

Ambiguities in the inversion of the Ocean Colour: Problems and Solutions

DEFOIN-PLATEL Michael^{1,2} and CHAMI Malik¹

¹Laboratoire d'Océanographie de Villefranche sur Mer
Université Pierre et Marie Curie-Paris6
Quai de la Darse, BP08
06230 Villefranche sur Mer, France
Email: dpm@obs-vlfr.fr

²Laboratoire d'Informatique Signaux et Systèmes de Sophia Antipolis
UNSA-CNRS
2000, route des lucioles, BP121
06903 Sophia Antipolis, France

Presented at SIRC 2006 - The 18th Annual Colloquium of the Spatial Information Research Centre
University of Otago, Dunedin, New Zealand
November 6th-7th 2006

ABSTRACT

The inverse problem of ocean colour consists in deriving the inherent optical properties (IOP) of marine particles from a reflectance spectrum measured at the sea surface. Such a problem is ill-posed or ambiguous because of the non-uniqueness of the solution; *i.e.* several combinations of IOP values can lead to a unique reflectance spectrum. Currently, great efforts are made in the development of inverse methods to accurately retrieve the IOP. However, much less studies have been devoted to the analysis of the ambiguities, which affect yet the error on the IOP retrieval. In this paper, a set of measures suitable to characterize the ambiguities of an inverse problem are introduced and used to describe the ocean colour problem. The ambiguities of the IOP retrieval are evaluated and their implications for inverse modelling are studied. Finally, different strategies to reduce the effects of ambiguities are discussed.

Keywords and phrases: Remote Sensing, Ocean Colour, Inverse Problem

1 INTRODUCTION

The inverse problem of ocean colour consists in determining biogeochemical parameters such as the chlorophyll *a* concentration from the upwelling radiance spectrum. Because the inherent optical properties (IOP) of the particles, namely the scattering, backscattering and absorption coefficients, are at the center of measured water leaving radiance and water constituents, the inverse problem is often examined as a two steps process: the derivation of IOP from the radiance, and then biogeochemical parameters from the IOP. Such an IOP-based inversion maximizes the information gained from remote sensing.

A number of methods for the IOP retrieval of each water component, namely phytoplankton (Chl), nonalgal particles (NAP), inorganic material (sed) and coloured dissolved organic matter (CDOM) from remotely sensed data have been proposed (Morel 1988, Chami & Robilliard 2002, Garver & Siegel 1997, Hoge & Lyon 1996, Roesler & Boss 2003, Schiller & Doerffer 2005). Equation 1 provides an example of relationships that is often used as a basis of many inversion algorithms:

$$R_{rs} = L_u / E_d \approx g \frac{bb_{tot}}{a_{tot}} \quad (1)$$

where R_{rs} is the subsurface remote sensing reflectance, defined by the ratio of nadir-viewed upwelling radiance L_u (in $W m^{-2} sr^{-1}$) to downwelling irradiance E_d (in $W m^{-2}$) just beneath the sea surface, bb_{tot} is the total backscattering coefficient (in m^{-1}), a_{tot} is the total absorption coefficient (in m^{-1}) and g is a proportionality factor (sr^{-1}). The major difficulty of solving the inverse problem of ocean colour rests on the fact that the

relationship between the R_{rs} and the IOP of each water component is mathematically not a bijection. Because of their additive property, IOP of each water component combined in different ways can lead to identical sum of IOP and thus to similar reflectance values. Furthermore, a given value of the ratio bb_{tot}/a_{tot} can be obtained from different values of bb_{tot} and a_{tot} , thus also generating a similar reflectance value as inferred by equation 1. Because the solutions are not unique, the inverse problem of ocean colour is said ambiguous or ill-posed. Therefore, fundamentally, both the inversion algorithm and the presence of ambiguities are sources of error in the retrieved IOP. Note that the error due to the occurrence of ambiguities also corresponds to the minimum error that can be made when inverting the reflectance, whatever the algorithm employed. It is the purpose of this paper to deeply analyze the non-uniqueness of the solution of the inversion problem and the resulting implications in terms of accuracy of IOP retrieval.

The paper is organized as follows. First, the radiative transfer simulations are described. Second, formal mathematical definitions of the ambiguities are provided and the distribution of the non-uniqueness of the solution within the dataset is characterized. Then the impact of the ambiguities on the accuracy of the retrieved optical parameters is evaluated. Finally, several strategies that might be relevant to tackle the ambiguity problem are discussed.

2 SIMULATION OF A SYNTHETIC DATASET

A great part controlling the success of an algorithm developed to address the inverse problem of ocean colour is its ability to correctly represent real-world conditions. Here, a synthetic dataset is used.

The simulations were performed using the OSOA radiative transfer model (Chami, Santer & Dilligeard 2001). The OSOA model solves the vector radiative transfer equation for the coupled atmosphere-ocean system using the successive orders of scattering method. Given a set of inherent optical properties in the water column, the OSOA model outputs the angular distribution of the radiance field. Standard atmosphere with tropospheric aerosols T70 (Shettle & Fenn 1979) having an optical depth of 0.2 at 555 nm (*i.e.* horizontal visibility of 23 km) was used to simulate the incoming solar light. The solar zenith angle was set up to 30 degrees. The nadir-viewed remote sensing reflectance was simulated just beneath the surface. The wavelengths used to compute the R_{rs} were: 412 nm, 443 nm, 490 nm, 510 nm, 555 nm, 620 nm and 665 nm. The set of wavelengths is referred to as \mathcal{L} . The inherent optical properties required for the simulation are the absorption coefficients, the scattering coefficients and the phase function of the particles. For computation of the IOP, a five components seawater model is considered. The five components are pure seawater, phytoplankton and their covarying particles (Chl), colored dissolved organic matter (CDOM), nonalgal particles (NAP) and inorganic particles (sed). In the notation, IOP_v is a vector composed by each component of the IOP, such that $IOP_v = \{a_{chl}, a_{CDOM}, a_{NAP}, b_{chl}, b_{sed}, bb_{chl}, bb_{sed}\}$. In this water model, the total absorption coefficient a_{tot} is the sum of contributions from all components. Since CDOM is supposed to be non-scattering material, the total scattering coefficient b_{tot} and backscattering coefficient bb_{tot} are the sum of all other components. We note also a_p , b_p and bb_p , respectively the absorption, scattering and backscattering coefficients of particles. The phase functions of particles are computed using Mie theory and different refractive indices for the phytoplankton and inorganic particles.

To feed the OSOA model with several thousands sets of IOP vectors a random assignment of each IOP value is performed. Nevertheless two different types of constraints have to be taken into account:

- the realistic constraints arise from the optical and biogeochemical properties of the oceanic constituents. Typically, these constraints consist in the selection of appropriate ranges of variation for each IOP, ranges of co-variation between IOP but also appropriate spectral dependencies for each IOP. The realistic constraints have to be respected to generate synthetic spectra similar to natural ones.
- the statistical constraints require that, in a ideal dataset, all the variables within the samples should be independent and identically distributed (Vapnik 1995). The notation *iid* is hereafter used to specify a dataset which is independent and identically distributed.

A huge number of reflectance spectra are simulated to cover a large variety of optical waters conditions. For convenience, the synthetic dataset containing $N = 10,000$ couples (R_{rs} , IOP) is hereafter referred to as DPC_{K2} .

3 AMBIGUITY OF THE INVERSE PROBLEM

The inverse problem of ocean colour is very complex mainly because this is an ill-posed problem with many ambiguities. Intuitively we understand that the ambiguous samples of a dataset are that for which several combinations of IOP values correspond to one “unique” R_{rs} spectrum. As an example, figure 1 shows the variations in the R_{rs} at 555 nm as a function of the total absorption, scattering and backscattering coefficient. It is clearly

observed that identical values of Rrs are obtained for different combinations of total IOP . It is obvious that this situation is problematic for inverse models since they may not be able to discriminate several similar Rrs spectra and so to output the desired IOP . Therefore one of the major issues of the inverse problem of ocean colour consists in a deep characterization of the ambiguities.

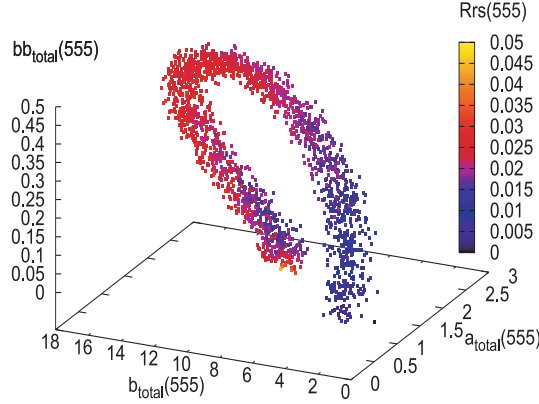


Figure 1: Example of ambiguities in the inverse problem of ocean colour. To a unique Rrs (555) value may correspond several triplets $(a_{tot}(555), b_{tot}(555), bb_{tot}(555))$. The units are: sr^{-1} for Rrs and m^{-1} for a_{tot} , b_{tot} and bb_{tot} .

3.1 Formal definitions of ambiguity

In order to perform a thorough analysis of the ambiguities present within the dataset, a formalism defining the ambiguities is required.

An objective measure of similarity is needed to determine uniqueness of several Rrs spectra. For that purpose, we define the *spectral distance* δ_{Rrs} between two Rrs spectra Rrs_1 and Rrs_2 :

$$\delta_{Rrs} = \max_{\lambda \in \mathcal{L}} \left(\left| \frac{Rrs_1 - Rrs_2}{\frac{1}{2}(Rrs_1 + Rrs_2)} \right| \right) \quad (2)$$

The equation 2 corresponds to the maximum of the absolute difference between Rrs_1 and Rrs_2 normalised to the average over these two spectra and computed for all wavelengths in \mathcal{L} . We note that this measure is symmetric and varies, in this study, in the range $[0, 2]$. Similarly, we define the spectral distance δ_{IOP} between two sets of IOP vector:

$$\delta_{IOP} = \max_{p \in IOP} \left(\max_{\lambda \in \mathcal{L}} \left(\left| \frac{IOP_{v1,p} - IOP_{v2,p}}{\frac{1}{2}(IOP_{v1,p} + IOP_{v2,p})} \right| \right) \right) \quad (3)$$

In equation 3, p is one element of the IOP vector (*i.e.* the absorption, scattering or backscattering coefficient of one of the seawater component). It is also possible to compute a spectral distance for each element of the IOP vector. This means that the maximum value in equation 3 is only searched over the set of wavelengths and not over the entire set of IOP_v . The spectral distance defined for each element of the IOP vector is referred using the notation of this element. As an example, the b_{chl} spectral distance between IOP_{v1} and IOP_{v2} is referred to as $\delta_{b_{chl}}(IOP_{v1}, IOP_{v2})$.

We define the set of *spectral neighbourhood* of one Rrs spectrum Rrs_1 within the DPC_{K2} dataset as:

$$N_{DPC_{K2}, \epsilon}(Rrs_1) = \{Rrs_2 | (Rrs_2, IOP_{v2}) \in DPC_{K2}, \delta_{Rrs}(Rrs_1, Rrs_2) < \epsilon\} \quad (4)$$

with a threshold value, typically the uncertainty in the Rrs measurements. In this paper ϵ is fixed to 5%, which is consistent with the accuracy provided by the current commercially available instruments. For clarity, DPC_{K2} and ϵ will be further omitted in the notation. Note that the definition given in equation 4 and other forthcoming equations could be used for any other dataset than DPC_{K2} . By construction all the elements of the set are said to be similar as Rrs_1 and so they will be considered as unique by any inversion algorithm.

We define the *inversion set* of a given Rrs spectrum Rrs_1 as:

$$S(Rrs_1) = \{IOP_{v2} | (Rrs_2, IOP_{v2}) \in DPC_{K2}, Rrs_2 \in N(Rrs_1)\} \quad (5)$$

Table 1: Ambiguity rate and mean ambiguity distance within DPC_{K2}

Data	R(DPC_{K2}) in %	$\bar{\Delta}(DPC_{K2})$ in %
IOP	92.00	165.76
a _{chl}	91.63	114.22
a _{NAP}	91.96	129.50
a _{CDOM}	91.38	86.30
b _{chl}	92.00	151.58
b _{sed}	91.97	124.06
bb _{chl}	91.97	153.04
bb _{sed}	91.92	119.11

The inversion set contains all the IOP vectors IOP_v for which the forward OSOA model outputs RRS spectra that are considered similar as Rrs_1 accordingly to the definition of the spectral neighbourhood of (equation 12).

A spectrum Rrs_1 within the dataset is considered as ambiguous if the spectral distance between the IOP vectors that belong to the inversion set $S(Rrs_1)$ is significant (typically greater than ϵ). To test whether Rrs_1 is ambiguous, the following function is used (equation 6):

$$TA_{IOP}(Rrs_1) = \begin{cases} true & \exists IOP_{v2}, IOP_{v3} \in S(Rrs_1) \mid \delta_{IOP}(IOP_{v2}, IOP_{v3}) > \epsilon \\ false & \text{otherwise} \end{cases} \quad (6)$$

Therefore, Rrs_1 is considered as an ambiguous spectrum when the predicate $TA_{IOP}(Rrs_1)$, which is called the *spectral ambiguity* of Rrs_1 , is true. It is also possible to test for the spectral ambiguity of Rrs_1 relatively to each of the elements of the IOP vector. As an example, the test for the b_{chl} spectral ambiguity is referred as $TA_{b_{chl}}(Rrs_1)$.

3.2 Analysis of the ambiguities within the dataset

We define the *ambiguity rate* R_{IOP} as the proportion of ambiguous samples within DPC_{K2} (equation 7):

$$R_{IOP} = \frac{1}{N} \sum_{(Rrs_1, IOP_v) \in DPC_{K2}} \begin{cases} 1 & \text{if } TA_{IOP}(Rrs_1) \text{ is } true \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

In equation 7, N is the number of element of DPC_{K2} . The ambiguity rate could be very helpful when trying to compare the ambiguities within different datasets and thus, to evaluate the impact of strategies to tackle the ambiguity problem, which will be discussed later. The ambiguity rates within the dataset DPC_{K2} is calculated for each element of the IOP vector and results are reported in table 1. The ambiguity rates show fairly similar values (around 92%). Therefore, almost all the trustworthy samples of DPC_{K2} have been considered to be ambiguous.

Because the spectral distance between the IOP vectors corresponding to ambiguous RRS samples may be highly variable (*i.e.* all the samples are not equally ambiguous), it is interesting to evaluate the maximum difference that exists between the IOP vectors belonging to the inversion sets. In this objective, we define the *ambiguity distance* Δ_{IOP} of a given RRS spectrum Rrs_1 as follows (equation 8):

$$\Delta_{IOP}(Rrs_1) = \max \{ \delta_{IOP}(IOP_{v1}, IOP_{v2}) \mid IOP_{v1}, IOP_{v2} \in S(Rrs_1) \} \quad (8)$$

We note that it is also possible to compute an ambiguity distance for each element of the IOP vector.

Figure 2 shows three similar RRS spectra. One of them (solid lines) is taken as the reference spectrum Rrs_1 and the two other spectra were selected within the spectral neighbourhood set $N(Rrs_1)$. The corresponding b_{chl} spectra that belong to the inversion set $S(Rrs_1)$ are plotted in figure 3. It is observed that highly different values of b_{chl} can lead to identical RRS values. In this example, the maximum bias between the b_{chl} spectra (*i.e.* the ambiguity distance) reaches 172.8%. Figures 2 and 3 clearly illustrate the necessity of quantifying the ambiguity distance for inverse modelling purpose, as it will be discussed later.

We have computed the mean ambiguity distance, noted $\bar{\Delta}_{IOP}$, over the 10,000 samples of the DPC_{K2} dataset. The mean ambiguity distance has to be understood as the average discrepancy between two sets of IOP_v that lead to similar RRS spectra. The results are reported in table 1. The mean ambiguity distance found for each IOP shows high values. However, a strong variability is observed depending on the inherent optical property that is studied. As expected from its definition, $\bar{\Delta}_{IOP}$ exhibits the highest value. We observe also that despite a considerable difference between the ranges of variation of b_{chl} and bb_{chl}, their mean ambiguity distance are

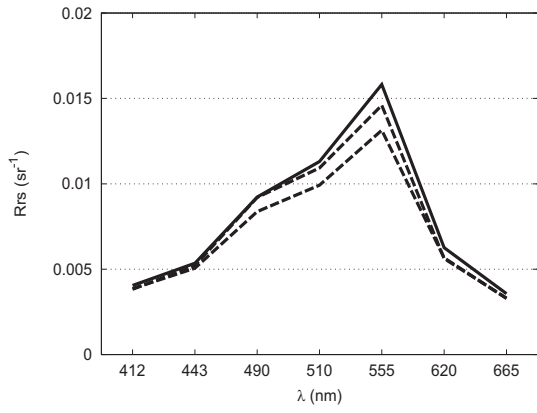


Figure 2: A set of spectral neighbourhood $S(Rrs_1)$ containing three Rrs spectra (sr^{-1}) with Rrs_1 as the solid line.

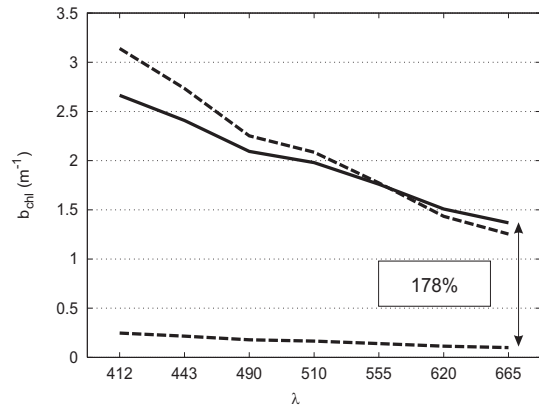


Figure 3: The corresponding inversion set $N(Rrs_1)$ containing three b_{chl} spectra (m^{-1}) with IOP_{v1} as the solid line.

almost equal. A similar remark can be made with regard to b_{sed} and bb_{sed} . This is not surprising since the backscattering coefficient is strongly correlated with the scattering coefficient.

We have computed the distributions of the ambiguity distances for all the samples of DPC_{K2} . One example is presented figure 4. The $\Delta_{a_{CDOM}}$ distribution is nearly uniform thus meaning that there is an equivalent amount of very ambiguous or slightly ambiguous samples. Note that the first class of the histogram in figure 4 corresponds to the unambiguous samples (with) and holds around 8% of the samples of DPC_{K2} , which is consistent with the value of 92% found for the ambiguity rate (table 1).

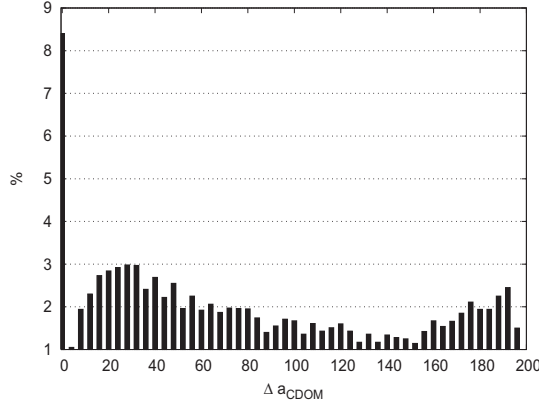


Figure 4: Distributions of ambiguity distances $\Delta_{a_{CDOM}}$ (dimensionless) within DPC_{K2} .

4 IMPACT OF AMBIGUITIES ON INVERSE MODELLING

We investigate now the influence of ambiguities on the performances of inversion algorithms. Most of the inversion algorithm uses least squares minimization to converge toward good solutions. It has been shown that, in presence of non-uniqueness of the solution, these types of inverse methods tend to output the average of the desired parameters (Bishop 1994). Thus, these models approximate the conditional mean $E(y|x)$, where y is the desired parameter and x is the input of the inverse model. When addressing the inversion of the ambiguous ocean colour problem, the conditional mean $E(IOP(\lambda)|Rrs(\lambda))$ will be approximated by most of the inversion algorithm. This situation is often dramatically problematic since the answer $E(IOP(\lambda)|Rrs(\lambda))$ can be, in the better case, only one element of the set of all the possible solutions but worth it can also be outside this set (if this latter is not a convex set). Thus, even when inverse models are based on realistic data, they could report outputs far from physically plausible solutions.

To illustrate this, we present in figure 5, the inversion of the spectral neighbourhood $N(Rrs_1)$ which was previously reported figure 2. A Multi-Layer Perceptron (MLP), also known as a Multi-Layer Feed-Forward Artificial Neural Network, which was trained on the DPC_{K2} dataset was used to perform the inversion. We can

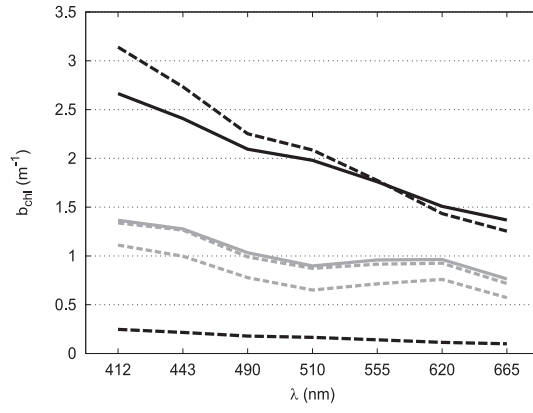


Figure 5: Comparison between an expected b_{chl} spectrum (black lines) and b_{chl} spectra (gray lines) predicted using a Multi-Layer Perceptron. The x-axis is the wavelength (nm^{-1}) and the y-axis is b_{chl} (m^{-1}).

see the three spectra of the inversion set $S_{b_{\text{chl}}}(Rrs_1)$ that should be retrieved and the three b_{chl} spectra that are actually predicted by the MLP. The three retrieved b_{chl} spectra from the inverse model are very similar each other but far from the desired spectra. This is not surprising since these retrieved b_{chl} spectra also correspond roughly to the average spectrum over the set of the three desired spectra accordingly to theory (Bishop 1994).

Thus, in the general case, the performance of an inverse model when trying to retrieve an inversion set $S(Rrs_1)$ from a neighbourhood $N(Rrs_1)$ depends on the dispersion of the spectra within the set $S(Rrs_1)$. We note D_{Rrs_1} the restricted dataset that is composed only of the n pairs (Rrs_2, IOP_{v2}) , where Rrs_2 belong to the set $N(Rrs_1)$ and IOP_{v2} belongs to the set $S(Rrs_1)$. We note also $E(S(Rrs_1))$ and $\sigma(S(Rrs_1))$ respectively the mean spectrum and the standard deviation of $S(Rrs_1)$. We note m the best inverse model that can be constructed on D_{Rrs_1} such that $m(Rrs_1) = E(S(Rrs_1))$. If the error function used to ensure the convergence of the algorithm is for example the root mean square error RMSE, then we have (equation 9):

$$\begin{aligned}
 \text{RMSE}[m, D_{Rrs_1}] &= \sqrt{\frac{1}{n} \sum_{(Rrs_2, IOP_{v2}) \in D_{Rrs_1}} (IOP_{v2} - m(Rrs_2))^2} \\
 &\approx \sqrt{\frac{1}{n} \sum_{(Rrs_2, IOP_{v2}) \in D_{Rrs_1}} (IOP_{v2} - m(Rrs_1))^2} \\
 &\approx \sqrt{\frac{1}{n} \sum_{(Rrs_2, IOP_{v2}) \in D_{Rrs_1}} (IOP_{v2} - E(S(Rrs_1)))^2} = \sigma(S(Rrs_1))
 \end{aligned} \tag{9}$$

Therefore, based on the equations 9, any inversion method using the standard RMSE to ensure the convergence of the algorithm within the restricted dataset D_{Rrs_1} can not report an error smaller than $\sigma(S(Rrs_1))$. Thus, the dispersion $\sigma(S(Rrs_1))$ of the IOP spectra within an inversion set is called the Minimum Inverse Problem Error (MIPE). Note that the unit of the MIPE is m^{-1} when retrieving inherent optical properties of marine particles.

Since the ambiguities are not equally distributed within the synthetic dataset DPC_{K2} and since the inversion sets hold different numbers of spectra with different distributions of their IOP values, it is only possible to provide an estimate of the MIPE for the entire dataset DPC_{K2} (equation 10):

$$\text{MIPE}(DPC_{K2}) = \text{RMSE}[m, DPC_{K2}] \approx \sqrt{\frac{1}{N} \sum_{(Rrs_1, IOP_{v1}) \in DPC_{K2}} \sigma(S(Rrs_1))^2} \tag{10}$$

The MIPE can be used to identify the most ambiguous regions within a dataset. In order to build two-dimensions maps of MIPE, we divided the whole dataset DPC_{K2} into a regular grid of 20×20 subsets according either to the Rrs or IOP values. Note that the number of elements differs within subsets, since Rrs or IOP values are not equally distributed, and some of them can even be empty. For each subset that hold a significant number of items (> 10), a local MIPE has been computed. Figure 6 shows a map of MIPE regarding the total

absorption coefficient a_{tot} . The IOP is plotted as a function of the R_{rs} and the wavelength 443 nm is used. The MIPE obtained regarding a_{tot} shows the highest values when the R_{rs} (443) is low and when the absorption coefficient is high. It is observed that the minimum error in the a_{tot} retrieval is within the range $[\pm 0.1 \text{ m}^{-1}, \pm 0.2 \text{ m}^{-1}]$ when the remote sensing reflectance is smaller than 0.02 sr^{-1} . When the particles are strongly absorbing, the R_{rs} values are so low that it is very difficult to clearly identify the contribution to R_{rs} from a_{tot} . As a result, the determination of a_{tot} is made more difficult. A similar reason conducted (O'Reilly, Maritorena, Mitchell, Siegel, Carder, Garver, Kahru & McClain 1988) to propose the use of the reflectance at 490 nm or 510 nm instead of the reflectance at 443 nm in the standard ocean colour algorithms such as those used for SeaWiFS to correctly retrieve the biomass concentration in the case of turbid open ocean waters. On the other hand, the MIPE significantly decreases as the reflectance increases. In this latter case, the contribution to R_{rs} (443) from a_{tot} is weak. The problem due to the low signal vanishes and the retrieval of a_{tot} is improved.

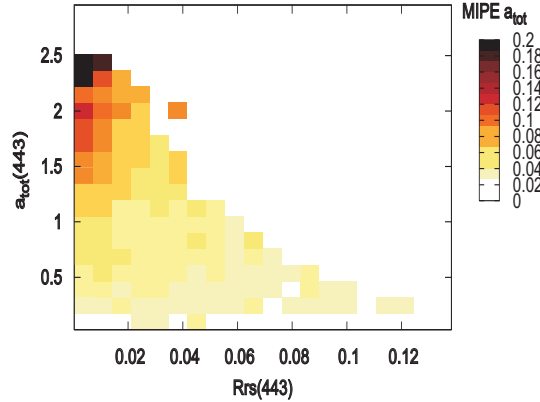


Figure 6: Map of Minimum Inverse Problem Error (m^{-1}) for a_{tot} expressed in terms of IOP vs R_{rs} .

5 DISCUSSION

We described the ambiguous aspect of the ocean colour problem and how the presence of ambiguities influences the performance of the inverse models. In this section, we discuss three different strategies that should be investigated to tackle the ambiguity problem.

5.1 Integration strategy

The integration strategy is concerned with the set of methods able to integrate the ambiguities into the inverse models. To overcome the non-uniqueness of the solution, the methods based on the integration strategy provide a global unique solution, even though this global solution consists in a distribution or a complex and structured collection of all the plausible solutions.

The first idea of the integration strategy is to ensure that the output of an inverse model (*i.e.* the retrieved IOP) is at least one of the plausible solutions instead of predicting the conditional mean. The underlying theory to achieve that is called learning with a distal teacher (Jordan & Rumelhart 1992). Such a theory has already been applied in the framework of the ocean colour inverse problem (Doerffer & Schiller 2000) or more generally in (Schiller 2003) to detect dramatic behaviours of models. It is worth noticing that the choice of this unique plausible solution among all the others is a critical issue that can be tackled using prior information; see for example the estimation of canopy variables in (Combal, Baret, Weiss, Trubuil, Mace, Pragnre, Myneni, Knyazikhin & Wang 2002).

A second idea of the integration strategy is to give an estimate of the distribution of the plausible solutions, *e.g.* with the probability density function. Gouveia et al. (1998) showed that ambiguities existing in inverse problems can be directly handled using the Bayesian approach (Tarantola 1987) which operates on data distribution. Therefore, Gouveia et al. (1998) concluded that there is no ill-posed inverse problem from a Bayesian viewpoint. Unfortunately, Bayesian-based statistical approaches, such as the Mixture Density Network (Bishop 1994) or Hierarchical Mixture of Experts (Jordan & Jacobs 1994) can not be systematically employed because they require both the construction of prior distributions of data and huge computational costs. Nevertheless interesting results have already been obtained for various real-world remote sensing inverse problems, for example in wind vectors retrieval (Cornford, Nabney & Williams 1999), ocean tomography inversion (Stphan, Thiria & Badran 1996), and atmospheric prediction of carbon monoxide (Hadji-Lazaro, Clerbaux & Thiria 1999).

In order to integrate the non-uniqueness of the solution into the inverse models, a third alternative in the integration strategy is to make use of a discrete representation, *e.g.* to output discrete lists of solutions for each IOP. To increase the level of confidence in the retrieved parameters, it is often possible to associate some probabilities of appearance to each of the proposed solution. In this context, Richaume et al. (2000) developed a technique based on the discretisation of the distributions using histograms. Such a technique might be applied regarding the ocean colour inverse problem.

5.2 Divide and Conquer strategy

The Divide and Conquer strategy, which is widely used in the computer science research field, can be applied to tackle the ambiguity problem. This may lead to transform a huge complex and strongly ambiguous inverse problem into numerous small unambiguous ones. We discern at least two ways of dividing the global problem.

The first way concerns the concept of transductive learning (local or personalized modelling) as opposed to inductive learning (global modelling); see (Kasabov 2005) for a complete review. In the local modelling approach, a set of local models are created from data, each representing a sub-space (cluster) of the problem space whereas a model is created for each input of the problem space in the personalized modelling approach. Different transductive learning techniques have already proved their efficiency in the Divide and Conquer strategy and should be used to eliminate the ambiguities, see for example the Polyhedral Mixture of Experts (Karniel, Meir & Inbar 1998) or the Evolving Connectionist System (Futschik, Reeve & Kasabov 2003).

The second way is much more concerned with physics and consists in the specialisation of inverse models to particular restricted sub-problems. Practically, inverse models specifically designed for given geographic areas and/or for given seasons and/or for particular waters types, *i.e.* CDOM or mineral dominated waters, might be satisfactorily applied to overcome the problem of ambiguities.

5.3 Enrichment strategy

The enrichment strategy proposes to drastically reduce the effects of ambiguities using ancillary data. We have identified at least three potential sources of ancillary information. The *in situ* measurements obtained during field experiments or using buoys moorings provide a significant amount of statistical ancillary information. This information can be processed off-line, during a long time period, to increase the global knowledge of the experts, for example to inform about seasonal or geographical variations of IOP or about prior distributions that can be critical in the Bayesian approach [Tarantola, 1987]. The global knowledge of the experts can then be used in real-time to resolve ambiguities of a given inversion or beforehand to improve the synthetic dataset creation process. The *in situ* collected data can also be used directly to adjust the inverse model or to select the most probable answer. This latter case can be envisaged only if the Rrs measurements of the database were collected at the same time period and same area as the Rrs spectrum to invert. The most likely solution of the inverse model is then the retrieved IOP that best matches the measured IOP of the database. Unfortunately, match-up events do not occur so often and can not be integrated routinely into an inversion algorithm. They can only be used as a validation step.

The second source of ancillary data relies on the use of additional physical information than that commonly measured in ocean color field experiment. In particular, the directional variations and the polarization of the reflectance might be relevant physical constraints that should contribute to significantly reduce the influence of the ambiguities on the retrieved solutions. Currently, the satellite sensor PARASOL (CNES) is the only one that is able to measure the directionality and polarization of the reflectance.

The last source of potential information that has to be investigated is the local context of the measurement. Indeed, it is most of the time clear that the different variables describing the physical phenomenon to study, in our case the set of all IOP, are temporally and/or spatially correlated. Thus, when inverting a given highly ambiguous spectrum for which multiple answers are produced by inverse models, it is promising to explore the spatial and/or temporal neighbourhood of the Rrs spectrum to invert to find the nearest non-ambiguous measurement. Then the ambiguity problem can be solved locally with a close-to-close approach. As an example, when computing the value of a pixel in a satellite image, the number of solutions might be significantly reduced accounting for the information provided by neighbouring pixels.

6 SUMMARY

Mathematical definitions were proposed to quantify the presence of ambiguities. Note that this is the first time, to our knowledge, that such formulations of the ambiguity problem are provided. The non-uniqueness of the solution of the inverse problem of ocean colour, so-called ambiguity problem, was then characterized. The rate of ambiguity of the remote sensing reflectance spectra within the dataset was high, around 90%. The influence of the

ambiguities on inverse modelling was studied. It was demonstrated that the fraction of the error that is ascribed to the ambiguities, so-called minimum inverse problem error, is equal to the dispersion of the plausible solutions of the problem. Practically, the ambiguity error is estimated calculating the standard deviation over all the IOP that lead to one given unique RRS spectrum. Based on this important result, it was shown that the minimum error made on the total absorption coefficient is the highest when the reflectance values in the blue are weak. Several strategies are finally discussed to tackle the ambiguity problem. Future efforts should be put in the application of these strategies to significantly enhance the performance of the inverse modelling in optically complex waters.

ACKNOWLEDGEMENTS

This work was funded by the Centre National d'Etudes Spatiales (CNES, France). We would like to thank Dr S. Verel for valuable discussion.

References

- Bishop, C. M. (1994). "Mixture density networks" *Technical report*. Aston University.
- Chami, M. & Robilliard, D. (2002). "Inversion of oceanic constituents in case I and II waters with genetic programming algorithms" *Applied Optics*. **41**(30): 260–274.
- Chami, M., Santer, R. & Dilligeard, E. (2001). "Radiative transfer model for the computation of radiance and polarization in an ocean-atmosphere system: polarization properties of suspended matter for remote sensing" *Applied Optics*. **40**(15): 2398–2416.
- Combal, B., Baret, F., Weiss, M., Trubuil, A., Mace, D., Pragere, A., Myneni, R., Knyazikhin, Y. & Wang, L. (2002). "Retrieval of canopy biophysical variables from bidirectional reflectance Using prior information to solve the ill-posed inverse problem" *Remote Sensing of Environment*. **84**: 1–15.
- Cornford, D., Nabney, I. T. & Williams, C. K. I. (1999). "Bayesian Inference for Wind Field Retrieval" *Neuro-computing Letters*. **26-27**: 1013–1018.
- Doerffer, R. & Schiller, H. (2000). "Neural Network for Retrieval of Concentrations of Water Constituents with the Possibility of Detecting Exceptional out of Scope Spectra" *IGARSS 2000*. pp. 714–717.
- Futschik, M. E., Reeve, A. & Kasabov, N. (2003). "Evolving connectionist systems for knowledge discovery from gene expression data of cancer tissue" *Artificial Intelligence in Medicine*. **28**(2): 165–189.
- Garver, S. & Siegel, D. (1997). "Inherent optical property inversion of ocean color spectra and its biogeochemical interpretation 1. Time series from the Sargasso Sea" *Journal of Geophysical Research*. **102**(18): 607–625.
- Hadji-Lazaro, J., Clerbaux, C. & Thiria, S. (1999). "An inversion algorithm using neural networks to retrieve atmospheric CO total columns from high-resolution nadir radiances" *Journal of Geophysical Research*. **104**: 23841–23854.
- Hoge, F. & Lyon, P. (1996). "Satellite retrieval of inherent optical properties by linear matrix inversion of oceanic radiance models: an analysis of model and radiance measurement errors" *Journal of Geophysical Research*. **101**(16): 631–648.
- Jordan, M. I. & Jacobs, R. A. (1994). "Hierarchical Mixtures of Experts and the EM Algorithm" *Neural Computation*. **6**(2): 181–214.
- Jordan, M. I. & Rumelhart, D. E. (1992). "Forward models: Supervised learning with a distal teacher" *Cognitive Science*. **16**: 307–354.
- Karniel, A., Meir, R. & Inbar, G. (1998). "Polyhedral mixture of linear experts for many-to-one mapping inversion" *ESANN'98*. Brussels, Belgium pp. 155–160.
- Kasabov, N. (2005). "Modelling and profile discovery in Bioinformatics: Global, local and personalised approach" *Pattern Recognition Letters*. .
- Morel, A. (1988). "Optical modeling of the upper ocean in relation to its biogenous matter content (Case I waters)" *Journal of Geophysical Research*. **93**(10): 749–768.
- O'Reilly, J., Maritorena, S., Mitchell, B., Siegel, D., Carder, K., Garver, S., Kahru, M. & McClain, C. (1988). "Ocean color algorithms for SeaWiFS" *Journal of Geophysical Research*. **103**(24): 937–953.
- Roesler, C. & Boss, E. (2003). "Spectral beam attenuation coefficient retrieved from ocean color inversion" *Geophysical Research Letter*. **30**(9): 1468.
- Schiller, H. (2003). "Neural net architectures for scope check and monitoring" *Computational Intelligence for Measurement Systems and Applications*.

- Schiller, H. & Doerffer, R. (2005). "Improved determination of coastal water constituent concentrations from MERIS data" *IEEE Transactions on Geoscience and Remote Sensing*. **43**(15): 585–591.
- Shettle, E. & Fenn, R. (1979). "Models for the aerosols of the lower atmosphere and the effect of humidity variations on their optical properties" *Environmental research papers*. **676**: 31.
- Stphan, Y., Thiria, S. & Badran, F. (1996). "Application of multilayered neural networks to ocean tomography inversions" *Inverse Problems in Engineering*. **1**: 181–304.
- Tarantola, A. (1987). *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation*. Elsevier. Amsterdam, Netherlands.
- Vapnik, V. N. (1995). *The nature of statistical learning theory*. Springer-Verlag New York, Inc.. New York, NY, USA.