REMOTE SENSING LETTERS
https://doi.org/10.1080/2150704X.2018.1553318

Taylor & Francis
Taylor & Francis Group

Cloud detection using infrared atmospheric sounding interferometer observations by logistic regression

Tengling Luo[a], Weimin Zhang[a,b], Yi Yu[a], Miao Feng[a], Boheng Duan[a] and De Xing[a]

[a]College of Meteorology and Oceanology, National University of Defense Technology, Changsha, Hunan province, China; [b]Laboratory of Software Engineering for Complex Systems, Changsha, Hunan province, China

ABSTRACT

Hyper-spectral infrared radiance data play an important role in cloud detection. To improve the cloud detection accuracy, this study proposes a novel cloud detection method based on the logistic regression model that uses the Infrared Atmospheric Sounding Interferometer (IASI) radiance data of four characteristic channels as the training features. Due to significant differences in the terrain between the land and the sea, the data from the oceans and continents are trained separately. Thereafter, the proposed scheme is verified and compared with existing methods. The results show that the accuracy of the proposed method (97% at sea and 88% on land) outperforms that of the existing Advanced Very High Resolution Radiometer (AVHRR)/IASI scheme (75% at sea and 55% on land). In addition, the proposed method uses only IASI observations as input and thus does not require the use of other auxiliary data.

ARTICLE HISTORY
Received 24 May 2018
Accepted 13 November 2018

1. Introduction

Hyper-spectral infrared sounders, such as the Atmospheric Infrared Sounder (AIRS) and the Infrared Atmospheric Sounding Interferometer (IASI), provide the most valuable observations for current numerical weather prediction (NWP) systems (Mcnally et al. 2006; Collard and Mcnally 2009; Hilton et al. 2009). Because a cloud is a blackbody that can absorb infrared radiation, assimilating infrared radiances is particularly difficult for infrared sounders because such radiances are easily contaminated by clouds. Substantial progress has been made in addressing these cloud-affected radiation problems (Pavelin, English, and Eyre 2008; Mcnally 2009; Bauer et al. 2011). Nevertheless, the use of the hyper-spectral IASI radiances still relies primarily on cloud-free data (Eresmaa 2015).

As clear-sky IASI radiance data are important for assimilation, the rapid and accurate detection of cloud-free IASI fields of view (FOVs) is a critical task. Four main cloud detection methodologies are employed for infrared sounders. (1) The first approach is the multi-threshold method, which directly detects clouds by employing the infrared brightness temperature and infrared brightness temperature difference threshold (Suseno, Yuga, and Yamada 2012; Jin et al. 2013). However, many cloud detection schemes contain multiple thresholds; thus, it is difficult to set suitable thresholds. (2) The second methodology is clear

CONTACT Tengling Luo luotl17@163.com College of Meteorology and Oceanology, National University of Defense Technology, Changsha, Hunan province, China

channel detection, which was proposed by McNally and Watts (Mcnally and Watts 2003); this method can select 'clear' channels from a cloudy sky based on assumptions about the radiative effects of clouds on the observations minus background departures. (3) The third group of methods is composed of artificial intelligence (AI) algorithms, which can automatically detect clouds after using a large number of samples for training. Examples of AI algorithms include the support vector machine (Li et al. 2015), scene learning technique (An and Shi 2017), deep extreme learning machine (Shi et al. 2016), fuzzy logic approach (Baum et al. 1996) and convolutional neural network (Johnston et al. 2017). However, most of these AI schemes use only cloud pictures, which cannot extract a sufficient quantity of effective cloud information from the instrument's channels, as input training features. (4) The fourth type of technique is the imager-based method. A few studies have employed high-spatial-resolution images to detect clouds (Lu 2007; Hagolle et al. 2010; Han et al. 2014), while other studies have improved the cloud detection accuracy of hyper-spectral infrared sounders by using collocated imager information. For instance, Li et al. used collocated high-spatial-resolution Moderate Resolution Imaging Spectroradiometer (MODIS) data to assist in AIRS sub-pixel cloud detection (Li et al. 2004). Another researcher employed the Advanced Very High Resolution Radiometer (AVHRR) for IASI sub-pixels to finish the cloud detection process (Eresmaa 2015). However, although imager-based algorithms have achieved remarkable successes in terms of cloud detection for hyper-spectral infrared observations, these methods are characterized by relatively high costs because they require the combination of two observational data streams, e.g., AVHRR/IASI and MODIS/Atmospheric Infrared Sounder (AIRS).

In this study, cloud detection is regarded as a binary classification problem, that is, the output can take only two values, '0' or '1', which represent the outcomes as cloudy or clear, respectively. Logistic regression is one of the most robust algorithms used to solve binary classification problems (Feng et al. 2014). The main idea of logistic regression is to evaluate the probability of a discrete outcome occurring based on a set of past inputs and labels. There are two advantages of The use of logistic regression to solve binary classification problems boasts two advantages: first, logistic regression exhibits a fast run-time (Komarek and Moore 2003); second, logistic regression is robust and can solve a binary classification problem in which a fraction of the training samples are corrupted (Feng et al. 2014).

In the past, the techniques used to detect clouds were based mainly on the multi-threshold scheme or the imager-based scheme (e.g., AVHRR/IASI, MODIS/AIRS) (Eresmaa 2015; Li et al. 2004). IASI level-2 cloud parameter products are also available for such research. Nevertheless, although these algorithms and products have produced remarkable achievements in terms of cloud detection with IASI data, the costs needed to apply these methods are relatively high, and these approaches suffer from a low accuracy. As an alternative, Rüthrich et al. (Rüthrich et al. 2013) proved that the clouds over the Tibetan Plateau can be detected effectively by logistic regression. Therefore, to make cloud detection more convenient and accurate, we propose a new approach based on the combination of logistic regression with IASI cloud detection. Four IASI FOV characteristic channels are used as the input features in the logistic regression model. Then, the proposed algorithm is compared with the traditional AVHRR/IASI cloud detection method and IASI level-2 products.

2. Materials and methods

2.1. IASI, AVHRR and the IASI cloud parameter product

The IASI is a hyper-spectral infrared sounder with 8461 spectral samples (Blumstein et al. 2004) that constitutes a key element of the payload on the series of Meteorological Operational satellite programme (METOP) polar-orbiting satellites (Blumstein and Phulpin 2007), which complete two ground observations per day, passing over East Asia at nearly 00:00 (UTC) and 12:00 (UTC) every day. The IASI acquires measurements in the infrared component of the electromagnetic spectrum at a horizontal resolution of 12 km over a swath width of approximately 2,200 km; accordingly, its measurements allow researchers to derive temperature and humidity profiles with a vertical resolution of 1 km and average accuracies of 1 K and 10%, respectively (Blumstein and Phulpin 2007). The infrared imager co-registers the IASI soundings with the AVHRR imager, another instrument on the MetOp-A satellite. The AVHRR is a six-channel scanning radiometer that provides three solar channels in the visible to near-infrared region and three thermal infrared channels with a spatial resolution of 1 km at nadir (Wang and Cao 2008) over a swath width of 2,928 km. To better understand the spatial differences between the IASI and AVHRR, the outgoing spectral radiances observed on IASI channel 921 (with a central wavenumber 875.0 cm^{-1}) and AVHRR channel 5 (833.3 cm^{-1}) are shown in Figure 1.

In this paper, we use the variable *Fractional Cloud Cover* from the IASI level-2 cloud parameter products as a training label. The IASI level-2 cloud parameter products are organized according to the orbit and are consequently split into different files every time when the IASI passes through the descending intersection point. The cloud parameter products contain the detection of clouds and dust aerosols in addition to fractional cloud cover, cloud top pressure and temperature and cloud phase data, which are retrieved from IASI sounder measurements. The retrievals are provided at the single IASI footprint resolution (which is approximately 12 km with a spatial sampling interval of approximately 25 km at nadir). These products are available via the European Organization for the Exploitation of Meteorological Satellites (EUMETSAT) data centre as IASI combined sounding products.

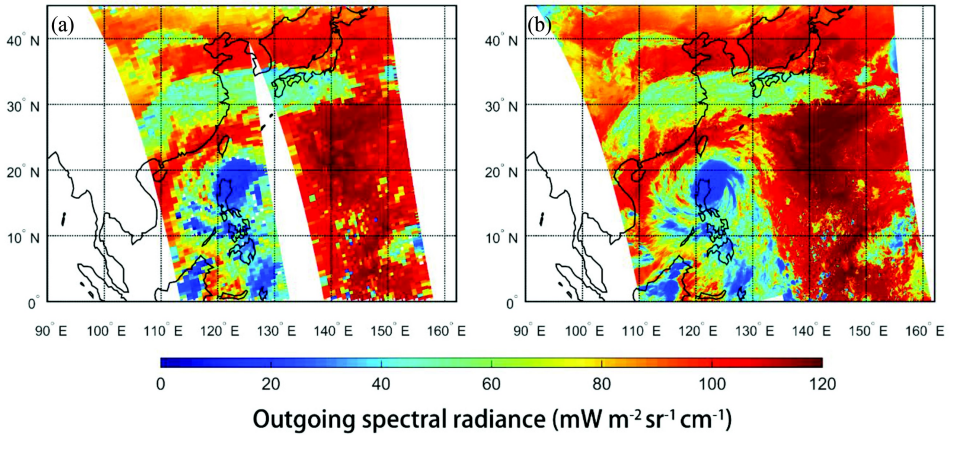

Figure 1. Outgoing spectral radiance (mW m^{-2} sr^{-1} cm^{-1}) measured by the Metop-A IASI and AVHRR imagers at 12:00 UTC on 19 October 2016. (a) Channel 921 (875.0 cm^{-1}) of the IASI; (b) channel 5 (833.3 cm^{-1}) of the AVHRR.

2.2. AVHRR/IASI cloud detection scheme

The AVHRR/IASI method, which was proposed by Eresmaa, is one of the cloud detection methods used by EUMETSAT (Eresmaa 2015). This cloud detection scheme collocates the IASI FOVs with the AVHRR FOVs and uses high-spatial-resolution AVHRR observations to detect the cloud information within the IASI FOVs.

Once an AVHRR pixel is collocated with an IASI footprint, the AVHRR radiation can be used to characterize the cloud characteristics in the IASI sub-pixel. In this study, the brightness temperatures of two AVHRR longwave channels (channel 4 and channel 5) are used as inputs. Then, three separate checks, namely, the homogeneity check, the inter-cluster consistency check and the background departure check, are performed (Eresmaa 2015). The IASI FOV is determined to be clear if it passes all three checks; otherwise, it is cloudy.

2.3. Logistic regression algorithm and datasets

2.3.1. Logistic regression algorithm

In statistics, logistic regression represents a regression model developed by David Cox (Cox 1959) where the dependent variables are categorical variables. The dependent variable of this article is dichotomous (i.e., binary), namely, the output has only two values, '0' and '1', which indicate whether the FOV of IASI is cloudy or clear, respectively.

In this study, there is a training set D composed of n observations, $\mathrm{D} = \{(x_i, y_i) |\ i = 1, 2, \cdots, n\}$, where $\boldsymbol{x}$ denotes a d-dimensional input feature, and y denotes the scalar label of $\boldsymbol{x}$. Here, $y = 0$ represents a cloudy IASI FOV, and $y = 1$ represents a clear IASI FOV. When given a feature $\boldsymbol{x}$, the probability that $\boldsymbol{x}$ is clear or cloudy is defined as p:

$$p(y = 1|\boldsymbol{x}) = \frac{1}{1 + \mathrm{e}^{-wx}} \tag{1}$$

$$p(y = 0|\boldsymbol{x}) = 1 - \frac{1}{1 + \mathrm{e}^{-wx}} \tag{2}$$

where $\boldsymbol{w}$ is the coefficient of feature $\boldsymbol{x}$. Based on the maximum likelihood, the control equation is:

$$L(\boldsymbol{w}) = \prod_{i=1}^{n} p(y_i|\boldsymbol{x}_i; \boldsymbol{w}) = \prod_{i=1}^{n} \left(\frac{1}{1 + e^{-wx_i}}\right)^{y_i} \left(1 - \frac{1}{1 + e^{-wx_i}}\right)^{1-y_i} \tag{3}$$

Additionally, we calculate the maximum $L(\boldsymbol{w})$ through iteration; after training, we can obtain the coefficient:

$$\boldsymbol{w} = \arg\max_{\boldsymbol{w}} L(\boldsymbol{w}) \tag{4}$$

Afterwards, we obtain the cloud detection predictor variable y :

$$y = \begin{cases} 1 \quad (\text{clear}) \quad if\ \boldsymbol{wx} \geq 0 \\ 0 \quad (\text{cloud}) \quad if\ \boldsymbol{wx} < 0 \end{cases} \tag{5}$$

Considering the significant terrain differences between land and sea, the coefficient $\boldsymbol{w}$ must be trained separately to obtain $\boldsymbol{w}_{\text{sea}}$ and $\boldsymbol{w}_{\text{land}}$.

2.3.2. Training features, training datasets and validation datasets

To ensure that the training features represent various levels of the atmosphere, we selected four characteristic channels, namely, channel 921, channel 386, channel 306, and channel 224, from among the 8461 channels of each IASI FOV as the training features. It should be noted that other operational channels can be used as training features; in this study, however, which is only a preliminary cloud detection experiment based on logistic regression, only four IASI channels are chosen. In future work, additional channels will be added.

The detailed information of these four characteristic channels is listed as follows. Channel 921 (with a central wavenumber of 875.00 cm^{-1}) is a window channel in which a portion of the electromagnetic spectrum can be transmitted through the atmosphere without any distortion or absorption. Channel 386 (741.25 cm^{-1}), channel 306 (721.25 cm^{-1}) and channel 224 (700.75 cm^{-1}) represent the temperature information of the lower, middle and upper atmosphere, respectively. Figure 2 (picture available online at https://gmao.gsfc.nasa.gov/forecasts/radmon_diagnostics.php) shows the weighting functions of the four characteristic channels that represent the sensitivity of each channel to the altitude. As shown in Figure 2(a), the weighting function of channel 921 reaches its maximum when it is at 1085 hPa, meaning that the observed radiance of channel 921 represents the surface information at 1085 hPa. Figure 2(b) through 2d show that when the weighting function reaches its maximum value, the corresponding height gradually increases. This demonstrates that channels 386, 306 and 224 represent the low-, middle-, and high-level information, respectively, of the atmosphere.

Additionally, we chose the IASI level-2 cloud parameter product variable *Fractional Cloud Cover* as the label for logistic regression. The *Fractional Cloud Cover* variable represents a percentage ranging from 0% to 100% that indicates whether the IASI FOV is either cloudy or clear (a *Fractional Cloud Cover* value of 0 indicates clear conditions, whereas a value of *Fractional Cloud Cover* value exceeding 0.8 indicates cloudy conditions).

In this paper, the training area was bounded by 95°E~155°E, 5°N~50°N; within this region, we selected 15,418 training samples on land and 26,656 training samples from the ocean as the training datasets.

For the validation datasets, we selected 5002 points on land and 10,012 points on the ocean from the typhoon Haima as the validation datasets. Moreover, all validation datasets were independent from the training datasets.

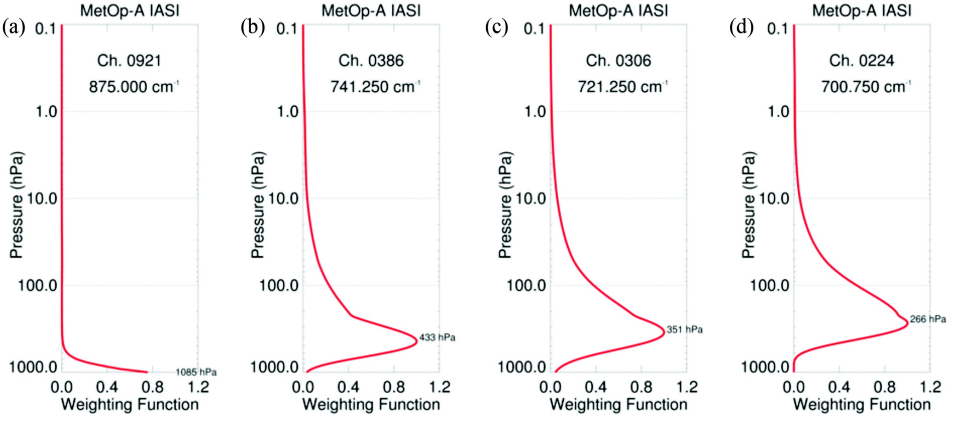

Figure 2. Weight functions of the four IASI characteristic channels. (a) Ch. 0921; (b) Ch.0 386; (c) Ch. 0306; and (d) Ch. 0224.

To evaluate the performance of the two methods, we calculated the probability of detection (POD), false alarm rate (FAR), and accuracy (ACC) (Han 2015)as follows:

$$\mathrm{POD} = H/(H+M) \tag{6}$$

$$\mathrm{FAR} = \mathrm{FA}/(\mathrm{FA}+H) \tag{7}$$

$$ACC = (H+\mathrm{CN})/(H+\mathrm{CN}+M+\mathrm{FA}) \tag{8}$$

where H is the number of actual cloudy FOVs that were correctly classified as cloudy, M is the number of cloudy FOVs that were incorrectly marked as clear (i.e., misses), FA is the number of clear FOVs that were incorrectly marked as cloudy (i.e., false alarm), and CN represents all remaining objects that were correctly classified as clear FOVs (i.e., correct negatives).

3. Experimental results

The typhoon Haima was used as the test object because the structure of its cloud is visually obvious; thus, it is suitable for estimating the accuracy of the abovementioned cloud detection schemes. We employed a true colour image obtained by the Advanced Himawari Imager (AHI) onboard Himawari-8 in addition to an AHI infrared channel image and the IASI level-2 *Flag Cloudiness* cloud product to evaluate the logistic regression method. The results of the cloud detection are shown in Figure 3. Figure 3(a) is the true colour image obtained by the AHI onboard Himawari-8. The typhoon vortex, which can be regarded as the structural feature of the cloud, is distinct within the typhoon area. Many thin clouds are distributed over the southern part of China and Japan in the non-typhoon region. Because some very thin clouds are not visible in the visible (VIS) band, the Himawari-8/AHI infrared channel was used to evaluate the cloud detection accuracy, as shown in Figure 3(b). As shown in Figure 3(c), the traditional AVHRR/IASI method shows different characteristics: the typhoon vortex structure has been destroyed, and the number of cloudy FOVs in the non-typhoon region is small

In the logistic regression algorithm, the cloud detection result depends on the trained coefficients. Here, the coefficients of the model on land are $\boldsymbol{w}_{\mathrm{land}} = (7.486, -26.966, 33.265, -17.704, 3.150)$, and those on the sea are $\boldsymbol{w}_{\mathrm{sea}} = (7.770, -7.320, 7.201, -2.419, 2.678)$. Figure 3(d) shows the result of the logistic regression algorithm. The structural characteristics of the vortices in the typhoon area are well preserved, and the distribution of clouds in the non-typhoon region also matches the *Fractional Cloud Cover* products more closely than the distribution obtained using the AVHRR/IASI algorithm.

Figure 3(e) depicts the results of the IASI *Fractional Cloud Cover* products; the white dots represent the cloudy FOVs (where the variable *Fractional Cloud Cover* is >0.8). As shown in Figure 3(e), the *Fractional Cloud Cover* products can accurately describe the distribution of clouds and the vortex structure of the typhoon. Because the *Fractional Cloud Cover* products are relatively accurate based on the validation report (Arriaga 2008), we regard them as the 'truth' in this study and use them to check the accuracy of the cloud detection methods. In addition, another IASI level-2 cloud product called the *Flag Cloudiness* was used, and we selected a *Flag Cloudiness* value of 4 for testing (i.e., the retrieved equivalent cloud amount exceeds 80%), as is shown in Figure 3(f).

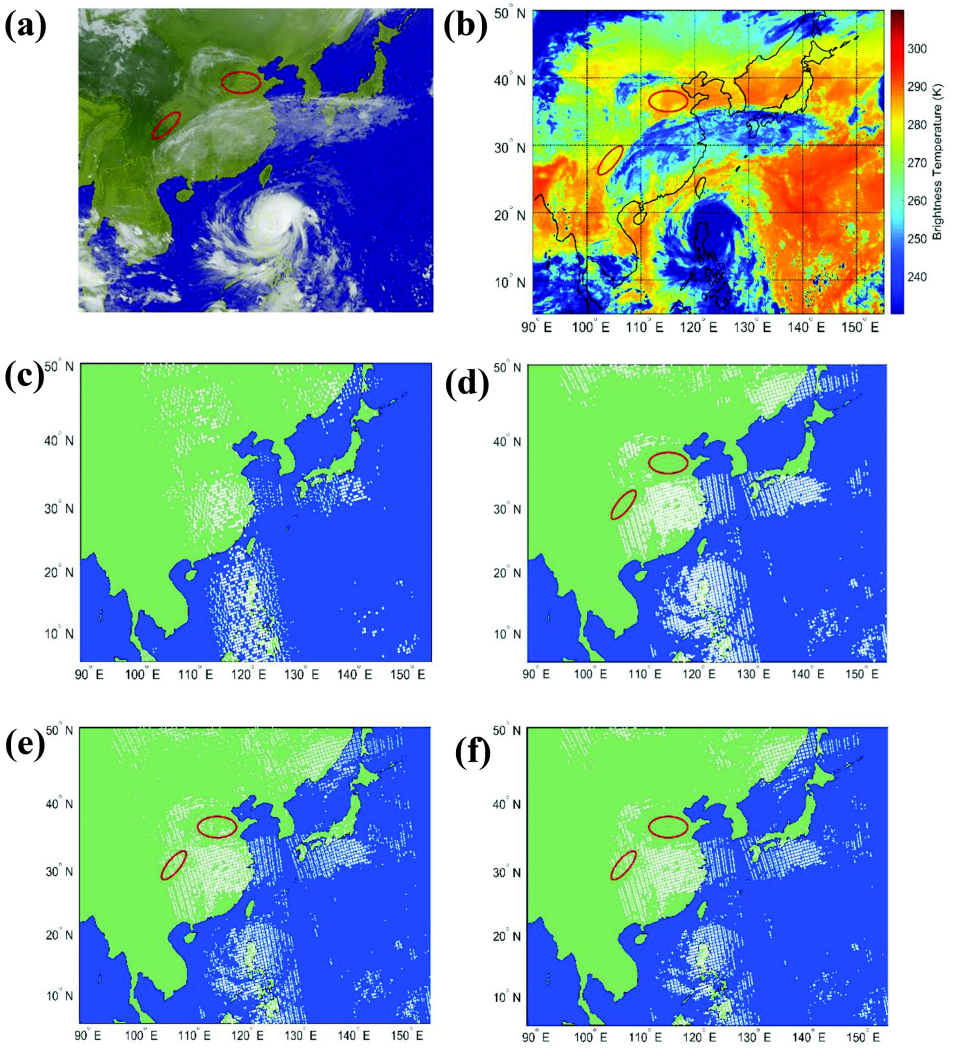

Figure 3. Cloud distribution of typhoon Haima at 12:00 UTC on 19 October 2016. The white dots represent the cloudy IASI FOVs. In the area demarcated by the red circle, the logistic regression method is superior to the level-2 products. (a) Himawari-8/AHI true colour image; (b) brightness temperature of Himawari-8/AHI channel 15 (807.7 cm^{-1}); (c) the results of AVHRR/IASI method; (d) the results of logistic regression scheme; (e) *Fractional Cloud Cover* products; (f) *Flag Cloudiness* products.

In the area demarcated by the red circle on the left, the Himawari-8/AHI true colour image (Figure 3(a)) and infrared channel image (Figure 3(b)) are clear; however, the IASI level-2 products (Figure 3(e,f) mistakenly detect the area as cloudy FOVs, whereas the logistic regression method (Figure 3(d)) can still correctly detect the area as clear FOVs. Moreover, in the area denoted by the red circle on the right, the Himawari-8/AHI VIS channel (Figure 3(a)), the infrared channel (Figure 3(b)), the level-2 *Flag Cloudiness* product (Figure 3(f)) and the proposed logistic regression method (Figure 3(d)) can detect the area as clear FOVs, but the level-2 *Fractional Cloud Cover* products mistakenly detect the area as cloudy FOVs. This result shows that the accuracy of the logistic regression method may exceed the accuracies of the IASI level-2 cloud products in some areas, although the proposed method utilizes the IASI level-2

products for its training datasets. Table 1 summarizes the values of the POD, FAR and ACC of the AVHRR/IASI and logistic regression methods on both land and sea. The POD of the AVHRR/IASI scheme (54% on land and 62% on sea) is much lower than that of the logistic regression scheme (81% on land and 94% on sea), which means that the AVHRR/IASI scheme misses more cloudy FOVs than does the logistic regression scheme. Moreover, the FAR of the AVHRR/IASI schemes is very high on both land and sea; on the former, the FAR is 0.46, which means that 46% of cloudy diagnoses are false alarms, and the FAR is slightly lower (0.43) on the sea. In addition, these results show that the logistic regression algorithm has a high accuracy (97% on sea and 88% on land) and that the AVHRR/IASI scheme has a much lower accuracy (75% at sea and 55% on land). It should be noted that the detection accuracies of both methods are lower on land. This may be because the land conditions are more complex than the sea conditions and because different surface types correspond to different surface emissivities, which may affect the detection results. When compared with the accuracy of the *Flag Cloudiness* product, the accuracy of the logistic regression method (95% at sea and 90% on land) is slightly higher than that of the *Flag Cloudiness* product (93% at sea and 88% on land); thus, the logistic regression method is superior to the *Flag Cloudiness* product in the left red area.

Figure 4 shows the values of the POD of clear and cloudy FOVs when the *Fractional Cloud Cover* variable is between 0.01 and 0.80. In Figure 4(a), 68% of the FOVs over the ocean and 60% of the FOVs on land are judged as clear FOVs when the *Fractional Cloud Cover* is between 0.01 and 0.2. As the *Fractional Cloud Cover* gradually increases, the POD of clear FOVs gradually decreases (Figure 4(a)), and the POD of cloud FOVs gradually increases (Figure 4(b)) both on sea and on land. This may be because the characteristics of clear FOVs are more obvious when the *Fractional Cloud Cover* is relatively small; thus, the possibility of detecting a clear FOV is greater, and the possibility of detecting a cloudy FOV is smaller. However, when the *Fractional Cloud Cover* is relatively large, the characteristics of cloudy FOVs are more obvious; thus, the probability of detecting a cloudy FOV is greater, and the probability of detecting a clear FOV is smaller.

Table 1. Values of the probability of detection (POD), false alarm rate (FAR) and accuracy (ACC) of the AVHRR/IASI and *Flag Cloudiness* products and the proposed logistic regression method for cloud detection at 12:00 UTC on 19 October 2016.

	AVHRR/IASI			Flag Cloudiness			Logistic Regression		
	POD	FAR	ACC	POD	FAR	ACC	POD	FAR	ACC
Land	0.54	0.46	0.55	0.81	0.08	0.88	0.84	0.05	0.9
Sea	0.62	0.43	0.75	0.9	0.05	0.93	0.92	0.04	0.95

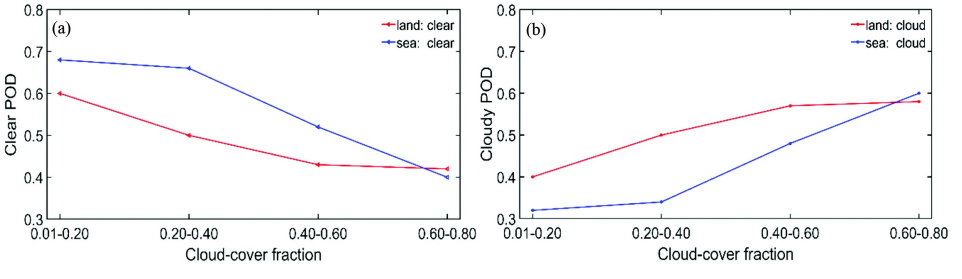

Figure 4. Values of the POD of clear and cloudy FOVs when the *Fractional Cloud Cover* is between 0.01 and 0.80. (a) clear POD; (b) cloudy POD.

In this study, the training sets for the logistic regression model originate from a small area. To affirm the efficacy of the proposed logistic regression algorithm, it is necessary to show that it works in many cases worldwide. Therefore, we employed two additional global IASI scan bands as test objects. Figure 5(a) shows the result of the logistic regression scheme, and Figure 5(b) depicts the result of the *Fractional Cloud Cover* products. Upon comparing Figure 5(a) with Figure 5(b), the logistic regression scheme performs well in the trained area (the red box); however, in the non-trained area, the cloud detection accuracy is reduced.

Table 2 summarizes the values of the POD, FAR, and ACC of the trained area and the non-trained area for the two different topographies. As shown in Table 2, the accuracy of the logistic method in the trained area is 90% at sea and 71% on land, while the accuracy in the non-training region is lower (73% at sea and 62% on land).

This result shows that the model trained by the data in a small area can effectively detect the trained area, but when the other non-trained area is detected, the accuracy of cloud detection will be reduced.

4. Discussion and conclusion

In this study, we proposed a novel cloud detection method based on the logistic regression model. In this method, we applied the radiances of four characteristic channels, namely, channel 921, channel 386, channel 306 and channel 224, in the IASI FOVs as the input features, and the *Fractional Cloud Cover* variable of the IASI level-2 cloud product was used as a label. Afterwards, the training set was divided into two parts, i.e., land and sea, and these parts were trained separately by a logistic model. In

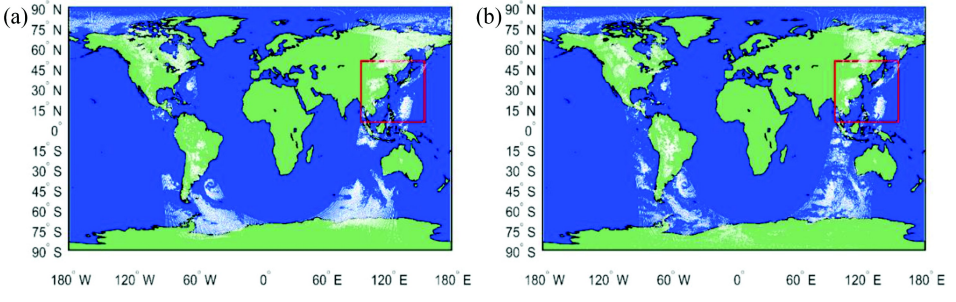

Figure 5. Cloud distributions of two global IASI scan bands during 00:00–03:00 UTC on 24 September 2016. The significance of the red rectangle is trained area. (a) The results of logistic regression scheme; (b) *Fractional Cloud Cover* products.

Table 2. Values of the probability of detection (POD), false alarm rate (FAR) and accuracy (ACC) in the trained area and non-trained area for the detection of clouds during 00:00–03:00 UTC on 24 September 2016.

	Within the trained area			Within the non-trained area		
	POD	FAR	ACC	POD	FAR	ACC
Land	0.72	0.14	0.71	0.65	0.52	0.62
Sea	0.92	0.09	0.90	0.78	0.32	0.73

addition, the proposed method was compared with the AVHRR/IASI method and IASI level-2 cloud products.

Compared with the AVHRR/IASI method, the logistic regression method can more accurately describe the distribution of clouds and the vortex structure of a typhoon, and its detection accuracy is significantly higher than that of the AVHRR/IASI method. However, both methods have a relatively lower accuracy when detecting clouds on land, which may be due to the complexity of the terrain. Compared with the IASI level-2 cloud parameter products, the results of the logistic regression method are closer to the true observations in some areas (i.e., the red circular areas in Figure 3). For the FOVs with *Fractional Cloud Cover* values between 0.01 and 0.8, the values of the POD of clear FOVs gradually decrease, while those of cloudy FOVs increase when the *Fractional Cloud Cover* increases. In addition, the model trained by data from a small area can detect the trained area effectively; however, when a non-trained area is detected, the cloud detection accuracy will be reduced.

The proposed method boasts two main advantages. First, the cloud detection accuracy of the proposed method is higher than that of the AVHRR/IASI method, and its accuracy may outperform that of the IASI level-2 products in some areas. Second, it is convenient to perform cloud detection by inputting only four channels into the trained model.

In this study, only four channels were used as training features. To improve the cloud detection accuracy, more channels will be selected as training features in the future. Moreover, to improve the cloud detection accuracy on land, new features, such as the surface emissivity and the normalized difference vegetation index (NDVI), will be added to training datasets in future research. Although the logistic regression model can accurately detect clouds in trained areas, the accuracy of the logistic regression model will decrease when the model is used to test a global case. To improve the global detection accuracy, we can obtain multiple models with different training datasets from certain areas of interest (e.g., over seas, deserts, grasslands, and polar areas). Furthermore, because cloud detection plays an important role in the assimilation of infrared hyper-spectral data, it will be possible to introduce the logistic regression model into the data assimilation system after we perform further experiments.

Funding

The authors acknowledge the support of the National Natural Science Foundation of China (41675097) and the National Natural Science Foundation of China (41375113).

ORCID

Tengling Luo http://orcid.org/0000-0002-1237-4064

References

An, Z., and Z. Shi. 2017. "Scene Learning for Cloud Detection on Remote-Sensing Images." *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing* 8 (8): 4206–4222. doi:10.1109/igarss.2016.7729176.

Arriaga, A. 2008. "Validation of the CO2 Slicing Method in the IASI L2 Processor" *EUMETSAT Meteorological Satellite Conference*.

Bauer, P., T. Auligné, W. Bell, A. Geer, V. Guidard, S. Heilliette, M. Kazumori, K. H. Min Jeong, C. Liu, and A. P. Mcnally. 2011. "Satellite Cloud and Precipitation Assimilation at Operational NWP

Centres." *Quarterly Journal of the Royal Meteorological Society* 137 (661): 1934–1951. doi:10.1002/qj.905.

Baum, B. A., V. Tovinkere, J. Titlow, and R. M. Welch. 1996. "Automated Cloud Classification of Global AVHRR Data Using a Fuzzy Logic Approach." *Journal of Applied Meteorology* 36 (36): 1519–1540. doi:10.1175/1520-0450(1997)036<1519:accoga>2.0.co;2.

Blumstein, D., and T. Phulpin. 2007. "In-Flight Performance of the Infrared Atmospheric Sounding Interferometer (IASI) on METOP-A." *Proceedings of SPIE* 6684. doi:10.1117/12.734162.

Blumstein, D., T. Maciaszek, T. Phulpin, and D. Simeoni. 2004. "IASI Instrument: Technical Overview and Measured Performances." *Proceedings of SPIE - the International Society for Optical Engineering* 5543: 196–207. doi:10.1117/12.560907.

Collard, A. D., and A. P. Mcnally. 2009. "The Assimilation of Infrared Atmospheric Sounding Interferometer Radiances at ECMWF." *Quarterly Journal of the Royal Meteorological Society* 135 (641): 1044–1058. doi:10.1002/qj.410.

Cox, D. R. 1959. "Corrigenda: The Regression Analysis of Binary Sequences." *Journal of the Royal Statistical Society* 21 (1): 238.

Eresmaa, R. 2015. "Imager-Assisted Cloud Detection for Assimilation of Infrared Atmospheric Sounding Interferometer Radiances." *Quarterly Journal of the Royal Meteorological Society* 140 (684): 2342–2352. doi:10.1002/qj.2304.

Feng, J., H. Xu, S. Mannor, and S. Yan. 2014. "Robust Logistic Regression and Classification." *Advances in Neural Information Processing Systems* 1: 253–261.

Hagolle, O., M. Huc, D. Villa Pascual, and G. Dedieu. 2010. "A Multi-Temporal Method for Cloud Detection, Applied to FORMOSAT-2, Venμs, LANDSAT and SENTINEL-2 Images." *Remote Sensing of Environment* 114 (8): 1747–1755. doi:10.3390/rs70302668.

Han, H., S. Lee, I. Jungho, M. Kim, M. I. Lee, M. Ahn, and S. R. Chung. 2015. "Detection of Convective Initiation Using Meteorological Imager Onboard Communication, Ocean, and Meteorological Satellite Based on Machine Learning Approaches." *Remote Sensing* 7 (7): 9184–9204. doi:10.3390/rs70709184.

Han, Y., B. Kim, Y. Kim, and W. H. Lee. 2014. "Automatic Cloud Detection for High Spatial Resolution Multi-Temporal Images." *Remote Sensing Letters* 5 (7): 601–608. doi:10.1080/2150704x.2014.942921.

Hilton, F., N. C. Atkinson, S. J. English, and J. R. Eyre. 2009. "Assimilation of IASI at the Met Office and Assessment of Its Impact through Observing System Experiments." *Quarterly Journal of the Royal Meteorological Society* 135 (639): 495–505. doi:10.1002/qj.379.

Jin, S., C. Homer, L. Yang, G. Xian, J. Fry, P. Danielson, and P. A. Townsend. 2013. "Automated Cloud and Shadow Detection and Filling Using Two-Date Landsat Imagery in the USA." *International Journal of Remote Sensing* 34 (5): 1540–1560. doi:10.1080/01431161.2012.720045.

Johnston, T., S. R. Young, D. Hughes, R. M. Patton, and D. White. 2017. Optimizing Convolutional Neural Networks for Cloud Detection. Paper presented at the Machine Learning on HPC Environments.

Komarek, P. R., and A. W. Moore. 2003. "Fast Robust Logistic Regression for Large Sparse Datasets with Binary Outputs." *Artificial Intelligence & Statistics*.

Li, J., T. J. James Gurka, S. W. Paul Menzel, and F. Sun. 2004. "AIRS Subpixel Cloud Characterization Using MODIS Cloud Products." *Journal of Applied Meteorology* 43 (8): 1083–1094. doi:10.1175/1520-0450(2004)043<1083:asccum>2.0.co;2.

Li, P., L. Dong, H. Xiao, and X. Mingliang. 2015. "A Cloud Image Detection Method Based on SVM Vector Machine." *Neurocomputing* 169: 34–42. doi:10.1016/j.neucom.2014.09.102.

Lu, D. 2007. "Detection and Substitution of Clouds/Hazes and Their Cast Shadows on IKONOS Images." *International Journal of Remote Sensing* 28 (18): 4027–4035. doi:10.1080/01431160701227703.

Mcnally, A. P. 2009. "The Direct Assimilation of Cloud-Affected Satellite Infrared Radiances in the ECMWF 4d-Var." *Quarterly Journal of the Royal Meteorological Society* 135 (642): 1214–1229. doi:10.1002/qj.426.

Mcnally, A. P., and P. D. Watts. 2003. "A Cloud Detection Algorithm for High-Spectral-Resolution Infrared Sounders." *Quarterly Journal of the Royal Meteorological Society* 129 (595): 3411–3423. doi:10.1256/qj.02.208.

Mcnally, A. P., P. D. Watts, J. A. Smith, R. Engelen, G. A. Kelly, J. N. Thépaut, and M. Matricardi. 2006. "The Assimilation of AIRS Radiance Data at ECMWF." *Quarterly Journal of the Royal Meteorological Society* 132 (616): 935–957. doi:10.1256/qj.04.171.

Pavelin, E. G., S. J. English, and J. R. Eyre. 2008. "The Assimilation of Cloud-Affected Infrared Satellite Radiances for Numerical Weather Prediction." *Quarterly Journal of the Royal Meteorological Society* 134 (632): 737–749. doi:10.1002/qj.243.

Rüthrich, F., B. Thies, C. Reudenbach, and J. Bendix. 2013. "Cloud Detection and Analysis on the Tibetan Plateau Using Meteosat and CloudSat." *Journal of Geophysical Research Atmospheres* 118 (17): 10082–10099. doi:10.1002/jgrd.50790.

Shi, M., F. Xie, Z. Yue, and J. Yin. 2016. Cloud Detection of Remote Sensing Images by Deep Learning. Paper presented at the Geoscience and Remote Sensing Symposium.

Suseno, D., P. Yuga, and T. J. Yamada. 2012. "Two-Dimensional, Threshold-Based Cloud Type Classification Using MTSAT Data." *Remote Sensing Letters* 3 (8): 737–746. doi:10.1080/2150704X.2012.698320.

Wang, L., and C. Cao. 2008. "On-Orbit Calibration Assessment of AVHRR Longwave Channels on MetOp-A Using IASI." *IEEE Transactions on Geoscience & Remote Sensing* 46 (12): 4005–4013. doi:10.1109/tgrs.2008.2001062.