Wong et al. is a well written paper that will be useful for the chemical-transport modeling community as well as the community making ozone flux observations. Wong et al. investigate four ozone dry deposition parameterizations used commonly in global scale chemical transport modeling, including how they compare against observations, how much interannual variability is simulated, and whether there is a trend in ozone deposition velocity over the past 30 years. The authors then examine the impacts of interannual variability and long-term trends in ozone deposition velocity may have influence ozone air quality. This paper is the first to examine such questions at the global scale. One of the paper's strengths is that it suggests that any one of four current parameterizations for ozone dry deposition is not necessarily better than another in terms of capturing observed ozone deposition velocities at several observational sites around the world. Another interesting finding is that even though the interannual variability in simulated ozone deposition velocity is muted compared to three ozone flux datasets with ~10 years of measurements, the simulated interannual variability has implications for simulated surface ozone concentrations.

Major issues
1. the linearity of response of surface ozone concentration to ozone deposition velocity is uncertain, but a major assumption in this study. i'm not convinced that the results from wong et al. 2018 are sufficient to warrant confidence in this assumption. one reason being that they were testing the response to surface ozone to LAI, which involves changes in several processes.
2. the authors' attribution of biases and intermodel differences are entirely speculative. there is no rigorous evaluation of the processes/aspects leading to differences. i tend to not be in favor of such speculation and I think it masks the strength of the model evaluation (that not any one parameterization is best or worst) and model intercomparison.

Minor issues
10: I tend to think the sinks of ozone are chemistry and dry deposition so "second largest sink" doesn't say much to me
15-16: "to drive four ozone dry deposition parameterizations"
62: I wouldn't say Silva & Heald 2018 is a review
66: "account for" is vague; in general this sentence implies canopy column models are better than big-leaf ones, which has yet to be shown in the literature
67: the authors said previously that reaction with BVOCs is a nonstomatal pathway so here saying that it is in addition to surface sinks is a little confusing
67-71: canopy column models still use resistance networks …
77-80: this has yet to be shown… these formulations can be variable across models …
80-88: the connection between these paragraphs (last sentence of previous one and first sentence of next one) could be articulated better
101: Hardacre et al. show factor of 2-3 differences across models - so are all models' seasonal cycles well represented? also I suggest changing "demonstrating" to "suggesting"
125: "unable" seems harsh; it doesn't seem Clifton et al. even tried to do this
128: cut "physics"
145: I find the placement/existence of this sentence strange. the authors don't investigate the same parameterizations that Wu et al. do.
153: refs for strong empirical relationship
162-173: I see that the authors have basically organized their parameterizations according to model (w/ exception of #2)
1) The GEOS Chem parameterization
2) Zhang parameterization
3) The CESM parameterization
4) The UKCA parameterization
I didn't realize this at first and the parameterizations chosen seemed quite strange. I would urge the authors to re-frame their parameterization description (but also noting that their parameterizations are not exact replicates of a given model)
175: It doesn't quite make sense to me that the authors say the Zhang parameterization is "open source" in one sentence and a couple sentences later say that implementing it required personal communication with Zhiyong and Leiming.
180: Given that GEOS Chem doesnt have a land surface model, I think the authors need to clarify how exactly Anet is calculated.
182-183: It's fine not to test Ra and Rb, but i suggest that the authors do not use this qualifier. This isn't well understood (Does Fares et al. even show this?)
188-9: has this model been evaluated? or used previously?
194-5: what are these variables used for?
195: presumably the authors' decisions about land type mapping (& differences for "W89" vs "Z03") impact the authors' results… one implication of this is that the authors' statement in the abstract or introduction that the only thing different across parameterizations is the model structure is not necessarily true
197: i would suggest cutting the "(eg. leaf physiological and soil hydrauilic constants)" - becoming more specific here doesn't help readers when the parameterizations are not laid out and we have no idea what these terms do/stand for
198: what's z0?
198: how is leaf wetness calculated? how is snow calculated?
203: how do the authors scale PFT-specific LAI? is there an established method of doing this? presumably this has implications for the findings
217: i think the authors need to articulate here or in the introduction the various effects that high CO2 may have on ozone dry deposition velocity and the various uncertainties in our understanding of CO2 fertilization (& reference previous work examining this)
229: is the proper/up-to-date way of referencing GEOS-Chem?
237: binned = jargon
243-246: discussing about dry deposition of other species and impacts on ozone requires introducing some concepts (or cutting talking about dry deposition of other species)
249-251: this seems like a strange choice to me. it's not differences in transport per se, it's differences in background ozone caused by changes in ozone dry deposition. why wouldn't the authors want to capture this? because it contributes to nonlinear responses to ozone dry deposition?
249: what is the baseline simulation?
254: Why not CLIM+LAI+CO2 as well?
261-3: How many sites does this cut?
265: Fractional coverage of what? (please spell out in text) Why are these figures shown? they are not very useful for the reader
270-1: Not sure what the point of this sentence is
273: it seems strange to me that the authors would generalize such as bias, given that its unclear if the bias is caused by a particular attribute of an land type or process, and that the land type-specific biases differ across the parameterizations
282: what does N=5 mean? 5 sites? 5 data points?
288: if the authors are implying ambient chemistry is happening then they should just say it
300: meaning that the authors do not leverage it
301-302: I'm not sure that the following lines illustrate this; in other words, i think BB "generally but not universally leads to improvements" is not supported by the actual findings – it seems to be for Z03 – but not for Wesely – which may suggest that we need to be paying attention to nonstomatal deposition estimates too.
313-4: what particular problem has been highlighted?
315: sampling biases meaning that the authors are not evaluating most locations on earth, right? the authors are sampling the time/place of the measurements
317-320: not sure what the point of this paragraph is. what is the hypothesis being investigated?
334-5: recommend that the authors don't speculate here or elsewhere
349-50: on a similar note as the above comment, how do the authors know this?
353: "is not desiccated"?
358: i don't think the authors show this; they just speculate that this is the cause.
368: will the authors more carefully articulate what Centoni finds so that the reader knows how to compare the findings
370: i assume that the authors are identifying the hot spot regions through their large delta O3. related: perhaps the authors are missing a delta on the v_d,i in Equation 3.
378: are the authors really "exploring the importance of seasonality in predictions of vd and their subsequent impact" with their current approach? (see comment below for line 404)
382-4: i suggest a semi colon connecting these two sentences
385: "shifts from the south to the north relative to July"
387: i'm not a fan of the authors' use of the term hydroclimate – it's vague – can the authors just say soil moisture or VPD or leaf wetness?
398: the suggestion that "hydroclimate [is] a key driver of process uncertainty" seems limited to the tropics/subtropics. am i correct in this interpretation? if so, this should be emphasized.
404: are the authors actually showing the impacts on seasonality? showing the impact in each season is not the same as showing the impact on seasonality (a couple of easy calculations could help here)
409: briefly describe this method such as the limitations/strengths of it
413: what trends? trends in meteorology, LAI, and/or CO2?
415: how is the annual change in vd estimated? is it using the Theil-Sen method? this part needs better explanation; the reader needs to at least have some concept of what the method used is
423-4: but they are small or nonsignificant per the first line of the paragraph?
439: or it may decrease as plants acclimate or as nutrients become limiting
452: assuming that ozone dry deposition should be a strong function of LAI
455: "complex"
466: "suggesting"
466: suggestion to cut "natural" here and in other spots - natural IAV has ambiguous meaning
475: heterogeneity?
478: soil moisture data?
480: refs for good performance at site level?
495-6: whether IAV in vd at Blodgett is caused by chemistry is unknown
491-497: steps on how authors calculated averages and CVs for long term data needed
499: Olivia has a new paper on this
526: a vague reference to an effort in asia doesn't do much to help the reader
527: "constrain"; why all of a sudden call it gaseous dry deposition?
528: what do the authors mean by reported? do they mean in the peer reviewed literature? there are many reasons why people report fluxes rather than deposition velocities in peer-reviewed publications, and previous work doesn't simply exist to provide deposition velocities for future model evaluation! many datasets are available by contacting the research groups that made them.
536: do the authors actually show that the four parameterizations differ most in leafy parts of the world? if not, i suggest rephrasing
542-544: is this something that is assumed widely?
543: demonstrates that
549: why "at least increase the spatiotemporal representativeness if not the absolute accuracy" - is there some limitation of the Ducker dataset that I am missing?
554-6: the authors could do a better job at illustrating why they are linking these two ideas
561-3: yet the authors barely make use of long-term datasets that are available!
583: what does low baseline vd actually mean?
586: v_d
587: do the authors mean the simulation year for the 30% testing?
588: is this somewhat inherently in the LAI product?
593-600: as is, this seems like a stretch to me
608: what is the difference between a model-observation integration and an empirical study?