

# 集合、同化思想在大气科学中的渗透

薛建军<sup>1,2</sup> 贾朋群<sup>1</sup> 肖子牛<sup>2</sup>

(1 中国气象局气象干部培训学院, 北京 100081;

2 中国科学院大气物理研究所大气科学和地球流体力学数值模拟国家重点实验室, 北京 100029)

**摘要:** 模式误差、初始误差、以及非线性系统的不稳定性共同导致了大气、海洋及其耦合系统预报预测的不确定。为了尽可能减少这些误差对数值预报的影响, 集合、同化的思想和方法“应需而生”并得以在现代天气气候预报预测中不断发展应用。本文简要回顾了集合预报、资料同化的理论方法的主要发展历程及应用, 介绍了ECMWF等机构在相关领域里的一些新的研发方向及进展。随着相关理论和技术的日臻成熟, 传统方法短期内难以取得新的重大进展, 而以人工智能等为代表的新兴技术与气象的融合应用受到关注并有望取得突破。

**关键词:** 数值预报, 集合, 同化, 人工智能, 气象

**DOI:** 10.3969/j.issn.2095-1973.2022.06.007

## Application and Progress of Ensemble and Assimilation for the Atmospheric Science

Xue Jianjun<sup>1,2</sup>, Jia Pengqun<sup>1</sup>, Xiao Ziniu<sup>2</sup>

(1 China Meteorological Administration Training Centre, Beijing 100081

2 State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029)

**Abstract:** The combined effects of model errors, initial condition errors and the instability of nonlinear systems lead to uncertainties in forecasting for the atmosphere, ocean and their coupled systems. In order to reduce the above negative effects on numerical weather prediction as much as possible, the methods of ensemble and assimilation have been proposed and continuously developed and applied in modern weather and climate prediction. This paper reviews the histories and applications of the main theories and methods for ensemble forecasting and data assimilation. We also introduce some progress and academic frontiers carried out by relevant institutions such as ECMWF. With the development of the theories and methods, limited space is left for traditional methods to achieve significant progress in the short term, while the fusion of emerging technologies such as artificial intelligence (AI) has attracted widespread attention and is expected to make breakthroughs.

**Keywords:** Numerical Weather Prediction (NWP), ensemble, assimilation, artificial intelligence (AI), meteorology

### 0 引言

大气科学作为一门较新的应用性学科, 它的学科独立、迅速发展与数学、物理、化学等学科的发展和相关技术的进步密不可分。自18世纪中叶以来, 在几次科技革命的推动下, 大气科学从定性描述到定量分析再到预报预测能力的日臻成熟<sup>[1-3]</sup>。特别是进入20世纪50年代以来, 伴随着数值预报概念的提出到其科学理论和技术方法的持续发展完善, 大气科学取得了迅猛发展。作为大气科学发展的重要标志, 数值预报早已引发了一场“静悄悄的革命”<sup>[4]</sup>, 它的巨大成功和飞速进步无疑是建立在多年稳步持续的科学认知和技术

进步的积累之上<sup>[4-5]</sup>, 这其中就包括集合、同化的科学理论及其技术方法在大气科学中的渗透和应用。

1963年, Lorenz提出了著名的混沌理论, 指出混沌系统由精确的确定性演化方程控制, 但具有不可预测且看似随机的行为<sup>[6-8]</sup>。而大气、海洋等具有很强的混沌特性, 导致预报预测时对初始状态具有高度的敏感性。众所周知, 从数值预报建立的科学理论与技术基础上看, 首先数值模式并不“完美”<sup>①</sup>, 其次一个真实的物理系统的状态也无法准确测量。加之大气、海洋等因观测误差、分析误差、站网分布不均等原因都会导致误差不可避免<sup>[9]</sup>。因此, 模式误差、初始误差、以及非线性系统的不稳定性共同导致了大气、海洋及其耦合系统预报预测的不确定<sup>[10-13]</sup>。因此, 为了尽可能

收稿日期: 2022年9月15日; 修回日期: 2022年10月28日

第一作者: 薛建军, Email: xuejianjun@cma.gov.cn

资助信息: 国家自然科学基金(42142009)

① 即便模式“完美”, 由于初值的误差及混沌特征, 预报预测仍然不可避免地存在预报误差。

能的减小预报预测的误差,集合预报、资料同化的科学理论与技术方法“应需而生”并得以在现代天气气候预报预测中不断发展应用。

## 1 集合预报

### 1.1 集合预报的提出

通常认为集合预报的思想是由Epstein和Leith提出的<sup>[8-12, 14-15]</sup>。在假定模式完美的情况下,找到一组合适的扰动,以“初值+扰动”进行预报来反映初值引起预报结果的不确定性,再通过集合平均保留各个预报最可能的共同结果,以减少初值带来的不确定性影响。在理论上可以用一个概率密度函数来描述这种不确定性<sup>[13, 16]</sup>, Epstein提出显式积分Liouville方程估计大气状态的概率密度分布,但在实际中几乎难以实现。随后, Leith建议采用蒙特卡洛(Monte Carlo Forecasting, MCF)方法,用随机函数产生扰动形成的初始场,实现了集合预报从理论迈向实际<sup>[10, 16-18]</sup>。

### 1.2 集合预报的发展

从时间上划分,一般可以将集合预报的发展历程划分为以下三个阶段<sup>[11-12, 17]</sup>。第一阶段,20世纪70—80年代,主要开展集合预报理论方法研究和数值试验探索。第二阶段,20世纪90年代,随着计算机技术,特

别是大规模高性能计算能力的提升,业务集合预报系统得以建立和发展。1992年12月7日,美国国家环境预报中心(NCEP)率先开展集合预报业务<sup>[19]</sup>,同年12月19日欧洲中期天气预报中心(ECMWF)的集合预报系统开始准业务运行并于18个月后(1994年5月)正式投入业务运行<sup>[19-20]</sup>。随后法国、英国、日本、澳大利亚、加拿大等国也建立了各自的集合预报系统<sup>[9, 11-12, 19]</sup>。第三阶段,20世纪90年代末以来,集合预报的研究更加深入,从初值的不确定性研究扩展到模式不确定性、集合预报产品解释应用技术等多个领域。集合预报速成为数值天气预报业务的核心,被全球多个国家和地区的业务、科研机构所采用,甚至成为全球各大气强国竞相发展的重要领域之一。表1,表2分别给出了世界气象组织(WMO)数值试验工作组对集合预报业务现状及发展态势的最新调查<sup>[21]</sup>,可以看到,国外主要的业务中心仍然在持续加紧其在全球或区域集合预报的工作部署和技术研发。

尽管我国的数值预报业务起步于1950年代并很快取得了一些具有国际影响的进展<sup>[22-24]</sup>,但直到改革开放之后,相关研究和业务应用才迎来了新的契机<sup>[24]</sup>。受限于计算资源等客观条件限制,国家气象中心在20世纪90年代中期引入新的高性能计算机,后于1996

表1 国外主要预报中心全球集合预报业务系统主要参数:水平分辨率(谱/格点)、垂直层数(L)、成员数(M)、预报时长及每天运行次数<sup>[21]</sup>

Table1 Operational system parameters of the global ensemble forecast of major forecast centers abroad: horizontal resolution (spectrum/grid), number of vertical layers (L), forecast duration, number of members (M), forecast duration and daily operation times<sup>[21]</sup>

| 预报中心(国家/地区)        | 2021   | 2022   | 2023   |
|--------------------|--|--|--|
| ECMWF(欧洲)          | TCo639 L137 D+15;<br>TCo319 L137 D+45;<br>耦合海洋-大气, 0.25°, 2次/天   | 51×TCo1023 或 TCo1279 L137 D+15;<br>100×TCo319 L137 D+45;<br>耦合海洋-大气, 0.25°, 2次/d | 50×TCo1023或TCo1279 L137 D+15;<br>100×TCo319 L137 D+45;<br>耦合海洋-大气, 0.25°2次/d |
| Met Office(英国)     | 20 km, L70, 7天;<br>M18(6h循环), M36后延产品  | 耦合海洋-大气20 km, L70; 10 d;<br>M18(6h循环), M36后延产品                                   | 耦合海洋-大气20 km, L70; 10 d;<br>M18(6h循环), M36后延产品                               |
| Météo France(法国)   | ~T <sub>L</sub> 1198c2.2L90(西欧7.5 km, 其他<br>37 km); M35, 4 d, 4次/d                                       | ~T <sub>L</sub> 1798c2.2 L105(西欧5 km, 其他<br>24 km); M35, 4 d, 4次/d               | ~T <sub>L</sub> 1798c2.2 L105(西欧5 km, 其它<br>24 km); M35, 4 dz.4次/d           |
| DWD(德国)            | 40 km(北大西洋、欧洲20 km) L90;<br>M40, 7.5 d(5天嵌套), 4次/d   | 26 km(欧洲13 km) L120; M40, 7.5 d(5d<br>嵌套), 4次/d                                  | 26 km(欧洲13 km) L120; M40, 7.5 d<br>(5 d嵌套), 4次/d                             |
| HMC(俄罗斯)           | T169L31+0.72°x 0.9° L28,<br>M22, 10 d, 1次/d<br>SLAV0.72×0.9L96; M60, 14 d, 1次/d                          | SLAV0.72×0.9L96; M60, 14 d, 2次/d   | SLAV 20 km, L96; M40, 14 d, 2次/d   |
| NCEP(美国)           | C384 L64, M31, D+16, 3次/d<br>D+35, 1次/d  | C384 L64, M31, D+16, 3次/d<br>D+35, 1次/d  | C384 L64, M31, D+16, 3次/d<br>D+35, 1次/d                                      |
| Navy/FNMOC/NRL(美国) | T359L60 M20, 16天, 2次/d   | T359L60 M20, 16 d, 2次/d  | TBD(T359L100或T425L100), M20,<br>16 d, 2次/d                                   |
| CMC(加拿大)           | 0.35°L80; M20, 16 d, 2次/d  | 0.35°L84; M20, 16 d; 2次/d  | 0.23°L84; M20, 16 d, 2次/d  |
| CPTEC/INPE(巴西)     | T126, L28; M15, 15 d, 2次/d   |  |  |
| JMA(日本)            | TL479 L128; M51, 11 d/ 5.5 d(4次/d);<br>TL479 L128, M51, 11~18 d, 1次/d;<br>TL319 L128; M25, 18~34 d, 2次/周 |  |  |
| KMA(韩国)            | 32 km, L70; M25, 12天, 2次/d   | 32 km, L70; M25.12 d, 2次/d   |  |
| IMD(印度)            | T1534, L64, 21M;<br>10 d, 2次/d   | T1534, L64, 21M;<br>10 d, 2次/d   | T1534, L64, 21M;<br>10 d, 2次/d   |

表2 国外主要预报中心区域集合预报业务系统主要参数: 水平分辨率(谱/格点)、垂直层数(L)、成员数(M), 预报时长及每天运行次数<sup>[21]</sup>

Table2 Operational system parameters of the regional ensemble forecast of major forecast centers abroad: horizontal resolution (spectrum/grid), number of vertical layers (L), forecast duration, number of members (M), forecast duration and daily operation times

| 预报中心(国家/地区)        | 2021                                       | 2022                                       | 2023                                     |
|--------------------|--|--|--|
| ECMWF(欧洲)          |  |  |  |
| Met Office(英国)     | 2.2 km, L70; 3M每小时, 18M后延产品; 5 d           | 2.2 km, L70; M每小时, 18M后延产品; 5 d            | 2.2 km, L70; 3M每小时, 18M后延产品; 5 d         |
| Météo France(法国)   | 2.5 km, L90; M16<br>2 d, 4次/d              | 1.3 km, L90; M16<br>d, 4次/d                | 1.3 km, L90; M16<br>2 d, 4次/d            |
| DWD(德国)            | 2.1 km, L65; M20<br>2 d×8                  | 2.1 km, L65; M40<br>2 d×8, 12 h×24         | 2.1 (1.0) km, L65; M40<br>2 d×8, 12 h×24 |
| HMC(俄罗斯)           |  | 2.2 km; M20<br>2 d; 2次/d                   | 2.2 km; M20<br>2 d; 2次/d                 |
| NCEP(美国)           | 16 km, M26 3.5 d 4次/d<br>3 km M10 2 d 4次/d | 16 km, M26 3.5 d 4次/d<br>3 km M10 2 d 4次/d | 3 km, M9, 18 h 24次/d, 2.5d 4次/d          |
| Navy/FNMOC/NRL(美国) | 4 km, 5 d M11, 2~4次/d                      | 4 km, 5 d, M21, 2~4次/d                     | 4 km, 5 d, M21, 2~4次/d                   |
| CMC(加拿大)           | 0.09°; M20, 3.5 d, 4次/d                    | 0.09°; M20, 3.5 d, 4次/d                    | 0.09°; M20, 3.5 d, 4次/d                  |
| CPTEC/INPE(巴西)     | 7~8 km; 多模式M3, 7~10 d                      |  |  |
| JMA(日本)            | 5 km; L76; M21, 39 h, 4次/d                 | 5 km; L96; M21, 39 h,<br>4次/d,             | 5 km; L96; M21, 39 h<br>4次/d             |
| KMA(韩国)            | 2.2 km, L70; M13, 2 d 2次/d                 | 2.2 km, L70; M13, 3 d, 2次/d                |  |
| IMD(印度)            |  |  |  |

年5月开发建立了一套中期集合数值预报系统<sup>[11, 13]</sup>。此后, 1998年6月在国产神威巨型计算机上建立了T106L19全球模式的中期数值天气集合预报系统, 并于2001年3月实现业务运行<sup>[16]</sup>。世纪之交, 在数值预报以引进为主转为自主开发为主的大背景下, 开启了我国新一代数值天气预报系统研发的新征程。经过多年努力, 我国自主研发的国家级全球(CMA-GEPS)和区域集合预报系统得以建立(CMA-REPS)<sup>[13, 24]</sup>, 实现了全球50 km, 15 d, 中国区域10 km, 3.5 d的集合预报。同时, 基于中国气象局次季节-季节-年际尺度一体化气候模式预测系统(CMA-CPS)构建了第三代气候集合预测模式业务系统, 实现大气45 km, 海洋0.25°×0.25°的S2S(0~60 d)和季节(未来13个月)气候预测<sup>[25-26]</sup>(表3)。

### 1.3 集合预报主要的技术方法

集合方法提出的背后, 是考虑到“单一”的初值/模式的不确定性难以避免, 而“一组”考虑了不同的初值/模式的不确定性影响得到的“一群”预报结果, 其大概率是要好于某单个结果。简单归纳起来, 集合预报的主要技术路线的思想如下<sup>[8-12, 27-28]</sup>:

(1) 假定模式完美, 主要解决初值带来的不确定性影响。“初值问题”的解决是集合预报提出并得以建立起来的重要标志。通过“初值扰动”, 生成一组不同考虑了不确定性的初值样本来驱动模式, 得到集合预报结果。由于早期的蒙特卡洛法(MCF)、时间滞后

表3 国家级集合数值预报系统概况<sup>[25, 26]</sup>

Table 3 Overview of CMA Ensemble Numerical Weather Prediction System<sup>[25, 26]</sup>

| 名称                                   | 概况   |
|--------------------------------------|--|
| 中国气象局全球集合预报系统(CMA-GEPS)              | 全球区域范围50 km分辨率<br>垂直87层, 模式顶高0.1 hPa<br>控制预报+30成员<br>15天预报<br>每天运行2次(00, 12UTC)  |
| 中国气象局区域集合预报系统(CMA-REPS)              | 中国区域范围10 km分辨率<br>控制预报+15成员<br>3.5天预报<br>每天运行2次(00, 12UTC)   |
| 中国气象局次季节-季节-年际尺度一体化气候模式预测系统(CMA-CPS) | 大气: T266 (~45 km), 垂直56层, 模式顶高0.1 hPa<br>海洋: 全球 0.25°×0.25°, 50层<br>S2S: 未来60天, 4个集合成员, 每周一、四运行<br>季节: 未来13个月, 21个集合成员每月运行 |

平均(LAF)等方法存在较大的不足, 难以满足实际业务应用需求<sup>[10, 16-18, 32]</sup>, 实践中常用的方法有:<sup>[8-12, 16-18, 28-31]</sup> 奇异向量法(SVs)、繁殖向量法(BVs)、观测扰动方法(PO)、集合转换(ET)和重新尺度化集合转换(ETR)和条件非线性最优扰动(CNOP)等。近年来集合卡尔曼滤波(EnKF)、集合转换卡尔曼滤波(ETKF)等数据同化的办法也有较多的应用。以上各方法特点及应用情况见表4。

(2) 假定初值精确, 致力于解决模式引起的不确定性影响。随着观测和同化等技术快速发展, 一些研究认为模式系统误差也是影响天气集合预报效果的主要原因, 比如模式中不同物理、化学过程、参数化方案等同样存在不确定性。集合预报从仅考虑初值扩展

表4 国外主要预报中心全球资料同化系统主要参数: 同化类型、水平分辨率(谱/格点)、垂直层数(L)、成员数(M)<sup>[21]</sup>  
 Table 4 2 Operational system parameters of the global data assimilation system of major forecast centers abroad: assimilation type, horizontal resolution (spectrum/grid point), vertical layer number (L), and number of members (M)<sup>[21]</sup>

| 预报中心(国家/地区)        | 2021   | 2022  | 2023  |
|--------------------|--|---|---|
| ECMWF(欧洲)          | OOPS/TCo1279 Hybrid-4DVar, 弱约束, 连续同化(集合资料同化 M50, TCo639) TCo1279 / TL399, L137 | OOPS/TCo1279 Hybrid-4DVar, 弱约束, 连续同化(集合资料同化 M50, TCo1023 或 TCo1279) TL511, L137 | OOPS/TCo1279 Hybrid-4DVar, 弱约束, 连续同化(集合资料同化 M50, TCo1023 或 TCo1279) TL511, L137 |
| Met Office(英国)     | Hybrid 4DVar (~40 km. L70), M44, EN-4D-ENVar                                   | Hybrid 4DVar (~40 km. L70), M44, EN-4D-ENVar                                    | Hybrid 4DVar (~40 km. L70), M100 EN-4D-ENVar                                    |
| Météo France(法国)   | ~4Dvar 6 h, 集合资料同化 M50, TL499C1L105  | ~4Dvar 6 h, 集合资料同化 M50, TL499C1L105   | ~4Dvar 6小时, 集合资料同化 M50, TL499C1L105   |
| DWD(德国)            | Hybrid 3DVar 13 km, EnKF, M40 (40/20 km) L90                                   | Hybrid 3DVar 13 km, EnKF, M80 (26/13 km), L120                                  | Hybrid 3DVar 13 km, EnKF, M80 (26 /13 km), L120                                 |
| HMC(俄罗斯)           | 3DVar, 0.25°×0.25°, L40  | 3DVar, 0.25°×0.25°, L40   | 3DVar, 0.25°×0.25°, L40   |
| NCEP(美国)           | Hybrid 4D EnVar, 25 km, L127 M80, LETKF  | Hybrid 4D EnVar, 25 km, L127 M80, LETKF   | Hybrid 4D EnVar, 25 km, L127 M80, LETKF   |
| Navy/FNMOC/NRL(美国) | Hybrid 4DVar, (T681L60/T119L100)   | Hybrid 4DVar, (T681L100/T119L100)   | Det. Hybrid 4DVar, (T681L100/T119L100)  |
| CMC(加拿大)           | Hybrid 4D EnVar, 0.35°, L84  | Hybrid 4D EnVar, 0.35°, L84   | Hybrid 4D EnVar, 0.35°, L84   |
| CPTEC/INPE(巴西)     | 3DVar, TQ0299 (~45 km), L64  |   |   |
| JMA(日本)            | Hybrid-4Dvar, TL319, L128, LETKF, M100   | Hybrid-4DVar, TL319; L128, LETKF, M100  | Hybrid-4DVar, TL319, L128, LETKF, M100  |
| KMA(韩国)            | Hybrid-4DVar 40 km, L70  | Hybrid-4DVar, 40 km, L70  |   |
| IMD(印度)            | Hybrid 4D EnVar, 25 km, L64, M80, EnKF   | Hybrid 4D EnVar, 25 km, L64, M80, EnKF  | Hybrid 4D EnVar, 25 km, L64, M80 EnKF   |

到了考虑模式不确定性的影响<sup>[10, 17, 28-29, 32]</sup>。具体的方法有:

多物理过程法。如加拿大气象中心(CMC)的全球集合预报系统采用多物理过程组合来减小物理过程不确定性引起的模式误差。如使用不同的对流、水平扩散、重力波拖曳、辐射和地形处理方案等<sup>[29, 33]</sup>。

随机物理过程法。在集合预报系统中的模式倾向方程引入了随机物理过程法,该方法假定模式的不确定性主要来自于参数化过程的不准确和数值模式的截断误差。如ECMWF在1998年10着手考虑随机参数化扰动方案(SPPT),2010年11月引入随机补偿方案(SPBS)<sup>[10, 20, 29]</sup>。我国的CMA-GEPS、CMA-REPS和CMA-CPS目前也采用了SPPT方案<sup>[13, 24, 26]</sup>。

多模式集合。考虑了不同模式误差的综合影响,除了天气预报,多模式集合预报方法也被应用于气候预测的集合预报中,并取得了一定的成功<sup>[10]</sup>。

然而,由于模式误差往往与初始误差导致的预报误差相互作用难以区分,加之模式误差来源广泛,从次网格物理过程的参数化、模式离散到计算误差都有可能且相互耦合,对集合预报模式不确定性的改善比对初值引起的误差难度更大。

(3)模式和初值都不够好,设法同时解决两者带来的不确定性影响。除了单独考虑初值误差或模式误差,理论上也可将两者同时考虑解决。比如:

多模式-多初值集合法,同时考虑不同物理过程和初值不确定性对集合预报的影响,但可能在统计上不显著,预报成员之间存在较大的差异成集合预报结果较大的系统偏差<sup>[8, 10, 12]</sup>。

随机全倾向扰动法(STTP)在数值模式倾向方程叠加随机强迫扰动,通过多次实现该随机扰动产生集合成员进行集合预报。NCEP从2010年开始使用此方法<sup>[10, 29]</sup>。

#### 1.4 ECMWF的集合预报

ECMWF作为最早开始发布业务集合预报的预报中心之一,经过近30年的发展,其集合预报技术和产品性能早已处在全球领先地位(图1)。从1992年以来,水平分辨率提升了20倍,从~320 km增加到~16 km;垂直分辨率提升了7倍,从19层增加到137层;预测时间从10 d延长到了46 d;集合成员个数从33个增加到51个;集合预报制作的频率也有所增加,当前还开展集合预报回报业务<sup>[20, 34]</sup>。2021年5月11日,在ECMWF对其IFS(Integrated Forecasting System)进行的系统升级中(IFS Cycle 47r2),将模式计算由传统的双精度(64位浮点计算)降低至单精度(32位浮点计算)以节省计算资源,提高处理速度(图2),将集合预报从91层的垂直分辨率提至137层,实现其全球模式确定性预报(HRES)、数据同化(EDA)和集合预报(ENS)达到统一的垂直分层。测试结果显示,升级后的集合预

报系统对自由大气整体的预报技巧提高了0.5%~2%，对50 hPa的平流层温度预报技巧提高了5%~20%，对热带对流层的预报技巧提高了6%。此外还提升了对热带气旋的预报技巧<sup>[35]</sup>。未来，ECMWF还将在2023年第一季度的Cycle 48r1计划中，将ENS的水平分辨率提升至9 km，以与HRES保持一致。同时将当前ENS extended作为一个单独的系统独立出来，维持36 km的水平分辨率和垂直137层的设置，以100成员每天从00UTC运行到第46 d<sup>[36]</sup>。这些改变无疑将加快ECMWF向“无缝的集合数据同化和预测系统”<sup>[37]</sup>迈出更加坚实地步伐。

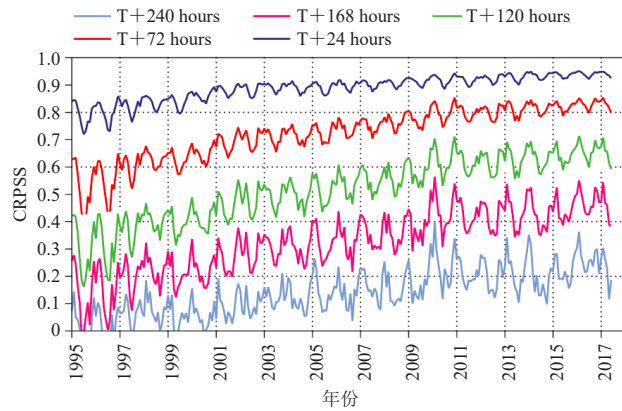


图1 1995—2017年，ECMWF北半球500 hPa位势高度集合预报CRPSS随时间的演变，预报时间为24、72、120、168和240 h（整体上呈上升趋势，每条线上波动与季节相关的可预报性有关，冬季比夏季更容易预测）<sup>[20]</sup>

Fig. 1 Time evolution, from 1 January 1995 to today, of the CRPSS of ENS forecasts for 500 hPa geopotential height over the Northern Hemisphere, for lead times of 24,72,120 168 and 240 hours. (The more or less regular pattern of peaks and troughs in each line stems from differences in predictability related to the seasons: winter weather tends to be more predictable than summer weather) <sup>[20]</sup>

## 2 资料同化

### 2.1 资料同化的概念及其理论发展

初值问题是数值预报的核心问题之一，在给定初值的条件下通过既定的模式进行数值求解以得到未来时刻的预报结果<sup>[17, 38-39]</sup>。资料同化是有效地初值形成方法，它能够“使用所有可用的信息，尽可能准确地估计大气运动的状态”<sup>[39-41]</sup>。先进的资料同化技术被认为是数值天气预报突破的关键因素之一<sup>[4, 17, 42-43]</sup>。资料同化通常包括资料预处理、客观分析和初始化等步骤，随着同化技术的不断发展，例如三/四维变分的应用，这些过程的界限越来越模糊，在实际中各步骤的功能已逐渐融合应用<sup>[17, 43]</sup>。

同其数值预报其他技术分支类似，资料同理论及

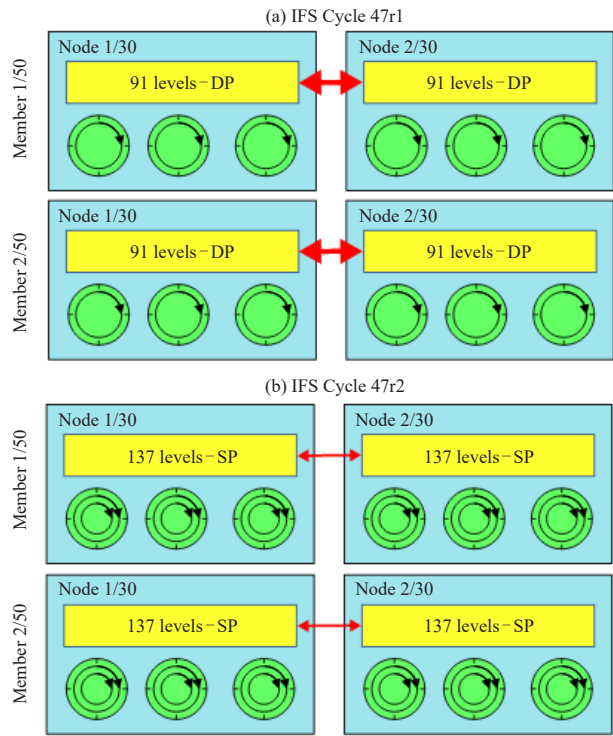


图2 IFS Cycle 47r1双精度 (DP) 和91层集合预测 (a) 和IFS Cycle 47r2单精度 (SP) 和137层集合预报 (b) 的计算效率比较 (单精度更快 (绿色圆圈) 并且减少了每个节点 (黄色框) 内存之间的数据传输 (红色箭头)) <sup>[35]</sup>

Fig. 2 The computational change from (a) IFS Cycle 47r1 with double precision (DP) and 91 levels in the ensemble forecast to (b) IFS Cycle 47r2 with single precision (SP) and 137 levels in the ensemble forecast allows faster core processing (green circles) and reduced data transfer (red arrows) between the memory on each node (yellow boxes)<sup>[35]</sup>

方法也是在实践中伴随着数值模式、综合观测和高性能计算等技术进步而不断发展。早在1922年Richardson把观测资料手工插值到网格点上作为数值预报的初始场。1950年代Charney、顾震潮等采用主观分析方法确定初值，这些也被看作是将同化的思想用于数值天气预报的主观分析中<sup>[39, 44]</sup>。20世纪50年代随着计算机和数值模式的发展，逐渐发展了气象领域中的客观分析方法 (OA)，而大气资料同化就源于数值天气预报的客观分析方法。60年代随着初始方程预报模式的出现，初值中不同变量间的动力学平衡成为关注的问题，客观分析所需的背景场亦可由模式预报提供，构成了预报—观测—预报的循环，资料同化的完整概念形成，并发展了统计插值 (又称最优插值) 的客观分析。70年代初伴随着全球模式的出现，提出了基于全球大气基本运动模态的非线性正规模式初始化方案。80年代以后资料同化的理论框架逐渐建立起来，这一时期卫星遥感资料的应用也极大地推动了同化理论与方法的

研究。80年代后期提出的变分同化方, 90年代中期又提出了基于集合预报的卡尔曼滤波方法等<sup>[17]</sup>。目前, 三/四维变分、混合变分、集合资料同化等方法以广泛用于业务资料同化系统<sup>[24, 42-43]</sup>, 全球主要业务中心也一直将资料同化系统的研发和升级作为关键核心业务能力之一(表5)<sup>[21]</sup>。

表5 集合预报初值扰动的主要方法及其应用情况<sup>[8-12, 16-18, 28-31]</sup>  
Table 5 Methods and applications of initial disturbance for ensemble forecasting<sup>[8-12, 16-18, 28-31]</sup>

| 名称   | 特点及应用情况  |
|--|--|
| 蒙特卡洛方法 (Monte Carlo method, MCF)   | 促成了集合预报从理论研究转向实际。由于不能反映模式分析误差的空间分布, 业务系统未采用  |
| 时间滞后平均方法 (Lagged Average Forecasting, LAF)   | Hoffman等在1983年提出, 难以产生高预报技巧的成员且对捕获随时间快速增长的误差能力有限   |
| 奇异向量法 (Singular Vectors, SVs)  | 由ECMWF提出并一直在业务系统中使用和不断改进的方法。如1998年演化奇异向量法 (evolved SV initial time SV, EVO-SVINI), 2010年, 集合资料同化一奇异向量法 (ensemble of data assimilation-initial time SV, EDA-SVINI)。我国的CMA-GEPS也采用了该方法 |
| 繁殖向量法 (Breeding vectors, BVs)  | 由NCEP发展并应用于其业务系统的, 用于产生集合该方法借鉴了资料同化技术, 通过模式的繁殖循环计算快速增长初始误差模态, 然后把得到的误差模态作为集合预报的初始扰动。该方法可以改善SVs未能加入模式物理过程的不足  |
| 观测扰动法 (Perturbation of Observation, PO)  | 由CMC发展和应用。该方法在资料同化期即引入集合预报思想, 在观测资料中加入具有独立同化分析周期的扰动, 由此产生独立的初始分析场, 与动力模式的协调性较差   |
| 集合转换法 (Ensemble Transform, ET)<br>重新尺度化集合转换范 (Ensembl Transform with Rescaling, ETR) | 该方法是对BV法的改进, 最早是Bishop和Toth在目标观测研究中提出的。2004年之前NCEP就开始了ET和ETR生成集合预报初始扰动的试验, 2006年5月ETR正式在NCEP投入业务使用   |
| 条件非线性最优扰动法 (Conditional Nonlinear Optimal Perturbation, CNOP)                        | 改善SVs不能充分反应非线性物理过程的影响, 描述了非线性模式中增长最快的初始扰动, 充分考虑了非线性的影响, 具有较好的应用前景, 特别是应用于集合预报对登陆我国的台风的路径预报有较显著改进。这一方法主要是由中国学者原创并持续发展   |
| 集合卡尔曼滤波 (Ensemble Kalman filter, EnKF)   | 通过观测、初始场和特定误差的分析以估计大气“最佳状态”, 代表了一种资料同化的方法。2005年1月, CMC开始使用集合卡尔曼滤波 (EnKF)   |
| 集合转换卡尔曼滤波 (Ensemble transform Kalman filter, ETKF)                                   | 充分考虑了集合预报系统和数据同化系统的一致性。我国的CMA-REPS使用了该方法   |

## 2.2 资料同化的方法

现代气象业务中资料分析同化的基本方法大致经历了以下发展<sup>[39-40, 42-44]</sup>。多项式函数拟合方法, 由Panofsky 1949年提出<sup>[45]</sup>, 是比较早的客观分析方法。逐步订正方法 (Successive Correction Method, SCM), Bergthorsson等<sup>[46]</sup>, Cressman<sup>[47]</sup>将它发展为一个业务客观分析方案, 用于美国当时的联合数值天气预报。最优插值方法 (Optimal Interpolation, OI), 经Gandin<sup>[48]</sup>全面地研制和开发, 并应用到苏联的客观分析中。变分方法 (Variational methods, Var) 如3DVar/4DVar是国际上大多数主要业务数值预报中心正在使用或使用过的业务方案<sup>[49-50]</sup>。集合卡尔曼滤波方法 (Ensemble Kalman Filter, EnKF) 在1994年由海洋学者Evensen<sup>[44]</sup>引入到资料同化领域, 逐渐成为同化技术发展的热点<sup>[17, 43]</sup>。

近年来, 中国科学家也在资料同化方法研究中以取得了一些代表性成果<sup>[24]</sup>, 如Wang等<sup>[51]</sup>提出的降维投影四维变分同化 (DRP-4DVar), Tian等<sup>[52]</sup>提出的集合四维变分同化方法 (NLS-En4DVar)。此外, 在业务同化系统中Han等<sup>[53-54]</sup>发展的有约束的卫星资料偏差订正技术 (CBC, Constrained Bias Correction) 作为重要的原创技术在CMA-GFS中取得显著成效, 并被ECMWF的同化系统引进和发展。尽管同化理论和技术方法也早已“今非昔比”取得了巨大进步, 但在数

值预报业务中的资料同化仍然面临一些问题, 比如观测算子、偏差、误差、代表性、稀疏化、背景误差协方差、强/弱约束、伴随问题等交织在观测、同化框架和模式约束等多个方面<sup>[55]</sup>。对这些问题的持续改善也期待着新的技术方法的出现和成熟应用。

## 2.3 ECMWF 资料同化新方向

ECMWF作为全球领先的气象中心, 也是资料同化方面的技术研发和业务应用的佼佼者。ECMWF第一个在业务中实践四维变分同化 (4DVAR) 技术, 其早在1987年制定的4年发展计划时开始酝酿用4DVAR来取代最优插值方法, 此后历经十年努力, 1997年11月实现了业务化运行<sup>[42]</sup>。自那时起ECMWF一直使用增量 (Incremental) 4DVar作为其同化系统核心算法<sup>[56]</sup>。

为了更加充分地利用当前的全球观测系统 (GOS) 产生连续的观测流, 减少“等待观测数据到达”的时间, 2019年ECMWF在同化系统中引入连续数据同化技术 (Continuous Data Assimilation)<sup>[56]</sup>。连续数据同化的优势是不必等待所有观测结果到达后才开始计算, 计算阶段与数据收集阶段重叠 (图3)。在运行中, 不再要求在固定的截止时间后停止观测进入, 而是在连续的外循环之间允许新的观测进入同化系统。由于每个外循环大约需要15 min才能完成, 因此在连续数据同化中, 有效截止时间延长了约25 min。在此框架

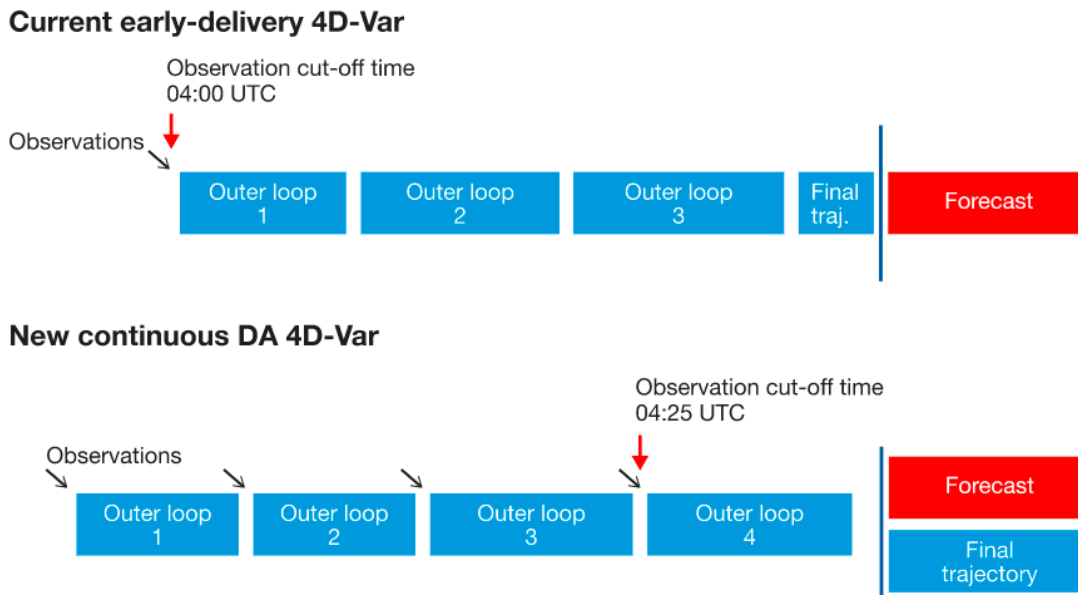


图3 早期的四维同化和当前四维连续数据同化示意 (黑色箭头表示在4D-Var在各阶段计算时获取的观测。蓝色竖线表示预报开始的时间并没有受到影响)<sup>[56]</sup>

Fig. 3 Schematic representation of the current early-delivery 4D-Var configuration and the proposed continuous DA 4D-Var configuration. (The black arrows indicate at which stage of the 4D-Var computations new observations are ingested. The blue vertical line denotes the time when the forecast computations start, which is unchanged)<sup>[56]</sup>

下可以利用比当前系统晚大约1.5 h的观测，以及在截止时间之后到达的观测。实验证实，连续数据同化反映能够更加充分的利用观测数据并产生更准确的分析结果，系统的可预测性提高了2~3个小时<sup>[56]</sup>。

此外，为改善海洋-大气之间不同系统同化时的初始化冲击，ECMWF还在积极推进耦合同化技术研发与应用。耦合数据同化能够使不同时间尺度和同化方法的地球系统各组成部分取得最大的一致性。ECMWF在2018年6月(IFS Cycle 45r1)和2019年6月(IFS Cycle 46r1)的两次系统升级中引入了大气、海洋和海冰的弱耦合数据同化。实验证实，弱耦合海洋-大气数据同化对热带和极地地区温度和湿度等大气变量有显著的改善<sup>[57]</sup>(图4)。

### 3 讨论及总结

自20世纪90年代以来世界主要气象强国开始投入了比以往更多的人力、财力、物力致力于加速发展数值预报业务体系，并取得显著成效<sup>[17]</sup>。然而随着集合预报、数据同化理论和方法的日臻成熟，在现有科学范式下难以或者需要花费更多的代价才能取得新的重要突破。初值难以精确，模式很难完美，可预报性问题短期内尚不能显著改善，数值天气预报仍然面临诸多难题<sup>[4, 58-59]</sup>。除了科学认识尚有不足，随着模式分辨率的提升、集合预报成员的增加，随之所需的计算资

源和电力消耗呈指数级增长<sup>[4]</sup>，这些技术、经济上的挑战也是业务数值预报系统发展不可忽视的因素。

由于传统方法短期内难以取得显著进展，世界气象强国更加重视新兴技术在数值预报领域的研究和业务布局<sup>[37, 60]</sup>。随着新一代信息技术和智能算法技术的不断进步，人工智能推动科技、产业发展和社会变革的巨大潜力得到全球更加广泛的认同。鉴于大气科学显著的非线性特征和当前气象领域典型的大数据特征<sup>①</sup>，因此利用机器学习、深度学习等人工智能的方法，借助海量的观测、模式数据来提升预报预测能力受到了前所未有的关注并取得了积极进展<sup>[61-70]</sup>。全球主要气象机构也敏锐地捕捉到了这一发展战略机遇，并逐渐将其视为气象科技进步的重要竞争领域之一。比如，ECMWF、美国国家海洋与大气管理局(NOAA)等竞相在其最新的发展战略中除了持续推进地球系统数值模式、资料同化等传统优势领域的布局外还进一步强化了人工智能等新技术在气象领域的科学研究与业务实践<sup>[71-72]</sup>。

大气科学的每一次飞跃都伴随着与其他科学技术进步的融合发展。集合、同化思想在大气科学中的渗透、应用和发展极大地推动了以数值预报技术为核心的现代气象业务发展。当前人工智能等新一代信息技术在气象及地球系统领域的融合应用也许能够较快

① 海量观测和数值模式数据，具备典型的大数据特征，即规模性(Volume)、多样性(Variety)、高速性(Velocity)和价值性(Value)。

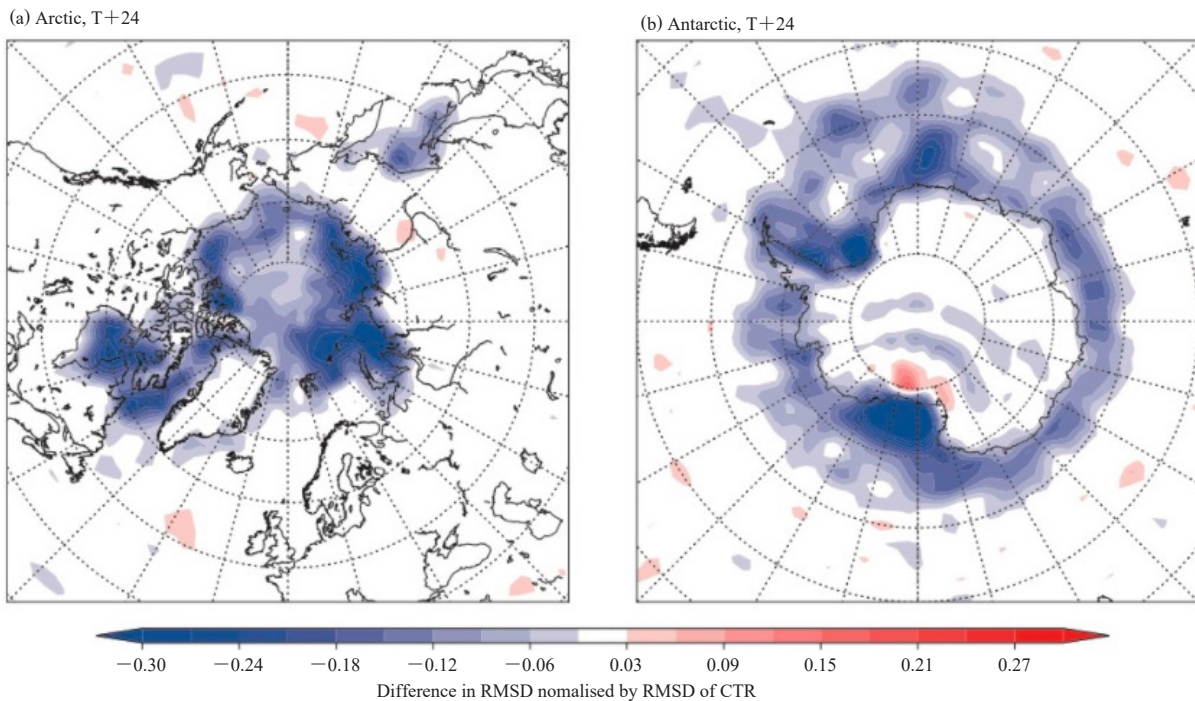


图4 北极 (a) 和南极 (b) 2017年6月9日—2018年5月21日耦合同化试验 (蓝色阴影表示使用弱耦合数据同化时预测与分析之间的差异较小)<sup>[57]</sup>

Fig. 4 The coupled data assimilation tests in the Arctic (a) and the Antarctic (b), for the period 9 June 2017 to 21 May 2018. (Blue shades mean that the differences between forecasts and the analysis are smaller when weakly coupled data assimilation is used)<sup>[57]</sup>

带来新的技术突破并大气科学的发展注入更多的创新源泉。

#### 参考文献

- [1] 约翰·M.华莱士,彼得·V.霍布斯,何金海,等. 大气科学(第2版): 中文版. 北京: 科学出版社, 2008.
- [2] 杨萍, 叶梦姝, 陈正洪. 气象科技的古往今来. 北京: 气象出版社, 2014.
- [3] 张静. 气象科技史. 北京: 科学出版社, 2015.
- [4] Bauer P, Thorpe A, Brunet G. The quiet revolution of numerical weather prediction. *Nature*, 2015, 525(7567): 47-55.
- [5] Alley R B, Emanuel K A, Zhang F. *Advances in weather prediction*. *Science*, 2019, 363(6425): 342-344.
- [6] ECMWF. 30 Years of ensemble forecasting and symposium for Prof. Tim Palmer. <https://events.ecmwf.int/event/315/>.
- [7] 吴增茂, 张新玲. 大气系统集合预报方法及应用研究的发展. *气象科技*, 2000, (1): 16-20.
- [8] 王大微, 陈德辉. 数值预报发展的新方向: 集合数值预报. *气象研究与应用*, 2007 (1): 6-12.
- [9] 段明铿, 王益兴. 集合预报方法研究及应用进展综述. *南京气象学院学报*, 2004 (2): 279-288.
- [10] 段晚锁, 汪叶, 霍振华, 等. 数值天气预报和气候预测的集合预报方法: 思考与展望. *气候与环境研究*, 2019 (3): 396-406.
- [11] 李泽椿, 陈德辉. 国家气象中心集合数值预报业务系统的发展及应用. *应用气象学报*, 2002 (1): 1-15.
- [12] 陈静, 陈德辉, 颜宏. 集合数值预报发展与研究进展. *应用气象学报*, 2002 (4): 497-507.
- [13] 陈静, 李晓莉. GRAPES全球/区域集合预报系统10年发展回顾及展望. *气象科技进展*, 2020 (2): 9-18.
- [14] Epstein E S. Stochastic dynamic prediction. *Tellus*, 1969, 21( 6): 739-759.
- [15] Leith C E. Theoretical skill of Monte Carlo forecasts. *Monthly Weather Review*, 1974, 102(6): 409-418.
- [16] 关吉平, 张立凤, 张铭. 集合预报研究现状与展望. *气象科学*, 2006 (2): 228-235.
- [17] 矫梅燕. 现代数值预报业务. 北京: 气象出版社, 2010.
- [18] 郑峰. 集合预报初值扰动在天气预报中的应用研究进展. *科技导报*, 2008 (19): 90-95.
- [19] 刘东贤, 张庆阳. 集合预报及其发展趋势. *中国科技信息*, 2007 (20): 268.
- [20] Roberto B, David R. 25 years of ensemble forecasting at ECMWF. <https://www.ecmwf.int/en/newsletter/153/meteorology/25-years-ensemble-forecasting-ecmwf>
- [21] WGENE Overview of Plans at NWP Centres with Global Forecasting Systems-2021. <https://wgne.net/nwp-systems-wgnetable/wgne-table/>.
- [22] 顾震潮. 我国数值预报的成就. *气象学报*, 1959 (3): 236-242.
- [23] Blumen W, Washington W M. Atmospheric dynamics and numerical weather prediction in the People's Republic of China. *Bulletin of the American Meteorological Society*, 1973, 54(6): 502-518.
- [24] 沈学顺, 王建捷, 李泽椿, 等. 中国数值天气预报的自主创新. *气象学报*, 2020 (3): 451-476.
- [25] 张志刚. 中国气象局数值预报业务发展概述. 2022年研究型业务(资料同化)培训班, 2022.
- [26] 陈静. 业务集合预报系统. 2022年研究型天气业务(数值预报)培训班, 2022.
- [27] 陈正洪, 丁一汇, 许小峰. 20世纪数值天气预报主要阶段与关键创新. *广西民族大学学报: 自然科学版*, 2016 (4): 28-33.
- [28] 杜钧, 李俊. 集合预报方法在暴雨研究和预报中的应用. *气象科技进展*, 2014 (5): 6-20.
- [29] 麻巨慧, 朱跃建, 王益兴, 等. NCEP、ECMWF及CMC全球集合预报业务系统发展综述. *大气科学学报*, 2011 (3): 370-380.
- [30] Mu M, Duan W S, Wang B. Conditional nonlinear optimal perturbation and its applications. *Nonlin. Processes Geophys.*, 2003, 10: 493-501
- [31] 穆穆, 周非凡. 基于CNOP方法的台风目标观测研究进展. *气象*

- 科技进展, 2015 (3): 6-17.
- [32] 高丽, 陈静, 郑嘉雯, 等. 极端天气的数值模式集合预报研究进展. 地球科学进展, 2019, (7): 706-716.
- [33] Houtekamer P L, Lefaiyre L, Derome J, et al. A system simulation approach to ensemble prediction. *Mon. Wea. Rev.*, 1996, 124(6): 1225-1242.
- [34] 陈超辉, 王勇, 杜钧, 等. 欧洲业务集合预报系统进展. 气象科技进展, 2020 (2): 19-29.
- [35] Mark R, Michael D, Düben P, et al. IFS upgrade provides more skilful ensemble forecasts. <https://www.ecmwf.int/en/newsletter/168/meteorology/ifs-upgrade-provides-more-skilful-ensemble-forecasts>.
- [36] ECMWF. Implementation of IFS cycle 48r1. <https://confluence.ecmwf.int/display/FCST/Implementation+of+IFS+Cycle+48r1>.
- [37] ECMWF. Strategy 2021-2030. <https://www.ecmwf.int/en/elibrary/19880-ecmwf-strategy-2021-2030>.
- [38] Bjerknes V. Das Problem der Wettervorhersage, betrachtet vom Standpunkte der Mechanik und der Physik. *Meteor. Z.*, 1903, 21.
- [39] 朱国富. 数值天气预报中分析同化基本方法的历史发展脉络和评述. 气象, 2015 (8): 986-996.
- [40] 朱国富. 理解大气资料同化的内在逻辑和若干共性特征. 气象, 2015 (8): 997-1006.
- [41] Talagrand O. Assimilation of observations, an introduction. *Journal of the Meteorological Society of Japan*, 1997, 75(1): 191-209.
- [42] 龚建东. 同化技术: 数值天气预报突破的关键: 以欧洲中期天气预报中心同化技术演进为例. 气象科技进展, 2013 (3): 6-13.
- [43] 熊春晖, 张立凤, 关吉平, 等. 集合-变分数据同化方法的发展与应用. 地球科学进展, 2013 (6): 648-656.
- [44] 官元红, 周广庆, 陆维松, 等. 资料同化方法的理论发展及应用综述. 气象与减灾研究, 2007 (4): 1-8.
- [45] Panofsky H. Objective weather-map analysis. *Journal of Application Meteorology*, 1949, 6: 386-392.
- [46] Bergthorson P, Dills B R. Numerical weather map analysis. *Tellus*, 1955, 7: 329-340.
- [47] Cressman G P. An operational objective analysis system. *Mon. Wea. Rev.*, 1959, 87: 367-374
- [48] Gandin L S. Objective analysis of meteorological fields. *Israel Program for Scientific Translations*, 1963.
- [49] Barker D, Huang W, Guo Y R, et al. A three-dimensional variational data assimilation system for MM5: implementation and initial results. *Monthly Weather Review*, 2004.
- [50] Rabier F, Järvinen H, Klinker E, et al. The ECMWF operational implementation of four dimensional variational assimilation. Part I: Experimental results with simplified physics <https://www.ecmwf.int/en/elibrary/11794-ecmwf-operational-implementation-four-dimensional-variational-assimilation-part-i>.
- [51] Wang B, Liu J J, Wang S D, et al. An economical approach to four-dimensional variational data assimilation. *Adv Atmos Sci*, 2010, 27(4): 715-727.
- [52] Tian X, Zhang H, Feng X, et al. Nonlinear least squares En4DVar to 4DENVar methods for data assimilation: formulation, analysis, and preliminary evaluation. *Monthly Weather Review*, 2018, 146(1): 77-93.
- [53] Han W. Constrained variational bias correction for satellite radiance assimilation. *Proceedings of the 19th International TOVS Study Conference*, 2014.
- [54] Han W, Bormann N. Constrained adaptive bias correction for satellite radiance assimilation in the ECMWF 4D-Var system. *ECMWF Technical Memorandum*, 783. ECMWF, 2016.
- [55] 韩威. 业务资料同化系统研发中的关键问题. 2022年研究型业务(资料同化)培训班, 2022.
- [56] Peter L, Massimo B, Elias H, et al. Continuous data assimilation for the IFS. <https://www.ecmwf.int/en/newsletter/158/meteorology/continuous-data-assimilation-ifs>.
- [57] Philip B, Patricia R, Hao Z. Coupled ocean-atmosphere data assimilation at ECMWF. <https://www.ecmwf.int/en/newsletter/160/meteorology/coupled-ocean-atmosphere-data-assimilation-ecmwf>.
- [58] Benjamin S G, Brown J M, Brunet G, et al. 100 Years of progress in forecasting and NWP applications. *Meteorological Monographs*, 2019, 59.
- [59] Randall D A, Bitz C M, Danabasoglu G, et al. 100 years of earth system model development. *Meteorological Monographs*, 2019.
- [60] NOAA. A report on priorities for weather research. [https://sab.noaa.gov/wp-content/uploads/2021/11/PWR-Report\\_24Nov2021\\_Draft.pdf](https://sab.noaa.gov/wp-content/uploads/2021/11/PWR-Report_24Nov2021_Draft.pdf).
- [61] 许小峰. 从物理模型到智能分析: 降低天气预报不确定性的新探索. 气象, 2018 (3): 341-350.
- [62] 陈云峰. 20世纪80年代人工智能在天气预报中的应用. 气象科技进展, 2018 (3): 88-91.
- [63] 孙健, 曹卓, 李恒, 等. 人工智能技术在数值天气预报中的应用. *应用气象学报*, 2021 (1): 1-11.
- [64] 唐伟, 周勇, 董昊, 等. 我国气象领域应用人工智能技术的现状和国际对比. 气象科技进展, 2019 (5): 55-56.
- [65] 吴灿, 戴洋, 何晓欢, 等. 气象和大气科学领域人工智能科学研究的国际态势分析. 科学咨询: 教育科研, 2021 (11): 1-5.
- [66] 张萌, 贾朋群, 王小光. 人工智能技术在大气科学领域的应用及其发展态势. 科学观察, 2020 (1): 22-29.
- [67] Reichstein M, Campus-Valls G, Stevens B, et al. Deep learning and process understanding for data-driven Earth system science. *Nature*, 2019.
- [68] Sun Z, Sandoval L, Crystal-Ornelas R, et al. A review of Earth artificial intelligence. *Computers & Geosciences*, 2022.
- [69] Irrgang C, Boers N, Sonnewald M, et al. Towards neural Earth system modelling by integrating artificial intelligence in Earth system science. *Nature Machine Intelligence*, 2021.
- [70] Boukabara S A, Krasnopolsky V, Penny S G, et al. Outlook for exploiting artificial intelligence in the Earth and environmental sciences. *Bulletin of the American Meteorological Society*, 2020: 1-53.
- [71] ECMWF. Machine learning at ECMWF: a roadmap for the next 10 years. <https://www.ecmwf.int/en/elibrary/19877-machine-learning-ecmwf-roadmap-next-10-years>.
- [72] NOAA. Artificial intelligence strategic plan. 2021-2025. [https://sciencecouncil.noaa.gov/Portals/0/Artificial%20Intelligence%20Strategic%20Plan\\_Final%20Signed.pdf?ver=2021-01-19-114254-380](https://sciencecouncil.noaa.gov/Portals/0/Artificial%20Intelligence%20Strategic%20Plan_Final%20Signed.pdf?ver=2021-01-19-114254-380).