

分类号 TP399

学号 08061149

UDC

密级 公 开

工学硕士学位论文

主成分方法在红外高光谱资料
同化中的应用

硕士生姓名 余 意

学 科 专 业 计算机科学与技术

研 究 方 向 大规模科学与工程计算

指 导 教 师 张卫民 研究员

国防科学技术大学研究生院

二〇一一年十一月

主成分方法在红外高光谱资料同化中的应用

国防科学技术大学研究生院

**Application of Principal Component
Algorithm to the Assimilation of High-resolution
Infrared Sounder Observations**

Candidate: Yu Yi

Advisor: Prof. Zhang Weimin

A thesis

**Submitted in partial fulfillment of the requirements
for the degree of Master of Engineering
in Computer science and Technology
Graduate School of National University of Defense Technology
Changsha, Hunan, P.R.China**

November, 2011

独 创 性 声 明

本人声明所呈交的学位论文是我本人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表和撰写过的研究成果，也不包含为获得国防科学技术大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示谢意。

学位论文题目： 主成分方法在红外高光谱资料同化中的应用

学位论文作者签名： 余意 日期： 2011 年 11 月 10 日

学位论文授权使用授权书

本人完全了解国防科学技术大学有关保留、使用学位论文的规定。本人授权国防科学技术大学可以保留并向国家有关部门或机构送交论文的复印件和电子文档，允许论文被查阅和借阅；可以将学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文。

(保密学位论文在解密后适用本授权书。)

学位论文题目： 主成分方法在红外高光谱资料同化中的应用

学位论文作者签名： 余意 日期： 2011 年 11 月 10 日

作者指导教师签名： 姚 斌 日期： 2011 年 11 月 10 日

目 录

摘 要.....	i
第一章 绪 论.....	1
1.1 卫星资料同化背景	1
1.2 卫星红外光谱垂直探测仪器.....	2
1.2.1 国外先进的红外高光谱探测器	2
1.2.2 我国卫星气象与红外探测仪	3
1.3 红外高光谱资料同化研究现状.....	5
1.3.1 红外高光谱资料的同化现状	5
1.3.2 问题的提出	6
1.4 本文的研究内容和创新.....	8
1.5 论文结构.....	9
第二章 红外高光谱主成分通道选择方法	10
2.1 主成分方法.....	10
2.1.1 主成分方法的原理	10
2.1.2 主成分的求解过程	11
2.1.3 主成分的重要性--主成分的贡献.....	14
2.2 主成分方法研究红外高光谱	15
2.2.1 求解红外高光谱主成分	16
2.2.2 选择光谱资料集	17
2.2.3 主成分 Jacobians 的非局部化.....	18
2.2.4 重构辐射率	21
2.3 其它红外高光谱信息表示方法.....	22
2.3.1 多种通道信息表示方法.....	22
2.3.2 直接通道选择方法.....	23
2.4 本章小结.....	24
第三章 基于主成分的辐射传输 PCRT 模式.....	26
3.1 大气辐射传输特性	26
3.1.1 大气的选择吸收性.....	26
3.1.2 大气辐射传输方程——卫星探测大气的理论基础.....	27
3.2 多种辐射传输模式	29
3.3 具有主成分特色的 PCRT 模式.....	31

3.3.1	PCRT 模式廓线集.....	32
3.3.2	PCRT 模式的 LBL 辐射率资料库	33
3.3.3	PCRT 模式的辐射传输方程	34
3.3.4	PCRT 模式的插值.....	35
3.3.5	主成分数目的确立.....	35
3.3.6	PCRT 模式的统计回归算法	37
3.3.7	预报因子	37
3.3.8	预报因子相关阈值.....	38
3.3.9	PCRT 模式预报 PC 谱线	39
3.4	PCRT 模式计算实例	40
3.5	本章小结.....	41
第四章	PCRT 模式在资料同化系统中的实现	42
4.1	四维变分资料同化简介.....	42
4.1.1	目标函数	42
4.1.2	观测算子	43
4.1.2	4D-Var 中的辐射传输模式.....	44
4.2	WRF 四维变分资料同化系统	45
4.2.1	WRF-DA 系统.....	45
4.2.2	设置 WRF 4D-Var 同化时间窗.....	45
4.2.3	WRF 4D-Var 的输入文件	46
4.2.4	同化红外高光谱的参数设置	46
4.3	PCRT 模式与 WRF 4D-Var 的对接与实现.....	48
4.3.1	PCRT 模式与 WRF 4D-Var 的对接.....	48
4.3.2	PCRT 模式的实现.....	48
4.4	本章小结.....	51
第五章	主成分方法同化红外高光谱试验	52
5.1	基于主成分的同化策略.....	52
5.2	误差分析.....	53
5.3	同化结果分析	54
5.4	PCRT 模式计算效率	56
5.5	本章小结.....	57
第六章	结论与展望	58
6.1	总结	58

6.2 未来工作展望.....	59
致 谢.....	60
参考文献	61
作者在学期间取得的学术成果.....	65

表 目 录

表 1.1 国外先进红外高光谱大气探测仪的光谱参数.....	3
表 1.2 ECMWF 2007 年 9 月业务同化 AIRS 和 IASI 的参数.....	6
表 2.1 AIRS 光谱通道的主要特征	15
表 3.1 目前 LBL 计算中包含的氟氯化碳和卤代烷浓度	33
表 3.2 选择预报因子相关阈值的最优值	39
表 4.1 WRF 4D-Var 的同化时间窗变量设置	46
表 4.2 WFR 4D-Var 输入文件	46
表 4.3 RT 指定卫星仪器信息的 namelist 变量.....	47
表 4.4 WRF 4D-Var 中部分 AIRS 变量设置	48
表 4.5 PCRT 模式预报因子参数设置.....	48
表 4.6 通道廓线数组.....	49
表 5.1 PCRT 模式相比 RTTOV 模式的计算增益	56
表 5.2 PCRT 模式与 RTTOV 模式的试验结果	57

图 目 录

图 1.1	对比使用 AIRS 前后的预报评分	5
图 1.2	对比 HIRS 的 19 个通道（左图）和 IASI 的 8461 个通道（右图）	6
图 2.1	按特征值递减顺序排序的 AIRS 主成分谱线.....	17
图 2.2	简化的四维变分同化公式.....	18
图 2.3	IASI 光谱的前 4 个主成分	19
图 2.4	IASI 前 4 个主成分的温度 Jacobians	20
图 2.5	IASI 前 4 个主成分的湿度 Jacobians	20
图 2.6	IASI 采样的原始光谱（上图）和重构的光谱（下图）	22
图 2.7	ECMWF 业务中同化的 AIRS 和 IASI 资料的最大通道数目	24
图 3.1	大气不同成分对辐射率的选择性吸收	26
图 3.2	薄层气体 ρ' 吸收辐射示意图	27
图 3.3	PCRT 模式的内部插值	35
图 3.4	对比不同数目主成分的重构误差（实线）和仪器噪声（点线）	36
图 3.5	IASI (a) 和 AIRS (b) 的拟合 rms 误差随阈值与预报因子分布情况.....	38
图 3.6	PCRT 模式逐层计算辐射率.....	40
图 4.1	4D-Var 中大气状态信息在控制变量与观测之间的传递	44
图 4.2	WRF 预报模式系统中的 WRF-Var	45
图 4.3	WRF 4D-Var 同化时间窗.....	45
图 4.4	WRF 4D-Var 调用 PCRT 模式流程图	48
图 4.5	PCRT 模式函数调用树	50
图 4.6	计算主成分算法 pcrt_PCcores	51
图 5.1	实际观测的仪器噪声和 200 个 IASI 主成分重构的辐射率噪声	53
图 5.2	AIRS 真实辐率和重构辐射率误差协方差矩阵.....	54
图 5.3	4D-Var 代价函数的 J_0 和 J_b 迭代图.....	55
图 5.4	使用 200 个（上图）与 400 个（下图）主成分理论效果.....	55

摘 要

近年来,随着先进的红外高光谱仪器的发展,红外高光谱资料已经成为数值天气预报(NWP)中心同化系统的重要部分。红外高光谱仪器不同于常规的遥感仪器,能够探测几千甚至上万光谱通道,提供了高精度和高垂直分辨率的大气状态信息。从这些大量通道中有效提取大气信息技术难度非常大,目前世界各NWP中心只使用近200个通道用于资料同化,其它通道因为仪器噪声等一系列原因被放弃使用。然而,红外高光谱的大量通道具有非常相似的光谱信号特征,在内容上相互之间高度相关,可以采用有效的方法将完整光谱信息压缩后用于同化。直接同化这些压缩信息能充分利用完整光谱有效信息且加快资料同化速度。

本文采用主成分方法有效提取红外高光谱通道信息,将红外高光谱资料的上千光谱通道压缩成几百个主成分谱线。通过求解光谱协方差矩阵获得光谱主成分,并且按特征值降序排列。最开始的几百个主成分包含了光谱中主要的大气状态信息而不是随机噪声,所以仅使用这几百个主成分就能表示上千红外高光谱通道中的有效信息,体现了主成分方法降低光谱噪声的潜力。

在数值预报资料同化系统中直接同化红外高光谱的主成分,要求使用有效的辐射传输(RT)模式从给定温度、水汽、臭氧、地表特征等变量的第一猜值场直接模拟主成分谱线。为此设计基于主成分方法的辐射传输(PCRT)模式,在已有的快速辐射传输模式中增加计算光谱主成分的模块,经过相关的回归训练测试,直接模拟主成分谱线或者将主成分谱线重构为类似“通道”特征的辐射率。相比模拟全光谱的快速辐射传输模式,PCRT模式不仅能够为任意输入的大气廓线模拟主成分谱线或者重构辐射率,而且PCRT模式具有主成分方法使问题空间维数下降的优点,能够更快速地模拟辐射率,获得更高的计算效率。

为了最终实现直接同化红外高光谱主成分,本文通过修改WRF四维变分资料同化(4D-Var)系统中有关辐射传输模式的接口,设置用户调用PCRT模式的方式,实现在4D-Var系统框架中嵌入PCRT模式,形成了具有同化红外高光谱资料主成分功能的4D-Var系统。通过多组4D-Var试验,验证了主成分方法应用于红外高光谱资料同化的效果。虽然,相比目前业务上对红外高光谱通道进行选择然后用于同化的方法,预报评分有所下降,但是相比完全不使用红外高光谱资料的同化试验,还是能产生明显积极的正效应。在保证同化效果不退化的情况下,通过设置主成分数目,得出PCRT模式相比欧洲中期天气预报中心(ECMWF)开发的RTTOV快速辐射传输模式能够获得5至21倍的计算增益。

关键词: 红外高光谱 通道 主成分 PCRT 模式 四维变分资料同化

ABSTRACT

In the recent years, the observations from high-resolution infrared sounders have become a major part of the assimilation systems in many numerical weather prediction (NWP) centers with the development of these sounders. Unlike the common sounders, high-resolution sounders can measure several thousands to a few thousands channels and can provide atmospheric information with high accuracy and vertical resolution. It is very hard to analysis the useful information from so many channels and, a large number of channels are improper for data assimilation for many reasons like instrument noise and etc., nearly only 200 channels are employed in the data assimilation system.

The spectral of high-resolution infrared sounders contain many channels that have very similar spectral signatures. Channels that similar in content are highly correlated to each other. The complete high-resolution infrared spectral with thousands of channels can be compressed by some methods for data assimilation. Also the compressed information can involve the whole spectral information in direct data assimilation system efficiently and improve the calculation speed.

Using principal component algorithm described in this paper can involve the channel information from the infrared hyper-spectral data efficiently which is compressed into several hundreds of principal components(PC). The principal components are generally calculated through the covariance of these spectral and they are in descending order according to the eigenvalues. These principal components are associated with real atmospheric variations rather random noise and the bulk of the atmospheric information in the spectrum is present in the first hundreds of PC. So only several hundreds of PC can represent the useful information involved in thousands of channels of the infrared hyper spectral data. This reflects the potential of principal components to reduce noise.

The direct use of PC for data assimilation in an NWP assimilation systems requires an efficient radiative transfer (RT) model that can simulate PC directly given first-guess fields of temperature, water vapour, ozone, surface properties and other variables, or even reconstruct any spectrum by using the PC to a assumed accuracy in a multiple linear regression algorithm. A principal component algorithm based radiative transfer (PCRT) model is designed in this article. The PCRT model is formed by adding some models corresponding to principal component method to the fast radiative transfer model which has been developed after many training and testing, then it can simulate principal component scores and consequently reconstruct radiance. Compared to the fast radiative transfer model simulating the whole spectral the PCRT model can simulate the PC or reconstruct radiance for any atmospheric profile inputted to the system, and it can simulate the radiance more effectively and achieve higher computing efficiency for it have the characters derived from the principal component algorithm which can reduce

the special dimension of the issue.

This article describes the implementation of the PCRT model in the four-dimensional variational data assimilation (4D-Var) system by modifying some interfaces in the 4D-Var system and setting the method to call the PCRT model for users, thus the 4D-Var system is formed with the function to assimilate directly the principal component of the high-resolution infrared spectral. Also we need to set the namelist parameters for high-resolution infrared spectral data and prepare the corresponding input files to assimilate directly high-resolution infrared radiance. Based on the result of many experiences we can draw a conclusion that the impact of the assimilation of PC from high-resolution infrared spectral is statistically neutral to the forecast score but it is significant positive compared the experiences without high-resolution data. By setting the number of the PC we can gain 5 to 21 times of computational efficiency of PCRT model over the fast radiative transfer model for TOVS/ATOVS (RTTOV) developed by the European Centre for Medium-Range Weather Forecasts (ECMWF) on the basis of not reducing the assimilation impact.

Key Words: High-resolution infrared spectral, Channel, Principal component, PCRT model, Four-dimensional variational data assimilation

第一章 绪 论

1.1 卫星资料同化背景

数值天气预报(NWP)问题是一个经典的偏微分方程的初、边值问题,必须先根据观测资料分析出初始场,然后用数值模式逐步积分得出预报,所以不论模式多么先进、计算机能力如何强大,高质量的初始大气状态场,仍是准确的数值天气预报的必要条件^[1]。近 10 年,数值预报模式已日趋完善,改进初值成了进一步提高数值预报水平的关键问题。然而,地球是以海洋为主的水环境,全球 80% 以上地区人类无法长期居住,这给人类获得可靠的大气初始状态带来了严重挑战。目前的无线电探空主要集中在亚洲、美洲和欧洲大陆地区,而占全球 70% 的广阔海洋地区几乎没有探测资料。因此在分析某一时刻的资料时,常规观测所提供的资料是不充分的,这势必对预报结果造成严重影响。

气象卫星在大气层外对地球进行观测,不受地表下垫面类型的限制,其观测资料对全球有较好的覆盖。卫星在轨道上不间断地对地球大气进行扫描,并且在轨道覆盖范围内,卫星观测数据的空间水平分辨率只与天顶有关,因此卫星在陆地、海洋、高原和冰川冻土等区域获得无差别的观测数据,极大地弥补了传统观测资料分布受人类居住条件的限制,真正意义上实时获得了全球覆盖均匀的大气状态三维分布资料。

目前,业务运行的气象静止或极轨卫星的探测器如 TOVS 或 ATOVS 上的 HIRS 仪器,采用的是滤光片分光技术,只有 18-20 个红外通道,通道宽度在几十个波数的数量级宽度^{[2],[3]},光谱分辨率很低。虽然此类仪器在提高天气预报水平中已经发挥了重要作用,但由于其光谱分辨率低,采用基于滤光片分光技术的探测仪器很难进一步提高垂直精度和垂直分辨率以满足数值天气预报进一步改进和发展的需求。

随着光学上关键技术的突破,发展高光谱分辨率红外垂直探测器成为可能,EOS(地球观测系统)三高(高精度、高光谱和高空间分辨率)遥感时代的到来,必然把大气遥感推向一个新的阶段^[4]。这些先进的大气红外探测仪拥有上千通道,谱宽在单波数的量级上,如 AIRS 光谱从 $3.7\mu\text{m}$ - $15.4\mu\text{m}$ (650 cm^{-1} - 2700 cm^{-1}) 有 2378 个通道^[5]。这些光谱提供了大量的大气状态信息,一定程度上具有为提高 NWP 模式提供高精度和高垂直分辨率信息的潜质^[6]。

卫星资料已经大量应用到数值预报上,据估计,目前预报所使用的资料中,卫星资料数量已经达到 85%,被数值预报的同化系统所吸纳的资料包括卫星云导风、降水估计、卫星反演的温度、湿度廓线(包括可降水量)、各通道的卫星辐

射率观测值、卫星微波传感器所观测到的海面风、SST 以及各种下垫面状况^[7]。这些被使用的资料约占现有卫星资料的 15%，也即有 85% 的现有资料尚未使用。这种情况对卫星红外高光谱资料的来说尤为突出，例如光栅式大气红外探测仪 AIRS (Atmospheric Infrared Sounder) 资料，欧洲中尺度数值天气预报中心 (ECMWF) 在业务上只从 2378 个通道最多选取其中的 155 个通道进行同化，其余通道都不用于资料同化。

卫星资料同化的前提是建立变分同化框架，美国在 21 世纪初就建立了基于 3D-Var 的 WRF (Weather Research and Forecasting model) 预报模式，为卫星资料的大量使用，特别是辐射资料的直接同化创造了条件。辐射率资料不是模式变量，无法为数值模式直接使用，必须进行卫星资料的数据同化才能与 NWP 模式的初值相互协调。变分同化方法采用复杂的观测算子即正向辐射传输模式，对给定模式初始参量，沿卫星观测方向，利用参数化的预报因子与透过率系数将分子辐射吸收传输机制用线性关系表达，表达了卫星辐射率（或亮温）与 NWP 模式变量间复杂的非线性关系。

1.2 卫星红外光谱垂直探测仪器

1.2.1 国外先进的红外高光谱探测器

近年来，为了提高探测精度和垂直分辨率以满足数值天气预报进一步改进和发展的需求，启动了一系列高光谱分辨率探测仪器的研制计划，卫星红外遥感技术取得了很大的发展。

2002 年 5 月 4 日在美国加州 Vandenburg 空军基地发射的 EOS Aqua 卫星平台上搭载的光栅式大气红外探测仪 AIRS 在 650cm^{-1} - 2700cm^{-1} 波数范围内构造了 2378 个光谱通道，光谱分辨率达到 $v/\Delta v=1200$ ，辐射测量绝对精度优于 0.2K，真正实现了高光谱高精度红外探测。AIRS 是一个扫描探测器，扫描宽度约 1650km，星下点分辨率约为 13km，垂直分辨率为 1km，由于 AIRS 具有高测量高精度和高光谱分辨率的特点，被用来探测精细的大气温度、湿度廓线、臭氧含量等^[8]。2003 年 10 月欧洲中期数值天气预报中心 (ECMWF) 开始应用 AIRS 辐射资料进行同化。这是在业务数值天气预报系统中第一次对红外高光谱遥感进行资料同化。第一次通过卫星上实现真正意义上的对大气海洋环境的高光谱分辨率大气红外探测，为我们提供了一个可以监测全球温度、水汽、臭氧等分布及变化的全新的机会。

美国下一代地球静止业务环境卫星 Geostationary Operational Environmental Satellite (GEOS) 上也将装有类似的高光谱红外探测仪器。GIFTS(Geostationary

Imaging Fourier Transform Spectrometer)为 NASA 新世纪计划地球观测 EOS 使命中搭载的高光谱、高空间分辨率红外干涉探测仪,运行分为两个阶段:第一阶段在美国上空进行试验;第二阶段为业务卫星,运行在印度洋上空,将实现大气的四维观测,1500 通道^[9]。

表 1.1 国外先进红外高光谱大气探测仪的光谱参数

遥感器 名称	分光 方式	光谱范围 (μm)	光谱分辨率 (cm^{-1})	星下点空间分 辨率(km)	灵敏度 (NET或NEN)	通道数
AIRS	光栅	长波15.4-8.80 中波8.22-6.20 短波4.61-3.74	0.55 1.2 2.0	13.5	0.15-0.35K (280K)	2378
IASI	干涉	长波15.50-8.26 中波8.26-5.0 短波5.0-3.62	0.25	12	0.20-0.35K (280K)	8461
CrIS	干涉	长波15.38-9.13 中波8.26-5.71 短波4.64-3.92	0.625 1.25 2.5	14	0.1-0.5K (250K)	1350
GIFTS	干涉	长波14.60-8.85 短波6.06-4.44	0.3 0.6	4	0.2 mW ($\text{m}^2\text{-Sr-cm}^{-1}$) 0.06mW($\text{m}^2\text{-Sr-cm}^{-1}$)	1500

欧洲 METOP 极轨气象卫星携带有高光谱大气红外干涉探测仪 IASI (Infrared Atmospheric Sounding Interferometer), 于 2006 年 10 月 19 日在 EUMETSAT MetOp-A 卫星中心发射。它采用麦克尔逊干涉技术实现高光谱卫星观测。在 $3.62\mu\text{m}$ - $15.50\mu\text{m}$ 红外光谱范围内,IASI 进行连续观测,而且每个通道的光谱分辨率相等 (0.25cm^{-1}),共 8460 个通道。2007 年 6 月 12 日 ECMWF 首次对 IASI 资料进行了同化试验。其后发射的 MetOp 卫星上 IASI 仪器和 NPOESS 卫星上的跨轨红外探测仪 (Cross-track Infrared Sounder, CrIS), 测量 650cm^{-1} - 2250cm^{-1} 范围内的 1350 个通道,光谱分辨率在 1.125cm^{-1} 和 4.5cm^{-1} 之间,将在可预知的时间范围内延续使用这些观测资料。这类仪器都是傅里叶转换分光仪。

高光谱红外探测仪 HES (HES—Hyperspectral Environmental Suite), 即 ABS 高级基线探测器 (Advanced Baseline Sounder) 将装载在未来的美国下一代静止业务气象卫星 GOES-R 及后续卫星上,按照目前的计划 GOES-R 将于 2012 年发射^[6]。表 1.1 列出了国外先进的红外高光谱探测仪的光谱参数。

1.2.2 我国卫星气象与红外探测仪

1961 年 1 月下旬,受强冷空气袭击,中国长江、黄河流域出现了严重的冰凌,致使华东、中南地区大范围有线通讯中断。此次气象事件后,周恩来总理指出:“要自力更生搞基本建设,要赶快改变落后面貌,搞气象卫星。”自此以

后,我国开始发展自己的气象卫星,截止到目前为止,已经发射了数十颗气象卫星,分别实现了极轨气象卫星和静止气象卫星的业务化运行,是继美国、俄罗斯之后第三个同时拥有极轨气象卫星和静止气象卫星的国家。我国的红外高光谱仪器目前还处于研发阶段,星载红外探测仪器还不属于高光谱仪器。下面简单介绍我国新一代风云气象卫星的研究进展、有效载荷以及探测仪器性能。

风云三号(简称FY-3)是在风云一号卫星基础上发展起来的我国新一代极轨气象卫星,首次携带11台探测仪器,具有光谱范围广(紫外、可见、红外和微波,FY-1仅可见光和红外)、探测通道多(共99个光谱通道,FY-1只有5/10个通道)和空间分辨率高(5个通道分辨率达到250m)等特点。在功能上和技术上向前跨越了一大步,具有获取全球、全天候、多谱段、三维、定量探测和地表、海洋及空间环境参数的能力,并与美国的新一代气象与环境监测卫星—NPOESS(National Polar Operational Environment Satellite System)、欧洲的新一代极轨气象卫星—METOP(Meteorological Operational Satellite)相当,世界气象组织已将FY-3卫星纳入新一代世界极轨气象卫星网,可提供高密度覆盖的全球观测。FY-3卫星装载的3台红外探测仪器:

(1) 可见光红外扫描辐射计--VIRR(Visible and Infrared Radiometer)

(2) 红外分光仪--IRAS(InfaRed Atmospheric Spectrometer):26通道的大气红外探测器,选择在 $15\mu\text{m}$ 大气吸收带和 $6.7\mu\text{m}$ 水汽吸收带的通道。采用滤光片的工作方式,扫描观测不同高度的地球大气结构,可以测量晴空条件下大气的温度结构和水汽的三维分布,与微波温度计和微波湿度计结合将获得三维大气温度、湿度廓线。

(3) 中分辨率光谱成像仪--MERSI(Medium Resolution Spectral Imager):10个可见光波段,空间分辨率为250m和1000m;7个近红外波段,空间分辨率为250m和1000m;2个短波红外波段,空间分辨率为1000m;1个热红外波段,空间分辨率为250m。

风云四号(简称FY-4)是我国第二代静止气象卫星,目前处于型号背景研究阶段,计划在2014年发射风云四号卫星第一颗星,将有效载荷成像辐射计、大气垂直探测器、闪电成像仪、太阳X射线成像仪和空间环境检测器。主要发展目标是:大大提高观测的时间分辨率和区域机动探测能力;提高扫描成像仪性能,以加强中小尺度天气系统的监测能力;发展大气垂直探测和微波探测,解决高轨大气三位立体遥感;发展闪电成像观测,加强雷电预警预报;发展极紫外和X射线太阳观测,加强空间天气监测预警。

FY-4采用三轴稳定平台,使得在高轨道卫星上装载大气垂直探测仪成为可能。FY-4干涉式大气垂直探测仪将摒弃滤光片技术,直接采用干涉分光技术,是

我国红外高光谱探测技术的一种跨越，技术难度较大。FY-4 闪电成像仪覆盖中国及邻近陆地和海洋区域，对闪电进行连续观测成像，提供实时、连续、可靠的闪电观测数据，将实现对中国及邻近区域闪电的业务化监测、预警和预报^[6]。

1.3 红外高光谱资料同化研究现状

1.3.1 红外高光谱资料的同化现状

红外高光谱仪器 AIRS 和 IASI 在资料同化系统中起非常关键的作用，能提供高精度度和高垂直分辨率的信息，但是从这些仪器中获得的大量资料带来了很多挑战，尤其在资料传输、资料存储和同化领域^[10]。目前欧洲中期天气预报中心 (ECMWF) 业务上使用的 AIRS 和 IASI 辐射率仅限于光谱长波区域选定的温度遥感通道和集中在红外光谱带非常有限的湿度遥感通道。IASI 带 3 (2000 cm^{-1} - 2760 cm^{-1}) 覆盖的短波光谱区域包含了很好的温度遥感通道，同样能够同化后用于数值天气预报 (NWP) 系统。IASI 短波带的通道与长波区域类似的通道相比拟，因为一系列原因没有得到充分利用。这些原因包括云监测和辐射传输偏差，局部热力学平衡偏离引起的昼夜资料可用性的变化，以及仪器噪声等^[11]。

通道的选择通常受陆地和海冰面的约束，或者根据质量控制来剔除一些观测。世界各气象中心在同化系统中均不使用对陆面敏感的通道（基于技术难度），但是对研究这方面同化工作感兴趣。各中心对卫星遥感资料进行了大幅度地稀疏（起码每 4 个像素仅使用 1 个），都使用选择的近 200 个通道，并且主要使用长

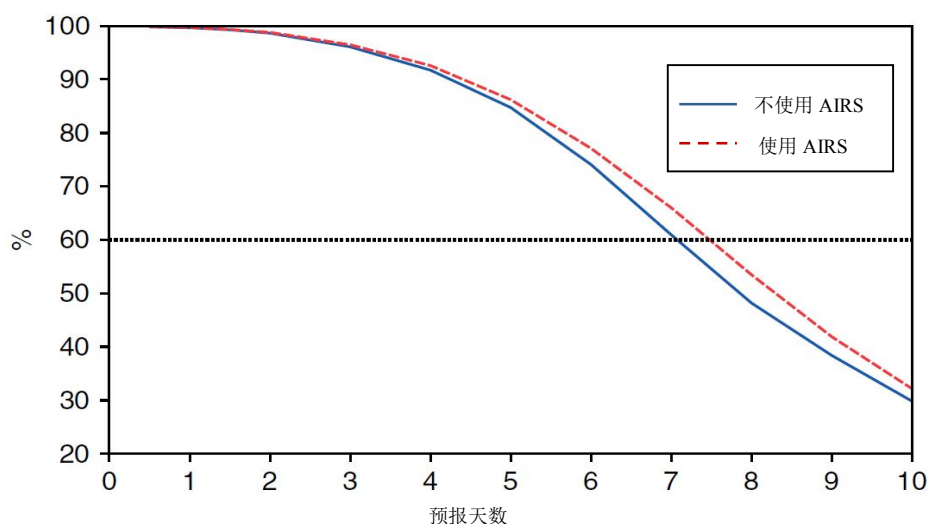


图 1.1 对比使用 AIRS 前后的预报评分

波 CO_2 带的通道。目前在业务中同化红外高光谱资料进入数值预报模式后获得了较好的预报评分，尤其在南半球常规探空资料相对稀缺的条件下，影响非常积极^[12]。图 1.1 描述了使用同化后的 AIRS 辐射率获得的南半球 500hPa 位势高度预报

反相关评分，相比业务分析中不使用 AIRS 资料，获得了明显的正效应。

1.3.2 问题的提出

高分辨率红外遥感光谱区域包含了许多具有非常相似光谱信号的通道，这些通道在内容上互相之间高度相关，而且存在大量的信息冗余。在辐射传输模式计算中，模拟完整光谱需要耗费大量的计算时间，同时对于资料同化而言，同化完整光谱代价很大并且效率很低。因此需要从红外高光谱资料中提取最具代表的通道，准确包含红外高光谱辐射率探测的信息，经过辐射传输模式处理以后进入资料同化系统，最终实现提高 NWP 模式的预报效果。

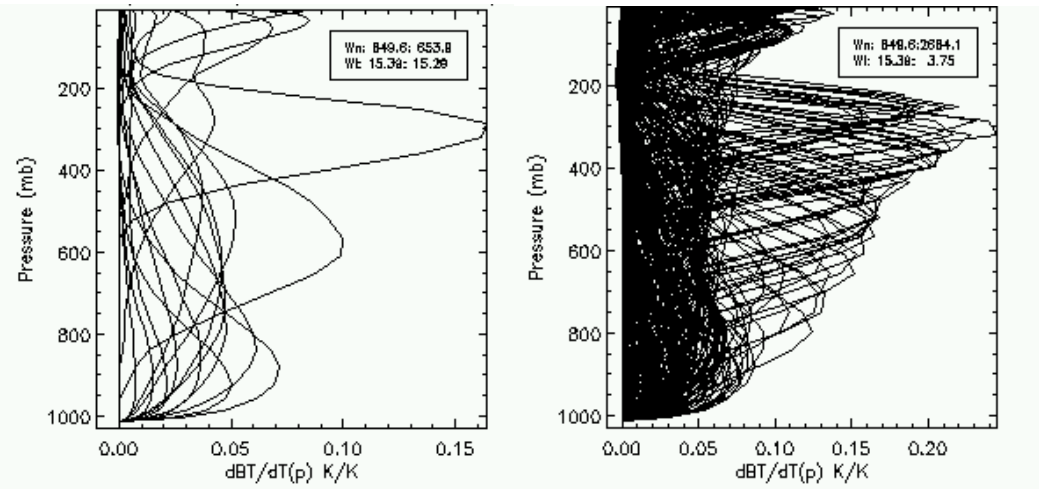


图 1.2 对比 HIRS 的 19 个通道（左图）和 IASI 的 8461 个通道（右图）

以前由于技术的限制，一般的卫星探测仪器只能提供数据量有限的通道，在这种情况下通道选择的原则是通过利用通道的光谱响应特性（如通道的中心频率、带宽等），控制通道的权重函数，使得观测结果能够最有效地对大气参数进行反演，此时求解的问题是欠定的。现在，大气红外探测仪具有几千甚至上万通道，问题的求解变成了超定的^[13]。图 1.2 显示了高分辨率干涉遥感仪（HIRS）探

表 1.2 ECMWF 2007 年 9 月业务同化 AIRS 和 IASI 的参数

	AIRS	IASI
获得的通道数	324	8461
检索的通道数	324	366
最大积极同化数	155	166
空间范围	海面 $\lambda < 14.14\mu\text{m}$	海面
光谱带	$15\mu\text{mCO}_2$ 带、长波窗口、 H_2O 带、 $4.5\mu\text{mCO}_2$ 带 ($\lambda > 4.45\mu\text{m}$)	$15\mu\text{mCO}_2$ 带
空间采样（稀疏到约 120km 之前）	最热 FOV（长波窗口） 每 AMSU FOV	探测器 1

测的 19 个通道和 IASI 探测的 8461 个通道。在 ECMWF 的业务同化中，基于每一个通道的 Jacobian 宽度、仪器噪声、冗余通道的干涉程度等因素，从 8461 个通道

中检索 366 重要的通道用于业务同化运行，如表 1.2 所示。与 HIRS 仅有 19 个通道相比，IASI 在业务操作中使用 366 个通道仍然显太多，由此引起新的问题，一个是计算量太大，另一个是通道之间存在相关性。计算量太大不利于业务应用。通道之间的相关性会造成反演或同化的不适定。如何尽可能多地提取有用信息，减小通道间的相关度，对于红外高光谱资料的同化是至关重要的。

对于高光谱红外探测器的通道选择国内外学者已经做了大量富有成效的工作。Rogers 提出基于信息熵的分步迭代法（以下简称为分步迭代法）进行通道选择^[14]。Florence 等实现了 Rogers 提出的分步迭代法进行通道选择的方法，并且提出了雅克比矩阵法，对比研究了这两种方法，得出的结论是：雅克比方法比迭代方法省时，但迭代方法的效果较好^[15]。最后综合了两者的优点，提出了常量迭代法，常量迭代法的目的是同一通道子集可用于不同的参考廓线，从而其实用性较强。Fourrie 提出手动选取通道的思想^[16]，主要采用实时数据评估了 Florence 提出的常量迭代法和自己提出的手动选取通道的方法，得出手动选取通道更符合实际情况。Andrew 和 James 从同化 AIRS 观测亮温的角度出发，采用分步迭代法，给出了所选的通道组合，白天选择了最优的 48 个通道，夜晚选择了最优的 62 个通道^{[17],[18]}。

杜华栋和张水平提出的通道选择方法的核心思想是借用 Florence 的基于信息熵的分步迭代法，刘辉提出的方法对基于权重函数所对应的峰值层进行通道选择。国内外通行的方法是基于信息熵的分步迭代法，此方法每次都是选信息量较大的通道，且前面选出的通道影响到后面的通道选择，但其存在的主要缺陷是在用迭代法进行通道选择时，依赖于背景误差和观测误差协方差矩阵的给定。然而，背景误差协方差的具体信息缺乏；分步迭代法每次是从未被选中的通道中选出一个，往往比较耗时。Florence 提出的雅克比矩阵法，Filipe 给出了具体的实施过程，首先，用背景误差和观测误差协方差矩阵对其雅克比矩阵进行标准化，然后，类似于权重函数的思想选每层峰值最大时所对应的通道。雅克比方法也依赖于背景误差和观测误差协方差矩阵的给定。在具体选取通道的过程中，只考虑了通道峰值最大贡献层，没有考虑通道之间的相关性。

针对上面方法的缺陷，本文采用主成分(PC)方法提取红外高光谱通道信息。主成分方法，既能保留最主要信息，又能得到影响及敏感性较大的通道，且去除了通道之间的相关性；直接对雅克比矩阵的结构进行分析，不需要过多的先验知识；可以同时考虑多个通道。理论上主成分方法利用这些通道之间的高度相关能够去掉观测中存在的大量不相关随机误差。基于 PC 方法过滤噪声，能够研究降低 IASI 带 3 仪器噪声后用于直接同化其主成分的可行性。

在 NWP 同化系统中直接使用 PC 要求有效的辐射传输(RT)模式(同样叫做

观测算子)从给定温度、水汽、臭氧、表面特性,或者未来使用的其它变量的第一猜值直接模拟 PC。ECMWF 同化卫星辐射资料采用的观测算子是基于 TOVS 资料的辐射传输(RTTOV)模式通过参数化大气透过率使模式足够精确和快速,以满足 NWP 近实时检测和同化卫星辐射率资料的要求^[19]。在业务上对于 AIRS 资料的同化,从 2378 个通道中选取 324 个通道,进行辐射传输透过率的计算^[20]。另外,ECMWF 为气候和空气质量中重要的示踪气体成分开发全面的检测和预报系统,在 RTTOV 输入状态向量中将可能包含更多的示踪气体的种类。实现这些要求将给 ECMWF 集合预报系统(IFS)带来大量计算负荷。因此更具计算效率的 RT 模式的有效性极可能成为同化高分辨率辐射资料的限制因素。

采用主成分方法分析红外高光谱资料的通道信息,研究基于主成分方法的辐射传输模式(PCRT)的研究与设计,并检验红外高光谱资料经主成分分析后在变分同化系统中获得的效果。尽管 PCRT 发展的基础理论是直接模拟 PC,注意到 PCRT 同样可以模拟光谱辐射率。因为主成分分析(使用一些线性组合的 PC 能够重构辐射率)中固有的维数减少,PCRT 比标准 RTTOV 辐射传输模式将能更有效地模拟辐射率。这方面的假设非常恰当,因为完全使用当前和未来高光谱分辨率遥感仪提供的大量信息要求使用更多通道中包含的有效信息。

1.4 本文的研究内容和创新

本论文实现了主成分方法在红外红光谱资料同化中的相关应用,分析了两种提取红外高光谱通道信息的方法,并测试了两种快速辐射传输模式 PCRT 和 RTTOV 的计算时间。详细探讨了 PCRT 模式辐射传输计算特征和开发测试过程,并且在四维变分资料同化系统框架中初步实现了 PCRT 模式。最后测试了 PCRT 模式相对于 RTTOV 模式的计算增益,并分析了经主成分方法提取的红外高光谱资料信息在同化试验中的效果及存在的问题,提出下一步研究计划。

主要工作包括:

(1) 通过研究目前卫星辐射率资料中最先进的红外高光谱资料,掌握其通道信息特征,了解数值天气预报中心对红外高光谱资料的同化现状和同化方法。

(2) 从光谱通道的角度,研究红外高光谱中包含的各种对应大气成分、结构等状态信息。用主成分方法分析红外高光谱资料,实现降低光谱噪声、消除使用通道之间的相关性,尽可能多地包含光谱中重要的信息。

(3) 研究实现基于主成分方法的快速辐射传输 PCRT 模式的设计与开发,从大气成分的吸收特性、透过率计算、光学厚度计算等环节与主成分的思想结合起来,形成能直接模拟主成分光谱的快速辐射传输模式,分析主成分中预报因子数目、主成分数目对误差水平以及计算开销的影响。

(4) 将 PCRT 模式在采用目前业务和研究中最先进的四维变分资料同化系统 (4D-Var) 框架中实现, 进行多组红外高光谱资料主成分的同化试验, 并分析主成分方法同化红外高光谱资料对预报评分的影响, 并计算 PCRT 模式相对 RTTOV 模式获得的计算增益。

1.5 论文结构

本文共分为六章:

第一章: 引言主要介绍课题背景, 国内外研究现状以及本论文的创新及论文结构;

第二章: 介绍主成分方法的原理及其提取红外高光谱通道信息的方法;

第三章: 介绍基于主成分方法的 PCRT 模式的设计思路;

第四章: 在四维变分资料同化系统框架中初步实现 PCRT 模式;

第五章: 主成分方法的测试与同化试验;

第六章: 结束语。概括全文, 总结课题工作和研究成果, 并展望下一步工作。

第二章 红外高光谱主成分通道选择方法

气象卫星被动遥感地球和大气（包括目标物与非目标物）所发射或反射的电磁波信息,但是为了得到目标物的信息，不能接受所有的电磁波，必须使用通道去进行遥测。卫星传感器只感应某一特定波段的电磁波，对其它波段的电磁波则一概过滤掉，这一特定波段就叫做通道^[21]。对应本文的应用，我们通常认为被同化的离散参量为通道，区别于广义的被仪器直接测量的专用通道。

由于在每一个拥有 8461 个通道的 IASI 光谱中通常只包含几十种独立信息，因此我们可以用一些独立的信息块描绘这些光谱，即使明显减少资料容量时，仍然能够保留主要信息。

为了有效使用上千通道遥感辐射率资料所包含的信息，必须研究描绘信息的方法，即用较小数目的通道遥感辐射率包含其中绝大部分信息。目前业务上普遍使用简单通道选择的办法：对通道进行简单的选择，对每一种仪器都采用两种选择通道的策略，即选择一部分常规检测通道和一部分积极用于同化的通道子集。检测通道不用于同化，但是可以检查不被积极同化的波段，由此从中可以洞悉仪器与模式的性能。这种方法由于删除了绝大部分通道，从几千通道中只选择几百个通道用作同化，因此这种方法不能在计算中包含绝大部份通道所包含的信息。

本章详细介绍主成分方法的基本原理，并分析主成分方法提取的红外高光谱信息的原理。主成分方法：应用仪器辐射率资料的子集，推断光谱变化的主成分，将上千通道光谱的信息压缩在几百个主成分中。为了对比，本章还简单介绍了目前业务使用的通道选择方法。

2.1 主成分方法

主成分分析法是从观测数据中获取主要信息的一种多变量统计方法。它用数目少得多的新变量代替原有观测变量，以寻找原观测变量之间的相关关系，且不损失原数据的主要信息^[22]。目前，主成分分析法被广泛应用于天体物理学、经济学、农业统计学等领域。

2.1.1 主成分方法的原理

主成分分析是把原来多个变量化为少数几个综合指标的一种统计分析方法，从数学角度来看，这是一种降维处理技术。假设现在需要研究 m 个变量，它可以是 m 个气象要素，也可以是气象要素场中的 m 个空间点的值^[23]。每个变量有 n 个观测量，每个数据元素为 x_{it} , $i=1,2,\cdots, m$; $t=1,2, \cdots, n$ 。写成矩阵形式为

$$X = (x_{it}) = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \quad (2.1)$$

如何从如此多变量的数据中抓住事物的内在规律性，要解决这一问题，自然要在 n 维空间中加以考察，这是比较麻烦的。为了克服这一困难，就需要进行降维处理，即用较少的几个综合指标来代替原来较多的变量指标，而且使这些较少的综合指标既能尽量多地反映原来较多指标所反映的信息，同时它们之间又是彼此独立的。那么，需要考虑如何选取这些综合指标（即新变量），显然，其最简单的形式就是取原来变量指标的线性组合，适当调整组合系数，使新的变量指标之间相互独立且代表性最好。

将 X 进行线性变换，组成 m 个新的变量 z_1, z_2, \cdots, z_m ，或者说，新变量是由原变量 x_1, x_2, \cdots, x_m 进行线性组合来构成。

$$Z = (z_{it}) = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1n} \\ z_{21} & z_{22} & \cdots & z_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ z_{m1} & z_{m2} & \cdots & z_{mn} \end{bmatrix} \quad (2.2)$$

$$\text{满足} \quad z_{it} = v_{i1}x_{1t} + v_{i2}x_{2t} + \cdots + v_{im}x_{mt} = \sum_{k=1}^m v_{ik}x_{kt} \quad (2.3)$$

$$i=1,2, \cdots, m \quad t=1,2, \cdots, n$$

全部写出，并用矩阵表示为

$$\begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1n} \\ z_{21} & z_{22} & \cdots & z_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ z_{m1} & z_{m2} & \cdots & z_{mn} \end{bmatrix} = \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1m} \\ v_{21} & v_{22} & \cdots & v_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ v_{m1} & v_{m2} & \cdots & v_{mm} \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \quad (2.4)$$

$$\text{或者} \quad Z = V^T X \quad (2.5)$$

$$\text{其中} \quad V = (v_1 \quad v_2 \quad \cdots \quad v_m) = \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1m} \\ v_{21} & v_{22} & \cdots & v_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ v_{m1} & v_{m2} & \cdots & v_{mm} \end{bmatrix}$$

(2.6)

是待定的。

2.1.2 主成分的求解过程

显然，不同的 V 矩阵，对应不同的线性变换。要给定一定的原则，使组成的新变量会有助于对气象要素场的分析。这些原则是：第一，要求新变量能最大限度地、集中地反映原 m 个变量的总方差。第二，要求这些新变量互相独立，这样就可以用较少的新变量描述原气象要素场的主要特征。这些新变量就是主成分。

下面就根据这些原则求出系数矩阵 V ，从而得到 m 个主成分 z_1, z_2, \dots, z_m 。

第一主成分的方差为

$$\frac{1}{n} \sum_{t=1}^n (z_{1t} - \bar{z}_1)^2 \quad (2.7)$$

据 (2.3) 式

$$\begin{aligned} \bar{z}_1 &= \frac{1}{n} \sum_{t=1}^n (z_{1t}) = \sum_{k=1}^m v_{1k} \cdot \frac{1}{n} \sum_{t=1}^n x_{kt} \\ &= \sum_{k=1}^m (v_{1k} \bar{x}_k) = v_{11} \bar{x}_1 + v_{12} \bar{x}_2 + \dots + v_{1m} \bar{x}_m \end{aligned} \quad (2.8)$$

代入 (2.7) 第一主成分方差

$$\begin{aligned} &\frac{1}{n} \sum_{t=1}^n (z_{1t} - \bar{z}_1)^2 \\ &= \frac{1}{n} \sum_{t=1}^n [v_{11}x_{1t} + v_{12}x_{2t} + \dots + v_{1m}x_{mt} - \\ &\quad (v_{11}\bar{x}_1 + v_{12}\bar{x}_2 + \dots + v_{1m}\bar{x}_m)]^2 \\ &= \frac{1}{n} \sum_{t=1}^n [v_{11}(x_{1t} - \bar{x}_1) + v_{12}(x_{2t} - \bar{x}_2) + \dots + v_{1m}(x_{mt} - \bar{x}_m)]^2 \\ &= \frac{1}{n} \sum_{t=1}^n \left[v_1^T \cdot \begin{pmatrix} x_{1t} - \bar{x}_1 \\ x_{2t} - \bar{x}_2 \\ \vdots \\ x_{mt} - \bar{x}_m \end{pmatrix} \right] \cdot \left[\begin{pmatrix} x_{1t} - \bar{x}_1 \\ x_{2t} - \bar{x}_2 \\ \vdots \\ x_{mt} - \bar{x}_m \end{pmatrix}^T \cdot v_1 \right] \\ &= \frac{1}{n} v_1^T \sum_{t=1}^n \begin{pmatrix} x_{1t} - \bar{x}_1 \\ x_{2t} - \bar{x}_2 \\ \vdots \\ x_{mt} - \bar{x}_m \end{pmatrix} \cdot \begin{pmatrix} x_{1t} - \bar{x}_1 \\ x_{2t} - \bar{x}_2 \\ \vdots \\ x_{mt} - \bar{x}_m \end{pmatrix}^T \cdot v_1 \\ &\quad \cdot \begin{pmatrix} x_{1t} - \bar{x}_1 \\ x_{2t} - \bar{x}_2 \\ \vdots \\ x_{mt} - \bar{x}_m \end{pmatrix} \cdot (x_{1t} - \bar{x}_1 \quad x_{2t} - \bar{x}_2 \quad \dots \quad x_{mt} - \bar{x}_m) \end{aligned} \quad (2.9)$$

而

$$\begin{aligned}
&= \frac{1}{n} \sum_{t=1}^n \begin{bmatrix} (x_{1t} - \bar{x}_1)^2 & \cdots & (x_{1t} - \bar{x}_1)(x_{mt} - \bar{x}_m) \\ (x_{1t} - \bar{x}_1)(x_{2t} - \bar{x}_2) & \cdots & (x_{2t} - \bar{x}_2)(x_{mt} - \bar{x}_m) \\ \vdots & \vdots & \vdots \\ (x_{1t} - \bar{x}_1)(x_{mt} - \bar{x}_m) & \cdots & (x_{mt} - \bar{x}_m)^2 \end{bmatrix} \\
&= \begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1m} \\ S_{21} & S_{22} & \cdots & S_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ S_{m1} & S_{m2} & \cdots & S_{mm} \end{bmatrix} \tag{2.10}
\end{aligned}$$

其中 $s_{ij} = \frac{1}{n} \sum_{t=1}^n (x_{it} - \bar{x}_i)(x_{jt} - \bar{x}_j)$, $i, j=1, 2, \dots, m$

$$S = (s_{ij}) \tag{2.11}$$

S 是样本方差矩阵。这样，第一主成分的方差化为

$$\frac{1}{n} \sum_{t=1}^n (z_{1t} - \bar{z}_1)^2 = \mathbf{v}_1^T S \mathbf{v}_1$$

要使上面表达式达到最大值，如果不对 \mathbf{v}_1 的模作一定限制，问题就毫无意义了。可限 $\mathbf{v}_1^T \mathbf{v}_1 = v_{11}^2 + v_{12}^2 + \cdots + v_{1m}^2 = 1$ 是一个单位向量，在此条件下求使 $\mathbf{v}_1^T S \mathbf{v}_1$ 达最大的向量 \mathbf{v}_1 ，这是一个求条件极值问题。可化为求

$$\Phi = \mathbf{v}_1^T S \mathbf{v}_1 - \lambda (\mathbf{v}_1^T \mathbf{v}_1 - 1)$$

的极值， λ 称为拉格朗日乘数。利用矩阵微分运算关系，令

$$\begin{aligned}
\frac{\partial \Phi}{\partial \mathbf{v}_1} &= 2S\mathbf{v}_1 - 2\lambda \mathbf{v}_1 = 0 \\
(S - \lambda I)\mathbf{v}_1 &= 0 \tag{2.12}
\end{aligned}$$

欲使 \mathbf{v}_1 有解，必须

$$|S - \lambda I| \mathbf{v}_1 = 0 \tag{2.13}$$

上式就是矩阵 S 的特征方程。问题转化为求 S 得特征值和对应的特征向量。而 m 阶矩阵有 m 个特征值 $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_m$ ，对应的特征向量记为 $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$ 。因为

$$\mathbf{v}_1^T S \mathbf{v}_1 = \lambda_1 \tag{2.14}$$

所以，选取特征方程中 m 个特征值中最大的特征值所对应的特征向量作为

$\mathbf{v}_1 = (v_{11} \ v_{12} \ \cdots \ v_{1m})^T$ ，就能使第一个主成分 z_1 的方差达到最大值 λ_1 。

再求第二个主成分 z_2 ，它的方差为

$$\frac{1}{n} \sum_{t=1}^n (z_{2t} - \bar{z}_2)^2 = \mathbf{v}_2^T \mathbf{S} \mathbf{v}_2$$

要求 z_1 和 z_2 不相关，即他们的协方差为 0。

$$\begin{aligned} & \frac{1}{n} \sum_{t=1}^n (z_{1t} - \bar{z}_1)(z_{2t} - \bar{z}_2) \\ &= \frac{1}{n} \sum_{t=1}^n \left[\mathbf{v}_1^T \cdot \begin{pmatrix} x_{1t} - \bar{x}_1 \\ x_{2t} - \bar{x}_2 \\ \vdots \\ x_{mt} - \bar{x}_m \end{pmatrix} \right] \cdot \left[\begin{pmatrix} x_{1t} - \bar{x}_1 \\ x_{2t} - \bar{x}_2 \\ \vdots \\ x_{mt} - \bar{x}_m \end{pmatrix}^T \cdot \mathbf{v}_2 \right] \\ &= \mathbf{v}_1^T \mathbf{S} \mathbf{v}_2 = \mathbf{v}_2^T \mathbf{S} \mathbf{v}_1 = \mathbf{v}_2^T \lambda_1 \mathbf{v}_1 = \lambda_1 \mathbf{v}_2^T \mathbf{v}_1 = 0 \end{aligned} \quad (2.15)$$

因为 $\lambda_1 \neq 0$ ，所以必须 $\mathbf{v}_2^T \mathbf{v}_1 = 0$ ，也就是两个向量互相正交。所以第二个主成分方差最大，并且与第一主成分正交，应有

$$Q = \mathbf{v}_2^T \mathbf{S} \mathbf{v}_2 - \lambda (\mathbf{v}_2^T \mathbf{v}_2 - 1) - 2\nu (\mathbf{v}_2^T \mathbf{v}_1)$$

达到极大。 λ 是拉格朗日乘数，对矢量 \mathbf{v}_2 微分，令其为 0，利用

$$\frac{\partial (\mathbf{x}^T \mathbf{A} \mathbf{x})}{\partial \mathbf{x}} = 2 \mathbf{A} \mathbf{x}$$

$$\text{得} \quad \frac{\partial Q}{\partial \mathbf{v}_2} = 2 \mathbf{S} \mathbf{v}_2 - 2 \lambda \mathbf{v}_2 - 2 \nu \mathbf{v}_1 = 0 \quad (2.16)$$

用 \mathbf{v}_1^T 左乘上式，并消去倍数 2，并注意 (2.15)，可得

$$\mathbf{v}_1^T \mathbf{S} \mathbf{v}_2 - \mathbf{v}_1^T \lambda \mathbf{v}_2 - 2 \nu \mathbf{v}_1^T \mathbf{v}_1 = 0 - 0 - \nu = 0$$

所以 $\nu = 0$ ，这样 \mathbf{v}_2 应满足的方程为

$$\mathbf{S} \mathbf{v}_2 - \lambda \mathbf{v}_2 = 0 \quad (2.17)$$

此方程与 (2.12) 完全一样。所以，不难看出，此时取 $\lambda = \lambda_2$ ， \mathbf{v}_2 为对应 λ_2 的特征向量，如此下去，采用同样的方法可以求出 m 个主成分，它是方差阵 \mathbf{S} 的第 m 个特征值所对应的特征向量，第 m 个主成分的方差是 λ_m 。

综上所述，求取主成分的方法是首先据 (2.11) 式求出样本协方差矩阵 \mathbf{S} 。然后，求 \mathbf{S} 的特征值和特征向量。特征值 $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_m$ ，对应的特征向量为 $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$ ，并组成矩阵 $\mathbf{V} = (\mathbf{v}_1 \ \mathbf{v}_2 \ \dots \ \mathbf{v}_m)$ 。最后，根据 (2.5) 式，由 $\mathbf{Z} = \mathbf{V}^T \mathbf{X}$ 求出主成分。

2.1.3 主成分的重要性--主成分的贡献

因为主成分的方差贡献是按 \mathbf{S} 矩阵的特征值大小排列的，所以，我们通常不

需要全部主成分，只需要前面几个。

定义：
$$C(k) = \lambda_k / \sum_{i=1}^m \lambda_i \quad (2.18)$$

为主成分 z_k 的贡献率， $C(k)$ 的数值越大，表明主成分 z_k 的贡献率越大，也就是它“综合反映” x_1, x_2, \dots, x_m 的能力越强。或者说，用 z_k 反映向量 \mathbf{x} 的能力越强，所以才把 z_k 称为主成分，也就是 \mathbf{x} 的主要部分。

前 p 个主成分的累积贡献率为：

$$C(p) = \sum_{k=1}^p \lambda_k / \sum_{k=1}^m \lambda_k \quad (2.19)$$

通常当 $p \ll m$ 时， $C(p)$ 就接近 100%，所以对于后面的主成分，因为贡献率很少，可以略去不用。

2.2 主成分方法研究红外高光谱

红外高光谱虽然拥有上千个光谱通道，但是这些通道中包含大量相似信息，因此从信息内容的角度分析，可以认为红外高光谱包含了远小于通道数目的独立信息块。表 2.1 描述了 2378 个 AIRS 光谱通道的主要特征^[24]。采用主成分的方法，从代表完整气象条件的光谱中，计算出主成分，并表示成数目为几百个的光谱，

表 2.1 AIRS 光谱通道的主要特征

通道号 范围	光谱范围 (cm ⁻¹)	光谱分辨率 (cm ⁻¹)	灵敏度 (K) NEDT(280K)	通道数	探测内容
1-130	649.61-681.99	0.41-0.46	0.321-0.697	130	温度
131-274	687.60-728.44	0.47-0.53	0.211-0.365	143	温度
275-441	728.06-781.88	0.55-0.63	0.219-1.113	166	表面
442-608	789.26-852.43	0.65-0.76	0.246-1.285	166	表面
609-769	851.49-903.78	0.57-0.64	0.125-0.544	160	表面
770-936	911.24-974.29	0.65-0.75	0.081-1.512	166	表面
937-1103	973.82-1046.20	0.76-0.88	0.072-0.266	166	臭氧
1104-1262	1056.10-1136.66	0.9-1.04	0.108-0.869	158	地表、臭氧
1263-1368	1216.97-1272.59	0.95-1.04	0.076-0.33	105	水汽
1369-1462	1284.35-1338.86	1.06-1.16	0.101-0.278	93	水汽、甲烷
1463-1654	1338.16-1443.07	0.95-1.11	0.072-0.148	191	水汽
1655-1760	1460.27-1527.00	1.15-1.25	0.071-0.22	105	水汽
1761-1864	1541.10-1613.86	1.28-1.41	0.12-0.47	103	水汽
1865-2014	2181.50-2325.06	1.75-1.98	0.086-0.248	149	温度、CO
2015-2144	2299.80-2422.85	1.74-1.94	0.13-0.243	129	温度
2145-2260	2446.20-2569.75	1.98-2.18	0.177-0.58	115	表面
2261-2378	2541.90-2665.24	1.94-2.13	0.302-1.026	117	表面

能开发使用红外高光谱资料的潜力，这样当大量减少使用资料时仍然能保持大量的信息。

2.2.1 求解红外高光谱主成分

主成分方法非常适合有效表示红外高光谱上千通道包含的信息，尽管相比全部通道的信息量稍小。这种技术使用上千光谱资料集代表完整大气状况，然后由光谱资料集计算出主成分^[6]。这些光谱是噪声标准化的，并且其中去掉总体的平均值。噪声标准化保证了对低信噪比的光谱区域噪声进行非选择性的拟合。使用辐射率而非亮温同样是正规的，因为仪器噪声在辐射率空间更恒定。另外，在信噪比非常低的区域，例如 IASI 光谱的短波部分，可能产生消极的辐射率不能由亮温代表^[25]。

如果这种辐射率的集合由 $n_{ch} \times m$ 的矩阵 X 代表（其中 n_{ch} 是通道数， m 是观测数，因此 X 的每一列代表一个独立的光谱），这些光谱的协方差 C 表示为

$$C = \frac{1}{m} XX^T = L \Lambda L^T \quad (2.20)$$

C
 X
 X^T
 L
 Λ
 L^T

其中， L 是主成分的矩阵， Λ 是相应的特征值。换句话说， X 能够通过奇异值分解为：

$$X = UWV^T \quad (2.21)$$

矩阵 U, W 和 V 的维数分别为 $n_{ch} \times q$ 、 $q \times q$ 和 $m \times q$ ，其中 q 是比 n_{ch} 和 m 的更小的数。因为（注意到 $V^T V = I$ ， $W = W^T$ ，并且，如果 $q = m$ ，所有 q 个数目以后的特征向量将是 0）：

$$n_{obs} \times L_q \Lambda_q L_q^T = XX^T = UWV^T VWU^T = UWWU^T \quad (2.22)$$

那么可以看作 U 等价于 L_q ， Λ_q 等价于（除以 m 比例因子） WW ，其中 L_q 和 Λ_q 对应于 L 和 Λ 的前 q 个特征向量和特征值。

选择采用协方差或者奇异值分解的方式计算主成分由问题的特殊性来决定。当需要全局样本集时，如果试图在样本集中尽可能多的多包含光谱数目，光谱数目通常达到上万。在这种情况下采用（和任意更新）协方差矩阵 XX^T 计算方式， X 矩阵可能变得非常大并且更有效，尤其在计算存储方面^[26]。

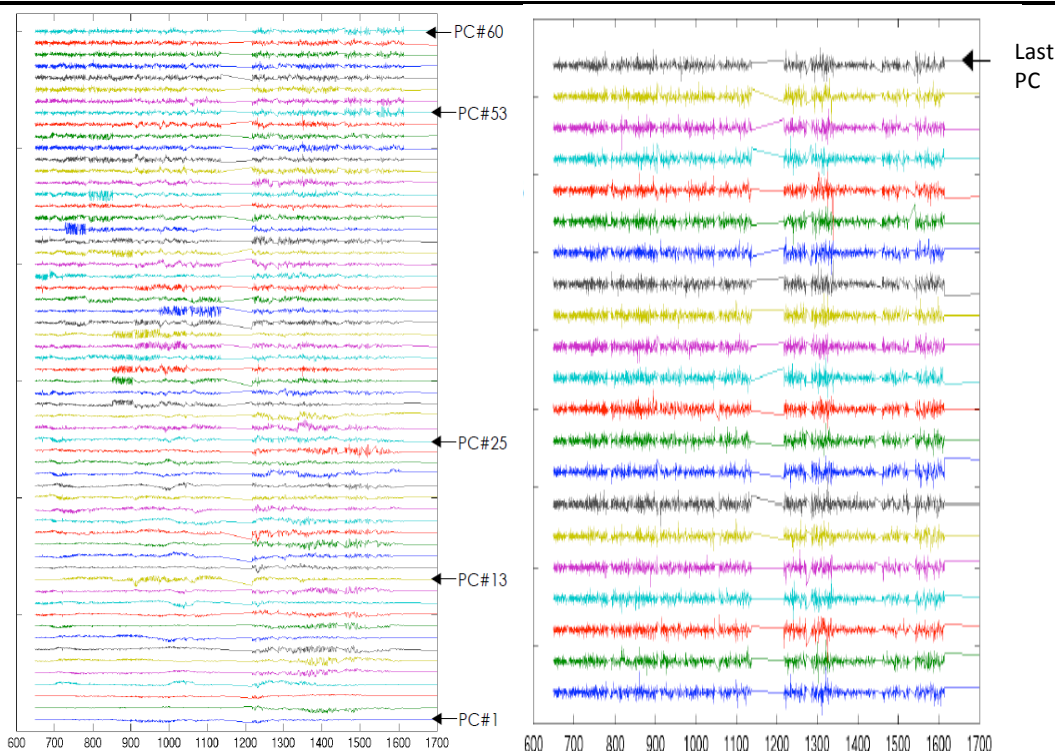


图 2.1 按特征值递减顺序排序的 AIRS 主成分谱线

对于辐射率变迹的干涉仪，需要考虑对噪声标准化光谱使用校正相关误差协方差矩阵的情况。辐射率变迹的过程中在相邻通道间引入了强误差相关。对噪声标准化 IASI 变迹辐射率使用对角误差协方差存在一个危险^[27]，即仪器分辨率尺度的光谱特征信号（这些在变迹过程被平滑）可能不能在主成分中表现出来。对于非变迹层光谱，误差协方差矩阵将对角化。

2.2.2 选择光谱资料集

主成分的振幅 p ，即通常所说的“谱线”，与观测辐射率 y 相关（观测辐射率是噪声标准化观测的辐射率，减去了平均观测），通过

$$p = L^T y \quad (2.23)$$

通常发现由全局产生的大约 200 个主成分能够获得原始光谱中包含的大气信号，剩余的成分大部分代表噪声。图 2.1 描绘了对应特征值递减顺序排列的 AIRS 主成分谱线。因此考虑最主要的主成分是遥感观测中有效信息通信的新基础。这可以用于多种目的：资料中心与 NWP 中心之间辐射率有效通信，NWP 模式资料同化系统中直接同化主成分辐射率。

本论文研究将主成分方法应用到红外高光谱资料的四维变分同化中。四维变分同化中的卫星资料观测算子 H 将大气状态变量 x 模拟成对应的辐射率或者等效亮温观测 y ，其代价函数可以简略描述如图 2.2 所示。

$$H(x) = \begin{bmatrix} \frac{\partial H_1}{\partial T_1} & \frac{\partial H_1}{\partial T_2} & \cdots & \frac{\partial H_1}{\partial T_n} & \frac{\partial H_1}{\partial q_1} & \frac{\partial H_1}{\partial q_2} & \cdots & \frac{\partial H_1}{\partial q_n} \\ \frac{\partial H_2}{\partial T_1} & \frac{\partial H_2}{\partial T_2} & \cdots & \frac{\partial H_2}{\partial T_n} & \frac{\partial H_2}{\partial q_1} & \frac{\partial H_2}{\partial q_2} & \cdots & \frac{\partial H_2}{\partial q_n} \\ \frac{\partial H_3}{\partial T_1} & \frac{\partial H_3}{\partial T_2} & \cdots & \frac{\partial H_3}{\partial T_n} & \frac{\partial H_3}{\partial q_1} & \frac{\partial H_3}{\partial q_2} & \cdots & \frac{\partial H_3}{\partial q_n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial H_m}{\partial T_1} & \frac{\partial H_m}{\partial T_2} & \cdots & \frac{\partial H_m}{\partial T_n} & \frac{\partial H_m}{\partial q_1} & \frac{\partial H_m}{\partial q_2} & \cdots & \frac{\partial H_m}{\partial q_n} \end{bmatrix} \quad (2.24)$$

其中：m 为通道个数；n 为模式层数；T 为温度；q 为湿度。

图 2.3-图 2.5 描绘了由 IASI 观测生成的前 4 个 IASI 光谱的主成分和相应的温度 Jacobians、湿度 Jacobians。可以看到主成分的 Jacobians 是极其非局部的，并且每一个对温度和湿度都有贡献。通过选择完整光谱的子集产生主成分，存在可行性即使 Jacobians 变得更局部化，并且温度和湿度信号能够分离到一定程度，但是采用少量复杂的方案例如重构辐射率的方式，能够使用主成分重构出更具真实通道特征的辐射率，下一小节中将做出解释。

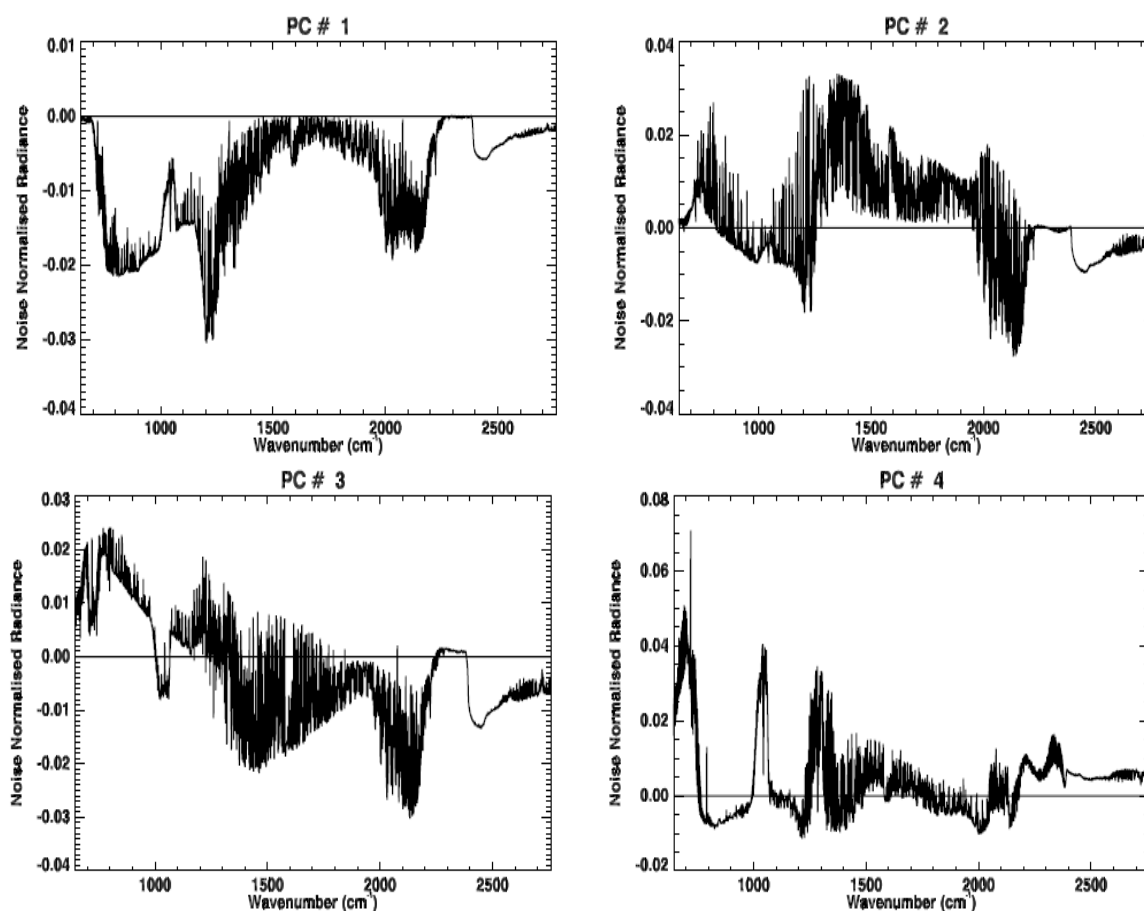


图 2.3 IASI 光谱的前 4 个主成分

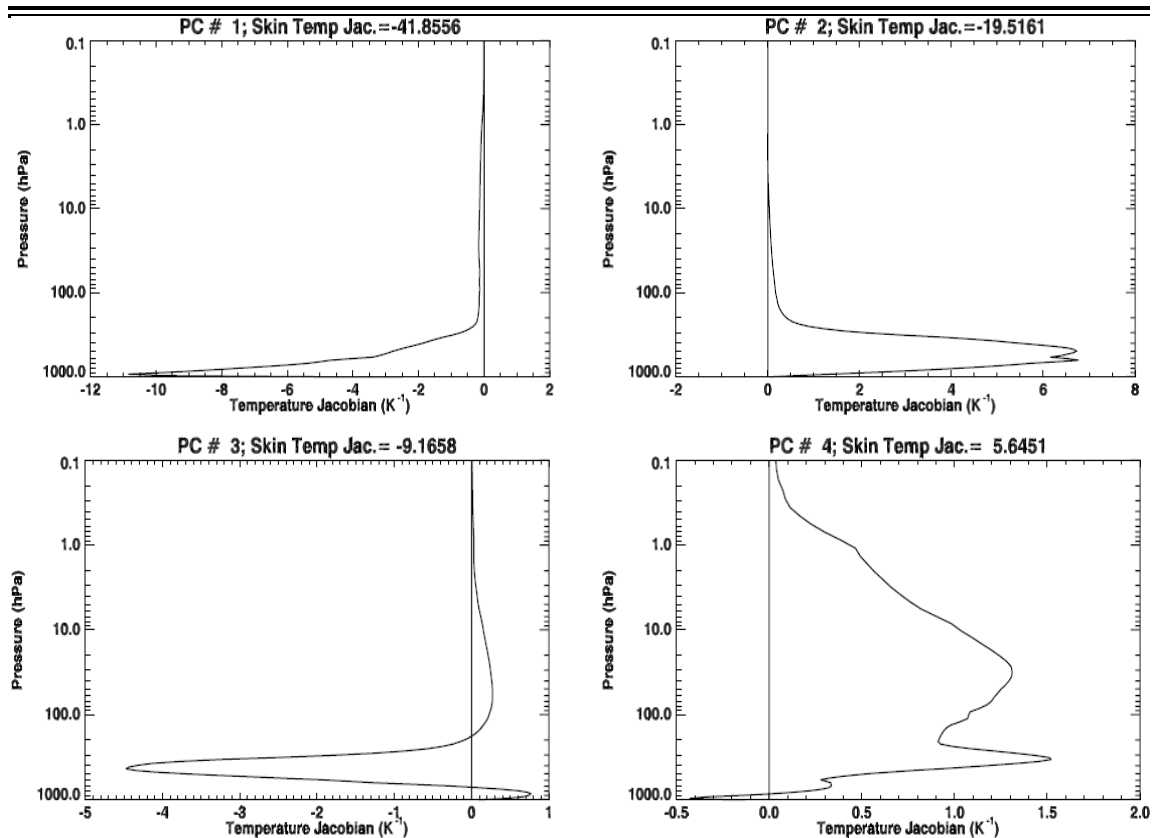


图 2.4 IASI 前 4 个主成分的温度 Jacobians

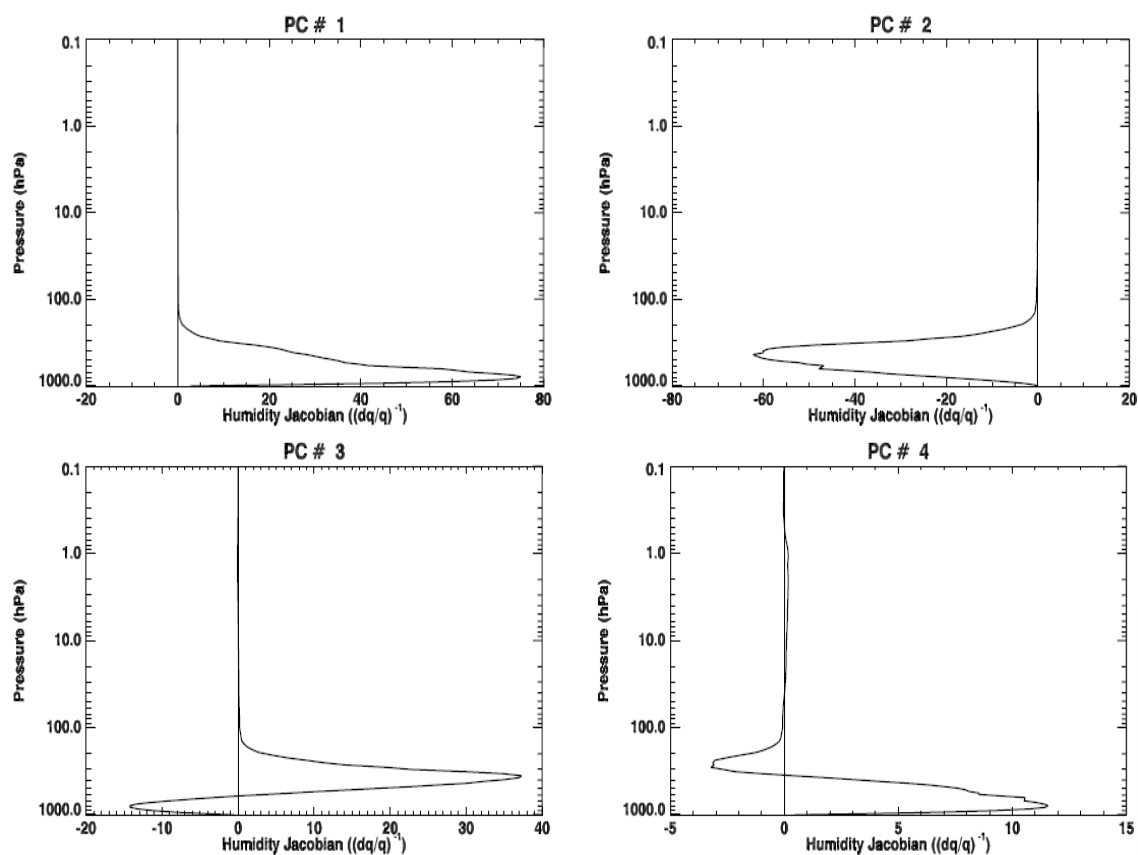


图 2.5 IASI 前 4 个主成分的湿度 Jacobians

在 NWP 中使用主成分最终的问题是监测。常规光谱观测 Jacobians 的相对局部性质允许通道与大气海拔和吸收种类之间的粗糙相关。这是非常有价值的，尤其可以用于诊断同化系统本身存在的问题并且区分这些缺陷是由模式产生还是观测产生。通过分析主成分的函数，可以推理出这种缺陷不是主成分方法引入的。因此，这绝不是一个不能克服的问题--光谱通道的选择与主成分的海拔可以同时检测--但是确实说明主成分对系统的影响远没有正常光谱辐射率直观^[28]。

2.2.4 重构辐射率

虽然采取一定的策略可以在同化系统中直接同化主成分谱线，但是正如上文所描述的由于主成分的 Jacobians 具有高度非局部化，可能在对流层的分析引入额外的平流层影响。针对主成分 Jacobians 非局部化这种情况，可以采用重构辐射率的方法解决这种问题。

重构的辐射率 \tilde{y} 由主成分振幅即主成分谱线 P_n 计算得到：

$$\tilde{y} = L_n p_n = L_n L_n^T y = y L_n L_n^T \quad (2.25)$$

这里下标“n”表明只考虑有前面 n 个主成分；其余的主成分则假设主要包含的噪声信息，y 表示噪声标准化的辐射率，即减去了观测的平均值。如果严格将 \tilde{y} 与 r 个通道子集对应，通过 L_n 上最开始的 L_m 个主成分（即只选择 $r \times n$ 矩阵的行向量表示 L_n 中研究所需的通道），这些通道将包含所有存在于 n 个主成分中的信息，提供 $L_n L_m^T$ 至少具有 n 个正奇异值。最小值的条件为 $r \geq n$ ，并且在实际应用中通常满足这个条件。

实际上由试验证明，例如，如果从最开始的 200 个主成分中产生辐射率，所获得整个重构的光谱精度能够达到机器的精度，即从这最开始的 200 个主成分‘通道’和 $L_m L_n^T$ 所生成的辐射率噪声水平相当于仪器噪声水平。因此对 650cm^{-1} 附近通道重构的辐射率能够测量对流层温度同时还能包含太阳光影响 2600cm^{-1} 附近近地面通道的信息。图 2.6 描述了从 IASI 光谱中采样的光谱和将主成分重构出的光谱。由图中可以看出重构光谱的性质和采样光谱的性质非常相似。因此可以将这些重构的‘通道’当作真实观测的辐射率用于资料同化。

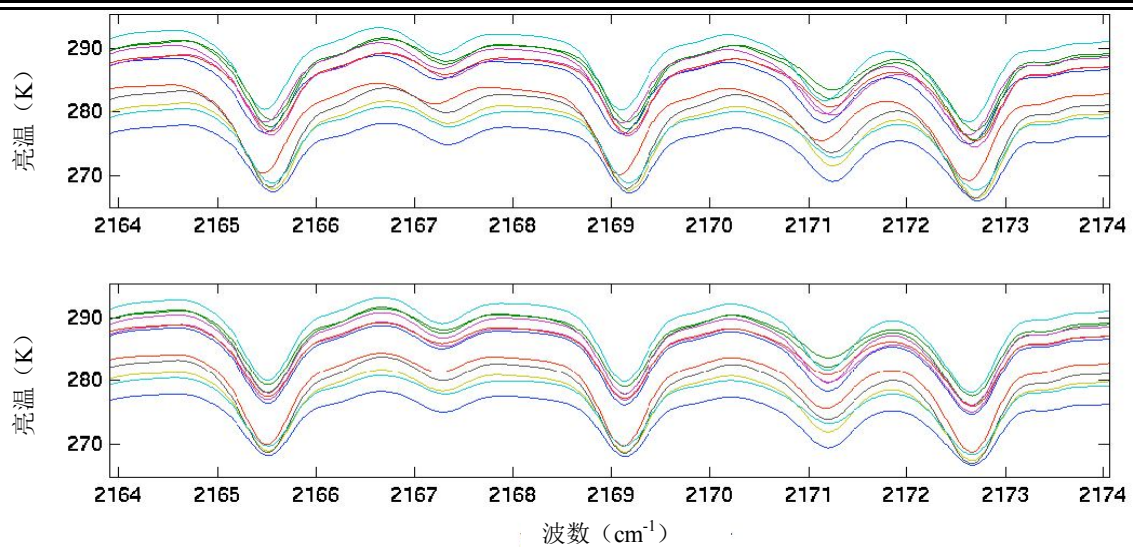


图 2.6 IASI 采样的原始光谱（上图）和重构的光谱（下图）

2.3 其它红外高光谱信息表示方法

2.3.1 多种通道信息表示方法

卫星传感器接收到的辐射一部分是目标物的辐射，并受到非目标物（背景）的影响和传输路径介质的影响，所以需要有效提取通道信息突出所要观测目标物的信息，将其它影响减到最小。

由于观测通道间存在相关性，给定了一套通道后，该套通道一般只包含有限的独立信息。通常，一套通道中有多少相互独立的信息片，就意味着该通道包含有多少“有效信息量”，曾庆存指出，并不能无条件地不断提高测量精度和增加通道数目而无限准确地间接推断出气温垂直分布^[29]。处理完整的红外高光谱没有意义，还可能在同化系统中增加不必要的计算负荷。

由于在每一个 IASI 光谱中通常只有几十种独立信息^[30]，因此可以用这几十种信息描述观测，当大量减少使用的资料容量时，仍然能够保存主体信息。为了有效使用上千通道遥感辐射率所包含的信息，必须研究有效刻画信息的方法，即通过使用较小数的通道辐射率包含绝大部分信息（这里通道我们指被同化的离散参量，可以是仪器探测的专门通道，也可以是前面提出的主成分谱线或者重构出的辐射率）。提取红外高光谱信息的方法除了上文描述的主成分方法，还包括通道选择方法、超级通道、反演方法。

- 直接通道选择方法：基于仪器噪声、Jacobians 宽度等原因从完整光谱中选取部分检测通道，再从这些检测通道中选择出积极用于同化的通道子集。
- 超级通道：对光谱中相似的通道算平均值以产生性能相似的一个通道，

去掉大量退化的随机噪声；

- 反演方法：用大气状态如温度廓线、湿度廓线等表示遥感光谱。

2.3.2 直接通道选择方法

直接通道选择方法根据使用目的选择两种类型的通道，一部分常规检测通道的，另一部分为积极用于同化的通道。检测通道不被同化，但是可以检测其中不被积极同化的波段，用以分析仪器与模型的性能。下面以 AIRS 和 IASI 为例，介绍通道选择方法的策略和步骤。

AIRS 从整个光谱中选择 324 个通道子集。选择通道时需要考虑每一个通道的 Jacobians 宽度、仪器噪声、其余通道对其的干涉程度，以及每一个检测通道的作用等因素。这样重新获得的光谱降低了观测光谱的噪声，达到仪器噪声的程度。

在 AIRS 的 324 个通道中，最多 155 个通道能够被同化（取决于同化系统中估计出的云顶高度）。下面列出不被积极同化的三个光谱区域：

- 波长小于 $4.45\mu\text{m}$ ，可能被白天太阳光辐射和 non-LTE（非局部热力平衡）条件下影响的区域（目前也不使用晚上的这些资料）；
- 9.5 微米左右的臭氧带；
- 大量分布于平流层的通道（因为模式等特殊原因不被使用）。

另外，一部分通道明显受一氧化碳和一氧化二氮的影响，产生了显著的偏移，例如，第 300 和第 453 两个通道变得非常嘈杂，也不用于同化。

对于 IASI，欧洲 NWP 中心通过 EUMETCAST 接收完整的 IASI 光谱，所以是在获得资料以后才选择通道。ECMWF 选择 366 个 IASI 通道，其中 300 个通道通过 GTS 参与世界气象组织的交换和发布^[31]。2007 年 Collard 对此进行了描绘。

选择 300 个通道的过程分为 3 个主要步骤：第一步：去掉基于经验标准的通道，以避免模式光谱中含有大量不确定性的光谱区域。第二步：主要的选择算法基于观测信息的内容，其中包含许多独立选择操作。第三步增加用于特殊目的的通道，这些通道不携带前面选择算法所选通道的信息。下面详细解释每一步的具体实现：

第一步为所有通道设置信号标识符，标识出通道内容，运行中根据这些标识来做出选择。首先去掉没有包含重要状态向量（CO、CH₄、和 N₂O）信息的通道。经过第一步，标识为 H₂O、O₃、太阳辐射或非局部热力平衡条件（non-LTE）效应等信息的通道也可能会在第二步的某些环节中被删除掉。

第二步包含了 Rodgers 提出的自动通道选择算法的几种运行方式。这些不同到的运行方式是为了保证，比如说，水汽通道（其温度 Jacobians 为峰值，对温度廓线也非常敏感，但是其雅克比值对假设的 H₂O 廓线同样非常敏感）不会被选择

用来提供温度信息。下面是通道选择算法的后续步骤：

- (1) 在 $15\mu\text{m}$ CO_2 带选择 30 个通道，只考虑其对温度信息的影响。
- (2) 接下来选择 36 个通道，其对流层雅克比峰值在 $15\mu\text{m}$ CO_2 的 707cm^{-1} - 760cm^{-1} 区域。根据 ECMWF 的经验得出，这个区域最能提高预报评分。
- (3) 同时考虑水汽和温度信息，并且包括了可用通道中 $6.3\mu\text{m}$ H_2O 带，选择接下来的 186 个通道。
- (4) 再选择 15 个通道仅提取臭氧信息。
- (5) 进一步添加 13 个通道用于提取被太阳光影响的信息。

第三步还包含了在窗口区域内额外附加 20 个通道，其弱吸收线用于云侦查和推导表层发射系数。这样一共选择了 300 个通道。再额外添加 66 个通道，分别为：与 HIRS 具有相似特征的通道(基于检测目的)；CNES 规定用于检测的通道； 707cm^{-1} - 760cm^{-1} 区域的 44 个通道，其峰值在低平流层产生^[32]。最后总共选择了 366 个通道。

选择的通道中只有一部分用于积极同化。图 2.6 显示了 ECMWF 业务同化系统 $15\mu\text{m}$ 带中 AIRS 和 IASI 仪器被同化通道的最大数目。直接通道选择方法比较简单，从完整光谱中选择出来，但是在执行选择时，是从剩下光谱中再选择出所需

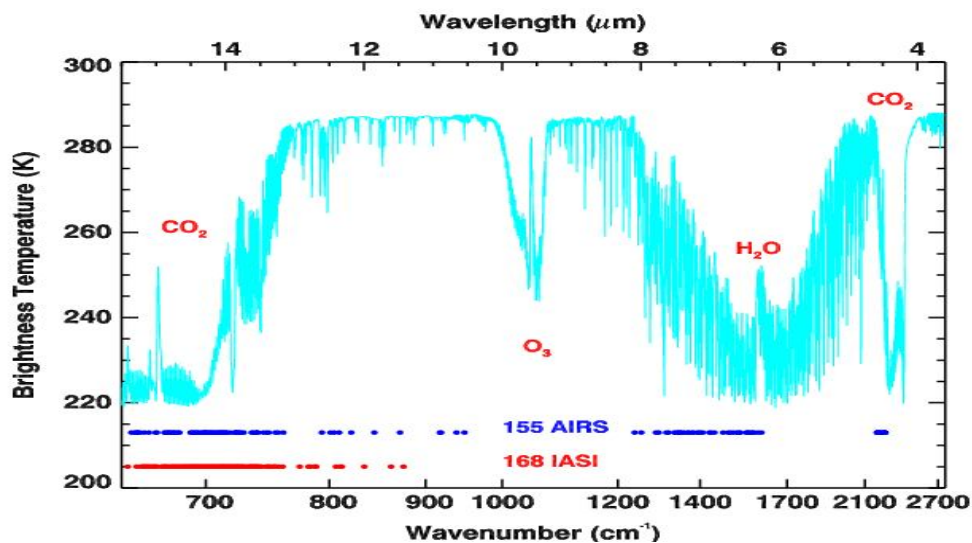


图 2.7 ECMWF 业务中同化的 AIRS 和 IASI 资料的最大通道数目

通道子集，这种选择操作的是一个比较费时的过程，另外对这些同化进行同化操作明显没有计算大多数未被选择通道所包含的信息^[33]。

2.4 本章小结

红外高光谱资料同化中往往需要分析许多变量指标，例如预报模式的第一猜

值构成了一组多变量指标，由于变量指标很多，增加了分析问题的复杂性。如何抓住主要的特点，用较少的指标代替原来较多的指标，又能综合反映原来较多指标的信息，主成分分析方法提供了有效的手段。

采用主成分的方法，从代表完整气象条件的光谱中，计算出主成分，并重构数目为几百个的光谱，将上千光谱通道信息压缩在这几百个光谱中，相当于滤去了完整光谱的噪声，且主成分之间的不相关特性，自然使各主成分之间信息不相关，从而消除了大量的信息冗余。主成分方法能使用红外高光谱资料的潜力，在大量减少使用的资料数量时仍然能保持大量有效信息。

第三章 基于主成分的辐射传输 PCRT 模式

大气辐射传输 (RT) 模式描述了卫星辐射率 (或亮温) 与温度、湿度等大气模式变量间的非线性关系, 是对任何种类卫星辐射率资料进行直接变分同化的前提和至关重要的部分。与应用卫星辐射率观测确定大气参数的反演过程相反, RT 模式输入预报模式大气状态廓线模拟卫星辐射率 (或亮温) 的计算过程一般称为正演^[34]。正演过程包括透过率计算、辐射率 (或亮温) 计算。目前有许多精确计算大气透过率的模式, 但这些模式一般基于逐线模式计算, 计算起来非常耗费时间, 在实际资料处理中不方便应用, 因此必须发展一套快速计算大气透过率的模式。为了满足这种要求, 本文在常规快速辐射传输模式的基础上应用主成分方法, 形成基于主成分的辐射传输 (PCRT) 模式。PCRT 模式对大量不同大气状态样本数据集计算逐线光谱资料库。从模拟辐射率的协方差矩阵特征向量中获得主成分谱线, 用作线性回归方案中的输入资料, 通过线性回归方案表示为依赖大气状态变量廓线预报因子的线性组合。预报因子构成了选定数目的多频辐射率, 这些辐射率由常规快速辐射传输模式计算得出。因此使用线性回归方案能够模拟主成分谱线, 从而实现为任意输入的大气廓线重构辐射率。

3.1 大气辐射传输特性

3.1.1 大气的选择吸收性

大气中的各个成分 (如 N_2 、 O_2 、 O_3 、 CO_2 和 H_2O) 对于太阳和地球辐射的削弱作用主要表现在两个方面, 一个是吸收, 一个是反射、散射。大气对辐射率的吸收具有选择性, 吸收的主要是紫外和红外部分, 对可见光的吸收较少, 如图 3.1

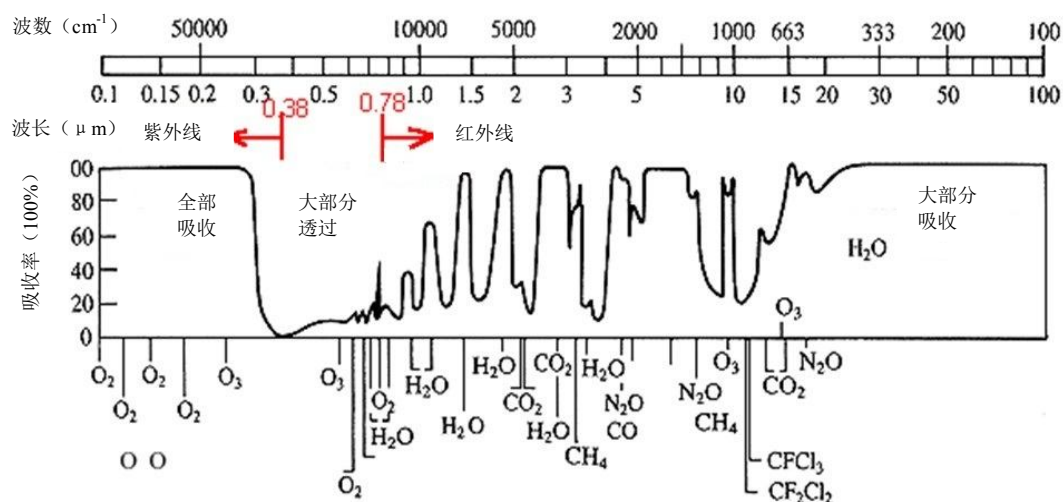


图 3.1 大气不同成分对辐射率的选择性吸收

所示。如：大气中的 CO_2 和 H_2O 会吸收太阳光中波长较长的红外线。这属于震动学知识。宏观物体都有自己震动的固有频率与波长，而且相对光波来说，这些波长较大。如果光波的频率越接近物体的固有频率，那么产生的共振现象越明显，光能转化也就越多，所以宏观物体吸收波长较长的红外波多一些，吸收短波较少。因此，大气的吸收能力与吸收物质及波长有关。吸收率为 0（即透过率为 1）的波段叫大气窗。吸收率为 1（即透过率为 0）的波段叫做大气吸收带^[21]。不同种类气体的吸收带是不同的。

3.1.2 大气辐射传输方程——卫星探测大气的理论基础

大气顶部所观测到的光谱辐射是由下面的辐射传输方程所决定的：

$$I(\nu) = B(\nu, T_s)\tau(\nu, p_s) + \int_{p_s}^0 B(\nu, T(p)) \frac{d\tau(\nu, p, w)}{dp} dp \quad (3.1)$$

其中, $I(\nu)$ 是卫星所测到的波数为 ν 处的辐射强度，等式右边第一项为地面放射的经大气吸收衰减后到达卫星的辐射，第二项为各层大气发出的经辐射衰减后到达卫星的辐射总和。 p 为气压， T 为温度， w 为吸收气体的混合比， τ 为波数为 ν 时从气压为 p 的一层大气到大气上界 ($p=0, \tau=1$) 的辐射透过率。在高度 z 上某一薄

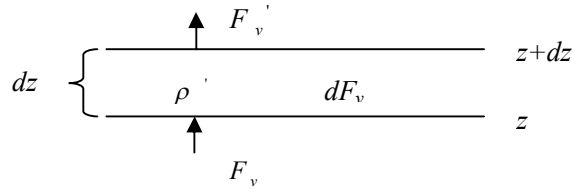


图 3.2 薄层气体 ρ' 吸收辐射示意图

层气体 dz 所吸收的辐射强度 dF_v 与射到 z 高度上的辐射强度 F_v 、薄层的厚度 dz 、吸收气体的密度 ρ' （图 3.2）以及状态 p ， T 有关，故

$$dF_v = -k(\nu, p, T)F_v\rho'dz$$

其中 k 为质量吸收系数。

改写方程 (3.1)，并从 z 到卫星高度 H 积分得

$$\int_z^H \frac{dF_v}{F_v} = - \int_z^H k(\nu, p, T)\rho'dz \quad (3.2)$$

由于 $\int_z^H \frac{dF_v}{F_v} = \ln \frac{F_{vH}}{F_{vz}}$ 代入式 (3.2) 得 $\ln \tau_v = - \int_z^H k(\nu, p, T)\rho'dz$

$$\text{即} \quad \tau_v = e^{- \int_z^H k(\nu, p, T)\rho'dz} \quad (3.3)$$

改用气压坐标，将

$$dz = -\frac{1}{\rho g} dp, \quad \frac{\rho'}{\rho} = w$$

代入式 (3.3) 得

$$\tau_v = e^{-\int_z^H k \rho' dz} = e^{-\int_p^0 \frac{\rho'}{\rho g} k dp} = e^{-\int_0^p \frac{w}{g} k(v, p, T) dp} \quad (3.4)$$

透过率 τ_v 与波数 ν 、气压 p 及吸收气体的密度廓线（即 w 的垂直分布）有关， CO_2 是十分均匀的气体，其混合比 w 近似于常数， $\tau_v \approx \tau_v(p)$ 。对于固定波数 ν ，透过 τ 就是气压 p 的函数。

为了推导辐射传输方程 (3.1)，可以认为卫星上所测得的辐射由以下两部分所组成：

- (1) 地面放射的经大气吸收衰减后到达卫星的辐射，记做

$$I_1 = B(\nu, T_s) \tau(\nu, P_s) \quad (3.5)$$

其中 T_s 为地面温度， P_s 为地面气压， B 为普朗克函数， $\tau(\nu, P_s)$ 为从地面 P_s 到卫星 ($p=0$) 的透过率。

- (2) 各层大气发出的辐射经衰减后到达卫星的辐射总和。对某一薄层大气向上经衰减后到达卫星的辐射为 $k(\nu, p, T) B(\nu, T) \tau(\nu, p) \rho' dz$ 。卫星测得各层大气所发射的辐射经衰减后到达卫星辐射的总和 I_2 为对 z 从 0 到 ∞ 的积分：

$$I_2 = \int_0^\infty k B \tau \rho' dz = - \int_{P_s}^0 k B \tau \frac{w}{g} dp \quad (3.6)$$

由式 (3.4) 对 p 求微商得：

$$\frac{d\tau_{\nu(p)}}{dp} = -\tau_{\nu}(p) \frac{w}{g} k \nu(p) \quad (3.7)$$

再代入式 (3.6) 即得

$$I_2 = \int_{P_s}^0 B(\nu, T(p)) \frac{d\tau}{dp} dp \quad (3.8)$$

因此

$$I(\nu) = I_1 + I_2 = B(\nu, T_s) \tau(\nu, p_s) + \int_{P_s}^0 B(\nu, T(p)) \frac{d\tau}{dp} dp,$$

即辐射传输方程 (3.1)。

由于方程 (3.1) 中的 $d\tau/dp$ 对 dp 气压层辐射贡献起到权重作用，故称为权重函数。各通道 (ν) 的权重函数都是随高度而变的，且不同的通道在不同的高度上有最大值，每个峰值即代表该波段向上辐射贡献的最大层次，故对某一波段而言，到达卫星的辐射主要来自于这一气压层（称为有效辐射层）。

3.2 多种辐射传输模式

由于红外高光谱遥感分辨率高，需要对不均匀的大气进行大量的辐射传输（RT）计算，所以如果使用传统的辐射传输模式将这些资料接入全数值天气预报资料同化系统中时将耗费较大的计算开销。目前，已经开发的 RT 模式为了满足计算开销上的约束，只选择使用红外高光谱仪器的部分通道子集，其它通道信息没有在同化系统中使用。

众所周知对不均匀大气进行逐线传输计算是一个非常耗时的过程。大气中的大量的气体分子在红外光谱区域具有大量的震荡-转动跃迁，或者单纯转动跃迁。分子线强度和形状是大气温度、气压和示踪气体垂直廓线的非线性函数。

常规模式将不均匀大气划分成大量薄层，然后对这些层使用合适的权重大气参数进行辐射传输计算。较高大气层的线宽度非常小，红外光谱区域的单色计算在 0.0001 至 0.0008 个波数间执行。例如，热辐射分光仪（TES）覆盖 650cm^{-1} 至 3050cm^{-1} 光谱区域，平均频率栅为 0.0004cm^{-1} ，因此在整个光谱范围内需要进行六百万次单色辐射传输计算。

多种方法可以实现高光谱资料处理过程中辐射传输计算时间最小化。TES 科学团队使用的方法：将吸收系数存储在庞大的上瞰表中作为单色频率栅位置的大气压和温度的函数。这样简单通过插值和相加能够计算特定大气层的光学厚度。这种方法避免了耗时的光谱线形和强度计算，但是正向模拟仍然必须对整个不均匀大气执行大量单色辐射传输计算，以获得大气顶部（TOA）辐射率。TES 科学算法利用 TES 仪器的高分辨率和高信息量的优势实现最小化计算代价，选择狭窄的微窗口用于反演特殊的示踪气体。然而这种方法可能是次优的，因为更多的通道不能用于求解整个所需的地球物理学参数^[35]。

另一种正向模拟方法通过有效快速参数化实现预报有效层的光学厚度。有效层光学厚度是通道平均值，它包含仪器线形函数（ILS）或者遥感响应函数（SRF）；因此每通道只需进行一次辐射传输计算。这种方法获得了有效的光学厚度，同时还获得不同大气层之间和不同大气气体之间所具有的光学厚度其它特征。下面的等式为计算层光学厚度的示例：

$$\begin{aligned}\tau_{Eff}^{Ch}(l) &= -\ln \frac{\int_{\Delta\nu} \phi(\nu-\nu') t^{Mono}(\nu, l) d\nu'}{\int_{\Delta\nu} \phi(\nu-\nu') t^{Mono}(\nu, l-1) d\nu'} \\ &= -\ln \frac{t^{Ch}(\nu, l)}{t^{Ch}(\nu, l-1)} = Func[\sec\theta, T_r, T_z(p, T_r, \dots)]\end{aligned}\quad (3.9)$$

其中 τ 是有效通道的层光学厚度， ϕ 是标准化 SRF， $\Delta\nu$ 是 SRF 的光谱范围， l 是大气层指标， t^{Mono} 和 t^{Ch} 分别为单色透过率和通道透过率（空间到层）。 T_r 是

层上的气压权重层温度比, T_z 是层平均温度对相关层温度之比, θ 是卫星天顶角, P 是气压。

目前针对卫星遥感应用开发了多种基于有效光学厚度的参数化快速模式。三种著名的快速模式为光学厚度透过率 (Optran) 模式, 卓越的辐射率传输算法 (ASRTA) 模式和 RTTOV 模式。这些方法已经成功的用于 HIRS、AMSU 和 AIRS 资料反演的业务算法中。

另外一类快速参数化方法通过使用一些代表性单色透过率或辐射率来预报通道透过率或辐射率, 例如相关 K 分布 (CKD) 模式、指数求和权拟合透过率 (ESFT) 模式, 辐射率采样模式 (RSM) 和最优光谱采样 (OSS) 模式。等式 3.10 和等式 3.11 表示了计算通道辐射率和透过率的方法。

$$R^{Ch}(\nu) = \int_{\Delta\nu} \phi(\nu - \nu') R^{Mono}(\nu') d\nu' = \sum_{i=1}^N w_i R_{\nu_i}^{Mono} \quad (3.10)$$

$$t^{Ch}(\nu) = \int_{\Delta\nu} \phi(\nu - \nu') t^{Mono}(\nu') d\nu' = \sum_{i=1}^N w_i t_{\nu_i}^{Mono} \quad (3.11)$$

等式中, $R^{Ch}(\nu)$ 和 $t^{Ch}(\nu)$ 分别代表中心频率 ν 的通道辐射率和透过率, w_i 是预先设定的单色辐射率或透过率。CKD 和 ESFT 都具有单色 (能够扩展包含多散射) 和非常快速 (每通道只需进行几个单色计算) 的优点, 都能为单一大气层和单一吸收气体执行计算。

然而, 这几种模式通常在单个大气层上训练, 然后通过假设垂直层之间的良好相关性和重叠气体不相关, 在之后的计算中引入通道透过率与气压、温度、气体含量的依赖关系, 再将整个计算扩展到给定精度的垂直不均匀大气中。学术界普遍将这种方法推广到遥感应用中。一种克服其缺陷的方法是在训练中同时包含层透过率和空间到层的透过率, 另一种方法是使用统计选择的 TOA 单色辐射率预报 TOA 通道辐射率 (即 RSM 方法)。RSM 具有对整个大气层一致性处理重叠气体和不均匀层的优点, 其缺点是需要大量近似 TOA 通道辐射率的单色辐射率, 以决定单色频率的位置。大气环境研究所 (AER) 开发的 OSS 方法克服了这方面的不足, 通过使用完整的 ESFT 最小化方案拟合 TOA 通道辐射率, 其特点是, 进行 1 至 15 个单色辐射传输计算以预报通道辐射率。OSS 同样能够预报层或者空间到层的透过率。Moncet 等将这种研究方法拓展成一个更完整的“蒙特卡罗”方法。OSS 快速模式计划用于处理跨轨红外遥感 (CrIS) 和未来 NPP/NPOESS 卫星上的先进技术微波遥感 (ATMS) 资料。

以上描述的所有快速模式都是基于通道的正向模式。这些模式使用大气温度和气体廓线的非线性函数, 或者使用简化的单色辐射率计算, 预报通道有效层的光学厚度、通道透过率或者通道辐射率。与之相反, 基于主成分的辐射传输 (PCRT) 模式不直接预报通道辐射率, 而是预报 PC 谱线, 相比通道数目具有更小的规模。

PCRT 模式利用主成分的正交特性将光谱信息（例如通道辐射率，透过率或者反射率）压缩成数目远小于通道数的少数重要的 PC 谱线。反过来，使用 PC 谱线通过线性组合可以计算通道辐射率。

3.3 具有主成分特色的 PCRT 模式

主成分方法通过检验多元资料集中包含的所有变量之间线性关系来实现问题维数的缩减。数据集的降维通过少数称为为主成分的不相关变量替换相关变量的原始序列。新获得的变量保存了原始资料中包含的大部分信息，这些变量能够有效描述资料。

红外高光谱中，存在大量性质相似的通道。例如，臭氧在 600cm^{-1} - 3000cm^{-1} 光谱范围内具有 100,000 多吸收线。大量的吸收线具有相同的洛伦兹或多普勒半宽，决定线密度和半宽的大气温度和压强非常相似。所有的单色辐射率是大气温度、湿度和示踪气体廓线的函数。独立信息块的数目远小于单色辐射率的数目，因此大量辐射传输计算是冗余的。基于主成分的辐射传输（PCRT）计算消除了这种冗余，将通道透过率或辐射率压缩成对应特征值递减顺序的主成分谱线，然后将这些主成分谱线用作线性回归方案中的输入资料，通过线性回归方案表示为依赖廓线预报因子的线性组合。预报因子由选定数目的多频辐射率构成，这些辐射率由标准的 ECMWF 的 RTTOV 快速辐射传输模式计算得出。因此使用线性回归方案能够模拟主成分谱线，从而为任意输入的大气廓线重构辐射率。

假设我们的数据集由 n 组观测组中的 m 个变量组成，构成 $n \times m$ 数据矩阵 X ，那么这些数据可以表示为向量形式 $\mathbf{x} = (x_1, x_2, \dots, x_m)^T$ （其中 T 表示转置）。如果 C 是数据矩阵 X 的 $m \times m$ 协方差矩阵，并且 A 是由行向量按特征值递减顺序排列的协方差矩阵特征向量形成的 $m \times m$ 矩阵，则向量的 PC， y 可写为：

$$\mathbf{y} = A\mathbf{x} \quad (3.12)$$

特征向量代表资料最大方差的方向，因此，每一个 PC 表示为由最大方差决定的变量的线性组合。PC 是正交的，因此，是线性无关的（这并不意味着统计学上的独立），并且与观测相对应的这些值称为 PC scores。

如果 λ_i 是对应第 i 个特征向量的特征值，那么值 $\lambda_i / \sum_{i=1}^m \lambda_i^2$ 给出第 i 个主成分表示的方差贡献率。由于定义矩阵 A 是正交的（因此非奇异），在此基础上，正交矩阵的逆等于它的转置，可以得出：

$$\mathbf{x} = A^T \mathbf{y} \quad (3.13)$$

用离散的形式描述为等式（3.12）和（3.13）：

$$y_{i,j} = \sum_{k=1}^m A_{i,k} x_{k,j} \quad (3.14)$$

$$x_{i,j} = \sum_{k=1}^m A_{i,k} y_{k,j} \quad (3.15)$$

其中, i 代表第 i 个值, j 代表第 j 个观测。在 PC 中, 使用前几个主成分 p , 小于 m , 代表数据中大部分方差。因此, 用前 p 个主成分代替 m 个变量, 实现降低维数的目的。在很多应用中, 维数的选择取决于主要 PC 所表示的总的方差, 它通常将依赖于资料的具体情况。

多元资料集包含了高分辨率的多色辐射率。对不同大气状态的采样 (即, 观测) 使用精确的逐线 (LBL) 模式计算获得这些多色辐射率。原则上, 可以将相应的逐线辐射率投影到从样本资料集的协方差矩阵获得的特征向量的固定偏差上, 实现为任意独立观测计算 PC 系数。然而, 这种算法包含了非常耗时的 LBL 辐射率计算, 辐射传输模式开发过程中最好不将它应用到准实时模拟 PC 谱线的业务运行计算中。在这种情况下, 需要开发更具计算效率的算法。同样的算法将考虑相对 LBL 模式所需的较小代价, 实现重构模拟的辐射率。

基于上面所讨论的, 开发基于主成分的辐射传输 PCRT 模式主要包含如下步骤: 生成 LBL 辐射率的样本资料库, 并且为任意不包含在样本资料库中的大气廓线开发预报精确 PC 谱线的算法。

3.3.1 PCRT 模式廓线集

廓线用于生成模拟的大气辐射率的多元资料集, 应该代表大气中温度、吸收量和表面参数的变化范围。为此, 本文使用 Chevallier 描述的廓线集中所包含的廓线。从大的廓线集中通过采样获得这些廓线, 由 ECMWF 预报系统 (cycle32R2) 于 2006 年 7 月到 2007 年 6 月期间运行的业务程序生成这个大的廓线集。除了温度、水汽和臭氧廓线以外, 廓线集还包含了地面特征的补充信息。

Chevallier 描述的廓线集包含了 12564 个廓线。正如 Chevallier 2002 年提出的, 通过随机采样能够实现廓线数目缩减达到 LBL 计算更容易处理的规模, 而这并不改变原始资料集中廓线的概率分布。这里随机采样任意 1000 条廓线, 另外再手动选择随机采样难以获得的部分廓线, 以覆盖资料集的极值。添加了这些额外的廓线以后, 最后总选定 1039 条廓线。快速模式的样本并不包含陆地表面观测, 所以只选择了海洋表面上的廓线。所有的廓线都插值到 Strow 定义的 101 层固定气压格点上, 并且用于执行 LBL 辐射率计算, 计算中采用大气顶层延伸至最接近表面气压值的气压格点。

除了温度和吸收量的垂直廓线以外, LBL 计算还要求表面温度、肤温、地面

气压和 10m 风速值。在我们选取的 1039 廓线集中，地面气压的值主要在 990hPa 和 1020hPa 之间变化，并且 940hPa 和 1040hPa 分别为最小和最大值。地面温度和肤温，在 273 和 303K 之间变化，较小的温度值 240K 主要分布在海洋表面上。在 LBL 计算的发射率计算过程中要求提供 10m 风速的值。我们选定的资料集中 10m 风速在 0ms^{-1} 的 40ms^{-1} 之间变化。然而，当估计的风速大于 15ms^{-1} 时，由于存在水面存在大量白色泡沫，导致 LBL 计算中的发射率计算模式可能失效，所以任意大于 20ms^{-1} 的值需要按比例缩小到小于 20ms^{-1} 。

最后，基于测试目的，我们选定了 1039 个相互独立廓线组成的较小资料库，用于训练 PCRT 模式。

3.3.2 PCRT 模式的 LBL 辐射率资料库

研究中采用美国大气和环境研究所 (ARE) 开发的 LBL 模式第 11.1 版 (LBLRTM_v_11.1) 计算所得的 LBL 辐射率资料库。计算中假设不考虑太阳辐射的海洋表面晴空情况。在覆盖 IASI 和 AIRS 5 个天顶角 (即具有相等正割间隔为 1 到 2 的角度) 观测的位置对每一个大气廓线计算 0.001cm^{-1} 分辨率的单色辐射率。这种选择对应总数为 5195 个观测。为了计算 LBL 计算中所需光谱发射率，我们修改了 Masuda 等 (1988) 模式，即使用 Wu 和 Smith (1995) 描述的方法引入了表面反射的发射效果，并且采用 Pinkley 和 Williamm (1976) 的盐分校正。

表 3.1 目前 LBL 计算中包含的氟氯化碳和卤代烷浓度

CFF-11 (ppbv)	CFC-12 (ppbv)	CCl_4 (ppbv)	CF_4 (ppbv)
246	540	74	90

LBL 计算包含了 17 种大气成分。只有 H_2O 和 O_3 可变。其它几种成分为常量，包括： CO_2 、 N_2O 、 CO 、 CH_4 、 NO_2 、 SO_2 、 NO 、 N_2 、 O_2 、 HNO_3 、 OCS 、 CCl_4 、 CF_4 、 CCl_3F (CFC-11) 和 CCl_2F_2 (CFC-12)。对于 CO_2 、 N_2O 、 CO 和 NO_2 采用 Matricardi 描述的气候廓线^[36]，而其它几种除了 HNO_3 和 NO 采用 MOZART 化学传输模式生成的廓线，其它则使用空军地球物理实验室 (AFGL) 大气成分廓线。 CCl_4 、 CF_4 、 CCl_3F (CFC-11) 和 CCl_2F_2 (CFC-12) 使用表 3.1 中罗列的值来反应天气体的浓度。

为了执行 LBL 计算，生成分子资料库，融合了从不同来源获得的线参数。这种方法基于 Matricardi 2007 年发布的结果，它表明在一定数目的光谱区域使用变化的线参数可能实现观测和模拟更好的一致性。混合资料库基于 AER 文件中的 LBLRTM_v11.1 包。这个文件主要来源于 HITRAN2004，并且包含了 2007 年 1 月 1 日更新的资料。然而，对于 1000cm^{-1} 和 1080cm^{-1} 之间的臭氧区域，使用 HITRAN_2000 的臭氧线参数， 1700cm^{-1} 和 2400cm^{-1} 之间的光谱区域，使用

GEISA_2003 中的水汽参数。

最后，我们对 IASI 和 AIRS 通道计算多色辐射率，使用当时的光谱响应函数的特性来引入多色辐射率。

3.3.3 PCRT 模式的辐射传输方程

PCRT 模式为选定的 1039 个廓线执行离线 LBL 计算生成多色辐射率资料集，对于不在选定廓线集中的廓线需要实时计算辐射率，扩充多元资料集。PCRT 模式采用近似的大气辐射传输方程：

$$L(\nu, \theta) = (1 - N)L^{Clr}(\nu, \theta) + NL^{Cld}(\nu, \theta) \quad (3.16)$$

其中 $L(\nu, \theta)$ 是波数为 ν ，天顶角为 θ 时大气顶部（TOA）向上的辐射， $L^{Clr}(\nu, \theta)$ 和 $L^{Cld}(\nu, \theta)$ 分别为晴空和云天大气顶的向上辐射率， N 为云量比例。

晴空 $L^{Clr}(\nu, \theta)$ 可写为：

$$L^{Clr}(\nu, \theta) = \tau_s(\nu, \theta)\varepsilon_s(\nu, \theta)B(\nu, T_s) + \int_{\tau_s}^1 B(\nu, T)d\tau + (1 - \varepsilon_s(\nu, \theta))\tau_s^2(\nu, \theta)\int_{\tau_s}^1 \frac{B(\nu, T)}{\tau^2}d\tau \quad (3.17)$$

上式右侧的一、二、三项分别为地表的出射辐射、地表反射的大气向下辐射和大气发射的辐射； T 是各层平均温度， T_s 是地表温度； $B(\nu, T)$ 是场温为 T 时的 Plank 函数， $\tau_s(\nu, \theta)$ 为地面至外空间的透过率， τ 为各层至外空间的透过率， $\varepsilon_s(\nu, \theta)$ 为地表发射率。

假设云顶发射率为 1 的云天辐射 $L^{Cld}(\nu, \theta)$ 定义如下：

$$L^{Cld}(\nu, \theta) = \tau_{Cld}(\nu, \theta)B(\nu, T_{Cld}) + \int_{\tau_{Cld}}^1 B(\nu, T)d\tau \quad (3.18)$$

其中， $\tau_{Cld}(\nu, \theta)$ 为云顶向外空间的透过率， T_{Cld} 为云顶温度。

若将大气辐射传输方程离散化，把大气顶至地面的大气分为若干薄层，并用谱带订正温度 T' 代替场温 T ，则卫星仪器第 i 通道的大气顶向上辐射为：

$$L_i = (1 - N)L_i^{Clr} + NL_i^{Cld} \quad (3.19)$$

其中晴空辐射 L_i^{Clr} 和云天辐射 L_i^{Cld} 分别为：

$$L_i^{Clr} = \tau_{i,s}\varepsilon_{i,s}B_i(T'_s) + \sum_{j=1}^{J_s} L_{i,j}^u + (1 - \varepsilon_{i,s})\left[\frac{\tau_{i,s}^2}{\tau_{i,j-1}\tau_{i,j}}\right] + L_i' \quad (3.20)$$

$$L_i^{Cld} = \tau_{i,Cld}B_i(T_{Cld}) + L_i'' + \sum_{j=1}^{J_{Cld}} L_{i,j}^u \quad (3.21)$$

这里的 $\tau_{i,j}$ 是模式层 j 至外空间在通道 i 谱响应区间内的透过率， L_i' 是地面至离地

最近的模式大气层 J_s 之间大气对 L_i^{Clr} 的最小贡献, L_i'' 是云顶至云顶以上最近的模式层 J_{Cld} 之间大气对 L_i^{Cld} 的贡献, 并定义

$$L_{i,j}'' = \frac{1}{2} [B_i(T_J') + B_l(T_{j-1}')](\tau_{i,j-1} - \tau_{i,j}) \quad (3.22)$$

3.3.4 PCRT 模式的插值

PCRT 模式使用内部插值方案允许在用户定义的层上输入廓线。用户定义的廓线插值到系数文件指定的层 (本文的例子中采用 101 层) 以执行光学厚度计算。为了确保 Jacobians 正确的映射回用户层上, 在用户指定的相同层上进行 RT 等式的积分, 将光学厚度插值回用户层。这种方式计算的预报因子变量与在训练过程中采用的预报因子变量不一致, 并且将在模拟的辐射率中产生误差。为了估计这些误差, 我们采用不相关数据集中的廓线以计算 IASI 预报因子变量。将 PCRT 光学厚度插值到大量 PCRT 系数中使用的 44 层气压格点, 以及专门为 IASI 设计的快速 RT 模式中的 90 层格点。

PCRT 辐射传输模式在每一次运行中, 对给定遥感的每一通道, 为输入的大气状态廓线计算大气顶部的辐射率或等价亮温 T^B 。在两个步骤中包含插值过程, 如图 3.3 所示, 一个是对廓线变量 X 进行插值, 一个是对光学厚度进行插值, 都采用分段权重积分插值算法。

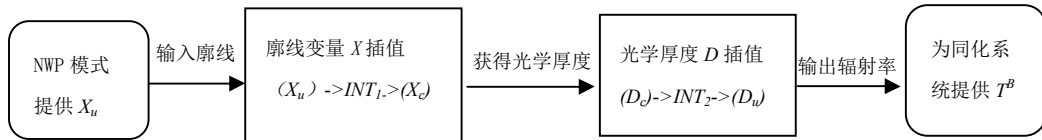


图 3.3 PCRT 模式的内部插值

其中括号表示矩阵, 为逐层的列向量, 这意味着 X_u 是一个标量, 下角标 u 和 c 分别表示用户层和系数文件定义的气压层。“INT”表示插值操作, 第二次插值即光学厚度到用户层的插值, 在所有的通道间循环进行。

这种插值方式能够灵活实现垂直坐标映射, 使用更平滑的插值, 改进光学厚度的计算, 用完整的预报模式温度廓线信息表示辐射传输积分中的源函数, 降低了接近大气顶层的不实际恒温的影响, 有效避免最近相邻对数线性插值算法中出现的“盲层”, 保持辐射率梯度的光滑结构特征^[37]。

3.3.5 主成分数目的确立

正如上文所描述的, 发展 PCRT 模式要求计算样本资料集协方差矩阵的特征向量, 并且为相同资料集中的每一个观测生成 PC 谱线。然后使用数目减少的主

成分 p ，重构给定精度的任意光谱。数目通常取决于观测的数目 n ，和资料集中的变量 m 。实际上， p 应该满足这样的条件，对于每一个通道 i ，辐射率重构误差 ε_i ，低于仪器噪声 η_i ：

$$\varepsilon_i = \sqrt{\frac{\sum_{j=1}^n (\sum_{k=1}^m A_{k,i} y_{k,j} - \sum_{k=1}^p A_{k,i} y_{k,i})^2}{n}} < \eta_i, \quad \forall i \quad (3.23)$$

注意在这节中辐射率用等价黑体亮温单位 (K) 表示。

试验中为不同的 p 值计算了辐射率的重构误差， ε_i 。图 3.3 显示了 IASI 随 PC 数目变化的辐射率重构误差。实线表示误差，点线表示仪器噪声。注意到与场景温度对应的平均辐射率，为每一个通道计算仪器噪声。在所有测试中，所有通道

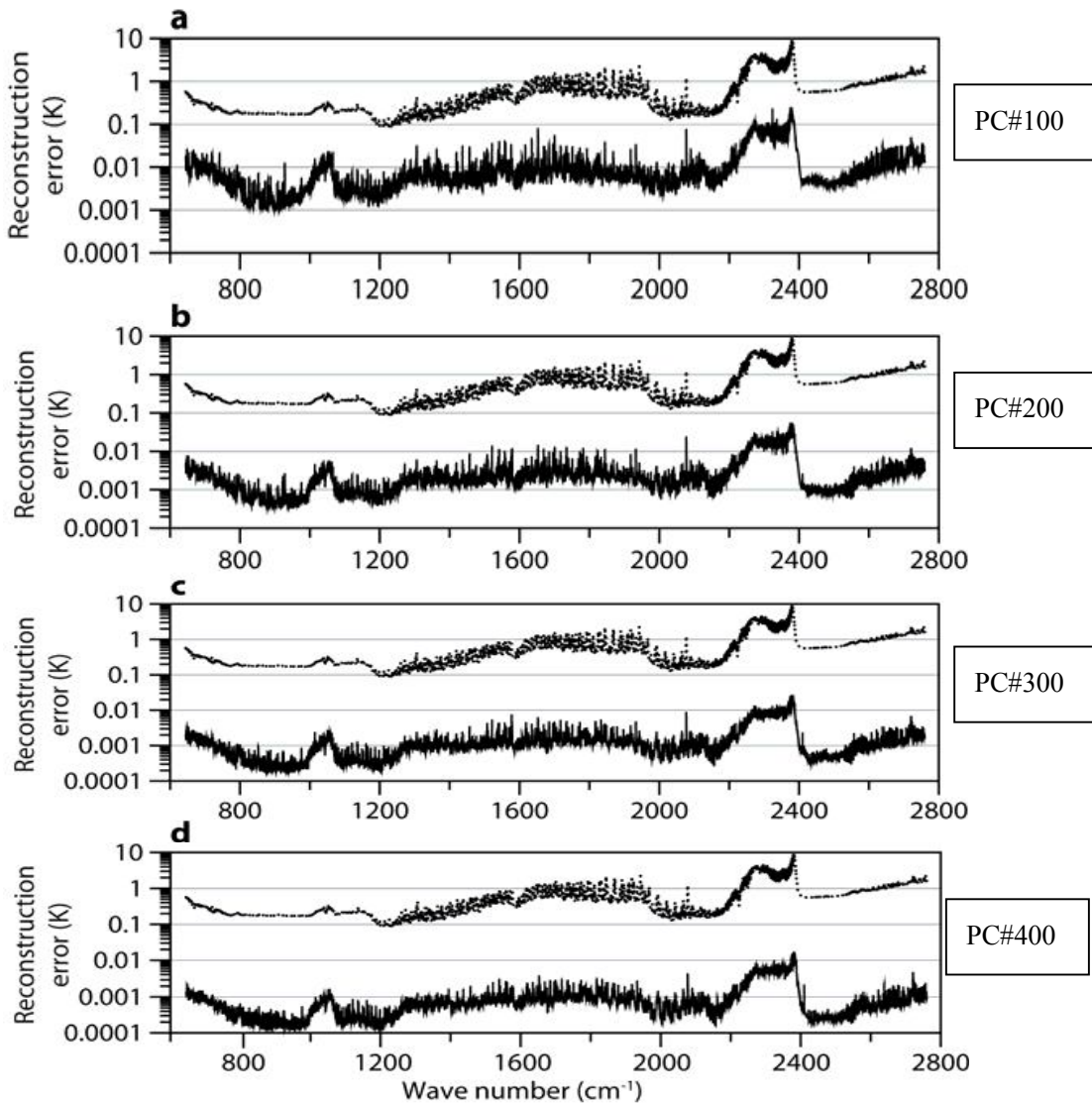


图 3.4 对比不同数目主成分的重构误差（实线）和仪器噪声（点线）

的重构误差都低于仪器噪声，并且随着主成分的数目增加而急剧下降。图 3.4 中的结果与样本集中的观测有关。重构独立集合中廓线的辐射率得到几乎同样的结

果。在 PCRT 模式中，主成分的数目是一个可调谐的参数。尽管通过试验表明使用少于 100 个主成分仍然满足等式 (3.23)，但是为了使整个 IASI 带 3 中重构误差在 LBL 模式误差中所占比率不会过大，因此规定使用的主成分数目不少于 100。

3.3.6 PCRT 模式的统计回归算法

在 NWP 中同化辐射率光谱的 PC 谱线要求为任意输入的廓线计算系数。这一节描述有效执行这个任务的算法。最终，从前面章节中描述的 LBL 辐射率的样本资料集中获得 PC 谱线值，用一个方程对这些值进行拟合。原则上，该方程能将主成分谱线和自变量（或者预报因子变量）联系起来。自变量包含了廓线相关的混合量，例如温度、吸收量。然而，第 3.3 节简要论述的主成分理论表明 PC 谱线和多色辐射率之间存在线性函数关系。很自然地，可以试着用一个多项式形式的方程来拟合，线性表示选定传感器的 m 个通道的多色辐射率。算法的效率由多色辐射率的快速计算决定。使用传统的快速辐射传输模式计算多色辐射率，虽然使用完全相同的 LBL 模式，但是分子资料库和垂直气压格点均按照前面 3.3.4 节中描述的设置来运行执行计算。前面描述 PCRT 模式的插值部分已经讨论了辐射率计算中使用不同气压格点的含义。这样我们的问题归结为下文定义的多元线性回归模式系数估计的计算：

$$y = \tilde{X}b \quad (3.24)$$

其中 y 是与变量相关的（即 LBL 光谱的主成分谱线）的 n 个观测的向量， \tilde{X} 是预报因子自变量的 n 乘 m 矩阵，而 b 是长度为 m 的待估未知系数向量。

在线性回归模式中，系数估计描述了被采样群体的特征，并且解释为群体正确特征的标准。然而，预报因子变量可能和其它预报因子变量或者变量的子集线性相关，这可能导致采样中估计的系数值与群体正确值明显不同。因为，很难从相同的群体中寻找一个采样来重复一个独立的采样。由此可见正确地估计系数非常重要，需要尽量减少预报因子变量之间的共线性相关效应。

3.3.7 预报因子

在 PCRT 模式的回归模式中，预报因子变量可以是 IASI 和 AIRS 光谱中的好几千个通道（分别是 8461 和 2378 个通道）。这些光谱含有许多具有非常相似光谱特征的通道。如上文所述，内容相似的通道在回归中可能产生问题，因为它们可能具有高度相关性。为了减少预报因子变量之间线性相关影响，需要对高度相关的通道进行聚类计算。本文采用 McMilin 和 Goldberg (1997) 提出的聚类方法。

聚类计算的步骤如下：

- (1) 计算样本廓线集通道辐射率的相关矩阵。
- (2) 寻找标准差最大的通道（主通道），以及在给定阈值上所有与主通道相关的通道。
- (3) 从选择的辐射率资料库中剔除第（2）步中获得的通道类别，并且重复这个过程直到不再发现更多的通道类别。
- (4) 当 $d < m$ 时结束，初始的 d_{init} 个最重要的通道减少为较小的 d_{final} 个通道，即 d_{final} 个回归方程中的预报因子。

3.3.8 预报因子相关阈值

最重要的通道数目和每一聚类中通道的数目及其与分布由规定相关程度的阈值决定。太高的阈值可能得到大量高度相关的主通道，而太低的阈值则可能道道太少的主通道。为了陈述这个问题，选定相关阈值的最优值，使用包含阈值选择的客观标准，最小拟合均方根（rms）误差定义为：

$$rms = \left(\frac{\sum_{i=1}^m \sum_{j=1}^n (R_{i,j}^{\text{LBL}} - R_{i,j}^{\text{PC}})^2}{mn} \right)^{1/2} \quad (3.25)$$

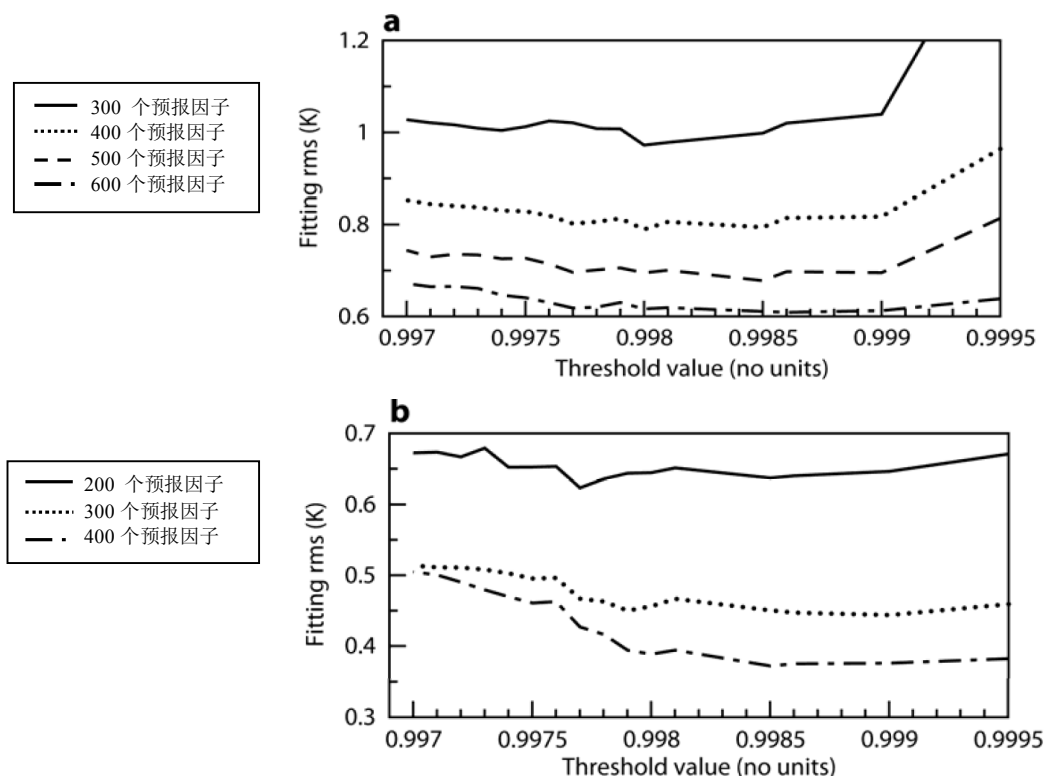


图 3.5 IASI (a) 和 AIRS (b) 的拟合 rms 误差随阈值与预报因子分布情况

其中 $R_{i,j}^{LBL}$ 是 LBL 辐射率, $R_{i,j}^{PC}$ 是 PCRT 模式辐射率。为了最小化辐射率重构误差的影响, 本文主要使用 400 个主成分算了 PCRT 模式辐射率。

对各种数目的预报因子变量和阈值评估了等式 (3.25) 的质量。尤其, 选择了 300、400、500、600 个 IASI 预报因子变量, 和 200、300、400 个 AIRS 预报因子变量, 阈值在 0.9972 和 0.9995 之间变化, 增量为 0.001。图 3.5 描绘了这些结果,

表 3.2 选择预报因子相关阈值的最优值

预报因子数	200	300	400	500	600
IASI 阈值	--	0.998	0.998	0.9985	0.9986
AIRS 阈值	0.9977	0.999	0.9985	--	--

显示出拟合误差随着阈值和预报因子数目变化。显然, 阈值的最优值 (即拟合误差曲线的最小值) 随着回归方案中预报因子变量数目变化。表 3.2 显示了本论文中使用的最优阈值。同样较容易看到一个事实, 拟合误差随着回归中预报因子变量数目下降。

3.3.9 PCRT 模式预报 PC 谱线

预报 PC 谱线是使用一些多色透过率或辐射率的非线性函数, 其优点为辐射传输方程将大气温度、湿度和示踪气体廓线转换为辐射率, 并且这些辐射率通过 PC 谱线线性表示为:

$$\bar{R}^{ch} = \sum_{i=1}^{N_{PC}} Y_i \bar{U}_i + \bar{\varepsilon} = \sum_{i=1}^{N_{PC}} \left(\sum_{j=1}^{N_{poly}} a_j R_j^{poly} \right) \bar{U}_i + \bar{\varepsilon} \quad (3.26)$$

其中, \bar{R}^{ch} 为通道光谱向量, \bar{U}_i 为 PC 向量, N_{PC} 为重要的 PC 数目。 Y_i 为 PC 谱线, 由多色辐射率 R^{poly} 的线性组合生成, a_j 为相应的权重。 $\bar{\varepsilon}$ 代表正向模式误差向量, 其值通常小于 0.04K。 PC 谱线 Y_i 是具有通道辐射率向量 R_j^{poly} 的 PC 向量 a_j 的点积。

$$Y_i = U_{N_{ch} \times 1}^T R_{N_{ch} \times 1}^{ch} = \sum_{j=1}^{N_{ch}} u(j,i) \times R^{ch}(j) \quad (3.27)$$

角标 i 表示第 i 个 PC 谱线, ch 表示通道指标, 上角标 T 表示转置, $u(j,i)$ 是 PC 矩阵 $U_{N_{ch} \times 1}^T$ 的元素。通道辐射率 $R_{N_{ch} \times 1}^{ch}$ 是通道频率范围内多色辐射率 $R^{ch}(j)$ 的线性组合。权重仅仅是多色频率栅位置的 ILS 或 SRF。预先计算的 PC 向量和 ILS (或 SRF) 都不从一个大气状态变化到另一个状态, 因此, PC 谱线是多色辐射率的线性函数 (等式 3.28)。

$$Y_i = \sum_{j=1}^{N_{ch}} u(j,i) \times \left[\sum_{k=1}^N \phi_k R^{poly}(k) \right] = \sum_{l=1}^{N_{poly}} a_l R^{poly}(l) \quad (3.28)$$

其中 $R^{poly}(k)$ 表示多色辐射率, ϕ 表示标准化 ILS (或 SFR), $u(j,i)$ 是权重。

3.4 PCRT 模式计算实例

模式输入的参数包括通道光谱、回归系数、不同大气气体吸收系数的特征向量。由于用于生成 PC 谱线的辐射传输计算为多频计算, 所以其辐射传输计算非常简单且直观。除了计算 PC 谱线和通道辐射率, PCRTM 同样能够提供分析的雅克比矩阵, 可做一维 (1-D) 物理反演或者三维或四维变分同化。

下面的例子显示了怎样计算晴空大气中的向上辐射率。首先, 由表面向上辐射率 (R_v^{up}):

$$R_v^{up} = \varepsilon_v B_v(T_s) \quad (3.29)$$

其中, ε_v 为频率为 ν 的表面发射率, B_v 为频率 ν 的假定量温时的普朗克函数。在这个例子中, 在表面肤温 T_s 情况下计算表面热辐射。遥感高度的向上辐射率成分、以及由其推导出来与单层温度相关 ($\frac{\partial R_v^{up}}{\partial T_l}$) 和单层示踪气体 (如 $\frac{\partial R_v^{up}}{\partial H_2O}$) 相关的

的导数, 如图 3.6 所示可以通过从底层向上的递归计算得到:

从底层到顶层范围内对单层 l 进行循环

$$\begin{aligned} \frac{\partial R_v^{up}}{\partial \tau_l^0} &= [B_v(T_l) - R_v^{up}] t_{nTop-1 \rightarrow l} \sec(\theta) \\ \frac{\partial R_v^{up}}{\partial T_l} &= \frac{\partial R_v^{up}}{\partial \tau_l^0} \frac{\partial \tau_l^0}{\partial T_l} + (1 - t_l) t_{nTop-1 \rightarrow l} \frac{\partial B_v(T_l)}{\partial T_l} \\ \frac{\partial R_v^{up}}{\partial H_2O_l} &= \frac{\partial R_v^{up}}{\partial \tau_l^0} \frac{\partial \tau_l^0}{\partial H_2O_l} \\ R_v^{up} &= R_v^{up} t_l + (1 - t_l) B_v(T_l) \end{aligned}$$

对下一层重复该计算

图 3.6 PCRT 模式逐层计算辐射率

其中, l 表示层的指标, t_l 和 $t_{nTop-1 \rightarrow l}$ 分别表示层透过率和顶层与当前层之间累积的透过率, θ 为遥感器天顶角。 $\frac{\partial \tau^0}{\partial T}$ 和 $\frac{\partial \tau^0}{\partial H_2O}$ 分别为由层光学厚度获得的与层温度

和层水汽含量有关的导数。 $\frac{\partial R_v^{up}}{\partial \tau_l^0}$ 为与层光学厚度有关的向上辐射率的导数。上

角标 0 表示在 0 天顶角 (即在天底方向) 位置计算的透过率。这两个量可以由上瞰表方法简单计算得到。计算完多色辐射率和相应的导数, 等式 3.28 可以用于预报 PC 谱线。如果需要通道辐射率, 可以用等式 3.26 将信息从 PC 范围转换到辐射率范围。

3.5 本章小结

基于主成分的 PCRT 模式从给定温度、水汽、臭氧、表面特性等变量的第一猜值，直接模拟主成分，实现高效准确的大气透过率计算。

PCRT 模式提高速度两个因素为：（1）PC 压缩后只需预报更少的预报量 Y_i ；（2）在选择多色频率范围内使用相关函数后，全部的预报因子数目减少。如果同化系统只使用 PC 谱线（几百个）而不是通道辐射率（几千个），第一个因素能够节省一个数量级的计算时间。第二个因素非常重要，因为 PCRT 只对几百个多色频率执行辐射传输计算，而基于通道的 RTMs 对几千个频率执行 RT 计算。

PCRT 模式的一个优点是不要求额外使用不同变迹函数用于模拟光谱。弱变迹光谱的 ILS 对每一个通道具有明显且较宽 ILS 光谱范围，基于这些通道的快速 RT 模式相比局部 ILS 更难处理。

由于主成分和回归方案中自变量之间存在高度的线性关系，使基于主成分的 PCRT 模式的逐线计算的辐射率具有更高的精度。有效使用基于主成分的快速辐射传输模式，能够促使充分利用主成分方法的降低噪声特征研究严重受仪器噪声影响的光谱区域，进而研究包含了这些红外高光谱区域信息的主成分谱线的四维变分资料同化效果^[38]。

第四章 PCRT 模式在资料同化系统中的实现

卫星资料的直接同化实际就是在观测算子中包含大气辐射传输 (RT) 正演模式将大气状态变量转化成辐射率, 然后采用变分等有效的数据同化方案同化“原始”形式或近“原始”形式的卫星资料的过程^[39]。实现卫星辐射率资料直接同化的可以归纳为三个要素: 1. 资料分析系统—变分技术; 2. 快速、精确的观测算子—快速辐射传输模式; 3. 资料。本章研究用主成分方法在卫星红外高光谱资料的直接同化中的应用, 则需要研究相应的几个方面: 四维变分资料同化 (4D-Var) 方法, 基于主成分辐射传输 PCRT 模式和经过主成分方法提取的红外高光谱资料信息。

4.1 四维变分资料同化简介

1969 年 Thompson 等指出, 必须要在整个时间序列的分析过程中保持变量的动力协调, 至此四维变分资料同化 (4D-Var) 的基本思想形成。四维变分资料同化简称四维变分同化, 它可以使用时间上分布的观测资料, 是在一个时间段上做同化, 其模式状态方程中包含了观测正演算子。4D-Var 有很多优点: 可同化多时刻的资料 (当前观测可对过去的分析结果产生影响); B 矩阵在同化窗口是隐式发展的, 可以考虑一点 flow-dependent; 可在代价函数中加上其它的弱约束项, 如 Courtier、Zou、Zupanski 等人, 在其文章中有描述^[40]。国内也有很多学者将 4D-Var 应用到实际天气分析中, 得到了一些有意义的结论。Awaji 等人将 4D-Var 用于海洋中。其不足之处在于, 目前为了减小代价虽然有了一些降阶和“自动伴随编译器”等简化办法, 但对于复杂模式而言, 由正模型的离散形式直接得到伴随方程的离散形式, 仍然是一个比较艰巨的任务; 另外, 在计算方面, 4D-Var 计算代价非常大。4D-Var 正是由于其高昂的计算代价目前没有被广泛推广到业务上, 但也得到了很大发展。随着研究的深入, 计算机水平的不断提高, 在不久的将来, 4D-Var 将成为资料同化方法的主流, 并被推广应用到实际业务中。

4.1.1 目标函数

四维变分中, 不同时间的观测资料 ($y_m^{\text{obs}} = y^{\text{obs}}(t_m)$, $m=0, 1, 2, \dots, M$) 可以同时影响初始时间的模式控制变量 ($x_0 = x(t_0)$)^[41]。四维变分在三维变分^[42]的观测算子 (H) 中加入数值天气预报模式 (M), 即

$$H(x_0) < \longrightarrow H(M(x_0)) \quad (4.1)$$

就可以得到四维变分的目标函数:

$$J(x_0) = J_b + J_o = \frac{1}{2}(x - x_b)^T B^{-1}(x - x_b) + \frac{1}{2} \sum_{m=0}^M (H_m(M_m(x_0)) - y_m^{obs})^T (O + F)^{-1} (H_m(M_m(x_0)) - y_m^{obs}) \quad (4.2)$$

适用于业务运行的增量形式的四维变分目标函数表示：

$$J(\delta x_0) = \frac{1}{2}(\delta x_0)^T B^{-1} \delta x_0 + \frac{1}{2} \sum_{m=0}^M (\tilde{y}_m - y_m^{obs})^T (O + F)^{-1} (\tilde{y}_m - y_m^{obs}) \quad (4.3)$$

$$\tilde{y} = H_m(M_m(x_b)) + H_m \delta x_m \quad (4.4)$$

$$\delta x_m = M_{m-1} \delta x_0 \quad (4.5)$$

式中，背景项 J_b 和观测项 J_o 分别表示模式分析相对背景场合观测资料的拟合程度。 $\delta x_0 = x_0 - x_b$ ， x_0 表示大气状态变量， x_b 表示大气状态的先验知识（称作背景场）， y^{obs} 是已知观测资料， H 为观测算子， B 为背景场 x_b 的误差协方差矩阵， O 为观测资料（ y^{obs} ）的误差协方差矩阵， F 为观测算子 H 的协方差矩阵。注意，黑体加粗的 $\mathbf{H} = \frac{\partial H(x)}{\partial x} \big|_{x=x_0}$ 是观测算子的切线性算子， H^T 为观测算子的伴随算子。

M_m 代表数值天气预报模式的切线模式或其它近似模式。常用的线性近似模式（ M_m ）可能比原预报模式（ M_m ）的分辨率低，或简化了物理过程参数化方案。

四维变分目标函数的梯度为

$$\nabla_{x_0} J = B^{-1}(x_0 - x_b) + \sum_{m=0}^M M_m^T H_m^T R^{-1} (H_m(M(x_m)) - y_m^{obs}) \quad (4.6)$$

式中， M_m^T 是数值预报模式的伴随算子（从 t_m 到 t_0 时的积分）。 $(O+F)^{-1}(H_m(x_m) - y_m^{obs})$ 是 t_m 时刻加到伴随模式变量上的强迫项， $\nabla_{x_0} J$ 是伴随模式积分的最终结果，即 \hat{x}_0 。

利用伴随模式计算四维变分目标函数梯度的方法最早是由 Le Dimet 和 Talagrand 引入到大气资料同化应用中的。Thépaut、Courtier 和 Navon 等先后发展了 ECMWF 和 NCEP 全球中期预报谱模式的伴随模式。正是有了计算梯度的这个有效方法，才使四维变分资料同化方法能够在现有的计算机上实现。

4.1.2 观测算子

本论文中研究卫星红外高光谱的直接四维变分同化，红外高光谱是卫星探测的辐射率资料，属于间接观测的非常规资料，需要观测算子插值模拟以后加入 4D-Var 系统进行计算。

观测算子，在观测的空间或者等价于模式的空间，将预报模式变量计算成等价于观测的量。如果观测资料是雷达探空资料，如风速、温度等，观测算子只需将预报模式得到的这些参量插值到相应的观测位置。如果观测资料是卫星探测的

辐射率,那么观测算子除了要进行预报模式空间到观测空间的插值,还需要对预报模式计算的风速、温度等参量进行复杂的辐射传输计算,将模式预报量转换成辐射率。辐射传输模式,针对观测为卫星辐射率资料,进行这种复杂的辐射传输计算。以前的辐射传输模式,只进行辐射传输计算,还不在其功能框架内包含插值的过程,不能等价于观测算子。现在研发的一些先进的快速辐射传输模式已经将这项功能实现在其结构框架内,包含插值和辐射传输计算两个部分,所以可以等价于观测算子。基于主成分的辐射传输模式 PCRT 模式,正如前文描述的,在进行辐射传输计算以外,也实现了更便捷有效的插值过程。

4.1.2 4D-Var 中的辐射传输模式

经过前面一节的描述,我们了解了观测算子和快速辐射传输模式的等价性,所以可以得出,我们主要需要关心 4D-Var 系统和辐射传输模式之间的关系,观测算子对整个 4D-Var 的作用和重要性。

辐射传输模式,表达了卫星辐射率(或亮温)与模式变量间复杂的非线性关系,可将相关的模式变量正演为卫星辐射率(或亮温),是卫星辐射率资料变分同化的关键环节^[43]。辐射传输模式有效地建立了预报模式变量和观测资料之间的联系。

同化辐射率资料过程中,大气状态信息在控制变量和观测辐射率之间转化关系如下:

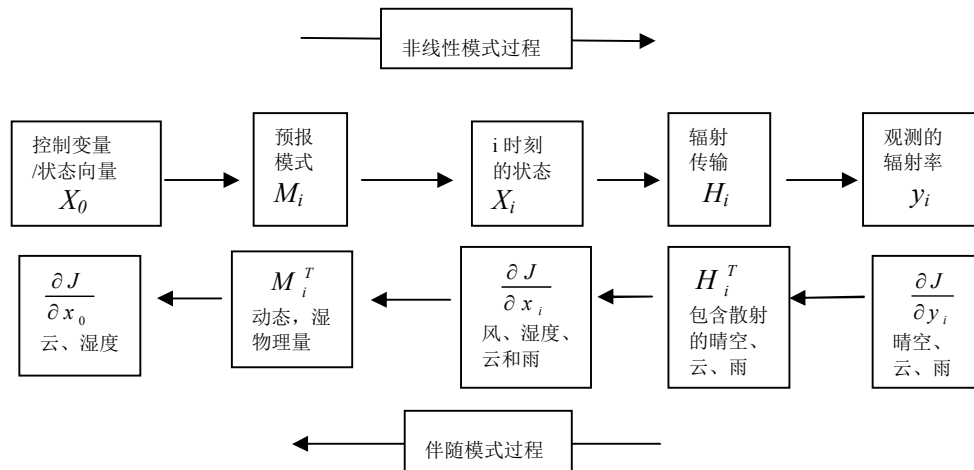


图 4.1 4D-Var 中大气状态信息在控制变量与观测之间的传递

观测辐射率通过辐射传输模式校正前一分析时刻的短期预报。在经典 12 小时四维变分同化循环中,ECMWF 采用 7,000,000 至 9,000,000 个观测校正 80,000,000 定义模式大气状态的变量,辐射传输模式,将对这些变量进行模拟。如此大的信息量,卫星辐射率观测算子在四维变分同化中占据十分重要的地位,且从计算量可以看出,其计算时间对整四维变分同化的时间代价影响也很大。

4.2 WRF 四维变分资料同化系统

4.2.1 WRF-DA 系统

美国国家大气研究中心 (NCAR) 已于 2011 年的 3 月 6 日推出天气研究与预报模式 (WRF) 的 3.3 版本。这一版本更新了其中变分同化系统 WRF-Var3.3, 简称 WRF 资料变分同化系统(WRFDA), 实现了四维变分资料同化 (4D-Var), 提供了同化卫星辐射资料的平台^[44]。图 4.2 描绘了 WRF 预报模式中 WRF-Var 系统:

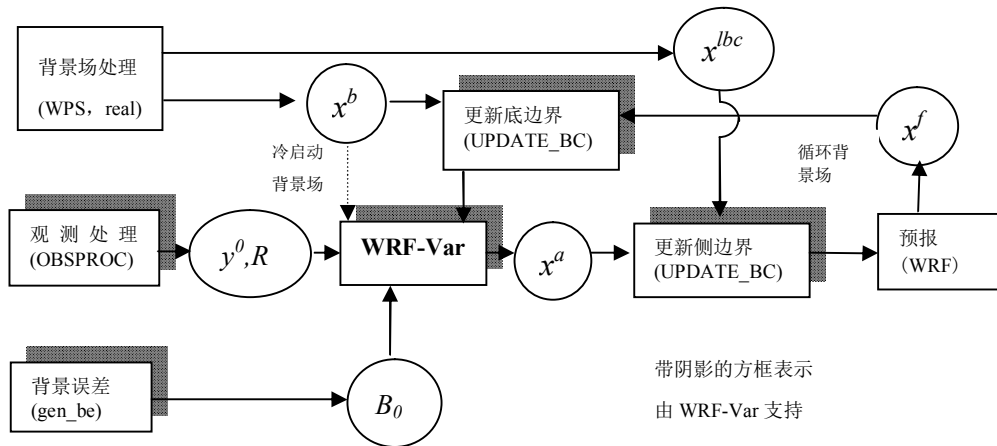


图 4.2 WRF 预报模式系统中的 WRF-Var

其中, x^b 表示从 WRF 预报或者 WPS 的 real 输出获得的第一猜值。 x^{bc} 表示由 WPS 的 real 输出的边界。 x^a 表示 WRFDA 获得的分析值。 x^f 表示 WRF 预报输出的值。 y^0 表示经过 OBSPROC 预处理的观测。 B^0 表示从 gen_be 获得的背景误差。 R 表示观测误差协方差^[45]。

4.2.2 设置 WRF 4D-Var 同化时间窗

WRF 4D-Var 同化时间窗为 6 小时。为了分析时刻 0h 准备观测文件, 那么 0h 到 6h 之间的观测都必须经过预处理, 而且如图 4.3 被划分为 7 个时间槽。注意图中, “Analysis time” 并非确切的分析时间 (0h), 它只表示为 WRF 4D-Var 变量列表文件 namelist 中设置的 time_analysis 变量。这个时间需要迟于确切分析时刻 3

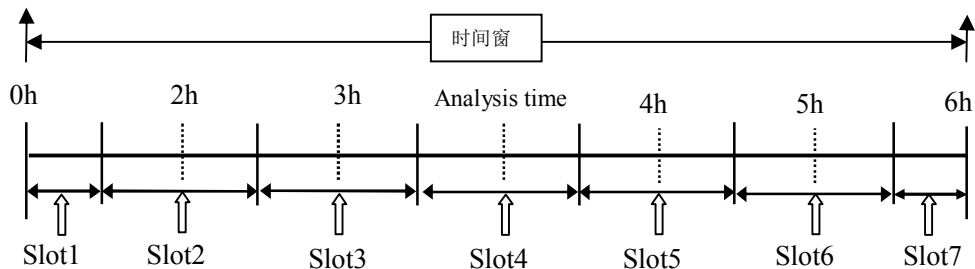


图 4.3 WRF 4D-Var 同化时间窗

个小时, 确切分析时刻仍然为 0h。这说明在设置这个变量时需要延迟 3 个小时

[46]。

当然，如果设置 `time_analysis` 为 4h，并且设置之后时间槽数目 `num_slots_past=4`，分析时刻之前时间槽个数 `time_slots_ahead=2`，那么得到的结果和图 4.3 一样。这两个变量只对 4D-Var 有效，用于规定时间槽的数目。例如，假如使用 2011 年 8 月 6 日的观测资料，真实分析时刻为 2011-08-06_12:00:00，我们需要对 OBSPROC 观测预处理变量列表的 `namelist` 做如下设置：

表 4.1 WRF 4D-Var 的同化时间窗变量设置

位置	变量名称	赋值	说明
&record1	obs_gts_filename	obs.2011080612	设置获得 2011 年 8 月 6 日 12 时的观测文件
&record2	time_window_min	2011-08-06 12:00:00	时间窗的起始时刻
	time_analysis	2011-08-06 15:00:00	设置文件需要的分析时刻
	time_window_max	2011-08-06 18:00:00	时间窗的结束时刻
&record9	use_for	4DVAR	指定 WRFDA 使用 4D-Var 同化资料
	num_slots_past	3	规定分析时刻之后 3 个时间槽
	time_slots_ahead	3	规定分析时刻之前 3 个时间槽

4.2.3 WRF 4D-Var 的输入文件

运行 WRF 4D-Var 需要 3 个输入文件：

- 1) 从 WPS/real（冷启动）输出的第一猜值和边界条件输入文件或者 WRF 预报（热启动）；
- 2) 观测资料数据文件(ASCII 格式，PREPBUFR 或者 BUFR 格式的辐射率)；
- 3) 背景误差估计数据文件（包含背景误差协方差）

表 4.2 WRF 4D-Var 输入文件

输入数据	格式	由以下文件生成
第一猜值	NETCDF	WRF 预处理系统（WPS）和 <code>real.exe</code> ，或者 WRF
观测	ASCII（也可能是 PREPBUFR）	观测处理（OBSPROC）
背景误差估计	二进制	WRFDA <code>gen_be</code> 程序/默认 CV3

目前，WRF 4D-Var 能够同化多种类型的资料，包括常规观测资料、卫星辐射率 BUFR 资料以及雷达资料，并且输入格式可以是 PREPBUFR 格式的资料或者 OBSPROC 处理的观测资料。

如果使用 PREPBUFR 格式的资料，需要在 `namelist.input` 输入文件的 `&wrfvar3` 中修改变量 `ob_format=1`，并且将资料链接为 `ob.bufr`。例如需要用 WRF 4D-Var 同化 12h 和 18h 的 PREPBUFR 资料，需要做如下链接：

```
> ln -fs $DATA_DIR/ob/2011080612/gds1.t12.prepbufr.nr ob01.bufr
> ln -fs $DATA_DIR/ob/2011080618/gds1.t18.prepbufr.nr ob02.bufr
```

4.2.4 同化红外高光谱的参数设置

首先需要链接执行文件，然后链接观测资料 and 第一猜值，背景场和 LANDUSE.TBL 等。运行红外高光谱资料最重要的是对 WRF 4D-Var 的变量列表输入文件 namelist.input 进行关键的设置，下面做简要的介绍：

1) 与 4D-Var 相关的设置：

&wrfvar1 中：var4d=true, var4d_lbc=true, var4d_bin=3600, ...。

2) 与辐射传输模式相关的设置：

直接同化辐射率的关键步骤是接入辐射传输模式，即 WRF 4D-Var 系统的卫星资料观测算子。WRF 4D-Var 中使用欧洲 EUMETSAT 开发的 RTTOV 模式或美国卫星资料同化联合中心 (JCSDA) 开发的 CRTM 模式，WRF 模式对这两个模式提供了方便快捷的用户接口。对此进行扩展，接入 PCRT 模式。在 namelist 中 RTM_OPTION，原本用 1 表示使用 RTTOV 模式，为默认值，2 表示使用 CRTM 模式，现在增设逻辑赋值 3 表示使用 PCRT 模式。这样 WRF 4D-Var 实现了这 3 种 RT 模式路径，或者不使用 RTM 模式（针对没有兴趣同化辐射率的用户）。这 3 种辐射传输模式，都能模拟大多数在轨卫星平台上不同仪器的观测辐射率资料。WRF 4D-Var 的一个重要的特征是，在对应于 namelist 设置的相应时刻动态的分配与辐射率同化相关的所有辐射率结构。需要在 &wrfvar14 中设置 4 个整型 namelist 参数在运行时刻指定同化的仪器：

表 4.3 RT 指定卫星仪器信息的 namelist 变量

变量	含义	说明
RTMINIT_NSENSOR	同化的遥感仪器的数目	例如总数赋值12，其中5个AMSUA；5个AMSUB；2个MHS；1个AIRS；1个SSMIS
RTMINIT_PLATFORM	同化的平台ID号数组	对应RTMINIT_NSENSOR的维度，1表示NOAA，9表示EOS，10表示METOP，2表示DMSP
RTMINIT_SATID	卫星ID号数组	与前面对应的仪器位于第几号卫星上，EOS平台的卫星范围为1-3
RTMINIT_SENSOR	传感器ID号数组	0,3,4,15,10,11分别表示HIRS, AMSUA, AMSUB, MHS, SSMIS, AIRS

例如同化 7 个卫星上的 14 个遥感仪器资料，namelist.input 作如下设置，且应严格按照这种写法赋值：

```
RTMINIT_NSENSOR = 14 # 6 AMSUA; 3 AMSUB; 3 MHS; 1 AIRS; 1 SSMIS
RTMINIT_PLATFORM = 1,1,1,1,9,10, 1, 1, 1, 1,1,10, 9, 2
RTMINIT_SATID = 15,16,18,19,2,2, 15,16,17, 18,19,2, 2, 16
RTMINIT_SENSOR = 3,3,3,3,3,3, 4, 4, 4, 15,15,15, 11, 10
```

WRF 4D-Var 能够同化所有遥感资料，前提是 PCRT 或者 CRTM 提供了相应的系数文件。另外，研究相应的资料接口、质量控制和偏差订正同样是正确同化这些辐射率资料的关键因素。

3) 下面以 AIRS 资料为例，WRF 4D-Var 部分相关变量的设置与说明如表 4.4

所示:

表 4.4 WRF 4D-Var 中部分 AIRS 变量设置

文件位置	变量名称	赋值	说明
&wrfvar4	use_airsretobs	true	读入相应的AIRS的BUFR文件
&wrfvar4	use_airsobs	true	从airs.bufr读入资料
&wrfvar14	rtminit_sensor	11	使用11号遥感器代表的AIRS资料
&wrfvar14	airs_warmest_fov	true	使用AIRS的窗口通道#914

4.3 PCRT 模式与 WRF 4D-Var 的对接与实现

4.3.1 PCRT 模式与 WRF 4D-Var 的对接

WRF 4D-Var 只支持两种辐射传输模式的接入，默认值 1 表示 RTTOV 辐射传输模式，2 表示 CRTM 的设置。在此进行扩展，用 3 表示接入 PCRT 辐射传输模式。现在将基于主成分的辐射传输 PCRT 模式安装在 WRF 4D-Var 接入辐射传输模式的对应位置，并对其接口做相应的修改，实现 WRF 4D-Var 变分同化系统与 PCRT 模式的对接。

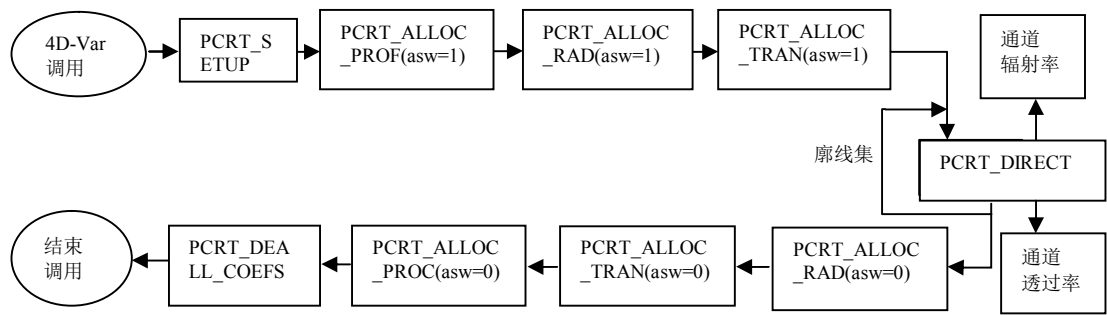


图 4.4 WRF 4D-Var 调用 PCRT 模式流程图

4.3.2 PCRT 模式的实现

4.3.2.1 重要参数设置

PCRT 模式使用多色辐射率计算主成分谱线，然后将这些主成分谱线用作线性回归方案的输入值，经过计算得到重构的辐射率。目前只能对晴空海表面上 AIRS

表 4.5 PCRT 模式预报因子参数设置

资料类型	变量	赋值	预报因子数目
AIRS	opts%addrefrac	1、2、3、4	300、400、500、600
IASI	opts%addrefrac	1、2、3	200、300、400

和 IASI 计算主成分谱线。执行时必须使用 PCRT 模式设置 opts%addrefrac 来选择预

报因子，可以进行如表 4.5 中的参数赋值设置预报因子。

使用 `opts%ipreg` 定义 PC 系数结构，从其中可以获得近似的通道列表，结构体 `pcrt_get_pc_predictindex` 获得输入的通道列表，输入的通道列表必须和预报因子列表一致，否则出错。

对于任何一种仪器类型，主成分的数目都可以在 1~400 之间变化，由 `pcrt_alloc_pccomp` 函数中的 `pccomp` 结构体定义，调用时，为所有廓线计算主成分。通过设置主成分数目和预报因子数目，调整模式的计算效率和计算精度。

PCRT 模式的要求额外的 `pccoef_metop_2_airs.dat` 和 `pccoef_eos_2_aisi.dat`。系数文件中包含了回归系数，用于预报主成分谱线和对大值为 400 的特征向量用于重构辐射率。为了计算多色频率预报因子，强制要求使用 LBLRTM 计算中相同的发射率和示踪气体廓线。系数文件中存储了固定不变的 CO_2 、 N_2O 、 CO 和 CH_4 廓线和 LBLRTM 发射率模式计算所需的回归系数。

通道廓线数组 `chanprof(:)` 用于声明计算的辐射率总数，数组里面的每一个元素用来表明通道指标和廓线指标，所有的通道必须按照廓线顺序排列，即先排列第一个廓线的所有通道，然后再排列第二个廓线的通道，如下表所示：

表 4.6 通道廓线数组

输入结构	AIRS（每次调用两个廓线）	IASI（每次调用两个结构）
Size (chanprof)	30	20
Size (profile)	2	2
Chanprof(:)%chan	1,2,...,15,1,2,...,15	1,2,...,10,1,2,...,10
Chanprof(:)%prof	1,1,...,1,2,2,...,2	1,1,...,1,2,2,...,2

4.3.2.2 PCRT 模式主程序介绍

基于主成分的 PCRT 模式中不仅包含了常规快速辐射传输模式的计算功能结构，最重要的是具有主成分特色的模块功能结构。这个结构包括两个部分：为一系列光谱集合生成主成分谱线的模块 `PCRT_SC`；使用由 `PCRT_SC` 生成的特征向量重构光谱辐射率的模块 `PCRT_RR`。

PCRT 模式的主函数 `PCRT_DIRECT` 为一个系数集合和通道廓线结构指定的廓线执行辐射传输计算，包含如下模块：

- (1) `pcrt_checkinput()`：用系数文件控制输入廓线；
- (2) `pcrt_setgeometry()`：从输入廓线计算辅助参数，如几何角度；
- (3) `pcrt_intavg_prof()`：将用户输入的廓线向 PCRT 层插值；
- (4) `pcrt_setpredictors()`：计算光学厚度和 PC 谱线预报因子；
- (5) `pcrt_opdep()`：计算光学厚度；
- (6) `pcrt_tran()`：计算透过率；
- (7) `pcrt_pc()`：计算 PC 谱线和重构的辐射率，其中包含了计算 pc 谱线函数

pcrt_PCcores(), 计算重构的辐射率函数 pcrt_reconstruct(), 重构辐射率转换成亮温的函数 pcrt_calcbt_pc();

(8) pcrt_emissivity(): 计算表面发射率;

(9) pcrt_intavg_chan(): 将光学厚度从 pcrt 系数层向目标层插值;

(10) pcrt_integrate(): 对辐射传输等式进行积分。

PCRT 模式的正向模式调用树如下图所示:

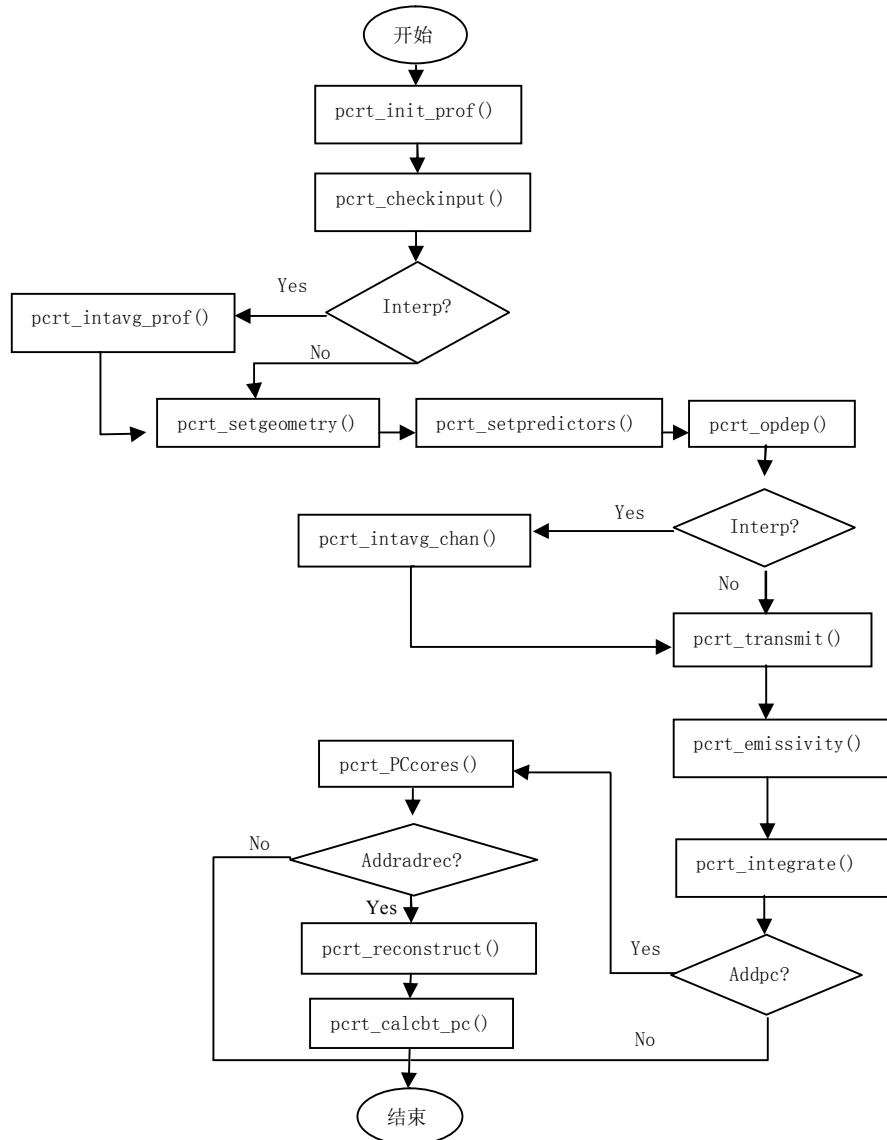


图 4.5 PCRT 模式函数调用树

PCRT 模式可以通过设置 opt%addpc 变量选择是否进行光谱主成分计算, false 表示直接输出模拟的完整光谱辐射率, true 表示进行主成分计算。当执行主成分计算时, 通过设置 opt%addradrec 变量选择是否直接输出主成分谱线, false 表示直接输出主成分谱线, true 表示继续使用主成分谱线重构辐射率, 最后再计算等价亮温。

4.3.2.3 pcrt_pc 模块介绍

PCRT 模式实现主成分计算思想的最关键模块^[47]是 `pcrt_pc`。这个模块从模拟的辐射率的协方差矩阵的特征向量获得 PC 谱线，然后将 PC 谱线输入到线性回归方案中，实现重构辐射率。其中包含了两个关键的部分 `pcrt_PCcores()` 和 `pcrt_reconstruct()`，分别实现了计算主成分谱线和重构辐射率。

下面以 `pcrt_PCcores()` 为例，描述其计算步骤如下：

Algorithm `pcrt_PCcores()`

输入: `nchannels`, `nPCcores`, `nprofiles`, `pccomp%PCcores`;

输出: `pccomp%PCcores(j)`

`pccomp%PCcores = 0._jprb`

`k = 1`

DO `prof = 1, nprofiles`

DO `i = 1, nchannels/nprofiles`

`chan = chanprof(i)%chan`

DO `j = 1 + (prof-1) * nPCcores/nprofiles, prof * nPCcores/nprofiles`

`chan_pc = chanprof_pc(j)%chan`

`pccomp%PCcores(j) = pccomp%PCcores(j) + &`

`&coef_pccomp%pcreg(opts%ipcreg)%coefficients(i, chan_pc) *`

`&radiancedata%clear(k) / coef_pccomp%noise(chan)`

ENDDO

`k = k + 1`

ENDDO

ENDDO

图 4.6 计算主成分算法 `pcrt_PCcores`

4.4 本章小结

这一章介绍了四维变分资料同化的框架结构，以及在 WRF 4D-Var 四维变分同化中实现基于主成分的辐射传输 PCRT 模式的方法。WRF 4D-Var 是一个先进而又庞大的系统，需要对众多参数作调整和设置才能实现接入基于主成分的辐射传输 PCRT 模式，并且通过修改辐射传输模式接口才能实现对 PRTM 模式的调用。对应第三章 PCRT 模式的计算流程，开发 PCRT 模式的软件模块，并重点设计其具有主成分计算的模块。PCRT 模式能够快速有效的将大气状态变量模拟成对应于观测的辐射率，而且其主成分的计算方式能够实现快速模拟，降低噪声的辐射率，方便四维变分同化系统利用大气状态信息调协控制变量。

第五章 主成分方法同化红外高光谱试验

主成分理论上能够消除红外高光谱中的大量冗余，有效提取整个红外高光谱资料中的信息^[48]。基于主成分的辐射传输 PCRT 模式在常规快速辐射传输模式的基础上加入主成分的计算思想，实现将大气状态变量模拟主成分谱线，或者再由主成分谱线重构出类似于通道特征的辐射率，为四维变分同化提供了很好的辐射率模拟值。通过同化试验可以验证主成分方法对研究红外高光谱资料方式带来的新突破。主成分的降噪特性，为我们提供了调查研究部分目前尚且不能使用的红外高光谱通道的机会^[49]。文章描述了多次同化试验的结果，验证主成分方法同化红外高光谱资料的效果，以及基于主成分的辐射传输 PCRT 模式的快速模拟红外高光谱主成分辐射率的特性。

5.1 基于主成分的同化策略

进行基于主成分方法的红外高光谱同化试验有 3 种策略：

- 1) 直接同化主成分谱线；
- 2) 将重构的辐射率代替真实观测用于同化；
- 3) 直接同化重构的辐射率。

第一种，直接同化主成分谱线。这种方案在主成分计算中需要正确选择光谱子带，才有可能降低 Jacobians 的非局部程度。在同化中需要采用具有主成分计算的辐射传输模式。由于主成分的 Jacobians 比原始通道更非局部化，更难分离垂直方向温度与湿度之间的信号。另外，只能使用云或地面上的通道，更难实现云检测方案，所以仅限于做晴空条件的同化试验。

第二种，对观测光谱进行主成分计算后再重构，将重构的辐射率作为“降噪”后的观测用于同化。这种方法由于主成分不能表示部分真实光谱的信号，将可能引入额外的误差。辐射传输模式可以不需要实现主成分计算方法，仍然用于计算模拟的辐射率。在误差特征不明显展开的情况下，可以使用资料提供者所提供的主成分。

第三种，直接同化重构的辐射率，即同化正向模拟重构的辐射率。这种方法需要基于主成分的辐射传输模式直接模拟重构的辐射率，获得的每一个重构的辐射率是原始光谱的线性组合。这种方法在对观测进行计算和辐射传输模拟计算时，都相应只使用部分重要的主成分，所以不会引入额外的误差。

基于以上的分析，我们采取第三种同化策略。在 WRF 4D-Var 系统中采用基于主成分的辐射传输 PCRT 模式，直接模拟重构的辐射率。

5.2 误差分析

如果观测向量误差为 ε_y ，主成分振幅的相应误差为：

$$\varepsilon_p = L^T \varepsilon_y \quad (5.1)$$

导致主成分空间中仪器误差协方差为

$$\text{Cov}(p) = E[\varepsilon_p \varepsilon_p^T] = L^T E L \quad (5.2)$$

其中 E 是原始观测的误差协方差。我们采用噪声标准化的观测， E 应该为单位矩阵，结果

$$\text{Cov}(p) = L^T L = I \quad (5.3)$$

然而，实际中用于正规化的噪声矩阵可能不同于从偏离单位矩阵 I 的 $\text{Cov}(p)$ 中获得的正确的矩阵，并且通常具有非对角元素。另外，误差预算的条件与仪器噪声无关，例如正向模式误差（在这种情况下计算 p 值的误差），将导致总的噪声与上面理想表达式的额外偏差。

应该注意到，就下面讨论的重构辐射率而言，正向模式误差没有通过等式 (5.3) 传播，并且应该 $\text{Cov}(p)$ 的附加条件。

重构辐射率的观测误差能够通过原始观测误差获得

$$\varepsilon_{\bar{y}} = L_m L_n^T \varepsilon_y + \varepsilon_R \quad (5.4)$$

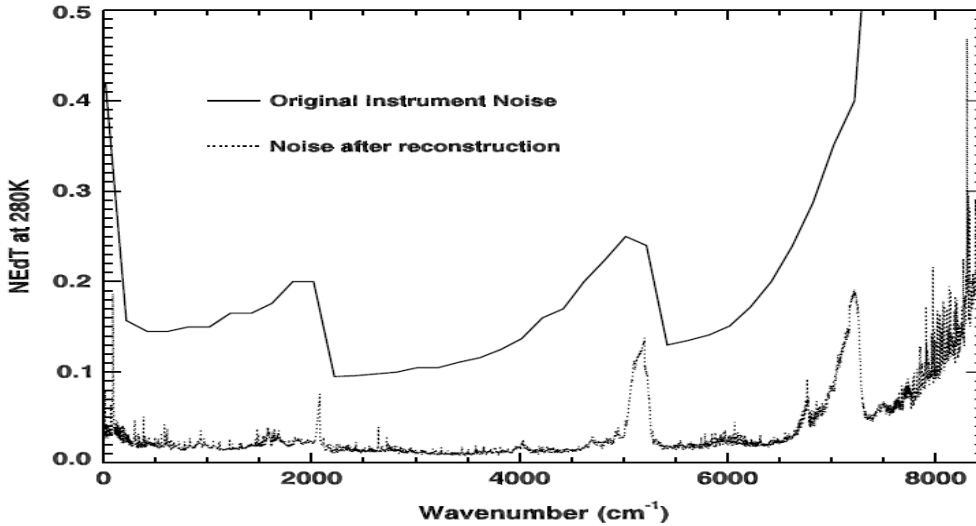


图 5.1 实际观测的仪器噪声和 200 个 IASI 主成分重构的辐射率噪声

通常，我们希望将重构辐射率当做真实辐射率的估值处理，并且去掉了没有被使用的主成分所代表的噪声。额外的误差项 ε_R ，是引入的重构真实（无误差）光谱的误差（即这种误差项来自于未使用成分中包含的真实大气信号）。如果明确使用正向模式计算重构的辐射率， ε_R 是 0（当然尽管仍然保留了正向模式误差）。如果假设重构辐射率是原始辐射率‘噪声净化’的替代，这一项将是非 0

的，尽管仔细选择主成分的数目和用于产生主成分的样本集，这一项应该低于等式最开始项的水平。

因此

$$\text{Cov}(\tilde{y}) = E[\varepsilon_{\tilde{y}} \varepsilon_{\tilde{y}}^T] = L_m L_n^T E L_n L_m^T + F_R \quad (5.5)$$

其中 E 是原始观测误差协方差矩阵（在噪声标准化辐射率单元）并且 F_R 是重构误差。同样，如果辐射率被真实噪声标准化， $E=I$ 。

由图 5.1 可以看出重构辐射率的噪声水平明显低于观测的仪器噪声水平。如果用通道误差相关矩阵图表示重构辐射率的通道相关误差，那么代表仪器噪声的对角线将明显简化，但是由于重构的通道是由整个光谱获得的，这将引入重构的通道误差相关^[50]。由图 5.2 可以分析这种现象，左图为真实辐射率 324 个通道子集的

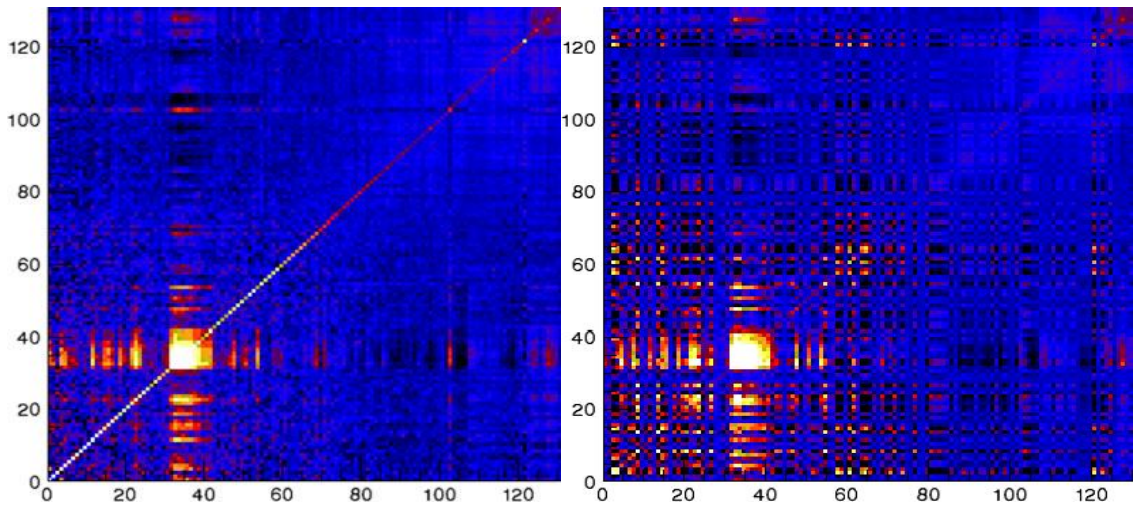


图 5.2 AIRS 真实辐射率和重构辐射率误差协方差矩阵

前 120 个通道的背景误差协方差矩阵，右图为相同投影向量的重构辐射率误差协方差矩阵。从图中可以看到代表仪器噪声的对角噪声明显下降，但是增加了部分相似通道的相关噪声。

5.3 同化结果分析

由模拟的 AIRS 辐射率（从观测资料的完整集合中获得，即晴空和有云场景）协方差中最开始的 200 个主成分计算得到重构辐射率。从 2378 个通道的 1688 个通道中获得的主成分，剔除了大量噪声‘间歇振荡’的通道和 $4.3\mu\text{m}$ 附近高度受白天非-LTE 影响的通道。所有资料都携带质量控制标识，它表明原始光谱是否重构到噪声范围内。

用 2011 年 7 月 1 日至 2011 年 8 月 10 日期间重构辐射率和真实辐射率做南半

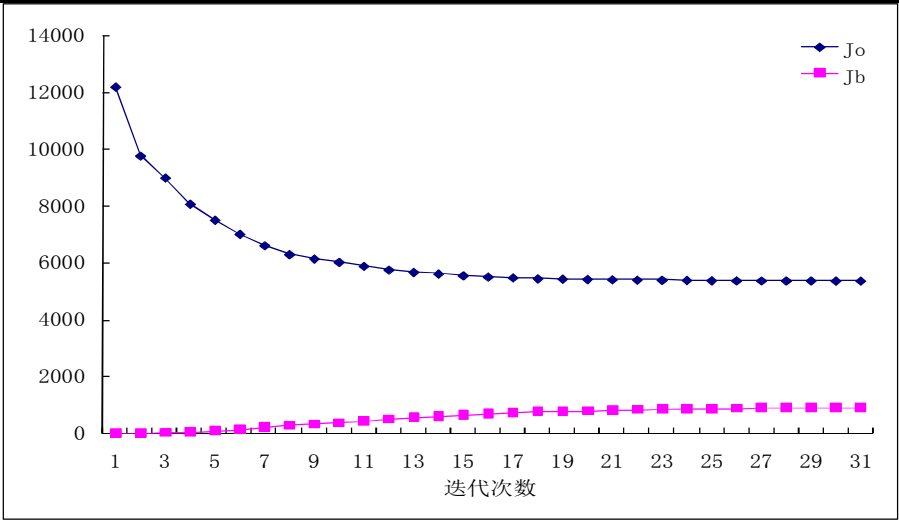


图 5.3 4D-Var 代价函数的 J_o 和 J_b 迭代图

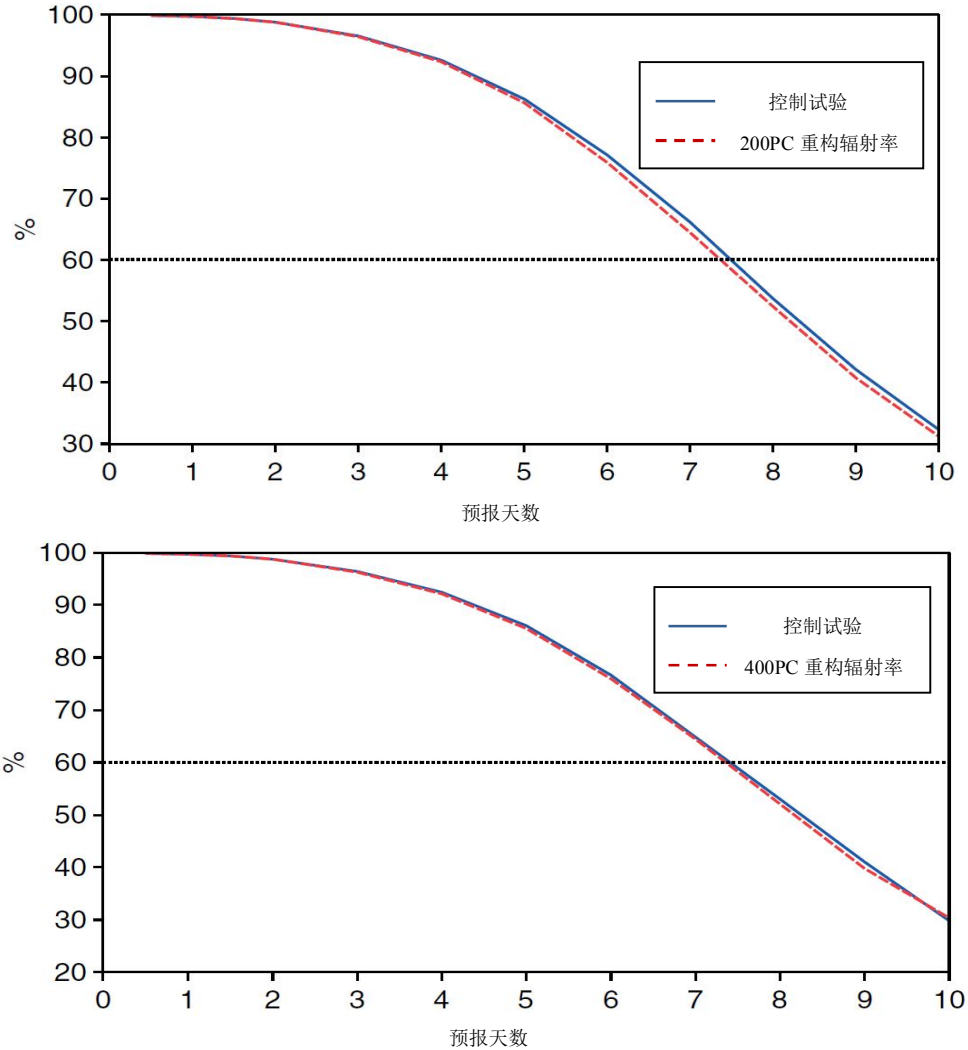


图 5.4 使用 200 个 (上图) 与 400 个 (下图) 主成分理论效果

球 500hPa 位势高度预报的反相关评分。图 5.3 显示了四维变分资料同化的代价函数随迭代次数递归下降情况。用重构的辐射率代替真实辐射率存在小幅度下降。

图 5.4 分别显示了理论上用 200 个和 400 个主成分重构辐射率（虚线）和控制实验（实线）对 500hpa 位势高度预报反相关评分产生的影响。预期实验中用 400 个主成分重构的辐射率相比 200 个主成分重构辐射率预报评分更接近控制试验，预报影响更积极。整体上，重构辐射率对南半球预报评分的影响是中性的。这些影响与真实辐射率同化差异不大，并且在南半球使用重构辐射率相比完全没有使用 AIRS 资料的影响仍是非常积极的（对比第一章图 1.1）。

5.4 PCRT 模式计算效率

如果我们固定预报因子变量的数目，对于不同的主成分数目可以估计出 PCRT 模式计算效率。将 RTTOV 模式与 PCRT 模式计算时间之比命名为计算效率的增益 g 。对于标准的 RTTOV 模式，大量的辐射率计算包含了层光学厚度的计算。单个通道光学厚度的计算要求大小等于 $l \cdot s \cdot pr$ 的线性运算，其中 l 是垂直的层数， s 是吸收种类的数目， pr 是每层的预报因子数目。通常， $l=100$ ， $s=7$ ， $pr=12$ ，因此 c 个通道的光学厚度的计算要求不少于 $o=c \cdot 8400$ 次运算。

对于 PCRT 模式， c 个通道的辐射率模拟包括自变量值的计算（ $d \cdot 8461$ 次运算）和辐射率重构 $p \cdot (c+d)$ 次运算，总共 $opc = d \cdot 8461 + p \cdot (c+d)$ 次运算。我们假设用 opc 除 o 得到 g 的保守估计为：

$$g \cong \frac{c}{d + p \cdot (c+d) / 8400} \quad (5.6)$$

可以直接计算选择 300、400、500 或 600 个预报因子时 $c=8461$ （即整个光谱）的 g 的值。计算效率的增益在 5 到 21 之间变化。使用等式（5.6），发现 IASI 辐射率重构中使用 200 个主成分时导致计算效率至少有 10 倍的增益。对于 AIRS，发现使用 100 个主成分得到至少 5 倍的增益。表 5.1 显示了对 IASI 执行计算时，设置不同预报因子与主成分数目获得的 PCRT 模式相对 RTTOV 模式的计算增益。

表 5.1 PCRT 模式相比 RTTOV 模式的计算增益

	300 预报因子	400 预报因子	500 预报因子	600 预报因子
100PC	20.92	16.74	13.95	11.95
200PC	16.64	13.85	11.86	10.37
300PC	13.81	11.81	10.32	9.16
400PC	11.80	10.29	9.13	8.20

由表中可以看到，在保证预报评分降幅较小的情况下，PCRT 模式的计算增益比较乐观，对 IASI 光谱采用 300 个预报因子，计算 100 个主成分，获得了 20.92 的计算增益，很大程度上比提高了计算效率。

实际试验中在 3.4Ghz 4 核处理器上用 PCRT 模式模拟 IASI 光谱 200 个主成

分，以及用 RTTOV 模拟 IASI 的 366 个通道辐射率需要的时间如下：

表 5.2 PCRT 模式与 RTTOV 模式的试验结果

	PCRT 模拟 200 个主成分	RTTOV 模拟 366 个通道
模拟时间	5 秒	1 分 19 秒

结果证明了 PCRT 辐射传输模式确实比 RTTOV 模式在模拟红外高光谱资料时，能获得较大的计算增益，达到了 15.8，能够为资料同化系统更快速地提供模拟辐射率资料，减小了辐射传输模式给资料同化系统带来的计算开销约束。

5.3 本章小结

将基于主成分的辐射传输 PCRT 模式在 WRF 4D-Var 资料同化系统中实现，进行多组主成分方法同化红外高光谱资料的试验。对比 PCRT 模式与 RTTOV 模式的资料同化系统，得出 PCRT 模式模拟的辐射率降低仪器噪声的性能优越，这方便对高度受仪器噪声影响的通道进行研究，但是由于每个主成分重构的辐射率都带表完整光谱的部分信息，所以又引入了通道间的相关。将主成分方法应用到资料同化红外高光谱系统中，影响是中性的。虽然相比目前业务中使用直接经过通道选取的红外高光谱资料做南半球 500hPa 位势高度预报^[51]，主成分方法的影响略有下降，但是相比完全不使用红外高光谱资料的业务预报明显属于正效应。在保证基本的同化质量条件下，分别采用不同数目的预报因子获得不同数目的主成分计算开销较小，相比 RTTOV 模式，PCRT 模式能获得 5 至 21 倍的计算增益，提高了计算效率，减少了辐射传输模式给同化系统带来的计算开销约束。

第六章 结论与展望

6.1 总结

本文使用主成分方法有效将红外遥感仪器的几千通道压缩到几百个类似通道的主成分。通过 PCRT 模式有效地实现将大气状态变量模拟成红外高光谱的主成分辐射率或者进而再重构成类似通道的辐射率。将 PCRT 模式与 WRF 4D-Var 系统对接起来实现, 实现在四维变分同化系统中才用主成分方法同化模拟的红外高光谱资料。为了实现具有主成分特色的辐射传输 PCRT 模式快速、有效的模拟红外高光谱主成分辐射率, 通过对 PCRT 模式选用特殊较小大气廓线集, 采用特殊的辐射传输计算方式, 并且根据其噪声水平, 设置需要计算的预报因子数目和主成分数目。在 WRF 4D-Var 对接 PCRT 模式, 修改调用 PCRT 模式的相应接口, 并且为同化红外高光谱资料设置其 `namelist` 参数, 实现了同化红外高光谱资料主成分辐射率。

主成分方法能够减少光谱中的大量仪器噪声, 方便我们研究通道选取方法中被放弃的通道用于资料同化的效果, 促使我们更进一步了解成千上万红外高光谱中包含的高精度, 高分辨率信息。

采用基于主成分方法的辐射传输 PCRT 模式有效模拟主成分辐射率, 或者将主成分重构出 Jacobians 局部化类似真实通道特征的辐射率, 实现用主成分方法提取模拟红外高光谱资料的信息。使用较小的廓线样本集生成辐射率资料库, 只对不在廓线样本集中的廓线执行辐射传输计算获得多色辐射率预报因子, 通过设置的预报因子进行回归计算, 得到模拟的主成分辐射率或重构辐射率, 这一系列因素使主成分方法具有非常高的计算效率, 获得相对目前业务上广泛使用的 RTTOV 较大的计算增益, 降低了辐射传输模式计算开销对资料同化系统的约束。

在先进的四维变分同化系统中, 同化红外高光谱主成分辐射率或重构的辐射率, 效果是中性的。虽然不能获得已经发布的预报更精确的预报影响, 但是我们应该了解到, 目前还不能证明在同化中添加更多的大量通道信息将对模式场产生更重大的积极影响, 而且对于 IASI 和 AIRS, NWP 模式业务同化的通道数目并不远小于重要的 $15\mu\text{m}$ CO_2 带的最大值^[52]。目前的业务同化中, 大量通道信息被放弃使用, 本文采用主成分方法的降噪特性提供了研究这些未开发通道信息的机会。

总之, 主成分方法包含了红外遥感完整光谱的信息, 并将这些信息表示成几百个数目的光谱。这种额外的信息具有优良的潜质, 促进我们对大气状态的认识

并且应用这些信息改进预报技巧，将来仍需要大量工作证明从这些压缩光谱中能够提取这些有效信息。

6.2 未来工作展望

开发基于‘块’的主成分计算，例如每 6 分钟对这个时段的观测资料计算主成分，并且专门为这些资料模拟出一个主成分的集合，确保计算中获得的特征向量能正确表达被压缩的光谱，并且这是一种更有效的压缩方式。在实现更密集的主成分计算时，如果发生资料通信过程，那么将不得不在计算过程中实现通信，且需要对其实现通信优化。如果需要更进一步需要提高主成分的计算效率，那么需要训练出一个不断变化的特征向量集合，适应基于主成分的快速辐射传输计算。从计算性能优化的角度，将基于‘块’的主成分计算实现并行化计算，势必更能满足实时资料同化的计算效率需求。

在搭建好的资料同化平台上，进一步研究先进的红外高光谱中蕴含的高精密度、高分辨率信息，调查红外高光谱各光谱带不能被充分用于预报的限制因素，区分仪器噪声，或者是其它的因素。进行多组同化试验，进一步研究更多种类红外高光谱仪器探测的光谱资料特征与同化方法。

致 谢

在我的课题和硕士论文完成之际，谨向在我攻读硕士学位的过程中曾经指导过我的老师，关心过我的朋友，关怀过我的领导，和所有帮助过我的人们致以崇高的敬意和深深的感谢！

衷心地感谢我的导师张卫民研究员。感谢张老师在我学习、工作和生活上给予的关心和指导，正是在张老师的谆谆教诲、悉心指导和严格要求下，我才得以顺利地完成任务的研究和论文的撰写。张老师工作繁忙，但对我课题的开题、研究，到最终论文的撰写，他都亲自过问，悉心指导，可以说我的课题凝聚着张老师的心血。张老师渊博的知识、精湛的学术水平、孜孜不倦的敬业精神让我受益匪浅。我很感激也觉得幸运可以成为张老师的学生。

感谢软件所宋君强老师对我们工作学习的支持与指导，感谢 611 教研室的曹小群老师、吴忠元老师、王舒畅老师、赵军老师、吴建平老师、张理论老师、王勇献老师对本课题的关心和指导。本课题的选择和展开是在他们的指导下完成的。在课题研究过程中，曹小群老师、吴忠元老师、王舒畅老师等不厌其烦地解答我提出的各种问题，指导我解决一个又一个难题。他们严格踏实的学术作风、极强的科研能力和在学术上的钻研精神都值得我认真学习。

感谢同门的贾雄、皇群博、覃耀磊同学，感谢师兄姚露、卢风顺、刘海、银福康、李宗哲、曹为、曾瑶源，师姐董蕴源、刘灿灿，他们对我学习的耐心帮助和细心指导让我受益匪浅。感谢已经毕业的尚卫卫、汤继飞、张开同学的入门指导，感谢师弟朱孟斌、李大力，同学郭松、熊海、杨博、夏榕泽、李正在学习上的支持，感谢我的室友孙冬婷、王志平、华楠楠、陈奎给予我生活上的帮助。

感谢湖南省气象局屈右铭同志的帮助，在他的指导下，我对 WRF 系统和 WRFDA 有了更深入的理解。他提供了部分运行 WRFDA 的脚本，促使了我掌握了运行 WRFDA 的相关技术。感谢他对我学习科研的指导。

感谢学员队唐政委，贺队长在日常生活的关怀，让我们可以安心工作。正是他们的付出让我们可以全身心投入学习。

感谢研究生院、计算机学院、硕士生队领导以及全体同学对我的关心和帮助。

永远感谢父母的养育之恩！在我二十多年的寒窗生涯中，父母孜孜不倦的教诲和鼓励是促使我不断进取并完成学业的重要精神动力。特别感谢我的公公、婆婆和我的老公桂猛，一直以来他们照顾好了家庭，给予我生活和学习上极大的关怀，让我全力以赴投入到学习当中，无论什么时候，他们都是我最坚强的后盾。

衷心感谢所有给予过我帮助和支持的人！

参考文献

- [1] Eugenia Kalnay. 大气模式、资料同化和可预报性. 北京: 气象出版社, 2005.
- [2] 黄静, 邱崇践, 张艳武. 晴空卫星红外模拟资料反演大气温度廓线的研究. 高原气象, 2005, 24 (6): 913-919.
- [3] 郭铤, 倾继祖. NOAA 卫星沙尘暴光谱特征分析及信息提取研究. 高原气象, 2004, 23 (5): 643-647.
- [4] 吕达仁, 王普才, 邱金桓. 大气遥感与卫星气象学研究的进展和回顾. 大气科学, 2003, 27 (4): 215-222.
- [5] Susskind J, Barnett C D, Blaisdell J M, et al. 2003. Retrieval of atmospheric and surface parameters from AIRS/AMSU/HSB data in the presence of clouds. IEEE Trans. Geosci. Remote Sensing, 41:390-409.
- [6] 官莉. 星载红外高光谱资料的应用. 北京: 气象出版社, 2007.
- [7] 方宗义, 许健民, 赵凤生. 中国气象卫星和卫星气象研究的回顾和发展. 气象学报, 2004, 60(5): 550-560.
- [8] Aumann H., Chahine M., Gautier C., et al. 2003. AIRS/AMSU/HSB on the Aqua mission: design, science objectives, data products, and processing systems. IEEE Trans. Geosci. Remote Sensing, 41:253-264.
- [9] Huang H., Revercomb H., Thom J., et al. Geostationary Imaging FTS(GIFTS) data proceeding: measurement simulation and compression. Proceeding of SPIE, 2000, 103-114.
- [10] Andrew C. Assimilation of AIRS and IASI at ECMWF. ECMWF Seminar on Recent development in the use of satellite observations in NWP, 2007, 127-150.
- [11] Marco M. A principal component based version of the RTTOV fast radiative transfer model. Quarterly Journal of the Royal Meteorological Society, 2010, 136:1823-1835.
- [12] Andrew C., McNally A., Hilton F., Healy S., Atkinson N. The use of principal component analysis for the assimilation of high-resolution infrared sounder observations for numerical weather prediction. Quarterly Journal of the Royal Meteorological Society, 2010, 136:2038-2050.
- [13] 官莉. 利用 AIRS 卫星资料反演大气廓线 1. 特征向量统计反演法. 南京: 南京气象学院学报, (29): 756-761.
- [14] 张建伟, 王根, 张华, 黄静, 陈靖, 吴玲玲. 基于主成分累积影响系数法的高光谱大气红外探测器的通道选择试验[J]. 大气科学学报, 2011, 34(1): 36-42.
- [15] Florence R., Forrie N., Currie D., et al. Channel selection methods for infrared atmospheric sounding interferometer radiances. Quarterly Journal of the Royal Meteorological Society, 2002, 128(581): 1011-1027.
- [16] Fourrie N., Jean-Noel thepaut research department. Validation of the NESDIS near real time AIRS channel selection. ECMWF Tech. Memo. European

-
- Center for Medium Range,2002.
- [17]Collard A., Saunders R., Cameron J., Harris B., Takeuchi Y., Horrocks L. Assimilation of data from AIRS for improved numerical weather prediction. Proceedings Of the 13th International TOVS Study Conference,2003.
- [18]James C., Andrew C., English S. Operational use of AIRS observations at the met office. 14th Internat Tovs Study Conference, Beijing,China,2005.
- [19]Sounders R., Martricardi M., Geer A. Rttov 9.1 user guide. NWP SAF report ,Met office,2008.
- [20]McNally A., Watts P., Smith J., Engelen R, Kelly G., Thepaut J., Matricardi M. The assimilation of AIRS radiance data at ECMWF. Quanlity Journal of the Royal Meteorological Society, 2006,132:935-957.
- [21]蒋尚城. 应用卫星气象学.北京大学出版社,2006.
- [22]胡庆军. 高等工程数学. 国防科技大学出版社.2006.
- [23]施能. 气象科研与预报中的多元分析方法. 北京:气象出版社,2005.
- [24]宋金暖. AIRS 资料反演晴空大气廓线方法的模拟研究.南京信息工程大学硕士学位论文.2007.
- [25]Hilton F., Collard A., Guidard V., Randriamampianiana R., Schwaerz M. Assililation of IASI Radiances at European NWP Centers. Proceedings of the ECMWF/EUMETSAT NWP-SAF Workshop on the assimilation of IASI in NWP,2009.
- [26]Huang H., Antonelli P. Application of principal component analysis to high-resolution infrared measurement compression and retrieval. Application Meteorological, 2001,40:365-388.
- [27]Martricardi M. An observation operator for the assimilation of principal component scores into a NWP system. Report EUMETSAT Comtract EUM/CO/07/4600000475/PS,ECMWF,Reading,UK.64pp,2009.
- [28]Fiona Hilton, Andrew Collard. Recommendations for the use of principal component-compressed observations from infrared hyperspectral sounders, Forecasting Research and Development Technical Report,2009.
- [29]曾庆存. 大气红外遥测原理. 北京: 科学出版社,1974.
- [30]Newman S., Hilton F., Clough S., Taylor J., Collard A., Smith W., Larar A. IASI validation studies. Proceedings of the ECMWF/EUMETSAT NWP-SAF Workshop on the assimilation of IASI in NWP, 2009,5:15-24.
- [31]Collard A., Matricardi M. Selection of a subset of IASI channels for near real time dissemination. Proceedings of ITSC-XIV, Beijing, China, 2005.
- [32]Collard A., McNally A. The assimilation of Infrared Atomospheric Sounding Interferometer radiances at ECMWF. Quanlity Journal of the Royal Meteorological Society,2009,135:1044-1058.
- [33]Hilton F., Atkinson N., English S., Eyre J. Assimilation of IASI at the Met Office and assessment of its impact through observing system experiment. Quanlity Journal of the Royal Meteorological Society, 2009,135:495-505.
- [34]潘宁, 董超华, 张文建, 黎光清. 变分同化及卫星资料同化. 气象科
-

-
- 技,2001,2:29-36.
- [35]Liu X., Smith W., D.K.Zhou., larra A. Pincipal component-based radiative transfer model for hyperspectral sensors: theoretical concept. *Applied Optics*,2006,45:201-209.
- [36]Marco M. Radiative transfer for IASI. ECMWF/EUMETSAT NWP-SAF Workshop on the assimilation of IASI in NWP,2009,5:25-38.
- [37]Yu Yi, Zhang Weimin, Huang Qunbo, Yao lu. A New Interpolation Algorithm in Fast Radiative Transfer Model. 2011 International Conference on Computer Application and System Modeling, 2011.
- [38]Tobin D., Antonelli P., Revercomb H., Dutcher S., Tuerner D., Taylor J., Knuteson R. Vinson K. Hyperspectral data noise characterization using principal compont analysis: Application to the atmospheric infrared sounder. *Journal Application Remote Sensing*,2007.
- [39]李娟, 朱国富. 直接同化卫星辐射率资料在暴雨预报中的应用研究. *气象*, 2008,34(12):38-43.
- [40]官元红,周广庆,陆维松,陈建萍. 资料同化方法的理论发展及应用综述. *气象与减灾研究*,2007,30(4):1-8.
- [41]邹晓蕾. 资料同化理论和应用(上册). 北京:气象出版社,2009.
- [42]Zhu jiang, Zhou Guangqing, Yan Changxiang, Fu Weiwei, You Xiaobao. A three-dimensional variational ocean data assimilation system: Scheme and preliminary results. *Science in China Serieces D: Earth Science*, 2006,49(11):1212-1222.
- [43]Sounders R., Matricardi M., and Brunel P. An improved fast radiative transfer model for assimilation of satellite radiance observations. *Quanlity Journal of the Royal Meteorological Society*,1999,144:1547-1558.
- [44]Huang X., Xiao Q., Barker D., Zhang xin., and et al. Four-Dimension Variational Data Assimilation for WRF: Formulation and Preliminary Results. *American Meteorological Society*,2009,137:299-314.
- [45]Huang X., Xiao q., Huang W., Barker D., Kuo Y., Michalakes J., Ma Z. The Weather Research and Forecasting Model Based 4-dimensional Variational Data Assimilation System. 2005.
- [46]ARW Version 3 Modeling System User'S Guide.
<http://www.mmm.ucar.edu/wrf/users>.
- [47]Harveman S., Thelen J., Taylor J., Keil A. The Havemann-Taylor Faster Radiative Transfer Code: Exact fast radiative transfer for scattering atmospheres using Principal Componets(PC).AIP Conference Proceedings, 2009, 1100:38-40.
- [48]Turner D., Knuteson R., Revercomb H., Lo C., Dedecker R. Noise reduction of Atmoshepric Emitted Radiance Interferometer(AERI) observations using principal component analysis. *Journal Atmospheric Oceanic Technology*, 2006,23:1223-1238.
- [49]Antonelli P., Revercomb H., Sromovsky L., W.L.Smith, Knuteson R.O., and et
-

- al. A principal component noise filter for high spectral resolution infrared measurements. *Journal Geophys. Res.*,2004:23102-23124.
- [50] Watts P., McNally A. The sensitivity of a minimum variance retrieval scheme to the values of its principal Parameter. In *Technical proceeding of the Fourth TOVS Study Conference*,1988,4:399-407.
- [51] Le Marshall, Jung J., Derber J., Chahine M., and et al. Improving Global Analysis and Forecasting with AIRS. *Bull. Amer. Metror. Soc.*,2006,87:891-894.
- [52] Andrew Collard. Assimilation of AIRS and IASI at ECMWF. *ECMWF Seminar on Recent development in the use of satellite observations in NWP*,2007:127-150.

作者在学期间取得的学术成果

- [1] Yu Yi, Zhang weimin, Huang qubo, Yaolu. A New Interpolation Algorithm In Fast Radiative. 2011 International Conference on Computer Application and System Modeling, 2011.EI 检索。
- [2] Huang Qunbo, Zhang weimin, Yu Yi. A Cloud Parameters in the Variational Assimilation. 2011 International Conference on Computer Application and System Modeling, 2011.EI 检索。
- [3] Jia Xiong, Liu Hai, Zhang weimin, Yu Yi. Design and Implementing of Meteorological Data Management System. 2011 International Conference on Computer Application and System Modeling, 2011.EI 检索。