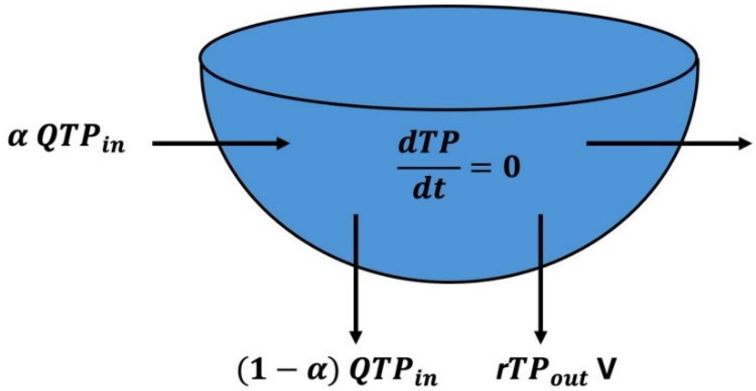
Updated phosphorus retention model for Danish lakes

A note on statistical modeling of total phosphorus retention under steady-state assumptions

Scientific note from DCE – Danish Centre for Environment and Energy

Date: 6 February 2026 | **08**

½



AARHUS UNIVERSITY
DCE – DANISH CENTRE FOR ENVIRONMENT AND ENERGY

Data sheet

Scientific note from DCE – Danish Centre for Environment and Energy

Category: Scientific briefing

Title: Updated phosphorus retention model for Danish lakes
Subtitle: A note on statistical modeling of total phosphorus retention under steady-state assumptions

Authors: Robert Ladwig, Martin Søndergaard, Peter Borgen Sørensen, Tuba Bucak and Liselotte Sander Johansson
Institutions: Aarhus University, Department of Ecoscience

Referees: Torben Linding Lauridsen (report), Zachary Paul Simpson (model code)
Quality assurance, DCE: Signe Jung-Madsen
Linguistic QA: Anne Mette Poulsen

External comment: The Agency for Green Transition and Aquatic Environment (SGAV). The comments can be found here

Claimant: The Agency for Green Transition and Aquatic Environment (SGAV)

Please cite as: Ladwig, R., Søndergaard, M., Sørensen, P.B., Bucak, T. & Johansson, L.S.. 2026. Updated phosphorus retention model for Danish lakes. A note on statistical modeling of total phosphorus retention under steady-state assumptions. Aarhus University, DCE - Danish Centre for Environment and Energy, 26 pp. – Scientific note no. 2026|08

Front page: Figure 1 in the note

Number of pages: 26

Contents

1 Preface

This report describes the development and performance of an updated model for total phosphorus retention in Danish lakes using a steady-state mass balance approach. This report has been prepared at the request by the Danish Agency for Green Transition and Aquatic Environment (Styrelsen for Grøn Arealomlægning og Vandmiljø, SGAV), which has also had the opportunity to comment on a draft of the report. Preliminary results have been presented and discussed at meetings with SGAV.

2 Summary

This report describes the development of an updated retention model for total phosphorus (TP) in lakes based on previous reports (Søndergaard et al. 2020; Trolle et al. 2015). In contrast to earlier works, the updated phosphorus mass balance model is statistically fitted to outflow TP concentrations. As phosphorus retention in lakes is challenging to measure and monitor, the mass balance model is used to infer phosphorus retention. The updated model demonstrates improved performance compared to the ones from previous reports, with an R^2 of 0.77 for predicting outflow TP concentrations. Further, a simplified exponential model relating phosphorus retention to lake area and outflow achieved an R^2 of 0.9, highlighting its potential for upscaling. Nonetheless, to improve future accuracy of lake phosphorus retention modeling, it is recommended to transition from steady-state to dynamic modeling on sub-annual time scales, which may better capture phosphorus retention and release in lakes (including internal loading).

3 Sammenfatning

Denne rapport beskriver udviklingen af en opdateret model for tilbageholdelsen af totalfosfor (TP) i søer. Rapporten er baseret på tidligere rapporter (Søndergaard et al. 2020; Trolle et al. 2015, men i modsætning til disse er den opdaterede fosformassebalancemodel statistisk tilpasset udløbs-TP-koncentrationer. Massebalancemodellen til at beregne fosfortilbageholdelsen anvendes, da fosfortilbageholdelsen i søer er vanskelig at måle og overvåge.

Den opdaterede model viser forbedret ydeevne sammenlignet med modellerne fra tidligere rapporter. Således opnås en R^2-værdi på 0,77 for forudsigelse af udløbskoncentrationer af TP. Desuden opnåede en forenklet eksponentiel model, der relaterer fosforretention til søareal og udløb, en R^2-værdi på 0,9, hvilket understreger modellens potentiale for opskalering. For fremover at forbedre nøjagtigheden af modellering af fosforretention i søer anbefales det dog at skifte fra steady-state-modellering til dynamisk modellering, og at det fortages på sæsonmæssige tidsskalaer, så det potentielt kan blive lettere at beskrive fosforretention og -frigivelse i søer (herunder intern belastning).

4 Background and objectives

The implementation of the EU Water Framework Directive requires that good ecological status is achieved in Danish fresh waters. The directive has been implemented in the River Basin Management Plans, which include 985 lakes. To meet the requirement for good ecological status, a wide range of measures has been established (see Executive Order no. 797 of 13 June 2023 on action programs for water body districts). Nonetheless, a large proportion of the lakes do not currently meet the set environmental objectives mainly due to a continued excessive input of nutrients from their catchments. In lakes, a portion of these external nutrients can be retained, while the remainder will be transported further downstream through the lake's outlet. An important aspect of defining targets for maximum phosphorus concentrations in lakes to achieve a good ecological state is to establish relationships between the phosphorus loading, the in-lake phosphorus concentrations, and the phosphorus fraction retained in the lake ecosystem. Phosphorus loading models, the basis for calculating retention, assume that a certain fraction of external total phosphorus load, TP_{in}, gets retained in a lake by relating its own total phosphorus concentration to the external load: $TP_{lake} \sim TP_{in}$.

This report aims to provide an update to the previous phosphorus retention model developed by Trolle et al. (2015). The previous model was based on an analysis of 23 lakes using a set of Vollenweider-type (OECD) models to relate inflow total phosphorus (TP) to in-lake TP concentrations by fitting various parameters to covariates such as hydraulic residence time. Trolle et al. (2015) re-arranged the best-performing Vollenweider-type models to quantify TP retention in lakes using the general form:

Eq. 1: $$\lambda = \frac{TP_{lake}}{TP_{in}} = \frac{a\frac{TP_{in}^{b}}{\left(1+\sqrt{t_w}\right)^{c}}}{TP_{in}}$$

where λ is the TP fraction in a lake, TP_{lake} is the in-lake TP concentration, TP_{in} is the inflow TP concentration, t_w is the hydraulic residence time expressed as volume over discharge, *V/Q*, and *a, b* and *c* are fitting coefficients.

To support lake management, exponential models of derived TP retention to hydraulic residence time were developed and grouped into deep and shallow lakes, with a mean depth threshold of 3 m to differentiate between the two categories. These derived TP retention models achieved an R^2 of 0.23 (Table 1).

Table 1. Retention models fitted by Trolle et al. (2015) in which $\lambda = \mathbf{TP_{lake}}/\mathbf{TP_{in}}$ with $\mathbf{TP_{lake}}$ as the in-situ TP concentration, $\mathbf{TP_{in}}$ as the inflow TP concentration, and t_w as hydraulic residence time.

Retention model	Form	R^2
All lakes	$\lambda = 0.371 t_w^{0.1477}$	0.19
Shallow lakes	$\lambda = 0.299 t_w^{0.0767}$	0.23
Deep lakes	$\lambda = 0.406 t_w^{0.1206}$	0.23

To inform the River Basin Management Plans (2027 -2033), this report will continue the development of TP retention models using longer time series data and a modified modeling approach. Statistical modeling of inflow to outflow TP is applied to improve estimation of TP retention in lakes over long time scales.

5 Methodology

Lake-specific data on inflow and outflow TP concentrations as well as outflow volume were extracted from the Vandløbsdata (NOVANA, stored in the database VanDa, vanda.miljoeportal.dk, and subsequently extracted from Overfladevandsdatabasen ODA, ODA.dk) and supplemented with data on lake volume, depth and area. Inflow and outflow TP concentrations as well as discharge were aggregated to annual averages. The dataset included 24 lakes with 367 lake-year combinations (from 1990-2021). Previous approaches excluded data prior to 2005 due to the uncertainty of TP inflows (i.e., past wastewater effluents causing internal TP release from the sediment years later, potentially invalidating steady-state assumptions). In this study, all time series data were included under the assumption that, mechanistically, random year effects in the statistical model would capture uncertainty in nutrient loads. Uncertainties also relate to TP analyses in the years from 2007-2017 (see more details Larsen et al., 2018, 2020).

This modeling is based on the TP mass balance in the modified form of:

Eq. 2: $$V\frac{dTP}{dt} = \alpha Q_{in}TP_{in} - Q_{out}TP_{out} - rTP_{out}V$$

where V is lake volume, α is an inflow multiplier accounting for settling of inflow TP_{in} in the littoral zone and/or near-shore wetland dynamics, Q_{in} is inflow discharge, Q_{out} is outflow discharge, TP_{out} is outflow TP concentration, and r is the retention coefficient (defined here as a loss term of TP). Note that the model does not include internal release of TP into the lake (Fig. 1). α is based on the concept of fast and slow settling of (often particulate) TP near a lake's inlet (Canfield and Bachmann 1981; Chapra 1982). Hence, fast settling TP is removed close to the lake inlet whereas slow settling TP from the inflow can be retained in a lake or flushed out.

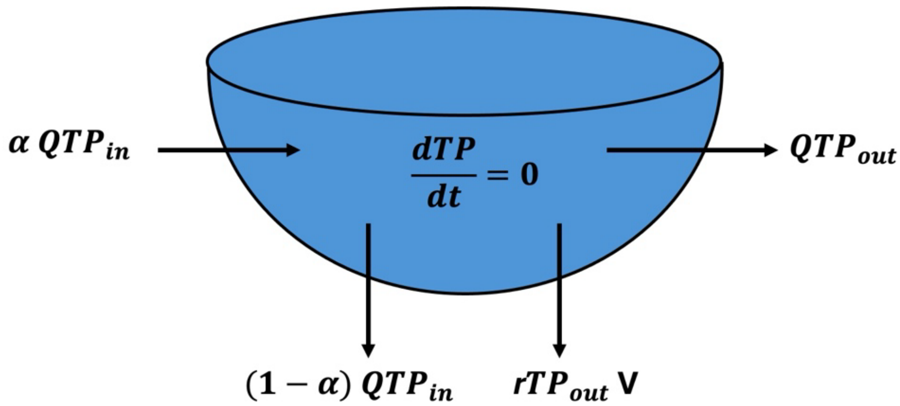


Figure 1. Conceptual sketch of the mass balance TP model. αQTP_{in} is the amount of (slow settling) TP entering the lake, QTP_{out} is the amount of TP leaving the lake, the lake can retain fast settling TP near the inlet $(1-\alpha)QTP_{in}$ as well as through retention, $rTP_{out}V$. It is assumed that no change in lake TP concentration over time occurs ($dTP/dt = 0$).

By assuming

- Steady-state conditions, $\frac{dTP}{dt} = 0$
- An annual balance of inflows to outflows, $Q_{in} = Q_{out}$
- In-lake TP, TP_{lake} or TP, equals TP_{out}

we can restructure Eq. 2 to calculate the projected outflow TP concentration $\overline{TP_{out}}$:

Eq. 3: $$\overline{TP_{out}} = \frac{(\alpha QTP_{in})}{Q+rV}$$

Equation 3 can be log-transformed[1] for more straightforward statistical modeling:

Eq. 4: $$\log\overline{TP_{out}} = \log(\alpha Q TP_{in}) - \log\left(Q + e^{\log r} e^{\log V}\right)$$

This transformation conditions the retention coefficient r to be always positive (hence, it acts as a sink of TP). Thus, retention as a process does not incorporate any release of TP from the ecosystem, e.g., internal TP loads.

The final model form includes random effects for lakes and years:

Eq. 5: $$\log\overline{\boldsymbol{TP_{out}}} = \log(\alpha \boldsymbol{Q TP_{in}}) - \log\left(\boldsymbol{Q} + \boldsymbol{e}^{\log \boldsymbol{r}} \boldsymbol{e}^{\log \boldsymbol{V}}\right) + \tau_{Lake} + \tau_{Year}$$

The model was run using the RTMB R-package (Kristensen et al. 2016; Thygesen et al. 2017). $\boldsymbol{TP_{out}}$ was estimated by fitting $\boldsymbol{\alpha}$ and r coefficients as well as random effects to measured $\boldsymbol{TP_{out}}$ values.

TP retention as a physical process can be estimated as:

Eq. 6: $$ret = rTP_{out}V$$

which quantifies the amount of TP retained by a lake. We also tested various alternative equations – such as introducing a second retention coefficient representing TP release from the sediment, and a model with sediment TP release acting as a function of lake area – which all proved to be less identifiable using the prescribed assumptions. In these cases, the statistical model failed to fit coefficients with a low uncertainty. We calculate a lake's gross TP retention as:

Eq. 7: $$ret_{gross} = rTP_{out}V + (1 - \alpha)QTP_{in}$$

assuming that the loss term near the inflow, $(1 - \alpha)\, QTP_{in}$, due to fast settling TP, also contributes to a lake's total retention of TP. This formulation makes the model more dynamic compared to the assumption that the lake is a completely mixed reactor. With α affecting inflow TP, the model accounts for spatial settling of especially particulate TP in littoral zones. This gross retention is used for the subsequent analysis.

We compared our estimates for lake TP retention against a simple first approximation of TP retention:

Eq. 8: $$ret_{pseudo} = \left(1 - \frac{TP_{out}}{TP_{in}}\right) TP_{in}\, Q_{in}$$

The first approximation of TP retention, ret_{pseudo}, approximates how much TP got retained in a lake using the observed data, hence it can be used to verify model performance. As it does not include any process descriptions except inflow and outflow dynamics of TP, its use for interpretating and upscaling TP retention processes is limited. Assuming that retained lake TP can be estimated from the ratio of outflow to inflow TP, substituting Eq. 3 into Eq. 8 yields the general steady-state mass balance equation:

Eq. 9: $$ret_{pseudo} = TP_{in}Q_{in} - TP_{out}\,Q_{out} = Q(TP_{in} - TP_{out})$$

[1] Note that all log transformations used in this report are natural logarithms.

We derive the TP fraction following Trolle et al. (2015) as:

Eq. 10: $$\lambda = \frac{\overline{TP_{out}}}{TP_{in}}$$

Note that Eq. 10 is conceptually similar to Eq. 1. By incorporating Eq. 3 into Eq. 10, we get:

Eq. 11: $$\lambda = \frac{\frac{(\alpha Q TP_{in})}{Q + rV}}{TP_{in}}$$

Further, when substituting $V = \tau_w Q$, we get:

Eq. 12: $$\lambda = \frac{\frac{(Q\alpha TP_{in})}{(Q + r\tau_w Q)}}{TP_{in}} = \frac{\frac{Q(\alpha TP_{in})}{Q(1 + r\tau_w)}}{TP_{in}} = \frac{\frac{\alpha TP_{in}}{1 + r\tau_w}}{TP_{in}}$$

Including, as a major difference to the model by Trolle et al. (2015), linear terms rather than non-linear dynamics such as in Eq. 1.

Model performance is evaluated using the root-mean squared error (RMSE), the Nash-Sutcliffe coefficient of efficiency (NSE) and the coefficient of determination, R^2. We applied the feature selection algorithm Boruta (Kursa and Rudnicki 2010) using random forest classifiers, to assess which covariates in a more extensive water quality and catchment dataset are most important to capture changes in gross retention, ret_{gross}. The extended dataset was derived from Ladwig et al. (2026). All data preprocessing as well as used variables are explained in detail in Ladwig et al. (2026). Derived ret_{gross} values were fitted using exponential models to the two most important covariates, lake area and lake outflow discharge (using all available data). Both covariates are related to the hydraulic residence time (t_w), with volume defined as area integrated over depth. However, by compressing volume and outflow into a single metric, information is lost. Therefore, by fitting three parameters (slope and two exponents, one for each covariate), a more informative simplified retention model can be constructed. To test the robustness of the exponential model fitting, a split-sample validation was run. The models were fitted on a training dataset and its performance through R^2 was evaluated on a separate testing dataset. This procedure was repeated multiple times with different splits between training and test datasets. Training to test ratios of 90-10, 70-30, 50-50, 30-70, 10-90 and 5-95 were evaluated over 50 iterations each.

6 Results and discussion

In this section, first the model itself – fitted coefficient values with uncertainty (*Model coefficients*) as well as model performance (*Model performance*) – are discussed. Afterwards, we discuss the model results:

- We quantify the TP fraction λ sensu Trolle et al. (2015) to highlight how the updated model compares with the former report (*TP fraction sensu Trolle et al. (2015)*) → this underscores how much TP is exported from the lake.
- We quantify the gross TP retention from the model to highlight lake-specific dynamics (*Modeled gross TP retention*), how gross TP retention relates to residence time (*Relationship of gross TP retention to residence time*), and to which other predictors the modeled gross TP retention is sensitive to (*Importance of predictors for gross TP retention*).
- Finally, based on the sensitivity analysis (*Importance of predictors for gross TP retention*), we develop exponential models of gross TP retention to discharge and lake area (*Exponential model of gross TP retention*) for upscaling.

6.1 Model coefficients

This subsection highlights the fitted parameters of the updated model (Eq. 5). Random effects due to lakes were greater than those associated with years, highlighting temporal similarities but lake-specific caveats (Table 2). After accounting for the proportions of explained variance (e.g., as $e^{Estimate^2}$), the lake random effect accounted for 59% of total variance (compared to 37% and 3% for the standard error of the observation estimate and year random effect, respectively). The retention coefficient, r, was estimated at 0.02 year^{-1}, ranging between 0.006 to 0.059 year^{-1}. α was estimated at 0.81, ranging from 0.74 to 0.87, suggesting that, for the majority of lakes, about 20% of inflowing TP is retained near the inlet, likely due to presence of macrophytes, fast settling or wetlands. These estimates of α are similar to previous TP retention modeling studies, as α often ranges from 0.55 to 0.70 (Khorasani and Zhu, 2021) as "[…] a significant proportion (30 – 45 %) of the TP loading into the lakes may be removed rapidly […]" (Khorasani and Zhu, 2021). Overall, standard errors for all estimates were relatively low.

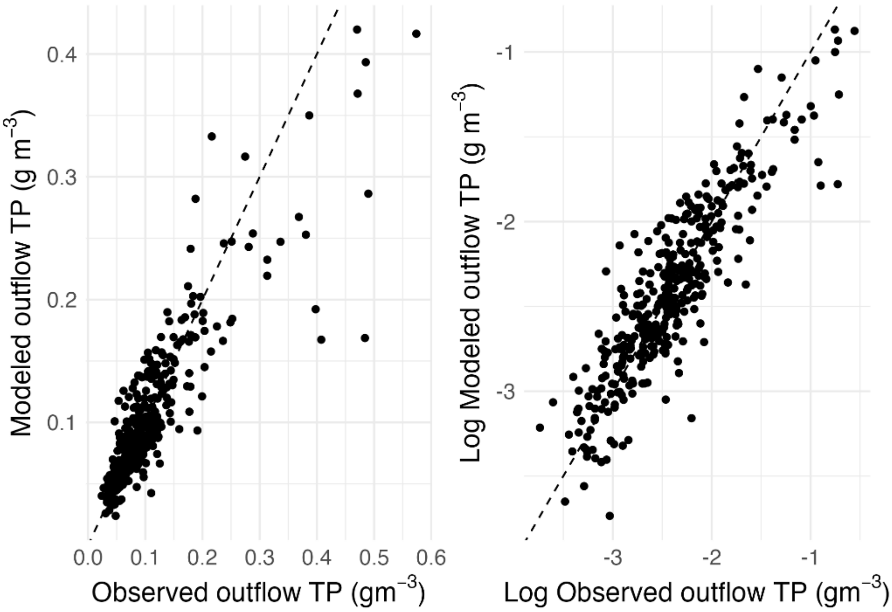
Table 2. Fitted model parameter estimates with standard errors.

Parameter	Estimate in log-space	Standard error
log Sd (standard deviation of random effect)	-1.301016	0.0400524
log τ_{Lake} (random lake effect)	-1.077266	0.1652361
log τ_{Year} (random year effect)	-2.595905	0.2808446
log r	-3.969478	1.084461
log α	-0.2097251	0.07902928

6.2 Model performance

This subsection demonstrates how well the model replicated observed TP concentrations. The projected $\overline{TP_{out}}$ obtained an RMSE of 0.038 mg TP L^{-1}, an NSE of 0.758 and an R^2 of 0.771 compared to TP_{out}, which indicate a well-performing model. However, the model underestimates high TP concentrations above 0.3 mg TP L^{-1} (Fig. 2).

Figure 2. Observed to modeled outflow TP concentration (24 lakes with in total 367 lake-years). Left are non-transformed values, right are log-transformed values. Lines represent a 1:1 linear regression.



6.3 TP fraction sensu Trolle et al. (2015)

Similar to Trolle et al. (2015), we quantified the fraction of TP exported from a lake, λ (Eq. 10), and found that most lakes have a λ below 1, indicating higher inflow TP concentrations than in-lake/outflow TP concentrations (Fig. 3). A λ above 1 would generally indicate that the lake acts as a source of TP, hence it releases more into the downstream environment than it receives from its inflow, suggesting that internal loads are important. Here, over time, λ slightly declines. As λ represents the ratio of outflow TP to inflow TP (again, assuming that outflow TP equals in-lake TP), this suggests that less TP from the inflow is accumulating in the water column, potentially due to increased settling. This can suggest improved water management or natural recovery with overall decreasing in-lake nutrient conditions as internal retention of TP is potentially increasing, hence a natural recovery of the system as it is able to bind incoming TP instead of releasing it. Correspondingly, the term (1- λ), that would quantify the fraction of TP retained in the lake, is increasing slightly over time.

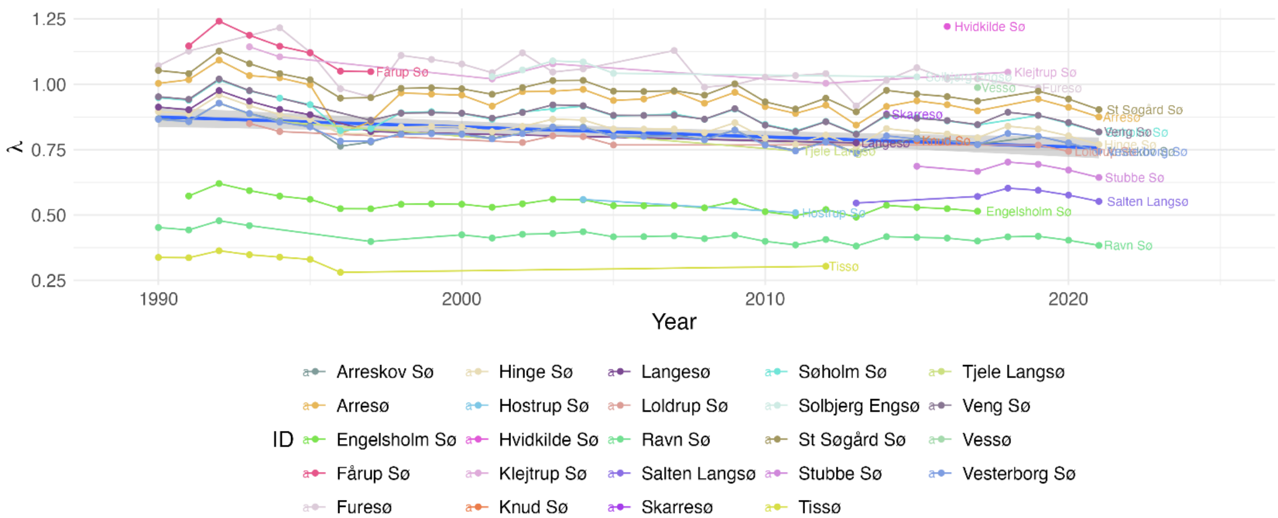


Figure 3. TP fraction exported from the lake, λ, over time for all 24 lakes. The blue line represents the average trend regression with confidence interval.

6.4 Modeled gross TP retention

This subsection discusses the projected gross TP retention rates derived from the fitted model for each lake over time. Model-derived estimates of gross TP retention, ret_{gross}, highlight lake-specific differences depending on size and contamination history (Fig. 4 A). The model projected most lakes to have a median gross TP retention of 1216 kg TP year^{-1} with a mean of 2338 kg TP year^{-1}. As reference, Søndergaard et al. (1993) quantified the TP retention for Lake Søbygaard at 3740 kg TP year^{-1}. In our data, Arresø, Tissø and Salten Langsø exhibit the highest gross retention, potentially overlapping with high external as well as internal TP loads. As gross TP retention scales with outflow TP concentration and lake discharge, high TP emerging from the lake is also increasing gross TP retention in the model. Note that the current model assumes only positive TP retention, meaning that the lake stores TP, on the modeled annual time scale. When normalised to area (Fig. 4 B), Veng Sø and Solbjerg Engsø had the highest gross areal retention.

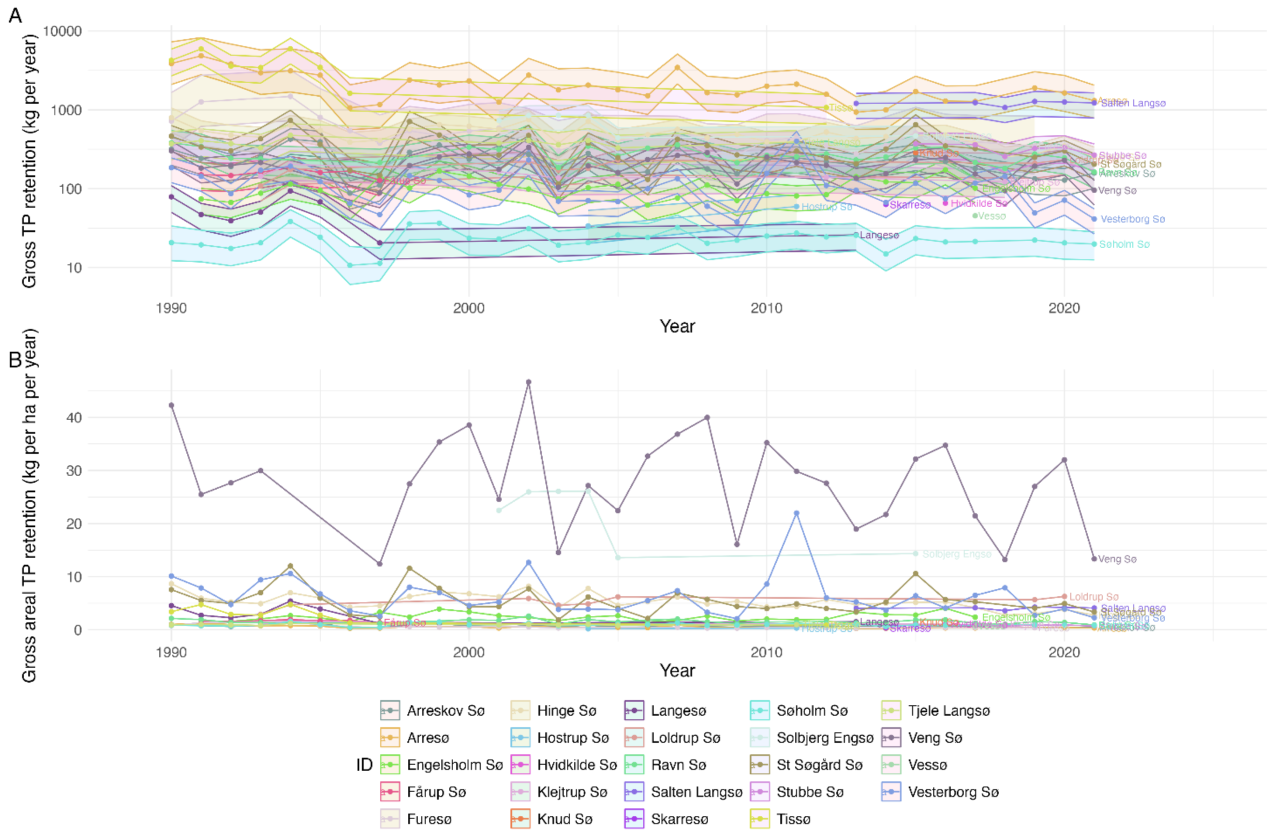


Figure 4. Modeled gross TP retention over time. Y-axis is log-scaled. Colored intervals represent the uncertainty around the mean projections. A is gross TP retention in kg per year over time. B is gross areal TP retention in kg per ha per year over time.

Comparing the model-derived gross TP retention, ret_{gross}, to an estimate of retention, ret_{pseudo}, underscores that the updated model overestimates the gross TP retention processes for some lake-years (Fig. 5), with an R^2 of 0.50. Plausible reasons for this are that model assumes a near-inflow settling of TP (which the pseudo estimate neglects), the steady-state assumption, and possibly due to negligence of internal TP release. This is in agreement with the model performance (Fig. 2) as the model underestimates high TP outflow concentrations, hence assumes higher gross TP retention.

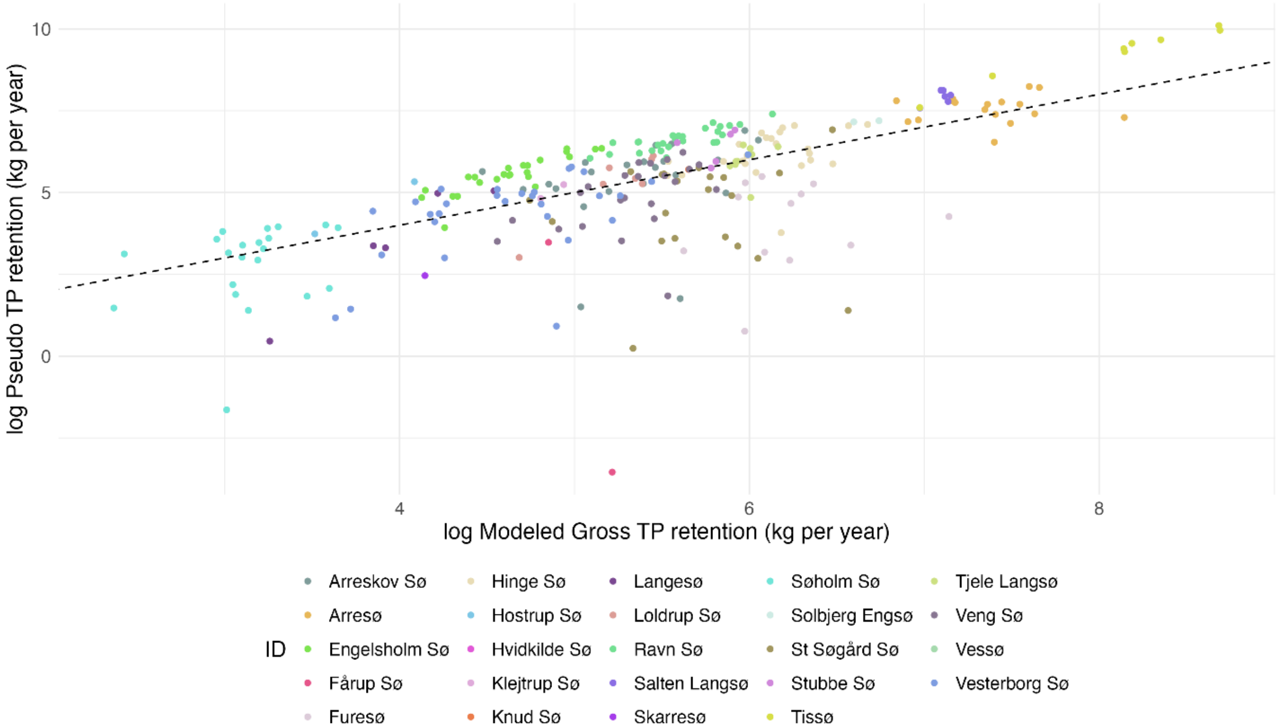


Figure 5. Log-transformed gross TP retention against pseudo-observed TP retention, ret_{gross}. Values are log-scaled.

6.5 Relationship of gross TP retention to residence time

In this subsection, we explore the relationship of gross TP retention to hydraulic residence time. In past reports, hydraulic residence time was identified as being a sensitive parameter to TP retention. In contrast to the results by Trolle et al. (2015), the model-derived gross TP retention did not exhibit a perfect linear relationship (when log-transformed) with hydraulic residence time (Fig. 6 A). This highlights that loss at the inflow is crucial for gross TP retention, which cannot be captured by using only hydraulic residence time. Still, larger lakes, and thus lakes with longer retention times, can retain more TP than smaller lakes (Fig. 6 A). Nonetheless, for individual lakes, this relationship seems to be inverse. Individual lake-years with lower discharge (hence higher residence times) exhibit a lower TP retention. Again, this highlights the modeled mechanism of inlet TP retention as with lower discharge, TP retention near the inlet also decreases. Accordingly, TP retention increases rapidly at low residence times and then plateaus at approx. 2.5 years (Fig. 6 B). When normalised by area (Fig. 6 C), log-transformed gross areal TP retention exhibits a linear relationship with residence time with a negative slope, underscoring that smaller lakes have higher areal retention than larger systems. Similarly, gross areal TP retention has an exponential relationship to residence time (Fig. 6 D). We conducted the follow-up analyses (Boruta analysis for predictor importance and exponential models) also for gross areal TP retention in the Supplementary material (Fig. S1 and S2).

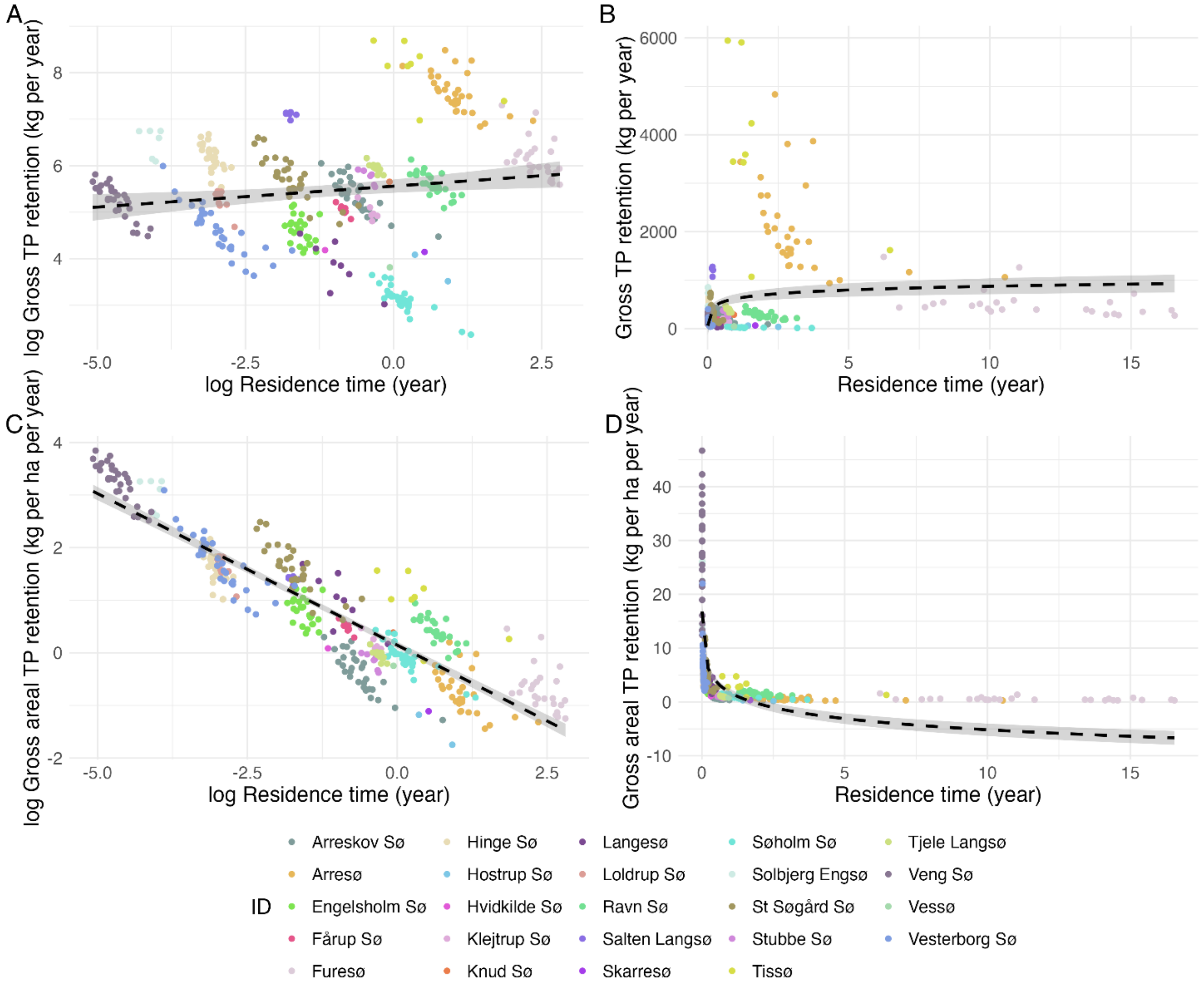


Figure 6. TP retention against residence time. A is log-transformed gross TP retention in kg per year against log-transformed residence time in years. B is gross TP retention in kg per year against residence time in years. C is log-transformed gross areal TP retention in kg per ha per year against log-transformed residence time. D is gross areal TP retention in kg per ha per year against residence time.

6.6 Importance of predictors for gross TP retention

To explore which covariates are the most important predictors for gross TP retention, a Boruta analysis was run (Fig. 7). The analysis showed that the most important predictors were discharge, Q, and area, A (for full explanation of all covariates, please see Ladwig et al., 2026), whereas residence time was not as important as expected. Following the principles of parsimony (using the hypothesis requiring the fewest assumptions), we developed exponential models relating gross retention to lake discharge and lake area for upscaling. Cheng et al. (2009) also highlighted that for shallow lakes (1.67 – 10.3 m depth), surface area was an important predictor.

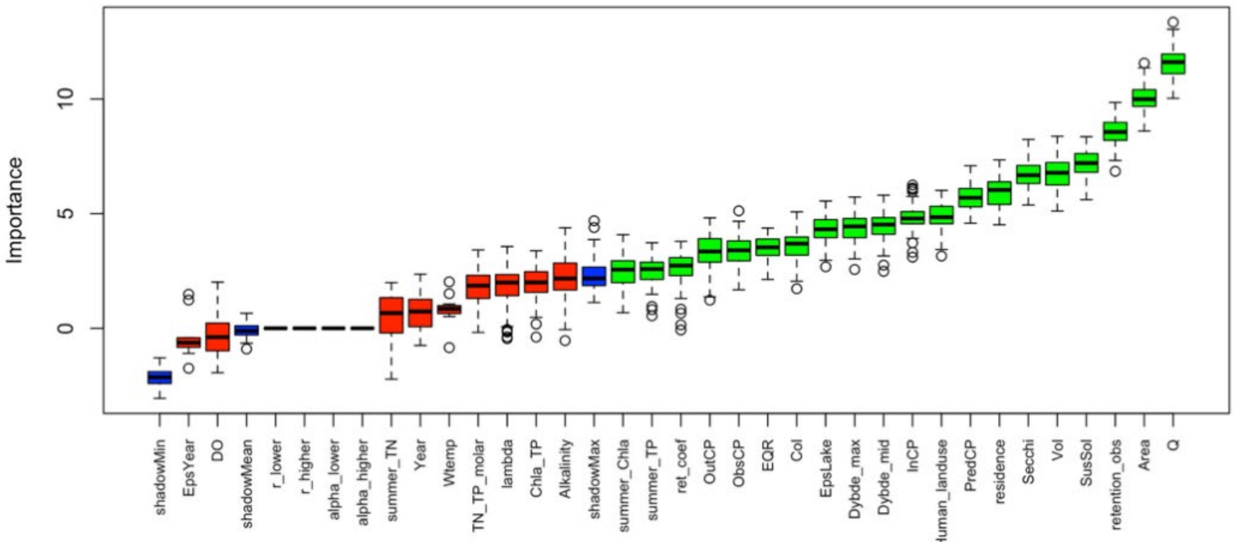


Figure 7. The importance of multiple water quality, lake characteristics, catchment and land use covariates on gross TP retention identified using Boruta analysis. Additional data are derived from the DCE report by Ladwig et al. (2026). In the boxplots, horizontal solid line indicates median of the distribution, the box represents the lower to upper quartile values of the data, the whiskers extend to the last data point beyond 1.5 * the interquartile range, circles represent outliers beyond this range.

6.7 Exponential models of gross TP retention

This subsection highlights the derived simpler models, which relate derived gross TP retention to available data such as discharge and area. Exponential models relating outflow discharge and lake area to gross TP retention were developed for all lakes, only shallow lakes and only deep lakes, respectively. These three models were further stratified by using either all data points (Fig. 8 left) or averaged lake responses (Fig. 8 right). The models all obtained an R^2 between 0.79 (shallow lakes) and 0.96 (deep lakes), and 0.9 when including all lakes. This is a substantial improvement compared to the previous report's R^2 of about 0.23. These equations may be cautiously used for upscaling to non-monitored lakes for which area and outflow discharge are known or can be derived. The exponents also highlight the more important role of discharge for shallow lakes and lake area for deeper lakes in controlling gross retention. Accordingly, shallow systems are mainly influenced by inflow hydraulics, whereas deeper lakes retain more TP due to their size and internal processes.

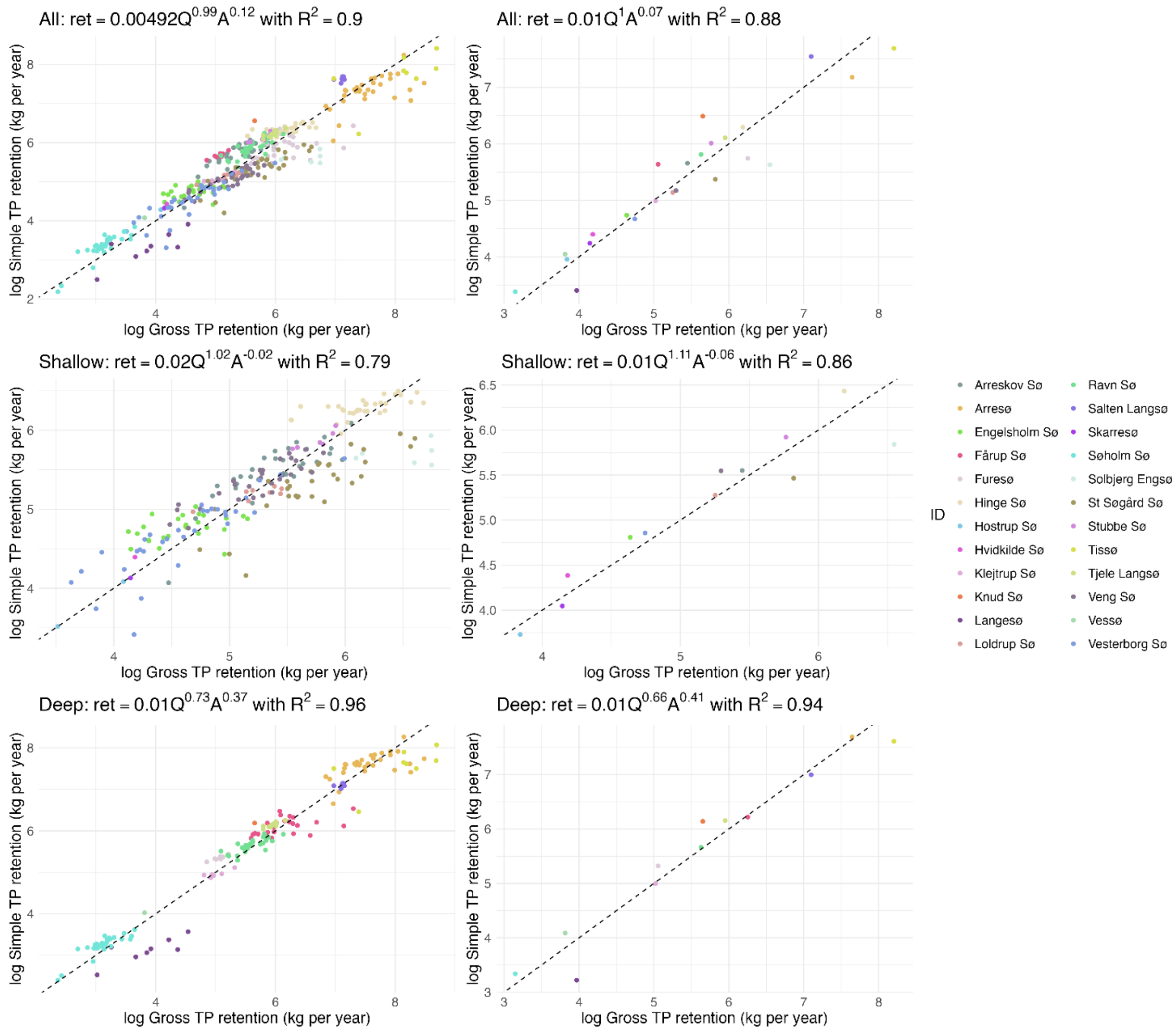
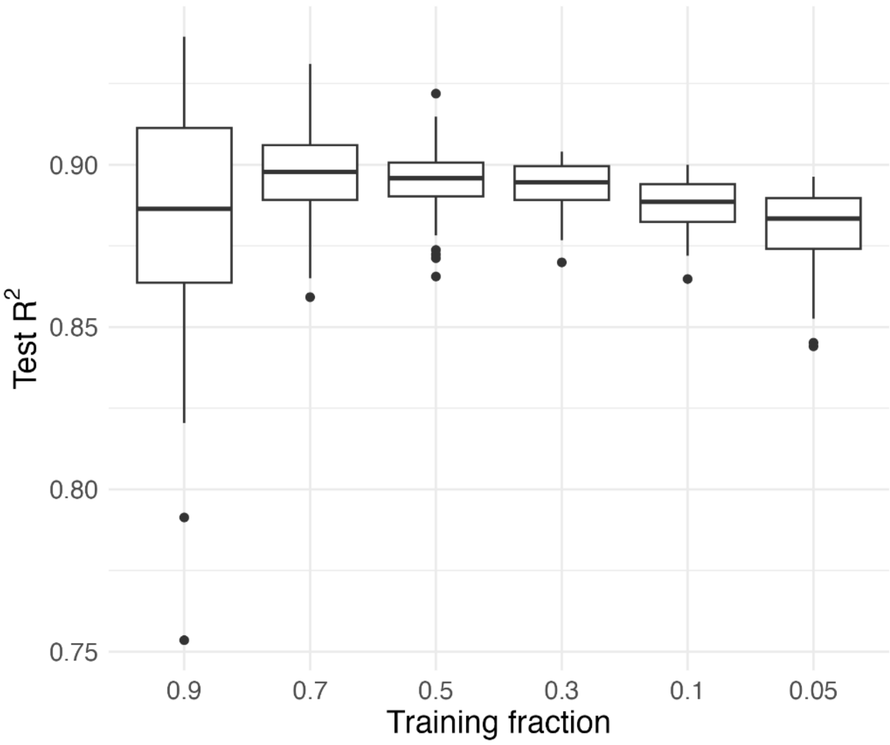


Figure 8. Log-transformed gross TP retention compared to predicted gross TP retention by an exponential model based on discharge and lake area, for all lakes, shallow lakes and deep lakes. The equation above the plots is the derived exponential model. Plots in the left column include all lake-years, plots in the right column show only lake-year averages.

This is in accordance with global studies (e.g., Kõiv et al., 2011) that lake volume, discharge and area are important predictors for TP retention with varying importance for shallow and deep systems.

To test the robustness of the exponential model fitting, a split-sample validation was run. The results show that exponential fitting is very robust with all training to testing ratios producing excellent model performance of R^2 over 0.85 (Fig. 9). Here, a training fraction of 0.9 (90-10 for training to test split) had the highest range of projected R^2 values (due to potential overfitting). Performance decreased with lower amount of training data but remained at a high level.

Figure 9. Performance (R^2) of exponential models relating gross TP retention to discharge and area at different training to test ratios using split-sample validation.



7 Conclusions

This report presents a modified modeling approach for calculating TP retention in lakes based on the general TP mass balance framework, which also formed the basis for previous reports. Compared to earlier models, the updated version shows an improved R^2 for both shallow and deep Danish lakes, and it is fitted using additional lake-years (Table 3). The simplified relationship between gross retention, area and outflow ($ret_{gross} \sim f(Q, A)$), can be cautiously applied to upscale gross TP retention in Danish lakes. Here, the Boruta method indicated that lake area (A), may be a better predictor of TP retention than lake volume (V). Exponential methods further validated this finding by showing higher exponents to the area term for deeper lakes compared to shallow ones. Nonetheless, we note that the exponential model using volume instead of area also achieved an R^2 of about 0.9, highlighting that both terms are related through depth, as $V = \int A\, dz$.

We note that, in contrast to gross TP retention, gross areal TP retention was highly sensitive to residence time (Fig. S1) and achieved slightly lower but still high R^2 of 0.52 (deep lakes) to 0.87 (shallow lakes) for exponential relationships of gross areal TP retention to residence time (Fig. S2). The analysis steps for gross areal TP retention were the same as for gross TP retention.

Table 3. Summary of the developed models for gross TP retention. TP is total phosphorus concentration (g TP m^{-3}) in inflow, in, or outflow, out. $\boldsymbol{\alpha}$ is an inflow multiplier (-). Q is outflow discharge (m^3 year^{-1}). V is lake volume (m^3), r is lake retention coefficient (year^{-1}), $\boldsymbol{ret_{gross}}$ is gross TP retention (g TP year^{-1}), A is lake area (m^2), and a, b and c are fitting coefficients

Model	Form	R^2
TP mass balance	$TP_{out} = \frac{(\alpha Q TP_{in})}{Q + rV}$ $ret_{gross} = rTP_{out}V + (1-\alpha)QTP_{in}$ with $r = 0.01888329$ and $\alpha = 0.8108071$	0.77
Exponential retention model (all lakes)	$ret_{gross} = aQ^bA^c$ with $a = 0.004918903$, $b = 0.994363471$ and $c = 0.1200764$	0.9
Exponential retention model (shallow lakes)	$ret_{gross} = aQ^bA^c$ with $a = 0.01907172$, $b = 1.022382$ and $c = -0.01609998$	0.79
	$ret_{gross} = aQ^bA^c$	0.96

Exponential retention model (deep lakes)	with $a = 0.008451743$, $b = 0.7321179$ and $c = 0.367592$

Current catchment models assume a TP retention in lakes of 4.5 kg ha^{-1} y^{-1} (Andersen et al. 2020). Using our dataset (all lake-years), we estimate a median area of about 111 ha (ranging from 7 to 3955 ha). Median lake outflow is about 10.7 million m^3 $year^{-1}$, ranging from 1.4 million to 80.5 million m^3 $year^{-1}$. Applying the exponential model for all lakes, we obtain a median areal retention of 2.31 kg TP ha^{-1} y^{-1} (ranging from 0.73 to 3.51 kg TP ha^{-1} y^{-1}). Hence, depending on the lakes, our derived model equation estimates lower areal TP retention values than the catchment estimate of 4.5 kg ha^{-1} y^{-1}. These values remain smaller than the estimated TP retention of 20 kg ha^{-1} y^{-1} in Lake Søbygaard (Søndergaard et al. 1993).

Trolle et al. (2015) developed a model formulation that related in-lake TP to inflow TP, which was replicated in this report through the TP fraction λ. In Trolle et al. (2015), this TP fraction was related to the hydraulic residence time to build a simpler model providing the fraction of TP leaving the lake. Hence, (1- λ) times the inflow TP concentration would provide the gross TP retention. The approach in this note directly uses gross TP retention as a function of discharge and area (both related to the hydraulic residence time), hence both formulations, λ and gross TP, are inherently related as:

Eq. 13: $ret_{gross} = (1 - \lambda)TP_{in}$

and

Eq. 14: $\lambda = 1 - \frac{ret_{gross}}{TP_{in}}$

All models abstract reality into a mathematical framework. Here, the model assumes steady-state conditions for the lakes and applies global values (a single value for all lakes in the dataset) for α and r. This implies that net inlet TP retention and in-lake TP retention are similar across lakes, only corrected by lake-specific lake discharge and volume, which is a very strong assumption. Especially the discharge is the most important covariate driving modeled gross TP retention as the settling of TP near the inlet takes up, for most lakes, over 80% of total gross retention. We also tested models including only the water column retention, hence without retention near the inlet, Eq. 6, but these models did vastly underestimate TP retention compared to the catchment estimate of 4.5 kg ha^{-1} y^{-1}. Note that the final fitted model underestimates high TP projections (Fig. 2) which can be an artefact of not accounting for internal TP release. Nonetheless, according to the model analysis of this study, near-inlet TP settling can be highlighted as an important part of TP retention in Danish lakes supporting the α-fraction hypothesis, hence that a potentially large pool of TP can rapidly settle near a lake's inlet (Khorasani and Zhu, 2021).

Though elegant and powerful in inferring potential gross TP retention, the current modeling approach is highly simplified. This is emphasized by its discrepancy with TP retention estimates used in catchment modeling, as mentioned above. The catchment estimate is calculated from annual mass balance of the most intensely monitored lakes. Future work should identify which estimates are more valid for TP retention Danish lakes. For this, it is recommended to develop dynamic and not steady-state TP retention models in future modeling work, incorporating past TP legacy in the sediment and the lake water column. This would improve the realism and accuracy of TP modeling for Danish lakes, especially within the context of the River Basin Management Plan and planned actions such as biomanipulation and/or lake restoration. Such a dynamic TP model, ideally operating on monthly time steps, would provide information on seasonal TP retention, TP release and TP legacy in lake sediments. It goes without saying that such a model would require substantial development time, but the foundations are already available, as dynamic modeling has already been applied to the dynamics of hazardous metals, an approach with obvious analogies to nutrient retention in lakes (Sørensen and Nielsen 2023).

8 References

Andersen, H.E., Heckrath, G. et al. (2010). Fosforkortlægning af dyrkningsjord og vandområder i Danmark. Aarhus Universitet, DCE – Nationalt Center for Miljø og Energi, Videnskabelig rapport nr. 397.

Canfield Jr., D.E. and Bachmann, R.W (1981). Prediction of total phosphorus concentrations, chlorophyll a, and Secchi Depths in natural and artificial lakes. Canadian Journal of Fisheries and Aquatic Sciences 38(4), 414-423.

Chapra, S.C. (1982). A budget model accounting for the positional availability of phosphorus in lakes. Water Research 16, 205-209.

Cheng, V., Arhonditsis, G.B. and Brett, M.T. (2009) A revaluation of lake-phosphorus loading models using a Bayesian hierarchical framework. Ecological Research 25. 59-76.

Khorasani, H. and Zhu, Z. (2021). Phosphorus retention in lakes: A critical reassessment of hypotheses and static models. Journal of Hydrology 603(126886). 1-15.

Kõiv, T., Nõges, T. and Laas, A. (2011). Phosphorus retention as a function of external loading, hydraulic turnover time, area and relative depth in 54 lakes and reservoirs. Hydrobiologia 660, 105-115

Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H. and Bell, B. M. (2016). TMB: Automatic differentiation and Laplace approximation. Journal of Statistical Software 70(5), 1-21.

Kursa, M.B. and Rudnicki W.R. (2010). Feature selection with the Boruta package. Journal of Statistical Software 36(11), 1-13.

Ladwig, R., Bucak, T., Johansson, L.S., Berthelsen, A. and Søndergaard, M. (2026). Nitrogen as a limiting nutrient in Danish lakes. Aarhus Universitet, DCE – Nationalt Center for Miljø og Energi, Videnskabelig rapport, in press.

Larsen, S.E., Tornbjerg, H., Søndergaard, M., Thodsen, H. and Blicher-Mathiesen, G. (2020). Forskelle i målt koncentration af totalkvælstof og totalfosfor i ferskvand ved at anvende de to oplukningsmetoder til organisk stof; autoklave- og UV-metode. Aarhus Universitet, DCE – Nationalt Center for Miljø og Energi, Videnskabelig rapport nr. 38.

Larsen, S.E., Windolf, J., Tornbjerg, H., Hoffmann, C.C., Søndergaard, M. and Blicher-Mathiesen, G. (2018). Genopretning af fejlbehæftede kvælstof- og fosforanalyser. Aarhus Universitet, DCE – Nationalt Center for Miljø og Energi, Videnskabelig rapport nr. 110.

Søndergaard, M., Nielsen, A., Levi, E.E. et al. (2020). Empiriske sømodeller for sammenhænge mellem indløbs- og søkoncentrationer af fosfor og kvælstof. Aarhus Universitet, DCE – Nationalt Center for Miljø og Energi, Videnskabelig rapport nr. 376.

Søndergaard, M., Kristensen, P. and Jeppesen, E. (1993). Eight years of internal phosphorus loading and changes in the sediment phosphorus profile of Lake Søbygaard, Denmark. Hydrobiologia 253, 345-356.

Sørensen, P.B., and Nielsen, A. (2023). Statistical model concept to quantify input and output of water, nitrogen and phosphorus for lakes with partly gauged watersheds. arXiv. https://arxiv.org/abs/2312.08120

Thygesen, U.H., Albertsen, C.M., Berg, C.W. et al. (2017). Validation of ecological state space models using the Laplace approximation. Environmental and Ecological Statistics 24, 317-339.

Trolle, D., Søndergaard, M. and R. Bjerring (2015). Sammenhænge mellem næringsstof-tilførsel og søkoncentrationer i danske søer. Aarhus Universitet, DCE – Nationalt Center for Miljø og Energi, Videnskabelig rapport nr. 138.

9 Supplement

The Boruta analysis for predictor sensitivity was ran also on gross areal TP retention (Fig. S1). This highlighted that residence time is the most sensitive covariate for gross areal TP retention. Accordingly, exponential models of residence time to gross areal TP retention were developed, which ranged from R^2 of 0.81 (all lakes) to 0.87 (shallow lakes) and 0.52 (deep lakes).

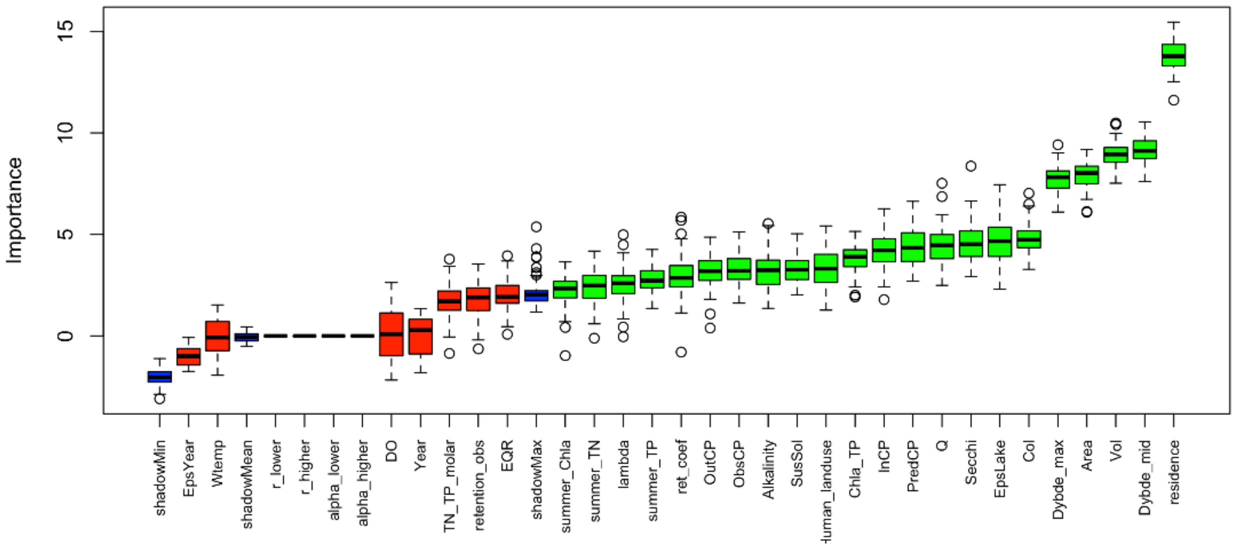


Figure S1. The importance of multiple water quality, lake characteristics, catchment and land use covariates on gross areal TP retention identified using Boruta. Additional data are derived from the DCE report by Ladwig et al. (2026). In the boxplots, horizontal solid line indicates median of the distribution, the box represents the lower to upper quartile values of the data, the whiskers extend to the last data point beyond 1.5 * the interquartile range, circles represent outliers beyond this range.

Figure S2. Log-transformed gross areal TP retention compared to predicted gross areal TP retention by an exponential model based on residence time, for all lakes, shallow lakes and deep lakes. The equation above the plots is the derived exponential model. Plots in the left column include all lake-years, plots in the right column show only lake-year averages.