Ocean Colour Climate Change Initiative (OC_CCI) - Phase Two

Product Validation and Algorithm Selection Report (PVASR)

Part 1 – Atmospheric Correction
Ref: D2.5
Date: 2015-12-23
Issue: 3.0
For: ESA / ESRIN
Ref: AO-1/6207/09/I-LG

PML | Plymouth Marine Laboratory

This Page is Intentionally Blank

SIGNATURES AND COPYRIGHT

Project	:	**Ocean Colour Climate Change Initiative (OC_CCI) - Phase Two**
Document Title:		**Product Validation and Algorithm Selection Report**
		Part 1 - Atmospheric Correction
Reference	:	**D2.5**
Issued	:	**2015-12-23**
Issue	:	**3.0**
Client	:	**ESA / ESRIN**
Authored	:	*Dagmar Müller*
		[Dr Dagmar Müller, Dr. Hajo Krasemann - HZG]
Reviewed	:	*John Swinton*
		[John Swinton - VEGA]
Approved	:	*S Sathyendranath*
		[Dr Shubha Sathyendranath - PML]
Address	:	Helmholtz-Zentrum Geesthacht (HZG)
		Max-Planck-Str. 1
		21502 Geesthacht
		Germany
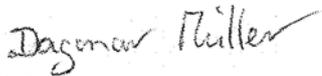
Document Change Log

Issue/ Revision	Date	Comment
0.1	2011-06-01	Initial Template
1.0	2011-10-31	Initial Version – Atmospheric Correction Algorithms
1.1	2011-11-08	Update for external release
1.2	2011-11-25	Updated with JRC review comments
1.3	2012-11-26	Updated following second iteration of Round Robin
1.4	2014-01-19	Updated following third iteration of Round Robin
2.0	2014-01-31	Amendment to version 1.3, 2nd RR, first automation partly introduced
2.2	2015-02-20	Major changes in following parts: OCCCI in-situ database v2.2, Band-shift protocol. - Round robin protocol: definition of flags, definition case-1 water. - Tool for inter-comparison. - Changes in AC processor versions, including system vicarious calibration.
3.0	2015-12-23	Major changes: OC_CCI in-situ database v3.0, band-shifting of in-situ data is revised thoroughly, new sensor VIIRS, different reflectance-normalisations are compared.

This Page is Intentionally Blank

Contents

Part I

Round Robin 3 - Atmospheric correction match-up comparison

The atmospheric correction match-up comparison is crucial in the data processing effort of the OC_CCI team. Many tasks have to be carefully fulfilled in order to provide data sets and methodologies, which allow the team to choose the best atmospheric correction for the generation of long-time products.

The basis of the entire analysis is the OC-CCI in-situ data base version 3.0, which is briefly introduced (chapter 1).

Band-shifting is necessary, as most of the in-situ remote sensing reflectance data sets do not match the spectral specifications of all the satellite sensors. In some detail, the methodology and validation of the band-shift is documented (chapter 2), as it has been revised thoroughly.

The round robin protocol defines, how satellite and in-situ data is selected and processed into a match-up data base (chapter 3). The criteria for the algorithm selection are fixed and presented in chapter 4.

In order to simplify the application of the round robin protocol on a given data set, a graphical user interface has been designed (chapter 5).

The results of the inter-comparison are summarised in chapter 6. The automatically generated documentation of statistical properties can be found in appendices A to P.

Some preliminary results of the inter-comparison on VIIRS data is presented in chapter 7.
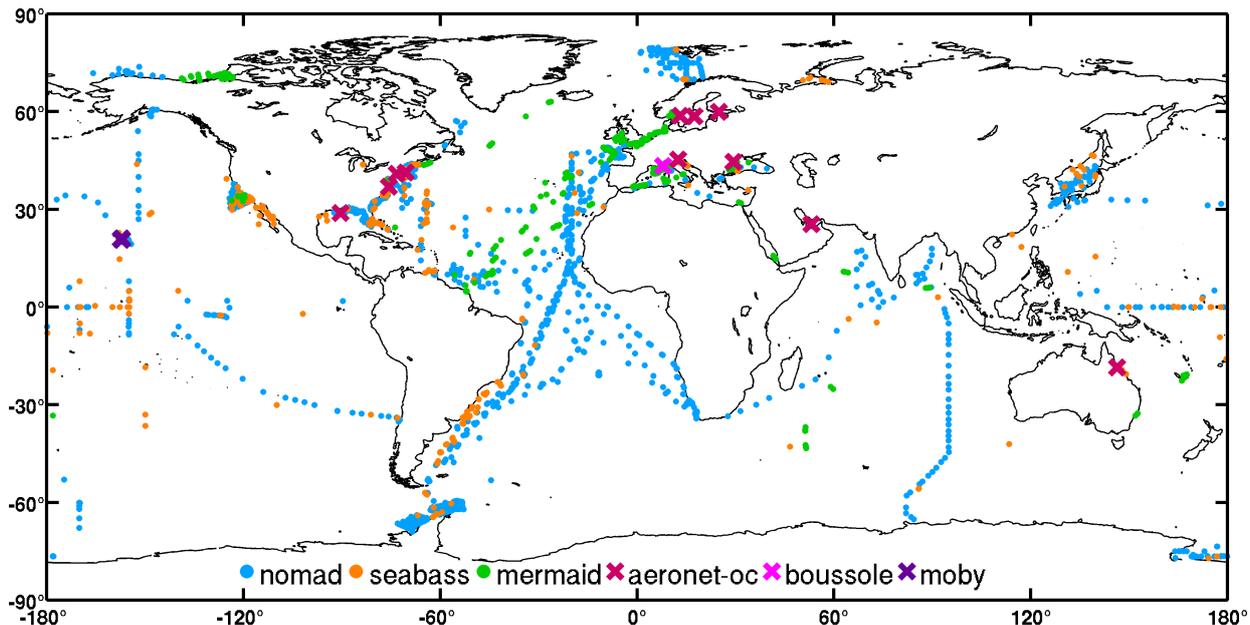
Figure 1.1: Global distribution of remote-sensing reflectance per data set in the final table. The origin of the data is described in colours. Points show locations where at least one observations is available. Crosses show sites where complete time series of remote-sensing reflectance were compiled.

1 In-situ database OC_CCI v3.0

The compilation of in-situ data used for the validation of the ocean colour products from the ESA Ocean Colour Climate Change Initiative (OC-CCI) project has been obtained from several sources. It spans between 1997 and 2015, and has a global distribution (Fig. 1.1).

Observations of the following variables were compiled: remote sensing reflectance, concentration of chlorophyll-a, inherent optical properties like phytoplankton absorption (a_{ph}), total absorption of detritus and gelbstoff (a_{dg}), backscattering of particles (bb_p) and the diffuse attenuation coefficient (k_d). Data were obtained from the following sources: MOBY, BOUSSOLE, AERONET-OC, SeaBASS, NOMAD, MERMAID, AMT, ICES, HOT, GEPCO. The data were acquired via the open internet services or from agreements with data providers. The acquired data either corresponds to data obtained from archives or from individual projects.

Methodologies were implemented for homogenisation, quality control and merging of all data sets. Apart from data reduction during quality control and conversion to standard variables, minimum changes were made on the original data. Full details can be found in the publication of Valente et al. [2015, submitted].

The final result is a merged table tuned for validation of satellite-derived ocean colour products and available in a text-format. Metadata of each in-situ measurement (original source, cruise or experiment, principal investigator) are propagated throughout the work and made available in the final table. This is an attempt to gather, standardise and merge several sets of high-quality bio-optical in-situ data. Using all sets of data in a single validation exercise increases the number of match-ups and enhances the representativeness of different marine regimes. By making available the metadata, it is also possible to analyse each set of data separately and explore its impact on ocean colour validation.

1.1 Normalisation of in-situ data

Three different normalisation approaches have been applied, in order to overcome the restriction to case 1 waters of the widely used f/Q methodology by Morel (Morel and Gentili [1996], Morel and Maritorena [2001], short: *fQMorel*). One approach is based on work by Lee et al. [2011] (short: *Lee2011*), while the other follows a methodology of Park and Ruddick [2005](short: *PR05*). As they require different numbers of available wavelengths, the number of spectra changes according to the applied normalisation procedure.

The effects of these normalisations can be summarized as:

- Satellite validation in case 2 waters is concentrated on AERONET-OC data, with a minor contribution

Normalisation	PR05		fQMorel		Lee2011	
site	Total	bs data	Total	bs data	Total	bs
abuAlB.	588	583	588	583	364	364
COVE	553	553	553	553	553	553
Gloria	1101	1099	1101	1099	1055	1054
GDT	1538	1520	1538	1520	1370	1354
HLH	1848	1768	1848	1768	1848	1768
LISCO	1090	1084	1090	1084	1090	1084
Lucinda	283	283	283	283	283	283
MVCO	3824	3823	3825	3824	2882	2882
Palgrunden	1186	1124	1187	1125	1187	1125
AAOT	8759	8757	8759	8737	5001	5001
WaveCIS	2183	2182	2184	2183	2183	2183
boussole	16728	3555	17422	3709	16142	3546
cc	308	1	320	13	314	313
mermaid	0	-	885	678	0	-
moby	5081	-	5081	-	5080	-
nomad	0	-	3326	2831	0	-
seabass	0	-	698	275	0	-

Table 1.1: Amount of normalised remote sensing reflectance spectra. Total: total number at a given site. bs data: bandshift needed. Band-shifting criteria are set to MERIS. Three different normalisations are used, which are f/Q by Morel, a model by Lee and a model by Park & Ruddick.

of COASTCOLOUR data.

- Lee2011 is closer to fQMorel in case 1 waters.

- PR05 is further apart from fQMorel in case 1 waters, but still within a 10% range.

- In case 2 waters, the methods give different results, but it is uncertain, which one is the most accurate.

Both new methods (Lee2011 and PR05) provide less Rrs in-situ data than fQMorel, which has also been applied to past versions of the database. Only MOBY and (to some degree) BOUSSOLE cover case 1 waters, while AERONET-OC and COASTCOLOUR (~300 measurements) provide data for case 2 waters. Other datasets are not normalised because of the unavailability of the viewing geometry (NOMAD) or because the Rrs data already comes normalized with the fQMorel method (SeaBASS, MERMAID).

Regarding the number of points that can be normalised from the available datasets (MOBY, BOUSSOLE, AERONET-OC, COASTCOLOUR), the differences can be dramatic (Table 1.1). PR05 is able to normalize ~100% of the available data (the required input is the chlorophyll concentration and the viewing geometry).

Lee2011 needs Rrs at four specific wavelengths (443, 490, 550-560, 667nm). The amount of data, which can be normalised, depends on the acceptable range of wavelength to these central values. Choosing a range of ±2nm, almost all AERONET-OC data is lost. Also MOBY data averaged to MODIS, MERIS and VIIRS central wavelengths, might be lost. With a range of ±4nm, Lee2011 looses 20% of AERONET-OC (17816 out of 22956), but normalises ~100% of the remaining datasets. According to Lee "drifting of 2-5 nm does not matter much, but depends on objectives. For instance, if it is to look decadal changes using data from different satellites, more precise tuning of this step might be necessary".

The comparison of fQMorel to PR05 and fQMorel with Lee2011 (±4nm) shows that Lee2011 is more similar to fQMorel than PR05 (fig. 1.2, example at 490nm). PR05 generates higher Rrs values than fQMorel. Typically the correction with fQMorel decreases Rrs in-situ (up to 20% in high chlorophyll-a, like Palgrunden with a mean chlorophyll concentration of 8±2 µg/l). With PR05 the correction is less strong than with fQMorel, which is most likely not applicable in the Baltic Sea with its case 2 waters.

1.2   Selection of Normalisation

After some discussion it has been decided to use the PR05 normalised in-situ data and combine it with all case 1 data (definition see section 2.2) from the NOMAD dataset, which is normalised by fQMorel.
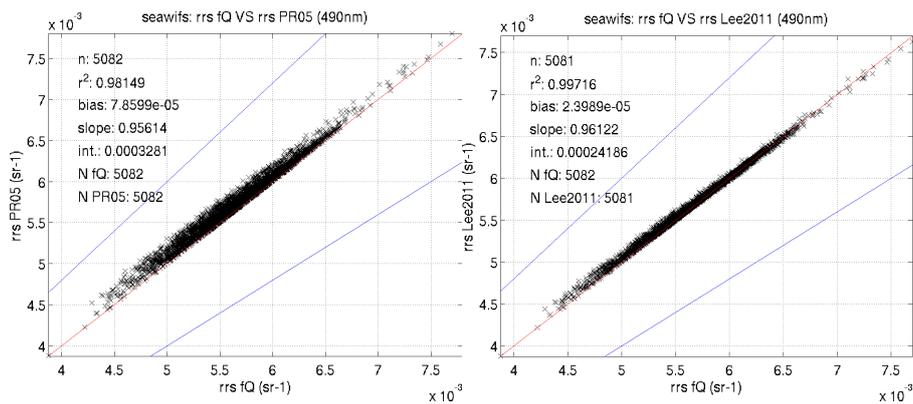
(a) MOBY (SeaWiFS bands)
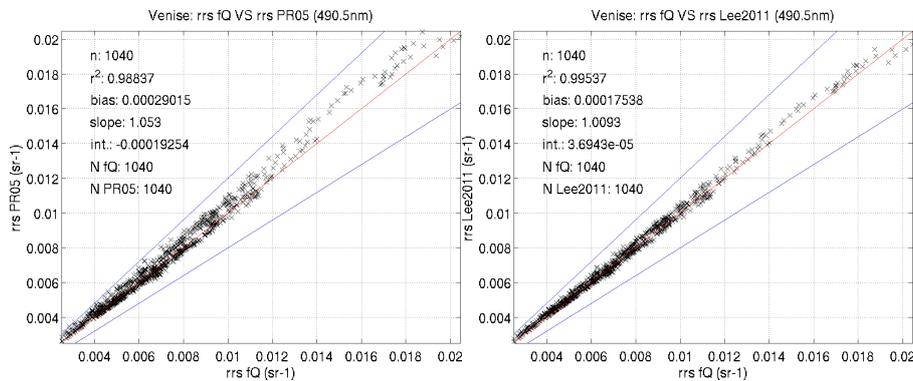
(b) AAOT
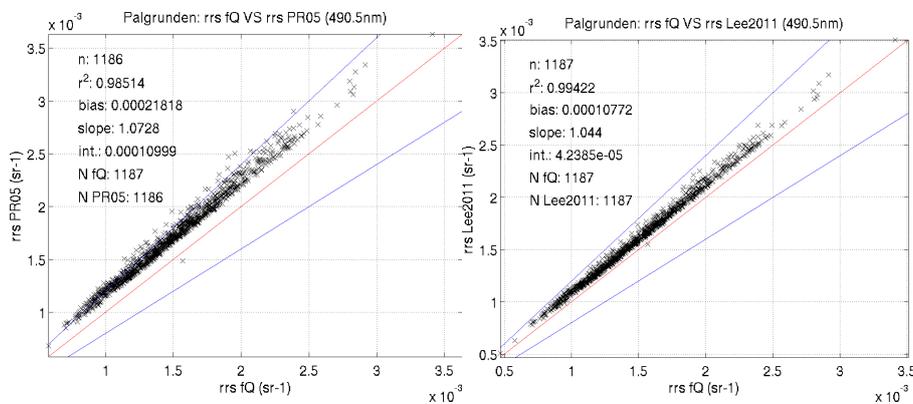
(c) Palgrunden
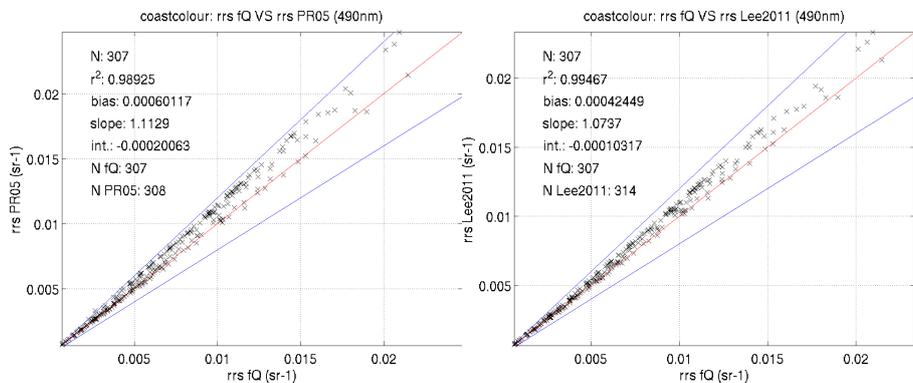
(d) Coast Colour dataset

Figure 1.2: Comparison of normalised in-situ remote sensing reflectances at 490nm. Lee2011 uses measurements within ±4nm of 443, 490, 550-560, 667nm.

7

2 Band-shifting the in-situ database

Starting from the OC_CCI in-situ database, it is necessary to tune it to the validation of each satellite sensor independently. By careful selection of spectra and a band-shifting procedure, the measured normalised remote sensing reflectances are converted to the appropriate wavelengths.

Summary

Between Round Robin 2 (RR2) and RR3, *major changes* have been implemented in the band shifting approach:

- Band shift relies on the all available measured spectral points instead of using only the preselected measurements within ±6nm of the satellite bands respectively.

- The models for the estimation of chlorophyll concentration and IOPs follow the parameterisation by Zibordi, but representative coefficients are derived directly from the OC_CCI in-situ data. Mainly the NOMAD dataset provides simultaneously IOP and Rrs measurements (normalised with fQMorel). For four water classes based on the Lee case 1 definition (section 2.2), these models are fitted to the in-situ data.

- During band-shift, the models are selected based on radiometric properties alone, which allows for variability of water classes at single sites.

2.1 Data selection

2.1.1 OC_CCI in-situ database version 3.0

The in-situ database combines remote sensing reflectances, which have been normalised with PR05. In addition, all measurements in case 1 waters of the NOMAD dataset (fQMorel normalisation) are included. The fQMorel approach has been developed for case 1 waters, so that no degradation of the in-situ is expected and the inconsistency in normalisation procedures is an acceptable trade off for more and globally distributed data points.

The amount of measured bands per spectrum can vary within a single site. For an overview, the histograms show the number of spectral points within each spectrum (Fig. 2.1, histograms to the left). The second histogram (Fig. 2.1, to the right) gives the amount of measurements per satellite band, which can differ in actual to nominal wavelengths up to 6 nm. The red columns highlight the amount of measurements, which deviate more than ±1 nm from the nominal wavelength and should be band-shifted.

The amount of spectra can change with the choice of a normalisation procedure, as these need different sets of available reflectance measurements in order to be applicable. Data points from NOMAD (and partially also from MERMAID and seabass) are included only in the fQMorel data set (tab. 1.1). They are key to the development of a new set of IOP model parameterisation. Even if PR05 might be the most appropriate approach to cover both case 1 and case 2 type waters, for this modelling step the fQMorel data remains the basis.

2.1.2 Selecting data for band-shifting

For the band-shifting the wavelengths up to 682 nm (previous PVASR version: 670 nm) are considered. The following criteria are used to select spectra which carry sufficient information, so that band-shifting can be applied successfully.

- If less than three measurements in this range (400-682 nm) exist and some of them need band-shifting, these measurements are discarded.

- If less than three measurements in this range exist and they do not need to be shifted, these measurements are collected in the result file. Missing values can not be estimated in a reasonable fashion.

- If all measurements in the range are available and do not need to be shifted, these data points become part of the result file immediately.

- If at least three wavelengths have been measured, but the value at a band close to the limits of the range is not available (e.g. Rrs at 412±6nm and 670±6 nm), the measurements are discarded.
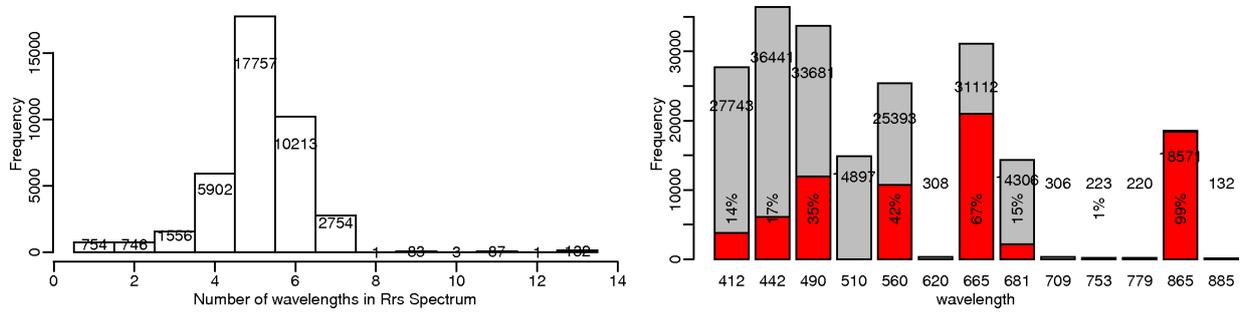
(a) MERIS
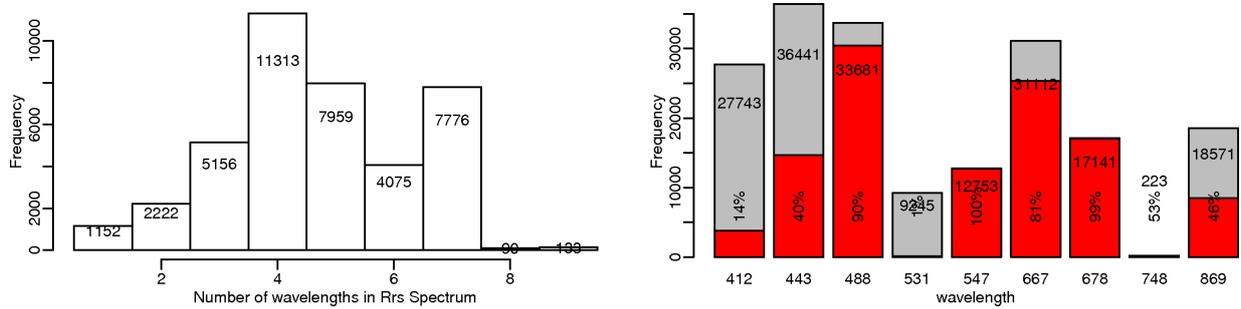
(b) MODIS
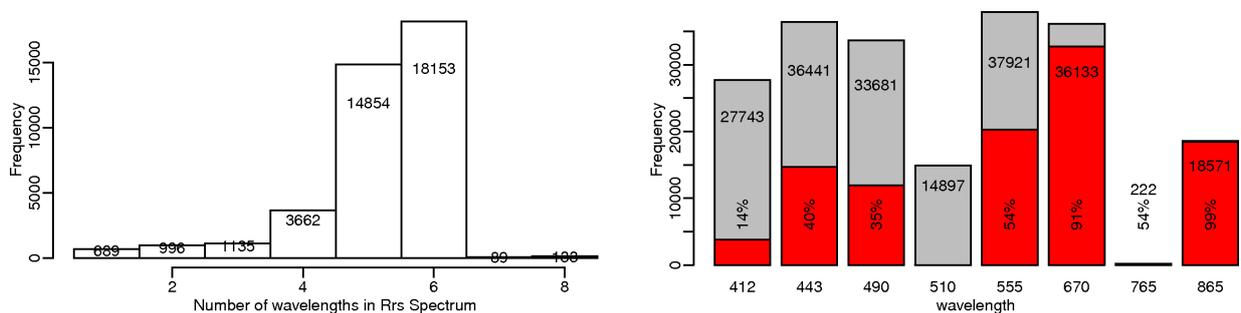
(c) SeaWiFS
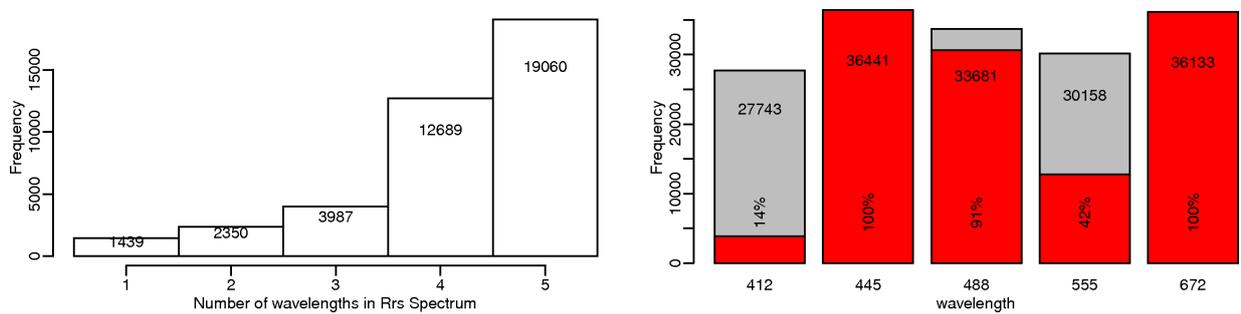
(d) VIIRS

Figure 2.1: Amount of spectra with measurements at N wavelengths. Amount of measurements at specific wavelengths. The histograms are based on measurements taken at ±6nm of the nominal central wavelength. (Normalisation: PR05)

9

The hyperspectral in-situ data from MOBY are folded with the spectral response functions of the different satellite sensors, so that band-shifting is not needed.

All other spectra undergo an iterative band-shift approach.

2.2 Definition of Case 1 and Case 2 waters

Since RR2 the identification of case 1 waters is handled by the following approach of Lee and Hu [2006].

It can be used to divide the not-case 1 waters into different groups as well, which provides a simple water classification based on remote sensing reflectance information only.

2.2.1 Remote-sensing criterion for Case 1 waters (Lee and Hu [2006])

This model uses four wavelengths of the remote sensing reflectance spectrum. With fixed parameters $\gamma = 0.1$ and $\eta = 0.5$ a range around a theoretically perfect case 1 water is constructed:

$$RR12_{CS1}(RR53) = 0.9351 + 0.113/RR53 - 0.0217/(RR53)^2 + 0.003/(RR53)^3$$

$$Rrs555_{CS1}(RR53) = 0.0006 + 0.0027 \cdot RR53 - 0.0004 \cdot (RR53)^2 - 0.0002 \cdot (RR53)^3$$

with $RR53 = Rrs555/Rrs490$ and $RR12 = Rrs412/Rrs443$.

The water is classified as case-1, if the spectrum fulfils the following criteria simultaneously:

$$(1 - \gamma)RR12_{CS1}(RR53) \leq RR12 \leq (1 + \gamma)RR12_{CS1}(RR53) \tag{1}$$

$$(1 - \eta)Rrs555_{CS1}(RR53) \leq Rrs555 \leq (1 + \eta)Rrs555_{CS1}(RR53) \tag{2}$$

The ratio RR12 represents the relative abundance of CDOM per chlorophyll (Carder et al. [1999]), RR53 is viewed as a measure of chlorophyll (e.g. Aiken et al. [1995]; O'Reilly et al. [1998]), and Rrs555 is a measure of particle backscattering (Carder et al. [1999]).

2.2.2 Remote-sensing criteria for Case 2 waters

Based on the inequalities which define the case 1 water (eq. 1 and eq. 2) it is a simple matter to specify further water types. Due to the availability of simultaneous in-situ measurements of IOPs and Rrs, it is reasonable to suggest the following three types, which focus on the CDOM to chlorophyll absorption ratio.

1. *Higher CDOM/Chl than case 1*: Only equation 1 is exploited. All spectra with $RR12 < (1-\gamma)RR12_{CS1}(RR53)$ are considered to be of this type.

2. *Lower CDOM/Chl than case 1*: Only equation 1 is exploited. All spectra with $RR12 > (1+\gamma)RR12_{CS1}(RR53)$ are considered to be of this type.

3. *Backscattering differs from case 1*: Although equation 1 is fulfilled, the spectrum shows either higher or lower backscattering than expected in case 1 waters ($Rrs555 > (1 + \eta)Rrs555_{CS1}(RR53)$ or $Rrs555 < (1 - \eta)Rrs555_{CS1}(RR53)$).

2.2.3 Example of classification

The case 1 model was applied to the band-shifted in-situ dataset and the two criteria (equation (1), equation (2)) are visualised as upper and lower boundaries around the "exact" case 1 model (blue line, figure 2.2). The relationship between RR12 and RR53 indicates the ratio of CDOM to chlorophyll concentration. Points below the lower boundary show an excess of CDOM, while there is a lack of CDOM modelled for those points above the upper boundary. The relationship between Rrs555 and the model of RR53 indicate the influence of particle backscattering. If points are above the upper boundary, there is a larger amount of particle scattering present than would be expected from a perfect case-1 water body. The boundaries are defined as $\pm 10\%$ deviation of the model RR12 and $\pm 50\%$ deviation of Rrs555, respectively.
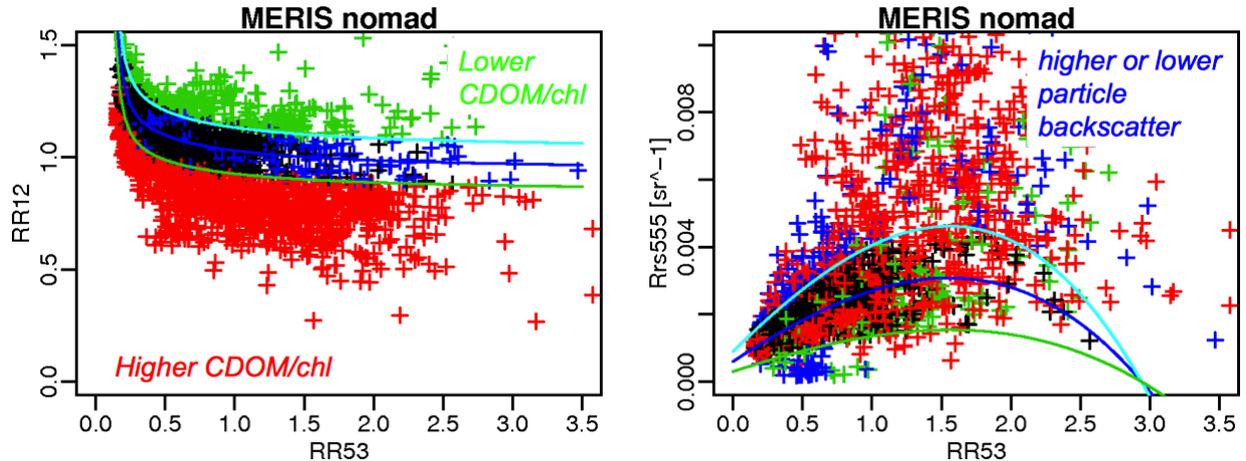
Figure 2.2: Dividing the NOMAD dataset into four water types: case 1 definition (Lee 2006) in black, higher CDOM/chl ratio (red), lower CDOM/chl ratio (green) and case 1 CDOM/chl but higher or lower backscattering (blue).

2.2.4 Effects on case 1 classification

Each spectrum of the OC_CCI v3.0 in-situ database (to central wavelengths of four sensors) is classified with the four water types based on Lee's case 1 definition during the iterative band-shift procedure. A considerable amount of spectra is classified as case 2 with higher CDOM/chl absorption (see fig. 2.3) for MERIS band-shifted in-situ data. Classifications based on data at SeaWiFS, MODIS and VIIRS wavelengths can be found in section 8.1. Only at Boussole and MOBY the case 1 conditions dominate the remote sensing reflectance spectra (table 2.1).

For the in-situ database, those band-shifted to MERIS wavelengths, 25.5% of the data points (8599 of 33748) are identified as strictly case 1 waters (MODIS: 26.4%, 8940 of 33909; SeaWiFS: 27.5%, 9309 of 33852; VIIRS: 18.6%, 5773 of 31037).

2.3 Revision of IOP model parameterisation

After applying the bandshift code of OC_CCI v2 (BSv2), strange results were found for some VIIRS examples (fig. 2.4), which led to an investigation on the influence of the underlying IOP models. The differences between BSv2 and BSv3 arise mainly from correcting an error in the adjustment factor. Although the sun irradiance is needed, if radiances are bandshifted, this is not correct for reflectances. As E0 is strongly structured, all newly bandshifted spectra are much smoother in their appearance than in BSv2.

In the course of the investigation, it became clear, that Lee's QAA model for IOPs is by design not suitable for an iterative approach. The model tries to fit the given Rrs spectrum exactly, so that the iterative process sticks with the initialised missing values. For example, if the spectrum is estimated between 410 and 670 nm in 5 nm steps, the iterative approach works fine with the Zibordi model (AAOT coefficients), but negligible changes are calculated with QAA (fig. 2.5).

Although the site-specific IOP model works well in this example (as this is the site, the model has been developed for), it has been investigated, how well the two sets of coefficients for AAOT and Gustav-Dalen-Tower (GDT) actually represent the inherent optical properties in the OC_CCI global in-situ dataset.

In order to take advantage of all combined measurements of Rrs and IOPs, which are in this case absorption of detritus and yellow substance, adg, and backscattering of particles, bbp, the database with f/Q-Morel normalisation is used. The global distribution of these data points is shown in Fig. 2.6.

As not all specified sites come with both types of measurements, we decided to use the Lee case 1 classification and split up the measurements into four water type classes: case 1 water, water with higher CDOM/Chl absorption, water with lower CDOM/Chl absorption, water with medium CDOM/Chl absorption but backscattering is not case 1 type (see section 2.2).

site	MERIS				MODIS				SeaWiFS				VIIRS			
	case 1	case 2 H	case 2 L	case 2 b	case 1	case 2 H	case 2 L	case 2 b	case 1	case 2 H	case 2 L	case 2 b	case 1	case 2 H	case 2 L	case 2 b
aoc Abu Al Bukhoosh	116	452	0	15	120	440	0	23	116	453	0	14	106	466	0	11
aoc COVE SEAPRISM	15	511	0	27	15	516	0	22	16	516	0	21	15	522	0	16
aoc Gloria	49	950	0	100	79	951	0	69	59	952	0	88	50	973	0	76
aoc Gustav Dalen Tower	107	1348	1	64	116	1371	0	33	103	1368	0	49	88	1387	0	45
aoc Helsinki Lighthouse	146	1567	1	54	134	1597	1	36	134	1593	1	40	122	1605	1	40
aoc LISCO	38	1015	0	31	36	1043	0	5	32	1042	0	10	25	1052	0	7
aoc Lucinda	0	281	0	2	0	281	0	2	0	281	0	2	0	281	0	2
aoc MVCO	123	3568	0	133	138	3595	0	91	124	3591	0	109	98	3640	0	86
aoc Palgrunden	51	1072	0	2	49	1076	0	0	50	1075	0	0	45	1078	0	2
aoc Venice	551	6191	1	1994	543	6246	0	1948	544	6317	1	1875	537	6511	0	1689
aoc WaveCIS Site CSI 6	221	1850	0	112	238	1868	0	77	226	1862	0	95	194	1909	0	80
boussole	2000	1282	465	7	2446	754	501	53	2161	1106	476	11	2190	1086	468	10
mermaid BioOptEuroFleets	0	25	0	8	0	25	0	8	0	25	0	8	0	26	0	7
mermaid SIMBADA	166	145	24	46	117	121	36	107	141	137	29	74	144	139	24	74
nomad	1205	1186	231	222	1221	1031	304	288	822	645	172	106	1230	1155	236	223
seabass val	119	68	40	48	101	67	57	50	51	17	17	14	114	70	43	48
moby	3692	7	2	1	3587	0	257	19	4730	0	348	3	815	0	175	1
Total	8599	21518	765	2866	8940	20982	1156	2831	9309	20980	1044	2519	5773	21900	947	2417

Table 2.1: Amount of case 1 and case 2 water spectra in OC_CCI v3.0 band-shifted database (Normalisation: fQMorel).
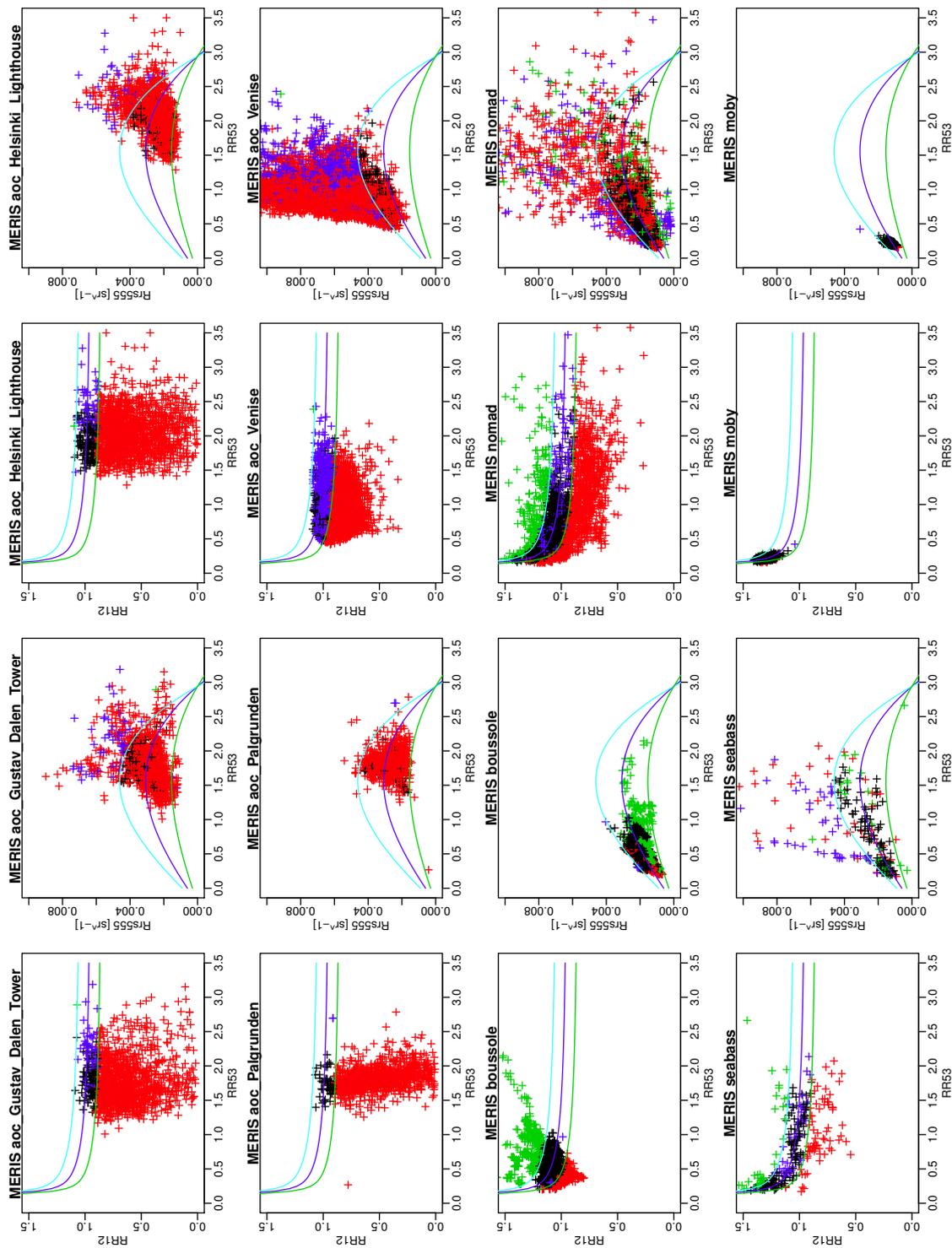
Figure 2.3: Classification of OC_CCI v3.0 in-situ database band-shifted to MERIS wavelengths per site, distinguishing four types of water: case 1 definition (Lee 2006) in black, higher CDOM/chl ratio (red), lower CDOM/chl ratio (green) and case 1 CDOM/chl but higher or lower backscattering (blue).

(a) Bandshift v2
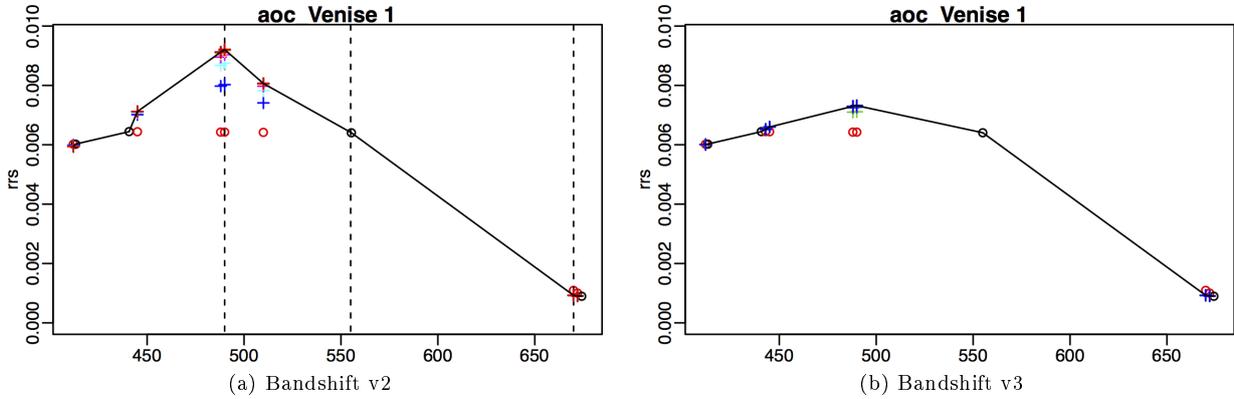
(b) Bandshift v3

Figure 2.4: Example of iterative bandshift on an aoc_AAOT spectrum at VIIRS wavebands. Black circles are measurements, red circles are starting points of the iteration process. The line connects measurements and results of the bandshift. Differences arise mainly from new IOP model and from an error in v2, which uses E0 in the adjustment, although this is not correct for reflectances. The jump in the blue is avoided in v3.

(a) Bandshift v3 beta, IOPs Zibordi coefficients for AAOT

(b) Bandshift v3 beta, IOPs QAA
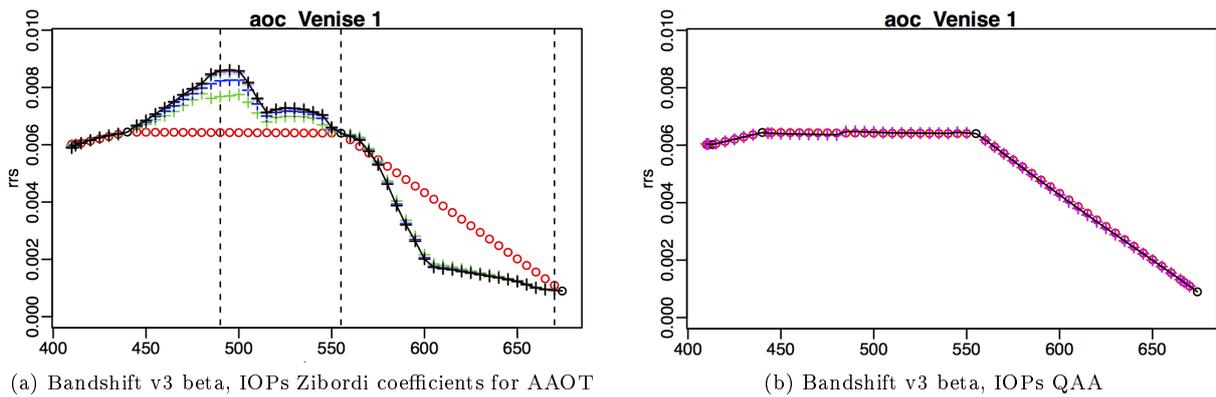
Figure 2.5: Example of iterative bandshift on a aoc_AAOT spectrum at full spectrum with 5nm steps. Black circles are measurements, red circles are starting points of the iteration process. (For reflectances: no E0 in the adjustment.) The QAA IOP model is designed to fit a given Rrs spectrum perfectly, therefore it is not suitable for the iteration approach.

14
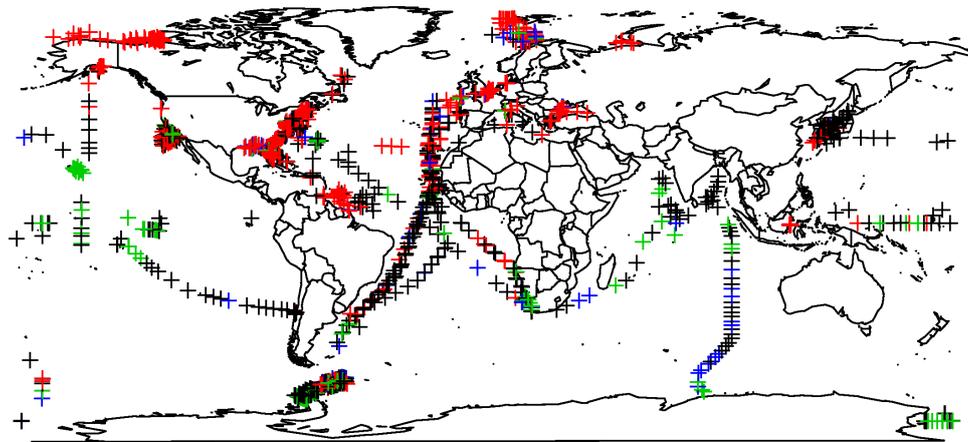
(a) Chlorophyll conc.
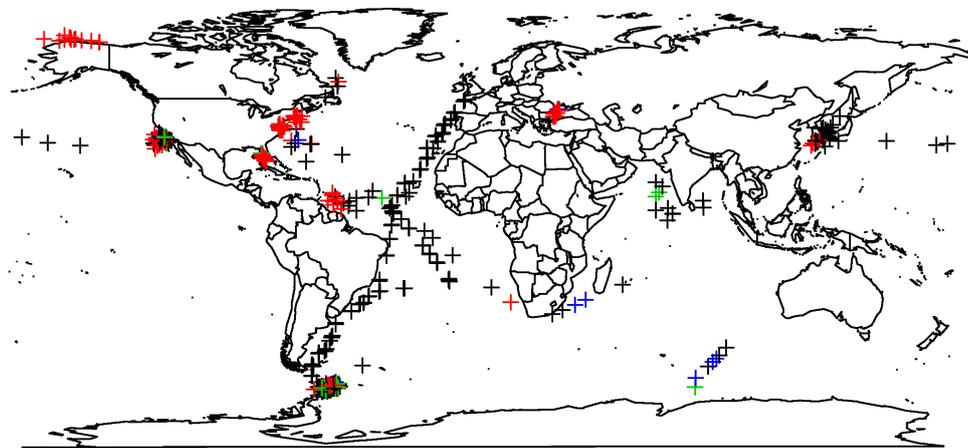
(b) adg
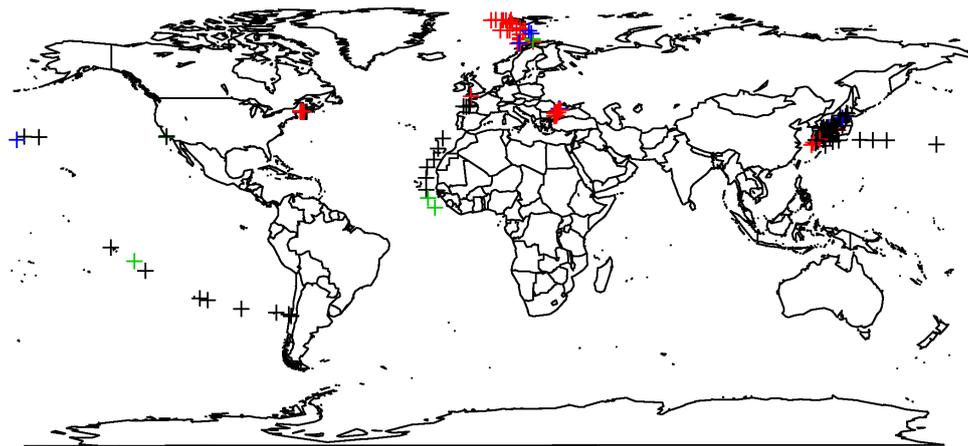
(c) bbp

Figure 2.6: Measurements of chlorophyll conc., adg and bbp with remote sensing reflectances in the OC_CCI in-situ database v3 (Normalisation f/Q Morel). Colours represent the four water classes: case 1 (black), higher CDOM/chl abs. (red), lower CDOM/chl abs. (green), case 1 abs but stronger scattering (blue).

	a0	a1	a2	a3
case 1	0.3559	-2.0932	0.7760	-2.0278
high CDOM/chl	0.1218	2.5334	0.9029	0
low CDOM/chl	0.5064	-1.728	-0.5352	1.3864
medium CDOM/chl, high backscatter	0.3352	-2.4827	0	0.6780

Table 2.2: Chlorophyll model. Parameter for four water classes.

	Slope	b0	b1
case 1	0.0132	-0.5475	-0.4802
high CDOM/chl	0.0141	0.0422	-0.7784
low CDOM/chl	0.0128	-0.3852	-0.6726
medium CDOM/chl, high backscatter	0.0134	-0.0061	-0.8737

Table 2.3: Parameterisation of a_{dg} model.

2.3.1 Chlorophyll concentration

The new chlorophyll models (yellow lines in fig. 2.7) give higher chlorophyll estimates than the AAOT or GDT parameterisation. The AAOT model does not represent the chlorophyll concentration to Rrs ratio of the OC_CCI in-situ data set very well. It is more similar to the all water samples, where the CDOM absorption is higher compared to a given chlorophyll absorption, than a case 1 model would suggest.

The chlorophyll concentration is modelled with a polynomial function of third degree with $R = \log_{10} Rrs490/Rrs555$

$$\log_{10} chl = a_0 + a_1 R + a_2 R^2 + a_3 R^3 \tag{3}$$

Only statistically significant coefficients are taken into account (tab. 2.2). Of course, in a direct comparison between measured and modelled chlorophyll concentration the new model (which is fitted directly to the data) leads to significantly better results than the polynomial model with the AAOT coefficients, which has been used before (Fig. 2.8). Especially smaller concentrations have been poorly represented before and heavily underestimated in the previous version.

2.3.2 Absorption of total dissolved matter, adg

In the same manner, the relationship between Rrs 490 and 670 nm ($R = \log_{10} Rrs490/Rrs670$) and absorption of detritus and gelbstoff a_{dg}, has been revised (tab. 2.3). It replaces the estimations of $a_y + a_{dp}$.

$$\log_{10} adg412 = b_0 + b_1 R \tag{4}$$

$$adg(\lambda) = adg412 \cdot \exp\left(-Slope \cdot (\lambda - 412)\right) \tag{5}$$

A comparison of modelled a_{dg} or $a_{dp} + a_y$ values against measured values is shown in Fig. 2.9. The model with AAOT coefficients tends to underestimate the absorption of a_{dg}.

2.3.3 Backscattering of particles

The backscattering is revised as well.

With $R = \log_{10} L_{wn}(490)/L_{wn}(555)$ the backscattering is modelled with an exponential decline (see table 2.4):

$$\log_{10} bb_p(510nm) = d_0 + d_1 \cdot R$$

$$\log_{10} bb_p(\lambda) = \log_{10} bb_p(510nm) \cdot (\lambda/510)^{-S_{bp}}$$

The previous model overestimated the measured particle backscattering (Fig. 2.10).
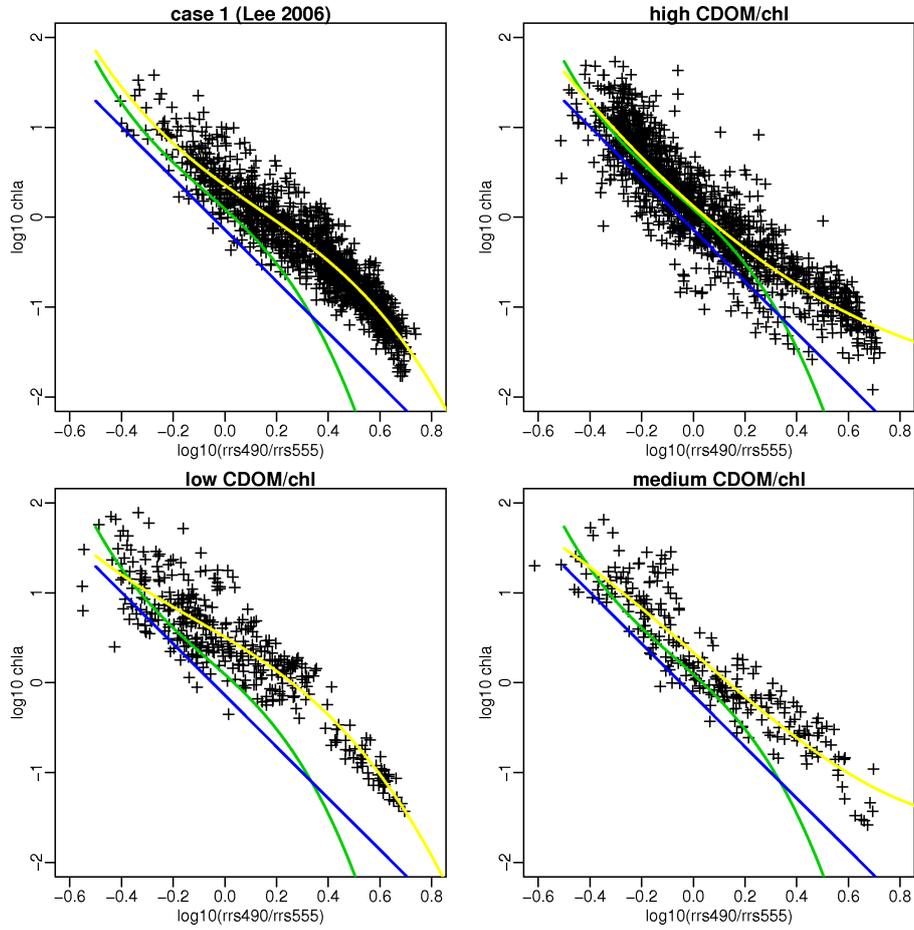
Figure 2.7: Chlorophyll models based on the OC_CCI v3.0 in-situ data and four water type classes (based on Lee definition). Green line represents the Zibordi AAOT model, the blue line the Zibordi GDT model. The yellow lines are the four models, which are used in the BS v3 for this study.

	S_{bp}	d_0	d_1
case 1	1	-2.3816	-1.3803
high CDOM/chl	1	-2.2632	-1.8359
low CDOM/chl	1	-2.4386	-1.2018
medium CDOM/chl, high backscatter	1	-2.2011	-2.0701

Table 2.4: Parameterisation of backscattering model.
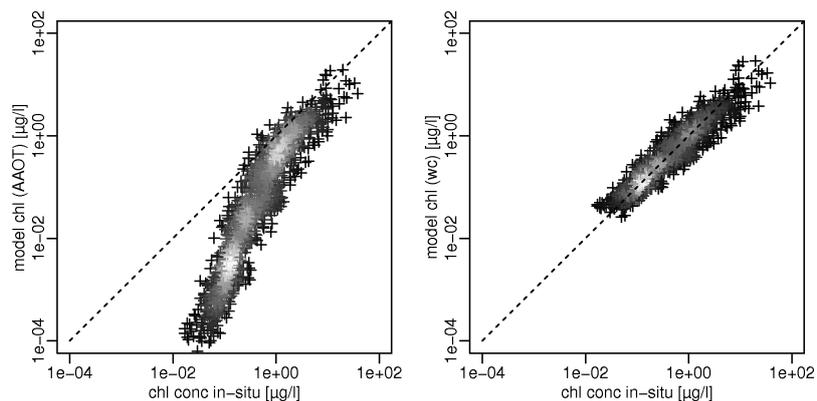
(a) case 1 Lee 2011
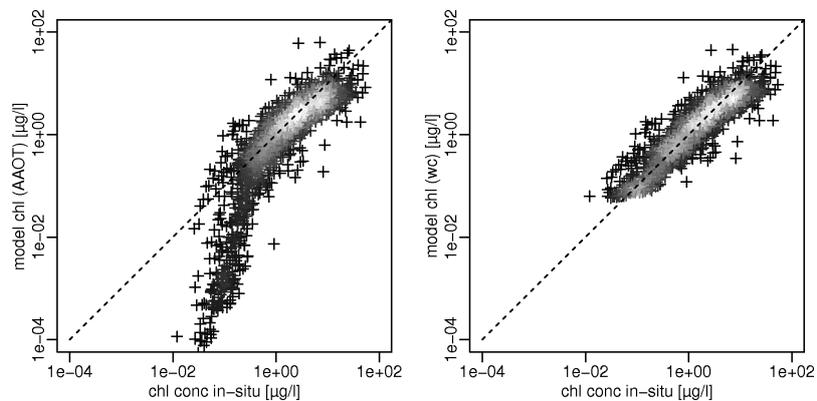
(b) higher CDOM/chl absorption
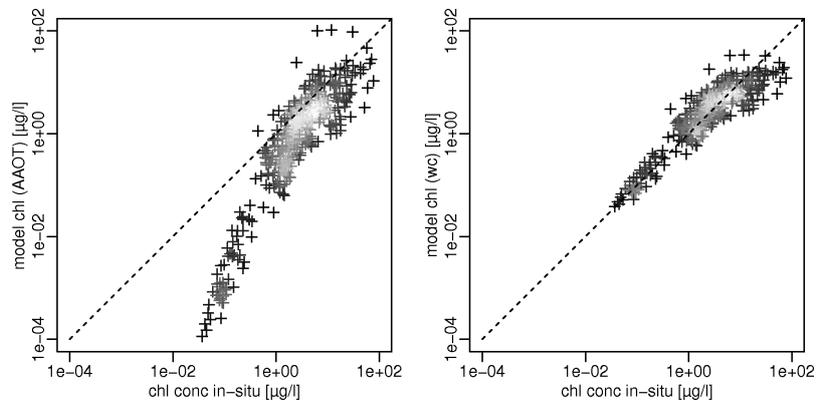
(c) lower CDOM/chl absorption
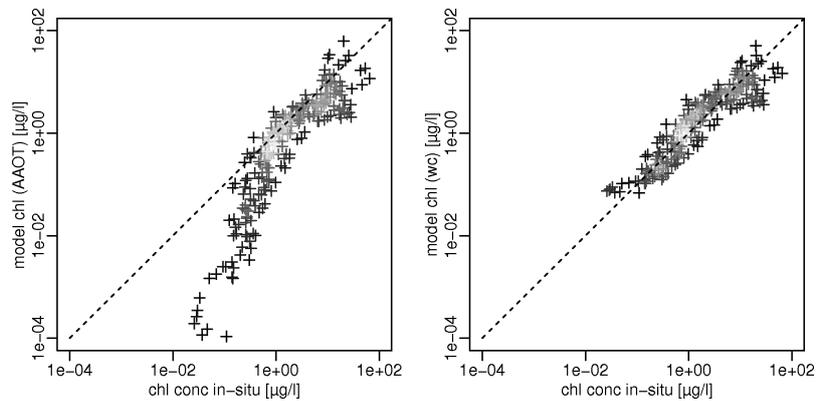
(d) medium CDOM/chl absorption, but higher scattering

Figure 2.8: Modelled chlorophyll concentration versus in-situ chla_fluor or chla_hplc.
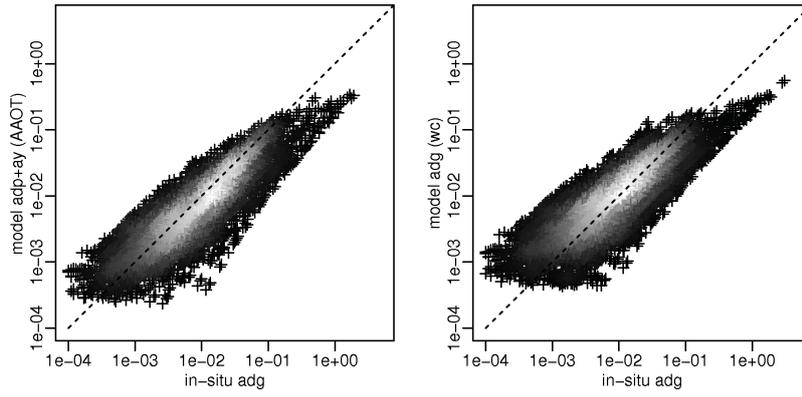
(a) case 1 (Lee 2006)
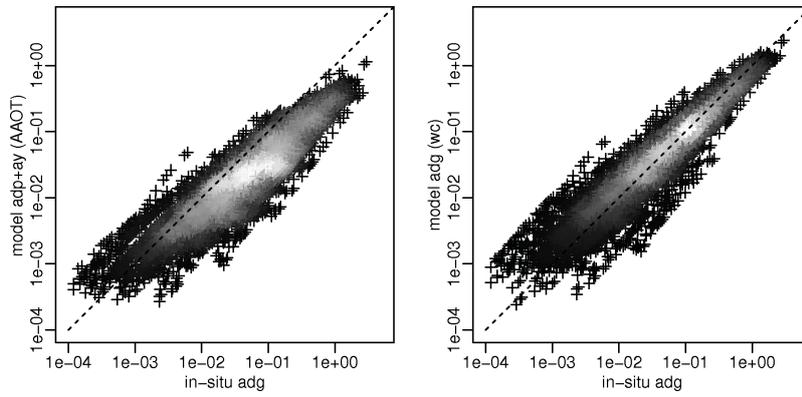
(b) Higher CDOM/chl absorption
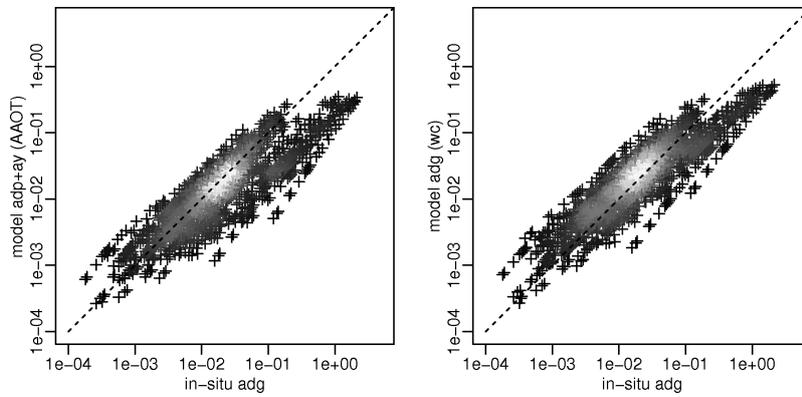
(c) Lower CDOM/chl absorption
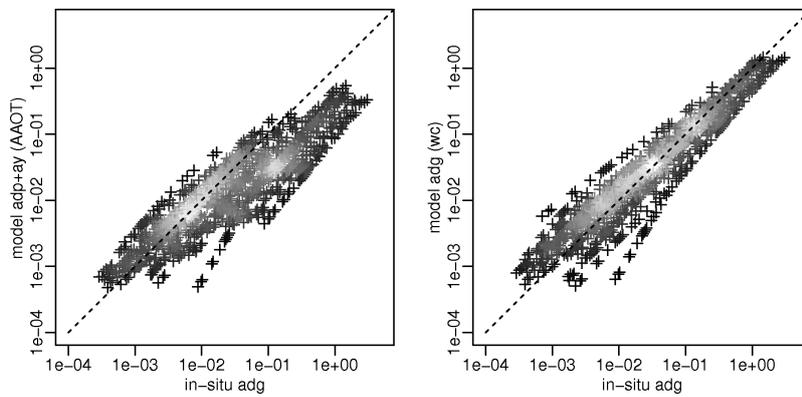
(d) Medium CDOM/chl absorption, but higher backscattering

Figure 2.9: Modelled a_{dg} or $a_{dp} + a_y$ (AAOT) versus in-situ a_{dg}. All spectral values are joined.
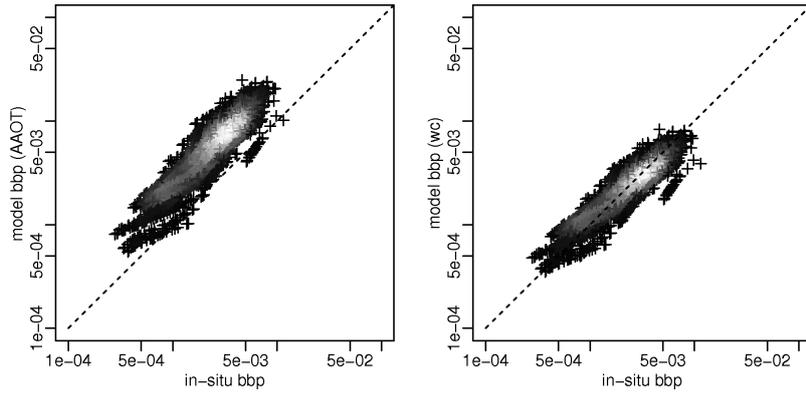
(a) case 1 (Lee 2006)
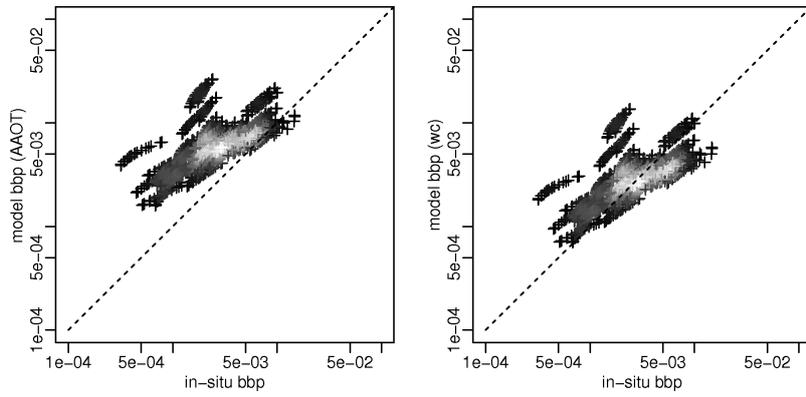
(b) Higher CDOM/chl absorption
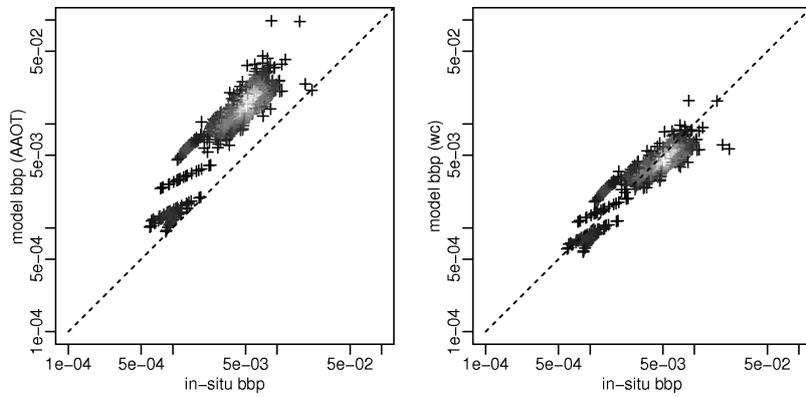
(c) Lower CDOM/chl absorption
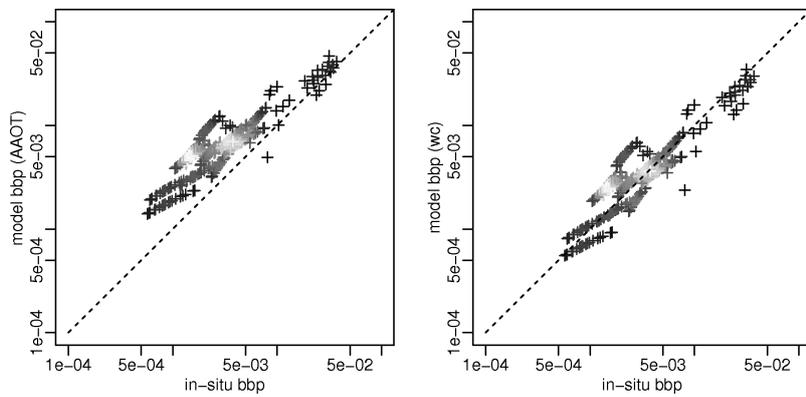
(d) case 1 CDOM/chl absorption, but higher backscattering

Figure 2.10: Modelled bb_p versus in-situ bb_p.

2.4 Band-shifting procedures

The bandshifting procedures need some estimation of IOPs and chlorophyll from the given remote sensing reflectances. Two different IOP models are implemented: QAA v5 and the parametric models from section 2.3. Three different types of corrections are implemented: two iterative corrections are based on f/Q tables (with corrections based on one or two nearest measurements to the target), while the third one uses the forward mode of QAA to derive correction factors without iteration.

The iterative approach is robust and works well even with very limited numbers of available measurements.

For each satellite sensor the nominal wavelengths of the spectral bands are given as

MERIS = 412, 442, 490, 510, 560, 620, 665, 681, 709, 753, 779, 865, 885 nm,

MODIS = 412, 443, 488, 531, 547, 667, 678, 748, 869 nm,

SeaWifS = 412, 443, 490, 510, 555, 670, 765, 865 nm,

VIIRS = 412, 445, 488, 555, 672 nm.

The models for IOPs and chlorophyll follow the descriptions in section 2.3.

2.4.1 Iterative method employing f/Q

The nominal wavelengths and those needed for IOP estimation are combined and the unique set of wavelengths is selected and sorted, while wavelengths above 682 nm are discarded. This limit was 678 nm in the earlier BS v2. Each spectrum is considered separately. Measurements at all available wavelengths are taken into account. The combination of nominal wavelengths and those needed for IOP modelling are initialised, either with actual measurements (within ± 1nm) or by linear interpolation between measured spectral points. Duplicates of wavelengths after copying values in the ± 1nm range are discarded, only the targeted nominal wavelength is kept. For each target wavelength, the two closest Rrs measurements are identified (λ_{start1}, λ_{start2}). From the initialised Rrs values the water class and the chlorophyll conc and IOPs are estimated.

The procedure follows Zibordi et al. [2009] to some length, but combines it with an iterative approach and modelling the Rrs

$$Rrs\left(\lambda_{target}, \lambda_{start}\right) = Rrs\left(\lambda_{start}\right) \frac{\frac{f}{Q_0}\left(\lambda_{target}, chl\right) \frac{b_b(\lambda_{target})}{a(\lambda_{target})}}{\frac{f}{Q_0}\left(\lambda_{start}, chl\right) \frac{b_b(\lambda_{start})}{a(\lambda_{start})}} \tag{6}$$

with $\frac{f}{Q}\frac{bb}{a}$ in order to allow for larger differences between start and target wavelength (communication Mélin).[1] All constants (f/Q, and water optical properties) are linearly interpolated towards the exact wavelength of the measurements.

The chlorophyll concentration is estimated from Rrs at 490 and 555 nm. Estimates for CDOM absorption a_y and non-pigmented particles absorption a_{dp}, combined in a_{dg} and pigmented particle (phytoplankton) absorption a_{ph}, particle backscattering b_{bp} are calculated and combined with tabulated water absorption a_w and backscattering b_w to total absorption and backscattering at all designated and measured wavelengths.

The number of iterations is restricted to a maximum of 20 initially. As a start wavelength, the two closest measured wavelengths to the target are selected. After calculating both estimates at the target wavelength, they are combined in a weighted (inverse distance) mean. Although differences between start and target wavelength can be rather large, the procedure converges almost always - an important improvement compared to the previous version, which suffered from divergent or alternating behaviour in the estimates.

$$Rrs\left(\lambda_{target}\right) = w_1 Rrs(\lambda_{start\,1}) + w_2 Rrs(\lambda_{start\,2})$$

The convergence limit is defined as the mean absolute deviation in Rrs between two iteration steps over the entire spectrum, which should be lower than $\epsilon = 10^{-9}$ and

$$\frac{1}{N_\lambda} \sum_{i=1}^{N_\lambda} \left\| Rrs\left(\lambda_i, n\right) - Rrs\left(\lambda_i, n+1\right) \right\| < \epsilon \tag{7}$$

---

[1]There has been an error in the previous version. Although the water leaving radiance is modelled as $L_{wn}\left(\lambda\right) = E_0\left(\lambda\right)\frac{f}{Q_0}\left(\lambda, chl\right)\frac{b_b(\lambda)}{a(\lambda)}$ and therefore includes E_0, the remote sensing reflectance is independent from E_0. $Rrs(\lambda) = L_{wn}\left(\lambda\right)/E_0\left(\lambda\right)$.

(a) Bandshift v3 beta, IOPs Zibordi coefficients for AAOT

(b) Bandshift v3, water class IOPs
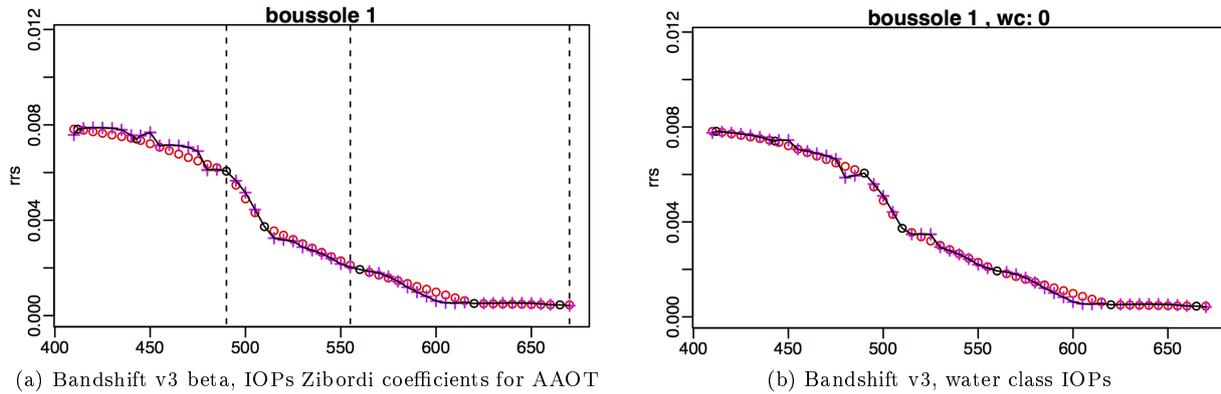
Figure 2.11: Example of iterative bandshift on a boussole spectrum at full spectrum with 5nm steps. Black circles are measurements, red circles are starting points of the iteration process. The AAOT IOP model covers the water type at BOUSSOLE not as well as the new model for water class 0 (which corresponds with the Lee case 1 definition)

If the series of chlorophyll values is converging but the convergence limit is not reached, the maximum number of iterations is raised stepwise by 20 iterations up to a total of 100. The values at the missing designated wavelengths are replaced by these estimates. If the convergence limit is not met, these values are used as the new input data for a next step in the iteration. At the moment, the estimated water class stays fixed throughout the iteration process.

For quality control purposes the band-shifting can be performed for each site or collection of measurements separately, and each Rrs spectrum is plotted with all the iterative steps in between. A second plot shows the estimated chlorophyll concentration per iteration step.

Example The new water class based IOP estimates lead to a bandshift, which gives physically reasonable results and reflects the OC_CCI in-situ database (example fig. 2.11).

2.4.2 Definition of band-shift quality flags

The outcome of this procedure is assigned a quality flag per spectrum. The first part of the flag string consists of a numerical character for each satellite band. The flag value codes the information about the band-shift procedure itself. Only measurement points can be start points. The considered start wavelength is the closest measured wavelength to the estimation point.

- Flag value 0: no band-shift necessary for this band, measurement has been within ±1 nm of the designated wavelength.

- Flag value 1: band-shift has been applied and distance between start and target wavelength has been ± 10nm.

- Flag value 2: difference between start and target wavelength is larger than ±10 nm.

The second part of the flag is introduced by an underscore "_". The flag value constitutes the state of the convergence of the error series.

- Flag value 0: the series converges strongly, i.e. the convergence limit is reached in less than 20 iterations.

- Flag value 1: the series converges, but the convergence limit is not reached after 20 iterations.

- Flag value 2: the series diverges.

In addition the number of iterations are kept in the final band-shifted dataset.

2.4.3 Selection of band-shifted data

In the next step, the band-shifted data is prepared for easy application.

- If the series of estimation diverges, only the original values are kept, while estimations are discarded.

- If the series converges, only estimations made within the ±10 nm interval between start and target wavelength are kept.

Bandshifted and original data are recombined, discarding all spectra with less than 3 spectral data points.
   The data points, which do not comply with these criteria, are set to NA - so each remaining data point in the band-shifted database can be used as they are.

2.5 Ideas for future versions

Further information on the band-shift procedures and their validation can be found in an up-coming technical note or publication.

2.5.1 Tested

- Test successive updating of band-shift results, using them as new starting values within the same iteration. This approach leads to divergence often and is therefore discarded.

- As QAA fits a given Rrs spectrum almost perfectly by design, it is not suitable for the iterative approach.

- Often IOP and Rrs data for sites is not available and therefore it is not feasible to provide site specific IOP models everywhere. The IOP model is no longer chosen according to sites, as the water characteristics can vary strongly at a single station over time.

- Instead, it has been decided to use a simple water type classification (Lee case1/case2 classification) based on Rrs measurements and IOP models have been fitted to these partial datasets.

- Update the water type classification with each iteration, as it might change with the on-going estimation of the spectrum. The water type is provided in the output.

2.5.2 Untested

- Use sensor response functions per band to convolute the spectral constants like extra-atmospheric solar irradiance E_0, water absorption and backscattering, and bi-directional correction f/Q.

- Instead of water types based on Lee case 1 definition, the optical water type classification of the OC_CCI project can be applied. If the number of in-situ measurements in each of the 14 classes is sufficiently large, the IOP model could be refined.

- The Lee case 1 and derived case 2 definitions need to be checked against the water type classification or spectra, which represent known water types. Both fixed factors in the Lee algorithm could be adjusted. In particular, the identification of scattering waters seems rather poor.

3 Round Robin 3 Protocol

Band-shifting of in-situ data to wavelegths of satellite sensors, as described in the previous section, is the first step towards a match-up database, which combines the in-situ data with satellite data. Now further measures can be taken.

Firstly, it is necessary to define the sets of quality flags per AC processor, which allow us to choose valid, high quality satellite products (section 3.1.2). In a second step, the aggregation methodology needs to be defined (section 3.2). In the course of the inter-comparison, different subsets of data points are used, which are selected by using individual and combined flags (IBQ: Individual Best Quality, CBQ: Common Best Quality) and are restricted to case-1 waters only (section 2.2).

3.1 Atmospheric Correction processors

For the intercomparison of MERIS based normalised water leaving reflectances, four atmospheric correction processors have been used.

These are

- MEGS 8.1, the ESA standard processor,

- l2gen version 7.2, the NASA processor,

- c2rcc v0.5 without normalisation, for MODIS and SeaWiFS. c2rcc v0.6 with normalisation and revised flags for MERIS. Inverse neural network.

- POLYMER version 3.4.

The development of a forward neural network has been discontinued. Its place has been taken by the appropriate inverse neural network. A fifth AC processor, siacs (FUBnn), is still under development.

3.1.1 Notes on l2gen processing

- SeaWiFS/l2gen gives only correct flags, if the data is processed from L1a to L2. Otherwise the straylight and HILT flag is wrongly raised in about 90% of all cases.

- MERIS/l2gen uses a parameter file with input parameter *rad_opt* set to 1, which ensures the application of the MERIS smile correction (similar to MEGS).

- MODIS, SeaWiFS and VIIRS are processed to L1c, so that radiometric corrections are applied correctly to the data.

3.1.2 AC processor quality flags

The screening of pixels has been revised in some parts.

Additional cloud flags are available for the four satellite sensors from the Idepix processor, which identifies land, cloud, snow/ice and mixed pixels. For further information, please refer to the Idepix ATBD.

All atmospheric corrections provide sets of quality flags, and in the aggregation process the individual flags and the information from Idepix are combined.

Changes in the AC flags are made for POLYMER, which discards the threshold at 412 nm, but allows for case 2 conditions.

- Idepix cloud mask (version 2.2.13, for all ACs): !F_LAND & !F_CLOUD & !F_SNOW_ICE & !F_MIXED_PIXEL

- Polymer: bitmask == 0 & bitmask == 1024

- c2rcc: v0.5 !(AC_NN_IN_ALIEN | AC_NN_IN_OOR); v0.6 !(Rtosa_OOR | Rtosa_OOS | Rwa_OOR)

- l2gen: !(LAND | CLDICE | SEAICE | HIGLINT | LOWLW | HILT | MAXAERITER | HISOLZEN | HISATZEN | NAVFAIL| ATMWARN | ATMFAIL | STRAYLIGHT)

- MEGS: !(CLOUD | LAND | ICE_HAZE | HIGH_GLINT | PCD_1_13 | PCD_19)

3.1.3 System Vicarious Calibration Gains

POLYMER v3.4 (PR05) (version 20150929)

- SeaWiFS: CALIB 1.0062 1.0007 0.9967 0.9921 1.0 1.0 1.0 1.0

- MODIS: CALIB 1.0128 1.0186 1.0 1.0152 1.0154 1.0153 1.0 1.0 1.0076 1.0 1.0 1.0 1.0 1.0

- MERIS: CALIB 0.9921 0.9949 0.9921 0.9950 0.9938 1.0023 0.9985 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0

- VIIRS: l2gen gains

3.2 Extraction, Aggregation and Filtering

3.2.1 Automated extraction with CalValus

After uploading the bandshifted in-situ data, the production with different AC processors and extraction of satellite data is done automatically on the CalValus system.

The extraction parameters are set to a macro pixel size of three pixels, and a maximum time difference between in-situ and satellite observation of three hours.

Only the first data point of overlapping macro-pixels is considered (overlap filter) and the macro-pixel has to be complete (here three by three pixels).

3.2.2 Aggregation and Filtering methodology

The aggregation process can vary quite strongly in terms of filtering techniques on the macro-pixel and the actual aggregation from the macro-pixel to a single value in the inter-comparison.

The extraction of satellite data consists of the (uneven) N by N pixel, for which the central pixel is closest to the in-situ location. If for a single satellite scene several in-situ measurements at the same location are available, only the one closest in time is considered. The overflight and the in-situ measurement have to occur within a ± 3 hours time difference.

A match-up point of good quality is created and selected by the following method:

- To each pixel in the macro-pixel, the defined AC flags are applied and only valid pixels are considered in the next steps.

- An outlier filter in form of a *standard deviation (σ) filter* is applied to the remaining pixels per wavelength in the macro-pixel. If the pixels are within $\mu_\lambda - f \cdot \sigma_\lambda \leq Rrs_n(\lambda) \leq \mu_\lambda + f \cdot \sigma_\lambda$, with factor $f = 1.5$ as default, they are kept as valid. This filter is applied to each wavelength independently, as noise can be wavelength dependent.

- If the number of valid pixels remains larger than half of the size of the macro-pixel, $N_{valid} > N^2/2$,

- the mean μ and standard deviation σ of the remaining pixels (per wavelength) is calculated and used to *test for spatial homogeneity*. If $\sigma_\lambda / \mu_\lambda < 0.15$, the macro-pixel is considered to be spatially homogeneous and the mean of the remaining valid pixels is a good representative of the entire macro-pixel measurement.

Due to the movement of the water body it seems necessary to account for the time difference between in-situ measurement and satellite overflight with the averaging of the homogeneous macro-pixel. But it is also possible to use the central pixel only. In this case, the instrument noise (or processor noise) becomes a larger issue, which can be more easily neglected in an averaging approach.

Even after aggregation further filtering might be considered, although this is a purely cosmetic step, which takes the sensitivity of the statistical parameters on outliers and the demand for gaussian distributed values into account, but is not driven by oceanographic or physical arguments.

3.3 IBQ, CBQ and case 1 definition

3.3.1 CBQ cautionary note

Although the CBQ (Common Best Quality) flags are the combination of all processor flags and for each processor the same pixels are chosen, the aggregated data points are not necessarily selected as good match-up points for all processors.

- During filtering, outliers may or may not exist for a certain processor, so that the amount of valid pixels can decrease below the limit for one processor but not for another due to processor specific noise in the data.

- The homogeneity criterion can lead to a different amount of available good match-up points, which is again dependent on the noise.

There is no filter in place, which selects only those aggregated spectral points, which are available for all processors.

3.3.2 CBQ and the case 1 selection

In Round Robin 2, the water type selection has been based on the satellite data.

In order to avoid differences in appointed water types, the classification is based on the in-situ data spectrum in Round Robin 3. Either during band-shift or by analysing the spectrum, the water type is assigned to the in-situ data and is independent from the AC retrieval.

3.4 Further investigation

- For inhomogeneous case 2 waters, the aggregation procedure could use a single centre pixel instead of the mean of the (valid) macro pixel.
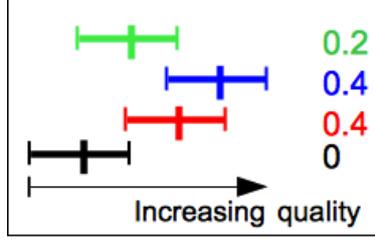
Figure 4.1: Scoring scheme relating points to the best instance with the highest quality (blue). Numbers are the appointed scores after normalisation.

4 Statistics and Scoring

4.1 Statistical parameters

Unlike in the first round robin where several more statistical parameters have been included in the selection process, their number has been reduced to four now. The remaining parameters are:

- RMSE (abs.)
- Bias (abs.)
- residual error (abs.)
- χ^2

The other parameters (relative RMSE, correlation coefficient, slope and intercept of linear regression, number of valid points, the number of spectra with a χ^2 lower than the 95% confidence level) are in some parts redundant. For MERIS seven wavelengths are considered, for MODIS and SeaWiFS six wavelengths are part of the study, for VIIRS five. As some wavelengths are very rarely measured in-situ, the χ^2 value is based on five wavelengths for all the sensors.

4.2 Scoring scheme

Most of the statistical properties come with a standard error or 95% confidence interval. All properties are transformed to negative oriented values, if necessary. To each property the evaluation scores are assigned by wavelength separately in the following manner (Fig. 4.1):

- The best algorithm is the one with the smallest value in the statistical property and receives 2 points.
- If the value corresponding to another algorithm falls within the confidence interval of the best, this algorithm is not significantly different from the best and receives 2 points as well.
- If the value of another algorithm lies outside the confidence interval of the best but their confidence intervals overlap, this algorithm receives 1 point.
- If the confidence interval of an algorithm does not overlap with the best algorithm, this algorithm receives 0 points.
- In order to weigh each wavelength equally the scores will be normalised, so that the sum of all points per wavelength and property over all algorithms equals 1.

All scores S are then summed up per wavelength and statistical property, which gives each of them equal weight. The measure of spectral shape, i.e. the mean χ^2 value, receives the same weight as a single waveband. The score is therefore multiplied by three (because there are three statistical parameters considered per wavelength), when added up to a total score.

$$S_{\text{total}}\left(\text{Processor}\right) \quad = \quad \sum_{i=1}^{7} S_{\text{RMSE.abs}}\left(\lambda_i\right) + S_{\text{res.error.abs}}\left(\lambda_i\right) + S_{\text{bias}}\left(\lambda_i\right) + 3 \cdot S_{\chi^2} \tag{8}$$

In favouring the best algorithm strongly, this scoring system tends to a non-linear behaviour. This approach has been preferred over one of relative scores, which considers all relationships to fixed limits per statistical parameters. It would have been necessary to define in absolutes of what is supposed to be a "good result". The choice would have been highly subjective. Another scoring approach which is applied in the comparison of in-water algorithms (Brewin et al. [2012]) uses the larger number of algorithms to its benefit. From over ten different retrieval algorithms for water constituents the mean of the statistical value under study and the confidence interval are calculated. All algorithms which perform within the confidence interval get the same score, while algorithms outside the interval receive a score of zero. With only four algorithms available this approach cannot be applied to the atmospheric correction comparison.
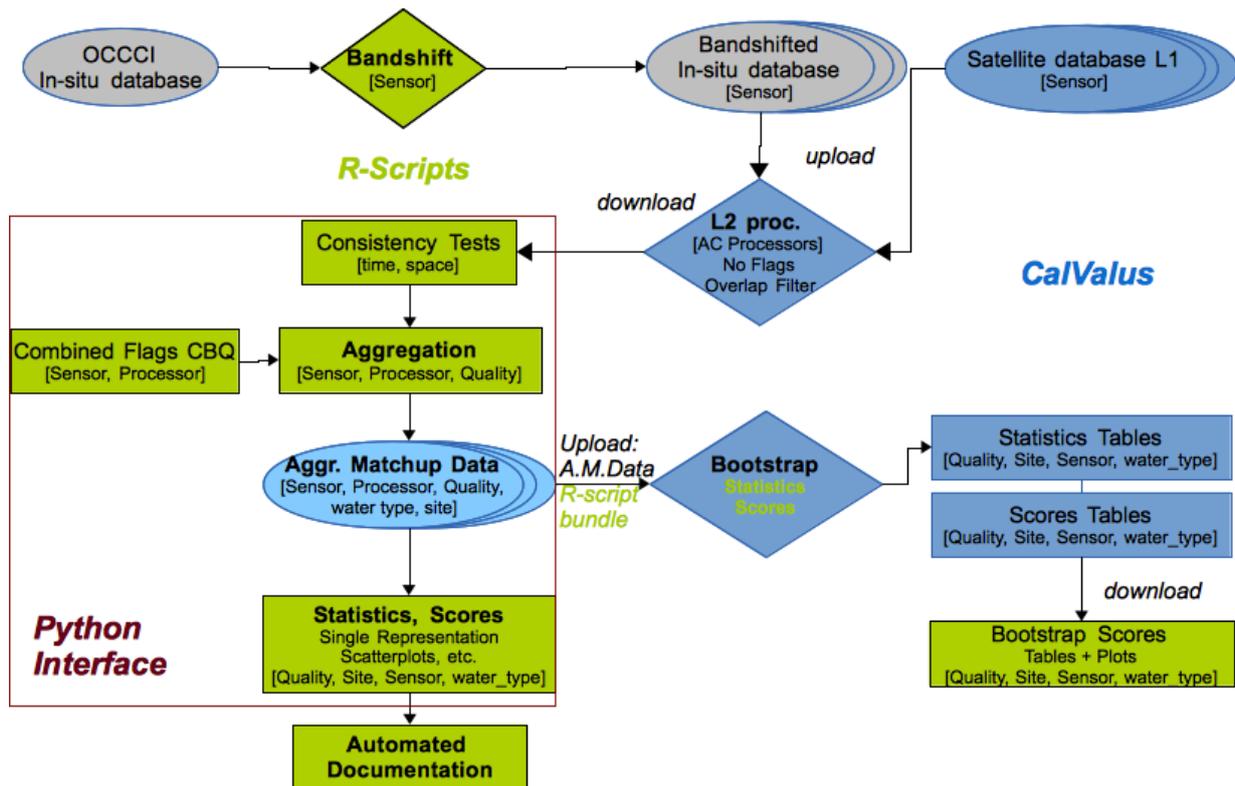
Figure 5.1: Round Robin software overview. Some software parts are implemented in R (green), others on the CalValus system (dark blue).

5 Round Robin automated tools

5.1 Overview

Several steps of the round robin procedure have been developed and implemented in different software parts (Fig. 5.1).

The user needs to perform the band-shift of the in-situ database and needs to apply an R-Script for this task.

The band-shifted in-situ database (including columns with latitude, longitude and time information, and a unique numbering of the data points for sorting purposes afterwards) has to be uploaded to the CalValus system. With the help of the match-up processor, the level 2 products are calculated based on a level 1 satellite database. Although there is a capability for processor dependent flags and aggregation in place on CalValus, we decided to use the overlap filter only, which eliminates duplicates in time for a single place and a single day. Only the in-situ measurement closest in time to the overflight is considered, which can be different for each satellite sensor.

The resulting tables with the L2 products are downloaded. It is necessary to change a single character in all POLYMER processed data files. The POLYMER processor includes a product "R'865", but the following procedures in R are not able to read in the apostrophe and would crash. Using the Python interface, the parameter files for the aggregation and statistics procedures can be easily set and run.

If changes to the statistics and scoring scheme have been made, it is necessary to upload a new software bundle to CalValus as well, which consists of exactly the same procedures as they have been used to analyse the single representation(s). Bootstrapping the statistics and the scores needs the aggregated match-up data to be uploaded to the CalValus system again. The statistical analysis results in an automatically generated documentation including scatterplots and tables of statistics and scores. The bootstrapping results need to be plotted with a separate script and are not yet automatically included in the documentation.

5.2 Round Robin Aggregation and Statistics Interface

Primarily the Python GUI for the aggregation and the statistical analysis is an interface, which allows the user to write R-script files with selected parameters (further on "parameter files") and invoke the R-scripts for the calculations.

The following sections give short summaries of the functionality of this analysis tool.

5.2.1 Data Selection (Fig. 5.2)

Sensor Four satellite sensors are implemented at the moment. If the selection is changed, the user should also press the "Initialise" button, which will then change the default settings of processor type, processor names, flags and variable names to the satellite specific default values.

Initialise button Pressing the "Initialise" button will change the default settings of processor types, processor names, flags and variable names to the selected satellite specific default values.

Processor - Extractions The default value "/path/" has to be replaced by the actual path to the appropriate CalValus extraction file, which contains the not aggregated 3x3 pixel data. The file can be selected by pressing the button and navigating with the file manager to the location. If the default value is not changed, the script assumes, that no data exists.

Processor type Select the processor type (POLYMER, MEGS, ForwardNN, l2gen, FUBnn). After changing the processor type the Processor name (processor specification), the flags and the water leaving reflectance variable names can be set to the specific default values by pressing the "Initialise" button. Default values will be set in all processor instances - if you have already made changes to one, they will be lost.

Processor spec Give a unique version or specification number to the processor. The name will be used throughout the entire analysis and appear in diagnostic plots. The user should choose a name without blanks and dots.

Flags The text field holds the flag expression in R syntax, which will be used to identify valid pixels.

Variables give the names of the products which are aggregated in coming calculations and are therefore part of the validation exercise. The default values are the normalised water leaving reflectances of the different atmospheric correction processors.

5.2.2 Data sorting (Fig. 5.3)

The extracted datasets can slightly differ in the order of extracted points. As the aggregated data will be combined in a single file, so that selecting subsets for the bootstrapping becomes easy, it is necessary, that the extractions are in the same order. Sorting can be quite time consuming and it is well advised to include a column of continuous numbering in the band-shifted in-situ dataset, so that the same measurements can be identified easily.

If the data can be sorted by such an identifier, the user can set the column name in the text field.

5.2.3 Aggregation parameter setup (Fig. 5.4)

The aggregated data is described by the following parameters:

Output folder can be selected from the file manager. All output files, which are the combined (and sorted) aggregated match-up data points from all AC processors, are written into this folder.

In-situ Variables holds the names of the columns with the in-situ data in the table. The default names are just influenced by the choice of the satellite sensor. The user has to make sure, that the number of in-situ variables matches the number of satellite products and that the names are given in the appropriate order. The R script does not check, whether names and content are consistent.

Column name In-situ sites

Column name In-situ time

Figure 5.2: Round Robin Python GUI Part 1- data selection. Default settings.

Figure 5.3: Round Robin Python GUI Part 2 - Sorting and addition of Idepix flags setup. Default settings.

Column name satellite time

N pixel gives the length of a side of the extracted square of pixels.

Filtered mean During aggregation the valid pixels can be filtered. If filtered mean is chosen, the values at pixels have to be within an interval of ± 1.5 standard deviation around the mean. The factor can be adjusted in the text field.

Spatial homogeneity check can be applied or not. If it is used, the number in the text field gives the threshold of the ratio of standard deviation to mean of the valid pixels.

Aggregation Sites The default settings are different for each satellite sensor. The combinations of sites starts with the global data set, which comprises all data points. Sites are excluded, which have been used for vicarious calibration. The combinations have been simplified to Global and Global-MOBY for all sensors.

Water type Three water type classifications are included in the script. "All" water types does not discriminate optical water types, whereas "strictCase1" selects only case 1 waters. The "case2" type chooses everything, that is not case 1. The definition of case 1 is selected in the following parameter

Definition Case1 The definition of case 1 can either be a very simplistic threshold, or the application of the definition by Lee either on the satellite data or on the in-situ data (see section 2.2). A third option, identification based on in-situ data, is recommended.

Quality distinguishes selections of individual or common best quality (IBQ, CBQ).

Filename addition If different versions of the aggregation are calculated, it might be helpful to add an identifier to the filenames.

Aggregation Parameters holds path and filename of the file with all the chosen parameters of this task. If an already existing file is chosen, it will be overwritten by pressing the "set aggr. parameter" button. If a parameter file already exists, it can be chosen with "Select". The entire steps for setting up the parameter file can be omitted, if such a parameter file exists and is to be used for the calculations. The user needs to be careful, as the entire text fields are not updated to the content of the parameter file and therefore may show quite different parameters than the ones which have actually been used for the calculations.

Run Aggregation Pressing this button starts the R script for the Aggregation and uses the specified parameter file.

There exists a short cut to run the R scripts: just select an existing parameter file and run the aggregation! Sometimes copying and adjusting the parameter file might be preferred to setting it up anew with the interface, especially if only small changes in parameters are made.

5.2.4 Statistics and Scores (Fig. 5.5)

Sensor is selected together with the one in data selection. As soon as the sensor is changed, the default names of the in-situ data is up-dated (both in this section and in the aggregation section).

Input folder is automatically changed to the aggregation output folder as soon as it has been selected.

File name addition is the same as in the aggregation section.

Output folder has to be selected.

In-situ Variables

In-situ Scale factor If the in-situ values are e.g. remote sensing reflectances instead of water leaving reflectances, they can be scaled by the giving multiplicative factor.

Outlier filter In the match-up data there may still exist significant outliers, which can influence the sensitive statistical parameters.

Parameters The statistical parameters, which appear in the table of the documentation, can be chosen here.

Figure 5.4: Round Robin Python GUI Part 3 - aggregation parameter setup. Default settings.

Score Parameters These statistical parameters are used for the scoring. (The script does not check whether the statistical parameters have been calculated. The user has to take care that the Score Parameters are a subset of the statistical parameters.)

Common Scale Site The scatterplots use a common scale per wavelength. The given site chooses the file on which the common scale will be based.

Aggregation Sites are the same as in the aggregation parameters.

Water Type is the same as in the aggregation parameters.

Quality is the same as in the aggregation parameters.

Scatterplots may or may not be created.

Spectral Statistic Plots may or may not be created.

Score Type can either be "AC", which corresponds to the method described above (see 4.2), or it can be "multiple".

Parameter File needs to be selected or created by choosing a path and filename and then use the "Set Statistics" button. This action will write a text file including all the information specified in the Statistics Parameter Setup section. Existing files will be overwritten without warning. As a shortcut it is possible to select an existing parameter file and calculate the statistics. Again, choosing an existing parameter file will not change the entries in text fields or selections!

Calculate Statistics Pressing this button will run the statistics script with the given parameter file (not with input from the interface itself). If changes in the interface text fields are made, a new parameter file has to be written.

Quit the python program.

Figure 5.5: Round Robin Python GUI Part 4- statistics parameter setup. Default settings.

6 Round Robin Results - Summary and Discussion

6.1 Match-up extraction with the CalValus System

In a first step, the band-shifted in-situ database is uploaded to the CalValus system. With the help of the match-up tool, the corresponding satellite data is identified and processed from level 1 to level 2. All parameters are given in section 6.2.

It is imperative for the SeaWifS processing with l2gen to calculate Level 2 products directly from level1a (NASA standard procedure). Otherwise you have to consider that a two-step approach from level 1a to level 1b to level 2 might lead to applying the vicarious calibration gains twice!

6.2 Set-up summary

In-situ database: OC_CCI v3.0 (section 1), normalisation following Park and Ruddick mostly, and f/Q Morel for case 1 observations in NOMAD dataset.

Band-shift: band-shifted iteratively (water type based IOPs, modelled with two closest wavelengths, weighted with inverse distance), $\Delta\lambda < 10nm$ (only closest wavelength), convergence limit $\epsilon = 10^{-9}$ (section 2)

Match-up protocol: ± 3h in-situ measurement and overflight, extract of full 3x3 pixels, centre closest to in-situ position, no overlap with other extractions.

Filtered mean: $\mu_\lambda - f \cdot \sigma_\lambda \leq Rrs_n(\lambda) \leq \mu_\lambda + f \cdot \sigma_\lambda$, with $f = 1.5$

Homogeneity criterion: $\sigma_\lambda/\mu_\lambda < 0.15$ and $N_{valid} > N^2/2$, with $N = 3$

Flag definition: see 3.1.2. Idepix cloud mask v2.2.13 is added to each processor.

Outlier filter: after aggregation is NOT applied.

Case 1 definition: Lee and Hu [2006]. (section 2.2) Selection is based on the in-situ data.

Case 2 definition: case 2 waters are defined as not being case 1 waters. (section 2.2)

Quality: Individual best (IBQ) and common best (CBQ)

Data selection: Sites which are used for vicarious calibration can be excluded from the in-situ database.

- Global: all sites, entire OC_CCI in-situ database
- Global-MOBY: all sites, but without measurements from MOBY.

Statistics and Scores: (section 4.1 and 4.2)

- RMSE
- Bias
- residual error
- χ^2

6.3 Results of statistics and scores

6.3.1 MERIS

Statistics The statistics for *IBQ* Global with all match-up data or restricted to case 1 data are nearly the same (figure 6.2a and figure 6.2b, left). Reducing the global dataset by the MOBY data raises the RMSE between 510 and 620nm. Throughout the entire spectrum POLYMER has the lowest RMSE and residual error. In the blue, the RMSE of c2rcc results decreases. The RMSE of all processors changes with the reduction of the data, which indicates that the quality of retrieved data in more complex waters is slightly reduced.

The overall bias is quite similar for all processors for the green to red bands, but in the blue POLYMER has a positive bias opposite to the negative bias in MEGS, l2gen and c2rcc. The absolute value of bias is similar

for all these three processors in the global dataset, the bias of POLYMER in the blue is smaller. If MOBY is excluded from the data, the biases of POLYMER, MEGS and l2gen increases in the blue, whereas the bias of c2rcc decreases. The bias of l2gen and MEGS also increases significantly at 620nm.

If the in-situ database is restricted to the case 2 waters, the bias is reduced significantly for c2rcc in the blue, while all other processors retrieve data with larger biases. The RMSE for MEGS results increases in the blue, but the strong error at 620nm for l2gen and MEGS gets reduced (figure 6.1c).

Although the general shapes of the spectral errors based on CBQ data remains the same compared to IBQ, the differences between processors become less pronounced (figure 6.2). POLYMER outperforms the other processors clearly over the entire spectrum with respect to bias and RMSE for Global, Global-MOBY, for the combination of water types and case 1 waters (figure 6.2b, figure 6.2a).

For case 2 waters, products of c2rcc have the smallest bias, but the RMSE is comparable for all processors, except in the blue. c2rcc performs better here.

Scores   The scores are based on the three statistical parameters RMSE, bias and residual error, which has been discussed in the previous section. In addition the χ^2 value of the spectra is evaluated as well.

For IBQ and CBQ with combined water types or the case 1 selection and in- or excluding MOBY data, POLYMER clearly performs best (figure 6.3).

In the CBQ/case 2 selection performances over the spectrum are more similar, which leads to the overlapping of the score distributions. The c2rcc results are slightly better. For the IBQ/case 2 data, all processors (except for MEGS) give a similar quality in the results.

At this stage, the case 2 capabilities of some processors are still under development. With the quite diverse behaviour of comparisons at different sites (in the study of system vicarious calibration), it has been decided to base any algorithm selection on the case 1 water selection.

Therefore, POLYMER is the preferred choice for the MERIS processing.
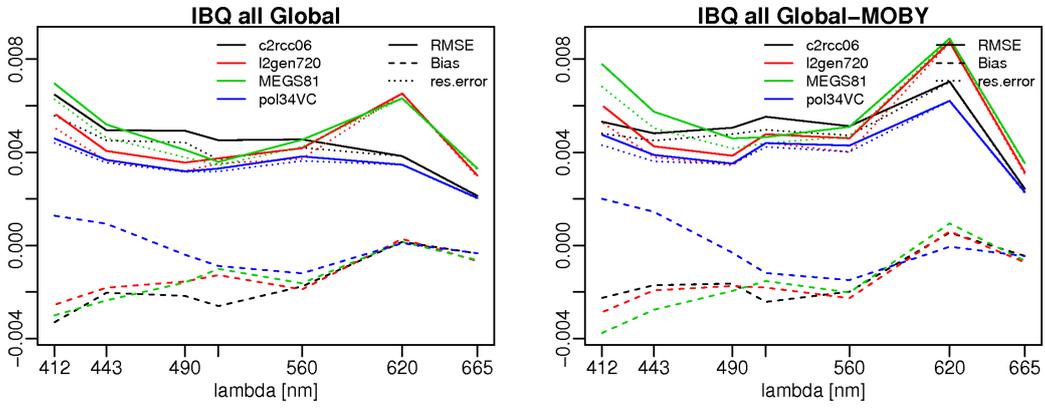
(a) All water types.
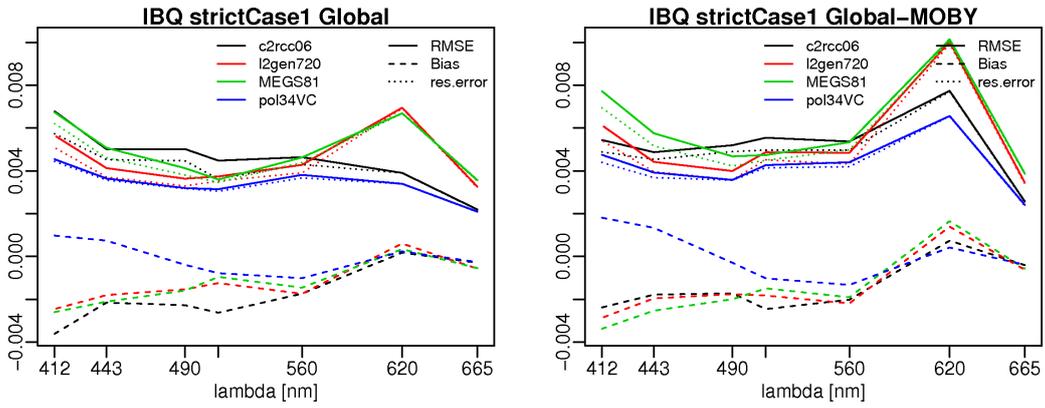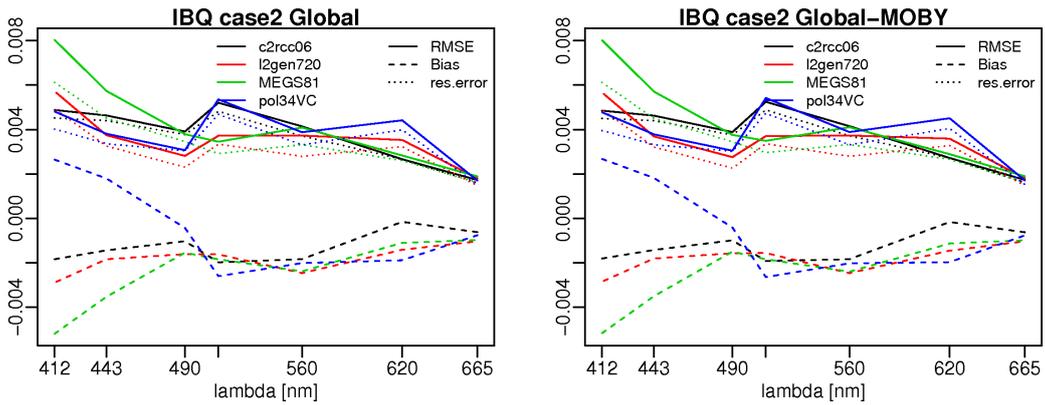
(b) Case 1.

(c) Case 2.

Figure 6.1: Spectral statistics IBQ. MERIS
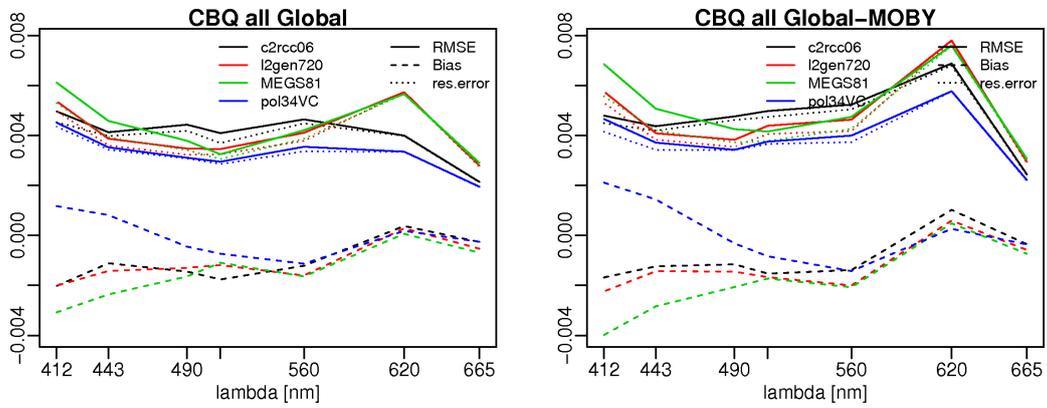
(a) All water types.
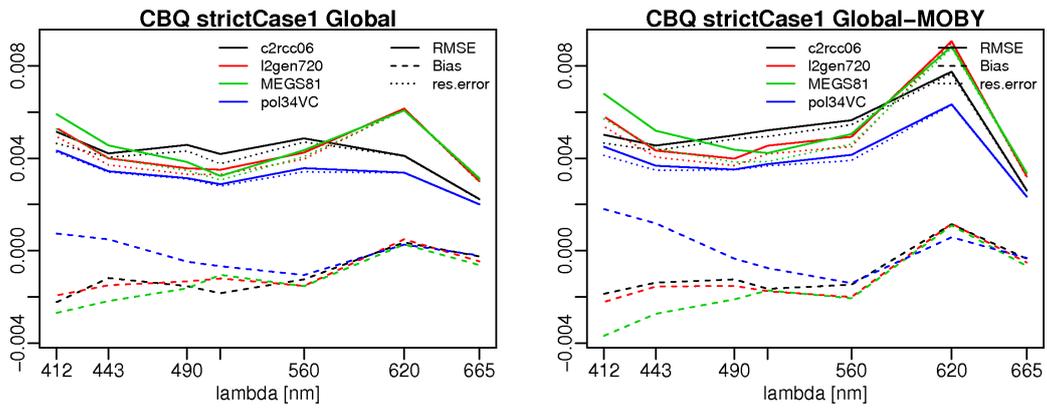
(b) Case 1.
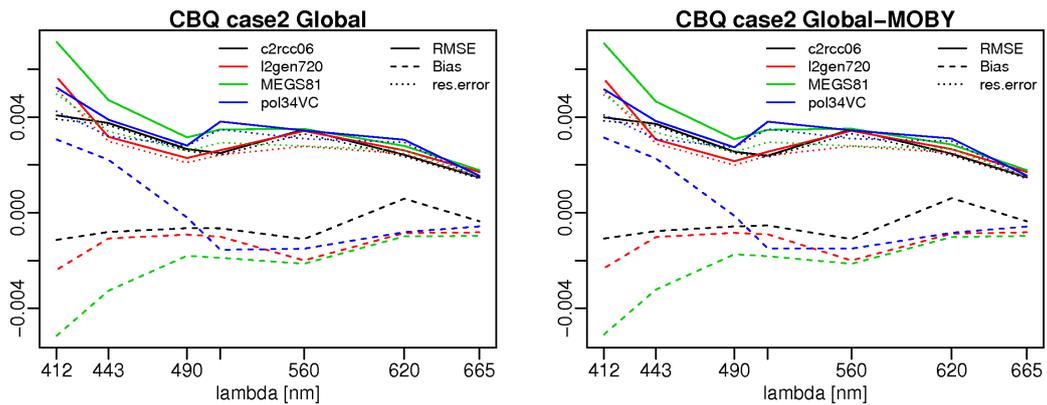
(c) Case 2.

Figure 6.2: Spectral statistics CBQ. MERIS
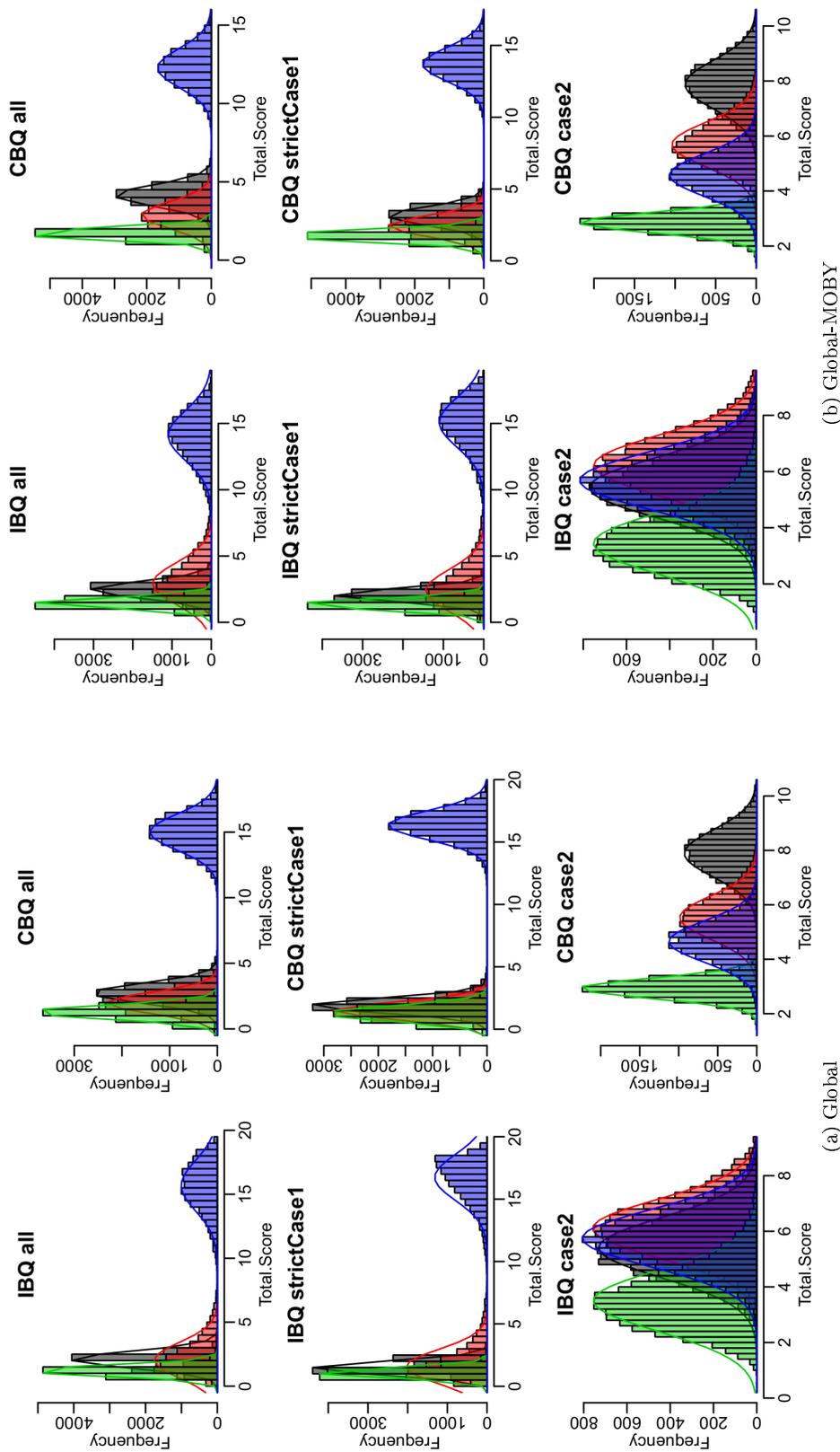
Figure 6.3: Bootstrapping MERIS. Colours refer to POLYMER (blue), MEGS (green), l2gen (red) and c2rcc (grey).

(a) Global

(b) Global-MOBY

6.3.2 MODIS

Statistics Based on the IBQ/all selection, RMSE of c2rcc results is always the largest. RMSE for l2gen and POLYMER are very similar. Differences can be seen in the spectral biases. While c2rcc has the largest bias in the blue, the bias is nearly zero from 488nm onwards. l2gen has a quite constant, but overall the largest bias over the entire spectrum. (figure 6.4a). If MOBY data is excluded, c2rcc performs better at 412nm in terms of RMSE. At 531nm RMSE raises for POLYMER and l2gen, which remain comparable. The bias of c2rcc gets lower at 412 but higher between 488 and 547nm. At 531nm all absolute values of biases increase.

For the IBQ/case 1 selection (figure 6.4b), RMSE drop to lower values for all processors, except for an increase of c2rcc results at 412nm. l2gen and POLYMER are quite similar in RMSE over the entire spectrum, with a slight advantage for POLYMER. The absolute spectral bias of POLYMER is the lowest, while the bias of c2rcc is comparable to l2gen except for a large deviation at 412nm. If MOBY data is removed, l2gen and POLYMER are very similar in RMSE, but c2rcc and POLYMER share the lowest spectral bias (except at 412nm).

For case 2 waters, the results are very similar to the global dataset without MOBY (figure 6.4c).

All CBQ/all and CBQ/case 1 selections (figure 6.5b and figure 6.5a) the spectral RMSE of POLYMER is the lowest (with small exceptions at 412nm). The bias of POLYMER is the lowest especially in case 1 waters excluding MOBY data. In that case the absolute bias of c2rcc increases strongly and the l2gen bias slightly.

Again, case 2 results are similar to those, which are based on the global dataset excluding MOBY data (figure 6.5c).

Scores For the IBQ/all selection the scores are similar for l2gen and POLYMER, favouring l2gen (figure 6.6). With the application of the CBQ selection, the order changes to highest scores for POLYMER in all water types. The influence of the common quality flags on the results is particularly strong in all datasets which include measurements from case 2 waters.

If data from MOBY is removed, the order in the scores stays the same. For CBQ the scores for l2gen and c2rcc are almost identical.

The algorithm selection is mainly based on the summary of statistics for case 1 waters.

Whereas the recommendation in RR2 has been to apply l2gen with MODIS data, here *in RR3 POLYMER is chosen for the processing.*
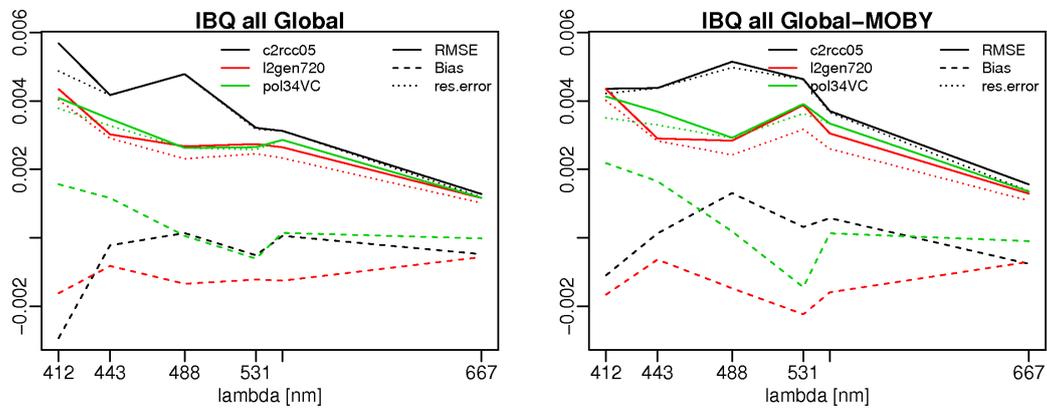
(a) All water types.
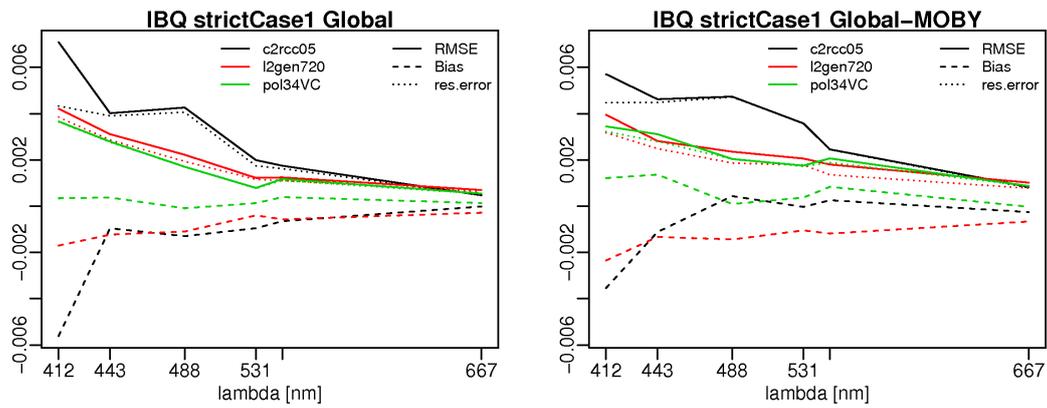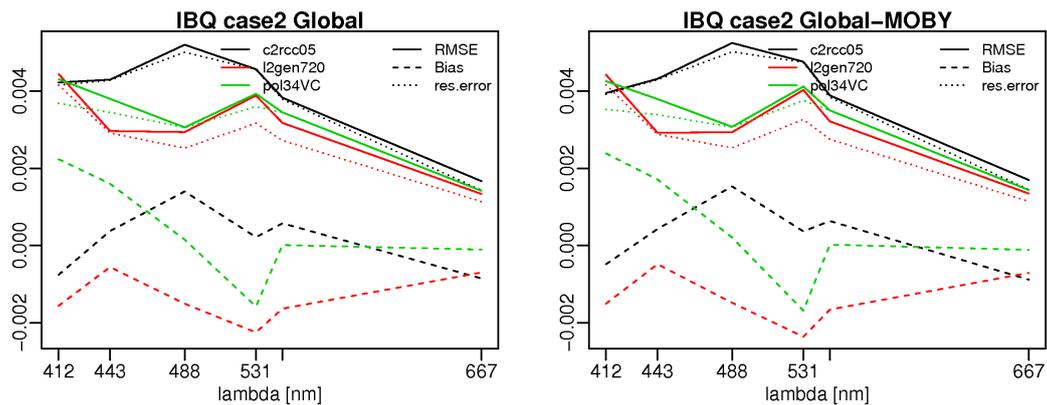
(b) Case 1.

(c) Case 2.

Figure 6.4: Spectral statistics IBQ. MODIS
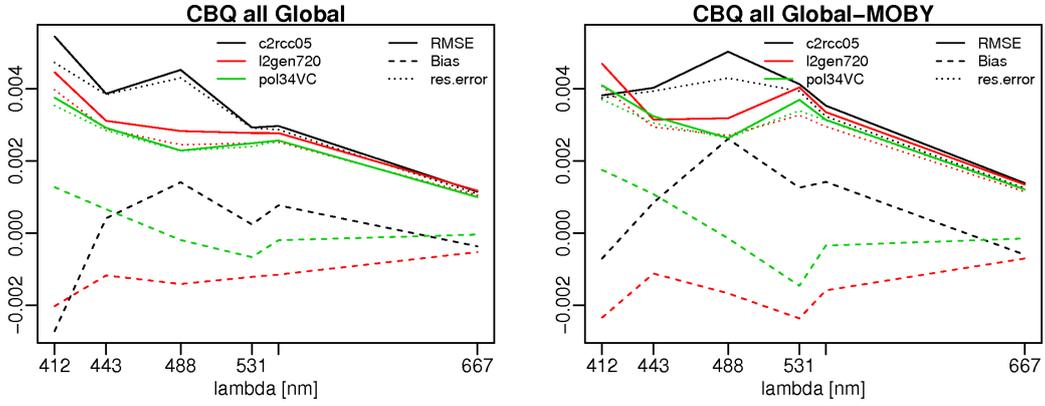
(a) All water types.
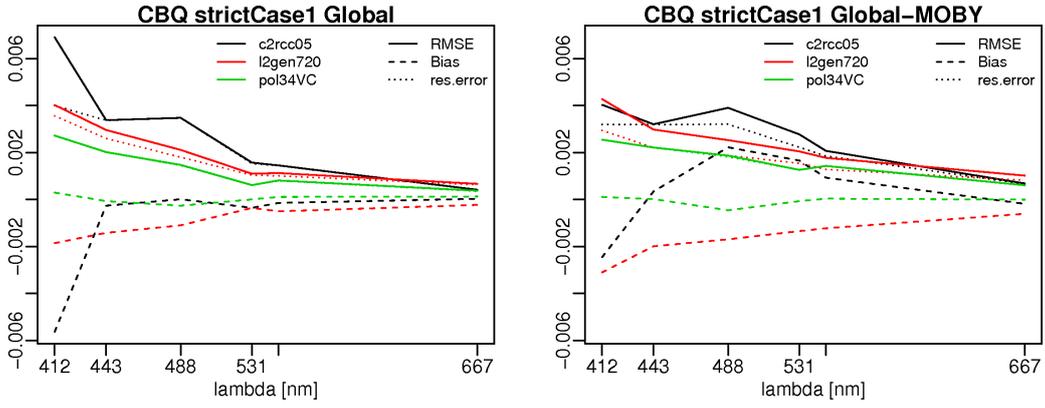
(b) Case 1.
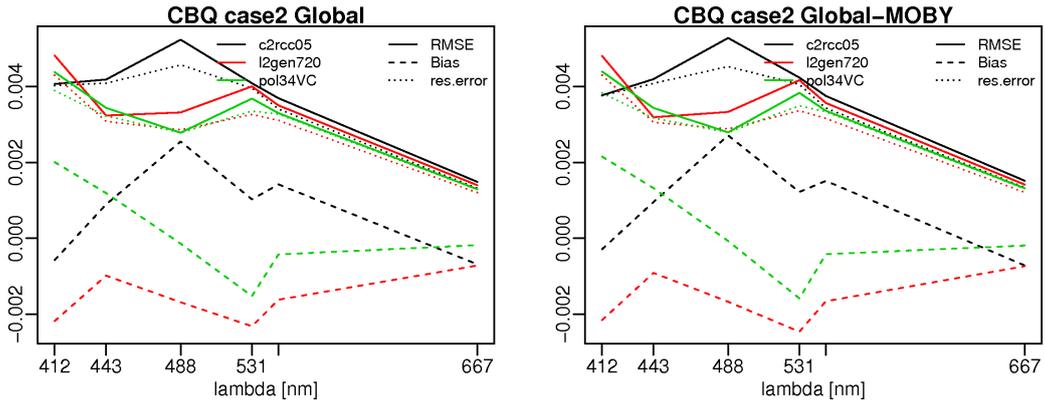
(c) Case 2.

Figure 6.5: Spectral statistics CBQ. MODIS

(a) Global

(b) Global-MOBY
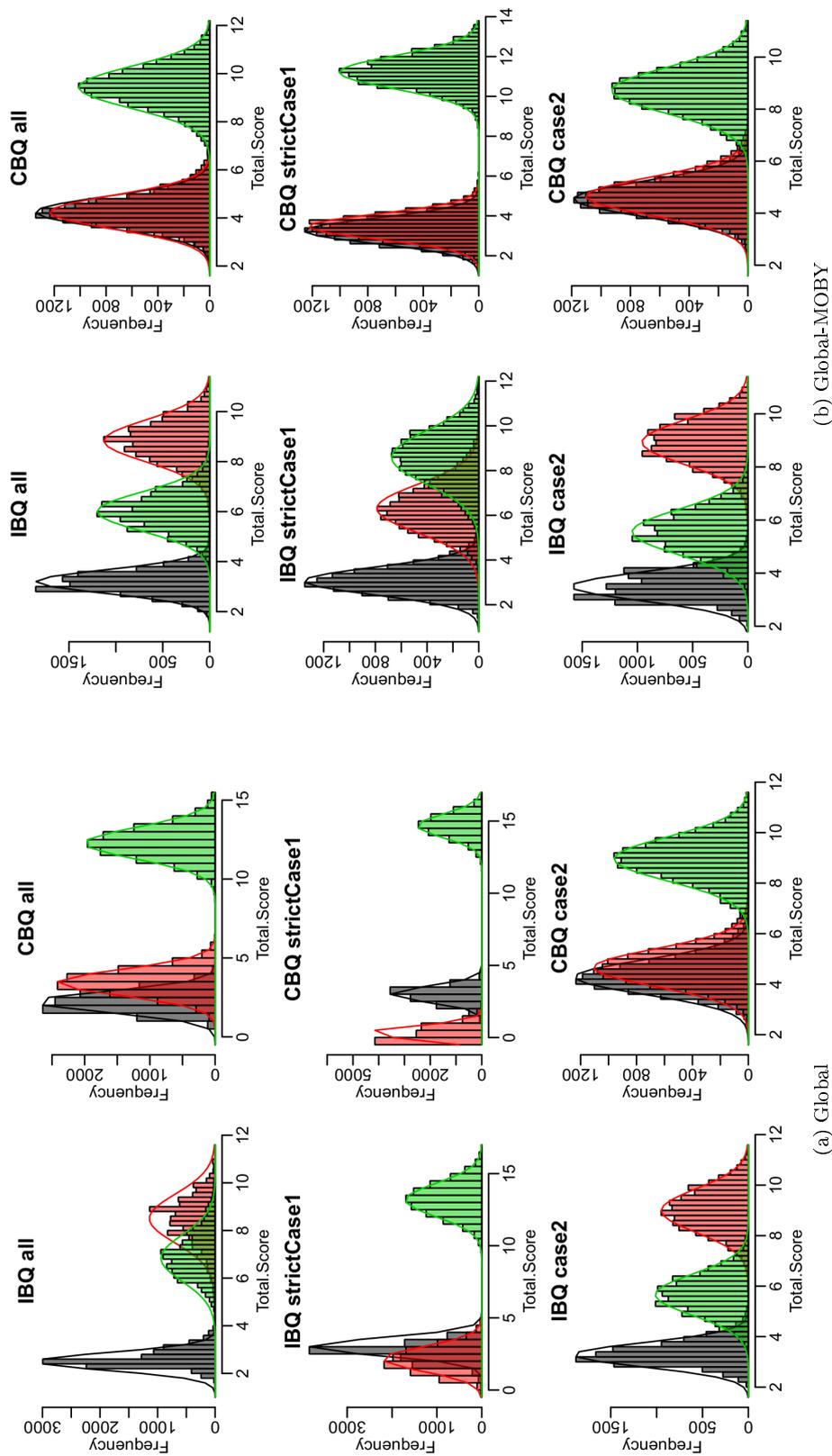
Figure 6.6: Bootstrapping the scores of MODIS match-up statistics. Colours refer to POLYMER (green), l2gen (red) and c2rcc (grey)

6.3.3 SeaWiFS

Statistics The spectral RMSE is smoothest for all AC processors in the IBQ/case 1 selection (figure 6.7, in particular figure 6.7b). Results of POLYMER and l2gen show very similar RMSE in case 1 conditions, but as soon as case 2 waters are included RMSE increases for POLYMER at 412 and 555nm. l2gen has a remarkably low spectral bias for all data selections which include case 1 water samples. In any case, c2rcc is outperformed in RMSE and bias by the other processors. Biases in case 1 waters remain stable for l2gen and POLYMER if MOBY data is excluded, but for c2rcc the bias at 555 and 670nm increases.

The spectral pattern of the statistical values based on the case 2 subset is similar to the subset of all water types without MOBY data (see figure 6.7c and figure 6.7a, right). Compared to case 1 conditions, the RMSE is higher at 490 and 555nm for all processors. The overall bias increases compared to case 1 waters.

Selecting the data with CBQ flags does not change these findings (figure 6.8).

Scores The scores reflect the poorer performance of c2rcc compared to the other AC processors correctly (figure 6.9). In concordance with the analysis of the spectral RMSE and bias, l2gen is clearly the best algorithm for all data sets, which include case 2 waters. If the study is restricted to case 1 conditions, POLYMER and l2gen perform quite similar. The scores are even in favour of POLYMER, if MOBY data is included in the study. Since RR2, the system vicarious calibration for POLYMER/SeaWiFS has been well established. If the algorithm selection is based on the case 1 selection, one would be free to chose between l2gen and POLYMER. Currently, l2gen is chosen as it is the selected algorithm for VIIRS, which leads to equal distribution of processors between the four satellite sensors. Having two POLYMER products (MERIS and MODIS) and two l2gen products (SeaWiFS, VIIRS) might lead to better consistency within the merged product, instead of using POLYMER/SeaWiFS and adding a different processing only at the end of the time series.
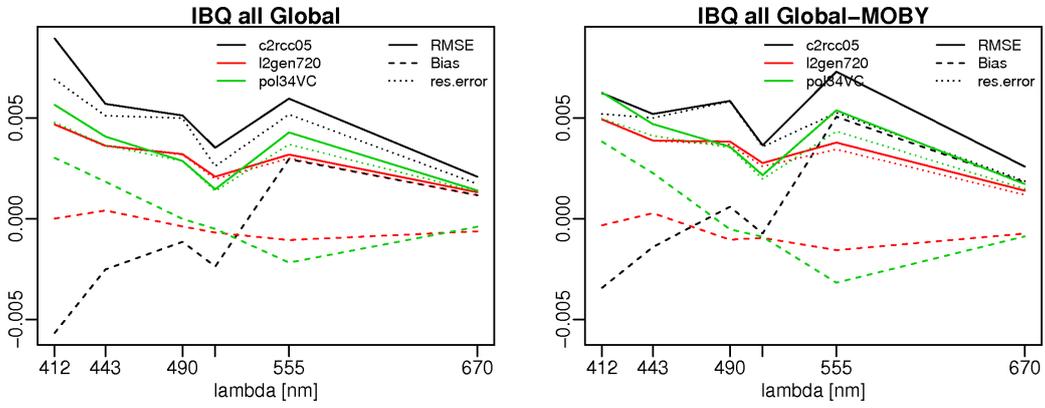
(a) All water types.
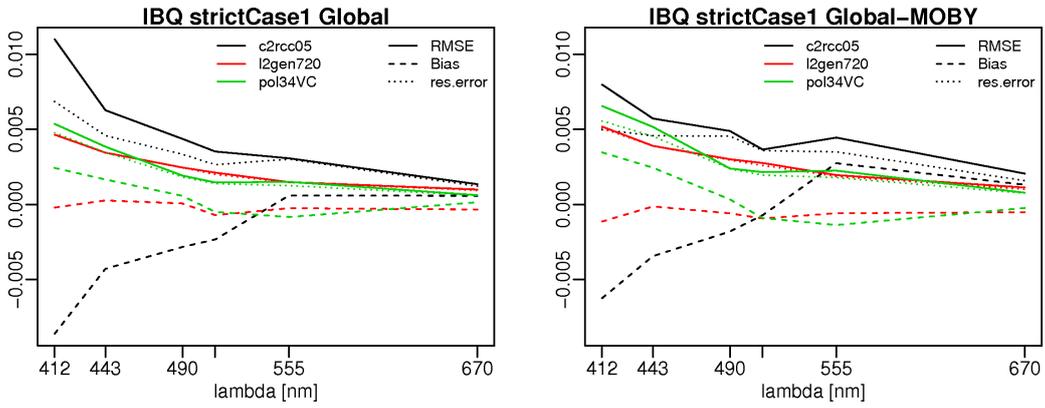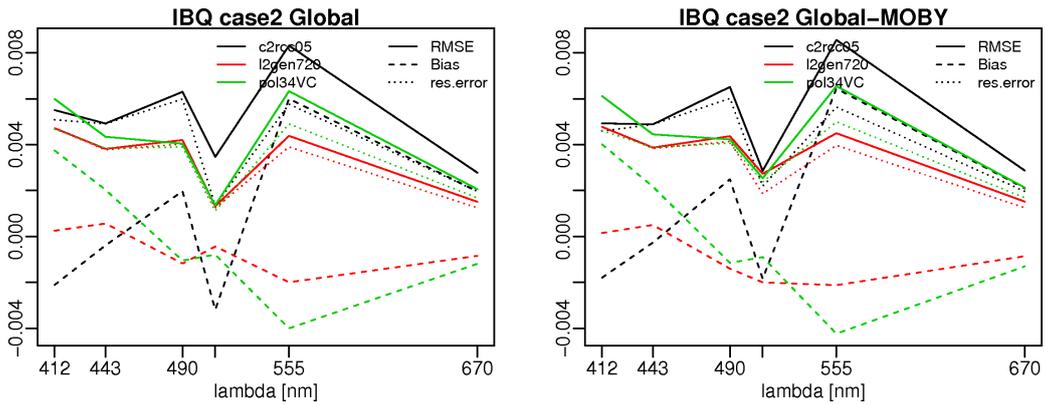
(b) Case 1.

(c) Case 2.

Figure 6.7: Spectral statistics IBQ. SeaWiFS
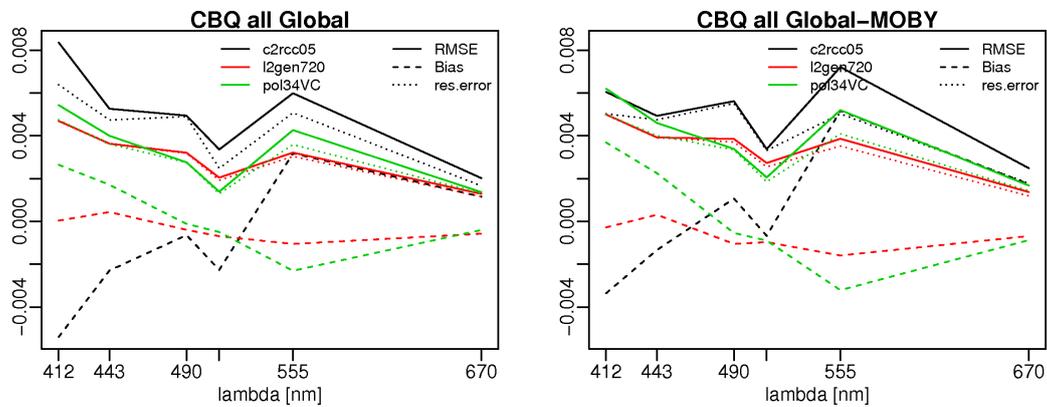
(a) All water types.
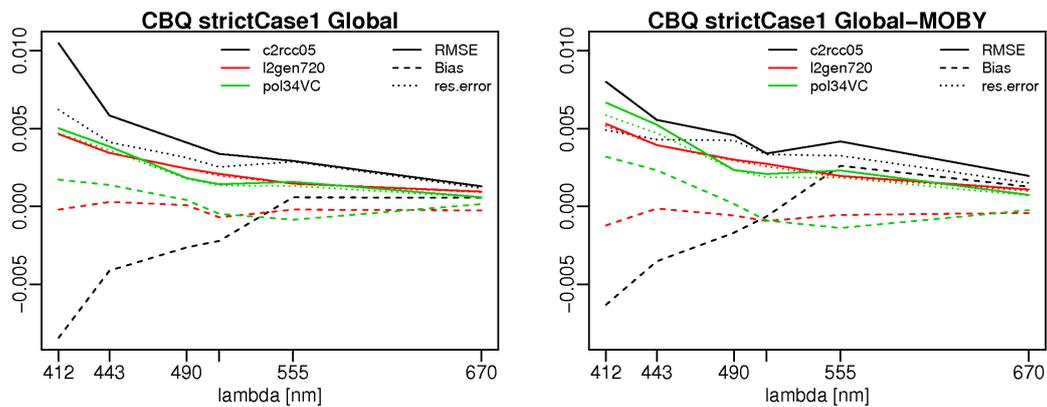
(b) Case 1.
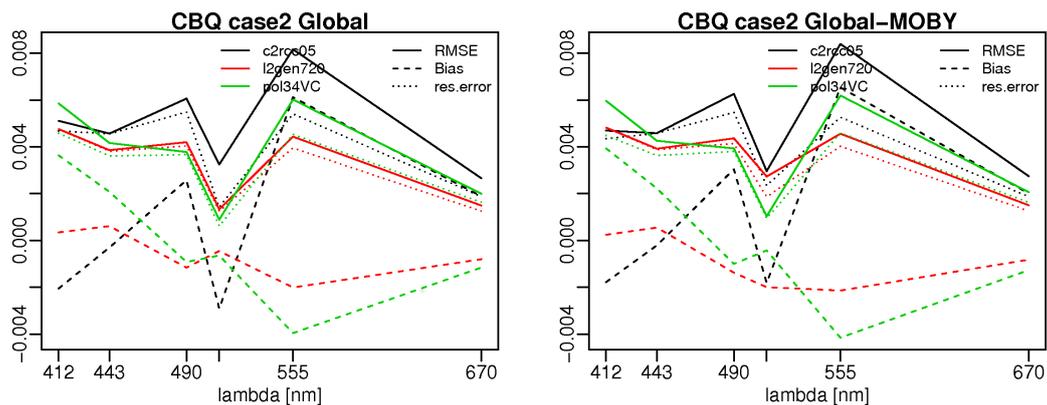
(c) Case 2.

Figure 6.8: Spectral statistics CBQ. SeaWiFS

(a) Global

(b) Global-MOBY
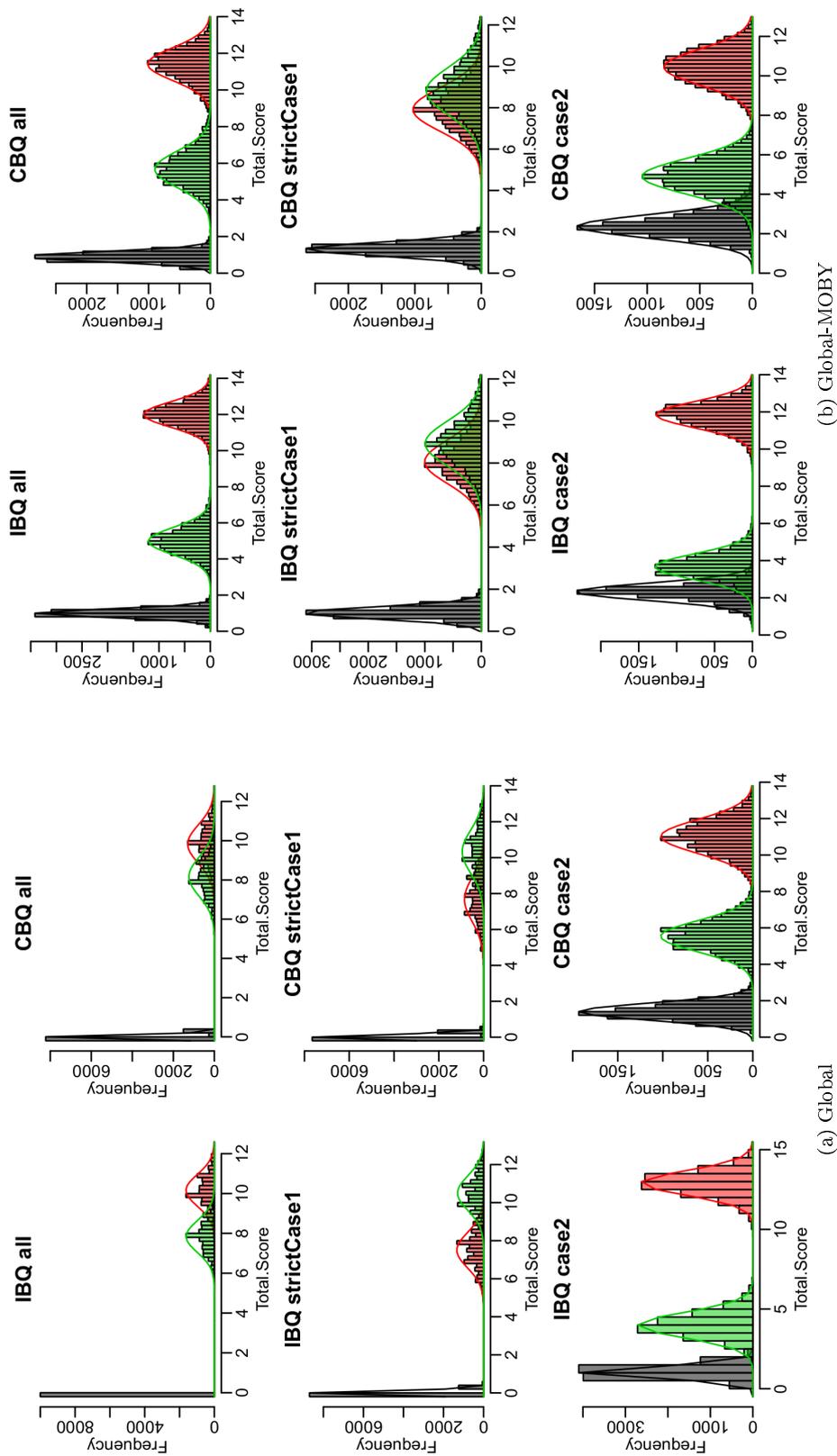
Figure 6.9: Bootstrapping the scores of SeaWiFS match-up statistics. Colours refer to POLYMER (green), l2gen (red) and c2rcc (grey)

7 Round Robin on VIIRS. Preliminary results.

"Like MODIS, the Visible Infrared Imaging Radiometer Suite (VIIRS) is a multi-disciplinary sensor providing data for the ocean, land, aerosol, and cloud research and operational users. VIIRS spectral coverage will allow for data products similar to those from SeaWiFS as well as SST, a standard MODIS product. SST is an Essential Climate Variable (ECV) and, through validation with instruments traceable to NIST standards, is a Climate Data Record. Also, as with SeaWiFS and MODIS, the VIIRS scan and orbit geometries will provide global coverage every two days.

The VIIRS design incorporates a SeaWiFS-like rotating telescope assembly which protects the optical components from on-orbit contamination. This will result in greater on-orbit stability than other designs. VIIRS also has a solar diffuser assembly with a stability monitor similar to MODIS for tracking on-orbit performance in visible wavelengths, and a MODIS-like black body calibration target for the infrared bands.

A two-day coverage is a general requirement for ocean ecology and carbon research because microscopic marine plant (phytoplankton) concentrations are highly variable, particularly in coastal zones. The VIIRS 750 m resolution across the entire scan will provide twice the coverage of MODIS and SeaWiFS, which is a substantial improvement for coastal and estuarine studies in particular. VIIRS also has shortwave infrared bands that can be used for turbid water aerosol corrections."[2]

7.1 Datasets for match-up comparisons

7.1.1 OC_CCI in-situ data version 3.0

As NPP has been launched in October 2011, the amount of possible match-ups with in-situ measurements is limited. The previous intercomparison work has shown that the results can be strongly influenced by the data selection and especially investigations in different water types might become unreliable.

In Round Robin 3 (November 2015) *915* match-ups for VIIRS/l2gen have been extracted, which is a reasonable number to start analysis even with differentiating of the water types.

There are about 1200 POLYMER extracts (IBQ) available, about 900 spectra are classified as case 2 and 300 as case 1.

7.1.2 Comparison with MODIS-A data

In order to create a larger match-up data base, the VIIRS Rrs data is also compared with MODIS-Aqua.

For the time range of January 2012 to December 2013 the granules of MODIS and VIIRS are selected, which fulfil the following criteria:

- maximum distance between nadir lines of satellite overpass is 250 km, the time difference should be less than 15 min and the observed location has between ±60° latitude.

- MODIS 3x3 macro-pixel have to fulfil the criteria of a good match-up point (see **??**).

- l2gen flags: !ATMFAIL !LAND !PRODWARN !HIGLINT !HILT !HISATZEN !COASTZ !STRAYLIGHT !CLDICE !COCCOLITH !HISOLZEN !LOWLW !CHLFAIL !NAVWARN !MAXAERITER !MODGLINT !CHLWARN !ATMWARN !SEAICE !NAVFAIL !FILTER !PRODFAIL

The amount of possible match-ups depends on the spatial and temporal distance (figure 7.1). As both satellites have very close orbits there is a distinct pattern occurring. The colour of the points is defined by the date of the overpass, ranging from blue in 2012 to magenta at the end of 2013. The points locations correspond to the nadir position of the first line of a scene, if this scene includes points, which fulfil the match-up criteria. The first nadir pixel does not have to fulfil these criteria itself.

Starting from this selections of scenes, macro-pixels are chosen from a 10km grid. There are three datasets available, which are positioned within a 300 km radius around Boussole (med), within a 250km radius around MOBY, or selected from the world wide data set by taking 1 macro-pixel out of 200 (figure 7.2).

The MODIS data is bandshifted to VIIRS wavelengths with the water type based IOP models and using only the closest wavelength (see section §2).

Although this "virtual in-situ match-up" data has been prepared, the extractions have not been processed.

---

[2]Introduction from: http://npp.gsfc.nasa.gov/viirs.html

(a) Max. distance 250 km, temporal distance 15 min

(b) Max. distance 100 km, temporal distance 15 min

(c) Max. distance 100 km, temporal distance 5 min
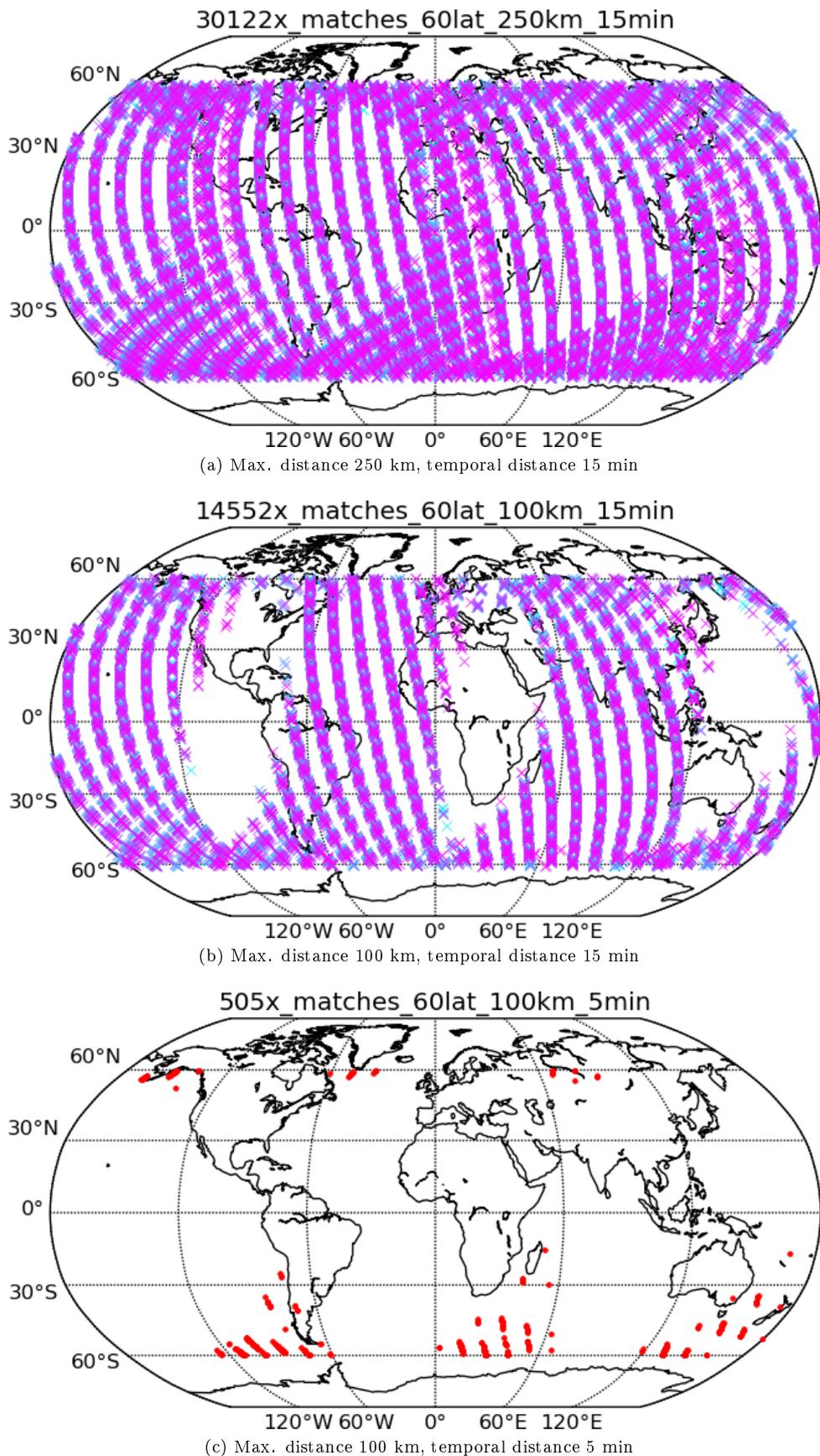
Figure 7.1: Overview of global match-up data of VIIRS and MODIS Aqua within different maximum differences in space and time of the nadir observation. Point positions are the first line nadir position, if match-ups which satisfy the definition are found within the entire scene. Colours change with time from blue (January 2012) to magenta (December 2013).

(a) Global match-up position.

(b) MOBY

(c) BOUSSOLE, med
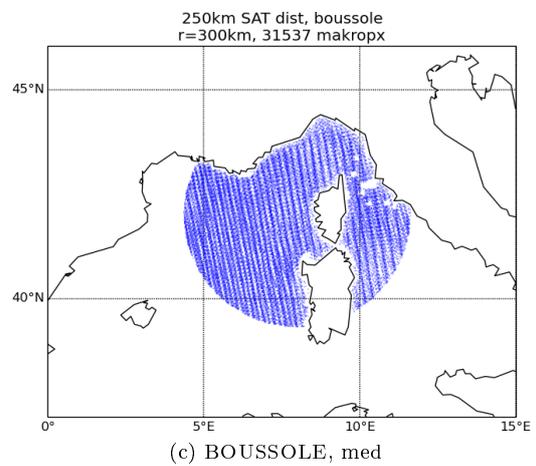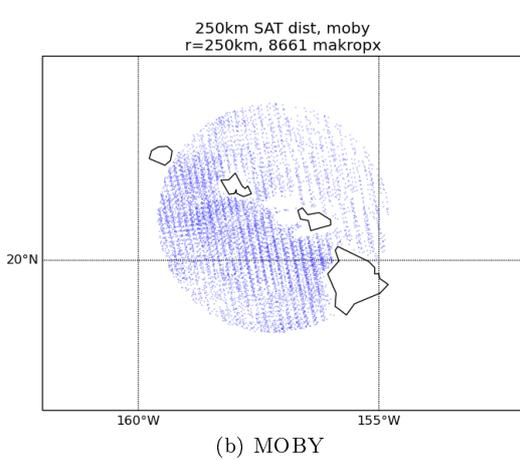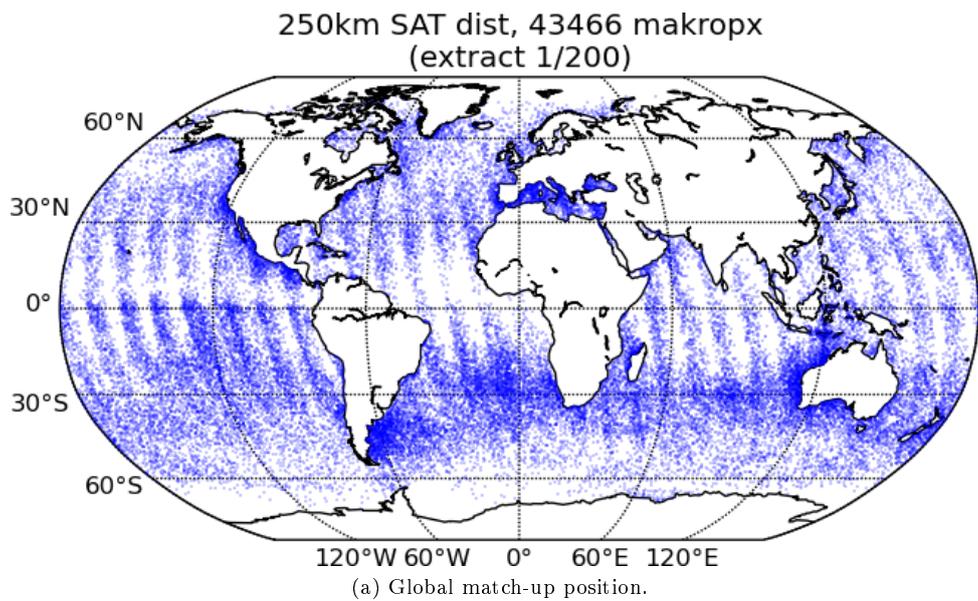
Figure 7.2: Overview of macro-pixel positions, which satisfy the match-up criteria and are chosen as data for the comparison.

7.2 Processing VIIRS

At the time of the Round Robin 3 (November 2015), the system vicarious calibration of VIIRS/POLYMER is still under investigation. The POLYMER processing uses the established l2gen SVC gains instead. Two different versions of the POLYMER processor are tested: the first one is based on radiometric models from Morel & Maritorena (polMM01), while the second one employs Park & Ruddick models (PR05). The POLYMER processing starts from L1c data, which needs to be processed from L1a.

An IDEPIX cloud mask has not been available yet, but it is in preparation.

Note   Starting from L1c introduces an error in the time stamp of the satellite pixel of +1 hour. Therefore in the time range of ±3hours not all extracts have been found for the POLYMER/VIIRS processing. This leads to a smaller amount of spectra in the selection of "common best quality" data, as were actually present in the OC_CCIv3.0 dataset. If the l2gen extracts are analysed on their own, 915 match-ups are found. By sorting and combining the data with the POLYMER extracts, only 347 match-ups of l2gen remain. It can be expected, that by correcting this time error, the amount of CBQ spectra would (at least) double.

7.3 VIIRS AC intercomparison

Statistics   If the dataset comprises all water types and all sites (IBQ all Global, figure 7.3a), l2gen has the lowest spectral RMSE followed by polMM01 and polPR05. On the other hand, considering the bias polPR05 and l2gen give the best results, while polMM01 has larger biases. If MOBY is excluded, the RMSE of l2gen products is almost stable, but it raises for the POLYMER spectra (except at 412nm, which shows lower RMSE). The spectral bias of l2gen remains the almost the same, absolute values increase slightly. The bias at 412 nm increases strongly for the POLYMER processors.

In case 1 waters (Global, IBQ, figure 7.3b) l2gen has still the lowest RMSE, but the RMSE of polPR05 is quite similar from 486 to 671 nm. In the blue, both POLYMER processors show larger RMSE and also biases. The bias of l2gen is spectrally flat and very close to zero. If MOBY is excluded, the RMSE of polPR05 at 443nm increases strongly, whereas values for l2gen remain stable. Still polPR05 and l2gen show very similar RMSE from 486nm onwards. The bias of polMM01 is here spectrally flat and close to zero, while polPR05 and l2gen show increasing (absolute) bias towards the blue.

In case 2 waters (IBQ, figure 7.3c) l2gen has the lowest spectral RMSE and the lowest absolute bias. Although RMSE of polPR05 is higher than of polMM01, the absolute bias of polPR05 is lower and closer to l2gen.

Changing the flagging of data to the common set of flags has a rather large influence on the statistics. The differences between the three AC processors becomes much smaller (CBQ, all global, figure 7.4a). But for the blue, the RMSE of polPR05 is the lowest. The bias of polPR05 and l2gen products is very similar, quite small and spectrally flat, whereas the products of polMM01 show larger biases at 412 and 551nm. If MOBY is excluded, the biases of the POLYMER processor change quite strongly in the blue. The spectral bias of l2gen products increases slightly over the entire spectrum.

As before, in case 1 waters, the spectral RMSE gets smoother (figure 7.4b). The POLYMER processors have the higher RMSE errors in the blue, but if MOBY data is excluded the error drops below the l2gen error, which remains quite stable. There is almost no difference in RMSE between the three processors from 486nm onwards. The bias of l2gen products is quite small and spectrally flat, whereas the POLYMER processors have similar larger biases in the blue. If the data from MOBY is excluded, the biases of polPR05 are the lowest.

In case 2 waters, the spectral RMSE of polPR05 is the lowest (but for 443nm). The spectral bias of l2gen is closest to zero.

Scores   In the IBQ selections, l2gen receives always the highest scores (figure 7.5). It has to be pointed out, that these statistics are based on about 340 spectra in contrast to 1200 spectra for the POLYMER processors.

In the majority of cases, polPR05 performs better than polMM01, especially if case 2 water samples are part of the in-situ data. The strong influence of the common set of flags is also reflected in the scores.

The results are still inconclusive. In an up-coming processing, VIIRS data is processed with l2gen.
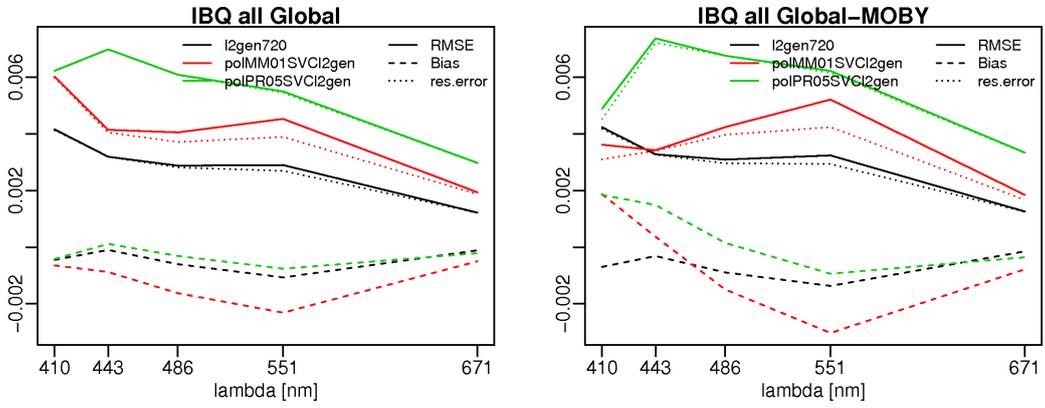
(a) All water types.
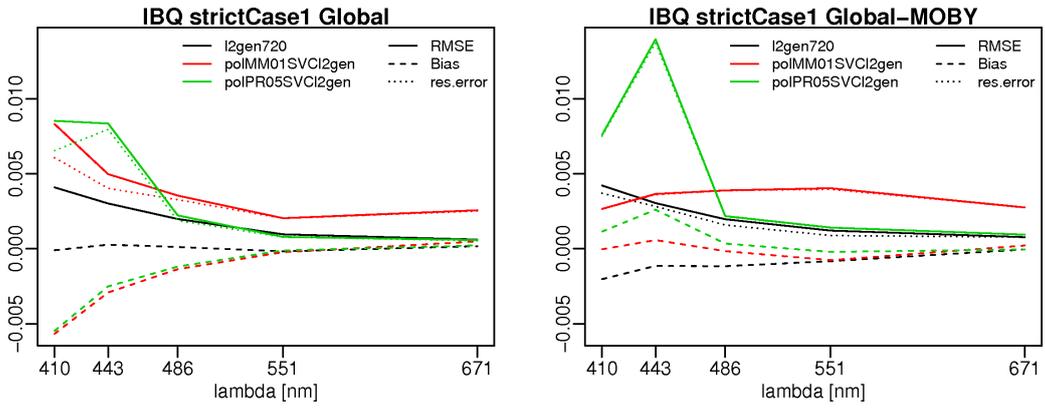
(b) Case 1.
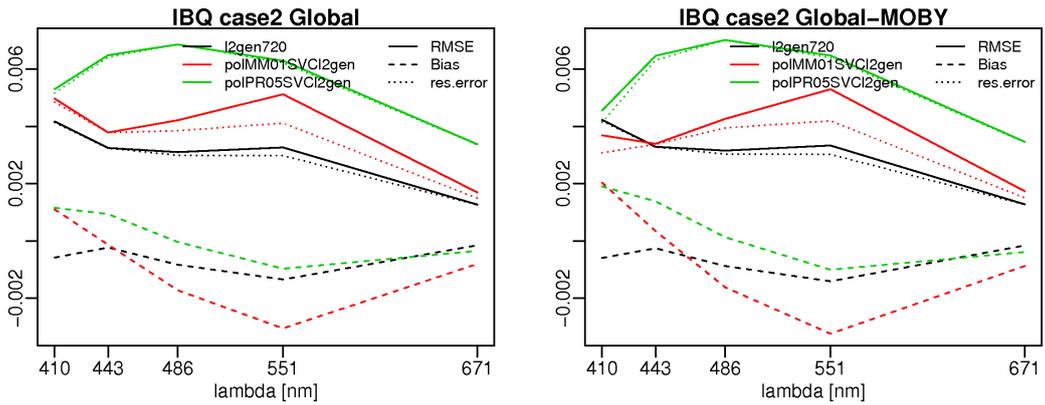
(c) Case 2.

Figure 7.3: Spectral statistics IBQ. VIIRS
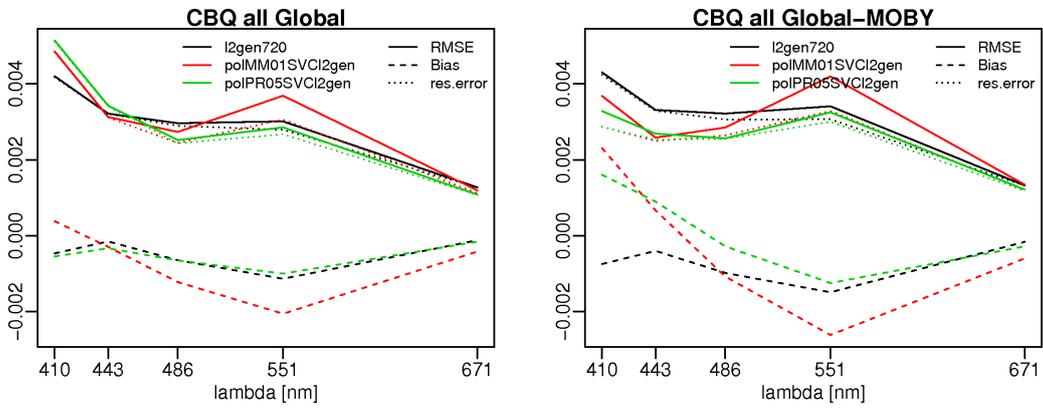
(a) All water types.
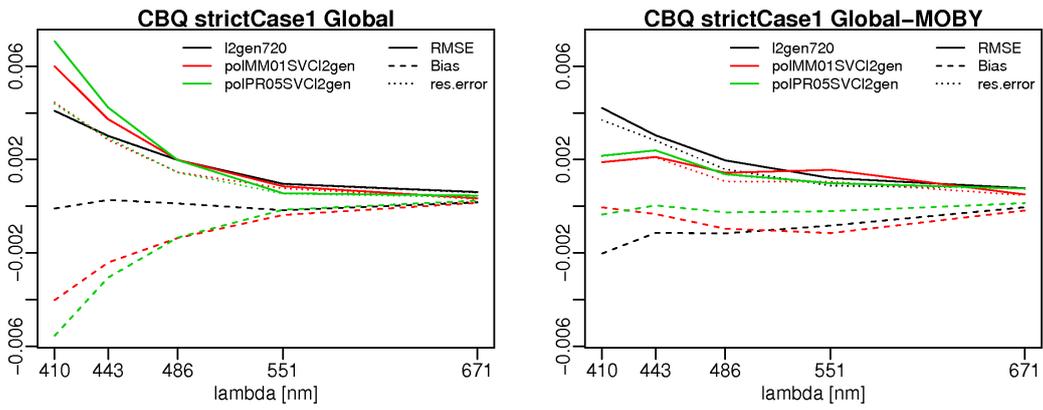
(b) Case 1.
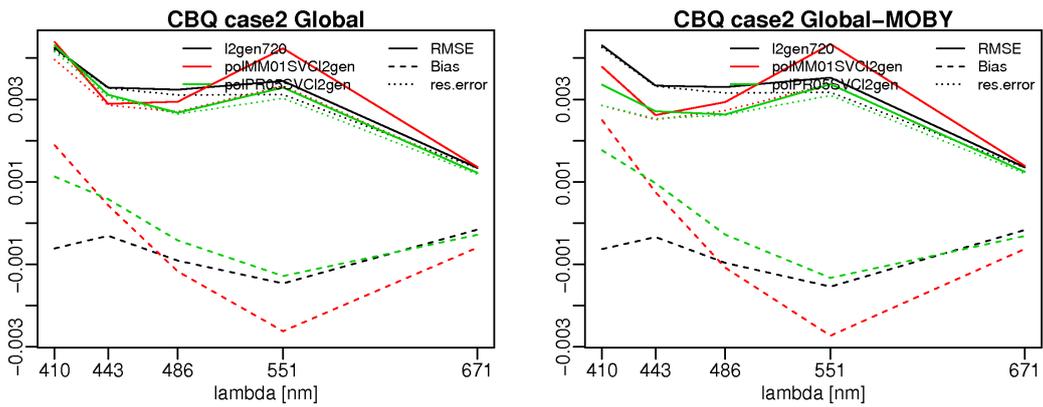
(c) Case 2.

Figure 7.4: Spectral statistics CBQ. VIIRS
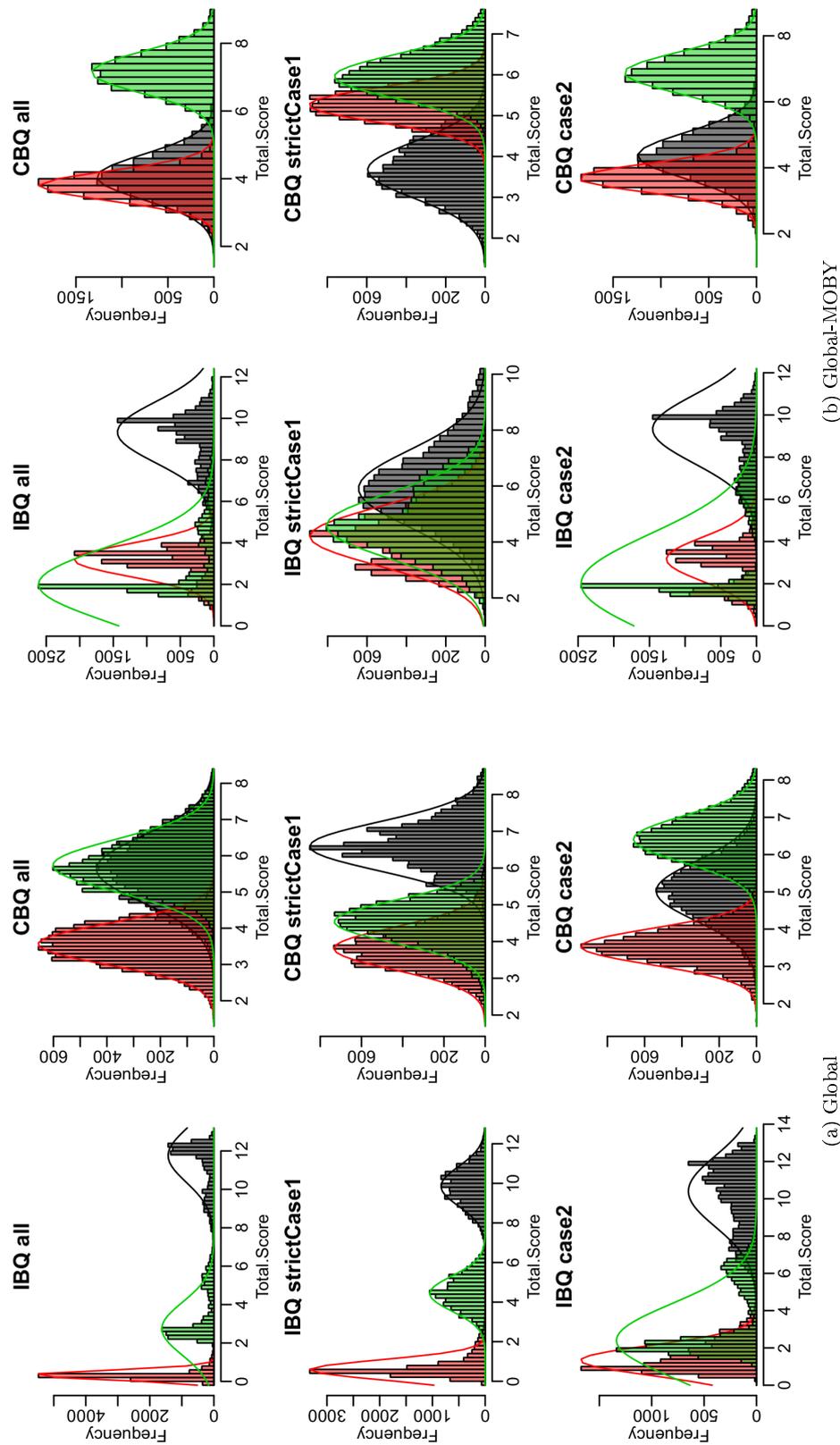
Figure 7.5: Bootstrapping the scores of VIIRS match-up statistics. Colours refer to l2gen (grey), POLYMER based on a Morel-Maritorena model (red) and POLYMER based on Park & Ruddick model (green)

(a) Global

(b) Global-MOBY

8 Illustration appendix

8.1 Water types based on band-shifted in-situ data
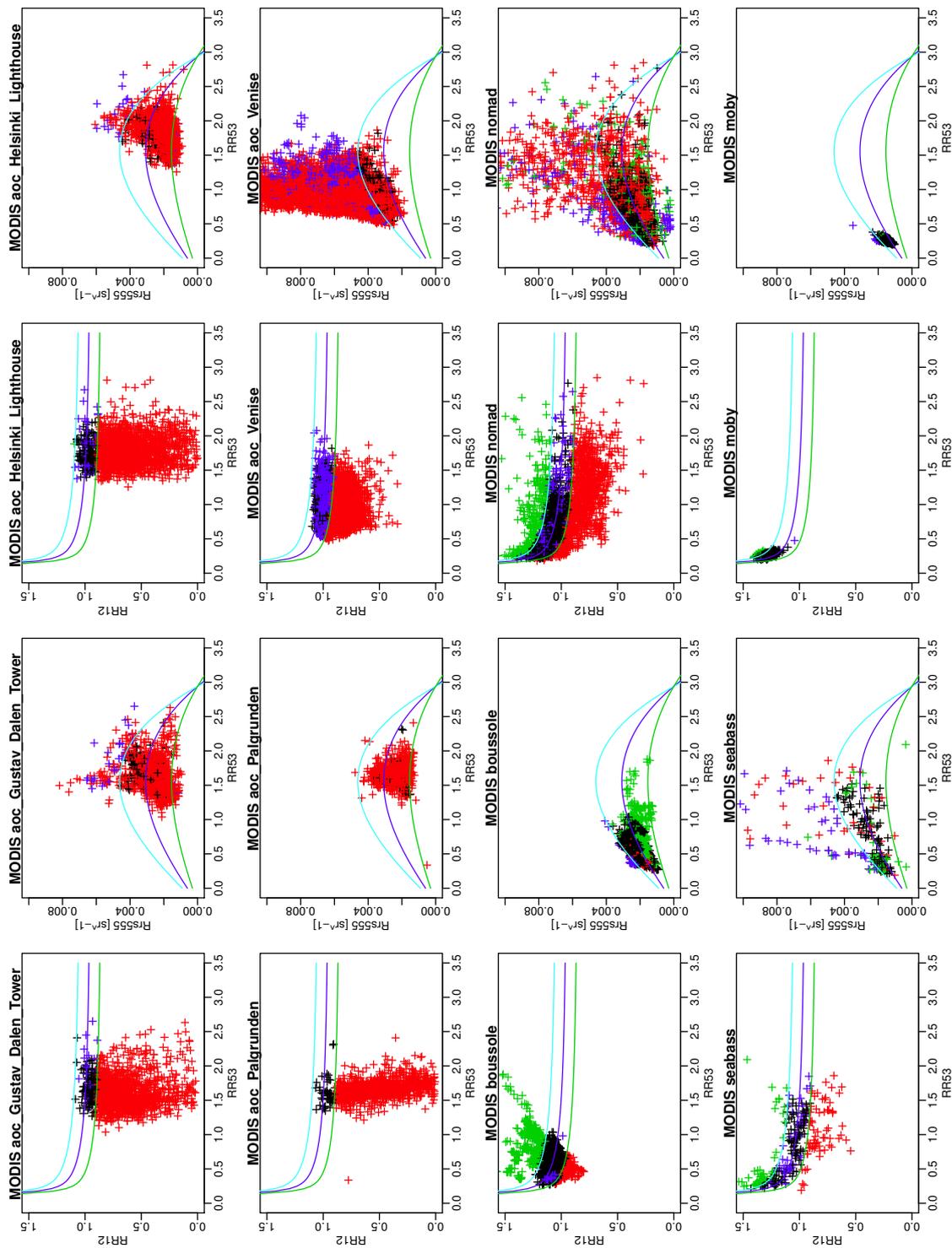
Figure 8.1: Classification of OC_CCI v3.0 in-situ database band-shifted to MODIS wavelengths per site, distinguishing four types of water: case 1 definition (Lee 2006) in black, higher CDOM/chl ratio (red), lower CDOM/chl ratio (green) and case 1 CDOM/chl but higher or lower backscattering (blue).
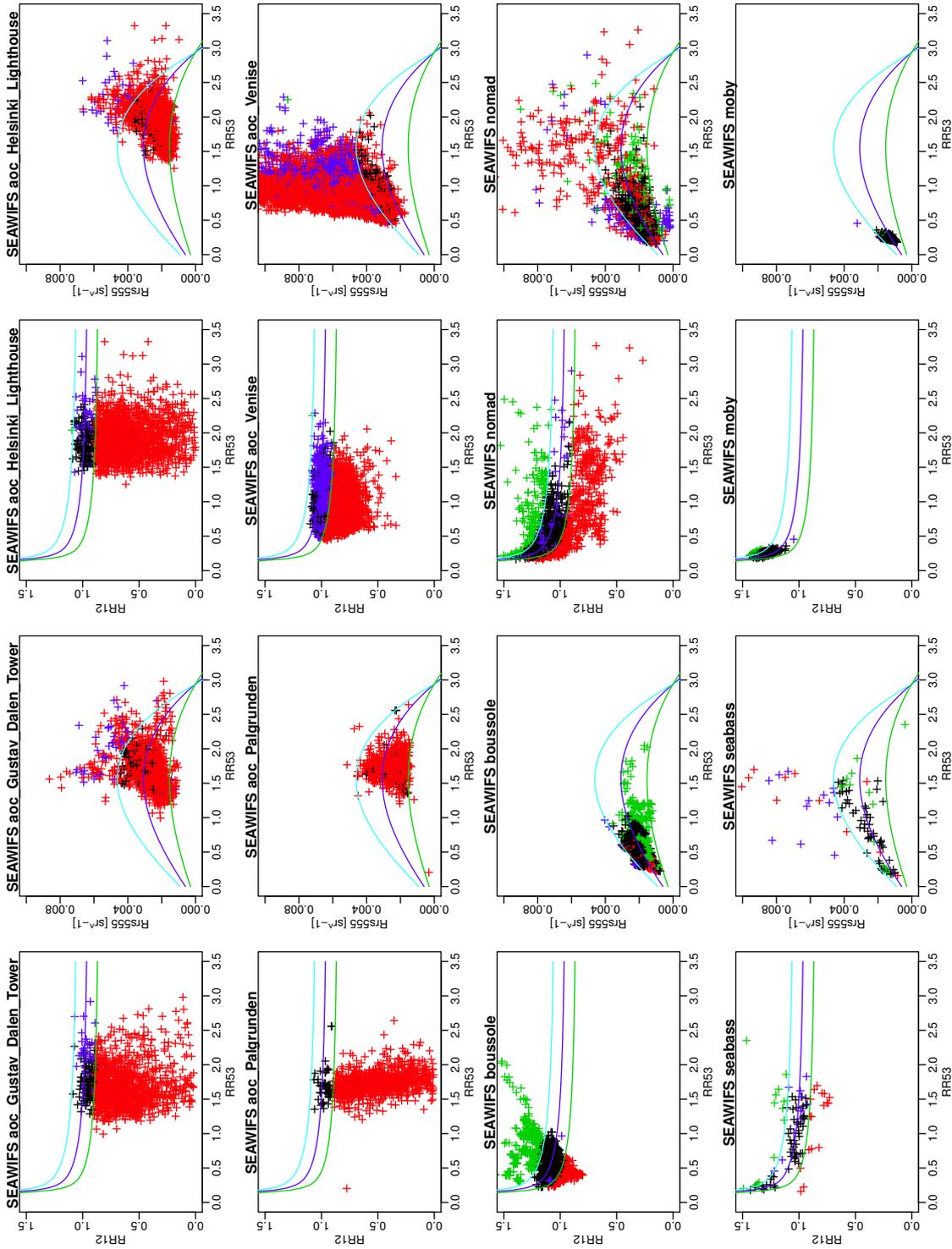
Figure 8.2: Classification of OC_CCI v3.0 in-situ database band-shifted to SeaWiFS wavelengths per site, distinguishing four types of water: case 1 definition (Lee 2006) in black, higher CDOM/chl ratio (red), lower CDOM/chl ratio (green) and case 1 CDOM/chl but higher or lower backscattering (blue).
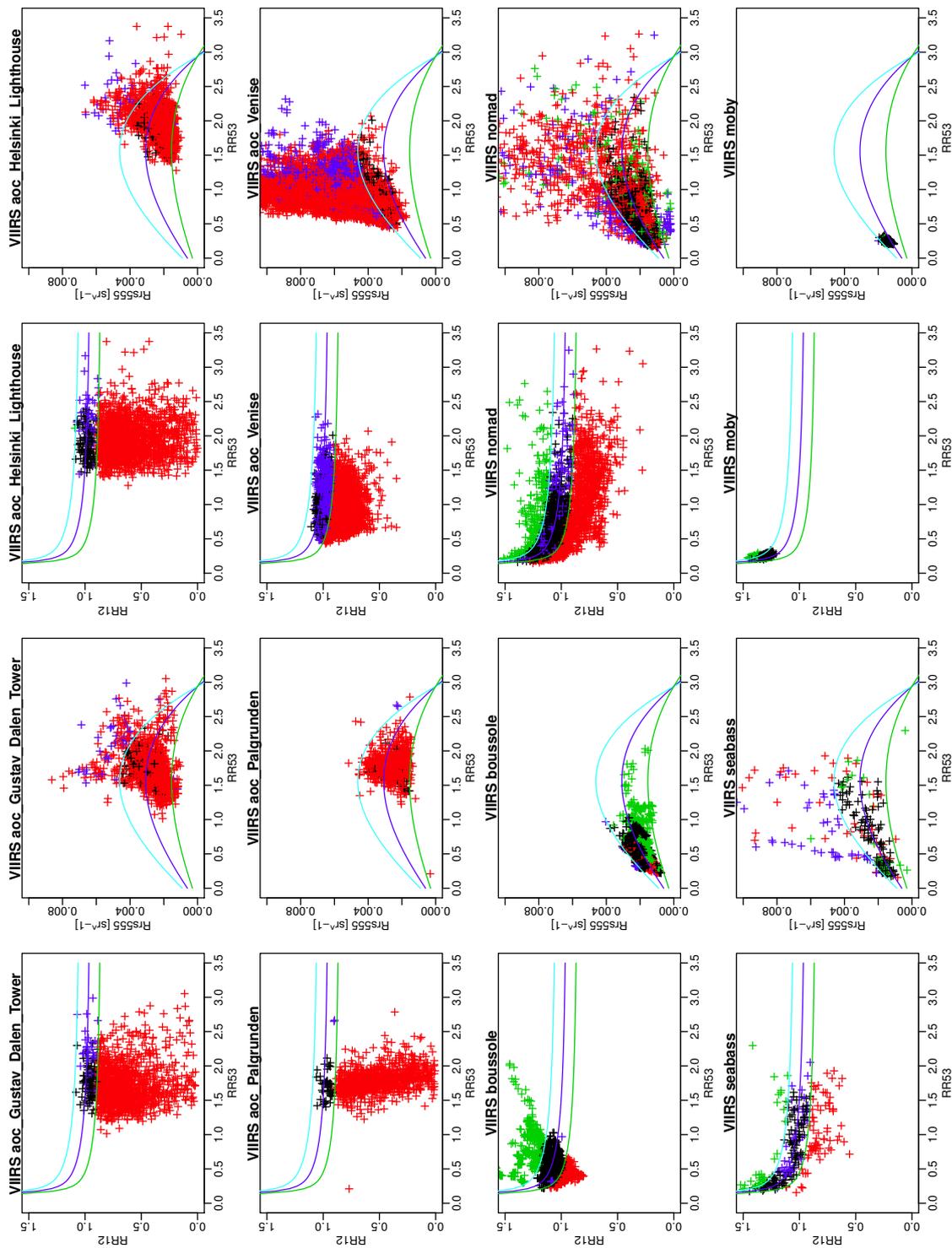
Figure 8.3: Classification of OC_CCI v3.0 in-situ database band-shifted to VIIRS wavelengths per site, distinguishing four types of water: case 1 definition (Lee 2006) in black, higher CDOM/chl ratio (red), lower CDOM/chl ratio (green) and case 1 CDOM/chl but higher or lower backscattering (blue).

References

James Aiken, Gerald Moore, Charles C. Trees, Stanford B Hooker, and Dennis K. Clark. The seawifs czcs-type pigment algorithm. Technical Report Vol. 29, NASA Technical Memorandum 104566, 1995.

R.J.W. Brewin, S. Sathyendranath, D. Müller, H. Krasemann, R. Doerffer, F. Mélin, C Brockmann, N. Fomferra, M. Peters, M. Grant, F. Steinmetz, P.-Y. Deschamps, J. Swinton, T. Smyth, P.J. Werdell, B. A. Franz, S. Maritorena, E. Devred, Z. Lee, Ch. Hu, and P. Regner. The ocean colour climate change initiative: a round-robin comparison on in-water bio-optical algorithms. *Remote sens. Environ.*, page in press, 2012.

K. L. Carder, F. R. Chen, Z. P. Lee, S. K. Hawes, and D. Kamykowski. Semianalytic moderate-resolution imaging spectrometer algorithms for chlorophyll a and absorption with bio-optical domains based on nitrate-depletion temperatures. *Journal of Geophysical Research*, 104(C3):5403–5421, 1999.

Zhong-Ping Lee, K. Du, K. Voss, Guiseppe Zibordi, B. Lubac, R. Arnone, and A. Weidemann. An inherent-optical-property-centered approach to correct the angular effects in water-leaving radiance. *Applied Optics*, 50(19):3155–67, 2011.

ZhongPing Lee and Chuanmin Hu. Global distribution of Case-1 waters: An analysis from SeaWiFS measurements. *Remote Sensing of Environment*, 101:270–276, 2006.

A. Morel and B. Gentili. Diffuse reflectance of oceanic waters. iii. implication of bidirectionality for the remote-sensing problem. *Appl. Opt.*, 35:4850–4862, 1996.

André Morel and Stéphane Maritorena. Bio-optical properties of oceanic waters: A reappraisal. *Journal of Geophysical Research*, 106(C4):7163–7180, 2001.

John E O'Reilly, Stéphane Maritorena, B Greg Mitchell, David A Siegel, Kendall L Carder, Sara A Garver, Mati Kahru, and Charles McClain. Ocean color chlorophyll algorithms for SeaWiFS. *Journal of Geophysical Research*, 103:24.937–24.953, 1998.

Young-Je Park and Kevin Ruddick. Model of remote-sensing reflectance including bidirectional effects for case 1 and case 2 waters. *APPLIED OPTICS*, 44(7), 2005.

A. Valente, V. Brotas, S. Sathyendranath, S. Groom, M. Grant, K. Voss, K. Hughes, G. Zibordi, J.-F. Berthon, E. Canuti, H. Klein, H. Loisel, M. Kahru, G. Mitchell, M. Ondrusek, S. Belanger, A. Poulton, L. Dransfeld, R. Barlowa, S. Gibb, S. Kratzer, V. Brando, D. Antoine, W. Balch, and K. Barker. A global in-situ bio-optical database for satellite validation and error characterisation. 2015.

Giuseppe Zibordi, Jean-François Berthon, Frédéric Mélin, Davide D'Alimonte, and Seppo Kaitala. Validation of satellite ocean color primary products at optically complex coastal sites: Northern adriatic sea, northern baltic proper and gulf of finland. *Remote Sensing of Environment*, 113:2574–2591, 2009.