

基于CMA-GFS全球模式实现国产数值预报的本地化应用

陈阳权

民航新疆空中交通管理局, 新疆 乌鲁木齐

收稿日期: 2022年8月4日; 录用日期: 2022年9月6日; 发布日期: 2022年9月14日

摘要

通过对CMA-GFS全球模式进行数据解码处理和二次研发, 利用MICAPS4对解码和二次开发数据进行检索和分析, 实现了其在新疆辖区的本地化应用。在实现过程中对CMA-GFS范围进行切割, 采用按需解码、多进程等方案有效提高了处理的效率, 满足预报对时效性的需求。CMA-GFS的解码数据和二次研发数据的种类丰富, 并生成了高空急流、颠簸指数、积冰指数、对流指数等具有航空气象特色的二次产品, 可为航空气象人员提供更专业的数据分析参考。通过与实况和多种全球模式预报进行对比, CMA-GFS的数值预报产品能够较好的描述气压系统和高空环流形势的演变, 作为气象预报参考是很有价值的。

关键词

Python, CMA-GFS全球模式, 数据解码, 二次研发

Localized Application of Domestic Numerical Forecast Based on CMA-GFS

Yangquan Chen

Xinjiang ATMB CAAC, Urumqi Xinjiang

Received: Aug. 4th, 2022; accepted: Sep. 6th, 2022; published: Sep. 14th, 2022

Abstract

Analysis and retrieval on MICAPS 4 through data decoding and secondary development of the CMA-GFS to realize localized application in Xinjiang. Cutting the range of CMA-GFS in the implementation process. The efficiency of processing is improved by on-demand decoding and multi-process to meet the demand for period of validity. There are various types of decoding data and secondary

R&D data of CMA-GFS, and generating secondary products with aviation meteorological characteristics such as high-altitude jet, turbulence index, icing index, convection index, etc., which can provide aviation meteorologists with more professional data analysis reference. Finding NWP products of CMA-GFS can better describe the evolution of the barometric system and the upper-air circulation situation by comparing it with the live and various global model forecasts. It is very valuable as a reference for weather forecast.

Keywords

Python, CMA-GFS, Data Decoding, Secondary R&D

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

数值预报是现代天气预报业务的基础, 无论客观预报还是人工预报, 均离不开数值预报的支撑, 如在降水、台风、强对流、空气质量等预报应用[1] [2] [3] [4], 还可用于风电风能预测、与其它系统融合分析[5] [6]等。目前气象业务主要的数值预报包括全球数值预报和中尺度数值预报, 常见的全球模式预报如欧洲中心全球模式预报(即 EC 细网格)、美国 NOAA 全球模式预报系统(GFS)、世界航空预报中心提供的全球模式预报(WAFS)、中国的全球模式预报系统(CMA-GFS)等, 常见的中尺度模式如 WRF、MM5 以及以此衍生而出其它中尺度模式预报系统等。

作为现代气象事业发展的最核心技术, 数值预报堪称气象领域的“国之重器”。但就新疆空管局气象部门而言, 业务中主要参考数值预报模式有 EC 细网格、GFS、WAFS 等, 如陈阳权等[7]利用 WAFS 全球预报产品构建了航路预报天气的客观预报系统, 为航空气象预报业务提供了有益的参考; 朱国栋等[8]利用随机深林对 EC 细网格数据进行释用, 为民航机场提供客观要素的指导预报。可见就新疆空管气象而言, 大量的研究还是以欧美模式为基础, 累积的预报经验也主要是这些模式, 缺乏国产数值预报的使用经验。而这些数值预报可能因各种不确定因素存在无法获取的风险, 可见研究和应用国产化的数值模式, 使用具有自主知识产权的数据, 既可牢牢牵住自主的“牛鼻子”, 摆脱受制于人的困境, 也可逐渐积累应用经验。

CMA-GFS (原名为 GRAPES-GFS)作为我国自主研发的新一代全球数值预报模式, 从初始场、变分同化、模式框架等核心技术均实现了自主化, 从 2016 年 6 月正式业务化来, 国内专家学者对其进行了大量的研究和应用, 如孔祥伟[9]、陈晓燕[10]等通过将 GRAPES-GFS 与其它全球数值预报模式结果进行对比分析, 得出 GRAPES-GFS 所欠缺的方面, 为模式改进指明方向; 李玉林[11] [12]等利用 GRAPES 数值模式计算结果来分析大气中水汽比含量特征, 在人工增雨业务中进行应用; 齐倩倩等[13]基于 GRAPES-GFS 进行误差诊断和预报能力分析表明该模式对东亚的预报技巧较高。可见, 通过持续的研究和应用, 经过多年的改进与实践, CMA-GFS 已经成为一套可信赖且完备的全球数值预报系统。

本文以 CMA-GFS 全球模式数据为基础, 通过对原始的模式要素进行解码, 并对多种对流指数、小时降水量、流场、急流、变温、变压、颠簸指数、积冰指数等多种物理要素进行二次计算, 将数据存储为 MICAPS4 的格点数据格式, 可使用 MICAPS 系统进行数据检索和查看。

2. CMA-GFS 介绍

CMA-GFS 全球模式从 2001 年开始组织数值预报团队启动研发, 2016 年实现业务化并面向全国下发产品。该模式每天运行 4 次, 分别是 00、06、12、18 UTC, 时间分辨率为 3 小时, 空间分辨率为 $0.25^{\circ} \times 0.25^{\circ}$, 预报时长为 10 天(240 小时), 提供了全球基础要素场和诊断要素场, 基础要素场主要包括各层次的高度场、温度场、东西风场、湿度场、比湿场、露点场及海平面气压场、10 m 东西风场、2 m 温度场、2 m 比湿场、2 m 露点场、降水量等, 诊断要素场主要包括 10 m 阵风、能量指数、最高最低温度、云量及各层次的涡度散度、垂直速度及相当位温等。单个文件约 430 M 左右, 每个起报时次接收到的数据量约 25 G 左右。

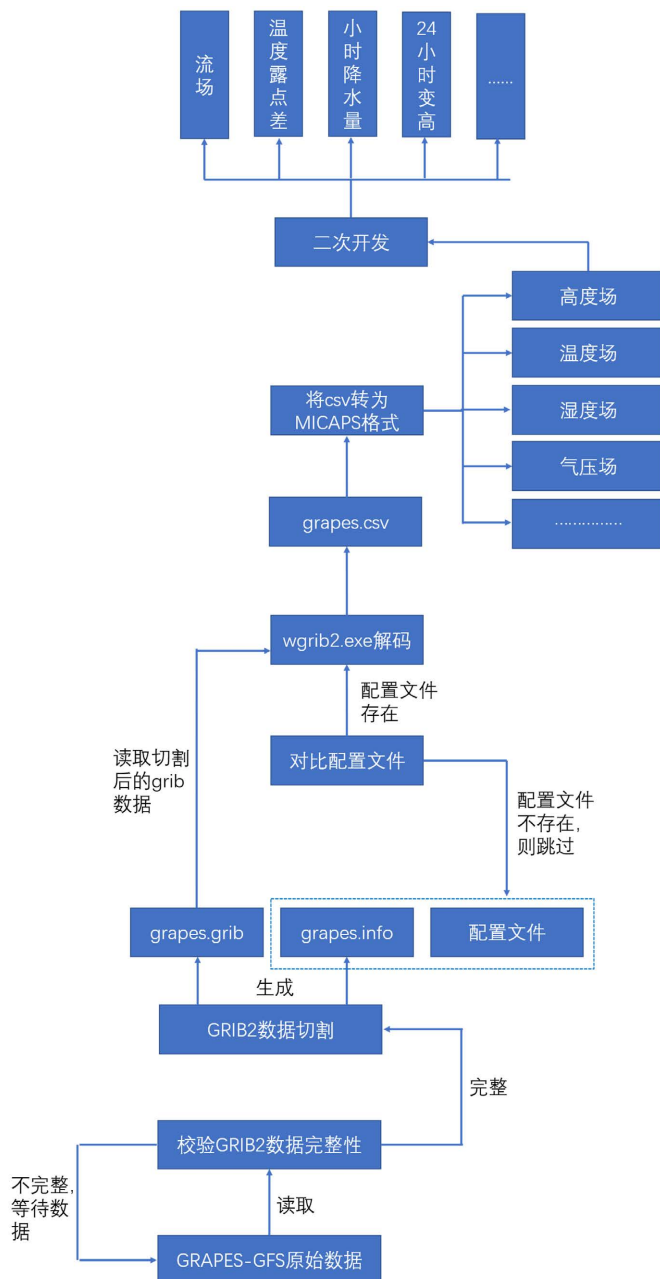


Figure 1. Data processing flow of a single process
图 1. 单个进程的数据处理流程

3. 本地化实现方案

新疆空管局接收到的 CMA-GFS 原始数据为 grib2 格式的格点数据，在实际使用中需要将 grib2 数据进行解码处理，将其转化为 MICAPS 格式产品进行使用。利用 MICAPS 系统的先天优势，可方便的将其与其它数值预报模式数据进行对比分析。

本方案使用 Python 语言设计解码和二次计算的核心程序。由于 CMA-GFS 数据量大，解码过程耗时，在设计解码程序时采用要素配置文件来驱动数据解码，并以多进程方式来提高数据处理的效率。

如图 1 所示为整个数据处理单进程的流程，当 CMA-GFS 原始数据到达本地后，本地处理程序轮询获取数据列表，读取数据列表的文件名和大小进行初步校验，将不符合初步判断的数据剔除列表，对通过初步校验的文件利用 wgrib2 程序判断数据完整性，将校验未通过的数据从数据列表中再次剔除，最终得到完整可用的数据列表，为解码做好准备，同时将校验通过的数据列表写入本地文件，下一次解码时已经解码的文件将不再处理。通过校验的数据进入解码环节，首先对解码数据的范围按实际需要进行切割，将切割后的 grib2 数据和数据索引存储在本地，后续对该 grib2 数据进行解码操作。在解码操作前，先读取解码要素配置文件与切割后的数据索引对比，对匹配的数据调用 wgrib2 应用程序将该要素解码为 csv 格式数据，并将 csv 格式数据写为 MICAPS 的第 4 类格式，从而得到原始数据中的要素产品，可在 MICAPS 中进行查看。基于 MICAPS 格式的基本要素数据，计算得到颠簸、积冰、对流、流场等二次产品，即可完成单个文件单个进程的数据解码。

3.1. CMA-GFS 原始文件的完整性校验

在进行解码前，需要校验解码原始数据文件的完整性，若原始数据文件不完整，则实际解码时跳过不完整的文件，可以有效防止在二次开发时数据处理错误的风险。数据校验可简单的采用 2 种方式，一是检测数据文件大小，CMA-GFS 单个文件约 430 M，若文件小于 400 M，可认为是不完整的数据；二是通过 wgrib2 来检测，使用“wgrib2 filename”命令，通过分析命令输出来判断数据完整性，若数据不完整，则输出的字符串中将出现“FATAL ERROR”，分析字符串是否出现“FATAL ERROR”即可判断文件的完整性。由于第二种方法 wgrib2 需要读取整个文件，效率较低，在实际业务运行中，可先判断文件大小筛除不符合大小的数据，若大小符合，再通过 wgrib2 的命令做进一步的判断。主要实现 Python 代码如下：

```
def check_data(file,size):
    """
    校验文件完整性,size 是文件大小最低限值
    """
    filesize=os.path.getsize(file)/1024/1024
    if filesize<size:
        logs(file +"文件大小与正常相差较大，不进行该文件处理！ \n")
        return (False,file)

#通过管道读取 wgrib2 的输出
    p=os.popen('wgrib2.exe '+file)
    pstr=p.read()
    if re.search(r'FATAL ERROR',pstr):
```

```
logs(file+"文件有错误或数据未取全！不进行处理\n")
    return (False,file)
```

```
logs(file+"文件校验成功*****\n")
    return (True,file)
```

3.2. 解码数据预处理

由于 CMA-GFS 是全球模式预报数据，其范围覆盖了全球区域，在实际应用时不需要查看全球的数据，对新疆而言，纬度范围东经 20°~70°、经度范围北纬 30°~140°的区域已经能够满足日常业务的需求，为了提高数据处理的效率，节省数据处理时间和存储空间，可按这个纬度进行切割，对于其它地区，可按需要给定经纬度进行切割。解码时只对切割后的区域格点数据进行解码处理，可大大提高数据处理的效率。

切割数据使用 wgrib2 进行，通过给定经纬度范围，调用 wgrib2 的“small_grib”即可生成给定经纬度的 grib2 格式数据。实现的主要 Python 代码如下：

```
def split_grib2(source_file,write_file,griblog_file,lat_limit, lon_limit):
```

```
    """
```

```
        将 grib2 数据进行切割，按经纬度经纬度切割后生成相应的文件和索引；
```

```
source_file 为源数据文件；
```

```
write_file 为切割后的文件；
```

```
lat_limit 为纬度范围如 0:90；
```

```
lon_limit 为经度范围如 0:180；
```

```
    返回值为 grib2 文件数据块数。
```

```
    """
```

```
    # 调用 wgrib2 的 API
```

```
os.system("wgrib2 "+source_file+" -small_grib "+lon_limit +
          " "+lat_limit+" "+write_file+">"+griblog_file)
```

```
infos = []
```

```
    with open(griblog_file, "r") as f:
```

```
infos = f.readlines()
```

```
    print(source_file+"数据共有"+str(len(infos))+“块数据”)
```

```
    return len(infos)
```

3.3. 配置文件驱动解码过程

在实际做数据处理前，通过 wgrib2-s 可查看 CMA-GFS 数据文件中所包含的要素信息，从要素信息中明确需要解码的要素数据，按这种方式按需解析数据，可大幅度提高数据处理的效率。在正式解码前，需要读取已经配置好的要素配置文件，并与在文件切割时生成 griblog_file 文件进行对比，精确查找需要解码的数据块。通过这种配置文件驱动的方式，可以按需解码需要的数据。配置文件的格式如下：

```
#要素表-要素名称:要素要素层次说明:要素说明(无以 1 个空格表示):micpas 说
```

```
#明:micaps 线条间隔:
#micaps 输出要素目录:原始输出还是二次开发(0 位原始数据, 1 为二次计算)
UGRD:10 m above ground: :10 米风 U 分量:4:U10M:0
VGRD:10 m above ground: :10 米风 V 分量:4:V10M:0
GUST:10 m above ground: :10 米阵风:4:GUST10M:0
VIS:surface: :能见度(km):4:VIS:0
```

以 VIS 为例,“VIS”为 CMA-GFS 的要素名,“surface”为要素所在的层次,“能见度(km)”为解码后写为 MICAPS 格式时的要素说明,“4”为 MICAPS 中等值线间隔,后一个“VIS”为输出的数据目录,0 表示原始数据解码。

读取配置文件的要素配置后,若在 griblog_file 中存在该要素,则对该要素进行解码。这里通过 Python 调用 wgrib2 的接口“-csv”,将该要素解码为 csv 数据格式作为中间数据存储在本地,然后读取 csv 中间数据,将该要素写为 MICAPS4 的第四类数据格式,即完成了该要素的解码工作。接口调用如下:

```
os.system(r"wgrib2 "+gribfile+r" -match ':'+reg+r"':'-csv "+csvfile)
```

其中 gribfile 即切割后的 grib2 格点数据,reg 为要素在 griblog_file 中的匹配字段,csvfile 即存储解码后的中间文件路径。

3.4. 要素二次开发计算

CMA-GFS 的原始数据要素已经能够满足大部分天气分析工作,如高度场、温度场、气压场、风场、湿度场、涡度、散度、稳定度、降水等要素数据。但在实际业务中,还需要对一些要素进行处理,以得到要素的合成量如流场、要素的变化量如变高、变温、变压等以及其它诊断量如对流指数、高空急流区、颠簸指数、积冰指数、逐 3 小时降水(总降水、大尺度降水及对流降水)等,这些诊断量为天气分析更加充分数据支撑。

3.5. 多进程处理

由于 CMA-GFS 数据量大,数据处理需要消耗大量的时间,若采用单个进程处理,数据处理效率较低。数据解码工作是典型的 IO 密集型计算,每个解码文件可作为单个进程进行解码,无需在不同文件之间共享数据,因此可将多个计算分配到不同的进程中同时处理。目前主流的计算机均具备多个处理器,若采用多进程同时处理数据的方式,可大大的提高数据处理的效率,节省数据处理时间。多进程处理时以计算机自带核心数减一的方式设计多进程处理方案,如 4 个处理器的计算机,则使用 3 个处理器做计算。在 Python 中的具体实现如下:

```
import multiprocessing
from functools import partial

def wgrib_func(latlimt,lonlimt,cfglist,api,file):
    """
    多进程调用解码函数
    """
    wgrb=wgrib.wgrib2()

    if __name__ == '__main__':
```

```

# 采用 4 个进程
pool = multiprocessing.Pool(processes = 4)
#使用偏函数向 wgrib_func 传入多个参数
pfunc = partial(wgrib_func, latlimt,lonlimt,cfplist,api)
# 使用 map 传入文件列表, map 阻塞进程
init_times=pool.map(pfunc,filelist)
pool.close()
pool.join()

```

上述 `wgrib_func` 函数为多进程调用的解码函数, `processes = 4` 即处理器核心数减一, 可按实际计算机的处理器进行设置, 或者使用 `multiprocessing.cup_count()-1` 进行计算。由于需要传入多个参数, 可使用偏函数 `partial()` 将 `wgrib_func` 前几个参数传入, 最后使用 `pool.map()` 传入最后的解码文件列表即可将任务分配到不同的处理器上处理, 可大大提高数据处理的效率。

4. CMA-GFS 全球模式产品分析

利用 Python 对 CMA-GFS 全球模式产品进行解码和二次计算可得到各类常规要素场、诊断要素场、航空气象要素场等预报产品, 这些预报产品在实际工作中可进行参考, 并与其它全球模式数据及实况数据进行对比分析, 通过对比分析和实时检验来了解 CMA-GFS 模式的预报性能, 累积预报经验。

图 2(a) 为 2021 年 10 月 13 日 08 时, CMA-GFS 的 500 hPa 分析场与实况的 500 hPa 高度场, 由图可见分析场与实况非常吻合, 分析场对欧亚大陆槽脊系统及西太平洋副高的描述与实际基本一致。图 2(b) 为 20 时的 500 hPa 高度场预报与实况的对比, 可见除了对高原西侧槽脊系统有一定的偏差外, 对欧亚中高纬槽脊系统及西太副高的预报与实况基本一致。

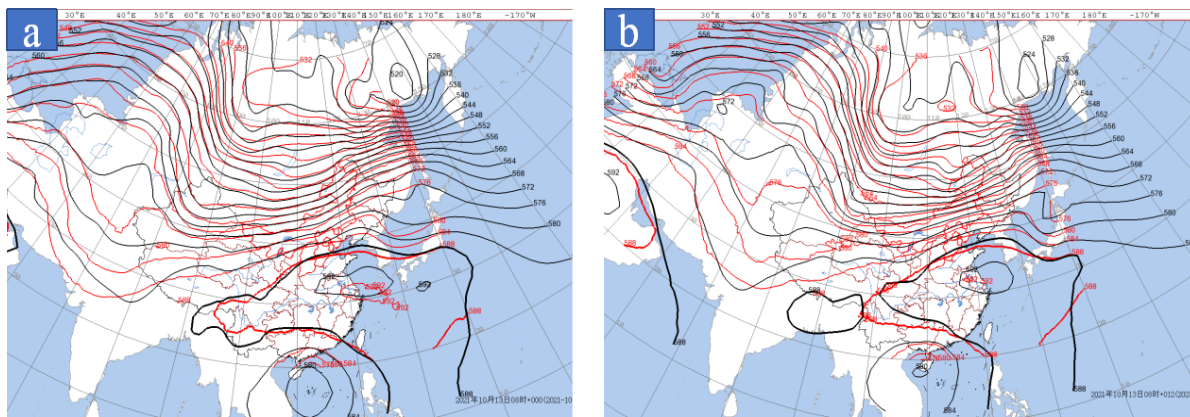


Figure 2. Comparative analysis of CMA-GFS and real 500 hPa height field (a) is the analysis field at 08:00, (b) is the forecast field at 20:00

图 2. CMA-GFS 与实况 500 hPa 高度场对比分析 (a) 为 08 时分析场, (b) 为 20 时预报场

图 3(a) 2021 年 10 月 13 日 08 时, CMA-GFS 分析场与实况的海平面气压对比分析, 由图可见分析场的高低形势与实况的形势基本吻合, 其对气压系统的强度和范围描述与实际基本一致。图 3(b) 为 20 时预报的海平面气压场与实况的对比分析, 由图可见高低形势的预报与实际比较吻合, 强度略在东部略偏弱, 西伯利亚高压强度和范围基本一致。

图 4 为 2022 年 8 月 16 日起报的 CMA-GFS、GFS、EC_THIN 等全球模式预报的 24 小时和 48 小时

位势高度预报, 以及与实况的对比分析。由图可见, 24 小时和 48 小时各模式预报对中高纬度地区的描述是基本一致的, 且对槽脊系统的位置和强弱预报与实况也基本一致。尤其是对影响我国的天气系统的预报 CMA-GFS 与其它模式预报基本一致, 对高原西部及中亚地区的低值系统预报略有差异。

通过上述对比分析发现, CMA-GFS 全球模式具备一定的预报性能, 可作为气象预报分析的有效参考。

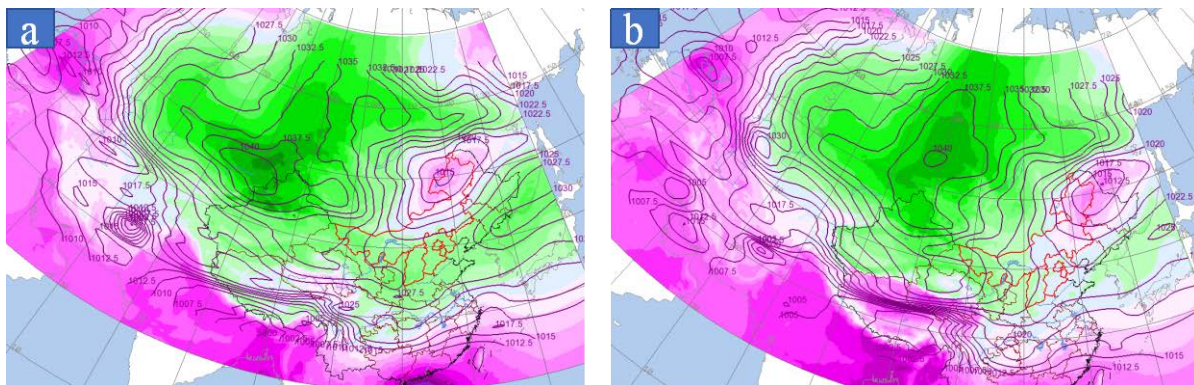


Figure 3. Comparative analysis of CMA-GFS and real sea level pressure field (a) is the analysis field at 08:00, (b) is the forecast field at 20:00

图 3. CMA-GFS 与实况海平面气压场的对比分析(a)为 08 时分析场, (b)为 20 时预报场

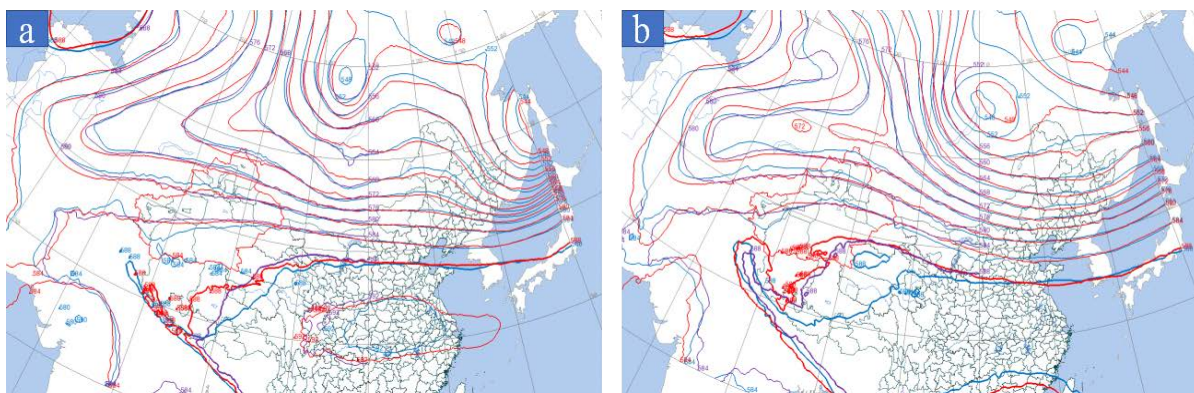


Figure 4. 24 hours (a) and 48 hours (b) 500 hPa geopotential height forecast of CMA-GFS (red contour), GFS (blue contour) and EC_THIN (purple contour) reported from 20:00 on August 16, 2022 and the corresponding live (black contour)

图 4. 2022 年 8 月 16 日 20 时起报的 CMA-GFS (红色等值线)、GFS (蓝色等值线)、EC_THIN (紫色等值线) 24 小时(a) 及 48 小时(b)的 500 hPa 位势高度预报以及对应的实况(黑色等值线)

5. 结论与讨论

通过对 CMA-GFS 全球模式进行数据解码处理和二次研发, 利用 MICAPS4 对解码和二次开发数据进行检索和分析, 实现了其在新疆辖区的本地化应用。在实现过程中通过对 CMA-GFS 范围进行切割、采用配置文件来驱动要素进行按需解码、多进程等方案, 有效提高了处理的效率, 满足预报对时效性的需求。CMA-GFS 的解码数据和二次研发数据的种类丰富, 并生成了高空急流、颠簸指数、积冰指数、对流指数等具有航空气象特色的二次产品, 可为航空气象人员提供更专业的数据分析参考。通过与实况和多种全球模式预报进行对比, CMA-GFS 的预报产品能够较好的描述气压系统和高空环流形势的演变, 作为气象预报参考是有价值的。

CMA-GFS 作为具有自主知识产权的国产数值预报, 经过多年的发展改进和应用, 已经能够为广大气象预报服务机构提供完备的数值预报产品。随着不断的应用、检验和改进, CMA-GFS 的预报性能定能不断提升, 广大预报人员也能够不断的累积该模式的使用经验, 届时才能够真正摆脱受制于人的困境。

参考文献

- [1] 张宸赫, 赵天良, 陆忠艳, 王东东, 阎琦, 周春晓, 苏航, 杜傢义. 基于数值模式和客观方法的辽宁地区空气质量预报检验研究[C]//中国环境科学学会. 中国环境科学学会 2022 年科学技术年会论文集(三). 2022: 458-469. <https://doi.org/10.26914/c.cnkihy.2022.016094>
- [2] 熊杰, 傅赐福, 于福江, 董剑希. 基于 ECMWF 细网格风场的台风“烟花”与飓风“艾达”风暴潮数值预报检验与分析[J]. 海洋预报, 2022, 39(3): 1-9.
- [3] 丛芳, 陈朝平. 四川两次副高边缘型暴雨的预报偏差分析及模式检验对比[J]. 中低纬山地气象, 2022, 46(3): 39-46.
- [4] 林振敏, 黄荣, 戚云枫, 曾小团. 广西对流尺度数值天气预报模式系统建设与效益评估[J]. 气象研究与应用, 2022, 43(2): 105-110. <https://doi.org/10.19849/j.cnki.CN45-1356/P.2022.2.18>
- [5] 周少辉. 荒漠草原地形风能资源评估和风速预报订正研究[D]: [硕士学位论文]. 南京: 南京信息工程大学, 2022. <https://doi.org/10.27248/d.cnki.gnjqc.2022.000397>
- [6] 王俊超, 王志斌, 赖安伟, 肖艳姣, 王珏. 基于数值预报与雷达外推预报融合方法的降水回波短临预报试验研究[J]. 干旱气象, 2022, 40(3): 485-499.
- [7] 陈阳权, 杜安妮, 丁旭, 张茜. 新疆区域航路天气要素分析及预报系统构建[J]. 气象水文海洋仪器, 2022, 39(2): 93-95. <https://doi.org/10.19441/j.cnki.issn1006-009x.2022.02.013>
- [8] 朱国栋, 孙建杰, 陈阳权, 王照刚. 随机森林方法在机场温度预测中的应用[J]. 气象科技进展, 2021, 11(4): 34-36.
- [9] 孔祥伟, 张君霞, 杨晓军, 符娇兰. 西北地区东部强降水大尺度数值模式预报空间偏差分析[J/OL]. 高原气象, 1-15. <http://kns.cnki.net/kcms/detail/62.1061.p.20220419.1816.014.html>, 2022-07-31.
- [10] 陈晓燕, 孔祥伟, 彭筱, 刘新伟, 吴晶, 任淑媛. 全球和区域数值模式在甘肃 2020 年汛期降水预报中的检验评估[J]. 干旱气象, 2022, 40(3): 524-535.
- [11] 李玉林, 周国强, 黄小晁. 利用 GRAPES 数值模式计算结果分析大气中水汽比含量特征[C]//中国气象学会. 第 28 届中国气象学会年会——S9 大气物理学与大气环境. 2011: 578-590.
- [12] 李玉林. 数值模式(GRAPES)在人工增雨作业应用研究[Z]. 江西省人工影响天气领导小组办公室, 2012-05-07.
- [13] 齐倩倩, 朱跃进, 陈静, 田华, 佟华. 基于 GRAPES-GFS 次季节预报的误差诊断和预报能力分析[J]. 大气科学, 2022, 46(2): 327-345.