

A Bayesian Cloud Mask for Sea Surface Temperature Retrieval

MICHAEL J. UDDSTROM, WARREN R. GRAY, RICHARD MURPHY, NILES A. OIEN, AND TALBOT MURRAY

National Institute of Water and Atmospheric Research, Wellington, New Zealand

(Manuscript received 29 January 1997, in final form 2 February 1998)

ABSTRACT

Bayesian methods are used to develop a cloud mask classification algorithm for use in an operational sea surface temperature (SST) retrieval processing system for Advanced Very High Resolution Radiometer (AVHRR) local area coverage (LAC) resolution data. Both radiative and spatial features are incorporated in the resulting discriminant functions, which are determined from a large training sample of cloudy and clear observations. This approach obviates the need to specify the arbitrary thresholds used by hierarchical cloud-clearing methods, provides an estimate of the probability that an instantaneous field of view is cloudy (clear), and allows the skill of different cloud discriminant models to be objectively analyzed.

Results show that spatial information is of particular importance in reducing the false alarm rate of the cloudy class. However, while the use of complex textural measures such as gray-level difference statistics—as opposed to simple statistics such as the standard deviation—improves the skill of nighttime cloud-masking algorithms, they are of little advantage during daytime hours.

Cloud mask discriminant models having similar high Kuipers' performance index scores (i.e., 0.935) are developed for both day and night satellite data from the Southern Hemisphere midlatitudes. Applied to LAC orbital (i.e., operational) data, the characteristics of the cloud masks appear to be similar to those derived from analysis of the training sample data. However, in this case, to enhance processing performance, a hybrid algorithm is employed—obviously cloudy instantaneous fields of view (IFOVs) are first removed via a gross threshold check and the Bayesian method applied only to the remaining IFOVs. This same (hybrid) algorithm is also applied to an ensemble of 30 days of AVHRR LAC data from the New Zealand region. Analysis of the resulting time-composited SST data (means and standard deviations) shows there is little evidence of a day–night bias in the performance of the Bayesian cloud-masking algorithm and that the resulting SST data may be used to determine the variability of oceanographic features.

Although this paper uses AVHRR data to demonstrate the principles of the Bayesian cloud-masking algorithm, there is no reason why the approach could not be used with other instruments.

1. Introduction

Accurate sea surface temperature (SST) measurements are essential for climate research since the SST controls the release of heat (a key factor in the energy budget) from the ocean to the atmosphere. Present SST data accuracy requirements for global climate research may be stated as ± 0.3 K or better over a length of 100 km and a timescale of days to weeks (Harries et al. 1986; Smith et al. 1996).

There have been a number of collocation studies of the global and regional accuracy of SSTs derived from the Advanced Very High Resolution Radiometer (AVHRR) sensors on board National Oceanic and At-

mospheric Administration (NOAA) satellites. For example, linear multichannel SST (MCSST) retrieval algorithms (McMillin and Crosby 1984) have regional accuracies of about 0.4–0.6 K (McClain 1989). Non-linear algorithms (Walton 1988) produce more accurate (than MCSST) results, especially in the presence of noise in the satellite data (McClain et al. 1990). However, as Gallegos et al. (1993) indicate, even operational systems such as those used by NOAA (McClain et al. 1985) show cloud contamination of the retrieved SSTs remains a problem, especially during night hours. To further improve the accuracy of SST data, new satellite instruments have been developed, such as the Along-Track Scanning Radiometer (ATSR) on the *ERS-1* and *ERS-2* satellites. The ATSR has very stable radiometric calibration, low noise detectors, and measures radiances from two atmospheric paths, leading to an expected SST accuracy of 0.3 K (Delderfield et al. 1986), which is in keeping with global climate research requirements. Some reported validation results show that ATSR data can be used to estimate SST to about 0.3 K precision (e.g., Forrester et al. 1993; Barton et al. 1995). Unfor-

Corresponding author address: Dr. Michael J. Uddstrom, National Institute of Water, and Atmospheric Research, P.O. Box 14 901, Wellington, New Zealand.
E-mail: m.uddstrom@niwa.cri.nz

tunately, the accuracy of “operational” SST data products is invariably poorer. For example, Jones et al. (1996) used analyses of SST variability to show that residual cloud contamination in the 0.5° spatially averaged ATSR “averaged SST” (ASST) data product over the South Atlantic leads to significant unphysical day–night biases and a consequent reduction in the accuracy of the SST data.

Due to skin (Hepplewhite 1989), diurnal thermocline (Wick et al. 1992), and emissivity effects (Smith et al. 1996), it is difficult to validate (see, e.g., Minnett 1991) satellite-derived SST data at the level of accuracy required for climate research. Nonetheless, undetected cloud in SST data generated by “operational” retrieval systems continues to be a problem since it introduces bias noise into the SST estimates.

Present operational SST cloud detection methods used by NOAA (McClain et al. 1985) for AVHRR, and the Rutherford Appleton Laboratory (e.g., Jones et al. 1996; Saunders and Kriebel 1988) for ATSR, are based on sequential decision tree algorithms. These methods together with the newly developed NOAA clouds from AVHRR (CLAVR) algorithm (Stowe et al. 1991) to be used (operationally) by NOAA in the future (McClain et al. 1990) use thermal, spatial coherence, reflectance, and multispectral features together with both dynamic and static threshold tests to determine whether a field of view is clear, partly cloudy, or cloudy. The details of these algorithms differ, significantly in some cases, but the fundamental ideas are similar. In a wider context, pattern recognition approaches using higher-order spatial textural information in satellite radiances have been employed (with varying levels of success) to determine cloud morphologies over complex and simple surfaces (e.g., Parikh 1977; Ebert 1987; Welch et al. 1988; Tovinkere et al. 1993; Karlsson 1994; Chou et al. 1994; Uddstrom and Gray 1996).

While these complex methods show significant cloud detection skill, they are not really suitable for determining cloud masks for operational SST processing since they are computationally expensive and yield low spatial resolution masks relative to the size of the radiometer instantaneous field of view (IFOV). Gallegos et al. (1993) have developed a cloud-masking algorithm that utilizes more sophisticated spatial measures than those used in present operational cloud detection algorithms, yet it is suitable for use in an operational AVHRR SST retrieval system. In particular, gray-level co-occurrence (GLC) (Haralick et al. 1973) measures of scene “texture” in AVHRR $0.6\text{-}\mu\text{m}$ reflectances are used during daytime hours, and thermal difference channels (i.e., $3.7\text{--}11$ and $11\text{--}12\ \mu\text{m}$) during nighttime hours. Again, using threshold methods, the texture features are combined with statistics of the 0.6- and $11\text{-}\mu\text{m}$ data to partition 3×3 local area coverage (LAC) IFOVs into clear or cloudy categories. These authors report that the resulting day and night algorithms performed “better” than those used operationally at NOAA

(McClain et al. 1985) and in Sea Space Corporation’s commercial software (which is similar to the NOAA algorithm). Likewise they state that their daytime results were superior to those from the threshold method proposed by Simpson and Humphrey (1990).

This paper presents a new type of operational cloud detection algorithm that uses spectral and or spatial (including second-order textural statistics) features in a Bayesian discriminant model to optimally detect the presence of cloudy and clear IFOVs. With this approach, the radiative and spatial information for some $n \times n$ IFOV data tile (typically $n = 3$) is utilized simultaneously and the probability that the tile is clear (or cloudy) estimated. A cloud mask may then be specified based on some threshold posterior probability that the tile is clear, thus minimizing “costly” misclassifications. This approach has the additional advantage that the skill of cloud discriminant models derived from different sets of radiative and spatial measures may be evaluated directly. However, while the Bayesian approach is optimal (assuming the necessary assumptions about the underlying probability distributions are satisfied), its absolute skill is determined by two requirements: 1) separability of the two classes in measurement feature space and 2) the span of the training sample of independent data used to specify the discriminant functions (Uddstrom and Gray 1996).

Here, the theoretical basis of the Bayesian cloud mask is developed in section 2, together with the definition of some potential radiometric and spatial features. Section 3 discusses the characteristics of the AVHRR training sample data and discriminant features, while section 4 describes the skill of several cloud mask discriminant models. Example AVHRR cloud masks and SST retrievals are shown in section 5. Section 6 summarizes the results of this research and indicates future directions.

2. Theoretical basis

Given the a priori knowledge that some observable belongs to one of m possible classes (ω_i , $i = 1, m$), a vector of measurements (\mathbf{f} , called the feature vector) that in some sense characterizes the observable, and the assumption that all misclassifications are equally costly, then the posterior probability $P(\omega_i | \mathbf{f})$ is the conditional probability that the observable belongs to class ω_i . Accordingly, the best classification of the observable is the one that maximizes this probability, that is, it belongs to class i if

$$P(\omega_i | \mathbf{f}) > P(\omega_j | \mathbf{f}), \quad \forall j \neq i. \quad (1)$$

From Bayes theorem

$$P(\omega_i | \mathbf{f}) = P(\mathbf{f} | \omega_i)P(\omega_i) / \sum_{j=1}^m P(\mathbf{f} | \omega_j)P(\omega_j), \quad (2)$$

where $P(\omega_i)$ is the a priori probability that the observable belongs to class i . If the underlying distribution is

multivariate Gaussian and the measurement errors are unbiased, a measure of the distance of feature \mathbf{f} from class i can be specified by the following discriminant function:

$$g_i(\mathbf{f}) = \log[P(\omega_i)/(2\pi)^{m/2}|\mathbf{S}_{f_i}|^{1/2}] - \frac{1}{2}(\mathbf{f} - \bar{\mathbf{f}}_i)^T \mathbf{S}_{f_i}^{-1}(\mathbf{f} - \bar{\mathbf{f}}_i), \quad (3)$$

where $\bar{\mathbf{f}}_i$ and \mathbf{S}_{f_i} are the mean (vector) and covariance (matrix) for class i and some set of measurable features. The posterior probability that the feature vector belongs to class i is then

$$P(\omega_i | \mathbf{f}) = \exp[g_i(\mathbf{f})] / \sum_{j=1}^m \exp[g_j(\mathbf{f})], \quad (4)$$

and the observable belongs to the class that satisfies Eq. (1).

In the case of determining a cloud mask for the SST retrieval problem there are just two possible classes—clear or cloudy—although the posterior probability for either class may not be unity, and the measurements could represent a partly cloudy observation. The accuracy, or skill, of such a cloud detection discriminant function will be dependent upon the separability of the two classes in feature vector space, which is a function of the class means and covariances and measurement noise.

For AVHRR measurements (Planet 1988), the elements of the feature vector (\mathbf{f}) could simply consist of spectral features such as the 0.6- and 0.9- μm reflectivities (which will be referred to as channels R_1 and R_2), the emissive and/or reflective components of the 3.7- μm channel (Stowe et al. 1991) (channels T_3 and R_3), and the emissive 11- and 12- μm (channels T_4 and T_5) measurements. Derived channels may also be used, such as thermal differences (e.g., $T_4 - T_5$ and $T_4 - T_3$) and reflectance ratios (e.g., $R_2/R_1 - 1$) (similar to the Q feature used by Saunders and Kriebel 1988). Spatial information may also be incorporated into the discriminant feature vector by considering measures of spatial texture over some data tile of $n \times n$ IFOVs. Typically, these might consist of a tile range (as used in CLAVR and APOLLO) or a standard deviation. However, higher-order spatial features such as gray-level difference and co-occurrence statistics (Parikh 1977) have also been shown to increase the skill of cloud detection and classification algorithms (Ebert 1987; Welch et al. 1988; Gallegos et al. 1993; Uddstrom and Gray 1996). In this paper gray-level difference (GLD) spatial statistical features will be considered (together with tile range and standard deviation) since these are computationally inexpensive to compute and are known to perform well compared to other methods (Weszka et al. 1976).

GLD features measure the local properties of the absolute differences between pairs of gray levels in an $n \times n$ IFOV image tile. A GLD probability density func-

tion for the image [$f(x, y)$, $x = 1, n$; $y = 1, n$] can be defined and has the form $P(m)_{d,\theta}$, where the m th entry is the relative frequency of occurrence of gray-level difference $m = |f(x, y) - f(x', y')|$ for pixels $f(x, y)$ and $f(x', y')$ separated by d pixels in direction q (relative to the image's x axis). If the texture is coarse, relative to distance d , then the differences m might be expected to be small and $P(m)_{d,\theta}$ will be large for small m and small for large m . Conversely, for fine texture, where d is comparable to the scale of the features in the image, $P(m)_{d,\theta}$ will be large for large m . Accordingly, measures of the texture in an image may be computed from estimates of the spread in $P(m)$ at different separations and angles. Conventionally, four measures are utilized, where $m = |f(x, y) - f(x', y')|$, and N is the number of gray levels.

1) The mean, expressed as

$$\mu_{d,\theta} = \sum_{m=0}^N mP(m)_{d,\theta}, \quad (5)$$

which, if small, indicates that the GLD values are concentrated near the origin (i.e., $m = 0$) and the texture is “coarse” relative to the spatial scale d .

2) The contrast, expressed as

$$\text{contrast}_{d,\theta} = \sum_{m=0}^N m^2P(m)_{d,\theta}, \quad (6)$$

or second moment of the $P(m)$ density function, measures the variability of the gray levels in the image.

3) The angular second moment,

$$\text{ASM}_{d,\theta} = \sum_{m=0}^N [P(m)_{d,\theta}]^2, \quad (7)$$

measures the degree of homogeneity in the image. A small ASM implies that the gray-level differences are all similar and the sampled area has textural variations on a spatial scale close to d , while a large ASM indicates that there are dominant gray tones present (e.g., when the GLD values are concentrated near $m = 0$).

4) The entropy is expressed as

$$\text{entropy}_{d,\theta} = - \sum_{m=0}^N P(m)_{d,\theta} \log P(m)_{d,\theta}. \quad (8)$$

This parameter indicates whether the “texture” is organized. This is largest when the $P(m)$'s are equal, or randomly distributed, but is small they are very unequal.

Each statistic is computed at four angles ($\theta = 0^\circ, 45^\circ, 90^\circ, \text{ and } 135^\circ$) and any number of pixel separations d . Directionally averaged (i.e., isotropic) measures of texture are computed by averaging the parameters over all four directions. For stable GLD statistics [i.e., for a well defined probability density function $P(m)_{d,\theta}$], the tile size n would typically be 8 or larger (for $d = 1$). Here

a tile size of 3 and $d = 1$ is used (the same value as that employed by Gallegos et al. 1993), implying that the resulting GLD statistics will be a rather poor representation of texture if it is spatially “coarse” relative to the tile size. However, even coarse GLD statistics may be useful if they are primarily used to aid the detection of cloud edges and clear ocean data.

Last, although GLD statistics are nominally defined on gray levels (or counts), calibrated, binned physical variables are used here. This approach ensures that spatial characteristics derived from different NOAA series AVHRR data may be combined (for those AVHRR instruments having similar filter functions). Bin sizes were determined by taking account of the expected noise in the measurements and the required dynamic range. For the reflectance and temperature samples, the bin size is 0.25 (percent or degrees kelvin). The reflectance ratio channel ($R_2/R_1 - 1$) uses a bin size of 0.005, while the difference channels ($T_4 - T_5$ and $T_4 - T_3$) are binned at 0.125-K resolution.

3. A priori data

Implementation of the algorithm suggested by Eq. (3) requires the specification of feature mean and covariance matrices for the cloudy and clear classes. Critically, the fidelity of the cloud detection algorithm depends upon the span of the a priori data (i.e., training sample) used to calculate these entities as well as the selection of feature vector elements. Measurement noise may also affect the separability of the classes in the (measurement) feature space.

To calculate the feature mean and covariance matrices required by the discriminant equations, LAC spatial resolution AVHRR data at full (10 bit) radiometric resolution were archived over a 16-month period for a number of subareas in the New Zealand–Australia region (see Fig. 1). The low (area SPCZ) and high (area 55S) latitude samples include tropical and sub-Antarctic waters together with radiances from warm and cold (respectively) low-level clouds. The area labeled EAC lies over the usual location of the retroflection of the East Australian Current and was included since this oceanographic feature can produce large spatial gradients in the SST—features that need to be recognized by the cloud detection algorithm as clear ocean data. Fortunately, a cloud classification of data on areas AKRADAR, WNRADAR, and CHRADAR was available to the authors, since these areas have been used to study cloud classification algorithms (Uddstrom and Gray 1996). Finally, data from area SALPEX_3 were added since this region incorporates subtropical water masses in the Tasman Sea, high SST spatial gradients on the Southland Front just east of the South Island (of New Zealand), and sub-Antarctic waters south of the subtropical front over the Chatham Rise (Heath 1985).

Using an analyst and the ENHANCE interactive computing environment (Kidson et al. 1992; Uddstrom and

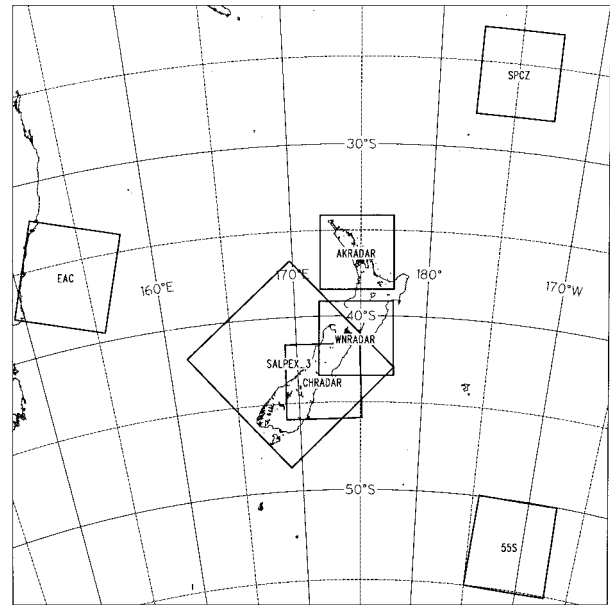


FIG. 1. Map showing areas over which training data were acquired.

Gray 1996), 3×3 IFOV data tiles from each of the archived areas were identified and labeled according to a warm ($T_4 > 0^\circ\text{C}$) cloud type. This approach, explained in detail in Uddstrom and Gray (1996), allows data to be specifically sampled from those cloud systems that have the potential to be mislabeled as clear by a cloud detection algorithm, for example, thin transmissive cirrus (Ci), stratus (St), and stratocumulus (Sc). Likewise, areas of high spatial gradient in cloud-free areas can be sampled. Last, cloud edges (i.e., of the order of 50% cloud cover) were sampled since any application of the algorithm will include data tiles that are partly cloudy or mixed, that is, include cloudy and clear IFOVs within a tile, as well as partially cloud-filled IFOVs.

For each sample and measurement “channel” (indicated in Table 1), tile mean, standard deviation, and range statistics together with GLD directionally averaged (i.e., isotropic) mean, angular second moment, contrast, and entropy statistics were computed.

All told, 2275 warm (i.e., $T_4 > 0^\circ\text{C}$) cloud, partly cloudy (i.e., mixed), and clear samples were labeled from the 16 months of data. Table 2 shows the distribution of these data by day and night. As might be expected, most emphasis in the labeling (i.e., training) phase has been placed on the nighttime data since are likely to be the more difficult to classify due to the reduced spectral information available.

Using Tukey box plots, the characteristics of selected features of these datasets are shown in Fig. 2. These diagrams suggest that the cloud detection problem, especially when the visible channels are unavailable, cannot easily be solved by the use of static thresholds (see also Stowe et al. 1991; Gallaudet and Simpson 1991). As evidenced in Fig. 2a the T_4 temperatures of Sc, St,

TABLE 1. Possible AVHRR cloud discriminant “measurement” channels.

Feature	Units	Explanation
R_1	%	Solar zenith angle-corrected channel 1 (0.6 μm) reflectance
R_2	%	Solar zenith angle-corrected channel 2 (0.9 μm) reflectance
$v_2 = (R_2/R_1 - 1)$		Visible ratio channel (good land/sea discrimination)
T_3	K	Channel 3 thermal (3.7 μm) channel (not used for daytime data)
T_4	K	Channel 4 thermal (11 μm) channel
T_5	K	Channel 5 thermal (12 μm) channel
$T_4 - T_3$	K	Stratus test difference channel (night only)
$T_4 - T_5$	K	Cirrus test difference channel (day and night)
I_3	mW/(sr m ² cm ⁻¹)	Channel 3 radiance
I_4	mW/(sr m ² cm ⁻¹)	Channel 4 radiance
I_5	mW/(sr m ² cm ⁻¹)	Channel 5 radiance

Ci, and mixed tiles intersect with the T_4 temperatures of cloud-free tiles. The difference channel ($T_4 - T_5$) (i.e., Fig. 2c) does not provide a simple discriminant for cloud-free tiles either. However, this feature is an excellent signal for transmissive cirrus since the emissivity of cirrus at 11 μm is lower than at 12 μm (Inoue 1985; d’Entremont 1986). The result is that ($T_4 - T_5$) may be quite large. Indeed, it is likely that the large positive ($T_4 - T_5$) samples in the CE (cloud edge) class of Fig. 2c indicate contamination of the sample by transmissive cirrus. This is not surprising since the CE class was labeled without recourse to a cirrus analysis (see Uddstrom and Gray 1996). The presence of cirrus in this class will cause no problems, however, since tiles including CEs are classified to the cloudy class. During nighttime hours, the temperature difference ($T_4 - T_3$) is near zero for clear samples, but it is positive for opaque water clouds such as stratus and fog since their emissivity at 3.7 μm is near 0.8, while at 11 μm it is close to 1.0 (Hunt 1973). However, mixed or open cell cumulus tiles (Fig. 2d), can have large negative ($T_4 - T_3$) differences when IFOVs are partially filled with cloud since the response of the blackbody function to temperature changes at 3.7 μm is approximately three times that at 11 μm . During daytime hours, the 0.9- μm reflectance (i.e., R_2) is generally an excellent discriminator of all but mixed tiles (Fig. 2b), where cloud shadows can lead to errors. The 0.9- μm channel is also less affected (than the 0.6- μm channel) by aerosol and Rayleigh scattering (Simpson and Humphrey 1990; Simpson and Gobat 1995).

The spatial statistics, Figs. 2e,f, suggest that a number of cloud types that are difficult to discriminate from

TABLE 2. Day and night, clear, cloudy, and mixed training sample sizes for 3×3 IFOV tiles of AVHRR data. In all cases the instrument scanangle was less than 50° , and for the daylight data they lie outside the sunglint region and have solar zenith angles less than 80° .

	Night	Day
Clear	1031	201
Cloudy	319	140
Mixed	247	248
Total	1587	589

radiative signatures alone can be discriminated using spatial information, for example, mixed (i.e., CE), Ci, and oCu tiles. In this regard the entropy statistic confirms that the no cloud (NC) and stratiform clouds show the most “organized” texture, while the cumuliform clouds and cloud edge samples show the least (that is, the GLDs are more randomly distributed).

4. Cloud model evaluation

The training sample dataset developed in section 3 may be used to test a number of possible Bayesian discriminant models, with the result that the usefulness of different radiative and spatial features can be evaluated objectively. These models may differ both in the definition of the radiative and spatial measures used, and the threshold posterior probability used to classify a sample (tile) as clear. They are also partitioned by day and night since day models can make use of the visible channels and night models the difference signal from the 11- and 3.7- μm channels. The 3.7- μm emissive data were not used in daytime models.

Results for nighttime cloud discriminant models are given in Table 3, while those for day data are in Table 4, where in each case the prior probabilities for the clear and cloudy samples are set to 0.5. The features used are indicated, while the skill of each discriminant function model is determined from its 2×2 (i.e., clear, cloudy) contingency table, using probability of detection (POD), false alarm rate (FAR), and the Kuipers’ performance index (KPI) (Murphy and Katz 1985) statistics. In this regard the KPI provides a relative measure of the accuracy of a discriminant function classifier compared to a random classification, thereby allowing the skill of different classifiers to be compared. The functional form of the KPI is

$$\text{KPI} = (\overline{\mathbf{cd}^T} - \bar{\mathbf{c}}\mathbf{P}^T)/(1 - \mathbf{PP}^T), \quad (9)$$

where \mathbf{c} is a row vector (c_1, \dots, c_m); $c_n = 1$ if the discriminant function predicts the observation belongs to class n , otherwise $c_n = 0$ for $n = (1, \dots, m)$; \mathbf{d} is the equivalent vector for the “true” class (i.e., cloudy or clear in this case); and \mathbf{P} is the vector of a priori probabilities for each of the m classes. The KPI has value 1 for a perfect classification and 0 for random classifications. The POD measures the accuracy of the discriminant function with respect to the number of ob-

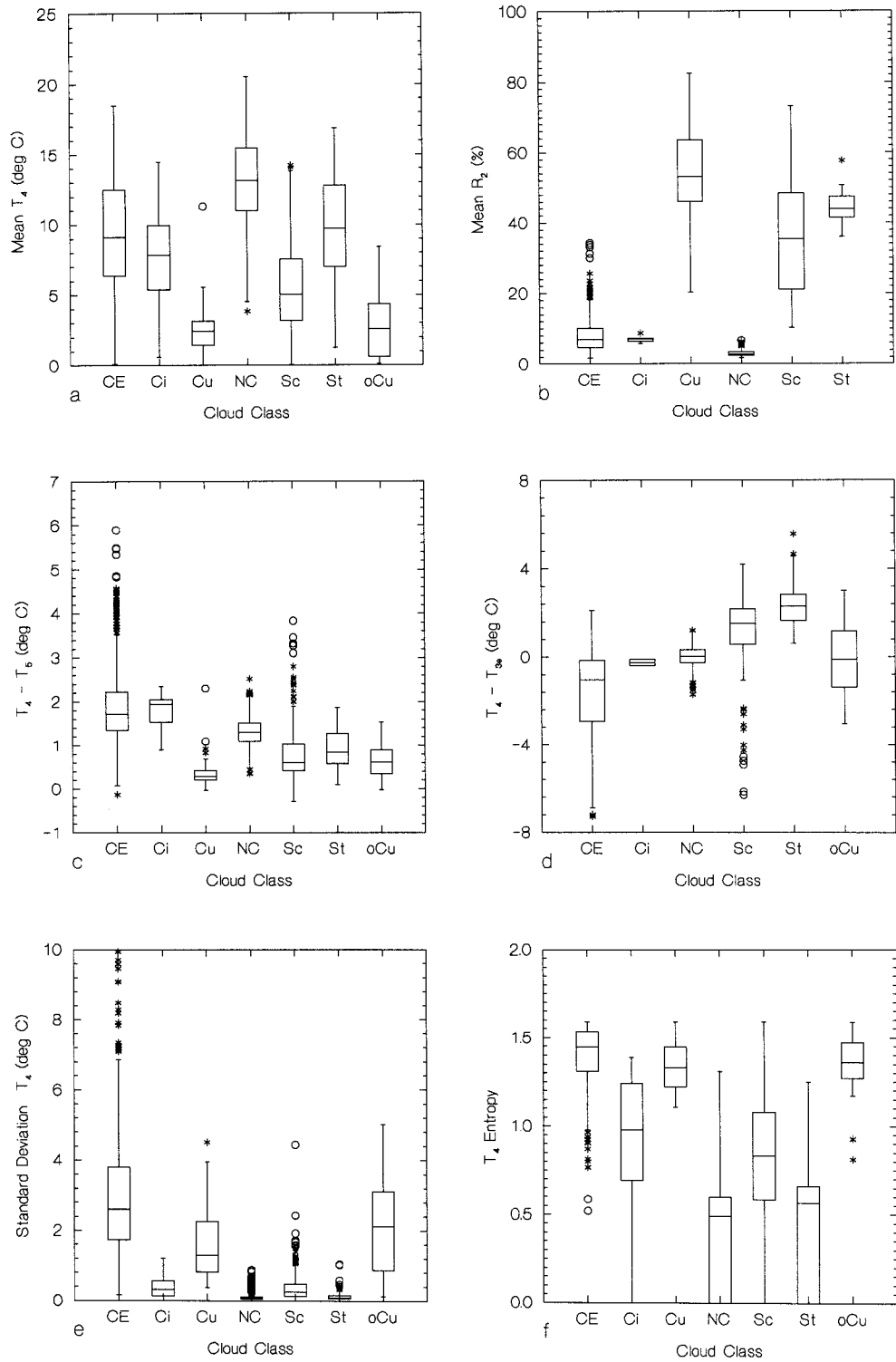


FIG. 2. Tukey box plot characteristics of 3×3 data samples making up the cloud discriminant a priori dataset. (a) Mean T_4 temperature (day and night data), (b) mean R_2 reflectance, (c) mean $(T_4 - T_5)$ difference (day and night), (d) mean $(T_4 - T_3)$ for night data, (e) T_4 standard deviation (day and night), and (f) T_4 isotropic entropy. Here CE, Ci, Cu, NC, Sc, St, and oCu refer to cloud edge, cirrus, cumulus, no-cloud, stratocumulus, stratus, and open cell cumulus, respectively. The horizontal line in the “box” represents the sample median and the upper and lower ends of the box are the hinges or medians of the remaining halves.

TABLE 3. Nighttime Bayesian cloud detection discriminant model skill. The features are calculated on 3×3 IFOV tiles; the sample size is 1598 (of which 1031 are clear, and 567 are cloudy or mixed). Samples were classified to the class with the greatest a posteriori probability. (POD is the probability of detection, FAR is the false alarm rate, and Kuipers' is the Kuipers' performance index.)

Model	Feature vector elements	POD (cloud)	POD (clear)	FAR (cloud)	FAR (clear)	Kuipers'
1	T_4	0.67	0.84	0.31	0.18	0.508
2	$T_4, \sigma(T_4)$	0.70	0.97	0.06	0.14	0.678
3	$T_4, (T_4 - T_5)$	0.73	0.86	0.26	0.15	0.592
4	$T_4, (T_4 - T_5), \sigma(T_4)$	0.73	0.98	0.05	0.13	0.711
5	$T_4, (T_4 - T_5), (T_4 - T_3)$	0.90	0.93	0.13	0.06	0.828
6	$T_4, (T_4 - T_5), (T_4 - T_3), \sigma(T_4)$	0.94	0.98	0.04	0.03	0.917
7	$T_4, (T_4 - T_5), (T_4 - T_3), \sigma(T_4), \sigma(T_4 - T_5)$	0.90	0.98	0.04	0.05	0.882
8	$T_4, (T_4 - T_5), (T_4 - T_3), \text{entropy}(T_4)$	0.97	0.97	0.06	0.02	0.936

servations of one of the m classes, while the FAR is related to the accuracy of the classification with respect to the number of times the event is selected by the discriminant function classifier.

The results for night models are presented first. Models 1, 3, and 5 indicate the cloud/clear detection skill that may derived from consideration of radiative features alone. The importance of the $3.7\text{-}\mu\text{m}$ channel is very evident, with a significant increase in the POD for both clear and cloudy (including mixed) samples and, relative to model 3 (which excludes $3.7\text{-}\mu\text{m}$ data) halved FARs for both classes. However, when a spatial statistic, the tile standard deviation, is added to any of these models (e.g., models 2, 4, and 6), the skill of each model is enhanced relative to the radiative-only models. Indeed, adding a spatial statistic to the simplest of all models (model 1, the $11\text{-}\mu\text{m}$ temperature) yields a greater improvement in the FAR for cloudy samples (i.e., leads to fewer clear samples being classified as cloudy) than the addition of further radiative information such as the 3.7- and $12\text{-}\mu\text{m}$ temperatures (cf. models 3 and 5). Because spatial gradients in SST are generally less than those in clouds (the exception being stratiform cloud), adding this information to the discriminant model aids the detection of clear samples. However, adding the radiative information from T_3 (i.e., models 5 and 6) does lead to a large improvement in the POD for cloud since it allows low level, generally stratiform water

clouds to be discriminated from the sea surface (cf. models 3 and 5), leading to fewer false clear classifications. Use of the tile range (i.e., tile maximum minus tile minimum) in place of tile standard deviation led to discriminant models having slightly lower (i.e., by 0.02) Kuipers' performance indices. Adding the standard deviation of the difference $T_4 - T_5$ degrades the skill of the cloud detection algorithm (cf. models 6 and 7) through misclassification of cloudy data. This feature is a measure of the variability in the atmospheric water vapor over the tile (which is near zero) and of cloud emissivity effects, which can be large for nonstratiform clouds. Consequently, addition of the standard deviation of the difference $T_4 - T_5$ leads to the misclassification of some tiles covered with stratiform cloud.

Replacing the tile standard deviation feature with the GLD isotropic entropy statistic further improves the skill of the model (cf. models 6 and 8). The FAR for cloud increases because more "clear" samples (i.e., those over high SST spatial gradient regions) are being classified as cloud; however, the FAR for clear samples decreases, indicating that the entropy statistic is of value for improving the detection of cloudy tiles, but makes little difference to the detection of clear tiles. Models utilizing the other GLD features were also tested. The GLD mean (μ) showed similar skill to that of the entropy statistic, but use of the contrast and angular second moment reduced the skill of the cloud mask models.

TABLE 4. Daytime (outside sunglint regions) Bayesian cloud detection discriminant models. The features are calculated on 3×3 IFOV tiles, the sample size is 590 (of which 201 are clear, and 389 are cloudy or mixed). Samples were classified to the class with the greatest a posteriori probability.

Model	Feature vector element	POD (cloud)	POD (clear)	FAR (cloud)	FAR (clear)	Kuipers'
9	R_2	0.65	0.99	0.01	0.41	0.630
10	R_2, T_4	0.72	0.99	0.01	0.36	0.705
11	$R_2, T_4, (T_4 - T_5)$	0.83	0.99	0.01	0.26	0.813
12	$R_2, \sigma(R_2)$	0.94	0.97	0.02	0.10	0.914
13	$R_2, T_4, (T_4 - T_5), \sigma(R_2)$	0.95	0.98	0.01	0.08	0.934
14	$R_1, T_4, (T_4 - T_5), \sigma(R_1)$	0.95	0.99	0.01	0.09	0.936
15	$v_2, T_4, (T_4 - T_5), \sigma(R_1)$	0.96	0.97	0.02	0.07	0.929
16	$R_2, T_4, (T_4 - T_5), \text{entropy}(R_2)$	0.94	0.99	0.01	0.10	0.930
17	$R_2, T_4, (T_4 - T_5), \sigma(R_2), \text{entropy}(T_4)$	0.96	0.99	0.01	0.07	0.951
18	$R_2, T_4, (T_4 - T_5), \sigma(R_2), \sigma(T_4)$	0.99	0.98	0.01	0.03	0.960

TABLE 5. Results for discriminant model 6 [i.e., feature vector $[T_4, (T_4 - T_5), (T_4 - T_3), \sigma(T_4)]$] and differing threshold posterior probability thresholds for a “clear” classification.

Posterior probability	Sample size	POD (cloud)	POD (clear)	FAR (cloud)	FAR (clear)	Kuipers'
0.50	1598	0.938	0.979	0.040	0.034	0.917
0.67	1568 (98%)	0.956	0.983	0.031	0.023	0.940
0.90	1495 (94%)	0.980	0.985	0.029	0.010	0.965
0.95	1432 (90%)	0.988	0.986	0.026	0.006	0.974
0.99	1238 (77%)	0.996	0.983	0.027	0.003	0.979

Almost identical observations can be drawn from the day model results given in Table 4. The inclusion of spatial features leads to significant improvements in the skill of the cloud detection model. Also, there appears to be some advantage to using the $0.9\text{-}\mu\text{m}$ (R_2) reflectance in the discriminant model rather than the $0.6\text{-}\mu\text{m}$ (R_1) measurement—contrary to the suggestion by Gallegos et al. (1993). Likewise, the use of ν_2 , as suggested by Gutman (1992), does not lead to an improvement in the skill of the discriminant model. However, the latter would be of greater value for the detection of land sea boundaries since it is a proxy for a vegetation index—but that is not a problem for SST processing, where the only requirement is that clear IFOVs be detected regardless of whether they are over sea or not.

Perhaps more interestingly, these results indicate that it is possible to specify day and night cloud detection models that show similar skill (cf. models 6 and 13), leading to the important conclusion that an appropriately defined Bayesian cloud detection algorithm should show little day/night bias.

The results in Tables 3 and 4 indicate the skill of discriminant models when samples are classified according to the class with the highest posterior probability, as specified in Eq. (1). However, for the SST retrieval problem, there is a high “cost” associated with misclassification of cloudy samples (i.e., large FARs for the clear class) since these lead (generally) to cold biases in the resulting SST data. A Bayesian cloud detection algorithm allows the use of a threshold posterior probability for clear classifications, thus reducing the incidence of (costly) misclassifications of cloudy samples. Table 5 shows the effects of changing the clear posterior probability threshold from 0.50 to 0.99 for model 6 of Table 3. Overall, the skill of the cloud discriminant model improves when the posterior probability is increased (note that the threshold for cloudy classifications remains at 0.5), leading to fewer clear false alarms (i.e., the FAR for clear classifications is reduced). An alternate understanding is that the probability of detecting a cloudy sample increases with the increasing clear posterior probability threshold. The only downside of this approach is that a fraction of the samples that is nominally clear [i.e., $P(\omega_{\text{clr}}|\mathbf{f}) > 0.5$] does not satisfy the posterior probability threshold. Here these may be relabeled as cloudy samples—although for the purposes

of scoring the contingency tables embodied in the results of Table 5, they are excluded. The second column in Table 5 indicates this effect. Evidently, increasing the posterior probability required for a clear classification to 0.95, leads to only 10% of data being unclassified. In the process, the Kuipers’ performance index increases from 0.917 to 0.974, a significant improvement.

5. Example results

The cloud-masking algorithm discussed above has been applied to *NOAA-11*, *-12* and *-14* data, using fully overlapping 3×3 LAC IFOV tiles in the satellite frame of reference. Although the Bayesian cloud detection algorithm can be applied to all data tiles, in practice, a hierarchy of gross tests are applied first. This approach saves considerable computer processing time. Because there is quasi-periodic noise in the $3.7\text{-}\mu\text{m}$ channel data (Walton 1988), two types of gross tests are applied.

The first of these is applied to each LAC IFOV T_4 (i.e., $11\text{-}\mu\text{m}$) temperature. Any IFOV having an $11\text{-}\mu\text{m}$ temperature less than 0°C is automatically labeled as cloudy. For daytime data, prior to the $11\text{-}\mu\text{m}$ test, a gross test is applied to each IFOV’s R_2 (i.e., $0.9\text{-}\mu\text{m}$) reflectance. The IFOV is labeled cloudy if the reflectance is greater than a threshold determined by the mean reflectance of the training sample’s clear data plus one standard deviation. This approach will discard data from the sunglint region, which may be clear. However, this is not considered serious for the present analysis since in practice the location of this region may be computed, and the R_2 test turned off there. The last gross check applied is to nighttime tile $T_4 - T_3$ mean difference. When this difference is greater than 1.25°C , the center IFOV in the tile is labeled cloudy by default. Tile means are used for this test in order to reduce the effect of the noise in the $3.7\text{-}\mu\text{m}$ data. Any tiles passing these tests (i.e., possibly clear) are then analyzed using the Bayesian discriminant functions. After the Bayesian cloud mask classifier has been applied and SSTs retrieved from the “clear” IFOVs, one last test is applied. The high-resolution SST is compared with a monthly 1° spatial resolution SST climatology derived from weekly Climate Analysis Center SST analyzed fields (Reynolds and Smith 1994). Any IFOVs having an SST that is more than 3°C colder than climatology are relabeled as

cloudy. The only other constraints applied during processing are to limit the scan angle to $\pm 50^\circ$ in order to preserve the spatial characteristics defined in the training sample, and for day data, the solar zenith angle to less than 80° . The latter is required to reduce the impact of cloud shadows, which can cause difficulties when low cloud lies adjacent too but on the antisolar side of high cloud.

a. Two example orbits

Figure 3 shows sections of two passes over the New Zealand region: a daylight pass from *NOAA-14* (orbit 9151) and a nighttime pass from *NOAA-12* (orbit 28438). The remapped images retain the full spatial resolution of the AVHRR LAC data (i.e., 1 km). In both cases warm marine stratus and stratocumulus are evident. Using the approach developed above, cloud masks and SSTs [using the Walton (1988) NLSST algorithm, and current NOAA coefficients] have been computed. The results are shown in Figs. 4 and 5. The daytime cloud mask (Fig. 4a) was calculated using the model 13 discriminant function (in Table 4) and a 0.90 posterior probability threshold for classification to the clear class. The particular test that determined whether a pixel was cloudy or not is also indicated in the figure. In this case 71.8% of IFOVs failed the R_2 or T_4 (i.e., 0.9 or 11 μm) gross tests (colored pink and green in the figure). Of the remaining 28.2% of IFOVs, 47.7% of these were determined to be cloudy by the Bayesian discriminant classifier (colored white). Evidently, the Bayesian classifier is especially useful in identifying cloud edges and regions of stratocumulus (e.g., north of 34°S and east of 175°E). Only 541, or 0.2% of the resultant “clear” IFOVs were relabeled cloudy by the SST climatology test (colored red). Accordingly 14.8% of all IFOVs were determined to be cloud free. The NLSSTs computed from the clear IFOVs are shown in Fig. 4b. With the exception of a few apparently cold SSTs on the northwestern edge (near 36.5°S , 177°E) of the cloud feature over East Cape (37.6°S , 178.3°E), there is no obvious cloud contamination in the SST data.

The nighttime cloud mask, estimated with discriminant model 6 (Table 3) and a 0.90 clear posterior probability threshold, is shown in Fig. 5a. In this case the $\pm 50^\circ$ scan angle limit is breached on the southwestern side of the image—with the result that a cloud mask is not estimated in that region. The T_4 (11 μm) gross test

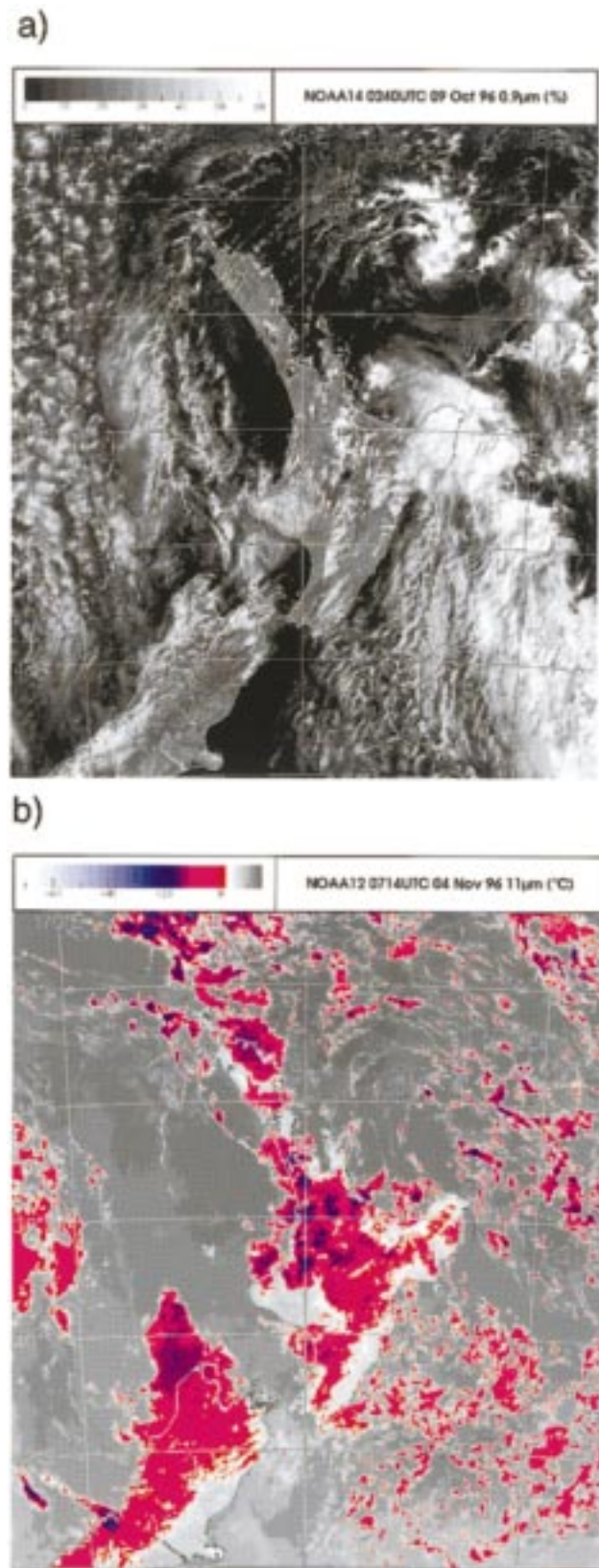
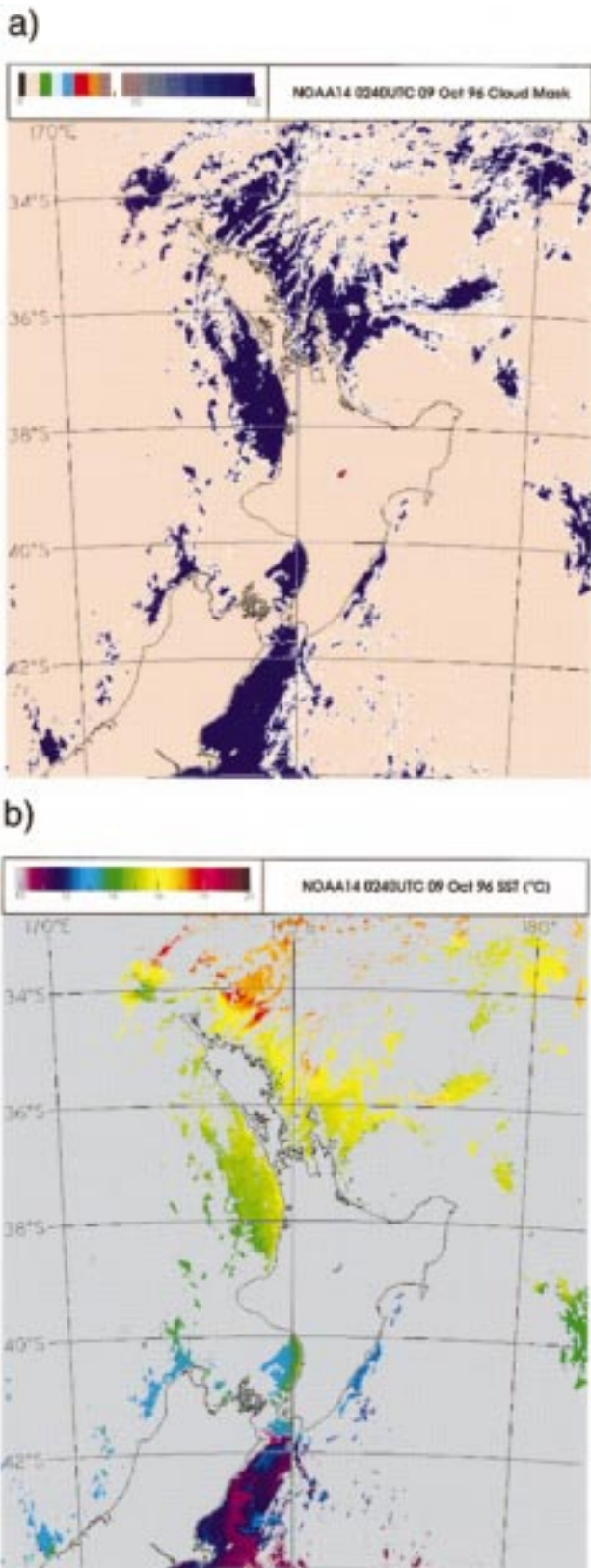


FIG. 3. Remapped AVHRR imagery for sections of two passes over the New Zealand region. The first is *NOAA-14* orbit 9151, 0238–0242 UTC 9 October 1996 and (a) the channel 2 (0.9 μm) solar-corrected reflectance. The second is *NOAA-12* orbit 28438, 0712–0716 UTC 4 November 1996 and (b) the channel 4 (11 μm) thermal image. These images are 1024×1152 pixels. For the thermal channel, temperatures warmer than 0°C are shown in gray shades, while colder temperatures are rendered in color.



(green) labeled 35.8% of the IFOVs cloudy, and the stratus [i.e., $(T_4 - T_3)$] test (light and midblue) found another 0.1% cloudy IFOVs. Of the 64.1% of IFOVs passing the gross tests, 56.8% of these were determined to be cloudy by the Bayesian classifier, leaving 27.7% of all IFOVs clear, when the 12.3% of IFOVs beyond the scan angle limits and the 0.1% of IFOVs failed by the SST climatology test (red) are included. With regard to the SST climatology test, no attempt is made to mask out land data—with the result that nearly all pixels re-labeled cloudy by the SST climatology test actually lie over “land.” Again the Bayesian classifier is of most value in identifying cloud edges and regions of isolated cumulus and stratocumulus. The resulting NLSSTs, estimated from the clear IFOVs are given in Fig. 5b. Like the daytime results there appears to be little to no evidence of cloud contamination in the SST data. The cold features in the vicinity of Cook Strait (41.5°S , 174.5°E) and east of Banks Peninsular (42.5°S , 174°E) are recognized oceanographic features (Heath 1985).

Cloud masks were also computed for these examples using differing posterior probability clear thresholds. Summary results are shown in Table 6. Increasing the posterior probability threshold for clear classifications from 0.50 to 0.90 reduces the yield of “clear” IFOVs available for SST retrieval by 6%. Remarkably this is the same value as obtained from the training sample (Table 5), perhaps indicating that the skill of the cloud mask discriminant algorithm increases in a similar way to that indicated in Table 5 and that the training sample satisfies the basic assumptions required to calculate the Bayesian discriminant functions.

b. Time-composited SSTs

While a full analysis of SST data products retrieved from AVHRR data using the Bayesian cloud mask lies outside the scope of this paper, the general characteristics of the resulting SST data have been determined using a 30-day sample (from 11 September 1996 to 10 October 1996) of NOAA-12 and -14 LAC data for the New Zealand region. Cloud masks for each of the more than 250 orbits were estimated using a clear IFOV posterior probability threshold of 0.90 and models 6 (Table 3) and 13 (Table 4), respectively, for day and night data. SSTs were retrieved using the NLSST algorithm of Wal-

←

FIG. 4. (a) Cloud mask for the NOAA-14 daytime pass of Fig. 3a. The threshold posterior probability for clear data was set at 0.90. Pixels colored pink failed the gross R_2 test, those colored green failed the gross T_4 test, and red pixels failed the SST climatology test. Light and midblue indicate those pixels that failed the $T_4 - T_3$ stratus test (only used at night, see Fig. 5). White pixels were classified cloudy by the Bayesian classifier, and the blue shades indicate the posterior probability (from 0.50 to 1.0) that the pixel is clear. Orange and mauve are not used. (b) The resulting NLSST retrieval for this orbit. In this image the cloud mask is colored gray.

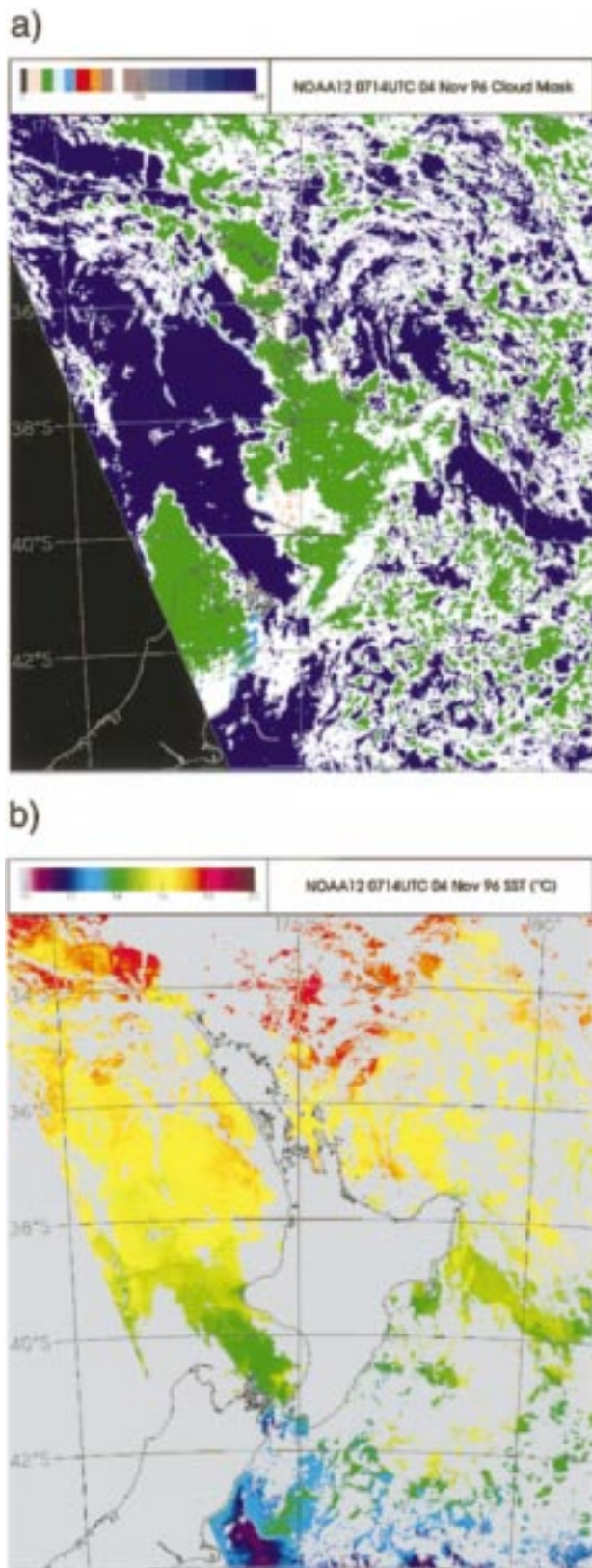


FIG. 5. Shows the same results as in Fig 4 but for the NOAA-12 nighttime pass of Fig. 3. The color scale is explained in the caption to Fig. 4. (a) The cloud mask image and (b) the image of NLSST retrievals.

TABLE 6. Summary statistics for two posterior probability thresholds. For each orbit the number of IFOV passing the gross checks (apart from the SST climatology test) are independent of the posterior probability (pp).

	Daytime orbit		Nighttime orbit	
	pp = 0.50	pp = 0.90	pp = 0.50	pp = 0.90
Percent of IFOVS failed by Bayes test	45.4	48.7	53.3	56.8
Percent of IFOVS failed by SST (gross) test	0.1	0.1	0.7	0.1
Total number of clear IFOV (%)	15.3	14.4	29.5	27.7

ton (1988), with a first-guess SST provided by the split-window multichannel SST (MCSST). For the reasons stated above, daytime data lying in regions where the solar zenith angle was greater than 80° or where data from scan angles were outside the range $\pm 50^\circ$ were not considered.

After retrieval, the SST data were remapped onto a 3504×3504 “pixel” Lambert conformal map projection at 1.1-km resolution, and successive pass sets mosaiced. Minor navigational errors in the resulting data were automatically corrected by analyzing the correlation between isotropic (Schowengerdt 1983) gradients in the SST data and prominent landmarks around the New Zealand, Australian (east), and New Caledonia coasts. The high-resolution SST data were time-composited for the period, and the mean and standard deviation data computed (Fig. 6). The maximum sample size at any “pixel” is 108, for the number of sets of passes processed from the 30-day period, while the minimum was set at 4.

The general characteristics of the oceanography in the New Zealand region are captured in Fig. 6. The East Australian Current is evident, with its retroflexion near $35^\circ\text{--}37^\circ\text{S}$, $150^\circ\text{--}155^\circ\text{E}$, likewise the Tasman Front and associated semipermanent eddies trending east northeast over the Lord Howe and Norfolk Ridges (Heath 1985). The East Auckland Current extends down the northeast coast of the North Island of New Zealand to the subtropical front over Chatham Rise (indicated by the platform in the 1000-m bathymetry extending east from the South Island to about 175°W). The Southland Front, east of New Zealand, can be seen above the shelf break indicated by the 1000-m bathymetry. Cold sub-Antarctic waters lie to the east of the Southland Front and south of Chatham Rise.

The standard deviation plot (Fig. 6b), equivalent to Jones et al.’s (1996) “variability plots,” indicates little evidence of cloud contamination in the composited SSTs. For most of the mapped area the standard deviation is less than 0.8°C . Larger standard deviations are, in general, associated with regions of higher oceanographic variability, such as in the East Australian Current and its retroflexion, along the Tasman Front and near the subtropical front east of New Zealand. Similarly

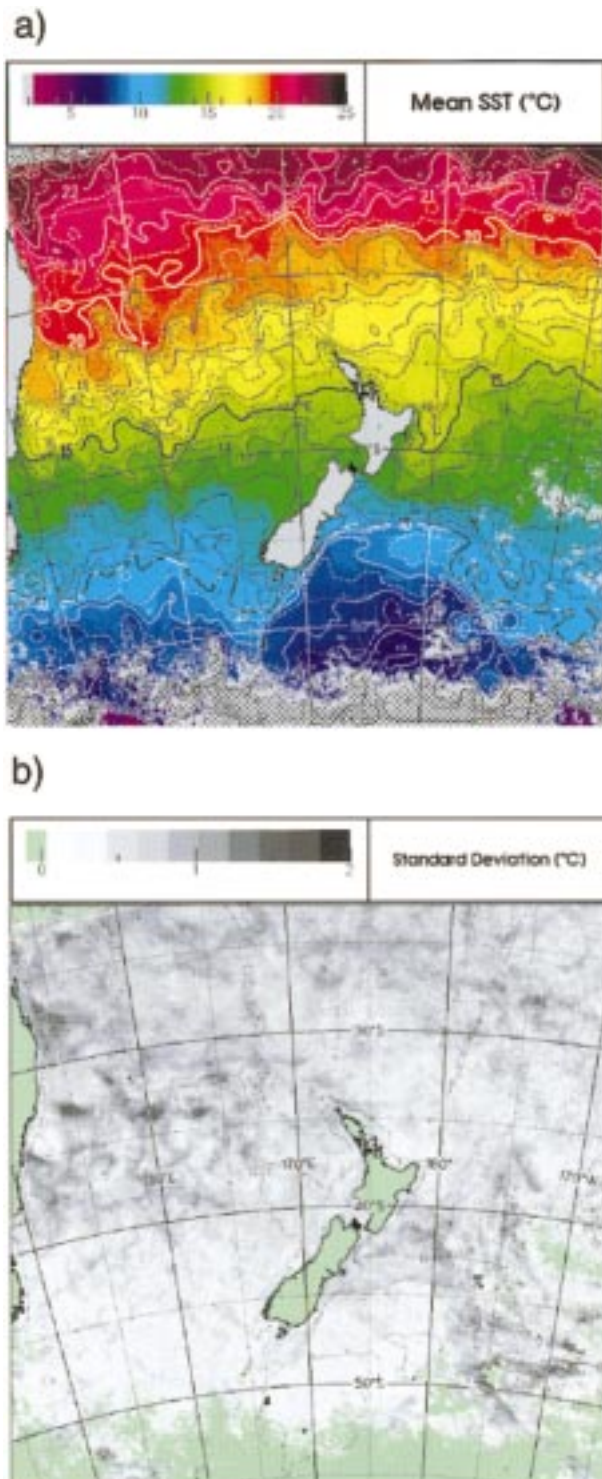


FIG. 6. (a) Mean and (b) std dev plots of NLSST data composited from 11 September 1996 to 10 October 1996. In (a) missing data are indicated in gray but green is used in (b). The maximum sample size at any data point is 108 and the minimum is 4. The SST “image” data have not been spatially filtered. The isotherm analysis was computed using the Cressman algorithm on a 16-km resolution grid.

there are regions of higher variability in tropical waters, for example, in the tropical convergence zone near 25°S, 170° to 180°E (Heath 1985). In the sub-Antarctic waters south and west of New Zealand, a region often covered in cloud, the variability is very low, mostly less than 0.5°C. However, east of 180°E and south of 45°S, the composited SST data show large variances. This region is known to exhibit complex oceanography (Bryden and Heath 1985), accompanied by substantial eddy activity, which is conjectured to be the source of the increased variance in the SST fields. However, further analysis of time series data will be required in order to prove this conjecture and demonstrate that the increased variance is not due to residual cloud contamination in the retrieved SSTs.

Separating the data into day and night samples, the standard deviations and their day–night differences were computed. These are shown in Fig. 7. Three points are apparent from consideration of this figure. First, the daytime standard deviations (Fig. 7a) have a greater range than those for the nighttime data (Fig. 7b). Evidently there can be little cloud contamination in the nighttime data since, as Jones et al. (1996) found in the ATSR ASST data product, residual cloud increases the variance in the composited data. The increased variability in the daytime data is likely a result of shallow surface heating due to insolation (i.e., the diurnal thermocline) together with variability in the surface wind fields over the compositing period. Second, many regions of high variability (i.e., greater than 0.8°C) can be found at the same locations in both datasets, for example, over the East Australian Current and in the Tasman Front. Last, when these two datasets are subtracted (i.e., Fig. 7c), it is clear that there is little day/night bias. Indeed, the mean difference in the standard deviations (day–night) is just +0.1°C, and the standard deviation of the difference is 0.25°C. A histogram of the differences (not shown) indicates that they are Gaussian and that there are no outliers. Further, the regions where this difference is positive tend to be associated with regions having increased daytime variability rather than nighttime variability. Finally, the day minus night mean SST differences were computed. These differences (not shown) indicate a 0.3°C warm bias in the daytime data and a standard deviation of 0.3°C, which is consistent with the idea of a diurnal thermocline (e.g., Ostapoff and Worthem 1974). There was no evidence of latitudinal or longitudinal trends in this bias.

All of these observations are consistent with the idea that the Bayesian cloud detection algorithm developed here shows little day/night cloud contamination bias.

6. Conclusions

A new operational AVHRR Bayesian cloud detection algorithm for SST retrieval has been presented and its properties demonstrated via analysis of model contin-

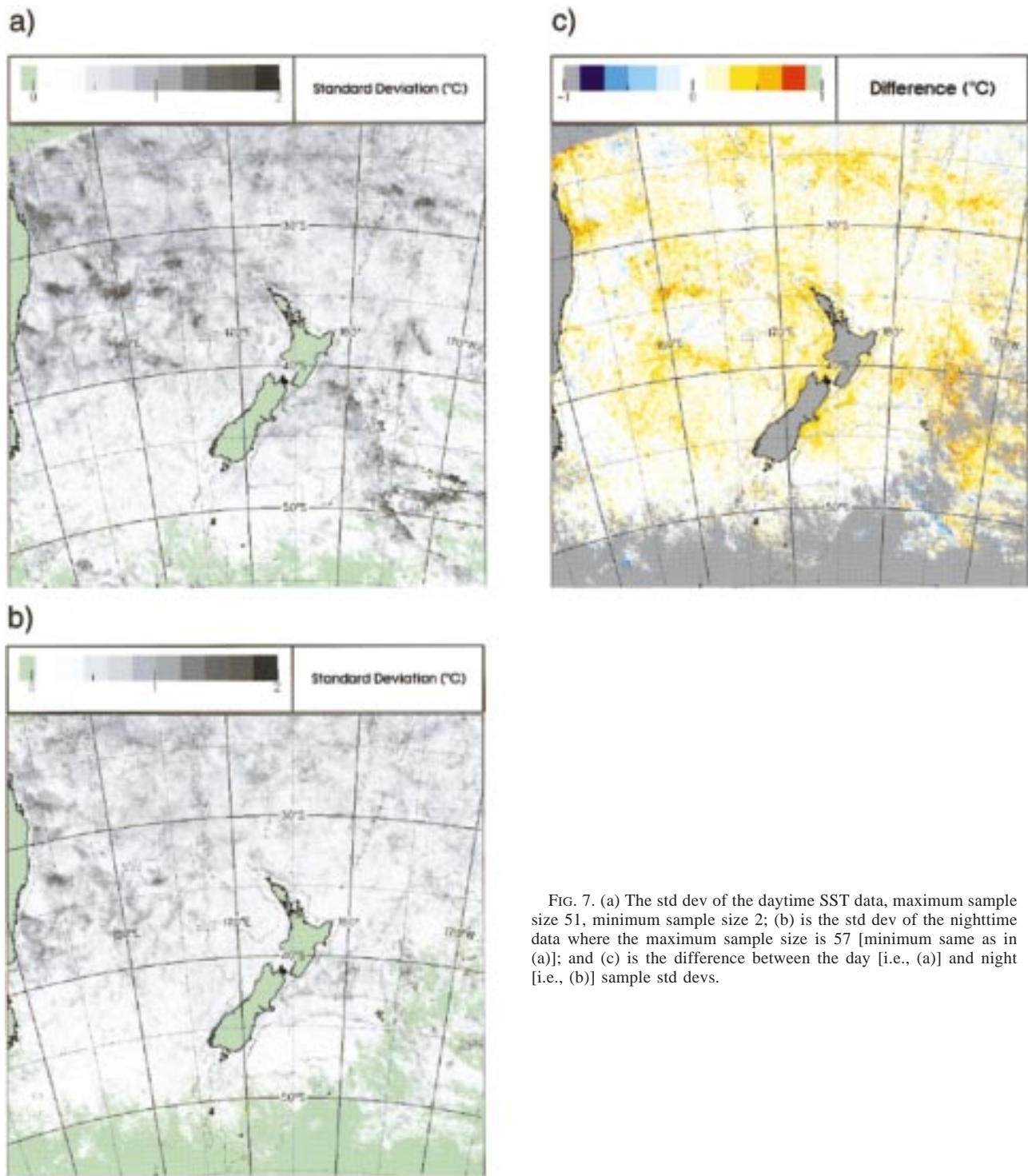


FIG. 7. (a) The std dev of the daytime SST data, maximum sample size 51, minimum sample size 2; (b) is the std dev of the nighttime data where the maximum sample size is 57 [minimum same as in (a)]; and (c) is the difference between the day [i.e., (a)] and night [i.e., (b)] sample std devs.

gency tables, sample orbits, and analysis of an ensemble of more than 250 LAC orbits. This approach differs from those used operationally elsewhere, in that it treats the identification of cloudy or clear IFOVs as a Bayesian classification problem. As a result all measurements,

both radiative and spatial, are used simultaneously in order to determine the probability that an IFOV belongs to the clear and cloudy states. This method obviates the need to establish the somewhat arbitrary static and/or dynamic thresholds used by hierarchical cloud-masking

algorithms, instead replacing that requirement with the need to establish appropriate mean and covariance matrices of the discriminant features of the two states. Here this problem has been resolved by considering the properties of various warm ($T_4 > 0^\circ\text{C}$) cloud types (i.e., Cu, Sc, St, Ci, and cloud edges) as well as cloud-free data tiles, at 3×3 IFOV (LAC) resolution. The benefit of this approach is that the radiative and spatial characteristics of each of the cloud types can be considered when formulating the cloud detection algorithm. Another feature of this approach is that the “skill” of different Bayesian discriminant models can be analyzed by applying each model to the training data and comparing the results. This is an appropriate approach if the training sample includes a large number of independent observations, which is satisfied here. The applicability of such results to real orbital data is, of course, dependent upon whether the training sample data span all situations encountered. While this condition is not assured, the sampling strategy used should result in training data that closely conform to this requirement. Using this approach, analysis of the skill of different Bayesian cloud detection models clearly demonstrates the critical importance of spatial (or textural) information in the formulation of any cloud detection model. While it is clear that radiative information is critical to the detection of some cloud classes such as stratus during nighttime hours, the spatial information significantly improves the detectability of clear IFOVs, regardless of the selection of spectral features. For the nighttime models tested, experiments with tile standard deviation and GLD measures of spatial information suggest that the GLD entropy statistic leads to the most skillful cloud detection model. However, the next most skillful model is that where the tile entropy is replaced by the standard deviation, a feature that can be computed at 50% of the computational cost incurred by the GLD calculation. For daytime models, there was no advantage in using entropy over the standard deviation spatial statistic. With regard to the visible channel radiative information, and contrary to Gallegos et al.’s (1993) findings, there appears to be no particular advantage in using the $0.6\text{-}\mu\text{m}$ (i.e., R_1) over the $0.9\text{-}\mu\text{m}$ (i.e., R_2) measurements. Use of a Saunders and Kriebel (1988)–like Q measurement (ν_2 , i.e., $R_2/R_1 - 1$) leads to a small, probably insignificant, degradation in the performance of the cloud detection model. Apparently, day and night Bayesian cloud detection models of nearly identical skill can be constructed.

While the Bayesian method bypasses the need for a hierarchical approach, gross tests are employed in order to reduce computational overheads. Results from two randomly chosen sample orbits demonstrate that the Bayesian algorithm is generally applied to between about 20% and 60% of all IFOVs. The actual percentage is, of course, dependent upon the meteorological situation and whether its a day or night orbit. Evidently the new algorithm is especially useful in identifying cloud

edges, scattered cumulus, and stratocumulus clouds. There is also some evidence that the training sample indeed spans the situations encountered with orbital data since adjusting the posterior probability threshold for clear IFOVs (on the orbital data) increased the percentage of cloudy IFOVs by the same percentage as that expected from the training sample. These results also show that SSTs retrieved from the cloud-free areas only very infrequently fail the monthly SST climatology quality control test, a further indication that there is little evidence of cloud contamination in the retrieved SST data.

The Bayesian cloud-clearing algorithm has also been applied to a contiguous 30-day sample of *NOAA-12* and *-14* (LAC) data from the New Zealand region. The resulting SST data were composited and mean and standard deviation statistics computed. When the day- and nighttime data from both spacecraft are combined, these statistics appear to reproduce the major oceanographic features of the New Zealand region, both in terms of the mean state and the centers of variability. There is little evidence of cloud contamination in either of these statistics. When the data are divided into day and night samples, the standard deviation data indicate that the daytime SSTs have the larger variability with a positive bias of 0.1°C (relative to night) and that the standard deviation of the difference is just 0.25°C . Locations showing larger standard deviations are generally found in the same locations on both the day and night samples and are associated with major oceanographic features. All of these observations suggest that there is little day/night bias in the cloud detection algorithm, confirming the results obtained from analysis of the performance of the discriminant models when applied to the training sample data.

Each of these results indicate that the Bayesian cloud discrimination algorithm presented here can identify the presence of cloud in AVHRR LAC resolution data observed during both night- and daytime with similar skill. Further, it is likely that the resulting SST data may be used to specify the variability of the SST field with some confidence. The algorithm is not computationally expensive, taking just 10–15-min CPU time to process one entire (LAC) orbit on a low-end Digital Equipment Corporation Alpha workstation (DEC 3000 Model 400). Further, although this paper uses AVHRR data to demonstrate the principles, there is no reason why the approach could not be used with other instruments.

In future research, LAC data from 1 January 1993 to the present will be reprocessed (at full resolution) using the cloud and SST algorithms discussed here. At the same time a Lagrangian archive of drifting buoy observations will be constructed together with the collocated AVHRR data, cloud mask, and retrieved SSTs for $1^\circ \times 1^\circ$ areas around each buoy location. A primary use of this high-resolution archive will be to allow evaluation of the relationship between cotemporal and co-spatial fisheries catch and effort data and oceanographic

thermal features. It will also allow a validated, high-resolution, SST climatology to be developed for the New Zealand region (i.e., the area shown in Fig. 6).

Acknowledgments. This research was supported by the New Zealand Foundation for Research Science and Technology under Contract CO1503.

REFERENCES

- Barton, I. J., A. J. Prata, and R. P. Cechet, 1995: Validation of the ATSR in Australian waters. *J. Atmos. Oceanic Technol.*, **12**, 290–300.
- Bryden, H. L., and R. A. Heath, 1985: Energetic eddies at the northern edge of the Antarctic Circumpolar Current in the Southwest Pacific. *Progress in Oceanography*, Vol. 14, Pergamon Press, 65–87.
- Chou, J., R. C. Weger, J. M. Lightenberg, K.-S. Kuo, R. M. Welch, and P. Breeden, 1994: Segmentation of polar scenes using multispectral texture measures and morphological filtering. *Int. J. Remote Sens.*, **15**, 1019–1036.
- Delderfield, J., D. T. Llewellyn-Jones, R. Bernard, Y. de Javel, E. J. Williamson, I. Mason, D. R. Pick, and I. J. Barton, 1986: The Along Track Scanning Radiometer (ATSR) for ERS-1. *Proc. SPIE-Int. Soc. Opt. Eng.*, **589**, 114–120.
- d'Entremont, R. P., 1986: Low- and midlevel cloud analysis using nighttime multispectral imagery. *J. Climate Appl. Meteor.*, **25**, 1853–1869.
- Ebert, E., 1987: A pattern recognition technique for distinguishing surface and cloud types in polar regions. *J. Climate Appl. Meteor.*, **26**, 1412–1427.
- Forrester, T. N., T. H. Guymer, and P. G. Challenor, 1993: Preliminary validation of ATSR sea surface temperatures near the Faeroes. *Proc. First ERS-1 Symp.—Space at the Service of Our Environment*, Cannes, France, European Space Agency, 807–811.
- Gallaudet, T. C., and J. J. Simpson, 1991: Automated cloud screening of AVHRR imagery using split-and-merge clustering. *Remote Sens. Environ.*, **38**, 97–121.
- Gallegos, S. C., J. D. Hawkins, and C. F. Cheng, 1993: A new automated method of cloud masking for Advanced Very High Resolution Radiometer full-resolution data over the ocean. *J. Geophys. Res.*, **98C**, 8505–8516.
- Gutman, G. G., 1992: Satellite daytime image classification for global studies of Earth's surface parameters from polar orbiters. *Int. J. Remote Sens.*, **13**, 209–234.
- Haralick, R. M., K. Shanmugam, and Its'hak Dinstein, 1973: Textural features for image classification. *IEEE Trans. Syst. Man Cybern.*, **SMC-3**, 610–621.
- Harries, J. E., D. T. Llewellyn-Jones, P. J. Minnett, R. W. Saunders, and A. M. Závody, 1986: Observations of sea surface temperature for climate research. *Philos. Trans. Roy. Soc. London*, **309A**, 381–395.
- Heath, R. A., 1985: A review of the physical oceanography of the seas around New Zealand—1982. *New Zealand J. Mar. Freshwater Res.*, **19**, 79–124.
- Hepplewhite, C. L., 1989: Remote observation of the sea surface and atmosphere: The oceanic skin effect. *Int. J. Remote Sens.*, **10**, 801–810.
- Hunt, G. E., 1973: Radiative properties of terrestrial clouds at visible and infrared thermal window wavelengths. *Quart. J. Roy. Meteor. Soc.*, **99**, 346–369.
- Inoue, T., 1985: On the temperature and effective emissivity determination of semi-transparent cirrus clouds by bi-spectral measurements in the 10- μ m window region. *J. Meteor. Soc. Japan*, **63**, 88–99.
- Jones, M. S., M. A. Saunders, and T. H. Guymer, 1996: Reducing cloud contamination in ATSR averaged sea surface temperature data. *J. Atmos. Oceanic Technol.*, **13**, 492–506.
- Karlsson, K.-G., 1994: Satellite-estimated cloudiness from NOAA AVHRR data in the Nordic area during 1993. SMHI Reports, Meteorology and Climatology, No. 66, 51 pp. [Available from Swedish Meteorological and Hydrological Inst., Box 923, S-601 76 Norrköping, Sweden.]
- Kidson, J., J. Couper, and M. Uddstrom, 1992: Interactive data visualization at the New Zealand Meteorological Service. Preprints, *Eighth Int. Conf. Interactive Information and Processing Systems for Meteorology, Oceanography, and Hydrology*, Atlanta, GA, Amer. Meteor. Soc., 132–134.
- McClain, E. P., 1989: Global sea surface temperatures and cloud clearing for aerosol optical depth estimates. *Int. J. Remote Sens.*, **10**, 763–769.
- , W. G. Pichel, and C. C. Walton, 1985: Comparative performance of AVHRR-based multichannel sea surface temperatures. *J. Geophys. Res.*, **90**, 11 587–11 601.
- , C. C. Walton, and L. L. Stowe, 1990: CLAVR cloud/clear algorithms and non-linear atmospheric corrections for multichannel sea surface temperatures. Preprints, *Fifth Conf. on Satellite Meteorology and Oceanography*, London, United Kingdom, Amer. Meteor. Soc., 133–138.
- McMillin, L. M., and D. S. Crosby, 1984: Theory and validation of the multiple window sea surface temperature technique. *J. Geophys. Res.*, **89C**, 3655–3661.
- Minnett, P. J., 1991: Consequences of sea surface temperature variability on the validation and applications of satellite measurements. *J. Geophys. Res.*, **96C**, 18 475–18 489.
- Murphy, A. H., and R. W. Katz, 1985: *Probability, Statistics and Decision-Making in the Atmospheric Sciences*. Westview Press, 545 pp.
- Ostapoff, F., and S. Worthem, 1974: The intradiurnal temperature variation in the upper ocean layer. *J. Phys. Oceanogr.*, **4**, 601–612.
- Parikh, J., 1977: A comparative study of cloud classification techniques. *Remote Sens. Environ.*, **6**, 67–81.
- Planet, W. G., Ed., 1988: Data extraction and calibration of TIROS-N/NOAA Radiometers. NOAA Tech. Memo. NNESS 107-Rev 1, 58 pp. [Available from U.S. DOC, NOAA/NESDIS, Washington, DC 20233.]
- Reynolds, R. W., and T. M. Smith, 1994: Improved global sea surface temperature analyses using optimum interpolation. *J. Climate*, **7**, 929–948.
- Saunders, R. W., and K. T. Kriebel, 1988: An improved method for detecting clear sky and cloudy radiances from AVHRR data. *Int. J. Remote Sens.*, **9**, 123–150.
- Schowengerdt, R. A., 1983: *Techniques for Image Processing and Classification in Remote Sensing*. Academic Press, 249 pp.
- Simpson, J. J., and C. Humphrey, 1990: An automated cloud screening algorithm for daytime Advanced Very High Resolution Radiometer imagery. *J. Geophys. Res.*, **95**, 13 459–13 481.
- , and J. I. Gobat, 1995: Improved cloud detection in GOES scenes over the ocean. *Remote Sens. Environ.*, **52**, 79–94.
- Smith, W. L., and Coauthors, 1996: Observations of the infrared radiative properties of the ocean—Implications for the measurement of sea surface temperature via satellite remote sensing. *Bull. Amer. Meteor. Soc.*, **77**, 41–51.
- Stowe, L. L., E. P. McClain, R. Carey, P. Pellegrino, G. G. Gutman, P. Davis, C. Long, and S. Hart, 1991: Global distribution of cloud cover derived from NOAA/AVHRR operational satellite data. *Adv. Space Res.*, **11** (3), 51–54.
- Tovinkere, V. R., M. Penaloza, A. Logar, J. Lee, R. C. Weger, T. A. Berendes, and R. M. Welch, 1993: An intercomparison of artificial intelligence approaches for polar scene identification. *J. Geophys. Res.*, **98**, 5001–5016.

- Uddstrom, M. J., and W. R. Gray, 1996: Satellite cloud classification and rain-rate estimation using multispectral radiances and measures of spatial texture. *J. Appl. Meteor.*, **35**, 839–858.
- Walton, C. C., 1988: Nonlinear multichannel algorithms for estimating sea surface temperature with AVHRR satellite data. *J. Appl. Meteor.*, **27**, 115–124.
- Welch, R. M., S. K. Sengupta, and K. S. Kuo, 1988: Marine stratocumulus cloud fields off the coast of southern California observed using LANDSAT imagery. Part II: Textural analysis. *J. Appl. Meteor.*, **27**, 363–378.
- Weszka, J. S., C. R. Dyer, and A. Rosenfeld, 1976: A comparative study of texture measures for terrain classification. *IEEE Trans. Syst. Man. Cybern.*, **SMC-6**, 269–285.
- Wick, A. A., W. J. Emery, and P. Schluessel, 1992: A comprehensive comparison between satellite-measured skin and multichannel sea surface temperature. *J. Geophys. Res.*, **97C**, 5 569–5 595.