Response to Referee #2

We thank the reviewer for the constructive assessment and for recognising the value of upscaling monitoring-station data to the national scale. We agree with the central methodological concern and outline below how we will restructure the treatment of climate scenarios and add the recommended trend analysis in the revised manuscript.

RC (general comment & L190–192): *For future predictions, water temperature predicted from various scenarios were aggregated to get their median on which analysis was done. According to the standard IPCC approach, the predicted temperatures should only be aggregated across different models for the same scenario. Please consider either selecting one scenario for this study or repeating the analysis separately for different scenarios.*

AC: We agree, and we thank the reviewer for this important point. In the revised manuscript we will adopt a single, consistent scenario throughout. The main text will be based exclusively on **SSP2-4.5**, using the ensemble of six GCMs (CMCC-ESM2, EC-EARTH3, GFDL-ESM4, MPI-ESM1-2-HR, MRI-ESM2-0, TAIESM1) together with the regional model ALADIN. All figures, tables, thresholds and heat-wave metrics will be recomputed on this basis, so that ensemble aggregation occurs strictly across models within one scenario, consistent with standard IPCC practice.

The original multi-scenario construct will not be discarded but moved to the Supplementary Material and reframed accordingly. There it will be presented only as a **scenario-spanning ensemble summary**, i.e. a risk-screening indicator across the range of available SSP projections, and explicitly *not* as a probabilistic central estimate. We will remove all wording (e.g. "multi-scenario ensemble median") that could imply a formal central tendency across scenarios.

RC (general comment & Section 3.5): *I suggest using trend analysis to consistently characterise the changes in annual number of days with different ranges of water temperature and assess their significance.*

AC: We will add a formal trend analysis applied consistently across all thermal metrics over 2016–2100, using the non-parametric Mann-Kendall test together with Sen's slope, computed per site with a per-model robustness check so that significance reflects agreement across ensemble members. Our analysis indicates, for example, that the number of days below 17 °C declines at approximately −3.67 days per decade with a significant trend at 100 % of sites, whereas heat-wave metrics show near-zero median slopes with significance at roughly 55–64 % of sites, which is reinforcing the manuscript's central message that chronic sub-lethal warming is more pervasive and more robust than acute events. These results will be reported in a new trend table and trend-map panel.

RC (data & code availability): *I also suggest making the code and data publicly available.*

AC: We agree and will make the processed data and analysis code openly available. Where specific input datasets cannot be redistributed we will state this explicitly and provide the source and access conditions rather than the raw data themselves.

Introduction

RC (L47–53): *Please check the references correctly. The papers cited do not support the statements in some cases. For example, introduction states "Yet, equally important and often overlooked are changes in cold-season thermal conditions during autumn and early spring, which govern the most temperature-sensitive life stages of salmonids: spawning, egg incubation, and fry emergence (Sadayappan and Li, 2025; Van Vliet et al., 2023)." Both cited papers deal with changes in river water temperature and not with how the changes impact aquatic life. Check the full manuscript to ensure the right citations are inserted in the right places.*

AC: We thank the reviewer for this careful check and agree. The two references support statements about thermal-regime change, not the ecological sensitivity of salmonid life stages. In the revised manuscript we will (i) keep these citations only where they support thermal-regime statements, and (ii) support the claim about temperature-sensitive life stages (spawning, egg incubation, fry emergence) with appropriate salmonid-ecology references. We will also audit citations throughout the manuscript to confirm that each supports the specific statement it accompanies.

RC (L91–92): *I suggest making the hypothesis clear. Currently, the phrasing suggests that the "increasing summer thermal stress" was quantified but the paper quantifies only the change in water temperature during Feb-May and Oct-Jan.*

AC: We agree the current phrasing overstates the scope. We will revise the objectives so they refer specifically to projected changes in the thermal regime, including the cold-season windows (Feb-May and Oct-Jan), rather than to "summer thermal stress" as a directly quantified impact. We will make explicit that the study quantifies changes in water temperature and threshold exceedance, not biological stress responses per se.

Methods

RC: *Please add the location of the monitoring sites in the Figure 1 map.*

AC: We will add the locations of the monitoring sites to the Figure 1 map.

RC: *Please add the number of data points available for each site in Table 1.*

AC: We will add the record length / number of available observations for each site to Table 1.

RC (L118): *Is the year 2002 or 2004? According to Table 1, the earliest sensor data is from 2004.*

AC: Thank you for catching this inconsistency. The correct start of the sensor record is 2004, consistent with Table 1; we will correct the text accordingly. In fact, monitoring records begin between 2004 and 2012 depending on the profile while all profiles share a common period of at least 2012–2022, which includes the recent warm summers (e.g. 2015, 2018), so the calibration captures comparable thermal extremes across sites.

RC (L125): *The use of visual inspection in data cleaning can compromise the reproducibility of the workflow. I suggest encoding the rules used to flag and remove data. For example, what temperatures were considered as "implausible" for particular river or what conditions - flat temperature for long period of time or sudden spikes, etc. – were considered to indicate sensor failure? This data cleaning workflow should be made available in code.*

AC: We agree. In the revised Methods we will specify the quality-control rules.

RC (L159–160): *This is already in "Results" and can be removed here.*

AC: We will remove these lines from the Methods, as they belong in Results.

RC (L161–165): *It is not clear for which purpose global model versus site-specific model was used. Why was the site-specific model not calibrated on part of data (70 or 80%) and tested on rest?*

AC: We thank the reviewer and will clarify the complementary roles of the two models in the revised Methods:

- The **global model** = a single transferable regression (T_water = 2.91 + 0.69 × T_air), is applied to any grid cell of the river network, including unmonitored reaches. It therefore underlies all spatial products (Figs. 1-3) and the river-length calculations, since covering the ungauged network requires one universal relationship.
- The **site-specific model** = parameterised per profile, provides the best local fit at each of the 35 monitored profiles and is used for all profile-level analyses (Figs. 4-9): threshold days, trends, and seasonal windows.

Regarding the calibration/validation design: the global model was evaluated with a split approach, calibrated on 25 profiles and validated on the 10 withheld profiles (R^2 = 0.93 on validation), precisely to test its transferability to unmonitored locations. The site-specific model deliberately uses the full record at each profile, because its purpose is not to test transferability (that is the role of the global model) but to capture the full range of local variability, including extremes. A within-site split would discard data needed to represent those extremes faithfully. We will state this rationale explicitly.

RC (L175): *Can you explain more about how these models were evaluated and summarize the evaluation metrics for different models?*

AC: We will add a summary table of calibration and validation metrics. In preparing this, we identified that the RMSE range reported in the current manuscript (1.5–3.1 °C) originated from a superseded model version (dated 10/2024) and is not consistent with the model actually used. We have recomputed all metrics consistently (full-year) and will report the corrected values:

- **Global model:** median RMSE 1.74 °C (range 1.34–3.43 °C)
- **Site-specific model:** median RMSE 1.49 °C (range 1.13–2.05 °C), outperforming the global model at 32 of 35 profiles

We thank the reviewer, as revisiting the evaluation for this response is what prompted us to detect and correct this inconsistency.

RC (L190): *Do you mean the 35 river profiles here?*

AC: Yes. We will clarify that this refers to the 35 river profiles.

RC (L190–192): *Why were model ensembles under different scenarios combined? For example, SSP1-2.6 will be very different from SSP5-8.5 and therefore should be analyzed separately. It makes more sense to get ensemble average across different models under same scenario but not across all model-scenario combinations.*

AC: This is addressed in our response to the general comment above: the revised manuscript uses a single scenario (SSP2-4.5), so ensembles are no longer aggregated across scenarios.

RC (L192–196): *These are repeated in the next section and can be removed here.*

AC: We will remove the duplicated text.

RC (L204): *Please add references to support this definition of river heat wave. Please add some rationale on why a fixed threshold was used.*

AC: We will support the heat-wave definition (water temperature ≥ 23 °C for ≥ 3 consecutive days) with literature and add the rationale. Consecutive days above a threshold as a heat-wave descriptor follows Turschwell et al. (2017). We use a fixed rather than percentile-based threshold deliberately, because 23 °C corresponds to a physiological limit for brown trout (cardiac failure at approximately 23.5 °C; Vornanen et al., 2014) rather than to local statistical variability; a fixed physiological threshold is more meaningful for a species-focused assessment. The ≥ 3-day minimum reflects that prolonged exposure is required for measurable effects on cumulative mortality and depletion of energy reserves. We will also note duration-dependence of thermal tolerance (Wehrly et al., 2007: tolerance declines from ~25.3 °C for 1 day to ~21 °C over 21–63 days).

RC (L207–214): *Please explain the reason behind choosing 12 °C and 10 °C as thresholds for seasonal analyses along with references.*

AC: We will add the ecological justification for both seasonal thresholds:

- **10 °C (ONDJ, incubation window):** Elliott & Elliott (2010) report optimal egg-incubation temperatures of 1–8 °C with declining survival above 7–8 °C; Réalis-Doyelle et al. (2016) show survival to first feeding falling to ~55 % at 10 °C.
- **12 °C (FMAM, emergence window):** Réalis-Doyelle et al. (2016) report a sharp decline in well-formed fry survival at 12 °C; Elliott & Elliott (2010) identify ~12.8 °C as the upper bound of effective growth.

RC (L215): *Please clarify whether the global or site-specific air-water temperature relationships were used to predict water temperature in different stretches of rivers. If site-specific relationships were used, how was the relationship for each river stretch or grid was selected from among 35 relationships.*

AC: Prediction across the river network uses the **global** model, i.e. the single grid-based relationship. The site-specific relationships exist only at the 35 monitored profiles and cannot be applied to ungauged reaches; there is therefore no selection among the 35 relationships for network-wide prediction. We will make this explicit in the Methods.

RC (L220): *The air temperature was from a gridded product, yet in figure 1c the maps show river network. What is the source of river network data and what is its resolution?*

AC: The river network is based on the watercourse-axis layer (*Osy vodních toků*) obtained from the official Central Register of Watercourses (*Centrální evidence vodních toků*, CEVT) of the Czech Republic, administered by the Ministry of Agriculture and the Ministry of the Environment (2025 version). As a vector dataset it has no raster spatial resolution; the source data correspond to an approximate map scale of 1:10,000. Water temperature was computed on the 500 × 500 m grid via the global air–water model, and each reach of the CEVT network was then attributed the value of the grid cell it intersects (a spatial join), so the linear features inherit the gridded output rather than being modelled as lines directly. We will add this description to the Methods.

Results

RC (L246–255): *Please add the number of regimes in different classes.*

AC: We will report the number of profiles falling into each thermal class in this section (and in the corresponding table).

RC (Figure 3): *Figure 3 is supposed to show binary condition – whether each grid/stretch has <= 350 days of <17 °C. But maps show the annual number of days with <17 °C. Please correct it. I suggest using two colors to indicate stretches that met this condition and those that didn't..*

AC: Thank you, this is correct. Figure 3 will be redrawn as a binary two-colour map distinguishing reaches that meet the ≥ 350-day threshold from those that do not, replacing the current continuous day-count rendering.

RC (Figure 4*): Consider adding a subpanel for > 23 °C as last row so that readers can get a full picture. For example, legend of figure 5 indicates that annual number of days with > 23 °C can be as high as 40 days in some river profiles. Therefore, they can be substantial for some river profiles.*

AC: We agree and will add a > 23 °C row to Figure 4, so the full range from sub-lethal to lethal conditions is shown consistently.

RC (Section 3.5): *Why were trends and significance of trend in different classes and thermal ranges not calculated? That could give a more comprehensive picture.*

AC: We agree this strengthens the analysis and, as noted in our response to the general comment, we will add a formal per-site trend analysis (Mann-Kendall with Sen's slope, 2016–2100) applied consistently across all thermal classes and ranges, reported with significance in a new trend table and trend-map panel.

RC (L284–286): *How was this trend calculated and over what time period? How can this trend indicate that the shift is already underway? Was it based only on historical data reconstructed from air-water temperature relationships?*

AC: We thank the reviewer for identifying this inconsistency. The rate of 1.8 days per decade is the projected trend over 2016–2100 (Sen's slope, per profile), not an observed historical trend; the word "already" was therefore imprecise. We will (i) state the period explicitly, (ii) reframe the statement as a projection, and (iii) ground the claim of spatial pervasiveness in the new trend-significance results rather than in a qualitative impression. The sentence will be revised to read:

"The rate of change in the 17–19 °C band was the strongest among all thermal classes, averaging 1.8 days per decade across profiles (range: −1.4 to 4.0 days per decade) over the projection period (2016–2100). This indicates that the shift from cold-water toward moderately warm conditions is projected to be spatially pervasive across the river network, affecting the large majority of profiles rather than being confined to individual warm rivers."

RC (L332–334, 339–340): *It is not clear how P25 (25th percentile) exceeding 0 indicates the signal is robust.*

AC: We accept this criticism. The P25 > 0 heuristic will be removed as an indicator of robustness. Robustness will instead be assessed by the formal trend analysis (Mann-Kendall significance and per-model agreement), which provides a defensible basis for the claim. Statements previously resting on P25 > 0 will be re-grounded in the significance results.

RC (L349–350): *This definition seems arbitrary. Can you please support this definition with some literature? Considering that > 23 °C is also a clear loss of thermal refugia, shouldn't this be days with water temperature exceeding 21°C rather than in the range of 21-23 °C?*

AC: We appreciate this point but respectfully retain the separate 21–23 °C band, on physiological grounds. Vornanen et al. (2014) identify qualitatively distinct cardiac states: at 21–23 °C the cardiac rhythmicity is disturbed but survival remains possible, whereas above ~23.5 °C cardiac arrest occurs and survival is time-limited. Merging the two ranges would erase precisely the distinction between chronic sub-lethal stress and acute lethal exposure on which the central argument of the manuscript rests. We will add these references and state the rationale explicitly. (The individual band thresholds are likewise literature-based: 17 °C - Jonsson, 2023; 19 °C, Borgwardt et al., 2020; 21–23 °C and > 23 °C, Vornanen et al., 2014.)

RC (L355–356): *The use of words "episodic" and "persistent" is misleading here. These are better descriptors of heat wave events, not annual number of days. I suggest changing the phrasings here.*

AC: We agree and will rephrase. We will reserve "episodic"/"persistent" for heat-wave event characteristics and describe the annual day-count metrics with more appropriate wording (e.g. chronic vs. acute exposure, or frequency-based terms).

RC (**357-358**): *How does P25 being close to zero indicate that stressful conditions are interannually variable?*

AC: We agree the inference was not justified and will remove it. Interannual variability will instead be characterised directly where relevant, and the robustness of trends will be reported via the significance metrics rather than percentile positions.

RC (Section 3.6): *Since this section discusses how number of days with water temperatures change annually, a figure showing annual changes can support them better than figure 4 showing decadal changes.*

AC: AC: We agree that the decadal boxplots emphasise distributional changes at the expense of the year-to-year evolution. In the revised manuscript we will make the annual dimension more visible, either by adding an annual time-series view of the key threshold-day metrics or, by clarifying in the text how the decadal aggregation relates to the underlying annual variability. We will choose the presentation that most clearly conveys the annual changes.

RC (L362–364): *This is premature as results on heat wave events have not even been presented yet.*

AC: We will move this statement so that it follows the presentation of the heat-wave results.

RC (L376–379): *Delayed response indicates a response to some driver, so please change the wording.*

AC: We will reword to avoid implying a mechanistic response to a driver, describing the temporal pattern in neutral terms.

RC (L385): *It is not clear how this was quantified. Was the first occurrence of those temperatures considered the onset of related stress? How was consistency calculated?*

AC: The reviewer is correct that "later and less consistent emergence" was an interpretive description rather than a computed metric; no onset-timing or consistency statistic was calculated. We will remove this wording and state only what the data support, cross-referencing the trend analysis where the weaker heat-wave signal is quantified. The passage will be revised to read:

"Heat-wave metrics increased later in the century and remained low in magnitude compared with the expansion of sub-lethal thermal exposure. In warm rivers, the median number of heat-wave events increased from zero during the baseline period to fewer than 0.5 events per year by 2070–2085, with upper-quartile values reaching approximately 0.6 events per year (Fig. 6). Because heat-wave events remain rare and spatially confined even at the end of the century, their trends are weaker and significant at fewer profiles than those of the sub-lethal indicators."

RC (L394–395): *There is no figure showing the duration of heat wave events to support this statement.*

AC: We agree. The maximum heat-wave duration has already been computed per profile for SSP2-4.5, so we will add it as a new panel (c) to Figure 6, alongside the existing panels for event count (a) and total heat-wave days (b). Because heat-wave occurrence is strongly concentrated in a small number of warm lowland rivers, we will present this panel as a spatial map, so that the statement about heat-wave duration is directly supported by both the magnitude and the location of the longest events.

Discussion

RC (L440): *It looks like there is a high increase in the number of days with 21–23 °C as well.*

AC: We agree and will acknowledge the substantial increase in the 21–23 °C band explicitly, so this range is given the same weight as the neighbouring classes in the discussion.

RC (L443): *I suggest using "less prevalent" rather than "episodic".*

AC: We will replace "episodic" with "less prevalent" here.

RC (L443–445): *This is a strong statement, considering this paper studies the changes in water temperature and not the impact of thermal changes on aquatic life. I suggest softening it and making it sound more like a hypothesis.*

AC: We agree and will soften this to a hypothesis, making clear that the ecological implication is inferred from thermal thresholds reported in the literature rather than demonstrated by our data.

RC (L446): *It is little confusing how the grid scale air temperature was used to get reach-scale water temperature. Please make this clear in methods.*

AC: We will clarify in the Methods that reach-scale water temperature is obtained by applying the global air-water regression to the gridded air-temperature product at each network cell (as detailed in our response to the L215 comment above).

RC (L469–470): *This line might be better placed in the next paragraph than here.*

AC: We will move this sentence to the following paragraph.

RC (Section 4.3): *This section discusses the spatial connectivity of thermally suitable river reaches. The paper currently summarizes percent length of river network with suitably cool temperatures but not their spatial connectivity. I suggest calculating an index that summarizes structural connectivity of thermal refugia in river network for different years or decades and its change over time to support the argument that spatial connectivity is decreasing.*

AC: We take the reviewer's point that our analysis quantifies the total length of thermally suitable network, not its physical connectivity. Rather than introduce a connectivity index that the thermal criterion alone cannot support reliably, we will revise Section 4.3 to state this limitation transparently and confine our claims to what the data measure. The relevant passage will be revised to read:

"By mid-century, suitable reaches are increasingly restricted to headwater systems. Because our analysis quantifies the total length of thermally suitable river network rather than its physical connectivity, we do not measure fragmentation directly; however, the progressive confinement of suitable conditions to spatially separated headwater areas is consistent with reduced connectivity between them. Whether these reaches remain mutually accessible depends on migration barriers and flow conditions not resolved here."

We will retain the general discussion of connectivity principles but frame a spatially explicit connectivity index as a priority for future work, noting that an index derived from the thermal criterion alone would overestimate accessibility by ignoring physical barriers.

RC (Section 4.4): *Local conditions can influence water temperature over the whole temperature range and not only above 23 °C. For example, groundwater-fed rivers are only cooler in summer but also warmer in winter. This should be made clear.*

AC: We agree and will revise Section 4.4 accordingly. We will make clear that local controls (groundwater influence in particular) act across the full thermal range rather than only at the high end, moderating both summer maxima and winter minima. Groundwater-fed reaches will be given as the explicit example: cooler than the network average in summer but warmer in winter.

RC: *The air-water temperature relationships derived from few rivers, when used to calculate temperatures in the whole network, can result in substantial uncertainties as the catchment characteristics of the 35 river profiles often might not cover the whole spectrum of conditions at national scale. This should be discussed too.*

AC: We agree and will add a dedicated paragraph on this upscaling limitation. The 35 calibration profiles may not span the full range of catchment characteristics present nationally, so predictions for reaches whose conditions are not represented in the calibration set carry additional uncertainty beyond the reported RMSE. We will state that network-wide results should therefore be interpreted as regional-scale patterns rather than reach-exact predictions. This point also connects to the concern raised by Referee #1 regarding the transferability of empirically fitted relationships to conditions outside the calibration range.

RC: *Is there any pattern in model performance among different river profiles that can shed some light on where (or under what conditions) the uncertainties can be higher?*

AC: We investigated this directly. We examined per-profile RMSE against profile characteristics and found no meaningful pattern: RMSE showed no correlation with altitude ($r = 0.07$) or median water temperature ($r = -0.18$), and was comparable across thermal groups (median RMSE 1.62, 1.45, and 1.46 °C for cold, intermediate, and warm profiles, respectively). We will add the following to Section 4.4:

"To assess whether model performance varied systematically across the network, we examined the relationship between per-profile RMSE and profile characteristics. RMSE showed no meaningful correlation with altitude ($r = 0.07$) or median water temperature ($r = -0.18$), and was comparable across thermal groups (median RMSE 1.62, 1.45, and 1.46 °C for cold, intermediate, and warm profiles, respectively). This indicates that model uncertainty is distributed relatively evenly across the thermal and altitudinal gradient, rather than being concentrated in a particular river type."

RC: *This paper considers 23 °C as the fixed threshold for detection of heat waves. Yet several riverine heat waves consider a flexible threshold since aquatic life in different rivers are acclimatized to different thermal conditions. Can you add some discussion on how using a variable threshold might change the picture?*

AC: We will add this discussion to Section 4.4. We will note that a percentile-based (variable) threshold rescales heat-wave detection to each river's local baseline, and would therefore flag events in cooler rivers that a fixed threshold does not, capturing local thermal anomalies rather than absolute exposure. We deliberately use a fixed threshold because our assessment targets a single species with a defined physiological limit (cardiac failure in brown trout at approximately 23.5 °C; Vornanen et al., 2014), which does not shift with local acclimation to the same degree as a statistical baseline does. We will make explicit that the two approaches answer different questions, local anomaly versus absolute physiological risk, and that a variable-threshold analysis would likely identify additional events in cooler headwater reaches while leaving our conclusions about absolute lethal exposure unchanged.

RC (Section 4.5): *Good call on recognizing the importance of stream flow on water temperature. I suggest editing this section to tighten it. For example, declining snowpack and increasing temperatures are reiterated multiple times.*

AC: We will tighten Section 4.5 and remove the repeated statements on declining snowpack and rising temperatures.

RC (L597–606): *This should either be a different section or integrated into the conclusion.*

AC: We will relocate this material, integrating it into the conclusion.

Conclusion

RC (L611–613): *The paper summarizes changes in water temperatures rather than ecological response to different types of thermal changes. Therefore, this statement is not supported by results. Please alter it accordingly.*

AC: We agree and will rephrase the concluding statement so that it refers to projected changes in the thermal regime and threshold exceedance, framing ecological consequences as inferred implications rather than demonstrated results.

Minor comments

RC (Figure 1): *I don't see a dotted line on subplot b? Also, Figure 1a is not discussed in the main text at all. Considering mentioning somewhere about how annual brown trout catch has been declining.*

AC: We will correct the Figure 1 caption to identify the dashed line as indicative only (it does not represent a fitted trend), and add it to subplot b if it is missing there. We will also add a sentence to the text referring to the long-term decline in brown trout catches shown in Figure 1a:

"Long-term angling records document a marked decline in annual brown trout catches in the Czech Republic over recent decades (Fig. 1a), consistent with the progressive deterioration of thermal habitat quantified in this study."

RC (L207): *there is a formatting error here*

AC: We will correct the formatting error.

RC (Figure 3, repetition): *Description repeats the same things twice.*

AC: We will remove the duplicated text in the Figure 3 caption.

RC (Figure 3, table): *The table is so small that it is hard to make out the numbers. Consider moving it to separate table. In general, font size on all figures can be increased.*

AC: We will move the embedded table to a standalone table and increase font sizes across all figures for legibility.

RC (L276): *This line sounds a bit awkward. By thermally exposed, do you mean rivers that are warmer (lowland) or those experiencing highest increase in days with water temperature > 17°C?*

AC: We mean the already-warmest rivers. We will replace the ambiguous term with an explicit description and add the supporting values. The sentence will read:

"The warmest profiles - Železný Brod (Jizera), Valašské Meziříčí (Rožnovská Bečva), and Jarcová (Vsetínská Bečva), each with a median water temperature above 9.9 °C - are projected to exceed 120 days per year above 17 °C by 2085."

RC (L282): *35 monitored profiles*

AC: We will correct this to 35 monitored profiles.

RC (L558): *Consider an alternative term for "stagnating".*

AC: We will replace "stagnating" with a more appropriate term.

RC (L565): *maybe "inverse relationship" instead of "negative relationship"*

AC: We will change "negative relationship" to "inverse relationship."

RC (Figure S1): *figure S1 is not cited anywhere.*

AC: We will add a citation to Figure S1 in the main text.