Cloud clearing techniques over land for land surface temperature retrieval from the Advanced Along Track Scanning Radiometer

Article

Accepted Version

To appear in the *International Journal of Remote Sensing*
Vol. 00, No. 00, 00 Month 20XX, 1–17

Cloud Clearing Techniques over Land for Land Surface Temperature Retrieval from the Advanced Along Track Scanning Radiometer

C. E. Bulgin[a]* and H. Sembhi[b] and D. Ghent[b] and J. J. Remedios[b] and C. J. Merchant[c]

[a] *University of Edinburgh, UK*; [b] *University of Leicester, UK*; [c] *University of Reading, UK*

(*Received 00 Month 20XX; final version received 00 Month 20XX*)

We present five new cloud detection algorithms over land based on dynamic threshold or Bayesian techniques, applicable to the Advanced Along Track Scanning Radiometer (AATSR) instrument and compare these with the standard threshold based SADIST cloud detection scheme. We use a manually classified dataset as a reference to assess algorithm performance and quantify the impact of each cloud detection scheme on land surface temperature (LST) retrieval. The use of probabilistic Bayesian cloud detection methods improves algorithm true skill scores by 8-9 % over SADIST (maximum score of 77.93 % compared to 69.27 %). We present an assessment of the impact of imperfect cloud masking, in relation to the reference cloud mask, on the retrieved AATSR LST imposing a 2 K tolerance over a 3x3 pixel domain. We find an increase of 5-7 % in the observations falling within this tolerance when using Bayesian methods (maximum of 92.02 % compared to 85.69 %). We also demonstrate that the use of dynamic thresholds in the tests employed by SADIST can significantly improve performance, applicable to cloud-test data to be provided by the Sea and Land Surface Temperature Radiometer (SLSTR) due to be launched on the Sentinel 3 mission (estimated 2014).

Keywords: Cloud Detection; Land Surface Temperature, Satellite Retrieval

1. Introduction

There is increasing interest in exploiting LST for numerical weather prediction and climate modelling. LST data are also used to understand regional surface energy balance and impacts on agriculture and floods. Satellite sensors provide LST data globally at higher spatial and temporal resolution than the sparse in-situ network of near surface air temperature measurements so are better suited for modelling and weather prediction purposes. Accurately discriminating cloud features is fundamental in LST retrievals as cloud-contaminated brightness temperatures can lead to large LST uncertainties (Jin and Dickinson, 2002; Simpson and Gobat, 1996). Undetected cloud can produce LST values as low as 230 K in the absence of snow or ice conditions (Wan et al, 2004). In other cases, undetected clouds can produce significantly small LST discrepancies that the cloud top temperature is classified as a valid LST (Neteler, 2010).

Cloud detection over land is challenging due to small-scale spatial structure and temporal variations in surface reflectance and emissivity as a function of largely dynamic vegetation and land cover change (Simpson and Gobat, 1996). Direct

*Corresponding author. Email: cbulgin@staffmail.ed.ac.uk

aerosol effects need to be accounted for which will modify observed brightness temperatures and top of atmosphere reflectance (Ramanathan et al, 2001; Kaufman, Tanre and Boucher, 2002). Traditionally, threshold based methods have been used to detect cloud over land (Zavody and Mutlow, 1995; Birks, 2007; Frey et al, 2008; Kriebel et al, 2010; Derrien and Le Gléau, 2011) but these are limited by the subset of the data on which they are trained and often performance is compromised under extreme atmospheric conditions or over particular land surface types (for example snow and desert surfaces). Offline diagnostic tests and previous studies eg. (Plummer, 2008) have indicated that the limitations in the standard cloud detection (SADIST) for Along Track Scanning Radiometer (ATSR) instruments are significant false flagging of surface features by the normalised difference vegetation index and missed cloud if only the gross cloud and thin cloud tests are applied.

In this paper we examine the performance of the SADIST cloud mask for the Advanced Along Track Scanning Radiometer (AATSR) instrument in comparison with newly developed cloud detection techniques. AATSR is a dual view, polar orbiting instrument with a local equator overpass time of 10.30. It made observations between 2002-2012 in seven channels centred at 0.55, 0.66, 0.87, 1.6, 3.7, 10.8 and 12 μm spanning the visible and infrared spectrum (ESA, 2007). The first algorithm compared against the SADIST cloud mask is based on the same tests, using modified thresholds to optimise performance. The second is a restricted Bayesian approach comparing observed brightness temperatures with probability distributions derived from radiative transfer model simulations and a local climatology. The third is a full Bayesian classification scheme previously implemented for cloud detection over ocean (Merchant et al, 2005; Embury and Merchant, 2012), and modified for application over land. Our performance assessment is limited to daytime imagery for which we have a manually classified dataset to use as reference (Gomez-Chova et al, 2007).

We describe the formulation of each algorithm in detail in Section 2. In Section 3 we use performance metrics to define classification skill with reference to a manually classified dataset of land-based daytime satellite imagery (Gomez-Chova et al, 2007). We also assess the impact of the cloud detection skill on the retrieved LST and consider performance as a function of surface biome. We conclude the paper in Section 4 with a discussion on the confounding factors in cloud detection over land including land surface type, diurnal variability in surface temperatures and the presence of aerosol.

2. Cloud Detection Algorithms and Test Data

In this section we discuss the six cloud detection algorithms evaluated in this paper. The first is the standard method (Birks, 2007) referred to as SADIST, and the second an adaptation of this using modified threshold tests (UOL_CLOUD_2). The third approach (UOL_CLOUD_3) is a restricted Bayesian method using radiative transfer simulations in the thermal channels. From here on in they will be referred to as UOL_2 and UOL_3. The final three algorithms are full Bayesian classification schemes producing a probability of clear sky which can be thresholded to provide a mask tailored to user requirements. These have been developed independently to the UOL_2 and UOL_3 algorithms. We consider the Bayesian calculation using infrared and visible channels individually and then combining the information from both. From this point onwards these algorithms are referred to as 'Bayes Thermal', 'Bayes Vis' and 'Bayes Joint'. Finally we describe the manually classified imagery

used to assess cloud detection performance.

2.1 *Standard ATSR Cloud Detection (SADIST)*

The SADIST cloud detection over land is an updated version of the scheme used over the ocean (Birks, 2007; Zavody and Mutlow, 1995), based on the heritage of the AVHRR Processing Over Land cLoud and Ocean (APOLLO) cloud mask (Saunders, 1986). Three tests are applied over land to daytime satellite imagery; the gross cloud test, the thin cloud test and the visible cloud test.

For the gross cloud test (GCT) a land pixel is flagged as cloud contaminated if the observed 12 μm brightness temperature is less than a reference value. The physical basis for this test is that for thick clouds, optical depth at this wavelength is sufficient to give a cold thermal signature with reference to a clear-sky brightness temperature. The thresholds for this test are provided in a look-up table (LUT) of climatological mean surface temperature as a function of latitude and month (Birks, 2007). These LUTs were constructed from ground station measurements made between 1961-1990 interpolated onto one degree latitude bands (New et al, 2002).

For the thin cloud test (TCT) the difference between the 11 and 12 μm brightness temperatures are compared to a threshold value as a function of the 11 μm brightness temperature; if the difference exceeds this threshold the pixel is classified as cloud. Ice particle absorption and emission are at a maximum at these wavelengths creating large brightness temperature differences in comparison with clear-sky conditions (Saunders and Kriebel, 1989). The detection thresholds were generated by Zavody and Mutlow, (1995) using radiative transfer modelling and radiosonde data.

The visible channel cloud test applies two normalised visible channel indices to each pixel within a 2D classification scheme (Birks, 2007). The normalised difference vegetation index (NDVI) and the normalised difference index (NDI) are calculated using AATSR reflectance:

$$NDVI = \frac{REF_{0.87} - REF_{0.67}}{REF_{0.87} + REF_{0.67}} \tag{1}$$

$$NDI = \frac{REF_{0.87} - REF_{0.55}}{REF_{0.87} + REF_{0.55}} \tag{2}$$

where REF_x denotes reflectance in the 0.55, 0.67 and 0.87 μm channels. The classification scheme uses the NDVI to NDI ratio to define twelve zones representing different surface types including clear and cloud-covered vegetation, desert and ice surfaces. A pixel is flagged as cloud if its locus falls within any cloudy zone. The ATSR standard cloud mask is calculated by collating the results of each cloud flag. If any one of the flags detects cloud, the pixel is masked. Plummer, (2008) gives an example of the SADIST cloud mask performance, demonstrating its tendency to falsely flag clear-sky pixels as cloud and the need to further develop cloud detection methods for ATSR over land.

2.2 *UOL_2 Threshold Based Cloud Detection*

The baseline for this scheme is the SADIST thermal cloud flags and a modified NDVI mask. The principle difference to SADIST is that UOL_2 exploits information extracted from radiative transfer simulations of clear sky radiances in the AATSR thermal channels to determine enhanced detection thresholds. These thresholds help to address inadequacies within the existing SADIST cloud screening whilst allowing consistency with the expected cloud flags to be processed for the Sea and Land Surface Temperature Radiometer (SLSTR) - the successor to AATSR (Birks and Cox, 2011).

For the calculation of enhanced thresholds, clear sky radiances are calculated for the AATSR thermal channels using the fast radiative transfer model RTTOV 10.2 (Hocking et al, 2011) using temperature and water vapour profiles from the European Centre for Medium Wave Forecasting (ECMWF) ERA-Interim analyses (Dee et al, 2011) as input. Surface emissivity is extracted from the global Co-operative Institute for Meteorological Satellite Studies (CIMSS) dataset at a spatial resolution of 0.05° (Seeman et al, 2008). These are derived from the Moderate Resolution Imaging Spectroradiometer (MODIS) emissivity product (MOD11) incorporating wavelengths between 3.6 and 14.3 μm using a baseline fit method to infill spectral gaps between emissivity bands available in MOD11. Here, emissivities for the AATSR channels are extracted from the dataset and re-gridded onto the ECMWF grid in the radiative transfer simulation.

To obtain enhanced detection thresholds a uniform random distribution of simulated clear-sky brightness temperatures in the 3.7, 11 and 12 μm channels covering the entire AATSR mission were averaged onto a global 1° latitude-longitude grid allowing the mean and standard deviation to be calculated. Clear-sky pixels were identified using the SADIST cloud mask. Similarly, the mean and standard deviation of the brightness temperature differences (11-12 μm, 11-3.7 μm and 3.7-12 μm) were averaged into 1 K bins as a function of brightness temperature from 200 - 310 K. These collectively formed the basis for new cloud detection thresholds for each of the thermal flags used in the SADIST mask. A description of the UOL_2 flags is given below.

The UOL_2 gross cloud test thresholds were calculated from look-up tables (LUTs) based on the gridded 12 μm mean brightness temperature ($\bar{BT}_{12}$) and standard deviation (σ_{BT12}) in place of the 1° latitude means used in the SADIST algorithm. Using the updated LUT, the detection threshold is calculated as:

$$Thresh_{12} = (\overline{BT}_{12(lat,lon)} + 3\sigma_{BT12(lat,lon)}) + 2.0 \quad (3)$$

If the 12 μm brightness temperature is less than this threshold the pixel is flagged as cloud. The threshold ($Thresh_{12}$) considers the empirical rule and assumes that any clear sky pixel should be within 3σ of the mean (expected) clear sky brightness temperature. Furthermore, a contingency brightness temperature shift of 2 K is added to maximise detection of the cloud feature (similar to the 2 K limit used in Plummer, (2008)).

The UOL_2 thin cloud test LUTs were generated using the mean and standard deviation of the 12 minus 11 μm brightness temperature differences, calculated as a function of the 11 μm brightness temperature between 200 and 310 K. A pixel is flagged as cloud if the threshold ($Thresh_{11-12}$) is exceeded.

$$Thresh_{11-12} = (\overline{BT_{11} - BT_{12}} + (3\sigma_{BT11-BT12})) \quad (4)$$

The UOL_2 NDVI cloud flag remains identical to SADIST but incorporates a clear-sky restoral to minimise false flagging of surface features. The mean clear-sky 11 μm brightness temperature which shows a strong contrast between clouds, land, coastlines, rivers and lakes is used to re-evaluate any pixel flagged as 'cloud' by the NDVI. If the brightness temperature exceeds the 11 μm threshold ($Thresh_{11}$) the pixel is restored to clear-sky.

$$Thresh_{11} = \overline{BT_{11}} + 2.0K \quad (5)$$

2.3 *UOL_3 Restricted Bayesian Infrared Cloud Detection*

The UOL_3 algorithm derives the pixel-level cloud mask using a combination of simulated brightness temperatures and observational climatology. The approach is equally valid for both day and night-time retrievals as the cloud masking is independent of visible wavelength information. The L1b AATSR data have geolocation information stored at tie-points every 25 km across-track and 32 km along-track; these tie-points form the vertices of so-called granules. Expected brightness temperatures are simulated from coincident clear-sky profile information for each granule. As with UOL_2, simulations are performed using RTTOV-10.2. In this context, coincident clear-sky brightness temperatures in each granule are derived by bilinear interpolation between surrounding ECMWF profile locations and a temporal interpolation between the 6-hourly analysis fields. The observational climatology is acquired for each 5x5° grid cell for each of the 27 biomes and diurnal conditions (day/night) required by an offline enhanced LST retrieval algorithm for AATSR (Ghent, 2012). Inaccuracies in the auxiliary data create large biases and absent values in the retrieved LST data. As such, the existing auxiliary land cover dataset has been superseded by a near 1 km variant of the GlobCover classification (Arino et al, 2007) with a better representation of bare soil classes. The mean and standard deviations for clear-sky conditions are stored in a LUT (fully described in the UOL_2 section).

To calculate the clear-sky probability at the pixel location a probability density function (PDF) assuming a normal distribution is constructed from the simulated mean brightness temperature for the corresponding granule and the standard deviation of the brightness temperature from the observational climatology from the corresponding 5x5 ° grid cell for the given month, biome and diurnal state as shown in Figure 1. Pixels are identified as cloud if the observed brightness temperatures are outside the 95% lower confidence limit on any test. For daytime observations, the cloud flag is set if either the observed 12 μm brightness temperature or 11 minus 12 μm temperature difference fall outside of the 95% confidence levels of the corresponding simulated PDFs. For night-time observations, the 12 μm brightness temperature and the 11 minus 3.7 μm differences are used.

For granules where insufficient profile data are available to simulate the expected brightness temperatures, or where incompatibilities between the atmospheric and surface states result in an RTTOV error (this only occurs in less than 10^{-4} of granules) then the individual pixel cloud flags are instead derived from the UOL_2 cloud masking routines for daytime and from the SADIST cloud masking routines

for night-time.

2.4 *Bayesian Thermal Channel Cloud Detection*

The infrared only Bayesian cloud detection is an extension of the methodology successfully employed in cloud detection over the ocean for the ATSR instruments (Merchant et al, 2005; Embury and Merchant, 2012). The derivation of the Bayesian cloud classification scheme is described more fully in Merchant et al, (2005) and here we describe the algorithm only briefly, focusing on the modifications made to apply the algorithm over land surfaces.

The Bayesian calculation gives a probability of clear sky $P(c|\boldsymbol{y}^o, \boldsymbol{x}_b)$, given the observation vector ($\boldsymbol{y}^o$) and prior knowledge of the background state ($\boldsymbol{x}_b$).

$$P(c|\boldsymbol{y}^o, \boldsymbol{x}_b) = [1 + \frac{P(\bar{c})P(\boldsymbol{y}^o|\boldsymbol{x}_b, \bar{c})}{P(c)P(\boldsymbol{y}^o|\boldsymbol{x}_b, c)}]^{-1} \tag{6}$$

where c and $\bar{c}$ denote clear and cloud conditions respectively. $P(c)$ and $P(\bar{c})$ are the prior probabilities of clear and cloudy conditions which are set to fixed values of 0.1 and 0.9. We use nadir view observations and the 11 and 12 μm channels only as these can be used for both day and night-time cloud detection.

$P(\boldsymbol{y}^o|\boldsymbol{x}_b)$ is the probability of the observations given the background state. For clear conditions, the probability density function (PDF) is assumed to be Gaussian and we use the fast forward model RTTOV 11 beta version (Hocking et al, 2013) to simulate clear sky brightness temperatures. For this study we have updated the RTM to version 11 as this edition now includes the ability to simulate visible channel radiances (discussed further in Section 2.5). Brightness temperature simulations between RTTOV 10.2 and RTTOV 11 are found to be consistent. The clear-sky probability $P(\boldsymbol{y}^o|\boldsymbol{x}^b, c)$ is defined as:

$$P(\boldsymbol{y}^o|\boldsymbol{x}_b, c) = \frac{e(-\frac{1}{2}\Delta\boldsymbol{y}^T(\boldsymbol{H}^T\boldsymbol{B}\boldsymbol{H} + \boldsymbol{R})^{-1}\Delta\boldsymbol{y}}{2\pi|\boldsymbol{H}^T\boldsymbol{B}\boldsymbol{H} + \boldsymbol{R}|^{0.5}} \tag{7}$$

where $\boldsymbol{H}^T\boldsymbol{B}\boldsymbol{H}$ is the error covariance in the background state and the $\boldsymbol{R}$ matrix contains the observation noise and forward modelling uncertainties.

The radiative transfer model (RTM) uses numerical weather prediction (NWP) data from the ECMWF ERA-Interim Reanalysis dataset as input, defining both surface properties and the atmospheric profile. The model is run at a spatial resolution of 20 km latitude by 25 km longitude and the brightness temperatures are interpolated to the pixel location from the four surrounding profiles. Skin temperature is used as the prior surface temperature with an error estimate of 3.5 K to account for diurnal variability.

Surface emissivity is defined using the UWIREMIS database which interfaces with the RTTOV 11 model. These data are provided as a function of latitude, longitude, month and viewing geometry, based on MODIS data between 2003 and 2009 (Borbas and Ruston, 2010). These are provided at a spatial resolution of 0.1° and are used to correct the interpolated brightness temperatures at the pixel location.

The GTOPO30 digital elevation map (Gesch and Greenlee, 1996) is a global composite of land surface topography and is used to make a correction to the

background surface temperature at the pixel location using the dry adiabatic lapse rate. In the Bayesian calculation, this discrepancy between the surface temperature at the pixel location and the profile location is incorporated in the surface temperature tangent linear for propagation into the brightness temperature probability density function (PDF).

For cloudy conditions, the PDFs are calculated empirically and ingested as look-up tables (LUTs). These LUTs are generated from the entire AATSR mission, based on the SADIST cloud screening and then iterated using the Bayesian cloud detection over the ocean. No textural component is used in the Bayesian calculation over land as clear-sky brightness temperatures are more spatially heterogeneous than over ocean surfaces. When assessing algorithm performance we place a threshold of 0.1 on the clear-sky probability to generate a binary cloud mask. The fact that this is a value where the classification works well suggests that there is a bias towards underestimating the clear-sky probability. This is likely to be the result of greater uncertainties in the prior NWP data used to model clear-sky radiances over land (compared to over ocean where we use a threshold of 0.9). We would expect that as these terms are refined, a more intuitive threshold will be appropriate.

2.5 *Bayesian Visible Channel Cloud Detection*

We pilot visible only cloud detection over land using the 0.6, 0.8 and 1.6 μm channels, developing cloudy PDFs using the pixels defined as cloud in the manually classified dataset (described fully in Section 2.7). The PDF dimensions are the three visible channel reflectances ranging from 0-1.2 with a bin size of 0.04. Due to the limited number of pixels available to populate the PDF, no segregation is made with respect to solar zenith angle. Figure 2 shows a two dimensional representation of the visible channel PDFs for each pair of constituent dimensions. The PDFs show the greatest reflectance in the 0.8 μm channel. The 'blocky' appearance of the PDF is the result of the input data volume limiting the data resolution.

For clear scenes we use RTTOV 11 to model the top of atmosphere (TOA) reflectance for the three visible channels. The surface reflectance in the RTM is specified using the bidirectional reflectance distribution function (BRDF) atlas (Vidot and Borbas, 2012). This is constructed in a similar way to the UWIREMIS atlas at a monthly temporal resolution and 0.1° spatial resolution. Surface BRDF is specified for all wavelengths between 0.4-2.5 μm as a function of latitude and longitude. We interpolate the atmospheric component of the TOA surface reflectance to the pixel location and correct for the local pixel BRDF.

For the visible channels we also use the manually classified dataset to define clear-sky pixels to calculate a bias correction for the RTM. For each clear-sky pixel we take the difference between the observed and modelled TOA reflectance and average these differences as a function of biome (see section 2.3 for a full description of the biome data). This bias correction is then applied to the TOA reflectance prior to the Bayesian calculation. We also use these data to calculate the error covariance between the observation minus model differences as a function of biome for each channel combination. This error covariance calculation accounts for both forward model error and the observation noise and is used to define the R-Matrix in the clear-sky PDF calculation (equation 7).

The probability of negative reflectance is strictly zero, but the PDF always has some non-zero probability in this unphysical domain, and sometimes this can be significant for example over dark surfaces such as lakes under low wind speeds or

forests. It is necessary, therefore, to renormalise the Gaussian distribution so that the integrated probability over the positive domain is unity. For computational speed, this is done approximately, focussing on the dimension in which the most significant unphysical probabilities are observed. To do this we intersect the PDF at the zero reflectance line in each dimension and calculate the fraction of the PDF (f_x) that crosses into unphysical values. For the dimension with the greatest fraction, the remaining PDF is multiplied by the probability sealing factor (α_x) using:

$$\alpha_x = \frac{1}{1 - f_x} \tag{8}$$

We use a threshold of 0.1 to generate a binary cloud mask consistent with the Bayes Thermal algorithm and uncertainties in modelling surface reflectance.

2.6 *Bayesian Joint Cloud Detection*

For the joint Bayesian cloud detection we combine the infrared and visible channels to calculate the clear-sky probability using the 0.6, 0.8, 1.6, 11 and 12 μm channels. We assume independence between the thermal and visible channel probabilities calculating the probability of the observations given the background $P(\boldsymbol{y}^o|\boldsymbol{x}_b)$ as:

$$P(\boldsymbol{y}^o|\boldsymbol{x}_b) = P(\boldsymbol{y}^o_{VIS}|\boldsymbol{x}_b)P(\boldsymbol{y}^o_{IR}|\boldsymbol{x}_b) \tag{9}$$

where the 'IR' and 'VIS' subscripts denote thermal infrared and visible channels respectively. For cloud we use the long-term thermal PDFs with the newly developed visible PDFs. For consistency we use a threshold of 0.1 on the clear-sky probability to generate a binary cloud mask although typically the probabilities are much higher when information from both the infrared and visible channels are included.

2.7 *Manually Classified Test Data*

Testing the relative performance of competing cloud detection algorithms is difficult in the absence of a perfect classifier. Some testing datasets have been developed using expert classification to provide reference data for a limited number of scenes. We use one such dataset generated during the development of a synergistic AATSR and MERIS cloud detection algorithm (Gomez-Chova et al, 2009) hereafter referred to as the 'manual mask'.

The classification process is semi-automatic, identifying pixel clusters on the basis of brightness at the visible and near-infrared wavelengths, water vapour and oxygen absorption features, whiteness and the normalised difference vegetation index (NDVI) (Gomez-Chova et al, 2007). These clusters are then manually classified by expert inspection to provide a binary cloud mask that can be used as a reference dataset, which represents our best understanding of the image classification for each scene.

The manual mask is applicable to co-located AATSR and MERIS images on a reduced resolution merged grid but is provided with a mapping of the AATSR pixels to the reduced grid. This is then used to construct a full resolution AATSR cloud mask for twenty-one scenes which forms the basis of all the performance

tests described in Section 3. The twenty-one scenes cover five principle locations: Tomsk in Russia, Ouagadougou in Burkina Faso, Mongu in Zambia, Abracos Hill in Rondonia, Brazil and Cart Site in Oklahoma.

3. Cloud Detection Algorithm Comparison

3.1 *Performance Statistics*

In this section we assess the performance of each cloud detection algorithm across twenty-one test scenes. These cover a range of latitudes (tropics, mid-latitudes), biomes (vegetated, bare soils, rivers, lakes as described in Table 2) and cloud types (opaque and semi-transparent). For each scene we calculate the performance of each cloud detection algorithm relative to the manual mask using the following statistical measures: the percentage of perfect classification (PP), the hit rate (HR), the false alarm rate (FAR) and the true skill score (TSS) (Bengtsson and Hodges, 2005; Mackie et al, 2010). PP is the percentage of all pixels correctly classified as either cloud or clear, the HR is the percentage of cloud pixels correctly identified, the FAR is the percentage of clear pixels erroneously flagged as cloud and the TSS for cloud detection is the HR minus the FAR.

Before presenting the overall performance statistics we consider two detailed case studies which exemplify some of the classification problems relating to each algorithm. The first, shown in Figure 3 is over Ouagadougou on 21st January 2003 which highlights issues with the classification of cloud edges and cloud shadow. The top six panels show the performance of the six cloud detection algorithms with reference to the manual mask. In this case the SADIST, UOL_2, UOL_3 and Bayes Vis algorithms all miss the cloud edges. In the Bayes Thermal, Bayes Joint and to a lesser extent in UOL_3, flagging of cloud shadow is evident. A surface feature is also flagged by the SADIST algorithm in the bottom left corner of the image. Diagnostic maps giving a performance breakdown for the individual SADIST tests (not shown here) indicate that the NDVI is prone to grossly over-flagging clear-sky pixels as cloud and that the gross cloud test and thin cloud test can often miss significant amounts of cloud. In this example the Bayes Thermal and Bayes Joint algorithms perform best.

Figure 4 shows the second case study over Mongu, Zambia on 19th July 2003. Here we have a thick cloud feature to the bottom right of the scene and a thin cloud feature in the top left. This scene is also affected by aerosol, most clearly seen in the false colour image to the centre and top right of the image. Here we see a typical case of SADIST over-flagging (Plummer, 2008). Some of this may be due to the presence of aerosol but other regions are surface features, for example rivers. In UOL_2 a lot of this over-flagging has been eliminated although some rivers are still erroneously flagged as cloud. All algorithms show some over-flagging around the thin cloud to the left of the image. UOL_3 and Bayes Thermal both pick out a feature with a thermal signature to the left of the image centre which in the false colour image looks like aerosol. The Bayes Thermal also over-flags other regions in the top right hand side of the scene which may be aerosol and surface features in the bottom left of the image (the possible reasons for which are discussed in Section 4). In the Bayes Vis some surface features are also identified. The surface feature over-flagging is less prevalent in the Bayes Joint but still evident.

A number of inter-related factors will affect the performance of the cloud detection including the presence of aerosol, cloud type and land surface type (biome) which we discuss in more detail in Section 4. Here we present the performance

statistics across all scenes and then recalculate the statistics by subdividing the scenes according to the background aerosol optical depth (AOD) taken from MODIS daily AOD fields (Acker and Leptoukh, 2007). These L3 data at 1x1° resolution are extracted over each scene domain and provide an indication of the average aerosol loading over the region on the given day. Attempts are not made to match observations at a pixel level as coincident L2 data are not always available. We classify 'clear scenes' as those with an AOD < 0.3, those with a low aerosol concentration with AOD between 0.3 - 0.4, and those with a high aerosol load with an AOD greater than 0.4. Table 3 shows the total number of clear and cloud pixels falling into each classification demonstrating that a significant number of pixels are included in each class.

Table 4 shows the performance statistics for all scenes. These clearly demonstrate the problems with the SADIST cloud detection algorithm which typically flags surface features as cloud giving the highest FAR (16.82 %) and the lowest TSS (69.27 %). The UOL_3 and Bayes Joint algorithms achieve significantly higher TSS's (77.93 % and 77.63 % respectively). The Bayes Joint algorithm has a higher FAR (10.82 %) than the UOL_3 (5.87 %) but also has a higher hit rate (88.45 % compared with 83.80 %). The UOL_2 algorithm using the modified thresholds from the standard mask also shows a significant improvement on SADIST with a TSS of 76.69 %.

Considering only the 'clear scenes' with AOD < 0.3 we see the best performance from all algorithms. UOL_3 has the highest TSS (78.31 %) closely followed by UOL_2 and Bayes Joint with TSS's of 77.74 % and 77.43 % respectively. The SADIST cloud mask over-flags surface features with a FAR of 17.34 % (also giving a high HR of 89.83 %). Bayes Thermal also has a tendency to over-flag with an FAR of 12.7 %. The hit rates for Bayes Thermal and Bayes Joint are 87.06 and 86.51 % respectively. Bayes Vis has the lowest hit rate at 79.38 %.

For low aerosol loadings (AOD 0.3 - 0.4) we see a reduction in the overall performance of all algorithms. Bayes Joint has the highest TSS (73.38 %) and SADIST the lowest (49.83 %). In all cases the FAR falls but the HR is also much lower. Bayes Joint has the highest hit rate (79.47 %) but for UOL_3 this falls to 66.22 %. The MODIS daily Angström exponent fields (Acker and Leptoukh, 2007) indicate that in these scenes the aerosol is predominantly in the coarse mode. Coarse mode aerosols such as dust are likely to reflect incoming solar radiation and may have a week cooling thermal signature. If the aerosol overlays the cloud it may change its apparent thermal and visible channel properties (as viewed from space) making it more difficult to detect. Overall cooling of the scene may lead to colder surface features such as rivers being less likely to be flagged, reducing the FAR.

For high aerosol loadings (AOD > 0.4) we see a tendency for all algorithms to over-flag. SADIST and the Bayes Thermal algorithms are most affected with FAR's of 30.28 % and 31.8 % respectively, although all algorithms have a FAR close to or exceeding 20.0 %. The HR also increases significantly showing the tendency of the cloud detection algorithms to flag everything in the presence of aerosol. The TSS scores typically fall between the clear scene and low aerosol loading values.

3.2 *LST Impacts*

Imperfections in the cloud detection algorithm applied to the offline enhanced LST dataset (Ghent, 2012) will result in spatial sampling uncertainties in the retrieved surface temperatures. At a pixel level, missed cloud will lead to a cloud contaminated LST, usually colder than the true surface temperature, whilst false

flagging of surface features will lead to a loss of data. Many data users will re-grid the full resolution data at a coarser spatial resolution where these uncertainties need to be correctly propagated. For missed cloud, a spatially averaged LST is likely to be colder than the true value, whilst for falsely flagged surface features the spatially averaged LST may be either warmer or colder depending on the surface temperature and biome of the pixels excluded and the extent of the false flagging.

No attempts have been made previously to physically quantify the spatial sampling uncertainty on LST due to errors in cloud detection over land. Here we take a first look at this problem using the manually masked LST field as our reference dataset. For each of the test scenes, we divide the image into 3x3 pixel boxes across which we take the average LST for all pixels classified as 'clear' by the algorithm tested. We then calculate the difference between the results from the manually masked data field and the tested algorithm. In order for a comparison to be made, at least three of the nine pixels used in each box have to be classified as clear. If less than three 'clear' pixels are available for one of the two cloud masks compared then the 3x3 pixel box is classed as 'over-flagged' (if the manual mask has three or more clear pixels) or 'under-flagged' (if the manual mask has less than three clear pixels). The target accuracy for AATSR LST retrievals during the day is 2.5 K globally (Llewellyn-Jones et al, 2001) so we impose a tolerance of plus or minus 2 K on the LST difference between the manually classified data and the tested algorithm. This tolerance is generous as other sources of uncertainty (instrument noise, a priori information and retrieval error) are likely to sum to around 1 K (personal communication, Darren Ghent). For those boxes where the LST difference exceeds the 2 K tolerance but both algorithms have sufficient clear-sky pixels available, we label these as 'large uncertainties' which would not be acceptable given the target accuracy.

Table 5 provides a summary of these results, again subdivided into categories determined by aerosol loading. Considering 'all' cases first we find that all algorithms increase the percentage of LST retrievals falling within the 2 K tolerance compared to SADIST. For UOL_2, UOL_3, Bayes Vis and Bayes Joint this is an increase of between 5-6.5 % in the total number of pixels with the highest percentage for UOL_3 (92.0.2 %). The large uncertainties for the SADIST mask are 0.71 % of boxes, falling to 0.31 % for the Bayes Joint algorithm. Boxes classified as 'over-flagged' or 'under-flagged' will also introduce a spatial sampling uncertainty. Overall, the differences between opaque cloud temperatures and surface temperatures are likely to be larger than the differences between surface temperatures across neighbouring biomes (in the case of falsely flagged pixels), although for semi-transparent cloud inter-biome differences could be greater. The preference in terms of LST retrievals is probably over-flagging (within reason as this also indicates a loss of data). We find that Bayes Joint gives the lowest under-flag percentage (2.11 % compared with 3.22 % for UOL_3). Both SADIST and Bayes Thermal have a significant problem with over-flagging (10.86 % and 10.49 % of boxes respectively).

For more pristine conditions where AOD < 0.3 the percentages of boxes within the 2 K tolerance are slightly higher for all algorithms. UOL_3 has the highest percentage (92.89 %) closely followed by Bayes Joint (91.68 %), UOL_2 (91.36 %) and Bayes Vis (91.07 %). Bayes Thermal and Bayes Joint have the fewest number of boxes with large uncertainties (0.25 % and 0.27 % respectively). Bayes Joint has a lower under-flag percentage compared to UOL_3 (2.58 % and 3.55 % respectively), but the UOL_3 over-flag percentage is lower (3.16 % compared with 5.47 %). SADIST has a high percentage of over-flagged boxes (11.03 %) but the

lowest under-flag percentage (1.91 %) due to its tendency to grossly over-flag.

For AOD values between 0.3-0.4 we see a slight increase in the number of boxes falling within the 2 K tolerance due to the reduction in the false alarm rate, despite the corresponding fall in the hit rate. The highest percentages are for UOL_3 and UOL_2 (94.6 % and 94.17 % respectively). The over-flag percentage drops for all algorithms with UOL_2 having the lowest value (0.6 %). The coincident fall in the hit rate seen in the performance statistics is reflected in the increased under-flag percentage. Bayes Joint is most resilient to this with an under-flag value of 2.33 % whilst for SADIST this reaches 5.14 %.

For AOD > 0.4 we see a reduction in the number of boxes falling within the 2 K tolerance. SADIST values fall from 86.59 % in the AOD < 0.3 case to 79.12 %. With the exception of Bayes Thermal which is very sensitive to the presence of aerosol, all other algorithms still outperform SADIST. Bayes Vis is the least affected by aerosol with 88.36 % of boxes falling within the 2 K tolerance compared with 91.07 % for AOD < 0.3 The large uncertainty percentage is highest for SADIST (1.02 %) and more than 50 % lower for all other algorithms with a minimum of 0.31 % for Bayes Joint. Bayes Joint has the lowest under-flag value (0.45 %) and SADIST the highest (1.37 %) indicative of significant over-flagging in all algorithms. Bayes Thermal and SADIST have the highest over-flagging values (19.9 % and 18.5 % respectively) and Bayes Vis the lowest (9.91 %).

3.3 *False Alarm Rate Analysis*

The false alarm rate (FAR) gives a direct measure of each algorithms ability to correctly identify cloud over a range of surface types and here we analyse performance as a function of biome. For each scene, the percentages of clear-sky pixels falsely flagged as cloud are calculated for four different biomes: urban areas, water bodies (including lakes and rivers), mosaic vegetation and bare soils (including mixed bare ground and entisols) as defined by the biome classification in the enhanced LST retrieval algorithm (Ghent, 2012). Only scenes in which there are more than 50 pixels for the given biome are included in the analysis.

Figure 5 shows that over the urban and artificial surfaces the behaviour of the SADIST cloud mask is most inconsistent covering the full range of possible FARs. With the exception of Tomsk on 12th August 2004 (scene 17), the Bayes Vis algorithm has a higher FAR than the Bayes Thermal algorithm showing that the change in surface reflectance is more significant than changes in emissivity for urban areas. The Bayes Joint, UOL_2 and UOL_3 algorithms consistently outperform the SADIST classification over this biome. For some scenes the FAR for UOL_2 is as much as 60 % lower than SADIST showing the benefit of the clear-sky restoral test. For Cart Site on 5th June 2006 (scene 6) there is high aerosol loading and all algorithms have a high FAR ranging between 45-60 %.

For waterbodies (lakes and rivers) we find in the scenes not affected by heavy aerosol loading the SADIST FAR is typically around 50 %. The Bayesian algorithms often don't perform as well as the UOL_2 and UOL_3 algorithms, likely due to the same reasons described above where water bodies are not adequately represented in the surface emissivity and reflectance atlases. For scenes affected by high coarse or mixed mode aerosol loading (Cart Site on 5th June 2006 - scene 6) all algorithms give high false alarm rates over waterbodies. For those where fine mode aerosol dominates (Mongu on 17th September 2003 - scene 15, and Tomsk on 23rd June 2005 - scene 18) the FAR is generally not as high.

We present a single vegetation class for comparison with bare soil areas com-

prising of 50-70% grassland, shrubland and forest and 20-50 % cropland. Over this biome we find that the SADIST false alarm rate tends to be more consistent but generally higher than the other algorithms (between 20-40 %) with the exception of in the presence of significant coarse mode aerosol where all algorithms have a higher FAR. False alarm rates for Bayes Joint, UOL_2 and UOL_3 are generally between 0-10% although Bayes Joint is occasionally higher if there is a problem with either the Bayes Vis or Bayes Thermal classifications.

Over bare soils fewer pixels are available so the results are harder to interpret but are included as a comparison against vegetated surfaces. Cloud detection over desert surfaces is a particularly difficult problem due to the high surface reflectance and large diurnal variation in surface temperature. FARs for all algorithms are below 40% in all but the aerosol affected scene (Ouagadougou on 15th July 2003 - scene 12), but the relative performance of each algorithm is mixed with SADIST often having lower values than the other algorithms. This analysis suggests that the UOL_3 and Bayes Joint algorithms outperform SADIST most significantly over urban and vegetated surfaces in the twenty-one scenes considered.

4. Discussion

In this paper we considered the impact of three key variables on cloud detection algorithm performance and the subsequent spatial sampling uncertainties in LST introduced by imperfect image classification. Cloud type, aerosol loading and land surface cover all affect performance and different algorithms respond in different ways to changes in these variables. We find that all six algorithms are less effective at discriminating between cloud and clear observations in the presence of aerosol. Performance is dependent on aerosol loading and is also likely to be affected by aerosol size, composition and altitude. Ideally for LST retrieval, masking algorithms should include an aerosol flag as undetected aerosol will exhibit a thermal signature. For low aerosol loadings where LST retrieval is made, the brightness temperature perturbation should be accounted for within the algorithm.

The UOL_2 and UOL_3 algorithms both work on a definition of 'clear-sky'. High aerosol loadings modify the observed brightness temperatures and top of atmosphere reflectance so that these no longer represent clear-sky conditions leading to increased flagging of pixels not containing cloud. For the Bayesian algorithms both clear and cloudy PDFs are represented in the calculation. The cloudy PDFs used in the Bayes Thermal classification are likely to contain some aerosol pixels flagged as 'not clear' increasing the likelihood of flagging aerosol as cloud. The Bayes Vis algorithm which does not include any aerosol pixels in the cloudy PDF as these are generated using the manually masked data shows the best performance under high aerosol loading. Aerosol may also have a smaller impact on top of atmosphere reflectance than brightness temperature explaining the increased performance of Bayes Vis relative to the other algorithms under these conditions.

Under high aerosol loading, for the Bayes Joint and UOL_3 algorithms which have the best cloud detection performance overall, less than 86% of observations fall within the 2 K tolerance imposed here on LST measurements. Further development is necessary to improve classifier performance under these conditions due to large increases in the false alarm rate. For UOL_3 there is the potential to include aerosol loadings in the radiative transfer simulations; and indeed this is intended for future evolution of this scheme. The Bayesian classifier can be developed into a multiway classifier as has been successfully trialled over high latitude sea-ice regions for AATSR (Bulgin et al, 2013) and over land for the SEVIRI instrument (Mackie,

2009). Inclusion of fine and coarse mode aerosol classes could significantly improve image classification.

Cloud detection performance is also affected by cloud type. Cloud edges and semi-transparent clouds are consistently difficult for cloud detection algorithms to identify and the Bayes Joint results in Section 3 show the value of using information at both visible and infrared wavelengths in these regions. For the Bayesian schemes, cloud detection in these regions could be improved with the development of new land based PDFs derived from multiple years of ATSR observations. The Bayes Joint results presented here are competitive using existing ocean based and proof of concept PDFs so improvement would be expected from the development of land based PDFs which include thin clouds and cloud edges over land surfaces. The UOL_2 and UOL_3 algorithms do not explicitly include cloud information so detection of semi-transparent cloud or cloud edges falling within the tail of the clear-sky distribution is difficult.

Biome also has an impact on the ability of the cloud detection algorithm to correctly identify clear-sky scenes. Although limited by the number of pixels available, and complicated by seasonal changes in land surface cover, the biome analysis in Section 3 gives some insight into algorithm performance over different land cover types. For the Bayes Thermal, Bayes Vis and Bayes Joint classifications, urban areas and waterbodies are unlikely to be well resolved in the surface emissivity and reflectance atlases used in the simulation of clear-sky observations, as these have a global resolution of 0.1° latitude and longitude (approximately 10 km^2 at the equator).

There are also some uncertainties associated with the NWP data used for Bayes Thermal, Bayes Joint and UOL_3 . In some cases the surface temperatures at the NWP tie points from which the simulated brightness temperatures are interpolated may not be representative of the surface temperature in the pixel to be classified, if the biome is significantly different. The Bayesian scheme includes a surface temperature correction for altitude in the surface temperature tangent linear after interpolating simulated brightness temperatures from the four surrounding NWP profile locations. For the UOL_3 algorithm linear interpolation is used but this will be changed to polynomial interpolation in future development work to reduce interpolation errors. Surface temperature data are also linearly interpolated in time, which will not capture the peak of the diurnal cycle (typically between 1300-1400 hours (Jin and Dickinson, 1999)) as they are available at 1200 and 1800 hours.

The comparisons presented here are limited to a small number of scenes (although comprising a larger total number of pixels than evaluated in previous studies) and further work is needed to fully assess algorithm performance globally. Analysis has also been limited to daytime imagery only, for which the reference dataset was available. Manual classification of satellite imagery to provide a reference dataset under night-time conditions is more difficult but comparisons could be made against in-situ data. These in-situ data could also be used to asses the long-term impact of each cloud detection algorithm on the retrieved LST.

5. Conclusions

For LST retrieval purposes, the Bayes Joint and UOL_3 cloud detection algorithms consistently outperform the SADIST scheme with significantly improved true skill scores. Under clear sky conditions these algorithms give smaller spatial uncertainties in retrieved LST than threshold based methods and Bayes Joint misses the fewest cloud pixels. Probabilistic based methods for cloud detection show the most

potential for correctly classifying images under extreme conditions including very wet or very dry atmospheres, snow/desert surfaces and biomes with large diurnal variability in surface temperature and both are recommended for future development. UOL_2 also shows significant improvements on the SADIST cloud detection scheme with the use of dynamic thresholds although is limited in extreme conditions by the difficulties of setting threshold values for the tests. It is however important in the context of data from the Sea and Land Surface Temperature Radiometer (SLSTR) due to be launched on the Sentinel 3 mission [estimated 2014] as the same cloud flags will be provided in the L1B data as those available for the ATSR instruments.

Evaluation of cloud detection algorithm performance is of critical importance to the development of LST data with low uncertainties from spatial sampling errors and is currently restricted by availability of perfectly classified data to use as a reference. To rigorously test cloud detection algorithms we recommend the development of a more extensive independent test database not used for algorithm development. Simpson, Schmidt and Harris, (1998) used a large test dataset to identify residual clouds in brightness temperature and reflectance data, in clear-sky scenes using different cloud detection algorithms over ocean, but this did not include a reference cloud mask. Cloud detection over land and ocean surfaces needs to be considered separately and the 'best' cloud detection is currently achieved by expert inspection. Although time consuming, this is advantageous to cloud detection development and essential for satellite retrievals with weather and climate applications. The test database should include observations covering a range of atmospheric conditions, cloud types, aerosol loading, biomes, solar and viewing geometry and include tools to analyse cloud detection performance.

6. Acknowledgements

We are grateful to Luis Gomez-Chova for providing manually masked data from the ESA SYNERGY and CCI Cloud projects for use as a reference dataset in this paper. We acknowledge the MODIS mission scientists at NASA for the provision of daily aerosol optical depth and Angström exponent fields. C. E. Bulgin was funded by the UK Natural Environment Research Council, grant reference NE/H004130/1.

References

Acker, J. G. and Leptoukh, G. 2007. "Online analysis enhances use of NASA Earth science data." *Eos, Trans. AGU.* 88(2): 14 -17.

Arino, O., Leroy, M., Bicheron, P., Brockman, C., Defourny, P., Vancutsem, C., Achard, F., Durieux, L., Bourg, L., Latham, J., Di Gregorio, A., Witt, R., Herold, M., Sambale, J., Plummer, S. and Weber, J-L. 2007. "GlobCover: ESA service for Global Land Cover from MERIS". *1-4244-1212-9/07 IEEE*, 2412-2415.

Bengtsson, L. and Hodges, K. I. 2005. "On the impact of humidity observations in numerical weather prediction." *Tellus A.* 57: 701-708.

Birks, A. R. 2007. "Improvements to the AATSR IPF relating to Land Surface Temperature Retrieval and Cloud Clearing over Land." AATSR Technical Note, Rutherford Appleton Laboratory, Chilton, Didcot, Oxfordshire, UK.

Birks, A., and Cox, C. 2011. "SLSTR: Algorithm Theoretical Basis Definition Document for Level 1 Observables." *STFC Report.*

Borbas, E. E. and Ruston, B. C. 2010. "The RTTOV UWiremis IR land surface emissivity

module." *EUMETSAT Satellite Application Facility on Numerical Weather Prediction*. NWPSAF-MO-VS-042. Version. 1.

Bulgin, C. E., Eastwood, S., Embury, O., Merchant, C. J., and Donlon, C. 2014. "The Sea Surface Temperature Climate Change Initiative: Alternative image classification algorithms for sea-ice affected oceans." *Remote Sensing of Environment*. http://dx.doi.org/10.1016/j.rse2013.11.022.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M.,van de Berg, L., Bidlot, J., Bormann, N., Delson, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Holm, E. V., Isaksen, L., Kallberg, P., Kohler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J. J., Park, B. K., Peubey, C., de Rosnay, P., Tavolato, C., Thepaut, J. N., and Vitart, F. 2011. "The ERA-Interim reanalysis: configuration and performance of the data assimilation system." *Quarterly Journal of the Royal Meteorological Society* 137: 553-597.

Derrien, M. and Le Gléau, H. 2011. "MSG/SEVIRI cloud mask and type from SAFNWC". *International Journal of Remote Sensing* 26(21): 4707-4732.

Embury, O. and Merchant C. J. 2012. "A Reprocessing for Climate of Sea Surface Temperature from the Along-Track Scanning Radiometers: A New Retrieval Scheme." *Remote Sensing of Environment*. 47-61. DOI: 10.1016/j.rse.2010.11.020.

European Space Agency. 2007. "EnviSat AATSR Product Handbook." Issue 2.2.

Frey, R. A., Ackerman, S. A., Liu, Y., Strabala, K. I., Zhang, H., Key, J. R. and Wang, X. 2008. "Cloud Detection with MODIS. Part I: Improvements in the MODIS Cloud Mask for Collection5". *Journal of Atmospheric and Oceanic Technology* 25: 1057-1072.

Gesch, D. and Greenlee, S. 1996. *GTOPO30 Documentation*. US Department of the Interior U.S. Geological Survey. http://eros.usgs.gov//Find_Data/Products_and_Data_Available/GTOPO30.

Ghent, D. 2012. "Land Surface Temperature Validation and Algorithm Verification." *ESA Report*.

Gomez-Chova, L., Munoz-Mari, J. Izquierdo-Verdiguier, E., Camps-Valls, G., Calpe, J. and Moreno, J. 2009. "Cloud Screening with Combined MERIS and AATSR Images." *IEEE International Geoscience and Remote Sensing Symposium*. IV 761.

Gomez-Chova, L., Campls-Valls, G., Guanter, L. and Moreno, J. 2007. "Cloud-Screening Algorithm for ENVISAT/MERIS Multispectral Images." *IEEE Transactions on Geoscience and Remote Sensing* 45(12): 4105-4118.

Hocking, J., Rayer, P., Saunders, R., Matricardi, M., Geer, A. and Brunel, P. 2011. "RTTOV v10 Users Guide." *EUMETSAT Satellite Application Facility on Numerical Weather Prediction*. NWOSAF-MO-UD-023. Version 1.5.

Hocking, J., Rayer, P., Rundle, D., Saunders, R., Matricardi, M., Geer, A., Brunel, P. and Vidot, J. 2013. "RTTOV v11 Users Guide." *EUMETSAT Satellite Application Facility on Numerical Weather Prediction*. NWPSAF-MO-UD-028. Version. 1.2.

Jin, M. and Dinkinson R. E. 1999. "Interpolation of surface radiative temperature measured from polar orbiting satellites to a diurnal cycle." *Journal of Geophysical Research* 104(D2): 2105-2116.

Jin, M. and Dickinson, R. E. 2002. "New observational evidence for global warming from satellite." *Geophysical Research Letters* 29(10).

Kaufman, Y. J., Tanre, D., and Boucher, O. 2002. "A satellite view of aerosols in the climate system." *Nature* 419: 215-223.

Kriebel, K. T., Gesell, G., Kästner, M. and Mannstein, H. 2010. "The cloud analysis tool APOLLO: Improvements and validations". *International Journal of Remote Sensing* 24(12): 2389-2408.

Llewellyn-Jones, D., Edwards, M. C., Mutlow, C. T., Birks, A. R., Barton, I. J., and Tait, H. 2001. "AATSR: Global-change and surface-temperature measurements from Envisat." *ESA Bulletin 105*.

Mackie, S. 2009. "Exploiting weather forecast data for cloud detection." *PhD Thesis*. University of Edinburgh, School of Geosciences.

Mackie, S. and Embury, O. and Old, C. and Merchant, C. J. and Francis, P. 2010. "Gen-

eralized Bayesian cloud detection for satellite imagery. Part 1: Technique and validation for night-time imagery over land and sea." *International Journal of Remote Sensing* 31(10): 2573-2594.

Merchant, C. J., Harris, A. R., Maturi, E., and MacCallum, S. 2005. "Probabilistic physically-based cloud screening of satellite infra-red imagery for operational sea surface temperature retrieval." *Quarterly Journal Royal Meteorological Society* 131: 25735-2755.

Neteler, M. 2010. "Estimating daily land surface temperatures in mountainous environments by reconstructed MODIS LST data." *Remote Sensing* 2: 333-351.

New, M., Lister, D., Hulme, M., and Making, I. 2002. "A high-resolution data set of surface climate over global land areas." *Climate Research* 21: 1-25.

Plummer, S. E. 2008. "The GLOBCARBON cloud detection system for the Along-Track Scanning Radiometer (ATSR) sensor series." *IEEE Transactions on Geoscience and Remote Sensing* 46(6): 1718-1727.

Ramanathan, V., Crutzen, P. J., Kiehl, J. T., Rosenfield, D. 2001. "Aerosols, Climate, and the Hydrological Cycle." *Science.* 294(5549): 2119-2124.

Saunders, R. W. 1986. "An automated scheme for the removal of cloud contamination from AVHRR radiances over western Europe." *International Journal of Remote Sensing* 7: 867-888.

Saunders, R. W., and Kriebel, K. T. 1988. "An improved method for detecting clear sky and cloudy radiances from AVHRR data". *International Journal of Remote Sensing* 9: 123-150.

Seeman, S. W., Borbas, E., E., Knuteson, R. O., Stephenson, G. R., and Huang, H. L. 2008. "Development of a global infrared land surface emissivity database for application to clear sky sounding retrievals from multispectral satellite radiance measurements." *Journal of Applied Meteorology and Climatology* 47: 108-123.

Simpson, J. J. and Gobat, J. I. 1996."Improved cloud detection for daytime AVHRR scenes over land." *Remote Sensing of Environment* 55(1): 21-49.

Simpson, J. J., Schmidt, A., and Harris, A. 1998. "Improved cloud detection in along-track scanning radiometer (ATSR) data". *Remote Sensing of Environment* 65: 1-24.

Vidot, J., and Borbas, E. 2012. RTTOV land surface VIS/NIR BRDF module.

Wan, Z., Zhang, Y., Zhang, Q., and Li, Z-L. 2004. "Quality assessment and validation of the MODIS global land surface temperatures." *International Journal of Remote Sensing* 25(1): 261-274.

Zavody, A. M. and Mutlow, C. T., Llewellyn-Jones, D., T., 1995. "A radiative transfer model for sea surface temperature retrieval for the along-track scanning radiometer." *Journal of Geophysical Research Oceans* 100(C1): 937-952.

Table 1. Details of the manually classified scenes used as reference data to calculate cloud detection algorithm performance

Scene Number	Location	Region	Date	Latitude Bounds	Longitude Bounds	Abs. Orbit No.	AOD
(1)	Abracos Hill	Brazil, S. America	06/06/2004	-13.03 - -8.7	-64.57 - 59.28	11858	< 0.3
(2)	Abracos Hill	Brazil, S. America	26/05/2005	-12.92 - -8.75	-64.54 - -59.29	17369	< 0.3
(3)	Cart Site	Oklahoma, N. America	09/08/2004	34.45 - 38.86	-101.18 - -94.61	12776	0.3-0.4
(4)	Cart Site	Oklahoma, N. America	16/05/2005	34.43 - 38.84	-101.19 - -94.62	16784	< 0.3
(5)	Cart Site	Oklahoma, N. America	25/07/2005	34.42 - 38.84	-101.19 - -94.63	17786	< 0.3
(6)	Cart Site	Oklahoma, N. America	05/06/2006	34.42 - 38.82	-101.18 - -94.62	22295	> 0.4
(7)	Cart Site	Oklahoma, N. America	08/10/2007	34.57 - 38.81	-101.14 - -94.63	29309	< 0.3
(8)	Ouagadougou	Burkino Faso, Eq. Africa	28/11/2002	11.15 - 13.97	-2.43 - 2.57	03897	0.3-0.4
(9)	Ouagadougou	Burkino Faso, Eq. Africa	01/12/2002	11.22 - 14.38	-3.85 - 1.23	03941	0.3-0.4
(10)	Ouagadougou	Burkino Faso, Eq. Africa	02/01/2003	11.15 - 13.98	-2.43 - 2.57	04398	0.3-0.4
(11)	Ouagadougou	Burkino Faso, Eq. Africa	21/01/2003	11.1 - 14.27	-3.16 - 1.92	04670	0.3-0.4
(12)	Ouagadougou	Burkino Faso, Eq. Africa	15/07/2003	11.23 - 14.18	-3.12 - 1.9	07176	> 0.4
(13)	Mongu	Zambia, South Africa	15/01/2003	-16.95 - -12.62	18.57 - 23.95	04583	0.3-0.4
(14)	Mongu	Zambia, South Africa	09/07/2003	-16.92 - -12.76	18.58 - 23.91	07088	< 0.3
(15)	Mongu	Zambia, South Africa	17/09/2003	-16.9 - -12.74	18.57 - 23.91	08090	> 0.4
(16)	Tomsk	Siberia, Russia	03/06/2004	54.21 - 58.84	80.01 - 89.8	11810	0.3-0.4
(17)	Tomsk	Siberia, Russia	12/08/2004	54.16 - 58.78	79.99 - 89.76	12812	< 0.3
(18)	Tomsk	Siberia, Russia	23/06/2005	54.13 - 58.74	79.97 - 89.72	17321	> 0.4
(19)	Tomsk	Siberia, Russia	06/10/2005	54.16 - 58.78	80.0 - 89.77	18824	< 0.3
(20)	Tomsk	Siberia, Russia	21/09/2006	54.26 - 58.72	80.05 - 89.73	23834	< 0.3
(21)	Tomsk	Siberia, Russia	02/08/2007	54.15 - 58.76	79.99 - 89.75	28343	< 0.3

Table 2. Dominant biomes for each of the locations represented by the reference data.

Location	Dominant Biome
Abracos Hill	Broadleaved evergreen and/or semi-deciduous forest and mosaic croplands.
Cart Site	Grassland, mosaic vegetation, mosaic forest and rainfed croplands.
Ouagadougou	Bare areas and grassland.
Mongu	Grassland, broadleaved deciduous forest, shrubland and mosaic cropland.
Tomsk	Needleleaved deciduous or evergreen forest, mosaic cropland, sparse vegetation, mosaic vegetation and rainfed croplands.

Table 3. Number of clear and cloud pixels in all scenes and the breakdown into 'clear', 'low aerosol' and 'high aerosol' scenes

Scene Definition	Cloud Pixels	Clear Pixels	Total Pixels
All	692023	2264737	2957187
AOD < 0.3	378985	1168468	1547880
AOD $0.3 - 0.4$	118132	733028	851160
AOD > 0.4	194906	3633241	558147

Table 4. Performance Statistics for each of the six cloud detection algorithms with reference to the manual mask segregated according to aerosol optical depth. 'PP' is the percentage of perfect classification, 'HR' is the hit rate, 'FAR' is the false alarm rate and 'TSS' is the true skill score.

All	PP	HR	FAR	TSS
SADIST	83.86	86.09	16.82	69.27
UOL_2	90.56	84.17	7.49	76.69
UOL_3	91.71	83.80	5.87	77.93
Bayes Thermal	85.55	88.61	15.39	73.22
Bayes Vis	88.17	80.59	9.52	71.08
Bayes Joint	89.01	88.45	10.82	77.63
AOD < 0.3				
SADIST	84.42	89.85	17.34	72.52
UOL_2	90.73	85.22	7.48	77.74
UOL_3	92.39	82.81	4.51	78.31
Bayes Thermal	87.24	87.06	12.70	74.36
Bayes Vis	88.45	79.38	8.61	70.77
Bayes Joint	89.84	86.51	9.07	77.43
AOD $0.3 - 0.4$				
SADIST	86.29	59.16	9.33	49.83
UOL_2	93.68	60.79	1.02	59.77
UOL_3	94.18	66.22	1.31	64.91
Bayes Thermal	87.20	79.32	11.53	67.79
Bayes Vis	89.04	61.62	6.54	55.08
Bayes Joint	91.90	79.47	6.09	73.38
AOD > 0.4				
SADIST	78.58	95.10	30.28	64.82
UOL_2	85.33	96.30	20.55	75.75
UOL_3	86.08	96.36	19.44	76.92
Bayes Thermal	78.35	97.25	31.80	65.46
Bayes Vis	86.06	94.45	18.44	76.01
Bayes Joint	82.29	97.66	25.96	71.70

Table 5. LST impacts with reference to the manual mask segregated according to aerosol optical depth.

All	2 K Tolerance	Large Uncertainties	Over-flag	Under-flag
SADIST	85.69	0.71	10.86	2.74
UOL_2	91.01	0.36	5.48	3.15
UOL_3	92.02	0.39	4.37	3.22
Bayes Thermal	87.02	0.32	10.49	2.18
Bayes Vis	90.79	0.49	4.90	3.82
Bayes Joint	90.60	0.31	6.98	2.11
AOD < 0.3				
SADIST	86.59	0.48	11.03	1.91
UOL_2	91.36	0.34	5.29	3.02
UOL_3	92.89	0.40	3.16	3.55
Bayes Thermal	88.54	0.25	8.61	2.60
Bayes Vis	91.07	0.45	4.28	4.20
Bayes Joint	91.68	0.27	5.47	2.58
AOD $0.3 - 0.4$				
SADIST	88.38	0.93	5.55	5.14
UOL_2	94.17	0.34	0.60	4.89
UOL_3	94.60	0.32	0.94	4.14
Bayes Thermal	89.47	0.39	7.73	2.41
Bayes Vis	91.86	0.60	2.73	4.81
Bayes Joint	93.70	0.40	3.57	2.33
AOD > 0.4				
SADIST	79.12	1.02	18.50	1.37
UOL_2	85.22	0.47	13.45	0.87
UOL_3	85.68	0.47	12.95	0.90
Bayes Thermal	79.04	0.39	19.90	0.67
Bayes Vis	88.36	0.45	9.91	1.28
Bayes Joint	82.88	0.31	16.37	0.45

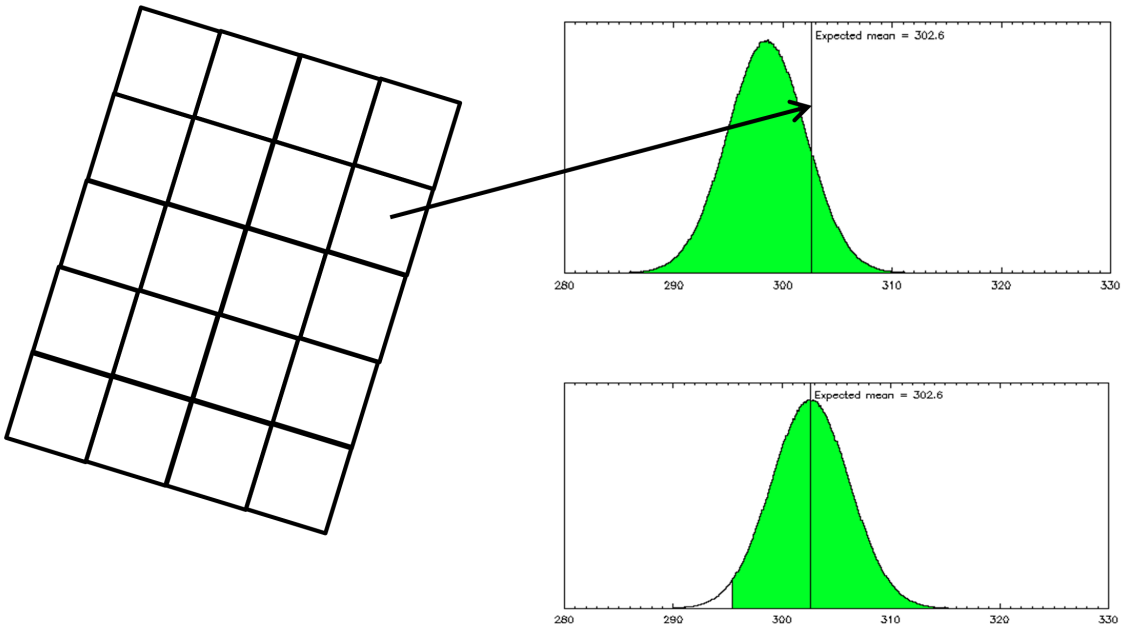

Figure 1. For each granule of an AATSR orbit (left) the expected 12 μm brightness temperature is simulated from coincident profiles. The PDF of observed 12 μm brightness temperatures for each biome-diurnal condition given the space and time position is also determined (top-right in green). This PDF is moved so that the mean equals the expected mean for the granule and the new PDF represents the expected clear-sky conditions (bottom-right in green).

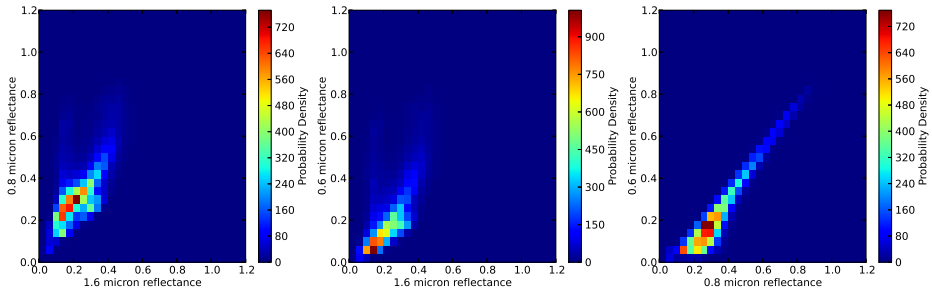

Figure 2. Proof of concept cloud PDFs for visible channel cloud detection. 2D slices of the 3D PDF are plotted to show the relationships between the 0.6, 0.8 and 1.6 μm channels.

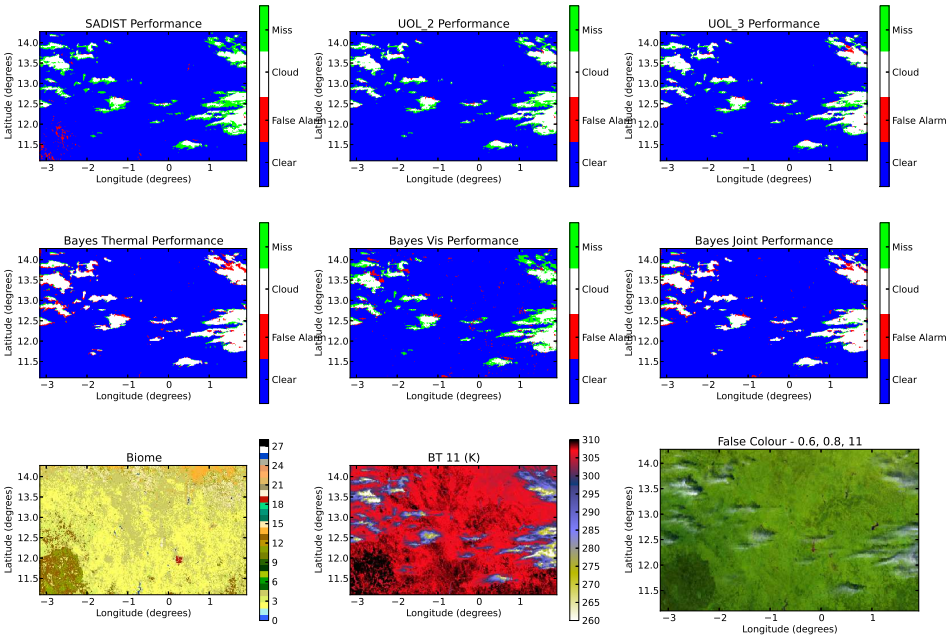

Figure 3. Cloud detection algorithm performance over Ouagadougou on 21st January 2003. From left to right, the top panels show the SADIST, UOL_2 and UOL_3 algorithms. The middle panels the Bayes Thermal, Bayes Vis and Bayes Joint algorithms, and the bottom panel the biome, 11 μm brightness temperature and a false colour images from the 0.6, 0.8 and 11 μm channels. The biomes are as follows: (0) Ocean, (1) Irrigated croplands, (2) Rainfed croplands, (3) Mosaic cropland, (4) Mosaic vegetation, (5) Closed to open broadleaved/deciduous forest, (6) Closed deciduous forest, (7) Open deciduous forest, (8) Closed needleleaved forest, (9) Open needleleaved forest, (10) Closed to open mixed forest, (11) Mosaic forest/shrubland, (12) Mosaic grassland, (13) Closed to open shrubland, (14) Closed to open grassland, (15) Sparse vegetation, (16) Regularly flooded forest (fresh water), (17) Regularly flooded forest (saline water), (19) Urban areas, (20) Mixed bare ground, (21) Entisols - Orhents, (22) Shifting sand, (23) Aridisols - Calcids, (24) Aridisols - Cambids, (25) Gelisols - Orthels, (26) Waterbodies, (27) Permanent snow/ice, (28) No data.

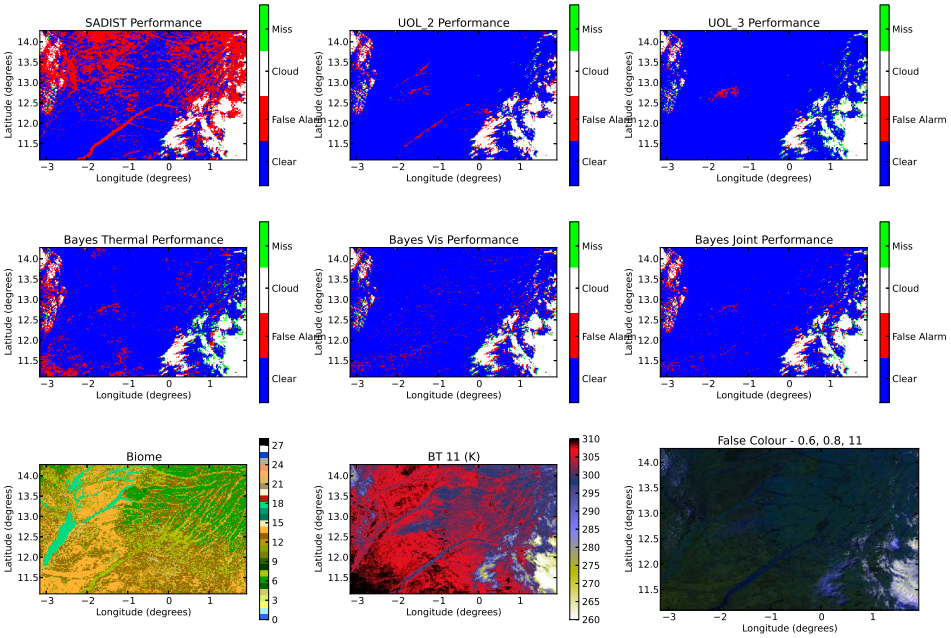

Figure 4. Cloud detection algorithm performance over Mongu, Zambia on 19th July 2003. From left to right, the top panels show the SADIST, UOL_2 and UOL_3 algorithms. The middle panels the Bayes Thermal, Bayes Vis and Bayes Joint algorithms, and the bottom panel the biome, 11 μm brightness temperature and a false colour images from the 0.6, 0.8 and 11 μm channels. Please refer to Figure 3 caption for biome definition.

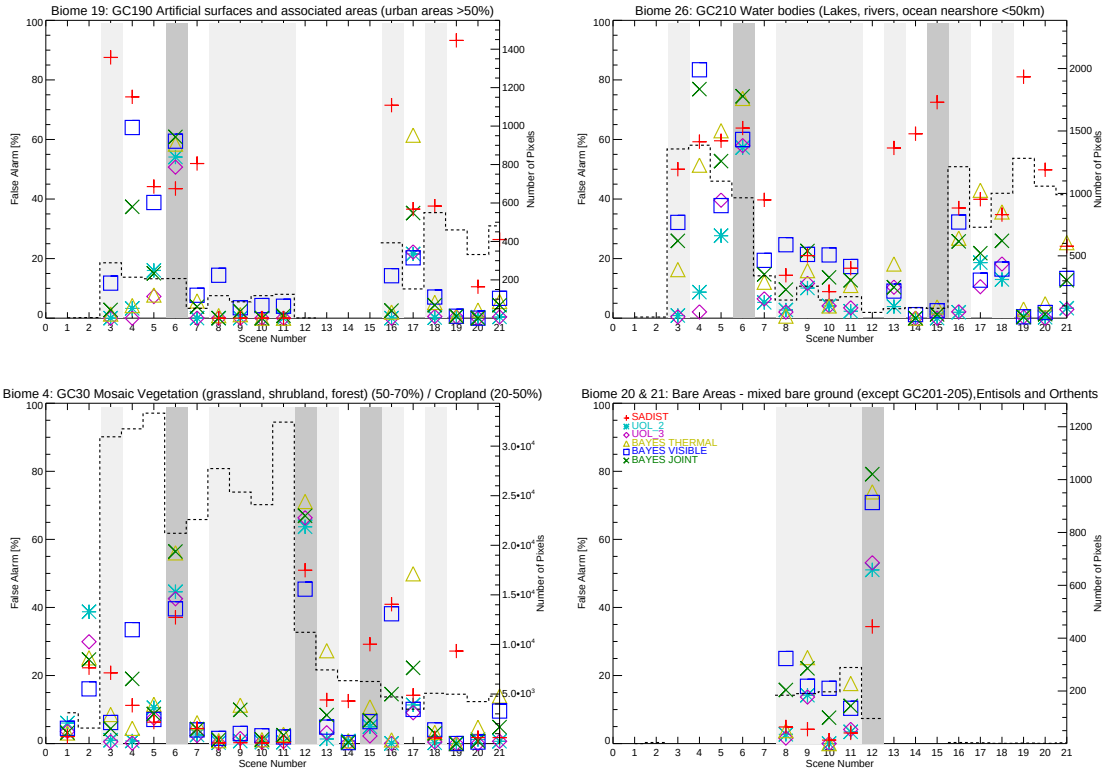

Figure 5. AATSR cloud detection algorithm performance over urban areas, waterbodies, mosaic vegetation and bare soils biomes. Scene numbers refer to the listing given in Table 1. Light grey shading indicates low aerosol concentrations (AOD 0.3-0.4) and dark grey shading high aerosol concentrations (AOD > 0.4).