Environ Monit Assess (2015) 187:364
DOI 10.1007/s10661-015-4585-4

Empirical and semi-analytical chlorophyll *a* algorithms for multi-temporal monitoring of New Zealand lakes using Landsat

Mathew G. Allan · David P. Hamilton · Brendan Hicks · Lars Brabyn

Received: 2 June 2014 / Accepted: 5 May 2015

Abstract The concentration of chlorophyll *a* (chl *a*; as a proxy for phytoplankton biomass) provides an indication of the water quality and ecosystem health of lakes. An automated image processing method for Landsat images was used to derive chl *a* concentrations in 12 Rotorua lakes of North Island, New Zealand, with widely varying trophic status. Semi-analytical and empirical models were used to process 137 Landsat 7 Enhanced Thematic Mapper (ETM+) images using records from 1999 to 2013. Atmospheric correction used radiative transfer modelling, with atmospheric conditions prescribed with Moderate Resolution Imaging Spectroradiometer (MODIS) Terra and AIRS data. The best-performing semi-analytical and empirical equations resulted in similar levels of variation explained (r^2=0.68 for both equations) and root-mean-square error (RMSE=10.69 and 10.43 μg L^{-1}, respectively) between observed and estimated chl *a*. However, the symbolic regression algorithm performed better for chl *a* concentrations <5 μg L^{-1}. Our Landsat-based algorithms provide a valuable method for synoptic assessments of chl *a* across the 12 lakes in this region. They also provide a basis for assessing changes in chl *a* individual lakes through time. Our methods provide a basis for cost-effective hindcasting of lake trophic status at a regional scale, informing on spatial variability of chl *a* within and between lakes.

Keywords Landsat · Bio-optical · Chlorophyll · Symbolic regression · Water quality · Remote sensing · Landsat · Semi-analytical

M. G. Allan (✉) · D. P. Hamilton · B. Hicks · L. Brabyn
Environmental Research Institute, The University of Waikato, Private Bag 3105, Hamilton 3240, New Zealand
e-mail: mat.g.allan@gmail.com

D. P. Hamilton
e-mail: davidh@waikato.ac.nz

B. Hicks
e-mail: hicksbj@waikato.ac.nz

L. Brabyn
e-mail: larsb@waikato.ac.nz

Introduction

Scientists and resource managers are required to assess lake water quality across wide temporal and spatial gradients. This information often forms the basis for decisions on the ecological status of a water body and its suitability for meeting a range of objectives. Historically, managers have relied on manual sampling, which can be time consuming and costly. More recently, satellite remote sensing has been adopted to provide cost-effective, high spatial resolution monitoring (e.g. Olmanson et al. 2008).

While remote sensing can only contribute direct information on optically active water quality constituents of water, it has the advantage of greatly increasing spatial resolution of monitoring compared with traditional in situ methods. In the past, Landsat has been the sensor of choice for monitoring inland water quality in small lakes because of its high spatial resolution. It has a freely available archive of data spanning more than 40 years. The Landsat 7 satellite has an Enhanced Thematic Mapper (ETM+) sensor with 30-m spatial

Published online: 19 May 2015

resolution, which compares favourably with ocean colour sensors (250–1000-m resolution), and provides sufficient resolution to detect spatial variations in optically active constituents (OACs), even for small lakes. Although Landsat remote sensing provides high spatial resolution, spectral resolution is low, which can result in individual bands containing spectrally opposing absorption and scattering features of OACs (Bukata et al. 1995). The low spectral resolution and signal-to-noise ratio (SNR) limit the complexity of any derived algorithms and reduce the ability to discriminate, for example, between chl *a* and suspended particles (SP) (Dekker and Peters 1993; Matthews 2011).

In a recent review of algorithms for the remote sensing of OACs (Matthews 2011), 27 studies were found to use Landsat ETM+ or Thematic Mapper (TM) satellite data to retrieve spatial distributions of chl *a*, turbidity, SP or Secchi depth. The review showed that in some instances, some of these constituents had high coefficients of determination, with sensor reflectance from satellites (e.g. r^2=0.99 for the retrieval of both SP (Dekker et al. 2002a, b) and chl *a* (Giardino et al. 2001)). Only a small number of examples exist, however, of remote sensing of inland water quality at landscape scales (e.g. Lillesand et al. 1983; Dekker et al. 2001a, b; Kloiber et al. 2002; Koponen 2006), and even fewer applications have used time series images to determine temporal changes in water quality (e.g. Dekker et al. 2001a, b; Olmanson et al. 2008). A major barrier to time series analysis of Landsat imagery is the time required to process large numbers of images. Automation of image processing could result in higher productivity, minimise user error and potentially enable near real-time image processing.

Of the Landsat studies cited above, only a few examine the underlying physical basis for the development of statistical algorithms (e.g. Dekker and Peters 1993; Gitelson et al. 1996; Brivio et al. 1997; Dekker et al. 2002a, b). The study of Allan et al. (2011) is typical of many empirically based Landsat studies of remote sensing of water quality, and, while useful for determining spatial distributions of chl *a*, derived algorithms have limited applicability beyond the spatial and temporal domain that was used for calibration against in situ samples, especially in more optically complex environments.

Bio-optical algorithms have potential to more clearly differentiate OACs. They have generally been applied to high spectral resolution satellite imagery (e.g. from Moderate Resolution Imaging Spectroradiometer (MODIS)), and to hyperspectral data, both of which allow precise measurement of spectral slopes. However, successful applications of simplified bio-optical models for single water quality parameter retrieval exist using broadband sensors where the OAC of interest dominate absorption and scattering. For example, semi-analytical models have been used to map SP concentrations using Landsat and Satellite Pour l'Observation de la Terre (SPOT) satellite data (Dekker et al. 2002a, b). The advantage of bio-optical models is that in situ data are not needed at the time of image capture, allowing for multi-site and multi-sensor comparisons through time. Dekker et al. (2002a, b) found that bio-optical algorithms for SP were more reliable and temporally robust than empirical algorithms. Their study also found that random point samples within the synoptic estimates of SP were within generally ±20–30 % of SP in situ grab-samples but could also deviate by as much as ±4000 %.

Inherent optical properties vary temporally and spatially (Kostadinov et al. 2010; Moisan et al. 2011; Devred et al. 2011), which presents a potential source of error when inverting semi-analytical models to determine optically active constituent concentrations. For example, variations in the phytoplankton pigment-specific absorption coefficient ($a^*\phi(\lambda)$) are related to phytoplankton pigment composition, cell size, packaging effect, light accumulation and nutrient limitation (Babin et al. 1993; Babin 2003; Bricaud 2004; Blondeau-Patissier et al. 2009). Variation of $a^*\phi(\lambda)$ causes non-linearity between light absorption and chl *a* concentration (Bricaud and Morel 1981). While there is considerable literature on variations in $a^*\phi(\lambda)$, the backscattering ratio $B_p\phi$ and the specific scattering coefficient $b^*\phi(\lambda)$ of phytoplankton are less well understood. They are determined by the size, physical structure and the outer coating of phytoplankton cells (Stramski et al. 2004). With increasing phytoplankton biomass, there is a greater cell wall surface area to produce scattering (Yacobi et al. 1995). The presence of gas vacuoles in some cyanobacteria has also been found to substantially increase the value of $b^*\phi(\lambda)$ (Dubelaar et al. 1987; Volten et al. 1998; Matthews and Bernard 2013).

The Rotorua lakes (Bay of Plenty region, North Island, New Zealand) include a number of lakes within a relatively small geographical area, which are subject to a regular programme of water quality monitoring (Burns

et al. 2005). They provide an ideal opportunity to conduct a regional-scale investigation of applications of remote sensing for determining water quality within and between lakes through time. In order to construct robust satellite-retrieval algorithms for chl *a*, validation is required for a time series of images. For algorithms to be robust, some knowledge of the physical mechanisms underpinning algorithm function is needed, especially in terms of identifying potential sources of error. With these factors in mind, the objectives of this study were to (1) determine the underlying physical mechanisms by which Landsat may be used to distinguish changes in chl *a* using a bio-optical model, (2) determine the accuracy of semi-analytical and empirical models for estimating chl *a* over a wide range of concentrations and (3) develop an automated image processing methodology to enable the processing of large numbers of images. A radiative transfer-based atmospheric correction was used with empirical and semi-analytical modelling used for chl *a* retrieval.

Methods

Study site

The Rotorua lakes (Fig. 1a) have a wide range of trophic states and mixing regimes (Burns et al. 2009) (Table 1) and lakes Rotorua, Rotoiti, Rotoehu, Rotomahana and Tarawera also have nutrient-enriched geothermal inputs (Hoellein et al. 2012). In August 2008, a diversion wall was constructed in Lake Rotoiti (Fig. 1b). Its purpose was to divert nutrient-rich water from Lake Rotorua, which enters Lake Rotoiti via the Ohau Channel inflow, away from the main body of Lake Rotoiti and towards the main outflow, Katituna River, adjacent to the inflow in the north-west region of the lake. Our study included Landsat imagery and in situ data captured before and after the construction of the diversion wall.

Satellite imagery and software

We used 137 Landsat 7 satellite images captured between 1999 and 2013 for this study. All image processing routines were automated using scripts written in Interactive Data Language (IDL) linked to Environment for Visualizing Images (ENVI) routines when necessary. All images were downloaded from the GloVis website created by USGS (http://glovis.usgs.gov/).

In situ data

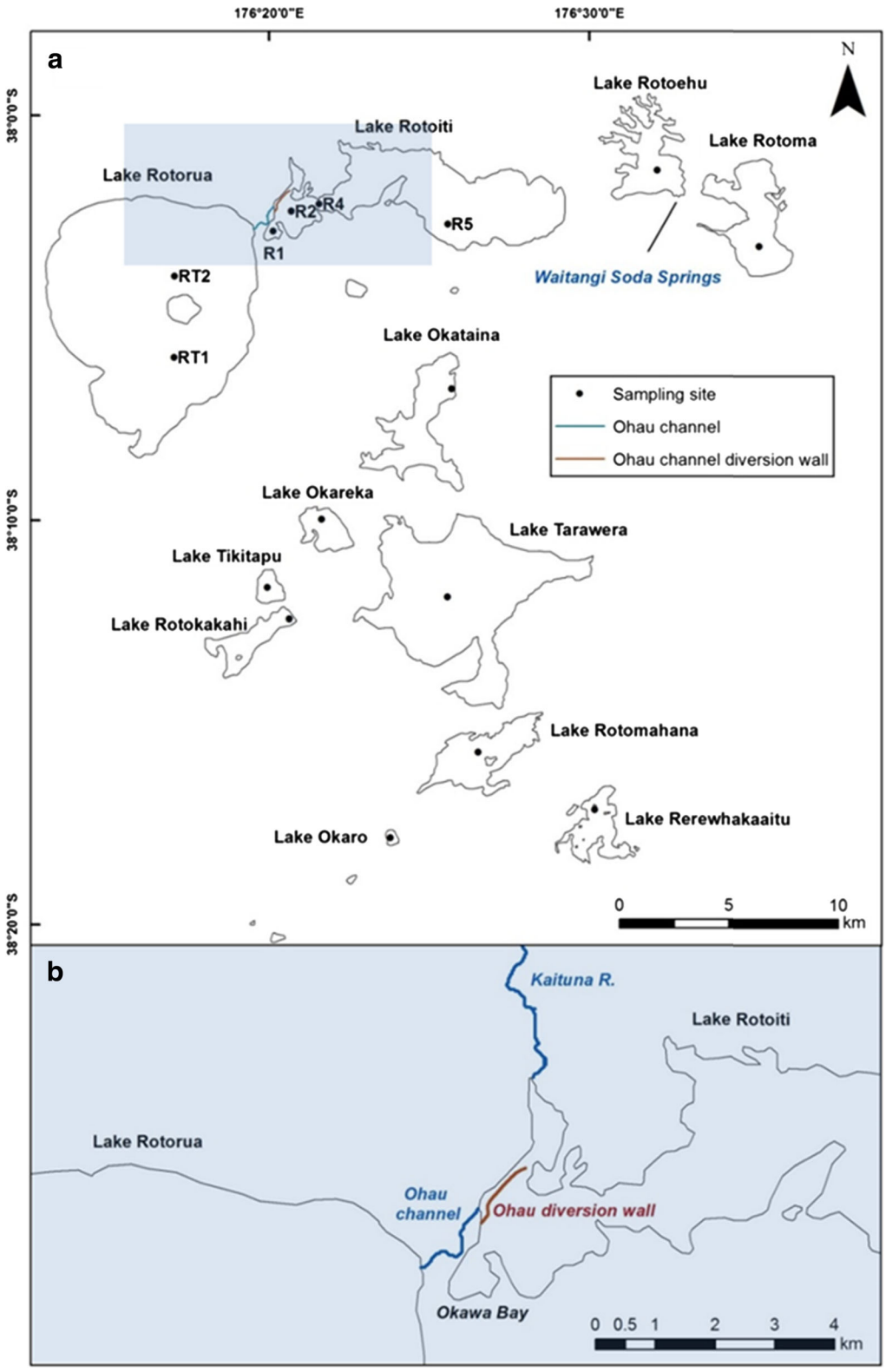
Monthly in situ samples of chl *a* for the Rotorua lakes were obtained from Bay of Plenty Regional Council. Surface samples are depth integrated (0- to 4–9-m depth in eutrophic lakes and 0- to 10–17-m depth in meso-oligotrophic lakes). From the satellite image capture dates (1999 to 2013), there were 27 in situ samples taken on the same date as in situ monitoring. In addition, all in situ samples corresponding to images from 25 January 2002 (NZST) and 24 October 2002 (n=29) were included, for which samples were collected within 5 days of image capture, except for Lake Rerewhakaaitu (in situ collected on 9 January 2002 and image capture on 25 January 2002).

All Lake Rotoma samples (n=38) captured within 1 month of satellite overpass were included in the algorithm development in order to ensure any derived models were suitable for estimation of chl *a* in oligotrophic lakes. The average in situ chl *a* in Lake Rotoma over the study period was 1.5 μg L^{-1}, and the maximum concentration was usually in winter or spring. Concentrations of chl *a* vary little in this lake (Paul et al. 2012), and therefore we assumed that data from satellite overpass dates could be matched to the corresponding monthly in situ samples.

Lake Rotorua is the most optically complex lake of the 12 Rotorua lakes, with particle resuspension previously observed during high winds (Stephens et al. 2004). Suspended particle concentration was not routinely measured during the study period in Lake Rotorua; however, a previous study (Vant and Davies-Colley 1986) using monthly samples from July 1983 to October 1984 gave a mean concentration of 6.7 mg L^{-1}, with range of 2.4–15.9 mg L^{-1}. Lake Rotoma is the most optically simple and has highest water clarity of the 12 lakes, and Lake Rotoehu exhibits the greatest spatial variability of chl *a*. These three lakes were chosen to assess the feasibility of developing time series analysis of chl *a* from remote sensing based on ground truthing with in situ data.

Image preparation

An automated image processing methodology was developed in order to allow efficient processing of the entire


176°20'0"E
176°30'0"E
38°0'0"S
38°10'0"S
38°20'0"S
a
N
Lake Rotoehu
Lake Rotoiti
Lake Rotoma
Lake Rotorua
R4
R2
R5
R1
RT2
RT1
Waitangi Soda Springs
Lake Okataina
Sampling site
Ohau channel
Ohau channel diversion wall
Lake Okareka
Lake Tarawera
Lake Tikitapu
Lake Rotokakahi
Lake Rotomahana
Lake Rerewhakaaitu
Lake Okaro
0 5 10 km
b
Kaituna R.
Lake Rotoiti
Lake Rotorua
Ohau channel
Ohau diversion wall
Okawa Bay
0 0.5 1 2 3 4 km

◀ **Fig. 1** **a** The Rotorua lakes' study site showing in situ sampling locations and **b** the extent of expanded area represented with shading from (**a**) to reveal the location of the Ohau diversion wall in Lake Rotoiti. Each in situ sampling location corresponds to a 5×5 matrix (30-m pixel) average from Landsat ETM+ satellite images

archive of Landsat ETM+ images for the Rotorua lakes. Procedures were created in IDL, which were all called from a main control procedure. The radiative transfer model was called from within IDL (6sv; Kotchenova et al. 2008), in the form of an executable which we compiled for a Windows operating system. Conversion from Landsat-scaled radiance to spectral radiance at the sensor aperture (L_λ; W m^{-2} sr^{-1} μm^{-1}) used a procedure (landsat_rad_sixs.pro) which read the Landsat metadata file in order to apply gains and offsets in the radiance conversion calculation. In order to generate inputs for atmospheric correction, this procedure also read the acquisition date and solar elevation angle from the Landsat metadata (to calculate solar zenith angle).

Atmospheric correction required the aerosol optical depth (AOD). In New Zealand, AOD is measured only at Lauder (860 km south-west of Lake Taupo in the North Island of New Zealand). Records of AOD from this station are among the lowest recorded globally, ranging from about 0.01 to 0.08 (Liley and Forgan 2009). To account for variations in AOD between Lauder and the Rotorua lakes, *Giovanni* Data and Information Services Center (NASA) was used to create time series outputs of AOD, water vapour (derived from MODIS Terra MOD08) and total column ozone (derived from MODIS Atmospheric Infrared Sounder (AIRS)). The global validation of MODIS AOD estimation error over land, ε, is given as

$$\varepsilon = \pm 0.03 + 0.05\mathrm{AOD} \quad \text{(Remer et al. 2005)} \qquad (1)$$

For instances when the MODIS Terra-retrieved AOD was negative, values were set to 0.05. Monthly mean values measured at Lauder range between 0.02 and 0.05 (Liley and Forgan 2009). It was assumed AOD over the Rotorua lakes would likely be higher than at Lauder; therefore, the value of 0.05 was chosen.

Following derivation of atmospheric correction parameters, the IDL procedure (landsat_rad_sixs.pro) was used to write input files for a radiative transfer model (6sv) for each band of Landsat. The 6sv input file contained the lake elevation (300 m), sensor elevation (705 km) and selected spectral band and aerosol model (continental). In addition, the bidirectional reflectance distribution function was chosen along with a non-homogenous target. The 6sv model was then run using the input files generated using an IDL procedure. The output files were then read

Table 1 Lake and catchment characteristics of the Rotorua lakes

Lake name	Lake area	Maximum lake depth	Mean lake depth	Annual mean chl *a*	Trophic state	Catchment pasture	Catchment indigenous forest/scrub	Catchment exotic forest
	km^2	m	m	$\mu g\ L^{-1}$		%	%	%
Okareka	3.4	33.5	20.0	3.9	Mesotrophic	37.8	51.6	7.6
Okaro	0.3	18.0	12.1	33.5	Supertrophic	90.6	2.1	6.3
Okataina	10.8	78.5	39.4	2.1	Oligotrophic	10.7	84.1	7.8
Rerewhakaaitu	5.3	15.8	7.0	2.9	Mesotrophic	75.3	7.2	15.2
Rotoehu	8.0	13.5	8.2	10.6	Eutrophic	34.2	33.4	32.0
Rotoiti	34.0	125.0	60.0	9.6	Mesotrophic	15.9	36.4	46.2
Rotokakahi	4.4	32.0	17.5		Eutrophic	26.3	16.6	57.1
Rotoma	11.1	83.0	36.9	1.3	Oligotrophic	23.4	46.0	26.7
Rotomahana	9.0	125.0	60.0	4.2	Mesotrophic	43.2	39.7	16.3
Rotorua	80.6	44.8	11.0	22.8	Eutrophic	51.8	25.1	14.3
Tarawera	41.3	87.5	50.0	1.5	Oligotrophic	19.7	62.4	16.0
Tikitapu	1.5	6.2	18.0	1.8	Oligotrophic	7.0	74.3	17.9

Source: Scholes and Bloxham 2008, Scholes 2011

(readsixs.pro) in order to convert radiance to water surface irradiance reflectance ($R(0+)$).

The 'Automated Cloud-Cover Assessment' of Irish et al. (2006) was implemented using an IDL procedure called acca.pro. Finally, a water-only image of $R(0+)$ was created using a mask that intersected lake polygons and a classification of Landsat band 5, in which water was differentiated from land, based on absorption of electromagnetic radiation in band 5. Surface irradiance reflectance $R(0+)$ was converted to subsurface irradiance reflectance ($R(0-)$) using an air-water interface parameter of 0.544 (Mobley 1994).

In order to fill missing data in Landsat images caused by the failure of the Scan Line Corrector (SLC), the ENVI routine DEM_BAD_DATA_DOIT was used in the IDL procedure repalcebaddata.pro. Delany triangulation was used to fill reflectance values less than −0.000001.

Algorithm development

Within the study, we tested a total of five different algorithms to estimate chl *a* from Landsat reflectance. Simplifications to three semi-analytical models were used to estimate chl *a* from Landsat B1, B2 and B3 $R(0-)$. Semi-analytical models were compared two empirical algorithms: one derived from symbolic regression, and one derived from regression between SP estimated from a simplified semi-analytical model and in situ chl *a*.

Forward and inverse bio-optical modelling

Forward bio-optical modelling was used to quantify the physical processes relating Landsat-measured reflectance and chl *a* concentration. The resulting equations were used to estimate chl *a* and SP using a simplified semi-analytical model of Landsat $R(0-)$.

Dekker et al. (1997) found the following bio-optical model to be suitable to estimate $R(0-)$ from absorption and scattering for turbid waters:

$$R(0-1) = r_1\left[\frac{b_b}{b_b + a}\right] \quad (2)$$

where r_1 depends on the anisotropy of the downwelling light field and scattering processes within the water, b_b (m^{-1}) is the total backscattering and a (m^{-1}) is total absorption. A value of r_1 of 0.31 was used in this study (cf. Gordon et al. 1988).

The absorption and backscattering coefficients are made up of the sum of individual optically active components:

$$b_b(\lambda) = b_{bw}(\lambda) + B_{bp}b^*{}_{\mathrm{SP}}(\lambda)C_{\mathrm{p}} + B_{b\phi}b^*{}_{\phi}(\lambda)C_{\phi} \quad (3)$$

$$a(\lambda) = a_w(\lambda) + \mathrm{C}_{\phi}a^*{}_{\phi}(\lambda) + C_{\mathrm{p}}a^*{}_{\mathrm{SP}} + a_{\mathrm{CDOM}}(\lambda) \quad (4)$$

$$a_{\mathrm{CDOM}}(\lambda) = a_{\mathrm{CDOM}(440)}a^*{}_{\mathrm{CDOM}}(\lambda) \quad (5)$$

where

(λ)	wavelength
$b_{bw}(\lambda)$	backscattering coefficient of water
$B_{bp}(\lambda)$	backscattering ratio of particulates
$b^*{}_p(\lambda)$	specific scattering coefficient of particulates
C_{p}	concentration of particulates
$B_{b\phi}(\lambda)$	backscattering ratio of phytoplankton
$b^*\phi(\lambda)$	specific scattering coefficient of phytoplankton
$a_w(\lambda)$	absorption coefficient of pure water
$C\phi$	concentration of chl *a*
$a^*\phi(\lambda)$	specific absorption coefficient of phytoplankton
$a^*{}_{\mathrm{p}}(\lambda)$	specific absorption coefficient of particulates
$a_{\mathrm{CDOM}(440)}$	coloured dissolved organic matter (CDOM) absorption at 440 nm
$a^*\mathrm{CDOM}(\lambda)$	specific absorption of CDOM

Values of $a_w(\lambda)$ and $b_{bw}(\lambda)$ were assigned to literature values (Morel 1974; Pope and Fry 1997). The bio-optical model was implemented using an optimised version of BIOPTI (bio-optical model for inland waters (version 1.0)) (Dekker et al. 2001b). BIOPTI uses inherent optical properties (IOPs) of Dutch inland waters (Dekker et al. 1997). Optimised parameters in the model included the phytoplankton-specific absorption coefficient $a^*\phi(\lambda)$, for which literature values for the cyanobacterium *Microcystis aeruginosa* (Zhang et al. 2012) were applied. This $a^*\phi(\lambda)$ value is only slightly higher than the value used in BIOPTI for cyanobacteria assemblages (Dekker 1993). *Microcystis aeruginosa* is a species of cyanobacteria commonly present in the Rotorua

lakes, and *Microcystis* spp. dominates the cyanobacteria genera found in lakes Rotoiti and Rotoehu (Wood et al. 2006). Recent work has found *M. aeruginosa* is one of the most efficient phytoplankton in backscattering light (Matthews and Bernard 2013; Zhou et al. 2012). The specific backscattering coefficient $b_b^*\phi(\lambda)$ of phytoplankton was parameterized within the range of measured and modelled values of *M. aeruginosa* derived from the literature (Matthews and Bernard 2013; Zhou et al. 2012). Matthews and Bernard (2013) used a value of $b_b^*\phi(510)$ of 2.1×10^{-3} m^2 mg^{-1} based on modelling of a heavily vacuolated cell, somewhat lower than that observed by Zhou et al. (2012) of 5.7×10^{-3} m^2 mg^{-1}. Matthews and Bernard (2013) found the spectral slope of $b_b^*\phi$ for *M. aeruginosa* was slightly downward sloping but became more positive towards the red band (700 nm). Zhou et al. (2012) also found $b_b^*\phi$ for *M. aeruginosa* was downward sloping. For our study, we assumed $b_b^*\phi$ was constant over each Landsat bandwidth and derived $b_b^*\phi$ by iteration, whereby the range of modelled and Landsat atmospherically corrected satellite B3 $R(0-)$ values was similar. The value of $b_b^*\phi$ adopted for our study was 3.8×10^{-3} m^2 mg^{-1} for the B1 band, 4.8×10^{-3} m^2 mg^{-1} for B2 and 3.5×10^{-3} m^2 mg^{-1} for B3.

The bio-optical simulations were run by varying chl *a* concentration using 30 concentration increments; 0–5.0 at 0.5 μg L^{-1} intervals, 5–23 at 2 μg L^{-1} intervals and 25–250 at 25 μg L^{-1}. Over these simulations, $a_{CDOM(440)}$ was fixed at 0.16 m^{-1} and SP concentration was fixed at 0.5 mg L^{-1} (SP and CDOM concentrations were not measured over the study period). While *aCDOM* (440) measurements were not available over our study period, previously in Lake Rotorua, the average $a_{CDOM(440)}$ was 0.23 m^{-1} and in Lake Rotokakahi 0.09 m^{-1} for 29 measurements between October 1983 and September 1984 (Davies-Colley and Vant 1987). The average value for these two lakes (0.16 m^{-1}) was assigned to $a_{CDOM(440)}$ values in the bio-optical model. Once forward bio-optical models were run for varying chl *a* concentrations, simplified relationships were fitted between chl *a* concentration and modelled Landsat $R(0-)$ in Landsat bands 1–3, in order to estimate chl *a* concentration from $R(0-)$ retrieved from time series satellite images.

In addition, bio-optical simulations of the effect of increasing SP concentrations were run by varying SP concentration within BIOPTI, with zero chl *a* concentration and $a_{CDOM(440)}$. The bio-optical simulations were run by varying SP concentrations from 0.5 to 9.5 at 1.0 mg L^{-1} concentration increments.

Statistical analysis

For all the tested algorithms, root-mean-square error (RMSE) and coefficient of determination (r^2) values were calculated between in situ-observed and model-estimated chl *a*. RMSE combines residuals from differences of estimated and observed values into a single measure of predictive power. It is defined as

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - x_i)^2}{n}} \quad (6)$$

where y_i is the observed value and x_i is the modelled value corresponding to a given time and location *i*.

Symbolic regression

Eureka Formulize is a scientific data mining software tool that detects mathematical patterns in experimental data. This software uses symbolic regression (Koza 1994) to search for equations that describe the mechanisms that produce the data. Symbolic regression uses evolutionary computational algorithms to search a space of mathematical expressions while minimising selected error metrics. Traditional linear and non-linear regressions fit parameters to a standard equation, whereas symbolic regression searches both the parameter space and the form of the equation simultaneously (Schmidt and Lipson 2009).

Symbolic regression was used in this study in order to include all the visible bands of Landsat in algorithm development. We selected the following mathematical operators for use in a symbolic regression equation: constant, integer constant, addition, subtraction, multiplication, division, exponential, natural log and power. The data were randomly shuffled and split into training and validation data sets.

Results

Bio-optical modelling simulations of the influence of phytoplankton on R(0−)

Subsurface irradiance reflectance ($R(0-)$) was modelled theoretically from the forward bio-optical model

(Fig. 2), varying chl *a* concentrations from 0 to 250 μg L^{-1} in 30 concentration increments (with fixed values of CDOM absorption of 0.16 m^{-1} and SP concentrations of 0.5 mg L^{-1}). This represents an idealised scenario as CDOM and SP concentrations vary naturally. Relative Landsat 7 spectral response functions ranging from 0 to 1 are shown for the B1, B2 and B3 bands. A water column with no chl *a* yields high $R(0-)$ at blue wavelengths compared to $R(0-)$ at red wavelengths. For Landsat B1, the bio-optical model gives an increase in $R(0-)$ with increasing chl *a* concentration due to higher phytoplankton backscattering than pigment absorption. For B2, there is a peak in reflectance at 550 nm coinciding with an increasing influence of phytoplankton backscattering and decreasing phytoplankton pigment absorption. Band 3 includes a local reflectance peak at c. 650 nm (due to locally decreased phytoplankton pigment absorption) and a reflectance trough at c. 680 nm caused by chl *a* absorption. A feature for wavelengths >680 nm is capture of the base of the reflectance peak at 704–714 nm, which results from very low phytoplankton absorption and a local increase in phytoplankton backscattering.

Semi-analytical relationships for estimation of chl *a* and SP from Landsat satellite data

The semi-analytical relationships produced by the bio-optical model were approximated using fitted exponential or polynomial relationships (Fig. 3). These algorithms were subsequently applied to estimate chl *a* concentrations (Fig. 4a–c). In addition to estimation of chl *a*, a semi-analytical relationship was used to estimate SP concentrations in the Rotorua lakes. First, theoretical

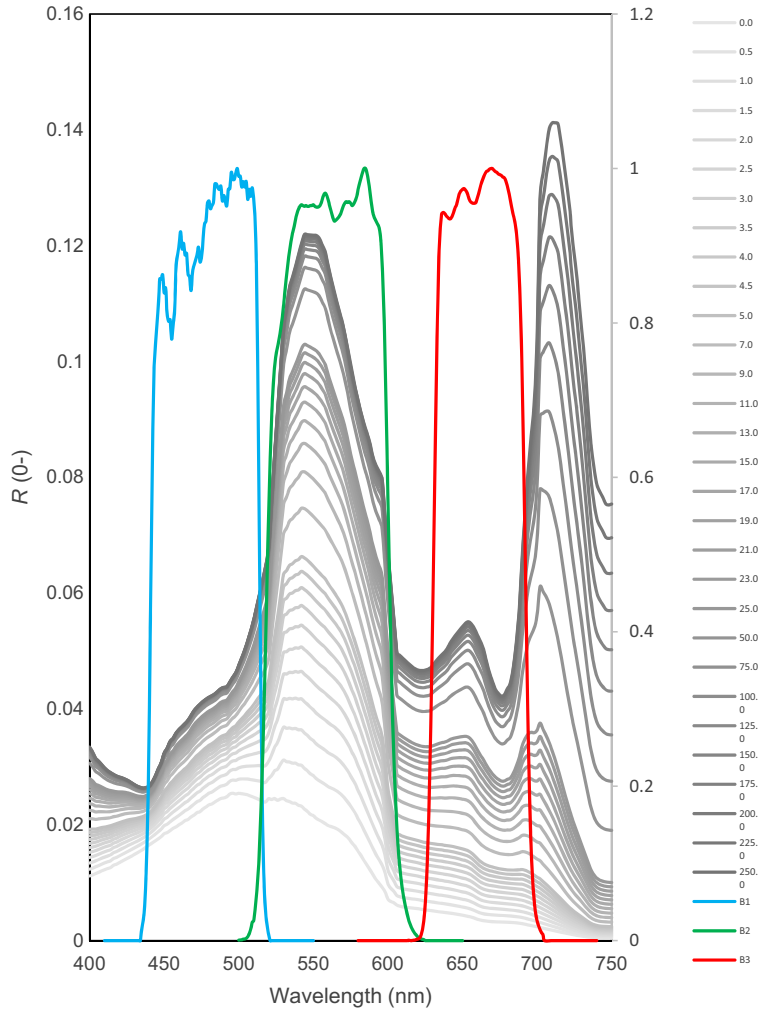


Fig. 2 Modelled subsurface reflectance ($R(0-)$) for varying chlorophyll (*chl*) *a* concentrations ranging from 0 to 450 μg L^{-1} with fixed coloured dissolved organic matter (CDOM) absorption of 0.16 m^{-1} and suspended particle (SP) concentration of 0.5 mg L^{-1}. The relative spectral response of Landsat bands B1, B2 and B3 is overlaid

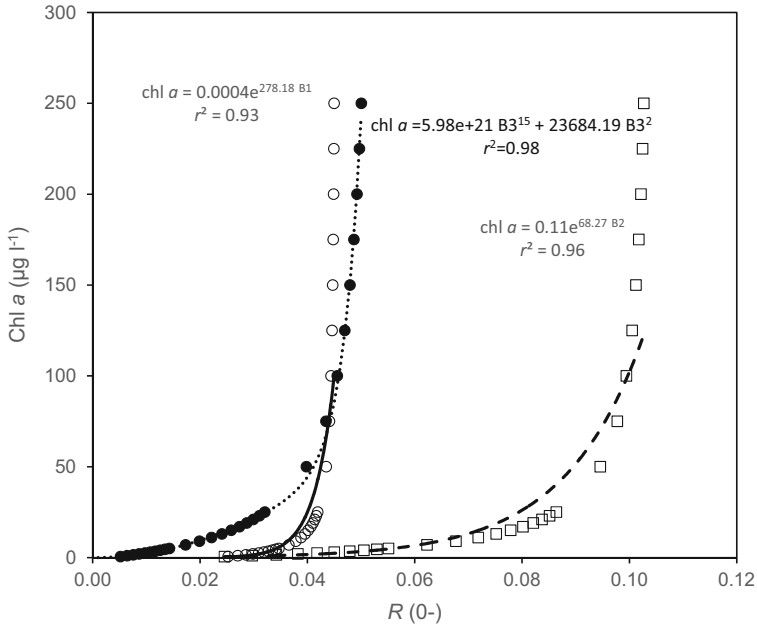


Fig. 3 The semi-analytical relationship between chlorophyll (*chl*) *a* concentrations and Landsat band-averaged subsurface irradiance reflectance *R*(0−) in B3 (*closed circles*), B2 (*open squares*) and B1 (*open squares*) based on individual analytical solutions. An exponential or polynomial relationship (*black line*) is used to fit the analytical relationship with closeness of fit given by r^2. These functions were used to estimate chl *a* from *R*(0−)

SP concentrations were plotted as function of the average of Landsat B2 and B3 *R*(0−) values from BIOPTI modelling (Fig. 4). A polynomial regression was used to approximate the semi-analytical relationship, and this function was then used to estimate SP concentration from average values of B2 and B3 Landsat *R*(0−). This method has been previously used for Frisian lakes (Dekker et al. 2002a, b) for estimation of SP. The SP estimated using the semi-analytical model was also used to estimate chl *a*, by fitting a polynomial relationship to in situ values of chl *a* concentration (Figs. 5 and 6c).

Statistical relationships and regression models between in situ chl *a* and Landsat B1-B4 reflectance

In order to determine which Landsat-derived parameters have potential for use in modelling chl *a*, values of r^2 were examined for relationships of in situ chl *a* to

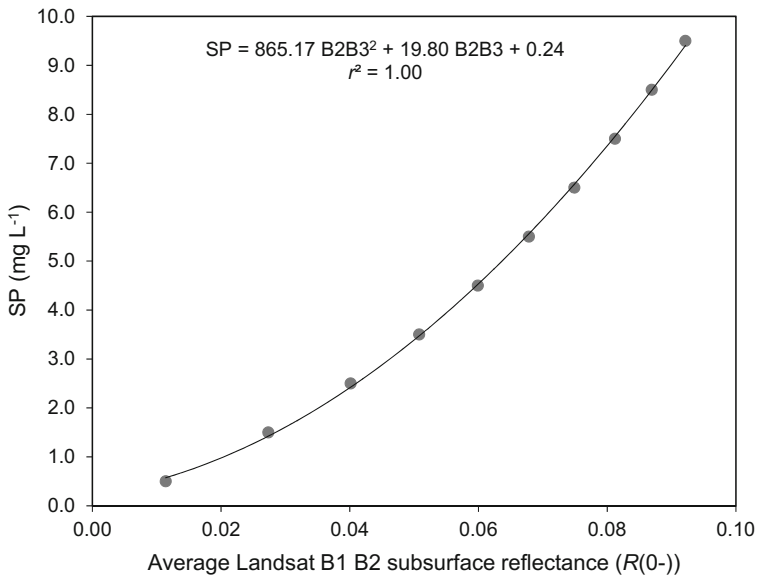


Fig. 4 The analytical relationship between suspended particle (SP) concentrations as a function of the average of Landsat B1 and B2 subsurface irradiance reflectance (B2B3) (*R*(0−)). A polynomial relationship (*black line*) is used to approximate the analytical relationship together with the r^2 value. This function was use to predict SP concentration from average B2B3 Landsat *R*(0−)

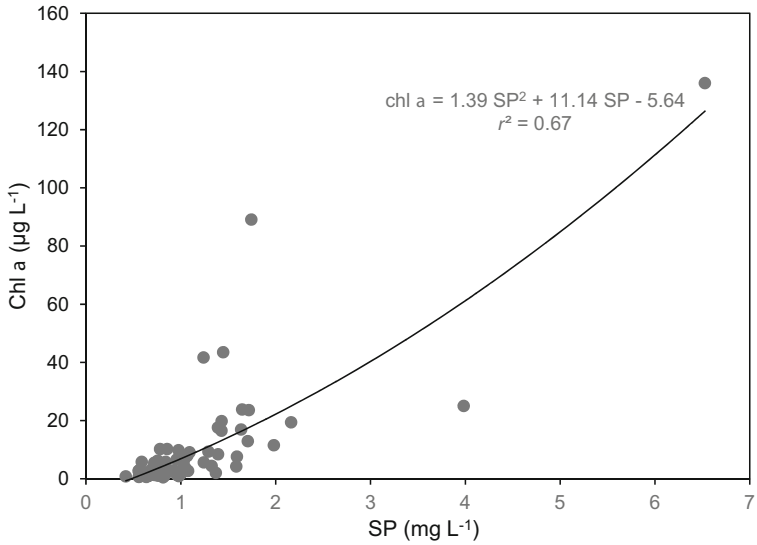


Fig. 5 Suspended particle concentration (SP) (mg L^{-1}) derived from the semi-analytical model plotted against in situ chlorophyll (chl) *a* (µg L^{-1}). A polynomial fit was used to estimate chl *a* from modelled SP

Landsat bands and commonly used band ratios (Table 2). The highest r^2 value was for ln(chl) vs. B3 (r^2=0.593). Chlorophyll *a* vs. B2 produced the second highest r^2 value of 0.582.

Of the semi-analytical algorithms applied to estimate chl *a*, the best-performing algorithm used B3, with r^2=0.68, and RMSE 10.69=µg L^{-1} and bias −2.44 µg L^{-1} (bias calculated as average absolute error) (Table 3, Fig. 6). The semi-analytical B2 had an RMSE 12.96 µg L^{-1}; however, it had a large bias of −6.02 µg L^{-1}, and the B1 algorithm produced a negative bias of −5.59 µg L^{-1}. Empirical algorithms for chl *a* estimation had slightly higher precision of chl *a* estimation than the semi-analytical algorithms. The empirical relationship for estimation of chl *a* from semi-analytically derived TSS had the lowest bias (0.01 µg L^{-1}). The chl *a* estimated from symbolic regression had the second lowest bias (−0.51 µg L^{-1}). The semi-analytical B3, chl-a-TSS and symbolic regression all had similar RMSE and r^2. In order to distinguish the best-performing algorithm, model performance was evaluated for these three algorithms for in situ chl *a* <5 µg L^{-1} (Table 4, Fig. 7). The symbolic regression algorithm between observed and estimated chl *a* had the lowest bias (1.07 µg L^{-1}) and the highest r^2 value (0.35). Based on these findings, the symbolic regression algorithm was applied to estimate chl *a* henceforth.

Influence of CDOM and SP on R(0−) and chl *a* retrieval

Running the bio-optical model with zero values of CDOM and chl *a*, and with SP concentrations ranging from 0 to 9 mg L^{-1}, revealed that increases in SP concentration increase *R*(0−) over all visible wavelengths (Fig. 8). The modelled effect of incremental increases in SP concentration on the accuracy of chl *a* estimation using the symbolic regression model and the B3 semi-analytical model is also large. Keeping chl *a* fixed at 30 µg L^{-1}, and increasing SP concentrations from 0 to 9 mg L^{-1} in steps of 0.2 mg L^{-1}, shows that increasing SP can result in positive errors in chl *a* estimation (Fig. 10b). However, the symbolic regression model has lower relative error than the B3 semi-analytical model. At SP concentration of 9 mg L^{-1}, the relative error in chl *a* using the symbolic regression model was 66.6 compared to 570.9 for the B3 semi-analytical model.

Running the bio-optical model with zero SP and chl *a*, and varying CDOM absorption from 0 to 1.8 m^{-1},

Fig. 6 Observed chlorophyll (chl) *a* (µg L^{-1}) (*y*) versus estimated chl *a* (*x*) using **a** semi-analytically modelled from B1 (SA-B1), **b** semi-analytically modelled from B2 (SA-B2), **c** semi-analytically modelled from B3 (SA-B3), **d** regression with semi-analytically estimated suspended particle concentration (chl a SP) and **e** using symbolic regression (Symbolic reg. chl *a*). Associated equations and statistics reported in Table 3

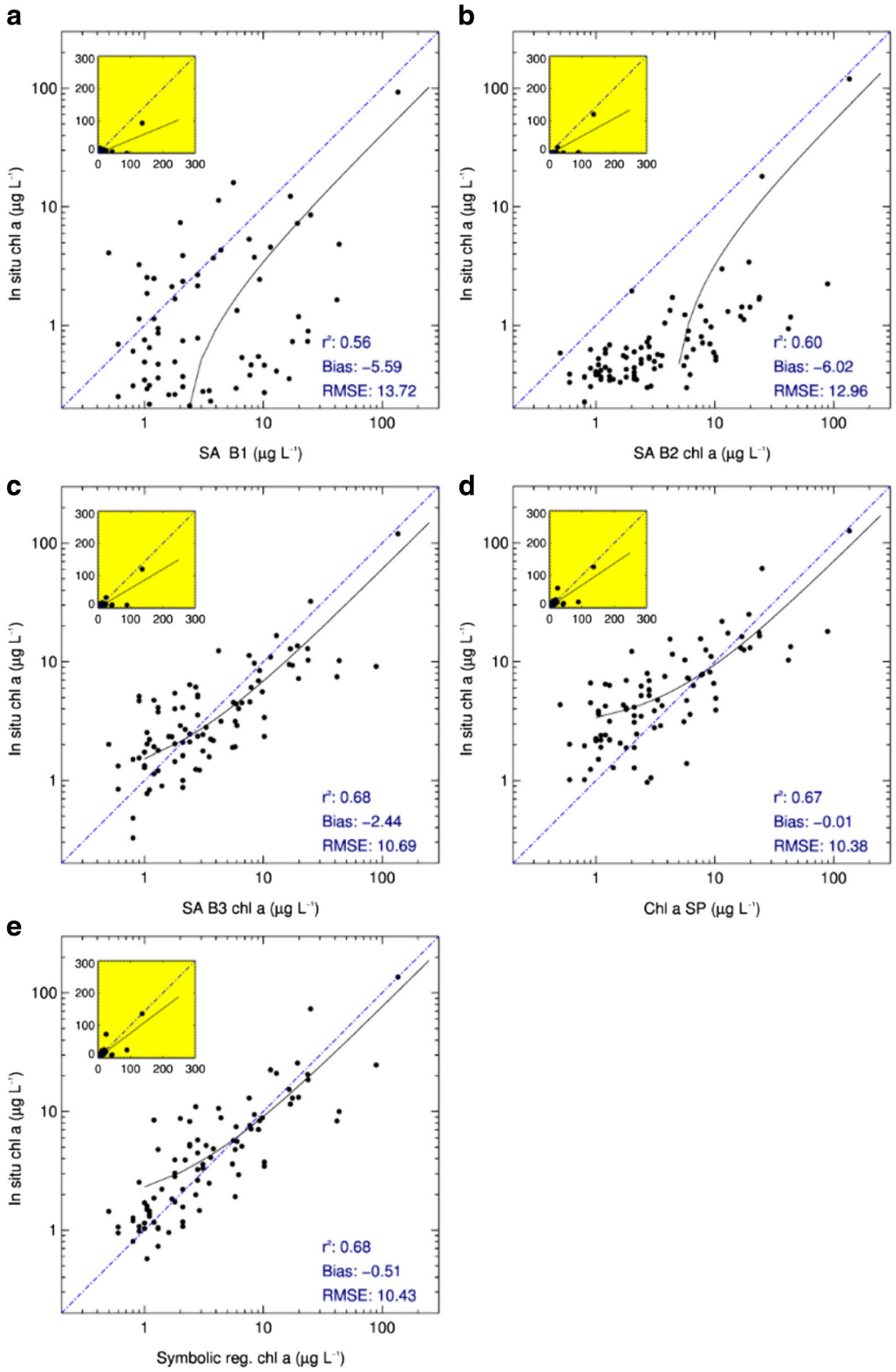

a
In situ chl a (µg L⁻¹)
SA B1 (µg L⁻¹)
r²: 0.56
Bias: −5.59
RMSE: 13.72
b
In situ chl a (µg L⁻¹)
SA B2 chl a (µg L⁻¹)
r²: 0.60
Bias: −6.02
RMSE: 12.96
c
In situ chl a (µg L⁻¹)
SA B3 chl a (µg L⁻¹)
r²: 0.68
Bias: −2.44
RMSE: 10.69
d
In situ chl a (µg L⁻¹)
Chl a SP (µg L⁻¹)
r²: 0.67
Bias: −0.01
RMSE: 10.38
e
In situ chl a (µg L⁻¹)
Symbolic reg. chl a (µg L⁻¹)
r²: 0.68
Bias: −0.51
RMSE: 10.43

Table 2 Linear regression relationships between in situ chlorophyll *a* and Landsat (n=88)

Parameter	Chl *a* r^2	ln(chl *a*) r^2
B1	0.105	0.076
B2	0.582	0.516
B3	0.504	0.594
B4	0.352	0.307
B1/B3	0.167	0.463
Ln(B1/B3)	0.234	0.512
ln(B1)	0.070	0.058
ln(B2)	0.408	0.565
ln(B3)	0.326	0.590
ln(B4)	0.269	0.310

All correlations are statistically significant at a significance level of 5 %

revealed that increases in CDOM absorption decrease $R(0-)$ (Fig. 9). Therefore, the presence of CDOM has the potential to introduce error and add negative bias in the estimation of chl *a* using both the semi-analytical and the symbolic regression algorithms. The modelled effect of CDOM absorption on the accuracy of the symbolic regression model and the B3 semi-analytical model is large. Keeping chl *a* fixed at 30 μg L^{-1}, and increasing CDOM absorption from 0 to 1.8 m^{-1} in 0.2 m^{-1} steps, shows that increasing CDOM can result in negative relative error in chl *a* estimates (Fig. 10a). However, the symbolic regression model has higher error. At CDOM absorption of 1.8 m^{-1}, the relative error in chl *a* using the symbolic regression model was −27.4, compared to a value of −4.0 of the B3 semi-analytical model.

Satellite estimation of chl *a* concentration

In lakes Rotorua, Rotoehu and Rotoiti, monthly in situ chl *a* concentrations were compared to chl *a* derived from the symbolic regression algorithm (Fig. 11). Although a direct comparison is not possible due to the time difference between in situ sample and satellite overpass, some general features can be identified. In Lake Rotoma, satellite estimations were usually of a similar magnitude to observed values for each date of comparison, and the general trend of winter maximum and summer minimum chl *a* is reflected in the estimations (Fig. 11a).

In Lake Rotoehu, chl *a* derived from the symbolic regression model was 134 μg L^{-1} on 5 February 2000, 160 μg L^{-1} on 6 January 2001 and 73 μg L^{-1} on 25 February 2002. These values were substantially higher than the corresponding in situ data (Fig. 11b). Although these values are very high, the maximum chl *a* measured in this lake was 197.7 μg L^{-1} on 26 April 1993. Surface distributions of chl *a* derived from symbolic regression are highly heterogeneous in Lake Rotoehu and appear to be influenced by wind speed and direction. High concentrations appear to occur in downwind locations in several cases (e.g. Fig. 12d–f). Lake-wide statistics (excluding a 100-m buffer zone from the shoreline) give a maximum-estimated chl *a* of 616 μg L^{-1} on 5 February 2000 (Fig. 12a), located in the southeast of the lake c. 500 m from the shoreline. This date also had the highest range of chl *a* values of 604 μg L^{-1}. The highest chl *a* concentrations were observed when wind speed was less than 4 m s^{-1} (e.g. Fig. 12a–f).

Table 3 Comparison of semi-analytical and empirical models for estimating chlorophyll (chl) *a* including semi-analytically modelled from B1 (SA B1), semi-analytically modelled from B2 (SA B2), semi-analytically modelled from B3 (SA B3), chl *a* estimated from regression with semi-analytically estimated suspended particle concentration (chl *a* SP) and using symbolic regression (Symbolic reg. chl *a*)

Algorithm	SA B1 chl *a*	SA B2 chl *a*	SA B3 chl *a*	Chl *a* SP	Symbolic reg. chl *a*
Equation	Chl a=0.0004 $e^{(278.18\ B2)}$	Chl a=0.1108 $e^{(68.265\ B2)}$	Chl a =5.98e+21 $B3^{15}$+ 23,684.19 $B3^2$	Chl a=1.39 SP^2+ 11.14 SP−5.64	Chl a=1.19+40,796.96 B2 B3−29,211.69 B1 B3
Intercept	−0.73	−2.26	0.93	2.71	1.57
Slope	0.42	0.55	0.60	0.67	0.75
r	0.75	0.78	0.83	0.82	0.82
r^2	0.56	0.60	0.68	0.67	0.68
p	0.000000	0.000000	0.000000	0.000000	0.000000
Bias	−5.59	−6.02	−2.44	0.01	−0.51
RMSE	13.72	12.96	10.69	10.38	10.43

Table 4 Comparison of semi-analytical and empirical models for estimating chlorophyll (chl) *a* <5 μg L^{-1} including semi-analytically modelled from B3 (SA B3), chl *a* estimated from regression with semi-analytically estimated suspended particle concentration (chl *a* SP) and using symbolic regression (Symbolic reg. chl *a*)

Algorithm	SA B3 chl *a*	Chl *a* SP	Symbolic reg. chl *a*
Intercept	1.17	0.95	−0.04
Slope	0.77	1.60	1.59
r	0.38	0.53	0.59
r^2	0.15	0.29	0.35
p	0.00345445	1.84774e−005	1.54972e−006
Bias	0.74	2.08	1.07
RMSE	1.97	3.27	2.45

In Lake Rotorua, the chl *a* derived from symbolic regression may have been affected by the presence of SP (Fig. 11c). This lake is large, shallow and wind-exposed where particle resuspension occurs. Figure 8 demonstrates that additions of SP increase $R(0-)$ at all visible wavelengths, which introduces the potential for error when using empirical algorithms. It should be noted that there was no relationship between chl *a* and hourly average wind speed at 1000 h on the overpass date. However on 4 May 2009, high wind speeds of 5.7 m s^{-1} occurred concurrently with chl *a* estimates of 77.7 μg L^{-1}. On this date concentrations of chl *a* in other Rotorua lakes were low or moderate (<15 μg L^{-1}), suggesting that SP may have been responsible for the elevated values of modelled chl *a*.

Chlorophyll *a* derived from the symbolic regression algorithm showed large spatial variations, both within and between lakes on 25 January 2002. Symbolic regression estimated within-lake spatial variation was high in lakes Rotoiti (0.5–198 μg L^{-1}) and Rotoehu (32–222 μg L^{-1}) (Fig. 13). In Lake Rotorua, chl *a* was elevated near the lake edge (affecting up to 10 pixels from the lake edge and corresponding to a depth of up to 2.5 m). Additional reflectance from bottom reflection or elevated levels of SP in extensive shallow areas in this lake may have contributed in part to the spatial variation of chl *a*. In other lakes, elevated chl *a* also often occurred near the lake edge; however, it generally affected only 1 or 2 pixels. In-lake variation of surface chl *a* derived from symbolic regression was highest for lakes Rotorua, Rotoiti and Rotoehu. In contrast, nearby lakes Rotoma and Okataina had low in-lake variation of chl *a* concentrations on all dates. Figure 14 shows chl *a* derived using the symbolic regression model on selected dates when there was high spatial variation of chl *a* within lakes. On 5 February 2000 (Fig. 14a), estimated chl *a* concentrations were >200 μg L^{-1} in lakes Rotorua and Rotoehu, while concentrations in Lake Rotoiti, including Okawa Bay, were <40 μg L^{-1}. On 6 January 2001 (Fig. 13b), chl *a* concentrations were >200 μg L^{-1} at site R1 (Okawa Bay) in Lake Rotoiti and in Lake Rotoehu, while in the main body of Lake Rotorua concentrations were <40 μg L^{-1}. On 19 March 2004 (Fig. 14c), there was high spatial variation of chl *a* in Lake Rotorua, particularly on the western shore. On 19 February 2011 (Fig. 14d), relatively high concentrations of chl *a* were observed at Lake Rotoiti site R1 compared with the main basin of Lake Rotoiti.

Discussion

In this study, a symbolic regression algorithm was used to estimate chl *a* from Landsat ETM+ subsurface irradiance reflectance. The semi-analytical algorithm using Landsat B3 demonstrated similar accuracy to the symbolic regression model over the entire dataset; however, the symbolic regression algorithm performed best for chl *a* concentrations <5 μg L^{-1}. At chl *a* concentrations <5 μg, however, only 30 % of the variation of in situ chl *a* concentrations was explained. However, we demonstrated the ability of the symbolic regression model to reproduce seasonal trends within in situ data for oligotrophic Lake Rotoma. Hindcasting of chl *a* concentration displayed large inter- and intra-lake variations that tended to be greater than those documented by water quality monitoring programmes at selected sites on the Rotorua lakes (Burns et al. 2009). We demonstrated the ability of the derived symbolic regression model to estimate chl *a* over landscape scales and long time periods, potentially capturing greater variability, which opens up new possibilities for applying remote sensing to lakes with sparse or no in situ monitoring history.

The symbolic regression model used all three Landsat visible bands. The ability of symbolic regression to discover both model form (e.g. mathematical operators and constants) and parameters resulted in a novel algorithm. The use of combinations of spectral bands, while increasing correlation coefficient values for relationships with chl *a*, can produce results that are difficult to interpret based on bio-optical theory (Dekker 1993), as noted here. However, some general features can be identified. Increasing concentrations of chl *a* are associated with a decrease in reflectance at blue wavelengths (B1)

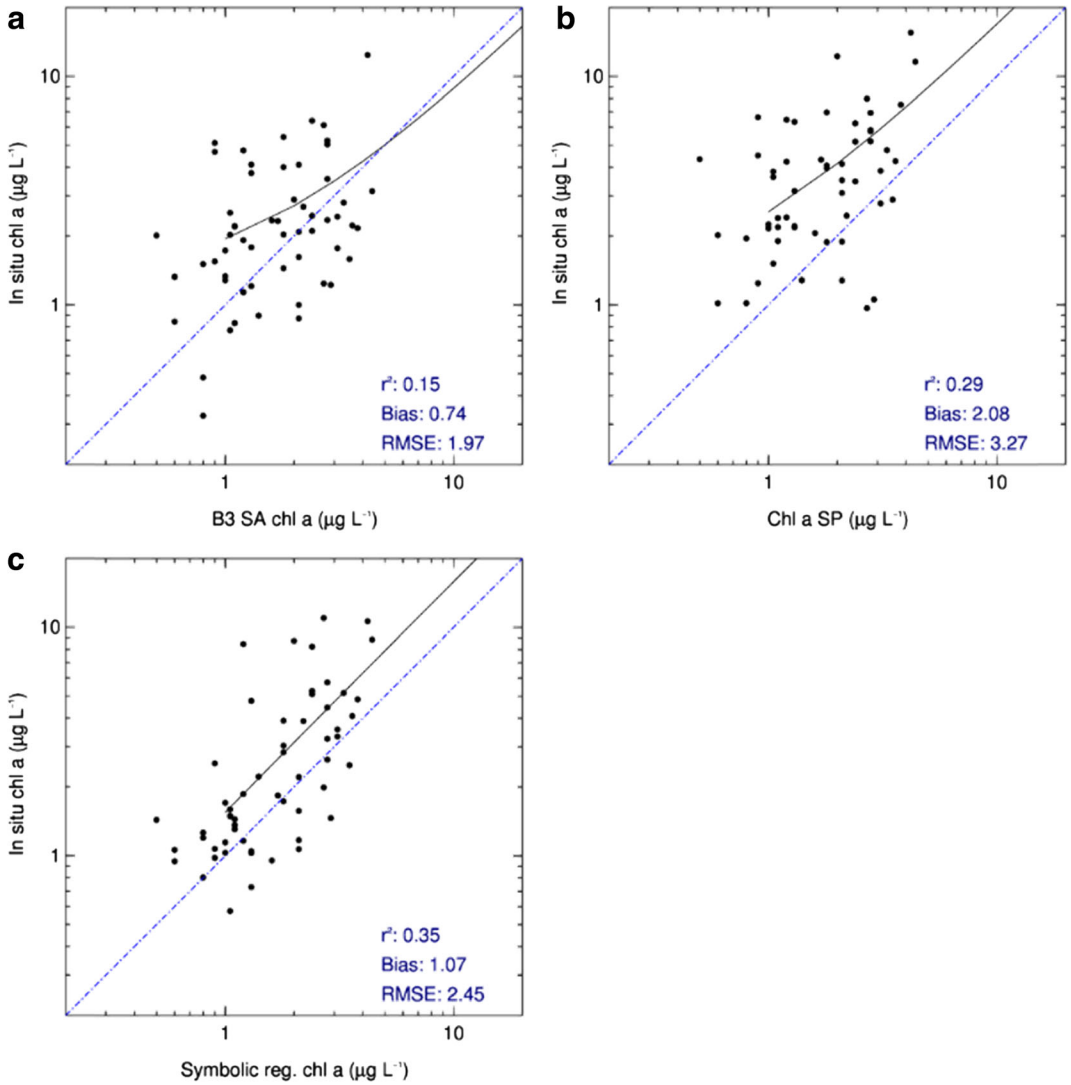


Fig. 7 Observed chlorophyll (chl) *a* (µg L^{-1}) (*y*) versus estimated chl *a* (*x*) for chl *a* concentration <5 µg L^{-1} using **a** semi-analytically modelled from B3 (SA-B3), **b** regression with semi-analytically estimated suspended particle concentration (chl *a* SP) and **c** using symbolic regression (Symbolic reg. chl a). Associated equations and statistics reported in Table 4

corresponding to a local pigment absorption maximum (Gitelson et al. 1993, 1996; Han and Jordan 2005). However, empirical and semi-analytical relationships established in this study indicate that reflectance at blue wavelengths (B1) increases with increasing chl *a* concentrations. We attribute this to efficient scattering of cyanobacteria species present in the Rotorua lakes, such as *M. aeruginosa* (Matthews and Bernard 2013; Zhou et al. 2012).

Bio-optical modelling indicated a positive correlation of B2 and B3 to chl *a* concentrations. Band 2 contains a local reflectance peak at 570–600 nm corresponding to a chl *a* and *b* absorption minimum. The Landsat B3 wavelength range includes a local reflectance peak near 650 nm which is caused by a combination of decreasing chl *b* and phycocyanin absorption and relatively low chl *a* absorption (in comparison to the chl *a* absorption maximum at 662 nm), and low absorption relative to scattering at these wavelengths (Dekker et al. 2002a). In addition, the B3 waveband captures the initial part of the reflectance peak centred at 690–720 nm (due to increased algal scattering). Chlorophyll fluorescence is often erroneously attributed to contributing to the 690–720-nm peak in eutrophic waters; however, it was not

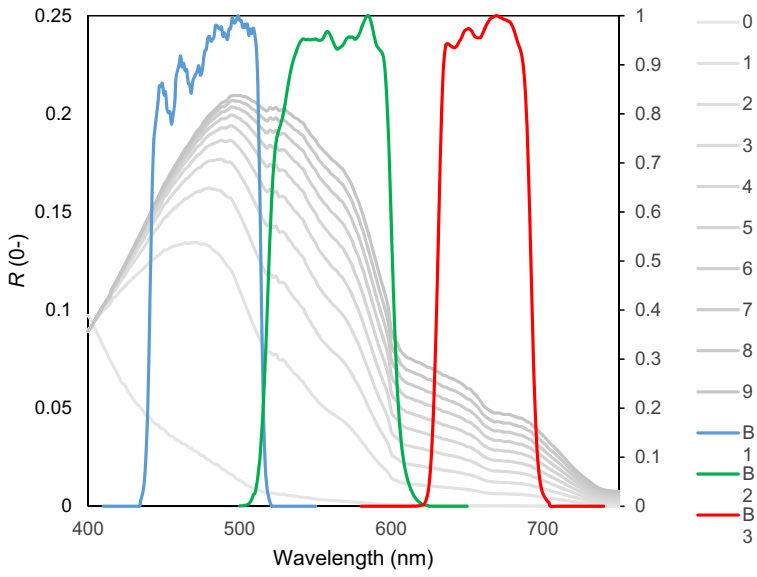


Fig. 8 Modelled subsurface irradiance reflectance ($R(0-)$) for varying suspended particle concentrations ranging from 0 to 9 mg L^{-1}, with zero coloured dissolved organic matter absorption and chlorophyll *a* concentration. The relative spectral response of Landsat bands B1, B2 and B3 is overlaid

simulated in the bio-optical model in this study, and it has previously been deemed insignificant compared with contributions from elastic scattering (Dekker et al. 2002a; Gilerson et al. 2007). The Landsat B3 wavelength range includes part of the reflectance trough centred at 624 nm (caused by chl *a* and phycocyanin absorption) and the entire trough centred at 676 nm caused by chl *a* absorption (Dekker et al. 2002a). Increasing chl *a* concentrations, however, still leads to increased reflectance at these troughs, although at 676 nm, the sensitivity is minimal (Gitelson et al. 2008). The net result of these spectrally counteracting forces is that increasing concentrations of chl *a* are associated with increased $R(0-)$ in Landsat B3.

We demonstrated there is potential for error in estimated chl *a* induced from independently varying concentrations of CDOM or SP when applying the symbolic regression algorithm and the semi-analytical algorithm. Concentrations of CDOM relatively low in Lake Rotorua and Rotoiti (0.16–0.23 m^{-1}) but may be

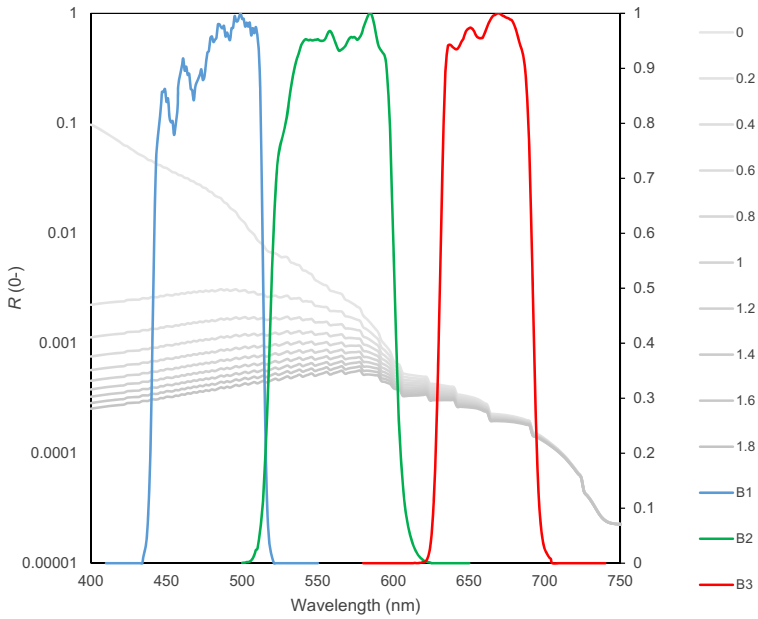


Fig. 9 Modelled subsurface irradiance reflectance ($R(0-)$) for varying coloured dissolved organic matter (CDOM) absorption ranging from 0 to 1.8 m^{-1}, zero suspended particle and chl *a* concentration. The relative spectral response of Landsat bands B1, B2 and B3 is overlaid

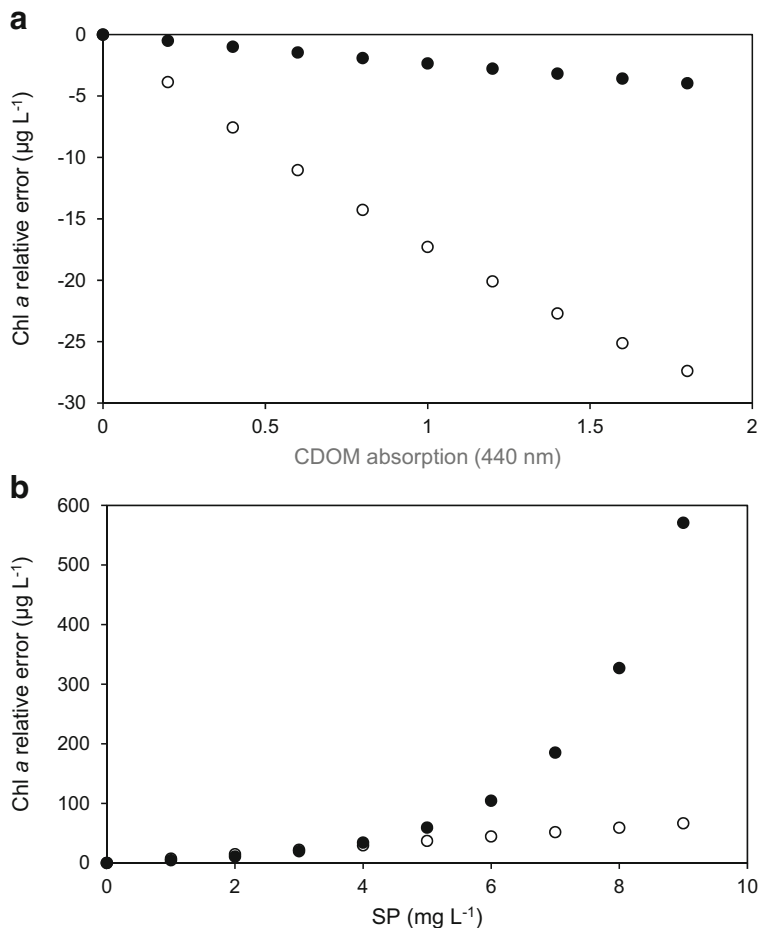


Fig. 10 Estimated chlorophyll (*chl*) *a* relative error using the semi-analytical B3 (*closed circles*) and symbolic regression (*open circles*), using simulated reflectance, with **a** varying CDOM absorption and **b** varying suspended particle (SP) concentration

higher in other lakes with forest-dominated catchments such as Tikitapu. Further investigation is needed regarding the spectral slope and concentration of CDOM in the Rotorua lakes. Lakes with high CDOM would likely require the use of bands in the red and near-infrared wavelengths, to avoid high CDOM absorption wavelengths (Gitelson et al. 2008); therefore, in these lakes, the semi-analytical algorithm using Landsat B3 may be less prone to error. The presence of independently varying concentrations of SP poses a greater potential for error, however, in hindcasting chl *a* estimates. The bio-optical model shows that the addition of small amounts of SP (e.g. c. 1 mg L^{-1}) increases $R(0-)$ in Landsat B1-B3, which increases potential for error in chl *a* hindcasts. Our study demonstrated that the presence of moderate amounts of SP causes large errors in the estimation of chl *a*. However, the symbolic regression algorithm was less prone to errors than the semi-analytical B3 algorithm.

A potential solution to errors in chl *a* hindcasts induced from variations in other optically active components is the use of bio-optical algorithms that allow simultaneous estimation of CDOM absorption, SP and chl *a* concentration. However, the Landsat series of satellite sensors possess spectrally broad bands which do not allow for a stable inversion with a bio-optical model. Further research will be critical to address shortcomings in broadband-based algorithms, but the limitations of the Landsat sensor will ultimately limit the ability of component discrimination (Matthews 2011). A recent review (Matthews 2011) revealed that there is currently no operational algorithm for remote sensing of chl *a* in inland waters using Landsat. This shortcoming has also been attributed to sensor-specific limitations including a low signal-to-noise ratio (SNR) (Matthews 2011). The application of three-dimensional hydrodynamic-ecological models to estimate CDOM and SP concentrations (Pahlevan et al. 2012) may provide

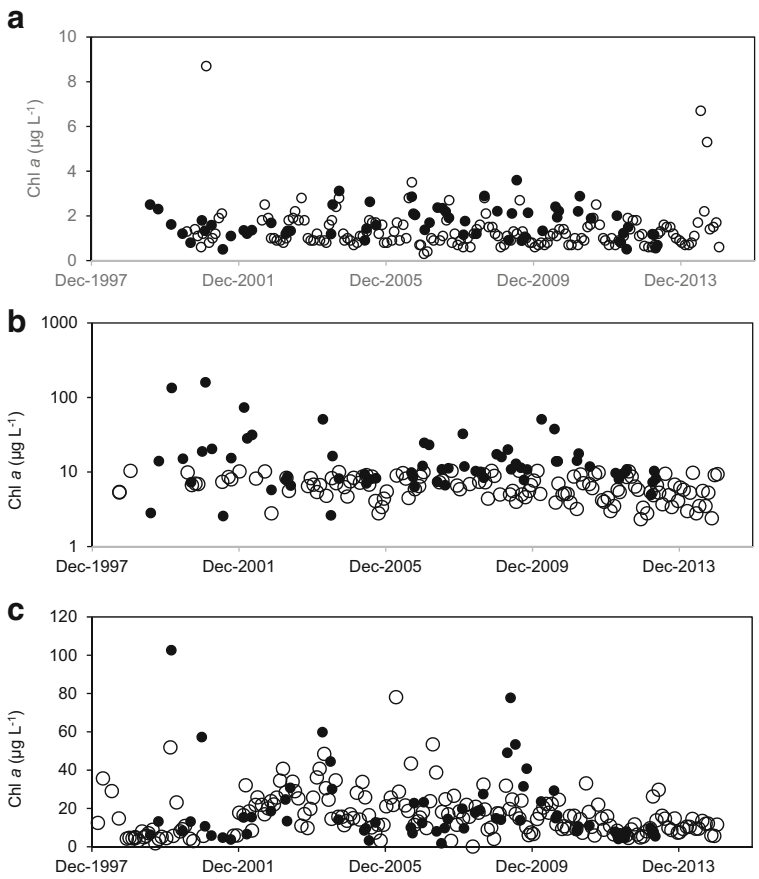


Fig. 11 Time series plots of estimated chlorophyll (chl) *a* (μg L^{-1}) from the symbolic regression (*closed circles*) and from all in situ data (*open circles*) for **a** Lake Rotoma, **b** Lake Rotoehu and **c** Lake Rotorua. Note that in most cases, dates of in situ and satellite sampling dates are not identical

sufficient temporal and spatial extrapolation in individual lakes to identify areas and times when CDOM and SP will most affect chl *a* estimates from Landsat sensors.

The simplified semi-analytical model using Landsat B3 performed similarly to the symbolic regression model (over the entire dataset) in terms of the accuracy of chl *a* hindcasts. The fact that the bio-optical model was parameterised with phytoplankton-specific IOPs or with values selected from the literature but for species commonly found in the Rotorua lakes is encouraging and may mean that this approach may be applied to other phytoplankton-dominated lakes where IOPs for phytoplankton species are known or have been previously documented. While bio-optical modelling is not applicable to simultaneous determination of independently varying optically active constituent concentrations based on Landsat data, it is useful in the simulation of the physical processes that result in varying $R(0-)$ with increasing chl *a* concentration, particularly in quantification of sources of error in empirical algorithms. In the Rotorua lakes, it is likely that there is both inter-lake and seasonal variability of phytoplankton-specific absorption and scattering coefficients based on the diverse assemblage of phytoplankton present (Paul et al. 2012). Phytoplankton-specific absorption, $a^*\phi(\lambda)$ depends not only on cell size but also on pigment composition and the packaging effect (Babin et al. 1993; Babin 2003; Bricaud 2004; Blondeau-Patissier et al. 2009). Therefore, more work is needed to quantify the effect of IOPs in the Rotorua lakes and apply specific IOPs that are representative of those found in the Rotorua lakes.

Variations in the specific scattering coefficient of phytoplankton are known to produce up to twofold differences in chl *a* concentrations retrieved from bio-optical models (Dierssen 2010). The backscattering ratio and the specific scattering coefficient of phytoplankton are a function of the size, physical structure and the outer coating of cells (Stramski et al. 2004). With increasing phytoplankton biomass, there is greater cell wall surface area, resulting in increased scattering

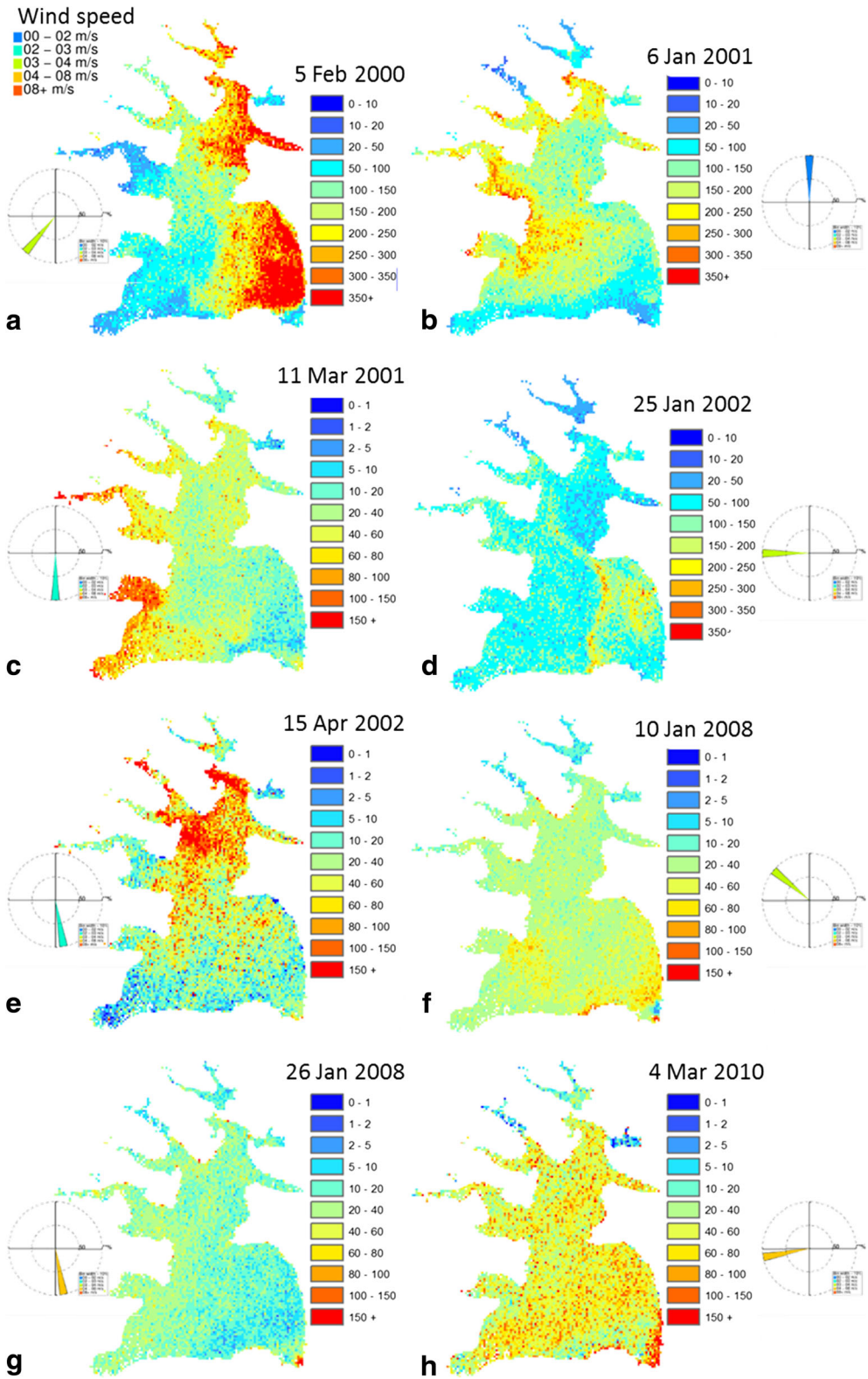


Fig. 12 Chlorophyll *a* concentration (μg L^{-1}) in Lake Rotoehu derived from the symbolic regression model

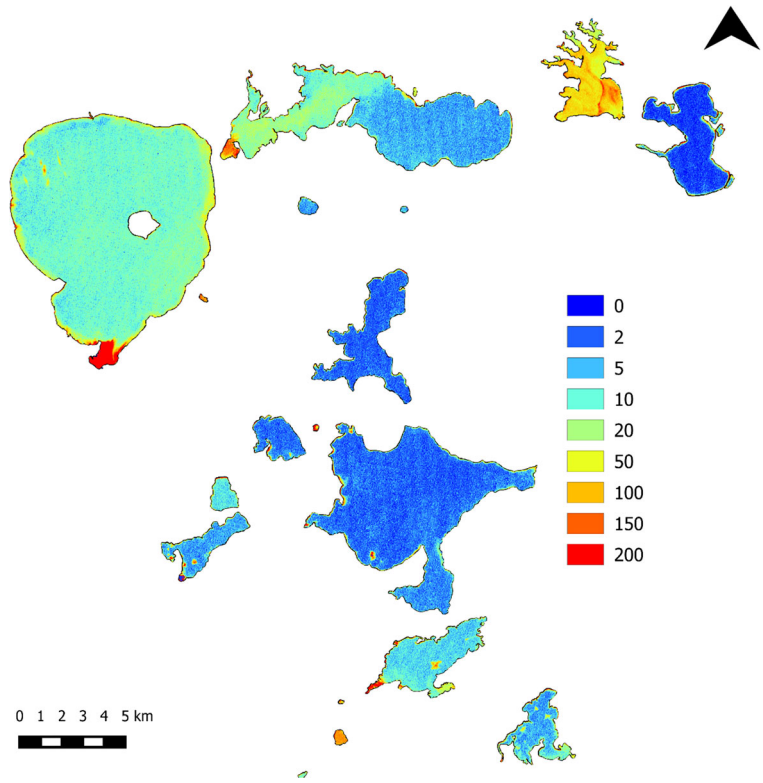


Fig. 13 Chlorophyll *a* (μg L^{-1}) derived from the symbolic regression model on 24 January 2002

(Yacobi et al. 1995). In addition, gas vacuoles present in some cyanobacteria have been found to be efficient at backscattering light (Dubelaar et al. 1987; Volten et al. 1998). Backscattering is inherently difficult to measure and there is considerable uncertainty in measured rates of backscattering by phytoplankton (Stramski et al. 2004). Variation in the backscattering ratio, and the specific scattering coefficient of phytoplankton that is not accounted for in bio-optical and empirical algorithms, has the potential to introduce error in estimates of chl *a*.

While remote sensing can provide information on temporal and spatial variations in chl *a* in surface waters, it does not capture variations deeper in the water column. It has been shown that in a homogenous water body, 90 % of reflectance emanating from beneath the water surface originates from water depths extending from immediately below the surface to the penetration depth where downwelling radiance decreases to 1/*e* of its surface value or c. 38 % (Gordon and McCluney 1975). This depth depends on the inherent optical properties of the optically active constituents and their concentrations, as well as the apparent optical properties of the ambient light field, and the layer from which the reflectance emanates can vary from as shallow as a few centimetres (Kuster 2004) to as deep as 60 m (Stramska and Stramski 2005).

Similar to what has been shown with oceanic distributions of chlorophyll (e.g. Hill and Zimmerman 2010), lakes Rotoma, Rotoiti and Tarawera have a deep chlorophyll maximum (DCM) which occurs when the lakes are stratified, and is closely associated with the thermocline depth, but is even more closely associated with the depth where light is 1 % of the surface value (i.e., the euphotic depth) (Hamilton et al. 2010). In Lake Rotoma, approximations of penetration depth of c. 5.5 m (PAR) from in situ measured light attenuation (Hamilton et al. 2010) have been recorded within the present study period, whereas in situ chl *a* concentrations are depth-integrated values from the surface to 17 m. Therefore, where chl *a* distributions are not homogenous over 0–17 m, remote sensing will not be representative of the in situ sample and estimates of chl *a* will not capture the DCM which has been recorded at depths c. 35 m in Lake

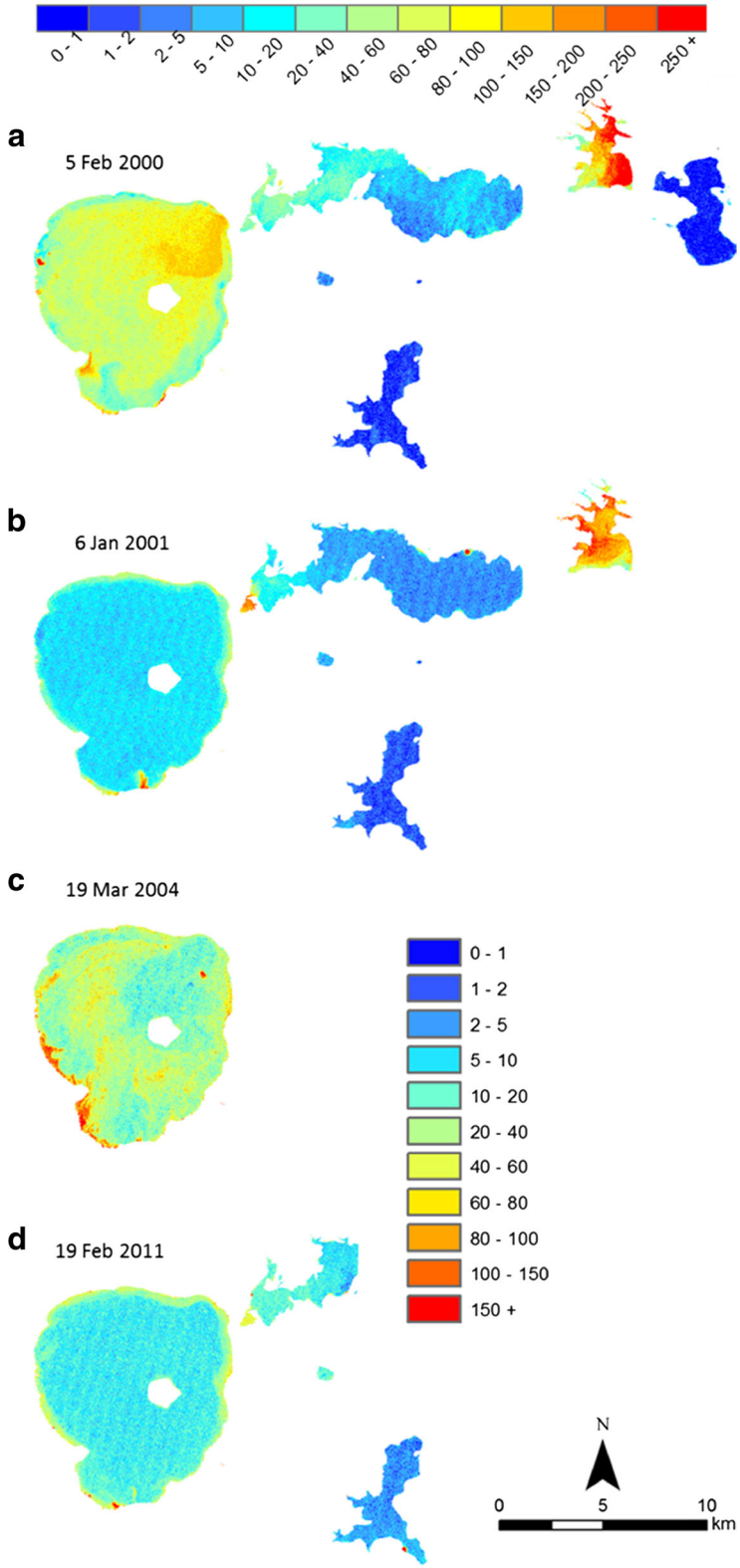


Fig. 14 Chlorophyll *a* (μg L^{-1}) derived from the symbolic regression model in north-western Rotorua lakes on selected dates with high spatial variation. Note that some lakes were removed due to cloud cover

Rotoma. For lakes with DCMs, the increasing development and deployment of real-time autonomous monitoring buoys with vertical chlorophyll fluorescence profiling capabilities is needed for effective monitoring of vertical chlorophyll distributions.

The use of a single station to monitor chl *a* in situ has been found to both over- and underestimate concentrations compared with remotely sensed chl *a* averaged over the lake area (Kallio et al. 2003; Kuster 2004). In Lake Rotehu, for example, complex spatial and temporal variations occur where the interaction of high cyanobacterial biomass can combine with temporary stratification events where buoyant cyanobacterial cells tend to aggregate at the water surface over durations of >2–3 h during calm conditions (Hamilton et al. 2015). The representativeness of grab samples from lakes can therefore be compromised by these strong gradients of chl *a* (Kutser et al. 2008). Remote sensing provides an opportunity to understand some of these horizontal complexities and, when combined with other tools such as three-dimensional hydrodynamic-ecological models (Chen et al. 2004) and real-time remote monitoring of chl *a* fluorescence (Hamilton et al. 2015) or phytoplankton-related IOPs (Babin et al. 2005), could provide valuable information for a deeper understanding of variations in phytoplankton biomass in all spatial and temporal domains.

Lake Rotoehu has high intra-lake variability of chl *a* based on the analysis of remotely sensed images. This lake has complex dendritic morphology, and frequent algal blooms give rise to the most heterogeneous distributions of chl *a* of all the Rotorua lakes, with a maximum satellite-estimated intra-lake range of 604 μg L^{-1} recorded in this study. Cyanobacteria blooms are commonly observed in summer (Burns et al. 2009), and satellite-estimated concentrations of chl *a* were often higher in downwind locations, as noted in other studies (e.g. Webster and Hutchinson 1994; Oliver and Ganf 2000; Oliver et al. 2012). The highest chl *a* concentrations were estimated during time periods when wind speeds were less than 4 m s^{-1}. Wind speeds above 2–3 m s^{-1} have been identified as critical to inducing entrainment of floating phytoplankton colonies into a turbulent surface layer (Webster and Hutchinson 1994), preventing the concentration of colonies on leeward shores. While the chl *a* values estimated via remote sensing were very high, other studies have found blooms containing *M. aeruginosa* can have chl *a* ranging from 70 to 1503 μg L^{-1} (Matthews and Bernard 2013). In Lake Rotoiti, chl *a* concentrations of up to 765 μg L^{-1} have been recorded in Te Weta Bay, corresponding to 430,000 cells ml^{-1} of *Microcystis* spp. (Wood et al. 2006).

The automated operational procedure developed in this study for remote sensing of chl *a* in the Rotorua lakes is applicable to other lakes; however, the symbolic regression model and the optimised B3 semi-analytical model are likely to be specific to the Rotorua lakes. The automation of procedures for retrieving chl *a* allowed for processing of large amounts of data and encompassed many operations which enabled a novel comparison of empirical and semi-analytical algorithms over a time series of images. However, as discussed above, the potential for error in chl *a* estimates must be carefully considered Landsat-based remote sensing of chl *a* still provides a valuable method for mapping the distribution of phytoplankton at a landscape scale, providing synoptic 'snapshots' and a cost-effective solution for examining temporal trends in water quality across a large number of lakes for which it may not be feasible to achieve complete coverage using a ground-based monitoring programme.

Acknowledgments Funding was provided by the Bay of Plenty Regional Council (BOPRC) and the Ministry of Business, Innovation and Employment (contract UOWX0505). This work benefited from participation in the Global Lakes Ecological Observatory Network (GLEON). We thank Bay of Plenty Regional Council for providing the measured data for water quality variables, in particular Paul Scholes, Glenn Ellery and Gareth Evans. Dr Matt Pinkerton (National Institute of Water and Atmospheric Research, New Zealand) provided technical guidance. Dr Hirokazu Yamamoto (Advanced Industrial Science and Technology, Japan) gave valuable feedback on atmospheric correction calculations. Richard Lamont (UOW) compiled 6sv for Windows.

References

Allan, M. G., Hamilton, D. P., Hicks, B. J., & Brabyn, L. (2011). Landsat remote sensing of chlorophyll a concentrations in central North Island lakes of New Zealand. *International Journal of Remote Sensing, 32*(7), 2037–2055.

Babin, M. (2003). Variations in the light absorption coefficients of phytoplankton, nonalgal particles, and dissolved organic matter in coastal waters around Europe. *Journal of Geophysical Research*, *108*(3211).

Babin, M., Therriault, J.-C., Legendre, L., & Condal, A. (1993). Variations in the specific absorption coefficient for natural

phytoplankton assemblages: impact on estimates of primary production. *Limnology and Oceanography, 38*, 154–177.

Babin, M., Cullen, J., & Roesler, C. (2005). New approaches and technologies for observing harmful algal blooms. *Oceanography, 18*, 210–227.

Blondeau-Patissier, D., Brando, V. E., Oubelkheir, K., Dekker, A. G., Clementson, L. A., & Daniel, P. (2009). Bio-optical variability of the absorption and scattering properties of the Queensland inshore and reef waters, Australia. *Journal of Geophysical Research, 114*(C05003).

Bricaud, A. (2004). Natural variability of phytoplanktonic absorption in oceanic waters: influence of the size structure of algal populations. *Journal of Geophysical Research*, 109, C11010.

Bricaud, A., & Morel, A. (1981). Theoretical results concerning light absorption in a discrete medium, and application to specific absorption of phytoplankton. *Deep Sea Research. Part A. Oceanographic Reseach Papers, 28*, 1375–1393.

Brivio, P. A., Giardino, C., & Zilioli, E. (1997). The satellite derived optical information for the comparative assessment of lacustrine water quality. *Science of the Total Environment, 196*, 229–245.

Bukata, R. P., Jerome, J. H., Kondratyev, K. Y., & Pozdnyakov, D. V. (1995). *Optical properties and remote sensing of inland and coastal waters* (p. 362). Boca Raton: CRC Press.

Burns, N., McIntosh, J., & Scholes, P. (2005). Strategies for managing the lakes of the Rotorua District, New Zealand. *Lake and Reservoir Management, 21*, 61–72.

Burns, N., McIntosh, J., & Scholes, P. (2009). Managing the lakes of the Rotorua District, New Zealand. *Lake and Reservoir Management, 25*, 284–296.

Chen, C., Wang, L., Ji, R., Budd, J. W., Schwab, D. J., Beletsky, D., Cotner, J. (2004). Impacts of suspended sediment on the ecosystem in Lake Michigan: a comparison between the 1998 and 1999 plume events. *Journal of Geophysical Research*, *109*, C10S11

Davies-Colley, R. J., & Vant, W. N. (1987). Absorption of light by yellow substance in freshwater lakes. *Limnology and Oceanography, 32*, 416–425.

Dekker, A. G. (1993). *Detection of optical water quality parameters for eutrophic waters by high resolution remote sensing*. Amsterdam: Free University.

Dekker, A. G., & Peters, S. W. M. (1993). The use of the thematic mapper for the analysis of eutrophic lakes: a case study in the Netherlands. *International Journal of Remote Sensing, 14*(5), 799–821.

Dekker, A. G., Hoogenboom, H. J., Goddijn, L. M., & Malthus, T. J. M. (1997). Relation between inherent optical properties and reflectance spectra in turbid inland waters. *Remote Sensing Reviews, 15*, 59–74.

Dekker, A. G., Peters, S. W. M., Vos, R., & Rijkeboer, M. (2001a). Remote sensing for inland water quality detection and monitoring. In A. van Dijk & M. G. Bos (Eds.), *GIS and remote sensing techniques in land- and water management*. Netherlands: Kluwer Academic Publishers.

Dekker, A. G., Vos, R. J., & Peters, S. W. M. (2001b). Comparison of remote sensing data, model results and in situ data for total suspended matter (TSM) in the southern Frisian lakes. *The Science of the Total Environment, 268*, 197–214.

Dekker, A. G., Brando, V. E., Anstee, J. M., Pinnel, N., Kutser, T., Hoogenboom, H. J., & Malthus, T. J. (2002a). Imaging spectrometry of water. In F. van der Meer & S. M. de Jong (Eds.), *Imaging spectrometry: basic principles and prospective applications* (pp. 307–359). Dordrecht: Kluwer.

Dekker, A. G., Voss, R. J., & Peters, S. W. M. (2002b). Analytical algorithms for lake water TSM estimation for retrospective analysis of TM and SPOT sensor data. *International Journal of Remote Sensing, 23*, 15–35.

Devred, E., Sathyendranath, S., Stuart, V., & Platt, T. (2011). A three component classification of phytoplankton absorption spectra: application to ocean-color data. *Remote Sensing of Environment, 115*, 2255–2266.

Dierssen, H. M. (2010). Perspectives on empirical approaches for ocean color remote sensing of chlorophyll in a changing climate. *Proceedings of the National Academy of Sciences of the United States of America, 107*, 17073–17078.

Dubelaar, G. B., Visser, J. W., & Donze, M. (1987). Anomalous behaviour of forward and perpendicular light scattering of a cyanobacterium owing to intracellular gas vacuoles. *Cytometry, 8*, 405–412.

Giardino, C., Pepe, M., Brivio, P., Ghezzi, P., & Zilioli, E. (2001). Detecting chlorophyll, Secchi disk depth and surface temperature in sub-alpine lake using Landsat imagery. *The Science of the Total Environment, 268*, 19–29.

Gilerson, A., Zhou, J., Hlaing, S., Ioannou, I., Schalles, J., Gross, B., & Ahmed, S. (2007). Fluorescence component in the reflectance spectra from coastal waters. Dependence on water composition. *Optics Express, 15*, 702–15721.

Gitelson, A., Garbuzov, G., Szilagyi, F., Mittenzwey, K., Karnieli, A., & Kaiser, A. (1993). Quantitative remote sensing methods for real-time monitoring of inland waters quality. *International Journal of Remote Sensing, 14*, 1269–1295.

Gitelson, A. A., Yacobi, Y. Z., Karnieli, A., & Kress, N. (1996). Reflectance spectra of polluted marine waters in Haifa Bay, Southeastern Mediterranean: features and application for remote estimation of chlorophyll concentration. *Israel Journal of Earth Sciences, 45*, 127–136.

Gitelson, A., Dallolmo, G., Moses, W., Rundquist, D., Barrow, T., Fisher, T., & Holz, J. (2008). A simple semi-analytical model for remote estimation of chlorophyll-a in turbid waters: validation. *Remote Sensing of Environment, 112*(9), 3582–3593.

Gordon, H. R., & McCluney, W. R. (1975). Estimation of the depth of sunlight penetration in the sea for remote sensing. *Applied Optics, 14*, 413–416.

Gordon, H. R., Brown, J. W., Brown, O. B., Evans, R. H., & Smith, R. C. (1988). A semianalytic radiance model of ocean color. *Journal of Geophysical Research, 93*(D9), 10909–10924.

Hamilton, D. P., O'Brien, K. R., Burford, M. A., Brookes, J. D., & McBride, C. G. (2010). Vertical distributions of chlorophyll in deep, warm monomictic lakes. *Aquatic Sciences, 72*, 295–307.

Hamilton, D., Carey, C., Arvola, L., Arzberger, P., Brewer, C., Cole, J., & Brookes, J. (2015). A Global Lake Ecological Observatory Network (GLEON) for synthesising high-frequency sensor data for validation of deterministic ecological models. *Inland Waters, 5*, 49–56.

Han, L., & Jordan, K. J. (2005). Estimating and mapping chlorophyll-a concentration in Pensacola Bay, Florida using Landsat ETM+ data. *International Journal of Remote Sensing, 26*, 5245–5254.

Hill, V. J., & Zimmerman, R. C. (2010). Estimates of primary production by remote sensing in the Arctic Ocean:

assessment of accuracy with passive and active sensors. *Deep-Sea Research Part I: Oceanographic Research Papers, 57*, 1243–1254.

Hoellein, T. J., Bruesewitz, D. A., & Hamilton, D. P. (2012). Are geothermal streams important sites of nutrient uptake in an agricultural and urbanising landscape (Rotorua, New Zealand)? *Freshwater Biology, 57*(1), 116–128.

Irish, R. R., Barker, J. L., Goward, S. N., & Arvidson, T. (2006). Characterization of the Landsat-7 ETM+ automated cloud-cover assessment (ACCA) algorithm. *Photogrammetric Engineering & Remote Sensing, 72*, 1179–1188.

Kallio, K., Koponen, S., & Pulliainen, J. (2003). Feasibility of airborne imaging spectrometry for lake monitoring—a case study of spatial chlorophyll a distribution in two meso-eutrophic lakes. *International Journal of Remote Sensing, 24*, 3771–3790.

Kloiber, S. M., Brezonik, P. L., & Bauer, M. E. (2002). Application of Landsat imagery to regional-scale assessments of lake clarity. *Water Research, 36*, 4330–4340.

Koponen, S. (2006). *Remote sensing of water quality for Finnish lakes and coastal areas*. Ph.D Thesis, Helsinki University of Technology: Finland.

Kostadinov, T. S., Siegel, D. A., & Maritorena, S. (2010). Global variability of phytoplankton functional types from space: assessment via the particle size distribution. *Biogeosciences, 7*, 4295–4340.

Kotchenova, S. Y., Vermote, E. F., Levy, R., & Lyapustin, A. (2008). Radiative transfer codes for atmospheric correction and aerosol retrieval: intercomparison study. *Applied Optics, 47*(13), 2215–2226.

Koza, J. R. (1994). Genetic programming as a means for programming computers by natural selection. *Statistics and Computing, 4*, 87–112.

Kuster, T. (2004). Quantitative detection of chlorophyll in cyanobacterial blooms by satellite remote sensing. *Limnology and Oceanography, 49*, 2179–2189.

Kutser, T., Metsamaa, L., & Dekker, A. G. (2008). Influence of the vertical distribution of cyanobacteria in the water column on the remote sensing signal. *Estuarine, Coastal and Shelf Science, 78*(4), 649–654.

Liley, J. B., & Forgan, B. W. (2009). Aerosol optical depth over Lauder, New Zealand. *Geophysical Research Letters, 36*(7), L07811.

Lillesand, T. M., Johnson, W. L., Deuell, R. L., Lindstrom, O. M., & Meisner, D. E. (1983). Use of Landsat data to predict the trophic state of Minnesota lakes. *Photogrammetric Engineering and Remote Sensing, 49*, 219–229.

Matthews, M. (2011). A current review of empirical procedures of remote sensing in inland and near-coastal transitional waters. *International Journal of Remote Sensing, 32*, 6855–6899.

Matthews, M. W., & Bernard, S. (2013). Using a two-layered sphere model to investigate the impact of gas vacuoles on the inherent optical properties of *Microcystis aeruginosa*. *Biogeosciences, 10*(12), 8139–8157.

Mobley, C. D. (1994). *Light and water: radiative transfer in natural waters*. San Diego: Academic Press.

Moisan, J. R., Moisan, T. A. H., & Linkswiler, M. A. (2011). An inverse modeling approach to estimating phytoplankton pigment concentrations from phytoplankton absorption spectra. *Journal of Geophysical Research, 116*, 1–16.

Morel, A. (1974). Optical properties of pure water and pure seawater. In N. G. Jerlov & E. Steemann Nielsen (Eds.), *Optical aspects of oceanography* (pp. 1–24). London: Academic.

Oliver, R., & Ganf, G. (2000). Freshwater blooms. In M. P. B. W (Ed.), *The ecology of cyanobacteria: their diversity in time and space*. (p. 149–194). Netherlands: Kluwer Academic Publishers.

Oliver, R., Hamilton, D., Brookes, J., & Ganf, G. (2012). Physiology, blooms and prediction of planktonic Cyanobacteria. In B. A. W (Ed.), *Ecology of cyanobacteria II*. (p. 155–194). Netherlands: Springer.

Olmanson, L. G., Bauer, M. E., & Brezonik, P. L. (2008). A 20-year Landsat water clarity census of Minnesota's 10,000 lakes. *Remote Sensing of Environment, 112*, 4086–4097.

Pahlevan, N., Garrett, A. J., Gerace, A. D., & Schott, J. R. (2012). Integrating Landsat-71 imagery with physics-based models for quantitative mapping of coastal waters near river discharges. *Photogrammetric Engineering & Remote Sensing, 78*, 1163–1174.

Paul, W. J., Hamilton, D. P., Ostrovsky, I., Miller, S. D., Zhang, A., & Muraoka, K. (2012). Catchment land use and trophic state impacts on phytoplankton composition: a case study from the Rotorua lakes' district, New Zealand. *Hydrobiologia, 698*, 133–146.

Pope, R.M., & Fry, E.S. (1997). Absorption spectrum (380–700 nm) of pure water. *II. Integrating Cavity Measurements*, 8710–8723.

Remer, L. A., Kaufman, Y. J., Tanre, D., Mattoo, S., Chu, D. A., Martins, J. V., & Holben, B. N. (2005). The MODIS aerosol algorithm, products, and validation. *Journal of the Atmospheric Sciences, 62*, 947–973.

Schmidt, M., & Lipson, H. (2009). Distilling free-form natural laws from experimental data. *Science, 324*, 81–85.

Stephens, S., Gibbs, M., Hawes, I., Bowman, E., & Oldman, J. (2004). *Ohau Channel Groynes*. NIWA Client Report: HAM2004-047. Prepared for Environment Bay of Plenty.

Stramska, M., & Stramski, D. (2005). Effects of a nonuniform vertical profile of chlorophyll concentration on remote-sensing reflectance of the ocean. *Applied Optics, 44*, 1735–1747.

Stramski, D., Boss, E., Bogucki, D., & Voss, K. J. (2004). The role of seawater constituents in light backscattering in the ocean. *Progress in Oceanography, 61*, 27–56.

Vant, W. N., & Davies-Colley, R. J. (1986). Relative importance of clarity determinants in lakes Okaro and Rotorua. *New Zealand Journal of Marine and Freshwater Research, 20*, 355–363.

Volten, H., Haan, J. D., & Hovenier, J. (1998). Laboratory measurements of angular distributions of light scattered by phytoplankton and silt. *Limnology and Oceanography, 46*, 1180–1197.

Webster, I. T., & Hutchinson, P. A. (1994). Effect of wind on the distribution of phytoplankton cells in lakes revisited. *Limnology and Oceanography, 39*, 365–373.

Wood, S. A, Briggs, L. R., Sprosen, J., G., Ruck, J. G., Wear, R. G., Holland, P. T., & Bloxham, M. (2006). Changes in

concentrations of microcystins in rainbow trout, freshwater mussels, and cyanobacteria in Lakes Rotoiti and Rotoehu. *Environmental Toxicology*, *21*(3), 205–222.

Yacobi, Y. Z., Gitelson, A., & Mayo, M. (1995). Remote sensing of chlorophyll in Lake Kinneret using high spectral-resolution radiometer and Landsat TM: spectral features of reflectance and algorithm development. *Journal of Plankton Research, 17*, 2155–2173.

Zhang, Y., Yin, Y., Wang, M., & Liu, X. (2012). Effect of phytoplankton community composition and cell size on absorption properties in eutrophic shallow lakes: field and experimental evidence. *Optics Express, 20*(11), 11882–11898.

Zhou, W., Wang, G., Sun, Z., Cao, W., Xu, Z., Hu, S., & Zhao, J. (2012). Variations in the optical scattering properties of phytoplankton cultures. *Optics Express, 20*(10), 11189–11206.