General author comment: We appreciate the time spent by the AE securing reviews and the consideration of our manuscript for potential publication in HESS. We also appreciate the helpful comments from the peer reviewers. We respond to these comments on a point-by-point basis below (our replies in blue italics). We feel these changes will improve our paper. At a high level, the revised manuscript will do a better job connecting our field data interpretation and modeling (present and future conditions). As a general response to some comments below, we note that (1) groundwater thermal modeling is far more robust than groundwater flow modeling or shallow soil temperature modeling (flashy signals are modulated) and (2) our focus in the modeling is to investigate long-term system response (sensitivity) to seasonal and decadal forcing rather than to exactly reproduce the conditions at our field site. This will be better articulated in our revised manuscript.

Reviewer 1

Overview: Thermal impacts of springs on coastal waters and the sensitivity of these springs to climate change are not well understood. To address this issue, this study used field study for a threatened coastal lagoon ecosystem in south-eastern Canada by pairing in-situ thermal and drone-based thermal imagery monitoring to estimate the discharge to the lagoon. It also applied a numerical model to relate measured spring temperatures to their respective aquifer depths, and to study long-term groundwater warming. The value of this study lies on providing some insights to coastal ecosystem management. I have some comments that may improve the quality of this article. Please see the detail as follows:

Thank for this accurate overview of our manuscript and your time in helping us improve our work.

Comment 1: There are two parts of this study: analyzing measurements and numerical modeling. I think the link between the two parts is that the model was employed to match the measurements to locate the aquifer depth that provides the water source to the lagoon. However, this link is not stressed in the text, so it looks like two separate studies. Most importantly, the major aim of the modeling (i.e., studying the sensitivity of groundwater temperature to climate change) is not related to the measurements analysis. I think the authors should work on the text more to link these parts to make them integrated.

This is a good point. The goal of the study was to look at both the present and future thermal impacts of these springs, and thus the field work (present) and modeling (present and future) are directly related. Also, the measured spring temperatures were used to infer the spring depths, which was a key factor in our numerical model (the linkage the reviewer alludes to above). However, we agree that these concepts should be tied together more closely in the text, and we will modify the introduction and methods to highlight this. We will also add a new methods figure that shows how the different aspects of the study (hydrology, drone sensing, radon, and numerical modeling) are integrated. We believe this will result in much stronger messaging and overall scientific narrative.

Comment 2: In regards to the hydrological modeling, some necessary uncertainty analyses is missing. Although two data sets of forcings were used, the assumptions and deficits of the hydrological model SHAW were not introduced and the related uncertainties or bias that may be derived from them were not analyzed. The authors need to discuss the uncertainties from many aspects (e.g., model, data, assumptions) and their possible influences to the results in the text to add the value of this manuscript.

In general, thermal modeling of hydrogeologic systems is far more robust than hydrogeologic modeling (e.g. water flux or head modeling) simply because the associated parameters (thermal properties) are far more constrained. This is particularly true below the shallow soil zone experiencing diel temperature fluctuations. However, we agree that adding another paragraph to the discussion text for the modeling will help with acknowledging some of the assumptions and uncertainty in the modeling approach. We will also refer to the rich literature on SHAW applications/limitations as this is one of the most commonly applied ground temperature models.

Comment 3: L105, "methods section": Too many words were used to introduce the monitoring software and system in section 3.1 and 3.2 which I think is not very relevant to the scientific topic. Is it possible simplify those sections and move some of the contents to SI?

Agreed – we will condense this text and move any tertiary points to the supplement. This is not fundamentally a study on thermal image analysis; rather that was just a step in our methods.

Comment 4: L230-231, "The conceptual … heat transport processes.": Please introduce more about the water and heat transport model. What key transport processes the model preserved?

We will add a couple more sentence on the surface energy balance and subsurface heat fluxes (conduction and advection) in the model. We will also add the governing subsurface heat transfer PDE in the main text or in the supplement. We will also emphasize this is a standard model.

Comment 5: L246-247, "A detailed description…detailed in Flerchinger (2017).": As mentioned above, a bit more about the SHAW model could be introduced in the text, rather than just refering another paper.

See above.

Comment 6: L258, "a daily resolution": Most land models use 1800s as the timestep. Is it a daily resolution too coarse for the soil moisture simulation?

This time step is pretty typical in groundwater temperature modeling (e.g. Langford et al., 2020, Groundwater) when sub-daily soil moisture and temperature fluctuations are not of interest (we are looking at more modulated seasonal or decadal signals). Although soil moisture plays a secondary role (e.g. in altering soil thermal properties) we do not need to resolve these changes at a high frequency. We will add one sentence indicating our justification for this time step.

Comment 7: L260-261, "The minimum and … RCP4.5 hindcast model": Why didn't use the historical reanalysis dataset as forcings? It would be more accurate than the model outputs.

It is a valid point that we could provide more details in the paper on our rationale for the dataset selection for the historic period forcing, although the cited Warner study reveals reasonable agreement between the two datasets. We will add text in the methods section for this point and explain our direction. It's worth noting that we do not have a direct long-term climate record in Basin Head, and that there are issues with either dataset given the reanalysis/statistical downscaling/modeling employed. Also, for our assessment we are not trying to reproduce or compare daily conditions but rather multi-year averages in modulated groundwater temperatures (see Table 2), and thus higher-frequency discrepancies between forcing data are less relevant for our modeling purposes.

Comment 8: L269-270, "(1) CNRM-CM5 … MRI-CGCM3, RCP8.5": What are the spatial resolutions for these model outputs and reanalysis data?

We will add more details to the revised manuscript. This location was taken within a ~10 x 6 km grid. We used the BCCAQv2 statistically downscaled data which is roughly 10x10km (downscaled from 100km).

https://climatedata.ca/explore/location/?loc=BAAHA&location-select-temperature=tx_mean&location-select-precipitation=prcptot&location-select-other=ice_days

Comment 9: L279-280, "The paired discharge … relationship for the lagoon.": I don't think this linear relationship is reliable enough based on only three sites.

Thank you for this comment. We agree that having three points is not ideal, and have noted this in the text (see L410). However, we do have a couple thoughts in reply to this:

1) *The relationship is not linear – it's a power relationship that appears linear on a log plot (see L280)*
2) *The relationship between plume size and flow rate is only valid for a single weather condition and tidal level. Also, many springs are only exposed for a short period at low tide. Thus, all of these points must be taken concurrently, and require flumes to be set up in each spring. Even with a large field team (about 6 people), we were only able to accurately gauge 3 springs at the same low tide point. Also, this event presented very ideal conditions (heat wave maximizing the thermal contrast, coincident with spring low tide exposing the most springs). If we were to get more points, we would have to return with a larger team for an entire new field campaign with very little chance of having the same ideal conditions. Thus, we do not think that collecting more data points to improve this relationship is feasible. Also, we are mostly interpolating (rather than extrapolating) with our plume Area-Q relationship as all but one spring had a smaller plume area than the largest spring we gauged (see supplement table). Thus, we think our approach is reasonable as a first-order assessment. We will add a couple more sentences highlighting the challenges of collecting more data points for this relationship, the unique area-Q relationship for given tide and weather conditions, the ideal environmental conditions during our study, and the limitations of our approach.*