

引用格式: ZHOU Houyu, DONG Qing, MENG Deli, *et al.* Reconstruction of remote sensing cloud cover over Tibetan Plateau based on MODIS data[J]. Remote Sensing Technology and Application, 2023, 38(5): 1136-1147. [周厚瑀, 董庆, 孟德利, 等. 基于 MODIS 数据的青藏高原遥感云量重构[J]. 遥感技术与应用, 2023, 38(5): 1136-1147.] DOI: 10.11873/j.issn.1004-0323.2023.5.1136

基于 MODIS 数据的青藏高原遥感云量重构

周厚瑀^{1,2}, 董庆^{1,3}, 孟德利^{1,2}, 赵文博^{1,2}, 边民^{1,2}

(1. 中国科学院空天信息创新研究院 数字地球重点实验室, 北京 100094;

2. 中国科学院大学, 北京 100049;

3. 中科卫星应用德清研究院 浙江省微波目标特性测量与遥感重点实验室, 浙江 湖州 313200)

摘要: 云是影响青藏高原能量平衡和地气过程的关键参量之一, 研究高原的云量对探讨青藏高原的气候变化具有重要意义。采用 2001~2020 年的 MODIS 总云量数据与 ERA5、CRA40 两套再分析资料的总云量数据, 以 3~11 月份 MOD06 云量资料为真值, 评价不同再分析资料在青藏高原的适用性。利用改进自编码器模型, 基于 ERA5 和 MOD06 重构了 1950~2020 年的 3~11 月的高原云量。结果表明: 在高原地区, ERA5 的云量值偏高, 而 CRA40 云量值偏低, 且 ERA5 与 MOD06 的相关性明显优于 CRA40 与 MOD06 的相关性; 通过相关系数、偏差、平均绝对误差、均方根误差四种评价指标, 发现改进自编码器模型在重构云量方面具有良好的效果, 且能模拟出青藏高原云量的变化趋势, 为研究青藏高原地区云量时空演变规律提供了可靠的长时序数据。

关键词: 青藏高原; 云量; MODIS; 再分析资料; 数据重构; 自编码器

中图分类号: TP79 **文献标志码:** A **文章编号:** 1004-0323(2023)05-1136-12

1 引言

青藏高原被称为“世界屋脊”, 作为世界的第三极, 其对气候变化的响应十分敏感^[1]。近 50 a 来, 青藏高原平均气候变暖的速率相对全球气候变暖速率更快^[2-3]。在此背景下, 青藏高原水循环过程及其相关要素发生了剧烈变化^[4]。云是地气系统的重要组成部分, 直接或间接的调节辐射通量、热平衡、水分循环等气候要素, 从而影响局部气候变化^[5]。在多种云参数中, 云量能更直观地反映气候变化特征^[6]。青藏高原独特的动力和热力作用, 致使高原云在很多方面具有高原气候特征, 因此分析青藏高原云量时空变化, 可以加深对高原复杂的气候系统的认识^[7]。

青藏高原地区多年平均云量表现为由东南向西北减少的基本空间格局^[8]。Bao 等^[9]利用 MODIS

云量产品, 研究了青藏高原 2003~2015 年的云量, 发现高原西部云量春季最高, 秋季最低, 高原东部云量春季和夏季较高。Ma 等^[10]利用 CERES 资料, 分析了青藏高原 2001~2019 年云量的特征和变化, 发现青藏高原地区以中高云为主, 且与总云量具有相似的季节变化和年周期。陈少勇等^[11]利用 1961~2000 年青藏高原 80 个测站的总云量资料, 发现青藏高原的总云量的稳定性夏季高于冬季, 东部高于西部。

目前, 青藏高原地区云量主要存在 3 种资料, 即卫星产品^[12]、再分析资料^[13]和气象站点观测数据^[14]。由于青藏高原地形复杂、环境恶劣, 气象站点稀少, 利用站点资料对整个青藏高原的云参数进行分析, 无法满足青藏高原整体性的研究^[15]。再分析资料如 ERA5、CRA40、JRA55 等虽时间序列较

收稿日期: 2022-06-30; 修订日期: 2023-06-25

基金项目: 国家自然科学基金项目“基于遥感数据的印太暖池区近 30a 海洋环境参数变化研究”(41876210)。

作者简介: 周厚瑀(1999—), 男, 江苏宿迁人, 硕士研究生, 主要从事气象遥感研究。E-mail: zhouhouyu20@mails.ucas.ac.cn

通讯作者: 董庆(1965—), 男, 河南信阳人, 研究员, 博士生导师, 主要从事遥感地学应用研究。E-mail: dongqing@aircas.ac.cn

长,但数据的质量和精度不能满足青藏高原地区云量精准研究的需要^[16]。而卫星云量资料具有覆盖范围广、时间连续性好、限制条件较少等优点^[17]。Kotarba等^[18]对比分析了MODIS云量和地面观测资料,证明在温带MODIS云量比其他卫星云量更接近真实值。刘健^[19]对比分析了青藏高原地区Patmos-X、MODIS、CLARA-A2卫星云量产品,发现MODIS具有更多的观测通道,且在高原复杂下垫面条件下,其稳定的仪器轨道和定标反映在长时间序列云量数据具有较好的均一性。

当前,对于青藏高原地区云量的研究虽然比较广泛,但仍存在一些不足。目前研究青藏高原云量多从单一数据出发,无法同时发挥多源数据的优势。如何基于高空间分辨率的遥感数据(如MODIS)和长时间序列的再分析数据,重构获取具有长时序和高精度的云量数据,成为了一个亟待解决的

难题。针对上述问题,本文对比分析了ERA5、CRA40两套再分析云量资料在青藏高原的适用性,并进一步利用改进的自编码器模型,采用质量较好的再分析资料ERA5与MOD06的数据,重构了青藏高原1950~2020年的3~11月份的云量,分析了高原云量的变化特征。

2 研究区与数据

2.1 研究区概况

青藏高原平均海拔4 000 m以上,南北方向近1 000 km,东西宽约2 500 km,面积约为250万km²,是世界上海拔最高的高原。其境内有一系列高大山脉,从北到南主要有祁连山、昆仑山、唐古拉山、冈底斯山、横断山脉和喜马拉雅山脉等^[20]。因地形复杂,高原内部为西风-季风共同控制区,气候差异很大。如图1所示,本文选取25~40°N,68~105°E区域为青藏高原的范围^[21]。

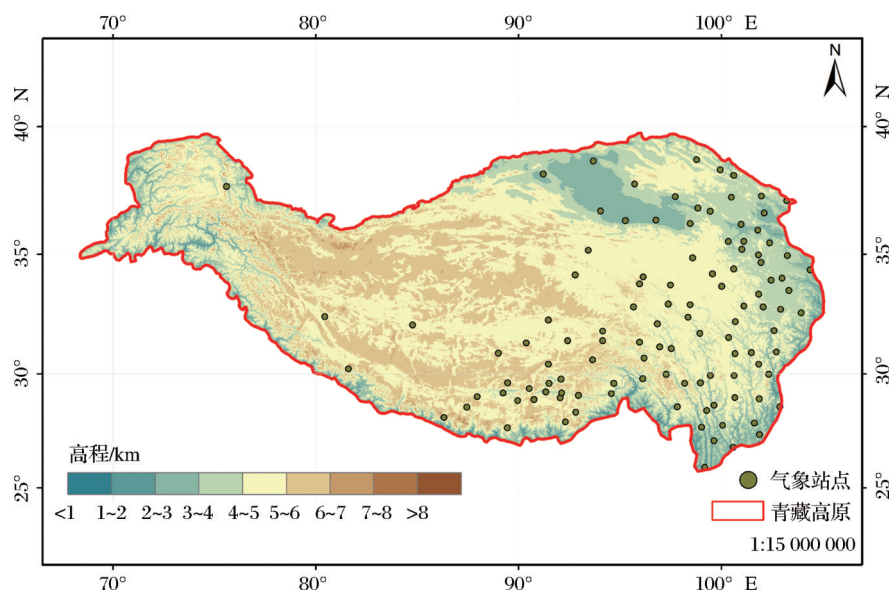


图1 研究区位置与地形

Fig.1 Geographical location and topography of the study area

2.2 数据介绍及预处理

2.2.1 Terra/MODIS(MOD06)

MODIS(Moderate Resolution Imaging Spectroradiometer)是搭载在Terra和Aqua卫星上的中等分辨率成像光谱仪,共有36个波段,光谱范围从0.4~14.4 μm。本文采用MODIS二级云产品,针对Terra和Aqua分别命名为MOD06和MYD06,包含云量、云顶高度、云相态等多种云宏观微观参数,其中云量产品的分辨率为5 km × 5 km^[22-23]。MOD06卫星云量资料具有高空间分辨率、反演精度高的优

势。本文利用MCTK插件对每日过境青藏高原的4~6景MOD06影像进行重投影、镶嵌、裁剪等处理,得到青藏高原上空的逐日云量数据,该云量数据为镶嵌后的Terra/MODIS(MOD06)云覆盖的平均值。

2.2.2 ERA5再分析资料

ERA5是由欧洲中期天气预报中心制作的第五代全球气候再分析产品,时间范围覆盖1950年至今^[24]。ERA5提供大气、陆地和海洋气候变量的产品,分辨率为0.125°,并使用从地面到80公里高度

的 137 个等压面来制作大气资料。这里采用 1950~2020 年的 3~11 月份的月尺度云量数据。

2.2.3 CRA40 再分析资料

CRA40 是由中国气象局生产的综合性的再分析资料,融合了地面气象站点的观测资料,在第三代大气再分析资料中对高原地区对云量的刻画较为可靠。CRA40 提供 1979 年至今的陆面及大气两类再分析产品,具有不同空间分辨率的数据^[25-26]。本文采用 2001~2020 年云量的逐月产品,空间分辨率为 0.5°。

2.2.4 ASTER GDEM

ASTER GDEM (Advanced Spaceborne Thermal Emission and Reflection Radiometer Global Digital Elevation Model) 为 DEM 数据,其全球空间分辨率为 30 m,垂直精度 20 m^[27]。ASTER GDEM 采用 UTM/WGS84 投影,数据覆盖范围为 83°S~83°N 之间的所有陆地区域,达到了地球陆地表面 99%。青藏高原地区 ASTER GDEM 的地形数据为研究青藏高原的云量提供理论支撑。

表 1 所采用的各数据集的基本参数简介
Table 1 Brief introduction of the basic parameters

数据集	数据来源	分辨率	时间长度	方法
MODIS	NASA	1 km	1999 年~至今	卫星反演
ERA5	ECMWF	0.125°	1950 年~至今	4Dvar
CRA40	CMA	0.50°	1979 年~至今	3Dvar
ASTER_GDEM	NASA	30 m	——	——

3 研究方法

研究技术路线如图 2 所示,主要包括以下内容:

(1) 根据表 1 可知,ERA5、MOD06、CRA40 的云量数据空间分辨率存在差异,本文将 MOD06 分别重采样至与 ERA5、CRA40 同一分辨率(0.125°、0.5°),利用气候倾向率等方法对比分析了 2001~2020 年间 ERA5、MOD06、CRA40 3 种云量资料在青藏高原地区的年际变化特征,并计算了 2001~2020 年间 ERA5 与 CRA40 的逐月数据与重采样后的 MOD06 逐月数据的相关系数,从不同时间尺度上评估两种再分析数据在青藏高原的适用性。

(2) 利用 2001~2010 年 3~11 月份的 ERA5 云量资料与 MOD06 云量资料训练本文所提出的改进自编码器模型(IMAE),重构 2011~2020 年的云量。将 2011~2020 年的利用改进自编码器模型重构的

云量通过相关系数(R)、偏差(Bias)、平均绝对误差(Mean Absolute Error, MAE)均方根误差(Mean Squared Error, MSE)与 MOD06 云量产品进行比较,验证模型的可靠性。

(3) 基于 2001~2020 年的 3~11 月份的 ERA5 云量资料与 MOD06 云量资料,通过已经验证完精度较好的改进自编码器模型(Improved-Autoencoder, IMAE),重构 1950~2020 年 3~11 月份青藏高原的云量。

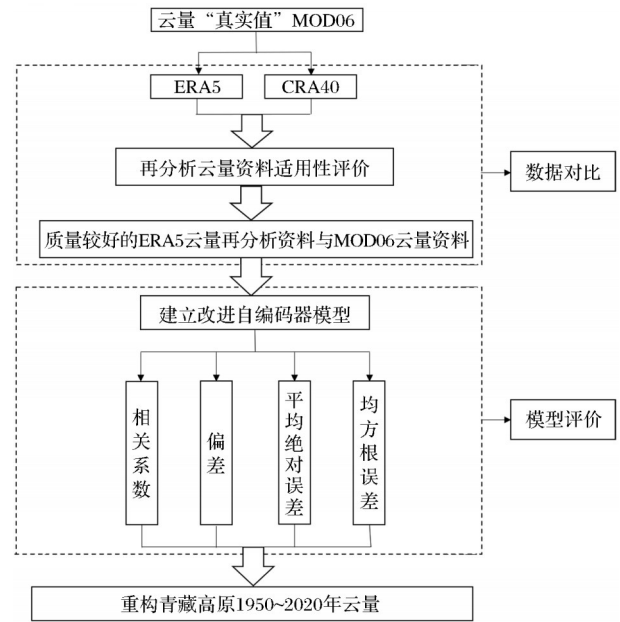


图 2 技术路线图

Fig.2 The flow chart of technology

3.1 云量资料适用性评价

本文利用青藏高原地区逐日云量数据分别计算月、季、年尺度的云量,公式如下:

$$y = \sum x/t \tag{1}$$

其中: y 为月、季、年尺度云量平均值; x 为日云量; t 为对应月、季、年的天数,本文对季节的划分为春季(3~5月)、夏季(6~8月)、秋季(9~11月)。

利用气候倾向率方法在进行云量线性回归分析,分析青藏高原地区云量变化情况,气候倾向率公式如下所示:

$$y_i = a + bx_i \tag{2}$$

其中: y 表示样本数为 n 的某一物理量; x 表示 y 所对应的时间样本个数; a 为回归常数; b 为回归系数^[28], b 值表示上升或下降的倾向度,即 $b > 0$ 时说明 y 值随时间 x 增加而增加,反之则说明 y 值随时间 x 增加而减小。

基于上述气候倾向率可分析藏高原地区不同

时间尺度上的云量变化情况,并通过计算相关系数分别评估ERA5、CRA40云量资料与云量“真实值”MOD06的相关性,进而评估ERA5与CRA40再分析资料在青藏高原地区的适用性。

3.2 云量重构方法

3.2.1 自编码器(Auto-Encoder)

自动编码器(Auto-Encoder)是一种特殊类型的神经网络结构(图3),它以无监督的方式进行训练,以获取输入数据低级别的形式。这些低级特征会被转换,用以映射实际数据^[29]。自动编码器的本质是一个回归模型,网络用来预测输入数据,由于网络只由几个神经元构成,因此会迫使网络产生有效的表示特征——将输入数据直接压缩成低编码的形式,以便解码器重现原始输入。自编码器采用非线性的映射,相对传统线性映射可以获取到图像更丰富的信息,对长时间序列数据重构具有较强的适用性,然而自编码器模型虽能完成图像的初步重构,但由于其使用全连接层进行下采样,导致图像的空间信息损失较为严重,因此本文需改进该模型以增强自编码器对图像特征的提取能力。

3.2.1 改进自编码器(Improved-AutoEncoder)

为了减少自编码器获取图像信息时产生的特征损失,本文将原自编码器模型的下采样方式由全连接操作改为卷积操作,该操作由设置卷积层实现。卷积层由若干卷积单元组成,卷积单元的参数由反向传播算法最佳优化得到。卷积运算的目的是提取输入数据的不同特征,一层卷积层只能提取一些低级的特征,更多层的网络能从低级特征中迭代提取高维语义特征,本文共设置四组卷积层,来提取输入图像的特征。对于二维图像,全连接层会

损失空间信息,而通过卷积操作,本文提出的改进自编码器能很好的保留二维信号的空间信息^[29],对于反映青藏高原内部云量差异具有重要意义。图4展示了本文改进的自编码器模型的结构,采用卷积方式,利用四组卷积对输入图像进行特征提取。在该模型中,epochs是对模型效果影响程度最大的参数,经过本文大量实验,当epochs处于0~25区间时,随着epochs取值增加,模型精度会提高,当epochs在25~50区间内,模型的精度增加逐渐放缓,达到50后趋于稳定,因此本文epochs取值为50。本模型采用均方误差(Mean Squared Error, MSE)作为损失函数,公式如下:

$$MSE = \frac{1}{m} \sum_{i=1}^m (y_i - f(x_i))^2 \quad (3)$$

其中: y_i 为真实值,本文指MOD06云量值, $f(x_i)$ 为预测值,本文指用改进自编码器模型重构出的云量值。本文基于2001~2010年的3~11月份ERA5与重采样后的MOD06云量资料作为训练样本,建立模型,重构出2011~2020年的云量资料IMAE-CF(Improved Autoencoder-cloud fraction, IMAE-CF),验证模型精度。利用2001~2020年的ERA5与MOD06作为训练样本,得出IMAE的权重函数,重构获取1950~2020年3~11月份的云量。

3.3 模型评价

为了评估青藏高原地区重构的云量IMAE-CF和云量真实值“MOD06”的有效性,综合考虑各类评价指标的适用性,选用了相关系数、偏差、平均绝对误差、均方根误差4个指标用于精度评价^[30]。

3.3.1 相关系数

皮尔森相关系数是一种线性相关系数,广泛应用于度量两个变量之间的相关程度,记为 R ,用来反

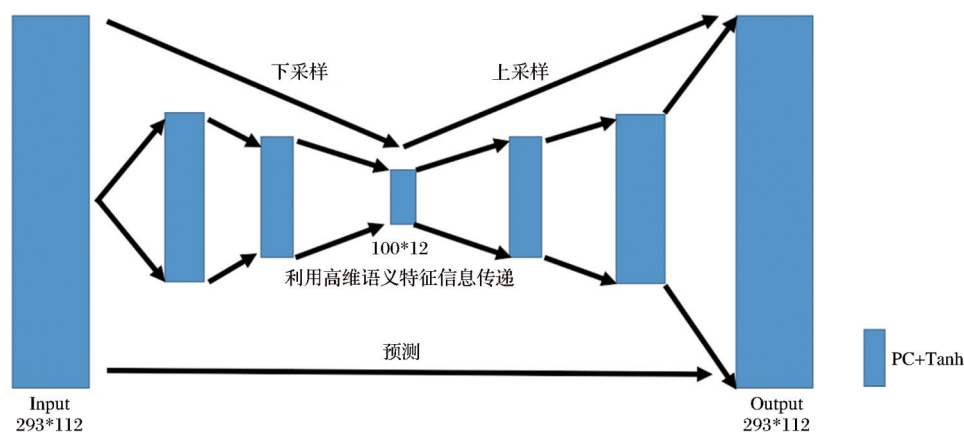


图3 自编码器原理图

Fig.3 The principle of Auto-Encoder

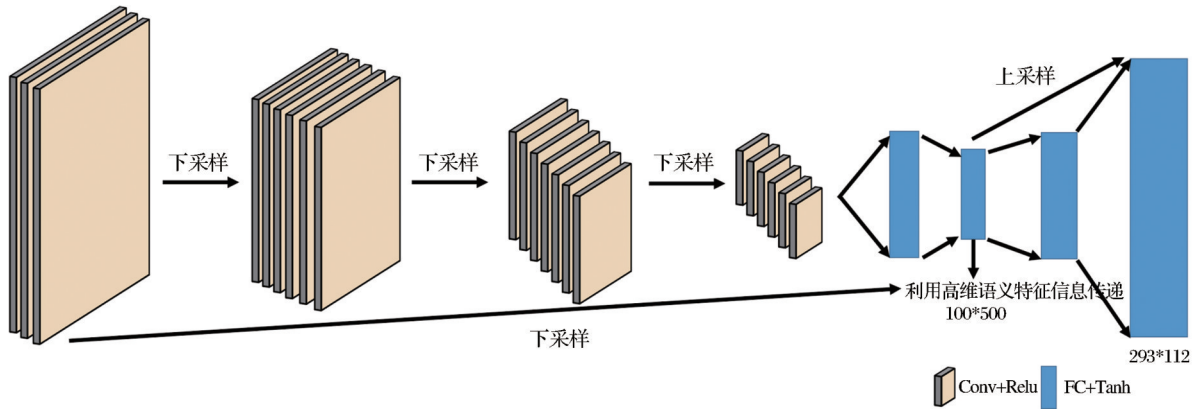


图4 改进自编码器原理图

Fig.4 The principle of improved Auto-Encoder

映两个变量的线性相关程度。计算公式如下：

$$R = \frac{\sum_{i=1}^n [x_i - \bar{X}_i][o_i - \bar{O}_i]}{\sqrt{\sum_{i=1}^n [x_i - \bar{X}_i]^2 \sum_{i=1}^n [o_i - \bar{O}_i]^2}} \quad (4)$$

其中： n 是参与比较的总样本数； x_i 为数据源； o_i 为要比较数据； \bar{X}_i 表示数据源的平均值； \bar{O}_i 表示要比较数据的平均值。本文 x_i 为MOD06云量资料， \bar{X}_i 表示MOD06的平均值， o_i 为需要与MOD06云量资料比较的CRA40云量资料、ERA5云量资料，以及改进自编码器所重构的云量值IMAE-CF， \bar{O}_i 为CRA40、ERA5、IMAE-CF的平均值，下述公式参数含义与式(3)一致。

3.3.2 偏差

偏差表示两种资料的平均偏离程度，计算公式如下：

$$Bias = \frac{1}{n} \sum_{i=1}^n (x_i - o_i) \quad (5)$$

3.3.3 平均绝对误差

平均绝对误差(MAE)指的就是模型预测值与样本真实值之间距离的平均值，反映模型预测值误差的实际情况，计算公式如下：

$$MAE = \frac{\sum_{i=1}^n |x_i - o_i|}{n} \quad (6)$$

3.3.4 均方根误差

均方根误差(RMSE)是预测值与真实值偏差的平方与观测次数 n 比值的平方根，常用来衡量观测值与真实值之间的偏差。如公式(7)所示：

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (x_i - o_i)^2}{n}} \quad (7)$$

4 结果分析

4.1 ERA5云量再分析资料有效性分析

本文将MOD06 3~11月份云量作为“真实值”，评价2001~2020年间3~11月份两种再分析资料在青藏高原地区的适用性^[18-19]。图5(a)和图5(b)分别

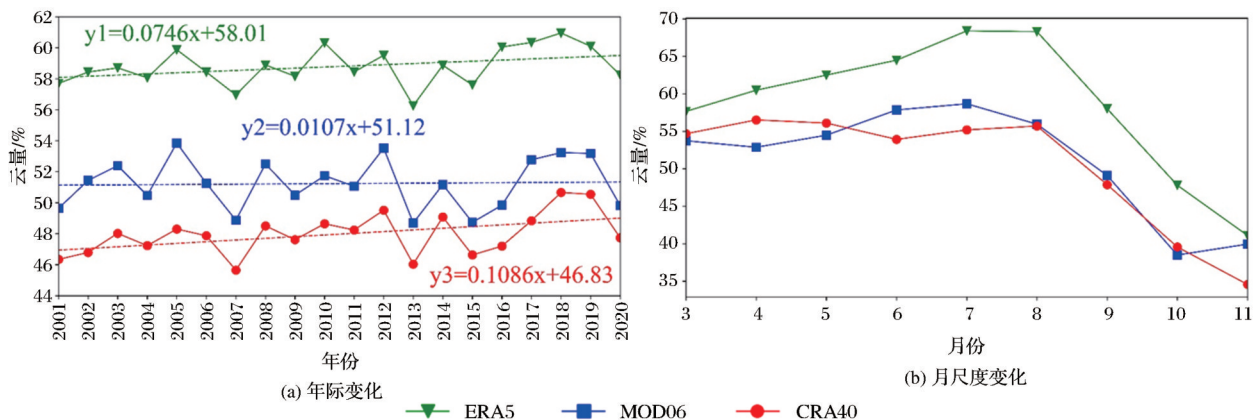


图5 3种资料2001~2020年青藏高原总体云量年际及月际变化(%)

(图(a)中虚线为线性拟合线,y1、y2、y3分别代表ERA5、MOD06、CRA40云量,x代表时间)

Fig.5 Annual and monthly variations of total cloud cover over the Tibetan Plateau from 2001 to 2020 based on three kinds of data (%)

展示了青藏高原2001~2020年3~11月份MOD06云量及ERA5、CRA40云量的年际变化及月尺度变化。可以看出,相比于MOD06数据,ERA5云量存在明显的高估问题,CRA40存在明显的低估问题。表2展示了两种再分析资料与MOD06云量资料的相对一致性,其结果均通过0.01的置信度检验。在3~11月份青藏高原地区,ERA5与MOD06云量相关系数均在0.6以上,且6~9四个月相关性最好,相关系数最高,分别达到0.82、0.82、0.80、0.85,其与MOD06的相关性明显优于CRA40云量资料。由上述分析可知,在云量值上两种再分析资料均存在不确定性,但ERA5与MOD06云量的相对一致性明显优于CRA40,表明ERA5再分析资料在刻画青藏高原云量的空间分布上更具优势,因此本文进一步利用ERA5与MOD06云量重构青藏高原云量。

表2 MOD06与CRA40、ERA5的1-12月份的相关系数
Table 2 The correlation coefficient between MOD06 and CRA40 and ERA5 from January to December

	3	4	5	6	7	8	9	10	11
CRA40	0.51	0.52	0.49	0.63	0.73	0.70	0.70	0.67	0.59
ERA5	0.61	0.67	0.68	0.82	0.82	0.80	0.85	0.77	0.64

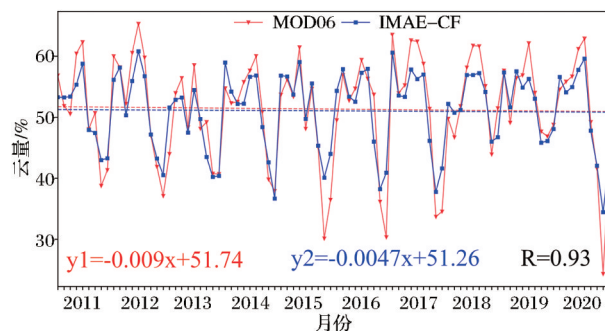


图6 2011~2020年3~11月份MOD06、IMAE-CF云量时间序列(图中虚线为线性拟合线,y1、y2分别代表MOD06、IMAE-CF云量,x代表时间)

Fig.6 Time series of MOD06 and IMAE-CF from March 2011 to November 2020

4.2 模型的月尺度评价

2011~2020年3~11月份MOD06云量资料以及模型预测的云量IMAE-CF的时间序列如图6所示可以看出,IMAE-CF云量资料较好地模拟了云量的变化趋势,与MOD06的变化趋势的相关性达到0.93,且极好的校正了ERA5云量资料值偏高的问题。云量资料均显示青藏高原云量3~11月份的基本呈现出3~5月份降低,5~7月份升高,至7月份达到最大值,8~11月份逐步降低的变化趋势。

图7(a)~(i)为2011~2020年3~11月份云量“真实值”MOD06、与模拟云量值IMAE-CF之间的密度散点图,其中横坐标代表MOD06云量资料,纵坐标代表模拟模型IMAE-CF,散点的颜色代表了该位置点的密度。可以看出,相较于ERA5云量资料与MOD06云量资料的相关系数,利用改进自编码器模型建立出的IMAE-CF云量资料与MOD06的相关系数在3~11月份均发生了较大提高,3~11月份相关系数均在0.8以上,6~9月份相关系数达到0.9以上,相比较于ERA5与MOD06的相关系数平均提高了20.72%。

图8(a)~图8(c)展示了2011~2020年10年校正前ERA5与云量真实值MOD06及校正后IMAE-CF云量与MOD06的Bias、MAE、RMSE的月际变化,其中横坐标代表时间,纵坐标代表不同评价指标。可以看出,本文提出的方法可以有效降低各项数据误差。其中,Bias在8~10月校正效果最好,而在3、11月校正效果相对较差。ERA5与MOD06的MAE、RMSE平均值分别为12.04%、15.06%,而模拟的云量值IMAE-CF与MOD06的MAE、RMSE平均值降低至7.26%、9.13%。RMSE与MAE的变化趋势基本一致,校正后的云量误差值夏季(6~8月)<春季(3~5月)<秋季(9~11月)。3~5月RMSE与MAE整体呈现下降趋势,5~7月呈上升趋势,然后下降至8月份,9~11月份呈上升趋势,至11月份达到峰值。各项指标显示在3、11月份校正效果相对较差,这可能是因为3、11月份青藏高原地区仍有少量积雪覆盖,使得原本ERA5数据与MOD06云量资料相对一致性较差,从而导致模型精度降低。

4.3 模型的季尺度评价

图9展示了2011~2020年青藏高原春、夏、秋三季的ERA5、MOD06、以及IMAE-CF的云量气候态,2011~2020年春季ERA5、MOD06、IMAE-CF云量气候态分别为60.40%、53.91%、54.47%,夏季云量值分别为67.06%、57.71%、55.21%,秋季云量值分别为49.65%、42.35%、43.44%。可以看出,MOD06与IMAE-CF在云量值上差异很小,且呈现出较好的空间一致性。云量分布总体呈现出春季(3~5月)为高原云量分布整体呈现东部高西部低的状态,夏季(6~8月)的分布特征为由东南向西北递减,秋季(9~11月)云量分布特征为东部高、西部低,且云量值最小。虽然ERA5与IMAE-CF空间

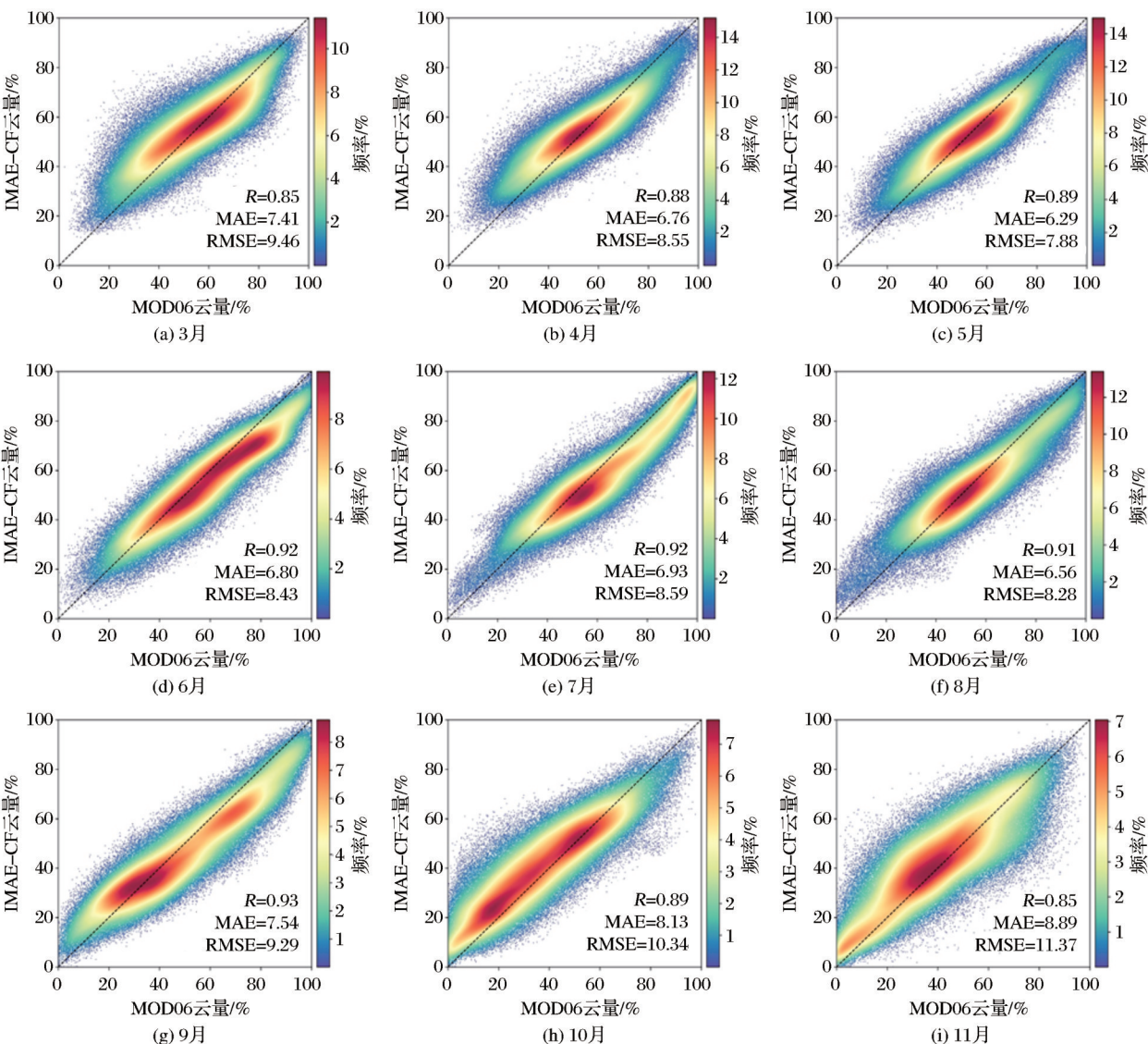


图 7 2011~2020 年 3~11 月份 MOD06 与 IMAE-CF 密度散点图 (图中黑线为对角线)

Fig.7 The density scatter of mod06 and IMAE-CF from March 2011 to November 2020

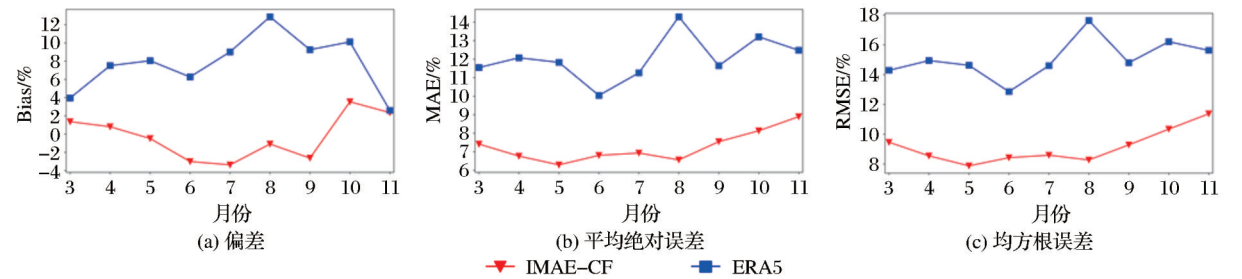


图 8 IMAE-CF、ERA5 与 MOD06 的 Bias、MAE、RMSE 变化图

Fig.8 The monthly variation of bias, MAE and RMSE

分布上表现出的分布趋势大致相同,但模型模拟的云量值在反映青藏高原云量内部差异的表现上明显优于ERA5 云量资料。

图 10 展示了春、夏、秋三季 MOD06 云量以及模拟云量 IMAE-CF 云量气候态的密度散点图,由图可知,春、夏、秋三季的相关系数分别达到 0.97、

0.99、0.98。可以看出模拟云量气候态与 MOD06 云量气候态具有高度相关性。

表 3 进一步量化了 2011~2020 年春、夏、秋三季 ERA5 与 MOD06 及 IMAE-CF 与 MOD06 的 Bias、MAE、RMSE。可以看出,原始 ERA5 云量资料在空间分布上与 MOD06 云量相对一致性较差,而模

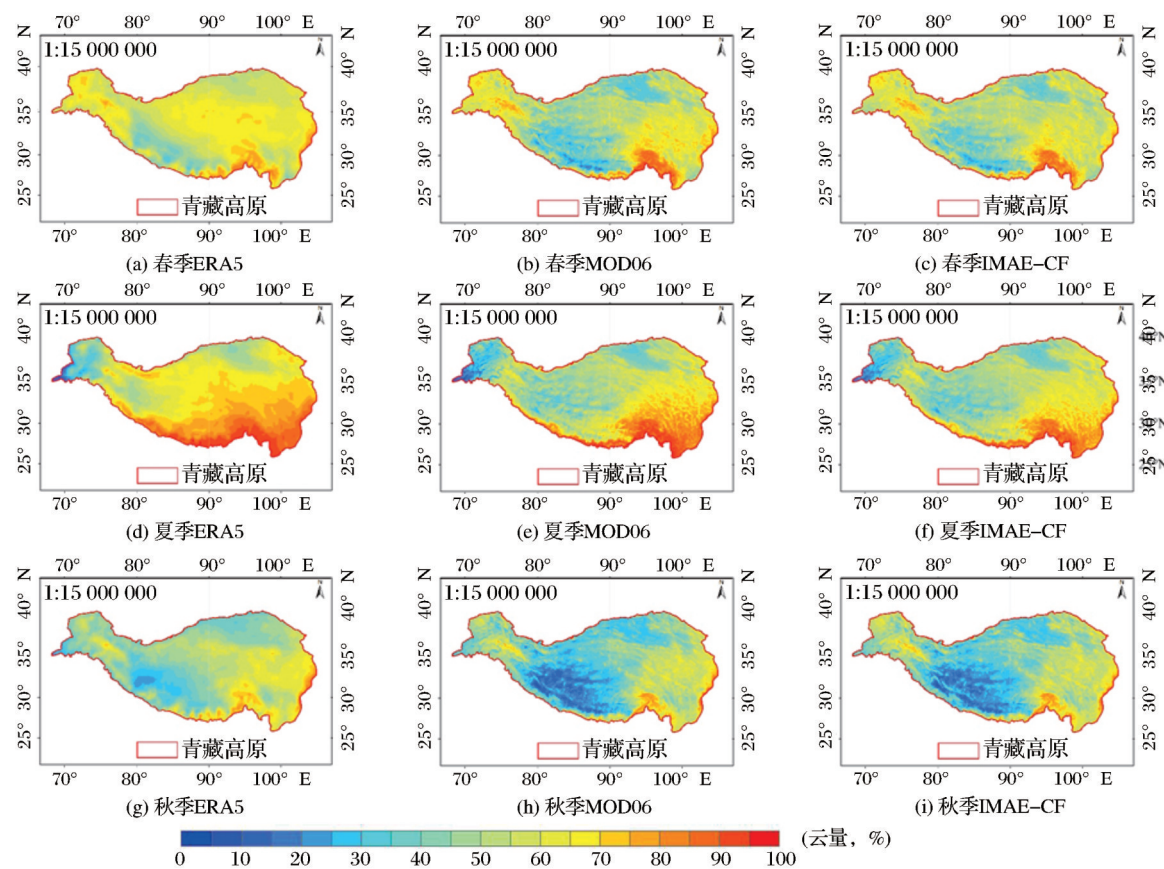


图 9 2011~2020 年 ERA5、MOD06、IMAE-CF 春、夏、秋云量气候态

Fig.9 Climatology of ERA5, MOD06 and IMAE-CF in spring, summer and autumn from 2011 to 2020

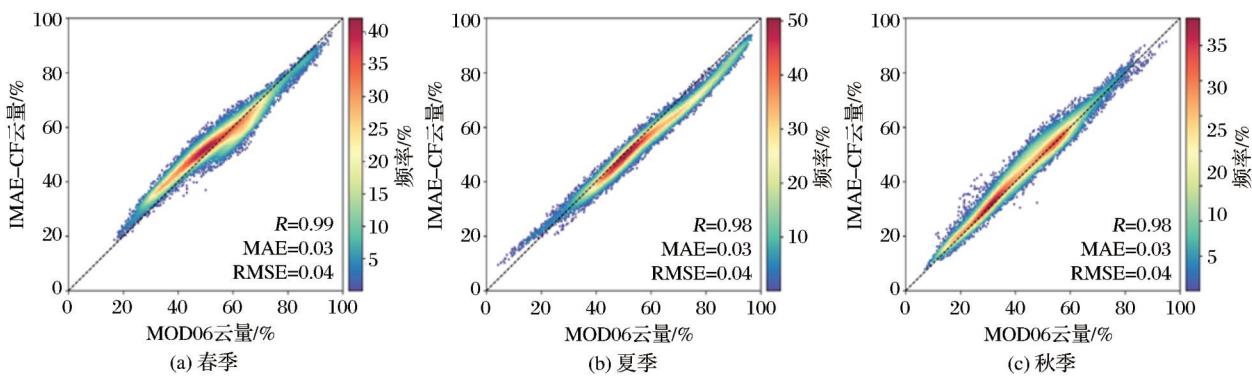


图 10 2011~2020 年春、夏、秋季 MOD06 与 IMAE-CF 密度散点图(图中黑线为对角线)

Fig.10 The density scatter of MOD06 and IMAE-CF in spring, summer and autumn from 2011 to 2020

拟后的云量资料与 MOD06 云量资料的相对一致性显著提高,Bias、MAE、RMSE 等误差均大幅度降低,进一步验证了模拟的云量资料在青藏高原云量内部差异的表现上优于 ERA5 云量资料。

4.4 重构长时间序列云量

将 2001~2020 年的 3~11 月份 ERA5 与 MOD06 云量资料作为训练样本,建立改进自编码器模型,并生成权重。输入 1950~2020 年的 3~11 月份 ERA5 云量资料,得到 1950~2020 年的 3~11

表 3 2011~2020 年 ERA5、IMAE-CF 与 MOD06 春夏秋三季云量气候态各项评价指标 (%)

Table 3 ERA5, IMAE-CF and MOD06 evaluation indicators of climatology in spring, summer and Autumn from2011 to 2020(%)						
检验指标	ERA5&MOD06			IMAE-CF&MOD06		
	春季	夏季	秋季	春季	夏季	秋季
Bias	6.50	9.38	7.32	0.65	-2.50	1.08
MAE	9.62	10.36	9.31	2.97	3.01	2.24
RMSE	11.78	12.87	11.38	3.61	3.65	2.93

月份云量资料 IMAE-CF。

图 11 展示了 1950~2020 年 3~11 月份每两年的 ERA5 云量与模拟云量 IMAE-CF 的时间序列,并叠加展示 2001~2020 年 3~11 月份每两年的云量“真实值”MOD06 的时间序列。

由图 11 可以看出云量呈现波动上升趋势, ERA5 虽云量值偏高,但对云量趋势的刻画较为准

确,模拟云量将 ERA5 云量值平均校正降低 13.85%,相关系数达到 0.92,证明了模拟云量的可靠性。由模型模拟出的 2001~2020 年 3~11 月份的时间序列与 MOD06 的时间序列相关系数达到 0.99。可以看出通过模型重构的云量既可以有效校正 ERA5 云量数值偏高的问题,又可以准确地反映出云量的变化趋势。

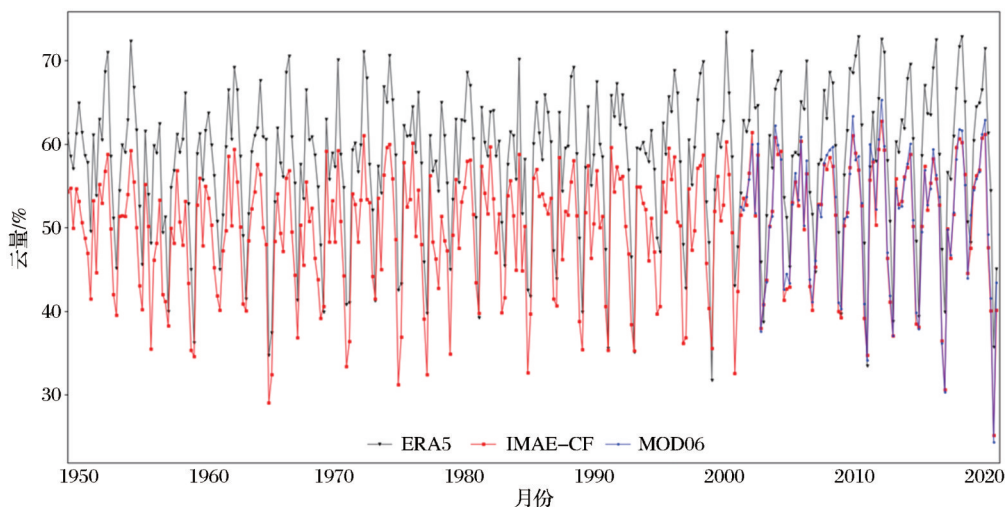


图 11 1950~2020 年 3~11 月份 ERA5 与 IMAE-CF 的时间序列,叠加 2001~2020 年 3~11 月份 MOD06 时间序列(每两年显示一次)

Fig.11 Time series of ERA5 and IMAE-CF from March 1950 to November 2020, Time series of MOD06 are overlaid (Display every two years)

图 12 展示了 1950~2020 年春、夏、秋三季模拟云量及 ERA5 云量分布情况。可以看出两种云量资料在春、夏、秋三季云量整体空间分布上表现大致相同,即春季云量东部高西部低,夏季云量由东南向西北递减,秋季云量为东部高、西部低。ERA5 云量资料在春、夏、秋三季均高估云量,而模拟云量 IMAE-CF 显示的青藏高原云量内部差异明显优于 ERA5,且极好地改善了云量高估问题。

5 结论与讨论

实验以 MOD06 云量为“真实值”,对青藏高原 2001~2020 年 3~11 月份的 ERA5、CRA40 云量资料进行适用性评价与分析,并基于相关系数(R)、偏差(Bias)、平均绝对误差(MAE)、均方根误差(RMSE)4 种评价指标评估本文所提出的改进自编码器模型的精度,最终重构 1950~2020 年青藏高原 3~11 月份云量。主要结论如下:

(1) 总体来看,青藏高原地区 2001~2020 年 3~11 月份 ERA5 云量资料偏高, CRA40 云量资料偏低; ERA5 云量与 MOD06 云量的相关性明显优于 CRA40 云量与 MOD06 云量的相关性,且在 6~9

月相关性最好,相关系数分别达到 0.82、0.82、0.80、0.85。

(2) 基于改进自编码器建立的云量重构模型可以有效校正 3~11 月份 ERA5 在青藏高原高估云量的问题。从相关系数(R)、偏差(Bias)、平均绝对误差(MAE)、均方根误差(RMSE)4 个方面评估模型精度,发现相较于原 ERA5 云量资料、本文所模拟的云量资料与云量“真实值”MOD06 更为接近,该模型极好地提升了相关系数、降低了各项误差。

本研究所提出的自编码器模型不仅对实现云量重构具有重要的意义,同时对实现不同气象要素如降水、水汽的重构也具有启示意义。利用多种不同类型的云量资料,重构了青藏高原云量,可为青藏高原提供了更为准确、时间序列更长的 3~11 月份的云量产品,对位置独特、气候多样、地形复杂的高原地区的气象研究具有十分重要的应用价值。但该研究仍存在不足,如青藏高原冬季由于积雪覆盖导致卫星云量产品质量较差,因此本文并未重构冬季云量,后续考虑增加降水、水汽、蒸散发等与云量直接相关的气候因素,减弱积雪对云量产品的影响,建立适用性更强的模拟模型,形成更为完备的

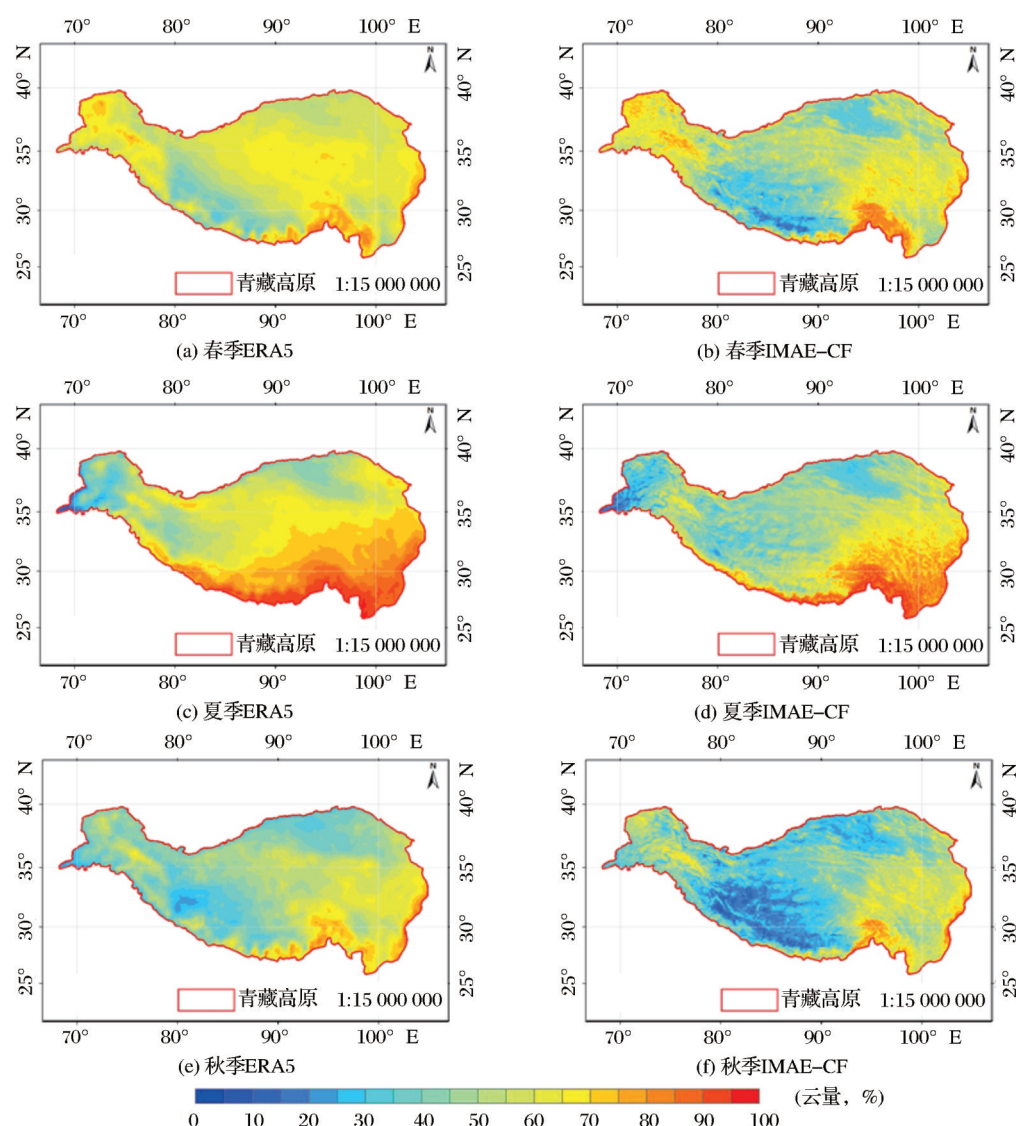


图 12 1950~2020 年 ERA5 与 IMAE-CF 春、夏、秋云量空间分布图

Fig.12 Climatology of ERA5 and IMAE-CF in spring, summer and autumn from 1950 to 2020

云量产品。

致谢:感谢 NASA、ECMWF、CMA 等单位提供各种资料数据,感谢中国—巴基斯坦地球科学研究中心提供科研条件完成此项研究!

参考文献 (References):

- [1] SEAGER R, NAIK N, VECCHI G A. Thermodynamic and dynamic mechanisms for large-scale changes in the hydrological cycle in response to global warming [J]. *Journal of Climate*, 2010, 23(17): 4651-4668.
- [2] CHEN H, ZHU Q, PENG C, *et al.* The impacts of climate change and human activities on biogeochemical cycles on the Qinghai-Tibetan Plateau [J]. *Global Change Biology*, 2013, 19(10): 2940-2955.
- [3] DUAN A, XIAO Z. Does the climate warming hiatus exist over the Tibetan Plateau? [J]. *Scientific Reports*, 2015, 5 (1): 1-9.
- [4] TANG Qihong, LIU Yubo, ZHANG Chi, *et al.* Research Progress on water vapor source change of precipitation in Qinghai-Tibetan Plateau and its surrounding areas [J]. *Transactions of Atmospheric Sciences*, 2020, 43(6): 1002-1009. [汤秋鸿, 刘宇博, 张弛, 等. 青藏高原及其周边地区降水的水汽来源变化研究进展 [J]. *大气科学学报*, 2020, 43(6): 1002-1009.]
- [5] ZHANG Xueqin, PENG Lili, ZHENG Du, *et al.* Temporal and spatial variation of total cloud cover over Qinghai-Tibetan Plateau from 1971 to 2004 and its influencing factors [J]. *Acta Geographica Sinica*, 2007, 62(9): 959-969. [张雪芹, 彭莉莉, 郑度, 等. 1971-2004 年青藏高原总云量时空变化及其影响因子 [J]. *地理学报*, 2007, 62(9): 959-969.]

- [6] HARRISON E F, MINNIS P, BARKSTROM B R, *et al.* Seasonal variation of cloud radiative forcing derived from the Earth Radiation Budget Experiment[J]. *Journal of Geophysical Research: Atmospheres*, 1990, 95(D11): 18687-18703.
- [7] DAI Jiaxian. The Climate of the Qinghai-Xizang Plateau. Beijing: China Meteorological Press[M]. Beijing: Meteorological Publishing House, 1990: 215-221.[戴加洗. 青藏高原气候. 北京: 气象出版社, 1990: 215-221.]
- [8] YOU Q, JIAO Y, LIN H, *et al.* Comparison of NCEP/NCAR and ERA-40 total cloud cover with surface observations over the Tibetan Plateau[J]. *International Journal of Climatology*, 2014, 34(8): 2529-2537.
- [9] BAO S, LETU H, ZHAO J, *et al.* Spatiotemporal distributions of cloud parameters and their response to meteorological factors over the Tibetan Plateau during 2003 - 2015 based on MODIS data[J]. *International Journal of Climatology*, 2019, 39(1): 532-543.
- [10] MA Q R, YOU Q L, MA Y J, *et al.* Changes in cloud amount over the Tibetan Plateau and impacts of large-scale circulation[J]. *Atmospheric Research*, 2021, 249: 105332.
- [11] CHEN Shaoyong, DONG Anxiang. Climate change and stability of total cloud cover over the Qinghai Xizang Plateau [J] *Arid Zone Research*, 2006, 23(2): 327-333.[陈少勇, 董安祥, 青藏高原总云量的气候变化及其稳定性[J]. 干旱区研究, 2006, 23(2): 327-333.]
- [12] NORRIS J R, ALLEN R J, EVAN A T, *et al.* Evidence for climate change in the satellite cloud record[J]. *Nature*, 2016, 536(7614): 72-75.
- [13] ZHANG Dejie, SHI Chunxiang, ZHANG Tao, *et al.* Comparative analysis of total cloud cover products from various data in China[J] *Plateau Meteorology*: 2022, 41(3): 803-813. [张德杰, 师春香, 张涛, 等. 多种资料的总云量产品在中国区域的对比分析[J]. 高原气象, 2022, 41(3): 803-813.]
- [14] DUAN A, WU G. Change of cloud amount and the climate warming on the Tibetan Plateau [J]. *Geophysical Research Letters*, 2006, 33(22): L22704. DOI:10.1029/2006GL027946
- [15] WANG J, DAI A, MEARS C. Global water vapor trend from 1988 to 2011 and its diurnal asymmetry based on GPS, radiosonde, and microwave satellite measurements[J]. *Journal of Climate*, 2016, 29(14): 5205-5222.
- [16] LI J, MAO J, WANG F. Comparative study of five current reanalyses in characterizing total cloud fraction and top-of-the-atmosphere cloud radiative effects over the Asian monsoon region[J]. *International Journal of Climatology*, 2017, 37(15): 5047-5067.
- [17] MA J, WU H, WANG C, *et al.* Multiyear Satellite and Surface Observations of Cloud Fraction over China[J]. *Journal of Geophysical Research Atmospheres*, 2014, 119(12): 7655-7666.
- [18] KOTARBA A Z. A Comparison of MODIS-derived cloud amount with visual surface observations[J]. *Atmospheric Research*, 2009, 92(4): 522-530.
- [19] LIU Jian. Comparative analysis of three satellite cloud cover data in the Qinghai Tibet Plateau[J]. *Journal of Remote Sensing*, 2021, 25(7): 1445-1459.[刘健. 三种卫星云量数据在青藏高原地区的比对分析[J]. 遥感学报, 2021, 25(7): 1445-1459.]
- [20] LIU X, CHEN B. Climatic warming in the Tibetan Plateau during recent decades[J]. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 2000, 20(14): 1729-1742.
- [21] ZHANG G, YAO T, XIE H, *et al.* Increased mass over the Tibetan Plateau: from lakes or glaciers?[J]. *Geophysical Research Letters*, 2013, 40(10): 2125-2130.
- [22] KING M D, PLATNICK S, MENZEL W P, *et al.* Spatial and temporal distribution of clouds observed by MODIS onboard the Terra and Aqua Satellites[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2013, 51(7): 3826-3852.
- [23] PLATNICK S, MEYER K G, KING M D, *et al.* The MODIS cloud optical and microphysical products: Collection 6 updates and examples from Terra and Aqua[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2016, 55(1): 502-525.
- [24] HERBACH H, BELL B, BERRISFORD P, *et al.* The ERA5 global reanalysis [J]. *Quarterly Journal of the Royal Meteorological Society*, 2020, 146(730): 1999-2049.
- [25] LIU Z Q, SHI C X, ZHOU Z J, *et al.* CMA global reanalysis (CRA-40): Status and plans[C]//Proc. 5th International Conference on Reanalysis. Rome, Italy: Nat. Meteor. Int. Canter, 2017: 13-17.
- [26] ZHAO D, ZHANG L, ZHOU T, *et al.* Contributions of local and remote atmospheric moisture fluxes to East China Precipitation estimated from CRA-40 reanalysis [J]. *Meteorol. Res.*, 2021, 35(1): 32-45.
- [27] KANG Xiaowei, FENG Zhongkui. Aster GDEM data introduction and program reading[J]. *Remote Sensing Information*, 2011(6): 69-72.[康晓伟, 冯钟葵. ASTER GDEM数据介绍与程序读取[J]. 遥感信息, 2011, (6): 69-72.]
- [28] WEI Fengying. Modern Climate Statistical Diagnosis and Prediction Technology(Second Edition)[M]. Beijing: Meteorologi-

- cal Publishing House, 2007:43-46[魏凤英.现代气候统计诊断与预测技术(第二版)[M].北京:气象出版社,2007:43-46.]
- [29] YUAN Feiniu, ZHANG Lin, SHI Jinting, *et al.* Overview of theory and application of self coding Neural Network[J]. Chinese Journal of Computers, 2019, 42(1): 203-230.[袁非牛, 章琳, 史劲亭, 等. 自编码神经网络理论及应用综述[J]. 计算机学报, 2019, 42(1): 203-230.]
- [30] SHENG Xia, SHI Yuli, DING Haiyong. Spatial downscaling of GPM precipitation over the Tibetan Plateau [J]. Remote Sensing Technology and Application, 2021, 36(3): 571-580. [盛夏, 石玉立, 丁海勇. 青藏高原 GPM 降水数据空间降尺度研究[J]. 遥感技术与应用, 2021, 36(3): 571-580.]

Reconstruction of Remote Sensing Cloud Cover over Tibetan Plateau based on MODIS Data

ZHOU Houyu^{1,2}, DONG Qing^{1,3}, MENG Deli^{1,2}, ZHAO Wenbo^{1,2}, Bian Min^{1,2}

(1.Laboratory of Digital Earth Science, Aerospace Information Research Institute, CAS, Beijing 100094 China;

2.University of Chinese Academy of Sciences, Beijing 100049, China;

3.Key Laboratory of Target Microwave Properties of Zhejiang, Deqing Academy of Satellite Applications, Huzhou 313200, China)

Abstract: The Tibetan Plateau (TP), with its unique climate characteristics and geographical pattern, plays an important role in global climate change. As an important part of the earth atmosphere system, cloud is key to affecting climate change. Cloud cover can more directly reflect the change of cloud. Therefore, it is of great significance to reconstruct a cloud cover product with longer time series and higher accuracy in the TP. In this paper, Considering the complex underlying surface types and geographical elevations in the TP, We select the cloud cover of MOD06, ERA5 and CRA40 from 2001 to 2020. We take the cloud cover of MOD06 from March to November as the true value and evaluate the applicability of the two reanalysis data in the TP through methods such as climate tendency rate and correlation coefficient. Based on ERA5 and MOD06, the improved auto-encoder model is used to reconstruct the cloud cover of the plateau from March to November of 1950 to 2020. The results show that the cloud cover of ERA5 is higher than that of MOD06, while that of CRA40 is lower than that of MOD06, and the correlation between ERA5 and MOD06 is obviously better than that between CRA40 and MOD06; The improved auto-encoder model evaluated by four evaluation indicators of correlation coefficient (R), bias, Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) has a good effect on cloud amount reconstruction. The correlation coefficient between cloud amount reconstructed by the improved auto-encoder model and MOD06 cloud amount data increases by more than 20% on average from March to November, and can simulate the change trend of cloud amount over the TP. The results provide reliable long time series data for studying the temporal and spatial evolution of cloud cover over the TP.

Key words: Tibetan Plateau (TP); Cloud over; MODIS; Reanalysis data; Data reconstruction; Auto-encoder