We would like to thank the referees for their constructive feedback. We appreciate the time and effort that was put into the review and we hope that our responses prove satisfactory. For clarity, the reviewer's comments are presented in black, with our responses in red.

N.B. The Atlas' technical report was cited as Malenfant et al., 2024 in the first version of the manuscript, but this will be changed to DPEH, 2024.

Best regards,

Gabriel Rondeau-Genesse, on behalf of all authors.

Referee #1

General comments:

#1-1. As the calculation of the hydrological drought indices is central for providing the results of this research, it is necessary to provide some more information about the HYDROTEL hydrological model, and how it was used / set up for this research. Readers unfamiliar with this model need some basic information about the concepts used in it, which then needs to be extended with the following: How many watersheds were actually simulated? (There is a mentioning of 10,000 watersheds and a threshold of 'larger than 50km², but it is not clear how many were finally simulated; all maps with indicator results show streams, but it is not clear how are they related to the simulated watersheds). Which exact outputs of the regional climate model were used? Were these only precipitation and temperature (for which the implemented bias corrections are discussed)? Were the watershed models set up with or without sub-basins? How were inputs and parameters specified (gridded or per sub-basin)? Which were dominant runoff components simulated (snowmelt? direct runoff? subsurface runoff (base flow, which would be important for drought conditions?)? In the 'Limitations' section the authors mention some deficiencies of the hydrological model used, and that the value of their article is primarily in the methodological framework that has been introduced, and in "producing plausible, physically coherent low-flow indicators" However, the above suggested information, in my view, is still needed for better understanding of the hydrological modelling component of this work.

We will clarify the text in that section and provide additional details on the hydrological model setup, based on the information below.

HYDROTEL (Fortin et al., 2001) is a distributed hydrological model that uses computational units called "Relatively Homogeneous Hydrological Units" (RHHUs). These units are determined based on land use, soil classification, and geographical characteristics such as slope and aspect. The average area of RHHUs in the Atlas is 11 km², with an associated hydrological network that consists of 28,035 river segments. For a given timestep, the streamflow at each river segment is the sum of incoming streamflow from upstream segments and runoff/baseflow from neighboring RHHUs. Although all 28,035 river segments are modeled, results are published for only 9,665 segments, based on the list of criteria mentioned in our manuscript. The vast majority of rejected river segments are due to the watershed size criterion.

The specific details regarding model configuration and the calibration setup for the Atlas are described in DPEH, 2024 (link, in French). A regional calibration was performed to optimize the average performance (calculated with a KGE objective function) across 70 watersheds, with validation over 151 stream gauges. The calibration includes two parameter sets—one for the north and one for the south of the St. Lawrence River—, each covering 35 watersheds. Six calibration strategies/model configurations were used. The full list of parameters and their values is available in Annex D of the Atlas' technical report (DPEH, 2024). Due to computational constraints, the "MG24HQ" configuration, recommended by the Government of Quebec for its superior performance in summer flows, was selected for our study (see Annex E of the report for validation results).

HYDROTEL is a physically based hydrological model, so all runoff components are simulated explicitly. The specific formulae can vary between model configuration and are listed in Table 16 (Annex D) of the technical report. The subsurface model, BV3C, is relatively simple and uses a cascade of three subsurface reservoirs, which is why that aspect was mentioned in the limitations section.

It will be made clearer that the only meteorological inputs for HYDROTEL are daily minimum and maximum temperature, and precipitation. Other state variables, such as snow and evapotranspiration, are calculated by the model itself.

#1-2. There is certain disconnection between the analytical work carried out to provide the results in terms of hydrological drought indices and the discussion part presented in section 4, especially sections 4.1 and 4.2. These sections mostly present information about drought impacts on ecosystems and socio-economic activities during the 2021 event. There are only some suggestions about how some of the calculated hydrological drought indices could be used to better assess drought impacts, but most of the content in these sections is a description of the impacts from 2021 drought. My suggestion would be to somewhat re-arrange the content in this article and to bring in the description of the 2021 drought in terms of impacts on ecosystems and socioeconomics much earlier (under 'Introduction', or under 'Methodology' as a separate section, or somehow combined with the section on 'Stakeholder consultation'). The argument can then be introduced that lack of adequate hydrological drought indices to assess future droughts under CC is an issue that this article addresses (There is even mentioning that the province of Quebec indeed does not have such indicators in line 60). The whole content of calculating these indices would then follow, and the Discussion section can be somewhat extended about the value / potential use of the proposed indicators. The need to move towards the next step, which is developing impact-related drought indicators could then be emphasized in that discussion. This is just a suggestion, and the authors may decide on a different approach to better connect the discussion part with the previous part of the article.

We agree that there is a disconnect between the results and the discussion. This comment and the suggestion are coherent with Comment #2-5. We will adjust the sections of our paper based on these recommendations.

Specific comments:

#1-3. Line 5, in the abstract: The statement "This approach allowed for enhanced collaboration with water management experts and other stakeholders to project the possible impacts of climate change on serious water deficits in Quebec." should be modified or removed. There is no evidence presented in the article that this actually occurred. (This comment is somewhat related to the general comment 2, mentioned above).

We agree that this might have been overstated. We will modify that line accordingly.

#1-4. Please insert somewhere in Section 1 (Introduction) or Section 2 (Methodology) a figure with an overview map of the actual study area, together with geographical features to which you are later on referring (e.g. St. Lawrence river / valley, Ottawa river, Lake of Two Mountains, etc.). This will be very useful for readers unfamiliar with the geography of Quebec.

We will either modify Figure 1 to provide a bigger map of the study area, complemented with the relevant geographical features, or add a new figure. In the latter case, this would be in the Supplements.

#1-5. Please make it clear in Section 1 (Introduction) that the focus of this research is hydrological drought (and the indices). When one reads this section, it is not always clear whether meteorological or hydrological drought will be the target of the analysis (later on it becomes clear).

We will add information to that effect in the last paragraph of the introduction.

#1-6. Line 60: Please add “and ecosystems” at the end of the sentence “…significantly impact human activities”, to be consistent with the content presented later.

This will be changed accordingly.

#1-7. Line 85: Please do not use ‘water deficits’, as it adds to the ambiguity regarding what is the focus of the cited study and of this study. The cited study is clearly about precipitation deficits (meteorological drought) and this study goes further in calculating hydrological drought indices based on streamflow data.

This will be changed accordingly.

#1-8. Line 95: Please us ‘Sect.’ or ‘Section’ consistently.

We will rework that sentence to circumvent the use of both forms, but this followed the journal guidelines: “The abbreviation "Sect." should be used when it appears in running text and should be followed by a number unless it comes at the beginning of a sentence.”

#1-9. Line 115: The statement “By combining the qualitative data from the questionnaire with the quantitative insights drawn from the storylines, it becomes possible to obtain a more comprehensive understanding of the potential impacts of future severe droughts” should be modified or removed. There is no evidence in the content of the article that such combination was performed. If it is a suggestion that this should be done in the future – the statement should be re-formulated. (This comment is related to the general comment 2, and specific comment 1, mentioned above).

That sentence will be removed or modified, depending on its state after the restructuration of the paper proposed in the Comment #1-2.

#1-10. Line 125: the sentence starting with “The retrospective analysis…” is not clear. Could you please expand it or re-formulate it? What is meant by 'combining streamflow measurements from various locations’? What was combined and how? Which 'gridded observation data were used for driving HYDROTEL’ (again, see general comment 1)?

The term “retrospective analysis” might not have been clear or appropriate. That sentence will be reworked alongside with the additions proposed in the Comment #1-1. Here are more details:

The historical streamflow data in the Hydroclimatic Atlas is reconstructed from 1970 to the present using six simulations (the six calibration strategies mentioned in our response to #1-1) of the HYDROTEL hydrological model driven by gridded observation data. These modeled streamflow are then corrected using an optimal interpolation method, which incorporates gauged streamflow data to adjust the modeled results, accounting for both model and observational uncertainties (Lachance-Cloutier et al., 2017). This process is similar to data assimilation techniques used in meteorological reanalyses.

#1-11. Line 150: The statement "… and, importantly for this work, potential evapotranspiration (PET)" is a bit surprising. Practically all hydrological models use PET as an input, so how can this be a particular feature of HYDROTEL that is important?

This will be rephrased. The main point is that HYDROTEL provides several options for computing Potential Evapotranspiration (PET), and that since PET is central to our methodology, the choice of computation method could potentially impact our results significantly (though, as stated at L.355, it does not). The intention was not to suggest that this is a particularly unique feature of HYDROTEL.

#1-12. Line 175: This approach is not entirely clear to me. The SPEI values will be comparable in terms of values, but if they come from two different distributions (from two different climates), how can then they be comparable? Can you please elaborate a bit?

We will revise that section to clarify our intentions.

By performing two calibrations—one for the past and one for the future—we can identify events that have a similar chance of occurrence and a similar progression throughout the year in both time periods. This allows us to link the historical extreme event (2021) to similarly rare and extreme future events. In flood terms, this would be akin to comparing the current 100-year flood to the future 100-year flood. Using two distributions makes this comparison more intuitive and meaningful.

#1-13. Line 225: So, there were 10 best analogues for the historical period *and* 10 best analogs for the future, and the difference between the averages of the two was calculated, right? Please clarify.

We will clarify this sentence, but this is correct.

#1-14. Lines 270 and 280: Would the results be sensitive to these weights values? Has this been tested? If not, perhaps it should be recommended?

We will add a clarification on this topic. Multiple tests were indeed conducted on the weighting scheme. Our results show that while the selected events themselves are sensitive to the weights and that the outcomes are thus not robust with a limited selection (e.g., 3 best analogues), the results are stable and robust when 10 or more analogues are used.

#1-15. Section 3.4 (Limitations). Perhaps it could be recommended to carry out similar analysis for individual basins, and not only for such a large area? This would bring the use of the indicators closer to actual water resources planning and management, in my view…

We can include recommendations on this in the limitations, as well as in the discussion if it is relevant after the restructuring proposed in the Comment #1-2. While we agree that this provides a broad overview, the distributed nature of HYDROTEL means that indicators are nonetheless computed at a local level, which we hope will have actionable value. A second phase of the project will begin in early 2025, focusing on further exploring the use of these storylines at the local scale, in collaboration with stakeholders in key watersheds affected by the 2021 drought.

#1-16. Figure 7. I don't see why this figure is necessary. Many drought impacts have been discussed, and not supported by similar generic information (and figures). Why is this generic view on drought impacts on river systems chosen to be presented? It is a bit of a distraction, in my view. I would suggest to present relevant figures for this particular research (perhaps Figure S2 from the supplementary material?)

We deemed that figure to be useful as a generic portrayal of the impacts of droughts on river ecosystems, given that a typical reader of this paper is likely to be less knowledgeable in that field. However, with the restructuration of that section proposed in the Comment #1-2, it is likely that this figure will have to be either moved to the Supplements or removed, since the associated text will be moved earlier.

We will consider the addition of Figure S2 to the main body.

Referee #2

General comments:

#2-1. Bias-adjustment of climate model: The authors mention that the detrended quantile mapping bias adjustment technique was applied for precipitation. It isn't clear whether the bias adjustment was applied to each ensemble member independently from each other? It is recommended that any parameters for bias adjustment applied to single-model-initial-condition large ensembles should be computed from the pooled ensemble rather than individual ensemble members to preserve the range of internal variability in the original ensemble. It would be good if the authors can clarify their approach.

A clarification will be brought to the text. The 50 members were indeed pooled together prior to bias adjustment.

#2-2. Hydrological model: more details of the hydrological model would be appreciated. For example, was the model calibrated to observed streamflow in any way and what is the spatial resolution of the hydrological model? Were there any statistical downscaling methods applied to further downscale the large ensemble data prior to hydrological modelling? Related to this, would the authors be able to provide some indication of model performance compared to observed flows at gauges (e.g. in the supplement)? There should also be some discussion of the simulated river flows driven by ClimEx over the historical period – do they exhibit similar hydrolgical behaviour to observed river flows and where does observed flows lie within the wider range of ClimEx simulated flows?

We will provide more details on the model setup. Refer to our responses to Comments #1-1 and #1-10 for answers to some of the questions raised.

While we likely won't include specific validation results in the paper, we will provide a more direct reference to that information when reworking the section. All validation results, including the KGE of the model and the RMSE of the minimum 7-day flow (7Qmin), are available in Annex E of the Atlas' technical report (DPEH, 2024; link, in French).

No additional spatial downscaling was applied to ClimEx data, as its spatial resolution (0.11°) is already close to the regridded observations used as a reference (10 km).

Direct comparisons between observed flow and simulated river flows driven by ClimEx are challenging due to the optimal interpolation method applied to historical flow (see our response to Comment #1-10 for more details). Figure 4 shows that there is a good agreement in river flows between observations and ClimEx analogues for 2021, but a few additional figures comparing the river flows between ClimEx and the observations over specific river segments could be added to the Supplements.

#2-3. SPEI calculation: Have the authors looked at how their results might change if the SPEI distribution for the future periods are fitted to the parameters calculated from the historical period? This could serve as a good sensitivity test for whether the decision to calculate SPEI separately for historical and future period is a valid approach.

Tests were conducted early in the project, but the results were difficult to apply effectively to the focus of our work (i.e., identifying future extremes that are more severe but still relatable to historical events). Please also refer to our response to Comment #1-12.

#2-4. Changes in drought analogues: as the authors note, there are interesting hydrological dynamics within the future analogues (such as changes in winter thaw and changes in the timing of low flow season) which deserves a bit more discussion. Could the authors provide a time series of P, PET and simulated flows for the baseline observed 2021 and for the top 10 future drought analogues (either at particular selected catchments or averaged)? The readers can then visualise these temporal changes easier.

We will consider adding such figure(s). Averaging over an area as large as Southern Quebec is likely meaningless for such an analysis, so we will select a few catchments that were impacted in 2021.

#2-5. Discussion: The current Discussion needs to be reframed. The section seems out of place and reads more like a separate literature review rather than an actual discussion of the results. Perhaps the authors could reframe this to link to the results of the paper (e.g. how these storylines can be used to further drive water quality models or how flow indicators from the drought analogues could be used to infer different drought impacts). Much of the socio-economic impacts of the 2021 drought could actually go in the introduction to highlight the severity of the event and motivate why the authors decided to choose the 2021 event as a case study. There should also be further discussion of the utility of the storyline approach and future work (e.g. are there limitations to the storyline approach? How can such an approach be used alongside other climate change projection products?)

This comment and the proposed solution align with Comment #1-2. We will restructure the sections accordingly.

Minor amendments:

#2-6. IPCC (2023) should be cited according to the official suggested citation according to the IPCC:

> Lee, J.-Y., J. Marotzke, G. Bala, L. Cao, S. Corti, J.P. Dunne, F. Engelbrecht, E. Fischer, J.C. Fyfe, C. Jones, A. Maycock, J. Mutemi, O. Ndiaye, S. Panickal, and T. Zhou, 2021: Future Global Climate: Scenario-Based Projections and NearTerm Information. In Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 553–672, doi:10.1017/9781009157896.006.

Noted. This will be changed.

#2-7. L80 – The authors may consider expanding the literature search and also refer to more recent studies which have also used an analogue approach to sample for drought/heatwave storylines – The following three studies may be useful, the first two for meteorological droughts and the third for hydrological droughts:

> Faranda, D., Pascale, S., and Bulut, B.: Persistent anticyclonic conditions and climate change exacerbated the exceptional 2022 European-Mediterranean drought, Environ. Res. Lett., https://doi.org/10.1088/1748-9326/acbc37, 2023.
>
> Liao, Z., An, N., Chen, Y., and Zhai, P.: On the possibility of the 2022-like spatio-temporally compounding event across the Yangtze River Valley, Environ. Res. Lett., 19, 014063, https://doi.org/10.1088/1748-9326/ad178e, 2024.
>
> Chan, W. C. H., Arnell, N. W., Darch, G., Facer-Childs, K., Shepherd, T. G., Tanguy, M., and van der Wiel, K.: Current and future risk of unprecedented hydrological droughts in Great Britain, Journal of Hydrology, 130074, https://doi.org/10.1016/j.jhydrol.2023.130074, 2023.

Thank you for the references. These will be consulted and added to the introduction.

#2-8. L108 and elsewhere – there are several mentions of a questionnaire that was given to stakeholders to gather information on the 2021 drought and to disseminate results. More discussion of what the outcomes from this questionnaire should be given. Could the questionnaire be published in the supplement as well for transparency?

As the results of the questionnaire were primarily used for the discussion, how we address this comment will depend on the restructuring proposed in the Comments #1-2 and #2-5. If it is still relevant, we will provide more information on the outcomes and some of the results.

We can add the questionnaire to the Supplements.

#2-9. L155 – what is MG24HQ?

We will rephrase it to be clearer. “MG24HQ” is the name of the specific calibration of HYDROTEL that was used for our study. Refer to our response to Comment #1-1 for details.

#2-10. Font sizes in all figures could be bigger. Consider showing a subset of the results and putting other sub-plots in the supplementary materials. For example, Figure 3 could be simplified by including SPEI-3 averaged over May-Oct rather than each month separately (the monthly plots could go in the supplementary materials).

Font sizes will be increased. For Figure 3, we can hide some of the SPEI-3 data and present the full figure in the Supplements. However, averaging SPEIs is generally not recommended. For other figures, we will simplify them where possible, though all indicators that are mentioned in the text should be present in the paper itself.

#2-11. The Limitations section should go within the Discussion section rather than Results.

This section was placed in the Results due to the nature of the Discussion section. However, with the proposed restructuring, this will no longer be an issue, and it will be moved as suggested.

Referee #3

General comments:

#3-1. The appropriateness of the approach is claimed, but hardly explained and/or compared with other approaches. Why can the extensive modelling approach add to the baseline method? Why would it be needed (or even allowed) to keep the storyline approach when running such models? The discussion might have revealed some of these issues, but paragraph 4 does not really engage with the climate stories/narratives as such.

The storyline approach is introduced in lines 65-85 of the Introduction, with specific examples from the literature. Some of the vagueness (as noted in the commented PDF) arose from the lack of established terminology for the various methods used to create storylines. However, a recent paper by Riglos et al. (2024) has since further categorized different types of climate storylines and narratives (https://www.sciencedirect.com/science/article/abs/pii/S1462901124001825). “Physical Climate Storylines” (PCS) (e.g., Hazeleger et al., 2015; Gessner et al., 2022; Matte et al., 2022) recreate past events in a modeling setup, then project those events into future climate conditions by modifying them within the model. Our methodology, along with that of van der Wiel et al. (2021), falls under “Scenario-based Approaches” (SBA). Instead of recreating and modifying past events using a climate model, SBAs involve identifying similar events within existing climate simulations. To find adequate analogues to the real event, SBAs typically rely on large ensembles.

This part of the Introduction will be reworked to incorporate Riglos et al. (2024) and other papers referenced by Referee 2.

#3-2. The text actually reads much like a rather traditional text on climate scenarios. It might be that the use of the model enforces this, but I found it a little tricky to find the added value of storylines - or even the difference with other ways of sharing scenarios. I would think that the text could be much the same without ever mentioning the term "storyline analysis".

As mentioned in the previous comment, storylines aim to extrapolate specific events, such as the 2021 drought, into future climate conditions to analyze potential differences in impacts. Another way to view them is as "temporal analogues." This approach contrasts with the more traditional method of producing climate scenarios through probabilistic ensembles of multiple climate models. Please also refer to my response to Comment #1-12.

We will revise the Discussion as recommended by all Referees to better highlight the value of the method and the potential applications of the proposed indicators.

#3-3. The method section seems to be incomplete. First, the method of storyline analytics is not discussed. How does one do it? Second, in the results section there seem to be several series of steps included that might need to be included in the method section.

How this is addressed will depend on the restructuring of the text, but Section 3.2 could be moved to the Methodology. Additionally, as suggested by Referee 1, Figure S2 will likely be moved from the Supplementary Materials to the main body of the paper to visually support the methodology used to construct the future analogues. See Comments #1-2 and #2-5.

#3-4. The usefulness of the approach is defended by claiming that the results can be used in discussions with stakeholders, with some details on stakeholders ideas in section 4 (which for some reason is presented as discussion and not as results). Any evidence for that claim is not offered, as the stakeholder data come from a survey that was done before the storyline process. As such, we read two different parts: one on climate scenarios and one on survey results. I could also imagine that in a dialogue with stakeholders, modelling scenarios, whether they are shared in storyline format or otherwise, might change the stories themselves. This can only be checked after engaging with the stakeholders. I would be very interested in that aspect.

This issue has also been noted in Comments #1-2 and #2-5, concerning the disconnect between our results and the discussion. By reworking the paper as suggested by both referees, we aim to clarify the added value of the storyline approach. The discussion will be revised to better align with our results and their potential applications. Please also refer to our response to Comment #1-15.

#3-5. In conclusion, I would think that this text is a little unbalanced in terms of explaining the usefulness of the method, in terms of added value of the modelling, and in terms of the rather different section 3 and 4. Section 4 does seem to miss the point on discussing results from Section 3 anyway. The storyline itself reads like a standard climate scenario explanation, which is useful, but NOT necessarily new enough to be published.

The restructuring of the paper should improve the flow of the text and reduce the disconnect between Sections 3 and 4.

Specific comments (from the PDF file):

L.5

We agree that this might have been overstated. We will modify that line accordingly.

L.25

We can rephrase that sentence to clarify that the issue is not that droughts in warm regions are easy to analyze (they are not), but rather that snow-driven recharge introduces an additional layer of complexity in cold regions.

L.70

Refer to the answer to Comment #3-1.

L.125

That section will be reworked to include more information on the model (see the response to Comment #1-1) and the sentence will be clarified. “Rivers significantly influenced by dams and other anthropogenic activities” refers to rivers where streamflow is actively managed on a day-by-day basis. Passive weirs and dams managed monthly are still included in the Atlas. While this does exclude several rivers, these are typically larger rivers with higher mean streamflow, making problematic low water levels less likely to occur (though not impossible). In any case, these rivers had to be excluded, as HYDROTEL is currently unable to account for them.

L.130

This will be rephrased. The term “significant” was used to indicate that a large number of simulations are needed (the higher, the better), rather than in its statistical sense.

L.185

Clarifications will be added to the text. Please refer to our response to Comments #1-12.

L.470

We agree that this might have been overstated. We will modify that line accordingly.