The University Of Sheffield.

Access to Electronic Thesis

University Library

The University of Sheffield

Wei Wang

Signal Processing Methods to Improve Ocean Surface Wave Estimation from a High Frequency Surface Wave Radar

Degree of Doctor of Philosophy

2011

Signal Processing Methods to Improve Ocean Wave Estimation from a High Frequency Surface Wave Radar

Wei Wang

School of Mathematics and Statistics

The University of Sheffield

Thesis submitted for the degree of

Doctor of Philosophy, December, 2011

Abstract

High frequency surface wave radars are operated as a remote sensor to measure ocean surface parameters to ranges exceeding 200-300 km from the coastline. The Bragg peaks in the power spectrum of backscattered radar electromagnetic signals from ocean waves reveal the Bragg resonant effect and the second order continuum reflect a hydrodynamic and electromagnetic modulation on the radio waves from the ocean waves. The power spectrum is thus utilized to invert the ocean wave directional spectrum by a non-linear integral equation. Further integrations of the ocean wave directional spectrum yield the estimates of wave parameters: significant waveheight, mean wave period, mean direction, directional spread, etc. Beside sea echoes, non-sea echoes or interferences (collectively termed 'clutter'), are also received by radar antenna receivers and included in the power spectrum. Clutters which occupy the second order continuum are treated as sea echoes in the inverse algorithm and cause inaccurate estimation of the ocean wave directional spectrum. Thus clutter mitigation is the main purpose of this thesis and is intensively investigated.

A three-step image processing approach is proposed in this thesis which could mitigate visible clutters, e.g. radio frequency interferences and ionospheric interferences, and also invisible clutters. The kernel implements a decomposition of the mixture space and then a projection of the mixture space into a desired subspace. In addition to this main approach, various signal processing methods are also investigated for improving the wave estimates, e.g. wavelet analysis, AR modeling, adaptive filtering algorithm. The clutter mitigation scheme is validated by operational use on a whole month of Pisces data and exhibits some improvements in the accuracy of wave estimates. To aid the operational use, a statistical pattern recognition method is also developed. Finally, the best schemes are chained together for a sequential operational use in terms of providing better wave estimation.

Acknowledgements

I would like to present my biggest thanks to my supervisor, Lucy Wyatt, for her excellent guidance, patience, and expertise. I'm also grateful to a number of people who have contributed, directly or indirectly, to this work: Hengyu Ke and Zezong Chen from Radiowave Propagation Laboratory at Wuhan University; Alan Zinober, Megan Jayne Hobbs, and Christopher Drew from the School of Mathematics and Statistics for their assistance, Mike Moorhead from the Neptune Radar Ltd. for his technical advice and help; the China Education Council for funding my Ph.D., and the University of Sheffield for granting me the studentship.

Finally, I would like to thank my parents who have always supported and encouraged me to work hard and overcome the stress experienced during the Ph.D. life.

Content outline

List of Figures

List of Tables

Abbreviation

AR: autoregressive
BIF: best individual features
CCA: canonical correlation analysis
CCR: correct classification rate.
CIT: coherent integration time
CODAR: coastal ocean dynamics applications radar
CWT: continuous wavelet transform
DA: discriminant analysis
db: Doppler bin(s)
DOA: direction of arrival
DS: Doppler spectrum
DWT: discrete wavelet transform
ED: eigenvalue decomposition
FA: factor analysis
FFT: fast Fourier transform
FMCW: frequency modulated continuous wave
FMICW: frequency modulated interrupted continuous wave
GDA: generalized discriminant analysis
GED: generalized eigenvalue decomposition
HFSWR: high frequency surface wave radar
HPF: high pass filter
ICA: independent component analysis
II: ionospheric interference
In-band: first and second order region of a Doppler spectrum
KDE: kernel density estimation
KMEANS: cluster the data into K groups
KNN: K nearest neighbour
LDA: linear discriminant analysis
LLP: linear preserving projection
LPF: low pass filter
LSE: least square error
MCOV: modified covariance algorithm
MTC: meteor trail clutter
MTI: moving-target indicator
MVDR: minimum variance distortionless response
Out-band: the remaining region in the Doppler spectrum excluding the in-band region.
PCA: principal component analysis
PDA: probability distance analysis
PISCES: Pisces radar
PRF: pulse repetition frequency
PSD: power spectrum density

PRI: pulse repetition interval
rb: range bin(s)
RD: Range-Doppler
RFI: radio frequency interference
SC STAP: stochastically constraint STAP
SED: sliding eigenvalue decomposition

sfs : superior first order peak

sfi : inferior first order peak

STAP: space-time adaptive processing
SVM: support vector machine
TREE: hierarchy tree
WERA: WEllen RAdar

Chapter 1 Introduction

With growing dependence of human activities on the ocean surface state, scientists are dedicated to understanding oceanic hydrodynamics on the surface using precise oceanic observations and mathematical models. In a macro sense, the mathematical foundation and observations can be used to improve ocean models, identify and track atmosphere-ocean interaction phenomena，and predict hurricane events. In a micro sense, they can be used in diverse applications, e.g. tracking and routing ships, improving the safety and efficiency of offshore industry operations, and managing fisheries. However, the ocean surface is hard to be observed completely and accurately because it is a stochastic field, with random waves superimposed on each other. Numerous methods to quantitatively measure the stochastic field have evolved from the most conventional instruments, e.g. current meters and wave buoys, to modern techniques, e.g. international ocean surface observation satellites. The complex topography inherent in coastal regions and the complex physical oceanography in the vast offshore regions add to the difficulty of remote sensing of the ocean surface.

High frequency surface wave radar (HFSWR) is a form of radar for the surveillance and sea-state mapping over the ocean surface on a large scale, to within 2-3 hundred kilometres offshore. Its defining characteristic is the transmission of vertically polarized electromagnetic radio waves in the 3-30 MHz frequency band, along the surface of the ocean. Vertical polarization is chosen for further propagation range, because the horizontal polarized radio wave doesn't propagate along the surface (Barrick and Fitzgerald 2000). During propagation, the radio waves interact with ocean waves with all wavelengths, and are much more strongly backscattered at near grazing angles by ocean waves which travel toward or away from the radar along the direction in which the radar is looking, if their wave length is exactly half of the radar wave length. This phenomenon is technically termed Bragg resonant scattering. The advantages of using HFSWR are: one, the low-loss propagation by the highly conductive ocean surface along the curvature of the air-water interface, allows for moderate transmitted power of radio waves and sensing beyond the horizon; two, the received radio waves bear a more direct relationship to the long waves of concern (gravity waves with wavelength 10-100m) than satellite-borne radar whose defining characteristic is the transmission of microwaves (wavelength is from several millimeters to meters) which penetrate the ionosphere and interact with ocean waves at a much smaller scale, e.g. ripples. The ocean wave spectrum that satellite-borne radar samples is on the tail of the ocean wave spectrum in the Pierson-Moskowitz wind-wave model, i.e. measuring mainly short surface ripples (Mollochristensen 1984), while the ocean wave spectrum that HF radar samples is the part where longer waves of concern reside; three, HF radar overcomes the limitation of conventional instruments, e.g. in situ buoys, that can provide only point measurements.

As a remote sensing tool, HF radar is capable of measuring oceanographic parameters at high temporal and spatial resolution. The backscattered radar signals are transformed into Doppler spectra (DS) by Fourier transform. The main characteristics in the spectrum are the two peaks due to Bragg resonant scattering and the second order continuum beside each peak associated with non-linear electromagnetic and hydrodynamic processes. The first order quasi-monochromatic Bragg lines are used to derive surface current and wind direction information. The second order continuum can be used for wind speed and wave estimation. The Seaview Sensing Ltd software package originally developed at the University of Sheffield has been used for this work for the transformation from DS to oceanographic parameters, i.e. wave, current, and wind information. (This software first identifies the first and second order regions of the DS and the noise level and then applies different algorithms to extract oceanographic parameters. An important parameter is used to evaluate the detection performance of the Bragg peaks: the relative divergence of the separation of the Bragg peaks from the expected (theoretical) value. In the rest of the thesis, I will refer to this package as Seaview.) Therefore, the quality of the DS has a vital impact on the precise estimation of the ocean surface state.

Since the second order region, generated by a complex wave-wave and ocean-atmosphere interaction, is easily contaminated by clutter, wave estimation has become the most challenging task facing every researcher in this field. Various clutters from known or unknown sources mixed in the radar measurements will degrade the quality of DS, resulting in errors in the estimations of the ocean surface state or even preventing any estimation. Thus, the improvement of wave estimation accuracy by clutter mitigation is the main goal of this thesis. The goal is implemented by undertaking three tasks as follows. The first task is the mitigation of radio frequency interference (RFI) and ionosphere interference (II). Among all the clutters, RFI and II are the two most frequently found and damaging clutters, in particular RFI in this application. Theoretical study on RFI and II clutters shows that they display unique and characteristic shapes in the Range-Doppler (RD) image (which is an image of DS along all detected range bins), i.e. RFI has a range-correlated shape and II a Doppler-correlated shape. These two clutters are called visible clutters in this thesis because they can be identified in the RD image. However some other clutters can be more damaging. They are wholly hidden inside the second order region and are thus invisible, e.g. ships. The mitigation of such invisible clutters is the second task. An impact of invisible clutter is to give either an overestimate or an underestimate of waveheight. If the invisible clutter is added to the superior Bragg peak, the resulting waveheight estimate will be underestimated. And if on the second order, the waveheight will be overestimated. To ameliorate this, different methods may be needed for these cases. The third and last task is a pattern recognition and classification of the Doppler spectra. By the classification, it can be determined whether an arbitrary Doppler spectrum is likely to lead to over or under prediction of the significant wave height, and hence an appropriate

method can be selected. An improvement in the availability and accuracy of wave measurements is the expected goal.

The data used in this thesis are provided by Pisces radar systems (Shearman and Moorhead 1988), which are in the format of DS. The data of another radar system WERA (Gurgel et al. 1999) are also available. WERA raw data are received signals from each antenna, which have not been differentially phased to form beams in various directions. Hence, they require a further processing termed beamforming before the DS transform. WERA DS are also available but they are on a grid so not so easy to do range-Doppler processing. No matter what radar system is adopted, DS are the final representation of data. If the mitigation of clutters in DS can be achieved for the Pisces dataset, the work in the thesis can be applied to other radar systems. For this reason beamforming methods are not discussed in-depth and the emphasis is put on the application to DS, although different clutter mitigation methods could be applied to un-beamformed data. A co-located wave buoy provides estimates of parameters of the directional frequency spectrum for the comparison and verification work in this paper. The whole structure of the thesis is explained below.

Chapter 2 presents an overview of the HFSWR system. The historical development of HFSWR systems is reviewed. The principles of HFSWR are introduced: Barrick's theory of the first- and second- order HF radar cross sections is presented; the linear frequency modulated continuous waveform is described for the radar transmitted signal; the transformation from radar received signal into DS is explained; the inversion algorithms to extract wave parameters are briefly described. The Pisces radar and Pisces dataset are both introduced. Beamforming techniques for array signals are briefly reviewed. The properties of the two main damaging clutters, i.e. RFI and II are examined in Chapter 3. Physical mechanisms that result in the existence of these two clutter sources in DS are explained. Note that, 'stationarity' is a term used in many places in this thesis. It is a term usually defined in two ways, wide sense or narrow sense. Wide sense stationarity means all the orders of statistics of a stochastic process are constant. Narrow sense stationarity only requires second order statistics to be constant. In this thesis, by stationarity, I mean the narrow sense.

A review of various signal processing schemes, to improve the quality of DS, is presented in Chapter 4. Three signal processing algorithms, i.e. wavelet analysis, autoregressive modelling, and adaptive filtering system, are examined and used for the detection of Bragg peaks, denoising, prediction of DS, and interpolation of missing data respectively. These applications can only help improve the quality of DS to a limited extent, because of a lack of in-depth mitigation of damaging clutters. The most efficient clutter mitigation scheme is a three-step image processing solution, which involves image recognition, segmentation and subspace projection. This solution is examined and demonstrated in detail in Chapter 5.

Chapter 5 divides all the clutters in the RD image into visible and invisible types. For visible ones, the most damaging clutters are RFI and II. The mathematical analysis of RFI and the clutter mitigation scheme based on the characteristic range-correlated shape are developed. This scheme is extended to cancel arbitrary visible clutters. Furthermore, invisible clutters are investigated in terms of their second order statistics. Blind source separation theory is used to mitigate those hidden clutters with two schemes and a priori knowledge of the buoy estimation. If the radar estimate is higher than the buoy, the over-estimation scheme is chosen, and if the radar estimate is lower, the under-estimation scheme is chosen. Three schemes are recommended for operational use. The improvement on some statistics of the wave parameters is presented. Note that robust tests of the statistical significance of these statistics are not done at the moment, so the conclusions are based on qualitative judgements.

Based on the study in chapter 5, an accurate prediction of the quality of DS, i.e. whether they are going to provide over- or under- estimations, is required, if buoy information is not available (in general buoy data are not available and, even if they are, they are one-location measurements and cannot cover the whole range of radar measurements). Pattern recognition and classification of DS are investigated in Chapter 6. Statistical pattern recognition based on Bayes theory and a geometric approach is examined separately. Since there are a large number of variates in a DS, feature selection and extraction methods are investigated for dimensional reduction. The canonical correlation analysis is demonstrated to show the best discriminating performance compared with other feature extractions methods. In the reduced feature space, modern classification techniques are examined. Support vector machine has proved to be the best classifier for this classification task. The best classification rate achieved for under and over estimation classes is around 76% and 86%. Future work is needed for improving the correct classification rate (CCR), the ratio of the number of correctly classified data to the number of the whole data to be classified.

Chapter 7 explains how the best algorithms recommended in the earlier chapters are brought together in a quasi-operational chain. The overall conclusions of the work in this thesis are presented in chapter 8.

Chapter 2 Principles of HFSWR and modeling of HFSWR signals

HFSWR systems are used to measure surface currents, wind, and waves in coastal regions. The radar system includes transmitting and receiving antennas. The transmitting antenna is used to transmit vertically polarized electromagnetic waves to propagate along the sea surface beyond the horizon. The receiving antenna is able to receive signals backscattered by ocean waves of half the electromagnetic wavelength, from a number of ranges depending on power, bandwidth and azimuth.

2.1 Historical review of HFSWR

The sea echo of HFSWR was first observed by the British defense radar network in World War II, and later confirmed by echoes found by scientists from New Zealand and Australia during ionosphere detection at the coast. For a relatively long time, people could not give a rational explanation for the source and cause of the signal. Crombie (1955) was the first one who proposed that this type of signal is induced by a backscattering mechanism of radar electromagnetic waves from ocean waves. Wait (1966) discovered and explained the Bragg backscattering mechanism in 1966. Barrick introduced the first and second order contribution of sea echo to the radar cross section and invented the first remote sensing tool for the sea surface: Coastal Ocean Dynamics Applications Radar (CODAR) for sea surface current measurements in 1977 (Barrick et al. 1977). Since then, the development of HFSWR speeded up and new inventions turned up all over the world. In the UK, examples of such systems are the Pisces, developed for long range wave measurement by Neptune Radar Ltd., Gloucester, from a University of Birmingham prototype (Shearman and Moorhead 1988) in 1987 and the Ocean Surface Current Radar (OSCR) (Prandle et al. 1992) developed in the Rutherford Appleton Laboratory at about the same time as Pisces; in America, Barrick further introduced the SeaSonde radar system (Paduan and Rosenfeld 1996); in Canada, the High Frequency Ground Wave Radar (HF-GWR), a typical multi-functional narrow band HFSWR was located in Newfoundland (Khan et al. 1994); in Germany, the Wellen Radar (WERA) was developed (Gurgel et al. 1999); in China, the Ocean State Monitor and Analysis Radar (OSMAR) was first developed in Wuhan University in 1993 and improved for long range measurements by the end of 1999 (Wu et al. 2001).

2.2 Principles of HFSWR

2.2.1 Barrick's equation

When HF radars transmit radio waves, they are scattered from ocean waves in all

directions, with some scattered toward the radio receiver. The largest contribution to the signal at the receiver has been shown to be due to scatter from ocean waves of half the radio wavelength traveling directly toward or away from the radar (Barrick 1972). This produces a peak in the DS of the demodulated backscattered signal at a frequency equal to the ocean wave frequency of this Bragg-matched wave. This frequency is positive if the wave is propagating toward the radar, and negative if the wave is traveling away. There is a second, usually smaller, peak in the DS at minus the frequency of the larger peak due to the Bragg-matched wave propagating in the opposite direction. These peaks are referred to as first-order peaks because they can be described by the first-order solution of a perturbation analysis of the interaction between electromagnetic and hydrodynamic waves (Barrick 1972). In this thesis, they are denoted as *sfs* and *sfi*, termed superior and inferior first order peak respectively. The rest of the power spectrum consists of a continuum, i.e. the second-order region of the spectrum (Barrick 1972), and a background noise floor. An example of a DS is given in Fig.2.1. The region between 0.4*(the frequency of the first order peak) and 1.6*(the frequency of the first order peak) is used for wave height and directional spectrum estimation. This is the region shaded in dark green in Fig.2.1. The choice of this range of the Doppler frequencies is associated with a linearization of the wave inversion algorithm. The light green is not currently used for oceanographic measurements (Wyatt 1989). The first order Bragg peaks are generally two orders of magnitude greater than the second order sea echo continuum.

HF radar measurements of ocean waves make use of a second-order theory for the hydrodynamic and electromagnetic processes that generate the back-scattered signal. HF radar systems for making measurements of sea-state parameters have developed over the last 25 years since Barrick (1972) derived theoretical formulations for the observed backscattered signals based on the directional spectrum of the sea surface. A first-order theory is the basis for surface current and wind direction measurements. Systems for making such measurements are now available commercially and have been used extensively providing detailed maps of surface currents in coastal waters. With respect to the second order theory, the ocean wave directional spectrum can be extracted from the second order sidebands by solving a non-linear integral equation (Wyatt 1995). Full mathematical details for the analysis of second order ocean wave interactions can be found in the work of Weber and Barrick (1977), and Weber and Barrick (1977). The second-order electromagnetic analysis is elaborated in Lipa and Barrick (1986) and is based on the method of Rice (1951).

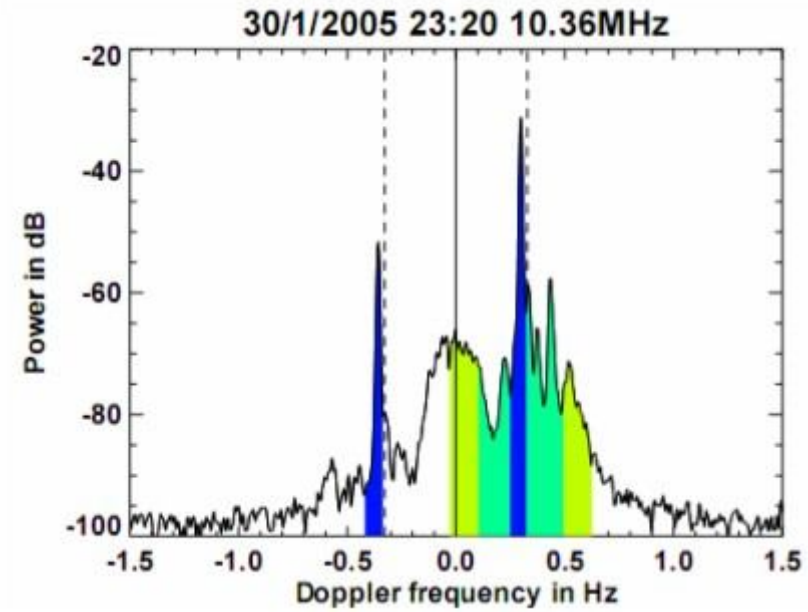


Fig. 2.1. Typical DS obtained at 23:20, 30/1/2005. First order peaks, shaded in blue and marked by the dashed lines are used for surface current and wind direction estimation. The second order continuum is shaded in dark and light green.

The two principal Barrick's equations describing the relationship between the power spectrum of the demodulated back-scattered signal and the directional wavenumber spectrum of the ocean are presented below (Barrick 1972). The one technically called 'first order equation' takes the form:

$$\sigma_1(\omega,\phi,d) = 2^6 \pi k_0^4 \sum_{m'=\pm 1} S(-2m'\vec{k}_0)\delta(\omega - m'\sqrt{2gk_0 \tanh(2k_0 d)}) \qquad (2.1)$$

where m' is the sign of the Doppler shift, $\vec{k}_0$ is the radar wave vector of magnitude k_0 and direction towards the scattering patch from the radar, $S(\vec{k})$ is the ocean directional wavenumber spectrum, and d is the water depth. ϕ is the radar bearing. This equation describes two peaks located at $\pm\omega_b$, where $\omega_b = \sqrt{2gk_0 \tanh(2k_0 d)}$, with amplitudes dependent on the amplitudes in the directional spectrum along the radar beam direction, towards the radar $S(-2k_0)$, and away from the radar S(+2 k_0).

The second-order contribution to the radar cross-section according to Barrick (1972) is given by:

$$\sigma_2(\omega) = 2^7 \pi k_0^4 \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \left|\Gamma(\bar{k}_1, \bar{k}_2)\right|^2 S(\bar{k}_1) S(\bar{k}_2) \delta\{\omega \pm \omega_1 \pm \omega_2\} dpdq \tag{2.2}$$

where the integration variables p and q are wavenumber components parallel and perpendicular to k_0, respectively; $\vec{k}_1$ and $\vec{k}_2$ are given by $\vec{k}_1 = (p - k_0, q)$ and $\vec{k}_2 = (-p - k_0, -q)$, so $\vec{k}_1 + \vec{k}_2 = (-2k_0, 0)$, which means a double scattering process leads to a Bragg resonance; $\omega_1 = \sqrt{g\left|\vec{k}_1\right| \tanh\left|\vec{k}_1\right| d}$, $\omega_2 = \sqrt{g\left|\vec{k}_2\right| \tanh\left|\vec{k}_2\right| d}$; '± 1' means the second-order contribution appears on the right and left of the first order peaks; Γ denotes the coupling coefficient, describing both the electromagnetic and hydrodynamic process that provide the second-order backscatter. The details can be found in Holden and Wyatt (1992).

2.2.2 **Waveform, range processing and Doppler processing**

Frequency modulated continuous waveform

Originally, HFSWR used a waveform of single-frequency pulses. The target's range is calculated by the time delay between the transmitted and received pulses. The pulse repetition interval (PRI) determines the maximum ambiguous range, i.e. $R_{\max} = \frac{cT_{PRI}}{2}$. The pulse length determines the range resolution. If a short pulse length is used, the range resolution is better. But since a high SNR is required, the transmitted peak power would need to be high to ensure enough energy is radiated onto the target. The fundamental quantity determining the range resolution is bandwidth (Skolnik 1990). Pulse compression methods, e.g. linear frequency modulation, are utilized to solve the problem by increasing the bandwidth of the pulse while using a long pulse; the requirements of good range resolution, moderate peak power, and good SNR are all satisfied. Hence, modern HF radar systems choose a form of frequency modulated continuous wave (FMCW) (Stove 1992). The range measurement is based on the difference of instantaneous frequency of transmitted and receiving radar waves. The FMCW provides good range resolution for a given transmission bandwidth, and allows for an output of constant power generated by the transmitter power amplifiers. The major limitation of FMCW is that it is easily jammed by pulsed radars with high power. That is why FMCW is generally not adopted by the military for surveillance (Bill 2008). The FMCW waveform, also called a chirp signal, is briefly formulated below, with the detailed theoretical analysis presented by Khan and Mitchell (1991). A review of numerous papers addressing different aspects of FMCW radars has been presented by Skolnik (1970).

Let us define the instantaneous frequency f_{FMCW} of the FMCW waveform, for a single repetition (a sweep is a PRI) as:

$$f_{FMCW}(t) = \frac{B}{T}t + f_c, \qquad |t| \le T/2 \tag{2.3}$$

where f_c is the waveform centre frequency; B is the bandwidth of the FMCW radar; T is the PRI. The phase of the signal with instantaneous frequency in Equ.(2.3) can be calculated as:

$$\phi(t) = 2\pi \int_{-T/2}^{T/2} f_{FMCW}(t)dt = \frac{\pi B}{T}t^2 + 2\pi f_c t, \quad |t| \le T/2 \tag{2.4}$$

Supposing the signal, $s(t)$, has amplitude, $A(t)$, a convenient complex notation is:

$$s(t) = A(t)e^{j\phi(t)} \tag{2.5}$$

Range and Doppler processing

When the received chirp signal is mixed with the transmitted chirp signal, and then passed through a low-frequency filter, the high-frequency components are removed and the baseband time series of the radar return is obtained. The range processing applies the Fast Fourier Transform (FFT) on the baseband signal in each sweep to generate a frequency spectrum. Since the frequency is linearly related to the time delay, i.e. range, of target echoes, target echoes from different ranges are separated in the frequency spectrum (or range spectrum). The number of range returns is determined by the frequency number, N_{FFT} (i.e. the size of FFT). This processing is also called the first time FFT. Suppose there are N_{CIT} sweeps in a coherent integration time (CIT), then there will be N_{CIT} range spectra. Coherence here means the phase shift between the received radar signal and the reference signal in each sweep is a fixed relationship. Within the CIT, the sea surface is assumed to be stable, and the received radar signals are reflected by the area of the 'same' sea state. (Barrick (1977) claimed that independent spectral samples are always obtained using a CIT of 200s in HF surface-wave radars for sea-state sensing. Different radars may have long or short CIT in their settings, depending on the stability of the sea surface, e.g. 100s if the sea is very unstable, and also on the Doppler frequency resolution, e.g. higher CIT gives better Doppler resolution which is good for wave estimation. That is why Pisces chooses 204.8s as its CIT for the estimation of oceanographic parameters. In addition, the impact of non-sea echo can thus be minimized by incoherent

integration). A target echo coming from a range bin consists of N_{CIT} temporal samples in a CIT. In the analysis given in section 5.1.2., it is found that the phase of this "time series" is linearly related to the velocity of the target. Then the second time FFT is applied to all the N_{CIT} samples for each range return to generate the power spectrum (usually and hereafter referred to as Doppler spectrum or just DS). This second time FFT is termed Doppler processing. After Doppler processing, the Doppler (velocity) information of radar echoes from a certain range can be obtained from the DS, as shown in section 5.1.2.

2.2.3 **Wave measurements**

HF radars can measure oceanic parameters, e.g. surface current, wind direction, wave height, wave period. Among them, the wave estimates have been found to be most challenging. An inversion algorithm for ocean wave directional spectrum estimation has been proposed (Wyatt 1989; Wyatt 2000; Wyatt et al. 2006), which allows for the evaluation of waveheight, wave period and wave direction. Although there are other inversion methods, I focus on this one because that is the one I will apply in this thesis.

Ocean wave directional spectrum

Wyatt (1990) extended the Chahine-Twomey relaxation method for the inversion of the atmospheric radiative transfer equation to provide an inverse solution to Barrick's integral equation describing the second order mechanism. The inverse algorithm is an improvement on the method introduced by Lipa and Barrick (1986) by allowing for the measurement of a wider range of ocean wave frequencies. Equ. (2.2) can be normalized using the Bragg frequency and wave number (Lipa and Barrick 1986) and expressed in the form of:

$$\sigma_2(\eta) = \int_{-\pi}^{\pi} K(\eta,\theta) S(k,\theta) d\theta \qquad (2.6)$$

where the wavenumber $k = k(\eta,\theta)$ is normalized by $2k_0$ (k_0 is the radar wavenumber), η is the Doppler frequency normalized by Bragg frequency, $S(k,\theta)$ is the ocean wave directional spectrum. With initialization and an iteration, the $S(k,\theta)$ can be inferred without any a priori information about the form of the solution. As the initial condition, a Pierson-Moskowitz wind-wave model is used to estimate the frequency spectrum and $\cos^4(\theta-\theta_w)/2$ is used

for the directional distribution with wind direction, θ_w, determined from the first order analysis (Wyatt et al. 1997). By the inversion algorithm described above, a dual HFSWR system is able to measure the directional spectrum of ocean wave energy at the frequencies with which it operates. Note that the inversion method doesn't depend on radio frequency, so it can be applied at any HF frequency (Wyatt and Holden 1992).

The solution of Equ (2.6), $S(k,\theta)$ (or $S(k_x,k_y)$), is on a wavenumber grid. This spectrum can be further converted into a form in terms of ocean wave frequency. The relationship between the directional frequency spectrum and the wave number spectrum can be expressed by:

$$S(f,\theta)=\frac{dk}{df}kS(k_x,k_y) \tag{2.7}$$

For a given wavenumber $\vec{k}$,(k_x,k_y), the corresponding angular frequency ω and ocean wave frequency f is defined by the dispersion relation:

$$\omega^2 = gk\tanh kd \qquad (2.8), \qquad f=\frac{\sqrt{gk\tanh kd}}{2\pi} \qquad (2.9)$$

where g is the acceleration due to gravity; k is the magnitude of the wavenumber vector $\vec{k}$; d is the water depth. According to the Equ.(2.9), the Bragg frequency f_B is $\pm\frac{\sqrt{2gk_0\tanh 2k_0d}}{2\pi}$. For the deep water case, f_B is approximately $0.102\sqrt{f_{radar}}$, where f_{radar} is the radar working frequency in the units of MHz.

Using Equ. (2.8) and (2.9), Equ.(2.7) is rearranged into the form:

$$S(f,\theta)=\frac{8\pi^2 f}{g}\sqrt{k_x^2+k_y^2}S(k_x,k_y) \tag{2.10}$$

$S(f,\theta)$ can be further expressed as the product of a magnitude function of frequency $S(f)$ and a directional distribution $\theta(f,\theta)$, given as:

$$S(f,\theta) = S(f)\theta(f,\theta) \tag{2.11}$$

$S(f)$ can be estimated by integrating the solution, $S(f,\theta)$ found from above.

$$S(f) = \int_0^{2\pi} S(f,\theta)d\theta \tag{2.12}$$

Hence the $\theta(f,\theta)$ is obtained by Equ.(2.11). Now, this directional distribution function is expanded into a Fourier series, as:

$$A_n(f) = \int_0^{2\pi} \theta(f,\theta)\cos n\theta d\theta \tag{2.13}$$

$$B_n(f) = \int_0^{2\pi} \theta(f,\theta)\sin n\theta d\theta \tag{2.14}$$

Equ. (2.15) below is used to estimate the mean direction spectrum versus frequency, $\theta^*(f)$, using the first two Fourier coefficient function $A_1(f)$ and $B_1(f)$ by:

$$\theta^*(f) = \arctan(\frac{B_1(f)}{A_1(f)}) \tag{2.15}$$

An example of the full ocean wave directional spectrum is shown below in Fig.2.2, where the directional spectrum is also represented by two separate spectra, i.e. a frequency spectrum $S(f)$ and a mean direction versus frequency spectrum $\theta^*(f)$ (termed direction spectrum for short hereafter). Fig.2.2(c) shows a lower and higher cut-off ocean wave frequency at 0.05 and 0.24 Hz respectively. The lower ocean wave frequency is determined by the Doppler frequency in the second order spectrum closest to the superior first order peak and the higher cut-off ocean wave frequency is determined by the Doppler cut-off frequency 1.6* f_B , where $f_B = 0.102\sqrt{f_{radar}}$, corresponding to a maximum wave frequency at 0.3Hz. Lipa (1977) has described how the DS can be split up into a key region close to the first order peak, (Doppler frequency from 0.6* f_B to 1.4* f_B), where the contributions are predominantly due to interactions between long and short waves (linear model adopted for the

inversion algorithm is invalid out of this region), and a region where radar observes short wave-wave interactions (the rest). This key region is extended to $0.4 * f_B$ to $1.6 * f_B$ by Wyatt (2000). Within this key region, swell (very long waves) information is contained in the frequencies very close to the first order peak. Wind-wave (long waves) information is contained in the rest of the second order spectrum. The higher the Doppler frequency displacement from the first order peak, the shorter wave is measured. Therefore, the highest Doppler frequency $1.6 * f_B$ gives the information of ocean waves of the highest detectable frequency. Several oceanic parameters are calculated from the estimated ocean wave directional spectrum, e.g. significant waveheight, mean wave period, mean direction.

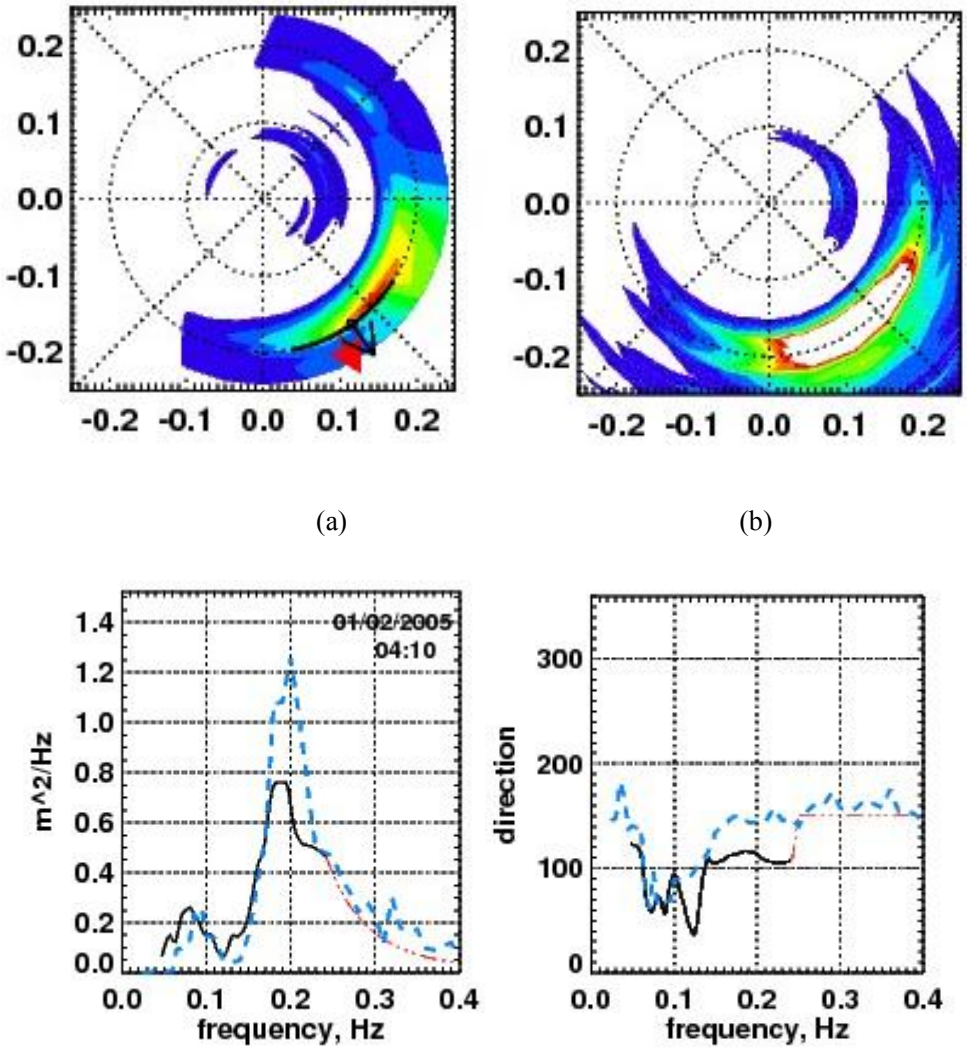


Fig.2.2 Ocean wave directional spectrum. Top panel-full ocean wave directional spectrum with the lowest level at 0.1*peak amplitude; the grid circles have 0.1 Hz steps, radial 45°steps; north is up, east right. The distance from the center gives the frequency. Long waves are near the center, shorter waves at perimeter: (a) radar, (b) buoy. The arrows in (a) show peak amplitude and frequency as measured by the radar (black) and the buoy (red). Bottom panel-(c) Frequency spectrum, (d) Direction spectrum (solid is the radar and dashed is the buoy). The scale in the LH axis is adjusted automatically in Seaview to centre the radar estimation. The frequency spectrum is in the unit of m^2/Hz because of the relationship between waveheight and frequency spectrum (refer to Equ. 2.16).

Significant waveheight

Significant waveheight is defined using the zeroth moment of the ocean wave frequency spectrum $S(f)$ in Equ. (2.16).

$$H_s = 4\sqrt{\int_{f_0}^{f_1} S(f)df} \tag{2.16}$$

where the frequency range of measurement is from f_0 to f_1, which covers the range of wave frequencies provided by the radar. The lower bound f_0 depends on the quality of the first order peaks, i.e. whether the first and second order are separated. The upper bound depends on radio frequency so will change for Pisces dataset. Typical values of f_0 and f_1 are 0.05 Hz and 0.22 Hz.

Fig.2.3 shows the radar measurement of significant waveheight in Feb., 2005, compared with buoy measurement.

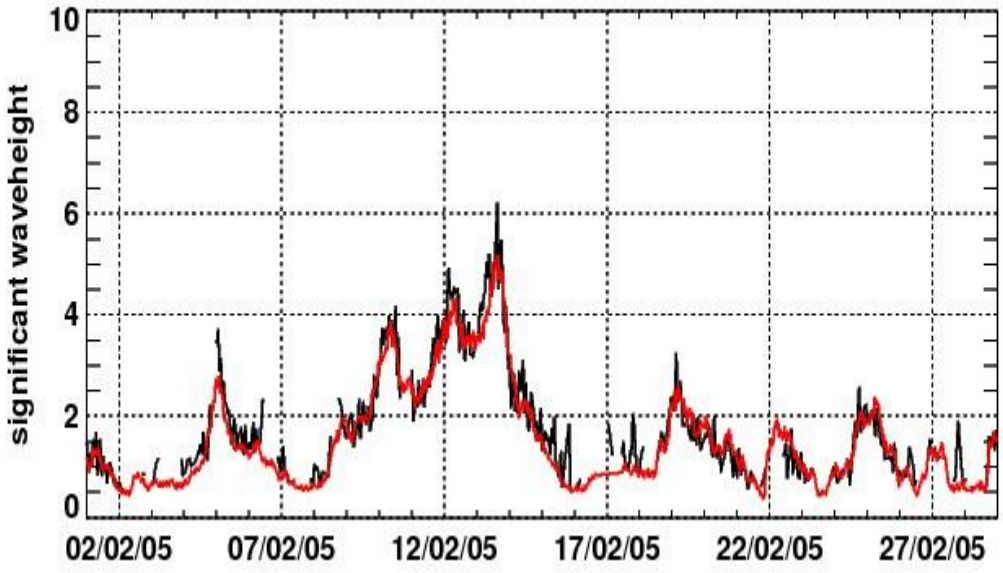


Fig.2.3. Significant waveheight (m) comparisons between buoy and radar for Feb 2005. Buoy measurement "red " and radar "black".

Mean wave period

Mean period is calculated using either the first or second moment of $S(f)$:

$$T_n = \left(\frac{\int_{f_0}^{f_1} S(f)}{\int_{f_0}^{f_1} f^n S(f)} \right)^{\frac{1}{n}} \qquad (2.17)$$

where $n = 1, 2$ indicate the first moment and second moment of the spectrum, i.e. T_1, T_2. Note that the 2nd moment of the spectrum, T_2, is particularly bad for Pisces comparisons with buoys since it emphasises the higher frequencies which the radar is not measuring. T_1 would be a better parameter for Pisces comparison in this work.

Mean wave direction

Mean wave direction θ_m is formulated as below:

$$\tan\theta_m = \frac{\int_{f_0}^{f_1} \sin\theta^*(f) S(f) df}{\int_{f_0}^{f_1} \cos\theta^*(f) S(f) df} \qquad (2.18)$$

where $S(f)$ and $\theta^*(f)$ are both determined from the directional spectrum $S(f,\theta)$ as explained above. Wyatt (Wyatt et al. 2006) found that $\theta^*(f)$ is rather more sensitive to noise in that spectrum than $S(f)$. It has been shown (Wyatt et al. 2006), that the accuracy of θ depends on waveheight, with much better estimates for waveheights greater than about 2m.

All these formulae have been utilized in the Seaview. This thesis focuses on the signal processing algorithms which could improve the quality of DS, so that the derived wave estimation by this software platform will be improved in both validity and accuracy. Although the first order region is important for current and wind estimation (the improvement in this part of a DS is also taken into consideration), the second order continuum is the region that directly reflects the ocean directional wave spectrum. Therefore, good SNR and lack of interference is needed. The biggest emphasis is put on this part.

2.3 Pisces radar data

Pisces (Shearman and Moorhead 1988) was developed by Neptune Radar Ltd., Gloucester, U.K., from a University of Birmingham prototype. It is designed to obtain ocean measurements up to about 200km. The Pisces radar system adopts a waveform of frequency modulated interrupted continuous wave (FMICW) (Khan and Mitchell 1991) to propagate electromagnetic waves at frequencies between 5 and 11 MHz with a mean power of up to 1200W. Shearman and Moorhead (1988) stated that this waveform has the advantages of modest peak power and sideband spread. In addition to a higher range resolution compared with pulse waves, they said that "FMICW modulation allows for mono-static operation without transmitter noise or power limitations and with the ability to focus the peak radar performance at the longest range". In Pisces (Wyatt et al. 2006), the bandwidth B is in the range of 8-250 kHz, providing range resolutions, ΔR, from 0.75 to 20km using $\Delta R = \frac{c}{2B}$.

Experiments using dual Pisces radar systems were implemented between 2004 and June 2005 to provide full directional wave spectrum and vector current measurements. One radar was deployed at Nabor Point, North Devon and the other one was located at Castle Martin, South Wales (see Fig.2.4) (the two datasets are referred to as 'np' data and 'cm' data respectively hereafter). Both have three fixed-beam positions with low sidelobes, which were used one after another sensing on each beam for about 19 minutes. Thus, the DS along each beam is measured hourly. Note that single radar operation measures the radial velocity of the current, but it is not capable of measuring the current vector velocity; single radar also measures the wind (from short wave) direction, but it cannot measure ambiguous wind direction, i.e. distinguish short waves traveling to the right or to the left of the beam. Dual radar systems are then required to resolve direction and amplitude ambiguities in wave and wind measurement and provide full vector surface currents (Wyatt 1987). The radar working frequency was varying from hour to hour in the 5-11MHz band to minimize the impact of interference as much as possible. Although observations from ranges within 300 km are all stored in Pisces dataset, the radar is configured to provide a maximum range of about 150 km for wave estimation. As a result, the intersection point at far range (No.3 in Fig.2.4 which is more than 200 km from each radar), did not produce valid data. At the same time of this trial, a directional waverider buoy was specifically deployed at 51° 9'.85N and 005° 21'.19W in the water depth of 68m to validate the radar data. This buoy was located at the position of No.4 in Fig.2.4, which is the intersection point of beam 1 from Castlemartin and beam 2 from Naborpoint, roughly range bin 5 ('range bin' and 'Doppler bin' will be termed 'rb' and 'db' for short hereafter) for each beam. Therefore, DS backscattered from rb 5 of both radars are used for the development of many algorithms in the following chapters using the knowledge of in-situ

measurements from the buoy. Note that, although wave buoy has a problem in measuring low frequency waves (Allender et al. 1989), they are assumed to be the reference data, i.e. accurate measurements, in this thesis.

The specification of the Pisces radar for this experiment is given in Table 2.1.

Table 2.1 Specification of Pisces radar system

Key parameters	**Setting**
Radar working frequency	5~11 MHz
Angular coverage	8° for each beam
Beam bearings (see Fig.2.4)	Nabor Point: 269°, 292°, 315° Castlemartin: 201.5°, 224.5°, 247.5°
Range coverage	0-300 km (0-150km for wave)
Range resolution	15 km
Bandwidth	10 kHz
Doppler coverage	±2.5 Hz
Doppler resolution	0.0048 Hz
Coherent integration time	204.8 s
Transmitter amplifier	3 kw peak / 900w mean
Transmit antenna	One 33-m-high log periodic
Receive antenna	16 elements

The pulse repetition frequency f_r is 5Hz, so the PRI T_{PRI} is 0.2s. The coherent integration time T_{CIT} is 204.8s. So the number of sweeps N_{CIT} is 1024. Note that averaging is need as part of the Doppler processing because the power spectrum estimated is from the scatter of a random ocean surface. Each times series of the range-separated data is about 19 minutes long and split up into overlapping T_{CIT} sections for coherent integration, which are then averaged to provide the DS used in this thesis. The overlapping ratio is set to 75% in Pisces. The number of sections in the Pisces average is 19. After an N_{CIT}-point FFT,

each DS has 1024 frequency bins. An alternative unit of the DS is Hz, i.e. 1 Doppler bin represents 0.0048 Hz. The bandwidth B is 10 kHz, giving a range resolution of 15 km. The maximum coverage is 300 km, so there are 20 range bins. The whole dataset used in the thesis are two months of data collected in February and March, 2005, which is stored in a 4-dimensional matrix: day, hour, range, and Doppler. The dataset observed at one day and hour consists of 20×1024 Doppler samples, i.e. $X \in \Re^{20 \times 1024}$, which is usually represented in a RD image in the following chapters.

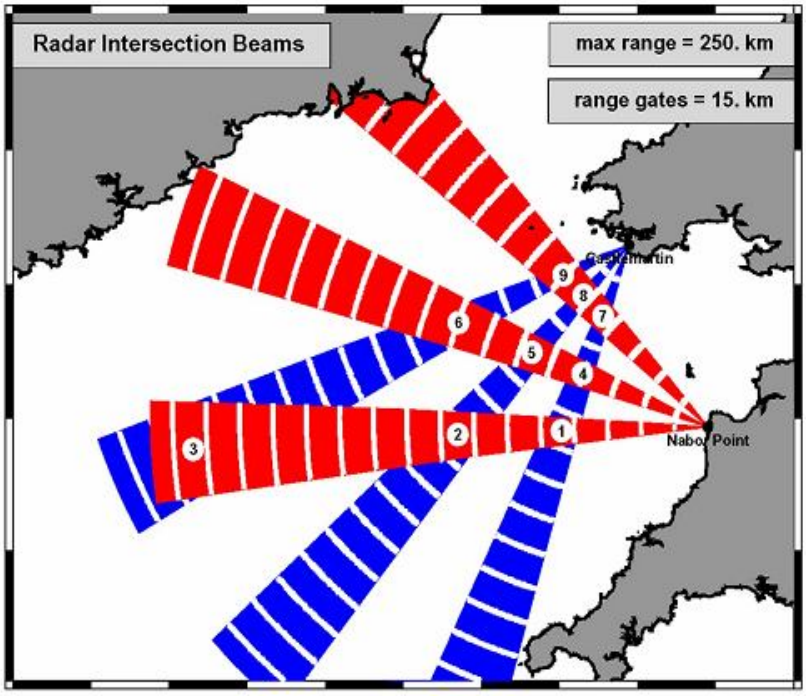


Fig.2.4. Radar sites of Nabor Point and Castlemartin, beams, range cells, and nine dual intersection points. The lower left-hand corner of the figure is at 50° N, 9° W, and the graticule increments are every 30'. (This figure is copied from Wyatt (2006), page 820).

2.4 Additional comments on beamforming

WERA is another important radar system commonly used around the world. In 1995, WERA was developed at the University of Hamburg by Gurgel (1999). This system transmits and receives frequency modulated continuous wave (FMCW). The chirp length of the FMCW is 0.26s. Typical bandwidth is 100 kHz, giving a range resolution of 1.5km. The grid resolution in the azimuthal direction, $\Delta\theta$, is typically 1^{o}. The WERA data available are a collection of observations from 16 antennas for all the range bins (80 range bins) over a CIT (2048 chirps). These data are the sorted range spectra after using the first time

FFT. At each antenna, the backscattered signal from a certain range bin is a superposition of backscattered waves from different azimuthal directions. Before Doppler processing, this array signal require beamforming for azimuthal resolution (one DS describes the power spectrum of the backscatter from a patch area of the sea in a certain range bin and a certain azimuth direction). Although the focus of this thesis is on the development of signal processing algorithms based on the DS, beamforming as a significant method for radar array signals is worth a brief review. It is not only a pre-processing for generating DS, but also is relevant to the removal of clutters in radar returns recorded from the antenna arrays.

Beamforming is used in antenna arrays for directional signal reception. Information from a linear array of antennas is combined in such a way that the expected radar return from a certain angle is preferentially observed. In HFSWR systems, beamforming can be performed in either an analog or digital way. For example, Pisces realizes beamforming in hardware by switchable cables used as phase shifters; WERA implements beamforming in software by conventional digital beamforming. In the following, the digital beamforming methods, conventional or adaptive, are briefly discussed for the application to array signals, like WERA.

Conventional beamforming uses a fixed set of weightings to combine the signals from the antenna elements in the array, primarily using the spacing of antennas and the direction of arrival (DOA) of waves. The feature of conventional beamforming is to pick up the desired signal. But it fails to suppress interferences. In contrast, adaptive beamforming techniques, e.g. minimum variance distortionless response (MVDR) (Capon 1969), generally combine the spacing and DOA with properties of the signals actually received by the array, mainly to improve rejection of unwanted signals from other directions. The advantage of Capon beamforming is to adaptively suppress interferences with high azimuth resolution and low sidelobes compared with conventional beamforming, as shown in Fig.2.5. The standard Capon beamformer assumes the signal of interest has a known DOA. However, in many cases, the DOA is not known. If there is any mismatch between the actual DOA of signal and the one assumed by the MVDR, the desired signal will be degraded, because it is treated as ‘noise’ coming from other directions.

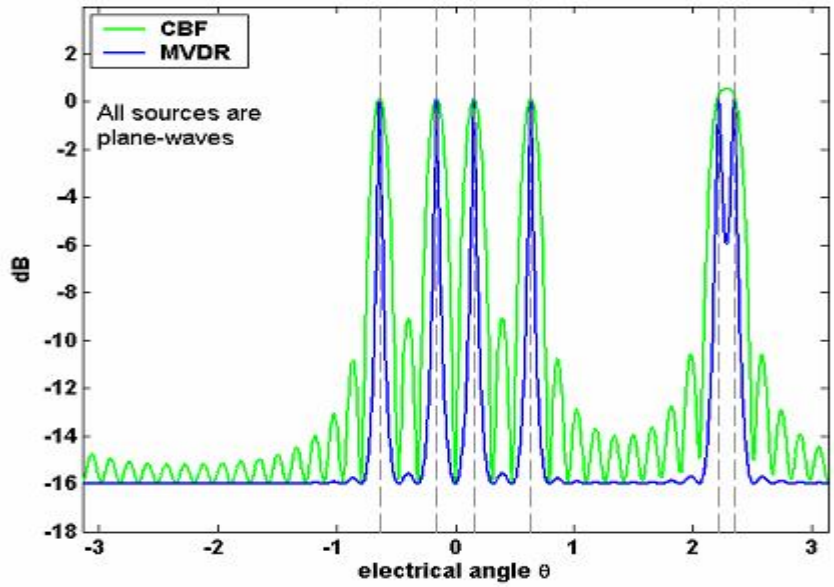


Fig.2.5 MVDR (blue) and conventional beamforming (green). (This figure is obtained from the presentation of Louis Scharf, Ali Pezeshki, and Barry Van Veen)

In radar application for ocean surface sensing, the direction vector of a wavefront is unknown in many scenarios. The estimated Bragg peaks may not match with the actual DOA exactly, due to multi-path, random scattering and medium, and array perturbations. This mismatch will result in the cancellation of desired sea echoes as well as the unwanted signals. In this regard, MVDR is not better than conventional beamforming. A robust beamformer that can combine the advantages and minimize the disadvantages of conventional and MVDR beamforming for the sensing of the ocean surface in HFSWR is still an open problem. In the field of HF skywave radar (skywave radar when used for ocean surveillance transmits 5-28MHz high frequency radio waves at an angle to the ionosphere, down from there to the sea where it is backscattered by targets via the ionosphere again to the radar), Anderson, Abramovich, and Fabrizio (1997) introduced a space-time adaptive processing (STAP) which combined MVDR beamforming with multiple constraints to eliminate 'nonstationary clutters' (ionospheric interferences). These linear constraints, either single or multiple, were investigated by Griffiths (1996). Fabrizio (2006) exploited STAP to cancel sidelobe RFI with spatial degrees of freedom and to reject main-beam RFI with temporal degree of freedom in skywave systems. Fabrizio (2008) addressed the mitigation of ionospheric interference spread in range and Doppler domain, and compared conventional and adaptive beamforming in terms of target detection for surface wave radars. The stochastically constrained (SC) STAP (Abramovich et al. 1998; Abramovich et al. 2000) is an adaptive beamforming method used in skywave radar systems. The data which are processed by SC STAP have been both beamformed and clutter-mitigated. SC STAP routines are operational

compared with “general fully adaptive” STAP (Fante and Torres 1995). Although STAP and SC STAP are both developed in the scenario of skywave applications, the signal models of sea echoes, ionosphere interference, target echoes, and noises are similar and thus their principles can be brought into the surface wave application. The disadvantage of SC STAP is its computation load which is much larger than conventional beamforming and MVDR. How to simplify the calculation of SC STAP and apply it into surface wave beamforming by appropriate constraints is worth study in the future.

The beamforming methods can be investigated on WERA raw data, which are the range-resolved data from each antenna. I have implemented the beamforming methods on WERA raw data. However, the generated DS using my beamforming process contain much higher noise than the DS beamformed by WERA software, mainly because the calibration step is not implemented. Beam-formed data are also available from WERA, but it has been gridded so the range-Doppler isolation is very inconvenient to reconstruct. The main aim of this thesis is to explore mitigation methods and diverting attention to beam-forming in order to also test these methods with WERA data was not possible in the time available. So the dataset used for the development of algorithms in chapter 4-6 is the Pisces dataset only, which have been beamformed in the hardware already.

Chapter 3 Properties of two main sources of clutter

In this chapter, two main contaminating clutters for the wave estimation, i.e. ionospheric interference and radio frequency interference, are investigated in terms of their physical formulation, mathematical model and properties. In this thesis, 'clutter' is a technical term referring to unwanted non-sea signals. More specifically, clutter will refer to both unwanted echoes, i.e. received radar waves backscattered from some non-sea source, and signals that have nothing to do with the signal transmitted by the radar, i.e. radio services.

3.1 Ionospheric clutter

Usually, the radio waves of HFSWR propagate along the surface of the ocean, but not all of the energy emitted by the radar propagates horizontally. There is also some skywave propagation resulting from the imperfect antenna system or irregular ground paths. This part of energy is directed upwards. HF radio waves, under certain conditions, may be reflected from different layers in the ionosphere depending on the electron density of them. In some cases, the energy reflected from the ionosphere returns to the radar receiver. For the ocean remote sensing purpose, echoes and radio frequency interferences propagated via the ionosphere are an unwanted signal, which is collectively called ionospheric clutter or ionospheric interference (II).

3.1.1 Description of the ionosphere

The ionosphere is the region of the Earth's atmosphere in the height range from 50-2000 km above the sea level. The physical property of this region is attributed to the radiation such as ultraviolet light from the Sun and cosmic rays from the universe, which ionize the atmospheric gases. Free electrons and ions exist in high density so that they can affect the properties of electromagnetic waves that are propagated within or through them. The profile of the ionosphere is considered as several stratified layers, denoted as D, E and F. This stratified structure of the ionosphere is highly variable and is often utilized for ionosphere-dependent communication over long distances. Rush (1986) presented a review of the properties of different layers in the ionosphere. The properties of D, E, and F layers, in an increasing order of their altitude, that he identified are briefly summarized in Table 3.1.

$N(z)$ in Table 3.1 denotes the electron density, which is a function of altitude z. $N(z)$ is an important feature of the ionosphere, which determines the reflection and refraction ability of each layer for HF radio wave propagation, and determines the II in the Doppler spectra. More specifically, this property is

described by the critical frequency f_0 of each layer, which defines the highest frequency of radio waves that can be reflected, or in another words, the lowest frequency that can be refracted through that layer, when the radio waves are propagated in the vertical direction. The critical frequency is related to the maximum electron density in the layer by the equation:

$$N_m = 1.24 \times 10^{10} \bullet f_0^2 \tag{3.1}$$

where N_m is the maximum electron density in the layer and f_0 is the critical frequency in MHz.

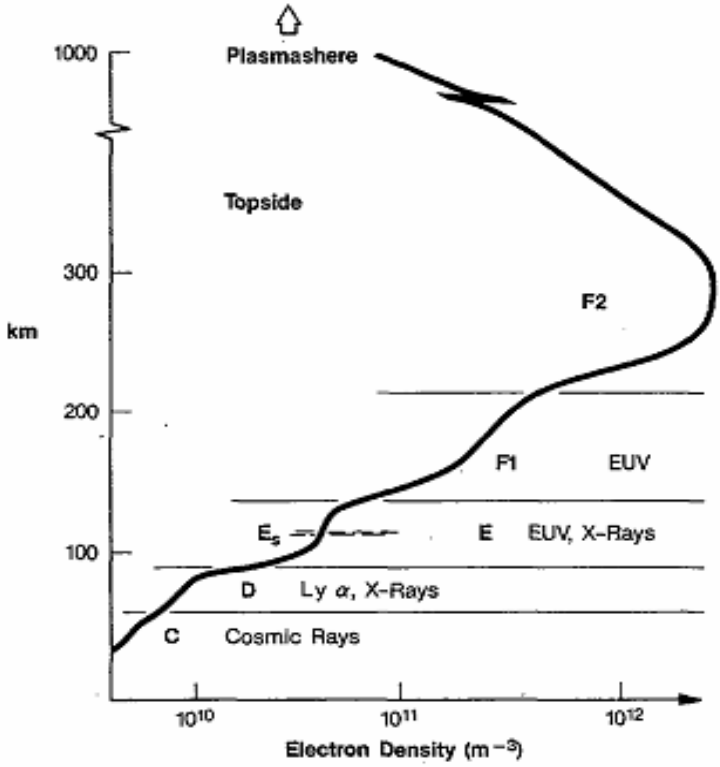


Fig.3.1. Vertical distribution of electron density in the ionosphere (Rush 1986). It is characteristic of mid latitudes at summer noon, solar maximum conditions.

Fig.3.1 shows a simplified vertical distribution of electron density in the ionosphere. Based on the understanding of properties of each layer in Table 3.1, their impact on the propagation of HF radio waves in 3-30M Hz and thus II is summarized as follows. The D layer does not contribute to the II, because it doesn't achieve sufficient electron density to reflect or even significantly refract HF radar waves. The key impact of the D region on HF radar systems is to absorb the transmitted radio waves in the day, via high collision frequency between the electrons and neutrons. This collision effect in E layer is not as dominant as in the D region. As a result, the E layer can reflect radio waves

lower than about 10 MHz and partly refract waves above 10 MHz. There is an additional layer known as the sporadic E, which appears in the altitude with maximum density within the E layer. Only the Es layer and F2 layer exist and have enough density to reflect HF radio waves at night. Beside reflection, E layer absorbs and refracts radio waves. If the refracted radio waves continue propagating upwards to the F layer, they will be either reflected by the F1 layer to reach the earth's surface at much further ranges, or absorbed by the F2 layer. The highest F layer is the region of the biggest electron density. The F1 layer has sufficient electron density to reflect radio waves in the high frequency band in the daytime. The effective collision frequency in the F2 layer is high. Therefore, the F2 layer shows more absorbing ability than reflection for HF radio waves.

Table 3.1 Properties of D, E, and F layers in the ionosphere. $N(z)$ is in unit (cm^{-3}).

Layer	Altitude/km	Appearing time & $N(z)$	Physical property
D	50-90	Daytime $N(z)$: $10^3 \sim 10^4$	electron density increasing rapidly with altitude; under strong influence of the sun; with maximum electron density occurring near local noon during summer ionization in 50-70km due to cosmic ray ionization in 70-90km due to solar X ray
E	90-140;	Mainly in daytime $N(z)$: $>10^4$	relevant to the solar activity; formed by ultraviolet radiation ionizing atomic oxygen; with maximum electron density occurring near noon and summer less collisions of particles than that in D layer;
Es	90-120;	From 10 minutes to a few hours in both day and night $N(z)$: 2×10^5	weakly correlated with solar activity 100m-2km thick and short-term; irregular, patchy, smooth and thin ionized gas; varying with location and high latitude. able to reflect HF and even VHF radio waves
F1	130-210	Day in summer $N(z)$: $10^5 \sim 10^6$	Disappear at night and winter directly controlled by the solar radiation maximum density appear at 1pm in summer
F2	250-500	All the time $N(z)$: 5×10^6	density is high in day and low at night; high in winter and lower in summer; influenced by neutral-air winds, electro-dynamic drift, ambipolar diffusion, and ionization day-to-day variation due to geomagnetic field

3.1.2 **Impact of ionosphere property on II signal processing**

HF surface wave radars treat the ionospheric echo as clutters. Both skywave and surface wave radar data can be severely contaminated by backscatter from the non-stationary ionospheric propagation channel, which leads to the failure of remote sensing of the ocean surface. Anderson and Abramovich (1996) discussed the diversity of physical mechanisms responsible for the non-stationarity for skywave radars, which can also be a problem of ionospheric echo in the HFSWR. The problems in the II signal processing in HFSWR can be summarized below:

- Nonstationarity over hour, day, season, and year increases the difficulty for the modeling of II
- Multiple simultaneous propagation modes cause different ionospheric Doppler shifts
- Large-scale spatial structure leads to the large spread of II in the Doppler domain
- Solar illumination gives rise to the day-night change of II in the DS
- High angle rays bring strongly range-dependent Doppler shifts
- Transient echoes from meteors, aurora, and other ionospheric sources also contribute to the clutter in the DS.

Given these problems, a number of signal processing algorithms are targeted at individual forms of signal degradation for skywave radars, especially the smearing due to phase path variations (Netherway et al. 1989), multimode (Anderson and Abramovich 1995), impulsive noise (Turley and Netherway 1990) and meteor echoes (Anderson and Godfrey). II Suppression algorithms for surface wave radars have also been exploited (Tian et al.; Wan et al. 2006). These techniques when applied individually can tackle the problems of phase path variation, multimode or impulsive noise successfully to some extent. However, if more than one problem occurs, none of these solutions solves all the problems simultaneously. A general algorithm which can optimize II-mitigation in some global sense is developed and proposed in chapter 5, without the need of analyzing and solving the problems due to diverse underlying physical mechanisms.

3.1.3 Impact of ionosphere properties on Doppler shift of II

Linear and nonlinear variation of phase path

The major reason that II gives spread Doppler shift is that the ionosphere is time-varying (Howland and Cooper 1993). As mentioned above, the electron density in the ionosphere determines the altitude of the reflecting point in the ionosphere. When the reflecting point moves upward or downward with time, the propagation path will change, and the Doppler shift of the ionosphere echo will

change as well. This Doppler shift Δf is formulated by: $\Delta f = -\frac{2f}{c}\upsilon\cos\theta$, as shown in Fig.3.2, where $\upsilon = \frac{dh}{dt}$ is the up-down velocity of the reflecting position; f is the radar transmitting frequency; c is the speed of light; θ is the incident angle; h is the reflecting height.

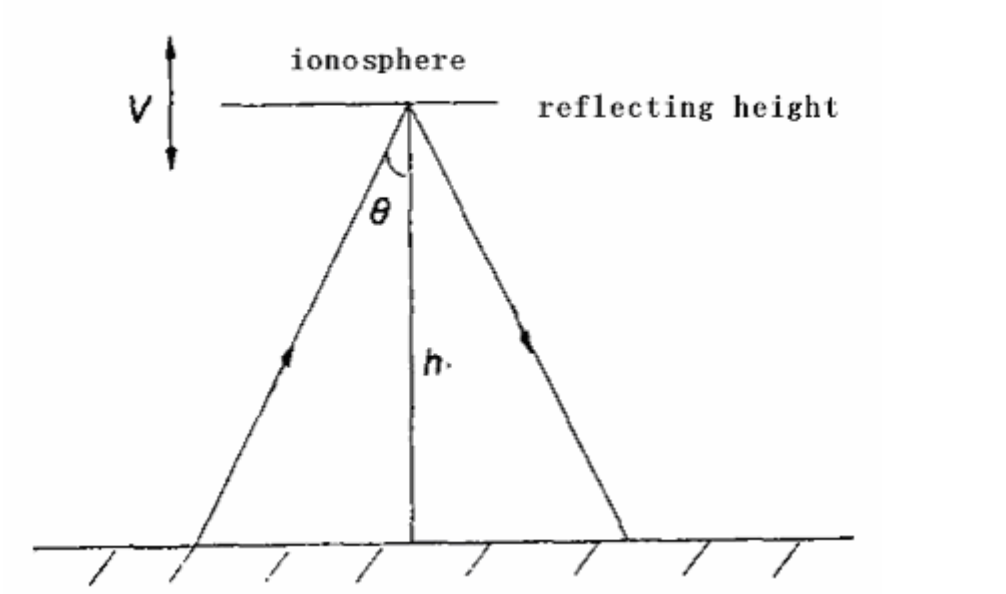


Fig.3.2 The Doppler shift of ionosphere interference generated by the vertical movement of the ionosphere (this figure is highly idealized and not the geometry of the radar).

Using this model, when the ionosphere is stable over a short time, say 10 s, the up-down speed υ is assumed constant. The typical Doppler frequency of ionosphere echo reflected from the Es layer is 0~0.1 Hz, from the F2 layer is 0.1~0.42 Hz, and from F layer is around 0.1 Hz. Since the coherent integration time used for sea sensing is around 200s, the ionosphere cannot be assumed moving with a constant speed. The phase path of the II is not linearly varied, so the Doppler frequency of ionospheric echoes will spread.

Multiple modes

When HF radar is working, any reflecting point in the ionosphere that connects the radar and the scattering sea area corresponds to a radio propagation path. Different reflecting points cause the difference in amplitude, phase, and Doppler shift of the II. During a long coherent integration time, the reflecting points move with changing speed, which contributes to the Doppler dispersion. Another parameter that can determine the reflecting point and thus the propagation time is the angle of incidence, θ. A larger θ usually leads to a higher reflecting point and a longer path, and the resulting II will appear at a greater range in the

Range-Doppler image.

3.1.4 Ionospheric interference in Pisces data

The ionospheric interferences coming from ionized layers may include specular or spread components. The specular components occupy small numbers of range and Doppler bins while the spread components can be large in both Doppler and range domains. II can also be a short term effect if it is due to the sporadic E layer. Checking the Pisces dataset, II are mostly found in RD images at night. The reason for rare II in the daytime is that the D layer absorbs most skywave propagation at that time (the use of lower HF band in Pisces may be a factor for rare II). In the nighttime, the D, E, and F1 layers, which are directly controlled by solar radiation, disappear. Without the D layer, the absorbing layer is gone. The possible ionosphere layers at nighttime are the sporadic Es and F2; the latter mainly absorbs sky-waves. Therefore, the main source of II in Pisces data is the Es layer. The II found is limited to one or two hours between 18:00:00-23:00 and 00:00-06:00, Doppler-correlated around rb 5, 10, and 15. It is also irregular, either spread all over the Doppler bins, or a small number of bins. These two features are a manifestation of the Es layer, because Es layer is short-term and irregular.

In conclusion, the variability of the ionosphere occurs for four possible reasons: the diurnal cycle of the D, E, and F layers, meteor trails, seasonal changes and the 11-year solar cycle. The properties of II are complex because of the diverse and changing properties of the ionosphere. Some II can persist for a long period of time, whereas others last for a short time. II in some occasional cases may have a comparatively higher Doppler shift and amplitude, while in others it does not. Some may be of high directivity and some are not.

3.2 **Property of radio frequency interference**

Radio frequency interference (RFI) produced by short-wave communication and broadcast is one of the major problems in remote sensing by HF radars, since it may mask the sea echoes in the first and second order region and thus degrade the quality of sea state estimation. Typically, RFI exhibits a nonstationary spatial structure, statistically characterized by a time-varying spatial covariance matrix within the radar CIT. The propagation media of the ionospheric layers may account for this physical phenomenon, in particular, the ionosphere D layer. The D layer usually plays the role of a shield against RFI in the daytime, since it vanishes in the nighttime. However, this is not always the case. In the Pisces dataset, dense RFI are found mainly in daytime, even just in two fixed hours, 10:00am and 13:00pm (this is partly due to radio frequency selection procedure), which implies that some radio station was broadcasting at that time period. RFI is also time-varying due to the variation in geometry between radar receiver and interference sources, and the impulsive nature of the sources.

Zhou (2005) proposed an analytical derivation of the RFI signal in the radar receiving data using a FMCW waveform, which has shown that after the first time and second time FFT, RFI are strongly correlated in the range domain. One single frequency RFI is manifested as two symmetric lines with respect to zero Doppler frequency along all range bins in the RD image. The RFI extends to all range bins with the same order of power while the sea echo's Bragg peaks attenuate proportionally to the fourth power of the ranges (Barrick 1972). Therefore, the correlation function of the interferences in the range domain is utilized for RFI cancellation. An investigation of the mitigation on RFI in the RD image together with II is presented at length in chapter 5, with detailed theoretical analysis and experimental validation. The property of RFI property is only briefly introduced here and shown by an example in Fig. 3.3, where RFI spreads over all the range bins.

Beside the II and RFI, there are some minor sources of clutters, such as transient lightning strikes and meteor reflections (in this thesis, meteor echo is not regarded as II), which contaminate a few sweeps in a CIT, and contaminate a small number of range bins (meteor reflections) or all the range bins within a sweep. Interferences due to human activities, e.g. moving targets like ships and planes are also clutters. Some of them are well localized in space, i.e. coming from the same direction and range.

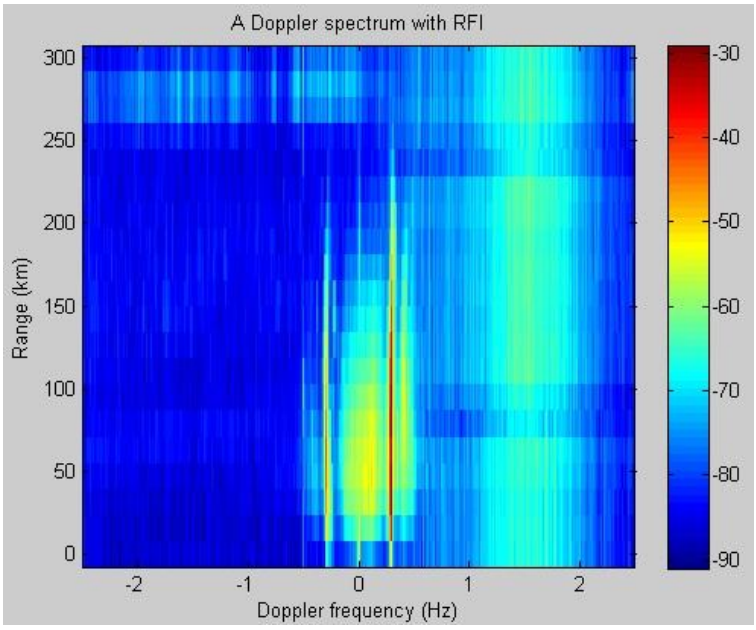


Fig.3.3. RFI in Range-Doppler spectrum. Strong RFI is located between Doppler frequency 1-2 Hz along all the range bins

Chapter 4 Review of signal processing schemes

Temporal and spatial inhomogeneities distort Doppler spectra and are a limiting factor for the measurement of oceanographic parameters. A lot of sources of clutter, e.g. atmosphere noise, radio frequency interference, ionosphere interference, wind turbines, measurement noise of the radar system itself, contribute to the unwanted inhomogeneities in a DS. In this chapter, various signal processing schemes are reviewed and explored in order to establish their abilities to minimize the impact of these external or internal influences on the quality of DS for ocean surface sensing. These schemes are used in detecting first order peaks, denoising additive noise, sea echo prediction, and supplementing missing DS.

Three signal processing schemes are briefly introduced in this chapter. Section 4.1 discusses the use of a wavelet analysis in the detection of Bragg peaks and denoising additive noise. Section 4.2 illustrates an application of an autoregressive model for signal prediction. Section 4.3 demonstrates an adaptive filtering system for supplementing missing data.

4.1 Wavelet analysis for Bragg peaks detection and denoising

4.1.1 Continuous wavelet analysis

Wavelet theory was first introduced in the 1980s and followed by several studies on its application in image and speech processing (Kronlandmartinet 1988) and ocean engineering (Gurley and Kareem 1999). In the past decades, it has become a prominent method in signal processing due to its novel multiresolution in time and frequency analysis. Wavelet analysis not only provides a frequency spectrum of a signal equivalent to the Fourier Transform, but also describes at what time the frequency is changed. In this method, a signal is decomposed by projecting it onto a scaled and translated version of a prototype mother wavelet. The mathematical definition is given as follows. The continuous wavelet transform (CWT) of a signal $f(t)$ is defined in terms of a mother wavelet function $\Psi(t)$ and two parameters, scaling a and translation b:

$$cwt_f(a,b) = \int_{-\infty}^{\infty} f(t)\Psi_{a,b}(t)dt \tag{4.1}$$

$$\Psi_{a,b}(t) = |a|^{-\frac{1}{2}} \Psi\left(\frac{t-b}{a}\right) \tag{4.2}$$

where $t-b$ determines the translation in time, $\frac{t}{a}$ determines the expansion of the wavelet function in scale. The two requirements of a mother wavelet function

are $\int_{-\infty}^{+\infty}\Psi(t)dt=0$, $\int_{-\infty}^{+\infty}\left|\Psi(t)^2\right|dt<\infty$, which imply that $\Psi(t)$ is a wave-like function which attenuates to zero (this is why 'wavelet' is named). Just as an infinitely oscillating sinusoidal wave is the basis function of the Fourier transform, $\Psi_{a,b}(t)$ - a translated and scaled $\Psi(t)$, is the basis function of a wavelet transform. The scale a determines the width of the wavelet. As the scale decreases, the wavelet gets narrower. It will include less of the time series, and the finer details stand out. On the contrary, if the scale increases, the wavelet is broad in time and therefore any local peaks in the time series get smoothed out. This property enables the wavelet transform to accurately detect signals that have spike-like features. Although the wavelet method is normally used for time dependent signals, in this thesis it will be applied to the Doppler spectrum for the detection of first order peaks

4.1.2 Discrete wavelet transform

The discrete wavelet transform (DWT) is another commonly used wavelet transform. It can largely reduce the computation load by using binary scales rather than continuous scales. According to Mallat (Daubechies et al. 1992; Mallat and Falzon 1998), the $L^2(\mathbb{R})$ space can be entirely and non-redundantly analyzed using the binary scale. The discrete wavelet transform (DWT) (Iyengar 1998), i.e. binary wavelet transform is defined as:

- Wavelet packet model, see Fig.4.1(a):

$$\begin{aligned} s_{AA_j}&=\sum_{n\in z}a(n-2k)s_{A_{j-1}},\ s_{DA_j}=\sum_{n\in Z}d(n-2k)s_{A_{j-1}} \\ s_{AD_j}&=\sum_{n\in z}a(n-2k)s_{D_{j-1}},\ s_{DD_j}=\sum_{n\in Z}d(n-2k)s_{D_{j-1}},\ j=1,2,...,J \end{aligned} \tag{4.3}$$

- Wavelet model, see Fig.4.1(b)

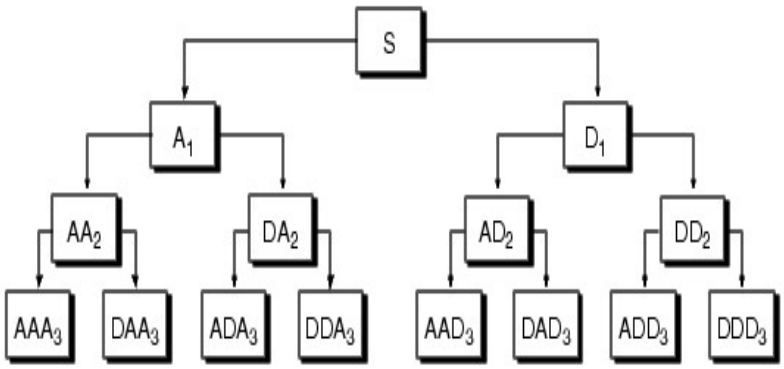
$$\begin{aligned} s_{A_j}&=\sum_{n\in z}a(n-2k)s_{A_{j-1}} \\ s_{D_j}&=\sum_{n\in Z}d(n-2k)s_{A_{j-1}},\ j=1,2,...,J \end{aligned} \tag{4.4}$$

where the original signal to be analyzed is $s(n)$; $a(n)$, $d(n)$ are coefficients of low-pass and high-pass filters; j represent the level of DWT and J is the maximum level of DWT; subscript A_j (the initial letter of 'approximation') and D_j (the initial letter of 'details') represent the low-frequency and high-frequency subsignal at level j respectively. The inverse wavelet transform

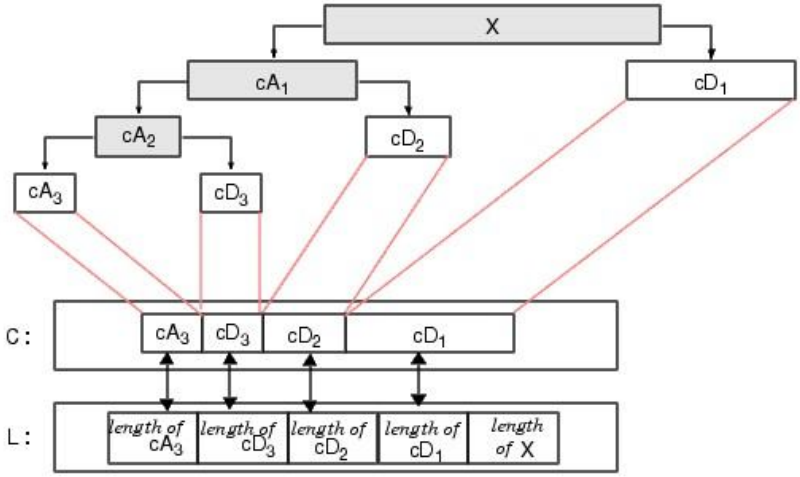
for Equ. (4.3) and (4.4) are defined in Equ. (4.5) and (4.6) respectively.

$$\hat{s} = \sum_{j=1}^{J} A^j + \sum_{j=1}^{J} D^j \qquad (4.5), \qquad \hat{s} = A^J + \sum_{j=0}^{J} D^j \qquad (4.6)$$

The difference between these two models is that the wavelet packet model offers a full decomposition on both 'approximation' and 'details' (see Fig.4.1(a)), while the wavelet model only decomposes 'approximation', i.e. cA_i in Fig.4.1(b).



(a)wavelet packet model



(b) wavelet model

Fig.4.1. 3-level DWT. 'A' is the initial letter for 'approximation', and 'D' is the initial letter for 'details'. (a): the wavelet packet model; (b): the wavelet model

There are three important issues to be addressed in wavelet analysis:

- Select a matched mother wavelet function
- Select a suitable model, i.e. wavelet or wavelet packet model
- Set the maximum level of wavelet decomposition

4.1.3 Discrete wavelet transform for Bragg peak detection

In radar signal analysis, Zhang (2007) introduced a method of wavelet transform in combination with independent component analysis (ICA) to detect targets masked in the time series of sea echoes. Instead of looking at the time series of sea echoes, here wavelet transform is applied to the DS for the detection of first order peaks. As explained in Chapter 2, the positions of the first order peaks are supposed to be $f_B = 0.102\sqrt{f_{radar}}$ in deep water. But in the actual measurements, the locations are shifted from the theoretical values, because the waves causing first-order scattering are superimposed on a sea surface which is physically moving due to surface currents. This offset of the Bragg peaks is used to estimate the radial component of the surface-current vector. In abnormal conditions, if strong clutters are mixed in the first order region and induce so called clutter peaks, the positions of the first order peaks will be confused with clutter peaks. This ambiguity will lead to miscalculations of the offset and incorrect current estimations. In addition, the wind direction will be mis-estimated, which results in error in the estimation of the ocean wave directional spectrum. Therefore an accurate detection of first order peaks is of significance.

The first order region in a DS is quickly varying while the second order continuum is slowly varying. According to wavelet theory, this spike-like part of the sea echo can be amplified by the wavelet transform while the much flatter part of the sea echo can be smeared out. After wavelet transform, the first order region is extracted. However, not only Bragg peaks but also clutter peaks in the first order region are extracted. To discriminate Bragg peaks from clutter peaks within the first order region, independent component analysis is introduced.

Independent component analysis is a technique used to separate independent sources linearly mixed in signals. Suppose that there are N independent signals $s_1, s_2, ..., s_N$ and N sensors to record the signals. The signal is recorded simultaneously at each sensor and denoted by $x_1, x_2, ..., x_N$, each coming from the N independent sources with different mixing ratios. The task of ICA is to separate the independent signals from the sensor signals. Cruces (2000) introduced an generalized kurtosis Fast-ICA method in the presence of noise. The advantages of it are one, the estimate is unbiased and two, the noise and signal interferences are minimized in the separated sources. Here，the ‘details’ at multiple levels are used as sensor signals. After ICA on them, the real Bragg peaks can be separated from the clutter peaks, due to the independence of clutter in both Doppler and Range domain. The algorithm uses the Cruces method for ICA in Matlab. The flow diagram for ICA is provided in Fig. 4.2, where A is the mixing matrix; S represents the matrix of independent sources; X is the

matrix of the mixture. $\hat{S}$ is the estimated matrix of independent components.

Note that an intrinsic feature of the ICA is the uncertainty of the order of the separated independent sources. This feature causes a problem of how to determine which IC contains Bragg peaks and which contains pseudo peaks. In the experiments, it has been found for many cases that the fourth IC contains the Bragg peaks and one of the other three ICs contains the pseudo peaks.

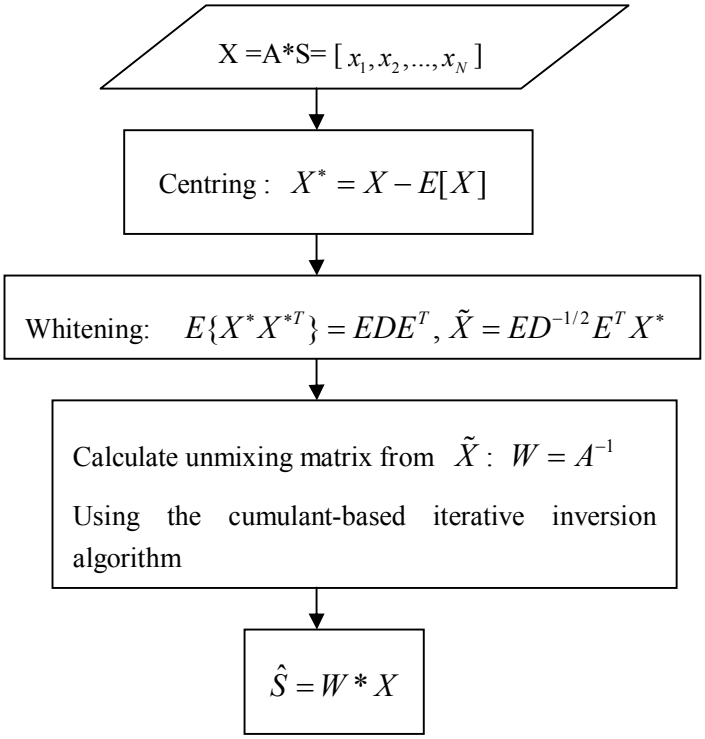


Fig. 4.2 Flow diagram of the ICA algorithm. A is the mixing matrix; S represents the matrix of independent sources; X is the matrix of the mixture; $\hat{S}$ is the estimated matrix of independent components.

In this application, the DWT setting is described as follows. 'db4' (Daubechies 2004), one of the Daubechies wavelet family of biorthogonal wavelet functions, is widely used in solving the problems of signal discontinuities, and selected as the wavelet basis function here. This wavelet model is selected because it is simpler than the wavelet packet model but sufficient enough for the separation of the first order region and the second order continuum (see Fig.4.5). The maximum decomposition level is chosen as 4 after testing the values from 1-7 (the maximum decomposition level defined by 'db4' is 7).

The procedure is summarised as follows:

1 A DS is chosen to be the original signal s.

2 The DS is decomposed by wavelet 'db4' at 4 levels.

2.1 $a(n)$ and $d(n)$, the coefficients of lowpass-filter and highpass-filter are calculated at each level in Matlab in terms of 'db4'.

2.2 'approximation' and 'details' are decomposed from s respectively using $a(n)$ and $d(n)$ at each level, under the wavelet model. ('approximation' contains the global information in the DS, whereas 'details' contains the local variations)

3 s is restored by choosing 'approximations' and 'details' at each level respectively (see Fig.4.4).

4 The first order region is extracted from sea echoes in the 'details' at level 4.

5 The 'details' at four scale levels are taken as sensor signals to the independent component analysis.

6 The output signals of independent component analysis are the independent sources, i.e. the Bragg peaks and clutter peaks are separated. The fourth IC contains the Bragg peaks and the first/second/third IC contains clutter peaks. The reason why this occurs is not clear at the moment, since the ICA algorithm has no particular meaning in the order of separated components.

A successful example following these procedures is described below. A DS is selected from rb 9, day 6, hour 20:10 (see Fig.4.3). It is shown in the DS that there are four peaks in the first order region. Before wavelet and ICA processing, I checked the positions of Bragg and clutter peaks in the RD image. Bragg peaks are distinct from clutter peaks in a RD image by that Bragg peaks are separated by twice Bragg frequency; correlated in range and occupy all range bins from 0 to 18; within 100 Doppler bins away from the zero Doppler frequency, while clutter peaks are correlated only in a few range bins. Fig.4.4 shows two strong clutters at Doppler bin 441 and 593 respectively at rb 9; the Doppler indices 458 and 574 are the real positions of Bragg peaks because they are correlated with range and occur within 100 Doppler bins beside 512 Doppler bin (zero Doppler frequency) . The frequency distance between Doppler bin 458 and 574 is 0.56 Hz, which is twice the Bragg frequency 0.28Hz, i.e. $f_B = 0.102\sqrt{f_{radar}}$ where $f_{radar} = 7.618$ MHz. Note that if the frequency separation is not twice Bragg frequency, this DS is abandoned as invalid data. By 4-level DWT decomposition using the Equ.(4.4), 4 pairs of 'approximation' and 'details' are generated, i.e. each pair is $\{cA_i, cD_i\}, i = 1,2,3,4$ as shown in Fig.4.1(b). Then by DWT reconstruction using $cA_i, i = 1,2,3,4$ and $cD_i, i = 1,2,3,4$ at each level respectively, eight signals are reconstructed, called 'approximation' at level 1-4

and ‘details’ at level 1-4. When Doppler spectra from 20 range bins are processed in this way one by one, four pairs of RD images after wavelet analysis are shown in Fig.4.5. The four RD images on the left and the right are reconstructed using ‘approximation’ cA_i and ‘details’ cD_i respectively. The bottom right plot in Fig.4.5 shows the signal ‘details’ restored at level 4, where the first order region is extracted without the second order continuum. Then the Doppler spectra of rb 9 in the right four RD images of ‘details’ are used as sensor signals, i.e. the input of ICA. The number of separated independent components (ICs) in the outputs of ICA is four, because the number of outputs is equal to the inputs, i.e. the maximum decomposition level. In this application, as mentioned before, the Bragg peaks are identified in the fourth IC and the clutter peaks are shown in the third IC. Fig.4.6 shows the IC of Bragg peaks at the top and the IC of clutter peaks below. It is shown that the dominant peaks in each plot display the accurate positions of Bragg peaks and clutter peaks. Another example is chosen from RD image of 11am, 16/02/2005. In this case, I only show the results in Fig. 4.7-4.9 because the procedures are all similar. These two examples are typical for a number of tests, which are not going to be presented here.

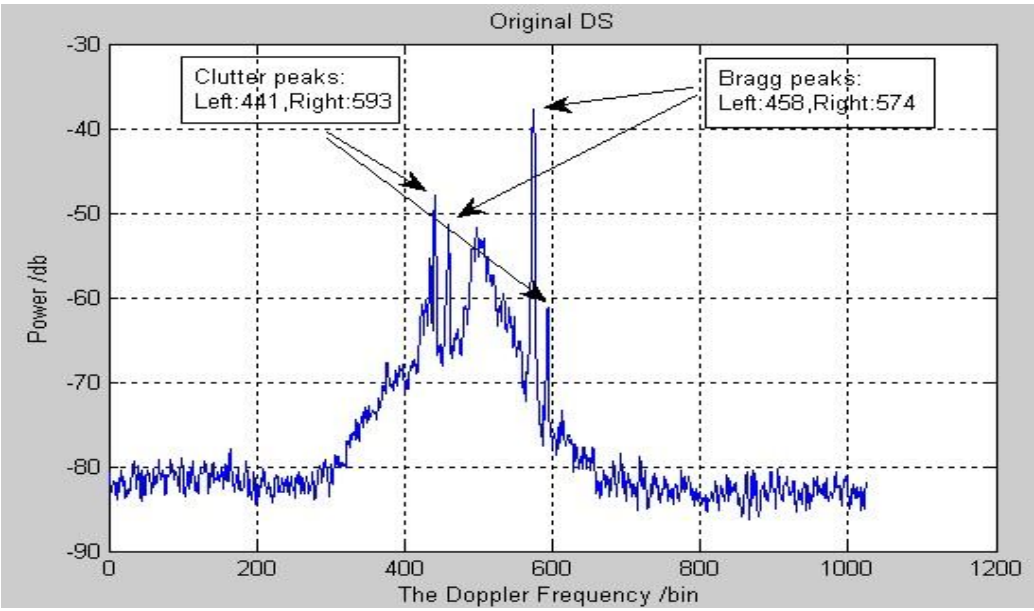


Fig.4.3 DS observed at 20:10, 06/02/2005. Four peaks in the first order region are pointed out by black arrows. The ‘441’ peak comes from clutter, but its amplitude is higher than the real first order peak at Doppler bin 458. Each Doppler bin represents 0.0048Hz. The frequency range is from -2.5Hz to 2.5Hz.

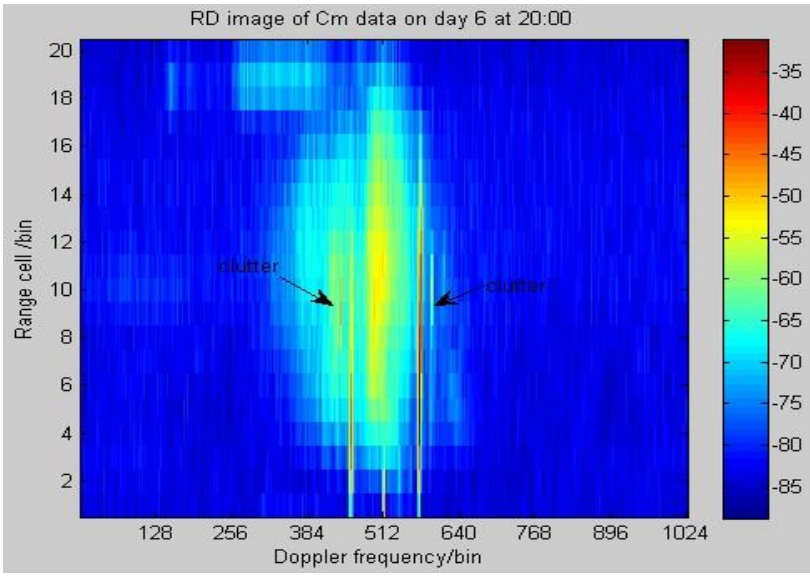


Fig.4.4 RD image of 'cm' data obtained at 20:10, 06/02/2005. The black arrows points at the clutter which leads to the faked peak in the DS of rb 9. Each Doppler bin represents 0.0048Hz. The frequency range is from -2.5Hz to 2.5Hz. Each range bin represents 15 km.

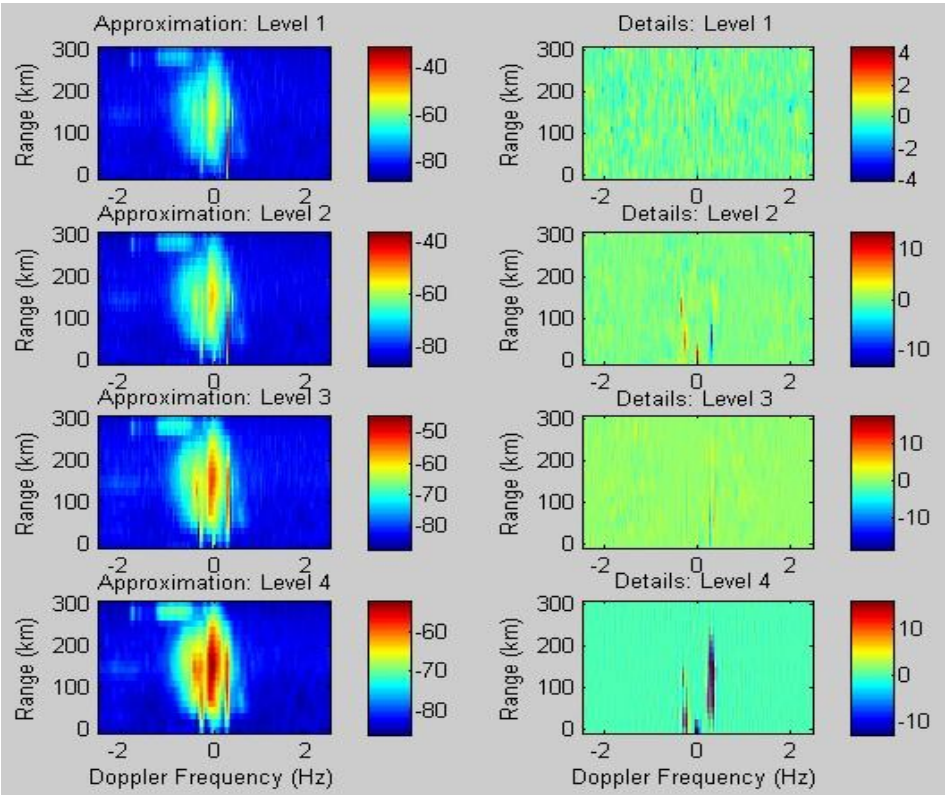


Fig.4.5 RD images of 'approximation' and 'details' by 4-level DWT. The four RD images on the left are reconstructed using 'approximation' at each level. The four images on the right are reconstructed using 'details' of each level. The first order region plus ground echo (zero Doppler backscatter) at short ranges is extracted in the subplot titled 'Details: Level 4'.

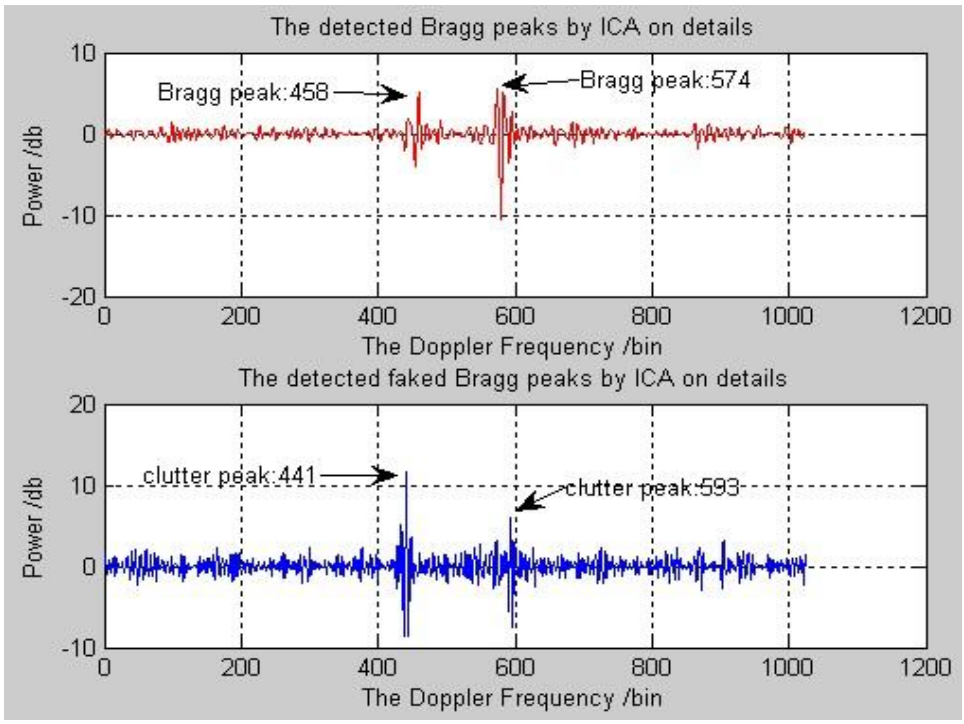


Fig.4.6. Two independent components in the output of ICA. In the top, the fourth independent component shows two peaks, which are the real first order peaks. In the bottom, the two dominant peaks in the second independent component show the positions of clutter peaks. Each Doppler bin represents 0.0048Hz. The frequency range is from -2.5Hz to 2.5Hz.

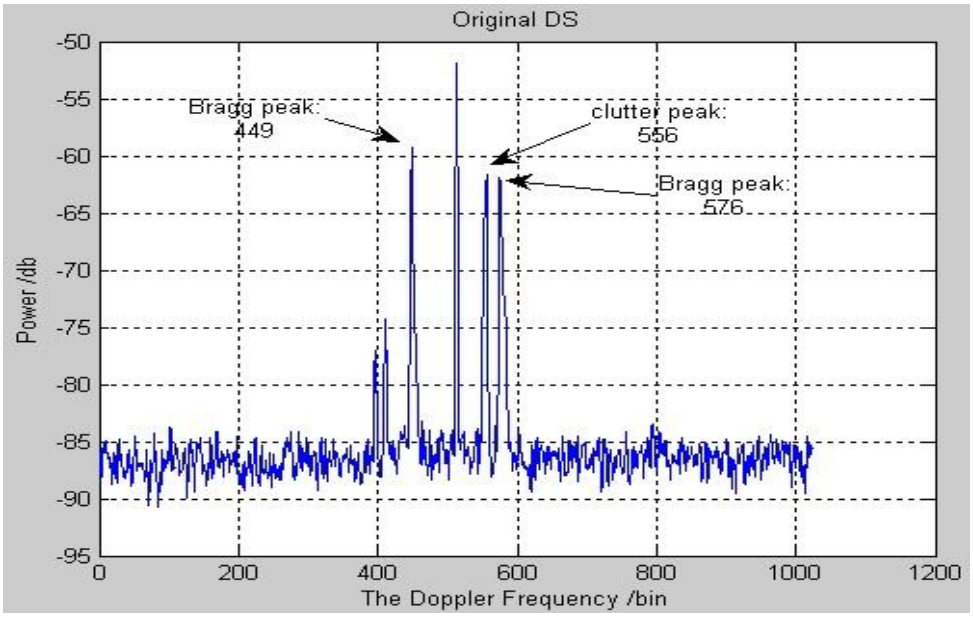


Fig.4.7 DS from rb 2 observed at 11:10, 16/02/2005. Three peaks in the first order region are pointed out by black arrows. The '556' peak comes from clutter, but its amplitude is larger than the real inferior first order peak at Doppler bin 576. Each Doppler bin represents 0.0048Hz. The frequency range is from -2.5Hz to 2.5Hz.

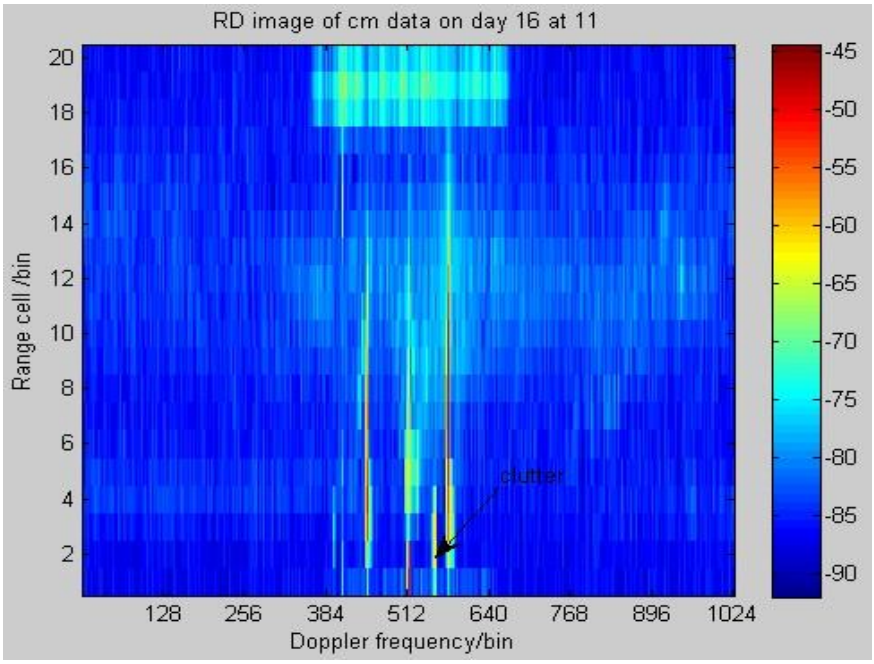


Fig.4.8 RD image of 'cm' data obtained at 11, 16/02/2005. The black arrows points at the clutter which leads to the confusing peak in the DS of rb 2. Each Doppler bin represents 0.0048Hz. The frequency range is from -2.5Hz to 2.5Hz. Each range bin represents 15 km.

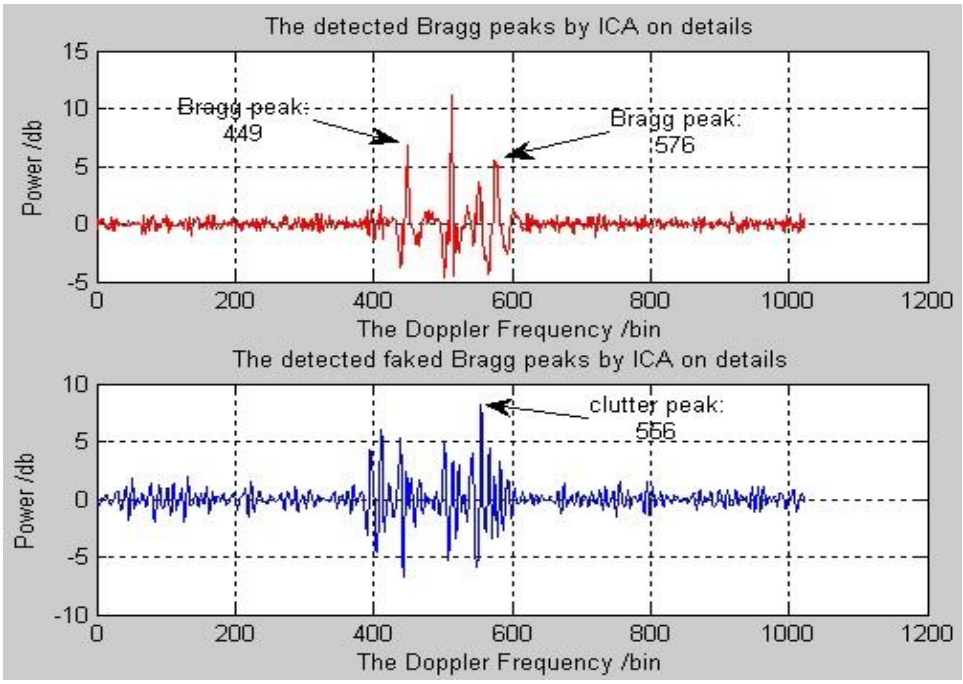


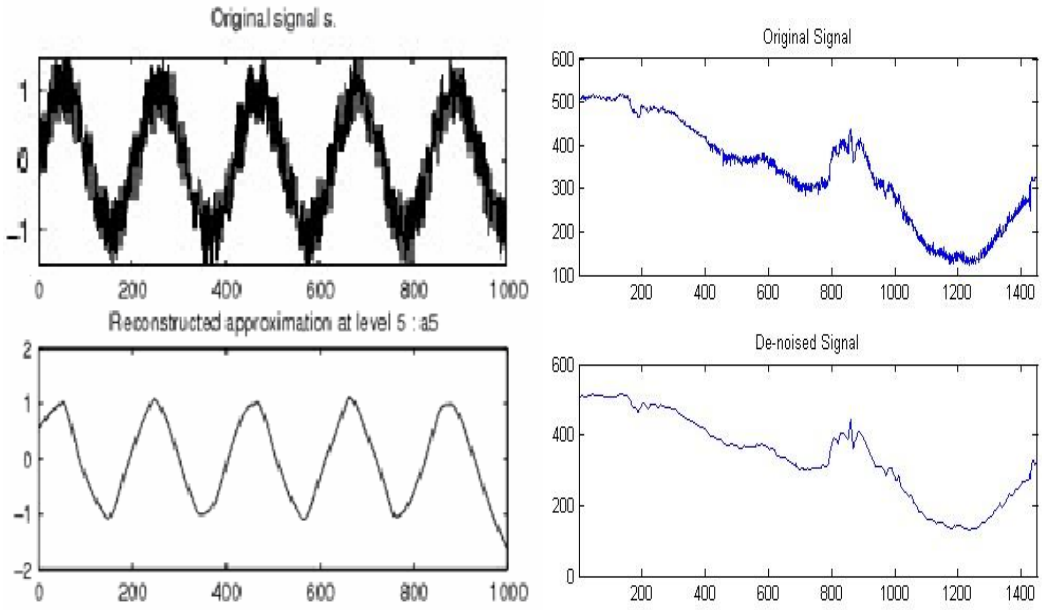
Fig.4.9. Two independent components in the output of ICA. In the top, the fourth independent component shows two peaks, which are the real first order peaks. In the bottom, the two dominant peaks in the second independent component show the positions of clutter peaks. Each Doppler bin represents 0.0048Hz. The frequency range is from -2.5Hz to 2.5Hz. Each range bin represents 15 km.

4.1.4 Discrete wavelet transform for denoising

Another common use of wavelet analysis is the recovery from noisy data, i.e. denoising additive noise. The procedures are explained as follows:

- A wavelet is chosen in the DWT and a signal is decomposed by the wavelet at level N.
- For each level from 1 to N, a threshold strategy (Guo et al. 2000) is selected and applied to the detail coefficients, i.e. $cD_1,...,cD_N \rightarrow c\hat{D}_1,...,c\hat{D}_N$. The coefficients of details under the threshold will be set to zero, while the ones above are kept.
- The signal is reconstructed using the original approximation coefficients of the level N, i.e. cA_N and the modified detail coefficients of levels from 1 to N, i.e. $\hat{s} = cA_N + \sum_{j=1}^{N} c\hat{D}_j$.

The noise added to a signal can be either white noise or non-white. The difference is that the power spectral density of the white noise is constant over all frequencies while that of the non-white noise is not. Fig.4.10(a)(b) show two examples for white and non-white noise denoising respectively (these two examples are copied from Matlab documentation). It is seen that the sinusoid function is recovered from the white noise while the highly perturbed electrical signal is recovered despite being perturbed by from noise with time-varying variance (the variance in the segment from 600-800 is higher than that in the segment from 1-200).



(a) white noise (b) non-white noise

Fig.4.10 White and non-white denoising. Top left: a sinusoid added with white noise; Bottom left: denoised sinusoid signal; Top right: an electrical signal added with non-white noise; bottom right: denoised electrical signal. Both figures are from Matlab 'help' menu.

In HF radar data, denoising is an important issue to be addressed because the signal level in far ranges is much lower than close ranges due to the attenuation. This can mask the first order peaks and limit the estimation of currents and wind. If the Bragg peaks in a DS from a far range is recovered by removing the variation of noise, the positions of Bragg peaks will be identified due to their more spike-like feature, which will make the current and wind estimation possible.

Here, the Matlab routine 'wden' is used as the denoising algorithm, which requires a few parameters: wavelet name, decomposition level N, threshold rule, threshold method. I have compared the denoising performance in experiments using various wavelets (see Appendix A), choosing levels from 1-N, choosing the four threshold rules. Two set of best parameters are chosen for denoising the DS in white and non-white noise cases. The parameters are given in Table 4.1.

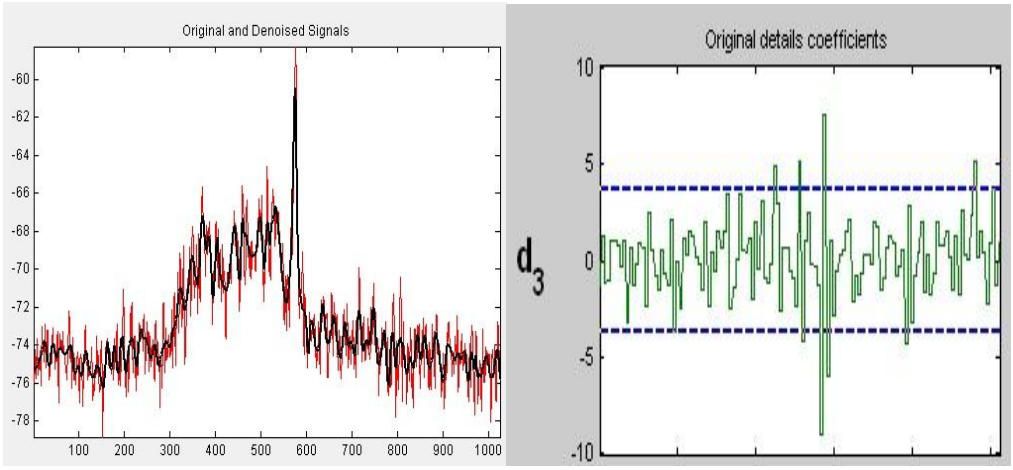
Table 4.1 Parameters for wavelet denoising algorithm

Noise type	Wavelet name	Level N	Threshold rule
White	'db7'	3	'sqtwolog'
Non-white	'rbio2.2'	3	'heursure'

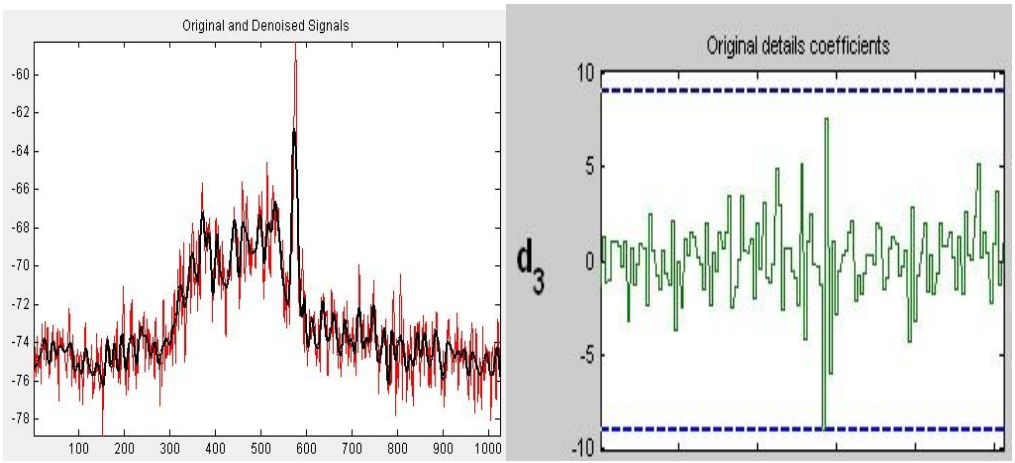
Note that: 'rbio2.2' is one of the reverse biorthogonal wavelet family; the three threshold rules are 'Stein's Unbiased Risk Estimate' (SURE), 'Fixed form threshold equal to sqrt(2*log(length(s)))' (sqtwolog), 'a mixture of 'SURE', and 'sqtwolog'' (heursure). The

definitions of them are given in Appendix A. SURE is a soft threshold, which works well for high SNR cases. If the SNR is very small, the SURE estimate is very noisy. 'heursure' is a combination of the former two methods that chooses SURE in high SNR and otherwise chooses the fixed threshold 2*log(length(s)). 's' is the noisy signal.

To facilitate the understanding of the threshold on coefficients of 'details' (noise), two examples are shown below in Fig.4.11 (In the denoising processing in Fig.4.12-4.14, the threshold is selected through the rule 'heursure' given in Table 4.1, which is conservative in high SNR and efficient in removing in low SNR). The original DS is a randomly chosen DS, called 's'. After wavelet transform, the signal 's' is decomposed as $s = cA_3 + cD_1 + cD_2 + cD_3$ (refer to the wavelet model in Fig.4.1b). Since the noise signal is always contained in the 'details' (high frequency component), cD_1 and cD_2 are both removed, i.e. the denoised signal $\hat{s} = cA_3 + cD_3$. cD_3 is kept because there is a peak in the Bragg frequency position. If it is removed, this essential information of Bragg peaks will be lost). For the cD_3, a low and high threshold values are set to demonstrate the impact of the threshold on the denoised signal. Fig.4.11 shows that the lower of the threshold, the more coefficients of 'details' kept.



(a) (b)



(c) (d)

Fig.4.11. (a)(c) Original and denoised DS ('black' is the denoised DS and 'red' is the original DS). (c)(d) The green curve shows the coefficients of cD_3, 'd3'; the blue dashed line represents the threshold. The threshold in (b) is lower than that in (d). Since the part of the green curve under the threshold will be set to zeros, there are more details kept in (b) (as seen in (a) at the Bragg peak frequency).

A noisy DS is selected from rb 13 in the RD image of 20:10, 06/02/2005. Checking the noisy DS in the left of Fig.4.11, it is found that the lower Bragg peak can easily be confused with the two peaks on its right. After non-white denoising, the recovered DS is displayed on the right of Fig.4.12. It is found that the SNR of the superior first order peak is increased by 5 dB and the inferior first order peak is identified due to the removal of the two noise spikes on its left. The original and the denoised DS were processed using Seaview, to identify the Bragg peaks, SNR of *sfs*, and SNR of *sfi*. It is found that the Bragg peaks are the same for both DS, i.e. 449 and 563, while the SNR of *sfs* is increased from 36.37 to 38.40 dB and the SNR of *sfi* is increased from 22.73 to 22.93 dB. To check if the Bragg peaks are correctly identified, the desired Doppler distance of the two peaks is calculated as $\Delta n_{Dop} = \frac{2f_B}{\Delta f_{Dop}}$, where Δf_{Dop} is the frequency in Hz per Doppler bin. Hz and f_B is the Bragg frequency. The radar working frequency is at 7.618 MHz, and so the Bragg frequency $2f_B$ is 0.5630 Hz. So $\Delta n_{Dop} = \frac{0.5630}{5/1024} = 115$. The estimated distance is (563-449), i.e.114. Hence, both Bragg peaks are correctly estimated. This example is demonstrated to show that

Bragg peaks become more ‘spiky’ after denoising due to the increased SNR.

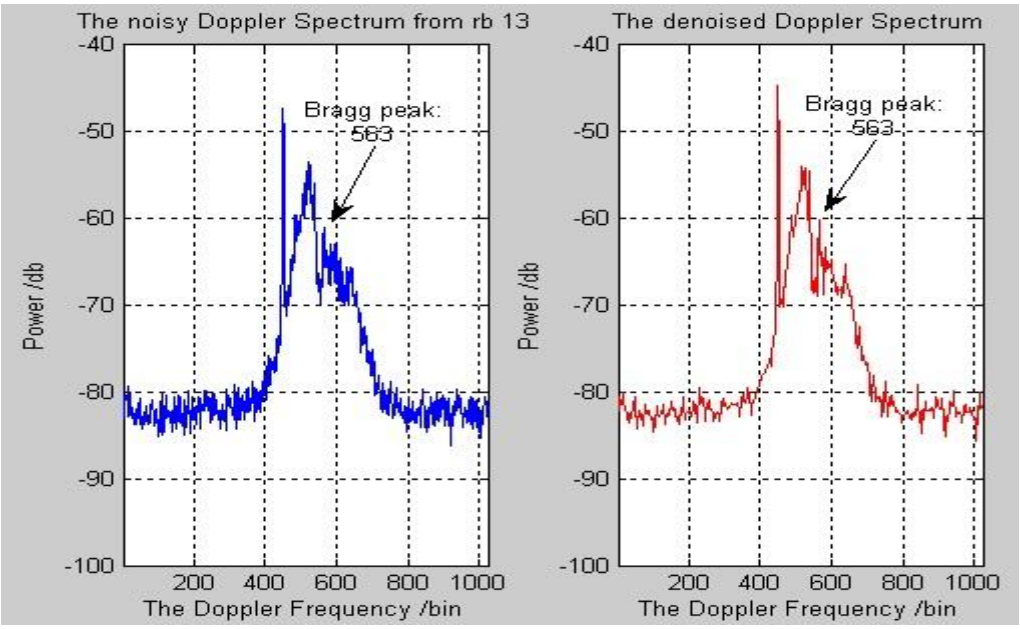
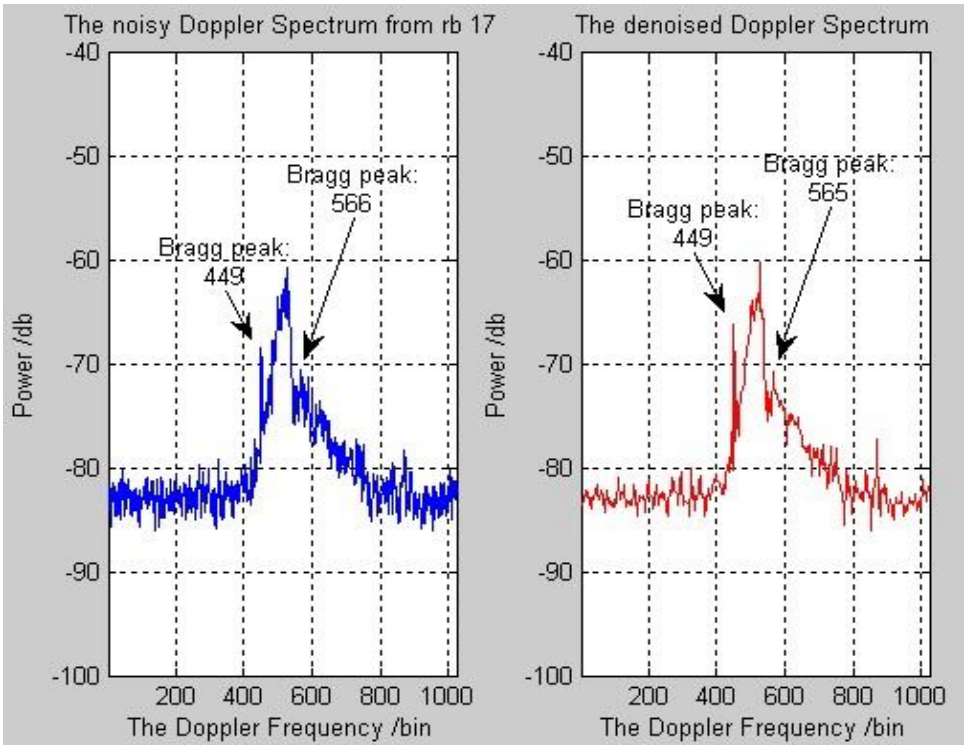
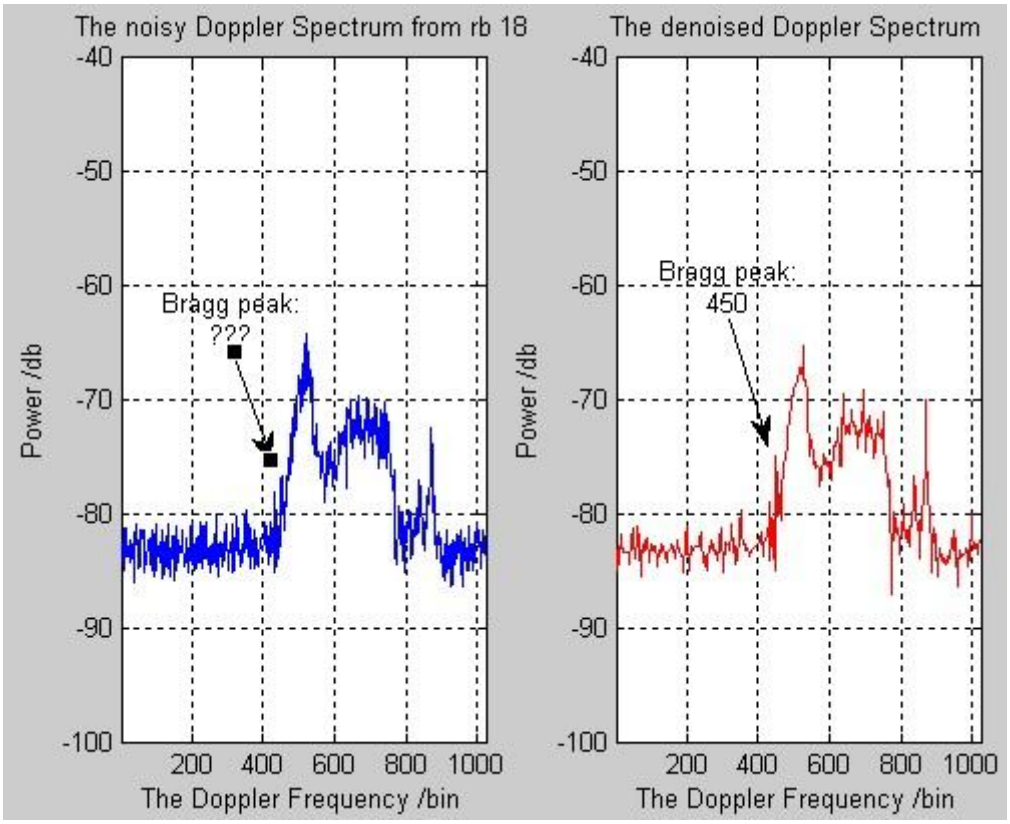


Fig.4.12 Left: original DS; Right: denoised DS. The noisy DS is obtained from rb 13, at 20:10, 06/02/2005. Each Doppler bin represents 0.0048Hz. The frequency range is from -2.5Hz to 2.5Hz.

Another example from further ranges and stronger noise is selected from the DS from rb 17 and 18 (255km and 270km offshore respectively) in the RD image of 20:10, 06/02/2005. After non-white noise denoising, the results are given in Fig.4.13(a)(b). It has been found by Seaview that the SNR of *sfs* from rb 17 is increased from 16.15 to 17.53 dB before and after denoising. The Seaview-identified Bragg peaks before and after denoising are the same for rb 13. For rb 17, the original Doppler distance of the peaks are (566-449), i.e. 117, while the distance of the denoised DS is (565-449), i.e. 116, which is closer to the desired value 115. For rb 18, the original data have no identified peaks, while the denoised have the *sfs* identified in 450 by Seaview.



(a)



(b)

Fig.4.13 Left: original DS; Right: denoised DS. (a) DS from rb 17, 255 km offshore; (b) DS from rb 18, 270 km offshore. Both DS are from 'cm' site, at 20:10, 06/02/2005. Each Doppler bin represents 0.0048Hz. The frequency range is from -2.5Hz to 2.5Hz.

The third example is selected from another day and hour, the DS from rb 13 in the RD image of 11:10, 25/02/2005. After non-white denoising, Fig.4.14 shows that the *sfi* detected by Seaview before and after denoising is very different.

The desired Doppler distance of Bragg peaks is 127 and it is 147 of the original and 127 of the denoised. This finding shows that the Bragg peaks are correctly identified by Seaview in the denoised DS. The white noise denoising algorithm is applied to all of the same data for comparison and its performance in identifying Bragg peaks is found to be worse than non-white because the noise beside Bragg peaks is Doppler-varying. So the plots are not presented here.

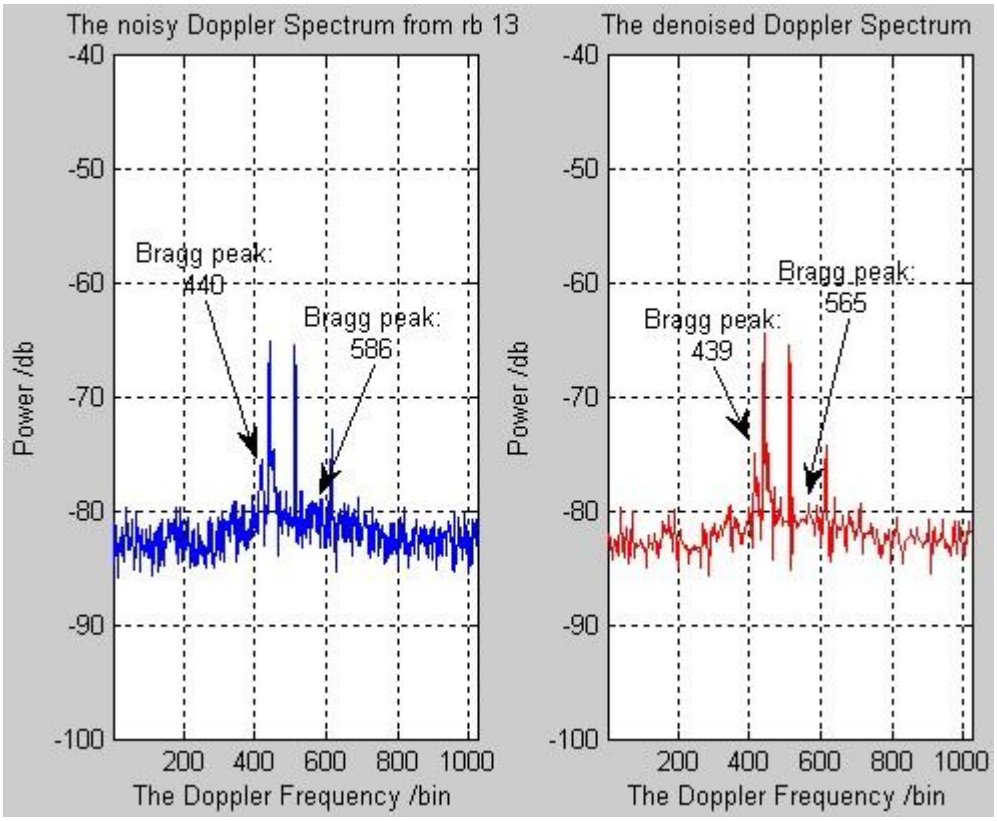


Fig.4.14 Left: noisy DS; Right: denoised DS. DS from rb 13, at 11:10, 25/02/2005 from 'cm' radar site. Each Doppler bin represents 0.0048Hz. The frequency range is from -2.5Hz to 2.5Hz.

4.2 Autoregressive modeling

In statistics and signal processing, an autoregressive (AR) model is a linear prediction formula that attempts to predict an output of a system based on the previous outputs. In 1927, Yule, a British statistician, introduced a model that used a linear regression method to search for periodicities in time series (Yule 1927). Yule's study was followed by Walker (1931) and the equations of AR model are called Yule-Walker equations in the form:

$$\hat{x}(n) = -\sum_{k=1}^{p} a_k x(n-k),\ e(n) = x(n) - \hat{x}(n)$$
$$x(n) = -\sum_{k=1}^{p} a_k x(n-k) + e(n) \tag{4.7}$$

where p is the order of AR model; $\hat{x}(n)$ is the predicted data of $x(n)$, using the previous p observed data $\{x(n-1),...,x(n-p)\}$; $e(n)$ is an error of this prediction. The parameters $a_1, a_2..., a_k,..., a_p$ are estimated using the least square mean by:

$$\varepsilon = E[e^2(n)] = E[(x(n) - \hat{x}(n))^2] = \min \tag{4.8}$$

If the signal $x(n)$ is a stationary stochastic process, then the coefficients $\{a_k\}$ are independent of time n.

Application for the prediction of the second order continuum

For remote sensing of the sea state, the power spectrum obtained from the AR model successfully provides stable spectral estimates from short-time datasets. Vizinho and Wyatt (1996) have compared AR modelling in the time series with the FFT for estimating the power density spectrum of HF radar backscatter. The property of less susceptibility to noise of a time-varying AR model has been validated in the application to nonstationary vibration signal analysis (Zhang et al. 2006). In this section, the DS is represented by a Doppler-varying AR model and examined in terms of whether it is less susceptible to noise than the original DS. More specifically, AR modelling is applied to predict the DS of sea echoes based on the estimated AR parameters.

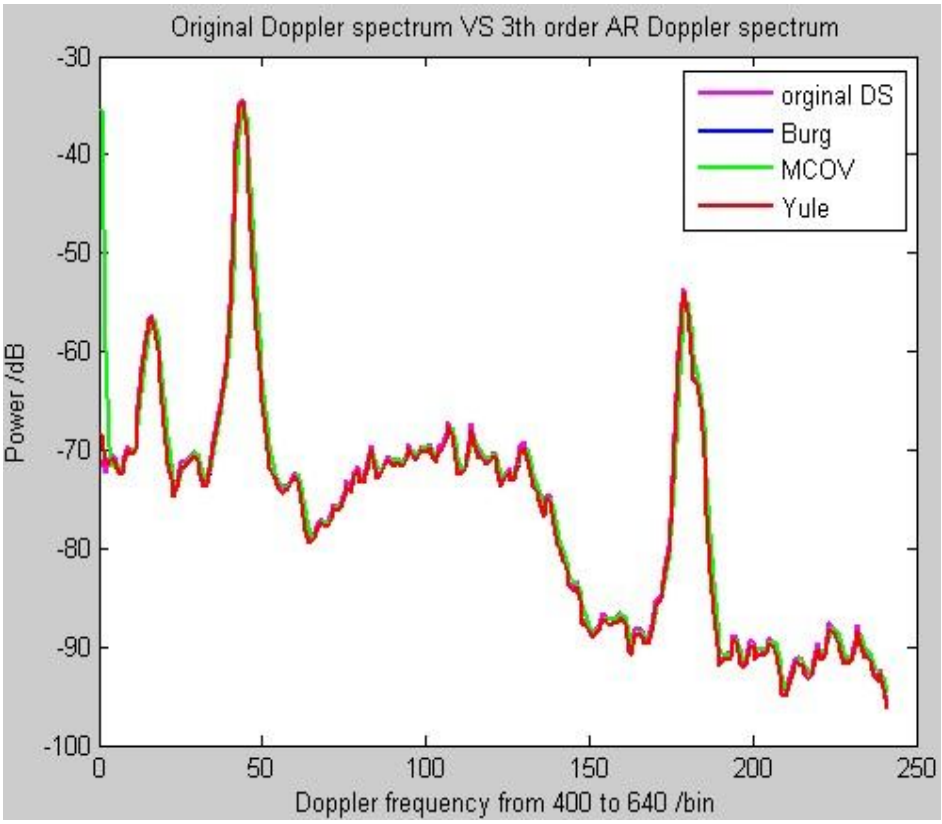
Let $x = [x(1),...,x(n),...,x(N)]^T$ be an N-element DS of ocean backscatter at one rb. x is modeled as the output of a p-order AR, driven by a white noise process $e(n)$. The objective is to compute the AR prediction of $x(n)$ by the past p values and p AR parameters, i.e. $\hat{x}(n) = AR(\{x(n-1), x(n-2),...x(n-p)\}, \{a_1,..., a_p\})$. There are two important issues to address in this prediction: calculating the AR parameters $\{a_1,..., a_p\}$ and choosing the optimal model order p.

The AR parameters can be estimated by Burg's algorithm, the Modified Covariance (MCOV) algorithm, and Yule's algorithm. The Burg method is a direct estimation of the AR coefficients. The modified covariance method is a least square linear prediction estimation. The Yule method is an efficient algorithm that solves the Yule-Walker equations by means of the Levinson-Durbin recursion. These three methods are compared and investigated comprehensively by Wear (Wear et al. 1995). There are several criteria for choosing the optimal order p in the autoregressive model. These methods include residual variance, final prediction error, Akaike information criterion, and minimum description length, all explained by Wear (Wear et al. 1995).

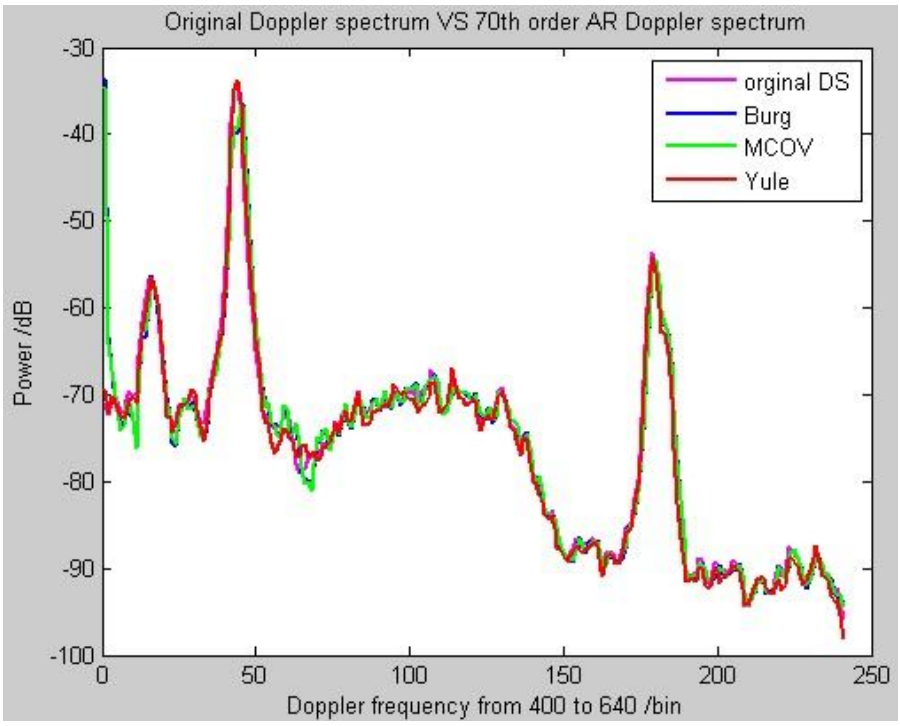
To select an appropriate method of parameter estimation and model order p for this application, experiments are applied to a randomly selected DS, obtained at 04:10, 01/02/2005. The order p is set as 3, 7, 20, 70 respectively for testing. The original DS and the predicted DS by three AR modelling techniques are shown in Fig. 4.15. It is seen that when the order is small, e.g. 3 or 7 or 20 (figures for p=7, 20 are similar to 3, so are not shown here), the errors in predictions of these three methods are all small (see plot (a)). But when $p = 70$, the errors from Burg's and MCOV's predictions are larger than that of Yule. The error is a measurement of the Euclidean distance between a segment of the original DS and the AR modelled DS. For Burg, MCOV, and Yule, the error is 29, 29 and 15 dB respectively. The errors in Burg's and MCOV's predictions also depend on the ratio of the order p to the length of the original signal. If the ratio is around or larger than 1/2, their performances degrade greatly, which can be seen in plot (c). The length of the original DS is 41 Doppler bins, and p is 20. The green (MCOV) and blue (Burg) curves tend to oscillate around the magenta one, but the red curve (Yule) is still smooth. From these results, the Yule estimation is found to be more stable than the other two techniques.
The three AR modelling techniques are examined with the power spectrum density (PSD) of the DS estimated by the classical periodogram using a Blackman window. In Fig.4.16, it is seen that the PSD of Burg and MCOV are less smooth than Yule for $p = 70$. The PSD of Burg and MCOV get smoother when p decreases.

From the experiments above, the relationship between the smoothness of PSD and the smoothness of DS is of interest. Looking at Fig.4.15 and Fig. 4.16, it can be seen that the smoother the PSD, the smoother the DS. Eventually, five features of the Yule method are identified from these Figs. First, the PSD of Yule approximates the periodogram in the lower frequency band better than MCOV

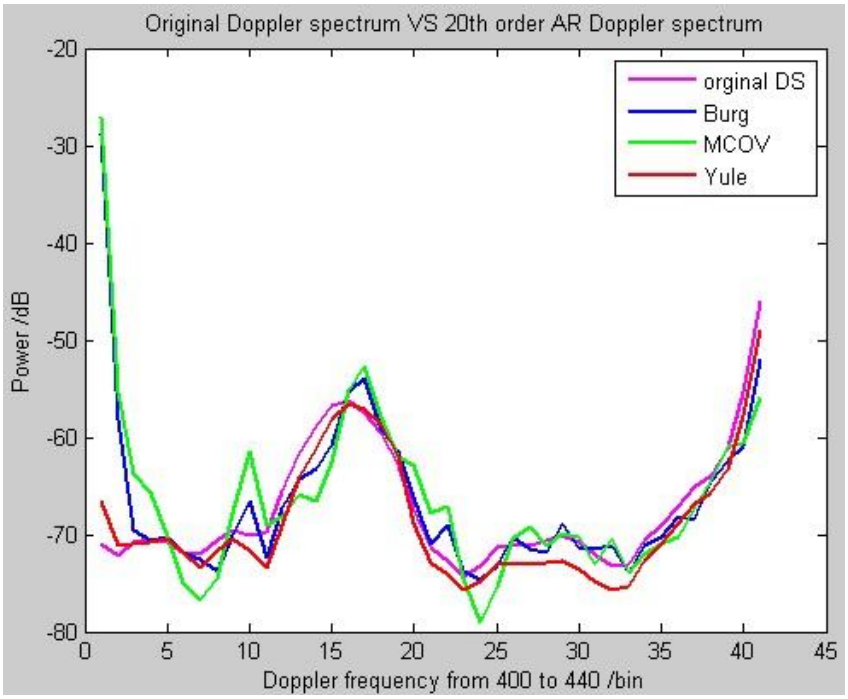
and Burg, which allows the Yule method to better preserve the main part of the original DS. Second, the PSD of Yule is higher than the periodogram and other two AR modelling methods over the higher frequencies band, which will distort the DS. At the moment, a clear criterion for an adequate fitting to the periodogram is not established, based on these two competing effects. Third, the PSD of Yule is smoother than periodogram, MCOV, and Burg, which indicates that the DS estimated by Yule is smoother than the original DS and the predictions by MCOV and Burg. This smoothing effect in the PSD may have a mitigation effect on the local variation of noise or clutters or signal. Fourth, Yule method is stably smooth regardless of p. Fifth, Yule shows better DC estimate than the MCOV and Burg, which is assumed due to the large bias of the first a few predicted values by MCOV and Burg which would change the mean of the predicted DS.



(a) 3-order AR model; signal length: 241

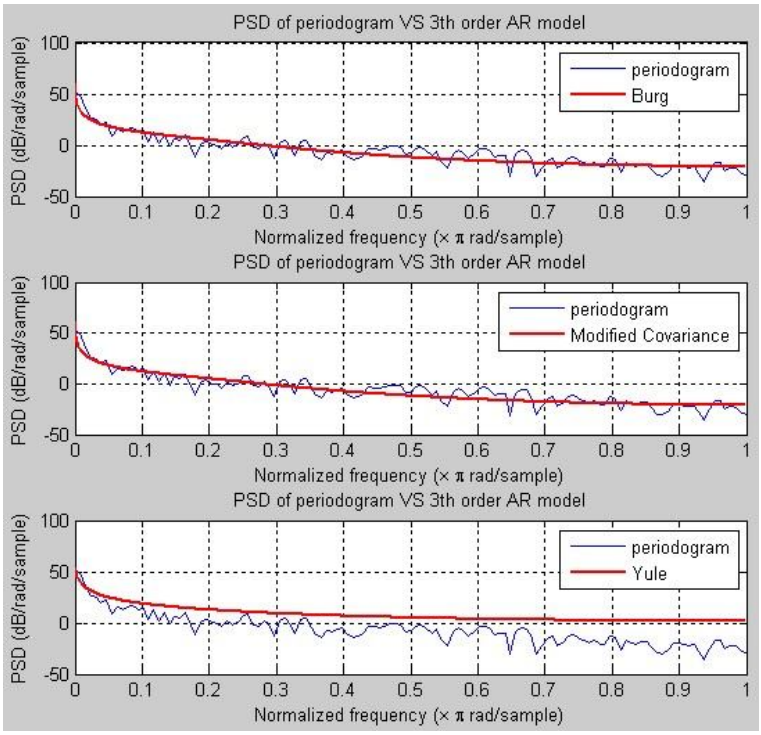


(b) 70-order AR model; signal length: 241

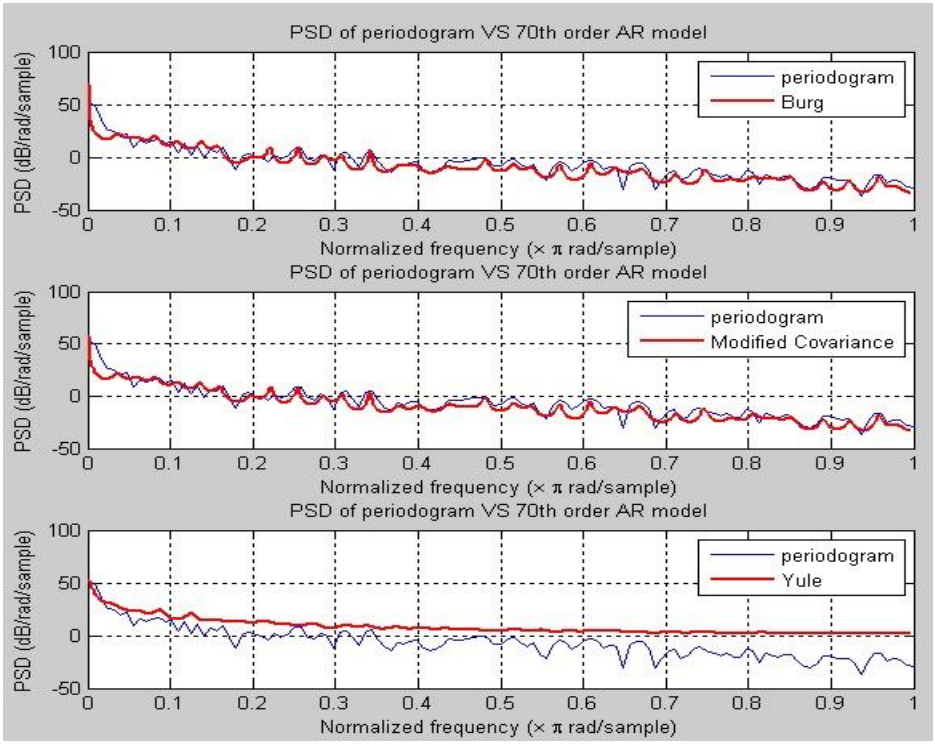


(c) 20-order AR model; signal length 41

Fig.4.15. Original DS from Doppler bins 400-640, (the first and second order region) compared with AR predictions by Burg, MCOV and Yule. The color for their curves is magenta, blue, green and red respectively. (a) 3-order AR model; signal length: 241 (b) 70-order AR model; signal length: 241 (c) 20-order AR model; signal length 41. Each Doppler bin represents 0.0048Hz. The frequency range is from -0.54Hz to -0.35Hz.



(a) 3-order AR model



(b) 70-order AR model

Fig.4.16. Power density estimations by periodogram using Blackman window, Burg, MCOV, and Yule. Red line shows PSD estimated by AR model and blue lines shows the classical periodogram.(a) 3-order AR model; (b) 70-order AR model

With the features identified above, Yule method for AR modelling is selected. The choice of order p will not be significant. AR order p is set to 7. The variation of this value has little change on the estimation of DS and also the wave estimation (this uses the DS). The improved estimation of the ocean directional wave spectrum by Yule DS is shown in Fig 4.17: the frequency spectrum derived by the new Yule DS looks better than the original radar estimation, although the direction spectrum is not changed much. Four parameters are used to quantify this improvement, shown in Table 4.2.

Table 4.2 Four parameters estimated by buoy, original DS, and AR DS. Significant waveheight h_s, mean period T_1, peak period T_p, and peak direction D_p.

Source	h_s /m	T_1 /s	T_p /s	D_p /°
Buoy	1.39	3.85	4.86	326.25
Original DS	1.04	5.575	5.35	297.18
AR DS	1.14	5.516	5.28	295.55

AR modelling has been tested on several data. The period and direction parameters are not changed much. As for the waveheight, some tests show some improvement while others don't. The example above is one of the good ones. The possible reason is AR modelling by Yule method can reduce local variation of the DS, which may be from sea echo or clutters. Yule AR model alone is not able to discriminate these two sources, but will smooth all. Therefore, this method requires a prior knowledge of the proportion of the clutter and sea echo in the segment of DS to be modelled. If the proportion of the clutter is much bigger than the sea echo, although the sea echo is mitigated along with the clutter, this trade-off will still improve the quality of DS. But this method still requires further in-depth exploration, e.g. the impact of overestimation of higher frequencies. Since the focus of this thesis is clutter mitigation not clutter smoothing for wave estimation, it is not discussed any more here.

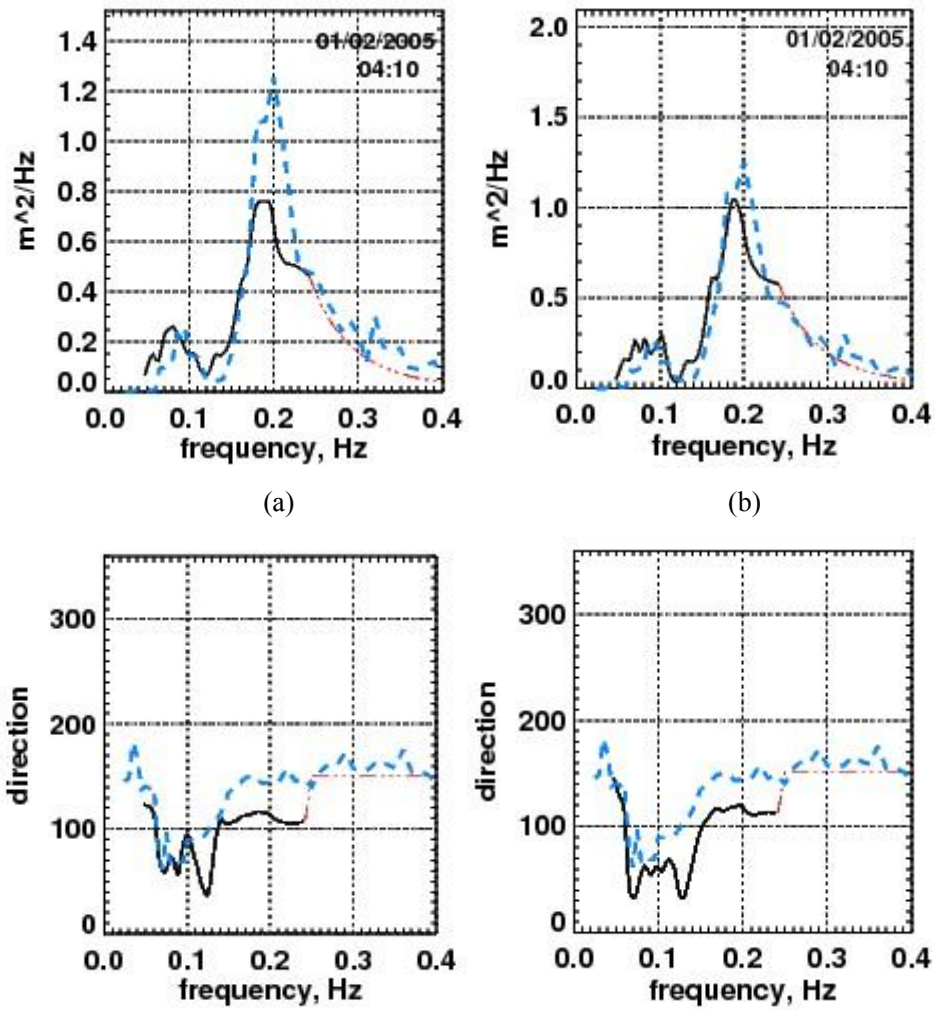


Fig.4.17. Comparisons of frequency spectrum and direction spectrum. (a) frequency spectrum: original (b) frequency spectrum: Yule; (c) mean direction: original (d) mean direction: Yule (solid is the radar and dashed is the buoy). In (a) and (b), the smaller peak at frequency 0.08-0.1Hz comes from the swell and the dominant peak at frequency 0.2Hz signifies the wind-driven waves.

4.3 Adaptive filtering system for filling missing gaps

Adaptive filters, aiming at transforming information-bearing signals into "improved" versions, "adjust their characteristics according to the signals encountered" (Regalia 2005). They are often preferred over fixed-characteristic filters, which are unable to adjust to changing signal conditions. Adaptive filters are widely used in signal restoration and interference cancellation. In HF radar remote sensing, radar data are sometimes missing, due to temporary failure of the hardware or software. Clearly, when the observations are missing, no oceanic parameters can be estimated, which causes inconsistent detection of the ocean surface in time and space. To fill these gaps, an adaptive filtering system is established that has an input DS $x(n)$, reference DS $d(n)$, and an adaptive

weighting operator $W(n)$. The operator adjusts its value based on the optimization of an objective function J, which is a function of the input and reference signal. The output of this system is the simulated missing data. Fig.4.18 shows the block diagrams of this adaptive filtering system. A general description of this application of adaptive filtering system is explained below, with the specific technical details of filling gaps described in subsection 4.3.2.

4.3.1 Adaptive filtering system

Let $x(n) = [x_1, x_2, ..., x_n ..., x_{1024}]^T, 1 < n \leq 1024$, represent the received DS, i.e. a mixture of sea echoes, and/or clutters, and additive noise. Let $s(n) = [s_1, s_2, ..., s_n ..., s_{1024}]^T, 1 < n \leq 1024$, represent the DS which is missing and going to be interpolated. Let $d(n) = [d_1, d_2, ..., d_n ..., d_{1024}]^T, 1 < n \leq 1024$, represent the reference DS, which is assumed to be very similar to $s(n)$. $x(n)$ is modified by an adaptive weighting operator W to estimate $s(n)$, with the knowledge of $d(n)$. The output of the weighting system is denoted by $\hat{s}(n)$. The error of estimation, $e(n)$, is calculated by subtracting $\hat{s}(n)$ from the reference $d(n)$. The objective function J of this adaptive filtering system is used to find an optimal weighting operator W, such that the average square error $\langle \underline{e}^2(n) \rangle$ is minimized (Equ.(4.11)). An adaptive searching algorithm is employed for this minimization problem.

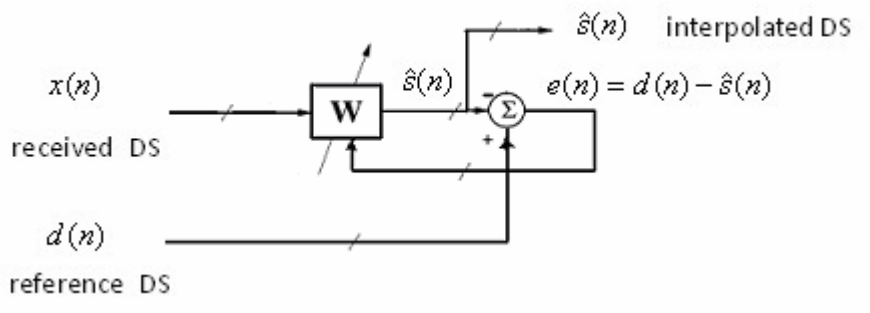


Fig.4.18 Block diagram illustrating the adaptive filtering system. $x(n)$ and $d(n)$ are the received and reference DS respectively. $e(n)$ is the error of estimation. $\hat{s}(n)$ is the output interpolated DS, which is missing. 'n' represents the Doppler index.

Ubiquitous among the different adaptive searching algorithms is the least-mean-square approach (Chambers 1994). The search starts with an initial solution. The weighting coefficients $W(n) = [w_1, ..., w_n, ..., w_N]^T$ are initialized with values of one. It then searches for the next solution generated by a small scalar change δ_n from the present solution and substitutes the solution (Lagarias et al. 1998). The formula goes like this:

$$w_n(l+1) = w_n(l) + \delta_n w_n(l), n = 1, ..., N \tag{4.10}$$

where l is the index of iteration times. The searching algorithm repeats this procedure until it finds the optimal weighting vector $W_{opt} = [w_1^*, ..., w_2^*, ..., w_N^*]^T$. It is called iterative improvement method when the substitution of the solution is allowed only where the objective function J improves. The algorithm uses the Lagarias (1999) method for the searching algorithm in the Matlab function named 'fminsearch'. The objective function is given as:

$$\begin{aligned} &J = \min_{W_{opt}} \{e^2(n) = |d(n) - \hat{s}(n)|^2\} \\ &such\ that: \hat{s}(n) = [w_1^* x_1, w_2^* x_2, ..., w_N^* x_N]^T \end{aligned} \tag{4.11}$$

4.3.2 Application to interpolating missing HF radar data

Although there are over 1000 papers covering a wide spread of applications of adaptive filtering, here I want to emphasize the special adaptation for

interpolating missing HF Doppler spectra for wave estimation. Generally in an adaptive filtering system, people have access to the received and reference signal. In this case of generating missing data, there is no access to $s(n)$, $x(n)$ or $d(n)$. The essential issue to address here is not the searching algorithm but the construction of the reference DS, $d(n)$, and the received DS, $x(n)$.

Construction of the reference Doppler spectrum $d(n)$

The reference signal $d(n)$ should be a DS with both good quality and the strongest similarity to the DS to be generated. An intuitive method is to search those Doppler spectra in adjacent hours to the missing data. There are two assumptions made in this method. The first is that the DS in adjacent hours have good quality. For example, if the data are missing because of strong contaminations of clutters, these clutters are assumed to be short-time events that damage radar observations in only a few hours. The DS observed in the remaining hours are not affected. The second assumption is the sea state in a few adjacent hours are rather stable and the wind direction doesn't vary much (the variation in wind direction will change the nature of the DS significantly), which indicates that Doppler spectra in adjacent hours have strong similarity to the missing data. To validate this assumption, both the sea state and wind direction in this application are examined. Fig.4.19 and 4.20 plot wind direction and waveheight observed by radar and buoy respectively for the whole month of February. In Fig.4.19, it is found that the wind direction is not always stable and varies with time sometimes. This implies that the assumption that wind direction doesn't change much should be checked individually in each experiment. Based on the time series of the waveheight in Fig.4.20, a further assumption can be made that the waveheight is stationary within three hours at the maximum by checking the autocorrelation function of the waveheight (see Fig.4.21) (autocorrelation within three hours is > 95%). Therefore, the reference DS should be selected from an optimal hour in the neighbouring three hours of the missing data. 'Optimal' here means that the hour should meet with three requirements.

1. This hour is close to the problematic hour and no more than 3 hours gap.
2. The wave spectrum derived from the DS of this hour is available.
3. The wind direction is stable.

This selected DS is denoted as $\hat{d}(n)$, but not used for the reference $d(n)$. This is because even though Doppler spectra in two continuous hours are similar, they will never be the same. So the real $d(n)$ should be $\hat{d}(n) \pm$ '*difference amount*'.

The 'difference amount' is termed as standard difference, and denoted by d_{std}. It is the estimated Euclidean distance between $\hat{d}(n)$ and the missing DS $\hat{s}(n)$. In practice, the distance of each pair of DS sections in two adjacent hours in Feb., ranging from $sfs-50$ to $sfs+49$, is calculated. This is done to avoid the impact of shifting Bragg peaks, due to the frequency selection procedure in Pisces, which changes with hour. In Fig.4.22, it is found that this Euclidean distance varies with hour and has a period of 24 hours, i.e. a day. It is also associated with the multiple frequencies used in one day at fixed hours, because the second order will be higher at higher radio frequency for the same waveheight. Hence, d_{std} is estimated by averaging the distances calculated using the previous 24 hours of the missing DS. This daily average instead of three hour average is taken also because a larger averaging range is assumed to avoid the case that the strong clutter in adjacent DS leads to the miscalculation of d_{std}. Taking d_{std} into consideration, the objective function (see Equ.(4.10)) is rewritten as:

$$\begin{aligned} J &= \min_{W_{opt}} \{e^2(n) = \left\| \left| \hat{d}(n) - \hat{s}(n) \right| - d_{std} \right|^2 \} \\ &such\ that: \hat{s}(n) = [w_1^* x_1, w_2^* x_2, ..., w_N^* x_N]^T \end{aligned} \tag{4.12}$$

Construction of the received Doppler spectrum $x(n)$

Compared with the construction of the reference DS, the received DS is less important, because it is going to be weighted by W_{opt}. In fact, the received DS $x(n)$ can be selected from any hour in that day. In practice, the received DS is selected with a goal to reduce the calculation time of the searching algorithms. This goal indicates that this DS should have a small Euclidean distance with the missing DS. Also, the selected DS should have the highest chance among other hours to have good quality. Fig. 4.24 shows the day-hour image of the Euclidean distance between the DS of rb 4 and rb 5 (choosing rb 4 and 5 because the buoy is located between rb 4 and 5). This has 24 hour bins in the X-axis and 28 day bins in the Y-axis. The colour of each cell shows the distance. The black cross indicates that the radar estimate is 30% less than the buoy estimate of waveheight in a cell (i.e. underestimation $\frac{hs\ of\ buoy - hs\ of\ radar}{hs\ of\ buoy} > 0.3$). On the contrary, the red dot is shown when the radar estimate is 30% larger (i.e.

overestimation $\frac{hs\ of\ buoy - hs\ of\ radar}{hs\ of\ buoy} < -0.3$). If there is no symbol, this means that the error of the estimate in this cell is in an acceptable range, i.e. $\pm 30\%$. The figure of 30% is used because DS observed in some hours outperform the others under this criterion. More specifically, the number of poor DS in the hours 00:10, 01:10, 02:10, 03:10, 04:10, 05:10, and 23:10 is between 5 and 7, less than the other hours. So the received DS is simulated using the average of the DS in hours 23:10-05:10. Note that 30% is used here as the threshold for poor data, while 20% is used as the threshold for good quality data elsewhere in this thesis.

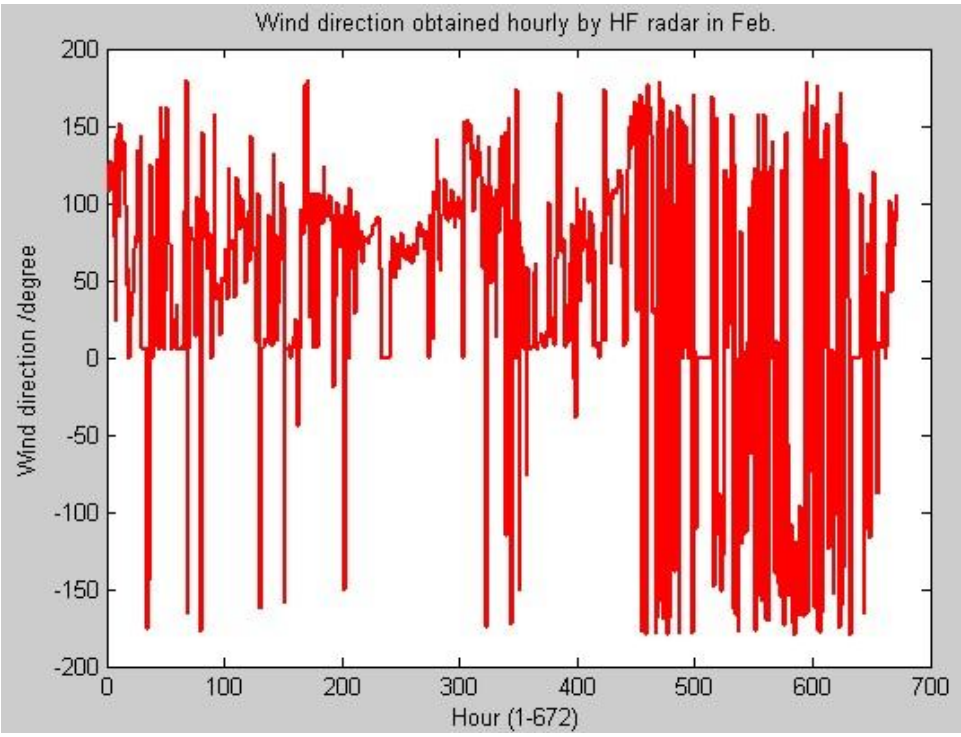


Fig.4.19 Wind direction estimated by radar in February. This time series of wind direction has 672 hour bins in the X-axis. The wind direction is in degrees.

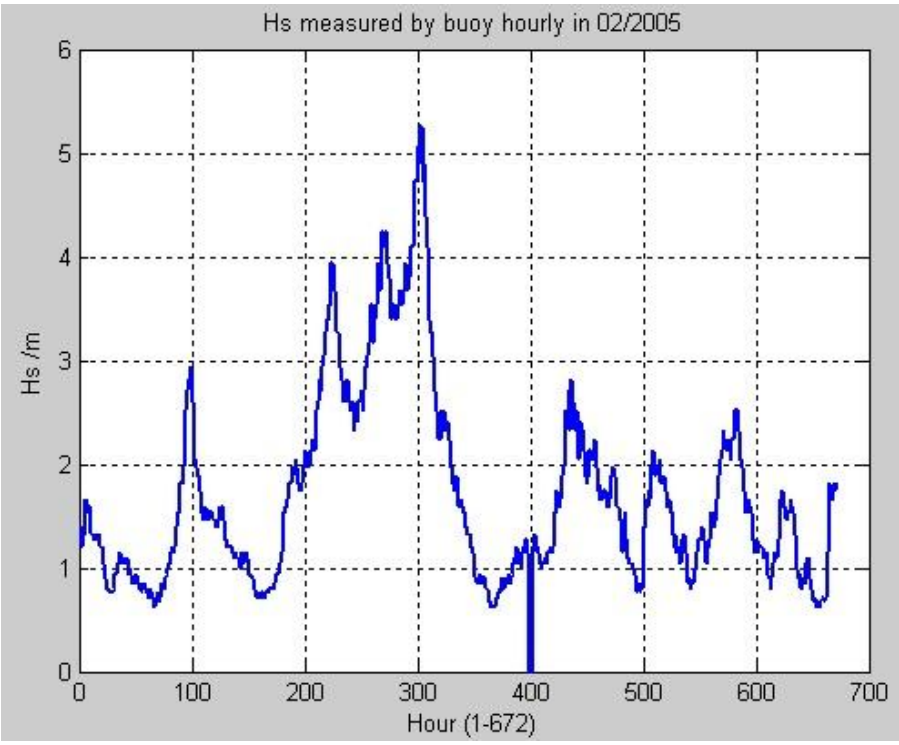


Fig.4.20. The buoy estimation of significant waveheight in Feb 2005, UK. This time series of waveheight estimates has 24×28=672 hour bins in the X-axis. The waveheight is in units of meters.

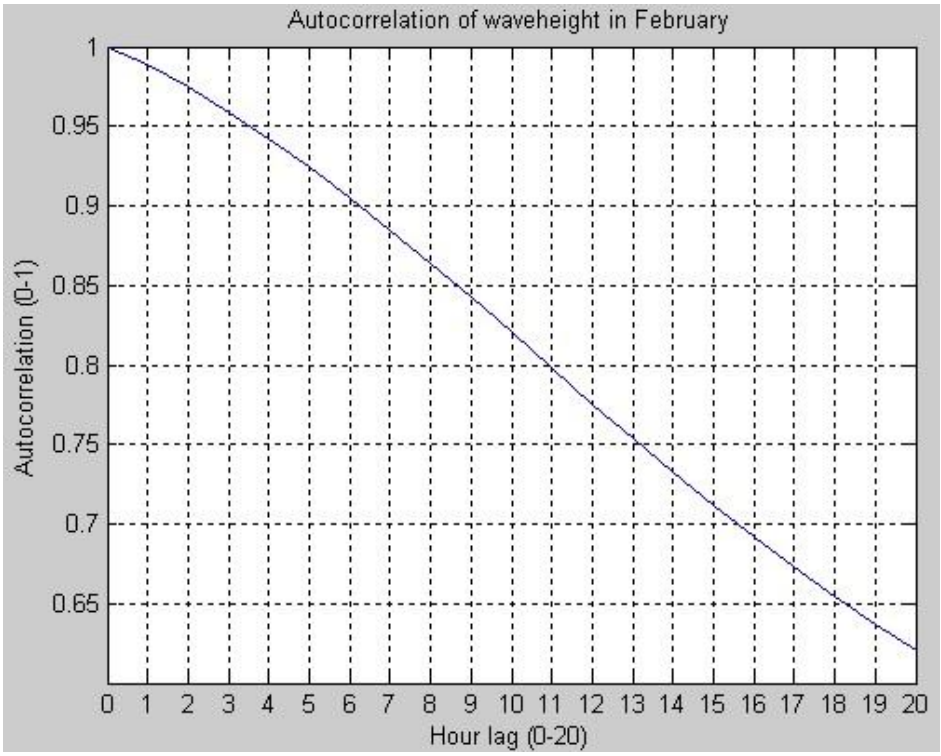


Fig.4.21 Autocorrelation function of waveheight based on the time series shown in Fig.4.20. The X-axis represents the hour lag; the Y-axis represent autocorrelation coefficient. As it shows, '3' is the maximum hour lag for two DS with the correlation in wave height larger than 95%. .

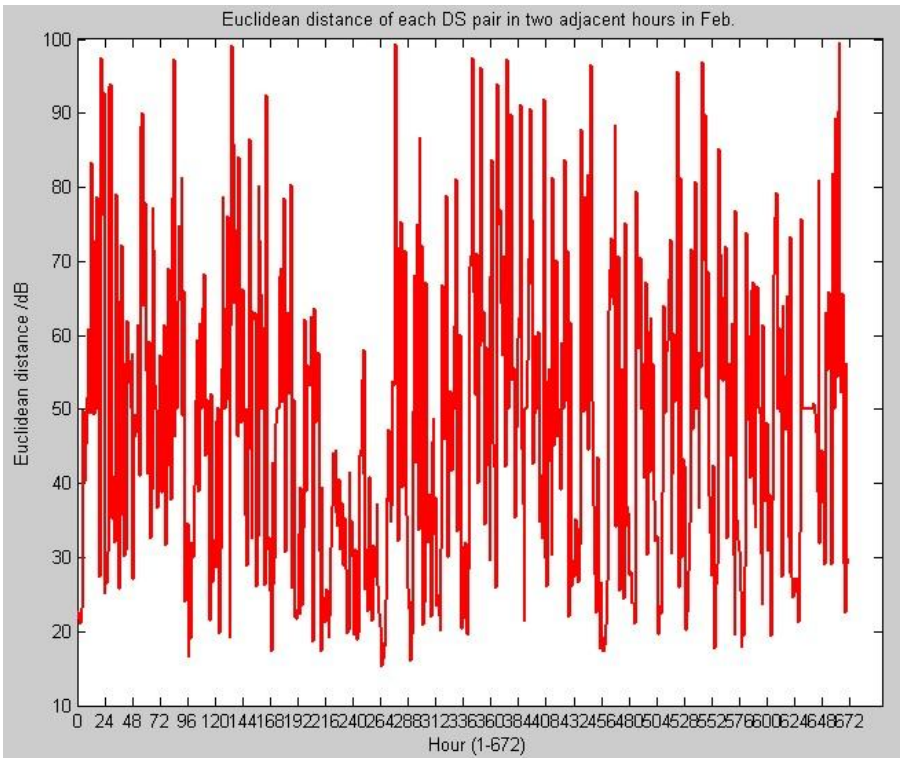


Fig.4.22 Euclidean distance of each DS pair in two adjacent hours from 00:10 01/02/2005 to 23:10 28/02/2005.

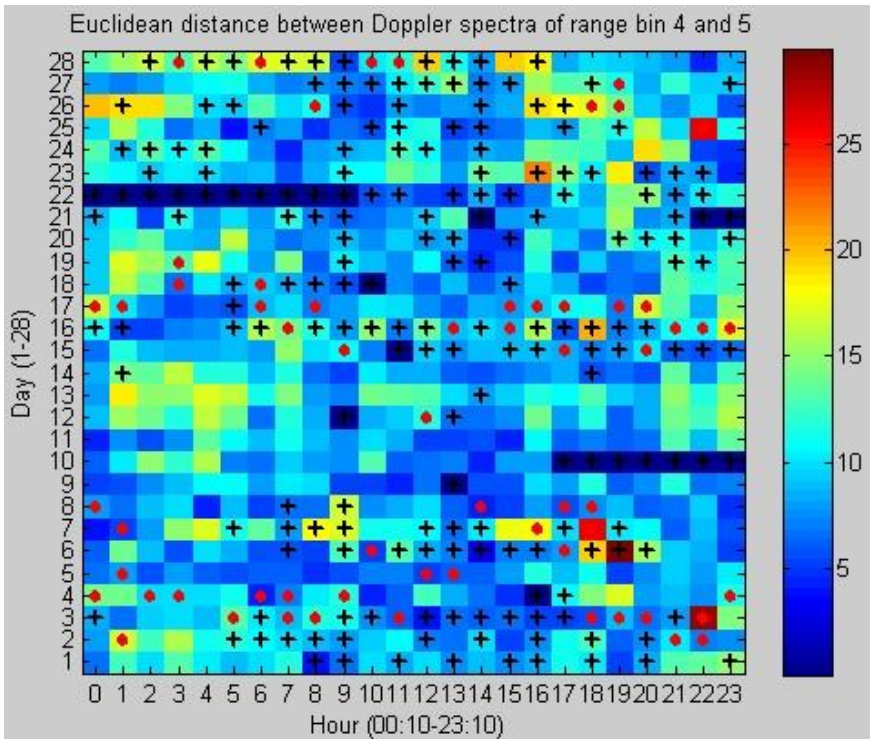


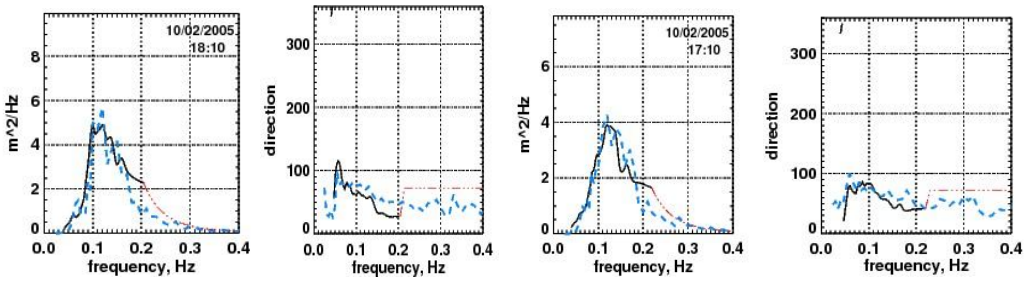
Fig.4.23. Day-hour image of Euclidean distance between DS of rb 4 and 5 for 'np' radar. Black cross: under-estimation; Red dot: over-estimation

Based on the methods explained above, the missing DS are supplemented as

follows. Note that what is described below is based on the 'np' dataset only. There are no missing data in the 'cm' dataset.

The first thing is to find in which day and hour the data are missing by checking the files. The missing DS are in continuous hours between 0 to 9 in day 22, between hours 17 and 23 in day 10, and 9-in-12 (the format is 'hour'-in-'day'), 13-in-9, etc. These missing DS correspond to the darkest blue cells in Fig.4.23, because the only reason for zero distance between DS of rb 4 and 5 is that they are missing (no two DS can be identical).

As an example, 18-in-10 is randomly selected from the missing set. For the reference signal construction, the available wave spectra from the radar in day 10 are found to be between hours 00:10 to 15:10. Among these hours, 15:10 is the closest one to 18:10. The assumption of stable wind direction needs to be checked. In day 10, the wind direction is found to be stable according to Fig.4.19, ranging from about 60° to 90° . So the DS of 15:10 is selected as $\hat{d}(n)$. Then the Doppler spectra observed between 00:10 to 05:10 and 23:10 (in a real-time procedure, DS from 23:10 would be chosen from the day before) are averaged and taken as the received DS, $\underline{x}$. After using the Lagarias searching algorithm (see section 4.3.1 and Equ. 4.12), the optimal weighting coefficient vector $W_{opt} = [w_1^*, ..., w_{1024}^*]^T$ is obtained, which makes the Euclidean distance between the estimated and the reference DS approximate d_{std}. The missing DS of 18-in-10 is estimated by $\underline{\hat{s}} = [w_1^* x_1, w_2^* x_2, ..., w_{1024}^* x_{1024}]^T$. Finally, the wave spectrum derived based on $\underline{\hat{s}}$ is shown in Fig.4.24(a). The buoy and radar estimates look similar in both the frequency and direction spectrum.

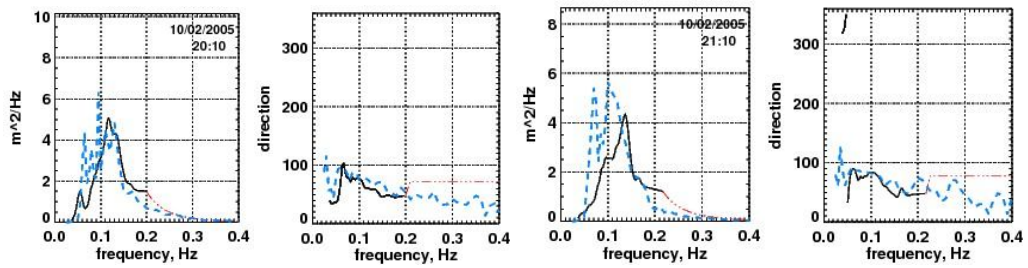


(a) (b)

Fig.4.24. (a) Left: Frequency spectrum; Right: direction spectrum from data of 18-in-10 (solid is the radar and dashed is the buoy); (b) data from 17-in-10.

The same procedures apply to generate the DS in 17-in-10. The optimal hour for this reference DS is selected from hour 14:10, since hour 15:10 was selected for 18-in-10 (the DS at hours 16:10-21:10 are all missing). The supplemented radar

wave estimate is shown in Fig.4.24(b). An agreement within ±20% is found between radar and buoy estimates of four oceanic parameters (see Table 4.3). The third and fourth examples are supplementing the data in 20-in-10 and 21-in-10. It may seem impossible to estimate the DS at 20:10 and 21:10, because the DS at hour 17, 18 and 19 are all missing. However, in this case, the DS of 20-in-10 and 21-in-10 are estimated using the newly supplemented Doppler spectra of hour 17 and 18 as the reference Doppler spectra respectively. The results are shown in Fig.4.25, where the two supplemented frequency spectra look less like the buoy estimate than was the case for 17-in-10 and 18-in-10, but the four parameter estimates are still in good agreement (see Table 4.3). The radar and buoy waveheights are 2.43m and 2.61m respectively for 20-in-10; 2.32 and 2.82 respectively for 21-in-10. Both radar waveheights are within 20% range of the buoy estimates. In addition, the directional spectra of both radar estimates still agree well with the buoy estimates. However, the buoy data at 20 and 21 both show the development of a swell component which this method is not capable of picking up.



(a)wave height: 20 (b) mean direction: 20 (c)wave height:21 (d) mean direction: 21

Fig.4.25. (a)(c) Frequency spectra (solid is the radar and dashed is the buoy); (b)(d) mean direction versus frequency (solid is the radar and dashed is the buoy). The radar estimates are derived based on estimated DS at 17 and 18.

As in last section, four parameters estimated by radar and buoy are compared to quantify the quality of the supplemented DS, i.e. significant waveheight hs, period T_2, peak period T_p, and peak direction D_p, and are shown in Table 4.3. Most of the parameters show that the supplemented missing data are in a good agreement with the buoy measurement, in particular h_s. The supplemented and reference DS are compared in terms of their radar-estimated and buoy-estimated parameters, e.g. 17-in-10 is compared with 14-in-10. It is found that this method at the moment cannot track the change of wave estimation by the buoy, and is limited to providing an approximated model of the missing data based on the closeness of the reference DS and the correct estimation of the standard difference d_{std}.

Table 4.3 Four parameters estimated by buoy and supplemented DS

Date	Source	h_s /m	T_1 /s	T_p /s	D_p / °
17-in-10	Buoy	2.61	5.69	8.39	247.5
	new DS	2.468	7.008	8.361	244.51
14-in-10	Buoy	3.04	6.24	8.04	270
	Ref DS	3.567	7.805	8.726	271.47
18-in-10	Buoy	2.71	5.95	8.58	253.12
	new DS	2.707	7.274	10.049	243.66
15-in-10	Buoy	2.93	5.95	8.80	258.75
	Ref DS	2.349	7.081	7.166	237.30
20-in-10	Buoy	2.61	6.54	10.05	264.38
	new DS	2.434	7.701	8.539	241.40
21-in-10	Buoy	2.82	7.23	10	270.00
	new DS	2.316	7.164	7.296	235.02

4.4 Summary

In this chapter three popular signal processing schemes have been reviewed and a suitable application of each method was considered. For example, the wavelet transform can detect the position of spike-like signals in a time series, so it was used to detect the accurate positions of Bragg peaks in a DS. AR modelling can predict signals based on the past records at a certain time, so it was used to predict DS. The adaptive filtering system can restore signals by finding an optimal weighting operator, which is calculated recursively according to some objective function. So it was used to supplement missing DS in the case that the radar operation is temporarily stopped for some reason or the short-term and strong clutters degrade DS in some hours. Results of each application were provided in each subsection.

So far, the procedures for supplementing missing data haven't been used for simulating a DS which is not missing but poor in quality in order to invert more accurate wave spectra. In that case, the way to construct the reference DS will have more possibilities, e.g. using the DS in adjacent range bins. This work will be left for future investigation.

There are limitations of each method tested in this chapter when it comes to the

improvement of wave measurements. The wavelet plus ICA method is only able to identify the positions of Bragg peaks, which increases the accuracy of the initial mean direction estimation that is substituted into the Pierson-Moskowitz spectrum (used for the initial condition of the inverse algorithm for the ocean wave directional spectrum). The non-white wavelet denoising method could make the Bragg peaks in the DS from far ranges more 'spiky', when the surrounding noisy signals are removed to some extent. Note that none of the white or non-white denoising methods are able to change the overall noise floor in the DS. The wavelet methods discussed above aim at improving the first order estimate, and have little influence on the second order region. Hence they are of limited use for improving the wave estimation. Using the AR model to regenerate a DS is only a preliminary attempt. It is still under development. Some unresolved issues are: how much it smooths a DS; what is the impact of overestimation on higher frequencies. Anyway, for a fundamental improvement of wave estimates, smoothing is not enough. These issues are interesting but are left for future investigation. Finally, an adaptive system was developed for the specific case that the DS is missing. This method was not intended for general processing, and required that the wind doesn't change much in the surrounding 3 hours of the missing DS.

Therefore, the applications above of the three signal processing schemes all have limited performance in separating the sea echoes from clutters in the second order region, which is needed for a significant improvement of the wave estimation. Another scheme called 'clutter mitigation scheme', which involves image recognition, segmentation and subspace projection, is developed in the next chapter and has shown improvements in wave estimation in a more robust and comprehensive way.

Chapter 5 Clutter mitigation schemes for Doppler spectra

As explained in Chapter 2, the radar power spectrum backscattered from the surface of the sea has been utilized to measure wave parameters. When strong and unknown clutters are mixed into the first- and second- order areas of the DS, parts which are essential for the wave estimation, the quality of the data will be significantly reduced. From a graphical point of view, all the clutters can be classified into two types: visible and invisible. In this chapter, image processing terminology will be used to describe features in the RD image.

Visible clutter refers to those non-sea amplitudes in the RD image that are clearly identifiable by eye. The most obvious examples of visible interferences are radio frequency interference or ionosphere interference, due to their characteristic regular distributions in the RD image. Some other examples are visible but irregular clutters, e.g. a moving ship or plane, meteor trails. In contrast, invisible clutter refers to those non-sea amplitudes masked in the first- and second- order region in the RD image. A visual presentation of these terms are shown in Fig. 5.1

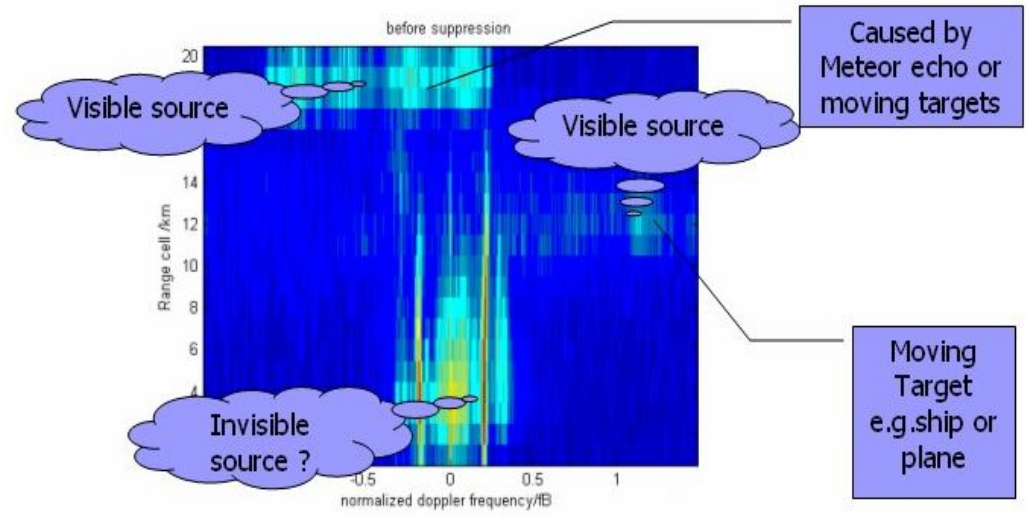


Fig.5.1 Range-Doppler image at 00:00 on Feb 1, 2005, Castlemartin, UK

In this chapter schemes are developed for visible and invisible clutter mitigation respectively. The visible scheme is examined firstly for the cancellation of the two main damaging sources of clutter, i.e. RFI and II, in order to obtain better wave estimation. Next, this RFI-mitigation scheme is generalized to mitigate visible clutters of arbitrary shapes. The generalized scheme is divided into two categories, i.e. unsupervised or supervised, depending on whether training data

are used or not. The unsupervised approach is developed for a case where no training (clutter-only) data are available. The unsupervised approach investigates four types of compositions of the compound signal, i.e. sea-only, clutter-only, sea with clutter (sea is stronger than clutter), sea with clutter (sea is weaker than clutter). The exceptional case that an unsupervised scheme cannot handle is sea with clutter but the signal amplitude of sea is almost equal to clutter. For this case, the supervised scheme is developed with an assumption that the clutters are distributed not only inside but also outside the sea echoes. This assumption indicates that training data containing clutters only outside the mixing area are available. However, if the assumption doesn't hold, i.e. the clutters are totally masked by the sea echoes, the supervised approach will fail. The invisible scheme is then developed for this case. Improvements in the radar estimation of the wave frequency spectrum are shown in the subsections 5.1.3.4, 5.2.5 and 5.3.5. Note that the extended visible clutter algorithm developed for canceling any shape of clutters is presented in this chapter mainly for demonstration. Visible clutter well away from the first and second order regions is of course not a problem for oceanographic parameter estimation, so no radar parameters are compared with buoy parameters to show the improvements in wave estimation. This generalized scheme can be used for the identification of the first order peaks in high noise conditions, which could have an impact on wave estimation because if Seaview is picking up the visible clutter peak instead of the first order peaks, some good quality second order information may be missing and hence wave measurements.

In this chapter all the clutter mitigation schemes except the generalized scheme are developed with the aim of improving the quality of Doppler spectra and thus the validity of wave estimation. All schemes follow three-step procedures: image recognition, segmentation processing and subspace projection. Section 5.1 firstly introduces a successful application of the visible clutter mitigation scheme to RFI and II, which are the dominant contaminating sources. Most of the material presented in this section is published in (Wang and Wyatt 2011). Section 5.2 extends the scheme designed specifically for RFI and II to any visible clutters, either regular or irregular shape. The last section 5.3 explores invisible clutter mitigation schemes to deal with invisible clutters in the first- and second- order areas.

5.1 Visible clutter mitigation on radio frequency interference

5.1.1 Introduction

For sea state monitoring, Wan (2006) noted that the three main sources of unwanted signals are ionospheric clutter, impulsive noise and radio frequency interference. RFI, together with a rough mitigation of II and meteor trails are taken into consideration here.

The cancellation methods for RFI reported to date fall into four categories. First, an intuitive approach is to shift the HFSWR working frequency to a frequency range that is not contaminated with RFI (Wyatt et al. 2006). However, this practice meets a problem in that it is sometimes difficult to find a free band with sufficient bandwidth (e.g. 30KHz) for operation in the 3-30MHz range that is crowded with broadcasting users. Second, Gurgel et al (2007) proposed an algorithm that removes all the signals except for the RFI, similar to switching off the transmitter, for RFI subtraction from the original radar signal. As Gurgel admitted, this algorithm doesn't remove RFI in all cases. It fails when there is strong RFI within the whole radar bandwidth. Third, adaptive beamforming algorithms have been developed using sub-arrays to get rid of RFI, because typically RFI has an obvious directivity property (Wan et al. 2005). For RFI cancellation, adaptive beamforming is basically a conventional signal processing method used in skywave radar systems. Fabrizio et al and Chan & Huang developed time-domain and Doppler domain cancellation techniques respectively in (Chan and Huang 1999; Fabrizio et al. 2004). Generally, three domains were exploited and compared to evaluate the performance of beamforming: time-frequency domain, range-Doppler domain and spatial domain (Xin et al. 2008). The beamforming methods were implemented using radar data, collected at each antenna receiver, that are omnidirectional in the azimuth domain. Data from such a system is available with the WERA (Gurgel et al. 1999). However the Pisces radar system provides data after analog beam forming, making it impossible to apply this approach (Wyatt et al. 2006). Even if the beamforming method is successfully carried out, no robust method has been developed to deal with the remaining RFI after beamforming. Fourth, Zhou (2005) proposed a cancellation solution to separate the dense RFI subspace from the noise subspace by orthogonal projection.

In this chapter, the fourth method is explored further and generalized for the mitigation of various complex cases of RFI, e.g. mixed with meteor trails clutter and ionosphere interference. The aim is to suppress RFI occupying the first- and second- order parts of the DS that are essential for oceanographic measurements. A mathematical model for RFI is demonstrated to further propose a robust solution including image recognition, segmentation, and subspace projection. Note that this method extends the subspace projection method proposed by Zhou (2005) to this three step scheme to allow an improvement of the wave estimation.

5.1.2 Mathematical Analysis of RFI

The characteristics of RFI in radar signals are determined by the waveform used by the radar system. In Pisces, the FMICW is adopted to measure both range and Doppler information (Mahafza 2000; Wyatt et al. 2006) and the mathematical model of RFI in the FMICW signal is analyzed below.

FMCW radar transmits radio waves, with the carrier frequency varying from f_1

to f_2. The bandwidth $B = \Delta f = f_2 - f_1$ is usually several kHz. In a sweep period *T*, the transmitted signal can be given:

$$S_T(t) = \cos(2\pi f_1 t + \pi\alpha t^2 + \varphi_1) \qquad 0<t<T \qquad (5.1)$$

where $\alpha = B/T$ is the linear frequency modulation rate; φ_1 is the initial phase. Assume there is a target moving with radial velocity of V away from the radar, then the target echo in the $(n+1)$ th sweep period can be written:

$$S_R(t) = S_T(t - t_d) \qquad 0<t<T \qquad (5.2)$$

Where $t_d = \frac{2R_0}{c} + \frac{2V}{c}nT + \frac{2V}{c}t,\ 0 < t < T$; R_0 is the initial distance of the target away from the radar; c is the speed of light. As Pisces is a FMICW radar system, both the transmitted and received signals are gated by a signal $g(t)$. The target echo expression is $S_{gR}(t) = (1 - g(t))S_T(t - t_d)g(t - t_d)$. Looking at this term "$[1 - g(t)]g(t - t_d)$", the impact of this gating signal is only to multiply the target echoes from different ranges with a different coefficient, indicating that the target echoes are recorded with different time periods. This impact won't change the frequency components of the target echoes. So the FMCW case is considered to simplify the analysis. When the received signal $S_R(t)$ is mixed with the local oscillator signal $S_T(t)$ and filtered by a low pass filter (LPF), the baseband signal $S_I(t)$ is given as:

$$S_I(t) = \cos(\omega_1(t - t_d) - \omega_1 t + \pi\alpha(t - t_d)^2 - \pi\alpha t^2) \qquad 0<t<T \qquad (5.3)$$

where $\omega_1 = 2\pi f_1$ is the angular frequency. The phase of the baseband signal is ϕ_{n+1}:

$$\phi_{n+1}(t) = -\omega_1 t_d - 2\pi\alpha t t_d + \pi\alpha t_d^2 \qquad 0<t<T \qquad (5.4)$$

For easy demonstration, let $a = \frac{2R_0}{c}, b = \frac{2V}{c}, d = \pi\alpha$, and substitute the value of $t_d = a + bnT + bt$ into Equ.(5.4), then Equ. (5.4) is rearranged into:

$$\begin{aligned}\phi_{n+1}(t) &= -a\omega_1 - bnT\omega_1 - bt\omega_1 - 2d(a+bnT+bt)t + d(a+bnT+bt)^2 \\ &= -bnT\omega_1 + d(bnT)^2 - 2d(t-a-bt)bnT + \phi_1 \qquad for \quad 0<t<T\end{aligned} \tag{5.5}$$

where $\phi_1 = -a\omega_1 + a^2 d - (\omega_1 b + 2ad - 2abd)t + (b^2 d - 2bd)t^2$ is the phase of the backscattered signal in the first sweep. Since $b = \frac{2V}{c}$ is normally very small, the terms in Equ.(5.5) containing b^2 and b without multiplied by ω_1 or nT are ignored. By this approximation, the phase ϕ_{n+1} becomes:

$$\phi_{n+1}(t) = -a\omega_1 + a^2 d - \omega_1 bnT + 2abdnT - 2bdnTt - (\omega_1 b + 2ad)t, \ 0<t<T \tag{5.6}$$

Substitute the values of a, b, d back, the target frequency in the $n+1$ sweep is:

$$f_{n+1} = \frac{1}{2\pi}\frac{d\phi_{n+1}}{dt} = -\frac{2V}{c}f_1 - \frac{B}{T}t_0 - \frac{2V}{c}Bn \tag{5.7}$$

where t_0 is the time to receive the target echo from the initial distance R_0, $t_0 = \frac{2R_0}{c}$. This frequency is made of three parts: the first term is due to the target velocity, which is the Doppler shift in Doppler processing; the second term is generated by the initial distance of the target away from the radar; the third term indicates that the target is moving from one sweep to the next sweep. Likewise, these analyses are applied to the RFI in the received signal. In the receiver, RFI is considered as a narrow band signal that can be modeled as a superposition of a series of sine functions with coherent frequencies:

$$S_J(t) = \sum_j A_j \cos(\omega_j t + \varphi_j), \ 0<t<T \tag{5.8}$$

where $\omega_j = 2\pi f_j$ is the jth single interfering frequency, φ_j is its initial phase, A_j is the time varying amplitude, which is approximated as a constant because its variation in frequency is small compared with the interfering frequency. For simplification, the RFI will be explained with respect to a single frequency.

In the (n+1) th sweep, this RFI can be written:

$$J_{n+1}(t) = A_j \cos[\omega_j(nT+t) + \varphi_j] \qquad 0<t<T \tag{5.9}$$

After mixing with the local oscillator signal, $S_T(t)$, and applying a low pass filter (LPF), the baseband signal is:

$$J_{n+1} = A_j \cos[2\pi f_d t - \pi\alpha t^2 + \varphi_d] \qquad 0<t<T \qquad (5.10)$$

where $f_d = f_j - f_1$ is the starting frequency and $\varphi_d = \varphi_j - \varphi_1 + 2\pi f_j nT$ is the initial phase. The frequency modulation rate is equal to that of the oscillator signal but of opposite sign.

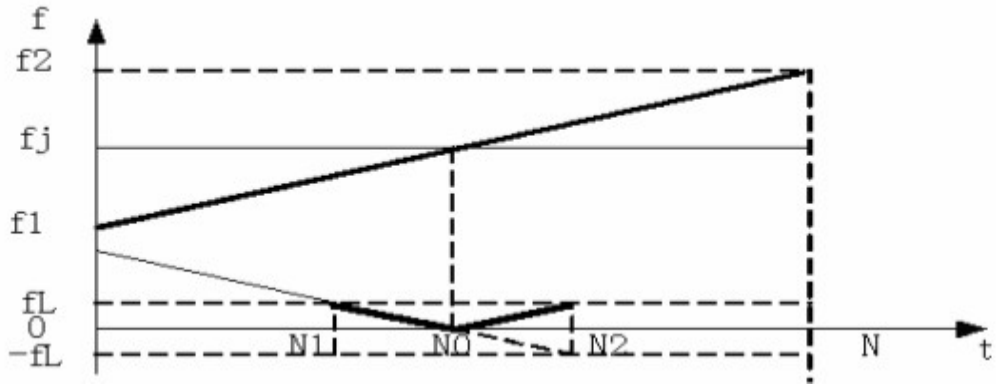


Fig. 5.2 Time-frequency property of a single frequency RFI mixed with the oscillator signal. f_L is the cutoff frequency of LPF and f_j is the RFI frequency. The FMCW is a chirp signal sweeping from f_1 to f_2. N0, N1 and N2 are samples in time. N is the number of samples in a sweep/chirp period. Suppose the sampling rate is ΔT, then $N * \Delta T = T$.

Fig. 5.2 shows that when the single frequency RFI enters the receiver, it becomes a linear frequency modulated signal. Its spectrum spreads over the whole receiver bandwidth. Due to the LPF, RFI just appears from the N1 to N2 samples in time for all the sweeps, as long as the frequency of RFI is stable. Notice that $N_2 - N_1 = \frac{2f_L}{\alpha} = \frac{2f_L N}{B}$, so for a certain radar system where N (the number of samples in a sweep period), B and f_L are known, the number of time samples that are occupied by RFI can be calculated.

Before further discussion of the behavior of RFI, the conventional radar signal processing for target detection is introduced, which usually consists of sequential Fast Fourier Transforms (FFT) called range processing and Doppler processing respectively.

As shown in Equ. (5.7), the frequency of the target echo is linearly related to the time delay or the distance of the target. In each chirp, for example the “n+1” chirp, the received signal is a superposition of target echoes backscattered from different ranges, or in other words, with different time delay. After the first FFT, the target echoes from different ranges are separated from each other in the frequency spectrum. This first FFT is thus called range processing and provides

the frequency spectrum (or range spectrum) of target echoes. The RFI in the $n+1$ sweep after range processing takes the form (Zhou et al. 2005):

$$J_{n+1}(f)=\frac{A_j}{2}\left\{e^{j\varphi_d}e^{j\frac{\pi}{\alpha}(f-f_d)^2}\int_{(f-f_d)/\alpha}^{T+(f-f_d)/\alpha}e^{-j\pi\alpha t^2}dt\right.$$
$$\left.+e^{-j\varphi_d}e^{-j\frac{\pi}{\alpha}(f+f_d)^2}\int_{-(f+f_d)/\alpha}^{T-(f+f_d)/\alpha}e^{j\pi\alpha t^2}dt\right. \tag{5.11}$$

where this integration is Fresnel integration and $|f|\le f_L$. When f_L (the LPF's cutoff frequency) is much smaller than f_j (the RFI frequency), the two integral terms above are approximately equal to $(1-j)/\sqrt{2\alpha}$ and $(1+j)/\sqrt{2\alpha}$ respectively. For any sweep the f_d and φ_d are constant, so RFI can be seen as the superposition of the same frequency components in all range bins.

Suppose I have N_{CIT} chirps in a CIT, then I have N_{CIT} samples as the "time series" for each target coming from a range. Hence, the phase of this "time series" is linearly related to the velocity of the target. The second FFT is applied to all the N_{CIT} samples for each range target (the number of the range targets is determined by the frequency number in the first FFT) to generate the DS. So the target echoes with different velocities from a certain range are separated from each other in the DS. Considering the RFI in N_{CIT} sweeps, assume the frequency components in the m th frequency bin are:

$$J_{n+1}(f_m)=\frac{A_j}{2}[e^{j\varphi_d}g_1(f_m)+e^{-j\varphi_d}g_2(f_m)] \tag{5.12}$$

where

$$g_1(f_m)=e^{j\frac{\pi}{\alpha}(f_m-f_d)^2}\int_{(f_m-f_d)/\alpha}^{T+(f_m-f_d)/\alpha}e^{-j\pi\alpha t^2}dt \tag{5.13}$$

$$g_2(f_m)=e^{-j\frac{\pi}{\alpha}(f_m+f_d)^2}\int_{-(f_m+f_d)/\alpha}^{T-(f_m+f_d)/\alpha}e^{j\pi\alpha t^2}dt \tag{5.14}$$

Since nT is an independent variable, as long as f_d is stable, $g_1(f_m)$ and $g_2(f_m)$ are also constant. But $\varphi_d=\varphi_j-\varphi_1+2\pi f_j nT$ is a function of nT, so the

FFT of $J_{n+1}(f_m)$ is also a superposition of single frequencies at $\pm f_j$. Since m is randomly chosen, the RFI is located in all range bins. The term RFI used in the rest of this thesis refers to the manifestation of the RFI in the processed radar data.

The RD image is used to represent the Doppler spectrum in the range and Doppler dimensions (Fig.5.1). The color scale is in dB in the image showing the amplitude of the Doppler spectrum. Based on the study of the mathematical model of RFI, RFI has a strong correlation in the range domain and a weak correlation in the Doppler domain. Therefore, a strong enough RFI would be very easy to detect, identify and subtract in the RD image.

5.1.3 Characteristics of RFI and signal processing schemes

The RFI can be investigated in the time, frequency or space domains. RFI has a strong temporal correlation that has been used for suppression in range processing (Luo et al. 2001). The solution in that paper requires a higher data rate than the HF radar system adopts. Another possible approach is to detect the range from the N1 to N2 samples (Fig. 5.2) in the time series that are contaminated by RFI, remove RFI and interpolate the gap in the time series of each sweep. However this approval has not been tested in this work. The limitations of such an approach are two: lack of efficient and well-proved interpolation algorithms; and it is only applicable for sparse RFI (the RFI is called sparse when it has low energy and short Doppler coverage; a dense case is if RFI has high energy and long Doppler coverage). For processing in the space domain, conventional adaptive beamforming fails to provide sea state measurements in the same direction as the RFI while new beamforming methods increase the computational load by adding stochastic constraints (Abramovich et al. 2000; Fabrizio et al. 2006). Therefore, a simple but efficient solution is investigated and recommended for practical use. This solution mitigates RFI by three consecutive sub-processings: image recognition, segmentation processing (not always a necessity) and subspace projection, motivated by the image and statistical characteristics of RFI in the spectrum based on the analysis in section 5.2.

There are two things to be noticed. One is that RFI patterns drawn from one area may not apply to another. For example, in data collected from the East Sea in China (Zhou et al. 2005), RFI becomes dense at night and sparse at daytime because of the disappearance of the D layer – the absorbing layer in the ionosphere. This pattern has not been identified in the Pisces dataset collected in UK, possibly because Pisces uses different frequencies at different times of day to avoid RFI as far as possible or the RFI itself is different here. At this different geographical location, stronger RFI at 10:00 and 13:00 in February 2005 due to short-wave communication or broadcast is found, an example of which can be seen in Fig. 5.3. This figure shows a Range-Doppler image with dense RFI. The

X-axis represents Doppler frequency in the range of -2.5~2.5 Hz, in units of Doppler bin. The color scale shows the power spectrum in dB. The Y-axis represents Range in the range of 0~300km, in units of Range bin. The dense interference covers the whole range axis and the Doppler domain between Doppler bin 1~150. The second one is that the distribution of RFI in the RD spectrum varies, e.g. dense or sparse, symmetric or asymmetric (the symmetric RFI means that it is approximately symmetric around zero Doppler frequency) as a result of different sources. The work described here is the first time a global solution to deal with all types of RFI has been developed

5.1.3.1 Image recognition

After Doppler processing, data collected by the Pisces radar are displayed in a RD image and sorted out by their image features. Fig 5.4 shows a comprehensive set of four types of RFI in pairs for comparison: RFI to the negative frequency side or positive frequency side of the sea scatter region; sparse or dense; singular or symmetric; out-band and in-band ("band" refers to the first- and second- order Doppler area of the sea echo).

In addition to RFI, meteor trails, ionospheric clutter, sea echo and target echoes all display their characteristic graphical features in the RD image. For instance, the area pointed by the black arrow in the center of Doppler domain and 18-20 bins of Range domain in Fig. 5.3 is due to meteor trails. So I classify different combinations of these clutters in the image into three levels of complexity: low, middle, and high (the definition of these classes is given along with the examples explained in the section 5.1.3.4). Based on the classification, RFI will be cancelled by segmentation and signal subspace projection discussed below. So far, image recognition has been realized simply by human intervention. Making it automatic is very important for any operational use but is beyond the scope of this work.

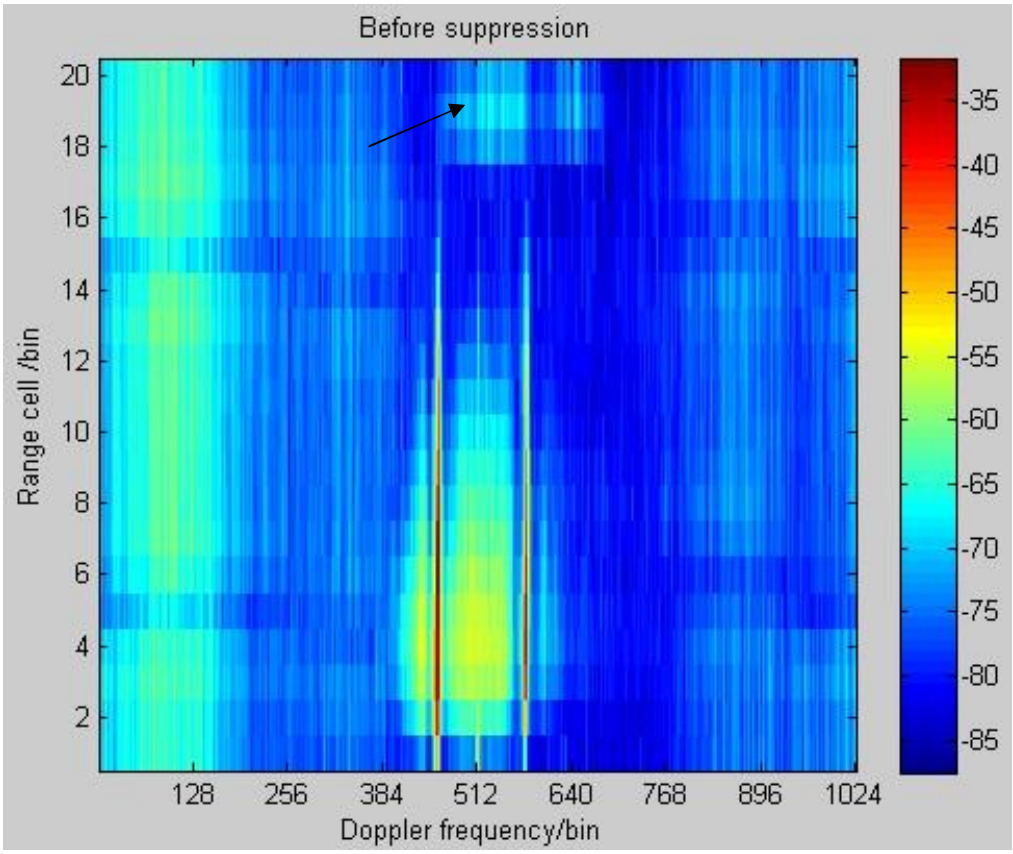


Fig. 5.3 RD image showing dense RFI. The horizontal axis represents the Doppler frequency and the vertical axis represents the range with a resolution of 15 km. The two dominant vertical lines between 400 and 600 Doppler bin indicate the two Bragg peaks. The area around the two peaks between 400 and 600 Doppler bin are 2nd order region. The area on the very end of the Doppler axis is the dense RFI. The color scale shows the power spectrum in dB. Each Doppler bin represents 0.0048Hz. The frequency range is from -2.5Hz to 2.5Hz. Each range bin represents 15 km. Data are taken from 'cm' site, at 13:00, 12/02/2005.

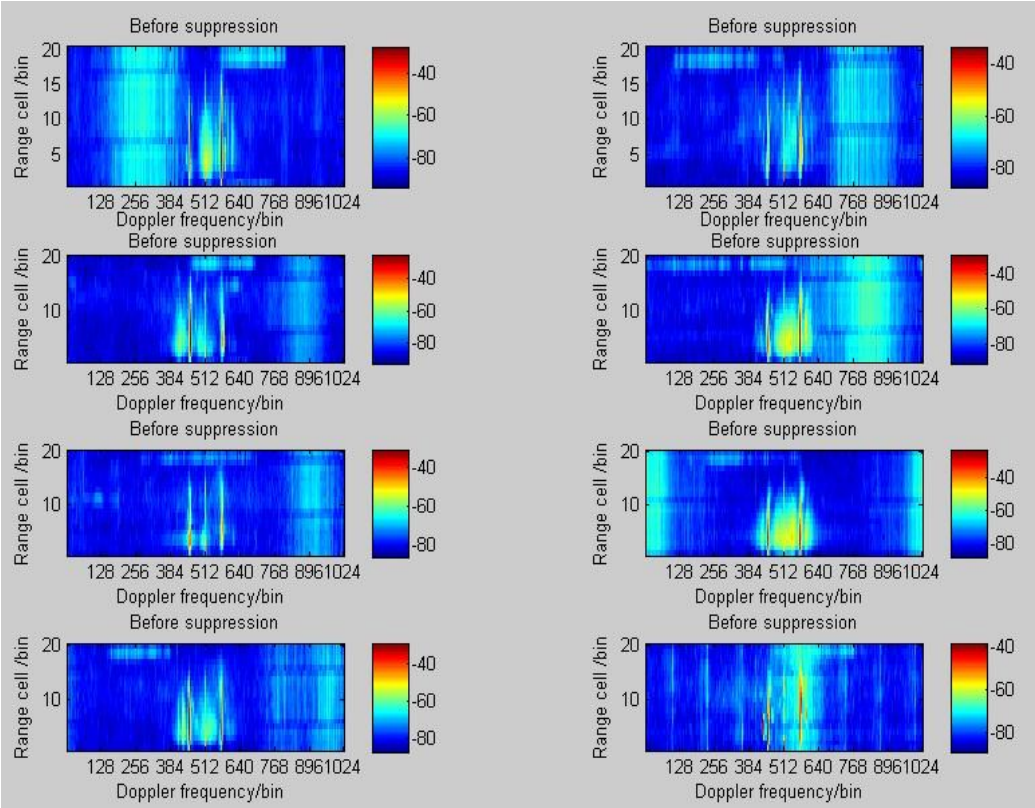


Fig 5.4. RD image comparison: top row: negative and positive; second row: sparse and dense; third row: singular and symmetric; bottom row: out-band and in-band. The color scale shows the power spectrum in dB. Each Doppler bin represents 0.0048Hz. The frequency range is from -2.5Hz to 2.5Hz. Each range bin represents 15 km. Data taken from Castlemartin in the United Kingdom at 10 am, 02, 06, 08, 14, 07, 13, 05, 08 in Feb. 2005

5.1.3.2 Subspace projection algorithm

In this subsection, it is demonstrated that there exists a subspace projection filter which maintains the sea echo and mitigates the RFI, regardless of the slight Doppler fluctuations of the filter. As discussed in the section 5.1.2, RFI has a strong range correlation and weak Doppler correlation, which features can be used as the basis of the subspace projection explained below. The segmentation method in subsection 5.1.3.3 is related to the subspace projection method, and is explained later.

This algorithm is similar to Zhou's method (2005) in the sense that they both use eigenvalue decomposition as the way to separate the signal and noise subspaces. The three differences are listed below.

1. This algorithm is developed for the purpose of wave estimation, so the signal model of the DS is analyzed in this scenario while not in Zhou's analysis.

2. Zhou uses the DS at far range to decompose eigenvectors while this algorithm

introduces a concept called the training matrix which does not necessarily have to be in far range; anywhere in the RD image can be chosen as the training matrix (this is explained in detail in section 5.2.1).

3. The correlation matrix of the RFI is estimated only in the range domain in Zhou's work while it can be estimated in both range and Doppler domain in this algorithm.

This algorithm proceeds as follows. Firstly select an area in the RD image with RFI alone to be the "training matrix", X_A. Secondly, select another area in the RD image with both RFI and sea echo to be the "processing matrix", X_B. Thirdly, construct the covariance matrix $\hat{R}_A$, $\hat{R}_B$ of X_A and X_B. Fourthly, apply eigenvalue decomposition analysis (EDA) on $\hat{R}_A$, to produce an eigenvalue matrix Λ and an eigenvector matrix V. Fifthly, project the processing matrix X_B into the eigenvectors V to get the RFI components in X_B and then subtract them from X_B to extract the sea echo.

A. Signal model

Let X be the complete RD matrix, i.e., $X \in R^{20\times 1024}$ and $X_A, X_B \subset X$. Let the scalar power value x_{rd} be the element corresponding to the *r*th range bin and the *d*th Doppler bin in X. In general, the Doppler power x_{rd} can be defined as the mixture

$$\begin{aligned} &x_{rd} = (s_{rd}) + (j_{rd}) + (c_{rd}) + n_{rd}, \\ &for\ r = 1,2,\ldots,20,\ d = 1,2,\ldots,1024 \end{aligned} \tag{5.15}$$

where s_{rd} is the sea echo signal; j_{rd} is the RFI signal; c_{rd} is the clutter; n_{rd} is additive white noise of power σ^2 in the Doppler domain. The terms in brackets are not always present.

B. Construction of training covariance matrix A and processing covariance matrix B

Training data are extracted from a sea-echo-free (sea echo is attenuated to some extent that its amplitude is similar to the noise level) area covering N_r range

bins and M_d Doppler bins in the RD image. Let $\underline{x}(d) = \underline{j}(d) + \underline{n}(d)$, a column vector of dimension N_r, be the RFI-plus-noise range profile at Doppler bin d. The variance of $\underline{x}(d)$ changes with Doppler frequency, d. In the Doppler frequencies occupied by RFI, the variance is small, due to the strong range correlation of the RFI, $\underline{j}(d)$. In practice, the training covariance matrix, i.e. the RFI-plus-noise covariance matrix, R_A, is approximated by the samples as

$$\hat{R}_A = \frac{1}{N_r} X_A{}^T X_A \tag{5.16}$$

where $X_A \in R^{N_r \times M_d}$ is the sample matrix selected from training area A. Each column (Doppler) of X_A is considered as a variable, i.e. $\underline{x}(d_m) = [X(r_1, d_m), X(r_2, d_m), X(r_3, d_m), \ldots, X(r_{N_r}, d_m)]^T$, $m = 1, 2, ..., M_d$. Each row (range) of X_A is considered as an observation, i.e. $\underline{x}(r_n) = [X(r_n, d_1)\ X(r_n, d_2)\ X(r_n, d_3), \ldots, X(r_n, d_{M_d})]$, $n = 1, 2, ..., N_r$. $\hat{R}_A$ is a symmetric matrix of dimension M_d, and $\hat{R}_A = \hat{R}_j + \hat{R}_n$.

Similarly, the sample estimate of the RFI-plus-sea-plus-noise covariance matrix is given by

$$\hat{R}_B = \frac{1}{N_r} X_B{}^T X_B \tag{5.17}$$

where $\underline{x}(d_m) = \underline{s}(d_m) + \underline{j}(d_m) + \underline{n}(d_m)$, $\underline{x}(r_n) = \underline{s}(r_n) + \underline{j}(r_n) + \underline{n}(r_n)$ and $\hat{R}_B = \hat{R}_s + \hat{R}_j + \hat{R}_n$.

C. Eigenvalue Decomposition

Eigenvalue decomposition is often used to divide the covariance matrix into two orthogonal subsignal spaces, namely “signal” and “noise” subspaces. Note that the term “signal” is not referring to a particular signal, but indicates the dominant subsignal space (dominant signal space is spanned by the eigenvectors corresponding to larger eigenvalues). For different purposes, either “signal” or “noise” can be extracted or subtracted.

I write $\hat{R}_A$, the sample covariance matrix of the radar return containing the RFI, and noise as follows:

$$\hat{R}_A = \hat{R}_j + \hat{R}_n = V \Lambda V^T = \sum_{i=1}^{M_d} \lambda_i \underline{v}_i \underline{v}_i^T \qquad (5.18)$$

where $\hat{R}_j$ and $\hat{R}_n$ are the M_d-order covariance matrix of "signal" and "noise" respectively; $V = [\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_{M_d}]$, $\underline{v}_i, i \in 1,2,\ldots,M_d$ is one of the M_d mutually orthogonal eigenvectors; $\lambda_i,\ i = 1,2,\cdots,M_d$ is eigenvalue corresponding to $\underline{v}_i$ and $\lambda_1 > \lambda_2 > \ldots > \lambda_{M_d}$. The first p eigenvalues represent the "signal space", and the remaining smaller eigenvalues are from the "noise space". Therefore, X_j (the RFI signal) and X_n (the noise signal) in X_A, can be obtained by:

$$X_j = V_p V_p^T X_A, \quad V_p = [\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_p] \qquad (5.19)$$

$$X_n = V_{N_r-p} V_{N_r-p}^T X_A, \quad V_{N_r-p} = [\underline{v}_{p+1}, \underline{v}_{p+2}, \ldots, \underline{v}_{M_d}] \qquad (5.20)$$

$$V_p \perp V_{N_r-p} \qquad (5.21)$$

Likewise, the sea echo can be extracted from the processing matrix X_B by

$$\hat{R}_B = \hat{R}_s + \hat{R}_j + \hat{R}_n \qquad (5.22)$$

$$X_j = V_p V_p^T X_B, \quad V_p = [\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_p] \qquad (5.23)$$

$$X_s = X_B - V_p V_p^T X_B \qquad (5.24)$$

where $\hat{R}_B$ is the sample covariance matrix of the radar return containing the RFI, noise and sea echoes; $\hat{R}_s$ is the sample covariance matrix of sea echo. In this application, Q, the number of non-zero eigenvalues is much less than M_d. p is set to '1', because the eigenvector $\underline{v}_1$ corresponding to the biggest eigenvalue spans the dominant subsignal space and this signal space is of interest.

So far, all the analysis above is based on X_A, which treats Doppler as variables,

so the dimension of $\hat{R}_A$ is M_d. If the range is taken as a variable, e.g. $\hat{R}_A = \frac{1}{M_d} X_A X_A^T$, the dimension of $\hat{R}_A, \Lambda$, and V will be N_r. Taking range as a variable is used to mitigate ionospheric clutter, because II is weakly correlated in range and strongly correlated in Doppler. The analysis for this case is similar and thus omitted. Examples are shown at the end of chapter 7 for the demonstration of future operational application.

The performance of this approach is critically determined by the accurate extraction of the RFI signal subspace. Generally, the RFI covariance matrix is estimated from far range bins, e.g. >150 km, where sea echo signals are sufficiently attenuated to be ignored. The bigger M_d we select, the better extraction of RFI. It is suggested to set M_d to 1024, which is called a 'full processing'. However, RFI-alone range bins over the whole Doppler domain are not always available, e.g. meteor trails echoes may occupy a few Doppler bins at far range bins. The solution for this problem is "patch processing", i.e. selecting a smaller Doppler range to avoid meteor trails.

5.1.3.3 Segmentation processing

Segmentation processing, which is not always necessary, is the second step in this 3-step clutter mitigation scheme. It is used to simplify the complexity of the RD image by removing unwanted contributions in the training area or processing area. Two window processing methods are developed for segmentation processing. The window sliding step by step along the Doppler/Range axis is called Doppler/Range- window. Four parameters are important: sliding length l, step length sl, range-window length wrl, Doppler-window length wdl. Basically, "window processing" is a sn step iterative self-training processing, where $sn = \frac{l}{sl}$. For each step, the training matrix is equal to the processing matrix, i.e. the same samples in the window area (a sub-matrix of X) are chosen to construct both $\hat{R}_A$ and $\hat{R}_B$. Then the dominant subsignal is extracted and subtracted by subspace projection.

5.1.3.4 Experimental validation

Data collected by the Pisces radar from Castlemartin (on the UK S Wales coast) in Feb. 2005 (Wyatt et al. 2006) after Doppler processing are used to test the performance of 3-step RFI clutter mitigation algorithm. The Doppler spectrum has a range of -2.5Hz to +2.5Hz, but the unit of Doppler bins is used here in the

RD image instead of Hz for the convenience of the explanation of the mitigation scheme.

In a low complexity case, the area beyond rb 15 in the RD image with only RFI is available. No segmentation method is needed here, because there is no other clutter. The mitigation result is shown in Fig.5.5(a)(b) (out-band RFI) and Fig.5.6 (in-band RFI). This algorithm works well both for a symmetric and non-symmetric, negative Doppler frequency and positive Doppler frequency, in-band and out-band distributed RFI. In the figures, the DS of rb 5 (about 75km from the radar site) is selected for more detailed comparison because the derived oceanographic estimates can be compared with that of a wave buoy located in the same area.

In a middle complexity case, Fig.5.7, the Bragg peaks extend to rb 18 and are not weak enough to be ignored, and there exists meteor trail clutter (MTC) from rb 18 to 20. An example is the dataset obtained from 10am, Feb 2nd. If I choose rb 15-20 and Doppler bin 1-1024 to construct X_A, the sea echo and MTC component are added into X_A and $\hat{R}_A$. To solve the problem, the unwanted MTC is mitigated by applying the "Doppler-window" method. I select an area with meteor trails only, $r = 18 \ldots 20, d = 568 \ldots 809,\ \ wrl = 3, l = 242,\ \ sl = 1,$ $wdl = 5$, in the RD spectrum. In the selection of the training matrix, sea echoes, the area of which are in Doppler bins from 430 to 620 and range bins from 1 to 17, is to be avoided. After the meteor trails are suppressed as shown in Fig 5.7(a), "full processing" is applied to an area, $r = 18 \ldots 20, d = 1 \ldots 1024, N_r = 3,$ $M_d = 1024$, in the RD spectrum to extract the RFI eigenvalues. The performance in the RD spectrum and the Doppler spectrum of rb 5 is shown in Fig. 5.7(a)(b). In this example, the RFI can be removed without removing the meteor trails, i.e. a 'patch processing' choosing $r = 18 \ldots 20, d = 1 \ldots 419, N_r = 3, M_d = 419$.

Removing the meteor trails is understood as a way of simplifying the middle complexity case to a low complexity case, which enables the whole Doppler range (1-1024) to be made use of for the construction of the training matrix, and thus gives a better removal of the RFI. The mitigation performance of another example is shown in Fig 5.8. These two examples differ in that the former is out-band and the latter is in-band RFI.

Low and middle complexity cases have been shown above, but some of them don't contribute to poor quality oceanographic estimates because the RFI is out-band. In the high level of complexity case, RFI occupies the first- and second- order Doppler areas, superposed with either range-correlated or Doppler-correlated clutters, e.g. MTC in far ranges and II in middle or close

ranges. The training matrix is obtained after "Doppler-window" or "Range-window" processing. The Doppler-window is used for the mitigation of range-correlated clutter like MTC or RFI. Notice that the length of the window is important. By varying the wdl length from 2 to 30, it was found that 3 gives the best suppression performance for many cases. Similarly, the Range-window method is used for the mitigation of Doppler-correlated clutter like ionospheric clutter. Varying wrl from 2 to 7, it was found that 5 is the best for many cases. Two examples are given. The first dataset I choose is 3am Feb.13th. The close range ionospheric clutter is mitigated by the Range-window, as is shown in Fig 5.9(a)(b). The second dataset is from 10am Feb. 28th. The Doppler-window was adopted for the mitigation of sparsely and evenly distributed RFI in the Doppler domain, shown in Fig 5.10(top). Fig 5.10 and 5.11 both used patch processing, but the difference is that the training matrix X_A of the latter excluded the Doppler area with a local clutter (the local clutter is the short line visible at Doppler bin 433 beside the sfi from rb 3 to 6). The clutter signal has a negative effect on wave data validity. When the clutter is mitigated together with RFI as shown in Fig 5.11 (middle), the performance of mitigation scheme is better. The buoy estimate (see Table 5.1 below) of the significant height h_s was 0.67m, while the original radar estimation was 6.24m. After suppression of RFI in Fig.5.10 and RFI-plus-target in Fig 5.11, the radar estimation was reduced to 2.17m and 0.67m respectively.

If I assume the most serious situation, which is not found in the data used in this thesis, but is possible in other circumstances, that meteor trail clutter, RFI, and ionospheric interference are superposed together in the first- and second-order Doppler areas, the segmentation scheme can also deal with that complicated case. The parameters and training matrix must all be correctly selected to ensure that the subsignals are well separated from each other.

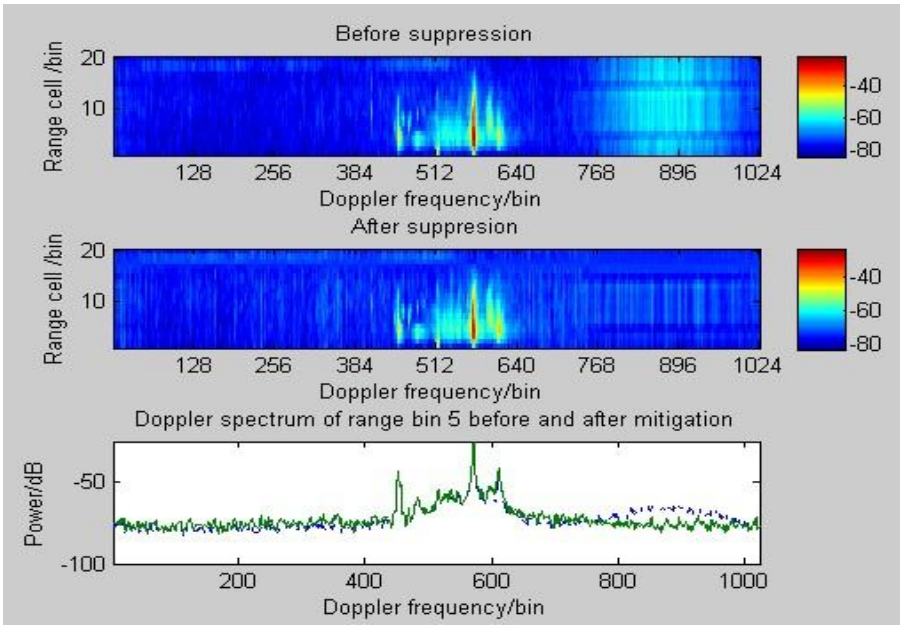


Fig.5.5a. RD image of low complexity. Top: before suppression; Middle: after suppression; Bottom: Doppler spectrum of rb 5. After mitigation: the solid line; Before mitigation: the dashdot line. The color scale shows the power spectrum in dB. Each Doppler bin represents 0.0048Hz. The frequency range is from -2.5Hz to 2.5Hz. Each range bin represents 15 km. The training matrix A is selected from rb 18 to 20. Data was taken at 10am 15/02/2005

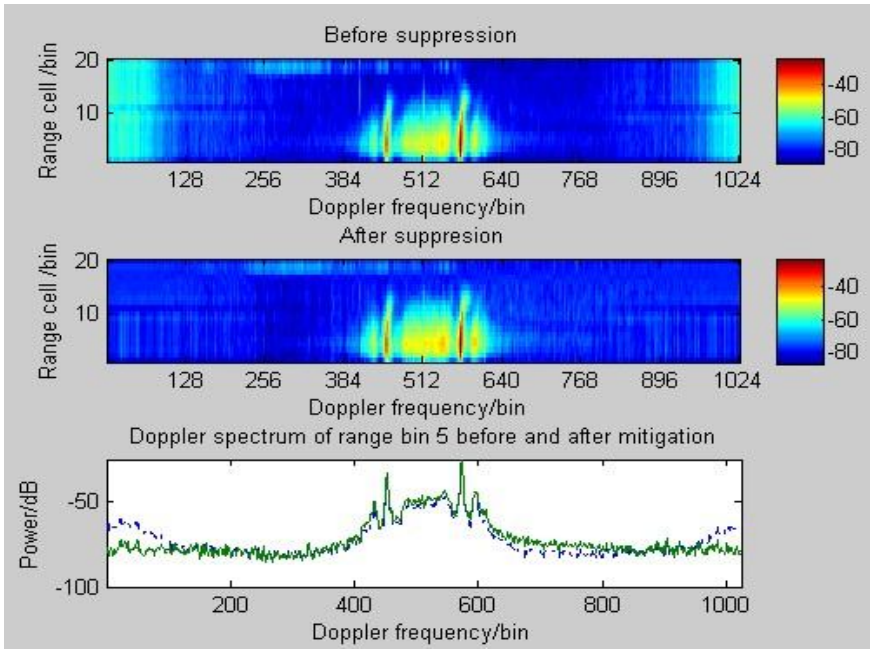


Fig 5.5b. RD image of low complexity. Top: before suppression; Middle: after suppression; Bottom: Doppler spectrum of rb 5. After mitigation: the solid line; Before mitigation: the dashdot line. The color scale shows the power spectrum in dB. Each Doppler bin represents 0.0048Hz. The frequency range is from -2.5Hz to 2.5Hz. Each range bin represents 15 km. The training matrix A is selected from rb 16 to 18. Data was taken at 10am 13/02/2005

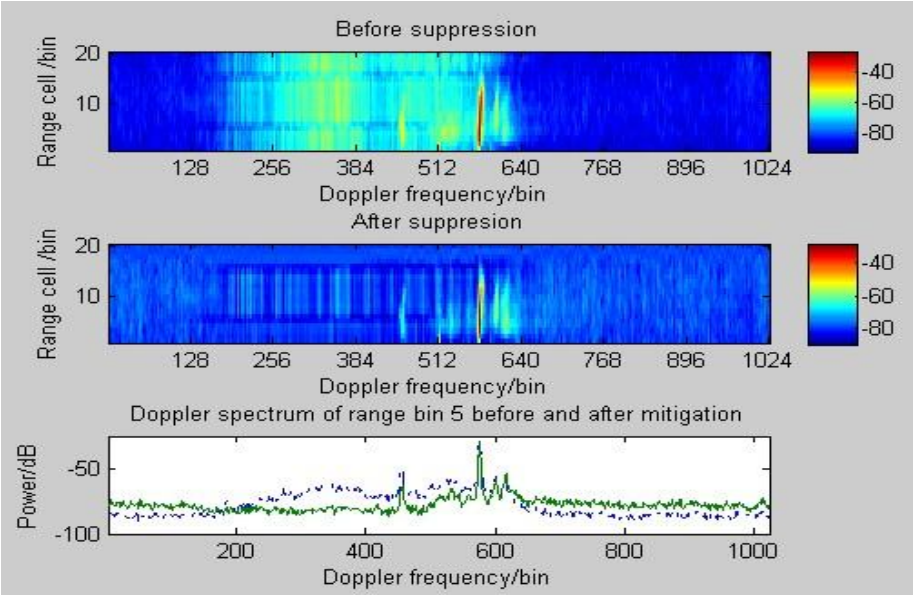


Fig 5.6 RD image and DS of low complexity. Top: before suppression of RFI; mid: after suppression of RFI; bottom: Doppler spectrum of rb 5. Before: the dashdot line; After: the solid line The color scale shows the power spectrum in dB. Each Doppler bin represents 0.0048Hz. The frequency range is from -2.5Hz to 2.5Hz. Each range bin represents 15 km. Full processing is performed. The training matrix A is selected from rb 18 to 20. Data was taken at 13pm 01/02/2005.

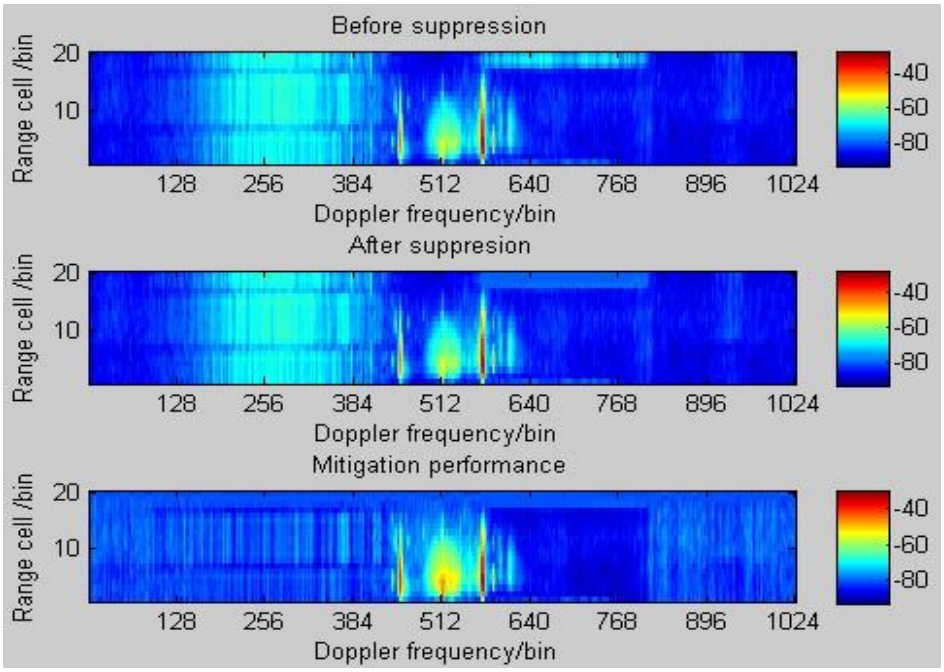


Fig 5.7a. RD image of middle complexity. Top: before suppression of meteor trails; middle: after suppression of meteor trails; bottom: after suppression of RFI. Range-sliding-window is performed. The color scale shows the power spectrum in dB. Each Doppler bin represents 0.0048Hz. The frequency range is from -2.5Hz to 2.5Hz. Each range bin represents 15 km. Data was taken at 10am 02/02/2005.

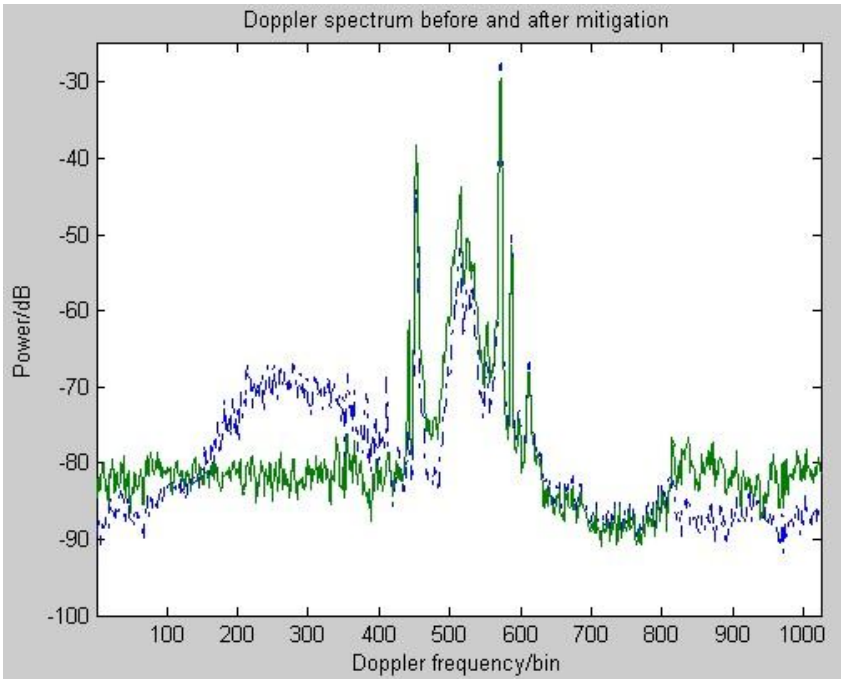


Fig 5.7b. Original and RFI mitigated DS. Dashdot line: RD before suppression of RFI; Solid line: RD after suppression of RFI. Each Doppler bin represents 0.0048Hz. The frequency range is from -2.5Hz to 2.5Hz. Patch processing is performed. Data was taken at 10am 02/02/2005.

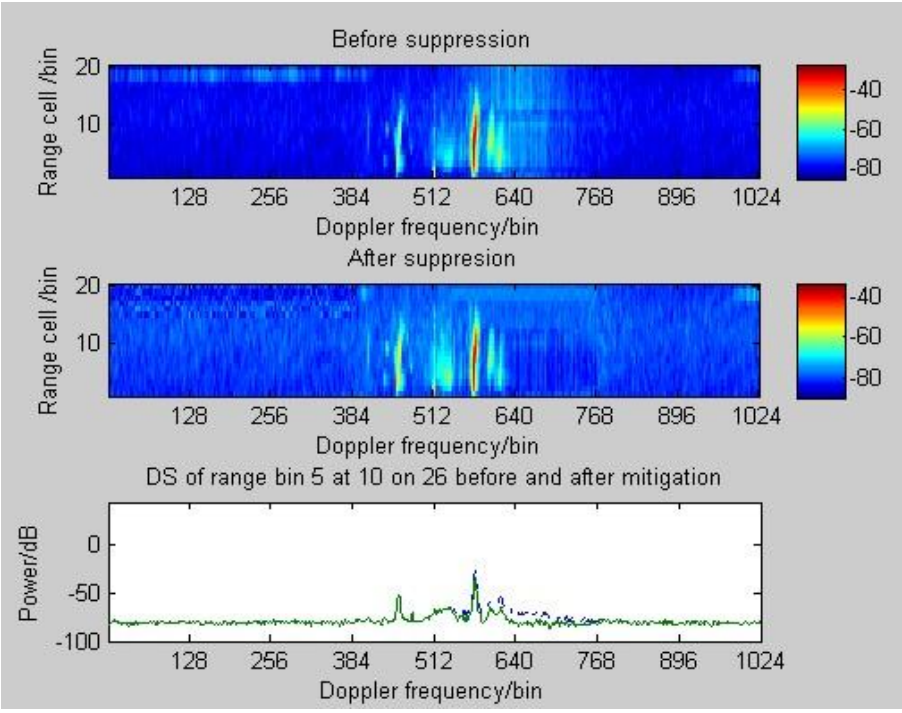


Fig 5.8. RD image and DS of middle complexity. Top: before suppression of RFI and meteor trail; mid: after suppression of RFI and meteor trail; bottom: Doppler spectrum of rb 5. Before: the dashdot line; After: the solid line. Doppler-window and patch processing are performed. Data was taken at 10am 26/02/2005.

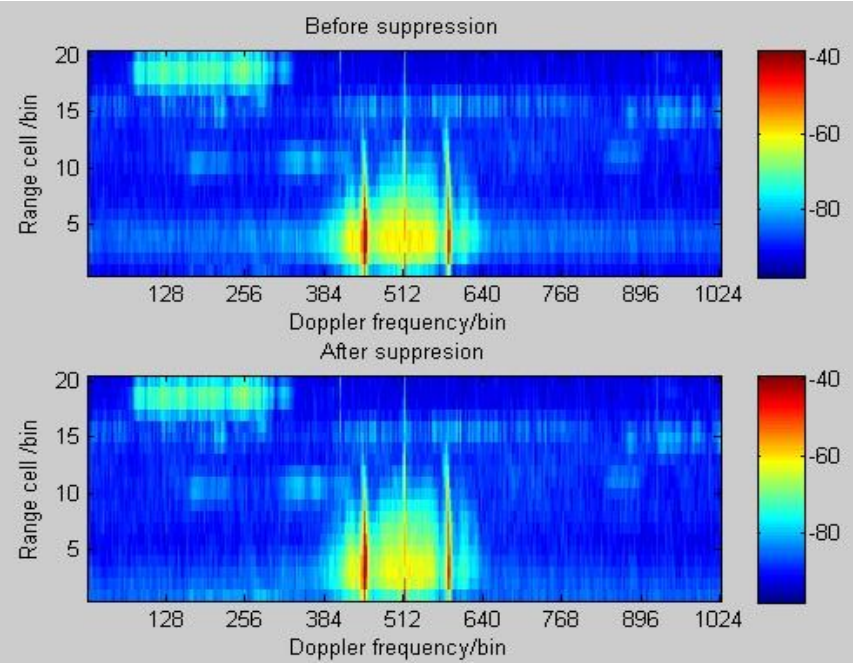


Fig.5.9a. RD image of high complexity. Top: before suppression of Ionospheric interference; Bottom: after suppression of ionospheric interference. The color scale shows the power spectrum in dB. Each Doppler bin represents 0.0048Hz. The frequency range is from -2.5Hz to 2.5Hz. Each range bin represents 15 km. Ionospheric interference mitigation is implemented by RD-Range-sliding window in rb 5. Data was taken at 3am 13/02/2005.

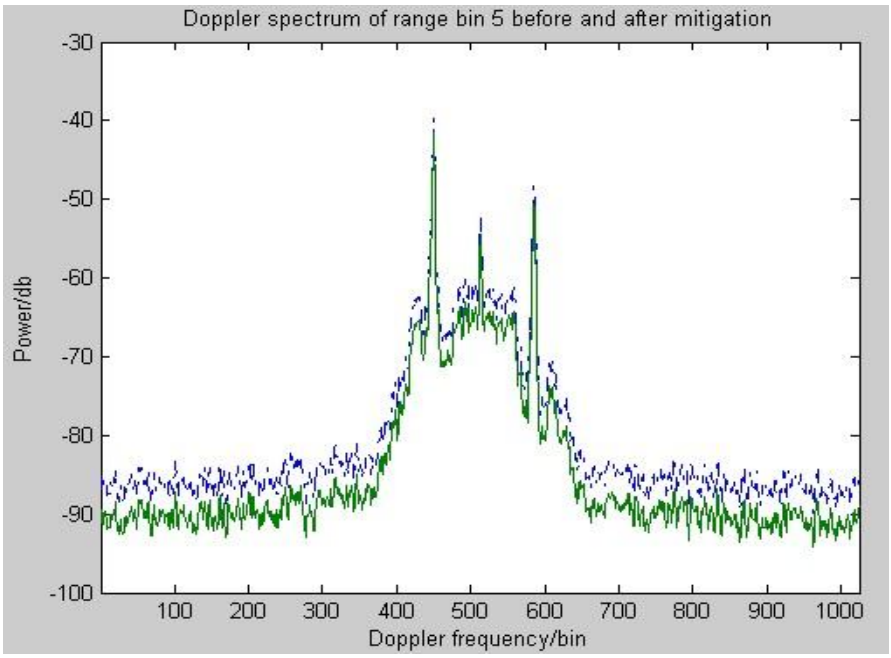


Fig 5.9b. Original and RFI-mitigated DS. Before: the dashdot line; After: the solid line. Each Doppler bin represents 0.0048Hz. The frequency range is from -2.5Hz to 2.5Hz. The difference between them is the mitigated ionospheric interference. Data was taken at 3am 13/02/2005.

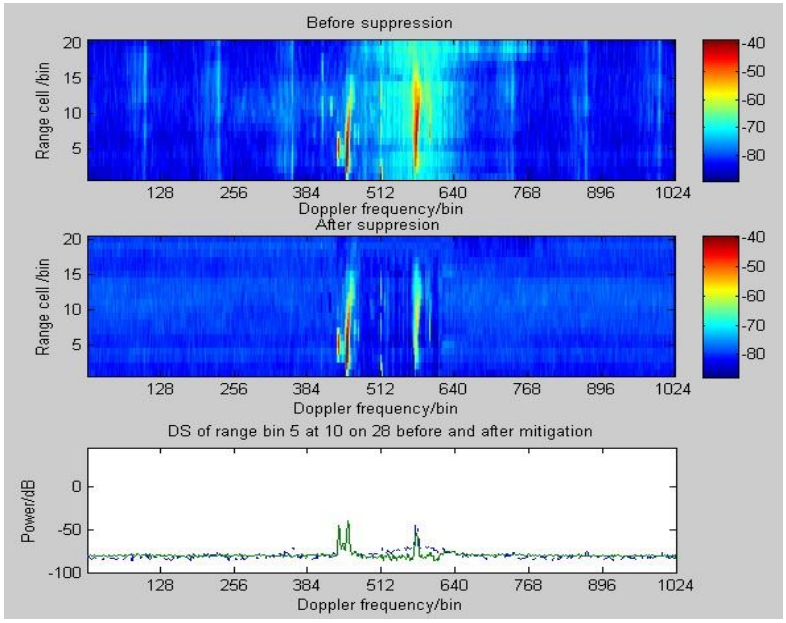


Fig 5.10. RD image and DS of high complexity. Top: before suppression of RFI; mid: after suppression of RFI; bottom: DS of rb 5. After mitigation: the solid line; Before mitigation: the dashdot line. The color scale shows the power spectrum in dB. Each Doppler bin represents 0.0048Hz. The frequency range is from -2.5Hz to 2.5Hz. Each range bin represents 15 km. Doppler-window and patch processing are performed. Data was taken at 10am 28/02/2005.

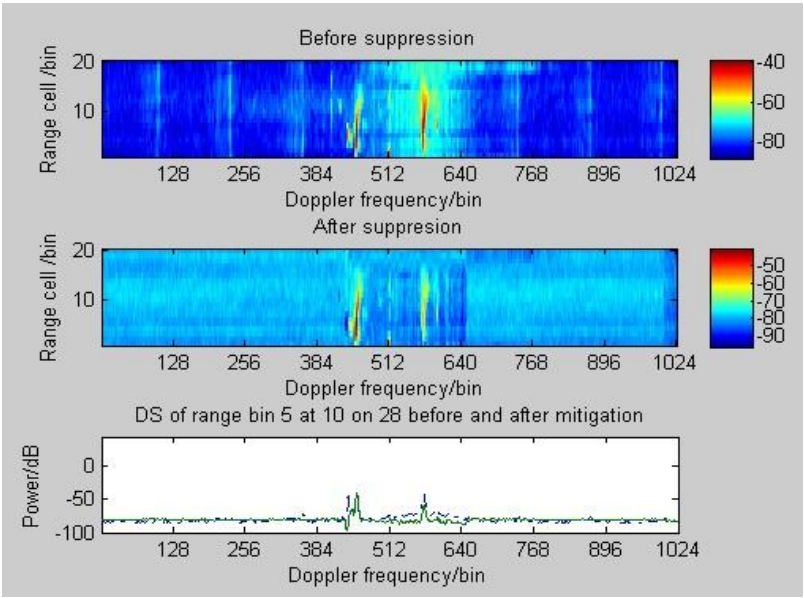


Fig 5.11. RD image and DS of high complexity. Top: before suppression of RFI and the target; mid: after suppression of RFI and the target; bottom: Doppler spectrum of rb 5. After: the solid line; Before: the dashed line. The color scale shows the power spectrum in dB. Each Doppler bin represents 0.0048Hz. The frequency range is from -2.5Hz to 2.5Hz. Each range bin represents 15 km. Doppler-window and patch processing are performed. Data was taken at 10am 28/02/2005.

5.1.3.5 Quantitative comparison

Radar data with good accuracy will give wave measurements that compare well with the buoy. In this comparison, single radar datasets are used because the mitigation of RFI is realized in the DS from a single radar site. It is more rational to validate the performance of the mitigation algorithm by comparing the single radar oceanographic parameters with the buoy estimates before and after mitigation. The radar data I have selected to illustrate performance are typical of all the data with RFI, and include all possible distributions of a mixture of RFI, meteor trails and ionospheric signals that were found in the data.

The quantitative comparison of the scheme is given in Table 5.1, where 'old' and 'new' denote the parameters estimated by the original and the processed radar data. Three parameters have been used: θ (the angle between peak wave direction and beam bearing), h_s (significant waveheight) and T_1 (mean period). The single radar provides empirical estimates of h_s and T_1 (Wyatt 2002). It has been shown that single radar estimates of h_s provide two alternative estimates: h_s per and h_s par. h_s per assumes ocean waves are propagating roughly perpendicular to the radar, while h_s par assumes ocean waves are propagating roughly parallel to the radar beam. To choose the correct estimate to compare with the buoy, I calculate the angle θ between the radar beam and the wave peak direction measured by the buoy. If $\theta < 45$, h_s par is adopted, while if $\theta > 45$, h_s per is selected. I have indicated this by putting the buoy data in the appropriate column.

Table 5.1. A comparison among parameters by Buoy, Radar old, and Radar new. θ-the angle between peak wave direction and beam bearing; h_s-significant waveheight; and T_1-mean period. The numbers in bold represent buoy estimates, and they are put in the column with the more accurate h_s option provided by a single radar.

Day/Hour	Data source	θ /°	T_1 /s	h_s per /m	h_s par /m	Fig
13/03am	Buoy	69	6.54	**3.67**		**5.9**
	Old		5.34	4.49	2.66	
	New		5.38	3.87	2.42	
28/10am	Buoy	63	6.24	**0.67**		**5.10/5.11**
	Old		8.75	6.24	3.57	
	New		4.76	0.64	0.95	
02/10am	Buoy	50	3.6	**1.05**		**5.7**
	Old		7.17	0.99	1.23	
	New		7.17	1.0	1.23	
26/10pm	Buoy	40	3.6		**1.1**	**5.8**
	Old		4.14	0.78	1.07	
	New		4.14	0.82	1.1	
01/13pm	Buoy	44	4.14		**1.33**	**5.6**
	Old		4.22	0.94	1.19	
	New		4.13	1.1	1.3	

As can be seen in the table, h_s has always been improved. For example, the buoy observation of h_s is 0.76m at 10am on Feb. 28th. The original radar estimation of h_s is 1.26m while it is reduced to 1.03m after suppression. The relative difference of the estimation is reduced from 88% to 54%. The improvement of h_s is not too much for some other cases. It is easy to understand that the RFI will not seriously impact the estimation accuracy when it is out-band, eg 02/10am (see Fig 5.7). Although the amount of data presented here is small, the cases are typical and it is reasonable to draw the conclusion that this three-step scheme will be effective to mitigate RFI and other clutters more generally.

5.2 Visible clutter mitigation on arbitrary clutters

An analysis of the mathematical model of RFI has been presented above. A

solution capable of dealing with the majority of complicated situations with RFI superposed by meteor trails, target and II is demonstrated. The processing procedures are image recognition, segmentation processing and signal subspace projection. Image recognition is the first and also the most important step, checking the features of the clutters in the problematic RD image. Segmentation processing is the second step. It removes non-RFI or non-II subsignals from a window area and slides along either the range or Doppler domains. The final subspace projection method is the essential one and is based on eigenvalue decomposition analysis. The differences found between buoy data, original radar data and processed radar data imply that the method is promising.

5.2.1 Introduction

After the detailed investigation of the three-step solution, i.e. image recognition, segmentation and subspace projection, has been given for the visible RFI mitigation, an extension of this solution to arbitrary visible clutter mitigation is demonstrated in this section. With this generalized algorithm, not only RFI and II, but other unknown and irregular-shape clutters in the RD image can also be mitigated. Generally speaking, any visible clutter in the RD image can be classified into one of the four types: range-correlated over many Range bins (see Fig. 5.17), range-correlated over few Range bins (see Fig. 5.18), Doppler-correlated over many Doppler bins (see Fig. 5.19), and Doppler-correlated over few Doppler bins (see Fig. 5.20). Detailed examples of each type of the RD image and their mitigation performance are given in the validation section 5.2.6.

The procedure is as follows.

1. An area from the RD image with the visible clutter is selected to be the "training matrix" or "clutter matrix", X_A.

2. Another area in the RD image with clutter of similar distribution as that in X_A is selected to be the "processing matrix", X_B (see Fig. 5.12). Any rectangular area (N_r range bins * M_d Doppler bins) in the RD spectrum can be selected to construct X_A and X_B. Note that X_A and X_B can be the same.

3. Covariance matrices $\hat{R}_A$, $\hat{R}_B$ of X_A and X_B are constructed.

4. Eigenvalue decomposition analysis is carried out on $\hat{R}_A$, to produce an eigenvalue matrix Λ and an eigenvector matrix V.

5. Selected eigenvectors in V are used to span the Eigenspace matrix $P_{eig} = VV^T$.

6. X_B is projected into the Eigenspace P_{eig} to generate the clutter subsignal space $P_b = VV^T X_B$.

7. P_b is subtracted from X_B to mitigate the clutter component.

The RD image is considered to be a superposition of clutter, sea echoes, interferences, and background noise floor (see section 5.1.3.2A). Hereafter, the model is simplified by considering interferences and clutters both as clutters, i.e. non-sea signals. Therefore, the model in matrix form is given as:

$$\begin{aligned} X_A &= X_c + X_{sea} + X_n \\ X_B &= X_c + X_{sea} + X_n \end{aligned} \tag{5.25}$$

where X_c represents the clutter signal matrix; X_{sea} is the sea echo matrix; the background noise floor is noted as matrix X_n.

This generalized clutter mitigation scheme can be classified into two categories, unsupervised or supervised, according to whether the training matrix equals the processing matrix or not, i.e. $X_A = X_B$ or $X_A \neq X_B$ respectively. For each category, algorithms are developed for two different classes, static or dynamic, according to whether the mitigation processing is a one-off or iterative processing. For example, the methods using Range-sliding window and Doppler-sliding window discussed in section 5.1.3.3 are in the dynamic class of the unsupervised category, because in every sliding step, the window is used as both A and B. The static algorithms can also be viewed as the kernel processing of the dynamic algorithms. The framework of the whole scheme is shown in Fig. 5.13.

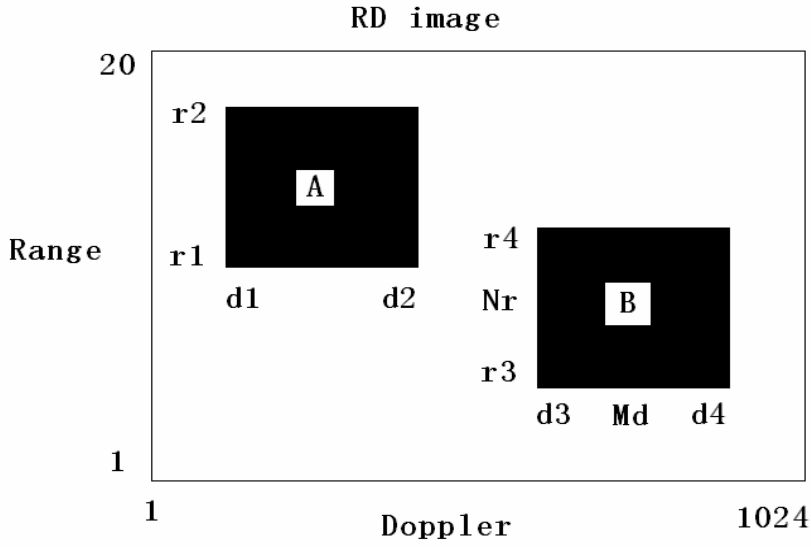


Fig. 5.12 Training region A defined by range bins r1 to r2 and Doppler bins d1 to d2; processing region B defined by range bins r3 to r4 and Doppler bins d3 to d4.

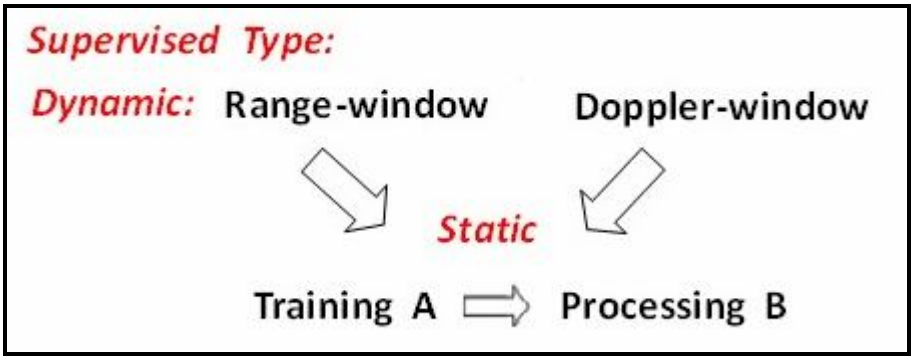


Fig.5.13 The framework of the generalized visible clutter mitigation scheme. Range-window is used for Doppler-correlated clutter and Doppler-window is used for range-correlated clutter. The framework of the unsupervised type is the same as the supervised except that $A = B$.

There are two static schemes implemented for the supervised clutter mitigation: for range-correlated clutter mitigation, e.g. RFI; for Doppler-correlated clutter mitigation, e.g. II. Their application has been demonstrated in section 5.1. There are four types of static schemes implemented for the unsupervised clutter mitigation. They are discussed in details in the next section.

5.2.2 Unsupervised static processing

Unsupervised static processing is carried out as follows. First, select an area in the RD image as the processing matrix X_B. X_B could contain clutter only $X_B = X_c + X_n$, or sea echo only $X_B = X_{sea} + X_n$ or clutter plus sea echo

$X_B = X_c + X_{sea} + X_n$. If $X_B = X_c + X_{sea} + X_n$, clutters can either be larger or smaller than sea echoes. Therefore, unsupervised static algorithms are designed in accordance with different compositions of X_B, for either extracting sea echoes or subtracting clutters. All of the algorithms follow the subspace projection theory explained in section 5.1.3.2. The four algorithms are explained below. They form the kernel function of unsupervised dynamic processing explained in section 5.2.3.

1. $\mathbf{X_B = X_c + X_n}$ where X_c is the dominant signal and the goal is to mitigate the dominant clutter.

The covariance matrix is constructed in the Doppler domain, i.e. $\hat{R}_B \in R^{M_d \times M_d}$, $M_d = d4 - d3$. The eigenspace P_{eig} is spanned by the eigenvector $\underline{v}_1 = [v_1, v_2, ..., v_{M_d}]^T$ that corresponds to the largest eigenvalue by $P_{eig} = \underline{v}_1 \underline{v}_1^T$. The dominant subsignal, X_c, is obtained by $P_{eig} X_B$. Then, the minor subsignal is calculated by subtracting the dominant signal $X_B - X_c$. The scale of $X_B - X_c$ is ±5 dB, representing some insignificant Doppler components from minor sources. Lastly, X_n is supplemented with a 'base' signal X_{base} to keep it in the same scale $X_n = X_B - X_c + X_{base}$. For this mitigation purpose, X_{base} is constructed using an area with the same shape as the processing area and without any obvious clutter and sea echo from the RD image.

2. $\mathbf{X_B = X_{sea} + X_n}$ or $\mathbf{X_B = X_{sea} + X_c + X_n, X_{sea} \gg X_c}$ where X_{sea} is the dominant signal and the goal is to extract the dominant sea echoes.

$\hat{R}_B$ and P_{eig} are constructed in the same way as above. The dominant subsignal, X_{sea}, is obtained by $P_{eig} X_B$.

3. $\mathbf{X_B}=\mathbf{X_c}+\mathbf{X_{sea}}+\mathbf{X_n}$ and $\mathbf{X_{sea}}>\mathbf{X_c}$ where X_{sea} is the dominant signal and the goal is to mitigate the clutters.

P_{eig} is spanned by the M_d-element eigenvector $\underline{v}_2$ that corresponds to the second biggest eigenvalue by $P_{eig} = \underline{v}_2 \underline{v}_2^T$. Then, the subdominant subspace of X_B is obtained by $X_c = P_{eig} X_B$. The dominant sea signal is calculated by subtracting the subdominant signal $X_{sea} + X_n = X_B - X_c$.

4. $\mathbf{X_B}=\mathbf{X_c}+\mathbf{X_{sea}}+\mathbf{X_{bg}}$ and $\mathbf{X_{sea}}<\mathbf{X_c}$ where X_c is the dominant signal and the goal is to extract the sea echoes.

P_{eig} is spanned by the eigenvector $\underline{v}_2$ as $P_{eig} = \underline{v}_2 \underline{v}_2^T$. The subdominant signal, X_{sea}, is obtained by $P_{eig} X_B$. Then the sea signal is constructed by supplementing the subdominant signal by the 'base' signal: $X_{sea} + X_n = P_{eig} X_B + X_{base}$. For this, X_{base} is constructed using (1) the average of the range profiles of X_B at each Doppler bin, if the clutter signal is Doppler-correlated; (2) the average of the Doppler profiles of X_B at each range bin, if the clutter is range-correlated.

5.2.3 Unsupervised dynamic processing

Unsupervised dynamic processing refers to sliding-window processing in either the Range or the Doppler domain step by step. In each step, it can perform a selected algorithm from the four unsupervised static methods explained above, based on the composition of X_B. These schemes are Range-window and Doppler-window processing according to the definition introduced in section 5.1.3.3, i.e. the window sliding step by step along the Doppler/Range axis. The good feature of these schemes, compared to the segmentation method introduced in 5.1.3.3, is the window processing in each step is generalized from the four algorithms in unsupervised static processing explained above. Therefore they are able to pre-filter clutters or pre-extract sea echoes. In terms of 'sliding', three parameters are important: sliding length l, step length sl, window length wl. The

unsupervised dynamic scheme is a sn step iterative processing, where $sn = \frac{l}{wl}$.

The framework of unsupervised dynamic schemes is similar to that of the supervised dynamic processing in Fig. 5.16. The only difference is the training area A is processing area B in unsupervised dynamic processing.

As discussed before, the window length wl is a key parameter and is determined by the amplitudes of the clutters in the whole sliding area. For the case that the clutter is evenly distributed (terms such as 'evenly, local, random' are given from the image-processing point of view) over a large scale, a larger window length wl can be chosen. A bigger step length sl is chosen correspondingly. For this work, sl is set to be equal to wl since no benefit was found for making $sl < wl$. For the case of local and random clutter distribution, a smaller wl is selected, and a small sl is also chosen ($sl = wl$).

5.2.4 Supervised static processing

Supervised static processing refers to a case where areas A and B are selected separately in the RD image. This case is taken into account because, if the amplitudes of clutters in B are at the same level as that of the sea echoes, i.e. $X_c \approx X_{sea}$, the unsupervised approach is not capable of differentiating the clutters from the sea echoes. The supervised approach thus is to first identify a training area A, which contains clutter only and where the clutter has similar distribution of amplitudes in A as in B, i.e. the source of the clutter in A is the same source of the clutter in B. In this application, the clutter in B is invisible, but it is visible in A. For an exact mitigation of clutters hidden in sea echoes, the training area A must be properly selected.

The clutters from different sources show various features in the RD image, i.e. small or large in scale, range or Doppler correlation. Considering this, all the possible features of clutters are investigated in the supervised static approach, and specific constructions of area A and B are provided to deal with these features. Generally speaking, depending on whether the clutter is range-correlated or Doppler-correlated, e.g. RFI is range-correlated and II is Doppler-correlated, A and B are constructed using Doppler and range as variables respectively. If Doppler is considered as variable, $d4 - d3 = d2 - d1$, i.e. the M_d of matrix A and B should be the same, while if the range is the variable, $r4 - r3 = r2 - r1$ (see Fig. 5.14).

To construct the training matrix X_A the area A is chosen to be as large as possible. The reason is the bigger the size of a matrix, the better representation of the signal space after eigenvalue decomposition. Therefore, instead of using a continuous area, the training area A can be constructed using two separate areas.

Fig. 5.15 shows the construction of training area A with one and two areas. Suppose there is a Doppler-correlated II in rb 3-5 over all Doppler bins. To mitigate this II, training area A is often selected from range bins 1-20 and Doppler bins 800-900 (without sea echoes) and processing B is selected from range bins 1 to 20 and range bins 1 to 1024 (see Fig.5.15(a)). Suppose another II occupied rb 11-13 over Doppler bins 800-900 in addition to the II in rb 3-5. The construction of matrix X_A must exclude the area rb 11-13, because the II in rb 3-5 is the clutter to be removed. Using the one area technique, rb 1-10, db 800-900 will be selected for X_A. Using two areas, two discontinuous areas can be connected by choosing rb 1-10 and rb 14-20, db 800-900. Due to the latter construction, the dimension of $\hat{R}_A$ is enlarged from 10 to 17 range bins, the dominant eigenvector spans a more characteristic signal subspace of II.

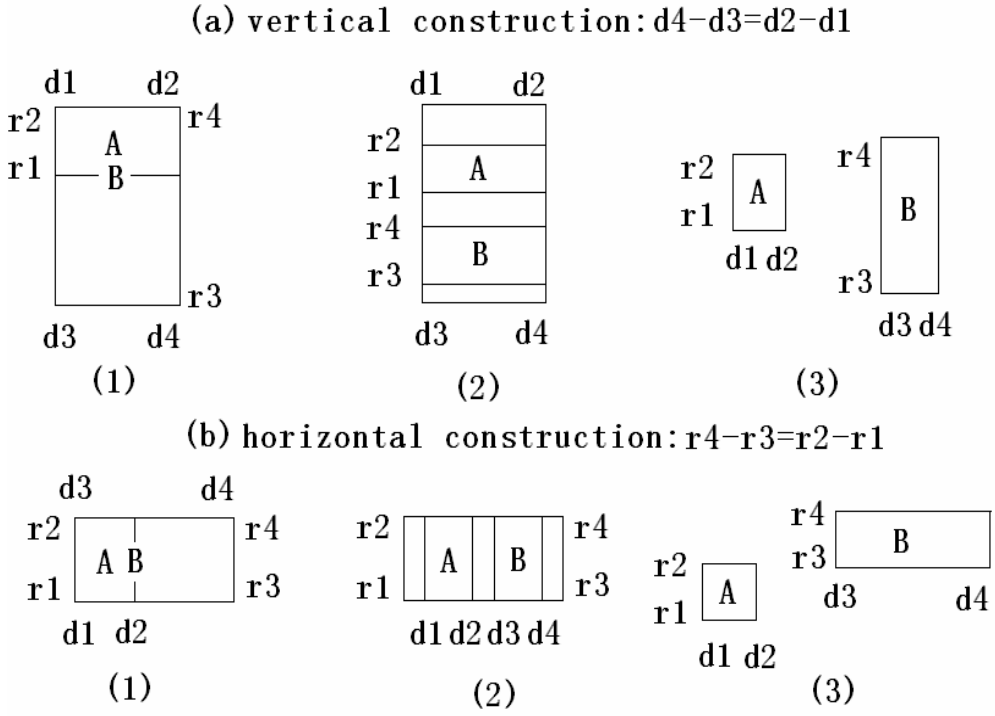


Fig.5.14 Six types of construction of training area A and processing area B. (a) shows the vertical construction and such that $d4-d3=d2-d1$: (1) A is inside B; (2) A is in the surrounding Range bins but the same Doppler region; (3) A is another area out of B. (b) shows the horizontal construction and such that $r4-r3=r2-r1$: (1) A is inside B; (2) A is in the surrounding Doppler bins but the same Range region; (3) A is another area out of B.

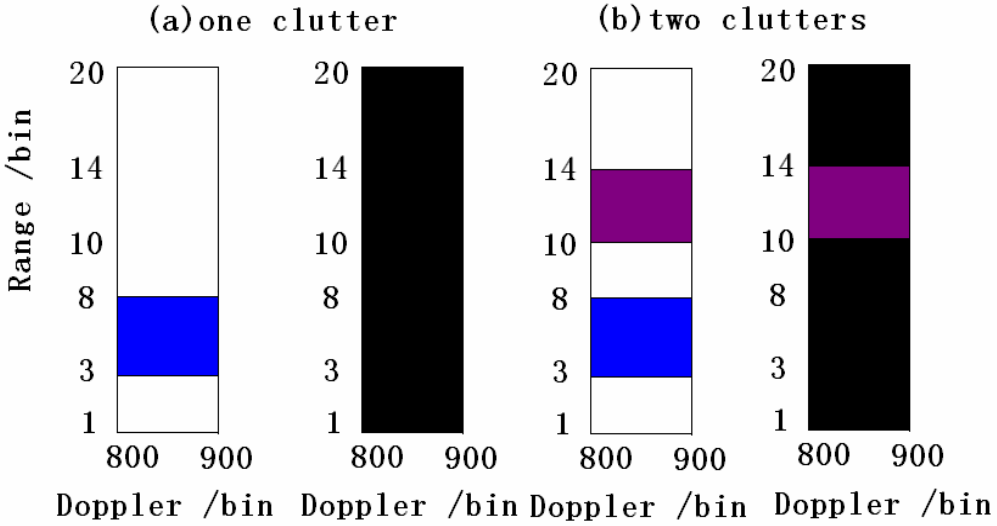


Fig.5.15 Using one and two areas to construct training matrix A. Blue and purple colors show two range-correlated II occupying different ranges. (a) and (b) show the composition of signals in the left plot. In the right plot, the black area is selected for the training matrix A. In (a), the area A is from rb 1-20 and db 800-900; in (b), the area A is from rb 1-10 and rb 14-20 and db 800-900, avoiding the ranges of the other II.

When the areas A and B are fixed, the second step is to construct the clutter subspace from A and the third step is to project the signal space of B into the orthogonal subspace of the clutter subspace. The second and third steps are performed according to procedures 3 to 7 explained in section 5.2.1.

5.2.5 Supervised dynamic processing

Similar to the definition of unsupervised dynamic processing, supervised dynamic processing refers to a sliding-window processing in either the Range or the Doppler domain step by step. In each step, it can perform one from the six optional kernel algorithms (using algorithms 0-2 for the visible case, using 3-5 for invisible case) integrated in Matlab routine, based on the composition of X_B.

The supervised dynamic processing requires four parameters defining A and B, i.e. r_1, r_2, d_1, d_3 for Doppler-window and r_1, r_3, d_1, d_2 for Range-window (see Fig. 5.16), together with three parameters defining the sliding window, i.e. sliding length l, step length sl, and window length wl. The Framework of this scheme is shown in Fig. 5.16. This approach is realized in the dynamic supervised Doppler-window and Range-window.

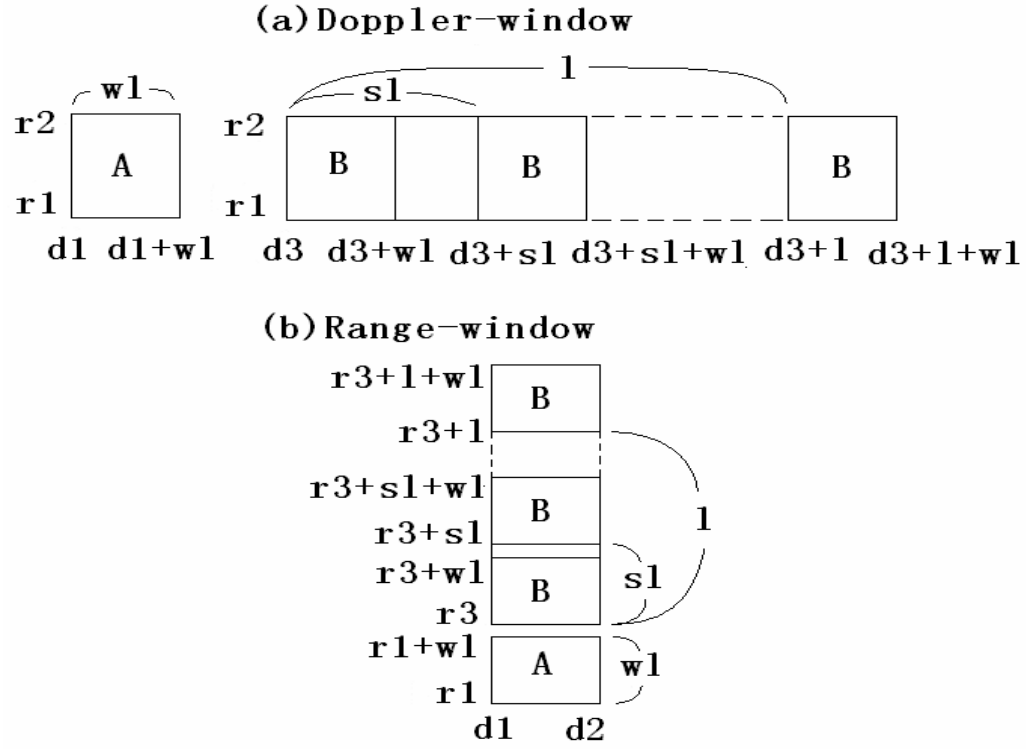


Fig.5.16. Supervised dynamic processing, where wl is the window length, sl is the step length, and l is the sliding length. (a) sliding window in the Doppler domain. The size of training region A is fixed and defined by range bins r1 to r2 and Doppler bins d1 to d1+wl. In the first sliding step, the processing region B is defined by range bins r1 to r2 and Doppler bins d3 to d3+wl. A Doppler-correlated or range-correlated static scheme is applied in this step. In the second sliding step, A is not changed and B moves sl Doppler bins to the positive Doppler direction, i.e. Doppler bins d3+sl to d3+$sl + wl$. The static scheme is applied again. This sliding process continues until B has stepped l Doppler bins to the positive Doppler direction. (b) sliding window in the range domain. The procedure is very similar except that the sliding direction is in range.

5.2.6 Experiments and validation

The proposed visible clutter mitigation scheme is tested using the Pisces dataset (section 5.1.3.4). The figures shown below are typical examples of each visible clutter mitigation. For a consistent demonstration of the clutter mitigation performance, those figures conform to the same format showing RD image before the mitigation in the top figure, RD image after the mitigation in the middle figure and the difference in the bottom figure. Different from the experiments and validations for the RFI-mitigation, this experiment is done for the validation of mitigation performance of any type of clutter in the RD image. These examples shown below may not necessarily contribute to the improvement of wave estimation, but provide the opportunity for application to other projects.

For example, in an application for ship tracking, the sea echoes are not wanted any more and become clutters for ship echoes detection. The sea echoes can be mitigated while the ship signals are retained by using this generalized scheme, as long as a proper training matrix A and processing matrix B are selected.

Section 5.2.1 noted that any visible clutter in the RD image can be classified into one of the four types: range-correlated over many Range bins (see Fig. 5.17), range-correlated over few Range bins (see Fig. 5.18), Doppler-correlated over many Doppler bins (see Fig. 5.19), and Doppler-correlated over few Doppler bins (see Fig. 5.20). The top figure in Fig. 5.17 displays a clutter that covers the whole Range axis and the 'band' region in the Doppler axis. In the middle and bottom RD images, it is found that the range-correlated clutter mixed with sea echoes is mostly eliminated, with some clutter residual left. This is because the clutter subsignal space expanded by the eigenvectors in the training matrix A is not exactly the same in the processing matrix B. The mitigation performance is better, i.e. the clutter residual is less, when the degree of similarity between the clutter distributions in the area A and B is higher. The assumption made in the subspace projection is that the clutter in the training area is distributed in the same way as it is in the processing area. The top figure in Fig. 5.18 shows a clutter that covers the area within few rb 17-20 and db 350-890. In the middle and bottom RD images, it is found that this clutter from meteor trails is eliminated. Note that the method used here is a static (one-off) unsupervised processing that extracts the dominant clutter signal and then subtracts it, while the method introduced in the section 5.1.3.4 is a Doppler-window processing, i.e. a dynamic unsupervised processing. The top figure in Fig. 5.19 shows a clutter that covers the area within rb 11-15 and many db 1-1024. In the middle and bottom RD images, it is found that the Doppler-correlated clutter due to II is eliminated largely. In some areas, e.g. rb 10-13 db 850-900, II is not removed completely, since the assumption that the clutter is homogenous over the whole Doppler axis is not held there. Inhomogeneous electron densities in the ionosphere (see chapter 3) is a main reason for unevenly distributed II in the RD image. The top figure in Fig. 5.20 shows a clutter that covers the area within rb 11-15 and few db 916-971. In the middle and bottom RD images, it is found that the patch-size range-correlated clutter is mostly eliminated. This patch is chosen just for the demonstration of a successful mitigation of an arbitrary patch size clutter.

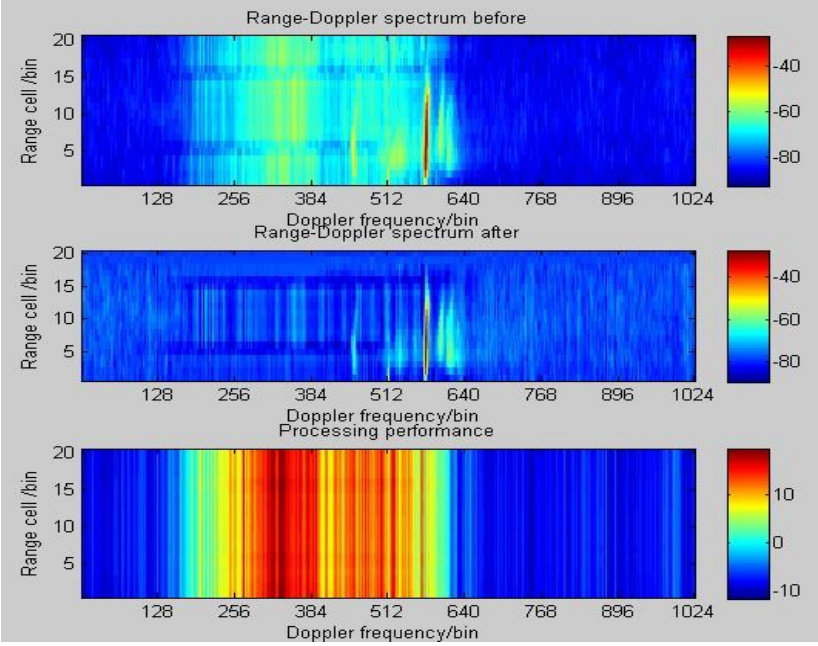


Fig. 5.17 Mitigation on clutters of the range-correlated over many Range bins shape. Top: before suppression; mid: after suppression; bottom: mitigation performance. The color scale shows the power spectrum in dB. Each Doppler bin represents 0.0048Hz. The frequency range is from -2.5Hz to 2.5Hz. Each range bin represents 15 km. The clutter covers the whole Range axis and the second order region of the sea echoes in the Doppler axis.

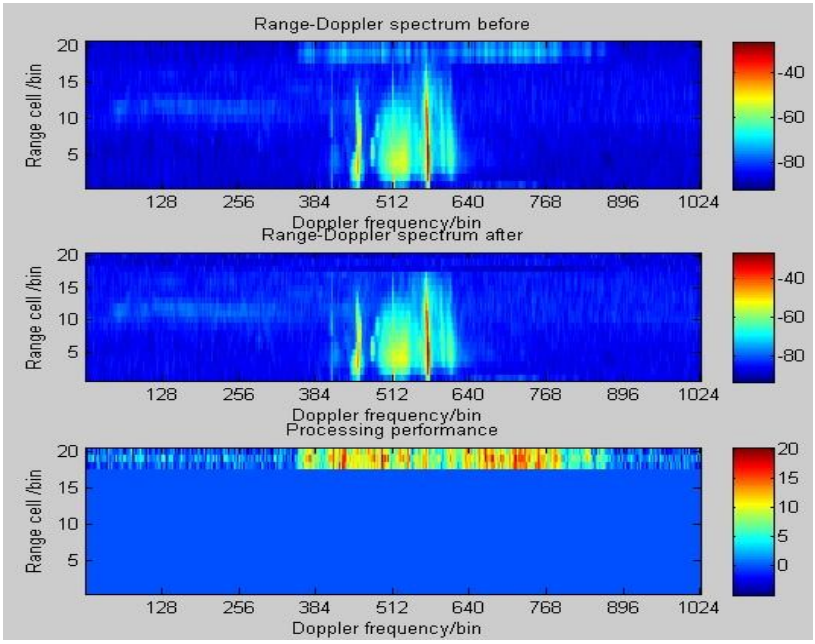


Fig. 5.18 Mitigation on clutters of the range-correlated over short Range bins shape. Top: before suppression; mid: after suppression; bottom: mitigation performance. The color scale shows the power spectrum in dB. Each Doppler bin represents 0.0048Hz. The frequency range is from -2.5Hz to 2.5Hz. Each range bin represents 15 km. The clutters cover the rb 17-20 and db 350-890.

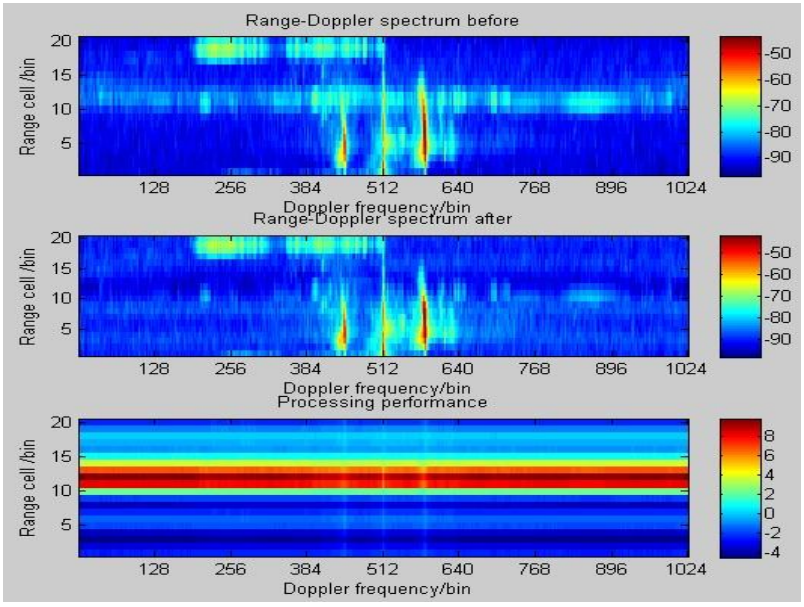


Fig. 5.19 Mitigation on clutters of the Doppler-correlated over long Doppler bins shape. Top: before suppression; mid: after suppression; bottom: mitigation performance. The color scale shows the power spectrum in dB. Each Doppler bin represents 0.0048Hz. The frequency range is from -2.5Hz to 2.5Hz. Each range bin represents 15 km. The clutter covers rb 11-14 and db 1-1024.

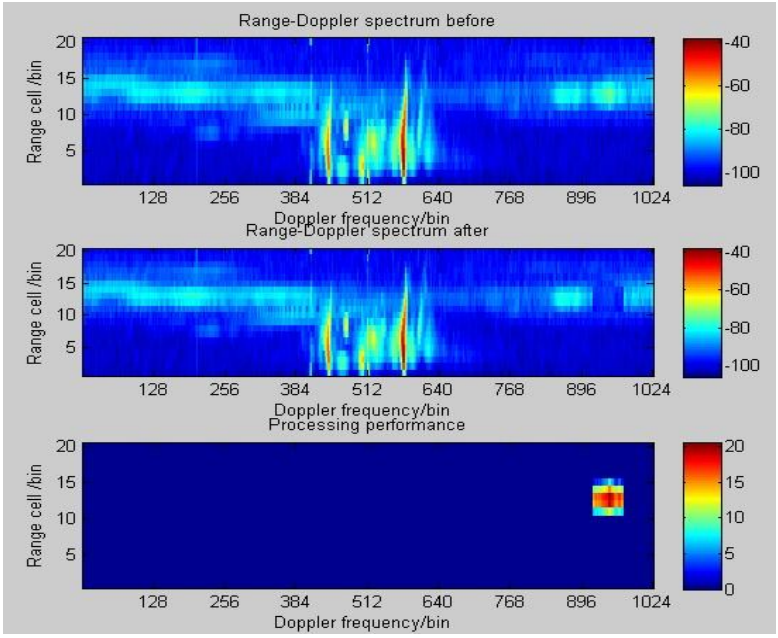


Fig. 5.20 Mitigation on clutters of the Doppler-correlated over short Doppler bins shape. Top: before suppression; mid: after suppression; bottom: mitigation performance. The color scale shows the power spectrum in dB. Each Doppler bin represents 0.0048Hz. The frequency range is from -2.5Hz to 2.5Hz. Each range bin represents 15 km. The clutter covers rb 11-15 and db 916-971.

In conclusion, this visible clutter mitigation scheme is capable of mitigating visible clutters of arbitrary shape, plus a few invisible cases if the clutter is partly hidden inside sea echoes and the part out of the sea echoes can be selected as training data. However, the visible mitigation scheme has limited improvement for the wave estimation if the clutter is totally masked in the sea echoes. For this case, invisible clutter mitigation is required.

5.3 Invisible clutter mitigation

The goal of invisible clutter mitigation is to minimize those invisible clutters, totally masked inside the second order region of sea echoes. In this section, algorithms that are developed and proposed for this goal are mainly in the category of blind source separation (BSS) (AbedMeraim et al. 1997). The objective of BSS is to estimate and separate various sources in the mixture signal even if they are not completely mutually statistically independent. In BSS, the various sources are unmixed by second-order statistics, if sources have non-vanishing second-order correlations. The BSS approach is often used to exploit the time-frequency diversity of signals. Here, it is used to exploit the Range–Doppler diversity of signals. By diversity, I mean different characteristics and features of the signal.

Subsection 5.3.1 formulates the invisible clutter separation problem; subsection 5.3.2, 5.3.3, and 5.3.4 demonstrates the application of eigenvalue decomposition (ED), generalized eigenvalue decomposition (GED), and sliding eigenvalue decomposition (SED) for invisible clutter mitigation respectively. Finally, the conclusion is given in subsection 5.3.5.

5.3.1 Formulation of the invisible clutter mitigation problem

Since the invisible clutters are mixed with sea echoes in both range and Doppler domains, the BSS problem can also be solved in these two domains separately. Let $\underline{x}(k)$ be a m - element vector. If k represents an index in Doppler, $\underline{x}(k)$ will be a range profile. If k represents an index in range, $\underline{x}(k)$ will be a Doppler profile. For both cases, $\underline{x}(k)$ is a mixture of signals from different sources. In this application, I consider sea echo as one source and all the other sources as invisible clutters. It is not necessary to separate all the invisible clutters, because the goal is to mitigate them. The invisible clutter mitigation problem is addressed by extracting one source from the mixture of two sources based on their different second order stationarity. Note that, 'second order stationarity' is a term usually used to describe whether the second order statistics of a time series is time-varying or not. In this section, the concept 'time series' is generalized to a range profile or a Doppler profile.

5.3.2 Eigenvalue Decomposition (ED)

The first class of approaches for the invisible clutter separation problem can extract stationary range profiles of sea echoes from nonstationary range profiles of clutters. Range profiles of sea echoes in the first- and second- order region are assumed to be stationary in the sense that their variances are not range-varying, particularly at the superior first order peak. However, those of the clutters are not. Clutters usually appear in one or a few range bins, because each range bin represents a distance of 15 km. Not many clutters have such large-scale distribution in the DS as to spread over many range bins (RFI has this feature, but is not consistent with the definition of invisible clutters). A method is derived based on the eigendecomposition of the covariance matrix of the range profile of the superior first order peak. The mixture signal is then projected into this eigenspace. Signals in the orthogonal subspace are thus obtained and subtracted from the mixture. Results shown in subsection 5.3.6 indicate that the method can improve the quality of DS if the radar estimates of waveheight are larger than the buoy, i.e. an overestimation case.

Let $X_A = [X(r_1, sfs), X(r_1+1, sfs), \ldots, X(r_2, sfs)]^T$ be an N_r-element range profile from r1 to r2 of the Doppler bin sfs, i.e. the training area A. The covariance matrix of the Range profile X_A is estimated by: $\hat{R}_A = E[X_A X_A{}^T]$. Let $X_B = [\underline{x}(d_3), \underline{x}(d_3+1), \ldots \underline{x}(d_4)], X_B \in R^{N_r \times M_d}$ be the processing matrix, containing N_r-element range profile of the Doppler bin d3 to d4. This area is selected from the second order continuum, essential for wave estimation. The covariance matrix of X_B is estimated by: $\hat{R}_B = E[X_B X_B{}^T]$. After the eigenvalue decomposition of $\hat{R}_A = V \Lambda V^T$, the stationary signal subspace is the eigenspace spanned by the biggest eigenvector $\underline{v}_1$, as $P_{eig} = \underline{v}_1 \underline{v}_1^T$. Then the projected sea signal space is obtained by $P_{eig} X_B$ with the clutter space $X_c = (I - P_{eig}) X_B$. Then the clutter-removed sea signal is obtained by: $X_{sea} = X_B - X_c$.

5.3.3 Generalized eigenvalue decomposition (GED)

The second class of approaches for the invisible clutter separation problem can extract nonstationary Doppler profile of sea echoes from their stationary clutter

counterparts. The training area is selected from the second order continuum inside *sfs*. The processing area is selected from the region including the *sfs* and the outside second order part (this processing region is a key area for waveheight estimation). The signal model of the two areas is described by: $X_A = X_{A\_sea} + X_{A\_c} + X_n$, $X_B = X_{B\_sea} + X_{B\_c} + X_n$, where X_{A_sea} and X_{A_c} refer to the sea echoes and clutters in the training data A respectively; X_{B_sea} and X_{B_c} are the counterparts from the processing B. In fact, the Doppler amplitudes of the first order region are much higher than these of the second order continuum. The variance of the Doppler file of sea in A is much lower than that in B. In contrast, the variance of the Doppler profile of invisible clutters that are hidden completely in the first and second order regions is not as strongly Doppler-varying as sea echoes, and can be assumed to be stationary. I derive a method based on the generalized eigendecomposition of two covariance matrices of the mixture in training area A and processing area B. Results shown in section 5.3.6 indicate that this method can improve the quality of Doppler spectra if radar estimates of waveheight are smaller than the buoy, i.e. an underestimation case.

Let $X_A = [\underline{x}(r_1), \underline{x}(r_1+1), \ldots \underline{x}(r_2)]^T, X_A \in R^{M_d \times N_r}$ be an M_d -element Doppler profile from r1 to r2, including the second order continuum inside *sfs*. Let $X_B = [\underline{x}(r_3), \underline{x}(r_3+1), \ldots \underline{x}(r_4)]^T, X_B \in R^{M_d \times N_r}$ be an M_d -element Doppler profile from r3 to r4, including the superior first order peak and the outside second order continuum. As said above, $X_A = X_{A\_sea} + X_{A\_c} + X_n$, $X_B = X_{B\_sea} + X_{B\_c} + X_n$. A visual representation of the concept of the parameters d1, d2, d3, d4, r1, r2, r3, r4 is shown in Fig. 5.12.

The Generalized Eigenvalue Decomposition (GED) method illustrated below is based on the assumption that the power of X_{sea} is σ_1^2 in the training area A and σ_2^2 in the processing area B. On the contrary, the power of X_c remains the same, σ_c^2 , in both areas. Therefore, the covariance matrix of X_A and X_B can be estimated as:

$$\hat{R}_A = E[X_A X_A{}^T] = \sigma_1^2 + \sigma_c^2 \qquad (5.22)$$

$$\hat{R}_B = E[X_B X_B{}^T] = \sigma_2^2 + \sigma_c^2 \tag{5.23}$$

where $\sigma_1^2 = E[X_{A\_sea} X_{A\_sea}{}^T]$, $\sigma_2^2 = E[X_{B\_sea} X_{B\_sea}{}^T]$, and $\sigma_c^2 = E[X_c X_c{}^T]$ This kind of second order non-stationarity happens at the boundary of the second order continuum and first order area, because the Bragg returns are normally much higher than the second order backscattering , i.e. $\sigma_1^2 < \sigma_2^2$, except for some extreme high sea states, i.e. $\sigma_1^2 \approx \sigma_2^2$.

The goal of the GED algorithm is to find an unmixing eigenvector $\underline{v}$ that can remove the stationarity clutters in the processing data, i.e. the stationary component is minimized. The GED algorithm is implemented by computing covariance matrices $\hat{R}_A$ and $\hat{R}_B$ first. Generalized eigenvalue decomposition of the two symmetric matrices determines the unmixing eigenvector for the optimal separation of superimposed non-stationary and stationary sources. The computation of the eigenvector associated with the minimal eigenvalue of the matrix pencil ($\hat{R}_B$, $\hat{R}_A$) is equivalent to the optimization:

$$\min_{\underline{v}} \frac{\underline{v}^T \hat{R}_A \underline{v}}{\underline{v}^T \hat{R}_B \underline{v}} \tag{5.24}$$

where $\underline{v}$ is the unmixing coefficient vector, minimizing the stationarity clutter in area A and B; The size of the two windows A and B, i.e. N_r and M_d, are free parameters that can be adjusted to fit with the kind of non-stationarity of sea echoes to be detected.

Algorithm: GED

1. select training matrix X_A and processing matrix X_B
2. estimate two covariance matrices of X_A and X_B by $\hat{R}_A = \frac{1}{N_r} X_A X_A{}^T$ and $\hat{R}_B = \frac{1}{N_r} X_B X_B{}^T$
3. check the minimum eigenvalue of matrix pencil $(\hat{R}_B, \hat{R}_A)$ (the definition is given in Appendix A) , i.e. $\hat{R}_A \underline{v} = \hat{R}_B \underline{v} \lambda$, to determine corresponding

eigenvector $\underline{v}$

4. estimate the unmixing eigenspace matrix $P_{eig} = \underline{v}\underline{v}^T$

5. extract sea echoes by transforming X_B by the unmixing matrix

$$X_{sea} = P_{eig} X_B$$

5.3.4 Sliding eigenvalue decomposition (SED)

Similar to the unsupervised dynamic processing for visible clutter mitigation, I can also address the invisible clutter mitigation problem by sliding window processing in the Doppler domain. Note that the sliding window in the range domain is not discussed because the minimum window width is 3 range bins while the invisible clutter is locally distributed, in one or few range bins (otherwise it would be visible outside the 1st and 2nd order regions).

Each sliding step performs a principal component extraction by eigenvalue decomposition. If the coverage of the sliding window is the first- and second-order region, the principal components will be the sea echoes, and the signal model will be $X = X_{sea} + X_c + X_n, X_{sea} > X_c$. Based on this sliding eigenvalue decomposition of the covariance matrices of the mixture in processing area B, SED performs the invisible clutter mitigation. Results shown in section 5.3.5 imply that this method improves the quality of DS, particularly for the wave direction.

Let the processing matrix $X_B = [\underline{x}(d3), \underline{x}(d3+1), \ldots, \underline{x}(d4)]$, $X_B \in R^{N_r \times M_d}$. Each column of X_B is a range profile of the Doppler bin from d3 to d4. At the ith siding step, the window matrix $X_i = [\underline{x}(d3+i-1), \underline{x}(d3+i), \underline{x}(d3+i+1)]$, $1 \le i \le sn$, is constructed, where sn is the number of sliding steps, and the window length is set to 3 (the minimal value for eigendecomposition). The same three important parameters defining the sliding window are sliding length l ($l = d4 - d3 + 1$), step length sl (sl=3), window length wl (wl=3). The goal of SED is to extract the eigenspace belonging to the sea echoes in each window i. The procedures of the SED algorithm are as follows.

Algorithm: SED

1. Form $X_i = \begin{pmatrix} X(r_3, d3+i-1) & X(r_3, d3+i) & X(r_3, d3+i+1) \\ \vdots & \ddots & \vdots \\ X(r_4, d3+i-1) & X(r_4, d3+i) & X(r_4, d3+i+1) \end{pmatrix}_{N_r \times 3}$. The covariance matrix $\hat{R}_i$ is estimated by $\hat{R}_i = E[(X_i - E(X_i)^T (X_i - E(X_i)]$.

2. This Hermitian matrix can be eigen-decomposed as $\hat{R}_i = \sum_{j=i}^{3} \lambda_j \underline{v}_j \underline{v}_j^T$, where $\lambda_j (j = 1,2,3)$ is the eigenvalue in descending order and $\underline{v}_j$ is the corresponding eigenvector.
3. The means of the range profile of each of the three Doppler bins are calculated as $\mu_{d3+i-1} = E[\underline{x}(d3+i-1)]$, $\mu_{d3+i} = E[\underline{x}(d3+i)]$, $\mu_{d3+i+1} = E[\underline{x}(d3+i+1)]$. The base signal X_{base} is the matrix form of the means of X_i.

$$X_{base} = \begin{pmatrix} \mu_{d3+i-1} & \mu_{d3+i} & \mu_{d3+i+1} \\ \vdots & \ddots & \vdots \\ \mu_{d3+i-1} & \mu_{d3+i} & \mu_{d3+i+1} \end{pmatrix}_{N_r \times 3} \tag{5.25}$$

4. The zero-mean $\tilde{X}_i$ is obtained by subtracting X_{base} from X_i as $\tilde{X}_i = X_i - X_{base}$.

5. Construct the dominant eigenspace by spanning the eigenvector $\underline{v}_1$: $P_{eig} = \underline{v}_1 \underline{v}_1^T$.

6. Project $\tilde{X}_i$ onto the dominant subspace P_{eig} to generate sea echoes $\tilde{X}_{i\_sea} = P_{eig} \tilde{X}_i$

7. Reconstruct X_{i_sea} by adding the mean, i.e. X_{base}: $X_{i\_sea} = \tilde{X}_{i\_sea} + X_{base}$

5.3.5 Experiments and validation

Figures of RD images used for validation of visible clutter mitigation scheme are not sufficient to display the mitigation performance of invisible clutter mitigation

schemes. The only way to validate the performance is to check the derived wave measurements. The ocean wave frequency spectrum, the direction spectrum (mean direction at each frequency) and derived waveheight, period and direction parameters are all used here and are good indicators of the performance of ED, GED, and SED. In a frequency spectrum, there are often two major components: wind sea and swell. Wind seas are generated by the local wind and swell by winds far from the measurement position that generate waves that propagate into the region. In this location (the Celtic Sea), wind seas have short wavelength and occupy the frequency band (0.1-0.3 Hz), while swell is waves with long wavelength and occupies the low frequency band (<0.1Hz). It will be demonstrated that the invisible schemes can improve the frequency spectrum estimation for these two components.

In this comparison, the dual radar data is used to validate the performance of the invisible clutter mitigation schemes for improving the ocean wave directional estimation, because the estimated $S(\vec{k})$ is inverted using the DS from both sites.

Since it is hard to determine which data contains invisible clutters, the three invisible schemes are applied to the whole dataset in February (the setting of the fixed parameters is explained in the examples below), to see the impact of these schemes on the estimated wave directional spectrum.

Comparing the new frequency spectra with the old ones, it has been found that for the data for which the significant waveheight is underestimated compared with the buoy, both GED and SED are useful to improve the accuracy. If the original radar waveheight was overestimated, only ED shows improvement of the accuracy but not for extreme cases (in those extreme cases, clutters are strong and large enough to be visible so it is not appropriate to use the invisible scheme). If the original data were good, for GED and ED, it has been found that the data may be made worse, better or be unaffected. This is because the invisible schemes are designed with the assumption that there is invisible clutter in the sea echo, i.e. all the data, either good or bad, are treated as problematic data. But for SED, it has been found that the good quality of the original data is preserved. Specifically it has been found that the ED scheme is often capable of reducing the swell component in a frequency spectrum if the swell is overestimated; the GED scheme is mainly capable of increasing the wind sea if it is underestimated, sometimes increasing the swell as well; SED is not error-oriented (i.e. the nature of the error in the data didn't impact the outcome of the algorithm, except that it doesn't work as well when there is an overestimation), and sometimes increases both components if they are underestimated.

Besides the waveheight, the modification of mean direction and mean period are also checked. ED and GED both improve the period parameter and worsen the direction. GED has better performance than ED by inverting more valid data i.e. 347>287. SED has shown the best result among the three schemes, by inverting the most valid data and the best improvement for mean direction and peak

direction.

These findings above are illustrated in Table 5.4 in section 5.3.6. The two schemes (ED and GED) can work better on the data with a priori knowledge that there is a particular type of error in the frequency spectrum. What's more, they should be applied to the DS which is responsible for the error. To find the error in the frequency spectrum is easy, since the buoy data is available, but to find the problematic DS is difficult. The current practice is to test recursively according to the procedure 2-5 below. One last thing to mention is that the testing starts from the DS for which the single radar waveheight indicates a problem. Since waveheight is the integral of the frequency spectrum, the error in a single radar waveheight is not able to reveal whether the problem is in the swell or wind sea but it is a good place to start.

Since SED is not error-oriented, its procedure is simpler than ED and GED. Examples of the SED scheme will be presented at the end of this section. For ED and GED scheme, the following procedure is adopted in each example:

1. Frequency spectra estimated by HF radar and the buoy are checked, in order to find whether there is an error in one of the three error types: wind sea error, swell error, wind sea plus swell error.
2. The problematic DS is selected from 'cm' and 'np' by comparing the buoy and original single radar estimates of waveheight to select the problematic DS initially. (If this DS is not the right one, at the next iteration try the DS from the other site.)
3. Select an appropriate scheme from ED, GED according to the error type.
4. For the selected scheme, appropriate values are chosen for the key parameters, i.e. $d_1, d_2, d_3, d_4, r_1, r_2, r_3, r_4$ (see the definition in Fig.5.12).
5. Show the impact of this scheme on the resulting frequency spectrum and direction spectrum. If the swell-peak is not reduced or the wind-sea is not increased to get closer to the buoy, change the DS and go to step 3.

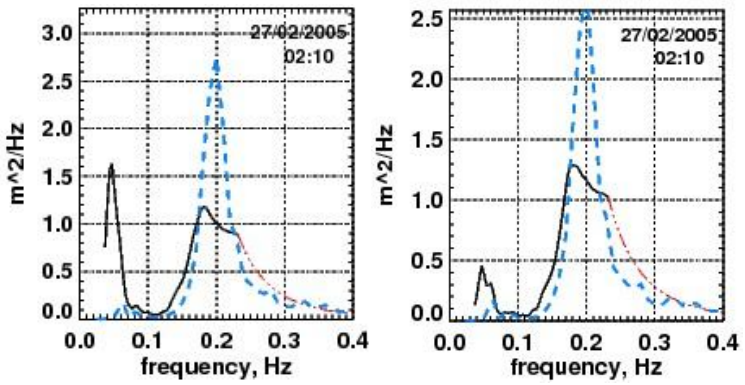
An example to show that the ED scheme can reduce the swell is chosen from the DS of rb 5 obtained at 02:10am, 27/02/2005, because in the frequency spectrum shown in Fig.5.21(a), the radar estimation indicates a wind sea with a peak at about 0.18Hz, and a swell with a peak at about 0.05Hz, while the buoy estimation shows mainly a wind sea. The problematic DS is firstly selected from 'cm', because the estimates of waveheight are 1.59m (buoy), 1.34m ('cm'), and 1.50m ('np') and 'cm' has larger error. Then, ED is applied to the 'cm' data, with the processed Doppler spectra shown in Fig.5.22. The scheme is applied to 'np' also, but the error between the buoy and radar estimate is not minimized, which means the error source is 'cm'. For ED, the parameters required are d_3, d_4, r_3, r_4

to define the processing area B, i.e. db $d_3 - d_4$ and rb $r_3 - r_4$. According to the theory, area B should include the first and second order region beside the superior first order peak sfs within detectable range bins ($R_{\max} = 150\,\text{km}$, so rb 10 is the maximum). The requirement for parameter setting is given by:

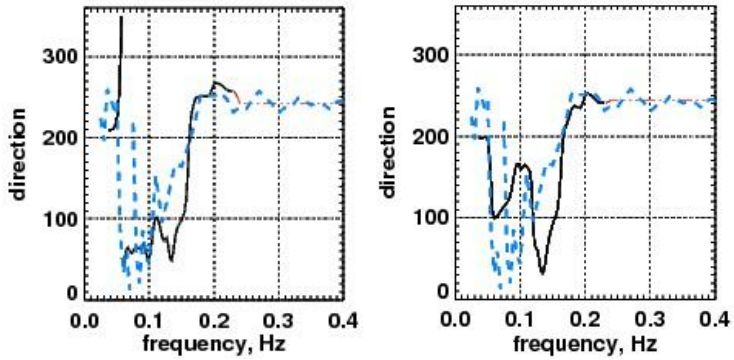
$$3 \le r_3 < r_4 \le 10, \quad 512 < d_3 < sfs < d_4 < 1024 \tag{5.26}$$

The setting of Doppler parameters of ED is determined by the position of sfs. As sfs is at db 585, after testing different sections of $d_3 - d_4$, the best section to minimize the swell error is 560-600, i.e. the adjacent second order continuum beside sfs. The reason to select rb 3 as the lower limit of range is because the second order continuum is low in the DS from rb 1 and 2. A further testing of range parameters is performed and shown in Fig.5.22. The modified DS are generated by the setting of rb 3-8 in (a) and rb 3-10 in (b) (if II appears around rb 9-10, rb 3-8 will be selected). Finally, d_3, d_4, r_3, r_4 are set to 560, 600, 3, 10. The frequency and direction spectrum modified by ED are shown along with the old ones in Fig.5.21.

It is seen from Fig.5.21 that the overestimated swell is reduced from 1.6 to 0.5 m^2 / Hz, while the wind sea is unchanged. As seen in Table 5.2, although the hs is not improved (because although the swell is overestimated, the total waveheight is underestimated), the other three parameters are all improved, especially the peak mean period T_p and peak wave direction D_p. When the parameters are changed from rb 3-8 to rb 3-10, this effect of reducing the swell and getting closer to the buoy data is stronger.

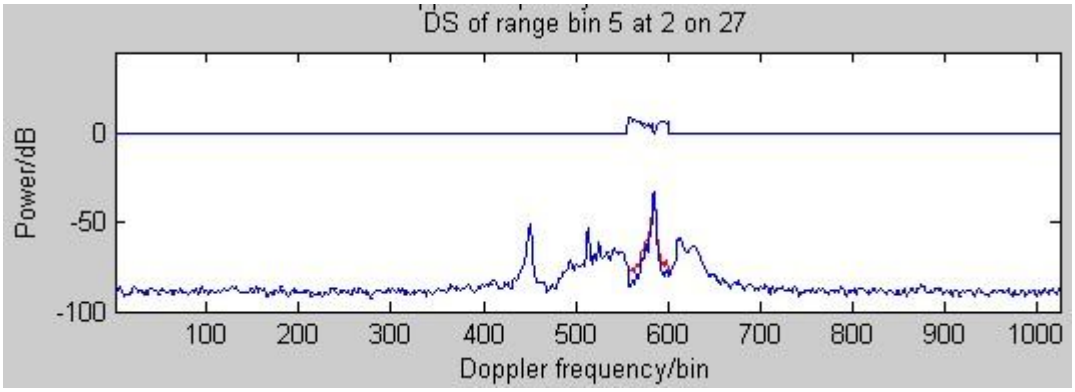


(a) (b)

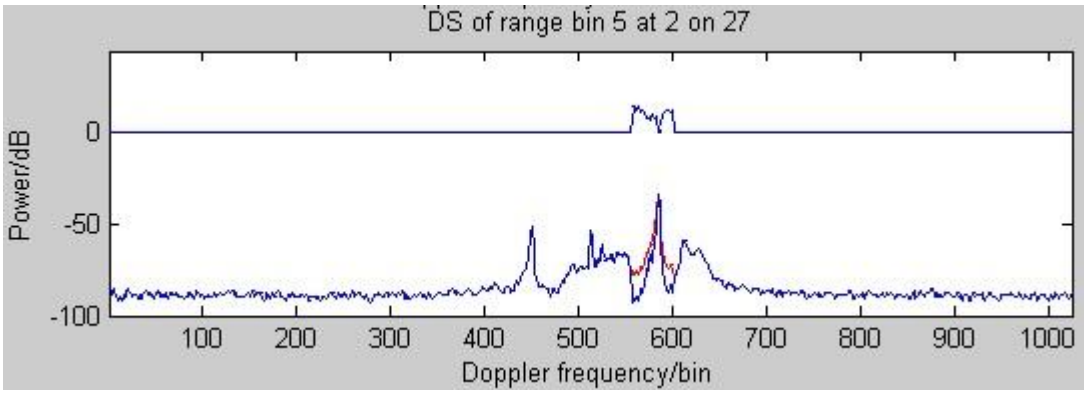


(c) (d)

Fig.5.21. Comparisons of frequency spectrum and direction spectrum before and after ED. (a) Frequency spectrum: old; (b) Frequency spectrum: after ED; (c) Direction spectrum: old; (d) Direction spectrum: after ED (solid is the radar and dashed is the buoy). The modified radar estimates are derived based on new 'cm' DS and old 'np' DS, obtained at 02:10, 27/02/2005.



(a) $r_3 = 3; r_4 = 8$



(b) $r_3 = 3; r_4 = 10$

Fig.5.22 DS from rb 5 of 'cm' RD image at 02:10, 27/02/2005 after ED. The original DS and processed DS are plotted in red and blue respectively. The error is plotted by the top blue line.

Table 5.2 Four parameters estimated by buoy, original DS, and DS after ED

Source	h_s /m	T_1 /s	T_p /s	D_p /°
Buoy	**1.59**	**4.14**	**5.21**	**73.12**
Original DS	1.38	6.19	21.12	57.66
DS after ED (rb 3-8)	1.30	5.83	5.35	66.48
DS after ED (rb3-10)	1.29	5.52	5.57	72.46

An example to show the performance of GED in increasing wind sea uses the DS from 04:10, 01/02/2005. Fig. 5.23(a) shows that the original radar estimation of the wind sea is lower than the buoy estimation. The waveheight estimated by buoy, single 'cm' and single 'np' are 1.39m, 0.95m, and 1.27m. Hence, 'cm' is tried first as the problematic data. For GED, both the training area A defined by

d_1, d_2, r_1, r_2, and the processing area B defined by d_3, d_4, r_3, r_4 are required. Area A is chosen from the second order continuum inside *sfs*. Area B is chosen from the outside second order continuum of *sfs*, including the superior first order region. Note that, the requirement for setting the parameters is given as:

$$r_1 = r_3, r_2 = r_4, d_4 - d_3 = d_2 - d_1, \quad 3 \le r_1, r_2, r_3, r_4 \le 10, \quad d_1 < d_2 = d_3 < sfs < d_4 \quad (5.27)$$

The setting of Doppler parameters of GED also depends on the position of *sfs*. Actually, the values of d_1, d_2 are determined by d_3, d_4. Here $sfs = 585$, the best d_3, d_4 have been found to be 577 and 597, so d_1, d_2 are set as 557, 577. The modified and original Doppler spectra are shown in Fig.5.24. It is found that the amplitude of *sfs* is reduced more if rb 3-10 is chosen rather than rb 3-8. The derived new frequency spectra using rb 3-8 and rb 3-10 are shown in Fig.5.23(b)(c) respectively, where it is seen that the underestimated wind sea is increased in both. GED using setting rb 3-10 has stronger impact on increasing this wind sea amplitude.

According to Table 5.3, the waveheight, peak period, and peak direction are improved. The mean period of 'rb 3-10' is worse than 'rb 3-8' because of the increased swell, while the waveheight of 'rb 3-10' is better than 'rb 3-8' by increasing the wind sea more.

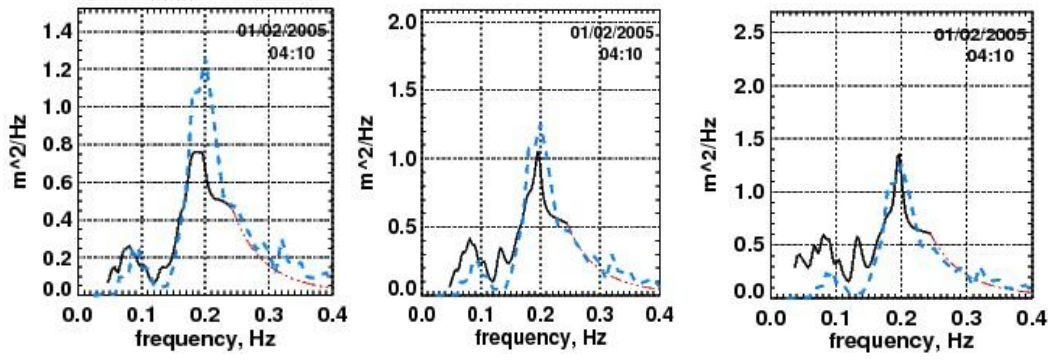
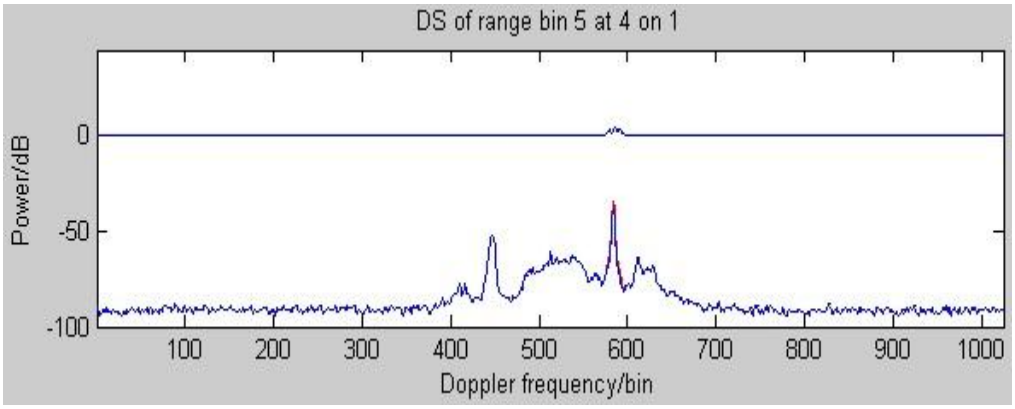
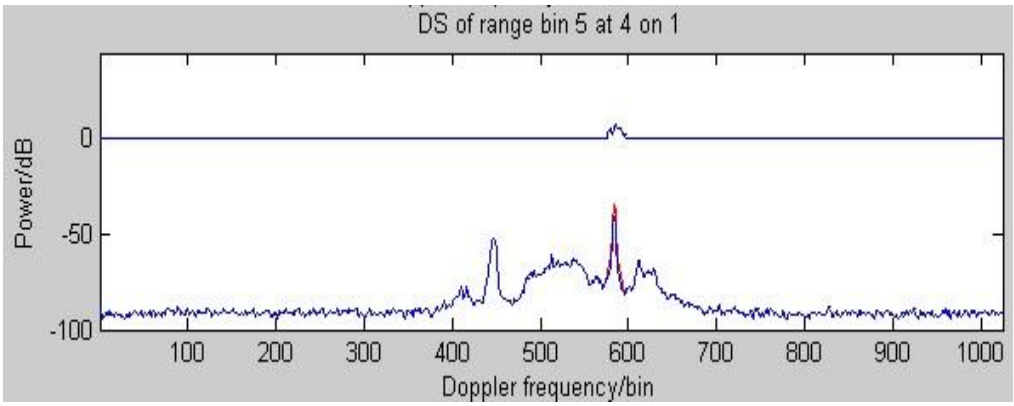


Fig.5.23. Comparisons of frequency spectrum before and after GED. Frequency spectrum inversed by new 'cm' and old 'np' at 02:10, 27/02/2005, the solid is the radar and the dashed is the buoy. (a) Frequency spectrum: old; (b) Frequency spectrum: rb 3-8; (c) Frequency spectrum: rb 3-10



(a) rb 3-8



(b) rb 3-10

Fig.5.24 DS from rb 5 of 'cm' RD image at 04:10, 01/02/2005 after GED. The original DS and processed DS are plotted in red and blue respectively. The error is plotted by the blue line.

Table 5.3 Four parameters estimated by buoy, original DS, and DS after GED

Source	h_s /m	T_1 /s	T_p /s	D_p /°
Buoy	**1.39**	**3.85**	**4.86**	**326.25**
Original DS	1.04	5.58	5.35	297.78
DS after GED (rb 3-8)	1.17	5.45	5.08	317.15
DS after GED (rb 3-10)	1.34	5.96	5.08	320.34

The two typical examples above show that the ED and GED schemes produce a decrease in the swell and an increase in the wind-sea in the frequency spectrum respectively. A question to answer is whether these two effects can be added by applying ED and GED one by one. After testing a number of data, the joint application of these two schemes is found to provide a combination of these two

effects regardless of the order. Examples of the joint mitigation performance are shown in Fig. 5.25. Note that, the final frequency spectrum shown on the third plot of each row is based on dual DS, either new ‘cm’ old ‘np’ (cnno), or old ‘cm’ new ‘np’ (conn), or new ‘cm’ new ‘np’ (cnnn). For example, if ‘cm’ causes the overestimated swell and ‘np’ causes the underestimated wind sea, ED will be applied to ‘cm’ and GED will be applied to ‘np’.

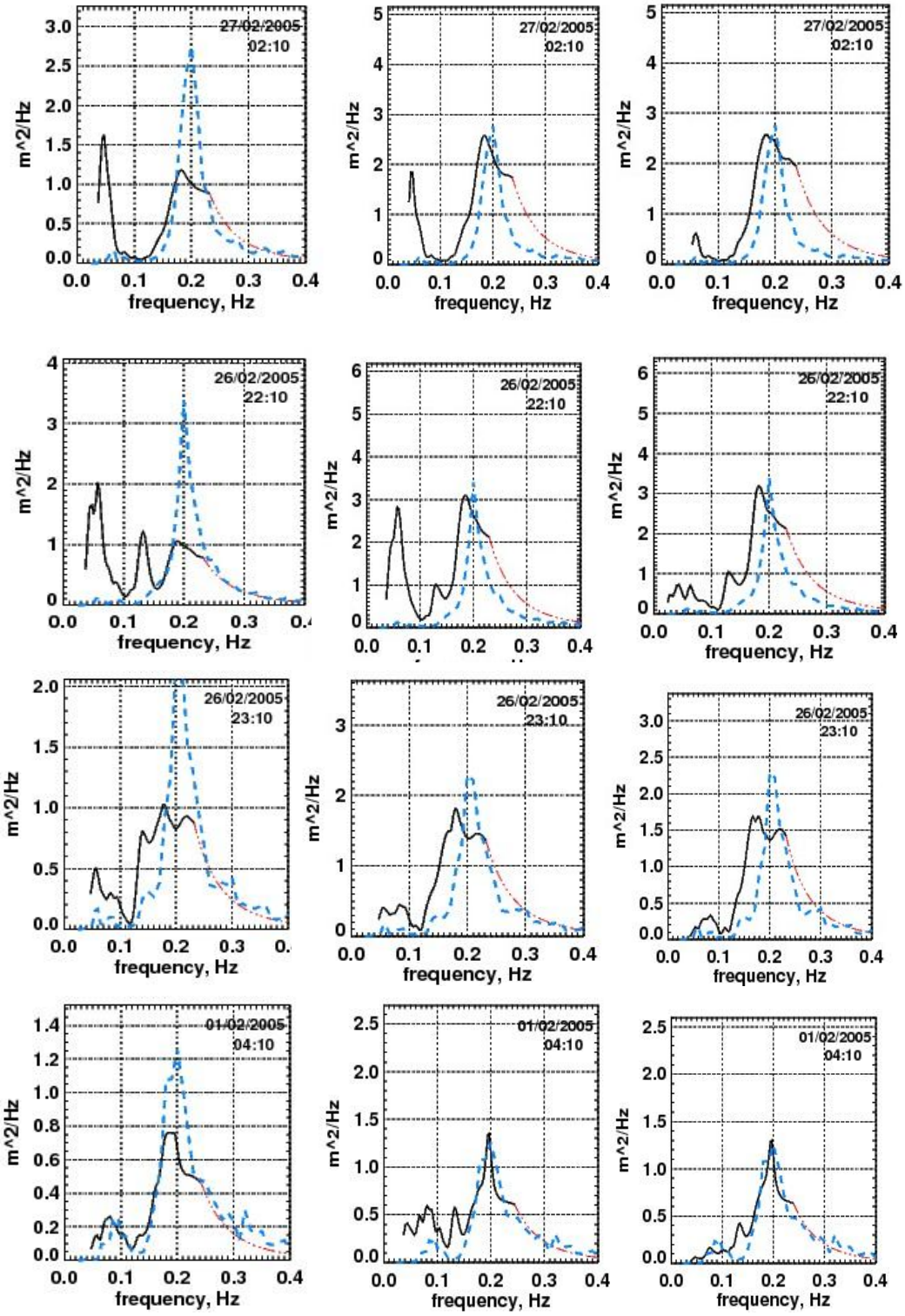
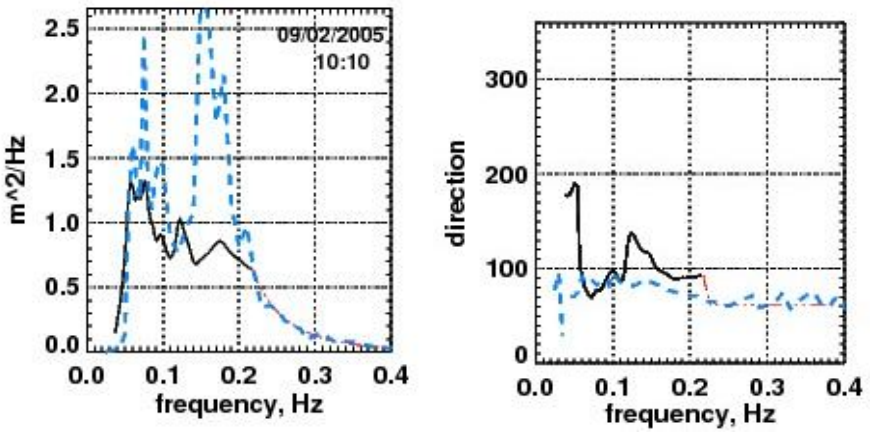


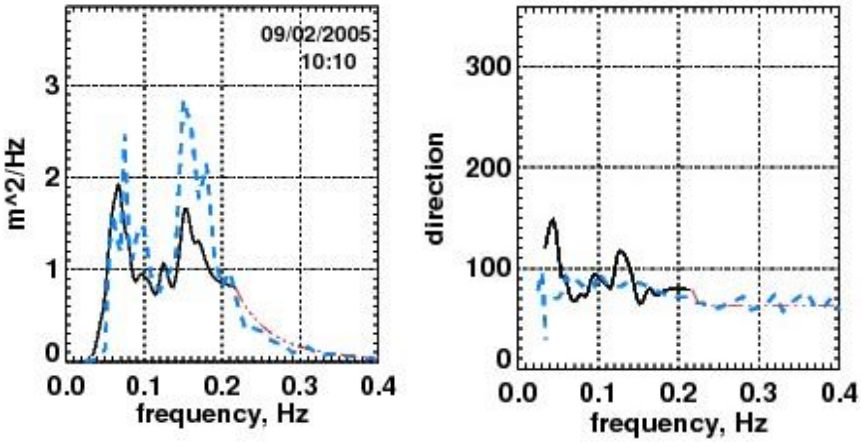
Fig.5.25. Comparisons of frequency spectrum before and after ED and GED Frequency spectra: radar (the solid) and buoy (the dashed). Left-before mitigation; Middle-after mitigation by GED; Right-after mitigation by ED.

SED is a sliding window processing that assumes that the dominant signals in each window are extracted. Two examples of SED scheme are shown in Fig.5.26-5.27. SED are tested on the data from 10:10 and 20:10, 09/02/2005.

These two figures show an improvement in both frequency and direction spectrum of radar estimation.

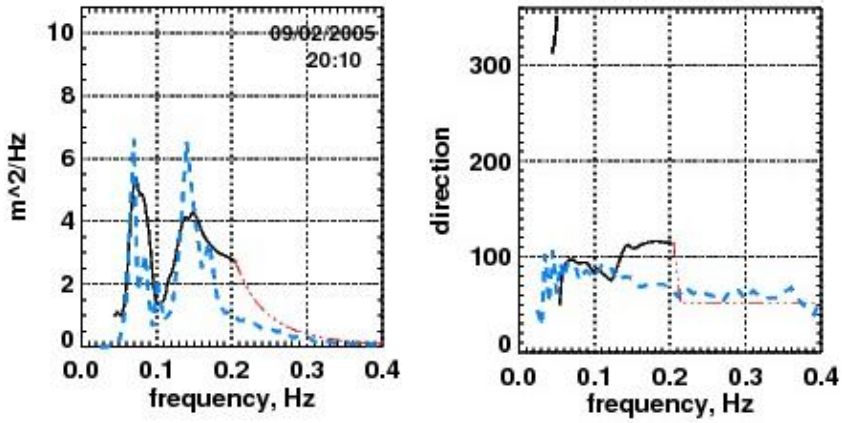


(a)

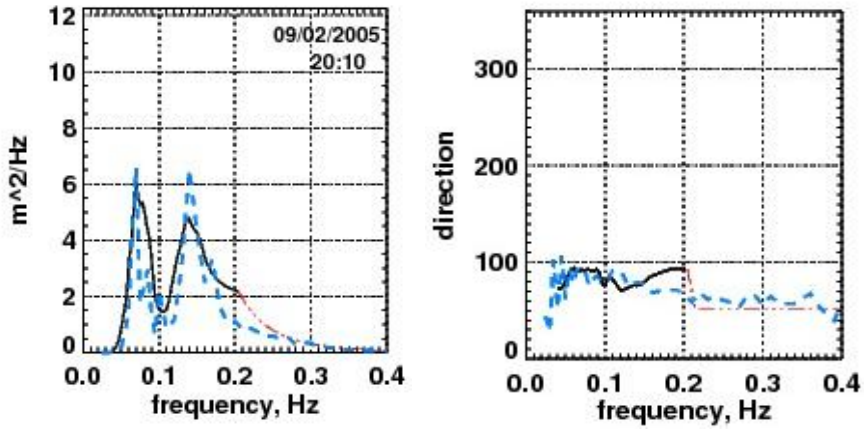


(b)

Fig.5.26. Comparisons of frequency spectrum and direction spectrum before and after SED. Frequency spectrum (left) and direction spectrum (right) at 10:10am, 09/02/2005, the solid is the radar and the dashed is the buoy. (a) Before mitigation by SED; (b) after mitigation by SED.



(a)



(b)

Fig.5.27. Comparisons of frequency spectrum and direction spectrum before and after SED. Frequency spectrum (left) and direction spectrum (right) at 20:10am, 09/02/2005, the solid is the radar and the dashed is the buoy. (a) Before mitigation by SED; (b) after mitigation by SED.

5.3.6 Operational application

So far case-by-case experiments have been shown. This section discusses the operational application of some suitable schemes developed in this chapter. There are three solutions recommended here. They are the SED scheme, the joint ED and GED scheme, and the unsupervised dynamic scheme (UDS). SED, ED, and GED are appropriate because they deal with invisible clutters in the in-band area, which in nature don't require the step of image recognition for clutter identification. The image recognition step is the main obstacle for visible clutter mitigation scheme to be applied operationally (as seen in section 5.1, the schemes are performed case by case). The UDS is also appropriate because it is both 'unsupervised' and 'dynamic', which means there is no need to find an appropriate training matrix and the construction for the processing matrix is simple and fixed.

Compared with ED and GED, SED is not error-oriented, which is found when it is applied operationally to all the data. What's more, SED won't degrade original good data, which makes it more suitable for operational use. For SED, the parameters required are d_3, d_4, r_3, r_4 , to define the processing area B (see Equ.(5.26)). The setting rule is simple, i.e. it should include the *sfs* and the second order continuum beside *sfs*. It has been checked that the *sfs* in this dataset is located in the range of rb 440-466, if it is the left peak; and rb 561-587, if it is the right peak. After testing different values of d_3, d_4, the setting

of parameters for operational use is: if $560 < sfs < 570$, $d_3 = 550, d_4 = 590$; if $570 < sfs < 590$, $d_3 = 560, d_4 = 600$; $r_3 = 3, r_4 = 8$. Note that a preprocessing is required if the sfs is on the left (the setting of parameters found above is for the case that sfs is on the right). In this case, to make use of the original setting, the sequence of this DS should be flipped. After SED, the DS will be flipped back. Experiments have shown that the performance on flipped DS is as good as on un-flipped DS. This technique is also used in the examples of ED and GED.

Another scheme which is suitable for operational use is the unsupervised dynamic scheme (UDS), i.e. a joint application of Range-window and Doppler-window (explained in section 5.2.3). The nature of these two schemes is the same as SED, i.e. principal components extraction, based on the assumption that the invisible clutters hidden inside the sea echoes are in a patch or spot shape and not as strong as sea echoes. If the window size is chosen small enough, small scale clutters can be eliminated out of the mixture during each sliding step. Thus, the window length wl, which is equal to the step length sl, is set to the minimum number 3. The difference between UDS and SED is that the former slides in both the Range and Doppler domain, while SED slides only in the Doppler domain. The parameters required for Range- and Doppler- window are $r_3(r_{begin}), r_4(r_{end}), d_3, d_4$ and $r_3, r_4, d_3(d_{begin}), d_4(d_{end})$ respectively. For operational use, the setting of them is fixed as $r_3 = 1, r_4 = 10, d_3 = 400, d_4 = 640$.

The third scheme for operational use is a joint application of ED and GED. The joint operational application is tested on all the data again, motivated by the improvement of each example in Fig.5.25. The setting of GED parameters for operational use is : if $560 < sfs \le 575$, d1=530, d2=560, d3=560, d4=590, r1=3, r2=8, r3=8, r4=8; if $575 < sfs < 590$, d1=540, d2=570, d3=570, d4=600, r1=3, r2=8, r3=8, r4=8. The setting for ED is the same as SED.

The performances of the three operational applications, together with the operational test explained at the beginning of section 5.3.5, are assessed by the statistics of significant waveheight, peak waveheight, mean period, peak period, mean direction, peak direction, comparisons summarized in Table 5.4. Note that, at the moment, I haven't tested the results for statistical significance and the conclusions are based on qualitative judgements. Notation 'cono', 'cnno', and 'cnnn' refer to dual sites operations of 'old cm old np', 'new cm old np', 'new cm and new np' respectively.

Before the discussion of successful applications, some unsuccessful attempts are worthy of consideration. When ED (EDcnno), GED (GEDcnno), and joint application of them (EDGEDcnno) are applied to 'cm' data in a whole month of Feb., the statistics of waveheight and direction are both worsened while period is not affected. The same result is found when they are applied to 'np' dataset and even worse result is found when they are applied to 'cm' and 'np' simultaneously. The reason is that the performance of ED and GED relies on knowledge of the error type in the frequency spectrum. They cannot be applied blindly to all data. A direct solution to identify the error type is not yet found, but an indirect method is established to predict whether the DS will result in an under, proper, or over estimation of the waveheight (this method will be investigated in chapter 6 at length). On the predicted DS, the ED and GED can work in a selective way as follows. If the DS is in the underestimation class, GED scheme will be applied since it can increase the waveheight by increasing wind seas (the power of wind seas is more dominant than swell in the frequency spectrum in most cases). If the DS is in the overestimation class, ED scheme could be used because it can reduce the amplitude of swell, which could to some extent reduce the waveheight. If the DS is in the proper class, no scheme is applied in order to preserve its good quality.

To validate the operational performance of ED and GED, DS in the whole Feb. are classified into under, proper and over estimation groups using buoy data. Then the ED and GED schemes are applied in a selective way explained above (EDGEDcnnnwbi). They exhibit much better operational performance than in the blind way (EDGEDcnno) and are also better than the other two recommended schemes in terms of improving waveheight estimation. The correlation coefficient (C.C.) of hs and peak hs is improved from 0.942 to 0.964 and from 0.871 to 0.898 respectively. The standard deviation (STD) of hs and peak hs is decreased from 0.37m to 0.28m and from 0.293m to 0.258m respectively. Wyatt (2006) has compared the Pisces radar measurements of the directional wave spectrum and derived oceanographic parameters with the estimations from a waverider buoy and estimations from the Met Office, U.K., operational wave model. During Jan. and Feb. in 2004, hs can be measured with useful accuracy (e.g. the C.C. and the STD with the buoy of 0.94 and 0.227m compared to the model values of 0.9 and 0.363m. The C.C. and the STD of peak hs with buoy is 0.88 and 0.341m respectively). It can be seen that the hs and peak hs estimated by the original Pisces data in 2005 are of similar accuracy to 2004. After clutter mitigation, radar data provide better accuracy for the significant waveheight. The number of valid available data is increased from 312 to 335. A disadvantage of this scheme is that it makes the estimate of direction worse.

Compared with significant waveheight, mean and peak period are much more sensitive to clutters in the radar backscatter and so are less accurate (in the same comparison using data in Jan. and Feb. in 2004, the C.C. of period between the

buoy and the radar is 0.59, much lower than that of the waveheight). Usually the period measurements are more accurate when waveheight is above 2 m. In low seas (< 2 m), spurious components in the directional spectrum may contaminate the estimation of this parameter. The best scheme among all to improve the statistics for period is UDS (UDScnnn) scheme. The C.C. of mean period T_1 is increased from 0.495 to 0.633. Both mean error (ME) and STD are decreased. The ME of peak period T_p is decreased from 1.608s to 1.531s. UDS will not only improve the estimate of the period, but it also preserves the quality of waveheight and improves the direction statistics, but the disadvantage is the number of valid data is not increased as much as the other two schemes.

As for the period, mean and peak direction are also sensitive to clutters in the radar DS and are more accurate in high seas (> 2 m). The best scheme to improve the direction estimate is SED (SEDcnnn). ME and STD of mean direction D_m are reduced from -14.35° to -11.79°, and from 43.07° to 39.91° respectively. Errors and STD of these sizes are commonly found in HF radar estimation of wave directions, e.g. in the same comparison using data from 2004, the ME and STD of mean direction with buoy is -13.61 and 47.0. An improvement has also been found in peak direction D_p, reducing ME and STD to some extent. After this scheme, waveheight is almost unaffected while the period is a little worse. A minus sign in ME is seen for all cases, suggesting a bias in wave direction estimation by radar, i.e. the waves estimated by radar are travelling more to the north than buoy estimates. But, a rigorous testing is needed in future to confirm this. This error is large, which shows the limit of HF surface wave radar accuracy in wave direction estimation. It is mainly accounted for by the angular spread of the radar, the spatial coverage of the intersection of dual radar beams (larger than the area of the buoy estimate), and the angular resolution in the inversion algorithm, where the number of different wavenumbers of ocean waves in a grid map is finite.

To sum up, for waveheight improvement, the joint application of ED and GED on classified DS is proposed. For period improvement, UDS scheme is advocated. For direction improvement, SED and UDS are both recommended. At the moment, UDS is recommended as an algorithm to improve all of these parameters together to some extent, although it is limited in improving the waveheight estimate.

Table 5.4 Comparison of statistics of waveheight, period, and direction estimates using a whole month data of Feb. in 2005. 'cono'- original DS from 'cm' and 'np'; 'cnnn'-new DS from 'cm' and 'np'; 'wbi'- with buoy information; C.C.- correlation coefficient; ME- mean error; STD- standard deviation; the numbers in bold are the estimates of these statistics by original radar data. For hs, the unit of ME and STD is m; for T_1, the unit of ME and STD is s; for D_m, the unit of ME and STD is degrees.

(a) Waveheight

Scheme	No. of valid data	hs			Peak hs		
		C.C.	ME	STD	C.C.	ME	STD
cono	312	**0.942**	**0.095**	**0.37**	**0.871**	**0.197**	**0.293**
EDcnno	287	0.934	0.072	0.39	0.871	0.187	0.287
GEDcnno	347	0.935	0.25	0.38	0.888	0.26	0.268
EDGEDcnno	365	0.935	0.336	0.41	0.875	0.294	0.298
EDGEDcnnnwbi	335	0.964	0.286	0.28	0.898	0.271	0.258
SEDcnnn	347	0.941	0.148	0.37	0.874	0.224	0.290
UDScnnn	322	0.946	0.123	0.36	0.89	0.204	0.277

(b) Period

Scheme	No. of valid data	T_1			T_p		
		C.C.	ME	STD	C.C.	ME	STD
cono	312	**0.495**	**0.766**	**1.555**	**0.316**	**1.608**	**4.57**
EDcnno	287	0.575	0.723	1.247	0.323	1.647	4.83
GEDcnno	347	0.582	0.77	1.291	0.310	1.628	4.88
EDGEDcnno	365	0.562	0.719	1.302	0.318	1.573	4.80
EDGEDcnnnwbi	335	0.485	0.770	1.560	0.300	1.59	4.70
SEDcnnn	347	0.459	0.79	1.681	0.276	1.572	4.85
UDScnnn	322	0.633	0.629	1.103	0.307	1.531	4.34

(c) Direction

Scheme	No. of valid data	D_m		D_p	
		ME	STD	ME	STD
cono	312	**-14.35**	**43.07**	**-17.37**	**51.31**
EDcnno	287	-18.224	46.77	-17.55	54.41
GEDcnno	347	1.124	60.94	-0.31	68.1
EDGEDcnno	365	-17.598	48.22	-19.64	57.64
EDGEDcnnnwbi	335	-16.086	51.81	-17.84	59.77
SEDcnnn	347	-11.793	39.57	-16.61	48.59
UDScnnn	322	-12.579	39.91	-17.11	48.85

5.4 Summary

The intention of this chapter has been to propose a general and robust solution to address the clutter mitigation problem in the RD image, in order to improve the wave estimation. The first problem addressed is the suppression of RFI and II, two major clutters preventing the DS from accurately inverting wave directional spectrum. RFI and II are found to be Doppler-correlated and range-correlated distributed in the RD image respectively. Using these typical graphical features, the RFI and II are characterized by a clutter-signal space in the training matrix, and then projected and removed from the processing matrix. The assumption made here is that the RFI and II in the training and processing data are homogeneous. The more the clutter meets this assumption, the less clutter residual is left.

This idea is then generalized to suppress arbitrary clutter due to the good feature of this solution that it doesn't require any information of the visible clutter itself and it just makes use of the clutter information contained in one area of the RD image to mitigate the clutter in another area, assuming that they are homogeneous. Hence, the construction of the training and processing areas becomes especially significant to the mitigation performance. If the clutter is homogeneous, a large training area will be selected, and a static (one-off) processing is enough. Otherwise, a small training area should be selected (the assumption is approximately held only in a small area) and adjacent to the processing area. Thus the mitigation processing should be done in a dynamic (window-by-window) way. The generalized scheme provided supervised, unsupervised, static, dynamic algorithms for a complete solution to mitigate clutter of arbitrary graphical feature, i.e. shape, correlation. If the clutter is not

mixed with the sea echo, supervised or unsupervised could both be appropriate to remove this clutter. If the clutter is mixed with the sea echo, then use supervised if the training data of the clutter is available, and use unsupervised otherwise. For the unsupervised method, four schemes are developed for four types of the constitution of the mixture signal (sea echo and clutter). Although only a few examples of original and clutter-mitigated DS have been included, many different examples have been tried which behave in much the same way.

The mitigation solution for RFI and II is a specific case of this generalized solution. The generalized scheme is able to deal with the RFI and II mitigation in any degree of complexity by removing the other clutters first, i.e. simplifying the complexity level of the RD image. Due to the generality, however, this scheme requires the input of several parameters, which has the major drawback that it requires intensive human invention. Up to now, the visible clutter mitigation scheme has not been operationally implemented. Future work for the development of the operational version can be focussed on the automatic identification of the clutter type, e.g. RFI or II, regular or irregular, Range-correlated or Doppler correlated, small scale or large scale, and the automatic setting of the parameters. Using the wave spectra measured by the buoy to simulate a DS and compare the simulated and measured DS to help identify the optimal parameters is also worth thinking about.

Another problem addressed in this chapter is the clutter which is small in scale and hidden in the in-band area. This problem is challenging in that no training data containing the clutter information is available. Invisible clutter mitigation schemes, ED, GED, and SED, are developed to solve this problem. Comparisons were conducted between ocean wave directional spectra from Pisces dual radar system and a waverider buoy. It has been found that the ED scheme reduces overestimated swell and GED could increase the underestimated wind sea. When the DS are grouped into under-, proper, and over- estimation classes by the buoy data, a joint application of ED and GED has shown a statistical improvement of the waveheight estimation. Unlike ED and GED, that have to be performed in a selective way, SED and UDS, which are independent of the DS, are tested operationally without any condition. The comparison of the statistics of waveheight, period, and direction using a whole month of (Feb, 2005) data before and after the operational mitigation processing provided evidence of the improvement. For waveheight improvement, the joint application of ED and GED on classified DS is proposed. For period improvement, UDS scheme is advocated. For direction improvement, SED and UDS are both recommended.

Several issues are not yet solved and worth further exploring:

(1) How to identify the problematic DS which contribute to the error in the swell or wind sea. This will help choosing the problematic DS and the matched scheme.

(2) How to improve the waveheight but also preserve or improve the quality of

direction and period.

(3) Is it possible to combine the advantages of the three operational schemes together?

(4) The optimal setting of parameters requires further consideration and investigation for various clutter or error cases.

(5) How to generate a scheme to automatically implement image recognition, i.e. identification of the visible clutter, and then selection of a matched visible clutter mitigation scheme and choice of appropriate parameters.

Within this chapter, all the methods proposed are based on the 20 DS from one beam, i.e. a RD image. For application to a complete dataset, e.g. Pisces data are obtained on 3 beams from each site, I would apply the methods beam by beam. The WERA radar provides DS that have been either interpolated or DFTed onto a rectangular geographical grid, so I really need to reprocess the data (as discussed earlier) using a different beamforming procedure to be able to access data on individual beams. Finally, as mentioned above, for the operational application of the joint ED and GED scheme, the determination of whether a DS is going to derive under-, proper-, or over- estimated significant waveheights is needed and will be elaborated in the next chapter.

Chapter 6 Doppler spectrum quality assessment schemes

6.1 Introduction

As discussed at the end of chapter 5, the performance of ED and GED schemes for wave estimation depends on knowledge of the quality of DS. By 'quality', I mean the ability to provide under, proper, or overestimation of waveheight. The quality of a DS is affected by many factors, including attenuation of transmitted radio waves, external clutters, internal noise, radar frequency, and atmosphere influences. More often than not, the common parameters that assess the quality of a DS are signal-to-noise ratio, noise level, first and second order moments etc. These measurements provide a rough separation of valid and invalid DS. However, they are not sufficient to predict the accuracy of the wave measurements. Therefore, this chapter proposes a statistical pattern recognition approach to model and classify all the DS into three classes, i.e. under, proper, and over classes, based on the buoy significant wave height estimates by extracting those spectra with radar wave height under 80%, within 20% or above 20% of the buoy height.

The statistical pattern recognition used here is a supervised approach with training and classification phases. A set of training data is assumed to have been provided, consisting of feature vectors that have been properly labeled with the correct class membership. Next, a learning procedure generates an optimal classifier which best classifies new data into the right class. This approach consists of three parts. The first part concerns preprocessing. A DS is a feature vector of D Doppler bins (D=1024). However, 1024-variate analysis is far beyond available computational power and unnecessary because Doppler bins outside the second order region are redundant for wave estimation. So the dimension of a feature vector is reduced from D to d by a preprocessing method. The second part is feature selection and extraction, aiming at finding the best feature space for classification. After this, the best features are extracted and the dimension is further reduced from d to 3. The third part is classification based on this feature space.

This chapter is outlined as follows: Section 6.2 shows the model for pattern recognition, which consists of three processing blocks: pre-processing block, feature selection/extraction block and classification block. Each block provides various processing options, the number of which is 2, 8 and 8 respectively. Section 6.3 describes the Density-based statistical classifier block by block. This first estimates the class conditional likelihood, a prior, and thus the posteriori, based on which classification is performed by Bayes rule. The feature selection methodology is described and the resulting classification accuracy is evaluated for different numbers of features. This approach is often chosen by pattern recognition researchers because it usually gives deeper insight into the structure

of the data space. However it fails when the classes are all mixed up in the feature space. Section 6.4 describes various methodologies for feature extraction. Section 6.5 elaborates various methodologies for classifier design, based on different proportions of training and testing sets. The indicator of the classification performance is the correct classification rate (CCR), namely the percentage of correctly classified testing data. The properties of various classifiers are discussed in terms of CCR, training speed, and stability.

The data-acquisition method is described in Chapter 2. The spectra are collected hourly from 20 range bins for N days in one month (N=28 for February and 31 for March). The dataset consists of $N \times 24 \times 20$ observations of 1024-bin DS.

6.2 Pattern recognition

There are several terminologies to make clear at the beginning. Firstly, a feature vector is a vector with each element being a random variable, namely a feature. It is a representation of real world objects and the choice of the representation strongly influences the classification performance. The criterion for choosing a feature is that the feature should behave as distinctively as possible among different classes, in order to ultimately yield a good classification result. In the context of this thesis, the original features I select are each Doppler bin in the DS, since they are the raw observations of the waves. To reduce the dimension of this large feature space, preprocessing is adopted to generate preprocessed features. Based on preprocessed features, the second and third terms, 'feature selection' and 'feature extraction', are used. Feature selection is selecting a subset of features in the original set of raw features, while feature extraction is transforming the original feature space into a new feature space with lower or higher dimensions. If the features are independent, the feature selection method is chosen, otherwise feature extraction will be more suitable (McLachlan and Wiley 1992).

A general model of the pattern recognition system is shown in Fig. 6.1. The classification and training phases are separated by a dashed line. The three parts of the pattern recognition approach introduced in section 6.1 are modularized in the three blocks.

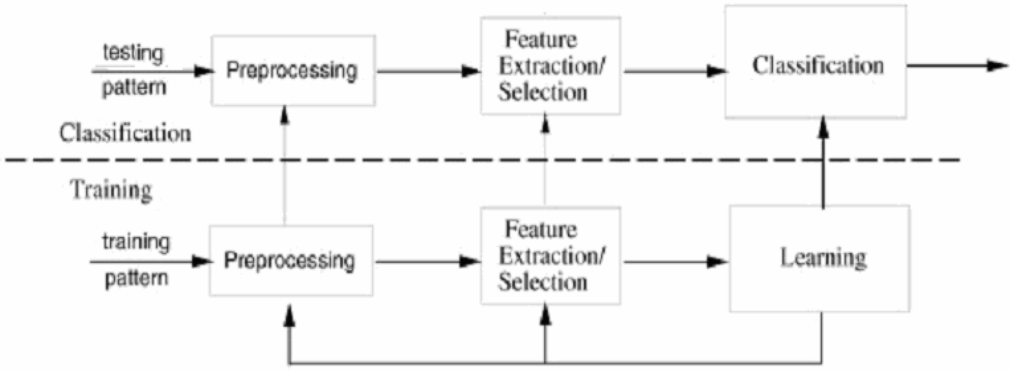


Fig. 6.1. A model for pattern recognition.

The first block is preprocessing, which is a module to preprocess the training or testing feature vectors in a 'sensible' way. The motivation for having this block is that the pattern recognition system is based on very few of the most important features, which characterize the class membership of the feature vectors. It is well known that the key wave information appears only in the first and second order region of a DS. Therefore, this preprocessing block first removes some redundant and irrelevant Doppler information which have a detrimental effect on the classifier performance. Two methods are developed to carry out this job (see Table 6.1). The first involves calculating 45 statistics of the first and second order continuum of a DS (described in Appendix A); the second just chooses the second order continuum beside the superior first order peak sfs, i.e. Doppler bins $sfs-50$ to $sfs+50$. These methods are capable of compressing essential information and removing redundant information respectively. In method 1, I first produce a reduced set of features which are not strongly correlated; the best features for classification are then selected from this reduced set. In method 2, the 100-bin Doppler section is selected as the set of features which has strong Doppler-correlation (excluding the first order region); the best features are transformed from this feature set by feature extraction. The relative merits of these approaches will be compared in the following sections.

The second block is Feature selection/extraction, which is a module to extract the best features. For this purpose, eight methods (see Table 6.2) are examined. Their performances for selecting or extracting the best features are discussed later in different subsections. In this chapter, the number of best features is no higher than 3, because 3 is the maximum dimension for visualization and also, as will be shown in Table 6.3, sufficient for classification compared with higher numbers.

Table 6.1 Eight options of feature selection/extraction on preprocessed feature space

Dimension reduction	Method name
Feature selection	Probability distance analysis (PDA)
Feature selection	Best Individual Feature (BIF)
Feature extraction	Principal Component Analysis (PCA)
Feature extraction	Factor Analysis (FA)
Feature extraction	General Discriminant Analysis (GDA)
Feature extraction	Canonical Correlation Analysis (CCA)
Feature extraction	Generalized Eigenvalue Decomposition (GED)
Feature extraction	Linear Preserving Projection (LPP)

The last block is learning or classification. Various classification methods (see Table 6.2) are examined and tested for this application. Their classification performances are compared and the optimal is chosen to predict the class membership of a DS, e.g. under, proper, or over estimation.

Table 6.2 Six options of classification on the best feature space

Classification Type	Classifier Description
KMEANS	Cluster analysis that partitions the observations into K clusters
SOM	Self organizing map
TREE	Hierarchy classifier
DA	Discriminant analysis
KNN	Classifying data using nearest neighbor method
SVM	Support vector machine

The focus of this chapter is on the statistical pattern recognition approach. Neural networks are also utilized in the classification block. The statistical pattern recognition is investigated in terms of two branches, i.e. the density-based approach and the geometric approach. Details of these methodologies are presented below.

6.3 Statistical pattern recognition based on the density-based approach

The derivation of the density-based (probability-based) pattern recognition is summarized as follows. Let $\underline{x} = (x_1, x_2, x_3)^T$ be a given feature vector to be assigned to one of c classes $\omega_1, \omega_2, \ldots \omega_c$ (in the specific case of Doppler pattern recognition, $\underline{x}$ represent a feature vector consisting of the best three features that are selected/extracted from a DS; $\omega_1, \omega_2, \omega_3$ represent under-, proper-, and over- estimation classes respectively). Let us have an a priori belief that the probability distribution function of each class is $p(\omega_i)$ and that a feature vector $\underline{x}$ belonging to that class has the likelihood $p(\underline{x} \mid \omega_i)$. Then, the posteriori (Kittler and Devijver 1981) is defined as $p(\omega_i \mid \underline{x}) \propto p(\omega_i) p(\underline{x} \mid \omega_i)$, based on which a boundary decision function is calculated. Bayes decision rule is defined to minimize a risk function, $R(\omega_i \mid \underline{x})$, for assigning feature vector $\underline{x}$ to class ω_i, i.e. to minimize the classification error. The risk function $R(\omega_i \mid \underline{x})$ is in the form of :

$$R(\omega_i \mid \underline{x}) = \sum_{j=1, j \neq i}^{c} L(\omega_i, \omega_j) \cdot p(\omega_j \mid \underline{x}) \tag{6.1}$$

where $L(\omega_i, \omega_j)$ is the loss incurred in misclassifying class ω_j into ω_i. For the simplest case, L is defined as $L(\omega_i, \omega_j) = \begin{cases} 1, i \neq j \\ 0, i = j \end{cases}$. Thus, the risk function is simplified to $R(\omega_1 \mid \underline{x}) = p(\omega_2 \mid \underline{x}) + p(\omega_3 \mid \underline{x}) = 1 - p(\omega_1 \mid \underline{x})$. The minimal risk rule can be viewed as maximizing the posteriori probability. Given that the prior probability is estimated by $p(\omega_i) = \dfrac{\text{no. of points in class } \omega_i}{\text{total no. of points}}$, the calculation of likelihood is the primary task for the density-based approach. An overview of solutions for likelihood estimation is shown in Fig. 6.2 (Jain et al. 2000).

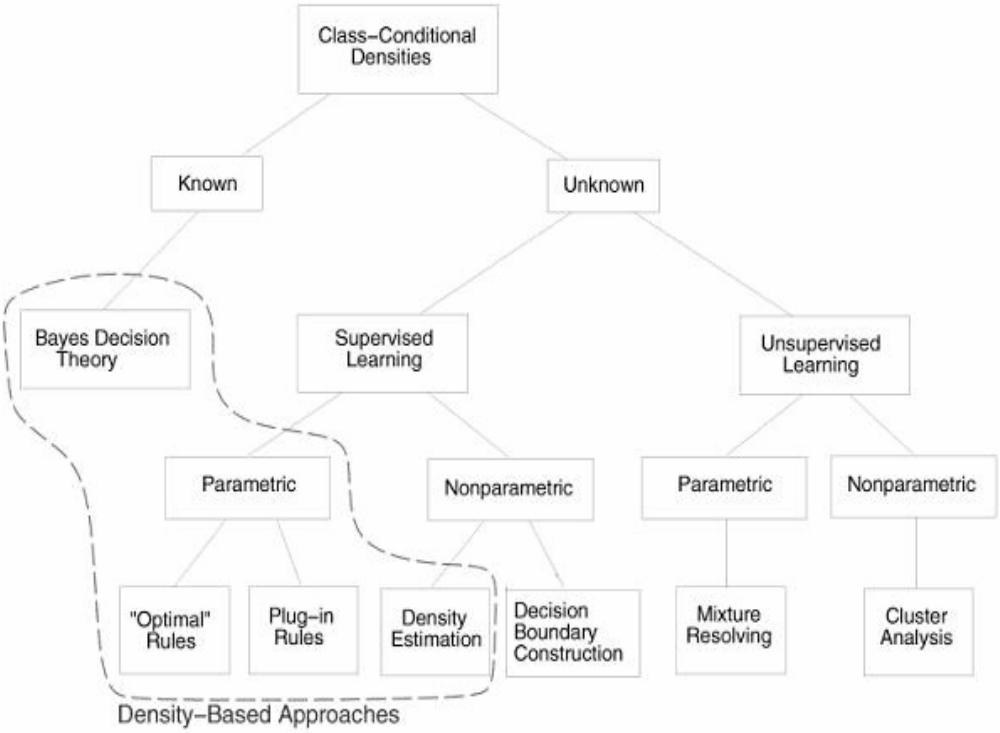


Fig.6.2 Framework of likelihood estimation (this figure is copied from (Jain et al. 2000)).

In this application, the form of likelihood is unknown but can be estimated from a set of training feature vectors. Wave estimations obtained from in situ buoys are used for the class labels of each training feature vector, i.e. 'Supervised Learning' block in Fig.6.2. Note that under the branch of 'Class-Conditional Densities-unknown-supervised learning', there are two categories, i.e. parametric and non-parametric, depending on whether the form of likelihood is known or not. In this section for density-based approach, both parametric and non-parametric methods are investigated. For the parametric case, the form of the likelihood is assumed in the form of a multivariate normal distribution but the parameters are unknown. A solution called Bayes 'Plug-in Rules' (Jain et al. 2000) resolve this problem by estimating these parameters, like means and variances, in the density functions using training feature vectors. For the non-parametric case, i.e. the form of the likelihood is unknown, Parzon window methodology (Wang et al. 2005) is used to estimate the density functions with kernel functions. Considering that these two solutions are under different hypotheses, some hypothesis tests must be performed in the feature space.

This density-based approach adopts the three-block model of pattern recognition (see Fig.6.1). The 'preprocessing' and 'feature selection/extraction' blocks implement the feature selection. The 'learning' block estimates the likelihood and posteriori of each feature vector. The class label, either under, proper, or over estimation, will be given to a feature vector if the posteriori of that class is maximum.

6.3.1 Feature selection

For the 45 weakly correlated statistics calculated in the preprocessing block, three feature selection methods are proposed. The first and second are termed 'probability distance analysis (PDA)' methods, while the third method is called 'best individual features (BIF)' (see Table 6.2). By PDA, the dimension of its feature space is reduced from 45 to 21. After BIF, this dimension is further reduced from 21 to 2 or 3.

The first method is to select features using probabilistic distance as a criterion of class separability. For a univariate x_j (the jth feature in a feature vector $\underline{x}$) and class ω_i, the feature x_j is not informative for classifying ω_i if $p(x_j) = p(x_j \mid \omega_i), i = 1,2,3$ ($p(x_j) = \sum_i p(x_j|\omega_i)p(\omega_i)$ is the unconditional PDF of x_j), which means the probability distribution of x_j has no separation ability for class ω_i. Otherwise, the bigger the L^1-norm (given an n-element vector $\underline{x}$, its L^1-norm is defined by $\|\underline{x}\| = \sum_{i=1}^{n} |x_i|$) distance of two PDF functions $p(x_j)$ and $p(x_j \mid \omega_i)$, the better class separation ability of the feature. In order to calculate this L^1-norm distance $d_i(x_j)$, the univariate PDFs $p(x_j)$ and $p(x_j \mid \omega_i)$ are estimated according to the procedures below.

1. The new feature vector after preprocessing is $\underline{x} = [x_1, ..., x_j, ...x_{45}]^T$, and the whole set is $X \in \mathbb{R}^{N \times 45}$, where N represents the number of feature vectors in a month's dataset (N=538 for Feb and 552 for March, 2005).

2. X is split into a training set X_{tr} and a testing set X_{te}.

3. In the training set, each feature vector is labelled by ω_i or simply i, with 1, 2, and 3 representing whether its wave heights are under- proper- or over-estimated compared to the buoy height respectively.

4. The labelled X_{tr} is classified into three subsets as $X_{tr} = \bigcup_{i=1}^{3} X_i, X_i \in \mathbb{R}^{n_i \times 45}$,

where X_i is the subset of class i; n_i is the number of feature vectors in X_i.

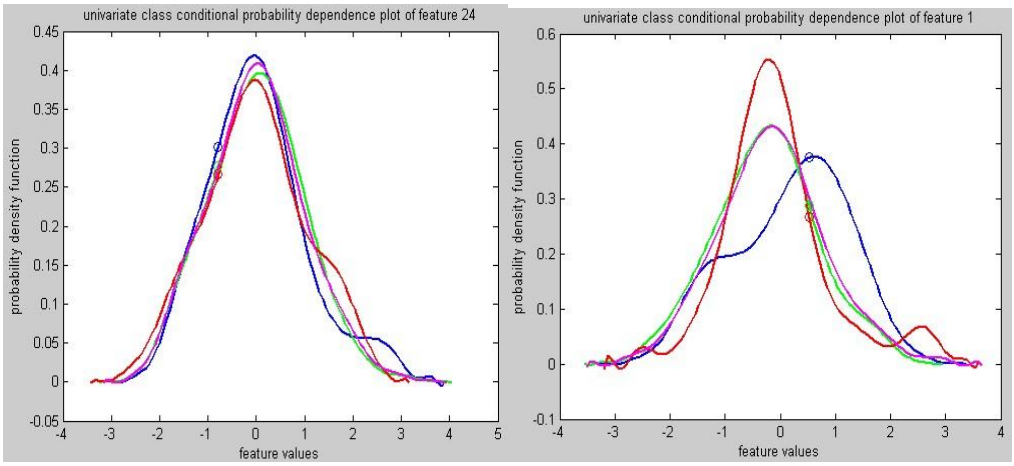
5. Univariate PDFs $p(x_j \mid \omega_i)$ and $p(x_j)$ are estimated using X_i and X_{tr} respectively, by a non-parametric univariate normal kernel function (this kernel method is explained in section 6.3.2; parametric estimation requires a hypothesis test to show the PDFs of features are normally distributed, while non-parametric doesn't require such a test).

6. By binning the values of x_j, I can produce the histogram of this feature $\hat{p}(x_j \mid \omega_i)$; I then find it by a polynomial to fit its representation, $\hat{\hat{p}}(x_j \mid \omega_i)$.

7. The optimal order of the polynomial fitting curve is set to 15, which provides the minimum root square error between $\hat{p}(x_j \mid \omega_i)$ and $\hat{\hat{p}}(x_j \mid \omega_i)$.

8. For the feature x_j, $1 \le j \le 45$, three continuous univariate likelihood $p(x_j \mid \omega_i)$ and one unconditional PDF $p(x_j)$ are estimated, each represented by 16 coefficients. In the Matlab program, they are stored in a matrix $P \in \mathbb{R}^{45\times4\times16}$.

Fig.6.3(a) depicts the univariate pdfs of feature 24, i.e. $p(x_{24})$ and $p(x_{24} \mid \omega_i), i = 1,2,3$ with magenta, blue, green and red respectively. Fib.6.3(b) depicts the curves for feature 1 with the same color setting. It is apparent that in Fig.6.3(a) the blue, green and red curves almost overlap the magenta curve, indicating that $d_i(x_{24}), i = 1,2,3$ are all approximately 0, i.e. this feature is not informative for all classes and should be eliminated from the feature set. In contrast, the univariate PDFs in Fig.6.3(b) display much larger probability difference. In this way, 24 features are eliminated out of 45, whose $d_i(x_j)$ are small.



(a) (b)

Fig. 6.3 Four univariate PDF curves: (a) Feature 24: completely overlapped; (b) Feature 1: half overlapped. Curves of likelihood of $\omega_1, \omega_2, \omega_3$ are in blue, green and red; the PDF curve of the mixture is in magenta. Blue, green, and red circles symbolize the probability values of class 1, 2, 3 for a certain feature value, eg. $x_{24} = -0.85$ in (a) and $x_1 = 0.82$ in (b).

The second feature selection method is an adjustment of method one, using the signs of the probability dependence of the 21 features to generate new features and the representative feature vectors of each class. The probability dependence function is defined by $\Delta d_i(x_j) = p(x_j) - p(x_j \mid \omega_i)$, where j is the index of feature and i is the index of the class. This method has two advantages: ignoring the impact of polynomial fitting error on the estimation of likelihood, and compressing the three probability dependence functions $\Delta d_i(x_j)$ of feature x_j into one function of x_j. The procedures for the second method are given below:

1. Let $\underline{x} = [x_1, x_2, \ldots, x_{21}]^T$ be a randomly selected feature vector. Each of the 21 elements is substituted into four univariate PDFs (the 15-order polynomial function), to generate four scalar probabilities $p(x_j \mid \omega_1)$, $p(x_j \mid \omega_2)$, $p(x_j \mid \omega_3)$ and $p(x_j), j = 1, 2, ..., 21$.

2. The three scalar signs of probability dependence of feature x_j are: $s_{ji} = sign(\Delta d_i(x_j))$, $i = 1, 2, 3$, such that $\Delta d_i(x_j) = p(x_j) - p(x_j \mid \omega_i)$.

3. s_{ji} has two possible values, -1, or 1. The three signs can be coded into one scalar with a base 2 by a formula: $x_j^s = 2^{2\times s_{j1}} + 2^{1\times s_{j2}} + s_{j3}$, where x_j^s compress the three signs information of probabilistic dependence of feature x_j into one new feature. This formula is designed with one condition that this is a one-one projection between a combination of signs (s_{j1}, s_{j2}, s_{j3}) and a value of x_j^s, i.e. 8 combinations correspond to 8 x_j^s.

4. Based on the 21 transformed features, a new feature vector $\underline{x}^s = [x^s{}_1, x^s{}_2, \ldots, x^s{}_{21}]^T$ is generated. The values of the new features are in a few fixed discrete numbers, e.g. if signs are (-1, 1, 1), the feature value is $x^s = 0.25 + 2 + 1 = 3.25$.

For a feature vector viewed as a point with 21 coordinates in a 21-dimension space, the new feature vectors are compressed into a few points rather than diversely or continuously distributed. If the best three features are selected out of 21, those points in the 3-dimension space belonging only to ω_i can be viewed as representative points of ω_i. The new testing feature vectors can be obtained after preprocessing, feature selection by method one and method two. If the testing feature vector overlaps the representative points of ω_i, they could be classified into class ω_i. Note that this method is not adopted in this feature vector recognition job but is explored as an extension to the first feature selection method. This is because a disadvantage of this method is that the number of feature vectors should be sufficiently large to estimate each univariate likelihood accurately. Otherwise, the representative points found in one dataset obtained from one month or one radar site cannot classify the testing feature vectors obtained in other cases, e.g. different month, radar, or sea area.

The third feature selection method is to fulfil the task of dimension reduction from 21 features to 3. Five criteria of class separability are used to rank key features from the 21 features, in order to find the optimal criterion and the best three features. Each independent evaluation criterion is for binary classification, so the selected best features are also class-conditional, e.g. feature 41 and 19 are the best features for classification of class 1 in most cases (see Table 6.3).

The third method uses three arguments in the Matlab algorithm: the criterion

name; R (an argument defines the correlation between the selected features and unselected features); N (the number of the most significant features). There are five criteria (Liu and Motoda 1998) (Theodoridis and Koutroumbas 1999) used to assess the significance of every feature for separating two labeled classes: 1) absolute value two-sample T-test with pooled variance estimate, 2) relative entropy, also known as Kullback-Lieber distance or divergence, 3) minimum attainable classification error or Chernoff bound, so-called Bhatttacharya method (see the details in Appendix A), 4) area between the empirical receiver operating characteristic curve (ROC) and the random classifier slope, 5) Wilcoxon-absolute value of the u-statistic of a two-sample unpaired Wilcoxon test, also known as Mann-Whitney. These are all possible methods implemented in the Matlab toolbox. If a large value of $R, 0 < R < 1$ is set, the features that are highly correlated with the selected good features are less likely to be included in the key feature list. This argument is especially useful for reducing the redundancy of strongly correlated features, i.e. the selected features are as independent as possible. The 21 features to be analyzed are statistics with different means and variances, and the latter two methods in the list above require normalization, applied to every feature to ensure comparability among different features.

A comparison of different feature selection processes with various settings of the arguments is shown in Table 6.3, where R is set to 1 for all experiments; 2, 3, 5, 8, 10 represent the dimensions of the selected feature space, N. In the 'Class' column, 1, 2, 3 are the class names for under-, proper-, and over-estimation respectively. 'NA' means 'not applicable'; this occurs when the matrix of the selected features becomes non-positive definite, and linear discriminant classification cannot proceed to get any result. It is noted that the best criterion is 'brattacharyya', with class-conditional CCR a little higher than the rest. The selected best features are features '19', '35', and '36', because they rank as the top three features by the criterion 'brattacharyya'.

Several interesting phenomena are noticed from the table.

1. As the number of best features increases from 2 to 3, 5, 8, 10, we tend to see small increase or in some cases slight decreases. Hence, there is no great advantage of using more than 2 or 3 features.

2. It has been found that even for the criterion 'brattacharyya', the CCR is no higher than 0.71. The classification performance of a dataset is determined by the statistics of the data. There is always an upper boundary of the CCR that a dataset can reach (Jain et al. 2000). 0.7 is shown to be an upper limit of the classification performance for this dataset. This is low compared to about 0.9 in other applications of such methods, e.g. classifying protein profiles (MathWorks 2011).

3. It has been found that those features selected by different criteria are fairly similar. The most commonly selected features are '25', '26', '35', '36', and

'19'. The index '25, 26' measures the ratio of Bragg peaks at rb 4 and rb 5 respectively (the intersection of dual radar beams is fixed and covers a larger spatial area than the buoy coverage. Rb 4 and 5 are both in the intersection with rb 5 closer to the position of the buoy). The physical meaning of the ratio is that it reveals the wind direction and is used in the wind direction estimation algorithm. Index '35, 36' measures the ratio of second order peaks at range bin 4 and 5 respectively. Index 19 is the covariance between the first order ratio and the second order ratio. No explicit physical meaning associated with the second order ratio and the covariance, but these statistics, in contrast with others, are more direct measurements of the wave-wave interactions and the Bragg backscattering.

Table 6.3 Correct classification rate using five criteria for best individual feature (BIF) selection. Class 1, 2, 3 represent under-, proper-, and over-estimation classes respectively. For each class, the best 2, 3, 5, 8, and 10 features are selected from the set of 21 features. Each number tells us the CCR for a class using a linear classifier based on the selected best features. The best features are listed in the BIF column in descending order.

Criterion	Class	N=2	N=3	N=5	N=8	N=10	BIF
Ttest	1	0.6952	0.6859	0.6933	NA	NA	41,19,1,4,35
	2	0.6524	0.6320	0.6413	0.6654	NA	35,41,1,4,36,19,40,7
	3	0.6283	0.6320	0.6338	0.6357	NA	36,23,35,19,10,43,40,38
Entropy	1	0.6952	0.6840	0.6914	0.6952	0.6933	19,41,4,6,1,35,7,38
	2	0.6375	0.6283	0.6636	NA	NA	30,41,19,35,4
	3	0.5799	0.5688	0.6097	NA	NA	30,4,3,21,36
Bhattach-arya	1	0.6840	0.7082	0.7045	NA	NA	19,35,41,36,25
	2	0.6283	0.6394	0.6599	NA	NA	35,36,19,41,25
	3	0.6301	0.6320	0.6301	NA	NA	36,35,23,19,26
Roc	1	0.6524	0.6840	0.6970	0.6952	NA	4,41,19,1,30,35,7,34
	2	0.6301	0.6320	0.6413	NA	NA	35,1,41,4,36
	3	0.6283	0.6320	0.6283	NA	NA	36,23,35,10,41
Wilcoxon	1	0.6245	0.6561	0.6487	NA	NA	4,23,35,38,6
	2	0.6375	0.6357	NA	NA	NA	35,4,36,
	3	0.6208	0.6245	0.6283	0.6245	NA	32,19,30,21,11,6,23,24

6.3.2 Likelihood estimation with the best 3 features

Given the selected best three features, some hypothesis must be tested to decide whether to use a parametric or non-parametric method for the likelihood estimation. If the likelihood can be assumed to be a multivariate normal distribution, parametric estimation will work; otherwise, non-parametric estimation will be used. So the hypothesis that the likelihood of the 3-feature feature vector obeys a multivariate normal distribution is tested using the whole

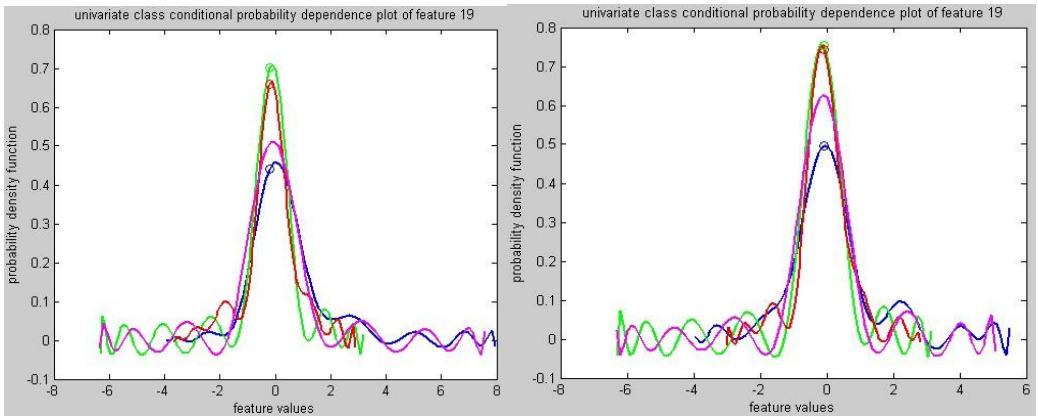
dataset.

Hypothesis tests for normality

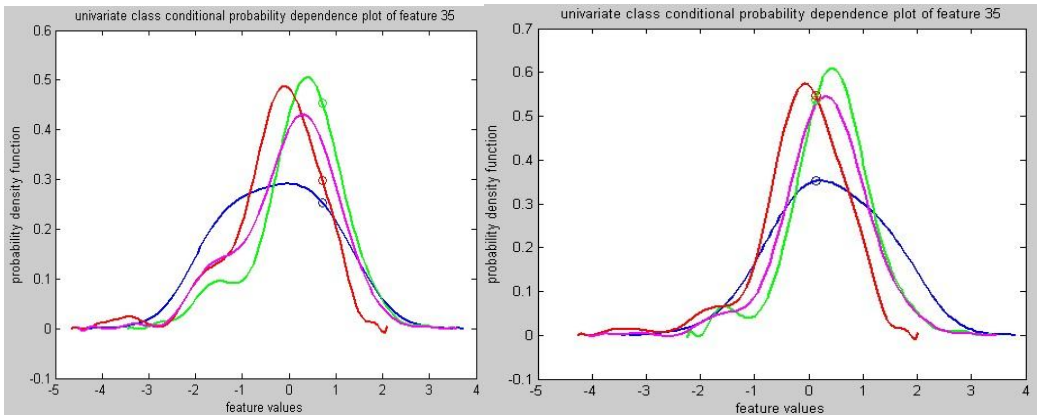
Burdenski (2000) stated that multivariate normality means all of the variables must be univariate normal, and all possible pairs of variables must be bi-variate normal. To check this, the following procedures are adopted.

1. Examine the three variables and check if they follow univariate normal distribution.
2. If the three features display univariate normality, examine the bivariate normality of each pair of variables in a scatter plot. If the scatterplot doesn't reveal an ellipical shape, consider replacing one of that pair of variables with another one. Alternatively, the variable can be transformed to be more normal by taking the square root, squares, the natural log or log-ten of the components. After the transformation, the skewness and kurtosis should be nearer zero.
3. If both univariate and bivariate normality are attained, examine the Mahalanobis distance by Chi-square scatterplot (Burdenski 2000). If it displays a fairly straight line, it suggests multivariate normality.

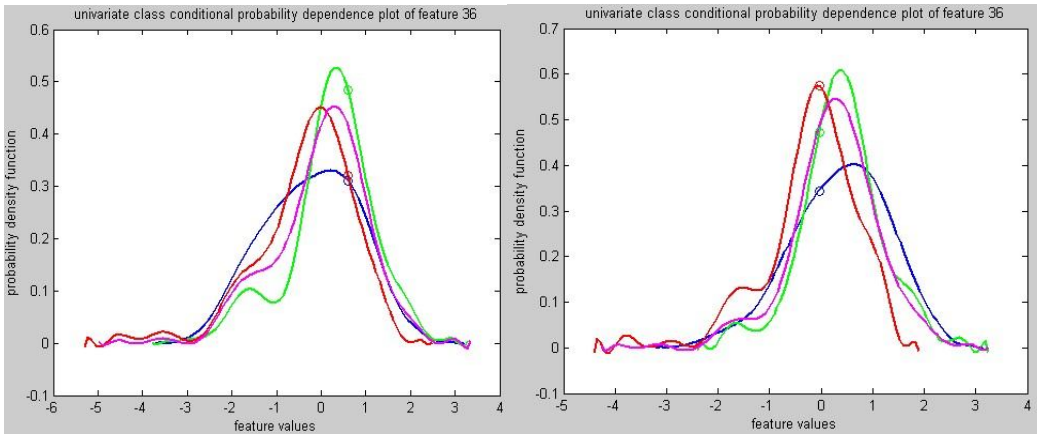
Univariate normality and bivariate normality are checked through Fig.6.4 and Fig.6.5 respectively. In Fig.6.4, different size of samples are chosen to estimate the univariate likelihood. Using the same color setting as before, the curve for $p(x_j)$ is plotted with magenta and the curves for $p(x_j \mid \omega_i)$, $i = 1,2,3$, $j = 19,35,36$ are plotted with blue, green and red respectively. It can be seen in plot (c) and (d) that none of the blue, green, red and magenta curves are Gaussian; when the sample size is doubled, the blue curve is more non-Gaussian. Fig 6.5 contains 9 subplots. Histograms of each univariate feature are shown in the diagonal subplots. The 6 off-diagonal subplots show all combinations of bivariate scatter plot. It is confirmed from the diagonal histograms that the three features don't follow normality, let alone multivariate normality. Hence, the parametric approach won't work.



(a) (b)



(c) (d)



(e) (f)

Fig.6.4 Univariate likelihood curves of feature 19, 35 and 36. The univariate likelihood in class 1,2,3 and the mixture are plotted with color blue, green, red, and magenta: Plot (a)(b) are from feature 19; plot (c)(d) are from feature 35; plot (e)(f) are from feature 36. In (a)(c)(e), the likelihoods are estimated with a whole month DS; in plot (b)(d)(f), the likelihoods are estimated with half month data. 'feature values' means the values of the feature variate.

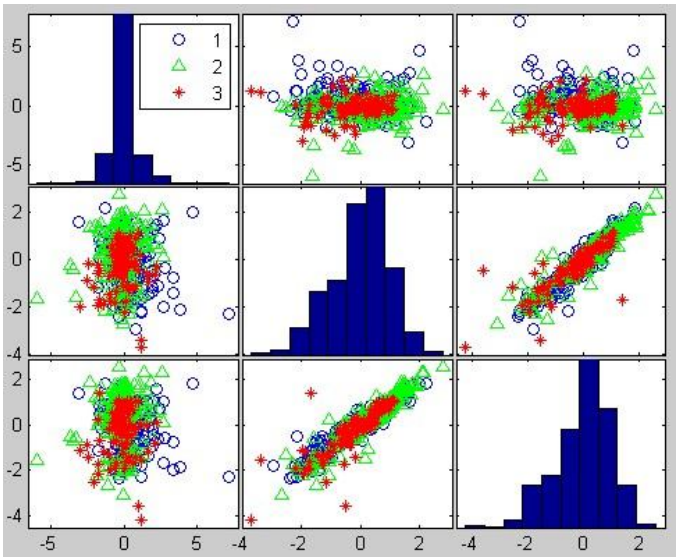


Fig.6.5 Scatter plot matrix with grouped features. This whole plot consists of 9 subplots. Histograms of univariate feature 19, 35, and 36 are shown in the diagonal subplots from top to the bottom. The 6 off-diagonal subplots show all combinations of bivariate scatter plot. Blue circle, green triangle and red stars represent feature vectors from class 1, 2, 3 respectively.

Non-parametric estimation of likelihood

Since the form of likelihood is not known and unable to be approximated using a multivariate normal distribution, a non-parametric solution is adopted. The most appropriate methodology to be used in a non-parametric case is the Parzen window approach (Botev et al. 2010), also known as multivariate kernel density estimation (KDE). Here, this algorithm is applied to estimate the likelihood of the multivariate feature vectors, more specifically tri-variate.

Let $Z \in \mathbb{R}^{538\times 3}$ be the feature space containing 538 feature vectors of 3 features. The kernel density estimation of the tri-variate likelihood, $p(\underline{z} \mid \omega_i)$, is given in (6.2). It expresses that to form a kernel density estimate, the mean of n_i kernels is taken, each kernel with mean $\underline{z}_j$ and bandwidth matrix H.

$$\hat{p}_H(\underline{z} \mid \omega_i) = \frac{1}{n_i}\sum_{j=1}^{n_i} K_H(\underline{z} - \underline{z}_j) \tag{6.2}$$

where n_i is the number of feature vectors in class ω_i; $\underline{z} = (z_1, z_2, z_3)^T, \underline{z}_j = (z_{j1}, z_{j2}, z_{j3})^T, j = 1,2,...,n_i, i = 1,2,3$; $H \in \mathbb{R}^{3\times 3}$ is the bandwidth matrix which is symmetric and positive definite; $K_H(\underline{z})$ is a

multivariate kernel function defined by:

$$K_H(\underline{z}) = |H|^{-\frac{1}{2}} K(H^{-\frac{1}{2}}\underline{z}) \tag{6.3}$$

where $|H|$ stands for the determinant of matrix H. Note that $K(\underline{z})$ is a non-negative multivariate function satisfying that $\int K(\underline{z})d\underline{z} = 1$; and the accuracy of KDE doesn't rely on the choice of the kernel function K, so the standard multivariate normal kernel is used and given as:

$$K(\underline{z}) = (2\pi)^{-3/2} \exp(-\frac{1}{2}\underline{z}^T\underline{z}) \tag{6.4}$$

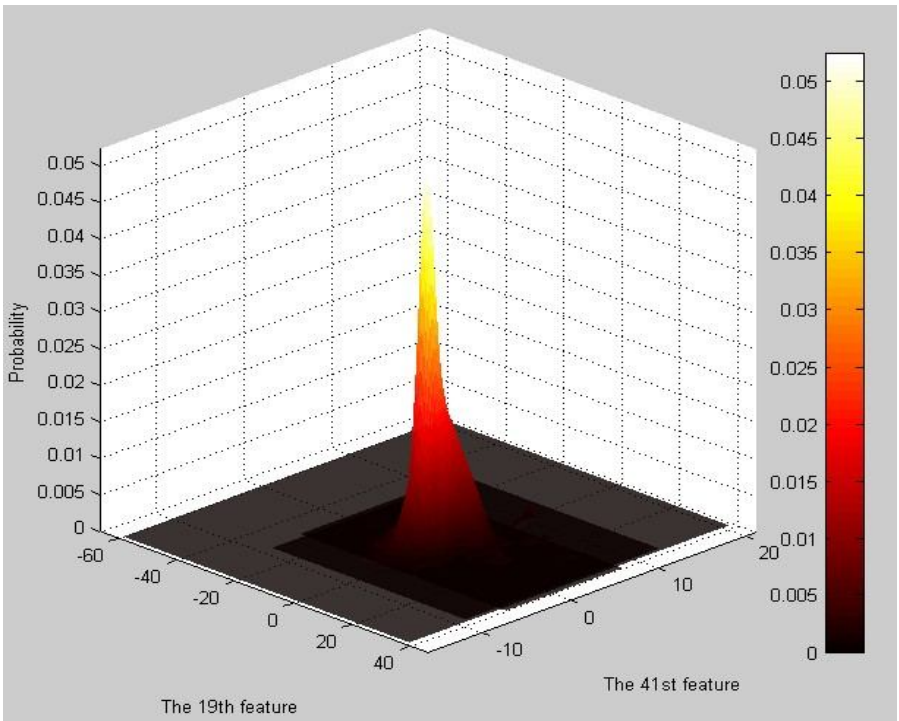
The very important factor in determining the shape and the performance of a KDE is the choice of bandwidth matrix (Wand and Jones 1995). Since the bandwidth matrix can be thought of as the scaling of the kernel function, it controls both the spread and the orientation of the kernel (the explanation of bandwidth matrix selection is provided in the Appendix A). The bandwidth matrix has a parameter of orientation. An option of H is the diagonal matrix, D, with positive entries on the main diagonal (Wand and Jones 1994), which is the most commonly used and is used in this work (the other two options are: $S = k\mathrm{I}$, a positive scalar times the identity matrix, and F, symmetric positive definite matrices).

As shown in Table 6.3, the correct classification rate has not much difference between 2- and 3- dimension feature space, so the problem of multivariate density function calculation can be simplified to bivariate density estimation. The KDE algorithms use the Botev el al (2010) method for univariate (for step 5 of the first method of feature selection) and bivariate KDE in a Matlab routine. Feature 19 and 41 are selected as the bivariates because they are the best two features for the classification of class 1, under most criteria. The bivariate likelihoods of all classes, i.e. $p([x_{19}, x_{41}]^T \mid \omega_i)$, are estimated and shown in Fig.6.6. The priori probability of each class $p(\omega_i)$ is estimated by $\frac{n_i}{N}$, where n_i is the number of feature vectors in class ω_i and N is the number of feature vectors in the whole training data. Finally the posteriori $p(\omega_i \mid [x_{19}, x_{41}]^T) \propto p(\omega_i)p([x_{19}, x_{41}]^T \mid \omega_i)$ is calculated. Using the Bayes decision rule of choosing the maximum of the three posteriori, the CCR has been found to be between 0.5 and 0.6.

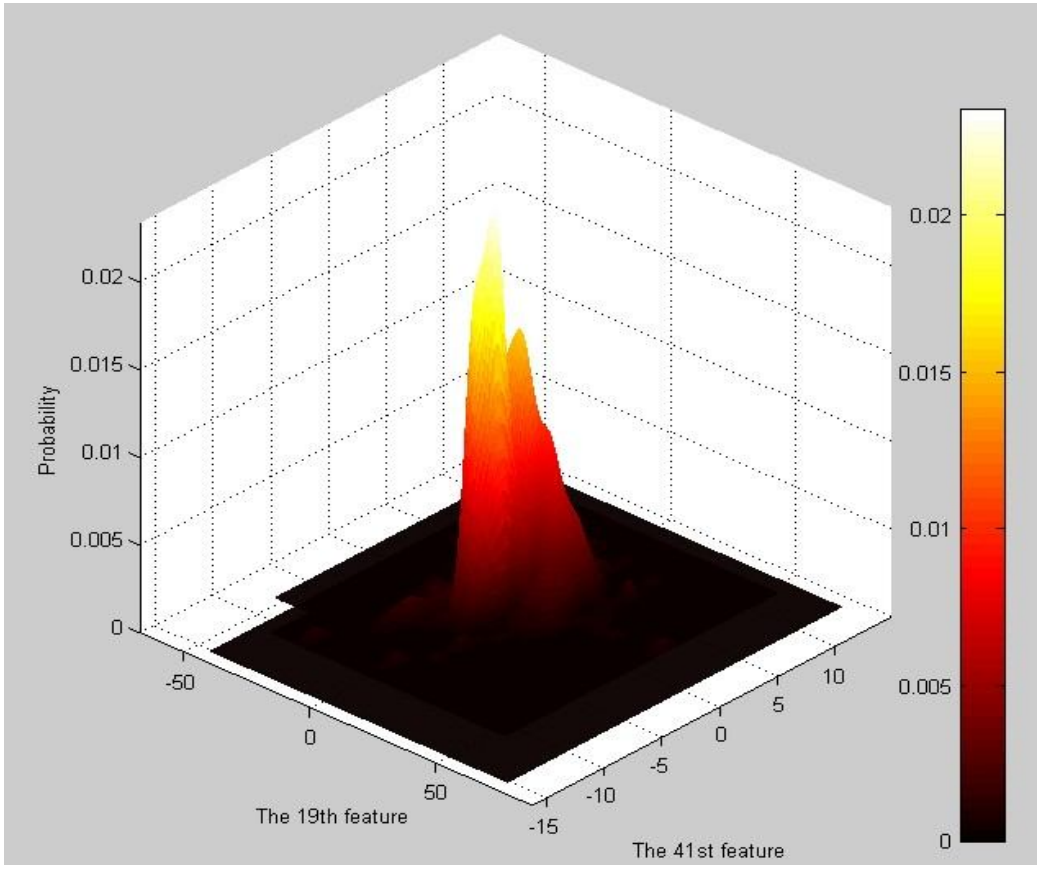
In conclusion, the density-based approach relies on the accuracy of multivariate KDE of likelihood and the estimation of a priori probability. There are two reasons to explain the poor classification performance of the Bayes approach.

One is the nature of the selected best features. Although feature 19 and 41 are the two most informative features in the 45 statistics, their bivariate probability dependence in February and March shows that they are both not sufficient for class separation, as shown in Fig.6.6 (the likelihood of class 1, 2 and 3 almost overlap). Another reason is the estimation of the priori probability of each class, which is estimated by the ratio between the number of feature vectors in each class and the total of training set. Looking at the dataset obtained in Feb. and Mar., the priori probabilities for under-, proper- and over-estimation classes are roughly 0.28, 0.54, 0.18 and 0.15, 0.56, 0.29 respectively. It indicates that the priori probability of each class varies with month. This variation will further change the estimation of the posteriori, and also change the final classification. Error will occur when using one month's data to classify DS collected in another month.

Theoretically speaking, as long as the probabilistic distance of the features (the values of the features of different classes have different probabilistic distributions) is big, and the sample feature vectors used for the training data are large enough to reflect the real priori and posteriori distribution, the classification performance should be very good. In this sense, if a whole year of data is used as the training dataset to classify the data in another year, the problem of the monthly-varying priori probabilities can be minimized to some extent. Whether 3 month, half a year, 1 year, 2 years or more are sufficient for the training size, is not clear so far and can be explored in future work. Nevertheless, the probabilistic distance of the feature space will not be improved much by enlarging the training size, because it is determined by the nature of the features. So the feature space is still the bottleneck in this Bayes approach.



(a)



(b)

Fig.6.6. Bivariate KDE $p([x_{19}, x_{41}]^T \mid \omega_i)$. The X and Y axis are the values of the 19th and 41st features. The vertical axis represents the probability. (a): data from Mar, 2005; (b): data from Feb, 2005. In each plot, the bivariate probability functions of class 1, 2, 3 are found to be mostly overlapping. The different dark square areas in the XY plane show the range of the distribution of each class. The colorbar shows the frequency.

6.4 Statistical pattern recognition based on the geometric approach

Since the performance of the density-based approach is not satisfactory, this section introduces the geometric, or distance-based, approach. By distance, I mean the Mahalanobis distance between feature vectors in the feature space, which involves the idea of covariance. A feature vector is a point in the feature space, with each feature as a coordinate. The ability to classify feature vectors by the geometric approach relies on the assumption that feature vectors belonging to different classes occupy compact subspaces and do not overlap in the feature space. The more distant the classes are from each other, the higher the chance of successful classification. Therefore, the goal of this approach is to find the feature space where the classes are maximally separated. In the same order as last

section, preprocessing, feature extraction and learning will be explained block by block.

6.4.1 Feature extraction

New feature vectors with 100 features are generated by the preprocessing block, which is the second order region beside the sfs of a DS, i.e. db $sfs-49$ to $sfs+50$. The 45 statistics used in density-based approach are not used as the input feature vector because the feature space selected from it is not very informative for the classification, and the features are weakly correlated while feature extraction is better applied to some strongly correlated features. Six feature extraction methods (see Table 6.1) are examined below in terms of their performance for discrimination in the extracted feature space. In this subsection, the goal is to find the best feature extraction method. By 'best', I mean two things: the feature vectors of different classes in the feature space overlap as little as possible; and the feature vectors in the same class fall as close together as possible.

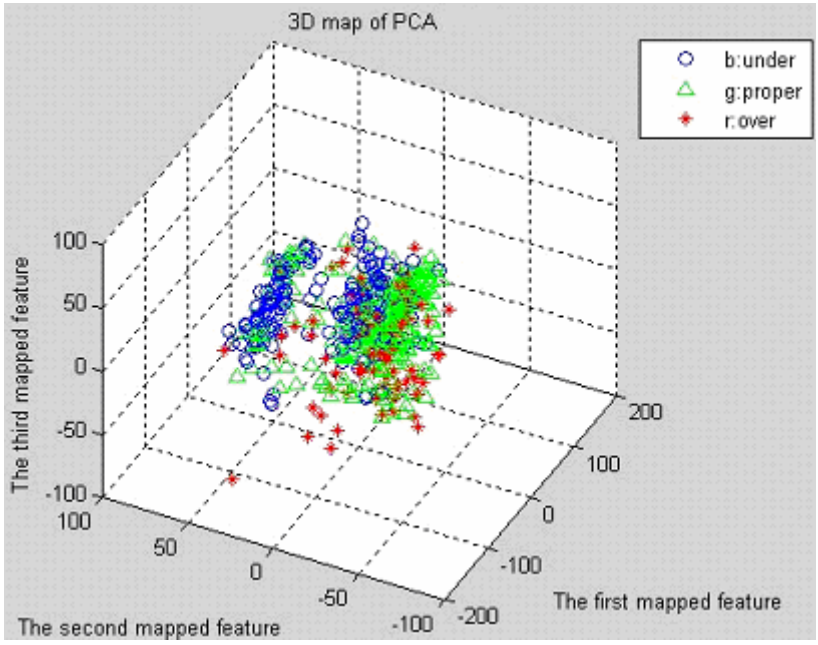
Principal Component Analysis and Factor Analysis

The definition, derivation and interpretation of principal component analysis (PCA) and factor analysis (FA) have been elaborated by Jolliffe (2005). There are two issues to address when using PCA and FA for feature extraction and classification. Firstly, using the correlation matrix or covariance matrix; secondly, the number of extracted Principal Components (PCs)/Factors.

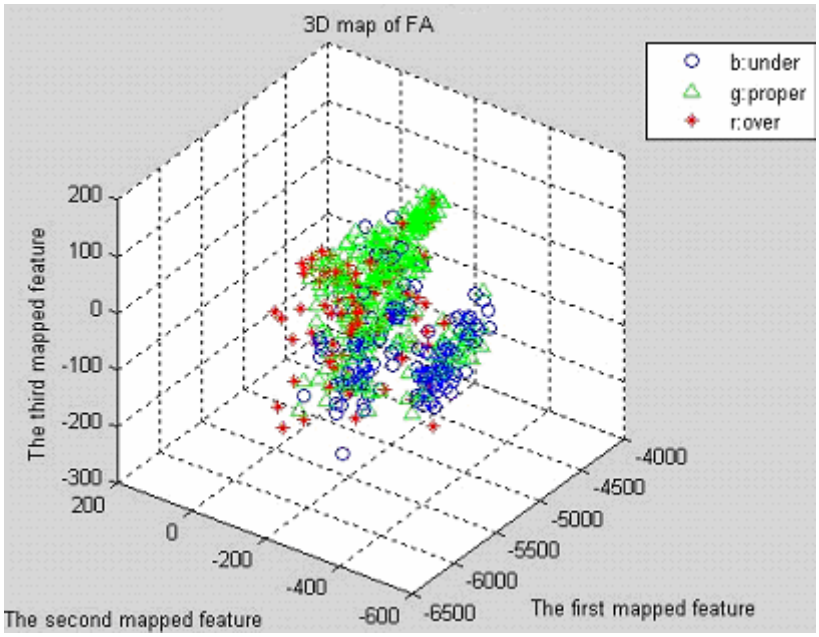
Using the correlation or covariance matrix depends on whether all measurements are made in the same units. For example, if the feature vector includes 45 statistics, which does not have the same scale of values, the correlation matrix will be preferred because it is desired to treat all variables on an equal footing while the covariance matrix gives greater coefficients to larger measurements, and smaller coefficients to smaller measurements. Here, the feature vector is a section of DS with the same value scale. Either matrix might be appropriate. The covariance matrix is chosen because it preserves the information on the standard deviation of the Doppler variables. Regarding the number of extracted features by PCA and FA, as said before, it is better to be lower than or equal to 3 for visualization. For PCA, if the number is set as 3, the first three principal components will be selected, while the remaining PCs will be discarded. To make sure that the discarded PCs are relatively insignificant, the variance distribution of all PCs is calculated. It is found that the first one, two, three, and four PCs account for 58%, 72%, 80%, and 83% respectively of the total variance of the feature vector. The variance of the fourth PC is smaller than one tenth of that of the first PC, while the second and third PC is around one fourth and one

seventh of that. It seems reasonable to select the first two or three PCs as extracted features. For FA, the feature vector with 100 features can be assumed to have 2, 3, or higher number of common factors. After factor analysis, the common factors are viewed as the extracted best features. It is found that the FA feature vectors assuming there are higher number of common factors are not separated better than assuming there are 2 or 3 common factors. The new feature vectors generated by PCA and FA are both presented in their 2-D and 3-D feature space map, which are compared in terms of their discrimination ability. The one with better class separation will be chosen for the feature space.

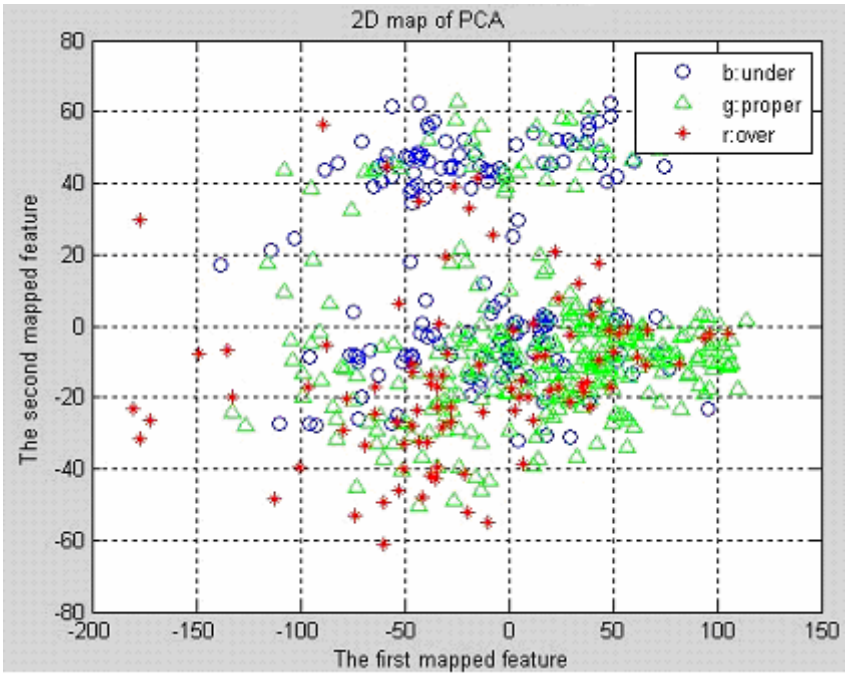
The 2-D and 3-D map of feature spaces generated by PCA and FA are shown in Fig.6.7. In the geometric approach, a feature vector is a point in the feature space. After PCA and FA are applied, each feature vector has two or three features, i.e. each point has two or three coordinates. The feature vectors belonging to class 1, 2, 3 are drawn with a blue circle, green triangle and red star respectively. From Fig.6.7, it can be seen that the feature vectors of class 1, 2, and 3 are still mixed in the FA map and the PCA map. An interesting behaviour is that the feature vectors are split into 2 clusters, each containing feature vectors from all 3 classes. It has been found that this is due to the position of the superior first order peak on either positive or negative Doppler frequency side, i.e. waves travelling toward or away from the radar respectively.



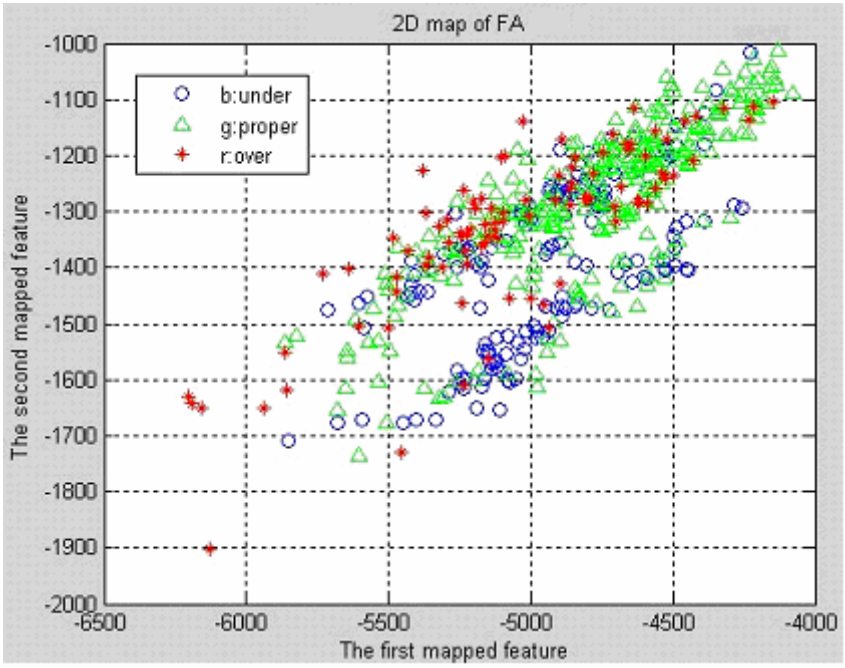
(a)


3D map of FA
b:under
g:proper
r:over
The third mapped feature
200
100
0
-100
-200
-300
200
0
-200
-400
-600
The second mapped feature
-4000
-4500
-5000
-5500
-6000
-6500
The first mapped feature

(b)


2D map of PCA
b:under
g:proper
r:over
The second mapped feature
80
60
40
20
0
-20
-40
-60
-80
-200
-150
-100
-50
0
50
100
150
The first mapped feature

(c)



(d)

Fig.6.7 3D and 2D map of PCA and FA based on data preprocessed by option 2. (a): 3D map of PCA; (b): 3D map of FA; (c): 2D map of PCA; (d): 2D map of FA. The under-, proper-, over- classes are plotted in blue, green and red

Canonical Correlation Analysis

The main derivation of canonical correlation analysis (CCA) is given by Preisendorfer and Mobley (Preisendorfer et al. 1988). In CCA, the variables are in two groups, and relationships are sought between these groups of variables. Suppose that $\underline{x}$, $\underline{y}$ are two feature vectors of d_1 and d_2 features respectively (in this problem, $\underline{x}$ refers to the Doppler section with 100 features, and $\underline{y}$ refers to the class labels of each $\underline{x}$). The objective of CCA is to find successively d pairs $\{\underline{a}_{k1}^T\underline{x}, \underline{a}_{k2}^T\underline{y}\}$ of linear functions of $\underline{x}$, $\underline{y}$, for $k = 1, 2, ..., d$, $d = \min[d_1, d_2]$. These pairs are called canonical variables, such that the correlation between $\underline{a}_{k1}^T\underline{x}$ and $\underline{a}_{k2}^T\underline{y}$ is maximized, subject to $\underline{a}_{k1}^T\underline{x}$ and $\underline{a}_{k2}^T\underline{y}$ are being both uncorrelated with previous derived canonical variables, $\underline{a}_{j1}^T\underline{x}$, $\underline{a}_{j2}^T\underline{y}$, j =1,2, . . . , (k - 1).

CCA can be used, considering $\underline{x}$ as a set of outputs of a system and $\underline{y}$ a set of

predictors or references, to find the optimal weightings of such a system. In this application the canonical features are linear functions of the original Doppler features, chosen to maximize the separation between classes. Specifically, the first extracted feature of CCA is the linear combination of 100-bin DS that has the maximum class separation ability (each Doppler bin is considered as a variable). The second extracted feature of CCA has the second maximum separation subject to it being orthogonal to the first feature, and so on.

The calculation of canonical variables (transformed features) of CCA are similar to that of principal components (PCs) in PCA in that they both look for a linear combination of the original variables. The difference between them is PCA looks for the combination of the original variables that has the largest possible variation; while CCA looks for the linear combination that has the largest separation between groups. There are two things to make clear in CCA. Firstly, the correlation rather than covariance is maximized. Second, the derived variables are uncorrelated.

A graphical representation of the feature space transformed by CCA is shown in Fig.6.8. The under-, proper-, over-estimation classes are plotted in blue, green and red colour, while the symbols for Feb data are circle, triangle and star, shown in plot(b). Note that the first canonical variable, $c1$, separates the 'under-estimation' class (which have high values of $c1$) from the other two classes. The second canonical variable, c2, reveals some separation between the 'proper-estimation' and 'over-estimation'.

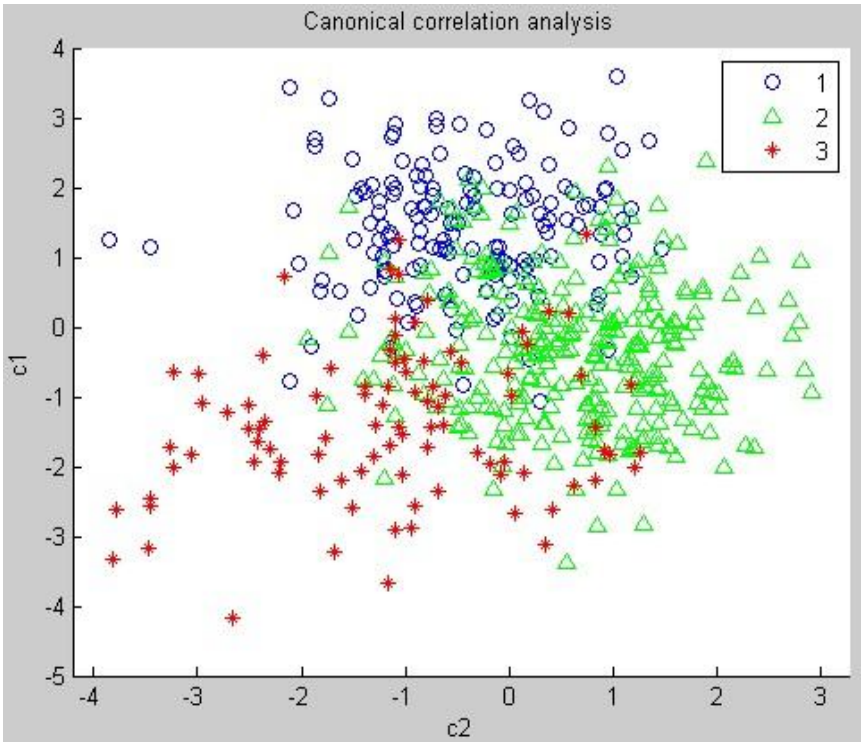


Fig. 6.8 CCA map from data obtained in Feb. The under-, proper-, over-classes are plotted in blue, green and red, while the symbols for data are circle, triangle and star.

Here, an important concept in the pattern recognition has to be introduced, i.e. the generalization ability. The low overlapping rate of separation performance shown in Fig.6.8 implied that the classification performance is optimized in the training set. However, this doesn't always result in a comparably good classification performance in the testing set. The reason is the model (the term 'model' here refers to the means of each class and the within-class, between-class covariance matrix of a dataset) is likely to be different from that in the testing data, if the training size is not large enough.

In this case, a small test is inserted here to check the generality of the CCA method. The training and testing procedures by CCA on February (the procedures for March are the same) data are explained as follows:

1. The means of c1 and c2 for class 1, 2, 3, which are calculated by half-month data and whole-month data respectively, are examined. It is noticed that the values of means don't vary much (see in Table 6.4). This is a necessary condition for good generality in CCA (CCA mapping matrix is calculated based on feature vectors centred at zero by subtracting their means).
2. The first 300 feature vectors (half month data), and the first 500 feature vectors in February are selected to train the model respectively, i.e. estimating mean, within- and between-class covariance.
3. The model is used to map the remaining feature vectors (testing data) into the CCA feature space.
4. Testing feature vectors in CCA feature space are labelled by buoy information and plotted. Feature vectors from different classes are checked to see if they are separated.

The results are shown in Fig. 6.9(a)(b). There are 538 feature vectors in February in total. Plot (a) shows the training feature vectors in the CCA map, and plot (b) shows the testing feature vectors grouped by the model in plot (a). The discrimination is not as satisfactory as in Fig.6.8. As said before, the half-month training set is not large enough to accurately estimate the model for a whole month. Then the training set is enlarged from 300 to 500 feature vectors and the remaining 38 feature vectors are mapped into the CCA feature space. The result is shown in plot (c)(d). It can be seen that the testing feature vectors belonging to proper and overestimation classes are roughly separated in the CCA map, but the separation of underestimation class is unconvincing due to only 1 feature vector in underestimation class in the testing set.

Table 6.4 Means of c1 and c2 for Class 1, 2, 3. c1 and c2 are the two features after feature extraction by CCA; Class 1,2, 3 are under-, proper-, and over- estimation respectively. Means of each feature in each class is calculated using a whole month of data (538 feature vectors) and the first half month data (269 feature vectors) respectively.

Month	February			
Class	Whole month		Half month	
	c1	c2	c1	c2
Class 1	1.8692	-0.8007	1.9725	-0.7663
Class 2	0.0685	0.5332	-0.0561	0.6952
Class 3	-1.1054	-0.6374	-1.4310	-0.8892

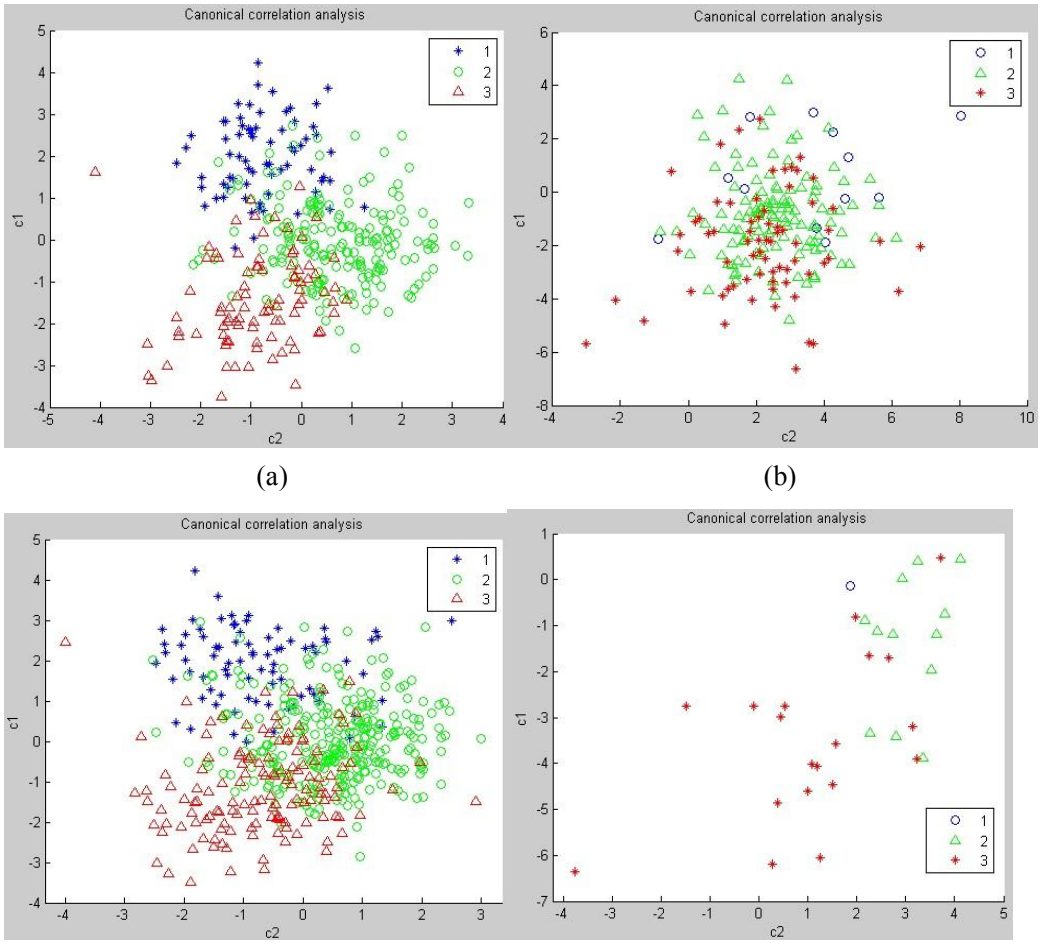
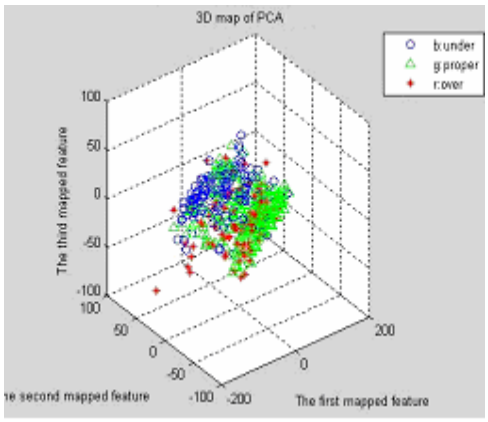


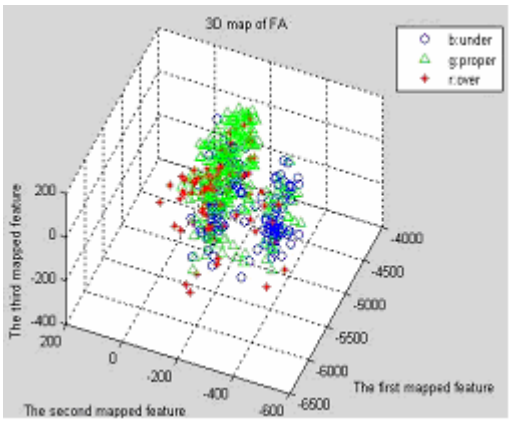
Fig.6.9 Training and testing procedures of CCA. Class 1, 2, and 3 represent under- (blue), proper- (green), and over-estimation (red) respectively. (a) the CCA feature space of the first 300 feature vectors in Feb, which are used together with the buoy labels of class membership to train the model, i.e. mean, between-class variance, within-class variance, for the feature extraction of the testing feature vectors in (b). (b) the CCA feature space of the remaining 238 feature vectors in Feb. using the model calculated using the 300-feature vector training set. (c) the same as (a) but using the first 500 feature vectors to train the model; (d) the CCA feature space of the remaining 38 testing feature vectors using the model calculated in the 500-feature vector training set.

The last three feature extraction methods used here are generalized eigenvalue decomposition (GED), locality preserving projections (LPP), and generalized discriminant analysis (GDA). GED has been applied for clutter mitigation in Chapter 5.3.3. Since GDA is a dimension expanding method, it is better applied to the 2-feature feature vectors when they are not linearly separable. For these three methods, the algorithms are fully developed by many researchers and

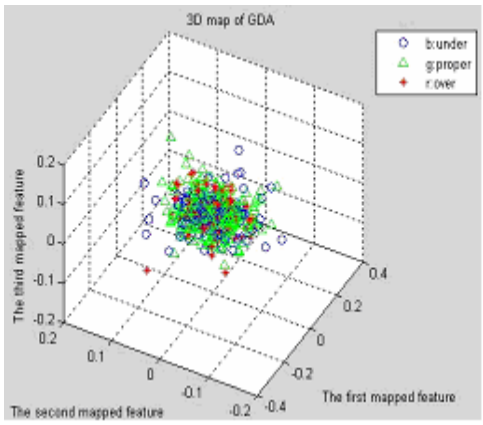
provided (LPP (Gui et al. 2009), GDA (Baudat and Anouar 2000)) in the Matlab toolbox, so they are not discussed further here. All six feature extraction methods listed in Table 6.1 have been discussed. Their performances are shown in Fig. 6.10 (a)-(f). In addition, some of these methods can be used in combination with others in a cascade way as shown in Fig.6.11.



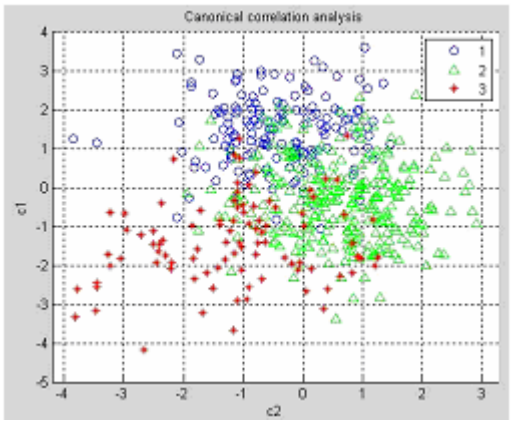
(a)



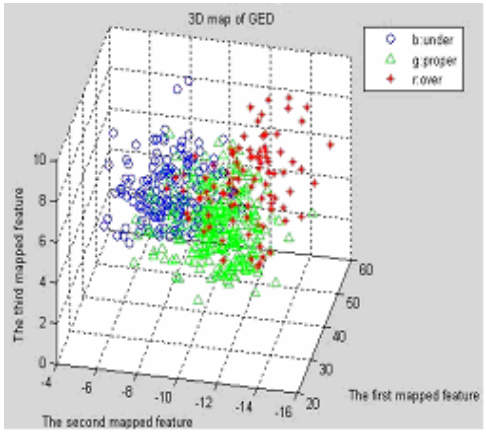
(b)



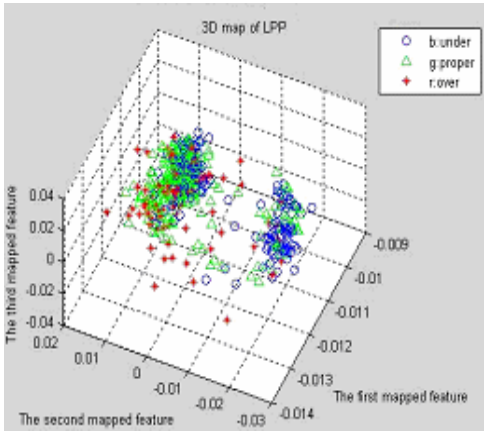
(c)



(d)



(e)



(f)

Fig.6.10(a)-(f) 3D map of feature space based on six independent feature extraction methods, listed in Table 6.1. The under-, proper-, over-estimation classes are plotted in blue circle, green triangle and red star. (a) PCA ; (b) FA; (c) GDA; (d) CCA; (e) GED; (f) LPP.

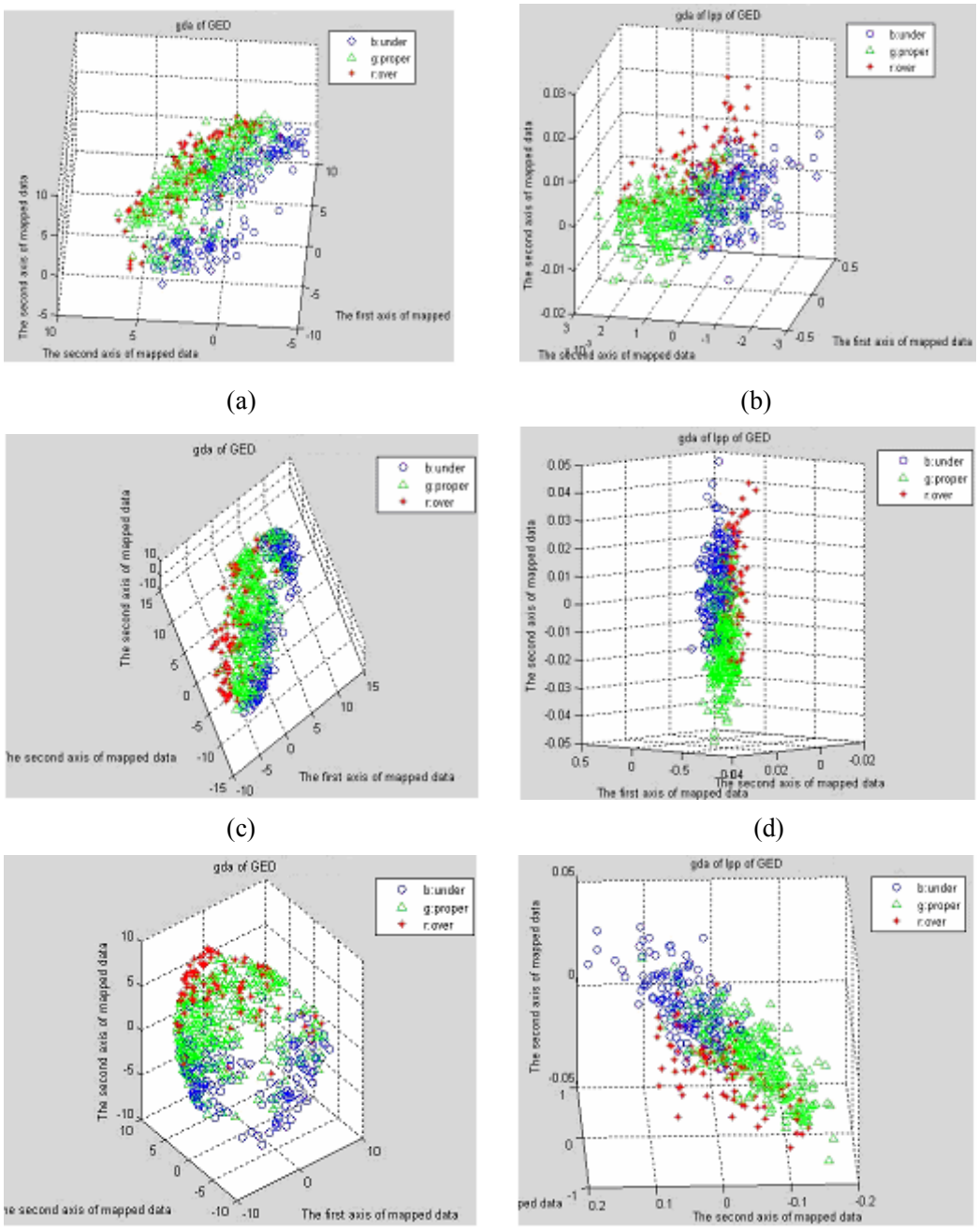


Fig.6.11(a)-(f) 3D map of feature space based on six combinations of feature extraction methods. The under-, proper-, over- classes are plotted in blue circle, green triangle and red star. GDA uses a Gaussian kernel function, the parameter of standard deviation of which is set by 0.5, 0.7, 1 respectively. (a) GDA of GED (std=0.5); (b) GDA of LPP of GED (std=0.5); (c) GDA of GED (std=0.7); (d) GDA of LPP of GED (std=0.7); (e) GDA of GED (std=1); (f) GDA of LPP of GED (std=1)

Comparing the performance of the six feature extraction methods and the BIF feature selection method, several statements could be made here.

(1) The BIF method (which is explained in Table 6.3 of section 6.3.1) could be used for both density-based and geometric approach, although the 3D map of its feature space is not shown here. Using BIF method, the most informative three features or Doppler bins for 'under-estimation' class are Doppler bins 'sfs-18', 'sfs-17', 'sfs-19'. For 'proper-estimation' and 'over-estimation' classes are Doppler bins 'sfs-26', 'sfs-25', 'sfs-24' and 'sfs-34', 'sfs-35', 'sfs-36' respectively. It indicates that a few consecutive Doppler bins within the second order continuum might be more key regions related to the wave estimation. This idea is checked using the BIF method on March data. The best features related to 'under-' ('sfs+1', 'sfs-49', 'sfs'), 'proper-' ('sfs-23', 'sfs+17', 'sfs-24'), and 'over-' ('sfs+1', 'sfs', 'sfs+38') classes are found to be different from those in February. It appears that the key region of the DS related to the wave estimation may also change with the oceanographic conditions of different months. The histogram of waveheights in Feb (642=28*24) and March (744=31*24) in 2005 are shown in Fig.6.12, which indicates a different distribution of waveheights. The relationship between certain Doppler bins and the waveheight estimate is left for further exploration.

(2) PCA and FA don't work well for this classification job, because the PCs accounting for the most variance don't relate to over or under estimation.

(3) The optimal independent feature extraction method is CCA, by which a majority of the three classes can be separated in a 2-dimensional feature space. On top of that, the feature extraction method of GED can be used with LPP and GDA in a chain. However, this chain method is not recommended because the computational load is much higher than CCA.

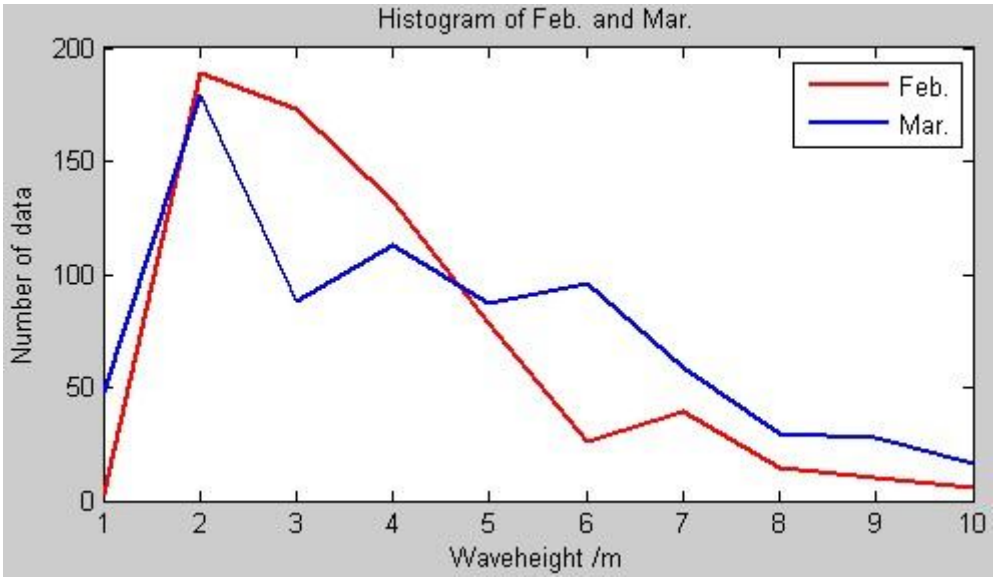


Fig.6.12 Histogram of waveheight in Feb (red) and March (blue).

6.5 Classifier design for pattern recognition

Given that specific clutter mitigation schemes have been proposed in chapter 6 for improving the accuracy of wave estimation, either for the overestimation or underestimation case, the main goal of this section is to classify the DS into under-, proper- and over- estimation classes. In the selected geometric approach, the preparation work including 'preprocessing' and 'feature selection/extraction' before classification has been explained above, and results in the extraction of the 'best' features. It has been shown above that CCA feature space is the best one. This section will focus on the last 'training' block, where classifiers are used to partition the feature space into subspaces occupied by feature vectors of different classes and assign 'unknown' (testing) feature vectors with accurate class labels. The performances of various classification techniques are compared and the optimal classifier is found.

Different classification methods have been tested in order to choose the one with the highest CCR. There are many perspectives from which to discuss classification techniques, e.g. supervised or unsupervised, statistical or neural network. Under the category of supervised classification, there are methods of TREE, SVM, KNN, DA. Under unsupervised branch, there are KMEANS and SOM. Under the category of neural network, there are SVM, SOM, and TREE. Under the statistical category, there is a density-based classifier whose classification is based on Bayes decision rule (see section 6.3.2), and the distance-based classifier, using feature extraction methods of PCA, FA, CCA, GDA, GED, LPP. Some of these methods are required to be used in a chain rather than independently. Sometimes the pattern recognition problem requires a combination of them.

Before the presentation of results of the application of these classifiers, several common issues are introduced here, which are adopted for all the classification methods. The testing data used for the comparison of various classifiers are 538 CCA feature vectors. Note that, these data are calculated using a whole month model. Hence, the best degree of separation of the testing feature vectors is assumed (this is because if the feature vectors are poorly separated, the performance of various classifiers on this feature space might all be poor and it would be difficult to compare and select a best one). This dataset is then split into training and testing sets by a certain proportion. The training and testing sets should be independent of each other, i.e. the two sets have no intersection. This requirement of independence is often ignored in practice but taken into consideration in this section. If the training set is small, the resulting classifier will not be very robust (as was proved in CCA testing). On the other hand, if the testing set is small, the confidence in the estimated CCR will be low. In this regard, different proportions are chosen, i.e. 0.2%, 1%, 2%, 7%, 19% and 56% of the dataset are the testing subset. The cross validation method (Schaffer 1993; Kohavi 1995) called 'Leave-M-out' is selected to split the available samples

(randomly choosing M out of the whole dataset to be the testing set and the rest to be the training set). The robustness of each classifier is validated by a 100-times iteration of the classification process.

A brief description and comparison of these methods is summarized in Table 6.5 followed by a more detailed discussion of each. Note that detailed derivations of those theories are not presented here but their references are given in Table 6.5. The programs for realizing these algorithms are adapted based on the Matlab toolbox.

Table 6.5 A comparison of various classification methods

Method	Property	Comments
KMEANS (G.A.F.Seber 1984)	Assign samples to the nearest class mean	No training needed; supervised best feature space construction needed; fast testing; metric dependent
SOM (Kohonen et al. 2009)	An artificial neuron network	No training needed; supervised best feature space needed; slow testing;
Tree (Breiman and Friedman 1988)	Finds a set of thresholds for a sequence of features	Iterative training procedure; training sensitive; fast testing;
DA (Mika et al. 1999)	Classifier using Malhoubis distance	Training needed; supervised best feature space needed; fast testing
KNN (Devijver and Kittler 1982)	Find the nearest k neighbors	Fast testing; supervised best feature space needed; for large dataset
SVM (Cristianini and Shawe-Taylor 2006)	Maximizes the margin between the classes by selecting a minimum number of support vector	Training needed; supervised best feature construction is optional; moderate testing speed; nonlinear; overtraining insensitive;

KMEANS

KMEANS clustering is very popular among methods in clustering analysis (Desarbo 1987). Cluster analysis deals with datasets in which the feature vectors are to be clustered into groups，i.e. those feature vectors falling close together are considered to belong to the same group. Since no class labels are known a priori, the clustered groups might not be the real classes. This method is often used to study the inherent structure of a dataset. Strictly speaking, KMEANS is half-unsupervised because it requires the knowledge of cluster centroid of each group. KMEANS clustering uses an iterative algorithm that assigns feature vectors to groups so that the sum of distances from each feature vector to its

cluster centroid, over all groups, is a minimum.

The clustering performance of KMEANS is tested on Fisher's iris data and HFSWR radar DS in a comparison. In 1920's, botanists collected measurements on the sepal length, sepal width, petal length, and petal width of 150 iris specimens, 50 from each of three species. The measurements are well known as Fisher's iris data (Fisher 1936). For both datasets, class labels are available, e.g., each sample in Fisher's data comes from a known species and the in situ measurements of the significant waveheight from wave buoys are obtained. The experiment goes like this:

1. Fisher's dataset uses feature selection for a 3D feature space construction, because Fisher showed that only three features are independent. Feature 'petal width' and 'petal length' are highly correlated(Fisher 1936).
2. Radar's dataset uses PCA feature extraction for a 3D feature space construction.
3. Both datasets are clustered by KMEANS without using class labels.
4. Both datasets are labeled by species and buoy information.
5. For each dataset, KMEANS clusters, termed artificial groups, were compared with labeled clusters, termed genuine groups, in their graphical representation of feature space, to see if they match with each other or not, and if feature vectors in different genuine groups are separated from each other.

The 3D representation of artificial groups and genuine groups based on Fisher's and Radar's datasets are shown in Fig.6.13(a)(b) respectively. Different colors and symbols are used to represent different classes. In Fig.6.13(a), the artificial groups almost agree with the genuine groups. In addition, three genuine groups are well separated due to the distinctive distributions of the 'petal length' feature. On the contrary, in Fig.6.14(b), the three KMEANS clusters not only do not agree with their genuine groups but also feature vectors of different genuine groups are mixed together. Looking at these two datasets, the features of Fisher's data are directly related to the true species. However, features extracted from DS measurements are related to the estimation of waveheights in a nonlinear and complex way, affected by hydrodynamic and electromagnetic mechanisms.

It is still early to draw the conclusion that KMEANS has no effect on clustering DS properly, because the clustering performance of KMEANS is also determined by the construction of its feature space. This has been proved by the Fisher's iris data. The genuine groups are well-clustered by KMEANS because the feature 'petal length' has different distribution of values for different classes. Therefore, the key to successful KMEANS clustering on DS is to find the optimal features. However, KMEANS, as a half-unsupervised method, is still worse than other supervised classifiers in classification, even based on the CCA feature space.

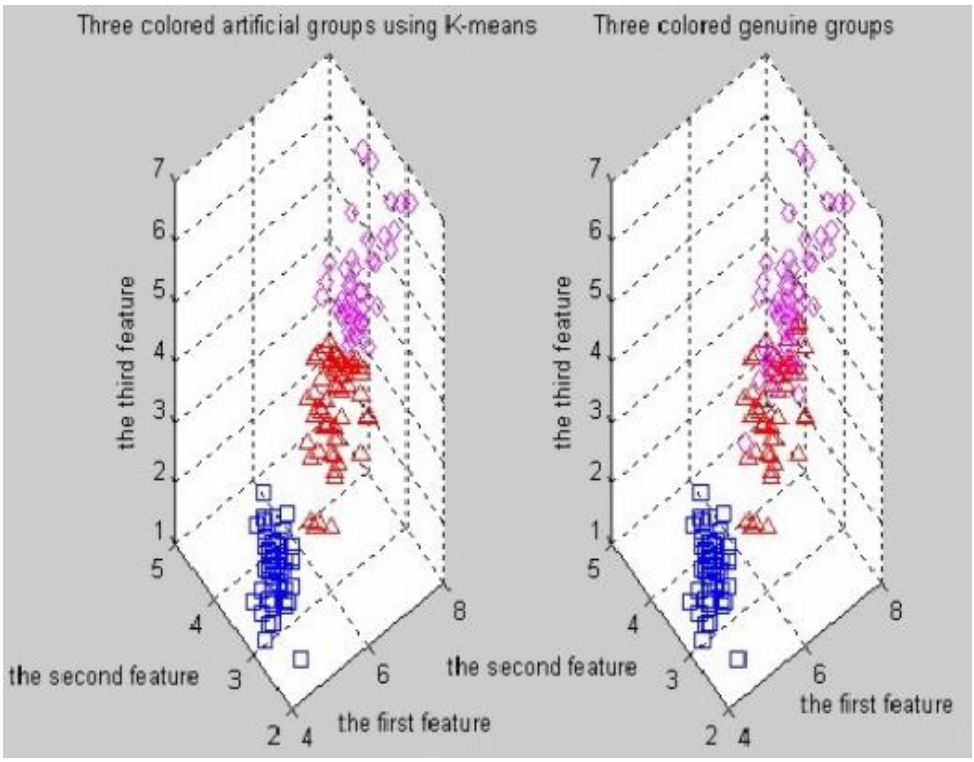


Fig. 6.13(a) 3D feature space of Fisher's iris data. Left: clustered by KMEANS; Right: clustered by labels. The first, second, and third feature axis are sepal length, sepal width, petal length respectively.

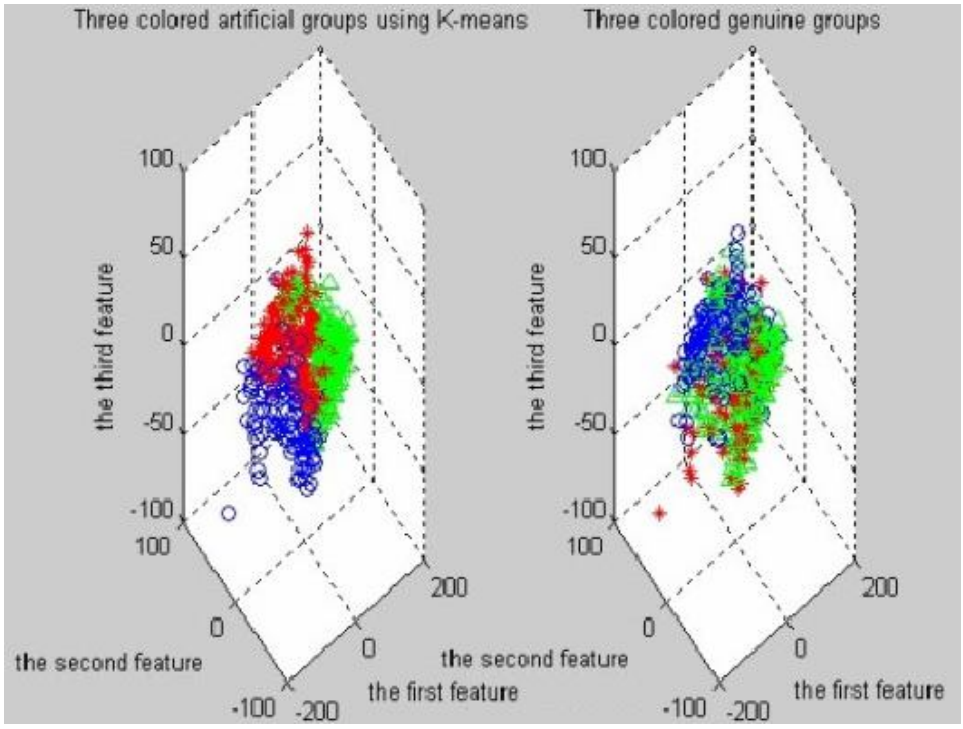


Fig. 6.13(b) 3D feature space of Radar's dataset. Left: clustered by KMEANS; Right: clustered by labels. The first-, second-, and third feature axis are the first, second, and third PC respectively.

Self Organizing Map Analysis

Self Organizing Map (SOM) is one of the most popular topics in the artificial neural network field. Neural networks generally can learn to detect regularities and correlations in their input and adjust their output responses to that input accordingly. SOM operates in two modes: training and mapping (Kohonen 1991). Training use input feature vectors to build the map. Then mapping automatically classifies a new input feature vector. A SOM comprises components called nodes. A position in the map space and a weight vector of the same dimension as the input data vector are associated with each node (e.g. the weight vector is 2-dimensional as in the CCA feature space). The nodes are usually arranged by a regular spacing in a hexagonal or rectangular grid. The SOM maps a higher dimensional input space to a two-dimensional map space. Suppose there is a high dimensional input vector $\underline{x}$, the procedure for SOM mapping is to find the node in the map with the closest weight vector to $\underline{x}$, and to assign the position coordinates of this node to $\underline{x}$.

In this application, the nodes of SOM can be considered as representative vectors of classes. Since the input radar data in the CCA feature space are clustered into three groups to some extent, the number of prototype nodes is set 3. The procedures of using SOM for classification are given below.

1. Initialize a SOM network by setting the number of prototype nodes as 3.
2. Mapping the data from CCA feature vectors into SOM feature vectors to find the weight vectors of three prototype nodes.
3. Calculate the distances of each feature vector to the three nodes to find the closest prototype node.
4. Label the feature vector by the membership of that closest prototype node

The classification performance of SOM on Radar feature vectors is shown below. Although SOM is an unsupervised learning method, its CCR is 0.747 based on CCA feature vectors, which is a well-separated feature space. Fig. 6.15 top shows a representation of the labeled CCA feature space before mapping. Class 1, 2, 3, i.e. under-, proper-, over-estimation are plotted in blue circle, green triangle, and red star. Fig.6.15 bottom shows the classified feature vectors labeled by SOM with the symbol ‘triangle’ and feature vectors labeled by buoy information with symbol ‘star’. A good match between SOM classification and real classes has been found, although there is a 25.3% misclassification

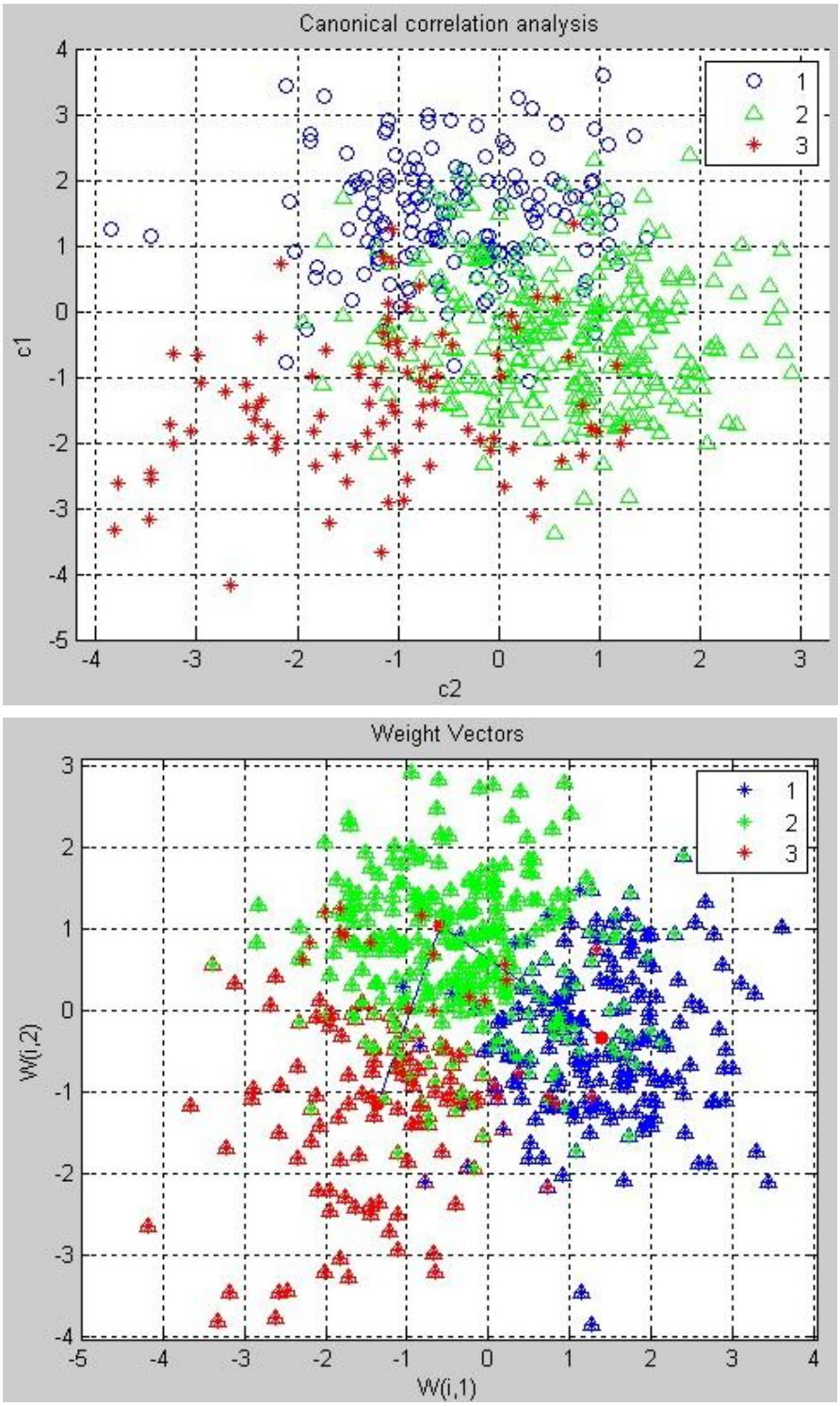


Fig 6.14. Top: CCA feature space; bottom: SOM weight space. In the top plot, feature vectors of the three classes are separated to a large extent. The plot below shows that the 'triangles', labelled by the SOM distances, match with the 'stars', labelled by buoy information.

Tree

Tree, also called decision tree, is a supervised and hierarchical classifier,

investigating classes in data by creating a decision tree. The decision tree is capable of finding a chain of threshold rules for defining each class, For example, a Doppler feature vector consisting of a number of features can be classified by a set of threshold rules, such as "if the first feature is less than 5.45, classify the data as over-estimation, otherwise continue testing other features". A big advantage of the decision tree is it does not require any assumptions about the distribution of the measurements in each group, i.e. the likelihood. The decision tree is first trained using training data and class labels to find threshold 'rules'. Note that these rules are not independent but linked together as leaf nodes of a tree. Next, a testing feature vector is checked by the rules of the tree from the top node. If the feature in the feature vector satisfies the rule, the tree classifier takes the left path, and if not it takes the right path. This checking procedure continues until it reaches a terminal node, which is labeled by a class name. The tree is trained based on the radar data preprocessed by option 1, i.e. 45 statistics, because the method is better applied to a feature vector with independent features. A trained tree classifier is shown in Fig 6.15.

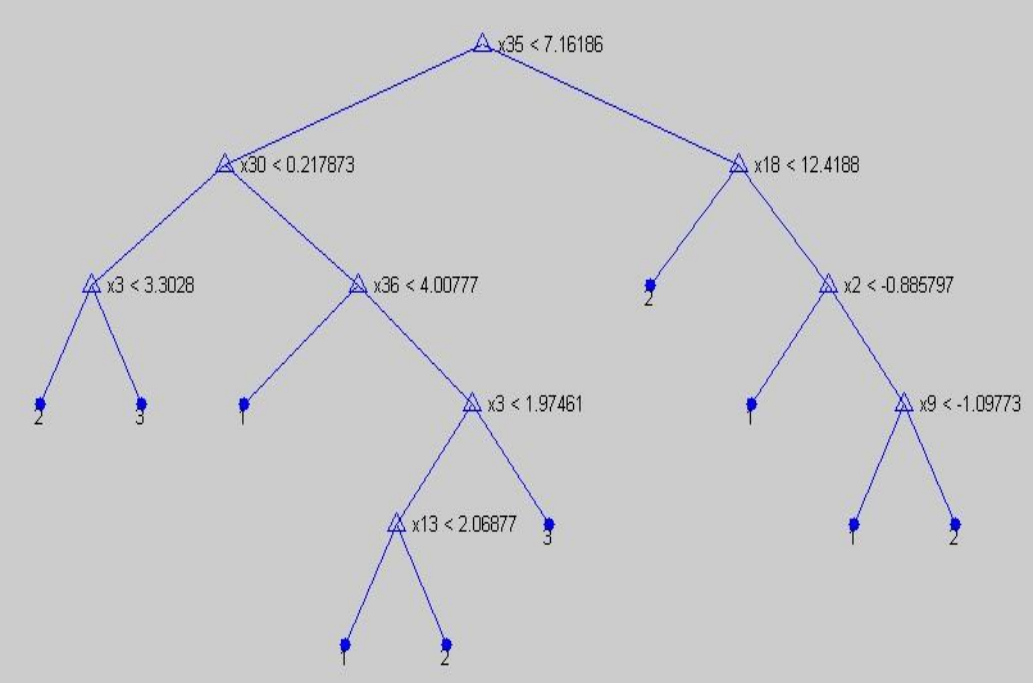


Fig.6.15 Tree classifier with threshold rules. The symbol 'triangle' represents a leaf node. Class '1, 2, 3' represent 'under-', 'proper-', and 'over-' estimation classes respectively. 'x41' represents the 41st statistic.

The classification performance of a tree classifier varies from 0.57 to 0.80, see Table 6.9. The number of feature vectors in the whole dataset is 538. The optimistic performance of a tree classifier is 80% by using 537 as training feature vectors. CCR decreases as the size of training set decreases. In conclusion, with a CCR of 56% for a half-size tree, the tree classifier is poor in predicting the quality of a DS. In addition, this supervised classifier is unable to classify radar data at another time and location because these hard-threshold rules are

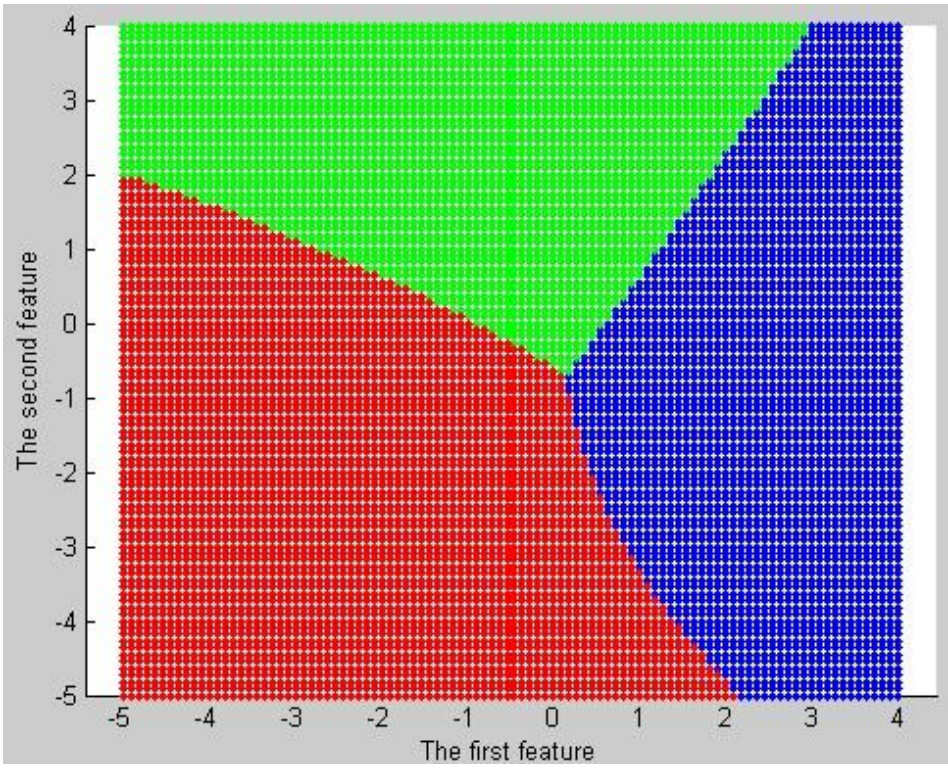
dataset-dependent.

Table 6.6 Tree classification verified by 'Leave-M-out' cross validation

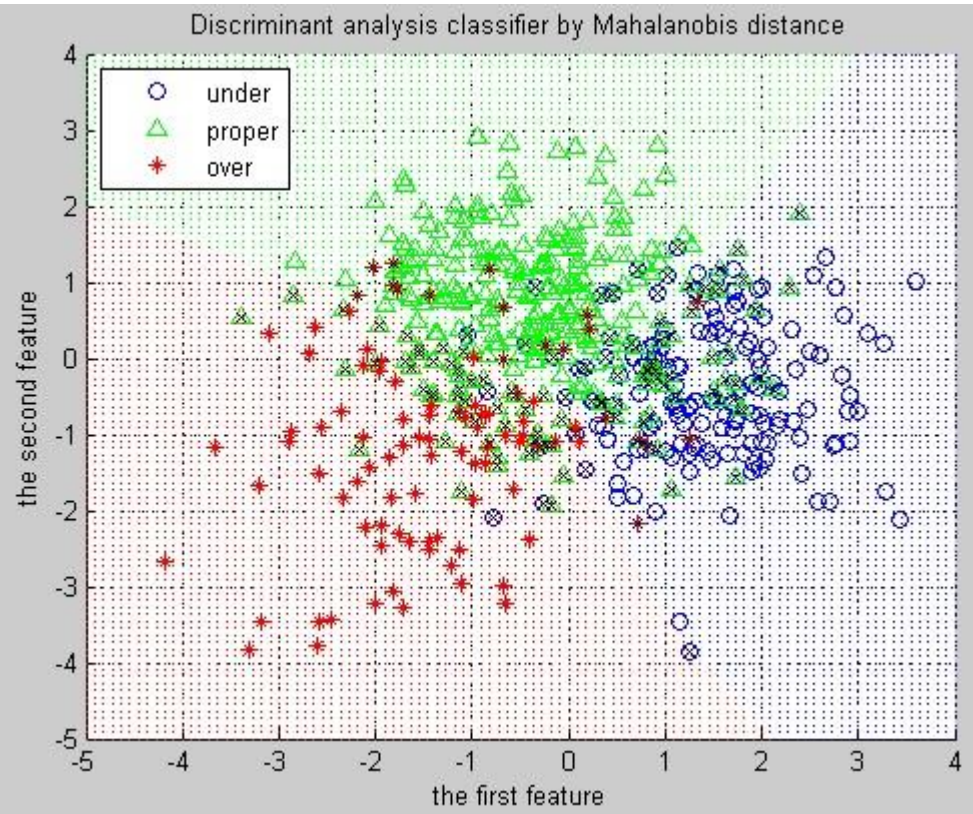
Training dataset number	Testing dataset number		CCR
358	300	(56%)	57%
438	100	(19%)	59%
500	38	(7%)	64%
528	10	(2%)	68%
533	5	(1%)	72%
537	1	(0.2%)	80%

Discriminant Analysis

The decision rule used in the discriminant analysis classifier is that each observation can be assigned to the class which has the minimum Mahalanobis (Mahalanobis 1936) distance $D_{mah} = (\underline{x} - \underline{\mu}_y)^T S^{-1} (\underline{x} - \underline{\mu}_y)$, where $\underline{x}$ is a feature vector in the testing set; $\underline{\mu}_y$ and S are the mean and covariance of training data of each class.



(a)



(b)

Fig 6.16. DA classification. (a) grid map; (b) classification map. Class 1, 2, 3 are plotted by blue-circle, green-triangle and red-star. The discriminant function uses Mahalanobis distances with stratified covariance estimates.

Fig 6.16 shows the class separation by the discriminant classifier. Plot (a) shows that the meshed grids map (the 2D feature space is digitized into a 10000*10000 grids map) is classified into three partitions by the training feature vectors. The training feature vectors are points plotted with blue circle, green triangle and red star for under, proper, and over estimation classes respectively in Fig.6.16(b). Here, each feature vector is considered as a geometrical point with two coordinates (features). After training, the grid map is used as a look-up table. During testing, the coordinates of a testing feature vector are calculated, so that the point is positioned in the grid map. The class membership is then obtained according to the 'look-up-table'. Those misclassified points are plotted by black cross symbol in Fig.6.16(b). The 'Leave-M-out' cross validation method is used to validate the classification performance of DA, which is listed in Table 6.10. Each percentage in the third column is an average of 100-time iteration results. It has been found that the CCR is rather stable for both 537 training patters and 358 training feature vectors.

Table 6.7 DA classification verified by 'Leave-M-out' cross validation.

Training dataset number	Testing dataset number: M		CCR
358	300	(56%)	1: 75.5%; 2: 75.0%; 3: 74.7%
438	100	(19%)	1: 75.5%; 2: 75.4%; 3: 74.3%
500	38	(7%)	1: 74.5%; 2: 75.8%; 3: 74.0%
528	10	(2%)	1: 72.9%; 2: 74.4%; 3: 74.8%
533	5	(1%)	1: 71.0%; 2: 77.8%; 3: 75.2%
537	1	(0.2%)	1: 67.0%; 2: 72.0%; 3: 72.0%

K Nearest Neighbor

In this problem of DS classification, a solution is to build up a collection of correctly classified feature vectors, and to classify each new feature vector using the evidence of nearby feature vectors. Fix and Hodges (1951) introduced such a non-parametric procedure, known as the k-nearest neighbor (KNN) rule. The definition of KNN rule is: an unclassified testing feature vector is assigned to the class most common amongst its K nearest neighbors (K is a positive and typically small integer). Cover and Hart (1967) provided a statistical justification of this procedure by showing that if the number N of training feature vectors and K both go to infinity in such a manner that $\lim_{K\to\infty, N\to\infty} K/N = 0$, the CCR of the KNN rule will approach the optimal Bayes CCR. However, if the sample is finite, the KNN rule will not be guaranteed to be the optimal way of using the information contained in the neighborhood of unclassified feature vectors.

An example of using KNN classification is shown in Fig. 6.17. The test feature vector (green dot with a question mark) should be classified either to the class of blue squares or to the class of red triangles. If K=3, it will be classified to the triangle class because there are 2 triangles and only 1 square inside the inner circle. If K=6, it will be classified to the square class, because there are four squares versus 2 triangles inside the outer circle.

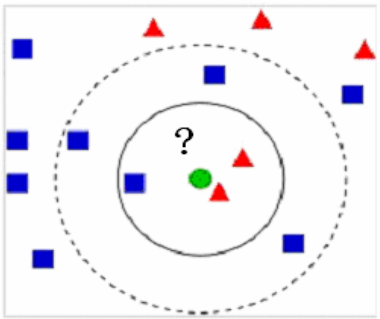
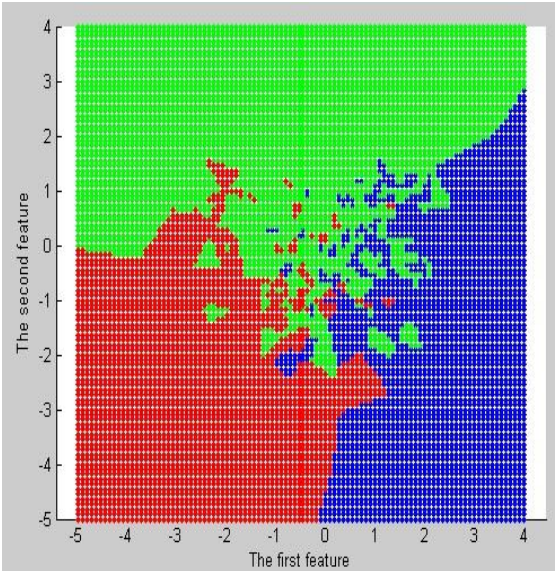
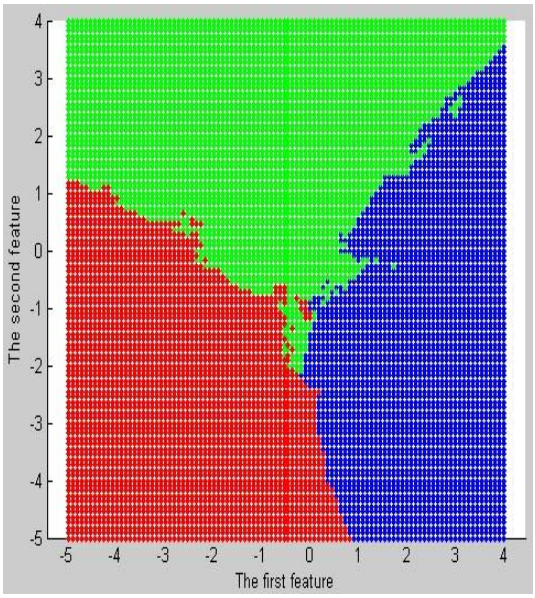


Fig. 6.17 KNN classification. Blue square and red triangle represent two classes. The green circle is classified into different classes if K is set differently.

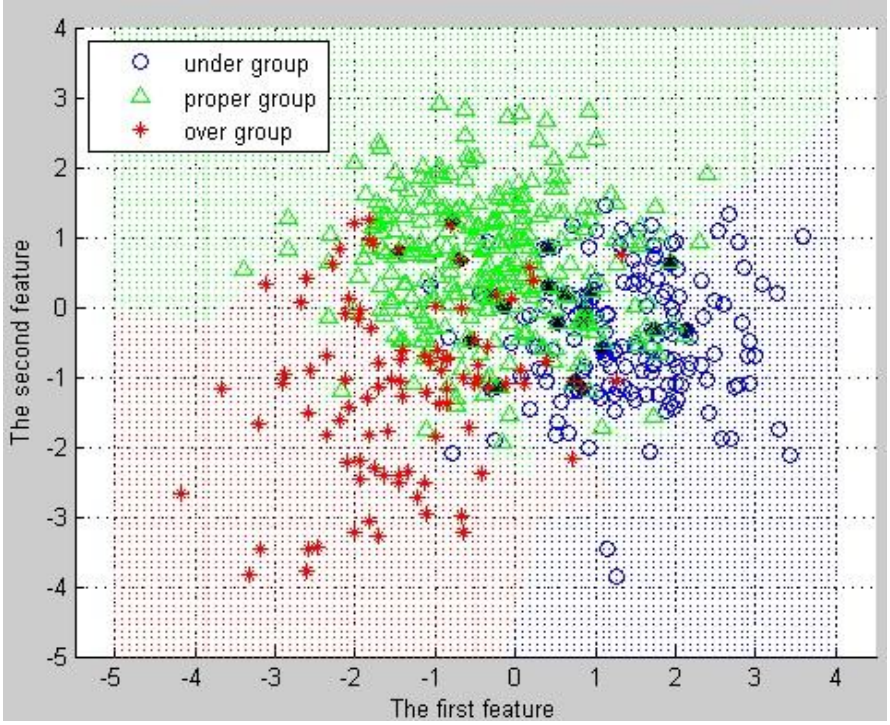
The KNN classifier is used for the classification of CCA feature vectors of DS, and the result is shown in Fig.6.18. By setting parameter K with values of 1 and 20, different partitions are shown in two grid maps. Fig.6.18(a) shows a detailed partition while Fig.6.18(b) displays a rather rough partition. As in the DA classifier case, the misclassified feature vectors are plotted by black cross symbol in Fig.6.18(c). The CCR of KNN classifier is validated by the 'Leave-M-out' cross validation method, and is shown in Table 6.8. The CCR roughly decreases from 76% to 68% as K decreases from 50 to 1. The maximum of K is set to 50, because higher values of this parameter are found to generate a lower CCR.



(a)



(b)



(c)

Fig.6.18 KNN classification on CCA features of Doppler spectra. (a) grid map (k=1); (b) grid map (k=20); (c) classification map. Black cross symbol represent those misclassified feature vectors.

Table 6.8 k-NN classification verified by 'Leave-M-out' cross validation

Training dataset number	Testing dataset number: M		CCR
358	300	(56%)	k=1, 68%; k=20, 76%; k=50, 76%
438	100	(19%)	k=1, 67%; k=20, 76%; k=50, 76%
500	38	(7%)	k=1, 68%; k=20, 75%; k=50, 76%
528	10	(2%)	k=1, 68%; k=20, 74%; k=50, 77%
533	5	(1%)	k=1, 66%; k=20, 75%; k=50, 75%
537	1	(0.2%)	k=1, 63%; k=20, 82%; k=50, 70%

The main drawback of the KNN rule is that it implicitly assumes the nearest neighbors of a feature vector point $\underline{x}$ should belong to the same class, and be contained in a relatively small region, so that there is sufficient resolution in the classification using different densities of classes in that region. However, the

distance between $\underline{x}$ and one of its closest neighbors can sometimes become very large outside the regions of high density. As a result, it can be questioned whether it is still reasonable in that case to give all the neighbors an equal weight in the decision, regardless of their distances to the point $\underline{x}$. Suppose the k nearest neighbors of $\underline{x}$ are $\underline{x}_1, \underline{x}_2, ..., \underline{x}_k$, and $d_1, d_2, ..., d_k$ are the corresponding distances arranged in increasing order. The class label of $\underline{x}_i$ is intuitively better to be given a greater weight than to the class label of $\underline{x}_j$, when $d_i < d_j$. Dudani (1976) has proposed to assign to the ith nearest neighbor $\underline{x}_i$ a weight w_i defined as: $w_i = \frac{d_k - d_i}{d_k - d_1}, d_k \neq d_1$, and if $d_k = d_1, w_i = 1$. Then, he classified the unknown feature vector $\underline{x}$ to the class for which the weights of the representatives among the K nearest neighbors sum to the greatest value. Dudani showed that this classification rule yielded higher CCR than those obtained using the unweighted KNN procedure for at least one particular dataset. However, whether to use a weighted or unweighted KNN is still an open problem. Some researchers (Bailey and Jain 1978; Jain and Mao 1991; Jain et al. 2005) have showed that, in the infinite sample case $(N \to \infty)$, the error rate of the traditional unweighted KNN rule is better than that of any weighted KNN rule. Others such as Macleod et al. (1987) presented arguments showing that this conclusion may not apply if the training set is finite. In this dataset, unweighted is better than weighted KNN, because there are few cases that the feature vector which is in the same class but in a low density region. The CCR is also stable at the level 76% (in Table 6.8, $K = 50$), varying the training set from 533 to 358.

Support Vector Machines classifier

Support vector machines (Cortes and Vapnik 1995) (SVM) have been introduced for solving pattern recognition problems. The standard SVM predicts, for each given input feature vector, which of two possible classes the input belongs to. SVM is thus a non-probabilistic binary linear classifier. Intuitively, a SVM model is a representation of the feature vectors as points in space, mapped so that the two classes are divided by as wide a gap as possible. Testing feature vectors are then mapped into that same space and labelled by a class name based on which side of the gap they fall. This method constructs an optimal separating hyperplane in a higher dimensional space into which the input data are mapped.

The largest distance that a hyperplane has to the nearest training data points of any class is called the margin, based on which a separation can be achieved. In general the larger the margin the higher the CCR of the classifier. The hyperplanes in a higher dimensional space are defined by the rule that the dot product of points on these hyperplanes with a vector in that space is constant. The quality and complexity of the SVM solution does not depend directly on the dimension of the input space (Cortes and Vapnik 1995).

Mathematically, the formulation is summarized by a set of equations of an optimization problem:

$$training\ set: D = \{(\underline{x}_i, c_i) \mid \underline{x}_i \in \mathbb{R}^p, c_i \in \{-1,1\}\}_{i=1}^n$$

$$Hyperplane: \underline{w} \cdot \underline{x} - b = 0$$

$$\because \quad for\ \underline{x}_i\ of\ the\ first\ class: \underline{w} \cdot \underline{x}_i - b \geq 1$$
$$for\ \underline{x}_i\ of\ the\ \text{sec}ond\ class: \underline{w} \cdot \underline{x}_i - b \leq -1$$
$$\therefore \quad c_i(\underline{w} \cdot \underline{x}_i - b) \geq 1,\ for\ all\ 1 \leq i \leq n$$

$$Maximize: \frac{2}{\|\underline{w}\|}(\text{margin}), subject\ to: c_i(\underline{w} \cdot \underline{x}_i - b) \geq 1$$

$$Lagrange\ multipliers: \min_{\underline{w},b} \max_{\underline{\alpha}} \{\frac{1}{2}\|\underline{w}\|^2 - \sum_{i=1}^{n} \alpha_i [c_i(\underline{w} \cdot \underline{x}_i - b) - 1]\}$$

In this application, there are three classes, i.e. over-, proper-, and under-estimation. The performance of the algorithm is tested on the two features extracted by CCA. There are 538 feature vectors, each is denoted by $\underline{x}_i \in R^2$, $i = 1,...,538$, and the associated class label c_i. c_i would be 1 if $\underline{x}_i$ is in the class ω, and -1 if not. The whole dataset is $X \in \mathbb{R}^{538\times 2}$. Validation of the SVM classification is via 100-time iterative splits of the data into testing and training sets. Since SVM is only capable of separating two classes, setting ω with 1, 2, 3, the whole dataset is rearranged into under or non-under, proper or non-proper, over or non-over classes respectively. For $\omega = 1$, each feature vector is either in under estimation or non-under estimation group. The hyperplane is searched using the least square error algorithm, which is both fast in speed and high in accuracy. An example of SVM classification is shown below in Fig.6.19, with the average and the standard deviation of CCR of each class listed in Table 6.9.

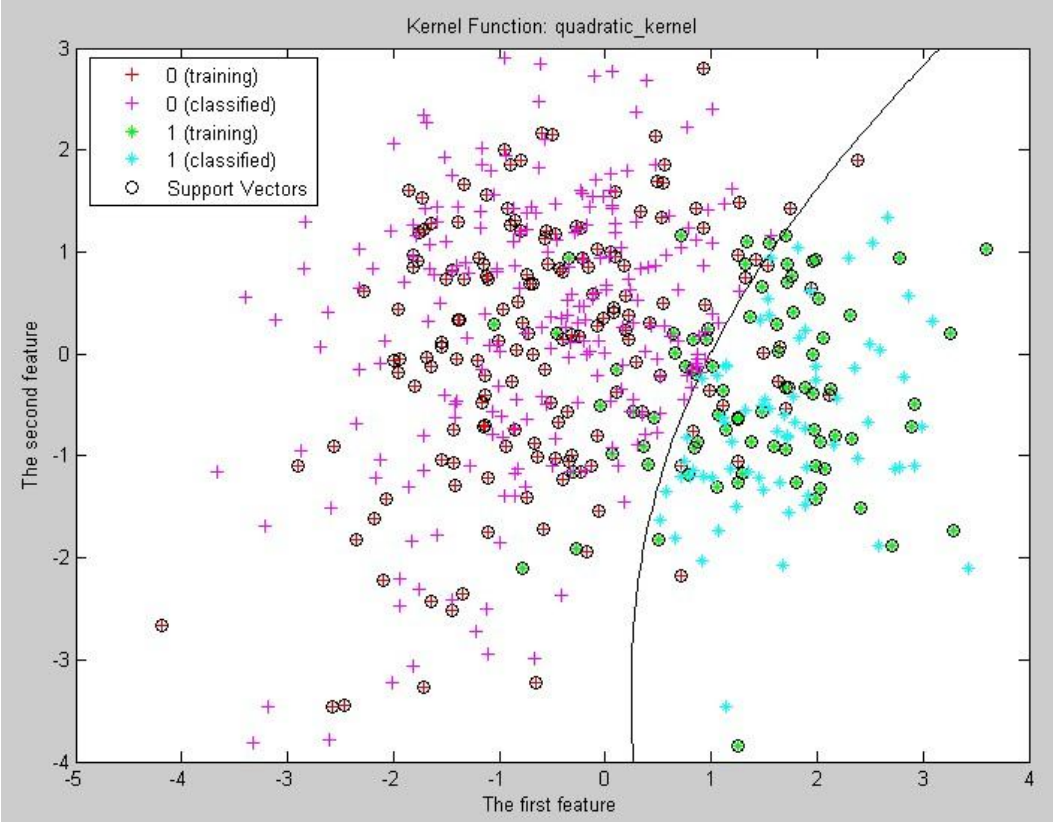


Fig.6.19 SVM classification. Half of the dataset is selected as training set. Red and magenta cross symbols represent training and testing feature vectors respectively belonging to non-underestimation class; green and blue star symbols represent training and testing feature vectors respectively belonging to underestimation class. The black circle symbolizes points which are support vectors.

Table 6.9 SVM classification by 'Leave-M-out' cross validation

Training set	Testing set	CCR
258	300 (56%)	**1**:85%, **2**:78%, **3**: 89%
438	100 (19%)	**1**:86%, **2**:79%, **3**: 90%
500	38 (7%)	**1**:87%, **2**:80%, **3**: 90%
528	10 (2%)	**1**:87%, **2**:80%, **3**: 90%
533	5 (1%)	**1**:87%; **2**:80%, **3**: 91%
537	1 (0.2%)	**1**:83%; **2**:81%; **3**:91%

Summary

There are several reasons for this comprehensive investigation of the various classification methods shown in Table 6.5 to solve a given classification problem. Firstly, each classifier is developed in a different context. For example, the tree is not a single set of clusters, as in KMEANS, but rather a multi-level hierarchy, where a cluster at a higher level comprises one or more sub-clusters. Secondly,

different feature sets can be used for the representation of the same problem. For instance, the 45 statistical features can be used to describe a DS, so can the extracted features of the first and second order region of a DS. For the 45 features of statistics, even if the three best individual features are selected, feature vectors of different classes still overlap in the feature space. The tree classifier which is based on a few decision threshold rules is more appropriate, because the statistics usually reveal generic information on different classes (although in this application, they are proved to be uninformative in section 6.3). Other classifiers, such as KMEANS, SOM, DA, KNN, and SVM are more concerned with the ratio of within-class and between-class distances. So before using these classifiers, the feature space must be properly constructed (the optimal feature space is that each class occupies a compact and non-overlapping subspace with a few outliers). These classifiers aim at finding representative feature vectors for each class, and then calculating the distance between a feature vector and the representative feature vectors. With minimum distance from a representative feature vector, a test feature vector is assigned to a class.

After comparing the CCR of four supervised classifiers in Table 6.6-6.9, it was found that SVM has the highest CCR, 87%, for underestimation, while the other three give similar values around 75%. However, in the comparison, SVM only considers two classes, while the other methods all try to identify three classes. Though in the above there has been no testing of how the other methods would perform if presented with the two-class problem, the expectation is that SVM would still out-perform them because it is a more sophisticated algorithm, and it is more able to deal with cases when classes are mixed by using decision boundaries of more complicated shapes than the other algorithms (Bennett and Campbell 2000; Buciu et al. 2006). Though the analysis indicates the superiority of SVM for the problem, it is nonetheless useful to have learnt the different properties of the other methods, which might be useful for solving other classification problems in future work. In addition to the high CCR for underestimation class, SVM shows more robust performance than the others for different proportion of points in the training and testing set; the computation speed of SVM is also higher than other neural networks. Therefore by classifying feature vectors into under and non-under classes, those underestimated DS will be identified. Thus, SVM is selected as the best classifier to solve this problem.

6.6 Validation of the whole pattern recognition approach with ED and GED scheme

After finding the optimal preprocessing, feature extraction method and classifier, the procedures for the classification of DS is performed as follows:

1. Preprocess all the 538 DS by selecting the first and second order region: $sfs - 50 \sim sfs + 49$

2. Obtain the label information from buoys for 538 feature vectors

3. Split the training and testing feature vectors by cross validation method

4. Training feature vectors and training labels are used to calculate the model of CCA mapping.

5. Both training and testing feature vectors are mapped into CCA feature space by the model.

6. By SVM, the class labels of testing data are estimated for under, proper, and over classes.

7. Compare the true class labels and the SVM labels of testing data, and calculate the CCR. The CCR is shown in Table 6.10. Each percentage is an average of a 100-time iteration.

Table 6.10 SVM classification by 'Leave-M-out' cross validation.

Training set	Testing set	CCR
258	300 (56%)	**1**:72%, **2**:60%, **3**: 80%,
438	100 (19%)	**1**:75%, **2**:64%, **3**: 84%,
500	38 (7%)	**1**:76%, **2**:65%, **3**: 85%,
528	10 (2%)	**1**:76%, **2**:65%, **3**: 86%,
533	5 (1%)	**1**:77%; **2**:64%, **3**: 85%;
537	1 (0.2%)	**1**:74%; **2**:67%; **3**: 87%

The goal of designing a recognition system is to classify unknown testing feature vectors which are likely to be different from the training feature vectors. Hence, the classification performance shown in Table 6.9 is higher than the performance shown in Table 6.10. The reason is attributed to the model estimated by the training set. In Table 6.9, the model, i.e. means, within- and between-covariances of each class, in the training set is the same as the model in the testing set (the training set is the testing set). In Table 6.10, the model is estimated from the training set and no buoy information of the testing set is used. The closer the model in the training and testing sets, the smaller the difference between the CCR in the two tables.

In the above, the joint application of ED and GED has been tested without the help of buoy data. Since the model needs to be trained, at least half a month of buoy data has to be available. In the classification, the first half month data were utilized for the training and the class membership of the DS in the remaining half month was predicted. After the classification, the operational application of ED and GED can be performed on the classified DS. Results have shown that within the testing DS, the number of DS which were classified as underestimation class was 61. Within the 61 DS, 46 are found to be truly from underestimation class, verified by the buoy data. At this stage, the joint ED and GED was applied only to those DS in underestimation class, and the non-underestimation DS were

preserved. Table 6.11 displays the resulting statistics of the oceanographic parameters after this operational application (EDGEDcnnnspr) on a small set of data, compared with the results of ED and GED using buoy information (EDGEDcnnnwbi) in the Table 5.4. As found before, EDGED application does not improve period and direction estimation. It is noticed that the improvement on the waveheight estimate is not as good as when using a whole month of buoy information, because the amount of processed data is small and about 25% 'under' DS were misclassified, which shouldn't therefore have been processed by the joint scheme. However, improvements in the estimation of the frequency spectrum by correctly classified and clutter mitigated DS were observed and are shown in Fig.6.20.

Table 6.11 Comparison of statistics of waveheight, period, and direction estimates in Feb. in 2005. 'cono'- original DS from 'cm' and 'np'; 'cnnn'-new DS from 'cm' and 'np'; 'wbi'- with buoy information; 'spr'- statistical pattern recognition; C.C.- correlation coefficient; ME- mean error; STD- standard deviation; the numbers in bold are the estimations of these statistics for original radar data.

(a) Waveheight

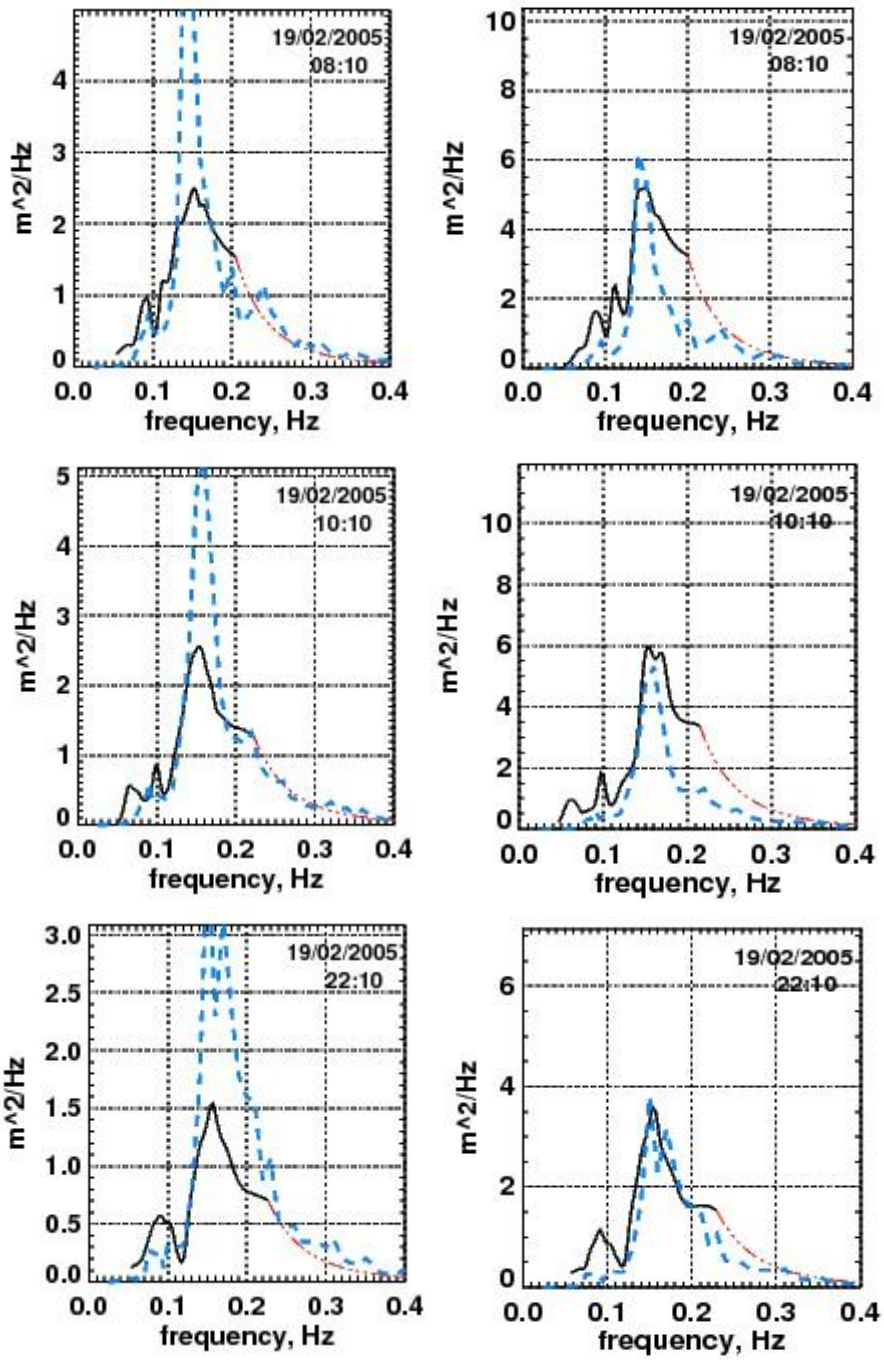
Scheme	No. of valid data	*hs*			Peak *hs*		
		C.C.	ME	STD	C.C.	ME	STD
cono	312	**0.942**	**0.095**	**0.37**	**0.871**	**0.197**	**0.293**
EDGEDcnnnwbi	335	0.964	0.286	0.28	0.898	0.271	0.258
EDGEDcnnnspr	328	0.949	0.147	0.37	0.878	0.217	0.282

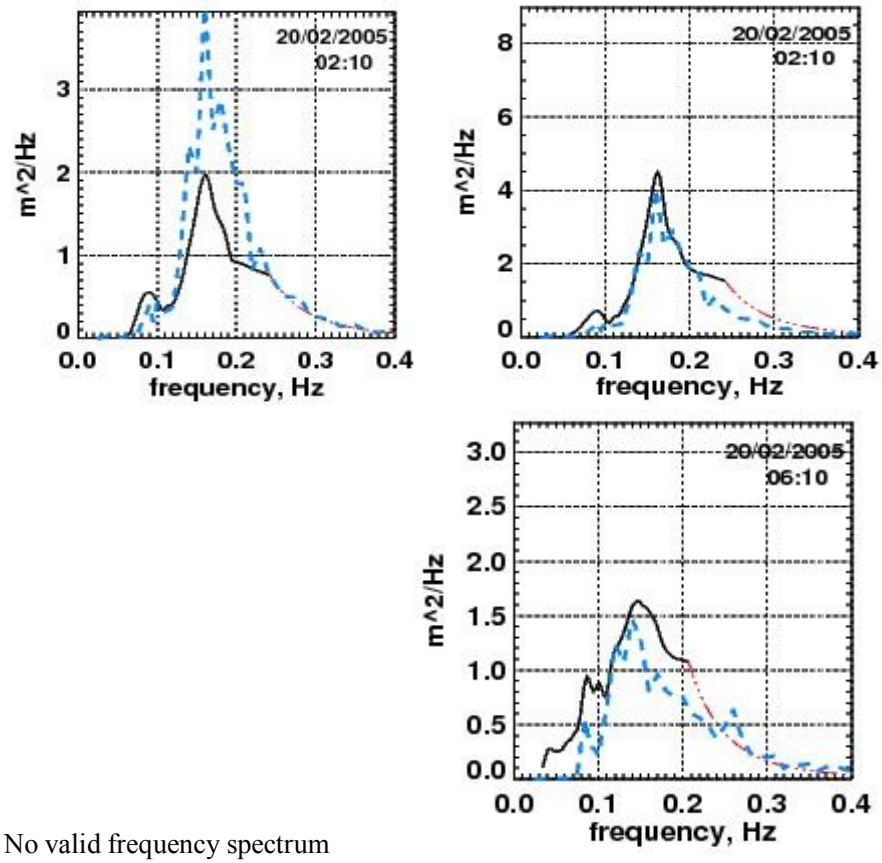
(b) Period

Scheme	No. of valid data	T_2			T_p		
		C.C.	ME	STD	C.C.	ME	STD
cono	312	**0.495**	**0.766**	**1.555**	**0.316**	**1.608**	**4.57**
EDGEDcnnnwbi	335	0.485	0.770	1.560	0.300	1.59	4.7
EDGEDcnnspr	328	0.499	0.770	1.526	0.310	1.59	4.62

(c) Direction

Scheme	No. of valid data	D_m		D_p	
		ME	STD	ME	STD
cono	312	**-14.35**	**43.07**	**-17.37**	**51.31**
EDGEDcnnnwbi	335	-16.086	51.81	-17.84	59.77
EDGEDcnnnspr	328	-14.95	44.28	-17.05	53.07





No valid frequency spectrum

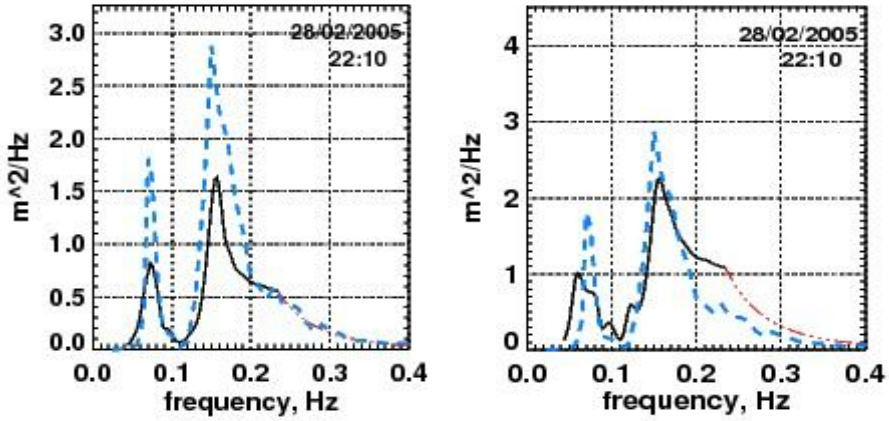


Fig.6.20 Frequency spectra. left: original DS; Right: new DS processed by ED and GED after classification

6.7 Summary

This chapter investigated the field of statistical pattern recognition for classifying a DS into one of three categories, under, proper, or overestimation of wave height.

A classical solution was attempted in section 6.3, which used 45 statistics as input features, non-parametric kernel density estimation of the multivariate likelihood, and a Bayesian decision rule for classification. To reduce the dimension of the feature space, the criterion of probability dependence was

examined on the 45 statistics. It was found that 24 statistics were independent of class and they were excluded from the feature set. The final best feature space was spanned by 3 individual statistics selected from the 21 statistics by the best individual feature algorithm. According to the hypothesis test of univariate and bivariate normality, it was found that the 3-feature feature vector doesn't follow a multivariate normal distribution. Hence, a non-parametric method was adopted for likelihood estimation. The resulting CCR of this Bayes density-approach was found to be around 60%, which is even lower than the CCR of the linear classification method in Table 6.3. The reason for the high misclassification lies in the selection of the feature set and the empirical estimation of class prior probability. This class prior probability can be influenced by many factors, e.g. the variation of oceanographic conditions, external clutters etc., This first attempt has demonstrated that the 45 statistics are not informative enough for classification.

Thus, another approach was investigated in section 6.4 and has shown satisfactory performance. This used the in-band 100-bin DS as input features, CCA feature extraction for the construction of the best feature space, and neural network SVM for classification. It has been found that using the in-band DS section directly is better than manipulating the DS by calculating the complicated statistics to solve this problem. It is also found that the PCA and FA methods have a much weaker ability to extract the best features than CCA and other combinations of methods, e.g. GDA, GED, and LPP. In the CCA feature space, feature vectors are found not to overlap, and feature vectors from the same class are closer; the three classes are roughly linearly separable. Quantitatively, the CCR of SOM, TREE, DA, KMEANS, and SVM were calculated and compared using the dataset that was split into training and testing sets. As the splitting ratio changed, the CCR of SVM was found to keep at a higher and more constant level than other classifiers, showing good robustness.

Finally, by means of this statistical pattern recognition solution, the testing feature vectors were classified particularly into the under-estimation class and the 'other' classes, because the joint ED and GED scheme mainly increases the underestimated DS. According to Table 6.10, the CCR for under-estimation DS is 76%, which means e.g. if the classifier says it is going to derive underestimated wave height, it has 76% chance to be true. In the validation of a combination of this classification and the joint ED and GED scheme, only half a month of DS were classified, because the other half month DS were used to train the model. 61 DS were identified as in the underestimated class and processed by the ED and GED scheme, but the improvement on the waveheight is not as noticeable as when the buoy information is used. In future work, the statistical pattern recognition method will be further developed in terms of the error type in the frequency spectrum. The application of the joint ED and GED to the data in additional months may then give a more significant result.

Chapter 7 Integration of the recommended signal processing methods

In chapters 1 to 6, various signal processing methods have been described to address different problems under different assumptions. This chapter summarizes all the recommended signal processing methods for operational use and shows their impact on selected wave parameters. The visible clutter mitigation schemes which require case-by-case human intervention are not included for comparisons of statistics of parameters, but an example of visible clutter mitigation is presented to show the potential of automation before further development.

7.1 Recommended operational procedures

Four operational procedures are recommended and listed as follows:

1. Denoise each DS in the RD image from rb 13 to 18 by non-white denoising using wavelet analysis in order to correctly identify the Bragg peaks, i.e. reduce the divergence of first order peaks estimated by Seaview.
2. Mitigate invisible clutters in each DS by an unsupervised dynamic scheme, UDS (see section 5.3.6).
3. Mitigate invisible clutters in each DS by a sliding eigenvalue decomposition scheme, SED (see section 5.3.6).
4. In another invisible clutter mitigation scheme, ED and GED, the pattern recognition scheme is performed to check which class the DS belongs to, under-, proper- or over-estimation. The procedure includes:
    1. Preprocess all data.
    2. Split the preprocessed feature vectors into a training set and testing set.
    3. Use the training set to estimate the CCA mapping matrix.
    4. Map the testing feature vectors into CCA feature space using the mapping matrix.
    5. Use SVM to classify DS for the underestimation class.
    6. The class membership is then determined.
    7. Perform joint ED and GED scheme on the underestimated DS.

Note that these four schemes are not implemented for all data in a sequence. The denoising method is an operational scheme applied individually to detect Bragg peaks, based on which Seaview could derive more wind, current and even wave parameters for far ranges. It cannot be applied operationally in sequence with invisible clutter mitigation schemes, since visible instead of invisible clutters are present in far ranges most of time. The last scheme, joint ED and GED, cannot be used sequentially with UDS or SED, because after the operation of UDS and SED, the DS will be smoothed to some degree, which will make invisible clutters harder to remove by ED and GED. UDS and SED are the only two schemes that can be sequentially applied to all data (see Table 7.2, which shows that their step-by-step operation is better than each individual one). In the next section, I

will analyze all the data (Feb. & Mar.) together with these four schemes. Since the statistical pattern recognition method at the moment is not robust enough to be used for two months of data, the buoy information is used for the classification in the EDGED scheme in the following comparison. Note that the operational application in section 5.3.6 is different from the one demonstrated here in that in section 5.3.6, only Feb. data are used to test the performance of each individual operational scheme, including the joint application of EDGED and the statistical pattern recognition method. No sequential operation of these schemes was taken into consideration at that time.

7.2 Operational results of recommended operational procedures

Table 7.1 shows improvements in the identification of Bragg peaks in Feb. and Mar. dataset, by operational scheme 1, the non-white denoising. The relative divergence:

$$r_{div} = \frac{\text{identified separation of Bragg peaks - theoretical separation of Bragg peaks}}{\text{theoretical separation of Bragg peaks}}$$

estimated by Seaview software, is examined on the original DS and the DS after denoising datasets. A value close to zero indicates better detection of Bragg peaks, so the absolute value of r_{div} is considered for the following comparison.

In the table, six parameters are used to present the improvements. n_1 is the number of DS which were taken as invalid data and cannot provide any wave parameter before denoising but can after denoising; n_2 is the number of DS for which $|r_{div}|_{before} < |r_{div}|_{after}$; n_3 is the number of DS for which $|r_{div}|_{after} < |r_{div}|_{before}$. The higher n_1, the better; the larger n_3 than n_2, the better. The other three parameters describe the divergence. $\overline{r}_{div}$ is the mean divergence of the n_1 identified DS; $S_{before} = \sum(|r_{div}|_{after} - |r_{div}|_{before})$, where $|r_{div}|_{before} < |r_{div}|_{after}$, measures how much better those original DS are identified by Seaview software than denoised DS. $S_{after} = \sum(|r_{div}|_{before} - |r_{div}|_{after})$, where $|r_{div}|_{after} < |r_{div}|_{before}$, measures how much better those denoised DS are identified than the original DS. Hence, the lower $\overline{r}_{div}$ and S_{before}, the better; the higher S_{after}, the better.

Table 7.1 Divergences before and after non-white denoising on two months data of Feb. and Mar.

Radar site	n_1	n_2	n_3	$\overline{r}_{div}$	S_{before}	S_{after}
Feb. & Mar. 'cm'	544	2414	2873	0.0666	51.4154	126.2653
Feb. & Mar. 'np'	439	2020	2218	0.0667	39.8330	97.5147

The table shows that the value of radar data for the estimation of oceanographic parameters is improved by this operational scheme, since n_1 more pairs of Bragg peaks are identified by Seaview software after denoising. The quality of the newly identified Bragg peaks is indicated by the small value of $\overline{r}_{div}$. Since $n_3 > n_2$, more Bragg peaks are correctly identified. Finally, since $S_{before} < S_{after}$, the amount of better identification is larger than the amount of worse identification.

Table 7.2 Comparison of statistics of waveheight, period, and direction estimates in Feb. and Mar. in 2005. 'cono'- original DS from 'cm' and 'np'; 'cnnn'-new DS from 'cm' and 'np'; 'wbi'- with buoy information; Percentage- percentage of valid data in the total dataset; C.C.- correlation coefficient; ME- mean error; STD- standard deviation. The numbers in bold are the estimates of these statistics from the original radar data.

(a) Waveheight

	Scheme	Percentage	*hs*			Peak *hs*		
			C.C.	ME	STD	C.C.	ME	STD
Feb. & Mar.	cono	48.8%	**0.910**	**0.132**	**0.40**	**0.809**	**0.295**	**0.328**
	UDScnnn	49.6%	0.912	0.150	0.40	0.817	0.297	0.320
	SEDcnnn	51.5%	0.911	0.176	0.40	0.820	0.322	0.329
	EDGEDcnnnwbi	49.9%	0.943	0.278	0.31	0.846	0.355	0.292
	UDSSEDcnnn	50.1%	0.913	0.209	0.40	0.830	0.336	0.324

(b) Period

	Scheme	Percentage	T_1			T_p		
			C.C.	ME	STD	C.C.	ME	STD
	cono	48.8%	**0.701**	**0.593**	**1.371**	**0.405**	**1.359**	**4.58**
Feb.	UDScnnn	49.6%	0.798	0.460	1.070	0.480	1.142	3.88
&	SEDcnnn	51.5%	0.794	0.476	1.081	0.412	1.129	4.30
Mar.	EDGEDcnnnwbi	49.9%	0.716	0.570	1.324	0.451	1.192	4.24
	UDSSEDcnnn	50.1%	0.792	0.468	1.072	0.428	1.121	4.19

(c) Direction

	Scheme	Percentage	D_m		D_p	
			ME	STD	ME	STD
	cono	48.8%	**-12.47**	**38.90**	**-15.64**	**47.06**
Feb.	UDScnnn	49.6%	-11.66	36.54	-15.26	44.19
&	SEDcnnn	51.5%	-11.66	35.37	-16.62	46.48
Mar.	EDGEDcnnnwbi	49.9%	-12.47	45.79	-15.07	51.34
	UDSSEDcnnn	50.1%	-10.14	34.56	-14.31	41.42

Improvements in statistics of oceanographic parameters by operational scheme bullets 2-4, i.e. UDS, SED, and EDGED based on pattern recognition, are presented in Table 7.2. Some qualitative judgements are made below in terms of the accuracy and validity of processed data.

- Waveheight

All schemes improve the C.C. statistic, although they worsen the M.E. statistic to some extent (for linear parameters, the C.C. and STD (rather than mean error) do provide useful information as the M.E. does for circular parameters (Barstow et al. 2005)). Among them, EDGED shows the biggest improvement of C.C. and STD. The fact that M.E. is worsened the most by EDGED is an expected result, because EDGED mainly deals with the underestimation problem (the solution for the overestimation problem is under development at the moment. The idea of this method being developed is to remove clutters from the second order or increase the SNR of Bragg peaks). When measurements with negative bias are improved but with positive bias are generally unaffected, the M.E. will be driven higher (a

positive M.E. means the radar measurement is bigger than the buoy measurements). The scatter plots of significant waveheight in 'Feb & Mar.' in Fig.7.1, estimated by original and EDGED processed radar data, have confirmed these qualitative remarks.

- Period

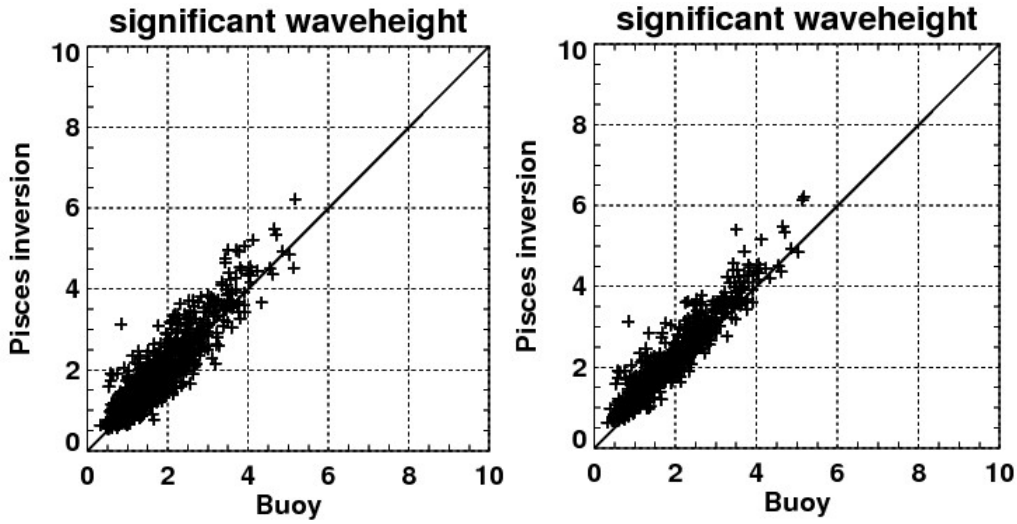
All schemes show improvements of the statistics of this parameter, with UDS and UDSSED being the best two schemes.

- Direction

UDS, SED, and UDSSED improve the direction statistics, with EDGED making the statistic of STD worse, especially for frequencies < 0.2Hz. This is probably associated with the impact of the ED scheme, which changes the spectrum at lower frequencies.

- Percentage of valid data

SED is the best scheme in increasing the percentage of usable Feb. & Mar. data.



(a) (b)

Fig.7.1 Scatter plots of significant waveheight. (a) original radar data; (b) radar data processed by EDGED scheme.

To conclude, all schemes, except EDGED, display overall improvements on those statistics presented above using Feb. & Mar. data. UDS displays more robust performance than UDSSED, SED and EDGED for period and direction estimation. SED is best at increasing the percentage of valid data. When the quality of oceanographic parameters and the percentage of valid data are both taken into consideration, UDSSED, a sequential operation of UDS and SED, improves the quality and validity at the same time, and thus is recommended to be the best operational scheme at the moment.

Last but not least, although the automation of visible clutter mitigation schemes

has not been implemented yet, a few attempts have been made. An example of using some selected visible clutter mitigation schemes sequentially to mitigate visible clutters is given below, with parameters set by human intervention (automation of the parameter setting is going to be developed in future). In the example, the RD image is selected from 'cm' site, 04:10, 28/02/2005, where there are meteor trails and II, which is commonly found in the dataset. The sequential visible clutter mitigation processing includes four steps:

Step 1: Mitigate II

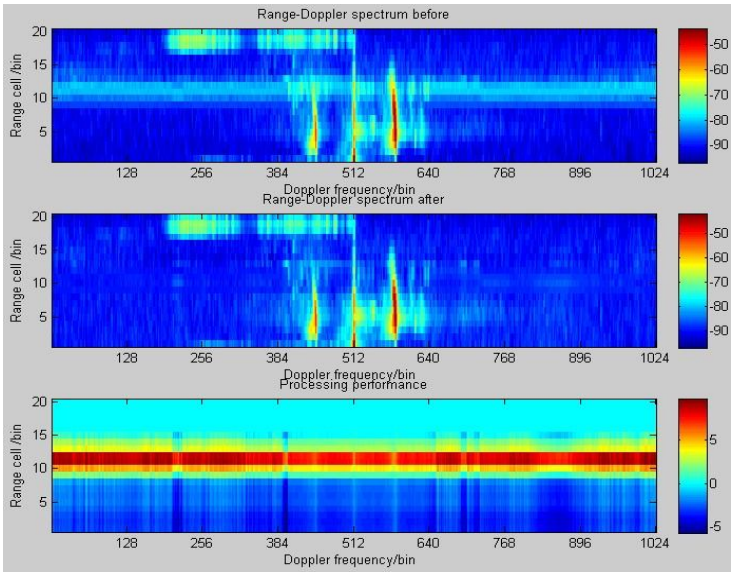


Fig.7.2 Mitigate II

Step 2: Mitigate residual II surrounding *sfs*

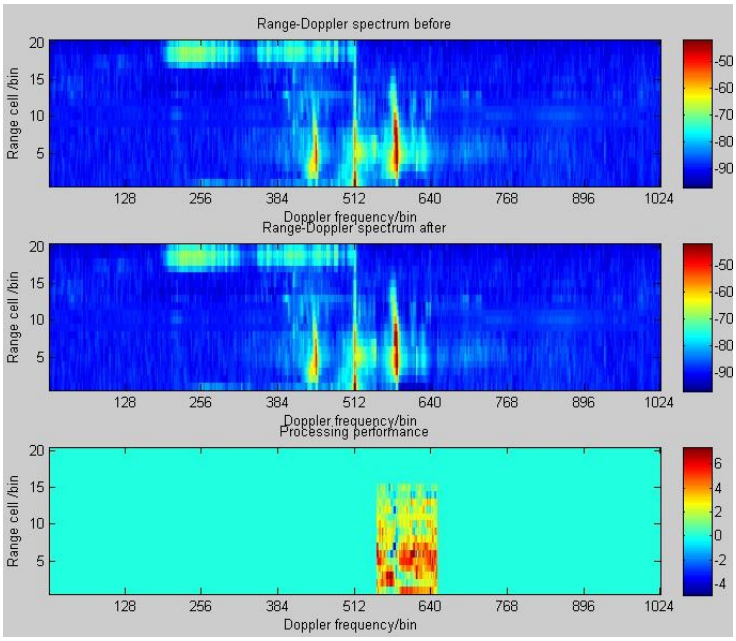


Fig.7.3 Mitigate residual II beside *sfs*

Step 3: Mitigate residual II surrounding the *sfi*

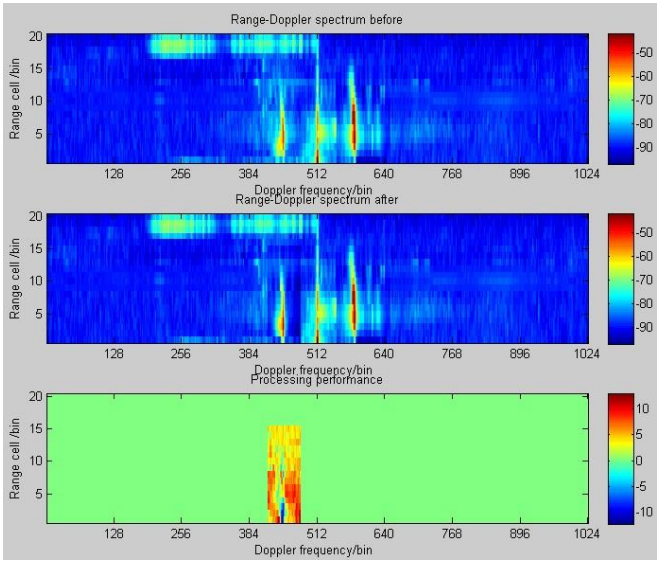


Fig.7.4 Mitigate residual II beside *sfi*

Step 4: Mitigate meteor trails

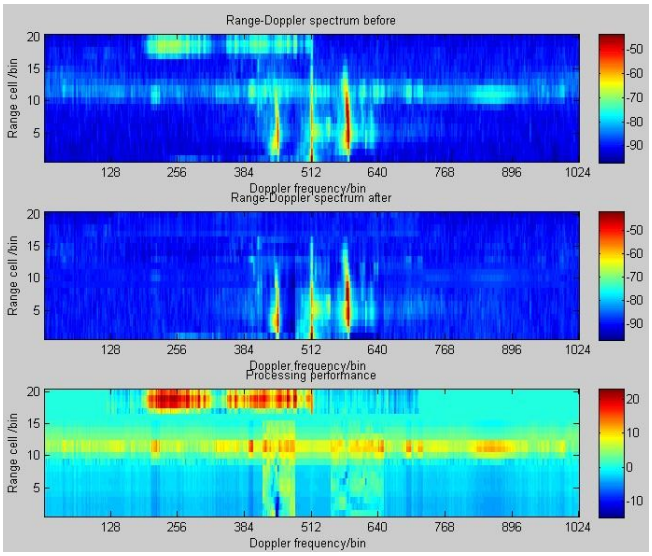


Fig.7.5 Mitigate meteor trails. Top: original RD image; Middle: RD image after clutters are mitigated; Bottom: mitigation performance.

From the bottom RD image in Fig.7.5, it can be seen that the meteor trails and inhomogeneous clutters that contaminate sea echoes are largely removed. In this example, although there is still substantial operator intervention for the setting of training area A and processing area B, the type of scheme has been fixed for mitigating specific clutter in each step. They are included in this chapter to show the potential and which part needs to be done (the automation of the image recognition) for the operational use of visible clutter mitigation schemes in future.

Chapter 8 Conclusions

A number of signal processing techniques have been explored, either in a preliminary or intensive sense, for the purpose of improving the ocean wave estimation from a HFSWR. The work in this thesis has been motivated by the strong negative impact of clutters contained in the DS for accurate wave estimation. A three step image processing approach has been developed in order to mitigate the effect of contaminations in the second order continuum caused by visible and invisible clutters, which can lead to invalid or inaccurate wave estimations. Some improvements in the quality of ocean wave measurements have been demonstrated.

The basic principles of HFSWR, i.e. the first and second order equations, the range and Doppler transform, were introduced at the beginning of chapter 2. The algorithm for inverting DS into ocean wave directional spectra was briefly explained; the methods for extracting wave parameters: significant waveheight, mean wave period, mean wave direction from the wave directional spectra were described. These parameters were used to validate the improvements in wave estimation by the proposed clutter mitigation approach and some other useful approaches developed in this thesis. The Pisces radar and the configuration of the dual radar experiment were discussed as the DS used in this thesis were available from this radar system.

Chapter 3 analyzed the properties of II and RFI because they have been acknowledged to be the two most contaminating clutters in the DS for ocean remote sensing. It has been found that the II in the Pisces dataset appears mainly at night time, caused by the sporadic E (E_s) layer in the ionosphere. This E_s layer is very unstable, and it can be a large zone or a small slice in space, homogeneous or inhomogeneous in electron densities. These physical features of the E_s layer explain the graphical features of II in the RD image, i.e. Doppler correlated, short in range, sometimes homogenous, sometimes not. Unlike II, RFI has been found mainly in the day time in this dataset, particularly in some fixed hours. RFI has nothing to do with the transmitted radar waves but is an external interference. When it is demodulated in the radar receiver, a single frequency RFI occupies a single Doppler frequency over all range bins. Theoretical analysis for the cause of the graphical feature of RFI in the RD image was presented in chapter 5, together with the image processing approach to mitigate it.

Chapter 4 mainly addressed the problems of finding the Bragg peaks and supplementing the missing DS. In HFSWR oceanography, the measurements of both surface current and wind direction rely on accurate positioning of the Bragg peaks. The first technique in chapter 4, wavelet analysis plus ICA algorithm, was developed to identify the frequency position of Bragg peaks in the case that they

are surrounded by pseudo-peaks from clutters. More often than not, the radar backscatter from far range is lower in power than from close ranges, due to stronger attenuation. So its SNR is sometimes too low to distinguish the Bragg peaks from surrounding noise. Wavelet analysis has also demonstrated its ability to identify the Bragg peaks in this case by getting rid of non-white noise. A second technique, AR modelling, was investigated to model the DS, which was conjectured as a way for smoothing the DS. The potential of this smoothing technique for reducing the clutter in a larger extent than the sea echo has not become fully evident yet. The key issue unresolved in this technique is how to ensure the assumption that the DS segment to be smoothed contains more clutter than sea echo. A third technique, an adaptive filtering system, was developed to supplement missing DS. It has been found that several DS are missing from the 'np' dataset, which obviously resulted in no estimations of the ocean wave spectra at these times. Experiments have shown that this technique is able to provide a reasonably accurate modelling of the missing DS by making use of an existing DS close to it and an expected difference between the existing and missing DS. However, it has also been found that this technique is limited in its ability to track the change of the sea state, since the supplemented DS is not a real measurement of the sea state.

The most important signal processing method recommended in this thesis, a three-step image processing for clutter mitigation was fully elaborated in Chapter 5. The three steps involve: image recognition, segmentation processing, and subspace projection. In image recognition, the graphical feature of the clutter in the RD image was analyzed, and the parameters that define the training and processing area were determined. In segmentation processing, Doppler- or range-window processing was performed to remove other out-band clutters for a better mitigation of the clutter in-band, e.g. removing meteor trails from far range to mitigate in-band RFI better. In subspace projection, an eigenvalue decomposition method was used to partition the processing matrix into a 'signal' and 'clutter' subspace. The orthogonality property of the eigenvectors was utilized to span the image of clutter and sea echo. The clutter image was removed and the sea image was retained.

The mitigation of II and RFI were examined in various complexity cases of the RD image in order to validate this approach. Since these clutters were found in a specific DS, a single radar estimate of the significant waveheight and period was more appropriate than dual radar estimates for the evaluation and some enhancement in their accuracy has been found. Beside II and RFI, mitigations for clutters, large or mall in area in the RD image, homogeneous or inhomogeneous in amplitudes, range-correlated or Doppler-correlated, were also taken into account in an extended image processing scheme. In each step except image recognition, this scheme offers different options, e.g. supervised, unsupervised, static, and dynamic, for the clutters in various cases,. A resulting drawback was the intensive human intervention needed in the setting of parameters and the

selection of a most appropriate option. This drawback prevents the operational use of this visible clutter mitigation scheme at the moment.

In addition to the visible clutter mitigation scheme, in many cases, there are no visible clutters that can be recognized in the RD image. Small size clutters which hide themselves in the in-band area are often ignored by researchers and are probably the most important in the case of wave estimation. Three invisible clutter mitigation schemes, ED, GED, and SED, were developed for this scenario. It has been found that ED scheme which lowered the second order side band beside the superior Bragg peak, could reduce the overestimated swell to some degree. GED which modified the Bragg peaks and their surrounding side bands has shown a consistent increasing effect on the underestimated wind sea in a large number of experiments. With the availability of buoy data, ED and GED can be applied operationally to underestimated DS and overestimated DS in a selective way. The potential of this joint approach to improve the waveheight estimation especially when they are underestimated became evident. SED, as a general approach, was not specific to the error in a frequency spectrum. Operational application of the SED scheme has demonstrated an effect in improving the estimate of the directional parameter. Along with these invisible clutter mitigation schemes, UDS, a dynamic unsupervised scheme, was applied to all the data in February operationally. Results have shown that this scheme improved the estimate of the period estimate to some extent. In the estimation of oceanographic parameters, period and direction are more susceptible to clutters than waveheight, and thus usually less accurately estimated in HF radars. Here, improvements on these two parameters are attributed to the minimizing effect on clutters by the SED and UDS schemes. Further development of these schemes will be pursued in future work for improving the accuracy of these two challenging parameters. Another important future development is the integration of the three individual improvements, e.g. waveheight, period, and direction, by one uniform scheme.

The objective of chapter 6 was to classify the DS into under-, proper-, and over-estimation classes. The aim is to classify the data and hence apply the mitigation methods for cases when no buoy data are available, which is the normal case. The concept of using 45 statistics to represent the feature space of DS observations was investigated. A probability dependence criterion was chosen to reduce the dimension of features. An individual feature ranking algorithm was adopted to select the best 3 features. The non-parametric KDE of likelihood of these 3-feature feature vectors was implemented. The CCR of the probability-based classifier was 0.5-0.6. The low percentage was attributed to the low discrimination power of features, and the error in the estimation of the posteriori and prior probability, which could not be solved if the size of samples was not sufficient large. Another geometric approach, which was more suitable for a smaller size of samples, was investigated. CCA feature extraction was utilized to transform the feature space from 100-bin second order continuum into

3-feature CCA feature vectors. The CCA method outperformed various feature extraction methods, e.g. PCA, FA, and GDA. SVM was chosen to classify the transformed feature vectors and displayed the highest CCR and robustness among various classifiers. Finally in chapter 6, half a month of data was used as training data, and the remaining DS were correctly classified into the underestimated class with a CCR of 75%. The joint ED and GED scheme was applied in conjunction with those classified DS and the impact on the statistics of the oceanographic parameters was examined. The improvement in the waveheight statistic was not as significant as when the DS were classified using the buoy data. Although the statistic didn't reveal much improvement (because the application on the misclassified DS had a negative effect on the overall improvement and also the number of processed DS was small), those frequency spectra, derived from the correctly classified and processed DS, matched much better with the buoy estimates. Future work in this area will be to find a more direct method to identify the error type in the frequency spectrum that is driven by a DS, and also to calculate a more robust CCA mapping matrix that could predict DS within a whole month, even a whole year. Finally, Chapter 7 integrated all the recommended algorithms in this thesis in a quasi-operational way.

HFSWR is now recognized as an efficient remote sensing tool for oceanographic parameter detection. Although I haven't completely solved the problem of clutter mitigation to improve wave measurement performance in this thesis, I have identified a number of methods that show promise and that, in some situations, have provided improved wave measurements. I have also identified methods that are less useful for this application and do not appear to be worthy of further consideration. The main outstanding further work suggested is as follows:

- Improve the adaptive filtering system in order for it to be applied to DS of poor quality
- Make the visible clutter mitigation schemes automatic in terms of image recognition and parameter setting for operational use
- Develop a uniform scheme that combines the three invisible clutter mitigation schemes for improving the accuracy of waveheight, period, and direction, together
- Further develop the statistical pattern recognition method for the identification of error type in the frequency spectrum

Appendix A

1. Wavelet functions

In chapter 4.1.4, a list of wavelets are tested to compared the performance of non-white and white denoising. These wavelets are provided by Matlab wavelet toolbox (MathWorks 2011).

Wavelet Family Short Name	Wavelet Family Name
'haar'	Haar wavelet
'db'	Daubechies wavelets
'sym'	Symlets
'coif'	Coiflets
'bior'	Biorthogonal wavelets
'rbio'	Reverse biorthogonal wavelets
'meyr'	Meyer wavelet
'dmey'	Discrete approximation of Meyer wavelet
'gaus'	Gaussian wavelets
'mexh'	Mexican hat wavelet
'morl'	Morlet wavelet
'cgau'	Complex Gaussian wavelets
'shan'	Shannon wavelets
'fbsp'	Frequency B-Spline wavelets
'cmor'	Complex Morlet wavelets

2. Threshold rules used in wavelet denoising

In chapter 4.1.4, a few threshold methods are adopted in wavelet white or non-white denoising. They are also included in the Matlab toolbox (MathWorks 2011)

1. SURE

'rigrsure' uses for the soft threshold estimator, a threshold selection rule based on Stein's Unbiased Estimate of Risk (quadratic loss function). One gets an estimate of the risk for a particular threshold value t. Minimizing the risks in t gives a selection of the threshold value.

2. sqtwolog

'sqtwolog' uses a fixed-form threshold yielding minimax performance multiplied by a small factor proportional to log(length(X)).

3. heursure

'heursure' is a mixture of the two previous options.

3. Matrix pencil (R_B, R_A) in generalized eigenvalue decomposition

Given two symmetric matrices R_A and R_B with the same size. R_B is supposed to be invertible. Let $\lambda = \frac{\underline{v}^T R_A \underline{v}}{\underline{v}^T R_B \underline{v}}$, then this quotient is minimized for eigenvector $\underline{v}$ of $R_B^{-1} R_A$ associated with the smallest eigenvalue λ.

Proof: minimize this quotient requires that the derivative of λ with respect to $\underline{v}$ equals zero:

$$\frac{(\underline{v}^T R_B \underline{v})(2R_A \underline{v}) - (\underline{v}^T R_A \underline{v})(2R_B \underline{v})}{(\underline{v}^T R_B \underline{v})^2} = 0$$

$$\Rightarrow \underline{v}^T R_B \underline{v} R_A \underline{v} = \underline{v}^T R_A \underline{v} R_B \underline{v}$$

$$\Rightarrow R_A \underline{v} = \left(\frac{\underline{v}^T R_A \underline{v}}{\underline{v}^T R_B \underline{v}} \right) R_B \underline{v} \qquad (\underline{v}^T R_B \underline{v} : scalar)$$

$$\Rightarrow R_A \underline{v} = R_B \underline{v} \lambda$$

$$\Rightarrow (R_B^{-1} R_A) \underline{v} = \underline{v} \lambda$$

Therefore, a generalized eigenvalue decomposition problem with a matrix pencil (R_B, R_A) is transformed into an eigenvalue decomposition problem of $R_B^{-1} R_A$. The solution $\underline{v}$ is an eigenvector of $R_B^{-1} R_A$ associated with the eigenvalue $\lambda = \frac{\underline{v}^T R_A \underline{v}}{\underline{v}^T R_B \underline{v}}$. The minimum is achieved for the smallest eigenvalue.

4. 45 statistics:

In chapter 6, there are two preprossing methods discussed. One of them is to calculate those important or informative statistics of each DS, in order to get rid of redundant information and thus reduce the dimension of the feature space. This preprocessing will facilitate the feature selection/extraction processing and lastly the classification. In this regard, 45 statistics are calculated and their meanings are specified below.

1-3: the slope of rb 3-5 in the range profile of *sfs*, outside second order peak of *sfs* (*sfso*), inside second order peak of *sfs* (*sfsi*)

4-7: db 400-640 of a DS is selected, 241 range profiles of rb 3-8 is obtained. The correlation

coefficient between the 241 range profiles and the range profiles of sfs, sfi, $sfso$, $sfsi$ are calculated. The principal component analysis is applied on the coefficients matrix with 4×241 elements, each column being a variable. The resulting first principal component has four elements and are taken as the 4th-7th statistics.

8-11: skewness of the range profile of rb 3-8 of sfs, sfi, $sfso$, $sfsi$

12-15: Kurtosis of the range profile of rb 3-8 of sfs, sfi, $sfso$, $sfsi$

16-19: the distance of range profile of sfs, sfi, and the distance of range profile of $sfso$, $sfsi$ are calculated respectively and denoted by gap_si and gap_soi. Correlation coefficient of the two distance profiles is calculated as the 16th statistic. The variances and covariance of the two distance profiles are calculated as the 17th, 18th, and the 19th statistic respectively.

20-24: the piecewise derivative of each two adjacent range bins in the range profile of gap_si

25-29: the Doppler power gap of sfs, sfi at range bin 4 and 5 are calculated as the 25th and 26th statistics. The variance, skewness and kurtosis of gap_si are calculated as 27th-29th statistics respectively.

30-34: the piecewise derivative of each two adjacent range bins in the range profile of gap_soi

35-39: the Doppler power gap of $sfso$, $sfsi$ at range bin 4 and 5 are calculated as the 25th and 26th statistics. The variance, skewness and kurtosis of gap_soi are calculated as 27th-29th statistics respectively.

40-42: the range profiles of sfs, sfi, $sfso$, $sfsi$ are all fitted with 3 order polynomial functions. The first order derivative of the range profile of sfs, $sfso$, $sfsi$ at rb 4 and rb 5 is calculated as 40th –42nd statistics respectively.

43-45: The second order derivative of the range profile of sfs, $sfso$, $sfsi$ at rb 5 and rb 7 is calculated as 43rd –45th statistics.

5. Bhattacharya rule for feature selection

The Chernoff performance bound (Carter et al. 2009) is associated with the Chernoff distance between two PDFs, say $f(x)$ and $g(x)$. The Chernoff distance is defined as:

$$D_{CH}(f,g) = -\log \int f(x)^{\alpha} g(x)^{1-\alpha} dx$$

where $0 \le \alpha \le 1$. Let $f(x)$ and $g(x)$ be the PDFs of two classes ω_i and ω_j respectively, i.e. $f(x) = p(x \mid \omega_i)$ and $g(x) = p(x \mid \omega_j)$. Note that x is a single feature rather than a multi-feature vector. As $D_{CH}(f, g)$ increases, the upper limit on the probability of classification error between points in classes ω_i and ω_j decreases. The bhattacharya distance is a special case of Chernoff distance, for which $\alpha = \frac{1}{2}$:

$$D_{Bha}(f, g) = -\log \int \sqrt{f(x)g(x)}dx$$

and bounds the classification error for feature selection. An ideal feature would ensure that its Bhattacharya distance among all classes is maximized, which gives minimal attainable classification error.

6. Bandwidth matrix selection in kernel density estimation

To understand the use of the bandwidth matrix, the idea of a bandwidth scalar is explained. For a univariate KDE, the bandwidth is a scalar, which determines the smoothness of the kernel function. The definition of a univariate KDE is given as:

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} K(\frac{x - x_i}{h})$$

where $K(\bullet)$ is the kernel function and $\int_{-\infty}^{\infty} K(x)dx = 1$; h is the bandwidth, $h > 0$, a smoothing parameter. A range of kernel functions is available. Usually, the standard normal function is chosen as the kernel function for mathematical convenience. Fig.A1 shows an example of the KDE estimation of an unknown density function, based on a finite data sample.

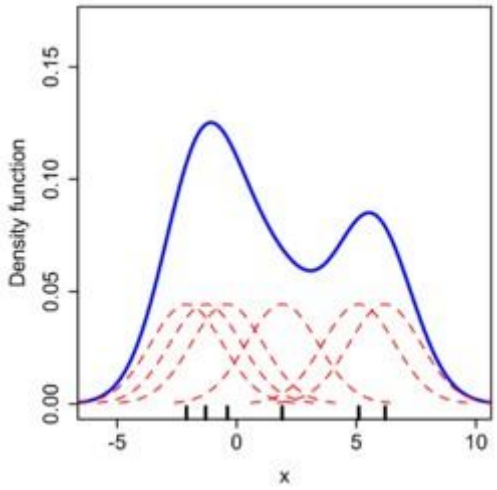


Fig.A1. KDE estimation of a PDF based on six data points: x1=-2.1, x2=-1.3, x3=-0.4, x4=1.9, x5=5.1, x6=6.2, with normal kernel function and a scalar bandwidth h. The data points are the rug plot on the horizontal axis. For estimation, a normal kernel with variance 2.25 is placed (indicated by the red dashed lines) on each of the data points xi. The kernels are summed to make the kernel density estimate (solid blue curve $\hat{f}_h(x)$). The fig. is copied from (Wikipedia 2011)

A choice of small bandwidth will lead to small bias and large variance of the KDE. Here, bias is defined by $E[f_h(x)]-E[\hat{f}_h(x)]$. There is a traid-off between the bias and the variance of the estimation. Fig.A2 shows an example of using different bandwidths.

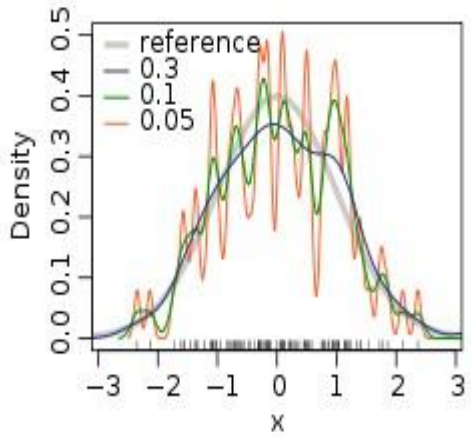


Fig.A2. KDE estimation of 100 normally distributed random numbers using different smoothing bandwidths. The red curve is made by a kernel with the smallest bandwidth. Its bias is the smallest but its variance is the largest. The blue curve is the smoothest and the closest to the true density function in gray, with a largest bandwidth number 0.3, but the bias is bigger than the other two choices. The fig. is copied from (Wikipedia 2011)

If univariate KDE is extended to the PDF of a multivariate $\underline{x}$ with d variables, the scalar bandwidth becomes the bandwidth matrix $H \in R^{d\times d}$. The key issue in multivariate KDE is the choice of the bandwidth matrix, which impacts the resulting estimation accuracy by controling the amount and orientation of the smoothing. Fig.A3 shows an example of how a bandwidth matrix influences the orientation of the kernel.

Fig.A3. Comparison of the three main bandwidth matrix parametrisation classes. Left: S, positive scalar times the identity matrix. Centre: D, diagonal matrix with positive entries on the main diagonal. Right: F, symmetric positive definite matrix. This figure is copied from (Wikipedia 2011)

The S type is a scalar times the identity matrix, so the numbers on the diagonal are the same, i.e. the same amount of smoothing applied in all variable directions, D type has different numbers on the diagonal, allowing different amounts of smoothing in each of the variables, and F type is a symmetric positive matrix, allowing arbitrary amounts and orientation of the smoothing. In this thesis, the D kernel is selected for mathematical convenience.

7. Correct classification rate using Bayes decision theory

In chapter 6, the bivariate class conditional PDFs, $p([x_1\ x_2]|\omega_i), i=1,2,3$, of each class are estimated by bivariate KDE. The prior probabilities, $p(\omega_i)$, are estimated by the ratio of the number of samples in each class to the total number. of samples. Then the posterior PDFs, $p(\omega_i|[x_1\ x_2])$, are estimated by $p(\omega_i|[x_1\ x_2]) = \dfrac{p([x_1\ x_2]|\omega_i)p(\omega_i)}{p([x_1\ x_2])}$. Next, each feature vector, $[x_1\ x_2]$, is classified to class ω_i if $p(\omega_i|[x_1\ x_2])$ is the maximum. The two features, i.e. x_1 and x_2, are selected from the feature set (19 25 26 35 36 41). These numbers are the index of statistics explained above, and they are selected out of the 45

statistics by the ‘BIF’ feature selection method. So the 6*6 matrix of correct classification rates are shown here.

	19	25	26	35	36	41
19	0	0.4547	0.4457	0.5272	0.5453	0.5399
25	0.4547	0	0.3605	0.5072	0.3587	0.3696
26	0.4457	0.3605	0	0.5036	0.3768	0.3659
35	0.5272	0.5072	0.5036	0	0.5072	0.5254
36	0.5453	0.3587	0.3768	0.5072	0	0.5507
41	0.5399	0.3696	0.3659	0.5254	0.5507	0

References:

AbedMeraim, K., et al. (1997). "Blind system identification." Proceedings of the Ieee **85**(8): 1310-1322.

Abramovich, Y. I., et al. (2000). "Stochastic-constraints method in nonstationary hot-clutter cancellation. II. Unsupervised training applications." IEEE Trans. Aerosp. Electron. Syst. **36**(1): 132-150.

Abramovich, Y. I., et al. (1998). "Stochastic-constraints method in nonstationary hot-clutter cancellation. I. Fundamentals and supervised training applications." Aerospace and Electronic Systems, IEEE Transactions on **34**(4): 1271-1292.

Allender, J., et al. (1989). The WADIC project: a comprehensive field evaluation of directional wave instrumentation, Elsevier. **16:** 505-536.

Anderson, S. J. and Y. I. Abramovich (1995). Radio oceanography : from weather forecasting to climatology. Proc. Workshop on Applications of Radio Science, Canberra.

Anderson, S. J. and Y. I. Abramovich (1996). Detection, classification and correction of ionospheric distortion in HF skywave (OTH) radar systems. Radar, 1996. Proceedings., CIE International Conference of.

Anderson, S. J., et al. (1997). "Stochastic constraints in nonstationary hot clutter cancellation." 1997 Ieee International Conference on Acoustics, Speech, and Signal Processing, Vols I - V: 3753-3756.

Anderson, S. J. and S. E. Godfrey Time-frequency signature of transient phenomena observed with HF skywave radar. Proc. Int. Conf. 'Modern Radar 94', Kiev.

Bailey, T. and A. K. Jain (1978). "Note on Distance-Weighted K-Nearest Neighbor Rules." Ieee Transactions on Systems Man and Cybernetics **8**(4): 311-313.

Barrick, D. (1972). Remote sensing of sea state by radar. **4**.

Barrick, D. and J. Snider (1977). "The statistics of HF sea-echo Doppler spectra." Antennas and Propagation, IEEE Transactions on **25**(1): 19-28.

Barrick, D. E. (1972). Remote sensing of sea state by radar. Remote sensing of the Troposphere. V. Derr, GPO, Washington, D.C. . **4**.

Barrick, D. E., et al. (1977). Ocean surface currents mapped by radar. **198:** 138-144.

Barrick, D. E. and R. Fitzgerald (2000). "The failure of “classic” perturbation theory at a rough Neumann boundary near grazing." Antennas and Propagation, IEEE Transactions on **48**(9): 1452-1460.

Barstow, S. F., et al. (2005). Measuring and Analyzing the directional spectrum of ocean waves.

Baudat, G. and F. E. Anouar (2000). "Generalized discriminant analysis using a kernel approach." Neural Computation **12**(10): 2385-2404.

Bennett, K. P. and C. Campbell (2000). Support vector machines: hype or hallelujah?, ACM. **2:** 1-13.

Bill, M. (2008). "THE DIFFERENCES BETWEEN PULSE RADARS AND FMCW ONES." from http://www.navigate-us.com/files/uploads/file/Review_5_Radar-1-1.pdf.

Botev, Z. I., et al. (2010). Kernel density estimation via diffusion, Institute of Mathematical Statistics. **38:** 2916-2957.

Breiman, L. and J. H. Friedman (1988). "Tree-Structured Classification Via Generalized Discriminant-Analysis - Comment." Journal of the American Statistical Association

83(403): 725-727.

Buciu, I., et al. (2006). "Demonstrating the stability of support vector machines for classification." Signal Processing **86**(9): 2364-2380.

Burdenski, T. (2000). "Evaluating univariate, bivariate, and multivariate normality using graphical and statistical procedures " Multivariate Linear Regression Viewpoints **26**(2).

Capon, J. (1969). "High-resolution frequency-wavenumber spectrum analysis." Proceedings of the IEEE **57**(8): 1408-1418.

Carter, K. M., et al. (2009). An information geometric approach to supervised dimensionality reduction. Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on.

Chambers, J. A. (1994). Normalised order statistic LMS adaptive filtering. Non-Linear Filters, IEE Colloquium on.

Chan, H. C. and E. K. L. Huang (1999). An investigation in interference suppression for HF surface wave radar. DREO Technical Report TR 2000-028, Defense Research Establishment Ottawa, Canada.

Cortes, C. and V. Vapnik (1995). "Support-Vector Networks." Machine Learning **20**(3): 273-297.

Cover, T. M. and P. E. Hart (1967). "Nearest Neighbor Pattern Classification." Ieee Transactions on Information Theory **IT-13**(1): 21-27.

Cristianini, N. and J. Shawe-Taylor (2006). An introduction to support Vector Machines: and other kernel-based learning methods, Cambridge university press.

Crombie, D. D. (1955). Doppler spectrum of sea echo at 13.56 Mc/s. **175:** 681?82.

Daubechies, I. (2004). Ten lectures on wavelets, Society for industrial and applied mathematics.

Daubechies, I., et al. (1992). "Special Issue on Wavelet Transforms and Multiresolution Signal Analysis - Introduction." Ieee Transactions on Information Theory **38**(2): 529-531.

Desarbo, W. (1987). "Cluster Dissection and Analysis - Theory, Fortran Programs, Examples - Spath,H." Journal of Classification **4**(1): 139-141.

Devijver, P. A. and J. Kittler (1982). Pattern Recognition: A Statistical Approach. Englewood Cliffs, NJ, Prentice/Hall.

Dudani, S. A. (1976). "Distance-Weighted K-Nearest-Neighbor Rule." Ieee Transactions on Systems Man and Cybernetics **6**(4): 327-327.

Fabrizio, G. A., et al. (2006). Stap for Clutter and Interference Cancellation in a Hf Radar System. IEEE Trans. Acoustics, Speech Signal Process., Toulouse.

Fabrizio, G. A., et al. (2004). "Robust adaptive beamforming for HF surface wave over-the-horizon radar." IEEE Trans. Aerosp. Electron. Syst. **40**(2): 510-525.

Fabrizio, G. A. and D. A. Holdsworth (2008). "Adaptive Mitigation of Spread Clutter in High Frequency Surface-Wave Radar." 2008 International Conference on Radar, Vols 1 and 2: 152-157.

Fante, R. L. and J. A. Torres (1995). Cancellation of diffuse jammer multipath by an airborne adaptiveradar. **31:** 805-820.

Fisher, R. A. (1936). "The use of multiple measurements in taxonomic problems." Annals of Eugenics **7**: 179-188.

Fix, E. and J.L.Hodges (1951). Discriminant analysis, nonparametric discrimination:Consistency properties. Randolph Field, TX,, USAF School of Aviation Medicine.

G.A.F.Seber (1984). Multivariate Observations. New York, Wiley.

Griffiths, L. J. (1996). "Linear constraints in hot clutter cancellation." 1996 Ieee International Conference on Acoustics, Speech, and Signal Processing, Conference Proceedings, Vols 1-6: 1181-1184.

Gui, J., et al. (2009). "Locality Preserving Discriminant Projections." Emerging Intelligent Computing Technology and Applications: With Aspects of Artificial Intelligence **5755**: 566-572.

Guo, D. F., et al. (2000). "A study of wavelet thresholding denoising." 2000 5th International Conference on Signal Processing Proceedings, Vols I-Iii: 329-332.

Gurgel, K. W., et al. (1999). "Wellen Radar (WERA): a new ground-wave HF radar for ocean remote sensing." J. Coastal Eng. **37**(3-4): 219-234.

Gurgel, K. W., et al. (2007). Radio frequency interference suppression techniques in FMCW modulated HF radars. Proc. IEEE OCEANS Conf. 2007, Europe.

Gurley, K. and A. Kareem (1999). "Application of wavelet transforms in earthquake, wind and ocean engineering." Engng Struct **21**(3): 149-167.

Holden, G. J. and L. R. Wyatt (1992). Extraction of sea state in shallow water using HF radar. **139:** 175-181.

Howland, P. E. and D. C. Cooper (1993). "Use of the Wigner-Ville distribution to compensate for ionospheric layer movement in high-frequency sky-wave radar systems." Radar and Signal Processing, IEE Proceedings F **140**(1): 29-36.

Iyengar, S. S. (1998). "Wavelets: An important tool of mathematical analysis for image/signal processing applications." Wavelet Applications V **3391**: 12-25.

Jain, A. K., et al. (2000). "Statistical pattern recognition: A review." Ieee Transactions on Pattern Analysis and Machine Intelligence **22**(1): 4-37.

Jain, A. K. and J. C. Mao (1991). "A K-Nearest Neighbor Artificial Neural Network Classifier." Ijcnn-91-Seattle : International Joint Conference on Neural Networks, Vols 1 and 2: B515-B520.

Jain, M. D., et al. (2005). "Nearest neighbor vector based palmprint verification." Proceedings of the Fifth IASTED International Conference on Visualization, Imaging, and Image Processing: 536-539.

Jolliffe, I. (2005). Principal Component Analysis, John Wiley & Sons, Ltd.

Khan, R., et al. (1994). Target detection and tracking with a high frequency ground waveradar. **19:** 540-548.

Khan, R. H. and D. K. Mitchell (1991). "Waveform analysis for high-frequency FMICW radar." Radar and Signal Processing, IEE Proceedings F **138**(5): 411-419.

Kittler, J. and P. A. Devijver (1981). "An Efficient Estimator of Pattern-Recognition System Error-Probability." Pattern Recognition **13**(3): 245-249.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection, Citeseer. **14:** 1137-1145.

Kohonen, T. (1991). "Self-Organizing Maps - Optimization Approaches." Artificial Neural Networks, Vols 1 and 2: 981-990.

Kohonen, T., et al. (2009). "On the Quantization Error in SOM vs. VQ: A Critical and Systematic Study." Advances in Self-Organizing Maps, Proceedings **5629**: 133-144.

Kronlandmartinet, R. (1988). "The Wavelet Transform for Analysis, Synthesis, and Processing of Speech and Music Sounds." Computer Music Journal **12**(4): 11-20.

Lagarias, J. C., et al. (1998). "Convergence properties of the Nelder-Mead simplex method in low dimensions." Siam Journal on Optimization **9**(1): 112-147.

Lagarias, J. C., et al. (1999). Convergence properties of the Nelder-Mead simplex method in low dimensions, Citeseer. **9:** 112-147.

Lipa, B. J. (1977). "Derivation of directional ocean-wave spectra by inversion of second radar echoes." Radio Science **12**: 425-434.

Lipa, B. J. and D. E. Barrick (1986). Extraction of sea state from HF radar sea echo-Mathematical theory and modeling. **21:** 81-100.

Lipa, B. J. and D. E. Barrick (1986). "Extraction of sea state from HF radar sea echo: Mathematical theory and modelling." Radio Science **21**: 81-100.

Liu, H. and H. Motoda (1998). Feature selection for knowledge discovery and data mining, Klumer Academic Publisher.

Luo, X., et al. (2001). "RFI suppression in ultra-wideband SAR systems using LMS filters in frequency domain." Electron. Lett. **37**(4): 241-243.

Macleod, J. E. S., et al. (1987). "A Reexamination of the Distance-Weighted K-Nearest Neighbor Classification Rule." Ieee Transactions on Systems Man and Cybernetics **17**(4): 689-696.

Mahafza, B. R. (2000). Radar systems analysis and design using matlab. Florida, CRC Press.

Mahalanobis, P. C. (1936). On the generalized distance in statistics. **12:** 49.

Mallat, S. and F. Falzon (1998). "Analysis of low bit rate image transform coding." Ieee Transactions on Signal Processing **46**(4): 1027-1042.

MathWorks. (2011). "Identifying Significant Features and Classifying Protein Profiles." from http://www.mathworks.com/products/bioinfo/demos.html?file=/products/demos/shipping/bioinfo/cancerdetectdemo.html#10.

MathWorks (2011). "R2011b documentation- Matlab Toolbox: wden."

MathWorks (2011). "R2011b documentation-Matlab Toolbox: waveinfo."

McLachlan, G. J. and J. Wiley (1992). Discriminant analysis and statistical pattern recognition, Wiley Online Library.

Mika, S., et al. (1999). Fisher discriminant analysis with kernels, IEEE**:** 41-48.

Mollochristensen, E. (1984). "Satellite Microwave Remote-Sensing - Allan,Td." Bulletin of the American Meteorological Society **65**(4): 386-387.

Netherway, D. J., et al. (1989). Reduction of Some Environmental Effects that Degrade the Performance of HF Skywave Radars**:** 288-292.

Paduan, J. D. and L. K. Rosenfeld (1996). Remotely sensed surface currents in Monterey Bay from shore-based HF radar (Coastal Ocean Dynamics Application Radar), American Geophysical Union. **101**.

Prandle, D., et al. (1992). "Tidal flow through the straits of dover." J. Phys. Oceanogr. **23**: 23-37.

Preisendorfer, R. W., et al. (1988). "The Principal Discriminant Method of Prediction - Theory and Evaluation." Journal of Geophysical Research-Atmospheres **93**(D9): 10815-10830.

Regalia, P. A. (2005). "Fundamentals of Adaptive Filtering [Book Review]." Control Systems Magazine, IEEE **25**(4): 77-79.

Rice, S. O. (1951). Reflection of electromagnetic waves from slightly rough surfaces, Wiley Subscription Services, Inc., A Wiley Company New York. **4**.

Rush, C. (1986). "Ionospheric radio propagation models and predictions--A mini-review."

Antennas and Propagation, IEEE Transactions on **34**(9): 1163-1170.

S. Cruces, et al. (2000). Blind Source Extraction in Gaussian Noise. proceedings of the 2nd International Workshop on Independent Component Analysis and Blind Signal Separation (ICA'2000), Helsinki, Finland.

Schaffer, C. (1993). "Selecting a Classification Method by Cross-Validation." Machine Learning **13**(1): 135-143.

Shearman, E. D. R. and M. D. Moorhead (1988). Pisces: A Coastal Ground-wave HF Radar For Current, Wind And Wave Mapping to 200 Km Ranges. Geoscience and Remote Sensing Symposium, 1988. IGARSS '88. Remote Sensing: Moving Toward the 21st Century., International.

Skolnik, M., I (1970). Radar handbook, McGraw-Hill.

Skolnik, M. I. (1990). Radar handbook, McGraw-Hill New York.

Stove, A. G. (1992). Linear FMCW radar techniques. **139:** 343-350.

Theodoridis, S. and K. Koutroumbas (1999). Pattern recognition, Academic Press.

Tian, W., et al. Ionosphere interference suppression for HFSWR using L-shape array. Signal Processing Systems (ICSPS), 2010 2nd International Conference on.

Turley, M. D. E. and D. J. Netherway (1990). OTHR signal reconstruction for data corrupted by impulsive noise**:** 203-209.

Vizinho, A. and L. R. Wyatt (1996). "Modern spectral analysis in HF radar remote sensing." Oceans '96 Mts/Ieee, Conference Proceedings, Vols 1-3 / Supplementary Proceedings: 1500-1503.

Wait, J. R. (1966). Theories of prolate spheroidal antennas(Prolate spheroidal antennas, discussing spheroidal wave functions, confocal plasma sheath effects and influence of DC magnetic field). **1:** 475-512.

Walker, G. (1931). "On periodicity in series of related terms." Proceedings of the Royal Society of London Series a-Containing Papers of a Mathematical and Physical Character **131**(818): 518-532.

Wan, X., et al. (2006). "Experimental trials on ionospheric clutter suppression for high-frequency surface wave radar." IEE Proc. Radar, Sonar and Navig. **153**(1): 23-29.

Wan, X., et al. (2005). "Adaptive cochannel interference suppression based on subarrays for HFSWR." IEEE Signal Process. Lett. **12**(2): 162-165.

Wand, M. P. and M. C. Jones (1994). "Multivariate Plug-in Bandwidth Selection." Computational Statistics **9**(2): 97-116.

Wand, M. P. and M. C. Jones (1995). Kernel smoothing. London, Chapman and Hall.

Wang, S. T., et al. (2005). "Possibility theoretic clustering." Advances in Intelligent Computing, Pt 1, Proceedings **3644**: 849-858.

Wang, W. and L. R. Wyatt (2011). "Radio frequency interference cancellation for sea-state remote sensing by high-frequency radar." Radar, Sonar & Navigation, IET **5**(4): 405-415.

Wear, K. A., et al. (1995). "A Comparison of Autoregressive Spectral Estimation Algorithms and Order Determination Methods in Ultrasonic Tissue Characterization." Ieee Transactions on Ultrasonics Ferroelectrics and Frequency Control **42**(4): 709-716.

Weber, B. L. and D. E. Barrick (1977). On the nonlinear theory for gravity waves on the ocean's surface. Part I: Derivations, American Meteorological Society. **7:** 3-10.

Weber, B. L. and D. E. Barrick (1977). "On the nonlinear theory for gravity waves on the ocean's

surface.Part I: Devations." J. Phys. Oceanogr. **7**: 3-10.

Wikipedia. (2011). "Kernel density estimation." from http://en.wikipedia.org/wiki/Kernel_density_estimation.

Wikipedia. (2011). "Multivariate kernel density estimation." from http://en.wikipedia.org/wiki/Multivariate_kernel_density_estimation.

Wu, S. C., et al. (2001). "Test of HF ground wave radar OSMAR2000 at the Eastern China Sea." Oceans 2001 Mts/Ieee: An Ocean Odyssey, Vols 1-4, Conference Proceedings: 646-648.

Wyatt, L. (1987). Ocean wave parameter measurement using a dual-radar system- A simulation study. **8:** 881-891.

Wyatt, L. R. (1989). Measuring the ocean wave directional spectrum with HF radar: THe inversion problem. Mathematics in Remote Sensing, Oxford, UK, Clarendon.

Wyatt, L. R. (1995). High order nonlinearities in HF radar backscatter from the oceansurface. **142:** 293-300.

Wyatt, L. R. (2000). "Limits to the Inversion of HF Radar Backscatter for Ocean Wave Measurement." Atmospheric and oceanic technology **17**: 1651-1665.

Wyatt, L. R. (2002). "An evaluation of wave parameters measured using a single HF radar system." Canadian J. Remote Sens. **28**(2): 205-218.

Wyatt, L. R., et al. (2006). "Operational wave, current, and wind measurements with the Pisces HF radar." IEEE J. Ocean. Eng. **31**(4): 819-834.

Wyatt, L. R., et al. (2006). Operational Wave, Current, and Wind Measurements With the Pisces HF Radar. **31:** 819-834.

Wyatt, L. R. and G. J. Holden (1992). Developments in ocean wave measurement by HF radar. **139:** 170-174.

Wyatt, L. R., et al. (1997). "Maximum likelihood estimation of the directional distribution of 0.53Hz ocean waves." Journal of Atmospheric and Oceanic Technology **14**: 591-603.

Wyatt.L.R. (1990). "A relaxation method for integral inversion applied to HF radar measurement of the ocean wave directional spectrum." Int. J. Remote sensing **11**(8): 1481-1994.

Xin, G., et al. (2008). Performance evaluation of radio frequency interference cancellation techniques in HFSWR. Proc. IEEE Radar 2008 Conf., Rome, Italy.

Yule, G. U. (1927). "On a method of investigating periodicities in disturbed series, with special reference to Wofer's sunspot numbers." Philosophical Transactions of the Royal Society of London Series a-Containing Papers of a Mathematical or Physical Character **226**: 267-298.

Zhang, L., et al. (2006). "Time-varying Autoregressive Model and Its Application to Nonstationary Vibration Signal Analysis." JOURNAL OF VIBRATION AND SHOCK **25**(6).

Zhang, S. W., et al. (2007). "A detection method of radar signal by wavelet transforms." Fourth International Conference on Fuzzy Systems and Knowledge Discovery, Vol 2, Proceedings: 710-714.

Zhou, H., et al. (2005). "Dense radio frequency interference suppression in HF radars." IEEE Signal Process. Lett. **12**(5): 361-364.