环境遥感

基于辐射传输模拟的神经网络云罩算法

云探测和筛选是获得许多卫星数据产品所需的至关重要的第一步。传统的基于阈值的云面具算法需要一个复杂的设计过程和微调每个传感器，他们有困难的地区部分覆盖的雪/冰。利用机器学习技术和耦合环境系统辐射传输建模的进展，我们开发了一种新的、无阈值的云面具算法，该算法基于广泛的辐射传输模拟驱动的神经网络分类器。CALIOP与MODIS数据配置的统计验证结果表明，其在不同生态系统上的性能是一致的，在中纬度积雪地区的冬季，其性能明显优于MODIS云罩(MOD35 C6)。使用较少的卫星信道数的仿真也显示了令人满意的结果，表明了该算法对于不同传感器配置的灵活性。与基于阈值的方法和以往的机器学习方法相比，这种新的云罩(i)不依赖于阈值，(ii)需要较少的卫星通道，(iii)在中纬度地区的冬季有更好的性能，(iv)可以方便地应用于不同的传感器。

1. 介绍

1.1。背景

可靠的云罩对于卫星遥感陆地、海洋或冰冻圈特性是必不可少的。由于云对短波和长波辐射的显著影响，将云像元错识别为地表或将云像元错识别为地表都会显著影响任何卫星遥感产品的质量。传统的云遮罩算法大多采用基于阈值的测试方法。这些算法包括自动云层评估(ACCA)算法(爱尔兰et al ., 2006)应用于陆地卫星ETM +传感器,云测试应用于MOD35算法(阿克曼et al ., 2010)中分辨率成像光谱仪(MODIS)传感器和AVHRR的云(CLAVR)(斯托et al ., 1999)以及其扩展CLAVR-x算法。这些算法通常使用阈值的组合测试,雇佣许多卫星频道位于可见(VIS),近红外(NIR)、短波红外(短波红外成像)和热红外(行动)波长范围(例如MOD35使用19乐队- 10反射率乐队和9热红外波段)探测云和雪/冰。这些测试中使用的阈值通常来自1)模型模拟，2)云/晴空场景的统计，3)专家经验。新算法，如fmask (Zhu and Woodcock, 2012;Zhu et al.， 2015)，采用基于对象的云和云阴影统计数据的动态阈值。在我们之前的工作(Chen et al.， 2014)中，我们开发了一种基于模型的动态阈值方法，并对其进行了测试，结果表明在冰雪覆盖的格陵兰高原上，与MODIS MOD35算法相比，该方法具有更好的性能。

由于云与雪/冰在可见光和近红外通道的光学特性相似，雪的检测一直是云掩膜算法设计的关键。利用VIS和SWIR数据绘制积雪覆盖指数是在20世纪70年代中期开发的。

Hall等(1995)引入归一化差分雪指数(NDSI)，利用MODIS数据绘制雪图。在此之前，Dozier(1987,1989)使用VIS/SWIR指数算法根据Landsat数据绘制雪图。大多数基于阈值的云掩码算法将在其处理链中使用NDSI (Ackerman et al.， 1998,2010;Irish等人，2006年;Zhu and Woodcock, 2012)，强调了雪探测的重要性，因为它的准确性也会影响云探测的准确性。

增强的计算能力和机器学习技术的改进使得机器学习算法，如决策树、逻辑回归、支持向量机和人工神经网络，可以用于云遮蔽和冰雪检测。Taravat等人(2015)使用多层感知器神经网络模型检测Landsat图像中的云。Hollstein等人(2016)比较了几种方法，包括决策树、经典贝叶斯、随机森林、支持向量机和随机梯度下降，应用于Sentinel-2多光谱仪器(MSI)图像。Hughes和Hayes(2014)使用美国地质调查局Landsat数据连续性任务(USGS LDCM)云覆盖评估数据子集训练的基于神经网络的方法(Scaramuzza et al.， 2012)，并与fmask (Zhu and Woodcock, 2012)进行比较，结果良好。贝叶斯方法比基于阈值的方法有了显著的改进。值得注意的是，基于贝叶斯统计方法的模型表明，模拟数据集可以作为一个预测器，以提高云检测精度。Merchant等(2005)首先将该方法用于海洋区域的云屏蔽，以获取海表温度。Bulgin et al.(2014)和Bulgin et al.(2018)将该方法扩展到陆地区域。在这些研究中，使用人工分类数据集进行验证。由Heidinger等人(2012)使用配置的AVHRR和CALIOP数据导出的自动贝叶斯分类器，显示了对基于阈值的方法的改进，以及在云掩蔽过程中获得不确定性的能力。本文还介绍了利用CALIOP数据计算后验云概率的方法。  
最近，支持向量机(SVM)方法已被用于最新的claudia a3算法(Ishida et al.， 2018)。高质量的训练数据集对于基于机器学习的方法和手工生成的数据集是必不可少的，例如ACCA参考数据集(Irish et al.， 2006)和由Hollstein et al.(2016)构造的Sentinel-2 MSI数据集，目前基于机器学习的云检测方案经常使用。在Ishida等人(2018)中，支持向量机分类的训练数据集也通过仔细检查典型的表面类型和消除不规则数据，从实际卫星测量中主观地选择。

1.2。传统方法的局限性

传统的基于阈值的云面具方法在冰雪覆盖地区仍然面临严重挑战，特别是在北极和亚北极地区，那里有频繁的温度反演(影响基于轮胎的测试)和中纬度地区，反射信号往往来自混合雪和植被/土壤覆盖的像素。为了处理这种复杂的地表条件，基于阈值的逻辑变得越来越复杂(可以在1-5 of Irish et al.， 2006)，通常需要大量的卫星通道。有时这些测试会产生相互矛盾的结果，需要额外的“清晰恢复测试”来避免错误分类(Ackerman等人，2010)。需要探测可能被雪覆盖的地区也增加了结果的不确定性。Wang et al.(2008)报道，在传统的基于阈值的方法(如MODIS cloud mask)产生的结果中，将积雪地区误分为“云”或“云”误分为“云”仍然是一个严重的问题，如第三节所示。

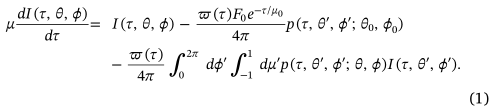
另一方面，机器学习方法一般不依赖于阈值，也不依赖于在云筛选前检测雪。然而，对人工生成数据集的依赖限制了基于机器学习算法的开发和操作使用。对数百万像素的数百张图像进行分类需要大量的人力资源，很难生成可靠的训练数据集。人工分类的图像数量有限，也很难覆盖所有可能的太阳/视图几何图形，这限制了训练过的算法的操作使用。最重要的是，手动分类的图像通常只能在发射后使用。这种情况阻碍了对算法性能的发射前评估，并使其在不同传感器上的应用变得困难。

2。在本文中，我们提出了一种新的基于机器学习的云和雪检测和识别方法，以克服以往方法的局限性。我们没有使用人工生成的数据集，而是模拟了机器学习算法所需的训练数据集。与根据实际测量手工生成的训练数据相比，模拟训练数据具有以下优点:  
•不需要人类识别数以百计、数百万像素的图像，大大节省了人力。

•训练样本的数量可以与期望/需要的一样大，这可以帮助避免过拟合问题，并用于充分挖掘机器学习技术的潜力。

•训练数据集可以覆盖所有可能的太阳/视图几何图形。

•该算法可以很容易地修改，以应用于不同的传感器;只需要新的训练数据集。  
为了创建这样的训练数据集，有必要考虑入射太阳辐射与不同类型的表面、气溶胶和云的相互作用。这一要求意味着使用全面的辐射传递模型是至关重要的。为了模拟复杂地表的反射率，我们构建了这样的模型;下一节将提供详细信息。2.1。辐射传输模拟  
为了模拟卫星仪器接收到的光信号，我们需要求解与光在耦合的大气表面系统中传播相关的辐射传输方程(RTE)。通过求解以下RTE，得到波长λ处的散射辐射I(τ，θ，ϕ):



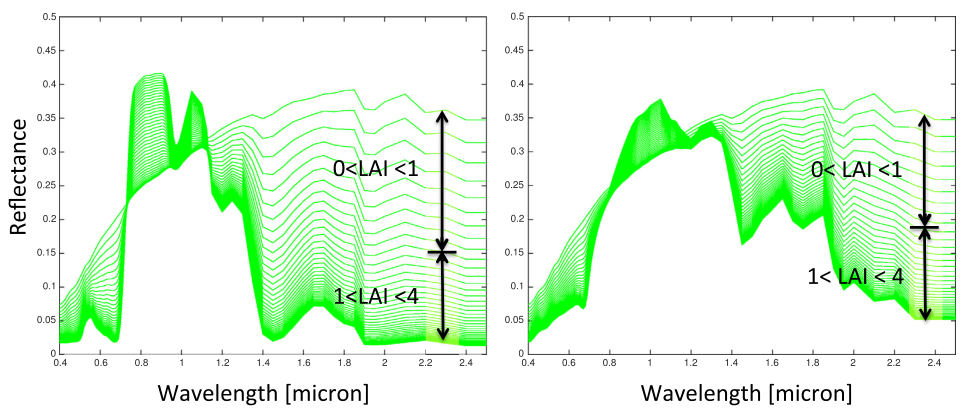
这里,F0is事件top-of-the-atmosphere (TOA)太阳能辐照度(正常梁),而微分光学深度dτ=−(α+β)dz,单一散射反照率ϖ=β/(α+β)=β/γ,和散射相函数p(τ,θ,ϕ;θ,ϕ)的固有光学特性(IOPs)散射/吸收介质。注意，我们分别用希腊字母α， β和γ=α +β来表示吸收系数，散射系数和消光系数。θ0和ϕ0代表太阳天顶角和方位角，μ0= cosθ0;θ '和ϕ '是散射事件前的传感器天顶角和方位角，θ和ϕ是散射事件后的对应角，μ= cosθ。在我们的训练数据集中，TOA双向反射系数(以下简称反射率)，定义为R(τ，θ，ϕ) =πI(τ，θ，ϕ)/F0 cosθ0，使用最新版本的畸变辐射传输模型(RTM) (DISORT 4.0, Lin et al.， 2015;stnes等，N. Chen等。R e m o t e S e n S i n g o f e n v i R o n m e n t 2 1 9(2 0 1 8) 6 2 - 7 1 6 1988 2017)采用副环带眼压方法由陈et al .(2017)在短波红外成像渠道提高准确性。

2.1.1。大气IOPs

我们使用美国标准大气成分资料(安德森et al ., 1986)分为14层提供输入基于MODTRAN乐队模型(例如,看到Stamnes et al。(2017)详情]产生吸收系数和光学深度由于大气痕量气体包括水,二氧化碳,O3, C H4,二氧化氮。为了解决IOPs的垂直变化，需要分层，经验表明，14层就足够了。分子(瑞利)散射光学深度计算从瑞利散射截面(Stamnes等人，2017)乘以空气的容重可从安德森等人(1986)。气溶胶的IOPs是由OPAC气溶胶模型的输出结果制成的(Hess, 1998)。假设液水云是由球形粒子的多分散组成的，IOPs是根据MieDebye理论(Mishchenko等人，2002年)利用Segelstein(1981年)的水的折射率计算的。对于冰云，IOPs是从Baum等人(2011)的大体积散射和吸收模型中的“一般习惯混合”模型中得出的。假定云的厚度为2公里。对于液态水云，云底高度假设为离地表2公里。对于冰云，假设云底高度为8千米，而不考虑表面高度。

2.1.2。表面IOPs

为了模拟不同地表类型的TOA反射率，我们使用土壤-叶片-冠层(SLC)模型(Verhoef and Bach, 2007)结合我们的扭曲RTM。采用SLC模型输出的双向反射分布函数(BRDF)作为畸变的下边界条件。图1显示了SLC模型模拟的下垫面土壤类型为1(代表耕作土壤类型)的不同绿色和棕色植被在最低点方向的大气底部(BOA)反射率。通过改变叶面积指数(LAI)、棕色植被分数(fb)或土壤类型等参数，可以模拟不同类型绿/棕色植被以及裸土的反射率。雪颗粒假设为冰球，其折射率为Warren和Brandt(2008)的冰。一旦确定粒径分布，可以通过米-德拜理论计算单色IOPs，或者通过有效雪粒粒径的参数化计算(Stamnes et al.， 2011)。



2.1.3。反射混合雪/植被土壤情况下和高海拔地区

为了更好地处理部分积雪覆盖的情况下,我们采用以下线性混合规则的反射与雪分数像素,f:

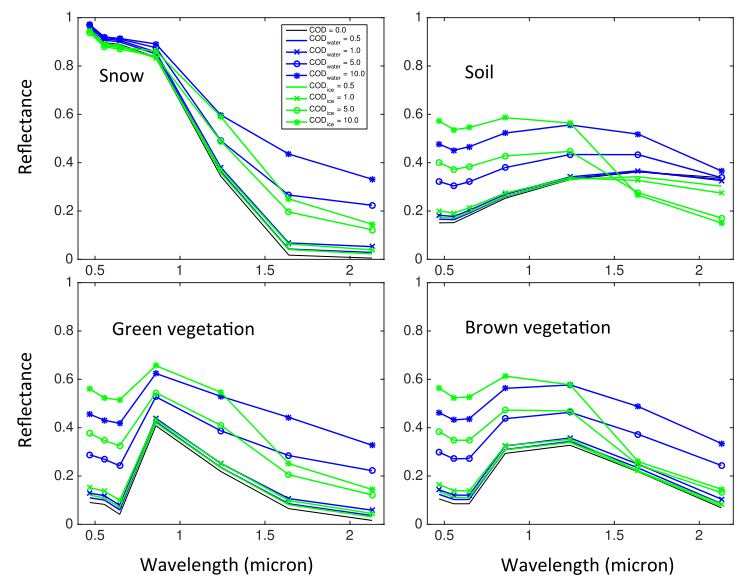


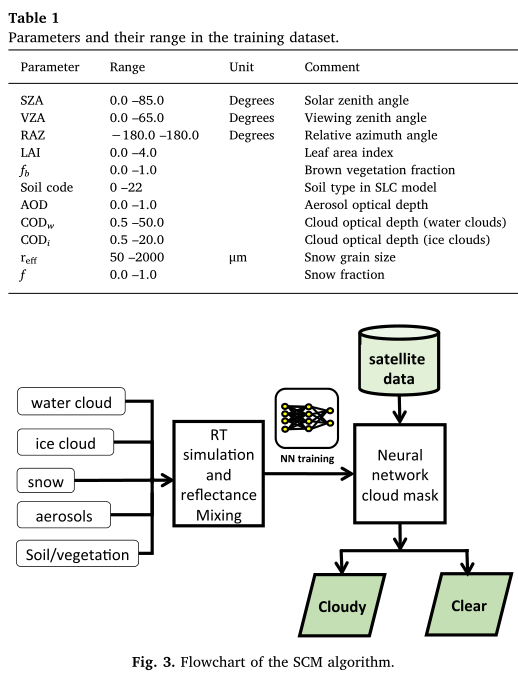
通过随机改变积雪系数f和雪/地参数，可以模拟各种雪混合植被/土壤情况下的TOA反射率。为了处理TOA反射率随地表高程的变化，我们假设地表高程在随机生成的0 ~ 1000 m高度处模拟TOA反射率。对于海拔非常高的雪区，如格陵兰岛和南极洲，我们在模拟中将地表海拔范围扩大到0至4000米。

2.1.4。用于晴空和多云情况下的反射率

图2显示了我们在不同的云光学深度(COD)下不同地表类型的辐射传输模拟中获得的TOA反射率的例子。可以看出，多云的情况有不同的反射率，这取决于下垫面的类型。由于高度保守的性质云的散射力和近红外波长区域(单散射反照率接近1云粒子,例如看到杨et al ., 2013),大量的太阳辐射将达到适度的表面甚至厚云(10)的光学深度和表面的反射信号将大大加剧TOA反射(看到小的例如第13章,2006)。因此，包含表面反射对于多云天空模拟是非常重要的。在最新的MODIS集合6云产品(Amarasinghe et al.， 2017)中，一个基于Cox-Munk的表面BRDF模型被用于解释来自海洋表面的重要贡献。在土地表面被认为作为传感器的反射,所以总反射率可以表述为一笔黑色表面的值加上一个代数修正项传感器的表面反照率成正比(见Eq。(5)在编写et al ., 2017),这可能导致重大错误当底层表面非常的双向反射各向异性(如雪)。在我们的辐射传输模拟数据集中，对植被、土壤和雪进行了严格的地表BRDF处理，避免了从陆地表面假设朗伯反射的潜在问题。TOA反射率对表面反射率的依赖性也意味着我们需要覆盖图1中所示的尽可能多的表面类型。底部大气(BOA)光谱反射率在最低点方向作为LAI的函数，绿色植被(左)和棕色植被(右)。在这些模拟中，入射太阳天顶角被设置为30°。(对于本图例中颜色引用的解释，读者可参考本文的web版本。)

图2所示。模拟了不同晴空(COD = 0.0)和多云情况(COD > 0.0)下MODIS TOA光谱在最低点方向的反射率。地表类型有雪(左上)、土壤(右上)、绿色植被(左下)和棕色植被(右下)。在555 nm波长下，云光学深度分别为0.5、1.0、5.0和10.0。在这些模拟中，入射太阳天顶角被设置为30°。(对于本图例中颜色引用的解释，读者可参考本文的web版本。)





建立一个可以代表大多数情况的综合数据集是可能的，这种多样性可能是对我们的机器学习方案的挑战。我们也可以观察到，对于非常薄的云(云的光学深度< 0.5)，反射率的变化与晴空情况相比通常是非常小的。这种情况表明，使用反射率通道进行云检测可能存在局限性，而且可能需要热红外通道来区分这些薄云与下表面。

2.2。神经网络训练  
我们基于机器学习算法的训练数据集由大量晴空和多云情况组成，设计用于覆盖尽可能多的表面类型和太阳/观看几何图形，以充分表示在自然界中遇到的可能组合。大气和地表参数如气溶胶/云光学深度、叶面积指数(LAI),分数布朗植被(fb),和雪颗粒被认为是自由参数允许在现实的范围不同(见表1)。晴空条件下大量不同病例随机选择来代表各种各样的土壤、植被和气溶胶组合的训练数据集。同样，对于任何给定的表面和晴朗的天空条件，大量的水/冰云的光学深度被随机选择来模拟相应的反射率。通过这种方式获得的6个卫星通道(MODIS/Aqua为0.47、0.55、0.66、0.86、1.24、2.13 μ m)的模拟反射率，以及太阳天顶、观测天顶、相对方位角、地表高程作为算法的输入参数。

通过这种方式，产生了超过2000万个样本，并用于训练二进制(cloudy/clear)神经网络分类器，使用一个简单的多层感知器方案，其中一个隐含层包含10个神经元。σ= 1/(1 + e−z)作为各层的激活函数。我们对我们的数据集进行随机排列，然后将其分为两部分:75%的总病例数用于训练，其余25%用于验证。经过足够的迭代次数(通常是200次左右)，训练和验证数据集的精度通常在98.5%到99.2%之间，这意味着我们在避免过拟合的同时获得了足够的精度。经过训练的神经网络可以在不到2秒的时间内处理一幅MODIS图像(通常包含2030×1354 = 2.7 M像素)。

表1列出了模拟的主要参数及其变化范围，图3给出了新型冰雪云罩(SCM)算法的流程图。

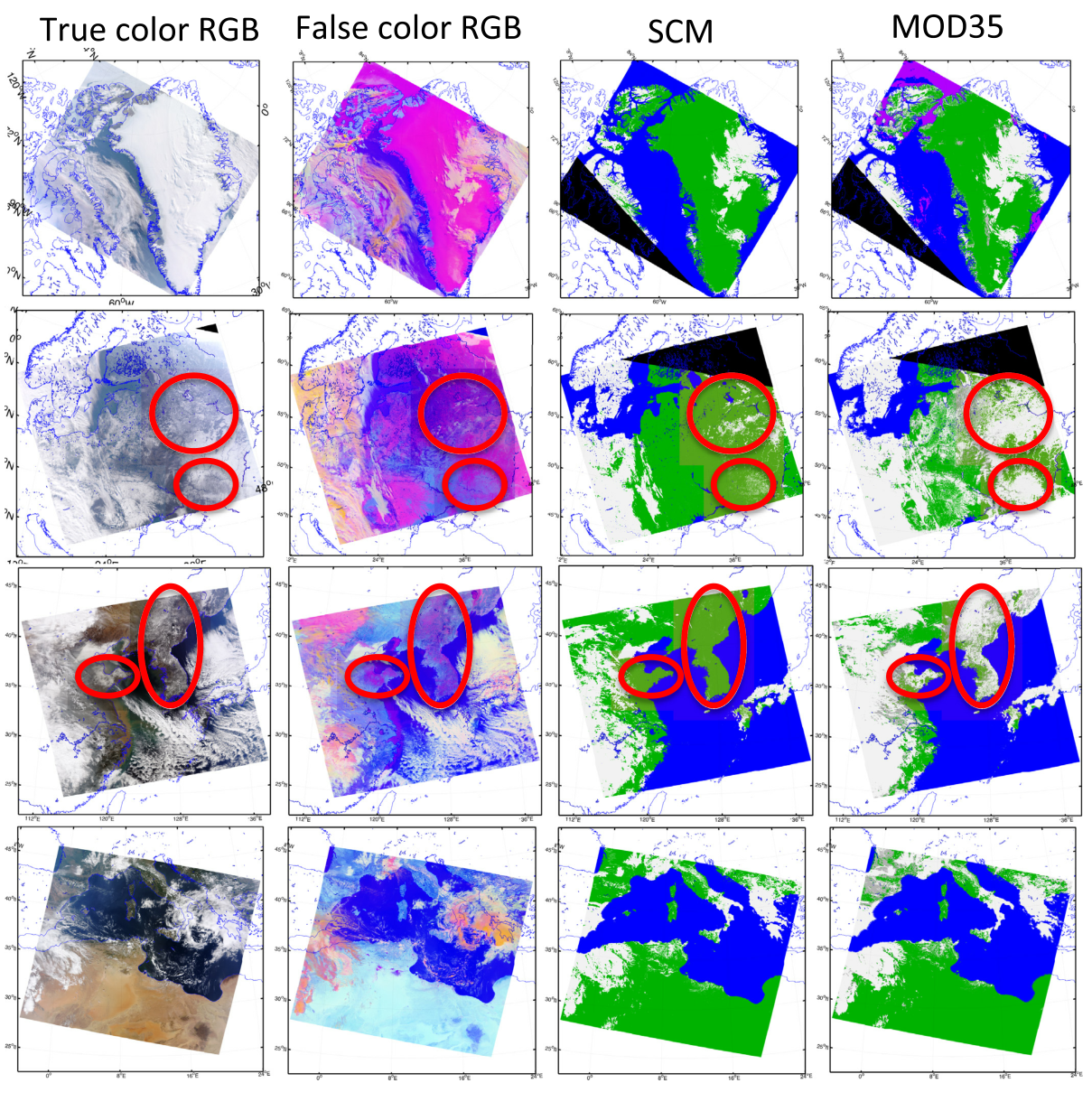
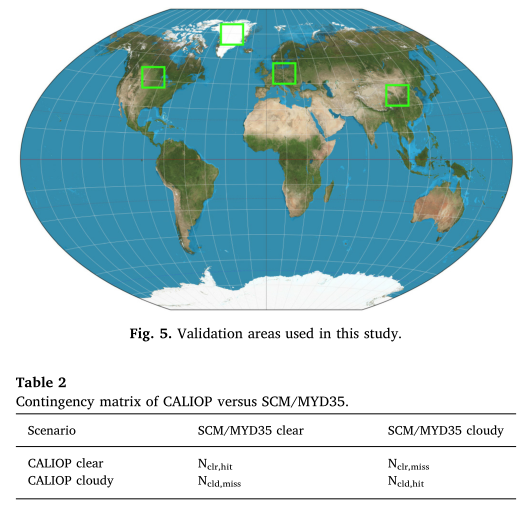


图4所示。MODIS图像的云掩模结果。从上到下:2015年7月9日格陵兰岛的Aqua MODIS图像;2008年1月18日，欧洲上空的Aqua MODIS图像;2003年1月24日东亚上空的Aqua MODIS图像;2009年9月26日摄于北非的Terra MODIS图像。利用0.65 μm和2.13 μm r反射率和10.8 μm亮度温度组成伪彩色RGB图像。多云的像素被标记为白色或灰色，晴空的陆地像素被标记为绿色，水域被标记为蓝色。水域上空的云没有标记。(对于本图例中颜色引用的解释，读者可参考本文的web版本。)

3.结果和验证

在本节中，我们将把训练好的神经网络分类器应用于MODIS图像，并验证其在不同陆地区域的性能。我们首先测试SCM，使用上述6通道配置，以模拟其在传感器上的性能，如GCOM-C(日本)上的第二代全球成像仪(SGLI)和Suomi国家极地轨道伙伴关系(Suomi NPP)气象卫星上的可见红外成像辐射计套件(VIIRS)。与MODIS相比，这些传感器缺少很多TIR波段，因此主要使用反射率波段实现的性能就显得尤为重要。然而，由于SGLI有两个热红外波段(11和12μ m)，我们也探索了采用与Wilson和Oreopoulos(2013)使用的动态阈值分裂窗口测试类似的优势，以提高对薄卷云的敏感性。



3.1。中纬度地区的结果:与MODIS图像的比较

图4给出了SCM生成的云检测结果的一些例子，并与MODIS云面具(MOD35 C6)采集集6生成的类似结果进行了比较。MOD35探测到的云显示为白色(肯定是多云)和灰色(可能是多云)。SCM算法的云检测是二进制的，这意味着只提供多云(显示为白色)和天空晴朗(显示为绿色)的陆地识别。图4中的伪彩色RGB图采用与Hutchison等人(2013)相同的配色方案(使用MODIS channel 1-6-32作为R-G-B)。在这个方案中，云通常看起来是白色的(温暖的低云)和黄色的(寒冷的冰云)，而被雪覆盖的区域通常看起来是粉红色的，因为它们的SWIR反射率相对较低。的比较,可以看到,在通用单片机和MOD35 C6 non-snow-covered土地区域有类似的云检测功能,包括植物的土地在欧洲和北非的撒哈拉沙漠(底部面板图。4)。这些比较表明,基于神经网络的SCM算法可以捕获各种土地表面的光谱特征包括各种类型的植被以及明亮的沙漠地区和区分云。对于许多以前使用的云遮罩算法来说，这种识别通常是困难的。对于格陵兰高原等纯雪覆盖地区，SCM和MOD35的性能也相似(图4顶部面板)，这意味着SCM采用的神经网络在高海拔雪覆盖地区表现良好。这两种算法在春季/秋季的性能下降可能是由于非常高的太阳天顶角(通常大于80°)和样本数量减少。然而，这种情况在中纬度地区的冰雪覆盖地区发生了变化。在图4中间的面板中，我们可以看到大量晴空雪像素(伪彩色RGB中的粉红色)被MOD35误分类为多云像素，而SCM提供了正确的识别。复杂的雪-草-土混合条件，给MOD35算法的阈值测试带来了相当大的困难，可以看出，错分类多发生在雪区边缘。

3.2。使用CALIOP进行统计验证

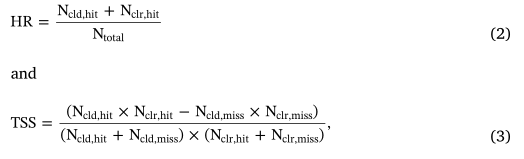
以往基于机器学习算法的验证仅限于基于图像的统计，基于使用从不同地点和季节选择的人类识别图像作为基准。由于人类需要分类的像素量很大，综合评价云面具算法的有效性，很难实现所需的时空覆盖。垂直偏振云气溶胶激光雷达(CALIOP)是搭载在CALIPSO卫星上的一种激光雷达，可提供气溶胶和云的高分辨率垂直剖面。CALIOP的活动云检测方案结合成对的Aqua MODIS数据提供了目前可用的最可靠的云掩膜结果评估。我们使用了配置CALIOP和Aqua MODIS测量数据，使用CALIOP检测到的多云/晴朗条件进行比较，与我们在被雪覆盖的格陵兰高原上采用的方式类似。因此，以CALIOP 1 km云检测结果作为基准。研究人员使用了2008年东亚、欧洲、北美和格陵兰岛的全年数据集。这些MODIS图像覆盖的大致面积用图5中的绿框表示，每个站点大约使用2000张MODIS图像。与我们在Chen等人(2014)中所做的类似，我们计算并比较了SCM和MOD35 (Aqua MODIS数据的MYD35)的命中率(HR)和hansen - kuiper (True)技能得分(TSS)。人力资源和技术支持的定义为

表2定义了nld、hit、Nclr、hit、Nclr、miss和nld、miss。图6和表3为SCM和MOD35的CALIOP验证结果。我们将图6中仅标注为“SCM (NN + only)”的神经网络得到的结果纳入其中。总的来说，SCM中的神经网络测试在无雪覆盖地区执行一致，达到约80%的HR和65%的TSS。加入热红外测试可以使HR和TSS提高5% ~ 7%，但性能略低于MOD35。这些结果表明，在非积雪覆盖地区的SCM有一定的改进空间，使用热红外波段范围内的额外测试可以显著改善薄云的识别，这是难以识别的反射基础方法。在冰雪覆盖的格陵兰高原地区，两种算法的表现非常接近，SCM (NN + BT)在夏季略有优势。加入热红外测试的优势相对较小，可能是由于使用了保守的阈值。结果表明，仅使用两个通道(10.8和12μ m)是不够的，可能需要更多的热红外通道(如3.7 μ m)来进一步改善结果。冬季月份HR和TSS的下降可能与样本量的减少和太阳天顶角的增大有关。最大的差异再次出现在冬季的中纬度地区，那里的雪经常与植被和土壤混合。MOD35的TSS得分由于晴空情况下较高的错误分类率而显著下降，这与我们基于图像的测试结果一致，表明MOD35处理复杂的雪混合植被/土壤场景有困难。需要注意的是，与SCM相比，MOD35有时具有较高的HR，但较低的TSS(例如在北美的1月份)。这种行为是由于人力资源的偏见,因为通常比晴空多云的情况下在我们的统计样本,和更全面的TSS捕获的增加,小姐,导致错误分类的阴天情况下的晴空MYD35 SCM在冬季的几个月相比,见表3。

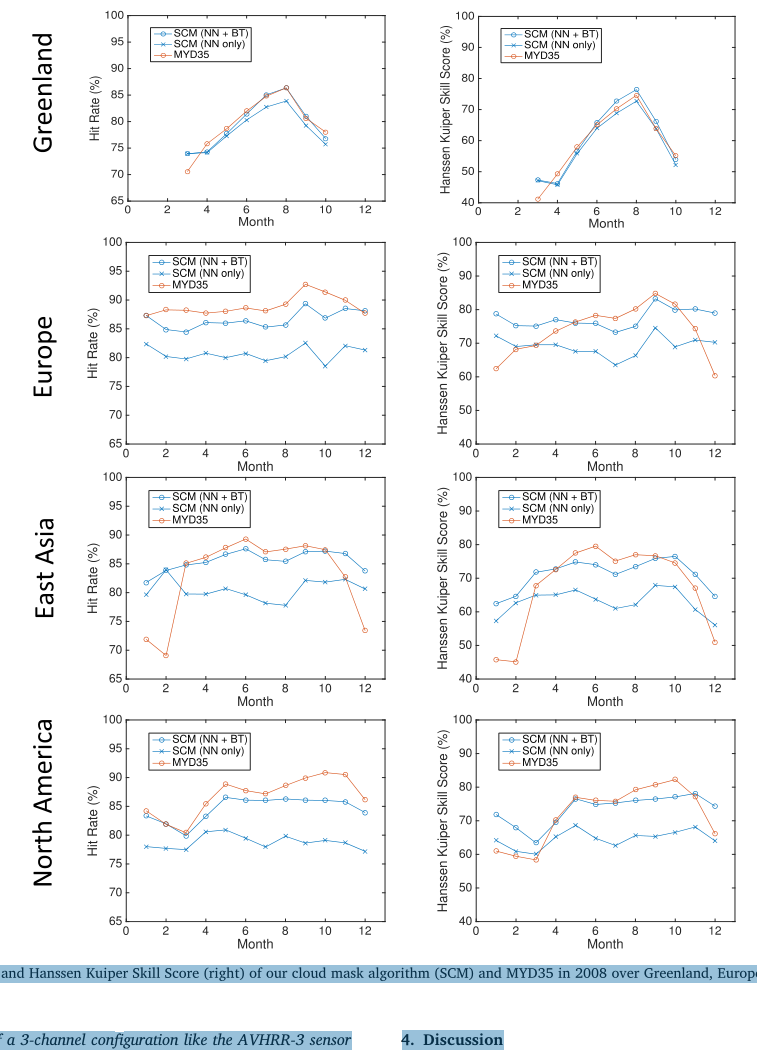


图6所示。2008年我们的云面具算法(SCM)和MYD35在格陵兰岛、欧洲、东亚和北美的命中率(左)和hansen Kuiper技能得分(右)。

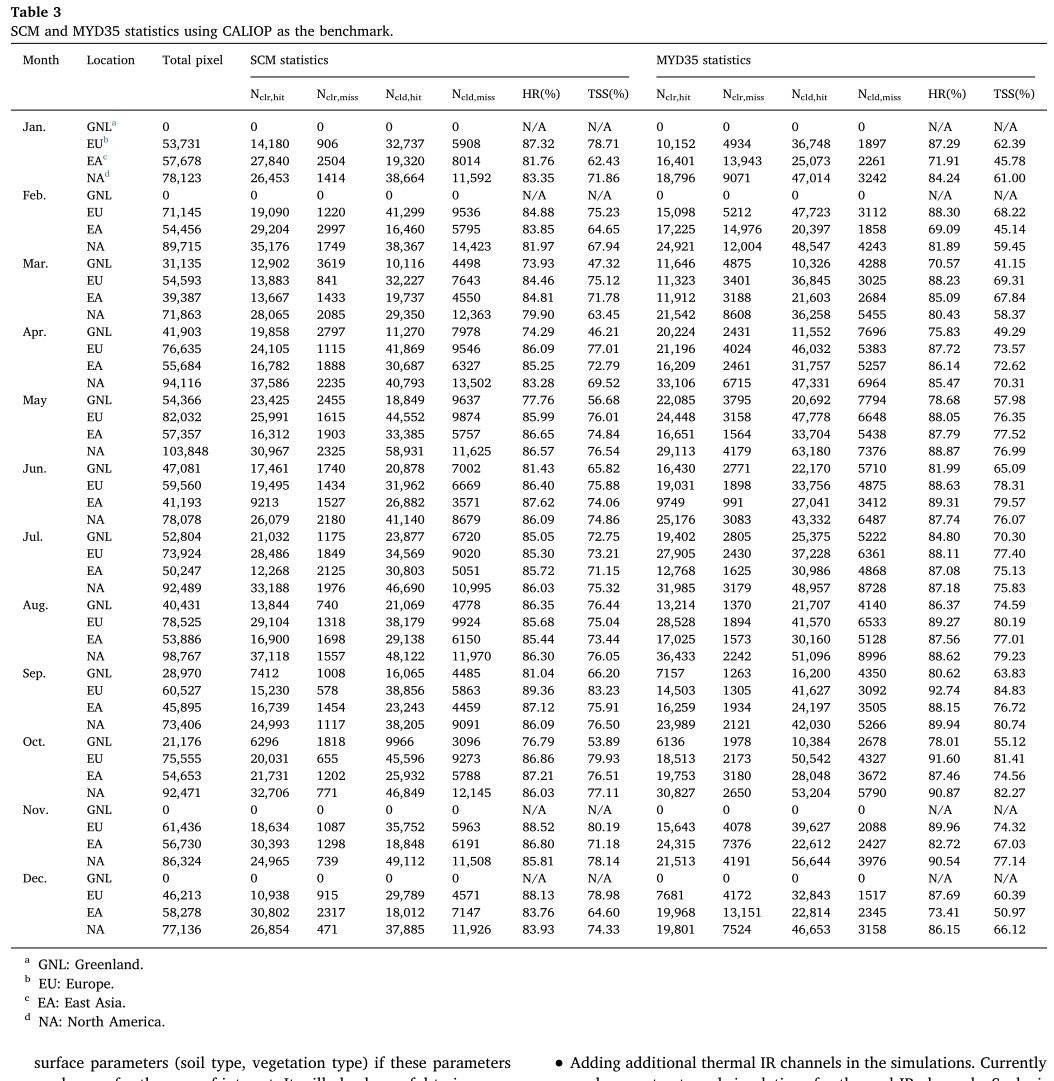
3.3。3通道配置的性能，如AVHRR-3传感器

最后，我们测试了一个特殊的配置，以研究单片机应用于其他传感器的灵活性。因此，我们没有使用上述6个MODIS反射率通道，而是尝试了3通道配置(0.47、0.66、2.13 μ m)来训练我们的算法，以模拟其在AVHRR-3等传统传感器上的应用。从图7可以看出，与6通道配置相比，这种3通道配置可以提供一致的性能，尽管性能稍差。这3个通道在中纬度地区冬季的结果仍然优于MOD35。使用模拟数据进行训练，使我们能够在卫星发射前评估基于机器学习技术的算法的性能，并探索用于遮蔽云的最有效的卫星通道组合。

4. 讨论

上面描述的SCM云筛选工具是我们第一次尝试使用一种基于综合辐射传输模拟和机器学习相结合的云筛选方案。根据我们迄今在此方法方面获得的经验，我们认为该方法有许多方面可以改进。这些包括:

* 根据当地大气和地面条件构建局部晴空模拟训练数据集。使用固定的大气成分剖面以及随机生成的表面属性(土壤类型、绿棕植被比)模拟当前训练数据集。因此，在本文中，我们描述了该方法的一般应用，使用一组通用的大气和地表参数来证明其在全球尺度上的有用性。在进一步的应用中，如果这些参数对于感兴趣的区域是已知的，则完全有可能构建一个使用本地表面参数(土壤类型、植被类型)的训练数据集。引入相对湿度和气压等参数，也有助于改善晴空模拟的动态范围。对于一年中给定时间的给定位置的应用，可以模拟一个高精度晴空辐射范围，以提高辨别能力
* 构建一个更真实的云天模拟训练数据集。在当前的模拟数据集中，为了简单起见，我们使用了一个固定的云高度(液态水云高于地表2.0公里，冰云高于地表8.0公里)。在未来的实现中，云的级别和属性可以变得更加灵活，以改善对多云/晴朗天空的判断。模拟数据集与人类识别数据集的结合使用也将是一个有趣的研究方向。
* 在模拟中增加额外的热红外通道。目前我们还没有对热红外通道进行模拟。这样的模拟将涉及各种不同的云的性质以及不同的表面发射率。红外热通道的使用不仅有助于提高对光学薄云的检测，而且将我们的方法扩展到夜间工作。
* 使用额外的机器学习技术来提高性能。在本文中，我们使用一个简单的感知器神经网络模型来进行云/晴空的判断，因为它易于训练和实现。其他方法，如袋装决策树、支持向量机和/或贝叶斯方法也可以使用。事实上，初步的测试表明，bagged tree模型可以达到比目前的神经网络方法更高的精度。目前正在利用卫星数据对这些模型进行测试/验证。



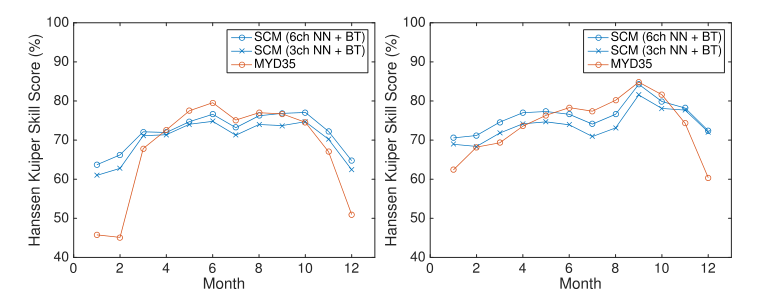


图7所示。我们的云面具算法(SCM)和MYD35在2008年在东亚(左)和欧洲(右)的Hanssen Kuiper技能得分。注意单片机在冬季的优越性能。

如上所述，我们的方法和算法当然可以改进，重要的是遥感社区的用户可以帮助进一步探索这一方法。我们计划在ESA的SNAP平台上创建当前算法的一个版本(作为“插件”)，从而使其可供遥感社区使用。我们还将把我们的培训数据集放在我们的网站(http://lllab.phy.stevens.edu)上，供那些对进一步探索这一方法感兴趣的人使用。

5. 总结

一种基于机器学习技术的云罩和雪检测算法(SCM)已经开发、描述和验证了Aqua MODIS/CALIOP在陆地地区的测量数据。SCM使用大量辐射传输模拟生成的模拟数据集来训练机器学习算法，而不是使用从实际测量得出的人类分类数据集。与MODIS cloud mask等传统方法或以往基于机器学习的算法相比，本算法具有以下特点:

* 它简化了测试逻辑，使用更少的卫星通道，同时能够在不同季节在不同类型的下垫表面上提供一致的性能，
* 晴空情况的误分类率较低，地表覆盖有雪和植被/土壤的中纬度地区冬季TSS得分明显较高
* 它在纯植被、土壤和积雪覆盖地区的表现类似于MODIS云罩
* 由于它完全依赖模拟数据进行算法训练，因此可以在卫星发射前很容易地将其修改为适用于新的传感器配置，以评估其性能。这一功能有助于在发射前探索使用哪些卫星通道来遮蔽云层和检索所需产品。

最后，我们应该指出JAXA的GCOM-C任务的目的是对碳循环和辐射收支进行全球长期观测(Imaoka et al.， 2010)。搭载GCOM-C的“Shikisai”卫星于2017年12月成功发射，并已开始数据传输。本文所述的云罩算法已在数据处理链中实现，将用于反演雪粒大小、杂质浓度、积雪覆盖程度等关键参数组成的冰冻圈产品。

感谢，鸣谢