

一种台风路径相似检索的算法研究

刘 勇 吴必文 王东勇

(安徽省气象台, 合肥 230031)

提 要: 台风路径可视为二维平面上的一段曲线, 根据两条台风路径曲线的相似离度可以判断其数值相似和形态相似的程度。由于直接应用台风基本资料进行相似离度计算有一定的难度和局限性, 需要研究利用相似离度原理判断两条台风路径曲线相似程度的计算方法。按照设计的算法处理台风基本资料, 可以确定两条台风路径曲线上的对应控制点, 得到两条曲线对应控制点之间的距离和距离偏差的方向, 最终只需在一个方向计算相似离度就能直接分析两条曲线的相似性。对相似台风进行检索的实例表明, 该算法是可行的, 能够从台风基本资料库中检索出与当前台风路径最相似的台风个例。

关键词: 台风路径 曲线 相似离度 算法

Algorithm Study on Similarity Retrieval of Typhoon Tracks

Liu Yong Wu Biwen Wang Dongyong

(Anhui Meteorological Observatory, Hefei 230031)

Abstract: A typhoon track can be considered as a curve on a two-dimension plane. According to the similarity deviation of two typhoon track curves, their numerical similarity and geometrical similarity can be estimated. There are still some difficulties and limitation when the original typhoon data are applied to the formula of similarity deviation. Based on the principle of similitude deviation, a brand new algorithm is designed to calculate the similar degree between two typhoon tracks. According to this algorithm, the control points on the two typhoon track curves are picked up, the distance between corresponding control point on the

资助项目: 安徽省人才开发基金项目(2002Z037) 资助

收稿日期: 2006年1月25日; 修定稿日期: 2006年5月25日

different curves and the direction of the distance deviation are calculated. As a result, the similar degree of two typhoon track curves can be easily estimated but need to calculate the similar deviation in one direction. Compared with the source typhoon track, the most similar destination typhoon track can be retrieved from the historical typhoon database by means of this new algorithm.

Key Words: typhoon track curve similarity deviation algorithm

引言

台风的移动路径,是各种影响台风移动的物理因子综合作用的结果。相似的台风路径,在一定程度上反映了影响台风移动的诸因子综合作用等效^[1]。当台风可能对某地区产生影响时,当地气象部门的业务人员为了做好决策气象服务工作,需要收集实时台风的路径预测信息,更需要分析与实时台风路径相似的历史台风个例情况,了解台风历史个例所导致的雨情和灾情等信息,以便做好气象服务工作。但是决策建议是否合理,气象服务是否成功,则与台风路径相似检索的准确程度密切相关。

关于台风路径的研究多集中于分析和预测台风路径,虽然有很多台风业务服务系统可以将单个或多个台风的路径显示出来^[2,3],但台风路径相似检索工作则不多见。所谓台风路径相似检索,就是针对某个台风,从历史资料库中查找出与其移动路径最相近的台风个例。魏应植曾将经、纬度坐标转换成计算机屏幕坐标,根据台风是否经过指定半径的搜索圆,确定台风是否经过某区域,寻找路径相似的台风^[4]。耿慧将热带气旋的活动分成三个区,根据当前位置、和过去12小时的移动方向,先找出位置相似的多个台风个例,再对不同区域的天气形势因子进行判断,找出路径较为相似的个例^[5]。本文介绍的是一种基于相似高度原

理,根据台风路径特点设计的判断两条台风路径曲线相似程度的计算方法。利用该算法,可以很方便地从台风基本资料库中检索出与当前台风路径最相似的台风个例。文中还对如何进一步提高台风路径相似检索的准确性进行了探讨。

1 基本资料库

台风基本资料包括台风的编号、名称、位置、强度等诸多信息。虽然台风路径相似检索重点考虑台风移动的路径,但为了确定多个相似个例中的最佳相似个例,仍然需要参考台风的移动速度、中心气压和近中心最大风速。为了检索路径相似的台风,基本资料库既要涵盖台风活动的相关数据,也要便于及时追加最新台风活动资料。因此,台风基本资料库建设采用了以下形式的数据结构:

[台风编号]

最大强度,中文名称,英文名称

时次1,经度1,纬度1,中心气压1,
近中心最大风速1

时次2,经度2,纬度2,中心气压2,
近中心最大风速2

……

时次N,经度N,纬度N,中心气压
N,近中心最大风速N

在上述结构中,台风如果没有中、英文名称可以空缺。

2 基本原理

台风路径相似检索其实是二维平面上曲线的相似检索。曲线相似检索既要考虑曲线的位置相近(数值相似),更要关注曲线之间的吻合程度(形态相似)。判定曲线相似检索的方法有很多种^[6,7],但常用的是相似高度法^[8],其基本原理如下:

假定正方形网格中有两条曲线*i*和*j*不完全重合,即存在数值差异和形态差异,两者的位置和形态如图1所示。

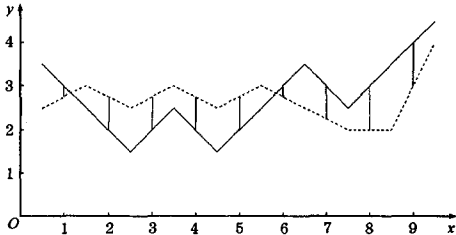


图1 台风路径曲线位置、形态示意

实线: 曲线*i*; 虚线: 曲线*j*

曲线*i*和曲线*j*在纵向的数值差异可以用下式计算:

$$D_{ij} = \frac{1}{M} \sum_{k=1}^M |y_{ik} - y_{jk}| \quad (1)$$

其中*M*表示两条曲线与纵线共同相交的次数。 y_{ik} 、 y_{jk} 分别为曲线和第*k*条纵线交点的高度值。 D_{ij} 是两条曲线在各纵线上绝对高度差值的总体平均,值越小,表明两条曲线在纵向数值上越接近。

由于曲线*i*和曲线*j*在纵向上的形态差异可以用下式计算

$$S_{ij} = \frac{1}{M} \sum_{k=1}^M |(y_{ik} - y_{jk}) - A_{ij}| \quad (2)$$

式(2)中

$$A_{ij} = \frac{1}{M} \sum_{k=1}^M (y_{ik} - y_{jk})$$

其中 A_{ij} 是两条曲线在各纵线上高度差

的数学期望。 S_{ij} 反映曲线在纵线上的高度差值与高度平均差值的离散程度,值越小,表明两条曲线在纵向形态上越相似。

相似高度定义为

$$C_{ij} = D_{ij} + S_{ij} \quad (3)$$

C_{ij} 由值相似系数 D_{ij} 和形相似系数 S_{ij} 共同决定,值越小,两条曲线的相似程度越高。

根据上述公式,可以计算两条曲线在横向的相似高度。

3 计算方法

从理论上讲,利用台风基本资料库中的数据,按照公式(1)~(3),就可以分别计算出当前台风路径曲线和被检索台风路径曲线在纵、横两个方向的相似高度,依据两条曲线的相似高度值,就可以判断两条曲线的相似程度。但事实上则会遇到两个方面的问题。一是计算公式要求两条曲线在纵线上或者横线上共同相交的次数相同,由于台风路径的多样性,这个要求实际上是难以满足的;二是在纵、横两个方向分别进行相似高度的计算将得到2个值相似系数、2个形相似系数,综合分析时有时会难以取舍。为了解决这两个问题,必须设计应用相似高度公式的计算方法。

根据相似高度的基本意义,两条台风路径曲线相似与否,在笛卡尔坐标系中必须从纵、横两个方向分别进行分析。但是从自然坐标系的角度来看^[9],如果能够按照一定的规则确定当前台风和被检索台风路径曲线上的特殊位置点(称为控制点),在获取了对应控制点的位置坐标后,只需在一个方向计算相似高度,就可以判断两条曲线的相似程度。

基于相似高度原理,检索路径相似台风的详细算法如下。

3.1 确定台风路径曲线上的控制点

台风路径曲线是由台风定时观测得到的位置依次连线构成的，由于台风的移动速度不均匀，在台风观测中得到台风位置并不是等间距的。按照图 2 所示的控制点选取方法就可以确定台风路径曲线上的控制点。

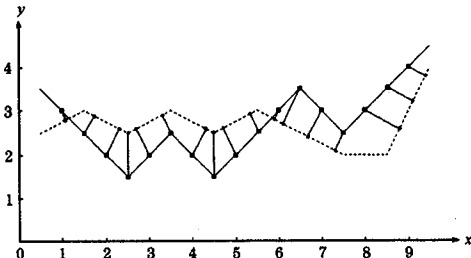


图 2 台风路径曲线控制点选取方法示意
实线：曲线 *i*；虚线：曲线 *j*

对于当前台风 *i* 可以通过插值的方法将台风路径划分成若干个等间隔的线段，再将线段两端所在的位置作为路径曲线上的控制点。对被检索的台风 *j*，可以在其路径上分别求出与当前台风的每个控制点距离最近的点，并将这些点作为路径曲线上的控制点。

在图 2 中，在曲线 *i* 上各控制点之间的距离都是相等的。曲线 *j* 上的控制点，实际上就是过曲线 *i* 的控制点向曲线 *j* 作垂线时的垂足点，该点和曲线 *i* 上对应控制点的距离最近。

从数值相似的角度来看，根据两条曲线各对应控制点之间的距离，按照公式 (1)，就能得到两条曲线数值相似的计算结果。

3.2 确定对应控制点之间距离偏差的方向

假定当前台风 *i* 路径曲线上某控制点 P_0 的坐标为 (x_0, y_0) ，相邻下一个控制点 P_1 的坐标为 (x_1, y_1) ，被检索台风 *j* 的路径曲线上与 P_0 相对应控制点 P_2 的坐标为 (x_2, y_2) 。

根据 P_0 和 P_2 两点的坐标，可以求出在控制点 P_0 上两条曲线之间的最短距离。为了确定这个最短距离的偏差方向，可以约定由 P_0 指向 P_1 的方向为正方向，如果 P_2 落在 P_0 和 P_1 两点所构成的矢量的右侧或正前方，则规定距离偏差的符号为负，反之规定距离偏差的符号为正。实际计算时，只需在坐标变换的基础上进行简单的判断即可。图 3 为坐标变换过程示意图。

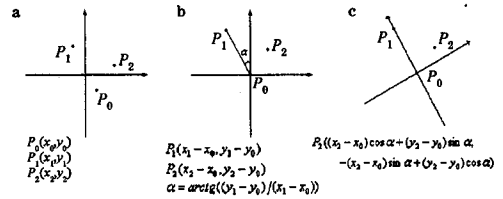


图 3 坐标变换过程示意

图 3a 为 P_0 、 P_1 、 P_2 三个点的原始位置和坐标位置。将原坐标系的原点平移到 P_0 点后，可求出新坐标系 Y 轴和 P_0 、 P_1 所构成的矢量夹角 α ，图 3b 为经过坐标平移后 P_1 、 P_2 的坐标位置以及夹角 α 的值。最后按逆时针方向旋转坐标系，使得 Y 轴方向和上述矢量方向重合，图 3c 为经过旋转的坐标系中 P_2 的新坐标位置。

根据约定，如果 P_2 在新坐标系中 $x > 0$ 或者 $x = 0, y > 0$ ，则这一对控制点上距离偏差的符号为负值，反之为正值。

从形态相似的角度来看，如果知道两条曲线各对应控制点之间的距离和距离偏差的方向（符号），根据公式 (2)，就可以进行两条曲线的形态相似计算。

3.3 计算两条曲线之间的相似离度

当前台风 *i* 路径曲线上的每一个控制点，被检索台风 *j* 的路径曲线上都有相对应的、距离最短的控制点。由于两条曲线对应控制点的坐标位置是确定的，所以可以确定

两条曲线各相应控制点之间的距离和距离偏差的符号。

分析公式 (1) 和 (2) 的右端可以发现, 决定相似离度值的仅仅是两条曲线在纵向的差值。两条曲线的控制点只要保持纵向差值不变, 无论是在纵向平移 (例如一端移至水平坐标轴), 还是在横向平移 (例如各控制点不等距) 都不会影响计算结果, 因此可以将当前台风 i 的路径曲线看作是水平坐标轴, 轴上有多个等距的控制点, 被检索曲线 j 上的控制点都位于经过水平坐标轴控制点并且与水平坐标轴垂直的直线上。这时曲线 j 上对应控制点的位置恰好就是两条曲线对应控制点之间的距离, 如果距离偏差的符号为正, 曲线在坐标轴的上方, 反之则在坐标轴的下方。

图 4 为根据图 1 中的原始数据经过上述处理后得到的两条台风路径曲线示意图。至此, 在确定了两条曲线各对应控制点之间的距离和距离偏差的符号后, 只需在一个方向应用相似离度的计算公式, 就能直接利用计算结果判断两条曲线的相似程度了。

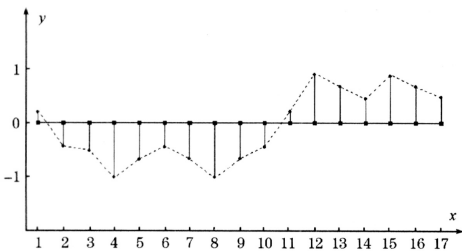


图 4 经过处理后的台风路径曲线示意
实线: 曲线 i ; 虚线: 曲线 j

4 台风路径相似检索方法的实现与应用

当前台风和被检索台风的路径资料都按上述计算方法处理之后, 在当前台风路径曲

线的每个控制点上, 所有被检索台风的路径曲线上都有了相对应的、距离最近的控制点, 根据这些控制点的坐标, 可以求出当前台风和所有被检索台风之间的相似离度值。在设定了需要选取的路径相似的台风数量后, 按相似离度值从小到大进行选取, 就可以得到路径相似的台风个例集合。

4.1 实例

笔者按照研究出的算法设计了“台风相似路径检索业务系统”, 台风基本资料由中国气象局上海台风研究所提供, 检索范围覆盖影响我国的西太平洋台风活动区域。该系统目前已在安徽省气象台投入业务运行。业务人员利用此系统, 可以很方便地从台风基本资料库中检索出与当前台风路径较为相似的若干台风个例, 并能逐个地进行显示和比较。为预报台风路径提供帮助。

图 5 为 2005 年 13 号台风“泰利”路径相似检索的结果。根据相似离度值的大小, 与 0513 号台风路径最相似的个例为 7613 台风, 第二相似、第三相似、第四相似的个例分别为 6911 台风、9215 台风和 7511 台风。从图中可以直观地看出, 虽然这些台风的起始位置各不相同, 但是在稍后阶段与当前台风的路径都是比较接近的, 从路径上看具有很高的相似性。系统还可以提供如表 1 所示

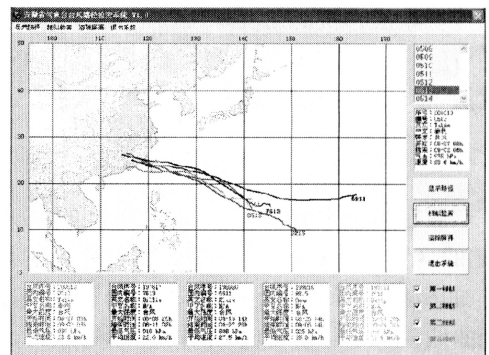


图 5 “泰利”台风路径相似检索结果

的当前台风和路径相似台风个例的有关基本数据, 这些信息可帮助预报员逐个分析相似

个例及其风雨特征。为进一步做好当地的台风预报和服务提供必要的信息基础。

表1 “泰利”台风及路径相似个例基本信息表

	当前台风	第一相似个例	第二相似个例	第三相似个例	第四相似个例
台风编号	0513	7613	6911	9215	7511
台风名称	Talim	Billie	Elsie	Omar	Betty
最大强度	超强台风	超强台风	超强台风	超强台风	台风
最低气压	935hPa	910hPa	888hPa	925hPa	948hPa
平均移速	23.6km·h ⁻¹	22.0km·h ⁻¹	27.5km·h ⁻¹	15.8km·h ⁻¹	21.4km·h ⁻¹
起始时间	08-27-08	08-03-20h	09-19-14	08-25-14	09-17-20
结束时间	09-02-08	08-11-02	09-27-20	09-05-14	09-24-02

4.2 应用

利用相似离度原理检索相似台风路径的算法, 同样适用于其它天气系统移动路径的检索。例如, 只要建立了风暴、沙尘暴、江淮气旋等天气系统中心移动路径的历史数据库, 根据实时路径的资料, 利用上述算法, 就能检索出系统移动的相似路径。

天气预报业务实践中, 预报员经验积累, 主要基于对典型天气形势和天气实况之间相互关系的综合分析, 预报员分析决策, 很大程度上依赖于对历史相似资料及天气状况的搜索^[10]。虽然当今数值预报的精度已经超越了预报员主观预报的水平, 但预报员的经验, 包括数值分析预报产品的应用经验, 在很长的时间内还是十分有用的^[11]。路径相似的天气系统对某地天气的影响有很多相似之处, 如果能将指定天气系统移动路径相似的个例提供给预报员, 既有助于其分析这些个例相对应的天气形势和风雨特征, 也可以对数值预报结果的可靠性进行检验。

5 结语与讨论

基于相似离度原理设计的检索台风相似路径的算法包括二个关键环节: 在当前台风

路径上和被检索的台风路径上确定对应的控制点; 分析两条曲线对应控制点之间距离和距离偏差的方向。经过这二个步骤处理之后, 只需在一个方向上计算相似离度, 就能直接利用计算结果判断两条曲线的相似程度。从实际应用的效果来看, 该算法是可行的, 不仅可以应用于台风路径的相似检索, 同样适合于其它天气系统移动路径的相似检索。

在实践中发现, 能否从台风基本资料库中检索出与当前台风路径相似程度很高的台风个例集合, 不仅取决于相似检索的算法, 与台风基本资料库中的个例数量、台风路径曲线的长度也有很大的关系。当台风路径曲线的长度较为适中或相对较短时, 由于控制点比较少, 很容易从基本资料库中找到多条相似的台风路径曲线, 检索出来的台风路径的相似程度就比较高。当曲线的长度相对较长时, 由于这类台风本身就比较少, 加上控制点多, 即使算法再好, 也很难找到路径比较相似的台风个例。应用此方法时, 可以将当前台风路径中对那些影响不明显段落(如最初阶段)截去, 仅仅对最接近当前时刻、影响最大的段落进行检索, 就可以明显提高路径检索的相似程度。

从气象服务的需求来看, 相似的台风不

仅仅是路径相似, 最大强度、中心气压和台风发生的时间等都应当相近, 因此预报员在确定路径最相似的台风集合时, 还需要参考与台风密切相关的其它要素。

致谢: 国家气象中心刘还珠女士对本研究工作进行了指导, 并对文章结构和方法应用提出了建设性的修改意见。中国气象科学研究院高志球先生、国家气候中心任福民先生也曾给予帮助。在此一并致谢!

参考文献

- 1 陈联寿, 丁一汇. 西太平洋台风概论 [M]. 北京: 科学出版社, 1979: 388.
- 2 林伟, 李开奇, 王志武等. 沿海台风决策服务系统研制 [J]. 气象科技, 2004, 32 (2): 129-131.
- 3 王东法, 徐霜芝, 何军等. 浙江省台风预警服务系统简介 [J]. 浙江气象, 1997, 18 (4): 29-32.
- 4 魏应植. 用 VB 作 50 年台风路径资料检索与分析 [J]. 成都气象学院学报, 2000, 15 (4): 317-321.
- 5 耿慧. 热带气旋过程 (动态) 相似路径选取 [J]. 气象, 1996, 22 (2): 48-49.
- 6 韩忠南, 俞善贤, 王耀生. 台风浪场及天气形势图的人工智能检索研究 [J]. 海洋预报, 1994, 11 (4): 1-5.
- 7 Liao T Arren, Zhang Zhiming, Mount Clauder R. Similarity measures for retrieval in case-based reasoning system [J]. Applied Artificial Intelligence. 1998, 6 (12): 267-288.
- 8 李开乐. 相似高度及其使用技术 [J]. 气象学报. 1986, 44 (2): 174-183.
- 9 朱乾根, 林锦瑞, 寿绍文等. 天气学原理与方法 [M]. 北京: 气象出版社, 1998, 38-39.
- 10 王萍, 孔秀梅, 杨洪敏等. 天气图相似检索研究 [J]. 天津大学学报, 2004, 37 (3): 264-268.
- 11 中国气象局科教司. 省地气象台短期预报岗位培训教材 [M]. 北京: 气象出版社, 1998: 1-13.