

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. 气象, 39(12):1609-1615.

热带气旋环境场相似检索方法的对比分析^{*}

鲁小琴 余 晖 赵兵科

中国气象局上海台风研究所,中国气象局台风预报技术重点开放实验室,上海 200030

提 要: 比较了欧氏距离、相似离度、相似系数、相关系数及最优相似系数等作为气象格点场相似性度量的适用性,结果发现以“距离”为主的算法无法准确地反映环境场中的一些特定系统的位置和强度,相似检索效果较差;而相似离度、相似系数、相关系数及最优相似系数可分别考虑两个场之间的形状和强度、空间分布及变化趋势,能找到较相似的场,可供业务参考。通过将不同指标检索到的相似环境场的演变对热带气旋(Tropical Cyclone, TC)路径的影响量引入到 TC 路径预报中,检验其业务应用效果,发现单纯移向的预报效果好于路径预报,路径预报中相似离度的效果稍好于其他方法,而移向预报中最优系数法最好。另通过相似环境场检索,能找到由相似环境场所对应的相似历史 TC,并依据不同度量性指标得到一个历史相似 TC 集,为台风业务、科研提供参考。

关键词: 环境场相似, 相似度量指标, 相似热带气旋

中图分类号: P458

文献标志码: A

doi: 10. 7519/j. issn. 1000-0526. 2013. 12. 009

Study on Similarity Retrieval Method for Ambient Field of Tropical Cyclones

LU Xiaoqin YU Hui ZHAO Bingke

Key Laboratory of Typhoon Forecast Technique, Shanghai Typhoon Institute, CMA, Shanghai 200030

Abstract: In this paper the suitability of Euclidean distance, similar disparity, similarity coefficient, correlation coefficient and best similarity coefficient used as a similarity measure in meteorological grid dataset is discussed. The results show that some special systems' position and intensity cannot be expressed correctly by distance algorithm. But using similar disparity, similarity coefficient, correlation coefficient and best similarity coefficient, two meteorological grid datasets' spatial distribution and the change trend are considered separately, so the similar ambient field can be found in the long history dataset, which can be applied in operation. In order to evaluate the similarity measures' operational application, the similar ambient field corresponding to tropical cyclone (TC) track's motion quantity is referenced in current source TC track forecast. The results show that 24 h moving direction forecast accuracy is better than that of 24 h track forecast and the similar disparity is better than other algorithm in track forecast, but the best similarity coefficient is better than other algorithm in moving direction forecast. By the query of similar ambient field, a similar TC ensemble corresponding to similar ambient field is found, which can be referenced in TC operational, forecast or research.

Key words: similar ambient field, similarity measure, similar tropical cyclone

^{*} 国家自然科学基金项目(40921160381)、公益性行业(气象)科研专项(GYHY201006008)、国家重点基础研究发展计划项目(2009CB421501)和地理信息科学教育部重点实验室开放研究基金资助项目(KLGIS2012A062)共同资助

2012 年 9 月 10 日收稿; 2012 年 12 月 13 日收修定稿

第一作者:鲁小琴,主要从事热带气旋资料信息化、遥感和 GIS 方面的研究工作. Email:luxq@mail. typhoon. gov. cn

引言

历史相似热带气旋(Tropical Cyclone, TC)检索可辅助进行路径、强度和风雨预报,为预报员所常用。2004 年由中国气象局下发全国受 TC 影响省份气象台站使用的《西北太平洋热带气旋检索系统》就具有对历史 TC 进行相似路径检索的功能,即查询与当前 TC 路径相似的历史 TC,为预报决策提供参考。在该系统中,王远飞等(2006)利用 GIS 的空间查询技术实现了 TC 路径相似检索,提供了一种基于地理空间相似的历史 TC 查询。但是,根据地理空间相似检索到的 TC 其未来发展趋势不一定和当前 TC 一致。主要原因在于,地理空间相似是定义在路径某一段中的一定宽度内通过的路径都为相似路径,但是前期路径相似的两个 TC,由于季节不同、环境场形势不同其未来发展趋势有可能完全不同。因此预报员更为关心的是相似 TC 的环境场是否相似。

要判断两个环境场是否相似,首先需要确定一个相似性度量,进而做比较。相似性度量(研究样品或变量之间的相似程度的数量指标)主要包含两种(John et al, 2009):一种是距离,它是将每一个样品看作 p 维空间的一个点,并用某种度量测量点与点之间的距离,距离较近的归为一类,距离较远的点应属于不同的类;另一种是相似系数,性质越接近的变量或样品,它们的相似系数的绝对值越接近于 1;而彼此无关的变量或样品它们的相似系数越接近于 0。由此产生的具体算法包含欧氏距离、马氏距离等距离类函数及相似系数、相关系数等系数类函数。

基于以上两种相似性度量,许多专家做了很多有关气象要素场的相似研究。李开乐(1986)最早提出引入一种衡量相似程度的统计量——相似离度,意思即相似性的差异程度。相比较距离算法仅考虑样本间值的差异,相似离度既考虑到样本之间的形相似情况,又体现了它们的值相似差异,是一个比较全面的相似标准,实际使用效果良好。罗阳(1996)提出了一个可反映波动三要素的新的相似性度量——最优相似系数,通过比较样本间平均值差值与波动振幅的相对大小,将平均值差异对样本相似性的影响引入到相似量中,克服了以往相似性度

量在考虑平均值影响时,需要知道样本极值的问题。罗阳等(2008)又在分别对不同类的场相似方法分析总结后,通过实际相似样本选择的对比分析和预报试验,证明最优相似系数的选取结果与人为的主观判定具有较高的一致性,预报能力也较强,优于其他相似量。钟元等(1991;2003)利用距离和相似系数构建综合相似判据,在 TC 相似预报中表现出了一定的预报技巧,但没有对多个度量性指标进行比较。郭达烽等(2005)在天气形势相似中分别进行相似距离和相似系数计算,并按最小相似距离和最大相似系数对历史样本排序,最后综合考虑两种排序结果,进一步优选出前 15 个与预报天气形势和天气系统最相似的历史样本。单九生等(2001)在做天气相似比较时主要使用了综合相似系数,即增加了一个“值”因子的相似系数来进行判别。万日金等(2006)比较了海明距离法、欧氏距离法、相似系数法、相似离度法等各方法的有效性,最后得到的海明距离法的效果最好,略高于相似离度法,但其主要分析是针对暴雨预报进行的。

因此到目前为止,在专门针对 TC 的业务预报中没有一个通用的、效果较好的环境场及 TC 相似判别方法。本文拟对 TC 环境场进行多方法的相似检索和综合分析,以期能得到一个有效、稳定的相似检索算法或方案,供业务参考;同时通过对 TC 相似环境场的检索分析能得到更多为 TC 科研和业务参考的信息。

1 资料介绍

环境场数据来自美国国家气象环境预报中心和国家大气研究中心(NCEP/NCAR)的 500 hPa 高度场再分析数据集,时间分辨率为 6 h,空间分辨率为 $2.5^{\circ} \times 2.5^{\circ}$;由于 TC 一般在海洋上活动,因此主要研究对 TC 走向起决定性作用的副高或西风槽等主要引导系统的活动区域,故设定参与计算的空间范围为 $60^{\circ}\text{N} \sim 10^{\circ}\text{S}$ 、 $100^{\circ}\text{E} \sim 160^{\circ}\text{W}$ 。试验时段为 2009—2010 年,历史库的时间范围为 1980—2010 年,逐日 00、06、12 和 18(UTC)共 4 个时次。热带气旋最佳路径资料使用中国气象局《热带气旋年鉴》资料,含每日 02、08、14 和 20(北京时)共 4 个时次的 TC 中心经度、纬度;其中 TC 编号为国内编号(YNN),而 TD(无编号)取当年 TC 的序号记录

(YYYYNN)。

2 相似试验

以 2009—2010 年有 TC 活动时的 500 hPa 高度场为目标环境场(以此环境场为目标,在历史库中查询与其相似的环境场),设定检索的历史库的时间范围为 1980—2010 年有 TC 活动日,分别以欧氏距离、相似离度、相似系数、相关系数和最优相似系数为度量性指标,检索逐时次的相似环境场。为了提高检索效率,首先利用当前时刻 TC 中心位置与历史 TC 中心位置间距离 $\leq 5^\circ$ 、当前时刻的月份与历史时刻的月份差 ≤ 2 个月进行筛选,对满足条件的样本进行相似度量计算,并确定利用该指标得到的相似场。

2.1 距离类指标的试验

循环计算目标环境场和历史库中场之间的欧氏

距离,距离最小者为与目标场最相似的场。结果发现,从人工主观判断来看,距离最小的场与目标场并不满足气象意义上的相似,即对 TC 移动起引导作用的副高、西风槽等系统的位置、强度差异较大。

欧氏距离公式:

$$d(\overline{x_i}, \overline{x_j}) = \left[\sum_{k=1}^p (x_{ik} - x_{jk})^2 \right]^{\frac{1}{2}} \quad (1)$$

式(1)表明欧氏距离反映的只是两个样本的空间距离而反映不出形状的差异,而形状差异往往是对 TC 有影响的系统的真实分布,同时也是预报员更为关心的(李开乐,1986;许映龙等,2011;涂小萍等,2010;钱燕珍等,2013;郭荣芬等,2013)。因此,通过简单求取距离大小来判断环境场的相似,不能达到目标(图 1)。

图 1 可见,就副高(588 等值线区域,下同)的位置和形状而言,目标场(图 1a)中副高分成两部分,最东端到达 165°E 附近;而相似场(图 1b)中副高为一整体,偏西,最东端至 140°E 附近,与目标场不相似。

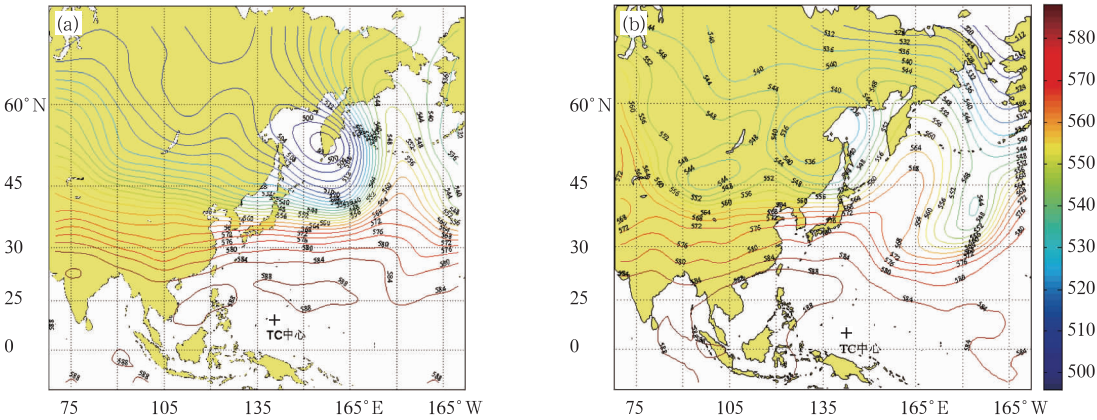


图 1 以欧氏距离作为度量性指标检索 1001 号 TC 在 2010 年 3 月 22 日 18 时的相似环境场
(a)目标场,(b)相似环境场(8602 号 TC 的 4 月 24 日 18 时)
(“+”为 TC 中心)

Fig. 1 Similarity ambient field of 1001 TC at 18:00 BT 22 March 2010 retrieved by Euclidean distance
(a) the objective ambient field, (b) the similarity ambient field (8602 TC at 18:00 BT 24 April 1986)
(“+” is the TC center position)

2.2 系数类指标的试验

分别利用相似离度、相似系数、相关系数及最优相似系数作为度量性指标,确定其对应的相似场。从多个个例的分析结果来看,各指标得到的相似场与目标环境场的整体形势配置比较相似,即关键系统的空间分布较一致;但是系统的强度有差别,且相似离度、相似系数检索的结果偏弱,而相关系数的结

果偏强。
由系数类函数中相似离度、相似系数、相关系数及最优相似系数的计算公式(罗阳等,2008)可知,相似离度、最优相似系数可以反映样本的“形”与“值”,故能较好地将对应场中系统的分布及强弱进行对比,进而找到比较相似的场;而相似系数和相关系数的检索以“形”似为主,效果与相似离度相当。以 2010 年 7 月 21 日 18 时为例,相似离度、相似系数、

相关系数及最优相似系数的相似环境场检索结果如图 2 所示。

由图 2 可以看出,目标环境场(图 2a)中副高断开成东西两部分,其东西轴线在 38°N 左右,西边部分的东边界在 147°E 附近,而东边部分的西边界在 167°E 附近;东部强,强度为 596 dagpm。检索结果中相似离度、相似系数及最优系数的结果相同,皆为 9606 号 TC 的 7 月 20 日 06 时,该时刻副高也断开成东西两部分,其东西两部分的轴线略倾斜,呈西南—东北分布,分别在 23°N 和 38°N 左右,西边部

分的东边界在 $147^{\circ}\sim 148^{\circ}\text{E}$ 附近,而东边部分的西边界在 $167^{\circ}\sim 168^{\circ}\text{E}$ 附近;东部最强,强度为 592 dagpm。总体而言相似场与目标环境场中的整体形势的配置非常相似,只是对副高强度的描述稍偏弱,及断开后西边部分的位置在南北分布上稍有差异。而相关系数法的检索结果为 2000 年 8 号热带低压的 7 月 16 日 12 时,该时刻虽然副高也断开成东西几部分,但其断开的位置及形状与目标环境场不同,强度也有差异,因此就这个例而言,效果不如相似离度、相似系数及最优系数。

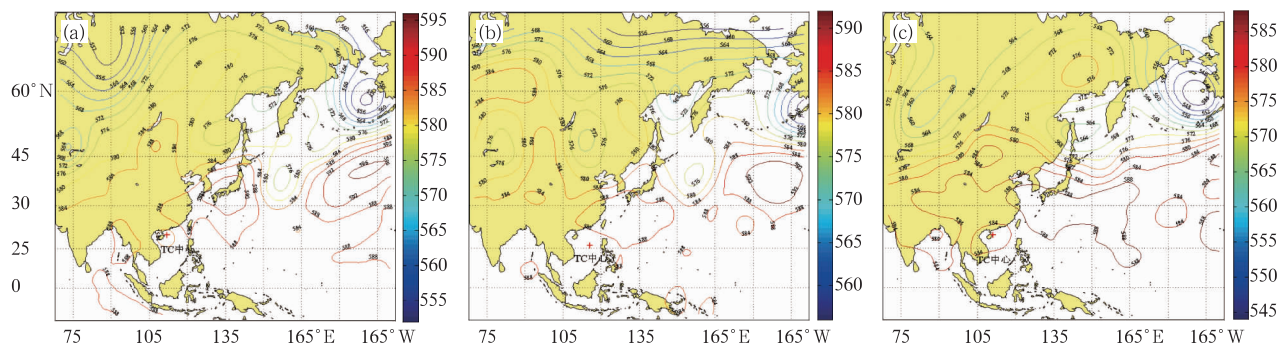


图 2 利用相似离度、相似系数、相关系数及最优相似系数作为度量性指标检索

2010 年 1003 号 TC 在 7 月 21 日 18 时的相似场结果

(a) 目标场, (b) 相似离度、相似系数、最优系数对 1996 年 9606 TC 在 7 月 20 日 06 时的相似检索结果,

(c) 相关系数对 2008 TD 在 2000 年 7 月 16 日 12 时的反演结果

(红色“+”为 TC 中心)

Fig. 2 Similarity ambient field of 1003 TC at 18:00 BT 21 July 2010 retrieved by similar disparity, similarity coefficient, correlation coefficient and the best similarity coefficient

(a) objective ambient field, (b) retrieved result of similar disparity, similarity coefficient, and the best similarity coefficient (9606 TC at 06:00 BT 20 July 1996), (c) retrieved result of correlation

coefficient (2008 TD at 12:00 BT 16 July 2000)

(The red “+” is the TC center position)

3 相似检索效果检验

在第 2 节中利用不同度量性指标检索到的相似场之间的对比判别是人工主观判断的,即主观比较主要影响系统的形状、位置、强度是否一致。但是在实际应用中,相似检索的目的是为了辅助业务预报。因此,可以依据检索到的相似环境场所对应的 TC 路径发展变化量来对目标环境场所对应的当前 TC 进行路径预报,以检验不同指标的实际业务应用效果。具体做法如下:

(1) 首先确定某一度量性指标(欧氏距离、相似离度和相似系数等),然后据此方法在历史库中对目

标环境场进行相似检索;

(2) 以检索到的相似环境场对应时刻的历史 TC 中心位置和其下一时次 TC 中心位置的差作为当前目标环境场所对应 TC 的未来移动量,进行路径预报;

(3) 依据最佳路径资料计算目标 TC 路径预报的误差。

基于 2009—2010 年试验样本数据,分别利用 5 类指标的相似检索结果进行 TC 路径预报。24 h 路径预报误差分布如图 3 所示。

图 3 可见,相似离度、相似系数、相关系数、最优系数及欧氏距离 5 类指标中,24 h 路径预报误差的差异不太大,四分位值分别为 201、210、216、212 和

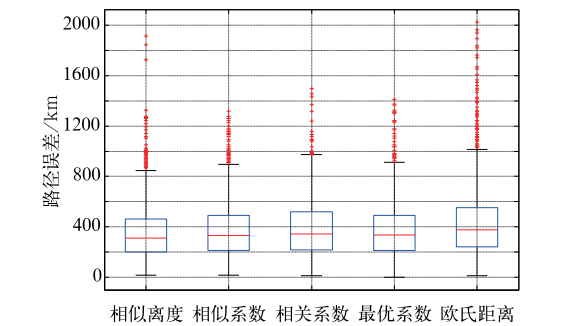


图 3 24 h 路径预报误差(样本 1104)

(蓝色框的下边界线为 4 分位线,上边界为 4 分 3 位线,蓝色框中的红色线为中位线,蓝框上方和下方的黑色短横线表示样本中去掉异常值后的极大值和极小值,最上方的红色“+”代表异常值)

Fig. 3 24 h forecast error of TC track (1104 samples)

(The lowest edge of the blue box is the quadrant line and the highest edge of the blue box is the third quadrant line; the red line in the blue box is the middle line; the black short lines over the blue box and under the blue box are the maximum sample value and minimum sample value separately, except the outliers; the red “+” samples are outliers)

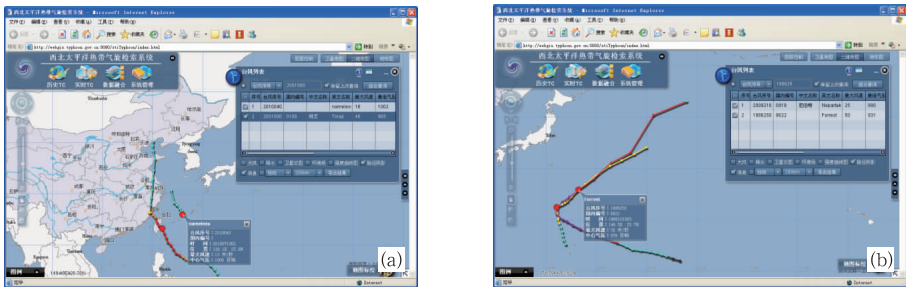


图 4 相似离度检索结果应用示例

(a) 201004 号 TD(2010 年 7 月 19 日 02 时)和 0108 号 TC(2010 年 7 月 29 日 20 时),

(b) 0919 号 TC(2009 年 10 月 11 日 08 时)和 8622 号 TC(1986 年 10 月 19 日 20 时)

Fig. 4 Application of retrieved results by similar disparity

对 4 号低压进行 24 h 路径预报,误差为 15 km,效果较好;而在图 4b 中,2009 年 0919 号 TC 在 10 月 11 日 08 时的相似环境场为 8622 号 TC 的 10 月 19 日 20 时,依据相同规则对 0919 做路径预报,误差达 1914 km。分析图 4b 中误差较大的原因,发现 0919 号 TC 在 10 月 11 日 08 时与 8622 号 TC 的 10 月 19 日 20 时后的 24 h 路径移动趋向是一致的,向东北移动;但是由于进入西风带,8622 号移动速度大大加快(6 h 移速最大达 $150\text{ km}\cdot\text{h}^{-1}$),远大于 0919 号(6 h 移速最大为 $15\text{ km}\cdot\text{h}^{-1}$),因此导致 24 h 路径预报误差大。据此可推断,基于相似环境

239 km,即 25% 的样本的误差在大约 200 km 以下;中位值分别为 312、331、345、336 和 376 km,也即 50% 的样本误差在大约 300 km 以下。据此分布可看出 24 h 路径预报结果中相似离度的效果较好,欧氏距离最差,相似系数、相关系数及最优相似系数的效果相当。由于 TC 移动路径受多个因素影响,除大尺度引导流外,还与地形、TC 本身的结构、移动速度等因素有关,因此本试验中只简单参考相似环境场的影响而进行 TC 路径预报的误差较大。但该方法可作为相似检索效果检验的一个客观判别方法。

综上,无论是人为主观判断,还是从业务应用的效果来分析,都表明相似离度是进行环境场检索的一个较好的度量性指标。但是从试验的 24 h 预报误差(以相似离度为例)来看,有的时次误差很小(图 4a),而有的时次误差很大(图 4b)。

图 4a 中,2010 年第 4 号热带低压 7 月 19 号 02 时的相似环境场为 2001 年 0108 号 TC 的 7 月 29 号 20 时,根据相似环境场对应 TC 的 24 h 移动量

场下对应 TC 的未来发展趋势较相似,那么对 TC 移向的预报精度应该高于路径预报。因此分析了 2009—2010 年所有样本的 24 h 移向预报误差(图 5)。

由图 5 可见相似离度、相似系数、相关系数、最优系数及欧氏距离 5 类指标中,四分位值分别为 12、12、14、11 及 17° ,即 25% 样本的预报误差在 $11\sim14^\circ$ 之间(欧氏距离除外),与 2011 年北京综合预报 24 h 移向的平均误差 13° 接近(陈国民等,2012),证明利用相似检索进行 TC 移向预报比路径预报精度稍高(路径误差的四分位值为 200 km,远差于业

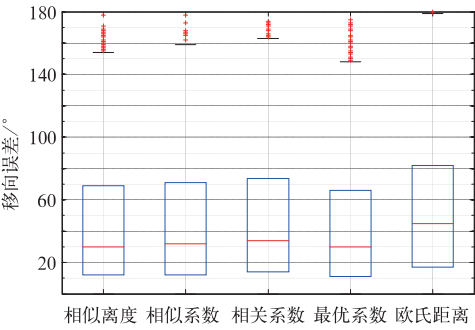


图 5 利用 5 类指标进行相似检索预报的 24 h 移向误差绝对值(样本数 1104)
(说明同图 3)

Fig. 5 Absolute forecast error of 24 h move direction by 5 types of similarity retrieved result forecast (1104 samples)
(The figure explanation is the same as Fig. 3)

务预报误差的 110 km), 可供 TC 业务预报参考。同时 5 类指标中, 最优系数效果稍好, 其次为相似离度、相似系数、相关系数, 欧氏距离最差。

综上, 5 类指标的相似检索在预报应用中表现为: 单纯移向预报的效果好于路径预报; 路径预报中相似离度的效果稍好于其他方法, 而移向预报中最优系数法最好。

4 相似检索的应用

根据第 3 节相似场检索效果检验发现, 检索到的相似环境场所对应的 TC 与当前目标环境场对应 TC 的未来发展趋势很相似, 因此依据环境场相似检索能找到未来发展相似的 TC, 可作为一种相似 TC 的检索方法。如: 依据相似环境场检索, 可查到

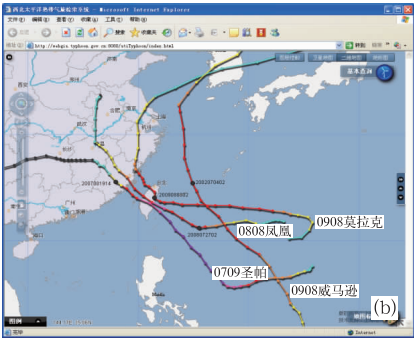
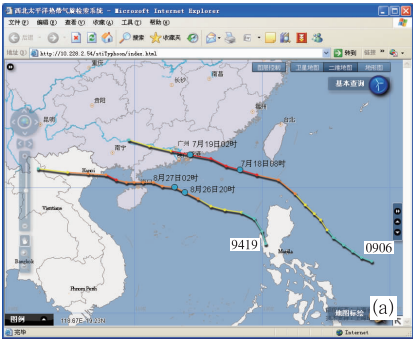


图 6 基于相似环境场检索到的相似路径示例图
(a) 0906 号 TC 的相似路径, (b) 0908 号 TC 相似路径
Fig. 6 Similar TC track based on the retrieved similar ambient field

0906 号 TC 在 7 月 18 号 08 时和 19 日 02 时的相似路径分别为 9419 号 TC 的 8 月 26 日 20 时和 27 日 02 时(图 6a), 据图 6a 可见二个 TC 的路径趋势非常一致, 能为 TC 路径预报提供参考。但是, 如果基于地理空间的路径相似检索方法进行相似 TC 检索, 由于 9419 号 TC 距 0906 号 TC 较远(平行距离 400 km 以上), 则不一定会被检索出来。由此可见, 基于相似环境场检索历史相似 TC, 能找到发展趋势相似的 TC, 对预报很重要。

同时, 利用不同的度量性指标所检索到的相似环境场对应的 TC 集, 对当前 TC 的未来移向有一定的指示意义。如 0908 号 TC 在 2009 年 8 月 8 日 2 时的相似环境场所对应的相似 TC(图 6b)。其中相似离度和相似系数的结果都为 0808 号 TC(环境

场相似时刻皆为 2008 年 7 月 27 日 02 时), 相关系数和最优相似系数的结果为 0709 号 TC(环境场相似时刻为 2007 年 8 月 19 日 08 时和 14 时), 欧氏距离的结果为 0205 号 TC(环境场相似时刻为 2002 年 7 月 4 日 02 时)。由图 6b 可见, 从相似时刻起, 0908 与 0709、0808 在 24 h 内有相同的移动趋势, 保持西北行; 而与 0205(北上)稍有差异。

5 结论与讨论

5.1 结论

通过上述分析, 可得到如下结论:

(1) 比较了欧氏距离、相似离度、相似系数、相

关系数及最优相似系数作为气象格点场的相似性度量的性能,发现以“距离”为主的算法无法准确反映一些系统的位置和强度,相似检索效果不好;而另 4 类算法可分别考虑两个场之间的形状和强度、空间分布或变化趋势,因此能找到较相似的场,可供业务参考。

(2) 通过将不同指标检索到的相似环境场的演变对 TC 路径的影响量引入到 TC 的路径预报中,然后分析其预报误差,进而检验不同指标的相似检索结果在业务上的实际应用效果,发现移向预测效果好于路径;路径预报中相似离度的效果稍好于其他方法,而移向预报中最优系数法最好。

(3) 基于相似环境场检索,能找到可参考的相似环境场及由相似环境场所对应的相似 TC,且此相似 TC 与目标场对应 TC 具有类似的发展趋势;根据不同指标可检索到一个相似历史 TC 集合,为预报服务。

5.2 讨论及下一步工作

在大量的分析样本中发现,排除移速、移向的影响后,不论哪一种方法都存在少量较大误差的样本。经人工检验,发现这些样本的相似场确实与目标场差异较大,表明在当前历史库中确实没找到可参考的相似场。那么如何判断当前查询到的相似环境场的可参考性是非常重要的。分析了每种方法的相似度量指标值与预报误差的关系,得到单纯从相似度量指标值很难判断其参考性的结论,因此下一步工作将就如何找到一个相似度量指标临界值展开;同时,本文只分析了各指标的 24 h 预报应用效果,实际上如何基于相似性指标进行 72 h 及更长时效的预报辅助是非常重要的。另外每种度量性指标都有

其优劣,如相似离度的应用在路径预报上效果最好,而最优系数则在移向的预报上效果最佳。因此在实际应用中要通过长期的业务预报效果总结规律,使其更好地为业务服务。

参考文献

- 陈国民,汤杰,曾智华. 2012. 2011 年西北太平洋热带气旋预报精度评定. 气象, 38(10): 1238-1246.
- 郭达烽,许爱华,肖安. 2005. 多级相似作温度精细化预报初探. 江西气象科技, 28(3): 22-26.
- 郭荣芬,肖子牛,鲁亚斌. 2013. 登陆热带气旋引发云南强降水的环境场特征. 气象, 39(4): 418-426.
- 李开乐. 1986. 相似离度及其使用技术. 气象学报, 44(2): 174-183.
- 罗阳. 1996. 一种新的相似性度量-高分辨相似系数. 空军气象学院学报, 17(1): 23-32.
- 罗阳,赵伟. 2008. 相似性度量研究及最优相似系数. 中国气象学会 2008 年年会天气预报准确率与公共气象服务分会论文集, 373-392.
- 钱燕珍,张程明,孙军波,等. 2013. 用数值预报释用方法做近海及登陆热带气旋强度预报. 气象, 39(6): 710-718.
- 单九生,张延亭. 2001. 江西省流域降水客观预报方法简介. 江西气象科技, 24(4): 9-13.
- 涂小萍,许映龙. 2010. 基于 ECMWF 海平面气压场的热带气旋路径预报效果检验. 气象, 36(3): 107-111.
- 万日金,何溪澄,林刚. 2006. 用动力相似方法预报广东省区域暴雨预报试验. 热带气象学报, 22(2): 198-202.
- 王远飞,陆涛,朱海燕,等. 2006. 基于 GIS 的热带气旋相似路径检索系统研究. 测绘科学, 31(5): 124-128.
- 许映龙,韩桂荣,麻素红,等. 2011. 1109 号超强台风“梅花”预报误差分析及思考. 气象, 37(10): 1196-1205.
- 钟元,胡波. 2003. 综合评估环境场影响的热带气旋路径客观相似预报模式. 热带气象学报, 19(2): 147-156.
- 钟元,祝荣霖. 1991. 中期(旬)天气客观相似预报方案. 气象学报, 49(4): 553-558.
- John A Richards, Jia Xiuping. 2009. 遥感数字图像分析(第四版). 张晔,张钧萍,谷延锋,等译. 北京:电子工业出版社, 191-193.