



Technische Universität München
Fakultät für Bauingenieur- und Vermessungswesen
Lehrstuhl für Methodik der Fernerkundung

Spectral Estimation for Synthetic Aperture Radar Tomography

Author: Xiaoxiang Zhu

Master Thesis

Earth Oriented Space Science and Technology – *ESPACE*

Supervisors:

Prof. Dr.-Ing. habil. Richard Bamler, TUM

Dipl.-Ing. Nico Adam, DLR

19 September, 2008

DECLARATION

This thesis is a presentation of my original research work. Wherever contributions of others are involved, every effort is made to indicate this clearly, with due reference to the literature, and acknowledgement of collaborative research and discussions. The work was done under the guidance of Professor Richard Bamler, at the Institute of Photogrammetry and Cartography, Technische Universität München.

Munich, September 19, 2008

Zhu Xiaoxiang

Abstract

Synthetic Aperture Radar (SAR) Tomography, a new and advanced technique in the field of SAR processing, is aimed at determining the 3-D reflectivity function from measured multi-pass SAR data. It is essentially a spectrum estimation problem as for a specific resolution cell the complex valued SAR measurements of a SAR image stack are actually the irregularly sampled Fourier transform of the reflectivity function in the elevation direction. The successful launch of the German high resolution SAR mission TerraSAR-X provides a new possibility to investigate this topic with high quality spaceborne data.

Within the framework of this master thesis, the spectrum estimation problem is formulated from a mathematical point of view. Different spectrum estimation strategies such as the Singular Value Decomposition (SVD) and Nonlinear Least Squares estimation (NLS) are evaluated and compared using both simulated data and TerraSAR-X data from the test site Las Vegas with special consideration of the difficulties caused by sparse and irregularly spaced sampling. The problem of ill-conditioning when using the Singular Value Decomposition is investigated and regularization tools (such as singular value truncation and Wiener filtering) are utilized to overcome this problem. For the sake of validation, the spectrum estimation results with TerraSAR-X data are compared to the probable ground truth.

Penalized model selection criteria such as the Bayesian Information Criterion (BIC), Akaike information criterion (AIC) and Minimum Description Length criterion (MDL) are implemented on the spectral estimates to determine the number of scatterers inside one resolution cell - which is necessary a prior knowledge for precise PSI displacement estimation. The probability of correctly detecting the number of scatterers and the accuracy of the corresponding elevation estimates are evaluated from simulated data. Finally, the model selection results with PS points of TerraSAR-X data are visualized in Google-Earth and the nature of PS pixel with multi-scatterers are discussed.

Key words: Synthetic Aperture Radar (SAR), tomography, Spectrum estimation, Singular Value Decomposition, Nonlinear Least Squares estimation, Wiener filter, multi-scatterers, Model selection, penalized likelihood criterion, TerraSAR-X

Table of Contents

Abstract.....	I
Table of Contents.....	II
1 Introduction.....	1
1.1 State of the Art.....	1
1.2 Motivation and Objectives	5
1.3 Introduction to TerraSAR-X Mission.....	6
1.4 Outline of the Thesis	8
1.5 Summary	8
2 Spectrum Estimation Strategies.....	9
2.1 Introduction to SAR Tomography	9
2.2 Interpolation Kernel and Corresponding Frequency Response.....	15
2.3 SVD Strategy.....	18
2.3.1 Mathematical Description	18
2.3.2 SVD for Spectrum Estimation.....	19
2.4 Nonlinear Least Squares Estimation	21
2.5 Summary.....	24
3 Spectrum Estimation Results with Simulated and TerraSAR-X Data	25
3.1 Data Simulation.....	25
3.2 Spectrum Estimation Results from Pure Simulated Data.....	27
3.2.1 Baseline Distribution	27
3.2.2 Number of Samples and Interference Problem	31
3.2.3 Performance of Grant Matrix	33
3.3 Spectrum Estimation Results for Semi-Simulated Data	35
3.3.1 Spectrum Estimation Results using the SVD	35
3.3.2 Wiener Filter.....	40
3.3.3 Spectrum Estimation Results using NLS	45
3.3.4 Comparison and Discussion	46
3.4 Spectrum Estimation Results with TerraSAR-X Data.....	48
3.5 Summary	50
4 Model Selection.....	51
4.1 Introduction to DLR PSI-GENESIS Processor	51
4.2 Model Selection Methods.....	52
4.2.1 Introduction to Model Selection.....	52
4.2.2 Model Selection Methods.....	53
4.3 Model Selection with SAR Tomography	57
4.4 Result with TerraSAR-X Data.....	65
4.5 Summary	68
5 Concluding Remarks	70
5.1 Discussion and Conclusion.....	70
5.2 Outlook	71

Terminology and Abbreviation	72
References	73
Acknowledgements	75

1 Introduction

Synthetic Aperture Radar (SAR) has played an important role in remote sensing since the 1980s. It has been demonstrated that it is able to reliably map the earth's surface and acquire information about its physical properties such as topography, morphology, roughness and the dielectric characteristics of the backscattering layer ([Bamler and Harl, 1998](#)). As an active sensor, SAR functions independently of solar illumination and is capable of penetrating clouds and (partially) vegetation canopy, soil and snow. However, it also has typical foreshortening, layover and shadowing problems.

In Earth observation, unique capabilities are associated with the use of remote sensing via synthetic aperture radars (SARs) and, particularly, with the extensions of SAR to interferometric modes (InSAR) and more generally to the joint use of coherent multiple acquisitions (PSI and SAR tomography etc.) ([Fornaro, 2005](#)).

The standard acquisition of a single SAR image is a two dimensional image of scene reflectivity in azimuth and range. InSAR techniques, which combine two or more complex-valued SAR images to determine geometric information about the imaged objects (compared to using a single image) by exploiting phase differences, have different applications according to the baseline type. With incidence angle difference (across-track InSAR), it is possible to get topography information and DEMs. Therefore, it is possible to reach the third dimension which is perpendicular to the azimuth and range plane. However, with one or two acquisitions, the reflectivity function along the third dimension is undetermined and multiple-scatterers within one resolution cell cannot be separated.

SAR tomography is a young technology which is based on multi-pass acquisitions to estimate the 3-D radar reflectivity function. This thesis investigates the spectrum estimation problem for SAR tomography with both simulated and real data, especially in case of small irregularly sampled datasets. Both deterministic and statistical methods are implemented and investigated. Additionally, the application of model selection to SAR tomography will be addressed.

1.1 State of the Art

InSAR and D-InSAR exploit the phase differences of coherently acquired SAR images in order to measure land surface deformations from space ([Bamler et al., 1998](#); [Massonnet et al., 1993](#)). Since the first spectacular results were published e.g. co-seismic deformation ([Gabriel et al., 1989](#)), the method has been well established in the geophysical community and complements GPS point measurements with two-dimensional displacement maps. These maps provide centimeter and even millimeter accuracy over areas of typically 50 km by 50 km and can be used to monitor deformations of the Earth's surface.

The accuracy of InSAR and D-InSAR is limited by temporal decorrelation of the surface and by electromagnetic path delay variations in the troposphere. The latter distortions can be reduced by temporal averaging of multiple interferograms which in turn reduces the temporal resolution (Hanssen, 2001). The Persistent Scatterer Interferometry (PSI) method uses stacks of 10-100 acquisitions of the same area, selection of long term coherent points – so-called Persistent Scatterers (PS) – and motion models in order to reduce the estimation error to well below 1 millimeter/year (Ferretti et al., 2001; Kampes, 2006; Adam et al., 2007). Persistent scatterers are bright points in the image that represent strong scatterers in the object space, e.g. metallic structures or retro-reflecting corners at a building. They do not decorrelate over long time spans. PSI is currently one of the most powerful space based geodetic measurement methods.

The PSI technique, however, can only be applied when sufficient points with long-term stable backscattering characteristics are found, i.e. when the spatial density of these PSs is large enough to represent the ground deformation pattern. An example of PSI analysis over an urban area is shown in Figure 1, where a deformation map of Las Vegas, USA is presented. Note the mm-scale of the deformation process.

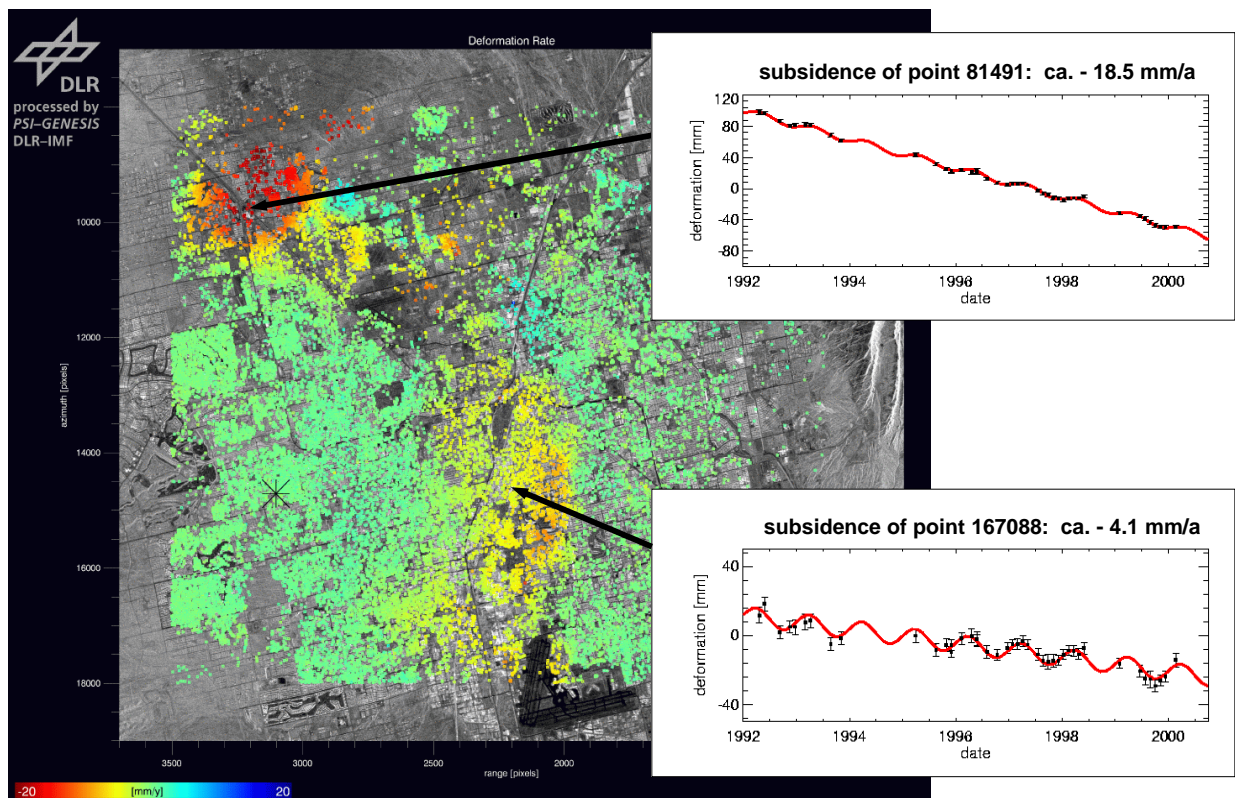


Figure 1.1: Example of a deformation map created from PSI for Las Vegas showing long term trends and periodic seasonal variations. Each of the colored points represents a time history over 9 years, obtained from 64 ERS data sets. The color reflects the linear component of the subsidence. The three ground water wells can be recognized, as well as two slight uplift areas (Adam et al., 2008).

The mathematical algorithms for PSI have been optimized for medium resolution SAR satellite systems (ERS, ENVISAT, 25 m resolution). With this kind of data the PS density is not high enough to achieve subsidence or deformation measurements for every building. Typically 100-450 PS/km² can be detected, i.e. on average a single PS in a square of 50-100 m side length. Only gross subsidence patterns like those in Figure 1 can be retrieved. Individual buildings can only be investigated opportunistically. Also, the limited resolution prevents accurate 3D localization of the scatterers. Often it cannot be distinguished whether a scatterer is part of a building and, hence, represents the subsidence of this building, or whether it is a double-bounce effect off the street-wall-interface and shows the subsidence of the pavement.

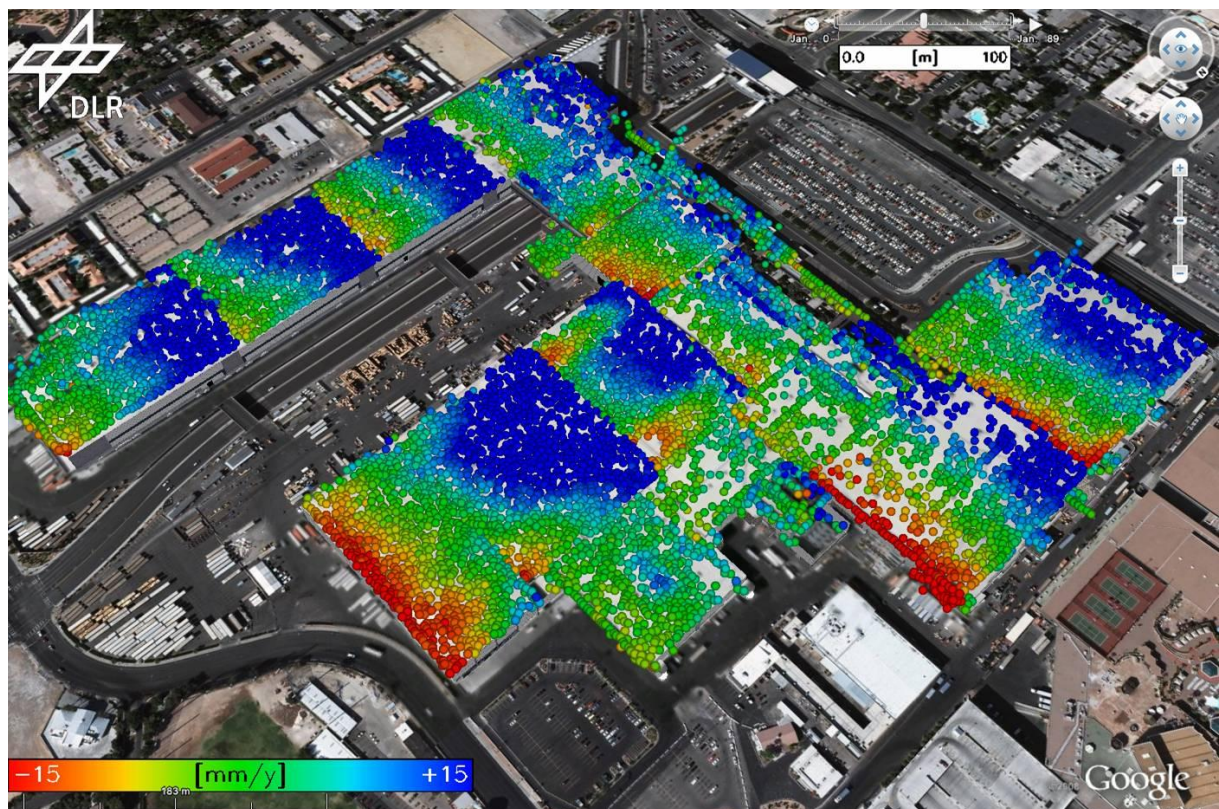


Figure 1.2: Example of thermal stress monitoring of a building in Las Vegas using the Persistent Scatterer technique with TerraSAR-X high resolution spotlight data (Adam et al., 2008). Only the linear motion component is retrieved and extrapolated to mm/y.

Recently, a new generation of SAR systems has been deployed in space, among them the first German SAR satellite TerraSAR-X with a spatial resolution of 0.6×1.1 m. This resolution has been made possible by virtue of the new spotlight imaging mode and the high bandwidth. Today, only the DLR-TUM group is able to exploit this special data for PSI measurement. The PSI density can be shown to rise by a factor of about 50-100 compared to the previously available data (Adam et al., 2008). Several tens of PS can now be identified on a single building. Figure 2 provides a preliminary building's deformation estimation from TerraSAR-X data applying the straight forward PSI processing technique. This opens the potential of retrieving for the first time deformation and structural stress of individual

buildings from space on a regular basis.

For a 4D (space-time) analysis each scatterer and its subsidence history must be accurately located in 3D space. In PSI this is done by exploiting the natural drift of orbits which results in a slightly different observation angle for each acquisition in the stack. A synthetic aperture in the elevation direction can be built up, comparable to the synthetic aperture in the flight direction of the SAR imaging principle. The fundamental difference, however, is that in the flight direction the data is sufficiently and regularly sampled according to the Nyquist theorem, while in the elevation direction sampling reflects random orbit locations and is sparse and irregular. A further complication is that in the multiple-scattering case the different scatterers may exhibit a different deformation history.

Imaging in the third dimension (elevation) is also referred to as “tomography”. While in the original references on SAR tomography ([Reigber et al., 2000](#)) an estimate of the reflectivity in elevation is derived, PSI needs to retrieve the coordinates of single points. Both methods are tomographic and are equivalent to spectral estimation, the first one being non-parametric and the latter parametric. For a full exploitation of 1 m resolution data in urban environments, the omnipresent ambiguities due to multiple-scattering must be resolved. Hence, multiple scatterers must be considered in the tomographic reconstruction and model selection must be applied in order to estimate the number of relevant scatterers ([Adam et al., 2005](#)).

Tomography, introduced to SAR in the early nineties, is a way of overcoming the limitations of standard two-dimensional (2-D) imaging by achieving, similar to Computed Axial Tomography (CAT), focused 3-D images. However, with respect to classical CAT, SAR tomography has a few additional difficulties. First of all, acquisitions are generally highly unevenly distributed in baseline: a classical Fourier-based inversion may decrease the performance with respect to a regularized inversion, ([G. Fornaro, 2005](#)). Second, 3-D data cannot be collected simultaneously, at least with existing satellites, but rather synthesized via repeated passes, well separated in time. Third, for spaceborne SAR the number of acquisitions depends directly on the number of passes, which means that for young SAR satellites such as TerraSAR-X the number of acquisitions can be very limited.

SAR tomography is addressed in very few scientific publications. The first experiments were carried out in the laboratory ([P. Pasquali, 1995](#)), under ideal experimental conditions, or by using airborne systems ([A. Reigber and A. Moreira, 2000](#)). The application of 3-D SAR tomography to spaceborne systems is limited ([G. Fornaro et al., 2005](#)), ([J. Homer, 2002](#)), ([Z. She, 2002](#)) and not yet well assessed until now. Notwithstanding, developments of SAR tomography for spaceborne systems would join the potentials of advanced imaging techniques to the synoptic view capabilities ([F. Gini et al., 2002](#)), and is fundamental to four-dimensional (4-D) (space time) SAR imaging, i.e. to techniques that not only separate point scatterers interfering in the same azimuth-range resolution cell, but also estimate their relative deformations ([G. Fornaro et al., 2006](#)).

1.2 Motivation and Objectives

Satellite remote sensing is becoming more and more important. It is able to recover contact-free information about objects. Synthetic Aperture Radar (SAR) along with its interferometric capabilities has many proven advantages compared to other techniques in the remote sensing sensor suite: SAR is independent of weather and sun illumination and due to phase measurements, it is even capable of measuring precise 3D localization and millimeter motions on the earth from space and reveals a previously unknown potential for earth observation.

SAR tomography (TomoSAR), is a consequent further development of SAR and InSAR: TomoSAR can retrieve the reflectivity distribution in the elevation direction (perpendicular to the range and azimuth plane), therefore, TomoSAR can better reconstruct 3D structures, distinguish multiple scatters within one resolution cell and allows to monitor with highest precision to 4D dynamic world extending the PSI technique.

In recent years, with ideal experimental conditions or by using airborne systems, some experiments for TomoSAR were carried out. However, for the spaceborne case, the application of 3-D SAR tomography to spaceborne systems is limited. Even though the importance of spaceborne TomoSAR has been recognized and the basic principle is well described and understood from theoretical point of view, there are not many examples from real data. The performance is limited either by satellite data acquisition or by the complexity of the problem itself.

Currently, there are several new SAR satellites, e.g. TerraSAR-X, COSMO-SkyMed, Radarsat-2, TanDEM-X, which are already launched or will be launched very soon. They provide new high quality data acquisitions for TomoSAR unavailable so far. The TerraSAR-X satellite, launched on June 15th 2007, provides high resolution of up to 1m in Spotlight mode. TomoSAR can indeed profit a great deal from such high resolution as the density of PS increase dramatically and the signal to clutter ratio has been improved significantly as well. With a short repeat cycle of 11 days, the stack can be built up rapidly. Needless to say, the launch of TerraSAR-X brings new blood to the development of spaceborne TomoSAR.

To summarize, practical demonstration on spaceborne TomoSAR is very important and challenging.

The research work in this thesis has been carried out with the following objectives.

- Analysis of existing tomographic algorithms:
An estimation theoretic framework is established to evaluate the applicability of existing tomographic algorithms in multiple scattering cases.
- Simulation of typical elevation apertures and parameter estimation:

Based on typical TerraSAR-X orbit histories, representative aperture sampling schemes are generated. Parametric and non-parametric tomographic reconstructions are simulated and evaluated.

- Implementation and assessment of model selection methods:

Bayesian, the minimum description length and the Akaike information criterion are implemented and tested on the simulated data in order to compare the performance to detect one, two or three dominant scatterers inside of a resolution cell.

- Application of the developed algorithms on real data:

Data of the Las Vegas testsite are preprocessed to form a PSI stack. The tomographic algorithms developed in the previous steps are applied on the real data. The detected single and multiple scatterers are compared to ground truth and validated.

1.3 Introduction to TerraSAR-X Mission

The TerraSAR-X satellite was launched on June 15th, 2007, from Baikonur in Kazakhstan. With its active antenna, the spacecraft acquires high-quality X-band radar images of the entire planet whilst circling the Earth in a polar orbit at 514 km altitude. TerraSAR-X is designed to carry out its task for five years, independent of weather conditions and sun-illumination, and reliably provides radar images with a resolution of up to 0.6×1.1m.

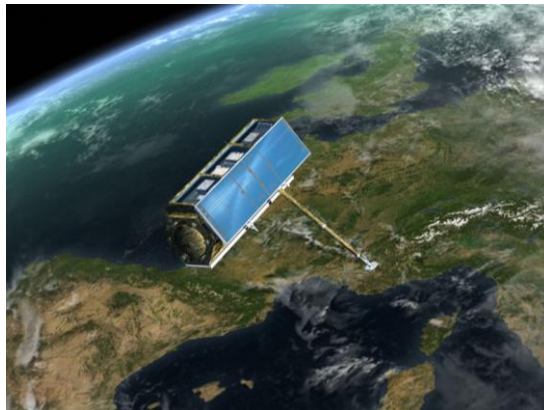


Figure 1.3: Artist's view: TerraSAR-X in space

Technical features include:

- Active phased array X-band SAR
- Single, dual and quad polarization
- Side-looking acquisition geometry
- Sun-synchronous dawn-dusk repeat orbit
- Repetition rate: 11 days; due to swath overlay, a 2.5 day revisit time can be achieved
- Orbit altitude range from 512 km to 530 km
- Three operational imaging modes:
 - SpotLight: up to 1m resolution, 10 km (width) x 5 km (length)

- StripMap: up to 3m resolution, 30 km (width) x 50 km (length)
- ScanSAR: up to 16 m resolution, 100 km (width) x 150 km (length)
- StripMap and ScanSAR: acquisition length extendable to 1,650 km
- 300 MHz high bandwidth providing a range resolution of 0.6m.

TerraSAR-X is the first German radar satellite to be implemented within a Public-Private Partnership (PPP) between the German Aerospace Center (DLR) and Europe's leading satellite specialist Astrium. DLR and Astrium share the costs of the development, construction and deployment of the satellite.

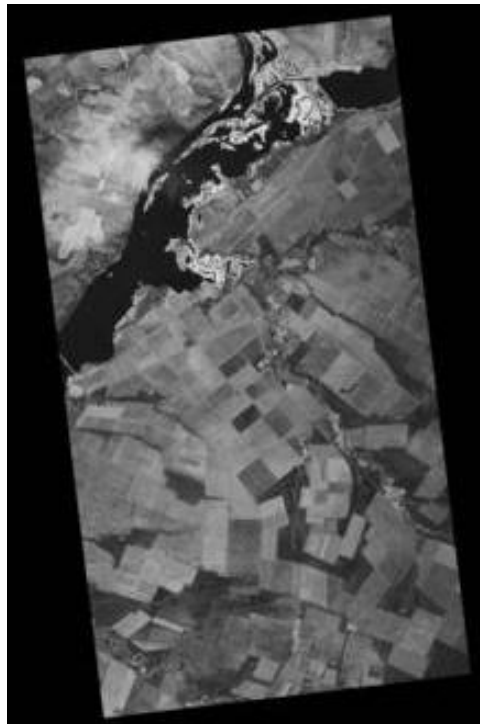


Figure 1.4: First scene acquired by TerraSAR-X only four days after its launch: Landscape near Volgograd, Russia.

In the future, the PPP-funded TanDEM-X, a twin satellite to TerraSAR-X, will enhance the mission. The satellite constellation will enable the generation of high-quality Digital Elevation Models (DEMs) on a global scale.

1.4 Outline of the Thesis

After the introduction, Chapter 2 introduces the basic principle of SAR tomography. The framework of SAR tomography is presented showing that the technique results finally in a spectral estimation problem. Afterwards, several spectrum estimation strategies, both deterministic and statistical, (including the Singular Value Decomposition and Nonlinear Least Squares estimation strategies) are explained in detail.

Based on the theory explained in Chapter 2, the effect of different criteria such as a random distribution of baselines, number of samples, amplitude of the signal, mutual inference, signal to noise ratio and performance of different spectrum estimation strategies are analyzed with simulated data in different conditions. The practical applicability of the developed algorithms is demonstrated using high resolution Spotlight data from TerraSAR-X of the Las Vegas test site.

SAR tomography can give information about the scatterer distribution within one resolution cell. It is capable of determining number of scatterers and the corresponding reflectivity. As an application of the development, Chapter 4 gives a short overview of the DLR PSI-GENESIS processor and describes how SAR tomography helps to improve the PSI processing by estimation of the scatterer configuration. Different model selection methods are discussed and the performances are investigated with simulated and real data. The model selection results are visualized in Google-earth.

Chapter 5 concludes this thesis, and attempts to provide an outlook of future research directions.

1.5 Summary

- 1) Spaceborne SAR tomography is a new and challenging technique which allows improving the PSI estimation. It can be considered as a straight forward extension of the PSI processing.
- 2) The subject of this thesis is the prototyping and test of spectral estimation methods for SAR tomography. The practical applicability of the developed algorithms is demonstrated on TerraSAR-X data even in irregularly sampled and small dataset.

2 Spectrum Estimation Strategies

This chapter presents some fundamentals of SAR tomography, interpolation and the use of SVD and Nonlinear Least Squares estimation (NLS) for spectrum estimation. The main purpose of this chapter is to develop and provide mathematical framework for the following chapter.

SAR tomography and its corresponding mathematical description are discussed in Section 2.1. Afterwards, the interpolation strategy for spectrum estimation and its corresponding frequency response kernel is described in Section 2.2. The singular value decomposition strategy for spectrum estimation is introduced in Section 2.3. Nonlinear Least Squares estimation which is a deterministic estimator is explained in Section 2.4. Finally, Section 2.5 summarizes the chapter.

2.1 Introduction to SAR Tomography

With applying a classical focusing algorithm, 2-D full resolution SAR images are obtained as the single-look complex (SLC) images. The complex valued measurement of a specific resolution cell of the SLC image represents the integral of the reflected signal along the elevation direction.

In Figure 2.1, the building at the centre is the famous hotel: Wynn Las Vegas. The orange line refers to the line of sight and the yellow lines represent the elevation direction for the respective single resolution cells. Figure 2.2 shows the corresponding TerraSAR-X high resolution Spotlight SAR image of that building where the brightness refers to the intensity of reflected radar signal. The pixels highlighted with yellow dots are nothing more than the integration of the reflected radar signal along the yellow lines i.e. in the elevation direction. Therefore, if there are two scatterers within such a resolution cell e.g. a scatterer on the building and on the ground, we can not distinguish them with a single SLC image.

SAR tomography retrieves the distribution of scatterers in the elevation direction and the corresponding reflectivity. In order to understand SAR tomography, we have to understand that SAR tomography is actually a spectrum estimation problem. The theoretical aspects of SAR tomography are quite well understood and there exist several publications ([Reigber and Moreira 2000](#), [Fornaro and Serafino 2003](#), [Fornaro and Lombardini 2005](#), [Lombardini 2005](#), [Fornaro 2006](#)). In this section, the theory of SAR tomography will be explained mathematically.



Figure 2.1: The famous hotel in Las Vegas: Wynn Las Vegas visualized in Google-Earth



Figure 2.2; Corresponding SAR image of Wynn Las Vegas. The brightness refers to the intensity of reflected radar signal received by SAR sensor.

Spaceborne SAR tomography is applied to determine the 3-D reflectivity function from repeat pass acquisition. With a single pass SAR image, with coordinates in the azimuth-range plane, the two dimensional reflectivity properties can be retrieved. Multi-pass SAR images have the potential of providing information in the third dimension.

Let us define the 3-D reference sensor frame of a single pass as:

- x: azimuth direction
- r: range direction
- s: elevation direction

where the origin is the position of the satellite.

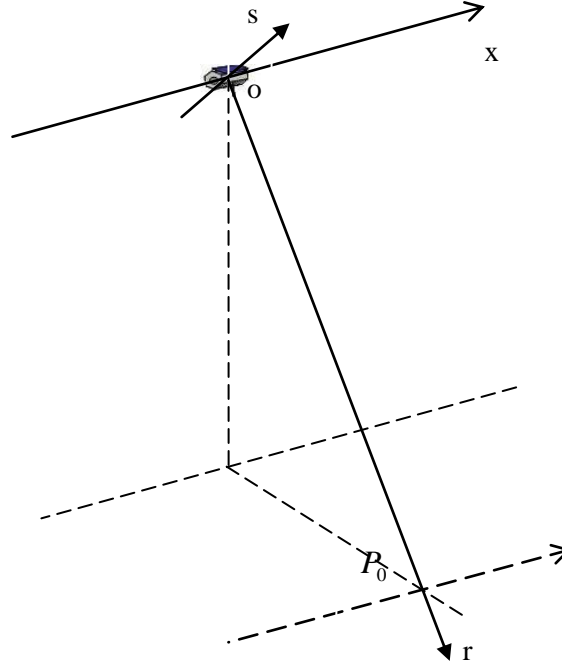


Figure 2.3: Sensor frame

Within this reference frame, $N+1$ SAR passes at coordinates (x, r) are made for each resolution cell. We are interested in the determining the s coordinates of each scatterer within one resolution cell and their corresponding reflectivities.

For the purpose of simplicity, we consider only one resolution cell and make the following assumptions.

- 1) All SAR images are co-registered.
- 2) The SAR sensor is observing ideal point scatterers.

With assumption 1, for a specific resolution cell, we shift the geometry from a 3D world to a 2D world. As depicted in Figure 2.4, let us set the pass labelled with n_M as the reference pass (master pass), then the reference frame can be specified with an origin which is the zero Doppler satellite position of pass n_M ; an r axis pointing to a reference point P_0 inside the resolution cell with zero elevation; and an s axis perpendicular to the r axis within the range-elevation plane. Note that the reference pass can be any pass. Therefore, we have $P_0 = (x_0, r_0, 0)$ with the target P lying at the position (x_0, r_0, s) . Under this definition, the azimuth coordinates of a specific target at zero Doppler position is $x = x_0$.

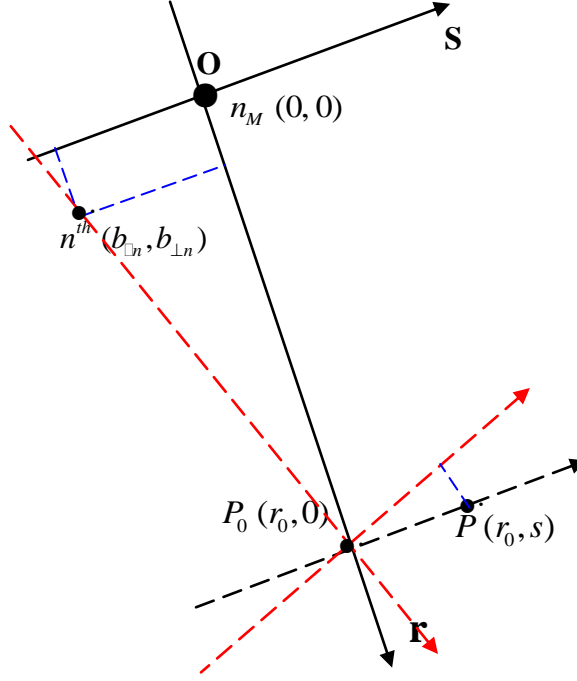


Figure 2.4: Two dimensional system frame

Assumption 2 is reasonable with DLR PSI-GENESIS which extracts phase and amplitude by Point Target Analysis (PTA). With this assumption, the post focusing 2-D point spread function (PSF) can be replaced by a 2-D Dirac function.

Using the 1st Born approximation, the obtained 2-D signal for the n^{th} antenna can be modelled as follows

$$\hat{\gamma}_n(x_0, r_0) = \iint dx dr f(x_0 - x, r_0 - r) \int ds \gamma(x, r, s) \cdot \exp[-j \frac{4\pi}{\lambda} R_n(r, s)], \quad n = 0, \dots, N \quad (2.1)$$

where x_0, r_0 are the azimuth and range coordinates associated with the focused data, λ is the operating wavelength, $\gamma(x, r, s)$ represents the 3D scattering model and $R_n(r, s)$ refers to the distance between the target and sensor. Within the 2-D sensor frame (RS plane) of the antenna n_M depicted in Figure 2.4.

$$R_n(r, s) = \sqrt{(r - b_{\parallel n})^2 + (s - b_{\perp n})^2} \approx |r - b_{\parallel n}| + \frac{(s - b_{\perp n})^2}{2|r - b_{\parallel n}|} \quad (2.2)$$

Here, $(b_{\parallel n}, b_{\perp n})$ is the position of the n^{th} antenna and (r, s) are the coordinates of the scatterer. $f(x_0 - x, r_0 - r)$ is the so called post focusing 2-D point spread function (PSF) which is given by

$$f(x_0 - x, r_0 - r) = \text{sinc}\left(\frac{x_0 - x}{\Delta x}\right) \text{sinc}\left(\frac{r_0 - r}{\Delta r}\right), \quad (2.3)$$

where Δx and Δr are the azimuth and range resolution, respectively.

Using assumption 2,

$$f(x_0 - x, r_0 - r) = \delta(x_0 - x) \delta(r_0 - r) , \quad (2-4)$$

$$\hat{\gamma}_n(x_0, r_0) = \int ds \gamma(x_0, r_0, s) \cdot \exp[-j \frac{4\pi}{\lambda} R_n(r_0, s)] \square h_n, \quad n = 0, \dots, N, \quad (2-5)$$

Then, $\gamma(x, r, s)$ and $R_n(r, s)$ only depend on the elevation since x_0 and r_0 are given by the focused position of the SAR images. By multiplying the received signal with a phase factor corresponding to the zero elevation reference point P_0 , we have.

$$g_n = h_n \exp[j \frac{4\pi}{\lambda} R_n(0)] = \int_{-a}^a \gamma(s) \exp[-j \frac{4\pi}{\lambda} (R_n(s) - R_n(0))] ds , \quad (2-6)$$

where,

$$\begin{aligned} R_n(s) - R_n(0) &= |r_0 - b_{\parallel n}| + \frac{(s - b_{\perp n})^2}{2|r_0 - b_{\parallel n}|} - (|r_0 - b_{\parallel n}| + \frac{(b_{\perp n})^2}{2|r_0 - b_{\parallel n}|}) \\ &= -\frac{b_{\perp n}}{|r_0 - b_{\parallel n}|} s + \frac{s^2}{2|r_0 - b_{\parallel n}|} , \end{aligned} \quad (2-7)$$

which is the range difference caused by a different sensor position.

$[-a, a]$ is the extent of the image scene in the elevation direction. The value of a mainly depends on the height of the building and the look angle.

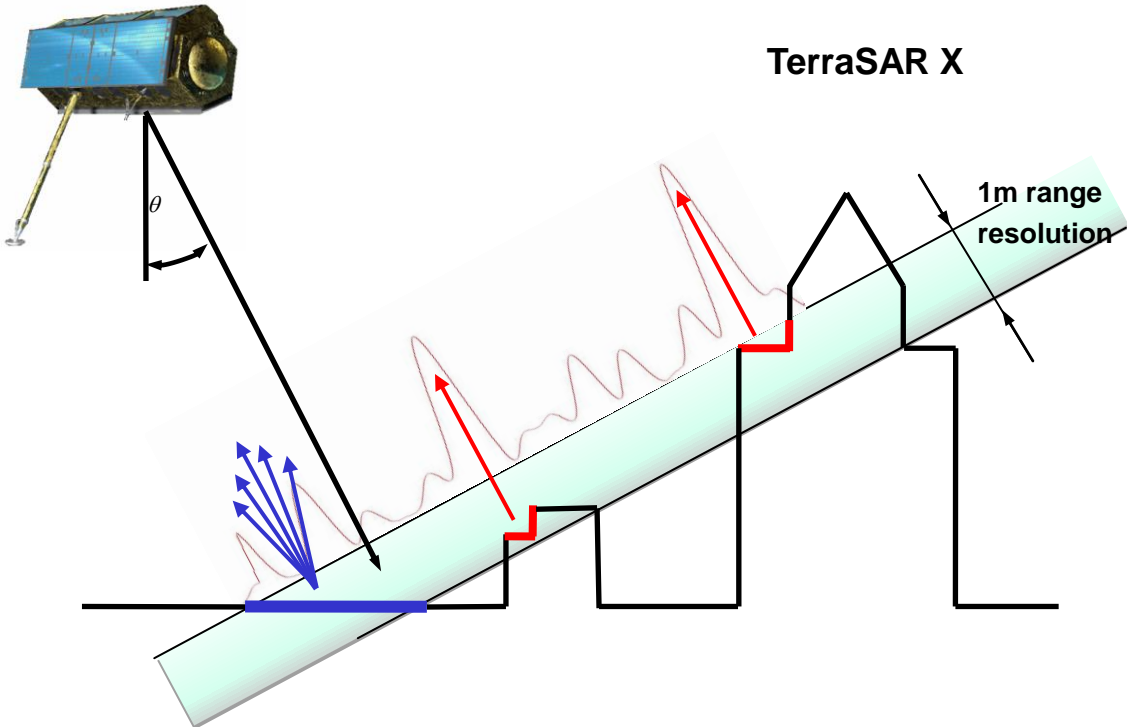


Figure 2.5: Two dominant scatterers inside the resolution cell

Inserting 2.7 into 2.6 gives:

$$g_n = \int_{-a}^a \gamma'(s) \exp[-j2\pi(-\frac{2}{\lambda} \frac{b_{\perp n}}{|r_0 - b_{\parallel n}|})s] ds, \quad (2-8)$$

where

$$\gamma'(s) = \gamma(s) \exp[-j \frac{4\pi}{\lambda} \frac{s^2}{2|r_0 - b_{\parallel n}|}] . \quad (2-9)$$

From which we can see that g_n is nothing else but the Fourier transform of $\gamma'(s)$ at position

$$f_n = -\frac{2b_{\perp n}}{\lambda|r_0 - b_{\parallel n}|} \approx -\frac{2b_{\perp n}}{\lambda r} .$$

From this, we can see that the effect due to the baseline in the range direction $b_{\parallel n}$ can be neglected. All passes can be assumed located along the elevation axis.

As the phase term of $\gamma'(s)$ only affects the phase of the final image and we are only interested in the amplitude which determines the power density, we ignore the phase term for simplicity:

$$g_n = \int_{-a}^a \gamma(s) \exp[-j2\pi f_n s] ds = FT[\gamma(s)]|_{f_n} = g(f)|_{f_n}, \quad n = 0, \dots, n_s, \quad (2-10)$$

which shows that the focused SAR image from the n^{th} pass for a specific resolution cell is nothing else but the Fourier Transfer of the reflectivity function in the elevation direction at the position f_n . Taking into account all $n_s + 1$ passes, we have a series which is irregularly sampled and depends on the baseline of each antenna relative to the master antenna frame (n_M). Therefore, in order to estimate the elevations of the scatterers, we have to estimate the peak in the frequency domain. Hence, SAR tomography is a spectrum estimation problem.

In order to understand this explicitly, we can also treat multi-pass data for a specific resolution as a signal in the elevation direction while a SAR image is a signal collection in the range and azimuth directions. Therefore, we also need an elevation direction focusing process and this is what SAR tomography does.

For uniform sampling at a sufficient rate, the discrete values of the reflectivity function along the elevation direction can be directly retrieved by a DFT. However, in our case the data is unevenly sampled and seldom passes over the area of interest. Fornaro et.al deal with this specific SAR tomography inverse problem for ERS data and obtained good results. However, compared to ERS, TerraSAR-X has far fewer acquisitions. Thus, we have to design an algorithm for the case of sparse irregularly sampled data.

Several possibilities are conceivable, both statistical and deterministic strategies. Interpolation, Singular Value decomposition and Nonlinear Least Squares Estimation are explained in detail in the following sections.

2.2 Interpolation Kernel and Corresponding Frequency Response

Interpolation is the process of determining the values of a function at positions lying between its samples. It achieves this by fitting a continuous function through the input samples.

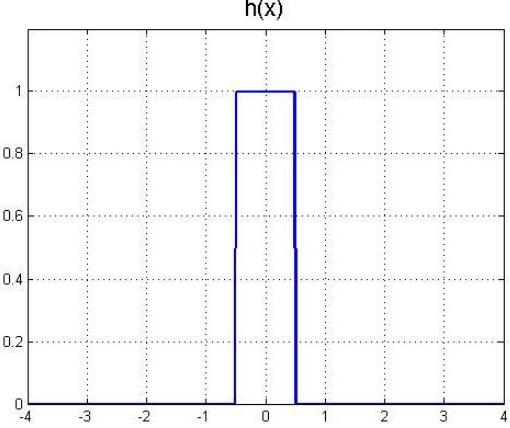
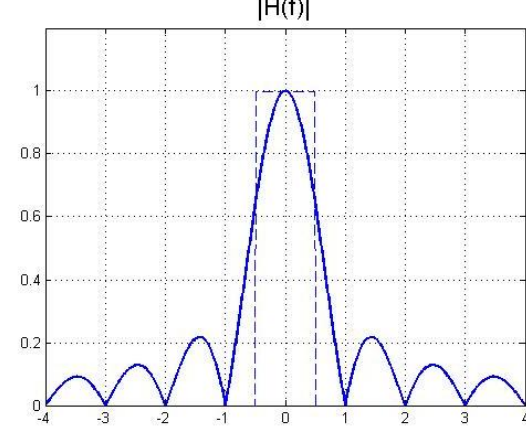
“Interpolation” has different meanings depending on the context. For continuous signals, it can be filling a gap. For discrete signals, it can be reconstruction of the original continuous signal which is sorted as polarimetric strategy; it can also be a change of the sampling pattern such as sampling rate conversion, resampling, geometric distortion, co-registration, shift in the sub-pixel domain, and so on. In our case, we wish to resample the series to a regular series and then estimate the spectrum using the resampled data.

The accuracy and computational cost of interpolation depends on the interpolation kernel. Here, the purpose of interpolation is to simulate an ideal low pass filter which is an infinite length sinc function in the time domain. Therefore, by comparing the frequency response with an ideal low pass filter, one can evaluate the performance of the interpolator. There are many different interpolation kernels such as nearest neighbourhood, linear interpolation, cubic convolution and B-splines. However, interpolation is only possible if a priori knowledge about the signal bandwidth is available. Here, it is assumed that the bandwidth of the low pass filter is equal to 1. For the purpose of performance evaluation, regular sampling is assumed in the following discussion.

1. Nearest Neighbour

The simplest interpolator from a computational standpoint is the nearest neighbor, where each interpolated point is assigned the value of the nearest sample point in the input image. The interpolation kernel and corresponding frequency response is represented in the following table. We can see that there is a large distortion of the signal spectrum and aliasing effects due to incomplete suppression of the spectral replicas.

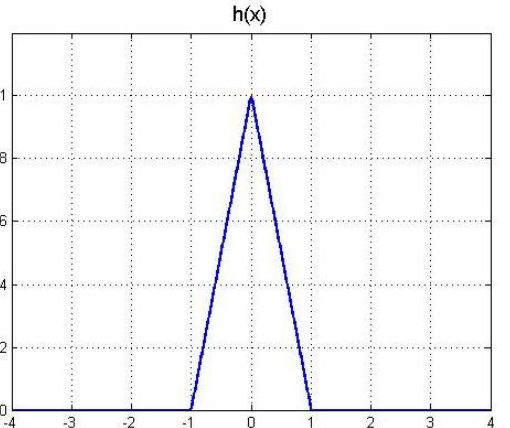
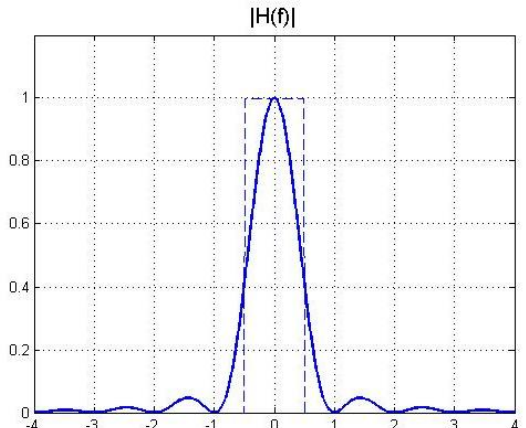
Table 2.1; The Nearest Neighbour interpolation kernel in the time and frequency domains

Time domain	Frequency domain
$h(x) = \begin{cases} 1 & 0 \leq x \leq 0.5 \\ 0 & 0.5 \leq x \end{cases}$	$H(f) = \text{sinc}(f)$
	

2. Linear Interpolation

Linear interpolation is a first degree method that joins two consecutive points in the input signal with a straight line. The interpolation kernel is a triangular function and its corresponding frequency response is the square of a sinc function. This kernel is also called the roof function or Bartlett window. The frequency response of the linear interpolation kernel is superior to that of the nearest neighbour interpolation function. The side lobes are less prominent. The performance is improved in the stop band. The pass band is moderately attenuated, resulting in smoothing.

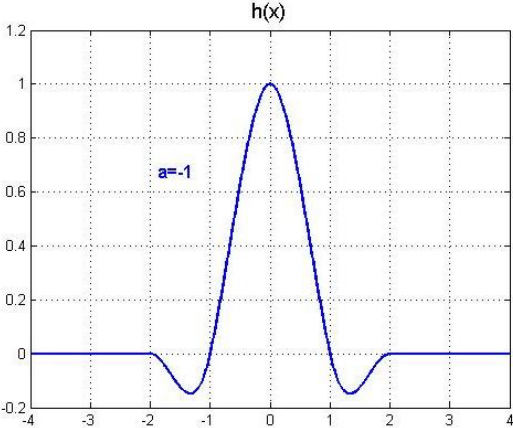
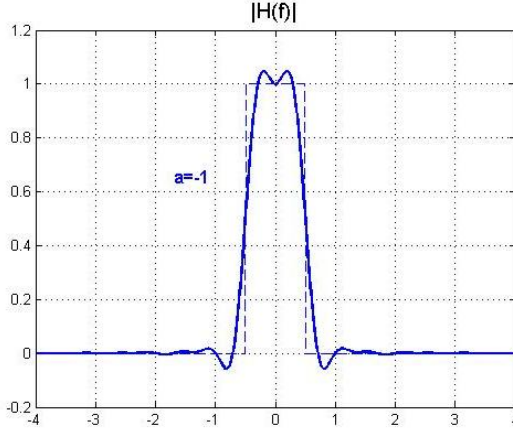
Table 2.2: The Linear Interpolation kernel in the time and frequency domains

Time domain	Frequency domain
$h(x) = \begin{cases} 1 - x & 0 \leq x \leq 1 \\ 0 & x \geq 1 \end{cases}$	$H(f) = \text{sinc}^2(f)$
	

3. Cubic Convolution

Cubic convolution is a third degree interpolation algorithm that well approximates the theoretically optimum sinc interpolation function. It is so called interpolation as “weighted truncated sinc” as it approximates the sinc interpolation function quite well. b is a parameter that can be controlled by the user. Different values of b suit different problems. For the case of $b = -1$, it approximates a sinc function with the same gradient at $|x| = 1$.

Table 2.3: The cubic convolution interpolation kernel in time and frequency domain

Time domain	Frequency domain
$h(x) = \begin{cases} (b+2) x ^3 - (b+3) x ^2 + 1 & 0 \leq x \leq 1 \\ b x ^3 - 5b x ^2 + 8b x - 4b & 1 \leq x \leq 2 \\ 0 & x \geq 2 \end{cases}$	$H(f) = \frac{3}{(\pi f)^2} (\text{sinc}^2(f) - \text{sinc}(2f)) + b \frac{2}{(\pi f)^2} (3 \text{sinc}^2(2f) - 2 \text{sinc}(2f) - \text{sinc}(4f))$
	

Based on the analysis above, we can see that in the case of regular sampling, the nearest neighbour is quite poor, linear interpolation gives a reasonable result at moderate cost, cubic convolution shows much better performance, but the cost is also relatively high.

However, everything discussed above is based on having enough samples. Therefore, interpolation in the data space can only be performed with a sufficient number of satellite passes. Otherwise, as we see from the frequency response, the pass band will be extended, leading to significant aliasing.

2.3 SVD Strategy

In linear algebra, the singular value decomposition (SVD) is an important factorization of a rectangular real or complex matrix, with several applications in signal processing and statistics. Applications which employ the SVD include computing the pseudo inverse, least squares fitting of data, matrix approximation, and determining the rank, range and null space of a matrix. SVD is a simple and valuable tool for analyzing image quality and the amount of independent information about the unknowns which can be reliably retrieved from observations in presence of noise. For spectrum estimation, it is generally possible to effectively overcome the associated problem due to nonuniform track distribution that may include significant noise propagation due to the ill-conditioned nature of the problem by Truncated Singular Value Decomposition (TSVD). In the following section, the mathematical description of the SVD and how to apply it to the spectrum estimation problem is presented.

2.3.1 Mathematical Description

Let $G \in \mathbb{R}^{m \times n}$ be a rectangular matrix with $m > n$. The SVD of G is a decomposition of the form (Hansen, 1992)

$$G = U \Sigma V^T = \sum_{i=1}^n u_i \sigma_i v_i^T, \quad (2-11)$$

where $U = (u_1, \dots, u_n)$ and $V = (v_1, \dots, v_n)$ are matrices with orthonormal columns,

$U^T U = V^T V = I_n$ and $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_n)$ has non-negative diagonal elements such that

$$\sigma_1 \geq \dots \geq \sigma_n \geq 0$$

The σ_i are the singular values of G while the vectors u_i and v_i are the left and right singular vectors of G respectively. The condition number of G is equal to the ratio σ_1 / σ_n .

From the relations $G^T G = V^T \Sigma^2 V$ and $G G^T = U^T \Sigma^2 U$ we see that the SVD of G is strongly linked to the singular value decompositions of the symmetric positive semi-definite matrices $G^T G$ and $G G^T$. This shows that the SVD is unique for a given matrix G -except for singular vectors associated with multiple singular values. In connection with discrete ill-posed problems, two characteristic features of the SVD of G are very often found.

- 1) The singular values σ_i decay gradually to zero with no particular gap in the spectrum.
An increase in the dimensions of G will increase the number of small singular values.
- 2) The left and right singular vectors u_i and v_i tend to have more sign changes in their elements as the index i increases, i.e. as σ_i decreases.

Although these features are found in many discrete ill-posed problems arising in practical applications, they are unfortunately very difficult or perhaps impossible to prove in general.

To see how the SVD gives insight into the ill-conditioning of G , consider the following relations:

$$\left. \begin{aligned} G v_i &= \sigma_i u_i \\ \|G v_i\|_2 &= \sigma_i \end{aligned} \right\} \quad i = 1, \dots, n, \quad (2-12)$$

We see that a small singular value σ_i , compared to $\|G\|_2 = \sigma_1$, means that there exists a certain linear combination of the columns of G , characterized by the elements of the right singular vector v_i , such that $\|G v_i\|_2 = \sigma_i$ is small. In other words, one or more small σ_i implies that G is nearly rank deficient, and the vectors v_i associated with the small σ_i are numerical null-vectors of G . From this and the characteristic features of G we conclude that the matrix in a discrete ill-posed problem is always highly ill-conditioned, and its numerical null-space is spanned by vectors with many sign changes.

The SVD also gives important insight into another aspect of discrete ill-posed problems, namely the smoothing effect typically associated with a square integrable kernel. Notice that as σ_i decreases, the singular vectors u_i and v_i become more and more oscillatory. Consider now the mapping Gx of an arbitrary vector x . Using the SVD, we get $x = \sum_{i=1}^n (v_i^T x) v_i$ and

$$Gx = \sum_{i=1}^n \sigma_i^{-1} (v_i^T x) u_i. \quad (2-13)$$

This clearly shows that due to the multiplication with σ_i , the high-frequency components of x are more damped in Gx than the low-frequency components. Moreover, the inverse problem, namely that of computing x from $Gx = b$ or $\min \|Gx - b\|_2$, must have the opposite effect: it amplifies the high-frequency oscillations in the right hand side b .

2.3.2 SVD for Spectrum Estimation

The SVD strategy can be used to solve the spectrum estimation problem under irregular sampling. Define the data space: $Y \subseteq C^{n_s+1}$ as a vector with n_s+1 elements for a specific pixel. Define the unknown space: $\Gamma \subseteq L(-a, a)$ which is the continuous reflectivity function of the target in the elevation direction with extent $(-a, a)$. With a compact operator L , we can switch from the unknown space to the data space as follows:

$$g = L\{\gamma\} \quad L\{\cdot\} = \int_{-a}^a \cdot \exp[j2\pi f_n s] ds \quad , \quad (2-14)$$

$$g_n = \gamma_1 \exp(j2\pi f_n s_1) + \gamma_2 \exp(j2\pi f_n s_2) \quad . \quad (2-15)$$

With an Adjoint operator L^* , the inverse process can be implemented:

$$\gamma = L^*\{g\} = \sum_{n=0}^N \exp[-j2\pi f_n s] g_n \quad . \quad (2-16)$$

The difficulty is that we cannot implement this operator directly as we have to either discretize the reflectivity function along the elevation direction or transfer to a regularly sampled data space and implement spectrum estimation. These two ways are essentially equivalent. Theoretically, we can use the Adjoint operator to transfer the few irregularly sampled data to the unknown space and then using the compact operator back to such a data space which is regularly sampled with the number of samples we want:

$$h_m = (LL^*g)_m \quad \text{with,} \quad b_{\perp n} = b_{\perp} \left(m - \frac{N}{2}\right) \quad f_m = \frac{2b_{\perp m}}{\lambda r} \quad m = 0, \dots, N \quad . \quad (2-17)$$

The tricky thing here is that the nonlinear operator can be replaced by a Grant matrix G with $G_{mm} = 2a \sin c[2a(f_m - f_n)]$ which means $h_m = \sum_{n=0}^{n_s} G_{m,n} g_n$ with $m = 0, \dots, N$ and $n = 0, \dots, n_s$, which can be proven as follows:

$$\begin{aligned} h_m &= (LL^*g)_m = (L\gamma)_m = \int_{-a}^a \gamma(s) \exp[-j2\pi f_m s] ds \\ \gamma(s) &= \sum_{n=0}^N g_n \exp[-j2\pi f_n s] \\ \Rightarrow \quad h_m &= \int_{-a}^a \exp[-j2\pi f_m s] \sum_{n=0}^N g_n \exp[-j2\pi f_n s] ds \\ &= \sum_{n=0}^N g_n \int_{-a}^a \exp[j2\pi(f_m - f_n)s] ds \\ &= \sum_{n=0}^N g_n \sin c[2a(f_m - f_n)] \\ \Rightarrow \quad \vec{h} &= \mathbf{G} \vec{g} \quad . \end{aligned} \quad (2-19)$$

Practically, error propagation is the main problem. However, the problem described by 2-19 fits exactly to 2-13. Therefore, instead of resampling by direct matrix multiplication, an SVD strategy will be implemented. As described in Figure 2.6, with simulated data, the Grant matrix can be generated first. Implementing the SVD algorithm on the G matrix, we can resample the data regularly. In order to solve the associating ill-conditioning problem,

regularization methods such as singular values truncation are implemented. Based on this resampled data, we can estimate the reflectivity function along elevation direction by transforming the “new” data to the spectral domain. With some model selection criteria, the number of scatterers and their relative position within one resolution cell along the elevation axis can be estimated.

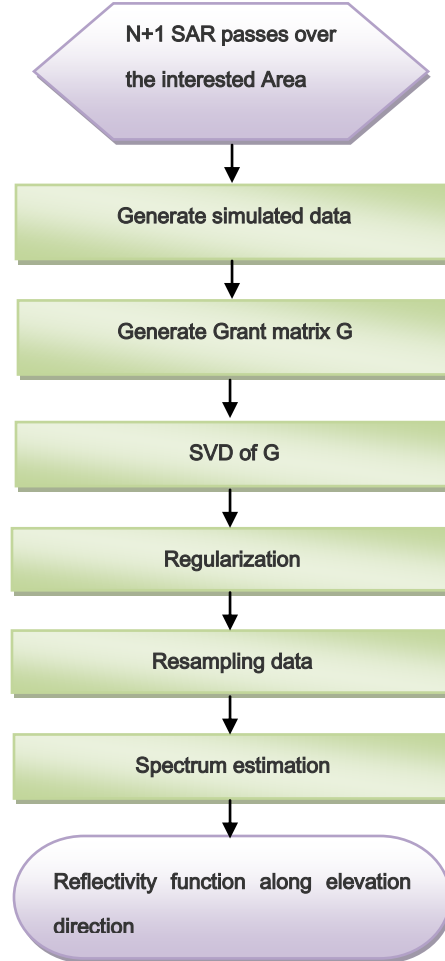


Figure 2.6: Flow chart of the SVD spectrum estimation strategy

2.4 Nonlinear Least Squares Estimation

Consider the noise-corrupted SAR observations of n_p complex-valued sinusoids, the observation equation is:

$$g_n = \sum_{k=1}^{n_p} \gamma_k e^{j2\pi f_n s_k} + v_n \quad n = 0, \dots, n_s \quad , \quad (2-20)$$

Where

γ_k	complex amplitude of k^{th} scatterer
s_k	elevation of k^{th} scatterer
$n_s + 1$	number of available data samples
v_n	observation noise, which is complex zero mean Gaussian noise
n_p	number of scatters within this resolution cell
f_n	frequency of sampled FT which depends on the baseline distribution
g_n	complex valued observation for n^{th} pass

As both the complex amplitude and elevation are unknown, the spectrum estimation problem is nonlinear. In order to solve nonlinear problem, there are two methods to reduce the complexity of the problem (Kay, 1993): transformation of parameters or separation of parameters

In the first case, we seek a one-to-one transformation of unknown parameters that produces a linear signal model in new space. After that, implement least square estimation to transformed unknowns. Finally convert back to the unknown parameters. This approach relies on the properties that the minimization can be carried out in any transformed space that is obtained by a one to one mapping and then converted back to the original space. The determination of the transformation, if it exists, is usually quite difficult. Suffice it to say: only a few nonlinear LS problems may be solved in this matter.

A second type of nonlinear LS problem that is less complex than the general ones exhibits the separability of property. Although the signal model is nonlinear, it may be linear in some of the parameters. That is true in our problem: the amplitude and elevation can be separated to the following form:

$$\vec{g} = H(\vec{s}) \cdot \vec{x} + \vec{u} \quad , \quad (2-21)$$

where

$$\vec{g} = \begin{bmatrix} g_0 \\ \cdot \\ \cdot \\ \cdot \\ g_{n_s} \end{bmatrix}_{(n_s+1) \times 1}, \quad H(\vec{s}) = \begin{bmatrix} e^{j2\pi f_0 s_1} & \dots & e^{j2\pi f_0 s_{n_p}} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ e^{j2\pi f_{n_s} s_1} & \dots & e^{j2\pi f_{n_s} s_{n_p}} \end{bmatrix}_{(n_s+1) \times n_p}, \quad \vec{x} = \begin{bmatrix} \gamma_1 \\ \cdot \\ \cdot \\ \cdot \\ \gamma_{n_p} \end{bmatrix}_{n_p \times 1}, \quad \vec{u} = \begin{bmatrix} v_0 \\ \cdot \\ \cdot \\ \cdot \\ v_{n_s} \end{bmatrix}_{(n_s+1) \times 1}.$$

Where H is a $(n_s + 1) \times n_p$ matrix depending on the unknown elevations of the scatterers. As this model is linear in amplitude and nonlinear in elevation, the LS error may be minimized

with respect to \vec{x} and thus reduced to a function of elevations only, which means a K dimensional search is needed. Since the object function:

$$J(\vec{s}, \vec{x}) = (\vec{g} - H(\vec{s})\vec{x})^T (\vec{g} - H(\vec{s})\vec{x}), \quad (2-22)$$

\vec{x} that minimizes J for a given \vec{s} is

$$\hat{x} = (H^T(\vec{s})H(\vec{s}))^{-1} H^T(\vec{s})\vec{g}. \quad (2-23)$$

The resulting error is

$$J(\vec{s}, \hat{x}) = \vec{g}^T [I - H(\vec{s})(H^T(\vec{s})H(\vec{s}))^{-1} H^T(\vec{s})] \vec{g}. \quad (2-24)$$

The problem now reduces to a maximization of $\vec{g}^T H(\vec{s})(H^T(\vec{s})H(\vec{s}))^{-1} H^T(\vec{s})\vec{g}$ over \vec{s} and a grid search can be used.

The following flow chart shows the Nonlinear Least Squares estimation procedure for spectrum estimation.

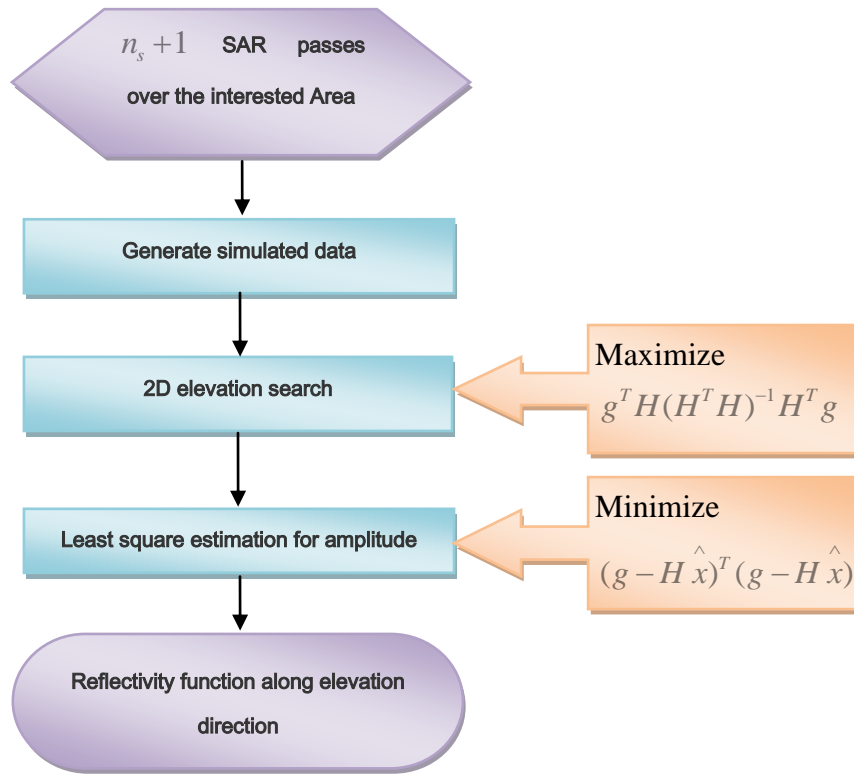


Figure 2.7: Flow chart of the Nonlinear Least Squares spectrum estimation strategy

To reduce the computational complexity, we assume that there are only two scatterers within one resolution cell. A two dimensional search is then required. After that the amplitudes can be retrieved by least square estimation. Finally we get the reflectivity function along the elevation direction.

When the noise is zero normally distributed and white, the NLS is also the MLE, the

theoretically optimum (achieves the CRB) for large samples. This method is very expensive as a multi-dimension search is required. When only two scatterers are assumed and with a small data stack, it is reasonable.

2.5 Summary

- 1) The fundamentals of SAR tomography were presented in this chapter. Its relation to spectrum estimation was discussed and the spectral estimators to be evaluated in this thesis were presented.
- 2) Although interpolation in data space is only possible when we have enough samples, for completeness, it is also considered in this thesis.

3 Spectrum Estimation Results with Simulated and TerraSAR-X Data

The spectrum estimators are evaluated on both simulated and real data. For simulated data, it is possible to compare the results with the simulated truth and effects caused by different factors can be separated. For real data, it verifies whether the strategies will be successful in practice.

Section 3.1 explains how the data are simulated. Spectrum estimation results from simulated data under different simulation condition are discussed in Section 3.2 and Section 3.3. Results with TerraSAR-X data at test site Las Vegas are discussed in Section 3.4. Section 3.5 summarized this chapter.

3.1 Data Simulation

The spectrum estimation algorithm is firstly evaluated on simulated data. For simplicity, we start with the case of two point scatterers within one resolution cell instead of continuous volume scattering.

The reflectivity function is then:

$$\gamma(s) = \bar{\gamma}_1 \delta(s - s_1) + \bar{\gamma}_2 \delta(s - s_2) \quad , \quad (3-1)$$

where s_1 and s_2 are the true elevations of the two scatterers.

In order to separate the effects of baseline distribution and noise, the complex-valued SAR data are simulated in two ways:

Table 3.1; Two data simulation cases

Case 1 (pure simulated)	Case 2 (semi-simulated)
No noise	Noise present
Simulated baseline distribution	Baseline distribution from real data

For case 1, the purpose is to investigate the influence of a random baseline distribution, the Grant matrix, the number of samples and the interference between scatterers. The data is simulated as follows.

Firstly, system parameters are initialized with respect to the ERS satellite mission:

Table 3.2: Initialized parameters for case 1

Distance from scene center	r	847.361 km
Uniform orthogonal baseline	b_{\perp}	30 m
Wavelength	λ	0.056m
Expected number of acquisitions	$N + 1$	99
Actual number of acquisitions	$n_s + 1$	15
Extent of image scene in elevation direction	a	$a_{nyq} / 2$

The focused data at (x_0, r_0) with different baselines along the elevation direction is then simulated as:

$$b_{\perp n} = b_{\perp} \left(n - \frac{N}{2} \right), f_n = \frac{2b_{\perp n}}{\lambda r} \quad n = 0, \dots, N$$

$$g_n = \bar{\gamma}_1 \exp(j2\pi f_n s_1) + \bar{\gamma}_2 \exp(j2\pi f_n s_2) \quad , \quad (3-2)$$

From the $N+1$ regular samples, $n_s - 1$ randomly chosen samples and two endpoints are taken. The reason for fixing the two endpoints is that the minimal distinguishable distance between two scatterers depends on the baseline range (i.e. the maximum difference between baselines). In this case, the baseline distribution is not strictly irregular, but regular with missing samples. Nevertheless, it is a reasonable approximation to the real world. With the baseline distribution in Table 3.2, the elevation resolution (the minimal distinguishable distance between two scatterers) can be computed with the following expression:

$$\text{baseline range} = (99 - 1) \times 30m = 2940 m;$$

$$\text{elevation resolution} = \frac{\lambda r}{2 \times \text{baseline range}} = 8.07m$$

For case 2, a real baseline distribution of TerraSAR-X Spotlight data is used. The main purpose here is to investigate the influence of noise. System parameters are initialized to those of the TerraSAR-X satellite mission instead of ERS.

Table 3.3: Initialization parameters for case 2

Distance from scene center	r	704.000 km
Wavelength	λ	0.031m
Actual number of acquisitions	$n_s + 1$	9
extent of image scene in elevation direction	a	depends on the test site

Subsequently, the focused data at (x_0, r_0) are simulated with the following baseline distribution in the elevation direction:

$$\vec{b} = [0.00 \ -98.17 \ -40.55 \ -94.99 \ -91.39 \ -107.18 \ 132.86 \ 107.719 \ -83.73] \text{ [m]}$$

With a base range of 240.0410 meter, the elevation resolution is 45.5 meter. Therefore, even from the theoretical point of view, it is impossible to distinguish two scatterers along elevation direction with a distance smaller than 45.5 meter and the only possibility to improve resolution is having more acquisitions with larger baseline range. Simulated data are then as:

$$f_n = \frac{2b_n}{\lambda r} \quad n = 0, \dots, n_s$$

$$g_n = \bar{\gamma}_1 \exp(j2\pi f_n s_1) + \bar{\gamma}_2 \exp(j2\pi f_n s_2) + \text{noise}, \quad (3-3)$$

The noise samples are independent, identically distributed (i.i.d.) complex zero mean and Gaussian. The SNR is defined as:

$$SNR = 10 \log_{10} \left(\frac{A^2}{\sigma_{noise}^2} \right), \quad (3-4)$$

where, SNR is the signal to noise ratio in dB, A is the amplitude of the signal and σ_{noise} is the standard deviation of the noise.

Results with simulated and semi-simulated data are discussed in Section 3.2 and 3.3.

3.2 Spectrum Estimation Results from Pure Simulated Data

As mentioned in Section 3.1.1, in this simulation, the baseline distribution is simulated in such a way that $n_s + 1$ passes are randomly chosen from an irregular baseline distribution with 99 samples. Since no noise is present, Nonlinear Least Squares estimation always gives the true value when the number of measurements is not less than the number of unknowns. Therefore, only the Singular Value Decomposition is evaluated without singular value truncation which is mainly applied to reduce noise. The main effect of the SVD strategy here is that the Grant matrix is used to resample the measurements. The extent of the image in elevation is assumed to be known.

3.2.1 Baseline Distribution

Figure 3.1 shows an example where 15 samples out of 99 were randomly chosen. The upper plot is the power spectrum of the measurements. In this example, two point scatterers are simulated which are located at 0m and 79.09m (about 10 resolution cells in between) along elevation axis. The corresponding reflectivity of the scatterers is 0.8 and 1 and the extent of the image is 197.7m. Red solid lines with stars are the normalized power spectra found using the SVD. As a reference, the power spectrum using all 99 samples, obtained using the DFT, is also shown with a blue solid line. For comparison, the normalized power spectrum

obtained from the DFT of the data with missing samples is shown with solid a green line (when a sample with a certain baseline is missing, zeros were inserted). The DFT of the full dataset can be treated as a reference which illustrates the best case while the DFT with 15/99 samples, which is the available dataset, shows the worst case.

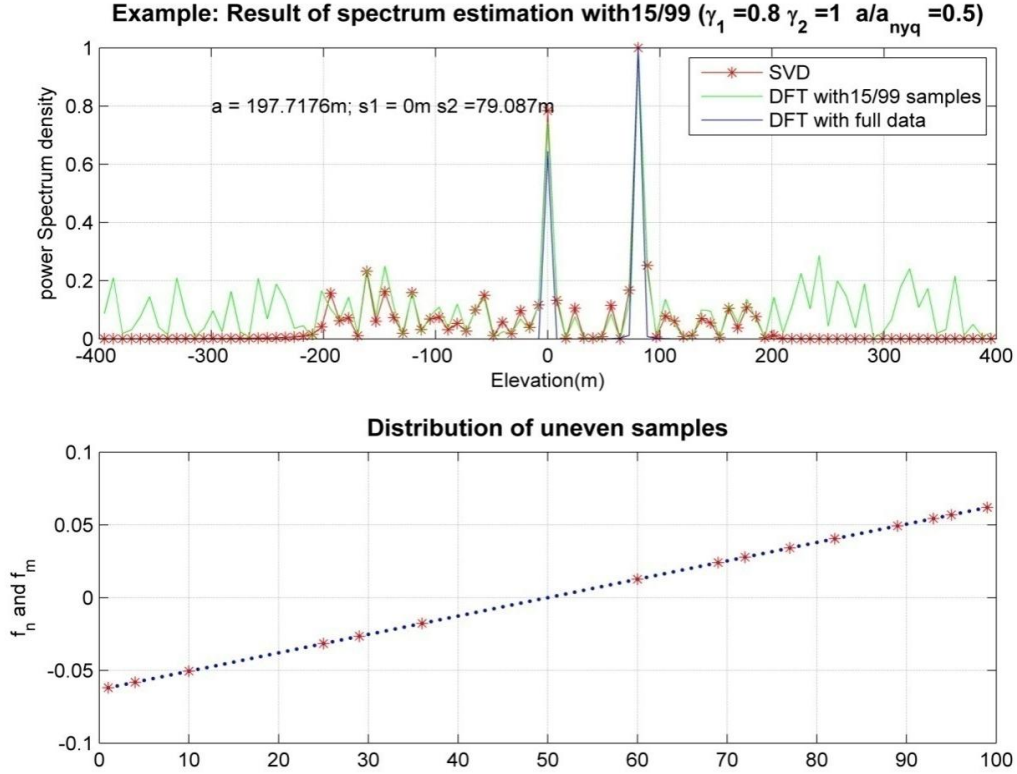


Figure 3.1: Example 1: spectrum estimation result with 15 samples out of 99 with favorable baseline distribution. The upper image refers to the normalized spectrum power; lower plots displays the distribution of uneven samples compared to the regularly sampled full data.

The SVD shows a better performance. With help of the prior knowledge about the extent of the object in elevation direction, the side lobes which may lead to false detection when the baseline distribution is not so favourable are suppressed. This is discussed in detail later in this section.

The relevant baseline distribution of this experiment is described in the lower plot. The blue dots refer to the uniformly sampled observations from which the nonuniformly sampled observations are randomly extracted. In this instance, the baseline distribution is relatively evenly distributed. Compared to the power spectrum obtained from the DFT of the full data, both the DFT with missing samples and the SVD perform well.

On the other hand, when the baseline distribution is not so favourably distributed, there may be a large divergence from the result obtained from the DFT of the full data. Figure 3.2 shows the result obtained for the same conditions but a different baseline distribution compared to the example showed in Figure 3.1. As visualized in the lower plot of Figure 3.2,

the baseline distribution is relatively uneven with both large gaps and very closely spaced samples. Due to the unfavourably distributed baseline, there are large sidelobes beside the main peak as shown in the upper plot of Figure 3.2. In this case, the spectrum of the signal is significantly distorted.

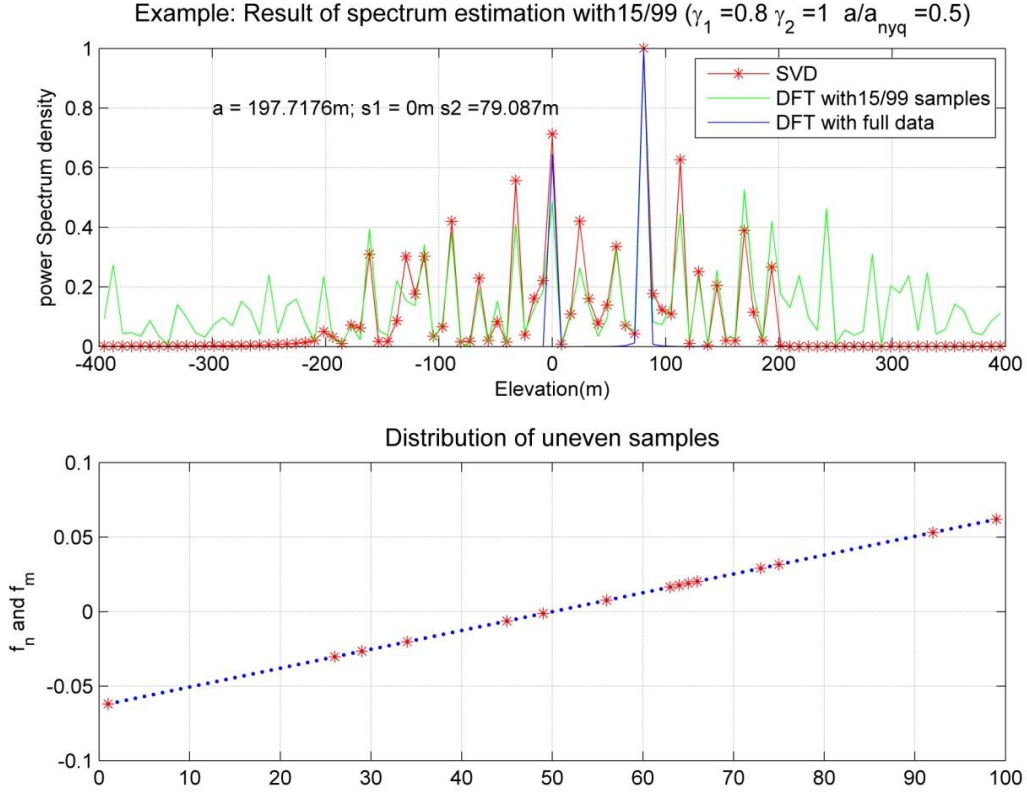


Figure 3.2: Example 2: spectrum estimation result with 15 samples out of 99 with unfavorable baseline distribution.

From the two examples above, we can see that the baseline distribution has a significant effect on spectrum estimation which cannot be ignored. In order to analyze the contribution of the baseline distribution, the ratio of the signal to strongest side lobe in different simulation condition is now investigated statistically.

The histogram of the signal to strongest sidelobe ratio of the estimated normalized power spectrum is plotted in Figure 3.3. The red solid line represents the histogram for the case of 1 scatterer. In this instance, the signal to strongest sidelobe ratio has a mean value of 3.85 and standard deviation 1.18. If only one scatterer is present, the mean signal to strongest sidelobe ratio is mainly due to the small dataset and nonuniform sampling while the standard deviation is mostly due to different baseline distributions. Therefore, we can conclude that the baseline distribution contributes significantly to the estimated power spectrum which is reflected in the histogram.

The blue solid line is for the case of 2 scatterers. In this instance, the signal to strongest sidelobe ratio has mean value 2.0 and standard deviation 0.6. A smaller mean signal to strongest sidelobe ratio and standard deviation are mainly due to the interference effect.

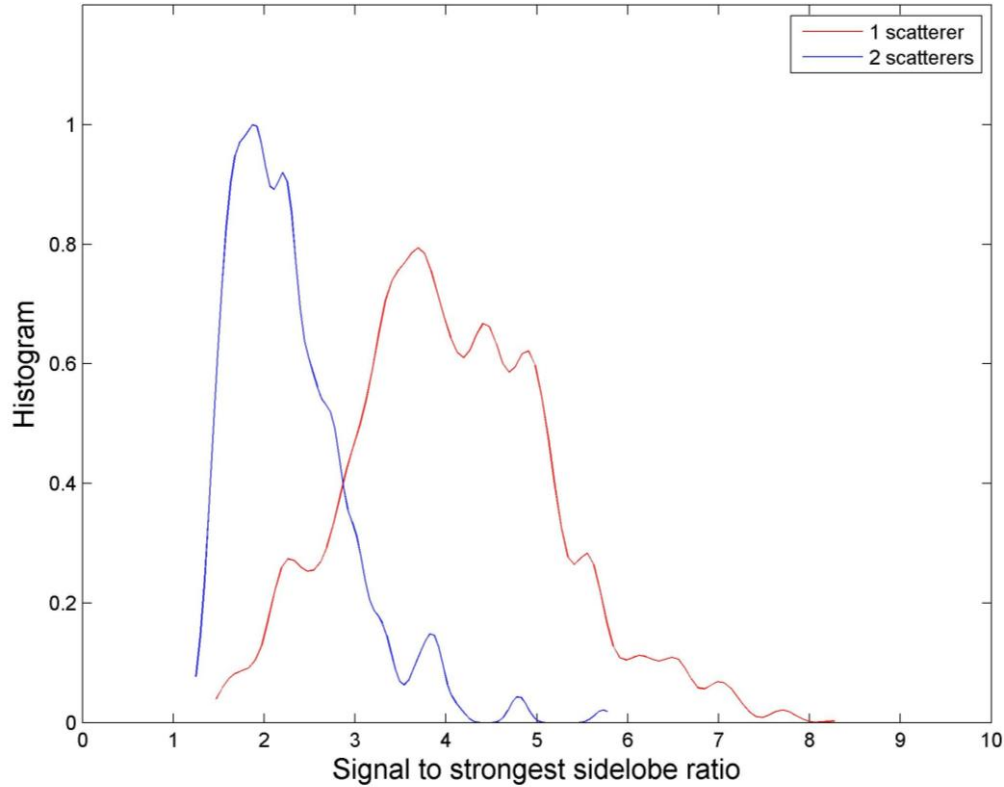


Figure 3.3: The histogram of signal to strongest sidelobe ratio of estimated normalized power spectrum

As singular value decomposition is mainly dealing with singular value and corresponding eigenvectors, it would be very interesting to see how the singular values are dependent on the random baseline distribution. Figure 3.4 shows the singular values behaviors due to different baseline configuration when fix a certain number of samples. From this viewgraph, we can see the spread of the singular values due to the random baseline distribution. For instance, when the number of samples is very small (e.g. 15), the singular value spread caused by a random baseline distribution is comparable to the magnitude of the singular value itself. This is exactly the reason why we choose to add no noise to the pure simulated data - as the baseline distribution has so large an influence on the singular values, it is very difficult to determine the singular value threshold and separate the effects due to the baseline distribution and noise. The conditioning of the problem is defined by the ratio between the largest singular values and smallest singular value. As we can see, the conditioning is much worse when there are more samples. However, it does not mean that more samples has worse estimation capability. As in case of more samples, on the one hand, we can find the useful singular values more clearly and thus set the threshold more easily. On the other hand, more samples means large baseline range, we can always get better resolution when there are more acquisitions available. therefore, better estimation capability

can be obtained if there are more samples available.

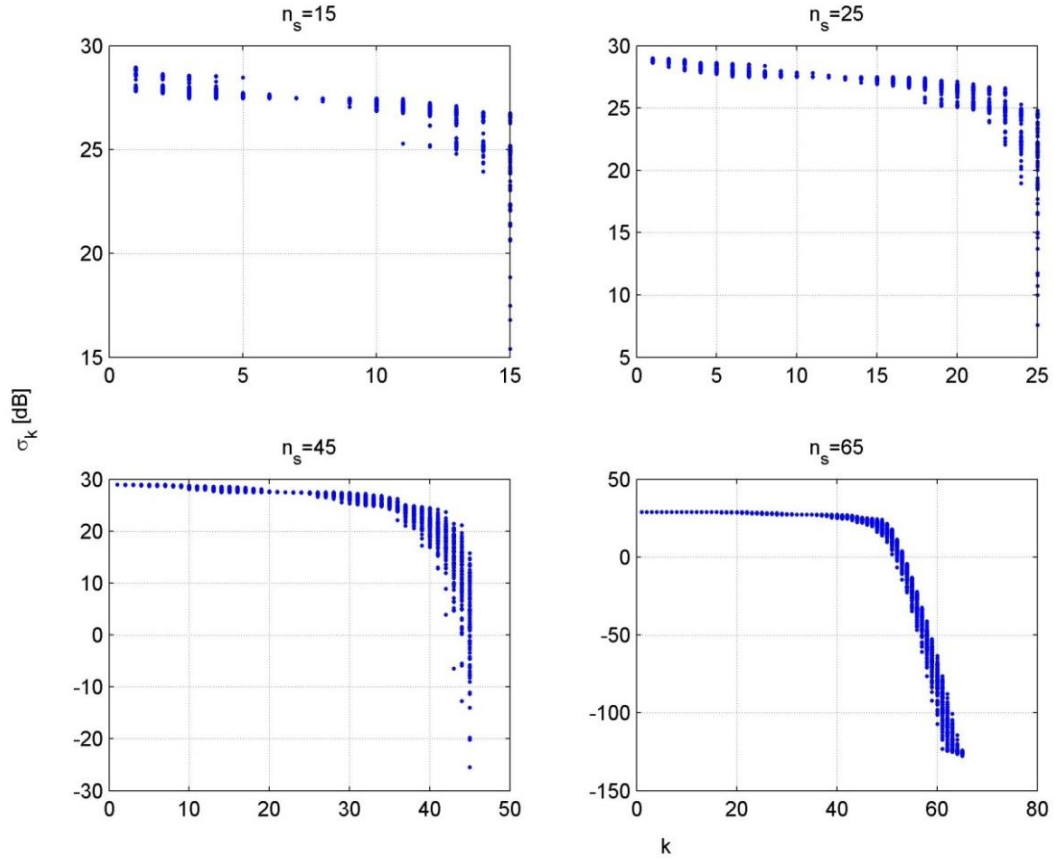


Figure 3.4: Singular values behavior vs. random baseline distribution

3.2.2 Number of Samples and Interference Problem

Since the number of samples starts of very small but increases continuously with the duration of the satellite mission, it is important to see how the performance improves as the number of samples is increased. Above all, it is important to determine the probability with which the scatterers can be correctly detected. As the full data and resampled data are sampled Fourier Transforms in the frequency domain at the same frequency position, a correct detection is defined as occurring when the detected peaks of power spectrum are located at the same position when compared to the result with full regularly sampled data. A Monte Carlo simulation with a randomly extracted baseline distribution was used to evaluate the detection rate. In this section, noise is not yet taken into account, thus the elevation can always be reliably detected when the number of samples is greater than 5 and only one scatterer is present. The case of one scatterer is not of interest, consequently, two scatterers are assumed.

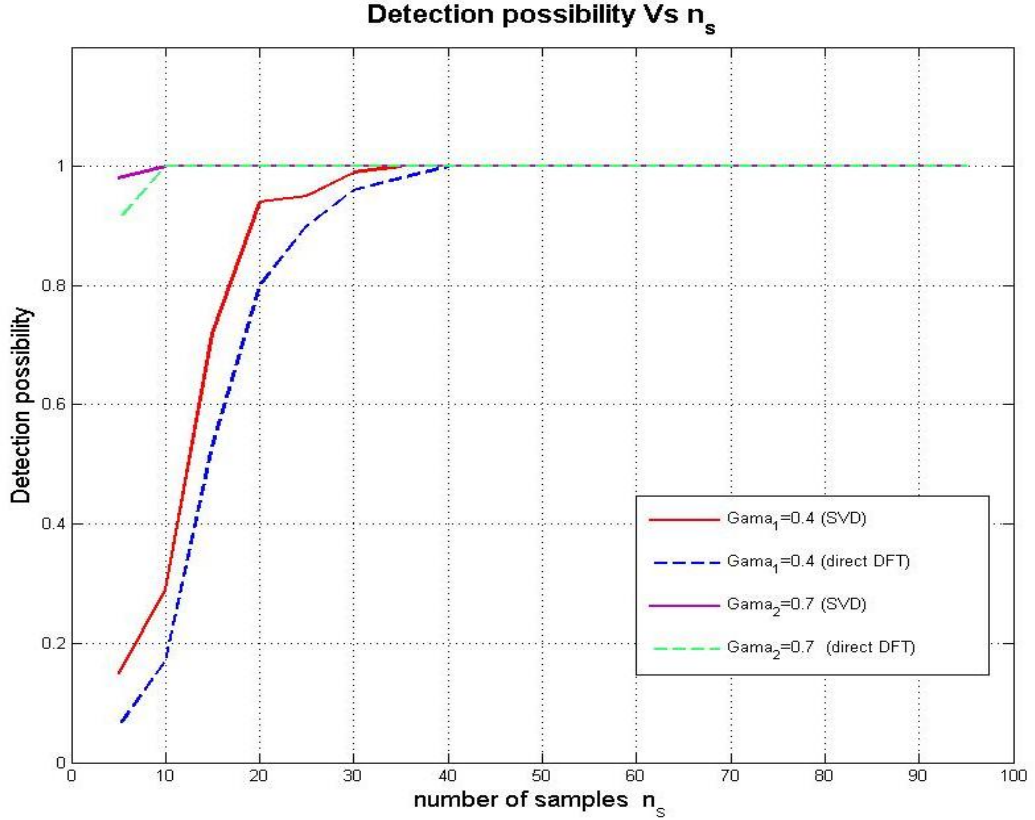


Figure 3.5: Detection possibility of the true elevation varies the number of samples

Figure 3.5 shows the detection rate versus the available number of samples when two scatterers are present in one resolution cell. The reflectivities of the scatterers are 0.4 and 0.7. We can see that the dominating scatterer with reflectivity 0.7 is reliably detected when only 5 samples are available. For the weaker scatterer when the number of samples is not sufficient, the detection rate is reduced due to interference from the other scatterer. This interference effect varies with the amplitude ratio and will be discussed in detail later in this section. For the weaker scatterer, when there are 5 samples available, the detection rate is only 10% using a direct DFT with missing data (blue dash line) and 15% using Grant matrix (SVD, red line). However, the detection rate for the weaker scatterer increases to 50% using a direct DFT with missing data and increases to 70% using the Grant matrix when the number of samples increases to 15. Therefore, we conclude that when the number of samples is not sufficient additional samples can improve the detection rate dramatically. When compared to the DFT with missing data, the SVD performs better, this will be discussed in detail later.

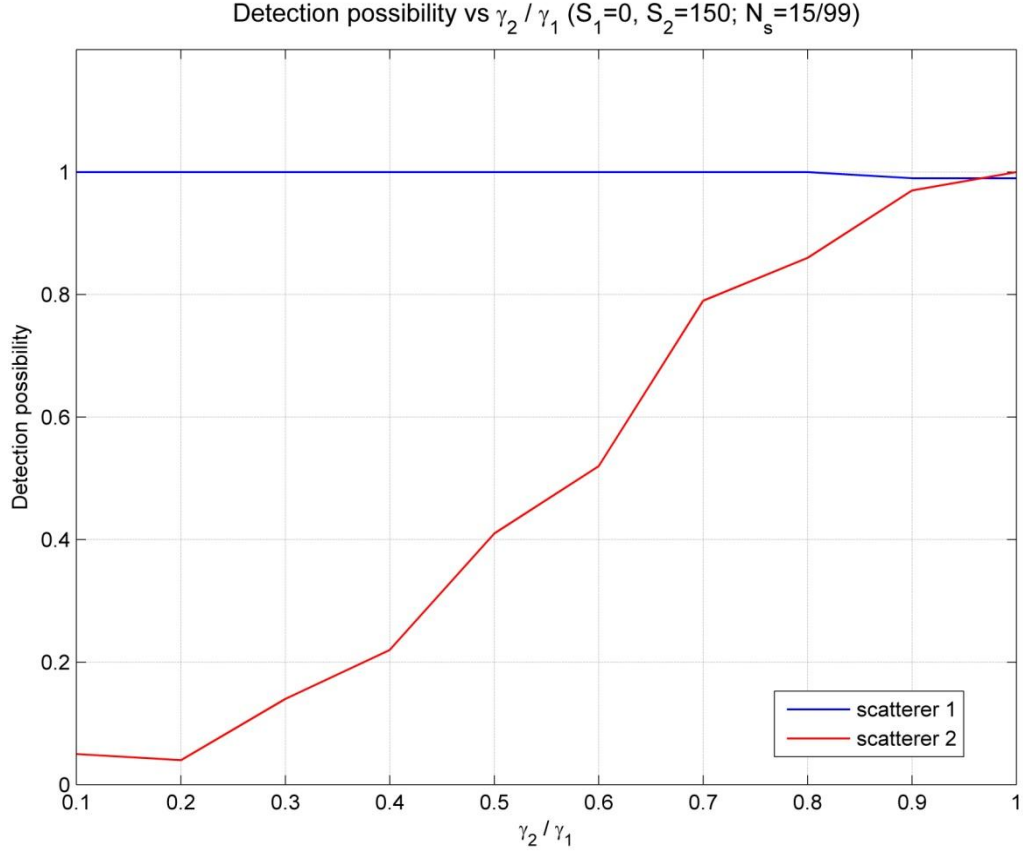


Figure 3.6: Detection possibility of the true elevation varies with the amplitude ratio between two scatterers

In order to investigate the effect of interference between the two scatterers, Figure 3.6 shows the detection rate for both scatterers versus the amplitude ratio between the two scatterers. The interference effect depends not only on the amplitude ratio but also on the relative position and number of samples due to the properties of the sinc function. In this plot the elevations 0m and 150m (those two scatterers are quite far away from each other with roughly 40 resolution cell in between), and number of samples (15/ 99) were fixed. As γ_2 / γ_1 increases toward 1, the detection rate also increases. Only when the amplitudes are comparable does the weaker scatterer have an effect on the detection rate of the stronger one. As plotted in Figure 3.5, when the amplitude of the weaker scatterer is smaller than half of the stronger one, the weaker scatterer may be hidden by the sidelobes of the stronger one.

3.2.3 Performance of Grant Matrix

As explained in Chapter 2, the Grant matrix is used to resample the data to a regular sampling rate. Therefore, in order to evaluate the performance of SVD resampling, it is very important to ensure that the use of the Grant matrix improves the quality of the data or at least has no negative effect. Therefore, detection rate using the DFT with missing samples and using the SVD is investigated as the amplitude ratio of the scatterers varies. In Figure 3.7,

the detection rate is shown as the amplitude of both scatterers varies.

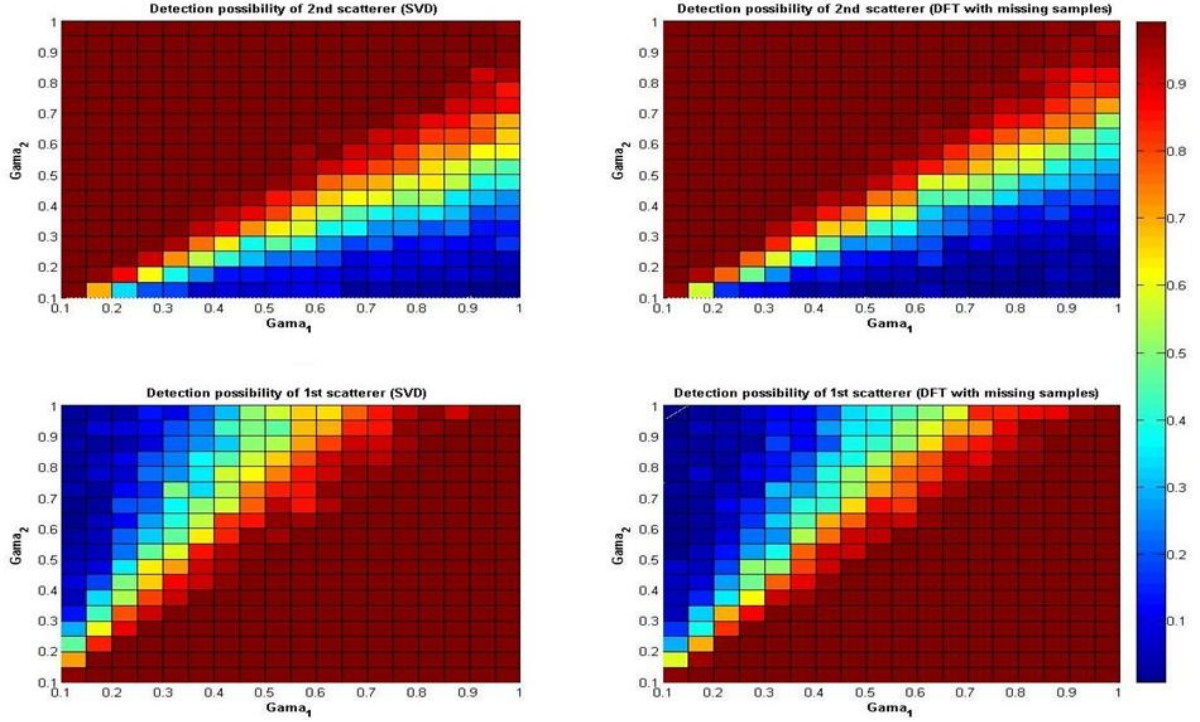


Figure 3.7: Performance of Grant matrix when the amplitude ratio between two scatterers varies

The upper left image shows the detection rate of the second scatterer using the SVD while the upper right image shows the results using a direct DFT. When one scatterer dominates (upper left corner and lower right corner of each image), the Grant matrix can not improve the detection rate of the weaker one as the interference dominates. However, when the two scatterers are comparable (the diagonal of each image), we can see that the Grant matrix shows some improvement in detecting the weaker scatterer. For instance, when $\gamma_2 = 0.1$ and $\gamma_1 = 0.2$, the detection rate of the second scatterer is more than 70% using Grant matrix and smaller than 50% using a direct DFT with missing data. The lower images show the same comparison for the first scatterer which shows the same behaviour as the upper images. Based on this analysis, we can say that the SVD will not reduce the quality of the data but improve it a little when the amplitude of the two scatterers is comparable.

3.3 Spectrum Estimation Results for Semi-Simulated Data

3.3.1 Spectrum Estimation Results using the SVD

For semi-simulated data, the baseline distribution is fixed to the real baseline distribution from a TerraSAR-X data stack as shown in the left image of Figure 3.8. Furthermore, normally distributed complex noise is present.

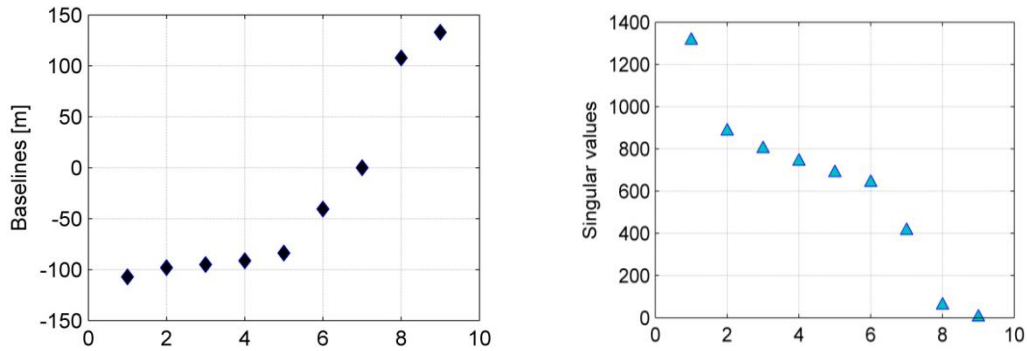


Figure 3.8: Baseline distribution and singular values

The typical singular value pattern when noise is present should be that they decrease slowly at first and then rapidly to the noise level with a sufficient number of samples. Figure 3.9 shows the singular values when 44 samples from the ERS satellite are available. From the plot, we can see that the singular values decrease very rapidly when the number of singular values is larger than 35. Eigenvectors and eigenfunctions associated with low singular values are unreliable as bases in the data and object space, in the sense that in these directions, the Grant matrix transfers only a small amount of information. The ill-conditioning is responsible for the instability of the solution and causes the presence of significant reconstruction distortions. Regularization techniques based on approximate solutions are in this case needed to limit the propagation of errors due to noise and obtain stable solutions.

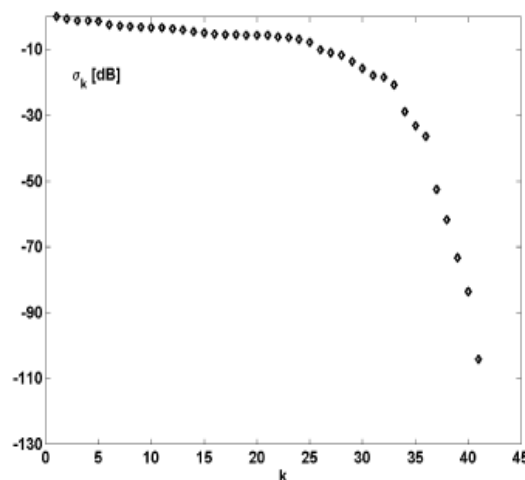


Figure 3.9 : Typical singular value pattern with scale in dB (source: Fornaro et. al, 2003)

When the number of samples is not sufficient, the behavior of the singular values is different as we can see from the left plot of Figure 3.8. The singular values decrease continuously but not rapidly and the minimal singular values is about 5 dB. Therefore, by cutting-off the singular values, there will be some infomation loss associates with noise compression. Thus, choosing threshold at which to cut off singular values becomes much more difficult. Dependent on the noise level, the threshold is determined through experiments. The good thing is when the data stack is small, the optimum number of singular values to retain is quite stable and not sensitive to the noise level.

As an example, two scatterers at position -100m and 180.5m (the distance between two scatterers is approximately 6 resolution cells) with amplitudes 0.8 and 1 (theoretical spectrum power should be 0.64 and 1) are simulated with SNR=4.24dB.

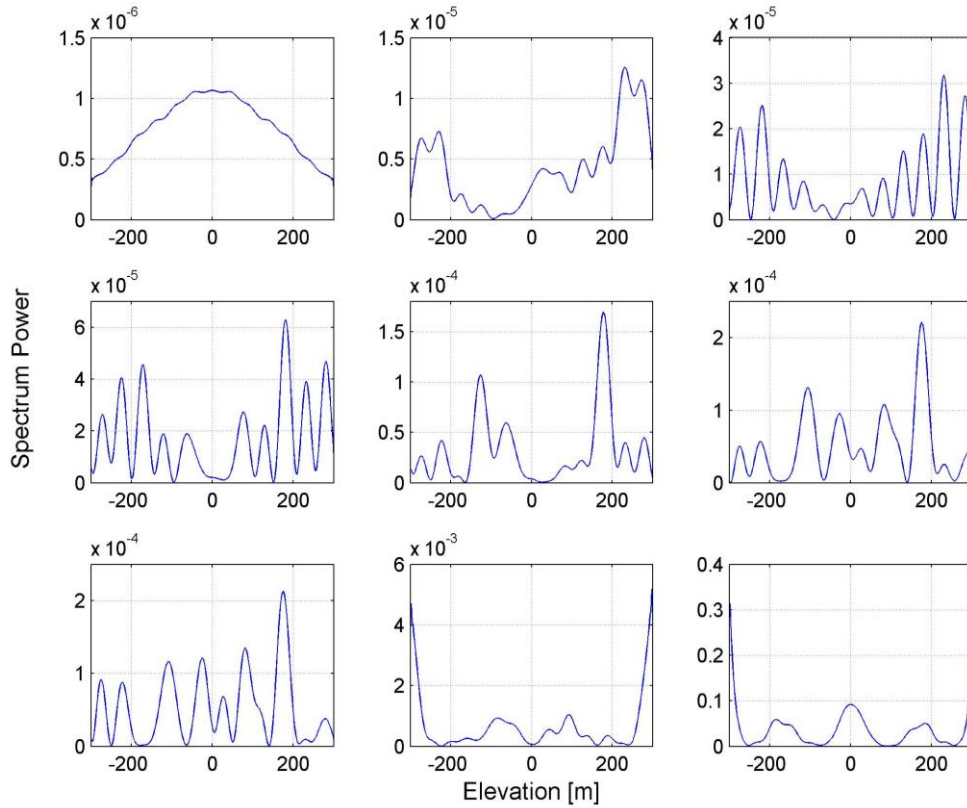


Figure 3.11: Example1: Result with different singular values cutting-off threshold (different number of singular values (1-9) are used; note the scale of each plot).

$$\text{SNR}=4.24\text{dB}, \gamma_1 = 0.8, \gamma_2 = 1, s_1 = -100 \text{ m}, s_1 = 180.5 \text{ m}$$

Figure 3.11 shows the estimated spectrum using different numbers of singular values. For example, the upper left plot shows the estimated spectrum when only the largest singular value is used and the upper middle plot shows the result for when 2 singular values are used; the rest may be deduced by analogy. It's worth to mention that in order to the contributions of different eigenvalues, the spectrum power is not normalized in this viewgraph. From the last two viewgraph, we can clearly see the effect of ill conditioning from the error propogation caused by small singular values. For the purpose of regularitization, one can easily see that when less than 5 singular values are used, too much information about the

signals is lost and when more than 7 singular values are used, too much noise is present. Only when the threshold is chosen in such a way that 5~7 largest singular values are retained, is a reasonable spectrum estimate obtained. The best result comes from using 5 singular values. From this example, we can see that the choice of a threshold is very critical. For this reason, instead of cutting off the singular values, a Wiener filter is implemented and discussed later in this section.

In order to show the effect of singular value truncation, the normalized power spectrum in case of noise free, noisy data and TSVD implemented on the noisy data is displayed in Figure 3.12. The noise free data in red can be treated as a reference; the green solid line shows the curve when direct matrix multiplication instead of an SVD (TSVD) are used; and the blue line shows the same as green curve the only difference being that the singular values are truncated according to the noise level. As for noise free data, there are two scatterers at positions -100 meter and 180.5 meter with reflectivity 0.8 and 1. Due to noise propagation, the weaker scatterer at position -100 meter (green line) is hidden by sidelobes. When the singular values are truncated, the peak at -100 meter is visible but slightly shifted from the true elevation.

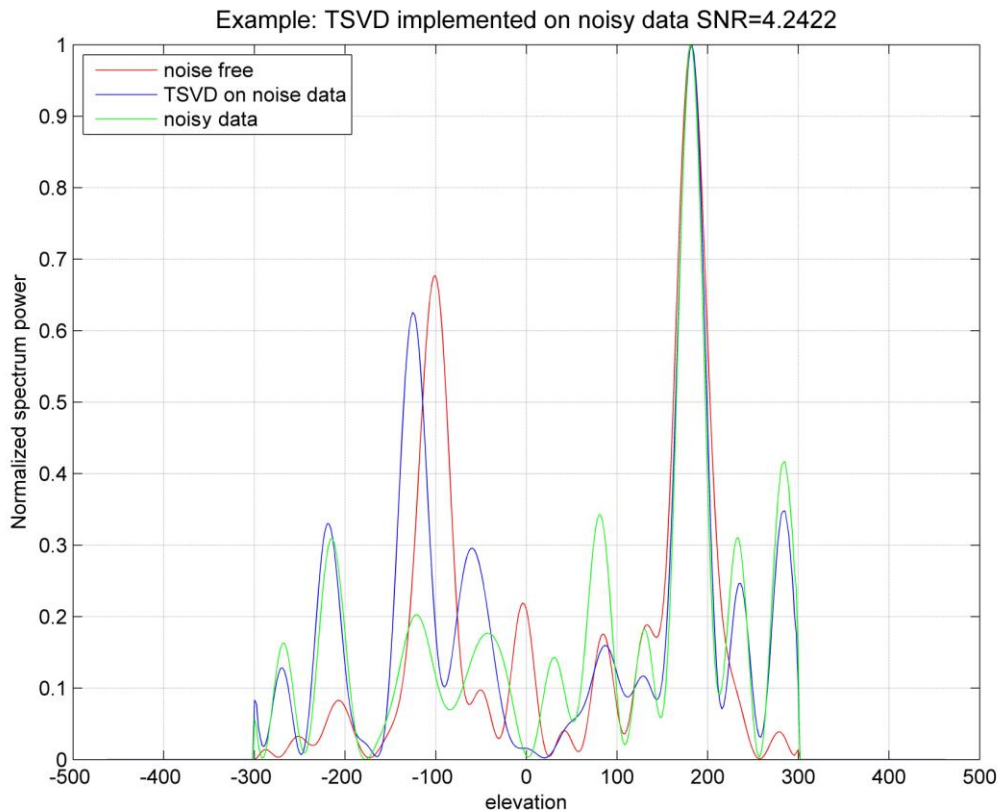


Figure 3.12: Example1: comparison between noise free, noisy data and TSVD

With this example, from one hand we can easily see the influence of regularization method on noise reduction and also the information loss which caused the peak position and amplitude to be shifted from the true position and amplitude. From other hand, we can see the big sidelobes present in the estimated spectrum. It would be very interesting to see the

behaviour of the signal to strongest sidelobe ratio varies with the position of single scatterer. Figure 3.13 shows the estimated normalized spectrum power for scatterer at different elevation position in case of noise free. From this viewgraph, we can see the magnitude of the strongest sidelobe is shift invariant and remains at the level of 0.3. If the noise or multiple scatterers are taken into account, the strongest sidelobe will become larger. Nevertheless, in our application, peaks with a magnitude smaller than 0.3 don't need to be considered as candidates of point scatterers.

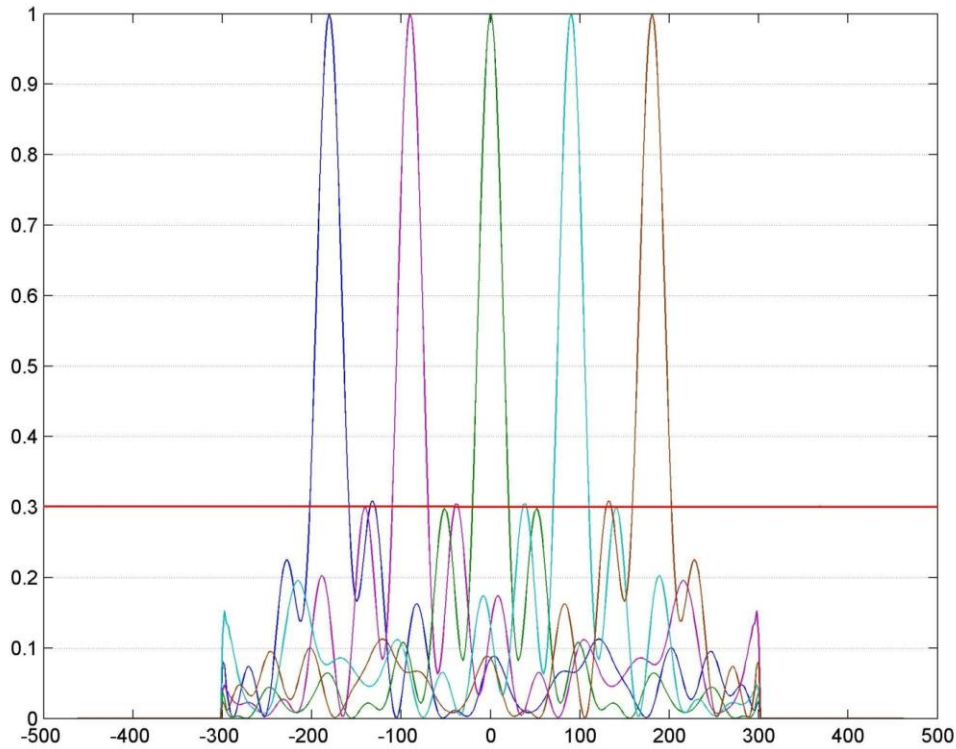


Figure 3.13: Estimation result for scatterers at different position in case of noise free. X axis refers to elevation in meter; y axis refers to the normalized spectrum power.

As another example, the situation that high rise buildings cause broad layover areas in the scene as a result of the high resolution is simulated. The building has an elevation of 440 meter (-220m ~ 220m) where ground is at zero elevation. Figure 3.14 shows the estimated spectrum. The X axis refers to true elevation of points on the building. The Y axis refers to the elevation of the estimated spectrum. The ideal image should be two highlight red lines (one horizontal and another oblique). The upper left image is the result with noise level SNR=0.89dB while from upper right image to lower right image accordingly refers to SNR=3.78dB, 9.81dB and noise free. From this Figure, we can clearly see the building structure and ground even at a very low SNR. When the distance between two scatterers is very small compared to the resolution (-45m ~ 45m), instead of two lines, one straight line with a smaller slope is present. The consequence of information loss can be clearly seen in the noise free image, the stronger power spectrum at the upper left corner and lower right corner do not correspond to reality. This is directly due to the singular value truncation and indirectly due to the interference between the two scatterers.

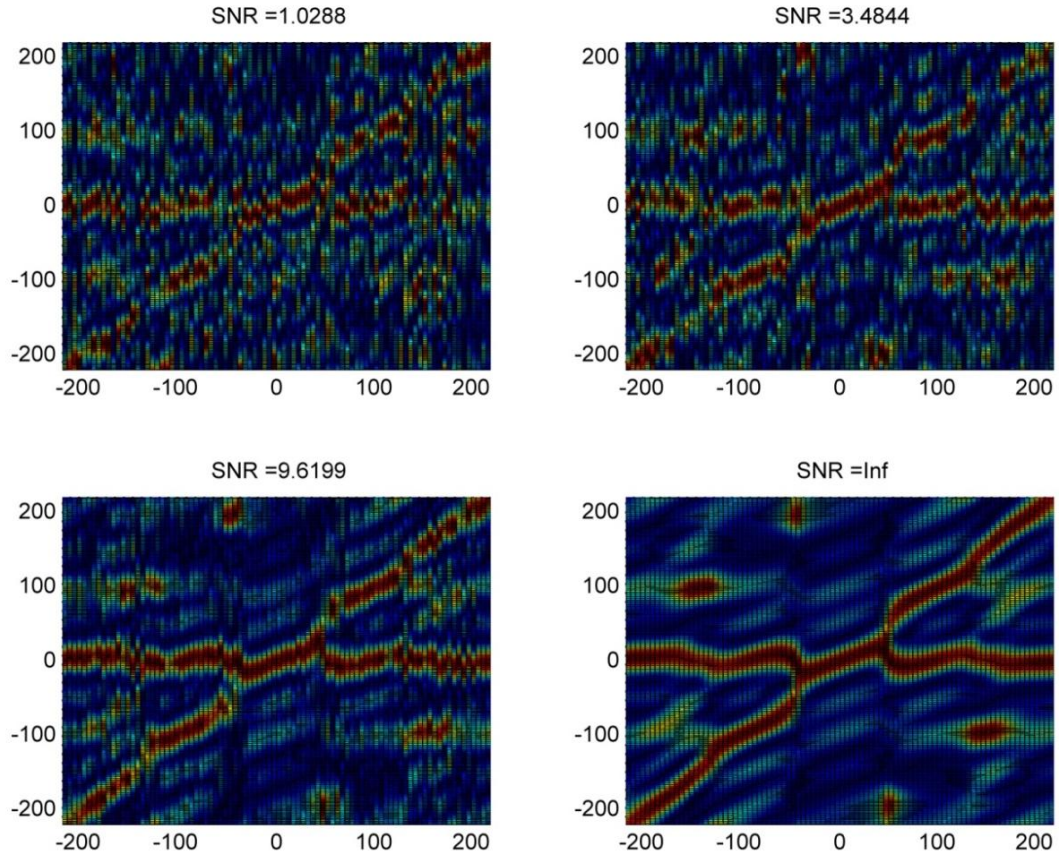


Figure 3.14: Example 2: The situation that high rise buildings cause broad layover areas in the scene as a result of the high resolution is simulated (by using TSVD).

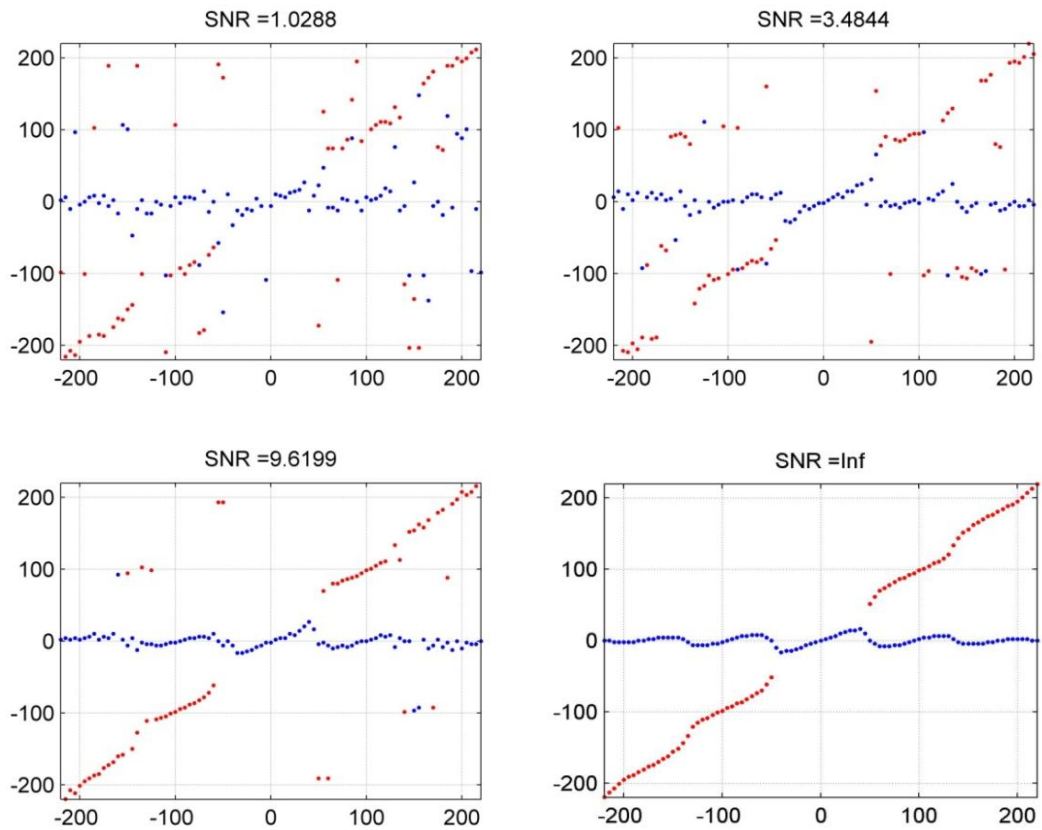


Figure 3.15: Example 2: elevation estimation result in different noise level (by using TSVD).

Figure 3.15 shows the estimated elevation from spectrum estimation based on two scatterers assumption. The estimator simply takes the two highest peaks of the normalized power spectrum. The red line refers to the building with height 440 meter (from -220 meter to 220 meter) while the blue line refers to ground at zero elevation. Even though only a simple elevation estimator is implemented, the estimation results are acceptable when compared to the truth. The TSVD gives good results especially in case of low SNR.

As mentioned above, the truncation threshold is critical when the number of samples is small. In order to improve performance, one possibility is to take more samples, which is time consuming. Another possibility is instead of truncating the singular values, to weight the singular values (filtering for the removal of noise from a “corrupted” signal). For this purpose, the Wiener filter, is discussed in the following section.

3.3.2 Wiener Filter

The Wiener filter is a filter proposed by Norbert Wiener during the 1940s and published in 1949. Its purpose is to remove noise by linear filtering in an optimal way and improve ensemble averaging by incorporating correlation information. The performance criterion is minimizing the mean square error based on the assumption that the signal $s(t)$ and additive noise $n(t)$ are stationary linear stochastic processes with known spectral characteristics or known autocorrelation and cross-correlation. The formula for the optimum filter is given by (William, H. Press, 1992):

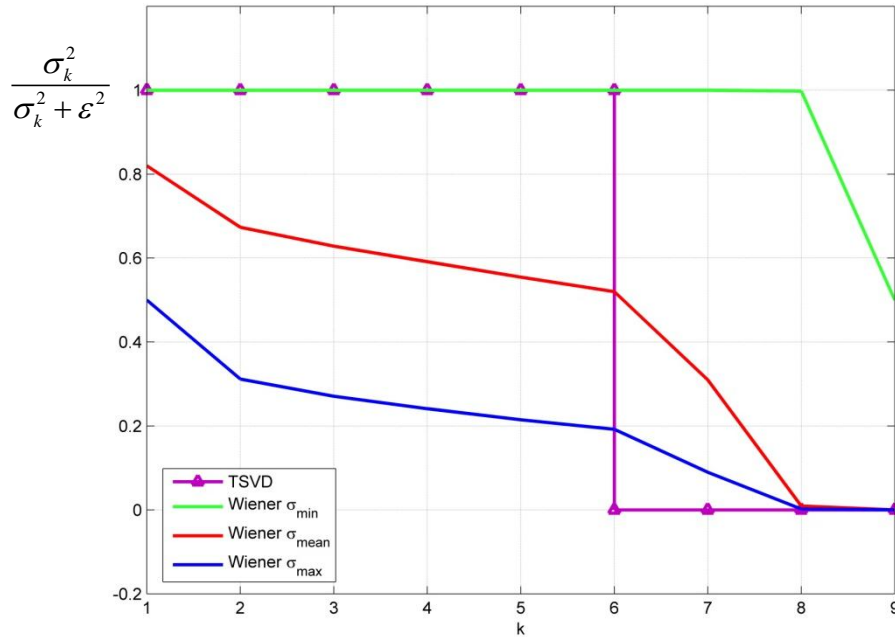
$$H_{wiener}(f) = \frac{|S(f)|^2}{|S(f)|^2 + |N(f)|^2}, \quad (3-6)$$

where $S(f)$ and $N(f)$ are the frequency response of the signal and noise. In our application, the implementation of the Wiener filter is formulated in table 3.4. Instead of truncation, the Wiener filter weights each singular value by a factor $\sigma_k^2 / (\sigma_k^2 + \varepsilon^2)$. σ_k is the k^{th} singular value, and ε is a parameter related to the noise level where smaller ε refers to high SNR. ε should be comparable to the singular values as too small an ε has only a slight regularization effect and too large an ε weights all singular values similarly. Therefore, different choices for ε ($\varepsilon = \sigma_{\min}$, $\varepsilon = \sigma_{\text{mean}}$ and $\varepsilon = \sigma_{\max}$) are discussed next.

Table 3.4: Formulation of Wiener filter in our application

SVD	$\gamma(s) = \sum_{k=0}^N \sigma_k^{-1} \langle \mathbf{g}, \mathbf{u}_k \rangle_Y \mathbf{v}_k(s)$
TSVD	$\gamma(s) = \sum_{k=0}^{n_{\text{cut}}} \sigma_k^{-1} \langle \mathbf{g}, \mathbf{u}_k \rangle_Y \mathbf{v}_k(s)$
Wiener filter	$\gamma(s) = \sum_{k=0}^N \frac{\sigma_k^2}{\sigma_k^2 + \varepsilon^2} \sigma_k^{-1} \langle \mathbf{g}, \mathbf{u}_k \rangle_Y \mathbf{v}_k(s)$

In order to visualize the difference between singular value truncation and Wiener filtering more explicitly, Figure 3.16 shows the different weighting on the eigenvalues using TSVD and Wiener filtering with different ε .


Figure 3.16: The Weighting factor of singular values (TSVD Vs Wiener filter)

For TSVD (purple solid line with triangle), when the singular value is larger than the threshold, a weight of 1 is used, otherwise the weighting factor is 0. For Wiener filtering, different weightings given by $\varepsilon = \sigma_{\min}$, $\varepsilon = \sigma_{\text{mean}}$ and $\varepsilon = \sigma_{\max}$ are shown. When $\varepsilon = \sigma_{\min}$, only the information carried by the smallest two singular values is retained. This is only suitable for extremely high SNRs as visualized in Figure 3.17. In the noise free case, the estimated elevation is perfect and without information losses. However, when some noise is adding to the measurements (even SNR=9.62 dB), the structure of the building (the oblique straight line) is completely hidden by the strong sidelobes caused by noise propagation at the position of the boundary of the extent of the images (here, around -220m and 220m as in

Figure 3.17). The structure of the ground is also very noisy with a large standard deviation.

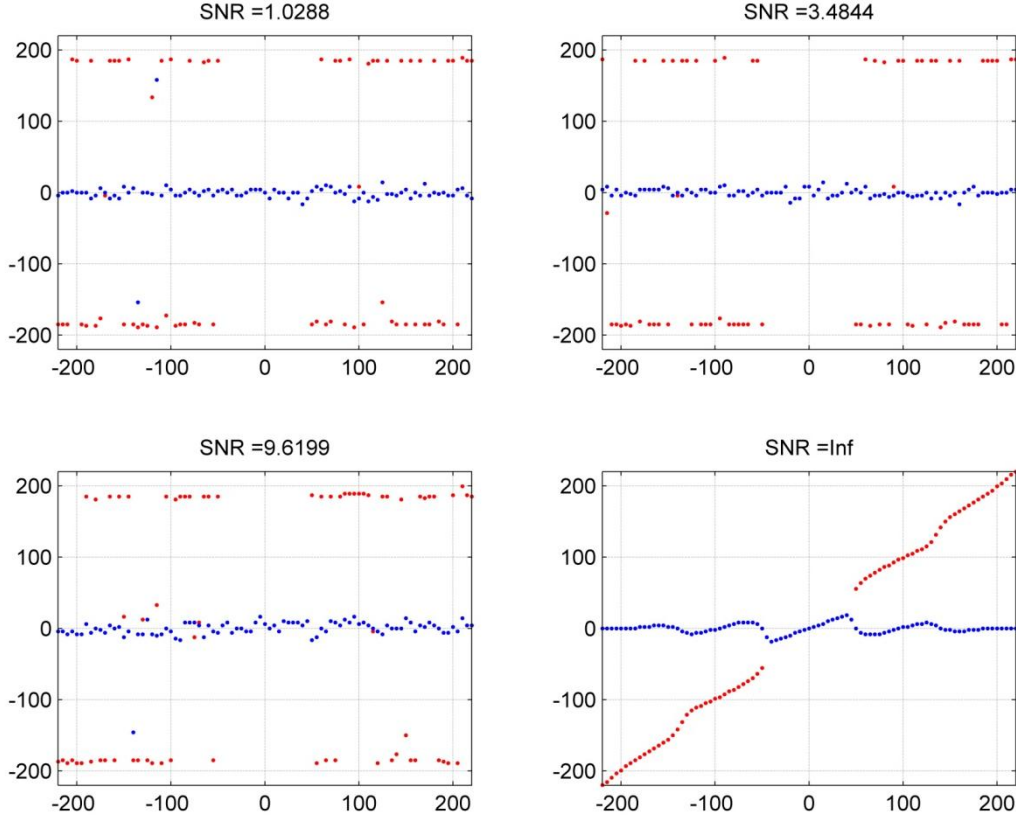


Figure 3.17: Example 2: elevation estimation result with Wiener filtering ($\varepsilon = \sigma_{\min}$)

The red solid line in Figure 3.16 represents a singular value weighting $\varepsilon = \sigma_{\text{mean}}$. Compared to the information carried in the largest singular values, there are some compression effects on information carried by all other singular values. Compared to the previous case, the weighting curve is more moderate which makes the estimator more stable. Figure 3.18 shows the estimation result for different SNRs. In this case, the performance of the Wiener filter is much better as noise has been effectively reduced. We can clearly see the structure of both building and ground. However, there is also some information loss which we can see from the noise free plot. Some scatterers, that should be located on the ground, turn out to be at around 100m and -100m. Figure 3.19 shows the results under the same conditions except that ε is set to be the maximum singular value. From this plot, we cannot see much difference compared to Figure 3.18. However, it is clear that the information loss here is not as severe as in Figure 3.18. The statistical characteristics of the estimation results for the various estimators and parameter settings are compared in detail in Section 3.3.4

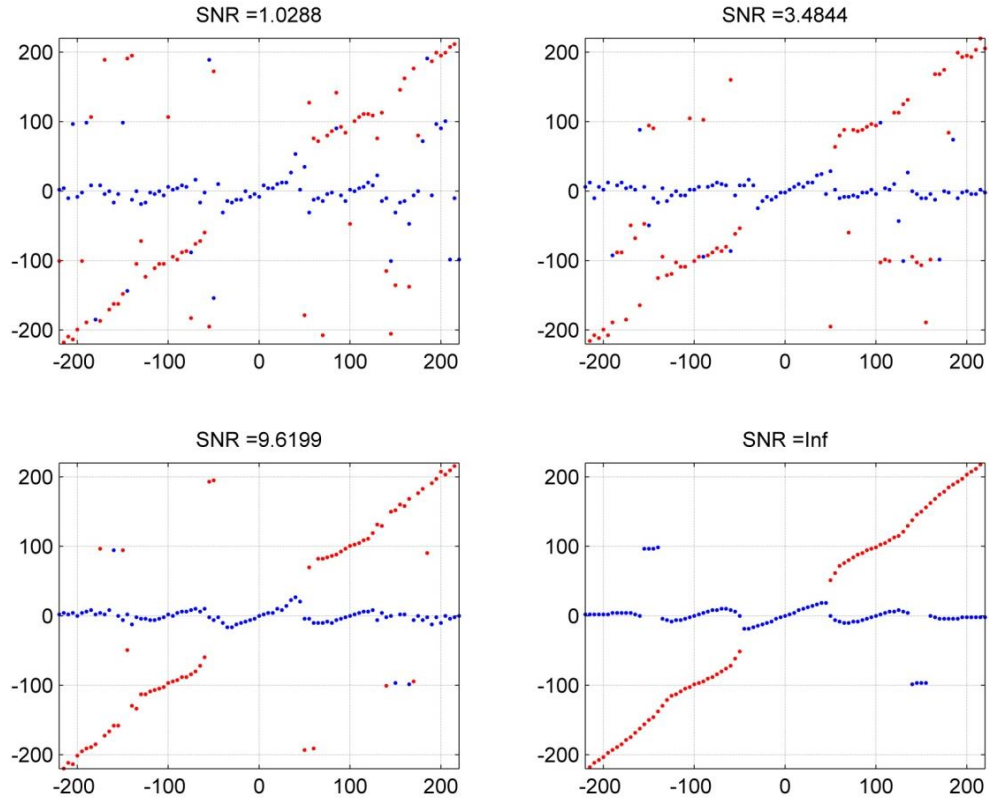


Figure 3.18: Example 2: elevation estimation result with Wiener filtering ($\varepsilon = \sigma_{\text{mean}}$)

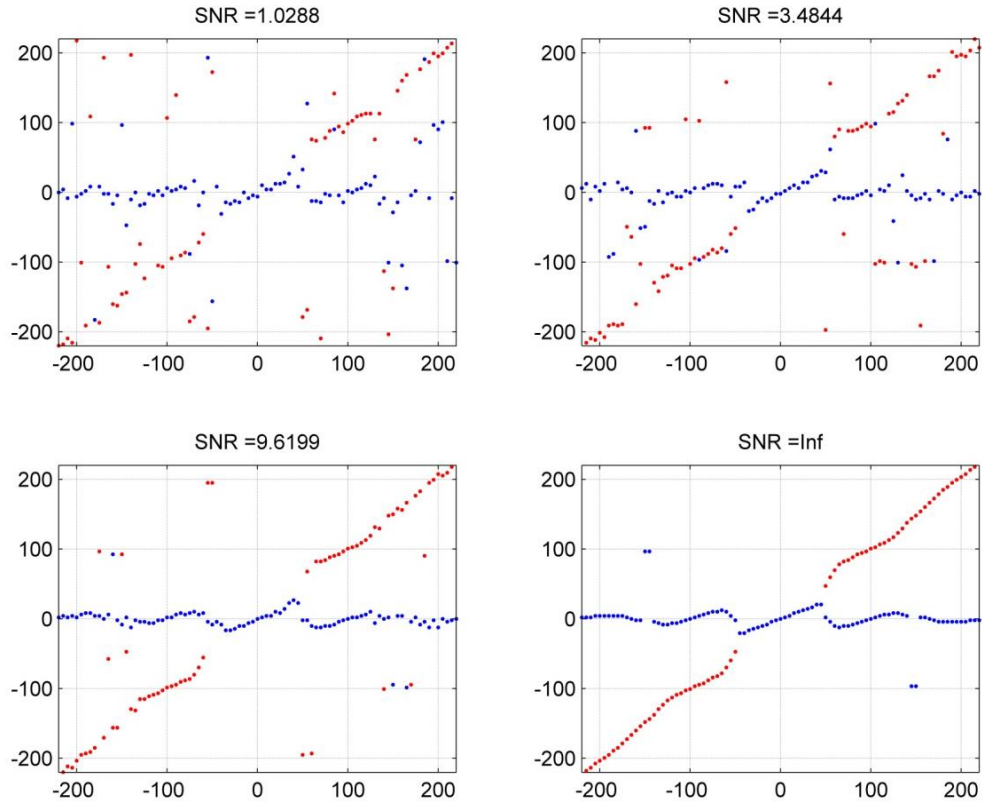


Figure 3.19: Example 2: elevation estimation result with Wiener filtering ($\varepsilon = \sigma_{\text{max}}$)

Even though via Wiener filtering the spectrum estimation is more stable, it may also decrease the estimation resolution. To highlight this effect, the baseline distribution is extended from 9 to 16 (7 new baselines are adding to the original baseline distribution of the real data) as shown in the left plot of Figure 3.20, the right plot shows the singular values corresponding to the new baseline distribution.

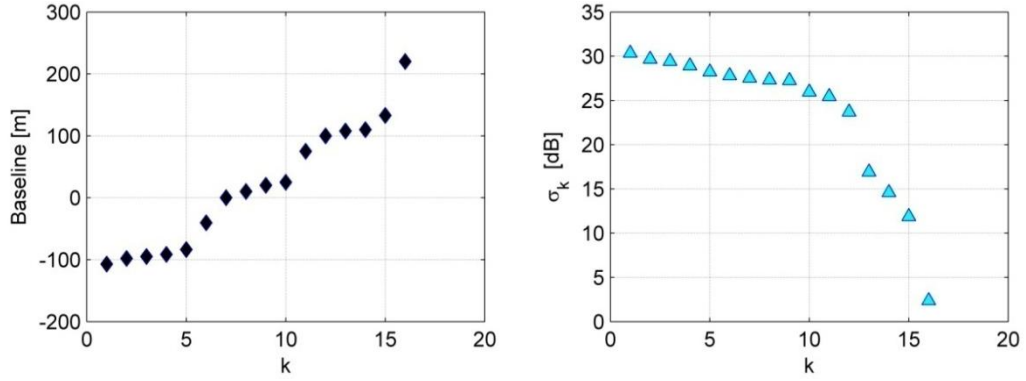


Figure 3.20: New baseline distribution and singular values

Figure 3.21 depicts the comparison between the Wiener filter and TSVD when different numbers of singular values (0~16) are used. In this case, signal reflected from two scatterers at elevation 0m and 50m with amplitude 0.9 and 1 respectively are simulated. In each subplot, the red solid line is the estimated normalized power spectrum for different truncation thresholds; the blue solid line is the estimated power spectrum using the Wiener filter.

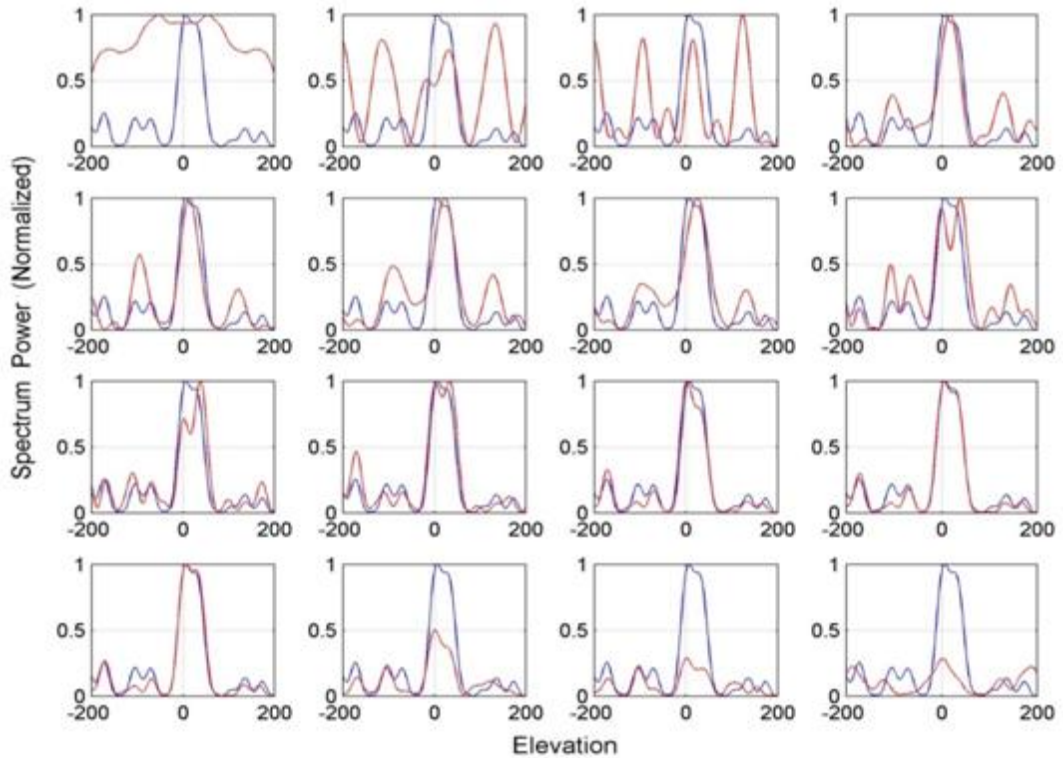


Figure 3.21: Example: by using Wiener filter the estimation resolution might be reduced. (Red: TSVD with different threshold; blue: Wiener filtering)

By analysing the baseline range, the two scatterers should theoretically be distinguishable. However, with Wiener filtering, those two scatterers can not be separated. Meanwhile, we do not need to specify such a threshold and the risk that useful information is removed is relatively small. By contrast, the TSVD can distinguish the two scatterers when the 8 largest singular values are used; however, it is extremely difficult to select the cutting off threshold in such a way that optimum number of singular values are used, as the singular values decrease not very rapidly and slight different threshold may lead to completely different result.

3.3.3 Spectrum Estimation Results using NLS

Figure 3.22 shows the results obtained by using NLS estimation under the same simulation conditions as in Figure 3.13. As explained in Section 2.4, a multi-dimensional search is required which is time consuming. In this example, a priori knowledge that two scatterers are present in a specific resolution cell is used. Thus, instead of a multi-dimensional search, only a two dimensional search in the elevation direction is required.

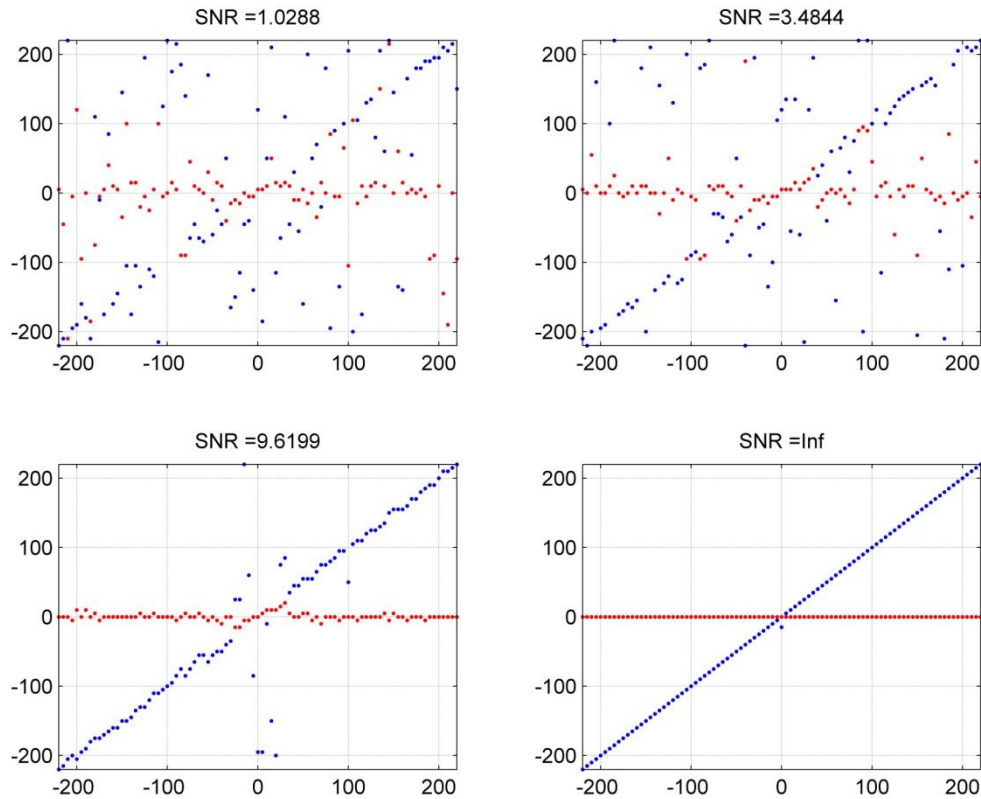


Figure 3.22: Example 2: elevation estimation result in different noise level (by using NLS). (Red: ground; blue: building)

From the upper left plot, we can see that the estimation result is not good when the SNR is very low. It performs worse compared to TSVD. When SNR=3.48dB, the performance of both strategies are comparable to each other. NLS has excellent performance when the SNR is 9.62dB or in the case of no noise. Compared to the result from TSVD, NLS does not show

any resolution limitation due to the baseline distribution, rather there is a strong dependency on the SNR. Even though NLS may provide excellent result when the SNR is relatively low, it is not practical as the critical SNR for PS points is 2dB.

3.3.4 Comparison and Discussion

Based on the analysis above, we can summarize the difference between the truncated singular value decomposition strategy and the Nonlinear Least Squares estimation strategy. First of all, SVD is a non-parametric method while NLS is parametric. As a multi-dimensional search is necessary for NLS, NLS is time consuming while SVD has much lower computational complexity. SVD does not need a priori model selection to determine the number of signals inside the specific resolution cell; the estimated power spectrum provides a possibility for model selection while NLS needs this prior knowledge even for formulating the observation equation. However, the resolution of SVD depends on the baseline range. For instance, in our simulation, only two scatterers with a separation larger than 45 meter can be distinguished. By contrast, NLS does not have resolution dependent on the baseline range from a theoretical point of view and extremely high resolution can be obtained when the SNR is very high. However, it works perfectly only with high SNR, in the case of low SNR, it works much worse than SVD. In practice, SNR=2dB would be the critical SNR for our application which is not high enough for NLS estimation. SVD has an ill-conditioning problem, thus regularization tools such as TSVD and Wiener filtering should be used. SVD can give good and stable performance with regularization. Taking all factors into account, SVD is a better choice here.

Table 3.5 summarizes the statistical properties (false alarm rate, mean of residuals and standard deviation of residual) of the results for example 2 by using different strategies (e.g. singular value truncation vs. Wiener filtering, SVD vs. NLS) or with different parameter settings (Wiener filter) for different SNRs. In the table, Wiener 1, Wiener 2 and Wiener 3 correspond to the Wiener filter with $\varepsilon = \sigma_{\min}$, $\varepsilon = \sigma_{\text{mean}}$ and $\varepsilon = \sigma_{\max}$ respectively.

Let us take the Wiener filter with different ε as an example. Due to the small regularization effect of Wiener 1, the estimation results for low SNR is catastrophic. For instance, with SNR=1.03dB, the false alarm rate of Wiener 1 is 40.45% which is very poor compared to Wiener 2 and Wiener 3. It only gives perfect results in a noise free situation. Wiener 2 and Wiener 3 have a similar performance for different SNRs. In addition, Wiener 2 has slightly better performance for low SNR. For instance, Wiener 2 has a false alarm rate of 29.87% when SNR=1.03 dB while Wiener 3 has a false alarm rate of 30.34%. On the other side, Wiener 3 shows better results in the noise free case due to a more moderate weighting of the singular values. In practice, Wiener 2 should be a better choice compared to Wiener 1 and Wiener 3.

Table 3.5: Comparison between diff. strategies or with diff. parameter setting

SNR [dB]	Method	False alarm (%)	Mean elevation estimate error [m]	Std of elevation estimate error [m]
1.03	TSVD	32.58	4.21	9.05
	Wiener 1	40.45	5.83	7.84
	Wiener 2	29.78	7.15	9.05
	Wiener 3	30.34	7.16	8.94
	NLS	46.63	1.84	8.57
3.48	TSVD	25.84	4.10	8.57
	Wiener 1	39.33	5.67	6.66
	Wiener 2	23.60	6.37	8.48
	Wiener 3	23.60	6.70	8.50
	NLS	36.52	0.66	8.11
9.62	TSVD	10.11	5.33	7.53
	Wiener 1	37.08	7.32	9.39
	Wiener 2	9.55	6.17	7.50
	Wiener 3	9.55	6.63	7.83
	NLS	6.74	0.09	4.94
INF	TSVD	0.0225	5.17	6.66
	Wiener 1	0.0112	5.12	7.02
	Wiener 2	0.0562	5.88	7.61
	Wiener 3	0.0337	5.99	7.72
	NLS	0	0.03	0.37

Based on the conclusion that Wiener 2 performs better than Wiener 1 and Wiener 3, it would be very interesting to compare Wiener 2 with the truncated SVD. Due to the moderate weighting, Wiener filtering appears to have a lower false alarm rate. For example, for an extremely low SNR of 1.03dB, TSVD has a 32.58% false alarm rate while Wiener 2 has a false alarm rate of only 29.78%. For the noise free case, Wiener 2 also shows better performance with a false alarm rate of 5.62%. Therefore, it is sufficient to comment that Wiener filtering which gives different weight according to the singular values has a better and more stable performance than simply truncating the singular values.

The argument that NLS shows best performance for high SNR is proved again. By looking at the mean value and standard deviation of the residuals, a slight bias between the true elevation and estimated elevation appears. The standard deviation, which represents the uncertainty of the estimated elevation, is within 10 meters uncertainty band. Compared to the 45 meter resolution, the estimation precision is excellent.

3.4 Spectrum Estimation Results with TerraSAR-X Data

For the purpose of validation, data of a selected urban area are pre-processed to form a PSI stack. The tomographic algorithm SVD discussed in the previous steps are applied and tuned to the real data. The detected single and multiple scatterers are compared to ground truth. A reference point P_{ref} on the ground is selected to compensate atmospheric effect and relative to this the resolution cells at points P_1 , P_2 and P_3 are estimated.

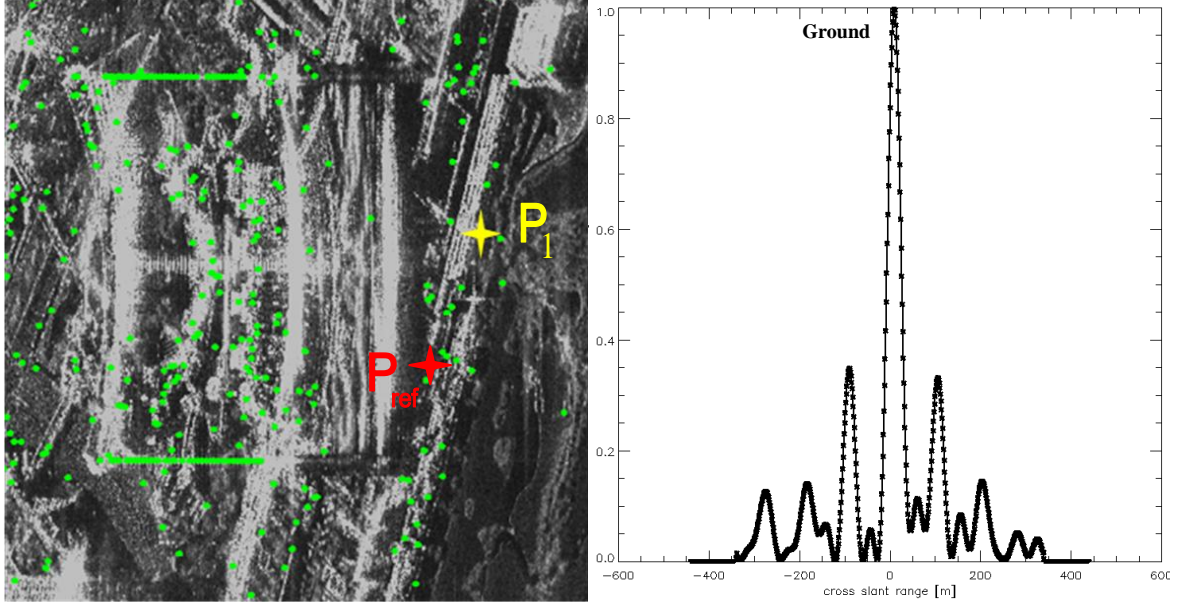


Figure 3.23: Example1: SVD implemented on real data

As visualized in Figure 3.23, the green points refer to the persistent scatterers. The reference point is chosen within the PSs which have higher signal to noise ratio. P_1 is a point on the ground as well and the reference point is selected as the closest point of P_1 . Therefore, the theoretical power spectrum should be a peak around zero elevation. The right plot shows the estimated normalized power spectrum obtained from the SVD strategy from which we can see that the estimated result fits to the ground truth. The sidelobes are quite small compared to the signal which is due to single scatterer (no interference) and the high SNR of PS point.

P_2 and P_3 are two point with similar height and likely have two scatterers within the same resolution cell (one from the ground and another from the building). The same point on the ground is chosen as the reference point; therefore, the estimated spectrum of those two points should show a similar pattern. As the surface of the building is made of high reflective material, the power spectrum should include a relatively weak peak near zero elevation and another relatively strong peak at a relatively high elevation. When checking the estimated result from the right plot of Figure 3.24 and Figure 3.25, we conclude that the estimate is consistent with the ground truth even though the sidelobes are quite high due to the

interference between the two scatterers and the limited number of available scenes. Due to the repeat cycle of only eleven days no problem is expected with typical data stacks.

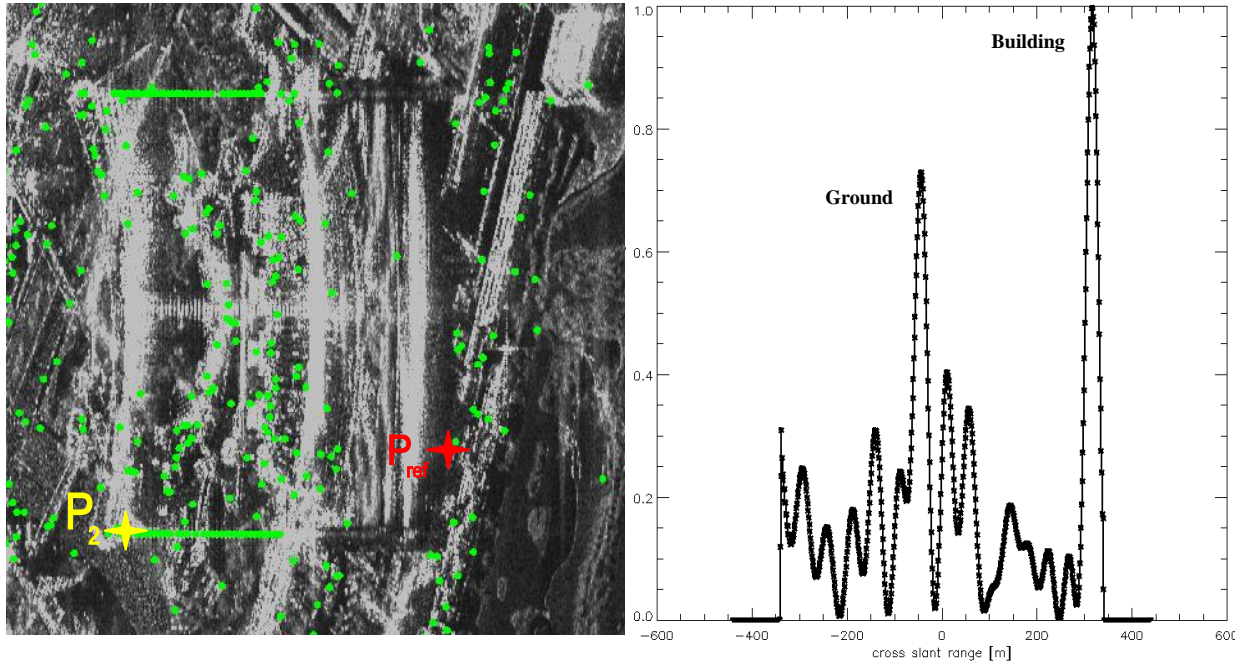


Figure 3.24: Example 2: SVD implemented on real data

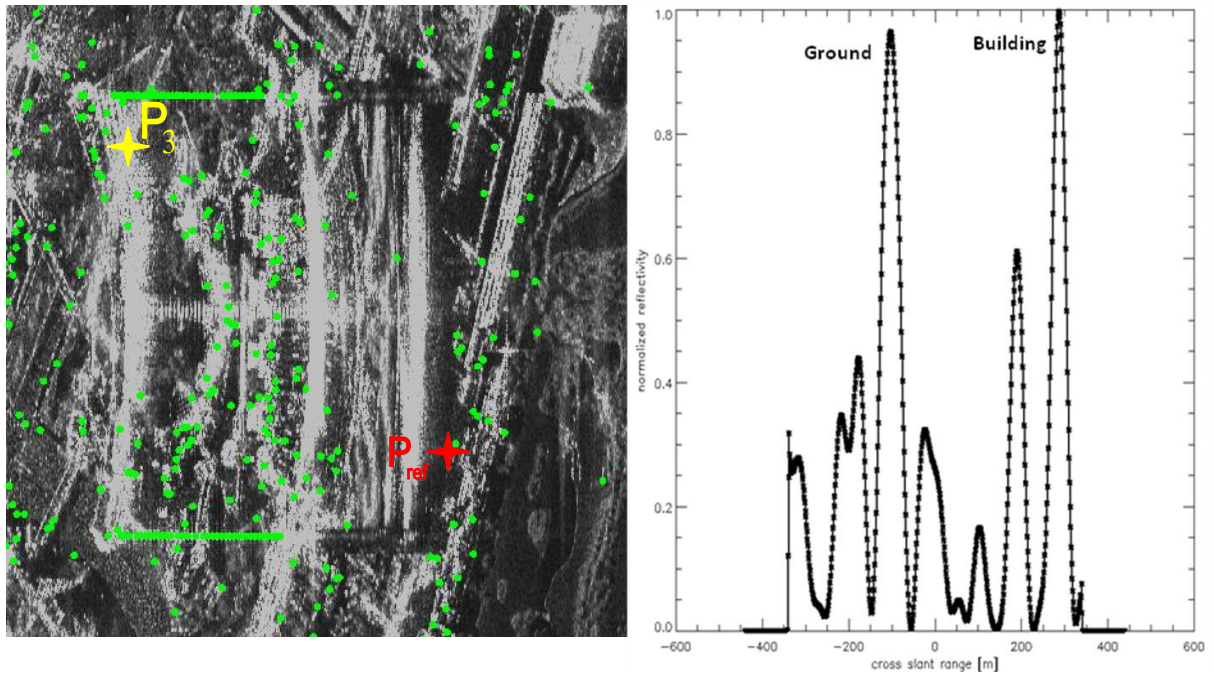


Figure 3.25: Example 3: SVD implemented on real data

3.5 Summary

- 1) For the purpose of performance investigation, the spectrum estimation strategies were applied to pure simulated data and semi simulated data.
- 2) The results for pure simulated data show that not only the number of samples but also the baseline distribution has a significant effect on the estimation result.
- 3) The performance of different strategies was mainly investigated based on the semi simulated data. As the SVD has an ill-conditioning problem under the influence of noise, regularization methods such as singular value truncation and Wiener filtering were implemented. Wiener filtering appears to be more stable and has better performance even though the resolution may be slightly reduced. NLS performs best for high SNR.
- 4) For the purpose of validation, spectrum estimation strategies including the SVD (singular value weighting using a Wiener filter) were evaluated on TerraSAR-X data and compared to the probable ground truth.

4 Model Selection

This Chapter gives a short introduction to the DLRs PSI-GENESIS processor and explains the necessity of model selection in SAR tomography for PSI techniques in Section 4.1. Several penalized likelihood model selection criteria such as Bayesian Information Criterion, Akaike information criterion, and Minimum Description Length criterion, which trade-off the accuracy and complexity of the model are explained in Section 4.2. Their performance is analyzed with simulated data with the historical baseline distribution of the TerraSAR-X satellite in Section 4.3. Finally, the model selection results with PS points of TerraSAR-X data are visualized in Google-Earth and the nature of PS pixel with multi-scatterers are discussed in Section 4.4. Section 4.5 summarizes the chapter.

4.1 Introduction to DLR PSI-GENESIS Processor

PSI-GENESIS is based on DLR's interferometry system GENESIS. The GENESIS system is optimized for operational DEM generation and has been used e.g. in the course of the Shuttle Radar Topography Mission (SRTM). The system is modular and consequently it could be easily updated for the Persistent Scatterer Interferometry (PSI) which allows the monitoring of subtle deformations of the Earth's surface. GENESIS and PSI-GENESIS are generic and support various SAR sensors e.g. ERS, ENVISAT/ASAR, RadarSAT, SIR-C/X-SAR, ALOS and TerraSAR-X.

A new challenge is the geometric complexity of radar mapping in urban areas. High rise buildings cause broad layover areas in the scene as a result of the high resolution. Figure 2.1 and 2.2 shows an example for the situation described above. Apparently, the radar returns from the ground and from the building interfere with each other. Classical interferometry can not resolve this ambiguity. Algorithms developed for PSI offer a solution ([Ferretti et. al, 2005](#)) ([Adam et. al, 2005](#)). The more natural solution is tomography ([Reigber, 2000](#)). Tomography which also is based on stacks of radar scenes offers a framework to separate the scatterers providing 3D scatterer locations and their reflectivity. Because the 3D localization of the estimates is essential the tomography principle needs to be considered in PSI processing. This is the reason several tomography algorithms are implemented and tested in the PSI-GENESIS system ([Adam et. al, 2008](#)).

For the purpose of application, instead of retrieving the reflectivity function along the elevation direction of a continuous volume scatterer, the case that several point scatterers are inside one resolution cell is more of interest. Persistent Scatterer Interferometry (PSI) is a revolutionary relative new technique which was introduced in the late 1990s for measuring ground displacements to mm level accuracy and over time periods previously unachievable

using conventional interferometry methods. It deals with persistent scatterers which are point scatterers which are stable during the whole temporal selection. The displacement rate is estimated according to the phase change of the same pixel which is measured with certain temporal baselines, perpendicular baselines and Doppler baselines. When the number of scatterers inside that resolution cell is not correct, the displacement estimation can be completely wrong. Thus, knowledge about the number of scatterers inside the resolution cell is very important. Tomography provides the possibility for model selection criteria to be applied to the estimated spectrum to separate and localize the PS in 3D space.

4.2 Model Selection Methods

4.2.1 Introduction to Model Selection

Model selection is the task of selecting a statistical model from a set of potential models, given data.

Model selection problems are encountered in many applications. In linear regression analysis, it is of interest to select the right number of nonzero regression parameters. With the smallest true model, statistical inferences can be carried out more efficiently. In the analysis of the time series, it is essential to know the true orders of an ARMA model. In problems of clustering, it is important to find the number of clusters. In signal detection, it is necessary to determine the number of true signals, and so on.

Taking model order determination as an example, the upper left figure shows experimental data to which we want to fit a harmonic polynomial. Now we come to the problem of what order of harmonic polynomial model should be used. If the order is too low, as showed in the upper right plot, the problem is under fitted and the residual of the model is too large. In contrast, when the order is too high, as showed in the lower right plot, the problem is over fitted. Even though the model fits the measurement very well, the model is too complicated. With the fourth order model, the result looks good. How can we decide which order of harmonic polynomial is the true one. This is a model selection problem.

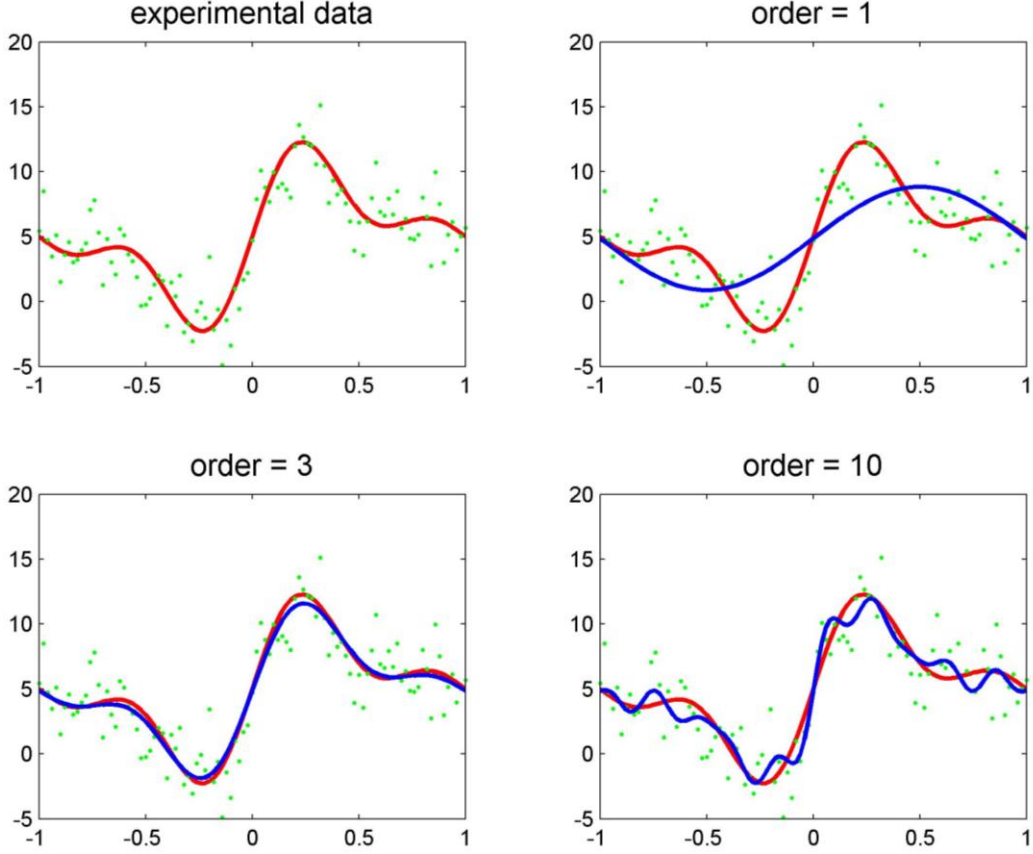


Figure 4.1: Example of harmonic polynomial order selection (red: true model; green: noisy measurements; blue: estimated model)

In this section, we survey the model selection discussed in statistical literature and mainly concentrate on those used in signal detection as detecting number of scatterers within one resolution cell is essentially a signal detection problem.

4.2.2 Model Selection Methods

Let k be a parameter which defines the complexity of the model which can be the number of parameters describing the signal, order of the model, number of signals, and so on. Let $\theta(k)$ be the unknown parameters or quantities. The relationship between $\theta(k)$ and the observed data y can be described by an observation model. The likelihood $p(y|\theta(k), k)$ will increase with increasing k , thus fitting the observations better. As a consequence, maximization of the likelihood function is useless for model selection. Instead of using only the likelihood as criteria, penalized likelihood criteria are used for model selection. The general form of penalized likelihood criteria is:

$$\hat{\theta}(k) = \arg \max_{\theta(k)} \{ \ln p(y|\theta(k), k) - C(\theta(k)) \} \quad , \quad (4-1)$$

In $p(y|\theta(k), k)$ is the likelihood and $C(\theta(k))$ is a complexity penalty, from which we can see that model selection is actually a trade-off between how good the model fits the data and the complexity of the model. A good model fits well to the observed data with small

description length. There are many types of penalized likelihood criteria, such as Bayesian Information Criterion (BIC), Cp criterion (Cp), Network Information Criterion (NIC), Akaike information criterion (AIC), subspace information criterion (SIC) and Minimum Description Length (MDL). Their principles are the same and the main difference comes from the penalty term. If this term only depends on the model dimension, then:

$$\hat{k} = \arg \max_k \{ \ln p(y | \hat{\theta}(k), k) - C(k) \} , \quad (4-2)$$

In other words, choose the best parameters for each k and then select among these models.

For the purpose of this thesis, Bayesian Information Criterion (BIC), Akaike information criterion (AIC) and Minimum Description Length (MDL) will be further discussed in this section.

▪ Bayesian Information Criterion (BIC)

The BIC is sometimes also named the Schwarz Criterion or Schwarz Information Criterion (SIC). It is so named because Gideon E. Schwarz (1978) gave a Bayesian argument for adopting it. The BIC is an asymptotic result derived under the assumptions that the data distribution is in the exponential family. The formula for BIC is:

$$\hat{k} = \arg \max_k \{ \ln p(y | \hat{\theta}(k), k) - \frac{k}{2} \ln n \} , \quad (4-3)$$

n refers to the number of samples. k is defined as above. The detailed derivation and performance of BIC is described in [\(Burnham, Kenneth P. und David R. Anderson, 2003\)](#).

For historical reasons, the BIC is defined by $\ln p(y | \hat{\theta}(k), k) - \frac{k}{2} \ln n$ multiplied by minus two:

$$\hat{k} = \arg \min_k \{ -2 \ln p(y | \hat{\theta}(k), k) + k \ln n \} , \quad (4-4)$$

Under the assumption that the model errors or disturbances are normally distributed, this becomes (up to an additive constant, which depends only on n and not on the model):

$$\hat{k} = \arg \min_k \{ n \ln \frac{RSS}{n} + k \ln n \} , \quad (4-5)$$

Where, RSS is the residual sum of squares from the estimated model. The following examples show how the penalized likelihood criterion works (BIC, AIC, MDL). The left image shows experimental data generated from a Gaussian mixture with 4 components. The left plot shows how the BIC value changes with the number of components. The preferred model is the one with the lowest BIC value.

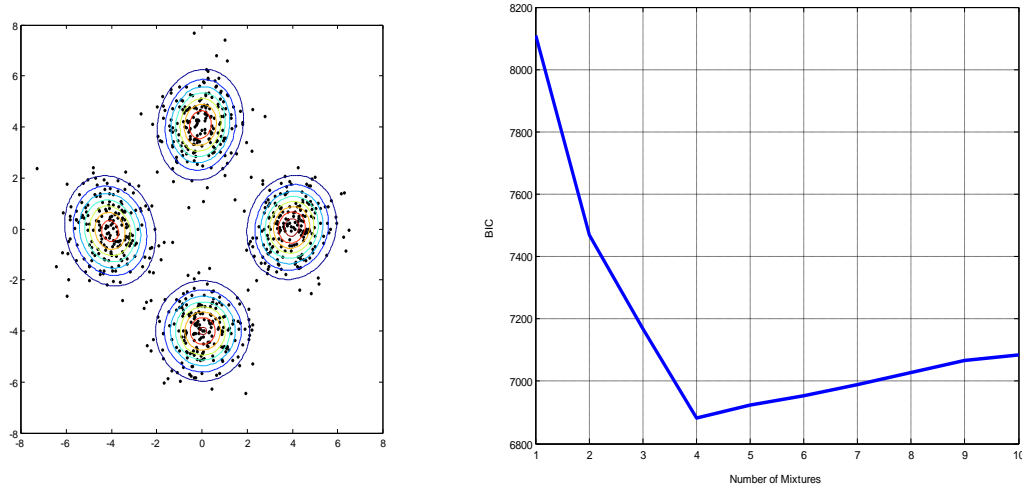


Figure 4.2: BIC application on model selection with 4 Gaussian Mixture Model
(Source: Joshua Broadwater, 2003)

BIC can measure the efficiency of the parameterized model in terms of predicting the data and penalizes the complexity of the model where complexity refers to the number of parameters in model.

▪ Akaike information criterion (AIC)

Akaike's (1973) seminal paper proposed the use of the Kullback-Leibler information or distance (information loss when the selected model is used to approximate the true model or the distance between the selected model and true model) as a fundamental basis for model selection. However, K-L distance cannot be computed without the complete information of the true model and the unknown parameters. Therefore, Akaike found a rigorous way to estimate K-L information, based on the maximum of the empirical log likelihood function. For the purpose of this thesis, a detailed derivation of the Akaike information criterion is given here. Books and papers on the derivation of AIC include Shibata (1983, 1989), Linhart and Zucchini (1986), Bozdogan (1987), and Sakamoto (1991). The formula for AIC is:

$$\hat{k} = \arg \min_k \{ -2 \ln p(y | \hat{\theta}(k), k) + 2k \} \quad , \quad (4-6)$$

The AIC methodology attempts to find the model that best explains the data with a minimum of free parameters. The AIC penalizes free parameters less strongly than the Schwarz criterion.

While Akaike derived an estimation of K-L information, AIC may perform poorly if there are too many parameters in relation to the size of the sample (Sugiura 1978, Sakamoto et al. 1986). Sugiura (1978) derived a second-order variant of AIC that he called c-AIC. Hurvich and Tsai (1989) further studied this small-sample bias adjustment, which led to a criterion that is called AICc,

$$AIC_c = AIC + \frac{2k(k+1)}{n-k-1} , \quad (4-7)$$

Unless the sample size is large with respect to the number of estimated parameters ($n/k < 40$), use of AIC_c is recommended.

▪ Minimum Description Length (MDL)

The minimum description length principle is a formalization of Occam's Razor in which the best hypothesis for a given set of data is the one that leads to the largest compression of the data. The MDL was introduced by Jorma Rissanen in 1978; it is an important concept in information and learning theory. "The fundamental idea behind the MDL Principle is that any regularity in a given set of data can be used to compress the data, i.e. to describe it using fewer symbols than needed to describe the data literally." ([Grünwald et al., 1998](#))

The MDL principle and Bayesian techniques are sometimes erroneously claimed to be the same. The broad MDL principle is sometimes also confused with an implementation of it as a particular and not so powerful model selection criterion BIC. The difference between the MDL and BIC is discussed in ([J. Rissanen, 2005](#)). If there is any Bayes principle it probably is associated with the paradigm prior-posterior probabilities in the Bayes' identity, or actually 'degrees of belief' since in the real case where none of the hypotheses is true the usual interpretation of probability such as 'the probability of the event that the hypothesis is true' is vacuous. Putting such religious sounding interpretations aside, the Bayes identity states:

$$Q(\theta | x) = \frac{P(x | \theta)w(\theta)}{\sum_{\eta} P(x | \eta)w(\eta)} , \quad (4-8)$$

where $w(\theta)$ is the prior probability of the hypothesis or model and θ is the parameter. $Q(\theta | x)$ the posterior, conditional on the data x . A fundamental objective in Bayesian statistics is to pick the hypothesis, labeled as $\bar{\theta}(x)$, which maximizes the posterior or the numerator, while the objective of the MDL principle is to maximize the code length for the data. It is true that we can encode the data with the 2-part code length

$$L(x, \bar{\theta}(x)) = -\log P(x | \bar{\theta}(x)) - \log w(\bar{\theta}(x)) , \quad (4-9)$$

given a class of models $\{P(\theta | x)\}$ and the prior $w(\theta)$, but this is not the shortest code length for the data. In fact, the following code length is shorter.

$$L(x) = -\log \frac{P(x | \bar{\theta}(x))}{P(\bar{\theta}(x))} - \log w(\bar{\theta}(x)) , \quad (4-10)$$

where,

$$P(\bar{\theta}(x)) = \sum_{y: \bar{\theta}(y)=\bar{\theta}(x)} P(y | \bar{\theta}(x))$$

Nevertheless, for the purpose of our application, once the penalized term is assumed to be the same, the MDL criterion is identical to the Bayesian Information Criterion which means Bayesian information is the shortest code as well. The MDL is then

$$\hat{k} = \arg \min_k \{ -\ln p(y | \hat{\theta}(k), k) + \frac{k}{2} \ln n \} , \quad (4-11)$$

The performance of the penalized likelihood criterions in this model selection problem are analyzed in Section 4.3.

4.3 Model Selection with SAR Tomography

In this section the focus is on the application of model selection criterion to semi-simulated data as explained in Chapter 3 which refers to the historical baseline distribution of TerraSAR-X satellite at the test site Las Vegas.

In the following, the residual sum of squares from the estimated model is computed in such a way that 1) the simulated observation are transformed to the frequency domain using the TSVD spectrum estimator as the red solid line showed in Figure 4.3, which is the normalized spectrum power of the signal spectrum. Here there are two scatterers within the resolution cell at positions -100m and 180.5m. With certain model assumption (number of scatterers), the elevation, amplitude and phase of the scatterers are estimated as the flow chart described in section 2.3.2.

The power spectrum estimated using the optimal model should then be a good fit to the true spectrum as showed in Figure 4.3. The blue solid line shows the power spectrum for different model orders. The upper plot compares the power spectrum of simulated noisy measurements and assuming a model order of 1 (1 scatterer). It is easy to conclude that the modeled power spectrum does not fit the data power spectrum at the position of -100m. The middle plot shows the same comparison under the assumption of 2 scatterers, with a significant improvement appearing. The lower plot is for an assumed model order of 3 (3 scatterers). The modeled power spectrum fits to the data power spectrum better but only slightly. Comparing the results with different model order assumptions, one would prefer to choose to believe 2 scatterers are present. This is a model selection process, using a model selection criterion one can determine the model order automatically.

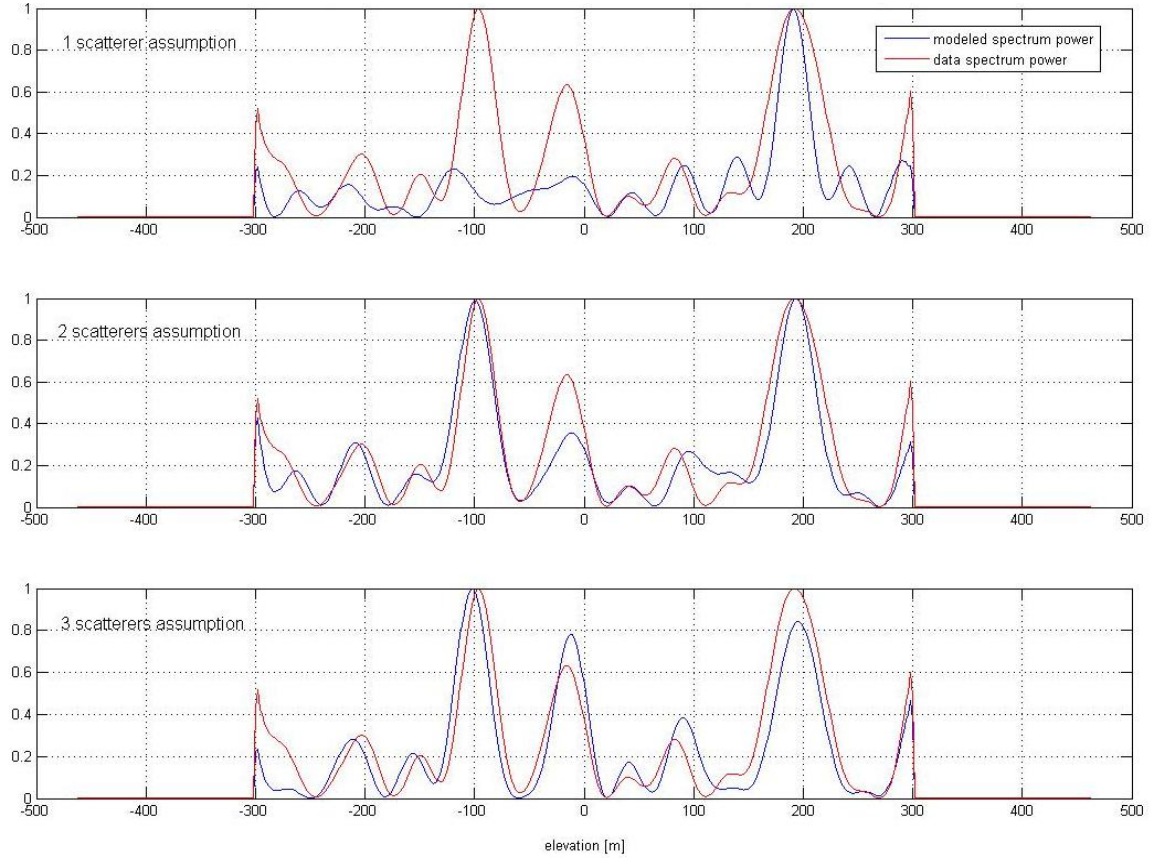


Figure 4.3: Example: two scatterers inside one resolution cell ($\text{SNR}=3.36\text{dB}$), the comparison between data spectrum power and modeled spectrum power under different model assumption.

One difficulty when using model selection criteria is how to compute the residual sum of squares. Because of the small datasets, irregular sampling and noise, the estimated result might be not reliable. Therefore, instead of evaluating the RSS in the data space, we shift to the spectrum domain. Based on the spectrum information, the RSS is computed by comparing the modeled spectrum and data spectrum. One can find similar approach in (Ferretti, 2005). Given the RSS, the information criteria can be computed. The model which minimizes the criterion is then an estimate of the true model order as shown in the flow chart in Figure 4.4.

It is necessary to assume that the number of scatterers inside one resolution cell is smaller than 4. For n scatterers, the model order $k=3n-1$ as for each scatterer the amplitude, elevation, and initial phase are the 3 unknowns which need to be estimated for complex data. Due to the normalization, one amplitude parameter is redundant.

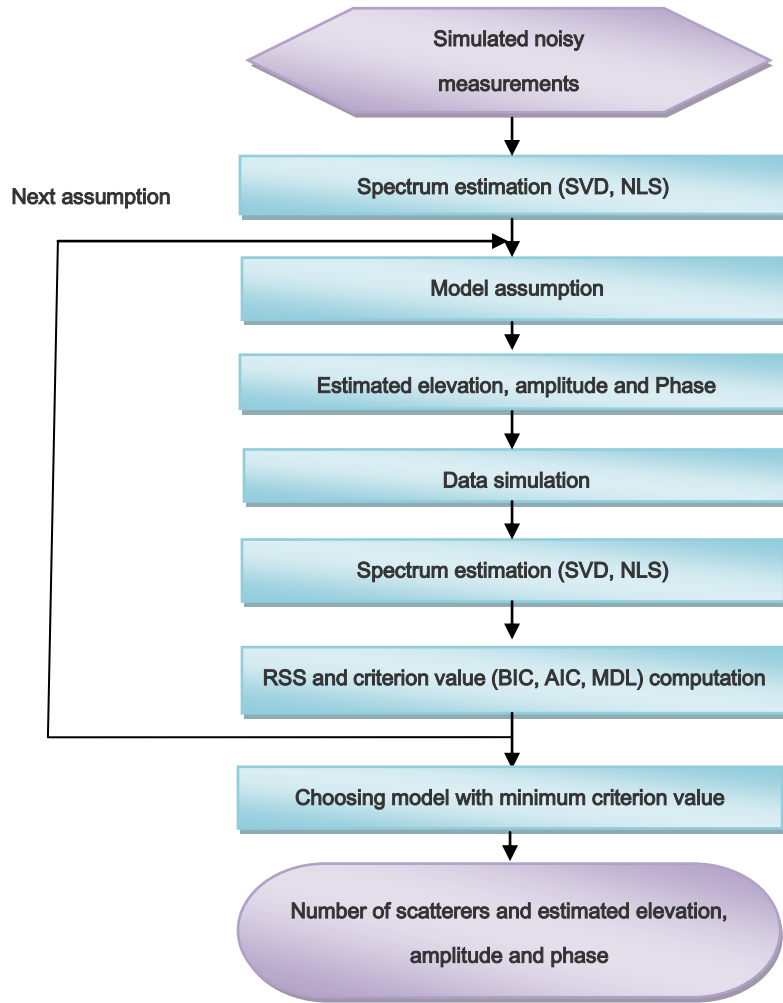


Figure 4.4: Flow chart of the model selection procedure for our problem

Figure 4.5 shows the criterion values varies with number of scatterers using different methods for the example showed in figure 4.3. All criterions suggest that there are two scatterers inside this pixel. In reality, the simulated truth is two scatterers at position -100m and 180.5m with amplitude 0.8 and 1. In this example, the selected model is fit to the true model. By comparing the criterion values for different models, we can see that Bayesian, MDL and AIC give similar penalty to the complexity of the model. AICc gives more penalties. As a general rule, Differences in criterion value (Δ) can be used to interpret strength of evidence for one model vs. another. A model with a Δ value within 1-2 of the best model has substantial support in the data, and should be considered along with the best model. A Δ value within 4-7 of the best model has considerably less support. A Δ value > 10 indicates that the worse model has virtually no support and can be omitted from further consideration.

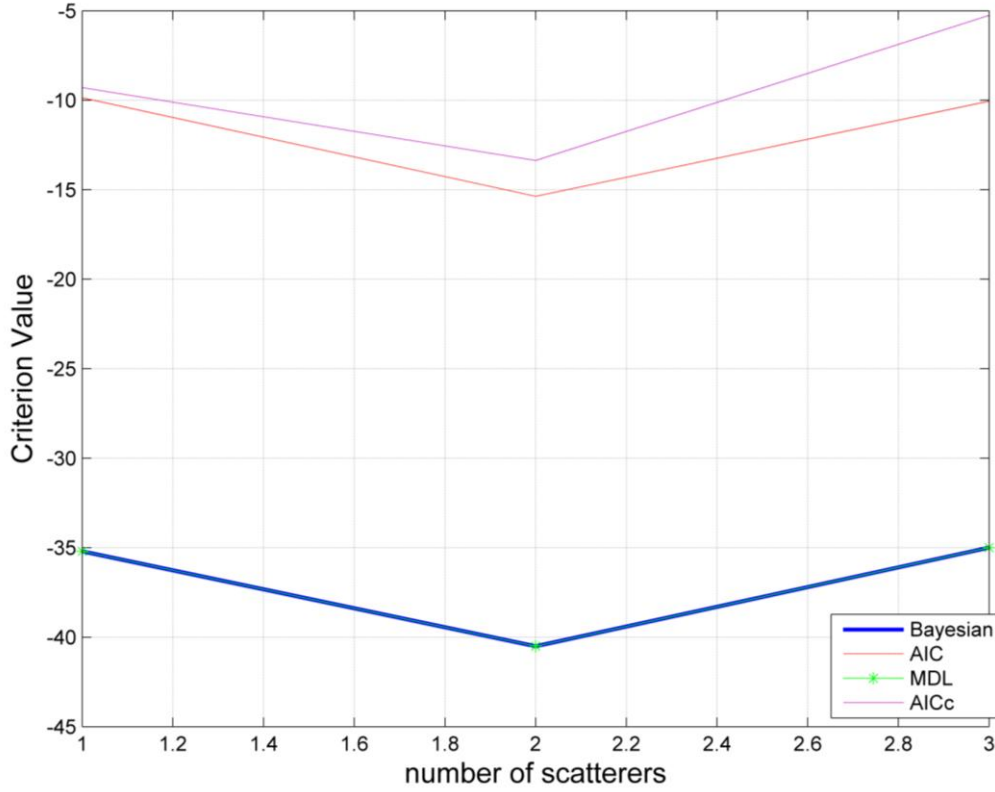


Figure 4.5: Criterion values for different model selection schemes

There is always the risk of selecting the wrong model, whether the model is selected manually and automatically. The performance of different model selection criteria with respect to detection rates and false alarm rates is discussed next.

▪ Detection Rates

An important characteristic for evaluating the performance of the model selection criteria is the detection rate which here refers to the probability of correctly detecting the number of scatterers. A Monte Carlo simulation with 1000 realisations was used to evaluate the detection rates of the above described model selection procedures.

Figure 4.6 shows the detection rate when only one scatterer is present. In this case, the results are quite good, and the detection rates are larger than 70% even with an SNR of 2dB. When the SNR is increased to 7dB, all methods have detection rates larger than 90%. On the other hand, the task is relatively easier when only one scatterer is present as there are no inference effects between scatterers. As the BIC more strongly penalises the complexity of the model compared to the AIC, it is more likely to select a lower model order from theoretical point of view. This is true in our experiments. The BIC (blue solid line) shows a higher detection rate. However, since the sampling size is very small, there are no big differences between them. As explained in Section 4.2, the MDL (green solid line with star) and BIC are identical in this application. The AICc (purple dashed line) penalizes the complexity of the model much more than the AIC (red dashed line). Therefore, it shows a much better performance when

the true model is very simple. However, when the true model becomes more and more complex, the criterion may not find a good fitting model.

The lower plot of Figure 4.6 shows the consistency of the different methods which means the probability that all methods have the same and correct outcome. From the plot, we can make the same comments as above - that these methods are all very similar to each other as the consistency is above 70%. This is easy to understand as all methods try to give a weight to the precision and complexity of the model and get an optimum.

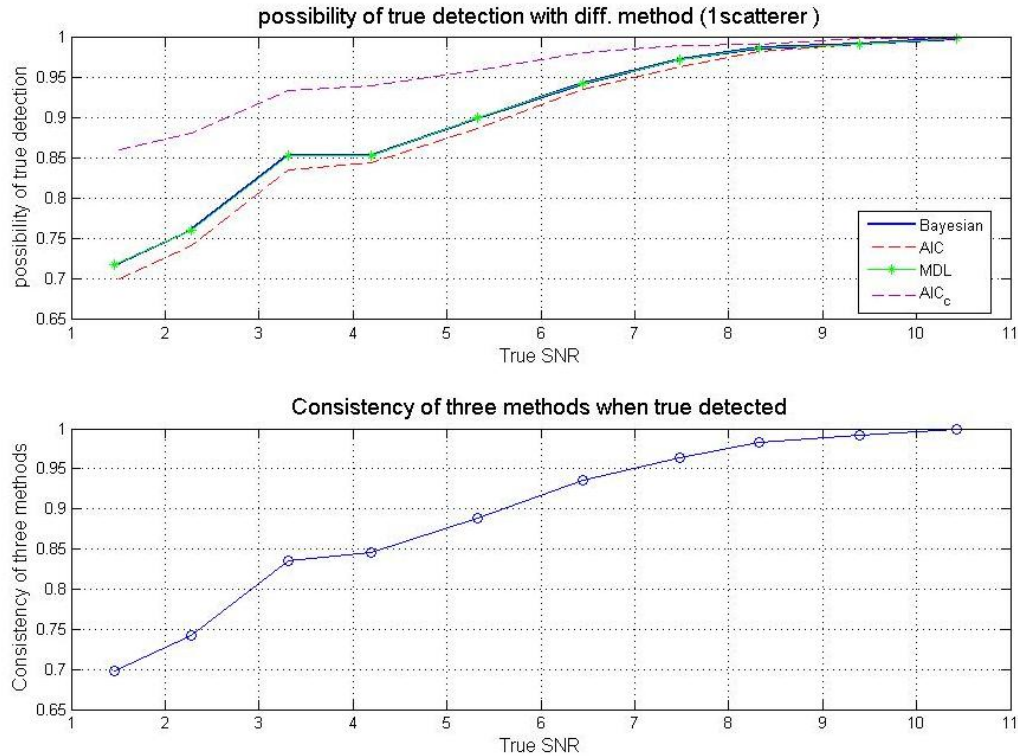


Figure 4.6: Detection possibility and consistency of different methods when only 1 scatterer is included

Figure 4.7 shows the detection rate for the case of two scatterers. As the penalty term of the AICc is too strong, the AICc is biased towards choosing a lower model order – and hence has a lower detection rate - when compared to the other criteria. However, the performance is still acceptable with a detection rate of about 60% at an SNR of 1dB. 2dB is the critical SNR for real data from TerraSAR-X. Accordingly, with AICc, the detection rates may reach 67%. The detection rates of the AIC, MDL and BIC are similar independent of the SNR. The AIC performs a little better when the SNR is very low in which case the RSS is very large. At the critical SNR of 2dB, the detection rate is 75% which is quite good. The consistency plot is again evidence for the arguments above.

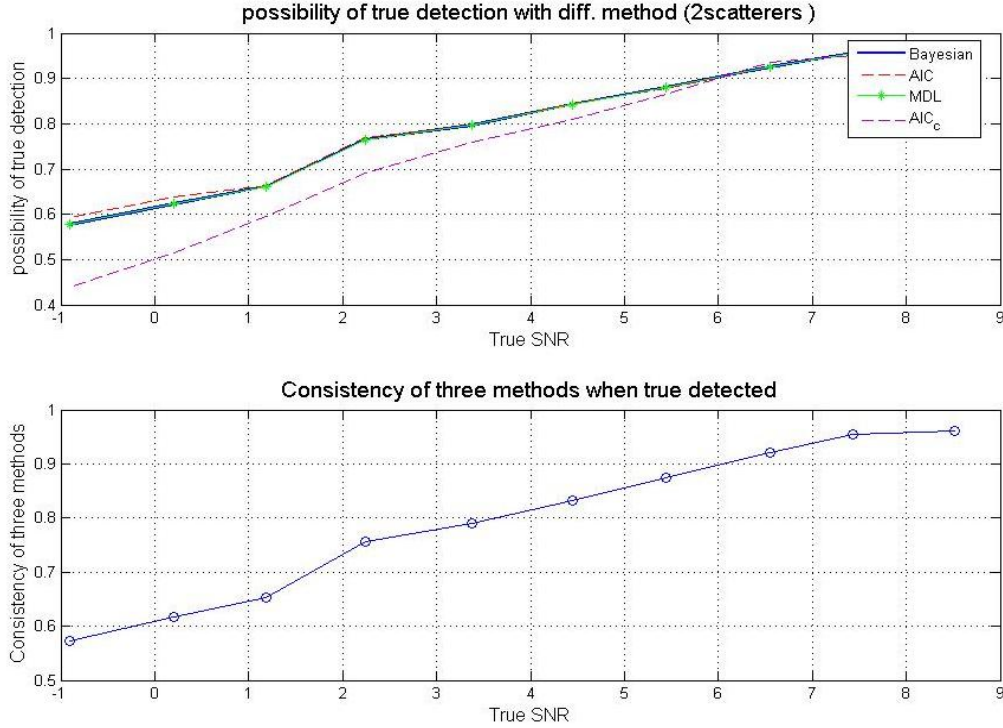


Figure 4.7: Detection possibility and consistency of different methods when only 2 scatterers is included

▪ Detection accuracy

For the model selection problem, not only the number of scatterers is of interest but also the elevation of the scatterers within one resolution cell as the final task is actually estimating the position of the scatterers and their reflectivity in the elevation direction. Therefore, how accurate the elevation of the scatterers can be estimated is also an important factor. The following figure shows a histogram of elevation estimates over 1000 Monte Carlo realisations. For the noisy measurements, two scatterers within one resolution cell, which are separately at the position of -100 meter and 180.5 meter, are simulated. The elevation estimation histogram is computed from experiments where the correct number of scatterers was detected. In this plot, what we are interested in is how accurate the elevation position can be estimated and the differences between different methods as well. Most elevation estimates are close to the true value and biased within an uncertain region, however, there are also some false estimates. For example, the histogram shows a small peak around zero elevation position which is false. Thus, it is very important to evaluate the estimation accuracy for the purpose of performance evaluation. As the basic idea behind these model selection criteria is very similar, the difference between them is too slight to see from the histogram. From the detection rate point of view, the BIC, MDL and AIC give similar results.

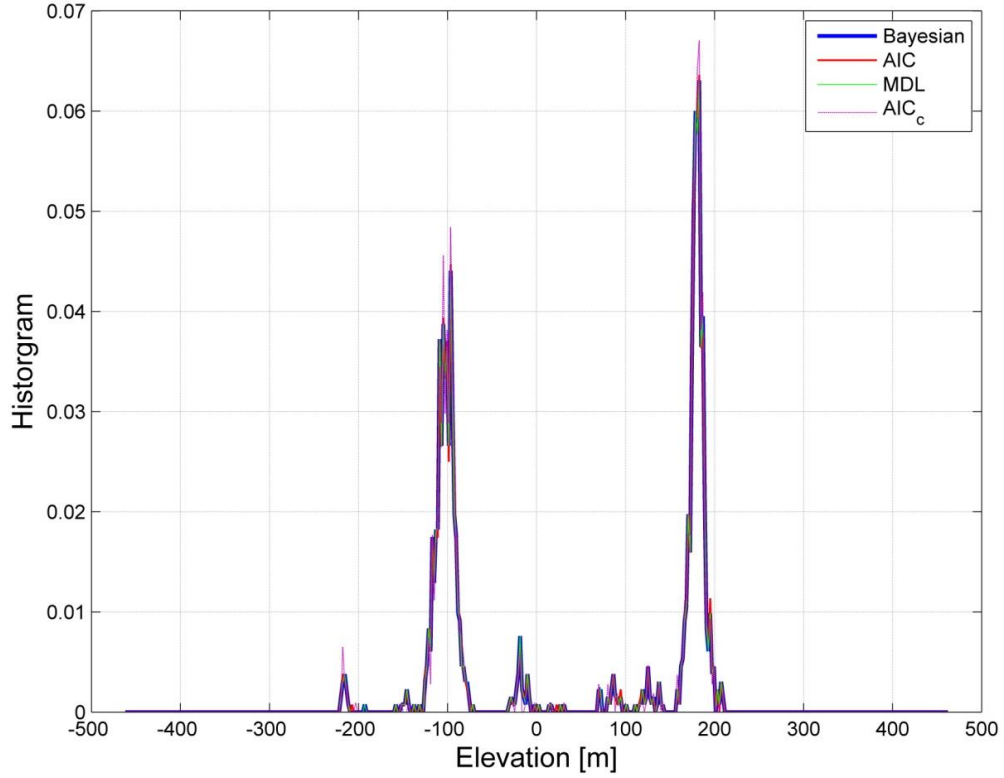


Figure 4.8: Histogram of estimated elevation with $SNR = 0.47dB$, the true elevation is -100 meter and 180.5 meter when the amplitude is 0.8 and 1. The corresponding detection possibility is: $Pd_Bayesian = 0.65$, $Pd_AIC = 0.66$, $Pd_MDL = 0.658$ and $Pd_AICc = 0.54$

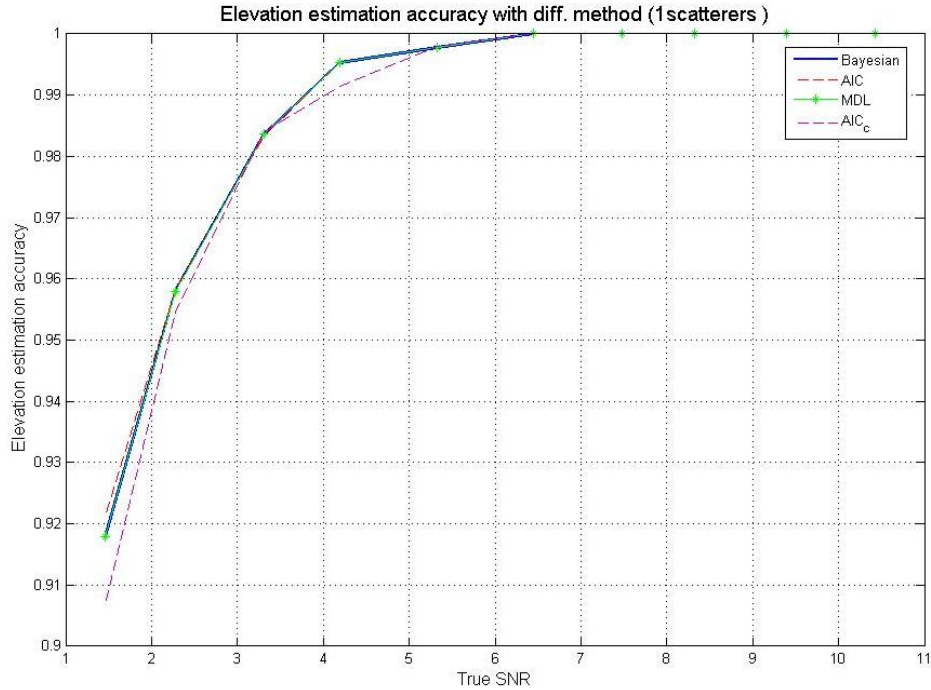


Figure 4.9: Elevation estimation accuracy with different method (1 scatterer at the position 100 meter are included)

Figure 4.9 shows the detection accuracy for the case of 1 scatterer. As specified in Chapter 3, the minimum distance that two scatterers can be distinguished is 45.46 meter. Thus, the

elevation of the scatterer is assumed to be accurately estimated when the estimated elevation is within this 45.4589 meter resolution cell. Based on this definition, the elevation estimation accuracy for 1 scatterer is higher than 93% for SNR=2dB for all methods. The estimation accuracy is close to 1 when the SNR is larger than 6 dB. However, the case of only 1 scatterer is not of so much interest as there is no inference between signals from different scatterers.

The estimation accuracy in the case of 2 scatterers is been illustrated in Figure 4.10. A correct estimation is assumed to occur only when the elevation of both scatterers is within the 45 meter resolution cell. As a result of the inference, the estimation accuracy is reduced slightly compared to Figure 4.10. Taking the case SNR=2dB with a single scatterer as an example, the probability that the elevation is correctly estimated is about 94%, for the case of two scatterers it is close to 90% for the BIC, MDL and AIC and 93% for the AICc. AICc shows higher accuracy as it most strongly penalizes the complexity of the model, and only when the two scatterer hypothesis is much more likely than the single scatterer hypothesis will two scatterers be detected. It is worth mentioning that whether the peak can be correctly detected or not also depends on the noise behavior. As mentioned above, the number of samples is too small to obey some statistical distribution. By means of the average SNR value out of thousand times experiments, the noise level may be consistent with the theoretical one (simulation input). However, for a single experiment, it is possible that the SNR value is much lower than the mean SNR in which case the signal may be distorted dramatically by higher level noise. Nevertheless, the estimation accuracy is quite good (0.93) even under the critical condition (SNR=2dB).

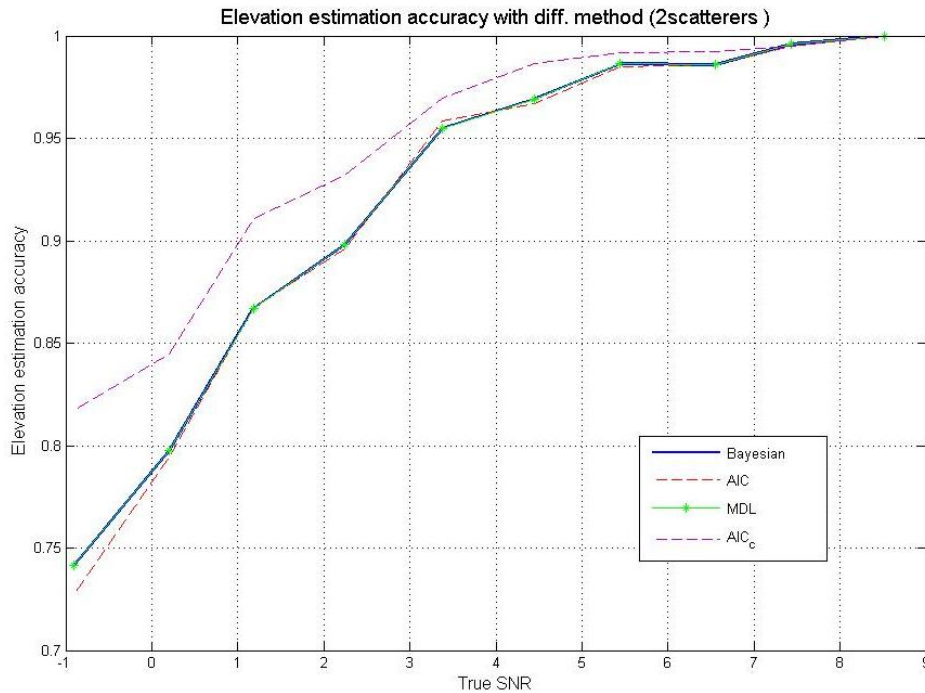


Figure 4.10: Elevation estimation accuracy with different method (2 scatterer at the position 100 meter and 180.5 meter are included)

In this section, model order selection schemes were implemented and tested on simulated data. Based on the analysis above, we conclude that model selection with SAR tomography shows high detection rate and elevation estimation accuracy for both single and multiple scatterers.

4.4 Result with TerraSAR-X Data

For the purpose of validation, the model selection criteria are evaluated on TerraSAR-X data. Figure 4.11 presents the result from 12 acquisitions (Three more acquisitions were obtained. Unfortunately the baseline range is not extended which means the 45 meter resolution remains). As mentioned above, the reference points are selected to penetrate the atmospheric effect. Conjugate multiplication in measurement space is equivalent to correlation in the spectral domain. Therefore, the noise level of the reference point has a dramatic influence on the performance of the spectrum estimator and thereby has a significant effect on model selection result as well. As a consequence, the yellow point which is supposed to be one strong scatterer on the ground is selected to be a reference point. With the reference point, the model selection criterion AIC is implemented on the PS points and the number of scatterers is retrieved and visualized. In Figure 4.11, the colour points visualized on gray image are PS points. Green points denote pixels with only one scatterer inside one resolution cell. Red points denote pixels with two scatterers and blue points three scatterers.

From the retrieved number of scatterers inside the specific resolution cell, most selected PS pixels contain only one scatterer. Pixels with two or more scatterers mainly appear in the area marked with the purple circles. From the texture in the SAR image, the marked region is the layover area of the higher part of the hotel and other lower structures beside the road. The strong scatterer on the ground and building are mapped onto one resolution cell. In this area, the model selection result fits to the ground truth, but as a result of high sidelobe caused by small number of samples, there are also several PS pixels referring to scatterers on small building which turn out to have three scatterers which is impossible due to the 45 meter resolution. This can also be caused by less a priori information about the extent of the SAR image which was selected according to the height of the Wynn Las Vegas. With help of Digital Elevation Models which do not need to be very precise, the estimation results are expected to be better.

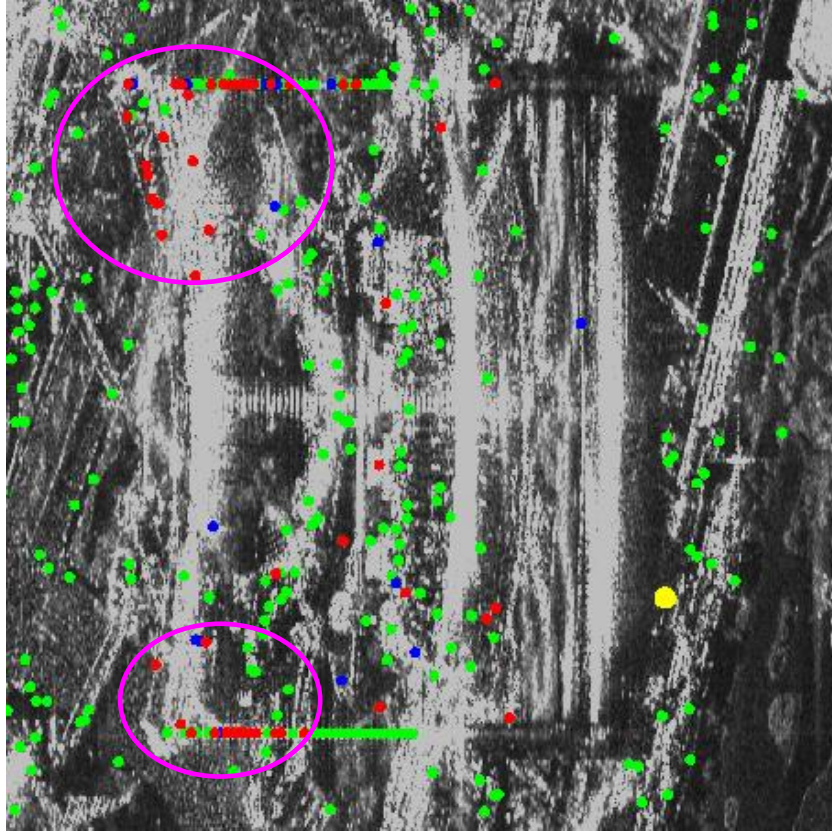


Figure 4.11: Model selection result using Akaike Information Criterion

In order to compare with ground truth, the estimated elevations of the scatterers inside each resolution cell of PS points are visualized in Google-earth. Figure 4.12 is the visualization of PS pixels with only one scatterer inside. Due to the Geo-coding, there is a slight bias between the visualized PS points and the 3D model of Google-earth. Nevertheless, we still can see the frame of the hotel especially the two sides of the hotel which have high PS density. As marked by circles, the retrieved height of the lower part of the building is excellent. The elevation estimation of points near ground is also quite stable which we can see from the localization of scatterers in front of the hotel. Within the area of the yellow circle appear several scatterers with similar height to the Wynn. This is probably due to the overlay area being wrongly geo-coded during PS processing which means the scatterer is actually located on top of the building.

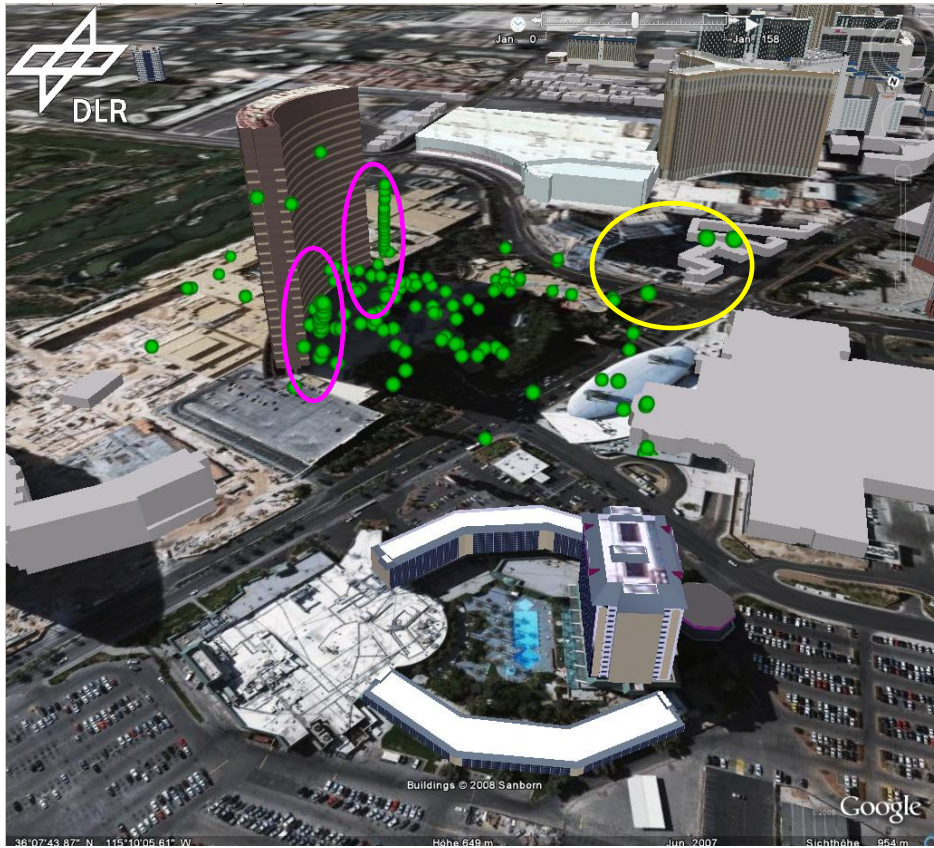


Figure 4.12: Visualization of PS pixels with only one scatterer per cell.

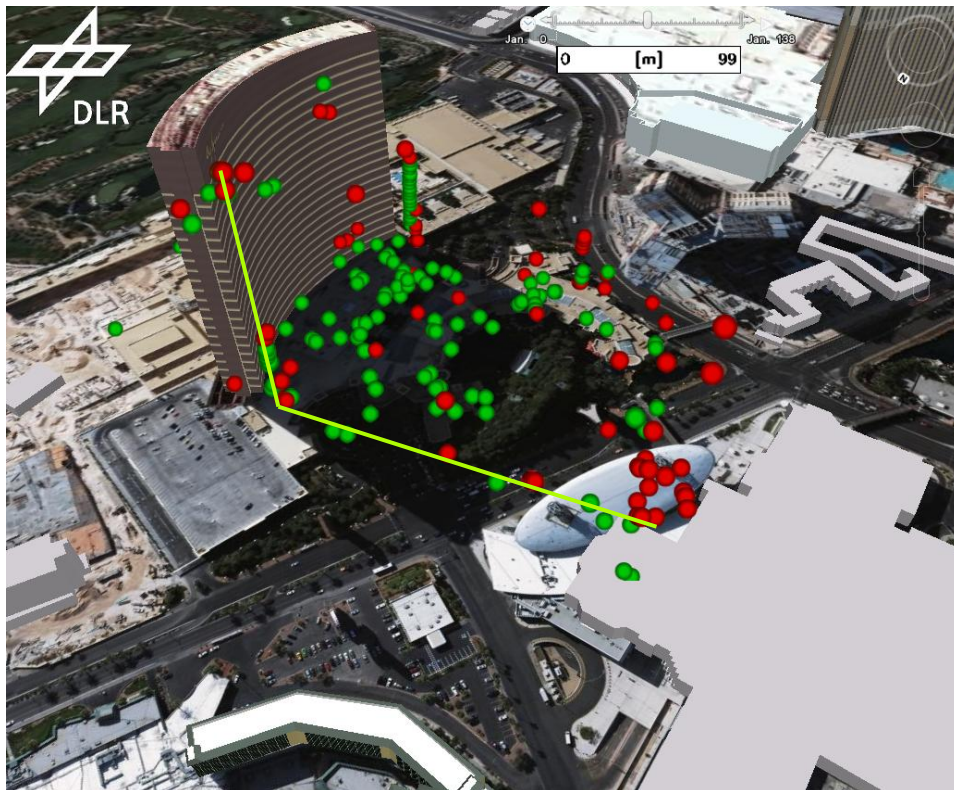


Figure 4.13: Visualization of PS pixels with one and two scatterers per cell.

The red points in Figure 4.13 are the visualized scatterers when the model two scatterers per cell are selected. As we expected, two scatters mapped into one resolution cell are

mainly caused by layover of the higher building and lower structures near the ground. However, as geo-coding of PS are based on the height estimation after the assumption that only one scatterer is present in one resolution cell, those pixels with two scatterers are either located at the building when the scatterer on the building is much stronger or located at the yellow line on the ground. In order to verify the position of the scatterers precisely, a second geo-coding procedure is required. However, as we can only retrieve the height difference between the two scatterers instead of the exact height, for this purpose, we need either additional information such as digital surface model or new algorithm to retrieve the height of both scatterers by combining the information from PS processing and Tomography.

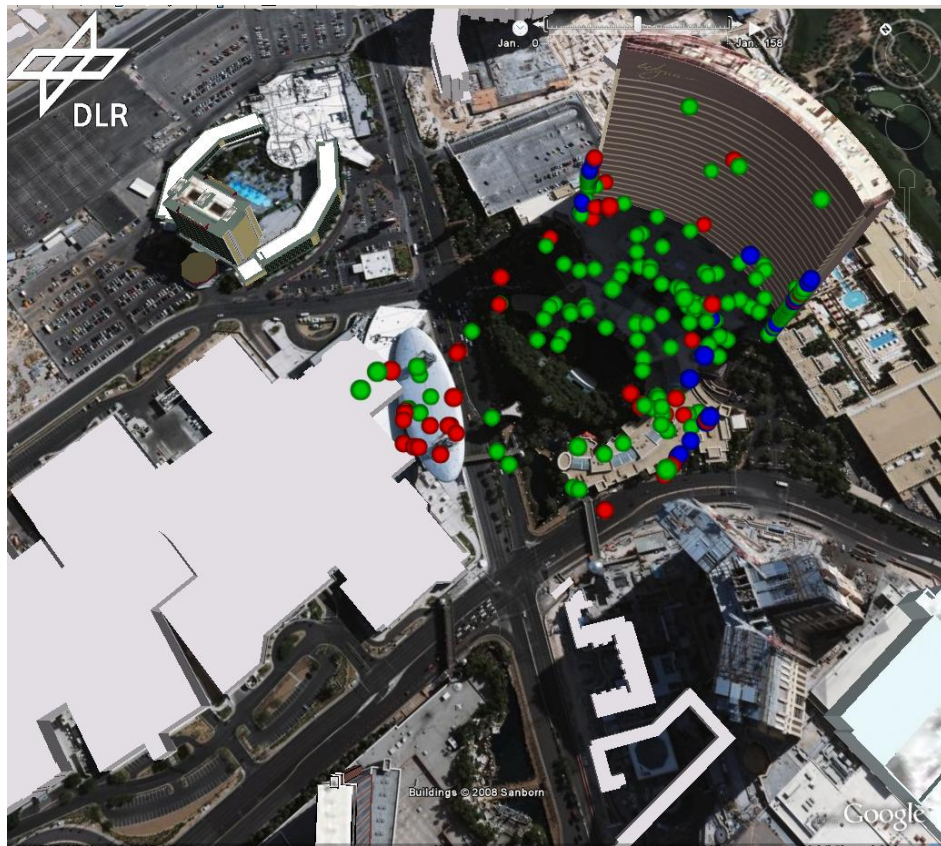


Figure 4.14: Visualization of PS pixels with one, two and three scatterers per cell

Figure 4.14 includes three scatterers per cell as well. The location of the pixels with three scatterers in the SAR image looks reasonable. However, the verification of the result is much more difficult.

4.5 Summary

- 1) The model order selection criteria which try to find a balance between how good the model fits to the data and the complexity of the model were described. With the aim of retrieving the true number of scatterers within one specific resolution cell, the model selection criteria were applied to the spectrum estimation results from simulated data for the purpose of performance evaluation. Good detection rates and elevation estimation accuracy were obtained.

- 2) Combining model selection criteria and spectrum estimation and applying them to real data was successful and reasonable results were obtained for the case of irregular sampling and small datasets. However, there is still much space for further improvement.

5 Concluding Remarks

5.1 Discussion and Conclusion

Based on the theoretical formulation of spectrum estimation problem for SAR tomography and its application on model selection, the simulation results, and results from TerraSAR-X high resolution data, we can draw the conclusion as the following:

- 1) SAR tomography, as a spectrum estimation problem, is capable to distinguish multi-scatterers within one resolution cell. The observations within the same resolution cell obtained by SAR sensor at different satellite positions is actually sampled FT of the reflectivity function along the elevation direction. Baseline distributions, which determine the frequencies where the FT of the reflectivity function are sampled, has a significant effect on the estimation results when a certain number of samples are available.
- 2) With the difficulties that only few and irregularly sampled datasets are available, the performance of spectrum strategies such as Singular Value Decomposition (SVD) and Nonlinear Least Squares estimation (NLS) were investigated based on the simulated data using baseline distribution of real data. With different noise level, spectrum estimation strategies show different performance. SVD has its particular ill condition problem due to noise propagation. Therefore, regularization tools, singular values truncation and Wiener filter, are implemented. Due to the small number of samples, truncated SVD has unavoidable information lose. By contrast, Wiener filter as an optimal filter shows a more stable and better performance even though the resolution might be slightly reduce due to the smoothing effect. As a deterministic method, NLS shows perfect performance under low noise level. However, it needs a priori knowledge about the number of scatterers and is time consuming since multi dimensional search is required.
- 3) For the purpose of validation, SVD, with singular values weighting by Wiener filter, is implemented to high resolution TerraSAR-X data at test site Las Vegas and compare to probable ground truth and shows reasonable results.
- 4) With the aim of application, model selection criteria, such as Bayesian Information Criterion (BIC), Akaike Information Criterion (AIC) and Minimum Description Length criterion (MDL), were implemented to the spectrum estimation result of SAR tomography to determine the number of scatterers within one resolution cell which is the important prior knowledge of precise PSI processing. By examining the detection possibility and detection accuracy under one and two scatterers simulation conditions,

this application are considered to be very successful with different noise levels. The model selection criteria are also applied to real data with 12 acquisitions, a reasonable result are obtained.

5.2 Outlook

After the investigation of this master thesis, more possibilities of extension of this work in the future would be outlined as follows:

- 1) Instead of implementing the discussed strategies on particular structure with choosing one reference point for the sake of penetrating the atmospheric effect, a reference network can be built for large-scale processing.
- 2) For further development and perfecting, Spectrum estimation strategies can be developed with help of Digital Elevation Model (DEM) as it can be used as a priori knowledge the extent of SAR image which may improve the elevation estimation stability and accuracy. It can be also treated as a reference to validate the estimate result.
- 3) The potential of more precise earth surface displacement might be realized by combining PSI techniques and SAR tomography which provides the possibility of 4D Space-Time Monitoring of Geodynamic Processes.

Terminology and Abbreviation

SAR	Synthetic Aperture Radar
InSAR	Interferometric Synthetic Aperture Radar
D-InSAR	Differential Interferometric Synthetic Aperture Radar
PSI	Persistent Scatterer Interferometry
TomoSAR	Synthetic Aperture Radar Tomography
CAT	Computed Axial Tomography
SLC	Single Look Complex
DEM	Digital Elevation Model
PSF	Point Spread Function
PTA	Point Target Analysis
SVD	Singular Value Decomposition
NLS	Nonlinear Least Squares
DFT	Discrete Fourier Transfer
SNR	Signal to Noise Ratio
RSS	Residual Sum of Square
ARMA	Auto Regressive Moving Average
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
SIC	Schwarz Information Criterion
MDL	Minimum Description Length
SIC	Subspace Information Criterion

References

- Adam, N., Bamler, R., Eineder, M., Kampes, B.: *Parametric estimation and model selection based on amplitude-only data in PS-interferometry*, in: Proc. of FRINGE 2005, Frascati, 2005
- Adam, N., Parizzi, A., Eineder, M., Corsetto, M.: *Persistent Scatterer Processing Validation In the Course of the Terrafirma Project*, in: Proc. FRINGE 2007 (ESA, Frascati), 2007
- Adam, N., Eineder M., Yague-Martinez N., Bamler R.: *High Resolution Interferometric Stacking with TerraSAR-X*, in: Proc. Int. Geosci. Rem. Sens. Symp. (IGARSS), Boston, 2008
- Bamler, R., Hartl, P.: *Synthetic Aperture Radar Interferometry*, Inverse Problems, vol.14, pp. R1-R54, 1998
- Bozdogan, H.: *Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions*. Psychometrika 52, 345-370, 1987
- Ferretti, A., Prati, C., Rocca, F.: *Permanent Scatterers in SAR Interferometry*, IEEE Transactions on Geoscience and Remote Sensing, vol. 39, No. 1, 2001
- Ferretti, A., Bianchi, M., Prati, C., and Rocca F.: *Higher-order permanent scatterers analysis*, EURASIP J. Appl. Signal Process., vol. 20, no. 20, pp. 3231–3242, 2005
- Fornaro, G., Serafino, F., and Soldovieri, F.: *Three-dimensional focusing with multipass SAR data*, IEEE Trans. Geosci. Remote Sens., vol. 41, no. 3, pp. 507–517, Mar. 2003
- Fornaro, G., and Serafino, F.: *Imaging Single and Double Scatterers in Urban Areas via SAR Tomography*, IEEE Trans. Geosci. and Remote Sens, Volume , Issue , 11-13 April 2007 Page(s):1 - 7
- Fornaro, G., Franceschetti, G., Perna, S.: *On center-beam approximation in SAR motion compensation*, Geoscience and Remote Sensing Letters, IEEE, 2006
- Fornaro, G., Serafino, F.: *Imaging of Single and Double Scatterers in Urban Areas via SAR Tomography*. IEEE Trans. Geosci. Remote Sensing, Vol. 44, no. 12, 2006
- Gabriel, A.K., Goldstein, R.M., Zebker, H.A.: *Mapping Small Elevation Changes over large Areas: Differential Radar Interferometry*, J. Geophys. Res. 94, pp. 9183-1919, 1989
- Gini, F. Lombardini, F.: *Multilook APES for multibaseline SAR interferometry*. Signal Processing, IEEE Transactions, 2002
- Grünwald, P., Kontkanen, P., Myllymäki, P., Silander, T. and Tirri, H.: *Minimum encoding approaches for predictive modeling*. In Proceedings of the Fourteenth International Conference on Uncertainty in Artificial Intelligence (UAI'98), pages 183-192. Madison, WI, USA, July 1998
- Grünwald, P., Pitt, M. A. and Myung, I. J. (eds.): *Advances in Minimum Description Length: Theory and Applications*, MIT Press, ISBN0-262-07262-9, April 2005
- Hanssen, R.: *Radar Interferometry: Data Interpretation and Error Analysis*, Kluwer Academic Publishers, Vol. 2, First Edition, 2001
- Hansen, P.C.: *Regularization Tools* [EB/OL], (2001-09) [2005-07-19]. <http://www.imm.dtu.dk/~pch/Regutools/>
- Homer, J., Longstaff, I. D., She, Z., and Gray, D.: *High resolution 3D imaging via multipass SAR*. IEE Proc., Radar Sonar Navig. 2002

- Kampes, B.: Radar Interferometry: *The Persistent Scatterer Technique, Remote Sensing and Digital Image Processing Series*, First Edition, Springer Verlag, New York, 2006
- Kay, S. M.: *Fundamentals of statistical signal processing*, PTR Prentice-Hall Englewood Cliffs, NJ, 1993
- Kenneth, P. Burnham and David, R. Anderson: *Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach* (Kindle Edition), Springer-Verlag, New York, 2003
- Linhart, H., Zucchini, W.: *Model selection*, New York: Wiley, 1986
- Lombardini F.: Differential Tomography: *A New Framework for SAR Interferometry*, in Proc. IGARSS, Toulouse, France, Jul. 2003, pp.1206–1208
- Massonnet, D., Rossi, M., Carmona, C., Adragna, F., Peltzer, G., Feigl, K., Rabaute, T.: *The displacement field of the Landers earthquake mapped by radar interferometry*, Nature 364, 138 - 142 (08 July 1993)
- Nariaki, S.: *Further analysts of the data by akaike' s information criterion and the finite corrections*, *Communications in Statistics - Theory and Methods*, Volume 7 , Issue 1, pages 13 – 26, 1978
- Pasquali, P. Pellegrini, R. Prati, C. Rocca, F.: *Combination of interferograms from ascending and descending orbits*, Geoscience and Remote Sensing Symposium, 1994
- Reigber, A., Moreira, A.: *First Demonstration of Airborne SAR Tomography Using Multibaseline L-Band Data*, GeoRS(38), No. 5, September 2000, pp. 2142-2152
- Reigber, A., Moreira, A.: *First demonstration of airborne SAR tomography using multibaseline L-band data*, IEEE Trans. Geosci. Remote Sensing, vol. 38, no. 5, pp. 2142-2152, 2000
- Rissanen, J.: *A short introduction to Model Selection*, Kolmogorov Complexity and Minimum Description Length, Available online at www.mdl-research.org, 2005
- Rissanen, J.: *Complexity and Information in Modeling*, Chapter IV in the book Computability, Complexity and Constructivity in Economic Analysis, (Editor: K Vela Velupillai ISBN 1405130784), Blackwell Publishings, Oxford, 2005
- Sakamoto, Y.: *Categorical Data Analysis by AIC*, KTK Scientific Publishers, Kluwer Academic Publishers, 1991
- Sakamoto Y., M. Ishiguro, G. Kitagawa: *Akaike information criterion statistics Reidel*, Dordrecht, The Netherlands, 1986
- Schwarz, G.: *Estimating the dimension of a model*. Annals of Statistics 6(2):461-464, 1978
- She Z., Gray D.A., Bogner R.E., Homer J.: *Three-dimensional space-borne synthetic aperture radar (SAR) imaging with multiple pass processing*. ID Longstaff - International Journal of Remote Sensing, 2002
- Shibata, R.: *A theoretical view of the use of AIC*, Pages 237-244 in O. D. Anderson (editor). Time series analysis: theory and practice. Elsevier Scientific Publ., North-Holland, 1983
- Shibata, R.: *Statistical aspects of model selection*. Pages 215-240 in J. C. Willems.(editor) From data to model. Springer-Verlag. London, 1989
- William, H. P., Saul, T. A., William, Vetterling, T. and Brian, P. F.: *Numerical Recipes in C*, (2nd Edition), New York, Press Syndicate of University of Cambridge, 1992
- Webpage: http://en.wikipedia.org/wiki/Model_selection

Acknowledgements

First of all, I would like to express my warm gratitude to my supervisor, Prof. Bamler, not only for his wonderful supervision and discussions but also for his great help during my master study. He was always there to listen and to give advice and showed me different ways to approach a research problem. He provides me the great opportunity for my further study as well. During the past two years, I admire him very much not only for his profound scientific knowledge, but also for his common touch and great patience to others. It is my prestige to be one of his students.

Second, I would like to express my gratitude to Nico Adam, who is most responsible for providing me technique support of this thesis. He taught me how to solve a technique problem, made me a better programmer, had confidence in me when I doubted myself, and brought out the good ideas in me. Special thanks go to Prof. Hinz, who is most responsible for helping me amend the writing of this thesis. He gave me great support during my study for a lot of things and makes me convince that being one of LMF is my wise choice and great honour. Thanks very much to Dr. Brcic; Ramon has been a friend and mentor. He always gives me good suggestions and provides me great support to correct my thesis.

Besides my advisors, I would like to thank Prof. Meng, thank her for recommending me to study in TUM and her selflessly help. I don't know how to express my thanks to her, only say thanks very much is not enough. I am also greatly indebted to many teachers in the past: Prof. Rummel who encouraged and motivated me a lot, Prof. Hogentobler, Prof. Stilla, Dr. Bosch, Dr. Schmidt, MS. Hedman and Dr. Gruber the list goes on and on. All the lecturers are very intelligent people, and I learnt from them more than I expected.

I am highly appreciated all the people of LMF and DLR, provide me a very nice study environment, including Dr. Eineder, Dr. Butenuth, Mr. Minet, Mr. Gernhardt, Mr. Parizzi, Ms. Cong, Mr. Frey, Ms. Lachaise, Mr. Gläßer, Mr. Martinez, Mr. Auer, Mr. Khadhra the list goes on and on. Meanwhile, I am highly appreciated to my teachers at NUDT who pay continuous attention to my study, including Prof. Liu Weidong and Prof. Xi Xiaoning.

Last but not least, I would like to thank my family: my parents, for giving me life in the first place, for educating me with aspects from both arts and sciences, for unconditional support and encouragement to pursue my interests; my sister, for listening to my frustrations, and for believing in me; my boyfriend, for sharing his life together with me and for giving me encouragement all the time.