

一种基于 MOST 的台风时空轨迹模型

地球与空间科学学院 2013 级 刘证源 程楚云 曹文溥

摘 要: 传统台风轨迹记录方式存在数据量大以及无法查询台风任意时刻位置等问题。本文根据台风轨迹位置和速度变化连续、规律性较强的特点, 基于 MOST 模型为设计了一个台风时空轨迹模型, 在保证数据精度的基础上减少台风轨迹数据量, 提供台风全时域位置查询功能。建模过程中, 设计了台风时空轨迹建模效果评价方法, 并对更新策略的阈值选取和阈值优化问题进行了研究。实例研究表明, 相关模型具有良好的可行性和适用性。

关键词: 移动对象时空模型; 轨迹建模; 更新策略; 模型评价; 台风

A Spatio-Temporal Typhoon Trajectory Model Based on MOST

Liu Zhengyuan, Cheng Chuyun, Cao wenpu,
School of Earth and Space Sciences, Peking University

Abstract: Traditional methods of typhoon trajectory recording output a large amount of data and cannot support location inquiry of typhoon at a specific time. Considering the movement of a typhoon is basically continuous and predictable, this paper promotes a Temporal-Spatial tractor model of typhoon based on MOST (Moving Objects Spatial Temporal), which reduces data volume of typhoons' trajectory records while keeping reasonable output precision and provides full-time domain location inquiry of typhoon location. A case study is carried out to verify the model, in which a method of evaluating typhoon trajectory modeling results is designed and, moreover, thresholds selection and optimization are discussed.

Key words: Moving Objects Spatial Temporal Model; trajectory modeling; update strategy; model evaluation; typhoon

台风是一种破坏力很强的灾害性天气系统, 经常造成巨大的生命和财产损失, 我国是台风灾害比较严重的国家之一(中国气象局, 2010; 国家气候中心, 1998)。建立台风轨迹监测记录, 为台风预报提供技术支持, 具有非常重要的研究和实用价值。随着天气雷达监测、地面自动气象站监测等技术的不断发展, 尤其是卫星遥感监测技术的发展和运用, 目前已经可以大量获取台风监测数据。目前普遍采用的按照一定时间或者一定空间间隔记录台风中心位置的点位记录方式, 存在数据量大、各记录点位数据无法高效拟合计算等问题, 对相关的台风分析和管理工作支撑明显不足。因此, 探索新的台风动态轨迹模型, 减少台风数据的存储量, 提供全时域位置查询服务, 提高数据分析能力和台风预测能力具有非常重要的意义。

移动对象建模与管理是 GIS 典型的时空数据模型问题, 国内外学者在这方面开展了大量研究。Wolfson(1997)最早提出移动对象时空(Moving Objects Spatial Temporal, MOST)模型用来描述当前和预计未来的移动情况, 通过引入动态属性概念, 将移动对象的位置表示为时间的函数, 移动对象正常运动时, 无须对其位

置信息反复更新。Guting 等(1998)基于抽象数据类型,在无限集上定义移动对象的概念模型和时空操作,但抽象层中的无限集合无法在计算机无法表达。为了解决这一问题,Forlizzi(2000)、Jin(2007)等研究了移动对象离散数据模型,将复杂的空间对象及移动轨迹分割为相对简单的离散片段。MOST 模型中最基本的创新思想是动态属性。对于动态属性 A,通过 3 个子属性来形式化表示 A.value、A.updatetime 和 A.function: A.value:类型是 Type; A.updatetime:一个时间值; A.function:函数,即 $f:\text{int} \rightarrow \text{Type}$,满足 $t=0$ 时, $f(t)=0$,语义为“A 在时间 t 时的值”;动态属性函数便可写成:

$$\text{Value}(A, t) = A.\text{Value} + A.\text{fuction}(t - A.\text{updatetime})$$

以此为基础,研究者对移动对象更新提出了一些方法和策略,以平衡存储移动对象数据通信量和精确度。Wolfson 等(1999)首先提出 dead-reckoning 更新策略来减少更新代价,分为 Speed Dead Reckoning 策略(SDR)、Adaptive Dead Reckoning 策略(ADR)和 Disconnection Detection Reckoning 策略(DDR)。Gowrisankar 等于 2002 年提出使用角度和线性偏移的 dead-reckoning 策略。Wolfson 等又于 2003 年引入 Deviation(偏离)并且与 Distance(距离)策略进行比较,保证移动对象更新精度。

在台风监测与研究工作中,3S 技术已经得到了广泛应用。美国、欧洲和印度等国家的气象部门已经在气象服务中广泛应用 GIS 和 RS 技术(朱海燕,2005);在中国,CHANG Liang(2005)基于 GIS 进行了中国东南沿海的台风模拟与风险评估,上海台风研究所建立了基于 GIS 的西北太平洋台风路径检索和分析系统。但截止目前,基于 GIS 的移动对象建模技术目前尚未在台风监测与研究工作中得到应用。

本文基于 MOST 开展台风时刻轨迹建模研究。台风显然可以看作一种移动对象,其运动具有较强规律性,通过选取合适的运动函数和数据更新策略,应当可较为准确地拟合台风轨迹,降低数据通信量和存储量,并提供一定的位置查询支持。

1 基于 MOST 的台风时空轨迹模型

本文根据台风运动具有位置和速度变化连续、规律性较强的特点,基于 MOST 设计了一个台风时空轨迹模型。

1.1 台风轨迹表达

MOST 模型引入了动态属性概念,将移动对象的位置表示为时间的函数,移动对象正常运动时,无须对其位置信息反复更新,仅当移动对象的运动状态发生异常(如某个参数超过指定阈值)时再进行更新(Wolfson, 1999)。基于 MOST 模型这一思想,我们可以把台风的位置作为动态属性,将其表示为时间的函数,即:

$$X(t) = X(t_0) + f(v(t), t - t_0) \quad (1)$$

其中, t 为当前时刻, t_0 为参考时刻(初始时刻或上一次进行更新的时刻), $X(t)$ 是台风在 t 时刻的位置, $X(t_0)$ 是台风在 t_0 时刻的位置, $v(t)$ 是速度函数即台风在 t 时刻的速度。

显然,如果速度函数 $v(t)$ 如果能准确表达,则可以使用积分的方法精确地算出台风在 t 时刻的位置 $X(t)$ 。当然在实际工作中速度无法做到精确表达,所以使用(1)式计算得到台风位置是近似值,具有一定的误差。而在一定的误差范围

内，可以直接使用（1）式计算得到台风的位置，当误差超过指定阈值时再触发数据库更新。

式（1）表达仅是台风轨迹的一个片段的运动函数。为了方便在计算机中表达 MOST 模型的轨迹，可将台风的总轨迹分割为相对简单的离散片段，每个片段所表示的信息和方法如下式所示：

$$SEG(i) = (StaticAttr, DynamicAttr(i), Period(i), F(i)) \quad (2)$$

其中，*StaticAttr* 是台风的静态属性集，包括台风的 ID、名字 NAME、台风记录点总数 POINTSUM 等；*DynamicAttr(i)* 是台风在第 *i* 段轨迹中的动态属性集，包括台风的位置坐标 $X(t)$ ，台风的运动速率 $v(t)$ ，台风的运动方向 $\theta(t)$ 、台风轨迹的计算位置 X_{cmpt} 与实际位置 X_{true} 的差值 $E(t)$ 等； $Period(i) = [t_s, t_e]$ ，是台风第 *i* 段轨迹的时间跨度，是一个左闭右开的区间，其中 t_s 为轨迹起点的时刻， t_e 为轨迹终点的时刻； $F(i)$ 为台风的运动函数（即（1）式），使用该函数可以计算出台风在当前生命周期中任意时刻的位置，同时可以计算短期内台风未来的位置，实现台风轨迹的短期预测。

台风的所有离散轨迹段组合在一起，构成台风的轨迹。在台风正常运动或运动参数误差小于一定阈值时，台风一直在某一轨迹段运行，台风的运动状态由当前轨迹段的运动函数直接计算出；当台风的运动状态突然发生变化或误差累积到一定阈值时，触发更新，台风进入下一轨迹段。

1.2 台风运动函数的选择

由于台风运动具有位置和速度变化连续且变化一般不发生突变的特点，本文采用线性函数作为全时域飓风轨迹模型的运动函数，即：

$$X(t) = X(t_0) + \overrightarrow{v(t)}(t - t_0) \quad (3)$$

1.3 更新策略设计

如前文所述，在台风正常运动或运动参数误差小于一定阈值时，台风一直在当前轨迹段运行；但当台风的运动状态突然发生变化或误差累积到一定阈值时，需要触发更新，即进入新的轨迹段。如果更新频率过高，即误差阈值选的过小，会大大增加数据的存储量和通信量；如果更新频率过低，即误差阈值选的太大，会导致轨迹模型的拟合精度迅速下降，难以满足精度的要求。因此，如何制定合适的更新策略和选取合适的阈值会直接影响建模的效果。综合以上考虑，本文在台风轨迹模型中同时设置了距离、速率、角度三个阈值，并设计了如下更新策略：

(1) 距离阈值 Th_d

显然台风在模型中的计算位置 and 实际位置的偏差直接影响轨迹的精度，故设置台风轨迹的距离阈值。计算当前时刻台风轨迹的计算位置 X_{cmpt} 与实际位置 X_{true} 的距离差值 $dis(X_{cmpt}, X_{true})$ ，如果此距离差值大于距离阈值 Th_d 时触发更新。

(2) 速率阈值 Th_v

为了防止模型出现台风的运动速率偏离实际过大的情况，设置台风运动的速率阈值。计算当前时刻台风运动的计算速率 V_{cmpt} 与实际速率 V_{true} 的差值 $|V_{cmpt} - V_{true}|$ ，如果此速率差值大于速率阈值 Th_v 时触发更新。

(3) 角度阈值 Th_θ

为了防止模型中因台风运动方向偏离实际导致短时间内模型预测出现严重

偏差的情况，设置台风运动方向的角度阈值。计算当前时刻台风运动的计算方向 θ_{cmpt} 与实际方向 θ_{true} 的差值 $|\theta_{cmpt} - \theta_{true}|$ ，如果此角度差值大于角度阈值 Th_{θ} 时触发更新。

上述三个阈值可以单独或者组合应用，结合式（3）完成对台风时空轨迹的建模工作。

2 实例建模结果

为验证上述台风时空轨迹模型的科学性和可行性，本文基于真实的台风轨迹进行了建模工作，并从轨迹压缩率和轨迹拟合精度的两个方面对于建模效果进行了评价和分析。实验环境为 Windows 10 系统和 Visual Studio 2013 集成开发环境，采用 C# 语言编写建模程序，数据库采用的是 Microsoft Access 数据库。

2.1 数据与建模

建模实验数据来自中国台风网，为 1949 年至 2014 年间太平洋西部共计 2212 条台风记录。每条台风记录是一个按时间顺序记录的台风中心位置的坐标序列，记录时间间隔为 6 小时。每个记录点除了有当前点位的经纬度坐标和台风的 ID、名称等信息外，还有此点位的日期、台风强度、风速、气压等数据。

轨迹曲折度是描述台风轨迹复杂程度的指标，定义如下：

$$\text{曲折度} = \frac{\text{台风轨迹总长度}}{dis(\text{轨迹起点, 轨迹终点})} \quad (4)$$

对于具有不同轨迹曲折度的台风，其建模效果可能不同。因此在实验中我们先对所有台风的轨迹曲折度大小进行了排序，并选取了曲折度最大的 4 条和最小的 4 条台风，作为实验的典型台风。这 8 条台风的详细数据如下表所示。

表 1 典型台风轨迹信息

台风名称	曲折度	记录点数
Washi	1.015	37
Sarika	1.035	28
Jelawat	1.040	124
Durian	1.045	121
Huko	2.447	46
Fitow	2.699	190
Jangmi	3.551	136
Tokage	3.565	25

国内外台风监测资料显示（许映龙等，2010），台风的平均移动速率为 10-20km/h，台风较高的 24 小时预报精度为 50-100km。故建模实验中选取距离阈值为 50km，速率阈值为 10km/h，角度阈值为 5°。对于每个台风，以轨迹起点的数据作为初始运动状态，对于之后的每个轨迹点，使用时空模型计算得到运动状态数据，看是否满足更新策略，即距离差值、速率差值和角度差值是否超过三个阈值，如果不满足，则舍弃该点的记录；如果满足，则触发更新，保留该点的记录并更新台风运动状态数据，进入下一轨迹段。最终，在一条轨迹建模完成后，将

记录全部存入数据库，并计算得到轨迹压缩率和轨迹拟合误差。

2.2 结果与分析

2.2.1 模型评价标准定义

(1) 轨迹压缩率

轨迹压缩率用于描述轨迹建模前后数据量减少的程度。本文所采用的压缩率定义如下：

$$\text{压缩率} = \left(1 - \frac{\text{建模后轨迹点数}}{\text{原始轨迹点数}}\right) \times 100\% \quad (5)$$

由于建模前后台风轨迹中点记录的格式可能有所不同，直接在压缩率计算公式采用建模前后文件的大小可能会受记录格式的影响，不能客观反映轨迹数据压缩情况，所以本文压缩率公式中采用的是轨迹记录点数。显然，建模后减少的点数越多，压缩率越大，台风数据量减少得越多。

(2) 轨迹拟合误差（即轨迹拟合精度）

轨迹拟合误差用于描述轨迹拟合的精度，显然误差越小拟合精度越大。本文定义距离偏移标准差和轨迹偏移面积两个量描述轨迹拟合误差。

① 距离偏移标准差：

$$\text{距离偏移标准差} = \sqrt{\frac{\sum \text{dis}(X_{\text{cmpt}}, X_{\text{true}})^2}{\sum \text{台风记录点数}}} \quad (6)$$

距离偏移标准差主要是描述时空模型轨迹上的点的偏移真实轨迹上的点的程度，单位为 km。

② 轨迹偏移面积：

$$\text{轨迹偏移面积} = \sum \text{模型轨迹与实际轨迹围成的多边形面积} \quad (7)$$

轨迹偏移面积主要是描述一整条时空模型轨迹偏离真实轨迹的程度，单位为 km^2 。

2.2.2 建模效果评价

8 条典型台风的时空轨迹建模效果如表 2 所示，所有台风的时空轨迹建模效果如表 3 所示。

表 2 典型台风时空轨迹建模结果

典型台风名称	曲折度	原始记录 点个数	建模后点 个数	轨迹压缩率	距离偏移标 准差(km)	轨迹偏移 面积(km^2)
Washi2005	1.015	37	23	37.84%	23.39	251.98
Sarika2011	1.035	28	23	17.86%	21.56	1291.73
Jelawat2000	1.040	124	82	33.87%	20.03	260.02
Durian2006	1.045	121	76	37.19%	22.64	226.86
Huko2002	2.447	46	44	4.35%	8.58	1424.40
Fitow2001	2.699	190	141	25.79%	13.17	105.74
Jangmi2008	3.551	136	106	22.06%	15.32	510.08
Tokage2011	3.565	25	14	44.00%	14.22	74.23

注：距离阈值：50km；速率阈值：10km/h；角度阈值：5°

表 3 所有台风时空轨迹建模结果

	条数	原始记录 点个数	建模后点 个数	轨迹压缩率	距离偏移标 准差(km)	轨迹偏移 面积(km ²)
所有台风	2212	188452	139603	25.92%	16.58	610.60

从表 2 和表 3 中可知, 8 条典型台风的轨迹压缩率在 4.35%-44.00%之间, 而所有台风的平均轨迹压缩率达到了 25.92%;, 8 条典型台风的距离偏移标准差在 8.58-23.39km 之间, 而所有台风的平均距离偏移标准差为 16.58km, 均远小于目前台风 24 小时预报精度, 完全达到了实际应用的要求。

显然, 随着阈值的放大, 误差和轨迹压缩率的变化趋势都是增大。从初步建模的实验结果可以看出, 阈值有进一步放大的空间, 以实现在保证距离偏移标准差低于台风 24 小时预报精度的同时提升轨迹压缩率。

3 关于阈值的进一步讨论

3.1 阈值组合优化

在采用距离, 速率, 角度组合而成的三阈值系统中, 不同的阈值选取的组合将得到不同的压缩率和误差。我们可以推测得知, 某一压缩率可以通过多个三阈值组合得到, 但这些组合导致的轨迹拟合程度及距离偏移标准差不尽相同。故而我们选择了多组三阈值组合, 并对共计 2122 条台风进行了模型组合的测试, 得到各模型组合下的压缩率及轨迹拟合偏差。并对能得到同样压缩率的阈值组合进行比较, 选择轨迹偏差最小的阈值组合作为该压缩率的最优三阈值组合。

为方便进一步探索, 各阈值的取值情况如下: 距离阈值从 0 km 至 99 km, 每 1 km 间隔有一个取值, 共计 100 个; 速度阈值从 0 km/h 至 31.2 km/h, 每 0.8 km/h 间隔有一个取值, 共计 40 个; 角度阈值从 0° 至 90°, 每 5° 间隔有一个取值, 共计 19 个。故通过排列组合, 可得到共计 76000 个阈值组合, 并将他们用于对数据库内台风的 MOST 建模, 得出其压缩结果。

我们得到的轨迹偏移标准差与压缩率在全部测试阈值组合下呈现面状分布, 见图 1。我们筛选出面状的下沿曲线作为最优阈值组合, 在这样的筛选下将保留同一压缩率下距离偏移标准差最小的样本, 组成最优多阈值组合。此时最优多阈值组合得到的误差与压缩率的关系见图 2。

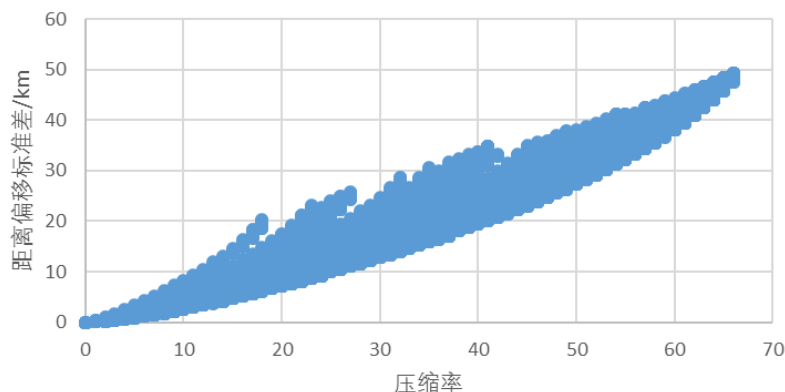


图 1 测试阈值组合下压缩率与距离偏移标准差关系

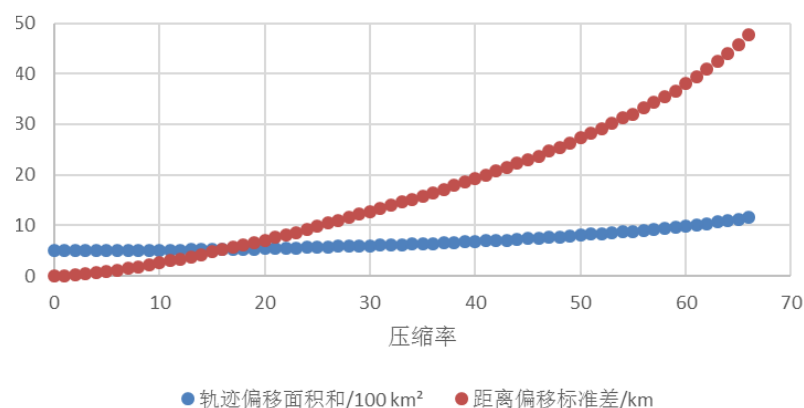


图 2 最优阈值组合的压缩率与轨迹误差关系

图 2 显示, 轨迹面积偏移量和轨迹偏移标准差均与压缩率保持了正相关关系, 即需要更大的压缩率的情况下, 相应的会带来较大的误差。我们尝试将两个轨迹偏移衡量指标分别与压缩率进行拟合, 可以得到较好的二次曲线关系, 残差值较小, 拟合度较高。具体可见图 3 中的拟合公式及拟合曲线。故而误差的累计并不是随着压缩率的减小线性增加的, 而是类指数级增长。而在压缩率较高的情况下, 可以观察到轨迹的距离偏移标准差的增长速度较二次拟合线要更快。从压缩效果上看, 在压缩率为 60%, 即数据量变为原数据量的 40% 的情况下, 距离偏移标准差为 40km, 仍小于预报精度 50-100km。压缩效果表现良好, 误差控制效果也较好。

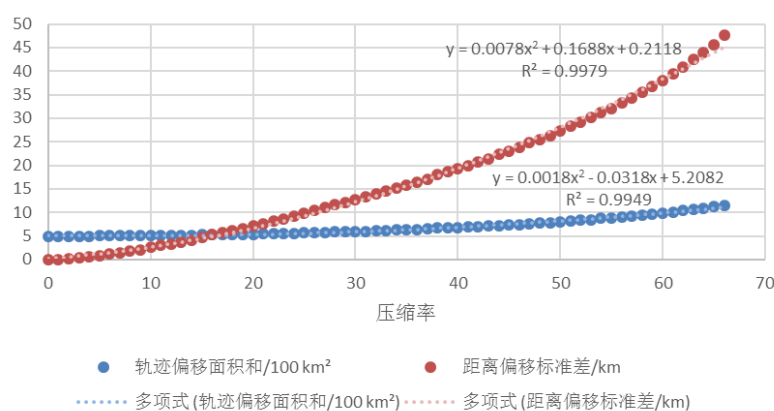


图 3 误差与压缩率关系拟合

3.2 各阈值与压缩率

我们尝试探寻最优阈值组合的选择规律, 故而将各阈值与压缩率的关系进行了进一步的探索, 见图 4。我们可以看到距离阈值与压缩率呈现较好的相关性, 通过拟合可得到三次多项式表达式进行二者关系的表达。速度阈值与角度阈值呈现出较为离散的情况。具体来看, 速度阈值仍呈现整体随着压缩率的上升而加大的情况。而角度阈值也表现为下沿数据点随着压缩率的上升而上升, 且所有数据点均处于 45° 对角线以上。

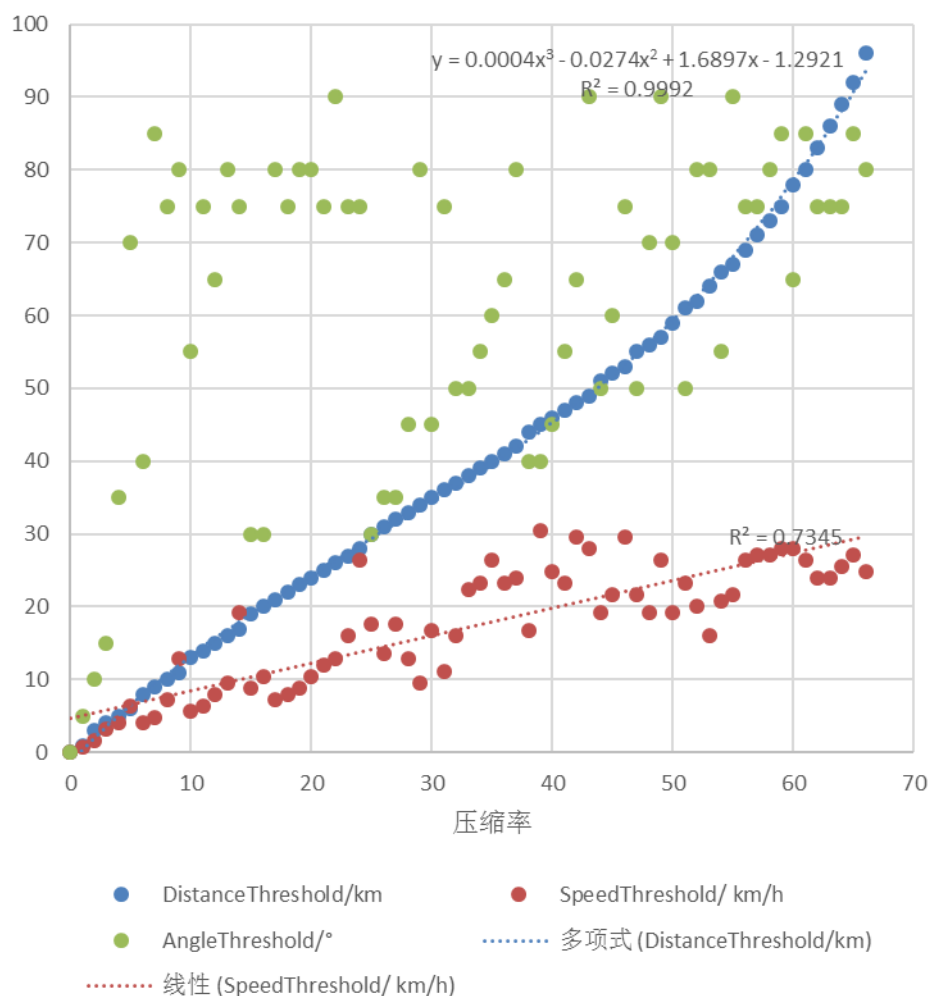


图 4 压缩率与各阈值关系

4 结论

针对传统的台风轨迹记录方式的不足，本文基于 MOST 设计了一个台风时空轨迹模型并进行了实验验证。实验结果表明此模型在保证数据精度的基础上减少了台风轨迹数据量，同时提供了台风位置全时域查询功能。在此基础上，本文还建立了关于台风时空轨迹建模效果的评价方法，并进行了三阈值系统的实验，对模型中更新策略的阈值选取和阈值优化的问题进行了研究。后续工作中拟开展对于台风进行聚类分析以对不同类别台风采用不同阈值的研究。

参考文献

- CHANG Liang, DUAN Zhongdong, OU Jinping. GIS applications in typhoon simulation and hazard assessment [J]. Harbin Inst Tech, 2005, 12 (4): 383 - 387.
- Erwig M, Güting R H, Schneider M, et al. Abstract and discrete modeling of spatio-temporal data types[C]. Proceedings of the 6th ACM international symposium on Advances in geographic information systems. ACM, 1998: 131-136.
- Forlizzi L, Güting R H, Nardelli E, et al. A data model and data structures for moving objects databases [M]. ACM, 2000.
- Gowrisankar H, Nittel S. Reducing uncertainty in location prediction of moving objects in road networks [C]. 2nd Int. Conference on Geographic Information Science (GIScience 2002), 2002.
- Güting R H, Böhlen M H, Erwig M, et al. A foundation for representing and querying moving objects [J]. ACM Transactions on Database Systems (TODS), 2000, 25(1): 1-42.
- Jin P, Yuchang Gong L Y. Design and implementation of a unified spatio-temporal data model [J]. Advances in Spatio-Temporal Analysis, 2007, 5: 67.
- Ouri Wolfson, Huabei Yin. Accuracy and Resource Consumption in Tracking and Location Prediction [C]. Lecture Notes in Computer Science 2750, 2003, 325-343
- Parent C, Spaccapietra S, Zimányi E. Spatio-temporal conceptual models: data structures+ space+ time[C]. Proceedings of the 7th ACM international symposium on Advances in geographic information systems. ACM, 1999: 26-33.
- Sistla A P, Wolfson O, Chamberlain S, et al. Modeling and querying moving objects[C]. 2013 IEEE 29th International Conference on Data Engineering (ICDE). IEEE Computer Society, 1997: 422-422.
- Wolfson O, Sistla A P, Chamberlain S, et al. Updating and querying databases that track mobile units [M]. Mobile Data Management and Applications. Springer US, 1999: 3-33.
- 国家气候中心. 全国气候影响评价[M]. 北京:气象出版社, 1998-2010.
- 汤杰, 陈国民, 余晖. 2010 年西北太平洋台风预报精度评定及分析[J]. 气象, 2011, 37(10): 1320-1328.
- 许映龙, 张玲, 高拴柱. 我国台风预报业务的现状及思考[J]. 气象, 2010, 36(7): 43-49.
- 中国气象局. 2003-2006 年热带气旋年鉴[CD]. 上海台风研究所制作, 2010.
- 朱海燕. GIS 空间分析方法在热带气旋研究中的应用 [D].上海:华东师范大学, 2005.