

# Shapley-模糊神经网络方法在华南台风卫星云图的长时效滚动预测中的应用\*

黄小燕<sup>1</sup> 何立<sup>1</sup> 赵华生<sup>1</sup> 黄颖<sup>1</sup> 吴玉霜<sup>2</sup>

HUANG Xiaoyan<sup>1</sup> HE Li<sup>1</sup> ZHAO Huasheng<sup>1</sup> HUANG Ying<sup>1</sup> WU Yushuang<sup>2</sup>

1. 广西壮族自治区气象科学研究所, 南宁, 530022

2. 广西壮族自治区气象台, 南宁, 530022

1. *Guangxi Institute of Meteorological Sciences, Nanning 530022, China*

2. *Guangxi Meteorological Observatory, Nanning 530022, China*

2019-09-26 收稿, 2020-12-03 改回.

黄小燕, 何立, 赵华生, 黄颖, 吴玉霜. 2021. Shapley-模糊神经网络方法在华南台风卫星云图的长时效滚动预测中的应用. 气象学报, 79(2): 309-327

Huang Xiaoyan, He Li, Zhao Huasheng, Huang Ying, Wu Yushuang. 2021. Application of Shapley-fuzzy neural network method in long-time rolling forecasting of typhoon satellite image in South China. *Acta Meteorologica Sinica*, 79(2):309-327

**Abstract** In order to make better use of the observational data of a large number of satellite cloud images to improve the forecasting ability of typhoon rain, increase the forecast accuracy of cloud system changes related to typhoon heavy precipitation, and increase the timeliness of the forecast for future cloud system changes, a model is constructed based on the cooperative strategy Shapley-fuzzy neural network for non-linear intelligent calculation of satellite cloud images of typhoons in South China. It is of great significance to enhance the practicability and timeliness of satellite cloud image data in the weather forecast of typhoon rain. Based on satellite cloud images of typhoon processes in South China during 2013–2016, an ensemble forecast method similar to the numerical forecast model is used to decompose the sample sequence of the cloud top brightness temperature of the satellite cloud images at 6 h intervals by the empirical orthogonal function, and the time coefficients decomposed are used as the forecast component in the cloud image forecast model. The development and changes of typhoon clouds are mainly affected by the atmospheric environmental flow field. Here, the physical quantity forecast products of the numerical forecast model are used as the forecast factors of individual forecast components of cloud images. And the stepwise variable selection algorithm of  $k$ -nearest neighbor mutual information estimation is adopted to realize the selection of related variables and the elimination of weakly related variables through a two-step process. The Shapley-fuzzy neural network ensemble forecast model of the corresponding time coefficient components is established, and the time coefficients and space vectors of each cloud image obtained by the forecast are further synthesized to reconstruct the forecast cloud image at the future time. In this way, the long-term and objective rolling prediction of the satellite cloud image at forecast lead time from 6 h to 72 h is realized. The prediction results of satellite cloud images of typhoons in southern China show that the cloud images predicted by the new scheme are highly correlated with the observed cloud images. The reconstructed predicted cloud image's basic contour, texture feature distribution, clarity, and cloud strength are relatively close to that of the observed cloud image. In addition, based on the forecast factors of the same cloud images, this paper further used the multiple linear regression scheme to predict the cloud images consistent with the new scheme for the same modeling and forecasting samples. Comparison of the results shows that this nonlinear forecasting model can better predict long-term development, movement, and trend of typhoon cloud cluster than the

\* 资助课题: 国家自然科学基金项目(41765002)、广西自然科学基金重点项目(2017GXNSFDA198030)、广西自然科学基金面上项目(2018GXNSFAA281281、2018GXNSFAA294128、2018GXNSFAA281229)、广西台风与海洋预报服务创新团队项目。

作者简介: 黄小燕, 主要从事智能计算预报技术在天气预报中的应用研究。E-mail: Gx\_huangxy@163.com

linear scheme. And the predicted cloud images are more consistent with the main features of the observed cloud images. The valid time of the cloud image forecast in this experiment has reached 72 h, which has a practical significance in operational weather forecasting.

**Key words** Satellite image, Typhoon in South China, Shapley-fuzzy neural network,  $k$ -nearest neighbor mutual information estimation, Multiple linear regression

**摘要** 为了更好地利用大量的卫星云图观测资料来提高台风暴雨的预报能力, 解决并提高对台风强降雨云系变化的预报精度, 延长对未来云系变化的预报时效, 构建基于合作对策 Shapley-模糊神经网络的华南区域台风卫星云图非线性智能计算滚动集合预测模型, 对增强卫星云图资料在台风暴雨天气预报中的实用性和及时性具有重要意义。依据 2013—2016 年华南区域台风影响过程的卫星云图, 采用类似于数值预报模式的集合预报方法, 通过对间隔 6 h 的卫星云图云顶亮温样本序列做经验正交函数分解, 将提取出的时间系数作为云图预报建模的预报分量。考虑台风云系的发展变化主要受云团环境物理量场的影响, 利用数值预报模式的物理量预报产品作为各预报分量的预报因子, 并采用  $k$ -近邻互信息估计的分步式变量选择算法, 通过两步过程实现相关变量的选择与弱相关变量的剔除, 分别建立相应时间系数的 Shapley-模糊神经网络集合预报模型, 进一步将预报得到的各时间系数与空间向量合成, 重构得到未来时刻的卫星云图预报图, 实现了云图 6—72 h 的长时效客观滚动预测。试验结果表明, 新方案所预测的云图与实况云图相关较高, 重构云图的基本轮廓、纹理特征分布、清晰度以及云系强弱方面都比较接近原始云图。另外, 研究进一步基于相同的云图预报因子, 针对同样的建模和预报样本采用多元线性回归方案进行和新方案一致的云图预测。对比结果表明, 这种非线性预报模型比线性方案能更好地预报未来较长时效台风云团的发展、移动的主要特征和变化趋势, 其预测的云图与实际云图的主要特征更相似。云图预报时效达到了 72 h, 具有业务实用价值。

**关键词** 卫星云图, 华南区域台风, Shapley-模糊神经网络,  $k$ -近邻互信息估计, 多元线性回归

中图法分类号 P456

## 1 引言

卫星云图以其时空分辨率高、覆盖面广的特点在气象领域得到了广泛应用, 成为除地面气象观测资料、空中风资料、空中温、压、湿资料等常规资料以外的重要的非常规气象资料。当前卫星云图的应用研究主要是在实时的监测分析和目标识别等方面 (Welch, et al, 1989; 张韧等, 2004; 王敏等, 2010; 张志清等, 2017; 李博等, 2018; 岳治国等, 2018; 郑建宇等, 2018; 丁璐等, 2018; 郝晓静等, 2019; 王新等, 2020; 崔林丽等, 2020; Zhang, et al, 2020), 人工判读仍是卫星云图分析的主要方法之一, 这既包含一定程度的主观因素, 不利于卫星云图丰富信息的充分提取, 同时也有碍天气预报制作的自动化与定量化。如果能实现对云图移动和变化状况进行一定时效的预测, 则可从一定程度上弥补实时卫星云图应用在时间上的缺陷, 能够显著增强卫星云图资料在天气预报中的实用性和及时性, 使卫星云图在灾害天气的预警、预报工作中发挥更大的作用。

中外对于云图的预测研究工作虽然已经进行了很久, 也取得了较大的研究进展, 但是主要以系统保持稳定为前提条件, 采用一些线性方法来基于

云图局部特征匹配和前后时次云图运动矢量关系进行线性外推, 在对云的发展趋势预测和预报时效上存在一定的不足, 限制了这些短时效云图预测的可用度 (Endlich, et al, 1971; Smith, 1975; Arking, et al, 1978; Li, 1998; Genkova, et al, 1999; 王登炎, 2000; 兰红平等, 2009; 蔡叔梅等, 2011; 梁立为等, 2015)。而且, 到目前为止, 对云图未来发展演变的预测研究大多是基于一些短时效 (1—3 h) 的预测, 超过 3 h 的更长时效的未来云图状况变化的云图预测理论和方法目前并不完善, 模型预测效果也不理想 (Hamill, et al, 1993; 白洁等, 1997; 王雷等, 1999; 龚克等, 2000; 陈刚毅等, 2005; 黄勇等, 2005; 刘科峰等, 2006, 2008; 王伟等, 2014)。

由于云的形成是一个非常复杂的过程, 云图随着时间的变化受到云内部和外部大气环境的综合影响, 因而云图与大气各要素场存在非线性关系, 而且云总是在不断发生移动、形变、扩展收缩甚至分裂、融合以及更复杂的生、消过程中, 使对云图追踪过程比计算机视觉中对一般的线性或近线性物体的追踪问题要难得多。因此, 传统的线性预测方法存在一定的局限性, 特别是超过 3 h 的未来云图状况变化, 仅依靠线性外推或不涉及云图移动变化的物理过程、动力和热力因子, 可能难以获得好的

预测结果。一些学者尝试运用非线性智能计算预报方法来解决这一问题,王继光等(2007)引入遗传算法构造了云图非线性动力预测模型,预报时效达到了42 h,云图预测模型能够较为客观合理地描述特定季节和区域云演变的基本结构,主要特征也描述较好。刘科峰等(2008)则提出了奇异值分解和径向基网络相结合的云图预测思想,对云图特征值进行动力模型重构和模型参数反演的云图演变非线性动力预测方法,该方法能合理地描述云运动的基本特征和演变趋势。但上述方法都没有考虑大气中水汽相态变化和云物理过程,制约了对云发展变化本质的刻画和描述。何如等(2010)、金龙等(2011)利用数值预报模式的物理量预报产品作为各预报分量的预报因子,并采用系统降维计算方法分别建立相应的时间系数神经网络预报模型,得出未来时刻的卫星云图预报图。但是这种非线性卫星云图预报模型在因子处理方法和预报模型上都还是处于探索阶段,云图的预测准确率也有待进一步的提高。因此,文中尝试根据云团发生、发展直至消亡的整个生命过程会明显受到大气环境场各个物理量因子影响的特点,并且云态变化更多的是非平稳、非线性的不规则变化,从云的演变过程与环境场影响因子的相关出发,建立二者的非线性统计关系模型,进一步设计采用 $k$ -近邻互信息估计的分步式变量选择算法(Kraskov, et al, 2004; 韩敏等, 2012),通过两步过程分别实现相关因子的选择与弱变量的剔除,并结合运用合作对策的Shapley值(Shapley, 1953; 雷勋平等, 2012)与模糊神经网络个体生成相结合的方法,设计一种新的非线性统计集合预报方法,来提高单一统计预报方程的预报技巧和预报稳定性,对卫星云图的未来发展演变进行预测,探索卫星云图较长时效的预报方法。

## 2 资料与预报建模设计

### 2.1 资料

华南区域受台风影响较为频繁,因此文中在建立卫星云图非线性集合预报模型时云图选用的是2013—2016年华南区域( $15^{\circ}$ — $30^{\circ}$ N,  $100^{\circ}$ — $120^{\circ}$ E)有台风影响过程的资料。

卫星云图资料选用的是FY-2E的云顶亮温(TBB)资料,云顶亮温是气象卫星获取的红外探测通道的数据,它是生成红外云图以及其他显示云图

中最原始的定量资料。云顶亮温是以相对于黑体温度而言,它的值一般低于 $0^{\circ}$ C,黑体温度越低表明云顶越高、对流越活跃,降水的潜在性越大。因此,可以根据云顶亮温以及一系列的相关数据,通过网络模型学习方法,为强降水的预测提供可能。文中云顶亮温投影方式为等经纬度投影;云图大小为 $401 \times 301$ 像素;星下点分辨率为 $0.05^{\circ} \times 0.05^{\circ}$ (5 km);采样间隔为6 h。

与华南台风卫星云图对应的预报因子来自于数值模式预报产品,选用欧洲气象资料中心(ERA-Interim)的全球再分析资料(<https://www.ecmwf.int/>),空间分辨率为 $0.75^{\circ} \times 0.75^{\circ}$ 经纬度,时间分辨率6 h(00、06、12和18时,世界时,下同),200、500、700、850 hPa 4个高度的温度、风场、位势高度场、速度场、涡度、散度、水汽通量、水汽平流、相对湿度、湿度以及各种物理量指数等,共4年(2013—2016年)。所选取的数值预报模式产品资料范围与云图资料范围一致。在卫星云图的物理量预报因子初选时,针对提取出的每张云图的 $n$ 个反映云图主要特征的预报分量,分别与每个要素场做场相关计算普查,通过设定适宜的阈值得到云图各个预报分量的初选预报因子。

### 2.2 预报建模设计

在华南区域台风卫星云图6、12、18、24、48和72 h共6个时效的滚动预报建模时,主要的设计思想包括构造一系列云图的时间序列,根据分辨率的不同每一幅云图包含了若干的像素点,若将每个点作为研究对象建立预测模型,计算量将非常庞大,同时也会增加模型的噪声干扰和计算误差。为此,根据Takens(1981)的相空间重构理论把原始的云图观测资料看成是某动力系统的离散值,求解与之相反的数值问题,即可重构出表现该云区移动的动力模型。即从云图资料序列中反演重构描述云区移动的动力预报模型。云图资料序列视为三维时间-空间场,研究引入经验正交函数分解(EOF)方法,对云顶亮温的数值矩阵采用EOF展开方法进行时空分离,将提取出来的空间场典型模态对应的时间系数序列作为模型输出的预报分量。进一步根据云图建模样本作EOF展开后的主分量的方差和累积方差贡献选出能够较好地反映云图主要特征的前若干个主分量对应的时间系数作为 $n$ 个预报分量,分别建立非线性集合预报模型。取这些空间

特征向量对应的时间系数作为建立云图预测模型的预报分量,可以减少噪声影响和误差积累,提高预报的精度。

云系的移动和生消变化等与云系周围大气环境条件密不可分,目前以欧洲预报中心为代表的数值预报模式产品的精度已经得到公认。在此合理地提取反映云系变化的主要特征,并找出这些主要特征与其未来变化密切相关的物理量预报因子,构建合适的数学预报建模方法,就可能建立较长时效的卫星云图预测模型。文中以欧洲预报中心数值预报产品为基础,构造对应的华南台风卫星云图预报因子。进一步采用 $k$ -近邻互信息估计的分步式变量选择算法,通过两步过程分别实现相关变量的选择与弱相关变量的剔除,从高维预报因子数据集中提取少数几个包含主要信息的变量作为卫星云图变化智能计算方法集成预报模型的输入,从而构建网络结构规模小、有效预报信息量大的非线性集成预报模型,提高卫星云图客观预报的精度。

目前的神经网络大多用于回归估计,集成的输出通常由各网络的输出通过简单平均或加权平均产生。当神经网络个体较多,集成规模较大时,优化权值会适得其反,适宜采用简单平均方法;而在集成规模较小,或者数据集中噪音较多时,优化权值将会提高学习系统的泛化能力(Sollich, et al, 1996)。Jin等(2008)也通过密度函数的积分推导,证明了当神经网络集成个体之间的差异越大时,神经网络集成预报的精度将越高。另外,由于云系移动或演变过程主要是受到云团内部和云系周围大气环境条件的综合影响,因此其变化特征主要为非平稳的非线性变化。为此,华南区域台风卫星云图的预测模型主要采用合作对策的Shapley值与模糊神经网络相结合的方法,设计一种新的非线性统计集合预报方法,确定模糊神经网络集成预报个体的权重系数,以此提高单一统计预报方程的预报技巧和预报稳定性。

最后,计算得到华南区域台风卫星云图独立样本序列的时间系数的预测值。与该云图样本通过EOF得到的空间特征向量经过时、空反演重构,得到云图独立样本二维云顶亮温矩阵的预测值,并对结果进行分析检验。图1为基于Shapley-模糊神经网络的华南区域台风卫星云图建模设计流程。

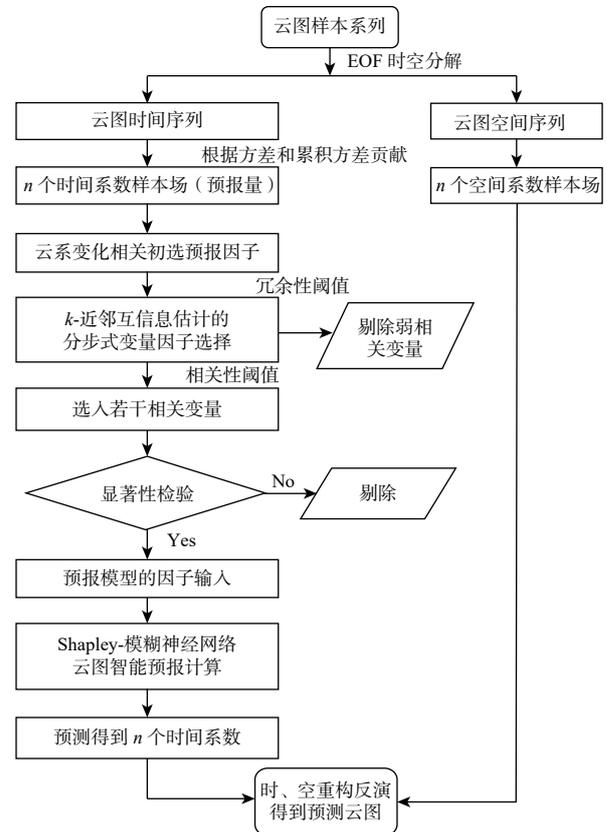


图1 基于Shapley-模糊神经网络的华南区域台风卫星云图建模设计流程

Fig. 1 Flow chart of the forecast model based on the Shapley-fuzzy neural network for satellite cloud images of typhoons in South China

### 3 方法

#### 3.1 $k$ -近邻互信息估计的分步式变量选择算法

针对每张云图都存在的高维因子集问题,采用 $k$ -近邻互信息估计的分步式变量选择算法,通过两步过程分别实现相关变量的选择与弱相关变量的剔除,替代原有的多元线性回归技术。该算法采用Kraskov等(2004)提出的基于 $k$ -近邻互信息计算方法,避免了直接对变量进行概率密度估计,适用于高维变量的互信息计算。其基本思想是:在 $X$ 和 $Y$ 构成的空间 $Z=(X,Y)$ 中,以 $\varepsilon_i/2$ 为点 $z_i=(x_i,y_i)$ 到其 $k$ -近邻的距离(通过最大值范数来度量), $\varepsilon_x(i)/2$ 为点 $z_i=(x_i,y_i)$ 到 $x$ 轴上相应点的距离,同理可得到 $\varepsilon_y(i)/2$ 。统计到点 $x_i$ 的距离严格小于 $\varepsilon_i/2$ 的数目( $n_x(i)$ ),同样对变量 $Y$ 作相同的处理得到 $n_y(i)$ 。则变量 $X$ 、 $Y$ 的互信息通过下式计算

$$I(X; Y) = \psi(k) - \langle \psi(n_x + 1) + \psi(n_y + 1) \rangle + \psi(N) \quad (1)$$

式中,  $N$  为随机变量  $X$  的样本数,  $\varepsilon_i (i = 1, 2, \dots, N)$  为任意小的数(文中取 0.01),  $\psi(*)$  为双  $\Gamma$  函数, 且满足  $\psi(1) = -0.5772516$ ,  $\psi(*+1) = \psi(*) + 1/*$ ; 符号  $\langle * \rangle$  表示对其中的所有变量  $i = 1, 2, \dots, N$  取平均; 同时, 近邻数目  $k$  的取值决定估计密度函数的光滑性,  $k$  值越大, 函数曲线越光滑, 同时互信息量  $I$  的统计误差越小, 而系统误差却相反。为了平衡这两种误差, 经过多次计算试验, 文中  $k=8$ 。

分布式算法的变量选择过程分为相关变量的选择与弱相关变量的剔除两个步骤, 理论上预报模型的输入矩阵只包含强相关变量, 而不应包含无关变量和弱相关变量。因此, 分步算法首先考虑输入变量与输出变量的相关程度, 分别得到强相关变量和弱相关变量的子集, 由于弱相关变量中包含冗余变量, 需要进一步对子集进行筛选, 剔除冗余变量。如果输入变量  $X_i$  满足

$$I(X_i; Y) > \delta_1 I(Y; Y) \quad (2)$$

式中,  $\delta_1 (\delta_1 \in [0, 1])$  为相关系数阈值, 则表明  $X_i$  包含了关于  $Y$  的一定量的信息, 即  $X_i$  为  $Y$  的相关变量。如果输入变量  $X_i$  不满足式(2), 可将  $X_i$  视为无关变量, 将其从输入子集中剔除。一般而言, 当特征维数较高时,  $\delta_1$  可以设置得较大, 从而有效降低特征维数。因为文中试验的预报因子特征维数较高, 经综合考虑和试验,  $\delta_1=0.35$  较合适。

通过  $k$ -近邻互信息估计的分步式变量选择算法从大量云图备选因子中选择出有限的预报因子后, 可以充分挖掘预报因子集的有效预报信息, 增强预报方法的物理基础, 提高预报模型的预报精度。

### 3.2 Shapley-模糊神经网络集合预报个体生成方法和原理

#### 3.2.1 模糊神经网络集合预报个体生成

考虑到云变化的复杂性和非线性, 在建立云图未来变化的预测模型时采用了模糊神经网络计算模型作为基础模型。模糊神经网络本质上也是一种多层前馈网络, 文中所采用的模糊神经网络结构为数值型 4 层前馈网络, 包括输入层、隶属度生成层、推理层和反模糊化输出层(王士同, 1998), 其中模糊规则形式为:

if  $x_1$  is  $\mu_{1k}$  and  $x_2$  is  $\mu_{2k}, \dots$ , then  $y = \omega_k, k = 1, \dots, m$

文中在模糊神经网络的隶属度生成层中, 采用

比较常见的高斯函数作为隶属度函数, 该函数具有较高的拟合和隶属性, 比较适合数据量比较大、预报精度要求较高的情况下运用。高斯隶属度函数仅由 2 个参数确定, 公式表达为

$$\mu_{ij} = \exp\left(-\frac{(x_i - a_{ij})^2}{\sigma_{ij}^2}\right) \quad (3)$$

$i = 1, 2, \dots, n; j = 1, 2, \dots, m$

式中,  $m$  是模糊分割数,  $a_{ij}$  和  $\sigma_{ij}$  分别确定隶属函数的中心和宽度。

在模糊逻辑中, 按照“和”运算, 一般有 3 种计算方法: ①按照取最小方式, ②按照相乘方式, ③按照相加方式。研究结果(Zeng, et al, 1996)表明, 相乘方式的性能优于取最小方式。文中在推理层中选择第②种方法, 即各节点的输出值分别为该节点所有输入的代数乘积。

$$\pi_j = \mu_{1j} \times \mu_{2j} \times \dots \times \mu_{nj} = \prod_{i=1}^n \mu_{ij} \quad (4)$$

$j = 1, \dots, m$

在输出层中, 采用如下的反模糊化网络输出。

$$y = \omega_1 \pi_1 + \omega_2 \pi_2 + \dots + \omega_m \pi_m \quad (5)$$

式中,  $\omega_j (j = 1, \dots, m)$  为连接权。

模糊神经网络是利用前馈网络的 BP 算法来训练调整参数, 在网络训练过程中, 设隶属度函数为式(3), 首先利用构建的预报因子学习矩阵对网络进行训练, 其主要计算步骤归结为

- (1) 设初始值  $i=1$ , 学习率  $\alpha=0.9$ , 网络训练误差为  $e_1$ 。
- (2) 初始时刻, 对网络的连接权、隶属函数的中心值及宽度值均用随机数进行初始化。
- (3) 采用 BP 算法对参数和权值进行训练调整。
- (4) 计算网络的实际输出与期望输出的误差  $e_2$ 。
- (5) 当  $e_2 > e_1$  时, 返回到步骤(3), 否则网络训练结束, 利用训练得到的网络参数和连接权进行预测, 得到预测个体。

#### 3.2.2 合作对策 Shapley 值在个体中的分配

多个模糊神经网络预测个体生成后会存在相互间的过拟合和泛化性问题, 为了改进模糊神经网络个体预报能力, 在进行华南区域台风卫星云图变化较长时效预测模型研究时, 主要设计采用 Shapley 值方法, 以组合预测有效度的平方和作为合作准则, 按合作对策的 Shapley 值在各模糊神经网络预报个体中进行分配, 从而确定各预报个体在模糊神

神经网络预报中的权重系数,提高模糊神经网络预报个体的差异度,建立一种新的卫星云图变化的机器学习客观集成预报模型。Shapley 值为如何决策一个  $n$  人讨价还价博弈中每个参与者所得的分配比例提供了一种很好的方法。近年来,学者们根据实际需要,对该方法做了相应改进(陈菁等, 2011; 陈启明等, 2012)。陈启明等(2012)将各单项预测模型看作组合预测合作对策的局中人,按合作对策的 Shapley 值确定各单项预测模型在组合预测中的权重系数。

具体假设存在  $m$  个模糊神经网络对某一个云图主分量  $Y_i(t=1, 2, \dots, N)$  进行了预测,其预测结果记为  $M=\{y_i, i=1, 2, \dots, m\}$ , 则  $M$  为组合预测方法的局中人集合。

设  $M$  的所有子集为  $2^M$ , 则  $M$  中的任一子集  $s \in 2^M$  形成组合预测方法的一个联盟,若干个局中人结成联盟后,这个联盟作为一个整体进行组合预测就是希望尽可能提高预测有效性。

设  $M=\{1, 2, \dots, m\}$ ,  $s \subseteq M$ ,  $v(s)$  为定义在  $2^M$  集合上的实值函数,令  $v(s)=F(s)$ , 满足  $v(\Phi)=0$ ,  $v(M) \geq \sum_{i=1}^m v(\{i\})$ , 则称组合预测方法为一合作  $m$  人对策。记为  $\Gamma=[M, v]$ 。  $v(s)$  称为  $m$  人对策的特征函数,其中,  $F(s)$  表示联盟  $s$  进行组合预测所得的预测有效度的平方和。

称  $v(s \cup \{i\}) - v(s)$  为第  $i$  种各单项预测方法对联盟  $s$  合作的“贡献”,其中  $s \subseteq M$ 。记  $\varphi_i(v)$  为第  $i$  种各单项预测方法的平均“贡献”,有

$$\varphi_i(v) = \sum_{i \in s} \frac{(m-|s|)! (|s|-1)!}{m!} [v(s) - v(s-\{i\})] \quad (6)$$

$i=1, 2, \dots, m$

式中,  $s$  为  $M$  中包含  $\{i\}$  的所有子集合;  $|s|$  为子集  $s$  中局中人的单项预测模型的个数。称  $\varphi(v) = (\varphi_1(v), \varphi_2(v))$  为合作  $m$  个模糊神经网络个体对策  $\Gamma=[M, v]$  的 Shapley 值,可以证明

$$\sum_{i=1}^m \varphi_i(v) = v(M) \quad (7)$$

式中,各单项模糊神经网络预报个体在组合预测方法中的平均贡献  $\varphi_i(v)$  之和等于合作的总成果,各单项模糊神经网络预报个体在组合预报方法中的加权系数的大小应根据合作中的平均贡献来确定。

考虑到  $v(M)$  为组合预测有效度,将  $\varphi_i(v)$  作如下归一化处理,可求得各单项模糊神经网络预报个体的权重系数

$$l_i = \varphi_i(v) / \sum_{j=1}^m \varphi_j(v) \quad (8)$$

显然,  $l_i \geq 0$ ,  $\sum_{i=1}^m l_i = 1$ 。式(8)即为所求的组合预测权系数的计算公式。用该式对华南区域台风卫星云图变化的模糊神经网络集成预报模型中的集成预报个体进行权重系数的计算,进而合理分配,提高个体之间的差异度,充分考虑了各单项预报模型对提高预测有效度的贡献是一种优性的组合预测,能较好地提高卫星云图预报的准确度。

## 4 华南台风卫星云图的非线性智能预报试验

### 4.1 预报建模样本和试验预报样本

台风所引发的强降水是目前大气科学研究的重点和难点,若能较为准确地预报未来云系的发生、发展情况,就能较好地判断台风强降水过程是否发生。因而,在建立卫星云图非线性集合预报模型时,主要选择的是华南区域有台风影响过程的云图资料作为预报试验对象。根据上述云图资料的选择标准以及云图资料的完整性,主要选择 2013—2015 年的台风过程 6 h 间隔共 196 张云图资料作为初次的预报建模样本建立预报模型,后续随着预报样本的增加,建模样本也相应增加。为了提高预报建模研究的客观性,以相同的标准,选择 2016 年的台风影响过程,共 86 张云图作为预报模型的独立样本进行实际的滚动预报检验。

### 4.2 预报对象的处理

由于每一张卫星云图的像素点较多(120701 个),如果对全部每个点都进行预报建模,计算量巨大,耗费时间长,对实际业务预报不适用。因此,采用自然正交函数分解方法对云图的实际云顶亮温数据场时间序列作时-空正交分解,将时、空要素场转化为若干空间的基本模态和相应的时间系数序列的线性组合,对时间系数进行建模预测,对预测后的时间系数与空间模态进行重构,得到新预测的卫星云图。

首先,采用 EOF 方法,对建模云图样本作分解

$$X = VZ \tag{9}$$

即任一时刻的云图可以表示为

$$X(x,y,t) = \sum_{i=1}^n v(x,y) \cdot \xi_i(t) \tag{10}$$

式中,  $v(x,y)$ 为 EOF 的主分量,  $\xi_i(t)$ 为相应于各主分量的展开时间系数,  $i$ 为自然正交展开的阶数。表 1 给出了云图建模样本作 EOF 展开后, 前 30 个主分量展开的方差和累积方差贡献。可以看到, 前 30 个主分量累积方差贡献接近 75%, 与原始云图的平均相关率为 81.3%。进一步增加 10 个主分量(第 31—40 个)进行计算分析可以发现, 其累积方差贡献仅提高 4.07 个百分点, 对应的相关率基本保持不变; 再继续增加到 50 个主分量, 累积方差贡献也仅增加了 3.09 个百分点, 相关率基本没有提高。分析可知, 后续的方差贡献率增加越来越缓慢, 但是增加的主分量越多, 意味着反演结果的误差也会增大。因此, 综合比较分析, 前 30 个主分量已经能够较好地反映云图的主要特征, 故取前 30 个主分量进行预报试验。

这 30 个主分量能否较好地还原出原始卫星云图的主要画面特征呢? 进一步采用分解出的这 30 个主成分的时间系数与空间相应的空间向量进行重构合成绘图, 并与原始卫星云图进行比较。运用 ENVI 软件进行卫星云图的绘制, 为了进行客观的对比分析, 在云图的颜色设定过程中, 以每一张原始图片的最大、最小值作为基准, 从 2.1 节介绍的云顶亮温的设定可知, 温度越低, 表明云顶越高, 对流越活跃, 从而降水的潜在性也越大, 因而温度低即值越小的数据用白色表示, 最大值设定为黑色, 从白色到黑色一共分为 256 级, 如此设定得到

每一张云图的标准色标。图 2 和 3 分别为 2013 年 1306 号台风“温比亚”和 2014 年 1409 号超强台风“威马逊”的 4 个时次的原始云图与提取的 30 个主分量向量合成云图的对比。可以发现, 前 30 个主分量描绘出的云图轮廓与原始云图的轮廓基本相似, 能较好地反映原始云图的主要特征。为此, 以这前 30 个主分量对应的时间系数  $T_1, T_2, T_3, \dots, T_{30}$  作为 30 个预报分量, 分别建立华南区域台风卫星云图的非线性集合预报模型。再根据式(10)将预报出的时间系数与相应的空间向量合成, 即可得出未来时刻的预报云图。

### 4.3 预报因子的处理

由云图观测资料可以清楚地看到, 大量的云系移动、生消变化都非常快, 特别是一些强对流云系, 可能在较短的时间内发生显著变化。这种云系随时间的快速变化, 其根本原因是云系内部的动力、热力、水汽状况发生了变化, 同时也与云系周围大气环境条件密不可分。因此, 要想能够有效地预测云系在相对长时间间隔(如 24 h 后)的状况, 必须在预报模型中考虑上述的云系内部、外部大气物理因子状况, 建立云系的移动、变化状况与导致云系变化的大气物理量因子的非线性关系。

按上述方法普查可以得到的初选云图预报因子数量众多, 如果全部将这些预报因子放入模型中进行建模, 则预报模型的输入结构将过于庞大, 容易出现过学习的现象。为了减少预报模型的输入节点, 同时又尽可能保留所有预报因子携带的预报信息, 在此采用  $k$ -近邻互信息估计的分步式变量选择算法, 通过两步过程分别实现相关变量的选择与弱相关变量的剔除, 对预报因子群进行降维去噪处

表 1 云图前 30 个主分量展开的方差和累积方差贡献 (%)  
Table 1 Variances and cumulative variance contributions of the first 30 principal components of the satellite cloud images (%)

项目	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
方差贡献	18.83	8.74	7.74	5.46	3.99	3.77	2.98	2.29	2.12	1.71
累积方差贡献	18.83	27.57	35.30	40.76	44.76	48.52	51.50	53.79	55.90	57.61
项目	PC11	PC12	PC13	PC14	PC15	PC16	PC17	PC18	PC19	PC20
方差贡献	1.68	1.54	1.38	1.24	1.05	1.02	0.95	0.93	0.89	0.77
累积方差贡献	59.29	60.83	62.21	63.44	64.49	65.51	66.47	67.40	68.29	69.06
项目	PC21	PC22	PC23	PC24	PC25	PC26	PC27	PC28	PC29	PC30
方差贡献	0.69	0.68	0.65	0.61	0.59	0.57	0.56	0.53	0.51	0.48
累积方差贡献	69.75	70.43	71.08	71.69	72.28	72.85	73.41	73.94	74.45	74.93

理。具体做法为

(1)采用式(1)对初选的云图预报因子群进行互信息计算。

(2)通过式(2)判定该预报因子与预报对象(云图时间系数)是否为相关变量,满足则选入,无关变量则剔除。

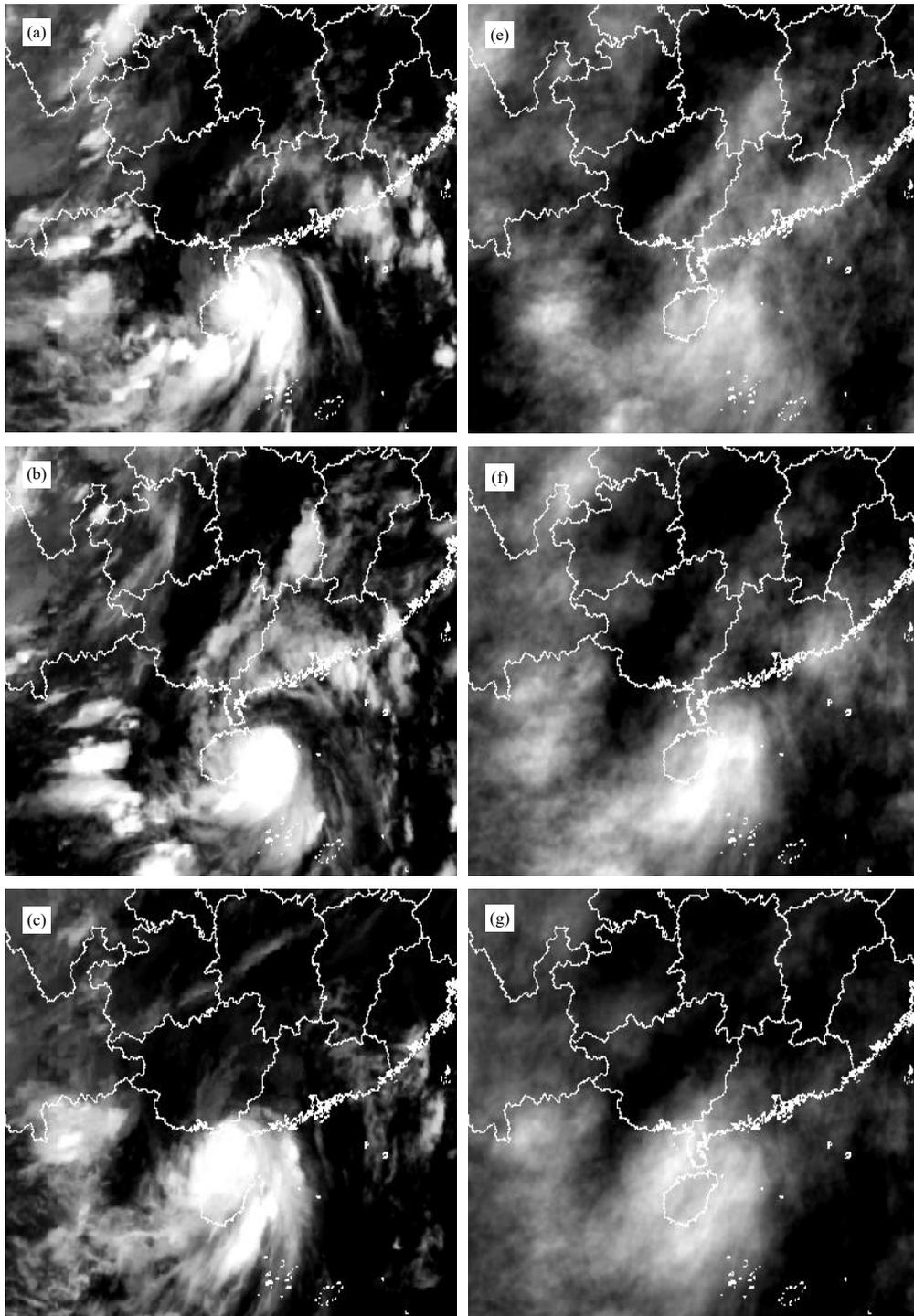
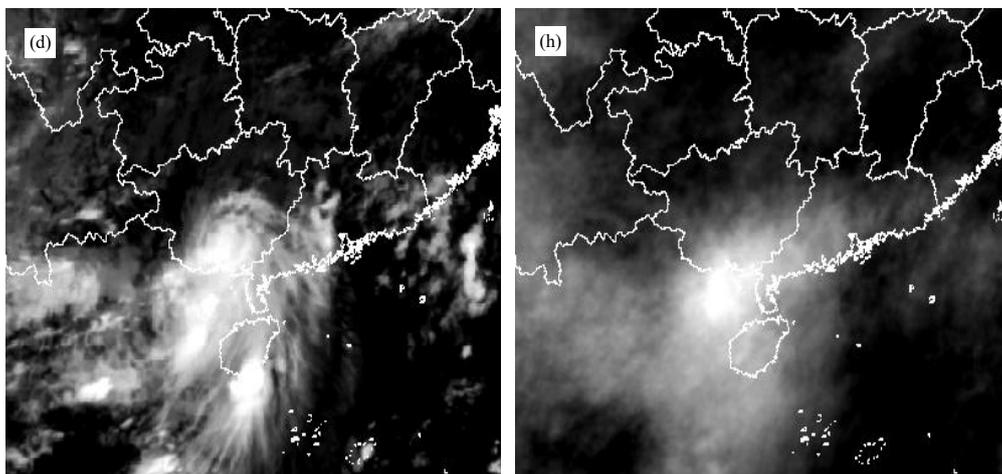


图2 台风“温比亚”4个时次(2013年7月1日12(a,e)、18(b,f)时,7月2日00(c,g)、06(d,h)时)实况云图(a-d)与30个主成分重构云图(e-h)

Fig. 2 Four times (12:00 (a, e), 18:00 (b, f) UTC 1 July 2013, 00:00 (c, g), 06:00 (d, h) UTC 2 July 2013) of observational cloud images (a-d) of Typhoon "Rumbia" and predicted cloud images (e-h) reconstructed by 30 principal components



续图 2

Fig. 2 Continued

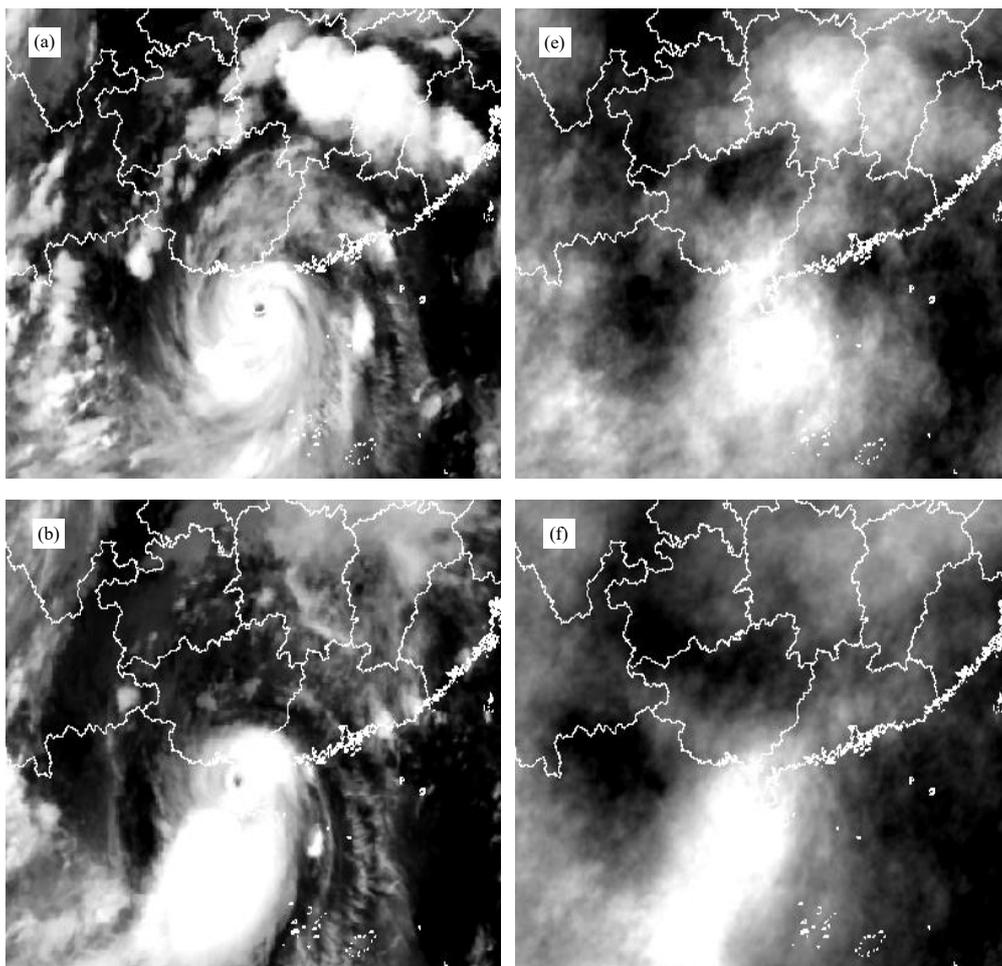
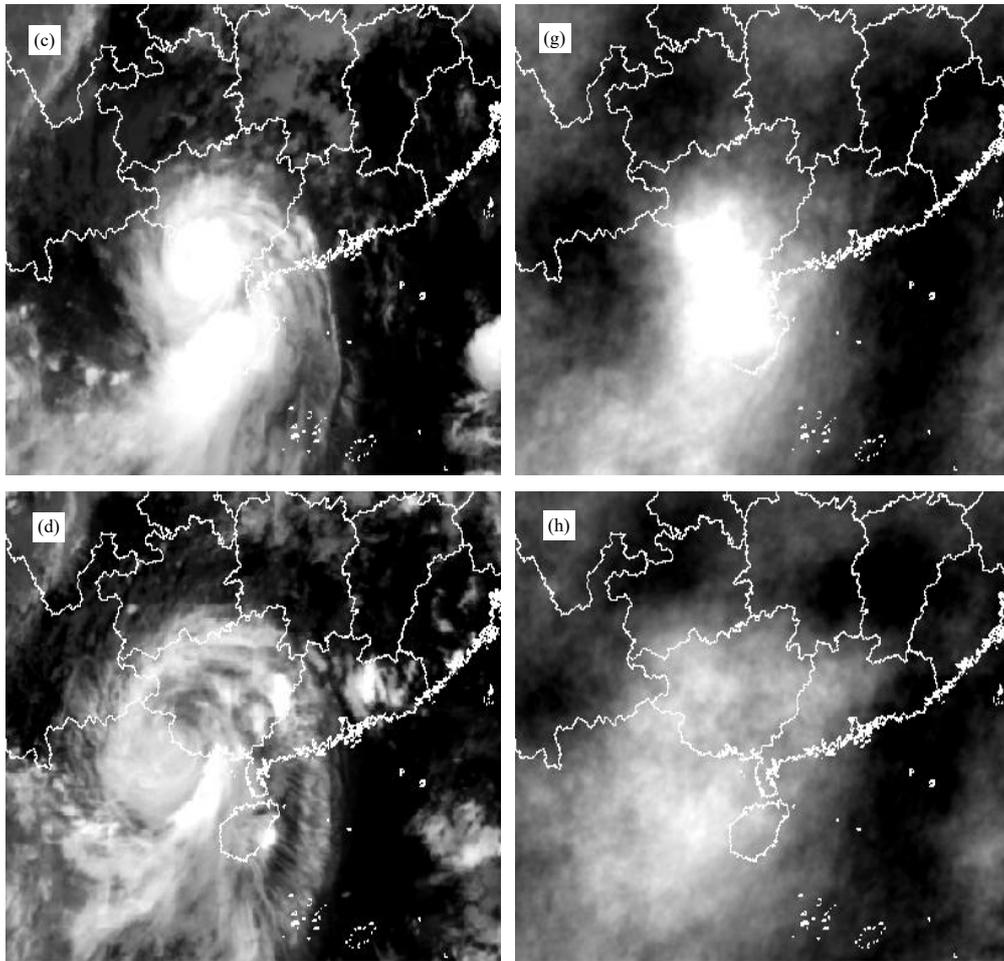


图 3 超强台风“威马逊”4 个时次 (2014 年 7 月 18 日 12 (a, e)、18 (b, f) 时、19 日 00 (c, g)、06 (d, h) 时) 实况云图 (a—d) 与 30 个主成分重构云图 (e—h)

Fig. 3 Four times (12:00 (a, e), 18:00 (b, f) UTC 18 July 2014, 00:00 (c, g), 06:00 (d, h) UTC 19 July 2014) of observational cloud images (a—d) of super typhoon "Rammasum" and predicted cloud images (e—h) reconstructed by 30 principal components



续图 3

Fig. 3 Continued

(3)对所有云图预报因子做 $k$ -近邻互信息估计的分步式变量选择算法筛选后得到的所有因子组合在一起构成新的云图预报因子集,新组成的高相关云图预报因子群进一步进行显著性检验,保留通过显著性检验的预报因子。

具体以6 h的预报时效为例,在对6 h的30个云图时间系数预报分量的智能计算预报模型研究时,首先根据欧洲预报中心数值预报模式6 h间隔的各种物理量相关要素场普查初选的云图预报因子。针对第1个云图时间系数,首先进行欧洲预报中心模式中的某一要素某一层级的每一个格点(共567个)分别与云图时间系数的相关计算,将成片( $\geq 10$ 个格点)稳定高相关(相关系数 $\geq 0.3$ )格点作为预报因子的选择区,在区内选3个相邻格点的最大平均值作为该相关区的代表值,作为待选因子。共在21个云图物理量场中提取预报因子,包括各

层次的散度场、垂直速度场、涡度场、比湿场等,以及各种云图与水汽相变相关的物理量指数——K指数场、 $ky$ 指数场,SI指数场等。经过初选因子的相关计算后,第1个云图时间系数预报分量的预报因子的相关系数最大(0.7042),入选的因子总数量超过2000个。可以发现,随着云图时间系数预报分量的增加,相关在不断下降,到第30个时,初选因子的最大相关系数仅为0.4033,相关程度下降近一半。分析可知,前3个云图时间系数预报分量对于云图的预报精度有更大贡献。

进一步用 $k$ -近邻互信息估计的分步式变量选择算法、设定 $\delta_1=0.35$ 阈值计算判识预报对象(云图时间系数)的高相关变量,此时第1个主分量的预报因子已经只剩下247个,而第30个主分量的数量仅为64个。进一步进行显著性检验后,第1主分量选入预报模型仅36个,而第30个主分量的因子数

量为 12 个。

值得注意的是, 进行显著性检验后的预报因子集可以很好地控制因子间的复共线性关系, 同时能分解出有物理意义的云图预报因子, 剔除影响预报效果的噪音因子, 从而控制选入最终预报模型的预报因子数量。一般来说, 适宜的预报因子矩阵结构进入预报模型能得到较高的预报精度, 但是模型输入过于庞大, 则容易产生过拟合等问题(金龙等, 2004)。

#### 4.4 Shapley-模糊神经网络华南台风卫星云图智能计算滚动预报试验

利用 3.2 节 Shapley-模糊神经网络集合预报建模的集合预报个体生成方法, 分别建立预报时效为 6、12、18、24、48 和 72 h 的各 30 个云图时间系数预报分量的智能计算集合预报模型。在此, 模糊神经网络预测个体生成的参数设置统一为: 各云图预报分量的预报因子数量为网络的输入节点, 输出节点为 1, 3 个推理层节, 网络训练次数设定为 1000 次, 学习因子取 0.9, 总体误差定为 0.001。训练过程中, 为了提高计算效率, 设定了两种终止训练的条件, 一是计算误差小于目标误差(0.001), 二是训练次数达到 1000 次, 计算过程满足其中的条件之一, 就终止网络训练。

具体以 6 h 预报时效为例, 针对卫星云图的第一个时间系数, 采用 Shapley 值方法, 以组合预测有效度的平方和作为合作准则, 按合作对策的 Shapley 值对各模糊神经网络预报个体进行合理分配, 确定各预报个体在集成预报中的权重系数, 提高集成预

报个体的差异度, 并充分考虑各单项预报模型对提高预测云图有效度的贡献, 从而建立云图第一个时间系数的预报模型进行云图预测建模。在计算过程中, 按照时间顺序, 逐次滚动进行, 即用建模样本(196 个)确立预报模型后, 采用该预报模型首先对第一个独立预报样本进行预报, 得到预测结果后, 加入到建模样本中(此时建模样本增加 1, 为 197 个), 接着对第二个独立样本进行预报检验, 依此类推, 完成最后一个(第 282 个)独立预报样本的预报, 如此就得到云图第一个时间系数的全部 86 个独立预报样本的预测值。按该方法和预报步骤, 对其余的 29 个云图时间系数分量进行预报, 得到最终的 30 个云图时间系数的全部独立预报样本的预测值, 使用该预测结果, 与相应的 30 个云图空间向量进行合成, 可以得到预报样本的每一张云图的预报结果。按照此方法, 完成所有预报时效的云图滚动预测。

#### 4.5 云图预测效果检验分析

##### 4.5.1 相关性评估分析

通过上述的建模预报试验, 可以分别得到预报时效为 6、12、18、24、48 和 72 h 的各 30 个时间系数预报分量的 Shapley-模糊神经网络集合预报结果, 分别与对应的空间系数进行合成, 得到预测云图。检验预报效果可用每一张预测云图与实况云图的像素点相关进行评估。这里进一步分别统计这 6 个预报时效的每一张预测云图的像素点与实况云图像素点的相关系数(图 4a)。可以看到, 新方案的 6 个预报时效的预报中, 这些华南区域台风影响

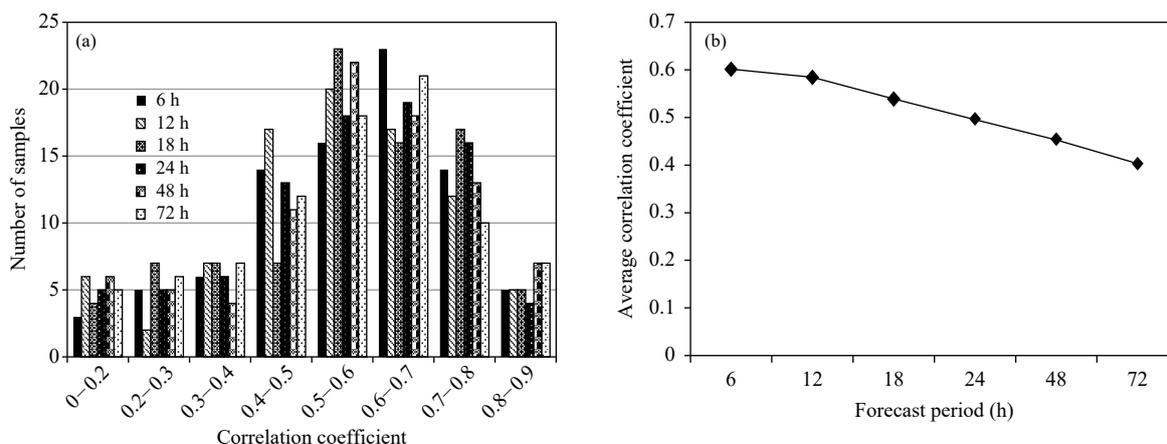


图 4 新方案预测云图与实况云图云顶亮温值的相关 (a. 相关系数分布, b. 平均相关系数)

Fig. 4 Correlation between cloud images predicted by the new scheme and cloud top brightness temperatures of the observed cloud images (a. distribution of correlation coefficient, b. average correlation coefficient)

过程 86 张预测云图与实况云图的相关系数主要分布在 0.4—0.8, 相关系数为 0.5—0.6 的个数以及 0.6—0.7 个数最多, 其次为相关系数 0.7—0.8 的个数, 0.4 以下相关较差的云图个数较少。总体而言, 新方案预测云图与实况的相关总体较好, 相关系数 0.5 以上的个数超过总样本的半数。

进一步分别统计 6 个预报时效的平均相关系数(图 4b)发现, 随着预报时效的延长, 相关逐渐下降, 这与实际情况吻合。主要原因与预报因子随着预报时效的延长相关逐渐减弱有关。但是值得注意的是, 云图预测时效达到了 72 h, 这样长的预报时效是有意义的。

为了客观分析基于  $k$ -近邻互信息估计的分步式变量选择的 Shapley-模糊神经网络集合预报方法对华南区域台风卫星云图预报的优越性, 将其预报结果与传统采用的线性方法进行比较。针对同样的建模样本和独立预报检验样本, 采用同样的方案选出的云图预报因子进行多元线性回归的预报建模运算。可以发现(表 2), 在每一张预测云图的像素点与实况云图像素点的相关系数分布场上, 多元线性回归方案的 6 个预报时效的预报中, 这些台风过程 86 张预测云图与实况云图的相关系数主要

分布于 0.3—0.7, 相关为 0.4—0.5 的个数最多, 其次为 0.5—0.6 个数, 0.7 以上相关较强的云图个数较少。对比分析可知, 在预测云图与实况云图的相关上, 新方案的相关性明显高于多元线性回归方案。6 个预报时效的平均相关系数的对比情况同样如此(图 5), 两个方案的预报结果随着预报时效的增加, 相关都是逐步下降的, 但是新方案的预报精度均明显高于多元线性回归方案。

另外, 两种预报方案都存在预测云图与原始云图相关系数低于 0.3 的预报较差个例, 以预报效果最佳的 6 h 预报时效进行分析发现, 新方案中这样的个例有 8 个, 其时序分布基本不连续。分析这些云图个例的特征可以发现, 共同表现为云型比较松散, 对流云体积小并且分布零散, 这些个例较多为台风前期发展和后期消散阶段, 说明预报方案对该时段分散凌乱云的预报能力相对较弱。而线性回归方案在这方面的表现却不太一致, 其预报较差的个例基本与新方案的个例不同, 9 个个例的时序连续分布较多。云型的分布特征差别也较大。后者中大多数云图的有云区和无云区的分布较为分明, 对流云体积较大, 云系处于较为旺盛的阶段, 但是线性回归方案对这些云系的预报能力却不足。在

表 2 新方案和多元线性回归方案预测云图与实况云图云顶亮温值的相关系数样本数对比(单位: 个)

Table 2 Comparison of the number of samples for the correlation coefficients of the cloud images predicted by the new scheme and by the multiple linear regression scheme with the cloud top brightness temperatures of observed cloud images (unit: piece)

相关系数	方案	6 h	12 h	18 h	24 h	48 h	72 h
0.0—0.2	新方案	3	6	4	5	6	5
	多元回归	5	7	7	7	7	7
0.2—0.3	新方案	5	2	7	5	5	6
	多元回归	4	5	6	6	6	7
0.3—0.4	新方案	6	7	7	6	4	7
	多元回归	13	13	13	10	11	13
0.4—0.5	新方案	14	17	7	13	11	12
	多元回归	17	14	14	20	18	24
0.5—0.6	新方案	16	20	23	18	22	18
	多元回归	20	19	17	21	20	17
0.6—0.7	新方案	23	17	16	19	18	21
	多元回归	19	19	20	15	13	9
0.7—0.8	新方案	14	12	17	16	13	10
	多元回归	6	7	7	5	7	6
0.8—0.9	新方案	5	5	5	4	7	7
	多元回归	2	2	2	2	4	3

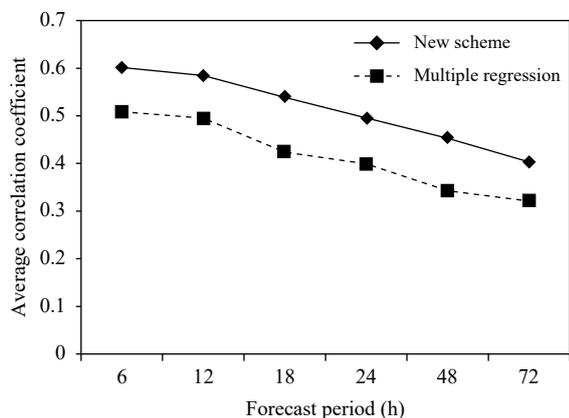


图5 不同预报时效新方案及多元回归预测云图与实况云图云顶亮温值的相关系数

Fig. 5 Distribution of correlation coefficient between the cloud images predicted by the new scheme and the cloud top brightness temperatures of the observed cloud images

预报因子一致的情况下,两种方案预报表现却几乎不一样,这除了与线性和非线性方案的预报性能有关外,是否还与各方案对入选的预报因素的敏感性

有关,值得进一步深入研究。

#### 4.5.2 实际预报个例分析

为了更直观地对比分析非线性新方案和线性方案预报性能的差异,进一步针对预报检验的独立样本进行预测云图与实况云图的绘制,绘制方案与4.2节介绍的绘制方案相同。图6和7为2016年影响华南区域的台风“妮姐”实况云图与预测重构云图的对比。其中图6为“妮姐”发展旺盛期的卫星云图,图7为其减弱接近消失的云图。

具体分析可知,图6a为2016年8月2日18时的实况云图,图6b为30个主成分未做预报的拟合云图,两者相似度很高,说明提取的30个主成分分量是可以很好地展示原始云图的主要特征的。图6c—h分别为预报时效6、12、18、24、48和72h的各30个时间系数预报分量的Shapley-模糊神经网络集合预测云图。可以看到,6个时效预测云图与实况云图在云的基本轮廓、纹理特征分布以及云系强、弱方面相似度都很高。特别是在台风的主体

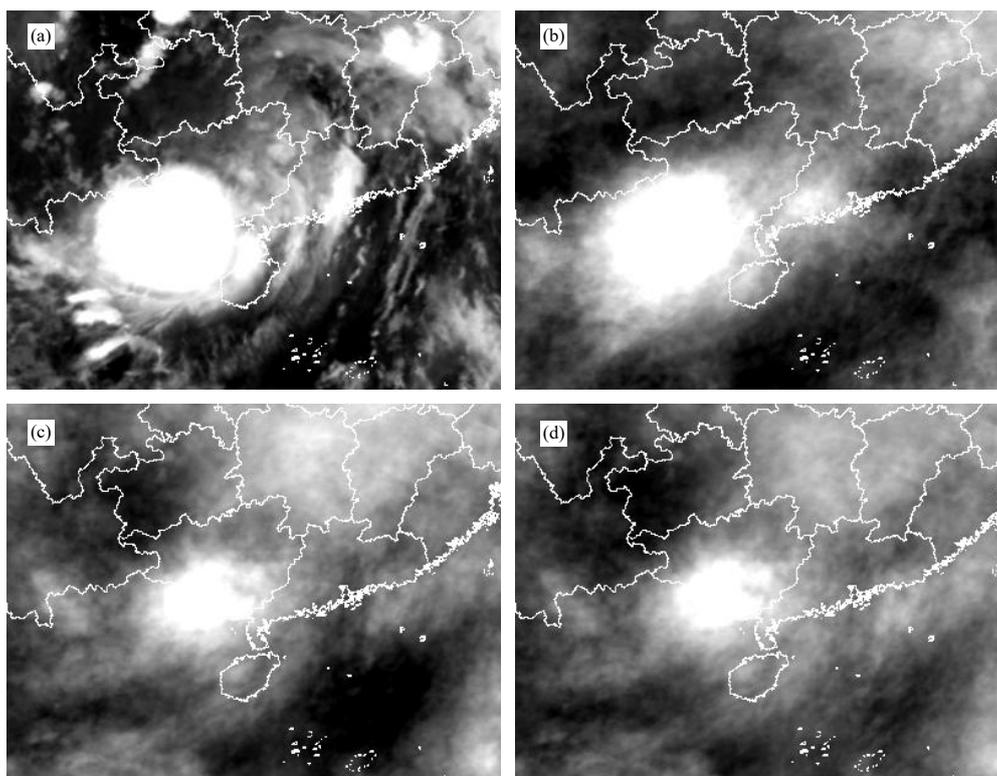
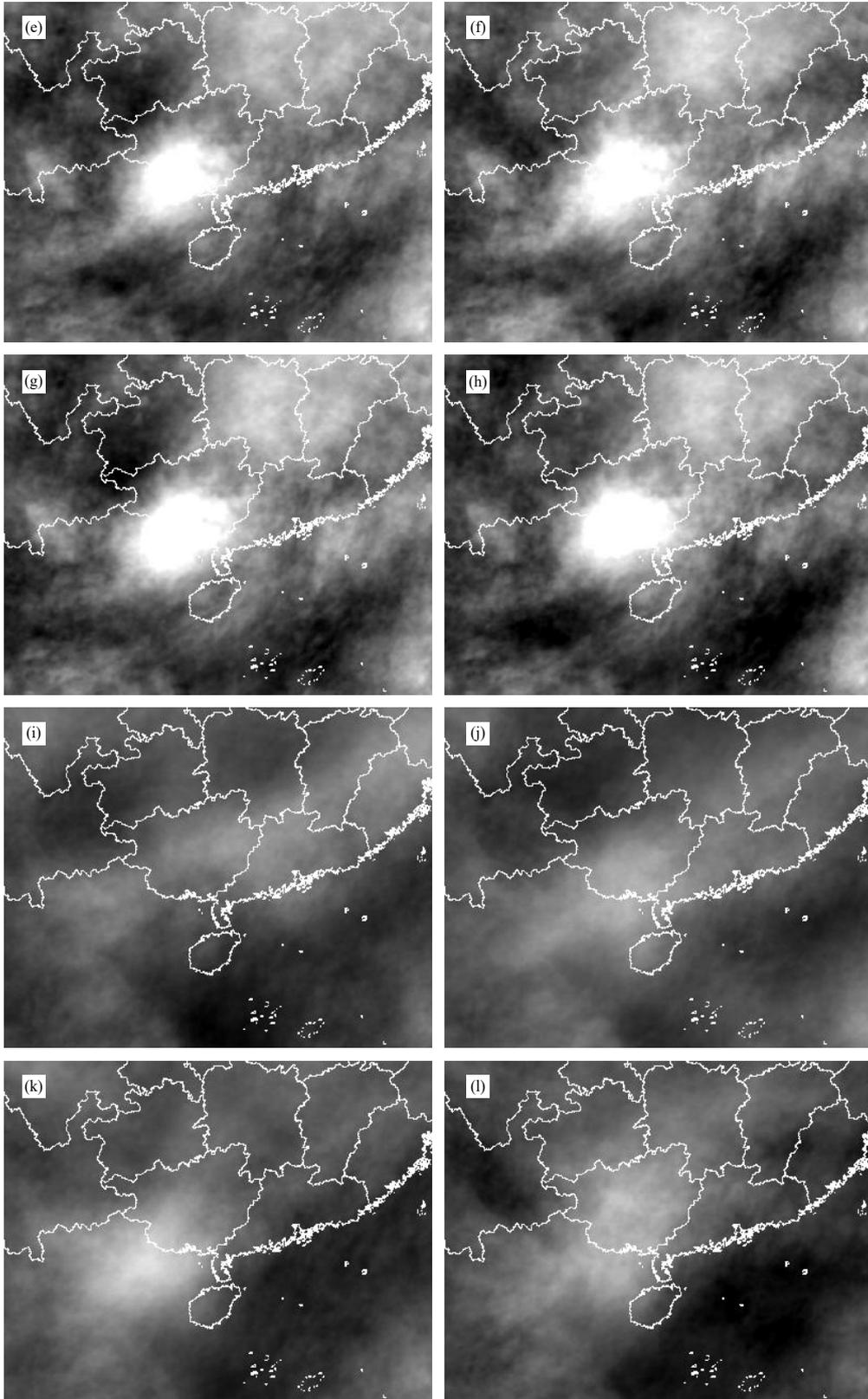


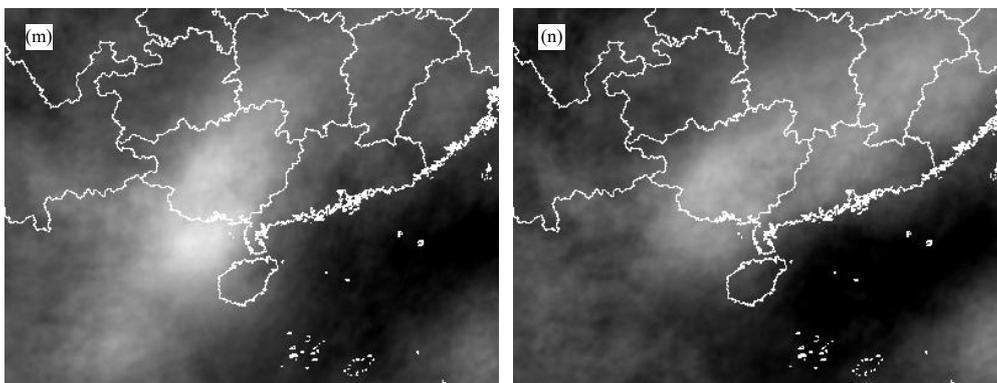
图6 台风“妮姐”实况云图与预测重构云图(2016年8月2日18时实况云图(a),30个主成分重构云图(b),非线性方案预测云图(c—h对应6—72h);线性方案预测云图(i—n对应6—72h))

Fig. 6 Observed and predicted cloud images of Typhoon "Nida" at 18:00 UTC 2 August 2016 (a. observed cloud images, b. cloud images predicted by 30 principal components, c—h. cloud images predicted by nonlinear scheme (6—72 h), i—n. cloud images predicted by linear scheme (6—72 h))



续图 6

Fig. 6 Continued



续图 6

Fig. 6 Continued

云团即广西西南部到越南和北部湾一带的云图的预报上, 预测云图基本都能有所体现, 形状较好, 强度也较接近实况。而多元线性回归方案所得到的预测云图(图 6i—n), 则明显在云系强弱上有所欠缺, 清晰度也不够, 即黑白不分明, 云系轮廓和纹理都较难分辨。

分析预测的像素点上的数值可以发现, 实况云图的最小值为 1760(对应白色, 云图发展旺盛, 对流较强), 最大值是 2940(对应黑色, 为无云状态)。新方案所预测的云图的像素值最小值的 6 个预报时效为 1793—1973, 最大值则分布在 2896—3043, 与实况云图在最小值上的差距为 33—213, 差距较小。

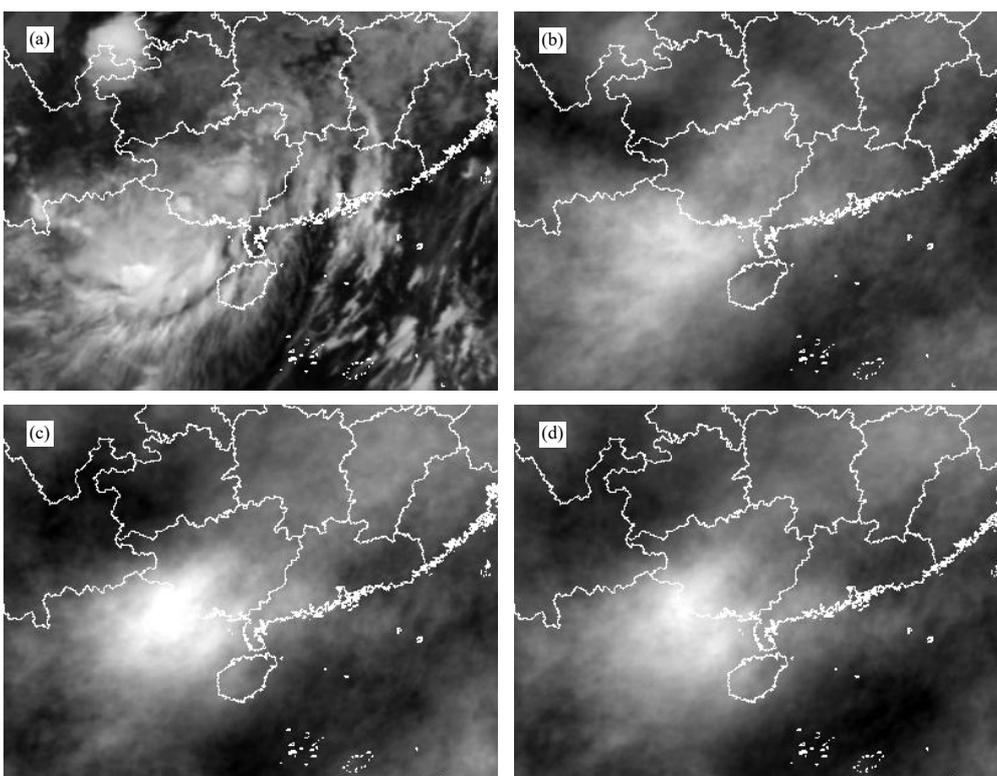
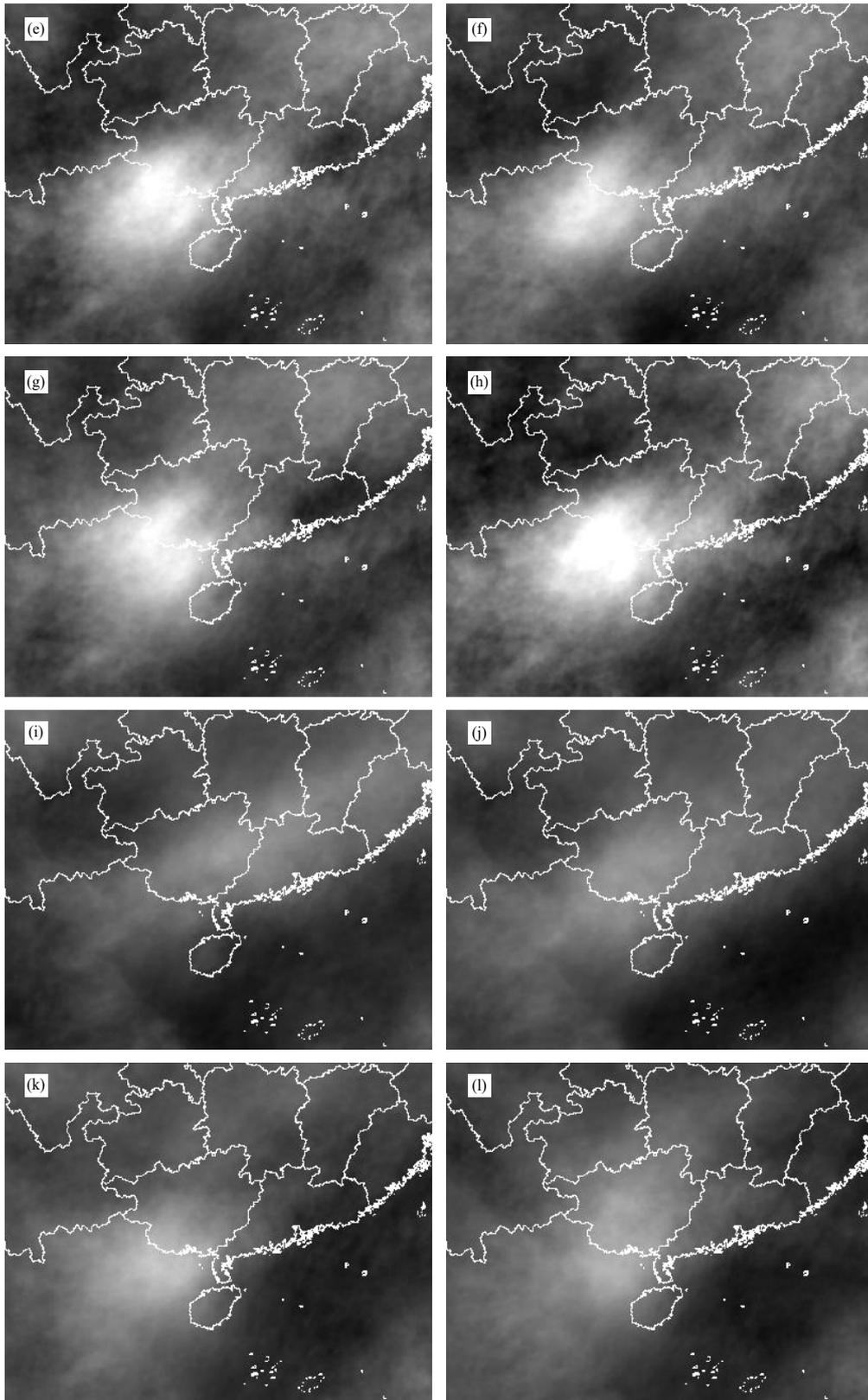


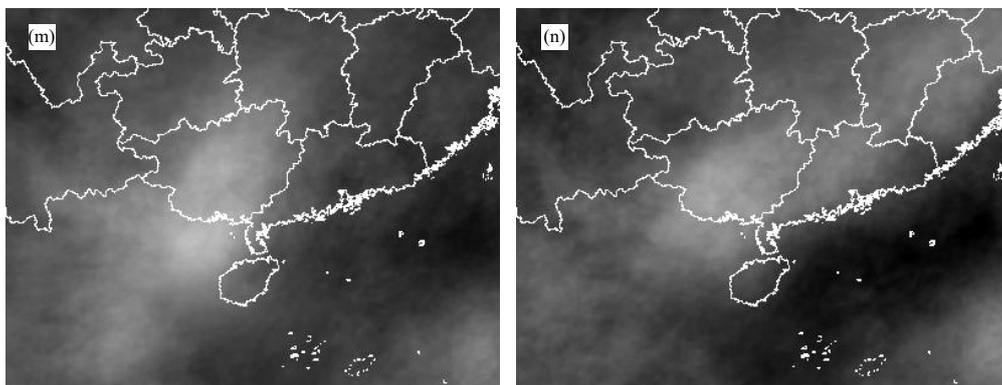
图 7 台风“妮姐”实况云图与预测重构云图(2016年8月3日00时)(a. 云图, b. 30个主成分重构云图, 非线性方案预测云图(c—h对应6—72h), 线性方案预测云图(i—n对应6—72h))

Fig. 7 Observed and predicted cloud images of Typhoon "Nida" at 00: 00 UTC 3 August 2016, observed cloud images (a), cloud images predicted by 30 principal components (b), cloud images predicted by nonlinear scheme (c—h correspond to 6—72 h), cloud images predicted by linear scheme (i—n correspond to 6—72 h)



续图 7

Fig. 7 Continued



续图 7

Fig. 7 Continued

因此,能较好地显示台风云图对流旺盛的区域。多元线性回归方案的最大、最小值的预报与实况云图差距较大,云图的像素值最小值的6个预报时效在2183—2392,最大值则分布在2814—2955,与实况云图在最小值上的差距较大,为423—632,因而在强度上预报都偏弱,不能较好地显示云图发展强盛的部分。

分析图7的情况亦如此,新方案6个时效预测云图(图7c—h)与实况云图在云的基本轮廓、纹理特征分布以及云系强弱方面相似度比多元线性回归方案(图7i—n)都更好。这是“妮姐”的一个消亡减弱阶段图,实况图上显示其较强的主体云团主要位于越南和北部湾上空,广西境内还存在一些残留云系。新方案6个时效预测云图(图7c—h)在位置上预报把握较理想,强度则比实况云图略强。而多元线性回归方案(图7i—n)在强度上的预报则同样表现偏弱,图片清晰度较差。总体上,新方案预测的重构云图效果优于多元线性回归方法。

## 5 结 语

为更好地利用大量的卫星云图观测资料提高对台风暴雨的预报能力,解决并提高对台风强降水云系变化的预报精度,加大对未来云系变化的预报时效,根据华南区域台风卫星云图生消、演变的非线性、时变性和非平稳性特点,从寻找对云团未来变化有显著影响的大气物理量预报因子出发,进行未来6—72 h共6个预报时效的云图变化趋势集合预报建模。

该方案的主要特点是:在云图预报建模过程中,采用经验正交函数分解方法进行华南台风云图

的特征抽取,并以时间系数作为预报建模的预报分量,建立从云图中抽取的主要特征向量与数值预报模式产品的非线性映射关系,使各预报分量预报模型具有很好的物理基础。并进一步对预报因子采用 $k$ -近邻互信息估计的分步式变量选择算法,通过两步过程分别实现相关变量的选择与弱相关变量的剔除,充分挖掘了预报因子集的有效预报信息。而在数学预报建模方法上,设计了类似于数值预报模式集合预报方法的Shapley-模糊神经网络集合预报方法。该方法不同于大气科学中传统的统计预报方法,是一种基于智能计算方法的非线性统计集合预报建模方法。从研究的华南区域实际台风卫星云图预测试验结果看,该方法对各预报分量进行实际预报时,不需要调整预报建模的各项参数,各预报分量预报模型具有很好的稳定性和普适性,这为开展云图预测提供了客观实用的新方法。

另外,还进一步做了基于相同的云图预报因子,针对同样的建模和预报样本采用多元线性回归方案进行和新方案一致的云图预测。对比结果表明,新方案预测的云图与实况云图相关高的样本更多,重构云图的基本轮廓、纹理特征分布、清晰度以及云系强弱方面都优于多元线性回归方案。

值得注意的是,文中的云图预报时效达到了72 h,具有业务实用预报意义。

## 参考文献

- 白洁,王洪庆,陶祖钰. 1997. GMS 卫星红外云图强对流云团的识别与追踪. 热带气象学报, 13(2): 158-167. Bai J, Wang H Q, Tao Z Y. 1997. Recognition and tracing of severe convective cloud from IR images of GMS. J Trop Meteor, 13(2): 158-167 (in Chinese)

- 蔡叔梅, 阮征, 陈钟荣. 2011. 基于自适应阈值的云团识别与追踪方法及个例试验. *气象科技*, 39(3): 332-338. Cai S M, Ruan Z, Chen Z R. 2011. Method for identifying and tracking cloud clusters based on adaptive threshold temperature and a case study. *Meteor Sci Technol*, 39(3): 332-338 (in Chinese)
- 陈刚毅, 丁旭羲, 赵丽妍. 2005. 用模糊神经网络自动识别云的技术研究. *大气科学*, 29(5): 837-844. Chen G Y, Ding X X, Zhao L Y. 2005. An automatic pattern recognition techniques of cloud based on fuzzy neural network. *Chinese J Atmos Sci*, 29(5): 837-844 (in Chinese)
- 陈菁, 代小平, 陈祥等. 2011. 基于改进的 Shapley 值法的农业节水补偿额测算方法. *水利学报*, 42(6): 750-756. Chen J, Dai X P, Chen X, et al. 2011. Calculation of compensation amount for agricultural water saving based on improved Shapley method. *J Hyd Eng*, 42(6): 750-756 (in Chinese)
- 陈启明, 陈华友. 2012. 一类组合预测模型的权系数确定的 Shapley 值方法. *安徽大学学报(自然科学版)*, 36(2): 29-34. Chen Q M, Chen H Y. 2012. Shapley value approach to determining the weights of a kind of combined forecasting model. *J Anhui Univ (Nat Sci Ed)*, 36(2): 29-34 (in Chinese)
- 崔林丽, 郭巍, 葛伟强等. 2020. FY-4A 卫星云顶参数精度检验及台风应用研究. *高原气象*, 39(1): 196-203. Cui L L, Guo W, Ge W Q, et al. 2020. Comparisons of cloud top parameter of FY-4A satellite and its typhoon application research. *Plateau Meteor*, 39(1): 196-203 (in Chinese)
- 丁璐, 童晓冲, 秦志远等. 2018. 三轴稳定型静止轨道遥感卫星指向确定的地标匹配方法. *测绘学报*, 47(11): 1506-1517. Ding L, Tong X C, Qin Z Y, et al. 2018. Landmark matching method about pointing to the ground of three-axis stabilization geostationary remote sensing satellite. *Acta Geod Cartogr Sinica*, 47(11): 1506-1517 (in Chinese)
- 龚克, 叶大鲁, 葛成辉. 2000. 卫星云图预测的运动矢量方法. *中国图象图形学报*, 5(4): 349-352. Gong K, Ye D L, Ge C H. 2000. A method for geostationary meteorological satellite cloud image prediction based on motion vector. *J Image Graph*, 5(4): 349-352 (in Chinese)
- 韩敏, 刘晓欣. 2012. 基于互信息的分步式输入变量选择多元序列预测研究. *自动化学报*, 38(6): 999-1006. Han M, Liu X X. 2012. Stepwise input variable selection based on mutual information for multivariate forecasting. *Acta Autom Sinica*, 38(6): 999-1006 (in Chinese)
- 郝晓静, 李清亮, 郭立新等. 2019. 基于气象卫星数据的我国沿海悬空波导反演方法初步研究. *电子学报*, 47(3): 600-605. Hao X J, Li Q L, Guo L X, et al. 2019. An preliminary research on inversion method of elevated duct from meteorological satellite observation over Chinese regional seas. *Acta Electron Sinica*, 47(3): 600-605 (in Chinese)
- 何如, 管兆勇, 金龙. 2010. 一种神经网络的云图短时预测方法. *大气科学学报*, 33(6): 725-730. He R, Guan Z Y, Jin L. 2010. A short-term cloud forecast model by neural networks. *Trans Atmos Sci*, 33(6): 725-730 (in Chinese)
- 黄勇, 孔庆欣, 郑兰芝. 2005. 基于最大互相关的对流云预测. *气象科学*, 25(4): 399-404. Huang Y, Kong Q X, Zheng L Z. 2005. Convective cloud forecast based on maximum interrelation. *Scientia Meteor Sinica*, 25(4): 399-404 (in Chinese)
- 金龙, 况雪源, 黄海洪等. 2004. 人工神经网络预报模型的过拟合研究. *气象学报*, 62(1): 62-70. Jin L, Kuang X Y, Wang H H, et al. 2004. Study on the overfitting of the artificial neural network forecasting model. *Acta Meteor Sinica*, 62(1): 62-70 (in Chinese)
- 金龙, 黄颖, 何如. 2011. 卫星云图的遗传神经网络集合预测模型研究. *计算机工程与应用*, 47(32): 231-235. Jin L, Huang Y, He R. 2011. Nonlinear ensemble prediction model for satellite image based on genetic neural network. *Comput Eng Appl*, 47(32): 231-235 (in Chinese)
- 兰红平, 孙向明, 梁碧玲等. 2009. 雷暴云团自动识别和边界相关追踪技术研究. *气象*, 35(7): 101-111. Lan H P, Sun X M, Liang B L, et al. 2009. An automatic tracking and recognition algorithm for thunderstorm cloud-cluster (TRACER). *Meteor Mon*, 35(7): 101-111 (in Chinese)
- 雷勋平, 邱 R. 2012. Shapley 值法的改进及其应用研究. *计算机工程与应用*, 48(7): 23-25. Lei X P, Qiu R. 2012. Research on improvement and application of Shapley value. *Comput Eng Appl*, 48(7): 23-25, 29 (in Chinese)
- 李博, 杨柳, 唐世浩. 2018. 基于静止卫星的青藏高原及周边地区夏季对流的气候特征分析. *气象学报*, 76(6): 983-995. Li B, Yang L, Tang S H. 2018. The climatic characteristics of summer convection over the Tibetan Plateau revealed by geostationary satellite. *Acta Meteor Sinica*, 76(6): 983-995 (in Chinese)
- 梁立为, 杨秀洪, 尹洁等. 2015. 一种云图的短时预测新方法探讨. *高原气象*, 34(4): 1186-1190. Liang L W, Yang X H, Yin J, et al. 2015. The new short-term cloud forecast method. *Plateau Meteor*, 34(4): 1186-1190 (in Chinese)
- 刘科峰, 张韧, 孙照渤. 2006. 基于交叉相关法的卫星云图中云团移动的短时预测. *中国图象图形学报*, 11(4): 586-591. Liu K F, Zhang R, Sun Z B. 2006. A cloud movement short-time forecast based on cross-correlation. *J Image Graph*, 11(4): 586-591 (in Chinese)
- 刘科峰, 张韧, 李文才等. 2008. 奇异值分解与神经网络结合的卫星云图云团移动预测. *解放军理工大学学报(自然科学版)*, 9(3): 298-301. Liu K F, Zhang R, Li W C, et al. 2008. Cloud cluster movement forecast technique of satellite cloud pictures based on singular value decomposition and artificial neural networks. *J PLA Univ Sci Technol*, 9(3): 298-301 (in Chinese)
- 王登炎. 2000. MCS 的形态特征和外推预报. *气象*, 26(8): 22-24. Wang D Y. 2000. Morphology feature and extrapolation forecast of MCS. *Meteor Mon*, 26(8): 22-24 (in Chinese)
- 王继光, 张韧, 郁文贤等. 2007. 基于卫星云图历史资料反演云团非线性预测模型. *国防科技大学学报*, 29(5): 41-47. Wang J G, Zhang R, Yu W X, et al. 2007. Non-linear forecast model of cloud clusters movement based on parameters retrieval of historical satellite cloud pictures time series. *J Natl Univ Defense Technol*, 29(5): 41-47 (in Chinese)
- 王雷, 黄培强, 江汉青. 1999. 利用卫星资料制作天气系统移动预报. *气象科学*, 19(3): 270-275. Wang L, Huang P Q, Shi H Q. 1999. The forecast of

- weather system motion using cross relations. *Scientia Meteor Sinica*, 19(3): 270-275 (in Chinese)
- 王敏, 韩雷. 2010. 基于 FY-2C 卫星的暴雨云团自动预警方法. *计算机工程*, 36(14): 241-242. Wang M, Han L. 2010. Automatic forecast method for heavy rainfall cloud clusters based on FY-2C satellite. *Comput Eng*, 36(14): 241-242 (in Chinese)
- 王士同. 1998. 神经模糊系统及其应用. 北京: 北京航空航天大学出版社.
- Wang S T. 1998. *Neuro-fuzzy System and Its Application*. Beijing: Beihang University Press (in Chinese)
- 王伟, 刘娟, 孟志斌. 2014. 基于时序遥感卫星云图的对流云团动态追踪预测. *电子学报*, 42(4): 804-808. Wang W, Liu J, Meng Z B. 2014. Identifying and tracking convective clouds based on time Series remote sensing satellite images. *Acta Electron Sinica*, 42(4): 804-808 (in Chinese)
- 王新, 唐世浩, 曹治强. 2020. 风云气象卫星“一带一路”热带气旋监测能力与最新进展. *海洋气象学报*, 40(2): 10-18. Wang X, Tang S H, Cao Z Q. 2020. Capability and latest progress of tropical cyclone monitoring over the Belt and Road area by FENGYUN meteorological satellites. *J Mar Meteor*, 40(2): 10-18 (in Chinese)
- 岳治国, 余兴, 刘贵华等. 2018. NPP/VIIRS 卫星反演青藏高原夏季对流云微物理特征. *气象学报*, 76(6): 968-982. Yue Z G, Yu X, Liu G H, et al. 2018. NPP/VIIRS satellite retrieval of summer convective cloud microphysical properties over the Tibetan Plateau. *Acta Meteor Sinica*, 76(6): 968-982 (in Chinese)
- 张韧, 王海, 孙照渤等. 2004. 双光谱卫星云图的模糊推理云分类. *防灾减灾工程学报*, 24(3): 257-263. Zhang R, Wang H, Sun Z B, et al. 2004. Cloud fuzzy inference and classification based on double-spectrum satellite images. *J Disaster Prev Mitig Eng*, 24(3): 257-263 (in Chinese)
- 张志清, 陆风, 方翔等. 2017. FY-4 卫星应用和发展. *上海航天*, 34(4): 8-19. Zhang Z Q, Lu F, Fang X, et al. 2017. Application and development of FY-4 meteorological satellite. *Aerosp Shanghai*, 34(4): 8-19 (in Chinese)
- 郑建宇, 刘东, 王志恩等. 2018. CloudSat/CALIPSO 卫星资料分析云的全球分布及其季节变化特征. *气象学报*, 76(3): 420-433. Zheng J Y, Liu D, Wang Z E, et al. 2018. Global distribution and seasonal variation of clouds observed from CloudSat/CALIPSO. *Acta Meteor Sinica*, 76(3): 420-433 (in Chinese)
- Arking A, Lo R C, Rosenfeld A. 1978. A fourier approach to cloud motion estimation. *J Appl Meteor*, 17(6): 735-744
- Endlich R M, Wolf D E, Hall D J, et al. 1971. Use of a pattern recognition technique for determining cloud motions from sequences of satellite photographs. *J Appl Meteor*, 10(1): 105-117
- Genkova I S, Pachedjieva B, Ganey G, et al. 1999. Cloud motion estimation from METEOSAT images using time mutability method//Proceedings of SPIE 3571, Tenth International School on Quantum Electronics: Laser Physics and Applications. San Jose, CA, United States: SPIE, 297-301
- Hamill T M, Neukorn T. 1993. A short-term cloud forecast scheme using cross correlations. *Wea Forecasting*, 8(4): 401-411
- Jin L, Yao C, Huang X Y. 2008. A nonlinear artificial intelligence ensemble prediction model for typhoon intensity. *Mon Wea Rev*, 136(12): 4541-4554
- Kraskov A, Stögbauer H, Grassberger P. 2004. Estimating mutual information. *Phys Rev E*, 69(6): 066138
- Li Z J. 1998. Estimation of cloud motion using cross-correlation. *Adv Atmos Sci*, 15(2): 277-282
- Shapley L. 1953. A value for n-person games//Kuhn H W, Tucker A W. *Contributions to the Theory of Games II*. Princeton: Princeton University Press, 307-317
- Smith E A. 1975. The McIDAS system. *IEEE Trans Geosci Electron*, 13(3): 123-136
- Sollich P, Krogh A. 1996. Learning with ensembles: How over-fitting can be useful//Touretzky D S, Mozer M C, Hasselmo M E. *Advances in Neural Information Processing Systems 8*. Cambridge, CO: MIT Press, 190-196
- Takens F. 1981. Detecting strange attractors in turbulence//Rand D, Young L S. *Dynamical Systems and Turbulence*, Warwick 1980. Berlin, Heidelberg: Springer, 361-381
- Welch R M, Navar M S, Sengup S K. 1989. The effect of spatial resolution upon texture-based cloud field classification. *J Geophys Res*, 94(D12): 14767-14781
- Zeng X J, Singh M G. 1996. Approximation accuracy analysis of fuzzy systems as function approximators. *IEEE Trans Fuzzy Syst*, 4(1): 44-63
- Zhang W J, Hui W, Lyu W, et al. 2020. FY-4A LMI observed lightning activity in super typhoon Mangkhut (2018) in comparison with WWLLN data. *J Meteor Res*, 34(2): 336-352