We thank Dr. Westhoff for his detailed review of our paper. We have tried to address each comment and improve our paper accordingly.

1: DTS measurements improved model performance

The authors claim that the model performance improved from a RMSE of 2 C for the 2001 model to 0.35 C when using DTS. These results are presented in section 4.2: Calibration results. However, to me it seems that no model parameter has been calibrated. The main difference with the 2001 model is the location of the upstream temperature boundary condition. Which caused indeed the improved model performance: the 2001 model had its upstream boundary condition about 10 km upstream of the currently investigated stream reach, while, when using DTS the upstream boundary was ~70 m upstream of the first node. With a longitudinal heating gradient of ~1 C/km one should have a really bad temperature model to have a RMSE of 2 C for the first km.

Calibrated parameters are described in the methods section (pg 10009-10010 and table 1) and include boundary condition flow rates, wind driven evaporative cooling (AA), thermal diffusivity of bed material (DIF) and bed albedo (BEDALB). However, we agree that attributing the improved model performance to the use of DTS is inappropriate because of the changed location of the boundary condition and model set up. Our statements relating DTS measurements to improved model calibration and performance will be removed in the revised manuscript. The paper will further be refocused to highlight the use of DTS to explore the variability of measured temperatures compared to results of more coarsely modeled rivers.

2: DTS can be used to post-process existing model results

I am not sure what the authors exactly mean with post-processing, but as far as I understand, the authors mean 'observing sub-grid heterogeneity'. This can indeed be done in this study since the model grid cells are much larger than the resolution of the observations. However, this is in most studies the other way around.

We will clarify our meaning in the text and revise our language to use the terms 'heterogeneity' and 'thermal variability' rather than 'post-processing'. As mentioned above, our intent is to demonstrate DTS data can be used to bracket the uncertainty and range of temperatures that occur in side channels, pools, and mixing zones not captured by coarse-grid models. In the Shasta River, a coarse grid model was used to evaluate promising habitat restoration alternatives to improve stream conditions for native fish. These included large scale alternatives such as relocating a major irrigation diversion, restoring Big Springs Creek which provides a significant inflow of relatively cool spring flow to the mainstem, and removing Dwinnell Dam which forms a small reservoir upstream. Previous work also considered smaller scale alternatives such as riparian shading and managing tail water return flow that typically introduces warm flows to the river. However, smaller scale trends could not be reproduced by the course grid model and thus field monitoring with DTS provides an opportunity to understand the small scale temperature dynamics and thermal variability to make restoration efforts more effective. Localized conditions related to flow and temperature can create barriers to fish passage or cause limiting conditions for spawning. Herein, DTS is used to explore temperature dynamics on a small spatial scale that can be used to inform interpretation of model results to fine tune restoration activities. We frame the discussion in the revised paper to reflect more of this background.

3: Air temperature may be a more important driver for water temperature than solar radiation

The authors come to this result by water temperature with both solar radiation and air temperature, where the latter correlates better. However, the authors say that understanding the causation is outside the scope of this article (P10015, L10) they have mentioned the reason for the good correlation already in Line 2 of the same page: 'solar radiation is the major factor influencing both air and water temperatures'. Air temperature is thus not a driver for stream water temperature, but subject to the same driver. The conclusion of the authors is thus erroneous.

We agree with the reviewer's statement that solar radiation is the primary driver for both air temperature and water temperature. Our observation is that the modeled water temperature correlates more closely with solar radiation whereas the measured water temperature correlates better with the air temperature. Therefore, the modeled effects of solar radiation on water temperature could be improved. We will do a major revision of the text per the reviewer's comment to clarify this point.

4. This research contributes to the literature by demonstrating the value of long-term DTS observations for model calibration and increased confidence in simulated temperatures

This claim is stated on P10002 L17-19, but nowhere in the manuscript the authors 'proof' this or even refer to it. In my opinion this claim may be only correct when over the cause of the observation period different processes occur that can then be parameterized separately. I do not see the benefit for the current study

We agree this statement early in the text is not returned to and 'proven'. It will be removed and instead the manuscript will focus on exploring the thermal variability captured by DTS (and which was not well-represented in the simulation model). Further we evaluate one dimensional thermal variability with DTS monitoring, and discuss how it can further refine understanding of potential habitat restoration activities on stream temperatures by identifying locations of local cool water refugia or reaches where stream temperatures limit native biota.

Reviewer #1 has additional concerns that some aspects of the temperature model have either not been explained or were used in a nonoptimal way. We provide quick answers to questions below, which we also incorporate into the manuscript text so that other readers do not have similar concerns and questions while reading this paper.

• Is RQUAL based on the advection-dispersion equation?

Yes, using the Holly-Preissmann numerical method. This will be made more explicit in the text.

• Are the locations of the nodes the centre of the grid cell or are they located at the upstream end of the grid cell? This may not be so important if the nodes would only be a few metres apart, but in the current setup it is important. It determines in which grid cell lateral flows come in and which DTS points should be averaged to have an observation for each grid cell.

They are in the center.

• I have some concerns about the 1 hour time step of the model. Flow velocities or water depths are not given, but if I assume a flow velocity of 0.1 m/s (water depth ~1m), the Courant number is 4 or 5 (based on a 80 m distance between the nodes). This means that on each time step, a water parcel (with a certain temperature) will, in reality, travel over 4 or 5 nodes, while in the model it only moves to the next downstream grid cell. Although the model may numerically be stable, the results may be less accurate.

We will explore this issue through sensitivity analysis and see if a smaller time step changes model results.

- **P10005, L16: Are the DTS measurements single or double ended?**

  They are single-ended, which is clarified in the manuscript

- **P10008, L9-10: Why are the DTS measurements not averaged over the full length of the grid cell?**

  Averaging them over the full length of the grid cell generally would increase the length of the DTS reach used to compare measured to modeled temperature at each node. Doing so however increases the Mean bias, RMSE, and MAE upwards of only 0.19^{o}C with the greatest difference (>0.1) occurring at nodes 9-11. The upstream nodes have no significant difference (<0.03oC different). This suggests that had the entire reach of the grid cell been used for calibration purposes, the model results would not have been significantly different. Table 2 will be updated to reflect DTS reaches corresponding to averaged temperature over the full length of the grid cell per the reviewers comment.

- **P10009, L14-17: This inflow may also be estimated by using a simple mass balance for temperature: Qdown Tdown= Qup Tup+ QPCO TPCO.**

  This is a good suggestion and was considered by the authors. However, since there was not a reliable time series of upstream flow measurements (Qup) due to macrophyte growth and inability to establish a rating curve, this introduced a second unknown making the mass balance approach unworkable. That said, Westhoff et al. (2007) includes an approach using measured temperatures to estimate the ratio of the flow of the lateral to the flow of the downstream mainstem.

$$\frac{Q_L}{Q_d} = \frac{T_d - T_u}{T_L - T_u}$$ L=lateral, d=downstream, u=upstream, T= temperature, Q=flow

Use of this approach will be further explored for inclusion in the revised manuscript to provide a basis for how the lateral inflow changes relative to the mainstem flow.

- **P10010, L22: The difference of 0.2 C can also be caused by the fact that the Hobo temperature logger that measures the upstream boundary condition has an error of 0.2 C compared to the DTS measurements.**
  This is a good point and we will add text indicating that the difference is within the uncertainty of the Hobo temperature sensors per reviewer's comment.
- **P10010, L16: The initial stream temperature is not important if the warming up period is long enough. For this model a warming up period as long as the travel time of a water parcel may already be long enough (except from some longer memory of the riverbed temperature).**
  Comparing modeled vs. measured temperatures at each node plotted with the initial stream temperature boundary condition (as recorded by the Hobo data logger), the first several nodes of the model track closely with the boundary condition. However by node 5 or 6 the modeled temperatures clearly deviate from the boundary condition temperatures and share daily peaks closely with measured temperatures. By and large, this paper evaluates modeled vs. measured temperatures with regards to thermal variability in the downstream portions of the modeled reach, far from the boundary condition where the model is simulating stream temperatures based on modeled physical processes.
- **P10011, L26-27: The claim that the authors were able to quantify the size of the mixing zone is a bit too strong: the stream is about 11 m wide and only one point over the width has been measured.**
  We will change the wording to state that the DTS measures elevated temperatures for about 40 meters downstream but that due to the limitations of the cable location and stream characteristics, this may not reflect the true length of the mixing zone.

  Also: downstream of the PCO there is a curve in the river which may cause the plume to go from the left to the right bank. Could it be that downstream of this curve the observed temperature returns to 'normal' values (see Fig 5a, at ~750 m).
  This is possible however it should be noted that the DTS cable goes from left bank to right bank before the curve. We will emphasize the uncertainty of the 1D approach and that we do not have enough information (2D measurements/models) to know whether the plume goes from left to right bank.
- **P10016, L20-21: Refer to Krause et al. (2012) who used two DTS line in one cross-section.**
  This is a good suggestion and we will add a reference to this work.

- **P10018, L1-4: I do not agree: measurements are needed to setup a model anyway, so why not immediately applying DTS to obtain these measurements.** We agree that in conducting field monitoring *prior* to modeling a system, choosing to use DTS could be of great value. However, our paper explores the case in which a course-grid model of the system has already been developed and used to assess restoration activities for improving flows and temperature conditions for fish species of concern. Where stream temperature models have already been developed for systems, use of DTS technology post-modeling has value in assessing small scale variability. In the case of the Shasta River, many restoration alternatives were taken off the table since the 2001 model (riparian shading in lower reaches, moving a major irrigation diversion…). From previous research, some restoration strategies were identified (managing cool-water springs) and subsequent monitoring (with DTS) can explore if thermal refugia exists locally or how tail water flows influence mainstem variability. We will add these thoughts to the discussion of the manuscript to clarify our intent and claim.

- **Figure 3: This figure is not of much use.**
  The figure will be removed.
- **Figure 4: Also add the time series of the boundary condition. This may explain the good fit and maybe even the differences in observed and simulated daily minimum temperatures.**
  The boundary condition will be added to this figure.
- **Figure 9: Which DTS points were used to represent node 9? Were some of these points already influenced by the PCO water?**
  This figure needs some correction and clarification. Since the lateral comes in to the model at node 9, node 10 should instead be used in the graph to reflect downstream conditions influenced by the PCO inflow. This does not drastically change the appearance of the graph but it will be corrected. This then is contrasted with measured temperatures at a location on the DTS in the mainstem but right at the confluence of the PCO. This location likely reflects mainly PCO water temperatures rather than mixed conditions, as captured by node 10 modeled temperatures. This demonstrates the wide variability of conditions in the mainstem resulting from the PCO inflow. It shows the extreme of the thermal variability. This explanation is added more explicitly to the manuscript.

References:

Westhoff, M. C., Savenije, H. H. G., Luxemburg, W. M. J., Stelling, G. S., van de Giesen, N. C., Selker, J. S., Pfister, L., and Uhlenbrook, S.: A distributed stream temperature model using high resolution temperature observations, Hydrol. Earth Syst. Sci., 11, 1469-1480, 2007.