Q. J. R. Meteorol. Soc. (2005), **131**, pp. 2735–2755 doi: 10.1256/qj.05.15

Probabilistic physically based cloud screening of satellite infrared imagery for operational sea surface temperature retrieval

By C. J. MERCHANT[1]*, A. R. HARRIS[2], E. MATURI[3] and S. MACCALLUM[1]

[1]*School of GeoSciences, University of Edinburgh, UK*
[2]*NOAA Co-operative Institute for Climate Studies, University of Maryland, USA*
[3]*National Oceanic and Atmospheric Administration, Camp Springs, USA*

(Received 3 February 2005; revised 23 June 2005)

SUMMARY

We propose and demonstrate a fully probabilistic (Bayesian) approach to the detection of cloudy pixels in thermal infrared (TIR) imagery observed from satellite over oceans. Using this approach, we show how to exploit the prior information and the fast forward modelling capability that are typically available in the operational context to obtain improved cloud detection. The probability of clear sky for each pixel is estimated by applying Bayes' theorem, and we describe how to apply Bayes' theorem to this problem in general terms. Joint probability density functions (PDFs) of the observations in the TIR channels are needed; the PDFs for clear conditions are calculable from forward modelling and those for cloudy conditions have been obtained empirically. Using analysis fields from numerical weather prediction as prior information, we apply the approach to imagery representative of imagers on polar-orbiting platforms. In comparison with the established cloud-screening scheme, the new technique decreases both the rate of failure to detect cloud contamination and the false-alarm rate by one quarter. The rate of occurrence of cloud-screening-related errors of >1 K in area-averaged SSTs is reduced by 83%.

KEYWORDS: ATSR AVHRR Cloud detection

1. INTRODUCTION

Sea surface temperature (SST, see appendix A for all acronyms) estimates are routinely produced from space-based sensors and are used to derive SST fields for use in NWP (as a lower boundary to the atmosphere) and in climate and oceanographic applications. Appropriate sensors operating at TIR wavelengths in adequately transmissive spectral windows (Deschamps and Phulpin 1980) can give relatively precise SST estimates (0.5 K or better) with high spatial resolution (few km), even from geostationary orbit (e.g. McClain 1989; Brisson *et al.* 2002).

Derivation of SST from TIR imagery consists of two steps: cloud screening (also known as cloud detection or cloud masking) and SST retrieval for those pixels deemed not to contain cloud. This article describes an innovation in cloud detection intended to achieve better results more flexibly, in the context of operational NWP.

The 'traditional' approach to cloud screening (e.g. Saunders 1986; Saunders and Kriebel 1988) involves a sequence of threshold tests in which spectral and/or spatial properties of the imagery are compared with ranges deemed to be characteristic of 'clear-sky' pixels; pixels whose properties are outside these limits are flagged as affected by cloud. This has been a successful approach during the past two decades of operational SST production, yet it can be criticized on several counts.

Firstly, threshold tests embody knowledge about indicators of cloudiness in a way that is not transparent. Thresholds are typically defined by an experienced interpreter of imagery in relation to sets of observations and/or simulations of satellite-observed radiance. Setting a threshold involves an expert judgement that may not be objective (in that a different expert would propose a different value), and that may be inconsistent

* Corresponding author: Crew Building, The King's Buildings, Mayfield Road, Edinburgh EH9 3JN, UK.
e-mail: c.merchant@ed.ac.uk

© Royal Meteorological Society, 2005.

between different threshold tests within the cloud-detection scheme. Related problems are that threshold schemes are often intricate and highly specific to a given sensor, and are therefore difficult to update and improve (especially if the original expertise is lost). If not updated, the performance of a static cloud-screening system will change over time in the face of changes in the observing system (e.g. drift in overpass time) or the environment (e.g. temperature trends). In this way, an artefact of the cloud screening is introduced into the data record, which is clearly undesirable in the context of climate monitoring, for example.

A second set of criticisms centres on the fact that the thresholds chosen are not logically connected to the size of effect (i.e. the bias in SST) resulting from non-detection and false detection. Different users may be tolerant of different levels of bias (or indeed scatter) in SST products; some may prefer to increase the coverage of SST at the expense of allowing some cloud-related bias. The trade-off implicit in a given SST product is fixed by the thresholds used, probably without ever having been quantified.

Thirdly, in the context of operational processing of imagery, not all the available prior information is used. Pre-determined threshold tests implicitly assume some sort of climatology as prior information, but forecast fields will generally have greater information content.

Recent work has brought in semi-probabilistic elements that go some way to addressing these criticisms. For example, a hierarchy of qualitative descriptors of SST confidence/quality is attached to products of the Ocean and Sea Ice Satellite Application Facility (Brisson *et al.* 2002). Such a scheme is thus more informative, but remains somewhat arbitrary. Forecast fields (of TCWV) are also used in this context to influence the thresholds (Le Gléau and Derrien 2000). Other approaches to setting thresholds (or 'decision boundaries') have been described by Murtagh *et al.* (2003) and by Di Vittorio and Emery (2002), among others.

In this paper, we attempt to lay down a fully probabilistic framework for cloud screening of TIR imagery in a rigorous way. In one sense, the work presented is hardly novel, being merely Bayes' theorem applied to cloud screening for SST. Furthermore, different Bayesian techniques have been applied to cloud screening before (e.g. Uddstrom *et al.* 1999; Murtagh *et al.* 2003). Nonetheless, we do show for the first time how to calculate a physically based estimate of the probability of each pixel in an image being clear sky. We begin in section 2 by outlining the theoretical basis and describing (section 3) how this relates to a typical case. We apply the method to example imagery in section 4. The discussion (section 5), with which this paper ends, argues that our approach not only promises better results, but has a number of other practical advantages for operational NWP centres.

2. Bayes' theorem applied to cloud screening

The cloud-detection problem is to deduce the likelihood of a pixel in an image being free of cloud, given the values of radiance of various TIR channels for the pixel (and perhaps other pixels in the image).

Prior knowledge is also available. From the time and geographical location of the observations, we are at least able to specify climatological values of surface temperature and atmospheric state. A 'stand-alone' cloud-detection scheme based on the theory developed below can make optimal use of climatology. In the context of NWP, however, we may assume we have a recent SST analysis and forecast fields of the atmosphere, which represent more informative prior knowledge.

Bayes' theorem for the probability of clear sky, given the observations and the background knowledge, amounts to (English *et al.* 1999, Eq. 4):

$$P(c \mid \mathbf{y}^{\mathrm{o}}, \mathbf{x}^{\mathrm{b}}) = \frac{P(\mathbf{y}^{\mathrm{o}} \mid \mathbf{x}^{\mathrm{b}}, c)P(\mathbf{x}^{\mathrm{b}} \mid c)P(c)}{P(\mathbf{y}^{\mathrm{o}} \mid \mathbf{x}^{\mathrm{b}})P(\mathbf{x}^{\mathrm{b}})}. \tag{1}$$

In Eq. (1), each P represents a probability or PDF as specified in its argument; c is the state of clear sky over ocean, $\mathbf{y}$ is the observation vector, $\mathbf{x}$ is the state vector, superscript o indicates 'observed', and superscript b indicates 'background' (i.e. prior knowledge). The definition of the elements of the observation vector and background state will vary according to the sensor and forward model respectively; a specific example follows in section 3.

For TIR imagers used in meteorology, clear-sky probability varies on length-scales from the pixel size ($\sim$1 km) and larger. This is much finer than the length-scales of variation in the atmospheric terms (other than cloudiness) in the background state ($\sim$100 km). Thus, to a good approximation on pixel-to-pixel scales, the background state can be assumed independent of clear-sky probability, and therefore $P(\mathbf{x}^{\mathrm{b}} \mid c) = P(\mathbf{x}^{\mathrm{b}})$, which further simplifies Eq. (1).

The term $P(\mathbf{y}^{\mathrm{o}} \mid \mathbf{x}^{\mathrm{b}})$ describes the PDF of the observations given the background state. We may decompose this PDF into the contribution from clear and cloudy conditions.

$$P(\mathbf{y}^{\mathrm{o}} \mid \mathbf{x}^{\mathrm{b}}) = P(\overline{c})P(\mathbf{y}^{\mathrm{o}} \mid \mathbf{x}^{\mathrm{b}}, \overline{c}) + P(c)P(\mathbf{y}^{\mathrm{o}} \mid \mathbf{x}^{\mathrm{b}}, c), \tag{2}$$

where an overbar signifies 'not'. This decomposition of $P(\mathbf{y}^{\mathrm{o}} \mid \mathbf{x}^{\mathrm{b}})$ into two terms will allow us to develop the approach of English *et al.* (1999) to the point that we are able to derive an objective estimate of clear-sky probability. Substitution of Eq. (2) into Eq. (1) and rearrangement gives the new equation we propose to estimate clear-sky probability:

$$P(c \mid \mathbf{y}^{\mathrm{o}}, \mathbf{x}^{\mathrm{b}}) = \left\{1 + \frac{P(\overline{c})P(\mathbf{y}^{\mathrm{o}} \mid \mathbf{x}^{\mathrm{b}}, \overline{c})}{P(c)P(\mathbf{y}^{\mathrm{o}} \mid \mathbf{x}^{\mathrm{b}}, c)}\right\}^{-1}. \tag{3}$$

To use Eq. (3) to estimate the clear-sky probability, we need to be able to evaluate or estimate the probabilities on the right-hand side of the equation. We need, first, a prior estimate of the probability of clear sky, $P(c)$; this could come from a cloud-cover climatology or from the cloud fraction in an NWP forecast. Of course, $P(\overline{c}) = 1 - P(c)$. The probability of the observation given the background state and clear-sky conditions, $P(\mathbf{y}^{\mathrm{o}} \mid \mathbf{x}^{\mathrm{b}}, c)$, can be estimated, as we show below, using a forward model for the observations together with knowledge of relevant error covariances. In principle, $P(\mathbf{y}^{\mathrm{o}} \mid \mathbf{x}^{\mathrm{b}}, \overline{c})$ could also be found by forward calculation. However, compared to clear-sky radiances, the variability of cloudy radiance observations is much greater and their distribution is dominated not by the parameters present in the background state, but by the details of the vertical cloud distribution. In the light of this, we describe below an approach that effectively uses a climatology of cloudy radiances to evaluate this term (although more sophisticated approaches can be envisaged in future).

3. Formulation for AVHRR-like sensors

To make the above outline more concrete, this section describes a formulation of this screening technique for SST sensors such as the AVHRR or ATSR (Edwards *et al.* 1990). We focus on the case of viewing a night-time scene, where typically three thermal-window channels (at nominally 3.7, 11 and 12 μm) are available.

(*a*) *The observation vector*

The brightness temperatures observed by the three thermal channels constitute three elements of the observation vector for a given pixel. We refer to this subset of the observation vector as $\mathbf{y}_\mathrm{s}$, where the s indicates 'spectral elements', because they are samples of the BT spectrum. Thus,

$$\mathbf{y}_\mathrm{s} = \begin{bmatrix} T_{3.7} \\ T_{11} \\ T_{12} \end{bmatrix},$$

where T_{ch} is the brightness temperature in channel ch.

In addition, we may use texture measures as elements in the observation vector. Since cloudy regions are usually more spatially variable than clear-sky regions, texture measures may be informative. Texture measures involve combinations of several pixels, which have the disadvantage that the resulting probability estimates are effectively at a lower spatial resolution than the imagery. However, the benefit of the extra information they embody may outweigh this disadvantage, as here (see below). We use a measure based on the r.m.s. deviations of the 3.7 and 11 μm BT, over a $\sqrt{m} \times \sqrt{m}$ pixel box centred on each image pixel. We use this to make an unbiased localized estimate of the population standard deviation (which we refer to as 'local standard deviation', LSD):

$$\mathbf{y}_\mathrm{t} = \begin{bmatrix} \sqrt{\dfrac{1}{m-1}\sum_{\mathrm{box}}(T_{3.7} - \overline{T_{3.7}})^2} \\ \sqrt{\dfrac{1}{m-1}\sum_{\mathrm{box}}(T_{11} - \overline{T_{11}})^2} \end{bmatrix}, \tag{4}$$

where $\mathbf{y}_\mathrm{t}$ contains the 'textural elements' of the observation vector. The greater non-linearity of the Planck function at 3.7 μm makes the LSD from this channel differently sensitive to sub-pixel and other forms of cloudiness than that from the 11 μm channel, which is why we use LSDs for both these channels. The 12 μm is similar to the 11 μm channel in this regard, and we omit it from the textural elements as not providing significant independent information.

Options other than LSD are possible for textural elements (Uddstrom *et al.* 1999). We have chosen LSD since it may readily be related to the physical noise characteristics of the sensor (see subsection 3(c)), and, with $m = 9$, is a relatively precise estimator while retaining a fairly high resolution.

(*b*) *The background state and forward model*

The background state, $\mathbf{x}^\mathrm{b}$, describes our prior knowledge of the state of the surface and atmosphere, i.e. it comprises surface temperature and atmospheric temperature and humidity profiles sufficient to allow estimation of clear-sky $\mathbf{y}_\mathrm{s}^\mathrm{b}$ with some known, reasonable error covariance. Operationally, this would be provided by previous SST analyses and by forecast fields. Here, we use ERA-40 fields from ECMWF (2000–05) at 1° spatial resolution, taking the fields for the nearest analysis time to the time of observation (i.e. within ±3 hours). We use the fast forward model RTTOV-7 (Saunders *et al.* 1999) to transform the ERA fields into clear-sky BTs at each grid point, and these are bilinearly interpolated in space to image pixel locations. The forward model for the texture elements is based on the sensor's radiometric noise levels (see the following subsection).

(c) *The clear-sky probability density function*

For clear skies, LSD may be assumed to be spatially uncorrelated with the central pixel's BT, so

$$P(\mathbf{y}^{\mathrm{o}} \mid \mathbf{x}^{\mathrm{b}}, c) = P(\mathbf{y}_{\mathrm{s}}^{\mathrm{o}} \mid \mathbf{x}^{\mathrm{b}}, c) P(\mathbf{y}_{\mathrm{t}} \mid \mathbf{x}^{\mathrm{b}}, c). \tag{5}$$

The spectral factor, $P(\mathbf{y}_{\mathrm{s}}^{\mathrm{o}} \mid \mathbf{x}^{\mathrm{b}}, c)$, can be found using a forward model, $\mathbf{F}$, to evaluate $\mathbf{y}^{\mathrm{b}} = \mathbf{F}(\mathbf{x}^{\mathrm{b}})$ and, assuming Gaussian background and observation errors (Rodgers 1976),

$$P(\mathbf{y}_{\mathrm{s}}^{\mathrm{o}} \mid \mathbf{x}^{\mathrm{b}}, c) = \frac{\exp\{-\frac{1}{2}(\mathbf{y}^{\mathrm{o}} - \mathbf{y}^{\mathrm{b}})^{\mathrm{T}}(\mathbf{H}'^{\mathrm{T}}\mathbf{B}\mathbf{H}' + \mathbf{R})^{-1}(\mathbf{y}^{\mathrm{o}} - \mathbf{y}^{\mathrm{b}})\}}{(2\pi)^{n/2}|\mathbf{H}'^{\mathrm{T}}\mathbf{B}\mathbf{H}' + \mathbf{R}|^{1/2}}, \tag{6}$$

where $\mathbf{H}' = \nabla_{\mathbf{x}}\mathbf{y}^{\mathrm{b}}$, $\mathbf{B}$ is the error covariance of the background state, $\mathbf{R}$ is the combined error covariance of the forward model and observations and n is the number of spectral elements in the observation vector.

To evaluate this equation, we need to know several parameters. The matrix $\mathbf{R}$ is the covariance matrix associated with the NEΔT (i.e. the radiometric noise) for each channel and with the forward modelling errors. Here, we assume radiometric noise and forward model error are uncorrelated between channels. The matrix $\mathbf{B}$ is, in principle, the covariance matrix of all the prior parameters obtained from the ERA data. For simplicity in this case-study, however, we have included in $\mathbf{B}$ only the two dominant terms in determining the BT of a thermal window channel: the surface temperature and the TCWV of the atmosphere. We estimate the background error variances in these parameters as $(1.2\ \mathrm{K})^2$ and $(15\%)^2$ respectively, when used as estimates at 1 km scales, and assume that these errors are independent. The gradients with respect to SST and TCWV of the background observations, which appear in $\mathbf{H}'$, we obtain by running the forward model with a perturbed surface temperature (original SST + 1 K) and scaled water vapour profile (original $\times$ 0.85). (Full calculation of $\mathbf{H}^{\mathrm{T}}\mathbf{B}\mathbf{H}' + \mathbf{R}$ using the tangent-linear functionality of RTTOV-7 will be more optimal, and will require forming the full covariance matrix of RTTOV-7 parameters from a covariance matrix for the ERA fields. Whether this extra computation is significantly beneficial is to be assessed in future work.)

Turning to the textural factor in Eq. (5), for a clear-sky scene uniform over a 3-pixel length-scale, the only contribution to LSD is from radiometric noise, ε^{o}, in the channels (plus any uncorrected striping or other artefacts). Away from oceanic thermal features, this is the situation for clear-sky observations at AVHRR-like resolution ($\sim$1 km). $P(\mathbf{y}_{\mathrm{t}} \mid \mathbf{x}^{\mathrm{b}}, c)$ then adequately resembles a Gaussian random variable that, for the channel ch, is distributed as $\varepsilon_{ch}^{\mathrm{o}} \pm \sqrt{2/(m-1)}\,\varepsilon_{ch}^{\mathrm{o}}$. For $m = 9$, the clear-sky LSD distribution therefore has a mean equal to the NEΔT and a standard deviation around this mean of half of the NEΔT. Under this approximate distribution, $\sim$99.9% of clear-sky LSDs will have values less than 2.5$\times$NEΔT.

Not all clear-sky areas are uniform over the 3-pixel length-scale, although it is a good approximation for most areas. Where there is a significant thermal gradient, $\mathrm{d}SST/\mathrm{d}x$, in the SST, this increases LSD. A marked front could consist of a thermal gradient $\sim$0.5 K km^{-1} maintained over a few kilometres. For a 1 km resolution image, this translates into an LSD

$$\sim\sqrt{\frac{3}{4}}\frac{\partial T_{ch}}{\partial SST}\frac{\mathrm{d}SST}{\mathrm{d}x} \sim 0.2 \text{ to } 0.5\ \mathrm{K}$$

(depending on the atmospheric attenuation). A typical assumption for AVHRR is that NEΔT $\sim$0.1 K, and so strong oceanic fronts can lead to clear-sky LSDs that are greater

than 2.5 × NEΔT. The probability that these are clear-sky is therefore underestimated when using the above distribution.

(Fronts are also far below the spatial resolution of the background field which is typically at ∼1°. This is why we use the relatively large value of $(1.2\ \mathrm{K})^2$ as the variance of the background SST. On the >100 km scale, the error in global SST analyses is more typically ∼0.5 K (Reynolds *et al.* 2002), but ocean thermal structure at smaller scales increases the error in using such a field as a prior for a 1 km pixel or a point observation. More spatially resolved SST analyses—such as those planned at ∼0.1° spatial resolution within the context of the Global Ocean Data Assimilation Experiment High Resolution SST project*—will allow the regions of potential ocean fronts in an image to be more precisely located a priori, and for the error assumed in background SST to be reduced elsewhere.)

(*d*) *The cloudy-sky probability density function*

Determining $P(\mathbf{y}^{\mathrm{o}} \mid \mathbf{x}^{\mathrm{b}}, c)$ is a more difficult task. This distribution is not likely to be Gaussian, and the impacts on BTs of the many possible vertical cloud distributions are large and variable.

We begin by making the same decomposition as was assumed for clear skies, that $P(\mathbf{y}^{\mathrm{o}} \mid \mathbf{x}^{\mathrm{b}}, \overline{c}) = P(\mathbf{y}^{\mathrm{o}}_{\mathrm{s}} \mid \mathbf{x}^{\mathrm{b}}, \overline{c}) P(\mathbf{y}_{\mathrm{t}} \mid \mathbf{x}^{\mathrm{b}}, \overline{c})$. This assumes that the texture of clouds is independent of their cloud-top temperature. This is reasonable since both stratiform and horizontally inhomogeneous clouds form at all tropospheric altitudes.

We consider first the joint PDF of the BTs. Figure 1 shows two dimensions of a three-dimensional distribution $P(\mathbf{y}^{\mathrm{o}}_{\mathrm{s}} \mid \overline{c})$ developed empirically from ∼750 globally distributed ocean scenes of ATSR-2. We use ATSR-2 because this series of sensors has very low NEΔT (∼0.05 K). To generate the distribution, the number of cloudy pixels with brightness temperatures in $(0.5\ \mathrm{K})^3$ bins was counted. Both forward (∼52° to 56°) and nadir (0° to 22°) views are combined in this distribution. Day scenes are not represented; channels around 3.7 μm are affected by reflected and scattered solar irradiance and have a different distribution. The standard ATSR-2 cloud mask was used to identify cloudy pixels for this purpose. The total number of cloudy pixels counted was $\sim 3 \times 10^8$.

$P(\mathbf{y}^{\mathrm{o}}_{\mathrm{s}} \mid \overline{c})$ is not equal to the distribution required by Eq. (5), namely $P(\mathbf{y}^{\mathrm{o}}_{\mathrm{s}} \mid \mathbf{x}^{\mathrm{b}}, \overline{c})$. The former distribution is not conditional on the background state. Clearly, the distribution of observations for cloudy pixels should depend on the background state; for example, cloud-top temperatures are usually lower than underlying SST, implying the distribution of cloudy BTs depends on the SST. An ongoing research effort aims to develop a method for making this PDF properly conditional. In the meantime, we use the approximation $P(\mathbf{y}^{\mathrm{o}}_{\mathrm{s}} \mid \mathbf{x}^{\mathrm{b}}, \overline{c}) \approx P(\mathbf{y}^{\mathrm{o}}_{\mathrm{s}} \mid \overline{c})$. With unbiased $\mathbf{x}^{\mathrm{b}}$ and realistic error covariance of $\mathbf{y}^{\mathrm{b}}$, this still gives reasonable results for clear-sky probability because the clear-sky PDF is much more localized (in BT space) than the cloudy-sky distribution.

We similarly developed a distribution for $P(\mathbf{y}_{\mathrm{t}} \mid \overline{c})$ empirically (Fig. 2). This distribution of LSD arises from real variability in cloudy scenes at ∼1 to 2 km spatial resolution convolved with ATSR-2 noise. NEΔTs for ATSR-2 channels are relatively low (∼0.05 K), so this distribution could be used for sensors with higher levels of noise by convolving it with appropriate synthetic noise. The distribution is dependent on the spatial scale, and therefore may not be valid for sensors whose spatial resolution is different (e.g. ∼5 km imagery from geostationary orbit).

* http://www.ghrsst-pp.org

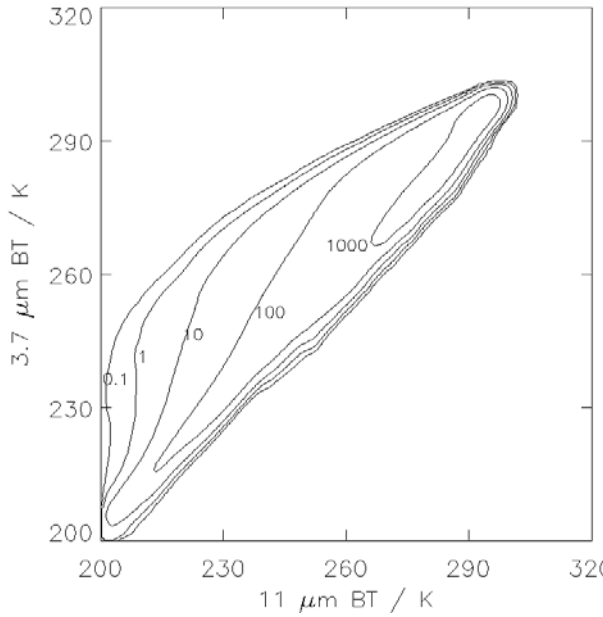

Figure 1. Joint probability density of 3.7 versus 11 μm brightness temperature (BT, K) for cloudy pixels, as estimated from ATSR-2 imagery screened by the standard mask. Contours are of probability per K^2 at intervals of powers of 10, labelled with values $\times 10^6$. The integral of the distribution is unity.

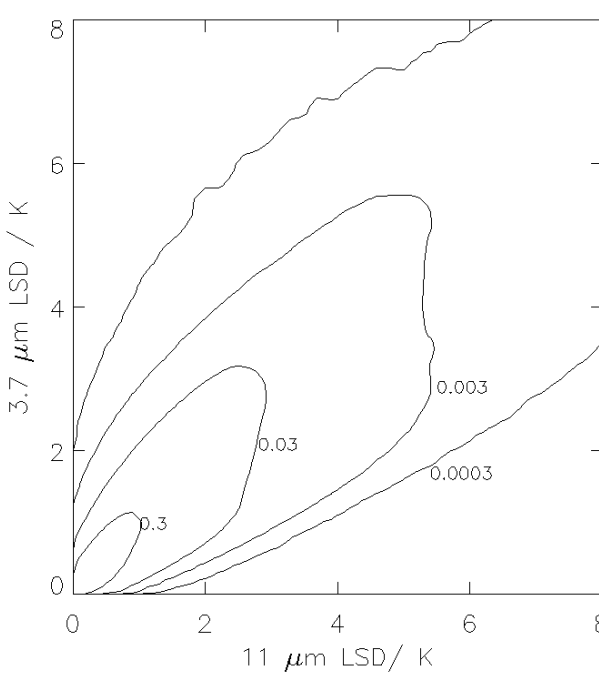

Figure 2. Empirical joint probability distribution for local standard deviation (LSD) of 3.7 and 11 μm brightness temperatures for cloudy pixels, estimated from the imagery of ATSR-2 (with low noise and ~1 km pixel resolution). Contour units are K^{-2}.

(*e*) *Prior probability of clear sky*

The prior probability of clear sky, $P(c)$, could be derived from climatological cloud cover or NWP fields. Here we use the mean ERA total cloud cover fraction over a 5° latitude–longitude box; it is not appropriate to use the values at 1° grid resolution because the time matching of analysis and satellite observation is ±3 hours in this experiment.

(*f*) *Evaluation of clear-sky probability*

At this point, all the PDFs and probabilities in Eq. (3) have been defined as described in the previous subsections. The clear-sky probability over a scene is now found by evaluating Eq. (3) for each pixel in turn.

4. Application to ATSR-2 imagery

As an example application, we use ATSR-2 images (independent of those used in section 3). A threshold-based cloud product is available with this imagery. This cloud product has been described by Zavody *et al.* (2000), and the performance of ATSR SSTs in validation indicates that the cloud product is generally effective (e.g. Merchant and Harris 1999; Parkes *et al.* 2000; Mathur *et al.* 2002). However, there are known problems with the cloud product (e.g. Simpson *et al.* 1998; Marsham 2003) which include examples of both over- and under-masking. For the sample of images discussed in this section, two of us (Merchant and Harris) have screened each image for clouds 'by eye' using thresholds defined by 'expert' judgement for each individual scene. In screening 'by eye' there is unavoidably a subjective element whose magnitude is indicated by the level of disagreement between the two 'experts'. This is discussed at greater length in subsection 4(b) below. We refer to the 'expert' cloud masks as 'tailored', and to the cloud product that accompanies the imagery as 'standard'.

(*a*) *Example image; Kuroshio Extension region*

We first of all present an example image in detail. Figure 3(a) shows an ATSR-2 nadir image at 3.7 μm observed over the Kuroshio Extension region at 1219 UTC on 19 August 1995. This image is 500 × 500 km, and the longitude–latitude coordinates are shown at each corner.

The sea surface appears warm with a strong oceanic front in the north and east (e.g. at the locations labelled A and B). Other smaller surface thermal features are also evident, such as at C. Clouds appear cold and there are regions of scattered cloud, some only just apparent (e.g. at D), as well as more extensive cloudy regions (e.g. at E). This image has been chosen as an example because its interpretation 'by eye' is relatively clear-cut, while it presents significant challenges to automated cloud screening, as discussed below.

Figure 3(b) shows the 11 μm channel brightness temperature at full resolution with, superimposed in black, the tailored cloud mask by Merchant. The image is cooler than at 3.7 μm because of greater atmospheric attenuation in this channel. (The northern portion of this image also shows evidence of clear-sky lower-tropospheric waves that are more apparent here for the same reason.) The tailored cloud mask successfully screens the cloud apparent in Fig. 3(a), with minimal masking of ocean features. Some misclassifications are evident on careful scrutiny of expanded colour-enhanced images, however. The agreement between the tailored mask of Merchant and the independent replication by Harris is 98.7%. The standard cloud mask

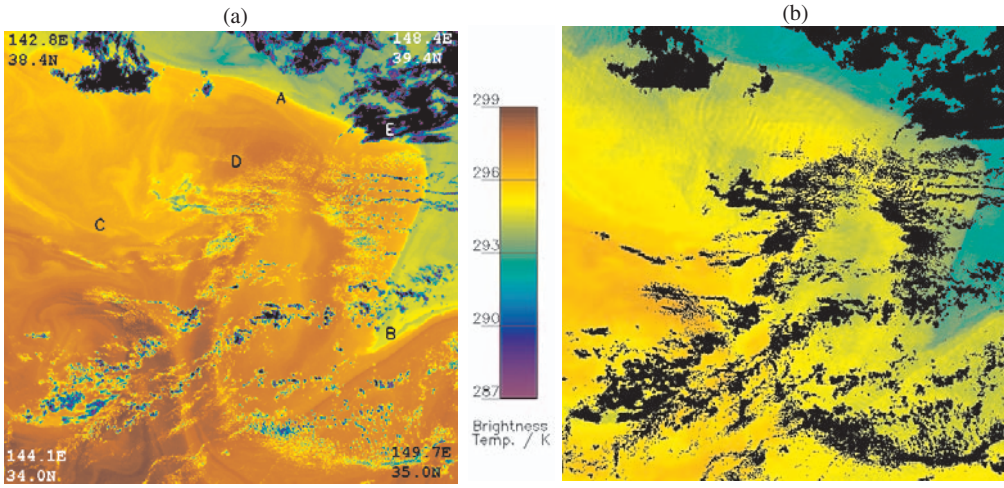

Figure 3. ATSR-2 imagery at 1 km resolution: (a) 3.7 μm brightness temperature (see text for meaning of labels A to E), and (b) 11 μm brightness temperature with tailored cloud mask superimposed in black.

(shown later in this subsection) agrees with the clear versus cloudy classification of Merchant's tailored cloud mask for 85% of pixels; the standard cloud mask over-masks some clear ocean on the cold side of the front and some frontal water, and categorizes as clear several areas of scattered cloud over the warmer part of the ocean.

An important contribution to the clear-sky probability estimate is the magnitude of the difference, $\mathbf{y}_s^o - \mathbf{y}_s^b$, i.e. the discrepancy between the forward-modelled BTs and those actually observed. This difference appears in the exponent of Eq. (6), and its magnitude squared in comparison with the covariance of the forward-modelled BTs determines the probability density of the observations. The magnitude, $|\mathbf{y}_s^o - \mathbf{y}_s^b|$, is plotted as Fig. 4. The most obvious features are pixels filled with cold cloud, which have large magnitudes of this difference and which appear grey in the figure. Two clear-sky areas labelled F and G have magnitudes of difference of order 2 K. These relatively large magnitudes are caused by significant differences between the background SST and the SST observed. The background SST is a smooth field that does not have any detailed frontal structure; it is too cold at F, and too warm at G. As commented above, this means that the assumed error in the background SST has to be large enough to accommodate frontal features that are resolved in the image; the scene-average SST retrieved from this image (see later) is 0.4 K higher than the background SST, while the standard deviation of that difference is 1.45 K (i.e. comparable to our assumed error in background SST of 1.2 K).

Around H, the magnitude of the $\mathbf{y}_s^o - \mathbf{y}_s^b$ difference is relatively small ($\sim$1 K) despite a too-low background SST, because the relatively small radiative impact of scattered (probably sub-pixel) clouds happens to offset the effect of the too-high background SST. This is a situation where the texture elements in the observation vector are needed to provide discrimination.

The LSD of the 3.7 μm brightness temperature using $m = 9$ is presented in Fig. 5. This texture measure is sensitive both to clouds, particularly cloud edges and sub-pixel clouds, and the oceanic thermal features. The maximum LSD is 7.7 K, but Fig. 5 is scaled between 0 and 1 K so that the LSD at the front (0.2 to 0.6 K) and of scattered cloud (0.3 K and greater) is contrasted clearly with the LSD in clear-sky areas away from

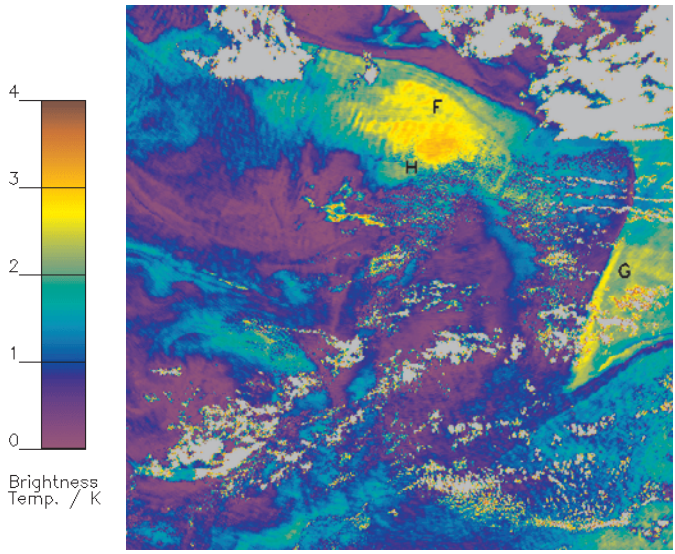

Figure 4. Magnitude of difference between spectral observation vector and corresponding clear-sky (forward-modelled) background vector, for the same scene as Fig. 3. The maximum value is 43 K (for a cold cloud-filled pixel), but in this image all values >4 K appear grey. (See text for meaning of labels F to H.)

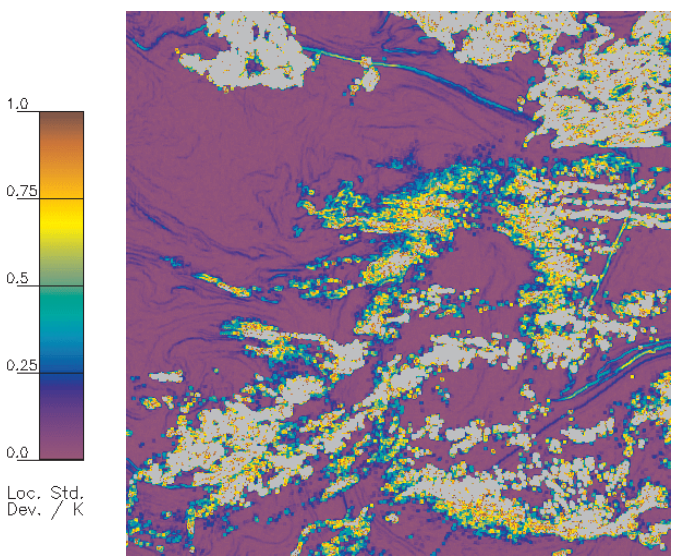

Figure 5. Local standard deviation of ATSR-2 3.7 μm channel image. Values >1 K appear grey.

fronts and clouds (~0.05 K). 99.7% of clear pixels that are not adjacent to clouds have LSD less than $2.5 \times \mathrm{NE}\Delta\mathrm{T} = 0.1$ K, which is a good agreement with the prediction of the 'forward model' for clear-sky LSD (section 3(c)) based on radiometric noise. Centre pixels that are clear sky but are adjacent to cloudy pixels have larger LSD, because cloudy pixels are present in the 3×3 box. The median LSD is 0.35 K for these

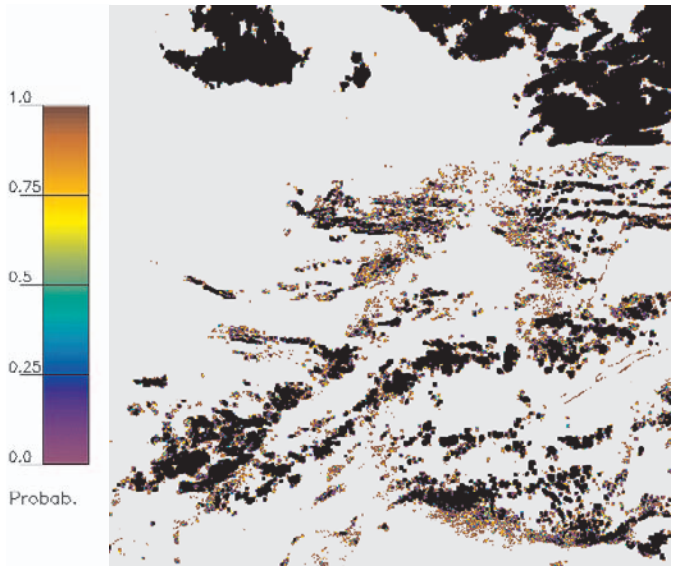

Figure 6. Clear-sky probability, estimated using the Bayesian approach, for imagery shown in Fig. 3. Values <0.4% appear black, and values >99.6% appear grey.

pixels; this is well outside the distribution expected from the 'forward model' for clear-sky LSD, and such pixels will tend to have underestimated clear-sky probability. In this sense, using the texture measure reduces the effective spatial resolution of the clear-sky probability estimate. The LSD of the 11 μm brightness temperature (not shown) is less sensitive to the front and to sub-pixel clouds and cloud edges in this scene.

The result of the Bayesian calculation of clear-sky probability is shown in Fig. 6. Most pixels are found to be very likely or very unlikely to be clear sky; only 7% of pixels yield estimated clear-sky probabilities between 0.01 and 0.99, with 19% less than 0.01 and 74% greater than 0.99. There is a good correspondence between masked pixels in Fig. 3(b) and low-probability pixels in Fig. 5. The mean clear-sky probability for pixels which are clear in the tailored mask is 0.97. Pixels flagged cloudy in the tailored mask have a mean clear-sky probability of 0.34.

The field of clear-sky probability can be transformed into a categorical clear–cloudy mask (like the tailored and standard cloud masks) by applying a threshold to the clear-sky probability estimate. This mask can then be compared to the tailored mask using measures familiar from forecast verification.

From Table 1, we then have:

- PP = Proportion of perfect classifications = $(h + z)/\tau$;
- HR = Hit rate = $h/(h + u)$;
- FAR = False alarm rate = $f/(f + z)$;
- TSS = True skill score = $h/(h + u) - f/(f + z)$.

Using a clear-sky probability threshold of 0.99 (with pixels above this being labelled 'clear'), the classification of 89.9% of pixels agrees with that of Merchant's tailored mask, i.e. that is the estimated PP measure. The various measures for this image are given in Table 2. The PP is not very sensitive to the value of the threshold on the probability. The HR is between 71% and 87%, depending on the probability threshold. (If this HR does not seem very high, bear in mind that this scene has been chosen

TABLE 1. VERIFICATION TABLE

	Tailored cloudy	Tailored clear
Bayesian cloudy	h	f
Bayesian clear	u	z

h is the number of hits (number of pixels correctly flagged as cloudy in the Bayesian mask, according to the tailored mask), f is the number of false alarms, u is the number of misses (i.e. undetected cloud), z is the number of correct not-cloudy pixels, and $\tau = h + z + f + u$.

TABLE 2. PERFORMANCE OF THRESHOLDED CLEAR-SKY PROBABILITY ('P-CLEAR') USED AS CLOUD MASK AND OF THE STANDARD CLOUD MASK, VERIFIED AGAINST A TAILORED MASK

Mask	PP	HR	FAR	TSS
P-clear < 0.9	88.8	71.3	4.4	66.9
P-clear < 0.99	89.7	78.0	5.6	72.4
P-clear < 0.999	90.5	86.8	8.1	78.7
Standard	84.6	67.1	8.5	58.5

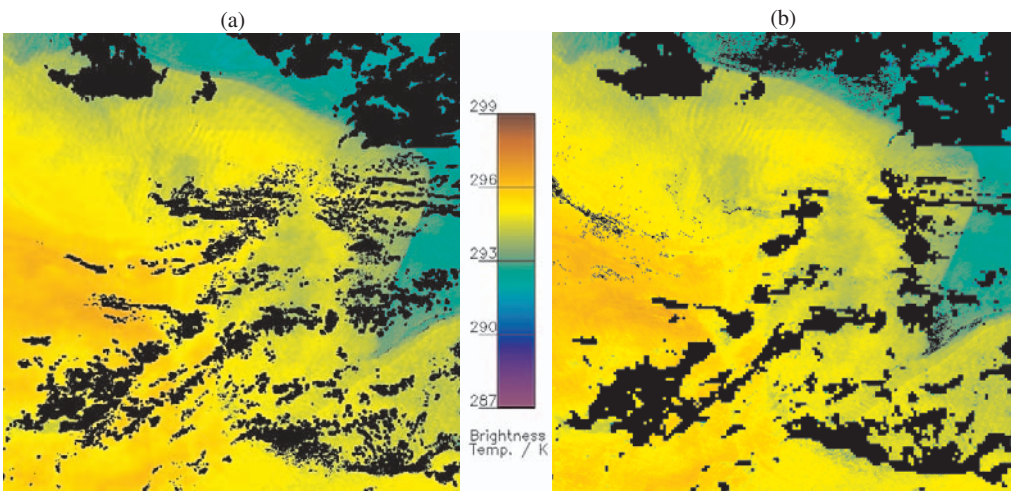

Figure 7. (a) 11 μm brightness temperature image screened by the Bayesian approach with a probability threshold of 0.99, and (b) the same image screened in the standard ATSR-2 product. These can be compared with the tailored mask (Fig. 3(b)).

because it has a lot of cloud that is relatively difficult to detect.) The visual comparison of the new mask (thresholded on 0.99) with the standard mask is favourable (see Fig. 7). Overall, the new mask is an improvement on the standard cloud screening. Of course, given our ability to use analysis fields rather than a fixed scheme, an improvement is to be expected. This illustrates the potential value of the new scheme in an operational context where forecast fields are available.

About a third of the wrongly classified pixels (in the new masks) are clear pixels that are adjacent to cloud. These are mostly misclassified as a result of the use of LSD in the observation vector. Could we do without the LSDs? An estimate of clear-sky probability can also be made using only the spectral elements of the observation vector (not shown). This reduces false detection (as cloudy) of clear pixels adjacent to cloud.

However, many areas of scattered cloud and of sub-pixel cloud at cloud edges are spectrally indistinguishable from clear sky given the uncertainty in the background state, and the fraction of pixels agreeing with the tailored mask falls to 84%. A measure of texture remains necessary in cloud screening of night-time imagery.

Some pixels that are cloudy yield clear-sky probabilities greater than 0.99, e.g. those around H in Fig. 4. Such cases present difficulties to automated cloud-screening systems. In Fig. 4, these pixels are seen to have BTs that are compatible with clear-sky pixels. The same pixels in Fig. 5 are seen to have LSDs that are comparable to those associated with modest thermal features of the ocean. In terms of BT and local texture, these pixels are insufficiently distinct from clear-sky pixels to be screened, except at the cost of incurring an unacceptable FAR. We detect these clouds on visual inspection by intelligently interpreting and interpolating larger-scale ocean temperature features; neither the traditional or Bayesian framework has a comparable metric at present.

The results described above are very much particular to this scene, and to the analysis fields used as prior information. Screening scattered, broken, low cloud of the sort prevalent in this scene is one of the more difficult challenges using 1 km resolution thermal imagery. In the next subsection, we apply similar analyses to a number of images with different cloud features.

(*b*) *Cloud screening a selection of images*

We developed a set of images for assessing the success of Bayesian cloud masking relative to the standard cloud masks as follows. We obtained a full day (28 January 1999) of ATSR-2 images, and, by examining 'quick looks', identified 59 images, each of size 500 × 500 km, covering a wide range of latitudes, ocean temperature structures and cloud formations. Merchant and Harris developed cloud masks for each of these images 'by eye', independently of each other and of the standard cloud mask. Details of the method are given in appendix B to this paper. After screening, the masks were jointly reviewed to assess the degree of agreement about interpretation of each image. In 16 cases there were identified features in the image that were differently interpreted. We retain for the analysis below only the 43 images remaining after excluding these; this selection represents the images where we consider the visual interpretation to be relatively reliable. The selection still includes 'difficult' cases such as thin cirrus, cloud systems with significant water vapour halos, and near-total cloudiness. Overall, the cloudiness of the selected images is 71%, and Harris flagged 2% more pixels as cloudy than Merchant.

For the selected 43 images, 96.5% of pixels have the same classification in each of the independent tailored masks. The analyses that follow use only these pixels in order to have very high confidence in the 'expert' classification for use as 'truth'.

Figure 8(a) shows the distributions obtained of estimated clear-sky probability $P(c \mid \mathbf{y}^{\mathrm{o}}, \mathbf{x}^{\mathrm{b}})$ (hereafter P_{c}), for the full estimate (including texture in the observation vector) and for the estimate using only BTs ('spectral-only'). The distributions are bimodal, as noted with reference to Fig. 6 above, with clearer separation between modes when texture is included as an observation.

We show the fraction of pixels that are 'truly' clear as a function of P_{c} in Figs. 8(b) and (c). These curves generally increase monotonically from low to high fractions as P_{c} increases, as should be expected. The correspondence is not one-to-one, which one might expect from a straightforward interpretation of P_{c}. This may reflect systematic biases in the tailored mask, the prior information, the cloudy PDFs, etc. Compared to the spectral-only curve, the curve for the full estimate is offset at low P_{c}

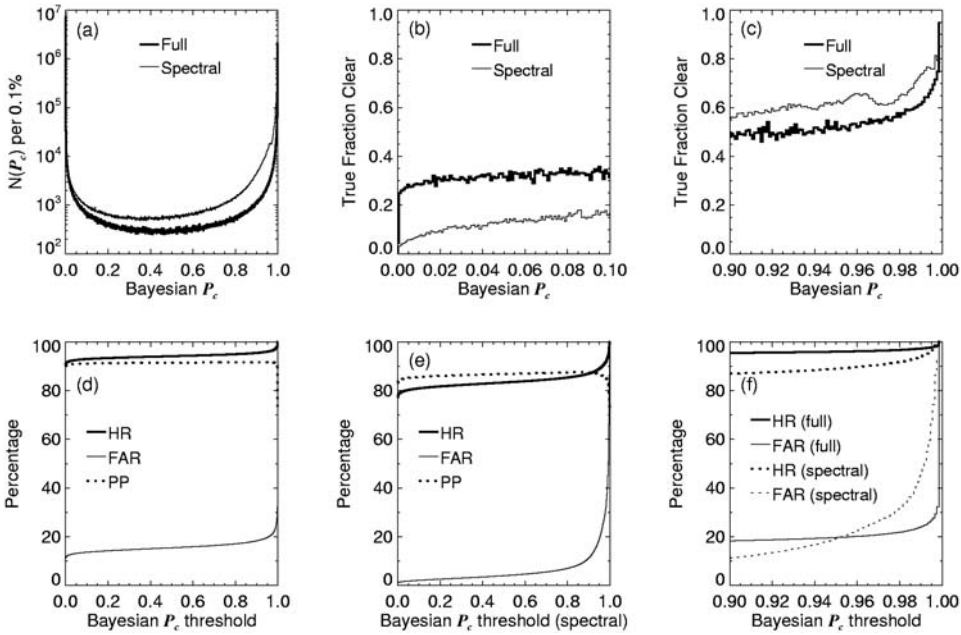

Figure 8. Results of estimated clear-sky probability, P_c, from 43 images: (a) distribution histogram of P_c when the observation vector includes brightness temperatures and LSD (Full), and brightness temperatures only (Spectral); (b) and (c) fraction of pixels 'truly' clear versus P_c (binned); (d) hit rate (HR), false-alarm rate (FAR) and proportion perfect (PP) as a function of the threshold used on P_c to create a binary mask for full observation vector; (e) as (d), but for spectral-only P_c; and (f) detail of top range of (d) and (e) for HR and FAR.

TABLE 3. PERFORMANCE OF THRESHOLDED CLEAR-SKY PROBABILITY ('P-CLEAR') USED AS CLOUD MASK AND OF THE STANDARD CLOUD MASK, VERIFIED AGAINST JOINT TAILORED MASKS FOR 43 SELECTED IMAGES

Mask	PP	HR	FAR	TSS
P-clear < 0.9	91.7	95.5	18.3	77.1
P-clear < 0.99	91.4	97.2	23.7	73.5
P-clear < 0.999	90.1	98.6	32.3	66.3
Standard	88.4	96.3	32.2	64.1

(a result of false alarms associated with the texture element) and rises more gradually at intermediate P_c (as result of the greater separation of classes from use of texture).

Figures 8(d) to (f) show the dependence of PP, HR and FAR on the threshold applied to P_c to create a binary mask, as verified against the classification in the joint tailored masks. Table 3 (analogous to Table 2) gives the numerical values for specific thresholds, and also the results for the standard cloud mask. These curves are instructive in deciding where to set this threshold for practical use (assuming most users will continue to require a binary cloud mask).

Considering first the FAR, we note (Fig. 8(d)) that, even at a very low threshold, it is over 10% for the full P_c estimate. This is the effect of using LSD over 3×3 pixel boxes as a texture observation: 42% of all clear pixels are adjacent to cloudy pixels in the joint tailored masks, and so a significant FAR is to be expected overall since pixels adjacent to cloud have an increased LSD. For the spectral-only P_c (Fig. 8(e)), the FAR is small (<10%) for thresholds on P_c of 0.9 or smaller. However, the spectral-only HR is too

low to be acceptable (<95%) unless a threshold >0.99 is used, but this in turn causes an unwelcome FAR of ~40% (Fig. 8(f)). Only by including the texture observations is a reasonable HR–FAR combination possible (by choosing a threshold on P_c of about 0.99; see also Table 3).

The performance of the new mask made by thresholding P_c at 0.99 is somewhat better than that of the standard screening. Both HR and FAR are improved (Table 3), as one would expect given the use of scene-specific prior information, in the form of ECMWF fields. Both the Bayesian and the standard schemes effectively consider the location of each pixel in 'BT space' relative to prior expectations. Both schemes also use LSD over 3×3 pixel boxes, and are therefore prone to elevated FAR, especially where there is broken cloud and/or strong gradients in SST. However, the FAR in the standard cloud mask, at 32%, is also increased by overdetection over cold SST features that are adjacent to warmer water (e.g. cold-core eddies, boundary current regions). In three of the 43 images, large areas ($>10^4$ km^2) of low SST under clear sky were erroneously flagged, by either the large-scale spatial coherence test or the $11/12$ μm thermal histogram test (see Závody *et al.* 2000, for descriptions of these tests); similar occurrences have also been observed by Simpson *et al.* (2005).

(*c*) *SST impact of cloud-screening errors*

Are the errors in screening in the new and standard cloud masks significant in terms of SST products from TIR imagery? Usually, the SST products used in NWP are at lower resolution than TIR imagery, and so we answer this from the viewpoint of cloud-screening effects on an averaged SST product. We calculate average SSTs (ASSTs) in 100×100 km cells (approximating the 1° latitude–longitude spatial resolution typically used in NWP SST analyses, e.g. Reynolds and Smith 1994), using the new, standard and tailored cloud masks. We retain only those cells where SSTs were obtained for at least 100 clear-sky pixels in the tailored mask to ensure that ASSTs are adequately precise, there being 869 such cells. The mean ASST over the cells is 290.60 K for the tailored mask, 290.56 K for the standard mask, and 290.54 K for the new mask. However, one should not infer from the smallness of these differences that the masks are equivalent for SST purposes, as explained below.

Figure 9(a) shows the distribution of difference between ASST derived using each of the automated masks and 'true' ASST derived using the joint tailored mask. This represents directly the range of impact of screening-related errors on ASSTs.

Firstly, we can see in Fig. 9(a) that the new screening is less prone to gross errors in ASST. The tails of the distribution are more populated in the case of the standard mask (the standard deviation of the distribution is 0.19 K for the new mask and 0.53 K for the standard mask). Using the standard mask, there are extreme errors in ASST of about 6 K (of both signs) and 3.5% of ASSTs are in error by more than 1 K in magnitude. In contrast, for the new mask the most extreme error has a magnitude of 3 K and only 0.6% of ASSTs are in error by more than 1 K.

Secondly, it turns out that ASSTs derived using the new mask are also more precise and less biased in the main peak of the distribution. A 'robust' measure of standard deviation (based on the spread of the middle 50% of data) is 0.07 K for the new mask and 0.09 K for the standard mask. The bias of the main peak is better estimated by the median of the distribution than by the mean. For the standard mask the median ASST error is –0.07 K, whereas for the new mask it is –0.03 K (a statistically significant difference at the 95% level of confidence).

Cloud-screening errors comprise both failure to detect (HR < 100%) and false alarms (FAR > 0%). Figures 9(b) to (e) separate out the consequences for ASST from

these different types of error. Figure 9(b) shows how different SSTs of pixels that are falsely flagged are from pixels in the same cell that are correctly passed as clear. The majority of the larger errors from the standard screening are warm ASST biases arising from incorrect screening of cold ocean areas, i.e. screening of truly clear pixels that are colder than their neighbours. Figure 9(c) shows how different are the SSTs of pixels that are incorrectly passed as clear from pixels in the same cell that are correctly passed as clear. The SSTs calculated for these truly cloudy areas are, as one would expect, lower than the nearby truly clear areas. They are higher in the case of the new mask (mean of distribution is –0.53 K) than in the case of the standard mask (–0.76 K). In other words, cloud-affected SSTs are less biased using the new mask.

Figures 9(b) and (c) are distributions of differences in calculated SST between wrongly classified and truly clear pixels. (These distributions do not correspond to the errors caused in ASST, because, when the ASST is calculated, the wrongly classified and truly clear pixels contribute to the cell average in proportion to their prevalence within the cell.) The ASST errors caused by the two types of classification error are distributed as shown in Figs. 9(d) and 9(e) for false alarms and failures to detect, respectively.

The purpose of this subsection has not been to criticize the standard cloud screening for ATSR-2, so much as to explore a concrete example of the effects of cloud-screening deficiencies on ASST products. The problems identified in previous sections with the standard ATSR-2 cloud screening have been shown to have significant adverse impacts on a small (but not negligible) fraction of 1° resolution ASSTs. With its more sophisticated and 'softer' use of prior information, the new Bayesian screening suffers fewer such effects, representing a useful improvement in outcome. This is promising. In this work, we have attempted to quantify and compare the errors arising from alternative cloud-screening systems against 'truth' (the joint tailored cloud masks). Establishing this 'truth' has been onerous, but an approach like this is needed to quantify screening-related errors in absolute terms.

5. Discussion and conclusions

In this paper, we present the theory for applying Bayes' theorem to cloud screening of TIR images in an operational context. The method develops the approach of English *et al.* (1999) to obtain a means to estimate objectively the likelihood of a pixel being cloud free in the light of the prior information available.

By choosing a threshold on that probability estimate of 0.99 to generate a binary mask, we obtain cloud masks that perform well compared to tailored masks (defined by expert judgement), and perform better than the traditional cloud-screening system used to provide the standard mask for the imagery. While most users will continue to require a binary cloud mask, more sophisticated uses of clear-sky probability can be envisaged in the future in the context of introducing data to assimilation systems—for example, using the clear-sky probability to refine error estimates for observations.

The new approach is rigorous, physically based and properly probabilistic, and, in that sense, arguably more aesthetically satisfying than threshold schemes. The overall improvements that we obtain from applying the new scheme are (i) a ~1% improvement in hit rate from ~96% to 97%, which of course corresponds to reducing the failure-to-detect rate by one quarter, and (ii) a reduction in false alarm from 32% to 24%. Furthermore, the remaining cloud-screening errors in the new scheme have demonstrably less effect on retrieved SST; screening-related errors in ASSTs, as measured

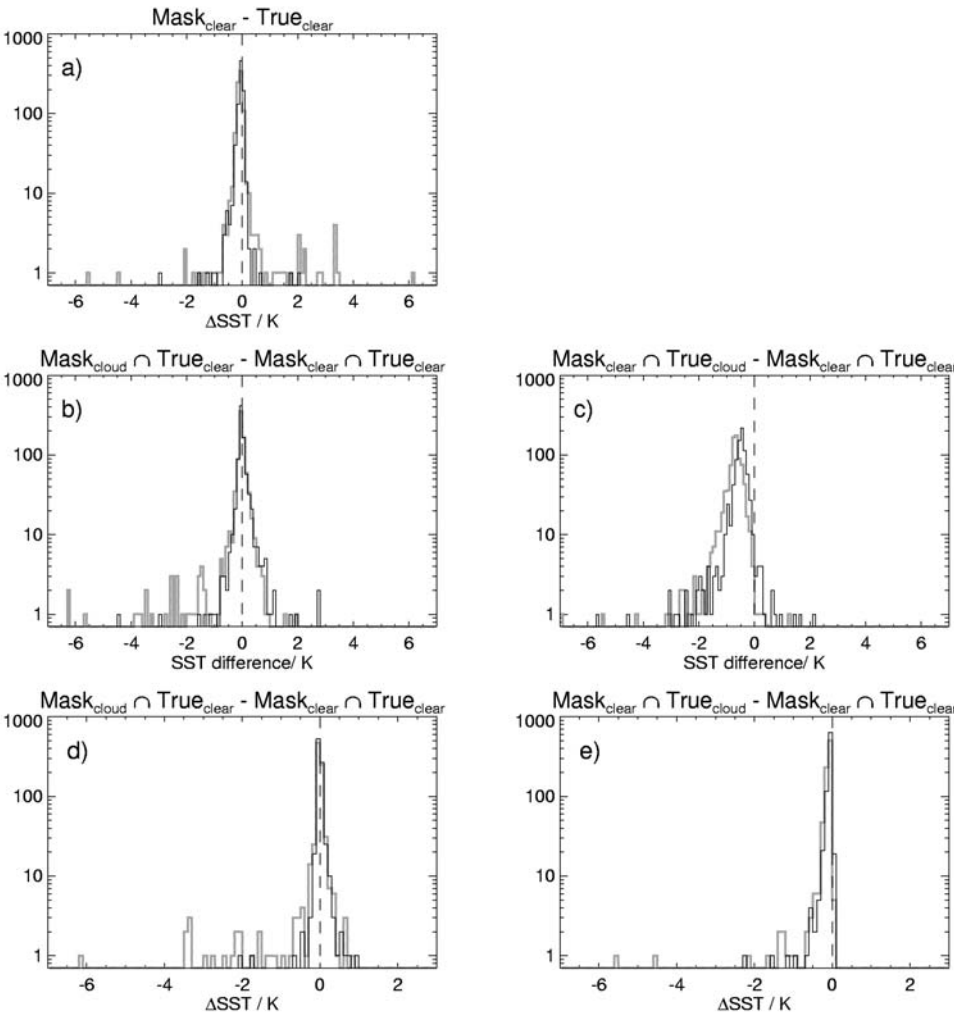

Figure 9. Effects of cloud-screening errors on 1° × 1° averaged SSTs for the new mask (solid lines, based on estimated $P_{clear} < 0.99$) and the standard mask (light grey lines), compared to a 'true' mask based on expert image interpretation. The SST retrievals are a weighted combination of three brightness temperatures. (a) Distributions of difference in cell-averaged SSTs between automated masks and true mask. This represents directly the influence of cloud-screening errors on spatially averaged SSTs, including both failures to hit and false detections of clouds. (b) Distribution of simple difference in cell-averaged SSTs for pixels that are falsely detected by the automated masks compared to those that are correctly identified as clear. This shows how different (in terms of SST) are the pixels that trigger false alarms from clear pixels within the same cell. (c) as (b), but for cloudy pixels that have not been detected by the automated masks compared to those that are correctly identified as clear. (d) and (e) are as (b) and (c), but the difference has been weighted to give the cell-averaged impact on SSTs (from false alarms and failures to detect, respectively).

by standard deviation, fell from 0.53 K to 0.19 K, mainly through a much lower frequency of errors in excess of 1 K. These are useful and significant improvements.

As noted previously by English *et al.* (1999), perhaps the most attractive aspects of this sort of approach are its generality and conceptual clarity. In effect, all the checks performed in a traditional cloud-screening system—checks on the plausibility and internal consistency of interpreting the observations as 'clear'—are marshalled

into a single assessment. Generality implies enhanced maintainability and adaptability for operational agencies when faced with new sensors, or in-flight changes such as loss of a channel. Provided the fast forward-modelling capability is in place, adapting the Bayesian approach can be straightforward. For example, at time of acceptance for publication, cloud-screening code based on this research is undergoing testing for operations of the constellation of GOES imagers at NOAA (using the same framework, but a different forward model and forecast fields from the NCEP). The time taken to screen the imagery is about twice that of the older threshold-based scheme, which is still well within the requirements for real-time operations.

Conceptual clarity helps when seeking to improve a cloud-screening system, especially if the original design expertise has been lost. Sophisticated screening systems can be rather opaque and delicate to change; 'tweaking' a test to solve a particular screening problem can have unforeseen consequences on screening elsewhere. In the Bayesian framework, improvements will come from using better prior information, better forward modelling, and/or better PDFs. Each of these elements can be examined systematically and independently when seeking to address any particular failure of screening.

There are two short cuts taken in the Bayesian scheme presented here: the use of a reduced state vector (only SST and TCWV) in the forward calculation of the Gaussian distribution of clear-sky brightness temperatures, and the use of global (non-conditional) cloudy-sky PDFs. Future work will assess the degree of improvement that can be brought by a more complete implementation in these areas.

Although the work presented in this paper focuses on night-time (TIR-only) imagery—generally the harder case—the principles also apply to reflectance channels in daytime imagery. Indeed, the system for the GOES SST processing (mentioned above) operates in daytime, using (at present) a highly simplified forward model for the reflectance channel (and dropping the 3.9 μm channel from the daytime observation vector). Similarly, although the work to date has been purely on cloud screening over ocean, work is under way to apply the technique to screening over land.

Another powerful approach to cloud screening has been described (Simpson *et al.* 2000) based on identifying the natural clusters of pixel properties in scenes, and then adaptively labelling these as clear or cloudy. This is interesting in that it seems to offer a 'prior-free' alternative; the classification emerges from the properties of the imagery. The labelling step, however, does assume that at least one cluster is clear sky, and there is ambiguity in the adaptive labelling in the night-time case. An interesting possibility is to tackle the labelling task on a properly probabilistic basis with the Bayesian approach described here, but applied to properties of classes rather than individual pixels.

Cloud screening of imagery is an important step in many applications of remote sensing. The Bayesian approach we present in this paper is well suited to the operational context, and indeed is the basis of a new cloud-screening system under development for GOES SST. Here we have demonstrated, on a sample of ATSR-2 TIR imagery, that a Bayesian estimation of clear-sky probability (albeit with some approximations) is effective for cloud screening, and have pointed to several promising directions for further research.

ACKNOWLEDGEMENTS

We thank the Rutherford Appleton Laboratory for providing the ATSR-2 data. We thank the British Atmospheric Data Centre and ECMWF for ERA-40 data.

S. MacCallum was funded by NOAA/NESDIS/ORA, and A. Harris is funded through the NOAA Cooperative Institute for Climate Studies. The contents of this paper are solely the opinions of the authors and do not constitute a statement of policy, decision or position on behalf of NOAA or the US government. We thank the anonymous reviewers of the manuscript for their contribution to improving its presentation.

APPENDIX A

TABLE A.1. LIST OF ACRONYMS

ASST	Averaged sea surface temperature
ATSR	Along-Track Scanning Radiometer
AVHRR	Advanced Very High Resolution Radiometer
BT	Brightness temperature
ECMWF	European Centre for Medium-range Weather Forecasting
ERA	ECMWF re-analysis
FAR	False-alarm rate
GOES	Geostationary Operational Environmental Satellite
HR	Hit rate
LSD	Local standard deviation
NCEP	National Centers for Environmental Prediction
NEΔT	Noise-equivalent differential temperature
NOAA	National Oceanic and Atmospheric Administration
NWP	Numerical weather prediction
PDF	Probability density function
PP	Proportion of perfect forecasts
RTTOV	Radiative Transfer for TIROS Operational Vertical sounder
SST	Sea surface temperature
TCWV	Total column water vapour
TIR	Thermal infrared
TSS	True skill score

APPENDIX B

Details of cloud screening 'by eye'

Each ATSR-2 'image' consisted of a 500 km $\times$ 500 km array of nadir BTs in three thermal channels. The locations of the screened images are shown in Fig. A.1. To determine the tailored cloud masks, both Merchant and Harris independently used a 'widget' implemented in Interactive Data Language. The widget allows a user to display any channel with adjustable stretch. The widget can also be used to apply a number of screening tests to different channels, while watching the change in the cloud mask as thresholds for these tests are varied. The tests available in the widget are:

(i) a 'gross threshold test' that masks 3.7 μm BTs below a threshold that can vary linearly over the image;

(ii) a similar test on 12 μm BTs;

(iii) a test masking pixels with a value of 3.7 μm BT minus 11 μm BT below a threshold;

(iv) a test masking pixels with a value of 3.7 μm BT minus 11 μm BT above a threshold;

(v) a test comparing a pixel's BT value with the median value of BT in a surrounding box (a texture measure that operates a full resolution and therefore does not overdetect in situations of broken cloud); and

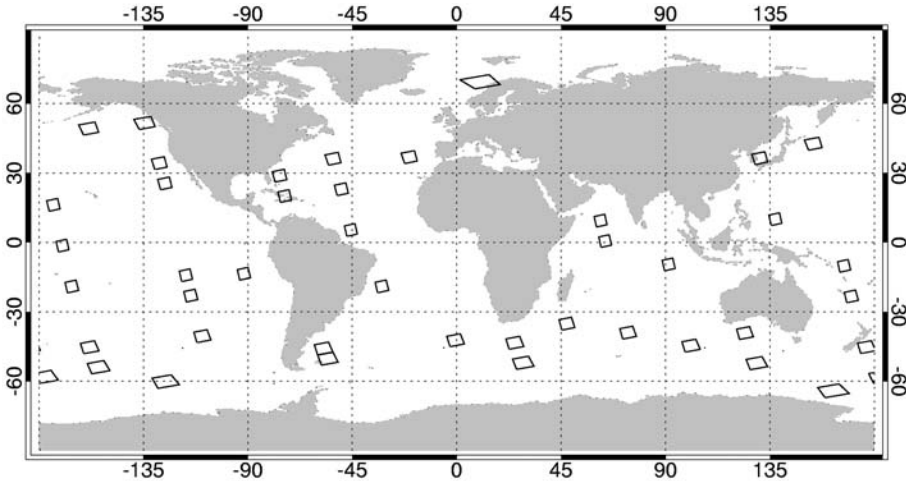

Figure B.1. Location of 43 images considered to be reliably screened 'by eye' and used for testing the new cloud-detection scheme.

(vi) a test in which a smoothed, interpolated 'background' 3.7 μm BT field is created from all the pixels not yet screened, so that pixels that deviate from this background by more than a threshold can be screened.

Thereafter, the user can edit the screening status of each individual pixel that has not been satisfactorily dealt with by the previous tests. Screening by eye is a laborious task, and despite using the widget to streamline the process, producing the cloud masks required about 25 hours of effort on the part of each investigator.

References

Brisson, A., Le Borgne, P. and Marsouin, A. 2002 Results of one year of pre-operational production of sea surface temperatures from GOES-8. *J. Atmos. Oceanic Technol.*, **19,** 1638–1652

Deschamps, P. Y. and Phulpin, T. 1980 Atmospheric correction of infrared measurements of sea surface temperature using channels at 3.7, 11 and 12 μm. *Boundary-Layer Meteorol.*, **18,** 131–143

Di Vittorio, A. V. and Emery, W. J. 2002 An automated, dynamic threshold cloud-masking algorithm for daytime AVHRR images over land. *IEEE Trans. Geosci. Remote Sensing*, **40,** 1682–1694

ECMWF 2000–2005 'ERA-40 project report series'. ECMWF, Reading, UK, available at http://www.ecmwf.int/publications/library/do/references/list/192

Edwards, T., Browning, J., Delderfield, D. J., Lee, K. A., Lidiard, R. S., Milborrow, R. S., McPherson, P. H., Peskett, G. M., Toplis, G. M., Taylor, H. S., Mason, I., Mason, G., Smith, A. and Stringer, S. 1990 The along-track scanning radiometer—measurement of sea-surface temperature from ERS-1. *J. Br. Interplanet. Soc.*, **43,** 160–180

English, S. J., Eyre, J. R. and Smith, J. A. 1999 A cloud-detection scheme for use with satellite sounding radiances in the context of data assimilation for numerical weather prediction. *Q. J. R. Meteorol. Soc.*, **125,** 2359–2378

Le Gléau, H. and Derrien, M. 2000 'Prototype scientific description for Meteo-France'. Report SAF/NWC/MFCMS/MTR/PSD, Issue 1, Rev. 1, available at http://www.meteorologie.eu.org/safnwc/publis/mtrsci10.pdf

McClain, E. P. 1989 Global sea-surface temperatures and cloud clearing for aerosol optical depth estimates. *Int. J. Remote Sensing*, **10,** 763–769

Marsham, J. 2003 'Lake temperatures—thermal remote sensing and assimilation into a lake model'. PhD thesis, University of Edinburgh, UK

Mathur, A. K., Agarwal, V. K. and Panda, T. C. 2002 Validation of ERS-1/ATSR-derived SST in Indian waters. *Int. J. Remote Sensing*, **23,** 5155–5163

Merchant, C. J. and Harris, A. R. 1999 Toward the elimination of bias in satellite retrievals of sea surface temperature. 2: Comparison with *in situ* measurements. *J. Geophys. Res.–Oceans*, **104,** 23579–23590

Murtagh, F., Barreto, D. and Marcello, J. 2003 Decision boundaries using Bayes factors: the case of cloud masks. *IEEE Trans. Geosci. Remote Sensing*, **41,** 2952–2958

Parkes, I. M., Sheasby, T. N., Llewellyn-Jones, D. T., Nightingale, T. J., Zavody, A. M., Mutlow, C. T., Yokoyama, R., Tamba, S. and Donlon, C. J. 2000 The Mutsu Bay experiment: Validation of ATSR-1 and ATSR-2 sea surface temperature. *Int. J. Remote Sensing*, **21,** 3445–3460

Reynolds, R. W. and Smith, T. M. 1994 Improved global sea-surface temperature analyses using optimum interpolation. *J. Climate*, **7,** 929–948

Reynolds, R. W., Rayner, N. A., Smith, T. M., Stokes, D. C. and Wang, W. Q. 2002 An improved *in situ* and satellite SST analysis for climate. *J. Climate*, **15,** 1609–1625

Rodgers, C. D. 1976 Retrieval of atmospheric temperature and composition from remote measurements of thermal radiation. *Rev. Geophys. Space Phys.*, **14,** 609–624

Saunders, R.W. 1986 An automated scheme for the removal of cloud contamination from AVHRR radiances over Western Europe. *Int. J. Remote Sensing*, **7,** 867–886

Saunders, R. W. and Kriebel, K. T. 1988 An improved method for detecting clear-sky and cloudy radiances from AHRR data. *Int. J. Remote Sensing*, **9,** 123–150

Saunders, R., Matricardi, M. and Brunel, P. 1999 An improved fast radiative transfer model for assimilation of satellite radiance observations. *Q. J. R. Meteorol. Soc.*, **125,** 1407–1425

Simpson, J. J., Schmidt, A. and Harris, A. 1998 Improved cloud detection in Along-Track Scanning Radiometer (ATSR) data over the ocean. *Remote Sensing Environ.*, **66,** 110

Simpson, J. J., McIntire, T. J. and Sienko, M. 2000 An improved hybrid clustering algorithm for natural scenes. *IEEE Trans. Geosci. Remote Sensing*, **38,** 1016–1032

Simpson, J. J., Tsou, Y. L., Schmidt, A. and Harris, A. 2005 Analysis of Along-Track Scanning Radiometer-2 (ATSR-2) data for clouds, glint and sea surface temperature using neural networks. *Remote Sensing Environ.*, in press

Uddstrom, M. J., Gray, W. R., Murphy, R., Oien, N. A. and Murray, T. 1999 A Bayesian cloud mask for sea surface temperature retrieval. *J. Atmos. Oceanic Technol.*, **16,** 117–132

Závody, A. M., Mutlow, C. T. and Llewellyn-Jones, D. T. 2000 Cloud clearing over the ocean in the processing of data from the Along-Track Scanning Radiometer (ATSR). *J. Atmos. Oceanic Technol.*, **17,** 595–615