SHAPE CLASSIFICATION OF ALTIMETRIC SIGNALS USING ANOMALY DETECTION AND BAYES DECISION RULE

J.-Y. Tourneret[(1)], *C. Mailhes*[(1)], *J. Severini*[(1)] *and P. Thibaut*[(2)]

[(1)] University of Toulouse, IRIT-ENSEEIHT-TéSA, Toulouse, France
[(2)] Collecte Localisation Satellite (CLS), Ramonville Saint-Agne, France
{corinne.mailhes, jean-yves.tourneret, jerome.severini}@enseeiht.fr, pierre.thibaut@cls.fr

1. INTRODUCTION

The use of altimetry measurements over ocean surfaces has been demonstrating its effectiveness for many years. Due to the improved ability of new altimeters to acquire return echoes from oceans, many efforts are now devoted to a better understanding of the signals near the coasts, in the hydrological basins and over land surfaces. The use of altimetry measurements over all these surfaces is now a well identified goal for present and future altimetry missions (conventional or not). Even though the physical processes that induce altimetric signals over land, coastal areas and inland water are different, the contamination of land signals in the altimetric measurements considerably damages the availability and the quality of the data in these cases. Consequently, it becomes crucial to be able to classify altimetric echoes with different shapes with two main objectives: the first objective is to propose dedicated algorithms (called retracking algorithms) able to extract the best geophysical information from each return echo, the second objective is to provide to the user an information about the signal shape giving him the level of confidence he can put on the various retracking algorithm output. A previous work presented in [1] addressed the problem of classifying altimetric signals according to the overflown surface. This paper shows that the methodology proposed in [1] can be modified for classifying altimetric signals according to their shapes.

2. ATIMETRIC SIGNAL MODEL AND PATTERN RECOGNITION SYSTEM

The objective of this paper is to propose a fast pattern recognition algorithm for classifying different shapes of altimetric signals. More precisely, the algorithm will assign a given altimetric signal to one of K classes denoted as $\omega_1, ..., \omega_K$. Each class ω_i is characterized by a template $\boldsymbol{T}_i = [T_i(1), ..., T_i(N)]$. The $K = 14$ class templates used in this study are depicted in Fig. 1. A given altimetric signal from class ω_i is supposed to be a noisy version of the corresponding template $\boldsymbol{T}_i$.

The template $\boldsymbol{T}_1$ associated to the first class results from a simplified formulation of Brown's model. Brown's model was initially studied in [2] and [3]. It has been shown to be appropriate to more than 95% of all altimetric waveforms backscattered from ocean surfaces [4]. The simplified formulation considered in this paper assumes that the received altimeter waveform is parameterized by three parameters: the amplitude P, the epoch τ and the significant wave height SWH. An altimeter waveform denoted as $s(t)$ can be classically written as

$$s(t) = \frac{P}{2}\left[1 + \text{erf}\left(\frac{t - \tau - \alpha\sigma_c^2}{\sqrt{2}\sigma_c}\right)\right] e^{-\alpha\left(t-\tau-\frac{\alpha\sigma_c^2}{2}\right)} + P_i \tag{1}$$

where

$$\sigma_c^2 = \left(\frac{\text{SWH}}{2c}\right)^2 + \sigma_p^2,$$

$\text{erf}\,(t) = \frac{2}{\sqrt{\pi}}\int_0^t e^{-z^2}\,dz$ stands for the Gaussian error function, c denotes the speed of light, α and σ_p^2 are two known parameters (depending on the satellite and on the altimeter) and P_i is the instrument thermal noise. The thermal noise can be classically estimated from the first data samples of $s(t)$ and subtracted from (1). As a consequence, the additive noise P_i can be removed from the model (1) with very good approximation. The received signal is sampled with the sampling period T_s, yielding

$$T_1(n) = \frac{P}{2}\left[1 + \text{erf}\left(\frac{u_n - \tau - \gamma\text{SWH}^2}{\sqrt{2}\sqrt{\mu\text{SWH}^2 + \sigma_p^2}}\right)\right] e^{v_n + \alpha\tau + \delta H^2}, \tag{2}$$

where $T_1(n) = s(nT_s) - P_i$ and the following notations have been used

$$\begin{aligned} u_n &= nT_s - \alpha\sigma_p^2, \quad v_n = -\alpha k T_s + \frac{\alpha^2\sigma_p^2}{2}, \\ \gamma &= \frac{\alpha}{4c^2}, \quad \mu = \frac{1}{4c^2}, \quad \delta = \frac{\alpha^2}{8c^2}. \end{aligned}$$

Note that the parameter P in (2) represents the amplitude of the waveform, the epoch τ corresponds to the central point of the "leading edge", while the significant wave height SWH is related to the slope of the "leading edge". The three parameters P, τ, SWH can be estimated from any altimetric signal from class ω_1 using the maximum likelihood estimator (MLE) [5]. The mean square error between the received altimetric signal and the estimated template $\boldsymbol{T}_1$ (obtained after replacing the unknown parameters P, τ, SWH by their MLEs) will be denoted as MSE.

The proposed pattern recognition system contains three different components referred to as anomaly detection, feature extraction and Bayesian classification. These components are detailed in the following subsections.

2.1. Anomaly detection

Anomaly detection has received a great attention in the literature (see for instance the recent survey of Chandola [6] and references therein). This paper concentrates on the one-class support vector machine (SVM) method [7, Chap. 8], [8] that has shown interesting properties in many applications. These applications include document classification [9] and audio signal segmentation [10]. The one-class SVM method is used here as a way of isolating Brown echoes (class ω_1) from abnormal echoes departing from the Brown model (classes $\omega_2, ..., \omega_{14}$). This step is interesting since it allows one to isolate very fast the large number of echoes that can be represented accurately by the Brown model. Only echoes declared as abnormal will enter the feature extraction and Bayesian classification blocks.

The anomaly detection procedure considered in this section associates to any altimetric waveform a 3 dimensional vector $\boldsymbol{x} = (P, \tau, \mathrm{SWH})$ composed of the altimetric signal amplitude P, epoch τ and significant wave height SWH. A training set $\chi = \{\boldsymbol{x_1}, ..., \boldsymbol{x_{N_t}}\}$ composed of N_t signals associated to class ω_1 is supposed to be available. This training set contains Brown echoes associated to real signals backscattered by ocean surfaces.

The first step of the one-class SVM approach maps the training data vectors into a feature space F via an appropriate transformation Φ. The transformation Φ is chosen such that the inner product between two transformed vectors $\Phi(\boldsymbol{x})$ and $\Phi(\boldsymbol{y})$ defines a kernel $k(\boldsymbol{x}, \boldsymbol{y}) = \langle \Phi(\boldsymbol{x}), \Phi(\boldsymbol{y}) \rangle$. This paper focuses on the Gaussian kernel defined as

$$k(\boldsymbol{x}, \boldsymbol{y}) = e^{-\frac{\|\boldsymbol{x}-\boldsymbol{y}\|^2}{\sigma^2}} \tag{3}$$

where the kernel parameter σ^2 has been optimized using the kernel-alignment criterion developed in [11].

The second step of the one-class SVM method determines a separating hyperplane between the data vectors of class ω_1 and the anomalies (belonging to classes $\omega_2, ..., \omega_K$). The separating hyperplane is the set of vectors $\boldsymbol{x}$ satisfying the equation $\langle w, \Phi(\boldsymbol{x}) \rangle - \rho = 0$. It is classically determined by minimizing the following criterion [8]

$$\frac{1}{2}\|w\|^2 + \frac{1}{\nu N_t}\sum_{i=1}^{N_t} \xi_i - \rho$$

for $w \in F, \rho \in \mathbb{R}$ and $\boldsymbol{\xi} = (\xi_1, ..., \xi_{N_t}) \in \mathbb{R}^{N_t}$ with the constraints $\xi_i \geq 0$ and $\langle w, \Phi(\boldsymbol{x_i}) \rangle \geq \rho - \xi_i$ for $i = 1, ..., N_t$. The slack variables ξ_i account for possible errors in the anomaly detection procedure. Indeed, $\xi_i > 0$ means there is an error in the classification of the training vector $\boldsymbol{x_i}$ whereas $\xi_i = 0$ means the vector $\boldsymbol{x_i}$ has been classified without error. The value of parameter ν is related to the fraction of possible outliers as discussed in [8].

2.2. Feature extraction

After the anomaly detection step, Brown echoes belonging to class ω_1 have been isolated (more than 95% of ocean waveforms should be classified as Brown echoes). The second step of the proposed pattern recognition system consists of classifying the remaining signals (which have not been identified as Brown echoes) in the $K - 1$ classes $\omega_2, ..., \omega_K$. The present study concentrates on altimetric waveforms registered by the Jason-2 satellite. Many features can be computed from an altimetric waveform for classification purposes. These features include statistical moments (mean, variance, skewness, kurtosis, ...), parameters related to the Brown model (significant wave height, backscatter coefficient, ...) or features related to the shape of the altimetric waveform (peakiness, rise time of the echo, ...). The resulting parameter vector will be detailed carefully in the final version of this paper. Following the ideas developed in [1], we propose to extract pertinent information from these features by using linear discriminant analysis (LDA). LDA consists of projecting any data vector $\boldsymbol{\theta}$ (containing the parameters of interest) onto appropriate axes (called discriminant axes). The resulting projected feature vector will be denoted as $\boldsymbol{\theta}_p$. The discriminant axes are defined as the eigenvectors $\boldsymbol{w}$ associated to the non zero eigenvalues of the following generalized eigenvalue problem

$$\boldsymbol{S}_B \boldsymbol{w} = \lambda \boldsymbol{S}_W \boldsymbol{w}, \tag{4}$$

where S_B and S_W are the between-class and within-class scatter matrices defined as

$$\boldsymbol{S}_B = \sum_{i=2}^{K} n_i \left(\boldsymbol{m}_i - \boldsymbol{m}\right) \left(\boldsymbol{m}_i - \boldsymbol{m}\right)^T, \quad \boldsymbol{S}_W = \sum_{i=2}^{K} \sum_{\boldsymbol{\theta} \in \Theta_i} (\boldsymbol{\theta} - \boldsymbol{m}_i)(\boldsymbol{\theta} - \boldsymbol{m}_i)^T,$$

and where Θ_i is the subset of the learning set containing the parameter vectors associated to the class ω_i, $\boldsymbol{m}_i$ is the average of these parameter vectors and $\boldsymbol{m} = \frac{1}{n} \sum_{i=2}^{K} n_i \boldsymbol{m}_i$ is the total mean vector with $n = \sum_{i=2}^{K} n_i$ (see [12, p. 117] for more details).

2.3. Bayes decision rule

The Bayesian classifier (BC) is optimal in the sense that it minimizes the probability of classification error (or an appropriate risk [12, p. 25]). The BC requires to define a loss function summarizing the cost of the different classification errors. In the case of a zero-one loss function (i.e., no loss to correct decisions and unit loss to any error), the BC reduces to the maximum a posteriori (MAP) rule which assigns a given waveform defined by the parameter vector $\boldsymbol{\theta}_p$ to class ω_i if

$$f\left(\boldsymbol{\theta}_p | \omega_i\right) P(\omega_i) > f\left(\boldsymbol{\theta}_p | \omega_j\right) P(\omega_j) \quad \text{for all } j \neq i$$

where $P(\omega_i)$ is the prior probability of the class ω_i and $f\left(\boldsymbol{\theta}_p | \omega_i\right)$ is the probability density function (pdf) of $\boldsymbol{\theta}_p$ conditional to the class ω_i. This study assumes that the different classes are equally likely (i.e., $P(\omega_j) = 1/(K-1)$ for all $j = 2, ..., K$). In this case, the BC reduces to the maximum likelihood classifier. The maximum likelihood classifier assigns $\boldsymbol{\theta}_p$ to class ω_i if $f\left(\boldsymbol{\theta}_p | \omega_i\right) > f\left(\boldsymbol{\theta}_p | \omega_j\right)$ for all $j \neq i$. We assume that the conditional pdfs $f\left(\boldsymbol{\theta}_p | \omega_i\right)$ are Gaussian (this assumption has been validated using different learning sets and will be illustrated in the final paper). Note that the statistical properties of the observed altimetric signals are more difficult to determine (the template is corrupted by multiplicative speckle noise with gamma distribution and by additive Gaussian noise). Thus, it is more complicated to derive the Bayesian classifier based directly on the altimetric signals.

3. SIMULATION RESULTS

Many experiments have been conducted to validate the proposed shape classification strategy. Because of space limitations, we concentrate in this summary on classification results obtained after anomaly detection and feature extraction (these two steps will be detailed in the final paper). These results have been obtained from a signal database constructed from Jason-2 altimetric signals. More precisely, $n_i = 40$ echoes have been manually selected for each class ($i = 1, ..., 14$) resulting in a total of $n_{\text{total}} = 560$ signals. The confusion matrix displayed in Table 1 shows the percentages of signals classified in each class. This confusion matrix has been obtained using the "Leave-One-Out" method [12, p. 485]. More precisely, $n_{\text{total}} - 1$ signals are used to train the classifier and the remaining signal is classified using the proposed classification strategy (feature selection + LDA + Bayesian rule). This operation is repeated n_{total} times and the confusion matrix is obtained after averaging the n_{total} classification results. The results depicted in Table 1 show the good performance of the proposed pattern recognition system for classifying shapes of altimetric signals.

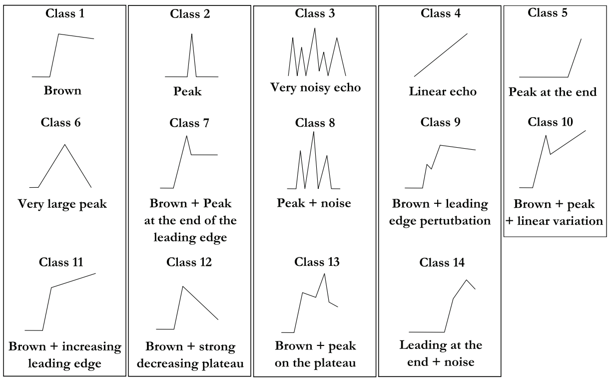

Fig. 1. Different shapes of altimetric signals to be classified.

	Predicted Class												
Actual Class ω_i	2	3	4	5	6	7	8	9	10	11	12	13	14
2	98	0	0	2	0	0	0	0	0	0	0	0	0
3	0	86	2	2	0	7	0	2	0	0	0	0	0
4	0	3	88	3	0	3	3	0	0	0	0	0	0
5	0	0	0	98	0	0	0	0	0	0	0	0	0
6	0	0	0	0	65	33	0	0	0	0	0	0	0
7	0	0	0	12	0	83	0	0	0	0	0	2	0
8	0	0	0	0	0	5	82	2	0	7	0	2	0
9	0	0	0	0	0	0	11	85	0	0	0	2	0
10	0	0	0	0	0	14	2	0	76	0	2	5	0
11	0	0	0	0	0	10	2	0	0	85	0	0	0
12	0	0	0	2	0	2	0	0	0	0	96	0	0
13	0	0	0	3	0	8	0	0	3	0	0	88	0
14	0	0	0	7	0	0	0	0	0	0	0	0	93

Table 1. Confusion matrix after anomaly detection.

4. CONCLUSIONS

This paper studied a pattern recognition system for classifying different shapes of altimetric signals. The system consisted of three steps, i.e., anomaly detection, feature extraction and Bayesian classification. The results obtained with the proposed system on real JASON-2 altimetric data are promising. The final paper will include a comparison with a classification strategy based on neural networks considered within the frame of the CNES project called PISTACH [4]. This project was aimed to improve altimetry products over coastal and hydrological areas.

5. REFERENCES

[1] J.-Y. Tourneret, C. Mailhes, L. Amarouche, and N. Steunou, "Classification of altimetric signals using linear discriminant analysis," in *Proc. IEEE IGARSS*, Boston, MA, 2008, vol. 3, pp. 75–78.

[2] G. S. Brown, "Average impulse response of a rought surfaces and its applications," *IEEE Trans. Antennas and Propagation*, vol. 25, no. 1, pp. 67–74, Jan. 1977.

[3] G. Hayne, "Radar altimeter mean return waveforms from near-normal-incidence ocean surface scattering," *IEEE Trans. Antennas and Propagation*, vol. 28, no. 5, pp. 687–692, Sept. 1980.

[4] P. Thibaut, "Contribution from pistach retracking team," in *Coastal Altimetry Workshop*, Pisa, Italy, Nov. 2007.

[5] J.-P. Dumont, *Estimation optimale des paramètres altimétriques des signaux radar Poseidon*, Ph.D. thesis, Institut National Polytechnique de Toulouse, Toulouse, France, 1985.

[6] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Computing Surveys*, vol. 41, no. 3, pp. 308–313, 2009.

[7] A. Smola and B. Scholkopf, *Learning with Kernels*, MIT press, Cambridge, MA, USA, 2002.

[8] B. Scholkopf, J. C. Platt, J. Shawe-Taylor, A. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Computation*, vol. 13, no. 7, pp. 1443–1471, 2001.

[9] L. M. Manevitz and M. Yousef, "One-class SVMs for document classification," *Journal of Machine Learning Research*, vol. 2, pp. 139–154, 2002.

[10] M. Davy and S. Godsill, "Detection of abrupt spectral changes using support vector machines. an application to audio signal segmentation," in *Proc. IEEE ICASSP-02*, Orlando, FL, USA, May 2002, pp. 1313–1316.

[11] N. Cristianini, J. Kandola, A. Elisseeff, and J. Shawe-Taylor, "On kernel-target alignment," in *Advances in Neural Information Processing Systems 14*, 2002, vol. 14, pp. 367–373.

[12] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, Wiley-Interscience, New York, 2nd edition, 2000.