Author response to Referee #1 comments for Manuscript #HESS-2020-222, "Variability in epilimnion depth estimations in lakes" by Harriet L. Wilson, Ana I. Ayala, Ian D. Jones, Alec Rolston, Don Pierson, Elvira de Eyto, Hans-Peter Grossart, Marie-Elodie Perga, R. Iestyn Woolway, Eleanor Jennings.

Thank you for your valuable comments which have enabled us to improve our study. Please find below our responses to each comment respectively. Updated text is in italics.

1. Referee comment: The article evaluates the epilimnion depth estimate in high-frequency data made by four different methods, including the effect of defining different thresholds that these methods require the user to choose. It also made this estimate using a hydrodynamic physical model. The aim of the study was to highlight the variability of the epilimnion depth estimates and how this variability could impact inferences about lake processes. This study draws attention to the need for researchers to unify their consensus on this topic, allowing comparisons between the results of different studies.

This is a very important study that addresses a hot topic in limnology and oceanography. With the increase in studies evaluating different bodies of water and the ease of obtaining increasingly sensitive measuring equipment, with the ability to perform and store an increasing number of measurements, we have entered the era of Big Data. With the increasing availability of data, the need for models capable of extracting the correct information from them also increases. The correct adjustment of these models depends on studies of this type.

Author response: Thank you for these positive comments. We greatly appreciate the enthusiasm for the study as a topic of current relevance.

2. Referee comment: I believe that the result obtained and discussed in the article mixed the estimates made in clearly stratified water columns, with estimates made in weakly stratified water columns with estimates made in water columns with the presence of multiple stratifications. Therefore, the described variability is not only due to the distinction between methods and limits, but also to the application of the methods under different conditions, and this should be clearer.

Author response: This is an important point by the reviewer, and we agree that both 1) the presence of multiple stratifications and 2) weakly stratified profiles, contribute to the large variability found between epilimnion depth estimates. In fact, it was *particularly* when water column profiles did not conform to the idealised three-layered structure, that we found the greatest divergence between epilimnion depth methods. However, we would argue it is not simply a matter of filtering out weakly stratified or multiple stratification profiles, since this would result in large data gaps, including in peak summer, and is limited by the same issues of subjectivity as the definition of the epilimnion depth itself. Instead, there is a need for users to acknowledge both the systematic differences between methods, and the application of the methods under different conditions, in respect to their specific purpose. We have addressed this point in the revised manuscript (please see Comment #3 for relevant edits to text).

3. Referee comment: P. 8 L. 281-283: I think that authors should make a clearer distinction between primary and secondary thermocline. For example, the method described by Read et al (2011) allows us to estimate the two thermoclines and, therefore, is more sensitive than the other methods, which do not consider this possibility. The comparison between the methods must consider the presence of these superficial micro-stratifications. For example, in the graphs b) of figure 4, there is clearly the presence of these microstratifications that are hampering some methods of identifying the main thermocline. In other words, it is not fair to fit a model that expects only one stratification with profiles that show various stratifications. It is obvious that the estimate will not be satisfactory. It is extremely necessary to make a pre-filter, removing the superficial layers of values before the estimate is made. This method was applied by Pujoni et al. (2019).

Author response: See the response to Comment #2 above and Comment #4 which are related. We do not think it is appropriate in this study to make a defined distinction between primary and secondary thermoclines, or filter/smooth out these stratifications for the following reasons; 1) determination of primary/secondary thermoclines is not objective and the selection of profiles requiring smoothing would demand further arbitrary thresholds (instead we argue that this topic deserves a separate study, investigating the full implications of different approaches), 2) in this study we aimed to estimate variability in epilimnion depth estimates between *common and existing* methods (rather than introduce new methods), where issues related to microstratification are a result/discussion point, 3) smoothing of these microstratifications may not be suitable for some applications, particularly where it is assumed that the epilimnion is isothermal and well-mixed. Nevertheless, we do fundamentally agree with the reviewer that this is an important issue that should be clearer in the text, therefore we propose editing the discussion to emphasise these points (see below). We also now cite the work of Pujoni et al., (2019) and Read et al., (2011) in this context.

L.388: '*The concept of the epilimnion, and more widely, the three-layered structure of a stratified lake, is fundamental in limnology. Yet, despite the ubiquity of the term, there is no objective or generic approach for defining the epilimnion and a diverse number of approaches prevail in the literature. In a comprehensive analysis of high-frequency, multi-year data from two lakes, this study has highlighted the extent to which common water temperature profile based epilimnion depth estimates differ. The level of variability in epilimnion depth estimates calculated using common methods and threshold values, was exceedingly high. This result calls into question the practice of arbitrary method selection and comparing findings between studies which use different methods or even just different thresholds. The magnitude of variability also casts ambiguity on the calculation of key biogeochemical and ecological processes in a lake that rest on the assumption that the layers of a lake are well defined, including calculations of metabolic rates, and oxygen fluxes (e.g. Coloso et al., 2008, Foley et al., 2012, Obrador et al., 2014, Winslow et al., 2016).*

In an idealised stratified profile, the epilimnion is portrayed as near-uniform in water temperature or density and clearly delineated from a well-defined metalimnion. However, many measured profiles, at least within this study, did not conform to this idealised three-layered structure. Instead the water columns were often more complex, including multiple pycnoclines and near-surface micro-stratification layers, or the boundaries between the epilimnion/metalimnion were blurred. One approach to this issue is to filter out appropriate water column profiles or apply functions that coerce the profile into the expected structure (Read et al., 2011, Pujoni et al., 2019, Gray et al., 2020). Filters, additional conditions or smoothing functions, however, may suffer from many of the same challenges as the estimation of the epilimnion depth, since they attempt to discretise data based on arbitrary criteria (Kraemer et al., 2020). For example, our analysis of temporally high resolution time series data emphasised that rather than jumping from states, such as stratified or isothermal, changes in the water *column occurred over an evolving continuum and often fluctuated between states. Similarly, the distinction between additional layers, such as the primary or secondary pycnocline, is fraught with the same issues of arbitrariness as discussed (Read et al., 2011). This study demonstrates that when epilimnion depth estimation methods, which are theorised for a three-layered water column, are applied to non-conforming water columns, they diverge widely on the location of the epilimnion depth, and at times, may not even be underpinned by the same theoretical assumptions. Since none of these methods can be considered the 'true' definition of the epilimnion depth, it is necessary to understand the degree to which methods differ. Improved understanding of their systematic differences will facilitate the use of methods that appropriately capture different processes, such as, air-water exchanges, thermocline entrainment or suspension of materials. Due to the realised complexities of observed and aggregated profile data, we may benefit from new approaches to water column discretisation that incorporate the vast proportion of profiles which do not conform neatly to the three-layered paradigm.'*

4. Referee comment: P. 8 L. 288-289: In this same line, we must discuss and define a threshold of what we call "homogeneous water column". I don't think it makes sense to compare the methods using profiles with low stability of the water column. If we no longer have a clear stratification, the methods should not be applied, as they will look for a thermocline that does not exist. I may be wrong, but the water column in the graphs c) in figure 4 is homogeneous and should not be subjected to comparison with these models.

Author response: This is another great point by the reviewer, and we agree that the stability of the water column has an influence on the results. Inherently, however, there are conditions within each of the methods, relating to the degree of stratification required to estimate the epilimnion depth. Method 1 has the precondition that the range in water density must be greater than the threshold value, else the epilimnion depth is assigned to the maximum lake depth. Similarly, Methods 2 and 3 have the precondition that the water density gradient must be greater than the threshold value. Finally, Method 4, the rLakeAnalyzer, is slightly different as it initially filters out profiles based on a 1°C water column range and will then identify the maximum density gradient regardless of the threshold value. As discussed inComment #3, a further stability-based condition could be introduced, but would suffer from being an arbitrary threshold to which the rest of the results would become dependent. Again, it is the fact that studies are currently using different approaches with different inherent stability thresholds that can contribute to the confusion caused when comparing studies, and this is one of the key points we are raising in this manuscript.

Note, though, that we did extensively investigate the use of pre-filters, including top-bottom density differences, water column total density gradient and Schmidt stability values. However, echoing the presented findings of the epilimnion depth analysis, we found different methods and thresholds largely altered the period that was deemed stratified. This analysis could not be justly presented in this study, without overly complicating the manuscript. In addition, they resulted in the removal of large amounts of data, even within peak summer, which is not suitable for analysis of mixing events for example. Finally, presenting the full time series results demonstrated the perils of using temporal means of epilimnion depth. For example, for calculating the summer mean epilimnion depth, the un-filtered mean will be influenced by periods of very deep epilimnion depth estimates (i.e. when the water column is nearly isothermal), while the filtered mean would not be representative of conditions during the full summer, but rather a subset of the stratified profiles.

5. Referee comment: P. 9 L. 350: Why did the authors use the range to estimate variability? The range is sensitive to outliers. Why not use standard deviation, which is a more robust estimate of variation?

Author response: The reviewer has highlighted an important point and in the revised document we will use inter-quartile range which more robust to outliers than range. To address this we update the methods, results (Table 2 and Figure 7), although the findings are very similar.

Table 2. Summary of statistics for each method, showing the mean (m), minimum (i.e. shallowest estimate) (m), maximum (i.e. deepest estimate) (m) and the interquartile range (m) of the April-October epilimnion depth estimates (summarised from the results shown in Fig.6a), and the mean Pearson's correlation coefficient (r) for each method, representing the mean correlation for all possible combinations between threshold values., for Lough Feeagh and Lake Erken.

	Lough Feeagh					Lake Eken				
Method	Mean (m)	Min (m)	Max (m)	IQR (m)	r	Mean (m)	Min (m)	Max (m)	IQR (m)	r
M1	-19.0	-4.6	-25.4	7.3	0.77	-10.0	-7.8	-11.2	1.3	0.92
M2	-35.9	-19.7	-41.4	6.5	0.48	-11.3	-8.4	-12.9	1.7	0.78
M3	-36.5	-22.4	-41.5	6.1	0.49	-11.8	-10.0	-13.0	1.2	0.82
M4	-21.1	-19.7	-21.5	0.3	1.00	-11.9	-10.1	-11.3	0.5	0.99

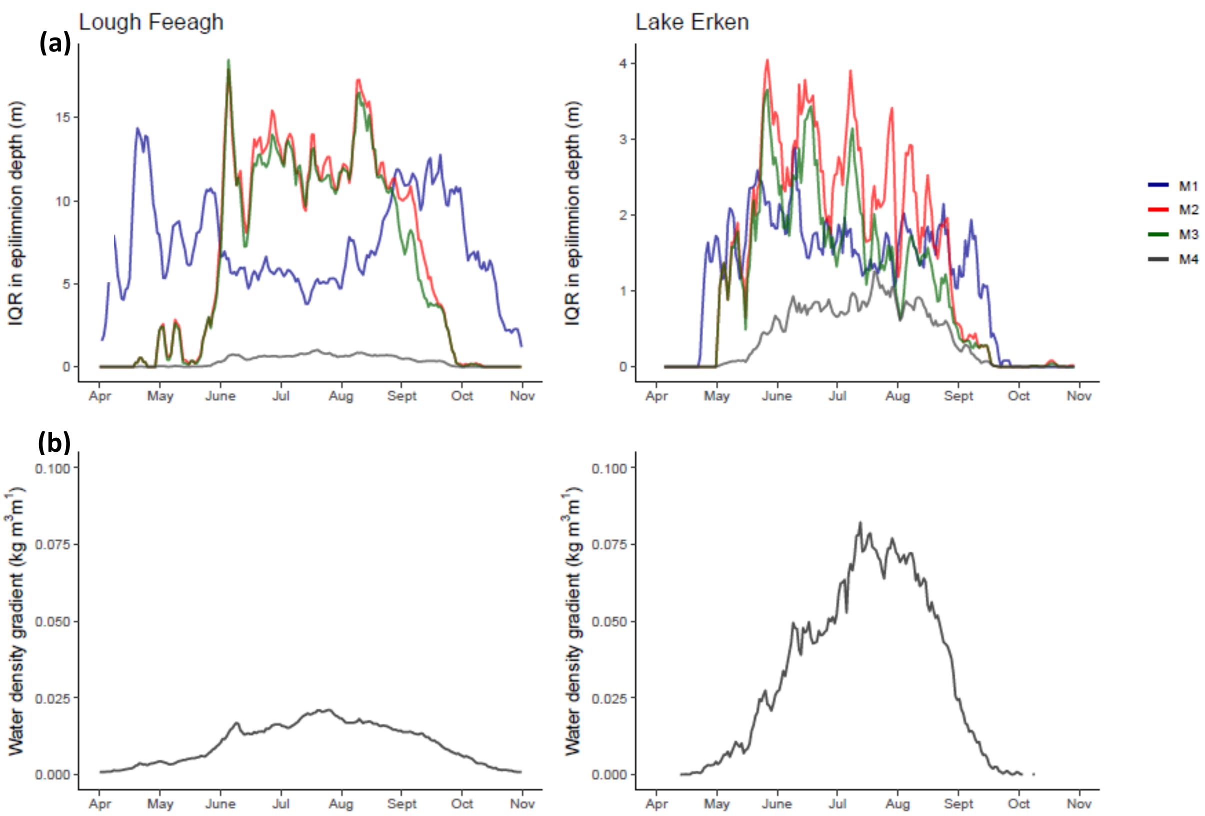


Figure 7. Inter-quartile range between the shallowest and deepest estimate for each method calculated from long-term daily mean epilimnion depth estimates for each Julian day, where a large range suggests high threshold sensitivity and a small range suggests low sensitivity **(a)**, and long-term daily mean water density gradient, calculated based on the surface and maximum measured depths **(b)**, for Lough Feeagh and Lake Erken.

L. 225-227: *'We also summarised these statistics for each method, showing the mean, minimum (shallowest), maximum (deepest) and interquartile range for each method, to demonstrate differences between methods. A large interquartile range in epilimnion depth estimates, indicated high sensitivity to the threshold value.'*

L. 317 – 323: *'For both lakes, the interquartile range in the mean Apr-Oct epilimnion depth estimates for each method was very high for M2, M1 and M3, indicating high threshold sensitivity in these methods. Method M4 had a substantially lower interquartile range than all other methods and a very high mean Pearson's correlation coefficient, indicating that both the mean value and the temporal pattern of the epilimnion depth were only weakly influenced by the threshold value. In both lakes, methods M2 and M1, where the epilimnion depth was defined from the surface downwards, had a higher interquartile range in estimates calculated with different threshold values, compared to methods M3 and M4, where the epilimnion was defined from the pycnocline upwards.'*

L.348 – 366: '*For all methods, threshold sensitivity fluctuated seasonally, although varied in pattern (Fig. 7). Threshold sensitivity was shown by the interquartile range between the shallowest and deepest epilimnion depth estimates calculated for all threshold values. In Lough Feeagh, M1 had a smaller range in epilimnion depth estimates during the peak summer months of June, July and August, compared with months when the onset and overturn of stratification commonly occurred. During periods of transient stratification, the stability of the water column was often low but frequent changes in the near-surface water density, induced large differences between estimates calculated using small thresholds compared with large threshold values. In contrast, methods M2 and M3 had the highest range in estimates occurring during the peak summer months. Even during peak summer in Lough Feeagh, gradients in the water column were relatively small (Fig. 7b), which resulted in a very large range between the smallest threshold values which found a near-surface epilimnion depth, and the largest thresholds that often found no epilimnion depth at all, therefore defaulting to the deepest depth. In Lake Erken, the water density gradients were typically much larger, and methods M1, M2 and M3 all peaked during May and June, when gradients in the water column were typically increasing but prone to fluctuations. For both lakes, M2 had typically a higher threshold interquartile range than M3 during peak summer and the overturn period, which was related to the common development of a secondary pycnocline. M4 produced much lower interquartile ranges in the epilimnion depth throughout the year, since as long as the 'mixed.cutoff' filter was met, the epilimnion depth was defaulted to the pycnocline if the threshold was not exceeded, thus largely reducing the ability for large differences to occur. The interquartile range in epilimnion depth estimates for M4 was highest during the peak summer months, which was when the epilimnion depth was typically shallowest and more frequently defined by the threshold value rather than defaulting to the pycnocline.'*

6. Referee comment: P. 10 L. 379-382: I would suggest showing some graphs of density profiles with the estimated depths of the methods so that it would be easy to see why there were such differences and whether one method made a "better" estimate than the other. I would suggest that the authors discuss a little about the visual assessment of profiles. Should we rely on this visual assessment to try to "correct" the biased estimates made by the models?

Author response: Visual assessment of profiles is certainly very helpful and is also commonly practised in limnology. We think that with 5 tables and 7 figures it may be excessive to add additional figures. Our intention with Figure 4 was to demonstrate to the user all methods/thresholds on three distinctly different profiles, however, visual assessment is not part of our result analysis. The focus for this study is for using high-frequency data and multi-lake analysis, and therefore the goal is to find methods that can be used systematically without being tailored to specific lakes.

We are interested however in visual assessment of epilimnion depth and we have conducted a survey investigating where limnologists visually identify the epilimnion depth using profiles from anonymous lakes. This is something we would like to publish at a later date as a short discussion paper, but is not appropriate for automated analysis of high frequency data.

Author response to Referee #2 comments for Manuscript # hess-2020-222, "Variability in epilimnion depth estimations in lakes" by Harriet L. Wilson, Ana I. Ayala, Ian D. Jones, Alec Rolston, Don Pierson, Elvira de Eyto, Hans-Peter Grossart, Marie-Elodie Perga, R. Iestyn Woolway, Eleanor Jennings.

Thank you for your valuable comments which have enabled us to improve our study. Please find below our responses to each comment respectively. Updated text is in italics.

1. Referee comment: The study addresses one of the fundamental paradigms of limnology, the three-layer structure of the stratified water column. Here, the authors compare different algorithms to quantify the depth of the epilimnion, defined as a well-mixed, homogenous surface layer. As a general and agreed on mathematical definition of the distinction between epilimnion, metalimnion and hypolimnion is missing, different arbitrary thresholds for the epilimnion depth were investigated on two lake systems. As this paper aims to quantify the variability of epilimnion depth estimations and the methodological differences between alternative algorithms, it is of huge interest for a wide audience of limnologists, water managers, oceanographers, modelers, and environmental engineers. The study design, methods and results are well explained, although some paragraphs should be improved. Overall, the results of the study are important for future limnological research and are challenging our current conceptual paradigm.

Author response: Thank you for these positive comments. We greatly appreciate the enthusiasm for the study as a topic of current relevance.

2. Referee comment: Quantifying the data variability by computing the range has potential shortcomings, e.g. bias by outliers, no information regarding the distribution of data. Specifically, in this study the authors did state the maximum and minimum values enabling every reader to calculate the range themselves. The authors should think about computing and stating alternative metrics like the variance, standard deviation and/or the interquantile range.

Author response: The reviewer has highlighted an important point and in the revised document we will use inter-quartile range which more robust to outliers than range. To address this we update the methods, results (Table 2 and Figure 7), although the findings are very similar.

Table 2. Summary of statistics for each method, showing the mean (m), minimum (i.e. shallowest estimate) (m), maximum (i.e. deepest estimate) (m) and the interquartile range (m) of the April-October epilimnion depth estimates (summarised from the results shown in Fig.6a), and the mean Pearson's correlation coefficient (r) for each method, representing the mean correlation for all possible combinations between threshold values., for Lough Feeagh and Lake Erken.

	Lough Feeagh					Lake Eken				
Method	Mean (m)	Min (m)	Max (m)	IQR (m)	r	Mean (m)	Min (m)	Max (m)	IQR (m)	r
M1	-19.0	-4.6	-25.4	7.3	0.77	-10.0	-7.8	-11.2	1.3	0.92
M2	-35.9	-19.7	-41.4	6.5	0.48	-11.3	-8.4	-12.9	1.7	0.78
M3	-36.5	-22.4	-41.5	6.1	0.49	-11.8	-10.0	-13.0	1.2	0.82
M4	-21.1	-19.7	-21.5	0.3	1.00	-11.9	-10.1	-11.3	0.5	0.99

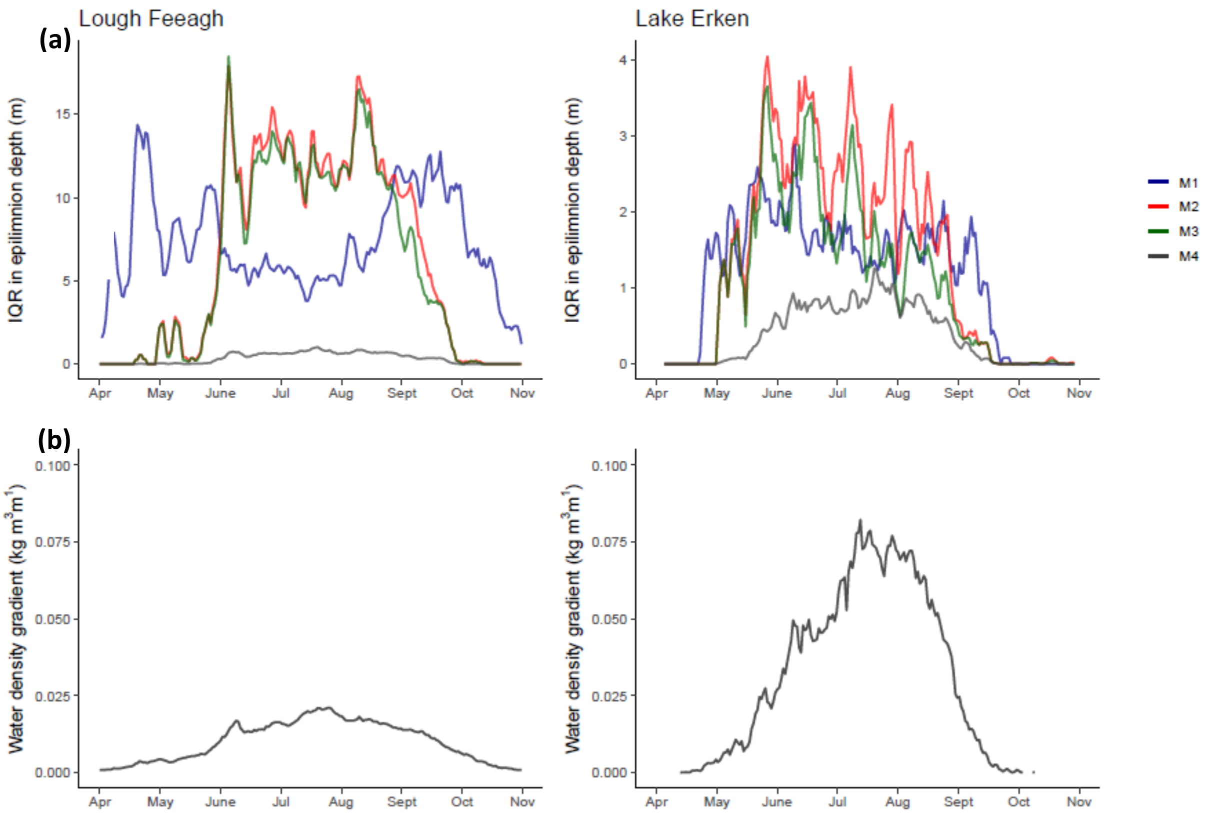


Figure 7. Inter-quartile range between the shallowest and deepest estimate for each method calculated from long-term daily mean epilimnion depth estimates for each Julian day, where a large range suggests high threshold sensitivity and a small range suggests low sensitivity **(a)**, and long-term daily mean water density gradient, calculated based on the surface and maximum measured depths **(b)**, for Lough Feeagh and Lake Erken.

L. 225-227: *'We also summarised these statistics for each method, showing the mean, minimum (shallowest), maximum (deepest) and interquartile range for each method, to demonstrate differences between methods. A large interquartile range in epilimnion depth estimates, indicated high sensitivity to the threshold value.'*

L. 317 – 323: *'For both lakes, the interquartile range in the mean Apr-Oct epilimnion depth estimates for each method was very high for M2, M1 and M3, indicating high threshold sensitivity in these methods. Method M4 had a substantially lower interquartile range than all other methods and a very high mean Pearson's correlation coefficient, indicating that both the mean value and the temporal pattern of the epilimnion depth were only weakly influenced by the threshold value. In both lakes, methods M2 and M1, where the epilimnion depth was defined from the surface downwards, had a higher interquartile range in estimates calculated with different threshold values, compared to methods M3 and M4, where the epilimnion was defined from the pycnocline upwards.'*

L.348 – 366: '*For all methods, threshold sensitivity fluctuated seasonally, although varied in pattern (Fig. 7). Threshold sensitivity was shown by the interquartile range between the shallowest and deepest epilimnion depth estimates calculated for all threshold values. In Lough Feeagh, M1 had a smaller range in epilimnion depth estimates during the peak summer months of June, July and August, compared with months when the onset and overturn of stratification commonly occurred. During periods of transient stratification, the stability of the water column was often low but frequent changes in the near-surface water density, induced large differences between estimates calculated using small thresholds compared with large threshold values. In contrast, methods M2 and M3 had the highest range in estimates occurring during the peak summer months. Even during peak summer in Lough Feeagh, gradients in the water column were relatively small (Fig. 7b), which* *resulted in a very large range between the smallest threshold values which found a near-surface epilimnion depth, and the largest thresholds that often found no epilimnion depth at all, therefore defaulting to the deepest depth. In Lake Erken, the water density gradients were typically much larger, and methods M1, M2 and M3 all peaked during May and June, when gradients in the water column were typically increasing but prone to fluctuations. For both lakes, M2 had typically a higher threshold interquartile range than M3 during peak summer and the overturn period, which was related to the common development of a secondary pycnocline. M4 produced much lower interquartile ranges in the epilimnion depth throughout the year, since as long as the 'mixed.cutoff' filter was met, the epilimnion depth was defaulted to the pycnocline if the threshold was not exceeded, thus largely reducing the ability for large differences to occur. The interquartile range in epilimnion depth estimates for M4 was highest during the peak summer months, which was when the epilimnion depth was typically shallowest and more frequently defined by the threshold value rather than defaulting to the pycnocline.'*

3. Referee comment: The manuscript stresses that the data used were sampled on a high frequency (1-2 min), but I do not see the advantage of using high-frequency data here in this study as diurnal epilimnion depth trends/oscillations were not discussed at all. At the end, as mostly seasonal means, ranges or data point fluctuations were discussed, short-term dynamics were neglected. Would a study that uses for example bi-weekly sampled vertical profiles over 50 years give the same results?

Author response: We used sub-daily data to aggregate to daily means, and then conducted the analysis on these daily values. There are, in fact, a number of extra (and interesting) complications when considering mixed depth on sub-daily timescales, such as seiche activity or diel cycles, which we believe would be worth addressing, but would require a whole new study. However, there are still important advantages of using high-frequency data even at the daily time-step. These are; 1) Daily water temperature estimates are still high-frequency (compared to fortnightly or monthly monitoring campaigns) and are commonly collected or calculated within models, thus making analysis particularly relevant, 2) Daily estimates tell us a lot more about the temporal evolution of stratification patterns than fortnightly or monthly one-off profiles, which is an important consideration in our results, 3) As an aggregate of finer resolute data (i.e. mean of sub-daily measurements), the daily data we use represents the variability within the day and are not affected by the time of detection as manually collected profiles would be, 4) Daily data produces many profiles to analyse (i.e. if we were using, say, monthly data, we would need to have 30 times as many years of data to get the same number of profiles), which allows us to be more confident in our results (i.e. more robust mean epilimnion depth, more validation in the variability we report and more robust in comparing the methods). To demonstrate this to the reader, we propose changing the following line in the introduction.

L.95: 'Although lower temporal resolution data is sufficient for investigating seasonal patterns, high-frequency data can be used to gain information on the level of day-to-day variability in epilimnion depth and demonstrates how methods perform over a continuum of water column conditions. In addition, through the vast number of measured profiles, high-frequency data offers a more robust comparison of methods, than previously demonstrated with manually collected datasets, and even when aggregated to the daily time-step is more representative of the sub-daily variability (Marcé et al., 2016).'

4. Referee comment: I'd advise the authors to discuss the challenge of multiple pycnoclines by microstratification more intensively in the manuscript. The occurrence of these profiles especially in Erken, which seems to behave more polymictic than Feeagh, is more interesting to real-world applications (and which method should be used then) than discussing shortcomings of the classical three-layer structure.

Author response: See the response to Comment #11 and Comment #20 which are related. We do not think it is appropriate in this study to make a defined distinction between primary and secondary thermoclines, or pre-filter out these superficial stratifications for the following reasons; 1) determination of primary/secondary thermoclines is not objective and the selection of profiles requiring smoothing would demand further arbitrary thresholds (instead this topic deserves a separate study to investigate the full implications of different approaches), 2) in this study we aimed to estimate variability in epilimnion depth estimates between *common and existing* methods (rather than introduce new methods), which highlighted the issue of micro-stratifications as enhancing variability, 3) smoothing of these micro-stratifications may not be suitable for some applications, particularly where it is assumed that the epilimnion is isothermal and well-mixed**.** However we do fundamentally agree that this is an important issue and would therefore intend to make the following additions/edits:

L.388: '*The concept of the epilimnion, and more widely, the three-layered structure of a stratified lake, is fundamental in limnology. Yet, despite the ubiquity of the term, there is no objective or generic approach for defining the epilimnion and a diverse number of approaches prevail in the literature. In a comprehensive analysis of high-frequency, multi-year data from two lakes, this study has highlighted the extent to which common water temperature profile based epilimnion depth estimates differ. The level of variability in epilimnion depth estimates calculated using common methods and threshold values, was exceedingly high. This result calls into question the practice of arbitrary method selection and comparing findings between studies which use different methods or even just different thresholds. The magnitude of variability also casts ambiguity on the calculation of key biogeochemical and ecological processes in a lake that rest on the assumption that the layers of a lake are well defined, including calculations of metabolic rates, and oxygen fluxes (e.g. Coloso et al., 2008, Foley et al., 2012, Obrador et al., 2014, Winslow et al., 2016).*

In an idealised stratified profile, the epilimnion is portrayed as near-uniform in water temperature or density and clearly delineated from a well-defined metalimnion. However, many measured profiles, at least within this study, did not conform to this idealised three-layered structure. Instead the water columns were often more complex, including multiple pycnoclines and near-surface micro-stratification layers, or the boundaries between the epilimnion/metalimnion were blurred. One approach to this issue is to filter out appropriate water column profiles or apply functions that coerce the profile into the expected structure (Read et al., 2011, Pujoni et al., 2019, Gray et al., 2020). Filters, additional conditions or smoothing functions, however, may suffer from many of the same challenges as the estimation of the epilimnion depth, since they attempt to discretise data based on arbitrary criteria (Kraemer et al., 2020). For example, our analysis of temporally high resolution time series data emphasised that rather than jumping from states, such as stratified or isothermal, changes in the water column occurred over an evolving continuum and often fluctuated between states. Similarly, the distinction between additional layers, such as the primary or secondary pycnocline, is fraught with the same issues of arbitrariness as discussed (Read et al., 2011). This study demonstrates that when epilimnion depth estimation methods, which are theorised for a three-layered water column, are applied to non-conforming water columns, they diverge widely on the location of the epilimnion depth, and at times, may not even be underpinned by the same theoretical assumptions. Since none of these methods can be considered the 'true' definition of the epilimnion depth, it is necessary to understand the degree to which methods differ. Improved understanding of their systematic differences will facilitate the use of methods that appropriately capture different processes, such as, air-water exchanges, thermocline entrainment or suspension of materials. Due to the realised complexities of observed and aggregated profile data, we may benefit from new approaches to water column *discretisation that incorporate the vast proportion of profiles which do not conform neatly to the three-layered paradigm.'*

5. Referee comment: Just to make future studies more concise, it would be better to investigate lakes whose monitoring programs did include vertical temperature, density and velocity profiles, e.g. estimating eddy diffusivities by ADCP. The addition of GOTM to this study to estimate turbulence seems a bit half-hearted as the calibration-validation was not described nor any figure showing the M5 results included in the manuscript.

Author response: While there would be some advantages to analyse lakes with vertical eddy diffusivity data, vertical eddy diffusivity measurements are fairly rare and often taken only for short periods of time, or only taken at specific depths in the lake. Therefore, it is likely that epilimnion depth estimates will continue to be derived using water temperature profile data, unless sensor technology evolves dramatically. To address the comments on the modelled turbulence section we have added some figures to supplementary (see Comment #32) as well as additional information on the calibration-validation (see Comment #15), and edited the relevant paragraph in the discussion (Comment #11).

6. Referee comment: L15-16: I'd recommend dropping the quotation marks around epilimnion and metalimnion in the abstract as it's a bit confusing to the reader. Further, as the first lines discuss the three-layered structure, I'd recommend shortly mentioning the hypolimnion here.

Author response: Agreed

7. Referee comment: L23: If needed you could also exchange 'approaches' and 'methods' with 'algorithms' throughout the manuscript.

Author response: Agreed.

8. Referee comment: L28: The phrase 'complex water column structures' is a bit confusing here. Are you referring to cases when the three-layer paradigm is violated? Maybe rephrase to complex thermal water column structure?

Author response: Agreed.

9. Referee comment: L35: The phrase 'less problematic' is quite vague, do you mean 'introduces less bias'?

Author response: Good point, we would like to simplify the sentence to;

L35. '*While there is no prescribed rationale for selecting a particular method, the method which defined the epilimnion depth as the shallowest depth where the density was 0.1 kg m^{-3} more than the surface density, may be particularly useful as a generic method.*'

10. Referee comment: L40: What do you mean by 'rapid gradients'? I'd recommend 'steep gradients'

Author response: Agreed.

11. Referee comment: L48-50: I'd recommend stressing these important statements more throughout the discussion at the end

Author response: Agreed. Please see the small edit below, in addition to updated discussion paragraph on L453.

L.48: '*The discretisation of these layers, however, is understood to be essentially theoretical, since micro-profile studies show that the conditions within layers are not uniform and exact cut-offs between layers do not necessarily exist (Imberger, 1985, Jonas et al., 2003, Tedford 2014, Kraemer et al., 2020).The definition of the epilimnion depth is thus inherently subjective, but has profound importance in limnology.'*

L453:*'Regardless of the method selected, however, all water temperature/density based methods are limited in their ability to indicate actual mixing processes. Our results using the lake modelled turbulence data demonstrated that even in a modelled environment, epilimnion depth estimates were inconsistent between the different methods and threshold values studied, and that turbulence based methods generally resulted in a shallower epilimnion depth estimate. These findings highlight the important but subtle difference between the layer detected by water density profiles (i.e. has been recently well-mixed and therefore has little resistance to further mixing due to the lack of density gradients), and the layer that is actively mixing, determined only through directly measured turbulence (Gray et al., 2020). Similarly, micro-profiling studies have shown that the actively mixing layer can be substantially shallower than the layer determined through water temperature profile data (McIntyre et al., 1993, Tedford et al., 2014). Micro-profile studies also demonstrate that within seemingly uniform layers there are micro-stratification layers, delineated by temperature differences as small as 0.02°C (Imberger, 1985, Shay and Gregg, 1986, MacIntyre, 1993; Jonas et al. 2003), which can be sufficient to isolate intermediate layers from atmospheric wind shear and cooling (Pernica et al., 2014). Although our results are not directly indicative of measured data, they demonstrate how even turbulence based methods are inherently arbitrary, as there is no objective threshold value (Monismith and Macintyre, 2009). Many of the ecological applications of the epilimnion depth have the underlying assumption that enough mixing is occurring in the epilimnion to keep the relevant organisms or particles suspended within the layer. Whether mixing is actually occurring, however, and to what extent, is not directly described by epilimnion depth estimations derived using water temperature or density profile data, and in fact, previous studies have found water density estimates of the epilimnion depth to be relatively poor indicators for the homogeneity of other ecological variables (Gray et al., 2020).'*

12. Referee comment: L74-75: Seems like sentencesL48-50 already explained that in reality there are no exact cut-offs, so how could there even be a consistent method used throughout limnology?

Author response: Interpretation of multi-lake studies or comparisons between separate studies, does require either consistent methods or an understanding of how/why the differences in methods will influence the results. We agree however that there cannot be an objective definition and therefore will change L74-75 to;

L.74: Despite the ubiquity of the epilimnion depth, there is no consistent method used in limnology.

13. Referee comment: L86: Is the vertical turbulence profile referring to a profile of turbulent eddy diffusivities? Here, the authors could also discuss field methods which measure turbulence in lakes, e.g. through velocity loggers. Or methods estimating diffusivities from water temperature profiles, e.g. gradient flux method by Heinz et al (1990) or heat budget method by Jassby and Powell (1975)

Author response: Yes, please see L86 rephrased below. Unfortunately, estimating vertical diffusivity from water temperature profiles as shown in Heinz et al., (1990) and Jassby and Powell (1975) can only estimate diffusivity below the epilimnion and below the photic zone. In addition, they can only estimate diffusivity over large temporal aggregates, depending on the accuracy of temperature sensors, and are also dependent on the (usually unknown) flux of heat with the sediment. Please also seeComment #11 for edits to discussion relating to turbulence.

L85: 'Compared with long-term water temperature datasets, there are relatively few turbulent eddy diffusivity measurements in lakes, typically using micro-profiling methods conducted over over a small time period (e.g. Imberger, 1985, Tedford, 2014). Other methods of estimating vertical eddy diffusivity, from water temperature data for example, as in the Jassby and Powell (1975) heat-flux method are restricted to use below the epilimnion and photic zone. Therefore, epilimnion depth definitions based on actual turbulence measurements are uncommon. Vertical turbulence profiles, however, as well as water temperature profiles, are estimated by some hydrodynamic lake models (Goudsmit et al., 2002, Dong et al., 2019). Such modelled data, therefore, offers a tool for assessing commonly used water temperature/density based methods in comparison to turbulence based methods. '

14. Referee comment: L116: I'd recommend moving the sentence "The lakes differ in many characteristics, [. . .]" to the beginning of the paragraph

Author response: Agreed.

15. Referee comment: 2.3 Simulated data: Tab. 1 suggests that the model was calibrated and validated. Which time periods were used? In this paragraph, the investigated fit criteria should also be mentioned. Also, in line 153, was the parameter of the minimum turbulent kinetic energy calibrated or, alternatively, which parameters affecting the min. turbulent kinetic energy were calibrated? This sentence is rather unclear.

Author response: Good point. We propose to update Section 2.3 with additional information on the calibration and validation process, in the text, and an additional Table in supplementary providing the model parameters and calibrated values, see below. The wnd_factor parameter is particularly important for the amount of turbulent kinetic energy available for mixing (Ayala et al. 2020).

L147 – 154: '*The Global Ocean Turbulence Model (GOTM), adapted for use in lakes, simulates small-scale turbulence and vertical mixing (Burchard et al,. 1999, Sachse et al., 2014, Moras et al., 2019, Ayala et al., in review) and was used to simulate daily profiles of water temperature (°C) and vertical eddy diffusivity (m-2 s -1) for Lake Erken and Lough Feeagh. A period of 4 years (1year spin-up followed by 3 years of calibration) was selected for calibration of GOTM, 2006-2009 for Lake Erken and 2008-2011 for Lough Feeagh. The model parameters that were calibrated were surface heat-flux factor (shf_factor), short-wave radiation factor (swr_factor), wind factor (wind_factor), minimum turbulent kinetic energy (k_min) and e-folding depth for visible fraction of light (g2) (See Apendix. Table S1.).'*

Table S1. Lake model parameters and calibrated values.

Parameter	Lake Erken	Lough Feeagh
Shf_factor	0.88	0.77
Swr_factor	0.98	0.93
Wind_factor	1.41	1.31
K_min	1.86e-6	3.48e-6
G2	2.14	0.56

'The validation period was 7 years (2010-2016) for Lake Erken and 4 years (2012-2015) for Lough Feeagh. For both calibration and validation, daily mean water temperatures were simulated when GOTM was forced using measured mean hourly. Model simulated profiles of mean daily water temperature were then compared to mean daily measured water temperature. During calibration the model was run approximately 5000 times to obtain a stable solution. Model performance was evaluated by comparing mean daily modelled and measured temperature profiles, the model efficiency coefficients used were percent relative error (PRE), root mean squared error (RMSE) and Nash-Sutcliffe efficiency (NSE) (Nash and Sutcliffe, 1970) (Table 2).'

Table 1. Lake model performance evaluation, showing the percentage relative error (%), root mean squared error (°C), and Nash Sutcliffe efficiency, for Lough Feeagh (profiles = 1016, years = 5) and Lake Erken (profiles = 1449, years = 7).

Statistic	Lough Feeagh		Lake Erken	
	Calibration	Validation	Calibration	Validation
PRE (%)	-0.48	0.47	-1.85	1.36
RMSE (°C)	0.67	1.18	0.53	0.55
NSE	0.97	0.92	0.98	0.97

16. Referee comment: 2.5 Analysis methods: This paragraph would benefit from sub-headings to make it easier for the reader to follow (similar structure as the Results would be beneficial).

Author response: We agree and can follow the same structure as the results.

17. Referee comment: L208: Here the colors are mentioned but the figure is not referenced. If you do not want to cross-reference the figure yet, maybe just write about 'color-coding'.

Author response: Agreed.

18. Referee comment: L232: 'logical and numerical schemes' is unclear to me. What do you mean by logical scheme here?

Author response: Good point, this is over-complicated language. We rephrase to:

L.231 'depending on the calculations used, may regularly encroach on the metalimnion (Lorbacher et al., 2006).'

19. Referee comment: L248: Is there a reason why the sensor deployment sensitivity was not tested on monitored data (e.g. by removing some loggers)?

Author response: Yes, we did that way also. However since Lough Feeagh was already quite coarse in vertical resolution, it did not seem like a fair comparison between the two lakes. The way we did it allowed us to see how the coarse resolution in Lough Feeagh could influence comparison between Erken, thus demonstrating that comparing lakes with different resolution will not be an equal comparison.

20. Referee comment: 3.2 Comparison between water density based methods: In my opinion, this paragraph is a crucial finding of the study as it discusses how deviations from the three-layer paradigm affected the results of the algorithms. Can you state how many profiles/observations points were either a) (classical paradigm), b) (multiple pycnoclines) or c) (weakly stratified)? Were most profiles following the three-layer structure or is the complex case b) dominating?

Author response: The difficulty with doing this is that there is no established method for classifying multiple pycnoclines or weak stratification. Thus, the results would be dependent on whichever arbitrary thresholds we used, and it is this dependence on arbitrary thresholds that we are highlighting in the study. Hence we do not think it would be appropriate within this paper. The topic is an interesting one though, and we think, worthy of a future study. However, we will dedicate a paragraph in the discussion to this consideration, which also addresses Comment #4 and Comment #11.

21. Referee comment: L301: Here it would be beneficial for the reader to state the mean epilimnion depth per lake plus the variance and the quantiles.

Author response: Agreed, see additional line below including the standard error and the inter-quartile range.

L301: '*The mean epilimnion depth estimate for all observed data, calculated with methods M1-M4 and all thresholds was -28.1 m (SEM =0.6 m, inter-quartile range = 19.0 m) for Lough Feeagh and -11.0 m (SEM = 0.1 and IQR = 2.3 m) for Lake Erken.'*

22. Referee comment: L330: Shouldn't it be '[. . .] epilimnion depth was identified at a depth above [. . .]' instead of "greater"? As everything is referenced to the surface, wouldn't greater mean deeper?

Author response: Agreed, thank you.

23. Referee comment: L332: It seems there's a word missing here

Author response: Agreed,

L332: '*For M4, the percentage of stratified days remained static regardless of the threshold value, because the epilimnion depth was detected for all profiles where the water column temperature range was more than 1 °C, regardless of the threshold used*.'

24. Referee comment: L336: Do you mean 0.025 kg/m3/m instead of 0.25 kg/m3/m which would be higher than the maximum investigated threshold?

Author response:. Thank you, updated to 0.025 kg/m3/m

25. Referee comment: L360: The phrase "[. . .] M2 had typically a higher threshold range than M3 [. . .]" confuses me. Do you mean that the range between the thresholds was higher?

Author response: Yes, rephrased:

L360: '*M2 had typically a greater range in epilimnion depth estimates than M3, when calculated with the same range of thresholds.'*

26. Referee comment: 3.4 Sensitivity of epilimnion depth: Can you test if the differences between the means were significant?

Author response: We consider that assessing whether the differences between the epilimnion depth estimates derived using high/low vertical resolution water temperature datasets were statistically significant would not be appropriate. This is because the lower resolution water temperature dataset is nested within the higher resolution dataset, and therefore the estimated MLD values based on these data would not be independent. Moreover, we know that these two datasets do differ, as we have induced that difference intentionally by filtering the higher resolution modelled data to obtain the lower resolution dataset. Our aim here was to quantify the change in epilimnion depth estimates for each method, accounting for when data might be available at high resolution intervals (our modelled 0.5 m data) or available using a deployment strategy similar to that currently employed in Lough Feeagh (mean of one sensor per 3 m). The result demonstrated how M1 in particular tended to be less sensitive to the vertical resolution of the input data than other methods, since the differences between the mean high and lower resolution data derived epilimnion depth estimates were smallest. In order to make the spread of the data more apparent to the reader, we have added the standard error for each of the means in Table 4 (see below). This will give a measure of the variability around the mean and allow a more informed assessment of the difference between the derived MLD values. For consistency, the standard error will also be added to Table 2, the table showing the observed results.

Table 4. Mean Apr-Oct epilimnion depth estimates (m) derived using high resolution and low resolution modelled water temperature data with standard error of the mean in brackets, and the difference calculated between the high resolution and low resolution estimate (m), for Lough Feeagh and Lake Erken.

	Lough Feeagh			**Lake Erken**		
Metho d	High resolutio n mean (m)	Low resolutio n mean (m)	Difference in mean epilimnio n depth (m)	High resolutio n mean (m)	Low resolutio n mean (m)	Difference in mean epilimnio n depth (m)
M1	-22.1 (0.7)	-22.2 (0.7)	0.1	-10.9 (0.1)	-10.9 (0.1)	0.0
M2	-31.7 (1.1)	-34.9 (1.2)	3.2	-11.9 (0.2)	-13.1 (0.2)	1.2
M3	-32.0 (1.1)	-35.2 (1.1)	3.2	-12.1 (0.2)	-13.1 (0.2)	1.0
M4	-22.1 (0.6)	-22.6 (0.6)	0.5	-11.3 (0.1)	-11.5 (0.1)	0.2

Table 2. Summary of statistics for each method, showing the mean (m) and standard error of mean in brackets, minimum (i.e. shallowest estimate) (m), maximum (i.e. deepest estimate) (m) and the interquartile range (m) of the April-October epilimnion depth estimates (summarised from the results shown in Fig.6a), and the mean Pearson's correlation coefficient (r) for each method, representing the mean correlation for all possible combinations between threshold values., for Lough Feeagh and Lake Erken.

	Lough Feeagh					**Lake Erken**				
Method	Mean (m)	Min (m)	Max (m)	IQR (m)	r	Mean (m)	Min (m)	Max (m)	IQR (m)	r
M1	-19.0 (0.8)	-4.6	-25.4	7.3	0.77	-10.0 (0.2)	-7.8	-11.2	1.3	0.92
M2	-35.9 (0.9)	-19.7	-41.4	6.5	0.48	-11.3 (0.2)	-8.4	-12.9	1.7	0.78
M3	-36.5 (0.8)	-22.4	-41.5	6.1	0.49	-11.8 (0.2)	-10.0	-13.0	1.2	0.82
M4	-21.1 (0.5)	-19.7	-21.5	0.3	1.00	-11.9 (0.1)	-10.1	-11.3	0.5	0.99

27. Referee comment: L405: The phrase 'particularly distinct systematic difference' seems a bit excessive

Author response: Good point, rephrased to;

L405: *'large systematic difference'*

28. Referee comment: L406: Is '[. . .] was equivalent to using different threshold values [. . .]' referring to density threshold values?

Author response: Agreed, rephrased to;

L406:*'Due to the non-linear relationship between water density and temperature, the use of water temperature was equivalent to using different density threshold values throughout the year, resulting in a distinct shift in the stratification period.'*

29. Referee comment: L424: Do you actually mean 'acceleration of epilimnion deepening' instead of shallowing here? Shouldn't the epilimnion deepen relative to the surface height during stratification onset?

Author response: Good point, this is not clear, we have rephrased to,

L.423 *'Alternatively, water temperature-based estimates typically resulted in earlier stratification, which could indicate a longer duration of phytoplankton in a shallower epilimnion.'*

30. Referee comment: L429: But these methods are detecting a layer that is specifically not isothermal relative to themselves, right?

Author response: Correct. To make this clearer we have rephrased.

L.429, *'An important difference was also found between 1) methods detecting the layer that is isothermal relative to the surface and 2) methods detecting the point that is isothermal relative to the steep gradients of the metalimnion.'*

31. Referee comment: L448: The phrase '[. . .] not be suitable for use with water density metric [. . .]' is unclear to me. Which water density metric are you talking about?

Author response: We are pointing out that use of water density metrics (e.g. rLakeanalyzer meta.tops) with a water temperature based stratification definitions (e.g. <1°C) is incompatible, and could lead to a seasonal bias. This is based on the results shown in Figure 3. We have rephrased this sentence as below.

L448. 'The results suggest that use of water density metrics, such as epilimnion depth estimates, in combination with traditional water temperature based definitions of stratification, are incompatible, given the non-linear relationship between temperature and density'

32. Referee comment: L453-466: A figure showing these results would be beneficial for the reader, or how M5 compares to the other models.

Author response: We propose to add a time series figure to the Supplementary Material showing the profile-based methods (M1-M4) and the turbulence method (M5), using all thresholds, for one year of data, as below.

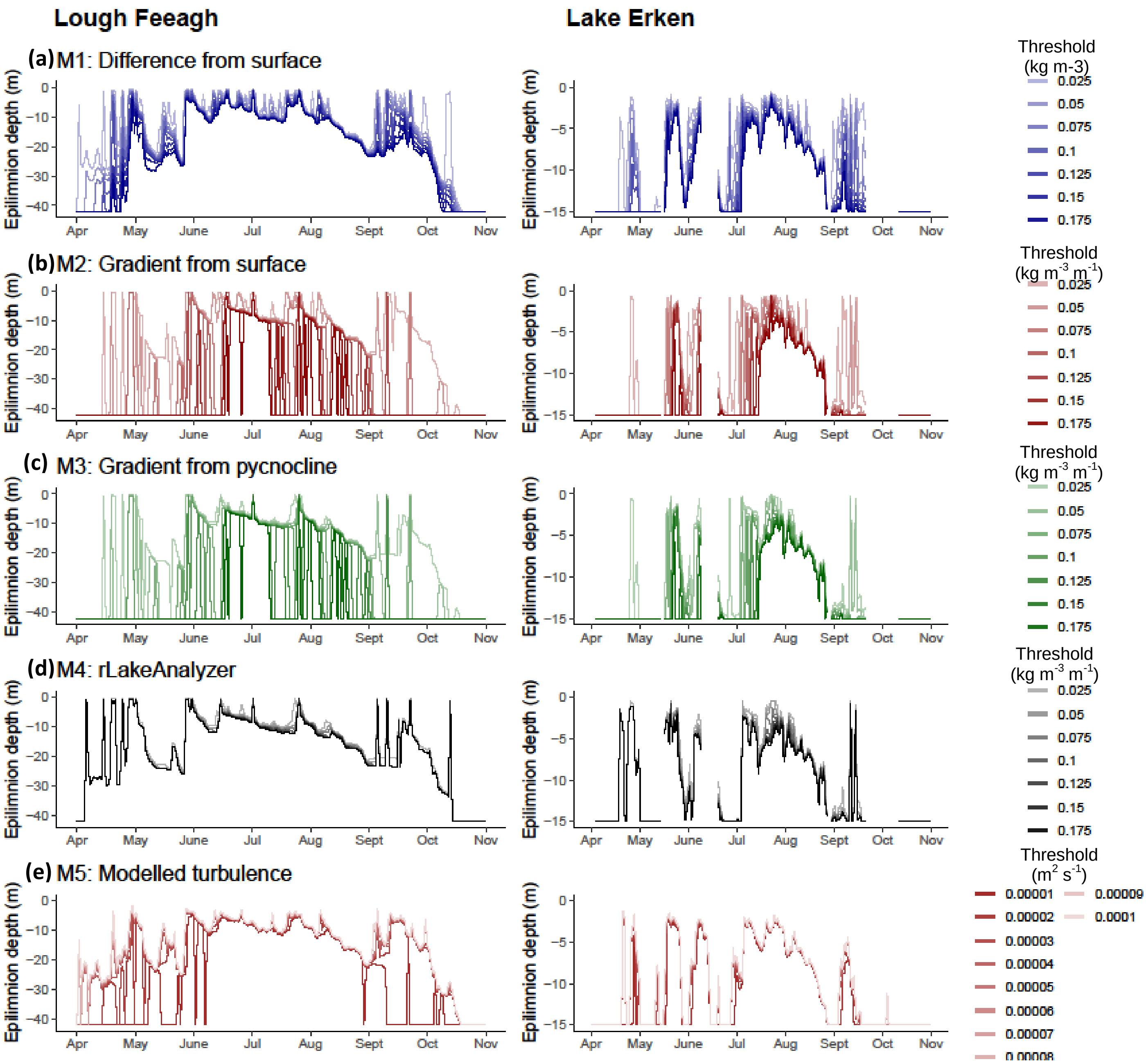


Figure S1: Daily epilimnion depth estimates using modelled data for 2016 from Lough Feeagh and Lake Erken, showing estimates from all profile based epilimnion depth methods, including M1, the absolute difference from the surface method **(a)**, M2, the gradient from the surface method **(b)**, M3, the gradient from the pycnocline method **(c)** and M4, the rLakeAnalyzer method **(d)**, as well as M5, the modelled turbulence based method **(e)**, calculated using the full range of thresholds, and for each lake.

33. Referee comment: L465: Are profile data here referring to water temperature profile data?

Author response: Yes, will be changed to read 'water temperature profile data' explicitly.

34. Referee comment: L515: I'd recommend not to write about 'problematic' here, maybe 'less bias', 'conservative assumption'?

Author response: L.515 '*less sensitive to vertical sensor resolution than the other methods*'

35. Referee comment: Table 3: Could you exchange the table with either a correlation matrix or correlation plot?

Author response: We propose to add a correlation matrix for each lake in the supplementary material (see Tables below) since they are very large tables, while keeping Table 3. as it is a useful summary for the main text.

Table S2: Correlation matrix between all methods and all threshold combinations for Lough Feeagh and Lake Erken.

Lough Feeagh

		Method 1								Method 2								Method 3								Method 4							
		0.03	0.05	0.08	0.1	0.13	0.15	0.18	0.2	0.03	0.05	0.08	0.1	0.13	0.15	0.18	0.2	0.03	0.05	0.08	0.1	0.13	0.15	0.18	0.2	0.03	0.05	0.08	0.1	0.13	0.15	0.18	0.2
Method 1	0.025																																
	0.05	.59																															
	0.075	.46	.82																														
	0.1	.42	.76	.95																													
	0.125	.41	.73	.91	.98																												
	0.15	.39	.71	.89	.95	.99																											
	0.175	.38	.69	.86	.93	.97	.99																										
	0.2	.37	.67	.85	.91	.95	.97	.99																									
Method 2	0.025	.41	.68	.73	.74	.76	.78	.79	.79																								
	0.05	.26	.48	.61	.65	.68	.70	.72	.74	.64																							
	0.075	.16	.30	.39	.43	.46	.49	.51	.53	.40	.67																						
	0.1	.11	.22	.29	.33	.36	.38	.40	.42	.30	.51	.77																					
	0.125	.08	.16	.22	.25	.27	.29	.30	.32	.22	.38	.58	.75																				
	0.15	.07	.13	.17	.21	.23	.24	.25	.26	.18	.31	.46	.60	.79																			
	0.175	.06	.11	.15	.17	.19	.21	.22	.23	.15	.26	.38	.49	.66	.83																		
	0.2	.05	.09	.13	.15	.17	.18	.19	.19	.13	.22	.33	.42	.56	.71	.86																	
Method 3	0.025	.39	.64	.68	.71	.74	.77	.78	.79	.94	.61	.43	.34	.26	.22	.19	.16																
	0.05	.25	.46	.58	.63	.66	.68	.70	.72	.61	.95	.71	.55	.42	.34	.28	.24	.61															
	0.075	.15	.28	.37	.42	.45	.48	.50	.52	.38	.64	.97	.80	.60	.48	.40	.34	.43	.69														
	0.1	.11	.21	.27	.31	.34	.36	.38	.40	.29	.49	.73	.95	.79	.63	.53	.45	.33	.53	.76													
	0.125	.08	.15	.20	.23	.26	.27	.29	.30	.21	.36	.54	.70	.94	.85	.71	.61	.25	.39	.56	.74												
	0.15	.07	.12	.16	.20	.21	.23	.24	.25	.17	.29	.43	.56	.75	.94	.88	.75	.21	.32	.45	.60	.81											
	0.175	.06	.11	.14	.17	.18	.20	.21	.22	.14	.24	.36	.47	.63	.79	.95	.90	.18	.27	.38	.50	.68	.84										
	0.2	.05	.09	.12	.14	.16	.17	.18	.18	.12	.21	.31	.40	.54	.68	.81	.95	.15	.23	.32	.43	.57	.71	.85									
Method 4	0.025	.40	.70	.82	.81	.78	.75	.72	.70	.58	.50	.36	.29	.22	.19	.17	.15	.62	.50	.36	.28	.22	.19	.16	.14								
	0.05	.39	.68	.79	.78	.75	.72	.69	.68	.55	.48	.36	.29	.23	.20	.18	.15	.59	.49	.36	.29	.23	.20	.17	.15	1.00							
	0.075	.39	.67	.78	.77	.74	.71	.68	.66	.53	.46	.34	.29	.23	.21	.18	.16	.58	.47	.35	.28	.23	.20	.17	.15	.99	1.00						
	0.1	.39	.67	.78	.77	.74	.70	.67	.65	.53	.45	.33	.27	.23	.21	.18	.16	.58	.46	.33	.27	.22	.20	.17	.15	.99	1.00	1.00					
	0.125	.39	.67	.78	.77	.73	.70	.67	.65	.53	.44	.32	.26	.22	.20	.18	.15	.58	.45	.32	.26	.22	.19	.17	.15	.99	1.00	1.00	1.00				
	0.15	.39	.67	.78	.77	.73	.70	.67	.65	.53	.44	.32	.26	.22	.20	.18	.15	.58	.45	.32	.25	.21	.19	.17	.15	.99	1.00	1.00	1.00	1.00			
	0.175	.39	.67	.78	.77	.73	.70	.67	.65	.53	.44	.31	.26	.21	.19	.17	.15	.58	.45	.32	.25	.21	.19	.17	.14	.99	1.00	1.00	1.00	1.00	1.00		
	0.2	.39	.67	.78	.77	.73	.70	.67	.65	.53	.44	.31	.26	.21	.19	.17	.15	.58	.45	.32	.25	.21	.18	.16	.14	.99	1.00	1.00	1.00	1.00	1.00	1.00	

Lake Erken

		Method 1								Method 2								Method 3								Method 4							
		0.03	0.05	0.08	0.1	0.13	0.15	0.18	0.2	0.03	0.05	0.08	0.1	0.13	0.15	0.18	0.2	0.03	0.05	0.08	0.1	0.13	0.15	0.18	0.2	0.03	0.05	0.08	0.1	0.13	0.15	0.18	0.2
Method 1	0.025																																
	0.05	.94																															
	0.075	.89	.97																														
	0.1	.85	.94	.98																													
	0.125	.82	.91	.96	.99																												
	0.15	.79	.89	.94	.97	.99																											
	0.175	.77	.87	.92	.96	.98	1.00																										
	0.2	.75	.85	.90	.94	.97	.99	1.00																									
Method 2	0.025	.85	.92	.92	.89	.87	.85	.83	.81																								
	0.05	.78	.87	.92	.94	.95	.95	.95	.94	.85																							
	0.075	.70	.78	.84	.87	.90	.92	.92	.93	.75	.90																						
	0.1	.64	.73	.78	.82	.85	.87	.88	.89	.70	.84	.94																					
	0.125	.59	.67	.72	.75	.78	.81	.83	.85	.64	.77	.86	.93																				
	0.15	.55	.62	.66	.70	.73	.76	.78	.80	.59	.71	.80	.87	.94																			
	0.175	.50	.57	.61	.64	.67	.69	.72	.74	.54	.65	.74	.80	.87	.93																		
	0.2	.45	.52	.55	.58	.61	.63	.66	.68	.49	.59	.67	.73	.80	.86	.93																	
Method 3	0.025	.75	.83	.86	.87	.88	.88	.88	.88	.84	.86	.82	.79	.73	.69	.64	.58																
	0.05	.72	.81	.86	.89	.91	.92	.93	.93	.78	.92	.89	.86	.81	.77	.72	.66	.91															
	0.075	.66	.74	.79	.83	.86	.88	.90	.91	.71	.85	.94	.92	.88	.84	.78	.72	.83	.92														
	0.1	.62	.70	.75	.78	.82	.84	.86	.87	.67	.80	.90	.96	.92	.88	.82	.77	.79	.88	.95													
	0.125	.58	.65	.70	.73	.76	.79	.81	.83	.63	.75	.84	.90	.97	.94	.88	.82	.73	.82	.89	.94												
	0.15	.53	.61	.65	.68	.71	.73	.76	.78	.58	.69	.78	.84	.91	.97	.93	.87	.68	.76	.84	.88	.94											
	0.175	.49	.56	.59	.63	.65	.68	.70	.72	.53	.64	.72	.78	.85	.91	.98	.94	.63	.71	.78	.82	.88	.93										
	0.2	.45	.51	.54	.57	.60	.62	.64	.66	.48	.58	.66	.72	.78	.84	.91	.98	.57	.65	.72	.76	.81	.87	.93									
Method 4	0.025	.78	.83	.86	.88	.88	.88	.88	.88	.77	.84	.82	.79	.74	.70	.65	.59	.90	.90	.84	.80	.74	.69	.64	.58								
	0.05	.76	.81	.83	.85	.86	.86	.86	.86	.74	.82	.81	.79	.75	.71	.66	.61	.88	.90	.85	.81	.76	.71	.65	.60	.99							
	0.075	.74	.79	.82	.83	.84	.84	.84	.84	.72	.80	.79	.78	.74	.71	.66	.62	.86	.88	.85	.81	.76	.71	.66	.61	.98	.99						
	0.1	.73	.78	.81	.82	.83	.83	.83	.83	.71	.79	.78	.76	.73	.70	.66	.62	.85	.87	.83	.80	.75	.71	.66	.61	.97	.99	1.00					
	0.125	.73	.77	.80	.81	.82	.82	.82	.82	.70	.78	.77	.75	.72	.70	.66	.61	.84	.86	.82	.79	.74	.70	.66	.61	.96	.98	.99	1.00				
	0.15	.72	.77	.79	.81	.81	.81	.81	.81	.69	.77	.75	.74	.71	.69	.65	.61	.84	.86	.81	.78	.73	.69	.65	.61	.96	.98	.99	1.00	1.00			
	0.175	.72	.76	.79	.80	.81	.81	.81	.80	.69	.76	.75	.73	.70	.67	.64	.60	.83	.85	.80	.77	.72	.68	.64	.60	.95	.98	.99	.99	1.00	1.00		
	0.2	.71	.76	.78	.80	.80	.80	.80	.80	.69	.76	.74	.72	.69	.67	.63	.59	.83	.85	.80	.76	.71	.67	.63	.59	.95	.98	.99	.99	1.00	1.00	1.00	

36. Referee comment: Fig. 2 is great, I like it a lot! It makes the whole study easier to understand.

Author response: Thanks!

37. Referee comment: Fig 3.: Why is there a thin blue shaded area below the red shade in Lake Erken?

Author response: That is a graphical issue and will be fixed in the revised version.

Additional References

Heinz, G., Ilmberger, J. and Schimmele, M., 1990. Vertical mixing in Überlinger See, western part of Lake Constance. Aquatic sciences, 52(3), pp.256-268.

Jassby, A. and Powell, T., 1975. Vertical patterns of eddy diffusion during stratification in Castle Lake, California 1. Limnology and oceanography, 20(4), pp.530-543.

Marcé, R., George, G., Buscarinu, P., Deidda, M., Dunalska, J., de Eyto, E., Flaim, G., Grossart, H.P., Istvanovics, V., Lenhardt, M. and Moreno-Ostos, E., 2016. Automatic high frequency monitoring for improved lake and reservoir management. Environmental Science & Technology, 50(20), pp.10780-10794.

Pernica, P., Wells, M.G. and MacIntyre, S., 2014. Persistent weak thermal stratification inhibits mixing in the epilimnion of north-temperate Lake Opeongo, Canada. *Aquatic sciences*, *76*(2), pp.187-201.

Shay, T.J. and Gregg, M.C., 1986. Convectively driven turbulent mixing in the upper ocean. *Journal of Physical Oceanography*, *16*(11), pp.1777-1798.

Tedford, E.W., MacIntyre, S., Miller, S.D. and Czikowsky, M.J., 2014. Similarity scaling of turbulence in a temperate lake during fall cooling. *Journal of Geophysical Research: Oceans*, *119*(8), pp.4689-4713.

Variability in epilimnion depth estimations in lakes

Harriet L. Wilson[1], Ana I. Ayala[2], Ian D. Jones[3], Alec Rolston[4], Don Pierson[2], Elvira de Eyto[5], Hans-Peter Grossart[6], Marie-Elodie Perga[7], R. Iestyn Woolway[1], Eleanor Jennings[1]

[1]Center for Freshwater and Environmental Studies, Dundalk Institute of Technology, Dundalk, Ireland
[2]Department of Ecology and Genetics, Limnology, Uppsala University, Uppsala, Sweden
[3]Biological and Environmental Sciences, Faculty of Natural Sciences, University of Stirling, Stirling, UK
[4]An Fóram Uisce, National Water Forum, Ireland
[5]Marine Institute, Furnace, Newport, Co. Mayo, Ireland
[6]Institute for Biochemistry and Biology, Potsdam University, Potsdam, Germany
[7]University of Lausanne, Faculty of Geoscience and Environment, CH 1015 Lausanne, Switzerland

Correspondence to: Harriet L. Wilson (wilsonh@dkit.ie)

Abstract. The "epilimnion" is the surface layer of a lake typically characterised as well-mixed and is decoupled from the "metalimnion" due to a rapid steep change in density. The concept of the epilimnion, and more widely, the three-layered structure of a stratified lake, is fundamental in limnology and calculating the depth of the epilimnion is essential to understanding many physical and ecological lake processes. Despite the ubiquity of the term, however, there is no objective or generic approach for defining the epilimnion and a diverse number of approaches prevail in the literature. Given the increasing availability of water temperature and density profile data from lakes with a high spatio-temporal resolution, automated calculations, using such data, are particularly common, and have vast potential for use with evolving long-term, globally measured and modelled datasets. However, multi-site and multi-year studies, including those related to future climate impacts, require robust and automated approaches algorithms for epilimnion depth estimation. In this study, we undertook a comprehensive comparison of commonly used epilimnion depth estimation methods, using a combined 17 year dataset, with over 4700 daily temperature profiles from two European lakes. Overall, we found a very large degree of variability in the estimated epilimnion depth across all methods and thresholds investigated and for both lakes. These differences, manifest over high-frequency data, led to fundamentally different understandings of the epilimnion depth. In addition, estimations of the epilimnion depth were highly sensitive to small changes in the threshold value, complex thermal water column structures and vertical data resolution. These results call into question the custom of arbitrary method selection, and the potential problems this may cause for studies interested in estimating the ecological processes occurring within the epilimnion, multi-lake comparisons or long-term time series analysis. We also identified important systematic differences between methods, which demonstrated how and why methods diverged. These results may provide rationale for future studies to select an appropriate epilimnion definition in light of their particular purpose and with awareness of the limitations of individual methods. While there is no prescribed rationale for selecting a particular method, the method which defined the epilimnion depth as the shallowest depth where the density was 0.1 kg m^{-3} more than the surface density, was shown to be overall less problematic than the other methods. may be particularly useful as a generic method.

1 Introduction

The "epilimnion depth", "mixed layer" or "top of the metalimnion" are common terms in limnology, typically referring to the deepest point of the surface layer of a stratified lake, characterised as quasi-uniform in terms of physical and biogeochemical properties, and overlying a layer of rapid steep vertical gradients. Incoming heat to a lake, received at the lake surface, expands water above 3.98 °C resulting in density stratification. Convective cooling at the surface and mechanical energy injected by the wind, drive vertical mixing (Wüest and Lorke, 2003). These competing surface fluxes result in a warm, well mixed layer of water that interacts dynamically with the atmosphere (Monismith and MacIntyre, 2008). The vertical propagation of energy manifested at the lake surface is constrained by the ~~sharp~~steep density gradients in the metalimnion, which act to decouple the epilimnion from the deep hypolimnion. As such, it has become foundational in limnology to consider a stratified lake as consisting of three well-defined layers, a turbulent epilimnion (diffusivity typically 10^{-5} - 10^{-2} m^2 s^{-1}), the stable metalimnion (5 x 10^{-8} - 10^{-6} m^2 s^{-1}) and the quiescent hypolimnion (3 x 10^{-6} - 10^{-4} m^2 s^{-1}) (Wüest and Lorke, 2009). The discretisation of these layers, however, is understood to be essentially theoretical, since micro-profile studies show that the conditions within layers are not uniform and exact cut-offs between layers do not necessarily exist (Imberger, 1985, Jonas et al., 2003, Tedford 2014, Kraemer et al., 2020).The definition of the epilimnion depth is thus inherently subjective, but has profound importance in limnology.

Quantifying the vertical extent of the epilimnion is crucial for understanding many of the physical, chemical and biological processes in lakes. Although the epilimnion is differentiated from the typically shallower layer that is *actively* mixing (Gray et al., 2020), the depth of the epilimnion indicates the volume and properties of the water that is influenced by air-water interactions. It is therefore essential for interpreting the physical response of lakes to long-term atmospheric changes (Lorbacher et al., 2006, Persson and Jones, 2008, Flaim et al., 2016), and extreme climatic events (Jennings et al., 2012, Calderó-Pascual et al., ~~in review~~2020) and is even required for predicting the local climate for very large lakes (Thiery et al., 2015). The epilimnion depth is also critical for the estimation of algal light availability, nutrient fluxes and epilimnetic water temperatures, which determine photosynthesis rates and establishes the basis of the food web in a lake (MacIntyre, 1993, Diehl et al., 2002, Berger et al., 2006, Bouffard and Wüest, 2018). The depth of the epilimnion is also used for estimating the transfer of oxygen, received at the lake surface, to deeper layers, sustaining aerobic life and preventing anoxia (Foley et al., 2012, Schwefel et al., 2016).

The increasing availability of high-frequency measured and simulated data, coupled with collaborative networks of lake scientists, offers a huge potential for broadening our understanding of the epilimnion depth. Water temperature profile data collected at high frequency intervals on automatic monitoring buoys in lakes are becoming increasingly available (Jennings et al., 2012, de Eyto et al., 2016, Marcé et al., 2016). In addition, the collation of these datasets globally through collaborative initiatives such as GLEON (http://gleon.org/) and NETLAKE (https://www.dkit.ie/netlake) (Weathers et al., 2013, Jennings et al., 2017), and modelling initiatives such as ISIMIP2b (Ayala et al., ~~in review~~2020), broadens the potential for long-term and multi-lake studies. However, these datasets also introduce new challenges for estimating metrics such as the epilimnion depth. Such large quantities of big data can limit users' capacity to examine individual profiles and therefore require robust, automated ~~approaches~~ algorithms with low computational expense (Read et al., 2011, Pujoni et al., 2019).

Despite the ubiquity of the epilimnion depth, there is no ~~objective or~~ consistent method used in limnology. The epilimnion depth can be defined in terms of many variables (e.g. water temperature, water density, turbulence estimations, surface fluxes, biogeochemical properties), represent different temporal scales of variability (e.g. inter-annual to sub-daily), and be calculated using a range of numerical approaches (e.g. sigmoidal functions, threshold ~~approaches~~algorithms) (Brainard and Gregg, 1996, Thomson and Fine, 2003, Kara et al., 2003, De Boyer et al., 2004, Lorbacher et al., 2006, Gray et al., 2020). A particularly common approach in limnology, due to the availability of the required data, is to define the epilimnion using water temperature profile data. However, inconsistencies exist between studies which use water temperature (e.g. Zorzal-Almeida et al., 2017, Strock et al., 2017), or water density (e.g. Read et al., 2011, Obrador et al., 2014). Often the epilimnion depth is defined as the location where the change in water temperature or density exceeds a user-defined threshold. However, studies vary in the value selected which may be defined in absolute units (e.g. Andersen et al., 2017) or gradients between consecutive sensors (e.g. Lamont et al., 2004). A particularly prevalent method in recent studies is the 'meta.top' function proposed in R package 'rLakeAnalyzer' (Read et al., 2011). In contrast, ~~vertical turbulence data are not routinely measured in lakes, and therefore~~epilimnion depth definitions based on actual turbulence measurements are uncommon~~. However, vertical~~ Compared with long-term water temperature datasets, there is relatively few turbulent eddy diffusivity measurements in lakes, typically using micro-profiling methods conducted over a small time period (e.g. Imberger, 1985, Tedford, 2014). Other methods of estimating vertical eddy diffusivity, from water temperature data, for example, the Jassby and Powell (1975) heat-flux method, are restricted to use below the epilimnion and photic zone. Vertical turbulence profiles, however, as well as water temperature profiles, are estimated by some hydrodynamic lake models (Goudsmit et al., 2002, Dong et al., 2019). Such modelled data, therefore, offers a tool for assessing commonly used water temperature/density based methods in comparison to turbulence based methods.

The diversity of epilimnion depth definitions and arbitrary selection process, suggests that methods may be used interchangeably, and are relatively insensitive to the threshold value used. However, recent studies have begun to recognise large inconsistencies between different definitions and the potential problems this may cause, although so far, in limnology, analysis has been restricted to a small number of manual profiles (Gray et al., 2020) and a limited number of methods (Pujoni et al., 2019). ~~As such, a systematic analysis of common epilimnion depth methods for use with high-frequency data is required to assess the agreement among methods.~~ Although lower temporal resolution data is sufficient for investigating seasonal patterns, high-frequency data can be used to gain information on the level of day-to-day variability in epilimnion depth and demonstrates how methods perform over a continuum of water column conditions. In addition, through the vast number of measured profiles, high-frequency data offers a more robust comparison of methods, than previously demonstrated with manually collected datasets, and even when aggregated to the daily time-step is more representative of the sub-daily variability (Marcé et al., 2016). Given the potential of multi-lake comparison and longitudinal studies, methods are required to perform consistently across temporal and spatial ranges, rather than being tailored specifically to one lake or period of time. Therefore, the sensitivity of different methods to temporal and spatial characteristics, such as water column structure and vertical resolution of data measurements is essential for assessing which methods are most suitable for future analysis (Fee et al., 1996, Thomson and Fine, 2003, Lorbacher et al., 2006, Pujoni et al., 2019).

In this study, we undertook an in-depth comparison of methods commonly used for the estimation of epilimnion depth using high frequency, multi-year data for water temperature profiles, collected with automated monitoring buoys from two European lakes, Lough Feeagh (Ireland) and Lake Erken (Sweden). In addition to estimates based on these measured data, we used simulated data output from a lake model to compare water temperature and turbulence based methods, and to assess the influence of vertical sensor resolution. The objectives of this study were to: 1) compare water temperature and water density based estimates of the epilimnion depth, 2) compare a range of common methods and threshold values, 3) assess the sensitivity of individual methods to the threshold value, the water column structure, and the vertical sensor resolution, and 4) to compare profile based methods to turbulence derived estimates using lake modelled data.

2 Methods

2.1 Study sites

We used data from two European temperate lakes, Lough Feeagh (53°56'N, 9°34'E) in Ireland and Lake Erken (59°51'N, 18°36'E) in Sweden (Fig.1.). The lakes differ in many characteristics, including depth, surface area and sensor deployment resolution, providing an opportunity to assess method performance in different lake specific conditions. Lough Feeagh is located on the west coast of Ireland and is a cold monomictic, oligotrophic and humic lake with a surface area of 3.9 km^2, maximum depth of 45 m and ~~average~~mean depth of 14.5 m (de Eyto et al., 2016). Lake Erken is located in east central Sweden near the Baltic coast and is a dimictic, mesotrophic, clear lake with a surface area of 24 km^2, maximum depth of 21 m and ~~average~~mean depth of 9 m (Yang et al., 2016). ~~The lakes differ in many characteristics, including depth, surface area and sensor deployment resolution, providing an opportunity to assess method performance in different lake specific conditions.~~ In addition, Lake Erken has a ~~much larger average~~substantially greater mean summer top-bottom density gradient (0.056 kg m^{-3} m^{-1}) compared to Lough Feeagh (0.016 kg m^{-3} m^{-1}).

2.2 Measured data

In this study, we used a total of 4783 daily water temperature profiles from Lough Feeagh (n = 2778) and Lake Erken (n = 2005). Profiles were collected at high frequency intervals on moored automatic monitoring buoys, and from these the ~~average~~mean daily profiles were calculated. On Lough Feeagh, vertical water temperature measurements were collected every 2 minutes for the period 2004-2017 at depths 0.9, 2.5, 5, 8, 11, 14, 16, 18, 20, 22, 27, 32, 42 m using submerged platinum resistance thermometers (PRTs) (PT100 1/10DIN, Lab Facility, Bognor Regis, United Kingdom) (de Eyto et al., 2016, 2020). On Lake Erken, temperature profile data were collected at 1 min intervals at depths 0.5 m to 15 m at 0.5 m intervals, using Type T thermocouple sensors using a Campbell scientific AM416 multiplexer and CR10 ~~datalogger~~data logger (Pierson et al., 2011). The topmost sensor data was excluded to match the topmost sensor in Lough Feeagh. In Lake Erken, the monitoring buoy was manually deployed each year prior to or just after the onset of stratification to avoid damage from the seasonal ice cover, and therefore the number of observations varied annually. To ensure data were consistent for both lakes, data were subset from 1st April to 31st October. To address the issue of large data gaps, years where less than 70% of the data between April to October were available (>150 days) were excluded from the analysis. The remaining years were 2004, 2005, 2006, 2011, 2012, 2013, 2014, 2015, 2016, 2017 for Lough Feeagh and 2002, 2005, 2008, 2009, 2014, 2015, 2017 for Lake Erken. Water density (kg m^{-3}) was calculated from water temperature (°C) using rLakeAnalyzer (Read et al., 2011), with the Martin and McCutcheon (1999) equation, assuming negligible effects of soluble material.

Meteorological data were required to drive a physical hydrodynamic model (GOTM; Global Ocean Turbulence Mode, Burchard et al,. 1999), including wind speed (m s^{-1}), atmospheric pressure (hPa), air temperature (°C), relative humidity (%), cloud cover (dimensionless, 0-1), short-wave radiation (W m^{-2}) and precipitation (mm day^{-1}). For Lake Erken, air temperature, wind speed and short-wave radiation were collected from the Malma Island meteorological station on the lake, at 1 min intervals and averaged to 60 min intervals. Mean sea level pressure, relative humidity and precipitation were measured at the Svanberga meteorological station located 400 m from the lake shore, at 60 min intervals. Cloud cover was recorded from Svenska Högarna Station, 69 km south-east of Lake Erken. In Lough Feeagh, wind speed, air temperature and short-wave radiation, mean sea level pressure, relative humidity and precipitation were measured in the meteorological station next to the lake (de Eyto et al., 2020). Cloud cover was recorded at Knock Airport, 50 km east from Lough Feeagh.

2.3 Simulated data

The Global Ocean Turbulence Model (GOTM), adapted for use in lakes, simulates small-scale turbulence and vertical mixing (Burchard et al,. 1999, Sachse et al., 2014, Moras et al., 2019, Ayala et al., ~~in review). GOTM~~2020) and was used to simulate daily profiles of water temperature (°C) and vertical eddy diffusivity (m^{-2} s^{-1}) for ~~1016 days in Lough Feeagh (2012-2016) and 1449 days in Lake Erken (2010-2016). For these simulations, the turbulent kinetic energy-dissipation (k-ε)~~Lake Erken and Lough Feeagh. GOTM was calibrated using 4 years of data (2006-2009 for Lake Erken and 2008-2011 for Lough Feeagh), including 1 year spin-up followed by 3 years of calibration. The calibrated model ~~was used, in combination with the~~

~~algebraic second-moment model (Canuto et al., 2001). The ACPy (Auto Calibration Python) program was used to calibrate the model and three non-dimensional scaling factors that affect the~~ parameters were surface heat-flux~~,~~ factor (shf_factor), short-wave radiation ~~input and~~ factor (swr_factor), wind ~~were calibrated, as well as parameters affecting the~~factor (wind_factor), minimum turbulent kinetic energy (k_min) and e-folding depth for visible fraction of light~~.~~ (g2) (see Table S1 in the Supplement for the calibrated values). During calibration, the model was run approximately 5000 times to obtain a stable solution. The validation period was 7 years for Lake Erken (2010-2016) and 4 years for Lough Feeagh (2012-2015). For both the calibration and validation, daily mean water temperatures were simulated when GOTM was forced using measured mean hourly data. Model simulated profiles of mean daily water temperature were then compared to measured mean daily water temperature profiles. Model performance was evaluated by comparing mean daily modelled and measured temperature profiles and the model efficiency coefficients used were percent relative error (PRE), root mean squared error (RMSE) and Nash-Sutcliffe efficiency (NSE) (Nash and Sutcliffe, 1970). Overall, there was a good model fit for both lakes (Table 1).

2.4 Definitions for the epilimnion depth

We selected four epilimnion depth definitions that are commonly used in limnology and that were computationally efficient for multi-year automated high frequency data. These methods we describe as profile based methods (M1 – M4) (Fig.2.). In addition, we calculated epilimnion depth using a method for modelled data only (M5). In our analysis, epilimnion depth was expressed relative to the water surface and is therefore always a negative value. The range of thresholds used for each method was selected based on the values found within the literature (see Table 1 in Gray et al., 2020). We made no assumption of the conditions below the deepest measured depth and therefore the deepest estimated epilimnion depth was limited to the maximum measured depth for each lake (42 m in Lough Feeagh and 15 m in Lake Erken).

2.4.1 Absolute difference from the surface method (M1)

In M1, the epilimnion depth was defined as the shallowest depth where the density was a given ‘threshold’ value more than the surface density (Fig. 2), with the surface density (ρ_1) approximated as the density at the topmost sensor deployment, 0.9 m in Lough Feeagh and at 1 m in Lake Erken. We used a linear interpolation method to estimate the epilimnion depth on a continuous depth scale for all methods (Read et al., 2011), which assumed a linear relationship of densities between the first measured depth which exceeded the threshold (z_{i+1}) and the preceding measured depth (z_i). The numerical scheme can be described as (using notation from Read et al., 2011);

$$z_e = z_i + ((\rho_1 + \Delta\rho) - \rho_i)(\frac{z_{i+1} - z_i}{\rho_{i+1} - \rho_i}), \quad (1)$$

where z is depth (m), ρ is water density (kg m^{-3}), and Δρ is the threshold value (kg m^{-3}). The threshold values for the absolute method, M1 only, ranged from 0.025 kg m^{-3} to 0.2 kg m^{-3} at intervals of 0.025 kg m^{-3}. For all methods excluding the rLakeAnalyzer method (M4), if the threshold value was not exceeded, the epilimnion depth was defaulted to the deepest value (Lorbacher et al., 2006). Epilimnion depth estimates calculated with water temperature used the same type of equation (Eq. 1) but with temperature rather than density and noting that temperature decreases with depth. The only threshold value used for temperature was 1 °C.

2.4.2 Gradient from the surface method (M2)

In M2, the epilimnion depth was defined as the shallowest depth where the density gradient between consecutive measured depths exceeded the threshold value. M2 can be described as,

$$z_e = z_{i\Delta} + \left(\Delta\rho/\Delta z - \frac{\partial\rho}{\partial z_{i\Delta}}\right)\left(\frac{z_{i\Delta+1} - z_{i\Delta}}{\frac{\partial\rho}{\partial z_{i\Delta+1}} - \frac{\partial\rho}{\partial z_{i\Delta}}}\right), \quad (2)$$

where $z_{i\Delta}$ is the midpoint between z_i and z_{i+1}, and $\frac{\partial\rho}{\partial z_{i\Delta}}$ is the density gradient between z_i and z_{i+1} and $\Delta\rho/\Delta z$ is the threshold value (kg m^{-3} m^{-1}). The threshold values for all gradient methods, (i.e. M2, M3 and M4), ranged from 0.025 kg m^{-3} m^{-1} to 0.2 kg m^{-3} m^{-1} at intervals of 0.025 kg m^{-3} m^{-1}.

2.4.3 Gradient from the pycnocline method (M3)

In M3, the epilimnion depth was defined as the deepest depth where the density between consecutive measured depths exceeded the threshold value, starting from the depth of the maximum density gradient (hereafter the 'pycnocline') as the reference depth, and moving to successively shallower measured depths. M3 can be described by,

$$z_e = z_{i\Delta} + \left(\Delta\rho/\Delta z - \frac{\partial\rho}{\partial z_{i\Delta}}\right)\left(\frac{z_{i\Delta} - z_{i\Delta+1}}{\frac{\partial\rho}{\partial z_{i\Delta}} - \frac{\partial\rho}{\partial z_{i\Delta+1}}\right). \qquad (3)$$

2.4.3 rLakeAnlayzer (M4)

In M4, the epilimnion depth was defined using the rLakeAnalylzer function 'meta.depths' (relating to output "meta.top"), which used the same numerical scheme as M3, Eq. (3), but differed in certain assumptions (Read et al., 2011). Firstly, in M4, the epilimnion depth was prohibited from extending below the depth of the pycnocline. Therefore, for profiles where the predefined threshold value was less than the maximum density gradient, the epilimnion depth defaulted to the maximum density gradient. This differed from the other methods where, for such profiles, the epilimnion depth was defaulted to the deepest measured depth. Secondly, a user defined filter ('mixed.cutoff' object) was used to remove profiles which were not sufficiently stratified to identify an epilimnion depth. We used the default filter value, which removed profiles where the overall water temperature range was less than 1°C. For the days which did not meet the filter value, and no epilimnion depth was identified, we set the epilimnion depth to the deepest measured depth (i.e. no epilimnion depth) to ensure each method had the same number of data points for comparison with other methods.

2.4.5 Modelled turbulence method (M5)

The modelled turbulence method (M5) used the GOTM lake model simulated profile estimates of vertical eddy diffusivity (m^2 s^{-1}). In M5, the epilimnion depth was defined as the first depth, from the lake surface, where the vertical eddy diffusivity fell below the predefined threshold value, and was described as;

$$z_e = z_i + \left(\Delta K_z - K_{z_i}\right)\left(\frac{z_{i+1} - z_i}{K_{z_{i+1}} - K_{z_i}}\right), \qquad (4)$$

where K_z is vertical eddy diffusivity (m^2 s^{-1}) and ΔK_z is the threshold value (m^2 s^{-1}). The thresholds ranged from 10^{-5} to 10^{-4} m^2 s^{-1} at intervals of 10^{-5} m^2 s^{-1}, based on the values described in Wüest and Lorke (2009) and MacIntyre and Melack (2009).

2.5 Analysis methods

2.5.1 Comparison between a water temperature and water density derived method

To compare water temperature and water density based estimates of the epilimnion depth, we used M1 only and used a water temperature threshold ~~value~~ of 1 °C with a density threshold of 0.1 kg m^{-3} for both sites. Firstly, we investigated the relationship between 1 °C and 0.1 kg m^{-3} throughout the year. To do this, we calculated the long-term ~~average~~mean water column temperature for each Julian day. For each day, we then calculated the change in density that would result from a 1 °C increase in the water temperature~~. We~~ and then subtracted 0.1 kg m^{-3} ~~from each Julian day value.~~. Positive values~~, shown in red~~, indicated that a 1 °C increase in the water temperature, resulted in a greater than 0.1 kg m^{-3} change in water density, while negative values, ~~shown in blue,~~ indicated a less than 0.1 kg m^{-3} change in water density. Secondly, we compared water temperature and water density based estimates of the epilimnion depth. To do this, we calculated the difference between the

~~average~~mean water density derived estimate and the water temperature derived estimate for each Julian day. Positive differences, ~~shown in red,~~ indicated that the water density derived estimate was shallower than the water temperature derived estimate, while negative values, ~~shown in blue,~~ were deeper. For all analysis of measured data, the total number of observations were used for Lough Feeagh (n= 2778, years = 10) and Lake Erken (n = 2005, years = 7).

2.5.2 Comparison between water density based methods (M1 – M4)

Following this, we ~~confined the analysis to comparing~~compared water density based epilimnion depth estimates, using all four methods: M1, M2, M3 and M4 and the range of thresholds described earlier. Using data from both sites, we considered overall variability (i.e. how much do estimates vary between all methods and all thresholds?), variability within each individual method using different threshold definitions (i.e. how sensitive are estimates to the threshold value selected?) and variability between methods (i.e. what systematic differences exist between pairs of methods?). Given that we had a total of 32 time series to compare, 4 methods each with 8 threshold values, it was necessary to compute summary statistics for each of them. Therefore, the following statistics were calculated for all 32 time series, for the period from 1st April to 31st October each year and then averaged across all years. Firstly, we calculated the ~~average~~mean epilimnion depth and presented the values for all methods and all thresholds values. We also summarised these statistics for each method, showing the ~~average~~mean, standard error of mean, minimum (shallowest), maximum (deepest) and interquartile range for each method, to demonstrate differences between methods. A large interquartile range in epilimnion depth estimates~~,~~ indicated high sensitivity to the threshold value. Secondly, we calculated the percentage of days with available data, where the epilimnion depth was detected above the deepest measured depth. This demonstrated differences between methods in regards to the stratified period. Thirdly, we calculated the percentage of days with available data where the epilimnion depth was detected above the maximum density gradient or pycnocline. By definition the epilimnion should have relatively small density gradients and should not be equal or deeper than the pycnocline, however automated methods, ~~depending on the logical and numerical schemes used~~have been found to ~~calculate them, may~~regularly encroach on the metalimnion (Lorbacher et al., 2006). We therefore used this metric to investigate how frequently epilimnion depth estimates calculated by each method erroneously extended into the metalimnion.

Pearson's correlation coefficients were also calculated for all possible combinations between the 32 time series, to quantify the degree of association between them, without using any estimates of significance (Thomson and Fine, 2003, Rivetti et al., 2016). ~~We~~The full correlation matrices were calculated and then for clarity, we presented only the ~~average~~mean Pearson's correlation coefficient for each method, representing the ~~average~~mean correlation for all possible combinations between threshold values. This indicated the extent to which changing the threshold value influenced the temporal patterns. We also presented the ~~average~~mean Pearson's correlation coefficients between each pair of methods (e.g. for all threshold combinations between M1 and M2 etc.) to demonstrate method agreement.

2.5.3 Sensitivity of epilimnion depth to water column structure and vertical sensor resolution

We also assessed the sensitivity of the profile based methods to changes in the water column structure and the vertical sensor resolution of measured data. For the water column structure sensitivity analysis, we calculated the long-term ~~average~~mean epilimnion depth estimate for each Julian day for all 32 method/threshold time series. For each method, using all thresholds, we calculated the range for each Julian day. The range in estimates was presented alongside the top-bottom density gradient for each Julian day, to investigate whether threshold sensitivity varied temporally and with water column structure. For the vertical sensor deployment resolution sensitivity analysis, we compared simulated water density profiles for both lakes at two different resolutions. High resolution data was resolved to 0.5 m for both lakes. Low resolution data were subset to ~~an average~~a mean of 1 sensor per 3 m, using the measured depths for Lough Feeagh, and data from 1, 2.5, 5, 8 and 13 m for Lake Erken. We then calculated the difference between the Apr-Oct ~~average~~mean epilimnion depth for the high and low resolution data. Methods where the high and low resolution data produced very different estimates were regarded as having high sensitivity to the vertical resolution of the data, while methods with small differences indicated low sensitivity. For all analysis using simulated data, the total number of observations were used for Lough Feeagh (n = 1016, years = 5) and Lake Erken (n = 1449, years = 7).

2.5.4 Comparison with modelled turbulence method (M5)

Finally, we assessed how each profile based method compared against the turbulence based estimates. For this analysis, both water density and vertical eddy diffusivity profile data were derived using the GOTM lake model. Then, using the same procedures as the measured data, we calculated the mean Apr-Oct ~~average~~ epilimnion depth for each method. We then calculated the difference between the turbulence method (M5) and each of the four profile based methods (M1- M4). We also presented the ~~average~~mean Pearson's correlation coefficients between each method and M5 (e.g. for all threshold combinations between M5 and M1, etc.). These results indicated the extent to which profile based methods were able to characterise active mixing penetration, within a hydrodynamic model setting, rather than confirming which method was more reliable for predicting the 'true' mixing depth.

3 Results

3.1 Comparison between a water temperature and water density derived method

There were large systematic differences between the epilimnion depth calculated using a water temperature based method compared to values calculated using a water density based method (Fig. 3). Due to the non-linear relationship between water density and water temperature, the difference in density induced by a water temperature increase of 1°C (water column ~~average~~mean) varied seasonally, with the pattern differing between sites (Fig. 3a). We found that on average, during the spring (April-May), when water column temperatures in both lakes were relatively low, a change of 1°C resulted in a water density change of less than 0.1 kg m^{-3}~~.~~, as shaded in blue. As a result of this anomaly, estimates of the epilimnion depth that were based on water temperature data were shallower compared to those calculated using the water density method (Fig.3b). In contrast, in general from June to October for both sites, a change of 1°C in water temperature induced a change in water density of greater than 0.1 kg m^{-3}, as shaded in red, which resulted in estimates of the epilimnion depth which were deeper when using water temperature compared to those estimated using water density. Based on the long-term daily ~~averages~~means, the differences in the estimates of epilimnion depth between the two methods ranged from 3 to 5 m for Lough Feeagh, and 2 to 4 m for Lake Erken.

3.2 Comparison between water density based methods (M1 – M4)

Inspection of water column profiles highlighted key differences in the performance of methods M1, M2, M3 and M4 (Fig. 4). In a stratified profile, with a well-defined three-layered water column profile, there was often strong agreement on the epilimnion depth between all methods and thresholds (Fig. 4a). In contrast, when the measured temperature profile was more complex, i.e. at times when there was some stratification close to the surface or when a secondary pycnocline had developed close to the surface, there was less agreement on the estimates of the epilimnion depth between methods (Fig. 4b). For such profiles, estimates of the epilimnion depth calculated with the absolute difference from the surface method, M1, were typically staggered at linear intervals along the profile depending on the exact threshold value. In contrast, the estimated epilimnion depth calculated using the gradient methods (M2, M3 and M4) had a tendency to cluster at discrete locations on the profile. Therefore, a small change in the threshold value induced either no difference at all in the epilimnion depth or at other times a very large difference. For profiles with low water column stability, there was particularly large differences in the estimated epilimnion depth calculated using different methods, reflecting differing underlying assumptions (Fig.4c). For M3, for example, the epilimnion depth was defaulted to the deepest depth when the threshold value was not exceeded, as was also the case for methods M1 and M2. In contrast, however, in M4, near-isothermal profiles often met the 'mixed.cutoff' filter condition (i.e. water column range > 1°C), whilst still not having sufficient density gradients to meet the user threshold value. As a result, in M4, the epilimnion depth was defaulted to the pycnocline, which, given the small density gradients, was often found at a very shallow depth.

Time series results demonstrated the extent of the variability in epilimnion depth estimates between all methods and thresholds (Fig. 5). Considering that all the time series estimates for Lough Feeagh (left-side) and Lake Erken (right-side) were presumed to estimate the same theoretical location, they would ideally all produce exactly the same temporal patterns. Instead, the differences were large enough to obscure the annual patterns and hinder the ability to compare between the two lakes. The overall mean epilimnion depth estimate using methods M1-M4 and all thresholds was -28.1 m (standard error = 0.6 m, IQR = 19.0 m) for Lough Feeagh and -11.0 m (standard error = 0.1 m, IQR = 2.3 m) for Lake Erken. The overall variability between all estimates was particularly high for Lough Feeagh, where the Apr-Oct ~~average~~mean epilimnion depth ranged by 36.9 m (-4.6 m to -41.5 m) while in Lake Erken, estimates ranged by 5.2 m ~~(from -~~(-7.8 m to -13.0 m) (Fig.6a).

There were evident systematic differences between methods. In both lakes, the ~~average~~mean Apr-Oct epilimnion depth for each method was shallowest for M1 and was on average shallower by 17.0, 16.6 and 2.2 m compared with methods M2, M3 and M4 in Lough Feeagh, and 1.2, 1.7, 0.8 m in Lake Erken (Table 2.). The minimum (shallowest) estimates of the Apr-Oct ~~average~~mean, for gradient methods (M2, M3 and M4) were comparable in magnitude to the maximum (deepest) estimate for the absolute difference from the surface method, M1. The ~~average~~mean Pearson's correlation coefficient between each pair of methods also demonstrated that certain method pairs had greater temporal agreement than other pairs (Table 3). The full correlation matrices are available in the Supplement (Table S2). Method pairs, M3-M2 and M4-M1 had particularly high Pearson correlation coefficients for both lakes, suggesting these methods produced similar temporal trends. In Lake Erken all method pairs had higher Pearson's correlation coefficients than Lough Feeagh.

The selection of a threshold value proved to be very important in the estimation of the epilimnion depth. For all methods, smaller threshold values produced shallower estimates of the ~~average~~mean Apr-Oct epilimnion depth while larger threshold values produced deeper estimates (Fig.6a). Methods with a large range between the shallowest (minimum) and deepest (maximum) estimate demonstrated high sensitivity to the threshold value (Table 2). For both lakes, the interquartile range in the ~~average~~mean Apr-Oct epilimnion depth estimates for each method was very high for M2, M1 and M3, indicating high threshold sensitivity in these methods. Method M4 had a substantially lower interquartile range than all other methods and a very high ~~average~~mean Pearson's correlation coefficient, indicating that both the ~~average~~mean value and the temporal pattern of the epilimnion depth were only weakly influenced by the threshold value. In both lakes, methods M2 and M1, where the epilimnion depth was defined from the surface downwards, had a higher interquartile range in estimates calculated with different threshold values, compared to methods M3 and M4, where the epilimnion was defined from the pycnocline upwards. M1, however, had higher ~~average~~mean Pearson's correlation coefficient than M2 and M3, indicating that the temporal pattern of the epilimnion depth was less influenced by the threshold value. In general, the threshold sensitivity of each method reduced with increasing threshold size. That is, the changes in the epilimnion depth occurring between threshold values decreased with increasing threshold value (Fig.6a). For example, for M2, the difference in the Apr-Oct ~~average~~mean epilimnion depth between the first two thresholds (0.025 and 0.05 kg m^{-3} m^{-1}) was much greater than the difference between the last two thresholds (0.175 and 0.2 kg m^{-3} m^{-1}), in both lakes.

The percentage of stratified days, defined as days where the epilimnion depth was identified ~~at a depth greater~~as shallower than the deepest measured depth, demonstrated the extent to which different methods/thresholds influenced ~~annual patterns~~ the stratified period (Fig.6b). For M4, the percentage of stratified days remained static regardless of the threshold value ~~for method M4 as an~~ , because the epilimnion depth was ~~defined whenever~~detected for all profiles where the water column temperature range was more than 1 °C, regardless of the threshold used. For all other methods, the number of stratified days decreased with increasing threshold value. For M1 the difference in stratified days between threshold values was small, compared to both gradient methods M2 and M3, particularly in Lough Feeagh. For example, in Lough Feeagh, for M3, the number of stratified days calculated using a threshold value of 0.~~25~~025 kg m^{-3} m^{-1} was 125, while for threshold values greater than 0.075 kg m^{-3} m^{-1}, the ~~average~~mean number of stratified days per annum decreased to less than 38 days.

The percentage of days where the epilimnion depth was located above the pycnocline, defined as days where the epilimnion depth was identified above the maximum density gradient, indicated that some methods may be less prone to erroneously estimating the epilimnion depth in the metalimnion, compared with others (Fig.6b). For both lakes, M1 had the highest number of days where the epilimnion depth was located above the pycnocline, suggesting that on average the method extended into the metalimnion less frequently than other methods. In Lough Feeagh, all gradient methods, M2, M3 and M4, had very high range occurring between the different threshold values. In Lough Feeagh, gradient methods calculated with a threshold value greater than 0.15 kg m^{-3} m^{-1}, resulted in an average of zero days where the epilimnion depth was located above the pycnocline.

3.3 Sensitivity of epilimnion depth to water column structure

For all methods, threshold sensitivity fluctuated seasonally, although varied in pattern (Fig. 7). Threshold sensitivity was shown by the interquartile range between the ~~shallowest and deepest~~ epilimnion depth estimates calculated for all threshold values. In Lough Feeagh, M1 had a smaller interquartile range in epilimnion depth estimates during the peak summer months of June, July and August, compared with months when the onset and overturn of stratification commonly occurred. During periods of transient stratification, the stability of the water column was often low but frequent changes in the near-surface water density, induced large differences between estimates calculated using small thresholds compared with large threshold values. In contrast, methods M2 and M3 had the highest interquartile range in estimates occurring during the peak summer months. Even during peak summer in Lough Feeagh, gradients in the water column were relatively small (Fig. 7b), which resulted in a very large ~~range~~differences between the smallest threshold values which found a near-surface epilimnion depth, and the largest thresholds that often found no epilimnion depth at all, therefore defaulting to the deepest depth. In Lake Erken, the water density gradients were typically much larger, and methods M1, M2 and M3 all peaked during May and June, when gradients in the water column were typically increasing but prone to fluctuations. ~~For both lakes, M2 had typically a higher threshold range than M3 during peak summer and the overturn period, which was related to the common development of a secondary pycnocline. M4 produced much lower ranges~~Compared with all other methods, M4 produced substantially lower interquartile range in the epilimnion depth throughout the year, since as long as the ‘mixed.cutoff’ filter was met, the epilimnion depth was defaulted to the pycnocline if the threshold was not exceeded, thus largely reducing the ability for large differences to occur. The interquartile range in epilimnion depth estimates for M4 was highest during the peak summer months, which was when the epilimnion depth was typically shallowest and more frequently defined by the threshold value rather than defaulting to the pycnocline.

3.4 Sensitivity of epilimnion depth to vertical sensor resolution

The vertical resolution of water density data was found to have a systematic influence on the estimation of the epilimnion depth for all methods (Table 4). Overall, the modelled higher vertical resolution data resulted in shallower estimates of the epilimnion depth, relative to the estimates made with the modelled low resolution data. For Lough Feeagh, the results showed that the annual ~~average~~mean Apr-Oct epilimnion depth estimate using high resolution data were on ~~average~~mean 0.1, 3.2, 3.2 and 0.5 m shallower than those using low resolution data for methods M1, M2, M3 and M4 respectively, while in Lake Erken they were 0.0, 1.2, 1.0, 0.2 m shallower. Methods M1 and M4 had substantially smaller differences between high and low resolution estimates compared with M2 and M3. In particular, M1 had almost no difference between high and low resolution data, indicating that this method had very low sensitivity to the vertical sensor deployment.

3.5 Comparison with modelled turbulence method (M5)

In general, the modelled turbulence method had very low sensitivity to the threshold value, compared with the profile based methods also calculated using modelled data. A time series comparison of all modelled results is available in the Supplement (Fig. S1). For both lakes, we found that the modelled turbulence method produced shallower estimates than modelled profile based methods (Table 5). In Lough Feeagh, the ~~average~~mean Apr-Oct epilimnion depth estimate using the modelled turbulence method M5 was -20.8 m, which was 1.3 m, 11.0 m, 11.2 m and 1.3 m shallower than methods M1, M2, M3, and M4 respectively, while in Lake Erken the M5 estimate was -11.0 m, which was 0.0 m, 1.0 m, 1.1 m and 0.4 m shallower. In both lakes, M1 had the strongest agreement with M5, demonstrated by both the ~~average~~mean difference (1.3 m in Lough Feeagh and 0.0 m in Lake Erken), and the highest Pearson's correlation coefficient in Lough Feeagh (r = 0.90) and Lake Erken (r = 0.89). This was followed by M4, which also had strong agreement with M5. In contrast, M2 and M3 had much weaker agreement with M5, in terms of both the Apr-Oct epilimnion depth estimate and the Pearson's correlation coefficients.

4 Discussion

The concept of the epilimnion, and more widely, the three-layered structure of a stratified lake, is fundamental in limnology. Yet, despite the ubiquity of the term, there is no objective or generic approach for defining the epilimnion and a diverse number of approaches prevail in the literature. In a comprehensive analysis of high-frequency, multi-year data from two lakes, this study has highlighted the extent to which common water temperature profile based epilimnion depth estimates differ. The level of variability in epilimnion depth estimates calculated using common methods and threshold values, was exceedingly high. This result calls into question the practice of arbitrary method selection and comparing findings between studies which use different methods or even just different thresholds. The magnitude of variability also casts ambiguity on the calculation of key biogeochemical and ecological processes in a lake that rest on the assumption that the layers of a lake are well defined, including calculations of metabolic rates, and oxygen fluxes (e.g. Coloso et al., 2008, Foley et al., 2012, Obrador et al., 2014, Winslow et al., 2016). ~~Ultimately, these results emphasise the limitations inherent to defining the epilimnion using water temperature profile data and on a temporally continuous basis. In an idealised stratified profile, the epilimnion is portrayed as near-uniform in water temperature or density and clearly delineated from a well-defined metalimnion. However, the vast majority of the measured profiles, at least within this study, did not conform to this idealised three-layered structure. In these cases, methods not only diverged on the location of the epilimnion depth but~~

~~also may not even be underpinned by the same theoretical principles. Since none of these methods can be considered the 'true' definition of the epilimnion depth, acknowledgement of the systematic differences between methods, manifest over high-frequency time series, is essential for selecting suitable methods for future analysis.~~

In an idealised stratified profile, the epilimnion is portrayed as near-uniform in water temperature or density and clearly delineated from a well-defined metalimnion. However, many measured profiles, at least within this study, did not conform to this idealised three-layered structure. Instead the thermal water column structure was often more complex, including multiple pycnoclines, near-surface micro-stratification layers, and blurred boundaries between the epilimnion/metalimnion. One approach to this issue is to filter out appropriate water column profiles or apply functions that coerce the profile into the expected structure (Read et al., 2011, Pujoni et al., 2019, Gray et al., 2020). Our analysis of temporally high resolution time series data emphasised, however, that rather than jumping from states, such as stratified or isothermal, changes in the water column occurred over an evolving continuum and often fluctuated between states. Similarly, the distinction between additional layers, such as the primary or secondary pycnocline, is fraught with the same issues of arbitrariness as discussed (Read et al., 2011). This study demonstrates that when epilimnion depth estimation methods, which are theorised for a three-layered water column, are applied to non-conforming water columns, they diverge widely on the location of the epilimnion depth, and at times, may not even be underpinned by the same theoretical assumptions. Since none of these methods can be considered the 'true' definition of the epilimnion depth, it is necessary to understand the degree to which methods differ. Improved understanding of their systematic differences will facilitate the use of methods that appropriately capture different processes, such as air-water exchanges, thermocline entrainment or the suspension of materials. Due to the realised complexities of observed and aggregated profile data, we may benefit from new approaches to water column discretisation that consider the vast proportion of profiles which do not conform neatly to the three-layered paradigm.

A ~~particularly distinct~~large systematic difference was found between water temperature and water density. Due to the non-linear relationship between water density and temperature, the use of water temperature was equivalent to using different density threshold values throughout the year, resulting in a distinct shift in the stratification period. Although water density gradients are driven by temperature changes in lakes and are also calculated from water temperature estimates, water density directly influences mixing processes and is therefore recommended for estimating the epilimnion depth (Read et al., 2011, Gray et al., 2020). The implications of using a water temperature based method may be particularly enhanced in Northern temperate lakes due to the large annual water temperature ranges (Maberly et al., 2020). Pronounced differences in the estimation of the epilimnion depth were also found within estimates derived using the same water density input data. Typically, for the range of common thresholds used in this study, the absolute difference from the surface method, M1, produced shallower estimates relative to gradient based methods. In addition, the difference between these methods was particularly large when the vertical resolution of the data was low. This suggests that studies using gradient based methods, particularly those using coarse vertical data, may have a deep bias relative to those using an absolute method, and consequently, were more prone to erroneously extending into the metalimnion. In addition, as may be expected, the use of larger threshold values also produced systematically deeper estimates of the epilimnion depth. Surprisingly, however, the magnitudes of these differences were on par with those occurring between methods. The implications of a shallow or deep bias may be far-reaching, particularly given that various biological and ecological metrics have already been found to be highly sensitive to changes in the epilimnion depth (Coloso et al., 2008, Gray et al., 2020). For example, a deeper estimate of the epilimnion depth would systematically lead to a larger ratio between the epilimnion and euphotic depth, compared with a shallower estimate, which if used to understand the development of a phytoplankton bloom, could lead to contradictory results (Huisman et al., 1999). Alternatively, ~~an acceleration of epilimnion shallowing during stratification onset, as found~~

~~with~~ water temperature-based estimates~~, might~~ typically resulted in earlier stratification, which could indicate a longer duration of ~~conditions necessary for~~ phytoplankton ~~growth~~in a shallower epilimnion. The implications of a seasonal or deep/shallow biases may be even more important for computing fluxes (e.g. oxygen or nutrients) between the epilimnion and the metalimnion, since both terms are influenced by the epilimnion depth (Giling et al., 2017, Gray et al., 2020).

An important difference was also found between methods detecting the layer that is isothermal relative to the surface and methods detecting the ~~top~~point that is isothermal relative to the steep gradients of the metalimnion, which has not been well considered in the literature. M1 and M2, defined from the surface downwards, were more prone to the detection of a shallow secondary pycnocline, compared with M3 and M4. Instead, M3 and M4, defined from the pycnocline upwards, prioritised the relative difference between the metalimnion and the surface. From a theoretical point of view, processes related to the air–water interface could be better suited to methods identifying the isothermal layer, while for processes related to the entrainment of deep water into the epilimnion are more suited to top of the metalimnion methods.

The selection of an epilimnion method also had surprisingly large consequences for understanding the stratification period, which is widely used for quantifying the impact of climate change on lakes (Livingstone, 2003, Butcher et al., 2015, Ayala et al., ~~in review~~2020). Notably, the ~~average~~mean epilimnion depth and number of stratified days calculated using M4, depended very little on the threshold value selected. Instead, the selection of the filter (defaulted to a water column range of > 1°C), which was unique to this method, determined the number of stratified days and largely influenced the other bulk statistics. This also resulted in the epilimnion being identified even when the threshold was not exceeded, which in some instances could have the effect of muting relative temporal changes in the epilimnion depth. In contrast, for the other methods, the threshold was used to determine whether the water column was considered to be stratified and therefore the stratification period was highly sensitive to the threshold value, similarly to the other bulk statistics. Ultimately, these results suggest that the stratification period calculated in different studies or for different regions cannot be compared unless identical definitions are used. The method most appropriate for identifying the stratified period has been considered in other studies (Woolway et al., 2014, Engelhardt and Kirillin, 2014) however our results offer some additional insights. ~~Although~~The results suggest that use of water density metrics, such as epilimnion depth estimates, in combination with traditional water temperature ~~thresholds are typical for defining the~~ based definitions of stratification~~ period in lakes, they may not be suitable for use with water density metric~~, are incompatible, given the non-linear relationship between temperature and ~~potentially introduce a seasonal bias~~density. In addition, estimations of the epilimnion depth, and the variability among definitions, may be particularly relevant for understanding the stratified period since it is often assumed that the onset of stratification marks the decoupling of the epilimnion from the deeper layers, thus determining the duration of nutrient limitations in the epilimnion and oxygen limitations in the hypolimnion (MacIntyre, 1993, Foley et al., 2012, Schwefel et al., 2016).

Regardless of the method selected, however, all water temperature/density based methods are limited in their ability to indicate actual mixing processes. Our results using the lake modelled turbulence data demonstrated that even in a modelled environment, ~~methods~~ epilimnion depth estimates were inconsistent ~~with~~between the different methods and threshold values studied, and that turbulence based methods~~, which~~ generally resulted in a shallower epilimnion depth estimate. ~~Although these results are not necessarily indicative of measured data, they do~~These findings highlight the ~~need for caution when interpreting water temperature/density derived epilimnion depth estimates. There is an~~ important but subtle difference between the layer ~~that~~detected by water density profiles (i.e. has been recently well-mixed and therefore has little resistance to further mixing~~,~~ due to the lack of density gradients~~,~~), and the layer that is actively mixing~~ and is~~, determined only through directly measured turbulence~~.~~ (Gray et al., 2020). Similarly, micro-profiling studies have shown that the actively mixing layer can be substantially shallower than the layer determined through water temperature profile data (McIntyre et al., 1993, Tedford et al., 2014). Micro-profile studies also demonstrate that within seemingly uniform layers there are micro-stratification layers, delineated by temperature differences as small as 0.02°C (Imberger, 1985, Shay and Gregg, 1986, MacIntyre, 1993; Jonas et al. 2003), which can be sufficient to isolate intermediate layers from atmospheric wind shear and cooling (Pernica et al., 2014). Although our results are not directly indicative of measured data, they demonstrate how even turbulence based methods are inherently arbitrary, as there is no objective threshold value (Monismith and Macintyre, 2009). Many of the ecological applications of the epilimnion depth have the underlying assumption that enough mixing is occurring in the epilimnion to keep the relevant organisms or particles suspended within the layer. However, whetherWhether mixing is actually occurring, however, and to what extent, is not directly described by epilimnion depth estimations derived using water temperature or density profile data, and in fact, previous studies have found water density estimates of the epilimnion depth to be relatively poor indicators for the homogeneity of other ecological variables (Gray et al., 2020). Therefore, epilimnion depth estimates derived from profile data may be more reliably used for indicating relative changes in mixing over time (Read et al., 2012, Calderó-Pascual et al., in review). 2020).

The selection of a suitable threshold value is far more important than previously attributed in limnology. In general, a suitable threshold is any value that can be reasonably considered as homogenous while also within the limit of sensor detection (De Boyer et al., 2004). However, all the threshold values used in this study met these criteria, yet produced fundamentally different epilimnion depth estimations and temporal patterns. Although, it may be unreasonable to suggest a 'universal' threshold value, a given study may find a threshold that is less problematic than other values. In general, we found that the sensitivity of the epilimnion depth to the threshold value decreased with the increasing size of the threshold. That is, for small thresholds the impact of changing the threshold value was greater than larger thresholds for the same incremental change. This may suggest that for studies using smaller threshold values, the results are more threshold dependent than those using large threshold values. However, larger threshold values had greater frequency of the epilimnion depth estimates being below the maximum density gradient, suggesting that larger threshold values tend to extend into the stable depths of the metalimnion more regularly, and hence somewhat explaining the lower threshold sensitivity. The trade-off between threshold sensitivity and encroachment into the metalimnion points towards a mid-range threshold, such as 0.1 kg m^{-3} or 0.1 kg m^{-3} m^{-1}, as potentially, being more reliable than large or small thresholds.

One of the main goals behind the global collection of high-frequency data in lakes is to understand how physical processes vary between lakes, which indicate how different lakes may respond to changing climatic conditions (Weathers et al., 2013, Kraemer et al., 2015, Woolway et al., 2019). In order to understand this, we require methods that perform consistently between lakes and over longitudinal scales. The differences between the two lakes studied, in particular variability in water column structure, the strength of density gradients and the vertical resolution of sensor deployment, influenced the level of agreement between epilimnion depth methods. Overall, Lake Erken had much greater agreement among methods than Lough Feeagh. In particular, we found this to be related to the difference in vertical resolution of the measured data between sites. Of all the methods considered in this study, our results suggest that the absolute difference from the surface method, M1, might be more useful as a 'generic' method, due in particular to the very low sensitivity to the vertical sensor resolution compared with all other gradient based methods. This finding is in agreement with previous oceanography studies that have similarly found gradient methods to be highly sensitive to vertical resolution (e.g. Lorbacher et al., 2006, Thomson and Fine, 2003). In addition, however, the performance and threshold sensitivity of all methods also fluctuated temporally as influenced by changes in the water column structure. Assessment of the uncertainty associated with epilimnion depth estimates may be useful, particularly for studies comparing the epilimnion depth between periods of time that vary in stratification intensity.

Although long-term epilimnion depth trends are only rarely reported directly (e.g. Hondzo and Stefan, 1993, Fee et al., 1996, Sahoo et al., 2013), they are embedded in our understanding of many climate related variables. For example, the epilimnion depth plays a key role in modulating the effects of eutrophication, browning and climate change on lake water surface and epilimnetic temperatures (Persson and Jones, 2008, Flaim et al., 2016, Strock et al., 2017, Bartosiewicz et al., 2019). As such, changes in the epilimnion depth may enhance or mute the effect of increasing incoming heat on water surface temperatures and therefore may be particularly important in explaining temporal and spatial anomalies in surface temperature trends. Given the results of this study it may be that long-term trends calculated using different metrics relate to fundamentally different parts of the water column that may be undergoing different changes due to climate change. Therefore, the strength, and even the direction, of long-term trends in the epilimnion may be highly dependent on the definition used (Yang and Wang, 2008, Somavilla et al., 2017).

5 Conclusion

This study has demonstrated the extent to which different definitions of the epilimnion depth lead to different locations of the epilimnion depth in the water column and produce very different and contradictory temporal patterns. These results have wide-reaching relevance in limnology, including for studies interested in metabolism, eutrophication, and hypolimnetic anoxia. The sensitivity of epilimnion depth methods to temporal and spatial characteristics, such as morphology, water column structure and vertical resolution of data measurements may also pose challenges for studies interested in long-term trends or global lake comparison studies. While there is no prescribed rationale for selecting a particular method, the M1 method, defined as the shallowest depth where the density was 0.1 kg m^{-3} more than the surface density, was shown to be ~~overall less problematic than~~particularly insensitive to the ~~other methods~~vertical sensor resolution of water temperature data, while the temporal pattern was relatively robust to changes in the threshold value, and therefore may be particularly useful as a generic method.

Code and data availability

The analysis codes and output data are stored in HydroShare http://www.hydroshare.org/resource/26dbc260405b4bb9b3ac16ec55432684. Source code of the model GOTM is freely available online at https://gotm.net/. The data used in this study from Lough Feeagh is available online at https://doi.org/10.20393/6C4760C2-7392-4347-8555-28BA0DAD0297.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

This research was carried out as part of the MANTEL Innovative Training Network which has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 722518. The long term monitoring program of Lough Feeagh is enabled by the field staff of the Marine Institute, and we gratefully acknowledge their support. The monitoring program at Lake Erken received financial support from the Swedish Infrastructure for Ecosystem Science (SITES~~)~~). RIW received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 791812

References

Andersen, M.R., Sand-Jensen, K., Woolway, R.I. and Jones, I.D.:Profound daily vertical stratification and mixing in a small, shallow, wind-exposed lake with submerged macrophytes, Aquatic sciences, 79(2), 395-406, https://doi.org/10.1007/s00027-016-0505-0, 2017.

Ayala, A. I., Moras, S., and Pierson, D. C.: Simulations of future changes in thermal structure of Lake Erken: Proof of concept for ISIMIP2b lake sector local simulation strategy, Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-335, ~~in review~~2020.

Bartosiewicz, M., Przytulska, A., Lapierre, J.F., Laurion, I., Lehmann, M.F. and Maranger, R.: Hot tops, cold bottoms: Synergistic climate warming and shielding effects increase carbon burial in lakes, Limnology and Oceanography Letters, 4(5), 132-144, https://doi.org/10.1002/lol2.10117, 2019.

Berger, S.A., Diehl, S., Kunz, T.J., Albrecht, D., Oucible, A.M. and Ritzer, S.: Light supply, plankton biomass, and seston stoichiometry in a gradient of lake mixing depths, Limnology and Oceanography, 51(4), 1898-1905. https://doi.org/10.4319/lo.2006.51.4.1898, 2006.

Bouffard, D. and Wüest, A.: Mixing in stratified lakes and reservoirs, In: Mixing and dispersion in flows Dominated by Rotation and Buoyancy, Springer, Cham, 61-88, https://doi.org/10.1007/978-3-319-66887-1_3, 2018.

Butcher, J.B., Nover, D., Johnson, T.E. and Clark, C.M.: Sensitivity of lake thermal and mixing dynamics to climate change, Climatic Change, 129(1-2), 295-305, https://doi.org/10.1029/2010WR009104, 2015.

Burchard, H. and Petersen, O.: Models of turbulence in the marine environment—A comparative study of two-equation turbulence models, Journal of Marine Systems, 21(1-4), 29-53, https://doi.org/10.1016/S0924-7963(99)00004-4, 1999.

Brainerd, K.E. and Gregg, M.C.: Surface mixed and mixing layer depths, Oceanographic Literature Review, 4(43), 325, https://doi.org/10.1016/0967-0637(95)00068-H, 1996.

Calderó-Pascual, M., de Eyto, E., Jennings, E, Dillane, M., Andersen, M., Kelly, S., Wilson, H., McCarthy, V.: Effects of consecutive extreme weather evnts on an Irish dystrophic lake: A detailed insight into physico-chemical and biological responses, Water, ~~in review~~2020.

Canuto, V.M., Howard, A., Cheng, Y. and Dubovikov, M.S.: Ocean turbulence. Part I: One-point closure model—Momentum and heat vertical diffusivities, Journal of Physical Oceanography, 31(6), 1413-1426, https://doi.org/10.1175/1520-0485(2001)031<1413:OTPIOP>2.0.CO;2, 2001.

Coloso, J.J., Cole, J.J. and Pace, M.L.: Short-term variation in thermal stratification complicates estimation of lake metabolism, Aquatic Sciences, 73(2), 305-315, https://doi.org/10.1007/s00027-010-0177-0, 2011.

Diehl, S., Berger, S., Ptacnik, R. and Wild, A.: Phytoplankton, light, and nutrients in a gradient of mixing depths: field experiments, Ecology, 83(2), 399-411, https://doi.org/10.1890/0012-9658(2002)083[0399:PLANIA]2.0.CO;2, 2002.

de Boyer Montégut, C., Madec, G., Fischer, A.S., Lazar, A. and Iudicone, D.: Mixed layer depth over the global ocean: An examination of profile data and a profile- based climatology, Journal of Geophysical Research: Oceans, 109(C12), https://doi.org/10.1029/2004JC002378, 2004.

de Eyto, E., Jennings, E., Ryder, E., Sparber, K., Dillane, M., Dalton, C. and Poole, R.: Response of a humic lake ecosystem to an extreme precipitation event: physical, chemical, and biological implications, Inland Waters, 6(4), 483-498, https://doi.org/10.1080/IW-6.4.875, 2016.

de Eyto, E., Dillane, M., Moore, T., Wilson, H., Cooney, J., Hughes, P., Murphy, M., Nixon, P., Sweeney, D., & Poole, R.: Lough Feeagh water temperature profiles, Marine Institute, Ireland. https://doi.org/10.20393/6C4760C2-7392-4347-8555-28BA0DAD0297, 2020.

Dong, F., Mi, C., Hupfer, M., Lindenschmidt, K.E., Peng, W., Liu, X. and Rinke, K.: Assessing vertical diffusion in a stratified lake using a three dimensional hydrodynamic model, Hydrological Processes, 34(5), 1131- 1143, https://doi.org/10.1002/hyp.13653, 2019.

Engelhardt, C. and Kirillin, G.: Criteria for the onset and breakup of summer lake stratification based on routine temperature measurements, Fundamental and Applied Limnology/Archiv für Hydrobiologie, 184(3), 183-194, https://doi.org/10.1127/1863-9135/2014/0582, 2014.

Fee, E.J., Hecky, R.E., Kasian, S.E.M. and Cruikshank, D.R.: Effects of lake size, water clarity, and climatic variability on mixing depths in Canadian Shield lakes, Limnology and Oceanography, 41(5), 912-920, https://doi.org/10.4319/lo.1996.41.5.0912, 1996.

Foley, B., Jones, I.D., Maberly, S.C. and Rippey, B.: Long- term changes in oxygen depletion in a small temperate lake: effects of climate change and eutrophication, Freshwater Biology, 57(2), 278-289, https://doi.org/10.1111/j.1365-2427.2011.02662.x, 2012.

Flaim, G., Eccel, E., Zeileis, A., Toller, G., Cerasino, L. and Obertegger, U.: Effects of re-oligotrophication and climate change on lake thermal structure, Freshwater Biology, 61(10), 1802-1814, https://doi.org/10.1111/fwb.12819, 2016.

Giling, D.P., Staehr, P.A., Grossart, H.P., Andersen, M.R., Boehrer, B., Escot, C., Evrendilek, F., Gómez- Gener, L., Honti, M., Jones, I.D. and Karakaya, N.: Delving deeper: Metabolic processes in the metalimnion of stratified lakes, Limnology and Oceanography, 62(3), 1288-1306, https://doi.org/10.1002/lno.10504, 2017.

Goudsmit, G.H., Burchard, H., Peeters, F. and Wüest, A.: Application of k-ϵ turbulence models to enclosed basins: The role of internal seiches, Journal of Geophysical Research: Oceans, 107(C12), 23-1, https://doi.org/10.1029/2001JC000954, 2002.

Gray, E., Mackay, E.B., Elliott, J.A., Folkard, A.M. and Jones, I.D.: Wide-spread inconsistency in estimation of lake mixed depth impacts interpretation of limnological processes, Water Research, 168, 115136, https://doi.org/10.1016/j.watres.2019.115136, 2020.

Heinz, G., Imberger, J. and Schimmele, M. Vertical mixing in Überlinger See, western part of Lake Constance, Aquatic Science, 52, 256–268, https://doi.org/10.1007/BF00877283, 1990.

Hondzo, M. and Stefan, H.G.,: Regional water temperature characteristics of lakes subjected to climate change, Climatic change, 24(3), 187-211, https://doi.org/10.1007/BF01091829, 1993.

Huisman, J.E.F., van Oostveen, P. and Weissing, F.J.: Critical depth and critical turbulence: two different mechanisms for the development of phytoplankton blooms, Limnology and oceanography, 44(7), 1781-1787, https://doi.org/10.4319/lo.1999.44.7.1781, 1999.

Imberger, J.: The diurnal mixed layer, Limnology and oceanography, 30(4), 737-770, https://doi.org/10.4319/lo.1985.30.4.0737, 1985.

Jassby, A. and Powell, T.:Vertical patterns of eddy diffusion during stratification in Castle Lake, California, Limnology and oceanography, 20(4), 530-543, https://doi.org/10.4319/lo.1975.20.4.0530, 1975.

Jennings, E., Jones, S., Arvola, L., Staehr, P.A., Gaiser, E., Jones, I.D., Weathers, K.C., Weyhenmeyer, G.A., ~~CHIU~~Chiu, C.Y. and De Eyto, E.: Effects of weather-related episodic events in lakes: an analysis based on high- frequency data, Freshwater Biology, 57(3), 589-601, https://doi.org/10.1111/j.1365-2427.2011.02729.x, 2012.

Jennings, E., de Eyto, E., Laas, A., Pierson, D., Mircheva, G., Naumoski, A., Clarke, A., Healy, M., Šumberová, K. and Langenhaun, D.: The NETLAKE Metadatabase—A Tool to Support Automatic Monitoring on Lakes in Europe and Beyond, Limnology and Oceanography Bulletin, 26(4), 95-100, https://doi.org/10.1002/lob.10210, 2017.

Jonas, T., Stips, A., Eugster, W. and Wüest, A., 2003. Observations of a quasi shear- free lacustrine convective boundary layer: Stratification and its implications on turbulence, Journal of Geophysical Research: Oceans, 108(C10), https://doi.org/10.1029/2002JC001440, 2003.

Kara, A.B., Rochford, P.A. and Hurlburt, H.E.: Mixed layer depth variability over the global ocean, Journal of Geophysical Research: Oceans, 108(C3), https://doi.org/10.1029/2000JC000736, 2003.

Kraemer, B.M., Anneville, O., Chandra, S., Dix, M., Kuusisto, E., Livingstone, D.M., Rimmer, A., Schladow, S.G., Silow, E., Sitoki, L.M. and Tamatamah, R.: Morphometry and average temperature affect lake stratification responses to climate change, Geophysical Research Letters, 42(12), 4981-4988, https://doi.org/10.1002/2015GL064097, 2015.

Kraemer, B.M.: Rethinking discretization to advance limnology amid the ongoing information explosion, Water Research, 115801, https://doi.org/10.1016/j.watres.2020.115801, 2020.

Lamont, G., Laval, B., Pawlowicz, R., Pieters, R. and Lawrence, G.A.: Physical mechanisms leading to upwelling of anoxic bottom water in Nitinat Lake, 17th ASCE Engineering Mechanisms Conference, 13-16, 2004, June

Livingstone, D.M.: Impact of secular climate change on the thermal structure of a large temperate central European lake, Climatic change, 57(1-2), 205-225, https://doi.org/10.1023/A:1022119503144, 2003.

Lorbacher, K., Dommenget, D., Niiler, P.P. and Köhl, A.: Ocean mixed layer depth: A subsurface proxy of ocean atmosphere variability, Journal of Geophysical Research: Oceans, 111(C7), https://doi.org/10.1029/2003JC002157, 2006.

Maberly, S.C., O'Donnell, R.A., Woolway, R.I., Cutler, M.E., Gong, M., Jones, I.D., Merchant, C.J., Miller, C.A., Politi, E., Scott, E.M. and Thackeray, S.J.: Global lake thermal regions shift under climate change, Nature Communications, 11(1), 1-9, https://doi.org/10.1038/s41467-020-15108-z, 2020.

MacIntyre, S.: Vertical mixing in a shallow, eutrophic lake: Possible consequences for the light climate of phytoplankton, Limnology and Oceanography, 38(4), 798-817, https://doi.org/10.4319/lo.1993.38.4.0798, 1993.

MacIntyre, S., and Melack, J.M.: Mixing Dynamics in Lakes Across Climatic Zones, in: Encyclopedia of inland waters, edited by: Linkens, G.E., Academic Press, Oxford, 603-612, https://doi.org/10.1016/B978-012370626-3.00040-5 , 2009.

Marcé, R., George, G., Buscarinu, P., Deidda, M., Dunalska, J., de Eyto, E., Flaim, G., Grossart, H.P., Istvanovics, V., Lenhardt, M. and Moreno-Ostos, E.: Automatic high frequency monitoring for improved lake and reservoir management, Environmental science & technology, 50(20), 10780-10794, https://doi.org/10.1021/acs.est.6b01604, 2016.

Martin, J.L., McCutcheon, S.C.: Hydrodynamics and Transport for Water Quality Modeling. Lewis Publ., New York, 1999

Monismith, S.G. and MacIntyre, S.: The surface mixed layer in lakes and reservoirs, in: Encyclopedia of inland waters, edited by: Linkens, G.E., Academic Press, Oxford 636-650, https://doi.org/10.1016/B978-012370626-3.00078-8 , 2009.

Martin, J.L. and McCutcheon, S.C.: Hydrodynamics and transport for water quality modelling, CRC press, 2018.

Moras, S., Ayala, A. I., and Pierson, D. C.: Historical modelling of changes in Lake Erken thermal conditions, Hydrol. Earth Syst. Sci., 23, 5001–5016, https://doi.org/10.5194/hess-23-5001-2019, 2019.

Nash, J.E. and Sutcliffe, J.V.: River flow forecasting through conceptual models part I—A discussion of principles, Journal of hydrology, 10(3), 282-290, https://doi.org/10.1016/0022-1694(70)90255-6, 1970.

Pernica, P., Wells, M.G. and MacIntyre, S.: Persistent weak thermal stratification inhibits mixing in the epilimnion of north-temperate Lake Opeongo, Canada, Aquatic sciences, 76(2), 187-201, https://doi.org/10.1007/s00027-013-0328-1, 2014.

~~Pujoni, D.G., Brighenti, L.S., Bezerra‐Neto, J.F., Barbosa, F.A., Assunção, R.M. and Maia‐Barbosa, P.M.: Modeling vertical gradients in water columns: A parametric autoregressive approach, Limnology and Oceanography: Methods, 17(5), 320-329, https://doi.org/10.1002/lom3.10316, 2019.~~

Persson, I. and Jones, I.D.: The effect of water colour on lake hydrodynamics: A modelling study, Freshwater Biology, 53(12), 2345-2355, https://doi.org/10.1111/j.1365-2427.2008.02049.x, 2008.

Pierson, D.C., Weyhenmeyer, G.A., Arvola, L., Benson, B., Blenckner, T., Kratz, T., Livingstone, D.M., Markensten, H., Marzec, G., Pettersson, K. and Weathers, K.: An automated method to monitor lake ice phenology. Limnology and Oceanography: Methods, 9(2), 74-83, https://doi.org/10.4319/lom.2010.9.0074, 2011.

Pujoni, D.G., Brighenti, L.S., Bezerra‐Neto, J.F., Barbosa, F.A., Assunção, R.M. and Maia‐Barbosa, P.M.: Modeling vertical gradients in water columns: A parametric autoregressive approach, Limnology and Oceanography: Methods, 17(5), 320-329, https://doi.org/10.1002/lom3.10316, 2019.

Rivetti, I., Boero, F., Fraschetti, S., Zambianchi, E. and Lionello, P.: Anomalies of the upper water column in the Mediterranean Sea, Global and Planetary Change, 151, 68-79, https://doi.org/10.1016/j.gloplacha.2016.03.001, 2017.

Read, J.S., Hamilton, D.P., Jones, I.D., Muraoka, K., Winslow, L.A., Kroiss, R., Wu, C.H. and Gaiser, E.: Derivation of lake mixing and stratification indices from high-resolution lake buoy data, Environmental Modelling & Software, 26(11), 1325-1336, https://doi.org/10.1016/j.envsoft.2011.05.006, 2011.

Read, J.S., Hamilton, D.P., Desai, A.R., Rose, K.C., MacIntyre, S., Lenters, J.D., Smyth, R.L., Hanson, P.C., Cole, J.J., Staehr, P.A. and Rusak, J.A.: Lake-size dependency of wind shear and convection as controls on gas exchange, Geophysical Research Letters, 39(9), https://doi.org/10.1029/2012GL051886, 2012.

Sachse, R., Petzoldt, T., Blumstock, M., Moreira Martinez, S., Pätzig, M., Rücker, J., Janse, J., Mooij, W. M., and Hilt, S.: 485 Extending one-dimensional models for deep lakes to simulate the impact of submerged macrophytes on water quality, Environ. Model Softw., 61, 410–423, https://doi.org/10.1016/j.envsoft.2014.05.023, 2014.

Sahoo, G.B., Schladow, S.G., Reuter, J.E., Coats, R., Dettinger, M., Riverson, J., Wolfe, B. and Costa-Cabral, M.: The response of Lake Tahoe to climate change, Climatic Change, 116(1), 71-95, https://doi.org/10.1007/s10584-012-0600-8, 2013.

Schwefel, R., Gaudard, A., Wüest, A. and Bouffard, D.: Effects of climate change on deepwater oxygen and winter mixing in a deep lake (Lake Geneva): Comparing observational findings and modeling, Water Resources Research, 52(11), 8811-8826, https://doi.org/10.1002/2016WR019194, 2016.

Shay, T.J. and Gregg, M.C.: Convectively driven turbulent mixing in the upper ocean, Journal of Physical Oceanography, 16(11), 1777-1798, https://doi.org/10.1175/1520-0485(1986)016%3C1777:CDTMIT%3E2.0.CO;2, 1986.

Somavilla, R., González-Pola, C. and Fernández-Diaz, J.: The warmer the ocean surface, the shallower the mixed layer. How much of this is true? Journal of Geophysical Research: Oceans, 122(9), 7698-7716, https://doi.org/10.1002/2017JC013125, 2017.

Strock, K.E., Theodore, N., Gawley, W.G., Ellsworth, A.C. and Saros, J.E.: Increasing dissolved organic carbon concentrations in northern boreal lakes: Implications for lake water transparency and thermal structure. Journal of Geophysical Research: Biogeosciences, 122(5), 1022-1035, https://doi.org/10.1002/2017JG003767, 2017.

Tedford, E.W., MacIntyre, S., Miller, S.D. and Czikowsky, M.J.: Similarity scaling of turbulence in a temperate lake during fall cooling, Journal of Geophysical Research: Oceans, 119(8), 4689-4713, https://doi.org/10.1002/2014JC010135, 2014.

Thiery, W., Davin, E.L., Panitz, H.J., Demuzere, M., Lhermitte, S. and Van Lipzig, N.: The impact of the African Great Lakes on the regional climate, Journal of Climate, 28(10), 4061-4085, https://doi.org/10.1175/JCLI-D-14-00565.1, 2015.

Thomson, R.E. and Fine, I.V.: Estimating mixed layer depth from oceanic profile data. Journal of Atmospheric and Oceanic Technology, 20(2), 319-329, https://doi.org/10.1175/1520-0426(2003)020<0319:EMLDFO>2.0.CO;2, 2003.

Woolway, R.I., Maberly, S.C., Jones, I.D. and Feuchtmayr, H.: A novel method for estimating the onset of thermal stratification in lakes from surface water measurements, Water Resources Research, 50(6), 5131-5140, https://doi.org/10.1002/2013WR014975, 2014.

Woolway, R.I. and Merchant, C.J.: Worldwide alteration of lake mixing regimes in response to climate change, Nature Geoscience, 12(4), 271, https://doi.org/10.1038/s41561-019-0322-x, 2019.

Winslow, L.A., Zwart, J.A., Batt, R.D., Dugan, H.A., Woolway, R.I., Corman, J.R., Hanson, P.C. and Read, J.S.: LakeMetabolizer: an R package for estimating lake metabolism from free-water oxygen using diverse statistical models, Inland Waters, 6(4), 622-636, https://doi.org/10.1080/IW-6.4.883, 2016.

Weathers, K.C., Hanson, P.C., Arzberger, P., Brentrup, J., Brookes, J., Carey, C.C., Gaiser, E., Gaiser, E., Hamilton, D.P., Hong, G.S. and Ibelings, B.: The Global Lake Ecological Observatory Network (GLEON): the evolution of grassroots network science, Limnology and Oceanography Bulletin, 22(3), 71-73, https://doi.org/10.1002/lob.201322371, 2013.

Wüest, A. and Lorke, A.. Small-scale hydrodynamics in lakes, Annual Review of fluid mechanics, 35(1), 373-412, https://doi.org/10.1146/annurev.fluid.35.101101.161220, 2003.

Wüest, A., and Lorke, A.: Small-Scale Turbulence and Mixing: Energy Fluxes in Stratified Lakes, Encyclopedia of Inland Waters, 628–635, https://doi.org/10.1016/B978-012370626-3.00084-3, 2009.

Yang, Y., Colom, W., Pierson, D. and Pettersson, K.: Water column stability and summer phytoplankton dynamics in a temperate lake (Lake Erken, Sweden), Inland waters, 6(4), 499-508, https://doi.org/10.1080/IW-6.4.874, 2016.

Yang, H. and Wang, F.: Revisiting the thermocline depth in the equatorial Pacific, Journal of climate, 22(13), 3856-3863, https://doi.org/10.1175/2009JCLI2836.1, 2009.

Zorzal-Almeida, S., Soininen, J., Bini, L.M. and Bicudo, D.C.: Local environment and connectivity are the main drivers of diatom species composition and trait variation in a set of tropical reservoirs, Freshwater biology, 62(9), 1551-1563, https://doi.org/10.1111/fwb.12966, 2017.

Tables

Table 1. Lake model performance evaluation, showing the percentage relative error (%), root mean squared error (°C), and Nash Sutcliffe efficiency, for Lough Feeagh (profiles = 1016, years = 5) and Lake Erken (profiles = 1449, years = 7).

Statistic	Lough Feeagh Calibration	Validation	Lake Erken Calibration	Validation
PRE (%)	-0.48	0.47	-1.85	1.36
RMSE (°C)	0.67	1.18	0.53	0.55
NSE	0.97	0.92	0.98	0.97

Table 2. Summary of statistics for each method, showing the ~~average (m),~~mean (m) and standard error of mean in brackets, minimum (i.e. shallowest estimate) (m), maximum (i.e. deepest estimate) (m) and interquartile range (m) of the April-October epilimnion depth estimates (summarised from the results shown in Fig.6a), and the ~~average~~mean Pearson's correlation coefficient (r) for each method, representing the ~~average~~mean correlation for all possible combinations between threshold values., for Lough Feeagh and Lake Erken.

	Lough Feeagh					**Lake ~~Eken~~Erken**				
Method	~~Average~~ Mean (m)	Min (m)	Max (m)	~~Range~~IQR (m)	r	~~Average~~ Mean (m)	Min (m)	Max (m)	~~Range~~IQR (m)	r
M1	-~~18.9~~19.0 (0.8)	-4.6	-25.4	~~20.8~~7.3	0.77	-10.0 (0.2)	-7.8	-11.2	1.3~~.4~~	0.92
M2	-35.9 (0.9)	-19.7	-41.4	~~21.7~~6.5	0.48	-11.3 (0.2)	-8.4	-12.9	~~4.5~~1.7	0.78
M3	-36.5 (0.8)	-22.4	-41.5	~~19~~6.1	0.49	-11.8 (0.2)	-10.0	-13.0	~~3.0~~1.2	0.82
M4	-21.1 (0.5)	-19.7	-21.5	~~1.8~~0.3	1.00	-11.9 (0.1)	-10.1	-11.3	~~1.2~~0.5	0.99

Table 3. ~~Average~~Mean of all Pearson's correlation coefficients calculated for each pair of methods (e.g. for all threshold combinations between M1 and M2 etc.), for Lough Feeagh and Lake Erken.

Method	Lough Feeagh r	Lake Erken r
M2-M1	0.35	0.77

M3-M1	0.33	0.74
M3-M2	0.55	0.8
M4-M1	0.68	0.81
M4-M2	0.29	0.71
M4-M3	0.3	0.75

Table 4. ~~Average~~Mean Apr-Oct epilimnion depth estimates (m) derived using high resolution and low resolution modelled water temperature data with standard error of the mean in brackets, and the difference calculated between the high resolution and low resolution estimate (m), for Lough Feeagh and Lake Erken.

Method	**Lough Feeagh**			**Lake Erken**		
	High resolution ~~average~~ mean (m)	Low resolution ~~average~~mean (m)	Difference in mean epilimnion depth (m)	High resolution ~~average~~ mean (m)	Low resolution ~~average~~mean (m)	Difference in mean epilimnion depth (m)
M1	-22.1 (0.7)	-22.2 (0.7)	0.1	-10.9 (0.1)	-10.9 (0.1)	0.0
M2	-31.7 (1.1)	-34.9 (1.2)	3.2	-11.9 (0.2)	-13.1 (0.2)	1.2
M3	-32.0 (1.1)	-35.2 (1.1)	3.2	-12.1 (0.2)	-13.1 (0.2)	1.0
M4	-22.1 (0.6)	-22.6 (0.6)	0.5	-11.3 (0.1)	-11.5 (0.1)	0.2

Table 5. ~~Average difference~~Difference in ~~average~~mean Apr-Oct epilimnion depth estimates (m) between each profile based method (M1-M4) calculated using lake modelled data and the modelled turbulence method (M5) and ~~average~~mean of all Pearson's correlation coefficients (r) calculated for each profile-based method and M5 (e.g. for all threshold combinations between M5 and M1 etc.), for Lough Feeagh and Lake Erken. Positive ~~average~~ differences indicate that the modelled turbulence method was shallower.

Method	Lough Feeagh		Lake Erken	
	Difference in ~~average~~mean epilimnion depth (m)	r	Difference in ~~average~~mean epilimnion depth (m)	r
M5-M1	1.3	0.90	0.0	0.89
M5-M2	11.0	0.55	1.0	0.73
M5-M3	11.2	0.54	1.1	0.72
M5-M4	1.3	0.88	0.4	0.85

Figures

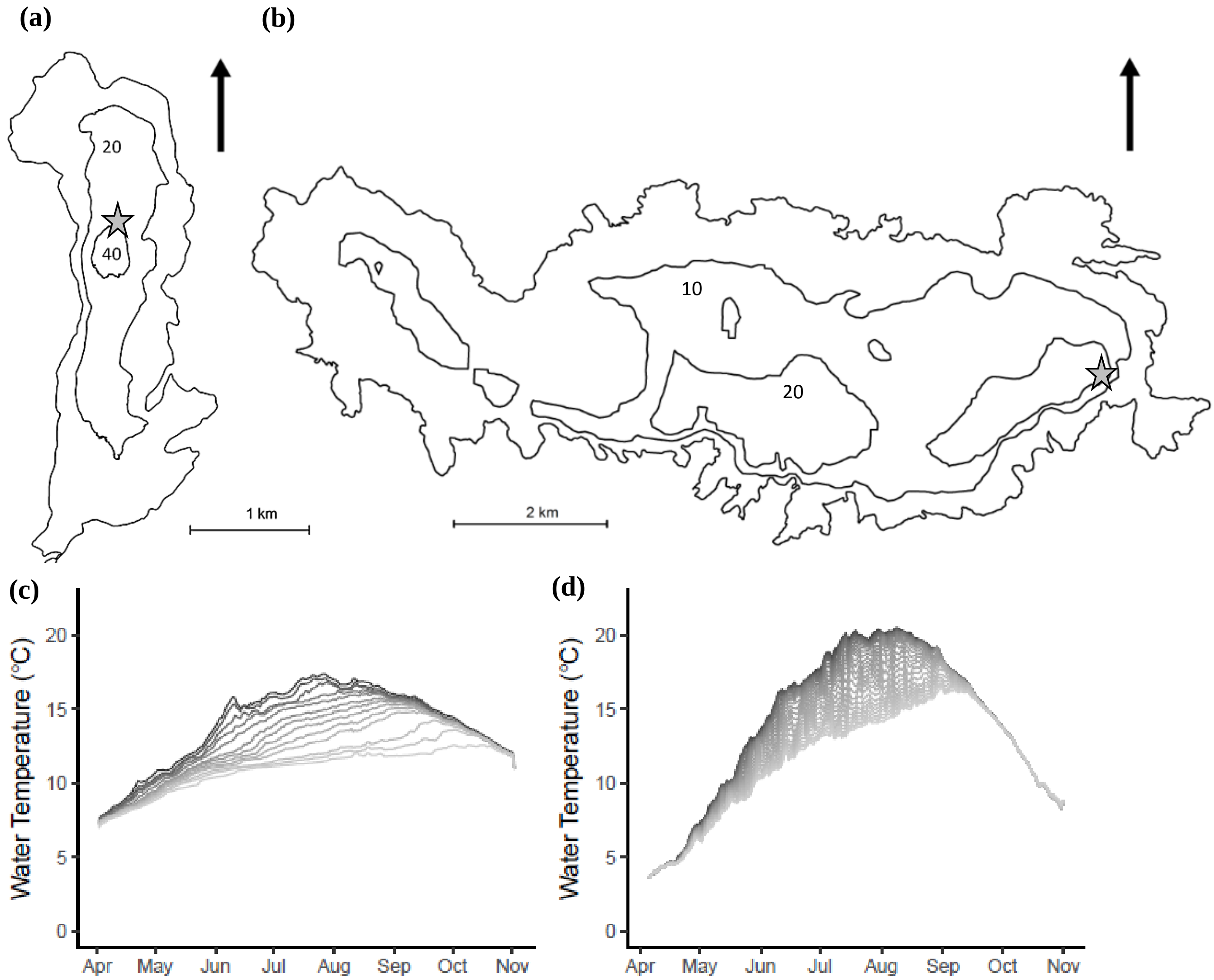


Figure 1. Bathymetric map of Lough Feeagh in Ireland **(a)** and Lake Erken in Sweden **(b)**, where the grey stars shows the locations of the automatic monitoring buoys used for measuring high-frequency water temperature profiles in both lakes, and long-term ~~average~~mean water temperature for each Julian day for all measured depths, for Lough Feeagh (profiles = 2778, years = 10) **(c)** and Lake Erken (profiles = 2005, years = 7) **(d)**.

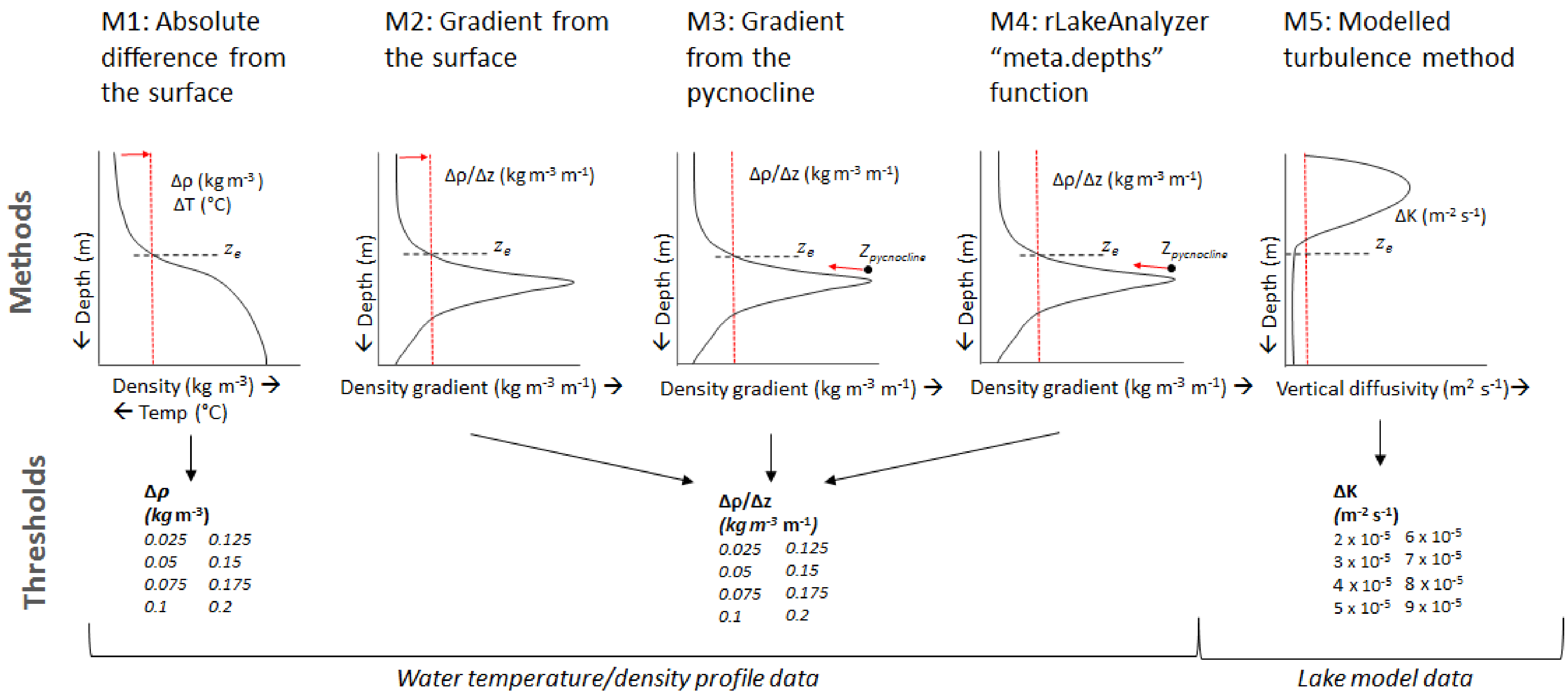


Figure 2. Schematic of epilimnion depth methods used in this study, including the range of threshold values for each method and the input data type (i.e. water temperature/density profile data or lake modelled data).

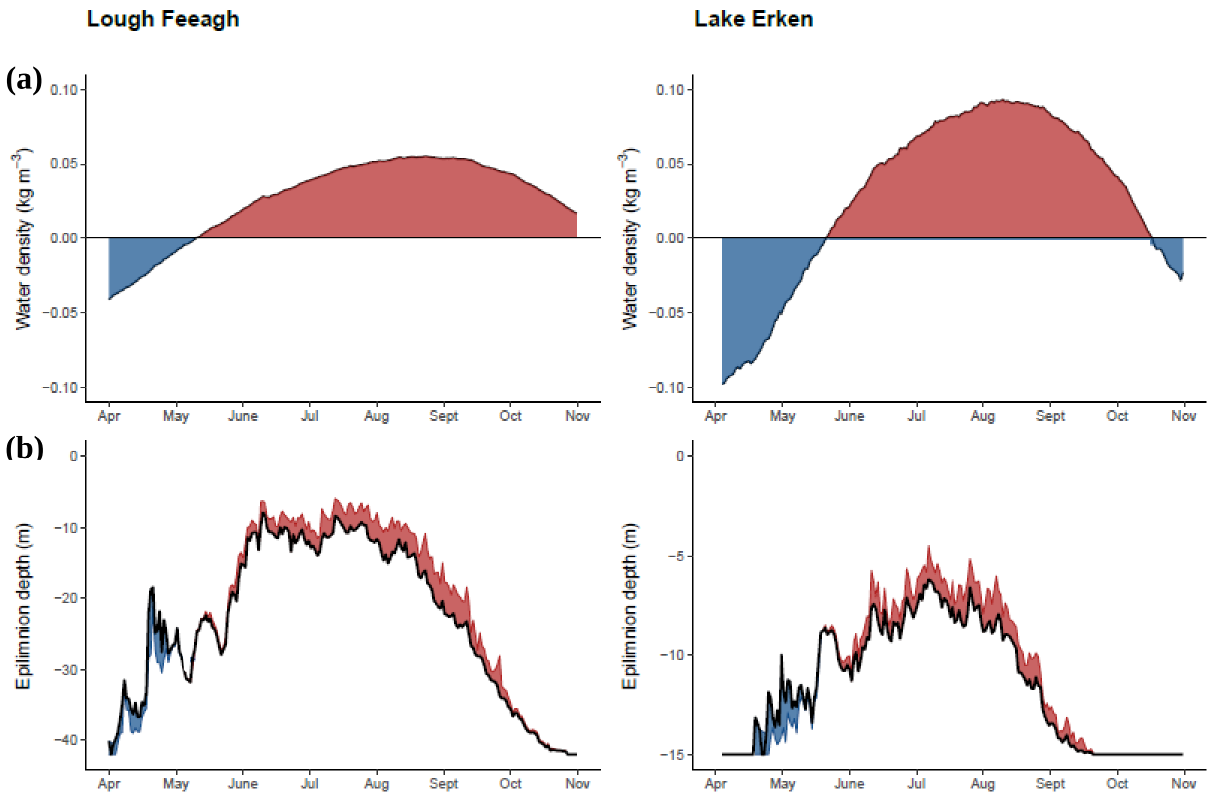


Figure 3. Long-term ~~average~~mean for each Julian day of the difference in water density (kg m^{-3}) induced by an increase of 1 °C in water temperature, relative to 0.1 kg m^{-3} (black line with the red shaded area demonstrating when the change induced by an increase of 1 °C change was greater than 0.1 kg m^{-3} and the blue shaded area for when it was less than 0.1 kg m^{-3}) **(a)**, and the long-term ~~average~~mean for each Julian day of epilimnion depth calculated using a water temperature threshold of 1 °C (the black line), compared to a water density threshold of 0.1 kg m^{-3} (shaded area, with the red shaded areas demonstrating when water density estimates were shallower and the blue shaded area for when they were deeper) **(b)**, for Lough Feeagh and Lake Erken.

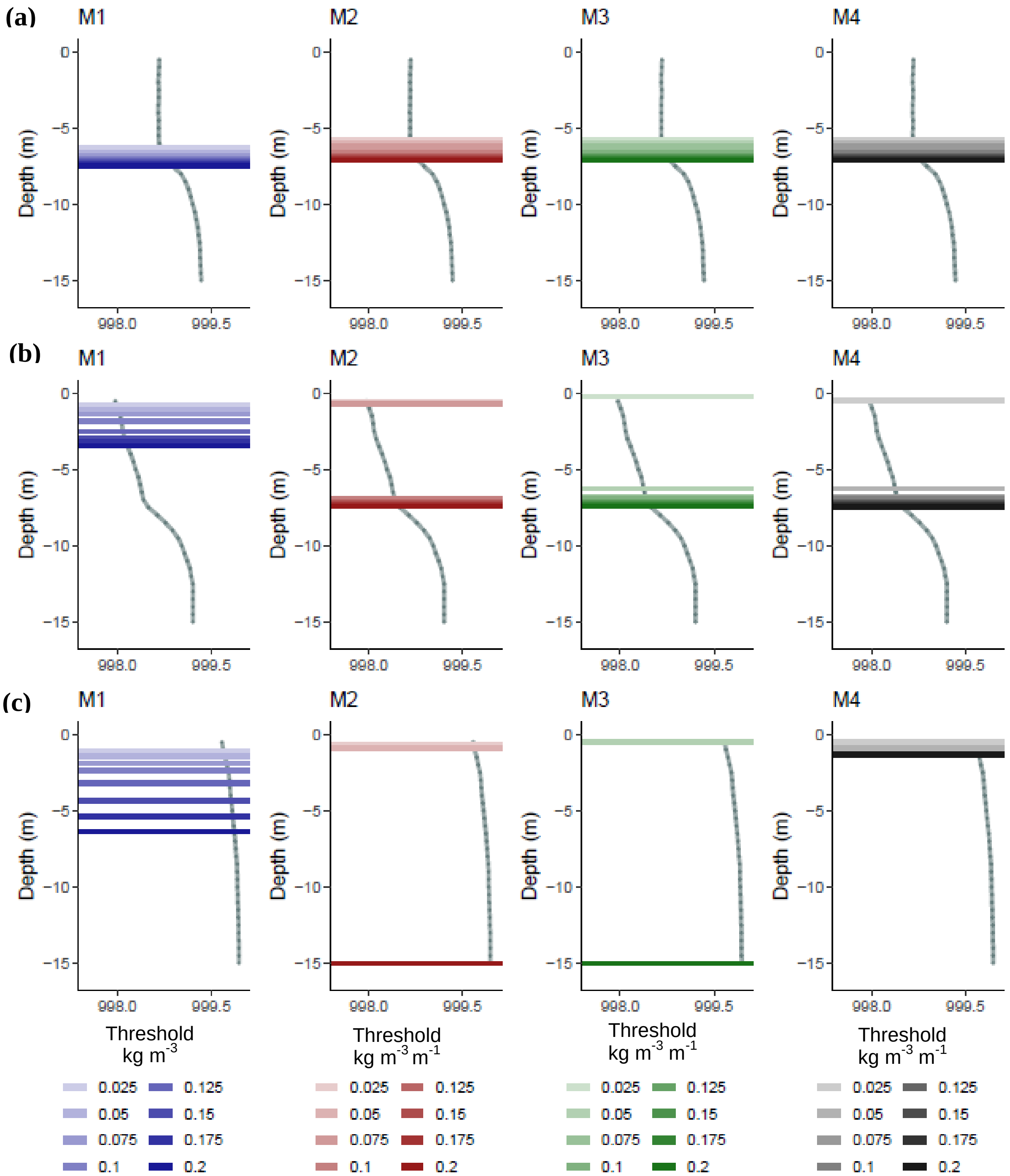


Figure 4. An example of water column profiles with the epilimnion depth estimates superimposed (horizontal lines) for all for all profile-based epilimnion depth methods calculated using the full range of thresholds for each. The water columns can be categorised as a three-layered water column structure **(a)**, an intensely stratified profile **(b)**, and a near-isothermal profile **(c)**, all from Lake Erken only.

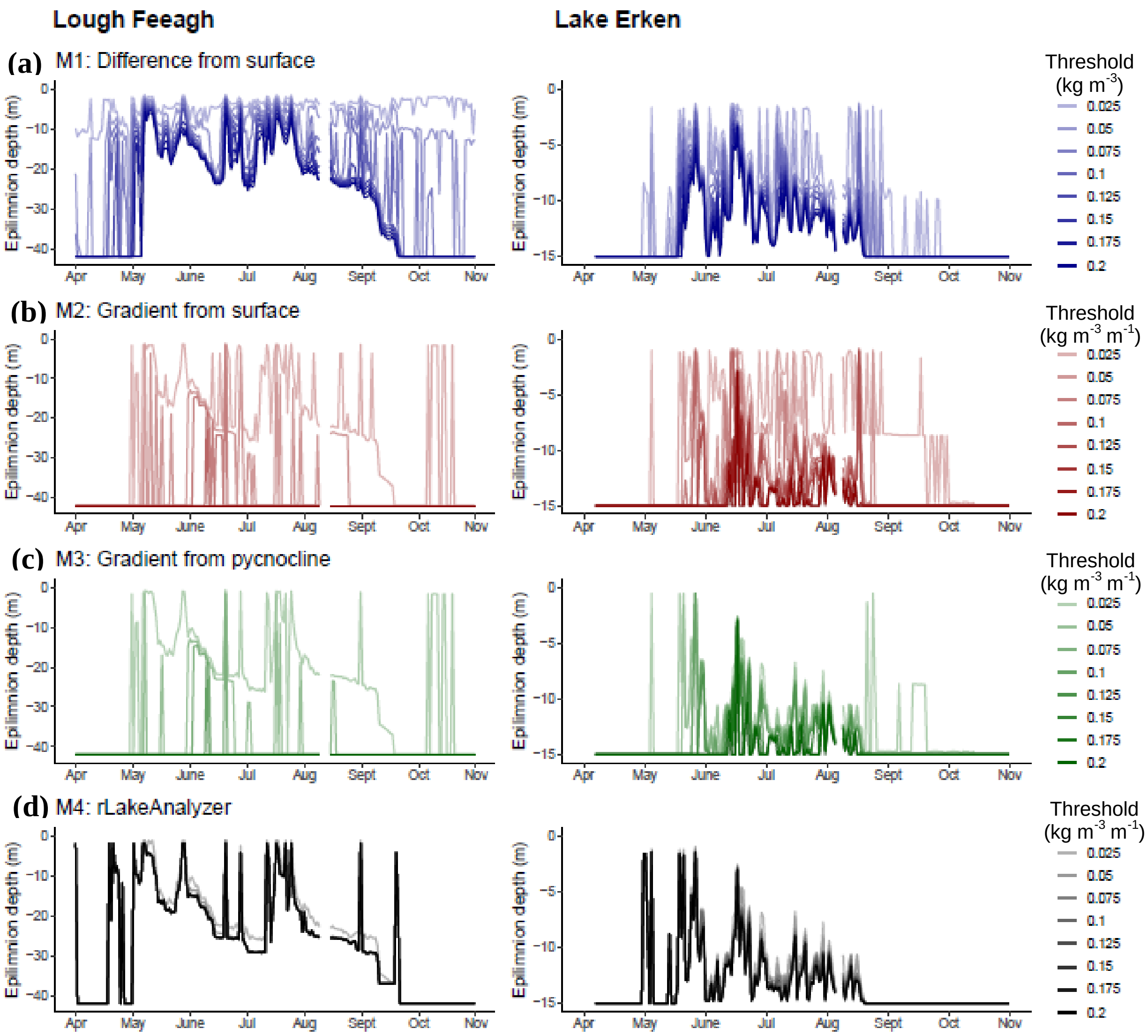


Figure 5. Daily epilimnion depth estimates using measured data for 2017 from Lough Feeagh and Lake Erken, showing estimates from all profile-based epilimnion depth methods, including M1, the absolute difference from the surface method **(a)**, M2, the gradient from the surface method **(b)**, M3, the gradient from the pycnocline method **(c)** and M4, the rLakeAnalyzer method **(d)**, calculated using the full range of thresholds, and for each lake.

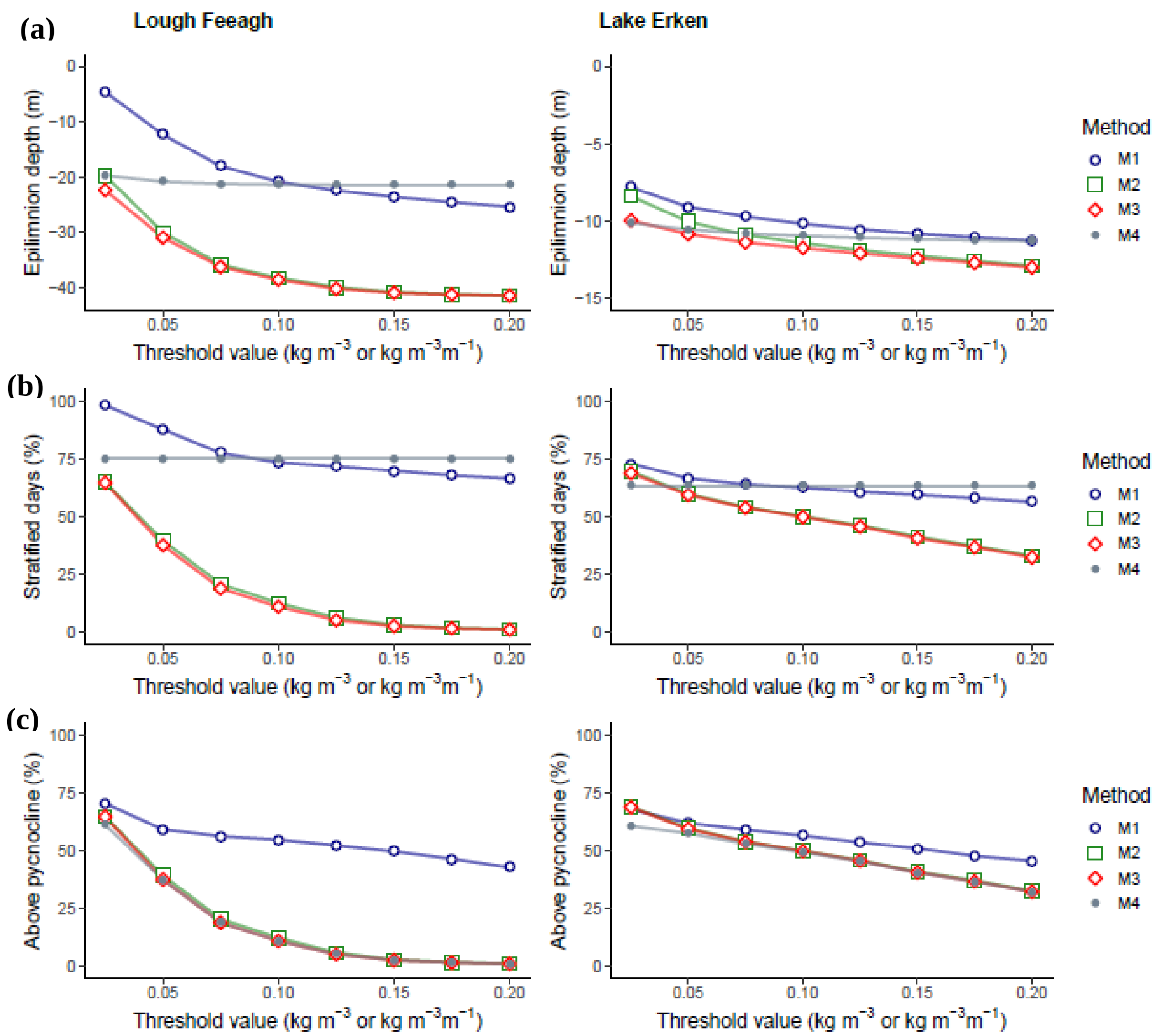
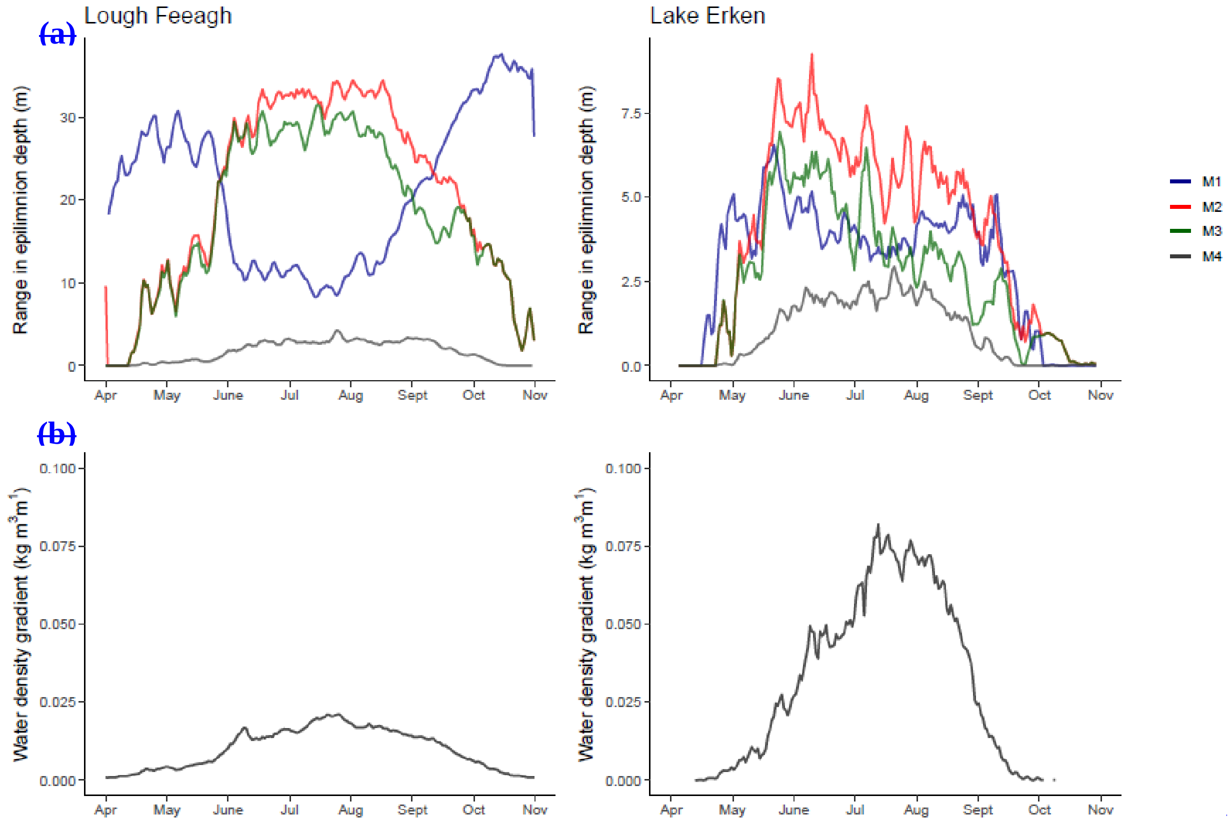


Figure 6. ~~Average~~Mean Apr-Oct epilimnion depth (m) **(a)**, percentage of stratified days, defined as days where the epilimnion depth was identified ~~at a depth greater~~as shallower than the lake maximum measured depth (%) (where a larger percentage value indicated a higher occurrence of days identified as stratified) **(b)**, and percentage of days where the epilimnion depth was above the pycnocline, defined as the number of days where the epilimnion was identified at a depth shallower than the maximum density gradient (where a larger percentage value indicated a lower occurrence of days erroneously extending into the metalimnion) **(c)**, for Lough Feeagh and Lake Erken.



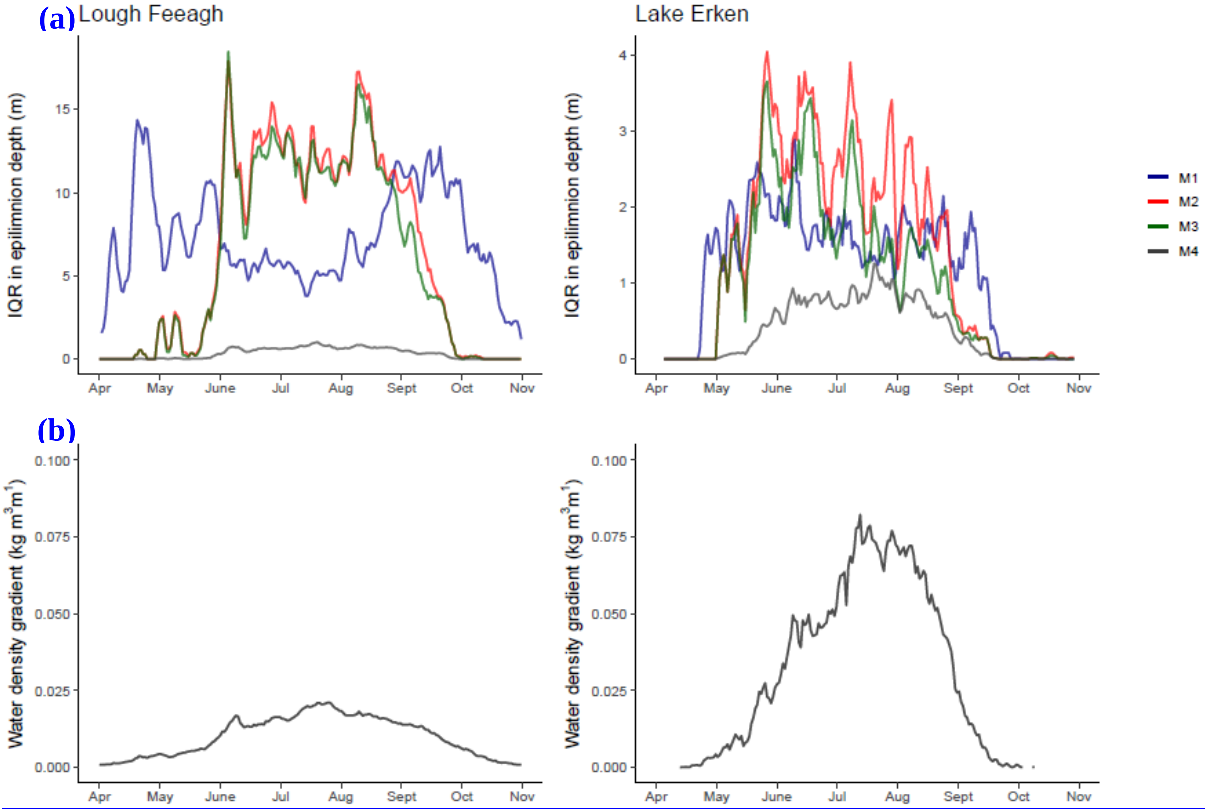


Figure 7. ~~Range~~Inter-quartile range between the shallowest and deepest estimate for each method calculated from long-term daily ~~average~~mean epilimnion depth estimates for each Julian day, where a large range suggests high threshold sensitivity and a small range suggests low sensitivity **(a)**, and long-term daily ~~average~~mean water density gradient, calculated based on the surface and maximum measured depths **(b)**, for Lough Feeagh and Lake Erken.

ADDITIONAL SUPPLEMENT DOCUMENT

Supplementary material of HESS-2020-222

Table S1. Lake model parameters and calibrated values.

Parameter	Lake Erken	Lough Feeagh
Shf_factor	0.88	0.77
Swr_factor	0.98	0.93
Wind_factor	1.41	1.31
K_min	1.86e-6	3.48e-6
G2	2.14	0.56

Table S2: Pearson's correlation coefficient (r) matrices for all methods and all threshold combinations for Lough Feeagh and Lake Erken.

Lough Feeagh

		Method 1								Method 2								Method 3								Method 4							
		0.03	0.05	0.08	0.1	0.13	0.15	0.18	0.2	0.03	0.05	0.08	0.1	0.13	0.15	0.18	0.2	0.03	0.05	0.08	0.1	0.13	0.15	0.18	0.2	0.03	0.05	0.08	0.1	0.13	0.15	0.18	0.2
Method 1	0.025																																
	0.05	.59																															
	0.075	.46	.82																														
	0.1	.42	.76	.95																													
	0.125	.41	.73	.91	.98																												
	0.15	.39	.71	.89	.95	.99																											
	0.175	.38	.69	.86	.93	.97	.99																										
	0.2	.37	.67	.85	.91	.95	.97	.99																									
Method 2	0.025	.41	.68	.73	.74	.76	.78	.79	.79																								
	0.05	.26	.48	.61	.65	.68	.70	.72	.74	.64																							
	0.075	.16	.30	.39	.43	.46	.49	.51	.53	.40	.67																						
	0.1	.11	.22	.29	.33	.36	.38	.40	.42	.30	.51	.77																					
	0.125	.08	.16	.22	.25	.27	.29	.30	.32	.22	.38	.58	.75																				
	0.15	.07	.13	.17	.21	.23	.24	.25	.26	.18	.31	.46	.60	.79																			
	0.175	.06	.11	.15	.17	.19	.21	.22	.23	.15	.26	.38	.49	.66	.83																		
	0.2	.05	.09	.13	.15	.17	.18	.19	.19	.13	.22	.33	.42	.56	.71	.86																	
Method 3	0.025	.39	.64	.68	.71	.74	.77	.78	.79	.94	.61	.43	.34	.26	.22	.19	.16																
	0.05	.25	.46	.58	.63	.66	.68	.70	.72	.61	.95	.71	.55	.42	.34	.28	.24	.61															
	0.075	.15	.28	.37	.42	.45	.48	.50	.52	.38	.64	.97	.80	.60	.48	.40	.34	.43	.69														
	0.1	.11	.21	.27	.31	.34	.36	.38	.40	.29	.49	.73	.95	.79	.63	.53	.45	.33	.53	.76													
	0.125	.08	.15	.20	.23	.26	.27	.29	.30	.21	.36	.54	.70	.94	.85	.71	.61	.25	.39	.56	.74												
	0.15	.07	.12	.16	.20	.21	.23	.24	.25	.17	.29	.43	.56	.75	.94	.88	.75	.21	.32	.45	.60	.81											
	0.175	.06	.11	.14	.17	.18	.20	.21	.22	.14	.24	.36	.47	.63	.79	.95	.90	.18	.27	.38	.50	.68	.84										
	0.2	.05	.09	.12	.14	.16	.17	.18	.18	.12	.21	.31	.40	.54	.68	.81	.95	.15	.23	.32	.43	.57	.71	.85									
Method 4	0.025	.40	.70	.82	.81	.78	.75	.72	.70	.58	.50	.36	.29	.22	.19	.17	.15	.62	.50	.36	.28	.22	.19	.16	.14								
	0.05	.39	.68	.79	.78	.75	.72	.69	.68	.55	.48	.36	.29	.23	.20	.18	.15	.59	.49	.36	.29	.23	.20	.17	.15	1.00							
	0.075	.39	.67	.78	.77	.74	.71	.68	.66	.53	.46	.34	.29	.23	.21	.18	.16	.58	.47	.35	.28	.23	.20	.17	.15	.99	1.00						
	0.1	.39	.67	.78	.77	.74	.70	.67	.65	.53	.45	.33	.27	.23	.21	.18	.16	.58	.46	.33	.27	.22	.20	.17	.15	.99	1.00	1.00					
	0.125	.39	.67	.78	.77	.73	.70	.67	.65	.53	.44	.32	.26	.22	.20	.18	.15	.58	.45	.32	.26	.22	.19	.17	.15	.99	1.00	1.00	1.00				
	0.15	.39	.67	.78	.77	.73	.70	.67	.65	.53	.44	.32	.26	.22	.20	.18	.15	.58	.45	.32	.25	.21	.19	.17	.15	.99	1.00	1.00	1.00	1.00			
	0.175	.39	.67	.78	.77	.73	.70	.67	.65	.53	.44	.31	.26	.21	.19	.17	.15	.58	.45	.32	.25	.21	.19	.17	.14	.99	1.00	1.00	1.00	1.00	1.00		
	0.2	.39	.67	.78	.77	.73	.70	.67	.65	.53	.44	.31	.26	.21	.19	.17	.15	.58	.45	.32	.25	.21	.18	.16	.14	.99	1.00	1.00	1.00	1.00	1.00	1.00	

Lake Erken

		Method 1								Method 2								Method 3								Method 4							
		0.03	0.05	0.08	0.1	0.13	0.15	0.18	0.2	0.03	0.05	0.08	0.1	0.13	0.15	0.18	0.2	0.03	0.05	0.08	0.1	0.13	0.15	0.18	0.2	0.03	0.05	0.08	0.1	0.13	0.15	0.18	0.2
Method 1	0.025																																
	0.05	.94																															
	0.075	.89	.97																														
	0.1	.85	.94	.98																													
	0.125	.82	.91	.96	.99																												
	0.15	.79	.89	.94	.97	.99																											
	0.175	.77	.87	.92	.96	.98	1.00																										
	0.2	.75	.85	.90	.94	.97	.99	1.00																									
Method 2	0.025	.85	.92	.92	.89	.87	.85	.83	.81																								
	0.05	.78	.87	.92	.94	.95	.95	.95	.94	.85																							
	0.075	.70	.78	.84	.87	.90	.92	.92	.93	.75	.90																						
	0.1	.64	.73	.78	.82	.85	.87	.88	.89	.70	.84	.94																					
	0.125	.59	.67	.72	.75	.78	.81	.83	.85	.64	.77	.86	.93																				
	0.15	.55	.62	.66	.70	.73	.76	.78	.80	.59	.71	.80	.87	.94																			
	0.175	.50	.57	.61	.64	.67	.69	.72	.74	.54	.65	.74	.80	.87	.93																		
	0.2	.45	.52	.55	.58	.61	.63	.66	.68	.49	.59	.67	.73	.80	.86	.93																	
Method 3	0.025	.75	.83	.86	.87	.88	.88	.88	.88	.84	.86	.82	.79	.73	.69	.64	.58																
	0.05	.72	.81	.86	.89	.91	.92	.93	.93	.78	.92	.89	.86	.81	.77	.72	.66	.91															
	0.075	.66	.74	.79	.83	.86	.88	.90	.91	.71	.85	.94	.92	.88	.84	.78	.72	.83	.92														
	0.1	.62	.70	.75	.78	.82	.84	.86	.87	.67	.80	.90	.96	.92	.88	.82	.77	.79	.88	.95													
	0.125	.58	.65	.70	.73	.76	.79	.81	.83	.63	.75	.84	.90	.97	.94	.88	.82	.73	.82	.89	.94												
	0.15	.53	.61	.65	.68	.71	.73	.76	.78	.58	.69	.78	.84	.91	.97	.93	.87	.68	.76	.84	.88	.94											
	0.175	.49	.56	.59	.63	.65	.68	.70	.72	.53	.64	.72	.78	.85	.91	.98	.94	.63	.71	.78	.82	.88	.93										
	0.2	.45	.51	.54	.57	.60	.62	.64	.66	.48	.58	.66	.72	.78	.84	.91	.98	.57	.65	.72	.76	.81	.87	.93									
Method 4	0.025	.78	.83	.86	.88	.88	.88	.88	.88	.77	.84	.82	.79	.74	.70	.65	.59	.90	.90	.84	.80	.74	.69	.64	.58								
	0.05	.76	.81	.83	.85	.86	.86	.86	.86	.74	.82	.81	.79	.75	.71	.66	.61	.88	.90	.85	.81	.76	.71	.65	.60	.99							
	0.075	.74	.79	.82	.83	.84	.84	.84	.84	.72	.80	.79	.78	.74	.71	.66	.62	.86	.88	.85	.81	.76	.71	.66	.61	.98	.99						
	0.1	.73	.78	.81	.82	.83	.83	.83	.83	.71	.79	.78	.76	.73	.70	.66	.62	.85	.87	.83	.80	.75	.71	.66	.61	.97	.99	1.00					
	0.125	.73	.77	.80	.81	.82	.82	.82	.82	.70	.78	.77	.75	.72	.70	.66	.61	.84	.86	.82	.79	.74	.70	.66	.61	.96	.98	.99	1.00				
	0.15	.72	.77	.79	.81	.81	.81	.81	.81	.69	.77	.75	.74	.71	.69	.65	.61	.84	.86	.81	.78	.73	.69	.65	.61	.96	.98	.99	1.00	1.00			
	0.175	.72	.76	.79	.80	.81	.81	.81	.80	.69	.76	.75	.73	.70	.67	.64	.60	.83	.85	.80	.77	.72	.68	.64	.60	.95	.98	.99	.99	1.00	1.00		
	0.2	.71	.76	.78	.80	.80	.80	.80	.80	.69	.76	.74	.72	.69	.67	.63	.59	.83	.85	.80	.76	.71	.67	.63	.59	.95	.98	.99	.99	1.00	1.00	1.00	

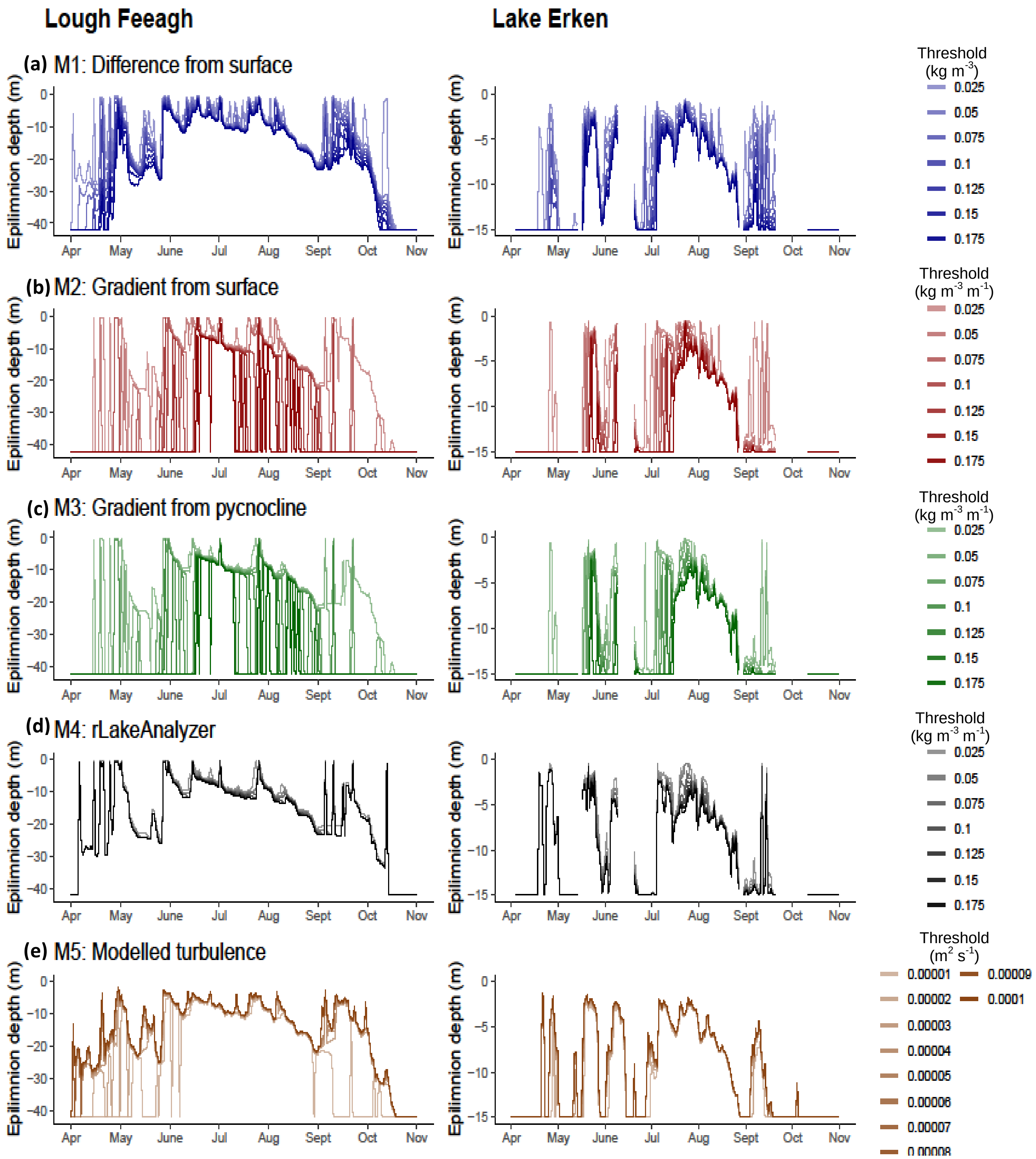


Figure S1: Daily epilimnion depth estimates using modelled data for 2016 from Lough Feeagh and Lake Erken, showing estimates from all profile based epilimnion depth methods, including M1, the absolute difference from the surface method **(a)**, M2, the gradient from the surface method **(b)**, M3, the gradient from the pycnocline method **(c)** and M4, the rLakeAnalyzer method **(d)**, as well as M5, the modelled turbulence based method **(e)**, calculated using the full range of thresholds, and for each lake.