MODELLING LIGHT CONDITIONS IN DANISH COASTAL WATERS USING A BAYESIAN MODELLING APPROACH

Model documentation

Scientific Report from DCE – Danish Centre for Environment and Energy No. 422 2021

AARHUS UNIVERSITY
DCE – DANISH CENTRE FOR ENVIRONMENT AND ENERGY

MODELLING LIGHT CONDITIONS IN DANISH COASTAL WATERS USING A BAYESIAN MODELLING APPROACH

Model documentation

Scientific Report from DCE – Danish Centre for Environment and Energy No. 422 2021

Jesper Philip Aagaard Christensen[1]
Nisha Shetty[1]
Nikolaj Reducha Andersen[1]
Christian Damgaard[1]
Karen Timmermann[2]

[1]Aarhus University, Department of Bioscience
[2]DTU, Institute of Aquatic Resources

AARHUS UNIVERSITY
DCE – DANISH CENTRE FOR ENVIRONMENT AND ENERGY

Data sheet

Series title and no.:	Scientific Report from DCE – Danish Centre for Environment and Energy No. 422
Category:	Scientific advisory report
Title:	Modelling light conditions in Danish coastal waters using a Bayesian modelling approach
Subtitle:	Model documentation
Authors:	Jesper Philip Aagaard Christensen[1], Nisha Shetty[1], Nikolaj Reducha Andersen[1], Christian Damgaard[1] and Karen Timmermann[2]
Institutions:	[1]Aarhus University, Department of Bioscience, [2]DTU, Institute of Aquatic Resources
Publisher:	Aarhus University, DCE – Danish Centre for Environment and Energy ©
URL:	http://dce.au.dk/en
Year of publication:	October 2021
Editing completed:	January 2021
Referee:	Jacob Carstensen
Quality assurance, DCE:	Anja Skjoldborg
External comments:	The Danish Environmental Protection Agency. The comments can be found here: http://dce2.au.dk/pub/komm/SR422_komm.pdf
Financial support:	The Danish Environmental Protection Agency
Please cite as:	Christensen, J.P.A., Shetty, N., Andersen, N.R., Damgaard, C. & Timmermann, K. 2021. Modelling light conditions in Danish coastal waters using a Bayesian modelling approach. Model documentation. Aarhus University, DCE – Danish Centre for Environment and Energy, 48 pp. Scientific Report No. 422 http://dce2.au.dk/pub/SR422.pdf
	Reproduction permitted provided the source is explicitly acknowledged

Abstract:	During this project, a number of Bayesian models were developed to model the light limiting depth distribution of eelgrass in Danish coastal waterbodies. Eelgrass is a marine ecosystem key species and regulated by the EU water frame directive. Here we model the maximum depth limit of eelgrass based on light limitation. We have assumed that the regional light limitation for eelgrass is around 16 % of surface irradiance. The light climate is calculated based on the national environmental monitoring data and modelled as a function of nutrient loadings, climate and physio-chemical measures of the water. In total, we have made models for 44 monitoring stations, covering 41 water bodies; 23 of the models include nutrient loadings from the local catchment and hence they can be used to regulate nutrient loadings from land.

Keywords:	Eelgrass, light limitation, depth distribution, Bayesian modelling, Water frame directive
Layout:	Graphic Group, AU Silkeborg
Front page photo:	Peter Bondo Christensen
ISBN:	978-87-7156-556-0
ISSN (electronic):	2245-0203
Number of pages:	48
Internet version:	The report is available in electronic format (pdf) at http://dce2.au.dk/pub/SR422.pdf

Contents

Summary

Bayesian hierarchical models were developed to predict light limitation depth for eelgrass in Danish waterbodies as a function of land-based nitrogen and phosphorus loadings, physicochemical and climatic predictors. The objective of the model development was to support the Danish implementation of the Water Framework Directive (WFD) by providing tools applicable for estimating potential depth distribution of eelgrass and subsequent to calculate maximum allowable nutrient input (MAI) to the Danish coastal waters covered by the WFD.

The applied light limitation depth was calculated from light profiles as the maximum water depth, where at least 16% of surface light was available for benthic primary production from March to October. Bayesian single-station models for the light indicator were developed for 44 Danish water quality monitoring stations representing 41 waterbodies.

For the resulting set of Bayesian models we found that nutrient input was selected as predictor-variable in 23 of the models. As expected, we found a negative slope coefficient between nutrient inputs and depth of light limitation depth in all but 4 stations.

Model evaluation plots and performance statistics revealed that most of the models could capture the levels and year-to-year variation in the depth of light limitation depth reasonably well, indicating that the models can produce reliable predictions of light conditions in Danish coastal waters.

Dansk resume

Vi har udviklet bayesianske hierarkiske modeller, der beskriver den lysbegrænsende dybde for ålegræs i danske vandområder med tilstrækkeligt datagrundlag, som funktion af landbaseret kvælstof- og fosforbelastning, fysisk-kemiske og klimatiske prædiktorer. Formålet med modeludviklingen var at tilvejebringe værktøjer, der kan anvendes til at estimere dybdeudbredelsen af ålegræs samt bestemme den maksimale næringsstoftilførsel, som vil sikre målopfyldelse (målbelastninger) til de danske farvande, der er omfattet af vandrammedirektivet.

Den anvendte indikator for lysbegrænsningsdybde blev beregnet ud fra lysprofiler, som den maksimale vanddybde, hvor mindst 16 % af overfladelyset var tilgængeligt for bentisk primærproduktion i den produktive periode fra marts til oktober. Vi har udviklet separate modeller for 44 danske overvågningsstationer i det nationale overvågningsprogram NOVANA, som repræsenterede 41 vandområder.

For enkeltstationsmodellerne fandt vi, at næringsstoftilførsel var den bedste variable til estimering af tilgængelighed af lys ved havbunden på 23 stationer. Som forventet fandt vi en negativ hældningskoefficient mellem næringsstoftilførsel og lysforhold ved havbunden for næsten alle stationer, med undtagelse af fire stationer.

Modelevalueringsplot og model-performance-statistik viste, at de fleste modeller kunne fange niveauer og år-til-år variationen i lysforhold ved havbunden i de enkelte vandområder, og at der var en rimelig god overensstemmelse mellem modelestimater og målinger. Dette indikerer, at modellerne kan anvendes til modelscenarier for lysforhold i danske kystvande.

1 Introduction

In this study, a Bayesian approach was used to develop light models to support the establishment of the Danish River Basin Management Plans 2021-2027 (RBMP 2021-27) as part of the implementation of the Water Framework Directive (WFD).

Light availability is crucial for marine ecosystem functioning and is, to a large extend, controlling the depth distribution of macrophytes (Binzer et al. 2006), including depth distribution of eelgrass (Short & Wyllie-Echeverria, 1996; Orth et al. 2006; Waycott et al. 2009; Krause-Jensen et al. 2011). The depth distribution of eelgrass is an intercalibrated indicator for the biological quality elements "angiosperms and macroalgae," and it is essential for assessing the ecological status of coastal waters as part of the WFD. The depth distribution of eelgrass responds to changes in eutrophication (van Katwijk et al. 2010, Vaudrey et al 2010), but often with considerable lag-times as deterioration and especially re-establishment of eelgrass is a lengthy process even when favourable environmental conditions are present (McGlathery et al 2012). Furthermore, the distribution of eelgrass depends on several factors, such as substrate conditions and the amount of grazers, for which forcing data is often lacking. As a consequence, statistical models are not suitable to describe and predict the depth distribution of eelgrass. Instead, light limitation depth is used as a proxy indicator for the potential depth limit of eelgrass. Sufficient light is a prerequisite for growth of benthic vegetation, but even if adequate light is available, benthic vegetation might be limited by other factors or delayed in the re-establishment.

Bayesian statistics has gained popularity in recent years, as modern computing power has facilitated the development of new algorithms and model types. Further, it was recommended, as a tool, by the panel of international experts evaluating models and methods used for the Danish RBMP 2015-2021 (Herman et al. 2017). In Bayesian statistics, parameter estimates are not considered point-estimates, but rather expressed as probability distributions, involving more uncertainty information. Parameter uncertainty is quantified using the prior knowledge from earlier studies or literature, along with the sample data, which is not the case in frequentist or classical statistics where parameter estimates are inferred from sample data only. For ecological models, it is essential to have information on the uncertainty in the parameter estimates and the resulting model predictions (Beck 1987; Ellison 1996; Omlin & Reichert 1999). Thus, the Bayesian perspective is more comprehensive, and incorporation of prior knowledge makes it more consistent in terms of the scientific process of progressive learning (Germano 1999) and the policy practice of adaptive management (Walters 1986). Bayesian statistics has been used previously in water quality modelling (Malve & Qian 2006; Gronewold et al. 2010; Cha et al. 2016).

The main purpose of this study was to develop reliable Bayesian models for the light indicator for as many Danish coastal water bodies as possible, allowing for quantification of the relationship between the response variable "light" and the predictor variables, both external and manageable such as nutrient inputs.

2 Materials and methods

2.1 Light indicator development (response variable)

Eelgrass light requirement

Availability of sufficient light is crucial for eelgrass to grow, and hence we can use light penetration as a proxy for potential eelgrass depth distribution. Several factors can affect the eelgrass distribution, but light availability sets the maximum depth for the eelgrass distribution (Duarte 1991; Olesen & Sand-Jensen 1993; Gallegos & Kenworthy 1996; Olesen 1996; Lee et al. 2007; Krause-Jensen et al. 2011). The indicator “light limiting depth” is used as a proxy for the indicator “eelgrass depth limit,” which has been intercalibrated within the WFD. This study uses monitoring data from the Danish environmental monitoring program to estimate the light limiting depth in Danish coastal waters. Other studies have shown that eelgrass growth requires between 11 % and 20 % of surface radiation (Duarte 1991; Krause-Jensen & Middelboe 2000; Nielsen et al. 2002; Olesen 1996; Short et al. 1995). To find the regional optimum, we used an optimization routine in R to fit the estimated light limitation depth to the actual measured maximum eelgrass depth distribution for each waterbody. The analysis showed an optimum at 16 % of surface light at the maximum depth, where the eelgrass distribution in each water body was observed. We also looked for a depth-dependent minimum light requirement, expecting lower light requirements for eelgrass with a lower depth distribution, which has earlier been suggested in the literature. We found an apparent trend in the light dependency with depth distribution of eelgrass, showing lower light requirements with increasing depth distribution, but the variation was very high. We didn’t find much change in light requirement with depth from four (covering the interval [3.5,4.5[) meters and up to eight meters; instead, the variation decreased with depth stabilizing around 16% of surface light at the deepest populations (Figure 2.1.). Since the goals for maximum depth distribution of eelgrass in more than 98 % of the Danish waters are deeper than 3.5 meters we use 16 % of surface irradiance as a general light requirement for eelgrass.

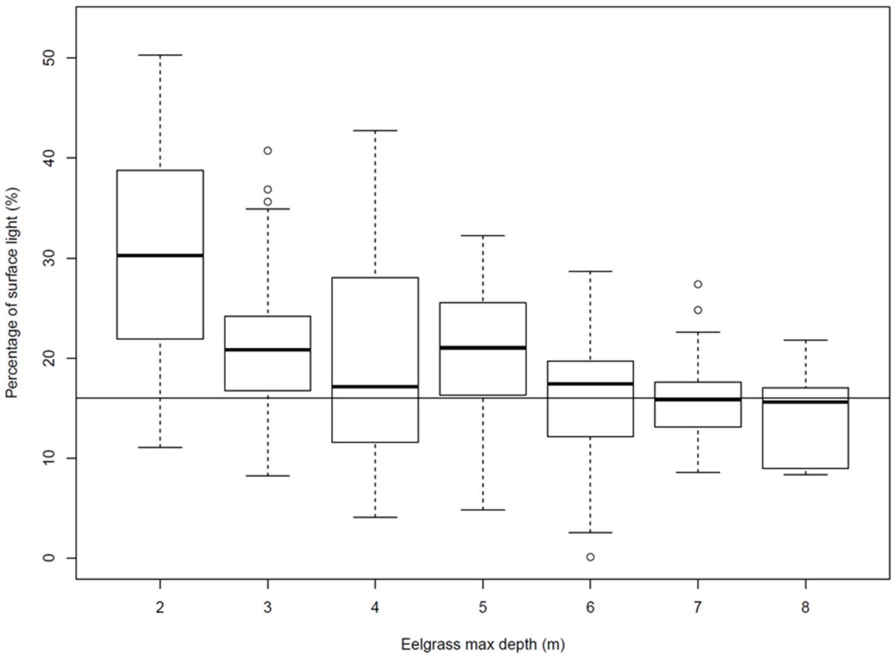


Figure 2.1. Boxplot showing the light requirement of eelgrass as a percentage of surface irradiance as a function of the maximum depth of the main distribution based on the data from the Danish monitoring and assessment program. The horizontal line shows 16 % of surface light, which is used here as the minimum light requirement.

Calculating the indicator

To calculate the light-limiting depth for the eelgrass depth distribution in Danish waters, we have first collected all quality-assured observations of light attenuation (Kd) and Secchi depths (SD) from all marine monitoring stations. At all stations and dates where both Kd and SD have been measured, a station and month-specific factor between Kd and SD is calculated

$$factor_{i,j} = Kd_{i,j} \cdot SD_{i,j}$$

where i is station and j is month.

This factor is then used to convert all SD measurements, where no Kd measurement has been made, to Kd. This is most relevant for older data, where fewer light attenuation measurements have been made.

As a result of the light requirement analysis, the potential eelgrass depth distribution – based on light availability was estimated as

$$Lz = \frac{-\ln(0.16)}{Kd}$$

where Lz is the light limiting depth in meters. Kd is the light attenuation coefficient calculated as

$$Kd = \frac{-\ln\left(\frac{Iz}{I0}\right)}{Z}$$

where Z is water depth in meters, $I0$ (mol photon m^{-2} d^{-1}) is surface irradiance, and Iz (mol photon m^{-2} d^{-1}) is the incoming radiation at water depth Z.

After converting all Kds to depths, we made linear interpolation between the observations with fewer than 60 days between them in order to calculate a time-weighted average. If a season in the dataset had more than 60 days between observations that season was not used. After interpolation, we calculated a mean value for the period of interest (March to October). This was done for every monitoring station and for every year with sufficient observations to cover the period following the aforementioned rule.

2.2 Data collection and preparation of predictor variables

Data for establishing time series of the predictor variables: "nutrient inputs," "Nutrient limitation," "salinity," "sea surface temperature" and "Buoyancy frequency (Brunt–Väisälä frequency)". All the data were obtained from the Danish National Aquatic Monitoring and Assessment Programme database (DNAMAP), whereas data for the climate predictor variable "irradiance" and "wind" were obtained from two weather stations in Copenhagen and Sprogø, respectively.

Table 2.1. Overview of the statistical models possible variables and some of the expected effects.

Variable	Positive effect (more light)	Negative effect (less light)
Nutrient input		Feed the productivity of the ecosystem and hence the productivity of potentially light absorbing and scattering organic matter
Nutrient limitation (nutrient concentration)	If the concentration of bioavailable nutrients are low the primary production is limited	
Salinity	Water of oceanic origin (more saline) are often more clear, in Danish waters	
Sea surface temperature	High temperature can affect the stratification of the water column and lead to a more stable water column with fewer suspended particles.	Water temperature affects a lot of processes that, either directly or indirectly can change the light penetration of the water higher temperature will increase both primary production and respiration of the system and increase the risk of oxygen depletion
Buoyancy frequency (Brunt–Väisälä frequency)	A stable water column (high BV) can lead to nutrient depletion of the photic zone and settlement of particles.	High BV can lead to increased growth of certain algae if nutrient are replete and increase the risk of hypoxia
Irradiance	UV light is known to breakdown coloured DOM and thus "bleach" the water	High incoming radiation can increase primary production under nutrient replete conditions
Wind	High wind speed can increase the amount of suspended particles in the water	

Data from DNAMAP were sampled with weekly to biweekly intervals, whereas irradiance and wind were sampled bi-hourly.

For all the semi-enclosed water bodies, nutrient inputs were provided as monthly values, obtained from integrating continuous flow measurements and nutrient spot samples from gauged catchments combined with model predictions from ungauged catchments. In the relatively open coastal areas, a larger but still local catchment was used.

Salinity and water temperature predictor variables were calculated as monthly means of salinity or temperature profiles averaged over the surface layer (0-10 m).

Buoyancy frequency was calculated as the Brunt-Väisälä buoyancy frequency (N) based on the difference between surface (0-1 m) and bottom density (1 m above bottom):

$$N = \sqrt{-\frac{g}{p_0}\frac{dp}{dz}}$$

where g is the regional gravitational constant (9.82 ms^{-2}), p_0 is the potential density (surface density, kg m^{-3}), dp is the difference between bottom and surface density (kg m^{-3}), dz is the depth difference between bottom and surface (m) and N is the buoyancy frequency (s^{-1}).

Three variables representing nitrogen limitation (Nlim) and phosphorus limitation (Plim and TPlim) were calculated based on inorganic nitrogen, inorganic phosphorus and total phosphorus concentrations. We chose P limitation based on both DIP and TP because some studies have shown that in some system TP is a better measure of limitation and bioavailable P than DIP (Ptacnik et al 2010, Bergström 2010). The limitation is estimated from a Monod kinetic function describing the growth rate µ as a function of substrate (S) and a half saturation constant (Ks) and based on literature values for half saturation constants (Fisher et al. 1988, Tyrell 1999, Fu et al 2005, Lewis and Wurtsbaugh 2008) The limitation is the described as 1 - μ, i.e. that very nutrient limited pelagic primary production gives values close to one whereas primary production in nutrient repleat conditions will give values close to zero. The nutrient limiting ($NutLim$) function is:

$$NutLim = 1 - \frac{S}{S + Ks}$$

The mechanical force of the wind on the water surface is proportional to the cubed wind speed (Alexander et al. 2000), and we, therefore, cubed the wind speed to obtain a relative measure on the wind energy delivered to the sea surface.

Irradiance data were obtained as half-hourly values of global irradiance from 1990 to 2017, and these were then converted to PAR values based on an algorithm from The University of Copenhagen. Data gaps were filled with data from Sprogø (SPØ) after adjustment of the level based on the maximum level of irradiance (0.96 of the level measured in Copenhagen). Data from the two sites in Copenhagen (HCØ and HBG) have the same level and slope, and the final unit is µmol photons m^{-2} s^{-1} calculated from global irradiance (W m^{-2}).

Data from 1990 to 1993 were hourly data, which were interpolated linearly to obtain 30 min intervals using the “Proc Expand” procedure in SAS. All values below 2 µmol photons m^{-2} s^{-1} were set to zero due to low sensor sensitivity within that range and problems in some years with a dark offset. This is significant in some winter months, where a dark offset may constitute a substantial part of the daily sum. Finally, the data were translated into monthly mean values. The remaining gaps were filled with average values for the same day and time from other years.

Model development was restricted to data from monitoring stations within the WFD zone of water bodies with at least 15 years’ data series between 1990 and 2018 ensuring that both year-to-year variations as well as potential longterm trends could be resolved. Only time series with a minimum of one bi-monthly observation was included to have enough data points to get robust monthly interpolated values.

The two predictor variables, nitrogen inputs, and phosphorous inputs, were highly correlated. To avoid potential collinearity between these predictor variables, principal component analysis (Wold et. al. 1987) was performed and the first principal component, which explained most of the total variation (an average of 91 %) was used instead of the original N and P inputs. This latent variable is referred to as "load." Predictor variables were measured in different units and scales. Therefore, they were standardized to a mean of zero and a standard deviation of one, making the variances of predictor variables comparable.

2.3 Bayesian approach

To quantify the relationship between the light and the predictor variables, i.e. nutrient input, chemical, physical and climate variables, we used a Bayesian modelling framework.

Bayesian inference is a way of combining information from data along with the prior knowledge from expertise, earlier studies, or literature. Bayesian statistics is based on principles of conditional probability. It interprets probability as a measure of believability or how confident we are in an event occurring.

Bayes' original theorem applied to point probabilities given as follows:

$$p(B/A) = \frac{p(A/B)p(B)}{p(A)}$$

The theorem illustrates that a conditional probability for event B given event A is equal to the conditional probability of event A given event B, multiplied by the marginal probability for event B and divided by the marginal probability for event A.

In other ways, Bayes' rule states how the prior information $p(B)$ and the likelihood $p(A/B)$ are combined to arrive at the posterior distribution $p(B/A)$. *p(A)* is often ignored since it is, in many cases difficult to calculate and can often be assumed constant.

Thus, we can write Bayes' rule as:

$$posterior \propto likelihood * prior$$

where *likelihood* is the likelihood function, which reflects information about the parameters contained in the data, and the *prior* is prior probability distribution, which quantifies prior belief about the parameters before observing data. *Posterior* is posterior probability distribution; prior distribution and likelihood are combined to form the posterior distribution, which describes total knowledge about the parameters after the data have been observed (Gelman et al. 2003; Glickman & van Dyk 2007).

In the Bayesian method, model parameters intercept and slopes are considered as random variables, and prior belief about these parameters, so-called prior probability distribution is assigned. Thus, the Bayesian approach incorporates prior understandings and evidence to produce new posterior beliefs. Additionally, Bayesian inference quantifies the uncertainty explicitly, which is appealing in environmental decision-making (Gelman et al. 2003; Ellison 2004; Clark 2005).

In the current study, light prediction models for 44 coastal monitoring stations were individually fitted. Single-station models were developed for all stations. Developing separate models for each monitoring station is essential, because many factors can affect the relationship between the model response, variable light limiting depth, and the predictor variables, to differ among monitoring stations. However, focusing on site-specific features increases the risk of overfitting. To minimize the problem of overfitting, we used WAIC (described below) to assess the number of variables in each model.

The Bayesian linear model used in this study summarized as shown:

$$Lightdepth_{observed_{ij}} \sim normal\left(Lightdepth_{true_{ij}}, \sigma\right)$$

Where $Lightdepth_{true_{ij}}$ is the mean, and σ is the standard deviation.

The Light depth data are nearly normal distributed; thus a normal distribution is used to model the light limitation depth.

The general model for the single-station light models:

$$Lightdepth_{true_{ij}} = \alpha_j + X_{1,ij}\,\delta_j^{x1} + X_{2,ij}\delta_j^{x2} \ldots + X_{n,ij}\delta_j^{xn}$$

where $Lightdepth_{true_{ij}}$ is i^{th} observation of light limitation depth from the j^{th} station.

α_j is the intercept term for single-station model.

δ_j^{x1} , δ_j^{x2} …δ_j^{xn} are single-station slope parameters for the n variables.

The non-informative prior distributions were used for model intercept and overall slope parameters:

$\alpha_j, \delta_j^{x1}, \delta_j^{x2} \sim N(0,1000)$

Standard deviation σ had non-informative prior:

$\sigma \sim Uniform(0,100)$

Initial values were defined for parameters as shown:

α_j = mean of $Lightdepth_i$

$\delta_j^{x1}, \delta_j^{x2}$… = 0 and

σ = Standard deviation of $Lightdepth_i$

Simulation of parameter posterior distributions was performed using Hamiltonian Monte Carlo (HMC) - a Markov chain Monte Carlo (MCMC) method. The MCMC method allows the user to sample all unknown parameters using joint posterior distributions that otherwise cannot be directly calculated (Gilks et al. 1996; Gamerman & Lopes 2006). In this study, sampling consisted of 10,000 iterations and 2,000 warmup iterations. Initially, two independent chains were used to sample from, but in the final models, one chain was used, which was sufficient. The mean value of the posterior samples was considered as the estimate of each parameter.

Convergence diagnostics such as effective number of samples, trace and posterior density plots were evaluated to ensure that sufficient number of chains was used, trajectory of the chain was stationary around the similar values, mixing was good and posterior had appropriate target distribution (McElreath 2016).

Statistical software SAS® (SAS 9.4, SAS Institute Inc, Cary, North Carolina, USA) was used for relevant data extraction from database and data management. Bayesian analysis was performed using R software (R Core Team 2018) using Rethinking (McElreath 2016) and RStan (Stan Development Team 2019) packages.

2.4 Choosing informative predictor variables

In this study, nine predictor variables were initially available for predicting light conditions, but Nlim and Plim was discarded in an initial screening based on correlations and linear relationships. However, not all predictor variables are of importance for light modelling at different monitoring stations. We used the method proposed by Lindeman, Merenda, and Gold (LMG) (Lindeman et al. 1980) to exclude predictor variables with no explanatory power for the light condition at a given station. The method uses sequential sums of squares from the multiple linear regression model, and overall assessment is obtained through the coefficient of determination (R^2) partitioned by averaging over orderings among predictors. R package '*relaimpo*' (Grömping 2006) was used to perform this analysis.

2.5 Model comparison and final model selection

Once the informative predictors were selected based on their relative importance in predicting light conditions, different combinations of important predictors were used in the Bayesian approach. The Watanabe-Akaike or widely applicable information criterion (WAIC) (Watanabe, 2010) was used to compare the model with different predictors. Lower WAIC values correspond to better model performance, and therefore models with the lowest WAIC were considered as the best model. However, for some monitoring stations, the final selected model was chosen based on the best balance between both WAIC value and R^2 value.

WAIC is regarded as an improvement on the deviance information criterion (DIC) for Bayesian models and is defined as (McElreath 2016):

$$\text{WAIC} = -2(\text{lppd} - \text{pWAIC})$$

where lppd is log-pointwise-predictive density, averaged over the posterior distribution, given by:

$$lppd = \sum_{i=1}^{N} \log \Pr(yi)$$

$\Pr(yi)$ is the average likelihood of the i^{th} observation of training sample.

pWAIC is the effective number of parameters, given by:

$$\text{pWAIC} = \sum_{i=1}^{N} \text{V}(yi)$$

$\text{V}(yi)$ is the variance in log-likelihood of the i^{th} observation of training sample.

2.6 Posterior predictive evaluation

Model performance for single-station models was assessed and quantified by comparing model predictions and observations of light conditions using a suite of quantitative and visual measures listed below.

- Time series plots of modelled and observed values of the light indicator (appendix A+B) together with the 95 % highest posterior density interval (HPDI) for each predicted observation obtained using samples from a posterior density.

- Plots and correlation analysis of observed vs. modelled light indicator values (appendix A+B) were used to evaluate the ability of the model to capture the variation in observed light conditions. The coefficient of determination R^2 (Table 2.a and 2.b) was used to quantify the variability captured by the model and an F test (table 2a and 2b) was used to determine if the correlation was significant.

- Time series plots of residuals (appendix A+B) were used to identify any non-random pattern and the ability to capture year-to-year variation. An F test (Table 2a and 2b) was used to test if the bias was time-dependent (i.e. if the regression between the residuals and time was significant).

- Plots of residuals vs observed light indicator values (appendix A+B) were used to detect if residuals were distributed as expected, i.e. "U-shaped" with smallest residuals close to the mean and larger residuals for small or large light depth values. The Shapiro-Wilk test for residuals was used to test if the residuals were normally distributed.

- Mean residuals (bias) were used to quantify any systematic deviation between model results and observations

- Root mean square error (RMSE) was used to assess the applicability of the model to capture high values or produced high values not reflected by the observations (Table 2a and 2b).

As there are no objective or formal criteria for determining when a model "is good enough," the following guidelines (Table 2.1.) were used to identify any potentially problematic areas that might influence overall model performance and the applicability of the model to perform model scenarios.

Table 2.2. Overview of statistical model assessment methods used for evaluation of model performance.

Statistical method	Description
Coefficient of determination (R^2) for the correlation between model results and corresponding observations	Quantify the variation captured by the model and should be as close to 1 as possible R^2 and the assessment criteria is only meaningful if there is "sufficient" variation in the observations
Significance test (F-test) for the correlation between model results and observations	Model results and observations should be significantly correlated
Average bias of model results and observation	Average bias identify any overall systematic deviation between model and observations and should be as close to 0 as possible
Significant test (F test) for the correlation between time and residuals	Correlation between time and residuals should be not significant as significant time trends in residuals could pose a problem for model scenarios
Root mean square error (RMSE)	RMSE should be as small as possible and not exceed the standard deviation of the observations
Shapiro-Wilk test for residuals	Test if residuals are significantly different from a normal distribution. NS implies that residuals are normally distributed as expected

3 Results and discussion

3.1 Bayesian Light limitation depth models

In total, 44 monitoring stations fulfilled the requirement of more than 15 years of regular monitoring, and data from these stations were included in the model development.

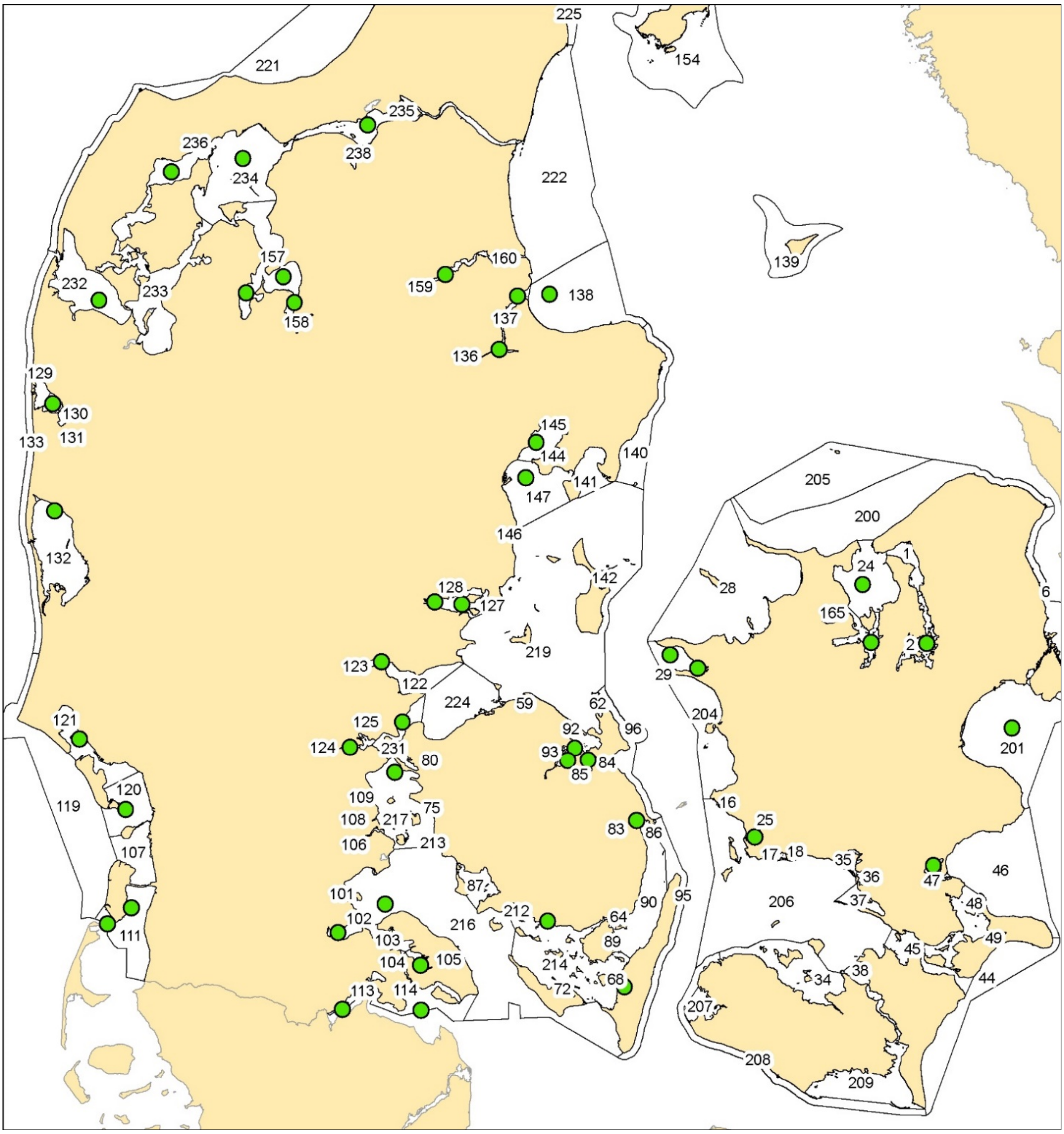


Figure 3.1. Map showing location of the 44 monitoring stations where monitoring data was used to develop single-station light models.

Single-station models were developed for 44 monitoring stations covering 41 water bodies (Figure 3.1.). The MCMC runs converged quickly with trace plot indicating that the centre of the chain appeared to be stationary around the similar values and with bell-shaped curves for the posterior distributions of parameters (not shown). Nutrient input was selected as predictor variables in 23 of the single-station models making it the second most common predictor variable after TPlim, which was a predictor in 26 of the single-station models.

The variable selection for single-station models was based on LMG and WAIC. LMG method using multiple linear regression and WAIC alone yielded the same list of important predictor variables for light depth prediction when compared for selected monitoring stations. For comparison reasons, a subset of stations was also tested with variable selection using partial least squared regression, which was used in similar work at an earlier stage. This method also yielded the same important variables as the selection methods used in this study. This indicates that variable selection was relatively insensitive to the choice of method.

3.2 Model fit and evaluation

The performance statistics for the 44 single-station models (Table 3.1.) showed significant ($p < 0.05$) correlations between model and observations for all the models, and R^2 values were in general acceptable ($R^2 > 0.3$). Mean residual (Bias), as well as RMSE, were low. The residuals were normally distributed except for three monitoring stations, and generally, no time trends in residuals were detected.

Visual inspection of the model performance plots (Appendix A) confirmed the overall good performance of the single-station models. Both time series plots as well as regression plots of observed vs. modelled light limitation depths showed that the models could capture the variation in the observation and regression lines of observed vs. model results were close to the 1:1 correspondence line.

Table 3.1. Summarized model performance statistics for the single-station models. R^2 is the coefficient of determination for the relationship between observed and modelled light conditions; RMSE is the root mean squared error; Shapiro-Wilk test is the normality test results for the residuals and F test for residual vs. time indicate if residuals are time dependent or not. Stars (*) indicate significance level, NS is "not significant"

Station name	Station ID	R^2	Mean residual (Bias)	RMSE	Shapiro-Wilk test for residual	F test for residual vs. time	Significant predictors
Kalø vig	ARH170002	0.33**	0,00212	0.62	NS	NS	Prin1, Wind, Sali
Århus bugt	ARH170006	0.22*	0,00096	0.65	NS	NS	Prin1, Wind
Hevring bugt	ARH190004	0.64***	-0,00097	0.36	NS	NS	Prin1, Sali, Temp, Irr
Randers ml	ARH230902	0.75***	0,00002	0.11	NS	NS	Prin1, TPlim, BV
Randers ydre	ARH230905	0.57***	0,00066	0.43	NS	NS	Prin1, Sali, BV
Nakkebølle fjord	FYN0018361	0.19*	0,00104	0.37	NS	*	Prin1, Temp
Lindelse nor	FYN0018571	0.93***	0,00012	0.21	NS	NS	Prin1, BV, Temp, Irr, Wind, Sali
Holcken havn	FYN0018752	0.27*	0,00003	0.23	NS	NS	Temp, Sali
Kerteminde fjord/kertinge nor	FYN0018843	0.16ns	0,00013	0.16	Yes	*	Temp
Bredningen, Lillebælt	FYN6100021	0.42***	0,00140	0.54	NS	NS	TPlim, Sali

Lillebælt nord - ved Frederecia	FYN6100051	0.74***	-0,00046	0.25	NS	NS	Prin1, TPlim, Sali
Nord for Als (Lillebælt vest)	FYN6200901	0.55***	0,00182	0.39	NS	NS	Prin1, TPlim, Wind
Odense ydre	FYN6900017	0.52***	0,00032	0.32	NS	NS	Prin1, TPlim, Wind
Odense indre	FYN6910008	0.48***	0,00004	0.26	Yes	NS	Prin1, TPlim, Sali,Temp
Mariager fjord	NOR5503	0.57**	0,00060	0.19	NS	NS	Prin1, TPlim, Sali
Vadehavet Grådyb - Ho bugt v Langli	RIB1610002	0.63***	0,00006	0.23	NS	NS	TPlim, Sali, Wind
Vadehavet Knudedyb	RIB1620014	0.62***	0,00050	0.39	NS	NS	Prin1, Sali, TPlim, Wind, Irr
Ringkøbing fjord nord	RKB1	0.81***	-0,00004	0.07	NS	NS	Prin1, Temp, TPlim
Nissum fjord	RKB22	0.93***	0,00001	0.07	NS	NS	Temp, TPlim, Wind
Køge bugt	ROS1727	0.35**	-0,00698	0.87	NS	NS	Wind, TPlim
Roskilde indre	ROS60	0.51***	0,00157	0.21	NS	NS	Sali, Wind
Lister dyb	SJY1	0.64***	0,00117	0.37	NS	NS	Wind, Sali, Temp
Augustenborg fjord	SJY12	0.44**	-0,00016	0.41	NS	NS	TPlim, BV, Irr, Wind
Aabenraa fjord	SJY15	0.51***	0,00380	0.44	NS	NS	TPlim, Sali, BV
Lister dyb v Rømø havneby	SJY3	0.39**	-0,00170	0.35	NS	NS	BV, Irr, Prin1
Flensborg fjord inder	SJYKFF2	0.48**	-0,00120	0.49	NS	NS	TPlim, Irr
Flensborg fjord yder	SJYKFF5	0.37**	0,00111	0.44	NS	NS	TPlim, Temp
Præstø fjord	STO0802008	0.26**	0,00260	1.04	Yes	NS	BV , TPlim, Prin1
Kolding fjord	VEJ0003350	0.35**	-0,00107	0.29	NS	NS	TPlim, BV, Sali
Vejle fjord	VEJ0004273	0.41***	0,00103	0.46	NS	NS	Prin1, Sali, BV, Temp, Irr
Horsens inder	VEJ0005790	0.27**	0,00041	0.39	NS	NS	TPlim
Horsens yder	VEJ0006489	0.33*	0,00355	0.48	NS	NS	Temp, BV, Sali
Nissum bredning	VIB3702-00001	0.68***	-0,00013	0.27	Yes	NS	Prin1, Temp, Wind
Løgstør bredning	VIB3708-00001	0.60***	-0,00009	0.44	NS	**	Sali, Temp, BV
Nibe bredning	VIB3711-00001	0.79***	0,00079	0.26	Yes	NS	Prin1, Sali, Temp, BV
Thisted bredning	VIB3723-00001	0.63***	0,00466	0.61	NS	NS	Sali, Temp, Wind
Skive fjord	VIB3727-00001	0.77***	-0,00060	0.13	NS	NS	Sali, Temp, Prin1, Irr
Lovns bredning	VIB3728-00001	0.71***	-0,00066	0.16	NS	NS	TPlim, Irr, Sali
Hjarbæk fjord	VIB3729-00001	0.87***	-0,00038	0.11	NS	NS	Prin1, Wind, Temp, Irr, TPlim
Isefjord dybt bassin	VSJ10003	0.36**	0,00051	0.46	NS	NS	Prin1, Wind, Temp, TPlim
Isefjord inderbredning	VSJ10006	0.57***	0,00084	0.26	NS	NS	Wind, Temp, Irr, TPlim
Kalundborg fjord yder	VSJ41007	0.57***	0,00008	0.42	NS	NS	Prin1, Temp, Irr, TPlim
Kalundborg fjord inder	VSJ41008	0.77***	0,00344	0.39	NS	NS	Wind, BV, TPlim, Sali
Skælskør fjord	VSJ51013	0.46**	-0,00251	0.42	NS	NS	Temp, Irr, TPlim

***p ≤ 0.001; **p ≤ 0.01; * p ≤ 0.05; . p ≤ 0.1; ns p > 0.1 (not significant).

Prin1 is the first principal component of nitrogen and phosphorus loadings (nutrient input); TPlim is a proxy for phosphorus limitation, Wind is cubed wind speed; Temp is sea surface temperature; BV is Brunt-Väisälä buoyancy frequency for the whole water column; Irr is incoming PAR radiation; Sali is salinity in the water surface (upper 10 m).

3.3 Model coefficients

Slope coefficients estimated from single-station models (table 3.2.) showed an overall negative correlation between nutrient inputs and light limitation depth, as expected. At four stations (Randers ydre, Lindelse Nor, Præstø Fjord and Skive Ford), the slope coefficient appeared to be positive.

Table 3.2. Estimated slope parameter for single-station models along with the standard deviation

Station name	Station ID	Coef ± SE for Prin1	Coef. ± SE for Temp.	Coef. ± SE for Sali	Coef. ± SE for BV	Coef. ± SE for Irr	Coef. ± SE for Wind	Coef. ± SE for TPlim
Kalø vig	ARH170002	-0.44 ± 0.18	NI	-0.2 ± 0.18	NI	NI	0.22 ± 0.17	NI
Århus bugt	ARH170006	-0.31 ± 0.15	NI	NI	NI	NI	0.24 ± 0.15	NI
Hevring bugt	ARH190004	-0.23 ± 0.14	-0.22 ± 0.16	0.22 ± 0.14	NI	0.2 ± 0.16	NI	NI
Randers ml	ARH230902	-0.07 ± 0.04	NI	NI	-0.04 ± 0.03	NI	NI	0.14 ± 0.04
Randers ydre	ARH230905	0.23 ± 0.15	NI	0.54 ± 0.17	-0.08 ± 0.15	NI	NI	NI
Nakkebølle fjord	FYN0018361	-0.15 ± 0.1	0.15 ± 0.1	NI	NI	NI	NI	NI
Lindelse nor	FYN0018571	1.07 ± 0.3	-0.55 ± 0.26	1.23 ± 0.29	0.48 ± 0.13	1.09 ± 0.24	-0.77 ± 0.16	NI
Holcken havn	FYN0018752	NI	0.1 ± 0.07	-0.09 ± 0.07	NI	NI	NI	NI
Kerteminde fjord/kertinge nor	FYN0018843	NI	0.07 ± 0.05	NI	NI	NI	NI	NI
Bredningen, Lille-bælt	FYN6100021	NI	NI	-0.16 ± 0.13	NI	NI	NI	0.43 ± 0.13
Lillebælt nord - ved Frederecia	FYN6100051	-0.36 ± 0.13	NI	-0.18 ± 0.11	NI	NI	NI	0.16 ± 0.11
Nord for Als (Lille-bælt vest)	FYN6200901	-0.16 ± 0.17	NI	NI	NI	NI	0.16 ± 0.14	0.35 ± 0.17
Odense ydre	FYN6900017	-0.18 ± 0.09	NI	NI	NI	NI	-0.09 ± 0.08	0.16 ± 0.09
Odense indre	FYN6910008	-0.2 ± 0.15	0.09 ± 0.1	-0.2 ± 0.14	NI	NI	NI	0.05 ± 0.14
Mariager fjord	NOR5503	-0.19 ± 0.09	NI	-0.11 ± 0.09	NI	NI	NI	0.13 ± 0.08
Vadehavet Grådyb - Ho bugt v Langli	RIB1610002	NI	NI	0.15 ± 0.07	NI	NI	-0.08 ± 0.06	0.17 ± 0.07
Vadehavet Knude-dyb	RIB1620014	-0.15 ± 0.12	NI	0.23 ± 0.12	NI	-0.2 ± 0.1	-0.23 ± 0.1	0.23 ± 0.11
Ringkøbing fjord nord	RKB1	-0.03 ± 0.02	0.05 ± 0.02	NI	NI	NI	NI	0.15 ± 0.02
Nissum fjord	RKB22	NI	0.05 ± 0.02	NI	NI	NI	-0.03 ± 0.02	0.22 ± 0.02
Køge bugt	ROS1727	NI	NI	NI	NI	NI	-0.54 ± 0.22	0.55 ± 0.22
Roskilde indre	ROS60	NI	NI	0.15 ± 0.06	NI	NI	0.12 ± 0.06	NI
Lister dyb	SJY1	NI	-0.22 ± 0.11	0.34 ± 0.11	NI	NI	-0.3 ± 0.11	NI
Augustenborg fjord	SJY12	NI	NI	NI	0.19 ± 0.13	0.22 ± 0.12	0.12 ± 0.12	0.29 ± 0.12
Aabenraa fjord	SJY15	NI	NI	0.1 ± 0.11	0.12 ± 0.11	NI	NI	0.45 ± 0.12
Lister dyb v Rømø havneby	SJY3	-0.08 ± 0.09	NI	NI	-0.22 ± 0.09	-0.15 ± 0.09	NI	NI
Flensborg fjord in-der	SJYKFF2	NI	NI	NI	NI	-0.1 ± 0.15	NI	0.43 ± 0.15
Flensborg fjord yder	SJYKFF5	NI	-0.14 ± 0.1	NI	NI	NI	NI	0.3 ± 0.1
Præstø fjord	STO0802008	0.21 ± 0.25	NI	NI	-0.51 ± 0.26	NI	NI	-0.44 ± 0.26
Kolding fjord	VEJ0003350	NI	NI	-0.08 ± 0.08	-0.07 ± 0.09	NI	NI	0.15 ± 0.09
Vejle fjord	VEJ0004273	-0.25 ± 0.13	0.2 ± 0.14	-0.19 ± 0.13	0.15 ± 0.14	-0.3 ± 0.15	NI	NI
Horsens inder	VEJ0005790	NI	NI	NI	NI	NI	NI	0.24 ± 0.09
Horsens yder	VEJ0006489	NI	-0.24 ± 0.16	-0.2 ± 0.17	-0.22 ± 0.17	NI	NI	NI

Nissum bredning	VIB3702-00001	-0.23 ± 0.06	-0.3 ± 0.06	NI	NI	NI	0.1 ± 0.07	NI
Løgstør bredning	VIB3708-00001	NI	-0.15 ± 0.12	0.51 ± 0.12	0.18 ± 0.12	NI	NI	NI
Nibe bredning	VIB3711-00001	-0.3 ± 0.09	-0.35 ± 0.08	0.22 ± 0.09	0.16 ± 0.08	NI	NI	NI
Thisted bredning	VIB3723-00001	NI	-0.38 ± 0.24	0.63 ± 0.25	NI	NI	-0.44 ± 0.23	NI
Skive fjord	VIB3727-00001	0.11 ± 0.04	-0.16 ± 0.04	0.21 ± 0.04	NI	0.11 ± 0.04	NI	NI
Lovns bredning	VIB3728-00001	NI	NI	0.14 ± 0.07	NI	-0.13 ± 0.07	NI	0.17 ± 0.07
Hjarbæk fjord	VIB3729-00001	-0.1 ± 0.07	0.15 ± 0.08	NI	NI	-0.18 ± 0.06	0.12 ± 0.06	0.21 ± 0.07
Isefjord dybt bassin	VSJ10003	-0.09 ± 0.12	-0.2 ± 0.11	NI	NI	NI	-0.16 ± 0.11	0.25 ± 0.12
Isefjord inderbred-ning	VSJ10006	NI	-0.27 ± 0.12	NI	NI	0.25 ± 0.12	0.13 ± 0.11	0.27 ± 0.1
Kalundborg fjord yder	VSJ41007	-0.28 ± 0.14	-0.22 ± 0.16	NI	NI	0.33 ± 0.14	NI	0.22 ± 0.13
Kalundborg fjord inder	VSJ41008	NI	NI	0.34 ± 0.15	-0.19 ± 0.15	NI	-0.24 ± 0.15	0.68 ± 0.15
Skælskør fjord	VSJ51013	NI	0.25 ± 0.17	NI	NI	0.19 ± 0.16	NI	-0.24 ± 0.16

Station Name and ID is the name of the water body and the monitoring station from which data were used to develop the models; Prin1 is the first principal component of nitrogen and phosphorous loadings (nutrient input); Temp is sea surface temperature; Sali is salinity in the water surface (upper 10 m); BV is Brunt-Väisälä buoyancy frequency for the whole water column; Irr is incoming PAR radiation; Wind is cubed wind speed; TPlim is a proxy phosphorous limitation. SE is standard error; NI: not included (respective predictor variable is not included for that station).

3.4 Model applicability

The developed Bayesian model framework provides light limitation depth models for 44 monitoring stations distributed in 41 coastal water bodies. The model performance statistics and evaluation plots indicate that most of the models can be used for scenario runs, at least when the scenarios are within or not too far from the model calibration area. As for all types of models, the uncertainty will increase when moving away from the calibration area.

Although the performance statistics for some of the models indicate potential problematic model performance, these models could still produce reasonable scenario results provided that the slope coefficients used in the scenarios are robust as determined by standard deviation. However, as a low model performance imply that important processes or mechanisms are not included in the model, this could potentially influence the reliability of model scenario results.

The models have been developed with the aim of producing nutrient input scenarios making the estimated “Load” coefficients (slope coefficient for the relation between nutrient loading and the light indicator) and associated uncertainty particularly important. Therefore, models could only be applied for nutrient input scenarios if nutrient loading was selected as predictor variable in the model. For models that do not contain load as predictor, the relation between nutrient input and light limitation depth have not been quantified. This does not necessarily imply that nutrient input and light penetration are not linked in that particular water body, but only that other factors that are not covered by the models are essential and that the available data do not support a quantification of the load coefficient. Although the slope coefficient for the Load predictor was not significantly different from 0 at all stations, load was selected as predictor variables and hence the information adds explanatory power to the model and the estimated slope coefficients should be used for running nutrient load scenarios. In total 16 of the 44 models were used for nutrient reduction scenarios, based on their performance.

To use the models for management scenarios, the loading slope coefficient (Load), that contains the combined effect of N and P loading, has to be separated into a nitrogen load slope (N-slope) and a phosphorus load slope (P-slope). The separation is based on the basic assumption that either N,P or both nutrients are limiting nutrients in the water bodies where we find a relationship between load and the extent of the light penetration. The nutrients are limiting when the dissolved inorganic fraction concentration is below the half-saturation concentration (*Ks*) for that nutrient. Both nutrients can be limiting simultaneously if they are below the threshold concentration. The limitation is described with a Monod growth kinetics equation

$$Lim_S = 1 - \frac{[S]}{[S] + Ks}$$

where $[S]$ is the concentration of the substrate (N or P), Lim_S is the limitation, and Ks is the half-saturation (or half velocity constant). The estimated limitation is calculated for both N and P based on the concentration of dissolved inorganic fractions of N and P (DIN and DIP). The limitation for each nutrient will then be weighted according to their estimated limitation from 0 to 1 (everything below 0.5 i.e., DIN or DIP concentration below half-saturation concentration, is considered as not limiting and assigned a 0). They are weighted, so the sum is 1:

$$Lim_{weightedX} = \frac{Lim_x}{Lim_N + Lim_P}$$

where $Lim_{weightedX}$ is the weighted limitation of X nutrient and Lim_N is the limitation of N and Lim_P is the limitation of P. Finally the prin1 slope is timed with $Lim_{weighted}$ to obtain the nutrient specific slope.

4 References

Alexander M, Scott J & Deser C. 2000. Processes that influence sea surface temperature and ocean mixed layer depth variability in a coupled model. Journal of Geophysical Research: Oceans 105: 16823-16842 http://onlinelibrary.wiley.com/doi/10.1029/2000JC900074/full.

Beck MB. 1987. Water quality modeling: A review of the analysis of uncertainty. Water Resources Research 23: 1393-1442.

Bergström AK. 2010 The use of TN:TP and DIN:TP ratios as indicators for phytoplankton nutrient limitation in oligotrophic lakes affected by N deposition Aquat. Sci. (2010) 72:277–281

Binzer T, Sand-Jensen K & Middelboe A-L. 2006. Community photosynthesis of aquatic macrophytes. Limnol Oceanogr. 51:2722–2733.

Cha YK, Park SS, Lee WH & Stow CA. 2016. A Bayesian hierarchical approach to model seasonal algal variability along an upstream to downstream river gradient. Water Resources Research 52(1): 348-357.
https://doi.org/10.1002/2015WR017327

Clark JS. 2005. Why environmental scientists are becoming Bayesians. Ecology Letters 8: 2-14.

Duarte CM. 1991. Seagrass depth limits. Aquatic Botany, 40(4), 363–377. https://doi.org/10.1016/0304-3770(91)90081-F

Ellison AM. 1996. An introduction to Bayesian inference for ecological research and environmental decision-making. Ecological Applications 6: 1036–1046.

Ellison AM. 2004. Bayesian inference in ecology. Ecology Letters 7: 509-520.

Fisher et al. 1988 FISHER, T. R., L. HARDING, D. W. STANLEY, andL. G. WARD. 1988. Phytoplankton, nutrients, and turbidity in the Chesapeake, Delaware, and Hudson estuaries.Estuarine, Coastal, and Shelf Science 27:61–93.

Gallegos C L & Kenworthy WJ. 1996. Seagrass depth limits in the Indian River Lagoon (Florida, U.S.A.): Application of an optical water quality model. Estuarine, Coastal and Shelf Science, 42(3), 267–288.
https://doi.org/10.1006/ecss.1996.0020

Gamerman D & Lopes HF. 2006. Markov chain Monte Carlo: Stochastic simulation for Bayesian inference. Boca Raton: Chapman & Hall/CRC. 342 pp.

Gelman A, Carli, JB, Stern HS & Rubin DB. 2003. Bayesian Data Analysis (2nd edition). Chapman & Hall/ CRC. 668 pp.

Germano JD. 1999. Ecology, statistics, and the art of misdiagnosis: The need for a paradigm shift. Environmental Reviews 7: 167-190.

Glickman ME & van Dyk DA. 2007. Basic Bayesian Methods. In: Ambrosius WT (Ed) Topics in Biostatistics. Methods in Molecular Biology™, vol 404. Humana Press, pp 319-338.

Grömping U. 2006. Relative Importance for Linear Regression in R: The Package relaimpo. Journal of Statistical Software 17(1): 1-27.

Gronewold A & Borsuk ME. 2010. Improving Water Quality Assessments through a Hierarchical Bayesian Analysis of Variability. Environmental Science and Technology. John Wiley & Sons, Ltd., Indianapolis, IN, 44(20):7858-7864.

Gilks WR, Richardson S, & Spiegelhalter DJ (Eds). 1996. Markov chain Monte Carlo in practice. Chapman and Hall / CRC, Boca Raton.

Herman P, Newton A, Schernewski G, Gustafsson B & Malve O. 2017. International evaluation of the Danish marine models. https://mfvm.dk/fileadmin/user_upload/MFVM/Nyheder/Bilag_1_Evalueringsrapport_om_de_danske_kvaelstofmodeller__10._oktober_2017-2.pdf

van Katwijk Marieke M., Bos Arthur R, Kennis Peer & de Vries Rob. 2010. Vulnerability to eutrophication of a semi-annual life history: A lesson learnt from an extinct eelgrass (*Zostera marina*) population. Biological Conservation. Volume 143, Issue 1. Pages 248-254.

Krause-Jensen D, Carstensen J, Nielsen SL, Dalsgaard T, Christensen PB, Fossing H & Rasmussen MB. 2011. Sea bottom characteristics affect depth limits of eelgrass *Zostera marina*. Marine Ecology Progress Series, 425, 91–102.

Lee KS, Park SR & Kim YK. 2007. Effects of irradiance, temperature, and nutrients on growth dynamics of seagrasses: A review. In Journal of Experimental Marine Biology and Ecology. Vol. 350, Issues 1–2, pp. 144–175.

Lindeman RH, Merenda PF & Gold, RZ. 1980. Introduction to Bivariate and Multivariate Analysis. Glenview, IL: Scott, Foresman and Company.

Malve O & Qian SS. 2006. Estimating nutrients and chlorophyll a relationships in Finnish lakes. Environmental Science & Technology 40: 7848-7853.

McElreath R. 2016. Statistical Rethinking: A Bayesian Course with Examples in R and Stan. CRC Press.

Nielsen SL, Sand-Jensen K, Borum J & Geertz-Hansen O. 2002. Depth colonization of eelgrass (*Zostera marina*) and macroalgae as determined by water transparency in Danish coastal waters. Estuaries, 25(5), 1025–1032. https://doi.org/10.1007/BF02691349

Omlin M & Reichert P. 1999. A comparison of techniques for the estimation of model prediction uncertainty. Ecological Modelling 115: 45-59.

Olesen B, & Sand-Jensen K. 1993. Seasonal acclimatization of eelgrass *Zostera marina* growth to light. Marine Ecology Progress Series, Vol. 94(1), 91-99.

Olesen B. 1996. Regulation of light attenuation and eelgrass *Zostera marina* depth distribution in a Danish embayment. Marine Ecology Progress Series, 134(1–3), 187–194.

Orth RJ, Carruthers TJB, Dennison WC, Duarte CM, Fourqurean JW, Heck KL, Hughes AR, Kendrick GA, Kenworthy WJ, Olyarnik S, Short FT, Waycott M & Williams SL. 2006. A global crisis for seagrass ecosystems. Bioscience 56: 987–996.

Ptacnik R, Andersen T, Tamminen T 2010 Performance of the Redfield Ratio and a Family of Nutrient Limitation Indicators as Thresholds for Phytoplankton N vs. P Limitation Ecosystems (2010) 13: 1201–1214

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Short FT, Burdick DM & Kaldy JE. 1995. Mesocosm experiments quantify the effects of eutrophication on eelgrass *Zostera marina*. Limnology and Oceanography, 40(4), 740–749.

Short FT & Wyllie-Echeverria S. 1996. Natural and human-induced disturbance of seagrasses. Environmental Conservation 23: 17–27.

Stan Development Team. 2019. RStan: the R interface to Stan. R package version 2.19.2. http://mc-stan.org/.

Tyrell T. 1999. The relative influences of nitrogen and phosphorus on oceanic primary production Nature 400: 525–531

Vaudrey J M P, Kremer J N, Branco B F, Short F T 2010. Eelgrass recovery after nutrient enrichment reversal. Aquatic Botany, 2010, 93(4), 237-243

Walters CJ. 1986. Adaptive Management of Renewable Resources. McMillan, New York.

Watanabe S. 2010. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. Journal of Machine Learning Research 11: 3571-3594.

Waycott M, Duarte CM, Carruthers TJB, Orth RJ, Dennison WC, Olyarnik S & Calladine A. 2009. Accelerating loss of seagrasses across the globe threatens coastal ecosystems. Proceedings of the National Academy of Sciences of the United States of America 106 (30): 12377–12381.

Wold S, Esbensen K & Geladi P. 1987. Principal component analysis. Chemometrics and Intelligent Laboratory Systems 2(1-3): 37-52.

Wold S, Sjöström M & Eriksson L. 2001. PLS-regression: a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems 58(2): 109-130.

Appendix A: Bayesian model evaluation for single-station model

Figures A1-A44: Bayesian model evaluation plots. In the time series plots, grey shaded areas represent 95 % highest posterior density interval (HPDI) for each predicted observation.

Figure A1. Kalø Vig, St. ARH170002

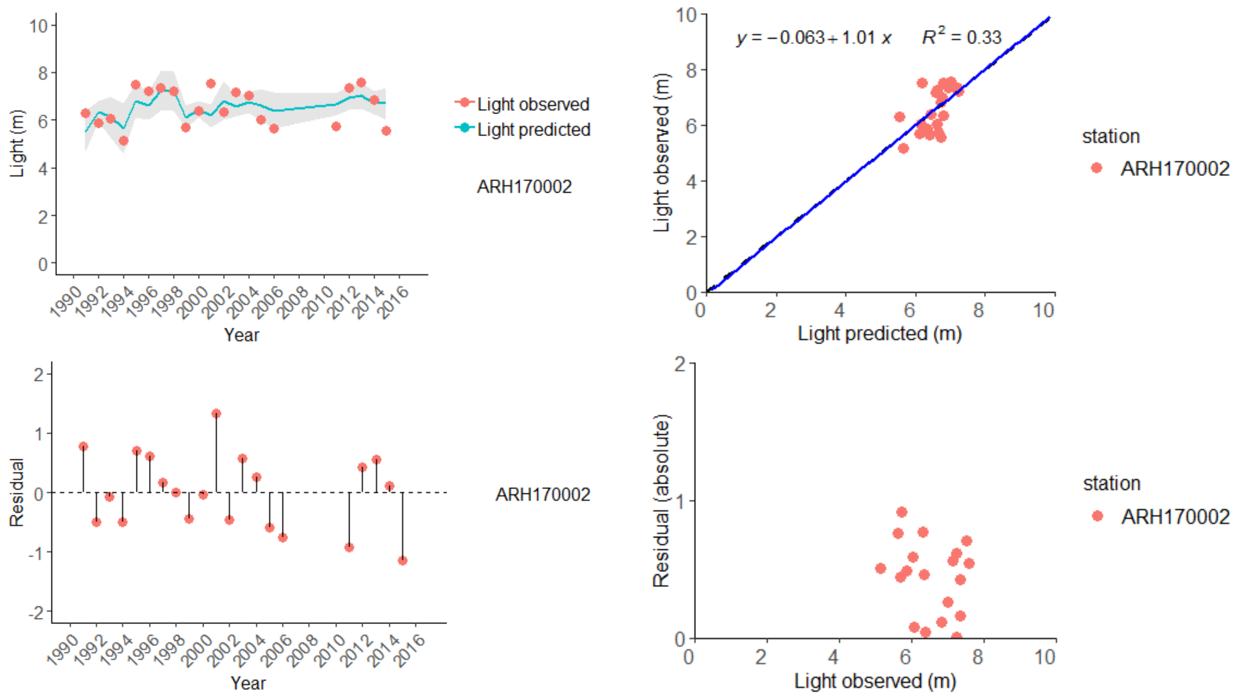


Figure A2. Århus Bugt, St. ARH170006

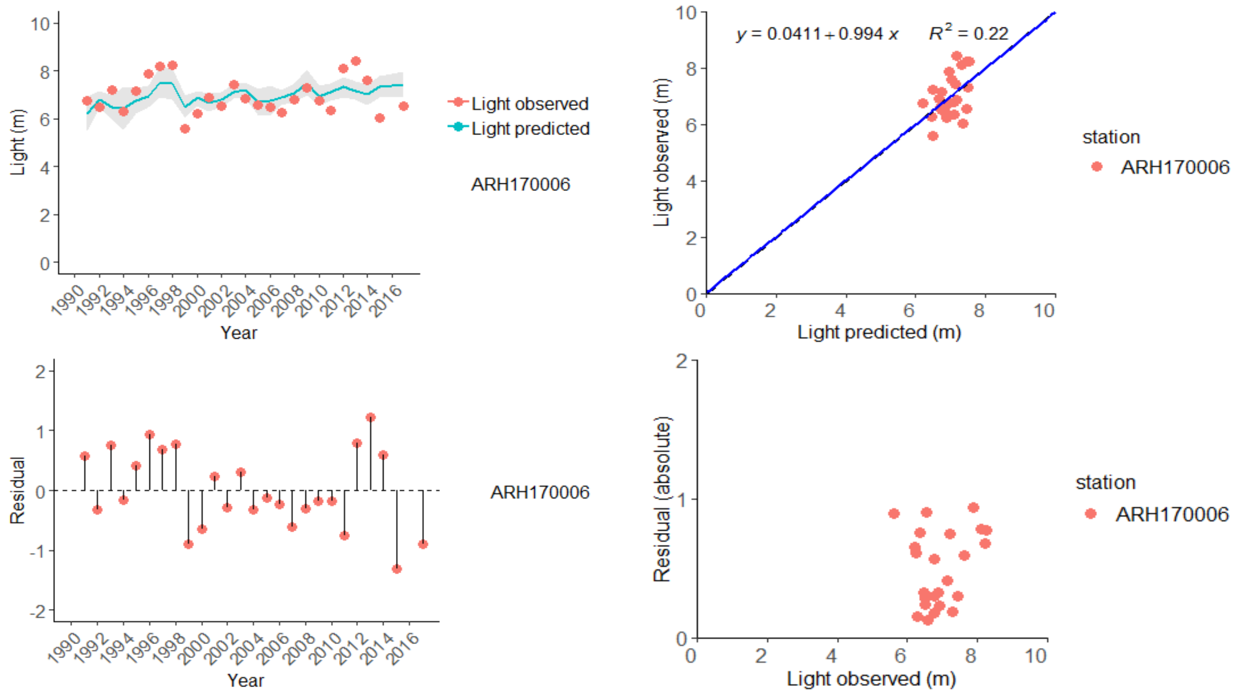


Figure A3. Hevring Bugt, St. ARH190004

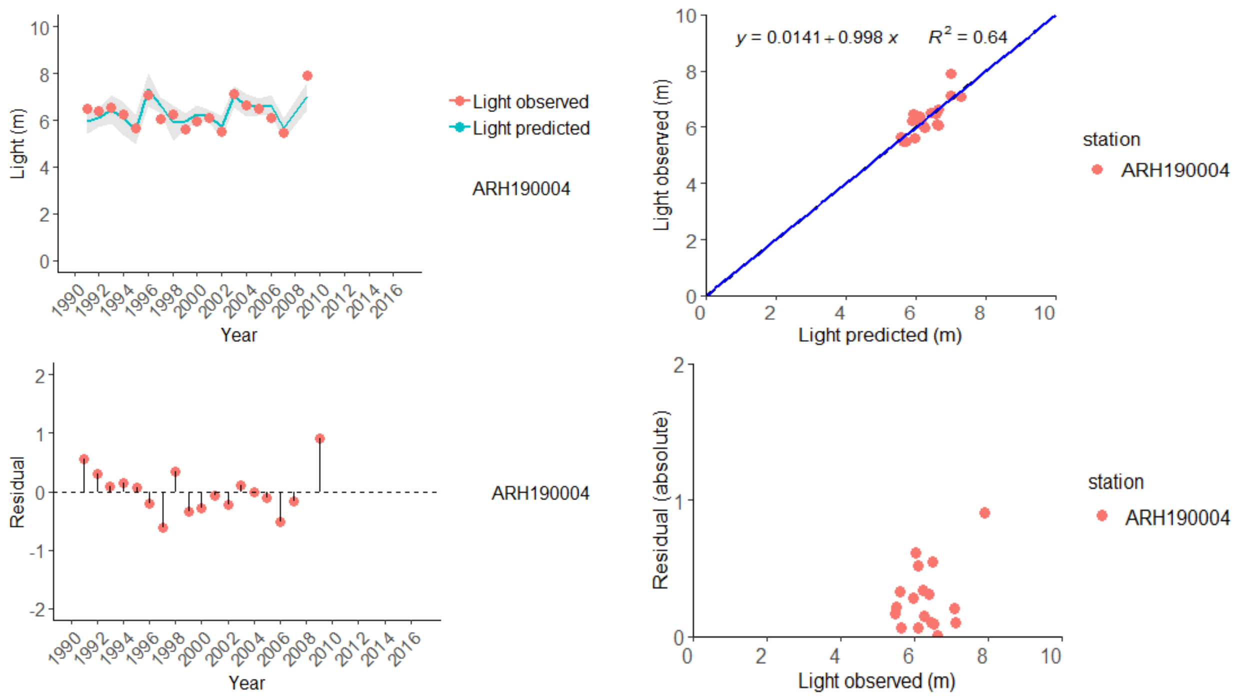


Figure A4. Randers ml, St. ARH230902

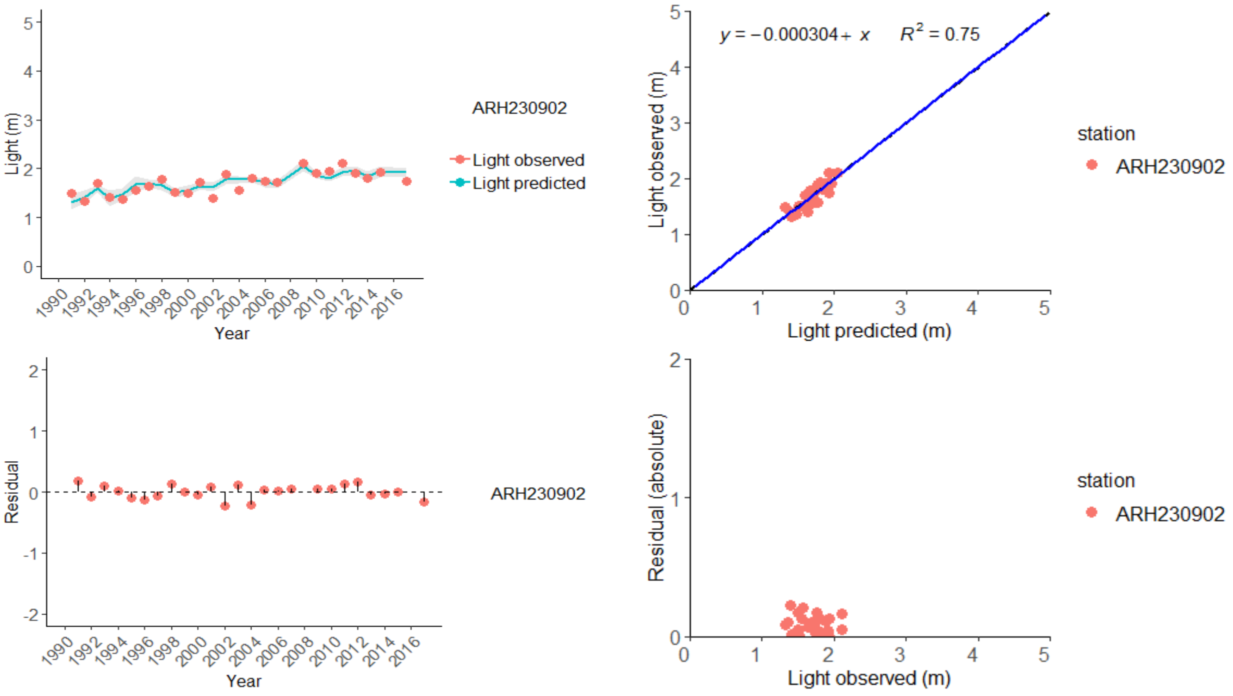


Figure A5. Randers ydre, St. ARH230905

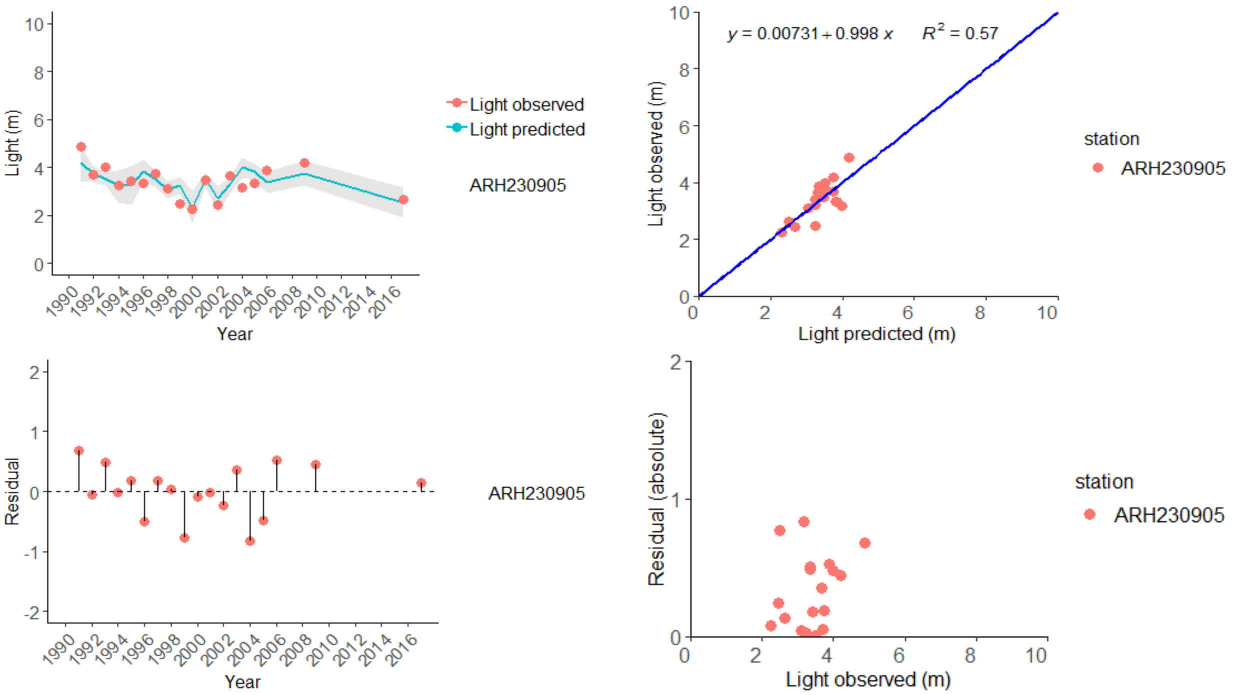


Figure A6. Nakkebølle Fjord, St. FYN0018361

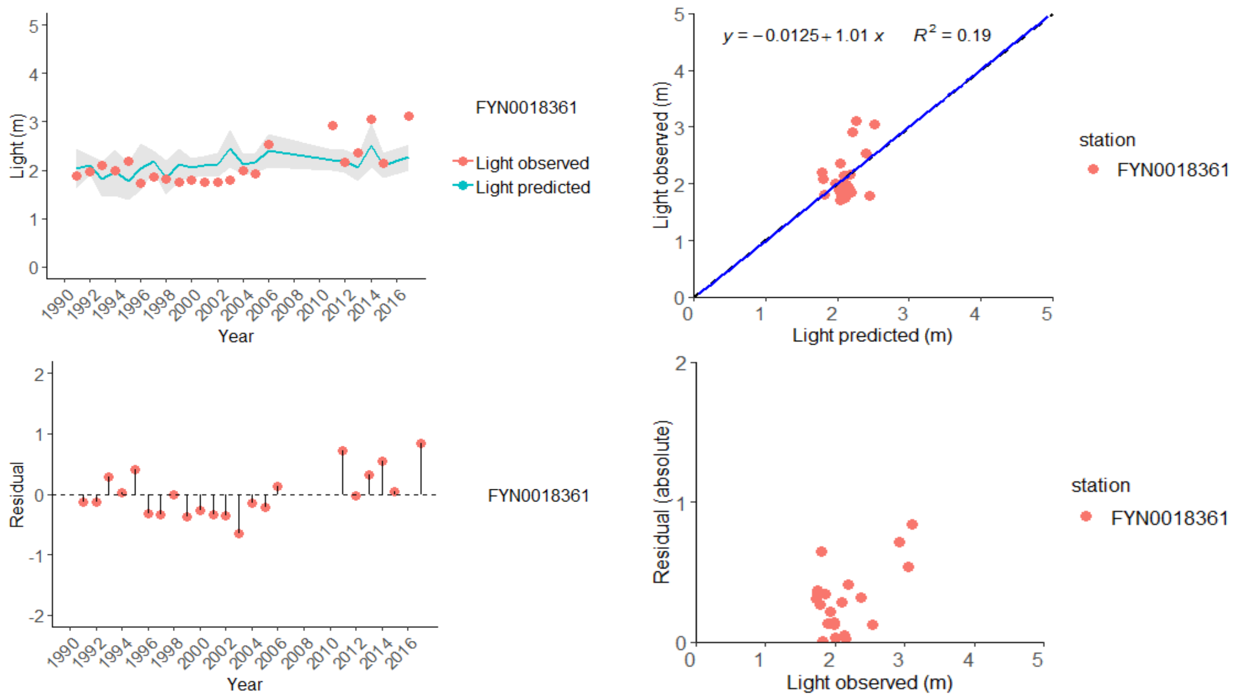


Figure A7. Lindelse Nor, St. FYN0018571

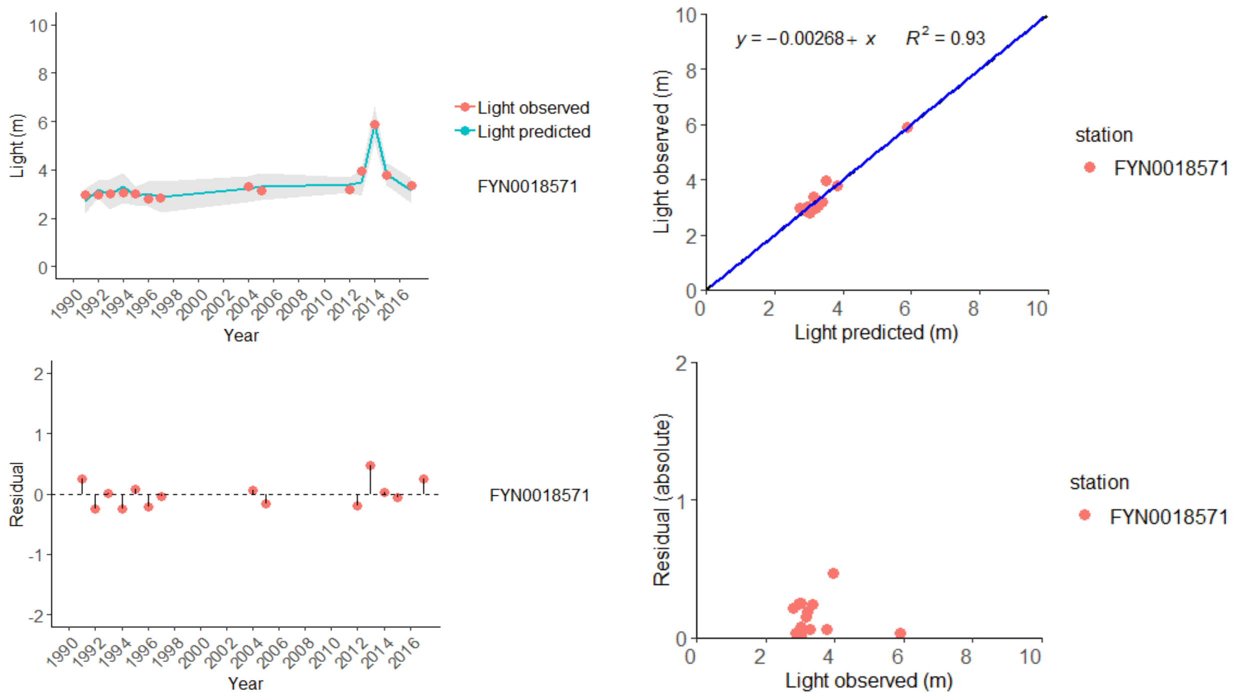


Figure A8. Holcken Havn, St. FYN0018752

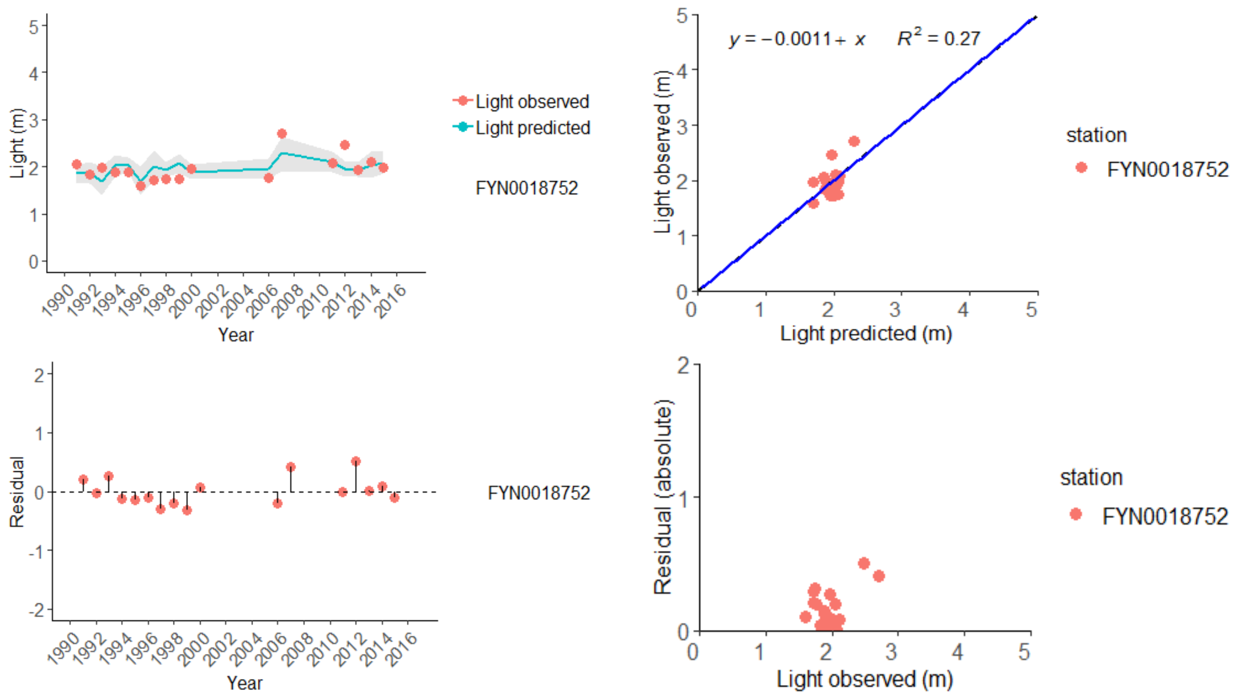


Figure A9. Kerteminde Fjord/Kertinge Nor, St. FYN0018843

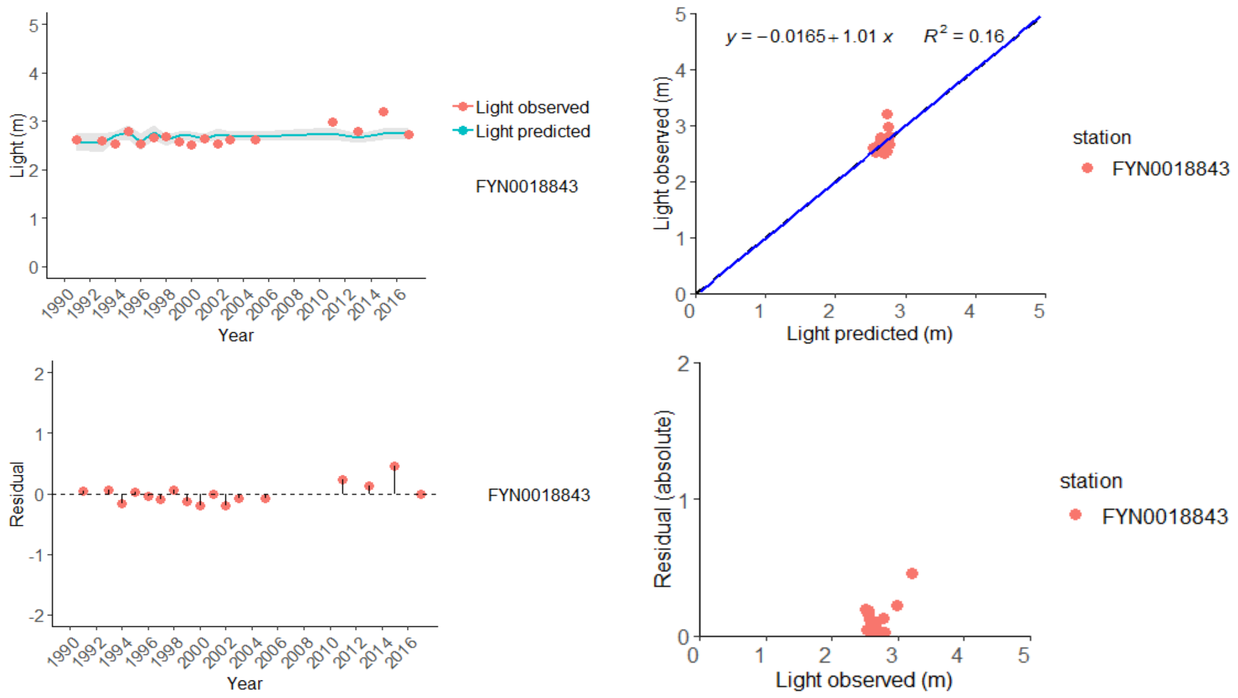


Figure A10. Bredningen, Lillebælt, St. FYN6100021

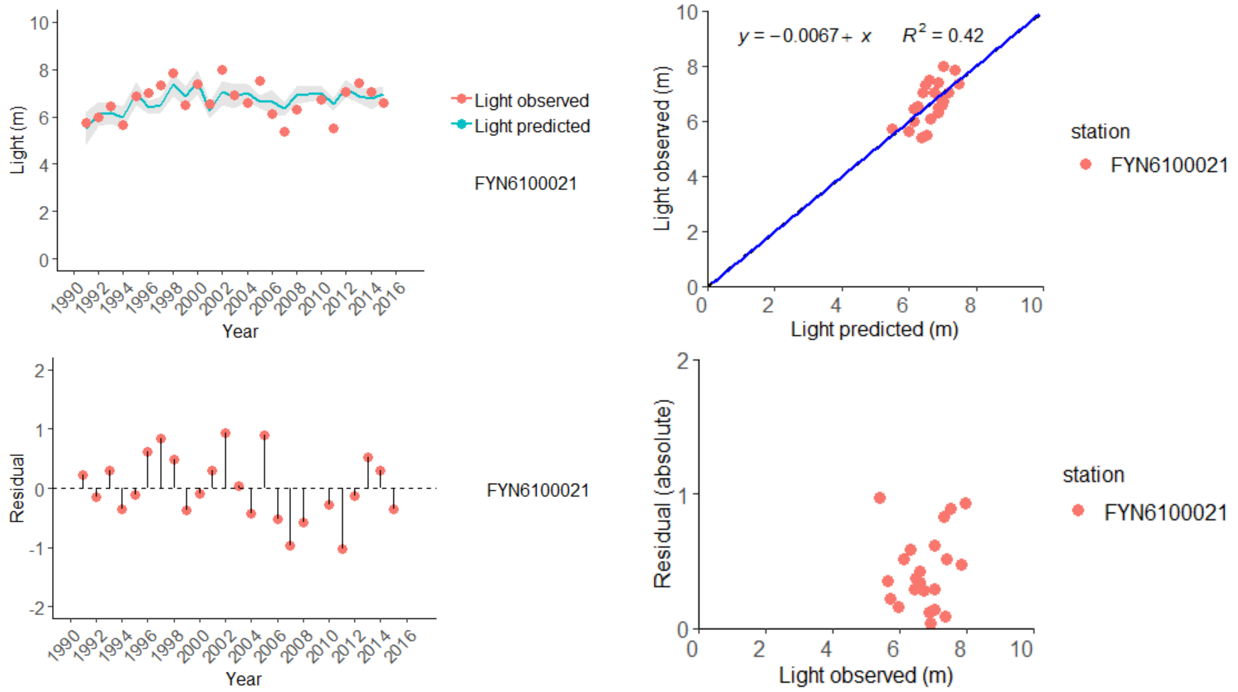


Figure A11. Lillebælt nord - ved Fredericia, St. FYN6100051

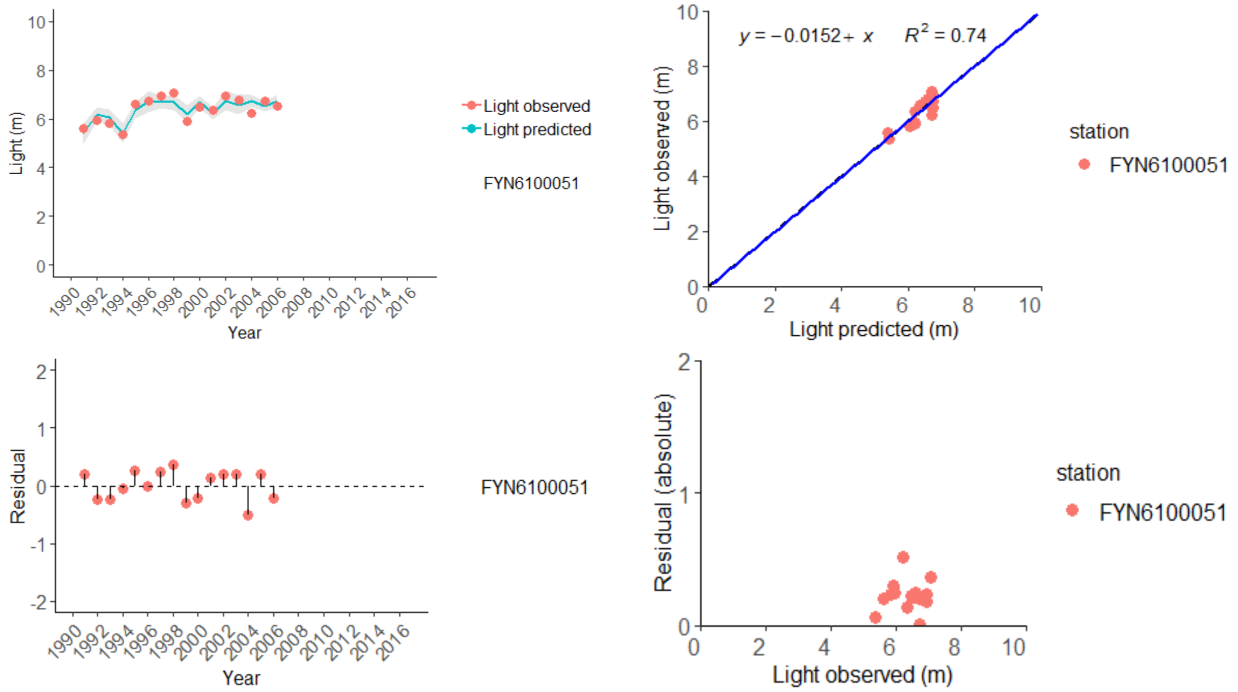


Figure A12. Nord for Als (Lillebælt vest), St. FYN6200901

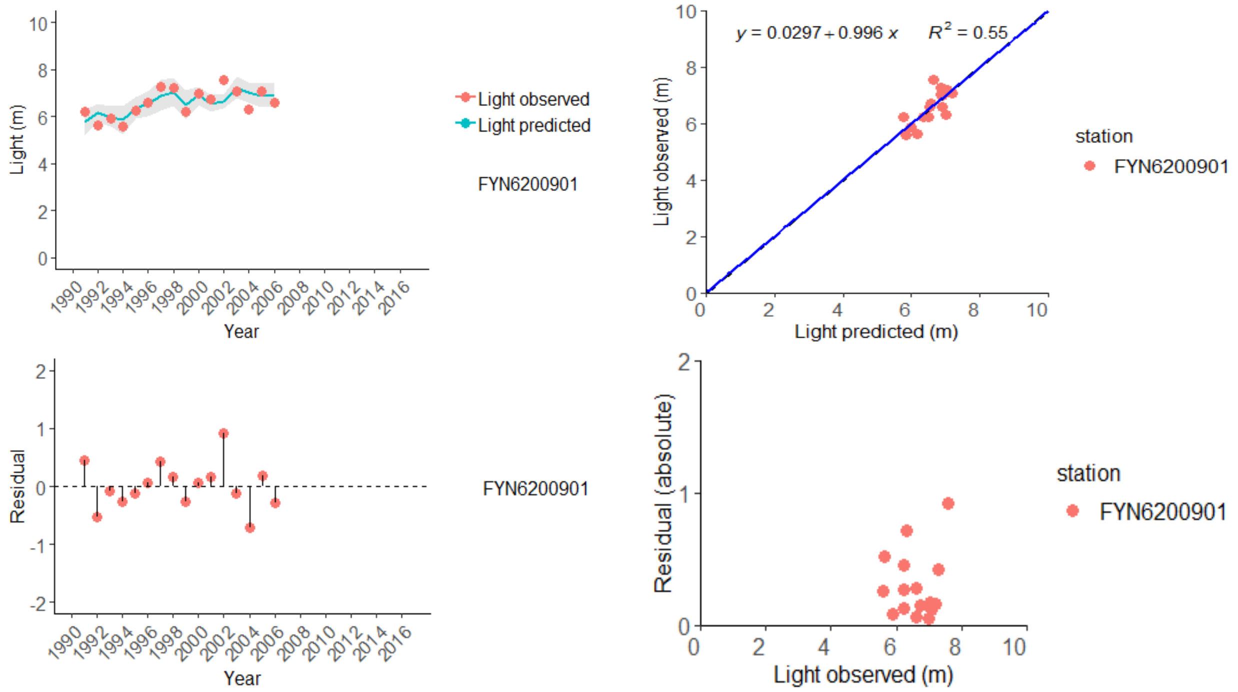


Figure A13. Odense ydre St. FYN6900017

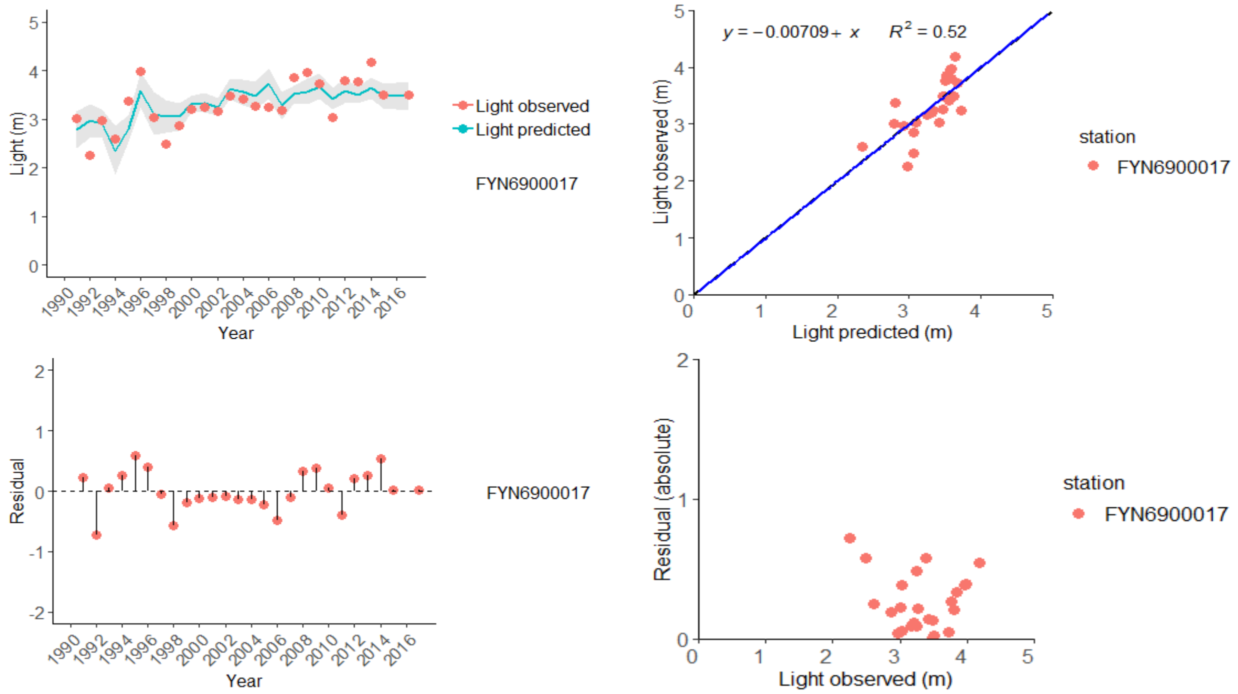


Figure A14. Odense indre, St. FYN6910008

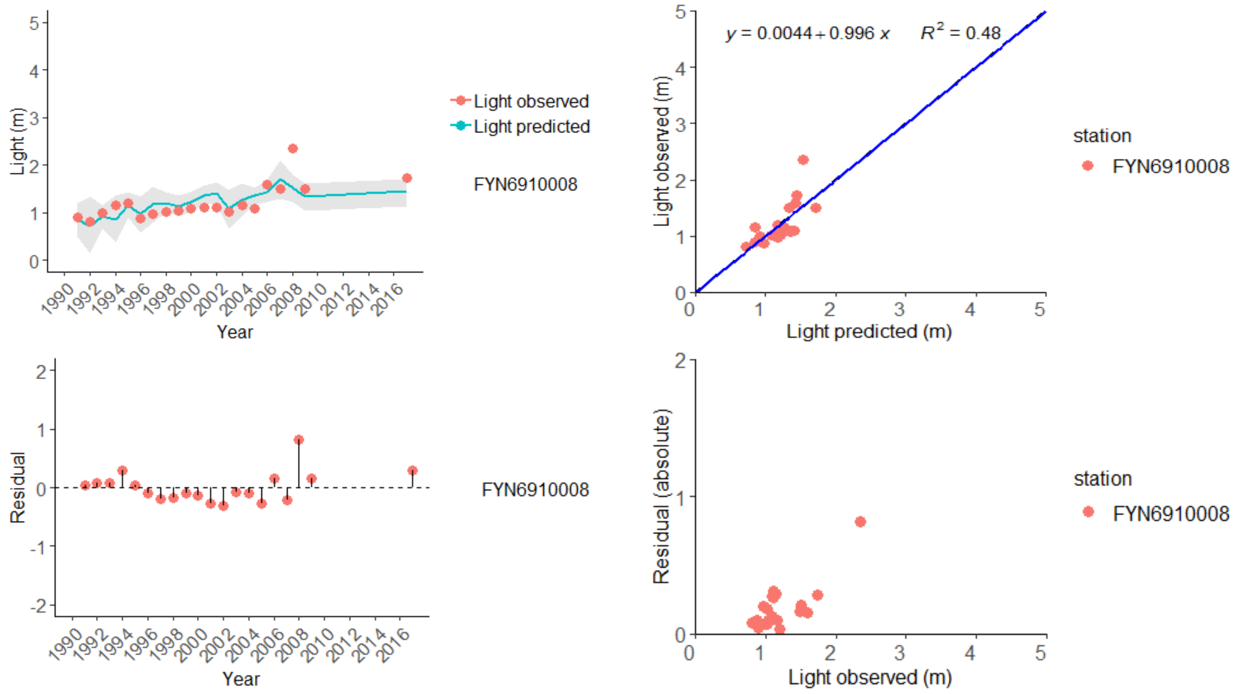


Figure A15. Mariager fjord, St. NOR5503

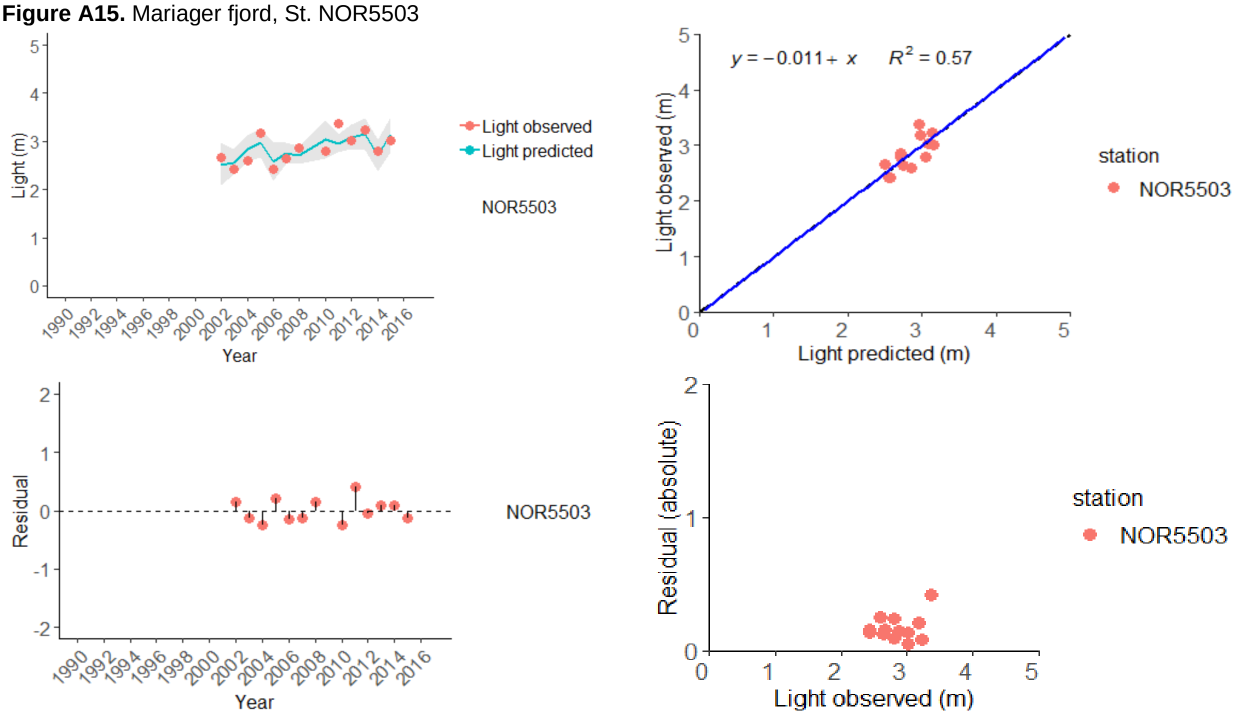


Figure A16. Vadehavet Grådyb - Ho Bugt v Langli, St. RIB1610002

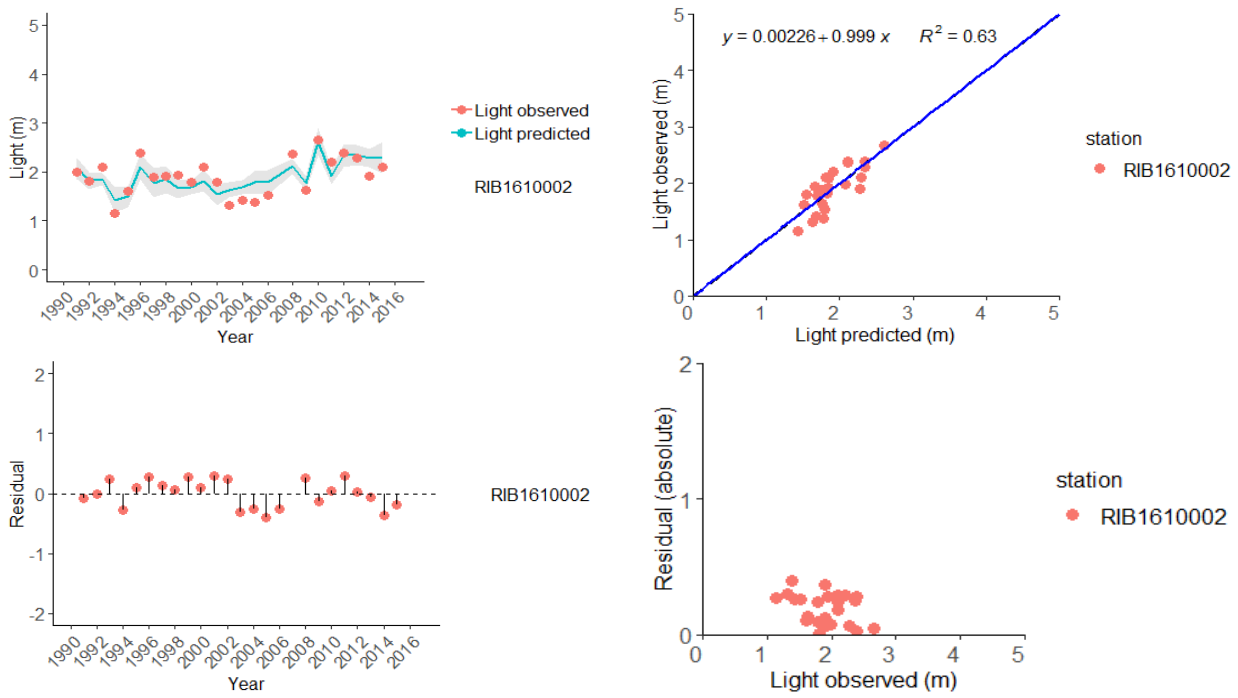


Figure A17. Vadehavet Knudedyb, St. RIB1620014

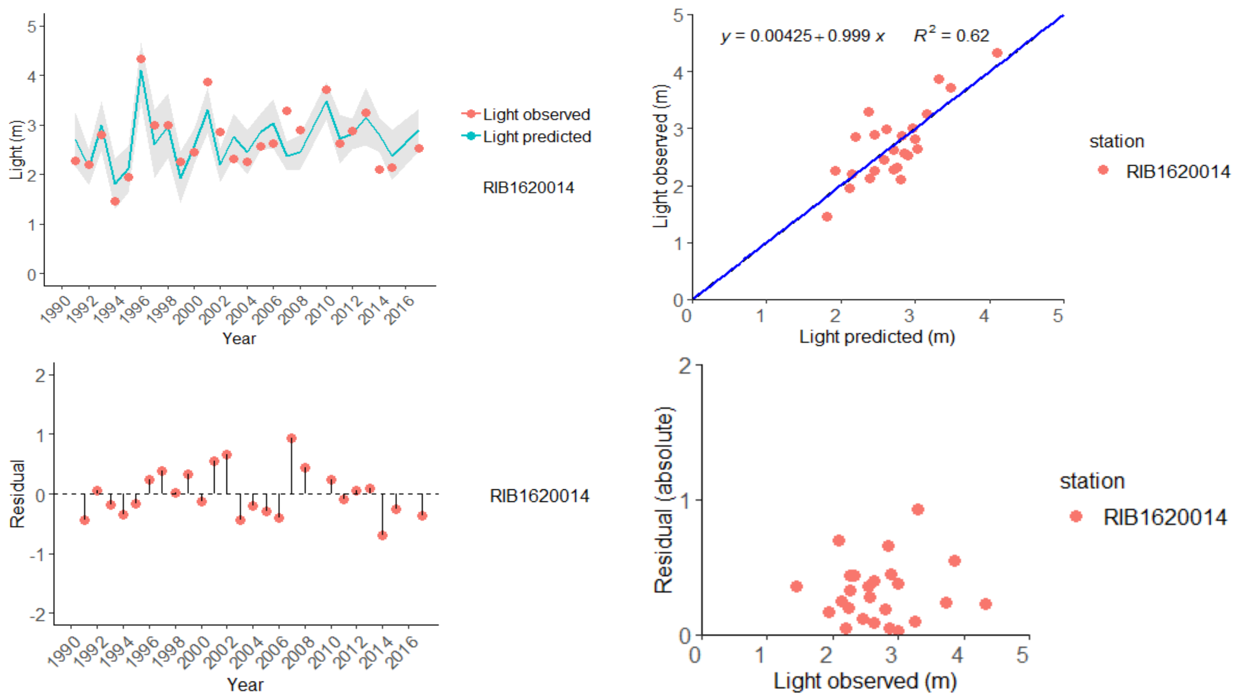


Figure A18. Ringkøbing Fjord nord, St. RKB1

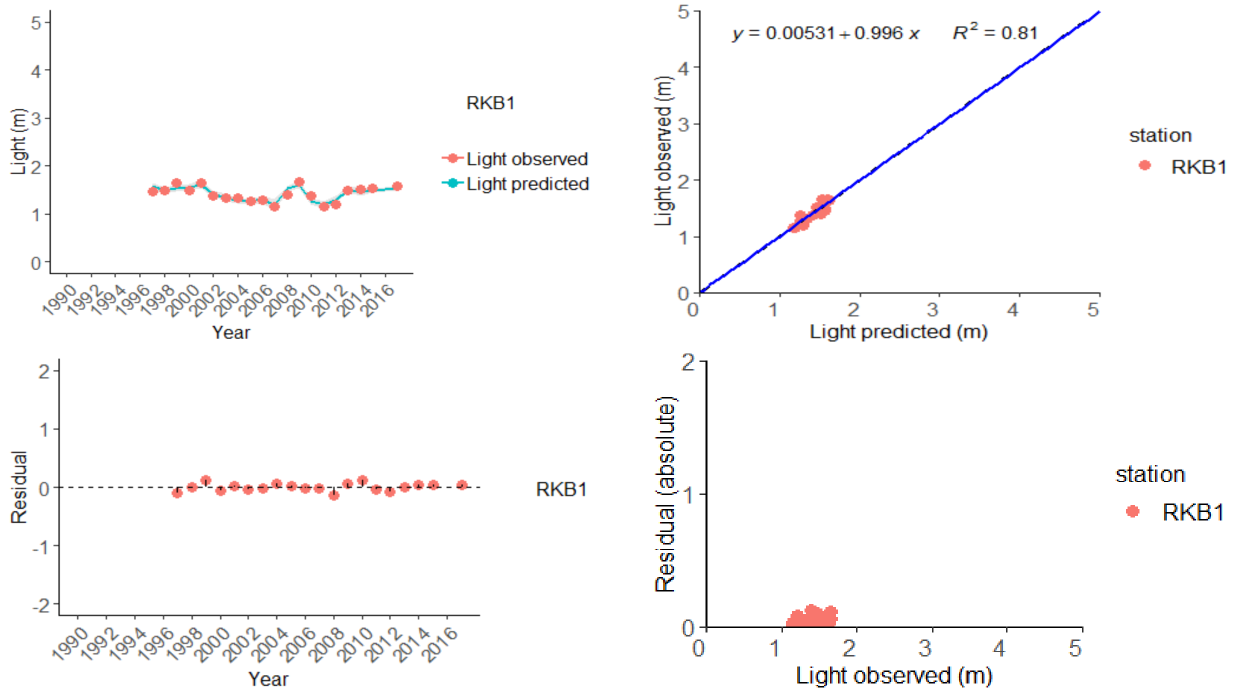


Figure A19. Nissum fjord, St. RKB22

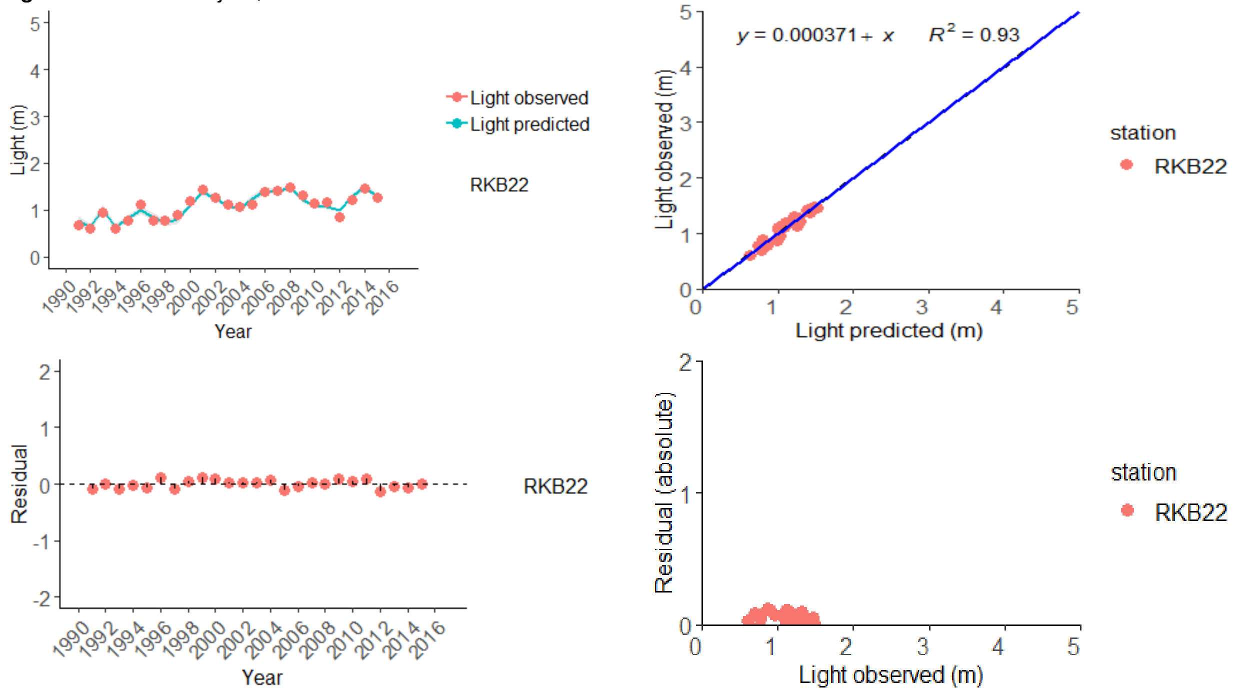


Figure A20. Køge Bugt, St. ROS1727

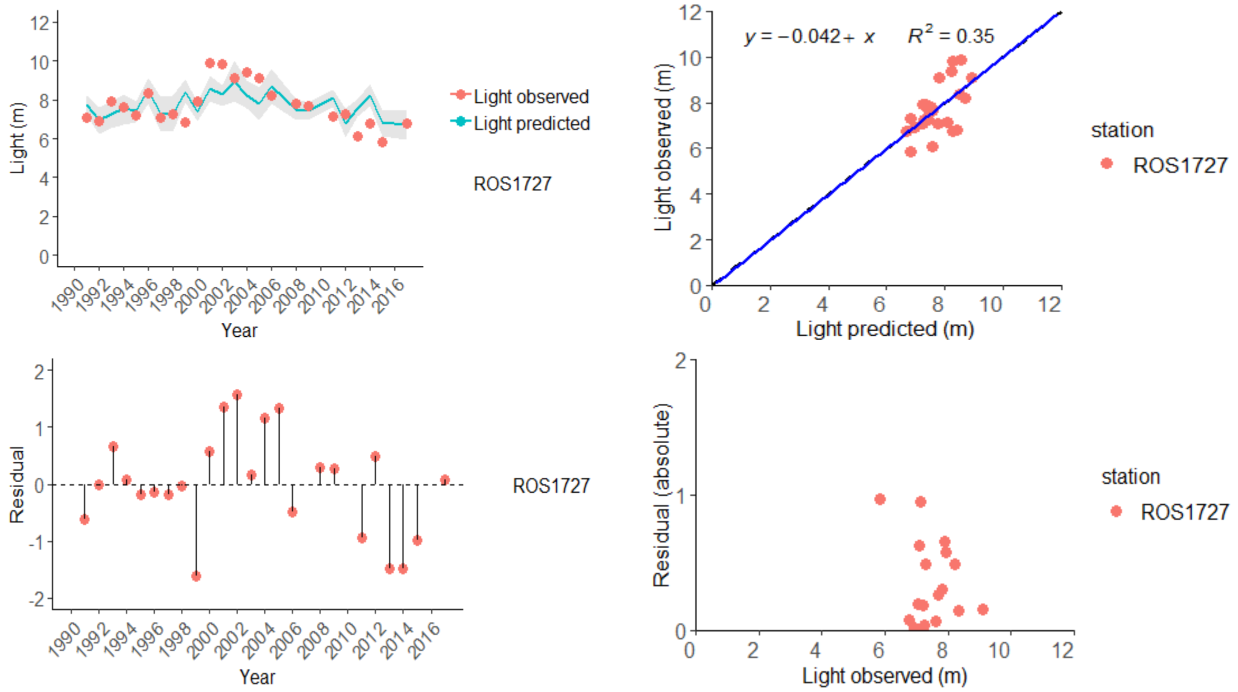


Figure A21. Roskilde indre, St. ROS60

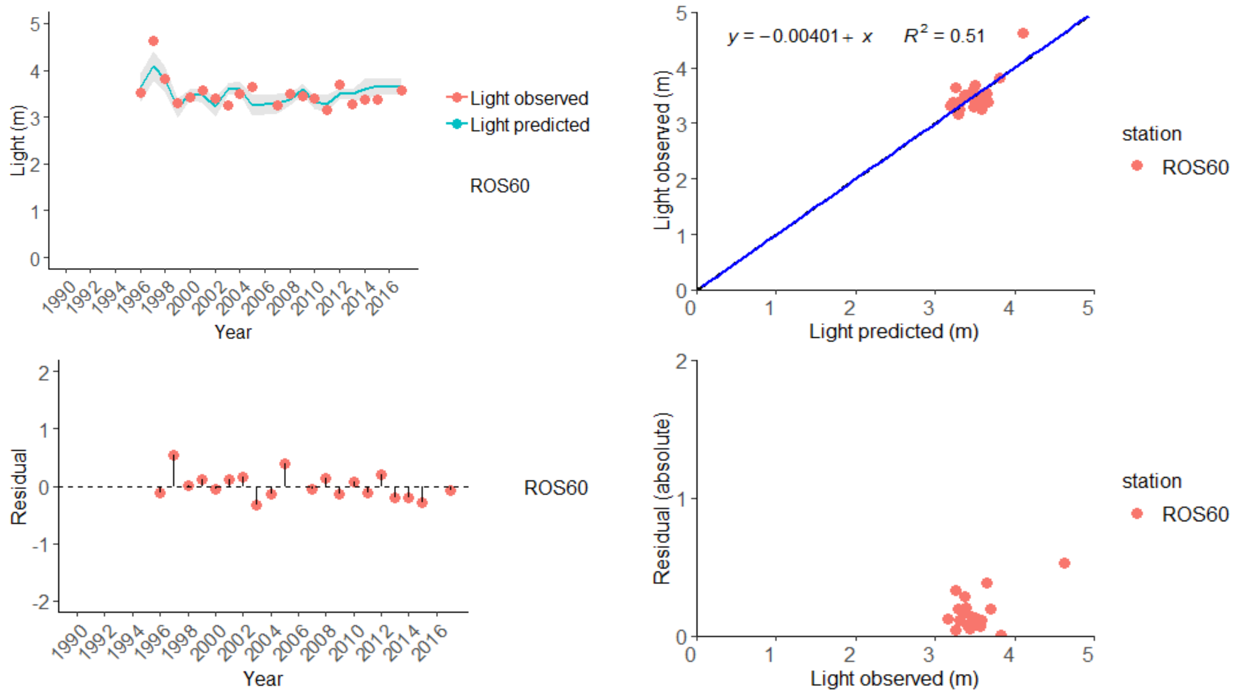


Figure 22A. Lister Dyb St. SJY1

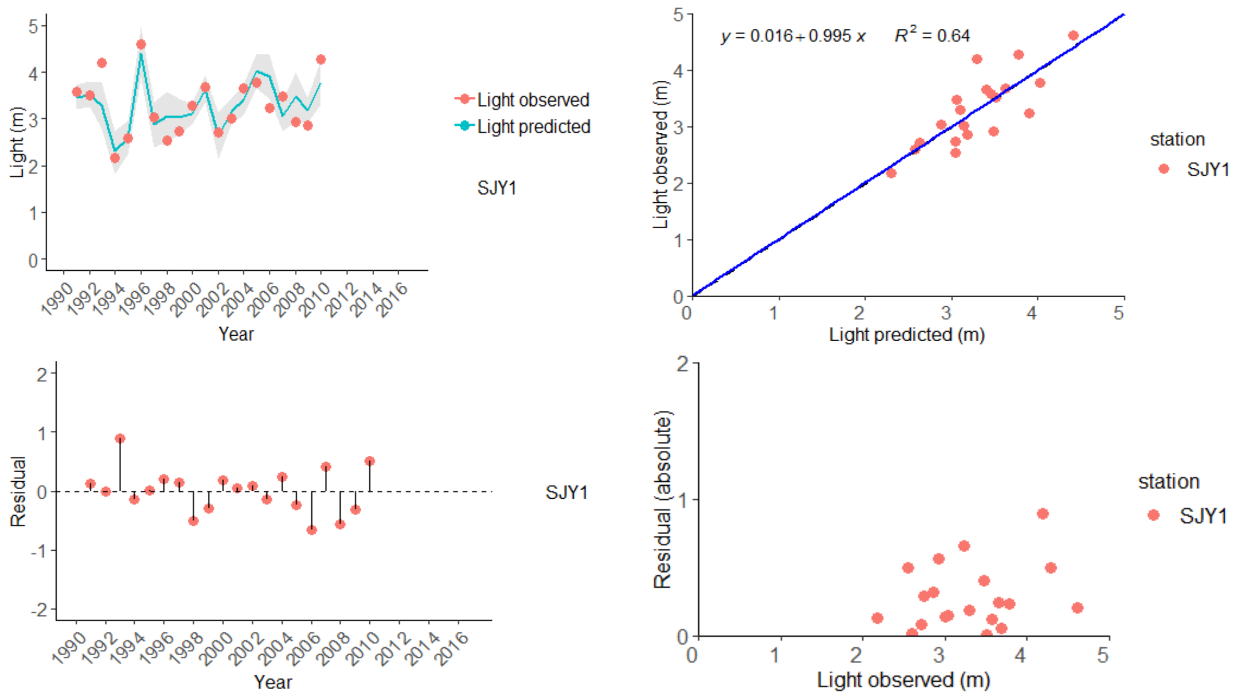


Figure 23A. Augustenborg Fjord, St. SJY12

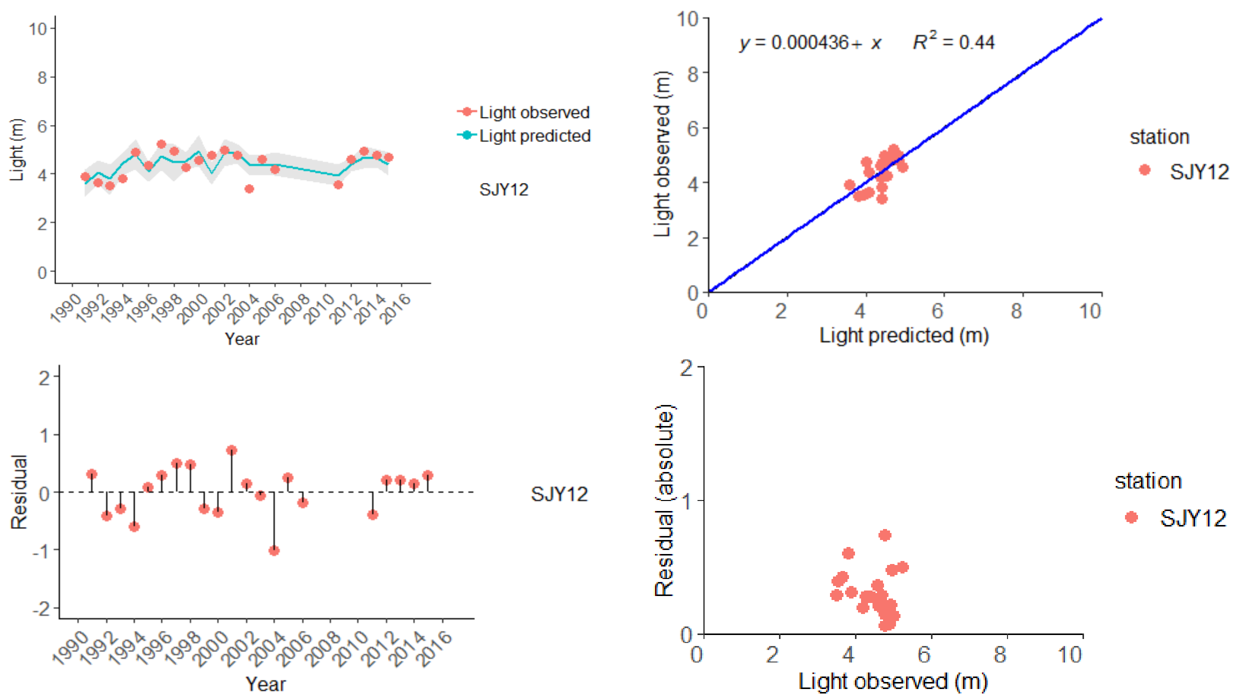


Figure 24A. Aabenraa Fjord, St. SJY15

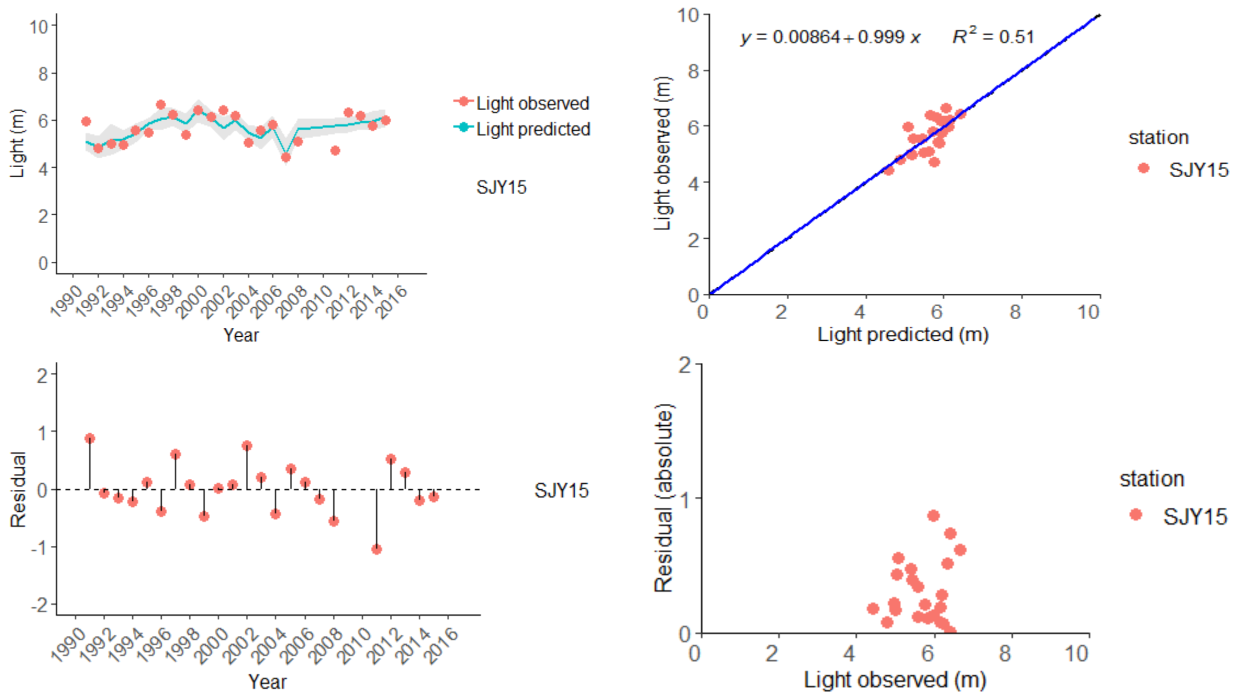


Figure 25A. Lister Dyb v Rømø Havneby, St. SJY3

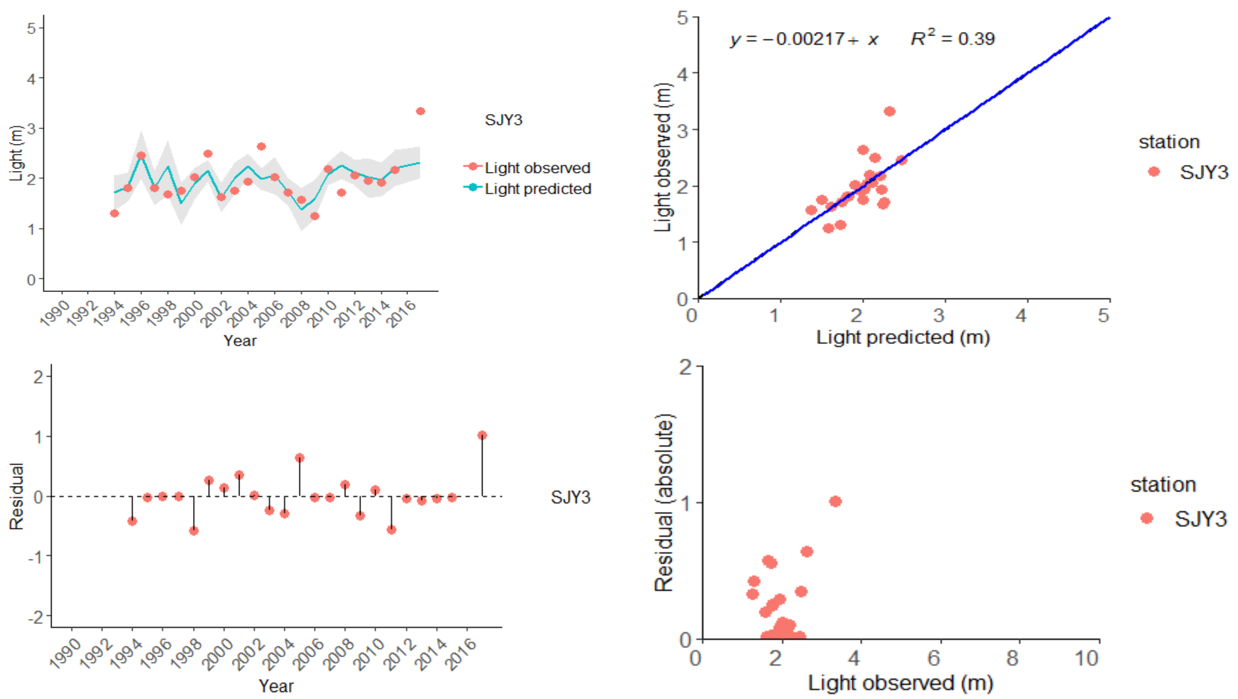


Figure 26A. Flensborg Fjord inder, St. SJYKFF2

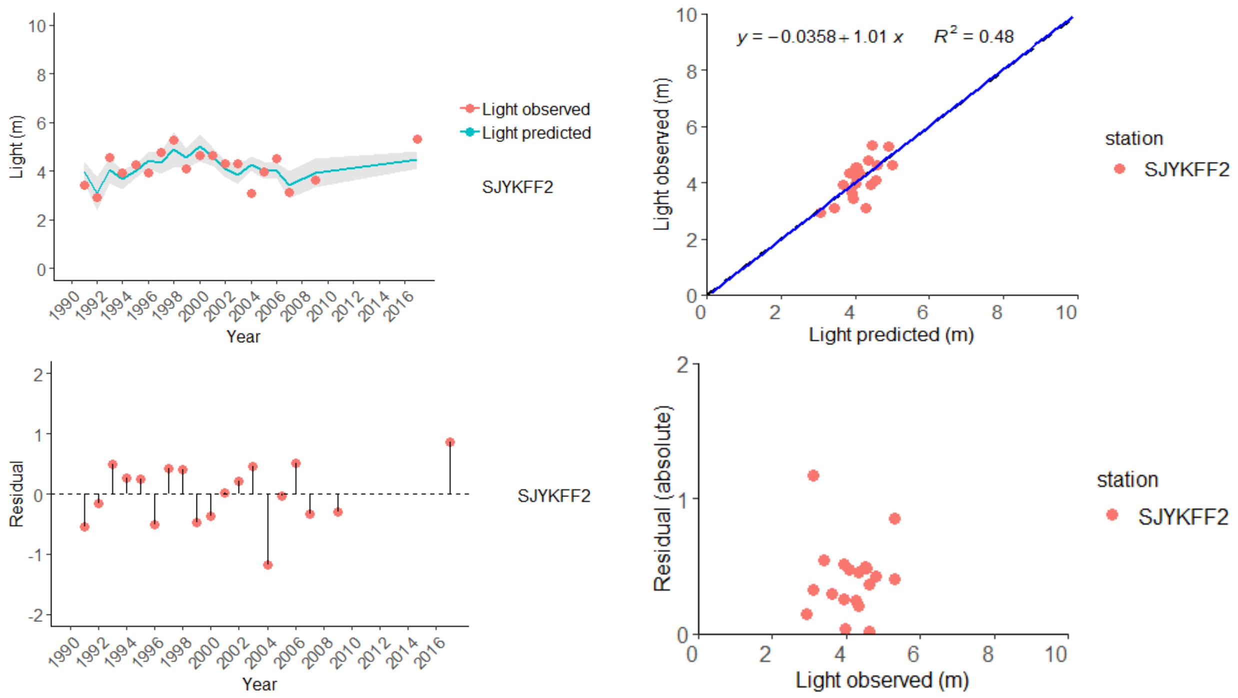


Figure 27A. Flensborg Fjord yder, St. SJYKFF5

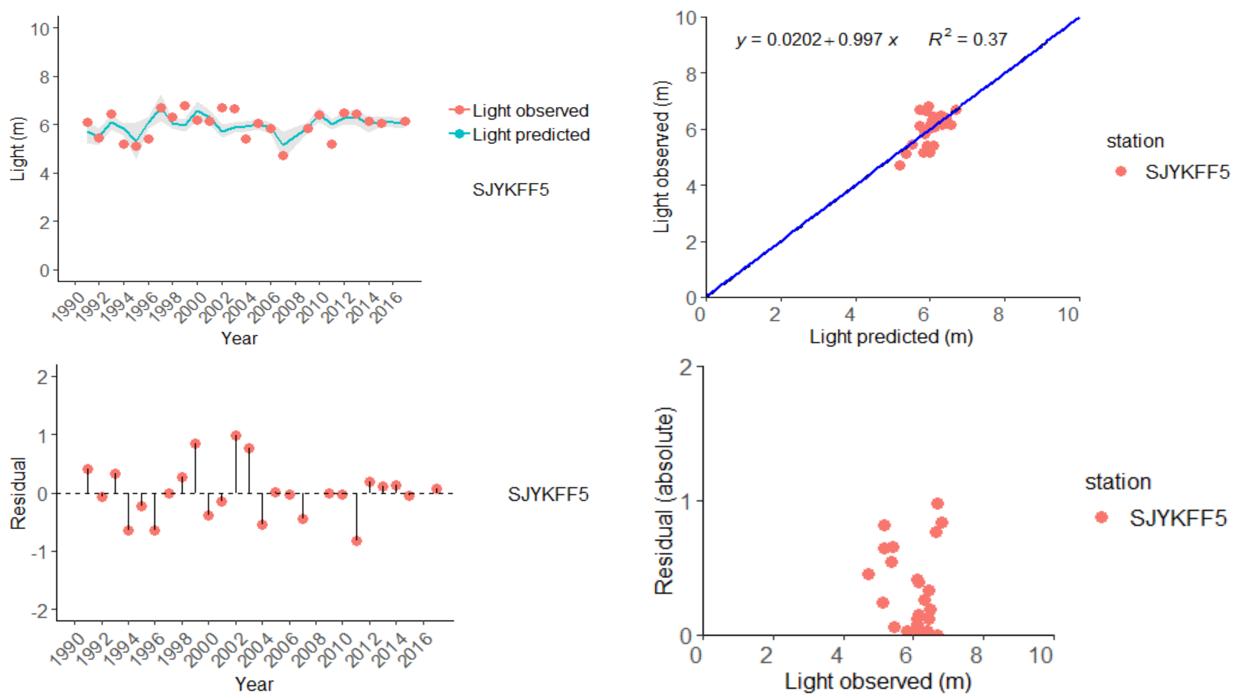


Figure 28A. Præstø Fjord, St. STO0802008

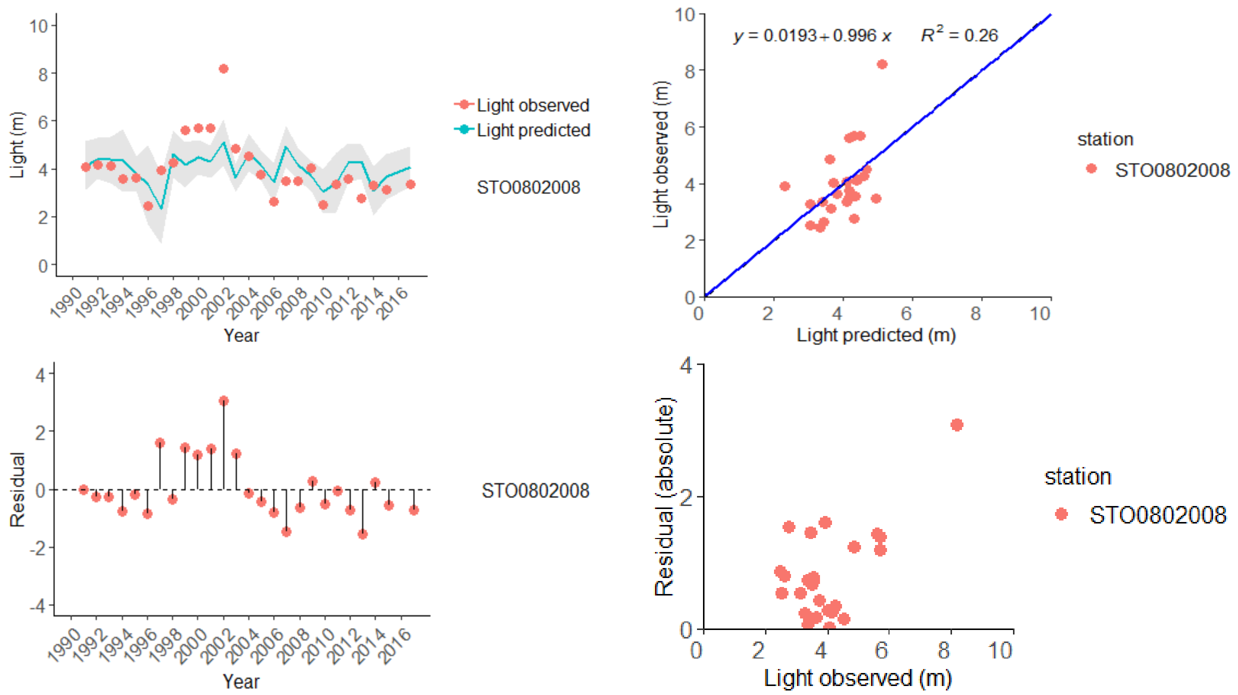


Figure 29A. Kolding Fjord, St. VEJ0003350

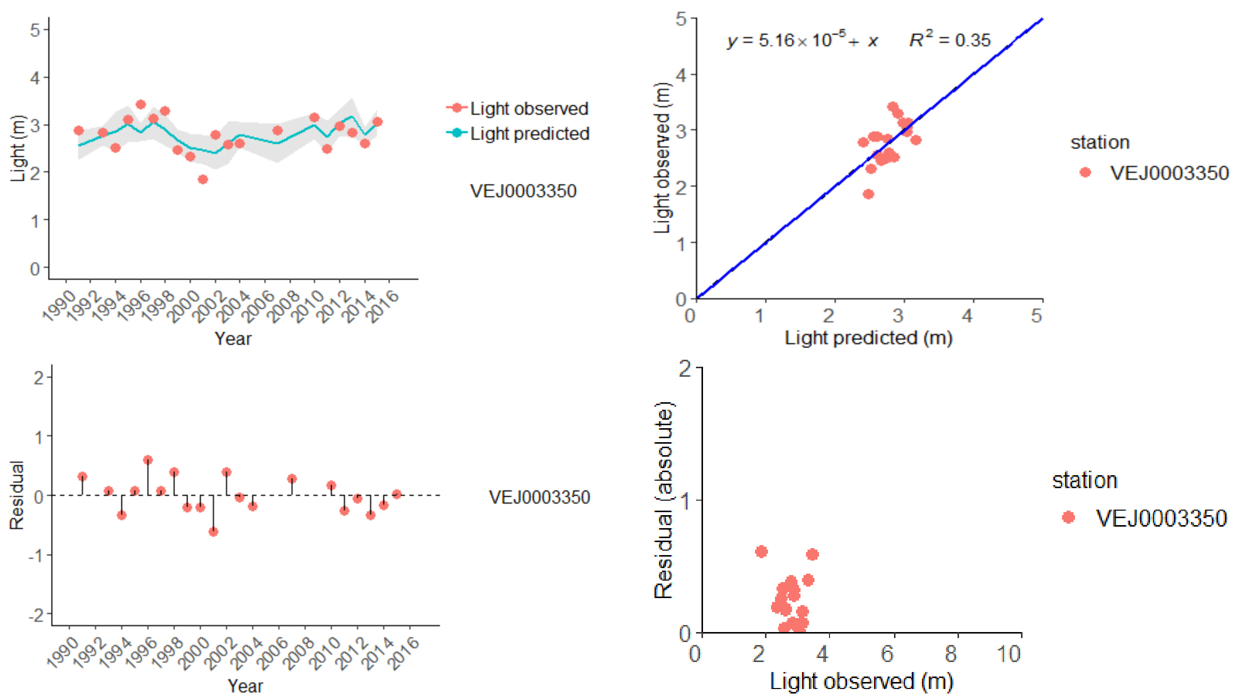


Figure 30A. Vejle Fjord, St. VEJ0004273

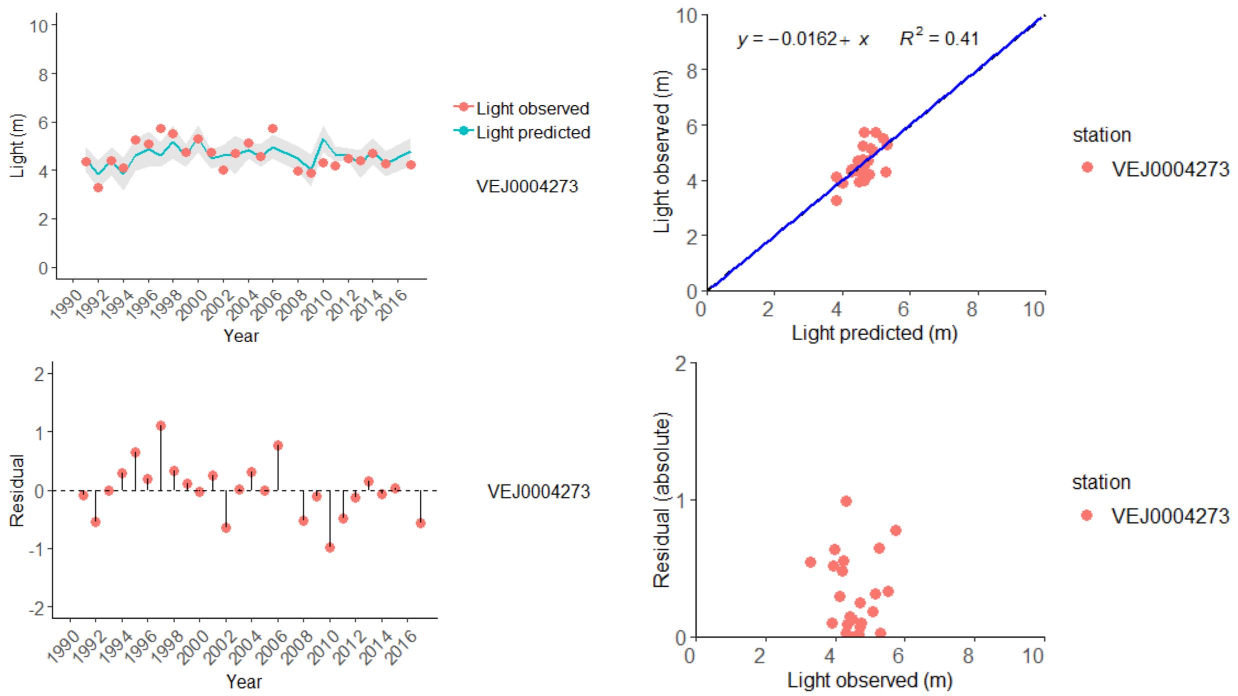


Figure 31A. Horsens inder, St. VEJ0005790

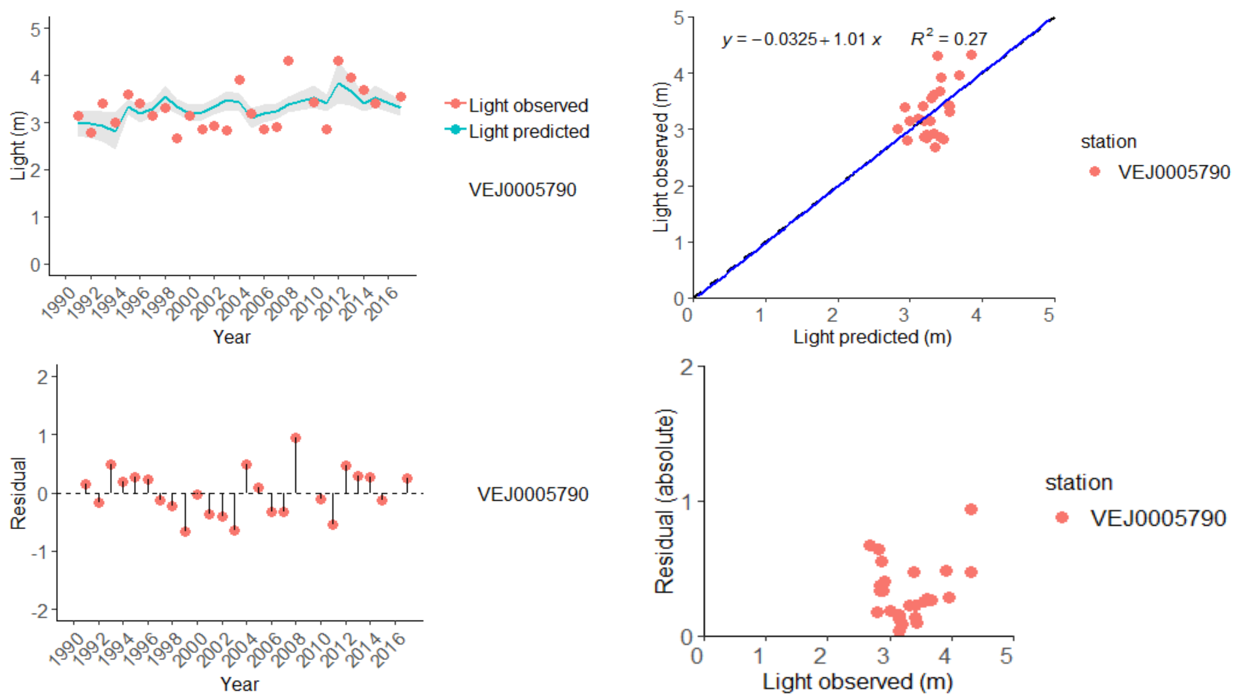


Figure 32A. Horsens yder, St. VEJ0006489

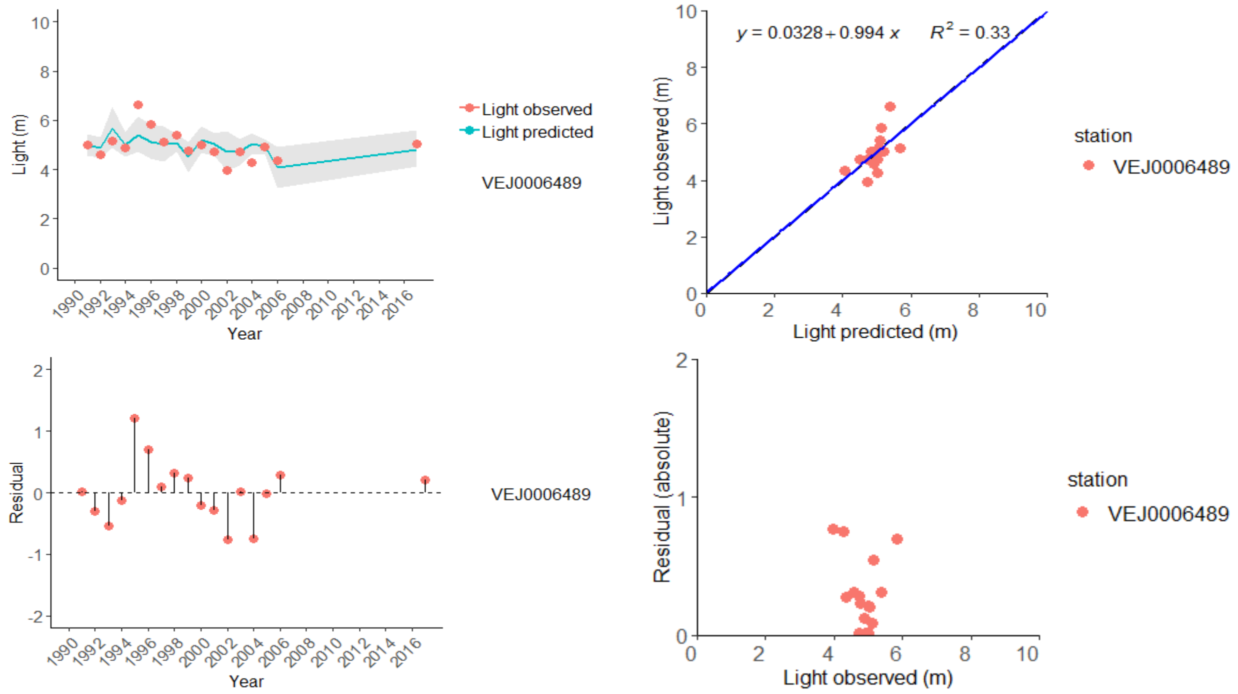


Figure 33A. Nissum Bredning, St. VIB3702-00001

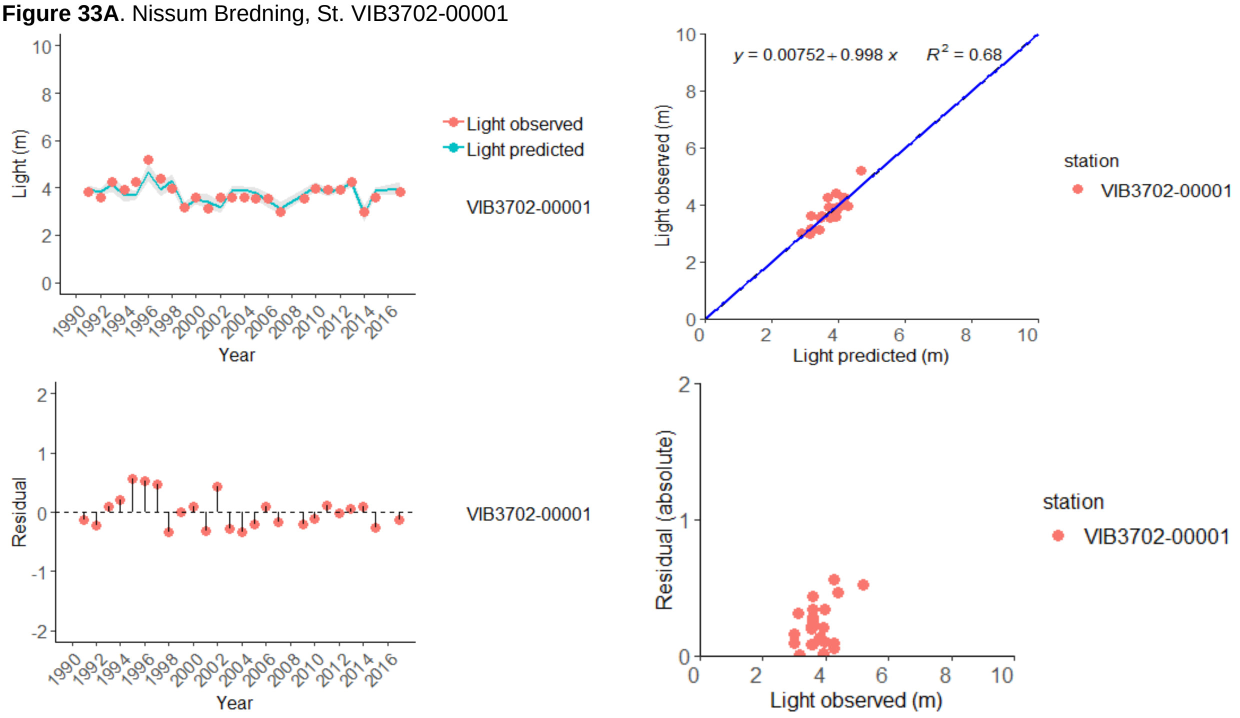


Figure 34A. Løgstør bredning, St. VIB3708-00001

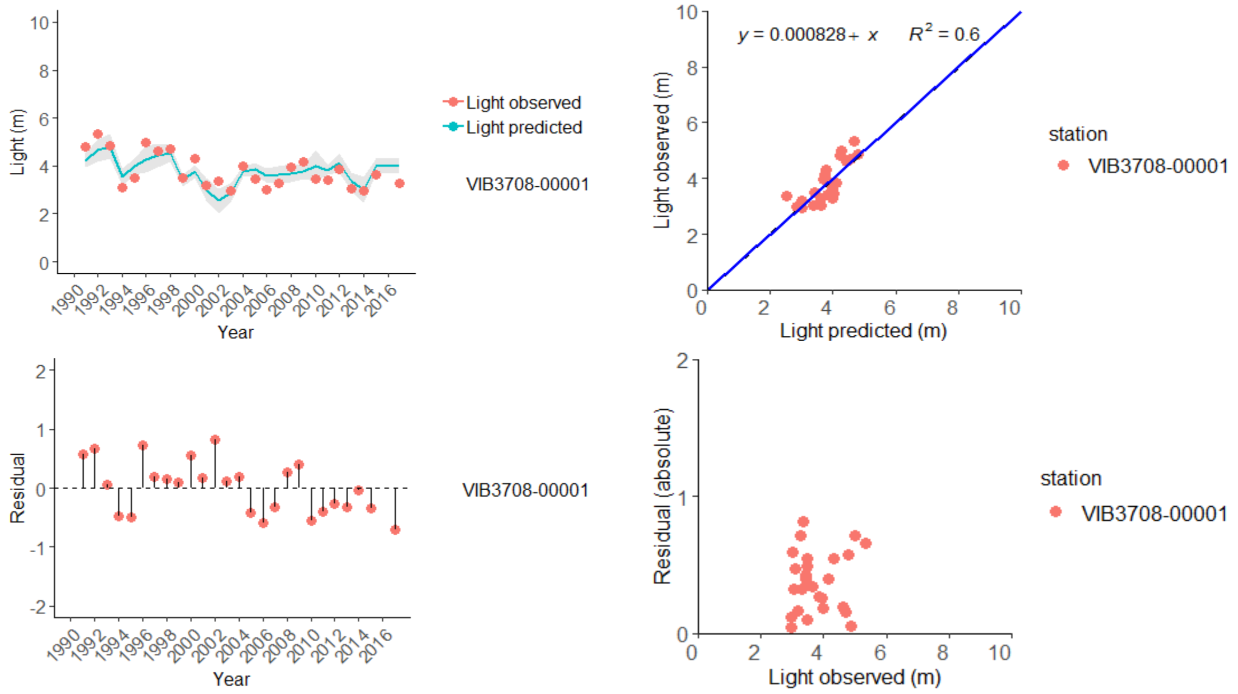


Figure 35A. Nibe Bredning, St. VIB3711-00001

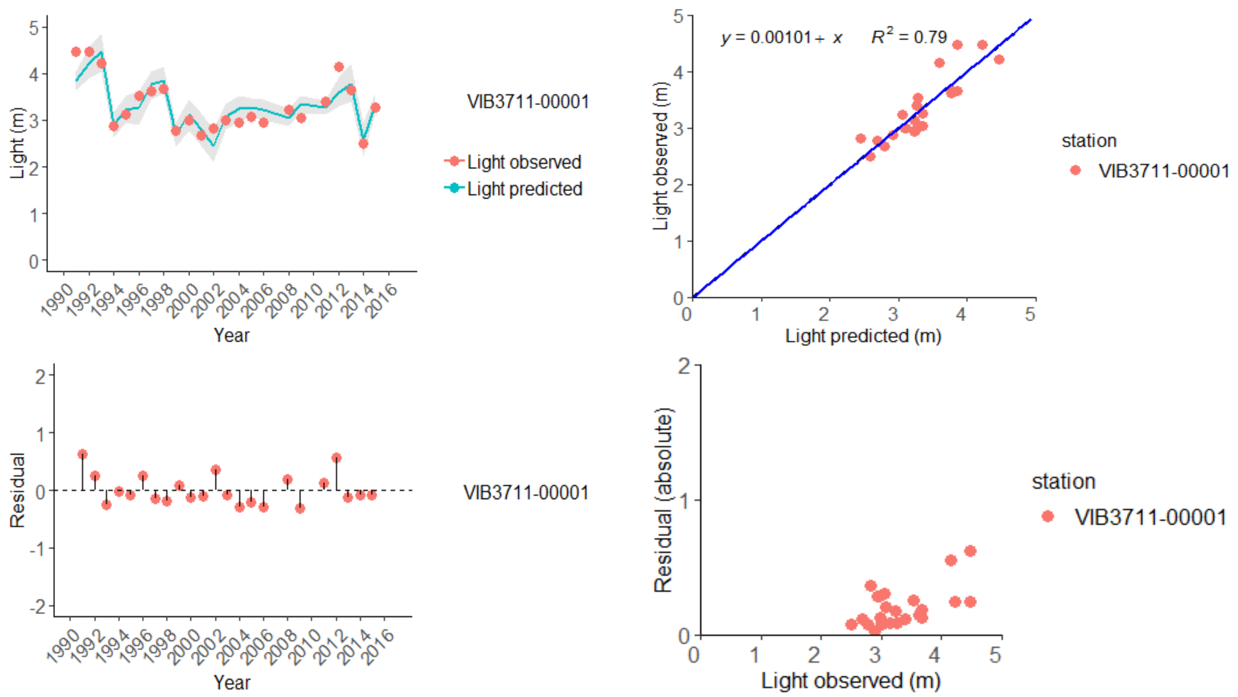


Figure 36A. Thisted Bredning, St. VIB3723-00001

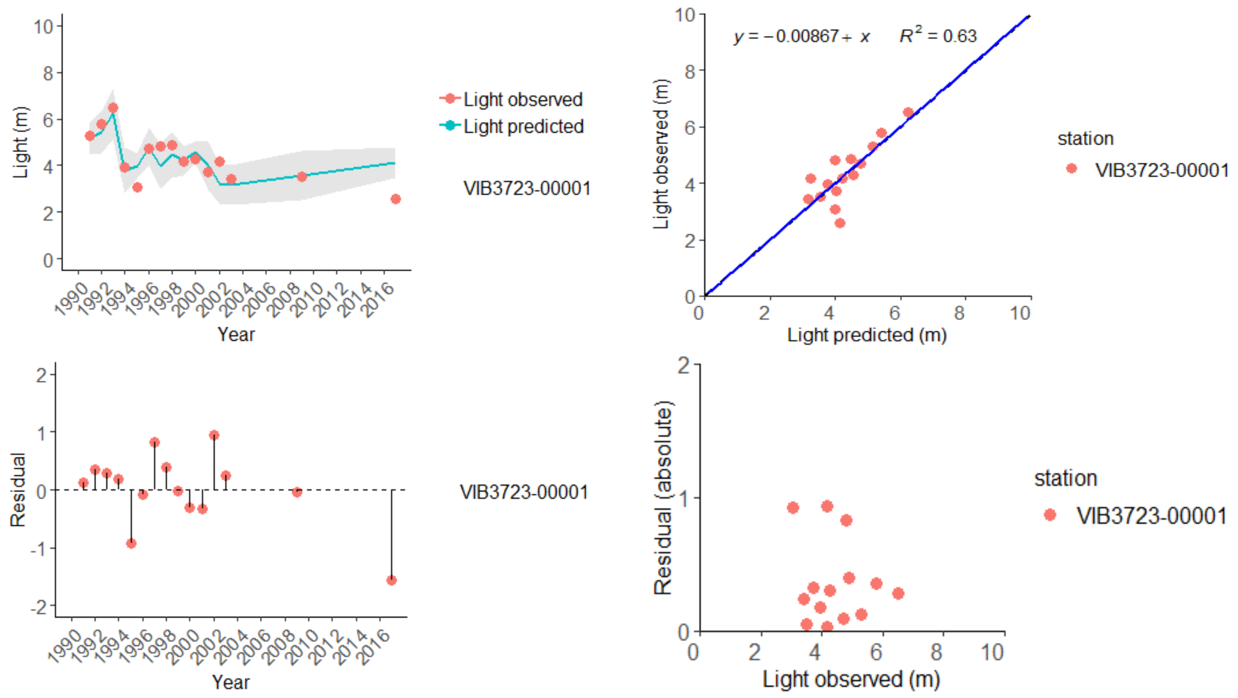


Figure 37A. Skive Fjord, St. VIB3727-00001

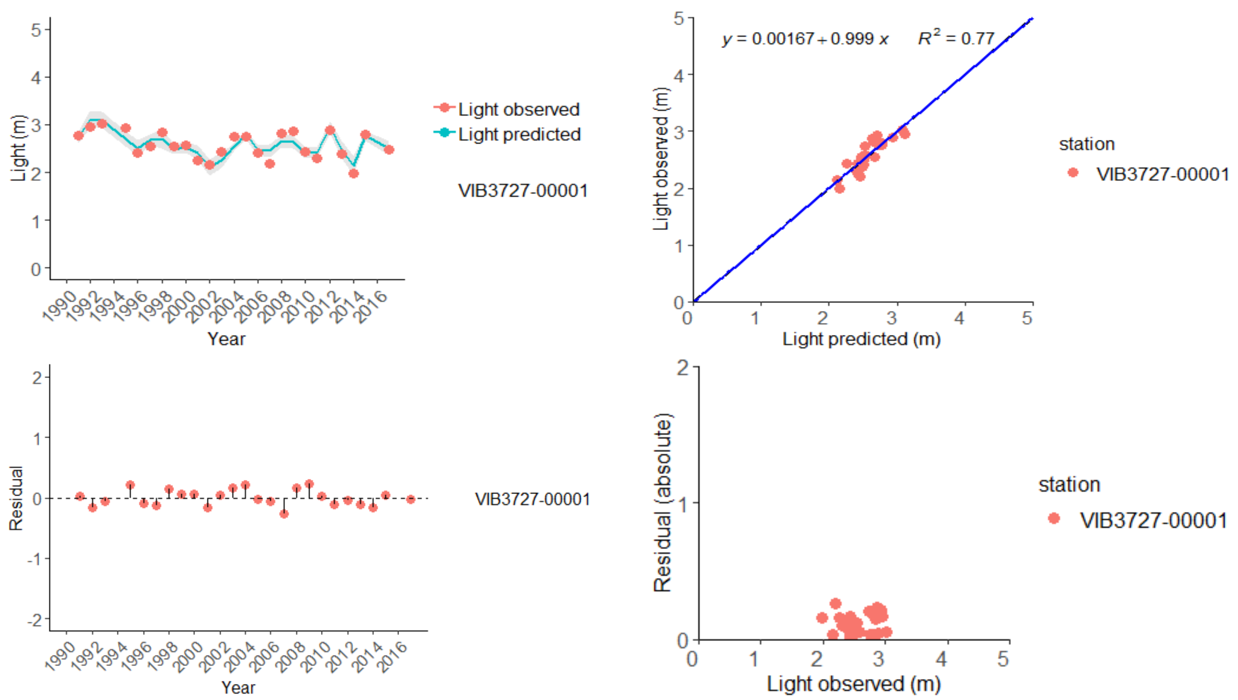


Figure 38A. Lovns Bredning, St. VIB3728-00001

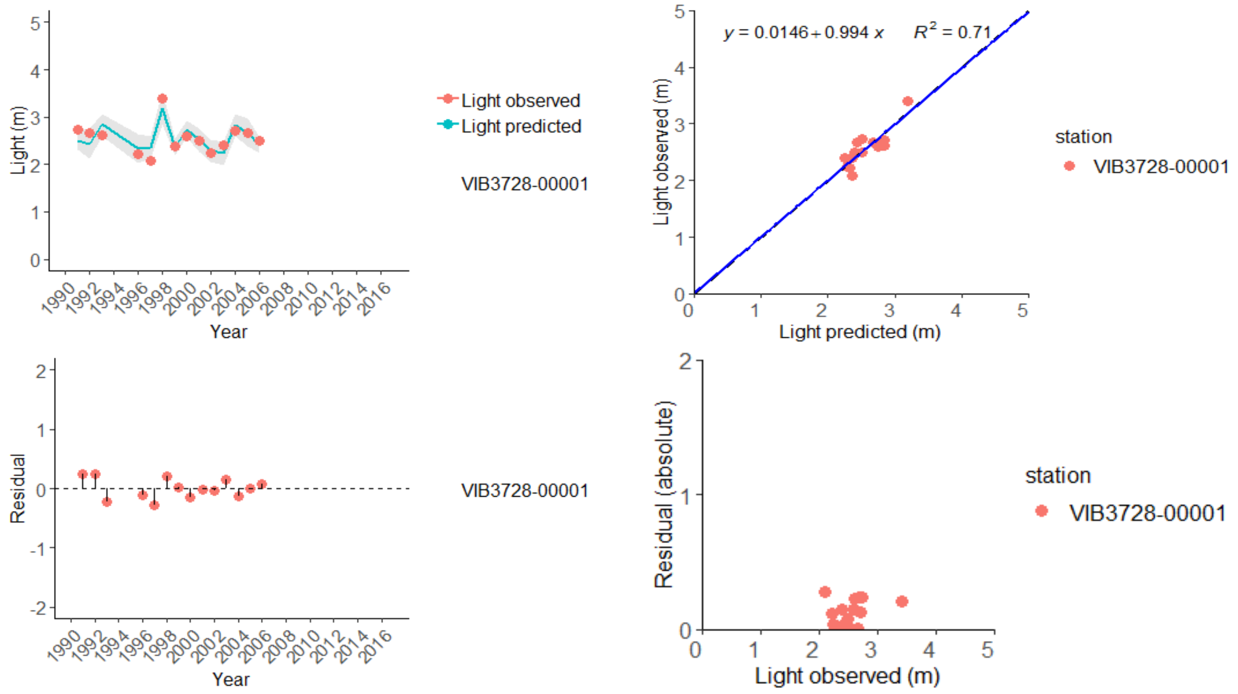


Figure 39A. Hjarbæk fjord, St. VIB3729-00001

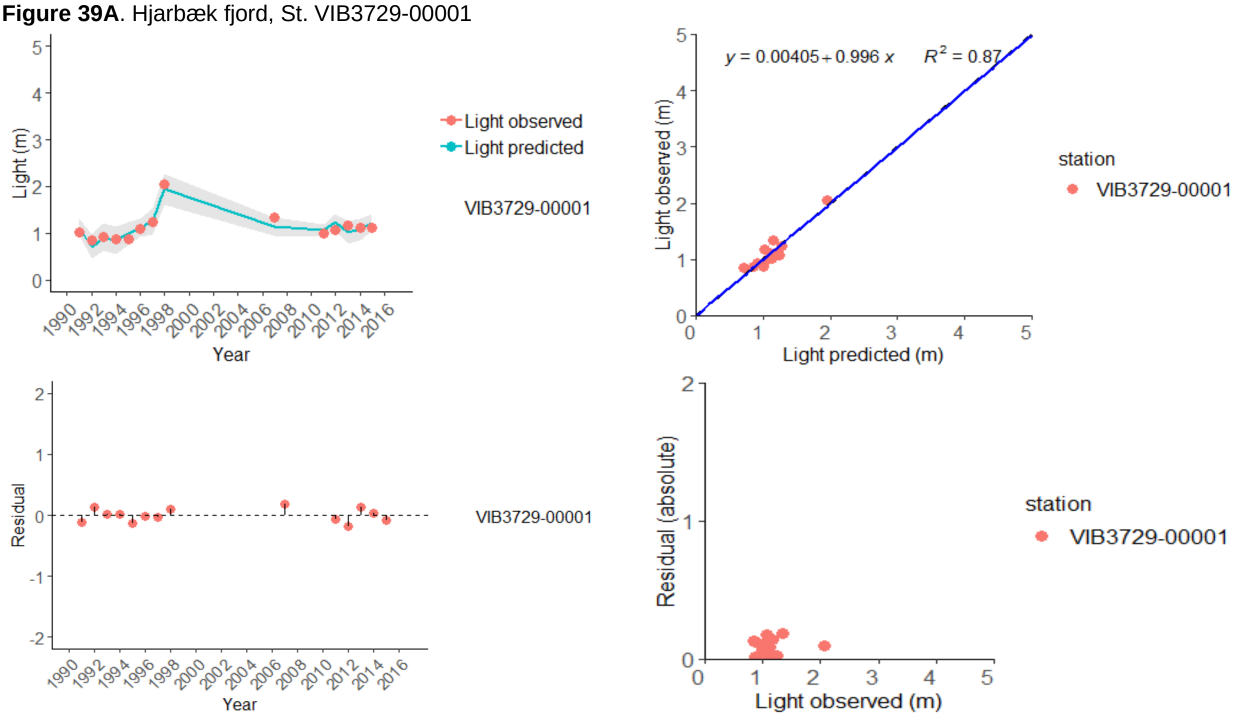


Figure 40A. Isefjord dybt basin, St. VSJ10003

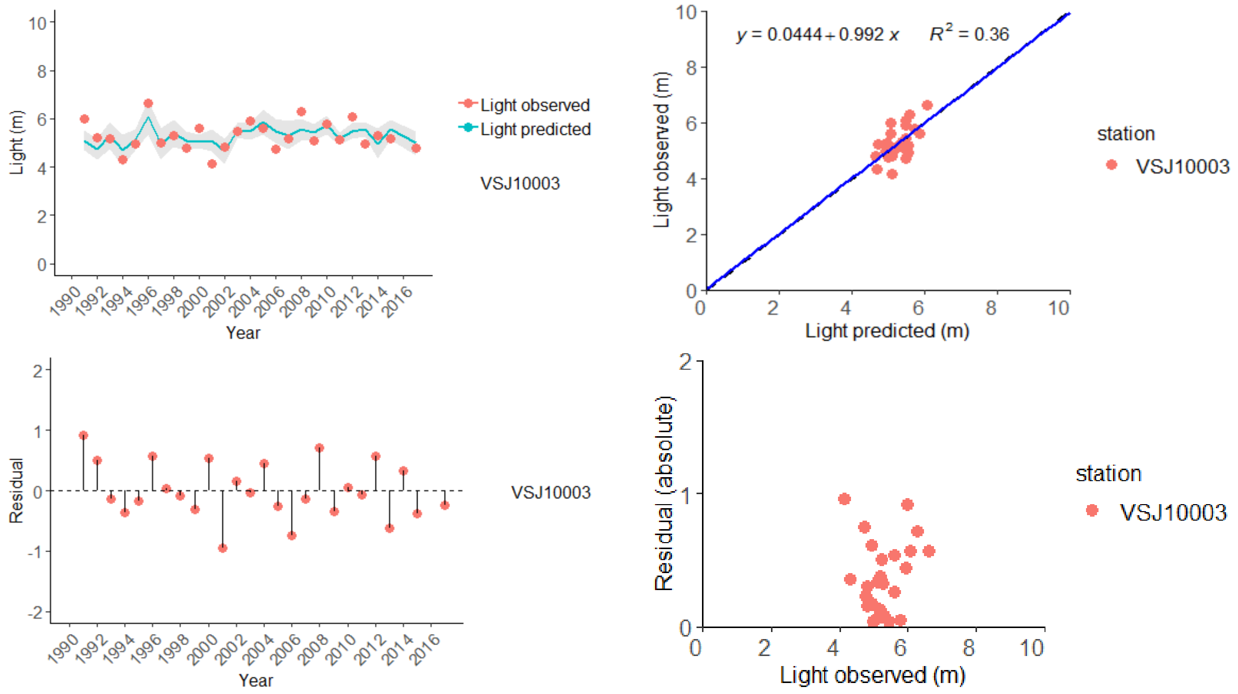


Figure 41A. Isefjord inderbredning, St. VSJ10006

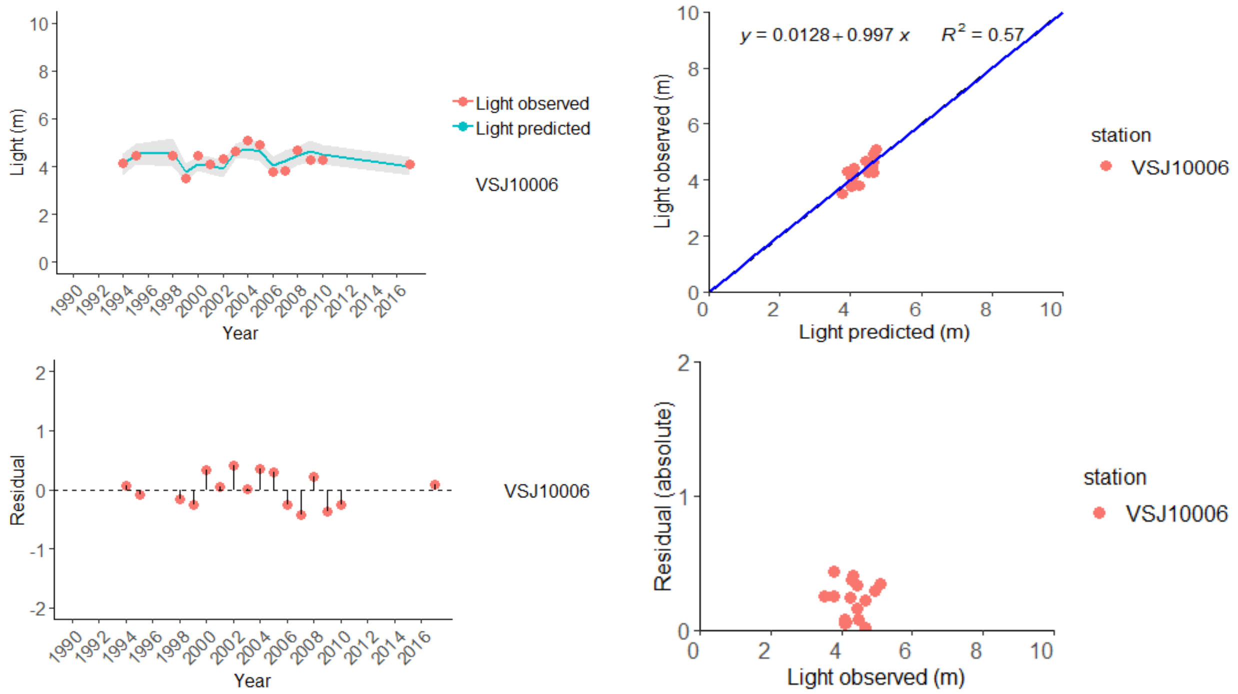


Figure 42A. Kalundborg Fjord yder, St. VSJ41007

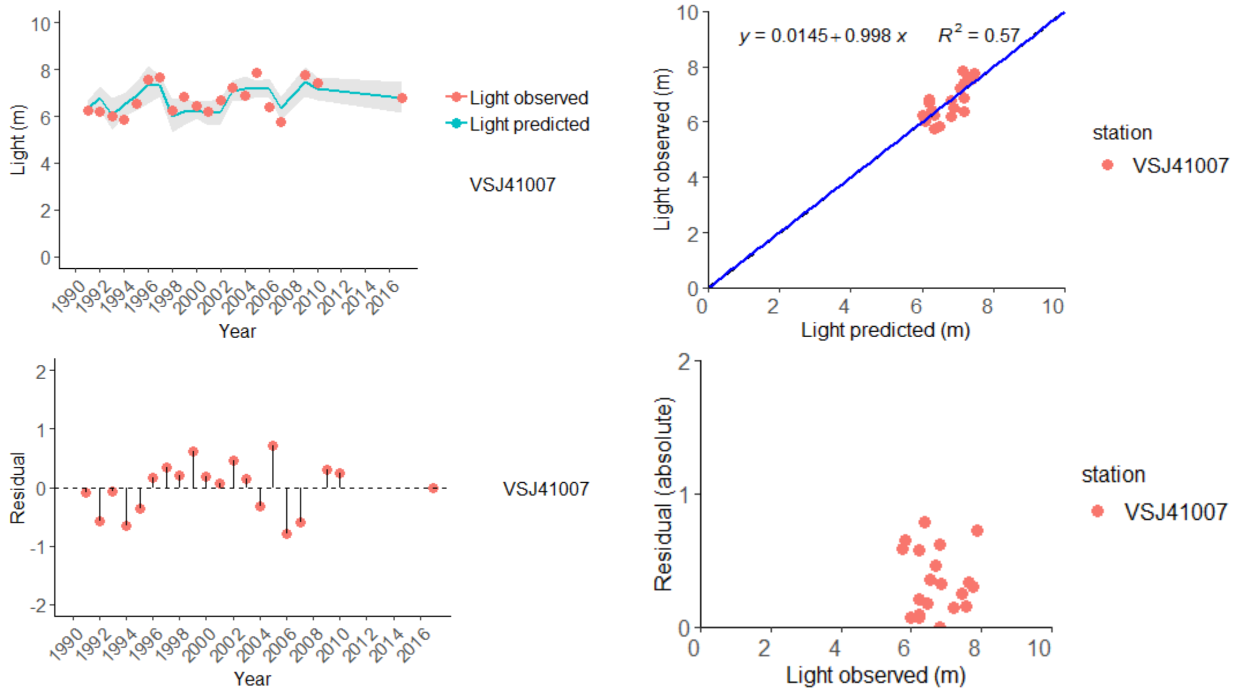


Figure 43A. Kalundborg Fjord inder, St. VSJ41008

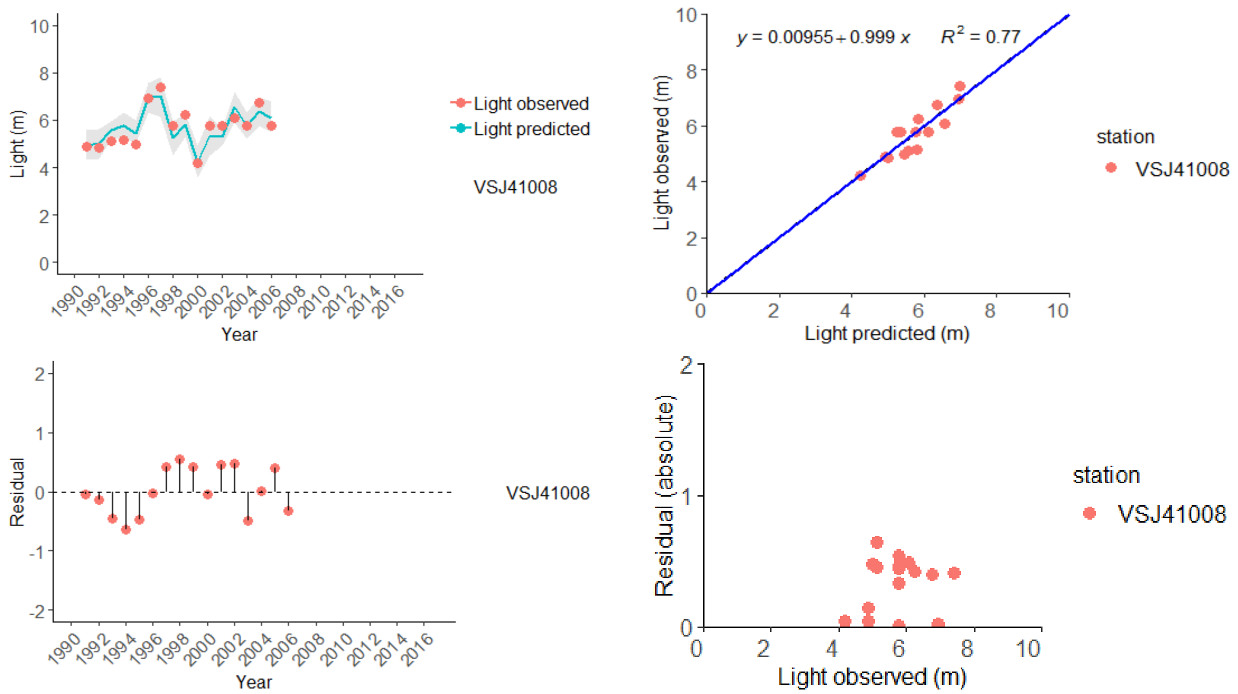
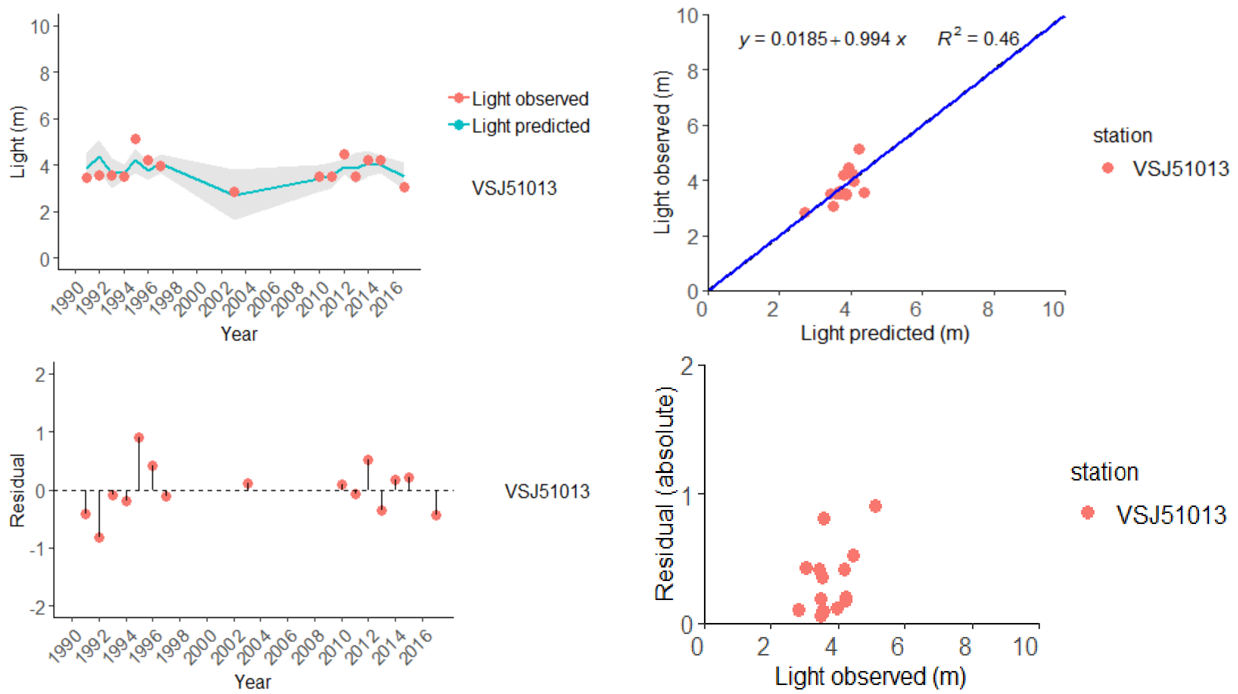


Figure 44A. Skælskør Fjord, St. VSJ51013



MODELLING LIGHT CONDITIONS IN DANISH COASTAL WATERS USING A BAYESIAN MODELLING APPROACH

Model documentation

During this project, a number of Bayesian models were developed to model the light limiting depth distribution of eelgrass in Danish coastal waterbodies. Eelgrass is a marine ecosystem key species and regulated by the EU water frame directive. Here we model the maximum depth limit of eelgrass based on light limitation. We have assumed that the regional light limitation for eelgrass is around 16 % of surface irradiance. The light climate is calculated based on the national environmental monitoring data and modelled as a function of nutrient loadings, climate and physio-chemical measures of the water. In total, we have made models for 44 monitoring stations, covering 41 water bodies; 23 of the models include nutrient loadings from the local catchment and hence they can be used to regulate nutrient loadings from land.

ISBN: 978-87-7156-556-0
ISSN: 2245-0203