

ERA-Interim和ERA5再分析数据在江苏区域的适用性对比研究

吕润清¹, 李响²

(1. 江苏省气象台, 江苏 南京 210008; 2. 国家海洋环境预报中心, 北京 100081)

摘要: 基于江苏省73个国家级自动气象观测站和3个高空探测站的观测数据, 对比研究了ERA-Interim和ERA5两套再分析数据在江苏区域的适用性。结果表明: 无论是地面气象要素还是高空气象要素, ERA5再分析数据的适用性均优于ERA-Interim再分析数据, 特别是ERA5再分析数据中的2 m气温及2 m相对湿度相对于ERA-Interim再分析数据体现出系统性的改进, 相关系数的提升超过0.4。两套数据的小时降水量与观测之间均存在相对较大的误差, 在实际使用中需谨慎。ERA5再分析数据与观测数据在一致性及误差方面存在较为显著的时空变化特征, 该数据在夏季的适用性低于其他季节, 在空间上夏半年体现了明显的南北分布特征, 与之对应, ERA-Interim再分析数据没有显著的时空分布特征。采用新一代的ERA5再分析数据对江苏区域的天气/气候进行研究/预报具有潜在的优势。

关键词: 再分析数据; 适用性; 误差时空分布

中图分类号: P73 **文献标识码:** A **文章编号:** 1003-0239(2021)04-0027-11

1 引言

人类社会的生产生活与各种天气/气候情况密不可分, 因此天气/气候学分析是保障社会发展的重要因素, 也面临巨大的挑战。高质量的气象数据是分析当前及过去天气现象和气候特征的重要基础。现有的气象观测资料受地域、时间和观测仪器的限制, 仅覆盖了部分海洋和极地等地区^[1], 其时间和空间的不连续性无法满足需求^[2]。

20世纪80年代后期, 随着计算机技术的发展, 基于求解动力方程的天气数值预报模式得到充分应用^[3]。研究人员利用数据同化技术将各种来源的地面、高空观测资料与数值模式的结果相结合, 得到格点化、连续长时间序列、高时空分辨率的历史资料数据集, 处理后的数据集称为再分析数据^[4]。再分析数据集综合了传统观测数据和数值模式数

据的优点, 弥补了单一观测资料时空分布不均的缺点, 增加了数值结果的准确性, 改进了对大气物理过程的描述, 能够充分反映气象要素的多种时空尺度变化特征^[5-6], 在各种天气/气候/海洋研究领域都发挥着重要作用, 如短时间内的强对流突发事件^[7]、短期天气预报、区域气候特征模拟^[8]、全球长时间气候变率、变化的监测和研究^[9]以及为海洋特征模拟提供驱动^[10]等。

美国、欧洲和日本等国家主要的业务/科研部门均开展了全球大气再分析资料的研究, 各类再分析数据集的时间跨度和分辨率逐步提高, 覆盖时间尺度从十几年到一百多年^[11]。例如: 美国国家环境预报中心(National Centers for Environmental Prediction, NCEP)、美国国家大气研究中心(National Center for Atmospheric Research, NCAR)和美国能源部(Department Of Energy, DOE)联合发布的NCEP/

收稿日期: 2020-04-21; 修回日期: 2020-09-07。

基金项目: 江苏省气象局青年基金项目(202001)。

作者简介: 吕润清(1994-), 女, 助理工程师, 硕士, 主要从事海洋气象预报研究。E-mail: runqinglyu@163.com

通讯作者: 李响(1986-), 男, 助理研究员, 博士, 主要从事海气相互作用研究。E-mail: lixiang@nmfnc.cn

NCAR (NCEP1)、NCEP/DOE (NCEP2)、NCEP/CFSR (Climate Forecast System Reanalysis, 气候预报系统再分析)等数据集^[12-14];美国国家航空航天局(National Aeronautics and Space Administration, NASA)提供的回顾性分析应用数据集(Modern-Era Retrospective analysis for Research and Applications, MERRA)^[15];欧洲中期天气预报中心(European Centre for Medium-range Weather Forecasts, ECMWF)推出的系列再分析数据:第一个全球大气研究计划(Global Atmospheric Research Program, GARP)全球实验(First GARP Global Experiment, FGGE)、15 a欧洲再分析数据(European Re-Analyses 15 years, ERA-15)、40 a欧洲再分析数据(European Re-Analyses 40 years, Era-40)和过渡期欧洲再分析数据(European Re-Analyses-Interim, ERA-Interim)^[16,17];日本气象厅(Japan Meteorological Agency, JMA)发布的25 a日本再分析数据(Japanese Re-Analysis 25 years, JRA-25)和55 a日本再分析数据(Japanese Re-Analysis 55 years, JRA-55)^[18-19]等。2014年,中国国家气象局(China Meteorological Administration, CMA)国家气象信息中心也开始研发中国全球再分析资料集(CMA's global atmospheric Re-Analysis, CRA)^[20]。

由于选用不同数值模式、观测资料以及同化和均一化过程带来的差异^[21],不同再分析数据集的可信度存在差异^[22-23]。2016年,ECMWF发布了使用新一代四维变分同化技术制作的第五代ECMWF再分析资料集(The fifth generation ECMWF Re-Analyses, ERA5),该数据集采用实时更新的方式提供1950年至今的高分辨率全球大气再分析数据^[24-25]。ERA-Interim与ERA5的适用性比较研究已在中国多个省份开展,孟宪贵等^[26]对比了山东省及周边的观测数据,结果表明ERA5在地面和对流层低层的相对湿度和风场上的表现明显优于ERA-Interim;朱景等^[27]利用浙江省71个气象自动观测台站的温度资料,证明ERA5对地表温度的模拟较ERA-Interim有明显改善。此外,同一数据集在不同区域的适用性也有较大差别,叶梦姝^[21]比较了全国降水资料与3套再分析资料,结果显示平原地区的再分析资料结果较山区更加可信,复杂地形区域对降水的空报和漏报情况都显著多于平原地区。

江苏位于 $116^{\circ}\sim 122^{\circ}\text{E}$, $30^{\circ}\sim 36^{\circ}\text{N}$ 之间,受东亚季风控制,兼具南北方气候特征,以淮河和苏北灌溉总渠一线为界,北部属于温带季风气候,南部属于亚热带季风气候。地理环境上,江苏为沿海省份,海岸线绵长,大陆地区整体地势平坦,地跨长江和淮河,域内湖泊众多。受季风和地理环境影响,江苏天气情况复杂多变,气象灾害相对频发,预报具有很大的挑战性^[28]。鉴于再分析数据集对天气/气候预测的重要性及其数据本身的不确定性,研判再分析数据在江苏区域的适用性及对相关的后续研究提供的理论依据,具有非常重要的意义。本文利用江苏省73个国家级自动气象观测站和3个高空探测站,对ERA-Interim和ERA5的地面和高空气象要素在江苏区域内的适用性进行分析对比,利用机器学习方法对再分析数据的误差进行分类,探究其主要的误差分布和偏离特征,最终给出这两种再分析数据在江苏区域的适用性情况,为使用该区域ERA5和ERA-Interim再分析数据集进行相关天气/气候研究及预报的气象科研人员提供借鉴和参考。

2 数据与方法

2.1 数据

观测数据为江苏省73个国家级自动气象观测站及3个高空探测站(南京、徐州和射阳)的观测数据,具体站点分布如图1所示。文中所用地面观测要素包括:海平面气压(单位:hPa)、10 m风向(单位:°)、



图1 江苏区域观测数据站点分布(红点所示为地面自动观测站,红五角星所示为探空站点)

10 m 风速(单位:m/s)、2 m 气温(单位:℃)、2 m 相对湿度(单位:%)和每 1 h 内降水累积量(单位:mm);高空观测要素包括:标准气压层(1 000 hPa、925 hPa、850 hPa、700 hPa、500 hPa、400 hPa、300 hPa、250 hPa、200 hPa、150 hPa 和 100 hPa)的位势高度(单位:位势米)、风向(单位:°)、风速(单位:m/s)、气温(单位:℃)和相对湿度(单位:%)。

由于江苏省在 2016 年进行了气象观测改革,对观测数据实施质量控制,剔除异常值,所以 2016 年后的数据质量结果更为可信。为了保证不同月份间对比的数据量保持一致,需要选取整年的观测数据。因此,选取观测数据时段为 2016—2018 年,其中地面要素为逐 10 min 质控后观测数据,高空要素为一日 2 次(世界时 00 时和 12 时,下同)的观测数据,总数据量达 11 449 758 个,具有一定的代表性。

再分析数据为 ERA-Interim 和 ERA5 再分析数据集。ERA-Interim 是 ECMWF 第 3 代再分析数据集,使用 ECMWF 集成预报系统(Integrated Forecast System, IFS) 31r2 循环(Cycle 31r2, Cy31r2)中的四维变分(Four-Dimensional Variational, 4D-Var)数据同化技术^[29]。该数据提供 0.125°~2.5° 多种空间分辨率,垂向上从地面起分为 60 层,分辨率最高可达 0.1 hPa,时间范围为 1979—2019 年,提供每日 4 次同化数据(世界时 00 时、06 时、12 时和 18 时)^[30]。ERA5 数据是 ECMWF 最新再分析数据集。作为 ERA-Interim 的替代品,ERA5 的时间和空间分辨率都得到提高,水平分辨率由 0.75° 提高至 0.281 25°,垂直分层由 60 层增至 137 层,垂向分辨率由 0.1 hPa 提高至 0.01 hPa(约为 80 km),数据的时间范围延展到从 1979 年开始到最近 5 d 内的实时数据(2020 年已发布将时间范围提前至 1950 年的初步版本),时间分辨率由每 6 h 提高至每 1 h^[24, 25]。此外,相比于 ERA-Interim 数据,ERA5 数据对观测资料的使用方法和物理过程描述都有显著的改进,数据同化技术使用的是 ECMWF 新一代 IFS 41r2 循环(Cycle 41r2, Cy41r2)的 4D-Var 技术^[31]。本文选取了 2016—2018 年同时间段的 ERA5 和 ERA-Interim 再分析数据。

2.2 方法

本文首先通过 Pearson 相关系数(Correlation

Coefficient, CC)衡量再分析数据和观测数据之间的一致性,并利用标准差(Standard Deviation, δ)和均方根误差(Root Mean Square Error, RMSE)量化偏离程度,具体计算公式如下:

$$CC = \frac{\sum_{i=1}^N (A_i - \bar{A})(E_i - \bar{E})}{\sqrt{\sum_{i=1}^N (A_i - \bar{A})^2 \sum_{i=1}^N (E_i - \bar{E})^2}} \quad (1)$$

$$\delta_A = \sqrt{\frac{1}{N} \sum_{i=1}^N (A_i - \bar{A})^2}, \quad \delta_E = \sqrt{\frac{1}{N} \sum_{i=1}^N (E_i - \bar{E})^2} \quad (2)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (E_i - A_i)^2} \quad (3)$$

式中, A 表示观测数据; E 为再分析数据; N 为样本数; $\bar{}$ 表示平均值。这 3 个统计量将通过泰勒图综合表示。标准化的泰勒图基于余弦定理,综合展示 CC、 δ 和 RMSE 的关系。以原点为圆心,方位角表示相关系数,距原点的距离为再分析数据和观测数据的 δ 比率;以观测数据点(Obs , 沿水平方向距离原点一个单位)为圆心的圆表示标准化 RMSE。因此距离观测数据点越近,说明再分析数据模拟效果越好^[32]。泰勒图中标准差比率为:

$$\delta = \frac{\delta_E}{\delta_A} \quad (4)$$

标准化均方根误差为:

$$RMSE' = \sqrt{\frac{1}{N} \sum_{i=1}^N [(E_i - \bar{E}) - (A_i - \bar{A})]^2} \quad (5)$$

同时,文中利用相对偏差(Relative bias)衡量误差偏离正负和分析误差来源,计算公式为:

$$Relative\ bias = \frac{\frac{1}{N} \sum_{i=1}^N (E_i - A_i)}{\frac{1}{N} \sum_{i=1}^N A_i} \times 100\% \quad (6)$$

为了充分利用有限的的数据,减少统计量计算误差,本文利用自助采样法(拔靴法, Bootstrap)^[33]计算上述统计量 95% 置信度的范围。在常规计算方法下,对于一组再分析数据和观测数据,统计量式(1)~(6)只能计算出对应的一个值,难以估计这个值的质量或者置信度。Bootstrap 方法作为一种增广样本统计方法,以原始数据为基础,通过重复抽样(抽样次数由用户指定),可以得到一组数据某统计量的分布特征(通过重复抽样得到某个统计量的多个

值)^[34],适用于本文统计量难以用常规方法进行假设检验的问题。在本文实际计算某统计量过程中,对于一组再分析数据和观测数据,规定通过 Bootstrap 方法得到某统计量 1 000 个样本,再以 1 000 个样本为基础,取具有 95% 置信度的范围进行检验,下文讨论的有关统计量都在 95% 置信区间内。

此外,本文利用无监督学习 *K* 均值 (*K-means*, *K* 为指定分类数目)聚类分析方法对各气象要素的误差进行分类。*K-means* 算法是一种基于空间欧式距离(空间中两个样本对象之间的真实距离)进行相似度划分的聚类算法,欧式距离越小的两个样本对象之间的相似程度越高^[35]。因其算法简捷高效,且可以指定分类数目,归类结果解释性强,此方法已被用于挖掘大量气象数据中深层次的关系。韩微等^[36]基于全国 50 a 的逐日气温观测资料,利用聚类分析方法将全国划分为若干区域,划分结果与按气候和地形划分结果相近。本文尝试利用再分析数据误差将观测站点划分为若干区域,各区域之内的误差最小化,同时区域之间的误差最大化,分类区域能具有一定地域特征。分类结果有助于描述

系统误差的空间分布,在使用再分析数据的过程中,注意系统误差对天气分析的影响,能够提高再分析数据利用率。

3 结果

3.1 地面要素的适用性及误差时空分布特征

3.1.1 地面要素的适用性

图 2 给出了 2016—2018 年 ERA5 和 ERA-Interim 数据中各地面气象要素的泰勒图。两套再分析数据对江苏区域的各地面观测要素均具有较好的表征能力,除小时降水量这一要素外,ERA5 的其他地面要素数据比 ERA-Interim 数据的适用性显著提升,尤其是 2 m 气温及 2 m 相对湿度。

相比 ERA-Interim 数据,ERA5 中海平面气压与观测数据的相关系数均高于 0.95 (ERA-Interim 在 0.9 附近),且均方根误差小于 0.25 hPa (ERA-Interim 在 0.25~0.5 hPa 之间)。ERA5 的 2 m 气温及 2 m 相对湿度相比于 ERA-Interim 表现出系统性改进,与观测数据的相关系数分别集中在 0.95 和 0.9 附近,

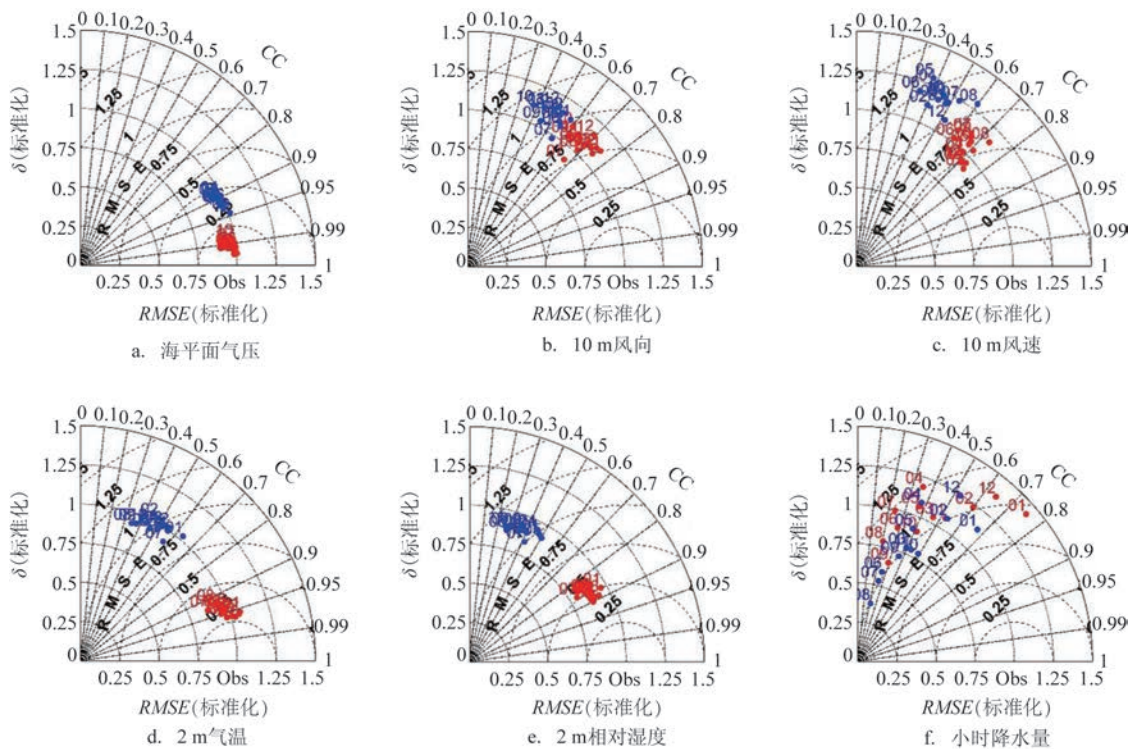


图 2 地面要素泰勒图(红色和蓝色散点表示 1—12 月 ERA5 和 ERA-Interim 再分析数据的相关系数、标准差和均方根误差的分布,散点旁数字表征月份,散点越靠近 Obs 表示再分析数据和观测数据间各项误差越小)

两个要素的相关系数相较于ERA-Interim与观测数据的相关系数0.5和0.4均提升了超过0.4,标准化均方根误差均减少了60%。ERA5数据的10 m风速风向相较于ERA-Interim数据也有所改进,与观测的一致性提高(相关系数提高约0.3),均值也更接近于观测。但是,上述地面要素在ERA5和ERA-Interim两套数据中的标准差比率没有显著差异,标准差比率均集中在1附近,ERA5和ERA-Interim均较准确地表征海平面气压、2 m气温和2 m相对湿度的振幅,ERA-Interim高估了10 m风速约15%,10 m风向的振幅也有约15%的偏离。两套数据的小时降水量相较于观测均有较大误差,相关系数分布较为离散,体现了该要素时间上适用性的差异,在实际研究/预报中需谨慎使用该要素。ERA5数据的地面要素适用性相较于ERA-Interim数据显著提升(除小

时降水量),可能主要得益于制作该数据的模式、同化系统和观测数据的改进。

需要注意的是,在统计学上,本文计算得到的相关系数只需高于0.104(根据数据量查表可得),即可认为在显著性水平0.001上显著相关^[37]。方差F检验^[37]表明,在显著性水平0.05上,两套再分析数据的海平面气压与观测数据相比离散程度没有显著差异,其他各要素离散程度都有显著差异。

3.1.2 相关系数和相对偏差的时空分布特征

两套再分析数据与观测数据的一致性及其误差的时空分布特征,对于再分析数据的合理使用非常重要。本文选择针对不同站点逐月的相关系数和相对偏差进行时空分布特征的分析。图3给出了两套再分析数据与站点观测数据之间相关系数的逐月变化。对于各地面气象要素,ERA5数据与观测

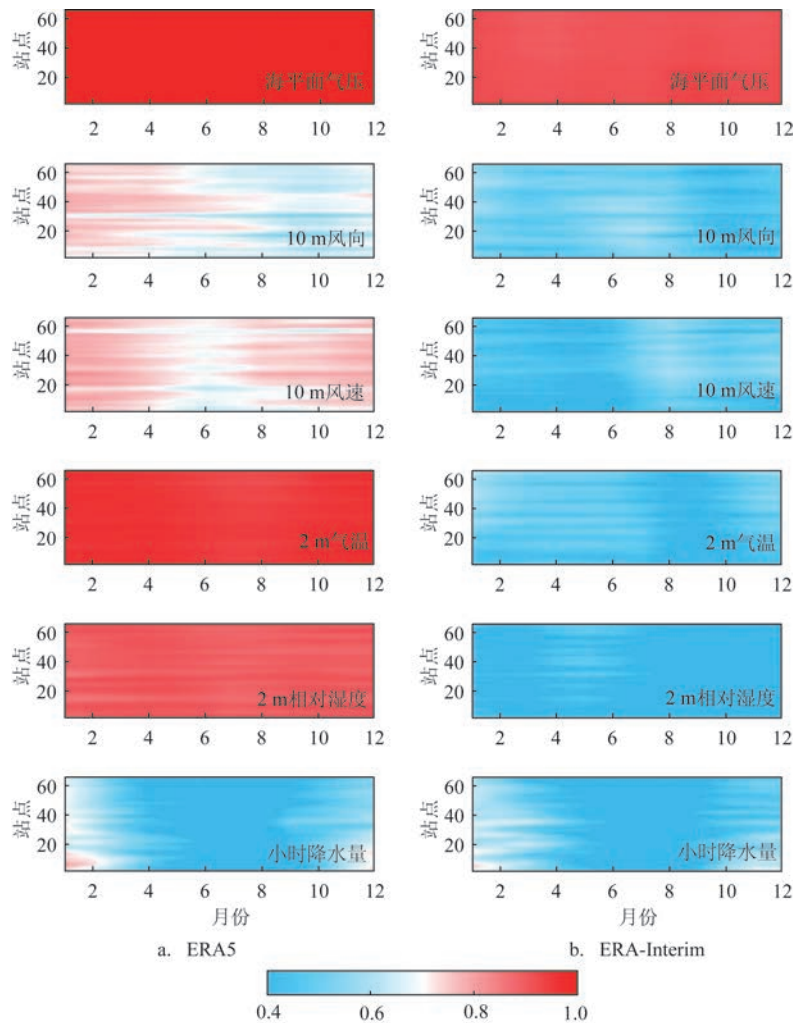


图3 地面要素站点相关系数逐月分布(各物理量单位见2.1部分)

数据的一致性要优于ERA-Interim数据,尤其是2 m气温及2 m相对湿度。ERA5再分析数据的各地面气象要素的相关系数季节差异显著,相关系数的低值区主要存在于夏季(6—8月)。小时降水量的相关系数差异较大,最好月份和最差月份相差达0.6,其他要素季节相差在0.2之内。此外,2 m相对湿度在1月和12月的相关系数与年均值相比偏低0.05,10 m风向在整个下半年(6—12月)相关系数明显偏低,比上半年偏低0.2。相较ERA5,ERA-Interim再分析数据与观测的相关系数的季节差异没有明显分布规律。海平面气压与观测数据的相关系数在9月比年均相关系数略偏高0.01。10 m风向、10 m风速、2 m气温和2 m相对湿度与观测数据的相关系数季节差异在0.2之内,10 m风向在3月、10月、11月

和12月的一致性较差,10 m风速在1—6月一致性偏低,2 m气温和2 m相对湿度在1—6月的相关系数较其他月偏高0.15,小时降水量的主要误差季节分布和ERA5相似。

图4给出了两套再分析数据与站点观测数据之间相对偏差的逐月变化。两套数据的相对偏差量级相近,且存在季节差异特征,不同要素的偏差的季节特征不同。但均值 t 检验^[37]显示,在0.05显著性水平上,ERA5再分析数据的海平面气压在冬季与观测相比没有显著差异,2 m气温部分站点在冬季与观测没有显著差异,其他气象要素均值与观测相比有显著差异;ERA-Interim的10 m风速和小时降水量与观测均值有显著差异,其他气象要素与观测相比只在部分站点有显著差异。两套数据相对偏

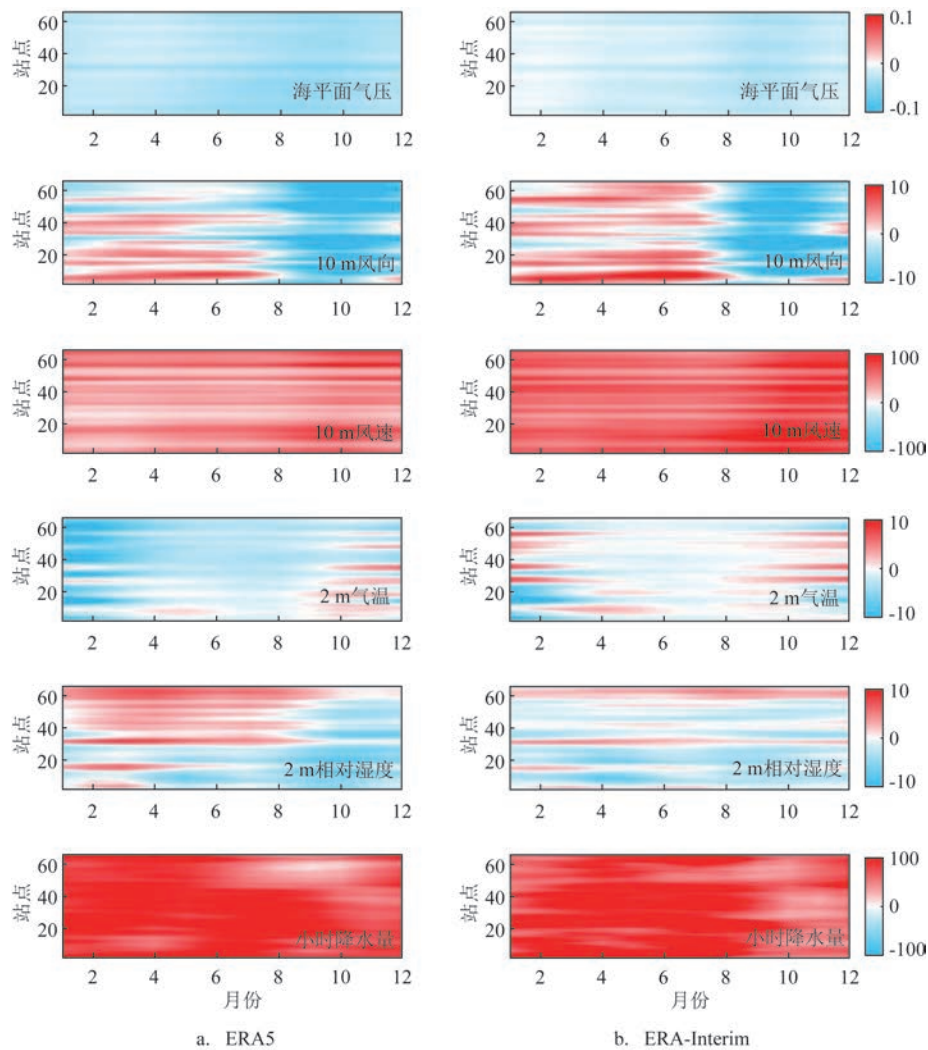


图4 地面要素站点相对偏差逐月分布(单位:%)

差的95%置信区间也显示,ERA5各气象要素的置信区间不包含0值,而ERA-Interim各气象要素相对偏差的置信区间在一半以上的站点跨0值分布。因此,ERA5再分析数据的相对偏差时空分布特征是可信的,在后续研究中有使用意义,而ERA-Interim相对偏差的时空分布特征在统计学上并不完全可信,需要注意。ERA5再分析数据与观测数据的相对偏差大致可以分为3个量级:海平面气压相对偏差在0.1%之内,低估了观测海平面气压值,且6—10月偏离程度略大;10 m风向、2 m气温和2 m相对湿度的相对偏差绝对值在10%之内,分别在1—7月、10—12月和1—7月表现为正偏差,其余月份相对偏差为负值;ERA5显著高估了10 m风速和小时降水量,10 m风速和小时降水量分别在1—7月和10—12月相对偏差较低。ERA-Interim再分析数据的相对偏差的季节分布特征与ERA5再分析数据类似,不同气象要素的偏离程度也相近,但2 m气温在1—4月为正偏差,2 m相对湿度在1—7月为负偏差。

导致这两套再分析数据适用性的季节差异的主要原因可能是江苏地区为显著的季风区域。受季风影响,冬季在强劲的西北气流控制下天气形势稳定,而夏季受东南气流影响,天气多变,强对流天气过程频发,且易受到台风等极端天气影响,多出现短时强降水^[7,28],导致再分析数据在夏季的适用性偏低。以季节差异明显的小时降水量为例,ERA5和观测的小时降水量降水分布直方图(图略)表明,ERA5数据的降水频次和一般降水量(小时降水量<20 mm)均高于观测,对一般降水存在空报现象,导致小时降水量被高估;同时,ERA5数据明显低估了短时强降水(小时降水量>20 mm)发生的概率,常漏报短时强降水,而短时强降水常出现在对流性天气多发的夏季,导致夏季的数据适用性较差。

图3和图4表明,两套再分析数据的适用性存在季节差异,同时也存在空间差异,不同观测站点间差异在夏半年(4—9月)尤为明显。为了更加直观地考察两套数据适用性的空间差异,采用无监督学习*K-means*聚类分析方法^[36]对73个观测站4—9月的误差进行分类,分类依据为各个站点地面要素的相关系数和相对偏差。

图5给出了ERA5再分析数据中各地面要素适

用性的空间差异。图中可以看出,ERA5再分析数据适用性的空间分布具有显著的地域特征,总体以长江为界将江苏省分为南北两个区域,两个区域间站点的相关系数差别不大,但是相对偏差有明显区别。其主要表现为10 m风向在南部地区(红点)为负偏差而在北部地区(蓝点)为正偏差,2 m相对湿度在南部地区为正偏差而在北部地区为负偏差。具体分类结果显示,南部区域海平面气压的相关系数相对较高,海平面气压、10 m风向和2 m气温的相对偏差为负,分别集中在-0.025 5%、-5.26%和-3.34%左右,10 m风速、2 m相对湿度和小时降水量的相对偏差为正,分别集中在59.14%、2.49%和56.63%附近。而在北部区域,2 m气温和2 m相对湿度的相关系数较高,海平面气压、2 m气温和2 m相对湿度的相对偏差为负,分别为-0.036%、-0.058 2%和-3.30%左右,10 m风向、10 m风速和小时降水量的相对偏差为正,分别集中在为2.22%、22.15%和90.83%附近。同样,对ERA-Interim再分析数据进行了聚类分析,结果显示为没有显著的空间分布特征(图略)。

3.2 高空要素的适用性

图6给出了两套再分析数据与江苏境内3个高空探测站2016—2018年气象探空数据的相关系数和相对偏差的对比。两套数据中的高空气象要素与观测较为一致,其中位势高度与气温比较准确,各层的风速、风向与观测相对一致,但是各层的相对湿度随高度的变化体现了较大的差异性。两套数据的适用性表现出相同的空间差异,其中,气温和风向的适用性在对流层低层和高层表现出明显空间差异,相对湿度在对流层中高层的适用性有明显空间差异。

ERA5数据各层位势高度与探空数据的相关系数都在0.95以上,相关系数从1 000 hPa开始随高度逐渐减小,在850~925 hPa之间达到局部极小值,然后逐渐回升,从150 hPa高度再度开始急剧减小,相关系数在对流层高层降至最小;相对偏差在1 000~700 hPa之间为负值,其随高度增加缓慢减小,在对流层中高层趋近于0。ERA5数据的风向和风速与探空数据的相关系数在0.7之上,基本随高度的增加而增大;相对偏差在1 000~700 hPa之间随高度增加而逐渐趋近于0,在对流层中高层基本保持不

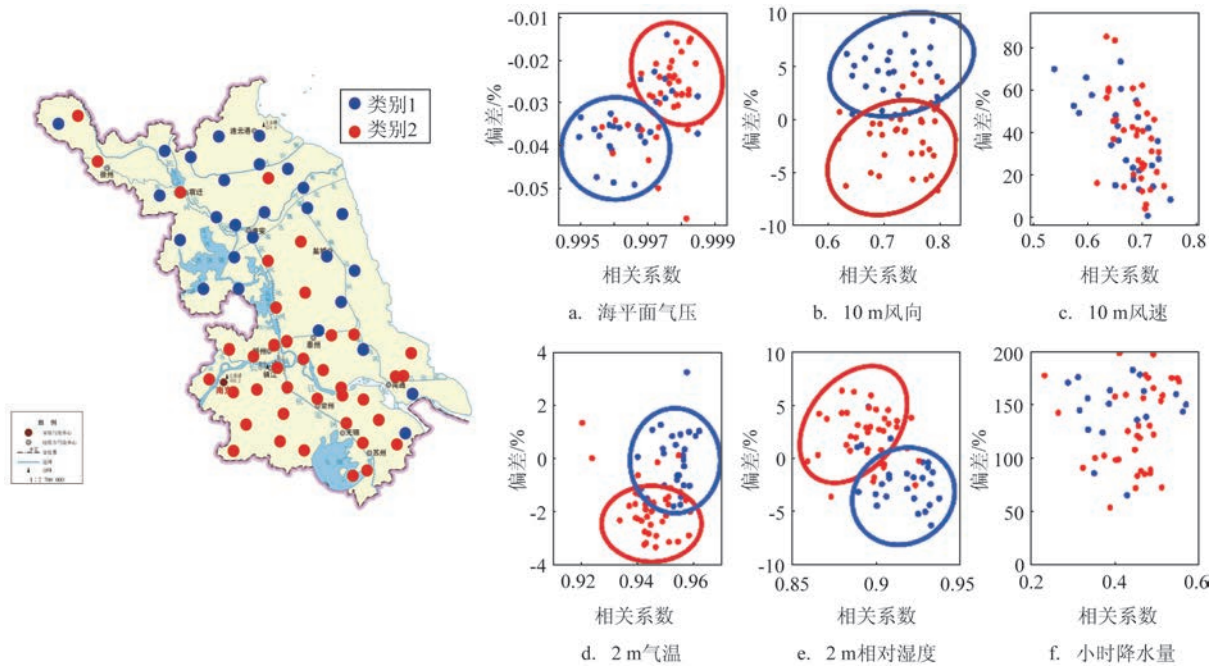


图5 ERA5地面要素适用性误差分类

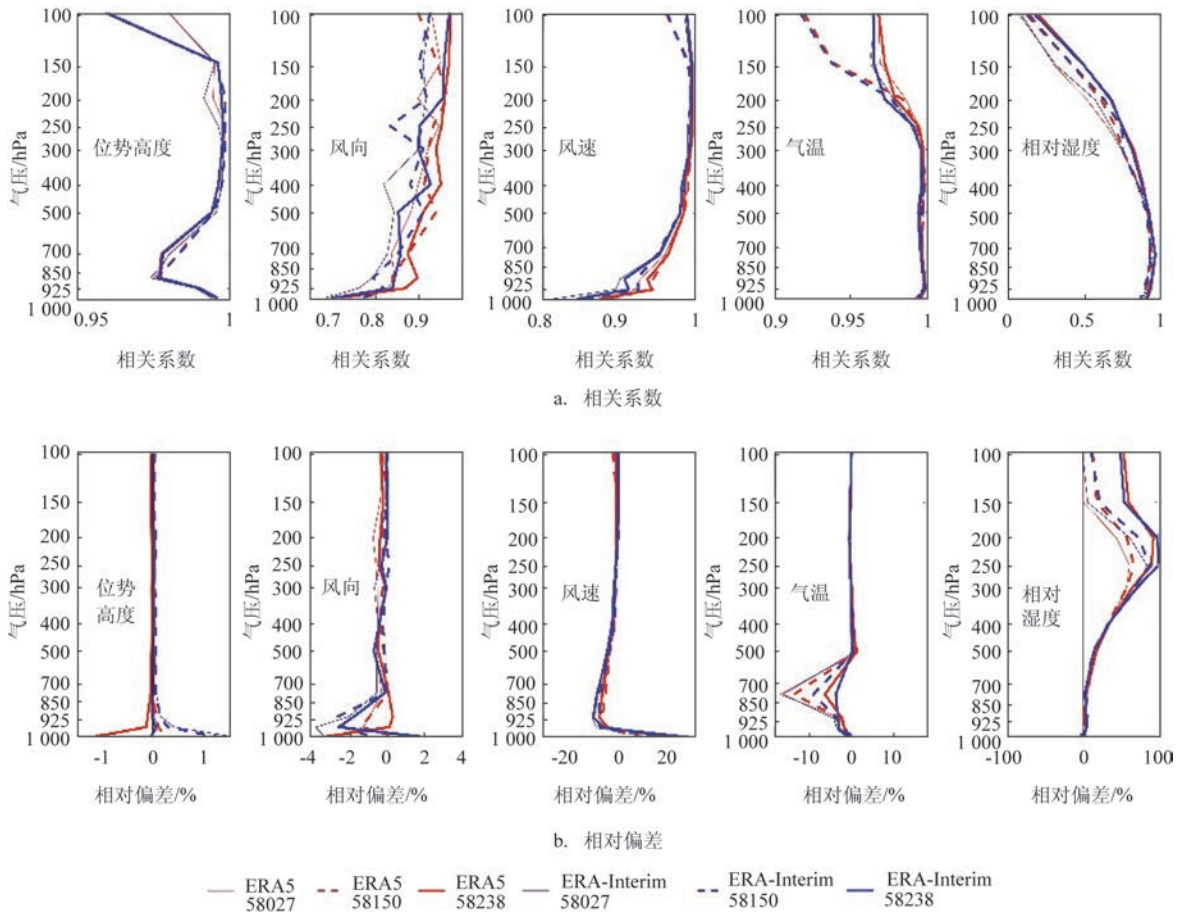


图6 高空要素相关系数和相对偏差对比图

变。ERA5的气温和相对湿度与探空数据的相关系数随高度增加而减小,其中相对湿度的相关系数变化显著,从0.9减小到0.2;气温的相对偏差在对流层中低层700 hPa达到极小值-17%,中高层接近于0,相对湿度则在对流层中高层有明显偏差,最大相对偏差可达100%。不同站点高空气象要素的适用性随高度变化的趋势基本一致,相比于站点58238,站点58150和站点58027的风向、风速、气温和位势高度在对流层低层和高空的适用性较差,相对湿度在对流层中高层与观测数据有更大差异。高空要素相比于地面要素没有明显季节变化特征(图略)。

两套再分析数据高空气象要素的相关系数和相对偏差随高度变化的趋势基本一致。相较于ERA-Interim数据,ERA5数据与探空数据的相关系数在风场和气温上的表现有所提升。风向在不同高度上均有改善,相关系数平均提高了0.3左右;风速的相关系数在对流层低层提高了约0.4;气温的相关系数在对流层高层提升约0.1;ERA5的位势高度和相对湿度与ERA-Interim相比没有明显优势。两套数据在不同气压层的相对偏差没有显著差异。

4 总结

本文利用江苏73个国家级自动气象观测站和3个高空探测站的观测数据对ERA-Interim和ERA5再分析数据的地面及高空气象要素在江苏省的适用性进行对比研究,主要结论如下:

(1)ERA5再分析数据的各地面要素的适用性均优于ERA-Interim再分析数据,其中2 m气温及2 m相对湿度体现了系统性的改进,相关系数提升超过0.4,但两套再分析数据的小时降水量均存在较大的误差。ERA5再分析数据与观测数据之间的一致性 & 误差均存在一定的时空分布特征。总体而言,ERA5中海平面气压的适用性在全年都有很好的表现,数据质量可靠,其他气象要素在夏季的适用性低于其他季节,空间上表征为在夏半年存在显著南北差异,10 m风向在江苏南部为负偏差而在北部为正偏差,2 m相对湿度在南部为正偏差而在北部为负偏差。与之相对的,ERA-Interim数据没有显著的时空分布特征。

(2)两套再分析数据的高空气象要素与观测数据均较为一致,其中位势高度与气温比较准确,各

层的风速、风向与观测数据相对一致,但是各层的相对湿度随高度的变化体现出较大的差异性,各个要素不同站点间适用性差异在对流层低层和高层比较明显。总体而言,ERA5再分析数据在各气压层与观测的一致性优于ERA-Interim再分析数据。

本文虽然取得了一定的研究成果,但是还存在一些不足。本文的分析工作,尤其是季节分析和聚类分析,是基于3 a(2016—2018年)的观测数据,其结果代表性有一定局限,我们将在以后的研究中进行更长时间序列的验证分析。同时,本文采用聚类分析方法对误差的空间分布特征进行了分析,但是对这种误差空间分布特征的成因未明确,在后续的研究中也将开展更多的机制分析工作,给出误差时空分布特征的成因。

参考文献:

- [1] 卢博,曹丛华,郭敬天,等.业务化海洋观测数据传输网络的设计与实现[J].海洋预报,2017,34(1):72-76.
- [2] 何金海,郭品文,银燕,等.大气科学概论[M].北京:气象出版社,2012:107-233.
- [3] Pielke R A. Chapter 1-Introduction[M]//Pielke R. A. Mesoscale Meteorological Modeling. San Diego: Academic Press, 1984: 1-2.
- [4] Bengtsson L, Shukla J. Integration of space and in situ observations to study global climate change[J]. Bulletin of the American Meteorological Society, 1988, 69(10): 1130-1143.
- [5] 程麟生.中尺度大气数值模式发展现状和应用前景[J].高原气象,1999,18(3):350-360.
- [6] 谢潇,何金海,祁莉.4种再分析资料在中国区域的适用性研究进展[J].气象与环境学报,2011,27(5):58-65.
- [7] 程胡华,闻斌,王益柏,等.NCEP/NCAR FNL资料在强对流天气中可信度的初步分析[J].气象与环境科学,2018,41(3):1-10.
- [8] 韦芬芬.不同再分析资料驱动场对我国夏季区域气候模拟的不确定性研究[D].南京:南京大学,2013.
- [9] 赵天保,符淙斌,柯宗建,等.全球大气再分析资料的研究现状与进展[J].地球科学进展,2010,25(3):242-254.
- [10] 周凯,陈希,李妍,等.两种海面风场的对比及对海浪模拟的影响[J].海洋预报,2012,29(3):9-14.
- [11] Compo G P, Whitaker J S, Sardeshmukh P D, et al. The twentieth century reanalysis project[J]. Quarterly Journal of the Royal Meteorological Society, 2011, 137(654): 1-28.
- [12] Kanamitsu M, Ebisuzaki W, Woollen J, et al. NCEP - DOE AMIP-II reanalysis (R-2) [J]. Bulletin of the American Meteorological Society, 2002, 83(11): 1631-1644.
- [13] Kalnay E, Kanamitsu M, Kistler R, et al. The NCEP/NCAR 40-year reanalysis project [J]. Bulletin of the American Meteorological Society, 1996, 77(3): 437-472.

- [14] Kistler R, Collins W, Saha S, et al. The NCEP - NCAR 50 - year reanalysis: monthly means CD - ROM and documentation[J]. Bulletin of the American Meteorological Society, 2001, 82(2): 247-268.
- [15] Rienecker M M, Suarez M J, Gelaro R, et al. MERRA: NASA's modern-era retrospective analysis for research and applications[J]. Journal of Climate, 2011, 24(14): 3624-3648.
- [16] Uppala S M, KÅllberg P W, Simmons A J, et al. The ERA-40 reanalysis[J]. Quarterly Journal of the Royal Meteorological Society, 2005, 131(612): 2961-3012.
- [17] Dee D P, Uppala S M, Simmons A J, et al. The ERA - interim reanalysis: configuration and performance of the data assimilation system[J]. Quarterly Journal of the Royal Meteorological Society, 2011, 137(656): 553-597.
- [18] Onogi K, Koide H, Sakamoto M, et al. JRA-25: Japanese 25-year reanalysis project—progress and status[J]. Quarterly Journal of the Royal Meteorological Society, 2005, 131(613): 3259-3268.
- [19] Ebita A, Kobayashi S, Ota Y, et al. The Japanese 55-year reanalysis “JRA-55”: an interim report[J]. SOLA, 2011, 7: 149-152.
- [20] 刘志权. 全球大气再分析技术研究 with 数据集研制实施方案(详细)[Z]. 国家气象信息中心, 2015.
- [21] 叶梦姝. 中国大气再分析资料降水产品在天气和气候中的适用性研究[D]. 兰州: 兰州大学, 2018.
- [22] 旷芳芳, 张友权, 张俊鹏, 等. 3种海面风场资料在台湾海峡的比较和评估[J]. 海洋学报, 2015, 37(5): 44-53.
- [23] 柳婧, 宋晓姜, 王彰贵. 中国近海 ASCAT 和 ERA-Interim 风场资料的评估[J]. 海洋预报, 2019, 36(1): 10-19.
- [24] Hersbach H, Dee D. ERA5 reanalysis is in production[R]. ECMWF Newsletter, 2016, 147(7).
- [25] Hersbach H, Bell W, Berrisford P, et al. Global reanalysis: goodbye ERA-Interim, hello ERA5[R]. ECMWF Newsletter, 2019: 17-24.
- [26] 孟宪贵, 郭俊建, 韩永清. ERA5再分析数据适用性初步评估[J]. 海洋气象学报, 2018, 38(1): 91-99.
- [27] 朱景, 袁慧珍. ERA再分析陆面温度资料在浙江省的适用性[J]. 气象科技, 2019, 47(2): 289-298.
- [28] 江苏省气象局. 江苏省天气预报技术手册[M]. 北京: 气象出版社, 2017: 1-21.
- [29] ECMWF. Part II: Data Assimilation[M]. Shinfield Park, Reading: ECMWF, 2008.
- [30] Berrisford P, Dee D P, Poli P, et al. The ERA-Interim archive Version 2.0[R]. Shinfield Park, Reading: ECMWF, 2011.
- [31] ECMWF. Part II: Data Assimilation[M]. Shinfield Park, Reading: ECMWF, 2016.
- [32] Taylor K E. Summarizing multiple aspects of model performance in a single diagram[J]. Journal of Geophysical Research: Atmospheres, 2001, 106(D7): 7183-7192.
- [33] Efron B. The jackknife, the bootstrap and other resampling plans [M]. Phil Adelpia: Society for Industrial and Applied Mathematics, 1982.
- [34] 谢益辉, 朱钰. Bootstrap方法的历史发展和前沿研究[J]. 统计与信息论坛, 2008, 23(2): 90-96.
- [35] Krishna K, Murty M N. Genetic K-means algorithm[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 1999, 29(3): 433-439.
- [36] 韩微, 翟盘茂. 三种聚类分析方法在中国温度区划分中的应用研究[J]. 气候与环境研究, 2015, 20(1): 111-118.
- [37] 魏凤英. 现代气候统计诊断与预测技术[M]. 2版. 北京: 气象出版社, 2007.

Comparison between the applicability of ERA-Interim and ERA5 reanalysis in Jiangsu Province

LYU Run-qing¹, LI Xiang²

(1. Jiangsu Meteorological Observatory, Nanjing 210008 China; 2. National Marine Environmental Forecasting Center, Beijing 100081 China)

Abstract: The applicability of ERA-Interim and ERA5 reanalysis in Jiangsu province is compared based on the observation data of 73 ground meteorological observation stations and 3 meteorological sounding stations. The results show that the applicability of ERA5 is higher than that of ERA-Interim for both ground and upper-air meteorological variables. In particular, the 2 m air temperature and relative humidity in ERA5 reanalysis shows a systematic improvement compared to ERA-Interim with the correlation coefficient increased by over 0.4. The

hourly precipitation of both reanalysis datasets reveal large errors compared to observations, and should therefore be used with caution. The ERA5 reanalysis and observation shows significant spatiotemporal variations in terms of consistency and error. The applicability of the ERA5 reanalysis in summer is lower than the rest of the year with a south-north spatial distribution characteristics, while the ERA-Interim reanalysis reveals no significant temporal and spatial distribution characteristics. Our results suggest that ERA5 is potentially a better choice for weather and climate research and prediction in the Jiangsu province.

Key words: reanalysis; applicability; spatiotemporal distribution of errors