Response to Anonymous Referee #1 comments of Manuscript ID WES-2022-55 entitled “Introducing a data-driven approach to predict site-specific leading edge erosion”

Thank you for taking the time to review our article. We have addressed your comments attentively, for which the details are provided below.

1. Define NEA and DKA domains more approppriately in the text.

   An additional description has been added, which reads as:

   “The DKA domain consists of an 800x600 model grid with 2.5 km grid spacing. Each model column in the grid has 65 vertical levels with hybrid coordinates that follow the terrain near the surface and fixed pressure levels at the top of model grid. The lowest model level is approximately 12 m above the surface and the highest level is at 10 hPa. The grid has a Lambert conformal conic projection with 25.0°E and 56.7°N as the respective reference longitude and latitude, and 8.2°E and 56.7°N as the central longitude and latitude. The NEA domain has a 1280x1080 model grid. It has the same grid spacing and vertical levels as the DKA domain. NEA has a Lambert conformal conic projection with 25.0°E and 60.0°N as the respective reference longitude and latitude, and 7.0°E and 60.0°N as the central longitude and latitude.”

2. It is unclear the reason for the α=0.16 adopted by the power law extrapolation. This extrapolation procedure should be further explained and, furthermore, the sensitivity of the measured quantities with respect to α should be investigated.

   First of all, the value used in the study was 0.143 and not 0.16 as stated in the manuscript. This has been corrected. Secondly, to answer your comment, as mentioned in Section 2.1, the mesoscale weather data was provided at different model level heights. For each site, the model level height closest to that of the corresponding hub height was used. On average, the absolute extrapolated distance (i.e., between model height and hub height) was 5.41 m and the maximum distance was 11 m. A sensitivity study was carried out and it was found that changes in the shear exponent had very small impact on the accumulated impingement which is heavily dominated by the rain rate. For shear exponents in the range of 0.143 ± 0.5, the maximum change in accumulated impingement was 1.7%. Since this sensitivity was performed for consistent shear exponents, the specific value is considered negligible for our study. It should also be mentioned that for the application of the model, the user can easily extrapolate using different shear exponents or even other extrapolation models. This information is now included in Section 2.1 to make it clearer for the reader/user.

3. The main input to the data-driven model is the accumulated rain impingement. Despite it does combine the effect of amount of rain and wind speed, it does not consider a third important parameter which is the rotational speed of the wind turbine. In the paragraph related to line 260, please expand the discussion while taking this argument into account. What would be the implication of including the rotation speed into the prediction models?

   The accumulated rain impingement is actually calculated based on the rotational speed, ω, as indicated explicitly in Eq. 3. The rotational speed directly affects the tip speed which is proportional to the rain impingement.

4. The authors discuss that the feedforward neural network with a 2-layer architecture with 5 neurons per layer + RELUs was enough and outperformed other methods such as PCE, support vector machine. The ability to learn non-linear relationships is not exclusive from FFNN, and the RELU effect to allow a linear piece-wise description of the neuron weights can be emulated by another model based on moving averages or any piecewise form of approximation, for instance. The authors are encouraged to improve all the discussion surrounding the choice for the FFNN, exposing also its main limitations and the results of the PCE that was largely investigated.

   The authors fully agree that the ability for FFNN to learn non-linear relations is not exclusive. The choice of model was based on the validation criteria described in the manuscript. A comparison of the error statistics is shown below: The section describing the model selection has been updated to include a better comparison and reads as:

   “It was chosen to use a simple feedforward neural network (FFNN) as the weak learners for the ensemble model. Several models were evaluated as candidates by means of the validation described previously, i.e., statistical performance and physical interpretability. The neural network was found to be the best model when comparing error statistics. As an example, the root-mean-square-error was found to be 30.0 % lower than the support-vector-machine, 36.4 % lower than polynomial chaos expansion (PCE) of first

model	fit	R^2	MAE	RMSE	MAPE	FFNN's relative [%] comparison R^2	MAE	RMSE	MAPE
FFNN	0.91x+0.02	0,8890	0,05	0,07	11,27	-	-	-	-
Linear regression	0.84x+0.05	0,7395	0,08	0,11	17,32	20,2	-37,5	-36,4	-34,9
Decision tree	0.89x+0.01	0,7347	0,08	0,12	17,17	21,0	-37,5	-41,7	-34,4
KNN	0.70x+0.03	0,6242	0,12	0,16	27,11	42,4	-58,3	-56,3	-58,4
PCE 1st order	0.85x+0.04	0,7409	0,08	0,11	17,50	20,0	-37,5	-36,4	-35,6
PCE 2nd order	0.89x+0.02	0,6961	0,09	0,13	19,33	27,7	-44,4	-46,2	-41,7
PCE 3rd order	0.86x+0.03	0,6659	0,10	0,14	23,13	33,5	-50,0	-50,0	-51,3
SVM	0.82x+0.05	0,8018	0,07	0,10	14,77	10,9	-28,6	-30,0	-23,7

order (46.2 % for second order) and 41.7 % lower than decision trees. For other error metrics the same behavior was observed. In addition, the neural network was also found best suited for capturing physical characteristics such as the incubation period. Specifically, it can be mentioned that PCE up to third order was evaluated in an exhaustive manner, i.e., by evaluating all the polynomial expansions repetitively. It was found that a first-order model performed best whereas higher-order models failed to represent the physicality requirements specified earlier.

When using ensemble techniques such as bagging, it is desired to use a high-variance model which is able to learn non-linear relations. This is because the variance is reduced through bagging and the bias is increased (or maintained). This trade is not unique for neural networks and other models such as those based on decision trees or expansion models can have the same ability."

5. Have the authors considered the use of classical surrogate models such as the Kriging method?

   Specifically, the Kriging method has not been considered as it only allows for interpolation within the convex hull whereas other regression techniques allow for extrapolation also. This inability would also cause errors in the ensemble training process as we here require extrapolation in some of the data splits when validating the against the test data. Simpler models, such as linear regression were also considered but outperformed by the FFNN. This information will be clarified in the description of the model selection and reads as:

   "Finally, it can also be mentioned that the ensemble training process used in this study, requires extrapolation in some of the data splits when validating against the test data, and though extrapolation is never recommended for ML applications, the FFNN has ability to do it anyway and performed better than SVM or models based on decision trees. It is the inherent definition of decision trees that makes them unsuited for extrapolation, e.g., the output of a decision tree is limited by the leaf nodes and can therefore never exceed the outer leaf nodes."

6. The authors should present one application case that applies the developed framework to predict damage in a new site within the covered region, focusing mainly on the workflow that would need to be followed. As the caption in Figure 3 describes, the trained model is evaluated on 99 new sites across the region of interest. This is our example of an application of the developed framework, in this case shown as the erosion maps in Figure 16. There we show simulated cases for the sites that were not used during training for different sequence lengths and initial damages. Ultimately, the trained erosion model is the backbone of the framework but it can have several different application depending on the inputs of the user.