This manuscript presents a detailed implementation of a 3D CFD model for a 30-km transect of the Columbia River. The model is calibrated with surface water elevations (SWE) and validated with SWE and velocity profiles. As part of this effort, an analysis of the model performance at three different time scales was performed: (1) short-term (2012), (2) middle-term (2013-2016), and (3) long-term (2018-2019).

The authors present a novel and practical approach to impose boundary conditions and calibrate roughness scales and shear stresses in large-scale 3D CFD models. After calibration and validation, the authors compared the new model results with previous modeling efforts in the study site. This analysis highlighted the importance of a detailed representation of 3D processes in channel flow modeling. Finally, the authors assessed the relative importance of dynamic and hydrostatic pressure along the reach, focusing on the potential implications for environmental and ecological functions in river systems.

**Summary:** Overall, this manuscript is well-written, clear, and represents an important contribution to the field of river hydraulics. The implementation approach is well-documented and reproducible. This will make for a significant contribution to GMD.

I include a commented pdf with editorial suggestions. In the following, I present two significant comments that require attention and a series of general observations.

**Major comments:**

1) The authors' argument for using a spatially-variable roughness ($k_s$) is that this parameter is expected to vary for a complex river system with heterogeneous bedforms and high curvature. I agree with this argument. However, the selection of roughness regions seems guided by geometric convenience and data availability and not the spatial variations that are reasonable controls for $k_s$. For instance, I would expect that a characterization of the different depositional environments and corresponding bedforms would provide a better guide for selecting roughness zones. The PNNL team has previously performed such classification for the study reach, which could be used. The main issue here is that given the complexity of this model, the authors are likely dealing with a non-unique solution, and the computational burden prevents them from analyzing this in detail. Some discussion in this regard would be valuable in this manuscript and essential.
to guide future model refinements. I suggest including some discussion in lines 415-422.

On a related note, I wonder how sensitive is the spatial variation of stage and velocity to the calibrated $k_s$ within a roughness region. Again, the authors assess this sensitivity at the point scale, where the SWE observations are available, but the overall response may be insignificant.

2) The idea of analyzing the relative importance of dynamic and hydrostatic pressure is an excellent illustration of the importance of these models to gain a mechanistic understanding of exchange processes along the sediment-water interface. I commend the authors for including this analysis! However, I expect the conclusions regarding the exchange to be incorrect. The reason for my skepticism is that the exchange process is driven by gradients in the pressure distribution and not by its magnitude. In other words, I suggest that the authors revise this analysis and focus on the spatial variability of the pressure gradients.

General comments:

- p7, l 180: For clarity, the roughness elements directly resolved are larger than 1m in "the vertical direction." To be precise, you could include "the vertical direction" in the text since the horizontal resolution is much lower (20m).
- p9, l 206: the velocity components for the outlet seem to have the x- and y-direction components mixed
- p 11, l 255: To better illustrate the potential presence of systematic biases, I suggest plotting error vs. stage. For example, the bias for low SWE in Figure 9 will be more evident with this metric.
- p 11, l 261 (and throughout the text): Using the tilde symbol "~" for value ranges is somewhat unconventional. I suggest using a dash "-".
- The authors use SWE to assess the model performance; however, this metric could be misleading, and I wonder if the water depth is a better alternative. In particular, when calibrating the roughness values, I expect the relative error in water depth to be a more reasonable measure of model performance.
- Labels and text in Figure 7 are hard to see.
- p 14, l 290: is it possible that the disagreement for high curvature results from using a constant roughness value for a region with varying depositional characteristics and bedforms?

Please also note the supplement to this comment: https://gmd.copernicus.org/preprints/gmd-2021-313/gmd-2021-313-RC3-supplement.pdf