

Authors' Response to Reviews of

Stochastic Modelling of Thermokarst Lakes: Size Distributions and Dynamic Regimes

Constanze Reinken

The Cryosphere, EGUSPHERE-2025-1817

RC: *Reviewer's Comment*, **CC:** Community Comment, **AR:** Authors' Response Manuscript Text

1. Anonymous Review 1

We again want to thank the reviewer for taking their time to read the manuscript and give such positive and constructive feedback. In the following, we list our responses to the reviewer's comments and detail any corresponding changes that we made to the manuscript.

1.1. Comment #1

RC: *The application of Geometric Brownian Motion (GBM) and Poisson processes to model lake area change and formation/drainage events is innovative and well-justified within the context of landscape-scale heterogeneity and stochasticity. This is a significant step beyond previous deterministic or analytical models.*

AR: We sincerely thank the reviewer for highlighting the novelty of our approach. We do not see any requirement for revisions based on this comment.

1.2. Comment #2

RC: *The three idealized regimes (Complete Drainage, Oscillation, Stabilization) effectively demonstrate the model's behavioral range and provide a useful conceptual framework for understanding possible long-term trajectories of thermokarst landscapes. The links to real-world examples for each regime are well-made.*

AR: We sincerely thank the reviewer for this positive feedback. We do not see any requirement for revisions based on this comment.

1.3. Comment #3

RC: *The current merging algorithm is identified correctly as a significant weakness. The assumption that merged lake area is the sum of the two original areas and that the new lake is perfectly circular is physically unrealistic and computationally expensive. It likely leads to a drastic overestimation of lake sizes and an underestimation of lake numbers.*

AR: We thank the reviewer for addressing this issue. As detailed in our responses to Reviewer 2, we investigated approaches to improve the merging algorithm and did not see a way to improve the merging representation without diverging from our general modelling concept and significantly adding to the complexity of the model and its degrees of freedom. Since our goal was to create a somewhat simple and computationally inexpensive model, we do not see a benefit in changing this and concluded that the development of a completely new

method would go beyond the scope of this study. As the reviewer mentions, we have already identified the current limitations in the manuscript and thoroughly discuss them.

While we did not change the way that two lakes merge, we did, however, introduce post-merge fission, i.e. the possibility of merged lakes to split again. The implementation of this new process makes the development of the lake size distribution and lake number as well as the effect of abrupt drainage on the system slightly more realistic, though effects on our simulations are small. For more information on this implementation, please refer to our response to Comment # 5 by Reviewer 2.

1.4. Comment #4

RC: *The high percentage of unusable data points (30-33%) in the remote sensing dataset severely impacts the robustness of the parameter estimation. This uncertainty propagates through the observation-based simulations and limits the confidence in the derived parameters ($\lambda_f, \lambda_d, \mu, \sigma$).*

AR: We thank the reviewer for highlighting this. We do not see any requirement for revisions based on this comment.

1.5. Comment # 5

RC: *A crucial result is that the stochastic component (σ) dominates the deterministic drift (μ) in the observed period. This implies that, for the 2000-2020 period in this region, random environmental variability was a stronger driver of annual lake area changes than any clear climate-driven trend, explaining the lack of strong correlations.*

AR: We thank the reviewer for highlighting this. We reviewed this finding after considering Comments # 7 and 8 by Reviewer 2, and did not find any new arguments against this conclusion. We do not see any requirement for revisions based on this comment.

1.6. Comment # 6

RC: *The inability to find significant correlations between model parameters and climate variables (TDD, P) is a notable negative result. While honestly reported, it highlights the current impossibility of confidently projecting lake dynamics under climate change scenarios with this model, as intended in the abstract. This is a major constraint on its immediate application in ESMs.*

AR: We agree that the mentioned limitation is a major constraint to the immediate application of our model into ESMs. Also in response to other comments, we revised the part of the "Discussion" that deals with the implications of our study for ESMs. Please refer to our response to community comment #1 by Dr. Elchin Jafarov to see the revised paragraph, which now highlights this issue even more.

1.7. Comment #7

RC: *The observation-based simulations project water area fractions increasing to over 50%, which is acknowledged as rare. This, combined with the high volatility, suggests the model parameters derived from 20 years of data may not be stable or representative of centennial-scale dynamics, potentially overestimating expansion.*

AR: We thank the reviewer for highlighting this take-away. We have added a sentence in the "Discussions" section to emphasize it further.

The simulated water areas in our observation-based experiments increase over the simulation period of 100 years to just over 50%, which is a relatively high value that is very rare in current Arctic landscapes. This further suggests that the model parameters derived from the 20-year observational dataset are likely not suitable for simulations on centennial scales, as they seem to overestimate expansion.

1.8. Comment #8

RC: *The suggestion that the idealized simulations could be interpreted as spanning 10 ka with a 10-year time step is helpful for context, but the parameters would then be "per decade." This should be stated explicitly in the text to avoid confusion (e.g., in Table 1, add a note: "Parameters are per year; for a 10-year time step interpretation, values would be per decade").*

AR: We agree that such a clarification would be helpful and added it to Table 2. Please note that we also adjusted the parameter values after making some changes to the model as part of the revisions.

Table 1: Hypothetical parameter values, fraction limit and model variant used for the three different idealized simulations. Parameters are per year; for a 10-year time step interpretation, values would be per decade.

Regime / Parameter	$\lambda_{f,v1/v2} [m^{-2}]$	$\lambda_{d,v1/v2} [m^{-2}]$	μ	σ	A_{lim}	variant
A. Complete Drainage	1e-9	2e-10	0.1	0.1	1	1
B. Oscillation	1e-9	2e-10	0.1	0.1	1	2
C. Stabilization	1e-9	2e-11	0.02	0.02	0.5	2

1.9. Comment #9

RC: *The comparison with the van Huissteden et al. (2011) model is good. The explanation for the differing results (their reliance on a prescribed river network vs. your data-driven drainage rates) is plausible and highlights the advantage of your approach, but also its current data dependency.*

AR: We thank the reviewer for this positive feedback. We do not see any requirement for revisions based on this comment.

1.10. Comment #10

RC: *The definition of "abrupt drainage" as a complete (>90% loss) and rapid event is clear. However, the discussion of results from other studies (Jones et al., 2011, 2020) that use different thresholds (e.g., >25% loss) is slightly confusing. A small table summarizing different study's definitions and converting their*

rates to a common framework would be helpful.

AR: We agree that a common framework for the comparison of drainage rates makes sense. To be in line with the threshold for lake drainage identification in the two mentioned observational studies (Jones et al. 2011 & Jones et al. 2020), we counted all lakes in our simulations that lost more than 25% of their area as a result of gradual drainage in addition to all abruptly drained lakes. We now mention this in the 'Results' section and have updated the corresponding part in the 'Discussion'. As argued in our initial replies, we do not see a need for a table, since both studies apply the same definition of 25% area loss constituting a drainage event.

Page 21, Lines 505 - 507 (in revised manuscript):

Ensemble runs with Variant 1 exhibited between 5 and 17 abrupt drainage events over the simulation period, while Variant 2 experienced 4 - 15. In addition, Variant 1 simulated between 5 and 18 lakes that lost more than 25% of their area until the end of the simulation through gradual drainage, whereas Variant 2 simulated between 4 and 20.

Page 26, Lines 607-618 (in revised manuscript):

The numbers of abrupt drainage events that we simulated using our model with observation-based parameterization are roughly in line with observed drainage events in Jones et al. (2020), where 98 drained lakes within a 30,000 km² area between 1955 and 2017 were found, which corresponds to an average drainage rate of 1.58 lakes per year. In this case, the definition of drainage events included all lakes that lost more than 25% of their surface area. In our model ensemble simulations we found between 4 and 17 abrupt drainage events for a 40x40 km cell over a period of 100 years and an additional 4 to 20 lakes having lost more than 25% of their area through gradual drainage. Extrapolating these numbers to a 30,000 km² this would be a drainage rate between ~~0.94 and 2.25~~ 1.5 and 6.98 lakes per year. ~~However, the definition of drainage events includes lakes that lost more than 25% of their surface area, whereas we only consider lakes to be subject to abrupt drainage when they drain completely.~~ While the data from Jones et al. (2020) and our parameterization data (Nitze and Nicholson, 2025) both cover a coastal plain, the drainage rates could still be specific to the different respective study regions in Alaska near Teshekpuk Lake and the Yana-Indigirka Lowlands. In Jones et al. (2011), the same definition of a more than 25% reduction in surface area was used to detect drainage events from high-resolution remotely sensed imagery from 1950/51, 1978 and 2006/07 for a 700 km² region on the Seward Peninsula, Alaska. An average drainage rate of 2.3 lakes per year was found, which is between ~~50 and 100~~ 14 and 66 times as high as the simulated ~~abrupt~~ drainage rates for our study region.

1.11. Comment #11

RC: *The axis titles of many figures are obscured, or the figures are not very clear, for example, in Figures 5, 6 and B1.*

AR: We sincerely thank the reviewer for pointing this out. We have revised the figures to conclude fully readable axis titles.

1.12. Comment #12

RC: *There are some spelling errors in the manuscript. For instance, the first reference should read "thermocarst lake" instead. The authors should carefully proofread the manuscript to avoid such errors.*

AR: We thank the reviewer for pointing this out. We proofread the manuscript and corrected any noticed spelling mistakes or other minor errors. All changes can be seen in the manuscript version with change tracking, except for corrections in the reference list, which can not be indicated due to technical reasons.

2. Anonymous Review 2

We carefully considered all recommendations for improvement of both the model and the analysis, but did not include all of them due to fundamental challenges with the implementation or because we feel that the changes would go beyond the scope of this study. As we argued in our initial responses, we believe that even without all of the changes, the paper still functions as a useful presentation of a novel modelling concept for one of the most difficult challenges in permafrost research and land surface modelling, even if this concept can still be improved upon as part of future research.

Any comments and suggestions that we have not implemented, we will gratefully keep in mind for our subsequent work on this topic. We want to express our sincerest appreciation to the reviewer for taking the time and effort to read the manuscript with such care and for giving such constructive, well-thought-out and helpful feedback.

In the following, we give point-by-point responses to the comments and explain any related changes that we made, or why we did not implement them.

2.1. Major Comment #1

RC: *There may be an issue with the GBM discretisation and Brownian scaling. Please implement the standard GBM increment for time step Δt . The text describes sampling the Normal with standard deviation equal to Δt (the time step). This should be $\sqrt{\Delta t}$. If steps remain annual, state $\Delta t = 1$ year explicitly and correct the noise scaling.*

AR: We sincerely thank the reviewer for pointing this out. As mentioned in our initial response, this was only a mistake in the text and not in our method. We have corrected the text accordingly:

Page 9, Line 245 (in revised manuscript):

We include the BM term $B(t)$ in our model using the *numpy* function *random.normal*, which randomly chooses a value from a normal distribution. In our case, we use time step ~~Δt~~ $\sqrt{\Delta t}$ of our simulation as the standard deviation for this distribution.

2.2. Major Comment # 2

RC: *Linear area-fraction scaling implies constant hazards per unit area and no explicit dependence on lake size, clustering, or covariates. A modest generalisation with state-dependent intensities $\lambda_f(\circ)$, $\lambda_d(\circ)$ (e.g. log-link GLM or Cox hazards) would capture simple nonlinear feedbacks while remaining identifiable.*

AR: We sincerely thank the reviewer for this suggestion and agree that the linear area-fraction scaling is a

simplification implying constant hazards that are unrelated to some potentially relevant factors. However, as we mentioned in our initial responses, we do not see a way to implement this without adding further degrees of freedom to the model and introducing another factor that requires parameterization, which is difficult to obtain considering the availability of long-term data products. We argue that our current linear scaling is a sufficient generalization at this point.

2.3. Major Comment # 3

RC: *Specify the boundary condition at zero area (absorbing vs. truncation) and whether very small areas are killed. This choice affects the conceptual split between gradual (GBM) and abrupt (Poisson) drainage under annual sampling.*

AR: We thank the reviewer for pointing out that this was not made clear enough in our initial version of the manuscript. We have added some additional explanation to the 'Method' section.

Page 9, Line 257-261 (in revised manuscript)

We treat gradual and abrupt drainage as two separate processes, since we assume them to have different triggers. We include gradual drainage in the GBM by allowing the drift parameter μ to be negative (eq. 7). This way, lake areas in the model can decrease from one year to another not only due to volatility but also due to changes in the sign of the drift. This type of drainage is often driven by slower processes such as evaporation, slow creation of sub-surface drainage channels as a result of seepage through the active layer, sedimentation or vegetation growth within the lake. Abrupt drainage, on the other hand, is triggered by more sudden events such as bank erosion or connection with an existing topographic drainage gradient, like a river or a neighbouring drained lake basin. In our case, we define abrupt drainage as a drainage event during which a lake drains completely in less than a year. ~~We use~~ Lakes that are subject to such an abrupt drainage event are removed from the pool of active lakes and will therefore stay drained throughout the rest of the simulation. Gradually drained lakes, on the other hand, are not removed and can still become subject to expansion or further gradual drainage. However, if an area has reached a value of zero due to gradual drainage, it will effectively stay zero for the rest of the simulation due to the nature of GBM (eq. 7). We model the number of abrupt drainage events per timestep with another Poisson process to simulate this with probability of k events [...]

2.4. Major Comment # 4

RC: *Beyond merging, lakes evolve independently. Introducing weak correlation (shared random environment / spatial frailties) or lightly correlated GBM shocks could capture hydrologic connectivity without materially increasing complexity.*

AR: We agree that incorporating weak correlations would not materially increase computational complexity of the model. However, as we mentioned in our initial responses, it would likely add another degree of freedom to the model. After further consideration, we still do not see a way to do this without adding more parameters that would need to be calibrated or parametrized using remote sensing data. We still believe, that it could be more beneficial to keep the current model version simple with as few parameters as possible, in order to make it more easily useful for conceptual and sensitivity studies, especially because the lack of robust parameterization of the model remains one of the largest limitations of our study. Since independent lake evolution is also a core assumption in the argumentation by Victorov et al. 2019, we think that it is a

justified first assumption. Testing weak correlation could be useful for future work on the model beyond this manuscript, however. We appreciate the reviewers input and will keep this in mind.

2.5. Major Comment # 5

RC: *The current merging rule can produce implausibly large single lakes and is computationally heavy. A stochastic, geometry-consistent alternative (continuum percolation / Boolean union-of-sets with polydisperse footprints), optionally with post-merge fission, would improve realism and cluster statistics.*

AR: As we also discuss in the manuscript, we strongly agree that the current merging algorithm limits the models applicability and the plausibility of simulation results. We also agree that an improvement of the merging algorithm should ideally cover three aspects: a stochastic approach to induce merging events in order to reduce computing time, geometry-consistency to avoid implausibly large single lakes, and post-merge fission that allows merged lakes to split again. However, we find that implementing a different method remains challenging. In the following, we discuss each of the mentioned aspects separately.

Stochastic approach:

Implementing a function that can calculate a merging probability at each time step instead of explicitly checking for overlaps between every two lakes, could indeed significantly reduce computing time and would make the spatial distribution of lakes unnecessary, further simplifying the model and saving memory.

We tested whether we can create an emulator function for merging probability by conducting several experiments, where we randomly distributed circles in a 2D plane with a defined cover fraction and number of circles (boolean disk models). For each combination of cover fraction and number of circles, we created 10 spatial distributions. While expanding lakes in our model can be assumed to be lognormally distributed following Victorov et al. (2019), the inclusion of formation and abrupt drainage and merging itself, can lead to different distributions. We therefore conducted the experiments for different distributions of circles. Two examples with lognormally distributed circles can be seen in Fig. 1.

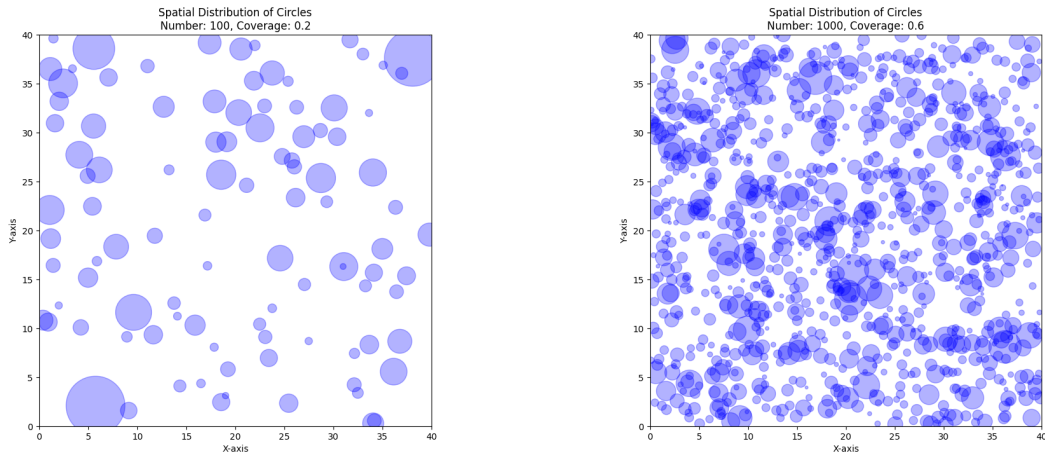


Figure 1: Example for randomly distributed circles in 2D plane with defined number and coverage. Circles sizes are lognormally distributed.

We then counted the overlaps of disks and calculated the mean across the 10 spatial model ensembles. Figure 2 shows the overlap numbers with different cover fractions and number of circles as an interpolated surface for each of the tested distributions. The number of overlaps, and therefore the merging probability, significantly differs depending on the size distribution of lakes at each time step. Creating a function that is simply depended on e.g. coverage and number of lakes would not be sufficient. Even including other statistical measures, such as mean and standard deviation of lake sizes, would theoretically not include all necessary information. Identifying the type of size distribution at each time step would be unfeasible.

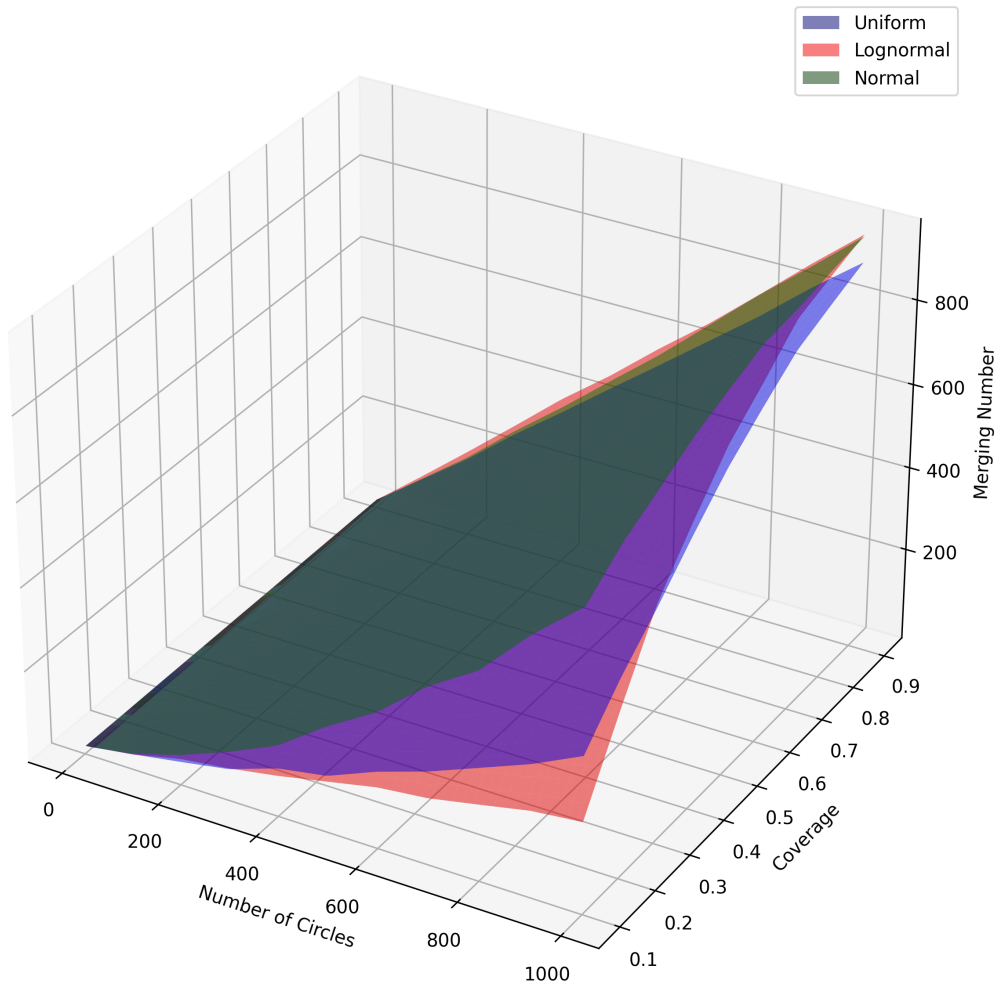


Figure 2: Interpolated surfaces for counted number of merges / overlaps for different cover fraction (coverage) and number of circles in three different boolean models with randomly distributed circles in a 2D plane. The random circles are assigned random radii taken from a uniform (purple), a lognormal (pink) and a normal (green) size distribution.

An additional challenge, is the fact that the merging probability in a dynamic system model does not only

depend on the immediate size distribution of lakes, but also on the merging events of the previous time step. Without including this effect, the merging probability would significantly overestimate merging events. However, quantifying this effect is difficult, and an implementation would require an expansion of the merging probability function to include at least one additional input variable.

Due to the mentioned challenges, we did not change the algorithm that determines the occurrence of merging events, and instead stayed with our initial approach of checking for overlaps of lakes explicitly.

Geometry:

We agree with the reviewer that a geometry-consistent approach would improve cluster statistics and make the model output more realistic. Currently, our merging algorithm simply calculates the sum of the areas of two merged lakes and essentially creates a new, perfectly round lake with it. In reality, merged lakes will neither be perfectly round nor are they guaranteed to have a surface area that equals the sum of the surface area of the previously separate lakes. A geometry-consistent approach could therefore consist of clustering merged lakes while keeping track of their individual geometries. However, after considering such an approach, we do not see a way to implement this within the framework of our model. An implementation would lead to the question of how lakes expand once they are in a cluster. Letting them expand separately would cause overlapping lake areas, while implementing a new lake expansion formulation for lake clusters would significantly complicate the model and require explicitly resolving spatial geometries, which would in principle violate one of the core goals of our study.

Post-merge fission:

We have implemented an algorithm that splits merged lakes once their area decreases to the sum of area that the initial lakes had when they merged. This change can make lake size distribution, lake numbers and the effect of abrupt drainage on the system slightly more realistic, but has relatively little effect on our simulation results. We have added a description in Section 2.1 'Model Framework'. Please see the text passage in the following.

Page 10, Line 289-294 (in revised manuscript):

If the area of a merged lake drains gradually, we include the possibility of post-merge fission, which allows merged lakes to split again. The model memorizes the areas of lakes at the time they merged. After drainage, it checks if any of the areas of merged lakes have reached a value between zero and the sum of the areas at time of merging. If that is the case, the lakes split and reassume their initial centre locations. The area is distributed among them according to their initial ratios. The lake that was incorporated into the bigger lake during the merging process and was removed from the pool of active lakes, gets reintroduced into that pool.

Besides the description of the newly implemented post-merge fission, we have not made any changes to the model and manuscript regarding the merging algorithm, as we believe that the limitations of the current implementation and challenges for different approaches are already sufficiently discussed.

2.6. Major Comment # 6

RC: *The birth–growth–death structure invites analysis:*

- *a size-structured Fokker–Planck (McKendrick–von Foerster + diffusion) for the area density $p(a,t)$ with integral birth/death terms;*

- *conditions for stationarity or self-similarity (lognormal-type tails) and closed-form moment dynamics;*
- *expectation dynamics for the total water fraction under an Alim ceiling.*

Even partial moment equations would substantively strengthen the mathematical core.

AR: We sincerely thank the reviewer for this input. We agree that further analysis in the form of partial moment equations or a Fokker-Planck equation would add to the mathematical merit of this study. We therefore included a partial differential equation for lake size distribution, which we derived by adding birth/death terms based on our assumptions on formation, drainage and merging to the Fokker-Planck equation for Geometric Brownian Motion (GBM). Since we do not see a way to directly incorporate this into the storyline of our manuscript, we have added it to the appendix. Please see the respective text passage in the following.

Page 32-33, Line 720-751 (in revised manuscript):

Appendix C: Partial Differential Equation for Lake Size Distribution

Lake expansion is modelled with GBM (eq. 7), which has the following corresponding Fokker-Planck equation (e.g. Stojkoski et al. 2020):

$$\frac{\partial}{\partial t} p(a, t) = -\mu \frac{\partial}{\partial a} a p(a, t) + \frac{\sigma^2}{2} \frac{\partial^2}{\partial a^2} a^2 p(a, t), \quad (\text{C1})$$

with $p(a, t)$ being the probability density for a lake having area a at time t , μ being the drift of the GBM and σ being the volatility. The formation and abrupt drainage processes in our model can be understood as birth and death terms and incorporated into the equation. The equation then gives the change of the number density of lakes $n(a, t)$ per unit area in a size bin around area a at time t rather than the probability density. Furthermore, the model includes a merging algorithm that influences $n(a, t)$. We expand eq. C1 to

$$\frac{\partial}{\partial t} n(a, t) = -\mu \frac{\partial}{\partial a} a n(a, t) + \frac{\sigma^2}{2} \frac{\partial^2}{\partial a^2} a^2 n(a, t) - D(a, t) + F(a, t) + M(a, t), \quad (\text{C2})$$

with $D(a, t)$ being the drainage term that gives the expected number of lakes of size bin around a removed due to abrupt drainage at time t , $F(a, t)$ being the formation term giving the number of lakes added through formation and $M(a, t)$ being the merging term giving the number of lakes removed or added through merging. Both abrupt drainage and formation are modelled with independent Poisson processes with effective rates $\Lambda_d(t)$ and $\Lambda_f(t)$. The rates can change over time due to climate dependence and because they are scaled by the water and/or drained area within the system at each timestep (see eq. 1 and 8). Abruptly drained lakes are chosen randomly according to the number of lakes obtained from the abrupt drainage Poisson process. Every size bin should therefore lose a number of lakes based on the fraction of removed lakes of the total expected number of lakes $N(t)$. The drainage term is therefore

$$D(a, t) = \frac{\Lambda_d}{N(t)} n(a, t). \quad (\text{C3})$$

Newly formed lakes all have the same initial area a_0 . The formation term can therefore be written as

$$F(a, t) = \Lambda_f \delta(a - a_0), \quad (\text{C4})$$

with $\delta(a - a_0)$ being the Dirac delta function (Dirac, 1927), which is 0 when $a \neq a_0$, making sure that newly formed lakes are only counted for the size bin around a_0 . When two lakes of areas a_x and a_y merge, the resulting area of the merged lake is $a_z = a_x + a_y$. If $R(a_x, a_y)$ is the rate at which two lakes of area a_x and a_y merge per timestep, then the expected number of lakes in the size bin around area a_x that is removed or added due to the merging is

$$M(a_z, t) = \frac{1}{2} \int_0^{a_z} R(a_y, a_z - a_y) n(a_y, t) n(a_z - a_y, t) da_y - n(a_z, t) \int_0^\infty R(a_z, a_y) n(a_y, t) da_y. \quad (\text{C5})$$

The first integral term $\int_0^{a_z} R(a_y, a_z - a_y) n(a_y, t) n(a_z - a_y, t) da_y$ describes the expected number of lakes that can merge and lead to a new summed area of a_z . The factor $\frac{1}{2}$ needs to be added to avoid double counting. The second term $n(a_z, t) \int_0^\infty R(a_z, a_y) n(a_y, t) da_y$ describes the expected number of lakes that had an area of a_z and were removed because they merged with a bigger lake. The probability or rate R of two lakes merging remains difficult to determine analytically, as it depends on the size distribution at that timestep as well as how many lakes have merged in the previous time step. The inclusion of post-merge fission, which is driven by area increases through the GBM, further complicates the changes of lake numbers.

2.7. Major Comment # 7

RC: *With annual data and gaps, the drift μ is weakly identified relative to the volatility σ . Consider a composite-likelihood or state-space formulation with uncertainty bands; Bayesian pooling (across neighbouring cells) can stabilise μ and the hazards.*

AR: We sincerely thank the reviewer for this insightful comment and recommendation. However, we currently do not see how the drift μ estimate can be stabilized further with the current datasets and without violating assumptions of our modelling approach. In our model, all lakes are assumed to follow the same GBM, with the volatility not only representing random annual fluctuations, but also variations between lakes. Considering this assumption, we believe that our estimates for μ and volatility σ obtained through our initial parameterization approach (Section 2.1.) are already maximum likelihood estimates. We also note that, before obtaining our estimates, we cleaned the dataset for obvious measurement errors as a result of data gaps. We don't see a way to distinguish remaining measurement error from natural volatility, even with e.g. Bayesian pooling. Furthermore, σ (which is essentially the noise) is itself a parameter in our model and linked to μ through GBM. Our parameterization method already contains a correction for the effect of σ on perceived μ , because they should not be viewed separately. Therefore, we are unsure whether the suggested methods would truly add to estimate robustness, and argue that our current approach of obtaining μ and σ together directly from the whole pool of data is sufficient for our purposes.

2.8. Major Comment # 8

RC: *Simple contemporaneous or one-year-lag correlations have low power and can miss nonlinearity. Try distributed-lag specifications, information-criteria-based lag selection, partial correlations (conditioning*

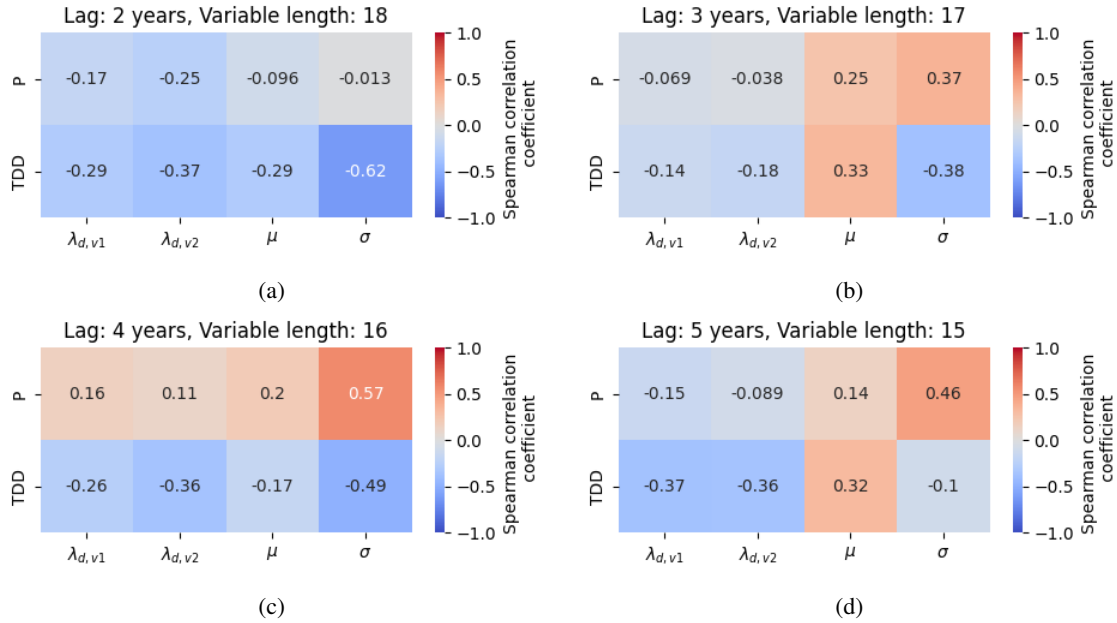


Figure 3: Spearman correlation coefficients between our parameters and climate variables TDD and P from the two, three, four and five years prior.

on antecedent water fraction), or spline thresholds before concluding there is no relationship.

AR: We thank the reviewer for this helpful comment. We understand that the current investigation of correlations between climate and our parameters is simplified and limited. We therefore considered and tested some of the mentioned methods.

Additionally to the lag of one year, we checked correlations for lags of up to five years. All correlation coefficients that we found were in a similar range as the ones we found with a lag of zero and one year. For instance, Fig. 3 shows the Spearman correlation coefficients for lags of two, three, four and five years. While it is reasonable that the system has some form of memory, there is no clear physical reason to believe that the climate of e.g. four years prior should influence lake dynamics directly. Therefore, we did not include this analysis in the manuscript. It is also important to note, that with larger lags, we reduce the length of the dataset and with it our confidence in the correlation analysis. We are not aware of any physical reasoning for distributed lags or clear criteria to base a lag selection on, which is why we did not implement this.

A visual assessment of data correlation for lags from zero to five years did not indicate that there are threshold to introduce regression splines, nor that there is a correlation between our parameters μ and σ and the antecedent water area fraction (i.e. water area fraction in the previous year) (Fig. 4).

We also tested partial correlations using the Python package *pingouin*. We controlled for either TDD , P or antecedent water area fraction. For parameters μ and σ , we found no indication of partial correlations, as all coefficients were below 0.6. High correlation could be found for λ_d and antecedent water area fraction, controlling for TDD of the previous year, as well as λ_d and P from the previous year, controlling for TDD (Fig. 5b). Figure 5 shows the residuals of the linear regressions that were analysed to obtain the partial correlation coefficients (Pearson). Figure 5a corresponds to the linear correlation between λ_d and antecedent

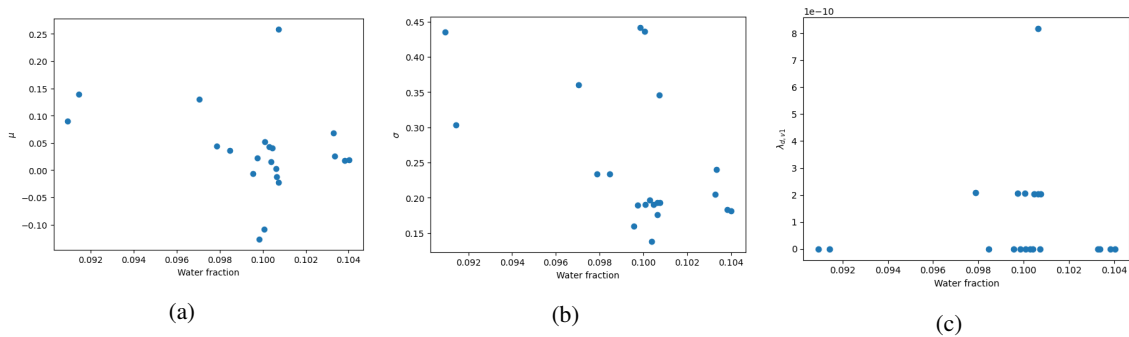


Figure 4: Parameters μ , σ and λ_d vs. antecedent water area fraction.

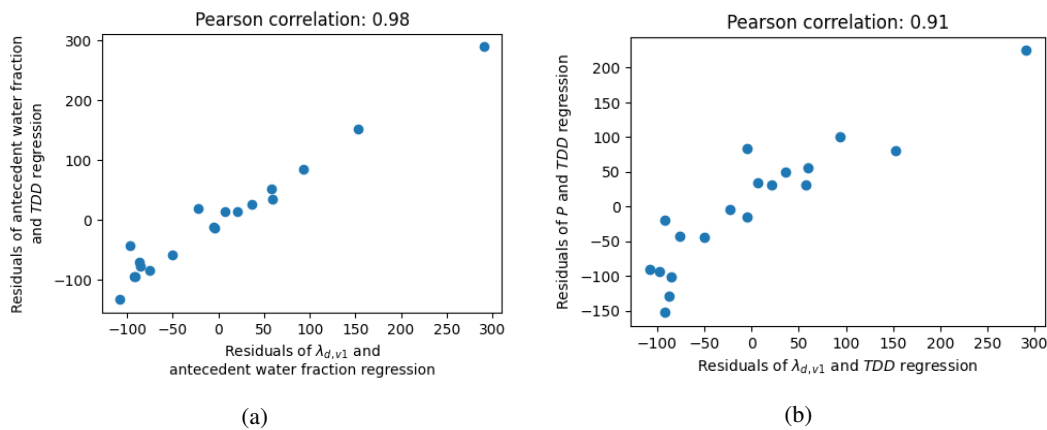


Figure 5: Partial correlations between λ_d and antecedent water area fraction controlling for TDD (a) and λ_d and P controlling for TDD with one year lag between climate variables and parameter. The partial correlations are measured by the linear correlation between the residuals of the linear regressions for λ_d with the goal variable and the residuals of the linear regressions for the goal variable and the control variable.

water area fraction, controlling for TDD . Figure 5b) corresponds to the linear correlation between λ_d and P controlling for TDD . While the coefficients are high (above 0.9), we believe that correlations between parameter λ_d and other variables should generally be handled with care due to the structure of the data. This parameter is simply a count of abrupt drainage events per year, scaled for the relevant area. Our timeseries of λ_d is based on several years with no lakes that drained under our definition of abrupt drainage, few years with one fully drained lake and one year with several lakes that drained. The one data point that is based on the year with several drained lakes can cause significant overestimations of correlations. Furthermore, partial correlations do not necessarily indicate a causal relationship, as far as we understand. They rather quantify remaining conditional linear association between two variables after conditioning on chosen control variables. We therefore think, that they have limited interpretability without a full picture of all relevant variables.

We have changed our wording regarding our results, to make it more clear that we do not conclude that there is no correlation, but rather that we cannot confidently say that there is one based on our analysis and due the size and structure of the observational data. And that we therefore do not have a basis to test a

climate-dependent model parameterization in this study. While an even more in depth look into e.g. partial correlations might make sense, especially with longer datasets, we think that it should include more variables and analysis tools, which would make it comprehensive enough to warrant a separate study.

Page 19, Line 468-469 (in revised manuscript):

~~No significant correlation was found between the climate variables and our parameter~~Based on our analysis and the available data, we cannot confidently determine a climate dependence of our parameters. Therefore, we do not have a robust basis for a climate dependent model parameterization.

2.9. Major Comment # 9

RC: *State the units of λ_f, λ_d (e.g. events per year per area) and specify how rates scale under aggregation/disaggregation so ESM tiles can ingest them consistently.*

AR: We thank the reviewer for this comment. In order to make the scaling of rates λ_f and λ_d more clear, we introduced the "effective rates" Λ_f and Λ_d that are already scaled and can be applied to the whole study area. Please see the revised text passages and equations below.

Page 6, Line 188-192 (in revised manuscripts):

Victorov et al. (2019) assume lake formation to be probabilistic and to occur independently on disjoint sites. This assumption of independence is in line with the fact that thermokarst happens due to a localized thawing process that depends on soil conditions at that particular point that are not directly influenced by thermokarst at a different point. When additionally assuming that the formation of the initial thermokarst depressions took place during a relatively short time or approximately at the same time, it can be shown that the number of depressions in an area obeys the Poisson law, meaning that the probability of k thermokarst depressions appearing during one year in A_f , which is the area available for formation, is

$$P_f(k, A_f t) = \frac{(\lambda_f A_f)^k}{k!} \frac{\Lambda_f(t)^k}{k!} e^{-\lambda_f A_f - \Lambda_f(t)}, \quad (1)$$

where λ_f is the formation rate, which $\Lambda_f(t)$ is the effective formation rate for the whole study area A . It is calculated as $\Lambda_f(t) = \lambda_f A_f(t)$, where $A_f(t)$ is the water area in the system at time t . The rate λ_f therefore corresponds to the average number of depressions per unit A_f and year $A_f(t)$ and unit t . An analysis of observational data from 16 sites showed strong agreement between the distribution of lake numbers and the Poisson distribution (Victorov et al., 2019).

Page 9, Line 261-265 (in revised manuscript):

We model the number of abrupt drainage events per timestep t with another Poisson process

$$P_d(k, A_d t) = \frac{(\lambda_d A_d)^k}{k!} \frac{\Lambda_d(t)^k}{k!} e^{-\lambda_d A_d - \Lambda_d(t)}, \quad (8)$$

with λ_d being the $\Lambda_d(t)$ being the effective abrupt drainage rate and A_d the area, by which λ_d is scaled. A_d for the whole study area A in year t , which is calculated as $\Lambda(t) = \lambda_d A_d(t)$. Here, $A_d(t)$ can either consist of only surface water across the region or the sum of water and drained area. With the this scaling, we make sure that the drainage rate is only above zero when lakes are present in the landscape, and that the drainage rate increases the more populated with lakes or depressions the area becomes. Since it is not clear, which of the two options for $A_d A_d(t)$ is more realistic, we implement two different variants that also concern the scaling of formation rate λ_f .

2.10. Major Comment # 10

RC: *Abrupt drainage = (near) complete loss within a year via the Poisson process; gradual drainage = negative GBM drift. With annual sampling, large partial losses (e.g. 60–90%) can be ambiguous; a competing-risks view with size-dependent abrupt-drainage hazard would help.*

AR: We thank the reviewer for this suggestion. As mentioned in our initial replies, we interpret this comment to refer to the analysis of observational data rather than model development or analysis of model results, since the model itself can keep track of whether a lake was abruptly or gradually drained. While we think that determining a size-dependent abrupt-drainage hazard from observational data is worth looking into, we also think that it is not a trivial task. Generally, further investigation of what information on lake dynamics could be obtained from the observational data would go beyond the scope of this manuscript and should rather be part of a separate study. We therefore added mention of this topic in the 'Discussion', but did not include it in our parameterization attempt. Please see the revised text passage in the following.

Page 27, Line 668-671 (in revised manuscript):

While the dependence on climate variables might become quantifiable for parameter μ and σ in a few decades with continued retrieval of remote sensing data, the same is likely not true for λ_f and λ_d . A robust estimation of these two parameters needs a number of observed formation and abrupt drainage events that satellite data will not be able to provide in the near future. In addition to this issue, our definitions of abrupt and gradual drainage make it difficult to differentiate between the two mechanisms in remotely sensed observational data. Large partial losses of lake areas can be interpreted as either process without additional analysis. Deriving a size-dependent hazard for abrupt drainage could be helpful, but also requires a large amount of data. It might therefore be necessary to think about calibration approaches beyond remotely sensed lake area timeseries.

2.11. Major Comment # 11

RC: *Provide a short Δt sensitivity (e.g. semiannual with rescaled noise) to test robustness of annual stepping.*

AR: We agree that an investigation of Δt sensitivity makes sense and thank the reviewer for this feedback. We provided a short Δt sensitivity study in the Appendix, which shows that our model's calculation of water and drained area fractions is largely time-step independent. Please see the respective text passage in the following.

Page 33, Line 752-763 (in revised manuscript):

Appendix D: Time Step Sensitivity

We conducted a time step sensitivity study to investigate the robustness of the default annual

stepping that we used in our simulations. To do this, we simulated ensembles of 10 members with Variant 1 of our model and time steps of $\Delta t_1 = 0.5$, $\Delta t_2 = 1$ and $\Delta t_3 = 2$. We used the parameterization seen in Table 2. No significant changes in model behaviour could be found. Figure 6a shows the ensemble means of water and drained fraction for the three different experiments, which all follow a somewhat similar trajectory. Any small-scale differences can be attributed to the stochasticity of the processes. Slight differences can be seen in the rate of lake number losses in Fig. 6b, with the simulations using a smaller timestep losing lakes more quickly. This can be explained by the implementation of the merging algorithm, which checks for overlapping lakes at every time step and therefore finds these overlaps earlier when the timestep is smaller.

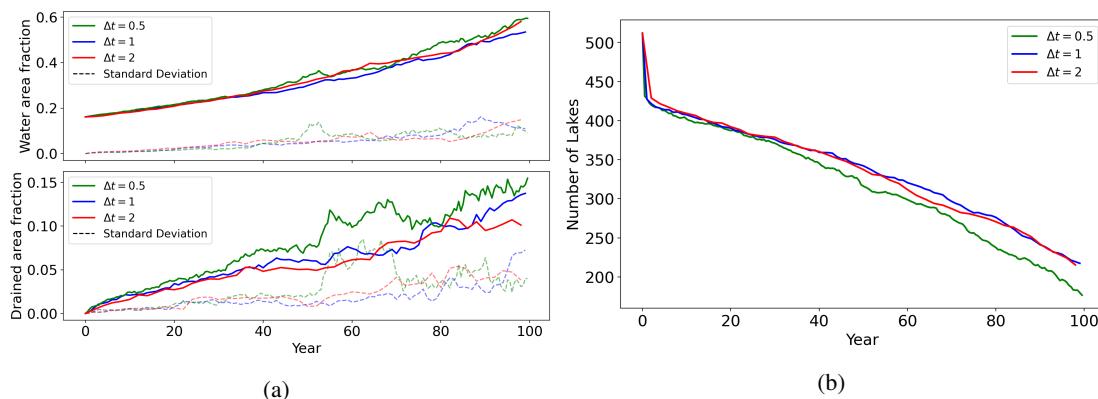


Figure 6: Model output for an ensemble simulation with Variant 1 and using the parameter values from Table 3 and time steps $\Delta t = 0.5, 1$ and 2 years: Time series of the ensemble means and standard deviation for water and drained area fraction (a) and the ensemble mean for lake number (b).

While the implementation of the merging algorithm has a slight time-step dependence when lake density is high, our model is time-step independent in its calculations of water and drained area fractions. It delivers similar results even with a semi-annual timestep. We want to note, however, that our model does not contain explicit seasonality and that we therefore do not recommend using it to investigate dynamics with temporal scales of less than a year.

2.12. Major Comment # 12

RC: *Complement the qualitative regimes with simple statistics (cluster count, Gini coefficient of lake areas, distribution quantiles) that distinguish regimes numerically.*

AR: We thank the reviewer for this suggestion. We focused on the Gini coefficient and calculated it over time for each Regime. We have added this figure to the Appendix together with a short discussion. While we agree with the reviewer that such statistical measures to distinguish the regimes are generally useful, we would like to note that the size distribution is strongly shaped by the merging algorithm that tends to lead to decreasing lake numbers and remains one of the model's biggest limitations. The mentioned measures need to be interpreted with that in mind and have somewhat limited informative value beyond that.

Page 34, Line 764-75 (in revised manuscript):

he Gini coefficient (e.g. Cowell and Flachaire, 2015) is a measure of inequality in a statistical distribution. A value of one indicates large inequality. For our model simulations, this would mean that there is only one active lake that contains all of the current water area of the system. A value of zero indicates perfect equality, where the water area is evenly distributed over several lakes. Figure ?? shows the Gini coefficients for the idealized simulations and regimes. Since Regime A and B were done with the same parameterization, they initially show a similar increase of the Gini coefficient while first lakes form and stochastic expansion and merging leads to increasing variety of lake sizes. Regime A quickly runs into a complete drainage of the landscape with no remaining active lakes. In Regime B, however, new lakes keep forming, leading to a changing Gini coefficient. Sudden drops of the Gini coefficient can be seen, where large lakes that have emerged through expansion and merging, drain and leave smaller lakes with a more equal size distribution. Regime C has slower expansion of lakes, which is why the Gini coefficient does not increase as steeply. In this regime, the model runs into a quasi-stabilization of lake areas, which is represented here by the Gini coefficient that oscillates around a value of 0.62 for roughly the last 670 years of the simulation.

Appendix E: Gini Coefficients for Idealized Simulations

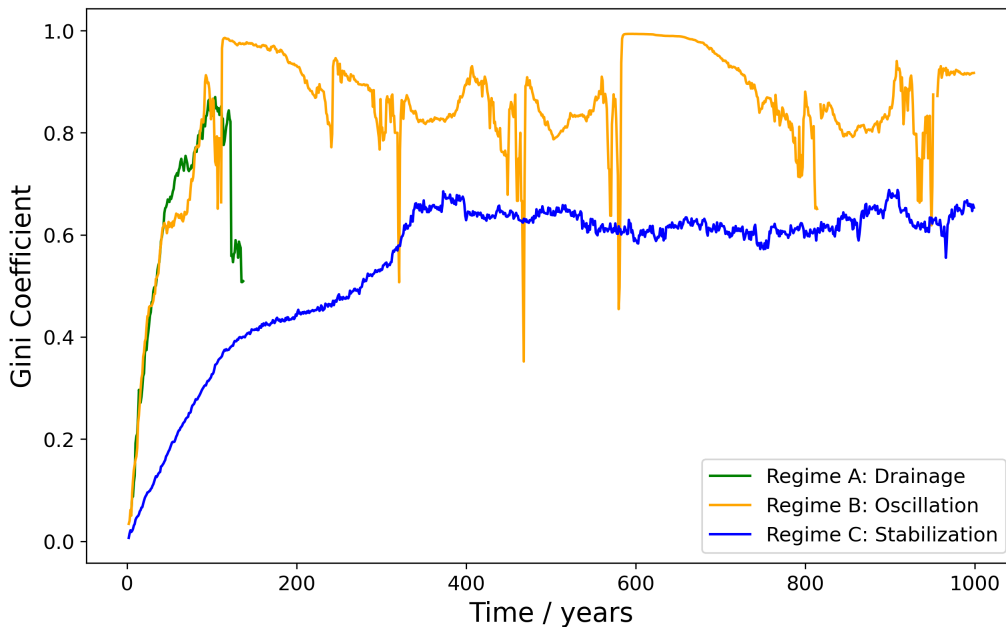


Figure 7: Gini coefficients of the three idealized simulations (Regime A: Drainage, Regime B: Oscillation, Regime C: Stabilization) as described in Sect. 3.1.

2.13. Minor Comment # 1

RC: *Bring the Variant 1/2 definitions forward and summarise implications in a small table. Clarify whether A_{lim} is a hard cap (projection) or a soft ceiling (e.g. logistic drift modulation).*

AR: We thank the reviewer for pointing out that the differences between the two model variants are not highlighted enough. We added the suggested table and believe that this change puts more emphasis on the definitions of the two variants.

Page 10-11, Line 301-302 & Table 1 (in revised manuscript):

We introduce an area fraction limit A_{lim} , which can stop lakes from growing any further. This limit represents the fraction of the study area that can theoretically be populated by lakes. When it is reached, no expansion of existing surface areas is allowed, but lakes can still shrink at that point. Expansion is prohibited until the area fraction is below the limit again. Formation, however, is not explicitly inhibited. Instead, the formation probability will naturally be zero at A_{lim} due to the implementation of the area-fraction scaling. In Variant 1 this includes drained lakes, meaning that the ratio of $A_{d,v1}(t)$ (eq. 9) and the study area needs to reach the limit for lakes to stop growing. In Variant 2, only the water area fraction influences lake dynamics. It is therefore the ratio of $A_{d,v2}(t)$ (eq. 10) to the study area, that needs to reach the limit. [Table 1 shows an overview of the main differences between the two model variants.](#) Values for the limit should be determined separately for different landscapes in the Arctic, considering the average ground ice content, topography and other landscape conditions. This task, however, is not trivial. For the purposes of this work, we mainly use a fraction limit of $A_{lim} = 1$ as default, essentially assuming that 100% of any simulated area can become covered with surface water.

[Table 2: The two model variants and their respective definitions for the areas used for scaling of formation and abrupt drainage rates \$\lambda_f\$ and \$\lambda_d\$ according to eq. 1 and 8, and the underlying assumptions, or rather implicit effects, on the formation and abrupt drainage possibilities.](#)

Variant	Relevant Area for A_{lim} and Scaling	Formation	Abrupt Drainage
1	$A_{f,v1}(t), A_{d,v1}(t)$ (eq. 9)	no new formation on drained area	region-wide probability increases with both drained and water area increase
2	$A_{f,v2}(t), A_{d,v2}(t)$ (eq. 10)	formation on drained area possible	region-wide probability increases with water area increase but independent of drained area

Furthermore, we have added some explanatory sentences answering the raised question on whether A_{lim} is a hard cap or not. Please see the revised text passage in the following.

Page 10, Line 296-299 (in revised manuscript):

We introduce an area fraction limit A_{lim} , which can stop lakes from growing any further. This limit represents the fraction of the study area that can theoretically be populated by lakes. [When it is reached, no expansion of existing surface areas is allowed, but lakes can still shrink at that point. Expansion is prohibited until the area fraction is below the limit again. Formation, however, is not explicitly inhibited. Instead, the formation probability will naturally be zero at \$A_{lim}\$ due to the implementation of the area-fraction scaling.](#)

2.14. Minor Comment # 2

RC: *Clearly state the initial area assigned to newly formed lakes.*

AR: We thank the reviewer for this comment and agree that the mention of the initial lake areas can easily be missed. To address this, we make sure to refer to the corresponding variable a_0 when mentioning the initial area, and added it to the list of the symbols. We also changed the area from 1 km^2 in our initial simulations to 1 ha to be in line with the resolution of the observational data used for parameterization.

2.15. Minor Comment # 3

RC: *Collect symbols/units in one table (including whether μ, σ are per-year) to aid reproducibility.*

AR: We agree that the addition of a table of symbols can make the manuscript more easily understandable and aids reproducibility. We added one in the appendix.

Page 30 (in revised manuscript):

Appendix A: List of Symbols

P_f, P_k	probability for formation or drainage events	
k	number of event occurrences	
λ_f, λ_d	formation and drainage rate	$\text{yr}^{-1}\text{m}^{-2}$
Λ_f, Λ_d	effective formation and drainage rate	yr^{-1}
A	area of study region	m^2
A_f, A_d	area relevant for formation and drainage	m^2
X	random variable	
t	time	yr
r	lake radius	m
j	year	
T	average temperature of water	$^{\circ}\text{C}$
c	specific heat	$\text{J kg}^{-1}\text{K}^{-1}$
α	share of heat in lake water leaving through lake sides	
ϵ_j^0	random episodic lake area change factors	
v	initial lake radius	m
f	stochastic transition function	
a	lake surface area	m^2
μ	drift	yr^{-1}
σ	volatility	yr^{-1}
B	standard Brownian Motion	

A_{lim}	area fraction limit	
R_{log}	logarithmic returns	
N	sample size	
P	total annual precipitation	mm
TDD	thawing degree days	°D

3. Community Comments by Elchin Jafarov

We sincerely thank Dr. Jafarov for taking the time to read our manuscript and give such constructive feedback. In the following, we give point-by-point replies to the comments and refer to the corresponding changes that we made to the manuscript.

3.1. Comment # 1

CC: *The authors briefly mention the challenges associated with incorporating phase change processes into their model. Expanding on this point would improve the manuscript, particularly by discussing how this stochastic model could be integrated into ESMs. For example, assuming a grid cell resolution of 0.5° or coarser, would the model estimate the ratio of land to water and apply distinct terrestrial and aquatic parameterizations based on that ratio? Providing more technical insight into how the model could be coupled with ESMs, specifically regarding the integration of thermal and hydrological processes, would be valuable.*

AR: We appreciate Dr. Jafarov highlighting the lack of technical details on ESM implementation in our manuscript. We have added a few more words on possible approaches and limitations. Please see the revised text passage in the following.

Page 21-24, Line 521-548 (in revised manuscript)

~~The model could provide ESMs with ESMs with~~ In principle, the model can provide information on water and drained area fractions in thermokarst-affected areas in response to a climate forcing. With this information, modelling of land-atmosphere fluxes of carbon, energy and water from permafrost-shaped landscapes could potentially become more accurate. Especially, estimates of net carbon emissions and the ratio between emitted methane and carbon dioxide could be improved, since new thermokarst lakes can be a significant regional methane source (e.g. Walter et al., 2007; Turetsky et al., 2020). Besides water and drained area fractions, our model can also provide the size distributions of lakes, which gives information on lake shoreline lengths and presumable lake depths. Furthermore, our model can keep track of the age of single lakes since the start of a simulation, which gives an indication of potential vegetation cover along the shoreline. These attributes have been shown to influence microbial decomposition processes and carbon emission pathways in waterbodies (Kutzbach et al., 2004; Juutinen et al., 2009; Knoblauch et al., 2015; Polishchuk et al., 2018; Rehder et al., 2021, 2023).

Some ESMs could use our model output to set the area of surface water in grid cells that have been previously identified to have thermokarst potential. For instance, JSBACH/ICON-Land (Re-

ick et al., 2021), ~~for instance, which is~~ the land component of the Max Planck Institute's Earth System Model ICON, ~~divides or JULES (Smith et al., 2022), the land component of the UK Earth System Model, divide~~ their grid cells into subgrid-scale tiles that are only defined by their fraction within the grid box. ~~It could use our model output to set the size of lake water tiles. Deriving~~ and include a representation of surface water. Theoretically, our model could be coupled asynchronously to such ESMs, calculating changes in water area fraction based on the ESM's climate variable output and provide it with the updated fraction as a boundary condition for the land-atmosphere interactions in the next time step. This could not only improve the simulated carbon exchange between land surface and atmosphere, but also the surface roughness as well as moisture and energy transfer. Furthermore, our implementation of abrupt drainage could improve projections of freshwater discharge into the ocean. Alternatively to an asynchronous coupling, a simple function of water area fraction in relation to changing climate variables that emulates our model responses, could provide a way to incorporate thermokarst lake dynamics directly into ESMs without significantly increasing computational costs. It is important to note, however, that prescribing water area fractions with an external model or through incorporation of one single emulator function could violate the ESMs water and energy balance. Calculations of energy and water cycle as well as coupling schemes would need to be adjusted within the ESM. Whether this is possible depends on the technical details of the ESM and its infrastructure. If the ESM allows, the coupling with our model could potentially also provide better representations of changes in surface roughness and seasonal energy and moisture transfer from land to atmosphere in thermokarst-affected regions. Furthermore, our implementation of abrupt drainage could improve projections of freshwater discharge into the ocean. However, a successful coupling requires an accurate parameterization of our model, which remains a challenge. Without it, possible model applications remain mostly conceptual and could, for instance, involve sensitivity studies investigating the relative importance of the resolved lake dynamic processes.

3.2. Comment # 2

CC: *Additionally, towards the end of the introduction, the authors refer to the abrupt thaw model introduced by Nitzbon et al. (2020). If the goal of this study is to model abrupt thaw processes, it would be helpful to clarify what new insights this model offers beyond those provided by Nitzbon et al. (2020). What additional understanding or capabilities does this approach contribute to the study of abrupt thaw? Furthermore, elaborating on how this model connects to the broader issue of permafrost carbon feedback, particularly in terms of coupling with lake carbon emissions, would strengthen the study's relevance to permafrost climate feedback.*

AR: We thank Dr. Jafarov for drawing attention to the fact that the differences between our model and the model by Nitzbon et al. (2020) were not made clear enough and that our placement of the description of this model within the text might have been confusing. We have added a few explanatory words and moved the paragraph in question to an earlier part of the 'Introduction'. We believe that through this altered structure, it now becomes clear that the limitation of the model by Nitzbon et al. (2020) is that it is not stochastic and needs initialization data for ground conditions that our model does not require. Please see the new placement in the following.

Page 4-5, Line 119-124 (in revised manuscript):

The lack of pan-Arctic models can be ascribed to the fact that thermokarst lakes are results of small-scale processes that depend on meter-scale surface and sub-surface heterogeneities, which

cannot be resolved in larger scale models. This scaling gap leads to a poor representation of permafrost-related processes and possible biases in climate models and ESMs. Even in light of current efforts to increase resolutions of ESMs, a deterministic and physics-based representation of thermokarst lake dynamics on the landscape or pan-Arctic scale remains challenging due to the lack of area-wide high-resolution data on the relevant soil conditions, particularly ground ice distribution. While significant advances in high-resolution remote sensing of the land surface have been made, such as the mapping of ice wedge polygons (Liljedahl et al. 2024), the detection of sub-surface heterogeneities and properties is still difficult.

In Nitzbon et al. (2020) CryoGrid3 was extended beyond the one-dimensional site level by using laterally coupled tiles. While this tiling approach allows for a representation of spatial heterogeneity in surface and subsurface conditions, pan-Arctic or landscape scale simulations would still require initialization data for these heterogeneities including the spatial distribution of ground ice, since each tile contains a one-dimensional vertical representation of the subsurface that needs to be parameterized. The one-dimensional, deterministic, and physics-based nature of CryoGrid3 also suggests that a high resolution would be necessary for reliable results with such large scale simulations, making it computationally expensive.

We have also added a few sentences on how the model connects to the broader issue of permafrost carbon feedback to the paragraph that discusses the possible usefulness to ESMs, which we have also extended in responses to Dr. Jafarovs first comment (see 3.1).

4. Additional Changes

4.1. Changes To Model Simulations:

We found an oversight in how our model calculated the drained area fraction, which led to significant discrepancies in the visible drained area in the spatial plots and the timeseries plots of drained area fraction. This also influenced the scaling of formation and drainage rates. We corrected this and added a more detailed explanation on how the fraction is calculated to the revised manuscript:

Page 10, Line 285-288 (in revised manuscript):

The influence of water and drained area is only incorporated implicitly. The areas $A_f(t)$ and $A_d(t)$ influence the number of formation or drainage events across a region, but not the spatial distribution of new lakes. It is therefore possible to see new lakes forming on drained area in both variants. In Variant 1, however, more drained area will lead to a smaller likelihood of lake formation and a higher likelihood of lake drainage. This effect does not exist in Variant 2, because in this model version these likelihoods are only dependent on the water area. For the calculation of drained area, this model version subtracts new water area from the existing drained area in proportion to the current drained area fraction, and does not consider the location of the new water area explicitly. For instance, in a cell with a drained area fraction of 0.3, 30% of new water is assumed to appear on this drained area. In Variant 1, none of this water is assumed to appear on drained area.

Together with the changes to the merging algorithm in response to the reviewer's comments, this has led to slightly different model behaviour and simulation results. We therefore had to adjust the parameterization

used for the idealized simulations:

Page 13, Table 2 (in revised manuscript):

Table 4: Hypothetical parameter values, fraction limit and model variant used for the three different idealized simulations. Parameters are per year; for a 10-year time step interpretation, values would be per decade.

Regime / Parameter	$\lambda_{f,v1/v2}$ [m^{-2}]	$\lambda_{d,v1/v2}$ [m^{-2}]	μ	σ	A_{lim}	variant
A. Complete Drainage	1e-9	1e-10 <u>2e-10</u>	0.30 <u>1</u>	0.02 <u>0.1</u>	1	1
B. Oscillation	1e-9	1e-10 <u>2e-10</u>	0.30 <u>1</u>	0.02	1	2
C. Stabilization	1e-8 <u>1e-9</u>	1e-12 <u>2e-11</u>	0.02	0.02	0.5	1 <u>2</u>

We also had to revise the figures and values in the text that show or describe our simulation results. Related to these changes, some wording in the "Results" and "Discussion" sections had to be adjusted as well. However, these changes do not affect our conclusions or the issues mentioned in the discussion. The revisions can be seen in the following. With the changes to the model, differences between the two model variants in the observation-based simulations are less clear than previously. We therefore removed two sentences in the "Results" section mentioning these differences.

Page 15-18, Line 408-458 (in revised manuscript):

Regime A: Complete Drainage

For a complete drainage regime, we use Variant 1. As described in Sect. ??, this variant's lake formation will become less likely the more the landscape is filled with water and drained area. At the same time drainage probability will increase. As a consequence, the system will likely reach a point where every lake has drained and no new lakes can form. In Fig. 8a an initial increase in water area fraction in the study region and number of lakes can be seen, as first lakes form in the previously lake-free system. In year ~~100~~70 of the simulation, a lake-filled landscape has established and first lakes have started to drain. A spatial representation of this can be seen in Fig. 8b. Newly formed lakes have been randomly distributed in the simulation area of 40x40 km. Since μ is positive, lakes have also expanded, causing them to reach different sizes and occasionally merge. With more larger lakes, abrupt drainage is more likely to result in stepwise decreases of water area fraction, as can be seen in Fig. 8a between year ~~240 and 241~~122 and 123 and 123. In year ~~240~~122 the system is mainly dominated by one large lake (Fig. 8c), which completely drains in year ~~241~~123. Lakes continue to drain until the water area fraction has reached zero and no active lakes remain (Fig. 8d). Since lake formation probability is scaled by $A_{f,v1}$ according to eq. 8 and ??, no new lakes can form after the drained area has reached a certain threshold. This leads to a relatively stable system, in which almost all area is covered by drained basins (Fig. 8a). While there are no known regions that are completely covered by drained lake basins, but do not have any extant lakes, there are areas with an especially high fraction of drained area, namely the Yamal Peninsula in Siberia (von Baeckmann et al., 2024) with 78% and the Yukon–Kuskokwim Delta (Jones et al. 2022). It can not be ruled out that the remaining lakes in these regions will also

drain in the future, without new lakes forming.

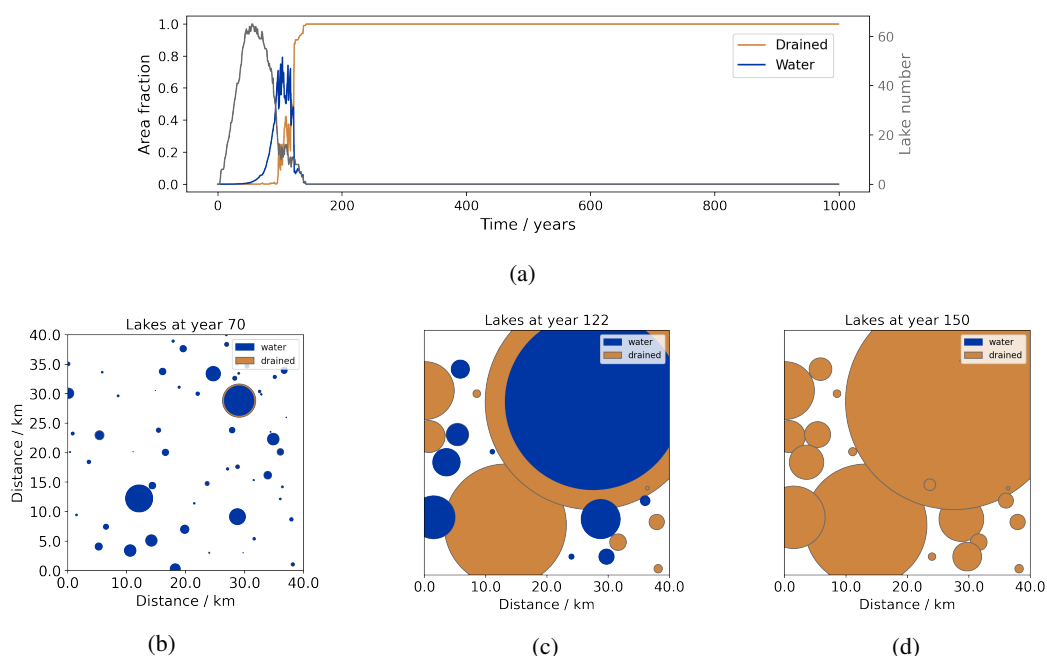


Figure 8: Model output for an idealized simulation for a complete drainage regime ("Regime A") using the parameter values from Table 4: Timeseries of water and drained area fraction as well as lake number (a) and spatial representation of lake distribution and drained area extent for simulation year ~~10070~~ (b), year ~~240122~~ (c) and year ~~900150~~ (d). Water area is displayed in blue and drained area in light brown.

Regime B: Oscillation

In order to achieve a regime, in which water and drained area oscillate with an inverse correlation, it is necessary for new lakes to be able to form on already drained area. We therefore use Variant 2 for this regime. Within the first ~~268~~^{ca. 80} years/iterations of the simulation, the system behaves similarly to regime A, while first lakes form and expand. Eventually, some become affected by abrupt drainage (Fig. 9a). The resulting landscape in year ~~10070~~ (Fig. 9b) resembles the landscape at the same time of regime A in Fig. 8b. A similar complete drainage regime as described for regime A happens between year ~~266 and 267~~¹⁰⁹⁻¹¹⁰. In this simulation, however, new lakes eventually appear, since Variant 2 scales the formation probability with $A_{f,v2}$ (eq. 8 & ??), which only contains the water area fraction. This means, that lake formation probability will rise when lakes drain and water area fraction decreases, regardless of large parts of the area being covered by drained basins. This behaviour repeats itself six times within the simulation period, as can be seen in Fig. 9a. Figure 9c represents a snapshot of the system at year ~~500315~~, where lake expansion and merging have lead to one big lake, which is close to draining and leading to the second of the abrupt system changes due to a stepwise decrease of water area fraction to almost zero. In year ~~900880~~ (Fig. 9d), a landscape with several newer differently sized lakes that resemble the lake distribution from year ~~10070~~ is visible (Fig. 9b). However, the lakes in year ~~900880~~ are already the ~~fourth~~^{sixth} generation of lakes. This type of landscape, where active lakes have appeared on

drained lake basins, has been found in several regions across the Arctic (Bockheim et al., 2004; Jones et al., 2012; Roy-Léveillé and Burn, 2017; Fuchs et al., 2019; Bergstedt et al., 2021). Interpreting the simulations as spanning 10 ka years and having a time step of 10 years, would mean that there are ~~four~~^{six} simulated lake generations over the course of 10 ka years, which is the same order of magnitude that has been observed, for example on the Seward Peninsula, Alaska, where Jones et al. 2012 found six generations of overlapping lakes that have formed over the course of the Holocene with the oldest being roughly 9 ka years old.

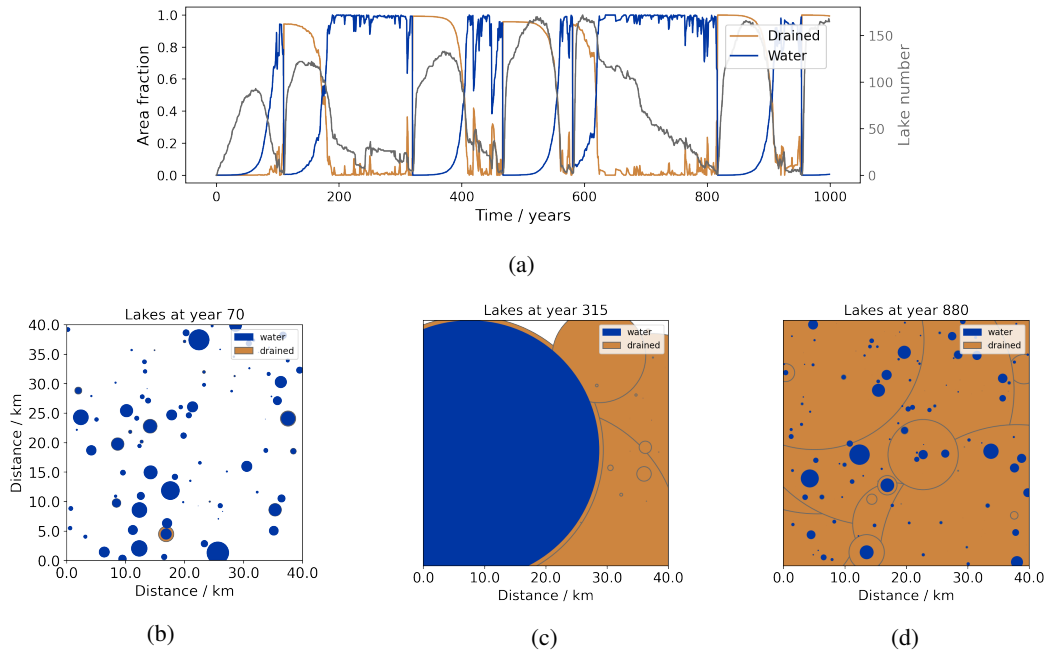


Figure 9: Model output for an idealized simulation for an oscillating regime ("Regime B") using the parameter values from Table 4: Timeseries of water and drained area fraction as well as lake number (a) and spatial representation of lake distribution and drained area extent for simulation year ~~100~~⁷⁰ (b), year ~~500~~³¹⁵ (c) and year ~~900~~⁸⁸⁰ (d). Water area is displayed in blue and drained area in light brown.

Regime C: Stabilization

Quasi-stabilization of the area fractions can occur when they have reached the pre-determined fraction limit, which keeps lakes from expanding further. For stabilization, drainage and formation rate also need to be chosen in a way that both processes are in balance at the fraction limit or their probabilities are close to zero. ~~In the former case~~^{Additionally}, lakes should not become too big, as abrupt drainage of a large lake would have a higher impact on the water area fraction than drainage of a smaller lake and would lead to more pronounced dips in water area fraction. To achieve a quasi-stabilization, we therefore also decrease μ and σ compared to the other regimes. ~~The high formation rate, still leads to a relatively steep increase in water area fraction in the beginning of the simulation (Fig. 4a).~~ When the sum of water ~~and drained~~ area fraction reaches the limit 0.5 ~~in Variant 2~~, lakes stop expanding (Fig. 10a). While lakes continue to form, they are also subject to abrupt drainage. This process happens often enough to counterbalance the water area increase

through formation, so that the ~~area fractions oscillate around~~ water area fraction oscillates at the fraction limit for the rest of the simulation. Figure 10b shows the landscape at year 50150, with many relatively small lakes. By year 200500 these have expanded and merged into bigger lakes. Additionally, some smaller lakes have formed and few lakes have drained (Fig. 10c). The larger size of the older lakes in year 900 (Fig. 10d) can mostly be explained ~~with~~ by the merging with newly formed lakes. In this regime, some individual lakes survive the simulation period while only slightly changing size. Thermokarst-affected regions with relatively stable lake area fractions and no large-scale trends over several decades have been identified using remote sensing data (e.g. Jones et al., 2009).

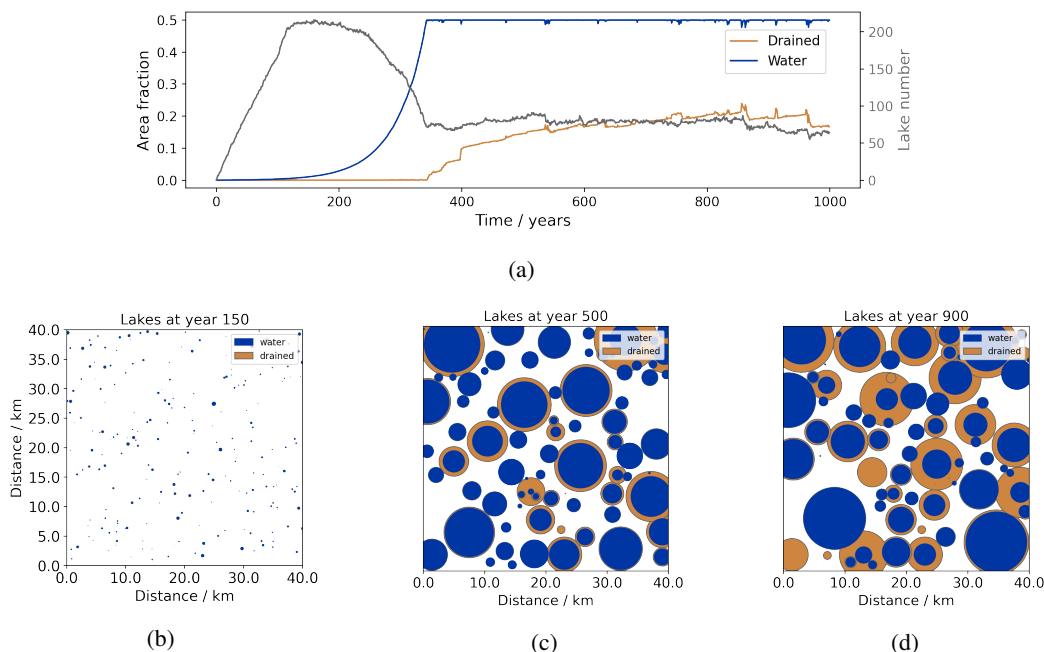


Figure 10: Model output for an idealized simulation for a stabilizing regime ("Regime C") using the parameter values from Table 3: Timeseries of water and drained area fraction as well as lake number (a) and spatial representation of lake distribution and drained area extent for simulation year 50150 (b), year 500 (c) and year 900 (d). Water area is displayed in blue and drained area in light brown.

Page 21, Line 497 (in revised manuscript):

Figures 11a and 12a show the simulated water and drained area fraction for each member in the ensemble for the two variants, as well as the ensemble mean and the standard deviation. On this timescale, distinct differences between the two variants can not be determined. The ensemble means of both variants showed an increase of water area fraction. In Variant 1, the fraction of water area in the study region increased by ~~228~~231 % from 0.16 to 0.520.53 during the 100 years of simulation. In Variant 2, the water area fraction increased similarly by 227241 %, resulting in a final water area fraction of 0.530.55 by year 2100. The yearly growth rate of lake radii averaged 0.450.53 m/a across all ensemble members for Variant 1 and 0.41only 0.33 m/a

for Variant 2. The standard deviation of water and drained area fraction in Variant 1 temporarily reached values of 0.14 and 0.16 and 0.07 , respectively. In Variant 2, they reached similar values of up to 0.15 and 0.05 , 0.14 and 0.07 . Ensemble runs with Variant 1 exhibited between 6 and 125 and 17 abrupt drainage events over the simulation period, while Variant 2 experienced 5 and 14 and 15 . In addition, Variant 1 simulated between 5 and 18 lakes that lost more than 25% of their area until the end of the simulation through gradual drainage, whereas Variant 2 simulated between 4 and 20 . With Variant 1, ~~between 284 and 429 merging events occurred~~ lake numbers decreased from the initial 512 lakes to between 161 and 266 lakes by the end of the simulation, while Variant 2 led to ~~268 - 402 merging events, meaning that the lake number decreased accordingly in the respective ensemble simulations~~ 94 - 254 remaining lakes. Figure 11b and 12b show the lake size distribution across all ensemble members with mean M and standard deviation SD at the start year 2000, year 2020 and year 2099. With both variants, the distribution flattened slightly and remained skewed to the right, while mean and standard deviation increased. A spatial representation from an example ensemble member for these three time points can be seen in Fig. 11c to 11e and Figures 12c to 12e. As explained in Sect. ?? such representations should be viewed as visual aids only. By year 2020 (Fig. 11d, 12d) some lakes had slightly decreased in size as a consequence of the relatively high volatility parameter σ . Most, however, had grown or merged. By year 2099 (Fig. 11e, 12e) in the two example simulations one or more especially large lakes had formed that dominate the landscape. ~~The mean drained area fraction in Variant 1 had an average value of 0.05 between years 2086 and 2100 (Fig. 7a), whereas Variant 2 had a drained area fraction that stayed below 0.03 throughout the simulation (Fig. 8a). This higher value in Variant 1 was mostly due to two ensemble members that showed a sudden increase of drainage area fractions as a result of an abrupt drainage event. This emphasizes the fact that Variant 1 will typically have a higher increase of abrupt drainage probability as the landscape becomes more populated with drained lake basins, because it considers drained area for the scaling of both formation and~~ ~~525 abrupt drainage rates.~~

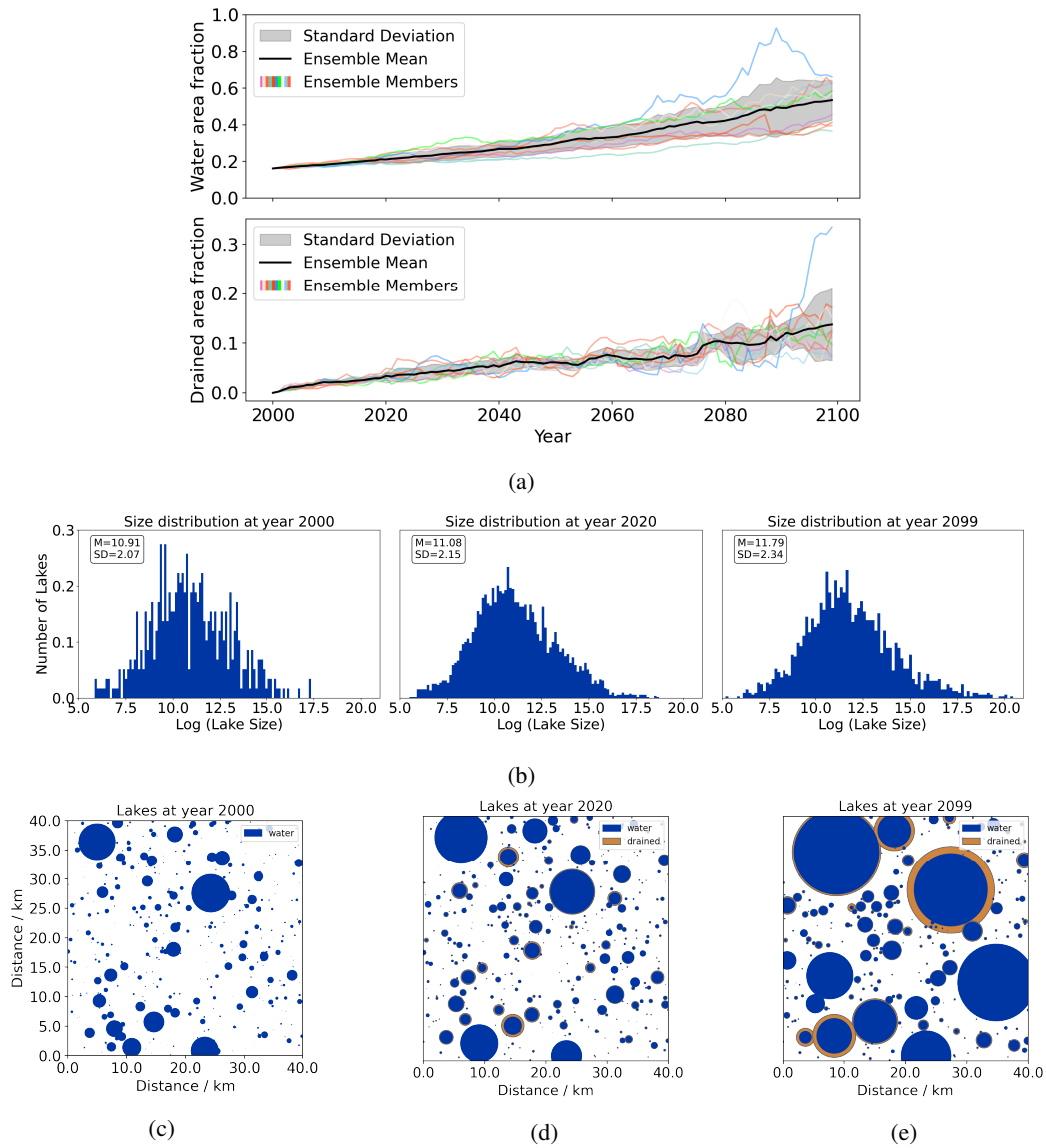


Figure 11: Model output for an ensemble simulation with Variant 1 and using the parameter values from Table 3: Time series of water (top) and drained area fraction (bottom) of all ensemble members as well as ensemble means and standard deviation (a). Histograms for size distribution across all ensemble members (b) at start of simulation (left), year 20 (middle) and year 2099 (right) with mean M and standard deviation SD . Spatial representation of lake distribution and drained area extent of an example ensemble simulation for start of simulation (c), year 20 (d) and year 2099 (e) with water area displayed in blue and drained area in light brown.

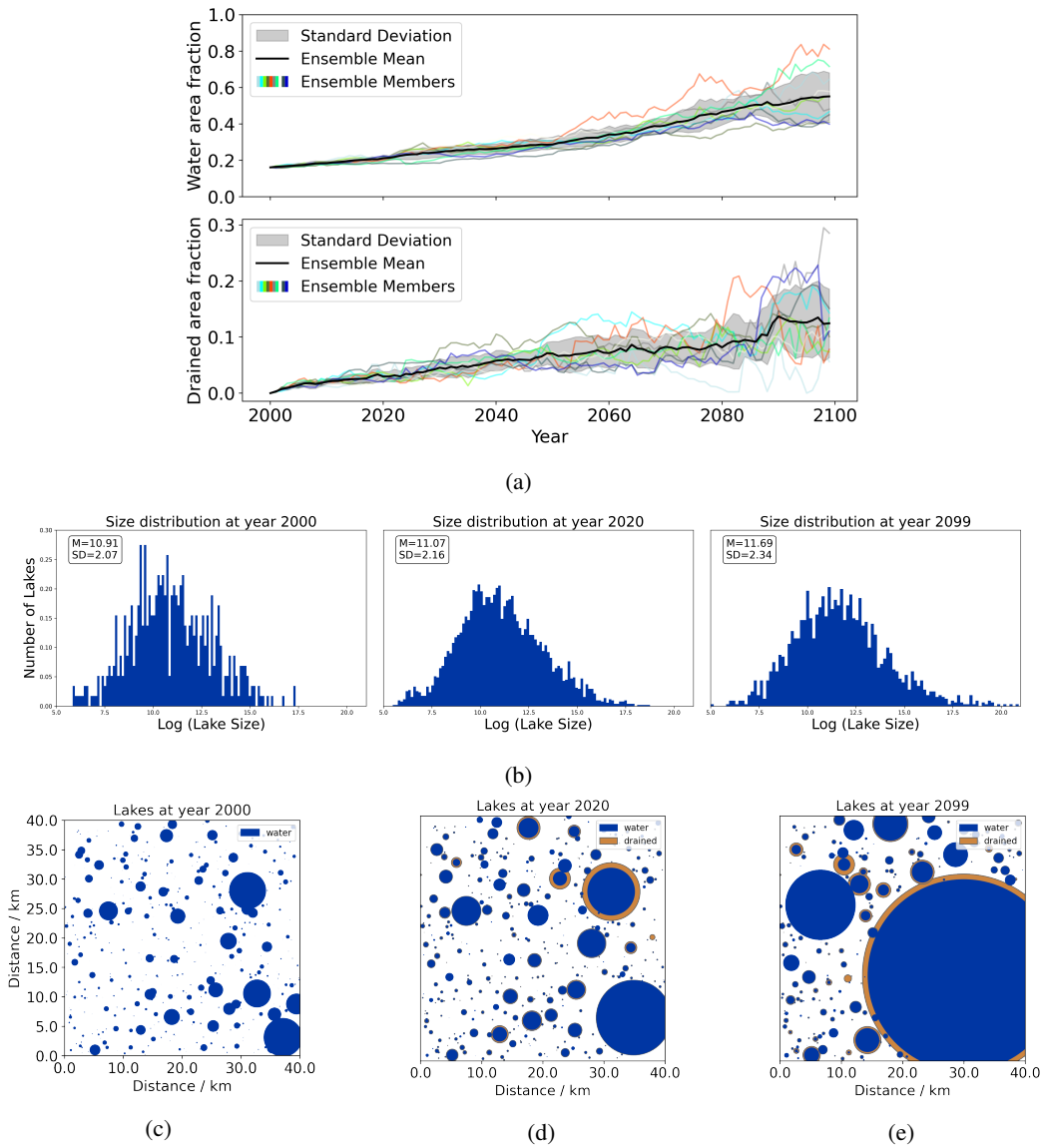


Figure 12: Model output for an ensemble simulation with Variant 2 and using the parameter values from Table ??: Time series of water (top) and drained area fraction (bottom) of all ensemble members as well as ensemble means and standard deviation (a). Histograms for size distribution across all ensemble members (b) at start of simulation (left), year 20 (middle) and year 2099 (right) with mean M and standard deviation SD . Spatial representation of lake distribution and drained area extent of an example ensemble simulation (c) for start of simulation (c), year 20 (d) and year 2099 (e) with water area displayed in blue and drained area in light brown.

4.2. Space-for-Time Approach Discussion

After some internal discussion, we added two sentences discussing the potential of a space-for-time approach for model parameterization.

Page 27, Line 650-654 (in revised manuscript):

Using a space-for-time approach for parametrization should also be handled with care for that reason. Still, such an approach could provide first estimates on how different climate states effect thermokarst lake dynamics. It could also help in identifying thresholds of climate variables that lead to a different process dominating the overall lake area trend in the system, and provide additional information for model calibration.

4.3. Additional References

Please also note that there have been several references added, but are not highlighted as new in the marked-up manuscript for technical reasons. We have listed them in the following:

V. Stojkoski, T. Sandev, L. Basnarkov, L. Kocarev, R. Metzler, Generalised Geometric Brownian Motion: Theory and Applications to Option Pricing, *Entropy*, Vol. 22, No. 12, p. 1432, 2020, <https://doi.org/10.3390/e22121432>

Paul A. M. Dirac, The physical interpretation of the quantum dynamics Free, 1927, *Proceedings of the Royal Society*, Vol. 2, p. 391-401, 1927, <https://doi.org/10.1093/rps/2.2.391>

Noah D. Smith, Eleanor J. Burke, Kjetil Schanche Aas, Inge H. J. Althuizen, Julia Boike, Casper Tai Christiansen, Bernd Eitzelmüller, Thomas Friborg, Hanna Lee, Heather Rumbold Rachael H. Turton, Sebastian Westermann, Sarah E. Chadburn, Explicitly modelling microtopography in permafrost landscapes in a land surface model (JULES vn5.4_microtopography), *Geosci. Model Dev.*, Vol. 15, p. 3603-3639, 2022, <https://doi.org/10.5194/gmd-15-3603-2022>

K. M. Walter, M. E. Edwards, G. Grosse, S.A. Zimov, F. S. Chapin, Thermokarst lakes as a source of atmospheric CH_4 during the last deglaciation, *Science* 318, 633-636, 2007, <https://doi.org/10.1126/science.1142924>

Z. Rehder, T. Kleinen, L. Kutzbach, V. Stepanenko, M. Langer, V. Brovkin, Simulated methane emissions from Arctic ponds are highly sensitive to warming, *Biogeosciences*, Vol. 20, 2837-2855, 2023, <https://doi.org/10.5194/bg-20-2837-2023>

Z. Rehder, A. Zaplavnova, L. Kutzbach, Identifying Drivers Behind Spatial Variability of Methane Concentrations in East Siberian Ponds, *Frontiers in Earth Science*, Vol. 9, 2296-6463, 2021, <https://doi.org/10.3389/feart.2021.617662>

S. Juutinen, M. Rantakari, P. Kortelainen, J. T. Huttunen, T. Larmola, J. Alm, J. Silvola, P. J. Martikainen, Methane dynamics in different boreal lake types, *Biogeosciences*, Vol. 6, 209-223, 2009, <https://doi.org/10.5194/bg-6-209-2009>

Y. M. Polishchuk, A. N. Bogdanov, I. N. Muratov, V. Y. Polishchuk, A. Lim, R. M. Manasypov, L. S. Shirokova, O. S. Pokrovsky, Minor contribution of small thaw ponds to the pools of carbon and methane in the inland waters of the permafrost-affected part of the Western Siberian Lowland, *Environmental Research Letters*, Vol. 13, 2018, <https://doi.org/10.1088/1748-9326/aab046>

L. Kutzbach, D. Wagner, E.-M. Pfeiffer, Effect of microrelief and vegetation on methane emission from wet polygonal tundra, Lena Delta, Northern Siberia, *Biogeochemistry*, Vol. 69, 341-362, 2004,

<https://doi.org/10.1023/B:BIOG.0000031053.81520.db>

C. Knoblauch, O. Spott, S. Evgrafova, L. Kutzbach, E.-M. Pfeiffer, Regulation of methane production, oxidation, and emission by vascular plants and bryophytes in ponds of the northeast Siberian polygonal tundra, *Journal of Geophysical Research: Biogeosciences*, Vol. 120, 2525-2541, 2015, <https://doi.org/10.1002/2015JG003053>

F. A. Cowell, E. Flachaire, Statistical Methods for Distributional Analysis, In: *Handbook of Income Distribution*, edited by Anthony B. Atkinson, F. B., vol. 2, pp. 359–465, Elsevier, 2015, <https://doi.org/10.1016/B978-0-444-59428-0.00007-2>

4.4. Supplementary Figures

We have corrected some variable names in Supplementary Figure A.1 and its caption:

Page 31 (in revised manuscript):

Appendix A: Modelling Scheme

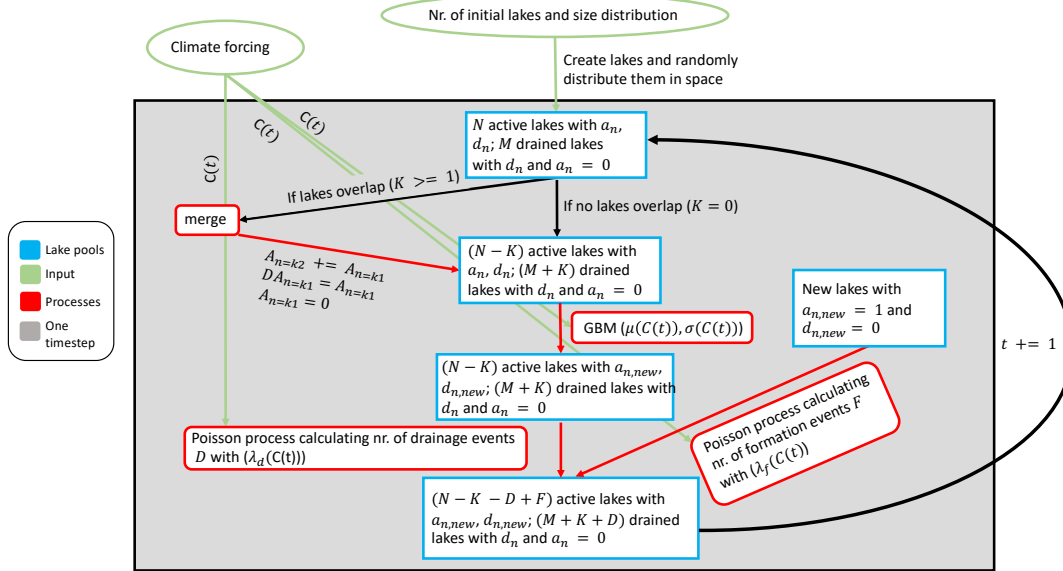


Figure 13: Modelling scheme. Green indicates input into the model, i.e. initialization data and a climate forcing. The climate forcing consists of an annual time series for precipitation P and thaw degree days (TDD). The grey box contains all operations that happen at each time step. Blue indicates the model objects, i.e. the active and inactive lake pool. Red indicates the main computations and processes, which are: merging of overlapping lakes, geometric Brownian motion (GBM) simulating lake expansion and gradual drainage, a Poisson process that calculates the number of drainage events, and a Poisson process that calculates the number of formation events, i.e. new lakes that are added to the lake pool. N is the initial number of active lakes at each timestep, M is the initial number of inactive lakes. A_n is the water surface area of the n -th lake, whereas DA_n is its drained area. $A_{n,new}$ and $DA_{n,new}$ are both of these areas after applying expansion or gradual drainage via GBM. K is the number of overlaps and therefore merging events. D is the number of drainage events, and F the number of formation events. $P(t)$ and $TDD(t)$ are $C(t)$ contains the values forcing in form of the a selected climate variables precipitation and thaw degree days variable at time t .