

國立臺灣大學理學院海洋研究所

碩士論文

Institute of Oceanography

College of Science

National Taiwan University

Master Thesis



高頻雷達觀測流場補缺值問題之探討

—以台灣東北部海域為例

A Study on Data Filling from Incomplete Dataset of HF

Radar Measured Ocean Currents

—A Case Study of the Flow Field Northeast of Taiwan

陳俞玟

Yu-Wen Chen

指導教授：王胃 博士、詹森 博士

Advisor: Joe Wang, Ph.D.; Sen Jan, Ph.D.

中華民國 102 年 7 月

July, 2013



國立臺灣大學碩士學位論文
口試委員會審定書

高頻雷達觀測流場補缺值問題之探討
—以台灣東北部海域為例

A Study on Data Filling from Incomplete Dataset of HF Radar
Measured Ocean Currents
—A Case Study of the Flow Field Northeast of Taiwan

本論文係陳俞鈺君（R00241101）在國立臺灣大學海洋所完成之
碩士學位論文，於民國 102 年 7 月 29 日承下列考試委員審查通過及口
試及格，特此證明

口試委員：

王 曾

（簽名）

（指導教授）

詹 森

楊 頌 堅

陳 米 丁

系主任、所長

戴 昌 鳳

（簽名）

（是否須簽章依各院系所規定）

誌謝



首先感謝指導教授王胄老師的耐心指導，無論是在學術研究、課業或是完成論文等方面都給予我極大的幫助，在研究態度、邏輯思考等方面也從老師的教誨中受益良多。感謝口試委員陳慶生老師、詹森老師以及楊穎堅老師提供許多珍貴的意見與錯誤的指正，使本文更加完善。另外還要感謝海洋所的老師們在課業上的教導，讓我學習到許多專業知識並奠定研究的基礎，由衷感謝。

感謝實驗室蔡雅鈴學姊、郭怡君學姊在日常生活以及學業上的照顧與解惑，也要感謝 122 實驗室的各位學長及助理在平日的協助，特別謝謝郭天俠學長在許多事情的幫助與鼓勵，尤其是時常督促我的研究進度，使本文能順利完成，非常感謝。同時要感謝我在碩士生涯中的好伙伴王心秀與廖康淵，不管是在課業、研究或是平日生活方面，都給予我相當多的幫助以及陪伴。還要感謝其他陪我度過這段研究生生活的人，感激不盡。

最後感謝我所有的家人以及朋友們，時常關心、鼓勵我，並陪伴、支持我完成學業，謝謝。

摘要



高頻雷達是近代廣用的一種即時監測海流的遙測工具，但高頻雷達訊號常因環境雜訊以及電離層等干擾影響，會造成在觀測區內出現海流觀測值缺漏的情況，因此如何有效填補缺漏值即成為海流監測即時作業一個重要的課題。本文使用臺灣大學海洋所、海軍軍官學校以及海軍大氣海洋局共同合作在臺灣東北部海域建置的 CODAR 雷達系統觀測資料做為研究的根據。

我們選用實向量經驗正交函數(real-vector EOF)及 Karhunen-Loève 展開法(KLE)兩種統計方法探討流場填補缺漏值問題。這兩種方法均是根據長期觀測資料之變量統計求出對應之特徵向量(eigenvector)作為模組(mode)基底，而觀測資料在各基底之投影即為該模組之振幅(amplitude)。經比較分析後，我們選用前 20 組模組(可解釋 96%以上的總變異量)，分別使用最小平方法(least square)與迭代法(iteration)以求解不完整資料情形下各模組之振幅，並以此進行不同個數缺漏值之填補實驗，以檢驗填補效果及填補誤差，藉以推測使用不同方法之缺漏個數上限。

填補實驗的統計結果顯示，以最小平方法進行填補實驗時，在缺漏個數小於資料點總數之 57%的情況下，其誤差隨缺漏點個數的變化並不大，若以迭代法搭配實向量 EOF 法或是 KLE 法分別進行填補，則誤差會隨缺漏點個數變多而呈線性增長，不過各方法在空間上之分佈皆甚為類似，在離雷達站較近處誤差較小；四種方法中以最小平方法配合實向量 EOF 法填補後之誤差相對較小，效果較佳。當缺漏個數為資料點總數之 57%時，最小平方法填補後所得之誤差開始變得明顯，當缺漏個數達資料點總數之 71%以上時，其變化趨勢會呈指數成長；而迭代法的實驗結果則是幾乎在所有缺漏情況下，其誤差皆大於最小平方法的誤差。在缺漏個數為資料點總數之 71%以上的情形時，這些方法均不適合用來填補缺漏值。

關鍵字：高頻雷達、CODAR、經驗正交函數、Karhunen-Loève 展開法、填補缺值

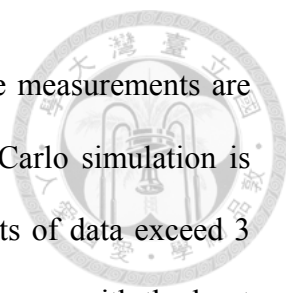
Abstract



The Coastal Ocean Dynamics Applications Radar (CODAR) is a High-Frequency (HF) radar system with compact antennas. Recently, CODAR becomes widely used, for monitoring ocean surface currents remotely in nearly real time; its ability of large coverage on ocean surface and high resolution both in time and in space, makes CODAR an ideal tool for the operational oceanography, such as now-cast of currents and data-assimilation tasks. However, environmental effects such as interferences from obstacles, and/or from ionospheric disturbances, often hampers or weakens the strength of CODAR system, might deteriorate the data quality of CODAR, inducing incomplete datasets with missing data or holes in the designated observation region. The development of appropriate methodology for filling missing data is therefore a necessity deserving for further studies, in prior to the development of operational framework.

In the context, we have analyzed a nearly two-year long dataset of CODAR, which was observed by two HF radars located at Suao and Han-Ben, respectively, and provided by the Surface Current Observations at North-East Taiwan (SCONET) project; basic statistics of the measurements show that the SCONET dataset satisfies the normality and weakly stationary condition. Further processing, by using both modal decomposition methods of Real-vector Empirical Orthogonal Function (REOF) and the Karhunen-Loeve Expansion (KLE), respectively, reveals that almost more than 96% of total variances of currents in the whole observation region can be interpreted by the first 20 modes of both methods. Therefore the first 20 modes of both methods are used for data reconstruction and data filling later.

An independent month-long time series of CODAR measurements is adopted for the data filling experiment, by which the incomplete dataset is generated by depleting



artificially assigned grid points in the original complete dataset, the measurements are therefore treated as the ground truth for later comparison. Monte Carlo simulation is used for the tests of data filling experiment when the missing points of data exceed 3 points. We have used both the EOF and the KLE methods, in accompany with the least square and the iteration procedures for the estimation of the amplitude of each modes, for the study of data filling experiment. Results show that the EOF method in accompany with the least square procedure is the best among the four methodologies, when the percentage of occurrence of the missing data is less than 57% of the whole dataset. However, all these four methods are not adequate for filling incomplete dataset, if the percentage of occurrence of the missing data exceeds 71%.

Key words: HF radar, CODAR, EOF, Karhunen-Loève expansion, data filling

目 錄



口試委員會審定書	i
誌謝	ii
摘要	iii
Abstract	iv
目 錄	vi
圖目錄	viii
表目錄	xii
第一章 緒論	1
1.1 高頻雷達測流簡介	1
1.2 臺灣東北海域 CODAR 系統簡介	4
1.3 缺漏值填補相關之文獻回顧	8
1.4 論文內容架構	9
第二章 資料統計特性	11
2.1 資料來源	11
2.2 時間域基本統計	18
2.3 頻率域分析以及濾潮處理	25
2.4 平穩性檢定	30
第三章 模分析方法與分析結果	35
3.1 經驗正交函數	35
3.2 Karhunen-Loève 展開法	47
3.3 以模重組法填補資料缺漏	58
3.3.1 最小平方法	58
3.3.2 迭代法	60

第四章 模擬實驗與結果討論	62
4.1 實驗設計	62
4.2 窮舉法實驗：任意 1~3 個資料點缺漏	66
4.3 蒙地卡羅法之實驗：多個資料點出現缺漏	74
4.4. 綜合討論	79
第五章 結論	85
參考文獻	87



圖目錄



圖 1-1 (a)雷達波與波浪作用示意圖;(b)上:無海流作用時的雷達回波頻譜圖,下: 有海流作用時的頻譜圖(縱軸表示回波能量密度,橫軸為頻率)(引自 Barrick et al., 1977)。	2
圖 1-2 美國加州利用 CODAR 雷達並配合 AVHRR 衛星在 Monterey Bay 觀測到的 海灣內及外海的海流和海表溫(引自 Paduan and Rosenfeld, 1996)。	3
圖 1-3 臺灣周圍海底地形變化示意圖(海底地形資料來源:國科會海洋學門資料 庫)。	4
圖 1-4 臺灣四周海域表層海流即時觀測網各 CODAR 雷達站站位分佈示意圖(站 位地點與名稱係參考國研院臺灣海洋科技中心海洋資料庫之 CODAR 資料庫)。	5
圖 1-5 蘇澳及漢本二雷達站(紅色方塊)及合成海流資料點位(深橘色十字)示意圖 (海底地形資料來源:國科會海洋學門資料庫)。	6
圖 1-6 合成海流資料點位之資料缺漏數量(組數)統計(2011 年 4 月 14 日至 2012 年 12 月 31 日,每組為一小時)。	7
圖 2-1 2012 年 4 月 30 日 1200UTC 蘇澳站短時間互能譜資料檔(CSS File)中三支 天線回波的自能譜(auto-spectra)。橫軸為頻率(中央為 0,左側為負,右側為正), 縱軸為與接收天線的水平距離(km),顏色越亮表示回波訊號越強。	12
圖 2-2 蘇澳站 2012 年 4 月 30 日 1200UTC 徑向海流分佈。徑向海流速度即是指 觀測區內某點海流流速指向該點與雷達站連線之射線上的流速分量,也就是 海流接近或遠離雷達站的流速分量。	12
圖 2-3 2012 年 4 月 30 日 1200UTC 的合成海流流速分佈圖,係由蘇澳與漢本二站 徑向海流資料經中控站系統結合而成。	13
圖 2-4 合成海流資料點位出現資料缺漏情況之統計(2011 年 4 月 14 日至 2012 年	

12 月 31 日)：(a)資料缺漏之百分比分佈；(b)不完整資料中缺漏點數之百分比統計直方圖；(c)長期統計資料缺漏百分比之平均日變化。.....	15
圖 2-5 資料點位編號之相對位置圖。.....	21
圖 2-6 觀測區內之平均流場及水深 20 公尺之長期平均流場(海底地形與長期平均流場資料來源：國科會海洋學門資料庫)。.....	21
圖 2-7 CODAR 觀測海流資料之標準差分佈：(a)東西向流速；(b)南北向流速。.....	22
圖 2-8 (a)東西向流速分佈圖；(b)南北向流速分佈圖；(c)東西向流速 P-P 圖；(d)南北向流速 P-P 圖。紅線表示常態分佈曲線。.....	24
圖 2-9 中央氣象局蘇澳站(121.87°E、24.59°N) 2011/8/16~2011/11/15 的(a)潮位觀測資料及(b)各分潮振幅。.....	25
圖 2-10 (a)第 5 點位、(b)第 69 點位之流速動能譜圖。.....	27
圖 2-11 潮流能量空間分佈圖：(a)全日潮分量；(b)半日潮分量。.....	28
圖 2-12 觀測區內第 1 個資料點海流資料(2011 年 8 月 16 日至 11 月 15 日)經調和分析後得到的各分潮潮流橢圓長軸係數。.....	29
圖 2-13 流速速率資料之平穩性檢定結果，彩色區塊表示通過檢定，表示資料具有穩定性，而黑色區塊表示未通過檢定：(a)平均值之檢定結果；(b)標準差之檢定結果。.....	32
圖 3-1 實向量 EOF 分析法重建流場之特徵值分佈圖與累積特徵值分佈圖：(a)全部 140 個模組；(b)前 20 個模組。.....	40
圖 3-2 實向量 EOF 法前 20 個模組重建流場與變動流場之變異數百分比空間分佈圖。.....	40
圖 3-3 實向量 EOF 分析法求出之第 1 模組：(a)模向量空間分佈；(b)模振幅之時間序列。.....	42
圖 3-4 實向量 EOF 分析法求出之第 2 模組：(a)模向量空間分佈；(b)模振幅之時	

間序列。	43
圖 3-5 實向量 EOF 分析法求出之第 3 模組：(a)模向量空間分佈；(b)模振幅之時間序列。	44
圖 3-6 原始觀測流場與實向量 EOF 法前 20 個模組重建流場之偏差均方根分佈圖。	46
圖 3-7 KLE 分析法重建流場之振幅平方和分佈圖與累積變異數分佈圖：(a)全部 140 個模組；(b)前 20 個模組。	51
圖 3-8 KLE 法前 20 個模組重建流場與變動流場之變異數百分比空間分佈圖。	52
圖 3-9 KLE 分析法求出之第 1 模組：(a)模向量空間分佈；(b)模振幅之時間序列。	54
圖 3-10 KLE 分析法求出之第 2 模組：(a)模向量空間分佈；(b)模振幅之時間序列。	55
圖 3-11 KLE 分析法求出之第 3 模組：(a)模向量空間分佈；(b)模振幅之時間序列。	56
圖 3-12 原始觀測流場與 KLE 法前 20 個模組重建流場之偏差均方根空間分佈圖。	57
圖 4-1 蒙地卡羅法前級試算結果：假設缺漏個數為 5，以最小平方法搭配實向量 EOF 法填補，橫軸為實驗次數，縱軸為填補值與真實值之誤差均方根平均值。	64
圖 4-2 任意 1 個資料點缺漏情形之填補實驗結果：(a)以最小平方法配合實向量 EOF 法；(b)以最小平方法配合 KLE 法；(c)以迭代法配合實向量 EOF 法；(d)以迭代法搭配 KLE 法填補之誤差均方根空間分佈。	70
圖 4-3 任意 1 個資料點缺漏情形之填補實驗結果(已扣除 4.1 節所述第 1 種誤差)：(a)以最小平方法配合實向量 EOF 法；(b)以最小平方法配合 KLE 法；(c)以迭代法配合實向量 EOF 法；(d)以迭代法搭配 KLE 法填補之誤差均方根空間分佈。	

佈。	73
圖 4-4 任意 10 個資料點缺漏情形之填補實驗結果：(a)以最小平方法配合實向量 EOF 法；(b)以最小平方法配合 KLE 法；(c)以迭代法配合實向量 EOF 法；(d)以迭代法搭配 KLE 法填補之誤差均方根空間分佈。	75
圖 4-5 任意 10 個資料點缺漏情形之填補實驗結果(已扣除 4.1 節所述第 1 種誤差)：(a)以最小平方法配合實向量 EOF 法；(b)以最小平方法配合 KLE 法；(c)以迭代法配合實向量 EOF 法；(d)以迭代法搭配 KLE 法填補之誤差均方根空間分佈。	76
圖 4-6 任意 35 個資料點缺漏情形之填補實驗結果：(a)以最小平方法配合實向量 EOF 法；(b)以最小平方法配合 KLE 法；(c)以迭代法配合實向量 EOF 法；(d)以迭代法搭配 KLE 法填補之誤差均方根空間分佈。	77
圖 4-7 任意 35 個資料點缺漏情形之填補實驗結果(已扣除 4.1 節所述第 1 種誤差)：(a)以最小平方法配合實向量 EOF 法；(b)以最小平方法配合 KLE 法；(c)以迭代法配合實向量 EOF 法；(d)以迭代法搭配 KLE 法填補之誤差均方根空間分佈。	78
圖 4-8 以實向量 EOF 法、KLE 法與最小平方法、迭代法進行填補不同缺漏個數實驗結果：(a)填補值與真實值之誤差均方根平均值；(b)填補值與 20 個模組重建流場之誤差均方根平均值分佈。	80
圖 4-9 以實向量 EOF 法、KLE 法與最小平方法、迭代法進行填補不同缺漏個數實驗結果：(a)填補值與真實值之誤差均方根標準差；(b)填補值與 20 個模組重建流場之誤差均方根標準差分佈。	81
圖 4-10 實向量 EOF 法與 KLE 法誤差均方根平均值之比較。	82
圖 4-11 F 檢測：比較(a)最小平方法；(b)迭代法之結果。	84

表目錄



表 2-1	合成海流資料中發生缺漏情形(共 429 組)之統計。	16
表 2-2	長期統計資料缺漏百分比之平均日變化。	17
表 2-3	各點位之東西向流速(U)與南北向流速(V)的基本統計。	19
表 2-4	全部資料東西向流速(U)與南北向流速(V)的基本統計資料。	23
表 2-5	各點位之東西向流速(U)、南北向流速(V)及速率(speed)的連個數。	33
表 3-1	原始觀測流場與實向量 EOF 法前 20 個模組重建流場之偏差均方根數值。	46
表 3-2	原始觀測流場與 KLE 法前 20 個模組重建流場之偏差均方根數值。	57
表 4-1	70 個資料點不同缺漏個數情況之變化種數。	63
表 4-2	填補任意 1 個資料點為缺漏情形之實驗結果。	68
表 4-3	填補任意 1 個資料點為缺漏情形之實驗結果(扣除空間上之不均勻性)。	71
表 4-4	顯著水準 $\alpha=0.05$ 時與各自由度相對應之信賴區間。	79
表 4-5	F 分配表中，顯著水準為 α 及自由度為 n 時的上、下限數值。	83

第一章 緒論



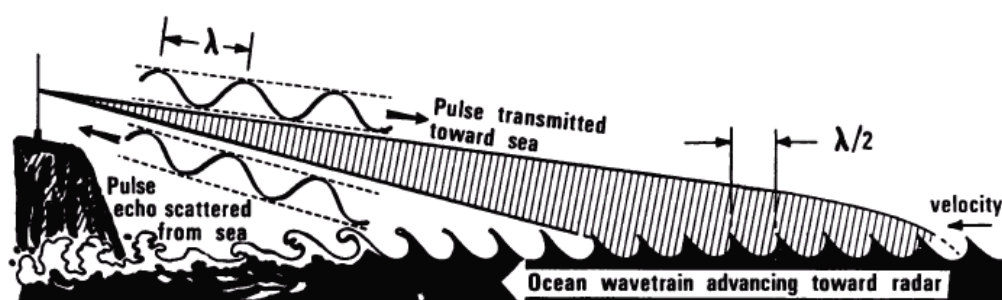
1.1 高頻雷達測流簡介

隨著文明以及經濟發展，人類社會對於海上活動的需求以及頻度均大幅增加。基於安全以及經濟效益等因素考量，海上活動諸如航運、遊憩、救難搜索或是污染追蹤等都需要參考海洋的資訊，特別是海面流場變化更是其中重要的一環。海表面流場除了受背景海流場之作用外，還會受到波浪、風場、潮汐等外在因素的影響，其變化十分複雜。海流觀測傳統上是利用漂流浮標(drifter)，或是錨碇(mooring)海流儀(如都卜勒流速儀 Acoustic Doppler Current Profiler, ADCP，或其他型式海流儀)等方法，缺點是在時間和空間尺度上難以同時兼顧。近代發展出的海流遙測技術，例如岸基高頻雷達(shore-based high-frequency radars)，則具有觀測資料空間涵蓋範圍大(尺度從數百公尺到數百公里)和時間解析度高(可以小至數分鐘)的優點，目前雖然其量測精確度尚不能與傳統方法相比，但對作業化海流觀測而言這已是一種可行的觀測工具(Barrick et al., 1977; Paduan and Rosenfeld, 1996)。

國外運用高頻(high-frequency, HF)雷達觀測海流之研究已發展了近 40 年之久(Teague et al., 1997)，其遙測原理如圖 1-1 所示：由雷達天線發射出的電磁波在傳播過程中接觸到海面粗糙構造時會產生後向散射(backscattering)，後向散射的回波其性質與會影響到海面粗糙度特性和分佈的物理因子(如波浪、海流等)之間有密切的關係，所以利用雷達回波的訊號可以反推出這些物理因子的分佈特性(Barrick et al., 1977; Paduan and Rosenfeld, 1996)。後向回波訊號主要是由布拉格散射(Bragg scattering)所造成，也就是當波長為 λ 的雷達波遇到海面上波長為 $\frac{n\lambda}{2}$ ($n=1, 2, 3, \dots$) 的水波時，雷達後向散射回波會發生建設性的疊加效應，因此回波信號就會較強(圖 1-1a)，其中尤以 $n=1$ 時為最強。當海面無流時，雷達回波頻譜圖上會在此 $\frac{\lambda}{2}$ 波長水波所對應之頻率處出現能量尖峰(如圖 1-1b 上圖)；但當海面有流時，由於海流

會對水波產生都卜勒頻移(Doppler shift)，因此在雷達回波頻譜圖上能量尖峰頻率亦會發生偏移(圖 1-1b 下圖)，故從雷達回波的頻移可以反算出表面海流的大小。不過單一雷達只能測到雷達天線徑向方向上的海流速度分量，需要二正交之分量才能結合成為海流速度向量，是以至少要結合兩座以上的雷達才能得到共同涵蓋範圍內的海流向量流場資訊(Barrick et al., 1977; Paduan and Rosenfeld, 1996)。

(a)



(b)

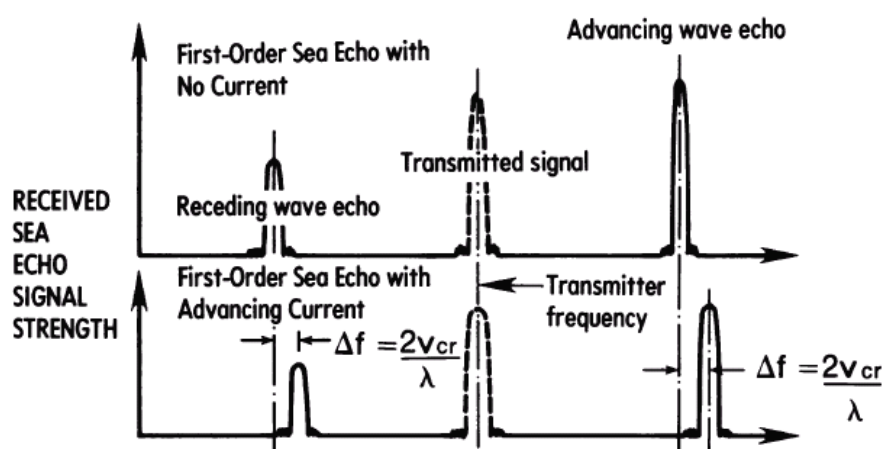


圖 1-1 (a)雷達波與波浪作用示意圖;(b)上:無海流作用時的雷達回波頻譜圖,下:有海流作用時的頻譜圖(縱軸表示回波能量密度,橫軸為頻率)(引自 Barrick et al., 1977)。

Emery et al. (2004)指出目前使用的高頻雷達測流系統大致可分成兩類:其一為波束形成(beam forming)雷達,例如 Wellen radar (WERA; Gurgel et al., 1999);另一種則是測向(direction finding)雷達,如 Coastal Ocean Dynamics Application Radar (CODAR)。其中 CODAR 系統因為體積小便於攜帶(Barrick et al., 1977),且較早商

業化(1984 年即已開始；參見 CODAR Ocean Sensors 網站，<http://www.codar.com/>)，目前使用較廣泛(Fang et al., 2011)。Paduan and Rosenfeld (1996)曾在美國加州的 Monterey Bay 地區利用 1994 年的 CODAR 雷達資料與 AVHRR 衛星資料所觀測到海灣及外海的海表溫相互比對(如圖 1-2)，其結果顯示 CODAR 測得的平均流場(黑色箭頭)在灣口處呈現一股往南的海流，而灣內的海表溫也同樣呈現往南方擴展的狀況，這種分佈態勢相當符合當地歷史水文資料統計特性，因此他們認為高頻雷達測流系統在長期觀測以及即時監測海流等作業方面具有很大的發展潛力。

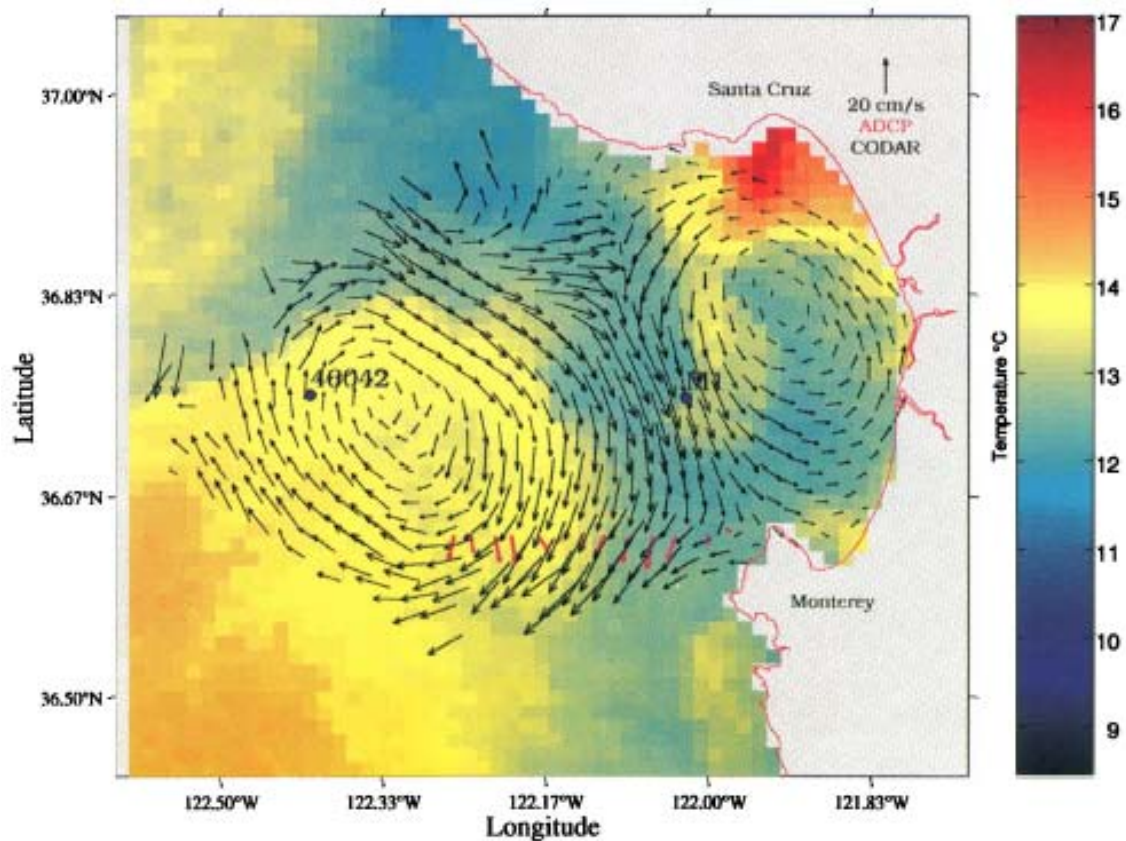


圖 1-2 美國加州利用 CODAR 雷達並配合 AVHRR 衛星在 Monterey Bay 觀測到的海灣內及外海的海流和海表溫(引自 Paduan and Rosenfeld, 1996)。



1.2 臺灣東北海域 CODAR 系統簡介

臺灣位於東亞海運樞紐，週遭有幾道重要海流，如黑潮、黑潮支流、大陸沿岸流、南海暖流等通過，另外鄰近海域海底地形崎嶇(圖 1-3)，風場多變，更增加了流況的複雜性(Jan et al., 2002)。由於運用 CODAR 雷達即時監測海流可以獲得空間以及時間解析度均甚高的流況資料，我國海洋學界多年前即倡議使用此系統來建立臺灣四周海域表層海流即時觀測網，所獲資料除可強化海洋資料庫外並可結合數值模式發展預報技術，以提供各界在研究、防災、救難及海上休憩活動等方面上的需求，目前觀測網建置作業已初步完成(環台各 CODAR 雷達站分佈如圖 1-4，參見國研院臺灣海洋科技中心網站，<http://www.tori.org.tw/>)。

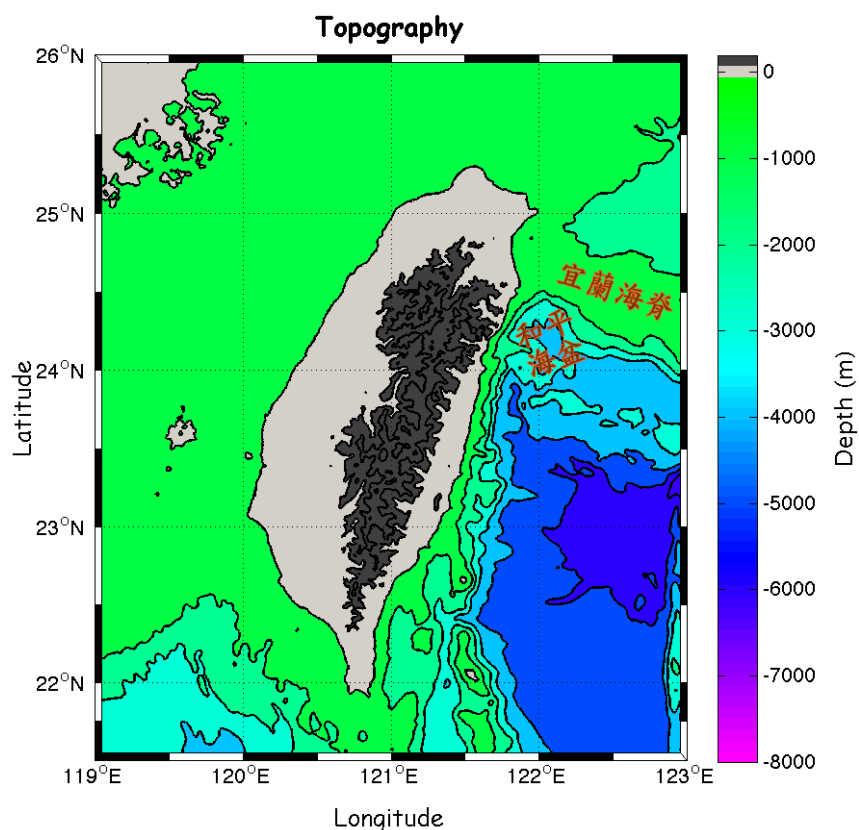


圖 1-3 臺灣周圍海底地形變化示意圖(海底地形資料來源：國科會海洋學門資料庫)。

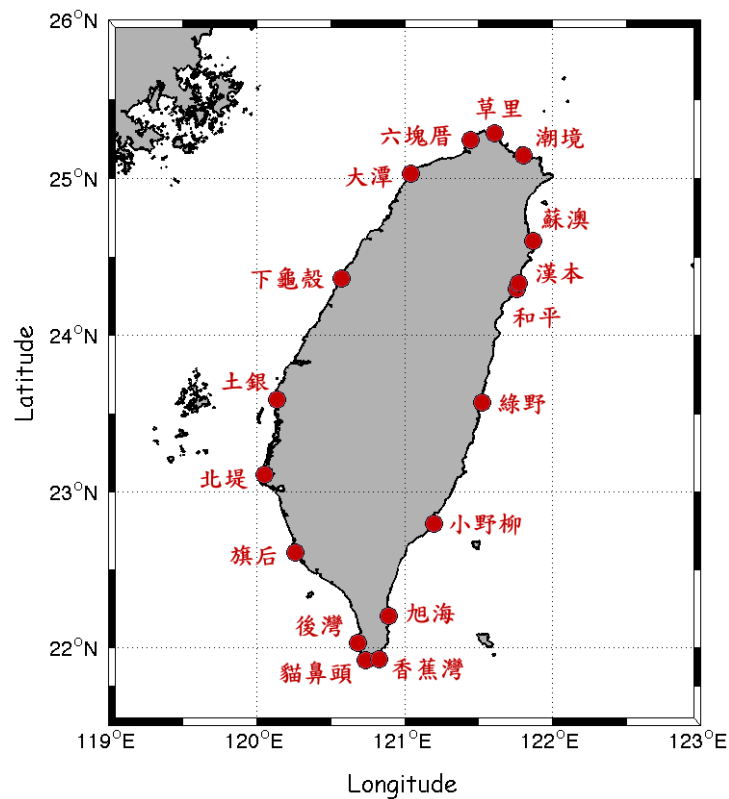


圖 1-4 臺灣四周海域表層海流即時觀測網各 CODAR 雷達站站位分佈示意圖(站位地點與名稱係參考國研院臺灣海洋科技中心海洋資料庫之 CODAR 資料庫)。

在圖 1-4 所示之環臺 CODAR 觀測站中，蘇澳與漢本二站係由國立臺灣大學海洋所、海軍軍官學校以及海軍大氣海洋局共同合作建置並維持，詳細情形可參見「臺灣東北海域海流觀測」(Surface Currents Observation North East of Taiwan, SCONET)網頁(<http://www-codar.oc.ntu.edu.tw>)。二站量測區涵蓋範圍包括臺灣東北部宜蘭海脊以及和平海盆交界處一帶海域，當地海底地形變化相當劇烈(圖 1-3)，亦是黑潮流經之處。二站中，蘇澳站(Suao)位於北方澳海邊的海岬上，經度 121.8825°E ，緯度 24.6002°N ，漢本站(Han-Ben)則位於和平溪北岸海邊，經度 121.7695°E ，緯度 24.3301°N ；二雷達站均使用長距型 CODAR 系統(Long-Range SeaSonde)，各具有一支直立的發射天線以及三支相互垂直的接收天線，所用頻率均為 4.3~5.4 MHz，空間解析度則取決於所用之頻寬，目前蘇澳與漢本二站均為 8.16 km (Fang et al., 2011)。

雖然長距型 CODAR 系統在最佳情況下理論上的量測距離可達 220 km，但 Fang et al. (2011)發現蘇澳、漢本二站均會受到與那國島(Yonaguni，與二站距離均為 110 km 左右)以及電離層干擾影響，尤其是後者在本地時間晚上 6 點到凌晨 2 點之間其效應最為嚴重。此外，雷達天線場型也會受到周遭環境的影響，若是雷達接收天線設置的位置不佳，例如附近有較高的樹木或是建築物等之類的高大物體，都會對天線場型產生不良效應(Kohut and Glenn, 2003; Ramp et al., 2008)，有可能導致系統在估算流速時發生誤差(Fang et al., 2011)。Fang et al. (2011)經由海流比對、校正，並依資料統計結果以及考慮兩站徑向交角大小等效應後，將蘇澳和漢本二站 CODAR 系統合成海流之產出結果限定在如圖 1-5 所示之 70 個站點內。在此觀測範圍內約有 26 個觀測點位於宜蘭海脊上，另外 44 個觀測點則位於較深處的和平海盆。為了簡化，以下將蘇澳、漢本二站 CODAR 系統之觀測海域簡稱為 SCONET 海域。

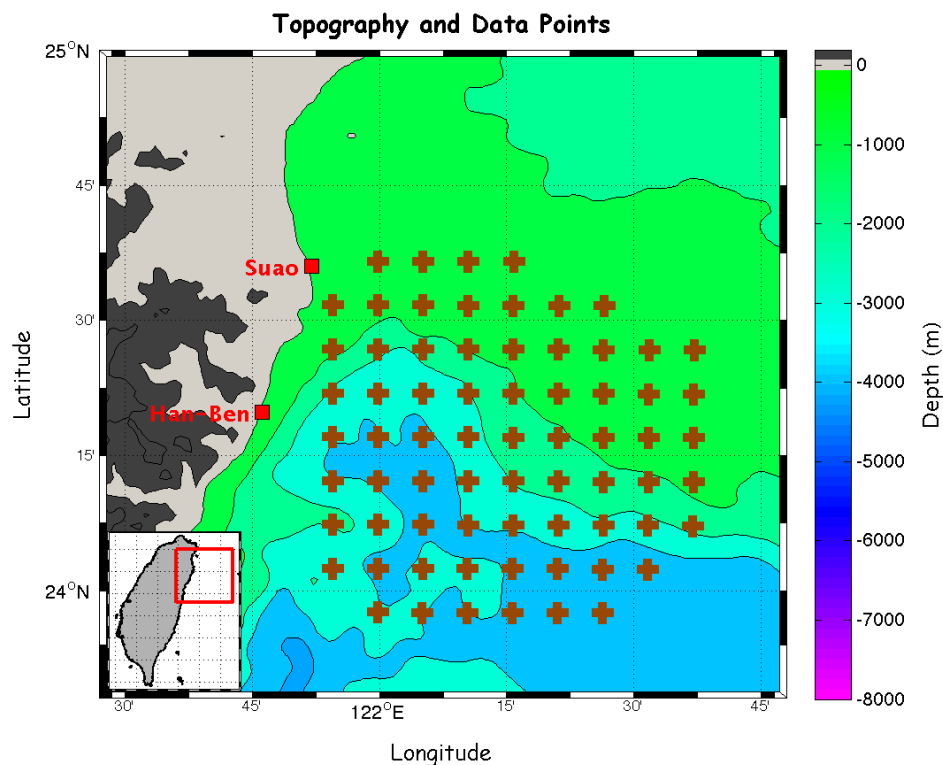


圖 1-5 蘇澳及漢本二雷達站(紅色方塊)及合成海流資料點位(深橘色十字)示意圖(海底地形資料來源：國科會海洋學門資料庫)。

蘇澳、漢本二站自 2011 年 4 月正式開始連線觀測以來，迄 2012 年 12 月所收到的海流資料中，不免會因雷達站斷電、儀器故障，或是電離層、環境雜訊干擾以及雷達波被遮蔽等因素影響，致使回波訊號太弱或接收不到雷達回波訊號，而造成此 70 個觀測點合成海流資料中不時會發生有些點位無資料的情形，此期間內各觀測點之資料缺漏數量統計如圖 1-6 所示。整體而言，此期間內每個觀測點缺漏資料之組數約介於 600~900 組之間，其數量與觀測點和雷達站之距離似乎呈正比關係，距離雷達站越遠處之資料缺漏情形則越嚴重(圖 1-6)。如何填補觀測資料缺漏值是發展高頻雷達海流現報作業化系統過程中一個實際的問題，因此也構成了一個有意義的研究課題。

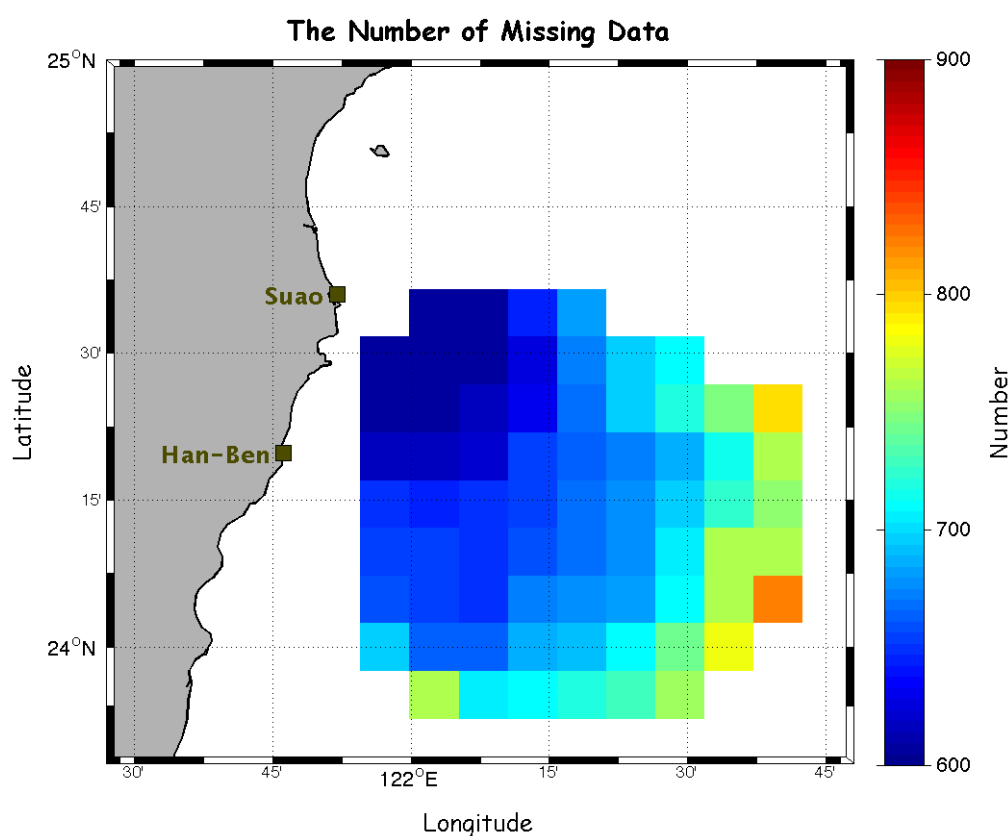


圖 1-6 合成海流資料點位之資料缺漏數量(組數)統計(2011 年 4 月 14 日至 2012 年 12 月 31 日，每組為一小時)。



1.3 缺漏值填補相關之文獻回顧

「如何填補缺漏資料(missing data)」是不同領域的資料分析，特別是對於遙測資料或是一些歷史水文資料而言，這是一個很常見的問題(Beckers and Rixen, 2003; Kondrashov and Ghil, 2006; Oliveira and Gomes, 2010)。造成觀測記錄中出現缺漏值的可能原因包括有儀器故障、有遮蔽物無法獲得資料(例如被雲覆蓋)、受到雜訊干擾等，特別是分析時空變化資料時，有缺漏值往往使傳統方法無法順利處理因而形成問題(Beckers and Rixen, 2003; Kondrashov and Ghil, 2006)。Han and Kamber (2006)曾提出資料清理(data cleansing)的概念，也就是在分析數據時要先進行包括填補缺漏值、消除雜訊及校正數據等前級處理過程，他們同時也提出了缺值填補方法的大方向，例如可以刪去有缺漏值的整筆資料、以平均值來填補，或者是使用以最有效的分析方法所得到之可能性最大的值來填補，後者亦是當下常用的補值方法。


填補缺值可以用參數相依模型(parameter-dependent models)作為模型基底(model-based)來推估缺漏值，例如使用最佳內差法(optimal interpolation)、經驗正交函數(empirical orthogonal function, EOF)、卡爾曼濾波(Kalman filtering) (Reynolds and Smith, 1994; Smith et al., 1996; Kaplan et al., 1997; Kondrashov and Ghil, 2006)，或是使用客觀分析法(objective analysis) (Bretherton et al., 1976)等方法。在這些方法中，經驗正交函數在物理海洋學以及大氣方面應用甚廣，例如 Boyd et al. (1994)以 EOF 法分析 Sargasso Sea 的 CTD 溫度資料，他們分別測試在不同深度有缺值時，使用前 5 個模組(mode)所合成的溫度資料與實際 CTD 量測之溫度資料間的誤差情形，結果發現結合水溫氣候值(climatological values)和 EOF 法合成的資料與 CTD 直接量測值間的誤差會比僅用水溫氣候值來推估之誤差要小；又如 Beckers and Rixen (2003) 選擇義大利半島和巴爾幹半島間的亞得里亞海(Adriatic Sea)中一塊區域做資料填補研究，此區域的 AVHRR 影像常受雲層及雜訊的影響，資料品質並不好，而利用 EOF 法則能夠對原始資料做過濾(filter)和內差補值(interpolation)處理；

另外 Alvera-Azcárate et al. (2005)也是使用亞得里亞海的 AVHRR 資料進行研究，運用在不同的雲覆蓋量，也就是有不同數量的資料缺值情況下，以 EOF 法填補之結果與原始資料相比較，發現填補結果之誤差雖會隨雲量變多而增加，但並不至於增加太多。

此外，Karhunen-Loève 展開法(Karhunen-Loève expansion, KLE) (Preisendorfer, 1988)是一種本質上與 EOF 法非常類似的方法，在流體力學以及工程領域亦有學者將其應用於填補缺漏值方面：例如 Bui-Thanh et al. (2004)使用在機翼表面量測到的壓力資料，經由 KLE 法填補缺漏資料之後，發現 KLE 法之重建資料與計算流體力學模式算出的結果相當吻合；而 Snihir et al. (2006)則是使用 KLE 法預測由不同電流將電池充電後所獲得的開路電壓(open-circuit voltage)曲線，其實驗結果認為此方法在電池管理系統(battery management systems)的降維技術(dimension reduction techniques)上有很大的潛力。

1.4 論文內容架構

由前節的文獻回顧可知，利用參數相依模型為基底之統計方法可以有效過濾、填補缺漏資料，這類方法亦有學者應用於高頻雷達資料上(例如 Emery et al., 2006)。將 CODAR 即時觀測資料裡的缺漏值適當填補後有助於海流資料分析，特別是資料同化(data assimilation)作業。Kohut et al. (1999)在十餘年前即倡議將 CODAR 的即時海流觀測資料進行資料同化後加入數值模式，他們認為如此當可有效提升海洋環境變化預報的正確性。由於經驗正交函數是物理海洋研究上常被使用的方法(Beckers and Rixen, 2003)，此法能有效地將資料特性用簡潔(compact)的形式呈現(Kaihatu et al., 1998)，故有不少學者將其應用在填補缺漏值方面(Boyd et al., 1994; Beckers and Rixen, 2003; Alvera-Azcárate et al., 2005)；另方面，與 EOF 法類似之 Karhunen-Loève 展開法則在工程計算或地球物理領域上使用甚多(Tiampo et al.,



2004; Dong, et al., 2006; Snihir et al., 2006)，但未見用於海洋資料處理方面，有鑑於此，本文將使用此方法處理 CODAR 觀測海流資料，並進一步與 EOF 法之結果相互比較。在以下的章節內，我們將分別探討以 EOF 法以及 KLE 法對 CODAR 海流資料進行模分解(modal decomposition)之效果，並比較以二法重組資料之優缺點，另外亦在不同缺漏值個數的情況之下，檢驗上述二法之填補結果成效與誤差大小，並推測使用此二法時缺漏資料點個數的上限。

本文架構如下，第一章簡介 CODAR 系統之運作原理以及臺灣東北海域 CODAR 觀測作業現況，另外也對前人如何應用統計方法填補缺漏值進行簡要的文獻回顧；第二章介紹 CODAR 雷達觀測海流資料之處理方法、即時資料傳輸流程以及資料統計特性；第三章敘述 CODAR 資料之模分解分析方法與分析結果，比較原始觀測流場與經過模分解後再重建的流場資料之差異度；第四章以真實觀測數據為根據，假設資料出現不同數量的缺漏情況，統計各不同填補方法所得結果與真實值之差異情形；第五章則為結論，比較各方法之優劣與限制後總結全文。

第二章 資料統計特性



在這一章內我們將簡介 SCONET 海域 CODAR 系統處理海流觀測資料的方法與傳輸流程，並說明資料的統計特性，包含平均流場、流速分量機率分佈情形等基本統計。另外，亦對流速資料之常態性以及平穩性進行了統計檢定。

2.1 資料來源

如前章所述，蘇澳與漢本二站均是使用長距型的CODAR高頻雷達測流系統，根據Fang et al. (2011)，二站CODAR系統之作業方式為在接收到回波信號後即先將其快速傅立葉轉換(Fast Fourier Transform, FFT)轉為能譜(spectrum)形式暫存，接著約經每15分鐘累積後自動產生一組互能譜資料檔(Cross Spectrum File, CSQ File)，以及約每30分鐘系統再將前後之CSQ File進行平均，產生一組稱為短時間互能譜的資料檔(Short Time Cross Spectrum File, CSS File) (如圖2-1)。為了有效提升訊噪比，CODAR系統又會以經驗時間長度(一般多使用3小時，但蘇澳與漢本二站係設定使用4小時，Fang et al. 2011)對所收集的CSS File再進行一次平均(Ramp et al., 2008)，並以此作為估算海流的基礎，二站CODAR系統每小時產出的流速資料即是根據這種平均後的能譜所估算，會以雷達站為中心按極座標方式輸出每間隔1度方位角以及徑向上每8.16 km一點之徑向海流資料檔(Radial Velocity File, RUV File) (如圖2-2)。最後再將不同雷達站的同步徑向海流觀測資料回傳到中控站系統並由後者做整合，產出合成表面海流資料檔(Total Velocity File, TUV File) (如圖2-3)，徑向海流資料檔與合成表面海流資料檔皆為每小時輸出一次。

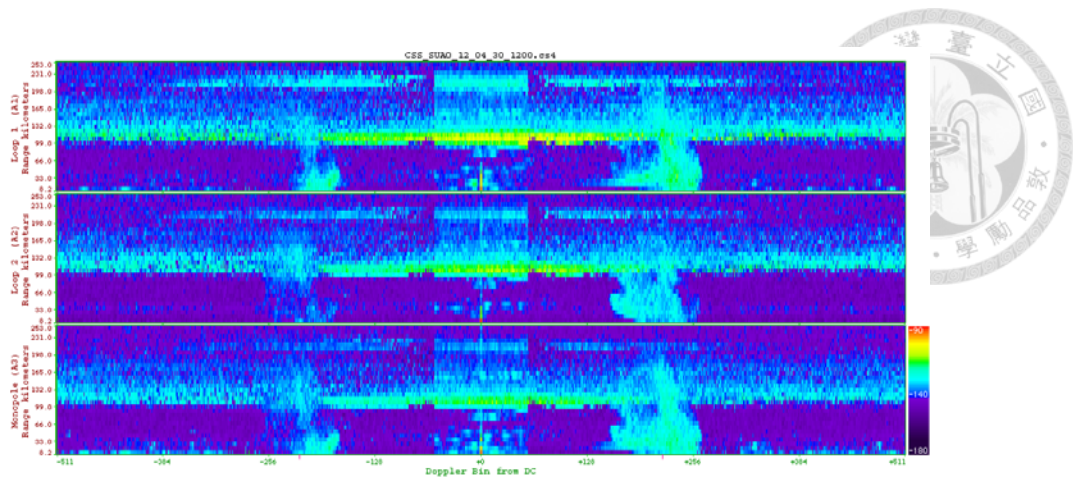


圖 2-1 2012 年 4 月 30 日 1200UTC 蘇澳站短時間互能譜資料檔(CSS File)中三支天線回波的自能譜(auto-spectra)。橫軸為頻率(中央為 0，左側為負，右側為正)，縱軸為與接收天線的水平距離(km)，顏色越亮表示回波訊號越強。

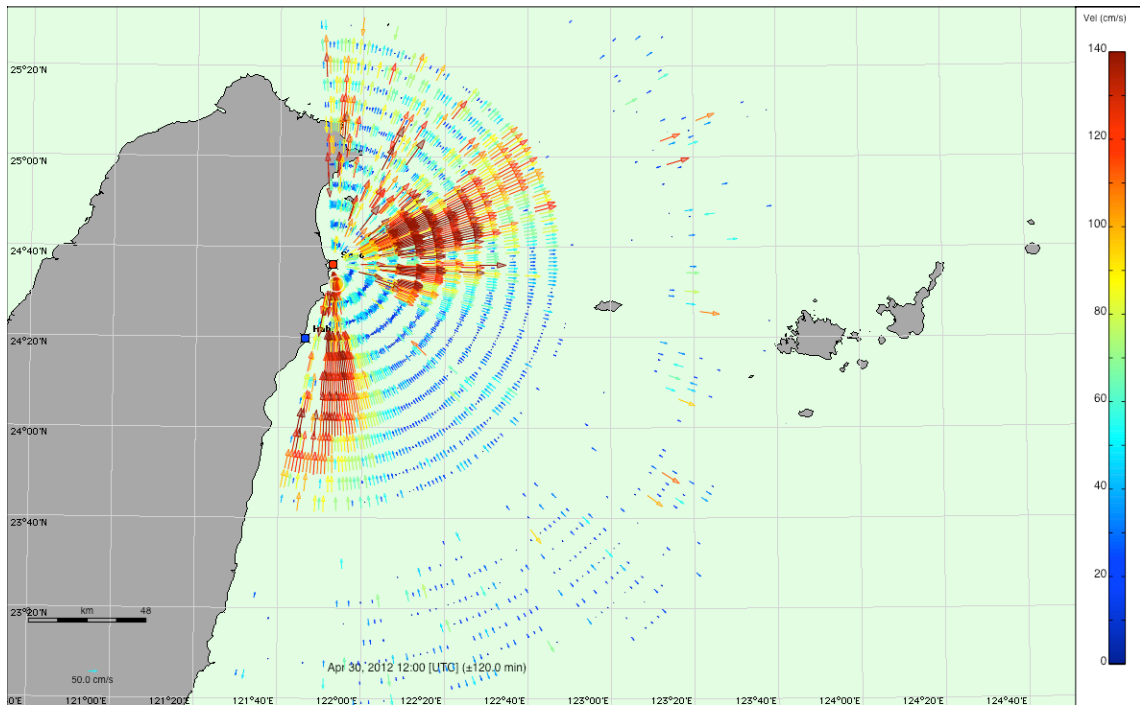


圖 2-2 蘇澳站 2012 年 4 月 30 日 1200UTC 徑向海流分佈。徑向海流速度即是指觀測區內某點海流流速指向該點與雷達站連線之射線上的流速分量，也就是海流接近或遠離雷達站的流速分量。

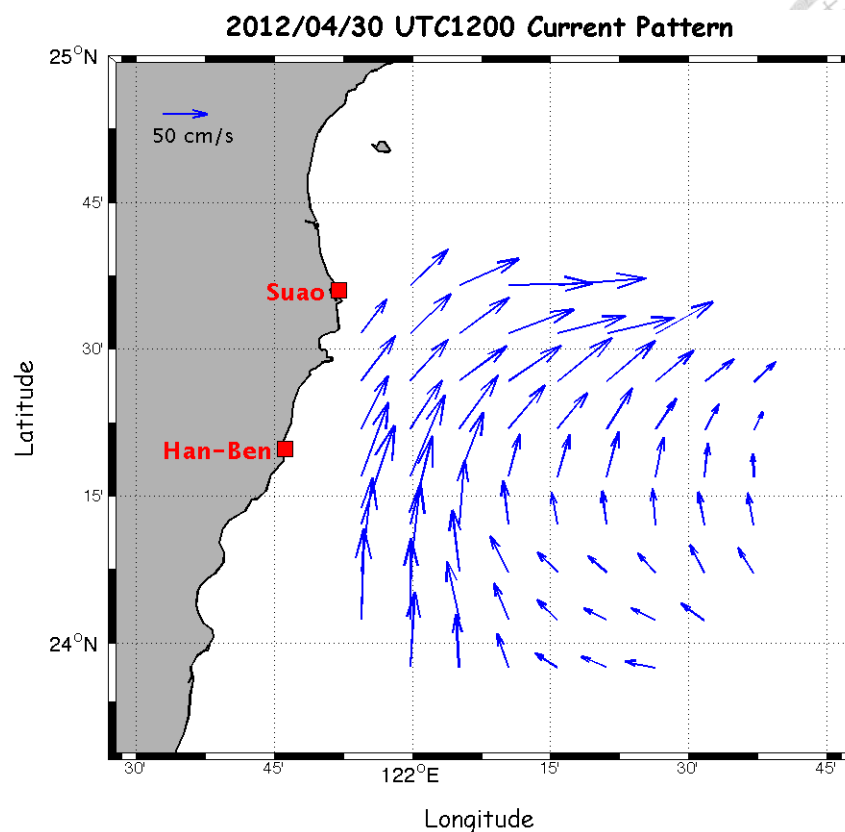



圖 2-3 2012 年 4 月 30 日 1200UTC 的合成海流流速分佈圖，係由蘇澳與漢本二站徑向海流資料經中控站系統結合而成。

Fang et al. (2011)將國科會海洋學門資料庫累積之歷史海流資料與蘇澳、漢本二 CODAR 長期觀測資料相比對，並參考二站互能譜資料中所記錄之資料品質參數統計，最後設定了資料品質相對較好且穩定的 70 個點(如圖 1-5 所示)作為中控站系統產生合成海流場之輸出範圍；我們在以下各章節內所使用的資料均為這類產品。本文分析所用資料是從 2011 年 4 月 14 日 2300UTC 開始，到 2012 年 12 月 31 日 2200UTC 為止，每小時一組(如圖 2-3 即為其中一例)，共計 15048 組。其中前 20 個月(即由 2011 年 4 月 14 日 2300UTC 至 2012 年 11 月 30 日 2300UTC 止)共 14305 組資料將用於 EOF 以及 KLE 模分析(詳見第三章)，而最後一個月(2012 年 12 月 1 日 0000UTC 至 2012 年 12 月 31 日 2200UTC 止)共 743 組資料則當作獨立樣本，作為填補缺值實驗之根據。



如 1.2 節所述，蘇澳、漢本二站有時會因各種因素影響會造成合成海流資料中出現某些點位無資料的情形。在 15048 組觀測資料中，各觀測點出現缺漏值之數量統計如圖 1-6，換算成百分比出現率則如圖 2-4(a)所示，可看出各點資料之最大缺漏率幾乎都在 5% 以下。此外，資料中完全沒有任何缺漏值的組數為 14014，約佔全部資料的 93.13%，有缺值情況的組數為 1034，共佔 6.87%；後者中，因儀器故障完全沒有任何觀測資料的組數則為 605，佔全部資料的 4.02%；而其餘的 429 組，資料缺漏量由 1 至 69 點，其百分比出現率(以總樣本數 429 計)分佈則如圖 2-4(b)所示，其中以每組缺漏一點的情形數量最多，共 126 次，佔有缺漏值資料組的 29.37%，缺二點佔 11.42%，缺三點則為 9.79%，顯示觀測值發生缺漏情形時，約有一半左右的情況都是在三點以內，至於其他缺漏數量情況詳見表 2-1。另一方面，將 15048 組資料中所有有缺漏值的組數(共 1034 組)按其每日出現之小時數另行統計，則每小時總共約發生 30~60 次缺漏，詳細數據見表 2-2。圖 2-4(c)為每日資料缺漏量百分比之逐時變化分佈，缺漏量之每小時出現率(以樣本數 1034 計)都在 6% 以下，但在每日 UTC 時間 10 時至 20 時(約本地時間晚上 6 點到凌晨 4 點)則比較容易發生資料缺漏情形，此時段內缺漏事件之出現率約為 4~6%，與 Fang et al. (2011) 所述 SCONET 海域 CODAR 系統受電離層干擾影響時間大致吻合。

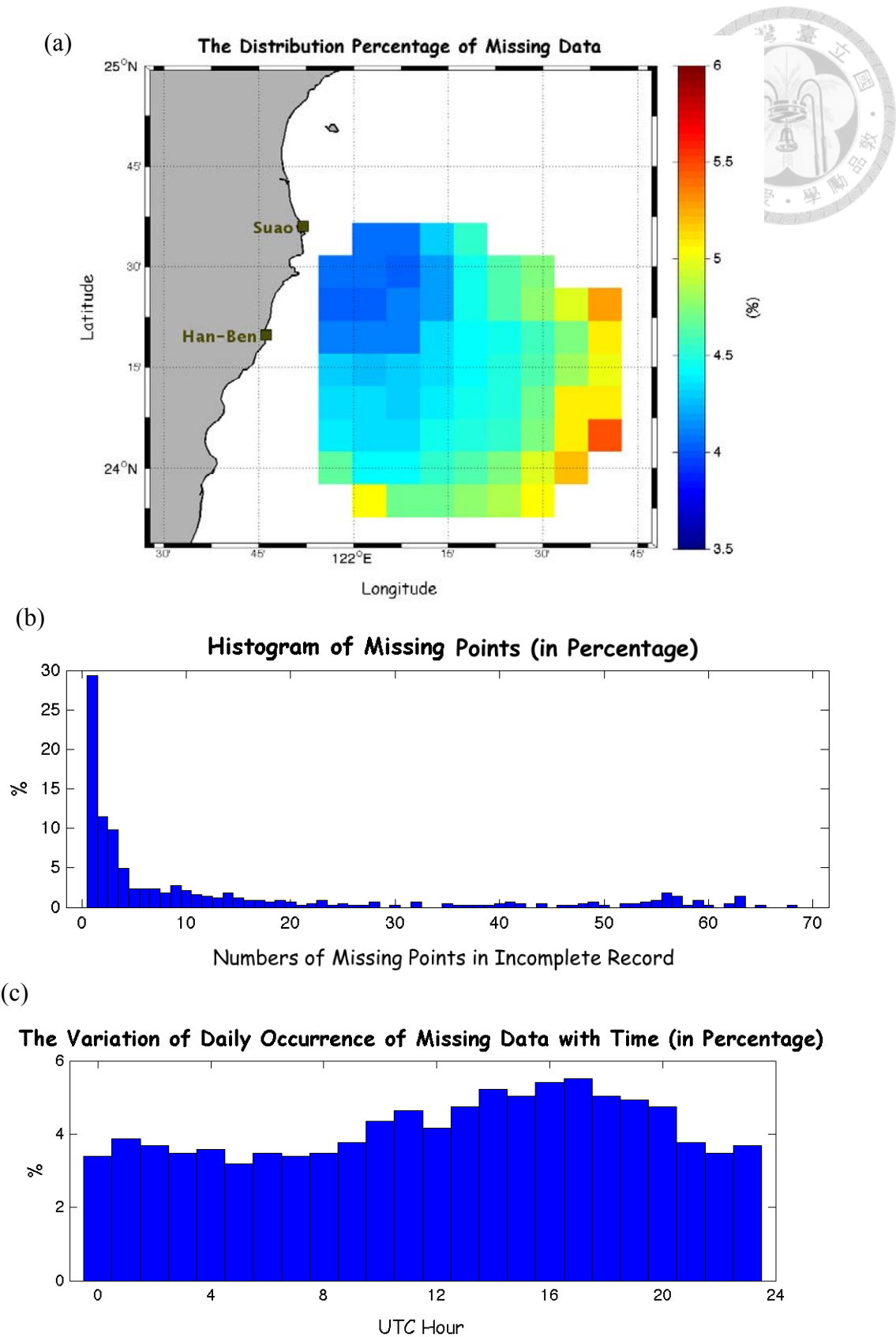


圖 2-4 合成海流資料點位出現資料缺漏情況之統計(2011 年 4 月 14 日至 2012 年 12 月 31 日)：(a)資料缺漏之百分比分佈；(b)不完整資料中缺漏點數之百分比統計直方圖；(c)長期統計資料缺漏百分比之平均日變化。

表 2-1 合成海流資料中發生缺漏情形(共 429 組)之統計。

缺漏值個數	組	%	缺漏值個數	組	%	缺漏值個數	組	%
1	126	29.37	24	1	0.23	47	1	0.23
2	49	11.42	25	2	0.47	48	2	0.47
3	42	9.79	26	1	0.23	49	3	0.70
4	21	4.90	27	1	0.23	50	1	0.23
5	10	2.33	28	3	0.70	51	0	0.00
6	10	2.33	29	0	0.00	52	2	0.47
7	10	2.33	30	1	0.23	53	2	0.47
8	8	1.86	31	0	0.00	54	3	0.70
9	12	2.80	32	3	0.70	55	4	0.93
10	9	2.10	33	0	0.00	56	8	1.86
11	7	1.63	34	0	0.00	57	6	1.40
12	6	1.40	35	2	0.47	58	1	0.23
13	5	1.17	36	1	0.23	59	4	0.93
14	8	1.86	37	1	0.23	60	1	0.23
15	5	1.17	38	1	0.23	61	0	0.00
16	4	0.93	39	1	0.23	62	2	0.47
17	4	0.93	40	2	0.47	63	6	1.40
18	3	0.70	41	3	0.70	64	0	0.00
19	4	0.93	42	2	0.47	65	1	0.23
20	3	0.70	43	0	0.00	66	0	0.00
21	1	0.23	44	2	0.47	67	0	0.00
22	2	0.47	45	0	0.00	68	1	0.23
23	4	0.93	46	1	0.23	69	0	0.00

表 2-2 長期統計資料缺漏百分比之平均日變化。

UTC Hour	No Missing	Missing	Missing (%)
0	592	35	3.38
1	587	40	3.87
2	589	38	3.68
3	591	36	3.48
4	590	37	3.58
5	594	33	3.19
6	591	36	3.48
7	592	35	3.38
8	591	36	3.48
9	588	39	3.77
10	582	45	4.35
11	579	48	4.64
12	584	43	4.16
13	578	49	4.74
14	573	54	5.22
15	575	52	5.03
16	571	56	5.42
17	570	57	5.51
18	575	52	5.03
19	576	51	4.93
20	578	49	4.74
21	588	39	3.77
22	591	36	3.48
23	589	38	3.68
Total	14014	1034	100.00





2.2 時間域基本統計


為了便於分辨與說明起見，首先我們將圖 1-5 所示之 SCONET 海域合成海流資料點位按照由下而上、由左而右之順序編碼命名(由 1 至 70)，其分佈如圖 2-5。接著將 2011 年 4 月 14 日至 2012 年 12 月 31 日期間此 70 個資料點之東西向流速(U)與南北向流速(V)分別求其平均值(mean)、標準差(standard deviation, SD)，以及中位數(median)，統計結果詳如表 2-3。

SCONET 海域內平均之觀測海流分佈情形如圖 2-6 所示，圖中深橘色箭頭即為平均海流流速向量，各點中最大流速為 84.87 cm/s，最小則為 28.79 cm/s，流向為北至東北，和我們所認知的黑潮流況相符。觀測區內平均海流流速較小處有二，其一是位於觀測區東南側，大約在 122.25°E~122.5°E 之間以及 24.25°N 以南的海域，另一處則是在宜蘭海脊上接近蘇澳的海域。圖 2-6 中以黑色箭頭表示的是海洋資料庫發佈的船載 ADCP 水深 20 公尺長期平均流場(1991 年~2008 年)，後者之海流流速範圍在 4.52~112.41 cm/s 之間，流向為東北，海流流速較小處亦位於 SCONET 海域的東南部與西北部。比較兩種不同來源的長期平均流場，可看出二者之間的差異在觀測區中部二者流速向量的偏差較小，且西部又較東部偏差小。二者之差異可能和船載 ADCP 與 CODAR 觀測資料之水深不同(前者水深約 20 公尺，後者約 1 公尺)有關。

計算出各資料點的平均海流流速後，再根據平均值計算各點位之海流分量標準差，可反映海流流速變動之幅度大小。東西向流速標準差(圖 2-7a)介於 17.53~38.62 cm/s 之間，此值在 SCONET 海域西北方較小、西南方較大，東部則大多介於 20~25 cm/s 之間；南北向流速的標準差分佈(圖 2-7b)則在 21.59~39.63 cm/s 之間，較小值出現在 SCONET 海域中南部，而北部的標準差大致上都較大。標準差與海流流速變動幅度成正比，故根據標準差的分佈可看出在 SCONET 海域西南部東西向流速變動較大，而在 SCONET 海域東北部則南北向流速變動較大。

表 2-3 各點位之東西向流速(U)與南北向流速(V)的基本統計。

D.P (cm/s)	U Mean	V Mean	U SD	V SD	U Median	V Median
1	12.00	79.81	33.75	31.53	13.30	81.67
2	13.94	75.82	27.32	31.39	14.75	78.24
3	19.57	68.34	21.70	31.41	20.03	71.68
4	17.56	50.32	19.45	30.57	17.26	53.41
5	20.50	36.27	18.22	31.36	19.99	37.48
6	29.03	29.53	17.57	30.98	29.74	28.84
7	33.86	24.26	19.16	29.53	36.26	23.70
8	16.62	78.57	38.62	29.76	18.05	80.30
9	12.10	75.96	30.20	29.65	12.78	77.00
10	14.78	72.80	25.95	29.19	14.82	73.65
11	18.61	65.24	24.23	28.47	18.91	66.65
12	23.19	53.89	20.31	30.09	23.32	55.75
13	26.51	36.21	17.53	30.84	26.35	36.22
14	35.21	31.23	17.84	30.53	35.82	30.08
15	37.21	27.80	18.99	29.93	40.00	27.23
16	34.50	24.48	20.44	28.59	38.06	26.17
17	2.50	64.09	34.73	26.26	3.77	64.99
18	3.47	63.15	26.50	25.78	3.44	63.86
19	9.45	61.66	23.65	25.36	8.76	61.96
20	16.82	58.02	22.70	24.72	15.68	57.99
21	25.11	51.55	21.76	27.58	25.12	50.48
22	34.53	42.75	18.85	28.79	34.09	41.30
23	41.67	34.90	17.96	30.18	42.41	32.60
24	41.25	32.84	18.74	29.63	43.77	32.44
25	40.49	27.51	20.80	31.67	43.54	28.79
26	-3.81	50.82	27.13	22.42	-3.62	50.93
27	0.22	50.74	21.57	22.61	-0.43	51.20
28	6.90	51.04	21.33	23.31	5.75	51.34
29	16.97	53.47	20.91	23.42	15.38	53.40
30	26.67	50.94	21.09	26.43	25.49	49.83
31	39.80	51.53	19.92	29.79	39.06	49.78
32	45.55	45.96	18.35	28.78	45.56	44.72
33	48.39	40.98	19.72	30.78	49.99	40.24
34	51.85	21.23	23.20	32.62	53.88	20.66



35	-13.10	32.50	23.02	21.59	-13.70	32.27
36	-3.52	37.79	20.19	22.75	-4.83	37.69
37	5.30	41.27	21.45	24.47	3.72	41.00
38	18.06	49.81	21.43	24.91	16.55	48.80
39	30.05	56.82	21.58	26.58	28.58	55.99
40	41.62	60.81	21.50	29.26	40.63	59.79
41	51.19	63.73	21.06	29.83	50.58	62.71
42	52.86	37.83	21.11	31.79	52.54	36.68
43	59.17	28.17	24.19	35.89	59.51	28.52
44	-11.01	28.63	21.14	22.69	-12.31	28.22
45	0.33	35.57	21.03	24.28	-1.46	35.02
46	9.09	42.28	22.39	26.82	6.95	40.50
47	21.49	54.90	22.38	27.46	19.68	53.25
48	32.33	63.51	22.48	28.07	30.83	62.43
49	41.98	66.65	22.83	30.69	40.73	65.79
50	52.02	67.06	23.07	30.89	50.87	65.34
51	57.22	43.65	23.31	33.48	55.87	42.95
52	-5.40	28.28	21.94	24.22	-6.86	26.42
53	6.97	40.82	23.09	27.23	5.14	38.58
54	15.16	50.25	23.40	28.38	13.24	48.20
55	24.71	60.51	23.05	28.94	23.19	59.22
56	32.24	63.92	22.83	30.36	30.67	62.44
57	39.48	58.70	23.17	31.38	37.89	56.84
58	48.39	55.69	23.80	33.10	46.65	53.82
59	56.75	53.37	24.63	34.57	54.72	52.51
60	13.29	49.49	25.03	30.38	11.33	47.09
61	19.09	56.12	24.57	30.93	17.15	53.77
62	25.53	61.20	23.62	31.47	24.13	59.64
63	30.49	56.46	22.78	30.66	29.15	55.45
64	36.45	47.42	23.25	32.78	34.77	45.84
65	44.64	43.01	23.65	36.52	42.56	41.26
66	20.42	56.33	25.20	33.73	18.50	53.68
67	25.36	55.48	23.55	31.61	24.07	53.80
68	28.31	44.89	22.45	31.12	27.26	44.49
69	33.55	35.81	22.99	35.47	32.24	35.30
70	41.01	34.54	23.10	39.63	38.93	32.29

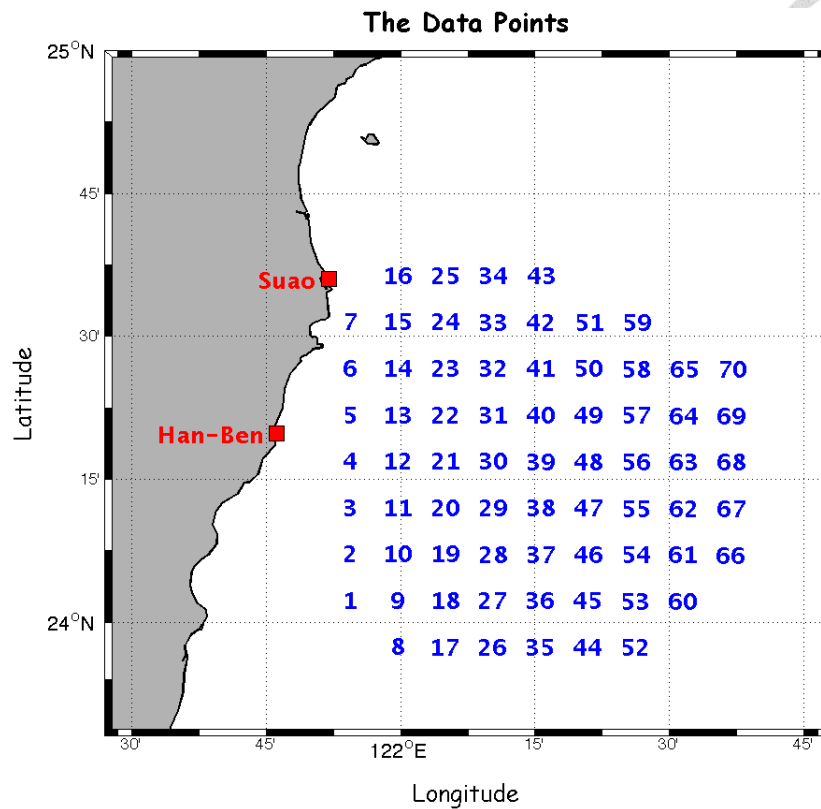


圖 2-5 資料點位編號之相對位置圖。

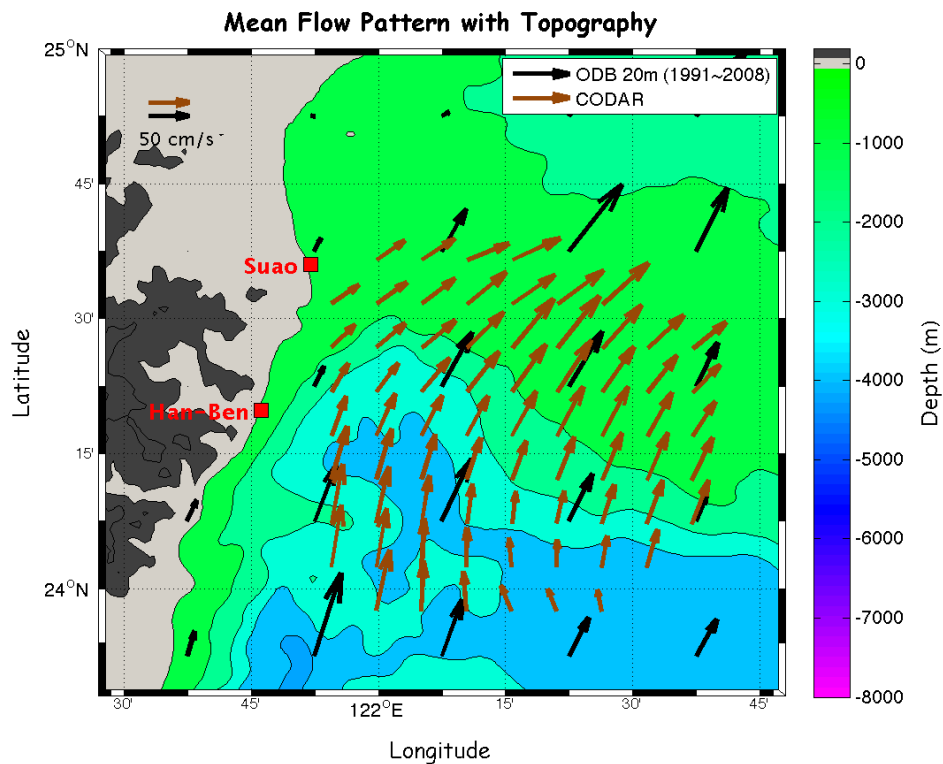


圖 2-6 觀測區內之平均流場及水深 20 公尺之長期平均流場(海底地形與長期平均流場資料來源：國科會海洋學門資料庫)。

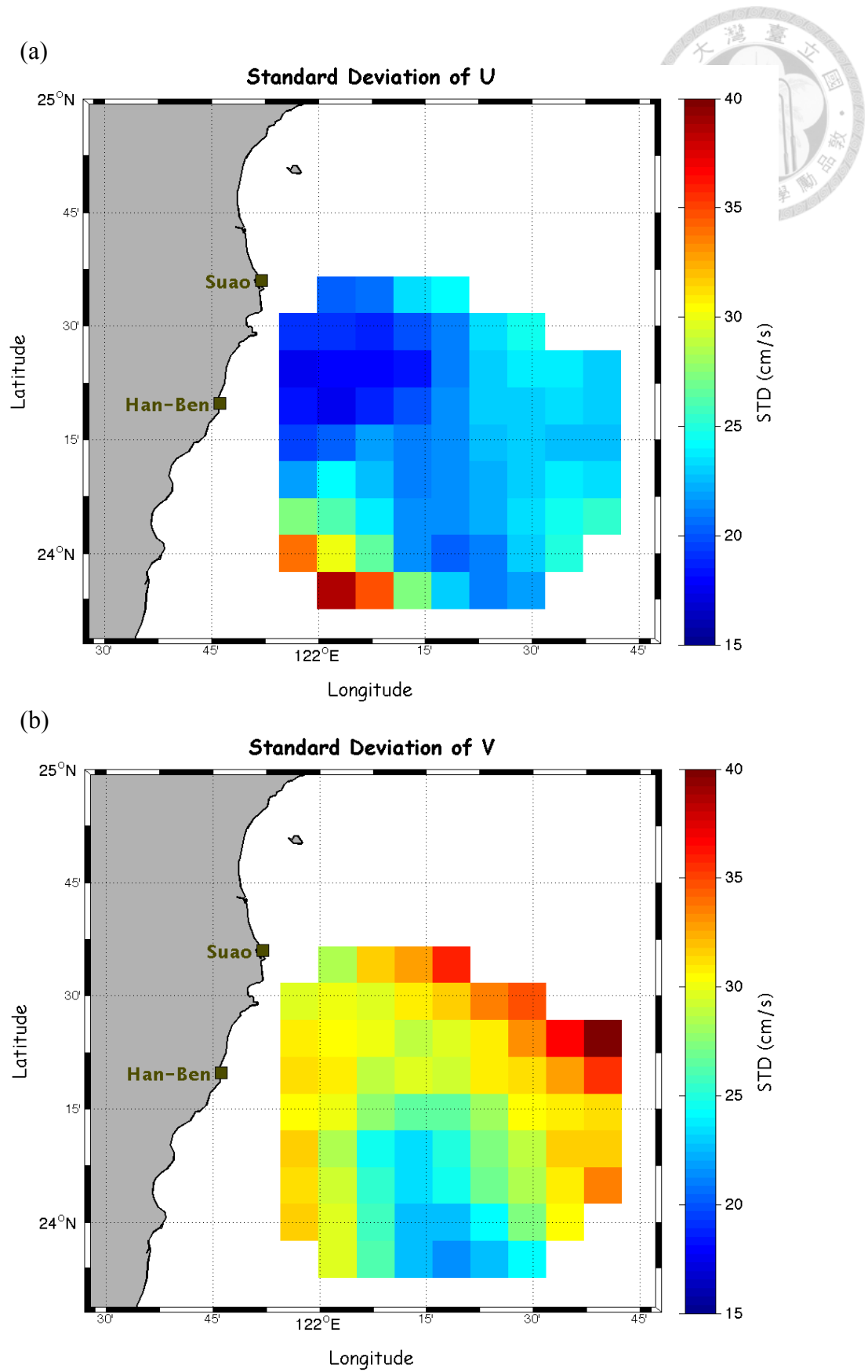


圖 2-7 CODAR 觀測海流資料之標準差分佈：(a)東西向流速；(b)南北向流速。

為了觀察 CODAR 流速觀測資料之機率分配情形，我們將 2011 年 4 月 14 日至 2012 年 12 月 31 日所有東西向流速與南北向流速資料統計分析，分別計算出各自的平均值、標準誤(standard error, SE)及中位數(如表 2-4)。另外，再根據表中之平均值以及標準誤，分別將東西向流速與南北向流速標準化(normalization)，以 Z 表示($Z = \frac{X - \mu}{s}$ ， X 為流速資料， μ 、 s 則分別是資料的平均值與標準誤)，並繪出機率分佈直方圖(圖 2-8a 與 b)，圖中直方圖是將標準化後的流速資料 Z 等分成 50 個區間後再繪出，每個區間大小(ΔZ)分別為 0.26(圖 2-8a)與 0.24(圖 2-8b)；紅色曲線為流速資料的平均值與標準誤所構成的常態分配曲線，也就是說，圖 2-8(a)與圖 2-8(b)的紅色曲線實際上分別是平均值為 25.65 cm/s、49.23 cm/s 以及標準誤為 28.86 cm/s、32.80 cm/s 的常態分配曲線。將此曲線與直方圖互相對照，可以從圖中看出東西向與南北向流速的分佈都相當近似於常態分配。此外，圖 2-8(c)與(d)則分別是東西向和南北向流速的 P-P 圖(probability- probability plot)，藍圈是標準化後流速資料 Z 與流速資料之累積分佈函數(cumulative distribution function, CDF)的分佈圖，若資料為常態分配，則藍圈的分佈情形應與紅色對角線吻合，由 P-P 圖中二者幾乎完全重疊可證明資料屬於常態分配。此外，將各點位之東西向、南北向流速同樣分別經標準化後以 Z 表示並觀察其分佈情形，70 個點位的資料皆為常態分配(因篇幅關係，此部份未列入論文)。

表 2-4 全部資料東西向流速(U)與南北向流速(V)的基本統計資料。

	Mean (cm/s)	SE (cm/s)	Median (cm/s)
U	25.65	28.86	26.39
V	49.23	32.80	48.74

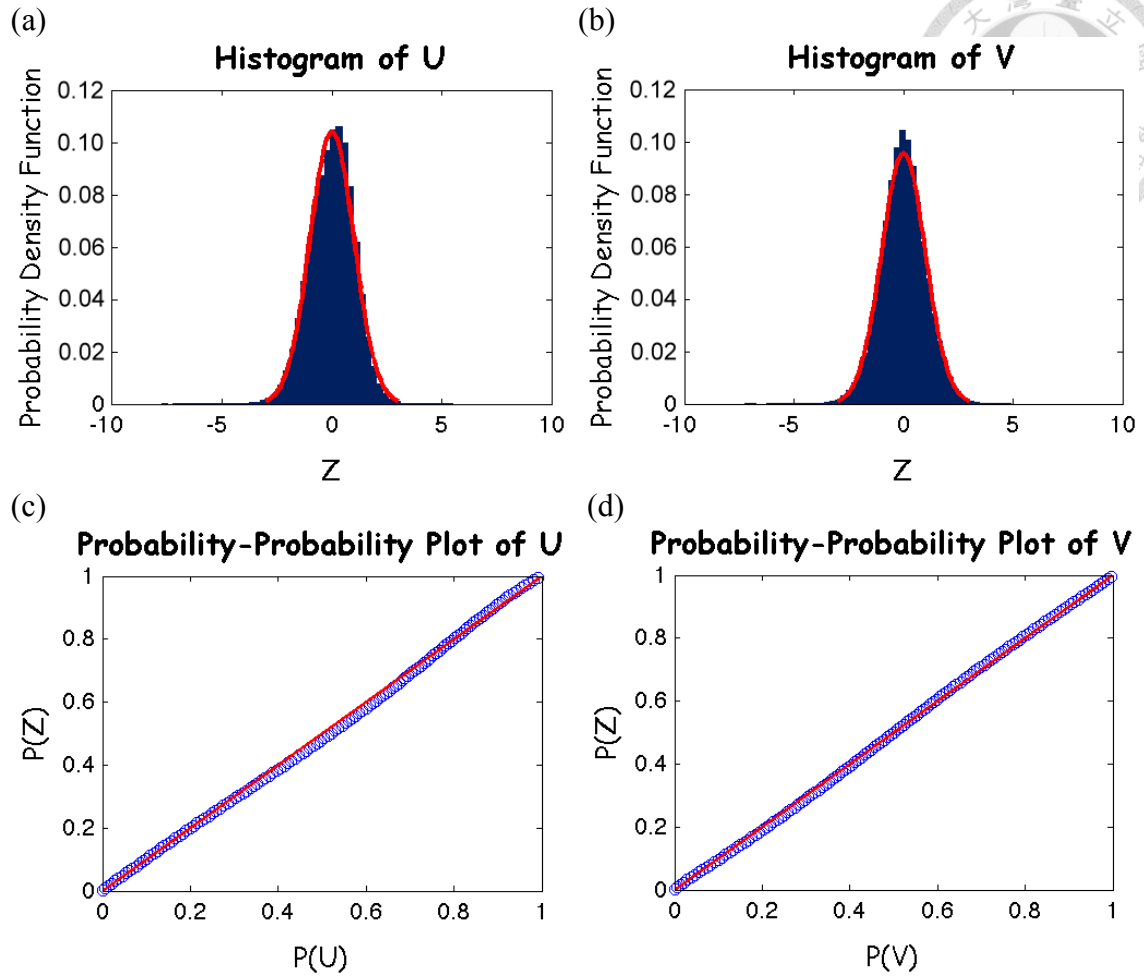


圖 2-8 (a)東西向流速分佈圖；(b)南北向流速分佈圖；(c)東西向流速 P-P 圖；(d)南北向流速 P-P 圖。紅線表示常態分佈曲線。



2.3 頻率域分析以及濾潮處理

潮汐與潮流是近岸海域裡重要的物理海洋現象；在淺水域裡潮流強度往往比背景海流還要強勁。由劉 (1999)知，位於臺灣東北方的 SCONET 海域潮汐主要是受混合潮型所主控，圖 2-9(a)為 2011 年 8 月 16 日至 11 月 15 日蘇澳港的逐時潮位曲線，可明顯看出本地潮汐 14 天週期的大小潮變化以及小潮時潮型以全日潮為主的特性。將這段潮位資料以 T-tide (Pawlowicz et al., 2002)進行調和分析，圖 2-9(b)即為主要一些分潮振幅分佈情形，根據 Defant (1961)的潮型指標(Form Number)，
$$F = \frac{K_1 + O_1}{M_2 + S_2}$$
，可算出蘇澳港的潮型指標 $F = 0.52$ ，因此可知鄰近蘇澳的 SCONET 海域潮汐應當係屬以半日潮為主的混合潮型。

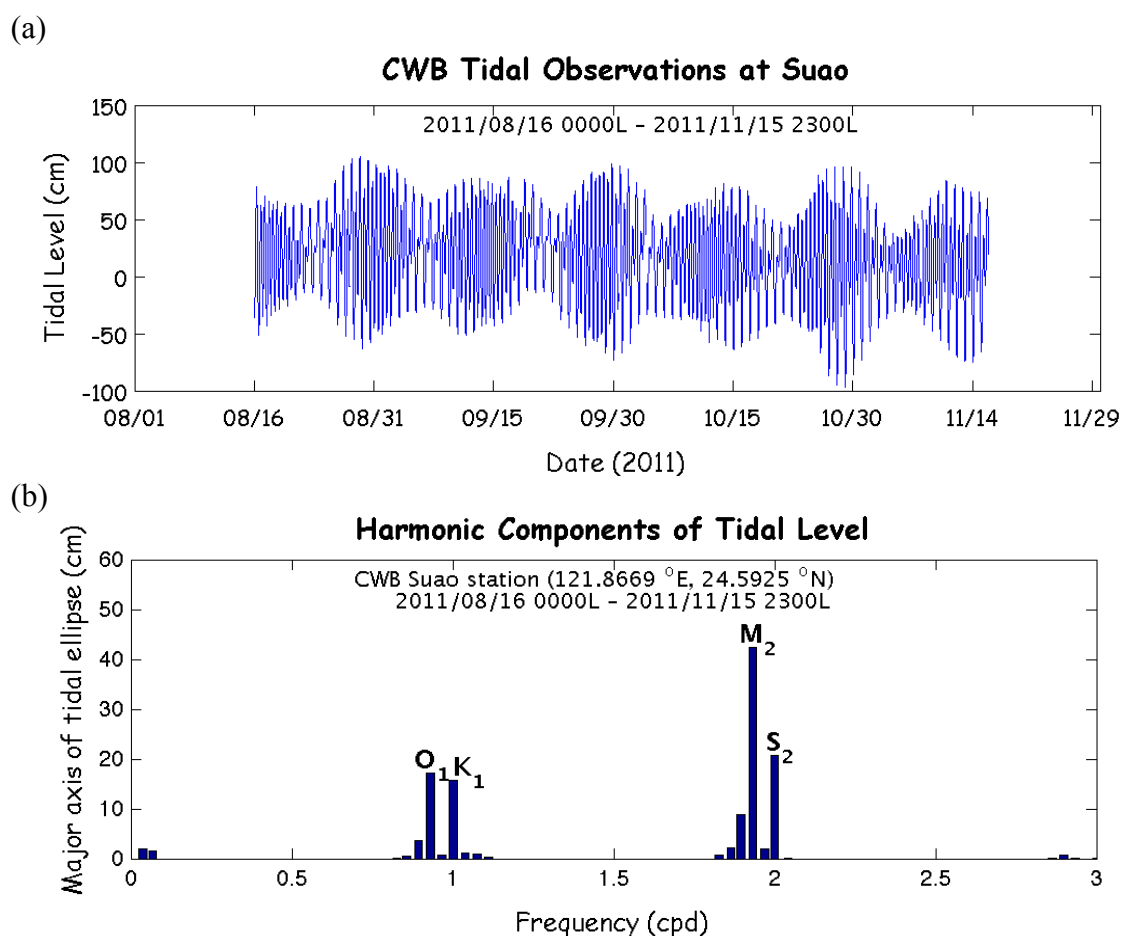



圖 2-9 中央氣象局蘇澳站(121.87°E、24.59°N) 2011/8/16~2011/11/15 的(a)潮位觀測資料及(b)各分潮振幅。

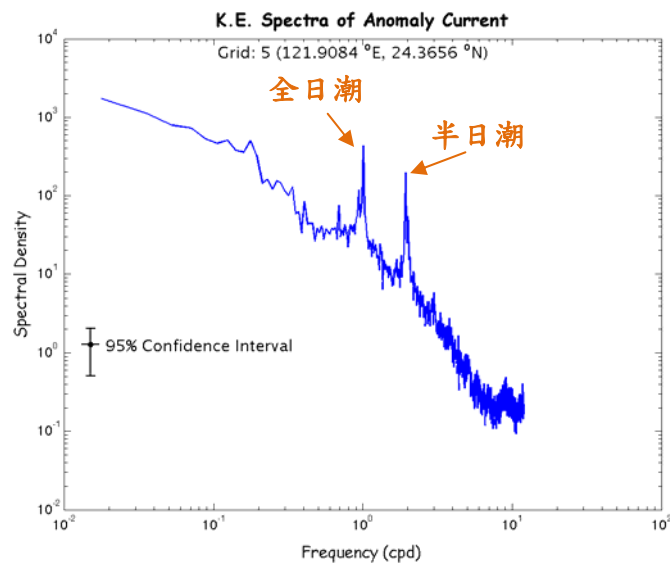


接下來進一步以能譜分析觀察各點位流速資料在頻率域之特性。首先將每個點位的 15048 筆東西向流速以及南北向流速資料皆分成 4 段(每段 4096 點)，分別扣除各段平均流速分量後，經快速傅立葉轉換(FFT)得出四組東西向和南北向流速之能譜密度(power spectral density)，然後將四組平均並在頻率域中每三個頻率相加平均及重新取樣，如此即得到自由度(degree of freedom)為 24 的動能譜(kinetic energy spectrum)，圖 2-10(a)為第 5 點位(相關位置參照圖 2-5)之動能譜，可以看出頻率 f 越低，能量越大，此外在週期約為 12 小時及 24 小時(即 2 cpd 和 1 cpd)處有明顯的峰值，分別代表全日潮與半日潮的成份，另外在 0.05~1 cpd 之間，動能譜有正比於 $f^{-1.1}$ 的趨勢，不過在所有 70 個點位中，這種特性之能譜僅出現在觀測區西部的資料點上，在觀測區東部(以圖 2-10b 為例)及南部，海流能譜中全日潮與半日潮分量之能量尖峰較不明顯，而在 0.05~1 cpd 之間其能譜能階則約正比於 $f^{-0.9}$ 。此外，將各資料點動能譜中與全日潮、半日潮頻率相對應的能量挑出並繪成水平分佈圖(圖 2-11)，可見全日潮潮流在觀測區東北部、西北部的能量較強(圖 2-11a)，半日潮潮流則是在北部的能量較大(圖 2-11b)，另方面半日潮潮流的能量在有些資料點上會比全日潮潮流還大，最大的大約可達全日潮潮流能量的兩倍以上，但在另外一些資料點上則又呈現相反的情形，這種能量分佈不均勻的特性可能和當地的內潮(internal tide)有關(Jan et al., 2012)。由動能譜可知，CODAR 觀測資料中低頻波動的能量較潮流為大。

若將各個資料點所觀測到的海流資料進行調和分析，可得出各資料點的分潮流橢圓長軸係數，圖 2-12 即為一例，該圖顯示第 1 個資料點位由 2011 年 8 月 16 日至 11 月 15 日資料所算出之潮流橢圓長軸係數，在全日潮及半日潮(O_1 、 K_1 與 M_2 、 S_2)以及低頻成份處潮流振幅均較大，全日潮及半日潮潮流橢圓長軸係數約在 10 cm/s 左右，至於其他點位之潮流橢圓長軸係數則多小於 10 cm/s。與表 2-3 相比可知本地潮流流速頗小於平均流速，當然其中也有部分原因可能係緣於 2.1 節

所述，為了提升訊雜比(Ramp et al., 2008)，我們所用之徑向海流資料檔是經由 4 小時平均處理後的 CSS File 計算得出，因此合成海流資料中週期低於 4 小時的訊號會被相當程度的抑制，而潮流成份也會比實際情形為弱(Fang et al., 2011)。為了去除潮流的影響，我們先將前 20 個月的原始觀測流場利用調和分析法求出各資料點位之潮流調和常數，再追算合成出該時間內各點位之潮流，然後將原始觀測資料扣除潮流，如此處理後即得出經過濾潮的海流資料，後者將使用於下一節之後的各項分析。

(a)



(b)

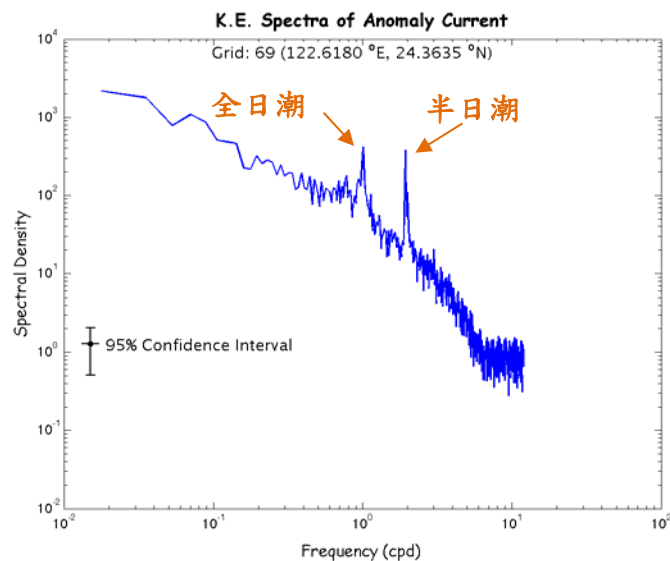
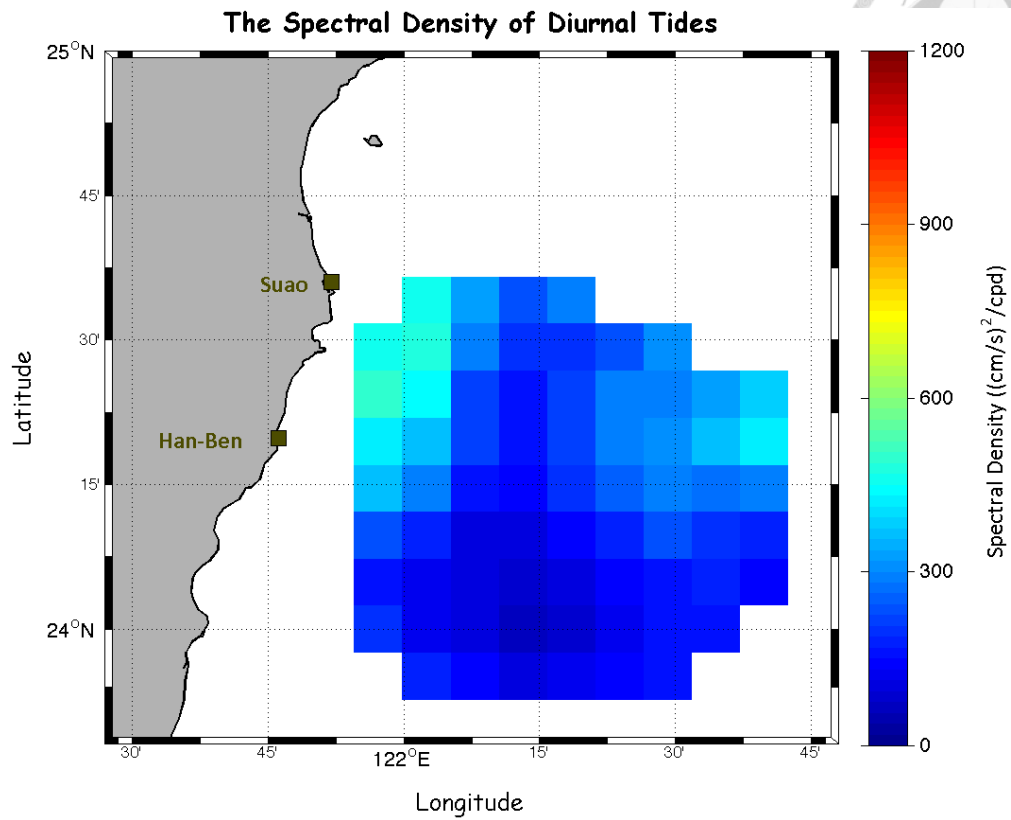


圖 2-10 (a)第 5 點位、(b)第 69 點位之流速動能譜圖。

(a)



(b)

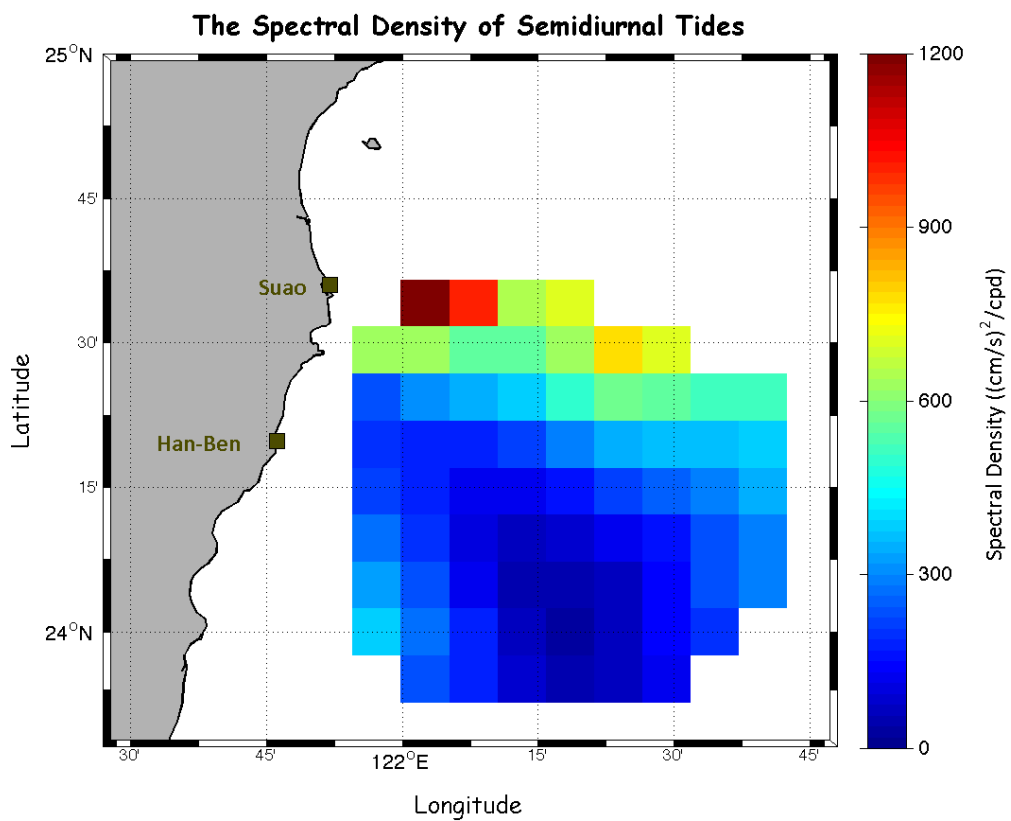


圖 2-11 潮流能量空間分佈圖：(a)全日潮分量；(b)半日潮分量。

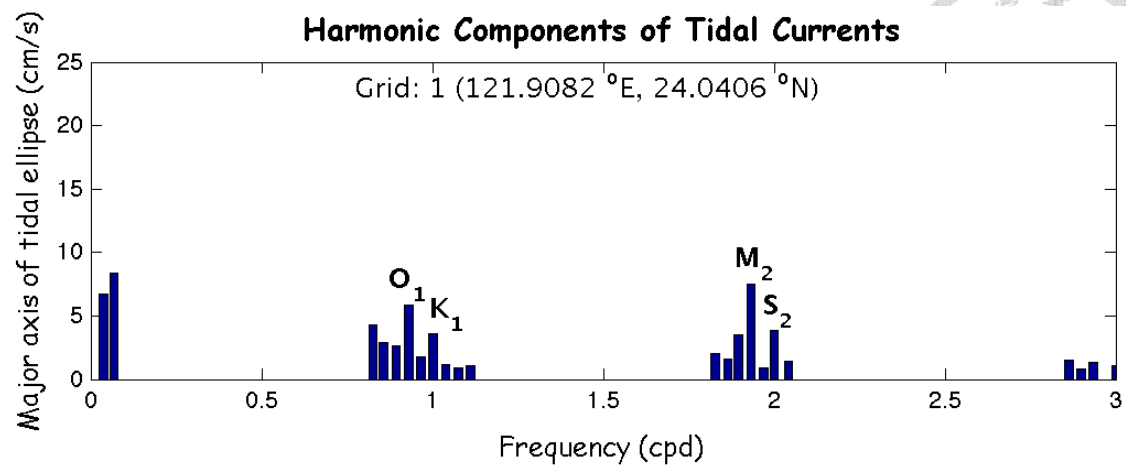


圖 2-12 觀測區內第 1 個資料點海流資料(2011 年 8 月 16 日至 11 月 15 日)經調和分析後得到的各分潮潮流橢圓長軸係數。



2.4 平穩性檢定

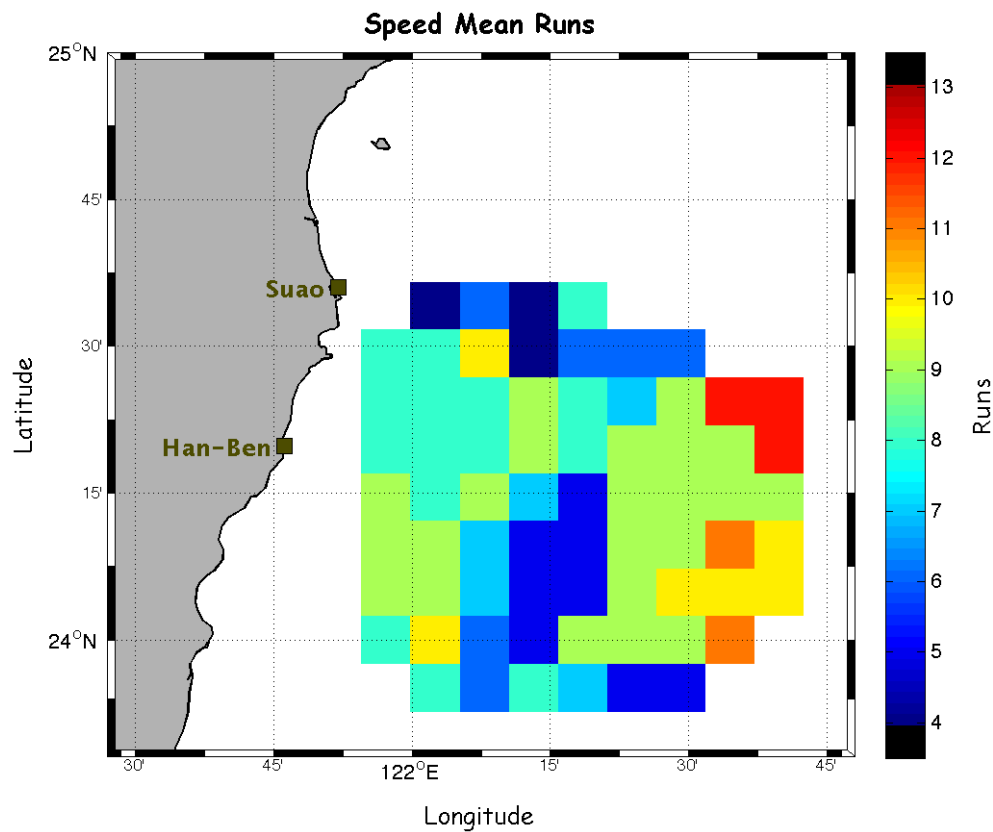
對我們在第四章將要探討的填補資料缺漏實驗而言，由於填補方式是藉助於實向量EOF法或KLE法(詳見第三章)得到的基底模組，但這些基底模組是根據「過去」歷史資料之統計求出，因此如果統計結果是非穩定的(nonstationary)，那麼後續的補缺過程在邏輯上便會喪失立論基礎，是以確定資料是否具備統計穩定性便是一項首要的工作。根據Bendat and Piersol (1971)，如果一個時序資料其資料總體平均(ensemble mean)以及共變異數(covariance)不隨起始時間的改變而變，而僅與各資料點之時間差有關，那麼這組時序資料即滿足弱平穩(weakly stationary)的假定。

Bendat and Piersol (1971)指出連檢定(run test)是檢定資料平穩性(stationarity)的一種作法。連檢定是一種無母數統計分析法(non-parametric statistics)，或稱不受分配限制的統計法(distribution-free)，按其作法我們的檢定步驟如下：(1)首先定義虛無假設(null hypothesis) H_0 為：資料是平穩的，而其對立假設(alternative hypothesis) H_1 則為：資料不是平穩的；(2)將前述經過濾除潮流後的14305組(約20個月)的CODAR海流觀測資料等分為16段；(3)求出每一段之平均值與標準差，如此就得出16個平均值與標準差的時序；(4)目測觀察此16個平均值與標準差之時序，如果沒有明顯的趨勢(trend)變化，那麼 H_0 就很有可能成立；(5)求出二序列數據的中位數，以及二中位數與二序列資料之交點，根據交點可將序列分成 R 個區間，也就是 R 個連，再由連分配(run distribution)表(Bendat and Piersol, 1971)檢定當設定之顯著水準(significance level)為 α 時，如果 H_0 成立，那麼連的個數 R 應介於 $r_{n;1-\alpha/2}$ 與 $r_{n;\alpha/2}$ 之間，反之若 R 落在 $r_{n;1-\alpha/2}$ 與 $r_{n;\alpha/2}$ 的範圍外，則 H_0 不成立；上述中 $r_{n;1-\alpha/2}$ 以及 $r_{n;\alpha/2}$ 為連分配表給定之上、下限數值。

如果 H_0 成立，根據 $\alpha = 5\%$ 之顯著水準以及 $n = N/2 = 16/2 = 8$ 查表後可得出連個數 R 應當介於 $r_{8;1-\alpha/2} = r_{8;1-0.05/2} = r_{8;0.975} = 4$ 和 $r_{8;\alpha/2} = r_{8;0.05/2} = r_{8;0.025} = 13$ 之間。表2-5即為針對SCONET海域內70個資料點之連檢定分析結果，表中數字分別為各資料點

東西向流速(U)、南北向流速(V)以及流速速率(speed)之平均值與標準差的連個數，表中第一行(D.P.)係指資料點位之編號(對應位置參見圖2-5)，第二至第四行是以平均值所計算出的連個數，最後三行則是用標準差所求出的連個數。圖2-13(a)與(b)則分別是以各資料點之流速速率平均值和標準差進行連檢定的結果，圖中彩色區域表示通過檢定(即各資料點的連個數介於4~13之間)，而黑色區域則表示不通過檢定，從二圖可知全部資料點的資料都通過檢定，即所有資料皆滿足弱平穩之假定，應當可以向外延伸用於觀測時段之外。

(a)



(b)

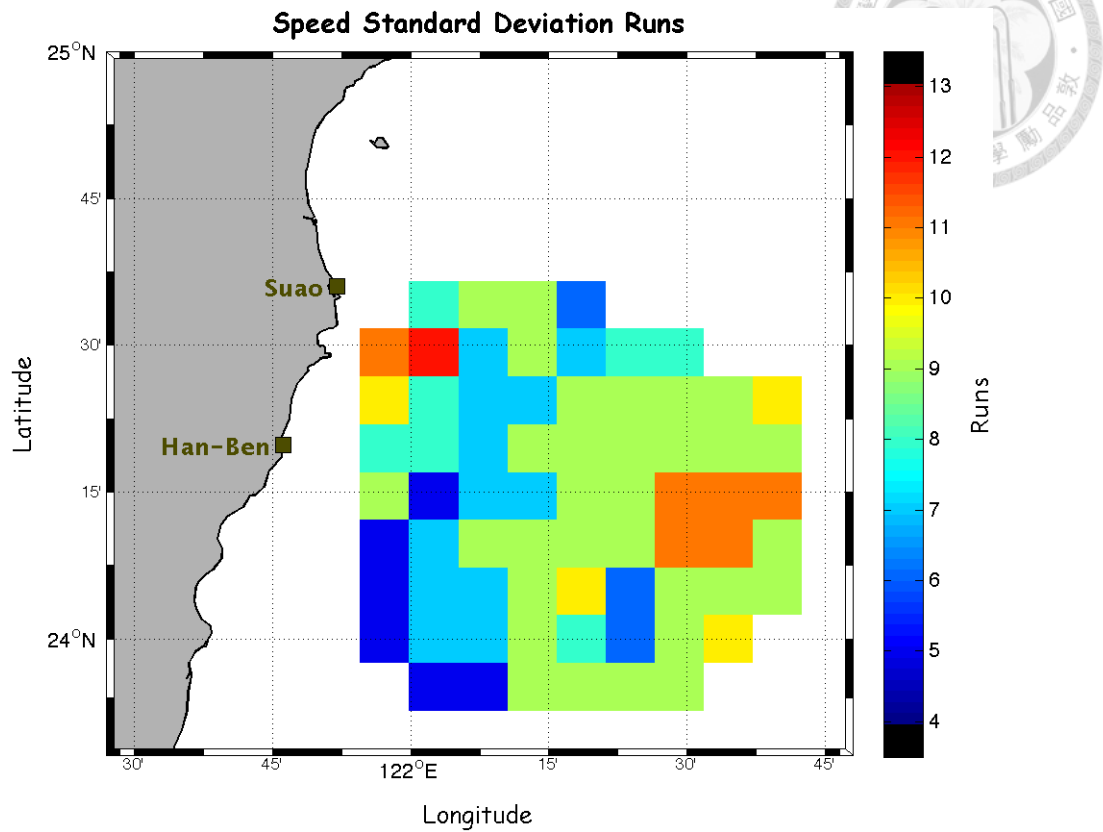
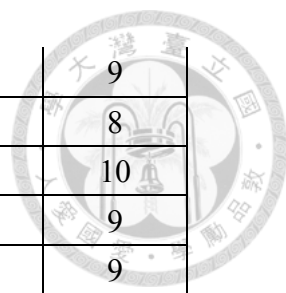


圖 2-13 流速速率資料之平穩性檢定結果，彩色區塊表示通過檢定，表示資料具有穩定性，而黑色區塊表示未通過檢定：(a)平均值之檢定結果；(b)標準差之檢定結果。

表 2-5 各點位之東西向流速(U)、南北向流速(V)及速率(speed)的連個數。

D.P.	U _{mean}	V _{mean}	Speed _{mean}	U _{SD}	V _{SD}	Speed _{SD}
1	9	8	8	8	8	5
2	9	10	9	8	8	5
3	9	10	9	5	8	5
4	7	7	9	6	9	9
5	9	7	8	4	7	8
6	11	11	8	8	7	10
7	9	9	8	8	10	11
8	5	10	8	8	8	5
9	9	10	10	4	8	7
10	9	10	9	7	9	7
11	7	9	9	7	9	7
12	9	9	8	6	9	5
13	11	7	8	4	9	8
14	11	7	8	6	11	8
15	13	11	8	9	10	12
16	13	11	4	7	10	8
17	4	10	6	6	8	5
18	11	10	6	5	5	7
19	9	10	7	7	9	7
20	7	9	7	4	7	9
21	7	7	9	6	9	7
22	7	7	8	6	9	7
23	13	7	8	7	9	7
24	13	11	10	9	11	7
25	13	9	6	7	9	9
26	6	12	8	8	11	9
27	9	10	5	5	11	9
28	9	11	5	5	11	9
29	9	11	5	4	9	9
30	7	7	7	4	9	7
31	9	7	9	6	9	9
32	10	5	9	6	9	7
33	10	9	4	7	9	9
34	11	7	4	7	9	9



35	7	9	7	5	9	9
36	9	11	9	5	11	8
37	9	11	5	5	9	10
38	7	11	5	5	9	9
39	7	9	5	4	9	9
40	9	9	8	6	9	9
41	10	9	8	6	9	9
42	10	7	6	6	4	7
43	10	7	8	7	6	6
44	9	8	5	5	9	9
45	10	8	9	5	7	6
46	12	8	9	5	9	6
47	10	10	9	5	9	9
48	10	10	9	5	9	9
49	10	11	9	4	9	9
50	10	9	7	6	9	9
51	10	7	6	6	4	8
52	10	10	5	5	7	9
53	10	6	9	9	9	9
54	10	10	10	7	9	9
55	10	10	9	7	9	11
56	10	12	9	7	9	11
57	10	11	9	4	9	9
58	10	11	9	4	7	9
59	10	9	6	6	8	8
60	10	6	11	9	10	10
61	10	10	10	7	11	9
62	10	10	11	7	11	11
63	10	10	9	7	11	11
64	10	10	9	4	11	9
65	10	11	12	6	12	9
66	10	10	10	7	11	9
67	10	10	10	7	11	9
68	10	10	9	7	9	11
69	10	10	12	7	12	9
70	10	9	12	6	12	10

第三章 模分析方法與分析結果



本章討論經驗正交函數與 Karhunen-Loève 展開法二種模分析法，簡要說明二種方法所分解出之模組成份、構成以及主要模組之物理意義。另外，亦分析以主要模組疊加重組之資料與原始資料之誤差及重組還原度。最後討論以模組重組法填補觀測缺漏值時之實際作法。

3.1 經驗正交函數

經驗正交函數(EOF)是物理海洋學常用的一種統計方法，係將資料分解成空間和時間上的許多成份，藉以探討影響資料行為的主要因素。這種方法在不同的科學領域裡有不同的名稱，例如海洋學和氣象學稱為經驗正交函數，統計學稱為主成份分析(Principal Component Analysis, PCA)，心理學或經濟學領域則使用因素分析(factor analysis)的名稱(Beckers and Rixen, 2003; Kerschen et al., 2005)。Kundu et al. (1975)提到 EOF 法是個不需要任何動力假設，只需要做資料統計的分析方法。近幾十年來，氣象學和海洋學方面更是廣泛地運用 EOF 法來分析資料(Beckers and Rixen, 2003)。由於 EOF 法可以從資料裡提取出主要的結構，能有效地將資料用簡潔的形式，也就是少量的主分量表現出來(Kaihatu et al., 1998)，因此成為簡化或過濾大量資料時常用的有用工具。EOF 法可以用於純量，例如海表溫(Kutzbach, 1967)，或向量資料，如風(Hardy, 1977)或海流(Kundu and Allen, 1976)等。

Kaihatu et al. (1998)指出，對於向量型態的資料，EOF 法有兩種不同的處理方式，其一如 Kundu and Allen (1976)係使用複數型態的 EOF 法，也稱為 CEOF (Complex Empirical Orthogonal Functions)法，是先將二維的水平流速向量以複數型態表示，接著再根據 EOF 法的步驟求解複數矩陣之特徵向量(各成份可能為複數)；另一種則是稱為實向量 EOF (real-vector EOF)法，此方法是使用向量原型直接進行分析，所求出之特徵向量各成份均為實數。由於 CEOF 法算出的各組特徵向量其方向並非固定值，但實向量 EOF 法卻不會有此種情況，且後者並不侷限於二維，

可以延展至 N 維向量(Kaihatu et al., 1998)，因此本文採用實向量 EOF 法來進行分析。

在 SCONET 海域內共有 70 個資料點(圖 1-5)，每個資料點所測到的都是一組二維的水平流速向量，在此我們採用的是濾潮後之流速 (U_i, V_i) , $i=1,2,\dots,70$ 。使用實向量 EOF 法計算時，需先將 (U, V) 減去各資料點的海流平均值 (\bar{U}, \bar{V}) ，以二者之差值 (u, v) 進行後續分析，為簡化計下文中 u 和 v 將簡稱為變動流速 (fluctuation)。

$$\begin{aligned} u(x_i, t_k) &= U(x_i, t_k) - \bar{U}(x_i) \\ v(x_i, t_k) &= V(x_i, t_k) - \bar{V}(x_i) \end{aligned} \quad (1)$$

在(1)式中， x_i 表示第 i 個資料點 ($i=1,2,\dots,70$) 的水平位置， t_k 表示第 k 小時 ($k=1,2,\dots,K, K=13280$)。由(1)式，共70個觀測點而每點又有兩個流速分量，因此可以組成一個 140×140 的共變異數矩陣(covariance matrix)，以 R 表示：

$$R = [R_{pq}] = \begin{bmatrix} \overline{u_1 u_1} & \overline{u_1 v_1} & \dots & \overline{u_1 u_M} & \overline{u_1 v_M} \\ \overline{v_1 u_1} & \overline{v_1 v_1} & & \overline{v_1 u_M} & \overline{v_1 v_M} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \overline{u_M u_1} & \overline{u_M v_1} & \dots & \overline{u_M u_M} & \overline{u_M v_M} \\ \overline{v_M u_1} & \overline{v_M v_1} & & \overline{v_M u_M} & \overline{v_M v_M} \end{bmatrix}, \quad p, q = 1, 2, \dots, N, \quad (2)$$

$N = 140, M = \frac{N}{2} = 70.$

式中，

$$\overline{u_i u_j} = \frac{1}{K} \sum_{k=1}^K u(x_i, t_k) u(x_j, t_k)$$

$$\overline{u_i v_j} = \frac{1}{K} \sum_{k=1}^K u(x_i, t_k) v(x_j, t_k)$$

$$\overline{v_i u_j} = \frac{1}{K} \sum_{k=1}^K v(x_i, t_k) u(x_j, t_k)$$

$$\overline{v_i v_j} = \frac{1}{K} \sum_{k=1}^K v(x_i, t_k) v(x_j, t_k), \quad i, j = 1, 2, \dots, M$$

K 為資料筆數。得出 R 後即可計算出 N 個特徵值(eigenvalue) $\lambda_n, n=1, 2, \dots, N$ ，以

及相對應的 N 組互相正交(orthogonal)的特徵向量(eigenvector) $\phi_n, n=1, 2, \dots, N$ ；即

$$\sum_{p=1}^N R_{pq} \phi_n^p = \lambda_n \phi_n^q, \quad n=1, 2, \dots, N; q=1, 2, \dots, N \quad (3)$$

$$\begin{aligned} \text{式中, } \phi_n &= [\phi_n^p]_{p=1, 2, \dots, N}^T = [\phi_n^1, \phi_n^2, \dots, \phi_n^N]^T \\ &= [\phi_n^u(x_1), \phi_n^v(x_1), \phi_n^u(x_2), \phi_n^v(x_2), \dots, \phi_n^u(x_M), \phi_n^v(x_M)]^T \end{aligned}$$

ϕ_n 又可稱為第 n 模之模組(mode)或基底向量(base vector)， T 表示轉置矩陣。得出

模組向量後，再計算變動流速對各模組向量之投影，令

$$w = [w_p] = [u(x_1), v(x_1), u(x_2), v(x_2), \dots, u(x_M), v(x_M)] (p=1, 2, \dots, N), \text{ 即:}$$

$$\begin{aligned} E_n(t_k) &= \sum_{p=1}^N w_p(t_k) \phi_n^p \\ &= [u(x_1), v(x_1), \dots, u(x_M), v(x_M)] \cdot \begin{bmatrix} \phi_n^u(x_1) \\ \phi_n^v(x_1) \\ \vdots \\ \phi_n^u(x_M) \\ \phi_n^v(x_M) \end{bmatrix} \\ &= \sum_{p=1, 3, 5, \dots}^N u\left(\frac{x_{p+1}}{2}, t_k\right) \phi_n^u\left(\frac{x_{p+1}}{2}\right) + \sum_{p=2, 4, 6, \dots}^N v\left(\frac{x_p}{2}, t_k\right) \phi_n^v\left(\frac{x_p}{2}\right) \\ &= \sum_{p=1, 3, 5, \dots}^N u\left(\frac{x_{p+1}}{2}, t_k\right) \phi_n^p + \sum_{p=2, 4, 6, \dots}^N v\left(\frac{x_p}{2}, t_k\right) \phi_n^p \end{aligned} \quad (4)$$

式中 E_n 即為 t_k 時間第 n 模組之振幅。由所有模組振幅之時序資料亦可再行合成變

動流速，即：



$$w_p(t_k) = \sum_{n=1}^N E_n(t_k) \phi_n^p$$

$$\begin{bmatrix} u(x_1) \\ v(x_1) \\ \vdots \\ u(x_M) \\ v(x_M) \end{bmatrix} = \sum_{n=1}^N E_n(t_k) \begin{bmatrix} \phi_n^u(x_1) \\ \phi_n^v(x_1) \\ \vdots \\ \phi_n^u(x_M) \\ \phi_n^v(x_M) \end{bmatrix}$$

$$u(x_i, t_k) = \sum_{n=1}^N E_n(t_k) \phi_n^{(2i-1)}$$

$$v(x_i, t_k) = \sum_{n=1}^N E_n(t_k) \phi_n^{2i}, \quad i = 1, 2, \dots, M \quad (5)$$

此外，數學上可證明 R 矩陣的跡數(trace)和等於 $\sum_{n=1}^N \lambda_n$ ，而由(3)式則可得出

$\sum_{n=1}^N \lambda_n = \sum_{p=1}^N R_{pp}$ ，因此特徵值與變動流速之總變異數(variance)則有如下(6)式之關係：

$$\begin{aligned} \sum_{p=1}^N R_{pp} &= \frac{1}{K} \sum_{k=1}^K \left[\sum_{p=1,3,5,\dots}^N u^2\left(\frac{x_{p+1}}{2}, t_k\right) + \sum_{p=2,4,6,\dots}^N v^2\left(\frac{x_p}{2}, t_k\right) \right] \\ &= \frac{1}{K} \sum_{k=1}^K \left[\sum_{p=1,3,5,\dots}^N u^2\left(\frac{x_{p+1}}{2}, t_k\right) \delta_{nn} + \sum_{p=2,4,6,\dots}^N v^2\left(\frac{x_p}{2}, t_k\right) \delta_{nn} \right] \\ &= \frac{1}{K} \sum_{k=1}^K \left[\sum_{p=1,3,5,\dots}^N \sum_{n=1}^N \sum_{m=1}^N E_n(t_k) E_m(t_k) \phi_n^{(2i-1)} \phi_m^{(2i-1)} + \sum_{p=2,4,6,\dots}^N \sum_{n=1}^N \sum_{m=1}^N E_n(t_k) E_m(t_k) \phi_n^{2i} \phi_m^{2i} \right] \quad (6) \\ &= \frac{1}{K} \sum_{k=1}^K \sum_{n=1}^N E_n^2(t_k) = \sum_{n=1}^N \lambda_n \end{aligned}$$

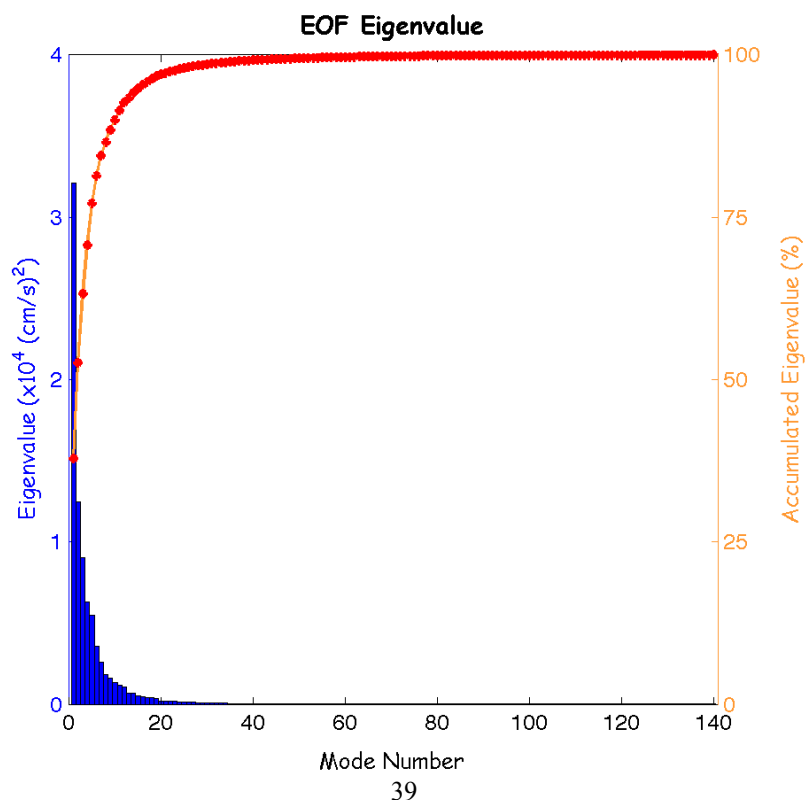
式中 $\delta_{nn} = \sum_{p=1}^N \phi_n^p \phi_n^p = 1$ ，是以特徵值代表可解釋之變異數。

將 2011 年 4 月 14 日至 2012 年 11 月 30 日完全無缺值且經濾潮後之 CODAR 資料(共 13280 組)按照(1)到(5)式進行分析，可以求出 140 個模組向量以及對應之

特徵值與各模組振幅。圖 3-1 為特徵值之分佈情形，圖右側軸則為各模組特徵值相對於全部 140 個模組特徵值總和之百分比，圖中之直方圖表示各模組可解釋之變異數的大小，橘線紅點則代表可解釋之變異數的累積百分比，從圖 3-1(a)可看出前 20 個模組即可解釋原始變動流速資料總變異數的 97.02%，而其餘的 120 個模組則只能影響剩下的 2.98%總變異數。至於前 20 個模組(圖 3-1b)中又以第 1 模組及第 2 模組可解釋總變異數的比例最大，二者合計約佔 53%，而第 3~20 模組之特徵值則呈指數衰減，此 18 個模組特徵值總和約佔總變異數的 43%。

此外，取前 20 個模組以重建合成變動流速資料，再分別計算各資料點所能還原的可解釋之變異數大小(如圖 3-2)，並依(6)式計算重建流場變異數與原始觀測流場總變異數之百分比稱為還原度，發現在 SCONET 海域之東部以及東南部還原程度較小，約為 94~97%，但在其餘區域還原度均在 97%以上。由此特性可知在離雷達站較遠處，第 21~140 模組對總變異數之貢獻度會比離雷達站近處稍大一些，因此造成前者還原度較小。

(a)



(b)

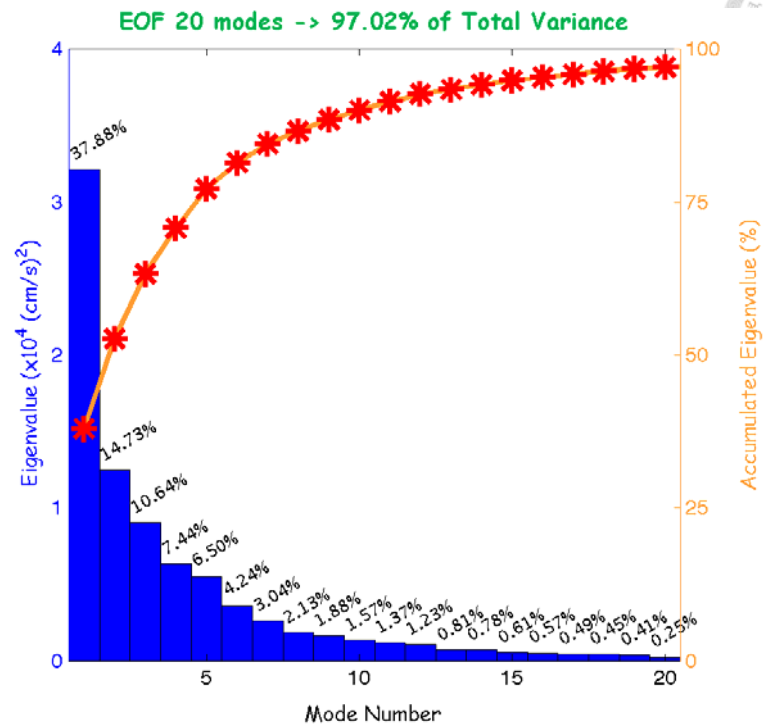


圖 3-1 實向量 EOF 分析法重建流場之特徵值分佈圖與累積特徵值分佈圖：(a)全部 140 個模組；(b)前 20 個模組。

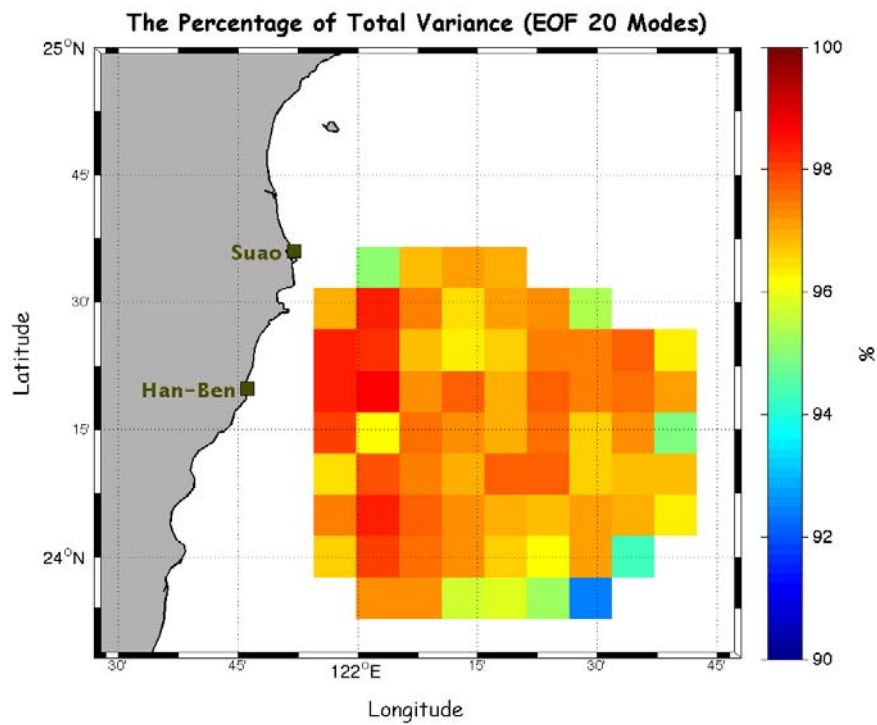


圖 3-2 實向量 EOF 法前 20 個模組重建流場與變動流場之變異數百分比空間分佈圖。


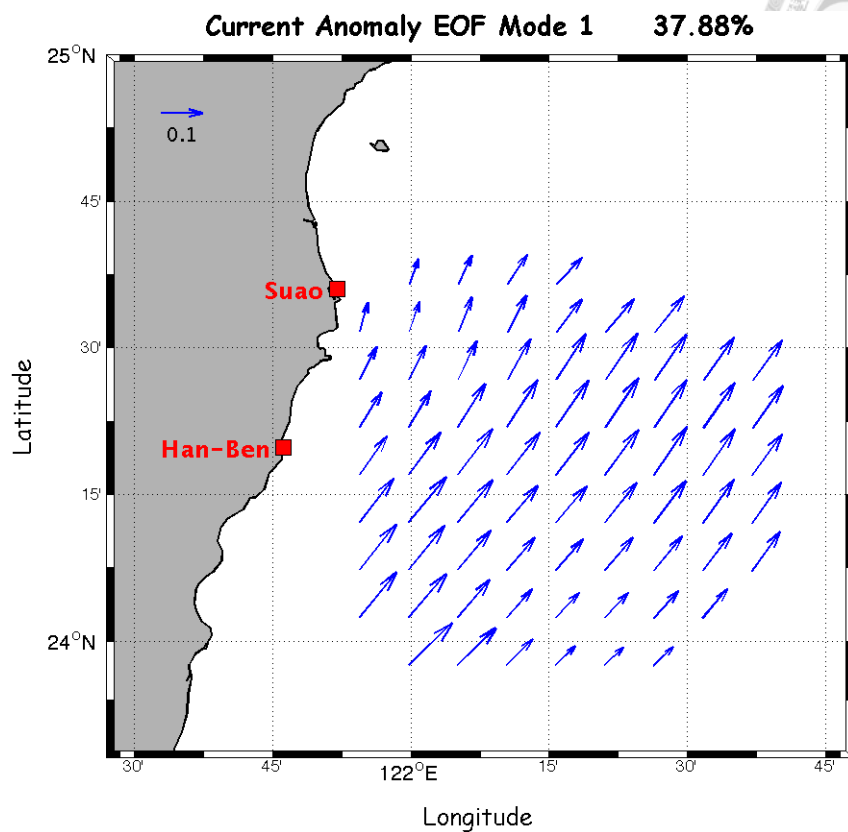


圖3-3~圖3-5顯示最主要的三個模組向量之水平分佈(因篇幅關係，其餘模組未列入)，各圖中(a)是各模組向量之空間分佈，(b)是該模組對應的振幅時間序列，將振幅時間序列與空間分佈相乘後，即是該模組所代表的流場向量隨時間的變化情形。在此三模組中，第1模組之特徵值(即其振幅之變異數)佔變動流速資料總變異數的37.88%，其模組向量之分佈態勢(圖3-3a)與平均流場甚為相似，流場向量的主要方向大致上是指向東北，只有在蘇澳附近較為偏北；再由圖3-3(b)來看，第1模組振幅為負值的時間主要約在10~2月之間，表示這段期間第1模組流場呈現的是西南向的海流，可能和這段期間冬季東北季風較強有關，因此第1模組所反映的應是平均流場之消弱或增長的效應。其次，第2模組特徵值佔變動流速資料總變異數的14.73%，圖3-4(a)是其模組向量之分佈情形，顯示似乎是一股沿宜蘭海岸的南向海流和另一股從花蓮北上的東北向海流在 24.25°N 附近會合之後向東流出的情形；圖3-4(b)則為對應之振幅時間序列，可發現在春夏季節(振幅多為正)此現象較明顯，至於秋冬時(振幅多為負)則表示為一股西向海流在接近臺灣海岸後於 24.25°N 附近分別向南、向北流。至於第3模組之特徵值則佔變動流速資料總變異數的10.64%，觀察圖3-5(a)與圖3-5(b)，顯示此模組係表示在 122.25°E 以西有一股由西南向東北的海流慢慢轉向東，最後再轉向南方的情形，猜測如果觀測範圍能夠更大，這種流況或許顯示的是一個渦流(eddy)之部份構造。

(a)



(b)

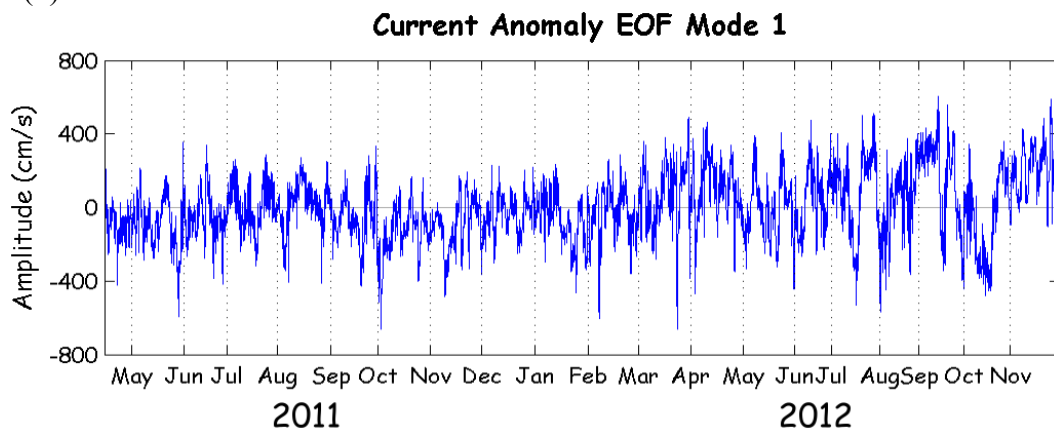
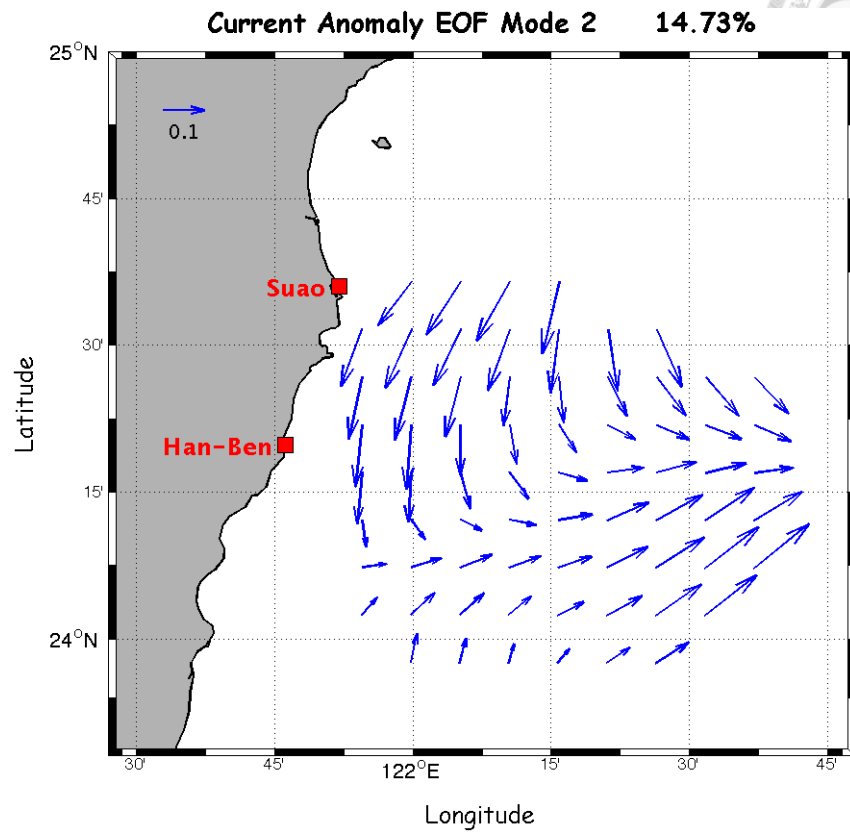


圖 3-3 實向量 EOF 分析法求出之第 1 模組：(a)模向量空間分佈；(b)模振幅之時間序列。

(a)



(b)

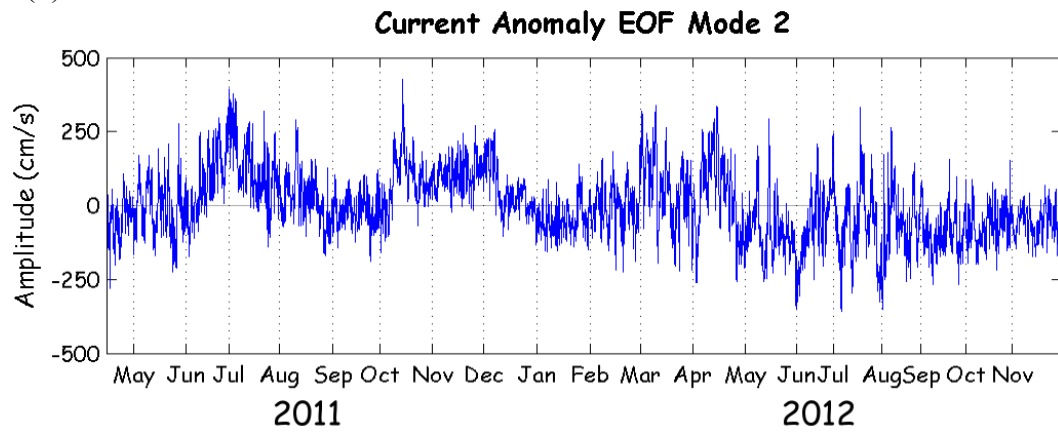


圖 3-4 實向量 EOF 分析法求出之第 2 模組：(a)模向量空間分佈；(b)模振幅之時間序列。

使用實向量EOF法分析除了可以找出影響流場的主要因素外，不同階的模組亦可反映不同尺度之運動成份，其中高階成份往往與尺度小的變化有關。當以有限模組數利用(5)式重建變動流場時，由於變動流速在各資料點上所能解釋的總變異數並不同，其中若某資料點之還原度不好，則可能是因雜訊干擾影響較明顯所致，因此利用模組重組亦可檢查資料品質，進而過濾掉小尺度運動之影響。在以下的分析我們使用前20個模組重建出流場，並計算重建流場與原始觀測流場二者間流速偏差的均方根(root mean square, RMS)，即：

$$RMS(x_i) = \sqrt{\frac{\sum_{k=1}^K [(u_R(x_i, t_k) - u_O(x_i, t_k))^2 + (v_R(x_i, t_k) - v_O(x_i, t_k))^2]}{K}} \quad (7)$$

上式中 u_R 、 v_R 表示重建流場的東西向、南北向流速， u_O 及 v_O 則分別是原始觀測流場的東西向與南北向流速， K 是資料筆數。(7)式可表示用前20個模組所重建之流場與原始觀測流場的偏差程度。

圖3-6顯示20個模組重建流場之偏差均方根分佈情形，表3-1則為其詳細數據(第一行為資料點編號，位置參照圖2-5)，整體而言偏差均方根大致介於3.68~8.40 cm/s之間，而70個點偏差均方根之平均為5.92 cm/s。從圖3-6中可看出SCONET海域外圍及北部資料點的偏差均方根較大，表示還原程度較差，而在中央區偏差均方根較小，還原程度較佳。

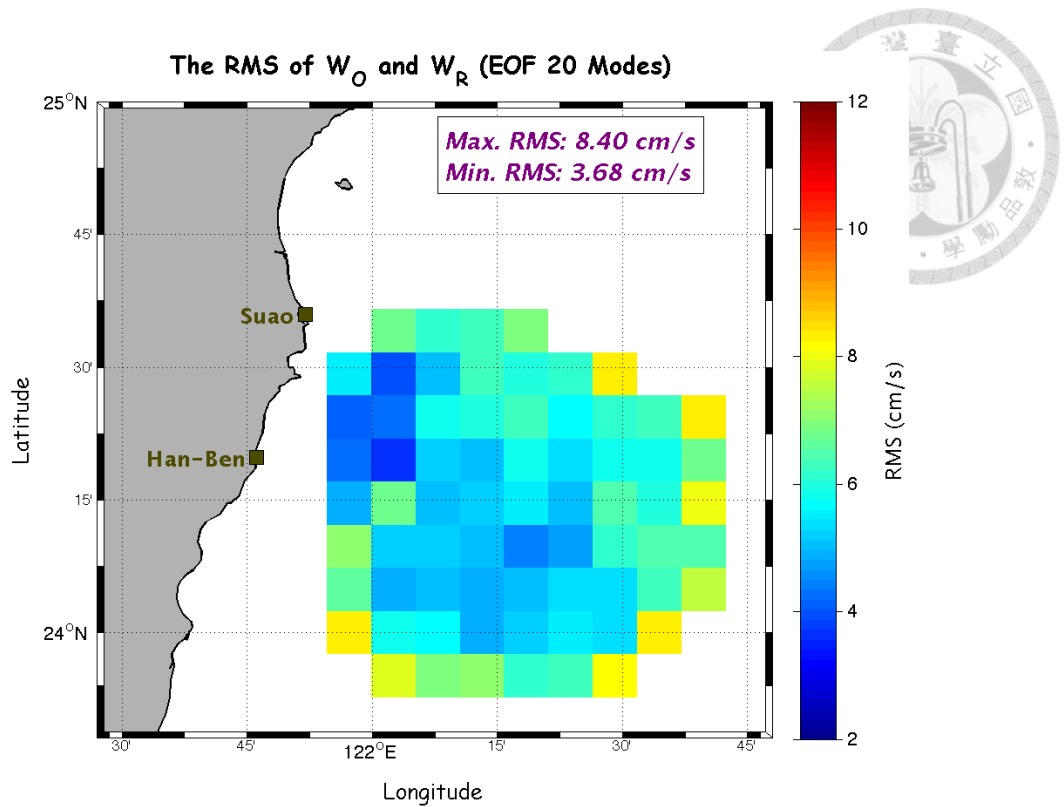


圖 3-6 原始觀測流場與實向量 EOF 法前 20 個模組重建流場之偏差均方根分佈圖。

表 3-1 原始觀測流場與實向量 EOF 法前 20 個模組重建流場之偏差均方根數值。

資料點 編號	RMS (cm/s)	資料點 編號	RMS (cm/s)	資料點 編號	RMS (cm/s)	資料點 編號	RMS (cm/s)	資料點 編號	RMS (cm/s)
1	8.36	15	4.03	29	5.06	43	7.00	57	5.79
2	6.55	16	6.82	30	5.17	44	6.25	58	6.09
3	7.02	17	6.87	31	5.03	45	5.53	59	8.38
4	4.83	18	5.65	32	6.00	46	5.43	60	8.40
5	4.33	19	5.03	33	6.29	47	4.69	61	6.31
6	4.14	20	5.15	34	6.30	48	5.07	62	6.47
7	5.46	21	5.10	35	6.09	49	5.29	63	5.97
8	7.85	22	5.26	36	5.13	50	5.73	64	5.88
9	5.76	23	5.76	37	5.11	51	6.21	65	6.29
10	4.83	24	5.03	38	4.47	52	8.13	66	7.59
11	5.13	25	6.08	39	5.47	53	5.31	67	6.47
12	6.78	26	7.05	40	5.81	54	5.42	68	7.95
13	3.68	27	4.95	41	6.30	55	6.16	69	6.78
14	4.31	28	4.82	42	6.02	56	6.50	70	8.38



3.2 Karhunen-Loève 展開法

Karhunen-Loève 展開法(KLE)亦稱為 Karhunen-Loève 分解(Karhunen-Loève decomposition, KLD)或是特徵正交分解(Proper Orthogonal Decomposition, POD)，在實際應用時其本質和前節所述之經驗正交函數類似，主要皆為將資料降階(order reduction)處理並提取(extraction)隱藏其中的特徵來表現資料以及進行分析(Preisendorfer, 1988; Kerschen et al., 2005)。這種方法最早是用在圖像辨識(pattern recognition)領域(Fukunaga, 1990)，之後則被應用到地球物理(Tiampo et al., 2004)以及紊流(turbulent flow) (Holmes et al., 1996)研究等方面。KLE 法和 EOF 法雖然概念相似，但兩者仍有差異，主要是 EOF 法使用共變異數矩陣，而 KLE 法則是使用相關矩陣(correlation matrix)計算特徵模組。

使用 KLE 法分析 SCONET 觀測區海流資料時，其主要步驟與前節之實向量 EOF 法相似，第一步亦是使用(1)式之變動流速(u, v)計算出(2)式之共變異數矩陣 R ，其矩陣大小亦為 140×140 ，接著將共變異數矩陣除以變動流速之標準差予以標準化成為相關矩陣 C ，即：

$$C_{pq} = \frac{R_{pq}}{\sigma_{pq}} \quad (8)$$

式中， σ 為標準差， C 以矩陣形式表示則為

$$C = [C_{pq}] = \begin{bmatrix} C_{11}^{uu} & C_{11}^{uv} & \cdots & C_{1M}^{uu} & C_{1M}^{uv} \\ C_{11}^{vu} & C_{11}^{vv} & & C_{1M}^{vu} & C_{1M}^{vv} \\ \vdots & & \ddots & \vdots & \\ C_{M1}^{uu} & C_{M1}^{uv} & \cdots & C_{MM}^{uu} & C_{MM}^{uv} \\ C_{M1}^{vu} & C_{M1}^{vv} & & C_{MM}^{vu} & C_{MM}^{vv} \end{bmatrix}, \quad p, q = 1, 2, \dots, N$$

$N = 140, M = \frac{N}{2} = 70.$

其中，

$$C_{pq}^{uu} = \frac{\frac{1}{K} \sum_{k=1}^K u(x_i, t_k) u(x_j, t_k)}{\sqrt{\frac{1}{K} \sum_{k=1}^K [u(x_i, t_k)]^2} \sqrt{\frac{1}{K} \sum_{k=1}^K [u(x_j, t_k)]^2}}$$



$$C_{pq}^{uv} = \frac{\frac{1}{K} \sum_{k=1}^K u(x_i, t_k) v(x_j, t_k)}{\sqrt{\frac{1}{K} \sum_{k=1}^K [u(x_i, t_k)]^2} \sqrt{\frac{1}{K} \sum_{k=1}^K [v(x_j, t_k)]^2}}$$

$$C_{pq}^{vu} = \frac{\frac{1}{K} \sum_{k=1}^K v(x_i, t_k) u(x_j, t_k)}{\sqrt{\frac{1}{K} \sum_{k=1}^K [v(x_i, t_k)]^2} \sqrt{\frac{1}{K} \sum_{k=1}^K [u(x_j, t_k)]^2}}$$

$$C_{pq}^{vv} = \frac{\frac{1}{K} \sum_{k=1}^K v(x_i, t_k) v(x_j, t_k)}{\sqrt{\frac{1}{K} \sum_{k=1}^K [v(x_i, t_k)]^2} \sqrt{\frac{1}{K} \sum_{k=1}^K [v(x_j, t_k)]^2}}, \quad i, j = 1, 2, \dots, M.$$

然後根據矩陣 C 可求出 N 個特徵值 $\Lambda_n, n=1, 2, \dots, N$ ，以及對應之 N 組互相正交的特徵向量 $\varphi_n, n=1, 2, \dots, N$ ，即：

$$\sum_{p=1}^N C_{pq} \varphi_n^p = \Lambda_n \varphi_n^q, \quad n=1, 2, \dots, N; q=1, 2, \dots, N. \quad (9)$$

$$\text{式中，} \quad \varphi_n = [\varphi_n^p]_{p=1, 2, \dots, N}^T = [\varphi_n^1, \varphi_n^2, \dots, \varphi_n^N]^T$$

$$= [\varphi_n^u(x_1), \varphi_n^v(x_1), \varphi_n^u(x_2), \varphi_n^v(x_2), \dots, \varphi_n^u(x_M), \varphi_n^v(x_M)]^T$$

φ_n 是第 n 模組之特徵向量，亦稱為模組或基底向量。最後計算變動流速

$$w = [w_p] = [u(x_1), v(x_1), u(x_2), v(x_2), \dots, u(x_M), v(x_M)] \quad (p=1, 2, \dots, N) \text{ 對各組特徵向}$$

量之投影作為振幅，即：



$$\begin{aligned}
A_n(t_k) &= \sum_{p=1}^N w_p(t_k) \phi_n^p \\
&= [u(x_1), v(x_1), \dots, u(x_M), v(x_M)] \cdot \begin{bmatrix} \phi_n^u(x_1) \\ \phi_n^v(x_1) \\ \vdots \\ \phi_n^u(x_M) \\ \phi_n^v(x_M) \end{bmatrix} \\
&= \sum_{p=1,3,5,\dots}^N u\left(x_{\frac{p+1}{2}}, t_k\right) \phi_n^u\left(x_{\frac{p+1}{2}}\right) + \sum_{p=2,4,6,\dots}^N v\left(x_{\frac{p}{2}}, t_k\right) \phi_n^v\left(x_{\frac{p}{2}}\right) \\
&= \sum_{p=1,3,5,\dots}^N u\left(x_{\frac{p+1}{2}}, t_k\right) \phi_n^p + \sum_{p=2,4,6,\dots}^N v\left(x_{\frac{p}{2}}, t_k\right) \phi_n^p
\end{aligned} \tag{10}$$

式中 A_n 代表 t_k 時間第 n 模組之振幅。此外，使用所有模組之振幅時序資料亦可以

重新合成出變動流速，即：

$$\begin{aligned}
w_p(t_k) &= \sum_{n=1}^N A_n(t_k) \phi_n^p \\
\begin{bmatrix} u(x_1) \\ v(x_1) \\ \vdots \\ u(x_M) \\ v(x_M) \end{bmatrix} &= \sum_{n=1}^N A_n(t_k) \begin{bmatrix} \phi_n^u(x_1) \\ \phi_n^v(x_1) \\ \vdots \\ \phi_n^u(x_M) \\ \phi_n^v(x_M) \end{bmatrix} \\
u(x_i, t_k) &= \sum_{n=1}^N A_n(t_k) \phi_n^{(2i-1)}, \quad i = 1, 2, \dots, M \\
v(x_i, t_k) &= \sum_{n=1}^N A_n(t_k) \phi_n^{2i}, \quad i = 1, 2, \dots, M
\end{aligned} \tag{11}$$

至於模組之振幅時序資料與變動流速總變異數二者間有相等之關係，即由(10)

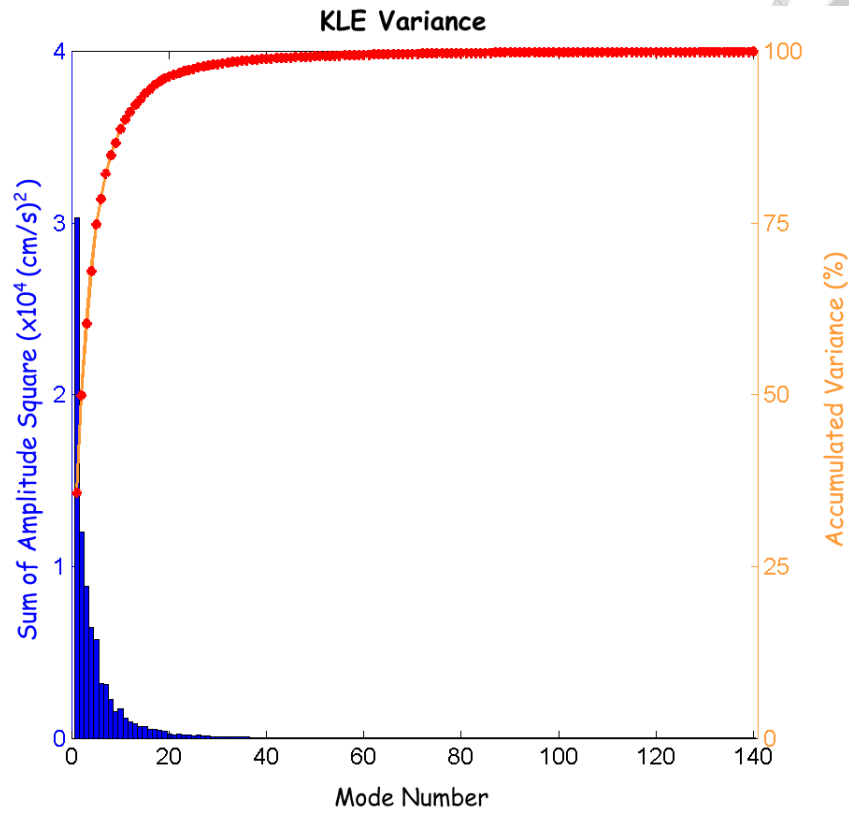
式可得出：

$$\begin{aligned}
\frac{1}{K} \sum_{k=1}^K \sum_{n=1}^N A_n^2(t_k) &= \frac{1}{K} \sum_{k=1}^K \left[\sum_{p=1,3,5,\dots}^N u^2 \left(x_{\frac{p+1}{2}}, t_k \right) \varphi_n^p \varphi_n^p + \sum_{p=2,4,6,\dots}^N v^2 \left(x_{\frac{p}{2}}, t_k \right) \varphi_n^p \varphi_n^p \right] \\
&= \frac{1}{K} \sum_{k=1}^K \left[\sum_{p=1,3,5,\dots}^N u^2 \left(x_{\frac{p+1}{2}}, t_k \right) \delta_{nn} + \sum_{p=2,4,6,\dots}^N v^2 \left(x_{\frac{p}{2}}, t_k \right) \delta_{nn} \right] \quad (12) \\
&= \frac{1}{K} \sum_{k=1}^K \left[\sum_{p=1,3,5,\dots}^N u^2 \left(x_{\frac{p+1}{2}}, t_k \right) + \sum_{p=2,4,6,\dots}^N v^2 \left(x_{\frac{p}{2}}, t_k \right) \right]
\end{aligned}$$

式中 $\sum_{p=1}^N \varphi_n^p \varphi_n^p = \delta_{nn} = 1$ ，是以振幅平方和能代表該模組可解釋之變動流速變異數。

將 2011 年 4 月 14 日至 2012 年 11 月 30 日(共 13280 組)經濾潮後的 CODAR 資料按照(1)、(2)以及(8)至(11)式進行分析，可以求出 140 個模組向量以及對應之各模組振幅。圖 3-7 為各模組振幅平方和之分佈情形，圖右側軸表示各模組振幅平方和與全部 140 個模組振幅平方和總和之百分比，圖中之直方圖表示各模組可解釋之變異數的大小，橘線紅點是指可解釋之變異數的累積百分比；從圖 3-7(a)即可發現使用前 20 個模組就能解釋所有原始變動流速資料總變異數的 96.43%，而其餘的 120 個模組僅能影響剩下的 3.57%總變異數，至於前 20 個模組(圖 3-7b)中第 1、第 2 模組可解釋總變異數的比例總和可達約 50%，第 3~20 模組之振幅平方和則呈指數衰減，此 18 個模組振幅平方和之總和約佔總變異數的 46%。

(a)



(b)

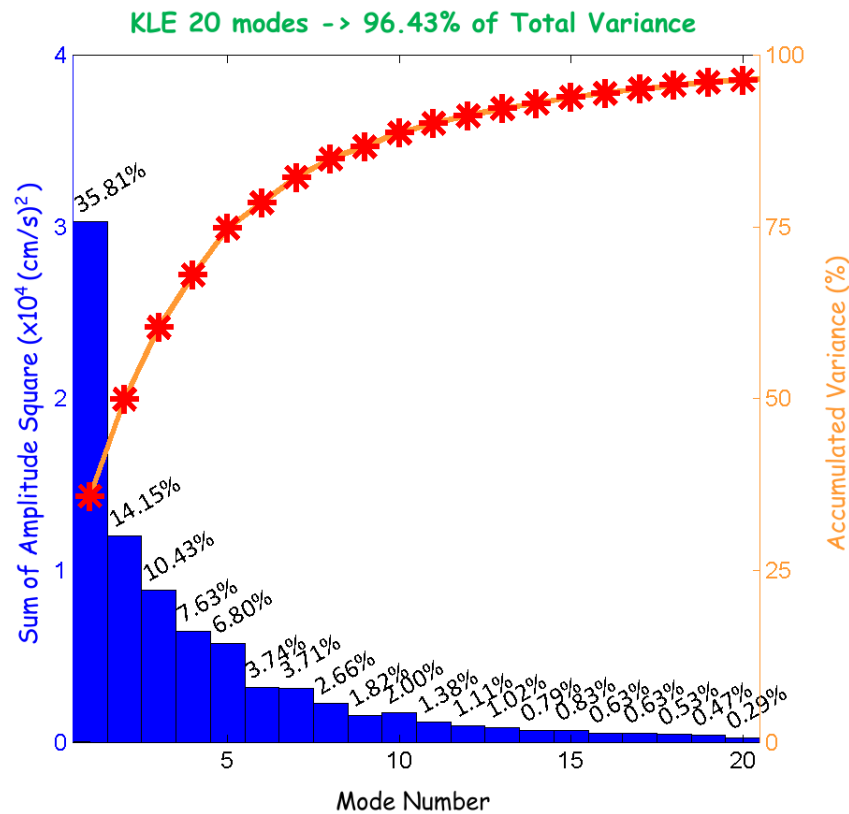


圖 3-7 KLE 分析法重建流場之振幅平方和分佈圖與累積變異數分佈圖：(a)全部 140 個模組；(b)前 20 個模組。

取前 20 個模組以重建合成變動流速資料，再分別計算各資料點所能還原的可解釋之變異數大小(圖 3-8)，並依(11)式計算重建流場變異數與原始觀測流場總變異數之百分比稱為還原度，可發現觀測區內大部分區域的還原度可達 96%以上，僅少數資料點小於 90%。細觀圖 3-8，可知還原度較差的資料點幾乎都分佈在觀測區外圍，也就是在這些資料點上高階之第 21~140 模組對總變異數的貢獻度較大，因此造成還原度較小。

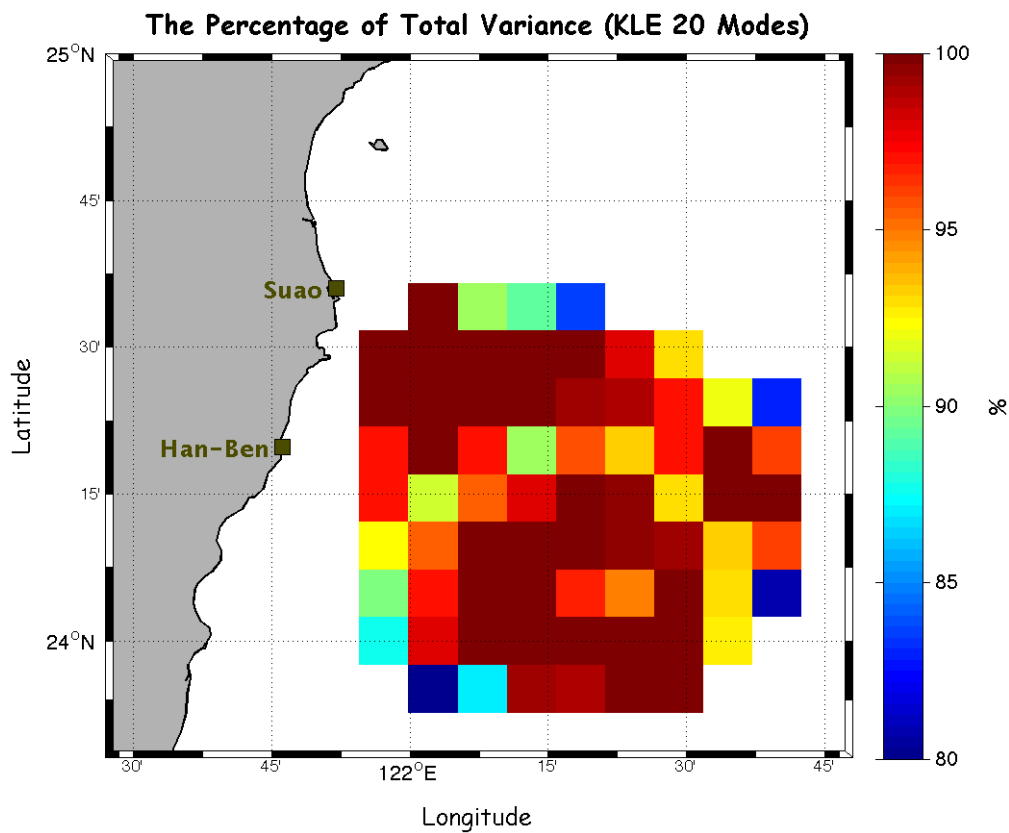
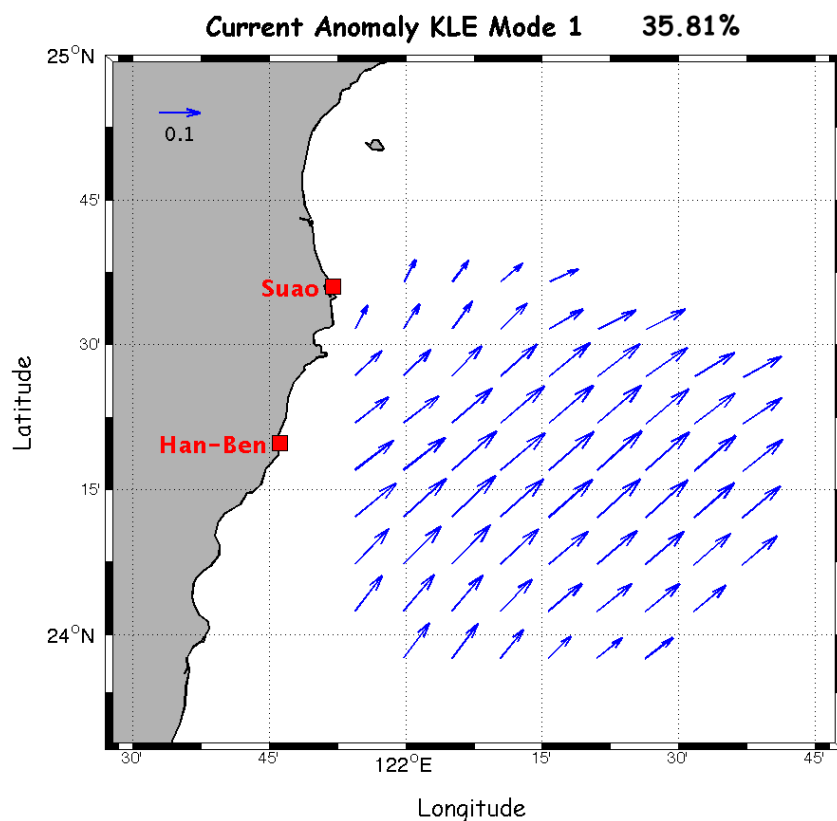


圖 3-8 KLE 法前 20 個模組重建流場與變動流場之變異數百分比空間分佈圖。

圖 3-9~圖 3-11 顯示 KLE 法最主要的三個模組向量之水平分佈(因篇幅關係，其餘模組未列入)，各圖中(a)是各模組向量的空間分佈，(b)圖則是該模組相對應之振幅時間序列，將振幅時間序列與空間分佈相乘後，即是各模組所代表的流場向量隨時間的變化情形。在此三個模組中，第 1 模組特徵值(即其振幅之變異數)佔所

有變動流速資料總變異數的 35.81%，其模組向量之分佈態勢如圖 3-9(a)，流場向量的主要方向為東北方向，與平均流場類似，不過在蘇澳附近流向較偏北；再由圖 3-9(b)來看，振幅為負值，即代表圖 3-9(a)之流向是相反的西南向海流，此情形之發生時間約在 10~2 月之間，可能與此段時間為東北季風盛行期有關，故此模組與實向量 EOF 第 1 模相似，亦應是反映出平均流場的消長效應。其次，第 2 模組特徵值佔變動流速資料總變異數的 14.15%，根據圖 3-10(a)的模組向量分佈，顯示似乎是一股從北方來且沿著宜蘭海岸的海流與另一股從南方來的沿岸流，於 24.25°N 處附近交會後往東流出之情形；觀察相對應之振幅時間序列(圖 3-10b)，此現象在春、夏二季(振幅多為正)似乎較明顯，秋、冬季(振幅多為負)大部分是反向，即有一股海流自東方沿 24.25°N 處附近在接近臺灣海岸後再分別轉向北方與南方流。至於第 3 模組之特徵值則佔變動流速資料總變異數的 10.43%，觀察圖 3-11(a)、圖 3-11(b)，顯示應為一股沿岸流向東流後逐漸轉向南方的情形。

(a)



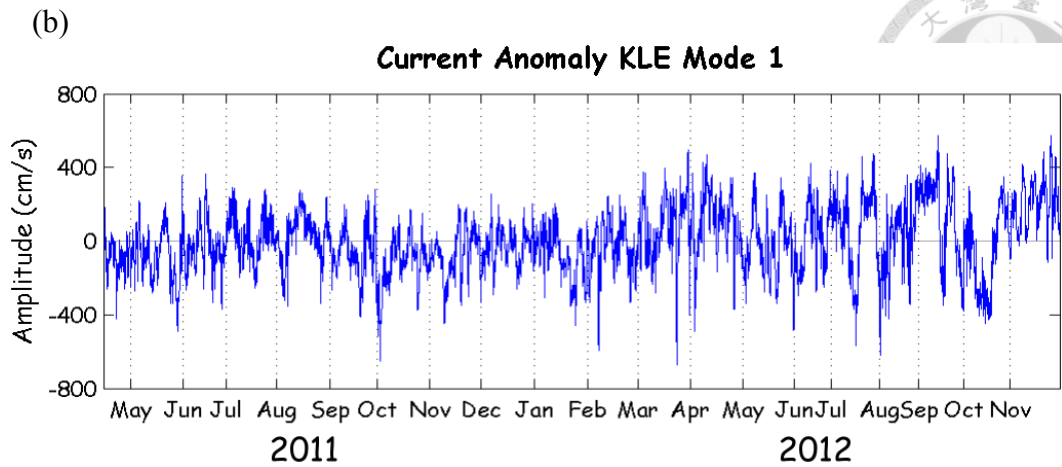
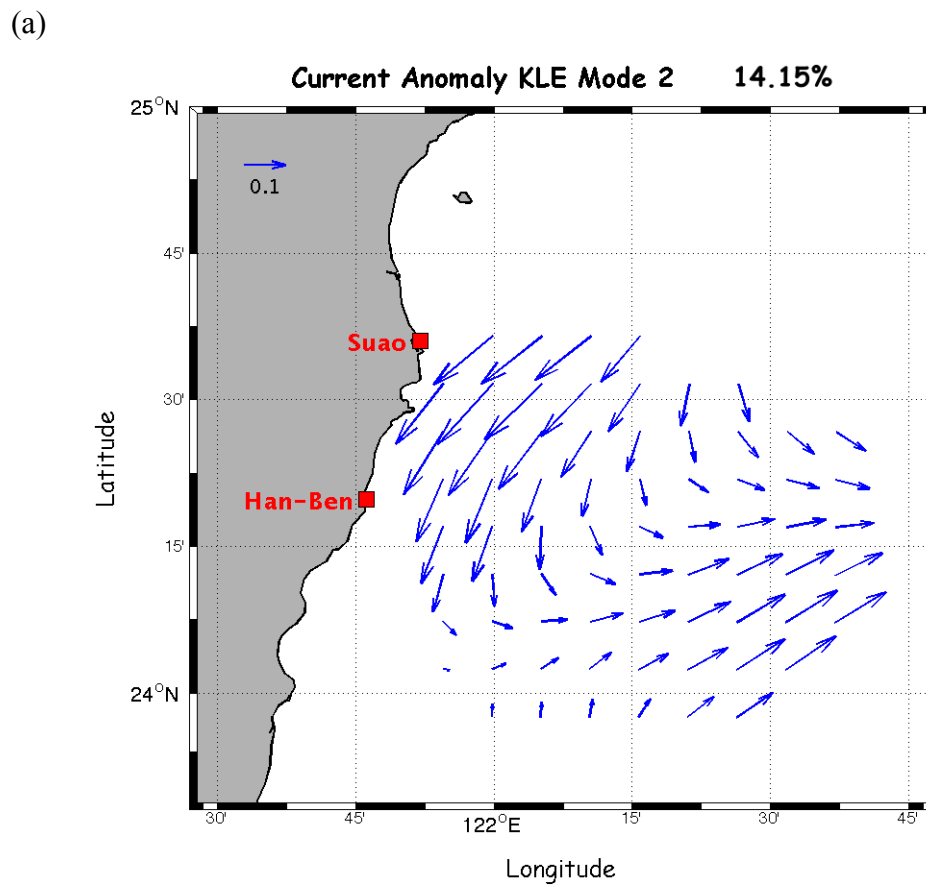


圖 3-9 KLE 分析法求出之第 1 模組：(a)模向量空間分佈；(b)模振幅之時間序列。



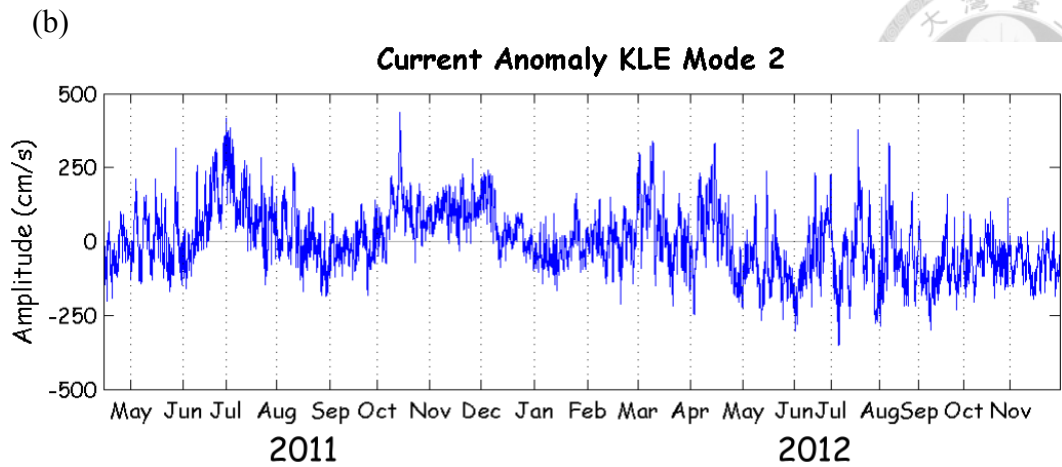
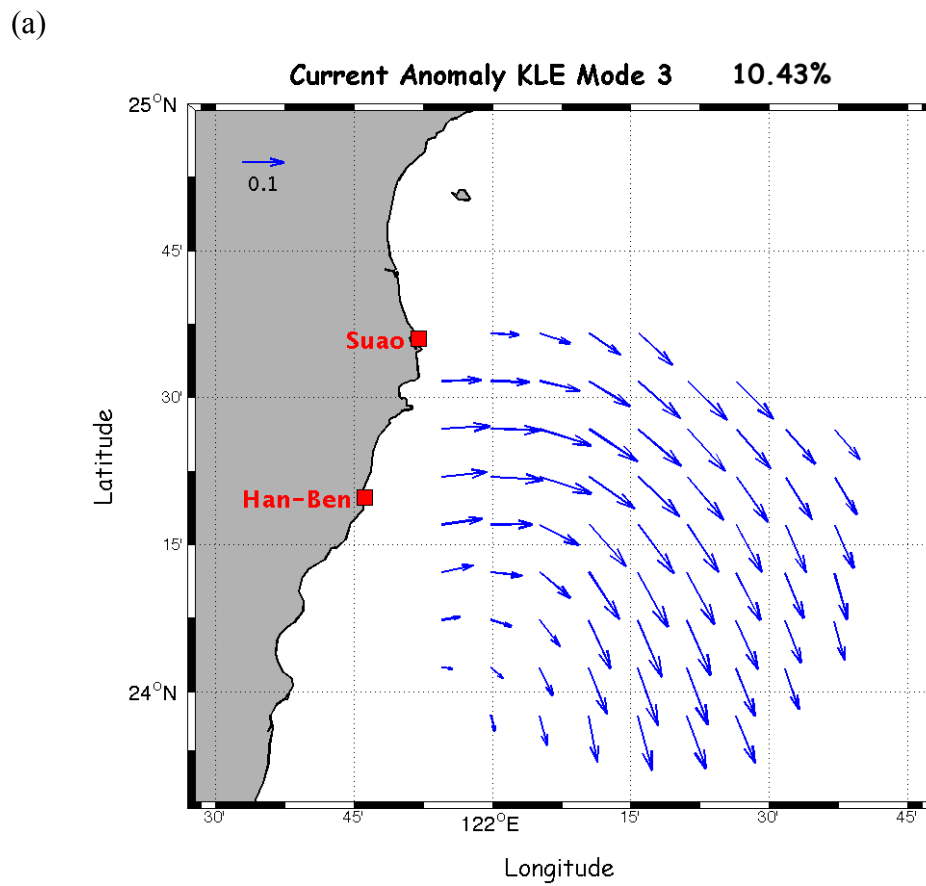


圖 3-10 KLE 分析法求出之第 2 模組：(a)模向量空間分佈；(b)模振幅之時間序列。



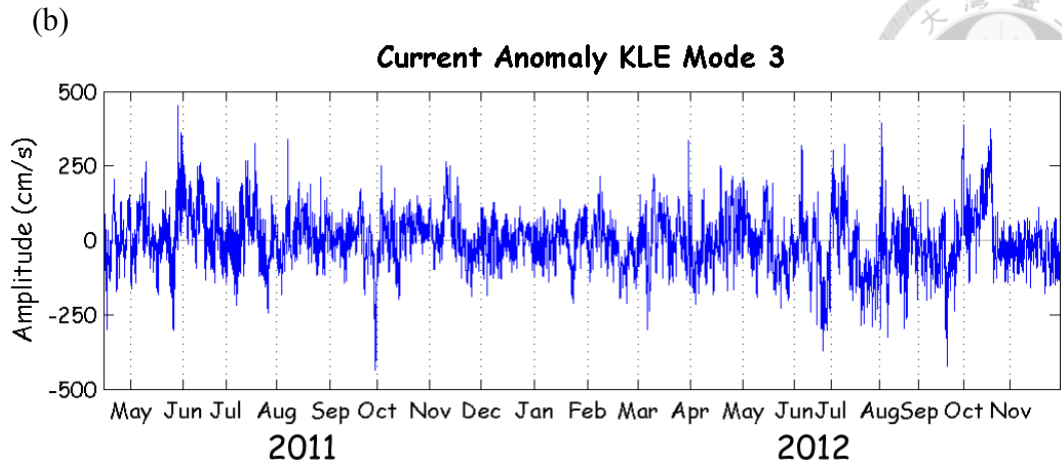


圖 3-11 KLE 分析法求出之第 3 模組：(a)模向量空間分佈；(b)模振幅之時間序列。

使用 KLE 法分析亦可以找出影響流場的主要因素，以及利用不同階模組反映出不同尺度之運動成份(高階模組成份與尺度小的變化有關)。當以有限模組數利用(11)式重建變動流場時，由於變動流速在各資料點上所能解釋之總變異數並不相同，其中若某資料點的還原度不好，則可能是因雜訊干擾影響較明顯所致，因此利用模組重組亦能檢查資料品質並濾除掉小尺度運動之影響。接下來我們亦使用前 20 個模組來重建流場，並利用(7)式計算重建流場與原始觀測流場之間流速偏差的均方根(RMS)，藉以觀察前 20 個模組所重建出的流場與原始觀測流場之間的偏差程度。

KLE 法前 20 個模組重建流場與原始觀測流場之間的偏差均方根其數據詳如表 3-2(表中第一行為資料點位編號，對應位置參照圖 2-5)。整體而言 KLE 法之偏差均方根範圍介於 3.79~11.38 cm/s，全部 70 個資料點偏差均方根之平均為 6.39 cm/s。圖 3-12 為 KLE 法偏差均方根之水平分佈情形，由該圖可看出大部分資料點的偏差均方根皆小於 7 cm/s，離雷達站較遠或分佈在 SCNET 海域外圍之資料點其偏差均方根較大，還原程度較差，而離雷達站較近處的偏差均方根則較小，還原程度較佳。

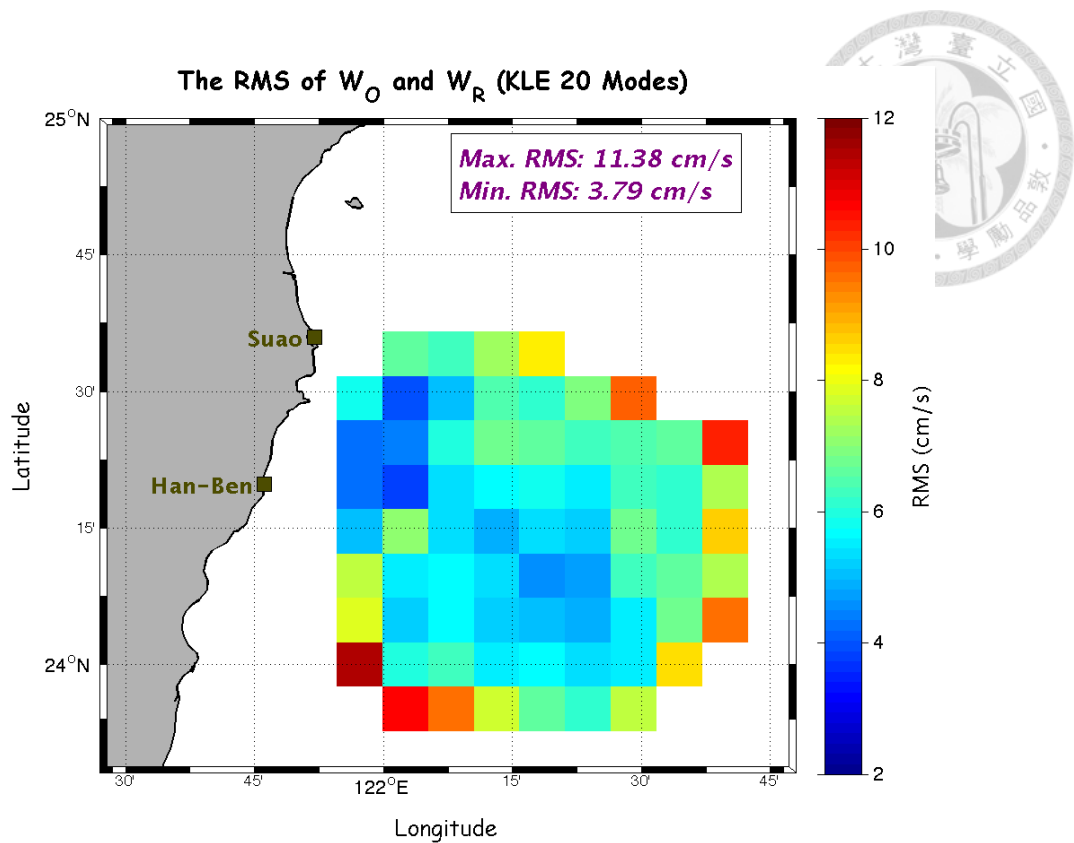


圖 3-12 原始觀測流場與KLE 法前 20 個模組重建流場之偏差均方根空間分佈圖。

表 3-2 原始觀測流場與 KLE 法前 20 個模組重建流場之偏差均方根數值。

資料點 編號	RMS (cm/s)	資料點 編號	RMS (cm/s)	資料點 編號	RMS (cm/s)	資料點 編號	RMS (cm/s)	資料點 編號	RMS (cm/s)
1	11.38	15	4.02	29	5.29	43	8.27	57	6.22
2	7.88	16	6.57	30	4.94	44	6.15	58	6.48
3	7.61	17	9.54	31	5.75	45	5.31	59	9.72
4	5.01	18	6.35	32	6.73	46	4.94	60	8.44
5	4.21	19	5.68	33	6.45	47	4.71	61	6.74
6	4.27	20	5.71	34	7.23	48	5.18	62	6.64
7	5.82	21	5.29	35	6.61	49	5.58	63	6.13
8	10.67	22	5.43	36	5.75	50	6.32	64	6.07
9	5.99	23	6.04	37	5.12	51	6.92	65	6.66
10	5.22	24	5.10	38	4.61	52	7.49	66	9.65
11	5.55	25	6.24	39	5.36	53	5.59	67	7.32
12	7.05	26	7.66	40	5.77	54	5.57	68	8.67
13	3.79	27	5.56	41	6.64	55	6.27	69	7.46
14	4.47	28	5.13	42	6.07	56	6.69	70	10.32



3.3 以模重組法填補資料缺漏

在觀測區內各點資料均完整無任何缺漏值的情形下，不論使用實向量 EOF 法或是 KLE 法，其所分解出的特徵向量(ϕ_n 或 φ_n)均能滿足正交條件，即：

$$\sum_{p=1}^N \phi_n^p \phi_m^p = \delta_{nm} = \begin{cases} 0 & \text{if } n \neq m \\ 1 & \text{if } n = m \end{cases}$$

但若資料不完整即有缺漏情況時，則 $\sum_{p=1}^{N-2S} \phi_n^p \phi_m^p$ 就可能不符合正交條件(假設 S 為資料缺漏之個數， $S \neq 0$)。因此在資料不完整有缺漏情況時，便不宜直接使用(4)式或(10)式來計算各模組特徵向量對應之振幅，而需使用最小平方法(least square)或是迭代法(iteration)求解(Boyd et al., 1994; Beckers and Rixen, 2003; Alvera-Azcárate et al., 2005)，以下將分別對兩種方法進行說明。

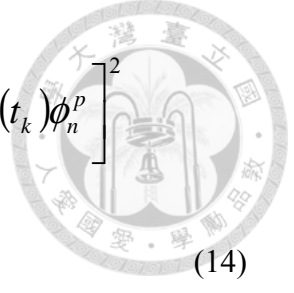
3.3.1 最小平方法

若將重建流場與原始觀測流場之變動流速分別以 (u_R, v_R) 與 (u_O, v_O) 表示， (u_R, v_R) 是藉由後者經模分析後得出之模組向量 ϕ_n 所推算出之流場(假設使用 L 個模組)， E'_n 是 ϕ_n 之振幅，為未知係數，即：

$$u_R(x_i, t_k) = \sum_{n=1}^L E'_n(t_k) \phi_n^p, \quad p = 2i - 1$$

$$v_R(x_i, t_k) = \sum_{n=1}^L E'_n(t_k) \phi_n^p, \quad p = 2i; i = 1, 2, \dots, M; M = 70. \quad (13)$$

令 $w = [w_p] = [u(x_1), v(x_1), u(x_2), v(x_2), \dots, u(x_M), v(x_M)]$ ($p = 1, 2, \dots, N$)，並以 $[w_p^O]$ 代表原始觀測流場、 $[w_p^R]$ 表示重建流場，由(13)式可知 $[w_p^O]$ 和 $[w_p^R]$ 的總誤差可表示為：



$$\begin{aligned}
\delta &= \sum_{p=1}^{N-2S} [w_p^O(t_k) - w_p^R(t_k)]^2 = \sum_{p=1}^{N-2S} \left[w_p^O(t_k) - \sum_{n=1}^L E'_n(t_k) \phi_n^p \right]^2 \\
&= \sum_{p=1}^{N-2S} w_p^O(t_k) w_p^O(t_k) - 2 \sum_{p=1}^{N-2S} \sum_{n=1}^L E'_n(t_k) w_p^O(t_k) \phi_n^p \\
&\quad + \sum_{p=1}^{N-2S} \sum_{m=1}^L \sum_{n=1}^L E'_m(t_k) E'_n(t_k) \phi_m^q \phi_n^p
\end{aligned} \tag{14}$$

令 $C_{xx} = w_p^O(t_k) w_p^O(t_k)$ 、 $C_{xn} = w_p^O(t_k) \phi_n^p$ 以及 $A_{mn} = \phi_m^q \phi_n^p$ ，則(14)式可改寫成

$$\delta = \sum_{p=1}^{N-2S} C_{xx} - 2 \sum_{p=1}^{N-2S} \sum_{n=1}^L E'_n(t_k) C_{xn} + \sum_{p=1}^{N-2S} \sum_{m=1}^L \sum_{n=1}^L E'_m(t_k) E'_n(t_k) A_{mn} \tag{15}$$

將(15)式對 $E'_n(t_k)$ 微分，並令

$$\frac{\partial \delta}{\partial E'_m(t_k)} = 0 \tag{16}$$

如此可得出 L 個聯立方程式：

$$-2 \sum_{p=1}^{N-2S} C_{xm} + 2 \sum_{p=1}^{N-2S} \sum_{n=1}^L E'_n(t_k) A_{mn} = 0, \quad m = 1, \dots, L. \tag{17}$$

由此得出之 E'_n 即為可使(14)式所示誤差為最小的解，由(13)式以及(17)式可得

$$w_p^R(t_k) = \sum_{n=1}^L E'_n(t_k) \phi_n^p = \sum_{m=1}^L C_{xm} \sum_{n=1}^L A_{mn}^{-1} \phi_n^p$$

$$\begin{bmatrix} u_R(x_1) \\ v_R(x_1) \\ \vdots \\ u_R(x_M) \\ v_R(x_M) \end{bmatrix} = \sum_{n=1}^L E'_n(t_k) \begin{bmatrix} \phi_n^u(x_1) \\ \phi_n^v(x_1) \\ \vdots \\ \phi_n^u(x_M) \\ \phi_n^v(x_M) \end{bmatrix} = \sum_{m=1}^L C_{xm} \sum_{n=1}^L A_{mn}^{-1} \begin{bmatrix} \phi_n^u(x_1) \\ \phi_n^v(x_1) \\ \vdots \\ \phi_n^u(x_M) \\ \phi_n^v(x_M) \end{bmatrix}$$

$$u_R(x_i, t_k) = \sum_{n=1}^L E'_n(t_k) \phi_n^{(2i-1)} = \sum_{m=1}^L C_{xm} \sum_{n=1}^L A_{mn}^{-1} \phi_n^{(2i-1)}$$

$$v_R(x_i, t_k) = \sum_{n=1}^L E'_n(t_k) \phi_n^{2i} = \sum_{m=1}^L C_{xm} \sum_{n=1}^L A_{mn}^{-1} \phi_n^{2i} \quad (18)$$

式中 A_{mn}^{-1} 為 A_{mn} 的反矩陣，如此可得到一個與原始觀測流場誤差最小的重建流場之變動流速 (u_R, v_R)。

3.3.2 迭代法

當變動流速資料有缺漏情況時，若欲使用迭代法求解，首先需將缺漏處以0取代，

假設在 x_1 點有缺漏之變動流速 $w^{(0)} = [nan, nan, u(x_2), v(x_2), \dots, u(x_M), v(x_M)] = [w_p^{(0)}]$ ，

其中 nan 表示為資料缺漏處，如果以0取代之，則為：

$$w'^{(1)} = [0, 0, u(x_2), v(x_2), \dots, u(x_M), v(x_M)] = [w_p'^{(1)}]$$

再將 $w'^{(1)}$ 進行實向量EOF法或KLE法分析，並根據(5)或(11)式合成新的變動流速資料 $w^{*(1)}$ (使用 L 個模組)：

$$w_p^{*(1)}(t_k) = \sum_{n=1}^L E_n^{(1)}(t_k) \phi_n^{p(1)} \quad (\text{實向量EOF法})$$

$$w_p^{*(1)}(t_k) = \sum_{n=1}^L A_n^{(1)}(t_k) \phi_n^{p(1)} \quad (\text{KLE法})$$

即可得第1次猜測之填補值 $[w_p^{*(1)}] = [u^{*(1)}(x_1), v^{*(1)}(x_1), \dots, u^{*(1)}(x_M), v^{*(1)}(x_M)]$ 。

接下來，缺漏處重新以第1次猜測之填補值取代：

$$w'^{(2)} = [u^{*(1)}(x_1), v^{*(1)}(x_1), u(x_2), v(x_2), \dots, u(x_M), v(x_M)] = [w_p'^{(2)}]$$

同樣再將 $w'^{(2)}$ 乘以實向量EOF法或KLE法之模組求出振幅 $E_n^{(2)}$ ，並合成新的變動

流速 $w^{*(2)}$ ：

$$w_p^{*(2)}(t_k) = \sum_{n=1}^L E_n^{(2)}(t_k) \phi_n^{p(2)} \quad (\text{實向量EOF法})$$

$$w_p^{*(2)}(t_k) = \sum_{n=1}^L A_n^{(2)}(t_k) \phi_n^{p(2)} \quad (\text{KLE法})$$

此為第2次猜測之填補值 $\left[w_p^{*(2)} \right] = \left[u^{*(2)}(x_1), v^{*(2)}(x_1), \dots, u^{*(2)}(x_M), v^{*(2)}(x_M) \right]$ 。

之後則是重複 y 次前述取代與合成2步驟，即：

$$w'^{(y)} = \left[u^{*(y-1)}(x_1), v^{*(y-1)}(x_1), u(x_2), v(x_2), \dots, u(x_M), v(x_M) \right] = \left[w_p'^{(y)} \right]$$

$$w_p^{*(y)}(t_k) = \sum_{n=1}^L E_n^{(y)}(t_k) \phi_n^{p(y)} \quad (\text{實向量EOF法})$$

$$w_p^{*(y)}(t_k) = \sum_{n=1}^L A_n^{(y)}(t_k) \phi_n^{p(y)} \quad (\text{KLE法}) \quad (19)$$

在迭代過程中，合成的填補值與原始流速之間的偏差均方根會隨迭代次數 y 的增加而逐漸收斂，我們設定之收斂標準為當第 y 次與第 $y-1$ 次的偏差均方根之差值小於0.001或迭代次數大於200時即停止迭代。

第四章 模擬實驗與結果討論



本章係根據第三章所介紹之方法設計資料填補實驗，並選用了一個 month 之獨立 CODAR 觀測資料作為檢驗測試樣本，討論以實向量 EOF 法以及 KLE 法填補不完整資料之成效、誤差大小以及限制。

4.1 實驗設計

我們選用 2012 年 12 月完全沒有任何缺漏點的 CODAR 觀測資料組(共 734 組)作為填補實驗之真實值。令 $w_k = [u_{1k}, v_{1k}, u_{2k}, v_{2k}, \dots, u_{Mk}, v_{Mk}]$ 代表 t_k 時間之 CODAR 觀測流場資料($k=1,2,\dots,734$)， u_{ik} 與 v_{ik} ($i=1,\dots,M$) 則為 t_k 時觀測到的東西向以及南北向流速分量， M 為 CODAR 觀測區資料點個數($M=70$)。填補實驗是將這些資料以人為方式產生資料缺漏，也就是人為移除某些資料點的流速資料，然後假設在這段期間內這些資料點上沒有觀測資料並以特定方法填補缺值，再將填補值與真實值相比對以判別誤差大小，這種方法也是文獻上常用的作法(如 Boyd et al., 1994; Beckers and Rixen, 2003; Alvera-Azcárate et al., 2005)。對於 SCNET 海域的 70 個觀測點，發生資料不完整時其缺漏個數從 1 至 69 皆有可能(圖 2-4b)，不過由於我們使用實向量 EOF 法與 KLE 法的前 20 個模組來填補資料空缺，當以最小平方方法求解時至少需要生成 20 個方程式，因此資料缺漏點的上限為 60。

由排列組合公式知，70 個資料點中缺漏 S 點之變化種類共有 $C_S^{70} = \frac{70!}{(70-S)!S!}$ 種組合情形，因此缺 1 點時共有 $C_1^{70} = \frac{70!}{(70-1)!1!} = 70$ 種變化，缺 2 點時有 $C_2^{70} = \frac{70!}{(70-2)!2!} = 35 \times 69 = 2415$ 種變化，而當資料缺漏個數為 3 點時則有 $C_3^{70} = \frac{70!}{(70-3)!3!} = 70 \times 23 \times 34 = 54740$ 種變化，至於其他大於 3 之情況則可參閱表 4-1。

由表 4-1 知，當 $S > 3$ 時變化種類之數量級即達百萬次以上，以我們現有的有限計算設備實在無法將所有情況一一計算，因此必須採用不同策略。在 $S \leq 3$ 時，我們

採用窮舉法，即將所有的變化情形均予以計算；當 $S > 3$ 時則採用蒙地卡羅法(Monte Carlo simulation, Metropolis and Ulam, 1949)。蒙地卡羅模擬法是一種利用隨機亂數來進行大量模擬實驗並解決問題的方法，若實驗次數越多則得到之結果會越趨近於理論值。實際作法為：當缺漏個數大於 3 時，我們使用 Matlab 的亂數產生器，依照缺漏個數的多寡隨機產生 1~70 間的整數作為缺漏資料點之編號，透過這種隨機方式製造資料空缺再予以填補。在實際計算前，我們先做了前級的模擬試算，其結果如圖 4-1，顯示在實驗次數少時填補值與真實值的誤差均方根並不穩定，但隨實驗次數變多(2000 次以上時)則會逐漸穩定且收斂，因此我們在缺漏個數為 4~60 時，係使用蒙地卡羅法執行 3000 次模擬實驗再求取填補值與真實值之偏差均方根，作為研判結果良窳與否之根據。

表 4-1 70 個資料點不同缺漏個數情況之變化種數。

缺漏個數	變化種數	缺漏個數	變化種數	缺漏個數	變化種數	缺漏個數	變化種數
1	70	16	2.48×10^{15}	31	7.14×10^{19}	46	3.51×10^{18}
2	2415	17	7.88×10^{15}	32	8.70×10^{19}	47	1.79×10^{18}
3	54740	18	2.32×10^{16}	33	1.00×10^{20}	48	8.58×10^{17}
4	9.17×10^5	19	6.35×10^{16}	34	1.09×10^{20}	49	3.85×10^{17}
5	1.21×10^7	20	1.62×10^{17}	35	1.12×10^{20}	50	1.62×10^{17}
6	1.31×10^8	21	3.85×10^{17}	36	1.09×10^{20}	51	6.35×10^{16}
7	1.20×10^9	22	8.58×10^{17}	37	1.00×10^{20}	52	2.32×10^{16}
8	9.44×10^9	23	1.79×10^{18}	38	8.70×10^{19}	53	7.88×10^{15}
9	6.50×10^{10}	24	3.51×10^{18}	39	7.14×10^{19}	54	2.48×10^{15}
10	3.97×10^{11}	25	6.46×10^{18}	40	5.53×10^{19}	55	7.21×10^{14}
11	2.16×10^{12}	26	1.12×10^{19}	41	4.05×10^{19}	56	1.93×10^{14}
12	1.06×10^{13}	27	1.82×10^{19}	42	2.80×10^{19}	57	4.75×10^{13}
13	4.75×10^{13}	28	2.80×10^{19}	43	1.82×10^{19}	58	1.06×10^{13}
14	1.93×10^{14}	29	4.05×10^{19}	44	1.12×10^{19}	59	2.16×10^{12}
15	7.21×10^{14}	30	5.53×10^{19}	45	6.46×10^{18}	60	3.97×10^{11}

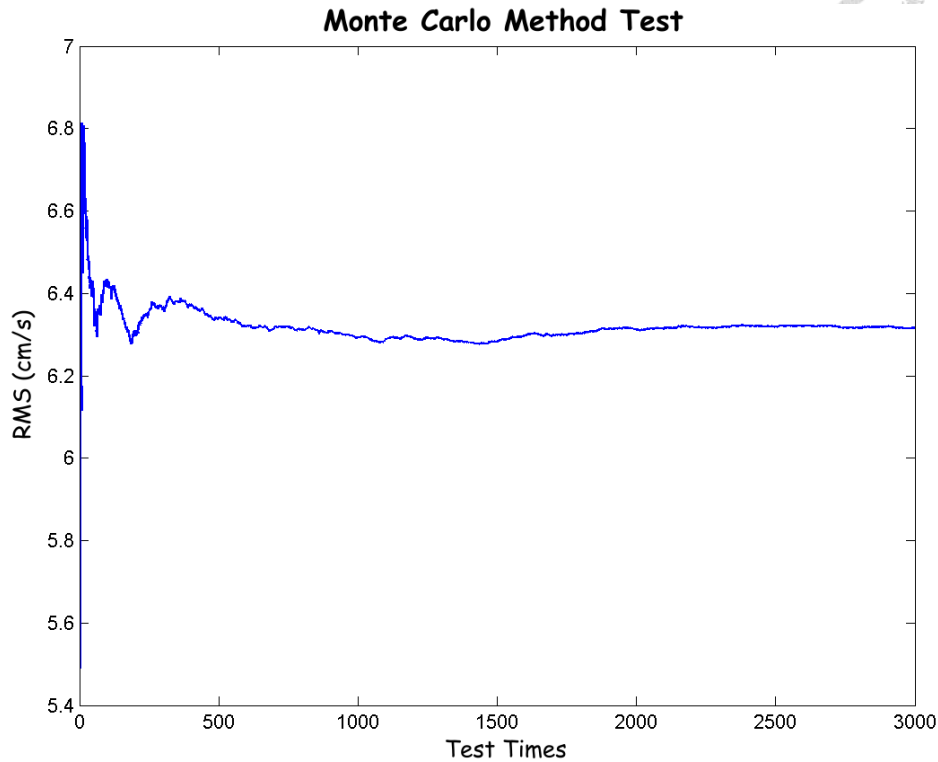


圖 4-1 蒙地卡羅法前級試算結果：假設缺漏個數為 5，以最小平方方法搭配實向量 EOF 法填補，橫軸為實驗次數，縱軸為填補值與真實值之誤差均方根平均值。

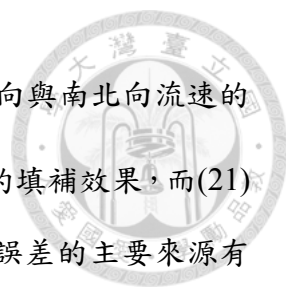
我們使用第三章所述的實向量 EOF 法、KLE 法以及最小平方方法、迭代法共 4 種方法，再按照資料缺漏個數分別利用窮舉法與蒙地卡羅法進行計算，實驗結果則以這一個月樣本資料中填補值與真實值之誤差均方根平均值作為判斷誤差大小的基準，根據(7)式可計算各資料點之平均誤差：

$$RMS(x_i) = \frac{1}{H} \sum_{h=1}^H \sqrt{\frac{\sum_{k=1}^K [(u_F(x_i, t_k) - u_O(x_i, t_k))^2 + (v_F(x_i, t_k) - v_O(x_i, t_k))^2]}{K}} \quad (20)$$

以及其缺漏情況之總體平均誤差：

$$RMS(H) = \frac{1}{M} \sum_{i=1}^M \sqrt{\frac{\sum_{k=1}^K [(u_F(x_i, t_k) - u_O(x_i, t_k))^2 + (v_F(x_i, t_k) - v_O(x_i, t_k))^2]}{K}} \quad (21)$$

式中 H 為不同缺漏個數的變化種類(缺漏數不大於 3 時， H 值參考表 4-1；缺漏個數大於 3 時則 $H = 3000$)， i 為填補資料點之編號， u_F 、 v_F 表示 2012 年 12 月東西



向、南北向流速的填補值， u_o 及 v_o 則分別是 2012 年 12 月東西向與南北向流速的真實值， K 為資料筆數，由(20)式之結果可研判在各觀測資料點的填補效果，而(21)式所算出的值則可以比較不同缺漏個數填補結果之好壞。實驗誤差的主要來源有二：(1)實向量 EOF 法與 KLE 法僅使用前 20 個模組重建流場，由變動流速總變異數(20 個模組之特徵值總和)來看本身即有約 3%的誤差；(2)使用不同方法填補缺漏值所造成之誤差。此外，由於模重組法係使用有限個模組重建流速資料，為了消除各資料點變異數本身在空間上就是不均勻之影響，我們亦將 2012 年 12 月無缺值之流速資料進行了同樣的模分析，並根據(5)式與(11)式利用前 20 個模組重建資料，再與填補值比較並另行統計分析。後二節將分別說明以窮舉法與蒙地卡羅法進行之補遺統計，最後比較並討論各種方法之填補結果。



4.2 窮舉法實驗：任意 1~3 個資料點缺漏

如前節所述，在資料缺漏個數為小於等於 3 時我們使用窮舉法計算觀測區內所有各種資料缺漏情形之填補結果，分別使用了實向量 EOF 法與 KLE 法，以及搭配最小平方方法與迭代法進行填補，然後再計算這一個月樣本資料中填補值與真實值之誤差均方根，以作為相互比較之依據。

當缺漏 1 點時，由表 4-1 可知這是最簡單的情形，在總資料點為 70 之狀況下總共只有 70 種變化，每次實驗即假設任意一個資料點為缺漏情形，共有 70 種填補實驗，每次實驗均是使用(20)式計算 (u_F, v_F) 與 (u_O, v_O) 之間的誤差均方根，並將其結果繪出以觀察各資料點之誤差在空間上的分佈情形。表 4-2 為最小平方方法與迭代法配合實向量 EOF 法、KLE 法進行填補缺漏實驗之誤差均方根計算結果，第一行為資料點位編號(相關位置參閱圖 2-5)，第二行(RMS_{LE})與第三行(RMS_{LK})分別為使用最小平方方法配合實向量 EOF 法與 KLE 法得出的誤差均方根，第四行(RMS_{IE})跟第五行(RMS_{IK})則分別是以迭代法搭配實向量 EOF 法與 KLE 法的計算結果，這四種誤差均方根在空間分佈上如圖 4-2 所示。圖 4-2 顯示以最小平方方法填補後之誤差均方根約介於 3.56~14.16 cm/s 之間。以迭代法填補後的誤差均方根則是在 5.05~22.39 cm/s 之間。同時，在距離雷達站較近處誤差均方根較小，在觀測區外圍則有少數資料點之誤差均方根會大於 12 cm/s。若以最小平方方法與迭代法之結果互相比較，顯然前者之誤差均方根較後者為小。另比較實向量 EOF 法與 KLE 法，可見最小平方方法搭配實向量 EOF 法會比 KLE 法誤差為小，但迭代法則以搭配 KLE 法誤差較小。

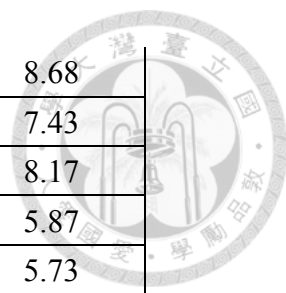
此外，為了消除因使用 20 個模組重建資料而造成各資料點可解釋變異數在空間上之不均勻性，我們將 2012 年 12 月(共 734 組)的 CODAR 觀測資料進行模分析再以 20 個模組得出重建流速 (u_R, v_R) ，計算其與填補值 (u_F, v_F) 之誤差均方根再另行統計(表 4-3)，如此可以濾除 4.1 節所述之第 1 種誤差來源，而所餘者僅為不同

填補方法所造成的誤差，此項結果之空間分佈如圖 4-3 所示。由圖 4-3 可明顯看出這四種方法之誤差均方根要比圖 4-2 所示者稍小，約在 2.64~20.12 cm/s 之間，同時圖 4-3(a)、(b)中僅有少數資料點的誤差均方根會大於 9 cm/s，而圖 4-3(c)、(d)則呈現觀測區外圍大部份資料點之誤差值均大於 9 cm/s。因此可知不同填補方法所產出之誤差始為主要的誤差來源。

當缺漏 2 點時，在總資料點為 70 的狀況下共有 2415 種填補實驗。當缺漏 3 點(假設任意 3 個資料點為缺漏情形)時則總共有 54740 種填補實驗(表 4-1)。在這兩種缺漏情況下，每次均以最小平方法與迭代法搭配實向量 EOF 法、KLE 法分別進行填補缺漏實驗，然後使用(20)式分別計算 (u_F, v_F) 與 (u_O, v_O) 、 (u_R, v_R) 之間的誤差均方根，這兩種情形之統計結果，其特性分佈非常類似於圖 4-2、圖 4-3(限於篇幅未列出)，同樣是在距離雷達站較近處誤差均方根最小，資料點越遠離雷達站則誤差均方根較大，特別是觀測區外圍東北角與西南角的少數資料點。此外，誤差均方根數值雖然會隨缺漏個數變多而有稍微增加的趨勢，但並未增加太多。若比較這四種方法對缺漏 2 點與 3 點填補實驗的效果，以最小平方法與實向量 EOF 法的組合進行填補後得到之誤差較小，填補效果較好，與缺漏 1 點的實驗結果相同。

表 4-2 填補任意 1 個資料點為缺漏情形之實驗結果。

資料點編號	RMS _{LE} (cm/s)	RMS _{LK} (cm/s)	RMS _{IE} (cm/s)	RMS _{IK} (cm/s)
1	13.06	14.02	22.39	17.60
2	9.39	9.52	9.31	10.19
3	8.57	9.54	10.79	11.45
4	4.80	4.92	6.60	5.56
5	4.18	4.11	6.58	6.53
6	4.31	4.35	6.73	7.11
7	5.82	6.42	7.63	7.51
8	10.57	11.57	19.27	18.11
9	5.72	5.45	7.94	8.21
10	5.39	6.88	8.10	9.91
11	4.86	6.41	8.30	10.38
12	8.51	9.47	10.83	11.36
13	3.56	3.72	5.05	5.16
14	4.28	4.49	5.29	5.79
15	4.04	4.10	6.31	6.12
16	7.22	7.93	10.91	11.33
17	7.19	8.82	10.37	10.65
18	5.24	6.35	6.02	6.22
19	4.77	6.07	5.91	6.92
20	4.71	5.49	5.99	6.27
21	5.30	5.54	8.08	9.11
22	4.42	4.75	6.38	7.51
23	5.67	6.14	7.11	8.49
24	6.03	6.07	7.72	8.12
25	7.81	8.88	13.24	14.57
26	7.13	7.91	7.94	7.68
27	4.59	4.91	6.20	5.94
28	5.11	5.49	7.00	7.64
29	4.76	4.55	7.40	6.53
30	4.62	4.76	7.83	8.29
31	4.56	5.35	7.54	9.53
32	5.60	6.63	6.24	6.40
33	5.54	5.61	6.48	6.56
34	6.86	7.63	12.53	13.16



35	6.15	6.23	8.71	8.68
36	5.37	5.75	6.81	7.43
37	5.57	5.97	6.58	8.17
38	4.65	4.72	6.03	5.87
39	4.93	5.62	5.60	5.73
40	5.34	5.46	5.85	5.93
41	5.60	5.98	5.71	5.24
42	5.98	6.87	8.83	7.76
43	7.83	8.35	16.89	15.51
44	7.31	6.39	9.17	9.69
45	5.08	4.75	5.98	5.87
46	5.52	4.72	6.38	6.95
47	4.70	4.68	6.61	6.18
48	5.25	5.35	6.66	6.65
49	5.14	5.13	6.50	7.15
50	5.39	5.73	6.58	6.55
51	6.94	7.00	11.75	10.44
52	9.36	9.62	10.53	11.36
53	4.97	5.23	6.30	6.05
54	5.67	5.74	6.82	6.38
55	6.76	6.77	8.58	8.09
56	5.47	5.74	6.81	7.27
57	5.48	6.17	7.41	7.95
58	6.04	6.67	7.75	8.10
59	8.96	9.86	12.20	12.15
60	8.75	8.09	10.91	11.61
61	5.92	6.19	7.82	8.24
62	6.45	6.52	8.17	8.43
63	5.35	5.50	7.07	6.86
64	7.04	6.98	8.87	8.66
65	6.85	6.91	9.36	9.40
66	9.10	10.48	15.43	15.03
67	6.91	7.36	9.41	9.25
68	7.41	8.17	9.67	8.90
69	7.69	9.13	12.87	13.05
70	12.48	14.16	17.79	17.52

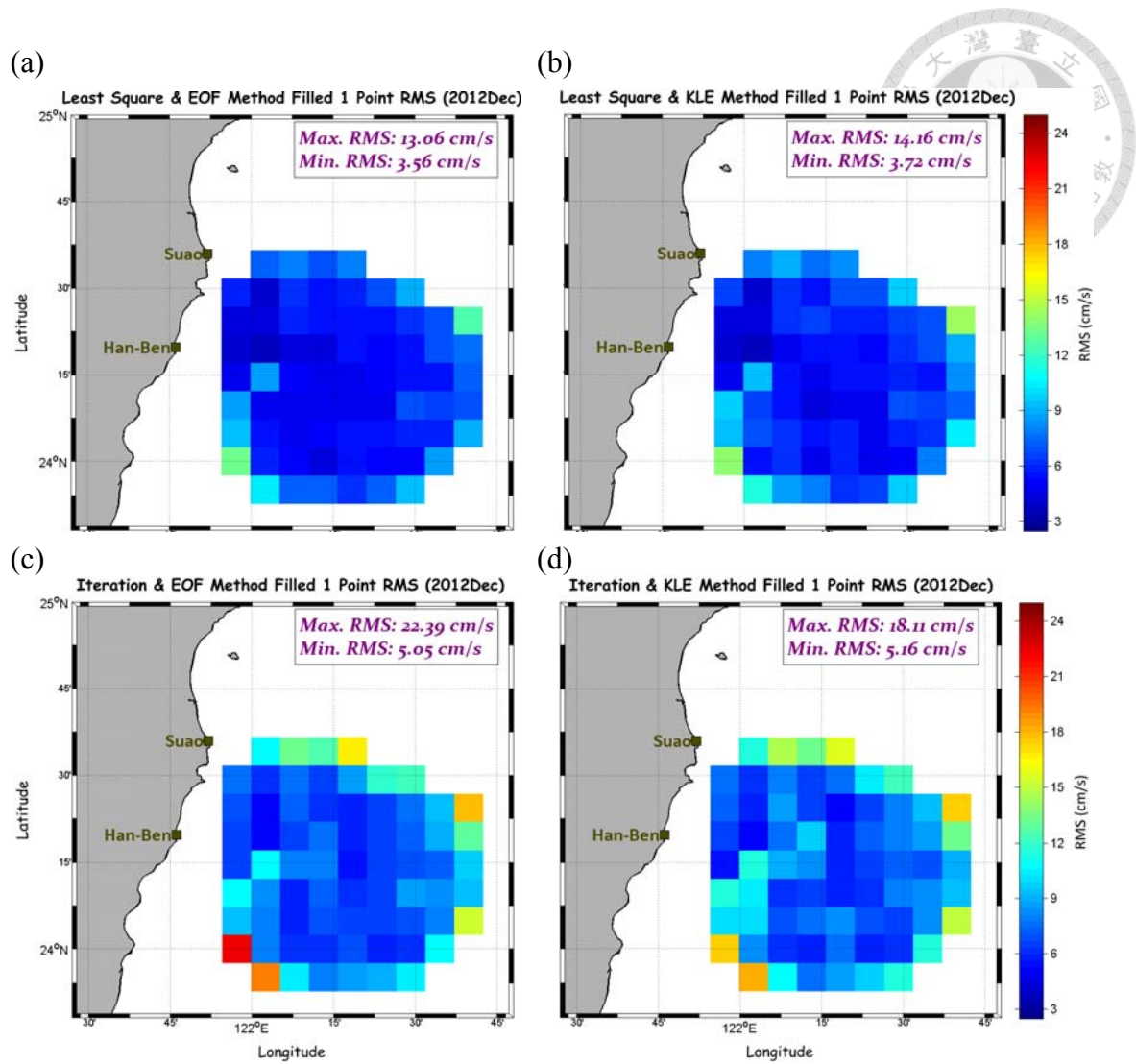
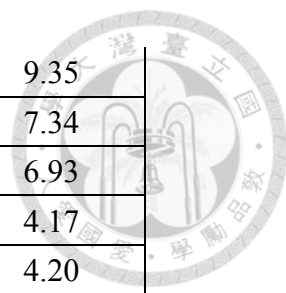


圖 4-2 任意 1 個資料點缺漏情形之填補實驗結果：(a)以最小平方法配合實向量 EOF 法；(b)以最小平方法配合 KLE 法；(c)以迭代法配合實向量 EOF 法；(d)以迭代法搭配 KLE 法填補之誤差均方根空間分佈。

表 4-3 填補任意 1 個資料點為缺漏情形之實驗結果(扣除空間上之不均勻性)。

資料點編號	RMS _{LE} (cm/s)	RMS _{LK} (cm/s)	RMS _{IE} (cm/s)	RMS _{IK} (cm/s)
1	11.83	10.77	20.12	14.55
2	8.18	8.76	8.24	8.98
3	7.53	7.20	9.19	9.60
4	3.60	4.81	8.34	6.59
5	3.62	4.15	7.13	6.89
6	3.33	4.51	5.68	5.75
7	4.13	6.21	7.89	7.41
8	8.89	8.60	17.62	15.91
9	3.88	3.76	7.56	7.63
10	4.17	5.70	7.26	8.99
11	3.78	5.08	7.24	9.23
12	7.79	7.79	9.90	10.17
13	2.83	3.28	5.38	5.44
14	2.98	4.02	4.80	4.82
15	2.64	3.50	5.69	5.50
16	4.90	5.82	11.02	11.14
17	5.54	6.80	9.65	9.36
18	3.42	4.85	5.09	5.05
19	3.17	4.93	4.20	5.34
20	3.04	4.44	4.62	5.06
21	3.91	5.45	8.23	9.39
22	2.88	3.75	6.07	7.23
23	4.16	4.95	7.02	7.79
24	3.95	4.31	6.87	7.28
25	6.08	6.13	10.49	11.39
26	4.89	6.19	8.30	7.57
27	2.98	3.39	5.27	4.91
28	3.49	4.34	5.38	6.22
29	3.73	3.96	5.91	4.95
30	3.43	4.63	6.93	7.55
31	3.16	4.74	7.15	8.92
32	4.49	4.79	6.78	5.73
33	4.05	3.76	6.49	6.34
34	4.60	5.27	10.64	10.98



35	4.36	5.84	9.63	9.35
36	3.85	5.49	6.87	7.34
37	4.12	4.90	5.40	6.93
38	3.53	4.23	4.70	4.17
39	3.80	5.28	3.76	4.20
40	3.69	3.98	5.31	5.51
41	3.92	4.28	5.17	4.22
42	4.38	4.73	8.90	7.31
43	5.40	6.06	15.69	13.94
44	5.41	6.15	9.92	10.48
45	3.04	3.35	5.02	5.13
46	3.57	3.43	5.58	6.43
47	3.68	4.21	5.47	4.82
48	3.91	4.62	4.73	4.73
49	3.28	3.39	5.38	6.23
50	3.62	4.63	6.60	6.22
51	4.65	4.70	10.87	9.12
52	6.88	9.33	11.16	12.31
53	3.47	4.06	6.21	5.93
54	3.68	4.09	5.75	5.10
55	5.64	5.57	6.83	6.24
56	3.68	3.33	4.02	4.50
57	3.83	4.49	5.97	6.59
58	4.45	5.09	7.66	7.43
59	5.94	5.66	10.35	8.95
60	5.38	4.90	8.93	9.44
61	4.21	4.30	6.42	6.64
62	4.43	4.96	6.35	6.24
63	3.44	3.37	5.06	4.95
64	5.15	4.54	7.13	6.83
65	4.16	4.25	7.96	7.73
66	6.18	6.81	12.57	11.85
67	4.54	3.64	8.79	8.18
68	4.95	5.51	9.85	8.26
69	4.86	5.52	9.49	9.20
70	10.22	11.04	13.21	13.02

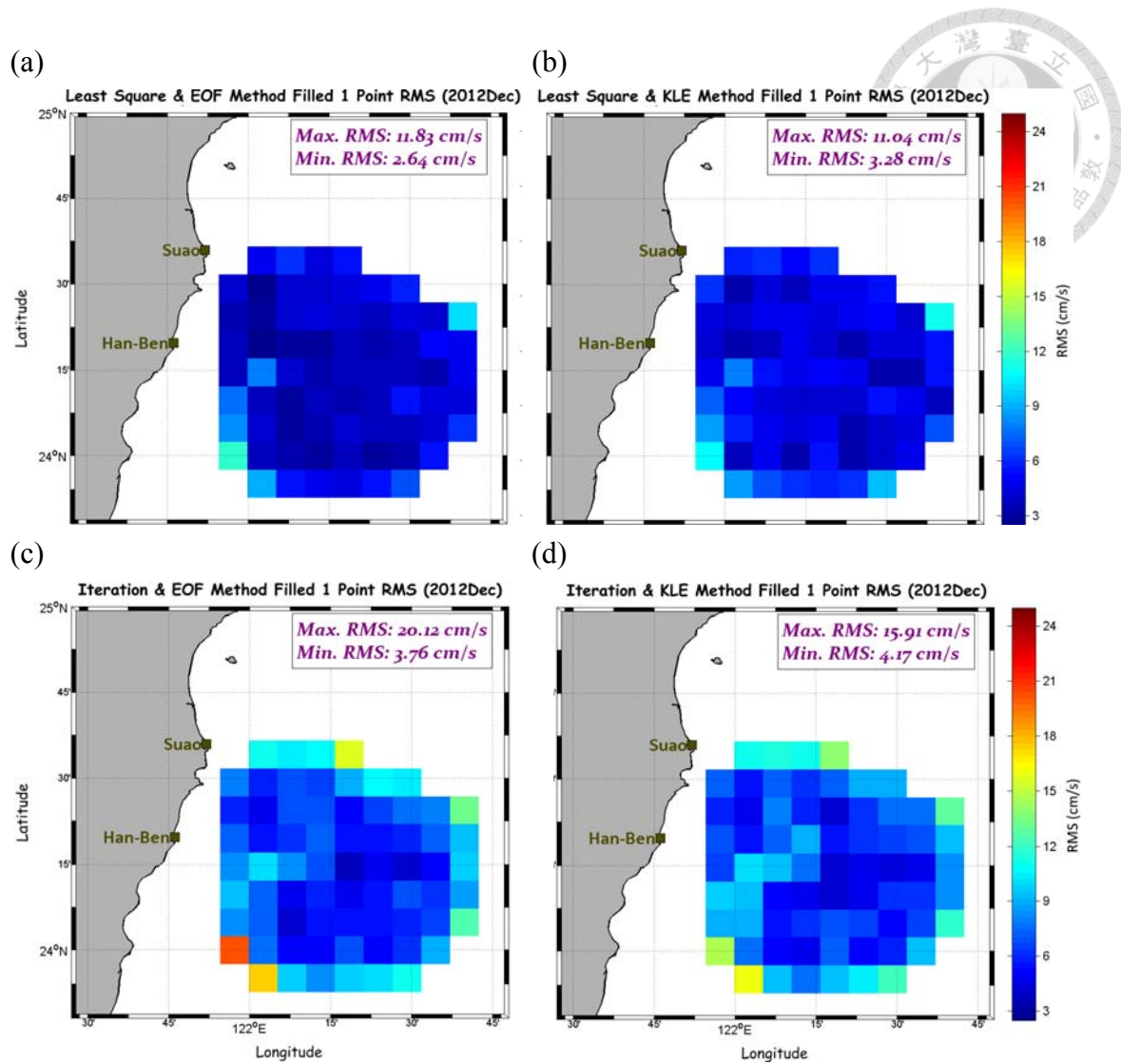


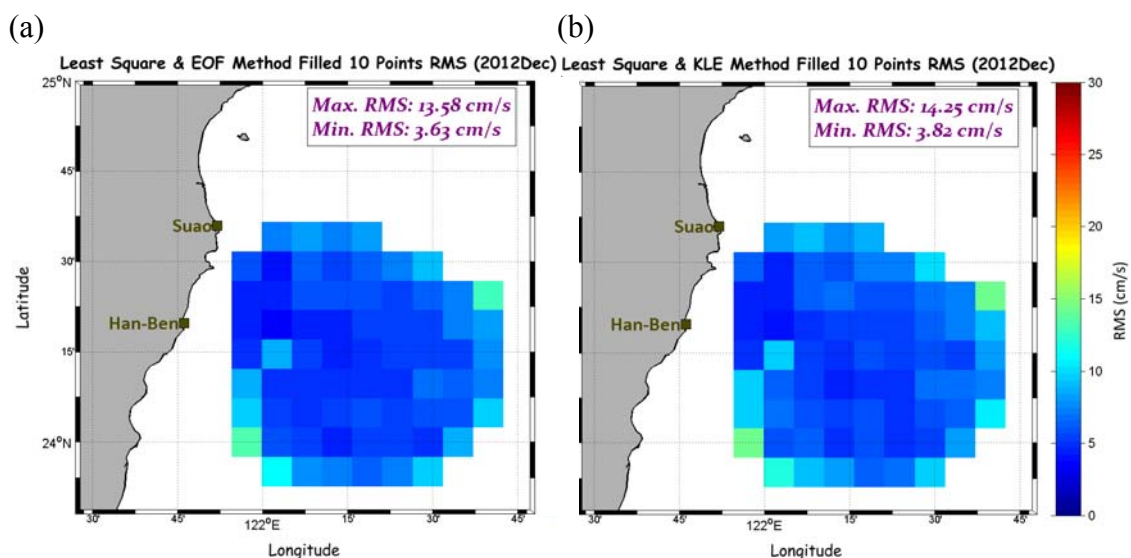
圖 4-3 任意 1 個資料點缺漏情形之填補實驗結果(已扣除 4.1 節所述第 1 種誤差)：
 (a)以最小平方方法配合實向量 EOF 法；(b)以最小平方方法配合 KLE 法；(c)以迭代法
 配合實向量 EOF 法；(d)以迭代法搭配 KLE 法填補之誤差均方根空間分佈。



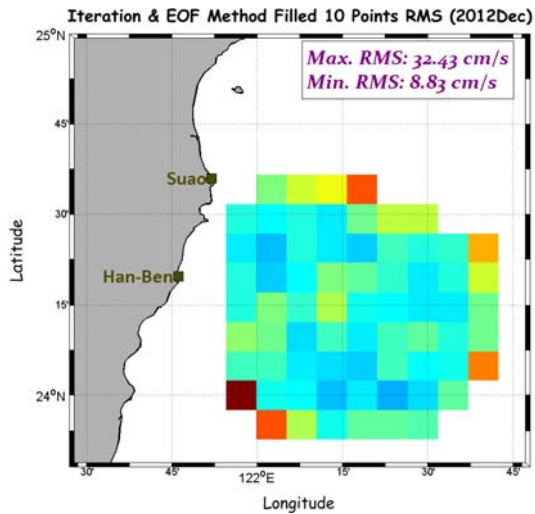
4.3 蒙地卡羅法之實驗：多個資料點出現缺漏

本節討論以蒙地卡羅法隨機設計缺漏資料點進行之填補實驗，在各種不同缺漏個數之情形下各情況均分別執行 3000 次(3000 種缺漏變化)模擬實驗，限於篇幅僅以缺漏個數為 10、35 的情況為例予以說明。

當缺漏 10 點時，以最小平方方法與迭代法分別配合實向量 EOF 法、KLE 法進行填補缺漏實驗之誤差均方根的計算結果其空間分佈如圖 4-4 所示；從圖中可發現若以最小平方方法填補則誤差均方根會小於 15 cm/s，而以迭代法填補後之誤差均方根則介於 8.78~32.43 cm/s 之間，明顯較前者大，同時後者亦顯示在離雷達站較近處誤差均方根較小，在觀測區外圍少數資料點的誤差均方根會大很多。比較最小平方方法與迭代法之結果，雖然前者之誤差均方根較小，但當再比較實向量 EOF 法與 KLE 法的填補結果時，則分別以前者與實向量 EOF 法、以及與 KLE 法這兩種組合的填補效果較好，四種組合中又以實向量 EOF 法與最小平方方法所得之誤差最小。



(c)



(d)

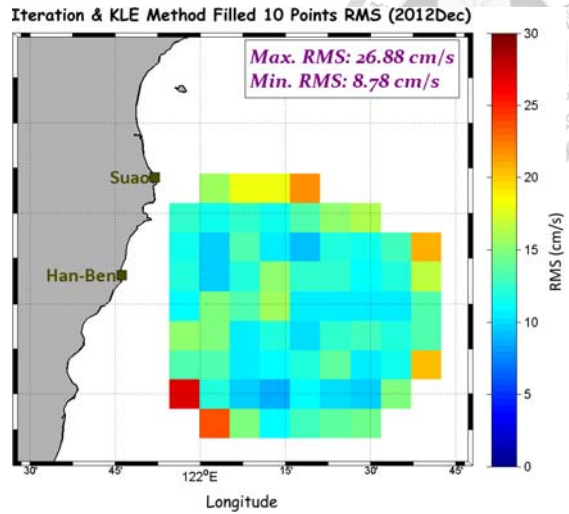


圖 4-4 任意 10 個資料點缺漏情形之填補實驗結果：(a)以最小平方配合實向量 EOF 法；(b)以最小平方配合 KLE 法；(c)以迭代法配合實向量 EOF 法；(d)以迭代法搭配 KLE 法填補之誤差均方根空間分佈。

為了消除各資料點因使用 20 個模組重建資料而造成變異數在空間上有不均勻性，將 2012 年 12 月(共 734 組)CODAR 觀測資料進行模分析，然後再計算填補值與 20 個模組重建資料二者間之誤差均方根，此結果即代表純由各填補方法所造成的誤差，圖 4-5 為此結果在空間上的水平分佈；圖中的誤差均方根數值與圖 4-4 相比稍小，以最小平方填補時，其誤差介於 2.78~12.37 cm/s，另以迭代法填補後，誤差範圍則約在 7.83~31.89 cm/s，不過大部分資料點的誤差皆小於 13 cm/s。由圖 4-4 與圖 4-5 可知，使用不同填補方法所造成之誤差始為實驗的主要誤差來源。

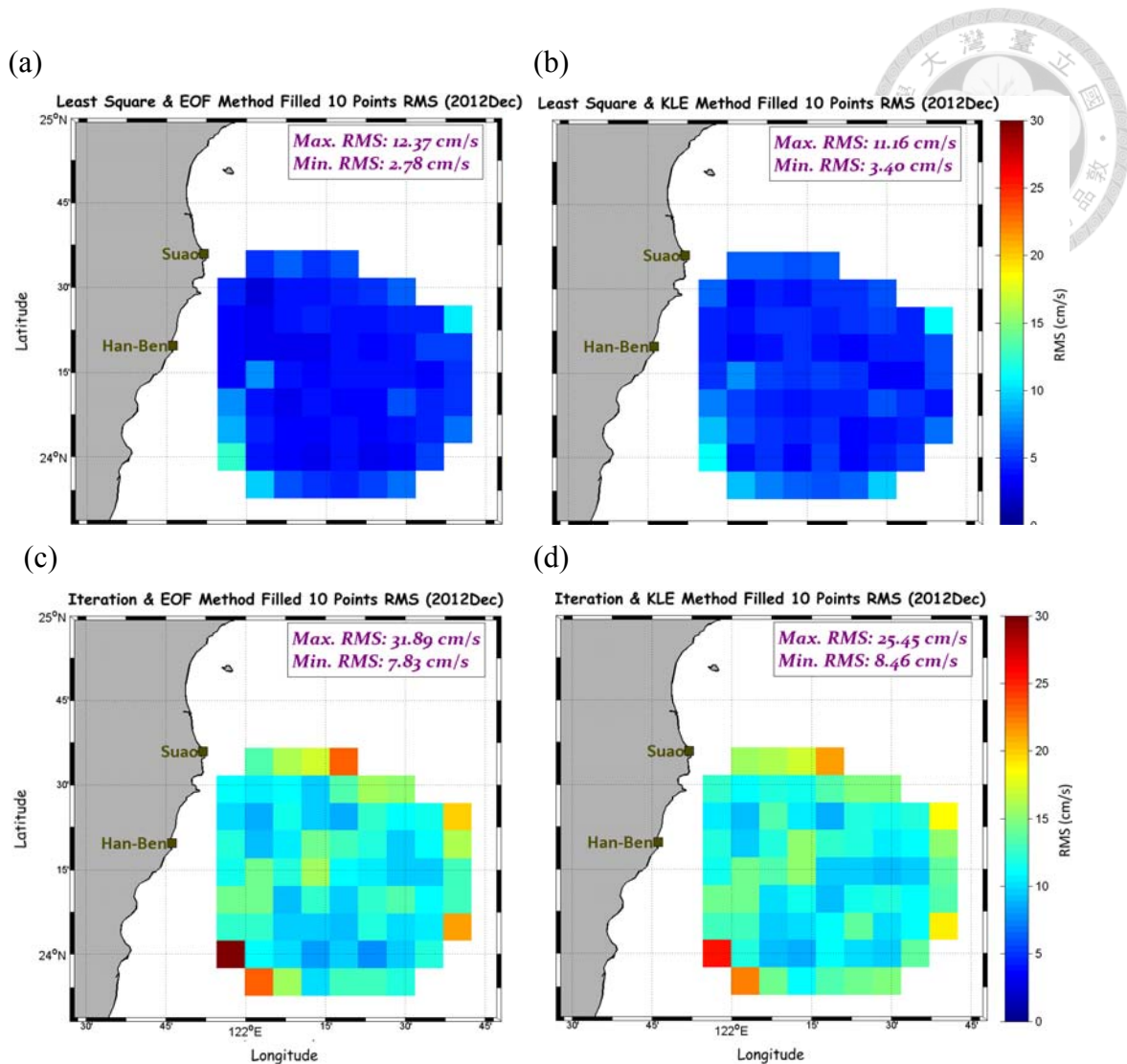


圖 4-5 任意 10 個資料點缺漏情形之填補實驗結果(已扣除 4.1 節所述第 1 種誤差)：(a)以最小平方方法配合實向量 EOF 法；(b)以最小平方方法配合 KLE 法；(c)以迭代法配合實向量 EOF 法；(d)以迭代法搭配 KLE 法填補之誤差均方根空間分佈。

當缺漏 35 點時，同樣計算四種方法之誤差均方根，其在空間上的水平分佈如圖 4-6，此時最小平方方法的誤差在 4.15~16.69 cm/s 之間，其範圍與圖 4-4(a)、(b)並沒有相差太多，同樣離雷達站越遠的資料點誤差均方根越大，尤其是觀測區東北與西南的少數資料點通常為誤差最大之處；不過使用迭代法填補缺漏資料點與缺漏 10 點的情況在空間上有較明顯的不同，在觀測區中南部的資料點比靠近雷達站的資料點誤差稍小，表示若是人為製造的缺漏點是在靠近雷達站的地方，則重建資料並填補後所計算得到之誤差均方根較大。此外，圖 4-7 為消除空間上變異數



不均勻性的填補結果，其誤差較小但空間分佈與圖 4-6 類似。整體而言，同樣是最小平方法搭配實向量 EOF 法的誤差均方根較小。

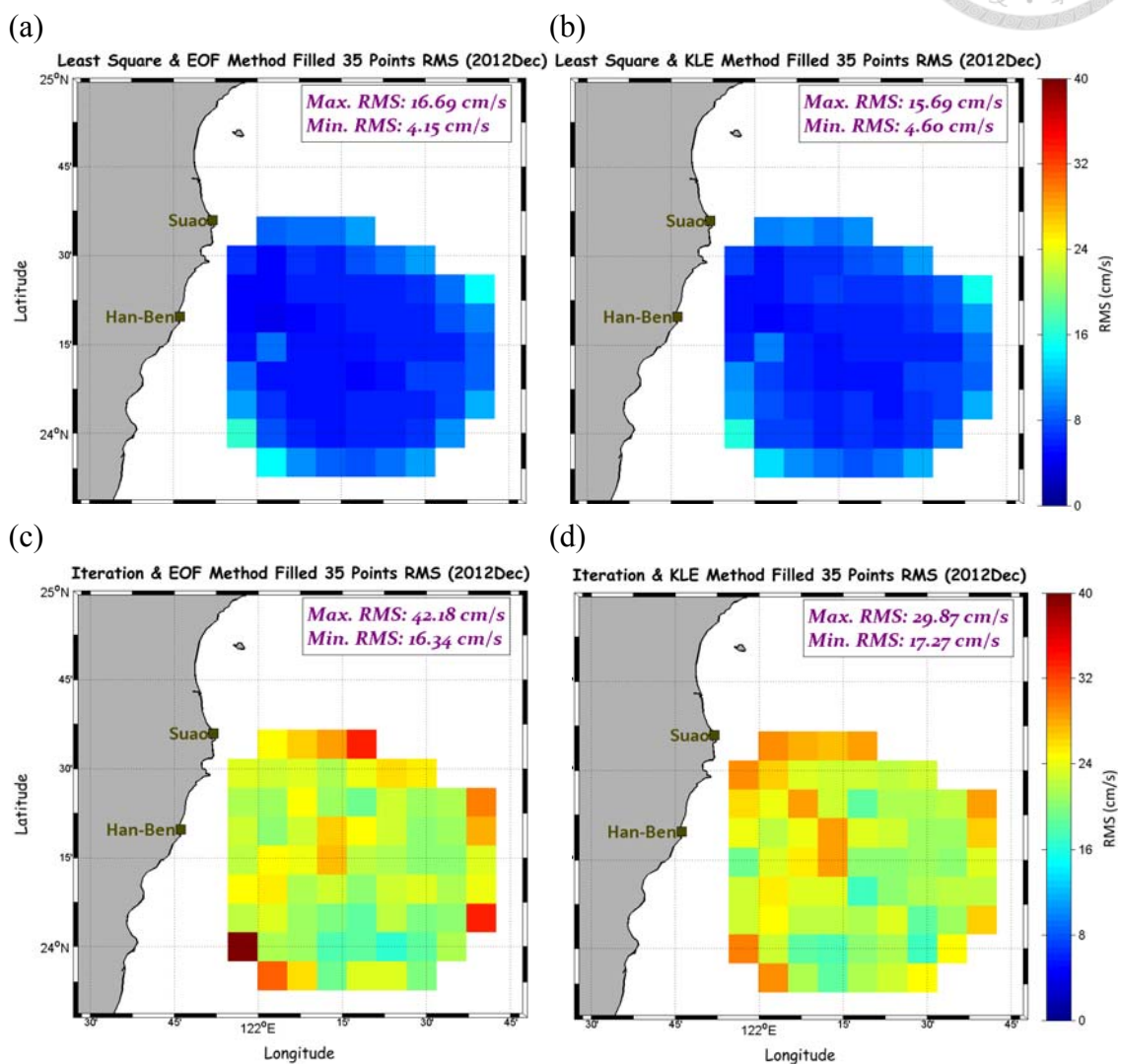


圖 4-6 任意 35 個資料點缺漏情形之填補實驗結果：(a)以最小平方方法配合實向量 EOF 法；(b)以最小平方方法配合 KLE 法；(c)以迭代法配合實向量 EOF 法；(d)以迭代法搭配 KLE 法填補之誤差均方根空間分佈。

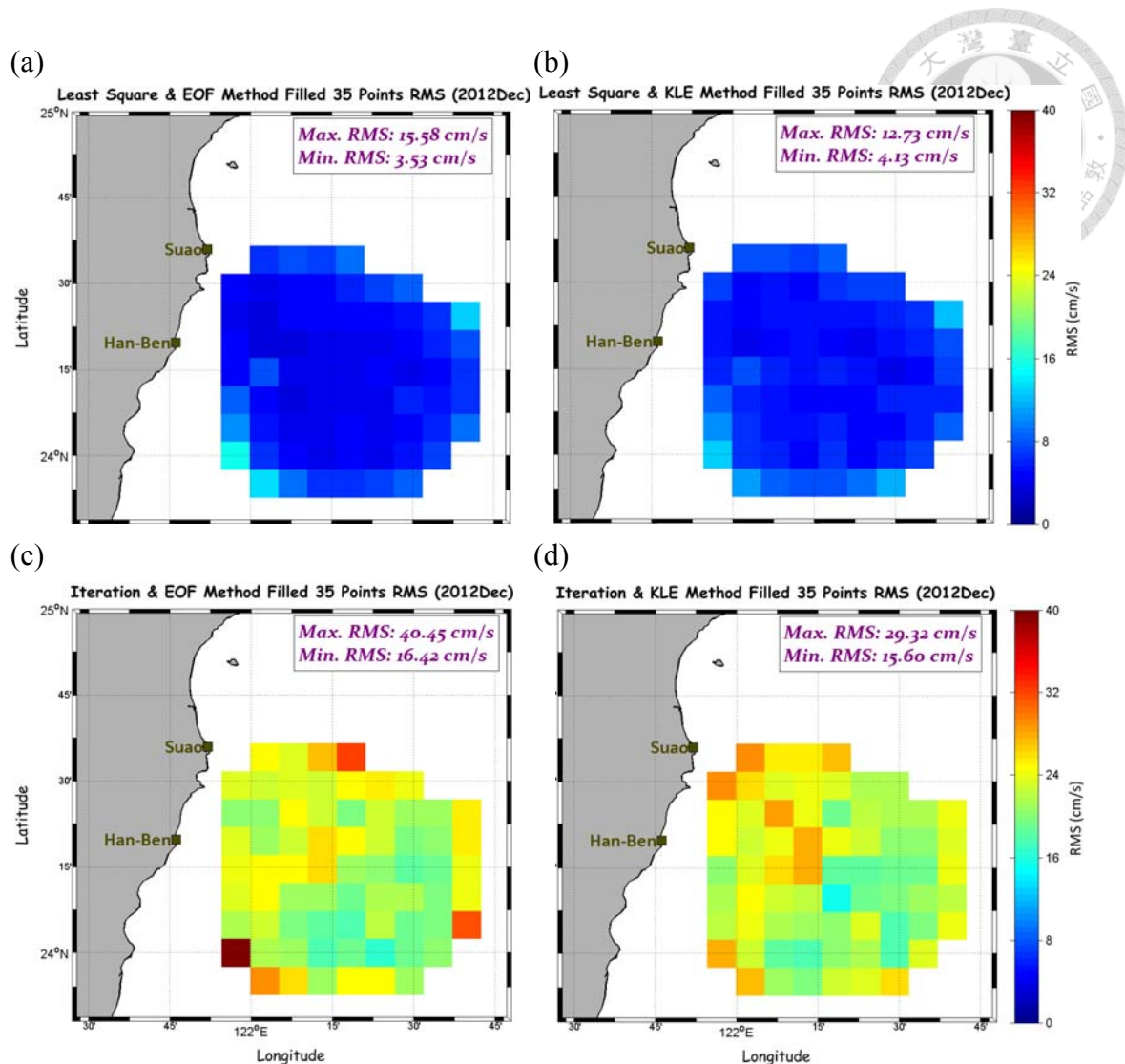


圖 4-7 任意 35 個資料點缺漏情形之填補實驗結果(已扣除 4.1 節所述第 1 種誤差)：
(a)以最小平方方法配合實向量 EOF 法；(b)以最小平方方法配合 KLE 法；(c)以迭代法
配合實向量 EOF 法；(d)以迭代法搭配 KLE 法填補之誤差均方根空間分佈。

而當缺漏個數越多，誤差均方根數值會逐漸增加，若以最小平方方法填補則每種缺漏情形前後之誤差的增幅不大，並在空間分佈上仍以靠近雷達站的資料點誤差較小；不過以迭代法填補時，其不同缺漏個數情形之誤差的差異較大，在空間上雖然是以靠近雷達站處誤差較大，但這是因為如果我們人為製造的缺漏點是在靠近雷達站的資料點，則所得到的誤差較大，填補效果較不好。若比較這四種方法的填補效果，則是以最小平方方法與實向量 EOF 法的搭配組合填補效果較好。



4.4. 綜合討論

若將不同缺漏個數填補實驗之誤差均方根互相比較，其結果如圖 4-8 所示，圖中標示了四種方法在各種缺漏個數情形之誤差均方根平均值，其信賴區間 (confidence interval) 範圍參照表 4-4，其估算公式(Bendat and Piersol, 1971)為：

$$\left[\bar{x} - \frac{st_{n;\alpha/2}}{\sqrt{H}} \leq \mu_x < \bar{x} + \frac{st_{n;\alpha/2}}{\sqrt{H}} \right] \quad n = H - 1 \quad (22)$$

式中 H 是樣本數(即實驗次數)， \bar{x} 、 s 分別是在各種缺漏個數情形下進行填補實驗後，由 H 次樣本中所得之誤差均方根的平均值與標準差， $t_{n;\alpha/2}$ 則是 t 分配 (t-distribution) 表中當自由度為 n 、顯著水準為 α 時的數值。

表 4-4 顯著水準 $\alpha = 0.05$ 時與各自由度相對應之信賴區間。

樣本數(H)	自由度($n = H - 1$)	信賴區間
70	69	$\bar{x} - 0.2384s \leq \mu_x < \bar{x} + 0.2384s$
2415	2414	$\bar{x} - 0.0399s \leq \mu_x < \bar{x} + 0.0399s$
54740	54739	$\bar{x} - 0.0084s \leq \mu_x < \bar{x} + 0.0084s$
3000	2999	$\bar{x} - 0.0358s \leq \mu_x < \bar{x} + 0.0358s$

綜合來說，圖 4-8(b)即填補值與 20 個模組重建值之誤差均方根平均值分佈，其數值皆比圖 4-8(a)填補值與真實值之間的誤差還小，不過分佈情況類似；由圖 4-8 可以發現，在缺漏個數介於 1~40 間時，以最小平方法填補的結果之誤差均方根分佈近似水平，表示填補誤差隨缺點數增加變化不大，但當缺漏個數大於 40 之後，誤差即開始逐漸變大，當缺漏個數大於 50 時，誤差會呈指數成長；而以迭代法進行填補實驗所得之結果，缺漏 1 點時之誤差就已經比最小平方法的誤差還大，並隨著缺漏個數增多而呈線性增長，圖 4-8(a)、(b)中迭代法誤差均方根平均值之斜

率分別大約為 0.37、0.38，且相鄰的任意兩種缺漏情形之誤差平均值皆大約相差 0.5 cm/s。

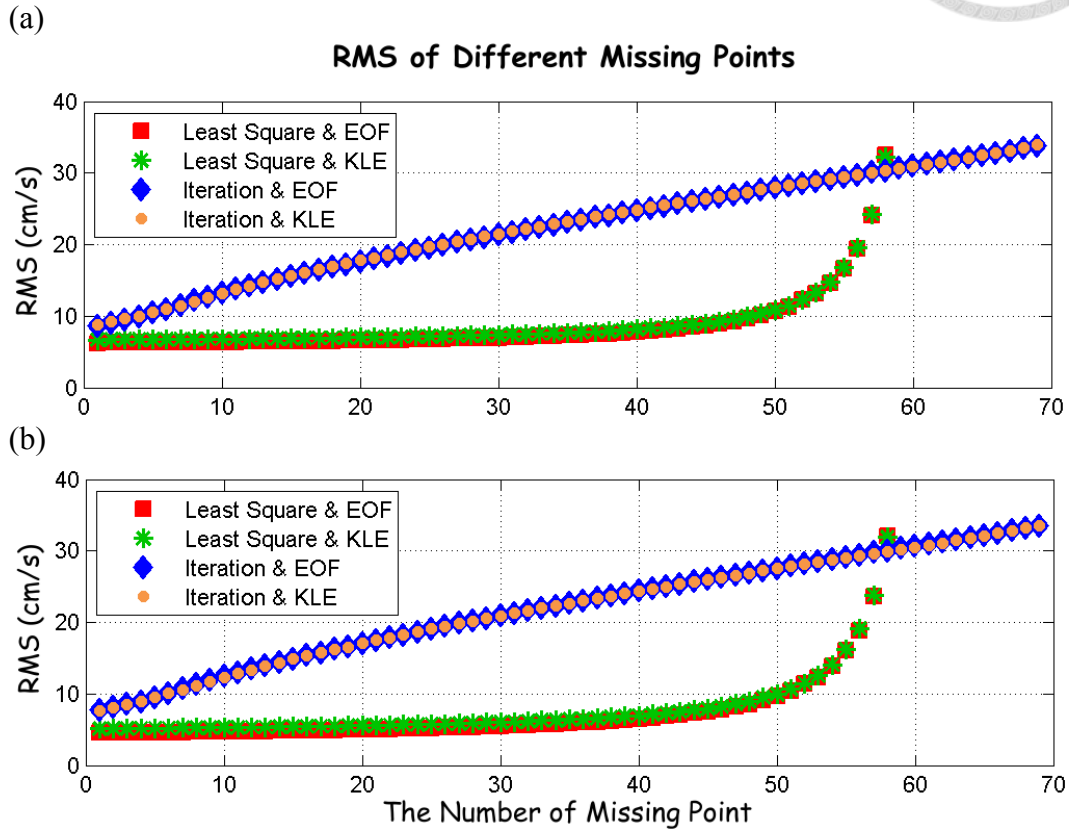


圖 4-8 以實向量 EOF 法、KLE 法與最小平方法、迭代法進行填補不同缺漏個數實驗結果：(a)填補值與真實值之誤差均方根平均值；(b)填補值與 20 個模組重建流場之誤差均方根平均值分佈。

理論上 H 次樣本得到之誤差均方根會呈現常態分佈，圖 4-8 為 H 次樣本的平均值，而 H 次樣本中與平均值的偏差程度可用標準差來觀察，與圖 4-8 相對應之各種缺漏情形的誤差均方根標準差如圖 4-9。圖 4-9(a)是填補值與真實值之誤差均方根標準差，當以最小平方法填補缺漏資料點時，可以發現標準差從 2 cm/s 左右開始逐漸減小，在缺漏個數為 20~30 之間有最小值，約 0.5 cm/s，之後又繼續增大，直到缺漏個數大於 50 時即呈現指數成長；另以迭代法填補缺值時，當缺漏個數為 1 的狀況下其標準差約 3 cm/s，隨著缺漏個數漸增，標準差會依斜率約 -0.04 的趨

勢逐漸變小。此外，圖 4-9(b)是由填補值與 20 個模組重建資料所算出之結果，亦有類似的分佈情形。如前述，標準差即表示誤差均方根的變動幅度，由此可知，當缺漏個數大約在 5~35 之間時，以最小平方法進行填補實驗所得到之誤差均方根分佈較集中，變動幅度不大，而當缺漏個數越大時則其誤差均方根分佈會越分散；若使用迭代法補遺，其誤差均方根分佈隨缺漏個數變多而越集中，即在 H 次實驗中，因為大部分資料點補遺後已有相當大的誤差，因此所得到的誤差數值都相當接近其平均值並在一定範圍內，因此使用此方法算出之標準差會隨缺漏點數增多而變小。

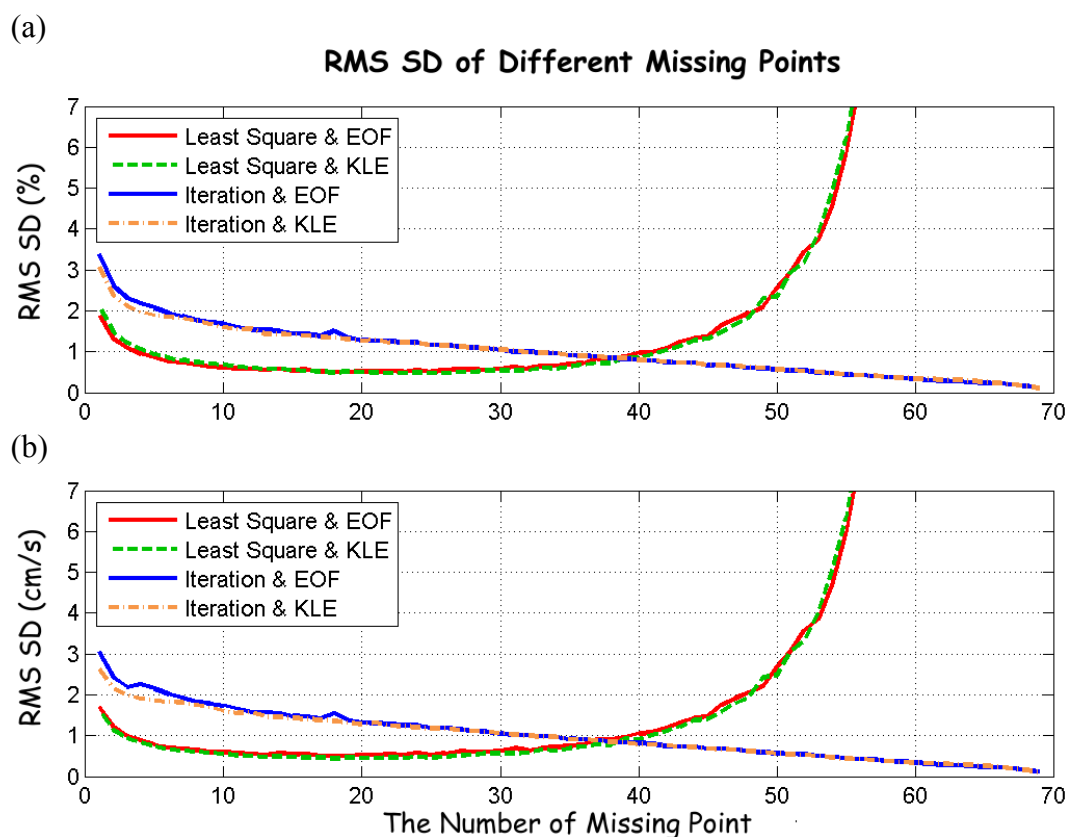


圖 4-9 以實向量 EOF 法、KLE 法與最小平方法、迭代法進行填補不同缺漏個數實驗結果：(a)填補值與真實值之誤差均方根標準差；(b)填補值與 20 個模組重建流場之誤差均方根標準差分佈。

接下來比較實向量 EOF 法與 KLE 法的填補結果，將圖 4-8 中二法之誤差均方根平均值互相比較(圖 4-10a)，可以發現使用最小平方方法填補缺漏時，實向量 EOF 法的誤差均方根平均值幾乎都小於 KLE 法的結果，最多其差值可達到 -8 cm/s 左右；但當使用迭代法時，則有可能是 KLE 法之結果較小，但其差值大約在 ± 0.3 cm/s 之間。

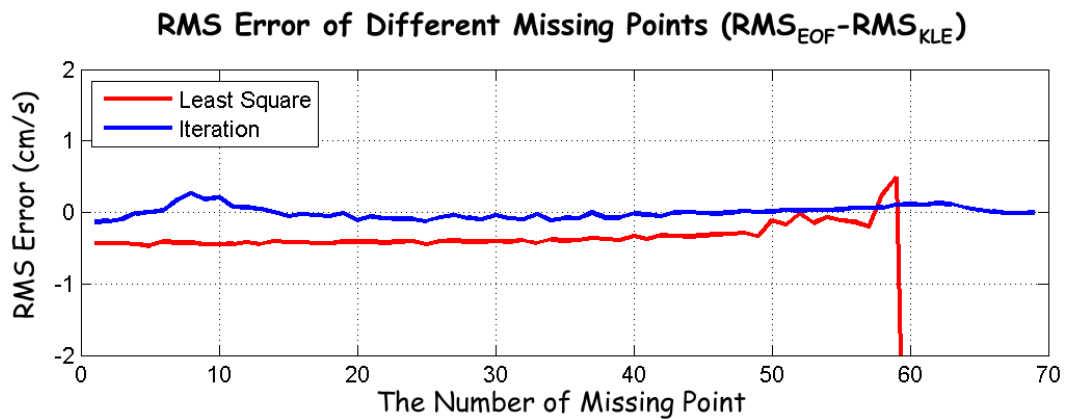


圖 4-10 實向量 EOF 法與 KLE 法誤差均方根平均值之比較。

由以上分析可知，若以填補值與真實值計算誤差均方根時，比較最小平方方法與迭代法的結果以前者之誤差較小，另比較實向量 EOF 法與 KLE 法，當搭配最小平方方法時是以實向量 EOF 法為佳，若搭配迭代法則有可能是以 KLE 法之誤差較小；再比較填補值與 20 個模組重建流場二者間之誤差均方根，則同樣是使用最小平方方法與實向量 EOF 法二者搭配時的填補效果較佳。此結論是僅以目測觀察實驗結果所得，但實向量 EOF 法與 KLE 法誤差均方根之間的差異相當小，較不易判斷上述結論是否正確，因此我們將使用 F 檢定(F-test)對實向量 EOF 法與 KLE 法的實驗結果進行檢測，以確認其正確性。

F 檢定法可以透過檢測實向量 EOF 法與 KLE 法實驗結果之變異數，推測在設定之顯著水準為 α 的情況時二者是否相同(Bendat and Piersol, 1971)；首先定義虛無假設 H_0 為 $s_{EOF}^2 = s_{KLE}^2$ ，而其對立假設 H_1 則為 $s_{EOF}^2 \neq s_{KLE}^2$ ，其中 s_{EOF}^2 是實向量 EOF



法在各種缺漏個數情形時誤差均方根平均值的變異數，而 s_{KLE}^2 則是 KLE 法的誤差均方根平均值變異數；接下來需計算二變異數之比值 F ：

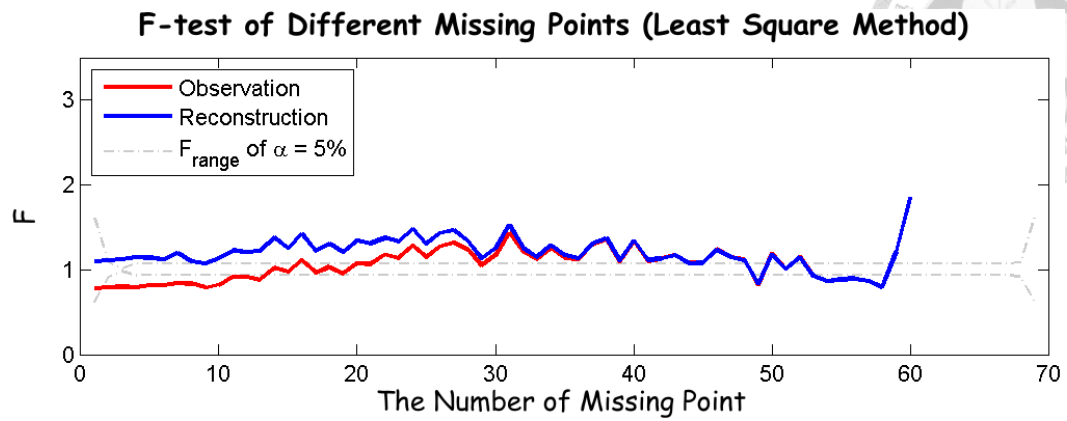
$$F = \frac{s_{EOF}^2}{s_{KLE}^2} \quad (23)$$

再由 F 分配(F-distribution)表中與顯著水準 α 及自由度 $n(n = H - 1, H$ 為樣本數)相對應的上、下限數值(表 4-5)，即可得知檢測結果。在顯著水準 $\alpha = 5\%$ 時其結果如圖 4-11，圖中紅線為填補值與真實值之誤差均方根結果，藍線為填補值與 20 個模組重建流場誤差均方根間的比較，灰色虛線則是 $\alpha = 5\%$ 的 F 值範圍，若由二者變異數計算得之 F 值在此範圍內，則 H_0 成立；由圖 4-11 可知在 $\alpha = 5\%$ 時，使用最小平方法時，在每種缺漏個數情形下實向量 EOF 法得出之誤差變異數幾乎均小於 KLE 法，而迭代法的檢測結果則在大部分情況時是 H_0 成立。因此可知在四種填補結果中以最小平方法搭配實向量 EOF 法的填補效果較佳。此外，由圖 4-8 之誤差變化情形亦可推測當缺漏個數大於 50(即缺漏量達 71%)時，本文所述的這些方法均不適合用來補遺。

表 4-5 F 分配表中，顯著水準為 α 及自由度為 n 時的上、下限數值。

樣本數(H)	自由度($n = H - 1$)	F
70	69	$0.62 \leq F < 1.61$
2415	2414	$0.92 \leq F < 1.08$
54740	54739	$0.98 \leq F < 1.02$
3000	2999	$0.93 \leq F < 1.07$

(a)



(b)

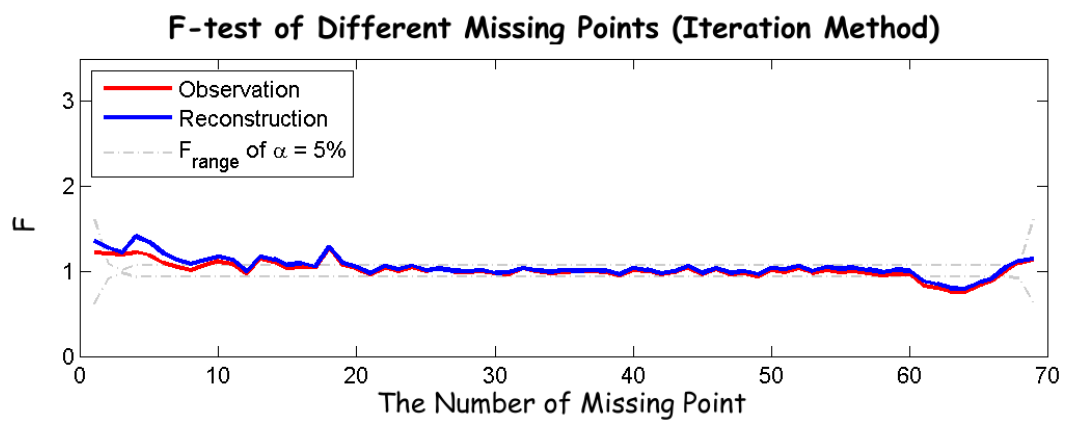


圖 4-11 F 檢測：比較(a)最小平方法；(b)迭代法之結果。


第五章 結論



高頻雷達是近代廣用的即時監測海流遙測工具，但雷達系統會因為儀器故障或訊號干擾等因素影響而不時造成觀測區內海流資料發生缺漏不完整的問題。蘇澳、漢本 CODAR 高頻雷達測站自 2011 年 4 月開始運作以來，迄 2012 年 12 月止在臺灣東北海域觀測並記錄 15048 組表面海流資料(每小時一組)，在這些資料中，各資料點發生缺漏資料之百分比出現率幾乎不超過 5%；整體而言，在此期間內如果扣除完全沒有任何缺漏以及完全沒有觀測資料兩種極端情況，則以觀測區內缺漏一個資料點的情形為最多，共有 126 組，佔所有有缺漏點情況的 29.37%；而缺漏 1~3 個資料點的比例則有 50.58%，即不完整資料的大部份情況缺漏個數都在 3 個以下；觀察每日出現資料缺漏的逐時變化，可以發現在本地時間晚上 6 點到凌晨 4 點間比較容易出現資料有缺漏的情形，此段時間內平均每小時資料缺漏之出現機率大約佔所有缺漏情形的 4~6%。

這些觀測資料之平均流向為北至東北，與我們所認知的黑潮流況相符；統計分析東西向以及南北向之海流流速分量資料，皆呈常態分佈；經由能譜分析，則可發現當地海流主要是受低頻運動的影響，其次則為潮流；由於潮汐運動不屬於隨機過程，因此使用調和分析法進一步將觀測資料中的潮流濾除，並以此濾潮後之資料進行平穩性檢定，結果呈現為具有弱平穩性，顯示這些時序資料之二階統計結果應當可以向外延伸用於觀測時段之外。

我們使用實向量 EOF 法以及 KLE 法分別分析濾潮後完整無缺值的 CODAR 觀測資料(由 2011 年 4 月 14 日至 2012 年 11 月 30 日，共 13280 組)，得出觀測場的特徵值與對應之特徵向量，並以此特徵向量做為基底，再求出流速資料在各基底的投影即可得出該模組之振幅，如此可將各基底分量重新合成並重建流場。經二法處理後，由實向量 EOF 法得到的前 20 個模組可解釋流速資料總變異量的 97.02%，而 KLE 法則可解釋 97.14%。此外，另計算由前 20 個模組重建流場與原始觀測流場間的流速偏



差均方根值，在70個資料點中，實向量EOF法之偏差均方根整體平均值為5.92 cm/s，而偏差均方根之最大與最小值分別為8.40 cm/s、3.68 cm/s，靠近二CODAR雷達站處均方根較小，但觀測區外圍及北部則較大。至於KLE法偏差均方根之整體平均值為6.39 cm/s，最大、最小值為11.38 cm/s與3.79 cm/s，同樣也是離雷達站較近處均方根較小、較遠處均方根較大。整體而言，此結果顯示離雷達站較遠處的資料點受到高階模組的影響較大，也就是說受到小尺度運動的影響較大。

此外，我們選用 2012 年 12 月(共 734 組)完整無缺值之 CODAR 觀測資料作為獨立樣本用以檢測不同方法填補不完整觀測資料之效果。我們分別使用了實向量 EOF 法以及 KLE 法(均使用前 20 個模組)並各自搭配最小平方法以及迭代法等共組成四種填補方法，依照人為製造的不同資料點缺漏個數進行填補實驗，並計算填補結果與原始觀測值之偏差均方根作為誤差用於比較填補效果。計算結果顯示，不管資料缺漏個數多寡與填補方法為何，誤差均方根皆是在靠近雷達站處較小，而在觀測區海域外圍少數的資料點處較大。若將四種資料缺漏填補方法之誤差均方根互相比較(圖 4-6)，可以發現以最小平方法進行填補實驗時，當資料點缺漏數小於 40 時其誤差均方根數值隨缺漏點個數的變化並不大，但當缺漏個數大於 40 後，誤差即開始明顯增大，當缺漏個數達 50 以上時，誤差則呈指數成長；若以迭代法填補時，其誤差均方根數值會隨缺漏點個數的增多而逐漸變大，並呈現線性增長的趨勢，其斜率皆約 0.37~0.38 左右。

綜合而言，當缺漏個數在 40 (57%)以下時，四種方法中以最小平方法搭配實向量 EOF 法進行填補之誤差均方根為最小，填補效果較佳；但在缺漏個數大於 50 (71%)時由於使用最小平方法填補後得到之誤差亦逐漸變大，故在此情況下不適合使用這些方法來填補缺漏值。

參考文獻



Alvera-Azcárate, A., A. Barth, M. Rixen, and J. M. Beckers, 2005 : "Reconstruction of incomplete oceanographic data sets using empirical orthogonal functions: application to the Adriatic Sea surface temperature", *Ocean Modelling*, **9**, 325-346.

Barrick, D. E., M. W. Evans, B. L. Weber, 1977 : "Ocean Surface Currents Mapped by Radar", *Science*, **198**, 138-144.

Beckers, J. M., and M. Rixen, 2003 : "EOF Calculations and Data Filling from Incomplete Oceanographic Datasets*", *Journal of Atmospheric and Oceanic Technology*, **20**, 1839-1856.

Bendat, Julius S., and Allan G. Piersol, 1971 : "Random data: analysis and measurement procedures", *John Wiley & Sons, New York*.

Boyd, Janice D., Eileen P. Kennelly, and Pavel Pistek, 1994 : "Estimation of EOF expansion coefficients from incomplete data", *Deep-Sea Research Part I: Oceanographic Research Papers*, **41.10**, 1479-1488.

Bretherton, Francis P., Russ E. Davis, and C. B. Fandry, 1976 : "A technique for objective analysis and design of oceanographic experiments applied to MODE-73*", *Deep Sea Research and Oceanographic Abstracts*, **23**, 559-582.

Bui-Thanh, T., M. Damodaran, and K. Willcox, 2004 : "Aerodynamic Data Reconstruction and Inverse Design Using Proper Orthogonal Decomposition", *AIAA Journal*, **42.8**, 1505-1516.

Dong, D., P. Fang, Y. Bock, F. Webb, L. Prawirodirdjo, S. Kedar, and P. Jamason,



2006 : "Spatiotemporal filtering using principal component analysis and Karhunen-Loeve expansion approaches for regional GPS network analysis", *Journal of Geophysical Research*, **111**, B03405, doi: 10.1029/2005JB003806.

Defant, Albert, 1961 : "Physical oceanography; volume 2", *Pergamon Press*.

Emery, Brian M., Libe Washburn, and Jack A. Harlan, 2004 : "Evaluating Radial Current Measurements from CODAR High-Frequency Radars with Moored Current Meters*", *Journal of Atmospheric and Oceanic Technology*, **21**, 1259-1271.

Emery, Brian M., Libe Washburn, Milton S. Love, Mary M. Nishimoto, and J. Carter Ohlmann, 2006 : "Do oil and gas platforms off California reduce recruitment of bocaccio (*Sebastes paucispinis*) to natural habitat? An analysis based on trajectories derived from high-frequency radar", *Fishery Bulletin*, **104.3**, 391-400.

Fang, Y.-C., J. Wang, Y.-J. Yang, J.-C. Mau, and M.-J. Huang, 2011 : "Preliminary Results of CODAR Surface Current Observations Northeast of Taiwan", *2011 International Workshop on Operating System for Marine Environment Monitoring and Forecasting, National Kaohsiung Marine University, College of Ocean Engineering, Kaohsiung*, 59-77.

Fukunaga, Keinosuke, 1990 : "Introduction to Statistical Pattern Recognition, Second Edition", *Academic Press*.

Gurgel, K.-W., G. Antonischki, H.-H. Essen, and T. Schlick, 1999 : "Wellen Radar (WERA): a new ground-wave HF radar for ocean remote sensing", *Coastal Engineering*, **37**, 219-234.

Han, Jiawei, and Micheline Kamber, 2006 : "Data Mining: Concepts and Techniques,



Second Edition", *Morgan Kaufmann Publishers*, 61-62.

Hardy, Donald M., 1977 : "Empirical eigenvector analysis of vector observations", *Geophysical Research Letters*, **4.8**, 319-320.

Holmes, Philip, John L. Lumley, and Gal Berkooz, 1996 : "Turbulence, Coherent Structures, Dynamical Systems and Symmetry", *Cambridge University Press*.

Jan, Sen., Joe Wang, Ching-Sheng Chern, and Shenn-Yu Chao, 2002 : "Seasonal variation of the circulation in the Taiwan Strait", *Journal of Marine Systems*, **35**, 249-268.

Jan, Sen, Ching-Sheng Chern, Joe Wang, and Ming-Da Chiou, 2012 : "Generation and propagation of baroclinic tides modified by the Kuroshio in the Luzon Strait", *Journal of Geophysical Research*, **117**, C02019, doi: 10.1029/2011JC007229.

Kaihatu, James M., Robert A. Handler, George O. Marmorino, and Lynn K. Shay, 1998 : "Empirical Orthogonal Function Analysis of Ocean Surface Currents Using Complex and Real-Vector Methods*", *Journal of Atmospheric and Oceanic Technology*, **15**, 927-941.

Kaplan, Alexey, Yochanan Kushnir, Mark A. Cane, and M. Benno Blumenthal, 1997 : "Reduced space optimal analysis for historical data sets: 136 years of Atlantic sea surface temperatures", *Journal of Geophysical Research*, **102.C13**, 27835-27860.

Kerschen, Gaetan, Jean-Claude Golinval, Alexander F. Vakakis, and Lawrence A. Bergman, 2005 : "The Method of Proper Orthogonal Decomposition for Dynamical Characterization and Order Reduction of Mechanical Systems: An Overview", *Nonlinear Dynamics*, **41**, 147-169.

Kohut, Josh, Scott Glenn, and Don Barrick, 1999 : "SeaSonde Is Integral to Coastal Flow Model Development", *Hydro International*, **3.3**, 32-35.

Kohut, Josh T., and Scott M. Glenn, 2003 : "Improving HF Radar Surface Current Measurements with Measured Antenna Beam Patterns", *Journal of Atmospheric and Oceanic Technology*, **20**, 1303-1316.

Kondrashov, D., and M. Ghil, 2006 : "Spatio-temporal filling of missing points in geophysical data sets", *Nonlinear Processes in Geophysics*, **13**, 151-159.

Kundu, Pijush K., J. S. Allen, and Robert L. Smith, 1975 : "Model Decomposition of the Velocity Field near the Oregon Coast", *Journal of Physical Oceanography*, **5**, 683-704.

Kundu, Pijush K., and J. S. Allen, 1976 : "Some Three-Dimensional Characteristics of Low-Frequency Current Fluctuations near the Oregon Coast", *Journal of Physical Oceanography*, **6**, 181-199.

Kutzbach, John E., 1967 : "Empirical Eigenvectors of Sea-Level Pressure, Surface Temperature and Precipitation Complexes over North America", *Journal of Applied Meteorology*, **6**, 791-802.

Metropolis, Nicholas, and S. Ulam, 1949 : "The Monte Carlo Method", *Journal of the American Statistical Association*, **44.247**, 335-341.

Oliveira, P., and L. Gomes, 2010 : "Interpolation of signals with missing data using Principal Component Analysis", *Multidimensional Systems and Signal Processing*, **21**, 25-43, doi: 10.1007/s11045-009-0086-3.

Paduan, Jeffrey D., and Leslie K. Rosenfeld, 1996 : "Remotely sensed surface currents in Monterey Bay from shore-based HF radar (Coastal Ocean Dynamics Application

Radar)", *Journal of Geophysical Research*, **101.C9**, 20669-20686.

Pawlowicz, Rich, Bob Beardsley, and Steve Lentz, 2002 : "Classical tidal harmonic analysis including error estimates in MATLAB using T_TIDE", *Computers & Geosciences*, **28**, 929-937.

Preisendorfer, Rudolph W., 1998 : "Principal Component Analysis in Meteorology and Oceanography", *Elsevier*.

Ramp, Steven R., Donald E. Barrick, Tomotaka Ito, and Michael S. Cook, 2008 : "Variability of the Kuroshio Current south of Sagami Bay as observed using long-range coastal HF radars", *Journal of Geophysical Research*, **113**, C06024, doi: 10.1029/2007JC004132.

Reynolds, Richard W., and Thomas M. Smith, 1994 : "Improved Global Sea Surface Temperature Analyses Using Optimum Interpolation", *Journal of Climate*, **7**, 929-948.

Smith, Thomas M., Richard W. Reynolds, Robert E. Livezey, and Diane C. Stokes, 1996 : "Reconstruction of Historical Sea Surface Temperatures Using Empirical Orthogonal Functions", *Journal of Climate*, **9**, 1403-1420.

Snihir, Iryna, William Rey, Evgeny Verbitskiy, Afifa Belfadhel-Ayeb, Peter H. L. Notten, 2006 : "Battery open-circuit voltage estimation by a method of statistical analysis", *Journal of Power Sources*, **159**, 1484-1487.

Teague, Calvin C., John. F. Vesecky, and Daniel M. Fernandez, 1997 : "HF Radar Instruments, Past to Present", *Oceanography*, **10.2**, 40-44.

Tiampo, K. F., J. B. Rundle, W. Klein, Y. Ben-Zion, and S. McGinnis, 2004 : "Using Eigenpattern Analysis to Constrain Seasonal Signals in Southern California", *Pure and*



applied geophysics, **161**, 1991-2003.

劉文俊，1999：”臺灣的潮汐”，自費出版。

