https://doi.org/10.5194/gmd-2024-174
Preprint. Discussion started: 20 November 2024

Geoscientific Model Development Discussions

Data-Informed Inversion Model (DIIM): a framework to retrieve marine optical constituents in the BOUSSOLE site using a three-stream irradiance model

Carlos Enmanuel Soto López[1,2], Fabio Anselmi[1], Mirna Gharbi Dit Kacem[1,2], and Paolo Lazzari[2]

[1]Università degli Studi di Trieste, 31127, Italy
[2]Istituto nazionale di oceanografia e di geofisica sperimentale - OGS, Trieste, 34010, Italy

Correspondence: Carlos Enmanuel Soto López (carlos.soto362@gmail.com)

Abstract. Within the New Copernicus Capability for Trophic Ocean Networks (NECCTON) project, we aim to improve the current data assimilation system by developing a method for accurately estimating marine optical constituents from satellite-derived Remote Sensing Reflectance. We developed and compared two frameworks by implicitly inverting a semi-analytical expression derived from the classical Radiative Transfer Equation. First, we used a Bayesian estimation, which provided retrievals of the optical constituents along with their uncertainties. Moreover, using historical in-situ measurements together with a Markov Chain Monte Carlo (MCMC) algorithm to adjust the model parameters, we were able to reduce the root mean square Error (RMSE) between the retrieved data and in-situ observations. Second, we employed the Stochastic Gradient Variational Bayes (SGVB) framework to efficiently approximate the Maximum Posterior (MAP) estimates of the optical constituents while simultaneously finding the Maximum Likelihood Estimate (MLE) of the model parameters. This approach resulted in faster computations of the optical constituents compared to Bayesian estimations, with equivalent RMSE values between the retrieved data and in-situ observations. We showed that both, the MCMC and SGVB based algorithms, were able to find sets of optimal parameters, which, due to correlations between them, are not unique. We conclude that both methods are consistent with the Radiative Transfer Equation. The first method provides reliable uncertainty estimations, while the second offers a faster alternative to standard inversion techniques, making it suitable for inversion and model optimization problems where MCMC algorithms are intractable.

1 INTRODUCTION

Operational systems, like Copernicus, use satellite-derived data, combined with data assimilation techniques, to obtain estimates of the marine ecosystem status. Traditionally, the assimilated variable is the chlorophyll retrieved data; nowadays, state of the art biogeochemical models are progressively including refined bio-optical models able to simulate optical variables such as Remote Sensing Reflectance, enabling the direct assimilation of multispectral reflectance measured by satellite sensors.

In this work, we aim to derive a framework to estimate the ocean inherent optical properties (IOPs), such as absorption and scattering coefficients, from measurements of satellite-derived apparent optical properties (AOPs), like irradiance and Remote Sensing Reflectance. The IOPs are of interest in their own, as they carry key information about ecosystem variables, such

https://doi.org/10.5194/gmd-2024-174
Preprint. Discussion started: 20 November 2024

as chlorophyll, which can be used as indicators of the trophic condition of large marine areas (Longhurst et al., 1996). Most importantly, the framework is intended to be employed as a module in a data assimilation schemes (Bruggeman et al., 2023), within an operational model services, to perform Remote Sensing Reflectance assimilation in a coherent way, providing an aligned forward and inverse procedure.

The retrieval of the IOPs of water bodies from measurements of the AOPs, is referred to as the inverse problem of ocean optics. This is crucially important since directly measuring IOPs with an extended spacial coverage is very difficult (Gordon, 2002).

The first step to compute the IOPs is to establish the forward relationship between the AOPs and the IOPs. In this context, the AOPs are described as a function of the IOPs using the Radiative Transfer Equation (RTE). Due to the complexity of the RTE, this computation is carried out in simple scenarios, resulting in simplified equations that can be solved analytically. Other approaches involve using semi-analytical equations or empirical relations, where the latter are combined with simplified expressions of the RTE. The inverse problem is solved using these forward computations to estimate the IOPs either explicitly, by analytically inverting the forward process (Zaneveld, 1989; Leathers et al., 1999; Tao et al., 1994; McCormick, 1996; Stramska et al., 2000; Salama and Verhoef, 2015; Lazzari et al., 2024), or implicitly, by using an estimate of the IOPs in the forward process and then iteratively adjusting the IOP values to match measurements of the AOPs (Gordon and Boynton, 1997; Boynton and Gordon, 2000; Michalopoulou et al., 2009; Salama and Verhoef, 2015; Erickson et al., 2023).

In this work, we focused on an implicit inverse method, where the forward model is the bio-optical model presented in Dutkiewicz et al. (2015) and described in section 3.1. The IOPs from the bio-optical model are the absorption, scattering and backward scattering coefficients of four optical-constituents: water, chlorophyll-α (whose increase or decrease is associated with changes in the density of phytoplankton) Chromophoric Dissolved Organic Matter, and Non Algal Particles. We focused in finding the density of these optical constituents, since we estimate the former IOPs as linear combinations of the latter. The model also depends on ad hoc parameters, originally computed as part of empirical relations from different studies (Morel, 1974; Aas, 1987; Dutkiewicz et al., 2015; Mason et al., 2016; Álvarez et al., 2023). We will optimize these parameters utilizing historical in-situ observations.

We compared two different frameworks. The first one is a Bayesian estimation, where we used a linearization of the forward process for estimating the uncertainties of the optical constituents, and Markov Chain Monte Carlo (MCMC) (Chib and Greenberg, 1995; Andrieu and Thoms, 2008) for the uncertainty of the parameters. This approach is described in section 4.

The second approach is based on the Stochastic Gradient Variational Bayes (SGVB) framework, introduced by Kingma and Welling (2013), and described in section 4.5. Allows for the estimation of parameters while also learning an estimate of the posterior distribution of the optical constituents. The idea is to approximate the probability distribution of the optical constituents given the satellite-derived Remote Sensing Reflectance using a neural network. This is the same framework used to train generative models known as Variational Auto Encoders (VAE), which also have been used to solve inversion problems (Zhong et al., 2019, 2021; Zhao et al., 2023; Shmakov et al., 2024). Originally proposed to solve inversion problems for cases when the posterior distribution is intractable (practically impossible to compute), this framework provides a fast way of estimating optical constituents, which are consistent with the forward model, and the in-situ observations.

https://doi.org/10.5194/gmd-2024-174
Preprint. Discussion started: 20 November 2024


Geoscientific Model Development Discussions

We employed three data sources covering a time span from 2005 to 2012: a dataset of historical satellite-derived Remote Sensing Reflectance, a dataset from the Ocean–Atmosphere Spectral Irradiance Model (OASIM, used as boundary conditions for the bio-optical model (Gregg and Casey, 2009)) and a set of in-situ measurements from the BOUSSOLE buoy, located in the Ligurian basin of the northwestern Mediterranean Sea (coordinates 7.54°E, 43.22°N) (Antoine et al., 2008). The description of the different datasets is presented in section 2.

2 DATA ACQUISITION

2.1 Ocean–Atmosphere Spectral Irradiance Model (OASIM)

As explained in Gregg and Casey (2009), "the OASIM model is intended to provide the surface irradiance over the oceans with sufficient spectral resolution to support ocean ecology, biogeochemistry, and heat exchange investigations, and of sufficient duration to support inter-annual and decadal investigations." From this model, we used the surface downward direct irradiance, the surface downward scattered irradiance, the surface photosynthetic available radiation, and the sun zenith angle, all for the coordinates at the BOUSSOLE buoy (Antoine et al., 2008), at wavelengths equal to 412.5 nm, 442.5 nm, 490 nm, 510 nm and 555 nm.

2.2 Satellite-derived Remote Sensing Reflectance

We used a Level 3 product provided by the E.U. Copernicus Marine Service Information (CMEMS). This is a merge of Level 2 Remote Sensing Reflectance from different satellite sources, as explain in Colella et al. (2023). This product provides pre-processed Remote Sensing Reflectance with daily resolution, spacial resolution of one kilometer, at six different wavelengths: 412 nm, 443 nm, 490 nm, 510 nm, 555 nm and 670 nm. Due to the fact that for oligotrophic and mesotrophic water, the absorption of water for wavelengths higher than 555 nm is dominant over the other constituents (Lee et al., 2002), we focus our attention on the data with wavelength less or equal than 555 nm. The values at the wavelengths 412 nm and 443 nm were assumed to be the same as the values with wavelengths at 412.5 nm and 442.5 nm in order to match the values computed with the OASIM model.

2.3 In-situ observations

We will use three sets of in-situ observations: chlorophyll-α, particulate backward scattering coefficient and downward light attenuation coefficient, acquired from the BOUSSOLE buoy (Antoine et al., 2008).

The three sets of measurements had 15 minutes resolution. We used only measurements between 10:00 and 14:00 GMT as representative. After filtering the data with an absolute tilt higher or lower than 10 degrees and dose reported at a depth more than 2 m below the nominal values (4 m and 9 m, depending on the instrument of measurement), we proceeded to average the daily values. Due to the noise in the measurements, before averaging, the downward light attenuation coefficient data was

https://doi.org/10.5194/gmd-2024-174
Preprint. Discussion started: 20 November 2024

Geoscientific Model Development Discussions

filtered with an analog high pass filter, using the package SciPy (Virtanen et al., 2020) from the programming language Python (Van Rossum and Drake, 2009), also filtering the noise with a frequency less than 4 hours.

Due to low vertical variability, the measurements of chlorophyll-α and particulate backward scattering coefficient were considered as the values just above the water-air interface. The former one had measurements at wavelengths equal to 442 nm, 488 nm, 550 nm and 620 nm.

Due to the height vertical variability of the downward light attenuation coefficient, the measurements were considered to be at a depth of 9 m, with values at the wavelengths 412 nm, 442 nm, 490 nm, 510 nm, 555 nm, 560 nm, 665 nm, 670 nm, 681 nm.

For the same reasoning described in section 2.2, we only used the values less or equal than 555 nm. The values at the wavelengths 412 nm, 442 nm, 488 nm and 550 nm were assumed to be the same as the values with wavelengths at 412.5 nm, 442.5 nm, 490 nm and 555 nm in order to match the values computed with the OASIM model.

3 BIO-OPTICAL MODEL

We now describe the Bio-optical model (Aas, 1987; Ackleson et al., 1994; Dutkiewicz et al., 2015; Álvarez et al., 2023), which details the interaction of the radiance with different constituents in the sea, called optical constituents. In section 3.1 we present the model of the water-leaving radiance, based on the classical Radiative Transfer Model (Dutkiewicz et al., 2015). In section 3.2, we use this model to compute the theoretical Remote Sensing Reflectance (R_{rs}^{MODEL}) (Aas and Højerslev, 1999). The aim of the inversion problem is to use this model, named forward model, and satellite measurements, to retrieve optical constituents that are consistent with future observations, for this end, we used historical in-situ observations described in section 3.3.

3.1 Radiative Transfer Model

To simulate the water-leaving radiance, we followed Dutkiewicz et al. (2015), using a one-dimensional, three-stream radiance model, where the vertical component of the radiance over the water column is decomposed into three interacting components following the system of equations,

$$\begin{aligned}
\frac{dE_{\mathrm{dir}}(h,\lambda)}{dh} &= -\frac{a(\lambda)+b(\lambda)}{\cos\theta}E_{\mathrm{dir}}(h,\lambda),\\
\frac{E_{\mathrm{dif}}(h,\lambda)}{dh} &= -\frac{a(\lambda)+r_s b_b(\lambda)}{v_s}E_{\mathrm{dif}}(h,\lambda)+\frac{r_u b_b(\lambda)}{v_u}E_{\mathrm{u}}(h,\lambda)+\frac{b(\lambda)-r_d b_b(\lambda)}{\cos\theta}E_{\mathrm{dir}}(h,\lambda),\\
\frac{E_u(h,\lambda)}{dh} &= -\frac{r_s b_b(\lambda)}{v_s}E_{\mathrm{dif}}(h,\lambda)+\frac{a(\lambda)+r_u b_b(\lambda)}{v_u}E_u(h,\lambda)-\frac{r_d b_b(\lambda)}{\cos\theta}E_{\mathrm{dir}}(h,\lambda).
\end{aligned} \tag{1}$$

These three equations describe how the vertical direct irradiance $E_{\mathrm{dir}}(h,\lambda)$ is attenuated by absorption, with $a(\lambda)$ the total absorption coefficient, and scattered into downward $E_{\mathrm{dif}}(h,\lambda)$, and upward irradiance $E_{\mathrm{u}}(h,\lambda)$, $b(\lambda)$ the total scattering coefficient, $b_b(\lambda)$ the total backward scattering coefficient, r_d, r_s and r_u the effective scattering coefficients normalized with respect

https://doi.org/10.5194/gmd-2024-174
Preprint. Discussion started: 20 November 2024

Geoscientific Model Development Discussions

to the backward scattering coefficients, $\cos(\theta)$, v_s and v_u the average cosines of the irradiance components, θ the Sun zenith angle, h the depth, and λ the wavelength.

Following Dutkiewicz et al. (2015), the values for r_d, r_s, r_u, v_s and v_u are approximated as constants (see Table 2). See Dutkiewicz et al. (2015), appendix B, for a derivation starting from the classical radiative transfer equation. For previous studies where similar transfer models have been used, see Aas (1987); Ackleson et al. (1994); Salama and Verhoef (2015); Álvarez et al. (2023) and Lazzari et al. (2024).

The total absorption and scattering coefficients are modeled as,

$$\begin{aligned} a(\lambda) &= a_{\mathrm{w}}(\lambda) + a_{\mathrm{phy}}(\lambda)\mathrm{chla} + a_{\mathrm{CDOM}}(\lambda)\mathrm{CDOM} + a_{\mathrm{NAP}}(\lambda)\mathrm{NAP}, \\ b(\lambda) &= b_{\mathrm{w}}(\lambda) + b_{\mathrm{phy}}(\lambda)C + b_{\mathrm{NAP}}(\lambda)\mathrm{NAP}, \\ b_b(\lambda) &= b_{b,W}(\lambda) + b_{b,\mathrm{phy}}(\lambda)C + b_{b,\mathrm{NAP}}(\lambda)\mathrm{NAP}, \end{aligned} \tag{2}$$

with chla, NAP and CDOM the concentration of the optical constituents Chlorophyll-α, Non Algal Particles and Chromophoric Dissolved Organic Matter respectively; $a_{\mathrm{w}}(\lambda)$ is the water-specific absorption coefficient, $b_{\mathrm{w}}(\lambda)$ and $b_{b,\mathrm{w}}(\lambda)$ the water-specific scattering and backward scattering coefficients, $a_{\mathrm{phy}}(\lambda)$ the chlorophyll-specific absorption coefficient of phytoplankton, $b_{\mathrm{phy}}(\lambda)$ and $b_{b,\mathrm{phy}}(\lambda)$ the carbon-specific scattering coefficients of phytoplankton (see Table 1), C the carbon concentration, which is derived as a function of chlorophyll and irradiance (Geider et al., 1997), with the chla:C ratio represented as a sigmoid curve dependent on Photosynthetic Available Radiation (PAR), as

$$C = \mathrm{chla} / \left(\Theta^{0}_{\mathrm{chla}} \frac{e^{-(\mathrm{PAR}-\beta)/\sigma}}{1+e^{-(\mathrm{PAR}-\beta)/\sigma}} + \Theta^{\mathrm{min}}_{\mathrm{chla}} \right), \tag{3}$$

with $\Theta^{0}_{\mathrm{chla}}$, β, σ, $\Theta^{\mathrm{min}}_{\mathrm{chla}}$ constant parameters (see Table 2), $a_{\mathrm{CDOM}}(\lambda)$, $a_{\mathrm{NAP}}(\lambda)$ and $b_{\mathrm{NAP}}(\lambda)$ the mass-specific absorption and scattering coefficients for CDOM and NAP respectably (Álvarez et al., 2023), with the latter calculated as,

$$\begin{aligned} a_{\mathrm{CDOM}}(\lambda) &= d_{\mathrm{CDOM}} e^{-S_{\mathrm{CDOM}}(\lambda-450)}, \\ a_{\mathrm{NAP}}(\lambda) &= d_{\mathrm{NAP}} e^{-S_{\mathrm{NAP}}(\lambda-440)}, \\ b_{\mathrm{NAP}}(\lambda) &= e_{\mathrm{NAP}} \left(\frac{550}{\lambda} \right)^{f_{\mathrm{NAP}}}, \end{aligned} \tag{4}$$

with S_{CDOM}, d_{CDOM}, S_{NAP}, d_{NAP}, e_{NAP}, f_{NAP} constant parameters (see Table 2), and $b_{b,\mathrm{NAP}} = b_{r,\mathrm{NAP}} b_{\mathrm{NAP}}$, with $b_{r,\mathrm{NAP}}$ the backscattering-to-scattering ratio of NAP.

3.2 Remote Sensing Reflectance

We used the system of equations in Eq. (1), subject to the boundary conditions

$$E_{\mathrm{dir}}(0,\lambda) = E^{\mathrm{OASIM}}_{\mathrm{dir}}(0,\lambda), E_{\mathrm{dif}}(0,\lambda) = E^{\mathrm{OASIM}}_{\mathrm{dif}}(0,\lambda), E_{\mathrm{u}}(\infty,\lambda) = 0, \tag{5}$$

with $E^{\mathrm{OASIM}}_{\mathrm{dir}}(0,\lambda)$, $E^{\mathrm{OASIM}}_{\mathrm{dif}}(0,\lambda)$, the direct and diffuse downward irradiance on the surface of the ocean. For this work, we used the values from the OASIM model (Gregg and Casey, 2009). By assuming an infinitely deep and homogeneous column

https://doi.org/10.5194/gmd-2024-174
Preprint. Discussion started: 20 November 2024

Geoscientific Model Development Discussions
Open Access
EGU

Table 1. Parameters dependent of λ used for the Radiative Transfer Model evaluation, with the water-specific absorption coefficient from Mason et al. (2016), the water-specific scattering and backward scattering coefficients $b_{\mathrm{w}}(\lambda)$, $b_{b,\mathrm{w}}(\lambda)$ with values interpolated from Morel (1974), the phytoplankton-specific absorption coefficient $a_{\mathrm{phy}}(\lambda)$ an average of the values for picophytoplankton, nanophytoplankton, diatoms and dinoflagellates, collected from literature in Álvarez et al. (2023), and the carbon-specific scattering and backward scattering coefficients $b_{\mathrm{phy}}(\lambda)$ $b_{b,\mathrm{phy}}(\lambda)$ from Lazzari et al. (2024).

λ [nm]	$a_{\mathrm{w}}(\lambda)$ [m^{-1}]	$b_{\mathrm{w}}(\lambda)$ [m^{-1}]	$b_{b,w}(\lambda)$ [m^{-1}]	$a_{\mathrm{phy}}(\lambda)$ [$\mathrm{m}^2(\mathrm{mgChla})^{-1}$]	$b_{\mathrm{phy}}(\lambda)$ [$\mathrm{m}^2(\mathrm{mgC})^{-1}$]	$b_{b,\mathrm{phy}}(\lambda)$ [$\mathrm{m}^2(\mathrm{mgC})^{-1}$]
412.5	0.00271	0.00535	0.002674	0.034	0.02102	5.38E-05
442.5	0.00574	0.00437	0.002184	0.04	0.02022	5.18E-05
490	0.0146	0.00284	0.001421	0.028	0.02054	5.26E-05
510	0.033	0.00247	0.001234	0.018	0.0205	5.25E-05
555	0.06098	0.00167	0.000836	0.009	0.01907	4.88E-05

Table 2. Parameters independent of λ used for the Radiative Transfer Model evaluation, r_d, r_s, r_u, v_s, v_u, S_{CDOM}, d_{CDOM} from Dutkiewicz et al. (2015) who took them from Aas (1987), Θ^0_{chla}, $\Theta^{\mathrm{min}}_{\mathrm{chla}}$, σ β computed as an empirical model from data in the BOUSSOLE Site (Lazzari et al., 2024), S_{NAP}, d_{NAP}, e_{NAP}, f_{NAP} and $b_{r,\mathrm{NAP}}$ from Álvarez et al. (2023), Q_a and Q_b from Aas and Højerslev (1999), and, T and γ from Lee et al. (2002).

Parameter name	Symbol	Value from literature	Units
Normalized effective scattering coefficient for direct irradiation	r_d	1.0	-
Normalized effective scattering coefficient for downward radiation,	r_s	1.5	-
Normalized effective scattering coefficient for backward radiation,	r_u	3.0	-
Average cosine for downward scattered radiation	v_s	0.83	-
Average cosine for upward scattered radiation	v_u	0.4	-
-	Θ^0_{chla}	0.03	$\mathrm{mgChla(mgC)}^{-1}$
-	$\Theta^{\mathrm{min}}_{\mathrm{chla}}$	0.005	$\mathrm{mgChla(mgC)}^{-1}$
-	σ	20	$(\mathrm{mmol})\mathrm{m}^{-2}\mathrm{s}^{-1}$
-	β	500	$(\mathrm{mmol})\mathrm{m}^{-2}\mathrm{s}^{-1}$
CDOM mass-specific absorption at 450 nm	d_{CDOM}	0.015	$\mathrm{m}^2(\mathrm{mgCDOM})^{-1}$
CDOM mass-specific absorption spectral slope between 350 and 500 nm	S_{CDOM}	0.017	nm
NAP mass-specific absorption at 440 nm	d_{NAP}	0.0013	$\mathrm{m}^2(\mathrm{mgNAP})^{-1}$
NAP mass-specific absorption spectral slope between 350 and 500 nm	S_{NAP}	0.013	nm
NAP mass-specific scattering at 550 nm	e_{NAP}	0.02875	$\mathrm{m}^2(\mathrm{mgNAP})^{-1}$
-	f_{NAP}	0.5	-
Backscattering-to-scattering ratio of NAP	$b_{r,\mathrm{NAP}}$	0.005	-
-	Q_a	5.33	-
-	Q_b	0.45	-
-	T	0.52	-
-	γ	1.7	-

https://doi.org/10.5194/gmd-2024-174
Preprint. Discussion started: 20 November 2024


Geoscientific Model Development Discussions
Open Access
EGU

of water (Ronald and Zaneveld, 1982), the system of equations can be solved analytically, with the final expression presented in Appendix A.

The Remote Sensing Reflectance $R_{rs}^{MODEL}(\lambda)$ can be computed from the solution $E_{\mathrm{u}}(0,\lambda)$ (Aas and Højerslev, 1999) as

$$R_{rs}^{MODEL}(\lambda) = \frac{\boldsymbol{E}_{u,\lambda}(0)}{Q(\theta)\left(\boldsymbol{E}_{\mathrm{dir},\lambda}(0) + \boldsymbol{E}_{\mathrm{dif},\lambda}(0)\right)} \tag{6}$$

with

$$Q(\theta) = Q_a e^{-Q_b \sin\left(\pi/180(90-\theta)\right)}, \tag{7}$$

Q_a and Q_b constant parameters (see Table 2).

Due to the interaction in the interface between the sea surface and the atmosphere, a correction has to be added to the R_{rs}^{MODEL} (Lee et al., 2002), with the relation,

$$\boldsymbol{R}_{rs,down}(\lambda) = \frac{R_{rs,up}(\lambda)}{T + \gamma R_{rs,up}(\lambda)} \tag{8}$$

where T and γ are constant parameters (see Table 2), $R_{rs,down}(\lambda)$ is the Remote Sensing Reflectance just under the sea surface, and $R_{rs,up}(\lambda)$ is the Remote Sensing Reflectance just up the sea surface.

Thus, the final expression for R_{rs}^{MODEL} is a model that depends on the optical constituents and the boundary conditions.

3.3 Model of the in-situ observations

We aim to model the chlorophyll-α as the retrieved quantity from the inversion problem. The particulate backward scattering coefficient ($b_{b,p}(\lambda)$) is modeled as the contribution to backward scattering from the phytoplankton and NAP,

$$b_{b,p}(\lambda) = b_{b,\mathrm{phy}}(\lambda)C + b_{b,\mathrm{NAP}}(\lambda)\mathrm{NAP} \tag{9}$$

where the carbon C is calculated as Eq. (3). The downward light attenuation coefficient (k_d) is computed by the relation,

$$E_{\mathrm{dir}}(h,\lambda) + E_{\mathrm{dif}}(h,\lambda) = (E_{\mathrm{dir}}^{\mathrm{OASIM}}(0,\lambda) + E_{\mathrm{dif}}^{\mathrm{OASIM}}(0,\lambda))e^{-k_d h}. \tag{10}$$

4 BAYESIAN INVERSE PROBLEM

The model for the Remote Sensing Reflectance (R_{rs}^{MODEL}) depends on the density of the optical constituents chla, NAP and CDOM. The inverse problem consists in retrieving these constituents from the forward model, and the satellite observations (R_{rs}^{OBS}). In section 4.1 we formalize the problem and introduce the nomenclature that is going to be used in the next sections, then in section 4.2 and 4.3 we introduce the Bayesian approach to solve the problem (Rodgers, 2000), as well as the approach used to optimize the model.

https://doi.org/10.5194/gmd-2024-174
Preprint. Discussion started: 20 November 2024

Geoscientific Model Development Discussions

4.1 Formal statement of the problem

We proceed to call $y \in \mathcal{Y}$ the set of satellite measurements, modeled with a forward model plus noise,

$$y(\lambda) = R_{rs}^{MODEL}(z, x(\lambda), \lambda; \Lambda) + \epsilon(\lambda), \tag{11}$$

where

$$x(\lambda) = (E_{\text{dif}}^{\text{OASIM}}(0,\lambda), E_{\text{dir}}^{\text{OASIM}}(0,\lambda), \theta, \text{PAR})$$

are available simulated quantities, $x \in \mathcal{X}$, gathered from the OASIM model,

$$\begin{aligned}\Lambda =& (r_s, r_u, r_d, v_s, v_u, a_{\text{w}}(\lambda), a_{\text{phy}}(\lambda), b_{\text{w}}(\lambda), b_{\text{phy}}(\lambda), b_{b,\text{W}}(\lambda), b_{b,\text{phy}}(\lambda),\\ & d_{\text{CDOM}}, S_{\text{CDOM}}, d_{\text{NAP}}, S_{\text{NAP}}, e_{\text{NAP}}, f_{\text{NAP}}, b_{r,\text{NAP}}, \Theta_{\text{chla}}^{0}, \Theta_{\text{chla}}^{\text{min}}, \beta, \sigma, Q_a, Q_b, T, \gamma),\end{aligned}$$

is a set of parameters, and

$$z = (\text{chla}, \text{NAP}, \text{CDOM}) \tag{12}$$

is a set of unknown or latent quantities $z \in \mathcal{Z}$, the optical constituents.

The inverse problem aims to retrieve the unknown quantity z^d given a set of measurements and OASIM-data $\{y_i^d, x_i^d\}_{i=0}^n$ performed during the day d, with n the number of measurements.

Since we have measurements for a discrete set of wavelengths (at a depth $h = 0$ m, with the exception of k_d, at a depth $h = 9$ m) the forward model is discretized as a five-dimensional vector, with each component representing values at different wavelengths. To distinguish between continuous functions and their respective discretization, λ is used as a subscript, e.g. $E_{dir,\lambda}$ represents a component of the five dimensional vector $\boldsymbol{E}_{dir}$ with magnitudes $E_{dir}(0,\lambda)$, with $\lambda = (412.5, 442.5, 490, 510, 555)$ nm. In similar fashion, $x_\lambda = (E_{\text{dif},\lambda}, E_{\text{dir},\lambda}, E_{\text{u},\lambda}, \theta, \text{PAR})$ is a component of the 5×5 tensor $\boldsymbol{x}$. Using this notation, the measurements and OASIM-data of the day d are written as $(\boldsymbol{y}^d, \boldsymbol{x}^d)$.

Noise is added to the model to account for the measurement uncertainty and the discrepancy between the forward model and the actual underlying function that governs the process, referred to as model uncertainty. In this work, we assumed that $\boldsymbol{\epsilon}$ is a random Gaussian variable with mean zero, and covariance Σ_ϵ.

As a consequence, the model of the measurement is a random variable with a Gaussian probability distribution

$$y \sim p_\Lambda(\boldsymbol{y}|z,\boldsymbol{x}) = \mathcal{N}(\boldsymbol{R}_{rs}^{MODEL}(z,\boldsymbol{x};\Lambda), \Sigma_\epsilon). \tag{13}$$

4.2 Bayesian approach to retrieve the latent variable

Under the Bayesian framework (Rodgers, 2000), the probability of the unknown quantity z, $p(z|\boldsymbol{y},\boldsymbol{x})$, given the true probability distribution of the measurement $p(\boldsymbol{y}|z,\boldsymbol{x})$, can be retrieved using the Bayes theorem,

$$p(z|\boldsymbol{y},\boldsymbol{x}) = \frac{p(\boldsymbol{y}|z,\boldsymbol{x})p(z|\boldsymbol{x})}{p(\boldsymbol{y}|\boldsymbol{x})}. \tag{14}$$

https://doi.org/10.5194/gmd-2024-174
Preprint. Discussion started: 20 November 2024

Geoscientific Model Development Discussions

Table 3. Root Mean Square Difference (RMSD) between in-situ measurements and the satellite measurements of R_{rs} in the Mediterranean Sea, obtained from a validation of the Copernicus Dataset (Colella et al., 2023).

$R_{rs,\lambda}$	RMSD($R_{rs,\lambda}$)
$R_{rs,412.5}$	1.5×10^{-3} sr
$R_{rs,442.5}$	1.2×10^{-3}sr
$R_{rs,490}$	1×10^{-3} sr
$R_{rs,510}$	8.6×10^{-4} sr
$R_{rs,555}$	5.7×10^{-4} sr

The probability distribution $p(z|\boldsymbol{y},\boldsymbol{x})$ is called the posterior probability distribution, or just the posterior, $p(\boldsymbol{y}|z,\boldsymbol{x})$ the likelihood, and $p(z|\boldsymbol{x})$ the prior probability distribution, or just the prior.

Since we are dealing with random variables, computing the posterior is equivalent to retrieving z. In the case when this computation is not possible, common approaches attempt to estimate the value of z that maximizes the posterior, named Maximum A-Posterior (MAP) estimate.

In the case of little knowledge of the value of z, it is common practice to use an improper prior, $p(z|\boldsymbol{x})$, as an uninformative prior, where each value of z is equally probable. This approach gives rise to the Maximum Likelihood Estimate (MLE).

For this work, we used a log-normal distribution prior (Campbell, 1995) for the latent variable z, with parameters μ_z, Σ_z. This is equivalent to making the change of variable $\tilde{z} = \log(z)$ with a Gaussian prior with mean μ_z and covariance Σ_z. With this prior, and the Gaussian likelihood which can be derived from the forward model $\boldsymbol{R}_{rs}^{MODEL}$, we can define the loss function

$$\begin{aligned}\mathcal{L}^{z,d}(\boldsymbol{y}^d,\boldsymbol{x}^d,\tilde{z}^d;\Lambda) &= -2\log(p_\Lambda(\tilde{z}^d|\boldsymbol{y}^d,\boldsymbol{x}^d)) \\ &= (\boldsymbol{y}^d - \boldsymbol{R}_{rs}^{MODEL}(e^{\tilde{z}^d},\boldsymbol{x}^d;\Lambda))^T\Sigma_\epsilon^{-1}(\boldsymbol{y}^d - \boldsymbol{R}_{rs}^{MODEL}(e^{\tilde{z}^d},\boldsymbol{x}^d;\Lambda)) + (\tilde{z}^d-\mu_z)^T\Sigma_z^{-1}(\tilde{z}^d-\mu_z) + c_0 \end{aligned} \quad (15)$$

with c_0 a constant. It can be shown that minimizing the loss function in Eq. (15), is the same as maximizing the posterior (Rodgers, 2000).

As an estimate of Σ_ϵ, we used a diagonal matrix, with elements equal to the square of the Root Mean Square Difference (RMSD) between in-situ measurements and the satellite measurements of R_{rs} in the Mediterranean Sea, shown in Table 3, obtained from a validation of the Copernicus Dataset (Colella et al., 2023).

For the prior parameters, we used $\mu_z = 0$ and $\Sigma_z = \mathbb{1}*\alpha$, with $\mathbb{1}$ a diagonal matrix of dimension 3×3, and α a hyperparameter to be determined. These parameters were chosen in order to recover the equivalent to a ℓ_2 regularization. In Appendix B we explain the criteria used to tune α.

https://doi.org/10.5194/gmd-2024-174
Preprint. Discussion started: 20 November 2024

Geoscientific Model Development Discussions
Open Access
EGU

To retrieve $\tilde{Z}^* = \{\tilde{z}^{d*}\}_{d=1}^{D}$, we want to minimize $\mathcal{L}^{z,d}$ with respect of $\tilde{z}^d$ for every day d. We can perform this retrieval for all the historical data by minimizing the loss function,

$$\begin{aligned}\tilde{Z}^* &= \text{argmin}_{\tilde{Z}} \mathcal{L}^z \\ &= \text{argmin}_{\tilde{Z}} \sum_{d=0}^{D} \mathcal{L}^{z,d}(\boldsymbol{y}^d, \boldsymbol{x}^d, \tilde{z}^d; \Lambda).\end{aligned} \tag{16}$$

4.3 Model optimisation scheme

We want to optimize the forward model $\boldsymbol{R}_{rs}^{MODEL}(\tilde{z}, \boldsymbol{x}^d; \Lambda)$ by adjusting the parameters Λ, using the in-situ observations $H^{\text{OBS}} = \{(\boldsymbol{kd}^{d,\text{obs}}, \boldsymbol{b_{bp}}^{d,\text{obs}}, \text{chla}^{d,\text{obs}})\}_{d=1}^{D}$, where D is the number of days, $\boldsymbol{kd}^{d,\text{obs}}$ is a vector of dimension five, $\boldsymbol{b_{bp}}^{d,\text{obs}}$ a vector of dimension three given that there are only in-situ observations for $\lambda = (442.5, 490, 555)\text{nm}$, and $chla^{d,\text{obs}}$ a scalar. Therefore, H^{OBS} is a set of nine dimensional vectors defined for each day.

In order to take into account missing data, we introduce the modeled observation function, H^{MODEL}, a nine dimensional vector analogous to H^{OBS}, and the presence-absence nine dimensional vectors I^d, with components equal to one if there are observations and zero otherwise.

Finally, we defined $\mathcal{H}$ as the scalar product between I^d and H^{MODEL} for each day,

$$\begin{aligned}\mathcal{H}(Z, \boldsymbol{X}; \Lambda) &= \{I^d \cdot H^{\text{MODEL}}\}_{d=1}^{D} \\ &= \{I^d \cdot (\boldsymbol{kd}(z^d; \boldsymbol{x}^d, \Lambda), \boldsymbol{b_{bp}}(z^d; \boldsymbol{x}^d, \Lambda), \text{chla})\}_{d=1}^{D}\end{aligned} \tag{17}$$

where $Z = \{z^d\}_{d=1}^{D}$ and $\boldsymbol{X} = \{\boldsymbol{x}^d\}_{d=1}^{D}$.

We aim to optimize the parameters Λ from the estimated probability distribution $p_\Lambda(\boldsymbol{y}|z, \boldsymbol{x})$, by finding the Maximum Likelihood estimate of the parameters, given the in-situ observations. Using a Gaussian Likelihood with unitary covariance, we arrive to an expression for the negative log likelihood equivalent to a ℓ_2 norm,

$$\mathcal{L}^H = \frac{\sum_{d=0}^{D} ||\mathcal{H}^d(Z^*; \boldsymbol{X}, \Lambda) - H^{OBS,d}||_2}{D} \tag{18}$$

where $||\cdot||_2$ is the ℓ_2 norm, taken over a space with a dimension equal to the number of observations performed during a day d. Minimizing this loss function would be equivalent to finding the Maximum Likelihood Estimate, but instead, in order to have an estimate of the uncertainty, we proceed to use a Markov State Monte Carlo algorithm to approximate the posterior probability distribution.

4.4 Optimization algorithm and Markov State Monte Carlo

We aim to use an algorithm that can estimate the latent variable z and his uncertainty, in an operationally efficient way. To do so, in the next section we describe how we minimized the loss function $\mathcal{L}^z$ described in section 4.2, which is equivalent to finding the MAP estimate.

https://doi.org/10.5194/gmd-2024-174
Preprint. Discussion started: 20 November 2024

Geoscientific Model Development Discussions

The final posterior distribution is not a Gaussian distribution, given that the forward model is a nonlinear function of z. Instead, to estimate his uncertainty in an operationally efficient manner, we approximated the posterior distribution with a Gaussian by linearizing the forward function around the MAP estimate.

In contrast, since the estimation of the mean and uncertainty of the parameters Λ is not intended to be part of an operational system, we approximate the posterior by using a Markov State Monte Carlo Algorithm described in section 4.4.2.

4.4.1 Estimation of the latent variable

We performed the minimization of $\mathcal{L}^z$ using the Adam algorithm, with a learning rate of 0.03 and all the default parameters from the library PyTorch (Paszke et al., 2019), with 90% of all the historical data per iteration, selected randomly across the entire time span. The remaining 10% was used as test set. A copy of the code availability for every algorithm described in this work is in Soto (2024).

After $\tilde{Z}^*$, the set of latent variables for the entire training set, has been retrieved, in order to estimate the uncertainty, we linearized $\boldsymbol{R}_{rs}^{MODEL}(e^{\tilde{z}^d}, \boldsymbol{x}; \Lambda)$ around $\tilde{z}^{d*}$, as

$$\begin{aligned}
&\boldsymbol{R}_{rs}^{MODEL}(e^{\tilde{z}^d}, \boldsymbol{x}; \Lambda) \\
&\approx \boldsymbol{R}_{rs}^{MODEL}(e^{\tilde{z}^{d*}}, \boldsymbol{x}; \Lambda) + \nabla_{\tilde{z}^d} \boldsymbol{R}_{rs}^{MODEL}(e^{\tilde{z}^d}, \boldsymbol{x}; \Lambda)|_{(\tilde{z}^d = \tilde{z}^{d*})} (\tilde{z}^d - \tilde{z}^{d*}) \\
&= \boldsymbol{R}_{rs}^{MODEL}(e^{\tilde{z}^{d*}}, \boldsymbol{x}; \Lambda) + \boldsymbol{K}(\tilde{z}^d - \tilde{z}^{d*}).
\end{aligned} \tag{19}$$

Then, as shown in (Rodgers, 2000) the covariance matrix of the approximate posterior can be written as

$$\Sigma_{\tilde{z}^{d*}} = (\boldsymbol{K}^T \Sigma_\epsilon^{-1} \boldsymbol{K} + \Sigma_z^{-1})^{-1}. \tag{20}$$

In this way, the uncertainty is computed as the root square of the diagonal elements of $\Sigma_{\tilde{z}^{d*}}$.

Then, since the resulting retrieved values $\tilde{Z}^*$ are normally distributed, $Z^* = \exp(\tilde{Z}^*)$ has a log-normal distribution and thus, the uncertainty can be computed with the 68% confidence interval (here we match the convention of using the standard deviation as uncertainty for variables with normal distribution).

The uncertainty for derived variables like kd and $b_{b,p}$ is computed with standard error propagation (Arras, 1998).

4.4.2 Markov State Monte Carlos algorithm for optimizing the model parameters

In order to estimate the posterior distribution of the parameters, $p(\Lambda|H^{\text{OBS}}, \hat{Z}, \boldsymbol{X})$, we used the Metropolis-Hasting Algorithm (Chib and Greenberg, 1995; Andrieu and Thoms, 2008). This approach returns samples from a probability density function $\pi(x)$ by defining a transition probability,

$$\begin{aligned}
&p(x,y) = q(x,y)\alpha(x,y), \; x \neq y \\
&\alpha(x,y) = \min\left[\frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}, 1\right] \text{ if } \pi(x)q(x,y) > 0, \\
&\qquad = 1, \text{ otherwise}
\end{aligned} \tag{21}$$

https://doi.org/10.5194/gmd-2024-174
Preprint. Discussion started: 20 November 2024

Geoscientific Model Development Discussions

where $q(x,y)$ is a proposed transition probability from x to y, and $\alpha(x,y)$ is the acceptance probability. With this definition, a sampling from $\pi(x)$ is performed by, first selecting an initial state x_0, then computing the probability $\alpha(x_0,x_1)$ of moving to a new state x_1 sampled with a probability density $q(x_0,x_1)$. If the movement is performed, then we sample the new state x_1 and repeat the process with the new state as the original state, if the movement is not performed, then we sample the original state and repeat the process.

Some drawbacks are known for this algorithm, for example the iterations have to be performed multiple times before the algorithm converges close to the mode of the distribution, or that successive iterations are strongly correlated, so many iterations have to be performed in order to obtain uncorrelated samples. These difficulties get increasingly stronger as the dimensionality of the sampling space gets bigger. In our case, to mitigate some of these effects, we choose to perturbed our parameters in such a way that we end up with a fourteen dimensional space.

A further complication is that the probability density that we want to sample depends on Z^*, the latent variable. This means that, each time we want to do an iteration of the Metropolis-Hasting Algorithm, we would need to find the MAP estimate of Z, increasing the computational time. To mitigate this problem, we proceed to use an estimate $\hat{Z}$, consisting of few iterations towards the MAP estimate.

Our model for the negative log likelihood is the loss function $\mathcal{L}^H$ described in 4.3, which give us the expression for the Likelihood

$$p(H^{OBS}|\Lambda,\hat{Z},\boldsymbol{X}) \propto e^{-\frac{1}{2}\mathcal{L}^H(H^{OBS},\hat{Z},\boldsymbol{X},\Lambda)}. \tag{22}$$

The density function, $\pi(x)$, that we want to sample from, is the posterior probability for the parameters. By using an improper prior, using the Bayes theorem, substituting the optimal Z^* with an approximation $\hat{Z}$, and using a symmetric probability $q(\Lambda_i,\Lambda_j) = \mathcal{N}(\Lambda_i,\alpha_q \times \mathbb{1})$, we arrive to the acceptance probability,

$$\alpha(\Lambda_i,\Lambda_j) = \min\left[e^{-\frac{1}{2}\left(\mathcal{L}^H(H^{OBS},Z^*,\boldsymbol{X},\Lambda_j)-\mathcal{L}^H(H^{OBS},Z^*,\boldsymbol{X},\Lambda_i)\right)},1\right], \tag{23}$$

where α_q is a hyperparameter, whose value is important for the convergence speed.

The perturbation of the parameters was performed in a non-standard way, since all the parameters Λ have an interpretation in the forward model. Consequently, we consider the literature values Λ^0 as close estimations of the optimal values and look for a set of parameters $\Lambda^* = \delta_\Lambda^T \Lambda^0$, where δ_Λ is a vector of small perturbations from the unity.

In order to preserve the shape of the absorption of chlorophyll as a function of the wavelength $a_{\text{phy}}(\lambda)$, the vector $\boldsymbol{a}_{\text{phy}}$ was perturbed as $\boldsymbol{a}^*_{\text{phy}} = \delta_{a_{\text{phy}}} \boldsymbol{a}_{\text{phy}}$, with $\delta_{a_{\text{phy}}}$ a learnable scalar and $\boldsymbol{a}_{\text{phy}}$ the literature values. The scattering coefficient of chlorophyll $b_{\text{phy}}(\lambda)$ was approximated as a linear function of λ, so we perturbed the tangent and the intercept of the linear interpolation. The backward scattering coefficient $b_{b,\text{phy}}(\lambda)$ was perturbed in the same manner. The parameters d_{CDOM}, $b_{r,\text{NAP}}$, S_{CDOM}, $\Theta^{\text{min}}_{\text{chla}}$, Θ^{0}_{chla}, β, σ, Q_a and Q_b perturbations consisted in a per parameter scalar multiplication. All the other parameters were left unperturbed.

In this manner, the perturbations δ_Λ were initialized with ones, then using alternate minimization (AM), alternating between finding the MAP estimate of Z^* and the MLE of the parameters, we reach values for δ_Λ that we expected were close to

https://doi.org/10.5194/gmd-2024-174
Preprint. Discussion started: 20 November 2024

Geoscientific Model Development Discussions

the mode of the posterior. Next, we used the Metropolis-Hasting Algorithm, waiting until the mean value of the different parameters converged. Since consecutive samples are strongly correlated, we compute the length such that consecutive samples weren't strongly correlated, in our case, two hundred and eighty iterations. Finally, we ran forty times the Metropolis-Hasting Algorithms, with random initializations, close to the AM output. For each of them, we drop the first two thousand iterations to ensure convergence, and sampled only non correlated values.

4.5 Neural Network Based Inversion Method: Data Informed Inversion Method (DIIM)

As the dimension of the posterior increases, MCMC methods became increasingly more challenging, and even point-ways estimates, like the one obtained with Alternate Minimization, could not converge, due to the non convexity of our models. As an alternative approach, we present a framework based on the Stochastic Gradient Variational Bayes (SGVB) framework (Kingma and Welling, 2013).

The SGVB framework considers a random latent variable $z \in \mathcal{Z}$ sampled from an unknown distribution $p_{\Lambda^*}(z)$, and a random variable $\boldsymbol{y} \in \mathcal{Y}$ sampled from a distribution $p_{\Lambda^*}(\boldsymbol{y}|z)$ conditional on the latent variable z. For example, $\boldsymbol{y}$ could be measurements from a known physical process, conditional on unknown physical hidden processes.

The aim is to efficiently approximate the Maximum Likelihood estimate of the parameters Λ,

$$\Lambda^* = \text{argmax}_\Lambda(p_\Lambda(y)). \tag{24}$$

For this end, the posterior probability distribution $p_\Lambda(z|\boldsymbol{y})$ is estimated as a parameterized function $q_\phi(z|\boldsymbol{y})$. It can be shown that finding Λ^* and ϕ^* such that

$$\begin{aligned}&\Lambda^*, \phi^* = \text{argmin}_{\Lambda,\phi}\mathcal{L}_{ELBO},\\ &\mathcal{L}_{ELBO} = -D_{KL}(q_\phi(z|\boldsymbol{y})||p_\Lambda(z)) + \mathbb{E}_{q_\phi(z|\boldsymbol{y})}[\log(p_\Lambda(\boldsymbol{y}|z))]\end{aligned} \tag{25}$$

were $D_{KL}(\cdot||\cdot)$ is the Kullback-Leibler divergence (DK divergence), an asymmetric, positively defined measure of the proximity between two probability distributions (Shlens, 2014), $p_\Lambda(z)$ is the prior distribution of the latent variable z, and $\mathbb{E}_{q_\phi(z|\boldsymbol{y})}[\cdot]$ stands for the expected value over the probability distribution $q_\phi(z|\boldsymbol{y})$, is approximately equal to finding the Maximum Likelihood estimate. This is due to the fact that $\mathcal{L}_{ELBO}$, where ELBO stands for "Evidence Lower Bound", is a lower bound of the data log-likelihood $\log p_\Lambda(\boldsymbol{y})$ (see appendix C).

If the SGVB framework is used with the neural networks as the approximate probability distributions $q_\phi(z|\boldsymbol{y})$, then the neural network architecture and minimization scheme is known as Variational Auto-Encoders (Kingma and Welling, 2013), where the model $q_\phi(z|\boldsymbol{y})$ is usually called the "Encoder", and $p_\Lambda(\boldsymbol{y}|z)$ the "Decoder".

Sohn et al. (2015) generalized this framework for what they called, Conditional Variational Auto-Encoders (CVAE), where the likelihood and posterior probabilities are allowed to be conditional distributions on a third set of random variables $x \in \mathcal{X}$, $y \sim p_\Lambda(y|\mathbf{z}, x)$, and $\mathbf{z} \sim q_\phi(\mathbf{z}|y, x)$. This is the final configuration we used, but instead of training a generative model as CVAE are usually used to, we used it to solve the inversion problem while simultaneously finding approximate values for the parameters Λ^*, as explained in section 4.5.1.

https://doi.org/10.5194/gmd-2024-174
Preprint. Discussion started: 20 November 2024

Geoscientific Model Development Discussions
Open Access
EGU

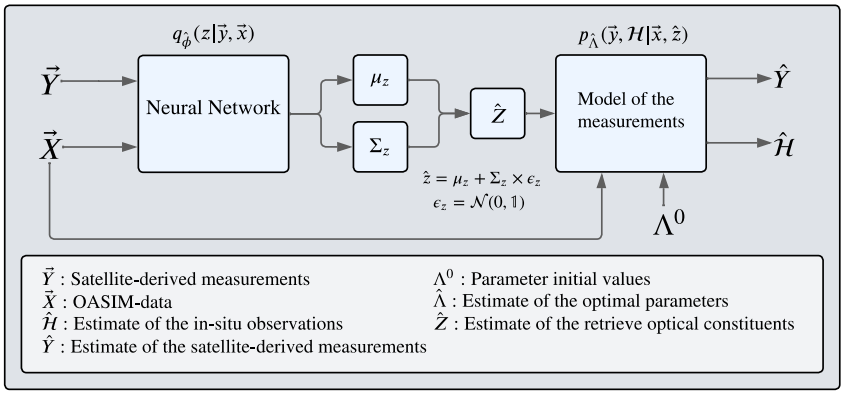


Figure 1. Diagram of the Stochastic Gradient Variational Bayes (SGVB) framework, adapted for the inversion problem, where the estimated $\hat{Z}$ is retrieve using a parameterized probabilistic function $q_\phi(z|\boldsymbol{y},\boldsymbol{x})$, which for our case, is a convolutional neural network (diagram in Fig. 2) and who's parameters ϕ are learned simultaneously as the parameters Λ.

4.5.1 Stochastic Gradient Variational Bayes framework for the Inversion Problem

CVAEs are commonly used to train a generative model $p_\Lambda(\boldsymbol{y}|z,\boldsymbol{x})$ from a probability distribution $p(\mathbf{z}|x)$ that is easy to sample, in order to generate samples that well approximate the target probability distribution (Doersch, 2021). They have been used to solve inverse problems, like image recovery (Zhong et al., 2019, 2021; Zhao et al., 2023), unfolding in high energy physics (Shmakov et al., 2024), among other applications. In contrast to previous applications of VAEs and CVAEs to inverse methods, in this work, instead of first training a CVAE with latent variables that lack a physical interpretation, we directly used the SGVB framework for the inverse method. Here, $p_\Lambda(\boldsymbol{y}|z,\boldsymbol{x})$ is the likelihood described in Eq. (13), where Λ represents the parameters of the forward function that we aim to optimize, and the latent variable $\mathbf{z}$ is the vector that we want to retrieve.

To do so, following the SGVB framework, we used a neural network $q_\phi(z|\boldsymbol{y},\boldsymbol{x})$ (diagram shown in Fig. 2) as an approximation of the posterior $p(z|\boldsymbol{y},\boldsymbol{x})$. As described in Kingma and Welling (2013), the $\mathcal{L}_{ELBO}$ is approximated by an empirical expression of the expectation value. The final expression for the loss function is,

$$\mathcal{L}_{ELBO} \approx \sum_{l=1}^{L} \log\left(p_\Lambda(\boldsymbol{y}|z_l,\boldsymbol{x})\right) + D_{KL}(q_\phi(z|\boldsymbol{y},\boldsymbol{x})||p_\Lambda(z|\boldsymbol{x})), \qquad z_l \sim q_\phi(z|\boldsymbol{y},\boldsymbol{x}). \tag{26}$$

The expression for $p_\Lambda(\boldsymbol{y}|z_l,\boldsymbol{x})$, used to learn Λ and ϕ, was different from the expression used in section 4.4 since to train the neural network, all the input data was scaled between zero and one. Also, we included a term which depends on H^{OBS} to

https://doi.org/10.5194/gmd-2024-174
Preprint. Discussion started: 20 November 2024

Geoscientific Model Development Discussions

include the in-situ measurements. The final model used was

$$\sum_{l=1}^{L}\log\left(p_\Lambda(\boldsymbol{y}|z_l)\right) = \sum_{l=0}^{L}(\boldsymbol{y}^d - \boldsymbol{R}_{rs}^{MODEL}(e^{\tilde{z}^d},\boldsymbol{x}^d;\Lambda))^T\Sigma_\epsilon^{-1}(\boldsymbol{y}^d - \boldsymbol{R}_{rs}^{MODEL}(e^{\tilde{z}^d},\boldsymbol{x}^d;\Lambda)) + (\mathcal{H}^d(e^{\tilde{z}^d},\boldsymbol{X};\Lambda) - H^{OBS,d})^T\Sigma_H^{-1}(\hat{\mathcal{H}}^d(e^{\tilde{z}^d},\boldsymbol{X};\Lambda) - H^{OBS,d}) \tag{27}$$

where Σ_ϵ^{-1} was the equivalent to the covariance matrix introduce in section 4.4, but transform according to the scaling, Σ_H^{-1} was chosen in order to have the equivalent to $\mathcal{L}^H$ from Eq. (18), L is the number of samples used per iteration, to approximate the expected value. We performed experiments with $L = 1$, $L = 10$ and $L = 100$. The performance of using higher values for L was not significantly higher, for which, we decide to use $L = 10$.

The neural network used was composed of two parts, one having as impute the mean μ_{q_z} and the other one, the covariance matrix Σ_{q_z} of a Gaussian probability distribution, in consequence, the DK divergence between two multivariate Gaussian is,

$$D_{KL}(q_\phi(z|\boldsymbol{y})||p_\Lambda(z)) = \frac{1}{2}\left[\frac{|\Sigma_z|}{|\Sigma_{q_z}|} + \mathrm{Tr}(\Sigma_{q_z}^{-1}\Sigma_z) + (\mu_{q_z} - \mu_z)^T\Sigma_{q_z}^{-1}(\mu_{q_z} - \mu_z) - \dim_z\right] \tag{28}$$

where $|\Sigma_z|$ stands for the determinant of the scaled covariance matrix used for the prior introduced in 4.2, Tr(A) stands for the trace of a matrix A, and $\dim_z = 3$, the dimension of z.

Finally, we added a ℓ_2 regularization for the parameters Λ, since it helped for the convergence of the Neural Network.

4.6 Architecture and Training of the Neural Network

As illustrated in Fig. 2, the Neural Network (NN) is composed of three sections, the first part has two hidden layers, whose function is to reduce the dimensionality of the input layer by projecting it into the space of the in-situ observations. To achieve it, this part was trained separately from the rest of the NN. This preprocessing was done to facilitate the convergence of the final output to values that are physically plausible. The second and third parts are the predicted mean of the latent variable μ_{q_z} and the Cholezky decomposition L_{q_z} of the covariance matrix $\Sigma_{q_z} = L_{q_z}^T L_{q_z}$. We choose as output of the NN the Cholesky decomposition, to ensure that the final covariance matrix is positive-defined.

To decide the best hyperparameters of the neural network, we used the library Ray Tune (Liaw et al., 2018), a python library designed for parameter tuning, together with the Bayesian Optimization HyperBand algorithm (Falkner et al., 2018) to search in the hyperparameter space. These include the number of hidden layers, the size of the hidden layers, the learning rate, the different moments for the Adam Algorithm used to train the neural network, and the size of the mini-batches.

In the same manner as with the MCMC algorithm, we used the same 90% of the data for training, from which, we select randomly five percent of it as validation for each iteration of the hyperparameter search.

Also, we did an exploration within the activation functions, finding the CELU activation function as the one that returned the best results. The CELU function is similar to the Rectified Linear Unit (ReLU) function, where instead of being the identity for positive inputs and truncating to zero for negative inputs, truncates to minus one for negative values and makes a smooth transition between the identity part and the truncation part (Barron, 2017),

$$CELU(x) = \max(0,x) + \min(0,\alpha_c e^{x/\alpha_c} - 1). \tag{29}$$

https://doi.org/10.5194/gmd-2024-174
Preprint. Discussion started: 20 November 2024

Geoscientific Model Development Discussions
Open Access
EGU

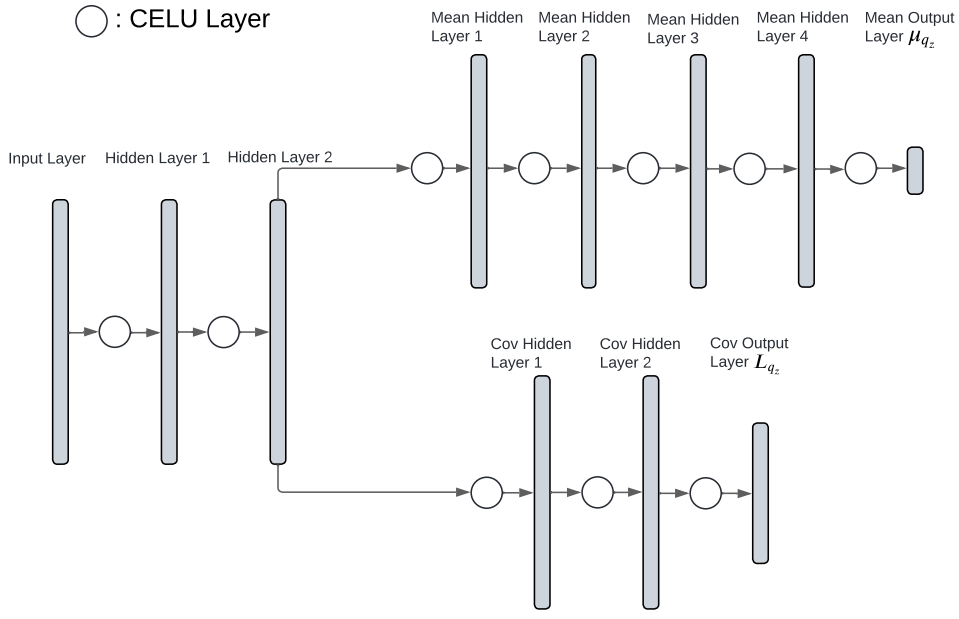


Figure 2. Diagram of the Neural Network (Soto, 2024) used as the parameterized probabilistic function $q_\phi(z|\boldsymbol{z},\boldsymbol{y})$. It is composed of three sections, the first two hidden layers reduce the dimensionality of the input layer by projecting it into the space of the in-situ observations. The output of the second layer is the input of the layers that learn the mean value of the latent variable μ_z, and dose that learn the Cholezky decomposition L_z of the covariance matrix. The dimension of the hidden layers and the number of hidden layers are tuned using Ray Tune (Liaw et al., 2018).

with α_c a hyperparameter also tuned with Ray Tune.

A diagram of the neural network $q_\phi(z,\boldsymbol{y},\boldsymbol{x})$ is presented in Fig. 2, which is part of the framework described in Fig. 1. To train the neural network first the measurements and OASIM-data $(\boldsymbol{X},\boldsymbol{Y})$ are passed to it returning an estimate for the mean and the covariance matrix of the latent variable Z. From these estimates, a random sample is computed, $\hat{Z} = \mu_z + \Sigma_z\epsilon_z$, $\epsilon_z \sim \mathcal{N}(0,\mathcal{I})$, and subsequently used as an estimate in the forward model $\boldsymbol{R}_{rs}^{MODEL}(e^{\hat{z}^d},\boldsymbol{x}^d;\Lambda)$, and with the observation function $\mathcal{H}(\hat{Z},\boldsymbol{X};\Lambda)$.

5 RESULTS

As described in section 4.4.2, we tuned twenty-six parameters, multiplying them by fourteen perturbation factors. Following Carmichael et al. (1997), the sensitivity of the Remote Sensing Reflectances, downward light attenuation coefficient and backward scattering coefficient can be computed by calculating the partial derivative with respect of the different parame-

https://doi.org/10.5194/gmd-2024-174
Preprint. Discussion started: 20 November 2024

Geoscientific Model Development Discussions
Open Access
EGU

ters ($\partial R_{RS}/\partial\delta_i$, $\partial kd/\partial\delta_i$, $\partial b_{b,p}/\partial\delta_i$), named the local sensitivity coefficients, and normalized with respect to the sensitivity coefficient (R_{RS}/δ_i, kd/δ_i, $b_{b,p}/\delta_i$) to obtained dimensional quantities. The results can be observed in Fig. 3.

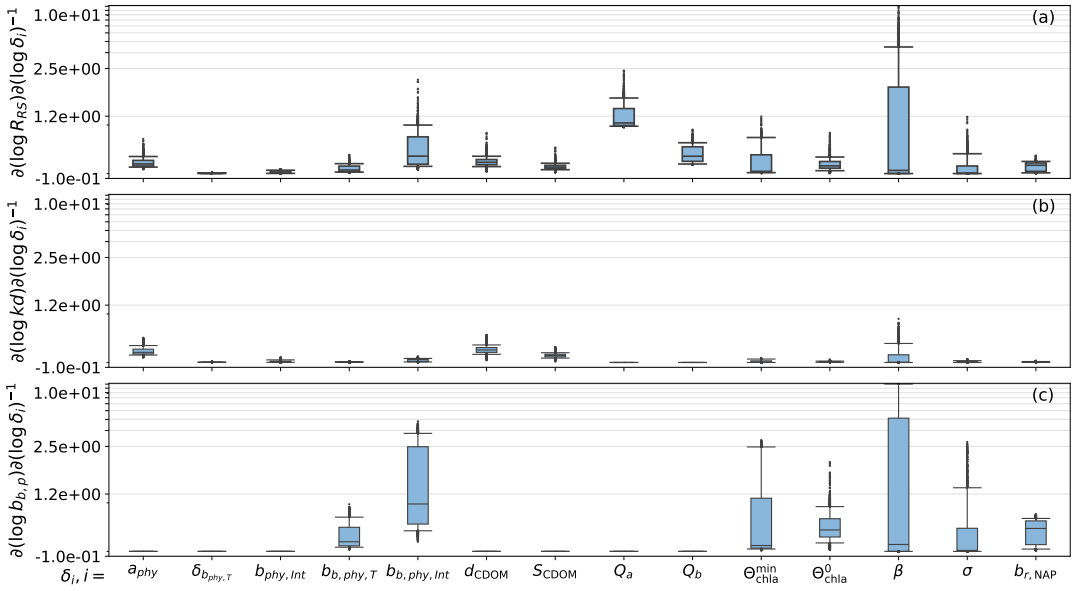


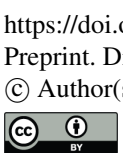
Figure 3. Sensitivity of (a) R_{rs}, (b) kd and (c) $b_{b,p}$ with respect to the perturbation factors δ_i evaluated at $\delta_i = 1$, the box plots represent the quartiles of the sensitivity for each day.

We noticed that R_{RS} and $b_{b,p}$ share a strong variability in the sensitivity with respect to the backward scattering coefficient of phytoplankton $b_{b,phy}$ and backscattering-to-scattering ratio of NAP $b_{r,\mathrm{NAP}}$ as well as the parameters $\Theta^{\mathrm{min}}_{\mathrm{chla}}, \Theta^{0}_{\mathrm{chla}}, \beta, \sigma$, which form part of the $chla : C$ ratio relation described in Eq. (3). This agrees with the seasonal variability in the abundance of the different phytoplankton functional types (Lazzari et al., 2012), as well as the variability in concentrations of pollution (Bodin et al., 2004). With this observation, we expect that using only one set of parameters for the full year would result in suboptimal predictions. Nevertheless, we proceed to find the optimal parameters that described the full historical data set.

To do so, we performed a MCMC algorithm as described in section 4.4.2. An example of the distribution obtained for each parameter can be observed in Fig. 4. The original values as well as the mean and standard deviation for the λ-dependent parameters can be appreciated in Fig. 5. Finally, the original values as well as the statistics obtained using the MCMC algorithm for the λ-independent parameters can be appreciated in Table 4.

In order to assess the normality of the final distributions, we performed a Kolmogorov-Smirnov test. Using a threshold value of 0.05, we arrived to the conclusion that most of the parameters agree with the hypotheses that they follow a Gaussian distribution, with the exception of β, and S_{CDOM}. For this reason, except the parameters just mentioned, we can use the standard deviation as a measure of the uncertainty. For β and S_{CDOM}, we reported the standard deviation as uncertainty, remarking that the final result didn't agree with the hypothesis that they followed a Gaussian distribution.

https://doi.org/10.5194/gmd-2024-174
Preprint. Discussion started: 20 November 2024



Geoscientific Model Development Discussions

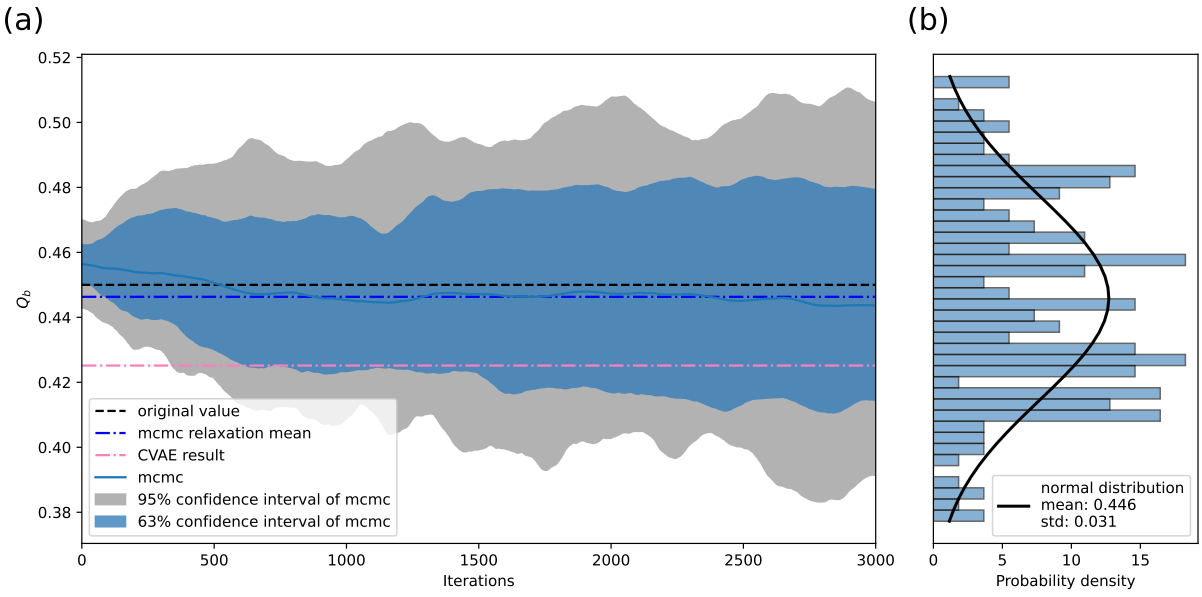


Figure 4. Result of the Metropolis-Hasting Algorithm for the parameter Q_b, using the transition probability shown in Eq. (23), with initial conditions close to the value obtained after performing Alternate Minimization. (a) Evolution of the parameter after each iteration of the algorithm, (b) final probability density estimated as a Gaussian distribution.

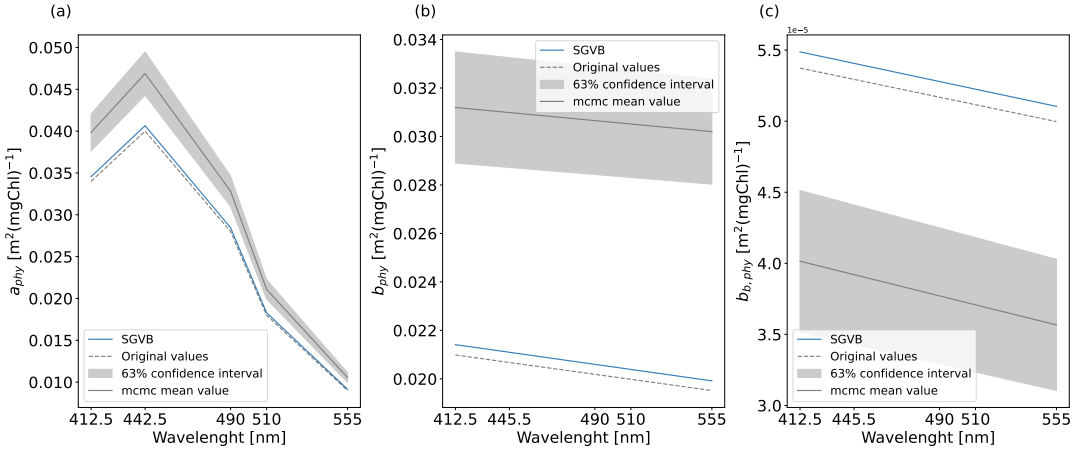


Figure 5. Original values (dashed line), final values using the SGVB framework (blue) as well as the mean and standard deviation (gray) for the λ-dependent parameters (a) absorption coefficient of phytoplankton $a_{phy}(\lambda)$, (b) scattering coefficient of phytoplankton $b_{phy}(\lambda)$ and (c) backward scattering coefficient of phytoplankton $b_{b,phy}(\lambda)$.

We also computed the covariance matrix between the perturbation factors δ_i, which can be appreciated in table 5. We observed strong correlations between d_{CDOM} and S_{CDOM} (evident due to the Eq. (4)), between β and S_{CDOM}, and in less degree

https://doi.org/10.5194/gmd-2024-174
Preprint. Discussion started: 20 November 2024

Geoscientific Model Development Discussions

Table 4. Original values, final values obtained using the SGVB framework, as well as the mean, standard deviation and Kolmogorov-Smirnov test coefficient for the sampling with the Metropolis-Hasting Algorithm for the λ-independent parameters.

	Original value	CVAE result	MCMC result	KS test for normality	KS p-value for normality
d_{CDOM} [$\text{m}^2(\text{mgCDOM})^{-1}$]	0.0150	0.0151	0.0146 ± 0.0020	0.0524	0.7525
S_{CDOM} [nm]	0.0170	0.0163	0.0122 ± 0.0020	0.1114	0.0348
Q_a	5.3300	5.9301	5.3145 ± 0.4657	0.0552	0.6932
Q_b	0.4500	0.4265	0.4459 ± 0.0314	0.0869	0.1683
$\Theta^{\text{min}}_{\text{chla}}$ [$\text{mgChla}(\text{mgC})^{-1}$]	0.0050	0.0050	0.0058 ± 0.0005	0.0755	0.3053
Θ^{0}_{chla} [$\text{mgChla}(\text{mgC})^{-1}$]	0.0300	0.0298	0.0301 ± 0.0024	0.0696	0.4019
β [$\text{mmolm}^{-2}\text{s}^{-1}$]	500.0000	616.2043	536.1500 ± 35.3500	0.1387	0.0038
σ [$\text{mmolm}^{-2}\text{s}^{-1}$]	20.0000	19.9735	24.3895 ± 1.8762	0.0927	0.1203
$b_{r,\text{NAP}}$	0.0050	0.0045	0.0052 ± 0.0005	0.0822	0.2180

Table 5. Correlation matrix between the perturbation factors δ_i, computed using the samples from the Metropolis-Hasting Algorithm.

$\delta_i, i =$	a_{PH}	$b_{\text{phy},T}$	$b_{\text{phy,Int}}$	$b_{b,\text{phy},T}$	$b_{b,\text{phy,Int}}$	d_{CDOM}	S_{CDOM}	Q_a	Q_b	$\Theta^{\text{min}}_{\text{chla}}$	Θ^{0}_{chla}	β	σ	$b_{b,\text{NAP}}$
a_{phy}	1.00													
$b_{\text{phy},T}$	0.15	1.00												
$b_{\text{phy,Int}}$	-0.16	-0.01	1.00											
$b_{b,\text{phy},T}$	0.05	0.21	-0.09	1.00										
$b_{b,\text{phy,Int}}$	-0.06	0.15	0.39	0.07	1.00									
d_{CDOM}	0.21	0.23	0.22	-0.02	0.43	1.00								
S_{CDOM}	0.24	0.31	0.24	-0.19	0.38	0.73	1.00							
Q_a	0.23	0.28	0.24	-0.15	0.22	0.41	0.62	1.00						
Q_b	-0.11	0.25	-0.01	0.32	-0.03	0.02	-0.02	-0.19	1.00					
$\Theta^{\text{min}}_{\text{chla}}$	0.08	0.11	-0.09	0.13	-0.12	-0.08	-0.18	-0.20	-0.06	1.00				
Θ^{0}_{chla}	0.06	-0.01	0.02	0.10	0.08	-0.17	-0.10	0.08	-0.10	-0.37	1.00			
β	-0.13	-0.21	-0.13	0.22	-0.21	-0.50	-0.73	-0.49	-0.05	0.11	-0.02	1.00		
σ	-0.03	-0.15	0.07	0.25	0.15	0.14	-0.04	-0.16	0.01	0.22	-0.06	0.08	1.00	
$b_{b,\text{NAP}}$	0.00	-0.11	-0.10	-0.20	-0.09	0.04	0.07	-0.03	0.13	0.11	-0.29	-0.06	0.07	1.00

between d_{CDOM} and $b_{b,\text{phy,Int}}$, S_{CDOM} and $b_{b,\text{phy,Int}}$, Q_a and d_{CDOM}, Q_a and S_{CDOM}, $\Theta^{\text{min}}_{\text{chla}}$ and Θ^{0}_{chla}, β and Q_a, and between β and d_{CDOM}.

As described in section 4.5, we also used the SGVB framework to find an optimal parametrization. To compare them with the output obtained with the MCMC algorithm, we evaluated the MAP estimates of the optical constituents z given each set of parameters. The results can be appreciated in table 4 and Fig. 5. Taking into account the uncertainty of the MCMC results, and

https://doi.org/10.5194/gmd-2024-174
Preprint. Discussion started: 20 November 2024

Geoscientific Model Development Discussions

using the 95% confidence interval, we see that only the parameters Q_a, Q_b, $\Theta^{\min}_{\text{chla}}$, Θ^{0}_{chla} and $b_{r,\text{NAP}}$ agree in the final values. Nevertheless, in order to assess the ability of the resulting parameters to optimize the forward model, we computed the MAP estimate of the latent variable z for the test set of the historical data, a set of data that was not used for training of the neural network in any way, or to estimate the parameters in any of the methods.

These estimates were computed using the literature values, the mean values obtained after performing the MCMC algorithm, and with the parameters obtained with the SGVB framework. We find that, in terms of the Root Mean Square Error (RMSE) between measurements and MAP estimates, values that can be appreciated in table D1, both methods optimized the forward model, and for some measurements, including the in-situ observations of chlorophyll, the parameters obtained with the SGVB framework outperform the MCMC results. We interpret the disparity between results using both methods, by arguing that the posterior probabilities that we are attempting to maximize are multi-modal, the Alternate Minimization in combination with the MCMC algorithm finds one set of locally optimal parameters, while the SGVB framework finds a different one.

The most noticeably advantage of using the SGVB framework, is that it provides an efficient way of computing estimates of the optical constituents z, which by construction, are also consistent with the forward model, with optimal RMSE between measurements and estimates. Since they are computed with a neural network, the computational time outperforms the standard implicit inversion methods, required in cases where the expression of the RTE is too complicated to invert it analytically. For completeness, in tables D2 and D3, we also present the Pearson correlation coefficients and the relative Median absolute deviation (rMAD) between measurements and estimates, and the timelines for the values estimated in Fig. D1, D2 and D3.

This timelines illustrates the ability of the inverse methods to describe the seasonal variability. We also notice that the uncertainty increase for days when the discrepancy between measurement and prediction is bigger, reaffirming the advantages of including uncertainty in the predictions. For the derived quantities $kd(\lambda)$ and $b_{b_p}(\lambda)$, for some cases, the standard error propagation method under-estimated the uncertainty, especially for $kd(\lambda)$ with λ greater than 510 nm.

Nevertheless, the values of R_{RS} computed with the estimated optical constituents are consistent with the satellite observations for all the wavelengths computed, as can be confirmed with the values of RMSE in table D1, rMAD in table D3 and correlation in table D2.

6 Discussion

In the last years, there has been an increasing number of applications of neural networks in earth sciences, like forecasts of the El Niño-Southern Oscillation (ENSO) by using historical simulations and a convolutional neural network (Ham et al., 2019), fusion of satellite data (Chapman and Charantonis, 2017; Denvil-Sommer et al., 2019; Bocquet et al., 2020), classification of regions on the ocean (Richardson et al., 2003; Saraceno et al., 2006), finding drivers of net primary productivity using self organizing maps (Lachkar and Gruber, 2012), reconstruction of oceanographic variables (Martinez et al., 2020; Pietropolli et al., 2022), classification of the anomalies of water leaving radiance (Mustapha et al., 2014), data reconstruction (Manucharyan et al., 2021; George et al., 2021), inversion of oceanographic variables (Brajard et al., 2006; Irrgang et al., 2019; Dessailly, 2012), pattern recognition (Maze et al., 2017; Jones et al., 2019; Jones and Ito, 2019; Boehme and Rosso, 2021; Desbruyères

https://doi.org/10.5194/gmd-2024-174
Preprint. Discussion started: 20 November 2024


Geoscientific Model Development Discussions

et al., 2021), forecast imposing physical constrains (De Bézenac et al., 2019; Erichson et al., 2019), increase of the resolution of modeling (Barthélémy et al., 2022), among others.

Our work makes use of a neural network to approximate the posterior probability distribution of optical constituents in the sea by employing the SGVB framework. As described in section 4.5.1, we minimized the ELBO loss function, which simultaneously optimized the forward model by finding the MLE of the parameters, deriving in-situ biogeochemical parameters for reflectance observations, linking the neural network procedure to an interpretable model. The novelty of our approach is the implicit inclusion of the model uncertainty in the results. In fact, the output approximates the true MAP estimate given a set of measurements.

On the other hand, our results with the SGVB framework under-estimated the uncertainty of the optical constituents, a computation that is of crucial importance for multiple applications, like objective comparison of simulations against observations, efficient assimilation of data with methods like the Kalman Filters, among others (Brankart et al., 2012). The requirement of reliable uncertainty estimations lead us to use only the point-ways estimate of the neural network. Furthermore, we explored the Bayesian approach, approximating the final posterior distribution of the optical constituents, $p_\Lambda(z|\boldsymbol{y},\boldsymbol{x})$, with a Gaussian probability distribution. This method returns estimates with reliable uncertainty estimations that can be used in real operational systems.

In particular, in addition to the optical constituents, our aim was to find the optimal model with respect to all the in-situ observations for the entire time span. This ambitious goal made the final results suboptimal for some individual measurements. For example Salama and Verhoef (2015) used a similar forward model with the aim of estimating the downward light attenuation coefficient at a wave-length of 490 nm, $kd(490)$, at different depths, obtaining a rMAD of 11.84%, while our result using the MCMC parameters presented a rMAD value of 22%. We noticed that by optimizing only one in-situ measurement, we could find a set of parameters that made that measurement more precise. Nevertheless, we decided to use the parameters presented, to balance the global accuracy. For example, in terms of the rMAD of the Remote Sensing Reflectance at a wavelength of 490 nm, $R_{rs}(490)$, we obtained a value of rMAD of 1.2%, outperforming previous works.

Our approach also differs from other works on Bayesian estimation of optical constituents (Gordon and Boynton, 1997; Boynton and Gordon, 2000; Michalopoulou et al., 2009; Erickson et al., 2023), since we are employing a three steam model, derived from the Radiative Transfer Model (Dutkiewicz et al., 2015), and using it to derived the in-situ observations for all the wave-lengths available. This feature allows scientists to understand the automatic learning process in terms of meaningful physical parameters.

The approach can be extended in different directions, in particular, the addition of more optical constituents, which will be facilitated once the information of the new satellite missions PRISMA (Hyperspectral Precursor of the Application Mission), with 12 nm spectra resolution ranging from 400 to 2500 nm, and PACE (Plankton, Aerosol, Cloud, Ocean Ecosystem) with a 5 nm resolution ranging from 350 to 890 nm, is used as input of the system.

https://doi.org/10.5194/gmd-2024-174
Preprint. Discussion started: 20 November 2024

Geoscientific Model Development Discussions
Open Access
EGU

7 CONCLUSIONS

By utilizing the Bayes theorem and linearizing the forward function, we achieved the inversion of the optical constituents, with an estimate of the uncertainty. The latter is fundamental for the assimilation of Remote Sensing Reflectance.

By using an MCMC algorithm, we computed a set of parameters that optimized the forward model, and showed that the solution was not unique by using the SGVB framework. Moreover, the SGVB framework can be used as an alternative to find point-wise estimates of optimal parameters, and also as an efficient way of computing point-wise estimates of the optical constituents.

Regarding the computational advantages of the SGVB framework, as long as the uncertainty is not required, it is the best option to estimate the optical constituents in operational systems. Nevertheless, the posterior probability learned by the neural network under-estimated the uncertainty of the result, which makes the MAP algorithm preferred when the uncertainty is a requirement. Since the computational time for the MAP estimate depends from the initial conditions, we proposed to use the SGVB estimates as initial conditions for the MAP algorithm, which, by making experiments with our current implementation, we found that is capable to reduce the number of steps more than 50%.

For future work, it would be important to apply and verify the accuracy of the approach with more optical constituents and to test Remote Sensing Reflectance assimilation in a biogeochemical model.

Code and data availability. The last version of the model is publicly available from GitHub at https://github.com/carlossoto362/OGS_one_d_model, under the Licence Apache, version 2.0. The version used to produce the results is archived on Zenodo, as are the input data and scripts to run the model and produce the plots for all the simulations presented in this paper, under the DOI 10.5281/zenodo.13741206 (Soto, 2024).

Two examples of how to download the data and use the code are available in google colab at

- Bayesian inversion example: https://colab.research.google.com/drive/1ROAXHcGPcubT5lr_1n1row_cYhTk8Qd-?usp=sharing
- Neural Network based inversion method: https://colab.research.google.com/drive/1wJwG-DsJ63Qy3z-jyIuX3tcyKxfrGlwd?usp=sharing.

The neural network architectures are stored in .pt files, which can be read as tensors using the PyTorch library from Python, and can be accessed with the command “torch.load(file.pt)”, the CVAE model used is in VAE_model/model_second_part_chla_centered.pt, which use VAE_model/model_first_part.pt. The historical data is in the folder npy_data, and can be access using the library NumPy from python with the command “np.load(file.npy)”; a description of each column is in the file README from the same folder.

https://doi.org/10.5194/gmd-2024-174
Preprint. Discussion started: 20 November 2024

Geoscientific Model Development Discussions

Appendix A

In this section, we expand the solution of Eq. (1) subject to the boundary conditions (5), under the homogeneity assumption. First, for simplicity, we re-write Eq. (1) as,

$$\frac{dE_{\text{dir}}(h,\lambda)}{dh} = -c_d(\lambda)E_{\text{dir}}(h,\lambda),$$
$$\frac{dE_{\text{dif}}(h,\lambda)}{dh} = -C_s(\lambda)E_{\text{dif}}(h,\lambda) + B_u(\lambda)E_u(h,\lambda) + F_d(\lambda)E_{\text{dir}}(h,\lambda),$$
$$\frac{dE_{\text{u}}(h,\lambda)}{dh} = -B_s(\lambda)E_{\text{dif}}(h,\lambda) + C_u(\lambda)E_u(h,\lambda) - B_d(\lambda)E_{\text{dir}}(h,\lambda),$$

subject to,

$$E_{\text{dir}}(0,\lambda) = E_{\text{dir}}^{\text{OASIM}}(0,\lambda), E_{\text{dif}}(0,\lambda) = E_{\text{dif}}^{\text{OASIM}}(0,\lambda), E_{\text{u}}(\infty,\lambda) = 0, \tag{A1}$$

were,

$$c_d(\lambda) = \frac{a(\lambda)+b(\lambda)}{\cos\theta},$$
$$C_s(\lambda) = \frac{a(\lambda)+r_s b_b()\lambda}{v_s},$$
$$B_u(\lambda) = \frac{r_u b_b(\lambda)}{v_u},$$
$$F_d(\lambda) = \frac{b(\lambda)-r_d b_b(\lambda)}{\cos\theta},$$
$$B_s(\lambda) = \frac{r_s b_b(\lambda)}{v_s},$$
$$C_u(\lambda) = \frac{a(\lambda)+r_u b_b(\lambda)}{v_u},$$
$$B_d(\lambda) = \frac{r_d b_b(\lambda)}{\cos\theta}. \tag{A2}$$

Equation (A1) is a linear system of Ordinary Differential Equations, which can be solved by, first solving the equation for $E_{\text{dir}}(h,\lambda)$, followed by solving the system of equations for $E_{\text{dif}}(h,\lambda)$ and $E_{\text{u}}(h,\lambda)$, taking the solution of $E_{\text{dir}}(h,\lambda)$ as the in-homogeneous part of the system of equations. The final expresion is,

$$E_{\text{dir}}(h,\lambda) = E_{\text{dir}}^{\text{OASIM}}(0,\lambda)e^{-hc_d},$$
$$E_{\text{dif}}(h,\lambda) = c^+e^{-k^+h} + x_{\text{dif}}E_{\text{dir}}(h,\lambda),$$
$$E_u(h,\lambda) = c^+r^+e^{-k^+h} + y_{\text{u}}E_{\text{dir}}(h,\lambda), \tag{A3}$$

https://doi.org/10.5194/gmd-2024-174
Preprint. Discussion started: 20 November 2024

Geoscientific Model Development Discussions

505 where,

$$
\begin{aligned}
c^{+} &= E_{\text{dif}}^{\text{OASIM}}(0,\lambda) - x E_{\text{dir}}^{\text{OASIM}}(0,\lambda),\\
k^{+} &= D - C_u,\\
r^{+} &= \frac{B_s}{D},\\
D &= \frac{1}{2}\left(C_s + C_u + \sqrt{(C_s+C_u)^2 - 4B_sB_u}\right),\\
x &= \frac{(-(C_u+c_d)F_d - B_uB_d)}{(c_d-C_s)(c_d+C_u)+B_sB_u},\\
y &= \frac{(-B_sF_d + (-C_s+c_d)B_d)}{(c_d-C_s)(c_d+C_u)+B_sB_u}.
\end{aligned}
\tag{A4}
$$

For completes, in the case when the expression $(c_d - C_s)(c_d + C_u) + B_sB_u = 0$, then the expression for c^{+} has to be exchange to $c^{+} = E_{\text{dif}}^{\text{OASIM}}(0,\lambda)$.

Appendix B: Tuning of the hyperparameter α

510 As seen in section 4, the final covariance matrix for the retrieve $\tilde{Z}^{*}$ depends on α thought Σ_z. We selected the value of α to fulfill two criteria, the final result for $\tilde{Z}^{*}$ should not depend on α, and the estimated uncertainty has to be close to the discrepancy between retrieved data and in-situ observations.

For this end, we defined the error of the forward model $\epsilon_{R_{rs}}(\alpha)$ as the Root Mean Square Difference between the satellite Remote Sensing Reflectance and the predicted by the model. We expect this quantity to not depend on α.

515 We also defined the error between the predicted uncertainty and the actual discrepancy between model and data $\epsilon_{\delta_{\text{chla}}}(\alpha)$, where the predicted uncertainty is estimated as the mean value of the standard deviation of the predicted chla^{MODEL}, and the discrepancy between model and data is estimated as the Root Mean Square Difference between chla^{OBS} and chla^{MODEL}.

We selected the parameter α that minimized this two errors.

We computed $\epsilon_{R_{rs}}(\alpha)$ and $\epsilon_{\delta_{\text{chla}}}(\alpha)$ for different values of α until the curve $\epsilon_{R_{rs}}(\alpha)$ flattens. With the errors computed, we
520 re-scaled the error functions $\epsilon_{R_{rs}}(\alpha)$ and $\epsilon_{\delta_{\text{chla}}}(\alpha)$ between zero and one in order to minimize both functions simultaneously by minimizing the loss function

$$
\mathcal{L}^{\alpha} = \overline{\epsilon_{R_{rs}}(\alpha)} + \overline{\epsilon_{\delta_{\text{chla}}}(\alpha)},
\tag{B1}
$$

where the line over the errors stand for the re-scaling. Figure B1 shows the final value of α selected, as a function of $\epsilon_{R_{rs}}(\alpha)$, $\epsilon_{\delta_{\text{chla}}}(\alpha)$ and $\mathcal{L}^{\alpha}$.

https://doi.org/10.5194/gmd-2024-174
Preprint. Discussion started: 20 November 2024

Geoscientific Model Development Discussions
Open Access
EGU

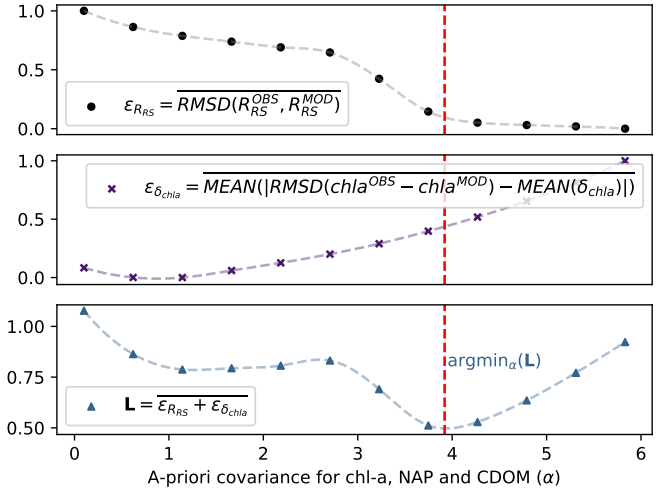


Figure B1. Illustration of how the hyper-parameter α was chosen. Using a higher α decreases the Root Mean Square Difference between the Remote Sensing Reflectance observed by satellite, and the one obtained with the model (a), but increases the error between the predicted uncertainty and the actual discrepancy between model and data (b). The value chosen was the one that minimized the $\mathcal{L}^{\alpha}$ loss function (c).

Appendix C

In this section, we show that $\mathcal{L}_{ELBO}$ is a lower bound of the data log-likelihood, first, we write the expression for the log-likelihood, by marginalizing over all possible values of the latent variable z

$$\log\left(p_{\Lambda}(y)\right) = \log\left(\int_{\mathcal{Z}} p_{\Lambda}(\boldsymbol{y}|z)p(z)dz\right), \tag{C1}$$

next we introduce the parameterized probability distribution $q_{\phi}(z|\boldsymbol{y})$

$$= \log\left(\int_{\mathcal{Z}} p_{\Lambda}(\boldsymbol{y}|z)\frac{q_{\phi}(z|\boldsymbol{y})}{q_{\phi}(z|\boldsymbol{y})}p(z)dz\right), \tag{C2}$$

finally, we use Jensen inequality to find a lower bound of the log-likelihood,

$$\begin{aligned}
&\geq \int_{\mathcal{Z}} \log\left(\frac{p_{\Lambda}(\boldsymbol{y}|z)p(z)}{q_{\phi}(z|\boldsymbol{y})}\right)q_{\phi}(z|\boldsymbol{y})dz \\
&= \int_{\mathcal{Z}} \log\left(\frac{p(z)}{q_{\phi}(z|\boldsymbol{y})}\right)q_{\phi}(z|\boldsymbol{y}) + \int_{\mathcal{Z}} \log\left(p_{\Lambda}(\boldsymbol{y}|z)\right)q_{\phi}(z|\boldsymbol{y})dz \\
&= -D_{KL}(q_{\phi}(z|\boldsymbol{y})||p(z)) + \mathbb{E}_{q_{\phi}(z|\boldsymbol{y})}[\log\left(p_{\Lambda}(\boldsymbol{y}|z)\right)] \\
&= \mathcal{L}_{ELBO},
\end{aligned} \tag{C3}$$

https://doi.org/10.5194/gmd-2024-174
Preprint. Discussion started: 20 November 2024

Geoscientific Model Development Discussions

Appendix D

In this section we include the Root Mean Square Error (RMSE), Pearson correlation coefficients (ρ) and relative Median Absolute Deviation (rMAD) for all the measurements and observations, using the MAP estimates with unperturbed parameters, MAP estimate with parameters from the MCMC algorithm, MAP estimate with parameters from the SGVB framework, and outputs from the SGVB framework. All the quantities are computed using only the test data, which is the 10% of the data that was not used in the MCMC algorithm nor in the training of the neural network. We also include supplementary figures, with the 2012 timelines comparing the historical data with the output of the MAP estimates and the SGVB framework. Finally, we include tables with the symbols used across this work.

Table D1. Root Mean Square Error between satellite and in-situ observations, and the modeled data using the Maximum A-posterior (MAP) estimate with unperturbed parameters, optimized parameters with the MCMC algorithm, optimized parameters with the SGVB framework, and modeled data purely with the SGVB framework.

Root Mean Square Error, RMSE(OBS,MOD) = $\sqrt{\text{MEAN}\left((\text{OBS}-\text{MOD})^2\right)}$				
	MAP with unperturbed parameters	MAP with MCMC parameters	MAP with SGVB parameters	SGVB output
$R_{RS,412.5}$	0.000319	0.000151	0.000218	0.000646
$R_{RS,442.5}$	0.000117	0.000115	0.000120	0.000553
$R_{RS,490}$	0.000246	0.000062	0.000090	0.000413
$R_{RS,510}$	0.000078	0.000065	0.000108	0.000242
$R_{RS,555}$	0.000067	0.000046	0.000044	0.000191
$k_{d,412.5}$	0.044358	0.044922	0.048587	0.049591
$k_{d,442.5}$	0.031283	0.028478	0.031518	0.033641
$k_{d,490}$	0.026799	0.023269	0.025459	0.026724
$k_{d,510}$	0.021233	0.018798	0.020363	0.020633
$k_{d,555}$	0.014163	0.013240	0.013648	0.014121
$b_{b,p,442.5}$	0.000811	0.000680	0.000549	0.000561
$b_{b,p,490}$	0.000640	0.000609	0.000650	0.000694
$b_{b,p,555}$	0.000497	0.000480	0.000551	0.000570
chla	0.406792	0.334460	0.299325	0.274626
Total	0.54740	0.46537	0.44123	0.42321

https://doi.org/10.5194/gmd-2024-174
Preprint. Discussion started: 20 November 2024

Geoscientific Model Development Discussions

Table D2. Pearson correlation coefficient r between satellite and in-situ observations, and the modeled data using the Maximum A-posterior (MAP) estimate with unperturbed parameters, optimized parameters with the MCMC algorithm, optimized parameters with the SGVB framework, and modeled data purely with the SGVB framework.

Spearman rank-order correlation coefficient ρ				
	MAP with unperturbed parameters	MAP with MCMC parameters	MAP with SGVB parameters	SGVB output
$R_{RS,412.5}$	0.98340	0.99730	0.99065	0.96225
$R_{RS,442.5}$	0.99710	0.99863	0.99738	0.97320
$R_{RS,490}$	0.98461	0.99737	0.99604	0.94149
$R_{RS,510}$	0.98336	0.99061	0.97535	0.79347
$R_{RS,555}$	0.97605	0.99331	0.98685	0.81142
$k_{d,412.5}$	0.79925	0.80544	0.80709	0.80948
$k_{d,442.5}$	0.88829	0.87788	0.88310	0.86992
$k_{d,490}$	0.84538	0.84020	0.84458	0.83217
$k_{d,510}$	0.85079	0.85049	0.85378	0.83961
$k_{d,555}$	0.66704	0.63792	0.66600	0.63839
$b_{b,p,442.5}$	0.65712	0.68309	0.70253	0.70185
$b_{b,p,490}$	0.53494	0.59123	0.61175	0.63218
$b_{b,p,555}$	0.65638	0.68457	0.69387	0.69072
chla	0.70004	0.83500	0.85845	0.88942
Total	11.52375	11.78302	11.86742	11.38557

https://doi.org/10.5194/gmd-2024-174
Preprint. Discussion started: 20 November 2024

Geoscientific Model Development Discussions
Open Access
EGU

Table D3. relative Median Absolute Deviation (rMAD) between satellite and in-situ observations, and the modeled data using the Maximum A-posterior (MAP) estimate with unperturbed parameters, optimized parameters with the MCMC algorithm, optimized parameters with the SGVB framework, and modeled data purely with the SGVB framework.

rMAD = MEAN($\lvert$OBS $-$ MOD$\rvert$/OBS)				
	MAP with unperturbed parameters	MAP with MCMC parameters	MAP with SGVB parameters	SGVB output
$R_{RS,412.5}$	0.046498	0.022824	0.023695	0.104053
$R_{RS,442.5}$	0.019152	0.020253	0.018087	0.092863
$R_{RS,490}$	0.048393	0.012281	0.017828	0.084470
$R_{RS,510}$	0.022621	0.019014	0.027908	0.056772
$R_{RS,555}$	0.028433	0.023294	0.018036	0.085871
$k_{d,412.5}$	0.275663	0.277158	0.307413	0.319424
$k_{d,442.5}$	0.238534	0.230608	0.243371	0.291961
$k_{d,490}$	0.240852	0.220262	0.229272	0.253470
$k_{d,510}$	0.175974	0.180554	0.176727	0.176630
$k_{d,555}$	0.106287	0.106784	0.104700	0.105321
$b_{b,p,442.5}$	0.350718	0.285072	0.244496	0.259961
$b_{b,p,490}$	0.345372	0.399523	0.545586	0.505087
$b_{b,p,555}$	0.371178	0.410633	0.578075	0.573776
chla	0.552390	0.572383	0.474861	0.283325
Total	2.82207	2.78064	3.01006	3.19298

https://doi.org/10.5194/gmd-2024-174
Preprint. Discussion started: 20 November 2024

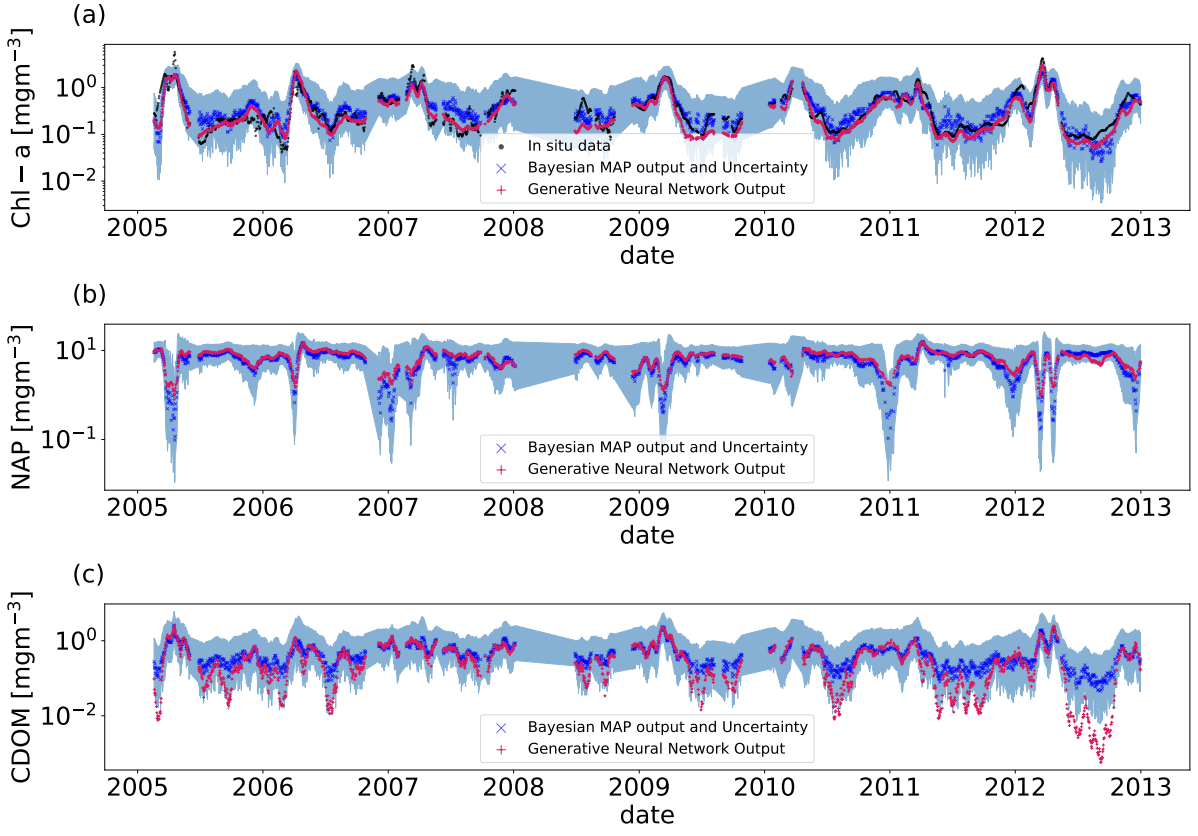


Figure D1. The 2012 time series for the chlorophyll-α (a), Non-Algal Particles (b) and Chromophoric Dissolved Organic Matter (c). For all the timelines, the black points are the in-situ observations from the BOUSSOLE buoy, the blue points are the MAP output with uncertainty (blue shadow), using the optimal parameters from the SGVB framework algorithm, and the red points are the output of the SGVB framework.

https://doi.org/10.5194/gmd-2024-174
Preprint. Discussion started: 20 November 2024

Geoscientific Model Development Discussions
Open Access
EGU

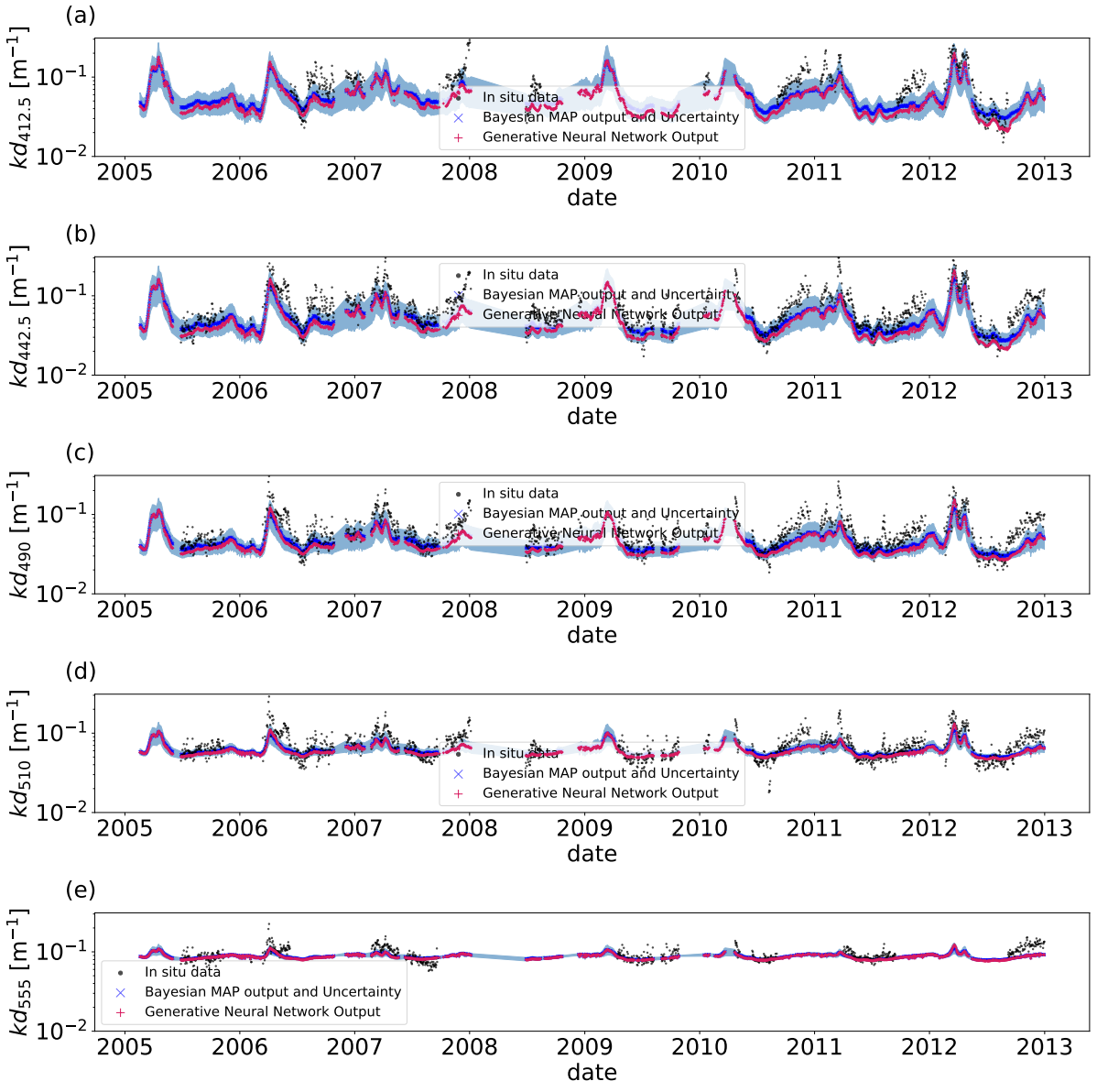


Figure D2. The 2012 time series for downward light attenuation coefficient ($k_d(\lambda)$), wavelengths $\lambda = 412.5$ (a), $\lambda = 442.5$ (b), $\lambda = 490$ (c), $\lambda = 510$ (d) and $\lambda = 555$ (e). For all the timelines, the black points are the in-situ observations from the BOUSSOLE buoy, the blue points are the MAP output with uncertainty (blue shadow), using the optimal parameters from the SGVB framework algorithm, and the red points are the output of the SGVB framework.

https://doi.org/10.5194/gmd-2024-174
Preprint. Discussion started: 20 November 2024

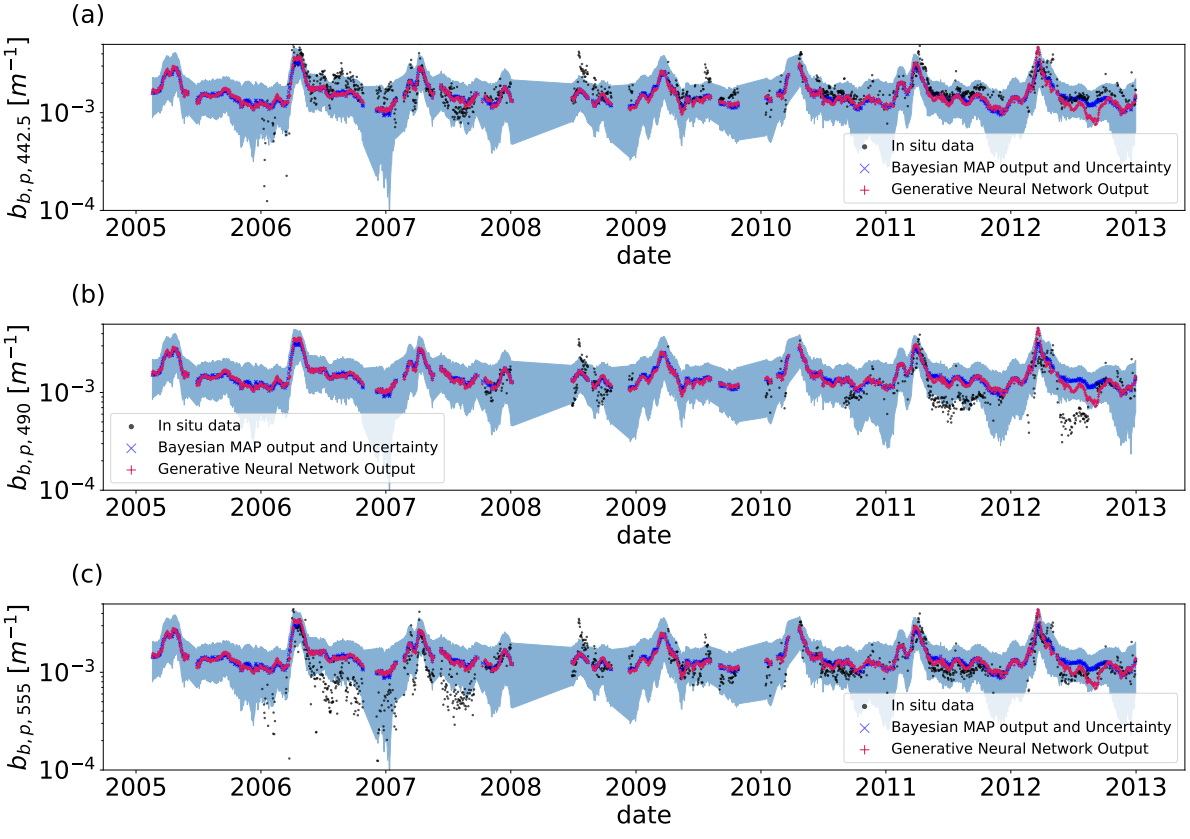


Figure D3. The 2012 time series for particulate backward scattering coefficient for the wavelengths $\lambda = 442.5$ (a), $\lambda = 490$ (b) and $\lambda = 555$ (c). For all the timelines, the black points are the in-situ observations from the BOUSSOLE buoy, the blue points are the MAP output with uncertainty (blue shadow), using the optimal parameters from the SGVB framework algorithm, and the red points are the output of the SGVB framework.

https://doi.org/10.5194/gmd-2024-174
Preprint. Discussion started: 20 November 2024

Geoscientific Model Development Discussions
Open Access
EGU

Table D4. Table of Symbols used for the Radiative Transfer Model.

Symbol	Meaning
E_{dir}	Vertical Direct irradiance
E_{dif}	Vertical scattered downward irradiance
E_{u}	Vertical scattered upward irradiance
θ	Sun zenith angle
h	depth at which a measurement is assumed to be taken.
λ	Wavelength at which a measurement is assumed to be taken.
$a(\lambda)$	Total absorption coefficient
$b(\lambda)$	Total scattering coefficient
$b_b(\lambda)$	Total backward scattering coefficient
w	Water
phy	Phytoplankton
chla	Chlorophyll-α
CDOM	Chromophoric Dissolved Organic Matter
NAP	Non Algal Particles
$a_{\text{w}}(\lambda)$	Water-specific absorption coefficient
$a_{\text{phy}}(\lambda)$	Chlorophyll-specific absorption coefficient of phytoplankton
$a_{\text{CDOM}}(\lambda)$	Mass-specific absorption coefficient of CDOM
$a_{\text{NAP}}(\lambda)$	Mass-specific absorption coefficient of NAP
$b_{\text{w}}(\lambda)$	Water-specific scattering coefficient
$b_{\text{phy}}(\lambda)$	Carbon-specific scattering coefficient of phytoplankton
$b_{\text{NAP}}(\lambda)$	Mass-specific scattering coefficient of NAP
$b_{b,\text{w}}(\lambda)$	Water-specific backward scattering coefficient
$b_{b,\text{phy}}(\lambda)$	Carbon-specific backward scattering coefficient of phytoplankton
$b_{b,\text{NAP}}(\lambda)$	Mass-specific backward scattering coefficient of NAP
PAR	Photosynthetic Available Radiation
$E_{\text{dir}}^{\text{OASIM}}(0,\lambda)$	Direct downward irradiance on the surface of the ocean, from the OASIM model
$E_{\text{dif}}^{\text{OASIM}}(0,\lambda)$	Scattered downward irradiance on the surface of the ocean, from the OASIM model
$E_{\text{u}}(\infty,\lambda)$	Scattered upward irradiance on the floor of the ocean
R_{rs}	Remote Sensing Reflectance
$b_{b,p}$	Particulate backward scattering coefficient
k_d	Downward light attenuation coefficient

https://doi.org/10.5194/gmd-2024-174
Preprint. Discussion started: 20 November 2024

Geoscientific Model Development Discussions

Table D5. Table of Symbols and notation used for the Bayes formalism.

Symbol	Meaning
$\boldsymbol{y}$	Vector, discretization of a continuous function in discrete values of λ
y_λ	Component of a vector with magnitude $y(\lambda)$
z^*	Optimal value of a retrieved quantity z, solution of a minimization problem
$\hat{z}$	Estimation of the optimal value of a quantity z
$\text{argmin}_z \mathcal{L}(z)$	Quantity z that minimized the loss function $\mathcal{L}$
$\text{argmax}_z p(y\|z)$	Quantity z that maximises the likelihood $p(y\|z)$
$\boldsymbol{y}^d$	Remote Sensing Reflectance data from day d
x^d	OASIM data from day d
z^d	Optical constituents from day d
Λ	Set of parameters from the forward model
$\boldsymbol{Y}$	Set of many days with Remote Sensing Reflectance data,which represents the train set when is used for training, and the test set when is used for testing
$\boldsymbol{X}$	Set of many days with OASIM data, which represents the train set when is used for training, and the test set when is used for testing
Z	Set of many days with retrieved optical constituents, which represents the train set when is used for training, and the test set when is used for testing
$p_\Lambda(\boldsymbol{y}\|z,\boldsymbol{x})$	Probability distribution of the variable $\boldsymbol{y}$ conditional on z, and $\boldsymbol{x}$, as a function of Λ
$\mathcal{N}(\mu,\Sigma)$	Gaussian probability distribution with mean μ and covariance matrix Σ
H^{OBS}	in-situ observations
H^{MODEL}	Model of the in-situ observations
$\mathcal{H}$	Observation operator, equal to H^{OBS} when there were observations, and zero otherwise
I^d	Presence-absence nine dimension indicator function
$\mathcal{L}^H$	Loss function used to minimize the distance between in-situ observations and predicted observations
$\mathcal{L}^z$	Loss function used to maximize the posterior probability $p_\Lambda(\tilde{z}^d\|\boldsymbol{y}^d,\boldsymbol{x}^d)$ for every day d
$\tilde{z}$	optical constituents with the change of variable $\tilde{z}=\log(z)$
Σ_ϵ	Covariance matrix of the Remote Sensing Reflectance
ϵ	Noise of the Remote Sensing Reflectance
δ_Λ	Perturbations on the parameters
$\nabla_{\tilde{z}^d}$	Gradient over every component of $\tilde{z}^d$
Σ_z	Covariance of the prior term associated to the optical constituents $\Sigma_z=\alpha\mathbb{1}$, $\mathbb{1}$ the identity matrix

https://doi.org/10.5194/gmd-2024-174
Preprint. Discussion started: 20 November 2024

Geoscientific Model Development Discussions
Open Access
EGU

Table D6. Table of Symbols and notation used for the SGVB formalism.

Symbol	Meaning
z	Latent variable sampled from an unknown distribution $p_{\Lambda^*}(z)$
$\boldsymbol{y}$	Random variable sampled from a known conditional distribution $p_{\Lambda^*}(\boldsymbol{y}\|z)$
$p_\Lambda(y)$	data-Likelihood of the parameter Λ
$p_\Lambda(z\|\boldsymbol{y})$	Posterior probability of the latent variable z
$q_\phi(z\|\boldsymbol{y})$	Estimate of the posterior probability of the latent variable z
$\mathcal{L}_{ELBO}$	ELBO loss function, where ELBO stands for "Evidence Lower Bound"
$D_{KL}(q_\phi(z\|\boldsymbol{y})\|\|p_\Lambda(z))$	Kullback-Leibler divergence between the two probability distributions $q_\phi(z\|\boldsymbol{y})$ and $p_\Lambda(z)$
$\mathbb{E}_{q_\phi(z\|\boldsymbol{y})}[log(p_\Lambda(\boldsymbol{y}\|z))]$	Expected value of $log(p_\Lambda(\boldsymbol{y}\|z))$ with respect to the probability distribution $q_\phi(z\|\boldsymbol{y})$

https://doi.org/10.5194/gmd-2024-174
Preprint. Discussion started: 20 November 2024

Geoscientific Model Development Discussions

Author contributions. CS implemented the code, performed the experiments, and wrote the first draft of the manuscript, MGDK did the data processing. PL and FA supervised the work. All authors collaborated on the design of the models and contributed to the manuscript's writing.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Financial support. This research has been supported by the European Commission HORIZON EUROPE Framework Programme (grant no. 101081273).

This research has been partly supported by the MED-MFC (Mediterranean – Monitoring Forecasting Centre) of the Copernicus Marine Service, which is implemented by Mercator Ocean International within the framework of a delegation agreement with the European Union (reference no. 21002L5-COP-MFC MED-5500).

https://doi.org/10.5194/gmd-2024-174
Preprint. Discussion started: 20 November 2024

References

Aas, E.: Two-stream irradiance model for deep waters, Applied Optics, 26, 2095–2101, 1987.

Aas, E. and Højerslev, N. K.: Analysis of underwater radiance observations: Apparent optical properties and analytic functions describing the angular radiance distribution, Journal of Geophysical Research: Oceans, 104, 8015–8024, 1999.

Ackleson, S. G., Balch, W. M., and Holligan, P. M.: Response of water-leaving radiance to particulate calcite and chlorophyll a concentrations: A model for Gulf of Maine coccolithophore blooms, Journal of Geophysical Research: Oceans, 99, 7483–7499, 1994.

Álvarez, E., Cossarini, G., Teruzzi, A., Bruggeman, J., Bolding, K., Ciavatta, S., Vellucci, V., d'Ortenzio, F., Antoine, D., and Lazzari, P.: Chromophoric dissolved organic matter dynamics revealed through the optimization of an optical-biogeochemical model in the NW Mediterranean Sea, Biogeosciences Discussions, 2023, 1–43, 2023.

Andrieu, C. and Thoms, J.: A tutorial on adaptive MCMC, Statistics and computing, 18, 343–373, 2008.

Antoine, D., Guevel, P., Deste, J.-F., Becu, G., Louis, F., Scott, A. J., and Bardey, P.: The "BOUSSOLE" buoy—A new transparent-to-swell taut mooring dedicated to marine optics: Design, tests, and performance at sea, Journal of Atmospheric and Oceanic Technology, 25, 968–989, 2008.

Arras, K. O.: An introduction to error propagation: derivation, meaning and examples of equation CY= FX CX FXT, Tech. rep., ETH Zurich, 1998.

Barron, J. T.: Continuously differentiable exponential linear units, arXiv preprint arXiv:1704.07483, 2017.

Barthélémy, S., Brajard, J., Bertino, L., and Counillon, F.: Super-resolution data assimilation, Ocean Dynamics, 72, 661–678, 2022.

Bocquet, M., Brajard, J., Carrassi, A., Bertino, L., et al.: Bayesian inference of chaotic dynamics by merging data assimilation, machine learning and expectation-maximization, FOUNDATIONS OF DATA SCIENCE, 2, 55–80, 2020.

Bodin, N., Burgeot, T., Stanisiere, J., Bocquené, G., Menard, D., Minier, C., Boutet, I., Amat, A., Cherel, Y., and Budzinski, H.: Seasonal variations of a battery of biomarkers and physiological indices for the mussel Mytilus galloprovincialis transplanted into the northwest Mediterranean Sea, Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology, 138, 411–427, 2004.

Boehme, L. and Rosso, I.: Classifying oceanographic structures in the Amundsen Sea, Antarctica, Geophysical Research Letters, 48, e2020GL089 412, 2021.

Boynton, G. C. and Gordon, H. R.: Irradiance inversion algorithm for estimating the absorption and backscattering coefficients of natural waters: Raman-scattering effects, Applied Optics, 39, 3012–3022, 2000.

Brajard, J., Jamet, C., Moulin, C., and Thiria, S.: Use of a neuro-variational inversion for retrieving oceanic and atmospheric constituents from satellite ocean colour sensor: Application to absorbing aerosols, Neural Networks, 19, 178–185, 2006.

Brankart, J.-M., Testut, C.-E., Béal, D., Doron, M., Fontana, C., Meinvielle, M., Brasseur, P., and Verron, J.: Towards an improved description of ocean uncertainties: effect of local anamorphic transformations on spatial correlations, Ocean Science, 8, 121–142, 2012.

Bruggeman, J., Bolding, K., Nerger, L., Teruzzi, A., Spada, S., Skákala, J., and Ciavatta, S.: EAT v0. 9.6: a 1D testbed for physical-biogeochemical data assimilation in natural waters, Geoscientific Model Development Discussions, 2023, 1–22, 2023.

Campbell, J. W.: The lognormal distribution as a model for bio-optical variability in the sea, Journal of Geophysical Research: Oceans, 100, 13 237–13 254, 1995.

Carmichael, G. R., Sandu, A., et al.: Sensitivity analysis for atmospheric chemistry models via automatic differentiation, Atmospheric Environment, 31, 475–489, 1997.

https://doi.org/10.5194/gmd-2024-174
Preprint. Discussion started: 20 November 2024

Chapman, C. and Charantonis, A. A.: Reconstruction of subsurface velocities from satellite observations using iterative self-organizing maps, IEEE Geoscience and Remote Sensing Letters, 14, 617–620, 2017.

Chib, S. and Greenberg, E.: Understanding the metropolis-hastings algorithm, The american statistician, 49, 327–335, 1995.

CMEMS: Mediterranean Sea, Bio-Geo-Chemical, L3, daily Satellite Observations (1997-ongoing). E.U. Copernicus Marine Service Information (CMEMS), doi: https://doi.org/10.48670/moi-00299, (Accessed on 11-13-2023).

Colella, S., Brando, V. E., Cicco, A. D., D'Alimonte, D., Forneris, V., and Bracaglia, M.: Ocean Colour Mediterranean and Black Sea Observation Product, ref: CMEMS-OC-QUID-009-141to144-151to154, url: https://catalogue.marine.copernicus.eu/documents/QUID/CMEMS-OC-QUID-009-141to144-151to154.pdf, 2023.

De Bézenac, E., Pajot, A., and Gallinari, P.: Deep learning for physical processes: Incorporating prior scientific knowledge, Journal of Statistical Mechanics: Theory and Experiment, 2019, 124 009, 2019.

Denvil-Sommer, A., Gehlen, M., Vrac, M., and Mejia, C.: LSCE-FFNN-v1: a two-step neural network model for the reconstruction of surface ocean p CO 2 over the global ocean, Geoscientific Model Development, 12, 2091–2105, 2019.

Desbruyères, D., Chafik, L., and Maze, G.: A shift in the ocean circulation has warmed the subpolar North Atlantic Ocean since 2016, Communications Earth & Environment, 2, 48, 2021.

Dessailly, D.: Retrieval of the spectral diffuse attenuation coefficient Kd (λ) in open and coastal ocean waters using a neural network inversion, Journal of Geophysical Research: Oceans, 117, 2012.

Doersch, C.: Tutorial on Variational Autoencoders, 2021.

Dutkiewicz, S., Hickman, A., Jahn, O., Gregg, W., Mouw, C., and Follows, M.: Capturing optically important constituents and properties in a marine biogeochemical and ecosystem model, Biogeosciences, 12, 4447–4481, 2015.

Erichson, N. B., Muehlebach, M., and Mahoney, M. W.: Physics-informed autoencoders for Lyapunov-stable fluid flow prediction, arXiv preprint arXiv:1905.10866, 2019.

Erickson, Z. K., McKinna, L., Werdell, P. J., and Cetinić, I.: Bayesian approach to a generalized inherent optical property model, Optics Express, 31, 22 790–22 801, 2023.

Falkner, S., Klein, A., and Hutter, F.: BOHB: Robust and efficient hyperparameter optimization at scale, in: International conference on machine learning, pp. 1437–1446, PMLR, 2018.

Geider, R., MacIntyre, H., and Kana, T.: Dynamic model of phytoplankton growth and acclimation: responses of the balanced growth rate and the chlorophyll a: carbon ratio to light, nutrient-limitation and temperature, Marine Ecology Progress Series, 148, 187–200, 1997.

George, T. M., Manucharyan, G. E., and Thompson, A. F.: Deep learning to infer eddy heat fluxes from sea surface height patterns of mesoscale turbulence, Nature communications, 12, 800, 2021.

Gordon, H. R.: Inverse methods in hydrologic optics, Oceanologia, 44, 2002.

Gordon, H. R. and Boynton, G. C.: Radiance—irradiance inversion algorithm for estimating the absorption and backscattering coefficients of natural waters: homogeneous waters, Applied optics, 36, 2636–2641, 1997.

Gregg, W. W. and Casey, N. W.: Skill assessment of a spectral ocean–atmosphere radiative model, Journal of Marine Systems, 76, 49–63, 2009.

Ham, Y.-G., Kim, J.-H., and Luo, J.-J.: Deep learning for multi-year ENSO forecasts, Nature, 573, 568–572, 2019.

Irrgang, C., Saynisch, J., and Thomas, M.: Estimating global ocean heat content from tidal magnetic satellite observations, Scientific Reports, 9, 7893, 2019.

Jones, D. and Ito, T.: Gaussian mixture modeling describes the geography of the surface ocean carbon budget., 2019.

https://doi.org/10.5194/gmd-2024-174
Preprint. Discussion started: 20 November 2024

Jones, D. C., Holt, H. J., Meijers, A. J., and Shuckburgh, E.: Unsupervised clustering of Southern Ocean Argo float temperature profiles, Journal of Geophysical Research: Oceans, 124, 390–402, 2019.

Kingma, D. P. and Welling, M.: Auto-encoding variational bayes, arXiv preprint arXiv:1312.6114, 2013.

Lachkar, Z. and Gruber, N.: A comparative study of biological production in eastern boundary upwelling systems using an artificial neural network, Biogeosciences, 9, 293–308, 2012.

Lazzari, P., Solidoro, C., Ibello, V., Salon, S., Teruzzi, A., Béranger, K., Colella, S., and Crise, A.: Seasonal and inter-annual variability of plankton chlorophyll and primary production in the Mediterranean Sea: a modelling approach, Biogeosciences, 9, 217–233, 2012.

Lazzari, P., dit Kacem, G., M, A. E., Chernov, I., and Vellucci, V.: Determination of Biogeochemical Properties in Sea Waters Using the Inversion of the Three-stream Irradiance Model, Scientific Reports, https://doi.org/https://doi.org/10.1038/S41598-024-71457-5, 2024.

Leathers, R. A., Roesler, C. S., and McCormick, N. J.: Ocean inherent optical property determination from in-water light field measurements, Applied Optics, 38, 5096–5103, 1999.

Lee, Z., Carder, K. L., and Arnone, R. A.: Deriving inherent optical properties from water color: a multiband quasi-analytical algorithm for optically deep waters, Applied optics, 41, 5755–5772, 2002.

Liaw, R., Liang, E., Nishihara, R., Moritz, P., Gonzalez, J. E., and Stoica, I.: Tune: A research platform for distributed model selection and training, arXiv preprint arXiv:1807.05118, 2018.

Longhurst, A., Sathyendranath, S., Platt, T., and Caverhill, C.: An estimate of global primary production in the ocean from satellite radiometer data, Oceanographic Literature Review, 2, 203, 1996.

Manucharyan, G. E., Siegelman, L., and Klein, P.: A deep learning approach to spatiotemporal sea surface height interpolation and estimation of deep currents in geostrophic ocean turbulence, Journal of Advances in Modeling Earth Systems, 13, e2019MS001 965, 2021.

Martinez, E., Gorgues, T., Lengaigne, M., Fontana, C., Sauzède, R., Menkes, C., Uitz, J., Di Lorenzo, E., and Fablet, R.: Reconstructing global chlorophyll-a variations using a non-linear statistical approach, Frontiers in Marine Science, 7, 464, 2020.

Mason, J. D., Cone, M. T., and Fry, E. S.: Ultraviolet (250–550 nm) absorption spectrum of pure water, Applied optics, 55, 7163–7172, 2016.

Maze, G., Mercier, H., Fablet, R., Tandeo, P., Radcenco, M. L., Lenca, P., Feucher, C., and Le Goff, C.: Coherent heat patterns revealed by unsupervised classification of Argo temperature profiles in the North Atlantic Ocean, Progress in Oceanography, 151, 275–292, 2017.

McCormick, N.: Analytical transport theory applications in optical oceanography, Annals of Nuclear Energy, 23, 381–395, 1996.

Michalopoulou, Z.-H., Bagheri, S., and Axe, L.: Bayesian estimation of optical properties of nearshore estuarine waters: A Gibbs sampling approach, IEEE transactions on geoscience and remote sensing, 48, 1579–1587, 2009.

Morel, A.: Optical properties of pure water and pure sea water, Optical Aspects of Oceanography, 1, 1974.

Mustapha, Z. B., Alvain, S., Jamet, C., Loisel, H., and Dessailly, D.: Automatic classification of water-leaving radiance anomalies from global SeaWiFS imagery: application to the detection of phytoplankton groups in open ocean waters, Remote sensing of environment, 146, 97–112, 2014.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S.: PyTorch: An Imperative Style, High-Performance Deep Learning Library, arXiv preprint arXiv:1912.01703, 2019.

Pietropolli, G., Cossarini, G., and Manzoni, L.: GANs for integration of deterministic model and observations in marine ecosystem, in: EPIA Conference on Artificial Intelligence, pp. 452–463, Springer, 2022.

Richardson, A. J., Risien, C., and Shillington, F. A.: Using self-organizing maps to identify patterns in satellite imagery, Progress in Oceanography, 59, 223–239, 2003.

https://doi.org/10.5194/gmd-2024-174
Preprint. Discussion started: 20 November 2024

Rodgers, C. D.: Inverse methods for atmospheric sounding: theory and practice, vol. 2, World scientific, 2000.

Ronald, J. and Zaneveld, V.: Remotely sensed reflectance and its dependence on vertical structure: a theoretical derivation, Applied Optics, 21, 4146–4150, 1982.

Salama, M. S. and Verhoef, W.: Two-stream remote sensing model for water quality mapping: 2SeaColor, Remote sensing of Environment, 157, 111–122, 2015.

Saraceno, M., Provost, C., and Lebbah, M.: Biophysical regions identification using an artificial neuronal network: A case study in the South Western Atlantic, Advances in Space Research, 37, 793–805, 2006.

Shlens, J.: Notes on kullback-leibler divergence and likelihood, arXiv preprint arXiv:1404.2000, 2014.

Shmakov, A., Greif, K., Fenton, M., Ghosh, A., Baldi, P., and Whiteson, D.: End-To-End Latent Variational Diffusion Models for Inverse Problems in High Energy Physics, Advances in Neural Information Processing Systems, 36, 2024.

Sohn, K., Lee, H., and Yan, X.: Learning structured output representation using deep conditional generative models, Advances in neural information processing systems, 28, 2015.

Soto, C.: carlossoto362/OGS_one_d_model: Inversion Model for 1D Bio-optical model of Remote Sensing Reflectance, https://doi.org/10.5281/zenodo.13741206, 2024.

Stramska, M., Stramski, D., Mitchell, B. G., and Mobley, C. D.: Estimation of the absorption and backscattering coefficients from in water radiometric measurements, Limnology and Oceanography, 45, 628–641, 2000.

Tao, Z., McCormick, N. J., and Sanchez, R.: Ocean source and optical property estimation from explicit and implicit algorithms, Applied Optics, 33, 3265–3275, 1994.

Van Rossum, G. and Drake, F. L.: Introduction to python 3: python documentation manual part 1, CreateSpace, 2009.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors: SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python, Nature Methods, 17, 261–272, https://doi.org/10.1038/s41592-019-0686-2, 2020.

Zaneveld, J. R. V.: An asymptotic closure theory for irradiance in the sea and its inversion to obtain the inherent optical properties, Limnology and Oceanography, 34, 1442–1452, 1989.

Zhao, Z., Ye, J. C., and Bresler, Y.: Generative Models for Inverse Imaging Problems: From mathematical foundations to physics-driven applications, IEEE Signal Processing Magazine, 40, 148–163, 2023.

Zhong, E. D., Bepler, T., Davis, J. H., and Berger, B.: Reconstructing continuous distributions of 3D protein structure from cryo-EM images, arXiv preprint arXiv:1909.05215, 2019.

Zhong, E. D., Lerer, A., Davis, J. H., and Berger, B.: Cryodrgn2: Ab initio neural reconstruction of 3d protein structures from real cryo-em images, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 4066–4075, 2021.