Reviewer 2: Anonymous

My main concerns are the following: i) the link between modelling results and CTD observations is weak, lacks description and a convincing discussion. ii) the calibration of the modelling chain lacks description and convincing results. iii) an uncertainty discussion of the modelling results is missing. iv) the structure of the ms is at some locations mixing methods, data description and results v) due to the concerns above the conclusions are vague, speculative and lack conviction.

Major concerns: 1. The linkage between FWV and FWC would require a thorough discussion of the mixing of freshwater into ocean water. Numerous paper exist on this topic but this ms fails to review the literature and discuss this complex topic in a convincing manner.

Response: The manuscript originally contained a more detailed explanation of oceanographic processes, with a more thorough discussion of freshwater mixing. However, it is not within the scope of this manuscript to review in great detail the theoretical extent of these oceanographic processes, nor is it especially relevant to the readers of The Cryosphere. Instead, we added a paragraph in Section 1 in the manuscript that focuses on the bioecological effects of stratification in Glacier Bay specifically, in order to address this comment and to keep the focus on the study area. The paragraph added to the intro section in the final draft of the manuscript is below in italics.

Long-term shifts in terrestrial freshwater storage and runoff can have significant implications for oceanographic stratification and circulation that moderate biogeochemical and ecological activity within Glacier Bay. Since Glacier Bay is a highly understudied, relatively remote national park, the complete freshwater budget for the bay cannot be quantified due to the lack of available data. However, seasonal trends in modeled freshwater runoff can be qualitatively compared with seasonal trends in broadscale oceanographic salinity records from 1993 to present collected by the U.S. National Park Service's Southeast Alaska Inventory & Monitoring Network (SEAN). This SEAN dataset served as the basis of the analysis performed by Etherington et al. (2007), which found positive correlations between phytoplankton biomass and stratification levels. The competing forces of macro-tidal flushing and strong stratification within the glacially-carved estuary generates temporally and spatially shifting trends in upwelling and nutrient availability (Etherington et al., 2007). Thus, accurate estimation of projection scenario runoff into Glacier Bay plays a paramount role in constraining future changes in water and nutrient circulation.

2. The description of the calibration is weak: if the authors claim to make realistic projections of future FWV into the ocean, I would expect a thorough discussion of the efficiency regarding snow melt, ice melt and rain runoff of the model; the necessity of multi dataset calibrations for hydrologic modelling under climate scenarios have been discussed in the literature.

Response: This comment from Reviewer 2 about the calibration description is warranted, and in the light of other reviewer's comments, the authors added a paragraph in Section 3.4 to clarify our calibration decisions and process. It should be noted that one of the primary reasons the authors chose to use the historical oceanographic dataset is because long-term observational datasets of stream flow and weather conditions within the boundary of Glacier Bay National Park do not exist. This is also why we chose to calibrate the model with observations and records from the nearby Mendenhall Glacier, as further explained in more detail in the new paragraph in Section 3.4. The paragraph added to the calibration section in the final draft of the manuscript is below in italics.

Recent studies (Beamer et al., 2016; Lader et al., 2016) have investigated the accuracy and biases of the MERRA reanalysis product in coastal Alaska compared to other reanalysis products such as ERA-Interim (Dee et al., 2011), CFSR (Saha et al., 2010), NCEP-NCAR (Kalnay et al, 1996), NARR (Mesinger et al., 2006), and others. Many SnowModel parameters were tested by doing a sensitivity analysis for each reanalysis product, including monthly precipitation adjustment factors, snow/rain temperature thresholds, snow and ice albedo factors, and more (see Beamer et al. (2016) their Table 2). For each of 4 reanalysis products, they calibrated model parameters based on observations of streamflow (Q) and glacier mass balance (B). The MERRA simulation Coefficient of Determination scores (r2) for glacier mass balance (B) and stream discharge (Q) for the Beamer et al. (2016) study were 0.80 and 0.95, respectively, and the Nash Sutcliffe Efficiency (NSE) scores were 0.67 and 0.91, respectively. While Beamer et al. (2016) identified the CFSR product as the 'best overall' for the GOA region, they found that MERRA was superior at the Mendenhall Glacier observational station, which is the closest calibration point (< 25 km) to GBNPP. For these reasons, in this study we rely on the model calibration of Beamer et al. (2016; their section 3.4) and we adopt their calibration parameters for SnowModel from their Table 2 and Table A1.

Long-term glacial mass balance programs and long-term streamflow gauge datasets do not exist within the GBNPP study area, thus constraining our ability to conduct additional calibration efforts. While the Mendenhall Glacier observation station is close in proximity to Glacier Bay, the glacier has receeded and thinned significantly since the early 1900's, glacial wastage is a significant component of annual streamflow (17%), and glacial meltwater contributes heavily to streamflow in the summer (50%; Motyka et al., 2003). As a result of these similarities in geography and hydrology, we rely on the calibration process, parameters, and best-performing reanalysis product (MERRA) from Beamer et al. (2016) for our study.

3. The uncertainty of the results are not discussed; are the projected changes significant? What is the uncertainty of the future scenarios?

Response: Please see the response to another comment from this Reviewer below about uncertainty and the variability of the historical simulations. Here, the Reviewer raises an important point, and included below is an analysis of the variability of the historical simulation results for the all watersheds in GBNPP. This table includes historical annual variability (+/- $\boldsymbol{\sigma}$ standard deviation in the table below) for the GBNPP study area and various grouped watersheds for the modeled output for runoff, snow precipitation, and SWE. This table includes the monthly climatological average (36-year) of the variable, the standard deviation of the 36-year monthly values of the variable, and the percentage of the climatological average represented by the +/- one standard deviation. The annual variability in the historical runoff averages 9% of the annual runoff when averaged over the entire GBNPP domain. The standard deviation can be calculated annually or monthly, by grouped or individual watershed, or over the entire domain for every variable. The reasoning for not including variability (***σ***) in every monthly figure is to cut down on visual clutter and also because the variability differs by a few percentage points from watershed to watershed.

Table: Historical variability (standard deviation of 36-year climatology; ***σ* (m)**) for annual runoff, annual snow precipitation, and annual snow water equivalence (SWE) spatially aggregated for each grouped and individual watershed.

	Annual Runoff			**Annual Snow Precipitation**			**Annual SWE**		
Watershed	**(m)**	***σ* (m)**	**(%)**	**(m)**	***σ* (m)**	**(%)**	**(m)**	***σ* (m)**	**(%)**
GBNPP	3.4	+/- **0.3**	9	2.0	+/- **0.3**	15	1.1	+/- **0.2**	18
North	3.1	+/- **0.3**	10	1.9	+/- **0.3**	16	1.1	+/- **0.2**	18
West	4.3	+/- **0.4**	9	2.1	+/- **0.3**	13	1.2	+/- **0.2**	17
West-Arm	3.4	+/- **0.4**	12	3.3	+/- **0.5**	15	1.9	+/- **0.3**	16
East Arm	3.6	+/- **0.4**	11	1.8	+/- **0.3**	17	1.0	+/- **0.2**	20

Additionally, Reviewer 2 asks about the uncertainty of the forecast scenario results. First, because of this comment and the comments of another reviewer, the authors chose to adopt the language of 'projection scenario', instead of 'forecast' scenario for the final draft of the manuscript. Adopting the term 'projection' is more in line with the original intent of the manuscript. As Reviewer 1 notes, "I would suggest using "projection" instead since this term acknowledges additional uncertainty involved in climate scenarios which arise from uncertain greenhouse gas emissions." We highlight our response to Reviewer 1 here: *We are not attempting to make a climate prediction (as defined in the provided weblink;* ***http://glossary.ametsoc.org/wiki/Climate_prediction****) of what will happen in the future in Glacier Bay. We are modeling one of the potential scenarios that may occur in the hydrology of the region that would accompany the RCP8.5 emissions scenario. For these reasons, the authors have chosen to change the term 'forecast scenario' to 'projection scenario', in every instance, throughout the manuscript. These changes are added to the final draft.* The authors have also added text in Section 3.5.1 to make this point clear in the final draft of the manuscript.

Lastly, the annual variability of the modeling results can be quantified, and the table above shows the historical variability (standard deviation) of runoff, snow precipitation, and SWE. However, there are many ways uncertainty affects the modeling results that are difficult to characterize and unrelated to annual variability. For example, spatially explicit modeling at 250m resolution is a simplification of a hydrologic system and environment that, in reality, operates at infinitely smaller scales. There are also environmental processes imperfectly described my model physics or imperfect model parameterizations that increase the uncertainty of the results. Additionally, a weather reanalysis product is used to force the model in the absence of long-term, in-situ weather station data, and there are errors and biases associated with reanalysis data assimilation and interpolation processes. The RCP 8.5 projection scenario then adds more layers of uncertainty, due to GCM model resolution and physics, and the associated likelihood of greenhouse gas concentrations in the future. Fully quantifying each of these uncertainties is not the aim of this study, and we acknowledge that this is an important limitation of the current study. We simultaneously think that since no long-term weather station records, stream flow gauges, or glacier mass balance programs exist within the study area, our physically-based, spatially explicit modeling approach is a valuable inquiry into the hydrology of Glacier Bay. It advances the current knowledge of a system that is not well characterized or measured, and it advances our understanding beyond the current literature.

We are not attempting to make a prediction of the likelihood of future hydrologic conditions, only describe the historical and projection scenario climatologies of these conditions. To address these concerns, we changed the language throughout the manuscript from ‘forecast’ to ‘projection scenario’ to clarify future scenario results, and we’ve added multiple clarifying sentences about the purpose and aims of the study (Abstract, sentence 3; Section 1, para 8; Section 3.5.1, para 2).

Specific comments: 1) Title: what is hydrologic diversity? This term is never mentioned in the ms accordingly it seems misleading to use it in the title. I would be helpful to have a title that reflects the content of the ms.

Response: Another reviewer commented on the ambiguity of this term in the title too. The title has been changed to *Seasonal Components of Freshwater Runoff in Glacier Bay, Alaska: Diverse Spatial Patterns and Temporal Change.*

2) Abstract: An introductory sentence explaining the problematic and the purpose of the ms is missing;

Response: The authors are comfortable leaving the abstract summarization of the research questions and problems primarily intact. Some of the language in the abstract has been changed to clarify the original intent of the manuscript. Additionally, the authors have added a single introductory sentence (in italics below) explaining the purpose of the manuscript.

The purpose of this study is to characterize the recent historical components of freshwater runoff to Glacier Bay and quantify the potential hydrological changes that accompany the worst-case climate scenario during the final decades of the 21st century.

L16 why "wide variety", the same "variety" exist in any glaciated catchment;

Response: To address this comment about a 'wide variety' vs 'variety', the authors removed 'wide' from the final version of the abstract.

L24: this sentence is redundant, as it does not contain any conclusive information about the study:

Response: After further review, the authors agree that this sentence is redundant, and the sentence has been clarified in the final version to reduce repeated ideas.

3) Introduction: Nice description of the study site; however, a description of the linkage between fresh water inflow in an ocean bay and the subsequent impacts on marine life is missing; also a review of the literature of intruding freshwater into water bodies would be helpful.

Response: The authors are attempting to balance the length of the manuscript and the scope of the ideas covered in the intro section. As previously mentioned, we added a paragraph in Section 1 to address some of the missing linkages between freshwater inflow and impacts on marine life.

pg3, L 15: "the goals are different": it would be helpful to outline the goals;

Response: To address this comment about the last paragraph in Section 1, and other related comments from other Reviewers, the authors added two sentences in this paragraph in the final draft to clarify the goals of the study.

L20: the results present do not convincingly present changes in the coastal runoff (see major concerns).

Response: The text in this line of the manuscript originally read '*The results of this study will add to the understanding developed by Etherington et al. (2007) and will provide constrained estimates of how much the coastal runoff in GBNPP will change in the future*.' After acknowledging the comments from this and other reviewers, the authors have changed this sentence to more accurately reflect the original intent of the manuscript.

The sentence now reads, '*These results will add to the understanding developed by Etherington et al. (2007) and will provide constrained estimates of potential changes in runoff in GBNPP under the RCP 8.5 projection scenario*.' This change in language is important, and the scrutiny of this sentence by the Reviewers makes sense. We are not attempting to make a climate prediction or define the likelihood of what will occur hydrologically in the future. We are saying that if the greenhouse gas concentrations associated with the worst-case scenario RCP 8.5 come to pass, the hydrology (runoff, precipitation, SWE, etc.) will be affected in the specific ways that are outlined in our results and discussion sections. This difference is important to note and we have made an effort to clarify this throughout the paper.

4) Methods1: pf4, L 8: model chain? Only two models are used, one for the reanalysis of the forcing data and the SnowModel.

Response: It is true that SnowModel and MicroMet are the only two models used for this study. Therefore, the words 'model chain', and every reference thereafter, were removed from the manuscript.

5) Methods2: I think the clarity of the ms would improve if methods and data were two separate chapter;

Response: The methods and datasets have been presented in the same section of the publication because the input datasets are closely linked to the modeling process. The authors think the article benefits from the inclusion of these two parts in the same chapter, and it is quite common for articles in The Cryosphere to present data and methods together. However, as noted by Reviewer 3, some of our methods were first presented alongside results in various results paragraphs. The authors have removed all of the text about methods from the results and placed them into the methods sections for clarity.

6) Methods 3: 3.4. describes in a very rudimentary way model calibration; r2 and NSE values are provided. Since the authors claim to provide an "added understanding" and "constrained estimates of how coastal runoff will change in the future" I would expect a thorough discussion on the efficiency of their modelling in regard to runoff, snow melt and ice melt contribution during the calibration period. If the calibration is not presented adequately, how can one trust in the results of future runoff?

Response: This comment makes sense, and the original version of the calibration description was lacking in clarity and length. As previously mentioned, we added description and details about the calibration decisions to Section 3.4 in the final draft.

7) Results: pg7,8: here results and methods to calculate the results are in the same chapter; I think a clear separation between methods and results would be helpful;

Response: This comment is similar to other Reviewer comments about the results section occasionally mixing descriptions of the methods into the text. We have made an effort to move all of these discussions about methods to the previous section whenever necessary in the final draft.

Additionally, the authors want to clarify that the descriptions of the results through Eq. 2, Eq. 3, and Eq. 4 are not equivalent to descriptions of methods. We are attempting to clarify exactly what is represented in the results figures, so that readers can easily understand what is meant by a changes in metrics like temperature, precipitation, and snowfall equivalent to total precipitation. These metrics, and changes in their values from a historical period to a projection scenario, are not methods for producing modeling results. They are metrics that describe the results produced by the simulations. There could be many ways to describe model output in terms of temperature, precipitation, and snowfall, and we are making an effort to be very clear about how we are presenting these changes in the results section. For these reasons, the authors think the descriptions of these equations belong in the results section and not in the methods section.

8) Figures 1, 2 and 5 (in total 8 maps) all show specific aspects of the study site; this information could be combined and presented in one or perhaps 2 large panels.

Response: Since Figure 5 is a visualization of the results of changes in glacier coverage between the historical period and the projection scenario, the authors think Figure 5 needs to remain on its own as a depiction of glaciers in the study area.

The nested watersheds study design, depicted in Figure 2, is complex to visualize in the same map. We are presenting a study design that includes different watershed scales, the entire study area as a group (GBNPP), the grouped watersheds that flow into Glacier Bay and the Pacific Ocean, the grouped watersheds in separate arms of Glacier Bay, and the individual watersheds that lie within these other areas. Visualizing these on the same map is likely to be difficult to interpret by readers and presenting them in separate panels, side-by-side was an important decision by the authors to clearly depict the study design. Additionally, due to another Reviewer's comment, we decided to add labels for Glacier Bay, the Pacific Ocean, and Icy Strait in Figure 2.

As for Figure 1, the authors think that it is necessary to show the larger region of interest, as well as the digital elevation model/bathymetry and intend to keep Figure 1 intact.

9) Figure 3: I do not understand why contour plots are used here; bars indicating the exact value of T and P change would be more helpful.

Response: Color heatmaps are one of the best ways to convey hundreds of values, along multiple axes through the use of a color gradient. The authors chose to use the heatmap to visualize these changes in temperature and precipitation because it efficiently and intuitively communicates 104 different values in a compact and clear package. Since this is a 3 panel figure, the authors think that displaying 312 bars or points representing these changes or 312 numbers in a large table would not clearly or succinctly communicate the changes in these 3 variables over time.

10) Figures 4, 6, 7, 8, 9 and 10: it would be helpful to add an uncertainty to each point; e.g. stdev from the mean over the 30 yrs (but this would only account for climatic availability); I recommend checking recent literature on this topic.

Response: We've chosen not to include the variability at each point for the modeling results in these figures because we want to simplify the visualization of the results. However, the authors have done some additional analysis to show the reviewers the standard deviation from the mean for some of the historical climatologies in runoff, snow precipitation, and SWE. See Table above for all of the grouped and aggregated watersheds and the authors' response to an earlier comment about uncertainty and variability.