

PACIFIC GAS AND ELECTRIC COMPANY
Wildfire Mitigation Plans
Rulemaking 18-10-007
Data Response

PG&E Data Request No.:	CalAdvocates_041-Q03		
PG&E File Name:	WildfireMitigationPlans_DR_CalAdvocates_041-Q03		
Request Date:	February 19, 2021	Requester DR No.:	CalAdvocates-PGE-2021WMP-07
Date Sent:	February 24, 2021	Requesting Party:	Public Advocates Office
PG&E Witness:		Requester:	Alan Wehrman

SUBJECT: RISK MODELING

The following questions related to PG&E's 2021 Wildfire Mitigation Plan (WMP) Update.

QUESTION 03

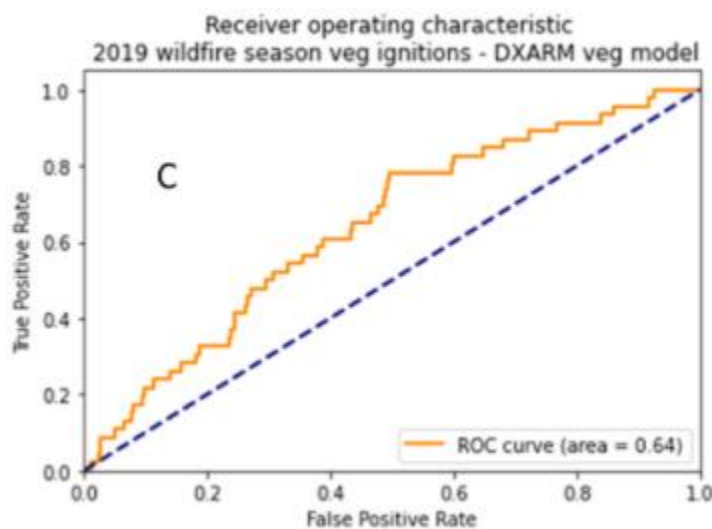
Pages 135-136 of PG&E's 2021 WMP state that the Vegetation Risk Model was modeled based on ignition data from 2015 to 2018, and tested on ignition data from 2019.

- a. How accurate was the Vegetation Risk Model during the tests using 2019 ignition data? Please provide quantifiable accuracy metrics, such as an R-squared value.
- b. How many ignitions from 2019 were used to test the Vegetation Risk Model?
- c. How many of the ignitions in part (b) were not accurately predicted by the Vegetation Risk Model?
- d. What changes, if any, did PG&E make following testing using 2019 ignition data, to improve the accuracy of the Vegetation Risk Model?
- e. Has PG&E tested the Vegetation Risk Model with 2020 ignition data? If so, please respond to parts (a) through (d) with respect to the 2020 ignition data.

ANSWER 03

- a) A Receiver-Operator Curve (ROC) has been used to measure the predictive performance of the Vegetation Risk Model. The interpretation of the ROC curves can be thought of in terms of the fraction of non-ignition locations you would need to harden (x-axis) to ensure that you harden some fraction of ignition locations (y-axis). The steeper the curve, the lower the overhead of work done that doesn't avoid ignitions. The dataset of 2015 – 2018 vegetation ignitions used to train the model achieved an Area Under the Curve (AUC) score of 0.73. The 2019 dataset was used as an out-of-sample test dataset to evaluate the model fit and achieved a score of 0.64 but a randomly withheld test sample from several years achieved a score of 0.72. The minimal reduction in AUC score between the training and testing datasets gives confidence that the model is not overfitting to the training dataset but also raises the possibility that the spatial pattern and other characteristics of 2019 vegetation-caused ignitions deviated slightly from 2015-2018. Based on this ROC result, workplans that follow the prioritization of can expect to address 50% of

vegetation ignition locations while addressing around 30% of non-ignition locations.



- b) The 2019 test dataset included 53 ignitions that were related to vegetation related events.
- c) We are predicting the probability of an ignition, not a binary classification of whether an ignition will occur. The ROC curves show the fraction of all ignitions above the threshold for all possible thresholds (y-axis) compared to the fraction of non-ignition locations also above each threshold (x-axis).
- d) The purpose of the out-of-sample testing with 2019 data was to assess the performance of the model as though 2019 had not happened yet – no information from 2019 informed the model that predicted for 2019. We therefore explicitly did not adjust the model in any way to improve the 2019 predictions. The testing dataset was used to ensure the model was not overfitting to the training dataset.
- e) A ROC curve for 2020 vegetation ignitions has not been performed yet. The data engineering to integrate the rigorously validated 2020 ignition dataset will not be completed to perform the analysis until the end of Q1 2021.