Response to Referees' Comments

We thank the referee for their comprehensive and constructive comments on our manuscript. Below, we address each specific point in turn:

This paper describes coastal ozone flux measurements made at a location on the south coast of the UK. The paper builds on previous techniques to process & understand the data including its uncertainty. The paper includes a comparison of the data to estimates from oceanic ozone deposition models. The way that the paper is written most benefits readers who are very familiar with oceanic ozone deposition measurements and models. I would urge the authors to make changes in order to expand the readership. One way of doing so would be to better characterize what they are doing (and why) before the results of a given analysis are presented. There are also a lot of figures and information to take in – is this necessary?

We have made efforts to broaden the article for the audience. The one- and two-layer models are now described thoroughly in the introductory section, and the behaviour of these models with respect to wind speed / friction velocity is mentioned as context for the discussion later in the paper. A description of the resistance model for deposition velocities is also included for some background. While we feel the figures presented are informative, some are less dependent on being presented alongside the text. As such what were Figures 3, 5, and 10 have been moved into a supplementary information document to accompany the paper.

My only major concern has to do with the footprint analysis, a large component of the paper. The footprint model used is for flat homogeneous terrain rather than heterogenous coastal site. I understand that a footprint model for the given land type may not be available, but I think the authors should explain more, with references, how a footprint model for a flat homogeneous terrain may or may not capture the footprint of a heterogeneous coastal site.

We have expanded the discussion of footprint limitations in line 331 as follows:

It is worth reiterating that the Kljun footprint model is designed for use in homogenous environments, which is not the case for our site. Furthermore, the double rotation applied to the wind data will result in varying pitch angles relative to the water surface, introducing a dependence of the footprint extent on this pitch angle. These limitations may be important for work relying on direct interpretations of the flux footprint, such as comparisons to emissions inventories (Squires et al., 2020; Vaughan et al., 2017). In contrast to an inventory comparison, we only use the flux footprint model to develop a strategy for robust data selection, and generate an aggregate footprint from several individual footprints. This approach follows the works of Amiro (1998), Göckede et al. (2006, 2008); Kirby et al. (2008), Metzger (2018) and Xu et al. (2018) who have demonstrated the utility of aggregation for deriving robust footprint-based metrics in heterogeneous environments.

Abstract:

Does the percentage of the flux footprint being water change with tide, or the size of footprint?

The percentage of the footprint over land will vary as the tide goes in and out, though it's true the major effects are from changes in wind and stability. We have qualified in the abstract that footprint size also has an effect.

Readers may not know the Fairall model well. Can the authors add some short description of this model to the abstract instead of, or in addition to, referring to the reference?

A brief description of the Fairall model has been included in the abstract

Can the authors clarify whether they are talking about fluxes or deposition velocities when they refer to 'deposition'? (this applies throughout the paper and the figures; I tend to think that 'deposition' refers to the flux')

'Deposition' has been specified to 'deposition velocity' or flux as appropriate throughout the document.

Line 26 – I think this a rather strong statement; only one paper suggests this

Altered the statement to convey that 25% is just an estimate, and the true value may be lower.

Line 31– briefly describe what is meant by 'atmospheric and surface resistance values'

Definitions for both atmospheric and surface resistance have been added in parentheses.

Line 31 – rephrase so as not to imply that we can't learn anything from these lab and box enclosure methods

This sentence has been reworded to properly convey the value of box-enclosure experiments in determining surface resistance values

Line 36 – references for this range of values? are the citations given in the previous sentences just for seawater?

All eddy covariance measurements referenced were over saltwater. We have qualified that the range of values given corresponds specifically to the eddy covariance measurements referenced in the prior sentence.

Line 52 – clarify the aspect of the depositional sink that needs to be better characterized, in line with the discussion in the previous paragraph; also, is it really a 'tropospheric ozone cycle'?

We have specified that it is the effects of wind speed and of the composition of the sea surface that are in need of better characterisation. The term 'cycle' was not appropriate – this has been changed to 'budget'.

Line 55 – not sure this is the right usage of the term 'natural variability'

Removed 'natural' instead encompassing more generally 'factors' that could affect uncertainty in the measurements

Line 74 – can the authors describe more clearly in the text what Figure 2 shows and what the author wants the reader to do in referring to all the parts

The 'parts' had been intended to aid convenience when referring to Figure 2. However, we realise this may be redundant given the presence of these labels within the Figure. These 'part' labels have been removed from the body of text, but left in Figure 2.

Line 105 – check sentence

Sentence completed – corrections not needed 'for determining an accurate ozone mixing ratio'.

Line 112 – what is 'dry ozone'?

Reworded to convey 'in the absence of water vapour'

Line 130 – there is a negative sign missing

Minus sign added, and rearranged for Flux on the left hand side.

Line 148-149 – what is 'contrary'? are the authors implying that the dependences of Chang and Helmig are incorrect?

The opposite is intended – rather the dependencies of Chang and Helmig cause us to be wary of our data at low wind speeds. The sentence has been reworded, qualifying that we observe an 'apparent increase' in deposition velocity at low wind speed, most likely from land interference rather than genuinely higher deposition over the water.

Line 150 – new paragraph starting at "Footprint analysis"

Separated into a new paragraph as requested

Line 159 – where is this estimate of roughness length from? is it appropriate for the location?

Eq. (12) is a rearrangement of the logarithmic wind profile equation to solve for roughness length (this has been added to the text for clarity). Due to the lack of roughness elements over the sea, the displacement height term has been omitted.

Line 170 –removal 'of'

'off' corrected to 'of'

Line 171 – clarify this sentence; what is the object of "contributing"?

Clarified 'contributing to the elevated surface <u>roughness</u> values'

Line 171-4 – can the authors clarify what they are doing here? are they further filtering their data based on the roughness lengths or not? if not, is the justification only that they don't want excessive data removal?

We confirm that roughness length has <u>not</u> been used as a filtering parameter – we merely note it as a potential alternative. Figure 5 shows that a roughness length filter of approximately $z_0 < 0.1m$ would only really exclude points already removed by the wind speed filter. Lowering that threshold would begin to remove several points across the full range of wind speeds, and we wish to retain as many points as possible for when data are later separated further into wind speed bins.

Line 175-180 – but does it mean anything for the authors' conclusions with regards to wind speed or friction velocity dependencies?

This criterion was in fact 0.10 m s^{-1}, not 0.15 m s^{-1} as stated. This has been corrected. A threshold of 0.15 m s^{-1} only makes a tiny difference too though: 0.001 cm s^{-1}.

This small difference is because the wind speed filter already applied removed the vast majority of very low u_* values, since they scale approximately linearly with each other over the ocean. The additional u_* filter made little difference (only removing 30 points), having little effect of the median given the number of data points clustered around that median. Giving values to 4 decimals would show the 0.10 m s^{-1} u_* filter would increase the median deposition from 0.00371 cm s^{-1} to 0.00373 cm s^{-1}, but a change of less than 1% cannot be considered significant here.

To avoid confusion, we have clarified that we are discussion a u_* filter in addition to the previously applied criteria, rather than as a substitute for wind speed.

Line 183 – what is being compared with the 20-min averaging?

60-minute averaging – added for clarity

Line 186 – "Flux and deposition velocity values"

'Deposition' added

Line 189-191 – say what this finding means

The Kolmogorov-Smirnov test shows that our distribution of values and the distribution where wind and ozone data are disjoined could not have been picked at random from the same distribution of values. Therefore, we are observing a flux that is statistically distinct from the noise in the measurements. A sentence has been added to clarify this.

Line 193-4 –say what this finding means

Clarified that the average flux value obtained was above the 2σ LoD, but with considerable uncertainty associated with it.

Line 220 – is there a reference for this equation?

This equation comes from the assumption that atmospheric surface stress and waterside surface stress are equal. This is the assumption made by Luhar et al. (2017), and the reference has been added.

Line 229 – is an assumption of constant Ts and [I-] fair? what's the 'relevant' time periods?

'relevant period' changed to 'April-May' to clarify that we are using the model values that span the period of the observations. SST and I are certain to vary, and to affect reactivity – the iodide especially will be the subject of future work where we quantitatively measure iodide (and other species) concentrations in the microlayer within the footprint area. A detailed set of these data were not yet available for this publication, so the data sources used by Luhar et al. have been used here as well for consistency.

Line 232 & 234 – what are the confidence intervals for m + b?

Standard errors have been added for the gradient and intercept of the linear fit of our deposition velocities against u_*. Similar values for the Fairall model are not readily available – they are not quoted when these values are given in the work of Helmig et al. (2012), and assessing the uncertainties in the original model are beyond the scope of this work.

Line 237 – in terms of 'remarkable' I recommend the authors remain objective

Language changed to be more objective.

Line 239 – why consider only iodide reactivity? and I'm not actually sure what this means – I thought the authors were fixing [I-]. Does this mean that the authors are only considering the temperature dependence of A? generally, it would help if the authors gave brief descriptions of the Fairall and Luhar models, otherwise the discussion is not very useful for readers who are not well versed in oceanic ozone dep models. the authors do this to some degree in the discussion, but it would be nice to have this information closer to the beginning of the article.

A sentence has been added to clarify that 'only iodide reactivity' is meant to convey that we are not attempting to add a quantitative reactivity term for organic material (as mentioned in section 4) or anything else. This means the model fits are for a single reactivity and temperature, and therefore a single A value, to examine wind speed / u_* dependence while other conditions are fixed at values typical for our site.
Regarding the models, a section introducing both models, their assumptions and dependencies has been added to the introduction to properly cover this ahead of the discussion.

Line 240 – while the Luhar model underpredicts vd, it doesn't seem like the variability in the Luhar model is necessarily off, or worse than Fairall. Can the authors provide quantitative metrics for how well these models fit the data?

Root mean square error (RMSE) and mean bias values have been added for both models. RMSE is high in both cases due to the large scatter in the data, but it is smaller for the Fairall parameterisation than for the 2-layer model. Mean bias is small for the Fairall parameterisation, and much larger for the two-layer model.

Line 244 – what is the object of amplifying?

Amplifying the potential influence 'of land deposition' - added for clarity

Line 245 – is this deposition velocity for grassland from the models used in Hardacre et al.? or some observations used in the Hardacre model evaluation? regardless, the authors need to clarify and discuss the high uncertainty in using this value, and use references for the observations at grasslands if they are using the observations. Generally, I'm not sure what we are learning from the analysis with the Hardacre grassland value.

The deposition velocity estimate is taken from the medians of the two datasets analysed by Hardacre et al. (2015), specifically Figures 4c and 4d. The inclusion of this quick calculation is intended to serve as a demonstration of how much influence the land could have on a coastal measurement if land exists within the flux footprint. This is then followed with an attempt at determining a more realistic value for our site given the land is not true 'grassland'. We have also qualified that the value we use is a median of the accumulated datasets used by Hardacre et al. (2015).

Line 251 – confidence intervals for the land and sea values?

Standard errors for the regression of land % with deposition added

Line 259 – I don't follow why ozone fluxes would be compared to emission inventories

Reference is made to inventories to highlight the kind of studies where more precise footprint areas are essential for lining up with sources. This is included to contrast with our work where the footprint is used more for quality control rather than trying to match up to specific sources and sinks.

Line 260 – in contrast to what? what do the authors mean by 'aggregates'?

Clarified 'In contrast to an inventory comparison, we only use the flux footprint model to develop a strategy for robust data selection, and generate an aggregate footprint from several individual footprints.'

Line 263 – why is this example 'extreme'? perhaps best to remain objective

We would maintain that this example is extreme, but realise we neglected to explain why – the tidal zone was very shallow in the work of Whitehead et al. (2009), meaning that at low tide the flux footprint was almost entirely over ~3 km of exposed seabed rather than water. These details have been added to justify the 'extreme'.

Line 265 – what could this mean in terms of the results? generally it might be better to have all the info about the tides in one paragraph, not two, with some of the info tacked on the end of a very long paragraph

The estuarine input to the coastal waters could change the chemical composition of the surface water, and thus its reactivity to ozone. Chemical analysis of the surface water does not form a part of this manuscript however, and will be a focus of our future work. A sentence has been added to clarify this. Additionally, 'Tidal influence' has been separated into its own subchapter, with both paragraphs merged within to distinguish it clearly from the prior section.

Line 266 – measurement height was adjusted how/where?

Clarified that we are referring to the measurement height used in flux and footprint calculations. The physical tower height was not changed, but the height of the tower above the water varied with tide, and this was the 'adjustment' made to the mean height above sea level to properly account for this change in footprints etc.

Line 270 – where the authors expecting to see a diurnal cycle? would be helpful if authors set the stage for describing this analysis more

Added that a diurnal was not expected – we merely provide the information given its presence in the discussion of Gallagher et al. (2001). The lack of a diurnal cycle also suggests land deposition to be minimal.

Line 273 – describe method of Langford briefly

The following sentence provides a brief description of the method. We feel a more in depth description would feel out of place here. The initial presentation of this and the theoretical flux uncertainty calculation methods are now introduced in section 2.5.

Line 285 – in what relationship?

Clarified that we refer to the relationship in equation 16.

Line 294 – similar to what literature? include references

Specified that we mean values for near-neutral conditions – references of Blomquist et al (2010) and Lenschow and Kristensen (1985) provided.

Line 295 – meaning that the authors do not use equation12 to calculate the integral timescale?

Equations (now updated) 17 and 18 are used with the peak of the co-spectrum in Figure 11 to determine our b value. This b value is then used in equation 18 with the wind and height data for each 20-minute period to estimate an integral timescale for each period.

Added that equations 17 and 18 are used with Figure 11 to determine the b value for clarity.

Line 299 – repeat empirical value here

Value added

Line 301– what do the authors mean 'they defined twice'?

Bad wording – clarified to '…Lenschow & Kristensen (1985) who multiplied the right-hand side of Eq. (16) by 2 to derive…'.

Line 302 – clarify here that talking about variability within the averaging interval

'within averaging intervals' added as requested.

Line 303-4 – this sentence confuses me. random instrument noise in the ozone measurement or the wind measurement?

This was poorly phrased – it has been changed to specify that noise is the ozone instrument is likely a large part of ozone variance.

Line 319 – say what the results with respect to block averaging vs. linear detrending means

Added a clause explaining that the use of linear detrending is not leading to large low-frequency information loss.

Line 324 – give the percentage for random uncertainty here

Percentage (85%) added, as well as clarification that this is a 2σ uncertainty.

Line 329 – does this choice of reaction-diffusion sublayer length have an impact on results? where is this estimate from?

It was erroneously stated in the original text that a fixed sublayer depth was used in this model estimate. That approach was investigated, but it is the <u>variable length</u>, parameterised from diffusivity and reactivity (the equation for which is now given as Eq. (11)) in the amended introduction) that was ultimately used. The script has been corrected to reflect this. For interest, the use of a fixed 3 µm layer rather than the variable layer (which works out as 4.2 µm) leads to a model estimate of 0.018 cm s^{-1}, up from 0.016 cm s^{-1}. This is a relatively small change in depth, given the range of 1.2 – 24 µm for waters varying 2-33 °C in temperature.

Line 333-4 – cut 'significantly'

'Significantly' removed

Line 353-5 – I'm confused by these sentences; rephrase

Reworded to reflect that the two-layer model gives values more similar in magnitude to our observations, but gives a wind speed dependence fundamentally different from some observed data.

Line 360 – why just discuss Helmig values here?

Comparison to previous values determined from tower-based eddy covariance measurements added (McVeigh, Whitehead).

Line 376 – give numbers here for instrument noise uncertainty

Limit of detection (0.113 mg m^{-2} h^{-1}) and noise level contribution to ozone variation (45-98%) added.

Line 378-9 – clarify what the authors mean by larger (longer or additional measurements or both?)

We intended for both – changed to reflect this – 'A longer dataset with more chemical composition variables'

Table 1 – say whether the data in the nth row is filtered by the criteria in the previous n-1 rows

Altered the table as per Reviewer 3's suggestion for each row to show only that filter from the total, with a row at the bottom showing the application of all values.

Figure 4 – it's so helpful here that the authors point out what the reader should be "getting" from this figure – can the authors do this for other figures?

Figure 1, 3, 7, 10, 12, 13, and 14 (as they appeared in initial submission) captions amended to clarify the 'take-away' message from the figure. As mentioned, Figures 3 and 10 are now moved to SI, now Figures S1 and S3 respectively.

Figure 5 – say what 'DoY' is

Plot x axis label changed to 'Day of Year 2018' (and moved to SI, Figure S2).

Figure 9 – instead of saying "points omitted" (which to me implies that the authors do not include the data in the averages), can the authors say something like "points outside the y axis range"?

'omitted' changed to 'beyond these y axis bounds'. Note, renumbered to Figure 7

Figure 12 – I don't know what I'm supposed to be looking at here/what this figure is telling me

The footprint plot is included to give an idea of the spatial area being observed over the course of these measurements. We realise the previous caption was unhelpful for anyone not familiar with contoured footprint plots, and as such has been updated to describe the bounds represented. Note, renumbered to Figure 9

Figure 14 – 'kaimal prediction' is not very clear

Changed to 'Expected cospectral shape predicted by Kaimal…' to better explain its use as an 'expected' reference point. Note, renumbered to Figure 11