

An Automated Image Analysis Framework for Model-based Feature Detection in Sparse Data

A Dissertation

Presented to

the Faculty of the School of Engineering and Applied Science

University of Virginia

In Partial Fulfillment

of the requirements for the Degree

Doctor of Philosophy

by

Andrea Vaccari

May 2014

Approval Sheet

This dissertation
is submitted in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

Andrea Vaccari

This dissertation has been read and approved by the Examining Committee:

Scott T. Acton, Adviser

Zongli Lin, Committee Chair

Jennie L. Moody

Robert M. Weikle, II

Stephen G. Wilson

Accepted for the School of Engineering and Applied Science:

James H. Aylor, Dean, School of Engineering and Applied Science

May 2014

Abstract

The development of novel remote sensing techniques, based on interferometric synthetic aperture RADAR (InSAR), currently allows for millimetric precision measurements of Earth surface deformation over time. One of the major challenges posed by these techniques, known as persistent scatterer interferometry (PSI), is the inherent sparsity of the data resulting from the RADAR scatterer selection process.

In this work we present an automated image analysis framework aimed at the detection of model-defined spatiotemporal features within sparse point cloud datasets, and we show how this framework can be tailored to the early detection of hazardous geophysical features within InSAR-derived data. In particular, we developed a spatiotemporal model describing the evolution of subsiding features, we verified its validity by using discrete element method simulations, and applied it, within our framework, to the early detection of sinkholes.

The ground truth dataset, used to develop the spatiotemporal model, was obtained by imaging a sinkhole prone area for a period of 70 months. The relevance of this dataset for our research is due to the fact that it contains four active features of which one (W1) collapsed before the data was taken, and one (W2) collapsed after the data was taken providing ground truth measurements.

We first approached the detection as a graph segmentation problem. By assigning each PSI scatterer within the ground truth dataset to a vertex and enforcing connectivity by the Delaunay triangulation, we obtained a graph that reflected the local neighborhood relationship. We then constructed an edge-weighting function designed to favor low weights for edges traversing boundaries of regions displaying signs of subsidence. The segmentation resulting from the application of the min-cut algorithm to this graph captured 27% and 94% of the collapsed area of W1 and W2 respectively.

Since we had at our disposal the time series of the displacements of each scatterer, we expanded our approach to leverage this information by developing a model-based spatiotemporal detection method. The parameters regulating the behavior of the model were used to generate a multidimensional parameter space that was then scanned with user-defined resolution. At each point, a spatiotemporal template was reconstructed based on the original model and the currently selected parameters. This template was used to analyze the point cloud dataset for regions with matching behavior. This method provided an improvement by identifying as high risk 52.6% and 81.6% of the collapsed area of W1 and W2 respectively against the values of 37.5% and 17.6% obtained from the graph cut approach. We also applied this method to a $40 \times 40 \text{ km}^2$ area of interest located in western Virginia. The ground validation on a subset of the detected features showed that 78% of the locations presented strong evidence of subsidence.

To improve on the computational burden imposed by the direct application of this exploration method with complex models over large datasets, we developed an activity detection approach where large datasets were subdivided into smaller blocks. The average and standard deviation of the displacements of the scatterers contained in each block were

used as elements of each block feature vector. Outliers in the feature space, corresponding to actively subsiding regions, were identified using their Mahalanobis distance. When applied to the ground truth dataset this screening method provided a $\times 15$ increase in the detection speed while still maintaining accurate results. To further reduce the impact of larger datasets and complex models, we introduced a second screening stage, based on the evaluation of the normalized mutual information between model and data, to pinpoint the location of features requiring full spatiotemporal analysis.

Finally, to leverage the inherent sparsity of the PSI data, we took advantage of the tools provided by the emerging field of graph signal processing and developed a graph-based scale space analysis approach that provided results comparable to those obtained by previous methods.

*All that is gold does not glitter,
Not all those who wander are lost;
The old that is strong does not wither,
Deep roots are not reached by the frost.*

J.R.R. Tolkien

Acknowledgments

I would like to thank my adviser, Dr. Acton for giving me the chance to pursue my interests and goals, trusting that I would be a good addition to the VIVA lab, and mentoring me during my journey. My lab mates at VIVA. I strongly believe that there is nothing more stimulating than sharing ideas and constructive criticism with a group of fellow researchers: lively discussion are at the heart of brilliant research. In particular, I would like to thank Michael "the Stüke" for the countless hours spent discussing options, brain storming, evaluating and testing approaches. Many of the ideas exposed in this work saw their first light during that time.

I am also grateful to my committee members, Dr. Lin, Dr. Moody, Dr. Weikle, and Dr. Wilson, for their time and support at every crucial step in my student career.

I would also like to thank my friends and colleagues who encouraged and supported my decision to go back to school, and my parents, for bringing me up an inquiring person and never preventing me from following my curiosity and desire for exploration.

Finally, I would like to dedicate this dissertation to my partner, Natalie, for her encouragement and support, and for being a constant source of inspiration.

Contents

Contents	viii
List of Tables	x
List of Figures	xi
List of Symbols	xviii
1 Introduction and technical background	1
1.1 Motivation	1
1.2 Monitoring Earth’s surface movements	6
1.3 RADAR for remote sensing applications	8
1.3.1 RADAR basics	8
1.3.2 Side-Looking Airborne/real-Aperture RADAR (SLAR)	17
1.3.3 Synthetic Aperture Radar (SAR)	21
1.3.4 Interferometric SAR (InSAR)	29
1.3.5 Differential InSAR (DInSAR)	33
1.3.6 Permanent Scatterer InSAR (PSInSAR) and SqueeSAR	35
1.4 Summary	42
2 Modeling	44
2.1 Introduction	44
2.2 Subsidence and sinkholes	45
2.2.1 Morphology	46
2.2.2 The Winkler sinkholes (Wink Sinks)	49
2.3 The model	52
2.4 Simulation	57
2.4.1 Model description	61
2.4.2 Results	64
2.5 Summary	68
3 Detection	70
3.1 Introduction	70
3.1.1 Graph theory primer	70
3.2 Graph cut approach	74
3.2.1 Results	79
3.3 Spatiotemporal matching	82
3.3.1 Introduction	82

3.3.2	Method	84
3.3.3	Results	88
3.3.4	Subsidence model discussion	103
3.4	Summary	107
4	Performance analysis	108
4.1	Introduction	108
4.2	Parameter space reduction	109
4.2.1	Results and comparison with spatiotemporal matching	113
4.3	Mutual information metric	116
4.3.1	Results and comparison with spatiotemporal matching	119
4.4	Summary	121
5	Graph signal detection	123
5.1	Introduction	123
5.1.1	Graph theory II	125
5.2	Detection and scale analysis	136
5.3	Results	142
5.3.1	Comparison with previous results	148
5.4	Summary	151
6	Conclusions and future work	153
6.1	Future work	154
A	Satellites and SAR instrumentation	157
A.1	European remote sensing satellites (ERS)	157
A.2	COSMO-SkyMed	160
A.3	Future of SAR	162
B	Project datasets	164
B.1	Precision of data	165
B.2	Wink Sinks	166
B.3	Virginia	169

List of Tables

3.1	Graph-cut sinkhole detection vs. collapsed ground truth (From [95] ©2012 IEEE)	81
3.2	Parameter space limits and increments for the artificial dataset	90
3.3	Parameter space limits and increments for the Wink Sink dataset	92
3.4	Graph-cut and spatiotemporal matching vs. collapsed ground truth	96
3.5	Parameter space limits and increments for the Virginia dataset	101
3.6	Categories of subsidence evidence based on ground inspection (From [78] ©2013 Taylor and Francis)	102
3.7	Results of ground validation on Virginia data (From [78] ©2013 Taylor and Francis)	103
A.1	ERS Orbital characteristics [32]	158
A.2	Characteristics of AMI in imaging mode [32]	159
A.3	COSMO-SkyMed constellation orbital characteristics [32]	160
A.4	Characteristics of the SAR-2000 in imaging mode [32]	162
B.1	Typical precision for PS located within 2 km from the reference point	165
B.2	Summary of the Wink Sinks dataset characteristics	168
B.3	ERS-1/2 LOS vector for Wink Sinks	168
B.4	Dates of the image acquisition for the Virginia dataset	170
B.5	COSMO-SkyMed LOS vector for Virginia	171
B.6	Summary of the Virginia dataset characteristics	171

List of Figures

1.1	Cumulative subsidence damage caused by sinkhole. Time periods on which estimates are based vary by State, and costs are not converted to constant dollars. Data not available for Puerto Rico, U.S. Virgin Islands, and Pacific Territories. (From [3] with the permission of FEMA)	2
1.2	(a) Landslide potential based on adverse formations associated with past landslide activity. (b) Landslide incidence and susceptibility in the conterminous United States. Data not available for Alaska, Hawaii, Puerto Rico, U.S. Virgin Islands, and Pacific Territories. (From [3] with the permission of FEMA) . .	3
1.3	RADAR resolutions. The highlighted area is know as the <i>resolution cell</i> which represents the resolution unit of a pulsed RADAR system: two objects within the same resolution cell cannot be distinguished. (Adapted from [23] and reproduced by permission from Author G. Richard Curry, Radar System Performance Modeling, <i>Second Edition</i> , Norwood, MA: Artech House, Inc., 2005. ©2005 by Artech House, Inc.)	9
1.4	Illustration of RADAR pulse transmission, reception and ambiguity case. . .	10
1.5	SLAR geometry. (Adapted from [28] with the permission of John Wiley & Sons, Inc.)	18
1.6	Intuitive approach to SAR azimuth resolution.	21
1.7	SAR azimuth resolution. (Adapted from [28] with the permission of John Wiley & Sons, Inc.)	23
1.8	SAR focused azimuth resolution. (Adapted from [28] with the permission of John Wiley & Sons, Inc.)	24
1.9	Example of possible SLAR and SAR distortions. From left to right we can see: (a) slant-range compression (the bottom part of the images shows the distorted image on top of the corrected one), (b) foreshortening, (c) layover, and (d) shadowing [38]. (With the permission of Natural Resources Canada (NRC). Reproduction has not been produced in affiliation with, or with the endorsement of, NRC.)	27
1.10	InSAR imaging geometry.	30

1.11	Example of two-pass DInSAR. (a) The amplitude image of an Single Look Complex (SLC) SAR image acquired on 4 October 1995 by the ERS-1 satellite over Peulik Volcano, Alaska. (b) The phase image of the SAR image acquired on 4 October 1995. (c) The phase image of an SLC SAR image acquired on 9 October 1997 by the ERS-2 satellite over Peulik Volcano, Alaska. The amplitude image is similar to that in (a) and therefore is not shown. (d) An original interferogram formed by differencing the phase values of two co-registered SAR images (b) and (c). The resulting interferogram contains fringes produced by the differing viewing geometries, topography, any atmospheric delays, and surface deformation. (e) An interferogram simulated to represent the topographic contribution in the original interferogram (d). (f) A topography-removed interferogram that was produced by subtracting the interferogram in (e) from the original interferogram in (d). The concentric pattern indicates approximately 17 cm of uplift centered on the southwest flank of Peulik Volcano, Alaska, which occurred during an aseismic inflation episode from October 1996 to September 1998. Each interferometric fringe (full-color cycle) represents 360° of phase change (or 2.83 cm of range change between the ground and the satellite). (From [50] with permission from the American Society for Photogrammetry & Remote Sensing, Bethesda, Maryland, www.asprs.org).	35
1.12	Comparison between (a) PSInSAR and (b) SqueeSAR when applied to a low populated area (near Winkler, Texas) (see section 2.2.2: <i>The Winkler sinkholes (Wink Sinks)</i>). The two images are showing the average scatterer velocity obtained by processing 22 images acquired between June 3, 1992 and February 21, 1998 by the ERS satellites 1 and 2. (Courtesy of TRE-Canada.)	41
2.1	Dissolution process for carbonate rocks [73]. (Courtesy of the U.S. Geological Survey)	46
2.2	Dissolution sinkholes [73]. (Courtesy of the U.S. Geological Survey)	48
2.3	Cover-subsidence sinkholes [73]. (Courtesy of the U.S. Geological Survey)	48
2.4	Cover-collapse sinkholes [73]. (Courtesy of the U.S. Geological Survey)	49
2.5	Upward propagating cover-collapse sinkholes [73]. (Courtesy of the U.S. Geological Survey)	50
2.6	Pointcloud representing the average subsidence velocity overlapped to an aerial of the interested area. Highlighted are the areas interested by the four subsidence phenomena. The reference point is identified in light green towards the bottom right corner.	51
2.7	Three temporal samples of the scatterers displacement over the Wink Sinks region. The samples provide the measured displacement after 1, 16 and 70 months past the first acquisition. Visible in the aerial layer are the actual collapsed region of Wink 1 (left) and Wink 2 (center).	52
2.8	Construction of displacement profile (top) based on data relative to a circular area surrounding Wink 2 for a single time frame (bottom). Dark markers represent scatterer affected by the largest displacement while blue/green those affected by the smallest. (From [78] ©2013 Taylor and Francis)	53

2.9	Different view of the profile stacks for Wink 2 (top) and Wink 3 (bottom): (a, d) spatiotemporal stack; (b, e) temporal evolution; (c, f) spatial evolution. (From [78] ©2013 Taylor and Francis)	54
2.10	Normalized Cross-Correlation between displacement profiles and fitted Gaussians for Wink 1 (blue), Wink 2 (green), and Wink 3 (red). (From [78] ©2013 Taylor and Francis)	55
2.11	Gaussian fit parameters versus time for Wink 1 (blue), Wink 2 (green), and Wink 3 (red): (a) amplitude and (b) sigma. (From [78] ©2013 Taylor and Francis)	56
2.12	Linear regression (dashed lines) of the amplitude versus time plots (solid lines) in Fig. 2.11 for Wink 1 (blue), Wink 2 (green), and Wink 3 (red). (From [78] ©2013 Taylor and Francis)	57
2.13	Illustration of the interaction forces considered in ESyS-Particles. (a) illustrates the basic frictional-elastic interaction between two unbonded particle, whereas (b) shows the more complex interactions accounted for when considering bonded particles.	60
2.14	Layered structure for the simulation of sinkhole formation with different type of top soil.	62
2.15	Simulation of a dissolution sinkhole. Three different stages of the simulation ($t = 0.2, 1.1, 20.0$ s) are shown in (a), (b), and (c) respectively. In (d) we illustrate the evolution of the displacement profile, in (e) the amplitude and in (f) the scale (sigma) of the best fitting Gaussian as functions of time.	65
2.16	Simulation of a cover-subsidence sinkhole. Three different stages of the simulation ($t = 0.2, 1.1, 20.0$ s) are shown in (a), (b), and (c) respectively. In (d) we illustrate the evolution of the displacement profile, in (e) the amplitude and in (f) the scale (sigma) of the best fitting Gaussian as functions of time.	66
2.17	Simulation of a cover-cover collapse sinkhole. Three different stages of the simulation ($t = 0.2, 1.1, 20.0$ s) are shown in (a), (b), and (c) respectively. In (d) we illustrate the evolution of the displacement profile, in (e) the amplitude and in (f) the scale (sigma) of the best fitting Gaussian as functions of time.	67
2.18	Simulation of a sinkhole forming under an hard rock top. Three different stages of the simulation ($t = 0.3, 1.1, 20.0$ s) are shown in (a), (b), and (c) respectively. In (d) we illustrate the evolution of the displacement profile, in (e) the amplitude and in (f) the scale (sigma) of the best fitting Gaussian as functions of time.	68
3.1	Example of basic graph.	74
3.2	Steps in the graph-cut process for Wink 2 and 4. (a) Original dataset representing the displacement of the scatterers for the acquisition taken 68.68 months after the reference image. Each point represent a vertex and is located according to its physical coordinates. (b) Delaunay triangulation defining the connectivity of the graph. (c) Result of the graph-cut algorithm. (d) Spline represented on top of the original data. (From [95] ©2012 IEEE)	80
3.3	Results of the graph cut algorithm applied to the Wink Sinks compared with ground truth data. (a) Wink 1. (b) Wink 2 and 4. (c) Wink 3.	81

3.4	Artificial test datasets and results of the detection algorithm applied to these datasets. (a) Time varying uniform spatial grid (shown at $t = 44.5$ months). (b) Time varying non-uniformly-sampled version of (a) (shown at $t = 44.5$ months). (c) Time fixed version of (b): the frame for $t = 44.5$ months is repeated for all the time intervals. (d-f) Result of the detection algorithm applied to (a-c). (g-i) Propagated version of (d-f). (From [78] ©2013 Taylor and Francis) . . .	89
3.5	Propagated results of our detection algorithm applied to the Wink Sinks. In (a) we show the propagated residual whereas in (b) we illustrate the propagated risk evaluated according to (3.16) (see section 3.3.3: <i>Virginia</i>). ((a) from [78] ©2013 Taylor and Francis)	93
3.6	Detailed results of our detection algorithm on the Wink Sinks. (a) Detection of Wink 1. (b) Detection of Wink 2 and Wink 4. (c) Detection of depression (potential Wink 3). (d) Propagated version of (a) with included outline of collapsed portion of Wink 1. (e) Propagated version of (b) with included outline of collapsed portion of Wink 2. (f) Propagated version of (c). (From [78] ©2013 Taylor and Francis)	94
3.7	Direct comparison between the (left) graph cut algorithm and (right) spatiotemporal matching approach for the Wink Sinks dataset with details for comparison with the collapsed sections of Wink 1 and Wink 2. In all the images the color mean high (red), moderate (orange) and slight (yellow) risk (see text for discussion).	95
3.8	Project area of interest $40 \times 40 \text{ km}^2$ located in Virginia, United States. Areas prone to sinkhole formation is highlighted in light blue while previously repaired sinkholes are indicated by dark red dots. (From [78] ©2013 Taylor and Francis)	97
3.9	(a) Processing flow diagram. (b) Virginia point cloud dataset showing location and average displacement velocity in mm/year of the PS. (c) Propagated residual map. (d) Propagated minimum-risk map. ((b) and (d) from [78] ©2013 Taylor and Francis)	99
3.10	Comparison between the results the proposed detection algorithm on the Wink dataset using variations on the model described in (3.8). For comparison, the location of the available scatterer is marked with black dots. The displayed map are not propagated to allow for better comparison. (a) Static model with fixed amplitude $\alpha = -104.63 \text{ mm}$ and $\sigma = 95 \text{ m}$. (b) Model with adapting amplitude α and fixed $\sigma = 95 \text{ m}$. (c) Model with fixed amplitude $\alpha = -104.63 \text{ mm}$ and adapting σ . (d) Fully adaptive model as described by (3.8). (From [78] ©2013 Taylor and Francis)	104
3.11	Comparison between different subsidence models performances. (a) Selection of on-target (green) and off-target (red) regions for each subsidence region in the Wink Sinks dataset. The dots represents the location of each scatterer and are color coded based on the displacement (in mm) measured at $t = 68.68$ months. (b) Quality factor (see text) evaluated for each of the presented models. (From [78] ©2013 Taylor and Francis)	106
4.1	Illustration of the distribution of the displacements for the entire dataset at $t = 1.15$ months (blue) and $t = 68.68$ months (red). (From [126] ©2013 SPIE)	110

4.2	(a) Example of possible subdivision of the Wink Sinks dataset into $N = 2601$ non-overlapping rectangular subsets identified by different cycling colors. The black rectangles show the regions where the subsidence activity is located. (b) Scatter plot of standard deviation vs. mean for the displacements within each block shown at $t = 1.15$ months (blue) and $t = 68.68$ months (red). (From [126] ©2013 SPIE)	111
4.3	(a) Scatter plot of the differential mean and standard deviation of the displacement for each block in the dataset. The outliers according to the Mahalanobis distance are identified in red. (b) The Wink Sinks dataset with identified (solid red) the block selected as outliers (active). For comparison we highlighted (green) the location of scatterers with an overall displacement larger than -50 mm. In light blue are indicated the block selected for analysis. (From [126] ©2013 SPIE)	112
4.4	Comparison between the residual maps resulting from application of the spatiotemporal matching algorithm to (a) the active regions and (b) the entire dataset. (From [126] ©2013 SPIE)	114
4.5	Results for the NMI analysis with progressively large size templates: (a) $\sigma = 50$ m, (b) $\sigma = 100$ m, and (c) $\sigma = 150$ m. In (d) we show the map resulting from the maximum intensity projection of the NMI along the scale (σ) axis. (From [126] ©2013 SPIE)	120
4.6	Comparison between (a) the spatiotemporal matching and (b) the NMI approaches applied to the active regions. (From [126] ©2013 SPIE)	121
5.1	Example of eigenvectors of the Laplacian for regular graphs. The top six plots represent eigenvectors for eigenvalues 0, 1, 3, 7, 40, and 79 for a cycle graph. For this graph, the eigenvector corresponding to the $u_{2k+1} = \cos(2\pi kn/N)$ and $u_{2k+2} = \sin(2\pi kn/N)$ for $k = 0, 1, \dots, N/2 - 2$ whereas $u_0[n] = 1$ and $u_{N-1}[n] = \cos(\pi n)$ (where n is the node "number"). The bottom six show eigenvectors 0, 1, 3, 4, 162, and 323 for a periodic lattice graph. The periodicity is obtained by folding the graph as a T^2 and connecting the folded ends. In both cases it is possible to see how a high frequency of change of sign between adjacent nodes corresponds to a higher eigenvalue.	133
5.2	From left to right, top to bottom, eigenvectors corresponding to eigenvalues 0, 1, 5, 10, 20, 50, 100, 200, and 400 for a graph obtained from the Wink 3 dataset. The verices are represented based on their actual coordinates. The connectivity is evaluated using the Delaunay triangulation and the weights assigned based on the actual physical distance between the points (Thicker edges correspond to higher weights). To improve readability of the plot, we simply show the sign domains (positive and negative) instead of the actual values.	134
5.3	Top: from left to right, three examples of artificial normalized Gaussian signals generated on the Wink 3 dataset with increasing value of sigma: 0.0005° , 0.0015° , 0.003° . Bottom: the corresponding spectra (only the first 500 eigenvalues) evaluated using (5.3).	135

5.4	Top: from left to right, three examples of artificial normalized Gaussian spectra (only the first 500 eigenvalues). Middle: the direct reconstruction of the spectra using (5.2). It is noticeable how a narrower spectrum will corresponds to a smoother signal. Bottom: the signal obtained from reconstruction of the spectra using (5.6) (centered at node 700).	140
5.5	Top row: the graph corresponding to the three datasets: (a) dataset I, (b) dataset II, and (c) dataset III. Second: the results of the normalized cross-correlation (5.9) with the template (5.11). Third: the template shifted to the maximum of the cross-correlation and scaled to minimize the least square error (shown) with the signal. Bottom row: same as third row with the difference that scale analysis (5.10) was applied to the template.	144
5.6	The signals represented in this figure correspond to those shown in Fig. 5.5. Here we represent the signals as function of the node index rather than as graph to favor a direct comparison. In the last two rows, where we display the template shifted at the node where the feature was detected and scaled to minimize the least square error with the original signal, the actual value of the LSE is also reported.	145
5.7	Graphical illustration of the implementation of (5.10) for the three dataset: (a) dataset I, (b) dataset II, and (c) dataset III. The top row shows the surface resulting from the analysis. The correlation coefficient is shown as function of node index and scale of the template spectrum. Below each surface, we display the correlation coefficient as function of the scale (one plot for node). Here it is possible to clearly identify the maximum that simultaneously provide identification of location and size of the best fitting template. In the case of Wink 3, since template and signal have opposite signs, we should identify the minimum.	146
5.8	Graphical illustration of the implementation of (5.10) for a dataset containing two features. (a) Representation of the graph signal (the underlying graph is not represented). (b) The graph signal illustrated as function of the node index n . (c) $\bar{\xi}_\sigma[n]$ as function of n and σ . (d) Maximum intensity projection of (c) along the scale axis represented on the base graph. (e) Maximum intensity projection of (c) along the scale axis represented as function of the node index to allow for direct comparison with (b). (f) Detections of features in the graph signal, obtained by thresholding (d) at a significance level of 0.75.	147
5.9	Comparison between spatiotemporal matching, NMI metric pre-screening, and graph scale methods over the Wink 3 area (dataset III). In (a) and (b) we show $\bar{\xi}_{max}[n]$ represented in two different color maps to favor direct comparison with the results of the spatiotemporal matching algorithm (c) and the mutual information pre-screening (d) applied to the same region (For more information about (c) and (d) see Fig. 4.6). In the bottom row we overlap the graph scale results to the spatiotemporal matching (e) and mutual information metric (f).	149
5.10	Comparison between results of (a) the graph cut approach, (b) the modified graph cut method, (c) the spatiotemporal matching algorithm and the graph scale analysis. The direct comparison is offered respectively in (d), (e), and (f).	150

B.1	Wink Sinks dataset at three different time slices.	166
B.2	Geometry of the ERS image acquisitions over the Wink Sinks area. (Courtesy of TRE Canada)	167
B.3	Virginia dataset at three different time slices.	169
B.4	Geometry of the COSMO-SkyMed image acquisitions over the Virginia area. (Courtesy of TRE Canada)	170

List of Symbols

(x_i, y_j)	Coordinates of vertex v_{ij} of a regular 2D graph embedded in an Euclidian space
A_e	Antenna effective aperture
A_i	Subset of vertices within a graph in graph-cut
A_i^c	Complement of the set A_i in V
B	Bandwidth for pulsed RADAR systems
B_n	Normal component of the InSAR baseling
$B_n^{(i)}$	Normal component of the baseline used to generate the i -th element of the InSAR stack
B_x, B_y	Orthogonal component of the InSAR baseline
C	A cut in graph-cut theory
D	Antenna directivity
D_A	Amplitude dispersion index
E	Graph incidence matrix
E	Set of edges of a graph
G	Antenna gain
$H(X)$	Entropy of the set X
H_0	Satellite elevation in InSAR
I	Mutual Information
I_i	Complex single look SAR image
K	Chirp rate for pulsed RADAR systems.
L	Discrete Laplace-Beltrami 2D operator
L_a	Size of the RADAR rectangular antenna in the azimuth direction

L_{synth}	Length of the SAR synthesized antenna
$M_n^{(t)}$	Mahalanobis distance of $\xi_n^{(t)}$ with respect to the ensemble of the $\xi_n^{(t)}$ for outliers detection
NMI	Normalized Mutual Information
$P_{r_{min}}$	Minimum detectable signal power for RADAR systems
P_r	Power reflected back to a RADAR system by a target
P_t	Peak transmitted power for RADAR systems
R	Range: Distance of an object from the RADAR system
R_0	Target range distance in InSAR
R_F	Far range for a SAR system
R_N	Near range for a SAR system
R_{max}	Maximum range for RADAR systems
R_{min}	Minimal measuring range for pulsed RADAR systems
R_m	Midswath slant range
R_u	Maximum unambiguous range for pulsed RADAR systems
$S(f)$	Spectrum of the waveform transmitted by a pulsed RADAR system
$S_{[i]}$	Graph shift operator
S_n	Blocks subdivision of dataset for outliers detection
T	Period of observation of a target returns in SAR systems
T_i	Time interval between the master and the i -th slave in the SAR stack
T_p	Pulse repetition interval for pulsed RADAR systems
V	Set of vertices (nodes) of a graph
V_s	RADAR platform velocity
$W_{\mathbf{x}}$	Window around pixel \mathbf{x} used to evaluate coherence $\gamma(\mathbf{x})$
W_{θ_A}	Azimuth spatial resolution for a RADAR rectangular antenna
W_{θ_E}	Elevation spatial resolution for a RADAR rectangular antenna
W_a	Size of the RADAR rectangular antenna in the elevation direction

W_g	Swath width
W_s	Swath width projection in the direction of the incident RADAR beam
ΔR	Difference in range distance of a target in InSAR
ΔR	Range resolution
ΔR_g	Ground range resolution
Δa	Ambiguous time for pulsed RADAR systems
Δt	Echo delay for pulser RADAR systems
Δv_{lm}	Mean relative velocity between PS candidates l and m
$\Delta x, \Delta y$	Spacing on regular 2D grid graph
$\Delta \phi$	Difference in phase of a target in InSAR
$\Delta \phi_{lm}^{(i)}$	Difference in phase between PS candidates l and m in the i -th differential interferogram
Ω_A	Antenna beamwidth
$\Sigma^{(t)}$	Covariance matrix of the set of $\xi_n^{(t)}$ obtained from subdivision of the point cloud dataset for outliers detection
α	Inter-category weight parameter for w_b in graph cut
$\alpha^{(i)}$	Atmospheric phase contribution in the i -th element of a SAR stack
β	Inter-category weight parameter for w_d in graph cut
\mathbf{A}	Adjacency matrix for a graph
\mathbf{D}	Degree matrix for a graph
$\mathcal{L}_{\mathbf{rw}}$	Random walk normalized graph Laplacian
$\mathcal{L}_{\mathbf{sym}}$	Symmetric normalized graph Laplacian
\mathcal{L}	Graph Laplacian
\mathbf{e}_i	i -th element of the standard (discrete-unit-interval) basis
\mathbf{f}	Graph signal
\mathbf{u}_ℓ	Eigenvector of the graph Laplacian corresponding to the λ_ℓ eigenvalue.
\mathbf{w}_m	m -th element of the DFT basis
$\boldsymbol{\tau}$	Graph template signal

χ_2^2	Chi-Square distribution with two degrees of freedom
δx	Along-track (azimuth) resolution
δ	Inter-category weight parameter for w_l in graph cut
ϵ_i	Phase error in the i -th element of a SAR stack due to potential error in the DEM
ϵ_{lm}	Relative elevation error between PS candidates l and m
η	Air/space-borne RADAR incidence angle
γ	Air/space-borne RADAR look angle
γ	Inter-category weight parameter for w_r in graph cut
$\gamma(\mathbf{x})$	Coherence of pixel \mathbf{x} between two SAR images
γ_{lm}	Overall coherence for PS candidates l and m
λ	RADAR operating wavelength
λ_0	First eigenvalue (after sorting) of the graph Laplacian
λ_ℓ	ℓ -th eigenvalue (after sorting) of the graph Laplacian
\mathcal{N}_i	Neighborhood of the vertex v_i in a graph
μ	Potential non modeled motion component for a target
μ_A	Average value of a pixel amplitude in a SAR stack
$\mu_n^{(t)}$	Average displacement within a block S_n for outliers detection
$\overline{\xi^{(t)}}$	Average of the set of $\xi_n^{(t)}$ obtained from subdivision of the point cloud dataset for outliers detection
ϕ	Round-trip phase change from RADAR to target
$\phi(x, \omega)$	Complex exponential element of Fourier basis
$\phi^{(i)}$	Residual phase for the i -th element of a SAR stack after removal of known phase contributions
ϕ_E	Elevation angle for ground based RADAR systems
σ	RADAR cross section
σ_A	Standard deviation of a pixel amplitude in a SAR stack
σ_ϕ	Phase dispersion for a pixel in a SAR stack
$\sigma_n^{(t)}$	Standard deviation of the displacements within a block S_n for outliers detection

τ_c	Effective duration of the pulse at the output of a matched filter
τ_p	Duration of pulse for pulsed RADAR systems
τ_r	Recovery time for monostatic pulsed RADAR systems
θ	Azimuth-north angle for ERS-1/2 observation of Wink Sinks scene
θ	Target broadside angle
θ_A	Azimuth angular resolution for a RADAR rectangular antenna
θ_E	Elevation angular resolution for a RADAR rectangular antenna
$\xi_n^{(t)}$	Point identified by the average displacement and standard deviation of the displacements, relative to time t , within the block for outliers detection
$a(t)$	Envelope of the transmitted signal for pulsed RADAR systems
c	Coherence of a scatterer in graph cut
c	Speed of light ($c \approx 3 \times 10^8$ m/s)
d	Displacement of a scatterer in graph cut
d_i	Degree of the vertex v_i in a graph
e_{ij}	Edge between vertices v_i and v_j of a graph
f_c	Carrier frequency of the transmitted signal for RADAR systems
f_r	Pulse repetition frequency for pulsed RADAR systems
$g(t)$	Output of a matched filter
h	Target elevation in InSAR
h_a	Altitude of ambiguity in InSAR
l	Physical distance between two scatterers in graph cut
n	General decorrelation noise term in PSInSAR
$p(d_i, g_i)$	Joint probability mass function of the sets D and G
$p(x_i)$	Marginals of the joint probability mass function for the set X
r_i	Potential motion of a target in range direction in the i -th element of a SAR stack
$s(t)$	Waveform transmitted by pulsed RADAR system
v_i	Component of the mean target velocity in the range direction for the i -th element of the InSAR stack

v_i	Vertex (node) of a graph
w_b	Edge weight associated with displacement range in graph cut
w_b	Edge weight associated with physical distance between scatterers in graph cut
w_d	Edge weight associated with displacement difference in graph cut
$w_{lm}^{(i)}$	Phase residual after coherence maximization for PS candidates l and m
w_r	Edge weight associated with scatterer coherence in graph cut
x	Along-track (azimuth) distance
ADI	Amplitude Dispersion Index
ALOS-2	Advance Land Observing Satellite-2 DAICHI-2
AMI	Active Microwave Instrument (SAR instrument onboard ESR satellites)
APS	Atmospheric Phase Screen
ASI	Agenzia Spaziale Italiana
AWGN	Additive White Gaussian Noise
COSMO-SkyMed	COntstellatIon of small Satellites for Mediterranean basin Observation
CSA	Canadian Space Agency
DEM	Digital Elevation Model
DEM	Discrete Element Method
DFT	Discrete Fourier Transform
DInSAR	Differential Interferometric Synthetic Aperture RADAR
DOT	Department Of Transportation
DS	Distributed Scatterer
DSM	Digital Surface Model
DTM	Digital Terrain Model
ERS	European Remote Sensing satellite
ESA	European Space Agency
FEMA	Federal Emergency Management Agency
FIR	Finite Impulse Response

FLH Office of Federal Lands Highway
GFT Graph Fourier Transform
GSP Graph Signal Processing
IIR Infinite Impulse Response
InSAR Interferometric Synthetic Aperture RADAR
JAXA Japanese Aerospace Exploration Agency
LOS Line Of Sight of the satellite
NRC Natural Resource Canada
PRF Pulse Repetition Frequency
PRI Pulse Repetition Interval
PS Permanent/Persistent Scatterer
PSI Persistent Scatterer Interferometry
PSInSAR Permanent Scatterer InSAR
RADAR RAdio Detection And Ranging
RCM Radarsat Constellation Mission
RITA Research and Innovative Technology Administration
SAR Synthetic Aperture RADAR
SLAR Side-Looking Airborne/real-Aperture RADAR
SNR Signal to Noise Ratio
TRE Tele-Rilevamento Europa
USDOT United Stated Department Of Transportation
VDOT Virginia Department of Trasportation

Chapter 1

Introduction and technical background

1.1 Motivation

Every year the U.S. Department of Transportation spends several millions of dollars to perform repairs to the infrastructure damaged by geophysical phenomena such as sinkholes and landslides [1]. The sudden and often unpredictable nature of these events is the main cause of the high repair costs and the disruption of the regular flow on the transportation network. In 1991, the Committee on Ground Failure Hazards Mitigation Research within the Division of Natural Hazard Mitigation of the Commission on Engineering and Technical Systems of the National Research Council published a report titled *Mitigating Losses from Land Subsidence in the United States* [2]. In this report it was stated that:

"More than 44,000 km² of land in 45 states in the United States have been lowered by the type of subsidence considered in this report.

[...] about 18 percent of the conterminous United States is underlain by cavernous limestone, gypsum, salt, or marble and is locally susceptible to catastrophic collapse into sinkholes.

Annual costs from resulting flooding and structural damage exceed \$125 million.”

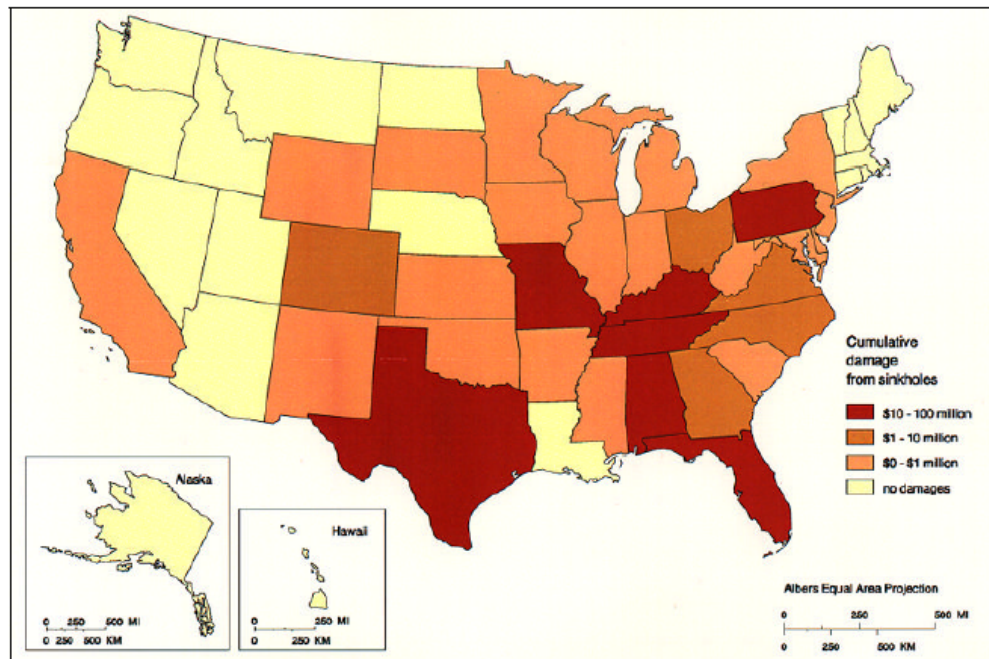


Figure 1.1: Cumulative subsidence damage caused by sinkhole. Time periods on which estimates are based vary by State, and costs are not converted to constant dollars. Data not available for Puerto Rico, U.S. Virgin Islands, and Pacific Territories. (From [3] with the permission of FEMA)

In 1997, the Federal Emergency Management Agency (FEMA) released a report titled *Multi-Hazard Identification and Risk Assessment* [3] where it summarized the results of the 1991 report with a chart (Fig. 1.1) describing the cumulative repair costs for damages caused by sinkholes in the United States.

In the same report it was stated that:

“Landslides occur in every State [...]. According to a 1985 study, roughly 40 percent of the U.S. population is exposed to direct and indirect effects of landslides.

Landslides have damaged or destroyed roads, railroads, pipelines, electrical and telephone lines, mines, oil wells, buildings, canals, sewers, bridges, dams, seaports, airports, forests, parks, and farms. [...] The best estimate of annual losses resulting from landslides in the United States are 25 to 50 lives and \$1 to \$2 billion in property damage.”

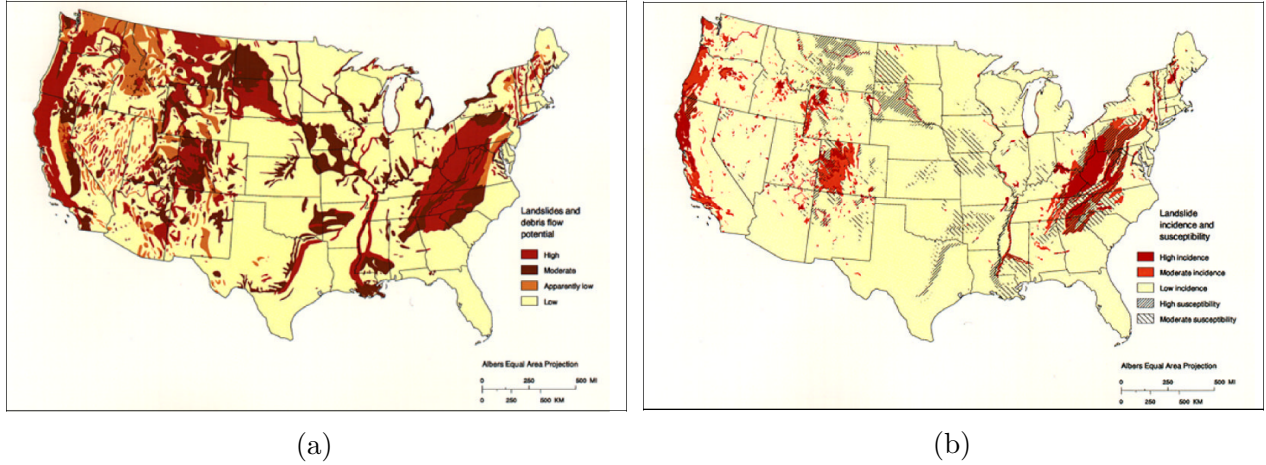


Figure 1.2: (a) Landslide potential based on adverse formations associated with past landslide activity. (b) Landslide incidence and susceptibility in the conterminous United States. Data not available for Alaska, Hawaii, Puerto Rico, U.S. Virgin Islands, and Pacific Territories. (From [3] with the permission of FEMA)

This was clearly summarized in the report by two charts (Fig. 1.2) displaying the potential and incidence of landslides.

Although sinkholes and landslides are caused by fundamentally different underlying geophysical phenomena, they have in common the fact that they both result in the movements of the Earth’s surface. In this work, we present our contribution to the development of early detection techniques, applicable to subsidence events, based on the use of novel remote sensing RADAR products capable of identifying minute changes in ground elevation.

Until recently, the main mitigation approaches have included careful mapping of areas known to be prone to subsidence phenomena, followed by expensive and time consuming

geophysical inspection techniques [4] such as ground penetrating radar [5], cone-penetrometer tests [6], standard penetration tests [7], and electrical resistivity tests [8]. Although these techniques can provide very detailed information about the ground structure and its changes over time, they are very labor intensive and extremely localized in nature thus seldom used as a routine surveying approach but rather for analysis once subsidence has been detected [9].

Starting in the early 1990s, mainly due to the deployment of the European Remote Sensing (ERS) constellation (see section [A.1: *European remote sensing satellites \(ERS\)*](#)) by the European Space Agency (ESA), civilian projects have been allowed increased access to synthetic aperture RADAR (SAR) instrumentation spawning the development of advanced imaging techniques such as interferometric SAR (InSAR).

There are several examples of successful applications of SAR and InSAR. The Office of Federal Lands Highway¹ successfully conducted several studies at the Cedar Pass area of Badlands National Park in South Dakota [10], Cimarron and Mesa Verde Slides in Colorado, Prosser Slide in Washington State [11], and the Amphitheatre Point Slide in California [12] using InSAR to detect ground deformations and identify specific areas requiring further ground-based surveys. Scientists at the U.S. Geological Survey routinely use InSAR to help correlate ground movements with seismic events and to detect fault zones as well as deformations due to volcanic activity [13]. InSAR has proved successful in long-term monitoring of ground subsidence in Sicily [14], where the phenomenon was eventually traced back to aquifer depletion through excessive groundwater extraction. Another InSAR analysis, relating hydrocarbon reservoir volumetric strain to pressure depletion, was applied to the

¹The office of Federal Lands Highway (FLH) is part of the US Department of Transportation (USDOT), Federal Highway Administration. The primary purpose of the FLH is to provide financial resources and technical assistance for a coordinated program of public roads that service the transportation needs of Federal and Indian lands.

extensive oil and gas fields in Oman with approximately 500 producer wells [15]. Here, the study of the horizontal and vertical components of the detected bowlshaped ground deformation areas provided indication of subsurface heterogeneity leading to detection of fault lines. The InSAR approach was also found to be suitable for monitoring wetlands in Floridas Everglades [16]. Here, the analysis of X-band (8-12 GHz) data from the TerraSAR-X satellite provided an indication of surface water level variations with accuracy of two to four centimeters. In a recent study conducted over sinkhole-prone regions at Wink and Daisetta in Texas, a combination of gravimetry and radar interferometry was used and the results indicated fair level of agreement between the two techniques [17] [18]. Another recent study [19] presented results of the use of InSAR for subsidence detection in the Carlsbad mining district (New Mexico). The results were obtained by applying differential InSAR (DInSAR) techniques to compare the phase between two images taken on July 2005 and March 2006.

The transportation community has shown a growing interest in the potential application of these techniques to the routine monitoring of key infrastructure features and their adjacent land corridor. These areas are often the point of origin of significant safety hazards to highway and railway operations [20] and the detection and analysis of precursory motion could identify potential landslides and sinkholes, reducing the maintenance costs² by allowing local Department of Transportation (DOT) to schedule preemptive maintenance and consolidation works. The limited occurrence of emergency repairs would result in lower prices for raw materials, increased safety for the public, and reduced disruption to the general traffic.

Because of this growing interest, the U.S. DOT Research and Innovative Technology

²Brian Bruckno, Virginia Department of Transportation research geologist, stated that just in central Virginia more than \$1.2M was spent in 2011 for the remediation of sinkholes and landslide with any single event costing in excess of \$25 k (personal communication, March 2013).

Administration (RITA) has funded several projects focusing on the use of remote sensing techniques to monitor the condition of the transportation network. In particular, the National Consortia on Remote Sensing in Transportation within the Commercial Remote Sensing & Spatial Information Technologies program consists of university-based research groups, including the Virginia Image and Video Analysis laboratory at the University of Virginia, leaders in the development of products and tools based on remote sensing technologies.

Within this framework, we will present here our contributions that were developed under the sponsorship of RITA and in collaboration with the Virginia Center for Transportation Innovation and Research and TRE Canada Inc. (a subsidiary of Tele-Rilevamento Europa) the world leader company in ground deformation monitoring services with InSAR technology. In particular, we will present a set of algorithms that, by making use of the last generation of interferometric products, capable of resolution down to the millimeter level, generate maps identifying locations where displacements of the Earth's surface conform to a predefined spatiotemporal model.

1.2 Monitoring Earth's surface movements

Topographic maps have been the main way to represent local and global elevation for thousands of years [21]. In the past decades, with the advent of digital techniques, elevation data have been recorded in the form of digital elevation model (DEM)³.

³DEM is a generic name for both digital terrain model (DTM) and digital surface model (DSM) with the former being a representation of the bare ground elevation and the latter including the height of any object present on the surface.

Before the development of remote sensing techniques, DEMs were mainly obtained thanks to measurements in the field. These long and painstaking campaigns were conducted while dealing with difficult logistics, safety concerns, and limited resources. As a result, often the collected data did not cover wide areas, the spatial sampling was not dense enough to provide detailed information about the topography, and campaigns were seldom repeated to evaluate ground deformation over time. Once radio detection and ranging (RADAR) instrumentation became available to the scientific community as a remote sensing tool, it sparked, over the last two decades, the development of new high resolution techniques. Today, we have the capability of mapping large regions (hundreds of square kilometers) with very high spatial resolution (few square meters), while measuring elevation with sub-meter and surface deformations with millimeter accuracy.

In order to generate the DEM of a region of interest, it is necessary to obtain three pieces of information:

- an accurate map of the observed scene allowing easy identification of the objects present within the imaged region
- high resolution measurements of the location of the objects in the scene with respect to a reference point (typically the measuring instrument)
- the exact location of the reference point with respect to a geographic coordinate system (geo-referencing)

In the rest of the chapter, we describe the basic concepts of RADAR and how new RADAR techniques were developed to address the requirements listed above. Starting from synthetic aperture RADAR (SAR) (see section 1.3.3: *Synthetic Aperture Radar (SAR)*), capable of

acquiring high resolution two-dimensional maps, through interferometric SAR (InSAR) (see section 1.3.4: *Interferometric SAR (InSAR)*), adding the third dimension and allowing the generation of DEMs, we will conclude with differential InSAR (see section 1.3.5: *Differential InSAR (DInSAR)*) and persistent scattering (see section 1.3.6: *Permanent Scatterer InSAR (PSInSAR) and SqueeSAR*) techniques, providing the capability of measuring DEM changes over time.

1.3 RADAR for remote sensing applications

1.3.1 RADAR basics

The fundamental idea behind RADAR was born in the late nineteenth century when interference was observed, due to a passing ship, while wireless signals were being transmitted [22]. It did not take long to develop the idea further for imaging purposes: coherent narrow-band radiation was pulsed in a specific direction and the time required for the echo to return measured, providing the detection and ranging capabilities.

While optical instrumentation relies on sunlight to reach Earth and illuminate a scene of interest, one of the main characteristics making RADAR attractive is that it provides its own illumination. Furthermore, when a scene is illuminated by sunlight, only the amplitude of the scattered light can be used as source of information about the imaged scene. This is because there is no fixed (in space or in time) phase relationship between the different infinitesimal sources of light within the Sun. On the contrary, a RADAR transmitter provides a coherent signal. In this case, both amplitude and phase delay information are available in

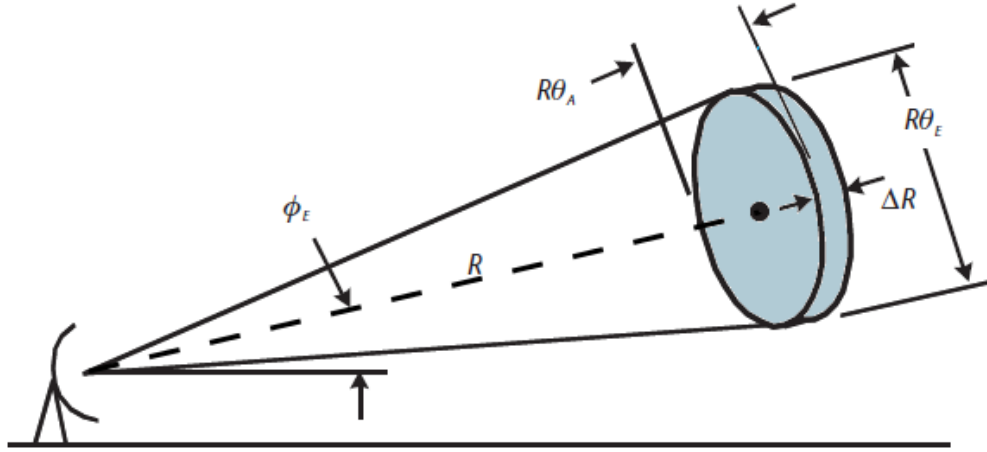


Figure 1.3: RADAR resolutions. The highlighted area is known as the *resolution cell* which represents the resolution unit of a pulsed RADAR system: two objects within the same resolution cell cannot be distinguished. (Adapted from [23] and reproduced by permission from Author G. Richard Curry, Radar System Performance Modeling, *Second Edition*, Norwood, MA: Artech House, Inc., 2005. ©2005 by Artech House, Inc.)

the return signals, allowing the development of the interferometric techniques on which SAR, InSAR, and the derived products we used in our research are based. Moreover, RADAR systems are typically operated at longer centimeter wavelengths that can easily penetrate clouds making for an attractive instrument that can be used at any time and almost in any weather condition.

In a RADAR system, a transmitter emits radio waves powerful enough that measurable amounts of radio energy are reflected from scattering objects back to a receiver, typically located at the same site as the transmitter (*monostatic* configuration). Considering the basic RADAR system illustrated in Fig. 1.3, we can define three directions: range, elevation and azimuth. *Range* is the direction of propagation of the RADAR waves (indicated by the dashed line). The distance of an object in this direction is called *range* and indicated as R . *Elevation* is the angle between a plane tangent to the Earth's surface at the location of the RADAR transmitter and the range direction and it is indicated as ϕ_E . The *azimuth* direction

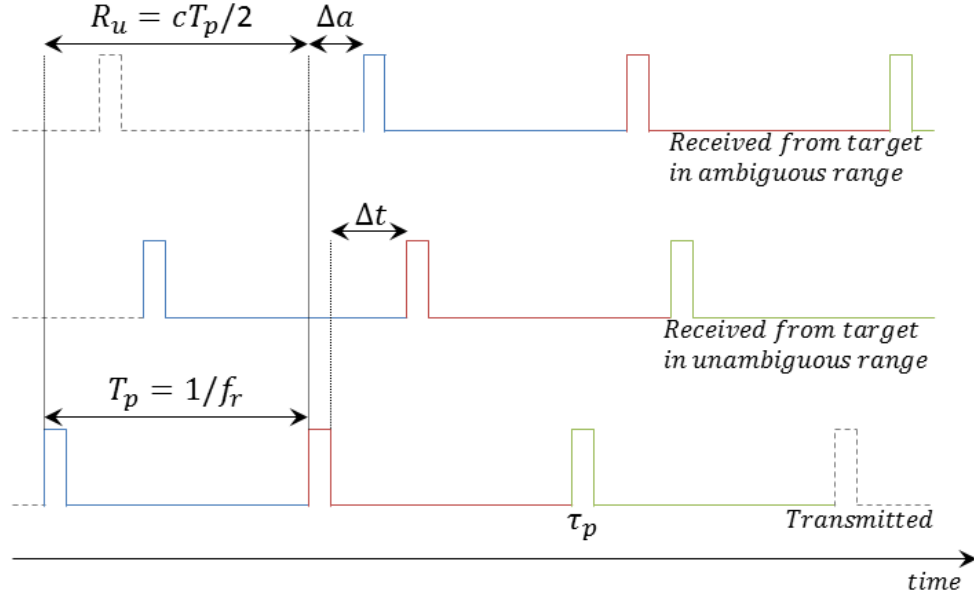


Figure 1.4: Illustration of RADAR pulse transmission, reception and ambiguity case.

(not illustrated) is defined as the angle between a reference direction in the tangent plane previously described and the range direction. These three coordinates are used to identify the spatial position of an object with respect to the known location of the RADAR system.

The fundamental process by which a RADAR measures the distance of an object is the evaluation of time delays. In the most basic design, a radio frequency source is connected to an antenna and emits a periodic sequence, of period T_p , of square pulses each of length τ_p . The period is known as *pulse repetition interval (PRI)* and its inverse (f_r) as *pulse repetition frequency (PRF)*. These values are important in determining the *maximum unambiguous range* which is the maximum range at which a target can be located: $R_u = cT_p/2 = c/2f_r$, where c is the speed of light in the medium. These concepts are summarized in Fig. 1.4 where we show two examples of echoes received from both an unambiguous and an ambiguous target (located at a distance larger than R_u). In the case of the unambiguous target, each echo is received before the following pulse is transmitted allowing for the correct evaluation

of the distance. In the ambiguous case, the echo from the "blue" pulse is received a time Δa after the transmission of the "red" pulse. Without other means to discriminate, the RADAR system will naturally interpret this echo as a response to the "red" pulse and associate it to an object closer than in reality. This ambiguity is commonly resolved by using a technique where the echoes obtained from observation made with several different pulse repetition frequencies are compared [24].

The maximum range defined by R_u does not take into consideration the actual physical dependencies of the transmitted power, the scattering characteristics of the target, its distance, and the sensitivity of the RADAR receiving systems. These properties are captured in what is known as the *radar (range) equation* which, in its most general form, can be written as⁴:

$$R = \sqrt[4]{\frac{P_t G^2 \lambda^2 \sigma}{(4\pi)^3 P_r}} \quad (1.1)$$

where P_t is the peak transmitted power, G is the antenna gain (a measure of how focused the transmitted power is in the elevation and azimuth direction), λ is the operating radar wavelength, and σ is the radar cross section (defined as the ratio between the power reflected from the target back to the radar (P_r) to the power density incident on the target).

If we identify with $P_{r_{min}}$ the minimum detectable signal power, then the maximum radar range R_{max} can be expressed as:

$$R_{max} = \sqrt[4]{\frac{P_t G^2 \lambda^2 \sigma}{(4\pi)^3 P_{r_{min}}}} \quad (1.2)$$

⁴For a detailed derivation see [25]

For monostatic pulse RADAR systems, where receiver and transmitter are co-located, there exists also a *minimal measuring range* which represents the minimum distance at which an object can be detected. This is due to the fact that monostatic systems cannot receive while they are transmitting a pulse and require a *recovery time* (τ_r) to switch from transmitting to receiving. Because of this, the RADAR is effectively blind to return signals from targets closer than $R_{min} = c(\tau_p + \tau_r)/2$.

When a target is within the minimum and maximum range limits, it is possible to evaluate its range R from the interval of time Δt elapsed between the pulse transmission and the echo reception: $R = c\Delta t/2$. In practical terms, the range corresponding to a certain delay is about 150 m for each microsecond of delay and, given the *single pixel* nature of a RADAR, the identification of the location in the azimuth and elevation directions is obtained from the actual orientation of the antenna.

Resolution

One of the key features of DEM is their resolution expressed as the minimum size of the ground element for which an elevation value is provided. This can be linked directly to the resolution of the instrument used to acquire the data. In particular, range and azimuth resolution are critical, in SAR systems, to generate precise maps that can, in turn, be used to obtain high quality DEM.

In the previously described RADAR system, we can define three resolutions, one for each direction. Together, these three values form what is known as *resolution cell* that, for a given RADAR systems, represents the minimum resolution unit (Fig. 1.3) with the implication that two objects cannot be distinguished if they are within the same resolution cell.

In order to differentiate between two targets separated in the range direction, a basic pulsed RADAR has to be able to completely receive the pulse echo from the closest one before being reached by the echo from the other. In other words, two echoes cannot overlap. In this case, the achievable resolution is given by:

$$\Delta R = c\tau_p/2 = c/2B \quad (1.3)$$

where ΔR is the range resolution, τ_p is the transmitted pulse duration, and $B = 1/\tau_p$ is the bandwidth of the RADAR system. The last relationship shows that higher range resolutions can be achieved by either decreasing the pulse duration or, equivalently, increasing the bandwidth. On the other hand, the average transmitted power, one of the key factors determining the signal to noise ratio (SNR) of the returned echoes, is given by the product of the pulse duration and its peak power. To maintain the desired return SNR, required to provide reliable detection and reduced false alarm rates, it would then be desirable to use longer pulses (or increase the peak power) while preserving the desired range resolution. Generating high peak powers at the high frequencies used in RADAR systems can be challenging. Two techniques known as *matched filter* [26][27], and *pulse compression* [25] were developed to allow the use of longer pulses while still achieving the desired range resolution.

A matched filter system is designed to search for a *known* signal within another one by evaluating the cross-correlation between the two. In the case of a RADAR system, the cross-correlation is computed between a replica of the transmitted pulse and the signal received during the inter-pulse period. This is typically achieved by performing the convolution of the received signal with the time-inverted complex conjugate version of the transmitted one. The

main advantage of this technique is that it maximizes the SNR at the filter output when the input can be characterized as a scaled version of the original signal corrupted by additive white gaussian noise (AWGN).

For a signal, we can write the response of the matched filter as:

$$g(t) = \int_{-\infty}^{+\infty} s^*(t' - t)s(t')dt' \quad (1.4)$$

where $s(t)$ represents the received signal shifted, for ease of notation, to be centered at the echo delay time (Δt). This equation is a special case of the *ambiguity function* which represents the output of a matched filter in the presence of Doppler shift (here ignored) in case of returns from a moving target [28].

The equivalent formulation in the frequency domain is given by:

$$g(t) = \int_{-\infty}^{+\infty} |S(f)|^2 \exp(-j2\pi ft)df \quad (1.5)$$

where $S(f)$ is the spectrum of the signal, and f is the frequency.

The output of the matched filter is maximized at the time when an echo of the transmitted pulse is received. Although the maximum occurs at a given instant, providing the location of an object in the range direction, the actual resolution (ability to distinguish between two subsequent echoes) is determined by the width⁵ of the power peak (the time width of $|g(t)|^2$) which corresponds to the effective width of the received pulse (τ_c) at the output of the matched filter receiver.

⁵The width is usually measured as the full width at half maximum - or 3dB-point - of the corresponding power peak.

Let us consider the response of this system when using two different type of pulses:

$$s(t) = a(t) \exp(j2\pi f_c t) \quad (1.6)$$

$$s(t) = a(t) \exp[j2\pi(f_c t + Kt^2/2)] \quad (1.7)$$

where f_c is the carrier frequency of the signal, t is the time, K is a constant, both amplitudes $a(t) = 1$ are considered constant, and the functions are defined over the interval $|t| \leq \tau_p/2$.

The signal represented by (1.6) is a burst of duration τ_p at the RADAR carrier frequency. The response to this signal can be calculated from (1.4) to be $g(t) = \tau_p(1 - |t/\tau_p|) \exp(j2\pi f_c t)$ with $t \leq \tau_p$. In this case, $|g(t)|^2$ is a quadratic function of width given by $\tau_c = \tau_p$ and the resolution is determined by the actual time width of the transmitted pulse.

The second equation (1.7) represents a signal where the frequency is linearly increasing (or decreasing depending on the sign of K) with time. This type of signal is known as *chirp* and K is called the *chirp rate*. One of the characteristics of a chirp is that the frequency spectrum is approximately a top-hat function of width (bandwidth) $B = |K|\tau_p$. In this case, from (1.5), we can see that the response of the matched filter is $g(t) = K\tau_p \text{sinc}(K\tau_p) \exp(j2\pi f_c t)$ with a width of $\tau_c = 1/B = 1/|K|\tau_p$.

The ratio between the actual duration of the pulse (τ_p) and the width at the output of the matched filter (τ_c) is known as the *compression ratio* and is given by $\tau_p B$ which, in the case of chirp signals, reduces to $|K|\tau_p^2$.

The use of matched filter receivers, not only provides the maximum available SNR in case of AWGN (effectively increasing the range of the system), but also decouples the pulse length from the achievable range resolution (effectively linking it to the bandwidth of the system).

As an example consider COSMO-SkyMed (see section A.2: *COSMO-SkyMed*), the satellite constellation providing the data for our research. The SAR instrument operates at a center frequency of $f_c = 9.6$ GHz (X-band) and can achieve chirp bandwidths up to 400 MHz (spotlight mode), obtaining a maximum theoretical range resolution of 0.5 m while using pulses τ_p with width between $30 \mu s$ and $80 \mu s$. Without pulse compression, the best achievable resolution in the range direction would be 1.5 km.

We have described how RADAR systems can achieve high range resolution. As we will see later, this is one of the two main coordinates used by SAR satellites to generate maps of the ground. The second coordinate, azimuth, will be identified by the direction of motion of the satellite. For the rest of this section, and in the following two, we are going to analyze the developments that allowed SAR instruments to achieve high resolution in this coordinate.

If we consider the system depicted in Fig. 1.3, assuming a rectangular antenna of sizes W_a in the elevation and L_a in the azimuth direction, the corresponding Fraunhofer (diffraction-limited)⁶ angular ($\theta_{E/A}$) and arc ($W_{\theta_{E/A}}$) resolutions are given by⁷:

$$\theta_E \propto \frac{\lambda}{W_a} \qquad W_{\theta_E} = R\theta_E \propto R\frac{\lambda}{W_a} \qquad (1.8)$$

$$\theta_A \propto \frac{\lambda}{L_a} \qquad W_{\theta_A} = R\theta_A \propto R\frac{\lambda}{L_a} \qquad (1.9)$$

where R is the distance in the range direction (slant range distance), and λ is the operating center wavelength of the RADAR. These angular resolutions define what is known as *RADAR*

⁶Satellite-based RADAR systems are typically operating in the Fraunhofer region which is characterized by $W^2/L\lambda \ll 1$, where W is the largest aperture size, λ is the operating wavelength, and L is the distance from the aperture. In the case of COSMO-SkyMed: $W = 5.7$ m, $\lambda = 3.125$ cm, and $L = 619.6$ km with a resulting ratio of 0.00168.

⁷A multiplicative factor of 0.886 is necessary if the standard 3dB-point (half power) is considered in the definition of beamwidth

beamwidth and, together with (1.3), they define the RADAR resolution cell (Fig. 1.3).

Modern RADAR antennae have better performance than simple rectangular apertures providing either higher directivity D and much narrower beamwidth ($\Omega_A = 4\pi/D$) or, as in the case of COSMO-SkyMed active antenna, steerable capabilities while maintaining near optimal performance [29][30]. Nevertheless, the basic principle of inverse proportionality between antenna size and resolution stands.

For a given antenna geometry, the size of the on-ground spatial resolution element (pixel) scales linearly with its distance from the antenna. In visible and infrared systems, where the wavelengths are of the order of $10^{-7} - 10^{-5}$ m, it is possible to achieve good resolutions even with modest size apertures on spaceborne instruments. In the case of RADAR systems, where the wavelengths are of the order of centimeters, this is not the case: to achieve the same spatial resolution of an optical system, the required aperture size would be hundreds of times larger, quickly becoming unfeasible.

As we will describe next, antenna synthesis techniques can be used to overcome this obstacle and provide good resolution, independent of the RADAR altitude, with reasonably sized antennas.

1.3.2 Side-Looking Airborne/real-Aperture RADAR (SLAR)

The first major breakthrough in the use of RADAR for ground monitoring was the development of the *side-looking airborne/real-aperture RADAR (SLAR)* which consisted in a RADAR system mounted on the side of an airplane.

In SLAR systems, the RADAR is typically pointed perpendicularly to the flight path

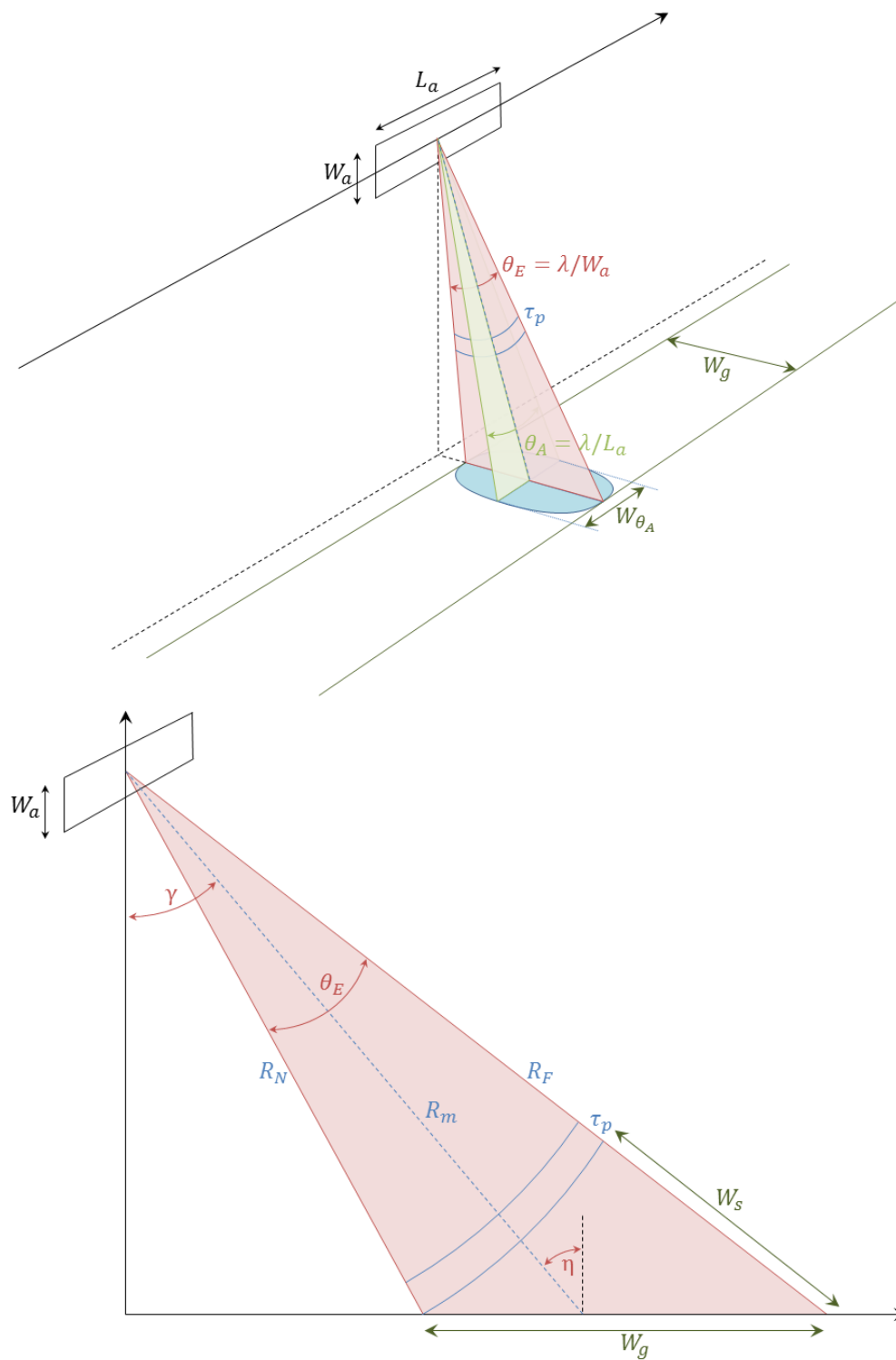


Figure 1.5: SLAR geometry. (Adapted from [28] with the permission of John Wiley & Sons, Inc.)

(sometimes the beam can be pointed forward or backward, in this case the angle between the center-beam and broadside is called the *squint angle*). The viewing direction is determined by the *look angle* (γ) which, in systems flying close enough to the surface (where Earth's curvature can be ignored), is considered identical to the *incidence angle* (η) (Fig. 1.5). In this configuration (often indicated as *strip-map* mode), the area imaged by a single pulse of the SLAR (also known as the *footprint*) is the projection of the antenna beam ($W_{\theta_E}, W_{\theta_A}$) onto the ground and is given by (W_g, W_{θ_A}), with

$$W_g = \frac{W_{\theta_E}}{\cos \eta} \approx \frac{\lambda R_m}{W_a \cos \eta} \quad (1.10)$$

$$W_{\theta_A} = R_m \theta_A \approx \frac{\lambda R_m}{L_a} \quad (1.11)$$

where W_g is the ground projection of W_{θ_E} (known as the *swath width*), and R_m is the midswath *slant range* (distance from the RADAR to the projection of the beam peak).

As the RADAR moves along in the azimuth direction, the footprint covers a swath of ground of width W_g and of a length that depends on the duration of the acquisition⁸. Within this swath, the resolution in azimuth and ground range are given respectively by W_{θ_A} and by

$$\Delta R_g = \frac{\Delta R}{\sin \eta} = \frac{c}{2B \sin \eta} \quad (1.12)$$

SLAR systems provided quite an improvement over existing ground mapping techniques, allowing the coverage of large regions with good resolution in the range direction. Nevertheless,

⁸In the early days of SLAR, mapping was cleverly obtained by recording the amplitude of the RADAR return on a photographic film where the advancement of the film was synchronized with the flight progression (azimuth direction) while the range direction was synchronized with the pulse delay [31].

the major limitation of such systems was due to the azimuth resolution which still directly depended on the physical size of the antenna and the wavelength of the RADAR. With resolutions of the order of tens of meters, when used from a low flying aircraft, if implemented with typical values for wavelength, antenna size, and range distance from space-borne systems, SLAR would give azimuth resolutions on the order of tens of kilometers: too coarse for most scientific requirements. For example, in the case of COSMO-SkyMed [32], with $W_a = 1.4$ m, $L_a = 5.7$ m, $R \approx 620$ km, $\lambda = 3.125$ cm, and $\eta = 38^\circ$, the theoretical size of the footprint would be $W_g \approx 18$ km, which is actually artificially broadened to ~ 40 km to allow uniform illumination of the ground in the range direction, and $W_{\theta_A} \approx 3.4$ km. In order to achieve an azimuth resolution similar to the one obtained from an aircraft, the requirements imposed on the azimuth dimensions of the space-borne antenna would be, once again, impractical.

The final step in ground resolution was achieved by implementing antenna aperture synthesis. This technique, at the basis of SAR, allows to obtain meter resolution images with modest size space-borne antennas. Again using COSMO-SkyMed satellites as an example, the actual SAR resolution for single-look complex images in strip-mode can be as low as $W_g = 3$ m, and $W_{\theta_A} = 3$ m with an antenna size of $W = 5.7$ m. Using advance synthesis technique - as in spot mode - where the antenna keeps tracking a localized area as the satellite moves in the azimuth direction, it is possible, for COSMO-SkyMed to achieve a resolution on ground of 50 cm.

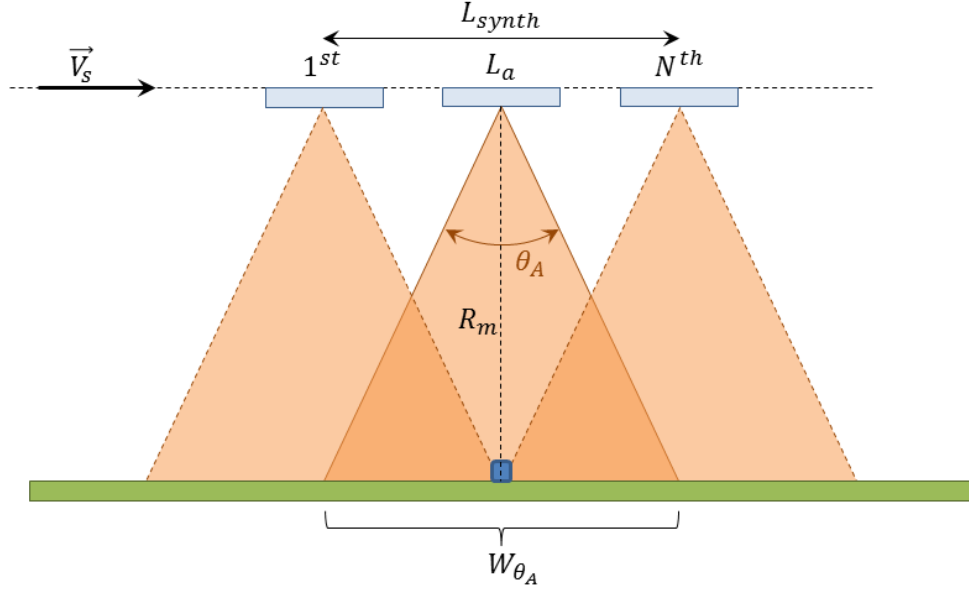


Figure 1.6: Intuitive approach to SAR azimuth resolution.

1.3.3 Synthetic Aperture Radar (SAR)

The ability of modern satellite RADAR systems to provide high resolution DEMs and track their changes over time (which is one of the key components at the basis of our research) can be tracked back to the development of SAR. The capability of this technology to image large regions with high resolution has been known since the early 50s when Carl Wiley observed a one-to-one correspondence between the azimuth coordinate of a target and the Doppler shift of the return signal⁹.

Before analyzing the details on how SAR systems achieve azimuth resolution theoretically equal to half the along-track (azimuth) size of the RADAR antenna ($L_a/2$), let's consider an intuitive approach.

⁹Carl Atwood Wiley is credited with being the first "[...] to have joined the aperture synthesis and image reconstruction techniques of radio astronomers with the techniques of microwave radiometry to achieve improved passive RADAR imaging." [33]. After successful experiments on what he called *Doppler RADAR Beam Sharpening*, he filed for a patent on August 13th, 1954 but the patent was placed under secrecy order for 11 years until it was finally granted on July 20th, 1965. [34]

As the platform carrying the RADAR moves along its track at a constant velocity, it emits a series of pulses according to the system PRF. Each pulse illuminates a footprint given by (1.10) and (1.11). Considering Fig. 1.6, we can look at each received echo as creating a virtual antenna with the location determined by the platform velocity V_s and the PRF. A target at a certain position along the azimuth direction will be illuminated by several pulses as long as it is within the diffraction-limited beamwidth of the RADAR antenna (W_{θ_A}). This is equivalent to saying that the target is "simultaneously" imaged by N virtual antennae (one for each pulse). The SAR processing achieves this "simultaneousness" by electronically synchronizing the echoes, effectively turning the N virtual antennae in a single synthesized antenna of length $L_{synth} = W_{\theta_A} \approx R_m \theta_A \approx R_m \lambda / L_a$. If we now consider the diffraction-limited resolution of an antenna of this length, we can evaluate the theoretical azimuth resolution of the SAR system: $\delta x \approx R_m \lambda / 2L_{synth} \approx L_a / 2$ (where the factor 2 in the denominator is required because of the signal round-trip). One of the amazing consequences of this result is that, to the first order, the resolution is independent from the distance between the RADAR and the imaged scene.

At the heart of SAR processing is the fact that two stationary targets, placed at two different angles with respect to the RADAR, will have two different velocities with respect to the moving platform and these velocities will change in a well determined manner over time.

Considering Fig. 1.7, a target at along-track (azimuth) distance x from the RADAR will exhibit a Doppler shift equal to [35]:

$$f_d = \frac{2V_s \sin \theta}{\lambda} \approx \frac{2V_s x}{\lambda R} \quad (1.13)$$

where V_s is the velocity of the RADAR platform, θ is the angle between broadside and the

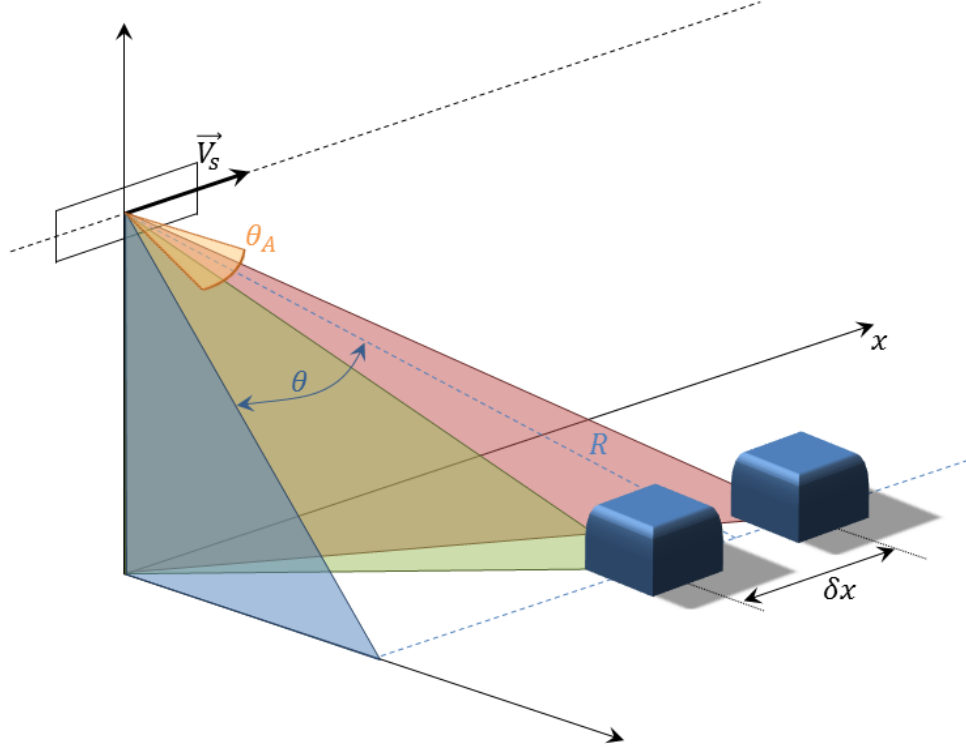


Figure 1.7: SAR azimuth resolution. (Adapted from [28] with the permission of John Wiley & Sons, Inc.)

target, and R is the slant range distance of the target. Inverting (1.13), we can observe that, as long as two objects are within the beam of the RADAR antenna (θ_A), the along-track resolution (δx) depends on the precision with which the Doppler frequency shift can be measured:

$$\delta x = \frac{\lambda R}{2V_s} \delta f_d \quad (1.14)$$

In general, the resolution to which the frequency of a periodic signal can be measured (*resolution bandwidth*) is proportional to the inverse of the period during which the signal is observed. In this specific case, the period during which the target is present within W_{θ_A} (Fig. 1.6) is $T = 1/\delta f_d = \lambda R/L_a V_s$ which, substituted in (1.14), gives the theoretical limit of $\delta x = L_a/2$.

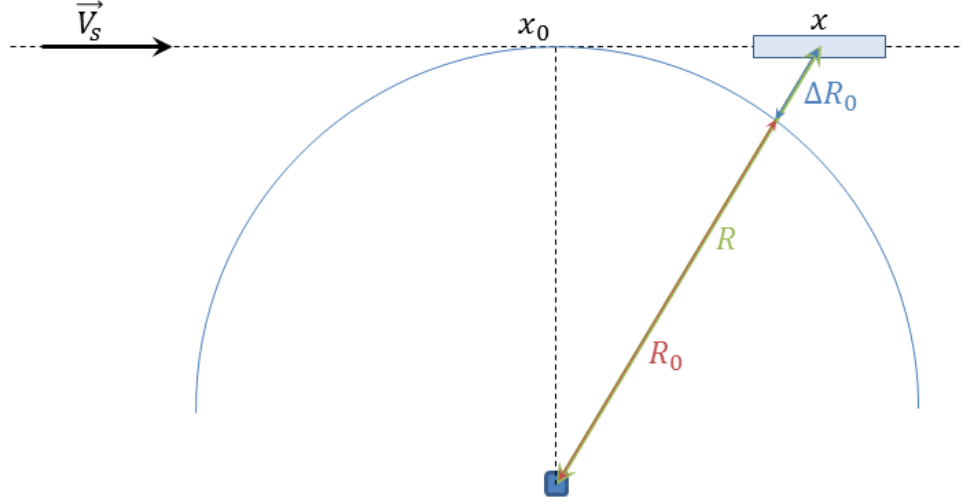


Figure 1.8: SAR focused azimuth resolution. (Adapted from [28] with the permission of John Wiley & Sons, Inc.)

This simplified derivation is not entirely correct since it is obtained under the assumption that a point target presents a constant Doppler shift for the entire time it is present within W_{θ_A} . In reality, the target will exhibit a range of frequency shifts (*Doppler bandwidth*) while traversing the antenna beam. To account for this, we can only use the previous approximation for the time when the Doppler shift can be considered constant, which reduces the available integration interval to $T = (\sqrt{\lambda R/2})/V_s$ [28] resulting, when substituted in (1.14), in what is known as the *unfocused SAR* azimuth resolution: $\delta x = \sqrt{\lambda R/2}$.

Unfocused COSMO-SkyMed would obtain an azimuth resolution of $\delta_x = 98.4$ m which is a vast improvement over the diffraction limited value but still far from the theoretical limit ($\delta x = 3$ m for an antenna size of $L_a = 5.7$ m).

The last step towards achieving the theoretical SAR resolution limit is to account for the entire Doppler shift. Following Fig. 1.8, we can describe the distance R between a SAR system, moving at constant speed V_s along a straight track, and a fixed target as $R = R_0 + \Delta R_0$ where R_0 is the minimum approach distance. We can also see that $R^2 = R_0^2 + (x - x_0)^2$ and,

in the approximation that $|x - x_0| \ll R_0$ (which is true for a typical antenna beamwidth in satellite systems), the distance between the radar and the target can be written as a Taylor series expansion assuming, without loss of generality, that $x_0 = 0$:

$$R \approx R_0 + \frac{x^2}{2R_0} + \dots \quad (1.15)$$

As the platform travels in the azimuth direction, the phase for the pulse round-trip can be written as [36]:

$$\phi(x) = -\frac{4\pi R(x)}{\lambda} \approx -\frac{4\pi}{\lambda} \left(R_0 + \frac{x^2}{2R_0} + \dots \right) \quad (1.16)$$

from which we can evaluate the change of frequency as

$$\frac{1}{2\pi} \frac{d\phi(x)}{dx} = -\frac{2}{\lambda R_0} x \quad (1.17)$$

The signal described by (1.16) and (1.17) can be interpreted as a chirp in the azimuth domain (x) with rate $K = -2/\lambda R_0$. We can then apply pulse compression and matched filter techniques: the reference function to use as correlation template is the expected Doppler frequency behavior and the "uncompressed pulse" is the length of the synthesized array: the length over which the target is illuminated by the radar (W_{θ_A}). The output of the azimuth processor, the equivalent of the compressed pulse, provides the final azimuth resolution:

$$\delta x = \frac{1}{|K|W_{\theta_A}} = \frac{\lambda R_0}{2} \frac{L_a}{\lambda R_0} = \frac{L_a}{2} \quad (1.18)$$

Ambiguities and distortions

From (1.18), it would seem that the azimuth resolution could be increased indefinitely by reducing the size of the RADAR antenna in the azimuth direction. This is obviously not the case. Size of antenna, azimuth resolution, and swath size are interrelated through the PRF (f_r).

To achieve the highest azimuth resolution, the SAR system has to sample the change in phase, related to the change in position, with a frequency higher than the Doppler bandwidth. Using (1.13) and observing Fig. 1.6, we can write this requirement as $f_{r_{min}} > 2V_s\theta_A/\lambda = 2V_s/L_a$. This limit, also known as *azimuth ambiguity*, states that, in order to provide adequate sampling for the Doppler shift, it is necessary to emit a pulse at least each time the platform moves by half the azimuth antenna size. This value regulates the minimum requirement for the PRF¹⁰.

To avoid *range ambiguities*, only one pulse can be impinging on the ground at any given time. Assuming that the spatial length of the pulse ($c\tau_p$) is much shorter than the difference between the far and near ranges ($R_F - R_N$) (Fig. 1.5), this requirement translates into $f_{r_{max}} < c/2W_s$, where W_s is the swath width projected in the direction of the incident RADAR beam. This poses a maximum limit to the PRF and links it to the swath size: the higher the PRF, the higher the azimuth resolution, the smaller the swath size.

¹⁰For an interesting analysis of different requirements leading to the same result for azimuth ambiguity see [37]

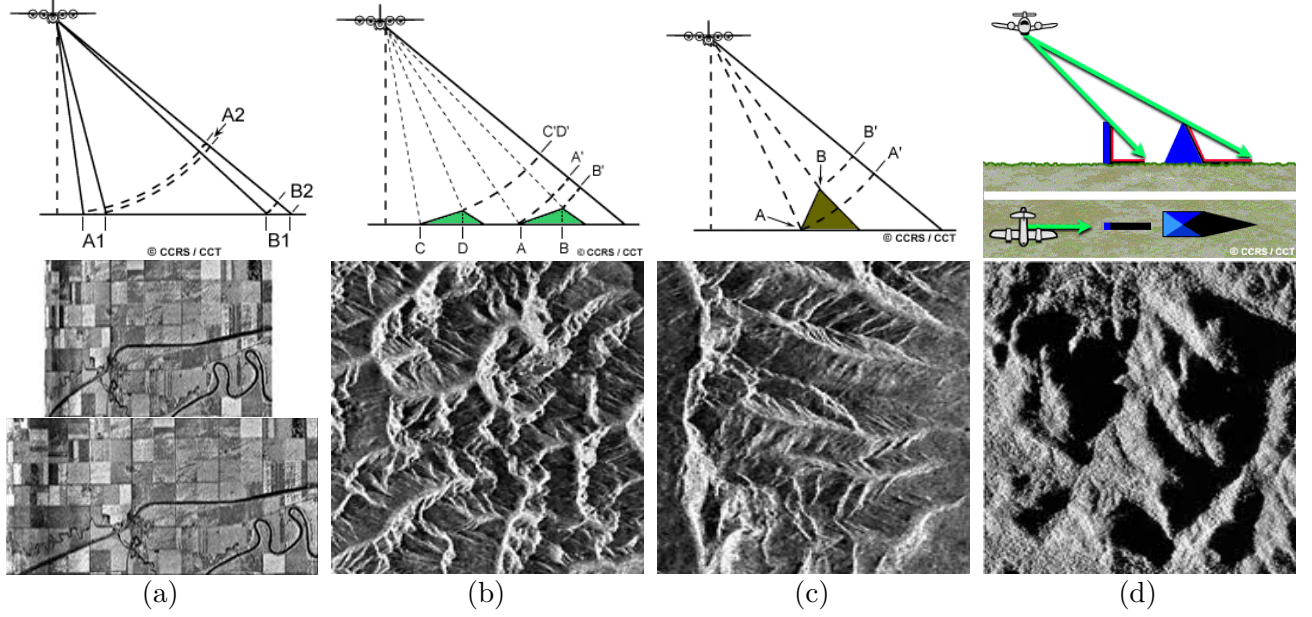


Figure 1.9: Example of possible SLAR and SAR distortions. From left to right we can see: (a) slant-range compression (the bottom part of the images shows the distorted image on top of the corrected one), (b) foreshortening, (c) layover, and (d) shadowing [38]. (With the permission of Natural Resources Canada (NRC). Reproduction has not been produced in affiliation with, or with the endorsement of, NRC.)

If we choose $f_{r_{min}} < f_r < f_{r_{max}}$, we can combine the two previously described requirements into a constraint on the RADAR antenna size by noting that $W_s = W_g \sin \eta$ and using (1.11):

$$A_a = W_a L_a > \frac{4V_s \lambda R_m \tan \eta}{c} \quad (1.19)$$

Even if f_r is chosen to remove range and azimuth ambiguities, there are still others that cannot be resolved with a single SAR measurement.

If we consider the looking geometry of a SLAR or SAR system (Fig. 1.5), it becomes apparent that several types of distortions affect the reconstructed image, potentially degrading the quality of the derived DEM. The main reason behind these distortions is the projection of the imaged scene in the azimuth/slant-range plane of the RADAR.

Every image is affected by the *slant-range scale distortion* (Fig. 1.9(a)) that causes a change of scaling as the imaged region moves from the close to the far range with respect to RADAR system. This error can be considered as a systematic error and can be removed during the geo-referencing correction since the imaging geometry is known.

More severe errors are those due to the relief displacement of the imaged scene (which is what we would like to measure in order to generate a DEM). In this case, the projection in the azimuth/slant-range plane results in distortions that cannot be corrected, and even in loss of data.

Foreshortening occurs when the RADAR images tilted structures. In the example illustrated in Fig. 1.9(b), although points C and D are in different geographical locations, due to the inclination with respect to the look angle, they are reached by the RADAR pulse simultaneously. When projected in the image plane, the corresponding points C' and D' will be considered at the same range distance. Due to the increased return signal, foreshortened regions will appear brighter than average in the RADAR image. Further inclination of imaged object toward the RADAR results in a phenomenon known as *layover* (Fig. 1.9(c)). In this case, the tip of the object (B) is imaged closer to the RADAR than the actual base. When the slope behind the tip of the imaged object is steeper than the slant angle, *shadowing* can be observed (Fig. 1.9(d)). In this extreme case, the shadowed area is not reached by RADAR signals resulting dark in the reconstructed image.

1.3.4 Interferometric SAR (InSAR)

Foreshortening and layover are both symptoms of the elevation ambiguity intrinsic in SAR systems: two targets at the same azimuth and range distance cannot be resolved. As we will see, this is the main reason why DEM cannot be generated from a single SAR image. A solution to this problem came with the development of a technique known as interferometric SAR (InSAR).

Before InSAR was developed, the only retained information, after forming the SAR image, was amplitude. The idea of considering also the measured phase came in the early 70s when RADAR interferometry was first used to map Venus surface reflectivity using two Earth-based antennae [39]. Few years later, a similar approach was implemented in a SLAR system by adding a second receiving antenna aboard an aircraft. The two antennae were positioned approximately vertically with respect to the ground and normal to the slant-range direction [31]. RADAR pulses were emitted by one of the antennae and the echoes received by both¹¹. A plot of the phase difference between the two received signals provided a graph where interference fringes could be directly correlated to the relative echo delays between the two antennae, in turn determined by the difference in elevation within the imaged scene.

InSAR is based on the fundamental idea behind interferometry: to compare the differences in the phase information between two complex signals. In each SAR image, every pixel (ground resolution element, typically a few tens of square meters in size or less) has an overall phase response that depends on the coherent superposition of several elemental responses generated by each of the scatterers present within its area, the distance of the pixel from

¹¹This observing configuration is known as *standard* mode in contrast to the *ping-pong* configuration where antennae transmit and then receive a pulse in turn. For satellite systems, most of the times, the latter is the typical configuration since the two interfering images are taken by the same satellite at two consecutive times.

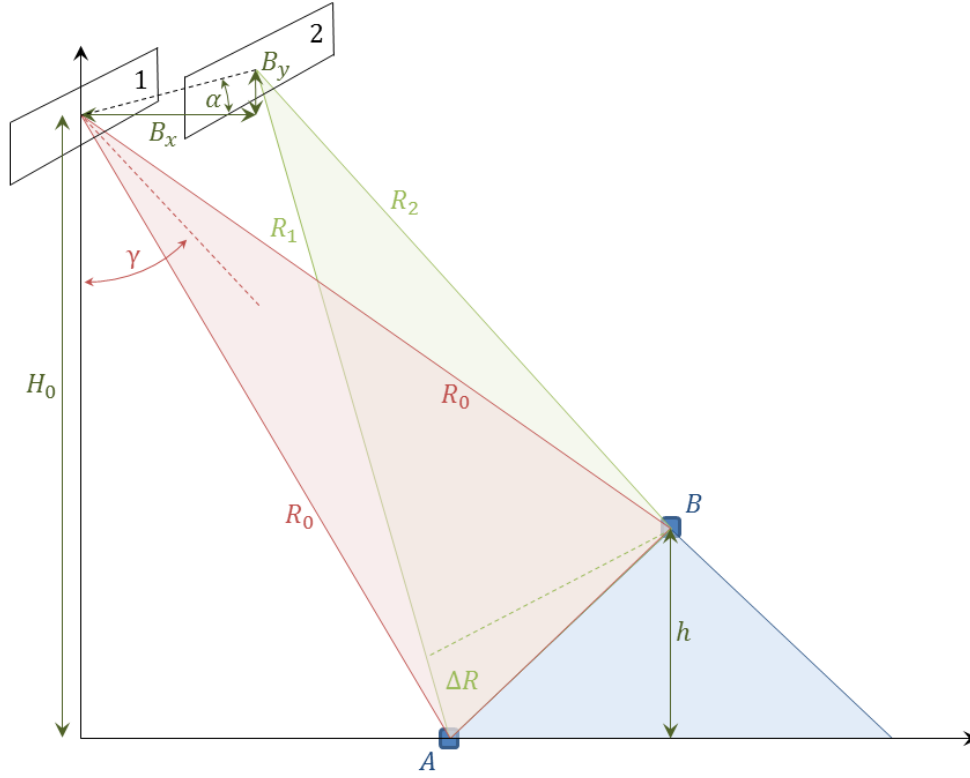


Figure 1.10: InSAR imaging geometry.

the satellite, and atmospheric phase delays. If two images are acquired within a short time interval, from different vantage points, and registered with each other, the difference in phase can be calculated and, together with the satellite attitude information relative to the two acquisitions, used to evaluate the position of each pixel with respect to the satellite [40].

With the added capability of measuring the elevation of each pixel, InSAR has become one of the most commonly used techniques in the generation of DEMs.

To better explain this imaging process, consider the simplified configuration illustrated in Fig. 1.10, where a slice is taken in the azimuth direction (considered perpendicular to the page).

Two targets (A and B) are at the same range and azimuth coordinates with respect to antenna 1, thus indistinguishable. When a second image is taken by antenna 2 from a

different viewing angle, it is possible to relate the elevation of each pixel to the difference in phase between the two SAR images using the distance between the two satellites (known as *interferometric baseline*). In particular, the elevation of a target (h) can be calculated as $h = H_0 - R_0 \cos \gamma$, where H_0 is the altitude of the RADAR system, R_0 is the range of the target measured by antenna 1, and γ is the look angle for the target that can be derived from $\Delta R = B_x \sin \gamma - B_y \cos \gamma = B \sin(\gamma - \alpha)$, where B_x and B_y (the orthogonal components of the baseline B) and α (the angle that the baseline makes with respect to an horizontal reference plane) depend on the geometry of the two observations, and ΔR is the difference in the target range as measured by the two antennae. Finally the difference in phase ($\Delta\phi$), obtained by multiplying the second SAR image by the complex conjugate of the first, can be used, with (1.16), to evaluate ΔR :

$$\Delta R = -\frac{\lambda}{4\pi} \Delta\phi \quad (1.20)$$

where the factor of 4π is used in the case of ping-pong (or repeated-track) measurements and 2π in the case of standard interferometry.

From this discussion we can see that, as the baseline increases, a given change in elevation will produce a higher change of phase. In other words, the longer the baseline, the higher the resolution in phase hence the precision in the elevation measurement. This is nicely summarized by the *altitude of ambiguity* [41]: $h_a = \lambda R \sin \gamma / 2B_n$, where B_n is the *normal baseline* (the component of the baseline perpendicular to the range direction). This is the

altitude difference that generates an interferometric phase change¹² of 2π within a range resolution element.

There is a limit to the maximum baseline that can be used since, as the normal baseline is increased, the phase change within a ground resolution element increases and the contributions from the elemental scatterers within the pixel start to cancel out reducing the overall coherence of the measured phase component. This phenomenon, known as *spatial baseline decorrelation*, is one of the potential problems affecting InSAR images [42] and can be understood by considering the fact that the phase in an interferogram is defined modulo- 2π (an integer number of RADAR wavelengths).

The ambiguity introduced by this phase wrapping has to be removed before the elevation can be calculated. This is done in a process known as *phase unwrapping*, one of the critical steps in the production of InSAR images [43], and is based on the ability to relate the phases of neighboring pixels to each other in order to reconstruct the expected continuous phase change. As the coherence is decreased, the ability to trace the phase relationship between adjacent pixels is reduced accordingly. For very large baselines, the images become completely decorrelated and interferometric reconstruction is no longer possible.

It is worth mentioning that other sources of decorrelation exist: *rotational decorrelation* caused by imaging a scene with different squint angles, and *temporal decorrelation* that occurs in repeated-track interferometry when two images are taken, not only from different vantage points, but also at different times. In this case, if the time interval between two acquisitions is significant, the pixels themselves might have changed their scattering characteristics, such

¹²The phase change is assumed to be due exclusively to the change in elevation and is evaluated after removal of the term induced by observation geometry (*orbital fringes*) and after projection of the interferogram from the azimuth-range plane to the local Earth's surface ellipsoid (*interferogram flattening*).

as in the case of seasonal changes in vegetation, or snow, coverage [42].

1.3.5 Differential InSAR (DInSAR)

In the previous few sections we have briefly described the fundamentals of RADAR and how this technology was developed into spaceborne SAR, that is now routinely used to generate high resolution complex images of large regions in a short time. We then focused on how interferometric processing added a third dimension to the imaged scenes allowing the generation of high quality DEMs.

Since the main focus of our research is to develop image processing techniques aimed at detecting changes in elevation indicative of the initial stages of potential geohazards, in the remaining sections of this chapter we are going to present recent advances in SAR image processing that lead to the development of new products that can provide measurement of elevation changes down to millimeter level accuracy.

In 1993, an article in Nature displayed for the first time the changes in the Earth's surface due to the Landers earthquake in California [44]. This focused the interest of the scientific community on the potential of a new technique that became known as differential InSAR (DInSAR).

Similarly to InSAR, *two-pass* DInSAR uses two SAR images, taken at different times (for example before and after the Landers earthquake¹³), to generate an interferogram which, as we have previously shown, contains all the information necessary to reconstruct a DEM of the region. The interferogram is then compared to a synthetic one generated from an existing

¹³This technique has also been successfully used to measure the effects of several recent devastating earthquakes: Haiti [45], L'Aquila [46], and Fukushima [47].

pre-event DEM obtained from external sources. The artificial interferogram is generated using the same satellite geometry under which the SAR images were acquired. This step is necessary to insure the removal of the interferometric fringes, due to the observation geometry, during the final differencing operation. The resulting differential interferogram contains the phase signature due to the changes in elevation [48] that, after the phase unwrapping step, can be used with (1.20) to evaluate the actual displacement.

If a DEM is not readily available or the resolution is not acceptable, the *three-pass* technique can be used. In this case, three different SAR images are acquired and one is chosen to serve as reference (*master*). After the two slaves are co-registered with the master, they are used to generate two interferograms. If any elevation change has occurred between the acquisition times, it will show in their difference [49].

These differential techniques have been used to monitor minute changes in dams, ground displacement due to underground excavations, glacier motion, coastline erosion, underground water extraction, and are routinely used to monitor the deformation due to changes in size in the calderas of volcanoes [50]. An example of this particular application is shown in Fig. 1.11.

To underline the importance of InSAR as a remote sensing technique, in 2004 JPL/NASA hosted a workshop with, among others, the goal of identifying the possible applications and benefits of satellite-based InSAR techniques. Between the several identified potentials was *topography and ground deformation*. The report evaluated at \$6.8 billion (yr.2000 USD) the yearly societal cost worldwide due to landslides and land subsidence alone (with a share of \$2.2 billion for the United States) and estimated the potential yearly saving accruing from InSAR use to be \$1.4 billion (\$430 million for the United States).

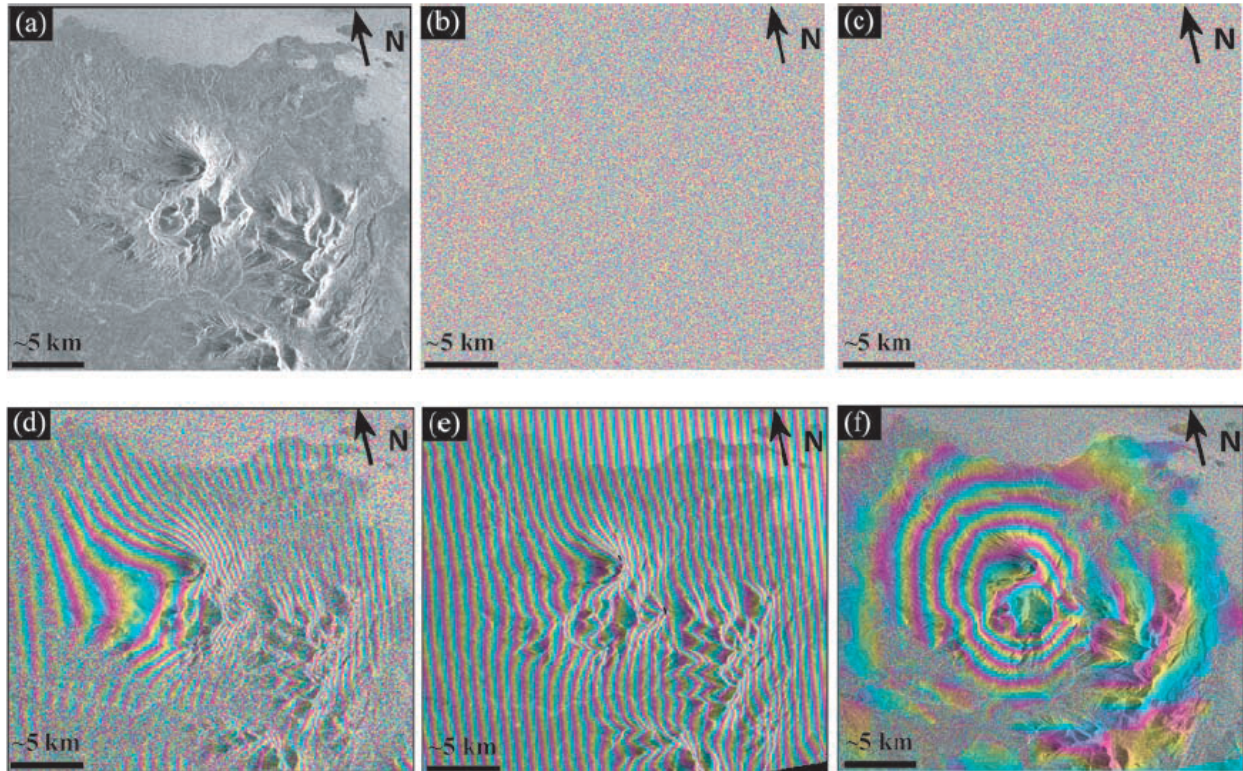


Figure 1.11: Example of two-pass DInSAR. (a) The amplitude image of an Single Look Complex (SLC) SAR image acquired on 4 October 1995 by the ERS-1 satellite over Peulik Volcano, Alaska. (b) The phase image of the SAR image acquired on 4 October 1995. (c) The phase image of an SLC SAR image acquired on 9 October 1997 by the ERS-2 satellite over Peulik Volcano, Alaska. The amplitude image is similar to that in (a) and therefore is not shown. (d) An original interferogram formed by differencing the phase values of two co-registered SAR images (b) and (c). The resulting interferogram contains fringes produced by the differing viewing geometries, topography, any atmospheric delays, and surface deformation. (e) An interferogram simulated to represent the topographic contribution in the original interferogram (d). (f) A topography-removed interferogram that was produced by subtracting the interferogram in (e) from the original interferogram in (d). The concentric pattern indicates approximately 17 cm of uplift centered on the southwest flank of Peulik Volcano, Alaska, which occurred during an aseismic inflation episode from October 1996 to September 1998. Each interferometric fringe (full-color cycle) represents 360° of phase change (or 2.83 cm of range change between the ground and the satellite). (From [50] with permission from the American Society for Photogrammetry & Remote Sensing, Bethesda, Maryland, www.asprs.org).

1.3.6 Permanent Scatterer InSAR (PSInSAR) and SqueeSAR

One of the major problem with interferometric images is decorrelation [42]. As we have already discussed, imaging baselines and acquisition intervals can affect spatial and temporal correlation. On the other hand, these parameters are often controllable by either accurately

selecting existing images from a historical database, or carefully defining the satellite attitude parameters for the observation run.

One unpredictable source of phase error is due to atmospheric inhomogeneities [51], in particular, to the turbulent motion of water vapor masses [52]. These errors are typically characterized by slow varying phase delays, known as *atmospheric phase screen (APS)*, superimposed on each SAR image [19] [53].

The InSAR image construction process, described in the previous section, can be extended to make use of many SAR images of a given scene with a technique known as *multi-pass* interferometry [54]. This approach begins with what became known as *SAR stack*, a large number of SAR images of the same scene acquired over time, processed to produced an *InSAR stack*. The processing of several images yields a dramatic reduction in the *speckle*¹⁴ and not only provides a measurement of the overall change occurred during the measurement, but also the series of partial changes between stack elements. Unfortunately, the introduction of atmospheric delays can result in severe long spatial wavelength distortions within the stack corrupting the final deformation measurements.

APS distortions can be quite large and are the main limitation to the achievement of the high resolution differential measurement required to characterize ground motion and identify conditions precursory to the development of geohazards. In order to limit these distortions, a set of new techniques, generally know as persistent scatterer interferometry (PSI or PSInSAR), were developed [55] where the density of measurement points is traded for precision.

¹⁴Speckle (noise) occurs with coherent radiation when the smallest resolution element is much larger than the wavelength of the radiation being scattered from the element. In this case, the measured signal is the superposition of all the scattering events occurring within the element itself and the final amplitude can vary from zero to their sum giving rise to the "granular" effect observed in SAR images.

The fundamental idea takes advantage of the fact that, while APS distortions show a strong spatial correlation within a single SAR image, they are not correlated from one image to the next in the stack whereas targets subject to motion show strong correlation in time (across the stack).

In general, the local *coherence* between two SAR images (I_1 and I_2) is defined as the absolute value of the complex cross-correlation coefficient ($\gamma(\mathbf{x})$) estimated over a small azimuth-range window (W_x) around each pixel (\mathbf{x}):

$$\gamma(\mathbf{x}) = \left| \frac{\langle I_1 I_2^* \rangle_{W_x}}{\sqrt{\langle |I_1|^2 \rangle_{W_x} \langle |I_2|^2 \rangle_{W_x}}} \right| \quad (1.21)$$

where the angled brackets represent averaging over window W_x defined at the pixel location \mathbf{x} . This is evaluated after the known phase components due to observation geometry and known terrain elevation are estimated and removed [40].

A coherence close to unity can be related to a low phase dispersion and is indicative of better traceability of the pixel phase behavior over time that, in turn, allows to discriminate between the phase contributions due to the APS and those due to the actual displacement.

The extraction of permanent/persistent scatterers (PS) begins with the selection of a stack of SAR images of a given scene and their registration to a single master chosen from the stack itself. The master image is identified by maximizing the expected coherence of the final InSAR stack [56]. During the registration process, the single SAR images are also radiometrically normalized to allow direct comparison of the amplitude returns. This is required because registration is followed by an amplitude analysis aimed at the selection of PS candidates: pixels in the SAR images for which the ratio between the mean μ_A and the standard deviation

σ_A of the time-series of the amplitude values is large. These pixels correspond to locations of high coherence and the ratio value, known as the *amplitude dispersion index* (D_A) is an indicator of the phase stability: $\sigma_\phi \cong \sigma_A/\mu_A = D_A$ [55].

The SAR stack is then corrected for phase components, due to the local topography, using an existing DEM, or one generated by the techniques previously described [53]. Each of the resulting images is then used, in conjunction with the master, to compute an InSAR stack with the phase difference of the i -th element of the stack ($\phi^{(i)}$) given, for each pixel, by [41]:

$$\phi^{(i)} = \frac{4\pi}{\lambda} r_i + \alpha^{(i)} + n + \epsilon_i \quad (1.22)$$

where r_i is the possible motion of the target in the range direction, $\alpha^{(i)}$ is the APS phase contribution, n is a general decorrelation white noise term, and ϵ_i is the phase error due to potential inaccuracies in the DEM evaluation and depends linearly on the normal component of the baseline ($B_n^{(i)}$) used to generate the i -th entry in the stack.

The first term on the right side of equation (1.22) is dependent on the actual pixel motion and can be written as:

$$\frac{4\pi}{\lambda} r_i = \frac{4\pi}{\lambda} v_i T_i + \mu \quad (1.23)$$

where v_i is the unknown component of velocity of the target in the range direction for the i -th frame¹⁵, T_i is the time interval between the acquisition of the master and the i -th slave image in the stack, and μ is a term due to possible motion not considered by the model

¹⁵Although the assumption of linearity for the target motion simplifies the problem, it is not required to converge to a solution and approaches allowing for different motion models have been developed [57].

(linear in this case). Substituting in (1.22) we obtain the general formulation:

$$\phi^{(i)} = \frac{4\pi}{\lambda} v_i T_i + \mu + \alpha^{(i)} + n + \epsilon_i \quad (1.24)$$

The approach taken by PSInSAR is to evaluate the different phase components exclusively on the grid generated by the PS candidates [58]. With this approach, for each entry i in the stack, we can write the phase difference between two candidate PS (l and m) as:

$$\Delta\phi_{lm}^{(i)} = \frac{4\pi}{\lambda} T_i \Delta v_{lm} + \Delta\epsilon_{lm} + \mu + \alpha^{(i)} + n \quad (1.25)$$

where Δv_{lm} is the mean relative velocity, ϵ_{lm} is the relative elevation error proportional to the orthogonal component of the baseline for the i -th interferogram in the stack ($\epsilon_{lm} \propto B_n^{(i)}$) [59], and $\Delta w_{lm}^{(i)} = \mu + \alpha^{(i)} + n$ is the phase residue.

This system of N equations in two unknowns (Δv_{lm} and $\Delta\epsilon_{lm}$) is not linear because of the phase wrapping that occurs in InSAR images.

If we consider that $\Delta\epsilon_{lm}$ can be estimated from the slope of $\Delta\phi_{lm}^{(i)}$ as a function of the normal component of the baseline $B_n^{(i)}$ and Δv_{lm} from the slope of $\Delta\phi_{lm}^{(i)}$ as a function of time T_i , then the solution can be found as a maximization problem for the overall coherence (γ_{lm}), limited to those PS candidates which are close enough that the phase error introduced

by their phase residue is less than π :

$$|\gamma_{lm}| = \left| \frac{1}{N} \sum_{i=1}^N \exp(j\Delta w_{lm}^{(i)}) \right| \quad (1.26)$$

$$\Delta w_{lm}^{(i)} = \Delta \phi_{lm}^{(i)} - \frac{4\pi}{\lambda} T_i \Delta v_{lm} - \Delta \epsilon_{lm} = \mu + \alpha^{(i)} + n \quad (1.27)$$

$$|\Delta w_{lm}^{(i)}| < \pi \quad (1.28)$$

Once the estimation has been carried out for all the possible combination of PS candidates satisfying the distance constraint, the APS is evaluated at the location of the PS candidates as:

$$\alpha^{(i)} = \Delta \phi_{lm}^{(i)} - \frac{4\pi}{\lambda} T_i \Delta v_{lm} + \Delta \epsilon_{lm} \quad (1.29)$$

interpolated, and removed from each SAR image in the stack [60].

This process can be repeated iteratively: select PS candidates based on their coherence values, eliminate the atmospheric phase contributions, and repeat until no more PSs are detected. As a result, if the SAR stack contains 45 or more images, it is possible to obtain a set of discrete permanent scatterer locations where the DEM can be evaluated with sub-meter accuracy and the history of change in elevation measured with millimetric precision for each time interval available in the original stack (see section **B.1: Precision of data**).

PSs identify pixels that display strong coherence and high phase stability under a wide range of imaging angles and baselines over the whole period during which they are observed. These pixels are often characterized by the presence of a strong sub-pixel point-scatterers providing strong retro-reflection (back-scattering) properties almost unaffected by spatial and temporal decorrelation. These types of targets are often associated with man-made or

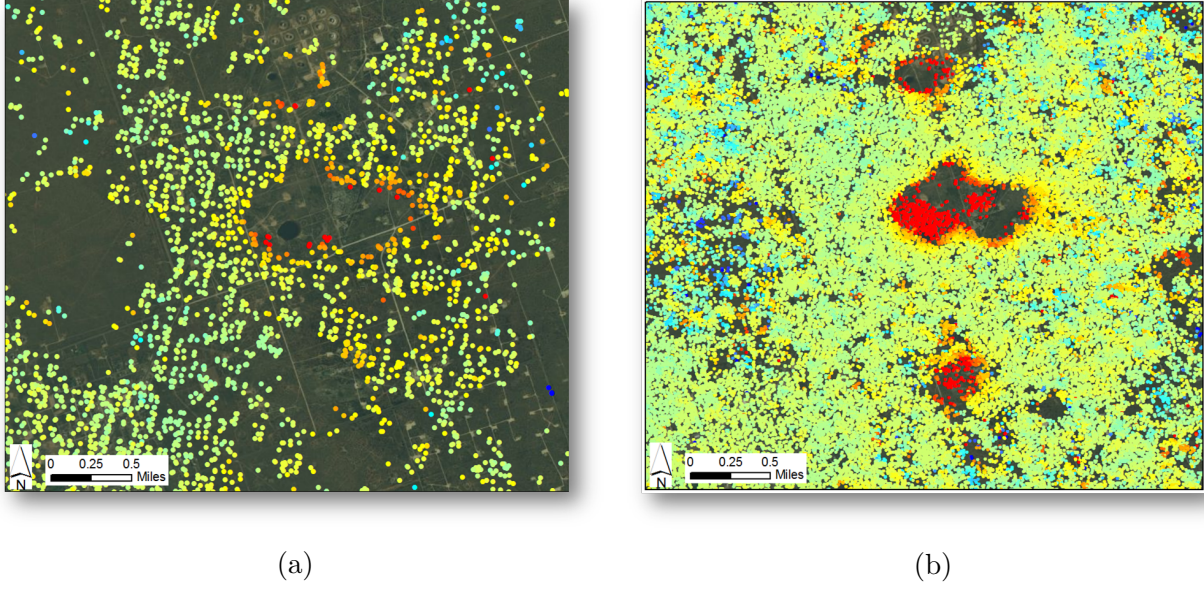


Figure 1.12: Comparison between (a) PSInSAR and (b) SqueeSAR when applied to a low populated area (near Winkler, Texas) (see section 2.2.2: *The Winkler sinkholes (Wink Sinks)*). The two images are showing the average scatterer velocity obtained by processing 22 images acquired between June 3, 1992 and February 21, 1998 by the ERS satellites 1 and 2. (Courtesy of TRE-Canada.)

exposed rock structures.

In rural areas, where most miles of the road network that can be affected by geophysical events are located, PSs are quite sparse. The SqueeSAR technique was developed to increase the density of scatterers for some of these regions [61].

In SqueeSAR, the concept of coherence is extended to the spatial domain: pixels, none showing strong enough individual temporal coherence to be classified as PS but displaying strong coherent behavior both spatially and temporally as an ensemble, are grouped in regions (with an associated *effective area*) and identified by a single scatterer placed in the center of mass of the region and called *distributed scatterer (DS)*.

The basic process by which SqueeSAR includes and extends PSInSAR is by searching the spatiotemporal neighborhood of each pixel within the SAR stack for a statistically homogeneous region. If we define $\mathbf{r}(x) = \{r_i(x)\}_{i=1,\dots,N}$ as the vector of the complex reflectivity

of a pixel x , measured for each image in the SAR stack, then two pixels x_1 and x_2 are defined statistically homogeneous if the null hypothesis that $\mathbf{r}(x_1)$ and $\mathbf{r}(x_2)$ are drawn from the same probability distribution function cannot be disproved to a certain level of significance [61]. This approach identifies pixels containing a permanent scatterer as individual one-dimensional temporal vectors whereas large areas of pixels exhibiting similar spatiotemporal reflectivity characteristics (the distributed scatterers) are grouped together. DSs are then combined into a single scatterer and, together with PSs, are processed by the PSInSAR algorithm described above.

When a stack of SAR images is processed using SqueeSAR, the resulting dataset is a non-uniform spatiotemporal point cloud dataset. In Fig. 1.12 we show an example of one time slice of such dataset where (a) illustrates the result of processing the SAR stack with PSInSAR whereas (b) shows the result of SqueeSAR. This dataset was generated over a scarcely populated area (an oil-field from early 1900) near Winkler, Texas (see section 2.2.2: *The Winkler sinkholes (Wink Sinks)*). These types of arid regions are ideal for the application of the SqueeSAR technique.

1.4 Summary

Reports by federal and state organizations and news broadcasts are a constant reminder of how expensive the outcomes of subsidence and landslides can be. The main goal of our research is to provide early detection of such events with a specific focus towards the transportation infrastructure.

To understand how remote sensing technology can be implemented to support the goal of

our research, we described how RADAR has evolved during the years to become one of the main tools capable of providing high resolution measurements of Earth's surface from spaceborne SAR instrumentation. We illustrated how interferometric techniques helped in the removal of inherent ambiguities in the SAR images, allowing measurement of the surface elevation and the production of high quality DEMs. We then examined differential interferometric techniques developed to measure changes in elevation over time and how unpredictable atmospheric phase contributions limited the achievable displacement measurement resolution. Finally we described the latest differential interferometric products, PSInSAR and SqueeSAR, which, by focusing the processing of SAR stacks on a subset of strong coherent scatterers, can remove atmospheric contributions and provide displacement time series for each of these points with resolutions down to the millimeter level. The dataset produced by SqueeSAR is a non-uniformly sampled, spatiotemporal point cloud and is the starting point of our analysis.

In the next chapter we will identify the phenomenon we are trying to detect and the methodology we used to derive a model that describes its spatiotemporal behavior. In the following chapters we will then delineate, discuss and compare the techniques we developed to detect such a model within the point cloud datasets.

Chapter 2

Modeling

2.1 Introduction

In the previous chapter we presented a brief history of how RADAR systems became the preferred technology to produce high resolution digital elevation models. We also described the newest interferometric techniques, based on the analysis of the phase stability of scatterers, and showed how these approaches can provide historical series of ground displacements with accuracies in the millimeter range. The goal of our work is to develop techniques to detect the early developmental stages of geohazards, potentially dangerous for the transportation infrastructure, within the point cloud datasets produced by these interferometric techniques.

In order to detect a feature within a signal, it is necessary to have a template, or model, of the of feature itself. In 1977 Tukey defined as *exploratory data analysis* a collection of techniques used to gain insight into data, uncover underlying structures, generate hypothesis, detect anomalies, and identify important measurements [62]. In our case, the data consist of a spatiotemporal point cloud dataset and, as underlying structure, we concentrated on

the detection and characterization of sinkholes because, differently from landslides and other similar geohazards, where the location of the event is identifiable (slopes), sinkholes occurrence is not predictable and requires both localization and characterization to be performed simultaneously.

In this chapter we will briefly describe the reasons behind the subsidence phenomenon and, in particular, we will analyze how sinkholes develop. We will then show how, starting from a SqueeSAR dataset, providing ground truth information about sinkhole morphology (see section 2.2.2: *The Winkler sinkholes (Wink Sinks)*), we developed a spatiotemporal model to serve as template for the detection techniques that we will illustrate in the following chapters. We will also provide a verification of the model by showing the results of four simulations of the development of different types of sinkholes and how the proposed model responds to each of them.

2.2 Subsidence and sinkholes

Due to their widespread occurrence, subsidence phenomena in karst terrain have been largely analyzed by the scientific community [63] [64]. Most of the existing approaches are aimed at the cataloging and characterization of already formed sinkholes [65] [66] [67] or the evaluation of risk maps derived from a probabilistic analysis of the location and frequency of existing occurrences [68] [69] and do not provide an evolutionary model for early sinkhole development. Some works have analyzed single sinkholes by evaluating the probability of collapse of an existing feature based on its specific characteristics [70], or with the development of few models aimed at the evaluation of the final overall maximum vertical displacement [71] or at

the forecast of the potential final size in karst terrain [72].

2.2.1 Morphology

The term *karst* refers to a specific topography where the carbonate bedrock has undergone a dissolution process leaving a porous landscape strongly coupled with an efficient underground drainage system. These types of terrains are often associated with well defined structures such as sinkholes, dry valleys, and cave systems with associated springs.

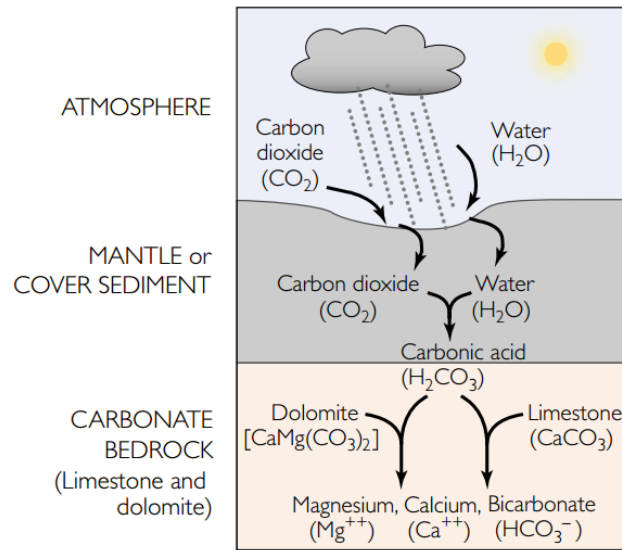
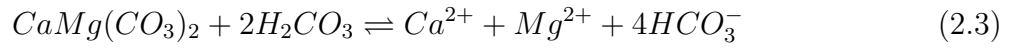


Figure 2.1: Dissolution process for carbonate rocks [73]. (Courtesy of the U.S. Geological Survey)

Carbonate rocks are defined as rocks containing more than 50% by weight of carbonate minerals (although typical value are often found to be above 90% [64]) which is a general term referring to any mineral containing the *carbonate ion* CO_3^{2-} . Most of carbonate rocks present in karst terrains are made of *calcite* ($CaCO_3$) (also known as *limestone*) and *dolomite* ($CaMg(CO_3)_2$), with calcite being the major component.

The main process behind the formation of karst terrain is the dissolution of the carbonate bedrock. As rain falls through the atmosphere and the soil, water (H_2O) combines with the soluble carbon dioxide (CO_2) creating *carbonic acid* (H_2CO_3). Carbonic acid flows through the cover sediment and reaches the carbonate bedrock where it interacts with the limestone and dolomite dissolving them into Magnesium and Calcium ions (Mg^{2+} and Ca^{2+}) and

bicarbonate (HCO_3^-) leaving behind the typical karst features (Fig. 2.1):



One interesting point to note is that reaction (2.2) and (2.3) are in equilibrium with (2.1). Because of this, an increase of available CO_2 in the atmosphere causes an increase of carbonic acid thus accelerating the dissolution of the carbonate bedrock.

Although every karst terrain possesses individual characteristics, engineering and geology literature contains a large number of attempts to characterize sinkholes based on the formation mechanism and the material which fails and subsides [63][64]. The six generally recognized categories (dissolution, collapse, caprock, dropout, suffosion, and buried) can be further grouped in three major types.

Dissolution sinkholes

In the case of dissolution sinkholes (Fig. 2.2), the rain water, rendered slightly acidic by the carbonic acid, percolates through the naturally occurring porosity of the carbonate soil slowly eroding it. As small depressions starts developing on the surface of the bedrock, erosion becomes more accentuated where water accumulates further increasing the dissolution process.

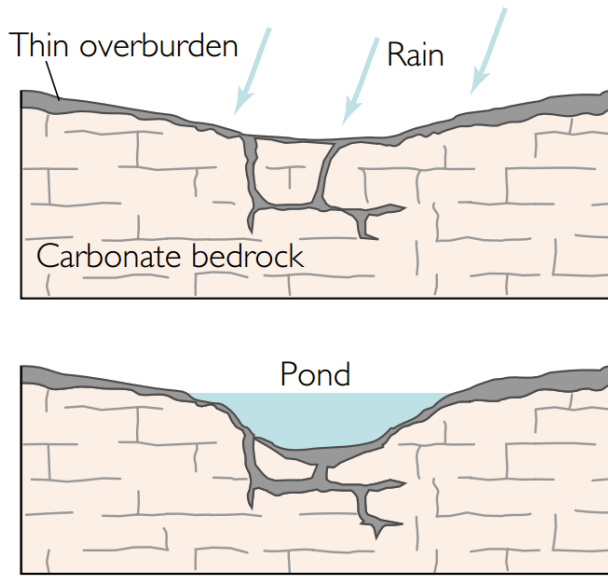


Figure 2.2: Dissolution sinkholes [73]. (Courtesy of the U.S. Geological Survey)

If present, the thin overlying soil will mirror the underlying bedrock surface. In other dissolution typologies, the joint of discontinuities in the bedrock will be widened due to the progressive dissolution creating a secondary porosity that can be eventually filled by sediment. This process typically results in the formation of gentle rolling hills and depressions possibly punctuated by rock pinacles due to local tectonic folding [74].

Cover-subsidence sinkholes

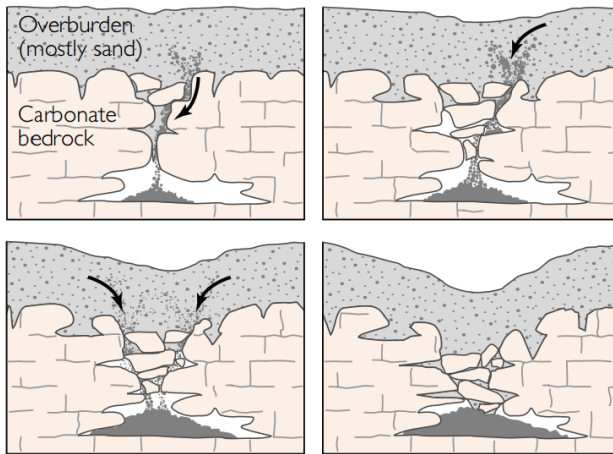


Figure 2.3: Cover-subsidence sinkholes [73]. (Courtesy of the U.S. Geological Survey)

Cover-subsidence sinkholes (Fig. 2.3) are similar to dissolution sinkholes with the difference that the layer on top of the carbonate bedrock is much thicker and mostly composed of low cohesion granular material with high sand content. The acidic water percolates through the top soil reaching the limestone layer where the dissolution process starts. As crevasses start forming, the

material from the top layers settles in the newly created spaces. As the process continues, more and more material settles originating a growing surface depression.

The morphology of cover-subsidence and dissolution sinkholes is generally slow and progressive and sudden collapse not very common [74].

Cover-collapse sinkholes

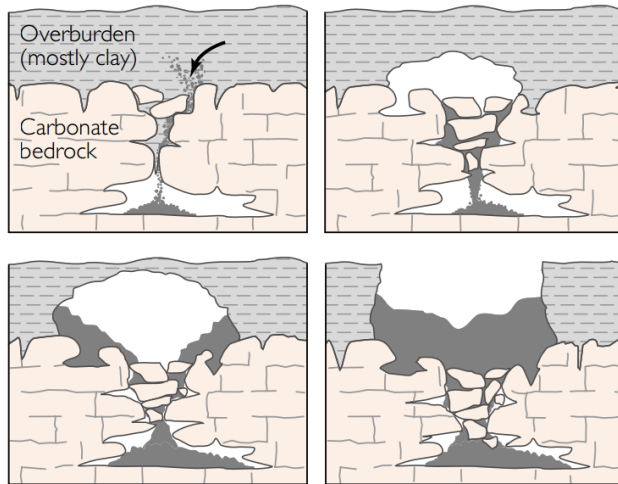


Figure 2.4: Cover-collapse sinkholes [73]. (Courtesy of the U.S. Geological Survey)

One of the most dangerous type of sinkhole is the cover-collapse (Fig. 2.4). With this morphology, a thick layer of highly cohesive material lays on top of a limestone bedrock. With time, acidic water infiltration starts to dissolve the carbonate rock and sediments from the cohesive layer slowly fill the new space. As this process continues, the top layer gets eroded forming an thinning arching

structure. As the cavity slowly progresses upwards the top layer narrows to the point where the strain due to the weight of the ceiling itself is such that the tensile stress is reached. Collapse in such circumstances can be immediate as we will see in the case of the Wink Sinks [74].

2.2.2 The Winkler sinkholes (Wink Sinks)

Although subsidence is a natural occurring phenomenon, there are several human activities that over the years have caused widespread events. Studies have shown that, for example,

inelastic compaction¹ has irreversibly altered the aquifer systems of 12 states causing large subsidence events such as the largest human alteration of Earth’s surface that occurred in San Joaquin Valley, California where, between 1925 and 1975, more than 5,200 square miles witnessed a subsidence in excess of 30 cm with a peak of 9 m near Mendota, CA [75]. More impressive, due to their catastrophic appearance, are ”man-made” cover-collapse sinkholes [76].

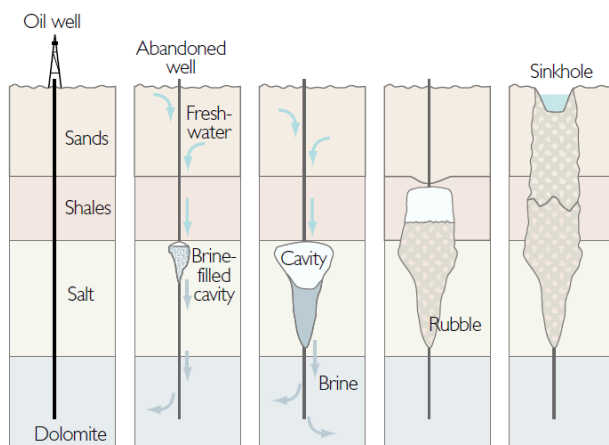


Figure 2.5: Upward propagating cover-collapse sinkholes [73]. (Courtesy of the U.S. Geological Survey)

The Wink Sinks refer to a group of sinkholes that developed in an area of Winkler County closed to Wink, Texas as final stage of an upward migrating dissolution cavity that originally developed more than 400 m underground [77]. The development of the cavity was probably caused by an imperfect plugging, applied to an oil well abandoned in the 1920s, that allowed for the circulation of wa-

ter along the well shaft. As the water reached soluble regions, the dissolution process started creating the original cavity.

Over time this developed according to the process previously described effectively creating an upward migrating cavity (Fig. 2.5) that resulted in the collapse of one of the sinkholes (Wink 1) in June 1980 and the second (Wink 2) in May 2002. Two more regions, Wink 4 and Wink 3, located respectively next to and south of Wink 2, are still developing.

¹Inelastic compaction occurs when, due to prolonged water extraction, the weight of the overlying rocks is no longer balanced by the porous-fluid pressure. Once this hydrostatic equilibrium is altered, the ground is compacted, effectively removing available volume from the aquifer. This process is irreversible [73].

The relevance of this 55 km² area to our project is due to the fact that one of the training dataset, that we purchased from TRE Canada to start the developments of the detection algorithm (see chapter 3: *Detection*), was obtained by imaging this area for a period of about 69 months between June 1992 and February 1998: after the collapse of Wink 1 but before the collapse of Wink 2 (see section B.2: *Wink Sinks*). The acquisition campaign conducted by the ERS satellites (see section A.1: *European remote sensing satellites (ERS)*) provided a stack of 22 SAR images that were processed with

the SqueeSAR algorithm resulting in a sparse, non-uniformly sampled point cloud dataset consisting of 93,513 points between permanent and distributed scatterers.

In Fig. 2.6 we show this point cloud dataset overlapped to the Wink Sink area. Each point represents a scatterer and is colored according to the average subsidence velocity (the point towards the bottom right corner, identified by the light green marker, is the reference point). Within this imaged area the four different regions interested by the subsidence phenomena are clearly visible.

Due to the timing between the acquisition and the collapse of Wink 2, this data is extremely important for our research since it can be considered ground truth and used to

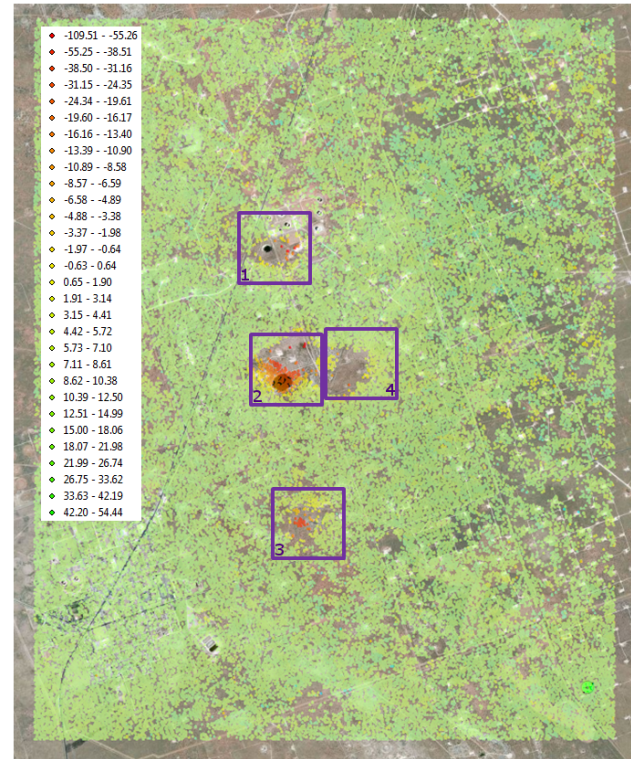


Figure 2.6: Pointcloud representing the average subsidence velocity overlapped to an aerial of the interested area. Highlighted are the areas interested by the four subsidence phenomena. The reference point is identified in light green towards the bottom right corner.

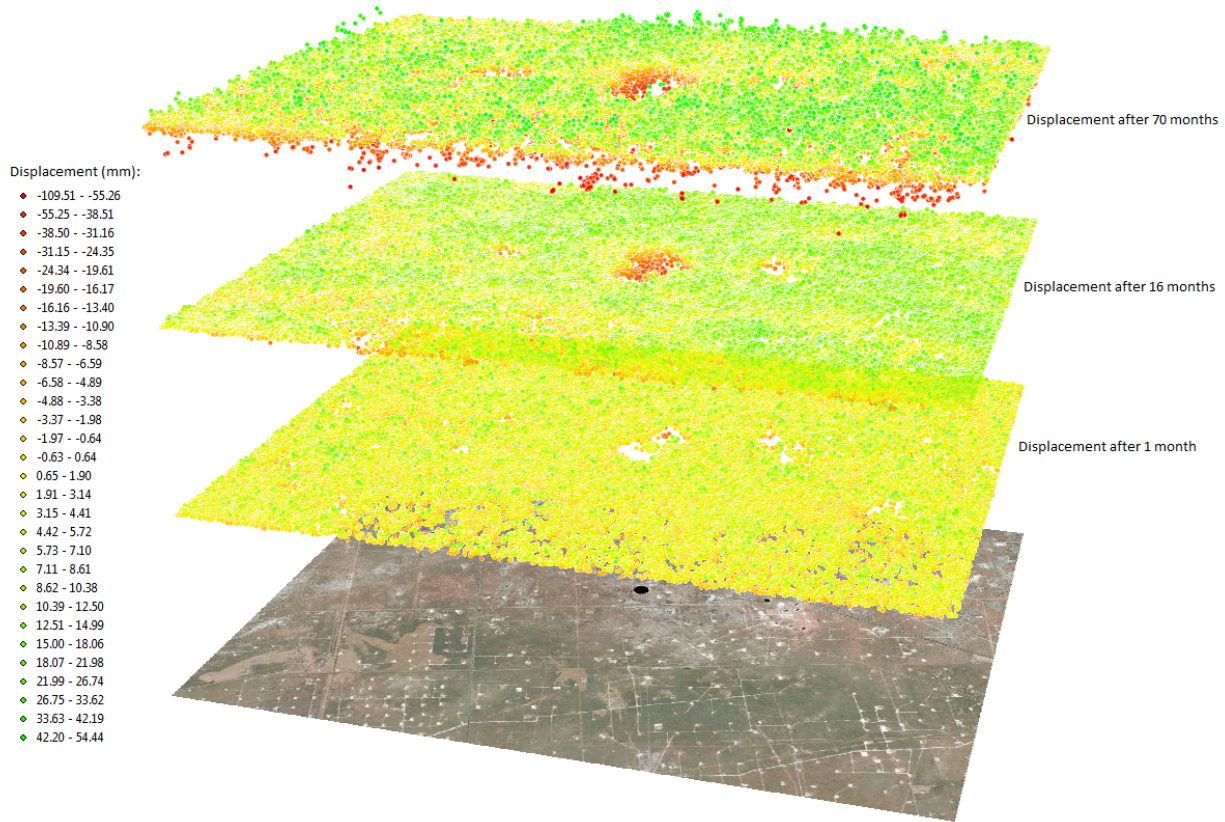


Figure 2.7: Three temporal samples of the scatterers displacement over the Wink Sinks region. The samples provide the measured displacement after 1, 16 and 70 months past the first acquisition. Visible in the aerial layer are the actual collapsed region of Wink 1 (left) and Wink 2 (center).

verify the effectiveness of the algorithms we developed.

2.3 The model

All the described sinkhole types have the common feature that the region affected by the subsidence process is much larger than the region where a visible sinkhole eventually develops. Even in the case of cover-collapse, usually forecasting evidence such as fissures, cracks and distortions typically appears before and on a larger area than the one eventually affected by the collapsing of the cave ceiling (Fig. 2.7). Because of this observation, we decided to start

analyzing the spatiotemporal behavior of the Wink Sinks data over larger regions surrounding the actual sinkholes.

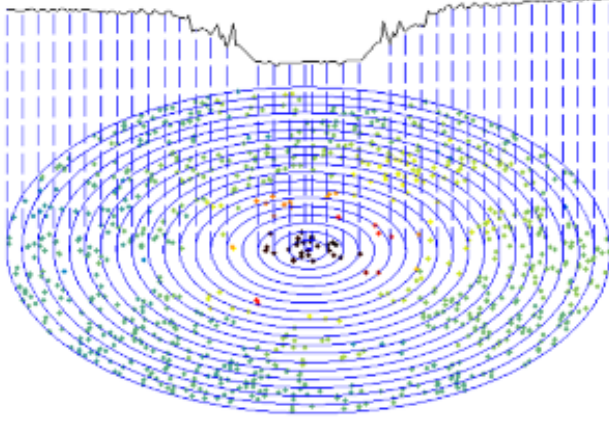


Figure 2.8: Construction of displacement profile (top) based on data relative to a circular area surrounding Wink 2 for a single time frame (bottom). Dark markers represent scatterer affected by the largest displacement while blue/green those affected by the smallest. (From [78] ©2013 Taylor and Francis)

We defined a circular area around each of the regions presenting strong subsidence features. The center was selected as the interpolated maximum subsidence based on the average velocity field. Within these regions, we defined a set of concentric annuli (5 m apart) that were used to bin the displacement data. For each one of the rings, a single average displacement value was calculated and used to construct the displacement profile (Fig. 2.8).

Preliminary analysis of the observed profile behavior for the Wink Sinks, especially Wink 2 and Wink 3 where the density of the scatterers was higher, suggested a strong Gaussian behavior. This is consistent with the transverse deformations observed in underground tunneling operations, where the vertical displacement of soft ground over the subsurface void is found to be well approximated by $d(x) = d_0 \exp(-x^2/2x_i^2)$, where x is the transverse distance from the tunnel centerline, d_0 is the maximum settlement, and $x_i = Kz$, where z is the depth of the tunnel, and K a constant that can be assumed to be 0.5 for cohesive and 0.25 for granular soils [79] [80] [81].

We repeated the binning of the data for each available time interval and plotted the results in temporal order (Fig. 2.9). The profiles seemed to preserve the Gaussian behavior over time

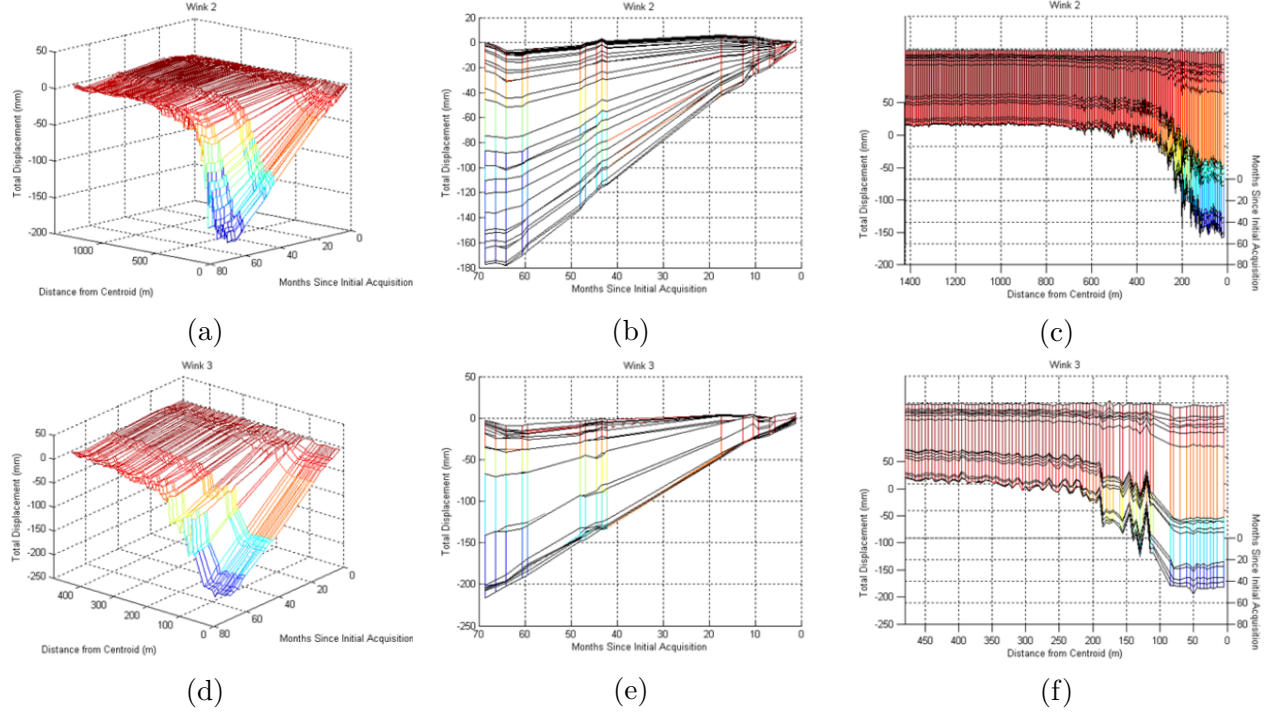


Figure 2.9: Different view of the profile stacks for Wink 2 (top) and Wink 3 (bottom): (a, d) spatiotemporal stack; (b, e) temporal evolution; (c, f) spatial evolution. (From [78] ©2013 Taylor and Francis)

and the changes within each of the bins (representing the concentric average displacement) showed strong linear amplitude behavior.

The general formulation for a two-dimensional elliptical Gaussian function, where the parameters might be time-dependent, is given by

$$g_{\mathbf{p}}(x, y, t) = A_t \exp \left\{ - \left[a_t(x - x_0)^2 + 2b_t(x - x_0)(y - y_0) + c_t(y - y_0)^2 \right] \right\} \quad (2.4)$$

where \mathbf{p} is the parameter vector defined as $\mathbf{p} = [A_t, (a_t, b_t, c_t), (x_0, y_0)]$, A_t is the time changing amplitude of the Gaussian, (x_0, y_0) its center (considered fixed in time), and (a_t, b_t, c_t) define time-varying orientation and ratio between the two elliptical axes². The

²For (2.4) to represent a two-dimensional Gaussian, the matrix defined by $\begin{bmatrix} a_t & b_t \\ b_t & c_t \end{bmatrix}$ has to be positive-definite.

latter can be defined as a function of the two main standard deviations ($\sigma_{x,t}^2$ and $\sigma_{y,t}^2$) and the counterclockwise angle with respect to the x axis (θ_t) by setting

$$a_t = \frac{\cos^2(\theta_t)}{2\sigma_{x,t}^2} + \frac{\sin^2(\theta_t)}{2\sigma_{y,t}^2} \quad (2.5)$$

$$b_t = \frac{\sin(2\theta_t)}{4\sigma_{x,t}^2} - \frac{\sin(2\theta_t)}{4\sigma_{y,t}^2} \quad (2.6)$$

$$c_t = \frac{\sin^2(\theta_t)}{2\sigma_{x,t}^2} + \frac{\cos^2(\theta_t)}{2\sigma_{y,t}^2} \quad (2.7)$$

The new parameter vector can then be expressed as $\mathbf{p} = [A_t, (\sigma_{x,t}, \sigma_{y,t}), \theta_t, (x_0, y_0)]$.

Considering the circular symmetry induced by the profile generation approach, equation (2.4) is greatly simplified and reduces to:

$$g_{\mathbf{p}}(x, y, t) = A_t \exp \left\{ - \left[(x - x_0)^2 + (y - y_0)^2 \right] / 2\sigma_t^2 \right\} \quad (2.8)$$

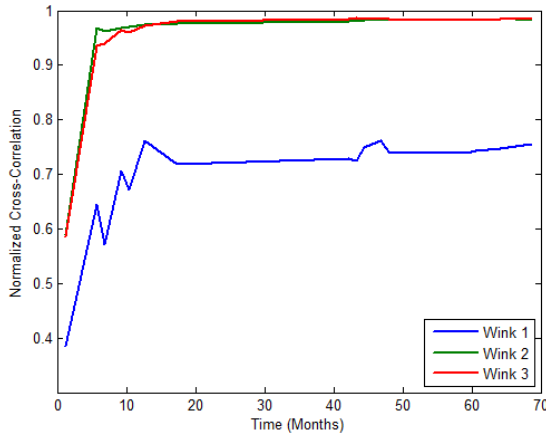


Figure 2.10: Normalized Cross-Correlation between displacement profiles and fitted Gaussians for Wink 1 (blue), Wink 2 (green), and Wink 3 (red). (From [78] ©2013 Taylor and Francis)

To derive a mathematical model of the displacement evolution, we fitted each profile with the projection of (2.8) over the plane of symmetry and evaluated A_t and σ_t . To substantiate the validity of this approach, we calculated the normalized cross-correlation between the displacement profiles and their corresponding Gaussian fit for each time t (Fig. 2.10).

The correlation coefficient for the first few

time frames ($t < 7$ months) did not show strong correlation due to the lack of significant changes in the subsidence however, as the time progressed and changes became more evident, a strong correlation developed with coefficients approaching unity for both Wink 2 and Wink 3. The reduced fit quality for Wink 1 is probably due to the fact that it had already collapsed and filled up with water before the SAR data was taken causing the lack of reliable scatterers within the central collapsed region; nevertheless, the surrounding data points still displayed a marked downward trend resulting in cross-correlation coefficients around 0.75.

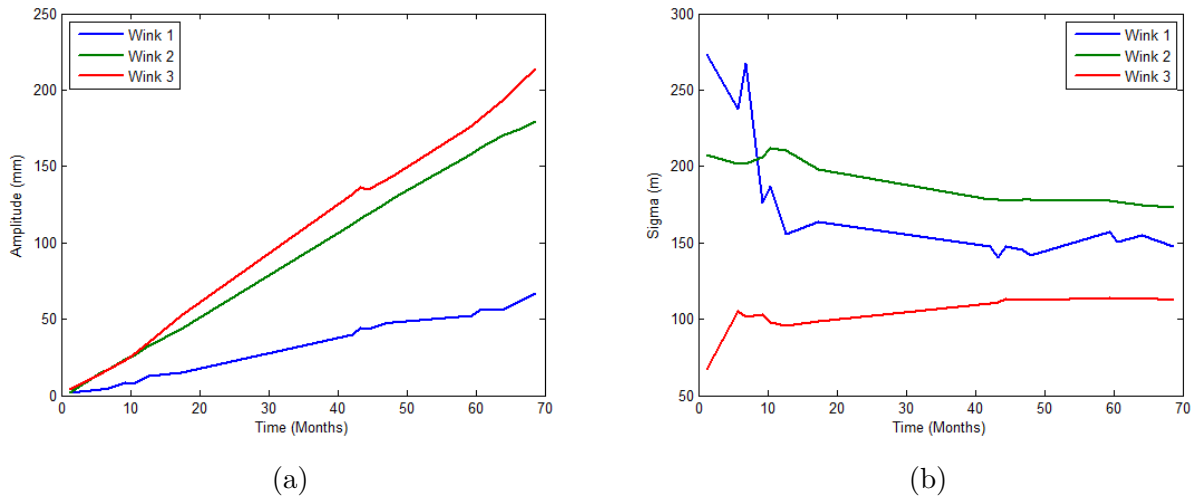


Figure 2.11: Gaussian fit parameters versus time for Wink 1 (blue), Wink 2 (green), and Wink 3 (red): (a) amplitude and (b) sigma. (From [78] ©2013 Taylor and Francis)

We then looked at the parameter vector behavior resulting from the fitting analysis to detect possible trends that would allow to write the explicit time dependency of A_t and σ_t . From Fig. 2.11 it can be noticed that the amplitude linear behavior is obvious for all the time samples (a), whereas σ_t only shows an asymptotic tendency to a fixed value (b) again attributable to the lack of strong subsidence features within the initial time frames. We analyzed the strength of the linearity by evaluating the linear regression of the amplitude

as function of time and calculating the normalized cross-correlation of the results with the original Gaussian amplitude data (Fig. 2.12).

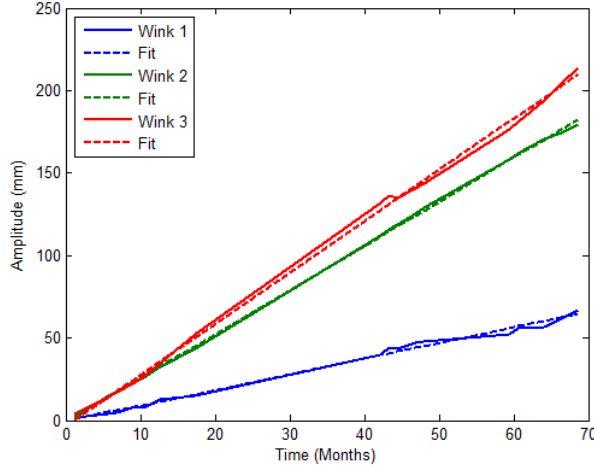


Figure 2.12: Linear regression (dashed lines) of the amplitude versus time plots (solid lines) in Fig. 2.11 for Wink 1 (blue), Wink 2 (green), and Wink 3 (red). (From [78] ©2013 Taylor and Francis)

For all the cases, the calculation yielded a coefficient of nearly one justifying the following assumption for the temporal behavior of the sinkhole amplitude: $A_t = A(t) = \alpha t$.

Based on our analysis, we modeled the spatiotemporal evolution of the sinkhole-induced subsidence as a Gaussian with constant σ and linearly increasing amplitude. Equation (2.8) was then modified to describe the displacement due to a growing sinkhole centered at a particular set of coordinates

$\mathbf{x}_0 = (x_0, y_0)$ as follows:

$$g_{\mathbf{p}}(\mathbf{x}, t) = \alpha t \exp \left[-(\mathbf{x} - \mathbf{x}_0)^2 / 2\sigma^2 \right] \quad (2.9)$$

Equation (2.9) represents the basic model we used in our detection and classification algorithms.

2.4 Simulation

The basic underlying phenomenon behind sinkhole formation is the slow erosion of sediments loosened up by the dissolution of carbonate rocks. These sediments are then flushed away through cracks in the bedrock underlying the top soil. As this process creates empty

underground spaces, larger rocks and boulders will dislodge from their location and fall to fill the vacuum. With time, the type and relationship between the different soil layers will dictate the observed sinkhole morphology.

To gain a better understanding of this process, and the interactions between the soil layers and the final sinkhole structure, and to verify the behavior described by (2.9), we created a model of a multi-layered soil and simulated a dissolution process by slowly removing material from the bottom of the stack.

The simulations were carried out using ESyS-Particle, an Open Source parallel computing software for particle-based numerical modeling written in C++ with a Python application programming interface [82].

ESyS-Particle is an example of what has become known as *molecular dynamics* or *discrete element method (DEM)* [83] [84]: a family of numerical computational methods for the solution of motion problems involving a large number of particles. In particular, ESyS-Particles implements a Lagrangian explicit finite-difference time-integration scheme.

The Lagrangian frame of reference refers to an approach where the physical properties of a material under analysis are described in term of coordinates that are fixed with respect to the material itself. In other words, the calculation point follows the material elements, or particles, as they move through space: velocities and positions are evaluated at each time step, and the instantaneous acceleration is obtained from the application of Newton's second law to each particle. In contrast, Eulerian methods fix a series of locations in space and study the evolution of physical quantities of interest at those locations as time progresses. In other words, as material moves across a specific location, quantities are calculated within fixed spatial boundaries. Lagrangian methods are typically used when it is possible to easily

identify elements within the material and describe the forces acting on and between them, such as in our case, whereas Eulerian methods are typically used when identification of particles is not easy as, for example, within fluids, or when the interest is focused on global properties such as density and pressure. There exists also hybrid methods that have been developed to address situations where the behavior transitions from one representation to the other [85].

The term *explicit* refers to the integration scheme used in the numerical simulation. Explicit methods evaluate the current state of the model based exclusively on the state at previous times. Similarly to what is observed for finite impulse response (FIR) and infinite impulse response (IIR) filter, explicit methods are typically more stable respect to their counterpart (implicit) which also include the current state as part of the integration, thus requiring an iterative approach until convergence is reached. The drawback to this gained stability is the requirement for a much shorter simulation time step to guarantee accurate results. As for the Lagrangian-Eulerian descriptions, also in the case of explicit-implicit integration schemes, hybrid methods have been developed [86].

Within the Lagrangian approach, each particle i is identified by the location of its center of mass (\mathbf{r}_i) and its orientation (ϕ_i) with respect to a given frame of reference, and the dynamic of the system is regulated by Newton's equations [87]:

$$\frac{\partial^2 \mathbf{r}_i}{\partial t^2} = \frac{1}{m_i} \mathbf{F}_i(\mathbf{r}_i, \mathbf{v}_i, \phi_i, \omega_i) \quad (2.10)$$

$$\frac{\partial^2 \phi_i}{\partial t^2} = \frac{1}{\hat{\mathbf{J}}_i} \mathbf{M}_i(\mathbf{r}_i, \mathbf{v}_i, \phi_i, \omega_i) \quad (2.11)$$

where \mathbf{F}_i and \mathbf{M}_i , the force and the torque, acting on the i -th particle through the mass m_i

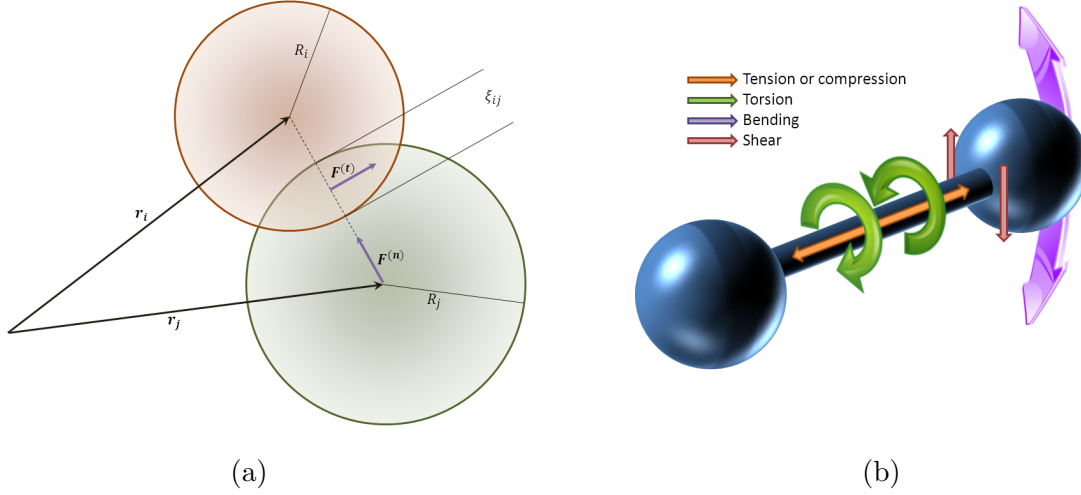


Figure 2.13: Illustration of the interaction forces considered in ESyS-Particles. (a) illustrates the basic frictional-elastic interaction between two unbonded particles, whereas (b) shows the more complex interactions accounted for when considering bonded particles.

and the tensorial moment of inertia $\hat{\mathbf{J}}_i$ respectively, are functions of the particle location \mathbf{r}_i , velocity \mathbf{v}_i , angular orientation ϕ_i , and angular velocity ω_i . In our case, due to the symmetry of our sinkhole model (2.9), we limited our simulation to a two-dimensional space for which both ϕ_i and \hat{J}_i are scalars.

A typical DEM integration cycle will consist of three basic steps:

1. Evaluate the total force and torque acting on a particle due to all the other interacting particles: $\mathbf{F}_i = \sum_{j \neq i} \mathbf{F}_{ij}$ and $\mathbf{M}_i = \sum_{j \neq i} \mathbf{M}_{ij}$
2. Calculate the instant acceleration and integrate (2.10) and (2.11) to evaluate the new position and velocity of each particle
3. Move time forward by a specified amount Δt

These steps will be repeated until the simulation is completed after a user specified time.

For our simulation we used rotating spherical particles. In this case, if i and j are two neighboring particles, they are considered to be interacting with each other if $\xi_{ij} =$

$R_i + R_j - |\mathbf{r}_i - \mathbf{r}_j| > 0$, where ξ_{ij} is known as *mutual compression*, R_i and R_j are the radii, and $|\mathbf{r}_i - \mathbf{r}_j|$ represents the distance between the centers of the two particles (Fig. 2.13).

If the interacting particles are unbonded, the force resulting from the interaction can be projected along the normal $\mathbf{F}^{(n)}$ and tangential $\mathbf{F}^{(t)}$ directions (Fig. 2.13(a)). ESyS-Particle uses frictional-elastic contact forces for these type of interactions. As a result, the normal force is computed using a linear elastic contact law, whereas the tangential force is calculated using a Coulomb friction law as described in [83]. In case of rotational bonded particles, the interaction are more complex (Fig. 2.13(b)) and the force resulting from the superposition of tension or compression, torsion, bending, and shear, is calculated according to the rotational particle bond formulation described in [88]. The inclusion of this type of bonded particles within our model, allows for the description of a cohesive/brittle top layer whose parameters can be modified to simulate a range of top soil characteristics, resulting in different type of sinkholes.

2.4.1 Model description

The model we developed consisted of a two-dimensional structure 150 m wide and 50 m deep. This volume was densely filled, with particles with random radii in the range 0.1-1.0 m, using a routine (GenGeo) provided with the ESyS-Particle distribution. The density was set to 2300 kg/m³, an average value between clay and sedimentary rocks³.

To minimize the influence of the model size on the simulation results, we enforced circular boundary conditions along the horizontal axis, periodically extending the structure both to

³The values used throughout the simulation are taken from the *Handbook of geotechnical investigation and design tables* [89].

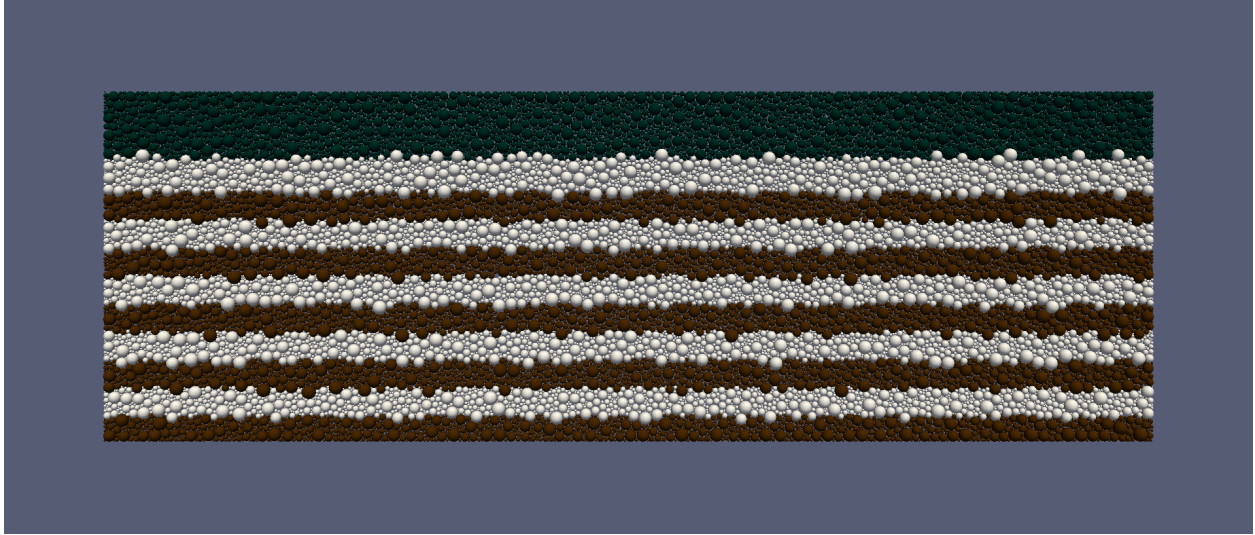


Figure 2.14: Layered structure for the simulation of sinkhole formation with different type of top soil.

the left and the right.

The basic model consisted of three main layers:

- **Bottom bedrock.** This was modeled by bonding the particles nearest to the bottom of the structure with the structure itself using a non rotational elastic link. The Young's modulus of this interaction was set to 15 MPa in line with values for harder rocks such as Granite and Sandstones. We also set a bond-breaking length of twice the maximum radius to allow for interactions simulating surface erosion of the bedrock layer.
- **Middle layer.** The middle layer (represented as alternating tan and brown bands in Fig. 2.14) was composed of unbonded rotational particles undergoing a frictional-elastic interaction. The physical properties for this group of particles were designed to simulate a mix of sand, large gravel, and rocks with a Young's modulus of 15 MPa, a shear modulus of 100 kPa, and static and dynamic friction coefficients of respectively 0.6 and 0.4. Given the large size of the particles, we simulated the high viscosity, that might

be encountered while interacting with finer materials, by introducing fairly high linear and rotational damping parameters, 1.5 and 0.1 respectively. These parameters are used by ESyS-Particle to generate a resistance to the motion proportional to the linear and rotational velocities. This has the result of effectively limiting the accumulation of kinetic energy by individual particles.

- **Top layer.** This layer (represented in green in Fig. 2.14) simulated the top soil and was composed of particles bonded to their immediate neighbors creating a continuous structure extended periodically by the circular boundary conditions. The Poisson's ratio and the angle of internal friction were kept constant to 0.25 and 30° , consistent with values for clay, whereas the Young's modulus and cohesive strength were varied for each simulation to mimic different types of soil behavior: from very hard to very soft clay.

For each simulation, the layered structure was allowed to settle under gravity, the only external force acting on the particles. When, according to the total internal kinetic energy, the equilibrium was reached, a small aperture in the middle of the bottom layer was opened allowing the outflow of the unbonded particles from the structure, effectively simulating the slow drainage of sediment through a crack in the bedrock layer.

Since one of the purposes of the simulations was to verify the model obtained from the analysis of the Wink Sinks dataset, we generated displacement profiles, similar to the one described in Fig. 2.8, by subdividing the top layer in 5 m intervals (the average size of the ground resolution element of a SAR image (see chapter A: *Satellites and SAR*

instrumentation)), and calculating the average displacement of the particles within each interval.

Following the process previously illustrated (see section 2.3: *The model*), we computed the best fitting Gaussian (2.9) for each of the displacement profiles, and evaluated the behavior of the amplitude and scale (sigma) parameters as function of time. We then verified the linearity of the amplitude behavior by evaluating the R^2 coefficient of the linear fit.

2.4.2 Results

Dissolution sinkholes

Dissolution sinkholes form when a small depression starts developing in the bedrock and the top layer smoothly mirrors the underlying surface. In our simulation environment it was not possible to simulate the slow action of water erosion on the bedrock. Instead we decided to simulate this process as a slow loss of loose material in a situation where the top layer was composed by a very cohesive/non-brittle soil with a Young's modulus and the cohesive strength of 15 MPa and 8 MPa respectively.

The results of the simulation are displayed in Fig. 2.15, where in (a), (b), and (c) we show three different time steps of the simulation. In (d) we show the evolution of the displacement profile with time. For each one of the profiles, we computed the best fitting Gaussian using (2.9) as template and plotted the calculated amplitude and sigma as function of time in (e) and (f) respectively.

The amplitude follows an almost perfect linear trend with a coefficient of $\alpha = -0.61 \text{ m/s}$ ($R^2 = 1.00$). It is also interesting to notice that the sigma obtained from this simulation

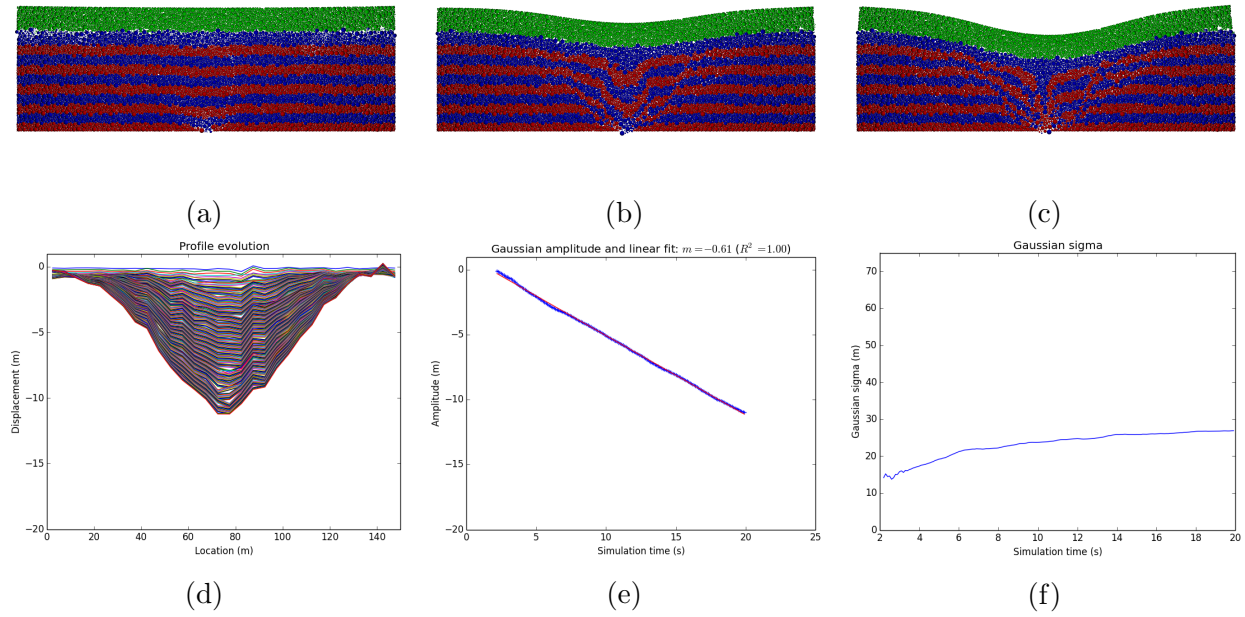


Figure 2.15: Simulation of a dissolution sinkhole. Three different stages of the simulation ($t = 0.2, 1.1, 20.0$ s) are shown in (a), (b), and (c) respectively. In (d) we illustrate the evolution of the displacement profile, in (e) the amplitude and in (f) the scale (sigma) of the best fitting Gaussian as functions of time.

follows a behavior that is comparable with the one obtained from the Wink Sinks analysis, displaying a trend towards an asymptotic value (Fig. 2.11).

Cover-subsidence sinkholes

Cover-subsidence sinkhole follow a similar pattern of dissolution sinkholes with the difference that the top layer is thick and composed of mostly granular material. To simulate this type of feature, we modified the parameters to remove the bonds between the particles in the top layer and replace them with the frictional-elastic properties of the middle layer.

The results of the simulation are displayed in Fig. 2.16.

Similarly to the previous results, the amplitude of the best fitting Gaussian follows a linear trend in time with a coefficient of $\alpha = -0.85$ m/s ($R^2 = 1.00$), and an asymptotic

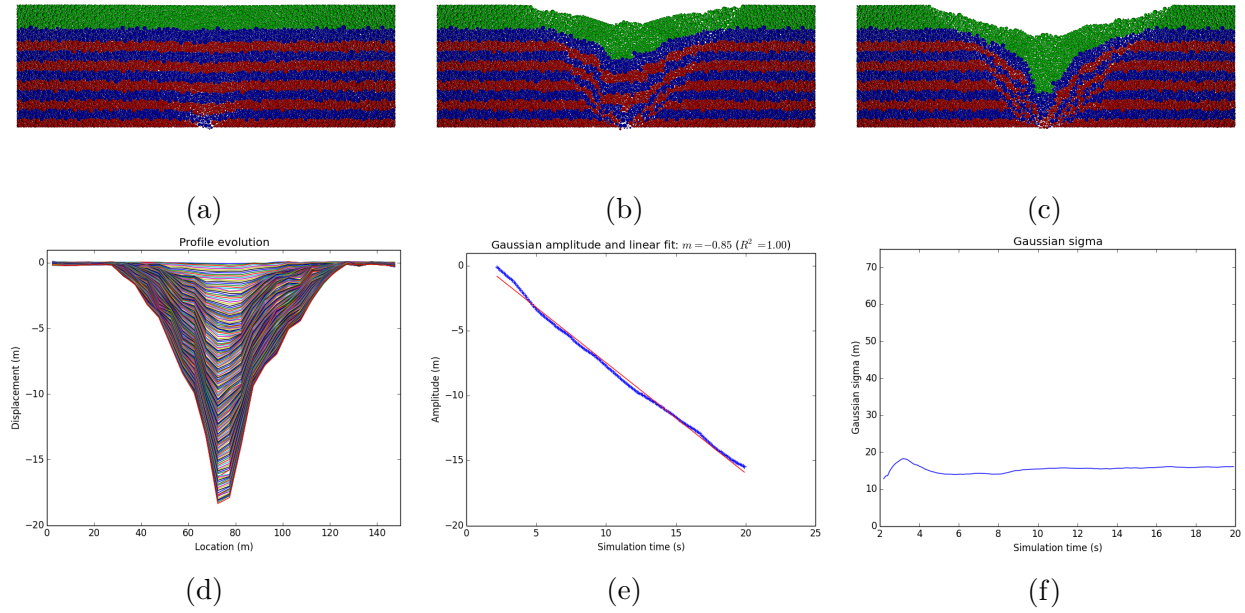


Figure 2.16: Simulation of a cover-subsidence sinkhole. Three different stages of the simulation ($t = 0.2, 1.1, 20.0$ s) are shown in (a), (b), and (c) respectively. In (d) we illustrate the evolution of the displacement profile, in (e) the amplitude and in (f) the scale (sigma) of the best fitting Gaussian as functions of time.

behavior can be observed for the Gaussian sigma as function of time.

Cover-collapse sinkholes

Cover-collapse sinkholes form when the top soil is composed of high cohesive brittle material. When the supporting force from the underlying soil is remove, the top layer typically collapses under his own weight.

We simulated this process for two different types of top layers. The first set of parameters was chosen to produce a behavior similar to a very stiff clay with a Young's modulus of 150 MPa and a tensile strength of 800 kPa. For the second simulation we further increased the stiffness of the material to prevent the collapse of the top layer and evaluate if, under these conditions, it was still possible to detect linear changes in displacement. Due to the thickness

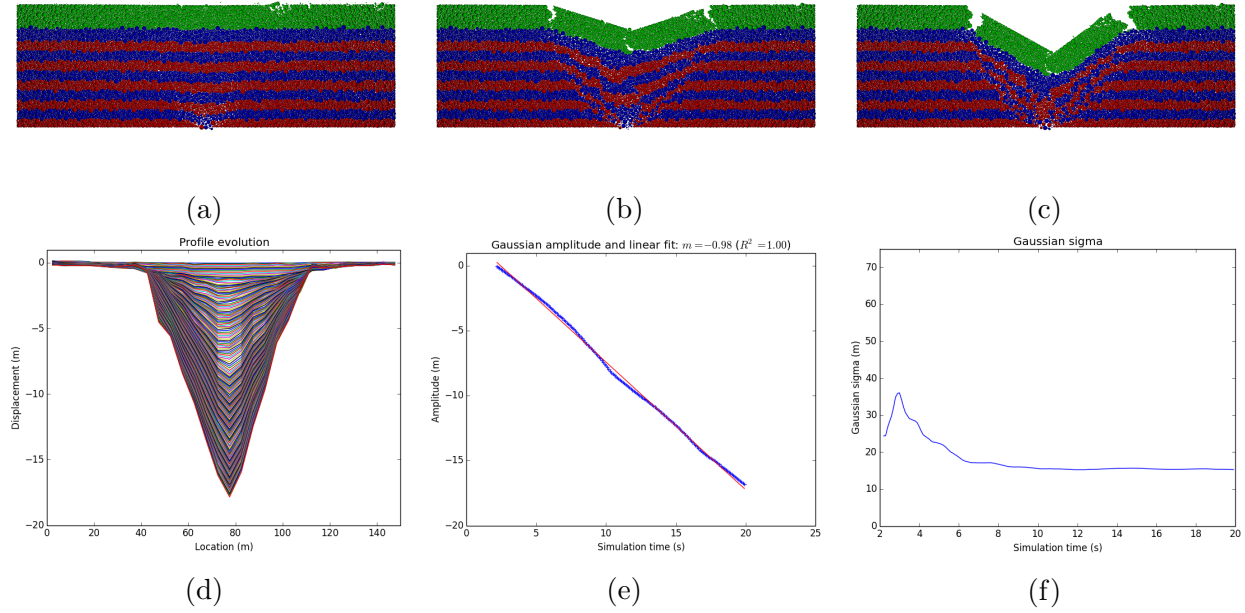


Figure 2.17: Simulation of a cover-cover collapse sinkhole. Three different stages of the simulation ($t = 0.2, 1.1, 20.0$ s) are shown in (a), (b), and (c) respectively. In (d) we illustrate the evolution of the displacement profile, in (e) the amplitude and in (f) the scale (sigma) of the best fitting Gaussian as functions of time.

of the top layer, we had to increase the values of the Young's modulus and the cohesive strength to 1.5 GPa and 8 MPa respectively. These values are more typical for sedimentary rock such as Sandstone and Shales than for clay.

Results for these two simulation are shown in Fig. 2.17 and Fig. 2.18 respectively. In the case of the collapsing sinkhole, we can still clearly observe the linear behavior of the amplitude of the Gaussian with time ($\alpha = -0.98$ m/s and $R^2 = 1.00$), and a sigma behavior resembling the previously observed asymptotic one. In the case of the non collapsing sinkhole, it is still possible to notice a fairly good linear trend with a lower growth rate ($\alpha = -0.12$ m/s and $R^2 = 0.9$), but the behavior of the Gaussian sigma is more erratic. This could also be influenced by the fact that, as it is possible to observe both in the actual simulation snapshot (a-c) and in the profile evolution (d), the size of the developing feature was quite large in

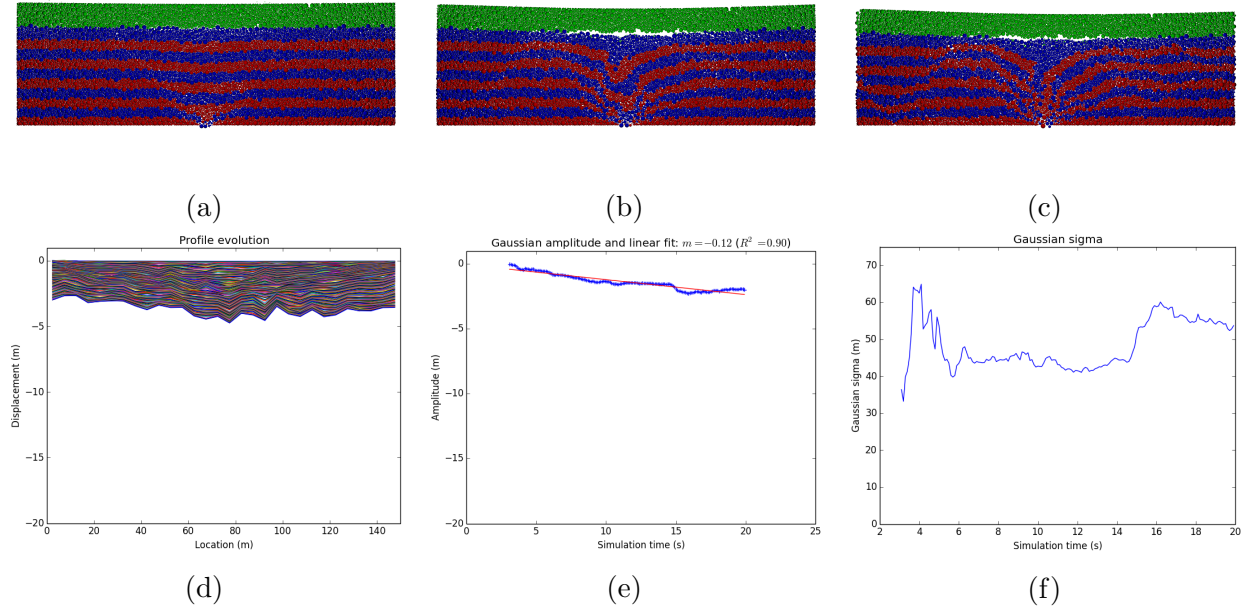


Figure 2.18: Simulation of a sinkhole forming under a hard rock top. Three different stages of the simulation ($t = 0.3, 1.1, 20.0$ s) are shown in (a), (b), and (c) respectively. In (d) we illustrate the evolution of the displacement profile, in (e) the amplitude and in (f) the scale (sigma) of the best fitting Gaussian as functions of time.

comparison with the actual simulation space and extended passed the edge of the model being effectively mirrored to the opposite side by the periodic boundary conditions.

2.5 Summary

In this chapter we presented some general information about the mechanism at the base of sinkhole development and how different interactions between soil layers result in different sinkhole morphologies.

We have also derived a spatiotemporal model for the development of a sinkhole by analyzing the behavior of the scatterer within the Wink Sinks dataset. The relevance of this dataset is dictated by the fact that one of the sinkholes, present within its area (Wink 2), collapsed after the satellite data was acquired, providing a ground truth measurement against which

the detection and characterization methods we will present in the following chapters are going to be tested.

We concluded the chapter by presenting the results of DEM simulations that we developed to verify the behavior of the spatiotemporal model for different types of sinkholes.

Chapter 3

Detection

3.1 Introduction

In the previous chapter we have described how, by analyzing the ground truth behavior of the features present within the Wink Sinks dataset (see section 2.2.2: *The Winkler sinkholes (Wink Sinks)*), we developed a model of the spatiotemporal evolution of sinkholes. In this chapter we will describe the process we followed, starting from a basic segmentation technique (graph cut), to develop of our main contribution: a framework for model-based feature detection in sparse spatiotemporal data.

3.1.1 Graph theory primer

Point cloud datasets are typically an unorganized collection of points obtained by some sampling process of an existing object. These type of datasets do not provide connectivity information or a manifold/mesh structure, making it hard to directly implement geometrical methods. Furthermore, due to the nature of the acquisition/processing techniques, often they

lack regular sampling, thus limiting the direct implementation of signal processing techniques [90]. Because of these characteristics, we first represented our datasets as *graphs*.

The starting point for a graph is a nonempty set $V = \{v_1, \dots, v_n\}$ whose elements are called *vertices* (or *nodes*). The *order* of the graph ($|V|$) is the number of vertices of a graph. Vertices can be connected with a collection $E = \{e_{ij} = (v_i, v_j) \mid v_i, v_j \in V\}$ of two-element tuples of V called *edges*. The *size* of a graph ($|E|$) is the number of edges in the graph. Two vertices are said to be *adjacent* if they are directly connected by an edge and the connecting edge is said to be *incident* onto the vertices it joins. The most basic definition of *graph* is then provided by a combination of the previous two set: $G = (V, E)$. This definition is still general enough to capture the abstract idea of graph: a set of vertices and the existence (or not) of a relationship between them (edges), but nevertheless it allows the development of an entire set of topological structures and operators [91]. What is still lacking is a description of the type and level of connectivity between vertices.

The next level of specialization comes with the definition of the type of edges: *undirected* simply represent connectivity between two vertices ($e_{ij} = e_{ji}$), whereas *directed* start from one node and end at another ($e_{ij} = v_i \rightarrow v_j$) and allow to represent concept such as "dependency". In this project we will be using undirected graphs without self-loops (vertices connected to themselves) or parallel edges (multiple edges connecting two adjacent vertices). Edges might also carry information about the level of connectivity. This is typically expressed by defining a *weight* (w_{ij}) associated with the edge e_{ij} . Weights are typically, but not necessarily, non-negative ($w_{i,j} \geq 0$). For an undirected graph it is usually required that $w_{ij} = w_{ji}$.

There are several different approaches that can be followed to define the connectivity of a graph [91] [92] [93]. In our case, since the data can be embedded in an Euclidian space,

defined by the geographical coordinates of the location of the scatterers, the weight definition is derived as some functional of the natural definition of physical distance. In other cases, such as abstracted graphs where, for example, nodes might represent features of a given dataset element, connectivity and distance might be designed to represent the more abstract concept of "is similar to", "is like", "interacts often with", etc. An example of this type of representation is provided by the *PageRank* algorithm at the basis of the Google search engine where the "[...]the number and quality of links to a page[...]" is used "[...]to determine a rough estimate of how important the website is." [94].

In many situations it is customary to base the connectivity of a graph on the concept of "similarity". In this case nodes, considered similar by some defined metric, will have incident edges with higher weights (typically normalized to 1) while very dissimilar nodes will have incident edges with weights closed to 0. This provides a natural concept of neighborhood by increasing the "strength" of the edges connecting similar nodes while allowing for easy removal (by thresholding, for example) of weak links. In some other situation, it is often more natural to consider the relationship between nodes that are "closed together" (*clusters*) introducing and underlying idea of an actual "distance" between vertices that follows the usual properties of a distance measure: $d_{ij} = d(v_i, v_j) = 0$, $d_{ij} = d_{ji}$, and $d_{ij} \leq d_{ik} + d_{kj}$. The use of either one approach or the other is dictated by the specific characteristics of the dataset and the goals of the representation.

One advantageous approach is to represent graphs as matrices. For a graph with n vertices, the $n \times n$ *adjacency matrix* (\mathbf{A}) is constructed such that the entry in row i , column j is 1 if the edge $e_{ij} = (v_i, v_j)$ is present and 0 if it is not. In case of weighted edges, the entry

will be the edge weight¹. Due to its construction, the adjacency matrix of an undirected graph is symmetrical and entries on the diagonal are zero if there are no self loops. The same representation can be used in case of weighted, directed graph in which case the entry in row i , column j is the weight of the directed edge $e_{ij} = v_i \rightarrow v_j$.

If we define as \mathcal{N}_i the *neighborhood* of vertex v_i (the set of vertices adjacent to v_i), then we can define the *degree* of the vertex v_i as:

$$d_i = \sum_{j \in \mathcal{N}_i} w_{ij} \quad (3.1)$$

With (3.1) we can proceed to define the *degree matrix* (\mathbf{D}) as the $n \times n$ diagonal matrix with the degrees d_1, \dots, d_n on the diagonal.

As an example, consider the weighted, undirected graph in Fig. 3.1. This can be completely described by its adjacency and degree matrices:

$$\mathbf{A} = \begin{array}{c} \text{nodes} \\ \begin{array}{c} A \\ B \\ C \\ D \\ E \end{array} \end{array} \begin{array}{ccccc} A & B & C & D & E \\ \left[\begin{array}{ccccc} 0 & 0.5 & 0 & 0 & 0 \\ 0.5 & 0 & 1.0 & 0.75 & 0 \\ 0 & 1.0 & 0 & 1.0 & 0 \\ 0 & 0.75 & 1.0 & 0 & 0.5 \\ 0 & 0 & 0 & 0.5 & 0 \end{array} \right] \end{array} \quad \mathbf{D} = \begin{array}{c} \text{nodes} \\ \begin{array}{c} A \\ B \\ C \\ D \\ E \end{array} \end{array} \begin{array}{ccccc} A & B & C & D & E \\ \left[\begin{array}{ccccc} 0.5 & 0 & 0 & 0 & 0 \\ 0 & 2.25 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 2.25 & 0 \\ 0 & 0 & 0 & 0 & 0.5 \end{array} \right] \end{array}$$

It is interesting to notice that, if we define the weighted version of the adjacency matrix as $\mathbf{W} = \mathbf{D}^{-1}\mathbf{A}$, the resulting operator represents a diffusion process from each vertex to its

¹The same representation can be used when using a distance instead of a weight (similarity) connectivity. In this case the weights will be replaced by distances and unconnected vertices will have a distance of ∞ .

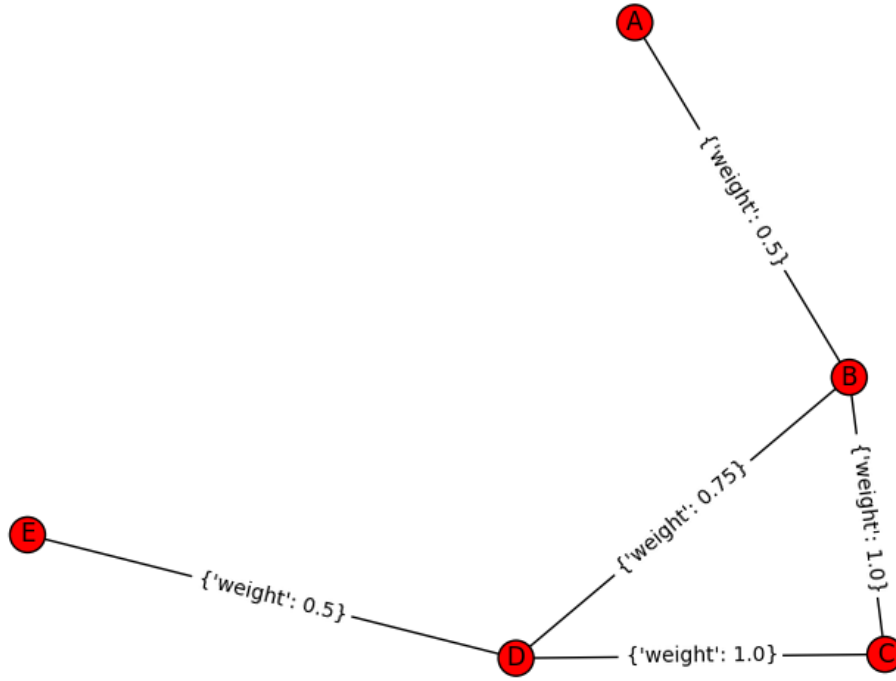


Figure 3.1: Example of basic graph.

neighbors determined by the weights of the edges. To verify that this is the case, let us first define a signal \mathbf{f} on a graph as one that only takes values on the nodes $f_i = f(v_i)$. This definition allows us to represent a signal on a graph as a vector of \mathbb{R}^n . If we now imagine to apply the \mathbf{W} operator to a *delta*, that assumes the value of 1 on a specific vertex and zero everywhere else ($f_i = f(v_i) = \delta_i$), we can see how this will be equivalent to distribute the value to the adjacent nodes based on the edge weights.

3.2 Graph cut approach

Armed with the basic notions of graph theory, we approached the problem of detecting sinkholes within the dataset as a segmentation problem on a graph [95].

In image processing, segmentation is often implemented as a preprocessing step to separate

objects from each other and from the common background [96] [97] [98]. Thresholding of the individual pixel intensity [99], detection of salient edges [100], statistical comparison between texture and color characteristics of different regions [101], and level set analysis [102] are just few examples of the techniques implemented to achieve the goal of identifying specific regions within an image. These techniques have expanded to include analysis of graphs representations and, under the collective name of *graph-cut* [103], are routinely implemented in medical imaging [104] [105], environmental sciences [106], and social networks [107].

We constructed our graph by assigning each scatterer, identified by the non-uniform, sparsely-sampled dataset obtained from the processing of the SAR stack by the SqueeSAR algorithm, to a vertex. We then embedded the graph in the Euclidian space determined by the physical coordinates of the scatterers. As connectivity we decided to implement the Delaunay triangulation [108] which maximized the minimum angles of the triangles produced and creates non-overlapping edges resulting in a connectivity construction that, away from the data boundaries, reflects local neighborhood relationships. Due to the construction process, at the boundary of the data, it is possible to observe very narrow and long triangles connecting vertices physically far from each other. We compensated for the potential inaccuracy due to these connections by introducing a weight term inversely proportional to the length of the edge thus minimizing the importance of longer edges (more about it later). Another advantage of the use of Delaunay triangulation is that, respect to other approaches such as thresholded Euclidian distance or k-nearest neighbors [91], it guarantees that the resulting graph is *connected* (give any two vertices there exists at least one path connecting them) thus eliminating the presence of isolated nodes or group of nodes.

If we consider a set of k subsets $\{A_i\}_{i=1,\dots,k} \subset V$, not necessarily disjoint, we can define as *cut* (C) the following measure of the "cost" of separating the subsets [92]:

$$C(A_1, \dots, A_k) = \frac{1}{2} \sum_{i=1}^k W(A_i, A_i^c) \quad (3.2)$$

where $A_i^c = V \setminus A_i$ is the complement of A_i with respect to V , and $W(A, B) = \sum_{v_i \in A, v_j \in B} w_{ij}$ is the sum of the weights of all the edges between the vertices in A and those in B .

In case the sets A_k are mutually disjoint ($\bigcup_k A_k = V$ and $A_i \cap A_j = \emptyset$ $i \neq j$), (3.2) reduces to the sum of the weights of the edges that connect the sets with their complements.

As for any other segmentation algorithm, the fundamental idea behind a *graph-cut* is to identify regions containing similar elements and separate them from each other and/or from a background. Within the graph theory framework, this translates into identifying cluster of nodes with similar properties and cuts that separate them from other cluster which are dissimilar. With this approach, the definition of the edge weights becomes critical. The ideal weighting function should be designed to simultaneously maximize intra-cluster and minimize inter-cluster weights. With the design goal achieved, the simple elimination of the lowest weight edges would result in the minimization of (3.2) and provide the desired result.

One approach could be to directly minimize (3.2), remove the detected clusters and then repeat iteratively on the remaining dataset until all meaningful clusters have been identified [109]. However once this approach is followed, it becomes apparent that a direct application of (3.2) favors cuts around smaller sets with the tendency of selecting single nodes since $W(A, B)$ increase proportionally to the number of edges cut. Typically, to increase spatial coherence and minimize the occurrence of isolated vertices cuts, (3.2) can be regularized by

including a term that favors larger subsets. The two main approaches used are *RatioCut* [110] and *Ncut* [111]: RatioCut normalizes the cut by dividing each sum element by the order (number of nodes) of A_i ($|A_i|$) whereas Ncut uses the sum of the degree of the vertices belonging to each A_i , also known as the *volume* of the subset defined as $\sum_{v_j \in A_j} d_j$.

After careful evaluation of the characteristics of our dataset, we decided to omit either of the normalization factors in the definition of our weights because we did not want to explicitly encourage spatial coherence. This decision was based on the fact that some points in our dataset (especially in the case of distributed scatterers) can represent large physical areas. If such points display trends that, in agreement with surrounding scatterers, accurately indicates that subsidence might be occurring at that location, we should not discourage a cut around them and, in the process, miss a potential sinkhole-forming region.

Graph cuts have been used extensively to pursue segmentation in images where intensity and location are the only information available for each pixel. Our datasets are unique in that they offer a large amount of information for each scatterer: displacement time series, coherence, velocity, acceleration, physical area represented, elevation, and location. Thanks to the richness of such feature set, we had a lot of flexibility in defining the weights to associate to each edge.

To achieve our goal of detecting subsiding regions, we defined independent weights for each of the factors that might be indicative of this phenomenon, with the intention of favoring a minimum for cuts occurring on the boundary of regions that showed sinkhole-forming behavior.

If we consider every edge e_{ij} , obtained from the Delaunay triangulation, for each of the incident vertices v_i and v_j , we can extract the displacement of the corresponding scatterers,

evaluated for one of the time slices t available in the SqueeSAR dataset ($d_i^{(t)}$ and $d_j^{(t)}$) and, from these, calculate the average displacement of the two scatterers $d_{avg}^{(t)} = (d_i^{(t)} + d_j^{(t)})/2$. The dataset also provides the physical coordinates of the two scatterers (used to evaluate the actual distance l_{ij}), and the coherence values c_i and c_j .

We can now define a displacement range of interest (d_{low}, d_{high}), that acts like a threshold for the acceptable average displacement values, and define a set of weight functions as:

$$w_b(v_i, v_j) = \begin{cases} 0 & \text{if } d_{low} < d_{avg}^{(t)} < d_{high} \\ \min \left(\left| d_{high} - d_{avg}^{(t)} \right|, \left| d_{avg}^{(t)} - d_{low} \right| \right) & \text{otherwise} \end{cases} \quad (3.3)$$

$$w_d(v_i, v_j) = \frac{1}{\left| d_i^{(t)} - d_j^{(t)} \right|} \quad (3.4)$$

$$w_r(v_i, v_j) = \frac{2}{c_i + c_j} \quad (3.5)$$

$$w_l(v_i, v_j) = l_{ij} \quad (3.6)$$

Each of these weights was designed to address a specific feature of interest. In particular w_b was designed to minimize the weight for edges connecting scatterers located within the displacement range of interest; w_d to reduce the cost for cutting steep edges, indicative of boundary regions; w_r to favor cuts through reliable edges (the higher the coherence of the two scatterer, the higher the quality of the displacement measurement); finally w_l was introduced to give priority to cuts that passed through shorter, and thus more accurate, edges rather than through sparse regions.

All these weights were normalized to the same mean and standard deviation and then combined into a final function using inter-category weighting parameters α , β , γ , and δ :

$$w_{ij} = \alpha \cdot w_b(v_i, v_j) + \beta \cdot w_d(v_i, v_j) + \gamma \cdot w_r(v_i, v_j) + \delta \cdot w_l(v_i, v_j) \quad (3.7)$$

After the weights were assigned to each edge by (3.7), we used an efficient implementation of the min-cut/max-flow algorithm² to find the minimizing cuts. Finally, to provide a visual representation of the segmentation results, for each cut we identified the midpoints of the removed edges and defined a contour using a spline interpolation.

3.2.1 Results

We applied the method to the Wink Sink dataset (see section B.2: *Wink Sinks*). As displacement values ($d_i^{(t)}$) we used the data selected from the latest available time slice in the dataset. For each scatterer, this corresponds to the cumulative displacement occurred after $t = 68.68$ months from the reference frame acquisition time. The range of the measured displacements for the entire set of scatterers was $[-289 \text{ mm}, 95.5 \text{ mm}]$ with an overall mean of -8.64 mm .

The inter-category weighting parameters were arbitrarily chosen to favor steep edges connecting high coherence scatterers exhibiting an average displacement between the values of $d_{low} = -115 \text{ mm}$, and $d_{high} = -99 \text{ mm}$: namely $\alpha = 0.1$, $\beta = 1$, $\gamma = 2$, and $\delta = 0.5$.

Fig. 3.2 shows the steps of the graph-cut approach. This illustration is restricted to the area of Wink 2 and Wink 4. It is interesting to notice that the removal of the regularization

²For details on the algorithm, its implementation and the wrapper we used see [112] [113] [114] [115].

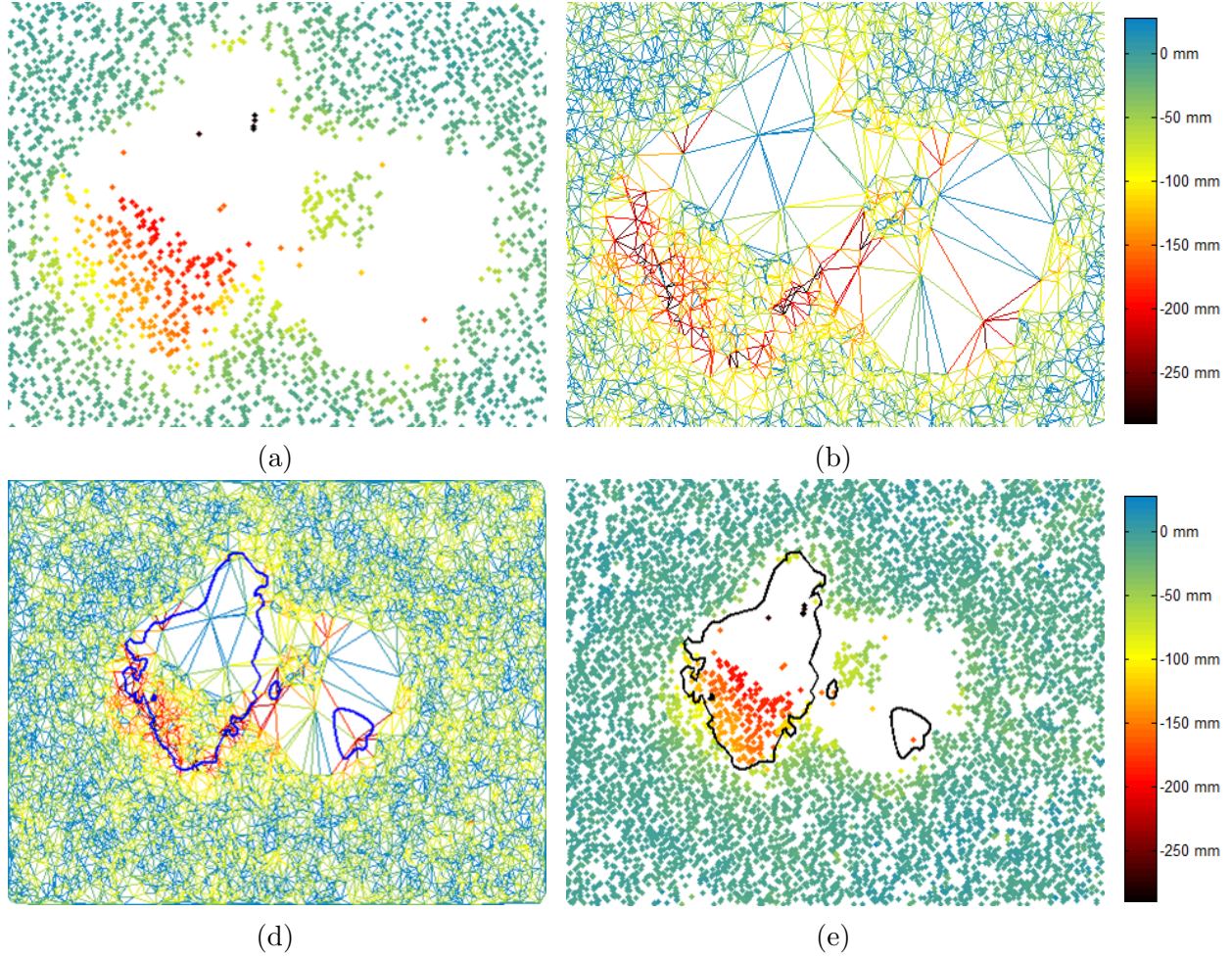


Figure 3.2: Steps in the graph-cut process for Wink 2 and 4. (a) Original dataset representing the displacement of the scatterers for the acquisition taken 68.68 months after the reference image. Each point represents a vertex and is located according to its physical coordinates. (b) Delaunay triangulation defining the connectivity of the graph. (c) Result of the graph-cut algorithm. (d) Spline represented on top of the original data. (e) Spline represented on top of the original data. (From [95] ©2012 IEEE)

term from (3.2), caused one of the cuts to converge to a minimum including a single node.

After analyzing the time series relative to the identified scatterer, we verified that this region was undergoing active subsidence. This was also confirmed by a recent study conducted in 2009, 11 years after the last frame in our dataset was acquired [17].

In Fig. 3.3 we show a direct comparison between the segmentation results for Wink 1 (a) and Wink 2 (b) and the respective ground truths. In (b) and (c) we also show the results of

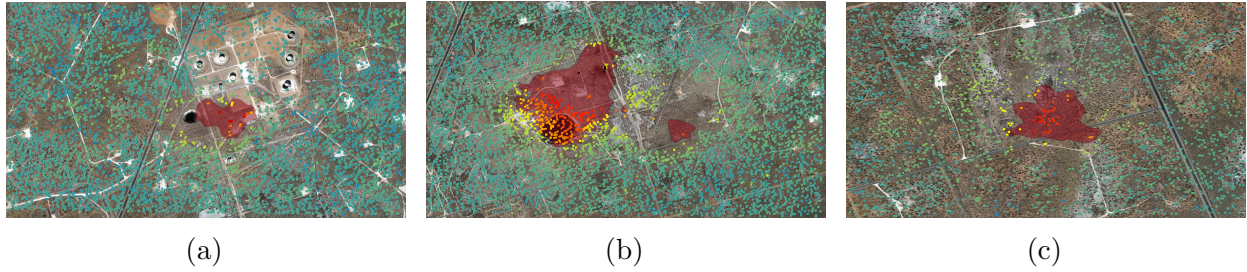


Figure 3.3: Results of the graph cut algorithm applied to the Wink Sinks compared with ground truth data. (a) Wink 1. (b) Wink 2 and 4. (c) Wink 3.

the algorithm when applied to Wink 3 and 4 although, since as May 2014 these sinkholes had not collapsed, no ground truth is available for direct comparison.

Table 3.1: Graph-cut sinkhole detection vs. collapsed ground truth (From [95] ©2012 IEEE)

	Total area	Segmentated	% detected
Wink 1	8,476 m ²	2,306 m ²	27.20%
Wink 2	5,098 m ²	4,775 m ²	93.67%

Table 3.1 provides a quantitative summary of the results shown in Fig. 3.3. The graph cut segmentation approach managed to capture about 27% of Wink 1 and 94% of Wink 2 collapsed sections. We did not capture a significant portion of Wink 1 most probably because the sinkhole had already collapsed and filled with water by the time the satellite data was acquired. As a results, it presented poor coherent scattering characteristics resulting in a very sparse distribution of scatterers.

One think to bear in mind when comparing the results with the size of the collapsed regions is that, often, the region affected by subsidence is much larger and not necessarily symmetrically placed with respect to the actual collapsed section [76].

3.3 Spatiotemporal matching

3.3.1 Introduction

Graphs provide a powerful tool to handle sparse non-uniformly sampled data such as ours but, up until recently, most of the graph cut applications have been directed to the analysis and segmentation of static datasets.

Some recent development have extended these algorithms to include temporal evolution. For example in [106] a stack of subsequent images obtained from the burnt area product (MCD64A1) [116] of MODIS³ is translated in a 3D (2D+time) graph by connecting neighboring pixels. Weights are then assigned to the edges based on the similarity of adjacent pixels intensity values whereas the temporal dimension is included by enforcing coherence between consecutive time slices. This is achieved assigning infinite weights to edges connecting pixels whose state changes from burnt to non-burnt, effectively preventing them from being cut. The minimization is then evaluated globally across the entire stack rather than one individual slices at the time.

A similar approach is used in [105] to evaluate the size of the patches of dryness developing over the iris of patients affected by dry-eye syndrome. An interesting development is the enforcement of decreasing monotonicity obtained by assigning an infinite weight to the temporal edges connecting pixels showing a growing wetness from one frame to the following one of the video. Once again, the minimization is evaluated globally for all the frames.

³The moderate resolution imaging spectroradiometer (MODIS) is 36-channels multi-spectrometer instrument flown aboard the Aqua and Terra earth observation satellites operated by the National Aeronautics and Space Administration.

A third example further extends the application of graph cuts to 4D datasets (3D+time) [104]. In this approach each voxel⁴ is connected both to its spatiotemporal neighbors with one set of edges and to two external nodes labeled "object" and "background" with another. The first set is weighted according to the intensities of the adjacent voxels, whereas the second connects each voxel to its neighbors, with zero weights, and to the external nodes with infinite weights. This serves as a mean to provide classification. After the user selects two initial set of seeding voxels, one belonging to the object and one to the background, the graph cut algorithm separates 4D volumes by enforcing spatiotemporal coherence and, as both connecting edges are severed, classifies each voxel as object or background based on the remaining connections to the external nodes.

All these approaches show the flexibility of graph cut algorithms to handle spatiotemporal problems. Nevertheless, they all rely on the ability to define weights that, while preserving coherence between nodes belonging to an *existing* object, can separate it from the background. On the contrary, in our datasets, the main value associated with each node is its displacement as function of time. Because of this, the object (subsiding regions) might *appear* at a certain spatiotemporal location and then evolve. Furthermore, as we have described in (see chapter 2: *Modeling*), what characterizes the signature of a sinkhole is a specific spatiotemporal *behavior*. In order to classify subsidences phenomena as sinkholes, it is necessary to distinguish between those showing a linear behavior from other displaying a different type of spatiotemporal evolution while, at the same time, identifying all subsiding regions. The method that we propose in this section achieves this result.

⁴ *Voxel* is the name typically assigned to a pixel within a 3D (volume) image.

3.3.2 Method

Our analysis shows that the subsidence in a sinkhole forming regions can be described as a Gaussian of fixed scale (standard deviation) with amplitude growing linearly in time (see section 2.3: *The model*):

$$g_{\mathbf{p}}(\mathbf{x}, t) = \alpha t \exp \left[-(\mathbf{x} - \mathbf{x}_0)^2 / 2\sigma^2 \right] \quad (3.8)$$

We would like to design a method that allows us to evaluate to what extent the available data conforms to our postulated spatiotemporal Gaussian sinkhole displacement evolution model.

The inherent spatial sparsity and non-uniformity of the data obviously presents the major difficulty in achieving the desired result. To address these issues, we decided to adapt the parameter space search approach suggested by the Hough transform [117] [118] by incorporating aspects of matched filtering [26] [27] and residual measurements.

The approach we developed [78] is based on the availability of a model describing the feature of interest. Starting from this representation, in our specific case the spatiotemporal behavior of the subsidence in sinkhole-forming regions illustrated by (3.8), we identify the key parameters and organize them in a vector. In our subsidence model, the parameter vector contains the location, the linear amplitude growth coefficient, and the fixed scale of the feature: $\mathbf{p} = [\mathbf{x}_0, \alpha, \sigma]$.

We employed the basic parameter search of the Hough transform; however, instead of the typical voting step, we applied a technique inspired by the matched filter approach where we replaced the correlation operation with a residual-based measure. This approach

provides the user with a finer control over resolution in the parameter space search, avoids the all-or-nothing nature of the decision threshold used in the Hough voting, and proves advantageous in case where one wishes to more robustly evaluate the matching where only a handful of points are available.

Specifically, this blended approach is implemented as follows: for each point in the parameter space identified by $\mathbf{p} = [\mathbf{x}_0, \alpha, \sigma]$, a template $g_{\mathbf{p}}(\mathbf{x}, t)$ is reconstructed corresponding to a specific instantiation of the spatiotemporal model (3.8). At this point we also identify a region of influence $R(\mathbf{p})$ over which the residual will be evaluated. The size is determined by the parameter \mathbf{p} and is limited to a circular area of radius 3σ centered at \mathbf{x}_0 . This provides a region that is large enough to include the significant part of the template (at distances greater than 3σ from \mathbf{x}_0 the amplitude of the template is less than 0.012% of the maximum) and small enough to limit the evaluation of the residual just to the relevant data points. The template is then used to scan the data point cloud $d(\mathbf{x}_i, t)$ and evaluate the match by calculating the residual $r(\mathbf{p})$ using a specific metric μ . Within the data point cloud, i identifies the coordinates of the N_R scatterers inside the region of influence $R(\mathbf{p})$, and t is one of the T time slices for which displacement information are available.

The metric μ is based upon the general definition of absolute residual $|d(\mathbf{x}_i, t) - g_{\mathbf{p}}(\mathbf{x}_i, t)|$ which is simply the absolute difference between the model-predicted and the measured displacement.

A measure of the match could be obtained by taking the ℓ^1 (or ℓ^2) norm of the residual vector, of size $N_R \times T$ and generated by considering the time series of all the points located within the influence region $R(\mathbf{p})$, and normalizing it by the total number of points considered. However, both of these norms provide a measurement of the actual *average* distance between

the template and the data. It is clear how this would be biased towards shallower features.

As an example, consider a situation where the data points identified by the influence region $R(\hat{\mathbf{p}})$ are compared against the corresponding template $g_{\hat{\mathbf{p}}}(\mathbf{x}_i, t)$ (where $\hat{\mathbf{p}} = [\hat{\mathbf{x}}_0, \hat{\alpha}, \hat{\sigma}]$) resulting in an ℓ^1 (or ℓ^2) residual $\hat{r}(\hat{\mathbf{p}})$. If we were to multiply each displacement value within $R(\hat{\mathbf{p}})$ by a factor β , the resulting ℓ^1 (or ℓ^2) residual would scale to $r'(\mathbf{p}') = \beta \cdot \hat{r}(\hat{\mathbf{p}})$ when considering the match with a new template $g_{\mathbf{p}'}$, where $\mathbf{p}' = [\mathbf{x}'_0, \alpha', \sigma']$, $\mathbf{x}'_0 = \hat{\mathbf{x}}_0$, $\sigma' = \hat{\sigma}$, and $\alpha' = \beta \cdot \hat{\alpha}$, while intuitively we would like both matches to provide an identical result since the *behaviors* are identical.

This would seem to suggest the use of classic normalizing approaches such as the normalized cross-correlation. The main drawback with this type of approaches is that the normalization step is performed in isolation for each of the terms independently rather than jointly, resulting in the removal of any global scale information. In our application, due to the nature of the model (3.8), where growth is based exclusively on change of scale linearly with time, this would effectively eliminate the temporal dependency.

To remove the potential bias toward smaller features and maintain temporal dependency, we designed a measure μ based on the notion of *proportional match*:

$$\mu(\mathbf{x}_i, t) = \min \left(\frac{|d(\mathbf{x}_i, t) - g_{\mathbf{p}}(\mathbf{x}_i, t)|}{\max(|d(\mathbf{x}_i, t)|, |g_{\mathbf{p}}(\mathbf{x}_i, t)|)}, 1 \right) \quad (3.9)$$

where $|d(\mathbf{x}_i, t) - g_{\mathbf{p}}(\mathbf{x}_i, t)|$ is the measure of the mismatch between the data point considered $d(\mathbf{x}_i, t)$ and the corresponding template value $g_{\mathbf{p}}(\mathbf{x}_i, t)$ evaluated at the same spatiotemporal coordinates (\mathbf{x}_i, t) .

This measure, which can be verified to be a metric, provides the desired amplitude scale

invariance and it guarantees that equivalent *behaviors* will results in equivalent measures independently of the amplitude of the considered template. This matches our final goal to provide detection and classification of features based on their behavioral similarity to the spatiotemporal model of the feature of interest. For data points where measured and template values have opposite sign, we limit the proportional measure (3.9) at one indicating maximum mismatch.

One of the issues deriving from the sparsity of the data is that the density of data points within the influence region is not constant. Under certain condition, it could be possible for the residual to be evaluated exclusively based on data concentrated in specific subsets of $R_{\mathbf{p}}$ leading to an incorrect results. To minimize the occurrence of such scenarios, we partition the influence region into three subregions based on the value of σ identified by \mathbf{p} :

$$R_1(\mathbf{p}) = \{\mathbf{x}_i \in d(\mathbf{x}_i, t) \mid 0 \leq \|\mathbf{x}_i - \mathbf{x}_0\| < \sigma\} \quad (3.10)$$

$$R_2(\mathbf{p}) = \{\mathbf{x}_i \in d(\mathbf{x}_i, t) \mid \sigma \leq \|\mathbf{x}_i - \mathbf{x}_0\| < 2\sigma\} \quad (3.11)$$

$$R_3(\mathbf{p}) = \{\mathbf{x}_i \in d(\mathbf{x}_i, t) \mid 2\sigma \leq \|\mathbf{x}_i - \mathbf{x}_0\| < 3\sigma\} \quad (3.12)$$

where $R(\mathbf{p}) = R_1(\mathbf{p}) \cup R_2(\mathbf{p}) \cup R_3(\mathbf{p})$.

The value of $\mu(x_i, t)$ is independently evaluated and normalized according to the number of spatial and temporal data points contained within each subregion:

$$r_k(\mathbf{p}) = \frac{1}{N_k \times T} \sum_t \sum_{x_i \in R_k(\mathbf{p})} \mu(x_i, t) \quad (3.13)$$

where N_k is the total number of points contained in $R_k(\mathbf{p})$ and t is the index into the total

number of time samples T available in the dataset. These partial residuals are then combined via a simple average. The result $r(\mathbf{p})$ is, in the case of our model, a four-dimensional matrix where each element represents the quality of the fit between the data and the spatiotemporal template $g_{\mathbf{p}}(\mathbf{x}_i, t)$.

3.3.3 Results

To validate the spatiotemporal approach, we tested the algorithm we developed against an artificial dataset, the Wink Sinks dataset (see section B.2: *Wink Sinks*), and the Virginia dataset (see section B.3: *Virginia*) where we verify our results against ground validation performed by Virginia Department Of Transportation (VDOT) geologists.

Artificial data

The first step towards validation of our algorithm was to test the approach on three artificial datasets designed to verify its efficacy in detecting data regions conforming to the specified model.

The coordinate section of the first dataset was obtained from a uniform spatial grid $\mathbf{x} = (x, y)$ with $-50 \text{ m} \leq x \leq 50 \text{ m}$ ($\Delta x = 2.5 \text{ m}$) and $-50 \text{ m} \leq y \leq 50 \text{ m}$ ($\Delta y = 2.5 \text{ m}$) for a total of 1,681 points for each of the 22 time intervals that were chosen to be consistent with the Wink Sink dataset: 0.00, 1.15, 5.75, 6.90, 9.21, 10.36, 12.66, 17.26, 42.18, 42.21, 43.33, 44.48, 44.52, 46.78, 46.82, 47.93, 47.97, 59.47, 60.62, 64.08, 66.38, and 68.68 months from the master reference (first acquisition). This grid was then used, at each time interval, to generate subsidence data using the model (3.8) with the parameter vector $\mathbf{p}_0 = [\mathbf{x}_0, \alpha, \sigma] =$

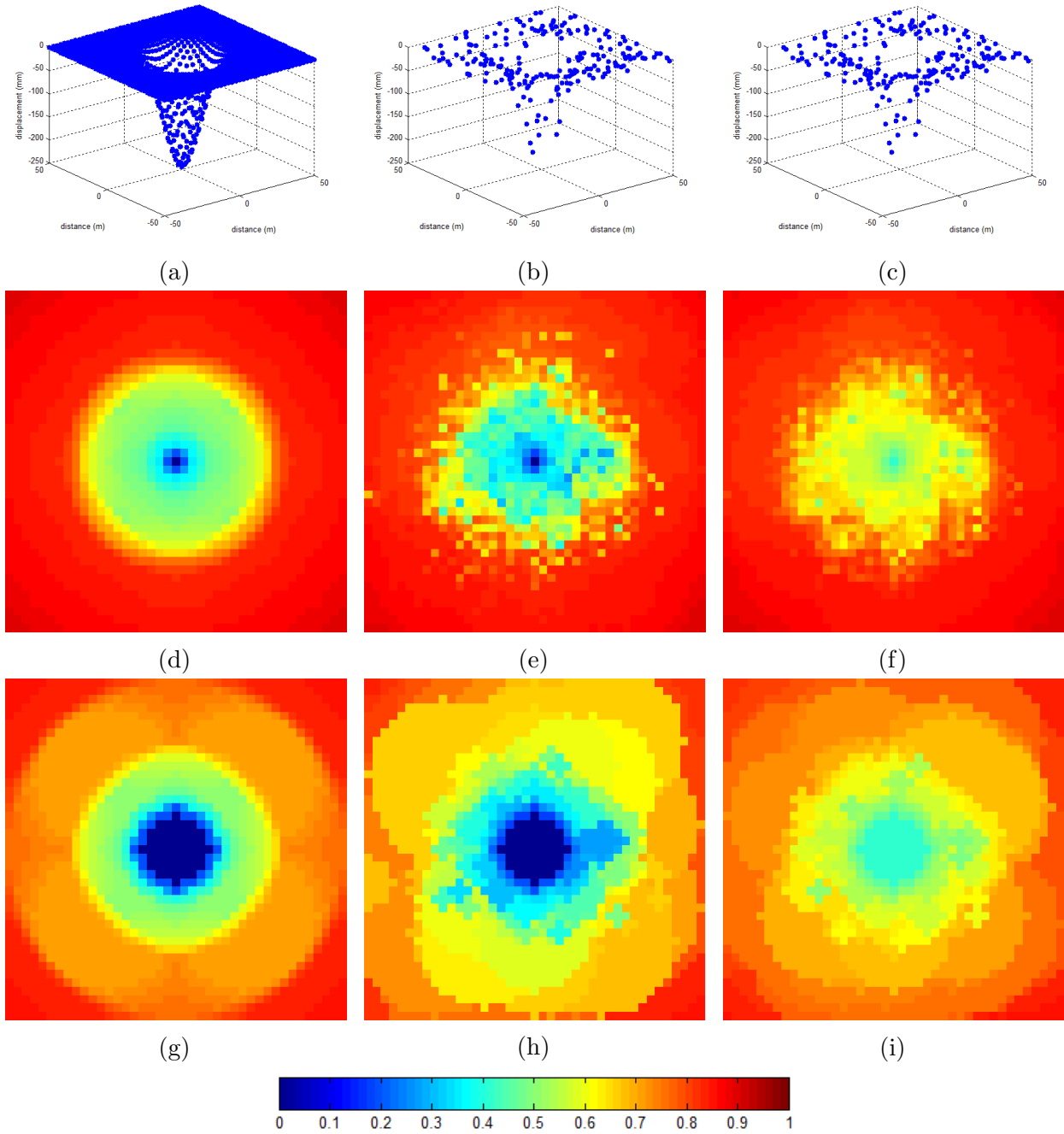


Figure 3.4: Artificial test datasets and results of the detection algorithm applied to these datasets. (a) Time varying uniform spatial grid (shown at $t = 44.5$ months). (b) Time varying non-uniformly-sampled version of (a) (shown at $t = 44.5$ months). (c) Time fixed version of (b): the frame for $t = 44.5$ months is repeated for all the time intervals. (d-f) Result of the detection algorithm applied to (a-c). (g-i) Propagated version of (d-f). (From [78] ©2013 Taylor and Francis)

$[(0 \text{ m}, 0 \text{ m}), -5.5 \text{ mm/month}, 10 \text{ m}]$. The resulting point cloud for $t = 44.5$ months is shown in Fig. 3.4(a).

The second dataset was generated in a similar fashion with the difference that, to increase similarity with the data processed by the SqueeSAR algorithm, the uniform spatial grid was replaced with a grid generated by 200 realizations of a bivariate uniform distribution: $\mathbf{x} \sim \mathcal{U}_2(-50, 50)$. The resulting point cloud for $t = 44.5$ months is displayed in Fig. 3.4(b).

For the last dataset, we extracted the data corresponding to $t = 44.5$ months from the second dataset and repeated it for each time interval. This type of data simulates a sudden change of elevation with no further subsidence. This corresponds to our model in the spatial coordinates, with the temporal coordinate being static. The resulting distribution, for any of the time intervals, is illustrated in Fig. 3.4(c).

We applied our approach to all three artificial datasets. The parameter space limits and increments were chosen to provide, for the first two dataset, a perfect overlap at \mathbf{p}_0 between template and data (Table 3.2).

Table 3.2: Parameter space limits and increments for the artificial dataset

Parameter	Minimum	Maximum	Increment
x_0	−50 m	50 m	2.5 m
y_0	−50 m	50 m	2.5 m
α	−10.05 mm/month	−0.05 mm/month	0.25 mm/month
σ	2.5 m	30 m	2.5 m

The four-dimensional result (one dimension for each parameter in the model) contains the residuals evaluated for each location of the parameter subspace identified by the selected parameter ranges and steps. In order to visualize this multidimensional result, we collapsed two of the dimensions onto the coordinate plane using a procedure adapted from the *maximum intensity projection* [119]. Since, in our case, we are interested in the minima of the residual,

we projected results according to:

$$m(x_0, y_0) = \min_{\alpha, \sigma} [r(\mathbf{p})] \quad (3.14)$$

The result of this projection is shown in Fig. 3.4(d)-(f).

As expected, our algorithm perfectly locates the template for the first two datasets, yielding, for both, a minimum residual $m(x_0, y_0) = 0$ at $(x_0, y_0) = (0 \text{ m}, 0 \text{ m})$. However, we also obtain an attenuated response for the fixed subsidence case (third dataset) with a minimum residual of $m(0 \text{ m}, 0 \text{ m}) = 0.42$.

This result shows that our measure complements strong detection capabilities for data regions behaving as the specified spatiotemporal model with the ability, typical of matched filter, to respond (at a lower level) also to data conforming simply to the general spatial shape of the model. This is desirable as these stable deformations might still be caused by subsidence phenomena and provide clues about the activity level of the region. At the same time, our approach provide the ability to distinguish between the evolving and non-evolving features by responding with two different levels.

Each minimum residual image also exhibit a response beyond the actual location of the best match. For the synthetic dataset, a clear radial fringing pattern is evident (Fig. 3.4(d)). These pattern corresponds to lower amplitude tangential fits between the data and smaller templates offset with respect to the center of the feature. This pattern is also visible for the second and third datasets (Fig. 3.4(e) and (f)) where the sparsity of the data allows for smaller (albeit still poorly fitting) templates to slightly register in emptier areas.

The map obtained using (3.14) only provides information about the location of the feature

and how well it is matched to the original model but it does not give any visual clue on the size of the matched template which is quite relevant in the case of our application. To allow for a direct visualization of the size of the detected feature, we associated a circular region to each minimum. The radius of this region is determined by the σ_m corresponding to the template that provided the best match at that particular location:

$$\sigma_m = \arg \min_{\sigma} \{ \min_a [r(\mathbf{p})] \} \quad (3.15)$$

In case of overlap between circular regions, visualization precedence is given to the region presenting the lowest residual. Examples of what we call *propagated maps* are shown in Fig. 3.4(g)-(i).

Wink Sinks

As second validation step, we applied our detection algorithm to the Wink Sink dataset (see section B.2: *Wink Sinks*).

Table 3.3: Parameter space limits and increments for the Wink Sink dataset

Parameter	Minimum	Maximum	Increment
x_0	0 m	6,670 m	10 m
y_0	0 m	8,290 m	10 m
α	-5 mm/month	-1 mm/month	0.25 mm/month
σ	5 m	185 m	10 m

The parameter space limits and increments were chosen to provide coverage of the entire dataset (Table 3.3).

As described in (see section 2.2.2: *The Winkler sinkholes (Wink Sinks)*), there are four major features in the dataset. Our algorithm was successful in detecting all four features

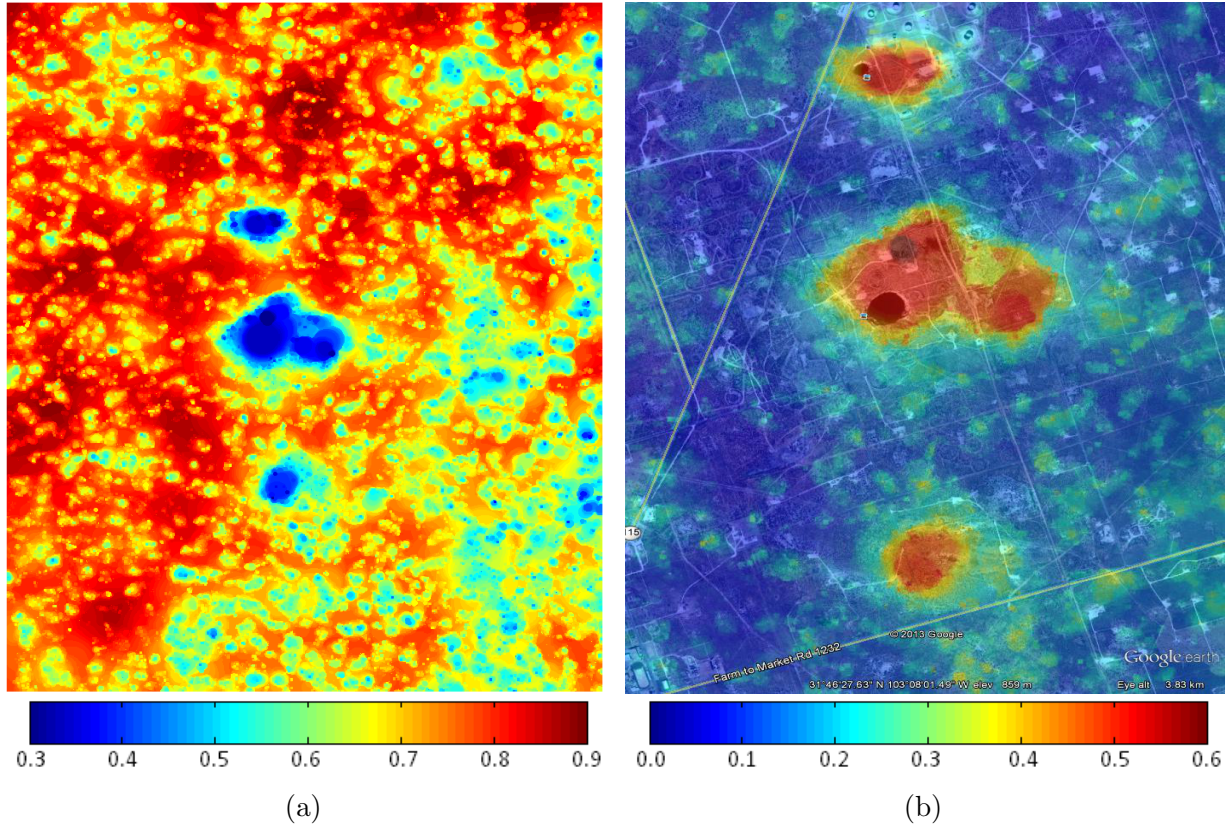


Figure 3.5: Propagated results of our detection algorithm applied to the Wink Sinks. In (a) we show the propagated residual whereas in (b) we illustrate the propagated risk evaluated according to (3.16) (see section 3.3.3: *Virginia*). ((a) from [78] ©2013 Taylor and Francis)

as it can be seen in Fig. 3.5. In this figure we also show the results of the propagated risk evaluated using (3.16) (see discussion below).

Details of the three regions are displayed in Fig. 3.6 where it is outlined (with a white contour) the collapsed region of Wink 1 Fig. 3.6(d) and Wink 2 Fig. 3.6(e). As we previously mentioned, the region affected by subsidence, due to the evolution of the sinkholes, is much larger than the actual collapsed section and our algorithm succeeded in detecting the extent of all the deformations including, as we mentioned when discussing the results of the graph cut algorithm, the region east of the collapsed section of Wink 2 (Wink 4) that in a recent study [18] was confirmed being a separate evolving depression.

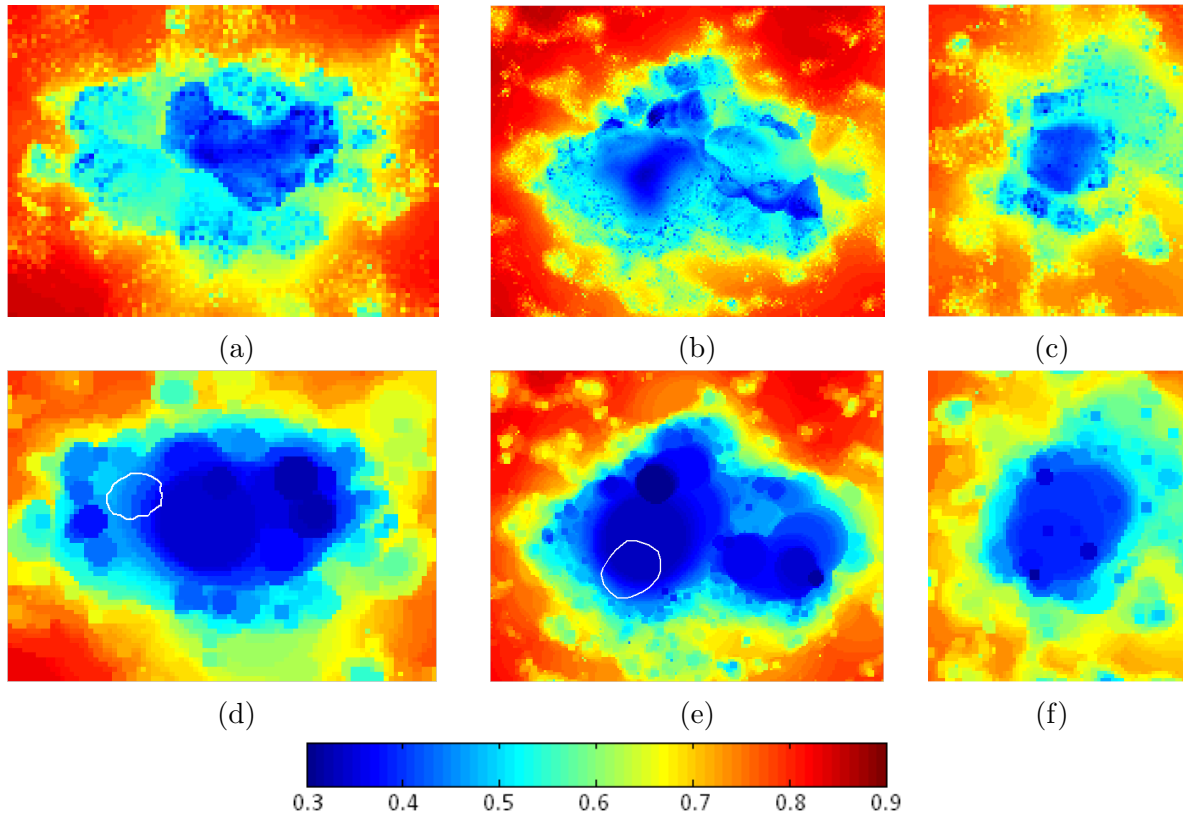


Figure 3.6: Detailed results of our detection algorithm on the Wink Sinks. (a) Detection of Wink 1. (b) Detection of Wink 2 and Wink 4. (c) Detection of depression (potential Wink 3). (d) Propagated version of (a) with included outline of collapsed portion of Wink 1. (e) Propagated version of (b) with included outline of collapsed portion of Wink 2. (f) Propagated version of (c). (From [78] ©2013 Taylor and Francis)

Comparison with graph cut approach

In Fig. 3.7 we present a direct comparison between (left) the graph cut and (right) the spatiotemporal matching approaches. The images relative to the spatiotemporal approach were obtained by binning the propagated risk, evaluated using (3.16), into the four ranges (severe, moderate, slight and none) that we will describe when discussing the results of the analysis of the Virginia data.

The graph cut algorithm does not naturally provide a range of values for the segmentation as the result is typically binary: object or background. To allow for direct comparison,

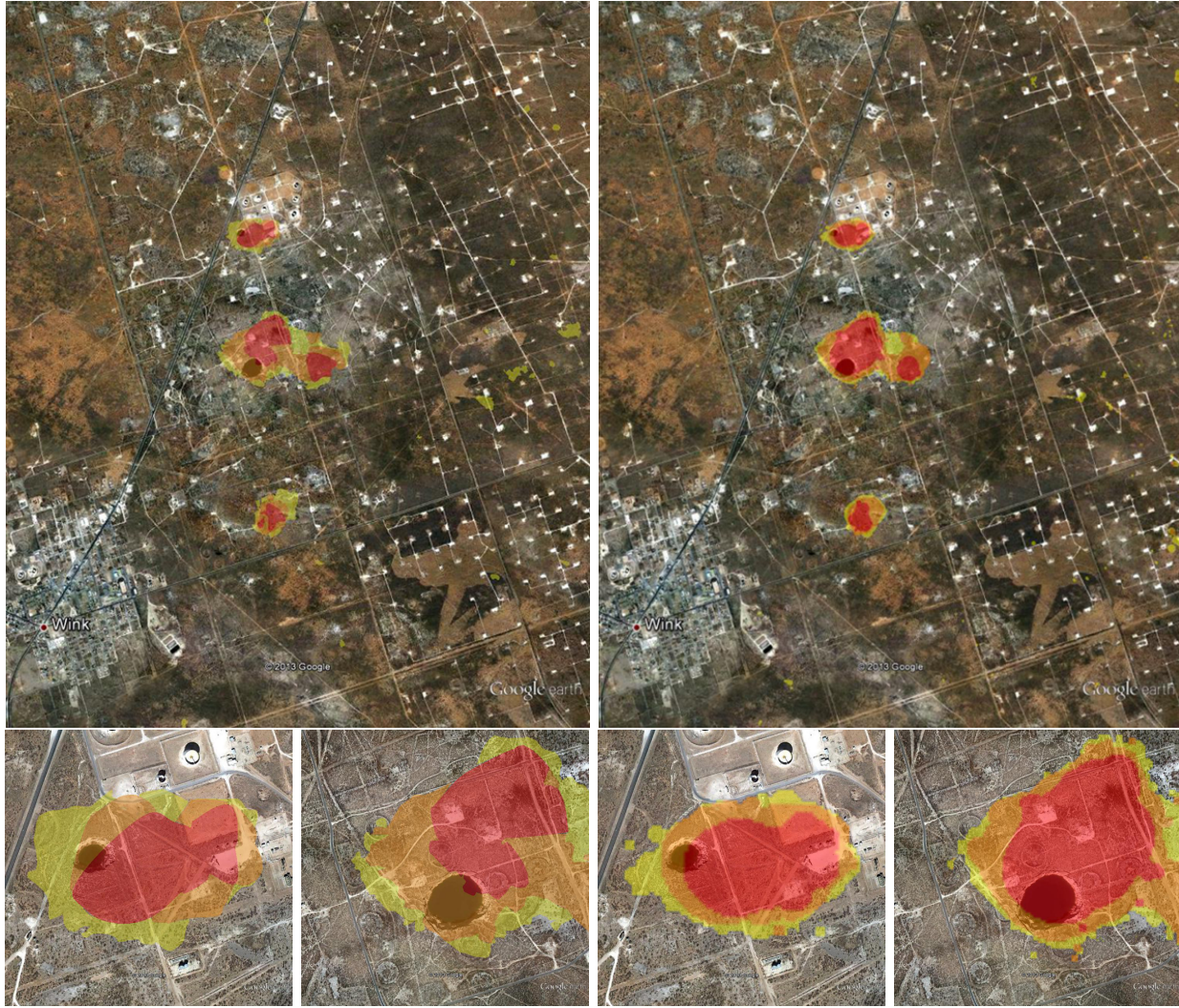


Figure 3.7: Direct comparison between the (left) graph cut algorithm and (right) spatiotemporal matching approach for the Wink Sinks dataset with details for comparison with the collapsed sections of Wink 1 and Wink 2. In all the images the color mean high (red), moderate (orange) and slight (yellow) risk (see text for discussion).

we modified the graph cut approach and iteratively run it for several average displacement window values ($[d_{min}, d_{max}]$) (see section 3.2: *Graph cut approach*) assigning proportionally higher risk to those regions that resulted segmented in multiple levels and for values of the average displacement indicating larger growth. Furthermore, to include the effect of the temporal evolution of the displacements, for each three consecutive time frames in the dataset we evaluated the average displacements and repeated the analysis just described.

The normalized results were then combined by averaging, and the binning evaluated using the same risk intervals used for the spatiotemporal data [120].

Table 3.4: Graph-cut and spatiotemporal matching vs. collapsed ground truth

<i>Approach</i>	Wink 1			Wink 2		
	<i>High</i>	<i>Moderate</i>	<i>Slight</i>	<i>High</i>	<i>Moderate</i>	<i>Slight</i>
Graph cut	37.5%	57.8%	4.7%	17.6%	82.4%	-
Spatiotemporal	52.6%	47.4%	-	81.6%	18.4%	-

Table 3.4 provides a quantitative summary of the results shown in Fig. 3.7. Both algorithms do a reasonable job in detecting these large features and the overall subsidence areas. It could be argue, especially in the case of Wink 2, that the spatiotemporal matching provides a better detection of the actual collapsed area although, as we previously mentioned, often the region affected by subsidence is much larger and not necessarily symmetrically placed with respect to the collapsed section. Nevertheless, in the case of Wink 2 (see section B.2: *Wink Sinks*), there are several scatterers, showing large displacement growth over time, placed right over the collapsed section that the graph cut approach fails to identify as high risk.

For large features, such as the Wink Sinks, both algorithms provide a reasonable detection but these are not the type of features we are interested in detecting. What we are looking for are the initial displacements that might be indicative of early subsidence development. In particular, we are interested in detecting scatterers undergoing a specific temporal displacement trend. Although the modified graph cut approach takes into account temporal variations by averaging cuts evaluated over the mean behavior of neighboring time slices, it still fails at discriminating between different temporal behaviors.

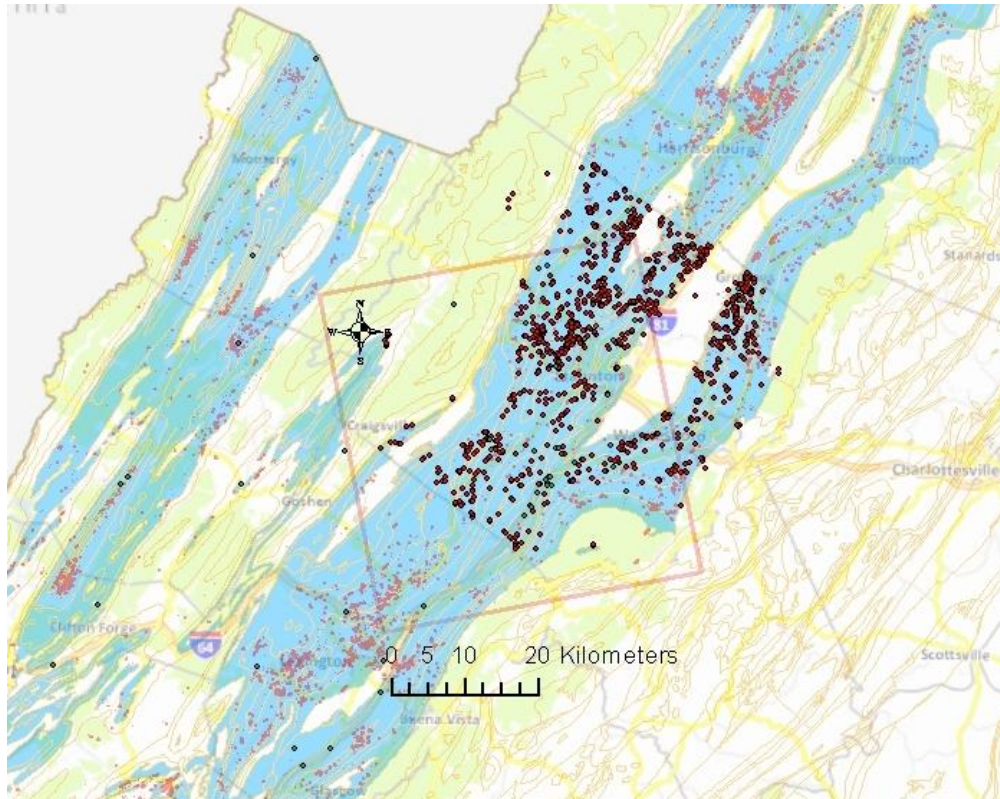


Figure 3.8: Project area of interest $40 \times 40 \text{ km}^2$ located in Virginia, United States. Areas prone to sinkhole formation is highlighted in light blue while previously repaired sinkholes are indicated by dark red dots. (From [78] ©2013 Taylor and Francis)

Virginia

As part of the project, we selected a region of about $40 \times 40 \text{ km}^2$ because of the diversity of geological conditions and history of subsidence.

The area, represented in Fig. 3.8, is centered roughly in the locality of Middlebrook, Augusta County, Virginia ($38^\circ 04' 24.79'' \text{ N}$ - $79^\circ 11' 30.88'' \text{ W}$) and is a tectonically complex area spanning the Valley and Blue Ridge physiographic provinces [121].

Geological ages ranging from Holocene sediments to Precambrian granulite gneiss [122] with frequent nonconformities, are represented in this area of study. The predominant tectonic framework consists of eastward-dipping thrusts faults and decollements related to

repeated orogenic cycles [123]. This area contains carbonate, non-carbonate clastic, and metamorphic terrains, resulting in both rock slope stability and karst geohazards. The karst areas range in age from Cambrian to Devonian and formed during the Taconic and Acadian Orogenies and their associated divergent and inter-orogenic periods. Karst lithologies consist mainly of limestone and dolostone while non-carbonate clastic lithologies consist of occasionally interbedded shales, slitstone, conglomerates, and sandstone, and the metamorphic lithologies consist of charnockite, granulite gneiss, quartzite, and greenschist and blueschist-grade metabasalt.

In central Virginia, the carbonate formations generally present a well-developed orthogonal joint system, which, combined with bedding-plane faulting and tectonic deformation, described above, create the potential for secondary porosity and preferential groundwater flow [124]. The majority of the karst features encountered within the Valley and Ridge physiographic province of Virginia are of the solution and cover subsidence sinkhole variety. As a consequence we expect a slowly forming typology progressively developing a karst topography reflecting the underlying bedrock geology and tectonics [125]. In Fig. 3.8 karst terrains are illustrated in light blue with areas of known sinkhole locations and previously repaired sinkholes identified by red markers.

One of the advantages of the algorithm that we developed is the level of flexibility it offers to include external information in the post-processing phase. Consider Fig. 3.9(a). This diagram illustrates the basic flow behind a typical implementation.

- **Feature identification.** The process starts by defining a specific feature of interest that we would like to detect within a dataset.

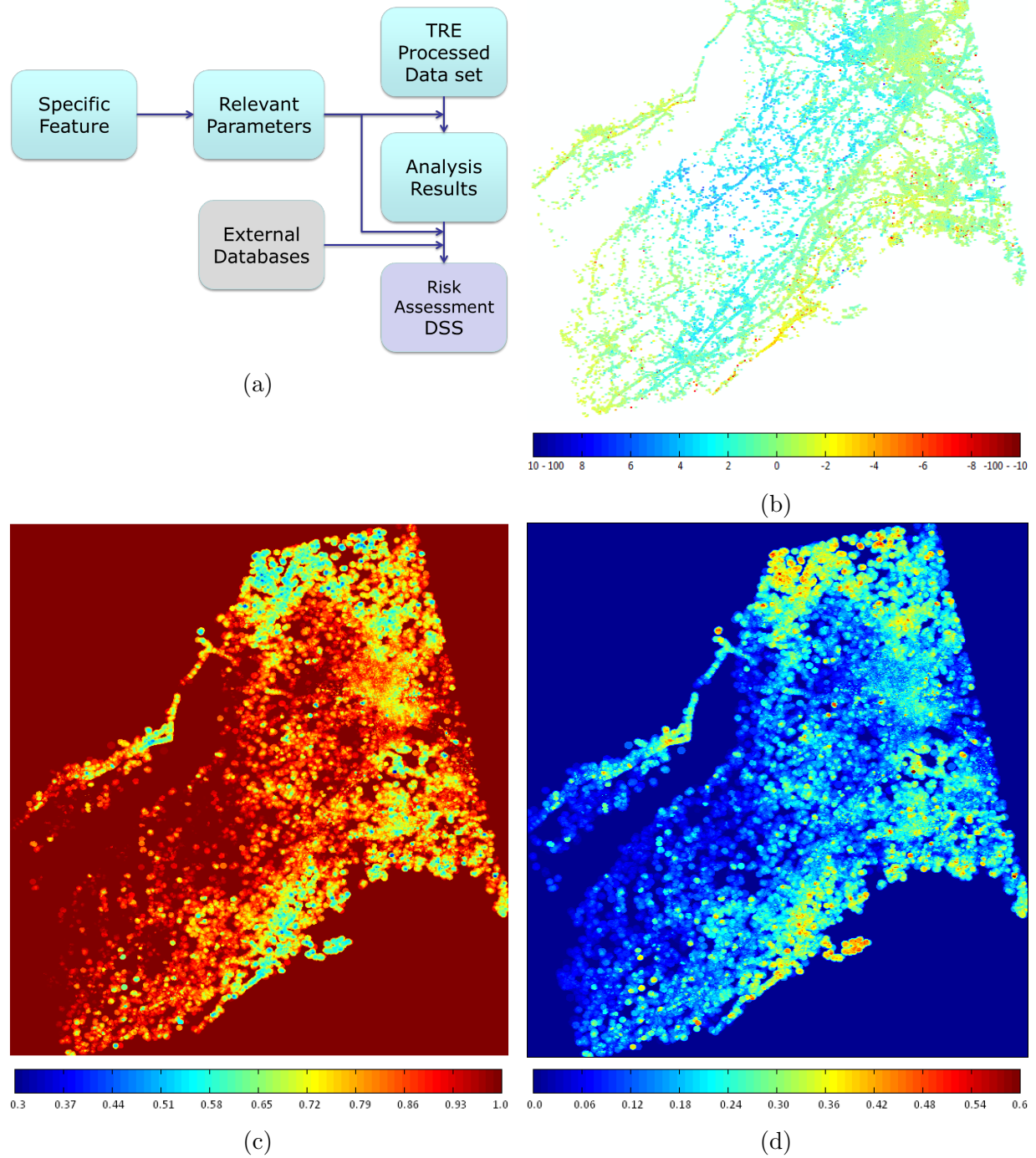


Figure 3.9: (a) Processing flow diagram. (b) Virginia point cloud dataset showing location and average displacement velocity in mm/year of the PS. (c) Propagated residual map. (d) Propagated minimum-risk map. ((b) and (d) from [78] ©2013 Taylor and Francis)

- **Parametrization.** The feature is analyzed and a parametrized model is created by identifying the n most relevant parameters.
- **Dataset acquisition.** A dataset is acquired from a remote sensing provider. In our case we use point cloud datasets from TRE Canada. One point to notice is that this process is not limited to remote sensing data and it can be easily extended to any point cloud dataset as long as the feature of interest can be represented by a parametrized model. Furthermore, the dataset is not required to be a point cloud, our approach can be used for uniformly sampled datasets such as stack of digital images.
- **Analysis.** After selecting ranges and steps for each of the n parameters, the dataset is analyzed. The result is an n -dimensional residual matrix. The minima in this matrix identify the parameter vectors that provide good model matching. In our specific case, once projected on the geographical coordinate space, the minima identify location where the measured displacement behave accordingly to our model. One of the advantages of this approach is that, for each residual value, all the n parameters are available. For our dataset, this means that information regarding location, size and growth speed are identified for both the best matches and every other geographic location.
- **External sources integration.** The extracted parameters can be used during post processing to generate secondary maps where external information are queried from existing sources using the parameters as indexes.
- **Decision support systems (DSS) - Risk assessment.** The ability to include external data in the generation of accessory maps facilitates the development of decision

support system algorithms. For example, a user could generate a map where actual residual values are weighted according to how close the potential sinkhole is to a major transportation highway or infrastructure or, as another example, generate a map where the risk function is defined based on the location with respect to known karst terrain and growth speeds or sizes that closely resembles that of existing features.

To illustrate this process, we used the Virginia point cloud dataset Fig. 3.9(b). Parameters were chose to provide coverage for the entire dataset extension (see section B.3: *Virginia*) and allow detection of significant subsidence (see Table 3.5).

Table 3.5: Parameter space limits and increments for the Virginia dataset

Parameter	Minimum	Maximum	Increment
x_0	0 m	43,235 m	2.5 m
y_0	0 m	51,768 m	2.5 m
α	-5 mm/month	-1 mm/month	0.25 mm/month
σ	5 m	100 m	5 m

We then analyzed the dataset with our approach and developed a very simple risk function based on the residual (Fig. 3.9(c)), and the minimizing amplitude values:

$$\rho(\mathbf{p}) = [1 - r(\mathbf{p})] \exp(1/\alpha_{\mathbf{p}}) \quad (3.16)$$

where $\alpha_{\mathbf{p}}$ is the α corresponding to the specific parameter space point \mathbf{p} considered.

The rationale behind the form of (3.16) is to provide an increased risk factor for those regions matching faster-growing templates (larger values of $|\alpha|$ in (3.8)) while reducing false detection by minimizing the risk factor in case of regions showing a slow subsidence grow rate.

Table 3.6: Categories of subsidence evidence based on ground inspection (From [78] ©2013 Taylor and Francis)

Categories	Infrastructure	Geomorphology
Absolute (A)	Cracks, settlement	Recent non-vegetated scarps
Strong (S)	Distortions or cracks	Overgrown scarps
Weak (W)	Repairs or cracks	Geomorphology indicates activity
Possible (P)	Near existing active region	In correct terrain, presence of pinnacles
None (N)	No or negative confirmation	No or negative confirmation

The risk map was then collapsed according to $m_\rho(x_0, y_0) = \max_{\alpha, \sigma}[\rho(\mathbf{p})]$ and propagated following $\sigma_{m_\rho} = \arg \max_{\sigma} \{ \max_{\alpha} [\rho(\mathbf{p})] \}$. The results are illustrated in Fig. 3.9(d). It is interesting to compare these results with the original scatterer displacement rates, in mm/year , shown in Fig. 3.9(a). Although, as expected, it is possible to identify a general correspondence between subsiding regions and areas of higher risk, the direct comparison highlights the ability of our method to distinguish between simple changes in elevation and subsidence following a modeled behavior.

We classified the regions within the propagated risk map into four separated categories according to the risk factor: severe ($m_\rho \geq 0.475$), moderate ($0.4 \leq m_\rho \leq 0.475$), slight ($0.35 \leq m_\rho \leq 0.4$), and no risk ($m_\rho \leq 0.35$). These ranges were selected based on the observed behavior of the risk function on the Wink Sinks dataset. Following this classification, we produced a list of potential subsidence candidates out of which we selected a sample of 32 locations for ground validation: seven identified as severe, 15 moderate, and 10 as slight risk.

To allow for a more direct comparison between the algorithm detection and the ground validation results, with the help of VDOT geologists, we classified the evidence of subsidence into five distinct categories based on the observed infrastructure deformation and geomorphology at the inspected locations. The categories and the associated evidences are listed in

Table 3.7: Results of ground validation on Virginia data (From [78] ©2013 Taylor and Francis)

Risk	Evaluated	A	S	W	P	N
Severe	7	4 (57%)	2 (29%)	-	-	1 (14%)
Moderate	15	8 (54%)	2 (13%)	2 (13%)	1 (7%)	2 (13%)
Slight	10	5 (50%)	4 (40%)	-	1 (10%)	-
<i>Total</i>	<i>32</i>	<i>17 (53%)</i>	<i>8 (25%)</i>	<i>2 (6%)</i>	<i>2 (6%)</i>	<i>3 (10%)</i>

Table 3.6.

Table 3.7 shows a direct comparison between the algorithm classification and the ground validation. These results show that 78% of the locations selected between those identified by our algorithm presents strong evidence of subsidence.

These results show that we can successfully isolate regions showing a specific type of subsidence behavior nevertheless, we would like to point out that, although definite signs of subsidence were identified during the ground validation phase for the Virginia results, it was difficult to assess the actual cause behind the subsidence. This is mainly due to the high sensitivity provided by the permanent scatterer technique: the largest observed displacements over the acquisition period were of the order of 30-40 mm. So, although the behavior of such regions showed a coherent movement in line with our model, it will require a longer observation campaign to precisely determine the cause of the observed subsidence and provide direct numerical comparison.

3.3.4 Subsidence model discussion

Since the detection algorithm we developed is independent of the specific parametric model we select as template, we can use it to indirectly evaluate the quality of the sinkhole model

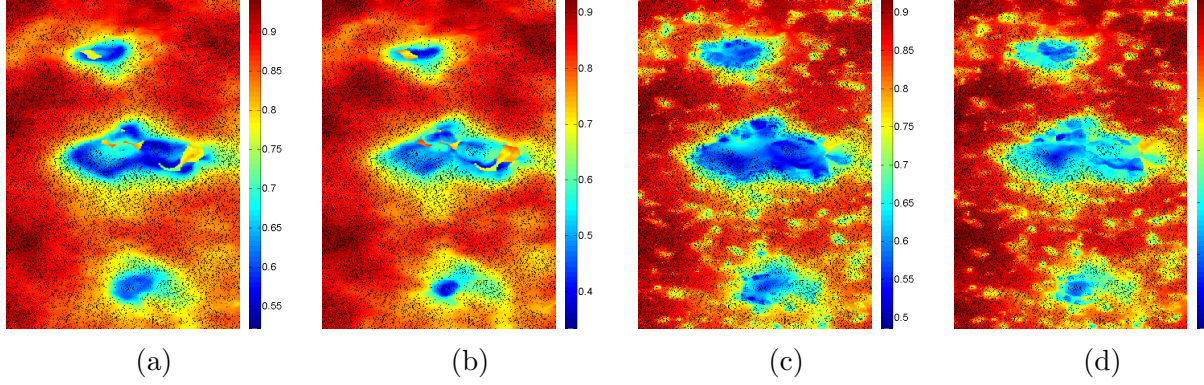


Figure 3.10: Comparison between the results of the proposed detection algorithm on the Wink dataset using variations on the model described in (3.8). For comparison, the location of the available scatterer is marked with black dots. The displayed maps are not propagated to allow for better comparison. (a) Static model with fixed amplitude $\alpha = -104.63$ mm and $\sigma = 95$ m. (b) Model with adapting amplitude α and fixed $\sigma = 95$ m. (c) Model with fixed amplitude $\alpha = -104.63$ mm and adapting σ . (d) Fully adaptive model as described by (3.8). (From [78] ©2013 Taylor and Francis)

itself (3.8) by comparing the results of the detection over the Wink Sinks area⁵ using modified versions of the subsidence model. The outcomes are illustrated in Fig. 3.10. As expected, reducing the degrees of freedom of the model results in an overall increased residual.

In Fig. 3.10(a), we present the results for a template where the amplitude and size were selected to be the mid-range values of the respective parameter space search ranges ($\alpha t = -104.3$ mm and $\sigma = 95$ m), effectively resulting in a model describing a fixed subsidence of a given amount not changing in time.

One of the major advantages of this fixed parameter configuration is the extreme reduction in computational burden, which allows the analysis of large datasets to be conducted in a relatively short time. The drawbacks are the loss in resolution (spatial details are averaged over the fixed size of the template) and the reduced ability to cope with low data density areas.

This is clearly visible in the top and center sinkhole regions where section with large residual

⁵The reasons behind the choice of the Wink Sinks dataset rather than, for example, the Virginia one, is once again dictated by the fact that it can be considered as "ground truth" (see section 2.2.2: *The Winkler sinkholes (Wink Sinks)*).

value appear in correspondence of areas with low scatterer density. The same issues are evident in Fig. 3.10(b), where the model was modified to allow the search for the best-fitting growth rate α while keeping the scale fixed ($\sigma = 95$ m).

Fast screening of large areas can also be obtained by using a model where the amplitude is kept constant at the expected mid-range ($\alpha t = -104.3$ mm in our case) while full parameter search is allowed for the scale. This effectively corresponds to a search for the best-fitting static model where the temporal evolution is ignored. Results for this model are shown in Fig. 3.10(c). In this case, the local spatial details are preserved allowing for localization of potential features but the temporal information is lost, greatly diminishing the classification capabilities of the detection algorithm as indicated by the overall higher residual values. Nevertheless, it is possible to envision a two stages process where large areas are first quickly scanned for spatial features coherent with the desired model and then only matching regions are selected for a full parameter search, potentially reducing the computational time (see chapter 4: *Performance analysis*).

Another interesting detail to notice is that only the fully adaptive model clearly identifies the region of subsidence corresponding to Wink 4 and separates it from the Wink 2 area. We believe this might be due to the fact that, although scatterers present between the two features are potentially aligned spatially to a subsidence model, their temporal behavior does not conform to that of a developing subsidence. This subtle difference can only be captured by a full adaptive model (3.8) that considers both spatial and temporal evolution of the displacement.

In order to quantitatively compare the performances of the different subsidence models variants described above, we selected two subregions for each of Wink 1, Wink 2, and Wink

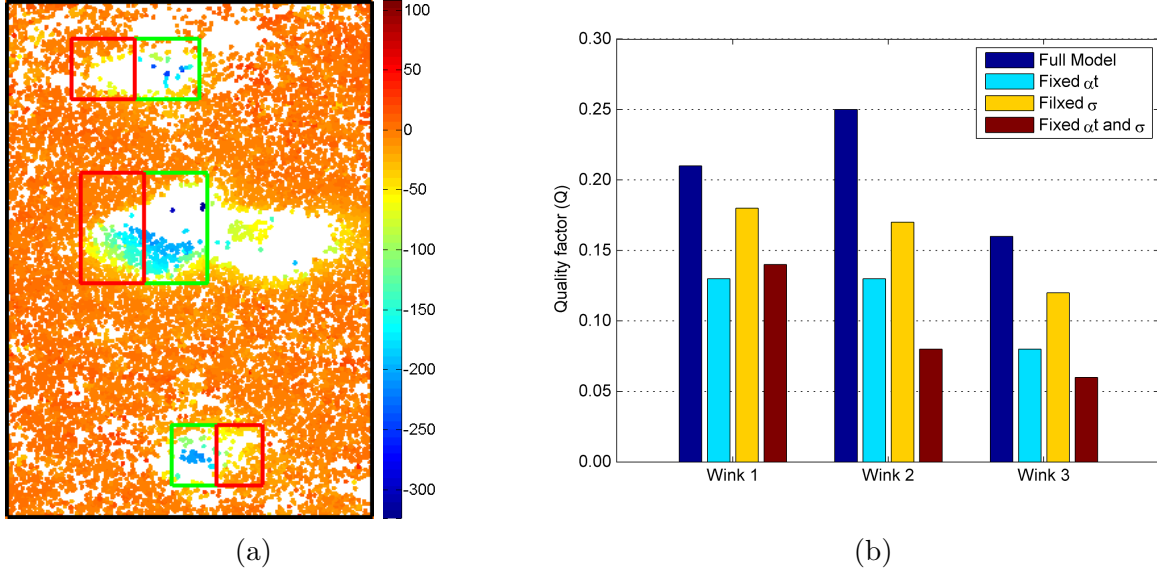


Figure 3.11: Comparison between different subsidence models performances. (a) Selection of on-target (green) and off-target (red) regions for each subsidence region in the Wink Sinks dataset. The dots represent the location of each scatterer and are color coded based on the displacement (in mm) measured at $t = 68.68$ months. (b) Quality factor (see text) evaluated for each of the presented models. (From [78] ©2013 Taylor and Francis)

3 (Fig. 3.11)⁶: the green box identifies the subregion where the potential center for the subsidence phenomenon might be located as extrapolated from the recorded displacements while the red box identifies a region where a lower level of subsidence is occurring.

Two desirable characteristics of a detection algorithm are to provide the lowest possible residual when on-target while achieving the largest possible discrimination between the on-target and off-target conditions. To measure the performance of our algorithm when using the different models, we defined a quality factor $Q = |\mu_{TON} - \mu_{TOFF}| / \mu_{TON}$, where μ_{TON} and μ_{TOFF} are the average residuals evaluated over the on-target and off-target areas respectively.

For each of the models and each of the sinkhole areas, we evaluated the quality factor and the results are illustrated by the histogram in Fig. 3.11(b), where it is clear that the model implementing full parameter search outperforms the other variants.

⁶The area corresponding to Wink 4 was not included in this evaluation due to the high sparsity of available scatterers.

3.4 Summary

We believe that the algorithm we developed, used in conjunction with a the spatiotemporal template described by (3.8), provides a valuable tool that can help in the detection of precursory subsiding motions that can be associated with the development of sinkholes but, as we have briefly touched upon in the model variants analysis, is quite expensive computationally.

In the next chapter we will discuss this issue and describe an approach that can potentially reduce the computational burden by pre-selecting areas of interest where the temporal behavior of the scatterers "differs" from the average behavior observed over the entire scene, and by providing a fast feature detection method with the goal of restricting the application of the spatiotemporal matching algorithm.

Chapter 4

Performance analysis

4.1 Introduction

The computational cost of the spatiotemporal matching algorithm depends on two main factors: the granularity of the parameter space and the number of points within the dataset.

If we consider the ranges and steps size required to provide an accurate analysis for the Wink Sinks (Table 3.3) and the Virginia (Table 3.5) datasets, we can round the number of elementary cells defined within each parameter space to approximately 180 million for the Wink Sinks and 122 billion for Virginia. At the same time, the two datasets contain a large number of points: 93,513 over 55.88 km² for each of the 22 time slices resulting in a total of about 2 million points (36,82 pts/km²) for the Wink Sinks and 296,121 over 1,600 km² for each of the 32 scenes for a total of about 9.5 million points (5,92 pts/km²) for Virginia¹.

¹If we envision an actual implementation of this approach at one of the State DOTs, we can imagine that a new set of scenes, potentially covering the entire State, will be acquired at regular intervals largely increasing both the ranges of the parameters and the number of data points.

The direct implementation of the spatiotemporal matching algorithm requires the evaluation of the residual measure (3.9), for each cell in the parameter space and for every data point within the region of influence (on average 17.9 thousands for the Wink Sinks and 800 for the Virginia dataset) resulting in a total of $2.9 \cdot 10^{12}$ evaluations for the Wink Sinks and $86.4 \cdot 10^{12}$ for Virginia.

Following these consideration, it becomes apparent that, with larger datasets and more complex models, the currently proposed approach would require increasingly long computation times to evaluate the residual matrix $r(\mathbf{p})$ for parameter spaces sampled densely enough to guarantee proper detection.

Although this burden can be mitigated by taking advantage of the highly parallel nature of the algorithm, or by implementing part of the calculation in the powerful graphics processing units available in nowadays video card, we devised an approach that can improve the detection performances.

4.2 Parameter space reduction

The proposed method [126] provides two tools that can be used, together or independently, to address the main bottlenecks of the spatiotemporal matching algorithm: the size of the parameter space and the expensive computation of the residual matrix $r(\mathbf{p})$. The first is based on a screening stage where areas that behave differently from the average behavior of the entire scene are detected, whereas the second provides faster features localization by implementing an approach based on the concept of mutual information.

In more detail, if we consider our datasets, for each point we have an associated displacement time series recording the changes in elevation with respect to the master reference measurement (for which $d(\mathbf{x}_i, 0) = 0 \forall \mathbf{x}_i$).

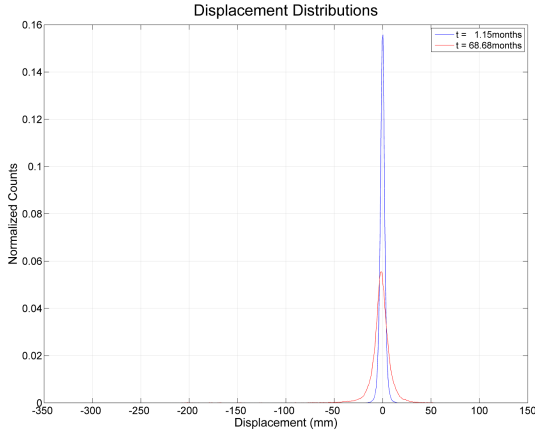


Figure 4.1: Illustration of the distribution of the displacements for the entire dataset at $t = 1.15$ months (blue) and $t = 68.68$ months (red). (From [126] ©2013 SPIE)

Looking at the normalized distribution of the displacements for the entire dataset at a given time, it can be observed that, starting from a $\delta(0) = 1$ at the reference time $t = 0$, as t increases, the distribution spreads due to changes in the displacements values. This is illustrated in Fig. 4.1 where two normalized

distribution are shown for $t = 1.15$ months (blue) and $t = 68.68$ months (red). After

careful examination, it can be seen that the latter has larger variance and that the mean has drifted towards the negative displacement values, in line with the development of subsiding features within the areas.

By analyzing the global distribution, it is difficult to evaluate size, location and amplitude of potential events due to the fact that the subsets of points displaying a subsiding trend is typically small and distributed in random patches within the dataset.

To increase detectability and provide localization, we divided the dataset into N smaller rectangular non-overlapping subsets, or blocks, $(S_n \ n = 1, \dots, N)$ spanning the entire dataset².

Fig. 4.2(a) shows a possible subdivision of the Wink Sinks dataset with $N = 2601$.

²The choice of subdivision is arbitrary although it should guarantee coverage of the entire dataset. Furthermore, the area covered by each block should be of the order of the size of the expected feature or smaller so insure that the statistical properties of *active* differ from those of *inactive* blocks (see explanation in text).

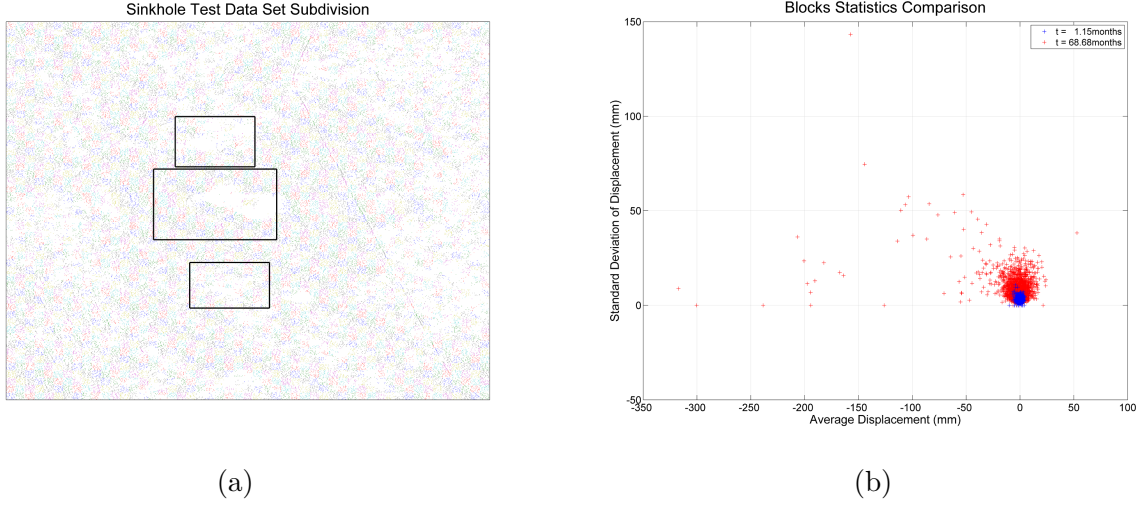


Figure 4.2: (a) Example of possible subdivision of the Wink Sinks dataset into $N = 2601$ non-overlapping rectangular subsets identified by different cycling colors. The black rectangles show the regions where the subsidence activity is located. (b) Scatter plot of standard deviation vs. mean for the displacements within each block shown at $t = 1.15$ months (blue) and $t = 68.68$ months (red). (From [126] ©2013 SPIE)

If we now consider each of the S_n obtained from the subdivision process, we expect the distribution of the displacements associated with active blocks to differ from that of inactive blocks where, with the term *active*, we indicate regions undergoing subsidence phenomena. In particular we expect an amplified version of the behavior observed for the entire dataset: the distribution for active blocks should show growing variance and an increasingly negative mean whereas, for inactive block, we expect the distribution to remain fairly narrow and show no preferential mean direction.

To better analyze this behavior, we evaluated the mean ($\mu_n^{(t)}$) and standard deviation ($\sigma_n^{(t)}$) for each block and represented them in a scatter plot (Fig. 4.2(b)) for two different times $t = 1.15$ months (blue) and $t = 68.68$ months (red). For the earlier time, all the blocks present similar statistics and are concentrated around the origin. As time increases, the statistic for the majority of the block (inactive) remains fairly clustered around the origin

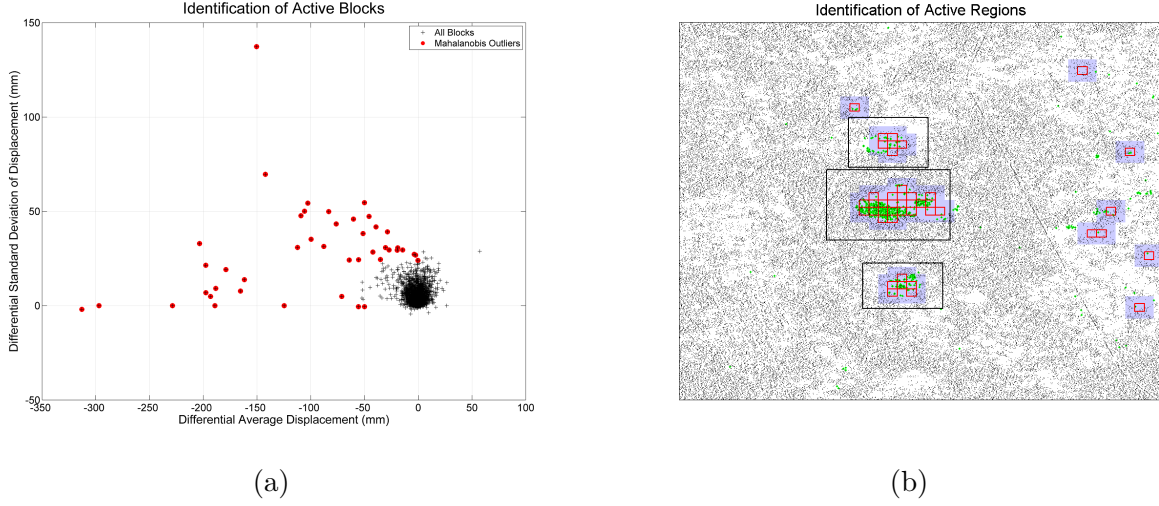


Figure 4.3: (a) Scatter plot of the differential mean and standard deviation of the displacement for each block in the dataset. The outliers according to the Mahalanobis distance are identified in red. (b) The Wink Sinks dataset with identified (solid red) the block selected as outliers (active). For comparison we highlighted (green) the location of scatterers with an overall displacement larger than -50 mm. In light blue are indicated the block selected for analysis. (From [126] ©2013 SPIE)

displaying a symmetric distribution respect to the mean and a slightly increased standard deviation whereas, for the active blocs, it drifts away from the origin towards negative means and largely increased standard deviations with a behavior indicative of developing subsidence.

To correct for displacement statistic potentially introduced by the SAR imaging process or by the SqueeSAR post processing and not due to developing features, we used the displacements measured at $t = 1.15$ months and subtracted them, on a block-by-block basis, from the measurements obtained at $t = 68.68$ months. The justification for this correction is that after a short time interval, in our case 1.15 months, we expect to observe almost no subsidence changes while potential systematic error introduced by the processing would already be present. The result is shown with black markers in Fig. 4.3(a).

Under the assumption that the majority of the blocks does not contain active features (which is usually the case when large areas are analyzed) and that the displacement distri-

butions represent natural statistics typical of each individual block characteristics, we can consider our data $(\xi_n^{(t)} = (\mu_n^{(t)}, \sigma_n^{(t)}))$ to follow a two-dimensional multivariate normal distribution. We can then look for outliers in our dataset and tag them as active blocks. To increase sensitivity, we carried this analysis for the displacement measured at $t = 68.68$ months.

Since our goal is to quickly identify possible active regions where the full spatiotemporal algorithm will be applied, between the many methods available for multivariate outliers detection [127], we opted for the Mahalanobis distance [128] that provides a fast measure of how dissimilar a test sample set is from an existing one.

The Mahalanobis distance estimates the sample mean vector $(\bar{\xi}^{(t)})$ and the covariance matrix $(\Sigma^{(t)})$ of the multivariate distribution given by the sample mean and standard deviation of our blocks S_n and, for each data point $\xi_n^{(t)}$ in the sample, evaluates a distance given by $M_n^{(t)} = \sqrt{(\xi_n^{(t)} - \bar{\xi}^{(t)})^T (\Sigma^{(t)})^{-1} (\xi_n^{(t)} - \bar{\xi}^{(t)})}$.

Given the above assumption on the distribution of the $\xi_n^{(t)}$, the Mahalanobis distance will follow a Chi-Squared distribution with two degrees of freedom (χ_2^2). We can then use a p-test to detect outliers. We chose a p-value³ of 0.01 as significance threshold⁴.

4.2.1 Results and comparison with spatiotemporal matching

In Fig. 4.3(a) we identify with a red-circled marker the $\xi_n^{(t)}$ selected as outliers by the outlined process. The corresponding blocks S_n are identified in Fig. 4.3(b) by red rectangle. Most of the selected blocks fall within the known active regions (identified by black rectangles) and

³The p-value represents the probability that the sample under analysis is coming from the considered distribution. In our case, the probability that a given $\xi_n^{(t)}$ is coming from the distribution generated by mostly inactive blocks.

⁴We use a significance threshold of 0.01 which, according to readily available critical value tables for the Chi-Squared distribution, means that, if $M_n^{(t)}$ is larger than 13.82, we can reject the hypothesis that the sample is coming from the inactive distribution and we can label it as an outlier.

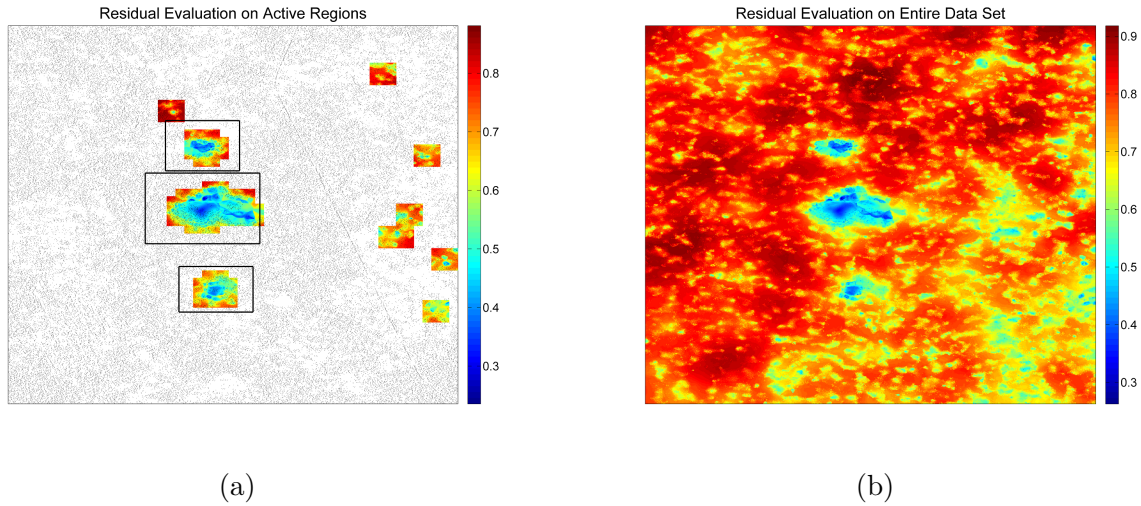


Figure 4.4: Comparison between the residual maps resulting from application of the spatiotemporal matching algorithm to (a) the active regions and (b) the entire dataset. (From [126] ©2013 SPIE)

near data points showing overall displacements larger than -50 mm (green). To insure proper coverage of the detected features, we extend the detection to include the direct neighbors of the identified block. The final area selected for subsequent analysis by the spatiotemporal matching algorithm is identified in light blue.

The speed up in the detection process is, to the first order, proportional to the ratio of the original area to the area covered by the blocks selected for processing. As shown in Fig. 4.3(b), this can be quite dramatic.

We verified the detection results by comparing the residual maps obtained by applying the spatiotemporal matching algorithm only to the active blocks and to the entire dataset (Fig. 4.4) and confirmed that the outliers detection step has correctly identified the active regions.

The main advantage of using this preprocessing step is to reduce the computational burden that would result from analyzing large datasets with complex models requiring several parameters. Furthermore this mechanism is independent of the complexity of the model and

directly compares the displacement statistic between blocks of data singling out, for further analysis, regions exhibiting a behavior different from the overall average.

The strength of this approach can also be one of its weaknesses. For the detection to work, it is required for subsiding regions to present a statistical signature different from the one of inactive blocks. If we consider that the nature of the datasets we are most likely to use is differential, we expect the detection algorithm to fail at differentiating between active and inactive regions during the initial acquisition period⁵. On the other hand the more images in a SAR stack, the higher the accuracy of the results obtained by the SqueeSAR algorithm and, typically, stacks are not processed until at least 15-20 frames have been acquired, with optimal results obtained with 30 or more frames. If we consider that, to increase phase traceability, a SAR frame is acquired every two weeks, we can expect that, after the one to two years necessary to build up the stack, relevant subsiding features will present a large enough signature to be detected.

Another option that can be considered, when small variation between blocks is expected, is to increase the p-value. This would effectively lower the threshold for the Mahalanobis distance and increase the number of blocks considered outliers. At the limit of zero distance, all blocks would be considered outliers and scheduled for full analysis, reverting back to the full spatiotemporal matching algorithm.

⁵We have noticed a similar behavior while trying to evaluate the spatiotemporal model of sinkholes (see section 2.3: *The model*).

4.3 Mutual information metric

Although the proposed method can provide a considerable increase in efficiency, with larger datasets and complex models, or in the case of low discriminating statistics between active and inactive regions, the screening stage can still identify large areas. Furthermore, the mere identification of a different statistical activity does not provide any information about the actual spatial distribution or temporal behavior of the scatterers. We would like to develop an approach to efficiently handle both these scenarios.

While discussing the property of the subsidence model (see section 3.3.4: *Subsidence model discussion*) we noticed that all the different variants provided a considerable decrease in computational time. This was mainly due to the fact that the parameter search over the amplitude and/or scale was replaced by a fixed value. We also noticed that, among the studied models, the one implementing a fixed amplitude template with adaptive scale showed the most promising results by both allowing the localization and preserving the scale of potential features, albeit losing temporal evolution information.

When we tested the performance of the scale-adaptive template, we assumed a certain fixed amplitude based on our a priori knowledge about the feature we were looking for. Obviously when analyzing general datasets we do not have this advantage. Furthermore, the proportional match measure we introduced (3.9) is designed to penalize, with high residual, any spatiotemporal mismatch between template and data.

To address both these issues, we designed an amplitude invariant approach, based on the concept of mutual information, that provides the advantages of a fast matching algorithm while preserving local scale information. This approach addresses both scenarios described

at the beginning of this section by providing a tool that can be either used to screen entire datasets to identify displacement regions spatially conforming to the desired model, or, in case of large datasets, as a follow up to the activity screening technique, previously illustrated, to provide information about the spatial behavior of the subsidence in the regions identified as active and further refine the selection of areas to be analyzed with the full spatiotemporal matching algorithm. In the rest of this section we will describe this approach and show its results on the Wink Sinks dataset when used as further refinement after the activity screening step.

Measures based on mutual information (I) and, in particular, normalized mutual information (NMI) are commonly used to evaluate hidden dependencies between datasets that might not be evident in covariance-based measures, such as correlation, which address linear dependency [129] [130]. Within our framework this feature can be used to evaluate the dependency between the expected displacements, according to a specific model, and the observed ones.

In our approach, we use the following definition for I and NMI [131] [132]:

$$I(D; G) = \sum_{d_i \in D} \sum_{g_i \in G} p(d_i, g_i) \log \frac{p(d_i, g_i)}{p(d_i)p(g_i)} \quad (4.1)$$

$$NMI(D; G) = NMI_{max}(D; G) = \frac{I(D; G)}{\max[H(D), H(G)]} \quad (4.2)$$

where D is the set of the actual displacement values d_i , G is the set containing the expected displacements g_i according to the subsidence model, $p(d_i, g_i)$ is the joint probability mass function, estimated using an optimized kernel method [133], $p(x_i)$ are the marginals evaluated

from the joint probability (where x_i can indicate either d_i or g_i), and $H(X)$ are the entropies of each dataset evaluated as $H(X) = -\sum_i p(x_i) \log(p(x_i))$.

To generate the *NMI* map, we proceeded similarly to the spatiotemporal matching approach but, taking advantage of the scale invariance properties of the *NMI*, we used the previously described variant of the subsidence model (3.8) where the dependence on time was removed. This resulted in a simplified parameter space including the spatial coordinates ($\mathbf{x}_0 = (x_0, y_0)$), the scale (σ), and a fixed amplitude (α) that can be arbitrarily chosen (we selected a value of $\alpha = -150$ mm).

Following the procedure outlined for the spatiotemporal matching algorithm, for each point in the now simplified parameter space, we defined a region of influence of 3σ around its coordinates (x_0, y_0) . After identifying the scatterers within this region, we defined two sets of data D and G . The first set (D) contained the scatterer displacements measured for a specific time slice of the point cloud dataset (in the case of the sinkhole dataset we selected $t = 68.68$ months where the subsidence was more pronounced)⁶. The second dataset (G) was generated by evaluating the simplified template at the spatial coordinates of the elements of D . In other words, for every \mathbf{x}_i in the region of influence, $D = \{d(\mathbf{x}_i, t)\}$ and $G = \{g_{\mathbf{p}}(\mathbf{x}_i, t)\}$ where $\mathbf{p} = [(x_0, y_0), -150 \text{ mm}, \sigma]$ and $t = 68.68$ months.

We then estimated the joint probability mass function $p(d_i, g_i)$, the two marginals ($p(d_i)$, and $p(g_i)$), and the entropies ($H(D)$ and $H(G)$). These quantities, together with (4.2), were used to evaluate the normalized mutual information now a function of the spatial coordinates and the size of the simplified template: $NMI = NMI(x_0, y_0, \sigma)$.

⁶Generally the last temporal slice available in datasets generated by differential measurements, such as SqueeSAR, is the one where it is possible to find the largest signatures of growing features.

The outcome of the process was a 3-dimensional matrix containing, at each location, the normalized mutual information between the actual and the expected displacement data, generated according to the template corresponding to the parameters associated with the location in the matrix. To generate the final map we used an approach similar to (3.14) and evaluated the maximum intensity projection of the matrix along the scale axis:

$$NMI(x_0, y_0) = \max_{\sigma} [NMI(x_0, y_0, \sigma)].$$

4.3.1 Results and comparison with spatiotemporal matching

We applied this method to the areas of the Wink Sinks dataset identified by the activity screening step. The NMI map was evaluated for each detect block using the time-fixed template with scale σ ranging within $[50 \text{ m}, 150 \text{ m}]$ with a step size of $\Delta\sigma = 1 \text{ m}$.

In Fig. 4.5 we show three NMI maps obtained at three different values of the scale and the final maximum intensity projection along the scale axis. It is interesting to notice how features of different size become salient at different template scales. This is a well known phenomenon that has been studied in depth in the field of image processing [134] and is at the core of several techniques known under the general name of *scale space analysis* [135]. In scale analysis, an image, considered at scale zero, is iteratively convolved with a Gaussian of fixed σ . As a result, at each step, the image is increasingly blurred and the large changes of intensity, that are typically associated with the presence of details, are increasingly smoothed (or diffused). As the iteration are repeated, larger regions start to get affected. In the limit, the entire image becomes uniform.

In our case, instead of convolution with a Gaussian, we evaluate the mutual information

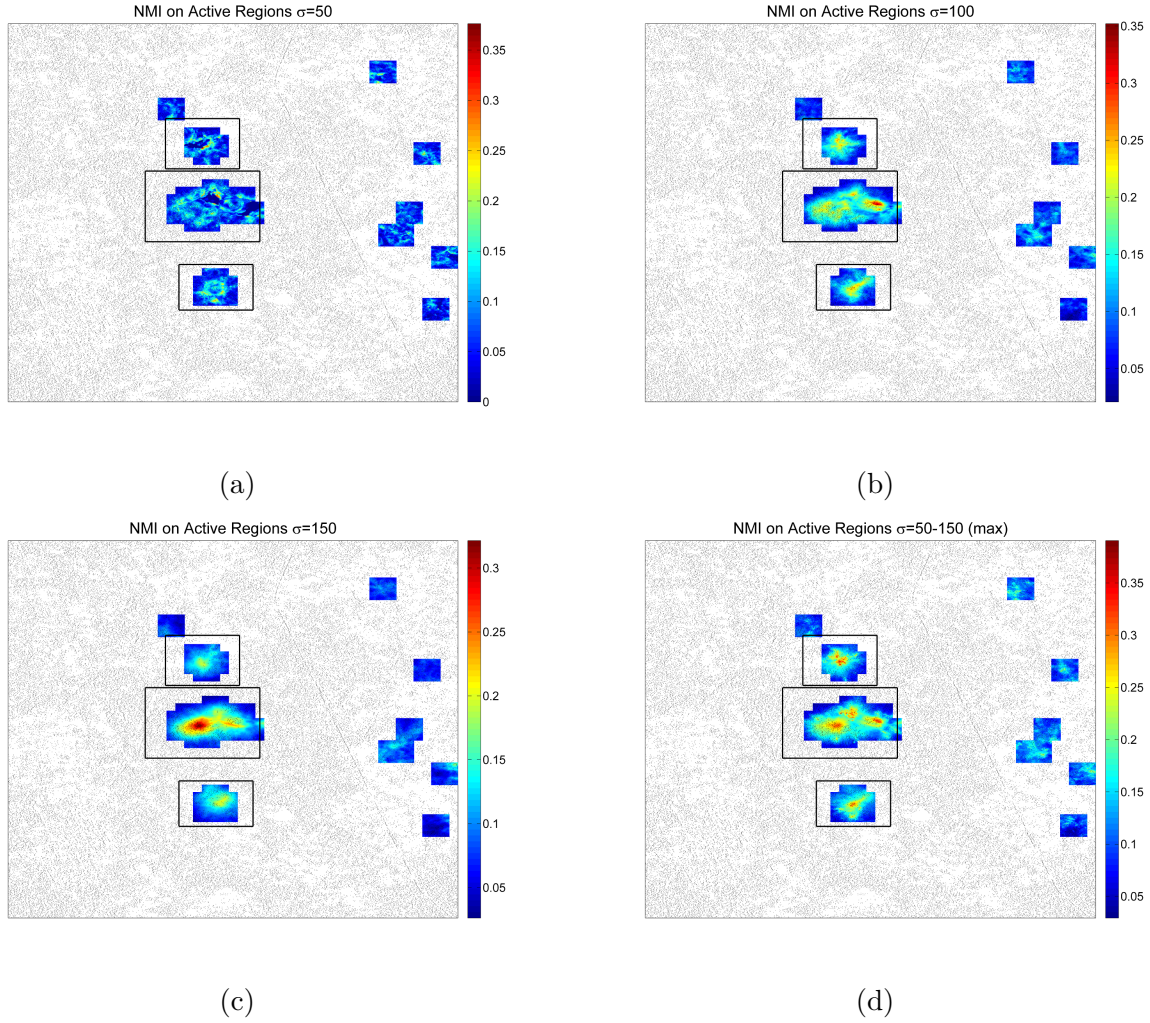


Figure 4.5: Results for the *NMI* analysis with progressively large size templates: (a) $\sigma = 50$ m, (b) $\sigma = 100$ m, and (c) $\sigma = 150$ m. In (d) we show the map resulting from the maximum intensity projection of the *NMI* along the scale (σ) axis. (From [126] ©2013 SPIE)

between a Gaussian template and the point cloud dataset and, instead of iterating with a fixed scale Gaussian, we use templates of increasing scale. The final projection step provides a synthesis of the process by showing, at each coordinate location, the value of the best fit obtained at that location insuring that, if a match was found it will show on the final map independently of its scale. To find out the actual scale at which the maximum occurred, we can simply use the $\arg \max$ function.

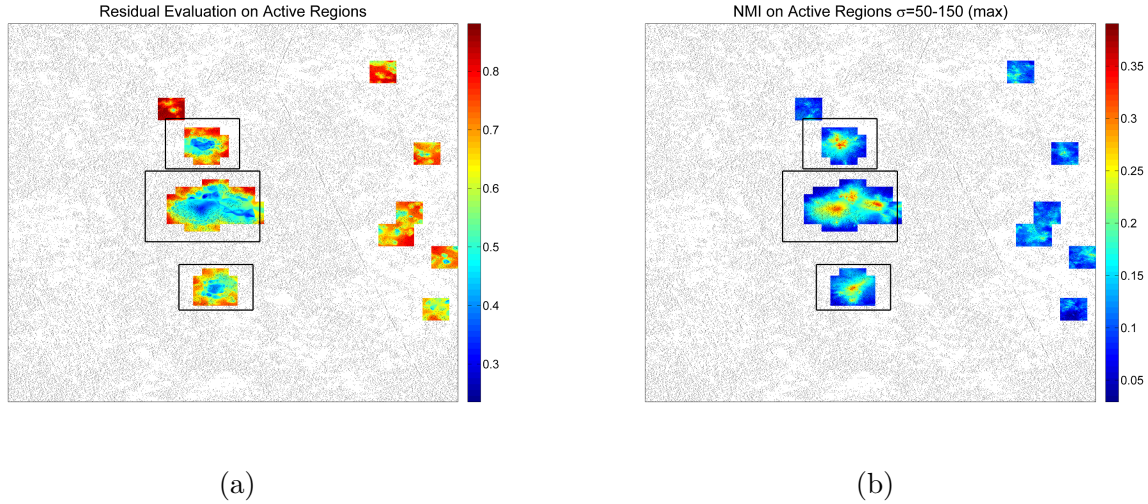


Figure 4.6: Comparison between (a) the spatiotemporal matching and (b) the *NMI* approaches applied to the active regions. (From [126] ©2013 SPIE)

In Fig. 4.6 se show the direct comparison between the results of the spatiotemporal matching algorithm and the *NMI* approach evaluated on detected active regions.

The alignment of the features detected by the two methods confirms that the *NMI* approach can be used, when required, as direct fast screening for entire dataset or as second-tier screening for regions identified as active by the parameter space reduction step. The final selection of the areas where the full spatiotemporal matching algorithm should be implemented could be achieved by simple thresholding of the *NMI* map at a predetermined level of significance.

4.4 Summary

In this chapter we described two techniques that can be used to restrict the area where the spatiotemporal matching algorithm needs to be applied. These two approaches, as well as the spatiotemporal matching method, require the implicit definition of a regular grid underlying

the physical location of the points in our dataset. This is achieved by specifying range and step size for the x_0 and y_0 parameters. To insure that each element in the parameter space includes at most one data point and that all element of the dataset are accounted for, the range is typically selected to cover the entire size of the scene and the step size is chosen to match the highest spatial resolution of the original SAR images. The resulting parameter space is sparsely populated in the spatial coordinates.

As an example, let us consider the spatial parameters for the Virginia dataset (Table 3.5). The ranges were set to $[0 \text{ m}, 43,235 \text{ m}]$ for x_0 and $[0 \text{ m}, 51,768 \text{ m}]$ for y_0 to provide coverage of the imaged area whereas the step size was set to 2.5 m for both to match the $3 \times 3 \text{ m}^2$ ground resolution of the COSMO-SkyMED constellation (see section A.2: *COSMO-SkyMed*). This resulted in a total of about 360 million cells. If we compare this with the actual number of permanent scatters in the dataset, 296,121 (see section B.3: *Virginia*), we can calculate that only 0.083% of the parameter space cells were actually occupied by data points.

In the next chapter we present a method that takes advantage of this inherent sparsity and implements a scale space analysis technique, similar to the one just illustrated, directly within the graph domain without defining any underlying grid.

Chapter 5

Graph signal detection

5.1 Introduction

In the previous chapters we have introduced a method to detect a specific behavior within a spatiotemporal point cloud dataset. In particular we developed an algorithm that can detect ground motion conforming to a model representing the early development of subsidence phenomena. We also identified one of the main drawbacks of our approach, the computational burden. We therefore developed two improvements that, if used as pre-screening stages, would limit the extent of the data to be analyzed.

By comparing the sparsity of the point cloud dataset, derived from the persistent scatterer processing, to the size of the parameter space required to properly identify features of interest, we noticed that only a minimal fraction of the parameter space cells is indeed occupied by actual physical scatters. This prompted us to evaluate alternative approaches to the detection problem.

In this chapter we will expand the background on graph theory we provided earlier (see

section 3.1.1: *Graph theory primer*) to include spectral tools, in particular the graph Fourier transform (GFT). We will then show how similarities between the discrete Fourier transform (DFT) and the GFT can be exploited to define a shift operator within the graph domain. This will allow us to develop a method that provides detection of features of interest, and their scale, directly within the graph without the need to identify and underlying grid as in our other approaches. We will then show the results of this method applied to three different dataset: an artificial one consisting of a Gaussian signal over a regular lattice graph, a Gaussian signal over the graph generated from the coordinates of scatterer within the Wink 3 area, and the actual Wink 3 dataset itself. For this latter case, we will compare the results with those obtained by the graph cut, spatiotemporal matching, and the normalized mutual information metric.

Approaches based on the eigenvalues and eigenvectors of the graph Laplacian (see next section), are generally known as *graph spectral theory* and have been intensively used to study the nature of graphs, their connectivity, and their topological characteristics [136] [137] [138].

More recently, with the increased interest in point cloud dataset applied to robotic vision, the focus has shifted from the analysis of graph properties toward the use of spectral techniques to detect and separate groups of vertices considered similar according to some metric, a process known as *spectral clustering* [139] [92] [140] [93].

In the last decade, with the introduction of social networks - where individuals are typically represented in graphs as multi-featured nodes connected by multi-layered edge structures - new analysis techniques, closely related to traditional signal processing, have been used to detect anomalous network dynamics, with emphasis on local changes in the edge density [141] [142], and transitional behaviors of single nodes with the intent of identifying patterns

and trends [143].

The emerging field of *graph signal processing (GSP)* [144] [145] has shifted the focus of spectral techniques from the analysis of local and global properties of the graph itself to that of the value assumed by *graph signals* on vertices, effectively turning the graph into an underlying connectivity space. The main tool at the basis of this new approach is the *graph Fourier Transform (GFT)* [146]. This novel technique, that is at the foundation of our approach, has sparked the development of graph-based traditional signal processing techniques aimed at the analysis of the local "frequency" behavior of graph signals. Here the graph signal is considered a function of the eigenvectors of one of the matrices representing the connectivity (typically, but not exclusively, the Laplacian). Multiscale geometric wavelets [147], spectral graph wavelet transform [148], and two-channels wavelet filter banks on bipartite graphs¹ [149] are just a few examples of the application of these techniques.

5.1.1 Graph theory II

An important matrix that is at the basis of both the field of *spectral graph theory* and GSP is the *Laplacian* [138]. The basic non-normalized graph Laplacian is defined as $\mathcal{L} = \mathbf{D} - \mathbf{W}$. The name is derived from the fact that the graph Laplacian performs over the graph similarly to the discrete difference counterpart. If we consider again the toy example that we showed

¹Bipartite graphs are graphs where vertices can be divided into two disjoint sets and edges only connect elements of different sets.

before (Fig. 3.1), we can write the Laplacian for that graph as:

$$\mathcal{L} = \begin{array}{c} \text{nodes} \\ \begin{array}{c} A \\ B \\ C \\ D \\ E \end{array} \end{array} \begin{bmatrix} A & B & C & D & E \\ 0.5 & -0.5 & 0 & 0 & 0 \\ -0.5 & 2.25 & -1.0 & -0.75 & 0 \\ 0 & -1.0 & 2 & -1.0 & 0 \\ 0 & -0.75 & 1.0 & 2.05 & -0.5 \\ 0 & 0 & 0 & -0.5 & 0.5 \end{bmatrix}$$

Let us consider a *graph signal* $f_i = f[i] = f[v_i]$ defined over an uniformly weighted *grid graph* ($w_{ij} = 1 \ \forall i, j$), where the nodes are placed on a regular 2D grid with equal spacing $(\Delta x, \Delta y)$ (assumed, without loss of generality, to be $(1, 1)$), and are connected with a 4-neighbors connectivity (NSEW) with each edge of equal weight w_{ij} (again assumed to be 1). With this regular construct, we can associate virtual coordinates to each node $v_i = (x_i, y_j)$ such that the neighbor to the north will be located at $(x_i, y_{j-1}) = (x_i, y_j - \Delta y) = (x_i, y_j - 1)$ and so forth.

If we then consider the result of applying the graph Laplacian \mathcal{L} to \mathbf{f} , we can write:

$$\begin{aligned}
-\mathcal{L}f_i &= -\sum_{j \in \mathcal{N}_i} w_{ij}[f_i - f_j] = \\
&= [f(x_i + 1, y_j) - f(x_i, y_j)] - [f(x_i, y_j) - f(x_i - 1, y_j)] + \\
&+ [f(x_i, y_j + 1) - f(x_i, y_j)] - [f(x_i, y_j) - f(x_i, y_j - 1)] = \\
&= [f(x_i + 1, y_j) - 2f(x_i, y_j) + f(x_i - 1, y_j)] + \\
&+ [f(x_i, y_j + 1) - 2f(x_i, y_j) + f(x_i, y_j - 1)] = \\
&= f(x_i + 1, y_j) + f(x_i - 1, y_j) + f(x_i, y_j + 1) + f(x_i, y_j - 1) - 4f(x_i, y_j) \approx \\
&\approx \frac{\partial^2 f}{\partial x^2}(x_i, y_j) + \frac{\partial^2 f}{\partial y^2}(x_i, y_j) = \nabla^2 f(x_i, y_j) = \nabla^2 f_i
\end{aligned}$$

with clear resemblance to the 2D discrete Laplace operator $\mathbf{L} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$ [97]. This resemblance is derived directly from the continuous Laplace-Beltrami differential operator. Sandryhaila and Moura write “[...]if the dataset is large and samples uniformly randomly a low-dimensional manifold then the (empirical) graph Laplacian acting on a smooth function on this manifold is a good discrete approximation that converges point-wise and uniformly to the elliptic Laplace-Beltrami operator applied to this function as the number of points goes to infinity.” [145]. In other words, if the graph is the result of the sampling of a continuous surface then the graph Laplacian provides a good approximation of what the behavior of the continuous Laplacian would be on the continuous surface.

In this interpretation, the quadratic form of the Laplacian operator (5.1) can also be considered as a measure of the smoothness of the surface represented by the graph: resulting in a small value if either \mathbf{f} has limited variability or the weights of the edges connecting regions,

or vertices, where the \mathbf{f} varies substantially, are small [92]:

$$\mathbf{f}^T \mathcal{L} \mathbf{f} = \frac{1}{2} \sum_{i,j=1}^n w_{ij} (f_i - f_j)^2 \quad (5.1)$$

where $\mathbf{f} \in \mathbb{R}^n$.

It is possible to observe the use of *normalized* versions of the Laplacian. There are two commonly used form of normalization: $\mathcal{L}_{\text{sym}} = \mathbf{D}^{-1/2} \mathcal{L} \mathbf{D}^{-1/2} = \mathbb{I} - \mathbf{D}^{-1/2} \mathbf{A} \mathbf{D}^{-1/2}$ and $\mathcal{L}_{\text{rw}} = \mathbf{D}^{-1} \mathcal{L} = \mathbb{I} - \mathbf{D}^{-1} \mathbf{A} = \mathbb{I} - \mathbf{W}$. If we remember the discussion about the diffusion operator \mathbf{W} (see section 3.1.1: *Graph theory primer*) we can understand why the second normalization is referred to as the *random walk Laplacian*. The choice of which version to use is determined by the problem at hand². In our case, since nodes have similar degrees, we opted for the unnormalized Laplacian.

As we mentioned in the introduction, in GSP there is a further "separation" between the actual graph and the signal. The graph is a projection of the samples of the original surface to an underlying hyperplane and the signal corresponds to the distance between the original samples and their projections. In this framework, the Laplacian matrix becomes a representation of the intrinsic characteristics of the graph, based exclusively on the established connectivity (topology). This underlying "fabric" is a quality determined by the graph itself and is decoupled from the values that a signal might assume at its vertices.

The two representations (sampled-surface and the lower dimensional projection) might differ depending on how the weights (or distances) between the vertices are defined and the

²In [92], the author argues that all the Laplacians will provide similar spectral results when the graph are fairly regular and every node has similar degree, whereas in case of high spread in the vertices degree statistic, the normalized versions are more robust in the sense that they guarantee convergence to the continuous case we described before as the number of nodes is increased.

smoothness of the original surface. We can imagine a case where, if the distances are defined from an Euclidian embedding of the surface, in the presence of large gradients, nodes that are far in the sampled surface representation might end up very close in the projected graph. This underlines the importance of the weight definition in graph theory.

It can be verified, by substitution, that the complex exponentials $\phi(x, \omega) = e^{j\omega x}$ are solutions of the eigenproblem $\nabla^2 \phi = \lambda \phi$ in \mathbb{R} . In this sense, we can also interpret the inverse Fourier transform of the frequency representation $\hat{g}(\omega)$ of a function $g(x)$ as the superposition of the eigenfunctions of the Laplace operator [148]: $g(x) = 1/2\pi \int \hat{g}(\omega) \phi(x, \omega) d\omega$, where the $\hat{g}(\omega)$ are obtained by the analysis step $\hat{g}(\omega) = \int g(x) \phi^*(x, \omega) dx$. Given the relationship described above between continuous and graph spaces, we can extend this approach to represent graph signals as superposition of the eigenvectors of the graph Laplacian.

For a connected graph with N nodes, we can write the discrete graph eigenproblem as $\mathcal{L}\mathbf{u} = \lambda\mathbf{u}$, where $\mathbf{u}_\ell \in \mathbb{R}^N$ is the eigenvector corresponding to the eigenvalue λ_ℓ .

Because of its definition, the Laplacian matrix is real and symmetric leading to a set of N eigenvalues $\{\lambda_\ell\}_{\ell=0,1,\dots,N-1}$ and the corresponding set of independent eigenvectors $\{\mathbf{u}_\ell\}_{\ell=0,1,\dots,N-1}$ which constitutes an orthogonal³ basis⁴ of \mathbb{R}^N .

The graph Laplacian eigenvalues have several interesting properties that provide insight about the graph connectivity. For example, in the case of connected graphs, the lowest eigenvalue (after sorting) is always zero ($\lambda_0 = 0$) and the corresponding eigenvector is constant. This can be verified by looking at each row of the Laplacian matrix which contains the degree of a vertex and the negative weights of its connections to the other vertices. Each row

³Normalization can be achieved by dividing each eigenvector by its norm.

⁴Some authors use the generalized eigenvectors derived from the Jordan decomposition of the adjacency matrix as basis [145] due to the wider range of type of graphs that can be analyzed with this approach. Since we will be using simple undirected graphs, we decided to follow the more common approach of the Laplacian.

will then sum to zero. Since this is true for every row, we see how the vector $\mathbf{1}^N$ becomes eigenvector with eigenvalue zero. Another interesting property of λ_0 is that its algebraic multiplicity is equal to the number of connected components of G . We use this as a check step in our approach, after determining the connectivity of the graph, to verify that we have a connected graph with a single component.

With an orthonormal basis at our disposal, we can now write the analysis and synthesis equations for the GFT [144]:

$$f[n] = \sum_{\ell=0}^{N-1} \hat{f}[\ell] u_{\ell}[n] \quad (5.2)$$

$$\hat{f}[\ell] = \langle \mathbf{u}_{\ell}, \mathbf{f} \rangle = \sum_{n=1}^N u_{\ell}[n] f[n] \quad (5.3)$$

where the convention is that the inner product⁵ is conjugate-linear in the first term, although in our case, all elements are real.

If we identify with \mathbf{U} the array composed of the eigenvectors as columns, we can rewrite the previous as $\hat{\mathbf{f}} = \mathbf{U}^T \mathbf{f}$ and $\mathbf{f} = \mathbf{U} \hat{\mathbf{f}}$. The reconstruction in this case is guaranteed by the fact that \mathbf{U} is an *orthogonal matrix* in which case $\mathbf{U}^T = \mathbf{U}^{-1}$ and $\mathbf{U} \mathbf{U}^T = \mathbb{I}$. Using this property, we can also prove that the Plancherel theorem applies to the GFT: $\langle \hat{\mathbf{f}}, \hat{\mathbf{g}} \rangle = \langle \mathbf{U}^T \mathbf{f}, \mathbf{U}^T \mathbf{g} \rangle = (\mathbf{U}^T \mathbf{f})^T \mathbf{U}^T \mathbf{g} = \mathbf{f}^T \mathbf{U} \mathbf{U}^T \mathbf{g} = \mathbf{f}^T \mathbf{g} = \langle \mathbf{f}, \mathbf{g} \rangle$.

If we consider the discrete Fourier transform (DFT), we can look at the analysis process as the projection of a signal in \mathbb{R}^N , originally defined on the discrete-unit-interval basis $\{\mathbf{e}_i = e_i[n] = \delta[i - n]\}_{i=1, \dots, N}$, onto the elements of the complete orthonormal DFT basis $\{\mathbf{w}_m = w_m[n] = \frac{1}{\sqrt{N}} e^{j2\pi mn/N}\}_{m=0, \dots, N-1}$. In the case of the GFT, we replace the DFT basis

⁵We use the natural definition of inner product for vector in \mathbb{R}^N : $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y}$.

with the one identified by the Laplacian eigenvectors.

While the concept of frequency (as number of oscillation within signal sample length) can be naturally defined within the DFT realm, given the underlying ordered connectivity of \mathbb{R}^N , it might not be immediately obvious that a similar property holds also for graphs.

In the case of graphs, the concept of frequency can be generalized to a relationship between neighboring nodes: low frequency signals show smooth changes between adjacent nodes and present few edges incident on vertices of opposite signs, whereas high frequency signals are characterized by a high number of adjacent node showing opposite signs. A numeric estimate of the frequency of a signal, can be obtained by counting the zero crossings within each of the eigenvectors as the eigenvalue is increased.

A more rigorous approach is based on the definition of the *oriented incidence matrix* (\mathbf{E}) where each row represents a vertex and each column an edge. If the edge e_{ij} exists then the entries corresponding to vertices i and j will have values -1 and +1 respectively⁶. For weighted graph, the ± 1 values are replaced by $\pm\sqrt{w_{ij}}$ (under the usual assumption that edge weights are symmetric).

⁶While the assignment of a sign to nodes occurs naturally in directed graph where the positive sign is associated to the arriving and the minus to the departing nodes, in the case of undirected graph, the orientation is arbitrary and sometimes both entries will be positive.

For the toy example in Fig. 3.1, the incidence matrix is given by:

$$\mathbf{E} = \begin{array}{c} \text{nodes} \backslash \text{edges} \\ \begin{array}{c} A \\ B \\ C \\ D \\ E \end{array} \end{array} \begin{bmatrix} e_{AB} & e_{BC} & e_{BD} & e_{CD} & e_{DE} \\ -\sqrt{0.5} & 0 & 0 & 0 & 0 \\ \sqrt{0.5} & -\sqrt{1.0} & -\sqrt{0.75} & 0 & 0 \\ 0 & \sqrt{1.0} & 0 & -\sqrt{1.0} & 0 \\ 0 & 0 & \sqrt{0.75} & \sqrt{1.0} & -\sqrt{0.5} \\ 0 & 0 & 0 & 0 & \sqrt{0.5} \end{bmatrix}$$

If we consider how \mathbf{E} behaves when applied to a signal, we see that it performs similarly to a derivative $(\mathbf{E}^T \mathbf{f})_{ij} = \sqrt{w_{ij}}(f_j - f_i)$ (this would definitely be the case if the weights were defined as $w_{ij} = 1/d_{ij}^2$ with d_{ij} representing a physical distance between the nodes). Following [150], the relative variation of a signal, given a particular graph G , can be evaluated as $\Delta_G^2 \mathbf{f} = \|\mathbf{E}^T \mathbf{f}\|^2 / \|\mathbf{f}\|^2$. If we observe that, independently from the chosen orientation of the edges, $\mathbf{E}\mathbf{E}^T = \mathcal{L}$, we can write $\Delta_G^2 \mathbf{f} = \mathbf{f}^T \mathcal{L} \mathbf{f} / \|\mathbf{f}\|^2$. Remembering the discussion on how the quadratic form of the Laplacian (5.1) is a measure of the smoothness of a signal, we can see how $\Delta_G^2 \mathbf{f}$ will be smaller for smoother signals and larger for high-varying ones, providing a measure of the variability (frequency) of the signal \mathbf{f} .

Fig. 5.1 illustrates this frequency behavior for some selected eigenvectors of two graphs: cycle and periodic lattice, where the similarity with the DFT is stronger due to the natural embedding of these regular graphs into the Euclidian space.

In our application graph geometries are more complicated. The graph for the Wink Sinks dataset (see section B.2: *Wink Sinks*), shown as the base graph for Fig. 5.2, was obtained by

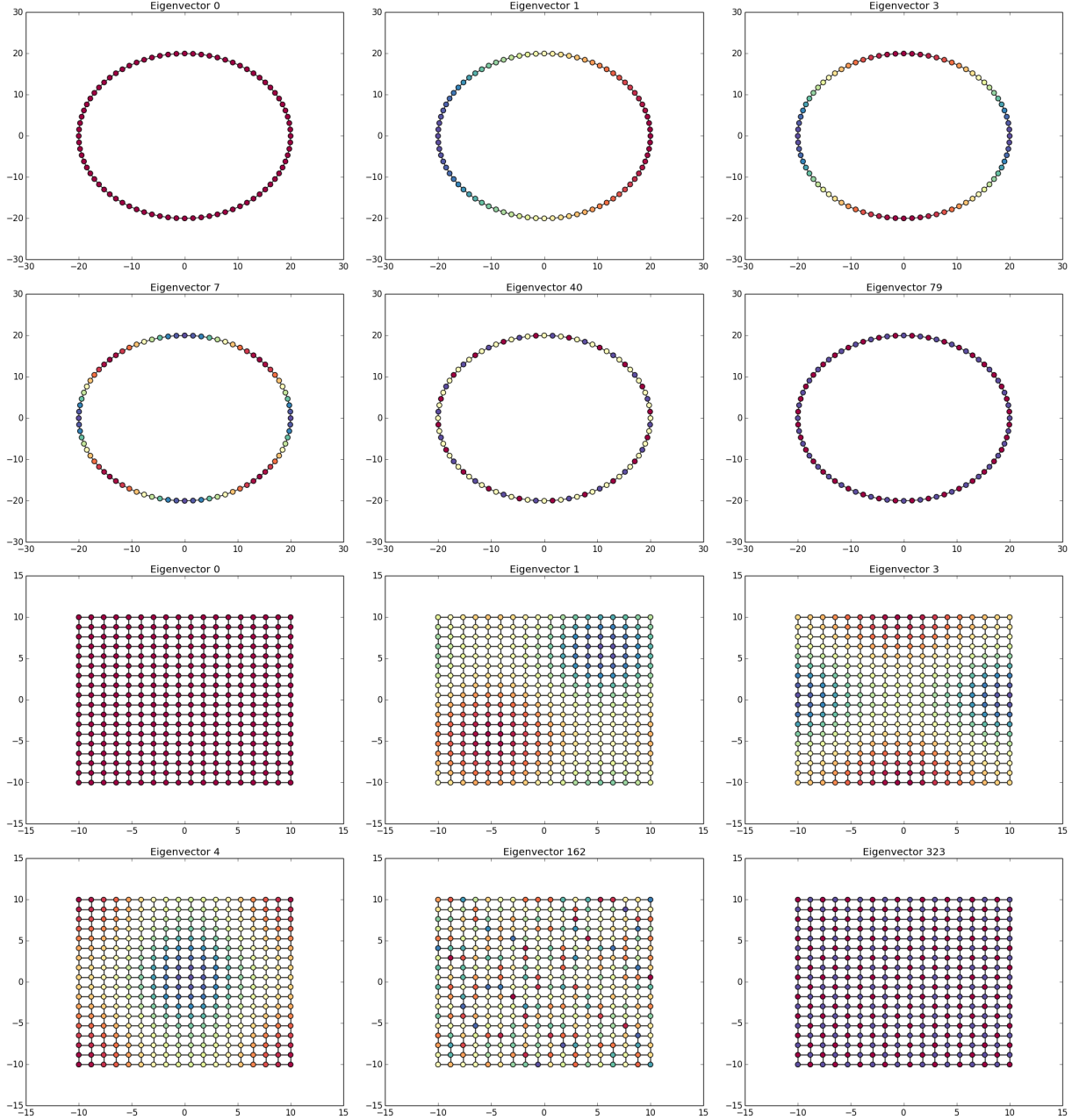


Figure 5.1: Example of eigenvectors of the Laplacian for regular graphs. The top six plots represent eigenvectors for eigenvalues 0, 1, 3, 7, 40, and 79 for a cycle graph. For this graph, the eigenvector corresponding to the $u_{2k+1} = \cos(2\pi kn/N)$ and $u_{2k+2} = \sin(2\pi kn/N)$ for $k = 0, 1, \dots, N/2 - 2$ whereas $u_0[n] = 1$ and $u_{N-1}[n] = \cos(\pi n)$ (where n is the node "number"). The bottom six show eigenvectors 0, 1, 3, 4, 162, and 323 for a periodic lattice graph. The periodicity is obtained by folding the graph as a T^2 and connecting the folded ends. In both cases it is possible to see how a high frequency of change of sign between adjacent nodes corresponds to a higher eigenvalue.

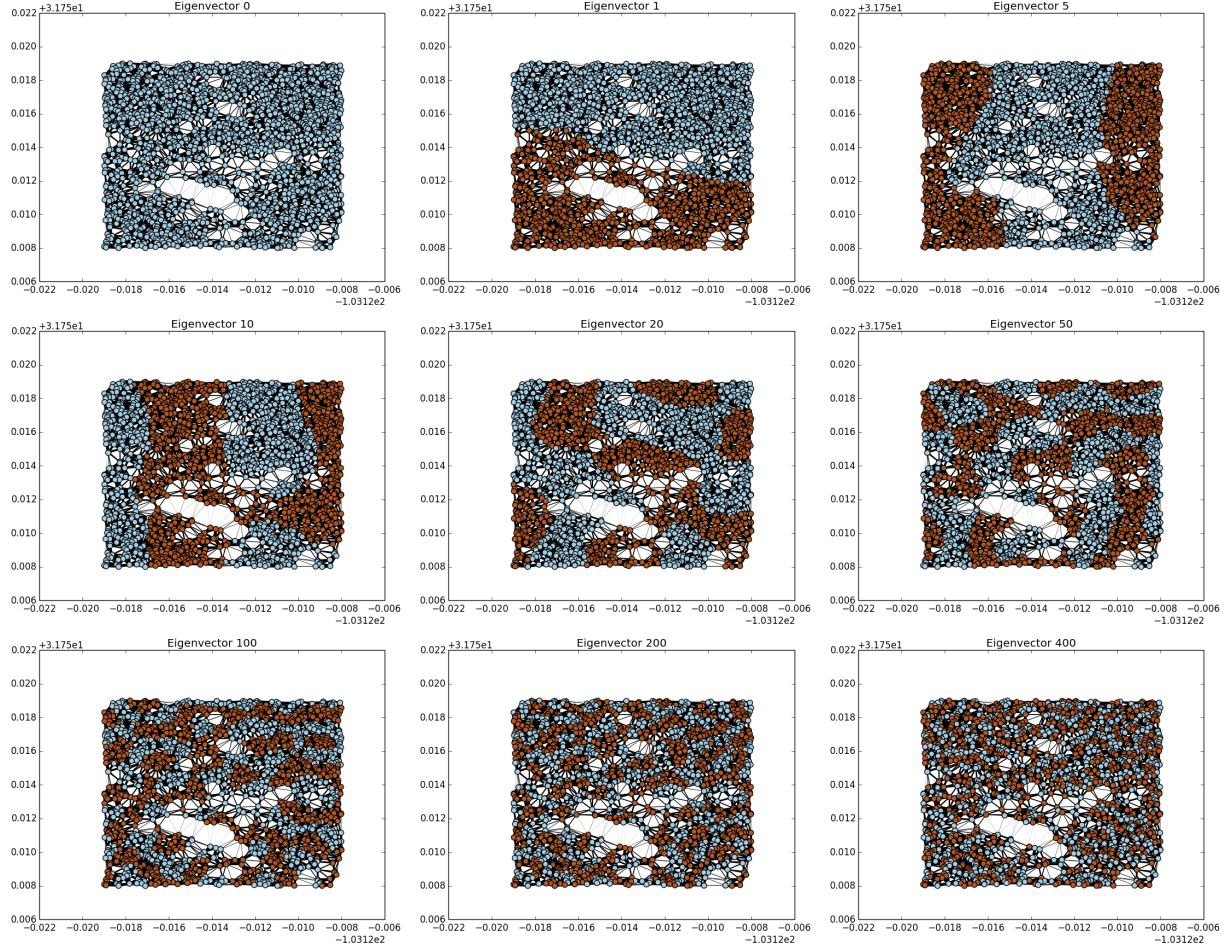


Figure 5.2: From left to right, top to bottom, eigenvectors corresponding to eigenvalues 0, 1, 5, 10, 20, 50, 100, 200, and 400 for a graph obtained from the Wink 3 dataset. The vertices are represented based on their actual coordinates. The connectivity is evaluated using the Delaunay triangulation and the weights assigned based on the actual physical distance between the points (Thicker edges correspond to higher weights). To improve readability of the plot, we simply show the sign domains (positive and negative) instead of the actual values.

associating a node to each scatterer, embedding the nodes into the Euclidian space defined by the coordinates of the corresponding scatters, connecting the nodes using the Delaunay triangulation (see section 3.2: *Graph cut approach*), and weighting the resulting edges exponentially, according to the physical distance between represented scatterers. Nevertheless, the eigenvectors of the resulting Laplacian continue to display the frequency behavior we illustrate above (Fig. 5.2).

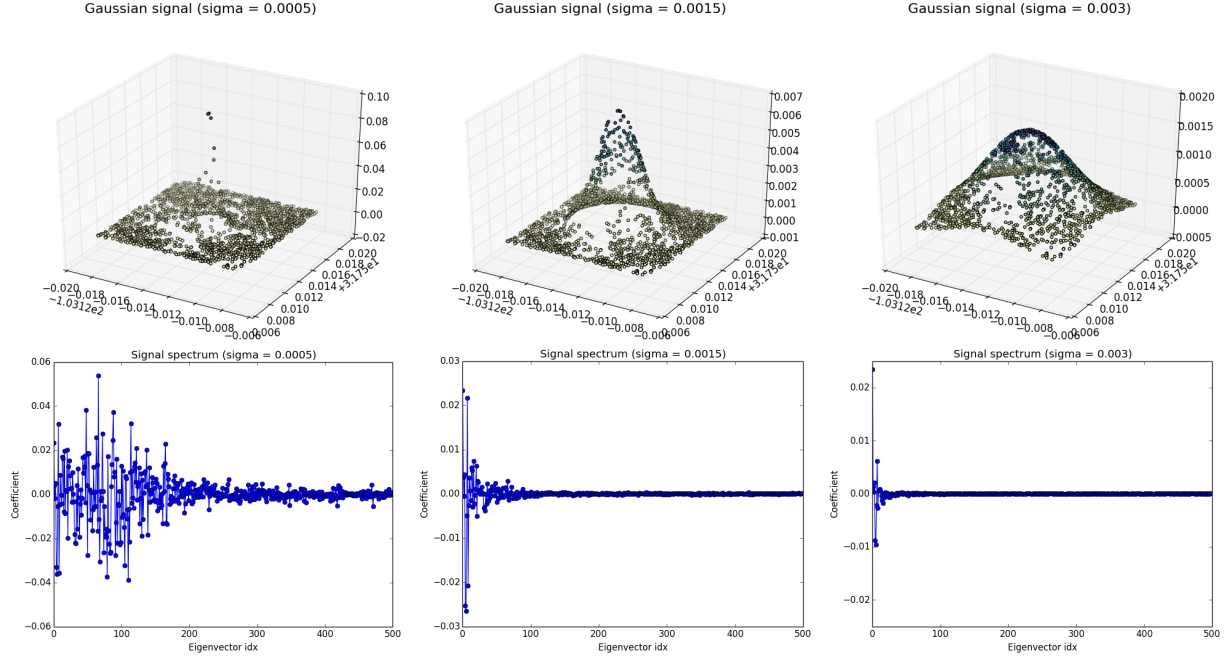


Figure 5.3: Top: from left to right, three examples of artificial normalized Gaussian signals generated on the Wink 3 dataset with increasing value of sigma: 0.0005° , 0.0015° , 0.003° . Bottom: the corresponding spectra (only the first 500 eigenvalues) evaluated using (5.3).

Because of this intrinsic frequency behavior associated with the eigenvector basis, it is natural to define the concept of *spectrum* of a graph signal as the representation of the GFT coefficient, obtained from the analysis step (5.3), as function of the sorted eigenvalues.

This representation follows the behavior we have learned to expect from the DFT: the larger the feature in the graph-signal space, the narrower the associated spectrum. This is illustrated in Fig. 5.3 where we show the GFT of Gaussian signals, of increasing scale, defined over the graph obtained from Wink Sinks dataset restricted to the region of Wink 3.

As equations (5.2) and (5.3) provide a convenient way of moving across representations, we can now use them to verify how far we can push the similarities between GFT and DFT.

5.2 Detection and scale analysis

To achieve our goal of detecting a feature in a graph signal, we would like to use an approach similar to the one used in matched filters: evaluate the cross-correlation between a signal (\mathbf{f}) and a template ($\boldsymbol{\tau}_\sigma$). The cross-correlation would then be repeated for templates of different scales (σ) and the optimal result could be obtained by employing the scale space analysis method outlined in the previous chapter, evaluating the maximum value across the scale axis.

In "regular" spaces, the cross-correlation is typically evaluated by the convolution between the signal and the time- or space-inverse of a specific template [26]. In turn, for computational efficiency, the convolution is evaluated as the element-wise multiplication of the Fourier transforms of the two signals after the appropriate zero-padding. Leveraging again the similarity between Fourier and graph transform, we can define the *graph convolution* between two graph signals $f[n]$ and $\tau_\sigma[n]$ as the inverse GFT of the product of their GFT representations [146]:

$$\xi_\sigma[n] = (\mathbf{f} * \boldsymbol{\tau}_\sigma)[n] = \sum_{\ell=0}^{N-1} \hat{f}[\ell] \hat{\tau}_\sigma[\ell] u_\ell[n] \quad (5.4)$$

or $\boldsymbol{\xi}_\sigma = \mathbf{U}[\text{diag}(\mathbf{U}^T \mathbf{f}) \mathbf{U}^T \boldsymbol{\tau}_\sigma]$, where $\text{diag}(\mathbf{x}) = \mathbf{I}\mathbf{x}$.

The result of this operation follows the parallelism with the DFT only on regular graphs such as the path, cyclic, and lattice. The main reason for the discrepancy lies on the fact that, although each eigenvector displays a frequency-like behavior, when compared with each other, this behavior lacks the regularity provided by the Fourier basis, except in the case of regular graphs. Nevertheless it is still possible to approach the problem as direct cross-correlation of

graph signals:

$$\xi_\sigma[n] = (\mathbf{f} \star \boldsymbol{\tau}_\sigma)[n] = \sum_{i=1}^N f^*[i] \tau_\sigma[n+i] \quad (5.5)$$

In order to perform this operation, we need to provide a meaningful definition of what *shifting* on a graph might mean.

We have previously discussed how the $\mathbf{W} = \mathbf{D}^{-1}\mathbf{A}$ matrix represent one of the natural processes by which "information" might travel from one node to the connected ones. The application of this operator, though, is closer to a diffusion and the repeated application of \mathbf{W} smooths the signal rather than shifting it. Following, once again, the similarities between Fourier and graph transforms, we can define a shift in the graph domain as the inverse transform of a "modulation" in the Laplacian eigenvector domain [146]. If we apply this operation to our template, we can write:

$$(S_{[i]}\tau_\sigma)[n] = \sum_{\ell=0}^{N-1} \hat{\tau}_\sigma[\ell] u_\ell[i] u_\ell[n] \quad (5.6)$$

Alternatively, we can define this *shift operator* as $\mathbf{S}_{[i]} = \mathbf{U}[\text{diag}(\mathbf{U}^T \boldsymbol{\delta}_i)] \mathbf{U}^T$, where $\boldsymbol{\delta}_i$ is a signal that is zero on every vertex except the i -th where it assumes the value of one, and, with the notation $\mathbf{S}_{[i]}$, we intend to underline the fact that this operation has to be understood as a *shift onto* a specific vertex i rather than a *shift of* i vertices⁷. Shifting to a node is effectively achieved by multiplying the spectrum of the signal to be shifted by the row of \mathbf{U} corresponding to the desired target node, followed by the inverse GFT.

⁷A factor of \sqrt{N} can be added to guarantee that the sum of the values is preserved by the operator. In our case, we will re-normalize the data after each shift.

Similar to what is expected for the DFT, if we consider a flat spectrum signal ($\hat{\phi} = \mathbb{1}^N$), the shift operator will reconstruct a delta signal at the desired node:

$$(S_{[i]}\phi)[n] = \sum_{\ell=0}^{N-1} \hat{\phi}[\ell] u_{\ell}[i] u_{\ell}[n] = \sum_{\ell=0}^{N-1} u_{\ell}[i] u_{\ell}[n] = \delta[n - i] = \delta_i[n] \quad (5.7)$$

This can be understood by noting that $u_{\ell}[i]$ and $u_{\ell}[n]$ are two rows of the orthogonal eigenvector matrix \mathbf{U} for which $\mathbf{U}\mathbf{U}^T = \mathbb{I}$.

Now that we have defined a graph shift operator, the most natural approach to implement the cross-correlation (5.5) would be to define a specific graph-domain template (τ_{σ}) that we are interested in detecting within the signal \mathbf{f} , and then evaluate the cross-correlations:

$$\xi_{\sigma}[n] = \langle \mathbf{f}, \mathbf{S}_{[n]}\tau \rangle = \sum_{i=1}^N f^*[i] (S_{[n]}\tau_{\sigma})[i] \quad (5.8)$$

for each vertex n in the graph, or $\xi_{\sigma} = \mathbf{f}^T \mathbf{U} [\text{diag}(\mathbf{U}^T \delta_n)] \mathbf{U}^T \tau_{\sigma}$. Once again, this direct approach does not work for a generic signal τ_{σ} on a generic graph.

This is, once more, a consequence of the representation of the graph signal in the Laplacian eigenspace, and is one of the major differences between the DFT and the GFT. In the DFT, the amplitude of the spectrum of a signal is invariant to translation and the phase is responsible for its reconstruction at the appropriate location. With the GFT, there is no phase. If we consider, for example, the eigenvectors illustrated in Fig. 5.2, we see that, if as signal is shifted on the graph before analysis, the two resulting spectra will be quite different while still maintaining the general width associated with the overall size of the signal (Fig. 5.3). When we apply the shift operator (5.6) to a signal, since there is no phase information, we

are effectively changing the amplitude of the coefficients required for the reconstruction of the signal. As a result, direct application of (5.6) to the spectrum of the desired template would cause, not only a shift, but also a distortion of its reconstructed signal. However, it has been shown that, if we were to start with a smooth spectral representation for $\hat{\tau}_\sigma[\ell]$, then the application of the shift operator (5.6) would result in a localized signal around the selected node [148].

Furthermore, it is possible to define a version of the uncertainty principle for graph signals and their spectra that provide an inverse proportionality link between the spread of the spectrum and the spread of the corresponding signal [150] [151]. This result, in conjunction with (5.3) defines the *windowed graph transform* that is used to analyze local graph "frequency" behavior. The approach consists in defining a localized spectrum, shift the resulting localized window on a specific vertex, point-wise multiply with the graph signal, apply the GFT to the resulting signal, and repeat the process for each node generating a plot similar to a spectrogram⁸.

We are going to use these two properties to define a smooth template of a given scale that we can use to evaluate (5.8) on our test data. One issue to be considered is that, by defining a graph signal starting from its spectrum and enforcing smoothness, we lose the possibility of controlling the specific shape because of the reasons listed above. Nevertheless, when starting from a Gaussian spectrum, the reconstructed signal resembles a smooth Gaussian centered on

⁸It could be argued that a direct definition of a window on a graph at each step without implementing any of the GFT could achieve the desired result without the use of this complicated mechanism. Although in principle this is true, there are several issues. First, defining a specific waveform and then using the location of the nodes as sampling points assumes that the graph is embedded in a space where it is possible to define such templates. The second problem is presented by the potential sparsity of the data. The graph-based approach automatically preserves the average value of the signal while shifting it over a specific vertex by implicitly considering density of nodes and weights of edges.

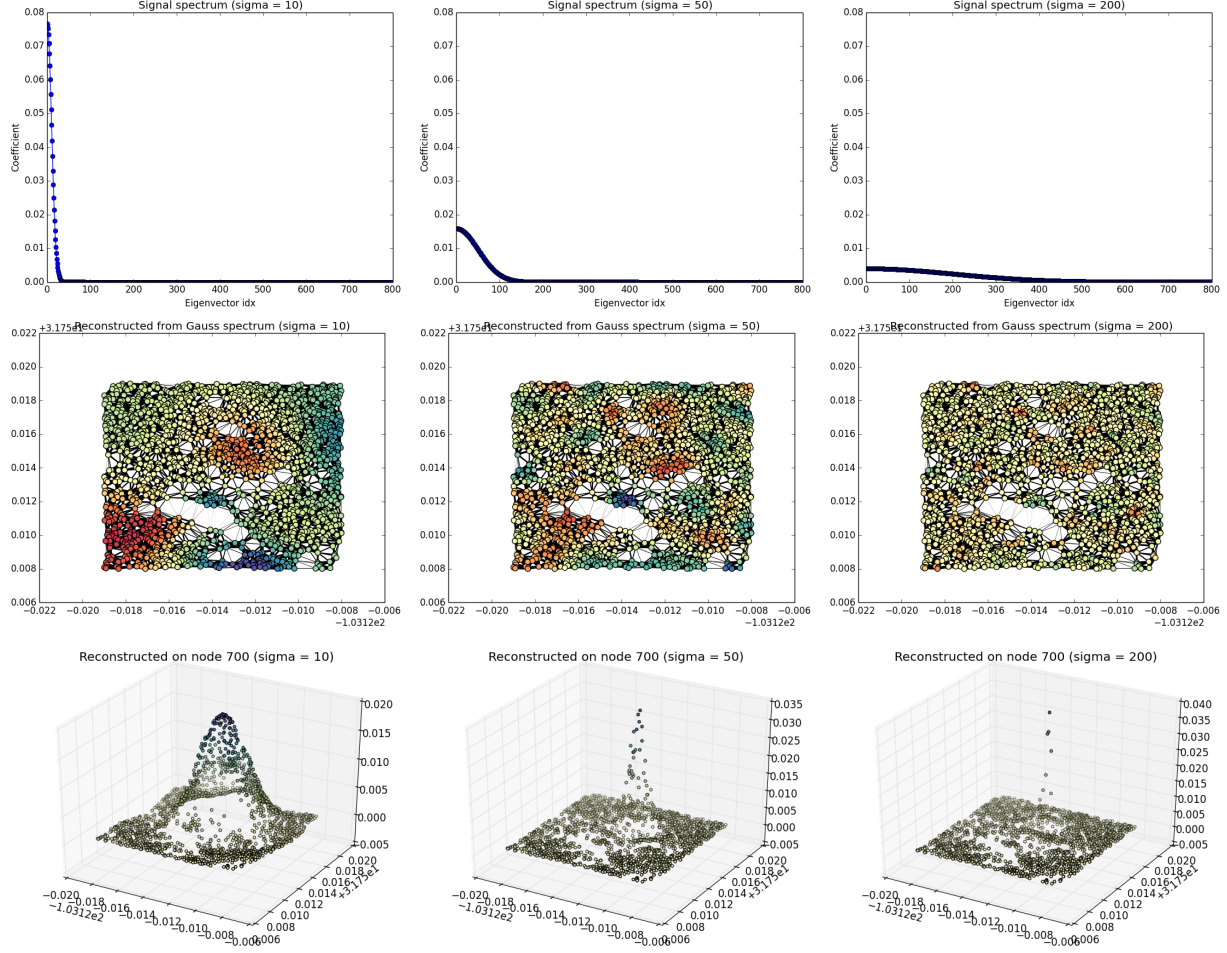


Figure 5.4: Top: from left to right, three examples of artificial normalized Gaussian spectra (only the first 500 eigenvalues). Middle: the direct reconstruction of the spectra using (5.2). It is noticeable how a narrower spectrum will correspond to a smoother signal. Bottom: the signal obtained from reconstruction of the spectra using (5.6) (centered at node 700).

the desired node. Gaussian-based objects are some of the most commonly used templates at the basis of several image processing approaches such as spot detection [152], edge detection [153], and scale-space analysis [134].

In (Fig. 5.4) we show an example of the template construction procedure for the graph obtained from the Wink 3 dataset. First we define a Gaussian spectrum ($\hat{\tau}_\sigma$) of a given scale σ (top). The middle line shows the synthesis (5.2) of a signal starting from the given spectrum. This signal is only presented to illustrate the results of the reconstruction since,

the shifting occurs in the eigenvalue domain. It is possible to notice that the narrower the spectrum, the smoother the reconstructed signal. The bottom row show the reconstruction of the template centered at node 700 ($\mathbf{S}_{[700]}\boldsymbol{\tau}_\sigma$).

We now have all the ingredients necessary to evaluate the normalized cross-correlation of the graph signals by applying a normalization factor to (5.8):

$$\bar{\xi}_\sigma[n] = \frac{\langle \mathbf{f}, \mathbf{S}_{[n]}\boldsymbol{\tau}_\sigma \rangle}{\|\mathbf{f}\| \|\mathbf{S}_{[n]}\boldsymbol{\tau}_\sigma\|} = \frac{\mathbf{f}^T \mathbf{U} [\text{diag}(\mathbf{U}^T \boldsymbol{\delta}_n)] \mathbf{U}^T \boldsymbol{\tau}_\sigma}{\|\mathbf{f}\| \|\mathbf{S}_{[n]}\boldsymbol{\tau}_\sigma\|} \quad (5.9)$$

Since the signals span the entire graph, the normalization of \mathbf{f} is intended to include only the region where the template is non-zero. This guarantees the localization of $\bar{\xi}_\sigma[n]$ around n .

Given a value for σ , $\bar{\xi}_\sigma[n]$ provides a measure of the match between the graph signal \mathbf{f} and the template $\boldsymbol{\tau}_\sigma$ shifted on node n . The maximum of this function will identify the location of the feature within the graph. To also extract information about the scale of the detected feature, we need to evaluate (5.9) for several values of σ . In this case, $\bar{\xi}_\sigma[n]$ is no longer a function of just n but a surface determined by both σ and n . The maximum of this surface provides both localization and size information for the best matching feature present in the dataset:

$$(n, \sigma)_{best} = \arg \max_{n, \sigma} \frac{\langle \mathbf{f}, \mathbf{S}_{[n]}\boldsymbol{\tau}_\sigma \rangle}{\|\mathbf{f}\| \|\mathbf{S}_{[n]}\boldsymbol{\tau}_\sigma\|} = \arg \max_{n, \sigma} \frac{\mathbf{f}^T \mathbf{U} [\text{diag}(\mathbf{U}^T \boldsymbol{\delta}_n)] \mathbf{U}^T \boldsymbol{\tau}_\sigma}{\|\mathbf{f}\| \|\mathbf{S}_{[n]}\boldsymbol{\tau}_\sigma\|} \quad (5.10)$$

Alternatively, if we expect the presence of multiple features, it might be more relevant to evaluate the maximum of $\bar{\xi}_\sigma[n]$ along the scale axis ($\bar{\xi}_{max}[n] = \max_\sigma(\bar{\xi}_\sigma[n])$), with an approach similar to the one described in (3.14). This will result in a graph signal indicating, for each

node, the best match obtained. This signal can then be thresholded at a predetermined significance level to detect the location and size of all the features present within the dataset.

5.3 Results

We tested the approach outlined in the previous section on three datasets.

Dataset I This dataset was generated artificially starting from a lattice graph of 1089 nodes (33×33) embedded in the interval $[-10, 10] \times [-10, 10] \in \mathbb{R}^2$. The nodes were then connected with a 4-connectivity (nearest neighbors - NSEW) with edges of weight $w_{ij} = 1$. The signal, a normalized Gaussian of $\sigma_I = 2.0$, was placed in the top right quadrant, centered on $(x_0, y_0) = (4, 4)$.

Dataset II The base graph for this dataset was generated using a subset of the Wink Sink point cloud (see section 2.2.2: *The Winkler sinkholes (Wink Sinks)*) restricted around the Wink 3 area (-103.139°E , 31.758°N to -103.128°E , 31.769°N). Each scatterer in the area was associated to a node and its coordinates used as natural embedding in \mathbb{R}^2 . The connectivity was obtained by applying the Delaunay triangulation (see section 3.2: *Graph cut approach*). Each edge was assigned a weight exponentially decreasing with the physical distance between the nodes. The value of σ was set to the average distance between adjacent nodes to insure that the strength of the edge reflected the average neighborhood properties. The resulting graph had 1826 nodes. The signal was obtained from a normalized Gaussian of $\sigma_{II} = 0.0007^\circ$, centered in the middle of the area at $(x_0, y_0) = (-103.134^\circ\text{E}, 31.764^\circ\text{N})$

Dataset III The base graph for this signal was obtained as illustrated for dataset II. The signal consisted of the actual displacement data for the scatterers used as vertices of the graph. Between the available time slices in the Wink Sinks point cloud dataset, we selected $t = 68.68$ months since, as previously discussed, it provides the largest displacement information.

These three datasets are illustrated in the top row of Fig. 5.5 as graph-signal and in the top row of Fig. 5.6 as function of the node index. This second representation is used because it favors the comparison of the results but it is strictly dependent on the node indexing.

To prove that we can effectively detect a signal within the graph, we evaluated the normalized cross-correlation (5.9) between each of the signals and a template defined by a smooth spectrum: a normalized Gaussian kernel of arbitrary size $\sigma = 40$:

$$\hat{\tau}_\sigma[\ell] = \frac{e^{-\ell^2/2\sigma^2}}{\sum_\ell e^{-\ell^2/2\sigma^2}} \quad (5.11)$$

In the second row of Fig. 5.5 and Fig. 5.6 we show $\bar{\xi}_\sigma[n]$ as function of n . For all three datasets, the approach succeeded in detecting the features within the graph since, comparing with the original signal, we can see that we achieved good discrimination between presence and absence of the feature. Nevertheless, since in our application we are also interested in extracting as many information as possible about the detected feature, we see that the evaluation of $\bar{\xi}_\sigma[n]$ only provides reliable localization. This is shown in the third row of Fig. 5.5 and Fig. 5.6 where the template is shifted at the node where the feature was detected and scaled to minimize the least square error with the original signal. The comparison between the scaled template (third row) and the original signal (first row), clearly shows that

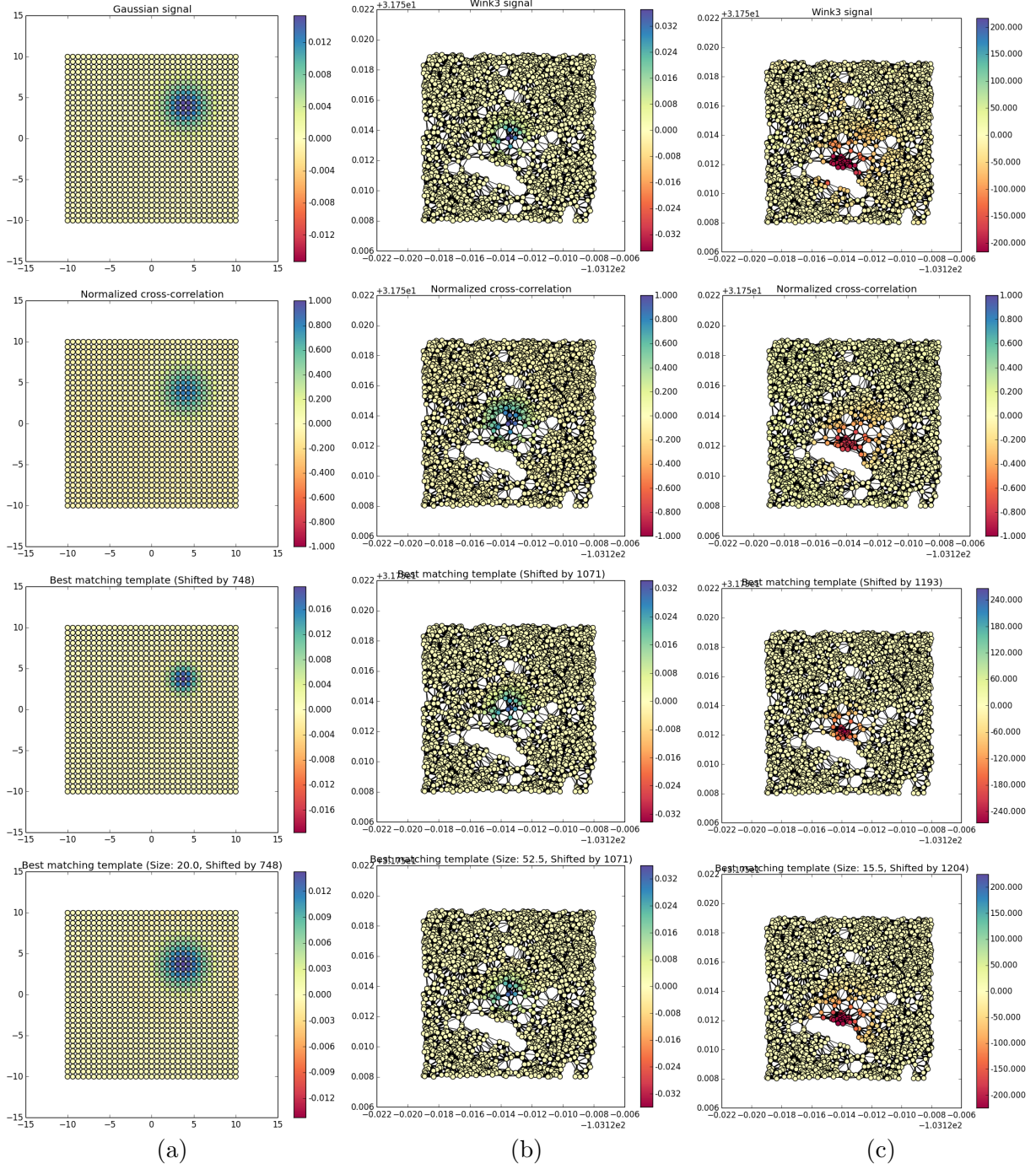


Figure 5.5: Top row: the graph corresponding to the three datasets: (a) dataset I, (b) dataset II, and (c) dataset III. Second: the results of the normalized cross-correlation (5.9) with the template (5.11). Third: the template shifted to the maximum of the cross-correlation and scaled to minimize the least square error (shown) with the signal. Bottom row: same as third row with the difference that scale analysis (5.10) was applied to the template.

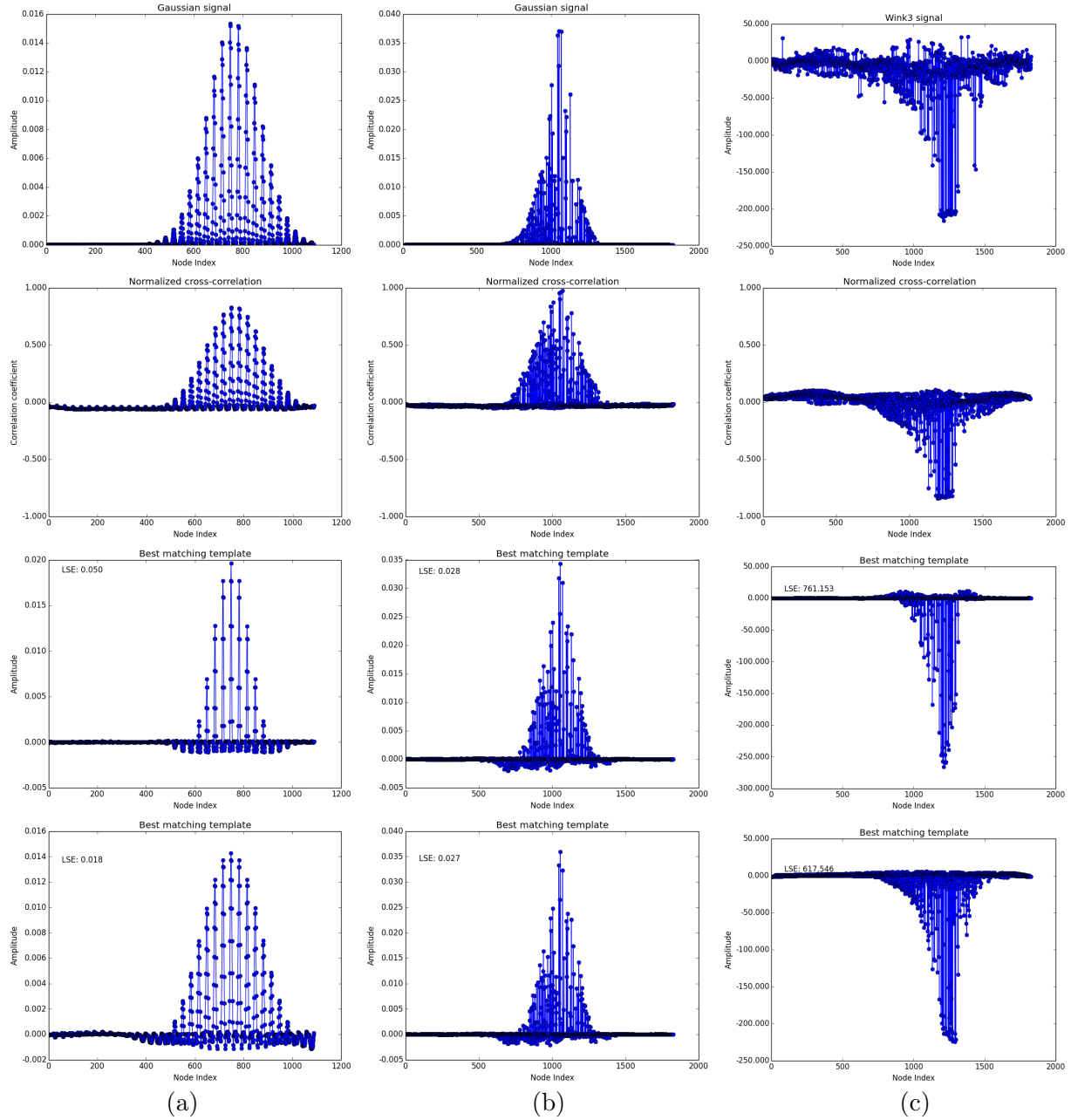


Figure 5.6: The signals represented in this figure correspond to those shown in Fig. 5.5. Here we represent the signals as function of the node index rather than as graph to favor a direct comparison. In the last two rows, where we display the template shifted at the node where the feature was detected and scaled to minimize the least square error with the original signal, the actual value of the LSE is also reported.

we cannot infer the scale of the feature from the template.

This issue is addressed by the scale space analysis approach described by (5.10) and the

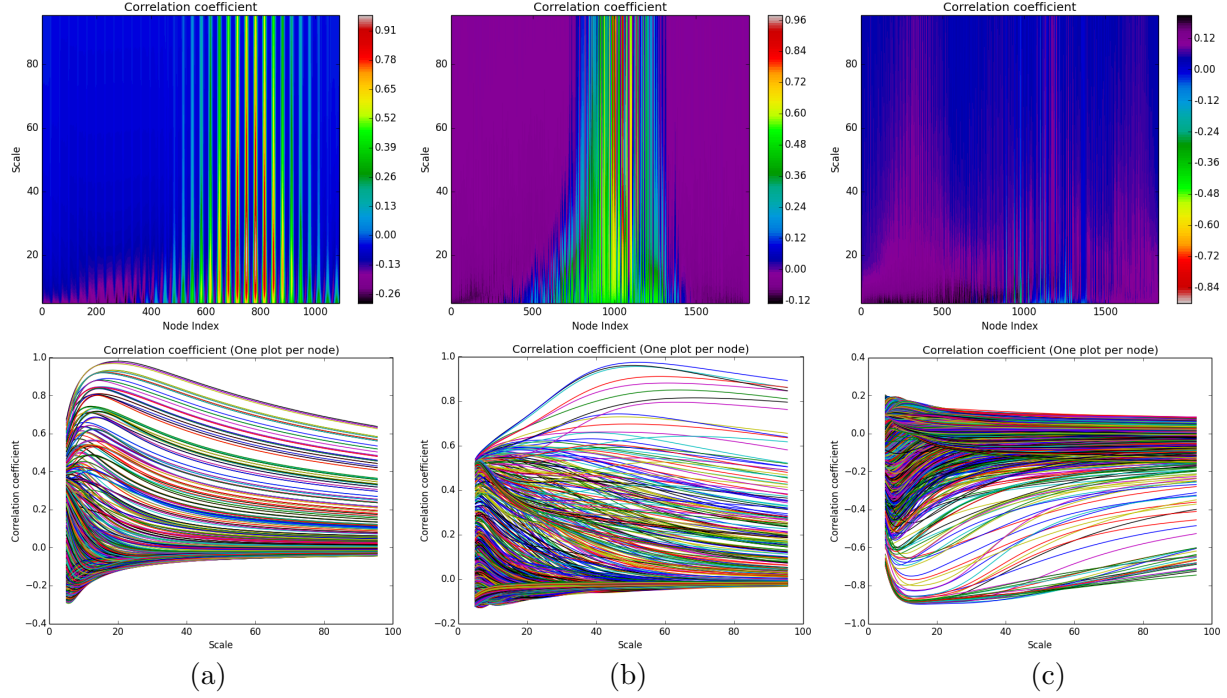


Figure 5.7: Graphical illustration of the implementation of (5.10) for the three dataset: (a) dataset I, (b) dataset II, and (c) dataset III. The top row shows the surface resulting from the analysis. The correlation coefficient is shown as function of node index and scale of the template spectrum. Below each surface, we display the correlation coefficient as function of the scale (one plot for node). Here it is possible to clearly identify the maximum that simultaneously provide identification of location and size of the best fitting template. In the case of Wink 3, since template and signal have opposite signs, we should identify the minimum.

process, for each of the datasets, it outlined in Fig. 5.7. The first row is a representation of $\bar{\xi}_\sigma[n]$ as function of the node index n (horizontal axis) and the scale σ (vertical axis) of the template spectrum $\hat{\tau}_\sigma[\ell]$. The evaluation of (5.10) is equivalent to finding the coordinates for which the surface achieves its maximum. This is better visualized in the bottom row where we plot, for each node, the value of the normalized cross-correlation (5.9) as function of the scale. In this representation the maximum⁹ is clearly identifiable for all three datasets.

Since, typically, there will be more than one feature we might want to detect within a dataset, we also tested the approach on a signal containing two features. The base

⁹In the case of Wink 3, since the template and the signal had opposite signs, the optimal match is located at the minimum.

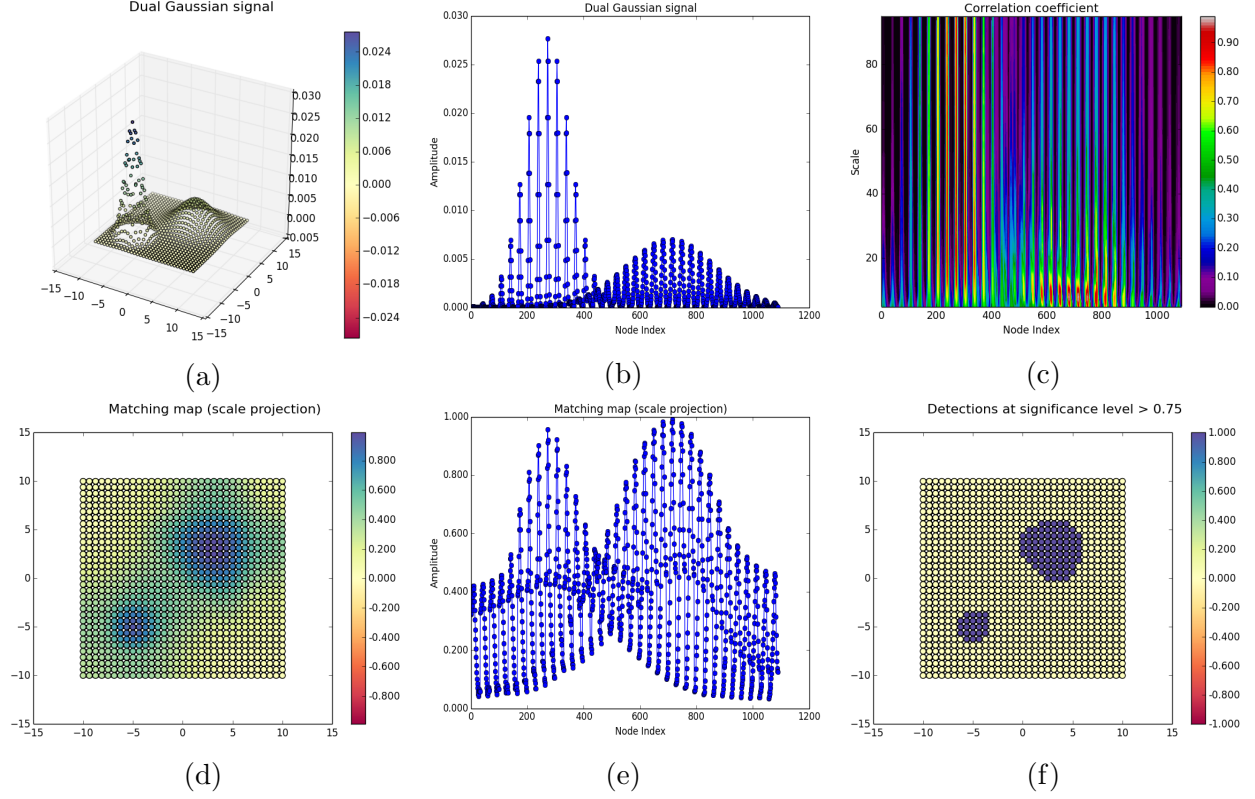


Figure 5.8: Graphical illustration of the implementation of (5.10) for a dataset containing two features. (a) Representation of the graph signal (the underlying graph is not represented). (b) The graph signal illustrated as function of the node index n . (c) $\bar{\xi}_\sigma[n]$ as function of n and σ . (d) Maximum intensity projection of (c) along the scale axis represented on the base graph. (e) Maximum intensity projection of (c) along the scale axis represented as function of the node index to allow for direct comparison with (b). (f) Detections of features in the graph signal, obtained by thresholding (d) at a significance level of 0.75.

graph and connectivity were generated as for dataset I whereas the signal consisted of two normalized Gaussian of $\sigma_a = 1.5$ and $\sigma_b = 3.0$ centered respectively at $(x_0, y_0)_a = (-5, -5)$ and $(x_0, y_0)_b = (3, 3)$.

In Fig. 5.8 we illustrate the results of the detection process. The graph signal, shown as discrete surface in (a) and as function of the node index n in (b), was analyzed using (5.9) for sizes of the template spectrum ranging from $\sigma_{min} = 5$ to $\sigma_{max} = 95$. The resulting $\bar{\xi}_\sigma[n]$ is shown in (c) as function of n and σ . It is possible to clearly identify the two maxima relative to location and scale of the two features. In (d) we show the graph signal corresponding

to the maximum intensity projection of (c) along the scale axis ($\bar{\xi}_{max}[n]$). The same signal is shown in (e) for direct comparison with the original signal (b). It is possible to notice that both location and size of the two features were correctly identified. Finally, in (f), we show a possible representation of the results where areas of interest are selected based on a predetermined significance level. In this particular case, we selected $\bar{\xi}_{max}[n] > 0.75$.

5.3.1 Comparison with previous results

In Fig. 5.9, we show a direct comparison between $\bar{\xi}_{max}[n]$ and the results obtained by the spatiotemporal matching method (see section 3.3: *Spatiotemporal matching*) and the normalized mutual information (NMI) metric pre-screening approach (see section 4.3: *Mutual information metric*), all applied to dataset III (Wink 3)¹⁰.

In the first row of Fig. 5.9 we present the same graph signal ($\bar{\xi}_{max}[n]$) with two different color maps to allow for better visual comparison with the results provided by the other methods. In (c) and (d) we show the results for the spatiotemporal matching algorithm and the NMI metric pre-screening approach respectively¹¹. Finally, in (e) and (f) we display a direct comparison between the results by aligning and overlapping the corresponding images. From this illustration, it is possible to see that all these methods, not only successfully identify the feature within the area, but also provide comparable information on its size.

For completeness, in Fig. 5.10 we show a comparison between the method illustrated in this chapter, the graph cut approach (see section 3.2: *Graph cut approach*), the graph cut

¹⁰We did not include the parameter space reduction technique (see section 4.2: *Parameter space reduction*) in this comparison because, differently from the other approaches, it does not provide information about the characteristics of the features, apart from a generalized localization within a larger area. Nevertheless, as previously described, this approach proved to be quite effective in reducing the size of the areas requiring full analysis.

¹¹See Fig. 4.6 for more details on these images.

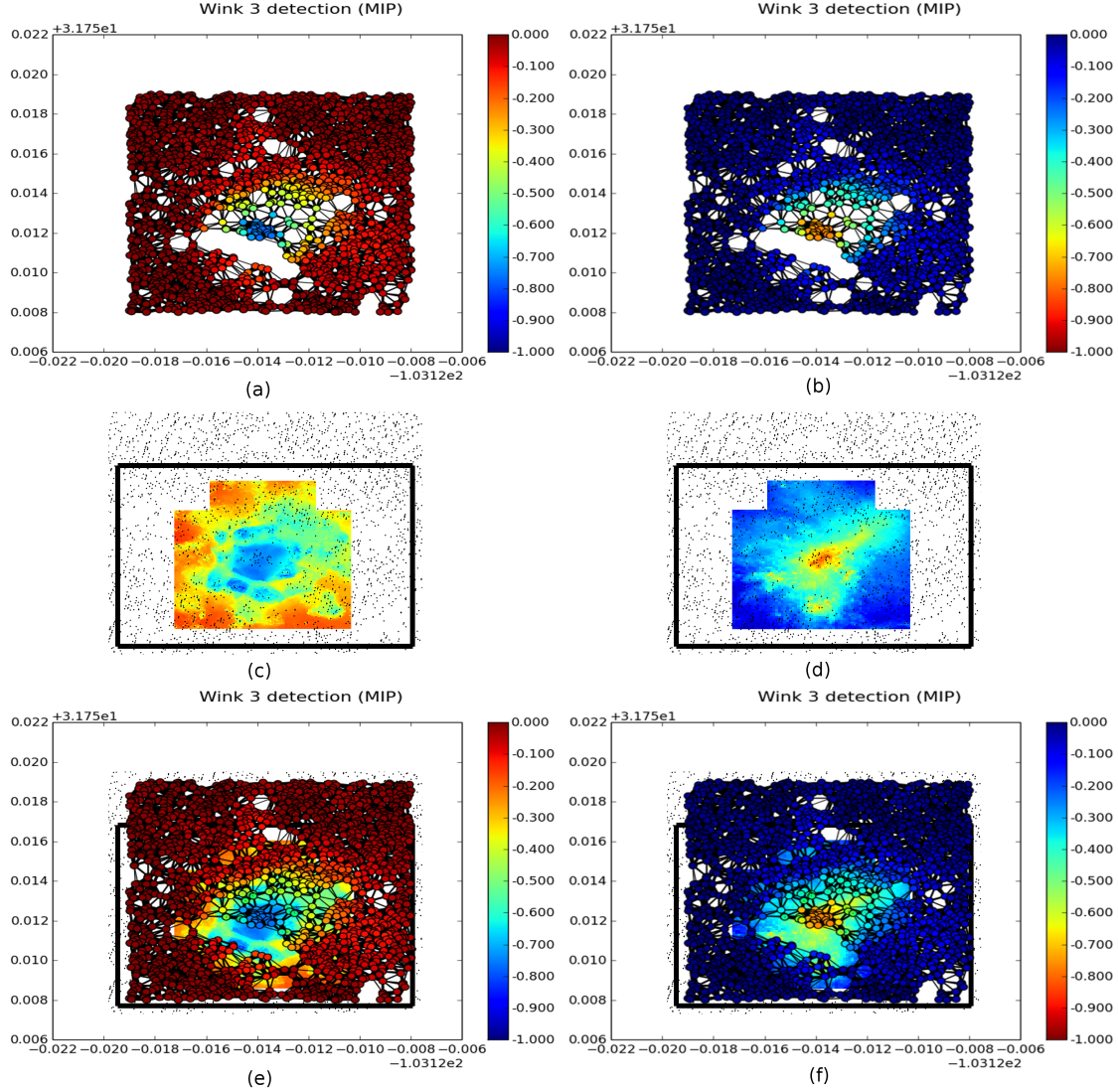


Figure 5.9: Comparison between spatiotemporal matching, NMI metric pre-screening, and graph scale methods over the Wink 3 area (dataset III). In (a) and (b) we show $\bar{\xi}_{max}[n]$ represented in two different color maps to favor direct comparison with the results of the spatiotemporal matching algorithm (c) and the mutual information pre-screening (d) applied to the same region (For more information about (c) and (d) see Fig. 4.6). In the bottom row we overlap the graph scale results to the spatiotemporal matching (e) and mutual information metric (f).

method modified to provide some sensitivity to the temporal behavior (see section 3.3.3:

Comparison with graph cut approach), and the spatiotemporal matching algorithm applied

to dataset III (Wink 3). In the first row we show (a) the result of the graph cut approach,

(b) the risk results for the modified graph cut and (c) the propagated risk as obtained from

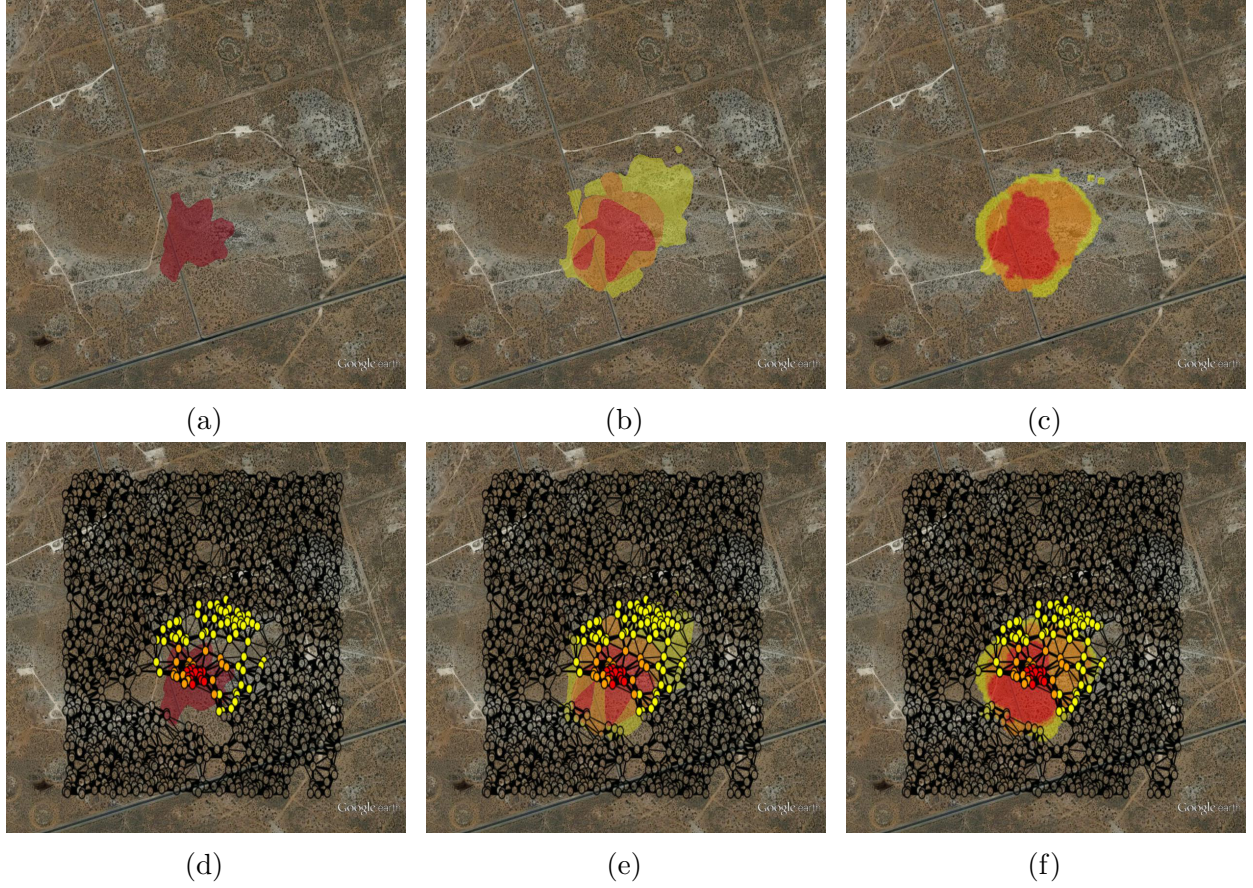


Figure 5.10: Comparison between results of (a) the graph cut approach, (b) the modified graph cut method, (c) the spatiotemporal matching algorithm and the graph scale analysis. The direct comparison is offered respectively in (d), (e), and (f).

(3.16) applied to the Wink Sinks dataset¹². The colors represent the detected level of risk: severe (red), moderate (orange), and slight (yellow). In the case of the graph cut approach (a), the entire detected area is considered high risk, since this approach only provides a logic label as output. In the second row we show $\bar{\xi}_{max}[n]$ overlapped to each of the previous results. We divided the values of $\bar{\xi}_{max}[n]$ in 4 bins to allow for direct comparison. The bins were chosen such that values in the interval $[-1, -0.75)$ corresponded to severe, $[-0.75, -0.5)$ to medium, and $[-0.5, -0.25)$ to slight risk. All the remaining data was considered as no risk.

¹²The comparison between the modified graph cut approach and the spatiotemporal matching method is argued with more detail in (see section 3.3.3: *Comparison with graph cut approach*).

It is important to underline that, although the results illustrated in Fig. 5.9 and Fig. 5.10 might seem to indicate that all approaches can offer satisfactory results in term of localization and scale identification, only the spatiotemporal matching approach provides the capability of discriminate features also based on the temporal behavior. Furthermore, since this method is model independent (see section 3.3.4: *Subsidence model discussion*), the information that can be extracted from the spatiotemporal approach is only limited by the set of parameters used in the model definition. In the case of features tens of meters in diameter and several meters deep, such as the Wink Sinks, and datasets that provide several years of displacement information, allowing for features to become relevant with respect to the surroundings, the added computational cost of providing temporal discrimination might not be required and simpler static model might suffice. If, on the other hand, we consider the fundamental motivation behind our research: the detection of **precursory** motion that might indicate the **early stages** of development of hazardous conditions for the transportation infrastructure, we see how the capability of detecting specific behaviors might become of critical importance in dataset covering limited temporal duration such as the Virginia one (see section B.3: *Virginia*).

5.4 Summary

In this chapter we presented an approach where the detection of features and their scale was conducted entirely within the graph-space. The development of this method was prompted by the inherent sparsity of the SqueeSAR data with respect to the granularity of the parameter space required, by the spatiotemporal matching algorithm, to provide an accurate analysis. In

particular the highest sparsity was caused by the underlying spatial grid required to identify the location of a feature. By providing a detection mechanism acting exclusively on the nodes of the graph, the need for this underlying grid was removed.

Although this approach provides good detection and the capability of identifying the scale of the detected features, it cannot yet provide temporal discrimination. Furthermore, it is based on the evaluation of eigenvalues and eigenvectors of matrices of size $n \times n$ where n is the number of nodes in the dataset. Several powerful computational approaches are available to address eigenproblems [154], nevertheless the memory requirement is most likely to quickly become an issue as the number of considered scatterer is increased. One approach, to address this issue, is to subdivide the dataset in smaller overlapping subsets and analyze them separately. The subdivision could be posed as a minimization problem where the goal is to identify a set of low weight cuts under the constraint that the separated components do not contain more than a given number of points. This is one of the area of research, together with the addition of temporal analysis, that we are planning to pursue in the future.

Chapter 6

Conclusions and future work

With the expected growing availability of SAR data (see section [A.3: *Future of SAR*](#)), we believe that the routine use of remote sensing data for surface displacement monitoring will become a reality. What we presented in this work is our contribution to the development of analysis techniques for the point cloud sparse dataset typically obtained from, but not restricted to, permanent scatterer interferometry.

In particular we presented a generalizable framework that allows the detection of a specific spatiotemporal behavior defined by a user-specified parametric model. The independence of the detection algorithm from the actual model provides this framework with the flexibility required for its potential implementation in the detection of a variety of features in different type of datasets.

By analyzing a ground truth point cloud datasets containing subsiding features, we also developed a model describing the spatiotemporal evolution of sinkholes. This model was then used, in conjunction with our detection framework, to study an area located in western Virginia. The ground validation on a subset of the features identified by our algorithm showed

that 78% of the locations presented strong evidence of subsidence.

Within this specific application, we evaluated the potential computational burden imposed by the direct implementation of the detection framework for complex models over large datasets and proposed two mitigating techniques to reduce the extent of the analysis region.

We then introduced a different approach that addressed the detection process from a graph-theoretical point of view, drawing from the similarities between the Fourier analysis in regularly-sampled spaces and what is becoming known as the graph Fourier transform. Based on this approach, we presented a detection and scale analysis algorithm where location and size of features were detected directly in the graph domain greatly reducing the size of the parameter space to be analyzed.

6.1 Future work

Although the spatiotemporal matching algorithm can be considered a mature approach, we would still like to evaluate possible algorithmic improvements and implementation strategies to speed up the computational time. The original algorithm was entirely developed in MATLAB [155] and, after a brief analysis on a subset of the Virginia dataset (see section B.3: *Virginia*), we evaluated that the projected run time for the whole dataset extent would have been 28 days. We then implemented the core functions in C++ and gained a factor of 10 on the total execution time.

Without diminishing the validity of the proposed parameter space reduction and mutual information analysis methods, we still believe that translating the algorithm to operate on a graphic processing unit would be an interesting research subject also because of the added

challenges due to the size of dataset and parameter space. Furthermore, a considerable increase in the computational speed would allow the use, for example, of complex subsidence models including different sinkhole symmetries.

A subject where we would like to concentrate our future efforts is the proposed graph detection approach. As already mentioned, one of the current main limitation to this approach, is the n^2 increase in the size of the Laplacian with the number n of nodes considered. We would like to address this problem by evaluating the effects that a sectioning of the graph might have on the detection method by answering questions such as: what is the relationship between the eigenvector of the global Laplacian and those of the Laplacians obtained from the smaller subsets? Is it possible to find an optimal cut that minimizes the distortion introduced by the subdivision? Can we stitch together the detection results evaluated on smaller subsets and, if possible, how does the compounded detection compares with the results evaluated directly on the entire graph?

Translating the detection problem to the graph domain has drastically reduced the size of the parameter space by leveraging the sparsity of the dataset. Nevertheless, this approach still requires to cycle through node indexes and scale sizes in order to test the match with all possible templates of interest. An interesting research subject would be to pose the entire detection problem as an optimization problem. In other words:

$$(n, \hat{\boldsymbol{\tau}}) = \arg \max_{n, \{\alpha_i\}} \frac{\langle \mathbf{f}, \mathbf{S}_{[\mathbf{n}]} \hat{\boldsymbol{\tau}} \rangle}{\|\mathbf{f}\| \|\mathbf{S}_{[\mathbf{n}]} \hat{\boldsymbol{\tau}}\|} \quad (6.1)$$

under a smoothness constraint for the spectrum of the template $\hat{\boldsymbol{\tau}}$.

Although the graph-domain approach allows for the detection of scale and location of

features present in the dataset, one of its main shortcomings is its inability to track temporal evolution. We would like research the possibility of extending the current method to include this ability. This would allow for a direct comparison of performances with the spatiotemporal matching algorithm.

Appendix A

Satellites and SAR instrumentation

In this appendix we provide technical information about the satellites that provided the data we used in our research as well as the instrumentation characteristics. We also identify future sources of data and describe new space agencies policies that might provide access to high resolution SAR data for low or no cost.

The two main datasets we used in our research were provided by two missions: the European Space Agency (ESA) European remote sensing satellite (ERS), and the Agenzia Spaziale Italiana (ASI) constellation of small satellites for mediterranean basin observation (COSMO-SkyMed).

A.1 European remote sensing satellites (ERS)

ERS-1 and -2 were launched into the same orbit in 1991 and 1995 respectively. The ERS-1 mission, sponsored by ESA, was the first Earth observation program designed to provide environmental monitoring in the microwave spectrum. The primary mission goals were to

provide data measurements of wind speed and direction at the ocean surface, waves length and direction, obtain high resolution mapping of land, ocean, ice, and costal zone, as well as detect changes in land and ocean surface processes. ERS-1 operated from 1991 until 2000 when the attitude control system failed, whereas ERS-2, launched in 1995 and was retired in 2011 after the orbit had decayed from the nominal 780 km to 573 km.

Orbital specification for ERS-1 and -2 are shown in Table A.1.

Table A.1: ERS Orbital characteristics [32]

Feature	Value		
Type	SSO, near circular		
Repeat cycle	3-day	35-day	168-day
Inclination	98.516°	98.543°	98.491°
Altitude	785 km	782 km	770 km
Orbits (cycle)	43	501	2411
Orbits (day)	14 $\frac{1}{3}$	14 $\frac{11}{35}$	14 $\frac{847}{2411}$

Among the several instrumental payloads, both satellites carried the AMI (active microwave instrument), a multimode RADAR operating in C-band (4-8 GHz). The three main modes of operation were directly related to the mission goals: image mode, wave mode, and wind mode. The data we used in our work was obtained in image mode

In AMI, a central local oscillator systems provided a reference at 123 MHz that was up-converted to the nominal RADAR operating frequency of 5.3 GHz. When in image mode, the transmitted pulse consisted of a 37.12 μ s chirp with a 15.55 MHz bandwidth transmitted with a tunable pulse repetition frequency ranging from 1640 Hz to 1720 Hz. Due to the fact that several pulses were transmitted by the satellite before the reception of the first echo from the ground, the system had to actively fine tune and keep track of the delays (typically 10 pulses long) to correctly correlate the received data.

The signal was amplified by a traveling wave tube amplifier and routed to the transmitting/receiving 10×1 m antenna, with a beam width of 5.4° in elevation and 0.288° in azimuth, and a gain of 40 dB.

Upon reception of an echo, the signal was down-converted, corrected for potential transmitting amplifier gain fluctuations, decompressed and routed, together with the reference replica, to a quadrature detector from which amplitude and phase information were sampled every 52.74 ns. The digitized signal was then transmitted to the ground segment where data was further processed to generate the various SAR products.

A summary of the most relevant features of this instrument is provided in Table A.2 (see [156] for more details).

Table A.2: Characteristics of AMI in imaging mode [32]

Feature	Value
Center frequency	5.3 GHz ($\lambda = 5.66$ cm)
Bandwidth	15.55 MHz
Polarization	Linear Vertical (LV)
PRF Range	1640-1720 Hz (2 Hz steps)
Long pulse	37.12 μ s
Compressed pulse	64 ns
Peak power	4.8 kW
Antenna size	10×1 m
Look angle	23° (mid-swath)
Spatial resolution	10-30 m
Swath width	100 km
Data rate	105 Mbit/s

For a brief period (October 1995 - June 1996) the two ERS satellites were joined in a tandem mission. In this configuration, the two instruments were flown in the same orbit with a delay (phasing) of 35 min. This configuration allowed the simultaneous observation of the same scene from two slightly different vantage points providing an interferometric baseline

with excellent temporal and spatial coherence resulting in the generation of very accurate DEM (~ 10 m).

A.2 COSMO-SkyMed

COSMO-SkyMed is a mission funded by the Agenzia Spaziale Italiana (ASI) and the Italian ministry of defense providing both civilian and defense imagery. The 4 satellites, currently orbiting Earth, started deployment on June 8, 2007 and ended on November 6, 2010 and a second generation upgrades has been approved to replace the existing satellites as they reach the projected end-of-life starting in mid 2015, providing continuity at least until 2023.

The mission, consisting of 4 medium-size satellites, deployed, under normal operation, equi-phased in the same orbital plane, is providing at least 4 acquisitions a day at the equator (one left- and one right-looking in ascending orbit, at approximately 6:00, and the same in descending orbit, at approximately 18:00 local time [157]). It was designed to cover a wide range of applications from defense and security to risk management and earth observation.

Nominal orbital information are provided in Table A.3.

Table A.3: COSMO-SkyMed constellation orbital characteristics [32]

Feature	Value
Type	SSO, circular
Contellation repeat cycle	16-day
Satellites	4
Phasing	90°
Individual satellite repeat cycle	5-day
Local Time of Ascending Node	6:00 (dawn/dusk)
Altitude	619.6 km
Orbits (cycle)	237
Orbits (day)	237 ^{13/16}

To achieve 3D SAR imaging of Earth's surface, the constellation can also be operated in *interferometric configuration* which can be implemented either as *in-plane*, where two satellites are in close proximity within the same orbital plane acquiring interferometric images in real time, or *dual-plane* where two of the satellites are placed in different orbit planes with 20'' separation [157].

The SAR system on board of the COSMO-SkyMed satellites (SAR-2000) can operate in three main modes: **spotlight**, providing the *enhanced spotlight mode* product, **stripmap**, resulting in the *himage* and *pingpong* modes, and **scanSAR**, allowing for *wideregion* and *hugeregion* modes. The data collected for our project was obtained in *himage* mode. In this configuration, the imaging method is similar to the one described for the ERS mission with the antenna pointed at a fixed direction with respect to the track. The swath width is about 40 km wide with a typical frame length, in the azimuth direction, of another 40 km corresponding to an acquisition time of 6.5 s. The pulse repetition frequency range is between 2905.9 Hz and 3874.5 Hz with a chirp duration of 35-40 μ s and a bandwidth of 65.64 MHz, at the far range (sampled at 82.50 MHz), and 138.60 MHz (sampled at 176.25 MHz) at the near range. The transmitted pulse is build in the time domain sample after sample (via look up tables) following a model that corrects for the distortion introduced by the hardware. The intermediate frequency is synthesized by single side band modulation with up-conversion to 2400 MHz followed by a $\times 4$ stage to the final X-band driver centered at a frequency of 9.6 GHz [157]. The receiver is a two stage superheterodyne scheme followed by quadrature demodulation operating at 2400 MHz [32].

Table A.4 presents a summary of the most relevant features the SAR-2000 instrument in imaging mode.

Table A.4: Characteristics of the SAR-2000 in imaging mode [32]

Feature	Value
Center frequency	9.6 GHz ($\lambda = 3.125$ cm)
Bandwidth	65.64 MHz
Polarization	Selectable: HH, HV, VH, VV
PRF Range	2905.9-3874.5 Hz
Long pulse	35-40 μ s
Compressed pulse	64 ns
Peak power	14 kW
Antenna size	5.7×1.4 m
Look angle	20-60°
Spatial resolution	3×3 m
Swath width	40 km
Data rate	600 Mbit/s

A.3 Future of SAR

There are several new satellite platforms, designed to produce InSAR raw data, being currently developed and launched by various space agencies. These systems will guarantee continuity in the availability of SAR data. Furthermore, several space agencies have committed to new data policies that will lower the cost of raw data in order to encourage value added processing, thus maintaining high interest for algorithms such as the ones described in this work.

The following is just a subset of the new missions that are planned for the next ten years.

Sentinel-1 [41], part of the Copernicus project [158], is a European Space Agency (ESA) mission comprising several satellites designed to provide data continuity for Earth observation missions with the final goal being the replacement of current older technology reaching retirement such as ERS (see section A.1: *European remote sensing satellites (ERS)*) and Envisat [32]. In particular *Sentinel-1* is a two satellite constellation that will provide C-Band SAR data continuity. *Sentinel-1A* was launched on April, 3 2014 while *Sentinel-1B* will follow in 2015. Both platforms are designed for a lifespan of 7 years with consumables for 12.

The mission will carry a C-Band (4-8 GHz) SAR instrument capable of generating strip-map images, with 80 km swath and a 5x5 m spatial resolution, from a 98.18° sun-synchronous, near-polar, circular orbit at an altitude of 693 km with a repeat cycle of 175 orbits in 12 days.

RADARSAT Constellation Mission (RCM) [159] is a Canadian Space Agency (CSA) mission, designed as the evolution of the RADARSAT program, to provide data continuity, improved operational use of the SAR payload, and system stability. The space segment is comprised of 3 equally-spaced satellites in a sun-synchronous circular orbit (dawn-dusk mission), at an altitude of 532 km with an inclination of 97.74° and a repeat cycle of 179 orbits in 12 days. The SAR instrument is a C-Band (center frequency 5.405 GHz) SAR with a chirp bandwidth of 100 MHz, designed to cover a swath width of 20-350 km with a spatial resolution of 5-50 m. The satellites are planned to be launched in 2018.

Advance land observing satellite "DAICHI-2" (ALOS-2) [160] is a Japanese Aerospace Exploration Agency (JAXA) mission with the objective of providing observations at various incidence angles, to accommodate interferometric analysis of pre- and post-disaster data. ALOS-2, expected to launch in 2014, will provide 14 days revisit time (213 orbits) while parked in a 97.74° sun-synchronous orbit at an altitude of 628 km. The L-band (1-2 GHz) sAR will provide a ground resolution element down to 3x3 m (range-azimuth) with a swath of 50 km.

As far as the processing costs of InSAR data with algorithms such as PSInSAR and SqueeSAR, we expect that, due to the increase availability of data, there will be an increasing competition between service providers that will result in better automation of the processing algorithms and cheaper stack processing.

Appendix B

Project datasets

The datasets we used for this work were provided by TRE Canada and consists of point clouds of both permanent and distributed scatterers obtained by processing the stack of SAR images using their proprietary algorithms PSInSAR and SqueeSAR (see section 1.3.6: *Permanent Scatterer InSAR (PSInSAR) and SqueeSAR*). In these datasets, each scatterer is identified by its location (in the case of PS) or the location of the center of mass of the coherence area (in the case of DS). Furthermore, for each scatterer, the datasets contain additional information that can be used during the analysis phase: elevation with associated standard deviation, coherence level, time series of the elevation changes, average displacement velocity, and acceleration with relative standard deviation.

This project was based on archival ERS (Wink Sinks) and ad hoc COSMO-SkyMED (Virginia) data acquired as part of the RITA funded project.

B.1 Precision of data

The parameters that are used to evaluate the quality of the processed data are:

- Precision of the estimated deformation rates
- Precision of the estimated elevations
- Precision of the geocoding

These element ultimately affect the final results and the subsequent secondary processing.

When a stack of at least 45 images is available for processing then the typical precision that can be applied to scatterers located within 2 km of the reference point¹ are indicated in Table B.1.

Table B.1: Typical precision for PS located within 2 km from the reference point

Measurement	Precision
Deformation rate	$< 1 \text{ mm/yr}$
Displacement error between contiguous SAR images	$< 5 \text{ mm}$
Elevation	$\pm 1.5 \text{ m}$
Positioning error along east direction	$\pm 3 \text{ m}$
Positioning error along north direction	$\pm 2 \text{ m}$

The values listed in Table B.1 can be affected, in both direction, by the following factors:

- Spatial density of the PS/DS (higher density produces higher precision)
- Quality of the RADAR targets (better SNR are preferable)
- Distance from the reference point
- Number of images processed

¹When a stack of SAR images is processed, the most coherent permanent scatterer is identified as reference.

- Period of time covered by the imagery
- Climatic conditions at the time of the acquisitions
- Temporal interval between subsequent acquisitions

Each dataset includes, for each scatterer, standard deviations of the displacement rates which provide a quantitative characterization of the precision of the measured values.

B.2 Wink Sinks

The Wink Sinks dataset was created using the TRE proprietary SqueeSAR algorithm applied to a stack of 22 ERS-1/2 (see section [A.1: *European remote sensing satellites \(ERS\)*](#)) images acquired between June, 3 1992 and February, 21 1998 over a sinkhole active region near Winkler, Texas (see section [2.2.2: *The Winkler sinkholes \(Wink Sinks\)*](#)).

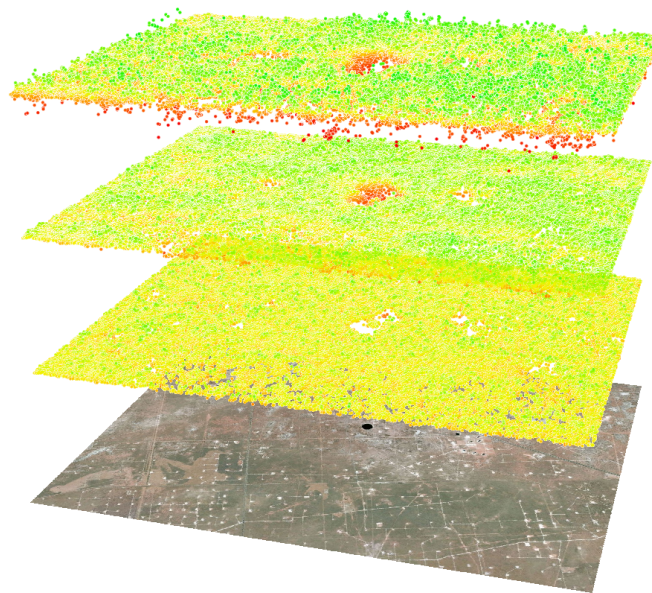


Figure B.1: Wink Sinks dataset at three different time slices.

The result consists in a point cloud dataset comprising 93,513 scattered identified over an area of 55.38 km², each with a 22-point time series. The dataset is characterized by several large temporal gaps including a 2 year gap between 1993 and 1995 and a 1 year gap between 1996 and 1997. As a result, there are few instances of phase unwrapping error of phase ambiguities in the data. It should be recalled that *phase unwrapping* is the process by which the original phase shift measured between SAR images is translated into ground deformation. This is a complex problem and often the global solution is not unique leading to location where ambiguities and error might occur.

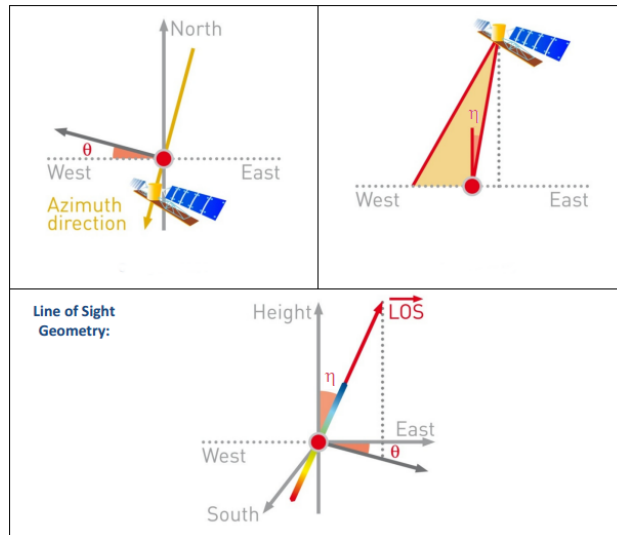


Figure B.2: Geometry of the ERS image acquisitions over the Wink Sinks area. (Courtesy of TRE Canada)

This is common in datasets with large time gaps since the elevation changes during the missing intervals might be large enough to reach the critical level of half wavelength necessary to skip a cycle. In this dataset post analysis was conducted at ambiguity points by looking at the time series of the displacements and evaluating which of the possible solutions best fitted the subsidence data across the large time gaps. Table B.2 summarized the main characteristics of the

Wink Sinks dataset.

All images for this dataset were acquired from a descending orbit with the satellite traveling from north to south and imaging to the west). The geometry of the image acquisition is shown in Fig. B.2 where $\eta = 22.05^\circ$ is the incidence angle formed by the incident beam and

Table B.2: Summary of the Wink Sinks dataset characteristics

Feature	Data
Satellite	ERS-1 and -2
Acquisition geometry	Descending
Analysis time interval	03 June 1992 - 21 February 1998
Number of scenes	22
Projection system used	Latitude, longitude
Datum	WGS 1984
Reference Point location	31°44'39.9654" N, −103°05'58.0518" E
Area of scene	55.38 km ²
Number of PS + DS	93,513
Average PS + DS density	1,689 ^{PS+DS} /km ²

the vertical to the imaged scene, and $\theta = 11.04^\circ$ represent the angle formed between the satellite azimuth traveling direction and the geographic north at the imaging time. From these values is possible to calculate the *line of sight* (LOS) versor:

Table B.3: ERS-1/2 LOS versor for Wink Sinks

Direction	Versor
North	-0.07191
East	0.36846
Vertical	0.92686

These values can be used to determine the sensitivity of the LOS to the vertical, east-west, and north-south directions. For example, the 22.05 off-nadir viewing angle is quite steep and gives rise to a versor value of $\cos \eta = 0.927$ (Table B.3). This implies that the measurement are very sensitive to vertical motion. At the same time, the sensitivity in the east-west (0.368) and in the north-south (-0.072) directions is quite low.

B.3 Virginia

This area of interest correspond to one full COSMO-SkyMed (see section A.2: *COSMO-SkyMed*) radar scene of approximately 1,600 km² in size covering most of Augusta county and a small part of Rock Bridge county in Virginia (see section 3.3.3: *Virginia*).

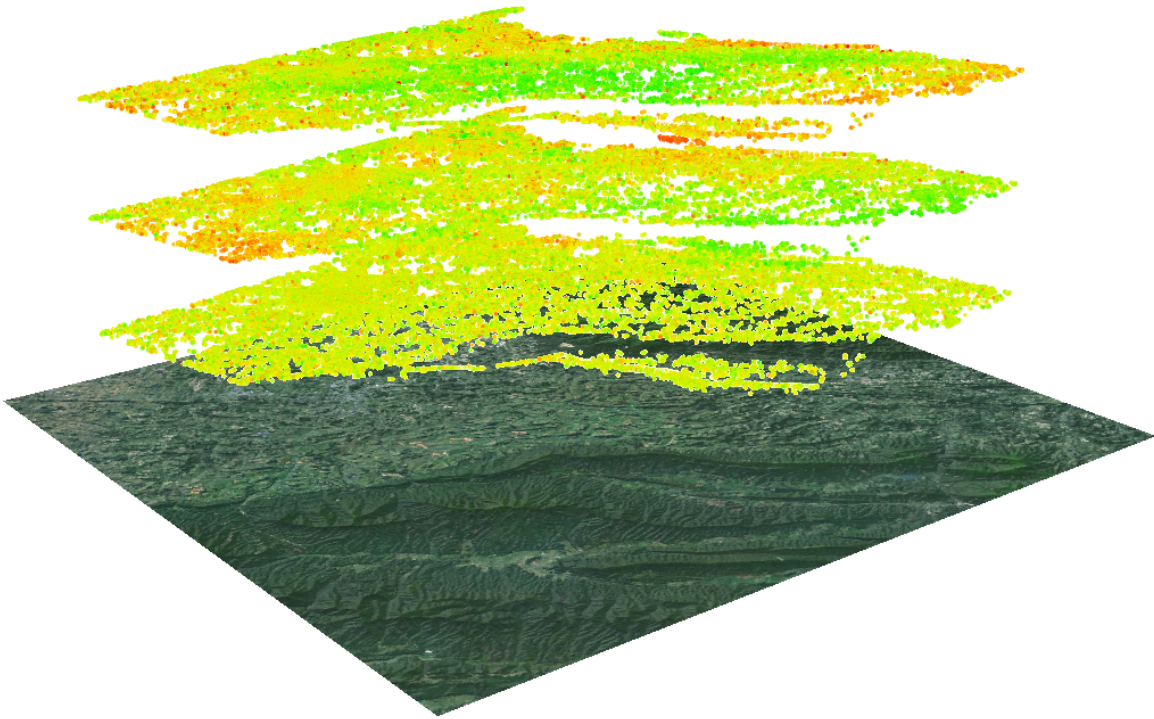


Figure B.3: Virginia dataset at three different time slices.

The RADAR data was acquired as a stack of 32 images between 29 August 2011 and 25 October 2012 (a total of 14 month of continuous monitoring). Images were acquired on a eight day interval for the first two months of data collection (September and October 2011) followed by a regular schedule of 16 days for the majority of the remaining 12 months with an additional set of short time acquisitions in June. Table B.4 shows the schedule of the acquisitions.

Table B.4: Dates of the image acquisition for the Virginia dataset

Number	Date	Number	Date
1	29 Aug 2011	17	23 Mar 2012
2	05 Sept 2011	18	08 Apr 2012
3	13 Sept 2011	19	16 Apr 2012
4	21 Sept 2011	20	02 May 2012
5	29 Sept 2011	21	18 May 2012
6	07 Oct 2011	22	11 Jun 2012
7	15 Oct 2011	23	12 Jun 2012
8	31 Oct 2011	24	15 Jun 2012
9	17 Nov 2011	25	19 Jun 2012
10	02 Dec 2011	26	05 Jul 2012
11	18 Dec 2011	27	21 Jul 2012
12	03 Jan 2012	28	06 Aug 2012
13	19 Jan 2012	29	22 Aug 2012
14	04 Feb 2012	30	23 Sep 2012
15	20 Feb 2012	31	09 Oct 2012
16	07 Mar 2012	32	25 Oct 2012

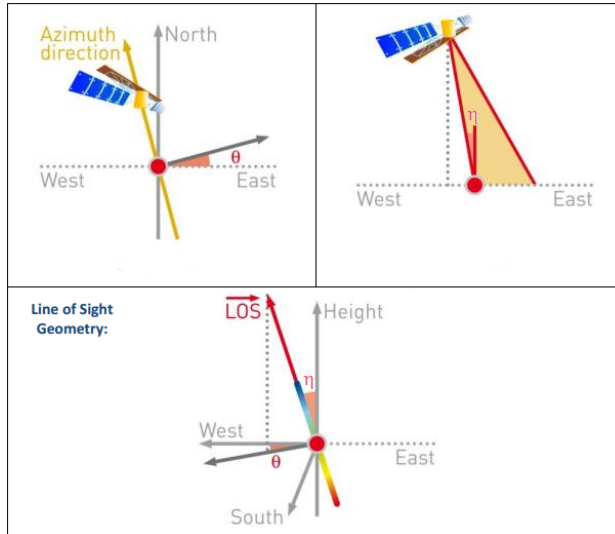


Figure B.4: Geometry of the COSMO-SkyMed image acquisitions over the Virginia area. (Courtesy of TRE Canada)

All images for this dataset were acquired from an ascending orbit with the satellite traveling from south to north and imaging to the east in beam mode H4-0B. This corresponds to COSMO-SkyMed stipmap himage mode with a near range angle of 20.05° and a far range angle of 23.5° [161]. The geometry of the image acquisition is shown in Fig. B.4 where $\eta = 22.24^\circ$ is the incidence

angle formed by the incident beam and the

vertical to the imaged scene, and $\theta = 12.51^\circ$ represent the angle formed between the satellite azimuth traveling direction and the geographic north at the imaging time. From these values

is possible to calculate the *Line Of Sight* (LOS) versor:

Table B.5: COSMO-SkyMed LOS versor for Virginia

Direction	Versor
North	-0.08893
East	-0.40081
Vertical	0.91183

Table B.6 summarized the main characteristics of the Virginia dataset.

Table B.6: Summary of the Virginia dataset characteristics

Feature	Data
Satellite	COSMO-SkyMed
Acquisition geometry	Ascending
Analysis time interval	29 August 2011 - 25 October 2012
Number of scenes	32
Georeferencing	PS aligned on orthophotos (1 foot resolution)
Projection system used	State Plane Virginia North FIPS 4501 (feet)
Datum	NAD 1983
Reference Point location	6744485.69 N - 11323234.49 E
Area of scene	617.8 sq. mile (1,600 km ²)
Number of PS + DS	296,121
Number of PS	166,348
Number of DS	129,773
Average PS + DS density	187 ^{PS+DS} /km ²

Both permanent (PS) and distributed (DS) scatterer were identified in the imaged region. Building, roads, fences and other man-made structures together with natural features such as rocks or exposed ground provide the source for most of the PS whereas DS corresponds to large areas (up to hundreds of square meters) and were most probably generated by large exposed areas such as bare ground or fallow fields and in general large patches of consistent and stable target. As previously mentioned, it is important to consider that while DS are represented as individual points for clarity of representation and ease of interpretation, these

measurement actually correspond to non-point-like features multiple pixel wide. In these measurements, the area of the DS varied from 76 to 891 m².

Bibliography

- [1] R. D. Miller, J. Xia, and D. Steeples, “Seismic reflection characteristics of naturally-induced subsidence affecting transportation,” English, *Journal of Earth Science*, vol. 20, no. 3, pp. 496–512, 2009, ISSN: 1674-487X. DOI: [10.1007/s12583-009-0042-z](https://doi.org/10.1007/s12583-009-0042-z).
- [2] Panel on Land Subsidence, Committee on Ground Failure Hazards Mitigation Research, Division of Natural Hazard Mitigation, National Research Council, *Mitigating Losses from Land Subsidence in the United States*. Washington, D.C.: The National Academies Press, 1991, ISBN: 9780309078726. [Online]. Available: http://www.nap.edu/openbook.php?record_id=1796.
- [3] Risk Assessment Branch, Mitigation Directorate, Federal Emergency Management Agency, *Multi-Hazard Identification and Risk Assessment (MHIRA)*. Washington, D.C.: Federal Emergency Management Agency, 1997. [Online]. Available: <https://www.fema.gov/media-library/assets/documents/7251>.
- [4] P. C. Sirles, *Use of Geophysics for Transportation Projects*, ser. NCHRP Synthesis of Highway Practice No.357. Washington, DC 20001 USA, 2006, p. 116, ISBN: 0309097649. [Online]. Available: http://onlinepubs.trb.org/onlinepubs/nchrp/nchrp_syn_357.pdf.

- [5] D. J. Daniels, "Ground penetrating radar," in *Encyclopedia of RF and Microwave Engineering*. John Wiley & Sons, Inc., 2005, ISBN: 9780471654506. DOI: [10.1002/0471654507.eme152](https://doi.org/10.1002/0471654507.eme152).
- [6] P. W. Mayne, *Cone Penetration Testing*, ser. NCHRP Synthesis of Highway Practice No.368. Washington, DC 20001 USA, 2007, p. 125, ISBN: 9780309097840. [Online]. Available: http://onlinepubs.trb.org...bs/nchrp/nchrp_syn_368.pdf.
- [7] C. Clayton, "The standard penetration test (spt): methods and use," Project Report, 1995. [Online]. Available: <http://eprints.soton.ac.uk/75036/>.
- [8] A. Samouëlian, I. Cousin, A. Tabbagh, A. Bruand, and G. Richard, "Electrical resistivity survey in soil science: a review," *Soil and Tillage Research*, vol. 83, no. 2, pp. 173–193, 2005, ISSN: 0167-1987. DOI: [10.1016/j.still.2004.10.004](https://doi.org/10.1016/j.still.2004.10.004).
- [9] S. Kruse, M. Grasmueck, M. Weiss, and D. Viggiano, "Sinkhole structure imaging in covered karst terrain," *Geophysical Research Letters*, vol. 33, no. 16, pp. 1–6, Aug. 2006, ISSN: 1944-8007. DOI: [10.1029/2006GL026975](https://doi.org/10.1029/2006GL026975).
- [10] S. A. Anderson, R. Surdahl, and B. Young, "InSAR evaluation of landslides and alternative transportation routes," in *Proc. of GeoTrans 2004 - Geotechnical Engineering for Transportation Projects*, Los Angeles, California, U.S.A.: American Society of Civil Engineers, Jul. 2004, pp. 2038–2047. DOI: [10.1061/40744\(154\)199](https://doi.org/10.1061/40744(154)199).
- [11] D. Power, J. Youden, J. English, K. Russell, S. Croshaw, and R. Hanson, "InSAR applications for highway transportation projects," Federal Highway Administration, Lakewood, CO, Tech. Rep. FHWA-CFL/TD-06-002, Apr. 2006.

- [12] J. Morgan, G. Falorni, A. A. Bohane, and F. Novali, “Advanced InSAR technology (SqueeSAR) for monitoring movement of landslides,” Federal Highway Administration, Lakewood, CO, Tech. Rep. FHWA-CFL/TD-11-005, Sep. 2006.
- [13] W.-L. Chang, R. B. Smith, J. Farrell, and C. M. Puskas, “An extraordinary episode of Yellowstone caldera uplift, 2004-2010, from GPS and InSAR observations,” *Geophysical Research Letters*, vol. 37, no. 23, pp. 1944–8007, Dec. 2010. DOI: [10.1029/2010GL045451](https://doi.org/10.1029/2010GL045451).
- [14] F. Canova, C. Tolomei, S. Salvi, G. Toscani, and S. Seno, “Land subsidence along the Ionian coast of SE Sicily (Italy), detection and analysis via small baseline subset (SBAS) multitemporal differential SAR interferometry,” *Earth Surface Processes and Landforms*, vol. 37, no. 3, pp. 273–286, Mar. 2012. DOI: [10.1002/esp.2238](https://doi.org/10.1002/esp.2238).
- [15] H. Klemm, I. Quseimi, F. Novali, A. Ferretti, and A. Tamburini, “Monitoring horizontal and vertical surface deformation over a hydrocarbon reservoir by PSInSAR,” *First Break*, vol. 28, pp. 29–37, May 2010.
- [16] S.-H. Hong, S. Wdowinski, and S.-W. Kim, “Evaluation of TerraSAR-X observations for wetland InSAR application,” *IEEE Trans. on Geoscience and Remote Sensing*, vol. 48, no. 2, pp. 864–873, Feb. 2010. DOI: [10.1109/TGRS.2009.2026895](https://doi.org/10.1109/TGRS.2009.2026895).
- [17] J. G. Paine, S. Buckley, E. W. Collins, C. R. Wilson, and W. Kress, “Assessing sinkhole potential at wink and daisetta using gravity and radar interferometry,” in *22nd Symposium on the Application of Geophysics to Engineering and Environmental Problems*. Forth Worth, Texas: Society of Expolration Geophysicists, Apr. 2009, ch. 57, pp. 480–488. DOI: [10.4133/1.3176733](https://doi.org/10.4133/1.3176733).

- [18] J. G. Paine, S. M. Buckley, E. W. Collins, and C. R. Wilson, "Assessing collapse risk in evaporite sinkhole-prone areas using microgravimetry and radar interferometry," *Journal of Environmental & Engineering Geophysics*, vol. 17, no. 2, pp. 75–87, Jun. 2012. DOI: [10.2113/JEEG17.2.75](https://doi.org/10.2113/JEEG17.2.75).
- [19] M. L. Rucker, B. B. Panda, R. A. Meyers, and J. C. Lommler, "Using InSAR to detect subsidence at brine wells, sinkhole sites, and mines," *Carbonates and Evaporites*, pp. 1–7, 2013. DOI: [10.1007/s13146-013-0134-9](https://doi.org/10.1007/s13146-013-0134-9).
- [20] A. Pigorini, M. Ricci, A. Sciotti, C. Giannico, and A. Tamburini, "Satellite remote-sensing PSInSAR technique applied to design and construction of railway infrastructures," *Ingegneria Ferroviaria*, vol. 9, pp. 729–757, Sep. 2010.
- [21] J. A. Harrell and V. M. Brown, "The world's oldest surviving geological map: the 1150 b.c. turin papyrus from egypt," *The Journal of Geology*, vol. 100, no. 1, pp. 3–18, 1992, ISSN: 00221376. [Online]. Available: <http://www.jstor.org/stable/30082315>.
- [22] M. Guarnieri, "The early history of radar [historical]," *Industrial Electronics Magazine, IEEE*, vol. 4, no. 3, pp. 36–42, 2010, ISSN: 1932-4529. DOI: [10.1109/MIE.2010.937936](https://doi.org/10.1109/MIE.2010.937936).
- [23] G. Curry, *Radar System Performance Modeling (Second Edition)*, ser. Radar System Performance Modeling. ARTECH HOUSE Incorporated, 2005, p. 394, ISBN: 9781580538169.
- [24] J. Y. N. Cho, "Multi-pri signal processing for the terminal doppler weather radar. part ii: rangevelocity ambiguity mitigation," *Journal of Atmospheric and Oceanic Technology*, vol. 22, pp. 1507–1519, 10 Apr. 2005. DOI: [10.1175/JTECH1805.1](https://doi.org/10.1175/JTECH1805.1).

- [25] B. R. Mahafza, *Radar Signal Analysis and Processing Using MATLAB*. Boca Raton, FL 33487-2742: Chapman and Hall/CRC, 2009, ch. 8, pp. 315–351, 504 pp., ISBN: 9781420066432. DOI: [10.1201/9781420066449.ch8](https://doi.org/10.1201/9781420066449.ch8).
- [26] G. L. Turin, “An introduction to matched filters,” *IRE Trans. on Information Theory*, vol. 6, no. 3, pp. 311–329, Jun. 1960, ISSN: 0096-1000. DOI: [10.1109/TIT.1960.1057571](https://doi.org/10.1109/TIT.1960.1057571).
- [27] ———, “An introduction to digital matched filters,” *Proceedings of the IEEE*, vol. 64, no. 7, pp. 1092–1112, Jul. 1976, ISSN: 0018-9219. DOI: [10.1109/PROC.1976.10274](https://doi.org/10.1109/PROC.1976.10274).
- [28] J. C. Curlander and R. N. McDonough, *Synthetic Aperture Radar: System and Signal Processing*, ser. Wiley Series in Remote Sensing and Image Processing. New York: Wiley, Nov. 1991, 672 pp., ISBN: 9780471857709.
- [29] P. Capece, “Active sar antennas: design, development, and current programs,” *International Journal of Antennas and Propagation*, vol. 2009, 2009. DOI: [10.1155/2009/796064](https://doi.org/10.1155/2009/796064).
- [30] A. Torre and P. Capece, “Cosmo-skymed: the advanced sar instrument,” in *Recent Advances in Space Technologies (RAST), 2011 5th International Conference on*, Jun. 2011, pp. 865–868. DOI: [10.1109/RAST.2011.5966966](https://doi.org/10.1109/RAST.2011.5966966).
- [31] L. Graham, “Synthetic interferometer radar for topographic mapping,” *Proceedings of the IEEE*, vol. 62, no. 6, pp. 763–768, Jun. 1974, ISSN: 0018-9219. DOI: [10.1109/PROC.1974.9516](https://doi.org/10.1109/PROC.1974.9516).
- [32] ESA. (2014). Eoportals directory: satellite missions database, [Online]. Available: <https://directory.eoportal.org/web/eoportal/satellite-missions>.

- [33] A. Love, “In memory of carl a. wiley,” *Antennas and Propagation Society Newsletter, IEEE*, vol. 27, no. 3, pp. 17–18, 1985, ISSN: 2168-0329. DOI: [10.1109/MAP.1985.27810](https://doi.org/10.1109/MAP.1985.27810).
- [34] C. A. Wiley, “Pulsed doppler radar method and apparatus,” pat. 3,196,436, 1965. [Online]. Available: <http://pdfpiw.uspto.gov/.piw?Docid=03196436>.
- [35] L. Cutrona, W. E. Vivian, E. Leith, and G. O. Hall, “A high-resolution radar combat-surveillance system,” *Military Electronics, IRE Transactions on*, vol. MIL-5, no. 2, pp. 127–131, 1961, ISSN: 0096-2511. DOI: [10.1109/IRET-MIL.1961.5008330](https://doi.org/10.1109/IRET-MIL.1961.5008330).
- [36] C. Oliver and S. Quegan, *Understanding synthetic aperture radar images*. Raleigh, NC 27613: SciTech Publishing, Jan. 2004, 479 pp., ISBN: 9781891121319. [Online]. Available: <http://app.knovel.com/hotlink/toc/id:kpUSARI005/understanding-synthetic>.
- [37] K. Tomiyasu, “Tutorial review of synthetic-aperture radar (sar) with applications to imaging of the ocean surface,” *Proceedings of the IEEE*, vol. 66, no. 5, pp. 563–583, May 1978, ISSN: 0018-9219. DOI: [10.1109/PROC.1978.10961](https://doi.org/10.1109/PROC.1978.10961).
- [38] N. R. Canada. (Oct. 30, 2013). Radar image distortion, [Online]. Available: <http://www.nrcan.gc.ca/earth-sciences/geomatics/satellite-imagery-air-photos/satellite-imagery-products/educational-resources/9325>.
- [39] A. E. E. Rogers and R. P. Ingalls, “Venus: mapping the surface reflectivity by radar interferometry,” English, *Science*, New Series, vol. 165, no. 3895, pp. 797–799, 1969, ISSN: 00368075. [Online]. Available: <http://www.jstor.org/stable/1727374>.

- [40] P. Rosen, S. Hensley, I. Joughin, F. K. Li, S. Madsen, E. Rodriguez, and R. M. Goldstein, "Synthetic aperture radar interferometry," *Proceedings of the IEEE*, vol. 88, no. 3, pp. 333–382, Mar. 2000, ISSN: 0018-9219. DOI: [10.1109/5.838084](https://doi.org/10.1109/5.838084).
- [41] A. Ferretti, A. Monti-Guarnieri, C. Prati, F. Rocca, and D. Massonnet, *InSAR principles: guidelines for SAR interferometry processing and interpretation*, K. Fletcher, Ed. Noordwijk, the Netherlands: ESA Publication, ESTEC, Feb. 2007, ISBN: 9290922338.
[Online]. Available: http://www.esa.int/About_Us/ESA_Publications/InSAR_Principles_Guidelines_for_SAR_Interferometry_Processing_and_Interpretation_br_ESA_TM-19.
- [42] H. A. Zebker and J. Villasenor, "Decorrelation in interferometric radar echoes," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 30, no. 5, pp. 950–959, Sep. 1992, ISSN: 0196-2892. DOI: [10.1109/36.175330](https://doi.org/10.1109/36.175330).
- [43] M. Richards, "A beginner's guide to interferometric sar concepts and signal processing [aess tutorial iv]," *Aerospace and Electronic Systems Magazine, IEEE*, vol. 22, no. 9, pp. 5–29, Sep. 2007, ISSN: 0885-8985. DOI: [10.1109/MAES.2007.4350281](https://doi.org/10.1109/MAES.2007.4350281).
- [44] D. Massonnet, M. Rossi, C. Carmona, F. Adragna, G. Peltzer, K. Feigl, and T. Rabaute, "The displacement field of the Landers earthquake mapped by radar interferometry," *Nature*, vol. 364, pp. 138–142, Jul. 1993. DOI: [10.1038/364138a0](https://doi.org/10.1038/364138a0).
- [45] E. Calais, A. Freed, G. Mattioli, F. Amelung, S. Jonsson, P. Jansma, S.-H. Hong, T. Dixon, C. Prepetit, and R. Momplaisir, "Transpressional rupture of an unmapped fault during the 2010 Haiti earthquake," *Nature Geoscience*, vol. 3, pp. 794–799, Oct. 2010. DOI: [10.1038/ngeo992](https://doi.org/10.1038/ngeo992).

- [46] S. Atzori, I. Hunstad, M. Chini, S. Salvi, C. Tolomei, C. Bignami, S. Stramondo, E. Trasatti, A. Antonioli, and E. Boschi, “Finite fault inversion of DInSAR coseismic displacement of the 2009 L’Aquila earthquake (central Italy),” *Geophysical Research Letters*, vol. 36, no. 15, pp. 1–6, 2009, ISSN: 1944-8007. DOI: [10.1029/2009GL039293](https://doi.org/10.1029/2009GL039293).
- [47] C. Bignami, S. Borgstrom, M. Chini, F. Guglielmino, C. Kyriakopoulos, D. Melini, G. Puglisi, V. Siniscalchi, and S. Stramondo, “Joint inversion of the 2011 Tohoku (Japan) earthquake from DInSAR and GPS data,” in *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*, Jul. 2012, pp. 1186–1189. DOI: [10.1109/IGARSS.2012.6351336](https://doi.org/10.1109/IGARSS.2012.6351336).
- [48] D. Massonnet and K. L. Feigl, “Radar interferometry and its application to changes in the earth’s surface,” *Reviews of Geophysics*, vol. 36, no. 4, pp. 441–500, 1998, ISSN: 1944-9208. DOI: [10.1029/97RG03139](https://doi.org/10.1029/97RG03139).
- [49] A. K. Gabriel, R. M. Goldstein, and H. A. Zebker, “Mapping small elevation changes over large areas: differential radar interferometry,” *Journal of Geophysical Research: Solid Earth*, vol. 94, no. B7, pp. 9183–9191, 1989, ISSN: 2156-2202. DOI: [10.1029/JB094iB07p09183](https://doi.org/10.1029/JB094iB07p09183).
- [50] Z. Lu, “InSAR imaging of volcanic deformation over cloud-prone areas - aleutian islands,” *Photogrammetric Engineering & Remote Sensing*, vol. 73, no. 3, pp. 245–257, Mar. 2007. [Online]. Available: http://info.asprs.org/publications/pers/2007journal/march/2007_mar_245-257.pdf.
- [51] Z. Li, E. J. Fielding, P. Cross, and R. Preusker, “Advanced InSAR atmospheric correction: MERIS/MODIS combination and stacked water vapour models,” *In-*

- ternational Journal of Remote Sensing*, vol. 30, no. 13, pp. 3343–3363, 2009. DOI: [10.1080/01431160802562172](https://doi.org/10.1080/01431160802562172).
- [52] R. M. Goldstein, “Atmospheric limitations to repeat-track radar interferometry,” *Geophysical Research Letters*, vol. 22, no. 18, pp. 2517–2520, 1995, ISSN: 1944-8007. DOI: [10.1029/95GL02475](https://doi.org/10.1029/95GL02475).
- [53] A. Ferretti, C. Prati, and F. Rocca, “Multibaseline insar dem reconstruction: the wavelet approach,” *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 37, no. 2, pp. 705–715, 1999, ISSN: 0196-2892. DOI: [10.1109/36.752187](https://doi.org/10.1109/36.752187).
- [54] S. Tebaldini and A. Monti-Guarnieri, “Methods and performances for multi-pass sar interferometry,” in *Geoscience and Remote Sensing New Achievements*, ed. by P. Imperatore and D. Riccio. Intech, 2010, ch. 18, ISBN: 978-953-7619-97-8. DOI: [10.5772/9112](https://doi.org/10.5772/9112).
- [55] A. Ferretti, C. Prati, and F. Rocca, “Permanent scatterers in sar interferometry,” *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 39, no. 1, pp. 8–20, 2001, ISSN: 0196-2892. DOI: [10.1109/36.898661](https://doi.org/10.1109/36.898661).
- [56] Array System Computing Inc. (under ESA contract number 20698/07/I-LG). (Feb. 28, 2013). Next ESA SAR Toolbox (NEST), [Online]. Available: <http://nest.array.ca/web/nest>.
- [57] A. Ferretti, C. Prati, and F. Rocca, “Nonlinear subsidence rate estimation using permanent scatterers in differential sar interferometry,” *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 38, no. 5, pp. 2202–2212, 2000, ISSN: 0196-2892. DOI: [10.1109/36.868878](https://doi.org/10.1109/36.868878).

- [58] C. Colesanti, A. Ferretti, C. Prati, and F. Rocca, “Monitoring landslides and tectonic motions with the permanent scatterers technique,” *Engineering Geology*, vol. 68, no. 12, pp. 3–14, 2003, Remote sensing and monitoring of landslides, ISSN: 0013-7952. DOI: [10.1016/S0013-7952\(02\)00195-3](https://doi.org/10.1016/S0013-7952(02)00195-3).
- [59] A. Ferretti, C. Prati, and F. Rocca, “Permanent scatterers in sar interferometry,” in *Geoscience and Remote Sensing Symposium, 1999. IGARSS '99 Proceedings. IEEE 1999 International*, vol. 3, 1999, pp. 1528–1530. DOI: [10.1109/IGARSS.1999.772008](https://doi.org/10.1109/IGARSS.1999.772008).
- [60] F. Rocca. (2010). Advance training course on land remote sensing - application for terrain motion, European Space Agency, [Online]. Available: <http://earth.esa.int/landtraining07/material.html>.
- [61] A. Ferretti, A. Fumagalli, F. Novali, C. Prati, F. Rocca, and A. Rucci, “A new algorithm for processing interferometric data-stacks: squeeSAR,” *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 49, no. 9, pp. 3460–3470, 2011, ISSN: 0196-2892. DOI: [10.1109/TGRS.2011.2124465](https://doi.org/10.1109/TGRS.2011.2124465).
- [62] J. W. Tukey, *Exploratory data analysis*, ser. Addison-Wesley series in behavioral science. Quantitative methods. Addison-Wesley Publishing Co., 1977, p. 688, ISBN: 978-0201076165, Massachusetts.
- [63] A. Waltham and P. Fookes, “Engineering classification of karst ground conditions,” *Quarterly Journal of Engineering Geology and Hydrogeology*, vol. 36, no. 2, pp. 101–118, May 2003.

- [64] T. Waltham, F. G. Bell, and M. G. Culshaw, *Sinkholes and Subsidence*. Chichester, United Kingdom: Springer Berlin Heidelberg, 2005, p. 382, ISBN: 3540207252, 9783540207252. DOI: [10.1007/b138363](https://doi.org/10.1007/b138363).
- [65] S. Filin, A. Baruch, Y. Avni, and S. Marco, “Sinkhole characterization in the dead sea area using airborne laser scanning,” English, *Natural Hazards*, vol. 58, no. 3, pp. 1135–1154, 2011, ISSN: 0921-030X. DOI: [10.1007/s11069-011-9718-7](https://doi.org/10.1007/s11069-011-9718-7).
- [66] F. Gutierrez, J. Guerrero, and P. Lucha, “A genetic classification of sinkholes illustrated from evaporite paleokarst exposures in Spain,” English, *Environmental Geology*, vol. 53, no. 5, pp. 993–1006, 2008, ISSN: 0943-0105. DOI: [10.1007/s00254-007-0727-5](https://doi.org/10.1007/s00254-007-0727-5).
- [67] Y. Gao, J. Alexander E.C., and R. Barnes, “Karst database implementation in Minnesota: analysis of sinkhole distribution,” English, *Environmental Geology*, vol. 47, no. 8, pp. 1083–1098, 2005, ISSN: 0943-0105. DOI: [10.1007/s00254-005-1241-2](https://doi.org/10.1007/s00254-005-1241-2).
- [68] J. P. Galve, F. Gutierrez, P. Lucha, J. Guerrero, J. Bonachea, J. Remondo, and A. Cendrero, “Probabilistic sinkhole modelling for hazard assessment,” *Earth Surface Processes and Landforms*, vol. 34, no. 3, pp. 437–452, Mar. 2009. DOI: [10.1002/esp.1753](https://doi.org/10.1002/esp.1753).
- [69] M. Lamelas, O. Marinoni, A. Hoppe, and J. Riva, “Doline probability map using logistic regression and GIS technology in the central Ebro basin (Spain),” English, *Environmental Geology*, vol. 54, no. 5, pp. 963–977, 2008, ISSN: 0943-0105. DOI: [10.1007/s00254-007-0895-3](https://doi.org/10.1007/s00254-007-0895-3).

- [70] C. E. Augarde, A. V. Lyamin, and S. W. Sloan, "Prediction of undrained sinkhole collapse," *Journal of Geotechnical and Geoenvironmental Engineering*, vol. 129, no. 3, pp. 197–205, 2003. DOI: [10.1061/\(ASCE\)1090-0241\(2003\)129:3\(197\)](https://doi.org/10.1061/(ASCE)1090-0241(2003)129:3(197)).
- [71] W. Z. Savage, "Prediction of vertical displacements in a subsiding elastic layer: a model for subsidence in karst terrains," U.S. Geological Survey, USGS Numbered Series, 1979, p. 16. [Online]. Available: <http://pubs.usgs.gov/of/1979/1094/report.pdf>.
- [72] E. A. Sorochan, G. M. Troitzky, P. D. Grigoruk, and V. V. Tolmachyov, "Mathematical modelling and prognose dimensions of karst sinkholes in multilayered soil bases of buildings and structures," in *Proc 6th International Conference on Numerical Methods in Geomechanics, Innsbruck, 1115 April 1988*, vol. 1, Rotterdam: A A Balkema, 1988, pp. 515–519. DOI: [10.1016/0148-9062\(90\)92739-2](https://doi.org/10.1016/0148-9062(90)92739-2).
- [73] U. G. Survey, *Land Subsidence in the United States (Circular 1182)*, D. Galloway, D. R. Jones, and S. Ingebritsen, Eds. Reston, VA: U.S. Geological Survey, 1999, ISBN: 0607926961. [Online]. Available: <http://pubs.usgs.gov/circ/circ1182/>.
- [74] D. F. Ritter, R. C. Kockel, and J. J. R. Miller, *Process Geomorphology*. Long Groove, IL: Waveland Press, Inc., 2011, p. 672, ISBN: 9781577666691.
- [75] J. F. Poland, B. E. Lofgren, R. L. Ireland, and R. G. Pugh, "Land subsidence in the san joaquin valley, california, as of 1972," U.S. Geological Survey, USGS Numbered Series, 1975, p. 77. [Online]. Available: <http://pubs.usgs.gov/pp/0437h/report.pdf>.
- [76] C. E. Jones and R. G. Blom, "Bayou corne, louisiana, sinkhole: precursory deformation measured by radar interferometry," *Geology*, vol. 42, no. 2, pp. 111–114, 2014. DOI: [10.1130/G34972.1](https://doi.org/10.1130/G34972.1).

- [77] K. Johnson, “Development of the Wink Sink in west Texas, U.S.A., due to salt dissolution and collapse,” *Environmental Geology and Water Sciences*, vol. 14, no. 2, pp. 81–92, Sep. 1989, ISSN: 0099-0094. DOI: [10.1007/BF01728499](https://doi.org/10.1007/BF01728499).
- [78] A. Vaccari, M. Stuecheli, B. Bruckno, E. Hoppe, and S. T. Acton, “Detection of geophysical features in InSAR point cloud data set using spatiotemporal models,” *International Journal of Remote Sensing*, vol. 34, no. 22, pp. 8215–8234, 2013. DOI: [10.1080/01431161.2013.833357](https://doi.org/10.1080/01431161.2013.833357).
- [79] M. Ahmed and M. Iskander, “Analysis of tunneling-induced ground movements using transparent soil models,” *Journal of Geotechnical and Geoenvironmental Engineering*, vol. 137, no. 5, pp. 525–535, May 2011. DOI: [10.1061/\(ASCE\)GT.1943-5606.0000456](https://doi.org/10.1061/(ASCE)GT.1943-5606.0000456).
- [80] K. Neaupane and N. Adhikari, “Prediction of tunneling-induced ground movement with the multi-layer perceptron,” *Tunnelling and Underground Space Technology*, vol. 21, no. 2, pp. 151–159, 2006, ISSN: 0886-7798. DOI: [10.1016/j.tust.2005.07.001](https://doi.org/10.1016/j.tust.2005.07.001).
- [81] R. B. Peck, “Deep sxcavation and tunneling in soft groun,” in *Proceedings of the Seventh International Conference of Soil Mechanics and Foundation Engineering*, O. Committee, Ed., Mexico City, 1969, pp. 225–290.
- [82] S. Abe, D. Place, and P. Mora, “A parallel implementation of the lattice solid model for the simulation of rock mechanics and earthquake dynamics,” English, *pure and applied geophysics*, vol. 161, no. 11-12, pp. 2265–2277, 2004, ISSN: 0033-4553. DOI: [10.1007/s00024-004-2562-x](https://doi.org/10.1007/s00024-004-2562-x).
- [83] P. S. Cundall and O. D. L. Strack, “A discrete numerical model for granular assemblies,” *Gotechnique*, vol. 29, no. 1, pp. 47–65, Mar. 1979. DOI: [10.1680/geot.1979.29.1.47](https://doi.org/10.1680/geot.1979.29.1.47).

- [84] O. R. Walton and R. L. Braun, “Viscosity, granular-temperature, and stress calculations for shearing assemblies of inelastic, frictional disks,” *Journal of Rheology*, vol. 30, no. 5, pp. 949–980, 1986. DOI: [10.1122/1.549893](https://doi.org/10.1122/1.549893).
- [85] K. Dimou, “3-d hybrid eulerian-lagrangian/particle tracking model for simulating mass transport in coastal water bodies,” Ph.D. Dissertation, Massachusetts Institute of Technology, 1992, p. 233. [Online]. Available: <http://dspace.mit.edu/handle/1721.1/28011>.
- [86] L. Noels, L. Stainier, and J.-P. Ponthot, “Combined implicit/explicit time-integration algorithms for the numerical simulation of sheet metal forming,” *Journal of Computational and Applied Mathematics*, vol. 168, no. 12, pp. 331–339, 2004, ISSN: 0377-0427. DOI: [10.1016/j.cam.2003.12.004](https://doi.org/10.1016/j.cam.2003.12.004).
- [87] T. Pöschel and T. Schwager, *Computational Granular Dynamics: Models and Algorithms*. Berlin, Germany: Springer, 2005.
- [88] Y. Wang, S. Abe, S. Latham, and P. Mora, “Implementation of particle-scale rotation in the 3-d lattice solid model,” English, in *Computational Earthquake Physics: Simulations, Analysis and Infrastructure, Part I*, ser. Pageoph Topical Volumes, X.-c. Yin, P. Mora, A. Donnellan, and M. Matsuura, Eds., Birkhuser Basel, 2006, pp. 1769–1785, ISBN: 978-3-7643-7991-9. DOI: [10.1007/978-3-7643-7992-6_4](https://doi.org/10.1007/978-3-7643-7992-6_4).
- [89] B. Look, *Handbook of geotechnical investigation and design tables*, 2nd. CRC Press/Balkema, 2014, p. 418, ISBN: 9781138001398.

- [90] L. J. Guibas. (Dec. 5, 2008). Detection of symmetries and repeated patterns in 3d point cloud data, Stanford University, [Online]. Available: http://videlectures.net/etvc08_guibas_dosarp/.
- [91] L. J. Grady and J. R. Polimeni, *Discrete Calculus*. Springer London, 2010, ISBN: 9781849962896. DOI: [10.1007/978-1-84996-290-2](https://doi.org/10.1007/978-1-84996-290-2).
- [92] U. von Luxburg, “A tutorial on spectral clustering,” English, *Statistics and Computing*, vol. 17, no. 4, pp. 395–416, 2007, ISSN: 0960-3174. DOI: [10.1007/s11222-007-9033-z](https://doi.org/10.1007/s11222-007-9033-z).
- [93] S. E. Schaeffer, “Graph clustering,” *Computer Science Review*, vol. 1, no. 1, pp. 27–64, 2007, ISSN: 1574-0137. DOI: [10.1016/j.cosrev.2007.05.001](https://doi.org/10.1016/j.cosrev.2007.05.001).
- [94] L. Page, “Method for node ranking in a linked database,” pat. 6,285,999, 2001. [Online]. Available: <http://pdfpiw.uspto.gov/.piw?Docid=6285999>.
- [95] M. Stuecheli, A. Vaccari, and S. Acton, “Graph cut segmentation of sparsely sampled images with application to insar-measured changes in elevation,” in *Image Analysis and Interpretation (SSIAI), 2012 IEEE Southwest Symposium on*, Santa Fe, New Mexico, U.S.A.: IEEE Computer Society, Apr. 2012, pp. 149–152. DOI: [10.1109/SSIAI.2012.6202475](https://doi.org/10.1109/SSIAI.2012.6202475).
- [96] F. Yi and I. Moon, “Image segmentation: a survey of graph-cut methods,” in *Systems and Informatics (ICSAI), 2012 International Conference on*, May 2012, pp. 1936–1941. DOI: [10.1109/ICSAI.2012.6223428](https://doi.org/10.1109/ICSAI.2012.6223428).
- [97] R. C. Gonzales and R. E. Woods, *Digital Image Processing*, 3rd. Upper Saddle River, New Jersey 07458: Pearson Education, 2011, ISBN: 9780133002324.

- [98] S. Raut, M. Raghuvanshi, R. Dharaskar, and A. Raut, “Image segmentation n.150; a state-of-art survey for prediction,” in *Advanced Computer Control, 2009. ICACC '09. International Conference on*, 2009, pp. 420–424. DOI: [10.1109/ICACC.2009.78](https://doi.org/10.1109/ICACC.2009.78).
- [99] N. Otsu, “A threshold selection method from gray-level histograms,” *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 9, no. 1, pp. 62–66, 1979, ISSN: 0018-9472. DOI: [10.1109/TSMC.1979.4310076](https://doi.org/10.1109/TSMC.1979.4310076).
- [100] J. Canny, “A computational approach to edge detection,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. PAMI-8, no. 6, pp. 679–698, 1986, ISSN: 0162-8828. DOI: [10.1109/TPAMI.1986.4767851](https://doi.org/10.1109/TPAMI.1986.4767851).
- [101] I. Karoui, R. Fablet, J. M Boucher, and J. Augustin, “Variational region-based segmentation using multiple texture statistics,” *Image Processing, IEEE Transactions on*, vol. 19, no. 12, pp. 3146–3156, 2010, ISSN: 1057-7149. DOI: [10.1109/TIP.2010.2071290](https://doi.org/10.1109/TIP.2010.2071290).
- [102] D. Mukherjee, N. Ray, and S. Acton, “Level set analysis for leukocyte detection and tracking,” *Image Processing, IEEE Transactions on*, vol. 13, no. 4, pp. 562–572, Apr. 2004, ISSN: 1057-7149. DOI: [10.1109/TIP.2003.819858](https://doi.org/10.1109/TIP.2003.819858).
- [103] Y. Boykov and G. Funka-Lea, “Graph cuts and efficient n-d image segmentation,” English, *International Journal of Computer Vision*, vol. 70, no. 2, pp. 109–131, 2006, ISSN: 0920-5691. DOI: [10.1007/s11263-006-7934-5](https://doi.org/10.1007/s11263-006-7934-5).
- [104] H. Lombaert, Y. Sun, and F. Cheriet, “Fast 4d segmentation of large datasets using graph cuts,” vol. 7962, 2011, 79622H–79622H–7. DOI: [10.1117/12.877505](https://doi.org/10.1117/12.877505).

- [105] T. Yedidya, P. Carr, R. Hartley, and J.-P. Guillon, “Enforcing monotonic temporal evolution in dry eye images,” in *Medical Image Computing and Computer-Assisted Intervention MICCAI 2009*, ser. Lecture Notes in Computer Science, G.-Z. Yang, D. Hawkes, D. Rueckert, A. Noble, and C. Taylor, Eds., vol. 5762, Springer Berlin Heidelberg, 2009, pp. 976–984, ISBN: 978-3-642-04270-6. DOI: [10.1007/978-3-642-04271-3_118](https://doi.org/10.1007/978-3-642-04271-3_118).
- [106] Y. Tarabalka and G. Charpiat, “A graph-cut-based method for spatio-temporal segmentation of fire from satellite observations,” in *Geoscience and Remote Sensing Symposium (IGARSS), 2013 IEEE International*, Jul. 2013, pp. 3494–3497. DOI: [10.1109/IGARSS.2013.6723582](https://doi.org/10.1109/IGARSS.2013.6723582).
- [107] A. K. Jain, “Data clustering: 50 years beyond k-means,” *Pattern Recognition Letters*, vol. 31, no. 8, pp. 651–666, 2010, Award winning papers from the 19th International Conference on Pattern Recognition (ICPR) 19th International Conference in Pattern Recognition (ICPR), ISSN: 0167-8655. DOI: [10.1016/j.patrec.2009.09.011](https://doi.org/10.1016/j.patrec.2009.09.011).
- [108] B. N. Delaunay, “Sur la sphère vide,” *Bulletin of Academy of Sciences of the USSR*, vol. 7, no. 6, pp. 793–800, 1934.
- [109] Z. Wu and R. Leahy, “An optimal graph theoretic approach to data clustering: theory and its application to image segmentation,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 15, no. 11, pp. 1101–1113, 1993, ISSN: 0162-8828. DOI: [10.1109/34.244673](https://doi.org/10.1109/34.244673).
- [110] L. Hagen and A. Kahng, “Fast spectral methods for ratio cut partitioning and clustering,” in *Computer-Aided Design, 1991. ICCAD-91. Digest of Technical Papers.*,

- 1991 *IEEE International Conference on*, 1991, pp. 10–13. DOI: [10.1109/ICCAD.1991.185177](https://doi.org/10.1109/ICCAD.1991.185177).
- [111] J. Shi and J. Malik, “Normalized cuts and image segmentation,” in *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, 1997, pp. 731–737. DOI: [10.1109/CVPR.1997.609407](https://doi.org/10.1109/CVPR.1997.609407).
- [112] Y. Boykov and V. Kolmogorov, “An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26, no. 9, pp. 1124–1137, 2004, ISSN: 0162-8828. DOI: [10.1109/TPAMI.2004.60](https://doi.org/10.1109/TPAMI.2004.60).
- [113] V. Kolmogorov and R. Zabih, “What energy functions can be minimized via graph cuts?” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26, no. 2, pp. 147–159, 2004, ISSN: 0162-8828. DOI: [10.1109/TPAMI.2004.1262177](https://doi.org/10.1109/TPAMI.2004.1262177).
- [114] Y. Boykov, O. Veksler, and R. Zabih, “Fast approximate energy minimization via graph cuts,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 23, no. 11, pp. 1222–1239, 2001, ISSN: 0162-8828. DOI: [10.1109/34.969114](https://doi.org/10.1109/34.969114).
- [115] S. Bagon, *Matlab wrapper for graph cut*, Dec. 2006. [Online]. Available: <http://www.wisdom.weizmann.ac.il/~bagon>.
- [116] L. Giglio, T. Loboda, D. P. Roy, B. Quayle, and C. O. Justice, “An active-fire based burned area mapping algorithm for the {modis} sensor,” *Remote Sensing of Environment*, vol. 113, no. 2, pp. 408–420, 2009, ISSN: 0034-4257. DOI: [10.1016/j.rse.2008.10.006](https://doi.org/10.1016/j.rse.2008.10.006).

- [117] P. V. C. Hough, “Machine analysis of bubble chamber pictures,” in *International Conference on High Energy Accelerators and Instrumentation*, Geneva, Switzerland: CERN, Sep. 1959, pp. 554–558.
- [118] D. Ballard, “Generalizing the hough transform to detect arbitrary shapes,” *Pattern Recognition*, vol. 13, no. 2, pp. 111–122, 1981. DOI: [10.1016/0031-3203\(81\)90009-1](https://doi.org/10.1016/0031-3203(81)90009-1).
- [119] J. Wallis, T. R. Miller, C. Lerner, and E. Kleerup, “Three-dimensional display in nuclear medicine,” *Medical Imaging, IEEE Transactions on*, vol. 8, no. 4, pp. 297–230, 1989, ISSN: 0278-0062. DOI: [10.1109/42.41482](https://doi.org/10.1109/42.41482).
- [120] M. Stuecheli, “Automatic sinkhole detection from satellite-based insar,” Master’s thesis, University of Virginia, 2013. [Online]. Available: <http://libra.virginia.edu/catalog/libra-oa:2845>.
- [121] E. Bingham, *Publication 105: physiographic diagram of Virginia [map]. 1:750,000*, Richmond, VA: Virginia Department of Mines Minerals and Energy, 1991. [Online]. Available: <https://www.dmme.virginia.gov/commerce/ProductDetails.aspx?ProductID=1435>.
- [122] M. J. Bartholomew, *Geology of the Greenfield and Sherando quadrangles, Virginia*. Richmond, VA: Commonwealth of Virginia, Dept. of Conservation and Economic Development, Division of Mineral Resources, 1977. [Online]. Available: http://www.dmme.virginia.gov/commercedocs/PUB_4.pdf.
- [123] E. K. Rader and G. P. Wilkesm, *Publication 163: geologic map of the Virginia portion of the Staunton 30x60-minute quadrangle [map]. 1:100,000*, Richmond, VA:

- Virginia Department of Mines Minerals and Energy, 2001. [Online]. Available: <https://www.dmme.virginia.gov/commerce/ProductDetails.aspx?productID=1443>.
- [124] R. M. Yager, C. I. Voss, and S. Southworth, "Comparison of alternative representations of hydraulic-conductivity anisotropy in folded fractured-sedimentary rock: modeling groundwater flow in the Shenandoah Valley (USA)," *Hydrogeology Journal*, vol. 17, no. 5, pp. 1111–1131, Jul. 2009. DOI: [10.1007/s10040-008-0431-x](https://doi.org/10.1007/s10040-008-0431-x).
- [125] R. C. Orndorff and K. E. Goggin, *Sinkholes and karst-related features of the Shenandoah Valley in the Winchester 30x60 quadrangle, Virginia and West Virginia [map]. 1:100,000. miscellaneous field studies MF-2262*, North Latitude: 3930'0"N (39.5000), South Latitude: 390'0"N (39.0000), East Longitude: 780'0"W (-78.0000), West Longitude: 7830'0"W (-78.5000), Washington, DC: U.S. Geological Survey, 1994. [Online]. Available: <http://pubs.er.usgs.gov/publication/mf2262>.
- [126] A. Vaccari and S. T. Acton, "Spatiotemporal gaussian feature detection in sparsely sampled data with application to InSAR," in *Algorithms for Synthetic Aperture Radar Imagery XX*, vol. 8746, Jun. 2013, 87460U–87460U–9. DOI: [10.1117/12.2020669](https://doi.org/10.1117/12.2020669).
- [127] I. Ben-Gal, "Outlier detection," in *Data Mining and Knowledge Discovery Handbook*, O. Maimon and L. Rokach, Eds. Springer, 2005, ch. 7, pp. 131–146, ISBN: 9780387244358. DOI: [10.1007/0-387-25465-X_7](https://doi.org/10.1007/0-387-25465-X_7).
- [128] P. C. Mahalanobis, "On the generalised distance in statistics," in *Proceedings National Institute of Science, India*, vol. 2, Calcutta, India: National Institute of Science of India Society of Exploration Geophysicists, Apr. 1936, pp. 49–55.

- [129] A. Kraskov, H. Stögbauer, and P. Grassberger, “Estimating mutual information,” *Physical Review E*, vol. 69, no. 6, p. 066 138, 2004. DOI: [10.1103/PhysRevE.69.066138](https://doi.org/10.1103/PhysRevE.69.066138).
- [130] W. Li, “Mutual information functions versus correlation functions,” English, *Journal of Statistical Physics*, vol. 60, no. 5-6, pp. 823–837, 1990, ISSN: 0022-4715. DOI: [10.1007/BF01025996](https://doi.org/10.1007/BF01025996).
- [131] A. F. McDaid, D. Greene, and N. Hurley, “Normalized mutual information to evaluate overlapping community finding algorithms,” *ArXiv e-prints*, 2011. [Online]. Available: <http://arxiv.org/abs/1110.2515>.
- [132] N. X. Vinh, J. Epps, and J. Bailey, “Information theoretic measures for clusterings comparison: variants, properties, normalization and correction for chance,” *The Journal of Machine Learning Research*, vol. 11, pp. 2837–2854, Oct. 2010. [Online]. Available: <http://jmlr.org/papers/v11/vinh10a.html>.
- [133] Z. I. Botev, J. F. Grotowski, and D. P. Kroese, “Kernel density estimation via diffusion,” *The Annals of Statistics*, vol. 38, no. 5, pp. 2916–2957, Oct. 2010. DOI: [10.1214/10-AOS799](https://doi.org/10.1214/10-AOS799).
- [134] J. J. Koenderink, “The structure of images,” *Biological Cybernetics*, vol. 50, no. 5, pp. 363–370, 1984, ISSN: 0340-1200. DOI: [10.1007/BF00336961](https://doi.org/10.1007/BF00336961).
- [135] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, “Multiscale vessel enhancement filtering,” in *Medical Image Computing and Computer-Assisted Intervention MICCAI98*, W. M. Wells, A. Colchester, and S. Delp, Eds., ser. Lecture Notes in Computer Science. Heidelberg: Springer Berlin Heidelberg, 1998, vol. 1496, pp. 130–137, ISBN: 9783540651369. DOI: [10.1007/BFb0056195](https://doi.org/10.1007/BFb0056195).

- [136] B. Mohar, “The laplacian spectrum of graphs,” in *Graph Theory, Combinatorics, and Applications*, Y. Alavi, G. Chartrand, O. R. Oellermann, and A. Schwenk, Eds., Wiley, 1991, pp. 871–898.
- [137] —, “Some applications of laplace eigenvalues of graphs,” in *Graph Symmetry: Algebraic Methods and Applications*, G. Hahn and G. Sabidussi, Eds., ser. NATO ASI Series C, vol. 497, Kluwer, 1997, pp. 225–275.
- [138] F. R. Chung, *Spectral graph theory*, ser. CBMS Regional Conference Series in Mathematics. Providence, RI: American Mathematical Society, 1997, vol. 92, p. 212, ISBN: 9780821803158. [Online]. Available: <http://www.math.ucsd.edu/~fan/research/revised.html>.
- [139] A. Y. M. I. J. Ng and Y. Weiss., “On spectral clustering: analysis and an algorithm,” in *Proceedings of Advances in Neural Information Processing Systems*, 2001, pp. 849–856. [Online]. Available: <https://papers.nips.cc/paper/2092-on-spectral-clustering-analysis-and-an-algorithm.pdf>.
- [140] U. von Luxburg. (Jul. 9, 2007). Lectures on clustering, University of Hamburg, [Online]. Available: http://videlectures.net/bootcamp07_luxburg_clu/.
- [141] B. Miller, N. Bliss, and P. Wolfe, “Toward signal processing theory for graphs and non-euclidean data,” in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, Mar. 2010, pp. 5414–5417. DOI: [10.1109/ICASSP.2010.5494930](https://doi.org/10.1109/ICASSP.2010.5494930).

- [142] B. Miller, M. Beard, and N. Bliss, “Matched filtering for subgraph detection in dynamic networks,” in *Statistical Signal Processing Workshop (SSP), 2011 IEEE*, Jun. 2011, pp. 509–512. DOI: [10.1109/SSP.2011.5967745](https://doi.org/10.1109/SSP.2011.5967745).
- [143] R. A. Rossi, B. Gallagher, J. Neville, and K. Henderson, “Modeling dynamic behavior in large evolving graphs,” in *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining*, ser. WSDM '13, Rome, Italy: ACM, 2013, pp. 667–676, ISBN: 978-1-4503-1869-3. DOI: [10.1145/2433396.2433479](https://doi.org/10.1145/2433396.2433479).
- [144] D. Shuman, S. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, “The emerging field of signal processing on graphs: extending high-dimensional data analysis to networks and other irregular domains,” *Signal Processing Magazine, IEEE*, vol. 30, no. 3, pp. 83–98, May 2013, ISSN: 1053-5888. DOI: [10.1109/MSP.2012.2235192](https://doi.org/10.1109/MSP.2012.2235192).
- [145] A. Sandryhaila and J. Moura, “Discrete signal processing on graphs,” *Signal Processing, IEEE Transactions on*, vol. 61, no. 7, pp. 1644–1656, Apr. 2013, ISSN: 1053-587X. DOI: [10.1109/TSP.2013.2238935](https://doi.org/10.1109/TSP.2013.2238935).
- [146] D. Shuman, B. Ricaud, and P. Vandergheynst, “A windowed graph fourier transform,” in *Statistical Signal Processing Workshop (SSP), 2012 IEEE*, 2012, pp. 133–136. DOI: [10.1109/SSP.2012.6319640](https://doi.org/10.1109/SSP.2012.6319640).
- [147] G. Chen and M. Maggioni, “Multiscale geometric wavelets for the analysis of point clouds,” in *Information Sciences and Systems (CISS), 2010 44th Annual Conference on*, Mar. 2010, pp. 1–6. DOI: [10.1109/CISS.2010.5464843](https://doi.org/10.1109/CISS.2010.5464843).

- [148] D. K. Hammond, P. Vandergheynst, and R. Gribonval, “Wavelets on graphs via spectral graph theory,” *Applied and Computational Harmonic Analysis*, vol. 30, no. 2, pp. 129–150, 2011, ISSN: 1063-5203. DOI: [10.1016/j.acha.2010.04.005](https://doi.org/10.1016/j.acha.2010.04.005).
- [149] S. Narang and A. Ortega, “Perfect reconstruction two-channel wavelet filter banks for graph structured data,” *Signal Processing, IEEE Transactions on*, vol. 60, no. 6, pp. 2786–2799, Jun. 2012, ISSN: 1053-587X. DOI: [10.1109/TSP.2012.2188718](https://doi.org/10.1109/TSP.2012.2188718).
- [150] A. Agaskar and Y. M. Lu, “An uncertainty principle for functions defined on graphs,” vol. 8138, 2011, 81380T–81380T–11. DOI: [10.1117/12.894359](https://doi.org/10.1117/12.894359).
- [151] ———, “Uncertainty principles for signals defined on graphs: bounds and characterizations,” in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, Mar. 2012, pp. 3493–3496. DOI: [10.1109/ICASSP.2012.6288669](https://doi.org/10.1109/ICASSP.2012.6288669).
- [152] T. Lindeberg, “Feature detection with automatic scale selection,” English, *International Journal of Computer Vision*, vol. 30, no. 2, pp. 79–116, 1998, ISSN: 0920-5691. DOI: [10.1023/A:1008045108935](https://doi.org/10.1023/A:1008045108935).
- [153] D. Marr and E. Hildreth, “Theory of edge detection,” English, 1167, vol. 207, The Royal Society, 1980, pp. 187–217. [Online]. Available: <http://www.jstor.org/stable/35407>.
- [154] G. H. Golub and H. A. van der Vorst, “Eigenvalue computation in the 20th century,” *Journal of Computational and Applied Mathematics*, vol. 123, no. 12, pp. 35–65, 2000, Numerical Analysis 2000. Vol. III: Linear Algebra, ISSN: 0377-0427. DOI: [10.1016/S0377-0427\(00\)00413-1](https://doi.org/10.1016/S0377-0427(00)00413-1).
- [155] MATLAB, “Version 8.0 (r2012b),” Natick, Massachusetts, 2012.

- [156] E. P. W. Attema, “The active microwave instrument on-board the ers-1 satellite,” *Proceedings of the IEEE*, vol. 79, no. 6, pp. 791–799, 1991, ISSN: 0018-9219. DOI: [10.1109/5.90158](https://doi.org/10.1109/5.90158).
- [157] ASI, “COSMO-SkyMed System Description & User Guide,” Agenzia Spaziale Italiana, Rome, Italy, Tech. Rep. ASI-CSM-ENG-RS-093-A, May 2007. [Online]. Available: http://www.e-geos.it/products/pdf/csk-user_guide.pdf.
- [158] ESA. (2014). Copernicus - the european earth observation programme, European Space Agency, [Online]. Available: <http://www.copernicus.eu>.
- [159] CSA. (Jan. 15, 2013). Radarsat constellation, Canadian Space Agency, [Online]. Available: <http://www.asc-csa.gc.ca/eng/satellites/radarsat/>.
- [160] JAXA. (Jan. 10, 2014). Alos-2 basic observation scenario, [Online]. Available: http://www.eoportal.org/directory/res_p1_Earthobservation.html (visited on 01/10/2014).
- [161] ASI/Telespazio. (2014). Cosmo-skymed faqs, [Online]. Available: http://www.e-geos.it/products/pdf/Cosmo_FAQ.pdf.