A New Statistical Modelling Approach to Ocean Front Detection from SST Satellite Images

Jo Hopkins[1] * Peter Challenor[2] and Andrew G.P. Shaw[2]

[1] *Proudman Oceanographic Laboratory, Liverpool, UK*

[2] *National Oceanography Centre, Southampton, UK*

Submitted December 2008

* *Corresponding author address:* Proudman Oceanographic Laboratory, Joseph Proudman Building, 6 Brownlow Street, Liverpool, L3 5DA, UK.

E-mail: jeh200@pol.ac.uk, Tel: +44(0)151 795 4859

ABSTRACT

Ocean fronts are narrow zones of intense dynamic activity that play an important role in global ocean-atmosphere interactions. Owing to their highly variable nature, both in space and time, they are notoriously difficult features to adequately sample using traditional *in-situ* techniques. In this paper we propose a new statistical modelling approach to detecting and monitoring ocean fronts from AVHRR SST satellite images that builds on the 'front following' algorithm of Shaw and Vennell (2000). Weighted local likelihood is used to provide a smooth, non-parametric description of spatial variations in the position, mean temperature, width and temperature change of an individual front within an image. Weightings are provided by a Gaussian kernel function whose width is automatically determined by likelihood cross-validation. The statistical model fitting approach allows estimation of the uncertainty of each parameter to be quantified, a capability not possessed by other techniques. The algorithm is shown to be robust to noise and missing data in an image, problems that hamper many of the existing front detection schemes. The approach is general and could be used with other remotely sensed data sets, model output or data assimilation products.

1. Introduction

The advent of Earth observation satellites in the 1970's has revolutionized the oceanographer's ability to study oceanic structures such as ocean fronts; the narrow regions marking the transition between two different water masses. Fronts are characterized by intense horizontal gradients in sea surface temperature, salinity, biological and/or chemical properties. They are highly dynamic zones continually changing through both space and time as the adjacent water masses are modified by across frontal mixing, vertical transport and air-sea interaction. Ocean fronts are climatologically important regions and play a substantial role in global ocean-atmosphere interactions. Of particular significance is the circumglobal frontal system of the Southern Ocean where intermediate water masses are formed (Spall 1995; Garabato et al. 2001), heat, salt, nutrients and momentum are redistributed, biological activity is stimulated (Moore and Abbott 2000), and atmospheric carbon dioxide is absorbed (Murphy et al. 1991; Currie and Hunter 1998, 1999). Variability of the strength and location of sea surface temperature fronts is also important to the coupling of winds and upper-ocean processes (O'Neill et al. 2003; Chelton et al. 2004).

The availability of a large and rapidly expanding data set of remotely sensed sea surface temperature (SST), altimetry and ocean colour has fueled a growing interest and demand for objective and automatic techniques to detect and monitor fronts. Accurate knowledge of frontal zones and how they change both temporally and spatially is important to many organizations worldwide with such diverse tasks as climate variability and monitoring, operational weather and ocean forcasting, validating ocean and atmospheric models, ecosystem assessment and fisheries research.

Fronts exhibit nonlinear flows and processes on a range of different temporal and spatial scales. Successful detection and monitoring therefore is a nontrivial problem. Capturing frontal features as they grow, merge, split, shrink and disappear is a considerable challenge. Complicating the task are the resolution limitations imposed by the instruments being used to image these phenomena. Noise and missing data owing to sensor inaccuracies, thermal calibration, atmospheric correction and cloud further add to the challenge of accurate detection.

Finding ocean fronts in remotely sensed imagery is a problem that has been approached in four different ways:

1. derivative based edge detection,
2. gradient magnitude thresholds,
3. statistical/probabilistic edge discrimination and classification, and,
4. surface fitting.

Derivative based edge detectors rely on locating gradient discontinuities in SST images that mark a sharp transition from cold to warm water (or *vice versa*). Examples include the Prewitt, Sobel, Kirsh, Roberts and LoG gradient operators (Simpson 1990). These methods, based on image processing techniques, however are generally not well suited to oceanographic remote sensing applications (Holyer and Peckinpaugh 1989); they struggle to discriminate between weak, small scale features and noise. Often a pre-filtering stage is required (e.g. Canny operator), but this smoothing blurs features and sharp gradients and makes subsequent edge detection more difficult. Setting an SST gradient magnitude threshold is a simple alternative way in which fronts may be mapped (Belkin and Gordon 1996; Kostianoy et al. 2004; Moore et al. 1997, 1999; Dong et al. 2006).

More sophisticated techniques involve the classification of pixels or windows of data and some form of statistical or probabilistic analysis to determine the presence of a front. Algorithms that fall into this category may be further subdivided: distribution diversity (entropy) based methods (Vazquez et al. 1999; Shimada et al. 2005); histogram analysis (Cayula and Cornillon 1992, 1995; Marcello et al. 2005); a clustering based approach (Holyer and Peckinpaugh 1989); and an examination of the moments combined with *a-priori* knowledge of the region (Gerson et al. 1979; Coulter 1983). Other edge detectors tested in an oceanographic frontal context include the wavelet based approach of Simhadri et al. (1998), mathematical morphology algorithms (Lea and Lybanon 1993; Krishnamurthy et al. 1994) and the Ordered Structural Edge Detector of Holland and Yan (1992). The histogram based Single Image Edge Detector (SIED) of Cayula and Cornillon (1992) has been used in many studies: Ullman and Cornillon (2001), Hickox et al. (2000) and Mavor and Bisagni (2001) to name but a few. Miller (2004) for example uses the SIED as the basis for constructing five day composite sediment, chlorophyll and thermal front maps which are then combined into a single multi-spectral image. This approach is ideal if a broad, perhaps exploratory study of the variability and relationship between the surface physical and biological properties within a region is required.

Shaw and Vennell (2000) use a surface fitting technique to 'follow' ocean fronts. An S-shaped function is fitted using least-squares to data extracted within a 20×30 km window centered at the front and orientated along its approximate direction. The approach is unique in the sense that it provides estimates not only of the front's location, but also of key frontal parameters: the mean temperature at the front, its width, and the temperature difference between the water masses on either side. Calculation of the orientation allows the extraction window to be stepped 2 km further along the projected path of the front resulting in a tracking routine across the image. One of the disadvantages of this technique is the bias introduced by user interaction. Since the algorithm is unable to track through cloud due to its field-of-view being limited to the extraction window, it is often necessary to break the front up into a series of segments, the processing of each requiring initialization parameters provided by the user. In an attempt to eliminate this bias Lou et al. (2005) apply a Prewitt gradient operator to automatically locate the front and initiate the 'front following' algorithm at each segment. The fixed window size used by Shaw and Vennell (2000) is a further limitation of the technique since it restricts the smallest resolvable feature to 20 km in an along-front direction. This is evident when the algorithm is compared to a 3×3 Prewitt edge detector. The 'front following' technique fails to capture the smaller scale features resolved by the gradient operator. The fixed size may be optimal for one segment, providing a sufficient amount of smoothing while not blurring oceanic features of interest, but be suboptimal for another where the length scale of features or number of available observations has changed.

In this paper we introduce a new front detection algorithm, based on Shaw and Vennell (2000), that targets a specific frontal structure, is robust to noisy and missing data and requires a minimum of user interaction. We extend the idea of statistical model fitting by using a weighted local likelihood approach to provide a smooth, non-parametric description of spatial variations in the position and strength of ocean fronts from remotely sensed SST images. A likelihood based approach allows us to quantify estimation uncertainty associated with each parameter. As yet, no other front detection technique is able to do this.

We illustrate the algorithm with data from the Southland Front, a localized section of the global Subtropical Front off the south-east coast of South Island, New Zealand. The Southland Front is a well defined boundary separating subantarctic and subtropical surface waters (Jillett 1969; Heath 1985; Shaw and Vennell 2001).

This paper is divided into a further six sections. Section 2 describes the data we use to illustrate the algorithm. In Section 3 the idea of maximum likelihood is introduced in the context of front detection and the mathematical function used to model the change in surface temperature across a front is described. Section 4 extends the idea of likelihood to regression models as a means of estimating spatial trends in frontal characteristics i.e. estimating any increases or decreases in the strength or temperature of the front as it's location changes. The performance and limitations of the algorithm are evaluated in Section 5 and the results compared to the 'front following' algorithm (Shaw and Vennell 2000) in Section 6. Conclusions are presented in Section 7 together with a discussion of the algorithm's advantages, limitations and potential future developments.

2. Data set

To illustrate the method we use a series of monthly composites of 4 km Pathfinder V5 AVHRR infrared SST data from around South Island, New Zealand. These were downloaded through NASA's Physical Oceanography Distributed Active Archive Center (PO.DAAC) POET data server (http://poet.jpl.nasa.gov/) in netcdf format. Observations are globally gridded into equiangle $0.044 \times 0.044^\circ$ pixels. Data were obtained for the period between January 1985 though to December 2005 (inclusive), providing a twenty-one year time series of 252 images that were used in the development of the algorithm. Only night-time overpasses were used to avoid any surface skin created by diurnal warming masking the true surface frontal structure.

3. A model ocean front: maximum likelihood

Remote thermal infrared and passive microwave sensors allow us to measure the rapid change in SST across ocean fronts. It is this surface expression that we aim to model. Suppose that $\mathbf{Z} = \{z^{(1)}, z^{(2)} \cdots, z^{(n)}\}$ is a vector of independently observed temperatures at right angles across an ocean front. The probability of these observations being drawn from a given model front may be expressed in terms of a likelihood function. This may be thought of as the formula for the joint probability distribution of the sample $\mathbf{Z}$. If $p(\mathbf{Z}; \boldsymbol{\theta})$ represents the probability density function of $\mathbf{Z}$ with a vector of unknown parameters $\boldsymbol{\theta}$, then:

$$\text{Likelihood} \equiv l(\boldsymbol{\theta}; \mathbf{Z}) \equiv p(\mathbf{Z}; \boldsymbol{\theta}) = \prod_{i=1}^{n} p(z^{(i)}; \boldsymbol{\theta}), \tag{1}$$

where n is the number of observations across the front. The aim of maximum likelihood estimation is to find the set of values of the unknown parameters $\boldsymbol{\theta}$, that given SST observations $\mathbf{Z}$, make the likelihood $l(\cdot)$ a maximum.

Now assume that each SST observation $(z^{(i)})$ is drawn from a normal distribution, $\phi(\cdot)$, with mean $\mathbb{E}(z^{(i)})$ and variance $\mathbb{V}(z^{(i)}) = \sigma^2$, assumed to remain constant for all i. A normal distribution is chosen in the absence of any other information about the observations and could be changed when necessary, for example where, because of incorrect flagging of scattered cloud, the distribution is skewed. If the expectation $\mathbb{E}$ is determined by the function $m(\mathbf{Y};\boldsymbol{\theta})$, where $\mathbf{Y} = \{y^{(1)}, y^{(2)} \cdots, y^{(n)}\}$ is a vector of known distances across the front corresponding to observations $\mathbf{Z} = \{z^{(1)}, z^{(2)} \cdots, z^{(n)}\}$, and the parametric model $m(\cdot)$ has a vector of unknown parameters, $\boldsymbol{\theta} = \{\theta_1, \theta_2, \ldots, \theta_q\}$, we may express the likelihood of observed temperatures $\mathbf{Z}$ by:

$$
\begin{aligned}
l(\boldsymbol{\theta}, \sigma; \mathbf{Z}) &= \prod_{i=1}^{n} \phi\left(\mathbb{E}(z^{(i)}), \mathbb{V}(z^{(i)})\right) \\
&= \prod_{i=1}^{n} \phi\left(m(y^{(i)};\boldsymbol{\theta}), \sigma^2\right) \\
&= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(z^{(i)} - m(y^{(i)};\boldsymbol{\theta}))^2}{2\sigma^2}\right]
\end{aligned}
$$

Sigma, σ, may be thought of as the standard deviation of noise about the model function. These concepts and assumptions are illustrated graphically in Figure 1a.

It is often more convenient to maximize the log of the likelihood. Letting $L(\cdot)$ denote the log of the likelihood, $\ln(l(\cdot))$:

$$
\begin{aligned}
L(\boldsymbol{\theta}, \sigma; \mathbf{Z}) &= \sum_{i=1}^{n} \ln\left\{\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(z^{(i)} - m(y^{(i)};\boldsymbol{\theta}))^2}{2\sigma^2}\right]\right\} \\
&= \sum_{i=1}^{n} \left[\ln(1) - (\ln\sqrt{2\pi} + \ln\sigma)\right] - \sum_{i=1}^{n} \frac{(z^{(i)} - m(y^{(i)};\boldsymbol{\theta}))^2}{2\sigma^2},
\end{aligned}
$$

Since $\ln(1) = 0$, and ignoring $-\ln\sqrt{2\pi}$ as an irrelevant constant,

$$
L(\boldsymbol{\theta}, \sigma; \mathbf{Z}) = -n\ln\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(z^{(i)} - m(y^{(i)};\boldsymbol{\theta}))^2. \tag{2}
$$

The maximum likelihood estimates $\hat{\boldsymbol{\theta}}$ and $\hat{\sigma}$ are attained when the rate of change of $L(\cdot)$, with respect to the unknown parameters $\boldsymbol{\theta}$ and σ, equals zero.

The function $m(\cdot)$ used to model the change in surface temperature across a front is required to be a sigmoid (S-shaped) function that is able to emulate the steep thermal gradient at the interface between two water masses with disparate surface temperatures. Previous work by Shaw and Vennell (2000) uses the hyperbolic tangent, and we adopt the same function here to represent a cross section of sea surface temperature observations.

$$\mathbf{Z} = m(\mathbf{Y};\boldsymbol{\theta}) + \epsilon = \theta_1 + \theta_2 \tanh\left[\frac{\mathbf{Y} + \theta_4}{\theta_3}\right] + \epsilon. \tag{3}$$

θ_1 is the front's mean temperature. $2\theta_2$ and $2\theta_3$ define the temperature difference and width respectively. θ_4 is a translation parameter determining the position of the front within an equiangle arc degree grid. The noise, ϵ, is assumed to be normally distributed with zero mean and standard deviation σ. Figure 1b is a graphical representation of the how the model parameters may be interpreted in a more physical sense.

Using this model, and setting $\boldsymbol{\theta} = \{\theta_1, \theta_2, \theta_3, \theta_4, \sigma\}$, the log-likelihood may now be expressed in full as follows:

$$L(\boldsymbol{\theta};\mathbf{Z}) = -n \ln \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^{n} \left(z^{(i)} - \left(\theta_1 + \theta_2 \tanh\left[\frac{y^{(i)} + \theta_4}{\theta_3}\right]\right)\right)^2. \tag{4}$$

The advantage of using maximum likelihood over other parameter estimation techniques such as least-squares are its statistical properties that allow the construction of confidence intervals around $\hat{\boldsymbol{\theta}} = \{\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \hat{\theta}_4, \hat{\sigma}\}$. Asymptotically (as $n \to \infty$), the maximum likelihood estimator is unbiased, has the smallest possible variance and is consistent. Thus as $n \to \infty$, it holds that $\mathbb{E}(\hat{\boldsymbol{\theta}}) = \boldsymbol{\theta}$, $\mathbb{V}\mathrm{ar}(\hat{\boldsymbol{\theta}}) = -\mathbf{H}^{-1}(\hat{\boldsymbol{\theta}})$ and therefore $\hat{\boldsymbol{\theta}} \approx \phi(\boldsymbol{\theta}, \mathbb{V}\mathrm{ar}(\hat{\boldsymbol{\theta}}))$, where $\mathbf{H}$ is a matrix of second order partial derivatives with respect to the unknown parameters.

That is, the distribution of estimates is asymptotically normal with the true (but unknown) parameters $\boldsymbol{\theta}$ as expectation. For small n this is an approximation (Davison 2003). The estimated variances of the maximum likelihood estimates are equal to the diagonal elements of, $-\mathbf{H}^{-1}(\hat{\boldsymbol{\theta}})$, the asymptotic variance-covariance matrix. The square root of the diagonal elements yields the standard errors. Given the asymptotically normal distribution of the maximum likelihood estimates, confidence intervals for $\hat{\boldsymbol{\theta}}$ may be constructed:

$$\hat{\boldsymbol{\theta}} - t_{\alpha,\upsilon} \cdot \mathbb{V}\mathrm{ar}(\hat{\boldsymbol{\theta}})^{\frac{1}{2}} \;\; ; \;\; \hat{\boldsymbol{\theta}} + t_{\alpha,\upsilon} \cdot \mathbb{V}\mathrm{ar}(\hat{\boldsymbol{\theta}})^{\frac{1}{2}},$$

where $t_{\alpha,\upsilon}$ is the t-value for a t-distribution with υ degrees of freedom for a $100(1-\alpha)\%$ confidence interval. υ is defined as $n-5$, the number of observations minus the number of unknowns being estimated.

Maximum likelihood parameter estimates are made based on the assumption that each set of SST observations ($\mathbf{Z}$) are taken at approximately right angles across the front and we may want to orientate our data to make this assumption more reasonable. In our illustration, the Southland Front approximately follows the 500 m isobath (Shaw and Vennell 2001), that runs south-west to north-east along the east coast of South Island, New Zealand. Observations from the original AVHRR SST image are therefore extracted within a rotated window and set within a new co-ordinate system before parameter estimates are made (Figure 2). In this way more accurate estimates of frontal characteristics are obtained. The limitations and complications of not specifically including an orientation parameter in the model function are discussed in Section 5.

A Newton-Raphson optimization scheme is used to calculate the maximum likelihood estimates $\hat{\boldsymbol{\theta}}$. Initial values are needed for parameters in the model to start the optimization. In our example we take the mean temperature (11.01°C), width (8.36 km), and temperature range (1.76°C) of the Southland Front as estimated by Shaw and Vennell (2001) as starting values for θ_1, $2\theta_2$ and $2\theta_3$. The position (θ_4) and standard deviation about the mean model (σ) are set initially to 0.5° and 0.15 respectively. To facilitate rapid and successful convergence a linear transformation is applied to all unknown variables by imposing upper and lower bound constraints (Table 1). The limits are determined by the resolution and accuracy of the data set, the size of the image and reasonable geophysical values expected for sea surface temperatures and frontal scales in the region.

Given that sea surface temperatures around New Zealand are unlikely to exceed 23°C throughout the year (Heath 1985; Chiswell 1994; Uddstrom and Oien 1999; Shaw and Vennell 2001), the mean frontal temperature (θ_1) is given an upper bound of 23°C. As in Shaw and Vennell (2000), a lower bound of 5°C is set for θ_1 so as to exclude any high level cloud that may have been missed by the cloud clearing algorithms. Taking into account the range of the seasonal SST cycle around New Zealand (Heath 1985; Chiswell 1994; Uddstrom and Oien 1999), an upper limit of 6°C was set for the temperature difference ($2\theta_2$) across the front. Any greater difference seems unreasonable and likely to include cloud. The lower bound for the temperature range was set at 0.1°C, the approximate relative accuracy of the Pathfinder data set (K.Casey, NOAA, personal communication). The maximum and minimum bounds for the frontal width ($2\theta_3$) are determined by the size (1°×2.8°) and resolution (0.044°) of the images. Note that only one front is allowed to exist. Each image is 1° wide, approximately 100 km (across-front scale), therefore an upper width range was set at 0.6°(∼60 km). This allows for roughly 20 km of data from each plateau region on either side of the front. Having a sufficient amount of data to identify these regions was found to be important by Shaw and Vennell (2000). The lower limit is set by the 0.044° (∼4 km) data resolution. The width and therefore gradient of features smaller than this can not be accurately resolved. Estimates of the temperature, temperature difference and position to within a pixel however may still be regarded as reliable, although the errors associated with them may not be. The frontal position (θ_4) is bounded by the limits of the extracted image.

Figure 3 is an example of the model function fitted to five different sets of SST observations taken across the Southland Front. 95% confidence intervals are given for each estimated parameter. Note how the estimated characteristics of the front vary between each profile.

4. A non-parametric model for spatial trend

The position, strength and temperature of fronts change across ocean basins in response to localized oceanic and atmospheric conditions and interactions such as: stratification, wind stress and mixing, sea level pressure, bathymetry, vorticity constraints, regional eddy activity, circulation dynamics and remote ocean-atmosphere forcing. To capture these spatial changes local likelihood, an extension of weighted local fitting techniques to likelihood based regression models (Tibshirani and Hastie 1987), is used to create a smooth, non-parametric description of spatial variations in frontal characteristics.

Assume that temperature observations $\mathbf{Z}_j = \{z_j^{(1)}, \ldots, z_j^{(n)}\}$ are available at fixed and known positions x_j along the front, where $j = \{1, \ldots, m\}$. Observations $\mathbf{Z}_j$ are realizations from the parametric model of Eq.(3):

$$\mathbf{Z}_j = m(x_j, \mathbf{Y}_j; \Phi(x_j)) = \theta_{1,j} + \theta_{2,j} \tanh\left[\frac{\mathbf{Y}_j + \theta_{4,j}}{\theta_{3,j}}\right] + \epsilon_j,$$

where ϵ_j is normally distributed random noise with a mean of zero and unknown variance σ_j. $\mathbf{Y}_j = \{y_j^{(1)}, \ldots, y_j^{(n)}\}$ is a vector of known distances across the front indexed by the superscript $i = \{1, \ldots, n\}$. We now assume that the unknown parameters $\boldsymbol{\theta}$ are themselves a smoothly varying function of x denoted by $\Phi(x_j)$. $\boldsymbol{\theta}_j = \{\theta_{1,j}, \theta_{2,j}, \theta_{3,j}, \theta_{4,j}, \sigma_j\}$ as before represent the front's mean temperature, the across-front temperature change, the width, position and standard deviation of noise about the model fit. We are interested in estimating the smooth function of parameters $\boldsymbol{\theta}$ as they vary along the front, i.e. the function $\Phi(x)$.

A standard approach would be to assume a parametric model for the form of $\Phi(x)$, such as a linear regression where $\boldsymbol{\theta}_j = \boldsymbol{\alpha} + \boldsymbol{\beta} x_j$. The likelihood equation $l(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \prod_{j=1}^{m} m(x_j, \mathbf{Y}_j; \Phi(x_j; \boldsymbol{\alpha}, \boldsymbol{\beta}))$ would then be solved to obtain parameter estimates $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$ and a fitted across-front temperature profile $\hat{\mathbf{Z}}_j = m(x_j, \mathbf{Y}_j; \hat{\Phi}(x_j))$. However, such a parametric approach is not justified in the case of frontal modelling; we can not presume to know how a parameter may vary in such a complex system.

In contrast, the local likelihood method assumes only that $\boldsymbol{\theta}$ is a smooth function of x and is an ideal alternative approach. We estimate the coefficients of the function $\Phi(x)$ locally at each discrete point x_j. The most basic case is where $\Phi(x_j)$ is assumed approximately constant at points close to x_j, i.e. no particular model for the behavior of the parameters near x_j. Denoting the log-likelihood associated with $m(\cdot)$ from the j^{th} set of temperature observations as:

$$L(\boldsymbol{\theta}_j;\mathbf{Z}_j) = -n\ln\sigma_j - \frac{1}{2{\sigma_j}^2}\sum_{i=1}^{n}\left(z_j^{(i)} - \left(\theta_{1,j} + \theta_{2,j}\tanh\left[\frac{y^{(i)}+\theta_{4,j}}{\theta_{3,j}}\right]\right)\right)^2, \tag{5}$$

the local likelihood estimator for $\boldsymbol{\theta}_j$ is of the form:

$$\hat{\boldsymbol{\theta}}_j = \max_{\boldsymbol{\theta}_j}\sum_{k=1}^{m} K(x_k - x_j;h)\cdot L(\boldsymbol{\theta}_j;\mathbf{Z}_k), \tag{6}$$

where $K(x_k - x_j; h)$ is a normal smoothing function such that $\sum_{k=1}^{m} K(x_k - x_j; h) = 1$, with bandwidth $h > 0$. The bandwidth controls the spatial smoothness of $\hat{\boldsymbol{\theta}}_j$. Note that j is fixed with k varying over the points $k = \{1,\ldots,m\}$. The estimator $\hat{\boldsymbol{\theta}}_j$ is the value of $\boldsymbol{\theta}_j$ which maximizes the weighted sum of likelihood contributions $w(x_k, x_j)L(\boldsymbol{\theta}_j;\mathbf{Z}_k)$ in which the weights $w(x_k, x_j) = K(x_k - x_j; h)$ are dependent upon the separation of x_k and x_j. The symmetry of the normal function ensures that most weight is given to the point of interest x_j. By solving each of these weighted local likelihood problems at each position x_j we obtain a series of smooth parameter estimates $(x_j, \hat{\boldsymbol{\theta}}_j)$, and a fitted set of temperature profiles $\hat{\mathbf{Z}}_j = m(x_j, \mathbf{Y}_j; \hat{\boldsymbol{\theta}}_j)$. Estimation uncertainty is quantified as before by constructing the variance-covariance matrix of estimates at each position x_j.

This local likelihood approach is a significant improvement over the standard maximum likelihood estimates (i.e. where $h = 0°$). The local likelihood estimates are constructed from a much larger set of observations and therefore have reduced confidence intervals. In addition, estimates in those regions with a sparsity or complete absence of temperature observations are made possible by drawing on surrounding information. The quality flagging system described in the appendix provides a means of identifying estimates made in these areas. The algorithms ability to deal with missing data is illustrated in Section 5.

a. *Bandwidth selection: likelihood cross validation*

The bandwidth h, which for the Gaussian kernel is the standard deviation, controls the width of the kernel and hence the smoothness of the fitted non-parametric regression. Larger values of h correspond to stronger levels of smoothing. Setting the bandwidth equal to zero, the local likelihood estimator reduces to the standard maximum likelihood estimate in Eq.(4) and parameters $\boldsymbol{\theta}_j$ are estimated using only data from points x_j. Taking $h \to \infty$ on the other hand sets the parameters constant globally with distance. Somewhere in between there is an optimal value of h, which may be considered as a measure of model order or complexity.

Likelihood cross-validation, an automatic method designed to determine the level of smoothing best supported by the available data (Silverman 1986) is used to select an optimal bandwidth h_{op} for each image. Figure 4 compares the fitted non-parametric trend for each model parameter using bandwidths $h = 0°$, $h = 0.025°$, $h_{op} = 0.11°$ and $h = 0.3°$. When h is small ($h = 0.025°$) too much of the high frequency variability introduced by the noise in the observations is modelled. When h is large ($h = 0.3°$) much of the mesoscale spatial structure is lost. The optimal bandwidth ($h_{op} = 0.11°$) captures mesoscale variability in the front's position, temperature and strength, while not over-smoothing and missing potentially important features.

Returning to our example, out of 252 monthly 4 km resolution images of the Southland Front the median value of h_{op} based on likelihood cross-validation is 0.095° ($\sim$10.6 km). A kernel with this bandwidth assigns the highest 95% of weights to observations within a 21.6 km window centred about the point of estimation. This distance is comparable to the expected length scale of physical processes in the region; the Baroclinic Rossby Radius around South Island, New Zealand is approximately 20 km (Chelton et al. 1998). We can be confident therefore that likelihood cross-validation is selecting a physically appropriate length scale for the smoothing parameter as well as one that balances the bias and variance of the estimates.

b. Local likelihood optimization

Newton-Raphson is used to find the local likelihood estimates $\hat{\boldsymbol{\theta}}_j$. Since this optimization technique can be sensitive to the initialization parameters (Gill et al. 1995), the bandwidth is initially set to zero and estimates from this simplified problem (i.e. no smoothing function) are then used as a starting point for the full local likelihood solution. Other more robust optimization routines such as the Simplex could be used where an initial guess close to the true parameter values is not possible. This may result however in a substantial increase in computing time. In an attempt to minimize the influence of areas where SST observations are: a. limited; b. unevenly distributed such that the frontal structure is not detectable; or c. where there is no discernible change in gradient, weightings additional to those supplied by the kernel function $K(\cdot)$ based upon the distance between x_k and x_j are introduced. In this way the quality or reliability of each likelihood contribution is also considered. Full details of the criteria used to assign these extra weightings are given in Hopkins (2008) and Appendix A.

Figure 5 shows the along-front trend in frontal characteristics for June 2004. This example demonstrates the algorithms ability to make estimates in regions where SST observations are unavailable. The size of the confidence interval increases where there is a lack of data.

5. Performance evaluation

Two aspects of the algorithms performance were tested using a simulated data set: 1. its ability to cope with noisy data; and 2. the implications of a meandering front. An artificial image was constructed with the same size ($1^\circ \times 2.8^\circ$) and resolution (0.044°) as the AVHRR SST images used in algorithm development. A straight front with constant parameters $\theta_1 = 12^\circ$C, $2\theta_1 = 2^\circ$C, $2\theta_3 = 20$ km, and $\theta_4 = 0.5^\circ$ was built centrally across this image.

The loss in accuracy and precision of estimates as the level of noise increases in an image is shown in Figure 6. Normally distributed random noise $\phi(0, \sigma^2)$ with σ between 0.05 and 0.55 was added to the base image and estimates with bandwidths $h = 0^\circ$ and $h = 0.15^\circ$ were made at each point x_j. For $\sigma = 0.05$ the front is very well defined. When $\sigma = 0.55$ no frontal structure can be distinguished through the noise. For a bandwidth of zero, the range of estimates increases rapidly as the amount of noise becomes more significant. This is brought under control by increasing the smoothing parameter to $h = 0.15^\circ$. Note that for $h = 0^\circ$, the standard deviation of noise is increasingly underestimated as the true value increases (Figure 6a, panel E). This is improved by using a bandwidth of 0.15° (Figure 6b, panel E). If the relative accuracy of AVHRR measurements between pixels is of the order 0.1°C we conclude that the local likelihood estimates are not overly sensitive to noise and errors in the AVHRR SST measurements.

The model function is fitted based on the assumption that the front is oriented east-west across the image (i.e. each cross section of observations ($\mathbf{Z}$) is at right angles to the front). In reality fronts meander and the angle at which each cross section bisects the front may change. The artificial front (with a width of 10 km) was rotated between 0° and 90°

from the horizontal. A front with an angle of 0° is bisected at right angles by each cross section. An angle of 90° represents a front orientated north-south across the image. Figure 7 shows how each parameter estimate (for $h = 0°$) deviates from its true value as the angle of rotation increases. Estimates of the mean temperature, temperature range and position remain unaffected by the orientation of the front. Estimates of the width and therefore the gradient are more sensitive. The estimated width exponentially increases as the angle steepens (Figure 7c). Beyond 25-30° the width is overestimated by >14% resulting in an underestimate of the gradient. Additionally, the standard deviation of noise about the model fit (Figure 7f) is slightly underestimated by 0-5% of the true value between 0° and 40°. This is consistent with results presented in Figure 6. At angles greater than 40° it is overestimated. This experiment highlights that using this version of the algorithm estimates of the width and gradient must be treated with caution in regions where the front meanders at angles greater than 25°. Note that if the front is not perpendicular to the axis along which the cross sections are taken then the errors resulting from any rotation are not represented in the algorithms error estimates.

The orientation of the front and hence the angle at which data $\mathbf{Z}_k$ are extracted, like the mean temperature and position etc. is an unknown variable, but not one that is easily incorporated into the optimization. Unlike the other parameters in vector $\boldsymbol{\theta}$ its value would need to be allowed to vary spatially between points k during the optimization such that it resulted in the smallest possible estimate of the width (or greatest gradient) at each position. This would make the optimization more complex. Furthermore, continually adjusting the extraction angle for vectors $\mathbf{Z}$ at each iteration would result in SST observations being used more than once. The effect on the underlying statistics of the technique of repeatedly using the same data is difficult to quantify. Maximum likelihood requires samples to be independent, if vectors $\mathbf{Z}_k$ and $\mathbf{Z}_{k+1}$ contain some of the same measurements then this requirement is clearly not met. Including a rotational parameter presents complications to both the optimization, programming and statistical assumptions of the technique. Possible solutions are discussed in Section 7.

One of the objectives in the development of a new front detection algorithm was to minimize the loss of information owing to regions occluded by persistent cloud cover. To test the robustness of the technique to data loss, SST observations are removed from a cloud free composite image of the Southland Front (September 1997). Instead of trying to simulate the real spatial patterns of cloud cover and contaminated SST retrievals, the patterns of missing data from a selection of other months provides a means of removing observations. Figure 8a shows along-front estimates of parameters for the cloud free image of September 1997. Note that the data missing in the top left hand corner is the Dunedin Headland. In Figure 8b a small (16%) percentage of the observations have been removed, mostly over the front itself. In Figure 8c over 40% of measurements are missing. Estimates of the front's position are most robust to a sparse data set, deviating very little from the original estimate made with a complete set of observations. Where observations have been removed the width is underestimated, although the estimate made with a complete data set falls, for the most part, within the 95% confidence interval. There are two peaks (>50km) in the estimated width. These points correspond to increases in the angle of the front relative to the horizontal and subsequent overestimation of the width.

6. Comparison to other techniques

The local likelihood algorithm is directly compared to the 'front following' algorithm of Shaw and Vennell (2000) as an independent means of validating its performance. In Figure 9 the location of the Southland Front on 28 March 1990 as estimated by both techniques from a daily 1 km resolution AVHRR SST image is shown. Also plotted are the fronts identified by the Single Image Edge Detector (SIED) (Cayula and Cornillon 1992).

The local likelihood and 'front following' algorithms produce very similar estimates of the front's position and both agree with the western most structure picked out by the SIED. The local likelihood estimates highlight more of the mesoscale variability than the 'front following' algorithm. This is likely due to the limited resolvable along-front length scale imposed by the 20 km wide moving window used by Shaw and Vennell (2000). There are gaps around 44.5°S and 45.8°S in estimates made by the 'front following' algorithm where the routine was unable to identify the front. There are no gaps in the local likelihood estimates although the location of the front over the Dunedin Headland is clearly incorrect. This however is identified by flagging those sets of estimates where one or more of the parameters in vector $\hat{\boldsymbol{\theta}}_j$ is on an upper or lower bound (see Table 1).

The SIED helps identify frontal structures further offshore and reveals a possible double structure to the Southland Front. North of Dunedin both the 'front following' algorithm and local likelihood estimates most closely follow the more shoreward of the two SIED fronts. This would suggest that the strongest of the two structures is found further west.

Table 2 compares the mean parameter estimates made by the 'front following' and local likelihood algorithms over three discrete sections A-C (shown in Figure 9). Identifying latitudinal sections over which both algorithms performed successfully is the best way to quantitatively compare the performance of the two techniques. Estimation of the standard errors using the local likelihood algorithm allows weighted mean estimates for each section to be calculated. Each estimate is weighted by the inverse of its own variance. No consideration of the reliability of estimates contributing to the mean is possible with the 'front following' technique. In all three sections estimates of the temperature (θ_1) and temperature range ($2\theta_2$) are comparable. Estimates of the width ($2\theta_3$) however are not in such close agreement. The disparity is most pronounced in section A, south of Dunedin, where the local likelihood algorithm estimates a tight meander in the position of the front (Figure 9). Using the 'front following' algorithm the width of the Southland Front over this section is estimated to be 8.10 km. The local likelihood approach returns a much greater estimate of 16.09$\pm$0.6 km. This is likely the result of overestimation due to the front's orientation. The moving extraction window used in the 'front following' algorithm is aligned normal to the front and therefore this technique does not suffer the same problem.

7. Discussion and conclusions

In this paper we have demonstrated how local likelihood may be used to help detect and characterize ocean fronts. The rapid change in SST between two different water masses on either side of a front is modelled using an S-shaped (tanh) function. The unknown parameters of this model are determined by maximizing a weighed sum of likelihood contributions from all available cross sections of SST observations in an image. A Gaussian smoothing function assigns weightings based on the distance of observations from the point of estimation. The bandwidth of the kernel function determines the smoothness of the fitted nonparametric regression, with larger values corresponding to stronger levels of smoothing. Likelihood cross validation is used to determine the optimal level of smoothing best supported by the available data. The weightings assigned by the kernel smoother are modified by an additional weighting based on an assessment of the quality of the likelihood contributions from each set of SST measurements. Estimation uncertainty is quantified by standard errors calculated from the variance-covariance matrix of each local likelihood solution.

a. Advantages and limitations

The local likelihood approach has both advantages and disadvantages over other front detection techniques. It should therefore be considered as an additional tool in the suite of existing SST front detection algorithms. In this section we discuss the merits, drawbacks and assumptions of the new technique.

The local likelihood algorithm is not overly sensitive to noise and regions with partially missing data. This means that a pre-filtering routine often necessary with other techniques (e.g. Canny edge detection, SIED) is not required. In this way all frontal structures within the original image are preserved. The technique targets a unique frontal structure rather than locating all gradient discontinuities in an image. Key estimates of frontal strength and temperature are obtained, important variables in terms of the structure and dynamics of the front and how it interacts with other oceanic phenomena and the atmosphere above. Crucially, uncertainty estimates are made which may be taken into account when results are used in further quantitative studies. However, using an S-shaped model function that allows for only one front is in some situations inappropriate. Fronts may bifurcate and then merge back together, creating multiple frontal structures. Fronts may also be embedded in a surrounding weak but non-zero gradient field for which the sigmoid function is not best suited. In these situations other techniques better equipped for finding multiple fronts rather than characterizing one particular feature may perform better. Similarly, the performance of the local likelihood algorithm deteriorates in areas of high meandering intensity and where small high gradient eddies are common. This is a further example of where alternative techniques need to be explored. Adding a second, third or fourth etc. function to the model would in theory allow more than one front in an image to exist. Further investigation into how such a model would perform in practice however is left for future research. Possible solutions to the problems caused by the front's orientation are addressed in Section 7b.

Currently, the new algorithm requires some initial knowledge of the approximate location and orientation of a front if it is to return good quality results. If little or no information about the area and front to be studied is available then the automated front detection techniques of Cayula and Cornillon (1992), Miller (2004) or Holyer and Peckinpaugh (1989) would be a useful starting point in making a first estimate of location and variability. From this, an appropriate range of values for the front's mean temperature, temperature difference and width could be estimated and used to initialize the optimization.

A normal distribution with standard deviation σ is assumed to model the noise about the model function. Scattered cloud that has not been correctly flagged by cloud screening algorithms will tend to skew this towards the colder side of the distribution. If there was a particular problem with cloud flagging then a more complex skewed distribution could be used. Similarly, σ may not always remain invariant across a front. It is allowed to vary along the front to take into account changes in meteorological conditions that may lead to variations in cloud contamination, but is not permitted to change across-front.

Applications of the algorithm developed here are numerous. Climate variability and monitoring studies, ocean forcasting, validating ocean models, ecosystem and fisheries research all require an accurate understanding of the spatio-temporal behavior of ocean frontal systems. The detailed results from such a front detection scheme would compliment *in-situ* data sets where variability of a front is difficult to resolve. It may also help put into a wider context findings from research cruises only able to sample a limited area.

Although the algorithm has been used with AVHRR SST data and illustrated with the Southland Front it is by no means limited to this type of data and location. It is adaptable to a wide variety of remotely sensed data sets (e.g. altimetry and ocean colour as well as SST), model output and ocean forecasts. Ocean colour and altimetry data may be used simply by adopting a model function that describes the change in colour or height across a front rather than the temperature. In some cases these data sets may be more reliable and appropriate indicators of frontal location.

b. Future developments

There are a number of improvements that could be made in future versions of the algorithm. The bandwidth of the smoothing function used to process each image is determined automatically using likelihood cross validation. An improvement on this scheme would be to introduce a variable bandwidth within each image. This would allow the smoother to adjust to localized variations in the density, distribution and quality of SST observations. For well defined structures that are not obscured by small scale clouds a smaller bandwidth would be favoured, minimizing bias in parameter estimates. Where the frontal structure is poorly defined or masked by cloud then a larger bandwidth would increase the weightings assigned to data further afield and decrease variance in the parameter estimates. Also, the along-front structure and variability of each parameter is different. Adopting different smoothing functions and bandwidths for each parameter may therefore be advantageous. The local likelihood framework however does not naturally allow the degree of smoothing applied to each parameter to be controlled because smoothing takes place in likelihood rather than in parameter space. Further investigation is needed to assess the possibilities of incorporating such flexibility.

The local likelihood algorithm only uses spatial information to estimate the position and characteristics of fronts in regions occluded by cloud cover. Incorporating temporal information about the structure and local motion of the front from images taken before and after the time of interest could improve estimates made in regions of very sparse data (Chin and Mariano 1997). If the position of a front is well defined at times $t-1$ and $t+1$, but obscured by clouds at time t, then estimates made at $t-1$ and $t+1$ could be used to constrain optimization at time t. Accounting for temporal evolution would be most effective when using daily or weekly images. Natural meso-scale variability (meanders, eddies) and the temporal smoothing inherent in monthly composite images means that the position and structure of a front in one month will not necessarily bear any resemblance to the location of features in the previous or following months. The advection and deformation of features over daily and weekly time scales is likely to be significantly less which would allow specific structures to be detected and matched between time frames.

Under the current design, changes in the orientation of the front can lead to overestimation of the width and subsequent underestimation of the gradient. Incorporating some form of angular dependency into the algorithm will be a high priority during future developments and will help to stabilize and correct estimates of the width. For angles less than 80° we have shown that the location of the front can be successfully estimated regardless of orientation. Knowing the relationship between orientation and the error in the estimated width, overestimation could be corrected for post-processing. A better solution would be to extract data $\mathbf{Z}_k$ at variable angles along the front during optimization. A standard edge detection technique could be used to locate the front and estimate its orientation providing the information needed to do this. Alternatively, the standard maximum likelihood model fit (no smoothing kernel) at each position k could be optimized with respect to the angle at which SST observations $\mathbf{Z}_k$ are extracted across the front to obtain the minimum possible estimate of the width (or maximum gradient). This information could then be used to extract vectors $\mathbf{Z}_k$ at optimal angles during the full local likelihood calculation. Both of these solutions would result in a certain number of SST observations being used more than once and the maximum likelihood assumption of independence between vectors $\mathbf{Z}_k$ violated. Strictly speaking however SST measurements in close proximity are not completely uncorrelated and the assumption of independently observed temperatures within each cross sectional vector made in Section 3 may be considered weak. Composite daily, weekly and monthly images are derived from mapping and averaging procedures applied to a number of individual satellite overpasses. This will introduce some level of dependence between measurements. In an attempt to take this into account the spatial decorrelation length scale, the distance at which data are no longer correlated, might be useful to consider when choosing the smoothing bandwidth. Despite these concerns over independence, it may be more appropriate to sacrifice precise statistical correctness in order to gain more realistic estimates of the front's strength.

Acknowledgments.

J. Hopkins was funded by the National Environment Research Council and undertook this work while studying for a doctorate at Southampton University. The authors would like to thank Dr Peter Miller for his helpful comments and application of the SIED in Figure 9. Thanks also to Prof Meric Srokosz, Dr Peter Cornillon and an anonymous reviewer for their useful suggestions and advise.

APPENDIX

Quality Flags and Weightings

The Gaussian kernel function $K(\cdot)$ distributes weightings to likelihood contributions $L(\cdot)$ at positions k based upon the distance between x_k and x_j. This is designed to ensure that data closest to the point of estimation j have most influence on the local likelihood parameter estimates. This weighting however does not take into account the quality or reliability of each likelihood contribution. There are a number of situations where downweighting of likelihood contributions $L(\cdot)$, in addition to weightings of the kernel $K(\cdot)$ is desirable.

In regions where there are a limited number of across-front SST observations, the algorithms ability to make reliable estimates of frontal characteristics is reduced. The quality of these estimates is somewhat dependent upon the amount of data missing and on how this percentage is distributed across the front. Consider the case where $h = 0°$. If a set of SST observations $\mathbf{Z}$ is reduced by 50%, where every other data point across the profile is absent, then the overall horizontal temperature structure is maintained and reliable estimates of the temperature and position etc. may be made. At the other extreme, if all the missing values are on one particular side of the front then vital information concerning one water masss is lost and estimates may not be a true reflection of the structure of the front at that point. Of course when there are no observations available then no estimate can be made. If when $h > 0°$ the data $\mathbf{Z}_k$ are located close to the point of estimation j and thus assigned a high weighting by the kernel $K(\cdot)$, localized sparsity and uneven distribution of observations at k may heavily bias the final local likelihood estimate away from the more realistic values supported by other neighbouring observations.

The second situation in which we might wish to downweight data is where there is no discernible change in gradient across the front, and the best fit to observations $\mathbf{Z}_k$ would be a straight line. This of course may be an accurate reflection of the state of the ocean; a strong subsurface front can exist without a marked surface thermal signature. Strong winds and mixing may break down the surface structure and increased solar insolation in the summer stratifies the water column resulting in an isothermal top layer that may become decoupled from and mask the sub-surface structure below (James et al. 2002). Alternatively it may be the case that gradient changes have been blurred through temporal smoothing in a composite image. This is particularly true of areas where the front has a high degree of spatial variability and an increased number of plumes, as is the case toward the north of South Island, New Zealand (Shaw 1998). A noisy set of observations may also make it difficult to pick out any rapid change in temperature between the two water masses. In all of these cases, the standard maximum likelihood estimate ($h = 0°$) of the width has a tendency to reach the upper bound of 60km (Table 1). As part of a local likelihood estimate this data would bias the width toward larger values.

Thirdly, we must be cautious where we estimate a very sharp decrease in temperature over spatial scales that we are unable to resolve. A lack of data directly over the front where SSTs are changing most rapidly often results in a very narrow estimate of the width. Unfortunately cloud cover over frontal regions is common and may not always be remedied by composite images.

Based on: 1. an evaluation of the optimizations convergence criteria, 2. the number of estimates free from their upper and lower bounds, 3. the value of the likelihood, 4. the residual sum-of-squares (RSS) of the model fit to data $\mathbf{Z}_j$, and 5. the percentage of missing data from $\mathbf{Z}_j$, a quality flag (QF) between 1 and 7 is assigned to each set of estimates $\hat{\boldsymbol{\theta}}_j^{[h=0]}$. Flag 1 is the highest quality and Flag 7 the poorest. Figure 10 is the flow chart that was used to assign each flag. The routines convergence criteria comprise an assessment of whether the function value and parameter estimates are converging on a solution, and whether the gradient vector at the solution ($\mathbf{g}(\hat{\boldsymbol{\theta}}_{sol})$) is zero (Numerical Algorithms Group 2006). Each flag is associated with a weighting (QW) between 0 and 1. This is combined with the weightings assigned by the kernel $K(\cdot)$ to produce a new weighting that takes into account both the location and reliability of each likelihood contribution $L(\cdot)$. If the quality weightings for each $\mathbf{Z}$ in an image are combined into a $1 \times m$ vector $\mathbf{QW}$, the j^{th} row of weightings of the final smoother matrix S_j is given by:

$$\mathsf{S}_j = \frac{K(x_k - x_j; h) \cdot \mathbf{QW}}{\sum_{k=1}^{m} [K(x_k - x_j; h) \cdot \mathbf{QW}_k]}. \tag{A1}$$

Normalization ensures that $\sum_{k=1}^{m} \mathsf{S}_{j,k} = 1$. Using weightings from the matrix S when calculating the likelihood and optimal bandwidth helps ensure a smooth non-parametric trend in along-front estimates of temperature etc. less likely to be interrupted by outliers. The criteria shown in Figure 10 may be adjusted as deemed necessary for different applications, or where a stricter or more leaniant weighting scheme is desired. Further discussion of the quality control criteria may be found in Hopkins (2008) and Numerical Algorithms Group (2006).

REFERENCES

Belkin, I. M. and A. L. Gordon, 1996: Southern Ocean fronts from the Greenwich meridian to Tasmania. *J. Geophys.Res-Oceans*, **101 (C2)**, 3675–3696.

Cayula, J. F. and P. Cornillon, 1992: Edge-detection algorithm for SST images. *J. Atmos. Oceanic Technol.*, **9 (1)**, 67–80.

Cayula, J. F. and P. Cornillon, 1995: Multi-image edge-detection for SST images. *J. Atmos. Oceanic Technol.*, **12 (4)**, 821–829.

Chelton, D. B., R. A. De Szoeke, M. G. Schlax, K. El Naggar, and N. Siwertz, 1998: Geographical variability of the first baroclinic Rossby radius of deformation. *J. Phys. Oceanogr.*, **28 (3)**, 433–460.

Chelton, D. B., M. G. Schlax, M. H. Freilich, and R. F. Milliff, 2004: Satellite measurements reveal persistent small-scale features in ocean winds. *Science*, **303 (5660)**, 978–983.

Chin, T. M. and A. J. Mariano, 1997: Space-time interpolation of oceanic fronts. *IEEE Trans. Geosci. Remote Sens.*, **35 (3)**, 734–746.

Chiswell, S. M., 1994: Variability in sea-surface temperature around New-Zealand from AVHRR images. *N. Z. J. Mar. Freshwater Res.*, **28 (2)**, 179–192.

Coulter, R., 1983: Application of the Bayes decision rule for automatic water mass classification from satellite infrared images. *Proc. 17th Int. Symp. Remote Sensing Environment*, **2**, 589–597.

Currie, K. I. and K. A. Hunter, 1998: Surface water carbon dioxide in the waters associated with the subtropical convergence, east of New Zealand. *Deep-Sea Res. Part I-Oceanographic Research Papers*, **45 (10)**, 1765–1777.

Currie, K. I. and K. A. Hunter, 1999: Seasonal variation of surface water CO_2 partial pressure in the Southland Current, east of New Zealand. *Mar. Freshwater Res.*, **50 (5)**, 375–382.

Davison, A., 2003: *Statistical Models*. 1st ed., Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, 726 pp.

Dong, S., J. Sprintall, and S. T. Gille, 2006: Location of the Antarctic Polar Front from AMSR-E satellite surface temperature measurements. *J. Phys. Oceanogr.*, **36**, 2075–2089.

Garabato, A. C. N., J. T. Allen, H. Leach, V. H. Strass, and R. T. Pollard, 2001: Mesoscale subduction at the Antarctic Polar Front driven by baroclinic instability. *J. Phys. Oceanogr.*, **31 (8)**, 2087–2107.

Gerson, D., E. Khedouri, and P. Gaborski, 1979: Location of oceanic fronts from digital satellite temperature data by automated pattern analysis. *Intergovernmental Oceanographic Commission Workshop Report No.17, Supplement 143-163.*

Gill, P., W. Murray, and M. Wright, 1995: *Practical Optimiztaion*. Academic Press, London, 401 pp.

Heath, R. A., 1985: A review of the physical oceanography of the seas around New- Zealand - 1982. *N. Z. J. Mar. Freshwater Res.*, **19 (1)**, 79–124.

Hickox, R., I. Belkin, P. Cornillon, and Z. Shan, 2000: Climatology and seasonal variability of ocean fronts in the east China, Yellow and Bohai Seas from satellite SST data. *Geophys. Res. Lett.*, **27 (18)**, 2945–2948.

Holland, J. A. and X. H. Yan, 1992: Ocean thermal feature recognition, discrimination, and tracking using infrared satellite imagery. *IEEE Trans. Geosci. Remote Sens.*, **30 (5)**, 1046–1053.

Holyer, R. J. and S. H. Peckinpaugh, 1989: Edge-detection applied to satellite imagery of the oceans. *IEEE Trans. Geosci. Remote Sens.*, **27 (1)**, 46–56.

Hopkins, J., 2008: Statistical Modelling and Variability of the Subtropical Front, New Zealand. Ph.D. thesis, Southampton University.

James, C., M. Tomczak, I. Helmond, and L. Pender, 2002: Summer and winter surveys of the Subtropical Front of the southeastern Indian Ocean 1997-1998. *J. Mar. Syst.*, **37 (1-3)**, 129–149.

Jillett, J., 1969: Seasonal hydrology of waters off the Otago Peninsula, south-eastern New Zealand. *N. Z. J. Mar. Freshwater Res.*, **3 (3)**, 349–375.

Kostianoy, A. G., A. I. Ginzburg, M. Frankignoulle, and B. Delille, 2004: Fronts in the Southern Indian Ocean as inferred from satellite sea surface temperature data. *J. Mar. Syst.*, **45 (1-2)**, 55–73.

Krishnamurthy, S., S. S. Iyengar, R. J. Holyer, and M. Lybanon, 1994: Histogram-based morphological edge detector. *IEEE Trans. Geosci. Remote Sens.*, **32 (4)**, 759–767.

Lea, S. M. and M. Lybanon, 1993: Automated boundary delineation in infrared ocean images. *IEEE Trans. Geosci. Remote Sens.*, **31 (6)**, 1256–1260.

Lou, X., W. Huang, A. Shi, and B. Fu, 2005: Satellite measurements of the Zhejiang-Fujian coastal front in the East China Sea with an improved front-following algorithm. *The International Society for Optical Engineering*, D. Lu and G. Matvienko, Eds., Bellingham, WA, Vol. 5832- Optical Technologies for Atmospheric, Ocean and Environmental Studies.

Marcello, J., F. Marques, and F. Eugenio, 2005: Automatic tool for the precise detection of upwelling and filaments in remote sensing imagery. *IEEE Trans. Geosci. Remote Sens.*, **43 (7)**, 1605–1616.

Mavor, T. P. and J. J. Bisagni, 2001: Seasonal variability of sea-surface temperature fronts on Georges Bank. *Deep-Sea Res. Part II-Topical Studies in Oceanography*, **48 (1-3)**, 215–243.

Miller, P., 2004: Multi-spectral front maps for automatic detection of ocean colour features from SeaWiFS. *Int. J. Remote Sen.*, **25 (7-8)**, 1437–1442.

Moore, J. and M. Abbott, 2000: Phytoplankton chlorophyll distributions and primary production in the Southern Ocean. *J. Geophys. Res -Oceans*, **105**, 28 709–28 722.

Moore, J., M. Abbott, and J. Richman, 1997: Variability in the location of the Antarctic Polar Front (90-20° W) from satellite sea surface temperature data. *J. Geophys. Res -Oceans*, **102**, 27 825–27 833.

Moore, J. K., M. R. Abbott, and J. G. Richman, 1999: Location and dynamics of the Antarctic Polar Front from satellite sea surface temperature data. *J. Geophys. Res -Oceans*, **104 (C2)**, 3059–3073.

Murphy, P. P., R. A. Feely, R. H. Gammon, D. E. Harrison, K. C. Kelly, and L. S. Waterman, 1991: Assessment of the air-sea exchange of CO_2 in the South-Pacific during austral autumn. *J. Geophys. Res -Oceans*, **96 (C11)**, 20 455–20 465.

Numerical Algorithms Group, 2006: NAG Libary Manual, Mark 21. E04- Minimizing or Maximizing a Function. Numerical Algorithms Group Ltd, Oxford, UK, http://www.nag.co.uk/numeric/fl/manual/html/FLlibrarymanual.asp.

O'Neill, L. W., D. B. Chelton, and S. K. Esbensen, 2003: Observations of SST-induced perturbations of the wind stress field over the Southern Ocean on seasonal timescales. *J. Climate*, **16 (14)**, 2340–2354.

Shaw, A. G. P., 1998: The temporal and spatial variability of the Southland Front, New Zealand using AVHRR SST imagery. Ph.D, University of Otago, Dunedin, New Zealand.

Shaw, A. G. P. and R. Vennell, 2000: A front-following algorithm for AVHRR SST imagery. *Remote Sens. Environ.*, **72 (3)**, 317–327.

Shaw, A. G. P. and R. Vennell, 2001: Measurements of an oceanic front using a front-following algorithm for AVHRR SST imagery. *Remote Sens. Environ.*, **75 (1)**, 47–62.

Shimada, T., F. Sakaida, H. Kawamura, and T. Okumura, 2005: Application of an edge detection method to satellite images for distinguishing sea surface temperature fronts near the Japanese coast. *Remote Sens. Environ.*, **98 (1)**, 21–34.

Silverman, B., 1986: *Density Estimation for Statistics and Data Analysis*. Monographs on Statistics and Applied Probability, Chapman and Hall, London.

Simhadri, K. K., S. S. Iyengar, R. J. Holyer, M. Lybanon, and J. M. Zachary, 1998: Wavelet-based feature extraction from oceanographic images. *IEEE Trans. Geosci. Remote Sens.*, **36 (3)**, 767–778.

Simpson, J. J., 1990: On the accurate detection and enhancement of oceanic features observed in satellite data. *Remote Sens. Environ.*, **33 (1)**, 17–33.

Spall, M. A., 1995: Frontogenesis, subduction, and cross-front exchange at upper ocean fronts. *J. Geophys. Res -Oceans*, **100 (C2)**, 2543–2557.

Tibshirani, R. and T. Hastie, 1987: Local Likelihood Estimation. *J. Amer. Statistical. Assoc.*, **82 (398)**, 559–567.

Uddstrom, M. J. and N. A. Oien, 1999: On the use of high-resolution satellite data to describe the spatial and temporal variability of sea surface temperatures in the New Zealand region. *J. Geophys. Res -Oceans*, **104 (C9)**, 20 729–20 751.

Ullman, D. S. and P. C. Cornillon, 2001: Continental shelf surface thermal fronts in winter off the northeast US coast. *Cont. Shelf Res.*, **21 (11-12)**, 1139–1156.

Vazquez, D. P., C. Atae-Allah, and P. L. L. Escamilla, 1999: Entropic approach to edge detection for SST images. *J. Atmos. Oceanic Technol.*, **16 (7)**, 970–979.

List of Tables

Table 1. Geophysical upper and lower bounds used to apply a linear transformation to unknown variables $\boldsymbol{\theta}$ using a $1^\circ \times 2.8^\circ$ window of data with a 0.044° pixel resolution.

Parameter	**Geophysical Lower Bound**	**Geophysical Upper Bound**
Mean Temperature (θ_1)	5°C	23°C
Temperature Range ($2\theta_2$)	0.1°C	6°C
Width ($2\theta_3$)	0.044° (∼4 km)	0.6° (∼60 km)
Position (θ_4)	-1	0
s.d (σ)	1e-06	6

TABLE 2. a. Mean parameter estimates made over sections A-C on 28 March 1990 (as marked in Figure 9) made by the 'front following' algorithm (Shaw and Vennell 2000). b. Weighted mean estimates ($\pm 2\times$standard error of the mean) over sections A-C using the local likelihood algorithm (Hopkins).

Latitude Section	**Parameter**	**a. Shaw and Vennell (2000)**	**b. Hopkins**
A	θ_1 (°C)	13.07	13.47±0.02
46.16-45.96°S	$2\theta_2$ (°C)	1.95	2.53±0.02
	$2\theta_3$ (km)	8.10	16.09±0.60
B	θ_1 (°C)	13.26	13.41±0.01
45.76-44.93°S	$2\theta_2$(°C)	3.24	3.00±0.01
	$2\theta_3$ (km)	12.85	9.30±0.21
C	θ_1 (°C)	13.09	13.39±0.03
44.88-44.75°S	$2\theta_2$ (°C)	2.63	2.44±0.03
	$2\theta_3$)	14.81	9.43±0.80

List of Figures

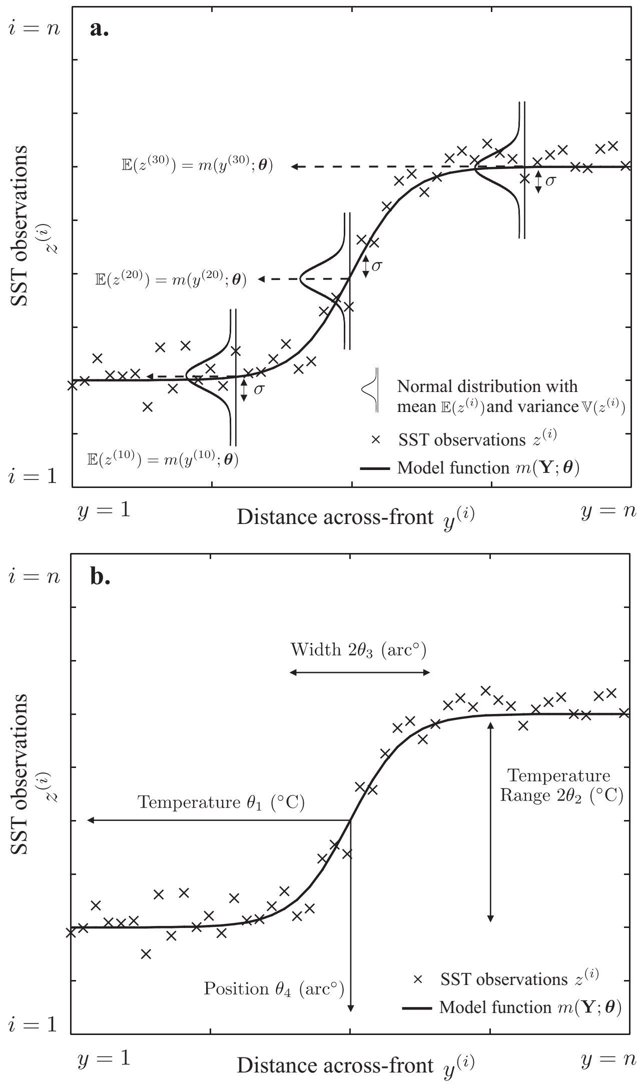

FIG. 1. a. Schematic illustrating the assumption that each SST observation $z^{(i)}$ at distance $y^{(i)}$ across the front is drawn from a normal distribution with mean $\mathbb{E}(z^{(i)})$ and standard deviation σ. The expectation $\mathbb{E}$ is determined by the model function $m(y^{(i)};\boldsymbol{\theta})$. b. Physical interpretation of parameters in model $m(\mathbf{Y};\boldsymbol{\theta}) + \epsilon = \theta_1 + \theta_2 \tanh\left[\frac{\mathbf{Y}+\theta_4}{\theta_3}\right] + \epsilon$.

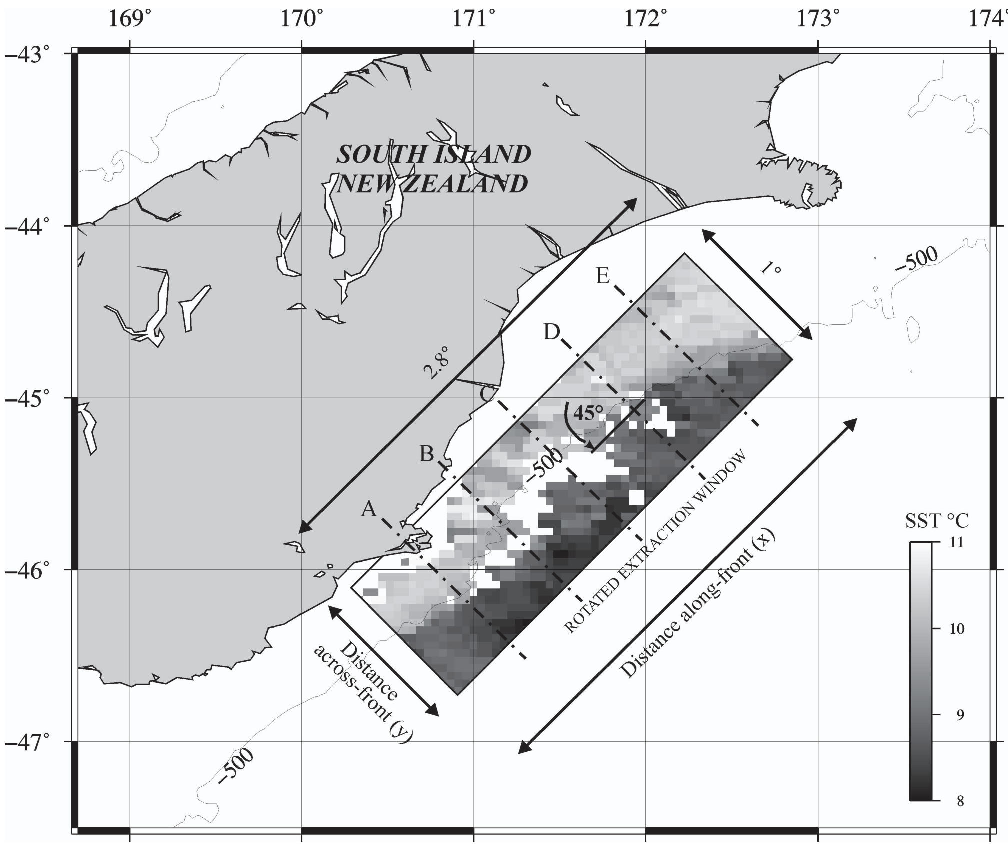

FIG. 2. A 1° by 2.8° window of data aligned approximately along the Southland Front off the east coast of South Island, New Zealand is extracted from each AVHRR SST image. The 500 m isobath is marked. Cross sections of SST observations A-E to which the model function is fitted in Figure 3 are shown.

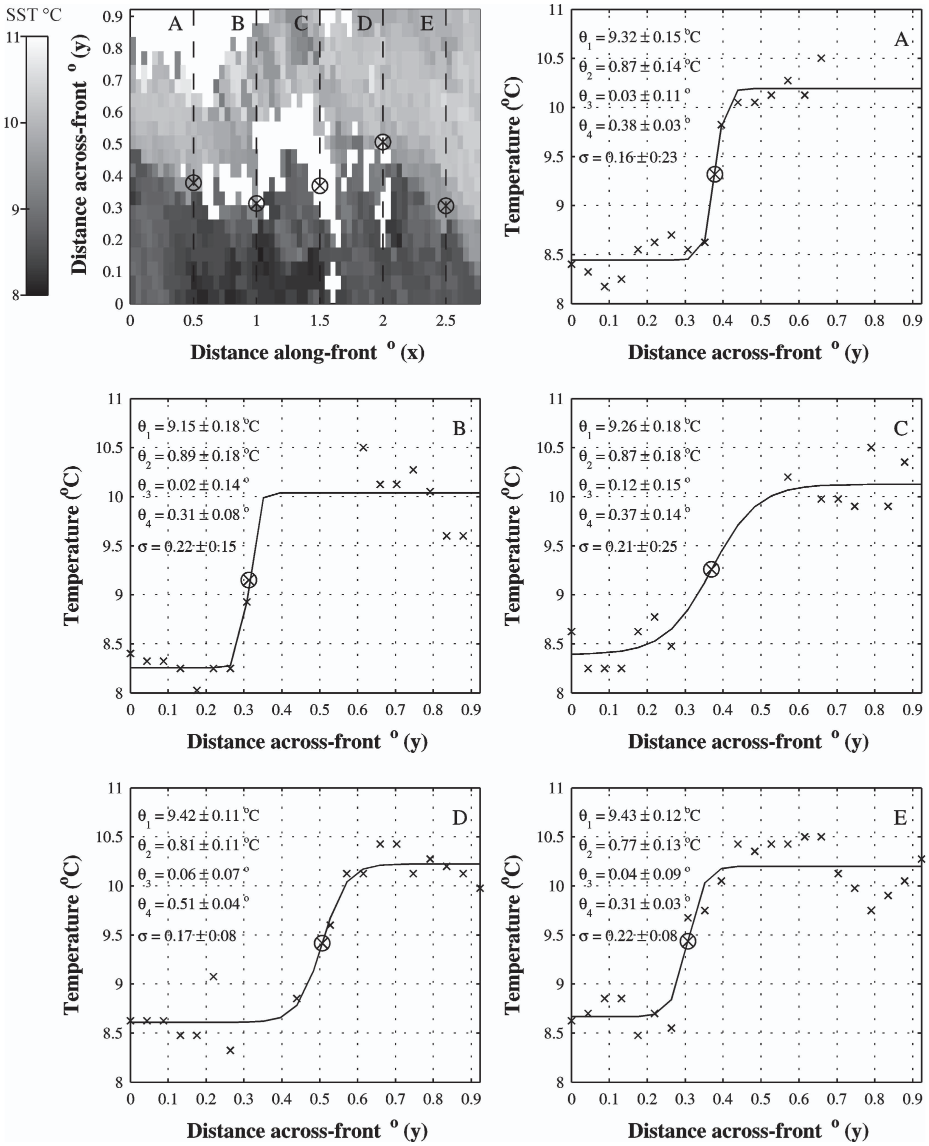

FIG. 3. Examples of the model function fitted to five different sets of SST observations (A-E) across a composite image of the Southland Front from June 2004. For each cross-section the observed temperatures are represented by black crosses and the fitted model (Eq.(3)) by a solid black line. $\otimes$ marks the estimated position. Also shown are the specific parameter estimates $\pm 95\%$ confidence intervals.

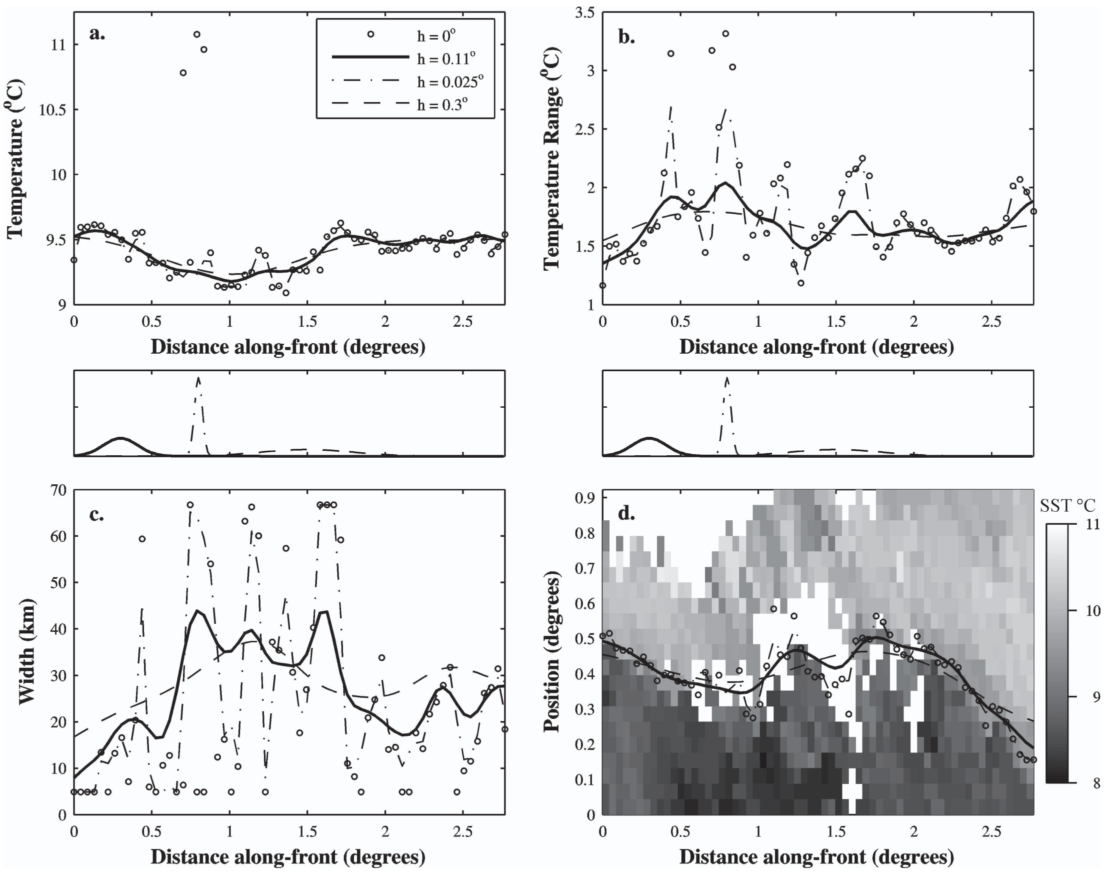

FIG. 4. Local likelihood parameter estimates along the Southland Front in June 2004 for bandwidths $h = 0°$, $h = 0.025°$, $h = 0.3°$ and $h_{op} = 0.11°$. The middle panels represent the width and normalized weightings of each gaussian kernel function. Note the tendency for peaks and troughs to be smoothed out as h increases. Weightings attached to observations $\mathbf{Z}_k$ become increasingly similar and the estimator approaches a fitted least-squares constant; its variance decreases while its bias increases. As h decreases the variance of the estimator increases, but its bias decreases and the non-parametric model is over-fitted.

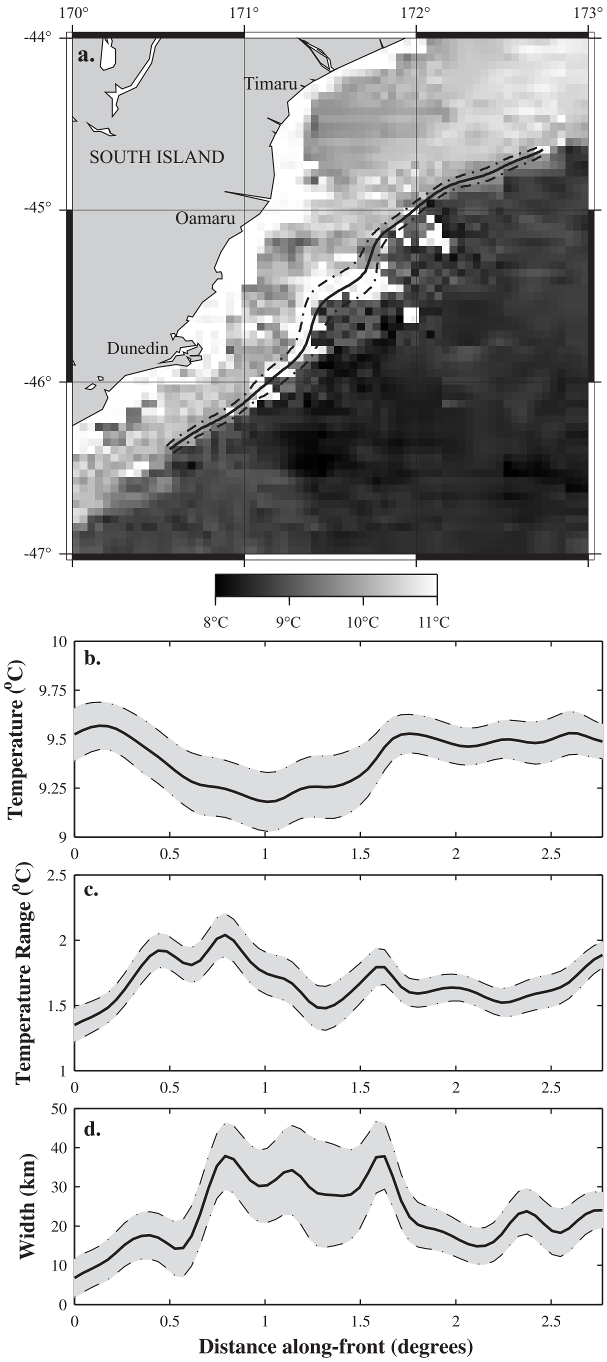

Fig. 5. Local likelihood parameter estimates ±95% confidence intervals (dashed lines) for the Southland Front in June 2004. h_{op}=0.11°.

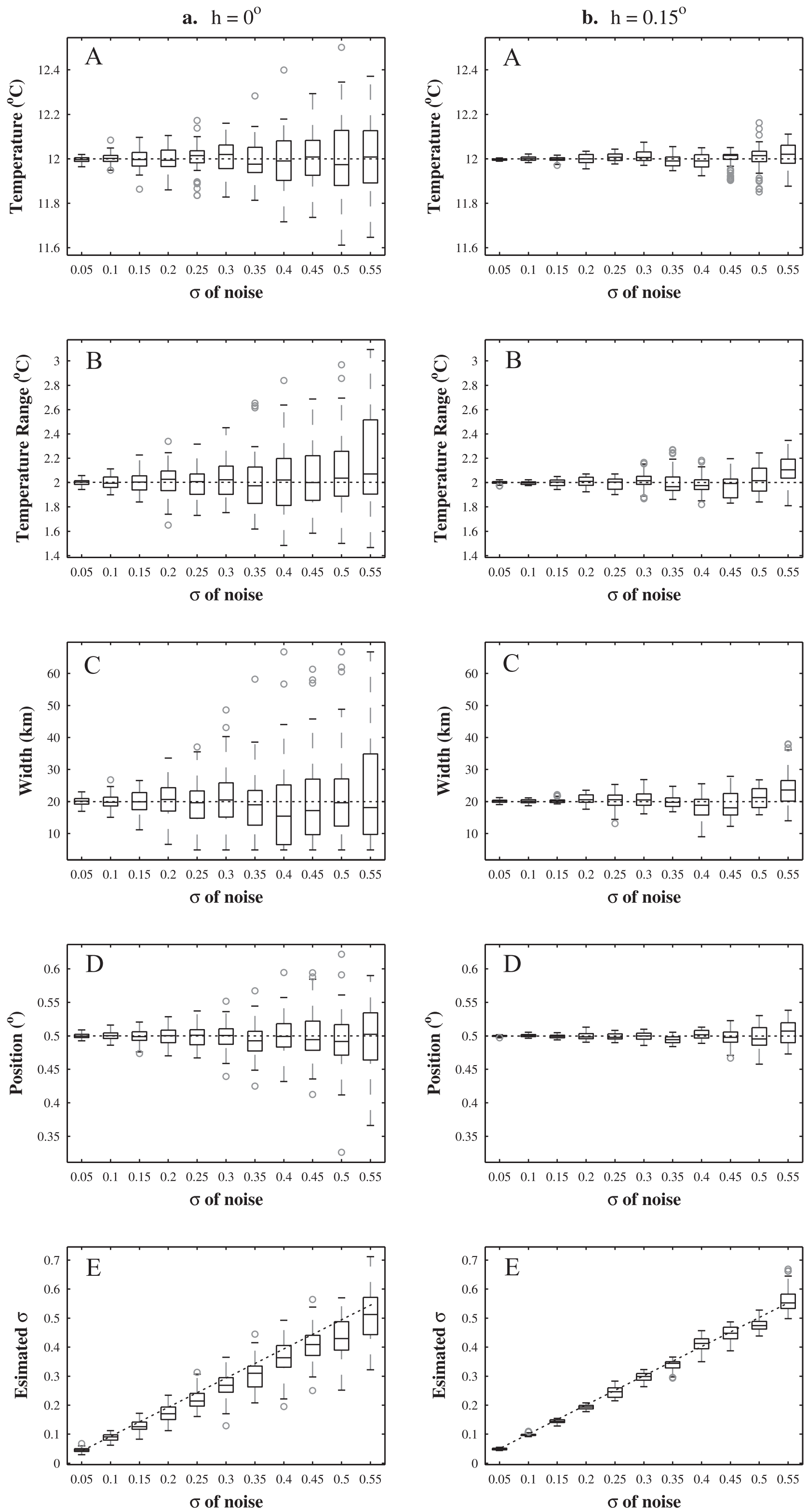

Fig. 6. Distribution of estimated temperature (A), temperature range (B), width (C), position (D) and standard deviation of noise (E) from a simulated data set with smoothing parameters a. $h = 0°$ (left), and b. $h = 0.15°$ (right). True parameter values are shown by the dashed line. Each box has lines at the lower quartile, median, and upper quartile values. Whiskers extend to the most extreme values within 1.5 times the interquartile range from the ends of the box. Outliers are marked by grey circles.

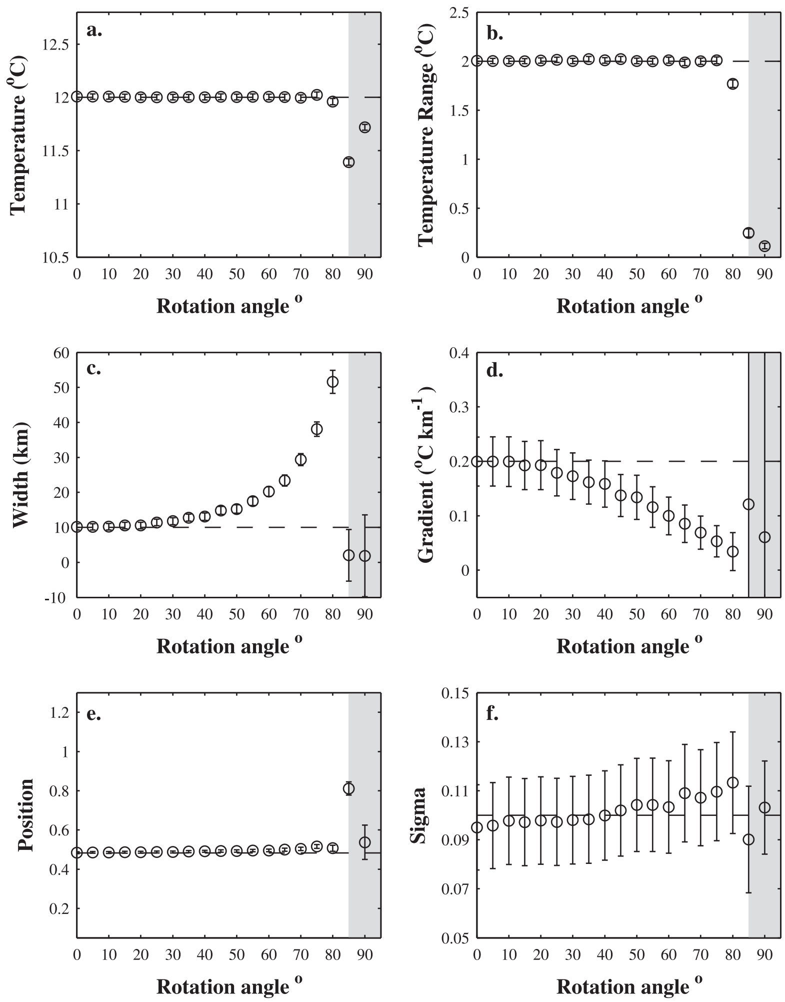

FIG. 7. Effect of frontal orientation on parameter estimates. Angle of rotation vs a. mean temperature (θ_1), b. temperature range ($2\theta_2$), c. width ($2\theta_3$), d. gradient ($\frac{2\theta_2}{2\theta_3}$), e. position ($\theta_4$) and f. standard deviation of noise about mean (σ). The dashed line indicates the true parameter value. The circles represent the mean estimated value at points across the image within the range $-0.1 > x_j < 0.1$. Error bars represent 2 × the mean of the standard errors of these estimates. The shaded gray box represents the angle past which the algorithm 'collapses'. Beyond 80° no useful output may be obtained. Eddies and meanders that loop back on themselves are not resolved.

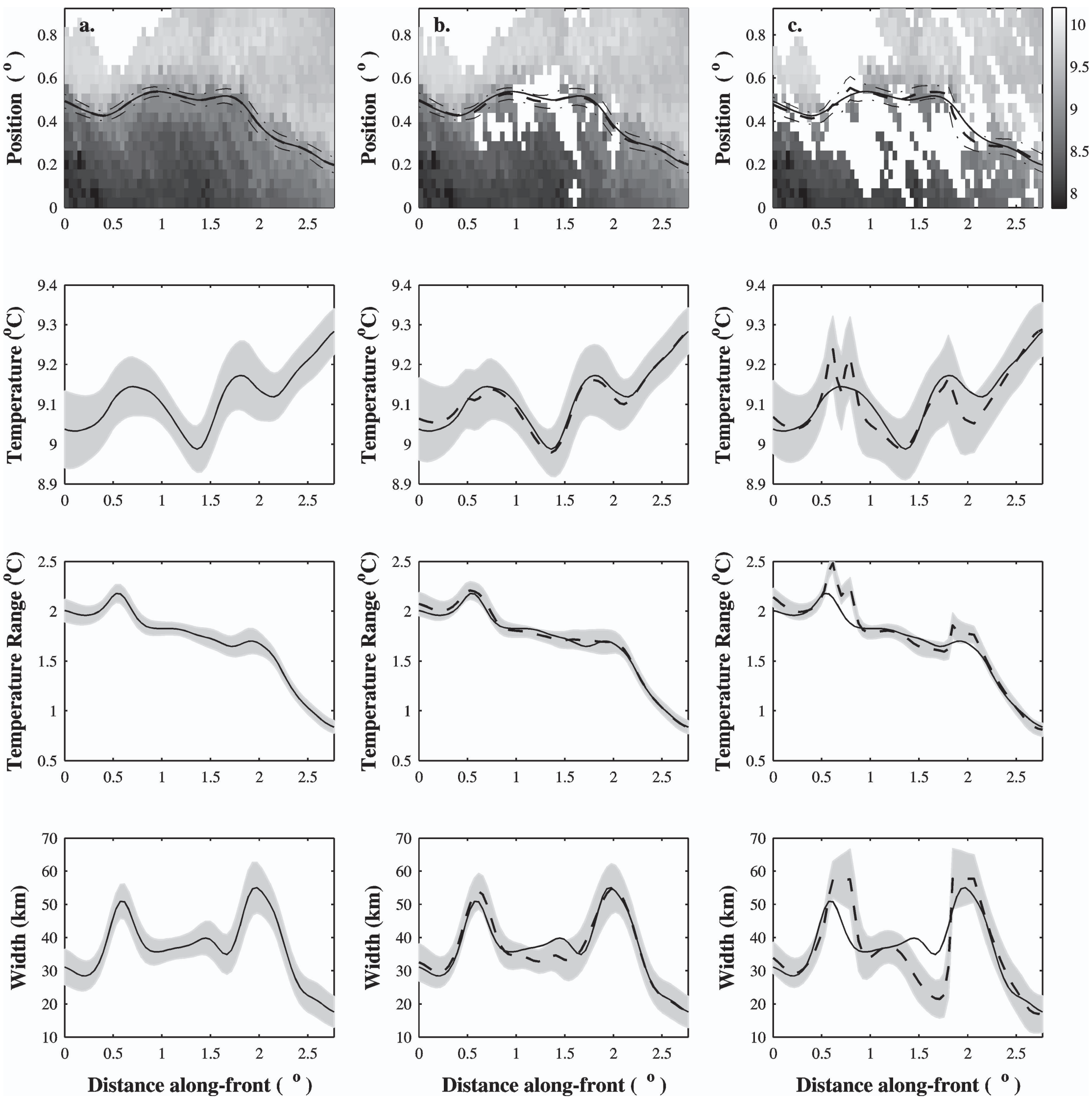

FIG. 8. a. Parameter estimates for September 1997 ±95% confidence intervals, b. Parameter estimates for September 1997 with 16% of observations removed (dashed line). Estimates made with the full data set (column a.) are marked with a solid black line, c. As for b. but with 43% of observations removed. n.b. missing data in the top left hand corner is the Dunedin Headland and accounts for 5% of percentages quoted.

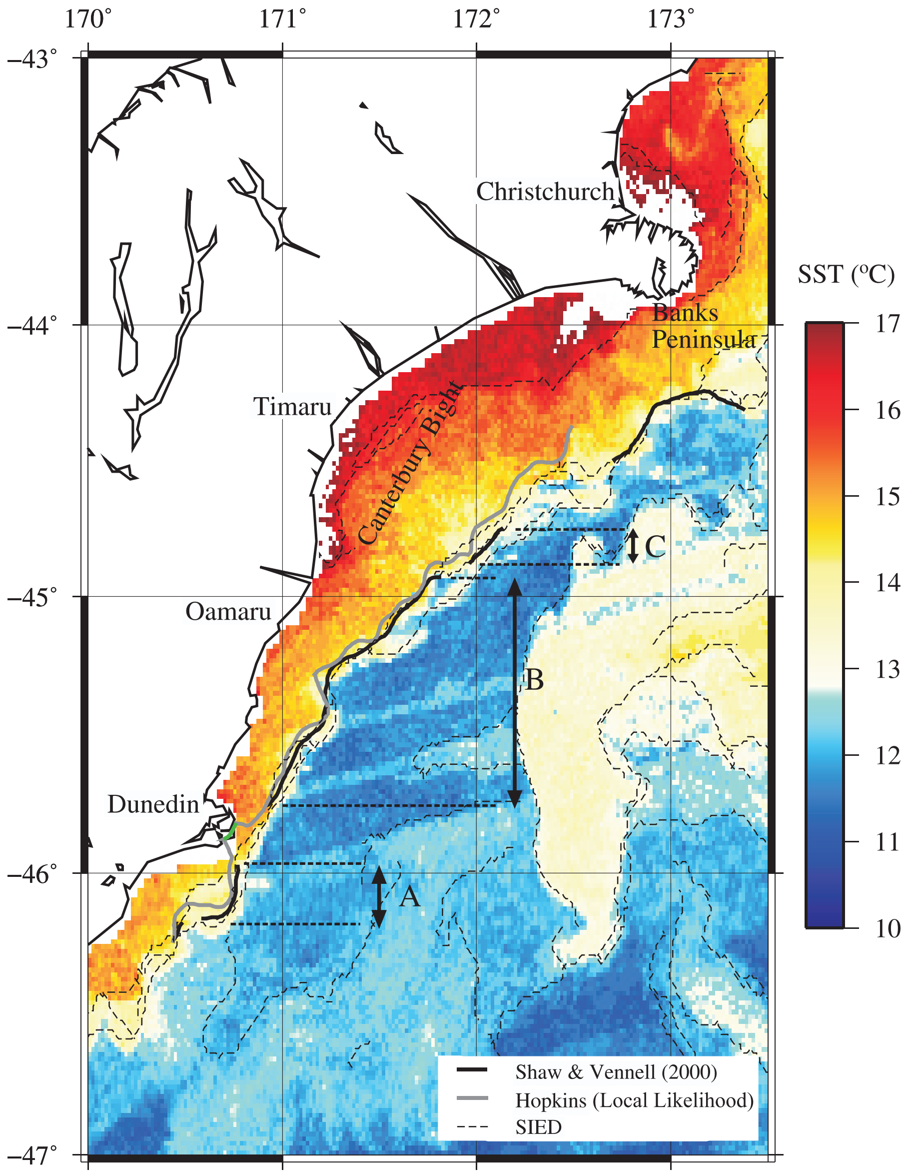

FIG. 9. Location of the Southland Front on 28 March 1990 from Shaw and Vennell (2000) compared with the estimated position using the local likelihood algorithm where $h_{op} = 0.025°$ (Hopkins). Gaps in the estimates made by Shaw and Vennell (2000) are where the algorithm failed to converge within the set parameter bounds. Local likelihood estimates that fall outside the range of acceptable values identified in Table 1 are identified in green. Fronts identified using the single image edge detection (SIED) of Cayula and Cornillon (1992) are also shown (minimum cross-front difference of 0.2°C). Sections A, B and C refer to estimates in Table 2.

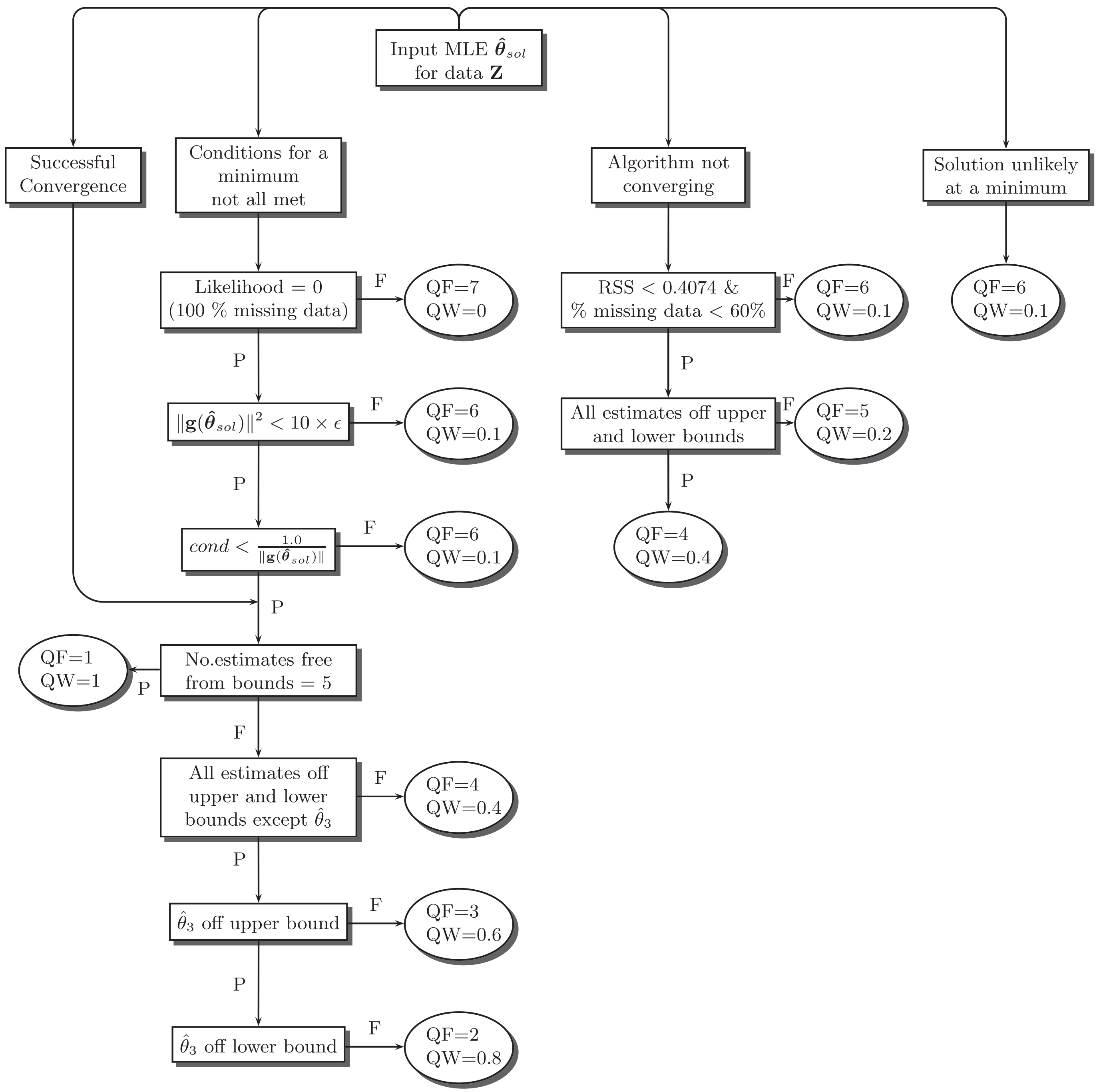

FIG. 10. Flowchart used to assign quality flags (QF) and weightings (QW) to each data set $\mathbf{Z}_j$ when $h = 0°$. P = Pass, F = Fail. The stringent accuracy demanded by the NAG optimization routine may result in an otherwise acceptable solution being reported as a failure. Therefore, when not all conditions for a minimim have been met the gradient vector $\mathbf{g}(\hat{\boldsymbol{\theta}}_{sol})$ and condition number ($cond$) of the matrix of second order derivatives at $\hat{\boldsymbol{\theta}}_{sol}$ are assessed. A small condition number indicates a high rate of convergence and an accurate estimate of $\mathbf{g}(\hat{\boldsymbol{\theta}}_{sol})$. ϵ is the machine precision and RSS the residual sum-of-squares of the model function fit.