, M. An aseismic slip pulse in northern Chile and along-strike variations in seismogenic behavior. *Journal of Geophysical Research: Solid Earth (1978–2012)* **111** (2006).
36. Haario, H., Saksman, E. & Tamminen, J. An adaptive Metropolis algorithm. *Bernoulli*, 223–242 (2001).

*Chapter 4***QUANTIFYING GROUND DEFORMATION IN THE LOS ANGELES AND SANTA ANA COASTAL BASINS DUE TO GROUNDWATER PUMPING****4.1 Introduction**

In regions over and adjacent to active aquifer systems, ground deformation can occur as a result of pumping of groundwater, long-term drought effects, heavy rainfall, and artificial recharge of the aquifers [1, 2]. Ground deformation is a response to changes in pore pressure in the aquifers, which changes the effective stress on the aquifer system's granular matrix and causes contraction or expansion of the aquifers. From a natural hazard perspective, land subsidence following groundwater withdrawal and compaction of aquifer systems is of particular interest when the subsidence is long-term, leading to increased strain on infrastructure and potential formation of earth fissures and surface faults [2]. Long-term subsidence can occur as a result of a slow decline in groundwater levels or permanent compaction due to effective stress levels exceeding preconsolidation stress levels [3]. More generally, ground deformation due to extraction of fluids in subsurface reservoirs can be described by short-term, elastic responses and long-term, inelastic compression or poroelastic rebound processes [1, 2]. Many municipal water districts closely monitor hydraulic head data to track pore pressures in order to maintain sustainable pumping practices and prevent effective stress levels within aquifers from exceeding their preconsolidation levels.

Many studies have shown measurements of ground deformation to be complementary to hydraulic head data for groundwater monitoring. Continuous monitoring of ground elevations at discrete points can be achieved through leveling and Global Positioning System (GPS) data. Additionally, GPS data provides measurements of horizontal motions, which can be useful for quantifying and modeling aquifer properties [2, 4]. Recently, interferometric synthetic aperture radar (InSAR) has proven to be a very useful remote sensing technique for acquiring spatially dense ground deformation measurements for deformation driven by hydrologic and geothermal fluid processes. Studies have used a small collection of interferograms to quantify seasonal deformation caused by the annual cycle of groundwater pumping and

recharge, as well as longer term subsidence from accumulated overdraft of aquifers [4, 5]. Another class of InSAR techniques uses “stacks” of many co-registered interferograms collected over a finite time period to construct a full time series of ground deformation. These techniques have been used to measure the evolution of land subsidence in the Santa Clara Valley in California [6], subsidence and uplift in Phoenix, Arizona [7], seasonal uplift and subsidence in the Los Angeles area [8, 9], etc.

Ground deformation within and around the Los Angeles area has been measured for several decades using GPS data from the Southern California Integrated GPS Network (SCIGN) and InSAR data. Historically, the primary driver for acquiring geodetic data was to quantify the rate of tectonic contraction across the region and the rate of elastic loading on potentially seismogenic faults such as those involved in the 1987 Whittier Narrows and 1994 Northridge earthquakes [8]. However, many of the geodetic signals used to study these fault systems are contaminated or completely obscured by non-tectonic processes such as groundwater pumping and oil extraction [4]. Several geodetic studies have thus aimed to quantify the total contribution of non-tectonic sources of deformation for the Los Angeles area.

Bawden *et al.* [4] used a series of interferograms from 1997 - 1999 to observe several anthropogenic deformation processes in the basins surrounding the Los Angeles area, including seasonal uplift and subsidence due to groundwater pumping in the Santa Ana Coastal Basin which is the primary source of groundwater for Orange County. The larger magnitude of the subsidence signal as compared to the uplift (60 mm for the former, 50 mm for the latter) implied a net subsidence signal in the basin thought to be due to inelastic compaction of lower permeability aquitards within the aquifer system. Watson *et al.* [8], Lanari *et al.* [9], and Zhang *et al.* [10] extended on the analysis by Bawden *et al.* [4] by including more interferograms over a longer time span. In particular, Lanari *et al.* [9] applied the small baseline subset (SBAS) algorithm [11] to produce a time series model from 1995 - 2002. Cross-correlation of the spatially varying time series with a reference sinusoid was performed to compute a time shift for each ground point. This time shift map revealed sharp boundaries for the region of the basin responding to the annual groundwater pumping and recharging, and heterogeneous time shifts within the basin also suggested lateral variability in hydraulic conductivity. The spatially dense measurements of ground deformation provided by InSAR time series can therefore elucidate subtle characteristics of time-dependent ground deformation

within the basin due to changes in aquifer pressure caused by anthropogenic and natural processes. In these studies, seasonal oscillations were generally quantified in an average sense under the assumption that the oscillations were purely sinusoidal with a period of 1 year.

In this study, we first explore the geologic and hydrologic setting of the coastal basins in the Los Angeles area. Following a brief discussion on how groundwater level changes drive ground deformation in Section 4.2, we examine the time history of groundwater levels for the main aquifer systems in Section 4.3 using hydraulic head data from the Water Replenishment District (WRD) in Los Angeles and the Orange County Water District (OCWD). We show head time series that exhibit highly complex time histories with time-varying seasonal variations and various longer-term trends related to background groundwater levels. We then perform an initial comparison between groundwater levels and surface deformation using GPS data. We demonstrate how deformation signals with short-term variations are driven by short-term variations in the principal aquifer while long-term deformation signals appear to be driven by long-term variations in the deep aquifer.

In Section 4.4, we perform an InSAR time series analysis on an expanded data set that includes interferograms from 1992 to 2011 to investigate the spatial variations in short-term, seasonal ground deformation. By using the assumption that seasonal ground motions can be represented by a linear combination of sinusoids, we can compare the timing and amplitude of peak seasonal deformation of ground points within the basin to study the effects of aquifer geometry and groundwater pumping practices on the ground response. This method of time series analysis is similar to the ones used in Bawden *et al.* [4] and Watson *et al.* [8]. However, since both hydraulic head and ground deformation time series show non-sinusoidal seasonal variations, in Section 4.5 we develop a new method for InSAR time series analysis that automatically decomposes the time series into generic long- and short-term signals. We compare the decomposed InSAR time series to all available monitoring wells to better understand the relationship between deformation signals of different durations and different segments of the groundwater system. Finally, in Section 4.6, we discuss how these results are related to groundwater management with regards to sustainable pumping practices, and we relate several of our InSAR observations to physical processes based on groundwater dynamics.

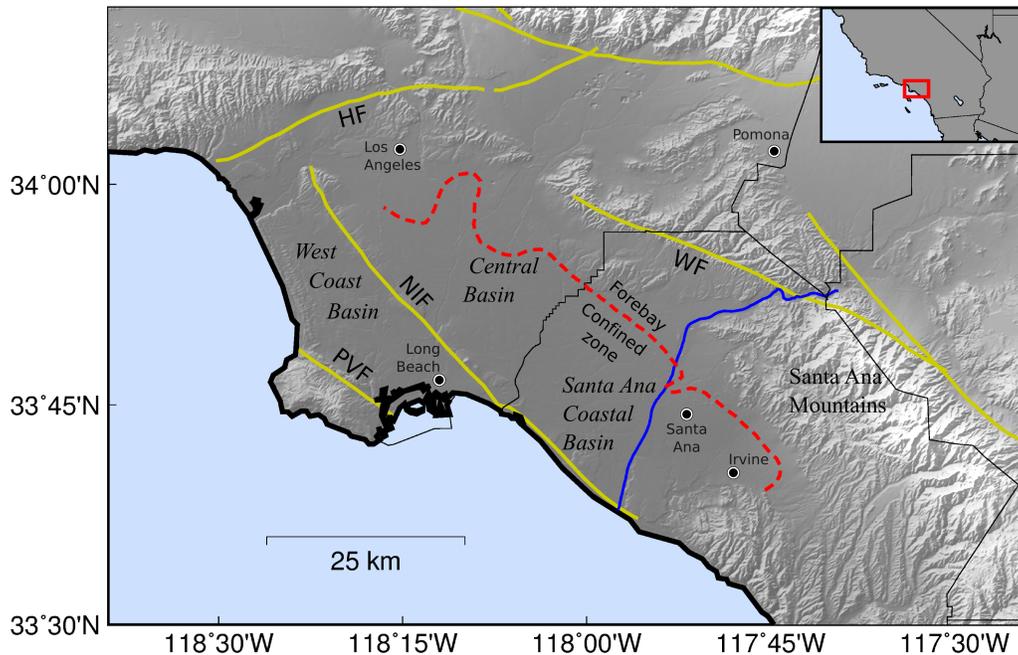


Figure 4.1: Location and tectonic setting of the Los Angeles Central and West Coast Basins and the Santa Ana Coastal Basin. The thick gold lines represent major faults in the area, including the Newport-Inglewood Fault (NIF), the Whittier Fault (WF), the Palos Verdes Fault (PVF), and the Hollywood Fault (HF). The thin black lines indicate county boundaries, and the thick blue line corresponds to the Santa Ana River. The dashed red line indicates the barrier between the forebay and confined areas of the groundwater system modified from estimates by the WRD and OCWD. The inset shows the location of the study area along the California coast.

## 4.2 Background

### Hydrology and Structure of the Los Angeles Basins

The Los Angeles area consists of several basins containing groundwater systems, including the Los Angeles Central and West Coast Basins and the Santa Ana Coastal Basin in Orange County (Figure 4.1). The Los Angeles Central Basin (hereafter referred to as the Central Basin) is separated from the West Coast Basin by the northern portion of the Newport-Inglewood Fault Zone (NIF), which acts locally as a barrier for fluid flow between the two basins. In this study, we focus on ground deformation within the Central and Santa Ana Coastal Basins since changes in groundwater levels are greater for those two basins. Both of these basins have groundwater systems with similar structural characteristics. These systems are lens-shaped (i.e., thicker in the central areas and thinner in the margins) and consist of freshwater-bearing

deposits with varying degrees of permeabilities [12]. The relative abundance of clay is the primary driver for determining the flow of groundwater through these systems. Clays generally have permeabilities that are much lower than those of sands and can restrict the flow of water in a given direction. The basins can thus be divided into forebay and confined areas. The forebay area occupies about 38 percent of the Santa Ana Coastal Basin and occupies the eastern region that meets the Santa Ana Mountains (Figure 4.1). Here, the shallower aquifers are known to be unconfined, i.e. vertical movement of water is not restricted by laterally extensive clay layers [13]. Groundwater recharge occurs mainly in the forebay area. The confined area (also known as pressure area) is considerably larger and extends from the western edge of the forebay area to the Pacific Ocean [12]. Here, there are laterally continuous thick layers of silt and clay that restrict vertical flow of groundwater, causing aquifers to be under confining pressure. While the actual groundwater system consists of several contiguous aquifer units, clay interbeds, and confining aquitards, the OCWD has developed a simplified model consisting of shallow, middle (principal), and deep aquifer layers (Figure 4.B1) [13]. The shallow aquifer generally spans depths of up to 200 feet for most of the basin, and is only unconfined in the forebay region. The principal aquifer, which supports over 90 percent of the basin pumping in Orange Country, is generally greater than 1000 feet thick for much of the basin and is semi-confined in the forebay. The semi-confined state allows recharge water to migrate downward into the principal aquifer. The deep aquifer, which is limited in production capability due to its depth and the presence of amber colored groundwater (which requires extra treatment to remove colors and odors from the water), defines water-storing units up to 2000 feet in depth in the center of the Santa Ana Coastal Basin. Since most of the groundwater production draws water from the principal aquifer, the observed ground deformation is presumed to be caused by pressure variations in the principal aquifer. Next, we will explore the physical relationships governing how pressure variations in aquifers can lead to ground deformation.

### **Groundwater Flow Theory and Aquifer Compaction**

The relationship between ground deformation and aquifer pressure for *confined* aquifers can be explained using groundwater flow theory based on the Principle of Effective Stress [14]. Effective stress can be expressed as

$$\sigma'_{ij} = \sigma_{ij} - \delta_{ij}p, \quad (4.1)$$

where  $\sigma'_{ij}$  and  $\sigma_{ij}$  are components of the effective and total stress tensors, respectively,  $p$  is the pore-fluid pressure, and  $\delta_{ij}$  is the Kronecker delta. Since groundwater

is a Newtonian fluid, the off-diagonal terms of the stress tensors corresponding to shear stress are equal for  $\sigma'_{ij}$  and  $\sigma_{ij}$  [2]. For aquifer layers that are nearly horizontal and are laterally extensive with respect to their thicknesses, changes in pore-fluid pressure gradients will be primarily in the vertical ( $\hat{z}$ ) direction. Therefore, assuming resulting strains are primarily in the  $zz$  component, we can simplify Equation 4.1 to

$$\sigma'_{zz} = \sigma_{zz} - p. \quad (4.2)$$

By assuming that changes in the total/overburden stress are negligible for our study period, changes in effective stress can then be simply expressed as

$$\Delta\sigma'_{zz} = -\Delta p. \quad (4.3)$$

We can relate pore-fluid pressure to hydraulic head, or water level, by knowing that hydraulic head is the sum of the pressure head and elevation head [15]:

$$h = \frac{p}{\rho_w g} + h_z, \quad (4.4)$$

where  $h$  is the total hydraulic head,  $\rho_w$  is the density of water,  $g$  is the gravitational acceleration constant, and  $h_z$  is the elevation head referenced to a given datum. We can then write Equation 4.3 as

$$\Delta\sigma'_{zz} = -\rho_w g \Delta h. \quad (4.5)$$

In order to relate pore-fluid pressure changes to ground deformation, we use the definition for one-dimensional skeletal compressibility,  $\alpha$ , as the ratio of vertical strain to vertical effective stress:

$$\alpha = \frac{-\Delta b/b}{\Delta\sigma'_{zz}}, \quad (4.6)$$

where the  $\Delta b$  is the change in thickness of a control volume with initial thickness  $b$  [2]. This equation thus relates the compaction and expansion of sediments to changes in effective stress; integrated compaction over the entire depth of the aquifer system is equivalent to what we measure as land subsidence [16]. The skeletal specific storage,  $S_s = \rho_w g \alpha$ , can be used to combine Equations 4.5 and 4.6 as

$$S_s b = S_k = \frac{\Delta b}{\Delta h}, \quad (4.7)$$

where  $S_k$  is the skeletal storage coefficient. Consolidation experiments on typical sediments have shown that specific storage for a given material can behave very

differently depending on whether the effective stress is above or below the pre-consolidation stress (i.e., the previous maximum effective stress). This boundary essentially separates the regimes of elastic and inelastic deformation depending on the head level. To account for these two regimes, two separate skeletal specific storages are used:

$$S_k = \begin{cases} S_{ke} & \text{for } \sigma'_{zz} < \sigma'_{zz(pre)} \\ S_{kv} & \text{for } \sigma'_{zz} \geq \sigma'_{zz(pre)}, \end{cases} \quad (4.8)$$

where  $S_{ke}$  and  $S_{kv}$  are the elastic and inelastic, or virgin, skeletal specific storage coefficients. In the inelastic regime when head drops below the preconsolidation level, irreversible loss of water storage occurs, and in the case where this state persists, the skeletal specific storage is expected to vary proportionally to the logarithm of the effective stress [2]. From a water management perspective, the preconsolidation level is used to define sustainable pumping rates in order for aquifer deformation to remain elastic.

### 4.3 Groundwater Level and Ground Deformation Time History

#### Monitoring Groundwater Levels with Hydraulic Head Time Series

Since the primary driver of ground deformation within the coastal basins is changes in aquifer pressure (Section 4.2), we gather data from several multi-port monitoring wells located within the coastal basins that measure hydraulic head levels at various aquifer depths. Here, we use twelve monitoring wells operated by WRD and thirty wells operated by OCWD (Figure 4.2). For both sets of wells, the depths of the sample ports allow us to determine hydraulic head levels for distinct aquifer units. For confined aquifers, the hydraulic head level corresponds to the piezometric (or potentiometric) surface which coincides with the hydrostatic pressure level of the water in the aquifer [1]. This surface is above the upper surface of the given aquifer layer and varies as water is added or removed to the system. For unconfined aquifers, the hydraulic head corresponds directly with the elevation of the water table. Groundwater levels as measured by hydraulic head time series for OCWD well station SAR-9 (located in the middle of the Santa Ana Coastal Basin over the confined aquifers) show large annual oscillations, particularly at the depth of the principal aquifer (Figure 4.3). These oscillations are caused by the annual cycle of groundwater recharge (artificial and natural) and withdrawal. From the early 1990s to the mid-2000s, groundwater levels typically peaked around March after the rainy season and are at their lowest towards the end of the summer during the heaviest periods of groundwater pumping. This time period also corresponds to

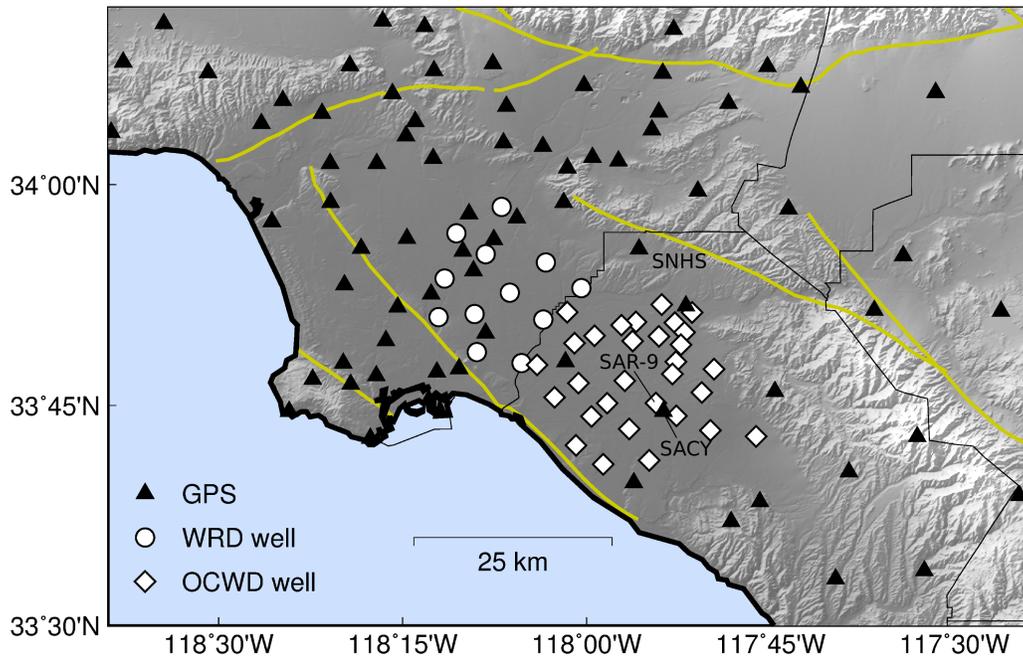


Figure 4.2: Distribution of GPS and well data used in this study. The white circles correspond to WRD wells used in this study while the white diamonds correspond to OCWD wells. The black triangles show the GPS coverage in this area.

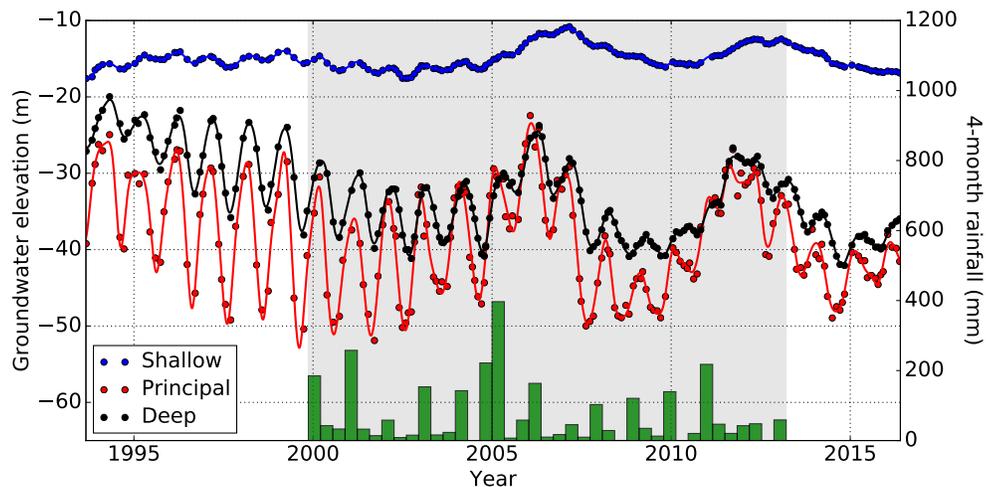


Figure 4.3: Hydraulic head time series for OCWD well SAR-9 for selected ports in the shallow, principal, and deep aquifers. The dots represent the raw head data while the solid lines are interpolated data using smoothing splines. The green bars represent the average rainfall in the Orange County area in 3-month intervals from 2000 to 2013. The light gray area indicates the time span for available rainfall data.

the largest annual fluctuations in groundwater levels due to water storage programs that encouraged increased groundwater pumping during the summer months when demand for imported water is higher [13]. In Orange County, this program, referred to as Short-Term Seasonal Shift (STSS), stopped after 2008, which is reflected in the lower amplitude oscillations in the hydraulic head time series. Superimposed over the annual water level fluctuations are longer-term variations. Generally, we can observe a decrease in overall water levels, which started in 1970 after the basin was essentially refilled after replenishment from Colorado River water [13]. We can also observe two transient increases in groundwater levels in 2005 and in 2012, both of which followed periods of heavy rainfall.

### Groundwater Diffusion and Varying Response Times

Since groundwater is pumped and recharged at discrete points within the Central and Santa Ana Coastal Basins, we expect spatial variations in hydraulic head due to groundwater flow in response to pressure gradients. Differences in permeabilities between sandy aquifer units and clay interbeds and aquitards can cause time delays in hydraulic head levels from one aquifer unit to another, as well as intra-aquifer delays. Additionally, changes in aquifer thickness can affect groundwater flow.

Three-dimensional flow of ground water in porous media can be described by the differential equation for a control volume [17, 18]:

$$\frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y} + \frac{\partial v_z}{\partial z} = -S_k \frac{\partial h}{\partial t} + Q(t, x, y, z), \quad (4.9)$$

where  $v_x$ ,  $v_y$ , and  $v_z$  are the rectangular components of the instantaneous bulk fluid velocity in the control volume, and  $Q(t, x, y, z)$  is a source term that can vary in time and in space. We can relate the fluid velocity to hydraulic head using Darcy's Law:

$$\begin{aligned} V_x &= -K \frac{\partial h}{\partial x} \\ V_y &= -K \frac{\partial h}{\partial y}, \end{aligned} \quad (4.10)$$

where  $V_x$  and  $V_y$  represent instantaneous averages of  $v_x$  and  $v_y$ , respectively, over the thickness of the aquifer, and the hydraulic conductivity,  $K$ , is assumed to vary in the  $x$  and  $y$  directions such that  $K = K(x, y)$ . We can integrate Equation 4.9 in the vertical  $z$  dimension for an aquifer of thickness  $b = b(x, y)$ , and for the purpose of this discussion, we will assume one-dimensional flow such that  $K = K(x)$ ,  $b = b(x)$ , and flow in the  $y$ -direction is constant [19]. We then obtain the following general

diffusion equation in one dimension for variable aquifer thickness and hydraulic conductivity:

$$K \frac{\partial b}{\partial x} \frac{\partial h}{\partial x} + b \frac{\partial K}{\partial x} \frac{\partial h}{\partial x} + bK \frac{\partial^2 h}{\partial x^2} = S_s b \frac{\partial h}{\partial t} - Q(t, x). \quad (4.11)$$

In the special case of flow through an aquifer with constant thickness and conductivity, Equation 4.11 can be written as [20]:

$$K \frac{\partial^2 h}{\partial x^2} = S_s \frac{\partial h}{\partial t} - Q(t, x). \quad (4.12)$$

For the following discussion, let us assume a point source,  $Q(t, x) = Q(t)\delta(x - x_0)$ , and assume  $Q(t)$  is periodic to represent groundwater recharge and pumping. Therefore, for a time varying source with a temporal frequency  $\omega$ ,  $Q(t) = \cos \omega t$ , the solution to Equation 4.12 will have the approximate form ([21])

$$h(x, t) \approx e^{-\sqrt{2\omega}|x-x_0|} \cos(\sqrt{2\omega}|x - x_0| \pm \omega t). \quad (4.13)$$

In the above equation, there is a frequency dependent attenuation term that would damp out the diffusion of the hydraulic head at higher source frequencies. In other words, head variations due to a point source perturbation in water pressure (such as a production well) would decay more rapidly away from the source for a faster withdrawal/pumping cycle. The  $\omega t$  factor in the periodic term in the above equation controls the diffusion speed, or hydraulic response time, of the pressure perturbation. For larger  $\omega$ , we would expect a shorter response time. By solving for the homogeneous solution to Equation 4.12, we can also estimate the material-dependent response time for confined aquifers as [22]:

$$T^* = S_s L_c^2 / K, \quad (4.14)$$

where  $T^*$  is the response time and  $L_c$  is a characteristic length for a specific aquifer unit. Thus, aquifers with higher hydraulic conductivity and/or smaller thicknesses would experience lower response times.

To assess timing differences in the vertical dimension, we use the SAR-9 hydraulic head data to estimate the amplitude and time to peak signal for the annual oscillations in head. Since the head data contain both seasonal variations and long-term trends, we model the time series as a linear combination of sinusoids with annual and semi-annual periods for the seasonal signal and third-order integrated B-splines (hereafter

referred to as  $B^i$ -splines; see Hetland *et al.* [23]) for the transient, long-term trends:

$$h(t) = \sum_{i=1}^2 \left[ a_i \cos \frac{2\pi}{T_i} t + b_i \sin \frac{2\pi}{T_i} t \right] + \sum_{j=1}^{32} c_j B^i(t - t_j), \quad (4.15)$$

where  $T_1 = 0.5$  years and  $T_2 = 1$  year, and  $B^i(t - t_j)$  represent the  $B^i$ -splines centered at time  $t_j$  [23]. Here, we partition the data time span into 32 evenly spaced knot times  $t_j$  so that the  $B^i$ -splines each have an effective duration of  $2 * (t_j - t_{j-1})$ . We estimate the coefficients  $a_i$ ,  $b_i$ , and  $c_j$  simultaneously using regularized least squares for the cost function

$$J(\mathbf{m}) = \underset{\mathbf{m}}{\operatorname{argmin}} \|\mathbf{h} - \mathbf{G}\mathbf{m}\|_2^2 + \mathbf{m}^T \mathbf{C}_m^{-1} \mathbf{m}, \quad (4.16)$$

where  $\mathbf{h}$  is the time series data,  $\mathbf{G}$  is the temporal design matrix containing the sinusoids and  $B^i$ -splines along the columns,  $\mathbf{m}$  is the vector of coefficients of the elements in  $\mathbf{G}$ , and  $\mathbf{C}_m$  is a prior covariance matrix for regularization. For this analysis, we set  $\mathbf{C}_m$  to be the identity matrix with an appropriate scaling coefficient. After estimating  $\mathbf{m}$  for each well, we can compute the amplitude and phase delay for each seasonal component as

$$A_i = \sqrt{a_i^2 + b_i^2} \quad (4.17)$$

$$\phi_i = \tan^{-1} \left( \frac{b_i}{a_i} \right), \quad (4.18)$$

where  $A_i$  is the amplitude and  $\phi_i$  is the phase delay, or time to peak signal. We repeat this procedure for the time series at each port depth to estimate the time to reach peak head levels for the seasonal signal between 1996 and 2006 (when the seasonal oscillations were most regular during the STSS program) and for the transient increase starting in 2004-2005 caused by heavy rainfall (Figure 4.4).

For both the annual and transient peak signals, we observe that water levels in the principal aquifer reach peak levels earlier than both the shallow and deep aquifers. The peak-to-peak seasonal increases in water levels are much higher in the principal aquifer, which is expected since 90% of groundwater pumping is from the principal aquifer. The aquitards cause noticeable delays in the timing of the peak signals (due to low permeabilities for clay) while also reducing the peak-to-peak amplitudes. The depth of the earliest time of peak signal generally varies from well to well, although

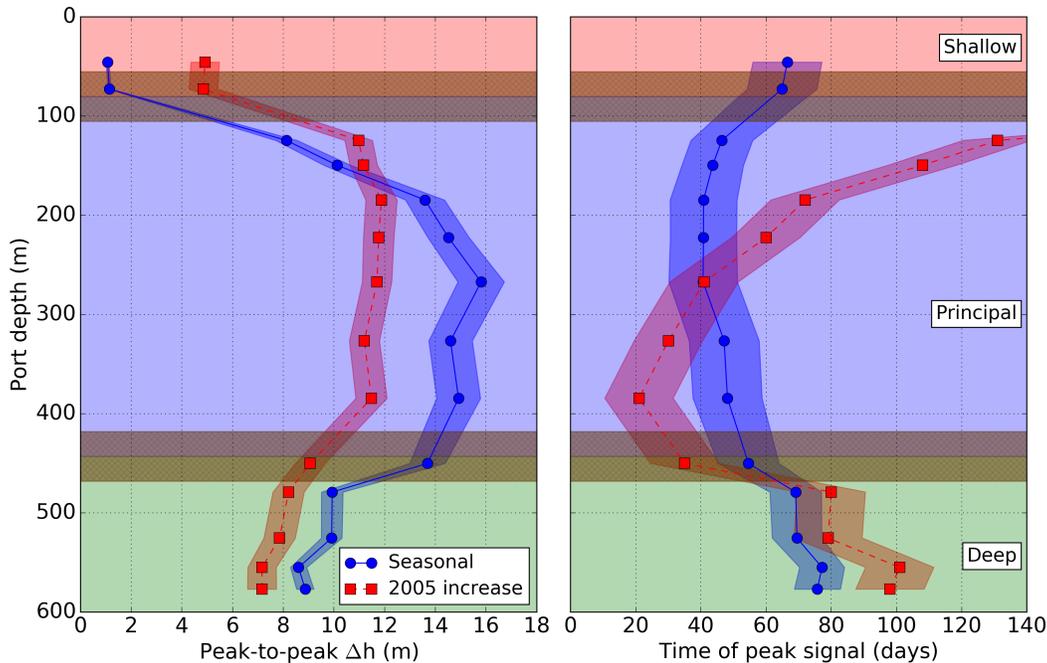


Figure 4.4: Depth profile of amplitude and phase delay of SAR-9 hydraulic head data. The blue lines correspond to the average seasonal oscillations while the red lines correspond to the uplift signal initiating in 2004/2005 due to a period of heavy rainfall. The horizontal shaded areas centered on the blue and red lines indicate the uncertainties associated with those values. The vertical shaded areas represent the depths corresponding to the three aquifer layers as estimated by the OCWD three-layer model, and the brown hatched regions indicate the approximate range of depths of the aquitards separating the aquifer layers.

the earliest times occur primarily in the principal aquifer. Spatial differences in the depths of the earliest peak times most likely represent spatial heterogeneity in conductivity due to the uneven distribution of aquifers and clay layers (Figure 4.B1).

The depth-dependent timing for the annual and transient signals are noticeably different. It appears to take longer for hydraulic head levels to equilibrate for the 2005 increase than for the annual cycle. These observations are consistent with the solution to the 1-D diffusion equation in Equation 4.13. The transient increase in groundwater levels for the 2005 heavy rainfall period has an effective period of  $\sim 4$  years while the seasonal cycle of recharge and pumping has a period of  $\sim 1$  year. The former has a slower diffusion speed but lower attenuation of head amplitudes within the aquifer whereas the latter has a higher diffusion speed but much more rapid attenuation of amplitude with depth. We conclude that the timing differences

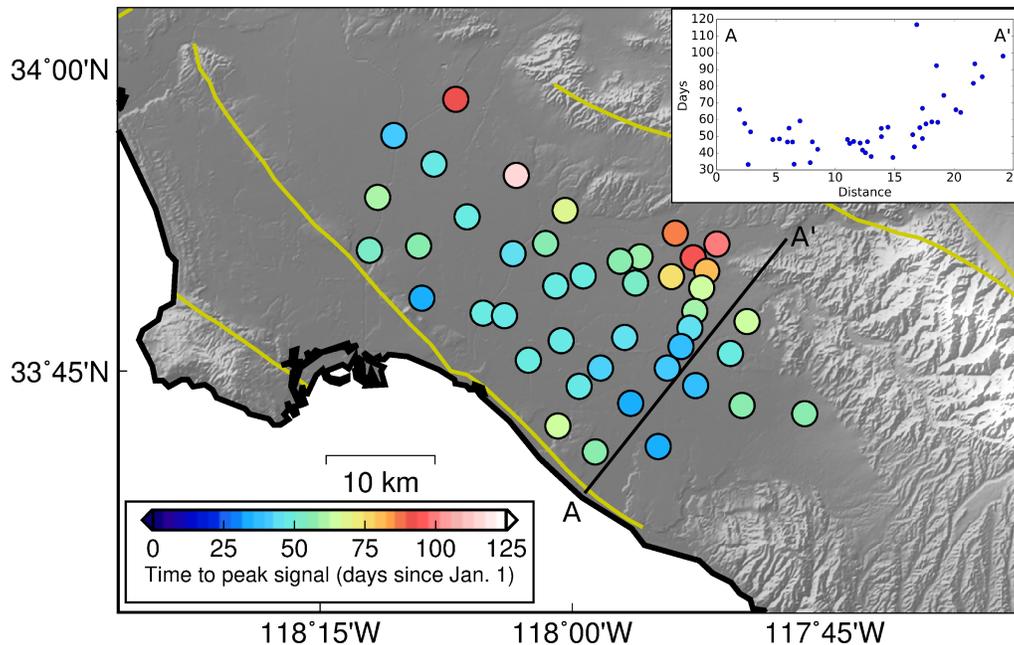


Figure 4.5: Time to peak seasonal signal for hydraulic head data from well ports located in the principal aquifer. Wells in the center of the basin reach their peak signals earlier in the year compared to wells in the margins of the basin. Inset shows timing of wells along the transect indicated by the A-A' line.

between these two processes is entirely due to variations in source behavior rather than any structural or material changes within the principal aquifer.

Since true groundwater flow is described by a three-dimensional diffusion process, we investigate timing differences in the horizontal direction by estimating the average time of peak annual recharge for all wells in the principal aquifer using the same modeling procedure in Equation 4.15. We can observe a clear propagation of time to peak signal from the center of the basin to the margins (Figure 4.5). The principal aquifer generally reaches peak groundwater levels between February and March in the central region of the basin while the margins peak between March and April. There is a steep gradient in timing in the forebay area of the Santa Ana Coastal Basin, which could represent a rapid change in hydraulic conductivity or thickness of the principal aquifer, which would act as an impedance to groundwater flow. Comparison of the propagation speeds of groundwater in the vertical and horizontal directions for the annual cycle reveals that propagation speeds are approximately one order of magnitude slower in the vertical direction than the horizontal direction (~30 m/day vs. ~300 m/day, respectively). This discrepancy is consistent with the general

observation that vertical conductivity is much lower than horizontal conductivity due to bedding laminae and the potential presence of laterally continuous clay beds within the main aquifer units that impede vertical groundwater flow.

### **Ground Deformation History**

From Equation 4.7, we know that ground deformation over a confined aquifer is approximately proportional to changes in hydraulic head under certain simplifying assumptions, mainly that aquifer or aquitard compression and expansion is elastic for effective stress levels less than the previous maximum effective stress. The proportionality constant in the elastic relationship is the skeletal storage coefficient,  $S_k$ . However, ground deformation can also be affected by inelastic compaction of clays in aquitards and interbeds. In this case, ground deformation will not be correlated with hydraulic head and is expected to vary exponentially with time [6]. The exponential relationship is derived from the theory of hydrodynamic consolidation and is used to describe the delayed response of fine-grained materials after effective stress levels have surpassed previous maximum levels.

Our approach is to start with the hypothesis that all deformation within the basin is elastic. For short-term signals driven by the annual cycle of groundwater pumping and recharge, this assumption is most likely to be true since typical production levels are a minor fraction of the storage capacity for the aquifer units in the basins [13]. Additionally, the common driving mechanism for inelastic deformation associated with compressible materials typically occurs over a long time span, although rapid, substantial stress increases can also lead to rapid inelastic deformation. In order to test the hypothesis of purely elastic deformation, we must compare short-term deformation signals with short-term variations in hydraulic head and long-term deformation signals with long-term variations in hydraulic head. As we discussed in Section 4.1, the deformation we observe on the ground surface is a result of the integrated compaction of aquifers and aquitards along the entire depth of the aquifer system. However, we also know that there is a time delay for a pressure perturbation to diffuse throughout an aquifer (e.g., Figure 4.4). This time delay is dependent on hydraulic conductivity along the diffusion path and the rate of pressure change at the pressure source (e.g., groundwater pumping rate). Therefore, different segments of the aquifers will be compacting/expanding at different times. By comparing ground deformation time series with hydraulic head time series at various depths, we can estimate the effective depth at which aquifer deformation is most correlated with ground deformation. In the ideal case, this depth would approximate the depth of a

hypothetical pressure source.

We examine vertical displacement data from the GPS station SACY, which is part of the SCIGN and is located about 1.4 km away from the SAR-9 well (Figure 4.2). We decompose both the GPS vertical displacements and the SAR-9 head time series into long- and short-term (seasonal) signals using a modified form of Equation 4.15. Instead of using sinusoids to model the seasonal signals, we use a linear combination of third-order B-splines (different than the integrated B-splines used for transient signals) to allow for seasonal signals with time-varying amplitudes and phase delays. We assign the temporal support of the B-splines such that the seasonal signal each year is described by a linear combination of B-splines spaced 0.2 years apart. With this approach, we could reconstruct seasonal signals with wide variations from year to year. In this study, we construct  $\mathbf{C}_m$  in Equation 4.16 such that  $B^i$ -splines are independent while B-splines are correlated with other B-splines that share the same centroid time within a year (e.g., B-splines centered in March are correlated with other B-splines centered in March). We assign the correlation strength for the B-splines to be exponentially decaying in time with a decay time of two years. The decay time was experimentally chosen in order to maintain the flexibility of the B-splines to model time-varying seasonal signals while still enforcing a level of coherency from year to year.

After decomposing the SACY vertical displacements and the SAR-9 hydraulic head time series at multiple port depths into long- and short-term components, we compute the Pearson correlation coefficients between each head time series and the GPS time series. We fit the correlation coefficient depth profiles with third-order polynomials in order to reduce the noise of the coefficient estimation and estimate the depth of maximum correlation between hydraulic head and ground deformation. We can use Equation 4.7 to estimate an  $S_k$  for each component to obtain scaled head time series. We note that these estimates of  $S_k$  will not reflect the true storage coefficient for a particular aquifer unit since ground deformation is caused by the vertically integrated response of the entire aquifer system.

For both the long- and short-term components, the vertical ground displacement data is well matched by hydraulic head variations at a given depth (Figure 4.6). The transition from regular groundwater fluctuations (caused by the annual cycle of recharge and pumping) to unsteady oscillations between 2008 and 2012 (caused by cessation of the STSS program) can be observed in the short-term GPS signal. Additionally, the long-term decrease in water levels and two heavy rainfall periods

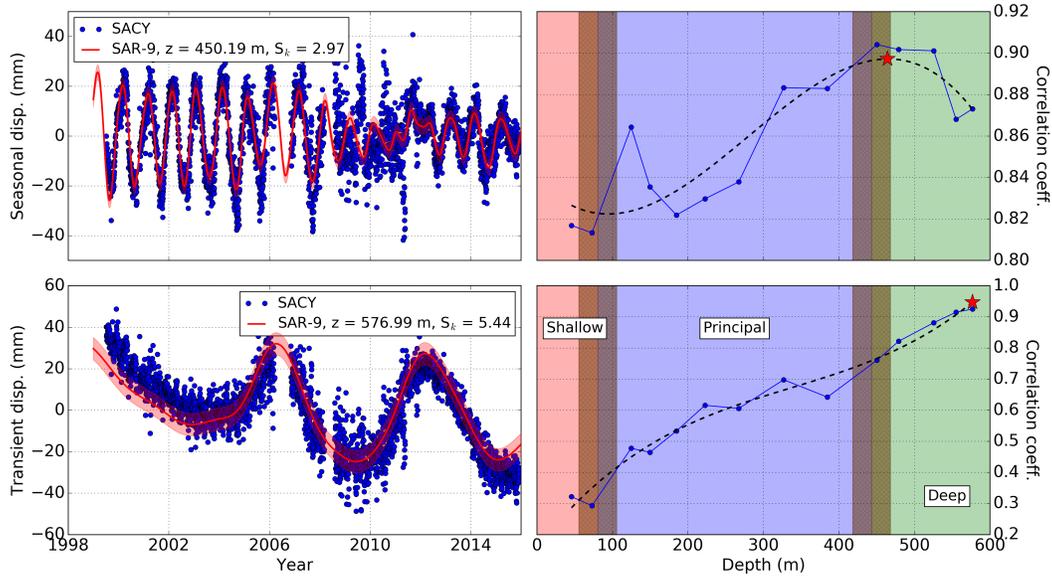


Figure 4.6: Cross correlation analysis between SAR-9 head data and SACY GPS data. The upper plots correspond to short-term, seasonal signals for both data sets while the lower plots correspond to long-term signals. We iterate over the time series at each port depth, compute the Pearson correlation coefficient between well and GPS data, and estimate the storage coefficient (scaling parameter) to best match the well and GPS data. The right plots show the Pearson correlation coefficient between the time series at each port depth and the SACY GPS time series. The dashed black line corresponds to a third-order polynomial fit to the correlation coefficients, and the red star indicated the location of maximum correlation. For the left plots, SACY data are shown with blue dots and the scaled well data for the depth of maximum correlation are shown solid red lines. The red shaded area represents the uncertainty in the scaling of the well data.

are manifested as vertical subsidence and transient uplift, respectively. We find that the seasonal ground deformation is best correlated with hydraulic head variations at a depth of about 450 m, which is the depth of the aquitard separating the principal and deep aquifers. Interestingly, the depth of best correlation for the long-term ground deformation is deeper ( $\sim 580$  m), and the estimated  $S_k$  for the deeper well is nearly twice the value of  $S_k$  for the shallower well (5.4 mm/m vs. 3 mm/m, respectively). The larger value of  $S_k$  is expected since the dynamic range of hydraulic head variations in the deep aquifer is lower than the principal aquifer. Overall, for both the long- and short-term signals, the GPS and scaled head time series are in very good agreement which suggests that at this location, the ground deformation is purely elastic and recoverable and no significant inelastic effects are present. In

the next section, we perform a similar comparison of hydraulic head to ground deformation using an 18-year InSAR time series. The spatially dense observations provided by InSAR allows us to assess the time-dependent ground deformation at every well location, which is an advantage over the sparse GPS network over the coastal basins.

#### **4.4 Central and Santa Ana Basin InSAR Time Series**

We use 165 SAR acquisitions from the European Space Agency ERS (European Remote Sensing) and Envisat satellites spanning from 1992 - 2011 to form 881 interferograms. The maximum perpendicular baseline (spatial separation between orbits) is 480 m, which is small for C-band SAR instruments. The temporal repeat times range from 35 to 210 days, although after 1995, the repeat times are generally 35 or 70 days which is sufficient to model most deformation signals observed in the GPS data. We use a coherence threshold of 0.4 to mask poorly resolved areas (such as over water) and any areas with unwrapping errors. Interferometric phase contributions due to topography are removed using a digital elevation model (DEM) produced by the Shuttle Radar Topography Mission (SRTM) with approximately 30 m spacing. We estimate and remove phase delays due to atmospheric effects using global atmospheric reanalysis data from the European Center for Medium-Range Weather Forecasts (ECMWF) [24]. We also remove long-wavelength signals due to orbital errors by estimating a two-dimensional linear ramp for each interferogram. Finally, we reference the time series to a  $400\text{ m} \times 400\text{ m}$  window co-located with the SCIGN GPS station SNHS which is located in an area of high coherence showing stable ground motion unaffected by groundwater pumping (Figure 4.2).

#### **Seasonal Amplitude and Phase Maps**

We first reconstruct the InSAR time series using the same time parameterization approach as Equation 4.15 where we assume the ground deformation can be described as a superposition of sinusoidal seasonal and transient effects and estimate the sinusoidal and  $B^i$ -spline coefficients independently for each pixel [23]. We limit our initial analysis of the InSAR time series to interferograms prior to 2008 to reduce the effect of the cessation of the STSS program on the time series reconstruction. In Sec. 4.A and 4.5, we will describe our method for performing a fully spatiotemporal time series analysis for the full time series to account for non-steady seasonal and transient deformation. At this point, our primary goal is to examine the characteristics of the steady seasonal deformation prior to 2008, and our experiments show that

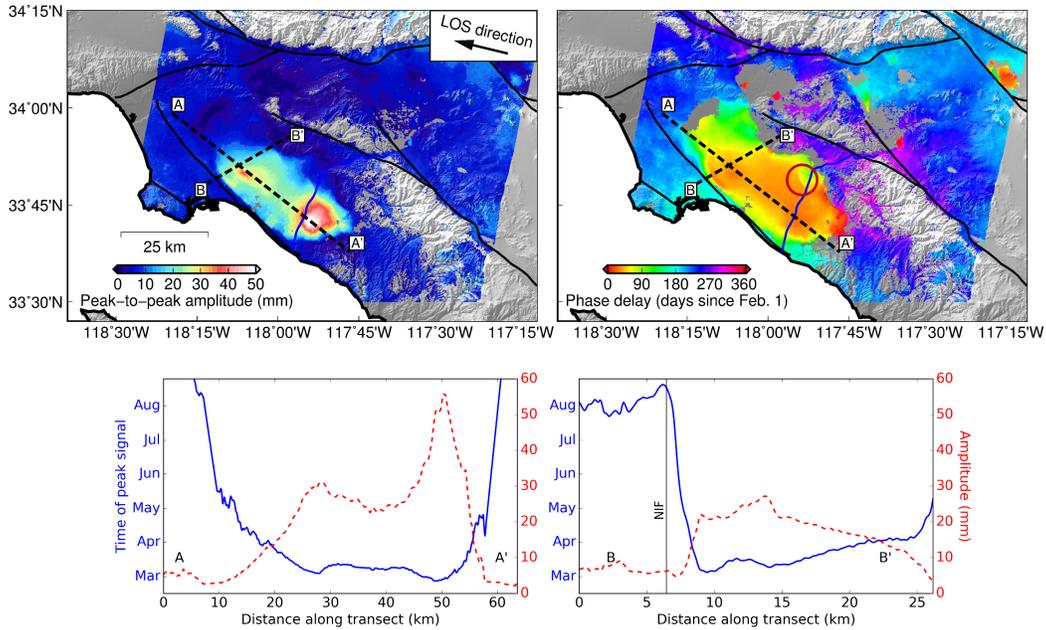


Figure 4.7: Maps of seasonal peak-to-peak amplitude and phase delay. The dashed black lines show the location of the transects used for the bottom plots. For the lower transect plots, blue lines correspond to the phase delay while the dashed red lines correspond to the peak-to-peak amplitude. The red circle in the map of seasonal phase delay shows the location of a discontinuity in amplitude and phase within the basin. The arrow indicates the satellite-to-ground line-of-sight (LOS) direction.

a pixel-by-pixel approach is suitable for estimating the coefficients of the sinusoidal components. We can then generate maps of seasonal amplitude and phase using Equation 4.17 for each pixel.

Maps of the estimated amplitude and phase for the annual (1 year period) seasonal signal between 1992-2008 show that most of the seasonal deformation is concentrated within the Los Angeles and Santa Ana coastal basins in the region corresponding to the confined aquifer (Figure 4.7). The maximum peak-to-peak amplitude is 5 cm in the southern end of the Santa Ana Coastal Basin, which agrees with the results obtained by Watson *et al.* [8] and Lanari *et al.* [9]. We can also observe a smaller pair of high-amplitude regions closer to Long Beach with amplitudes of 3 cm. The seasonal amplitude decreases rapidly outside of the confined aquifers, particularly in the western edge of the basin bounded by the Newport-Inglewood Fault (NIF) where the fault is an effective barrier to across-fault fluid flow. This effect can also be observed in the map of the seasonal phase where the ground east of the NIF has a peak signal in March whereas the ground west of the NIF has a peak

signal in July. The seasonal amplitude decreases rapidly to the north towards Los Angeles, although the seasonal phase map still shows a coherent boundary which may delineate the effective northern boundary of the confined aquifers.

In general, the seasonal amplitude appears to be inversely correlated with the seasonal phase, i.e. higher amplitude areas peak earlier in the year, which suggests that groundwater dynamics in the basin follow a standard diffusion process. The central areas of the basin experience the highest amplitudes and earliest peak times, and we observe delays as one approaches the margins of the basin. From Section 4.3, we observed a similar delay in hydraulic head from the well data, suggesting that the main driver of the delay in ground deformation from the central region of the basin to the margins is the time delay necessary for aquifer pressures to equilibrate in the horizontal direction. Furthermore, in the northwest area of the basin where aquifer thicknesses are relatively constant, we can observe an exponential decline in seasonal amplitude and linear variation in seasonal phase away from an amplitude peak (A-A' transect in Figure 4.7), which agrees with the diffusion solution for periodic head variations in Equation 4.13. We also observe a sharp discontinuity in both the seasonal amplitude and phase maps on the eastern edge of the basin where the Santa Ana River enters the forebay region indicating some form of impediment to groundwater flow. Here, the peak amplitude occurs in March on the west side of the discontinuity and in May on the east side. While no known faults exist in this area, this area corresponds to the approximate boundary between the forebay and confined zones. Comparison of the seasonal phase map to the thickness of the principal aquifer (using the aquifer model developed by OCWD) shows that the phase discontinuity is also roughly coincident with a rapid change in depth of the principal aquifer in the OCWD model (Figure 4.8). In Section 4.6, we explore dynamic models investigating the impact of a rapid change in aquifer thickness on groundwater flow.

### **Groundwater Pumping and Seasonal Ground Deformation**

We expect that groundwater pumping practices would have a strong impact on the amplitude of seasonal ground deformation, particularly during the STSS pumping program. Hydraulic head in a confined aquifer near a production well will experience a drawdown during periods of groundwater pumping [1, 25]. To explore the impact of pumping on head levels, we use groundwater production time series for 250 OCWD production wells that measure total groundwater pumping on a monthly basis. We create a 50×50 uniform grid where the dimension of each grid cell is

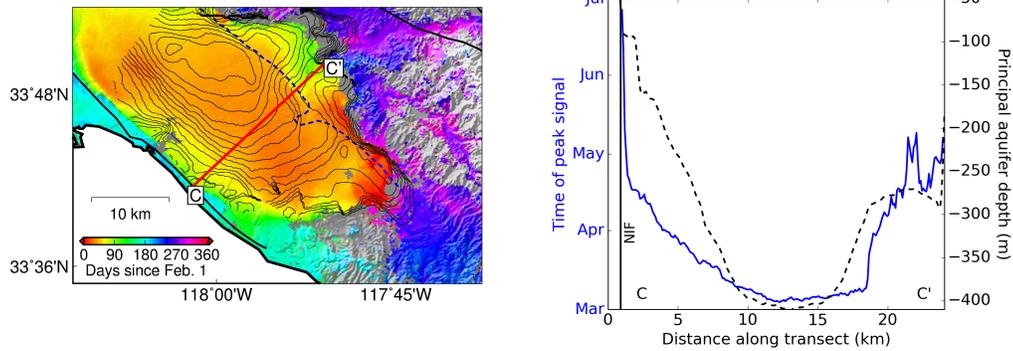


Figure 4.8: Map of seasonal phase delay (left) with contour lines corresponding to the depth of the bottom of the principal aquifer. The dashed blue line corresponds to the approximate boundary between the forebay and confined areas as defined by OCWD. The right plot shows the transects (C - C') for the seasonal phase (solid blue) and aquifer depth (dashed black).

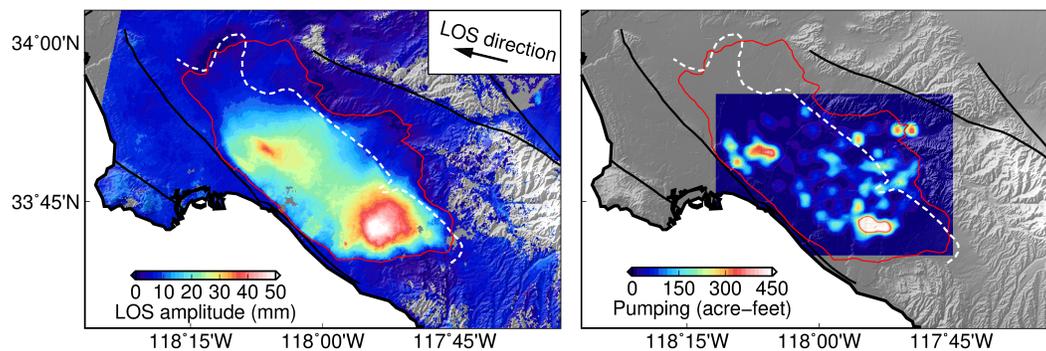


Figure 4.9: Maps of seasonal peak-to-peak LOS amplitude (left) and seasonal groundwater pumping (right). The locations with the highest seasonal pumping correspond to the locations of the highest seasonal ground deformation within the pressure area. The red outline encompasses the seasonal ground deformation as observed in the InSAR map of seasonal phase delay. The dashed white line indicates the approximate boundary between the forebay and confined areas.

approximately  $0.6 \times 0.6$  km and compute the amplitude of total seasonal groundwater pumping in each grid cell. We can clearly observe an association between the high ground deformation areas and areas with high seasonal groundwater production (Figure 4.9). The pair of high-amplitude ground deformation regions near Long Beach directly correspond to two regions of concentrated groundwater pumping. Additionally, the amplitude of seasonal groundwater pumping is positively correlated with the amplitude of seasonal ground deformation. However, not all areas of

high groundwater pumping are co-located with high ground deformation, particularly in the forebay regions. In those regions, the shallow aquifers are unconfined, and the principal aquifer is only semi-confined [13]. Therefore, large fluctuations in groundwater levels will only result in minimal ground deformation.

#### **4.5 Simultaneous Time Varying Seasonal and Long-Term Ground Deformation Using InSAR Time Series Analysis**

In order to study the full complexity of ground deformation in the Central and Santa Ana Coastal Basins, including the time-varying seasonal deformation and long-term subsidence, we now decompose the full InSAR time series from 1992-2011 into seasonal and transient components using a new method for geodetic time series analysis that extends on the sparse regularization methods of Chapter 2 by incorporating spatial coherency into the time series reconstruction (Section 4.A). This method estimates the coefficients of a temporal dictionary for every pixel simultaneously, resulting in a very large regularized least squares problem that uses data from all interferograms. The regularization function for the model parameters is a combination of a sparsity-inducing  $\ell_1$ -norm (to enhance interpretability of the time series model) and an  $\ell_2$ -norm (for parameters that may be correlated in space or time). We adapt a distributed convex optimization algorithm, the alternating direction method of multipliers (ADMM), to solve this problem in parallel using hundreds of CPU processors. This algorithm allows us to solve a least squares problem with several million parameters in only a few minutes; the end result is a self-consistent time series model that can be decomposed into long- and short-term signals of various timescales.

Prior to the time series decomposition, we limit the area of analysis to the coastal basins and downsample the pixels to a spacing of approximately 200 meters in order to reduce the computational load. We then populate the temporal dictionary with third-order B-splines with timescales of 0.2 years to model the seasonal ground oscillations caused by the annual cycle of groundwater pumping and recharge. Similar to the decomposition of the SACY GPS data in Section 4.3, we include coherency between B-splines that share the same centroid time within a year (in addition to the spatial coherency). We also include B<sup>i</sup>-splines in the temporal dictionary to model long-term, transient signals.

To validate the estimated InSAR time series model, we compare the decomposed long- and short-term signals with long- and short-term signals measured by three

GPS stations within the Santa Ana Coastal Basin. We use the method of Riel *et al.* [26] to perform the time series decomposition for the GPS time series independently from the InSAR time series. As with the InSAR data, we use B-splines for seasonal signals and  $B^i$ -splines for long-term signals with a sparsity-promoting regularization scheme to limit the total number of  $B^i$ -splines needed for reconstruction of the GPS data. We then project the three-component GPS time series onto the radar LOS direction. The long- and short-term signals from the two data sets are in very good agreement, even during time periods where we have a gap in temporal coverage of SAR acquisitions (Figure 4.10). As expected from the SAR-9 head time series, the short-term, seasonal ground deformation is relatively steady from 1996 to 2000, followed by a slight decrease in amplitude of oscillations from 2000 to 2007. After 2007-2008, the seasonal signal changes its temporal pattern significantly as a result of the cessation of water storage programs that encouraged regular, increased groundwater pumping during the summer months, which is also measured by the GPS stations. The long-term signal once again shows long-term subsidence which is interrupted by a 2-year uplift period associated with heavy rainfall between 2004-2005. In the following sections, we use the spatial continuity of our InSAR time series model to examine the spatial behavior of the unsteady seasonal signals, as well as the long-term subsidence and rainfall-induced uplift.

### **Short-Term Basin Deformation**

We isolate the short-term basin deformation using the B-spline coefficients estimated from the InSAR time series analysis and compare the March to September ground deformation for three different years: 2004, 2008, and 2009 (Figure 4.11). For 2004, the March to September basin subsidence is in good agreement with the seasonal amplitude map in Figure 4.7, which is expected since the STSS program is still active during this time period. Peak subsidence is greater than 40 mm in the southeast zone of the Santa Ana Coastal Basin. However, in 2008, we observe a significant reduction in ground deformation throughout the whole basin. Peak subsidence only reaches 10-15 mm primarily in areas closer to the coast and in the center of the basin. In 2009, the spatial pattern changes again, with peak subsidence now reaching about 30 mm in a concentrated zone in the southern edge of the basin close to the main source of groundwater pumping. For both 2008 and 2009, we see that most of the seasonal ground deformation closer to Long Beach in the western half of the basin disappears, and most of the deformation in the forebay region is minimized as well.

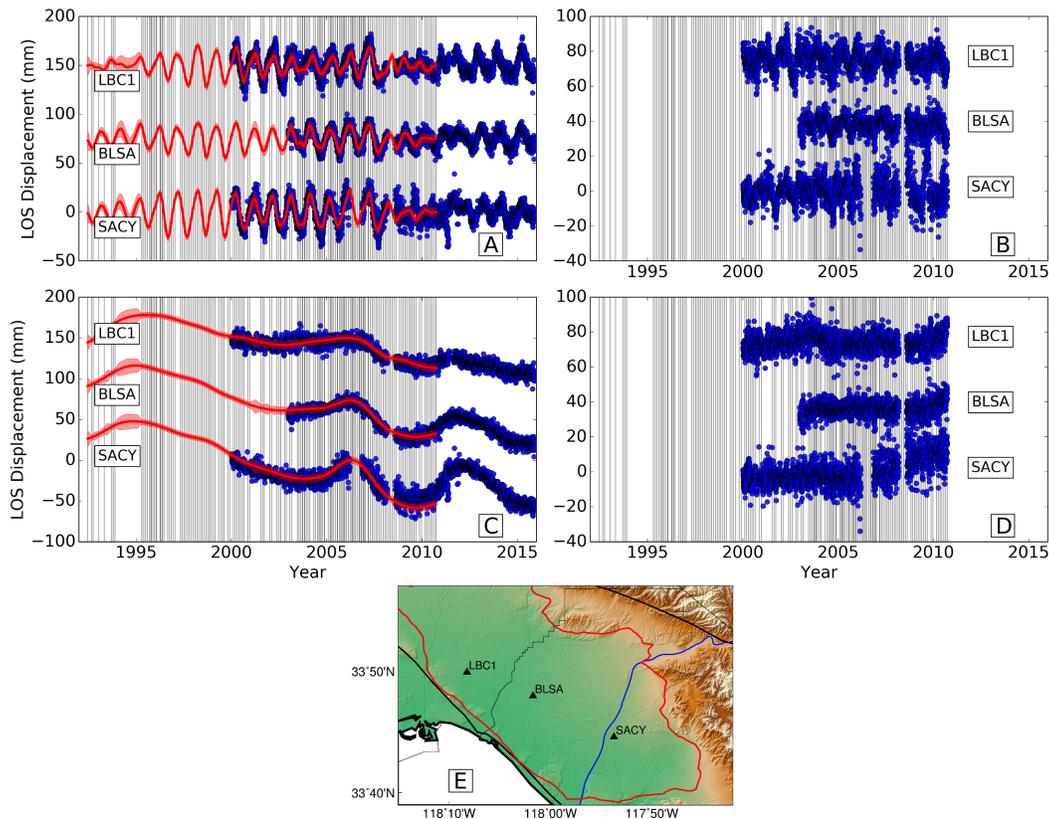


Figure 4.10: InSAR time series reconstruction compared with three GPS stations (LBC1, BLSA, SACY) located in the basin. The GPS data are projected onto the radar LOS direction. Both the InSAR and GPS time series have been decomposed into short-term, seasonal signals (A) and long-term, transient signals (C). GPS data are shown with blue dots, and InSAR data are shown with solid red lines. The shaded red regions in (A) and (B) represent the uncertainties in the reconstructed InSAR time series. Note the increase in uncertainty during time periods with no SAR acquisitions. Thin vertical black lines correspond to the SAR acquisition times. (B) and (D) show the residuals between the GPS and InSAR data in (A) and (C), respectively. The locations of the GPS stations are shown in (E).

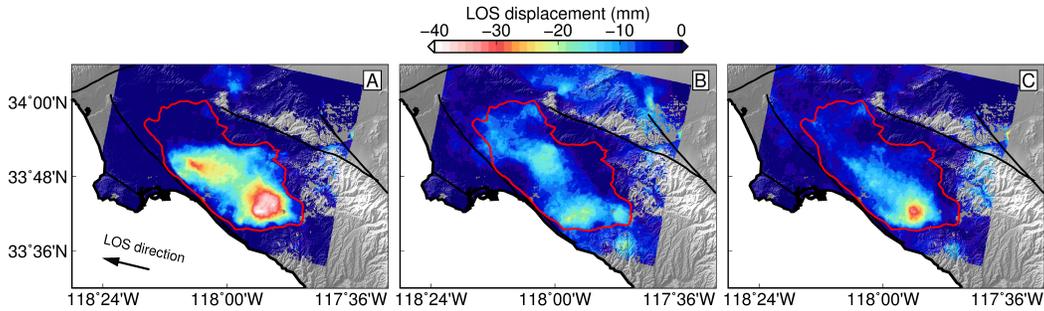


Figure 4.11: Maps of March to September LOS deformation for three different years: (A) 2004; (B) 2008; and (C) 2009. The 2004 ground deformation is consistent with the maps of seasonal amplitudes, whereas the 2008 and 2009 maps show starkly different spatial patterns and amplitudes due to cessation of the STSS program.

### Long-Term Basin Deformation

By examining the spatial distribution of the  $B^i$ -spline coefficients as estimated by the ADMM time series analysis, we can determine the onset times and durations of the dominant long-term, transient signals in our time series. The sparsity-promoting regularization forces most of the coefficients to be zero except for those associated with several longer-term deformation signals observed primarily within the coastal basins (Figure 4.B2). Firstly, we detect a  $\sim 2.5$ -year subsidence signal starting in 2007 that is contained within the region of the basin defined by the phase delay of the seasonal deformation. From the comparison between GPS and hydraulic head data in Section 4.3, it is likely that a large fraction of this observed subsidence is due to an overall reduction in groundwater levels in the principal and deep aquifers. The subsidence was preceded by a 2-3 year uplift signal in the southeastern portion of the Santa Ana Coastal Basin starting in late 2004, which we observed in the data for GPS station SACY and is due to a period of heavy rainfall during the winter of 2004 - 2005 (Figure 4.3). This period of heavy rainfall has also been associated with transient uplift due to recharging aquifers in the nearby San Gabriel Valley [27].

Profiles of the deformation patterns for the three main long-term events (1995 - 2004 subsidence, 2004 - 2007 uplift, and 2007 - 2010 subsidence) suggest that the two subsidence periods are connected and act as a single subsidence process that is distinct from the uplift period due to heavy rainfall (Figure 4.12). The subsidence is strongest in the center of the basin and is roughly coincident with the area of high seasonal deformation. Additionally, we observe that the subsidence rate increased by a factor of 2 after 2007. The strongest subsidence signal actually occurs within

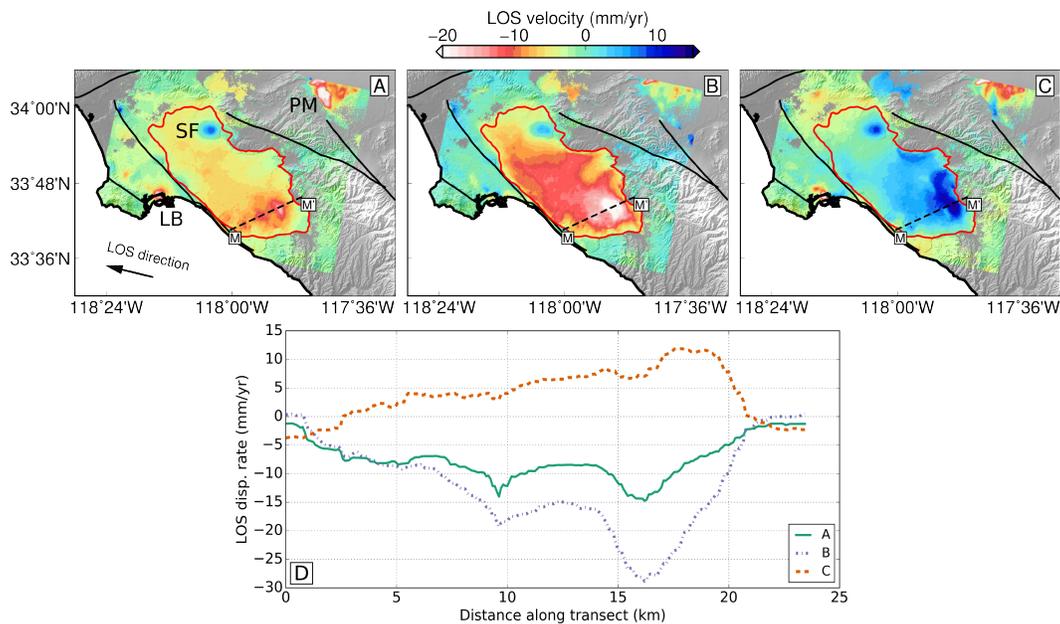


Figure 4.12: (A)-(C) show maps of subsidence and uplift rates for the time periods 1995 - 2004 (A), 2007 - 2010 (B), and 2004 - 2007 (C). The labels in (A) indicate the subsidence signature in Long Beach (LB) due to the Wilmington oil Field, uplift in the Sante Fe Springs Oil Field (SF), and subsidence in Pomona Basin (PM) due to groundwater pumping. The transect profile is shown by the black dashed line, M - M'. (D) The solid green line corresponds to the transect for the 1995 - 2004 time period, and dashed blue line corresponds to 2007 - 2010, and the red dashed line corresponds to 2004 - 2007. The transect data show very consistent deformation rate profiles for the two time periods associated with subsidence, whereas the uplift profile has a distinctly different profile.

two narrow regions of the southeastern basin with length scales of approximately 3 - 5 km. In these regions, subsidence exceeded 100 mm between 1995 and 2004 and reached 80 mm between 2007 and 2010. The rest of the subsidence is broadly distributed throughout the rest of the basin. During the uplift event, which was caused by heavy rainfall between 2004 and 2005, most of the uplift is concentrated on the eastern margin of the basin where we observe ~30 mm of uplift. This area of the basin is where the Santa Ana River flows into basin and is close to the main point of entry for artificial recharging operations. This area is also coincident with the forebay area of the basin which is a mix of confined (principal) and semi-confined (shallow) aquifers [13]. The concentration of uplift in this region supports recharge in the shallow and principal aquifers due to heavy rainfall.

We can also observe long-term signals from 1995 to 2004 not associated with

groundwater activities in the coastal basins, such as uplift in the Sante Fe Springs Oil Field and subsidence in the Wilmington Oil Field in Long Beach. For the latter, episodic subsidence has been observed due to historically high oil production despite modern re-pressurization efforts [28]. For the former, the uplift mechanism is unclear since extraction rates have generally been higher than injection rates which would usually lead to net subsidence [4]. We can also observe subsidence in the Pomona Basin due to groundwater pumping. We do not observe large tectonic signals from fault slip due to the high subsidence rates in the basin. The NIF is perhaps the best observed fault system in this data set due to high interferometric correlation, and it is estimated to have a slip rate of approximately 1 mm/yr which is likely to be obscured by the groundwater-driven subsidence in the basin [4, 10].

The modeled transient time series for points within the narrow, high-subsidence areas exhibit much more rapid subsidence than representative points in other regions of the basin (Figure 4.13). For the eastern high-subsidence zone (represented as point (i) in Figure 4.13), the subsidence accelerates after 2007 at a higher rate than the western high-subsidence zone (point (iii)), which can also be observed in Figure 4.12. By comparing the spatial pattern and location of these high-subsidence areas to the seasonal amplitudes and time to peak seasonal signal, it is clear that the subsidence process for these areas is distinct from the normal hydrological cycle of groundwater recharge, pumping, and steady decline. In fact, the seasonal time series for points within and adjacent to the high-subsidence areas are nearly identical (Figure 4.13C). We had previously observed that the long-term ground deformation for the GPS station SACY can be fully explained by long-term changes in groundwater levels in the principal and deep aquifers (Figure 4.6). Unfortunately, SACY and other OCWD wells lie just outside the high-subsidence regions, preventing us from determining whether rapid decline of water levels is responsible for the large ground deformation there. However, the small spatial wavelength of these features suggests that the subsidence mechanism is not directly related to water and could be due to long-term, inelastic compaction of aquitards or laterally compact clay lenses in the aquifers.

The long-term subsidence signals for both 1995 - 2004 and 2007 - 2010 and the seasonal ground deformation amplitude show relatively consistent spatial patterns characterized by peak ground deformation in the southeast area of the basin near Irvine and Santa Ana and larger deformation in the center of the basin as compared to the margins. By scaling the seasonal amplitude map by a factor of -4 (to allow the seasonal amplitudes to best match the subsidence values from 1995 - 2004), we

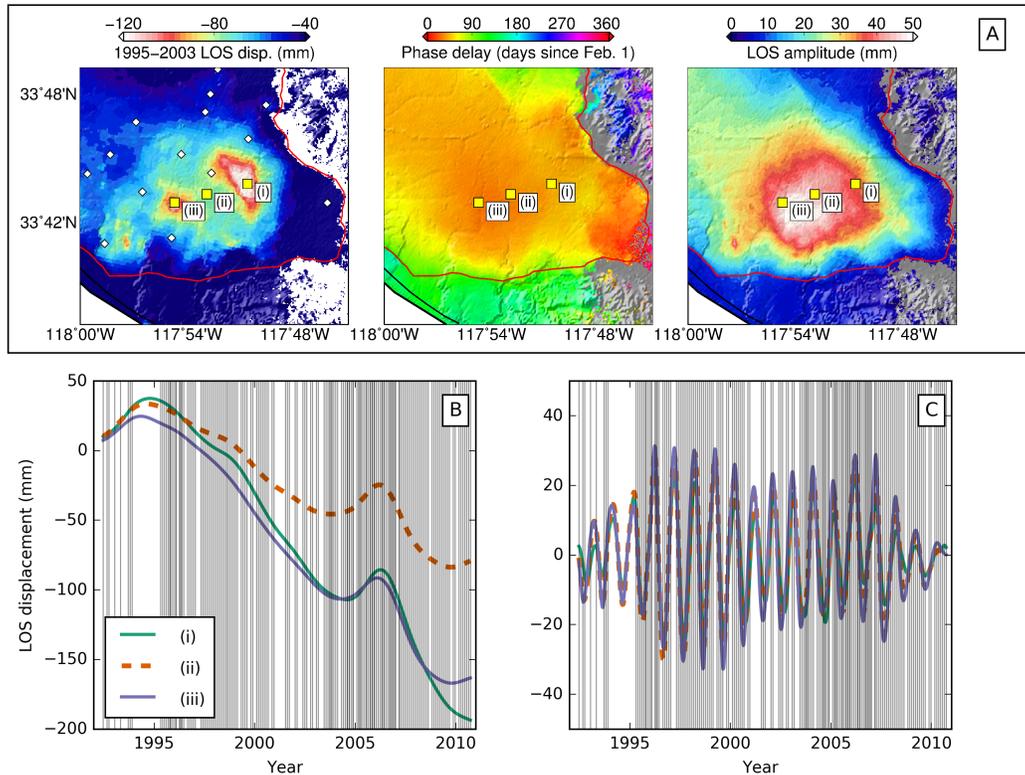


Figure 4.13: (A) Close-up maps of long-term LOS displacement from 1995 - 2004, seasonal phase delay, and seasonal peak-to-peak amplitude. The white diamonds in the subsidence map correspond to large-scale OCWD production wells. The yellow squares show the locations of the time series in (B) and (C). (B) shows the time series for the long-term displacements while (C) shows the time series for the short-term, seasonal displacements.

can create a map of *residual* subsidence to isolate the differences between the long- and short-term signals (Figure 4.14). The residual subsidence along the margins of the basin show that the spatial extent of long-term ground subsidence is larger than the extent of seasonal ground oscillation. Additionally, the two zones of positive residuals co-located with the zones of peak seasonal groundwater pumping also show that the subsidence signal is primarily driven by a longer spatial wavelength decline in groundwater levels and is not directly responding to yearly pumping variations. Of the two narrow zones of high-subsidence, only the larger zone is noticeable in the residuals, suggesting that the smaller subsidence zone may be related to long-term pumping trends. The persistence of the larger high-subsidence zone in the residuals supports the idea that this feature is caused by long-term, inelastic compaction of clay lenses since this localized area does not correspond to

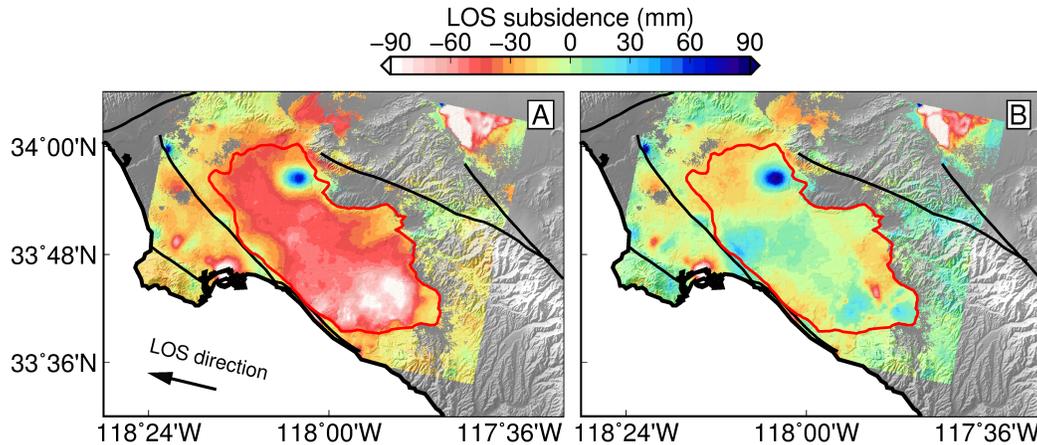


Figure 4.14: Comparison between long-term LOS subsidence from 1995 - 2004 (A) and residual subsidence after removing a synthetic subsidence signal corresponding to the scaled seasonal displacement (B).

any large-scale production wells.

### Ground Deformation vs. Hydraulic Head for the Santa Ana Coastal Basin

In Section 4.3, we compared hydraulic head time series from OCWD well SAR-9 with vertical ground deformation time series from GPS station SACY and found that short-term, seasonal ground deformation was most correlated with short-term head variations near the aquitard layer separating the principal and deep aquifers. On the other hand, long-term ground deformation was most correlated with long-term head variations in the deep aquifer. We expand on that analysis by comparing InSAR-derived ground deformation to the OCWD wells within the deforming areas of the coastal basins (we exclude the WRD wells in this analysis because the data only span back to the year 2000). For each well location, we iterate over the hydraulic head time series for each port depth and compute the correlation coefficient with the InSAR-derived ground deformation. We then use a third-order polynomial to estimate the depth of maximum correlation for both seasonal and long-term deformation and estimate  $S_k$  using the well port closest to the maximum correlation depth.

For both seasonal and long-term signals, the depths of the well ports most correlated with ground deformation are deeper in the center of the basin than in the margins, which roughly follows the geometry of the principal aquifer (Figure 4.15). For seasonal signals, the depth of best correlation is generally at or shallower than the depth of the aquitard separating the principal and deep aquifers, similar to the

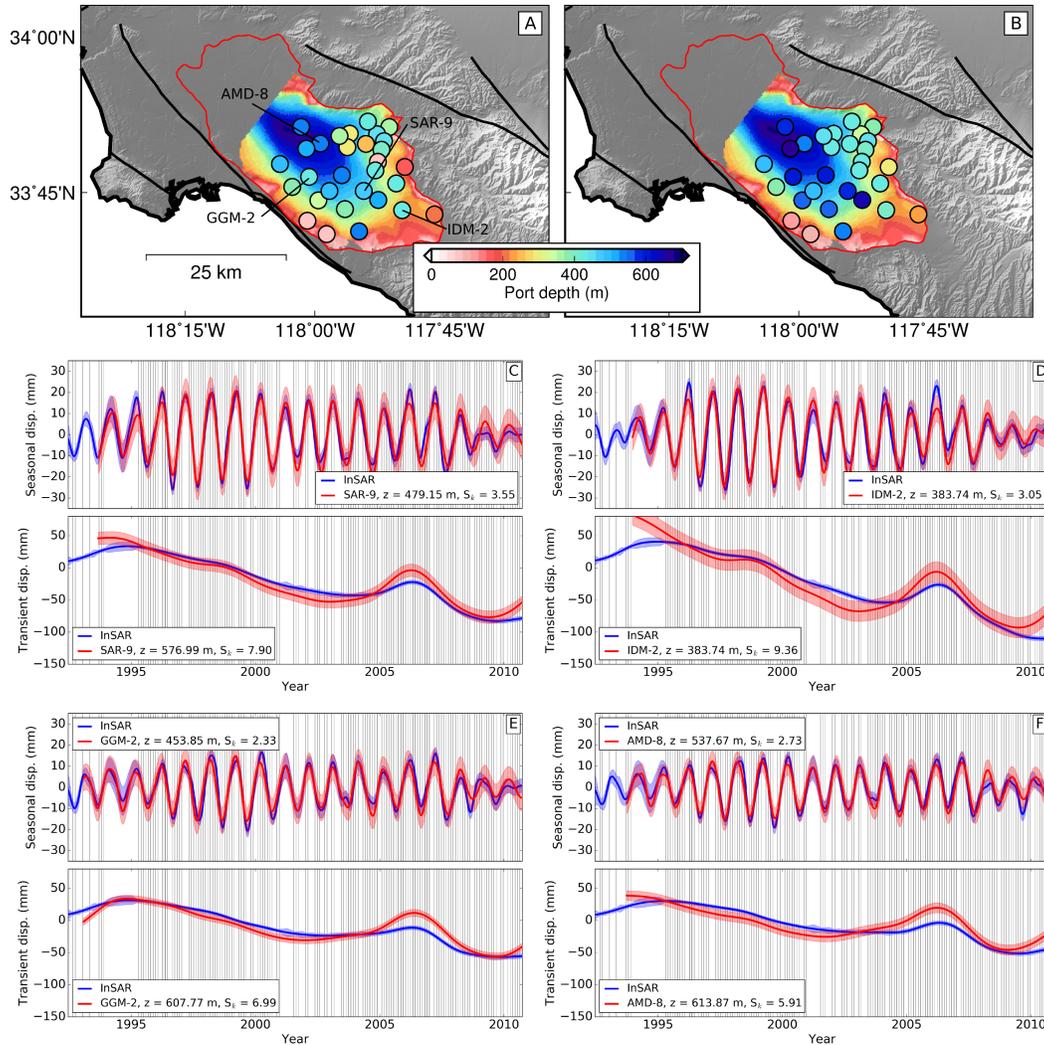


Figure 4.15: Depth of maximum correlation between hydraulic head data and InSAR displacement are denoted with colored circles for OCWD wells for seasonal signals (A) and long-term signals (B). The background color in the basin corresponds to the depth of the bottom of the principal aquifer. (C)-(F) show the InSAR time series (blue) and scaled head for the well port with maximum correlation (red) for select wells in the basin.

SACY and SAR-9 comparison. However, the depth of best correlation for the long-term signals is systematically deeper than the aquitard, particularly in the center of the basin. These depths occur primarily within the deep aquifer, and the long-term hydraulic head time series require larger values of  $S_k$  to match the ground deformation. We note that since monitoring wells do not fully penetrate the entire aquifer system, the true depth of best correlation for the long-term signals may be deeper than the ones shown here. Both seasonal and long-term ground deformation are matched very well with the scaled head data, which supports the hypothesis that the observed ground deformation within the coastal basins can be fully attributed to an elastic response to changing groundwater levels (with the exception of the narrow, high-subsidence zones discussed in the previous section).

## 4.6 Discussion

### **InSAR and Groundwater Management**

The total level of seasonal groundwater pumping by OCWD is well correlated with seasonal changes in hydraulic head in the principal aquifer. As a consequence, the areas with the highest levels of seasonal groundwater pumping directly correspond to the areas with the highest amplitude of seasonal ground deformation (Figure 4.9). We also observed that after the cessation of the STSS program in 2008, the spatial pattern of ground deformation during the summer months changed drastically. To compare the overall relationship between groundwater pumping and basin deformation, we use the long- and short-term InSAR results to construct time series of spatially integrated basin deformation. Similarly, we use the OCWD production wells to construct a time series of total groundwater production, and we decompose the production time series into long- and short-term signals using the same method we used for the SAR-9 time series.

As expected, the seasonal basin deformation is out of phase with the seasonal groundwater production by about 4-5 months (Figure 4.16). The basin surface reaches peak levels around March after several months of low groundwater production and heavier rainfall. Peak groundwater production then occurs in July/August, although production is generally high between May to September during the STSS program. Peak production is then followed by maximum ground subsidence around September. The cessation of the STSS program meant that production levels are more evenly distributed throughout the year. The long-term subsidence signals are relatively consistent with the long-term variation in groundwater production. From 1995 to 2002, a steady increase in production corresponded to steady subsidence

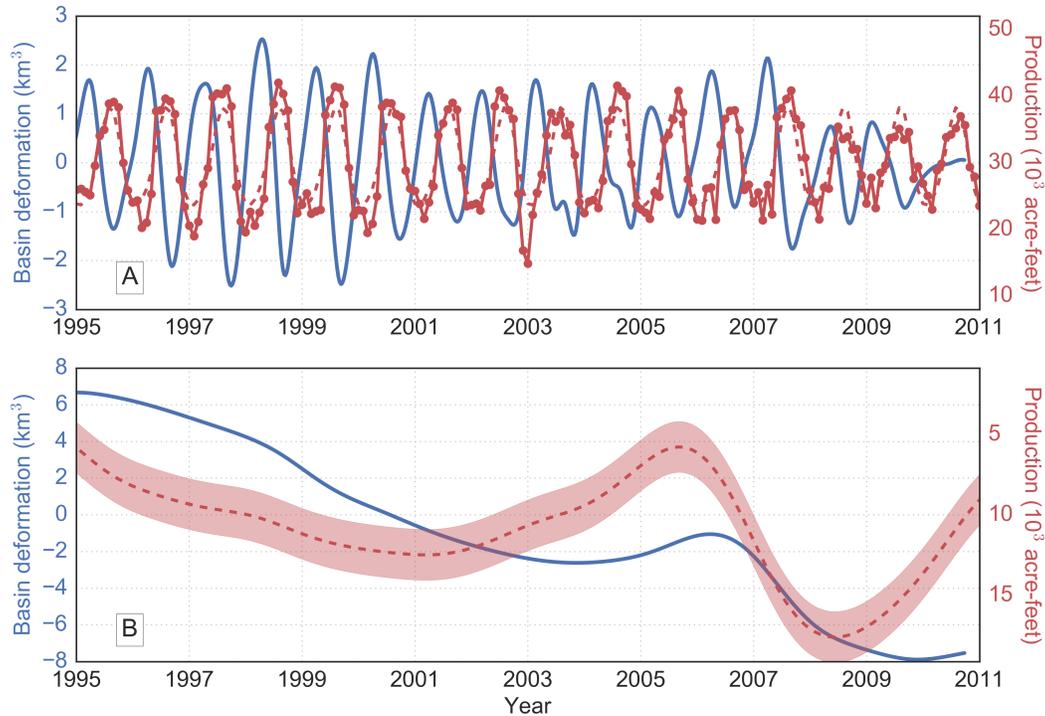


Figure 4.16: Time series of total volumetric basin deformation vs. total OCWD groundwater production for seasonal signals (A) and long-term signals (B). The volumetric basin deformation was computed by summing the product of pixel area and displacement at each pixel. The dashed red lines correspond to a model of the production time series consisting of annual and semiannual sinusoids for (A) and  $B^i$ -splines for (B). In (B), the data are not shown for visual clarity, and the red shaded area indicates the 1-sigma scatter of the data.

within the basin. The basin uplift initiating in 2005 is a net response to heavy rainfall (Figure 4.3) and decreased production, both of which led to historically high water levels and basins at near-full levels [13]. The increased basin subsidence rate following the uplift appears to be caused by an increase in production rate starting in late 2005. Therefore, we conclude that the majority of both the long- and short-term ground deformation are elastic responses to changes in head levels caused by changes in total groundwater production and recharge. The consistency between deformation and head levels, and the lack of evidence of any large-scale inelastic compaction, requires that storage coefficients remain constant in time, which means that stress levels within the aquifer and aquitards have remained above their preconsolidation levels since the 1990s.

### Estimating Depth-Dependent Storage Coefficients

Estimation of storage coefficients for the multi-aquifer systems in the Central and Santa Ana Coastal Basins is complicated by the fact that ground deformation is the integrated compaction with depth of all aquifer and aquitard layers. For many of the monitoring wells in the groundwater basins, the ports do not span all of the main aquifer layers (shallow, principal, and deep), so we cannot fully assess the contribution of each layer to the observed ground deformation. Nevertheless, we can compare ground deformation at specific points to wells with ports that span all layers, such as the GPS station SACY and the OCWD well SAR-9. For a vertical profile of the multi-aquifer system discretized onto a uniform grid, we can express the ground deformation at the surface as:

$$\Delta d = \sum_{i=1}^5 S_{s_i} \sum_{j=1}^{N_i} \Delta h_j \Delta z, \quad (4.19)$$

where  $S_{s_i}$  is the specific storage coefficient for the  $i$ -th layer which consists of  $N_i$  elements of uniform thickness  $\Delta z$ , and  $h_j$  is the hydraulic head for the  $j$ -th element in layer  $i$ . Here, we assume five total layers (three aquifer layers and two aquitards) with depths specified by the OCWD aquifer model. Since monitoring wells only coarsely sample the aquifer layers, at each observation epoch we interpolate the hydraulic head to a regular depth grid before computing  $\sum_j \Delta h_j$  for each layer. We can then estimate the specific storage for all layers using linear least squares. We constrain the estimated specific storage values to be above a small negative value to enforce positivity while allowing for errors due to the depth interpolation scheme.

For the seasonal ground deformation, compaction of the aquitard separating the principle and deep aquifers provides the dominant contribution to ground deformation due to the high estimated specific storage (Figure 4.17). We can use Equation 4.7 to convert the specific storage values to storage coefficients by multiplying by the thicknesses of the corresponding layers, and we find that the storage coefficient for the aquitard is  $\sim 2$  mm/m, similar to the value we estimated in Section 4.3 for the SAR-9 head time series best correlated with the SACY GPS displacements. For the long-term ground deformation, the estimated storage coefficient for the deep aquifer is  $\sim 8$  mm/m, which is slightly higher than the value we estimated in our previous analysis. The higher value here is required to compensate for the small negative values for specific storage in the top three layers, which are non-physical but necessary for numerical stability. The very small values for storage coefficients we estimate for the shallowest layers are most likely due to the limited resolving

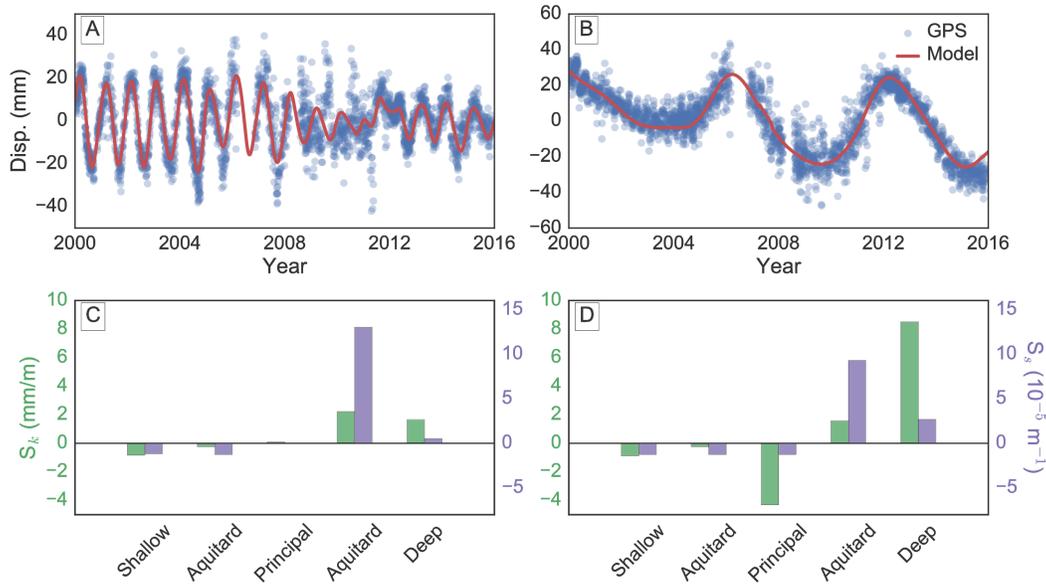


Figure 4.17: Estimation of storage coefficients for each aquifer and aquitard layer using vertical ground displacements from GPS station SACY and hydraulic head from OCWD well SAR-9. GPS data are shown in blue dots while the model (summation of hydraulic head in each layer scaled by a storage coefficient) is shown in red for short-term (A) and long-term (B) variations. (C) and (D) show the estimated storage coefficients in green and specific storage in magenta for short- and long-term signals, respectively.

power of surface deformation measurements. Hydraulic head variations have the lowest amplitudes in the shallower layers (Figure 4.4), resulting in an overall lower contribution to surface deformation. The constrained least squares procedure we utilize here will thus tend to encourage small specific storage values for those layers.

Both short- and long-term signals require a high specific storage for the aquitard separating the principal and deep aquifers which is consistent with a higher concentration of clay. For the long-term signal, a higher specific storage was estimated for the deep aquifer, resulting in a significantly larger storage coefficient due to the large thickness of the deep aquifer. However, the larger storage is most likely not a result of any physical change in storage coefficient between the two signals. Instead, the higher value is a consequence of a larger contribution of compaction of the deep aquifer to the observed ground displacement due to the lower temporal frequency of the diffusion process. From Equation 4.13, a lower temporal frequency will result in a lower attenuation of amplitude for a given periodic pressure perturbation and a slower diffusion speed. Thus, the long-term signal causes a greater overall change in

pore pressure for the deep aquifer than the short-term signal which attenuates more quickly. The relatively large specific storage required for the deep aquifer implies a higher bulk clay content as compared to the sand-dominant principal aquifer.

For comparing ground deformation to hydraulic head throughout the groundwater basins, we used an alternative approach of estimating a specific depth for each well that resulted in the highest correlation between deformation and head. We found that the short-term, seasonal ground deformation was most correlated with wells that were located close to or above the aquitard between the principal and deep aquifers. The high similarity between short-term head time series in the principal aquifer (Figure 4.6) and fast diffusion speeds (Figure 4.4) indicate that seasonal ground deformation can probably be explained as a combination of bulk deformation of the principal aquifer and compaction of the adjacent aquitard. We also found that a large area of the long-term subsidence signal was most correlated with wells located in the deep aquifer, which supports our conclusion that the lower temporal frequency of the long-term pressure perturbation causes greater pressure changes in the deep aquifer. Therefore, we can interpret the long-term decline in head level in the deep aquifer as a delayed response to long-term variations in groundwater levels, which are driven by a combination of long-term variations in groundwater production and isolated periods of heavy rainfall. The time delay is due to the lower permeability of the aquitard separating the two aquifers.

The narrow, high-subsidence regions shown in Figures 4.12 and 4.14 may correspond to inelastic compaction of laterally finite clay lenses in one of the aquifer layers, particularly the larger of the two subsidence regions. The high amplitude of subsidence may also indicate that these clay lenses are substantially thicker than other compressible clay units within the aquifers. This region does not correspond to any large-scale production wells, so the higher subsidence rate is most likely not caused by pumping activity since the 1990s. Instead, we are most likely observing a combination of inelastic compaction starting several decades earlier and elastic subsidence due to long-term groundwater decline. While we do not have well data directly over those regions to perform a cross-correlation analysis between hydraulic head and ground deformation, we did observe that the spatial signature of those regions does not appear in the map of seasonal amplitude of ground deformation.

### **Aquifer and Fault Structure from InSAR Seasonal Deformation**

The annual cycle of groundwater pumping and recharge, which was relatively constant from the early 1990s to 2008 due to the Short Term Seasonal Shift (STSS) program, resulted in regular ground deformation which oscillated in tandem with groundwater levels. In particular, we found that the short-term, annual ground oscillations were most correlated with fluctuations in hydraulic head in the principal aquifer. The areas with the highest seasonal amplitudes in ground deformation were co-located with the areas of highest seasonal groundwater pumping. Additionally, we found that the amplitude of ground deformation was positively correlated with the total amount of groundwater pumping during the summer months, as well as the level of hydraulic head drawdown as a result of the pumping.

Both the seasonal amplitude and phase maps suggest groundwater dynamics that follow a standard diffusion process where groundwater flows towards the zones of high pumping, leading to a drawdown in hydraulic head. However, we also observed sharp discontinuities in the seasonal phase due to the NIF and also within the basin away from any known fault systems (Figure 4.7). For the discontinuity within the basin, the seasonal phase changed from mid-March on the west side to late-April on the east side with an amplitude decrease of about 4 cm. From Equation 4.7, we know that a rapid change in ground deformation can be caused by either a rapid change in hydraulic head, aquifer/aquitard thickness, or hydraulic diffusivity. We observed that the thickness of the principal aquifer changed relatively rapidly in the location of the phase/amplitude discontinuity, which could mean that aquifer thickness controls the diffusion of hydraulic head throughout the aquifer. While the effectiveness of faults as barriers to groundwater flow is known, the effect of aquifer geometry on groundwater flow is still relatively uncertain.

To investigate the effects of faults and aquifer thickness on observed ground displacement, we perform a series of numerical experiments based on the groundwater flow diffusion equation (Equation 4.11). We perform numerical simulations for three different scenarios: 1) constant hydraulic conductivity and aquifer thickness; 2) constant hydraulic conductivity and a step decrease in aquifer thickness; and 3) a narrow zone of low hydraulic conductivity and constant aquifer thickness. The last scenario is representative of a narrow fault zone characterized by low permeability fault gouge. Equation 4.11 can be discretized using finite differences (central difference in space and forward difference in time) to obtain an explicit Euler update scheme to integrate the hydraulic head forward in time. We construct the  $x$ -domain

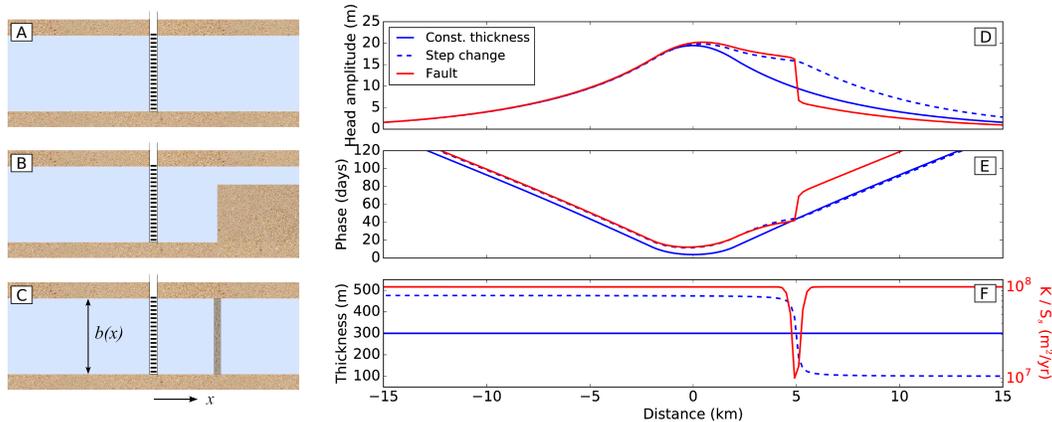


Figure 4.18: (A)-(C): model setup for one-dimensional diffusion simulation with a periodic pressure source placed in the center of the domain. The three different models are: (A) constant hydraulic conductivity and aquifer thickness; (B) constant hydraulic conductivity and a step decrease in aquifer thickness; and (C) narrow fault zone of low hydraulic conductivity and constant aquifer thickness. (F) shows the specific values for thickness and hydraulic conductivity for the three different scenarios. The amplitude and phase of the simulated hydraulic head in response to the periodic pressure source are shown in (D) and (E), respectively. The colors corresponding to the different models are: (A) blue, (B) dashed blue, and (C) red.

to span from -40 to 40 km and choose a spatial resolution and time step in order to maintain solution stability. We place a sinusoidal source in the middle of the domain with a period of 1 year to simulate groundwater pumping and artificial recharge and set the head boundary conditions at the edges of the domain to be 0. We run the simulation for each scenario for several years and compute the amplitude and phase of the head response for each point in the domain.

The simulation results show that discontinuities in the amplitude and phase of the hydraulic head are most likely due to the presence of a fault rather than any rapid change in aquifer thickness (Figure 4.18). Both the fault and rapid decrease in aquifer thickness cause an amplification of the amplitude of the head on the side closer to the source. However, the fault causes a step decrease in the amplitude (as observed in the InSAR results) while the thickness change maintains a continuous amplitude profile. Furthermore, only the fault causes a step increase in the phase. In general, in our experiments, we found that changes in aquifer thickness had only a small effect on the diffusion results.

Therefore, we believe that the discontinuity observed in the InSAR seasonal ampli-

tude and phase is caused by the presence of an unmapped fault. The nearest fault system is the Peralta Hills Fault, a reverse fault system located 10 km north of Santa Ana, California and about 5 km away from the observed discontinuity [29, 30]. The main fault is north-dipping with the last rupture possibly occurring in the Holocene, and its surface trace follows the curved boundary between the Santa Ana Coastal Basin and the adjacent Santa Ana Mountains. However, the extent of the fault system from the Peralta Hills into the alluvial plains is uncertain and not well mapped at depth. The discontinuity that we observe in our InSAR results could thus correspond to a blind branch of the Peralta Hills Fault with no observable surface trace. The discontinuity is also roughly coincident with the Anaheim Nose, an anticline with a north-west strike parallel to the Peralta Hills Fault [31]. However, the anticline is a broad feature with a crest between 5000 - 10000 feet below the ground surface. Therefore, it is unlikely to affect groundwater flow in any observable manner.

#### **4.7 Conclusions**

We investigated the spatiotemporal ground deformation of the Los Angeles Central and Santa Ana Coastal Basins using a combination of GPS and C-band InSAR time series. We introduced a new method for decomposing both time series into long- and short-term signals in a spatially consistent manner using a distributed algorithm designed to solve large scale, regularized least squares problems very efficiently. This particular data set provided a unique opportunity to apply the time series analysis method on data with both time-varying seasonal signals and subtle transients caused by unknown geophysical processes. The resultant time series decomposition allowed us to isolate the short-term ground deformation caused by annual variations in hydraulic head due to groundwater production practices that emphasized pumping during the summer months. The 18-year timespan of the InSAR data and the improved time series analysis resulted in spatially continuous maps of seasonal amplitude and phase of ground deformation with improved spatial resolution over previous studies. In particular, we were able to detect fine-scale features in the seasonal deformation including spatially varying pore pressure diffusion effects and groundwater flow obstruction due to a potential unmapped fault. The reconstructed long-term signals revealed subsidence over a wide area within the basins, as well as a 2-3 year uplift signal starting in 2004-2005 caused by a period of heavy rainfall. By comparing the time series results with hydraulic head data provided by WRD and OCWD, we were able to determine that ground deformation within the basins can be almost completely explained as an elastic response to head variations caused

by groundwater production practices. Furthermore, we found that head variations in different parts of the aquifer system were responsible for the different timescales of ground deformation. Groundwater production is heavily monitored by agencies such as WRD and OCWD through a vast collection of monitoring well data. However, our results, in combination with previous InSAR studies over groundwater basins, show that analysis of InSAR time series data can be a useful tool for assessing the sustainability of pumping practices, and the continuing availability of data from new InSAR missions can be exploited by the hydrology community to aid groundwater monitoring over this area.

#### 4.A Appendix A: Spatiotemporal InSAR Time Series Analysis

Discussion about the need for extracting the spatiotemporal variability of deformation, and why we can't use methods like PCA/ICA for complex deformation processes.

In Riel *et al.* [26], a new method for detecting transient signals in geodetic time series was proposed that modeled time series as a linear combination of displacement functions chosen from a *dictionary* (matrix) of functions that resemble secular, seasonal, and transient signals. The displacement functions for transient signals were third-order time-integrated B-splines ( $B^i$ -splines) which exhibit one-sided behavior of a particular timescale. For reconstruction of transient signals of unknown onset times and durations, a highly overcomplete, non-orthogonal collection of  $B^i$ -splines of various timescales are used to populate the dictionary. Then, transient detection becomes a least squares minimization of a joint cost function that consists of a data misfit term and a sparsity-inducing regularization term to limit the total number of dictionary elements needed to reconstruct the original time series [26, 32, 33]. For a geodetic time series  $\mathbf{d} \in \mathbb{R}^{M \times 1}$  and a dictionary  $\mathbf{G} \in \mathbb{R}^{M \times P}$  consisting of  $P$  elements, the resulting cost function would be:

$$\varphi(\mathbf{m}) = \|\mathbf{G}\mathbf{m} - \mathbf{d}\|_2^2 + \lambda\|\mathbf{m}\|_1, \quad (4.20)$$

where  $\mathbf{m}$  is the vector of coefficients for each element in  $\mathbf{G}$  and the  $\ell_1$ -norm regularization term (scaled by the parameter  $\lambda$ ) minimizes the number of non-zero values in  $\mathbf{m}$ . Enforcing a sparse set of dictionary elements is critical for proper decomposition of any time series into its relevant timescales and enhancing interpretability of the non-zero elements in  $\mathbf{m}$ .

Riel *et al.* [26] also introduced a spatial weighting scheme for analyzing a network of GPS stations through an iterative re-weighting approach that ensures that the

estimated  $\mathbf{m}_i$  for a station  $i$  is consistent with  $\mathbf{m}_j$  from its neighboring stations. The “consistency” is controlled by an exponential distance weighting function between station  $i$  and stations  $j = 1, \dots, N, j \neq i$  in a network of  $N$  GPS stations. For InSAR time series analysis, an identical procedure can be used where each pixel in an interferogram is a “station”, and spatial weighting functions can be computed for each pixel and the other pixels in the interferogram. However, for an interferogram with  $N$  pixels, this approach would require the evaluation of  $N^2$  weighting functions, which would be very computationally intensive. An alternative approach would be to formulate a *simultaneous* estimation problem where we estimate  $\mathbf{m}_i, i = 1, \dots, N$  all at once in a large least squares problem. In this approach, the global linear model for  $K$  interferograms,  $\{\mathbf{G}\}\{\mathbf{m}\} = \{\mathbf{d}\}$ , would be constructed as:

$$\begin{bmatrix} \mathbf{G}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_1 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{G}_1 \\ \hline \vdots & \vdots & \vdots & \vdots \\ \hline \mathbf{G}_K & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_K & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{G}_K \end{bmatrix} \begin{bmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \\ \vdots \\ \mathbf{m}_N \end{bmatrix} = \begin{bmatrix} d_1^1 \\ d_1^2 \\ \vdots \\ d_1^N \\ \hline \vdots \\ \hline d_K^1 \\ d_K^2 \\ \vdots \\ d_K^N \end{bmatrix}, \quad (4.21)$$

where  $\mathbf{G}_k$  is the  $k$ -th row of the temporal dictionary corresponding to interferogram  $k$ ,  $\mathbf{m}_i$  is the coefficient vector corresponding to the  $i$ -th pixel, and  $d_k^i$  is the  $i$ -th pixel of interferogram  $k$ . The problem size becomes  $\{\mathbf{G}\} \in \mathbb{R}^{(N \cdot K) \times (N \cdot P)}$ ,  $\{\mathbf{m}\} \in \mathbb{R}^{(N \cdot P \times 1)}$ , and  $\{\mathbf{d}\} \in \mathbb{R}^{(N \cdot K \times 1)}$ , where  $\{\mathbf{d}\}$  is a “flattened” 1-D vector containing the entire InSAR stack.

### Large scale time series analysis using the alternating direction method of multipliers

Applying the cost function in Equation 4.20 to the linear model in Equation 4.21 would require solving a very large convex optimization problem for a length  $(N \cdot P)$  solution vector  $\{\mathbf{m}\}$ . For most modern InSAR time series, this problem cannot successfully be solved on a single computer. However, very powerful distributed optimization algorithms have been developed in recent years that take advantage of rapidly improving cluster computing frameworks to minimize the computational requirement of any one compute node. One such algorithm is the alternating

direction method of multipliers (ADMM) described by Boyd *et al.* [34]. The most general form of the problem solved by ADMM is

$$\begin{aligned} & \text{minimize } f(\mathbf{m}) + g(\mathbf{z}) \\ & \text{subject to } \mathbf{A}\mathbf{m} + \mathbf{B}\mathbf{z} = \mathbf{c} \end{aligned} \quad (4.22)$$

with optimization variables  $\mathbf{m}$  and  $\mathbf{z}$  and parameter arrays  $\mathbf{A}$ ,  $\mathbf{B}$ , and  $\mathbf{c}$ . Here, the primary variable we want to solve for,  $\mathbf{m}$ , has been split into two parts,  $\mathbf{m}$  and  $\mathbf{z}$ , such that the convex functions  $f$  and  $g$  are *separable* across the splitting [34]. This separability is crucial for exploiting large scale parallel computing frameworks. The constraint function is then used to enforce consistency between  $\mathbf{m}$  and  $\mathbf{z}$ . We form the augmented Lagrangian

$$L_\rho(\mathbf{m}, \mathbf{z}, \mathbf{y}) = f(\mathbf{m}) + g(\mathbf{z}) + \mathbf{y}^T(\mathbf{A}\mathbf{m} + \mathbf{B}\mathbf{z} - \mathbf{c}) + (\rho/2)\|\mathbf{A}\mathbf{m} + \mathbf{B}\mathbf{z} - \mathbf{c}\|_2^2, \quad (4.23)$$

where  $\mathbf{y}$  is a *dual* variable (from dual ascent optimization methods) and  $\rho > 0$  is a penalty parameter. ADMM performs the following iterations:

$$\mathbf{m}^{k+1} = \underset{\mathbf{m}}{\operatorname{argmin}} L_\rho(\mathbf{m}, \mathbf{z}^k, \mathbf{y}^k) \quad (4.24)$$

$$\mathbf{z}^{k+1} = \underset{\mathbf{z}}{\operatorname{argmin}} L_\rho(\mathbf{m}^{k+1}, \mathbf{z}, \mathbf{y}^k) \quad (4.25)$$

$$\mathbf{y}^{k+1} = \mathbf{y}^k + \rho(\mathbf{A}\mathbf{m}^{k+1} + \mathbf{B}\mathbf{z}^{k+1} - \mathbf{c}). \quad (4.26)$$

For the sparsity-inducing regularization problem in Equation 4.20,  $f$  is the misfit cost function and  $g$  is the  $\ell_1$ -norm regularization term. In this case, we would solve Equation 4.22 with  $\mathbf{A}$  and  $\mathbf{B}$  set to the identity and negative identity matrices, respectively, and  $\mathbf{c}$  set to 0.

The original ADMM formulation takes advantage of the separability of the convex functions  $f$  and  $g$  by splitting the optimization problem in one of two ways: 1) splitting across the data, or 2) splitting across the parameters [34]. The former is useful for a modest number of parameters and a very large number of training data while the latter is useful for the reverse situation. In Section 4.A, we will show how we can combine both splitting approaches in a consistent way to efficiently analyze InSAR time series. For the current discussion, we will proceed with the second approach by splitting across the parameters for our linear model  $\{\mathbf{G}\}\{\mathbf{m}\} = \{\mathbf{d}\}$ . For this approach, we partition the variable vector  $\{\mathbf{m}\}$  as  $\{\mathbf{m}\} = (\mathbf{m}_1, \dots, \mathbf{m}_N)$  where the individual  $\mathbf{m}_i$ 's do not need to be the same sizes. Correspondingly, we partition

$\{\mathbf{G}\} = (\mathbf{G}_1, \dots, \mathbf{G}_N)$  such that  $\{\mathbf{G}\}\{\mathbf{m}\} = \sum_{i=1}^N \mathbf{G}_i \mathbf{m}_i$ . This partitioning allows us to formulate the separable cost function

$$\text{minimize } \left\| \sum_{i=1}^N \mathbf{G}_i \mathbf{m}_i - \{\mathbf{d}\} \right\|_2^2 + \lambda \sum_{i=1}^N \|\mathbf{m}_i\|_1.$$

The ADMM problem in Equation 4.22 can then be re-formulated as:

$$\begin{aligned} & \text{minimize } \|\mathbf{z}_i - \{\mathbf{d}\}\|_2^2 + \lambda \sum_{i=1}^N \|\mathbf{m}_i\|_1 \\ & \text{subject to } \mathbf{G}_i \mathbf{m}_i - \mathbf{z}_i = \mathbf{0}, \quad i = 1, \dots, N, \end{aligned} \quad (4.27)$$

where the variables  $\mathbf{z}_i$  are introduced with the same length as the global data vector  $\{\mathbf{d}\}$ . Each update of Equations 4.24 - 4.26 can be done in parallel where a global reduction step (i.e., an *Allreduce* summing operation) is performed prior to the  $\mathbf{z}^{k+1}$  update to combine the individual  $\mathbf{G}_i \mathbf{m}_i$  predictions. Each individual worker would also need a copy of the InSAR stack contained in the global  $\{\mathbf{d}\}$  vector.

Up to this point, the formulation of the ADMM problem has only considered the  $\ell_1$ -norm sparsity-inducing regularization function. However, we still need to enforce spatial coherency of the  $\mathbf{m}$  vectors for each pixel as is done for GPS networks. One straightforward way to achieve this is to augment the regularization function  $g$  with an additional function that penalizes the spatial structure of the estimated  $\{\mathbf{m}\}$  vector through a covariance matrix,  $\mathbf{S}$ , such that:

$$g_i(\mathbf{m}_i) = \lambda \|\mathbf{m}_i\|_1 + \beta \|\mathbf{S}_i \mathbf{m}_i\|_2^2,$$

where  $\beta$  is an additional penalty parameter controlling the strength of the spatial penalty. This joint  $\ell_1$ - and  $\ell_2$ -norm regularization function is called the *elastic net* problem and has proven to be very useful for a large number of machine learning and statistics problems [35, 36]. The original elastic net formulation where  $\mathbf{S} = \mathbf{I}$  is used for selecting sparse *groups* of predictors whereas the  $\ell_1$ -norm-only regularization tends to select one predictor per group. This group selection is very useful for our case since we want to choose a sparse set of  $B^i$ -splines that are spatially grouped together. Appropriate construction of  $\mathbf{S}$  can then enforce spatial coherency beyond the original elastic net.

In practice, performing ADMM for InSAR time series analysis methods is tractable due to the sparsity structure of the  $\{\mathbf{G}\}$  array where the vast majority of its elements are zero. We can use powerful sparse linear algebra libraries to efficiently perform

the parallel ADMM iterations. To maintain tractability, we also wish to construct the spatial covariance matrix  $\mathbf{S}$  with a sparse number of non-zero elements. If we were to apply the spatial weighting function discussed for GPS networks, we would have a dense  $(N \cdot P) \times (N \cdot P)$  covariance matrix. Instead, since interferograms are continuous images of ground deformation, we can use a discrete Laplacian smoothing operator to construct  $\mathbf{S}$  such that the spatial distribution of a  $B^i$ -spline of a particular onset time and duration is smooth. Thus,  $\mathbf{S}$  is very sparse, and we can enhance the inherent grouping effect of the elastic net with a simple smoothing operation.

### Recursive time series analysis with ADMM

In the previous section, we discussed how we can split the global optimization problem across parameters in order to use parallel processing computing frameworks. However, each worker still needs to have a copy of the entire InSAR stack in  $\{\mathbf{d}\}$ . When the number of interferograms becomes large, or when each interferogram has a large number of pixels, this requirement can be computationally taxing. We can mitigate this effect by processing the InSAR stack in a *sequential* or *recursive* manner, i.e. one interferogram at a time. This sequential approach has been discussed in applications that aim to estimate sparse signals in a real-time manner as data are acquired [e.g. 37, 38]. In addition to real-time estimation capabilities, these applications also require low memory usage in deployment situations where onboard memory is limited. For our purposes, performing time series analysis on an interferogram-by-interferogram basis using the ADMM method while splitting across parameters greatly reduces the computational burden on any one compute node.

Following the approach outlined by Angelosante *et al.* [37], we can re-write the misfit cost function,  $(\|\{\mathbf{G}\}\{\mathbf{m}\} - \{\mathbf{d}\}\|_2^2)$ , as  $(\{\mathbf{m}\}^T \mathbf{R}_K \{\mathbf{m}\} - 2\{\mathbf{m}\}^T \mathbf{r}_K)$  where

$$\mathbf{R}_K = \sum_k^K \{\mathbf{G}\}_k^T \{\mathbf{G}\}_k, \quad \mathbf{r}_k = \sum_k^K \{\mathbf{G}\}_k^T \{\mathbf{d}\}_k, \quad (4.28)$$

and the  $k$  subscript indicates the subset of the  $\{\mathbf{G}\}$  and  $\{\mathbf{d}\}$  arrays corresponding to interferogram  $k$ , and  $\mathbf{R}_K \in \mathbb{R}^{(N \cdot P) \times (N \cdot P)}$  and  $\mathbf{r}_K \in \mathbb{R}^{(N \cdot P) \times 1}$ . We can then sequentially update  $\mathbf{R}_K$  and  $\mathbf{r}_K$  for a new interferogram by using

$$\mathbf{R}_{K+1} = \mathbf{R}_K + \{\mathbf{G}\}_{K+1}^T \{\mathbf{G}\}_{K+1}, \quad \mathbf{r}_{K+1} = \mathbf{r}_K + \{\mathbf{G}\}_{K+1}^T \{\mathbf{d}\}_{K+1}. \quad (4.29)$$

For a given InSAR time series of  $K$  interferograms, we can recursively build  $\mathbf{R}_K$  and  $\mathbf{r}_K$  without needing to load the entire time series into memory and distribute copies of the time series to each worker. Now, the ADMM minimization problem for the elastic net regularization with partitioning across parameters becomes:

$$\begin{aligned} & \text{minimize } \left[ \{\mathbf{z}\}^T \mathbf{R}_K \{\mathbf{z}\} - 2\{\mathbf{z}\}^T \mathbf{r}_K \right] + \sum_{i=1}^N \left[ \lambda \|\mathbf{m}_i\|_1 + \beta \|\mathbf{S}_i \mathbf{m}_i\|_2^2 \right] \\ & \text{subject to } \{\mathbf{m}\} - \{\mathbf{z}\} = \mathbf{0}. \end{aligned} \quad (4.30)$$

In the above equation, note that  $\{\mathbf{z}\}$  and  $\{\mathbf{m}\}$  are both vectors of length  $(N \cdot P)$  and are variables in the global optimization problem. By maintaining linear separability in the regularization function, the  $\{\mathbf{m}\}$ -update using Equation 4.24 can still be performed in parallel. While the update of  $\{\mathbf{z}\}$  using Equation 4.25 is in the global domain,  $\mathbf{R}_K$  is still very sparse and we can make use of parallel sparse linear algebra packages to efficiently perform the update. Here, we use the Portable, Extensible Toolkit for Scientific Computation (PETSc) [39] in conjunction with the Multifrontal Massively Parallel Sparse direct Solver (MUMPS) [40].

### Selection of penalty parameters

Equation 4.30 requires selection of two penalty parameters for the  $\ell_1$ - and  $\ell_2$ -norms,  $\lambda$  and  $\beta$ , respectively. One possible approach would be  $K$ -fold cross validation, which would involve randomly partitioning the data vector into  $K$  subsets. Then, for a given  $\lambda$  and  $\beta$ , we would solve for  $\{\mathbf{m}\}$  using  $K - 1$  subsets for training and use the last subset to test the out of sample performance of the trained model. We would then iterate over the  $K$  subsets such that each subset is used as the testing set. The combination of  $\lambda$ ,  $\beta$  that gave the lowest testing area over the  $K$  folds would then be the optimal penalty parameters. However, this two-dimensional exploration of penalty parameters could be quite costly for a large number of data and parameters. Alternatively, we can reduce the computational cost by using an *independent* source of data as a testing set and not have to perform a two-dimensional grid search  $K$  times. In this work, we can use data from GPS stations within the coastal basins and outlying areas as an independent test of predictive power of our InSAR time series model. Thus, we perform a single two-dimensional grid search for  $\lambda$  and  $\beta$  and choose the combination that results in a time series model that best agrees with the GPS data.

### Discussion on Recursive ADMM for InSAR Time Series Analysis

A high fraction of InSAR time series methods and their applications have relied on a pixel-by-pixel approach where the time history of each pixel is solved independently of its neighboring pixels. This approach is simple, easily parallelizable, and has proven to be useful for a large number of ground deformation studies. However, these pixel-by-pixel approaches implicitly ignore spatially correlated data errors, such as those from atmospheric effects which can be coherent over length scales of tens to hundreds of kilometers [41, 42]. Omission of these errors from the time series analysis can lead to significant changes in the time series model parameters and, perhaps more importantly, their uncertainties. The formulation of the ADMM minimization problem allows us to directly incorporate data covariance matrices for these spatially correlated data errors. The recursive update in Equation 4.29 would be modified by inserting a covariance matrix between the  $\{\mathbf{G}\}_k$  and  $\{\mathbf{d}\}_k$  multiplications. Accounting for these errors, in conjunction with enforcing spatial coherency in the elastic net regularization scheme, greatly improves the robustness of the time series model to data artifacts such as unwrapping errors and partial scene coverage due to incomplete SAR frames.

Recursive processing of the interferograms is highly suitable for real-time analysis of InSAR data. The shorter repeat times of modern SAR platforms has lead to an unprecedented volume of data being collected to monitor ground motion with high temporal resolution. However, most modern software packages designed for InSAR time series analysis reprocess the entire time series each time a new interferogram is acquired. This strategy is not sustainable in terms of computational time and resources. By maintaining a “state” representative of the current time series model (encapsulated within the  $\{\mathbf{m}\}$  vector), and by keeping the state invariant to any specific observation epoch, we can efficiently update our time series model by recursively updating the ADMM arrays ( $\mathbf{R}_K$  and  $\mathbf{r}_K$ ) and rapidly estimating a new state,  $\{\mathbf{m}\}$ . For non-real-time applications, the recursiveness allows for simple parallelization of interferogram assimilation since each worker can compute their own  $\mathbf{R}_K$  and  $\mathbf{r}_K$ , and a global reduction operation can combine the arrays before estimation of the time series model.

While the combination of partitioning across the global parameter vector and recursive updating of the ADMM arrays ( $\mathbf{R}_k$  and  $\mathbf{r}_K$ ) greatly increases the computational efficiency of ADMM for InSAR time series analysis, we are still limited by the requirement that every pixel in an interferogram requires its own set of temporal

dictionary elements. Analysis of a high spatial resolution InSAR stack with a hundred or more dictionary functions leads to a very large number of parameters and memory requirements of several tens to over a hundred gigabytes. However, optimized computing frameworks that specialize in distributed storage and distributed processing, such as Apache Hadoop or Apache Spark, can easily handle such memory requirements by intelligently loading and caching problem data. Thus, future work involves migrating our ADMM implementation to an appropriate distributed processing framework.

## 4.B Appendix B: Supplemental Figures

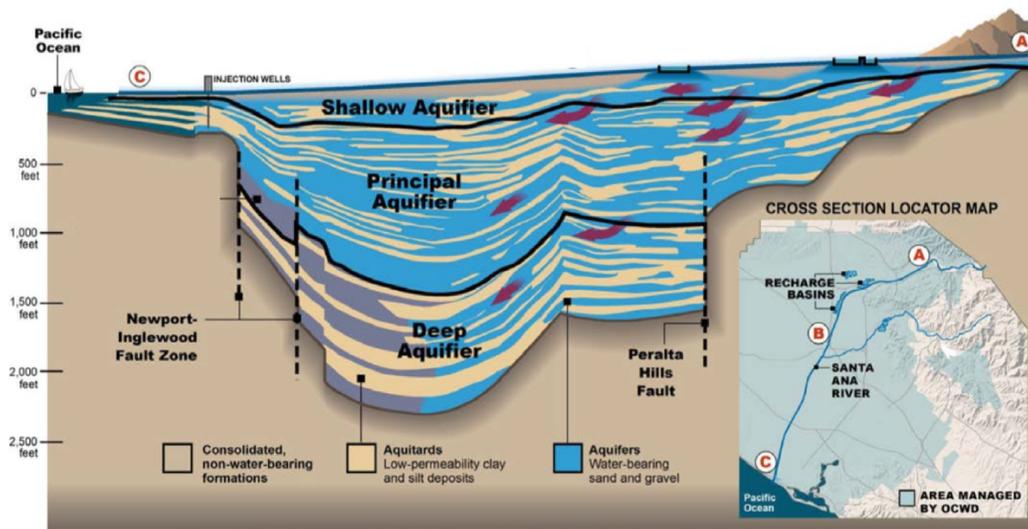


Figure 4.B1: Cross-section of the Santa Ana Coastal Basin aquifers and partitioning into three-layer model (modified from Woodside & Westropp [13]). The distribution of aquifers and interleaving aquitards can be aggregated into a simplified three-layer model where the aquitards separating the three layers are laterally extensive.

## References

1. Todd, D. K., Mays, L. W., *et al.* *Groundwater hydrology* (Wiley New York, 1980).
2. Galloway, D. L. & Burbey, T. J. Review: regional land subsidence accompanying groundwater extraction. *Hydrogeology Journal* **19**, 1459–1486 (2011).
3. Wilson, A. & Gorelick, S. The Effects of Pulsed Pumping on Land Subsidence in the Santa Clara Valley, California. *Journal of Hydrology* **174**, 375–396 (1996).
4. Bawden, G. W., Thatcher, W., Stein, R. S., Hudnut, K. W. & Peltzer, G. Tectonic contraction across Los Angeles after removal of groundwater pumping effects. *Nature* **412**, 812–815 (2001).
5. Amelung, F., Galloway, D. L., Bell, J. W., Zebker, H. A. & Lacznik, R. J. Sensing the ups and downs of Las Vegas: InSAR reveals structural control of land subsidence and aquifer-system deformation. *Geology* **27**, 483–486 (1999).
6. Chaussard, E., Bürgmann, R., Shirzaei, M., Fielding, E. & Baker, B. Predictability of Hydraulic Head changes and Characterization of Aquifer-System and Fault Properties from InSAR-derived Ground Deformation. *Journal of Geophysical Research: Solid Earth* **119**, 6572–6590 (2014).

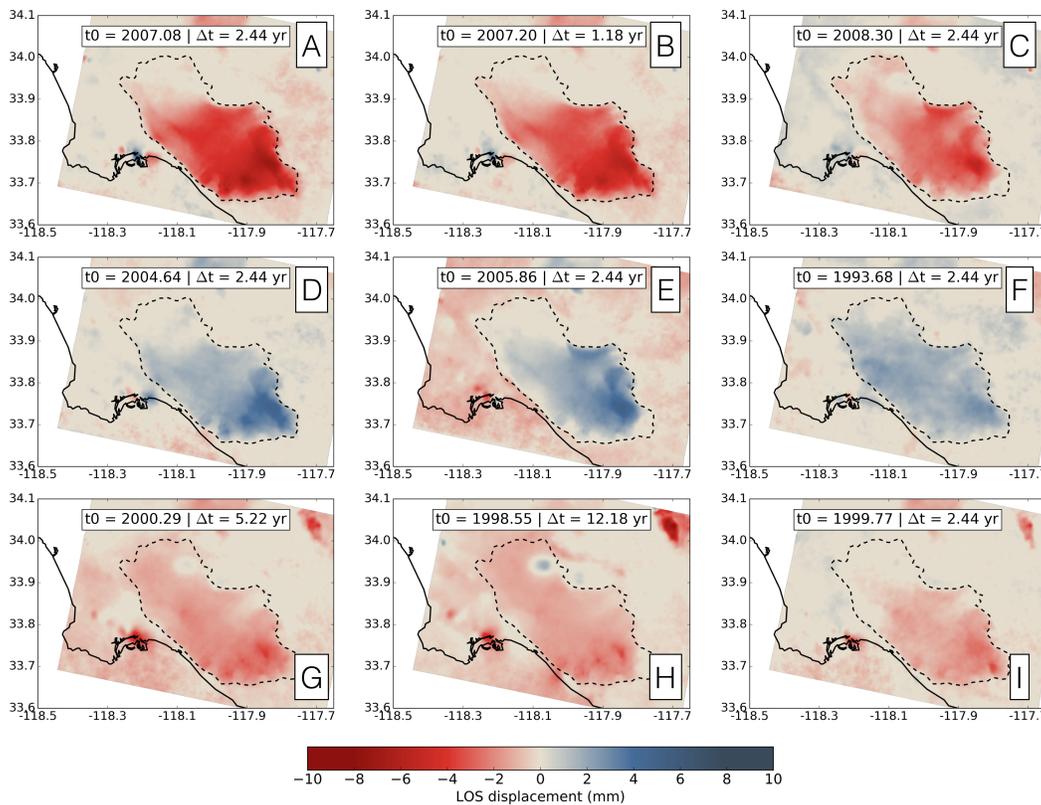


Figure 4.B2: Spatial distribution of largest  $B^i$ -spline coefficients determined from the recursive ADMM InSAR time series analysis. For each image, the color at each pixel corresponds to the amplitude of the specified  $B^i$ -spline. The largest estimated coefficient is associated with a rapid subsidence signal with a time scale of 2.4 years centered in the beginning of 2007 (A). We can also observe high values for the uplift event due to heavy rainfall in (D) and (E). In (H), we see that the uplift in the Santa Fe Spring oil field and subsidence in the Pomona basin are well described by a long-term  $B^i$ -spline centered on 1999.

7. Miller, M. M. & Shirzaei, M. Spatiotemporal characterization of land subsidence and uplift in Phoenix using InSAR time series and wavelet transforms. *Journal of Geophysical Research: Solid Earth* **120**, 5822–5842 (2015).
8. Watson, K. M., Bock, Y. & Sandwell, D. T. Satellite interferometric observations of displacements associated with seasonal groundwater in the Los Angeles basin. *Journal of Geophysical Research: Solid Earth* **107** (2002).
9. Lanari, R., Lundgren, P., Manzo, M. & Casu, F. Satellite radar interferometry time series analysis of surface deformation for Los Angeles, California. *Geophysical Research Letters* **31** (2004).
10. Zhang, L., Lu, Z., Ding, X., Jung, H.-s., Feng, G. & Lee, C.-W. Mapping ground surface deformation using temporarily coherent point SAR interferometry:

- Application to Los Angeles Basin. *Remote Sensing of Environment* **117**, 429–439 (2012).
11. Berardino, P., Fornaro, G., Lanari, R. & Sansosti, E. A new algorithm for surface deformation monitoring based on small baseline differential SAR interferograms. *Geoscience and Remote Sensing, IEEE Transactions on* **40**, 2375–2383 (2002).
  12. Thiros, S., Bexfield, L., Anning, D. & Huntington, J. Conceptual understanding and groundwater quality of selected basin-fill aquifers in the Southwestern United States: U.S. Geological Professional Paper 1781 (2010).
  13. Woodside, G. & Westropp, M. *Orange Country Water District Groundwater Management Plan 2015 Update* tech. rep. (Orange County Water District, 2015).
  14. Terzaghi, K. v. Die berechnung der durchlassigkeitsziffer des tones aus dem verlauf der hydrodynamischen spannungserscheinungen. *Sitzungsberichte der Akademie der Wissenschaften in Wien, Mathematisch-Naturwissenschaftliche Klasse, Abteilung IIa* **132**, 125–138 (1923).
  15. Calderhead, A., Therrien, R., Rivera, A., Martel, R. & Garfias, J. Simulating pumping-induced regional land subsidence with the use of InSAR and field data in the Toluca Valley, Mexico. *Advances in Water Resources* **34**, 83–97 (2011).
  16. Chen, J., Knight, R., Zebker, H. A. & Schreüder, W. A. Confined aquifer head measurements and storage properties in the San Luis Valley, Colorado, from spaceborne InSAR observations. *Water Resources Research* (2016).
  17. Jacob, C. *et al.* Engineering Hydraulics. *Flow of Ground Water*, 321–386 (1950).
  18. Keranen, K. M., Weingarten, M., Abers, G. A., Bekins, B. A. & Ge, S. Sharp Increase in Central Oklahoma Seismicity Since 2008 Induced by Massive Wastewater Injection. *Science* **345**, 448–451 (2014).
  19. Hantush, M. S. Flow of ground water in sands of nonuniform thickness: Part 2. Approximate theory. *Journal of Geophysical Research* **67**, 711–720. ISSN: 2156-2202 (1962).
  20. Fetter, C. W. *Applied Hydrogeology* (Prentice Hall, 2000).
  21. Guenther, R. B. & Lee, J. W. *Partial Differential Equations of Mathematical Physics and Integral Equations* (Courier Corporation, 1996).
  22. Alley, W. M., Healy, R. W., LaBaugh, J. W. & Reilly, T. E. Flow and Storage in Groundwater Systems. *Science* **296**, 1985–1990 (2002).
  23. Hetland, E., Musé, P., Simons, M., Lin, Y., Agram, P. & DiCaprio, C. Multi-scale InSAR Time Series (MInTS) Analysis of Surface Deformation. *Journal of Geophysical Research* **117** (2012).

24. Jolivet, R. *et al.* Improving InSAR geodesy using global atmospheric models. *Journal of Geophysical Research: Solid Earth* **119**, 2324–2341 (2014).
25. Middleton, G. V. & Wilcock, P. R. *Mechanics in the Earth and Environmental Sciences* (Cambridge University Press, 1994).
26. Riel, B., Simons, M., Agram, P. & Zhan, Z. Detecting transient signals in geodetic time series using sparse estimation techniques. *Journal of Geophysical Research: Solid Earth* **119**, 5140–5160 (2014).
27. King, N. *et al.* Space geodetic observation of expansion of the San Gabriel Valley, California, aquifer system, during heavy rainfall in winter 2004–2005. *Journal of Geophysical Research: Solid Earth (1978–2012)* **112** (2007).
28. Hauksson, E., Göbel, T., Ampuero, J.-P. & Cochran, E. A Century of Oil-field Operations and Earthquakes in the greater Los Angeles Basin, Southern California. *The Leading Edge* **34**, 650–656 (2015).
29. Center, S. C. E. D. *Historical Earthquakes and Significant Faults in Southern CA* <<http://scedc.caltech.edu/significant/index.html>> (2012).
30. Myers, D. J., Nabelek, J. L. & Yeats, R. S. Dislocation modeling of blind thrusts in the eastern Los Angeles basin, California. *Journal of Geophysical Research: Solid Earth* **108** (2003).
31. Bjorklund, T. The Whittier Fault Trend: Cross Sections, Structure Maps, and Well Tops in the Major Oil Producing Area of the Northeastern Los Angeles Basin. *Amer. Assoc. Petrol Geol., Search and Discovery, Article* **10038** (2003).
32. Tibshirani, R. Regression Shrinkage and Selection Via the Lasso. *J. R. Stat. Soc. B* **58**, 267–288 (1996).
33. Chen, S. S., Donoho, D. L. & Saunders, M. A. Atomic Decomposition by Basis Pursuit. *SIAM Journal on Scientific Computing* **20**, 33–61 (1998).
34. Boyd, S., Parikh, N., Chu, E., Peleato, B. & Eckstein, J. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning* **3**, 1–122 (2011).
35. Zou, H. & Hastie, T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67**, 301–320 (2005).
36. Jenatton, R., Audibert, J.-Y. & Bach, F. Structured variable selection with sparsity-inducing norms. *The Journal of Machine Learning Research* **12**, 2777–2824 (2011).
37. Angelosante, D., Bazerque, J. A. & Giannakis, G. B. Online adaptive estimation of sparse signals: Where RLS meets the  $\ell_1$ -norm. *Signal Processing, IEEE Transactions on* **58**, 3436–3447 (2010).

38. Kopsinis, Y., Slavakis, K. & Theodoridis, S. Online sparse system identification and signal reconstruction using projections onto weighted balls. *Signal Processing, IEEE Transactions on* **59**, 936–952 (2011).
39. Balay, S., Gropp, W. D., McInnes, L. C. & Smith, B. F. *Efficient Management of Parallelism in Object Oriented Numerical Software Libraries in Modern Software Tools in Scientific Computing* (eds Arge, E., Bruaset, A. M. & Langtangen, H. P.) (Birkhäuser Press, 1997), 163–202.
40. Amestoy, P. R., Duff, I. S., L'Excellent, J.-Y. & Koster, J. A fully asynchronous multifrontal solver using distributed dynamic scheduling. *SIAM Journal on Matrix Analysis and Applications* **23**, 15–41 (2001).
41. Emardson, T., Simons, M. & Webb, F. Neutral atmospheric delay in interferometric synthetic aperture radar applications: Statistical description and mitigation. *Journal of Geophysical Research: Solid Earth* **108** (2003).
42. Lohman, R. B. & Simons, M. Some thoughts on the use of InSAR data to constrain models of surface deformation: Noise structure and data downsampling. *Geochemistry, Geophysics, Geosystems* **6** (2005).

*Chapter 5*

## CONCLUSIONS

In this thesis, our initial goal was to develop an automated transient detection algorithm for GPS networks that would detect anomalous signals generated by any number of geophysical processes. In the early stages of development, several different methods had been proposed that were based on *statistical* analysis of geodetic data, such as covariance descriptor analysis, principal component analysis, Hidden Markov modeling, etc. These methods are non-parameteric in nature and are intended to be flexible in order to detect any type of transient signal. However, these methods often make limiting assumptions and can be sensitive to observation noise. Therefore, a key motivation was to develop a robust detection algorithm that could utilize our prior knowledge about most transient signals observed in geodetic data, which is that they are non-repeating, isolated events that often have a characteristic temporal evolution.

The temporal parameterization approach that we introduced in Chapter 2 made of use of an overcomplete dictionary of integrated B-splines ( $B^i$ -splines) of varying onset times and durations. We can view this parameterization as a restriction of the model space of detected transient signals to signals that resemble slow steps in time. Additionally, the spatial weighting that we introduced in Chapter 2 would be a further restriction of the model space by encouraging spatial coherency between transient signals. This restriction of the model space is actually a way to improve robustness in the detection performance and reduce the sensitivity to observation noise and station-specific signals. The overcompleteness of the dictionary compensates for the model space restriction by allowing for a wide range of onset times and durations. The final key ingredient to our approach was the sparsity-enforcing  $\ell_1$ -norm regularization that minimizes the total number of  $B^i$ -splines needed to reconstruct a transient signal, enhancing the interpretability of our results and improving the performance of the spatial weighting.

When we first attempted to apply this technique to InSAR time series, we quickly ran into computational difficulties due to the nature of the spatial weighting scheme and the number of iterations generally required to achieve convergence in the time series reconstruction. The simultaneous inversion of the time series of all the interfero-

gram pixels using the elastic net regularization approach partially overcame these difficulties by using a single prior covariance matrix to enforce spatial coherency of  $B^i$ -splines used to reconstruct transient signals. However, practical performance was limited by the serial convex optimization solver used to solve the elastic net problem and limited computer memory on a single compute node. The distributed computing capabilities of the alternating direction method of multipliers (ADMM) discussed in Chapter 4 was crucial for allowing the transient detection framework we developed to be feasible and fast by harnessing the power of modern high performance computing resources. Furthermore, the formulation of the recursive ADMM approach to InSAR time series analysis was important for handling InSAR time series with a large number of interferograms, such as the 881 ERS and Envisat interferograms used for the study of ground deformation in the Los Angeles basin. This formulation could prove to be very useful for current and future InSAR missions where the temporal sampling period can be as low as 8 days, resulting in very large InSAR time series that need to be analyzed in near real-time for natural hazard monitoring. On a similar note, this formulation can be used for large GPS networks where covariances between east, north, and up components can be enforced to perform three-dimensional daily hazard monitoring over wide regions. For example, monitoring ground deformation over the Cascadia subduction zone for detection of slow slip events would require  $\sim 450000$  model parameters (3000 temporal parameters for  $\sim 150$  stations), which is well within the computational capabilities of small cluster computing systems.

Throughout the progression of the development of this transient detection framework, we gradually realized that the problem of transient detection was nothing more than a subproblem of a robust, generic time series decomposition problem. Here, decomposition means separating a time series into temporal components corresponding to distinct physical phenomena, such as secular plate motion, seasonal deformation due to earth tides, semi-periodic motion due to hydrologic processes, transient motion due to a slow slip event, long-term transient displacements due to magma chamber inflation or groundwater withdrawal, etc. Therefore, an implicit result of successfully detecting and reconstructing transient signals is a more accurate quantification of the *other* signals recorded in the geodetic data set. In areas overlying active plate boundaries, proper detection of transient signals results in a better estimate for the secular rate. In areas with time-varying seasonal signals, such as in the coastal basins near Los Angeles, proper characterization of long-term transient subsidence and uplift allowed us to accurately isolate the motion due to

the annual cycle of groundwater pumping and recharge.

In the time series analysis framework that we presented in this thesis, much of the transient detection performance depends on proper construction of the temporal dictionary and proper selection of regularization parameters. For the former, the dictionary should include a set of functions sufficient to reconstruct all possible temporal signals expected for a given time series. Sparsity-inducing regularization is optimal when a given signal can be reconstructed with a sparse set of functions from the dictionary. For most applications, an *overcomplete* dictionary containing  $B^i$ -splines with long- and short-term timescales can adequately describe realistic transient signals. Short-term  $B^i$ -splines should have timescales equal to the most rapid transients expected in a time series. In situations where a transient is not well described by a combination of  $B^i$ -splines (e.g., a signal with curvatures of different signs), the reconstruction accuracy may be diminished, but the overall interpretability of the solution can still be exploited by examining the timescales and amplitudes of the non-sparse coefficients.

For selection of regularization parameters, in Chapter 2, we presented a Bayesian formulation of the optimization problem that treats the regularization parameter as a stochastic hyperparameter following a gamma distribution. Thus, the problem of time series decomposition and selection of hyperparameters can be performed simultaneously by using a probabilistic framework. In this way, one can incorporate more accurate error models for observation noise that could exhibit any arbitrary covariance structure. One important advantage to this approach is that one can obtain a full posterior distribution for the coefficients of the temporal dictionary, allowing for quantification of realistic uncertainties and assessment of the temporal model prediction power.

In Chapters 3 and 4, we demonstrated how a diverse set of geodetic data is necessary for investigating the properties of any geophysical process that causes measurable surface deformation. In the case of the Bárðarbunga caldera collapse, we found that the suite of InSAR data from the international constellation of satellites complemented with three component GPS data was necessary to constrain a consistent model of magma withdrawal, rifting, and subsidence of the ice overlying the caldera. While spatial diversity of data was obtained by combining GPS data with interferograms formed from multiple look directions, temporal diversity was obtained through interferograms with a wide range of repeat times. Specifically, the one-day repeat interferograms over the ice-covered caldera were invaluable for formulating

the model of aseismic deformation on ring faults due to a deflating magma chamber. The longer repeat time interferograms on the ice-free ground adjacent to the caldera were then used to estimate the magma chamber volume change and the amount of opening along the emplaced dike. If more interferograms were available, a more consistent way to utilize temporal diversity would be a time series approach similar to the study of the Los Angeles coastal basins in Chapter 4.

The model of aseismic deformation on ring faults driven by deflation of a deeper magma chamber was developed entirely with multi-temporal interferograms and moment tensor solutions that were publicly available during the eruption sequence and at the time of writing. A more recent study utilized a larger and more diverse set of geodetic and seismic data to analyze the interaction between magma chamber deflation and magma flow within the regional scale dike [1]. Specifically, geobarometry, aircraft-based altimetry, high-precision earthquake locations, and GPS observations directly on the ice-covered caldera were used to determine the geometry of the magma chamber, its depth, and the orientation of the ring faults associated with the piston overlying the chamber. The chamber depth was estimated to be between 8-12 km for a point pressure source, and subsidence of the ice overlying the caldera was controlled by slip on an outward-dipping ring fault on the north side of the caldera and an inward-dipping ring fault on the south side of the caldera. The large CLVD components for the caldera rim seismic events were hypothesized to be caused by failure of support structures within the magma reservoir. Therefore, the model of magma chamber deflation, crack collapse within the chamber, and aseismic slip on ring faults that we presented in Chapter 3 is largely supported by the updated data and analysis used in Gudmundsson *et al.* [1]. While the outward-dipping orientation of the northern ring fault differs from what we originally proposed, the main conclusion that aseismic slip dominates subsidence over the caldera is still valid.

For the study on ground deformation in groundwater basins in the Los Angeles area, we only used interferograms from one look direction, so spatial diversity was limited. However, ground subsidence and uplift is primarily in the vertical direction, so interferograms from one line-of-sight was sufficient for our purposes. The real strength of the data set was the large number interferograms available to form a densely sampled time series that we could decompose into long- and short-term components. Therefore, the temporal diversity of the InSAR data is incorporated into the reconstruction of the time series. We then demonstrated how the decomposition into the different temporal components was important for understanding

the groundwater dynamics of the aquifer system and the influence of groundwater pumping practices. The discovery that different parts of the aquifer system were responsible for different components of ground deformation will be important for aiding water districts in maintaining sustainable pumping practices. Future geodetic studies over this area should include data from recent InSAR satellites, which have a wider range of look directions, radar wavelengths, and repeat times, which should allow for quantification of the full three-dimensional motion of the ground within all groundwater basins.

Looking forwards, the techniques and applications presented in this thesis are a step forward towards a modern time series analysis framework that can ingest observations from any geodetic data source on a day-to-day basis. Throughout this work, we placed an emphasis on robustness and scalability of the transient detection procedure. These characteristics will be particularly important for ingesting data from the future international constellation of SAR satellites, which will provide systematic observations of all deforming regions on Earth with low latency between observation times. The increased temporal coverage and higher quality of observations will greatly enhance studies of surface deformation but will pose substantial challenges with regards to data processing and assimilation.

The methods presented here can be applied to a data set of any size if access to a high performance computing cluster is available. However, even if resources are limited to a single compute node, we could still make use of standalone distributed computing resources such as Apache Hadoop or Apache Spark which are designed to minimize the memory requirements of computing tasks by caching to disk any data not used for a given active operation. On the other end of the spectrum, the prevalence of graphical processing units (GPUs) in high performance computing clusters could be exploited to dramatically reduce computational costs and/or increase the temporal model complexity. For the latter case, that would imply that we could have a much larger temporal dictionary for a larger number of ground points, thus increasing both the temporal and spatial resolution of a time series model.

Finally, we believe that any practical time series analysis method used to monitor potentially hazardous anomalous signals should incorporate many different methods for time series decomposition, such as the previously mentioned statistics-based models. In addition to providing independent results to cross-validate the techniques presented in this thesis, these different methods could complement one another and be used to obtain a more thorough understanding of the spatiotemporal characteris-

tics of observed geophysical processes. As an example, consider a volcano monitoring system that utilizes observations from GPS and InSAR data. ADMM-based transient detection would be used to track the coefficients of non-zero  $B^i$ -splines in a temporal dictionary. In the event that a spatially coherent set of non-zero  $B^i$ -splines is detected due to movement of magma beneath the ground surface, a decomposition algorithm such as principal component analysis could be run on the same data to check if the spatial pattern of the  $B^i$ -splines is described by the spatial pattern of one or more of the largest principal components. Furthermore, *ensemble* methods can be used to combine the temporal signal of the largest  $B^i$ -splines with the temporal signal of the largest principal components to obtain a modeled transient signal with potentially improved prediction performance and greater robustness.

## References

1. Gudmundsson, M. T. *et al.* Gradual Caldera Collapse at Bárðarbunga Volcano, Iceland, Regulated by Lateral Magma Outflow. *Science* **353**, aaf8988 (2016).