

Ocean Sci. Discuss., referee comment RC1  
<https://doi.org/10.5194/os-2021-83-RC1>, 2021  
© Author(s) 2021. This work is distributed under  
the Creative Commons Attribution 4.0 License.

## **Comment on os-2021-83**

Anonymous Referee #1

---

Referee comment on "Using machine learning and beach cleanup data to explain litter quantities along the Dutch North Sea coast" by Mikael L. A. Kaandorp et al., Ocean Sci. Discuss., <https://doi.org/10.5194/os-2021-83-RC1>, 2021

---

Review of 'Using machine learning and beach cleanup data to explain litter quantities along the Dutch North Sea coast' by Kaandorp et al.

### Summary

The manuscript uses machine learning techniques to formulate a regression model to investigate the influence of a range of hydrodynamic, atmospheric and coastline geometry factors obtained from publically available data and re-analysis products, as well as particle tracking model runs based on publically available hydrodynamics model output, on the density distribution of (plastic) beach litter on the Dutch coast. The regression model is 'trained' using several years of data from beach cleaning surveys. The main explanatory factors identified are negative correlations with tidal elevation variability and maximum tidal amplitude, and a correlation with coastline orientation relative to residual currents. The particle tracking model runs did not contribute significantly to the predictive capability of the regression model. A trimmed-down version of the regression model, using only the most important contributing factors, was used to generate a 5-year average litter concentration distribution along the Dutch coast.

### General remarks

This is a very well-written manuscript, that I have read with interest. It mobilises a variety of different data, and uses sophisticated statistical techniques to analyse them together. It addresses a clear need to improve understanding and predictive capability of beaching plastics in the marine environment. Having said that, the manuscript is not completely convincing, and leaves me wondering about a few things:

1. By which mechanism do tidal variability and amplitude control the beach litter distribution? Do higher variability and amplitude promote removal of previously beached material and in that way define a lower equilibrium concentration? Or does it simply spread the litter more in the cross-shore direction making it harder to detect in the beach surveys? How might this be illustrated? Or is the preference of the regression model for tidal variability and amplitude an artifact of the method (see also next points)?

2. A regression model cannot be used to prove causality. Indeed, the authors acknowledge that many of the factors investigated are not independent, as well as admitting that there are many regression factors compared with the number of available data points. I think that the authors can do more to make their case:

a) they can 'cripple' the regression model by removing (some of) the most important factors (rather than those that do not matter much) and show that the resulting model has significantly reduced predictive capability (as opposed to shifting predictive power to a different factor);

b) they state that the particle tracking model does not add to the predictive capability as it includes most of the factors from others sources incorporated in the regression model. So they can treat the particle tracking model as an independent (causal!) experimentation tool, and validate its results against the field data as well as cross-comparing with the results of the regression model, and subsequently disable the most important processes suggested by the regression model to directly illustrate their influence on the results.

3. Have all potential causal factors been included in the analysis? The particle tracking results suggested that rivers may be the dominant source of beach litter. One factor I'm missing (and that may be as important or more important than tidal variability) is the proximity of the nearest riverine source (in the upstream direction). Should this not be added as a regression factor? It's possible that in the current setup the influence of a factor like this is attributed to the tidal factors which have a spatial gradient in the same direction.

4. Is the regression model as good as it seems? I would like to see direct comparisons of predicted annual distributions with the observed annual spatial distributions in Appendix A. I'm aware that these data were used for training so it's not a validation, but it would illustrate to what extent the regression model captures the temporal and spatial variations, which are not represented in Figure 4.

I would like to challenge the authors to address these points, which would significantly improve the manuscript, and hence recommend major revisions.

#### Specific comments

l 10-12: recommendations for removal. You currently don't really give any, please remove this sentence from the abstract or work out clear recommendations and state them specifically here.

l 77 Some sites were sampled multiple times per year. Please describe how these were treated in the analysis. Would this bias the results and how?

Table 1. daily mean currents. This would alias in a tidal component. What would be the influence on the results? Also, are the bathymetries used to produce the various products used consistent?

Figure 2 / rivers: why has only a subset of rivers been included? What has driven the decision to include certain rivers but not others? Please include a list of rivers used in an appendix?

equation 1. In reality, I would expect  $\tau_{\text{beach}}=F(x,y,z,t)$ ? Please discuss why a constant was chosen?

Figure 6 and associated text: it would be helpful to have an indication what the principle components may dominantly represent if that's at all possible?

l 67 month: which month of the year?

l 82 August: why one month? Why this month?

Technical corrections

I 5. what: which

I 6 remove might

I 6 variability: of what? types/size/volume/??

I 10 what: which

I 15 increase: release?

I 15 need of: need for

I 21 other sinks: such as?

I 23 ...plastic items by removal, ...

I 25 ...reduction of new plastic waste...

I 75 weighing devices: scales

I 80 mean currents: surface? depth-averaged?

Figure 1. Please include a scale vector for the currents.

Figure 1 caption. Please include references to the data sources. Add a period to the end of the caption.

| 107 there's a space between models and the period ending the sentence.

| 108 the daily-mean ocean-surface currents

| 109-111 include references to the data sources?

| 134 radius of 50km: from where?

Many of the figures: please re-consider the choices of colours to facilitate colour-blind people.

Figure 2: please include axes.

| 187 what: which

| 188: distance: between what?

| 345 less: fewer

| 403 insights into which processes may be causing

| 404 and which length scales should be