

孙虹雨,王迪,刘成瀚,等.ERA5 再分析资料在辽宁省的适用性分析与订正[J].沙漠与绿洲气象,2023,17(4):101-110.

doi: 10.12057/j.issn.1002-0799.2023.04.014

开放科学(资源服务)标识码(OSID):



ERA5 再分析资料在辽宁省的适用性分析与订正

孙虹雨^{1,2},王 迪³,刘成瀚²,张 燕⁴,王式功^{1,5*}

(1.成都信息工程大学大气科学学院/高原大气与环境四川省重点实验室,四川 成都 610225;2.辽宁省气象台,辽宁 沈阳 110166;3.河南省气象台,河南 郑州 450003;4.凉山州气象局,四川 凉山彝族自治州 615000;
5.成都平原城市气象与环境四川省野外科学观测研究站,四川 成都 610225)

摘 要:基于 2009—2018 年辽宁省 62 个地面站和 9 个探空站(省内 4 部,省外 5 部)资料,提取了温度、湿度、气压、风场等常用物理量,采用统计分析方法对 ERA5 资料在辽宁省的适用性进行了分析,并尝试使用机器学习的方法订正 ERA5 资料的偏差。结果表明:在地面观测要素分析中,2 m 温度相关系数普遍较高,均方根误差普遍偏低,辽河流域及其西侧地区效果好于其他地区;10 m 风速对比结果总体稍差, U 分量相关系数总体低于 V 分量;ERA5 资料质量有明显的月变化特征,相关系数总体呈现春秋季节高于夏冬季,且相关系数总体较高的季节,均方根误差也普遍偏高。在高空观测要素分析中,辽宁中部地区的数据质量高于东西部,温度的平均相关系数最高,相对湿度相关系数平均较低,均方根误差总体较大;相对湿度的相关系数在春夏之交时最低,夏季上升较快,夏秋之交时达到最高,相关系数快速上升的月,均方根误差也呈现出快速上升的趋势;高空 U 风场和 V 风场的均方根误差总体相差不大,且均方根误差月变化一致,总体经向风(V 风场)质量低于纬向风(U 风场)。通过机器学习的订正方法有效提升了地面温度和相对湿度的应用能力,使 ERA5 地面温度资料的均方根误差缩小 0.5~1.0 °C,地面相对湿度资料的均方根误差缩小最高可达 26%。

关键词:再分析资料;适用性;均方根误差;相关系数;机器学习

中图分类号:P413

文献标识码:A

文章编号:1002-0799(2023)04-0101-10

由于大气实际探测方式较多,观测时间不统一,空间分辨率不均匀,不利于历史资料统计分析。20 世纪 90 年代之后,美国、欧洲和日本等推出了自己的“再分析资料”,较好地解决了常规观测资料历史反演的困难。随着当前探测方式和探测资料的不断

增多,资料同化技术的迅速提升,计算机性能的高速发展,给“再分析资料”注入了新的活力,使其在时间分辨率和空间分辨率上有了质的提升。欧洲中期天气预报中心(ECMWF)是较早开展再分析资料研发的机构,最新的 4D-Var 方法也是最早投入业务运用。时至今日,ECMWF 推出了五代再分析资料产品,分别为:第一代再分析产品(FGGE)^[1]、第二代再分析产品(ERA-15)^[2]、第三代再分析产品(ERA-40)^[3]、第四代再分析产品(ERA-Interim)^[4]和第五代再分析产品(ERA-5)^[5]。第五代再分析产品从数据量、时空分辨率以及同化方法上都有了较大的提高。再分析资料质量如何,能否反映历史真实的大

收稿日期:2022-11-19;修回日期:2023-03-21

基金项目:海南省南海气象防灾减灾重点实验室开放基金项目(SCSF202007);辽宁省气象局科学技术研究项目(ZD202224)

作者简介:孙虹雨(1985—),女,高级工程师,主要从事短期天气预报研究。E-mail:81196822@qq.com

通信作者:王式功(1955—),男,教授,主要从事环境气象和天气预报技术方面的研究。E-mail:wangsg@cuit.edu.cn

气状况,近年来备受关注,国内外很多专家和学者针对不同的再分析资料做过很多质量评估工作。孟宪贵等^[6]通过对 ERA5 资料的实用性进行了分析,指出 ERA5 资料的海平面气压和 2 m 温度与实况相关性好于 2 m 相对湿度和 10 m 风场。朱景等^[7]利用浙江省 71 个台站观测的日平均气温和地表(0 cm)温度资料对两套再分析资料(ERA5 和 ERA-Interim)进行评估,指出两套资料总体实用性均较高。秦艳慧等^[8]研究表明再分析资料能较好地再现高原多年冻土区的地表温度基本特征,并能较好地描述地表温度的季节变化。滕华超等^[9]和白磊等^[10]通过分析 ERA-Interim 风速再分析资料的适用性,指出 ERA-Interim 风速再分析资料整体具有较高的可信度,且 ERA-Interim 再分析资料的气压、温度的可信度要优于 NCEP/NCAR 数据,但局部有差异。支星等^[11-12]对再分析资料年平均高空温度在中国地区的可信度进行系统分析,指出三种再分析资料(NCEP/NCAR、ERA 和 JRA)高空温度较探空资料高空温度偏低,ERA 和 JRA 在对流层中下层与探空资料更加接近。陈艳春等^[13]分析了五种再分析资料在本地的适用性,指出在环渤海区域 10 m 风场中,JRA 资料与观测站的相关系数最高,ERA 资料的均方根误差最小。袁松等^[14]指出 NCEP 资料中温度、相对湿度、 U 风场和 V 风场绝对差值随地点、高度、季节变化较小。高淑新等^[15]和谢潇等^[16]指出,不同再分析资料存在一定的共性,但不同地区不同时间尺度下差异显著,辽宁西部地区再分析资料可用性差异较大。郑艳萍^[17]通过研究 ERA5 再分析资料在广东省的适用性指出,ERA5 再分析地面资料的海平面气压、温度表现较好,相对湿度表现稍差。王秀明等^[18]选取 60 个超级单体风暴个例,研究再分析资料在我国强对流天气环境分析中的适用性,指出 NCEP 再分析资料计算的对流有效位能、水汽参数与观测值差异较大,用 K 指数、温度直减率来分析大气层结稳定度效果较好。夏凡等^[19]通过对 WRF 预报的 2 m 温度进行验证,指出 08 时的预报效果好于 20 时,且白天预报效果好于晚上。许敏等^[20]和李翔翔等^[21]指出,GRAPES 和 T639 模式对大雨及以上降水和小雪预报效果较好,T639 模式 TS 评分高于 GER、JPN 等模式;通过订正 JAR55 再分析资料,其地面温度均方根误差得到有效缩小。陈艳春等^[22]运用环渤海气象台站资料与 NCEP/NCAR 等 5 种再分析资料研究其适用性问题,指出 ERA-Interim 资料水平分辨率较高,强风时可重点参考。邓小花等^[23]也指出各种再分析资料所

使用的同化方案、数据来源、质量控制算法及相关的偏差校正方法不同,应从经验出发分析各类再分析资料在不同方面的优缺点。

由于大气的状态和变化具有很强的地域性,已有的研究成果可为科研和业务应用提供思路和参考依据,部分成果和参数在辽宁省不一定适用。本文将重点分析 ERA5 资料在辽宁地区的适用性,并根据 ERA5 资料的误差结果,运用机器学习的方法进行订正,缩小 ERA5 资料与实况资料的差距,为 ERA5 资料在辽宁适用性的研究和应用提供借鉴和参考。

1 资料和方法

1.1 资料来源

1.1.1 ERA5 再分析资料

为便于计算和提高运算效率,结合辽宁省边界,资料范围为 $38^{\circ}\sim 44^{\circ}\text{N}$, $118^{\circ}\sim 126^{\circ}\text{E}$ 。ERA5 再分析资料选取业务中常用资料包括地面温度、露点、风向和风速资料,高空选取不同高度的气压、位势高度、温度、露点、风向和风速资料。为尽可能保证评估的普适性和代表性,选用 2009 年 1 月 1 日—2018 年 12 月 31 日(北京时,下同)共 10 a 资料;地面和高空时间分辨率为 1 h,空间分辨率为 $0.25^{\circ}\times 0.25^{\circ}$,高空选用 11 个标准层(1 000、925、850、700、500、400、300、250、200、150、100 hPa)。

1.1.2 实况观测资料

辽宁省共有国家地面站 62 个,探空站 4 个,站点分布如图 1 所示。文中涉及的地图,均来自于“标准地图服务网”下载的中国标准地图,审图号为 GS(2020)4619 号。实况资料时间为 2009 年 1 月 1 日—2018 年 12 月 31 日共 10 a 资料,地面观测资料选取 62 个国家站的温度、露点、风向和风速资料,探空资料选取气压、高度、温度、露点、风向和风速资

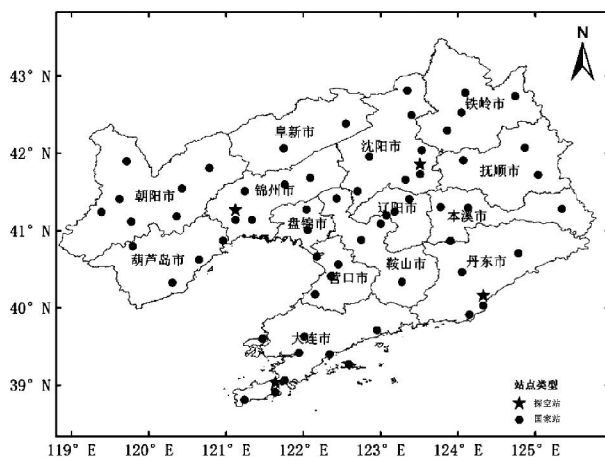


图1 辽宁省国家气象观测站和探空站分布

料。由于 ERA5 再分析资料的时间分辨率为 1 h, 所以地面资料也选用逐小时资料进行匹配; 高空选用每日 2 次(08、20 时)的探空资料。所有观测资料来自辽宁省气象局, 本文使用的站点均为考核站点, 数据质量可信度较高。

1.2 评估方法

为分析和检验实况观测资料和 ERA5 再分析资料的差异, 很多专家运用不同的方法将再分析资料匹配到实况观测站点上, 其中有双线性插值法和邻近点匹配法等。朱景等^[7]通过研究指出, 双线性插值会带来插值误差, 而且邻近点匹配法误差相对较小; 所以本文将采用邻近点匹配法, 将 ERA5 再分析资料和地面及高空实况观测资料进行匹配。为定量评估再分析资料的性能, 选用均方根误差和相关系数来检验再分析资料对实况观测的偏离程度。相关系数(R_R)公式:

$$R_R = \frac{\sum_{i=1}^N (X_{\text{ERA5},i} - \bar{X}_{\text{ERA5}})(X_{\text{obs},i} - \bar{X}_{\text{obs}})}{\sqrt{\sum_{i=1}^N (X_{\text{ERA5},i} - \bar{X}_{\text{ERA5}})^2} \sqrt{\sum_{i=1}^N (X_{\text{obs},i} - \bar{X}_{\text{obs}})^2}}, \quad (1)$$

均方根误差(R_{MSE})公式:

$$R_{\text{MSE}} = \sqrt{\frac{\sum_{i=1}^N (X_{\text{obs},i} - X_{\text{ERA5},i})^2}{N}}. \quad (2)$$

式中: N 为样本数量, i 为每一个样本, X_{ERA5} 为 ERA5 再分析资料, X_{obs} 为实况观测资料。

1.3 机器学习订正方法

为提升 ERA5 资料的适应性, 减小资料与实况资料的均方根误差, 提高相关系数, 本文采用机器学习的方法对 ERA5 地面温度和相对湿度资料进行订正。机器学习模型选用线性回归中的岭回归算法(Ridge), 采用“K 折交叉验证”的方法将训练集和评估集随机分离, 运用“10 折交叉验证”的方法评估模型准确度。岭回归算法最大的优势是可用于共线性分析的有偏估计回归方法, 通过损失部分信息的代价, 获取更真实可靠的回归方法。本文采用 Python 的 Sklearn 库包, 调用岭回归算法模型来实现^[24]。

2 结果分析

2.1 地面气象要素适用性分析

2.1.1 地面气象要素相关系数分布

由全省站点相关系数分析可知, 2 m 温度相关系数 >0.99 的站点大多集中于辽河流域及其西侧地

区, 大连、葫芦岛地区相关系数相对较低, 但均 >0.94, 也属强相关。2 m 露点温度相较于 2 m 温度, 相关系数 >0.99 的地方较少, 主要分布于沈阳和铁岭地区, 但 ≥0.98 的地区较多, 在全省均有分布; 2 m 露点温度相关系数低值区依然集中于大连和葫芦岛地区。地面气压相关系数均较高, 平均值高达 0.994, 仅有个别站点相关系数出现 <0.95 的情况, 说明 ERA5 再分析资料在地面气压的反演中表现优越。10 m 风速的对比中发现相关系数总体下降, 平均相关系数仅 0.684, 但辽河流域及两侧地区依然是相关系数高值区, 抚顺、丹东、葫芦岛地区最低, 仅达到 0.55~0.60; 大连、本溪部分地区较低。由于风向为 0°~360°, 对比验证不便, 故采用 U 、 V 分量形式对比验证, U 分量代表纬向风, V 分量代表径向风。由 10 m 风场 U 、 V 分量相关系数可知, U 分量相关系数明显低于 V 分量, 这和均方根误差 U 分量高于 V 分量有明显不同, 在业务应用时要特别注意; 且 U 分量相关系数平均值仅 0.666, V 分量相关系数平均值仅 0.818, U 、 V 分量相关系数高值区主要位于辽河流域地区, 丹东、本溪、朝阳地区较低, 最低仅 0.44; V 分量相关系数虽然较 U 分量高, 但 >0.90 的站数仅 13 个, 高相关系数区位于辽河流域及以西区域, 东部站点的相关系数明显偏小。

2.1.2 地面气象要素均方根误差分布

由全省站点均方根误差分布可知, 2 m 温度(图 2a)的均方根误差总体 <2 m 露点温度(图 2b), 2 m 温度均方根误差最低为 1.56 °C, 平均为 2.0 °C, 而 2 m 露点温度均方根误差最低为 1.89 °C, 平均为 2.4 °C; 从空间分布也能直观看出, 2 m 温度的均方根误差在 1.5~2.0 °C 分布较广, 特别是辽河流域以西最为集中, 且仅有个别点均方根误差 >2.6 °C; 2 m 露点温度均方根误差在 1.5~2.0 °C 的仅有 4 个站点, 均分布于辽宁北部的沈阳和铁岭地区, 均方根误差在 2.1~2.5 °C 分布最广, 大连地区的均方根误差较大且分布集中。由地面气压的均方根误差分布(图 2c)可知, 最低为 0.46 hPa, 均方根误差大值区集中于辽宁的东部和西部, 误差最高为本溪桓仁站(32.2 hPa), 均方根误差较低的地方主要集中于辽河流域及其西侧地区。

10 m 风速(图 2d)均方根误差高值和低值在全省分布较为分散, 全省 14 个地市中, 均有误差高值和低值区。经统计发现, 风速误差最低为 1.2 m/s, 最高为 2.9 m/s, 平均为 1.76 m/s, 总体表现良好; >2.0 m/s 的站点多分布于沿海地区, 推测可能是受

海洋影响,ERA5再分析资料反演质量较差造成。根据 U 、 V 风场空间分布可知(图2e、2f),辽宁中北部地区 U 风场均方根误差低于 V 风场,10 m风场 U 分量均方根误差平均值为1.5 m/s,较10 m风场 V 分量总体较低,特别是辽宁中北部地区均方根误差均 <1.5 m/s,大连、朝阳、葫芦岛地区误差相对较高,大连地区的误差最高,但大多 ≤ 2.8 m/s;10 m风场 V 分量误差平均值为2.0 m/s,高误差地区全省分布较为分散,误差 >2.5 m/s的地区位于大连、抚顺、营口和葫芦岛地区。

2.1.3 地面气象要素相关系数及误差逐月变化特征

为研究地面各观测要素和ERA5资料对比状

况,根据各要素在不同月份的平均相关系数和平均均方根误差对比特征,绘制了月变化分布图(图3)。选取了相对湿度(RH)、风速(V)、 U 风场(u)、 V 风场(v)分量等4个月变化特征相对明显的要素,其他月变化特征不明显的要素在此省略。根据相关系数逐月分布可以看出地面相对湿度的相关系数要高于风速、 U 风场和 V 风场,各要素月变化特征明显;1—4月相关系数呈波动上升趋势,4月达到最大值后开始下降,7月达到最低,而后开始上升至10月;总体呈现春秋季节平均相关系数好于夏冬季。因相对湿度(RH)平均均方根误差较大,为清楚显示风速、 U 风场、 V 风场月变化特征,故在图3均方根误差图中右

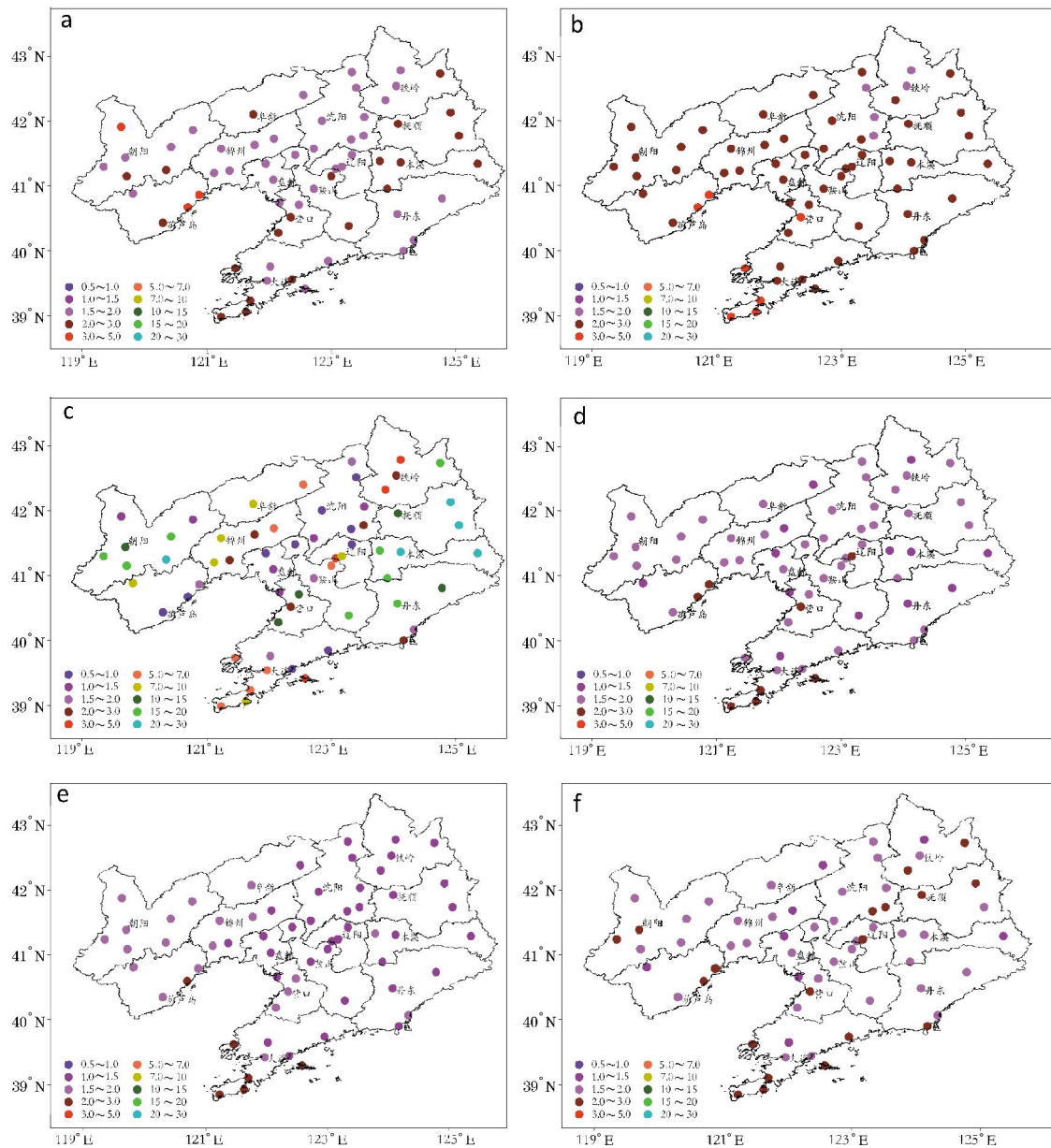


图2 地面气象要素均方根误差空间分布

(a为2 m温度,b为2 m露点温度,c为地面气压,d为10 m风速,e为10 m风场 U 分量,f为10 m风场 V 分量)

侧单独列出相对湿度坐标轴(红色)。由平均均方根误差月分布(图 3b)可以看出,相对湿度的均方根误差最高,所有要素的均方根误差月变化特征明显,春季最高,夏季最低;相关系数高的季节,均方根误差也偏高。 U 、 V 风场的月变化中, U 风场平均相关系数低于 V 风场,但风场的平均均方根误差为 1.773, U 风场平均均方根误差也小于 V 风场;而风速的平均相关系数仅有 0.670,在业务和科研使用中需要注意。

2.2 高空气象要素适用性分析

辽宁省共有 4 个探空气象站,分别为沈阳、大

连、丹东和锦州站,为保证研究区域的全面性,增加辽宁省周围的 5 个探空站进行补充分析,分别是吉林省的长春和临江站,内蒙古的通辽和赤峰站,以及河北省的乐亭站。

2.2.1 高空气象观测要素相关系数分布

根据探空站的 ERA5 再分析资料和实况资料对比相关系数分析发现(图 4),总体上温度的相关系数最高,在高空 300 hPa 以下,相关系数均超过 0.95,300 hPa 以上丹东站和锦州站相关系数略有降低,锦州站在 150 hPa 处最低,为 0.74。从温度相关系数随气压的分布廓线(图 4a)可知,锦州站最低、

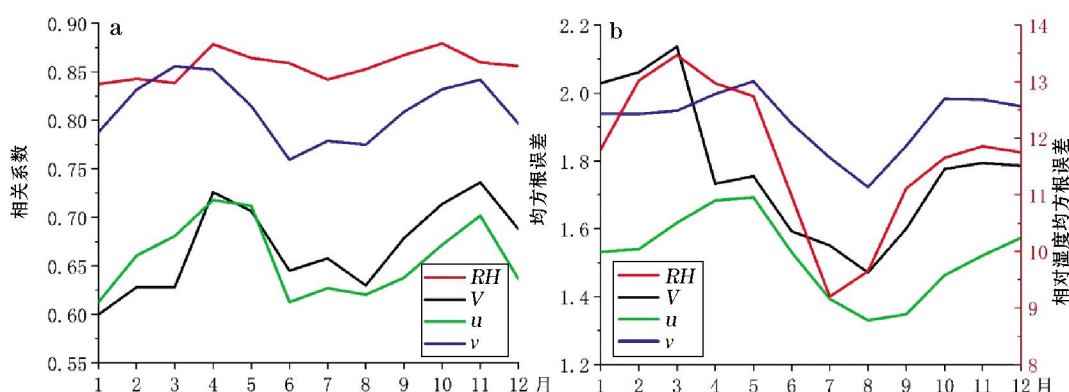


图 3 辽宁省所有地面站点各要素平均相关系数(a)和平均均方根误差(b)逐月分布

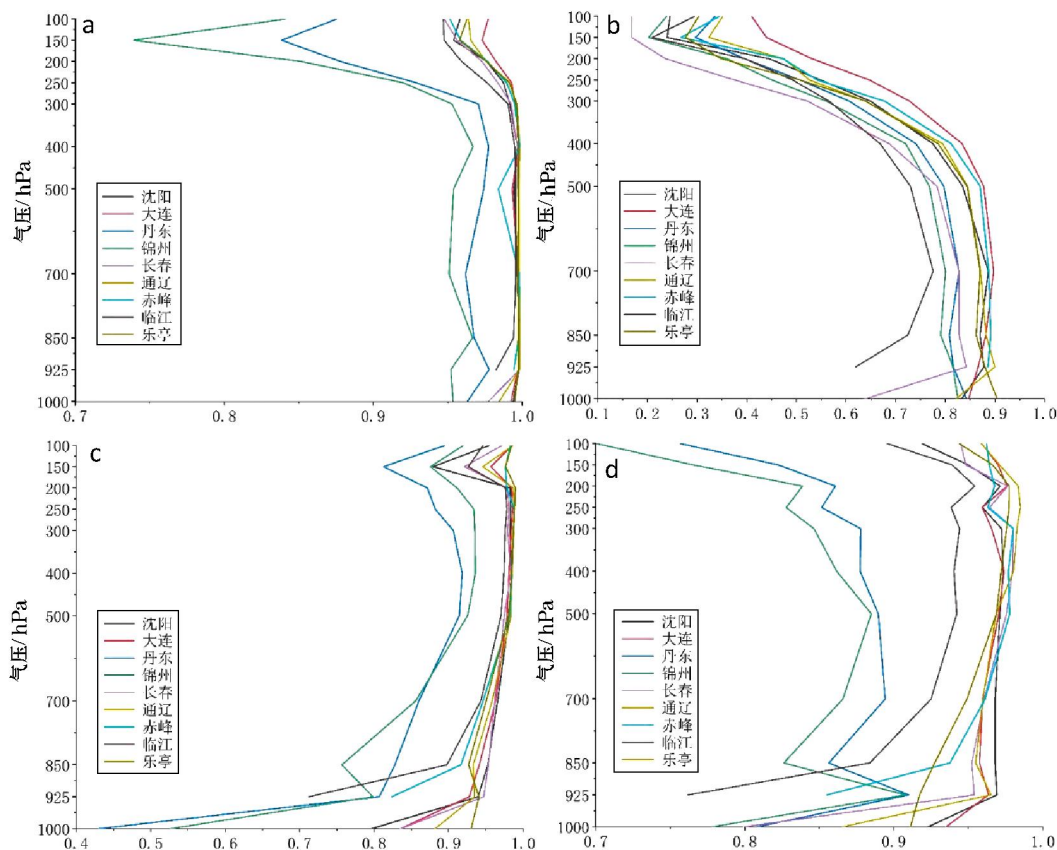


图 4 高空气象观测要素相关系数分布
(a 为温度,b 为相对湿度,c 为 U 风场,d 为 V 风场)

丹东站次之,其他站点分布较为一致,且相关系数随高度变化较小;平均相关系数最高为大连站,达到0.991,其次是通辽站;ERA5高空资料在辽宁中部地区的质量要高于东西部地区,这可能和地形有关,辽宁中部为辽河流域平原地带,海拔较低,东西部分别为山区和丘陵地带,地形可能影响了ERA5资料同化。而ERA5资料的相对湿度(图4b)质量相比于温度稍差,多数站在低层相关系数高于高层,丹东和锦州站相关系数在中层及以下明显较其他站点偏小,大连站相关系数在中层突然大幅度减小,其他层结相关系数较高;相对湿度的平均相关系数最高是大连站,为0.723,其次是通辽站,为0.675,最低是临江站,为0.547;相对湿度的诊断在天气预报中至关重要,在日常业务和研究中,若使用ERA5再分析资料的相对湿度数据时,应谨慎使用。由 U 风场的相关系数(图4c)分布可知,中低层以上相关系数整体较高,且越往高空相关系数基本呈现升高趋势(除150 hPa异常偏低);相关系数在低空普遍较低,1 000 hPa丹东和锦州站偏低尤为明显;除丹东和锦州站,其他站在中层及以上相关系数均超过0.9,相关性较强,说明ERA5资料对东西方向的风整体质量较高,优于

南北方向。相较于 U 风场, V 风场(图4d)相关系数整体偏低,特别是丹东和锦州站的 V 风场相关系数较低,且两站从中层到高层相关系数逐渐减小,100 hPa达到最小,此时锦州站相关系数仅为0.701。除丹东和锦州站外,其他站点的 V 风场相关系数总体较高,整层平均相关系数均超过0.9,平均相关系数最高是大连站,为0.963。

2.2.2 高空气象观测要素均方根误差分布

高空ERA5资料和实测资料对比均方根误差分布如图5所示。根据温度均方根误差分布(图5a)可知,ERA5资料在丹东站和锦州站的均方根误差最大,与这两地的相关系数较低相一致;总体呈现中层均方根误差最小,低空均方根误差整体较高;最大温度均方根误差出现在锦州站的950 hPa,为4.089;各层平均均方根误差最高的为锦州站,均方根误差为3.134,其次为丹东站,最小是通辽站,为0.904。相对湿度的均方根误差(图5b)总体较大,从低空到300 hPa均方根误差上升较快,在300 hPa达到最大,各站在300 hPa平均均方根误差为35.540;从各站平均均方根误差看,通辽站最小,为21.933,最大是临江站,为25.096;在分析天气时,如需研究相对

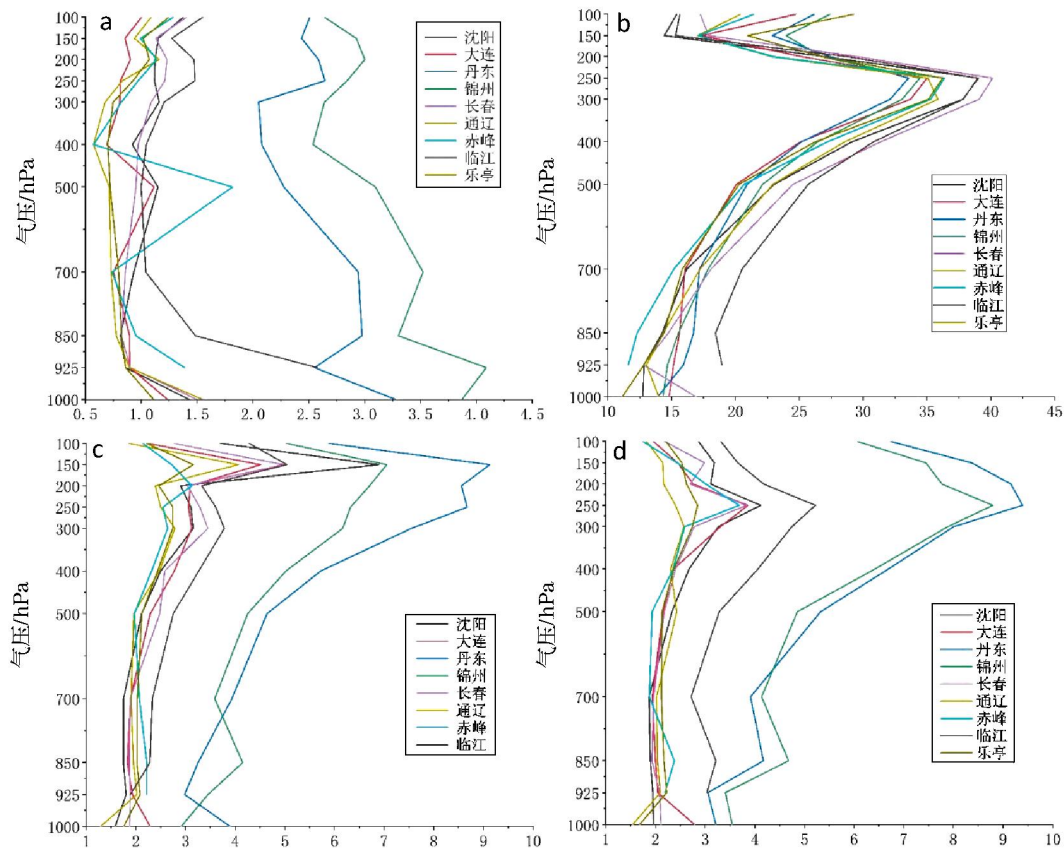


图5 高空观测要素均方根误差分布
(a为温度,b为相对湿度,c为 U 风场,d为 V 风场)

湿度或湿度相关物理量时要特别注意。由 U 风场均方根误差分布(图 5c)可知,误差较大的站点出现在丹东站和锦州站,其中丹东站误差最大,平均均方根误差,为 5.838,其次是锦州站为 4.968,平均均方根误差最小是通辽站,为 2.290;由垂直分布可以看出,大多数站点的均方根误差集中在 2 左右,且越往高空,误差越大,在 150 hPa 误差急剧升高达到最大;误差最大值是 150 hPa 的丹东站,为 9.127。由 V 风场的均方根误差垂直分布(图 5d)可知,误差最大的站点依然出现在丹东站和锦州站,且平均均方根误差值明显超过 U 风场,丹东站平均误差最大,锦州站次之,最大误差出现在 250 hPa 的丹东站,为 9.392;大多数站点的各层均方根误差也都集中于 2 附近,这种分布方式和 U 风场类似,但 V 风场的总平均误差要高于 U 风场;从低层到高层, V 风场均方根误差呈现上升趋势,但在 250 hPa 达到最高,再往高空误差下降,这种分布和 U 风场既有相似之处,又有区别;再次说明,ERA5 资料总体经向风(V 风场)质量低于纬向风(U 风场)。

2.2.3 高空气象要素相关系数及误差逐月变化特征

根据所有探空站点实测资料,与 ERA5 资料各层次一一匹配,绘制了温度(T)、相对湿度(RH)、 U 风场(u)和 V 风场(v)平均相关系数和平均均方根误差逐月分布图(图 6)。由于相对湿度的相关系数较其他要素低,均方根误差较其他要素高,特将相对湿度用单独坐标轴绘制(右侧红轴)。由相关系数分布(图 6a)可知,相对湿度相关系数最低,但其有很强的月变化特征,呈现“单峰单谷”型,从年初相关系数迅速下降,到 4 月达到最低(相关系数为 0.636),然后从 6 月开始快速上升,8 月达到最高,相关系数为 0.750,之后继续下降;可见相对湿度的平均相关系数在春夏之交时最低,夏季相关系数上升较快,夏秋之交时相关系数达到最高。平均相关系数最高的为温度,月变化较小,各月平均相关系数均在 0.99

以上。 U 风场平均相关系数是明显高于 V 风场,这和上文的研究结果相一致, U 风场全年月变化不及 V 风场明显, V 风场相关系数从年初开始升高,5 月达到最高(相关系数为 0.947),而后缓慢下降。从平均均方根误差的分布(图 6b)来看,相对湿度误差分布依然具有很强的月变化特征,呈现“单峰”型,6 月最高(均方根误差为 27.892),1 月最低(均方根误差为 21.524),在相关系数快速上升的月(6—8 月),平均均方根误差也呈现出快速上升的趋势。温度的平均均方根误差最小,年平均值为 1.6,且波动较小。 U 风场和 V 风场的均方根误差相差不大, U 风场稍低一些,而且 U 、 V 风场平均均方根误差月变化较一致,季节性变化明显,冬季最高,夏季最低。

2.3 基于实况资料的 ERA5 再分析资料订正

为提升 ERA5 再分析资料的业务和科研应用,结合前面的分析发现,ERA5 资料温度的数据质量相对较高,而相对湿度的数据质量相对较低,鉴于文章篇幅有限,本次选择了两种具有代表性的要素进行试验订正,订正方法选择机器学习模型,将地面 ERA5 资料的温度和相对湿度进行订正。将每一个国家站点的 ERA5 与实况资料的均方根误差(以下简称为“原误差”)减去机器学习模型订正结果资料和实况资料的均方根误差(以下简称为“订正后误差”),得到均方根误差之差空间分布(图 7)。

由国家站点的地面温度均方根误差之差可知,原误差平均值为 2.1,订正后误差平均值为 1.9,说明基于岭回归算法的机器学习订正方法有效缩小了 ERA5 资料与实况资料的差距;由各站点地面温度原误差与订正后误差空间分布(图 7a)可知,正值误差越大说明机器学习订正算法效果越好,负值越大说明机器学习订正算法起到反作用越大,差值大部分集中在 0~1.0,0~0.5 分布最多,证明机器学习订正算法可有效减小误差 0.5~1.0 °C,其中辽河平原地区提升效果明显;辽宁南部的部分地区订正有反

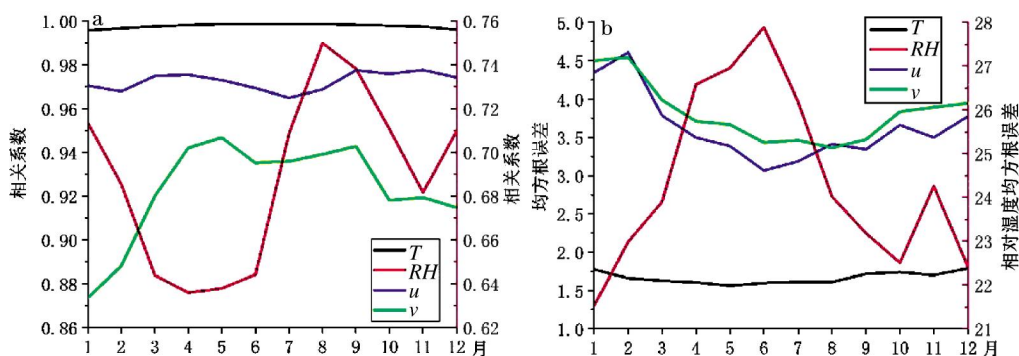


图6 辽宁省所有探空站各气象要素平均相关系数(a)和平均均方根误差(b)逐月分布

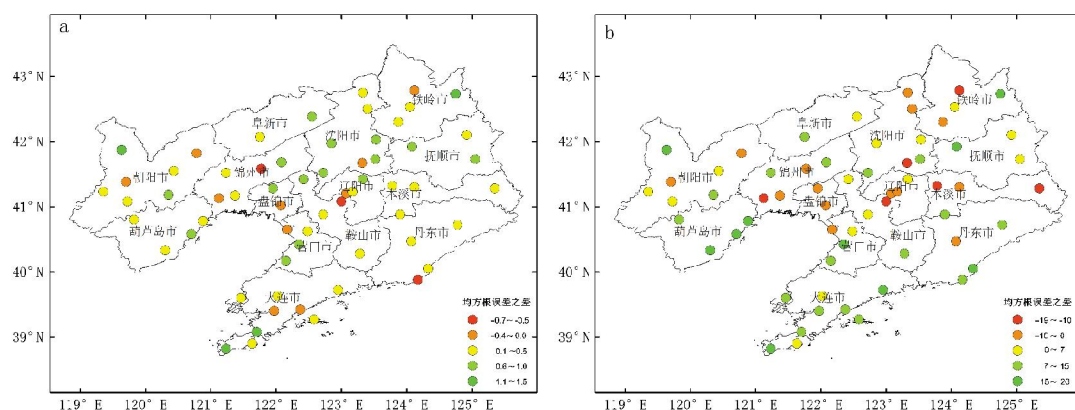


图7 各国家站点地面温度(a)和相对湿度原误差(b)与订正后误差之差空间分布

作用,分布比较分散,但均 $\leq -0.7^{\circ}\text{C}$,业务运用或科研中注意适当调整。由国家站点的地面相对湿度均方根误差之差可知,原误差平均值为11.7,订正后误差平均值为10.3,订正效果好于地面温度;由各站点地面相对湿度原误差与订正后误差空间(图7b)分布可知,相对湿度的订正最好的地区普遍位于辽宁的南部地区,大连、营口、丹东地区分布最集中,辽西的葫芦岛地区也有较多分布,误差最高可降低26%;辽宁中部地区的本溪、辽阳地区订正反作用较为集中,但起到反作用的站点数明显低于正作用的数量。从机器学习的岭回归算法订正结果的综合分析可知,本方法可有效提升ERA5资料的应用能力,有效缩小与实况资料的差距。

3 结论与讨论

(1)地面要素相关系数中2 m温度整体较高,大多集中于辽河流域及以西地区,2 m露点温度相关系数稍低于2 m温度,但大多超过98%,2 m温度均方根误差<2 m露点温度;地面气压相关系数均较高,平均值高达0.994,辽宁的东部和西部地区均方根误差较大。高空站点温度的相关系数最高,低层均方根误差整体高于中高层,相对湿度相关系数整体较低,均方根误差较大。

(2)地面10 m风速平均相关系数整体相对较低,辽河流域及两侧地区依然是相关系数高值区,均方根误差全省分布不均;地面风场 U 分量相关系数明显低于 V 分量,且辽宁中北部地区 U 分量均方根误差低于 V 分量。高空站点辽宁中部地区的数据质量要高于东部、西部地区; U 风场的相关系数从中低层以上整体比较高,且往高空相关系数呈现升高趋势,均方根误差集中在2左右,且越往高空,误差越大;ERA5资料在高空总体经向风(V 风场)质量低

于纬向风(U 风场)。

(3)地面相对湿度、风速、 U 风场和 V 风场平均相关系数和均方根误差月变化特征明显,相对湿度相关系数和均方根误差整体最高,相关系数总体呈现春秋季节高于夏冬季,所有要素的均方根误差分布呈现春季高、夏季低的状态,且相关系数高的季节,均方根误差也偏高; U 风场相关系数低于 V 风场, U 风场均方根误差也低于 V 风场。

(4)高空资料逐月分布中,温度的相关系数最高,均方根误差最小; U 风场平均相关系数明显高于 V 风场;相对湿度相关系数在春夏之交时最低,夏季相关系数上升较快,夏秋之交时达到最高;而相对湿度在相关系数快速上升的月,均方根误差也呈现出快速上升的趋势。 U 风场和 V 风场的均方根误差相差不大,且均方根误差月变化相一致。

(5)通过机器学习的订正方法可有效提升地面温度和相对湿度的应用能力,ERA5地面温度资料的均方根误差减小主要集中在辽河流域,误差减小0.5~1.0 $^{\circ}\text{C}$;地面相对湿度资料的均方根误差减小主要分布于辽宁南部地区,降低误差最高达到26%。

参考文献:

- [1] BENGTSSON L. FGGE 4-dimensional data assimilation at ECMWF [J]. Bulletin of the American Meteorological Society, 1982, 63: 29-43.
- [2] GIBSON J K, HERNANDEZ A, KALLBERG P, et al. The ECMWF reanalysis project (ERA)-plans and current status [R]. Amer Meteor Soc, 1994.
- [3] UPPALA S M, KALLBERG P W, SIMMONS A J, et al. The ERA-40 reanalysis [J]. Quarterly Journal of the Royal Meteorological Society, 2005, 131(612): 2961-3012.
- [4] SIMMONS A, UPPALA S, DEE D, et al. ERA-Interim: new ECMWF reanalysis products from 1989 onwards [J]. ECMWF Newsletter, 2007, 10: 25-35.

- [5] HERSBACH H.The ERA5 atmospheric reanalysis[C].AGU Fall Meeting, 2016.
- [6] 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.
- [7] 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.
- [8] 秦艳慧,吴通华,李韧,等.ERA-Interim 地表温度资料在青藏高原多年冻土区的适用性[J].高原气象,2015,34(3):666-675.
- [9] 滕华超,陈艳春,汤子东.ERA-Interim 风速再分析数据在山东省的适用性评估[J].山东气象,2016,36(1):27-31.
- [10] 白磊,王维霞,姚亚楠,等.ERA-Interim 和 NCEP/NCAR 再分析数据气温和气压值在天山山区适用性分析[J].沙漠与绿洲气象,2013,7(3):51-56.
- [11] 支星,徐海明.3 种再分析资料的高空温度与中国探空温度资料的对比:年平均特征[J].大气科学学报,2013,36(1):77-87.
- [12] 支星,徐海明,米卫红,等.中国东北和西北地区三种再分析资料与探空资料的温度对比[J].气象与环境学报,2014,30(6):147-157.
- [13] 陈艳春,王娜,顾伟宗,等.环渤海区域再分析资料地面风速场的适用性对比分析[J].海洋气象学报,2017,37(1):67-72.
- [14] 袁松,程华,王东勇,等.模式再分析与实际探空资料的对比分析[J].气象科学,2012,32(1):62-67.
- [15] 高淑新,宋晓巍,李琳琳,等.CLIGEN 天气发生器在中国东北三省模拟温度的适用性评价[J].气象与环境学报,2019,35(4):77-84.
- [16] 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.
- [17] 郑艳萍.ERA5 再分析资料在广东省的适用性初步分析[C]//中国气象学会.第 35 届中国气象学会年会 S20 深度信息化:应用支持与智能发展.中国气象学会,2018:7.
- [18] 王秀明,俞小鼎,朱禾.NCEP 再分析资料在强对流环境分析中的应用[J].应用气象学报,2012,23(2):139-146.
- [19] 夏凡,杨晓霞,刘诗军,等.WRF 模式 2 m 温度在山东预报效果的评估[J].气象与环境学报,2019,35(5):10-17.
- [20] 许敏,丛波,刘艳杰,等.廊坊地区 5 种数值模式降水预报性能检验与评估[J].气象与环境学报,2016,32(1):09-15.
- [21] 李翔翔,黄淑娥,杨军,等.多种再分析地面气温资料在江西省的适用性[J].气象科技,2020,48(6):877-886.
- [22] 陈艳春,王娜,顾伟宗,等.环渤海区域再分析资料地面风速场的适用性对比分析[J].海洋气象学报,2017,37(1):67-72.
- [23] 邓小花,翟盘茂,袁春红.国外几套再分析资料的对比与分析[J].气象科技,2010,38(1):1-8.
- [24] 魏贞原.机器学习 Python 实践[M].北京:电子工业出版社,2018:211.

Applicability Analysis and Revision of ERA5 Reanalysis Data in Liaoning Province

SUN Hongyu^{1,2}, WANG Di³, LIU Chenghan², ZHANG Yan⁴, WANG Shigong^{1,5}

(1.School of Atmospheric Sciences, Chengdu University of Information Technology/Plateau Atmosphere and Environment Key Laboratory of Sichuan Province, Chengdu 610225, China;

2.Liaoning Meteorological Observatory, Shenyang 110166, China;

3.Henan Meteorological Observatory, Zhengzhou 450003, China;

4.Liangshan Meteorological Bureau, Liangshan 615000, China;

5.Chengdu Plain Urban Meteorology and Environment Sichuan Provincial Field Scientific Observation and Research Station, Chengdu 610225, China)

Abstract In order to systematically study the applicability of ERA5 reanalysis data in Liaoning province, this study extracted temperature, humidity, pressure, wind field and other commonly used physical quantities from 62 ground stations and 9 sounding stations (4 within the province and 5 outside the province) in Liaoning province during 2009–2018. Statistical analysis was used to analyze the applicability of ERA5 data in Liaoning province, and machine learning methods were attempted to correct the deviations in ERA5 data. The results show that in the analysis of ground observation elements, the correlation coefficient of 2 m temperature is generally high, and the root mean square error is generally low, meanwhile, the effect is better in the Liaohe River Basin and its western region than in other areas; The comparison results of 10 m wind speed are generally poor, and the correlation coefficient of U -component is generally lower than that of V -component; There is a clear monthly variation in the quality of ERA5 data, with the correlation coefficient generally higher in spring and autumn than in summer and winter. In seasons with high correlation coefficients, the root mean square error is generally high. In the analysis of sounding observation elements, the data quality in the middle of Liaoning province is higher than that in the eastern and western regions; The average correlation coefficient of temperature is the highest, while the average correlation coefficient of relative humidity is low, and the root mean square error is generally large; The correlation coefficient of relative humidity is lowest in the transition period between spring and summer, rises rapidly in summer, and reaches the highest in the transition period between summer and autumn. In months with a rapid increase in correlation coefficient, the root mean square error also shows a rapid increase; The root mean square errors of upper U wind field and V wind field are basically the same, and their monthly variation are consistent. The quality of meridional wind (V wind field) is lower than that of zonal wind (U wind field) in general. The correction method of machine learning effectively improves the application ability of ground temperature and relative humidity, reduces the root mean square error of ERA5 ground temperature data by 0.5~1.0 °C, and reduces the root mean square error of ground relative humidity data by up to 26%.

Key words reanalysis data; applicability; root mean square error; correlation coefficient; machine learning