Chapter 3

RETRIEVAL OF THE PIGMENT CONCENTRATION IN CASE I WATERS

3.1. Introduction

It has been recognised that Artificial Neural Network (ANN) techniques have a good potential to derive the water constituents both in Case I and Case II waters [*Keiner and Brown*, 1999; *Schiller and Doeffer*, 1999; *Gross et al.,* 2000]. ANNs have several distinct advantages. First, ANNs may be used to approximate the non-linear relationship between observations and target parameter(s) without explicitly knowing their direct functional dependence. Second, ANNs have the ability to generalise even in the presence of noisy data. Third, once an ANN is created, the processing time for parameter retrieval is short as compared to the inverse modelling techniques.

The number of *in-situ* data sets combining pigment concentration and concomitant measurements of the oceanic light field is still rather limited. Besides, the distribution of pigment concentrations in the available data sets is rather inhomogenous and thus does not meet the requirement for successful training of an ANN. Therefore, the training data used in this study is obtained from simulations of radiative transfer in the atmosphere-ocean system. RT predicts the light field based on physical models describing the interaction of matter and light. If all processes relevant to the radiative transfer in the atmosphere-ocean system were known, the light field could be predicted for arbitrary combinations of water constituents concentrations, sea surface state, atmospheric composition and observation geometry. The key problem here is to exactly define the inherent optical properties (IOPs) of sea water and to relate these to the parameters of interest (such as the pigment concentration). The IOPs of pure sea water are meanwhile well known. Well established parameterisations also exist for the absorption coefficient of marine particles in Case I waters as function of the pigment concentration and the absorption of coloured dissolved organic matter (CDOM). Recently, a new model for the phase function of marine particles was derived by optimising the agreement between RT simulations of the remote sensing reflectance with the corresponding SeaBAM data [*Zhang et al.,* 2003]. This availability allows to simulate the light field in Case I waters as required for this study.

In the present work, a method is proposed for the retrieval of the pigment concentration in Case I waters on a global scale, based on Artificial Neural Network (ANN) techniques. Input to the presented method is the spectral remote sensing reflectance just above the sea surface. The resistance against atmospheric correction errors which is required for methods working at sea

level is assured by adding the previously determined appropriate noise level to the data set used for the training of the ANNs.

3.2. Data Sets

Three different data sets are used in this study, relating remote sensing reflectance (SeaBAM) or hemispherical reflectance (COASTLOOC) to the pigment concentration:

1. a synthetic data set ("*training data*") derived from RT simulations to train the different ANNs,
2. the SeaBAM *in-situ* data ("*validation data*") to evaluate the performance of each individual ANN and such to identify the most appropriate one with respect to pigment retrieval,
3. the COASTLOOC *in-situ* data ("*test data*") to assess in how far the ANN-based pigment retrieval scheme is applicable to data other than those contained in the SeaBAM data set.

These three data sets are described in more detail in the following subsections.

3.2.1. Training Data

The synthetic data set used for the training of ANNs was created using the computer code MOMO [*Fell and Fischer*, 2001] which has been described in Section 2.3. In order to speed up the radiative transfer simulations and to make it convenient to compare simulations with *in-situ* measurements, a number of assumptions and simplifications were made:

- atmospheric Rayleigh scattering according to a constant surface air pressure of 1013.25 hPa,
- atmospheric scattering by maritime aerosol assuming a constant aerosol optical depth of 0.1 at 550 nm,
- no atmospheric absorption,
- flat air-sea interface,
- no vertical stratification of the ocean,
- no inelastic scattering inside the water body (*i.e.* no Raman scattering, no chlorophyll-*a* fluorescence, no CDOM fluorescence).

For any study based on RT simulations of the oceanic light field, detailed knowledge on the IOPs of the oceanic constituents is required. The IOP models used in this study are summarised in Table 3.1. Based on the above assumptions and the selected IOP models, simulations of remote sensing reflectance at nadir direction were made for:

- 18 solar zenith angles between 0° and 87°,
- 5 wavelengths: 412, 443, 490, 515, and 555 nm,
- 300 pigment concentrations between 0.025 and 35 mg m^{-3}.

In the SeaBAM data set (see below), no information is given on the solar zenith angles at which measurements have been taken. Therefore, the simulated remote sensing reflectances were averaged over solar zenith angles between 0° and 70° to construct a synthetic data set where the relationship between R_{RS} and pigment concentration is independent on the viewing geometry. The wavelengths chosen for the simulations correspond to the central wavelengths of spectral channels commonly used for space-borne or *in-situ* ocean colour measurements. The method derived herein may therefore be applied to most ocean colour instruments. A logarithmic distribution of the pigment concentrations has been selected for the simulations, thus, each order of magnitude within the considered pigment concentration range is represented with a similar number of cases.

Table 3.1. Parameterisations of inherent optical properties of pure sea water and water constituents used of Case I waters.

Constituents	IOP	Parameterisation or measurement	Reference
Pure sea water	Absorption	Directly measured	Pope and Fry [1997]
	Total scattering	$b_w(\lambda) = 0.00288(\frac{\lambda}{500})^{-4.32}$	Morel [1974]
	Phase function	$p_w(\lambda) = 0.06225(1 + 0.835\cos^2\theta)$	Morel [1974]
Marine particulate matter	Absorption	$a_p(\lambda) = A_{ap}(\lambda)[Chl]^{B_{ap}(\lambda)}$	Bricaud et al. [1998]
	Total scattering	$b_p(\lambda) = A_{bp}[Chl]^{B_{bp}}$, A_{bp}=0.252 and B_{bp}=0.635	Loisel and Morel [1998]
	Phase function	$p_p = f(Chl, \lambda)$	Zhang et al. [2003]
CDOM	Absorption	$a_y(\lambda) = a_y(443)\, e^{-S_y(\lambda-443)}$, $S_y = 0.014$	Bricaud et al. [1981]
		$a_y(443) = 0.00348 + 0.5 * a_p(443)$	Zhang et al. [2003]

3.2.2. Validation Data: the SeaBAM Data Set

The performance of each individually trained ANN is assessed using the SeaBAM data set. The SeaBAM data set is a compilation of measurements and derived parameters taken at 919 stations on occasion of various oceanographic campaigns. 900 stations are considered as being located in Case I waters. The SeaBAM data set covers a pigment range between approx. 0.03 and

30.0 mg m^{-3}, and provides remote sensing reflectance values into nadir direction at six wavelengths (412, 443, 490, 515, 555, 670 nm). It is considered to contain the most comprehensive information on remote sensing reflectance values and concomitant pigment concentrations in Case I waters on a global scale. Detailed information on the SeaBAM data set is shown in Table 3.2. It is publicly available from the SeaBASS (SeaWiFS Bio-optical Archive and Storage System) web site (http://seabass.gsfc.nasa.gov).

Table 3.2. Data sources and characteristics of the SeaBAM data set, reproduced after *O'Reilly et al.* [1998].

Campaign name	Provider / PI	Area	Time	Data sets	Water type
BBOP92-93	D. Siegel	Sargasso Sea	monthly, 1992-1993	72	1
BBOP94-95	D. Siegel	Sargasso Sea	monthly,1994-1995	67	1
WOCE	J. Marra	50°S-13°N,88°-91°W 10°S-30°N,18°-37°W	March 1993 April 1993	70	1
EQPAC	C. Davis	0.140°W	March, Sept. 1992	126	1
NABE	C. Davis	46°N, 19°W	April 1989	40	1
NABE	C. Trees	46°-59°N, 17°-20°W	May 1989	72	1
CARDER	K. Carder	North Atlantic Pacific Gulf of Mexico Arabian Sea	Aug. 1991 July 1992 April 1993 Nov. 1994, June 1995	87	1
CALCOFI	G. Mitchell	California Current	Aug. 1993-July 1996	303	1
MOCE 1	D. Clark	Monterey Bay	Sept. 1992	8	1
MOCE 2	D. Clark	Gulf of California	April 1993	5	1
North Sea	R. Doerffer	55°-52°N, 0°-8°E	July 1994	10	2
Chesapeake Bay	L. Harding	~37°N, 75°W	April, July 1995	9	2
Canadian Arctic	G. Cota	~74.38°N, 95°W	Aug. 1996	8	1
AMT	G. Moore	50°N-50°S	Sept. 1995, April 1996	42	1
Total				919	

3.2.3. Test Data: the COASTLOOC Data Set

The COASTLOOC (Coastal Surveillance Through Observation of Ocean Colour) data set relates the subsurface hemispherical reflectance at 13 wavelengths between 412 and 865 nm to concomitantly measured IOPs and water constituents concentrations [*Babin*, 2000]. Most of the 424 stations visited during the COASTLOOC campaigns have been gathered in Case II waters in different European coastal areas, except for 93 stations located in Case I waters in the Atlantic Ocean and Mediterranean Sea. The measurements at 67 of the 93 Case I water stations fulfilled the requirements in terms of completeness and accuracy, and were retained for further processing in the frame of algorithm development for Case I waters. Table 3.3 provides more information on this data set.

Table 3.3. Regional distribution of the COASTLOOC stations located in Case I waters.

Campaign	PI	Area	Time	Case I stations
COASTLOOC 1	M. Babin	NE Atlantic Ocean	April 1997	34
COASTLOOC 3	M. Babin	Northern Adriatic Sea	July, August 1997	6
COASTLOOC 4	M. Babin	Mediterranean Sea (Golfe du Lion)	October 1997	12
ALMOFRONT 2	H. Claustre	Mediterranean Sea (Alboran Sea)	January 1998	41
Total				93

3.3. ANN-based Pigment Retrieval in Case I Waters

In the present study, a three-layer MLPs ANN is used to derive the functional relationship between the spectral remote sensing reflectance into the nadir direction and the pigment concentration. The architecture of the ANN has been described in Section 2.4.

3.3.1. Parameters affecting the performance of an ANN

The performance of a trained MLP depends mainly on three limiting factors: information content of the offered input, number of the neurons in the hidden layer, noise level added to the training data set. This is elucidated in the following subsections.

3.3.1.1. Information content of the offered input

To determine which and how many spectral bands or band ratios are best suited for pigment retrieval, nine combinations of input data (listed in Table 3.4) were tested. Of the nine listed cases, the first three are combinations of remote sensing reflectances at different wavelengths. The other six cases are combinations of remote sensing reflectance ratios. The dimensionality of the input data determines the number of neurons in the input layer.

3.3.1.2. Number of the neurons in the hidden layer

How many hidden neurons should be used? Generally speaking, if too few hidden neurons are used, high training error and high generalisation error may be obtained due to underfitting and high statistical bias. If too many hidden neurons are used, low training error may result, but there may be high generalisation errors due to overfitting and high variance [*Geman et al.*, 1992]. Besides, the training process becomes very slow. In this study, the performance of ANNs with 4, 9 20, and 30 hidden neurons has been tested.

3.3.1.3. Noise level added to the training data set

Through adding noise to synthetic training data, the robustness of the trained ANN with respect to noisy input data is increased. However, adding too much noise will obviously lead to a high generalisation error [*Koistinen and Holmström*, 1992]. To determine the appropriate noise level, 5 %, 10 %, 20 %, and 30 % noise was added to the synthetic training data used as input to the ANNs (R_{RS} or R_{RS} ratios). Since no specific noise model was available, a simple approach was chosen: the noise in different samples and in different elements of a sample is assumed to be independent, its distribution is assumed to be random and to have zero mean.

3.3.2. ANN Training

A synthetic data set for ANN training has been generated from the RT simulations outlined in Section 3.2. The synthetic data set is composed of 300 remote sensing reflectance spectra, corresponding to 300 pigment concentration values. Using a higher number of samples to train the ANNs did not significantly improve their performance. The main reason for this is the relative simplicity of the IOP models which are all parameterised as function of the pigment concentration alone in combination with the constant observation geometry.

For each of the nine input combinations listed in Table 3.4, training data sets of four different noise levels (5%, 10%, 20%, 30%) were used to train ANNs with 4, 9, 20, or 30 neurons in the hidden layer. This gives a total of 9×4×4 = 144 combinations, each of which is represented by an individually trained ANN. One specific training set has been derived for each combination of input parameters and noise level, giving a total of 9×4=36 training data sets. Each training data set is then used to train the 4 ANNs characterised by their different number of hidden neurons.

Table 3.4. Performance with regard to pigment retrieval of spectral band combinations used as input to ANNs.

Case No.	Input	Neurons in hidden layer	Noise level (%)	Training data (synthetic)		Validation data (SeaBAM)	
				RMSE	r^2	RMSE	r^2
1	$R_{RS}412$, $R_{RS}443$, $R_{RS}490$, $R_{RS}515$, $R_{RS}555$	4	30	0.156	0.971	0.271	0.835
2	$R_{RS}443$, $R_{RS}490$, $R_{RS}515$	4	30	0.199	0.953	0.311	0.785
3	$R_{RS}443$, $R_{RS}490$	4	30	0.210	0.948	0.322	0.782
4	$R_{RS}412/R_{RS}555$, $R_{RS}443/R_{RS}555$, $R_{RS}490/R_{RS}555$, $R_{RS}515/R_{RS}555$	20	5	0.0983	0.988	0.159	0.925
5	$R_{RS}443/R_{RS}555$, $R_{RS}490/R_{RS}555$, $R_{RS}515/R_{RS}555$	30	5	0.0787	0.993	0.148	0.934
6	$R_{RS}443/R_{RS}555$, $R_{RS}515/R_{RS}555$	30	5	0.0781	0.993	0.156	0.929
7	$R_{RS}443/R_{RS}555$, $R_{RS}490/R_{RS}555$	4	20	0.127	0.981	0.155	0.930
8	$R_{RS}443/R_{RS}555$	9	10	0.0924	0.990	0.156	0.929
9	$R_{RS}490/R_{RS}555$	9	5	0.0886	0.982	0.167	0.919

3.3.3. Determining ANN Architecture and Noise Adding for Optimal Pigment Retrieval

In order to find the ANN best suited for pigment retrieval, the ANN pigment forecasts were compared to the SeaBAM data by two error measures. First, root mean square error (RMSE) defined by:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}[\log_{10}(CHL_i^{\,D}) - \log_{10}(CHL_i^{M})]^2}\,, \tag{3.4}$$

where *CHL* represents the pigment concentration, and the superscripts *D* and *M* indicate derived and measured values. Second, the square of the Pearson's correlation coefficient r^2. The pigment concentrations are log-transformed prior to calculating RMSE and correlation coefficients. The optimum number of hidden neurons and the appropriate noise level with regard to the above two error measures are given in Table 3.4 for each of the nine input combinations.

Using the remote sensing reflectance at five wavelengths between 412 and 555 nm as input to the ANN (case No. 1) results in r^2 = 0.835 between the ANN pigment forecast and the corresponding SeaBAM values (RMSE = 0.271). Using one or more remote sensing reflectance

ratios as input to the ANN (cases No. 4-9), the r^2 values are in any case above 0.919, and the RMSE values below 0.167. The performance of the ANNs using remote sensing reflectance ratios as input is thus considerably higher than that of ANNs using absolute remote sensing reflectance values. The explanation for this effect is that spectrally correlated noise is partly cancelled out through division of the remote sensing reflectance at two wavelengths.

Of the six combinations of remote sensing reflectance ratios taken as input to the ANN, the best results (lowest RMSE and highest r^2) were obtained using the three remote sensing reflectance ratios $R_{RS}443/R_{RS}555$, $R_{RS}490/R_{RS}555$, and $R_{RS}515/R_{RS}555$ as input (case No. 5). Interestingly, this case is not the one with most spectral ratios as input (case No. 4). Even the performance of case No. 8 with only one spectral ratio ($R_{RS}443/R_{RS}555$) as input is better than that of case No. 4 using four spectral ratios, and close to that of case No. 6 using three spectral ratios. Generally speaking, the higher the number of the spectral bands, the more information is available, and the higher should be the accuracy of the ANN forecast. On the other hand, inclusion of noisy data will reduce the ANN forecast performance.

3.3.4. Best ANN for Pigment Retrieval

From the above results, the following optimum ANN architecture has been selected: three layers, three neurons (plus one bias parameter) in the input layer, 30 neurons (plus one bias parameter) in the hidden layer, one neuron in the output layer. The appropriate noise level to be added to the synthetic data is 5 %. Due to the chosen architecture, the input to the neural network is given by the four element vector:

$$I = [R_1, R_2, R_3, B], \tag{3.5}$$

where the bias parameter B is always set to one, and the elements R_1, R_2, and R_3 are given by:

$$R_1 = 0.05 + 0.305 \times (\ln (R_{RS}443 / R_{RS}555) + 0.829) , \tag{3.6a}$$

$$R_2 = 0.05 + 0.411 \times (\ln (R_{RS}490 / R_{RS}555) + 0.473) , \tag{3.6b}$$

$$R_3 = 0.05 + 0.805 \times (\ln (R_{RS}515 / R_{RS}555) + 0.382) . \tag{3.6c}$$

Appropriate pre-processing is essential for a successful training of ANNs: the transformations (3.6a)-(3.6c) have the aim a) to represent the roughly linear dependence of the logarithms of colour ratio and pigment concentration and b) to map the input data to an interval confined to the range [0.05, 0.95]. Choosing this interval instead of [0.0, 1.0] allows for pigment retrieval even if the input values are slightly outside the range of the synthetic training data set. However, such extrapolation should be handled with great care. The temperature constant of the sigmoid was given a value of $c_t = 1.0$. The elements of the weight matrices

$$W^{IH} = \begin{pmatrix} W^{IH}_{1,1} & \cdots & W^{IH}_{1,4} \\ \cdots & \cdots & \cdots \\ W^{IH}_{31,1} & \cdots & W^{IH}_{31,4} \end{pmatrix} , \tag{3.7a}$$

and

$$W^{HO} = \left(W_{1,1}^{HO} \quad \ldots \quad W_{1,31}^{HO} \right), \tag{3.7b}$$

the one-element output vector $\vec{O} = [O_1]$ of the ANN is re-transformed into the pigment concentration by:

$$Chl = exp\ ((O_1 - 0.05) / 0.124 - 3.71)\ . \tag{3.8}$$

3.4. Evaluating the Performance of the ANN-based Pigment Retrieval Algorithm

3.4.1. Assessing the Performance of the Trained ANN

The potential of the selected ANN for pigment retrieval from real measurements is assessed in three steps by applying it a) to the synthetic data ("*training data*") used for the ANN training, b) to the SeaBAM data ("*validation data*") used to determine the ANN architecture and noise level to be added to the input data, and c) to the COASTLOOC data ("*test data*") which have not been used for the ANN development.

3.4.1.1. Performance with respect to the training data

The ANN pigment concentration forecasts using simulated remote sensing reflectance ratios as input data are compared to the pigment concentrations used as input for the corresponding RT simulations. As shown in Figure 3.1, the inversion is successful with regard to the synthetic training data set (r^2 = 0.993, RMSE = 0.0787 for the log-transformed pigment concentration). Figure 3.1 also reveals a limitation of the synthetic training data set: For small pigment concentrations, the dependence of the simulated remote sensing reflectance on the pigment concentration is weak. The ANN may therefore not distinguish between pigment concentrations less than approx. 0.06 mg m^{-3}, although the training data comprises pigment concentrations between 0.025 and 35 mg m^{-3}.

3.4.1.2. Performance with respect to the validation data

In a second step, the ANN is applied to the *in-situ* measurements of the remote sensing reflectance contained in the SeaBAM data set (900 stations in Case I waters, Figure 3.2A). Here again, the pigment concentrations derived from the SeaBAM reflectance data agree well (as compared to empirical algorithms, see section 3.4.2) with the corresponding *in-situ* measurements (r^2 = 0.934, and RMSE = 0.148 for the log-transformed pigment concentration). This is not surprising, since the SeaBAM data have a) been used to derive the back scattering model for marine particles used for the RT simulations [*Zhang et al.,* 2003], and b) to select the most appropriate ANN architecture and noise level.

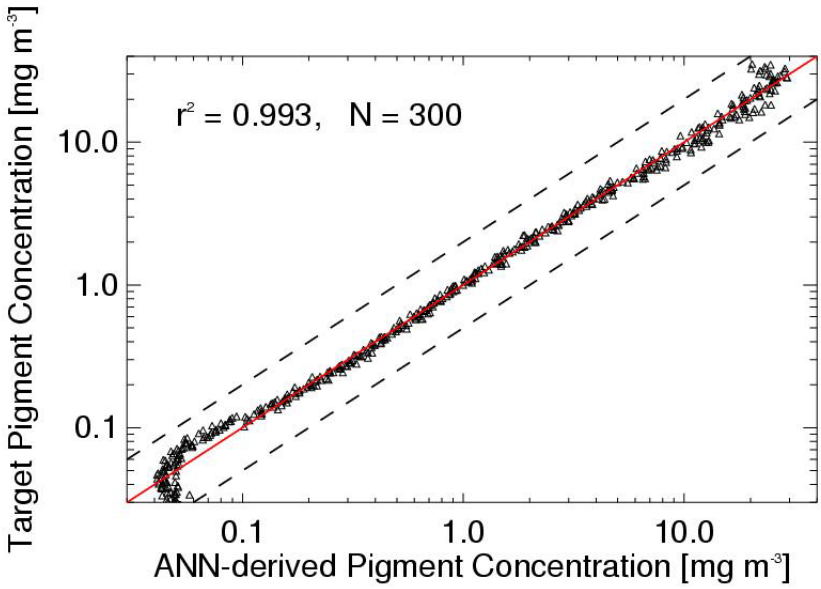

Figure 3.1. Scatter plot showing the performance of the ANN for the synthetic training data set. The target pigment concentration designates the pigment concentration used as input to the RT simulations. The dashed lines indicate the factor 2 error margin.

3.4.1.3. Performance with respect to the test data

In a third step, the ANN is applied to 67 COASTLOOC Case I reflectance spectra (Figure 3.2B). Here, it is assumed that the ratio of the hemispherical reflectance at two wavelengths is identical to the corresponding remote sensing reflectance ratio. Satisfactory (again as compared to the empirical algorithms) performance (r^2 = 0.892, RMSE = 0.219 for the log-transformed pigment concentration) is observed, even though the COASTLOOC data are totally independent from the SeaBAM data and have not been used in any respect to derive the ANN. This is a hint that the retrieval method presented herein has a potential to be applied on a regional or global scale. However, ANN-derived pigment concentrations from the COASTLOOC reflectance data seem to be systematically overestimated by about 30 %. Since a similar effect is also observed when applying empirical algorithms, the observed behavior appears to be data driven. The instrumental problems are mainly suspected to be reason behind these deviations, since the radiometer used for all COASTLOOC and ALMOFRONT campaigns showed a significant degradation in several spectral channels which might not have fully been compensated by the applied correction procedure. Besides, systematic differences between corresponding spectral ratios of the remote sensing reflectance and the hemispherical reflectance might also contribute to the observed deviations.

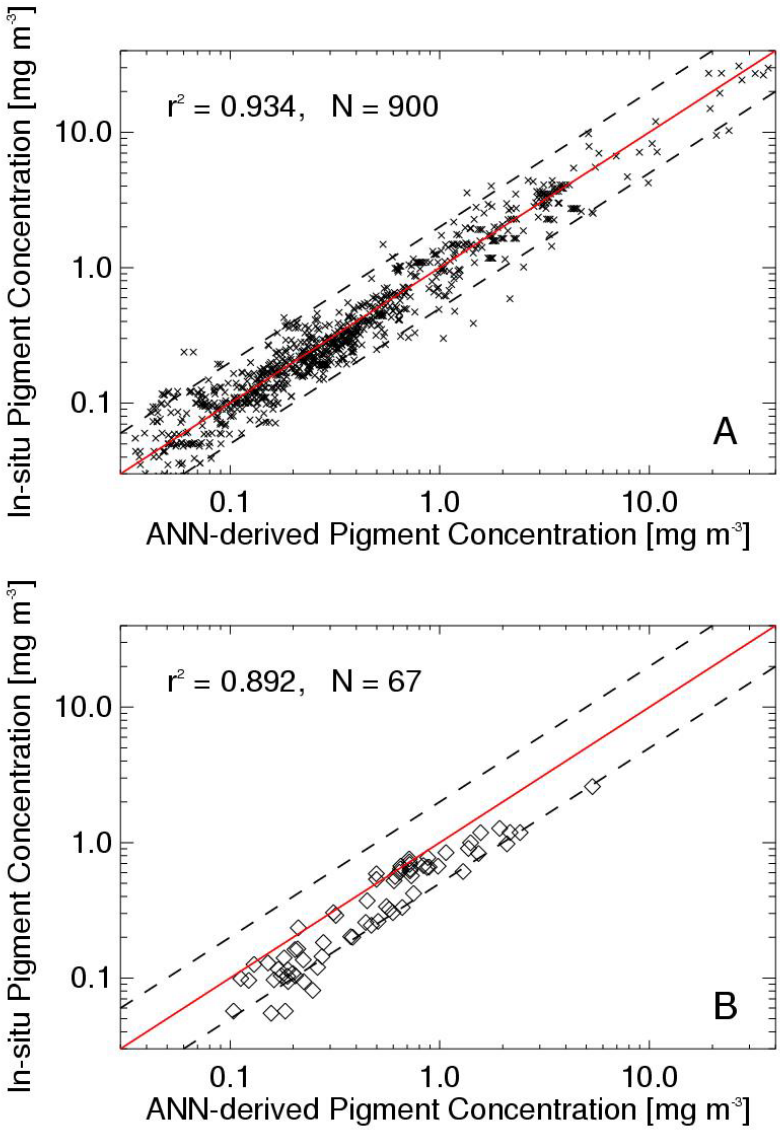

Figure 3.2. Comparison of ANN-derived pigment concentration to corresponding *in-situ* measurements for the SeaBAM (A) and COASTLOOC (B) data sets. The dashed lines indicate the factor 2 error margin.

3.4.2. Comparison with Existing Empirical Algorithms

In order to further evaluate the performance of the trained ANN, it is compared to the most successful empirical algorithms compiled in *O'Reilly et al.* [1998], listed here in Table 3.5. The algorithms to be compared are applied to the reflectance data from 900 SeaBAM Case I stations. The performance of the different algorithms is given in Table 3.6 and is depicted in Figure 3.3. The ANN has the highest r^2 and lowest RMSE of all compared algorithms. The relative success of the ANN-based pigment retrieval is partly explained by the fact that the underlying IOP models represent the SeaBAM data well. A further reason is that it uses more spectral information than do the empirical algorithms. The performance of an ANN using one colour ratio

($R_{RS}443/R_{RS}555$) slightly drops (r^2 = 0.929 instead of r^2 = 0.934, RMSE = 0.156 instead of RMSE = 0.148); however, it is still above that of the corresponding empirical algorithms (e.g., MOREL-3: r^2 = 0.918, RMSE = 0.175, OC2B: r^2 = 0.924, RMSE = 0.156).

Table 3.5. Empirical ocean colour algorithms based on remote sensing reflectance [*O'Reilly et al.,* 1998].

Algorithm	Functional form	Band ratio
POLDER	$C=10^{(a_0 + a_1*R + a_2*R^2 + a_3*R^3)}$ a = [0.438, -2.114, 0.916, -0.851]	$R = \log_{10} (R_{RS}443 / R_{RS}565)$
CalCOFI 2-band cubic	$C=10^{(a_0 + a_1*R + a_2*R^2 + a_3*R^3)}$ a= [0.450, -2.860, 0.996, -0.3674]	$R = \log_{10} (R_{RS}490 / R_{RS}555)$
MOREL-3	$C=10^{(a_0 + a_1*R + a_2*R^2 + a_3*R^3)}$ a = [0.20766, -1.82878, 0.75885, -0.73979]	$R = \log_{10} (R_{RS}443 / R_{RS}555)$
MOREL-4	$C=\exp^{(a_0 + a_1*R + a_2*R^2 + a_3*R^3)}$ a = [1.03117, -2.40134, 0.3219897, -0.291066]	$R = \ln (R_{RS}490 / R_{RS}555)$
OC2	$C=10^{(a_0 + a_1*R + a_2*R^2 + a_3*R^3)} + a_4$ a = [0.3410, -3.001, 2.811, -2.0410, -0.040]	$R = \log_{10} (R_{RS}490 / R_{RS}555)$
OC2B	$C=10^{(a_0 + a_1*R + a_2*R^2 + a_3*R^3)} + a_4$ a = [0.1909, -1.9961, 1.3020, -0.5091, -0.0815]	$R = \log_{10} (R_{RS}443 / R_{RS}555)$
OC4	$C=10^{(a_0 + a_1*R + a_2*R^2 + a_3*R^3)} + a_4$ a = [0.4708, -3.8469, 4.5338, -2.4434, -0.0414]	$R = \log_{10} ([R_{RS}443>R_{RS}490>R_{RS}510] / R_{RS}555)$

Table 3.6. Performance of the ANN-based pigment retrieval scheme as compared to selected empirical algorithms when applied to the SeaBAM and COASTLOOC data sets.

Algorithm	SeaBAM Case I (N = 900)		COASTLOOC Case I (N = 67)	
	RMSE	r^2	RMSE	r^2
ANN	0.148	0.934	0.219	0.892
POLDER	0.256	0.919	0.423	0.898
CalCOFI 2 band cubic	0.186	0.918	0.221	0.863
MOREL-3	0.175	0.918	0.278	0.896
MOREL-4	0.198	0.906	0.243	0.875
OC2	0.162	0.918	0.181	0.860
OC2B	0.156	0.924	0.230	0.893
OC4	0.151	0.928	0.198	0.873

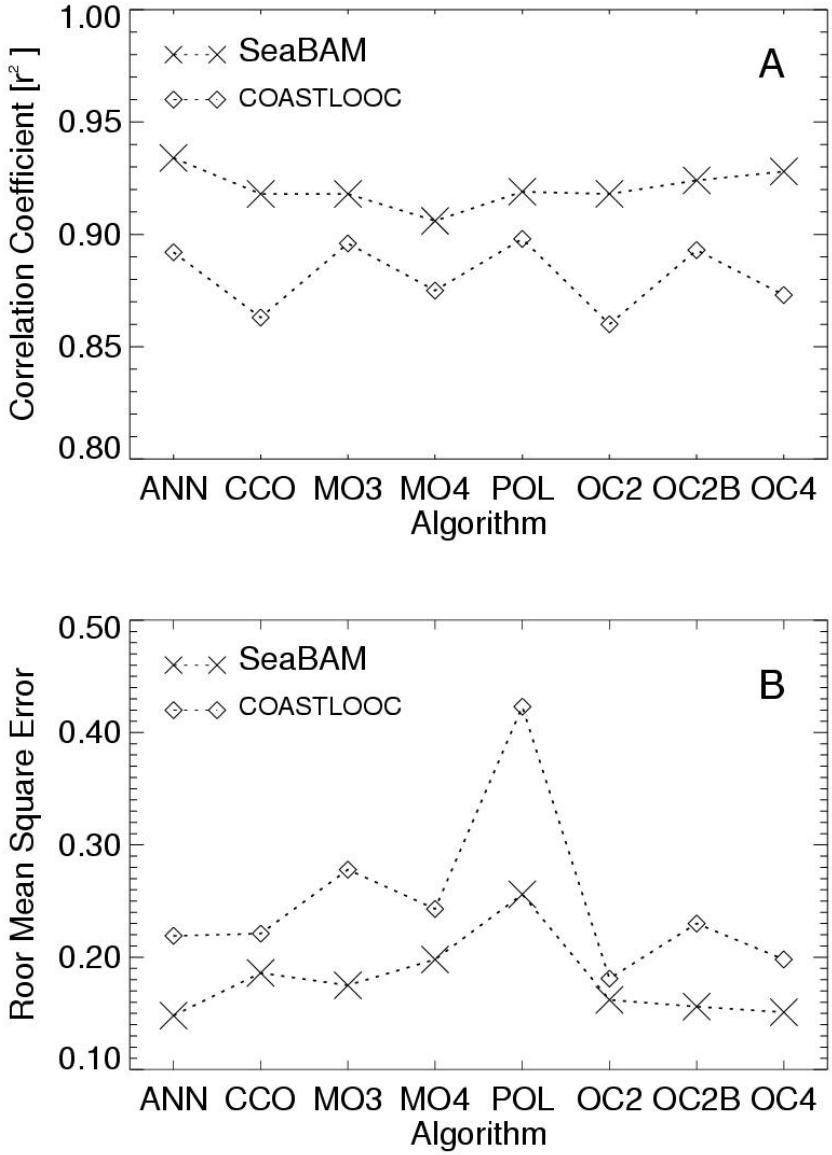

Figure 3.3. Performance of the ANN-based pigment algorithm presented in this study as compared to selected empirical algorithms from *O'Reilly et al*. [1998], using square of correlation coefficient r^2 (A) and root mean square error RMSE (B) of derived *vs*. actually measured pigment concentration as measure of success. The algorithms were applied to the SeaBAM (×) and COASTLOOC (◇) data sets.

3.4.3. Resistance against Noise

In order to compare the resistance against noise of the trained ANN to that of the empirical algorithms, random noise of 0~30 % was added to the SeaBAM remote sensing reflectance ratios before offering them as input to the pigment retrieval algorithms. Comparing RMSE and r^2 (Table 3.7), one observes that the ANN-based algorithm has a higher performance against noise than do have the empirical algorithms (Figure 3.4). This is a fundamental advantage when applying algorithms to real measurements.

Table 3.7. Performance against noise of the ANN-based pigment retrieval scheme as compared to empirical algorithms. All algorithms were applied to SeaBAM remote sensing reflectance ratios to which different levels of noise have been added. The following acronyms are used: POL = POLDER, CCO = CalCOFI two-band cubic, MO3 = MOREL-3, MO4 = MOREL-4, OC2, OC2B and OC4 = Ocean Chlorophyll algorithms 2, 2B and 4 [*O'Reilly et al.,* 1998].

Noise		ANN	POL	CCO	MO3	MO4	OC2	OC2B	OC4
0 %	RMSE	0.148	0.256	0.186	0.175	0.198	0.162	0.156	0.151
	r^2	0.934	0.919	0.918	0.918	0.906	0.918	0.924	0.928
5 %	RMSE	0.150	0.258	0.189	0.177	0.203	0.167	0.158	0.154
	r^2	0.932	0.917	0.915	0.916	0.901	0.913	0.921	0.925
10 %	RMSE	0.155	0.263	0.197	0.181	0.213	0.179	0.164	0.162
	r^2	0.927	0.912	0.906	0.911	0.890	0.901	0.915	0.918
20 %	RMSE	0.174	0.281	0.227	0.200	0.251	0.226	0.186	0.192
	r^2	0.910	0.892	0.871	0.891	0.850	0.850	0.893	0.886
30 %	RMSE	0.197	0.310	0.271	0.227	0.304	0.264	0.220	0.225
	r^2	0.888	0.861	0.818	0.859	0.790	0.811	0.856	0.847

3.5. Conclusions

In this study, a method for pigment retrieval from ocean colour in Case I waters has been derived and evaluated. The retrieval method is derived by applying ANN techniques to a set of remote sensing reflectance spectra typical of Case I waters, which have been obtained from RT simulations and the IOP models compiled in Table 3.1.

The ANN employed in this study has three layers: one input layer consisting of three neurons (plus one bias parameter), one hidden layer consisting of 30 neurons (plus one bias parameter), and one output layer consisting of one neuron. The three neurons in the input layer correspond to the remote sensing reflectance ratios $R_{RS}443/R_{RS}555$, $R_{RS}490/R_{RS}555$, and $R_{RS}515/R_{RS}555$. Applying the trained ANN to the Case I SeaBAM data gives a correlation between predicted and measured pigment concentrations of $r^2 = 0.934$ and a root mean square error of RMSE = 0.148; applying it to the Case I COASTLOOC data which have not been used to derive the phase function model of marine particles used for the RT simulations, results in $r^2 = 0.892$ and RMSE = 0.219.

The performance of the ANN-based pigment retrieval scheme is comparable to the most successful empirical algorithms: applying e.g., the SeaWiFS algorithm OC4 to the SeaBAM data set gives $r^2 = 0.928$, and RMSE = 0.151 as compared to $r^2 = 0.934$ and RMSE = 0.148 for the

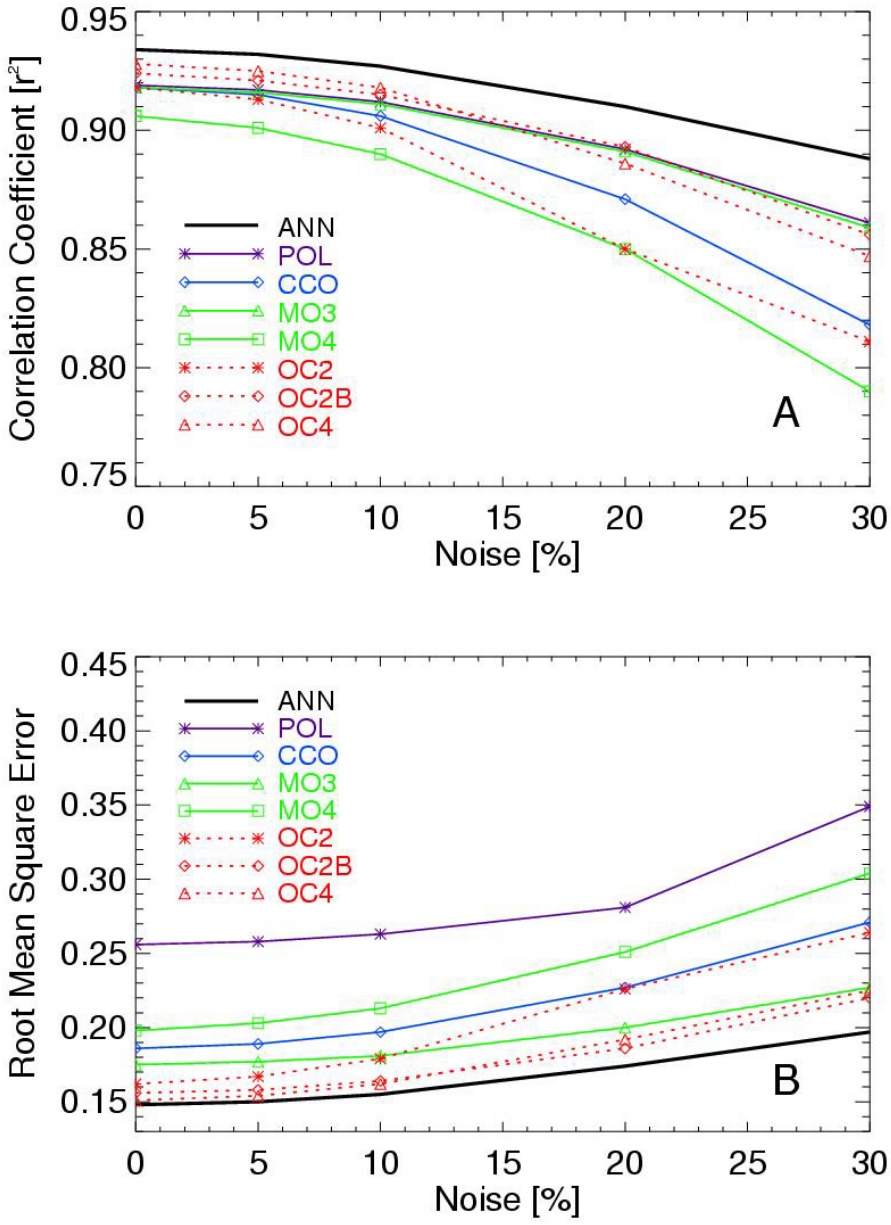

Figure 3.4. Performance against noise of the ANN-based algorithm presented in this study and selected empirical algorithms, using square of correlation coefficient r^2 (A) and root mean square error RMSE (B) of derived *vs*. measured pigment concentrations as measure of success.

ANN-based algorithm. The resistance against noise of the ANN-based algorithm is superior as compared to the empirical algorithms: adding 20 % random noise to the SeaBAM remote sensing reflectance ratios before offering them to the pigment retrieval algorithms, r^2 of OC4 drops from 0.928 to 0.886, while only going down from 0.934 to 0.910 for the ANN-based algorithm. Correspondingly, the RMSE of OC4 is increased from 0.151 to 0.192, while only rising from 0.148 to 0.174 for the ANN-based algorithm. It is this resistance against noisy input data which make ANN-based pigment retrieval schemes a very suitable tool for studying the marine

environment from space-borne Earth Observation data. This is especially true considering the fact that the presented method in principle allows the inclusion of viewing geometry, sun position, rough air-sea interface, etc. In fact, all processes may be considered for which a physical or statistical model exists that can be integrated into the RT code.