



# 一种基于降噪自编码神经网络的积雪产品去云方法

## 摘要

目前应用最为广泛的积雪覆盖区域图(SCA)可由中分辨率成像光谱仪(MODIS)获取,常被用于积雪覆盖时空变化的研究中.由于受云遮挡的影响,MODIS积雪产品存在较大区域的数据缺失.为了消除云遮挡的影响,本文构建一种降噪自编码神经网络模型,建立雪粒径与复杂地形、土地覆盖类型之间的复杂的映射关系,实现云下积雪参数的补全,提高积雪产品的覆盖面积.本文选取开都河流域为研究区域,将MODIS反演得到的积雪产品数据与地形地物数据结合,并通过降噪自编码神经网络(Denoising Autoencoder Artificial Neural Network)、极值雪线法相结合的方法来定量地回归补全高山复杂地形下由于云覆盖导致的积雪缺失数据,从而得到无缺失的逐日雪盖数据.其中,降噪自编码神经网络融合多特征数据,建立地形特征与雪粒径数据之间的非线性映射关系,从而来补全云层下的雪粒径数据;极值雪线法主要用来去除低海拔地区误报值,进一步提高雪盖提取精度.采用MODIS积雪产品对去云结果开展精度验证,本文所提出的去云方法的精度超过86%,有效地提高了雪盖提取精度.因此,本文所提的算法可以有效地去除复杂地形区域的云覆盖.

## 关键词

降噪自编码神经网络;极值雪线法;复杂地形;去云

中图分类号 P426.63+5;P407.8

文献标志码 A

收稿日期 2021-06-10

资助项目 国家自然科学基金(41875027,41871238)

作者简介

张永宏,男,博士,教授,主要从事大气遥感检测、图像处理分析研究.zyh@nuist.edu.cn

1 南京信息工程大学 自动化学院,南京,210044

2 南京信息工程大学 大气物理学院,南京,210044

3 南京信息工程大学 电子与信息工程学院,南京,210044

## 0 引言

作为现代气候系统的五大圈层之一的冰冻圈,其中的积雪分布最为广泛,并且是高海拔地区农田灌溉以及居民生活的主要用水来源<sup>[1-2]</sup>,过量的积雪也会引发积雪灾害<sup>[3]</sup>.新疆作为我国冰雪储存量最为丰富的地区,其积雪的储存量位居全国之首,约占全国积雪资源的三分之一<sup>[4]</sup>.而位于新疆天山南坡的开都河流域,是山间盆地的典型区域,冰雪融水与降水是当地的主要水资源补给,其中积雪的变化直接影响着开都河流域的径流变化<sup>[5]</sup>.

为了获取开都河流域长时间、大范围的积雪覆盖信息,中分辨率成像光谱仪(Moderate-resolution Imaging Spectroradiometer, MODIS)积雪产品逐步得到应用,对于该仪器的诸多研究<sup>[6-7]</sup>表明MODIS数据在探测积雪信息方面具有很强的能力.卫星遥感所提供的多尺度、长时间序列、高空间分辨率且时间连续的卫星数据,对于定量、定性监测积雪信息与变化具有重要且积极的作用.但同时,MODIS作为一颗光学卫星,云遮挡限制了该产品的使用.诸多科研人员对积雪影像中的云像元去除进行了大量的研究,可以概括为:多传感器融合法、时间或空间过滤法、模型驱动法<sup>[8]</sup>.

多传感器融合的去云方法融合光学遥感数据与微波遥感数据,利用微波对云的穿透性,实现对云层下积雪信息进行获取<sup>[9-11]</sup>.时间或空间过滤法是另一种常用的去云方法.时间与空间过滤法可以单独使用,但通常综合使用来达到更佳的去云效果<sup>[12-15]</sup>.然而这些方法在积雪积累期,山地的云覆盖持续时间较长,去云精度降低,并且受季节、地形特性等环境因素影响.

模型驱动法主要依赖于建立积雪与其他环境因素之间的关系模型.Jain等<sup>[16]</sup>探讨了喜马拉雅地区积雪与海拔高度之间的相互关系,对积雪覆盖面积(SCA)与积雪持续时间(SCD)进行估计,得到积雪覆盖面积与积雪持续时间随海拔的变化趋势.Kour等<sup>[17]</sup>评估了喜马拉雅山西部奇纳布河盆地的积雪覆盖面积百分比与海拔、坡度以及坡向之间的关系,结果显示积雪覆盖面积百分比与地形之间存在着密切的关系.侯海艳等<sup>[18]</sup>将新疆北部地区观测资料建立人工神经网络,通过优选模型较为合理地估算了北疆地区的积雪深度信息.

本研究提出降噪自编码神经网络与极值雪线相结合的方法对MODIS积雪数据中的云覆盖像元进行去除.本文以MODIS数据反演

所得的积雪产品为研究对象,采用回归模型从定量角度对积雪信息进行分析估计,从而从积雪数据中去除云覆盖像元,并生成研究区域的逐日连续的积雪覆盖数据。

## 1 研究区域概况

如图 1 所示,研究区域位于新疆天山南坡的开都河流域(82.94°~86.02°E,42.19°~43.35°N)。天山地区的降水主要受西风气流和北冰洋气流的水汽影响,开都河也是我国最长的内陆河流——塔里木河的四大主要源流之一,属于暖温带大陆性干旱气候<sup>[19]</sup>。

开都河流域北高南低,地形复杂,流域上游段海拔高度一般在 3 000 m 以上,区域内最高山峰为克勒代乌拉,海拔 4 679 m。海拔 3 600 m 以上地域常年积雪,地形以山脉为主,且地势险峻、人烟罕至,积雪零碎,该区域内观测站稀少<sup>[20]</sup>。在开都河流域腹地,卫星数据反演的精度相对较低,且在云遮挡情况下由于缺少地面站数据而导致验证困难。所以本文基于卫星遥感数据利用神经网络的方法,对气象要素以及地形参数进行联合训练来估计缺失的积雪信息,以期有效地减少卫星遥感影像中的云干扰,研制无云或少云的雪盖数据。

## 2 数据来源

### 2.1 雪粒径数据

雪粒径是描述积雪微观物理性质的一项重要参

数,其时空分布作为融雪径流模型、雪化学模型以及气候模型的输入参数,对融雪进行评估具有重要意义<sup>[21]</sup>。本研究主要以反演所得积雪产品作为研究对象,基于量化积雪数据进行模型搭建,估计被云遮挡的雪粒径数值,从而达到对积雪产品的去云处理效果。许多研究表明,基于 MODIS 500 m 分辨率数据能够较理想地模拟雪粒径<sup>[22-24]</sup>。本文所选用的数据包括 2000—2020 年空间分辨率为 500 m 的 MODIS 反演得到的逐日雪粒径原始数据<sup>[25]</sup>,数据相关参数如表 1 所示。

表 1 基于 MODIS 反演的雪粒径数据说明

Table 1 Description of snow grain size data retrieved from MODIS

属性	变量描述
数据类型	0~1 000 $\mu\text{m}$ (0:没有数据;其他:雪粒径大小)
数据格式	TIFF
空间范围	82.33°~88.12°E;42.03°~44.37°N
空间分辨率	500 m
时间范围	MOD:2000年2月27日至2020年5月7日 MYD:2002年7月4日至2020年5月9日
时间分辨率	1 d

### 2.2 数字高程数据

数字高程模型(Digital Elevation Model, DEM),是地表地貌信息的数字化表达,利用 DEM 作为输入数据源,是实现积雪识别的一种有效的方法<sup>[26]</sup>。积雪日数和 DEM 之间存在着明显的正相关关系,积雪的

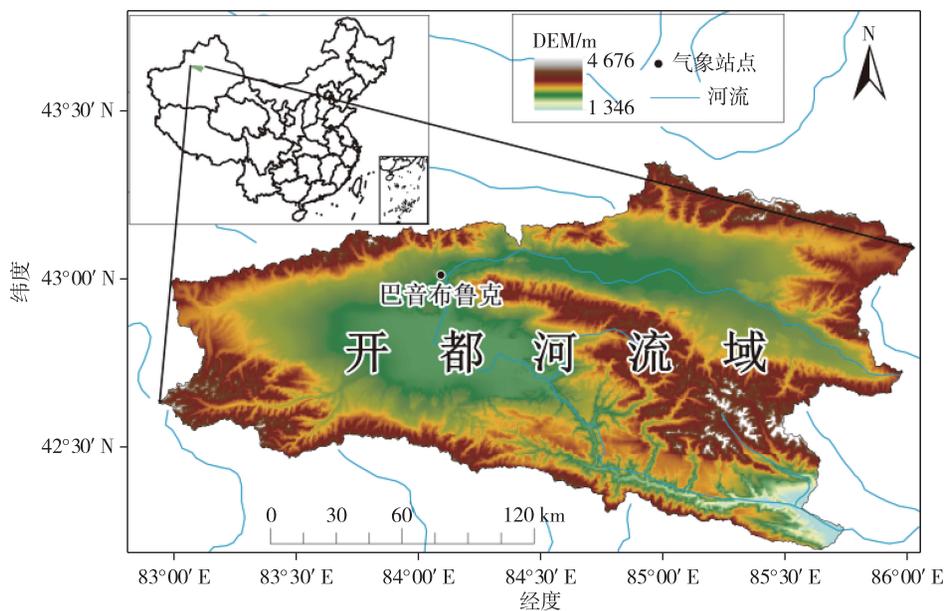


图 1 研究区概况

Fig. 1 Map showing the overview of the research region

空间分布变化受海拔高度差异因素的影响<sup>[27]</sup>.本文所使用的是来源于地理空间数据云的 ASTER GDEM V2 数据集.通过 Arcgis 软件对其进行数据预处理,包括对原始 DEM 数据进行镶嵌、投影,按照所研究开都河流域的矢量图进行裁剪处理,同时采用双线性插值方法将裁剪所得的 DEM 数据的空间分辨率由 30 m 上采样为 500 m,最后得到分辨率为 500 m 的开都河流域 DEM 数据.

### 2.3 土地覆盖类型数据

雪粒径在时空关系上的变化与气象条件、下垫面性质密切相关.采用 MODIS 卫星反射率数据的监督分类得到的土地覆盖类型产品(Land Cover data) MCD12Q1 地物数据,其为分辨率 500 m 的土地覆盖类型描述数据,共包含 17 个土地覆盖类型.研究区植被多为旱生植物,且植被种类较多<sup>[28]</sup>,主要以草地为主,在盆地平原存在较大面积湿地,在高海拔山岭主要分布大量的雪和冰.

## 3 研究方法

### 3.1 基于极值雪线法的高程滤波模型

为了减少本文降噪自编码神经网络回归模型对较低海拔分类误差较大的后果,采用极值雪线法进行处理.在高山地区,海拔通常被认为是雪盖分布的主导因素<sup>[29]</sup>,而极值雪线法正是通过对无云像元的海拔信息对云覆盖进行剔除的.

雪线法通常基于最小积雪像元的海拔作为陆地线,最大积雪像元的海拔作为雪线.处理时将低于陆地线海拔地区的云覆盖分类为陆地,将高于雪线海拔地区的云覆盖分类为雪盖,而位于陆地线与雪线之间范围的云覆盖不进行分类<sup>[30]</sup>.雪线法去除精度相对较高,但是,雪线法的除云精度会随着云覆盖比例的增加而降低<sup>[31]</sup>.除此之外,区域雪线法也常用于去除一定海拔范围内的云覆盖,该方法中的陆地线和雪线是由所有陆地像元以及所有积雪像元海拔高度的平均获得<sup>[32]</sup>,相较于极值雪线法,该方法不适用于本研究区域,因为均值雪线法极易受到高海拔与低海拔地区无云像元的影像,从而使结果精度大大降低.极值雪线法公式如下:

$$S(x, y, t) = 0, \quad H(x, y) < H_{\min}(t), \quad (1)$$

式中:  $S(x, y, t)$  为当天( $t$ )的  $x, y$  位置上像元的积雪覆盖,即  $S(x, y, t) = 0$  时为陆地;  $H(x, y)$  为  $x, y$  位置上像元的海拔高度;  $H_{\min}(t)$  为当天( $t$ )的陆地线阈值,即积雪覆盖像元的最低海拔高度.考虑到本研究

地区主要位于高海拔地带,为了减少对高海拔的雪数据缺失像元误分类为积雪,对雪线则设定为最高海拔.本文的方法是找到雪覆盖的最低海拔高度,即设定为当前单景数据的陆地线.

### 3.2 基于降噪自编码神经网络的雪粒径回归模型

降噪自编码神经网络是一种无监督的前馈神经网络学习模型<sup>[33]</sup>,其输出是输入的非线性映射,并且能够对数据的深层特征进行自动提取,实现数据特征降维处理以及自主学习特征映射关系<sup>[34]</sup>.自编码神经网络具有三层结构,分别是输入层、隐藏层和输出层.第一层包含 4 个节点,每个节点均与每个输入相连;以此类推,下一层与上一层相连.由输入层到隐藏层是编码过程,其将输入  $\mathbf{x}(t)$  映射到隐空间,即编码为

$$\mathbf{h} = \delta(\mathbf{W}\mathbf{x} + \mathbf{b}). \quad (2)$$

由隐藏层到输出层是解码过程,解码过程是将潜在向量  $\mathbf{x}$  作为输入从低维空间中重建输入,即解码为

$$\hat{\mathbf{x}} = \delta'(\mathbf{W}'\mathbf{x} + \mathbf{b}'), \quad (3)$$

式中:  $\delta, \delta'$  分别为编码与解码过程的激活函数;  $\mathbf{W}, \mathbf{W}'$  分别为编码与解码过程的权重;  $\mathbf{b}, \mathbf{b}'$  为解码过程的偏置;  $\mathbf{x}$  作为输入,  $\hat{\mathbf{x}}$  是解码的输出,可以认为是输入  $\mathbf{x}$  的预测值.根据经验与输入数据特点,本文采用 Relu 激活函数,相较于 Sigmoid 激活函数收敛速度更快,使神经网络具备了非线性特征学习能力.本文降噪自编码神经网络模型包括多特征数据和神经网络两个主要部分,基本结构如图 2 所示.

采用深层自编码器(Deep Autoencoder, DAE),在编码器和解码器之间堆叠更多隐藏层,网络模型输入数据包括雪粒径、高程、坡向、坡度和地类数据样本.从输入到输出的维度分别为  $256 \rightarrow 128 \rightarrow 64 \rightarrow 128 \rightarrow 256$ .自编码器通常会学习到低维度数据信息,该低维度信息通常可以提供更好的输入表示<sup>[35]</sup>,进行特征选择和特征提取,去掉数据集中夹杂的噪声,在相对较小的数据集上有较高的鲁棒性<sup>[36]</sup>.同时为了防止模型过拟合,将模型的 dropout 参数设置为 0.4,这可以通过阻止神经元的激活状态来提高神经网络的性能<sup>[37]</sup>.模型训练参数的批处理大小为 1 000,最大训练次数为 200 个 Epoch.根据经验选择学习率为 0.000 1,使用 Adam 优化算法对网络参数进行优化.

对于训练过程中常采用的损失函数是均方根误差(RMSE,其量值记  $\eta_{\text{RMSE}}$ )、平均绝对误差(MAE,

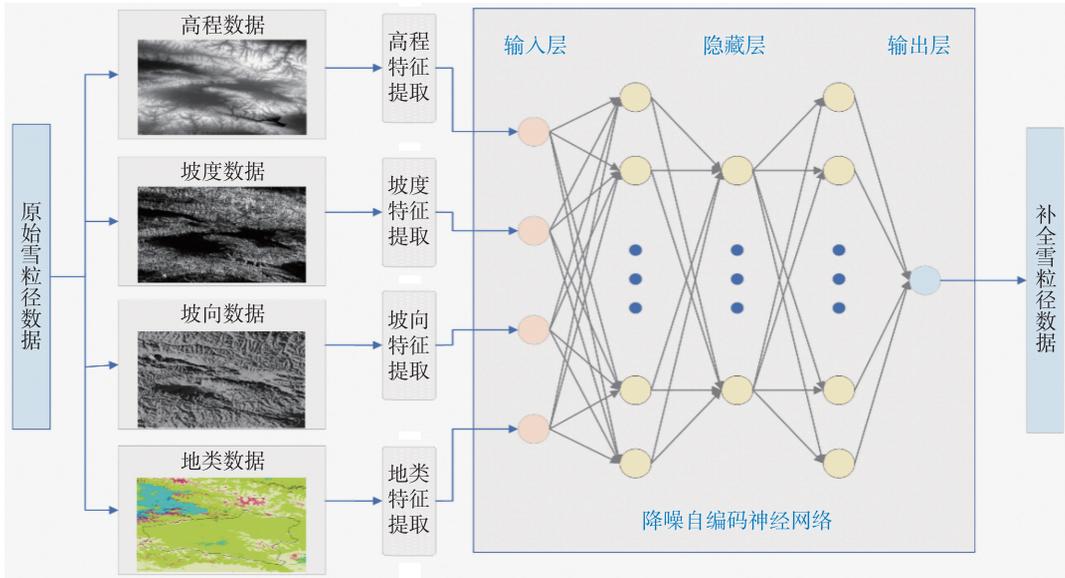


图2 降噪自编码神经网络模型

Fig. 2 Denoising autoencoder artificial neural network model

其量值记为  $\eta_{MAE}$ ), 各计算公式为

$$\eta_{RMSE} = \sqrt{\frac{1}{N} \sum_{n=1}^N \|x_n - \hat{x}_n\|^2}, \quad (4)$$

$$\eta_{MAE} = \frac{1}{N} \sum_{n=1}^N \|x_n - \hat{x}_n\|, \quad (5)$$

式中:  $n$  为观测样本数;  $x_n$  为模型的模拟值;  $\hat{x}_n$  为模型的真实观测值. 当 RMSE 和 MAE 的值均为零时, 认为模型的模拟性最好. 通过不断迭代训练自编码神经网络来学习和更新权重矩阵和偏置向量, 将给定的输入向量  $x$  与估计向量  $\hat{x}$  进行计算比较, 通过不断迭代使损失函数达到最小化来完成神经网络的学习过程.

## 4 结果与分析

### 4.1 积雪产品去云结果

#### 4.1.1 单景雪粒径去云结果

为了验证本文方法对缺失雪粒径数据估计效果, 选择云量较少且大范围出现降雪天气的一天进行研究, 以 2018 年 5 月 25 日为例. 按照训练集 80% 占比以及测试集 20% 占比, 对降噪自编码神经网络进行网络训练, 训练过程如图 3 所示.

从模型的训练误差曲线与测试损失曲线可以看出, 在经过 200 次迭代 (Epoch) 训练后, 模型的训练损失与测试损失达到较低且波动较小, 网络达到了最佳的训练结果且模型收敛速度很快, 说明本文的降噪自编码神经网络能够有效地建立多特征输入数

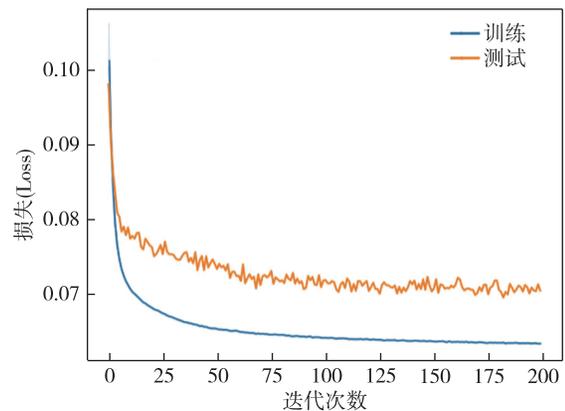


图3 神经网络训练与测试损失曲线

Fig. 3 Training and testing loss curves of neural network

据与雪粒径之间的映射关系以达到较高的去云精度. 在经过 3.1 节的极值雪线法处理后, 将模型输出按照预定的栅格位置信息填充到原始雪粒径影像, 如图 4 所示, 最终得到完整的雪粒径数据影像.

图 4a 与 4b 分别是雪粒径去云前后的数据影像, 其中数据包括云覆盖 (黑色)、积雪覆盖 (雪粒径大小范围从  $0 \sim 1000 \mu\text{m}$ ) 以及陆地部分 (绿色). 与原始的 MODIS 雪粒径产品相比, 经过本文的降噪自编码神经网络以及极值雪线法处理后, 有 10.6% 的缺失数据被分类为陆地, 积雪的覆盖率增加了 10.15%, 积雪覆盖有了明显增加.

为了验证本文所提方法的有效性, 在相同数据集的情况下, 分别与回归决策树 (Classification And

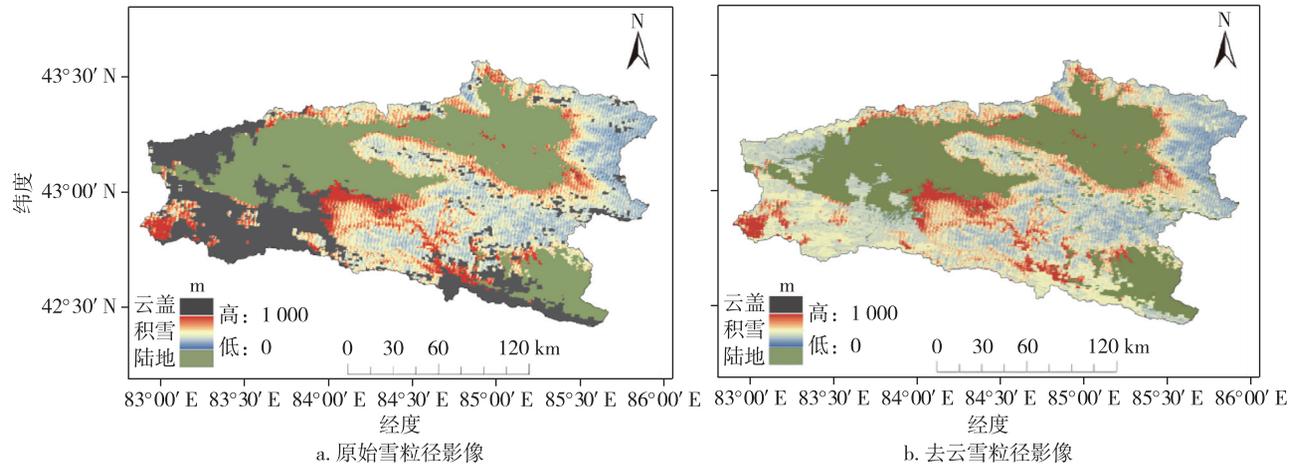


图4 2018年5月25日雪粒径去云实验结果

Fig. 4 Raw data (a) and cloud removed (b) snow grain size images over the Kaidu River basin on May 25, 2018

Regression Tree, CART)<sup>[38]</sup>、K最近邻(K-Nearest Neighbor, KNN)<sup>[39]</sup>、随机森林(Random Forest, RF)<sup>[40]</sup>、岭回归(Ridge Regression, RR)<sup>[41]</sup>以及支持向量回归(Support Vector Regression, SVR)<sup>[42]</sup>进行试验对比.表2列出了在本文所使用的相同数据集上,各种回归模型方法的均方根误差(RMSE)与平均绝对误差(MAE).

表2 在相同数据集上不同回归模型方法的评估指标

Table 2 Evaluation indicators of different regression models on the same dataset

回归模型	RMSE	MAE
CART <sup>[38]</sup>	30.621	22.470
KNN <sup>[39]</sup>	25.698	19.132
RF <sup>[40]</sup>	26.236	19.595
RR <sup>[41]</sup>	25.103	19.821
SVR <sup>[42]</sup>	24.952	18.867
本文方法	21.674	16.340

从表2中的对比结果可以看出,6种方法的模型精度有所不同,其中回归决策树的误差最大.因为回归决策树在模型训练过程中容易发生过拟合现象,且在进行属性划分时,不同的判别准则会导致不同的属性倾向而使精度降低.相比较其他5种机器学习方法,本文所提方法对雪粒径数据的去云结果整体上明显改善,且RMSE与MAE均最低.以上结果表明,本文所提出的降噪自编码神经网络方法在对雪粒径去云效果上优于传统的机器学习方法.

#### 4.1.2 2000—2019年积雪覆盖分析

本文选取实验数据时,在训练阶段按照不同数据缺失率进行处理,训练集与测试集比例为8:2.在

实验中,模型的估计值与真实值之间的误差越小,说明模型回归效果越好.实验选取测试误差(Test Loss)为MAE,作为主要评价指标对模型的回归效果进行评价,实验结果如图5所示.

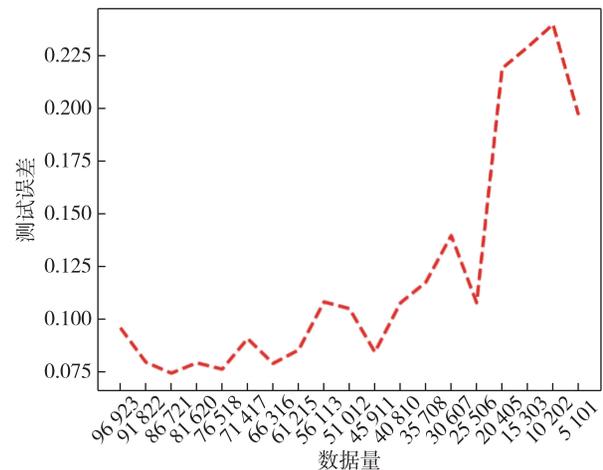


图5 不同数据量情况下,训练阶段估计数据的测试误差

Fig. 5 Test loss of data estimated during training under different missing rates

由实验结果可以看出,随着数据量的降低,整体的测试误差呈现上升趋势,在数据量为25 506及以下时训练阶段的测试误差过大.考虑到模型回归结果受限于雪盖率大小,为了提高长时间序列积雪覆盖分析的准确率,本文通过模型训练分析筛选积雪数据量大于25 506的单景数据进行估计处理.由于无法获取部分原始MODIS数据,导致2000—2019年部分数据缺失,缺失数据天数如表3所示.

图6显示了在全年雪粒径数据基础上获得的2000—2019年积雪日数的空间分布情况(数据从

2000年6月1日到2020年5月7日).为了便于对降雪进行研究,选取一个雪年作为时间单位,即从当年6月1日至次年5月31日.

通过对积雪产品去云前后雪空间分布的对比,可以看出2010年是积雪覆盖日数最多的一年.开都河流域中部所处的天山山脉的山间盆地,即被阿尔宾山分开的两个盆地,包括小尤尔都斯盆地、大尤尔都斯盆地,在盆地地区的积雪日数整体相对较少,盆地周围被海拔范围在2 500~4 000 m的山脉所环绕,而海拔较高的山脉地区的积雪日数相对较多.每年积雪日数在200 d以上的地区主要分布在北部的天山主脉的依连哈比尔尕山、中部的阿尔宾山和南部的库鲁克塔格山3 000~4 600 m的高海拔范围.

去云雪粒径数据集提供了包括降雪量和雪覆盖天数在内的不同的估计数据,这对水文和气候模型的研究具有很大的作用.为了进一步分析开都河流域积雪覆盖特征,对图6中原始MODIS与去云结果进一步研究,分别得到2000—2019年开都河流域全范围的年最大、最小以及平均积雪日数,结果如表3

所示.

从表3中可以看出,本文方法在对积雪的去云处理后,年平均积雪日数有明显增加,较原有有云MODIS数据增加了38~53 d.

对原始MODIS雪粒径数据集与本文的去云雪粒径数据集,在年平均积雪覆盖范围进行比较,时间范围是2000—2019年,结果表明,原始MODIS雪粒径数据集和去云雪粒径数据集在积雪覆盖范围上有很大的不同.如图7所示,本文研究方法所得去云数据的雪覆盖率明显大于原始MODIS数据的雪覆盖率,去云数据的雪覆盖基本保持在30%~37%,而原始MODIS数据的积雪覆盖率则仅为13%~17%.前后数据对比表明,使用本文去云雪粒径数据集对积雪量化方面的研究具有重要的参考价值.

#### 4.2 结果验证

由于开都河流域气象站稀少,仅有一个位于天山山脉中部的巴音布鲁克气象站,来自该区域的地面观测气象数据有限,因此在验证时采用了MODIS的积雪覆盖产品.本文选择2018年观测云量较少的

表3 原始MODIS雪粒径产品和本文去云雪粒径产品的年最大、最小、平均积雪日数比较

Table 3 Comparison of annual maximum, minimum and average snow days between original MODIS product and the proposed cloud-removal snow grain size product

年份	缺失数据量	原始 MODIS			去云		
		最小年积雪日数/d	最大年积雪日数/d	年平均积雪日数/d	最小年积雪日数/d	最大年积雪日数/d	年平均积雪日数/d
2000	131	0	169	44	33	225	87
2001	138	0	181	39	32	226	83
2002	121	0	199	33	32	239	76
2003	115	0	197	37	34	246	80
2004	109	0	195	39	43	253	84
2005	124	0	194	44	40	238	88
2006	126	0	188	34	29	231	73
2007	108	0	202	37	39	253	83
2008	124	0	191	36	35	236	84
2009	117	0	185	39	29	238	80
2010	120	0	196	53	45	239	92
2011	119	0	194	35	48	243	88
2012	138	0	198	36	28	224	77
2013	120	0	193	43	45	238	83
2014	126	0	176	44	36	234	89
2015	147	0	181	39	36	213	77
2016	137	0	183	41	47	228	90
2017	137	0	195	37	28	224	79
2018	129	0	180	41	43	236	92
2019	107	0	186	41	29	234	85

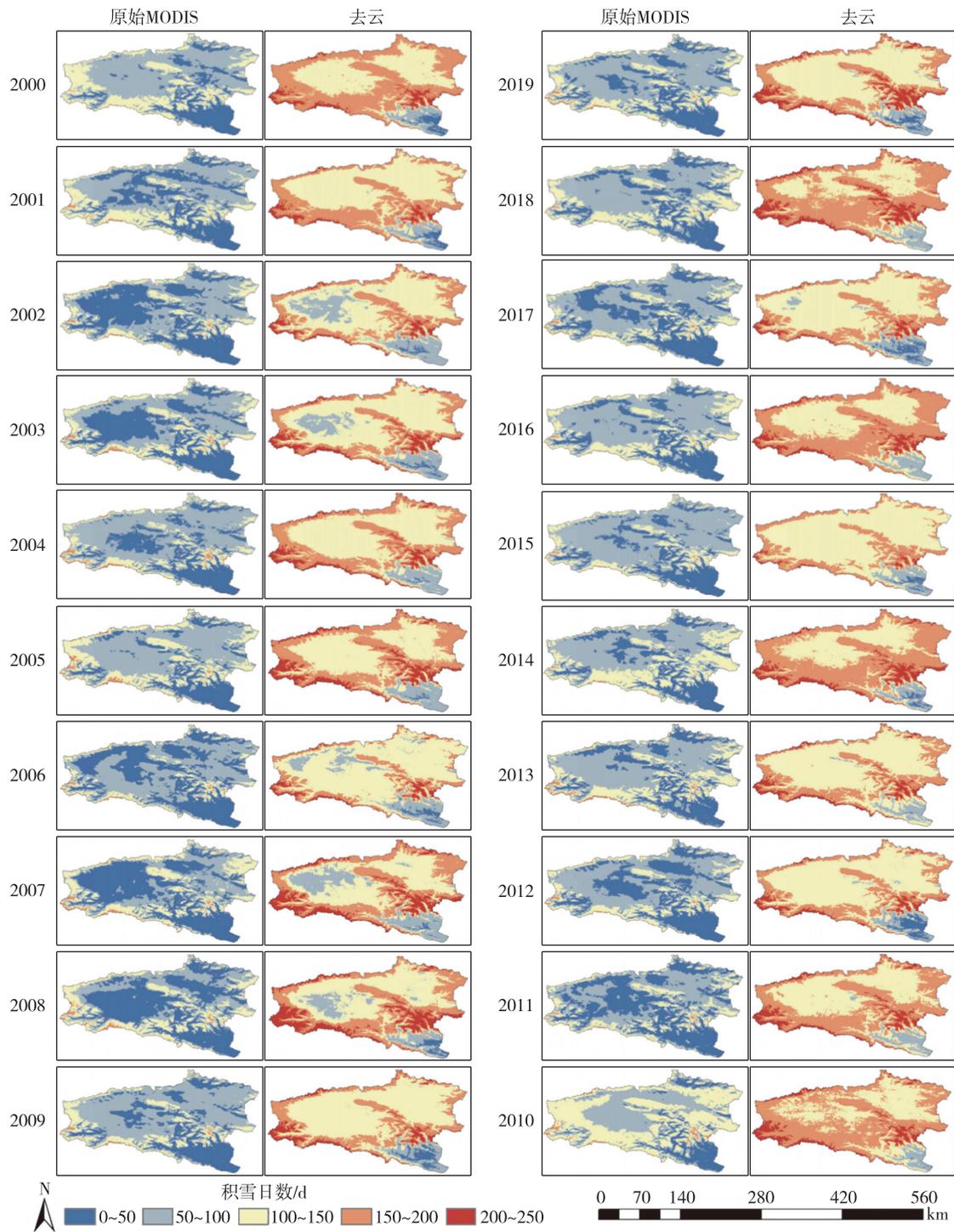


图 6 原始 MODIS 积雪产品和本文去云雪粒径产品年积雪日数比较

Fig. 6 Comparison of annual snow days between original MODIS product and the proposed cloud-removal snow grain size product

4 天作为验证对象.在验证时,采用本文方法对云覆盖导致的积雪缺失数据进行处理,分别从正确分类以及错误分类的结果上进行验证.正确分类包括积雪分类为积雪、陆地分类为陆地;错误分类包括积雪分类为陆地、陆地分类为积雪.采用本文所提出的方法对所有的云覆盖导致的缺失数据进行去云补全处

理,结果如表 4 所示.

图 8 显示了 2000—2019 年 Terra 卫星 MOD、Aqua 卫星 MYD 的雪粒径数据的月平均积雪覆盖率,以及采用日结合方法和本文方法之后的月平均积雪覆盖率.原始的 MODIS 积雪产品的月平均雪盖率与季节因素有较大的相关性,冬季的积雪覆盖率

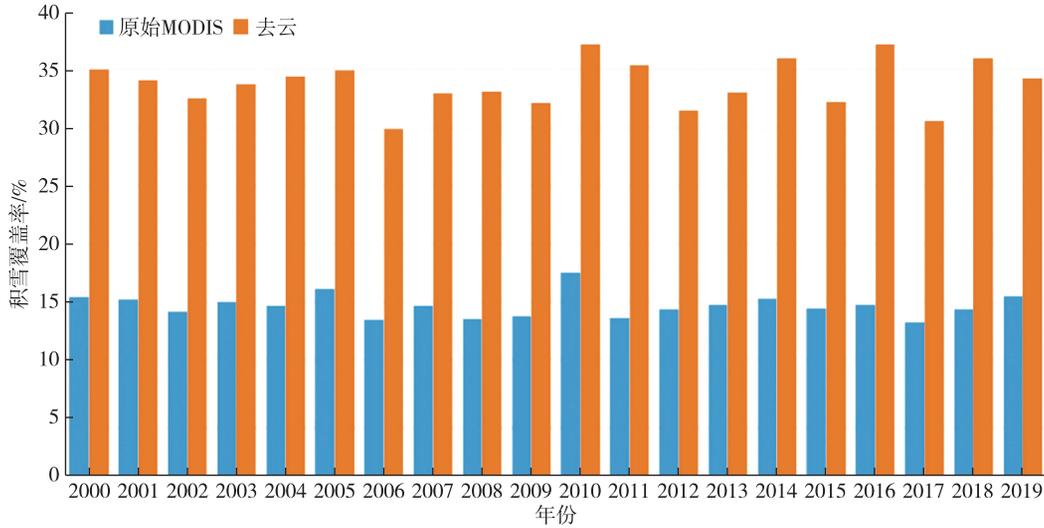


图7 原始合并 MODIS 和去云积雪产品年平均积雪覆盖率比较

Fig. 7 Comparison of annual snow days between original merged MODIS product and the proposed cloud-removal snow cover product

表4 去云的分类结果验证

Table 4 Verification of classification results after cloud removal

日期	云覆盖率	数据补全率	正确分类		错误分类	
			积雪→积雪	陆地→陆地	积雪→陆地	陆地→积雪
2018-03-21	12.93	100	87.34	3.63	2.01	7.02
2018-04-12	16.66	100	86.75	3.41	2.76	7.08
2018-05-25	21.40	100	76.21	9.82	4.21	9.76
2018-10-25	22.32	100	70.20	16.08	5.29	8.43

最高而夏季的积雪覆盖率最低.可以看出 Terra 卫星 MOD 数据的月平均积雪覆盖率高于 Aqua 卫星 MYD, 主要原因是 Aqua 卫星的过境时间晚于 Terra 卫星, 并且在白天随着时间的推移云覆盖在渐渐增加.本文方法相较于原始 MOD 数据, 月平均积雪覆盖率增加了 3.65%~50.35%, 尤其是在 11 月至次年 3 月期间, 积雪覆盖率有明显增加.

### 5 结论

本研究的目的是从原始 MODIS 雪粒径产品中去除云覆盖区域, 以获取开都河流域的去云积雪信息, 从而提出一种适用于高山复杂地形区域的积雪去云模型.本文提出了降噪自编码神经网络与极值雪线法相结合的模型来从 MODIS 积雪产品中去除云覆盖, 并对结果进行了验证.本文主要通过无监督的降噪自编码神经网络来提取地形数据中能够映射出原始雪粒径的抽象的特征, 对训练过程中误差进行分析, 模型拟合精度较高、误差较小.同时, 本文使用单景数据进行数据集划分与模型训练, 克服了开都河流域气象站观测数据受限导致真值标签不足的问题.

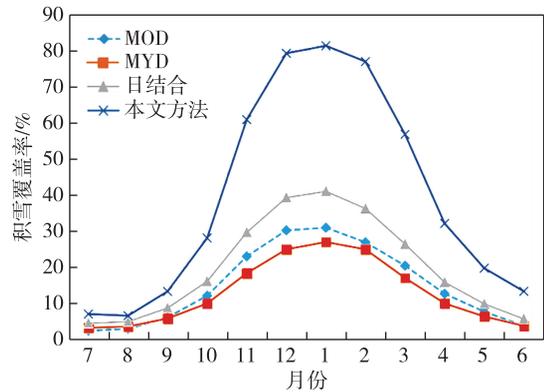


图8 2000—2019 年原始数据产品以及不同处理方法的月平均积雪覆盖率比较

Fig. 8 Comparison of monthly average snow cover rate among raw data products and different processing methods for period 2000–2019

问题.经过验证表明, 与其他去云方法相比, 本文方法对云覆盖的去除效果表现更优, 效果更加稳健.采用本文方法处理 2000—2019 年包含云覆盖的原始 MODIS 积雪产品, 即可以制作 500 m 空间分辨率的去云雪粒径数据.利用本文所提方法, 可以估计开都

河流域积雪的动态变化,对认识该区域积雪的长期变化具有重要的意义.

## 参考文献

### References

- [ 1 ] Singh P, Arora M, Goel N K. Effect of climate change on runoff of a glacierized Himalayan basin [ J ]. *Hydrological Processes*, 2006, 20( 9 ): 1979-1992
- [ 2 ] Kling H, Fürst J, Nachtnebel H P. Seasonal, spatially distributed modelling of accumulation and melting of snow for computing runoff in a long-term, large-basin water balance model [ J ]. *Hydrological Processes*, 2006, 20( 10 ): 2141-2156
- [ 3 ] Gao J L, Huang X D, Ma X F, et al. Snow disaster early warning in pastoral areas of Qinghai province, China [ J ]. *Remote Sensing*, 2017, 9( 5 ): 475
- [ 4 ] 李良序, 黄镇, 傅华. 新疆雪盖特征分析 [ J ]. *新疆气象*, 2001( 5 ): 21-23  
LI Liangxu, HUANG Zhen, FU Hua. Snow cover feature in Xinjiang [ J ]. *Bimonthly of Xinjiang Meteorology*, 2001( 5 ): 21-23
- [ 5 ] 蓝永超, 钟英君, 吴素芬, 等. 天山南、北坡河流出山径流对气候变化的敏感性分析: 以开都河与乌鲁木齐河出山径流为例 [ J ]. *山地学报*, 2009, 27( 6 ): 712-718  
LAN Yongchao, ZHONG Yingjun, WU Sufen, et al. Sensitivity of mountain runoff of rivers originated from the south slope and the north slope of the Tianshan Mountain to climate change: taking mountain runoff of Urumqi river and Kaidu river for example [ J ]. *Journal of Mountain Science*, 2009, 27( 6 ): 712-718
- [ 6 ] Parajka J, Blöschl G. Validation of MODIS snow cover images over Austria [ J ]. *Hydrology and Earth System Sciences*, 2006, 10( 5 ): 679-689
- [ 7 ] Wang X W, Xie H J, Liang T G. Evaluation of MODIS snow cover and cloud mask and its application in northern Xinjiang, China [ J ]. *Remote Sensing of Environment*, 2008, 112( 4 ): 1497-1513
- [ 8 ] Tran H, Nguyen P, Ombadi M, et al. A cloud-free MODIS snow cover dataset for the contiguous United States from 2000 to 2017 [ J ]. *Scientific Data*, 2019, 6: 180300
- [ 9 ] Liang T G, Zhang X T, Xie H J, et al. Toward improved daily snow cover mapping with advanced combination of MODIS and AMSR-E measurements [ J ]. *Remote Sensing of Environment*, 2008, 112( 10 ): 3750-3761
- [ 10 ] Gao Y, Xie H J, Lu N, et al. Toward advanced daily cloud-free snow cover and snow water equivalent products from Terra-Aqua MODIS and Aqua AMSR-E measurements [ J ]. *Journal of Hydrology*, 2010, 385( 1/2/3/4 ): 23-35
- [ 11 ] 包慧漪, 张显峰, 廖春华, 等. 基于 MODIS 与 AMSR-E 数据的新疆雪情参数协同反演研究 [ J ]. *自然灾害学报*, 2013, 22( 4 ): 41-49  
BAO Huiyi, ZHANG Xianfeng, LIAO Chunhua, et al. Co-inversion of snow parameters in Xinjiang based on MODIS and AMSR-E data [ J ]. *Journal of Natural Disasters*, 2013, 22( 4 ): 41-49
- [ 12 ] Poggio L, Gimona A. Sequence-based mapping approach to spatio-temporal snow patterns from MODIS time-series applied to Scotland [ J ]. *International Journal of Applied Earth Observation and Geoinformation*, 2015, 34: 122-135
- [ 13 ] Parajka J, Pepe M, Rampini A, et al. A regional snow-line method for estimating snow cover from MODIS during cloud cover [ J ]. *Journal of Hydrology*, 2010, 381( 3/4 ): 203-212
- [ 14 ] Hall D K, Riggs G A, Foster J L, et al. Development and evaluation of a cloud-gap-filled MODIS daily snow-cover product [ J ]. *Remote Sensing of Environment*, 2010, 114( 3 ): 496-503
- [ 15 ] Li X H, Fu W X, Shen H F, et al. Monitoring snow cover variability ( 2000 - 2014 ) in the Hengduan Mountains based on cloud-removed MODIS products with an adaptive spatio-temporal weighted method [ J ]. *Journal of Hydrology*, 2017, 551: 314-327
- [ 16 ] Jain S K, Goswami A, Saraf A K. Role of elevation and aspect in snow distribution in western Himalaya [ J ]. *Water Resources Management*, 2009, 23( 1 ): 71-83
- [ 17 ] Kour R, Patel N, Krishna A P. Effects of terrain attributes on snow-cover dynamics in parts of Chenab basin, western Himalayas [ J ]. *Hydrological Sciences Journal*, 2016, 61( 10 ): 1861-1876
- [ 18 ] 侯海艳, 侯金亮, 黄春林, 等. 基于人工神经网络和 AMSR2 多频微波亮温的北疆地区雪深反演 [ J ]. *遥感技术与应用*, 2018, 33( 2 ): 241-251  
HOU Haiyan, HOU Jinliang, HUANG Chunlin, et al. Retrieve snow depth of north of Xinjiang region from ARMS2 data based on artificial neural network technology [ J ]. *Remote Sensing Technology and Application*, 2018, 33( 2 ): 241-251
- [ 19 ] 张一驰, 李宝林, 程维明, 等. 开都河流域径流对气候变化的响应研究 [ J ]. *资源科学*, 2004, 26( 6 ): 69-76  
ZHANG Yichi, LI Baolin, CHENG Weiming, et al. Hydrological response of runoff to climate variation in Kaidu catchment [ J ]. *Resources Science*, 2004, 26( 6 ): 69-76
- [ 20 ] 杨建国, 杨建伟, 王振龙, 等. 开都河流域水文特性分析 [ J ]. *水利科技与经济*, 2012, 18( 7 ): 10-13  
YANG Jianguo, YANG Jianwei, WANG Zhenlong, et al. Hydrological characteristics of Kaidu river basin [ J ]. *Water Conservancy Science and Technology and Economy*, 2012, 18( 7 ): 10-13
- [ 21 ] 姜腾龙, 赵书河, 肖鹏峰, 等. 基于实测数据的不同雪粒径光谱分析 [ J ]. *冰川冻土*, 2009, 31( 2 ): 227-232  
JIANG Tenglong, ZHAO Shuhe, XIAO Pengfeng, et al. Spectral analysis of different snow grain sizes based on field measurement [ J ]. *Journal of Glaciology and Geocryology*, 2009, 31( 2 ): 227-232
- [ 22 ] Painter T H, Rittger K, McKenzie C, et al. Retrieval of subpixel snow covered area, grain size, and albedo from MODIS [ J ]. *Remote Sensing of Environment*, 2009, 113( 4 ): 868-879
- [ 23 ] 郭忠明, 王宁练, 毛瑞娟, 等. 基于 MODIS 反演祁连山七一冰川雪粒径 [ J ]. *地理科学*, 2013, 33( 3 ): 378-384

- GUO Zhongming, WANG Ninglian, MAO Ruijuan, et al. Retrieved of snow grain size from MODIS over Qiyi glacier, Qilian Mountain [J]. *Scientia Geographica Sinica*, 2013, 33(3): 378-384
- [24] 魏亚瑞,郝晓华,王建,等.基于MODIS数据的北疆积雪黑碳和雪粒径反演及时空变化分析[J].*冰川冻土*, 2019, 41(5): 1192-1204  
WEI Yarui, HAO Xiaohua, WANG Jian, et al. Retrieval and analysis of spatiotemporal variation of snow black carbon and snow grain size in northern Xinjiang based on MODIS data[J]. *Journal of Glaciology and Geocryology*, 2019, 41(5): 1192-1204
- [25] Wang J G, Feng X Z, Xiao P F, et al. Snow grain-size estimation over mountainous areas from MODIS imagery [J]. *IEEE Geoscience and Remote Sensing Letters*, 2018, 15(1): 97-101
- [26] 张彦丽,张丽萍.基于Sentinel-2时序数据的山区积雪识别与面积变化[J].*生态学杂志*, 2020, 39(8): 2810-2820  
ZHANG Yanli, ZHANG Liping. Snow cover identification and area change in mountainous regions based on Sentinel-2 time series data [J]. *Chinese Journal of Ecology*, 2020, 39(8): 2810-2820
- [27] 肉克亚木·艾克木,玉素甫江·如素力,玛地尼亚提·地里夏提. 博斯腾湖流域积雪的时空变化特征及其与气候因子的关系[J].*水生态学杂志*, 2020, 41(4): 9-18  
ROUKEYAMU Aikemu, YUSUFUJIANG Rusuli, MADINIYATI Dilixiati. Spatiotemporal changes of snow cover in the Bosten Lake basin and its relationship with climate factors[J]. *Journal of Hydroecology*, 2020, 41(4): 9-18
- [28] 买托合提·阿那依提,玉素甫江·如素力,麦麦提吐尔逊·艾则孜,等.新疆开都河流域主要地貌形态特征研究[J].*冰川冻土*, 2014, 36(5): 1160-1166  
MATOHTI Anayit, YUSUFUJIANG Rusuli, MAMAT-TURSUN Eziz, et al. Study of the landform features in the Kaidu River basin of Xinjiang, China [J]. *Journal of Glaciology and Geocryology*, 2014, 36(5): 1160-1166
- [29] Tong J, Déry S J, Jackson P L. Topographic control of snow distribution in an alpine watershed of western Canada inferred from spatially-filtered MODIS snow products [J]. *Hydrology and Earth System Sciences*, 2009, 13(3): 319-326
- [30] Gafurov A, Bárdossy A. Cloud removal methodology from MODIS snow cover product [J]. *Hydrology and Earth System Sciences*, 2009, 13(7): 1361-1373
- [31] Li X H, Jing Y H, Shen H F, et al. The recent developments in cloud removal approaches of MODIS snow cover product [J]. *Hydrology and Earth System Sciences*, 2019, 23(5): 2401-2416
- [32] Dariane A B, Khoramian A, Santi E. Investigating spatiotemporal snow cover variability via cloud-free MODIS snow cover product in central Alborz region [J]. *Remote Sensing of Environment*, 2017, 202: 152-165
- [33] Takaki S, Yamagishi J. A deep auto-encoder based low-dimensional feature extraction from FFT spectral envelopes for statistical parametric speech synthesis [C]//2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). March 20 - 25, 2016, Shanghai, China. IEEE, 2016: 5535-5539
- [34] 刘自然,李谦,颜丙生,等.堆叠稀疏自编码深度神经网络算法及其在滚动轴承故障诊断中的应用[J].*机床与液压*, 2020, 48(23): 208-213  
LIU Ziran, LI Qian, YAN Bingsheng, et al. Application of depth neural network algorithm with stacked sparse auto-encoder in rolling bearing fault diagnosis [J]. *Machine Tool & Hydraulics*, 2020, 48(23): 208-213
- [35] Xu Q Y, Wu Z, Yang Y Q, et al. The difference learning of hidden layer between autoencoder and variational autoencoder [C]//2017 29th Chinese Control and Decision Conference (CCDC). May 28 - 30, 2017, Chongqing, China. IEEE, 2017: 4801-4804
- [36] 胡海洋,张力,李忠金.融合自编码器和 one-class SVM 的异常事件检测[J].*中国图象图形学报*, 2020, 25(12): 2614-2629  
HU Haiyang, ZHANG Li, LI Zhongjin. Anomaly detection with autoencoder and one-class SVM [J]. *Journal of Image and Graphics*, 2020, 25(12): 2614-2629
- [37] Hinton G E, Srivastava N, Krizhevsky A, et al. Improving neural networks by preventing co-adaptation of feature detectors [J]. *Computer Science*, 2012, 3(4): 212-223
- [38] Zekić-Sušac M, Has A, Knežević M. Predicting energy cost of public buildings by artificial neural networks, CART, and random forest [J]. *Neurocomputing*, 2021, 439: 223-233
- [39] Barrash S, Shen Y N, Giannakis G B. Scalable and adaptive KNN for regression over graphs [C]//2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP). December 15 - 18, 2019, Le Gosier, Guadeloupe. IEEE, 2019: 241-245
- [40] 任家东,刘新倩,王倩,等.基于KNN离群点检测和随机森林的多层入侵检测方法[J].*计算机研究与发展*, 2019, 56(3): 566-575  
REN Jiadong, LIU Xinqian, WANG Qian, et al. A multi-level intrusion detection method based on KNN outlier detection and random forests [J]. *Journal of Computer Research and Development*, 2019, 56(3): 566-575
- [41] Rumere F O, Soemartojo S M, Widyaningsih Y. Restricted ridge regression estimator as a parameter estimation in multiple linear regression model for multicollinearity case [J]. *Journal of Physics: Conference Series*, 2021, 1725: 012021
- [42] Al-Qatf M, Yu L S, Al-Habib M, et al. Deep learning approach combining sparse autoencoder with SVM for network intrusion detection [J]. *IEEE Access*, 2018, 6: 52843-52856

## Cloud removal for snow products based on denoising autoencoder artificial neural network

ZHANG Yonghong<sup>1</sup> CHEN Shuai<sup>1</sup> WANG Jiangeng<sup>2</sup> ZHU Linglong<sup>3</sup> CHEN Shiwei<sup>1</sup>

1 School of Automation,Nanjing University of Information Science & Technology,Nanjing 210044

2 School of Atmospheric Physics,Nanjing University of Information Science & Technology,Nanjing 210044

3 School of Electronics & Information Engineering,Nanjing University of Information Science & Technology,Nanjing 210044

**Abstract** Snow cover is one of the important parameters in the study of hydrometeorology. At present, the most widely used Snow Cover Area (SCA) can be obtained by Moderate-resolution Imaging Spectroradiometer (MODIS), which is often used in the study of temporal and spatial changes of snow cover. However, large area snow data missing existed in MODIS snow cover products due to the cloud occlusion. To address this, we take the Kaidu River basin as the research region, and combine the snow product data retrieved from MODIS carried on the Terra and Aqua satellites with the topographic feature data, then use a denoising autoencoder artificial neural network and the extreme snow line method to quantitatively complement the snow data loss caused by cloud occlusion in complex alpine terrain. The denoising autoencoder artificial neural network combines multi-feature data to establish a nonlinear mapping relationship between topographic features and snow grain size, which is then used to complement the missing snow grain size data. The extreme snow line method is used to remove the false report value in low altitude area and obtain the snow cover image with high precision. In contrast verification, the accuracy of the proposed cloud removal method is over 86%, which effectively improves the snow cover detection. Therefore, the approach proposed in this paper can effectively remove cloud occlusion from snow products in complex terrain areas.

**Key words** denoising autoencoder artificial neural network; extreme snow line method; complex terrain; cloud removal