

张仙¹ 吴琮² 陈以祺² 黎亚少² 王巍巍³

基于 Holt-ARIMA-Lagrange Multiplier 的 CWC 信息熵时序研究

摘要

降水云系的发展过程及其特征分析,是云降水物理学中的一个重要问题.本文选取一次云发展过程中的 700 hPa 云水含量 (Cloud Water Content, CWC) 和大气垂直方向上气流速度 (Ω , OMG) 的 1 h 值,以信息熵来度量 CWC 空间分布的混沌程度,辅以 OMG 的时间变化来判断云的发展,并提出了一种基于多尺度分解、Holt 模型、自回归滑动平均模型 (Autoregressive Integrated Moving Average Model, ARIMA) 和 Lagrange Multiplier 的组合预测方法.结果表明:1) CWC 熵具有非线性和非平稳性;2) 在云的不同发展阶段,北方 CWC 熵序列的均值都小于南方,方差普遍大于南方;3) OMG 区域均值与 CWC 熵的小波低频重构的极值点在时间上有很好的对应关系,相近的极值点在南方中占 50%,在北方中占 83.3%,表明 CWC 熵可以在一定程度上反映云系的发展;4) CWC 熵序列往往具有多种时间尺度特征,故进行多尺度分解之后再组合建模的 Holt-ARIMA-Lagrange Multiplier 模型比单一预测方法、单层分解的预测模型更优,准确率提高 3% 以上.

关键词

云水含量;信息熵;小波分解;经验模态分解 (EMD);自回归滑动平均模型 (ARIMA);Holt 两参数指数平滑法

中图分类号 P457.6

文献标志码 A

收稿日期 2022-06-22

资助项目 国家自然科学基金 (42075068, 41975176, 41975087);国家重点研发计划重点专项 (2018YFC1507905)

作者简介

张仙,女,工程师,研究方向为气候统计分析.332941628@qq.com

吴琮(通信作者),女,研究方向为气候预报和应用统计学.1836950110@qq.com

0 引言

云是自然界水循环的有形结果,内部存在复杂的反馈机制.云水含量能够反映云系的组织和形态,其信息熵能够描述云系的自组织状况.段海霞等^[1]通过涡旋自组织动力学研究了西北地区几次降水过程,得到了暴雨过程中降水云团自组织预报判断的依据.张小娟等^[2]于 2019 年利用中尺度数值模式 WRF 的数值模拟,结合卫星资料等,分析了冰雹云系的发展演变特征及其云物理结构.陈逸伦^[3]基于云系和降水系统的整体性,揭示了云团时空变化特征.学者们也对中国不同地区云水量的时空分布及变化趋势进行了分析,云水量存在季节变化特征,南多北少^[4-5].云水含量是气候数值模拟的预报参量之一,也是研究云系内部发展的重要参数.

实测数据概率分布的离散性较高,故信息熵成为度量云水含量不确定程度的有效工具.1995 年,张学文等^[6]提出熵气象学方法.王惠娟^[7]提出基于信息熵的 PM_{2.5} 浓度的不确定研究,在小范围内同时考虑了指标的变化规律和空间分布.Berta 等^[8]研究了熵的不确定度以及测量的可逆性.李丽娟等^[9]总结了熵不确定度研究的发展历史和进展,对不确定关系进行推广,得到了更加普适的数学关系表达式.

气象数据通常具有非平稳、波动大等特点,在时间序列分解研究方面,单一的分解误差常常较大.由于不同分解方法各有优势与不足,采用混合多尺度分解方法,再对分解后的分量序列分别进行预测,然后集成,最终的预测效果较好^[10].例如,自回归滑动平均模型 (Autoregressive Integrated Moving Average Model, ARIMA) 适用于平稳和非平稳序列,故 ARIMA 具有广泛的适用性.Geetha 等^[11]通过 ARIMA 模型预测了飓风在未来一段时间内的发展趋势;Xiong 等^[12]提出一种混合建模框架,将区间 Holt 指数平滑法和多输出支持向量回归相结合,发现对线性趋势数据的预测效果较好;汪漂^[13]提出混合区间多尺度分解模型,将区间序列分解成区间趋势和残差,分别利用 Holt、支持向量

1 中国民用航空华北地区空中交通管理局气象中心,北京,100621

2 南京信息工程大学 数学与统计学院,南京,210044

3 南京信息工程大学 大气与环境实验教学中心,南京,210044

机和 BP 神经网络预测;Wu 等^[14]提出一种基于二次分解的 AQI 组合预测的优化算法,综合考虑了多种影响因素.针对非线性、非平稳、多时间尺度的时间序列,采用多尺度分解与组合预测的方法能够有效提高模型的预测性能^[15-17].

本文选取一次云发展过程中的 700 hPa 云水含量的信息熵,借助分量分解、Holt、ARIMA 模型和 Lagrange Multiplier 集成等方法,建立一种适用于非线性、非平稳气象数据的组合预测模型,以期云物理学中云系组织形态演变规律研究提供帮助.

1 数据和描述性统计分析

1.1 研究数据

选取 2020 年 9 月 30 日至 10 月 3 日中国部分地区(96.62°~143.39°E, 26.16°~53.13°N, 图 1) 700 hPa 云水含量(Cloud Water Content, CWC)和大气垂直方向上气流运动速度(Airflow Velocity in the Vertical Direction of the Atmosphere, OMG)的 1 h 数据,表 1 列出数据提取的阈值.共有 96 个时间点,时间 1 表示 2020 年 9 月 30 日 0 时,时间 2 表示 9 月 30 日 1 时,……,时间 96 表示 10 月 3 日 23 时,期间有一次降水过程.研究数据来自外场观测资料和 Weather Research and Forecast Model 模式资料,范围包括东北地区、华北地区、中部地区和绝大部分西南地区等,以 40°N 为界的南、北各呈现出一大片云,格点为 6 km×6 km.

表 1 云系发展情况研究变量

Table 1 Variables in research of cloud development

名称	变量	单位	阈值
700 hPa 云水含量	CWC	g/kg	≥ 4
大气垂直方向上气流速度	OMG	hPa/s	> 0

大气垂直方向上的运动与云中水汽的凝结密切相关,直接影响云的发展.根据 OMG 来划分 4 d 内云系发展阶段,如图 2a 所示,根据北方大气垂直方向上气流速度均值的变化情况,以一个周期运动为一个阶段,将 4 d 内云的整体变化分为 5 个阶段.分别是:第 1 阶段:0~27;第 2 阶段:28~38;第 3 阶段:39~76;第 4 阶段:77~86;第 5 阶段:87~96.如图 2b 所示,根据南方大气垂直方向上气流速度均值的整体变化情况,将 4 d 内云的变化分为 6 个阶段.分别是:第 1 阶段:0~24;第 2 阶段:25~47;第 3 阶段:48~62;第 4 阶段:63~74;第 5 阶段:75~85;第 6 阶段:86~96.

1.2 CWC 信息熵的概况

为了研究云系如何组织和发展,信息熵是一个重要的工具.熵的本质内涵是变化和不确定度,熵越小,其有序度越高,反之则混乱程度越高^[6].为了研究云水含量 CWC 的混沌程度,本文通过其信息熵值来进行接下来的分解与预测研究.

为了反映云系状况,针对每个时刻 CWC 超阈值的空间分布数据,计算信息熵值,连接所有时刻的熵

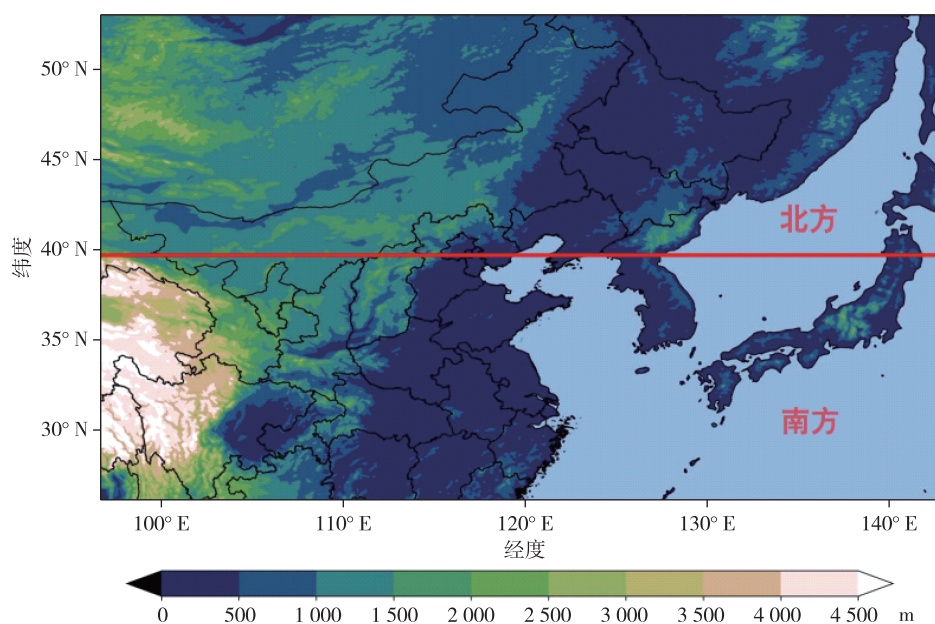


图 1 研究区域

Fig. 1 Study area

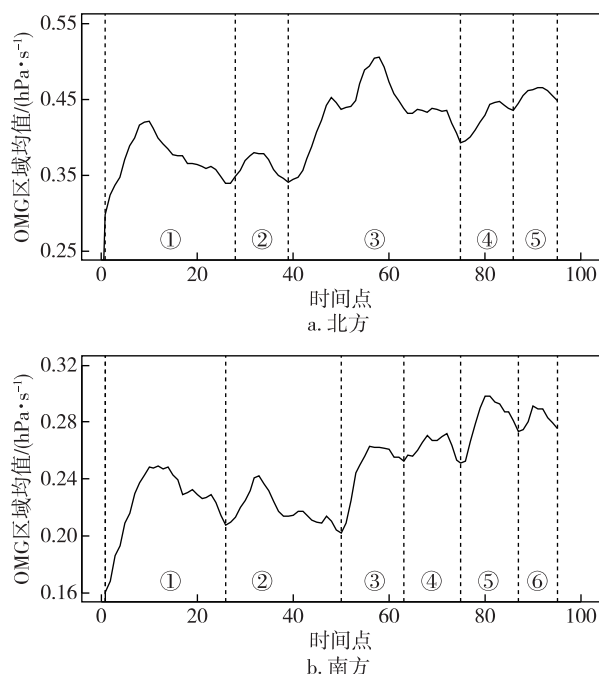


图2 2020-09-30—10-03 北方和南方的 OMC

Fig. 2 OMC values in the north (a) and the south (b) from Sept. 30 to Oct. 3 of 2020

值绘制时序图(图3).南方 CWC 熵整体上大于北方,说明南方 CWC 空间分布数据的不确定度更大,并且周期性较北方明显.通过箱线图(图4)对比南、北方 CWC 熵的离散情况以及异常值,南方异常值数据较多且南分布较为集中.

进行 ADF (Augmented Dickey-Fuller) 平稳性检验,得到北方数据的 P 值为 0.58,布尔判定结果为 $h=0$,南方数据的 P 值为 0.81,布尔判定结果为 $h=0$,说明南、北方 CWC 熵序列均是非平稳的.进一步,进行 BDS 检验(表2),最大嵌入维数设为 6,BDS 统计量的值随着嵌入维度的增加而增加,且 P 值都小

于 0.01,说明南、北方 CWC 熵都有显著的非线性.

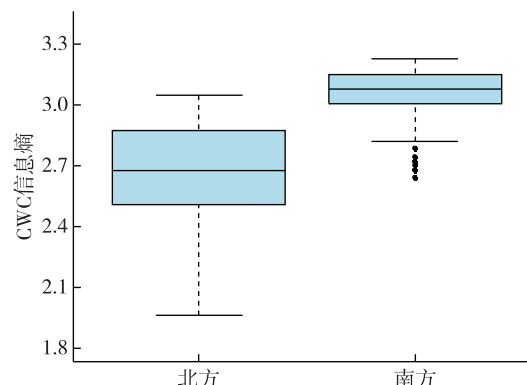


图4 CWC 信息熵箱线

Fig. 4 Boxplot of CWC information entropy

表2 CWC 信息熵的 BDS 检验

Table 2 BDS test of CWC information entropy

维度	BDS 统计量		标准误差		Z 统计量		P 值	
	北方	南方	北方	南方	北方	南方	北方	南方
2	0.12	0.16	0.01	0.01	16.91	16.17	<0.01	<0.01
3	0.20	0.27	0.01	0.02	17.59	16.93	<0.01	<0.01
4	0.24	0.34	0.01	0.02	18.10	17.92	<0.01	<0.01
5	0.27	0.38	0.01	0.02	19.23	19.12	<0.01	<0.01
6	0.28	0.40	0.01	0.02	20.71	20.59	<0.01	<0.01

依据图2划分的云系发展阶段,绘制箱线图来对比南、北方云在不同发展阶段下的差异.由图5可知,北方的中位数呈先下降后上升趋势,南方先升后降又升.随着云系的发展,南方离散程度持续下降,而北方变化不大.

由表3可知,在云系的不同发展阶段,北方 CWC 信息熵均值都小于南方,方差普遍大于南方.北方最大均值出现在第5阶段,为 2.92,最大方差在第

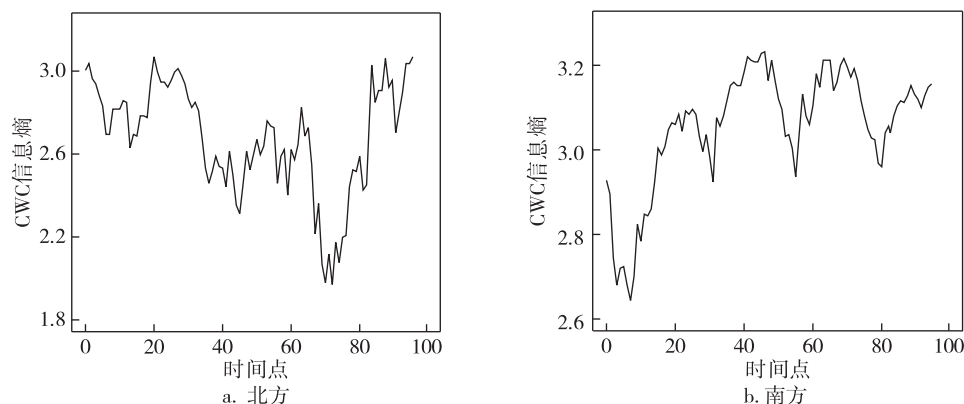


图3 2020-09-30—10-03 北方和南方的 CWC 信息熵

Fig. 3 CWC information entropy in the north (a) and the south (b) from Sept. 30 to Oct. 3 of 2020

3 阶段,为 0.05;南方最大均值在第 4 阶段,为 3.18,最大方差在第 1 阶段,为 0.02.

表 3 不同发展阶段下 CWC 信息熵的均值和方差
Table 3 Average and variance values of CWC information entropy at different stages

阶段	北方			南方		
	时间点	均值	方差	时间点	均值	方差
第 1 阶段	0~27	2.85	0.01	0~24	2.88	0.02
第 2 阶段	28~38	2.71	0.03	25~47	3.12	0.01
第 3 阶段	39~76	2.45	0.05	48~62	3.09	0.01
第 4 阶段	77~86	2.62	0.04	63~74	3.18	<0.01
第 5 阶段	87~96	2.92	0.01	75~85	3.04	0.01
第 6 阶段				86~96	3.13	<0.01

2 CWC 信息熵的多尺度分解

CWC 信息熵是非线性、非平稳的时间序列,且云系中存在复杂的物理变化过程.为了研究时间序列的内部信息与序列结构,通常将 CWC 熵在一组基上展开,进行多尺度分解.不同的多尺度分解方法有不同的参数设置和层数,本文综合考虑不同分解方法的优缺点,选择小波分解^[18]和经验模态(Empirical Mode Decomposition, EMD)分解.小波分解保留信号时域上的特征和频域上的分辨率,但是基函数以及分解层数需要自行设置.EMD 分解不需要提前设置基函数和分解层数,具有很好的自适应能力,但存在边缘效应.因此,本文综合运用了小波分解和 EMD 分解两种方法.

2.1 小波分解与重构

选择 sym4 作为小波分解的基函数,分解层数设

为 6 层,结果如图 6,其中, $S_1, S_2, \dots, S_5$ 表示 CWC 信息熵分解后的分量,RES 为趋势序列,趋势序列反映整体趋势.虽然将 CWC 信息熵值分解成平稳有规律的序列,但是分解的层数太多,单独研究每个序列过程较为繁琐,导致很难给出所有分量的物理解释,并且会加大后期预测的难度,故需要重构分解后的序列.将 S_1, S_2, S_3 3 个部分叠加作为高频项 a_1 ,代表 CWC 信息熵的较大波动, S_4, S_5 叠加作为低频项 a_2 ,代表 CWC 信息熵的较小波动,RES 为分解后的趋势序列,重构结果如图 7 所示.

计算重构序列的 3 个辅助指标.方差贡献率是重构后序列的方差与原序列方差之比,衡量分量对原序列的贡献率.平均周期定义为各个分量样本数与极大值或者极小值点个数之比,用来表征序列的周期长短,平均周期的单位是点数.相关系数则是描述各个分量与原序列的线性相关程度.计算结果如表 4 所示,无论南方还是北方,3 个辅助指标均为趋势项>低频项>高频项.

表 4 CWC 信息熵的小波重构辅助指标值
Table 4 Auxiliary index value of wavelet reconstruction of CWC information entropy

辅助指标	方差贡献率		平均周期		相关系数	
	南方	北方	南方	北方	南方	北方
高频项	0.12	0.20	12	14	0.35 *	0.45 *
低频项	0.20	0.30	24	32	0.56 *	0.60 *
趋势项	0.59	0.45	96	96	0.83 *	0.71 *

注: * 表示通过了显著性水平 $\alpha=0.05$ 的显著性检验.

此外,OMG 区域均值能够在一定程度上反映云的发展阶段,例如:出现极大值点往往表示云系处于

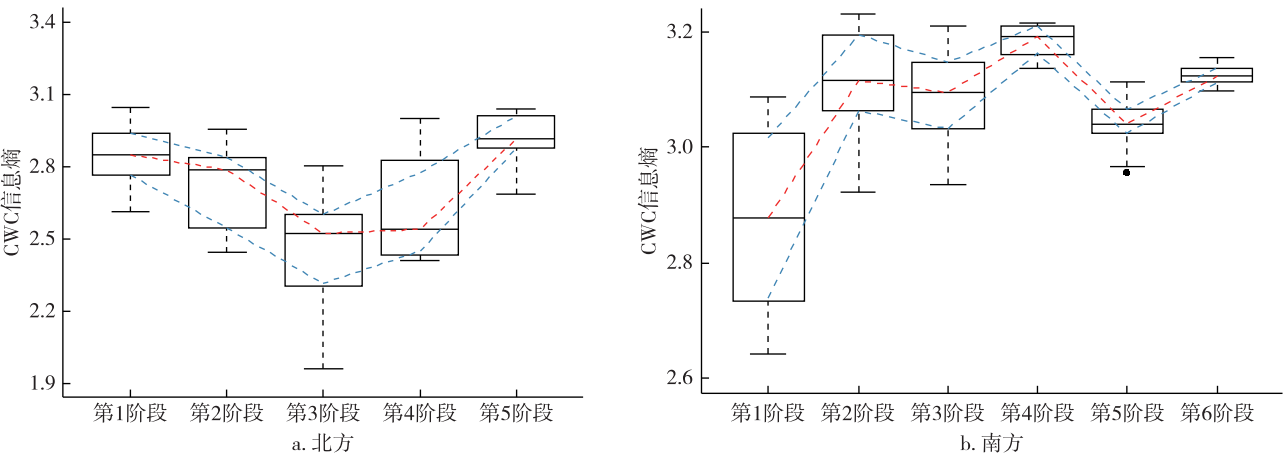


图 5 不同阶段下北方和南方 CWC 信息熵的箱线图

Fig. 5 Boxplots of CWC information entropy in the north (a) and the south (b) at different stages

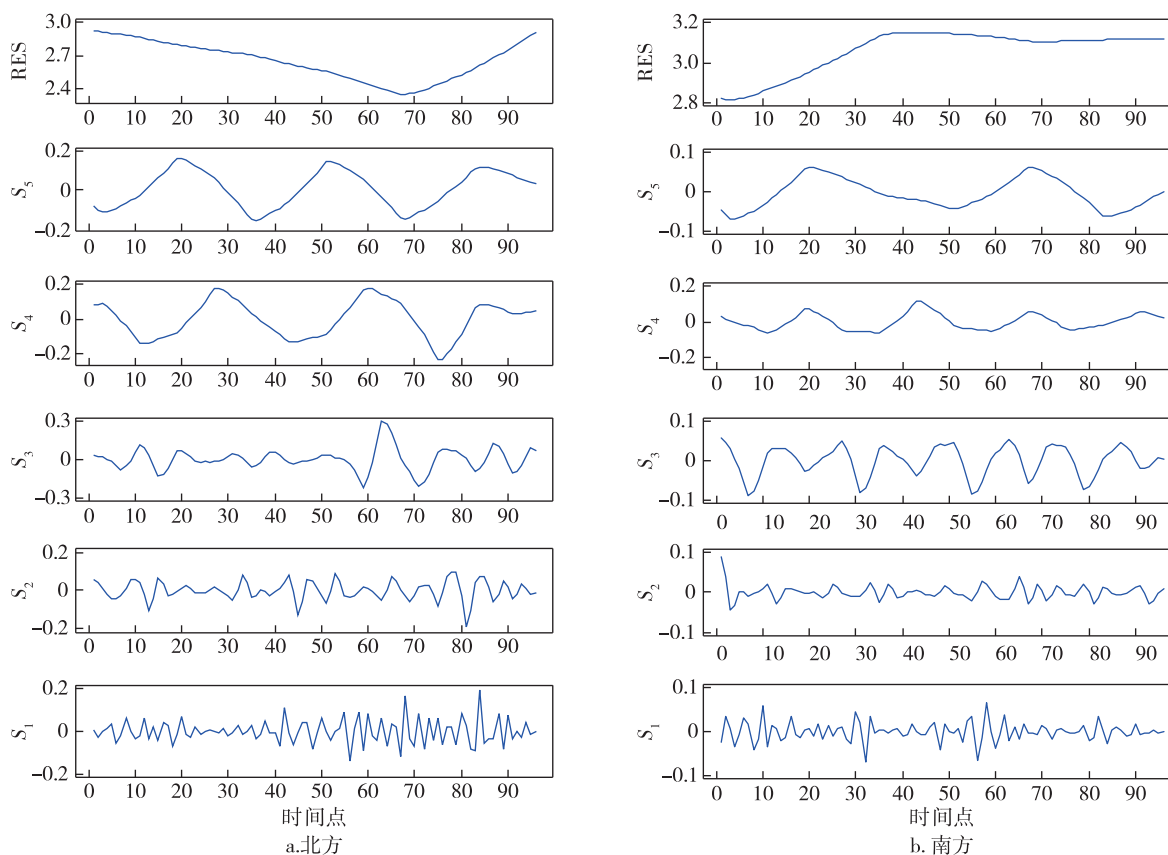


图6 北方和南方 CWC 信息熵的小波分解

Fig. 6 Wavelet decomposition of CWC information entropy in the north (a) and the south (b)

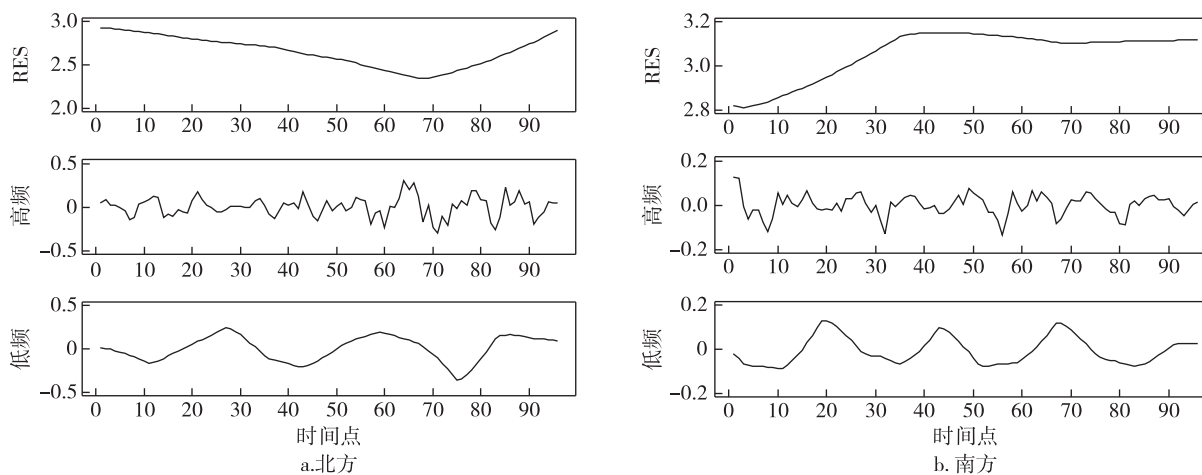


图7 北方和南方 CWC 信息熵的小波重构

Fig. 7 Wavelet reconstruction of CWC information entropy in the north (a) and the south (b)

发展阶段,出现极小值点表示云系处于消散阶段. CWC 熵值重构低频项的极值如果也能出现相应的信号,说明利用 CWC 熵值开展研究是具有信息挖掘价值的,有助于判断云系发展阶段.OMG 区域均值与 CWC 信息熵小波分解的重构低频项极值点相近,甚至有些是重合的,如表 5 所示.相近,这里指时间相

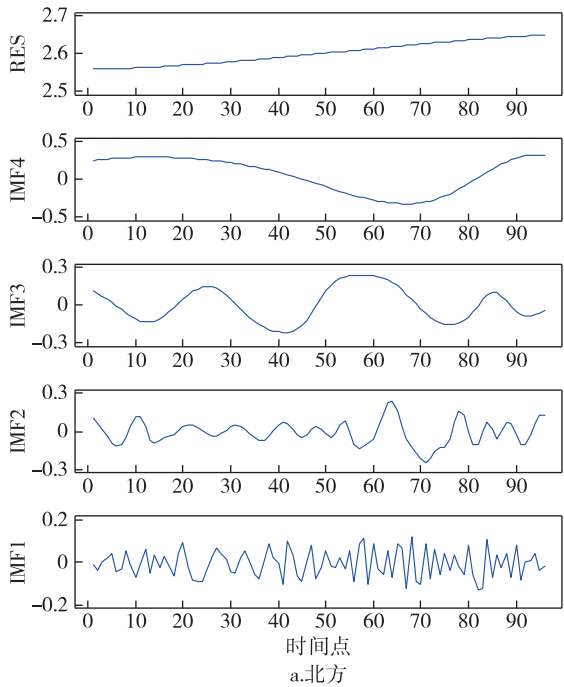
差在 2 个时间点以内.对于 OMG 均值序列,CWC 信息熵的小波分解低频重构的相近极值点为南方占 44.4%,北方占 55.5%;对于 CWC 熵值重构低频项,南方 OMG 均值的极值点的相近占 50%,北方占 83.3%.这里的百分比,反映的是 CWC 熵值重构低频项与 OMG 均值的极值点的信号响应程度.

表 5 OMG 均值与 CWC 信息熵小波重构低频的极值点
Table 5 Extreme points of OMG mean and CWC information entropy low frequency of wavelet reconstruction

指标	极值点出现时间
南方 OMG 均值	17、26、 35 、 51 、62、75、 83 、88、 92
南方 CWC 熵值重构低频项	12、20、 35 、44、 51 、68、 82 、 92
北方 OMG 均值	10 、 27 、34、 40 、52、 60 、 74 、87、93
北方 CWC 熵值重构低频项	12 、 27 、 42 、 60 、 74 、82

2.2 EMD 分解与重构

EMD 分解方法^[19]需要设置停止条件 defstop = [0.05,0.5,0.05],迭代次数为 1 000 次,结果如图 8 所示,其中 IMF1,IMF2,...,IMFn 是分解后的序列,



RES 为分解后的趋势序列。
考虑到分解层数太多不利于预测,将 IMF1 和 IMF2 两个部分叠加作为高频项 a_1 ,IMF3 和 IMF4 叠加作为低频项 a_2 ,RES 为分解后的趋势序列,重构如图 9 所示。

由表 6 可见:1) 方差贡献率:低频项>趋势项>高频项;2) 周期:趋势项>低频项>高频项.相关系数都通过了 $\alpha=0.05$ 的显著性检验,南方 CWC 熵:趋势项>低频项>高频项;北方 CWC 熵:低频项>高频项>趋势项.运用不同分解方法,各个指标方差贡献率和相关系数的计算结果的排序可能存在差异,周期的排序没有发生变化。

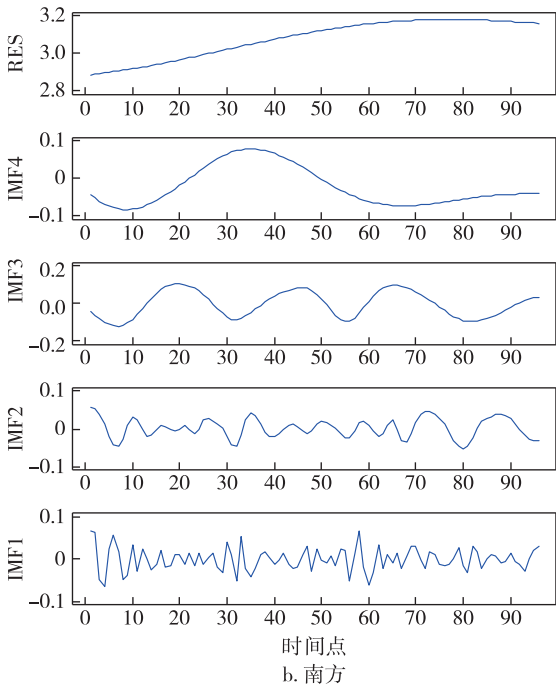


图 8 北方和南方 CWC 信息熵的 EMD 分解

Fig. 8 EMD decomposition of CWC information entropy in the north (a) and the south (b)

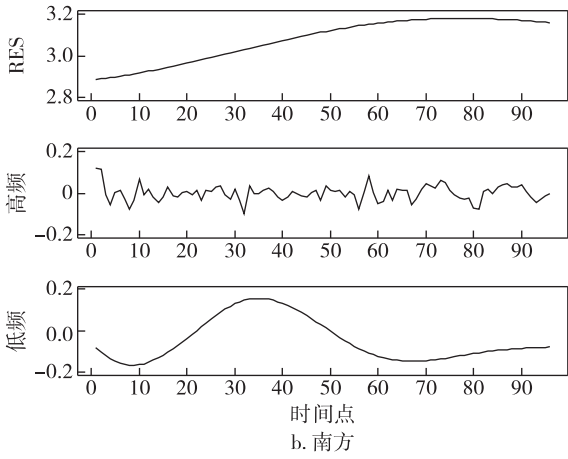
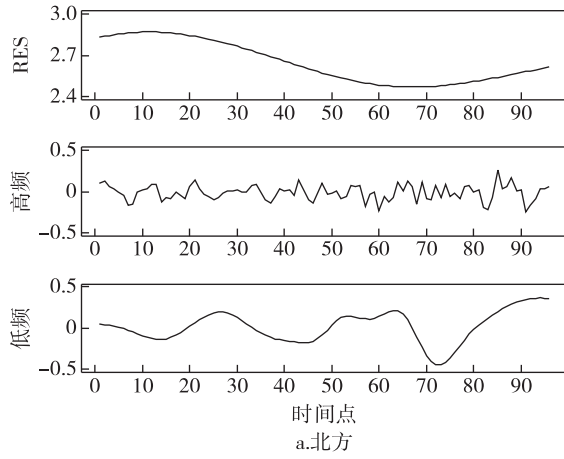


图 9 北方和南方 CWC 信息熵的 EMD 重构

Fig. 9 EMD reconstruction of CWC information entropy in the north (a) and the south (b)

表6 CWC信息熵的EMD重构项辅助指标
Table 6 EMD reconstruction auxiliary index values
of CWC information entropy

辅助指标	方差贡献率		平均周期/h		相关系数	
	南方	北方	南方	北方	南方	北方
高频项	0.07	0.13	8	8	0.25 *	0.58 *
低频项	0.54	0.54	48	32	0.33 *	0.73 *
趋势项	0.51	0.32	96	96	0.71 *	0.39 *

注: * 表示通过了 $\alpha=0.05$ 的显著性检验.

3 CWC信息熵的组合预测

3.1 趋势项的Holt预测

采用Holt预测^[20]分解后的趋势序列,ARIMA模型^[21]预测分解后的高频序列和低频序列,最后借助Lagrange Multiplier法^[22]将结果进行集成.使用Holt平滑方法研究CWC信息熵的趋势项,观察两种分解方法下云水含量信息熵重构后的趋势项,发现存在明显的趋势,因此采用Holt方法进行预测.

设 X_t 为原始时间序列,三个分量分别为 L_t , T_t 和 S_t ,对应的公式为

$$\begin{cases} L_t = a \frac{X_t}{S_{t-m}} + (1-a)(L_{t-1} - T_{t-1}), \\ T_t = b(L_t - L_{t-1}) + (1-b)T_{t-1}, \\ S_t = s \frac{X_t}{L_t - L_{t-1}} + (1-s)S_{t-1}, \end{cases}$$

其中, a, b, s 是平滑系数且 $0 \leq a, b, s \leq 1$. 参数设置如表7所示,趋势序列的拟合效果较好(图10).

表7 Holt指数平滑的参数

重构方式	a	b	s
北方小波重构	0.16	0.25	0.5
北方EMD重构	0.12	0.30	0.5
南方小波重构	0.20	0.50	0.5
北方EMD重构	0.13	0.30	0.5

3.2 高频项和低频项的ARIMA预测

对CWC信息熵分解后的高频项和低频项的时间序列进行Box.test白噪声检验^[23]和单位根平稳性检验^[24].表8显示,南、北方小波重构和EMD重构后的高频项均是平稳的非白噪声序列,故不需要差分;北方小波重构和南方EMD重构的低频项均是非平稳序列,需要进行差分,且两个序列的一阶差分即可满足序列的平稳性.综合考虑两种情况选择最优模型,结果如表9所示.

选择CWC信息熵前66个数据进行训练建模、后30个的数据进行预测,结果如图11和12所示.ARIMA模型对于波动情况预测较好,但是对峰值、谷值等预测精度还有待提高,低频项的训练效果明显优于高频项.

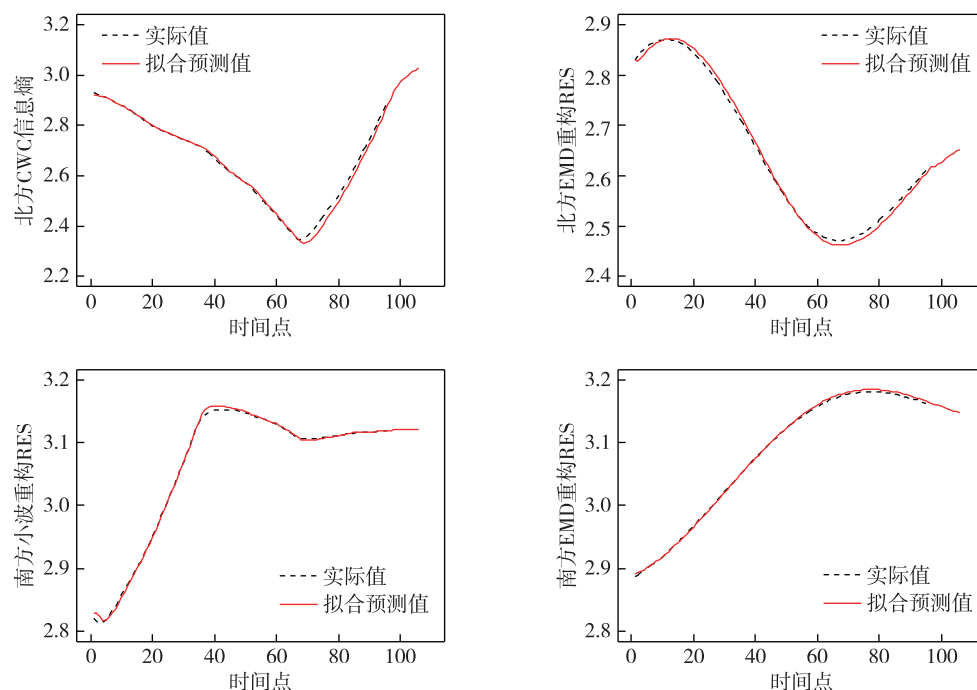


图10 趋势项Holt预测

Fig. 10 Holt forecast results of trends

表 8 高频项/低频项的平稳性和白噪声检验

Table 8 Stationarity test and white noise test of high- and low-frequency sequences

指标	北方				南方			
	小波重构		EMD 重构		小波重构		EMD 重构	
	高频	低频	高频	低频	高频	低频	高频	低频
ADF 单位根	<0.05	0.08	<0.05	0.02	<0.05	0.02	<0.05	0.38
Box. test 白噪声	<0.05	<0.05	<0.05	<0.05	<0.05	<0.05	<0.05	<0.05

表 9 高频项和低频项的预测模型

Table 9 Prediction models for high- and low-frequency sequences

趋势项	小波重构	EMD 重构
北方高频	AR(1)	ARMA(4,3)
南方高频	MA(2)	MA(1)
北方低频	ARIMA(4,1,0)	ARMA(4,2)
南方低频	AR(4)	ARIMA(1,1,0)

3.3 Lagrange Multiplier 组合预测

根据上文建立的 Holt 和 ARIMA 的单项预测,基于 Lagrange Multiplier 法建立组合预测模型. X_t 为 CWC 信息熵实际值, $\hat{X}_{1t}$ 为基于小波分解的序列预测值, $\hat{X}_{2t}$ 为基于 EMD 分解的序列预测值, $t = 1, \dots, n$, 则有 $\hat{X}_t = \lambda_1 \hat{X}_{1t} + \lambda_2 \hat{X}_{2t}$, 误差为 $e_t = X_t - \hat{X}_t$, 满足:

$$\begin{aligned} \min z &= \sum_{t=1}^n e_t^2, \\ \begin{cases} \lambda_1 + \lambda_2 = 1, \\ 0 \leq \lambda_1, \lambda_2 \leq 1. \end{cases} \end{aligned} \tag{1}$$

基于 Lagrange Multiplier 法,确定组合预测模型的最优权重:

$$\lambda_1 = \frac{\sum_{t=1}^n e_{2t}^2 - \sum_{t=1}^n e_{1t}e_{2t}}{\sum_{t=1}^n e_{1t}^2 + \sum_{t=1}^n e_{2t}^2 - 2 \sum_{t=1}^n e_{1t}e_{2t}}, \tag{2}$$

$$\lambda_2 = \frac{\sum_{t=1}^n e_{1t}^2 - \sum_{t=1}^n e_{1t}e_{2t}}{\sum_{t=1}^n e_{1t}^2 + \sum_{t=1}^n e_{2t}^2 - 2 \sum_{t=1}^n e_{1t}e_{2t}}, \tag{3}$$

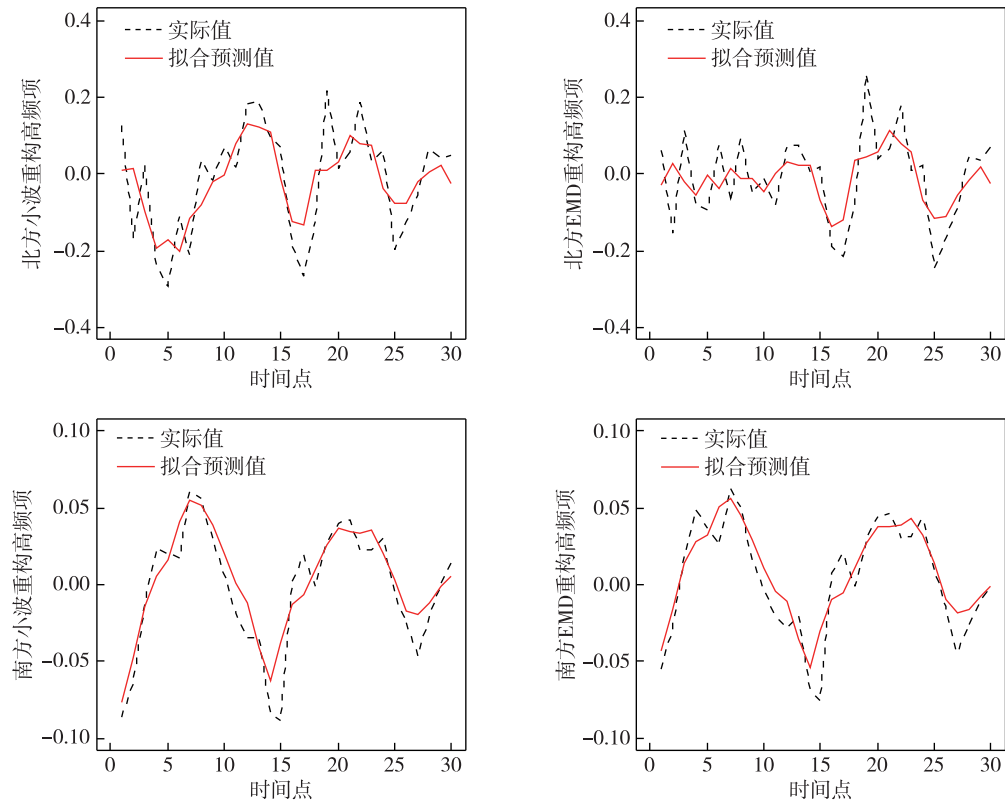


图 11 高频项 ARIMA 预测

Fig. 11 ARIMA forecast results of high-frequency sequences

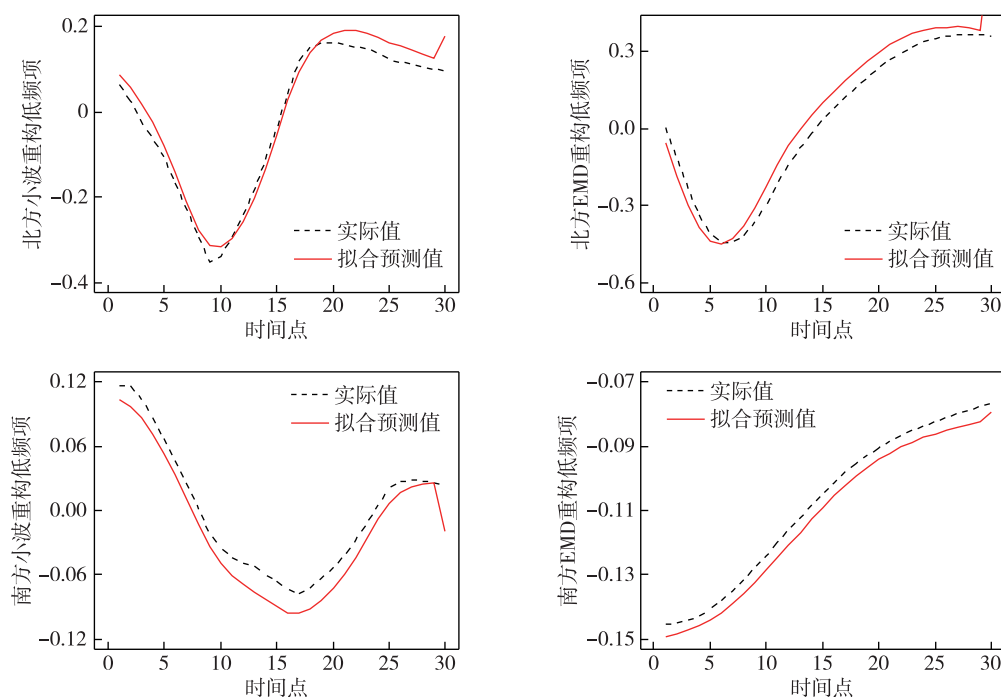


图 12 低频项 ARIMA 预测

Fig. 12 ARIMA forecast results of low-frequency sequences

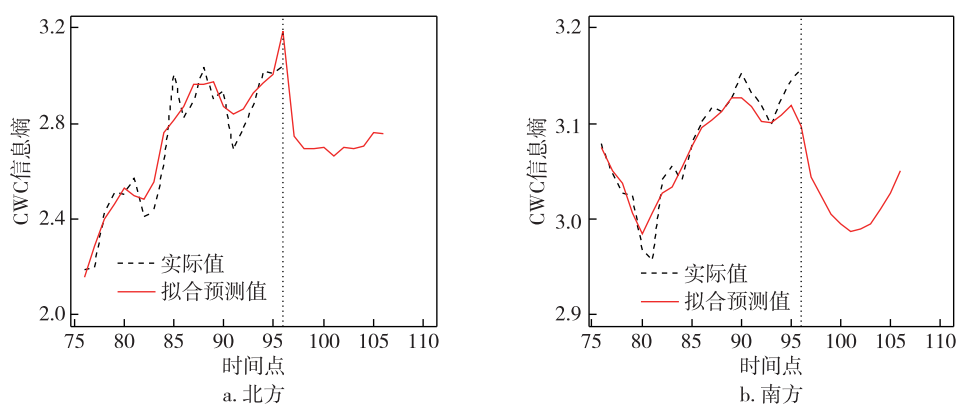


图 13 CWC 信息熵的组合预测结果

Fig. 13 Combined prediction results of the CWC information entropy

其中, $e_{1t} = X_t - \hat{X}_{1t}$, $e_{2t} = X_t - \hat{X}_{2t}$.

计算 Lagrange Multiplier 方法的权重, 北方 CWC 信息熵的权重为 $\lambda_1 = 0.482$ 和 $\lambda_2 = 0.518$, 南方的为 $\lambda_1 = 0.644$ 和 $\lambda_2 = 0.356$. 基于前 66 个数据训练建模, 对后 30 个数据(时间 67~96)的趋势项、高频项和低频项的分量分别进行预测, 将分量预测值叠加, 得到对应分解下的预测值, 并对未来 10 h(时间 97~106)进行预测, 整体预测效果较好(图 13). 本文提出的“分解-重构-组合”的思想是可行的, 将时间序列分解成趋势项、高频项和低频项, 能够有效地提取序列的不同频次的特征, 并获得较好预测效果.

3.4 组合模型性能分析

选取 3 个误差评价指标评估预测模型的性能, 拟合优度 R^2 取值范围为 0~1, 越接近于 1 拟合效果越好, 平均绝对百分比误差 MAPE 和均方根误差 RMSE 的范围都是 $[0, +\infty)$, 越小越好^[25]. 由表 10 可见, 不同分解方法下, 组合预测误差有一定的差异. 北方和南方 CWC 信息熵的组合模型, EMD 分解下的 RMSE 和 R^2 都更优, 小波分解下的 MAPE 更优. 不同的分解方法各有优劣, 故本文选择两种分解方法进行研究.

表 10 误差评价指标
Table 10 Error evaluation index

指标	北方		南方	
	小波分解	EMD 分解	小波分解	EMD 分解
RMSE	0.23	0.22	0.33	0.25
MAPE	0.17	0.37	0.21	0.24
R^2	0.81	0.82	0.90	0.93

此外,对建立的 Holt-ARIMA-Lagrange Multiplier 组合模型进行对比分析.模型 1 未对 CWC 信息熵进行分解,直接采用 ARIMA 模型预测;模型 2 采用 BP 神经网络^[26]预测;模型 3 和模型 4 分别使用一种多尺度分解,然后对分解后的趋势项序列进行 Holt 预测,高频和低频项序列进行 ARIMA 预测,最后将单项预测结果相加,得到最终预测结果;模型 5 为本文提出的组合模型.由表 11 可见,本文建立的组合模型对南北方 CWC 信息熵的预测效果最好.模型 1 对不分解序列进行预测,根据准确率等于 1 减去误差进行计算,只有 52.33%.对比模型 1、3、4,对序列进行多尺度分解后可以提高预测准确率.对比模型 3、4、5,综合两种分解方法的预测效果优于单一分解,模型准确率提高了 3%~4%.

表 11 不同预测模型评价指标
Table 11 Evaluation indexes of different prediction models

预测模型	R^2	MAPE	RMSE
1: ARIMA	0.52	0.34	0.31
2: BP 神经网络	0.34	0.37	0.43
3: 小波分解-Holt-ARIMA	0.81	0.27	0.25
4: EMD 分解-Holt-ARIMA	0.82	0.27	0.22
5: Holt-ARIMA-Lagrange Multiplier *	0.85	0.21	0.18

4 结论

在降水云系的发展过程中,700 hPa 云水含量是表征云的生成、发展、消散等阶段的重要云物理量.本文提出混合多尺度分解的 Holt-ARIMA-Lagrange Multiplier 组合模型,对 2020 年 9 月 30 日—10 月 3 日一次降水云系发展过程中的 CWC 空间分布信息熵的 1 h 值进行时间序列研究,不仅能够定量衡量云发展过程中的 CWC 的混沌程度及其变化规律,也能够对于云系的自组织状况进行初步分析.

1) 计算 CWC 信息熵的 1 h 值,发现,以 40°N 为界,北方和南方的标准差分别为 0.26 和 0.13,偏度分别为 -0.67 和 -1.23,峰度分别是 2.95 和 4.北方波动性较大,南方更加左偏且分布更加陡峭.在时间

序列特征上,南北方的 BDS 统计量 P 值均小于 0.01,ADF 统计量的 P 值均大于 0.05,说明南北方 CWC 信息熵序列均是非线性且非平稳的.

2) 在云的不同发展阶段,CWC 信息熵的特征各不相同.北方的中位数先降后升,南方先升后降又升.随着云系的发展,南方离散程度持续下降,而北方变化不大.北方方差最大为 0.05,均值最大为 2.92;南方方差最大为 0.02,均值最大为 3.18.

3) OMG 均值与 CWC 信息熵的小波低频重构的极值点存在着一定的对应关系,设定 ± 2 h 为时间相近的范围,则相近的极值点在南方云中占 50%,在北方云中占 83.3%,说明 CWC 熵可以在一定程度上反映云系的发展.

4) 综合运用小波分解和经验模态分解,并重构 CWC 信息熵的分量,基于 ARIMA 模型预测重构后的高频项和低频项、基于 Holt 模型预测趋势项,最后基于 Lagrange Multiplier 法集成 CWC 信息熵的预测值,建立 Holt-ARIMA-Lagrange Multiplier 组合模型.通过实证分析,本文提出的组合模型比单一预测方法、单层分解的预测模型的准确率提高了 3%以上.

参考文献

References

- [1] 段海霞,李耀辉,张强,等.西北区域几次暴雨过程中的自组织现象[J].高原气象,2011,30(4):890-900
DUAN Haixia, LI Yaohui, ZHANG Qiang, et al. Phenomenon of vortex self-organization of several rainstorm processes in northwest region of China[J]. Plateau Meteorology, 2011, 30(4): 890-900
- [2] 张小娟,陶玥,刘国强,等.一次冰雹天气过程的云系发展演变及云物理特征研究[J].气象,2019,45(3):415-425
ZHANG Xiaojuan, TAO Yue, LIU Guoqiang, et al. Study on the evolution of hailstorm and its cloud physical characteristics[J]. Meteorological Monthly, 2019, 45(3): 415-425
- [3] 陈逸伦.基于多源卫星数据的云团和雨团识别及其特征研究[D].合肥:中国科学技术大学,2019
CHEN Yilun. Identification of cloud clusters and rain cells and their features from multi-satellite observations[D]. Hefei: University of Science and Technology of China, 2019
- [4] 程敬雅,游庆龙,蔡森.全球云水量气候分布及变化趋势特征分析[J].气候与环境研究,2021,26(5):541-555
CHENG Jingya, YOU Qinglong, CAI Miao. Climatic distribution and trend characteristics of global cloud water content[J]. Climatic and Environmental Research, 2021, 26(5): 541-555
- [5] 刘菊菊,游庆龙,周毓荃,等.基于 ERA-Interim 的中国

- 云水量时空分布和变化趋势[J].高原气象,2018,37(6):1590-1604
- LIU Juju, YOU Qinglong, ZHOU Yuquan, et al. Spatio-temporal distribution and trend of cloud water content in China based on ERA-Interim reanalysis[J]. Plateau Meteorology, 2018, 37(6): 1590-1604
- [6] 张学文, 马力. 熵气象学简介[J]. 气象, 1995, 21(1): 52-56
- ZHANG Xuwen, MA Li. Entropy meteorology[J]. Meteorological Monthly, 1995, 21(1): 52-56
- [7] 王惠娟. 基于信息熵的大气 $PM_{2.5}$ 浓度的不确定性研究[D]. 武汉: 武汉理工大学, 2016
- WANG Huijuan. The research of $PM_{2.5}$ atmospheric concentration uncertainty based on information entropy[D]. Wuhan: Wuhan University of Technology, 2016
- [8] Berta M, Wehner S, Wilde M M. Entropic uncertainty and measurement reversibility[J]. New Journal of Physics, 2016, 18(7): 073004
- [9] 李丽娟, 明飞, 宋学科, 等. 熵不确定度关系综述[J]. 物理学报, 2022, 71(7): 25-41
- LI Lijuan, MING Fei, SONG Xueke, et al. Review on entropic uncertainty relations[J]. Acta Physica Sinica, 2022, 71(7): 25-41
- [10] 车金星. 复杂数据的变量选择与预测方法研究[D]. 西安: 西安电子科技大学, 2019
- CHE Jinxing. Variable selection and forecasting method for complex data[D]. Xi'an: Xidian University, 2019
- [11] Geetha A, Nasira G M. Time series modeling and forecasting: tropical cyclone prediction using ARIMA model[C]//2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom). New Delhi, India. IEEE, 2016: 3080-3086
- [12] Xiong T, Li C G, Bao Y K. Interval-valued time series forecasting using a novel hybrid Holt^I and MSVR model[J]. Economic Modelling, 2017, 60: 11-23
- [13] 汪漂. 混合区间多尺度分解的区间时间序列组合预测[J]. 运筹与管理, 2021, 30(10): 159-164
- WANG Piao. An interval time series combination forecasting approach based on hybrid interval multi-scale decomposition[J]. Operations Research and Management Science, 2021, 30(10): 159-164
- [14] Wu Q L, Lin H X. A novel optimal-hybrid model for daily air quality index prediction considering air pollutant factors[J]. Science of the Total Environment, 2019, 683: 808-821
- [15] Monjoly S, André M, Calif R, et al. Hourly forecasting of global solar radiation based on multiscale decomposition methods: a hybrid approach[J]. Energy, 2017, 119: 288-298
- [16] 李栋, 薛惠锋, 张燕. 基于经验模态分解的降水量组合预测模型[J]. 计算机仿真, 2019, 36(3): 458-463
- LI Dong, XUE Huifeng, ZHANG Yan. Combined forecasting model of precipitation based on empirical mode decomposition[J]. Computer Simulation, 2019, 36(3): 458-463
- [17] Xiang Y, Gou L, He L H, et al. A SVR-ANN combined model based on ensemble EMD for rainfall prediction[J]. Applied Soft Computing, 2018, 73: 874-883
- [18] 庞金凤, 刘波, 张波, 等. 基于小波分解的沙尘天气发生日数预测组合模型研究: 以 2008—2016 年策勒沙漠-绿洲过渡带沙尘天气发生时序为例[J]. 气象, 2019, 45(5): 651-658
- PANG Jinfeng, LIU Bo, ZHANG Bo, et al. Study on the combined model of forecasting the days of sand-dust weather based on wavelet decomposition; taking the time series of dust weather in the transitional zone of Qira desert-oasis during 2008-2016 as an example[J]. Meteorological Monthly, 2019, 45(5): 651-658
- [19] Chen Q C, Wen D, Li X Q, et al. Empirical mode decomposition based long short-term memory neural network forecasting model for the short-term metro passenger flow[J]. PLoS One, 2019, 14(9): e0222365
- [20] 杨国华, 郑豪丰, 张鸿皓, 等. 基于 Holt-Winters 指数平滑和时间卷积网络的短期负荷预测[J]. 电力系统自动化, 2022, 46(6): 73-82
- YANG Guohua, ZHENG Haofeng, ZHANG Honghao, et al. Short-term load forecasting based on Holt-Winters exponential smoothing and temporal convolutional network[J]. Automation of Electric Power Systems, 2022, 46(6): 73-82
- [21] 刘金培, 汪漂, 黄燕燕, 等. 基于区间时间序列小波多尺度分解的组合预测方法[J]. 统计与决策, 2020, 36(19): 5-9
- LIU Jinpei, WANG Piao, HUANG Yanyan, et al. Combined forecasting method based on interval time series wavelet multi-scale decomposition[J]. Statistics & Decision, 2020, 36(19): 5-9
- [22] Yang B, Xiang X Q, Kong W Z, et al. Adaptive multi-task learning using Lagrange multiplier for automatic art analysis[J]. Multimedia Tools and Applications, 2022, 81(3): 3715-3733
- [23] 李为东, 李莉, 徐岩. 基于时间序列分析的北京地区 $PM_{2.5}$ 浓度研究[J]. 运筹学学报, 2018, 22(2): 115-126
- LI Weidong, LI Li, XU Yan. The concentration research of $PM_{2.5}$ in Beijing with time series analysis[J]. Operations Research Transactions, 2018, 22(2): 115-126
- [24] 左秀霞. 带高次趋势项的 ADF 单位根检验[J]. 数量经济技术经济研究, 2019, 36(1): 152-169
- ZUO Xiuxia. ADF unit root test with high order trend term[J]. The Journal of Quantitative & Technical Economics, 2019, 36(1): 152-169
- [25] 丁严, 许德合, 曹连海, 等. 基于 CEEMD 的 LSTM 和 ARIMA 模型干旱预测适用性研究: 以新疆为例[J]. 干旱区研究, 2022, 39(3): 734-744
- DING Yan, XU Dehe, CAO Lianhai, et al. Applicability of the LSTM and ARIMA model in drought prediction based on CEEMD: a case study of Xinjiang[J]. Arid Zone Research, 2022, 39(3): 734-744
- [26] Li X N, Cheng X, Wu W J, et al. Forecasting of bioaerosol concentration by a back propagation neural network model[J]. Science of the Total Environment, 2020, 698: 134315

Information entropy time series of CWC based on Holt-ARIMA-Lagrange Multiplier

ZHANG Xian¹ WU Qiong² CHEN Yiqi² LI Yashao² WANG Weiwei³

1 Meteorological Center of Air Traffic Regulation of Civil Aviation in North China, Beijing 100621

2 School of Mathematics and Statistics, Nanjing University of Information Science & Technology, Nanjing 210044

3 Experimental Teaching Center for Meteorology and Environment, Nanjing University of
Information Science & Technology, Nanjing 210044

Abstract The development process and characteristic analysis of precipitation cloud system is an important issue in the field of cloud precipitation physics. Here, the 700 hPa Cloud Water Content (CWC) and the 1h value of air-flow velocity (ω , OMG) in the vertical direction of the atmosphere are used to measure the chaos degree of CWC distribution via the information entropy and judge the cloud development via OMG time series, hence a combined prediction model is proposed based on hybrid multi-scale decomposition, Holt model, Autoregressive Integrated Moving Average model (ARIMA) and Lagrange Multiplier. The results show that, the CWC entropy has nonlinear and non-stationary characteristics; the clouds over the north have smaller means of the CWC entropy sequence and larger variance compared with those over the south regardless of the cloud development stage; a good temporal corresponding relationship is found between the regional average OMG and the extreme point reconstructed by the wavelet low-frequency of the CWC entropy, and close extreme value points account for 50% in clouds over the south and 83.3% in clouds over the north, showing that CWC entropy can somehow reflect the cloud development; the multiple timescale features of CWC entropy sequences make the multi-scale decomposed Holt-ARIMA-Lagrange Multiplier model more accurate than the single prediction method and single-layer decomposed prediction model, with accuracy improvement of more than 3%.

Key words cloud water content (CWC); information entropy; wavelet decomposition; empirical mode decomposition (EMD); autoregressive integrated moving average (ARIMA); Holt two-parameter exponential smoothing