LIKELIHOOD RATIO STATISTIC FOR INFERRING THE UNCERTAINTY OF SATELLITE DERIVED BATHYMETRY

Sicot Guillaume[⋆] *Ghannami Mohamed Ali*[⋆] *Lennon Marc*[†] *Thomas Nicolas*[†] *Loyer Sophie*[‡]

[⋆] ENSTA Bretagne - M^3 Team - Lab-STICC, UMR-CNRS 6285, Brest, France
[†]Hytech-imaging, F-29200 Brest, France
[‡]Shom, Brest, France

ABSTRACT

We consider the study of performing likelihood-based inference in Lee's radiative transfer semi-analytical model. This model is widely used for the inversion of bathymetric products in coastal areas. We perform assessment of uncertainty on retrieved bathymetry in terms of hypothesis testing. Uncertainty results are compared to Monte Carlo simulations showing that inference based on the likelihood ratio statistic is much more reliable than the Wald statistic. Our results suggest that the Wald statistic based on Fisher observed information outperforms the Cramèr-Rao lower bound in evaluating the estimation uncertainty. Confidence intervals derived from likelihood ratio statistic reveal the asymmetry of the maximum likelihood uncertainty arguing that variance-based uncertainties may be irrelevant for inferring uncertainty on bathymetric products.

Index Terms— Radiative transfer, bathymetry, uncertainty, Likelihood inference

1. INTRODUCTION

Since the seminal work introduced by Lee and al 1998 [1], satellite multi-spectral imagery has been widely exploited to derive simultaneously, various products (bathymetry, bottom classification, chlorophyll-A concentration ..) in coastal areas. The inversion of these geophysical parameters from remote-sensing reflectance measured above the surface Rrs is achieved through a semi-analytical radiative transfer model. This approach has the virtue of not requiring in-situ measurements which are mandatory for calibrating empirical methods. Uncertainty on bathymetry retrievals from satellite imagery is a crucial information for performing quality control on bathymetric products and data assimilation methods. Various studies have been carried out in order to assess uncertainty of ocean colour parameters but only few ones have studied the propagation of R_{rs} uncertainty through the inversion of Lee's model on the retrieved depth. In particular, [2] has investigated minimum bounds of uncertainty on retrieved parameters and demonstrated that the Maximum Likelihood Estimator (MLE) optimally achieves these bounds. However, this optimal behaviour is only guaranteed asymptotically for large samples[3], whereas in practice, the uncertainty of the estimator is required at the pixel level for a single spectral observation. To our knowledge, no research has been conducted to evaluate the performance of test statistics derived from likelihood asymptotic inference theory as to yielding uncertainty on depth estimation. This paper provides a preliminary study of the contribution of first order likelihood inference in constructing confidence intervals on bathymetric retrievals both from hyperspectral and multispectral data.

Section 2 describes Lee's model for light propagation in shallow waters, as well as the statistical modelling approach and inference methods applied in this study. Simulations and methods are described in section 3. We discuss results in terms of contributions, limitations and future perspectives of our research in Section 4.

2. MODELLING APPROACH AND INFERENCE

We introduce a new notation which is more relevant in the statistical inference framework. In our study, we consider a non-linear model $\eta(.)$ which maps a parameter vector θ consisting of water column parameters to the data space.

The physical model η is designed to explain a set of observations at various wavelengths λ_i through a given parametrisation θ.

$$\eta(\theta) = [\eta(\theta, \lambda_1), \eta(\theta, \lambda_2), ..\eta(\theta, \lambda_m)]$$

This model η corresponds to the radiative transfer model we introduce in the next section.

2.1. Physical model

We use the semi-analytical model developed in Lee *et al*. 1999. The model describes the remote-sensing reflectance under the water surface as the sum of two contributions: water column reflectance and bottom reflectance; both reflectance spectra are absorbed exponentially with the water column depth z.

$$r_{rs}(\lambda) = r_{rs}^{dp}(\lambda)\{1 - e^{-K_u^C(\lambda).z}\} + \frac{\rho(\lambda)}{\pi} e^{-K_u^B(\lambda).z} \quad (1)$$

The different attenuation coefficients $K_u^C(\lambda), K_u^B(\lambda)$ are expressed as functions of 3 optically active water constituents : chlorophyll-A concentration denoted chl, color dissolved organic matter denoted $cdom$ and finally total suspended matter indicated as tsm. The complete model that maps the parameters $(chl, cdom, tsm)$ to the attenuation coefficients is based on the developments in [4],[5] and [6].

The seabed reflectance $\rho(\lambda)$ is modeled as a sum-to-one linear mix of sand and vegetation reflectances : $\rho(\lambda) = \alpha_M r_{sand}(\lambda) + (1 - \alpha_M) r_{veg}(\lambda)$.

The remote-sensing reflectance above the surface $R_{rs}(\lambda)$ is modeled as :

$$R_{rs}(\lambda) = \frac{\zeta^{dp} r_{rs}(\lambda)}{1 - \gamma^{dp} r_{rs}(\lambda)} \quad (2)$$

Considering θ = $(chl, cdom, tsm, z, \alpha_M)$, our physical model $\eta(\theta, \lambda)$ is simply the remote-sensing reflectance modeled for the parameter vector θ at wavelengths λ.

2.2. Statistical model and parameter estimation

The evaluation of the uncertainty in the bathymetry requires prior knowledge of the noise affected by the measurement, which takes into consideration instrument errors and environmental noise. The mathematical description of the noise makes it possible to implement a statistical model, supporting the propagation of the measurement uncertainty up to the estimated parameters. Previous studies [2],[7] assumed an unbiased measurement with a Gaussian distribution on the remote-sensing reflectance products. Likewise, we model the measured subsurface reflectance as a vector $y = [y(\lambda_1), y(\lambda_1), ..., y(\lambda_N)]$ that follows a multivariate Gaussian distribution :

$$y(\lambda) = \eta(\theta) + \epsilon$$

Where ϵ is a realization of a zero mean Gaussian noise described by the covariance matrix $\Sigma = E[\epsilon.\epsilon^T] = E[(y - \eta(\theta))(y - \eta(\theta))^T]$.

The likelihood function of a parameter vector θ is defined as the joint density function of the data is:

$$L(\theta; y) = \frac{1}{\left(\sqrt{2\pi}\Sigma\right)^N} \exp - \frac{[\eta(\theta) - y]^T \Sigma^{-1} [\eta(\theta) - y]}{2} \quad (3)$$

For a given set of observations y, an optimal estimate $\hat{\theta}(y)$ can be computed, such that $\eta(\hat{\theta})$ fits the data y the most. The log-likelihood is defined as the logarithm of the likelihood $l(\theta; y) = log(L(\theta; y))$. This study focuses on likelihood-based inference, therefore the MLE is our estimator. That is the parameter $\hat{\theta}$ that maximizes the likelihood 2.2 for a given observation y over the parameter space.

$$\hat{\theta}(y) = \underset{\theta}{\operatorname{argmax}}\, l(\theta; y)$$

As the sample size tends to infinity, the MLE has a an asymptotic normal distribution centered on the true value of the unknown parameter [8]. In addition, it has the minimum variance achieved among unbiased estimators which is equivalent to the Cramèr-Rao lower bound. Such attractive properties have made the MLE a very popular estimator.

For an additive Gaussian noise, it can be easily shown that the MLE is equivalent to minimizing the Mahalanobis quadratic distance:

$$\hat{\theta}(y) = \underset{\theta}{\operatorname{argmin}} (y - \eta(\theta))^t \Sigma^{-1} (y - \eta(\theta))$$

Next section addresses the theoretical background which deals with evaluating confidence intervals derived for this estimator.

2.3. Likelihood inference and confidence intervals

In statistical inference both the measurements y and the estimator $\hat{\theta}$ are considered as realizations of random variables. The unknown θ is considered as a deterministic variable. The aim of statistical inference is to draw statistical conclusions about the estimate. Essentially, likelihood inference seeks to evaluate the propagation of the measurement noise to the retrieved parameters by describing the probability distribution of the MLE. It is a standard measure to carry out inference on a single parameter of interest regardless of the other parameters known as nuisance parameters usually denoted by λ. In our study the water column depth z is the parameter of interest and λ corresponds to the rest of the parameters. Hence $\theta = (z, \lambda)$ with λ = $(chl, cdom, tsm, \alpha_M)$. The nuisance parameters, are in a way, marginalized by replacing λ by the maximum likelihood $\hat{\lambda}(z)$ for a fixed z in the likelihood related quantities. This gives rise to the profile likelihood defined as:

$$l_p(z) = \max_{\lambda} l(z, \lambda) = l(z, \hat{\lambda}(z)) \quad (4)$$

In a sense, knowledge about the parameter z is summarized in the profile likelihood as corresponding optimal values for λ are traced in the parameter space.

We also define the matrix j as the observed Fisher information matrix, *i.e* the opposite of the hessian matrix of the log-likelihood $-l_{\theta\theta}(\theta; y)$. It can be shown that for additive multivariate Gaussian noise:

$$j_{uv}(\theta; y) = (\eta(\theta) - y)^T \Sigma^{-1} \frac{\partial^2 \eta}{\partial \theta_u \theta_v} + i_{uv}(\theta; y) \quad (5)$$

u, v denote the index of the element located in the u^{th} row and the v^{th} column of the j matrix. $i = E[j(\hat{\theta}; y]$ is the expected Fisher information which is given by $\frac{\partial \eta}{\partial \theta}^T \Sigma^{-1} \frac{\partial \eta}{\partial \theta}$ [8]. We also define the profile observed information function for the parameter of interest as $j_p(z) = -\partial^2 l_p(z)/\partial z \partial z^T$. $j_p(z)$ can be computed directly from the inverse of the full observed information matrix j as $j_p(z) = \{[j(z, \hat{\lambda}(z))]^{-1}\}_{zz}$ where zz indicates the diagonal value at parameter z. Likelihood inference for a single parameter of interest, in our case z, is generally based on the following statistics [9]:

$$t(z) = j_p(\hat{z})^{1/2}\{\hat{z} - z\}, \tag{6}$$

$$r(z) = sign(\hat{z} - z)\sqrt{2\{l_p(\hat{z}) - l_p(z)\}} \tag{7}$$

If some regularity conditions are satisfied and for large samples, these statistics under the true hypothesis are characterized by an asymptotic standard normal distribution. The statistic t , called Wald statistic relies on the Fisher observed information j which can also be replaced by the expected information i. The Wald statistic is another way of stating the asymptotic behavior of the MLE estimator for large samples.

The statistic r is commonly called the likelihood-ratio statistic. In contrast to the Wald statistic based on local curvature of the log-likelihood, the likelihood ratio statistic r yields global information on the parameter z.

These so-called statistics are typically applied in hypothesis testing (deriving p-values) and confidence intervals construction. For a given significance level α , the $(1-\alpha)\%$ confidence interval is provided by the values of z that translates r and t in the quantiles $\{q_{\alpha/2}, q_{1-\alpha/2}\}$ of the standard normal distribution. For the Wald statistic, this is straightforward as the interval can be directly computed from $\hat{\theta}$ and $j_p(\hat{z})$ as $\hat{\theta} \pm q_{\alpha/2}.j_p(\hat{z})^{\frac{-1}{2}}$. On the other hand, in the likelihood ratio statistic, no analytical solution exists for (7) and confidence intervals require a nonlinear optimization in order to compute l_p, trace r numerically over a suitable grid $[z^-, z^+]$ and interpolate the resulting curve $r(z)$ at the values $q_{\alpha/2}, q_{1-\alpha/2}$.

3. DATA AND METHODS

In this section we describe the different simulations intended to analyse and validate our results. Optimisation method is briefly presented before addressing the assessment of the different test statistics.

3.1. Simulation

Simulated spectra can be generated from noise covariance matrix Σ and a given parametrisation of the model $\eta(\theta)$. Two types of measurements are simulated: hyperspectral and multispectral. The former is based on an ideal sensor with 61 bands ranging from 400 nm to 1000 nm with a resolution of 10 nm. The latter is constructed from the former thanks to Sentinel-2 MultiSpectral Instrument (MSI) Spectral Response Functions (S2-SRF) provided by the European Space Agency (ESA). Only the first 7 bands were considered in simulating multispectral data. The viewing zenith angle and the subsurface solar zenith angle were set to 0° and 30°, respectively. In order to evaluate the quality of the derived uncertainty with respect to different environmental conditions, different water column and seabed scenarios presented in table 1 are investigated. The influence of chl and $cdom$ is not adressed in the analysis. The tsm parameter is varied giving rise to two types of water columns: clear waters (**CW** : $tsm = 0.1g.m^{-3}$) and relatively turbid waters (**CW** : $tsm = 3g.m^{-3}$). Additionally, two types of seabed are considered: a relatively dark bottom (**DB** : 25% of sand versus 75% of vegetation), and a relatively white bottom (**WB** : 75% of sand versus 25% of vegetation).

Parameters	Chl (mg/m^3)	CDOM(m^{-1})	TSM(g/m^3)	α
CWWB	0.03	0.01	0.1	0.75
TWWB	0.3	0.01	3	0.75
CWDB	0.03	0.01	0.1	0.25
TWDB	0.3	0.01	3	0.25

Table 1. Parameters of the water column used in analysis.

As for the noise covariance matrix, we consider a spectrally uncorrelated noise with identical magnitude $5.10^{-4}\%$ on the remote-sensing reflectance. That is to say : $= \sigma^2.I_m$ with $\sigma = 5.10^{-4}$, I_m is the identity matrix and m is the number of bands.

3.2. Optimisation

We use the local optimizer based on trust-region method supplied in **scipy least-square** package to optimize for the MLE in equation 2.2 [10]. Initial tests prove this method is superior to the textbfscipy Levenberg–Marquardt implementation which cannot incorporate bounds on parameters. We consider the following subset of $\mathcal{R}^5 : [0, 5] \times [0, 1] \times [0, 10] \times [0, 30] \times [0, 1]$ for bounding the optimisation. Throughout the analysis, optimization is performed using a fixed initialisation point $\theta_{init} : \{0.01, 0.01, 0.01, 1, 0.5\}$. The same optimization procedure is applied to derive the profile likelihood in 4.

3.3. Evaluation of the test statistics

We simulate water types described in table 1, as well as for both sensors : hyperspectral and Sentinel 2. For a given depth z_t, 5000 noisy spectra are generated from the model mean in order to compute 95% confidence intervals and validate the theoretical results.

The performance of likelihood-based inference is evaluated by means of simulation compared to hypothesis testing results.

We evaluate the performance of the test statistic in rejecting the null hypothesis $z = z_t$. Statistical hypothesis testing is another way of perceiving confidence interval and more convenient to compare the performance of different statistics. In short, we compare the null hypothesis $z = z_t$ with the alternative hypothesis $z = \hat{z}$ through the test statistics r and t. The probabilities of observing at least as extreme as the test statistics at the null hypothesis ($r(z_t)$ and $t(z_t)$) are called p-values. If the p-values are small compared to α, we reject the null hypothesis. Equivalently, the null hypothesis is rejected, if for a given significance level, the test statistic at the null hypothesis is not within the range $q_{\alpha/2}, q_{1-\alpha/2}$. For each parametrisation in table 1 and each sensor, we compute the null hypothesis rejection rate for the confidence level 95%. Ideally, it is desired to have rejection rates close to the nominal value 5%.

We also pay special attention to the profile likelihood for the different water-types and how confidence intervals can be derived from the r statistic.

4. RESULTS AND DISCUSSION

In the following, hypothesis testing results for a confidence level of 95% are presented and discussed in order to highlight and compare the performance of the studied test statistics for different configurations. Later, we detail and analyse the likelihood ratio r statistic results.

4.1. Hypothesis testing results

Table 2 summarises simulation results carried out at the nominal level of $\alpha = 5\%$ for the parametrisations presented in table 1 and for two null-hypothesis on the depth parameter : $z_t = 3m$ and $z_t = 10m$.

parametrisation	Hyperspectral			Sentinel 2		
	r	t_O	t_E	r	t_O	t_E
CWWB-3m	4.6%	2.4%	2.3%	4.3%	17.5%	14.9%
CWDB-3m	5%	6%	7.9%	4.2%	13.1%	14.4%
TWWB-3m	3%	4.7%	12.5%	2.5%	7.5%	36.3%
TWDB-3m	2.7 %	4.8 %	19.5%	3.9%	13.4%	34%
CWWB-10m	3.9%	7.2 %	11.9%	2.6 %	3.4%	32.9%
CWDB-10m	3.9%	7.2 %	11.9%	2.6 %	20%	32.9%
TWWB-10m	1.8%	7.1%	22.5%	2.2%	2.6%	64.7%
TWDB-10m	2.24 %	10.4%	25.9%	1 %	42.3%	55.4%

Table 2. Null rejection rates of the tests at a nominal value of $\alpha = 5\%$ for both r statistic and Wald statistics : t_O derived from observed information and t_E derived from expected information.

Results demonstrate that the likelihood ratio r has a better performance than the Wald statistics in most cases yielding rejection rates close to 5%. In general, rejection rates are below the nominal level suggesting that the confidence interval yielded by r is rather conservative. On the other hand, Fisher based statistics tend to be liberal. It is also worth noting that t_O and t_E have their best performance in the hyperspectral case for relatively clear shallow waters. Furthermore, t_O, and t_E show liberal rejection rates as the bottom becomes less visible in the case of "TWDB" for example. For the multispectral case, rejection rates exceed the nominal level in most cases for the Fisher information based statistics. The behaviour of Wald statistics can be explained by the poor quadratic approximation of log-likelihood by the Fisher information matrix. The t_O based on the observed information performs better than t_E providing rejection rates (7%) relatively close to the nominal level. In other words, the realization of the curvature of the log-likelihood is more appropriate than its expectation as to providing robust confidence intervals for clear waters and hyperspectral observation. These results allowed direct comparisons for likelihood ratio statistic against Fisher information-based statistics in different configurations.

4.2. Likelihood profile and r statistic

Figure 1 illustrates the likelihood profile as well as the corresponding r statistic for a simulated hyperspectral observation with **CWWB** water type and a 3m water column depth. Notice that the likelihood has the shape of an inverted parabola around the maximum. The sharper the curvature of the l_p profile, the smaller the confidence interval and the more accurate our estimate.

The bounds of 95% confidence intervals on z are simply the intersection between r and the corresponding confidence values for a standard normal distribution: ± 1.96. Notice that the MLE estimate corresponds to the maximum of the profile likelihood and the zero of r. This is an indicator that z is to some extent identifiable as no other candidate has greater likelihood was found through the profiling process. Empirical 95% quantiles computed from simulation agree with the likelihood ratio derived confidence intervals.

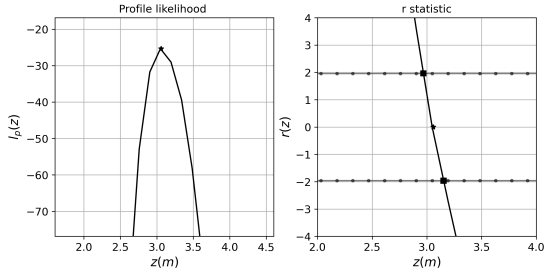

Fig. 1. On the left: profile likelihood for hyperspectral configuration and a depth of 3m simulated observation. On the right panel: construction of the 95% confidence intervals from the corresponding r statistic. The square dots are indicators for the empirical 95% quantiles estimated from simulation.

Both curves of r and l_p can give valuable information on the asymmetry of uncertainty. The interval obtained in this case is nearly symmetrical suggesting that the model behaves well for this specific configuration. The assumption of the

normality of $\hat{z}$ is more or less relevant given the parabolic shape of the likelihood profile.

The asymmetry is strongly pronounced for difficult cases such as turbid waters and dark bottoms where identifiability of parameters in the model is not guaranteed. To highlight this asymmetry, we plot the likelihood profile as well as the corresponding r curve for CWWB" and "TWWB" waters in the figure 2.

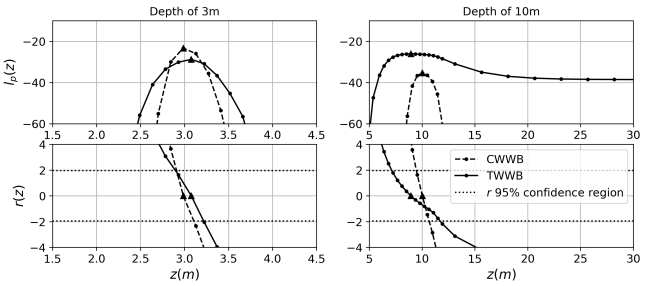

Fig. 2. Profile likelihood for hyperspectral configuration and dark seabed. Triangular markers indicate the location of the MLE

We report that the asymmetry is often related to an upper bound that is farther apart from the estimate compared to the lower bound.

Overall, the intersection between the curve $r(z)$ and the confidence values ± 1.96 was always guaranteed for clear waters in our analysis. For turbid or relatively deep cases, the upper bound may require a computation of the likelihood profile very far from the estimated value $\hat{z}$. Interestingly, we come across cases where the intersection does not exist in the definition domain of z for the upper bound. Figure 3 shows an example of this case for "TWDB" waters with 10m . This case seems to indicate that the model structure allows ensuring that z is greater than a value with a level of confidence, here 95%, but it does not allow to provide an upper bound for this level of confidence.

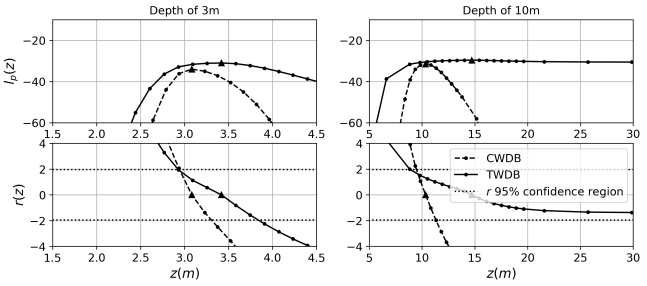

Fig. 3. Profile likelihood for hyperspectral configuration and dark seabed. Triangular markers indicate the location of the MLE

5. CONCLUSION

Likelihood based inference was investigated as a tool to estimate uncertainty on satellite derived bathymetry. The results obtained show that the variances estimated from the Fisher information may be uninformative of the bathymetry uncertainty especially for multispectral data and optically complex waters. Two main reasons explain these results: the complexity of the model and the low sample size. In contrast, the 95% confidence intervals derived from the likelihood ratio yielded an overall better uncertainty estimation for both hyperspectral and multispectral types of measurement. In addition to being operationally conceivable, the method proved to be more appropriate for capturing the asymmetry of the uncertainty and delivering valuable information on the identifiability of the depth parameter.

6. REFERENCES

[1] C.Mobley R. Steward J. Patch Z. Lee, K. Carder K, "Hyperspectral remote sensing for shallow waters: Deriving bottom depths and water properties by optimization," *APPLIED OPTICS*, vol. 38, June 1999.

[2] S. JAY, "Predicting minimum uncertainties in theinversion of ocean color geophysicalparameters based on cramer-rao bounds," *OPTICS EXPRESS*, vol. 26, January 2018.

[3] O.E. Barndorff-Nielsen D.R. Cox, *Inference and Asymptotics*, March 1994.

[4] A. Morel H. Claustre A. Bricaud, M. Babin, "Variability in the chlorophyll-specific absorption coefficients of natural phytoplankton: Analysis and parameterization," *JOURNAL OF GEOPHYSICAL RESEARCH*, vol. 100, pp. 13,322–13,323, July 1995.

[5] G. Ferrari N. Claustre A. Bricaud N. Obolensky GHoepffner M. Babin, D. Stramski, "Variations in the light absorption coefficients of phytoplankton, nonalgal particles, and dis-solved organic matter in coastal waters around europe," *OPTICS EXPRESS*, vol. 108, July 2003.

[6] V. Fournier-Sicre F.Fell D. Stramski M.Babin, A.Morel, "Light scattering properties of marine particles in coastal and open ocean waters as related to the particle mass concentration," *Limnology and Oceanography*, vol. 48, pp. 843–859, 2003.

[7] B. Koetzc S. Phinn J. Hedley, C. Roelfsemab, "Capability of the sentinel 2 mission for tropical coral reef mapping and coral bleaching detection," *Remote Sensing of Environment*, February 2012.

[8] L. Steven M. Kay, *Fundamentals of Statistical Signal Processing: Est imat ion Theory*, 1993.

[9] A. C. Davison A. R. Brazzale and N. Reid, *Applied Asymptotics Case Studies in Small-Sample Statistics*, March 2007.

[10] T. F. Coleman M. A. Branch and Y. Li, "A subspace, interior, and conjugate gradient method for large-scale bound-constrained minimization problems," *SIAM Journal on Scientific Computing*, vol. 100, pp. 1–23, 1999.