

# SpecWa: Spectral remote sensing data and chlorophyll *a* values of inland waters

(<https://doi.org/10.5880/fidgeo.2020.036>)

Philipp M. Maier<sup>1</sup> and Sina Keller<sup>1</sup>

<sup>1</sup> *Karlsruhe Institute of Technology, Institute of Photogrammetry and Remote Sensing, Karlsruhe, Germany*

## Table of contents

|                                                                              |   |
|------------------------------------------------------------------------------|---|
| 1. Licence .....                                                             | 1 |
| 2. Citation.....                                                             | 1 |
| 3. Abstract.....                                                             | 2 |
| 4. Data description.....                                                     | 2 |
| 4.1. Description of the equipment.....                                       | 2 |
| 4.2. Experimental setup.....                                                 | 3 |
| 4.3. Measured water quality parameters .....                                 | 4 |
| 4.4. Site properties.....                                                    | 4 |
| 4.5. Pre-Processing of the spectrometer data.....                            | 5 |
| 5. File description.....                                                     | 6 |
| 5.1. File inventory .....                                                    | 6 |
| 5.2. Description of data tables .....                                        | 6 |
| 5.2.1. '2018_spectral_reference.csv' and '2019_spectral_reference.csv' ..... | 6 |
| 5.3. Acknowledgments.....                                                    | 7 |
| 6. References .....                                                          | 7 |

## 1. Licence

Creative Commons Attribution 4.0 International License (CC BY 4.0)



## 2. Citation

**When using the data please cite:**

Maier, Philipp M.; Keller, Sina (2020): SpecWa: Spectral remote sensing data and chlorophyll *a* values of inland waters. GFZ Data Services. <https://doi.org/10.5880/fidgeo.2020.036>

**The data are supplementary to:**

Maier, P. M. and Keller, S. (2019a). Estimating chlorophyll *a* concentrations of several inland waters with hyperspectral data and machine learning model. ISPRS Ann. Photogrammetry, Remote Sensing and Spatial Information Science IV-2/W5, pp. 609–614. <https://doi.org/10.5194/isprs-annals-IV-2-W5-609-2019>

Maier, P. M. and Keller, S. (2019b). Application of different simulated spectral data and machine learning to estimate the chlorophyll *a* concentration of several inland eaters. In: 2019 10th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS), Amsterdam, Netherlands, 2019, pp. 1-5, <https://doi.org/10.1109/WHISPERS.2019.8921073>

### 3. Abstract

This dataset contains data of measurements at several inland water bodies in the region around Karlsruhe, Germany, including spectral data and mainly chlorophyll *a* values. The measurements are motivated by the idea to link remote sensing data and in-situ acquired data for an estimation of selected water parameters solely based on spectral data with data-driven machine learning approaches. To address this topic, we monitored the selected inland water bodies with a spectrometer as a remote sensing device and two in-situ measurement devices to collect among others the chlorophyll *a* values during the summers of the years 2018 and 2019. The eleven selected water bodies are characterized by a relatively small size and a variety of different chlorophyll *a* concentrations.

In sum, the dataset of the year 2018 consists of 1,305 datapoints and the dataset of the year 2019 includes 2,830 datapoints. Each datapoint is defined by the spectral data in the range of 389 nm to 910 nm and all in-situ measured reference values such as the chlorophyll *a* concentration of the respective year. Depending on the applied in-situ measurement device, the chlorophyll *a* concentration or additional water parameters were either measured directly on-site or taken as water samples and analyzed in the laboratory.

We provide the data in nearly raw format including information about the water bodies. The data are organized along the year of the measurements. In addition, exemplary scripts for reading and processing the data are included.

### 4. Data description

We conducted measurements at several inland water bodies around Karlsruhe, Germany, between June and October in the years 2018 and 2019. The locations of the selected water bodies are given in Table 2. The measurements are motivated by the idea to link remote sensing data and in-situ measurements of water bodies for an estimation of selected water parameters solely based on spectral data. To address this topic, we monitored with a RoX spectrometer (JB Hyperspectral, 2020) as remote sensing device and two in-situ measurement devices: the AlgaeLabAnalyser and the AlgaeTorch (bbe moldaenke, 2020a and b). These two devices measure concentrations of specific water parameters such as chlorophyll *a*, diatoms, or green algae. The water bodies, that we have chosen, are characterized by a relatively small size and a variety of different concentrations of the respective water parameters.

Maier & Keller (2019a) have already shown on the 2018-subset of the SpecWa dataset that the chlorophyll *a* concentration is predictable by the spectral data with machine learning models (Maier et al., 2019a). Moreover, Maier & Keller (2019b) demonstrated that scaling this high resolved dataset to different spectral satellite resolutions still leads to feasible results when estimating the chlorophyll *a* concentration with machine learning models (Maier et al., 2019b).

#### 4.1. Description of the equipment

To measure the spectral data of the water bodies, we applied a RoX spectrometer (see JP Hyperspectral Devices, 2020) covering a spectral range of 341 nm to 1014 nm with an intermediate sampling interval of 0.65 nm. It consists of two fiber optic cables. One fiber optic cable has a cosine receptor at its top with a field of view of 180°, while the other one has a field of view of 25°. The integration time of the sensor is determined by the incoming radiation on the cosine receptor. This is very useful for measuring under various angles of the sun or changing cloud conditions. The final spectral result is the reflectance, the ratio of the reflected radiance to the incoming irradiance on the cosine receptor for each wavelength. The RoX spectrometer was calibrated by the company in the laboratory against an Ulbricht sphere.

The water parameters were monitored by the AlgaeLabAnalyser (bbe moldaenke, 2018) in 2018 and by the AlgaeTorch (bbe moldaenke, 2019) in 2019. The AlgaeLabAnalyser is a spectral fluorometer. It measures chlorophyll *a* in a range of 0 µg/L to 200 µg/L with a resolution of 0.01 µg/L. In addition, it is able to distinguish the following algae classes: green algae, diatoms, cryptophytaceae, and

cyanobacteria. Moreover, it measures CDOM (bbe moldaenke, 2018). The AlgaeLabAnalyser is a device for measuring in the laboratory. Therefore, we collected samples at the water bodies. In the laboratory, each of the samples was analyzed in the device where they are stimulated by various LEDs. The AlgaeTorch, however, uses a measurement technique to monitor chlorophyll *a* directly in the water. It measures the chlorophyll *a* concentration in the range of 0  $\mu\text{g/L}$  to 200  $\mu\text{g/L}$  with a resolution of 0.01  $\mu\text{g/L}$ . The device is placed in the water and after a measurement time of about 15 seconds, the values of the concentrations are stored and displayed. It measures the intensity of fluorescence by the stimulation of the algae with different LEDs. In addition, it measures turbidity and the proportion of chlorophyll related to cyanobacteria (see bbe moldaenke, 2019).

## 4.2. Experimental setup

The intention of the measurements was to collect hyperspectral on-site data of different water bodies as well as the corresponding near-time water parameters on local scale. For the recording of the hyperspectral data, we installed the RoX spectrometer on a tripod. The fiber cable with the cosine receptor was directed perpendicular to the sky while the bare fiber cable pointed to the water surface, hence in the opposite direction (see Figure 1). Depending on the water body, we applied two kinds of installation for the tripod and the spectrometer: when monitoring a natural water body with flat littoral zone, we installed the tripod as far inside the water as it was possible with respect to the water level (see Figure 1, left), otherwise when monitoring an artificial water body without a flat littoral zone, we installed the tripod outside the water (see Figure 1, right). In addition, we aimed to minimize the influence of the ground during the hyperspectral data acquisition as well as other influencing factors such as shadows of surrounding trees or buildings.

The sampling frequency of the RoX was set to 15 seconds in case of the year 2019 measurements and to 20 seconds in case of the year 2018 measurements. That sampling frequency resulted in three to four spectra per minute. During the measurements in the year of 2018, we took a reference water sample for the measurement of the water parameters every fifth minute which we later analyzed with the AlgaeLabAnalyser in the laboratory. During the measurements in the year of 2019, we sampled in-situ measurements of the water parameters approximately every second minute with the AlgaeTorch. The water samples were collected in a depth of 20 centimeters below the water surface and as close as possible to the spectrometer measurements. Thus, we ensured to record the same water conditions with the remote sensing sensor and the in-situ probes. At each water body, we switched the devices to different spots in the water. Furthermore, we aimed to conduct our measurements under optimal conditions with clear sky and without clouds, nevertheless during a few measurements slight cloud cover occurred.



*Figure 1: Measurement setup of the RoX spectrometer at a natural water body (left) and an artificial pond (right), taken from Maier & Keller (2019a).*

### 4.3. Measured water quality parameters

Table 1 summarizes the measured water parameters. Besides the chlorophyll *a* concentration which is represented by the different algae species, we monitored turbidity and CDOM. Since the calibration range for the chlorophyll *a* concentration is between 0 µg/L and 200 µg/L, we removed the chlorophyll *a* values above 200 µg/L. Note that we monitored different water parameters during the measurements in the years of 2018 and 2019 due to the characteristics of the two different in-situ measurement devices.

Table 1: Description of the measured water parameters of the experiment during 2018 and 2019.

| Parameter            | Unit | Description                                    | Year        |
|----------------------|------|------------------------------------------------|-------------|
| Chlorophyll <i>a</i> | µg/L | Measured concentration of chlorophyll <i>a</i> | 2018 & 2019 |
| Cyanobacteria        | µg/L | Measured concentration of cyanobacteria        | 2018 & 2019 |
| Green algae          | µg/L | Measured concentration of green algae only     | 2018        |
| Diatoms              | µg/L | Measured concentration of diatoms only         | 2018        |
| Cryptophytaceae      | µg/L | Measured concentration of cryptophytaceae      | 2018        |
| Turbidity            | FTU  | Measured turbidity                             | 2019        |
| CDOM                 | µg/L | Measured concentration of CDOM                 | 2018        |

### 4.4. Site properties

We monitored different chlorophyll *a* concentrations at several water bodies during the selected months in the years 2018 and 2019. During the measurements in the year 2018, we selected eleven different inland waters while during the measurements in the year 2019, we focused on six specific water bodies. However, we increased the measurement frequency of these six water bodies in the year 2019 compared to the measurement frequency of the year 2018.

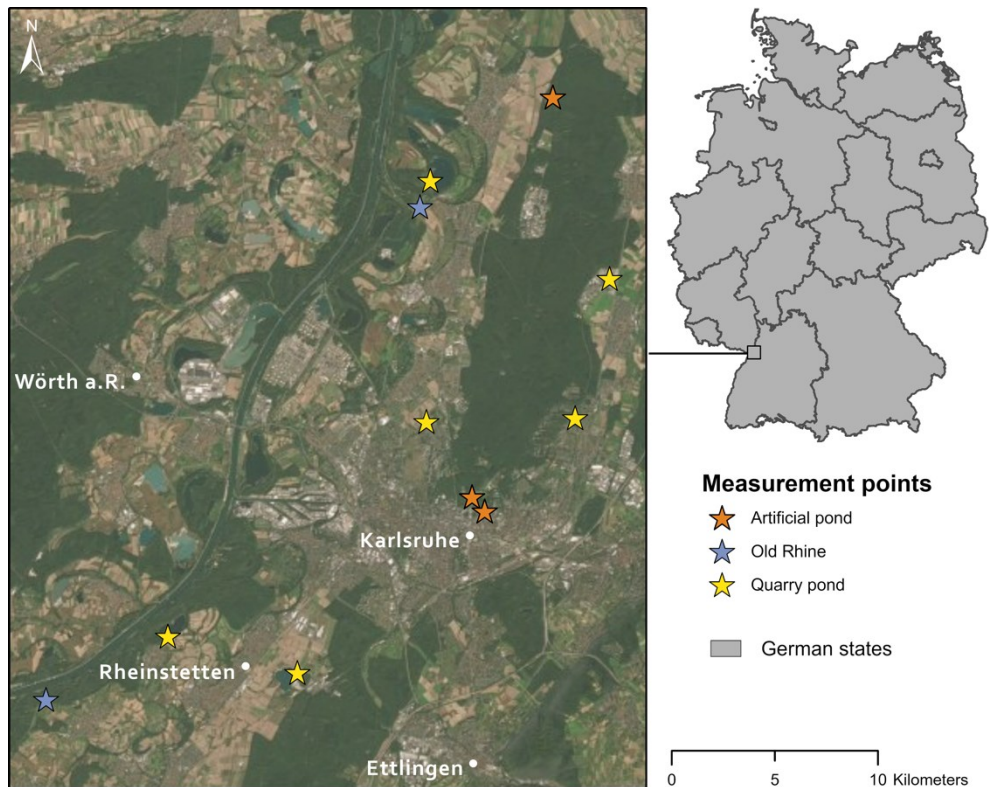


Figure 2: Map of the study area including the water bodies which were monitored. Map projection: WSG 1984.

In the entire dataset, we identified the three artificial ponds as eutrophic water bodies with the highest concentrations of chlorophyll *a*. On the contrary, the quarry pond “Leopoldshafen” is characterized by a very low chlorophyll *a* concentration. The chlorophyll *a* concentration of the other water bodies lays in-between.

*Table 2: Brief description of the water bodies including an overview of the measurement setup. The number of datapoints refers to the number of spectral RoX measurements assigned to the respective values of the water parameters.*

| Name                       | Lon    | Lat     | Number of datapoints | Depth of measurement in m | Chlorophyll <i>a</i> range in µg/L |
|----------------------------|--------|---------|----------------------|---------------------------|------------------------------------|
| ap_castle garden karlsruhe | 8.4049 | 49.0170 | 1,322                | 1 to 2                    | 6.3 to 171.1                       |
| ap_KIT                     | 8.4104 | 49.0129 | 907                  | 0.5 to 1                  | 22.2 to 199.6                      |
| ap_tmb                     | 8.4401 | 49.1312 | 60                   | 2 to 3                    | 61.0 to 100.5                      |
| old_rhine_au               | 8.2193 | 48.9590 | 21                   | 2 to 3                    | 4.7 to 9.1                         |
| old_rhine_leopolds hafen   | 8.3822 | 49.0999 | 8                    | 0.5 to 1                  | 9.8 to 11.0                        |
| qp_blankenloch             | 8.4648 | 49.0794 | 494                  | 0.5 to 3                  | 2.4 to 21.0                        |
| qp_epple                   | 8.3288 | 48.9667 | 42                   | 1 to 3                    | 1.6 to 13.0                        |
| qp_ferma                   | 8.2724 | 48.9771 | 20                   | 1 to 3                    | 3.3 to 6.4                         |
| qp_heide                   | 8.3850 | 49.0385 | 221                  | 1 to 3                    | 1.7 to 16.5                        |
| qp_leopoldshafen           | 8.3867 | 49.1074 | 105                  | 1.5 to 3                  | 0.0 to 8.7                         |
| qp_waldstadt               | 8.4498 | 49.0396 | 485                  | 1.5 to 3                  | 0.0 to 17.0                        |

#### 4.5. Pre-Processing of the spectrometer data

As a first pre-processing step, we dropped the spectral data at the edge of the spectral range of the RoX spectrometer due to e.g. sensor noise and atmospheric disturbances. This results in spectral data in the range of 389 nm to 910 nm. Furthermore, we excluded spectra which we identified as outliers such as spectra including waves at the water surface or spectra measured during cloudy conditions. To identify the outlying spectra, we first calculated the median spectrum for each waterbody at each specific measurement date. Then, we compared the respective spectrum to the median spectrum and excluded it if the specific spectrum was higher than twice of the median spectrum or lower than half of the median spectrum. Since the measurement conditions were carefully chosen, the number of outliers is low.

In the last pre-processing step, the spectra were assigned to the measured values of the water parameters using the respective time stamp. Since the sampling frequency of the spectral RoX sensor was higher than the sampling frequency of the probe measurements, some spectra were aligned to the same probe measurements. In total, the dataset for the year 2018 consists of 1305 datapoints (see ['2018\\_spectral\\_reference.csv'](#)) and the dataset for 2019 consists of 2380 datapoints (see ['2019\\_spectral\\_reference.csv'](#)). We define one datapoint as the spectral data in the range of 389 nm to 910 nm and all respective reference values such as the chlorophyll *a* concentration and turbidity. The spectral data and the measured values of the water parameter of the given \*.csv-files can be read in with the python script ['read\\_in\\_dataset\\_water.ipynb'](#) or the R-script ['read\\_in\\_dataset\\_water\\_R.ipynb'](#).

## 5. File description

In this section, the details to the provided files of the “Spectral remote sensing data and chlorophyll a values of inland waters” dataset are provided.

### 5.1. File inventory

- `'2018_spectral_reference.csv'` and `'2019_spectral_reference.csv'` contain the spectral remote sensing data and the measured values of the water parameters of the respective water body and measurement experiment.
- The spectral data and the measured values of the water parameter of the given \*.csv-files can be read in with the python script `'read_in_dataset_water.ipynb'` or the R-script `'read_in_dataset_water_R.ipynb'`.
- In addition, a `'SpecWa_dataset_bibliography.bib'` file is provided.

### 5.2. Description of data tables

The spectral remote sensing data and the measured values of the water parameters of the water bodies are stored in two files according to the years.

#### 5.2.1. `'2018_spectral_reference.csv'` and `'2019_spectral_reference.csv'`

These files contain the spectral RoX data and the values of the respective water parameters of the water bodies.

| Column header       | Unit                   | Description                                                                                                                                              |
|---------------------|------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------|
| datetime            | YYYY-MM-DD<br>HH:mm:SS | Measurement date and time of the spectrometer data                                                                                                       |
| year                |                        | In 2018, the water parameter values were measured with the AlgaeLabAnalyzer while in 2019, the water parameter values were measured with the AlgaeTorch. |
| name                |                        | Name of the water body                                                                                                                                   |
| lon                 | dd                     | Longitude                                                                                                                                                |
| lat                 | dd                     | Latitude                                                                                                                                                 |
| chl_a               | µg/L                   | Measured concentration of chlorophyll a                                                                                                                  |
| cyano               | µg/L                   | Measured concentration of cyano bacteria                                                                                                                 |
| green_algae         | µg/L                   | Measured concentration of green algae, available for 2018                                                                                                |
| diatoms             | µg/L                   | Measured concentration of diatoms, available for 2018                                                                                                    |
| cryptophyta         | µg/L                   | Measured concentration of cryptophyta, available for 2018                                                                                                |
| turbidity           | FTU                    | Measured turbidity, available for 2019                                                                                                                   |
| CDOM                | µg/L                   | Measured concentration of CDOM, available for 2018                                                                                                       |
| 389.35 to<br>910.32 |                        | Reflectance between 0 and 1 at the respective wavelength of 389.35 nm to 910.32                                                                          |

### 5.3. Acknowledgments

We thank our project partner bbe moldaenke for lending us the AlgaeLabAnalyser and the AlgaeTorch in the context of the WAQUAVID project funded by the German Federal Ministry of Education and Research (BMBF). We also thank Stefan Hinz for his support of our study and our scientific assistants Philipp Wagner, Eva Klier, Ian Steegmayer, and Oliver Sefrin as well as our trainee Sophia Helgert for their support during the field work.

### 6. References

- bbe moldaenke (2018) AlgaeLabAnalyzer: <https://www.bbe-moldaenke.de/de/produkte/chlorophyll/details/algaetorch.html> (last accessed 29<sup>th</sup> July 2020)
- bbe moldaenke (2019) AlgaeTorch: <https://www.bbe-moldaenke.de/de/produkte/chlorophyll/details/algaelabanalyser.html> (last accessed 29<sup>th</sup> July 2020)
- JB Hyperspectral Devices (2020) RoX – Reflectance box: <https://www.jb-hyperspectral.com/products/rox/> (last accessed 29<sup>th</sup> July 2020)
- Maier, P. M. and Keller, S. (2019a). Estimating chlorophyll *a* concentrations of several inland waters with hyperspectral data and machine learning model. ISPRS Ann. Photogrammetry, Remote Sensing and Spatial Information Science IV-2/W5, pp. 609–614. <https://doi.org/10.5194/isprs-annals-IV-2-W5-609-2019>
- Maier, P. M. and Keller, S. (2019b). Application of different simulated spectral data and machine learning to estimate the chlorophyll *a* concentration of several inland eaters. In: 2019 10th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS), Amsterdam, Netherlands, 2019, pp. 1-5, <https://doi.org/10.1109/WHISPERS.2019.8921073>