Response to Reviewers

Thank you to the reviewers and editor for careful review of our paper "Quantifying Small-scale Temperature Variability using Distributed Temperature Sensing and Thermal Infrared Imaging to Inform River Restoration". We are confident that it will improve the paper.

Reviewers unanimously recommended further analysis for our paper. Following the HESS interactive review process, in which we publish a response to review comments and the editor reviews then invites a revised manuscript or rejects the paper (step 5 here: https://www.hydrology-and-earth-system-sciences.net/peer_review/interactive_review_process.html), we organize this response into

1) a description of additional analyses that we will complete if invited to submit a revised manuscript, and
2) detailed responses to all other reviewer comments that we have addressed in the manuscript because it was easier for us to make changes as we responded to reviewer comments. Reviewers' recommendations are in black and our responses are in blue.

Additional Analyses of DTS and TIR Temperature Variability

All reviewers have responded positively to using distributed temperature sensing (DTS) and thermal infrared (TIR) sampling to understand spatial stream temperature variability over a range of scales and comparing those data with modeled stream temperatures to identify possible locations and extents of thermal refugia for Lahontan cutthroat trout (LCT), noting that this is a unique and needed line of research. However, all reviewers recommended further analyses to move beyond summary statistics. We agree with the reviewers that additional analyses are possible with our existing data and that they will significantly improve the paper so that it is advances understanding of thermal refugia connectivity for cold-water fish species. If we are invited to submit a revised version of the manuscript, we will add the following analyses:

1) Quantify the percentage of DTS and TIR temperatures that exceed 21 °C (temperatures preferred by adult LCT), 24 °C (chronic 7-day upper thermal limit), and 28 °C (lethal threshold for LCT) in space and time. We will include the temperatures and/or percentages by which the simulation model under-or over-estimates measured temperatures to better understand the times and locations that 1D modeling succeeds of fails for trout. This will also allow for a direct comparison of DTS and TIR for quantifying the amount and connectivity of cool-water refugia (recommended by Reviewers 1 and 4).
2) Compare left and right streambank DTS measurements. This will evaluate lateral temperature variability at a 1 m scale. We will then evaluate temperature ranges at different spatial scales (e.g., 1 m, 10 m, 100 m, 300 m) and discuss implications to interpret 1-D model results (recommended by Reviewers 2 & 4).
3) Similarly, we will compare TIR measurements at various scales (e.g., lateral pixels and longitudinal pixels of increasingly longer reaches) to complete a similar evaluate of spatial scale with TIR data. We have the raw TIR data which we will use rather than the summary points from the Watershed Sciences (2012) analysis (Reviewer 2).
4) We will add the TIR imagery with a DTS deployment (e.g., overlay the data or show side by side panels to Figure 3 and/or 4 to compare DTS and TIR data and results for quantifying the amount and connectivity of thermal refugia (Reviewer 3).

These analyses will advance our work beyond summary statistics to evaluate temperature ranges and alternative spatial scales so that we can better compare sampling and modeling methods and discuss the pros and cons of each method.

Authors Response to Reviewer 1

General Comments:

The paper addresses relevant scientific questions focusing on the importance of spatial and temporal fine scale variations in thermal habitat for Pacific salmon. The introduction could focus more on setting up the problem of warming waters for Pacific salmon in the southern extent of their range and on the importance of small scale variability in temperature for fish survival. Examining more recent papers that reference Sutton et al 2007 could help in this effort. Combining DTS, TIR, and one-dimensional temperature modeling in an effort to better manage and restore river systems is unique.

We added the following text and citations to the first paragraph in the introduction: "*Trout and salmon avoid heat stress by sheltering in pockets of cold water when stream temperatures are near upper thermal tolerances (Dunham et al. 2003; Sutton et al. 2007) and climate change is anticipated to increase stream temperatures in summer, fall, and winter (Isaak et al. 2012). Recent research has quantified when and where cold-water fish need thermal refugia (Brewitt and Danner 2014), estimated the size of thermal refugia and the distance between refugia (Fullerton et al. 2018), demonstrated how fish use thermal refugia (Frechette et al. 2018), and measured the length of time that fish can survive between refugia (Pepino et al. 2015). Where stream temperatures are warming or where cold-water fish species are at the southern extent of their range, assessing stream temperatures at small temporal and spatial scales is thus important to quantify stream temperature heterogeneity and best manage complex and variable riverine habitats (Vatland et al. 2015).*"

The methods were outlined reasonably well; however there were some points of confusion. It is possible that a cartoon schematic of the deployment of the DTS could be beneficial.

We show the layout of the fiber optic cable in Figures 3 and 4? Do you mean that you'd like to also see the location of the DTS and calibration baths? If so, we can add those to the figures.

The DTS itself had "two channels"and at times the wording became confusing because the DTS was deployed in a river channel.

We changed the wording to 'instrument channel' when discussing the DTS channels to avoid confusion.

Results largely consisted of summary statistics of the temperature data and modeled temperature. It would be instructive to do more analysis that relates to the Pacific salmon of interest and its 28 degrees C threshold. Additional investigation might be done to research the following questions: 1) What percentage of the deployment was 28 degrees C exceeded at the mainstem site? 2) When the temperature was exceeded, did the model over or underestimate the actual value and by what percentage? 3) Is there any way to go beyond the summary statistics of the dataset?

The conclusions reached are clear; however, they seem fairly straightforward in that a one-dimensional model would certainly miss fine scale variation that occurs in a real river channel. I think it could be shown more clearly and in further detail the times when the one-dimensional model succeeds or fails for the Pacific salmon.

Thank you for your specific suggestions. See the proposed 'Additional Analyses of DTS and TIR Temperature Variability' section above for our response.

How certain is that the results presented would not be different for a deployment time later in the summer? How might the results and discussion change for a period later in the summer?

We added the following text to the limitations section to address this comment: "*We deployed the DTS during mid-summer when we anticipated stream temperatures would be warmest (and when the DTS*

was available) as worst-case scenario for thermal refugia and connectivity. Additional research is needed to quantify how results would change for deployments earlier or later in summer."

Key numerical results could be added to the abstract.
We added numerical results of model fit with observed data and can add additional numerical results following the addition of new analyses suggested by all reviewers.

At times the writing does not appear to have a singular voice.
We read through the manuscript to improve writing and maintain a singular voice throughout the paper.

Specific Questions:
pg. 3, Line 30: In what way do the diversions and return flows influence the temperature? It could be beneficial to explicitly state they warm the waters.
Diversions and return flows usually warm the river, but do not always warm the river. For example, one of our findings is that water in the Wabuska Drain return flow can be cooler than the rest of the river. For that reason, we left our original sentence unchanged.

pg. 4, Line 10: What percentage of the deployment period did temperatures exceed 28 degrees C? Is this a common problem in the watershed (are there locations where it is always greater than 28 degrees C) or does the problem occur only during discrete heatwave events?
See proposed Additional Analyses of DTS and TIR Temperature Variability section above for our response.

pg. 4, Line 30: Is it necessary to describe how DTS works in this level of detail?
We omitted a couple sentences in this paragraph to remove unneeded details and improve readability.

pg. 8, Line 28: How were key features added to the model output?
We clarified that key features were georeferenced so that the modeled reached that contained the features could be identified. This section now reads:
River features like agricultural return flows, diversions, beaver dams, and seeps were georeferenced so that the modeled reach that contained those features could be identified. Measured DTS and TIR temperature ranges were applied to the model outputs to estimate small-spatial scale variability within each 300 m modeled reach. (starting page 9, line 30).

pg. 9, Line 29: It would be useful to show that there was not a large warming even that occurred between the two deployments at the two sites. Weather conditions should be ruled out as a factor of influence on the two deployments.
We will add a supplemental figure of air temperature data to show that there was not a large warming event that occurred between the 2 deployment events and to rule out weather conditions as an influence on the deployments.

Specific Notes:
pg. 2, Line 3: Therefore, assessing stream temperatures...
Text revised to this suggested wording.

pg. 2, Line 9: Stream temperature models are a useful tool for river management because they help...
Text revised to this suggested wording.

pg. 2, Line 23: ...spatially-continuous stream surface temperatures. However, it... What does "it" refer to?
Text revised to this suggested wording. We substituted 'TIR data' in place or 'it'.

pg. 3, Line 31: Interactions between... "between" compares two things while "among" compares more than two things
Thank you. We substituted the word 'among' in place of 'between'.

pg. 4, Line 12: Low instream flows from surface water diversions have also caused the Walker Lake level to decline
We kept our wording for consistency throughout of manuscript. We do not refer to Walker Lake as 'the Walker Lake' elsewhere in the paper.

pg. 5, Line 25: Verb tense was changed to present tense.
We double checked that our methods are described in past tense. Accuracy of data loggers are in present tense because it always holds true.

Authors Response to Reviewer 2

In this manuscript, high spatial resolution temperature observations are used to investigate small scale spatial variability in stream temperatures. These are placed on top of simulated stream temperatures with grid cells of 300m, leading to the conclusion that at moments when the simulated temperature exceeds the critical temperature (for LTC) of 28 C, there are always pockets of water that are cool enough to serve as refuges. Although these findings are potential useful, this manuscripts reads a bit as an engineering report with the focus on 'improving' an existing 1D 300m-resolution temperature model. There is little in-depth science involved: the results section is mainly a summing up of observations at the 300 m scale, while more quantitative assessment of the uncertainties/sensitivities of the used approach is missing.

Because I still think the available data is sufficient to improve the manuscript significantly, I am advising major revisions. What in my opinion is needed is a more in-depth analysis about spatial variability over a whole range of scales. For example: one can compare the left and right bank DTS measurements for each cross section as the smallest scale (_1m). Subsequently, the temperature range for progressively larger scales can be obtained. Finally, the implication for the model scale (300m) can be used as a practical case study. Similar things can be done for the TIR observations. Although it was not clear to me if the 10 points which were used to obtain the summary points were measured in 1 cross section or if they were measured over 300m, with a spatial resolution of 0.6 m, there seems to be enough data to do the same analysis for the TIR data.
Thank you for your detailed and specific suggestions. See the proposed 'Additional Analyses of DTS and TIR Temperature Variability' section above for our response and outline to improve our paper following your recommendations.

Line by line comments:
P2,L29: Westhoff et al. (2007) did their research in the Maisbich river in Luxembourg.
We apologize for our error. We have checked our references and revised this sentence to read: "DTS data were used to calibrate and validate a 1.3 km physically-based, one-dimensional stream temperature model of the Boiron de Morges River in southwest Switzerland (Roth et al., 2010) and a 580 m river reach in Luxembourg's Maisbich River (Westhoff et al. 2007)."

P3, L3: Neilson et al. (2010) did indeed not use these measurements, but they did model the temperature of transient storage zones.
This comment may be an error. We do not cite Neilson et al. 2010 on page 3, L3 and the Neilson et al. 2010 paper we cite is a different one from the reference the reviewer provided. We cite:

Neilson, B. T., Hatch, C. E., Ban, H. and Tyler, S. W.: Solar radiative heating of fiber - optic cables used to monitor temperatures in water, Water Resour. Res., 46(July 2009), 1–17, doi:10.1029/2009WR008354, 2010.

P3,L10: isn't spatial variability not the same as temperature range?
We included this phrase to specify that we quantify spatial variability as temperature range rather than other metrics like distributions. We changed the wording to "… evaluate small-scale stream temperature variability, quantified as the range of stream temperatures…," to clarify our intent.

P3, L23: please also indicate stream width and depth estimates.
We added width and depth so this sentence now reads "*The Walker River is a desert stream with mean annual flows of 15.5 – 30 m3/s, width of approximately 7.6 m and depth of about 32.9 cm.*"

P4, L31: change to: the majority of the reflected energy has it original wavelength.
Text revised to incorporate reviewer's suggestion.

P6, L25: Are these 10 locations spread over 300 m, or 10 locations covering one crosssection? And why are only values in the centre of the channel taken? Pools etc, are generally located on the sides of the channel. The spatial variability will probably be larger when these values on the side of the channel will be taken into account as well.
We will analyze the raw data versus using Watershed Sciences' summary points to address this comment. See further details in the 'Additional Analyses of DTS and TIR Temperature Variability' section above.

Section 3.4: Here, many statistical parameters are explained. As a reader I found it difficult to remember what the meaning is of each statistic. Furthermore, the symbols defined here hardly ever comes back in the results section. My suggestion here is to only define a few of them, and in the results section clearly state if a spatial or temporal range is discussed.
We followed your suggestion to remove many parameters. We clarified the statistical parameters that we kept. Overall this should clarify this section and remove some redundancy and unneeded detail.

P7, L16: use subscripts for TSavg: now it reads as T*S*a*v*g (the same for all other equations). Instead of 'avg' one can also use a bar on top of T
We have removed avg in the notation and use bars over the top of the variable instead.

P8, L1: what does "top of hour" mean?
We changed text to say 'hourly' instead of 'top of the hour' throughout the manuscript.

P8, L10: If I understood it correctly, these summary points were the mean, min and max. So why are there "on average" 3 summary points? This implies that at some location you have less and at others you have more summary points.

We will analyze the raw data versus using Watershed Sciences' summary points to address this comment. See further details in the 'Additional Analyses of DTS and TIR Temperature Variability' section above.

P8, L15: L is river length, isn't it?
Yes. We clarified the text to read "*where $\overline{T}_L$ is average stream temperature for the length of the East, West, or mainstem Walker River, L*".

P8, L28: What do you mean with "extrapolate model outputs"?
We re-wrote that section to clarify our meaning. It now reads "*River features like agricultural return flows, diversions, beaver dams, and seeps were georeferenced so that the modeled reach that contained those features could be identified. Measured DTS and TIR temperature ranges were applied to the model outputs to estimate small-spatial scale variability within each 300 m modeled reach*."

P8, L31: the key features are the same as the river features discussed in the previous lines, aren't they?
We changed 'key features' to 'river features' to clarify our meaning.

P11, L15: I guess you mean that the backwater location is subject to a higher thermal mass (or lower heat capacity)?
We revised this sentence to read "*The warmer temperatures occurred in an unshaded, shallow, backwater location subject to lower heat capacity*".

P11, L33: with a single snapshot in time one cannot conclude that the stream is cooling. One can only see that it is cooler downstream, but this can also be caused by warm water that only travels slowly downstream.
We revised this text to read "*Although Wabuska Drain was receiving agricultural returns and therefore contributing warm water, rather than the cool water observed during times with limited agricultural returns, a 4.5 km stretch of river downstream from the Wabuska Drain was 1 °C cooler than the segment of river upstream of Wabuska Drain*".

Section 5: Discussion: I miss a discussion on the difference between TIR and DTS measured ranges: which one has the largest range and highest temperature, and why? Although both methods measured at different periods in time, it must be possible to say something about it.
We added the following text to address this concern "*Overall, DTS measured a larger temperature range than TIR imagery in both the East Walker River and mainstem river (Tables 2 and 3) because the fine DTS spatial resolution could measure temperatures that varied spatially over short distances where beaver dams or return flows existed. The warmest temperatures were measured by TIR imagery in the East Walker River, but by DTS in the mainstem, indicating that both methods are useful, and more widespread data collection of longer DTS deployments or repeated TIR collection would further improve results.*" (starting page 16, line 9)

P14, L4-5: Why is the change in temperature over time minimized on warm days? I would presume that on such days the rate of change is at its maximum, since also the daily amplitude is the largest on these days
We omitted to words 'warm, clear' to clarify that we minimized stream temperature changes through time by collected TIR imagery on days with similar weather conditions.

P14, L8-9: Can you quantify the effect of stratification?

We noted that quantifying temperature range from vertical stratification is outside the scope of this paper.

P15, L6: The definition of temperature range is defined as Tmax-Tmin, so it can never be a negative value.
We changed '*temperature range*' to '*temperatures*' to distinguish meaning from our earlier definition.

Section 5.4: this belongs in the conclusion (or summary) section.
We moved this section to the summary section.

Figures, general comment: add the caption of the figures also right below the figures: I know this will happen once the manuscript is published, but as a reviewer, I found it annoying that these captions were only listed in the text.
We will add captions to the figures when we submit a revised manuscript.

Fig. 3b: reduce the scale on the vertical axis to max 2 C?
We kept the scale on the y-axis to max 8 C so the figure is comparable with Figure 4b. We felt that different scales between Figures 3b and 4b would be misleading to readers.

Fig.5: In my opinion, there is little added value to the panels b-d. I also wouldn't call them "summaries" anyway: it is just a different way of plotting the same results.
We removed panels b-d from this figure.

Fig.7: What was the time for the simulated temperature? I am also wondering how variable the range over time is. This is especially interesting since the TIR data is from a completely different year than the DTS measurements. The legend of panel b can also be improved: What is the dark blue line, how are the average temperatures obtained?
We added the time for the simulated temperature to the caption and reworded it so it now reads "*Locations of river features that affect stream temperatures in the Walker Basin (a). Warmest predicted RMS stream temperatures for June 29, 2015 (6:00 pm) with estimated temperature ranges by river feature using DTS data from June 29, 2015 at the warmest observed time (3:15 pm) and TIR data from July 18 and 24 - 26, 2012) (b).*" We removed the average temperatures to simplify the figure.

The DTS measurements give spatio-temporal ranges and the TIR measurements give spatial variability for one time period. TIR and DTS data were acquired at different times. We have overlain both DTS and TIR onto the same model run because it reduced data/complexity and led to largely the same results. We could overlay these data onto the RMS model output for matching time periods (i.e., TIR for 2012 model run and DTS for 2015 model run) and separate Figure 7b into 2 figures if this is misleading. We like this figure because it is our perception that managers make decisions as new data is analyzed and shared without always developing contracts for additional model runs. But as we've said, we can omit it if reviewers feel it is misleading.

Authors Response to Reviewer 3

Hydrology and Earth System Sciences manuscript HESS-2018-441 describes the result of a study to quantify small-scale temperature variability in the Walker River, Nevada, using a combined approach leveraging DTS, thermal infrared (TIR) remote sensing and one-dimensional river temperature

modelling. DTS and TIR data were incorporated with coarser temperature model predictions to identify potential thermal refugia habitat that may otherwise be missed from lower-resolution temperature model outputs. While the general idea of combining TIR, DTS and model predictions is laudable (and there is definitely a need for this type of research), I felt that the manuscript was generally too descriptive, with a lack of formal hypothesis testing or deeper analysis. Furthermore, the method by which the TIR/DTS data was combined with the temperature model predictions (through simple addition of temperature ranges to the temperature model-derived thermal long profiles) was rather simplistic and does not appear to offer a substantial benefit over the use of the DTS/TIR data alone. Despite these issues, I support the general concept of the manuscript and believe that the idea of combining DTS, TIR and model predictions has merit, but I feel that significant revisions (including a much more in-depth treatment/analysis of the data beyond the scope of what would normally be considered 'major revisions') is needed prior to publication.

Please see the proposed 'Additional Analyses of DTS and TIR Temperature Variability' section above for our outline to improve data analysis in our paper.

—General comments—

-In section 3.4 (and table 1), it is quite difficult to remember/follow the names of the different variables. Is there any way of streamlining/simplifying this?

We removed many parameters and clarified the statistical parameters that we kept. Overall this should clarify this section and remove some redundancy and unneeded detail.

-It's probably not necessary to give all of the detail about the DTS system (ie. 2 channels, cable stress, etc) in the methods and results sections. Consider removing all superfluous information to streamline the manuscript.

We removed unneeded DTS details and text to streamline the manuscript.

-It would be nice to see some of the TIR imagery to allow for comparison with the DTS data; given the large time difference between the TIR and DTS data, the information in table 3 is of limited use.

We kept Table 3 since it provides a succinct summary of max, min, and average TIR measured temperatures and ranges. We will add TIR imagery to the revised manuscript (see the additional analyses section above).

-Use of the word 'RMS' (ie. the name of the model) is confusing in places when you are also talking about RMSE. Consider putting 'RMS' in italics or similar to avoid confusion.

We italicized 'RMS' throughout the manuscript to avoid confusion with RMSE.

-There is very little information about the RMS model in terms of inputs, implementation, etc. I appreciate that this information is already given in the other references provided, but it would nonetheless be useful for see it outlined in this MS (maybe a short paragraph explaining model function, input data, etc).

We improved the description of the RMS model to include description of inputs, model function, and implementation. Section 3.3 now reads:

"RMS is a 1-dimensional hydrodynamic and water quality model which solves the St. Venant equations for conservation of mass and momentum and the Holly-Priessmann mass transport equation (Hauser and Schohl, 2002). RMS has a hydrodynamic module and a water quality module, which are run sequentially (Hauser and Schohl, 2002). Input requirements for the hydrodynamics module are channel geometry, roughness coefficients, boundary conditions and initial surface water elevations. Outputs are velocity and depth at each model node which are input into the water quality module. Additional inputs

for the water quality module include weather data, riparian shading estimates, boundary temperatures and initial water temperature. Outputs are hourly stream temperatures.

Previous research provided modeled streamflows and stream temperatures for one wet (2011) and three dry (2012, 2014, 2015) irrigation seasons (April 1-October 31) (Elmore et al. 2016; Null et al. 2017). The model was developed to simulate stream temperatures from environmental water purchases that alter thermal mass to improve habitat for native organisms and connect Walker River and Lake. Irrigation season (April 1–October 31) was modeled because that is the time period that environmental water purchases occur from irrigators. A total of 305 km of the East Walker, West Walker, and mainstem Walker Rivers were represented in RMS at an hourly time step and 300 m modeled reach spatial resolution. As a 1-dimensional model, each reach has a homogenous temperature throughout the reach. Walker River modeled extent includes the East Walker River downstream of Bridgeport Reservoir (river km 243 to 117), the West Walker River downstream of Topaz Reservoir (river km 60 to 0) and the mainstem Walker River to Walker Lake (river km 117 to 0) (Fig. 1). For additional model detail see Elmore et al. (2016) and Null et al. (2017)."

-It's not immediately clear in the manuscript how you 'apply' the DTS/TIR temperature data to the model. From my reading, you simply add/subtract the temperature range to the model outputs. Is this correct? If so, I would have thought that this process (ie. calculating temperature variability from DTS and TIR and simply adding/subtracting to/from the model outputs) is accompanied by a range of issues given that the DTS data is essentially a measure of spatio-temporal temperature range, whereas the TIR data only gives spatial variability. Also, given that the TIR data and DTS data were acquired at different times, it does not seem appropriate to overlay these data onto the RMS model output for the same point in time (eg. fig 7).

We changed this wording to "*Measured DTS and TIR temperature ranges were added to model outputs to estimate small-spatial scale variability within each 300 m modeled reach. This provided an estimate of spatial variability missing in the model output that is needed to gain insight into potential habitat availability at smaller-spatial scales.*" The DTS measurements give spatio-temporal ranges and the TIR measurements give spatial variability for one time period. TIR and DTS data were acquired at different times. We have overlain both DTS and TIR onto the same model run because it reduced data/complexity and led to largely the same results. We could overlay these data onto the RMS model output for matching time periods (i.e., TIR for 2012 model run and DTS for 2015 model run) and separate Figure 7b into 2 figures if this is misleading.

-What is the reason for the large temperature discrepancy between the two river banks? This kind of information is quite interesting and could benefit from a thorough treatment in the manuscript.
Thank you for this suggestion. We will add it to the revised manuscript. See the 'Additional Analyses' section above.

-Section 3.4 of the methodology is difficult to follow. Given that this section contains the majority of the analyses covered in the results and discussion sections, it would be beneficial to restructure this section to make the subsequent results and discussion easier to follow.
We completed a major revision of this section, removing redundancy and unnecessary detail. We hope this section is easier to follow now.

—Introduction—
P2 L1: Would be good to have 1-2 sentences of more general information on the importance of river temperature in the context of climate change.

We added the following phrase to the 1st sentence: "…and climate change is anticipated to increase stream temperatures in summer, fall, and winter (Isaak et al. 2012)."

P2 L24: It would be a good idea here to qualify the point about TIR only providing a single snapshot in time by saying something along the lines of 'unless acquired multiple occasions'.
We added this phrase.

—Methods—
P4 L22-28: These lines are potentially redundant and could be moved or redistributed elsewhere in the methods section.
We omitted these lines and added the pertinent information that was not redundant elsewhere in the methods.

P4 L30-31 and P5 L1-5: I'm not sure that a description of how DTS works is necessary; consider removing this paragraph (or at least, shortening it to one sentence)
We shortened this section so it is now 2 sentences: "*DTS units measure temperatures by sending a laser pulse down a fiber-optic cable and timing the return signal. Although the majority of the reflected energy has its original wavelength, a portion of the energy is absorbed and re-emitted at both shorter (Anti-Stokes backscatter) and longer (Stokes backscatter) wavelengths. Temperatures along the cable are determined from the Stokes/Anti-Stokes ratio (Selker et al. 2006).*"

P5 L6-13: Consider adding a figure to illustrate the deployment of the DTS. I appreciate that this is partially covered in fig 1 (large scale) and fig 3/4, but it would be nice to see a full resolution map showing the DTS installation.
We show the layout of the fiber optic cable in Figures 3 and 4? Do you mean that you'd like to also see the location of the DTS and calibration baths? If so, we can add those to the figures.

P5 L18-22: Is this information necessary? Consider removing.
We omitted lines 19-22 (line numbering from original document).

P5 L28: This sentence formulation is slightly difficult to follow. Do you mean to say that the calibration process consisted of using a linear transform to correct the DTS based on the difference between the DTS and thermocouple temperatures in the ice bath?
We revised this sentence to read: "Calibration used a linear transformation to correct the DTS data based on the difference between the DTS and thermocouple temperatures."

P6 L18-32: There isn't any mention here of the winter TIR data collection flights which you subsequently refer to later on in the manuscript. The manuscript only appears to have dates and hydrometeorology information for the summer flights. I appreciate that the winter data is only used for locating seeps, but it would be good to talk about the winter flights here first.
We discuss the winter and summer flights in the first sentence of Section 3.2.

P6 L32: Although I managed to find the Watershed Sciences documents online, they were quite difficult to track down. It would therefore be good to give some further brief details of the flights, for example RMSE or R2 of TIR data vs. logger values, etc.
We added the following sentences to describe TIR validation. "*TIR radiant temperatures were validated with 28 Hobo Pro and iButton kinetic sensors. For the river extent used here, TIR data were within 0.5 °C of the instream sensors except for one location in the East Walker River where two instream sensors*

were 1.7 °C and 3.3 °C cooler than radiant TIR temperature, and one location in the West Walker River where an instream sensor was 1.1 °C cooler than radiant temperature."

P7 L7: Sentence ('each reach [: : :] throughout the reach') is a little clunky – consider rewriting.
We revised this sentence to read: "As a 1-dimensional model, each reach has a homogenous temperature."

P8 L28: This sentence ('To extrapolate model outputs: : :') is difficult to follow; it is difficult to understand exactly what you are doing here. Can you think of a different way to explain this step in the methodology?
We've revised this sentence to read "River features like agricultural return flows, diversions, beaver dams, and seeps were georeferenced so that the modeled reach that contained those features could be identified."

P8 L32 and P9 L1: What do you mean by 'were developed in 2012'? Do you mean that diversions/return flows were implemented in 2012 in the model, or in reality? Or just that they were mapped/identified?
We changed wording from 'developed in 2012' to 'identified in 2012'.

P8 L2-4: If you wanted to exhaustively identify all seeps, I wonder if a better practice would have been to combine data from the winter and summer survey flights? Also, please give some detail about how the seeps were identified. Was it from manual photo interpretation? Were aerial photos acquired simultaneously to aid the interpretation process?
Due to the low water in July, the majority of locations on the Walker that showed groundwater activity in the winter were dry at the time of the summer flight. Seeps were identified from cooler stream temperatures that could not be attributed to shadows, cutbanks, or vegetation. We added text describing this to Section 3.4.

—Results—
P9 L26-30: These sentences would be better suited to the discussion section.
We moved the last sentence to the Discussion (section 5.2). We left the sentences about the dam releases so the results make sense to readers unfamiliar with this system.

P10 L20: I'm not sure what you mean here when you say that the daily max/minimum temperature changed little when analysed with the Walker River. Do you mean to say that a lack of large-scale temperature variability in the study reach masks considerable localised variability in areas like the Wabuska drain?
Yes. We revised the sentence to reflect your suggested wording. "When the 20 m section of the Wabuska Drain return flow canal (shown approximately at distance 110 – 175 m in Fig. 2b) was analyzed with the mainstem Walker River, daily minimum and maximum temperatures did not change because reach scale temperature variability masks localized variability in areas like the Wabuska Drain."

P10 L25-30: Some of this material might be better suited to the discussion section.
We moved this text to Section 5.2 in the Discussion.

P11 L32-33 and P12 L1-5: This would also be more suitable to the discussion.
We moved this text to the discussion.

—Discussion—

P13 L23-31 and P14 L1-9: I'm not really sure what information this 'limitations' section adds to the manuscript. There are clearly limitations when conducting an approach such as this combining TIR, DTS and modelling. However, this section reads more like a list of problems associated with TIR and DTS data collection (which have already been well established in the literature) rather than a critical appraisal of the inherent difficulties of combining and comparing these types of data with 1D temperature models, which would be much more interesting (and potentially useful for the reader).

We largely rewrote the limitations section. We briefly summarized problems associated with TIR and DTS data collection which have been previously discussed in the literature. Then we highlighted the difficulties of comparing datasets with different spatial and temporal resolutions and that were collected in different years.

We added the following text: "*Obtaining small-scale spatial and temporal stream temperatures and comparing them to model results has a number of limitations. First, the spatial and temporal resolution vary between DTS and TIR datasets and with simulated model results. TIR imagery represents a single point in time unless costly flights are repeated. DTS measurements are dense (1 meter in these deployments) with a 15 minute temporal resolution, but are limited by cable length and field crews to monitor the deployment. Comparing measured data to hourly model results with 300 m spatial resolution further reduces the number of comparable observations. Second, DTS and TIR measurements were collected in different years because we used existing TIR imagery collected as part of the Walker Basin Project, a multi-partner comprehensive effort to sustain the basin's economy, ecosystem, and lake. Future studies could collect data specifically to overlap in time and space; however, opportunistically using existing data for re-analysis and to improve model result interpretation and river management is a laudable goal that may reduce the cost of river science and management. Multi-year, multi-partner river monitoring, modelling, and management is common in large, important, or complex river basins. This research highlights the differences in temperature variability given alternative sampling methods.*"

—Figures—
Figure 7(b). As discussed above, it is not clear how you combine the RMS stream temperature data with DTS data from June 2015 and TIR data from July 2012. Surely this mixing up of different dates and times means that the temperature ranges from the DTS cannot be comparable to those from the TIR? Also, to what does the 'average temperature' refer? Is it from the DTS data (temporal) or is it average temperature (spatial) calculated as the mean of all pixels covering the refugia/beaver dam, etc?

The DTS measurements give spatio-temporal ranges and the TIR measurements give spatial variability for one time period. TIR and DTS data were acquired at different times. We have overlain both DTS and TIR onto the same model run because it reduced data/complexity and led to largely the same results. We could overlay these data onto the RMS model output for matching time periods (i.e., TIR for 2012 model run and DTS for 2015 model run) and separate Figure 7b into 2 figures if this is misleading. We removed the average temperature to clarify the figure.

Authors Response to Reviewer 4

The article compares temperature data from DTS and thermal imagery within a stream reach where one-dimensional temperature modelling was conducted. The study takes place in the Walker Basin, NV, where stream temperatures can exceed thermal tolerance for native, threatened trout. This study provides a unique combination of methods and is useful in understanding the pros and cons of each. The study is relevant for work that seeks to restore habitat for fish in streams where temperatures rise above the thermal tolerance of threatened cold-water species.

There are a few places where the manuscript could be improved. The authors generally summarize the differences between the methods, but they could take it to the next level with doing a more quantitative analysis comparing the different methods. At the end, there is not a strong conclusion, or list of pros and cons, comparing DTS vs TIR methods for validating or contributing to stream modelling efforts. What would the authors decide to use - DTS or TIR - based on the results of this paper?

Second this paper has the data to conduct analyses that look at the availability of different water temperatures that are relevant for trout. For example, the authors focus on 28C as a thermal cut off for LCT, but these fish may become thermally stressed at much lower temperatures (e.g., 21 or 22C). Additionally, very cold temperatures in the summer may not be ideal for growth. The authors could consider adding an analysis that describes the area of stream that is available to LCT below different cut offs, e.g., below 18C, below 21C, below 25C, and below 28C, and the distance between those areas. The DTS and TIR methods identify river features that the one dimensional model doesn't, but how important/large/cold is the water provided by these features? The authors could also compare how the different methods perform in quantifying the amount and connectivity of thermal refugia.

Thank you for your detailed and specific suggestions. See the proposed 'Additional Analyses of DTS and TIR Temperature Variability' section above for our response and outline to improve our paper following your recommendations.

One way that could improve the flow of the results section would be to put the comparison of all three methods up front, and then describe the results of how temperatures differ in the basin second. These are currently split into different sections, with the temperatures differences in the basin described following each method.

Thank you for this suggestion. We will re-evaluate and improve the organization of our paper following completion of the additional analyses of DTS and TIR data.

Additionally, the description of the DTS channels can be a bit confusing because there are river channels and DTS channels – is there another descriptor that could be used in place? The focus on the Wabuska diversion can also be confusing, because it was flowing during the time window of one method and not flowing for the time window of the other method.

We changed the wording to 'instrument channel' when discussing the DTS channels to avoid confusion. The Wabuska Drain is important to our findings, so we left it in the manuscript. We have tried to be clear about when it was flowing and when it was not flowing, but had standing water.

Finally, the introduction and discussion could use more context for why micro thermal regufia is important for cold water fish and trout. The author should look for more recent citations of Sutton et al. 2007, Brewitt and Danner 2014, etc.

We added the following text and citations to the introduction: "*Trout and salmon avoid heat stress by sheltering in pockets of cold water when stream temperatures are near upper thermal tolerances (Dunham et al. 2003; Sutton et al. 2007) and climate change is anticipated to increase stream temperatures in summer, fall, and winter (Isaak et al. 2012). Recent research has quantified when and where cold-water fish need thermal refugia (Brewitt and Danner 2014), estimated the size of thermal refugia and the distance between refugia (Fullerton et al. 2018), demonstrated how fish use thermal refugia (Frechette et al. 2018), and measured the length of time that fish can survive between refugia (Pepino et al. 2015). Where stream temperatures are warming or where cold-water fish species are at the southern extent of their range, assessing stream temperatures at small temporal and spatial scales is*

thus important to quantify stream temperature heterogeneity and best manage complex and variable riverine habitats (Vatland et al. 2015)."

We also added the following citations to the discussion: "*Future research is needed to validate temperature ranges by river feature at the watershed-scale, evaluate how fish use thermal refugia, and to improve understanding of the resiliency of thermal refugia with anticipated climate change (Fullerton et al. 2018; Frechette et al. 2018; Stevens and DuPont 2011; McCullough et al. 2009).*"

Specific Comments:
Pg 2 Line 5. It could be useful to explain here explicitly what you mean by one dimension. (Obvious to stream temp. modelers but less intuitive to managers). Do you mean longitudinal interpolations along the stream?
We changed this sentence to "However, one-dimensional stream temperature models *that estimate longitudinal stream temperature changes and* that are applied at the watershed-scale are poor predictors of thermal micro-habitats" to describe one-dimensional stream models for readers unfamiliar with modeling.

Pg 2 Line 13-15. Similar comment as above – what does 2, vs 3 dimensions mean for streams? Length, width, depth, and in that order?
1-, 2-, and 3-dimensional stream temperature models typically represent longitudinal, lateral, and depth dimensions, in that order. To clarify, we revised the sentence to "… one-dimensional because they are less data intensive and more computationally efficient than two- or three-dimensional models *that also vary by width and depth*…".

Pg 2 Lines 23-25. Switch the order of these sentences. Describe how TIR has been used to locate various stream features, but then the downside is that it is a single snapshot in time.
Done.

Pg 3 Lines 2-5: This sentence is key in setting up what your study contributes to the ones that you cite, but it isn't clear. These methods have been used to calibrate reach scale models but hasn't been used to quantify temperature ranges within model reaches? The difference in the wording is very slight, and needs to be clarified.
Yes, we mean that DTS and TIR have been used to calibrate models, but haven't been used to quantify temperature ranges within the model reaches. We have rephrased the sentence to clarify our meaning: "*While DTS and TIR have been used to calibrate reach-scale models, no studies have used DTS and TIR to quantify temperature ranges by river feature within model reaches, and use to that information to estimate likely temperature ranges at the watershed scale*."

Pg 3 Lines 9 – 10 Objective #3 is has circular wording, it's not clear what the objective is. Is the goal to identify features with greater temperature ranges because they have variable temperatures? Maybe the objective is just to identify those features?
The objective is simply to identify the features. We omitted the phrase "…*due to spatially variable temperatures*" to clarify meaning.

Pg 3, Lines 11-18 I suggest reducing this description of theWalker Basin to 1 sentence, maybe 2 sentences here. You get into more detailed description in the next section. For example, environmental water purchases seem relatively uncommon, and you do them more justice in the next section.

We omitted 2 of the sentences so now we have 2 sentences here: "*Nevada's Walker Basin was the study watershed and is representative of other arid and semi-arid watersheds in western USA where cold water species, like trout and salmon, are temperature-limited. River restoration is ongoing in the Walker Basin and there is a clear need to understand small-scale stream temperature ranges in different different river features (e.g., pools, confluences) to identify temperature barriers to migration and better interpret watershed-scale model results.*"

Pg 4 Second paragraph – It would be nice to include some description about how narrow the range of LCT is. It makes your study more special.
We added the historical range of LCT. This sentence now reads: "*The historical range of LCT is the Lahontan Basin in eastern California, southeastern Oregon, and northern Nevada*, although LCT are limited in their native range by warm stream temperatures, low streamflows, and low dissolved oxygen in the Walker River (Coffin and Cowan 1995; USFWS 2003)."

Pg 4 Line 8 – Are all these study sites in CA? For a European journal, the places should be better located.
These sites are in Nevada and Oregon. We've added the state in which they're located to the text.

Pg 4 Line 14 - This sentence is a little unclear as to what it means in context of the previous sentence. Because the lake is inhabitable, it means the lake and stream systems are disconnected?
We deleted this sentence.

Pg 4 Line 17 – Restore to tolerable salinity ranges? Restore is a broad word. The discussion of salinity is really interesting, but perhaps not relevant to your stream temperature focus.
We revised the sentence to "to restore Walker Lake to salinity to tolerable levels". We kept the discussion of salinity and Walker Lake to add context to the research.

Pg 4 Line 22 – Can you be more specific about what a 'dry year' means in this basin? I.e., received <25% of historical rainfall, or stream base flows were at XX% of the average flows for that time of year? (Especially since you do have USGS gauge data)
We removed this section following the recommendation of another reviewer. However, we added that in dry year 2012 snowpack was 50% of normal and in dry year 2015 snowpack was 5% of normal to the DTS and TIR subsections.

Pg 4 Line 26 – Comparing measured and modeled data? (rather than between)
We changed "between" to "comparing".

Pg 6 Line 31 – the East Walker has more flow than the mainstem – presumably because of diversions? Remind the reader of that again
Diversions vary streamflows in both the East and West Walker Rivers so that some days flows are higher in the East Walker River and some days they are higher in the West Walker River. We do not want to imply that West Walker flows are always lower than East Walker flows, so we made no changes to the text.

Pg 9 Line 30 – was the Elmore et al. 2015 study also conducted in the Walker R basin? This point may be better situated in the discussion.
We moved this sentence to the Discussion (section 5.2).

Pg 9 – Fig 2 caption – A reminder of what is the Wabuska Drain (not-flowing, but standing water from an ag ditch) would help the reader in the caption.
We added a description of Wabuska Drain, so the caption now reads: "Stream temperatures measured for the length of the DTS cable at East Walker River (a) and mainstem Walker River (b) DTS sites. Wabuska Drain, *which was not flowing but had standing water during sampling*, is located at cable distance 110-175 m in the mainstem Walker River site (b)."

Pg 10 – Fig 3 – The sub-panels in Fig 3 should be labelled. In looking at the figure, it's not clear why these two features are called out. It looks like there is more variation between river left and river right than anything else. Fig 3 b could probably be moved to supplemental.
We amended Figure 3 caption to include that insets show details of spatial temperature variability. When we complete the analysis of temperature differences between the right and left banks we can change the inset or remove the downstream-most inset, if needed. We moved figure 3b to supplemental information.

Pg 10 – Why do Fig 3 and Fig 4 visualize different times, 5:30 pm vs 3:15 pm?
Figures 3 and 4 visualize each reach at their maximum daily temperature. We felt this was more representative and comparable than picking an arbitrary time that is the same for both reaches.

Pg 10 Lines 19-31 – How different would this look if the Wabuska drain was flowing? Is it more often flowing or not flowing in the summer months?
We do not know how different results would be if the Wabuska Drain were flowing. Thus we added a line to the limitations section: "*Additional research is needed to quantify how results would change when the Wabuska Drain is flowing, or for deployments earlier or later in summer.*" The Wabuska Drain is usually flowing during summer, except during dry/very dry years when irrigators use more groundwater than surface diversions. The majority of years in the past decade have been dry/very dry years.

Pg 10 – Fig 4 – It would be nice to add a line or circle indicating where the beaver dam is on that reach of stream. Fig 4 b could also be moved to supplemental. However it might be worth exploring why there is day-to-day variation in the temperature ranges – could plot the data against air temperature range for the day?
When we complete our additional analysis we will add them to the map to label the beaver dam locations. We moved figure 4b to supplemental information.

Pg 11 Lines 10 – 11 Have LCT been observed using the Wabuska drain when it is flowing, or when stream temps aren't so high at the mouth?
LCT have not been observed near the Wabuska Drain that we are aware of.

Pg 11 Lines 31-32 Are there studies that show that LCT cannot move through warm temperatures? Maybe save the temperature implications for fish for the discussion, when you can do a more thorough review of studies on LCT (and related salmonids) and their thermal tolerance. They may be able to move through a small area of warm temperatures. What may be more important is the extent of cold water refuges, which could be quantified with the given data.
We moved all implications for fish to the discussion so that we could put the findings into context with the literature.

Pg 11 – Lines 33-34 – Make it clear that the cool water from Wabuska was observed with the other method/time window when stream temps were monitored.
We added 'during the TIR flight' to the sentence to clarify.

Pg 12 – Fig 5 – Similarly to the other figures, the sub-panels b,c,d could be moved to supplemental since they are redundant with the main figure. Alternatively, they may be a better way to present the data in the main figure, so you could consider dropping panel a.
We removed panels b-d from this figure.

Pg 13 – Lines 16-19 – Model estimated within 1C – that's pretty good! But under estimating by 2.5C is less desirable – could be a point to discuss.
We discuss this in section 5.3 where we add "*TIR stream temperature measurements in the lower reaches of the mainstem Walker River were considerably warmer than simulated results and remain near LCT lethal thermal threshold for an additional 45 km than was previously estimated.*"

Pg 14 – Lines 13-15, Maybe it didn't exceed 28C, but trout can be thermally stressed at much lower temperatures.
We added that trout can be thermally stressed at lower temperatures to the sentence.

Pg 14 Line 23 – Is this study from the Walker R Basin, or generally citing groundwater depletion from drought?
This study is from the Walker Basin. We made no change since the sentence specifically discusses interflow to the Wabuska Drain.

Pg 14 Line 30 – How migratory are LCT? Would they historically migrate between lakes and streams, and was that during the summer? If not, then you could shift your focus to movement and opportunities for longitudinal connectivity rather than migration.
LCT historically migrated between Walker River and Lake. However, we changed this sentence to focus on aquatic habitat longitudinal connectivity because it is more broadly applicable for cold water habitat and species.

Pg 15 Lines 9-10 Describing that there is a range in temperatures doesn't necessarily mean that there is thermal refugia. How cold is the water around these features? If it has high variability but it still warm, then it may not provide refuge for LCT.
We added and subtracted the temperature ranges that we measured from the modeled values. It doesn't mean there is always thermal refugia, but is suggests there may sometimes be refugia. We modified our wording to "*suggests that cool-water refugia may sometimes exist*" so that we do not oversell our findings.

Pg 15 Lines 27-29 This would be a good place to give a nod to the work that has evaluated how fish use thermal refugia (do a forward search on Sutton et al 2007).
We added 4 references (Fullerton et al. 2018; Frechette et al. 2018; Stevens and DuPont 2011; McCullough et al. 2009) to give a nod to current research on how fish use thermal refugia.