

一种基于图像指纹历史相似天气快速检索技术及其应用*

唐健¹⁾ 代刊¹⁾

¹⁾(国家气象中心, 北京 100081)

摘要: 针对天气预报业务对历史相似个例参考的持续需求, 以及传统检索方法在效率、客观性方面的不足, 文章提出了一种基于图像指纹的历史相似天气快速检索技术。该技术首先借鉴图像学的指纹相似性检测算法, 构建了气象要素场相似度量 and 单要素检索算法; 基于天气学原理, 融合高空环流与地面要素场的耦合特征, 设计了多要素协同检索算法, 以期更准确地识别与目标天气过程相似的历史降水场及对应环流背景。基于上述技术, 采用“云+端”架构构建了历史相似天气检索平台并在中央气象台投入业务试用, 已在多次天气过程预报中展现出较好的应用潜力与实用价值。统计检验结果表明, 该方法检索的相似个例在距平相关系数和均方根误差等指标上均显著优于气候态。最后对未来算法优化、客观评估及深化应用等方面进行探讨。

关键词: 历史相似天气; 相似检索算法; 多要素检索; 天气预报; 图像指纹

Fingerprint-Based Historical Analog Weather Retrieval Technique and Its Application

TANG Jian DAI Kan

National Meteorological Centre, Beijing 100081

Abstract: Addressing the persistent demand for historical analogue references in operational weather forecasting and the deficiencies of traditional retrieval methods in efficiency and objectivity, this paper proposes a rapid retrieval technology for similar historical weather events based on image fingerprinting synergy. This technology employs fingerprint similarity detection algorithms from imageology and establishes a similarity quantification method and single-variable synergistic query algorithm. Guided by synoptic principles and considering the coupling characteristics between upper-air circulation and surface variables, we design a multi-variable synergistic query algorithm to more accurately identify historical precipitation fields and the corresponding circulation patterns similar to a target weather process. Based on these techniques, a historical similar weather retrieval platform is developed using a "Cloud + Client" architecture, which enables efficient retrieval of historical analogues. The platform has been operationally

* 中国气象局重点创新团队智能预报技术团队项目 CMA2022ZD04 和新疆维吾尔自治区重点研发专项 (2022B03027-2) 共同资助

第一作者: 唐健, 主要从事集合预报、诊断分析预报技术研究。

通讯作者: 代刊, 主要从事集合预报、定量降水预报、预报检验评估研究。E-mail: daikan@cma.gov.cn

trialed in the National Meteorological Centre, demonstrating good application potential and practical value in forecasting several weather events. Statistical verification indicates that the similar cases retrieved by this method significantly outperform climatology in terms of anomaly correlation coefficient and root mean square error. Finally, this paper discusses future development directions for technology, particularly in algorithm optimization, objective evaluation, and deepening application.

Key words: historical analog weather, similarity retrieval algorithm, multi-variable retrieval, intelligent weather forecasting, platform architecture

引言

在气象预报业务实践中,参考历史相似天气过程是预报员积累预报经验、构建预报思路的常用且有效途径(陈兴芳和赵振国,2000)。即便在数值天气预报(Numerical Weather Prediction, NWP)和人工智能气象模型等技术迅猛发展的今天,历史相似个例及其衍生的相似预报方法依然具有不可替代的参考价值,尤其对于提升预报员的综合分析能力和应对极端天气事件的预报水平具有重要意义;而通过合理设计的相似算法,则可以从相似个例中转化或者提取具有业务价值的预报参考信息(张静和何俊仕,2012;钱磊等,2022;李颖等,2023)。然而,在实时业务中,如果完全依靠预报员的经验和记忆,手动从海量的历史气象资料中查找与当前或未来天气过程相似的历史个例,不仅效率低下、主观性强,且难以满足业务对时效性、客观性和准确性的要求。

为解决这一难题,国内外学者已开展了诸多研究,尝试将历史相似天气个例的检索自动化并应用于业务实践,相关研究大致可归为两类。一是通过人工整理和经验归纳,构建特定区域或特定类型(特别是灾害性天气)的历史天气个例库,并记录其关键特征,如影响系统、强度、路径、发生时间等,辅以相应的观测和预报数据,供后续检索(董芹等,2009;黄成南等,2013;焦圣明等,2017;孙卓等,2020)。这类方法高度依赖于历史个例的质量,且构建和维护成本较大。二是通过计算机算法自动匹配历史相似场。传统方法多采用基于格点场的直接比较,如计算欧氏距离、空间相关系数、相似离度等指标来衡量天气场的相似性(鲁小琴等,2013;李开乐,1997;Monache et al,2013)。然而,随着历史数据量的爆炸式增长和气象要素场的高维特性,此类方法的计算和存储开销巨大,难以满足业务实时性的需求。

近年来,部分研究者借鉴了图像处理领域的图像指纹(Image Fingerprinting)

或感知哈希 (Perceptual Hashing) 技术, 将二维或三维气象要素场视为图像, 通过提取其压缩性特征来实现快速检索 (Franch et al, 2019)。这种方法通过将高维天气场映射到低维特征空间, 再利用局部敏感哈希 (Locality Sensitive Hashing, LSH) 等近似最近邻搜索技术, 可以显著降低计算复杂度, 同时较好地保留天气形势的整体形态特征 (闫旭等, 2022)。例如, Raoult et al (2018) 提出了一种基于小波变换的指纹索引方法, 用于快速检索相似天气场。该方法通过对气象场进行小波压缩, 生成指纹并建立索引, 显著提高了检索效率; 闫旭等(2022) 利用 LSH 算法对 1980—2019 年再分析资料构建指纹索引, 实现了相似天气形势的快速检索, 单次检索时间虽有大幅缩短但仍需近 4 分钟, 这对于高时效性的业务需求而言仍有提升空间。在天气形态的深度学习相似识别方面, 周必高等(2022) 利用改进 VGG16 模型实现了副热带高压结构的相似检索, 验证了以图像特征衡量大尺度环流相似度的可行性。

尽管图像指纹方法展现出巨大潜力, 但在气象领域的应用仍面临挑战: (1) 历史样本依赖性: 若当前天气型在历史记录中缺乏相似先例, 方法有效性会受限, 这也是相似查询方法的天生缺陷, 很难克服; (2) 多要素、多层次信息表达: 如何在紧凑的指纹中有效融合多个关键气象要素及不同垂直层次的信息, 是一个需要深入探索的问题; (3) 空间与尺度不变性: 传统的图像指纹方法对天气系统的空间平移、旋转和尺度变化可能不够鲁棒, 可能低估某些形态相似但位置或强度略有偏差的天气系统的相似性; (4) 近似搜索的潜在遗漏: 近似哈希机制可能导致部分最优匹配被遗漏, 需要设计补偿策略。尽管存在上述挑战, 图像指纹技术在气象相似预报中的应用前景广阔 (郝翠等, 2022; 乔丹杨等, 2023; 任福民等, 2023; 赵亮等, 2024; 黄晓远等, 2025)。它不仅能辅助预报员快速做出判断和订正, 还能借鉴集合预报的思想, 提供多样化的历史相似成员, 为量化预报不确定性提供新思路 (Kratzert et al, 2021)。

基于人工智能的相似性检索方法在气象领域得到了广泛应用, 尤其是在极端天气预测和历史天气模式匹配方面。如 Chattopadhyay et al (2020) 提出了一种基于胶囊神经网络的类比预报框架, 该框架通过识别与当前天气模式相似的历史大气环流模式, 成功预测了北美地区的极端天气事件, 准确率达到 80%。Hu et al (2023) 引入了一种基于深度学习的相似性度量方法, 通过训练神经网络生成

潜在特征，克服了传统方法中对预测因子的选择和优化的瓶颈，从而提高了近地面变量的预测准确性。这些研究表明，人工智能方法在气象领域的相似性检索中具有巨大的潜力。通过构建高效的相似性度量和匹配算法，可以实现对历史天气模式的快速检索，为极端天气的预警和研究提供支持。

基于上述背景，本研究在现有图像指纹技术基础上，以提升业务应用中的实用性为目标，提出一种改进的相似天气检索方法。该方法充分考虑天气学机理，特别是高空环流场与地面要素场的协同作用，构建了多要素检索框架。基于此框架，已设计和实现了“云+端”架构的业务化平台，在中央气象台开展了业务试用并取得了良好的反馈。本文重点对该方法的核心算法、应用效果等进行阐述，并对未来发展方向进行展望。

1. 数据

构建准确、全面的实时和历史样本库是相似天气检索技术的基础。本研究主要使用模式再分析资料和数值预报产品，涵盖高空环流形势和地面气象要素等。

1.1 再分析数据

历史样本库主要基于欧洲中期天气预报中心（ECMWF）提供的第五代全球大气再分析资料 ERA5 (Hersbach et al, 2020)，其空间分辨率为 $0.25^\circ \times 0.25^\circ$ ，时间分辨率为 1h。本研究选取的地面要素主要包括：降水、2 m 气温、2 m 露点温度、10 m 风场、海平面气压、总云量、总可降水量；高空要素主要包括：200、500、700、850、925 hPa 等不同层次的位势高度、风场、气温、相对湿度、垂直速度、位涡和散度等。

为了与国家级天气预报业务中逐日降水预报的时段划分（08 时至次日 08 时及 20 时至次日 20 时；全文均为北京时）保持一致，将 ERA5 逐小时降水数据累积处理为这两个时段的 24 h 累计降水量。选用 ERA5 降水作为历史样本，主要基于以下考虑：一是 ERA5 同化了大量的地面和卫星观测数据，能够较为真实地反映历史降水过程的强度和空间分布，具有接近“观测真值”的参考价值；二是再分析降水本身兼具分析场特征和一定的模式预报特征，在未来平台进行实时相似个例匹配时，以其作为“历史参考场”与数值模式预报结果进行比对，相较于实况降水而言可提供更具业务参考意义和可解释性的相似性评估。

本文使用的 ERA5 数据时间范围为 1959 年 1 月 1 日至 2024 年 12 月 31 日

(所建平台中所有数据为实时更新)。高空环流场数据选择与 24 h 累计降水量相应时段匹配的中间时次，进行环流形势与地面降水的联合相似性分析。

1.2 预报数据

为了使平台具备实时相似个例检索与评估能力，本文使用了中央气象台业务接收的 ECMWF-IFS (Integrated Forecasting System) 高分辨率确定性数值预报产品。选用的气象要素与 ERA5 再分析数据基本保持一致，包括降水预报和多层次高空要素预报。该数据空间分辨率为 0.125° 。在数据处理方式上，同样将预报的 24 h 累计降水按天气预报业务中逐日降水时段进行逐日累计处理，以对齐预报时次。预报数据的时间范围为 2017 年 1 月 1 日至 2024 年 12 月 31 日。

1.3 检验与评估数据

在对检索出的历史相似个例进行客观评估时，以气候态的日降水和环流数据作为参照基准。本文采用的降水气候态数据来源于全球降水测量计划(GPM) 的卫星反演降水产品 (IMERG) 2001—2022 年日降水气候平均数据集(Huffman et al, 2020, <https://gpm.nasa.gov/data/imerg/precipitation-climatology>)，数据集基于多源卫星微波观测数据融合处理，具有 0.1° 的空间分辨率和日时间尺度的长期平均降水信息，适用于全球范围的降水气候特征分析。环流背景的气候态数据则选用 NCEP/NCAR 再分析 1981—2010 年月平均资料 (Kalnay et al, 1996; <https://psl.noaa.gov/data/gridded/data.ncep.reanalysis.html>)。该数据涵盖自 1948 年以来的大气各层标准等压面要素，空间分辨率为 2.5° ，能有效反映中高层环流形势的月平均气候态特征，被广泛应用于气候诊断和大尺度天气系统分析。

2. 关键技术

历史相似天气快速检索过程中的核心技术主要包括两方面：一是构建单要素相似天气检索的基础算法(以下简称单要素检索)，用于度量单个气象要素场(如降水场或任意单要素场)之间的相似程度；二是在此基础上，引入多要素检索策略(以下简称多要素检索)，通过综合考虑多种相关气象要素的信息，实现多变量联合检索下的相似度评估与个例筛选，以更贴近天气学分析的综合视角。

2.1 历史资料库构建

基于 1.1 节所述的 ERA5 再分析数据为历史实况样本以及 ECMWF-IFS 模式预报数据为预报样本，选取关注的降水量及其相关的关键高空、地面气象要素，

经过时段累积、空间插值统一格点等预处理，构建结构化的历史资料库，为后续的相似检索提供数据支撑。该资料库不仅存储要素场的原始格点数据，还预先计算并存储了每个要素场的“指纹”信息，以加速检索过程。

2.2 基于感知哈希的格点场指纹检索

单要素检索算法的目标是，给定一个目标天气要素场和特定区域，系统能够从海量历史资料库中快速返回与之最相似及其他多个高度相似的历史个例。本研究借鉴了计算机图像相似性检索中广泛应用的感知哈希算法 (Perceptual Hashing Algorithm, pHash)。该算法的核心思想是为每个图像（在此即为天气要素场）生成一个紧凑的“指纹” (fingerprint) 字符串，通过比较不同图像指纹间的汉明距离 (Hamming distance) 来判定它们的相似性。汉明距离越小，表明图像内容越接近。以一个二维气象要素场为例，感知哈希算法的基本流程通常包括以下步骤：

(1) 尺寸归一化与预处理：将输入的要素场数据插值或裁剪到统一的尺寸（如 8×8 像素，或根据要素场特性选择如 32×32 像素或 64×64 像素等更合适的尺寸以保留更多细节）。此步骤旨在消除尺度差异，并降低计算复杂度。

(2) 灰度化处理：将归一化后的要素场进行灰度化处理。随后，可以采用离散余弦变换、小波变换特征提取方法，选取其中的低频分量，因为这些分量通常代表了要素场的主要结构信息。例如，对 8×8 像素的灰度图进行离散余弦变换，然后保留左上角的若干低频系数。对上一步选取的特征系数计算平均值或中位数。

(3) 二值化：比较每个特征系数与步骤 (2) 计算得到的均值。若系数大于或等于均值，则对应指纹位记为 1；若小于均值，则记为 0。

(4) 构建指纹：将所有比较结果串联起来，形成一个二进制字符串，进而得到该要素场的感知哈希指纹。

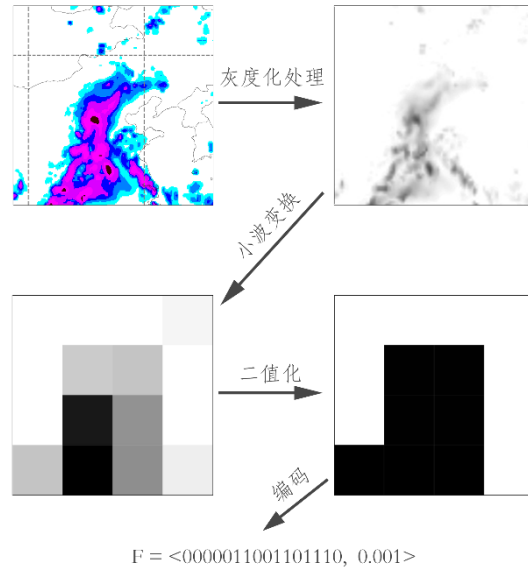


图 1 指纹构建过程示意图：以降水场为例

Fig. 1 Schematic diagram of the fingerprint construction process based on precipitation field

本研究中，定义拟构建要素场 f 的指纹为 F ， $F(f) = \langle s, r \rangle$ 。式中： s 为通过上述感知哈希过程生成的、表征要素场空间形态特征的二进制指纹字符串； r 则为表征要素场整体强度特征的某个标量值（例如，区域平均降水量、最大降水量，或降水场的某个统计矩）。图 1 显示了以降水场为例的指纹构建过程。如图所示，将降水场输入后，按上述步骤，最终编码后生成感知哈希指纹 $\langle 0000011001101110, 0.001 \rangle$ ，其中二进制字符串“0000011001101110”表征降水场空间形态特征，0.001 表征降水场整体强度。

按上述方法，得到某个要素场的指纹 $F_1 = \langle s_1, r_1 \rangle$ 和 $F_2 = \langle s_2, r_2 \rangle$ 后，指纹之间的相似度 $\text{Similarity}(F_1, F_2)$ 通过基于 s_1 和 s_2 汉明距离的形态相似性和基于 r_1 和 r_2 差异的强度相似性进行计算，具体公式如下（Raoult et al, 2018）：

$$\text{Similarity} = (1 - \alpha) \times \frac{\text{hamming}(s_1, s_2)}{\text{nbits}} + \alpha \times |r_1 - r_2| \quad (1)$$

式中： $\text{hamming}(s_1, s_2)$ 表示两个等长字符串 s_1 和 s_2 之间的汉明距离，即对应位置不同字符的个数，例如字符串“011011101”“110001001”之间的汉明距离为 4； α 为权重系数，用来调整形状特征和强度特征在总相似度中的比重；nbits 为形状特征的位数。

2.3 单要素与多要素相似度算法

任何复杂的天气过程，尤其是暴雨、强对流等灾害性天气，其发生发展都是

特定尺度天气系统在有利的环流背景下相互作用的结果 (Doswell III, 1987)。因此, 仅依赖单一要素进行相似性检索, 易出现难以全面捕捉天气过程本质特征的情况。例如, 若仅基于单一降水场查询, 可能返回的相似个例虽然降水分布相似, 但其对应的天气尺度或大尺度环流背景可能差异显著, 从而导致预报参考价值降低。反之, 若仅依赖单一环流场 (如 500 hPa 位势高度场) 查询, 则可能难以筛选出与目标降水分布也匹配的个例。为解决这一问题, 本研究设计了多要素检索的概念, 旨在综合评估多个关键气象要素场的相似性。

具体而言, 定义两个或多个气象要素 (例如, 降水场、850 hPa 风场、500 hPa 位势高度场等) 的协同相似度 (total similarity, St) 为各单要素相似度的加权集成。计算公式如下:

$$St = \sum_{i=0}^n \alpha_i * S_i \quad (2)$$

$$\sum_{i=0}^n \alpha_i = 1, \alpha_i \geq 0$$

式中: 相似度 S_i 是第 i 个要素场采用 2.2 节单要素检索基础算法计算得到的相似度得分, α_i 是分配给第 i 个要素场的权重系数。

权重系数可根据天气学经验由预报员人工赋值, 例如在暴雨预报中, 可给予降水场和低空急流场更高的权重; 或者, 在缺乏先验知识时, 可取所有要素的平均权重, 并加以人工进行迭代调整。研究中发现, 权重系数的选择对最终检索结果具有较强的敏感性。不同的天气类型或不同的预报关注点 (如降水落区、降水中心强度等), 可能需要不同的权重组合才能获得最优的匹配效果。未来可以探索基于机器学习方法自适应优化权重分配的策略。

通过对所有历史个例计算其与目标场的协同相似度, 并进行排序, 即可返回协同相似度最高的若干个历史个例作为最终的综合匹配结果。

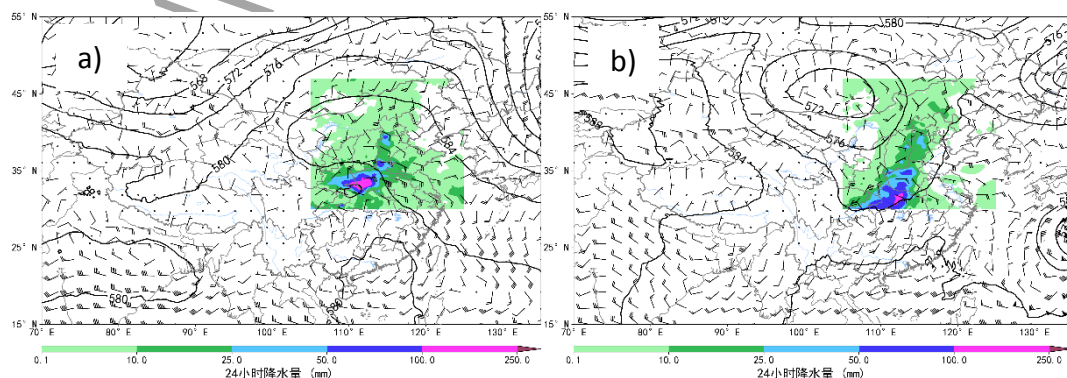
2.4 单要素与多要素检索效果对比

为了直观展示多要素检索相较于单要素检索的区别, 本文以“75·8”河南特大暴雨事件为例, 选取 ERA5 再分析数据中的 1975 年 8 月 7 日 20 时至 8 日 20 时的 24h 累计降水量作为输入目标。

首先进行单要素检索，仅使用 24 h 累计降水场作为输入（图 2a），从历史库中检索。需要说明的是，图 2a 中还额外显示了对应的中间时刻高空环流场，以方便与不同检索方法得到结果进行对比。图 2b 为返回结果中相似度最高的结果，即 2015 年 7 月 14 日 20 时至 15 日 20 时降水过程，从雨带的空间分布和强度上不难发现，其与输入降水场具有一定的相似性，可见西南—东北向带状降水，但目标区域受 500hPa 低涡控制，而输入原始场则高层受高压坝、低层受台风残涡北侧偏东气流控制。即单要素检索所得个例，降水场形态相似，但环流场不相似。

接下来进行多要素检索。从天气学角度判断，将 500hPa 位势高度、低层 850 hPa 风场等动力因子作为辅助协同检索因子，与降水场共同纳入相似性检索。24h 累计降水量、850hPaU 风、850hPa V 风、500hPa 位势高度的权重分别设置为 0.3，0.2，0.2，0.3。图 2c 显示多要素检索返回的最相似个例，为 2000 年 7 月 4 日 20 时至 5 日 20 时。可见，加入动力因子后，多要素检索所得个例，既保留了雨带的相似性，低层风场也更为相似。

进一步加入湿度场因子，用 700hPa 相对湿度场替换 850hPaU 风，进行多要素检索。24h 累计降水量、850hPa V 风、700hPa 相对湿度、500hPa 位势高度的权重分别设置为 0.3，0.2，0.2，0.3。如图 2d 所示，返回的最相似个例为 1963 年 8 月 5 日 08 时至 6 日 08 时，此次检索返回的个例不仅在降水分布上保持了较高的相似性，对应环流背景也与输入源更为接近，“75·8”暴雨为台风残涡北上（丁一汇，2015），而“63·8”暴雨则是西南低涡北上与太行山地形共同影响。这样的检索结果更能满足预报员从天气系统和动力机制角度理解和借鉴历史个例的需求，整体相似度和预报参考价值得到提升。



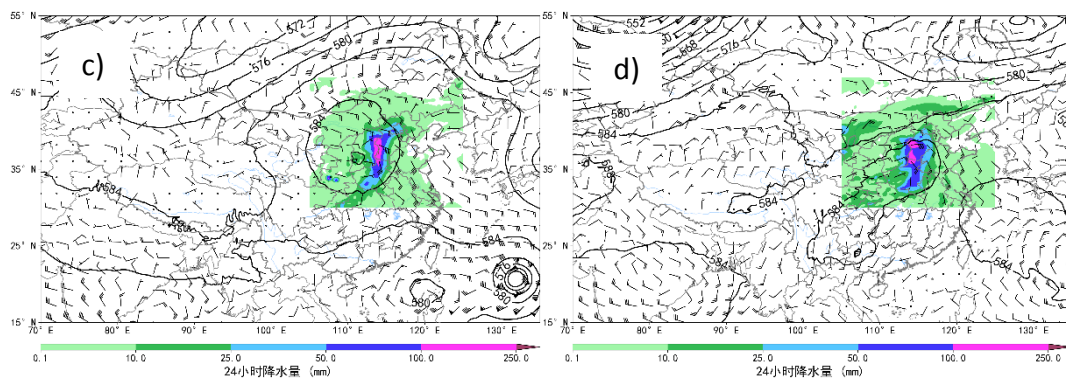


图 2 以“75·8”河南特大暴雨过程的 24h 累计降水量为 (a) 目标输入, (b) 单要素和 (c, d) 多要素检索得到的最相似个例及其相应的 500hPa 位势高度 (等值线, 单位: dagpm) 和 850hPa 风场 (风羽)

(a) 目标输入: 1975 年 8 月 7 日 20 时至 8 日 20 时

(b) 单要素 (24 h 累计降水量) 检索: 2015 年 7 月 14 日 20 时至 15 日 20 时,

(c) 多要素 (增加动力因子) 检索: 2000 年 7 月 4 日 20 时至 5 日 20 时,

(d) 多要素 (进一步增加湿度场因子) 检索: 1963 年 8 月 5 日 08 时至 6 日 08 时

Fig. 2 (a) The 24 h accumulated precipitation during the August 1975 Henan extreme rainstorm taken as target input, and the most similar cases retrieved by (b) single element and (c, d) multiple elements and their corresponding 500 hPa geopotential height (contour, unit: dagpm) and 850 hPa wind field (barb)

(a) target input: from 20:00 BT 7 to 20:00 BT 8 August 1975, (b) single element (24 h accumulated precipitation) search: from 20:00 BT 14 to 20:00 BT 15 July 2015, (c) multi-element (adding power factor) search: from 20:00 BT 4 to 20:00 BT 5 July 2000, (d) multi-element (further adding humidity field factor) search: from 08:00 BT 5 to 08:00 BT 6 August 1963

需要指出的是, 在特定需求场景下, 例如在决策气象服务中, 由于决策部门往往更关注历史上降水量级相近的相似过程及其灾害影响, 以降水场为输入的单要素检索对于快速定位历史个例并对比仍具有应用价值和参考意义。

3. 平台实现

历史相似天气检索平台采用“云+端”架构进行设计与开发, 前端采用 Vue.js 构建, 后端基于 Python 开发, 并通过 Docker 部署以确保环境一致性与系统的可移植性; 数据接口层对接天擎大数据平台统一管理多源气象数据; 历史资料库

及预计算指纹信息存储于 PostgreSQL 高性能关系数据库中；确保历史相似天气检索系统在业务环境中的实用性、可维护性及未来推广性。

平台的网页用户界面设计力求直观易用，主要划分为 4 个功能区域(图 3)。

(1) 检索参数输入区。用户在此可指定目标天气过程的实况或预报时段，选择数据来源。用户可通过预设区域列表选择目标区域，指定参与多要素检索的各要素分配权重。(2) 目标个例实时显示区。根据检索参数输入区条件，系统会实时加载并显示目标天气过程的要素场空间分布，便于用户在检索前对目标个例进行确认和初步分析。(3) 检索结果筛选与信息展示区。用户可在此设置更多筛选条件，如相似个例与目标个例的最小时间间隔，以避免检索到过于临近的、非独立事件，特定年份范围、月份范围、模式起报时间、预报时效等，对初步检索结果进行精细化过滤。同时，该区域也会显示检索任务的进度、检索到的相似个例数量等信息。(4) 相似个例结果展示区。以瀑布流形式展示最终筛选出的历史相似个例，包括其发生时间、与目标个例的相似度得分、关键要素场的对比图像等。系统按需异步加载并显示更多结果图像，确保用户获得完整信息的同时优化页面加载性能。

通过并发测试对平台及后台算法、接口性能进行了测试。设计 50 个并发用户，每秒启动 5 个并发用户执行检索操作，并根据设置的时间间隔发起重复请求，持续运行 5 分钟观察平台的压力表现，结果显示在 50 个并发压力情况下，平台及接口响应时间保持稳定，平均响应时间在 0.011 s 左右，最大约 0.015~0.020s；测试期间成功率 100%（共发送检索命令 15734 次），平均吞吐率 11480.33 字节/s，平均点击率 45.74 次/s，未见错误与异常波动。说明算法复杂度与系统架构在目标并发与数据规模下具有充足的性能余量与可扩展性。

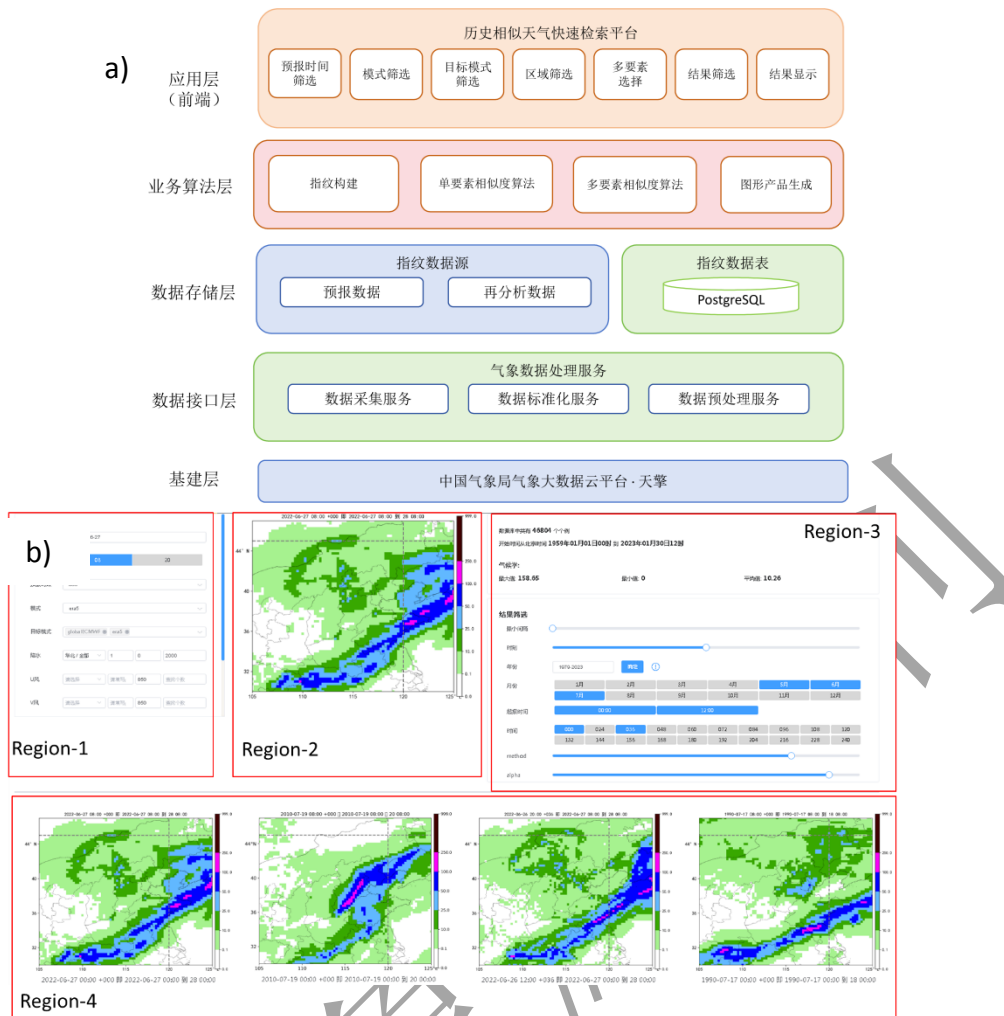


图 3 历史相似天气检索平台的 (a) 功能结构图和 (b) 4 个区域界面展示
 检索参数输入区 目标个例实时显示区 检索结果筛选与信息展示区 相似个例结果展示区

Fig. 3 (a) Platform function structure diagram and (b) display of four regional web interfaces of the historical analog weather retrieval platform

4. 应用及效果

选取 2023 年“23·7”华北特大暴雨实际预报个例进行验证，并开展了批量统计检验，此外对平台的业务试用反馈进行概述。

4.1 2023 年“23·7”华北特大暴雨个例

此次过程受到台风“杜苏芮”残余环流北上的影响（杨舒楠等，2023；张芳华等，2023）。选取 2023 年 7 月 29 日 20 时起报的 ECMWF-IFS 模式 36 h 预报时效的 24 h 累计降水（即 7 月 31 日 08 时至 8 月 1 日 08 时）预报场和中间时刻 850 hPa 风场、700 hPa 风场以及 500 hPa 位势高度四个要素。根据天气学经验，对于此类台风残余低压北上引发的持续性暴雨，中低层环流引导和高层形势配置均十分重要。基于此设置相应权重，24h 累计降水量、850hPa U 风、

700hPa 相对湿度、500hPa 位势高度的权重分别为 0.3, 0.2, 0.1, 0.4。在多要素协同配置下, 平台返回的最相似历史个例为 1996 年 8 月 4 日 08 时至 5 日 08 时的天气过程(图 4)。

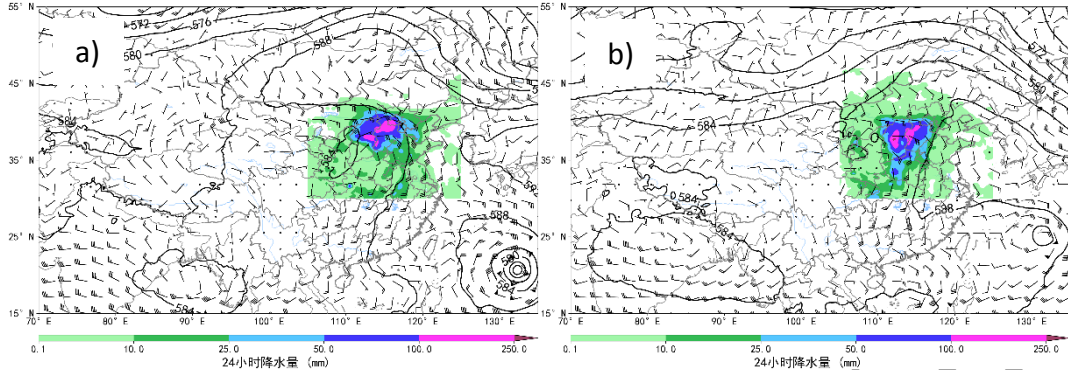


图 4 “23·7” 华北特大暴雨个例检验: 24 h 累计降水量(填色)及相应的 500hPa 位势高度(等值线, 单位: dagpm)和 850hPa 风场(风羽)

(a) 目标输入: 2023 年 7 月 30 日 08 时至 31 日 08 时, (b) 多要素检索: 1996 年 8 月 4 日 08 时至 5 日 08 时

Fig. 4 (a) The 24 h accumulated precipitation (colored), 500 hPa geopotential height (contour, unit: dagpm) and 850 hPa wind field (barb) from the case study of the July 2023 extreme rainstorm in North China

(a) target input: from 08:00 BT 30 to 08:00 BT 31 July 2023, (b) multi-element search: from 08:00 BT 4 to 08:00 BT 5 August 1996

从事后检验和天气学分析角度来看, 多要素检索所返回的个例(图 4b)与目标场的综合相似性较好。从降水形态与落区上看, 1996 年个例的降水主雨带也位于华北地区, 强度和范围与“23·7”过程的预报场具有较好的一致性。从影响系统与环流配置来看, 在 1996 年个例的同期天气形势中, 同样存在台风(9608 号)残余的深厚低值系统缓慢北上, 其东侧有副热带高压坝稳定维持, 引导气流将水汽持续输送至华北地区, 西风带系统相对偏北。这种环流配置与“23·7”过程中“杜苏芮”残余低压北上、副高异常强盛西伸, 呈现“方头”状、东侧热带系统水汽接力输送等关键特征高度吻合。这种特定的环流形势导致了河南、河北等地的持续性强降水过程。

该个例检验表明, 相似天气检索技术可以有效定位具有相似天气学意义和动力成因的复杂天气过程, 能够帮助预报员从更深层次理解当前预报形势, 并借鉴历史经验。

4.2 统计检验

为定量评估该相似检索技术的有效性, 本文对一段时间内的检索结果进行了统计检验。选取 2023 年 1 月 1 日至 12 月 31 日作为批量试验期, 每日以 ECMWF-IFS 业务预报的 36h 预报时效的 24h 累计降水场作为目标, 分别进行单要素(仅 24h 累计降水量)检索和的多要素(24h 累计降水量、500hPa 位势高度、850hPa 风场)检索, 共计 365 个样本。24h 累计降水场表征降水落区和强度, 是相似个例

检索的核心对象；850hPa 风场能够反映低层水汽输送和辐合条件；500hPa 位势高度场则代表中层环流背景、槽脊形势及副热带高压等大尺度系统特征。

检验时，以常用的距平相关系数(ACC)和均方根误差 (RMSE)，以及第 2 节相似度为主要评估指标；以 24h 累计降水量及相应中间时次的 500hPa 位势高度为评估对象。在 ERA5 再分析数据中检索相似历史个例，从单要素检索、多要素检索结果中分别选取相似度排名前 10%的个例，计算这些相似个例的平均要素场与目标预报场之间的 ACC 和 RMSE，同时，计算目标预报场与 1.3 节所述同期气候态场之间的 ACC 和 RMSE，作为参照基准，计算相似检索的结果较之气候态的改善率 (ACC 提升率/RMSE 降低率)。结果如表 1 所示。

表 1 2023 年批量试验期相似检索结果与气候态对比
Table 1 Comparison between similarity retrieval results of batch test period in 2023 and climatology

检索方法		评估指标				相似度	
		统计量	ACC (24h 累计降水量)	RMSE/mm (24h 累计降水量)	ACC (500hPa 位势高度)		RMSE/ dagpm (500hPa 位势高度)
单要素检索	目标与前 10%相似个例		0.454	0.01	0.956	43.40	96%
	目标与气候态		0.242	1.74	0.92	82.78	
	相似检索较气候态的改善率		87.6%	99.4%	3.91%	47.6%	
多要素检索	目标与前 10%相似个例		0.337	0.01	0.91	37.35	94%
	目标与气候态		0.201	2.69	0.870	194.33	
	相似检索较气候态的改善率		67.6%	99.6%	4.6%	80.8%	

从表 1 可以看出，单要素检索和多要素检索得到的前 10%相似个例的相似度均值分别达到了 96%和 94%以上。较高的相似度得分，说明算法能够稳定地找到高度匹配的个例。两种检索方式得到的相似个例与目标预报场的 ACC 和 RMSE，都显著优于气候平均场的表现。其中，单要素检索较之气候态，24h 累计降水量的 ACC 提升了 87.6%，RMSE 降低了 99.4%；多要素检索也表现出显著改进。这表明本文的相似检索方法能够有效找到比平均气候状况更接近当前预报形势的历史天气过程。

单要素检索与多要素检索对比，在 24h 累计降水量的 ACC 方面，单要素检索略高于多要素检索，这可能是因为单要素检索更专注于优化降水本身的相似性。然而，在 500hPa 位势高度场的 RMSE 方面，多要素检索比单要素检查更优（较气候态的结果降低逾 80%），这符合多要素协同设计的目标，即更好地匹配整体环流形势。需要注意的是，此处的统计结果可能受到所选协同要素、权重分配以及评估指标的综合影响。

4.3 业务效果

基于本文的相似检索技术开发的历史相似天气检索平台，自 2022 年 7 月起在中央气象台投入业务试用，现已稳定业务运行，并已推广到天津、宁夏等多个省级气象部门。预报员可以通过该平台进行交互式的历史相似个例检索，快速获取当前预报形势与历史上发生的相似天气过程之间的对比信息。从前期业务应用效果来看，该平台已展现了良好的辅助预报效果。

例如，2022 年主汛期多次强降水预报过程中，5 月 9—14 日，西南地区东部、江南、华南受东移高原槽和西南涡共同影响出现的强降水天气过程（韩旭卿和张涛，2022），6 月 26—30 日，云南、四川盆地经江汉、黄淮至东北地区南部受东移高空槽与副高北上影响出现的大范围强降水天气过程（孟庆涛和于超，2022）中，平台均通过提供历史相似暴雨个例，有效支持了预报员研判强降水过程落区、强度以及持续时间。又如，2023 年 1 月 12 日北京初雪过程（荆浩等，2023），高空受 500 hPa 短波槽、850 hPa 高度上反环流的共同影响，对流层中低层偏南风输送暖湿空气到北京。在该次降雪过程预报前，平台检索到的相似历史降雪个例为预报员提供了关于降雪起始时间、相态转换、主要降雪时段和量级等方面的有益参考。再如，在 2023 年 9 月 23 日杭州亚运会开幕式预报（孙杭媛等，2023）服务中，通过对 1959—2022 年 9 月 23 日前后 15d 共 1984 个样本进行相似检索和筛选，确定了以 588 dagpm 线和低层偏东风为重要检索因子，返回的 18 个较相似结果中只有 3 个出现了弱降水，为开幕式精准预报和风险提醒预报提供了历史经验支撑。

上述应用实例显示了历史相似天气检索平台在日常预报业务，特别是在复杂和极端天气事件预报中的实用价值和积极作用。平台可以支撑预报员更高效地从海量历史数据中挖掘有价值的信息。

5. 结论与讨论

本文针对天气预报业务中对历史相似个例检索的迫切需求,以及传统人工查找和简单算法匹配的不足,开展了历史相似天气快速检索技术的研究与平台应用工作。主要结论如下:

(1) 本研究将图像处理领域的指纹相似性检测算法应用于海量历史再分析资料和实时数值预报要素场的相似性度量,引入的单要素相似天气检索的基础算法能够将高维天气场压缩为低维指纹,实现相似天气的有效检索。构建了基于降水、环流等多个关键要素场的相似度进行加权集成的多要素检索,能够更全面、更准确地定位与目标天气过程相似和降水场相似的历史个例。个例分析和统计检验表明,多要素检索在捕捉天气系统整体配置和动力成因相似性方面优于单要素检索。

(2) 基于“云+端”架构,开发了历史相似天气检索平台,平台具备友好的用户交互界面、灵活的检索参数设置、高效的后台处理能力以及清晰的结果可视化功能。已在中央气象台投入业务试用,在多次业务预报服务中取得了良好的应用效果,获得了预报员的积极反馈。

(3) 本研究发展的历史相似天气快速检索技术及平台,能够帮助预报员从海量历史数据中高效、相对准确地检索出与当前预报形势相似的天气过程。不仅有助于预报员建立合理的“预报预期”,加深对特定天气类型的理解,丰富预报经验,从而提升预报能力,还在极端天气事件的预报和应对中具有重要的科学参考价值和业务应用潜力。

尽管历史相似天气检索平台在前期业务试用中取得了一定成效,但仍需在以下几个方面进一步完善。首先,当前相似天气检索算法对区域大小和位置较为敏感,虽已按气候区划建立分区模型,但增加了系统复杂性,未来需研究更具自适应性的区域处理方法和更加鲁棒的相似性度量算法。其次,检索结果的相似程度评判仍主要依赖预报员主观经验,现有 ACC 和 RMSE 等统计指标不足以完全反映关键天气特征的相似性,需构建更全面细致的客观评估指标体系。第三,在查询算法优化方面,可探索结合深度学习方法提升对天气特征捕捉能力,并针对不同天气过程类型研究动态权重分配策略,引入专家知识或机器学习方法进行权重自适应优化。第四,需结合可解释性人工智能技术增强算法透明度,融入气象动力

学约束以确保检索结果的物理合理性。最后，应深化系统应用价值，探索与数值预报和 AI 气象模型的结合、气候变化背景下的应用适配，以及构建天气过程知识图谱、历史灾害性天气个例库等，为预报业务和气象知识传承提供更有力的支持。

致谢：感谢国家气象中心林建在本文撰写中提供的指导。

参考文献

- 陈兴芳, 赵振国. 2000. 中国汛期降水预测研究及应用 [M]. 北京: 气象出版社, 153-157. Chen X F, Zhao Z G. 2000. A Study of Precipitation Prediction in Flood Season Over China and Application [M]. Beijing: China Meteorological Press, 153-157 (in Chinese).
- 丁一汇. 2015: 论河南“75.8”特大暴雨的研究:回顾与评述[J]. 气象学报, (3): 411-424. Ding Y H. 2015: On the study of the unprecedented heavy rainfall in Henan Province during 4-8 August 1975: review and assessment[J]. Acta Meteor Sin. 73(3): 411-424 (in Chinese).
- 董芹, 霍焱, 董喜春, 等. 常州市灾害性天气查询系统[J]. 气象科技, 2009, 37(4): 457-461. Dong Q, Huo Y, Dong X C, et al, 2009. Inquiry System of Disastrous Weather in Changzhou. [J] Meteor Sci Technol, 37(4): 457-461 (in Chinese).
- 韩旭卿, 张涛, 2022. 2022年5月大气环流和天气分析[J]. 气象, 48(8): 1070-1076. HAN X, ZHANG T, 2022. Analysis of the May 2022 Atmospheric Circulation and Weather [J]. Meteor Mon, 48(8): 1070-1076 (in Chinese).
- 郝翠, 张迎新, 徐路扬, 等. 2022. 改进最优集合预报方法对北京极端气温和风速预报的检验[J]. 暴雨灾害, 41(4): 467-476. DOI: 10.3969/j.issn.1004-9045.2022.04.011 HAO C, ZHANG Y X, XU L Y, et al., 2022. Verification of improved analog ensemble methods for forecasting extreme temperature and wind speed in Beijing [J]. Torr Rain Dis, 41(4): 467-476. DOI: 10.3969/j.issn.1004-9045.2022.04.011 (in Chinese).
- 黄成南, 炎利军, 莫满堂, 等. 广东灾害天气历史个例库系统的设计与实现[J]. 广东气象, 2013, 35(3): 68-73. Huang C N, Yan L J, Mo M T, et al, 2013. Design and implementation of disaster weather case database system in Guangdong Province [J]. Guangdong Qixiang, 35(3): 68-73 (in Chinese).
- 黄晓远, 李旭, 杜梦莹, 等. 2025. 融合相似预报方法在陇东南短期强降水预报中的应用[J]. 高原气象, 44(1): 214-223. Huang X Y, Li X, Du M Y, et al, 2025. Application of fusing analog method in short-term heavy precipitation forecast in southeastern Gansu [J]. Plateau Meteor, 44(1): 214-223 (in Chinese).
- 焦圣明, 郑媛媛, 王宏斌, 等. 2017. 灾害性天气个例库智能分析系统的设计与实现[J]. 气象, 43(3): 354-364. Jiao S M, Zheng Y Y, Wang H B, et al. 2017. Design and Implementation of an Intelligent Analysis System for Disaster Weather Cases [J]. Meteor Mon, 43(3): 354-364 (in Chinese).
- 荆浩, 赵桂洁, 于波, 等. 2024. 北京一次罕见1月初雪过程的复杂降水相态成因分析. 气象, 50(8): 905-916. DOI: 10.7519/j.issn.1000-0526.2024.040101. Jing H, Zhao G J, Yu B, et al, 2024. Analysis of Complex Precipitation Types During a Rare First Snow Process in January in Beijing. Meteorological Monthly, 50(8): 905-916. DOI: 10.7519/j.issn.1000-0526.2024.040101 (in Chinese).
- 李开乐, 1997. 多层次格点资料相似离度动态综合做广东多站点长期趋势预报[J]. 热带气象学报, 13(2): 180-185. Li K L, 1997. Long-term multi-station tendency forecast for the province of Guangdong using dynamic combination of multilevel grid data in terms of similar discrete values [J]. J Trop Meteor, 13(2): 180-185 (in Chinese).
- 李颖, 金茹, 许雯, 等. 2023. 基于路径与环境场最优赋权的致灾相似台风检索方法研究[J]. 水利水电技术, 54(7): 14-26. Li Y, Jin R, Xu L, et al, 2023. Study on the optimal weighting impact of typhoon track and environmental field in similar typhoon retrieval model for similar disasters [J]. Water Resour Hydropower Eng, 54(7): 14-26 (in Chinese).
- 鲁小琴, 余晖, 赵兵科, 2013. 热带气旋环境场相似检索方法的对比分析[J]. 气象, 39(12): 1609-1615. Lu X Q, Yu H, Zhao B K, 2013. Study on similarity retrieval method for ambient field of tropical cyclones [J]. Meteor Mon, 39(12): 1609-1615 (in Chinese).

孟庆涛,于超,2022.2022年6月大气环流和天气分析[J].气象,48(9):1209-1216. MENG Qingtao,YU Chao,2022.Analysis of the June 2022 Atmospheric Circulation and Weather[J].Meteor Mon,48(9):1209-1216 (in Chinese).

钱磊,邱学兴,郑淋淋,等,2022.基于降水空间分布相似的最优集成降水预报及其检验[J].暴雨灾害,41(3):324-335. Qian L, Qiu X X, Zheng L L, et al, 2022. Prediction and test of optimal integrated precipitation based on similar spatial distribution of precipitation[J]. Torr Rain Dis, 41(3): 324-335 (in Chinese).

乔丹杨,卢晔,贺瑶,2023.相似预报法在关中西部强对流预报中的应用和检验[J].陕西气象,(2):8-13. Qiao D Y, Lu Y, He Y, 2023. Application and test of analog forecast to forecasting severe convection in the western Guanzhong[J]. J Shaanxi Meteor, (2): 8-13 (in Chinese).

任福民,贾莉,吴彩铭,等,2023.相似预报原理之再认识:动力统计相似集合预报理论及其应用[J].气象学报,81(2):193-204. Ren F M, Jia L, Wu C M, et al, 2023. Advances in dynamic-statistical analog ensemble forecasting and its application to precipitation prediction of landfalling typhoons: a renewed understanding[J]. Acta Meteor Sin, 81(2): 193-204 (in Chinese).

孙杭媛,陈淑琴,马依依,等,2023.杭州亚运会期间降水天气及其环流形势分析[C]//浙江省气象局,浙江省气象学会.精彩亚运数智气象——亚洲气象保障服务技术交流会议论文集编.,120-130 (in Chinese).

孙卓,高嵩,李江波,等,2020.基于MICAPS4的灾害性天气个例检索与训练系统设计及应用[J].暴雨灾害,39(2):207-212. DOI: 10.3969/j.issn.1004-9045.2020.02.012 Sun Z, Gao S, Li J B. 2020. Design and application of disastrous weather case retrieval and training system based on MICAPS4[J]. Torrential Rain and Disasters, 39(2): 207-212. DOI: 10.3969/j.issn.1004-9045.2020.02.012 (in Chinese).

闫旭,段勇,才奎志,等,2022.基于局部敏感哈希的天气形势场图像相似预报[J].图像与信号处理,2022,11(4):183-190. <https://doi.org/10.12677/JISP.2022.114018>. Yan X, Duan Y, Cai K Z, et al. 2022. Similarity Prediction of Weather Situation Field Image Based on Locality Sensitive Hashing, 11(4): 183-190 (in Chinese).

杨舒楠,张芳华,胡艺,等,2023.“23·7”华北特大暴雨过程的基本特征与成因初探[J].暴雨灾害,42(5):508-520. Yang S N, Zhang F H, Hu Y, et al., 2023. Analysis on the characteristics and causes of the "23·7" torrential rainfall event in North China[J]. Torr Rain Dis, 42(5): 508-520(in Chinese).

张芳华,杨舒楠,胡艺,等,2023.“23·7”华北特大暴雨过程的水汽特征[J].气象,49(12):1421-1434. Zhang F H, Yang S N, Hu Y, et al,2023.Water Vapor Characteristics of the July 2023 Severe Torrential Rain in North China[J].Meteor Mon,49(12):1421-1434(in Chinese).

张静,何俊仕,2012.模糊可变集合方法在降雨相似性分析中的应用[J].水文,32(4):44-48. Zhang J, He J S, 2012. Application of variable fuzzy sets in rainstorm similarity analysis[J]. J China Hydrol, 32(4): 44-48 (in Chinese).

赵亮,魏铁鑫,王丽荣,等,2024.基于多维特征的区域性暴雨过程相似性检索方法研究[J].热带气象学报,40(6):966-973. Zhao L, Wei T X, Wang L R, et al, 2024. Similarity retrieval method for regional rainstorm processes based on multidimensional features[J]. J Trop Meteor, 40(6): 966-973 (in Chinese).

周必高,鲁小琴,郑峰,等,2022.基于改进的VGG16模型的副热带高压相似识别及应用评估[J].气象,48(12):1608-1616. Zhou B G, Lu X Q, Zheng F, et al, 2022. Research on the similarity recognition and application evaluation of subtropical high based on modified VGG16 model[J]. Meteor Mon, 48(12): 1608-1616 (in Chinese).

Chattopadhyay A, Nabizadeh E, Hassanzadeh P, 2020. Analog forecasting of extreme-causing weather patterns using deep learning[J]. J Adv Model Earth Syst, 12(2): e2019MS001958.

Doswell III C A, 1987. The distinction between large-scale and mesoscale contribution to severe convection: a case study example[J]. Wea Forecasting, 2(1): 3-16.

Franch G, Jurman G, Coviello L, et al, 2019. MASS-UMAP: fast and accurate analog ensemble search in weather radar archive[J]. Remote Sens, 11(24): 2922.

Hersbach H, Bell B, Berrisford P, et al, 2020. The ERA5 global reanalysis[J]. Quart J Roy Meteor Soc, 146(730): 1999-2049.

Hu W M, Cervone G, Young G, et al, 2023. Machine learning weather analogs for near-surface variables[J]. Bound-Lay Meteor, 186(3): 711-735.

Huffman G J, Bolvin D T, Braithwaite D, et al. Integrated Multi-satellite Retrievals for the Global Precipitation Measurement (GPM)

Mission (IMERG). *Advances in Global Change Research*. Cham: Springer, 2020(67): 343-353.

Kalnay E, Kanamitsu M, Kistler R, et al, 1996. The NCEP/NCAR 40-year reanalysis project[J]. *Bull Amer Meteor Soc*, 77(3): 437-472.

Kratzert F, Klotz D, Hochreiter S, et al, 2021. A note on leveraging synergy in multiple meteorological data sets with deep learning for rainfall–runoff modeling[J]. *Hydrol Earth Syst Sci*, 25(5): 2685-2703.

Monache L D, Eckel F A, Rife D L, et al, 2013. Probabilistic weather prediction with an analog ensemble[J]. *Mon Wea Rev*, 141(10): 3498-3516.

Raoult B, Di Fatta G, Pappenberger F, et al, 2018. Fast retrieval of weather analogues in a multi-petabytes archive using wavelet-based fingerprints[C]//18th International Conference on Computational Science–ICCS 2018. Wuxi, China: Springer: 697-710.

