Wildfire Risk Assessment Prioritization Calculations for Distribution and Transmission

Pacific Gas and Electric Company®

Risk Management

Wildfire Risk Assessment Prioritization Calculations for Distribution and Transmission

Table of Contents

PGE-DIXIE-NDCAL-000004257

PGE-DIXIE-NDCAL-000004259

1 Overview

1.1 Document Usage

This document is to be utilized as a user guide for use with Wildfire Risk Assessment Prioritization Output_Consolidated_190223_vF.xlsxStarting with Section 2, each section is a standalone description of the methodology in the associated tab within the excel output file. For example, Section 2 for the Risk Prioritization Methodology relates to the associated tab within file Wildfire Risk Assessment Prioritization Output_Consolidated_190223_vF.xlsx in the tab labeled “Dx-EVM”. The tab details are also described within the associated procedure sections.

Each section also describes each of the tabs, columns and associated calculations. Section 1.2 of this document describes the high-level prioritization approach, and Section 1.3 includes Definitions.

1.2 Prioritization Overview

This document describes how wildfire risk prioritization was calculated for activities planned on distribution and transmission circuits as shown in the output file Circuit Prioritization Output_Consolidated_190223_vF.xlsx. In the effort to prioritize circuits for wildfire mitigation activities, an analysis was performed to develop a risk based prioritization for:

1) Enhanced vegetation management (EVM) on distribution circuits,
2) System hardening (SH) on distribution circuits,
3) Accelerated inspection and repair work on distribution circuits
4) Accelerated inspection and repair work on transmission circuits

For each prioritization, an analysis was performed to determine a wildfire risk score for each circuit based upon different components of risk. Wildfire Risk is calculated by three components: likelihood of failure, likelihood of spread and consequence, and egress. These three components can be defined as follows:

1) *Likelihood of failure*: Relative risk of a circuit causing an outage / ignition
2) *Likelihood of wildfire spread and consequence score*: Relative ability for an ignition to spread and quantity of homes or timber affected if an ignition were to occur
3) *Egress score*: Potential ease of accessing or exiting a community in case of a mass evacuation during a catastrophic event

Availability of data at the time of model development and an evolving perspective of key wildfire risk variables resulted in some variations among the models. The methodology, risk components, data used, and outputs for each model are explained further in detail throughout this document.

1.3 Definitions

a. **Accelerated Inspection** refers to the increased frequency of regulated inspection outside of typical inspection timeframes
b. **Egress** refers to the potential ease of accessing or exiting a community
c. **Enhanced Vegetation Management (EVM)** refers to increased vegetation management program as a result
d. **Likelihood** refers to the relative risk in a given timeframe NOTE: For likelihood of wildfire spread and consequence score, this includes the quantity of homes or timber impacted should an ignition even occur

e. **Risk Tiers** refers to the relative rank of each risk component score for the circuit to the total set of scores in the prioritization model. A risk component or overall risk prioritization score can receive a risk tier score of "High", "Medium", or "Low".

f. **HFTD** refers High Fire Threat Districts (HFTD) established by the CPUC in Fire Map 2. These districts are Tier 3, Tier 1, and Zone 1.

g. **System Hardening (SH)** refers to the protection of distribution systems from vegetation damage via improved distribution network design

2 Risk Prioritization Methodology for EVM

The risk-based prioritization of distribution circuits for Enhanced Vegetation Mitigation (EVM) work in 2019 was based upon a model comprised of three components: 1) likelihood of failure, 2) likelihood of wildfire spread and consequence score, and 3) egress score. The likelihood of ignition was calculated using a regression methodology, the likelihood of wildfire spread and consequence was based upon model outputs from a third-party fire model, and the egress factor was calculated through understanding the at risk population's ability to exit an area via the town's current road infrastructure. Multiplying these three components resulted in a wildfire prioritization score for each circuit.

The output for this prioritization model is contained in Circuit Prioritization Output_vConsolidated_190223_vF.xlsx, tab 'Dx-EVM.' The following sections provide further detail on the data shown in this tab.

2.1 Circuit Information (Columns C-E)

This section provides the circuit name (feeder name), district, and number of Tier 2 and Tier 3 miles for each circuit included in the analysis.

NOTE: A primary step for this analysis was mapping all of PG&E's tree records (over 3 million trees) to each circuit for PG&E. This was conducted by geo-locating each tree to the nearest circuit. If a circuit located in Tier 2 or Tier 3 did not have any trees mapped to its location, it was not considered for this analysis. This step resulted in approximately 100 circuits (~2,000 circuit miles) not being accepted for further analysis.

2.2 Likelihood of Failure Calculation (Columns F-K)

Likelihood of failure (Column K) was determined using a regression analysis based upon a set of independent variables (Columns F-I) and a dependent variable (Column J). The regression analysis, using a logistic regression, was used to determine what characteristics of the circuit (independent variables) were more likely to indicate the likelihood for a failure to occur (dependent variable).

2.1.1 Likelihood of Failure Calculation – Independent Variables (Columns F-I)

a. Independent variables are the factors that may indicate the occurrence of the dependent variable. For this analysis, several different factors were considered for the circuit level analysis. These factors are:

   i. **Tier 3 miles divided by HFTD miles** - Tier 3 miles divided by HFTD miles

ii. **HFTD miles divided by total OH miles** - HFTD miles divided by total OH miles
iii. Health Score – Conductor age (15%) + Size & wire type (20%) + Number of splices (20%) + High probability of corrosion and Copper wire (15%) + Potential for snowpack (10%) + i2t issues exist (15%) + % loading of conductor (5%)
iv. Environmental Score - Wildfire probability score based on various environmental factors
v. Age Score – Score based on conductor installation year
vi. Size Score – Score based on: wire type, size, description, category, and miles of line
vii. Splice Score – Score based on the number of splices on a particular phase
viii. Corrosion Score – Score based on the probability of corrosion + % of conductor being copper wire
ix. I^2T Score – Score based on whether i^2t issues exist
x. Loading Score – Score based on the % of loading on a conductor
xi. % Aluminum reinforced material - % of the HFTD miles that utilize aluminum reinforced materials
xii. % Aluminum material - % of the HFTD miles that utilize aluminum materials
xiii. % Copper material - % of the HFTD miles that utilize Copper materials
xiv. Conductor size – Average size of the conductor (1-9)
xv. Wind Score - Based on GIS climate layer added (PG&E climate hazards)
xvi. Company related and equipment failure outages - Company related and equipment failure sustained outages between 2015 and 2017
xvii. Vegetation related outages - Vegetation related sustained outages between 2015 and 2017
xviii. **Number of trees per mile** - Tagged trees // Tagged trees per mile
xix. Number of total tagged trees - Tagged trees // Tagged trees per mile
xx. Number of high risk tree species per mile - Tagged high risk trees per mile

b. Drivers listed from items *i* to *xv* were obtained from PG&E's STAR model. The STAR model is a tool that PG&E developed internally for asset risk scoring. Within this tool, data is characteristics for each conductor segment of a circuit are captured. Some of these characteristics (e.g. age, conductor size, etc.) are used to calculate certain scores such as "Health Score" in item *iii*. These factors were aggregated at the circuit level.

c. Drivers listed from items *xvi* to *xvii* were obtained from PG&E's ILIS outage database. For this analysis, the number of outages for each circuit were tabulated for each out the outage categories listed.

d. Drivers listed from items *xviii* to *xx* were obtained from PG&E's tree record database maintained by the vegetation management team. This database maintains records of all trees inspected or worked on by PG&E in its service territory.

e. All data was only for assets within High Fire Threat Districts (HFTD), Tier 2-3 and zone 1 from the CPUC Fire Map 2. All data used in this analysis was provided by PG&E.

f. The items in **bold** in the list above were found to be the most significant variables in predicting the dependent variable. As a result these are listed in Columns F-I. More detail is provided on these variables in section 2.1.3.

2.1.2 Likelihood of Failure Calculation – Dependent Variable (Column J)

a. The occurrence(s) of an ignition on each circuit or protection zone was selected as the dependent variable (the variable to predict). If a circuit had one or more ignitions over the time period selected, that circuit or protection zone received a value of 1 as the dependent variable. If no ignition was recorded for the circuit, then it received a value of 0 as the dependent variable.

b. Data used to determine the dependent variable was the PG&E CPUC reportable ignitions from July 2013 to April 2018 that occurred in the HFTD (Tier 2, Tier 3, and Zone 1) within PG&E's service territory. This expanded data set was used to maximize the available dependent variable data points.

2.1.3 Likelihood of Failure Calculation – Calculated Likelihood (Column K)

a. The set of independent variables and dependent variable for each circuit was entered into a regression model. In order to determine the regression type, multiple regressions were investigated to see which best represented the desired outcomes. The regressions investigated included logistic regression, random forest model, Poisson model, negative binomial model, zero inflated variable model, and linear regression. A logistic regression model was chosen for this methodology because 1) the dependent variable is binary and 2) and the output of a logistic regression is a predicted probability that a circuit or protection zone will be a value of 1, or in this case an ignition.

b. From the logistic regression model for EVM at the circuit level, the following independent variables were found the be most significant in predicting an ignition:

   i. Number of trees per mile (Column F)
   ii. Number of total tagged trees (Column G)
   iii. HFTD miles divided by total OH miles (Column H)
   iv. Tier 3 miles divided by HFTD miles (Column I)

c. The resulting likelihood percentage (Column K) output from the model for each circuit and protection zone is one of the three components in the risk-based prioritization score.

2.3 Likelihood of wildfire spread and consequence (Column L)

This score was developed by a third-party firm, Reax Engineering. This score was developed following a similar fire modeling methodology to one that influenced the development of the High-Fire Threat Districts included in CPUC's Fire Map 2, Data Source 4. Although this component is shown as one number, this value is derived two separate outputs 1) likelihood of wildfire spread and 2) wildfire consequence score.

2.3.1 Likelihood of wildfire spread

a. Reax Engineering developed a spread score by using such factors as:

   i. Fuel type and density (grass vs. brush)
   ii. Topography (slope and natural fire breaks)
   iii. Weather / Wind
   iv. Distance from fire station / air suppression bases (speed to suppression)

2.3.2 Wildfire Consequence score

a. Reax Engineering developed a consequence score by using such factors as:

   i. Density of population
   ii. Density of structures
   iii. Potential negative impact to natural resources

2.3.3 Combined Risk Score

a. An output from the Reax Engineering model was a combined likelihood of wildfire spread and consequence score, which was used for this risk calculation. These scores were calculated as output at 300m resolutions throughout the PG&E's service territory around PG&E assets.

2.3.4 Circuit score calculation

a. Given that circuits can span many miles and these spread scores can change across the span, the mode score (the score that occurs most often) along the circuit was chosen as the score for the total circuit shown in Column L.

b. For further reading on the calculations, methodology, and output of the Reax Engineering model, see Reax Engineering whitepaper, Data Source 3.

2.4 Egress Score (Column M)

Egress is focused on the potential ease of accessing or exiting a community in case of a mass evacuation during a wildfire. At each census location (town and incorporated community) in California, an assessment was performed that considered the population and the road density, while taking into consideration the type of roads.

2.4.1 Approach

a. Egress is focused on the potential ease of accessing or exiting a community in case of a mass evacuation during a wildfire and included factors such as:

   i. Population of towns and unincorporated communities
   ii. The road density for each community by road type
      i. Highways / Interstates
      ii. Country roads

iii. Residential roads

2.4.2 Data Details

a. The egress data includes all incorporated and census designated locations in California including roughly 1,500 locations. Egress data was integrated and overlayed with each other using GIS. The road segment lengths were calculated within GIS in miles for all 2 million road records.

2.4.3 Methodology

a. At each census location (town and incorporated community) in California, an assessment was performed that considered the population and the road density, while taking into consideration the type of roads.

1. Population and proximity to High Fire Threat Districts (HFTD)

i. To assess the population size for each location in proximity to a HFTD, the census population counts were used and superimposed with the HFTDs designated areas on CPUC's Fire Map 2.

2. Road Density

i. The analysis examined the total miles of road in each location using three categories of roads identified from the 2017 National Transportation Dataset.

ii. Highways

iii. Country Roads

iv. Residential Roads

2.4.4 Egress Calculation

a. The egress score for each location is calculated by dividing population from section 2.4.3.1 above and dividing by road density in section 2.4.3.2 above.

b. With these location egress scores, the five nearest locations to each conductor segment of a circuit are identified within a 100 mile radius. The egress score of each location is then divided by the distance of the location to the conductor segment. The average of these five distance-adjusted scores is then calculated to determine an egress score for the conductor segment.

c. A circuit egress score is calculated by taking the mile-weighted average egress scores across each conductor segment. As a final step, the log of this score was calculated. This score results in the third component of the wildfire risk prioritization score calculation.

d. For further information on the Egress score, see Egress Analysis working paper, Data Source 1

2.5 Wildfire risk prioritization score (Column N)

The final score for each circuit was calculated by multiplying the three components: likelihood of failure, the likelihood of wildfire spread and consequence score, and egress score. With this score, circuits were prioritized for EVM work with those with higher scores receiving a higher priority than those with a lower score. In some cases, PG&E made changes to the prioritization order of circuits based on operational factors (e.g. land and environment, safety, planned projects, geographic access, weather, government relations, and customer communications).

2.6 Risk Tiers (Columns O-Q)

The last columns in the worksheet indicate the relative rank of each risk component score for the circuit to the total set of scores. If one of the risk components (likelihood of failure, the likelihood of wildfire spread and consequence score, and egress score) has a score in the top third of total scores for all the circuits for that risk component, a value of “High” is shown. The middle third receives a value of “Medium”, and the bottom third receives a value of “Low.”

3 Risk Prioritization Methodology for System Hardening

The risk-based prioritization of distribution circuits for System Hardening (SH) work in 2019 was based upon a model comprised of three components: 1) likelihood of failure, 2) likelihood of wildfire spread and consequence score, and 3) egress score. The likelihood of ignition was calculated using a regression methodology, the likelihood of wildfire spread and consequence was based upon model outputs from a third-party fire model, and the egress factor was calculated through understanding the at risk population's ability to exit an area via the town's current road infrastructure. Multiplying these three components resulted in a wildfire prioritization score for each circuit.

For SH work is planned and conducted at a protection zone (segments between two protection devices) level. Accordingly, slight adjustments and considerations were made in the model to generate prioritization results at the protection zone for SH rather than at the circuit level as conducted for EVM.

The output for this prioritization model is contained in Wildfire Risk Assessment Prioritization Output_Consolidated_190223_vF.xlsx, tab 'Dx-SH.' The following sections provide further detail on the data shown in this tab.

3.1 Circuit Information (Columns C-F)

This section provides the protection zone name, circuit name (feeder name), district, and the number of Tier 2 and Tier 3 miles for each protection zone included in the analysis.

3.2 Likelihood of Failure Calculation (Columns G-N)

Likelihood of failure (Column N) was determined using a regression analysis based upon a set of independent variables (Columns G-L) and a dependent variable (Column M). The regression analysis, using a logistic regression, was used to determine what characteristics of the circuit (independent variables) were more likely to indicate the likelihood for a failure to occur (dependent variable).

3.2.1 Likelihood of Failure Calculation – Independent Variables (Columns G-L)

a. Independent variables are the factors that may indicate the occurrence of the dependent variable. For this analysis, several different factors were considered for the circuit level analysis. These factors are:

  i. **Tier 3 miles divided by HFTD miles** - Tier 3 miles divided by HFTD miles
  ii. **HFTD miles divided by total OH miles** - HFTD miles divided by total OH miles
  iii. Health Score – Conductor age (15%) + Size & wire type (20%) + Number of splices (20%) + High probability of corrosion and Copper wire (15%) + Potential for snowpack (10%) + i2t issues exist (15%) + % loading of conductor (5%)
  iv. **Environmental Score** - Wildfire probability score based on various environmental factors
  v. Age Score – Score based on conductor installation year
  vi. Size Score – Score based on: wire type, size, description, category, and miles of line

vii. Splice Score – Score based on the number of splices on a particular phase
viii. Corrosion Score – Score based on the probability of corrosion + % of conductor being copper wire
ix. I^2T Score – Score based on whether i^2t issues exist
x. Loading Score – Score based on the % of loading on a conductor
xi. % Aluminum reinforced material - % of the HFTD miles that utilize aluminum reinforced materials
xii. **% Aluminum material -** % of the HFTD miles that utilize aluminum materials
xiii. % Copper material - % of the HFTD miles that utilize Copper materials
xiv. Conductor size – Average size of the conductor (1-9)
xv. Wind Score - Based on GIS climate layer added (PG&E climate hazards)
xvi. **Company related and equipment failure outages** - Company related and equipment failure sustained outages between 2015 and 2017
xvii. **Vegetation related outages** - Vegetation related sustained outages between 2015 and 2017
xviii. Number of trees per mile - Tagged trees // Tagged trees per mile
xix. Number of total tagged trees - Tagged trees // Tagged trees per mile
xx. Number of high risk tree species per mile - Tagged high risk trees per mile

b. Drivers listed from items *i* to *xv* were obtained from PG&E's STAR model. The STAR model is a tool that PG&E developed internally for asset risk scoring. Within this tool, data is characteristics for each conductor segment of a circuit are captured. Some of these characteristics (e.g. age, conductor size, etc.) are used to calculate certain scores such as "Health Score" in item *iii*. These factors were aggregated at the protection zone level.

c. Drivers listed from items *xvi* to *xvii* were obtained from PG&E's ILIS outage database. For this analysis, the number of outages for each protection zone were tabulated for each out the outage categories listed.

d. All data was only for assets within High Fire Threat Districts (HFTD), Tier 2-3 and from the CPUC Fire Map 2. All data used in this analysis was provided by PG&E.

e. The items in **bold** in the list above were found to be the most significant variables in predicting the dependent variable. As a result these are listed in Columns G-K. More detail is provided on these variables in section 3.1.3.

3.2.2 Likelihood of Failure Calculation – Dependent Variable (Column M)

a. The occurrence(s) of an ignition on each circuit or protection zone was selected as the dependent variable (the variable to predict). If a protection zone had one or more ignitions over the time period selected, that protection zone received a value of 1 as the dependent variable. If no ignition was recorded for the protection zone, then it received a value of 0 as the dependent variable.

b. Data used to determine the dependent variable was the PG&E CPUC reportable ignitions from July 2013 to April 2018 that occurred in the HFTD (Tier 2, Tier 3, and Zone 1) within PG&E's service territory. This expanded data set was used to maximize the available dependent variable data points.

3.2.3 Likelihood of Failure Calculation – Calculated Likelihood (Column N)

a. The set of independent variables and dependent variable for each protection zone was entered into a regression model. In order to determine the regression type, multiple regressions were investigated to see which best represented the desired outcomes. The regressions investigated included logistic regression, random forest model, Poisson model, negative binomial model, zero inflated variable model, and linear regression. A logistic regression model was chosen for this methodology because 1) the dependent variable is binary and 2) and the output of a logistic regression is a predicted probability that a circuit or protection zone will be a value of 1, or in this case an ignition.

b. From the logistic regression model for SH at the protection zone level, the following independent variables were found the be most significant in predicting an ignition:

    i. Environmental Score (Column G)
    i. % Aluminum material (Column H)
    ii. Vegetation related outages (Column I)
    iii. Company related and equipment failure outages (Column J)
    iv. HFTD miles divided by total OH miles (Column K)
    v. Tier 3 miles divided by HFTD miles (Column L)

c. The resulting likelihood percentage (Column N) output from the model for each circuit and protection zone is one of the three components in the risk-based prioritization score.

3.3 Likelihood of wildfire spread and consequence (Column O)

This score was developed by a third-party firm, Reax Engineering. This score was developed following a similar fire modeling methodology to one that influenced the development of the High-Fire Threat Districts included in CPUC's Fire Map 2, Data Source 4. Although this component is shown as one number, this value is derived two separate outputs 1) likelihood of wildfire spread and 2) wildfire consequence score.

3.3.1 Likelihood of wildfire spread

Reax Engineering developed a spread score by using such factors as:

    i. Fuel type and density (grass vs. brush)
    ii. Topography (slope and natural fire breaks)
    iii. Weather / Wind
    iv. Distance from fire station / air suppression bases (speed to suppression)

3.3.2 Wildfire Consequence score

Reax Engineering developed a consequence score by using such factors as:

i. Density of population
ii. Density of structures
iii. Potential negative impact to natural resources

3.3.3 Combined Risk Score

An output from the Reax Engineering model was a combined likelihood of wildfire spread and consequence score, which was used for this risk calculation. These scores were calculated as output at 300m resolutions throughout the PG&E's service territory around PG&E assets.

3.3.4 Circuit score calculation

a. Given that protection zones can span many miles and these spread scores can change across the span, the mode score (the score that occurs most often) along the protection zone was chosen as the score for the total circuit shown in Column O.

b. For further reading on the calculations, methodology, and output of the Reax Engineering model, see Reax Engineering whitepaper, Data Source 3.

3.4 Egress Score (Column P)

Egress is focused on the potential ease of accessing or exiting a community in case of a mass evacuation during a wildfire. At each census location (town and incorporated community) in California, an assessment was performed that considered the population and the road density, while taking into consideration the type of roads.

3.4.1 Approach

Egress is focused on the potential ease of accessing or exiting a community in case of a mass evacuation during a wildfire and included factors such as:

a. Population of towns and unincorporated communities
b. The road density for each community by road type
    i. Highways / Interstates
    ii. Country roads
    iii. Residential roads

3.4.2 Data Details

The egress data includes all incorporated and census designated locations in California including roughly 1,500 locations. Egress data was integrated and overlayed with each other using GIS. The road segment lengths were calculated within GIS in miles for all 2 million road records.

3.4.3 Methodology

At each census location (town and incorporated community) in California, an assessment was performed that considered the population and the road density, while taking into consideration the type of roads.

1. **Population and proximity to High Fire Threat Districts (HFTD)**
    i. To assess the population size for each location in proximity to a HFTD, the census population counts were used and superimposed with the HFTDs designated areas on CPUC's Fire Map 2.

2. **Road Density**
    i. The analysis examined the total miles of road in each location using three categories of roads identified from the 2017 National Transportation Dataset.
    ii. Highways
    iii. Country Roads
    iv. Residential Roads

3.4.4 Egress Calculation

a. The egress score for each location is calculated by dividing population from section 2.4.3.1 above and dividing by road density in section 2.4.3.2 above.
b. With these location egress scores, the five nearest locations to each conductor segment of a circuit are identified within a 100 mile radius. The egress score of each location is then divided by the distance of the location to the conductor segment. The average of these five distance-adjusted scores is then calculated to determine an egress score for the conductor segment.
c. A protection zone egress score is calculated by taking the mile-weighted average egress scores across each conductor segment of the protection zone. As a final step, the log of this score was calculated. This score results in the third component of the wildfire risk prioritization score calculation.
d. For further information on the Egress score, see Egress Analysis working paper, Data Source 1

3.5 Wildfire risk prioritization score (Column Q)

The final score for each protection zone was calculated by multiplying the three components: likelihood of failure, the likelihood of wildfire spread and consequence score, and egress score. With this score, protection zones were prioritized for SH work with those with higher scores receiving a higher priority than those with a lower score. In some cases, PG&E made changes to the prioritization order of circuits based on operational factors (e.g. land and environment, safety, planned projects, geographic access, weather, government relations, and customer communications).

3.6 Risk Tiers (Columns R-T)

The last columns in the worksheet indicate the relative rank of each risk component score for the protection zone to the total set of scores. If one of the risk components (likelihood of failure, the likelihood of wildfire spread and consequence score, and egress score) has a score in the top third of total scores for all the protection zones for that risk component, a value of "High" is shown. The middle third receives a value of "Medium", and the bottom third receives a value of "Low."

4 Risk Prioritization Methodology for Distribution Enhanced and Accelerated Inspection Process

The risk-based prioritization of distribution circuits for Distribution Enhanced and Accelerated Inspection Process work in 2019 was based upon a model comprised of three components: 1) likelihood of failure, 2) likelihood of wildfire spread and consequence score, and 3) egress score. The likelihood of ignition was calculated using a regression methodology, the likelihood of wildfire spread and consequence was based upon model outputs from a third-party fire model, and the egress factor was calculated through understanding the at risk population's ability to exit an area via the town's current road infrastructure. Multiplying these three components resulted in a wildfire prioritization score for each circuit.

The output for this prioritization model is contained in Wildfire Risk Assessment Prioritization Output_Consolidated_190223_vF.xlsx, tab 'Dx-Inspection.' The following sections provide further detail on the data shown in this tab.

4.1 Circuit Information (Columns C-G)

This section provides the circuit name (feeder name), circuit ID, region name, district, and the number of Tier 2 and Tier 3 miles for each circuit included in the analysis.

4.2 Likelihood of Failure Calculation (Columns H-AM)

Likelihood of failure (Column AM) was determined using a regression analysis based upon a set of independent variables (Columns H-AK) and a dependent variable (Column AL). The regression analysis, using a random forest model, was used to determine what characteristics of the circuit (independent variables) were more likely to indicate the likelihood for a failure to occur (dependent variable).

4.2.1 Likelihood of Failure Calculation – Independent Variables (Columns H-AK)

a. Independent variables are the factors that may indicate the occurrence of the dependent variable. For this analysis, over 1,000 independent variables were analyzed to determine which set of variables best predicted the dependent variable. This set consisted of the variables from the following categories:

   i. Conductor characteristics
   ii. Transformer characteristics
   iii. Fuse characteristics
   iv. Protective characteristics
   v. Switch characteristics
   vi. Pole characteristics
   vii. Capacitor bank characteristics
   viii. Notification data

b. The preceding data was gathered at the asset level. For this analysis, this data was aggregated to the circuit level to determine counts of assets, counts of asset characteristics, percentage of assets, or percentage of assets characteristics at the circuit level as necessary for each variable.

c. All data was only for assets and notification data within High Fire Threat Districts (HFTD), Tier 2-3 and Zone 1 from the CPUC Fire Map 2, from 2015 – 2017. All asset data and notification data used in this analysis was provided by PG&E

d. The top 10 independent variables included in the model ranked by importance are as follows:

   i. Conductor: Broken or Damaged – Notification describe signs of conductor being broken or damaged
   ii. Material AR – Percent of Miles – % of line with Aluminum Reinforced material
   iii. Conductor Size 6 – Percent of Miles - % of line with Conductor size 6
   iv. Conductor Size 4 – Percent of Miles - % of line with Conductor size 4
   v. Material C – Percent of Miles - % of line with Copper material
   vi. Pole Species – Percent Western Red Cedar - % of Non-Steel poles with Western Red Cedar species
   vii. Transformers: Rated kVA – 15 – Transformers with rated kVA of 15
   viii. Total Transformers – Number of transformers on line
   ix. Total Poles – Number of poles on line
   x. Poles: Original Treatment Cellon Gas – Number of poles treaded with Cellon Gas

e. However, the 30 variables were found to be significant in predicting the dependent variables. These variables are the ones listed in Columns H-AK in the worksheet.

4.2.2 Likelihood of Failure Calculation – Dependent Variable (Column AL)

a. The number of outages on each circuit was selected as the dependent variable (the variable to predict) as a proxy to estimate circuits with a higher likelihood to fail, which could possibly lead to an ignition. Specifically, a circuit was deemed to have a higher likelihood to fail if the circuit scored in the top quartile outages per circuit mile of all circuits in PG&E's service territory in Tier 2 and Tier 3 areas.

b. Data used for this analysis was total unplanned outages from 2015-2017 in Tier 2 and Tier 3 areas from ILIS outage data.

c. To define the values for the dependent variable for each circuit, a top quartile circuit was calculated as a circuit if the circuit had greater than 1.167 outages per mile. If a circuit scored in the top quartile, it received a 1 as the dependent variable. If it did not score in the top quartile, it received a 0.

d. The dependent variable for the distribution likelihood calculation was the number of unplanned outages between 2015 and 2017. Circuits were given a value of 1 if they had outages per mile in the top quartile (greater than 1.167 per mile) and a zero if not.

4.2.3 Likelihood of Failure Calculation – Calculated Likelihood (Column AM)

a. The set of independent variables and dependent variable for each circuit was entered into a Random Forest model. In order to determine the regression type, multiple regressions were investigated to see which best represented the desired outcomes. The regressions investigated included logistic regression, random forest model, Poisson model, negative binomial model, zero inflated variable model, and linear regression. A random forest is an ensemble machine learning method that can be used for classification or regression. It constructs a multitude of decision trees when training the model and outputs the mode classification (classification) or the mean prediction (regression) of the individual trees. A random forest model was chosen because it is a very robust modelling method. Random Forest Models are preferred for large data sets with a large number of fields and are able to deal with unbalanced data and correct for overfitting of individual decision trees. Based on the random forest model regression, coefficients were assigned by circuit in order to determine likelihood

b. In this context, the random forest model used a classification model to calculate the likelihood of each circuit having a high likelihood of failure (a dependent variable value of 1) or not (a dependent variable value of 0). See Section 4 for further discussion on the random forest model.

c. The model determines 1) which set of independent variables best predicts the dependent variable 2) the significance level of each independent variable and 3) the likelihood that the each circuit has a high likelihood of failure. The key output is the likelihood of failure for each circuit calculated as a percentage.

d. The likelihood percentage output from the model for each circuit is one of the three components in the Wildfire Risk calculation.

4.3 Likelihood of wildfire spread and consequence (Column AN)

This score was developed by a third-party firm, Reax Engineering. This score was developed following a similar fire modeling methodology to one that influenced the development of the High-Fire Threat Districts included in CPUC's Fire Map 2, Data Source 4. Although this component is shown as one number, this value is derived two separate outputs 1) likelihood of wildfire spread and 2) wildfire consequence score.

4.3.1 Likelihood of wildfire spread

Reax Engineering developed a spread score by using such factors as:

i. Fuel type and density (grass vs. brush)
ii. Topography (slope and natural fire breaks)
iii. Weather / Wind
iv. Distance from fire station / air suppression bases (speed to suppression)

4.3.2 Wildfire Consequence score

Reax Engineering developed a consequence score by using such factors as:

i. Density of population
ii. Density of structures
iii. Potential negative impact to natural resources

4.3.3 Combined Risk Score

An output from the Reax Engineering model was a combined likelihood of wildfire spread and consequence score, which was used for this risk calculation. These scores were calculated as output at 300m resolutions throughout the PG&E's service territory around PG&E assets.

4.3.4 Circuit score calculation

a. Given that circuits can span many miles and these spread scores can change across the span, a mileage-weighted score for each circuit was calculated for each circuit. This was conducted by finding the score from the Reax model for each conductor segment of the circuit, then aggregating the score at the circuit level at on a mileage-weighted basis.

b. For further reading on the calculations, methodology, and output of the Reax Engineering model, see Reax Engineering whitepaper, Data Source 3.

4.4 Egress Score (Column AO)

Egress is focused on the potential ease of accessing or exiting a community in case of a mass evacuation during a wildfire. At each census location (town and incorporated community) in California, an assessment was performed that considered the population and the road density, while taking into consideration the type of roads.

4.4.1 Approach

Egress is focused on the potential ease of accessing or exiting a community in case of a mass evacuation during a wildfire and included factors such as:

b. Population of towns and unincorporated communities
c. The road density for each community by road type
    i. Highways / Interstates
    ii. Country roads
    iii. Residential roads

4.4.2 Data Details

The egress data includes all incorporated and census designated locations in California including roughly 1,500 locations. Egress data was integrated and overlayed with each other using GIS. The road segment lengths were calculated within GIS in miles for all 2 million road records.

4.4.3 Methodology

At each census location (town and incorporated community) in California, an assessment was performed that considered the population and the road density, while taking into consideration the type of roads.

3. **Population and proximity to High Fire Threat Districts (HFTD)**
    i. To assess the population size for each location in proximity to a HFTD, the census population counts were used and superimposed with the HFTDs designated areas on CPUC's Fire Map 2.

4. **Road Density**
    i. The analysis examined the total miles of road in each location using three categories of roads identified from the 2017 National Transportation Dataset.
    ii. Highways
    iii. Country Roads
    iv. Residential Roads

4.4.4 Egress Calculation

a. The egress score for each location is calculated by dividing population from section 2.4.3.1 above and dividing by road density in section 2.4.3.2 above.
b. With these location egress scores, the five nearest locations to each conductor segment of a circuit are identified within a 100 mile radius. The egress score of each location is then divided by the distance of the location to the conductor segment. The average of these five distance-adjusted scores is then calculated to determine an egress score for the conductor segment.
c. A circuit egress score is calculated by taking the mile-weighted average egress scores across each conductor segment. As a final step, the log of this score was calculated. This score results in the third component of the wildfire risk prioritization score calculation.
d. For further information on the Egress score, see Egress Analysis working paper, Data Source 1

4.5 Wildfire risk prioritization score (Column AP)

The final score for each circuit was calculated by multiplying the three components: likelihood of failure, the likelihood of wildfire spread and consequence score, and egress score. With this score, circuits were prioritized for EVM work with those with higher scores receiving a higher priority than those with a lower score. In some cases, PG&E made changes to the prioritization order of circuits based on operational factors (e.g. land and environment, safety, planned projects, geographic access, weather, government relations, and customer communications).

4.6 Risk Tiers (Columns AQ-AT)

The last columns in the worksheet indicate the relative rank of each risk component score for the circuit to the total set of scores. If one of the risk components (likelihood of failure, the likelihood of wildfire spread and consequence score, and egress score) has a score in the top third of total scores for all the circuits for that risk component, a value of “High” is shown. The middle third receives a value of "Medium", and the bottom third receives a value of "Low." Column AT follows the same methodology for risk component, however, this indicates the relative rank of the circuit's wildfire risk score.

5 Risk Prioritization Methodology for Transmission Enhanced and Accelerated Inspection Process

The risk-based prioritization of distribution circuits for Distribution Enhanced and Accelerated Inspection Process work in 2019 was based upon a model comprised of three components: 1) likelihood of failure, 2) likelihood of wildfire spread and consequence score, and 3) egress score. The likelihood of ignition was calculated using a regression methodology, the likelihood of wildfire spread and consequence was based upon model outputs from a third-party fire model, and the egress factor was calculated through understanding the at risk population's ability to exit an area via the town's current road infrastructure. Multiplying these three components resulted in a wildfire prioritization score for each circuit.

The output for this prioritization model is contained in Wildfire Risk Assessment Prioritization Output_Consolidated_190223_vF.xlsx, tab 'Tx-Inspection.' The following sections provide further detail on the data shown in this tab.

5.4 Circuit Information (Columns C-G)

This section provides the circuit name (feeder name), circuit ID, district, and the number of structures in scope for each circuit included in the analysis.

5.5 Likelihood of Failure Calculation (Columns G-P)

Likelihood of failure (Column AM) was determined using a regression analysis based upon a set of independent variables (Columns G-N) and a dependent variable (Column O). The regression analysis, using a logistic model, was used to determine what characteristics of the circuit (independent variables) were more likely to indicate the likelihood for a failure to occur (dependent variable).

5.5.1 Likelihood of Failure Calculation – Independent Variables (Columns G-N)

a. Independent variables are the factors that may indicate the occurrence of the dependent variable. For this analysis, over 1,000 independent variables were analyzed to determine which set of variables best predicted the dependent variable. This set consisted of the variables from the following categories:
   i. # of structures
   ii. # of notifications
   iii. # of ignitions
   iv. Frame type (Ex. Dead End, Horizontal)
   v. Conductor Size
   vi. Conductor Type
   vii. Proximity to Coastline

viii. Steel vs. Non-Steel

b. The preceding data was gathered at the structure level. For this analysis, this data was aggregated to the circuit level to determine counts of structure, miles of conductor characteristics, or percentage of structure characteristics at the circuit level as necessary for each variable.

c. Any circuit within Tier 2, Tier 3, or Zone 1 were considered in scope. Additional scope was included for (1) structures immediately outside of HFTD (within .2 miles of the nearest non-steel structure or within .5 miles of the nearest steel structure) and (2) structures outside of HFTD but within an 80%+ quartile of relative spread risk (See 4.II - Likelihood of Spread and Consequence) along a circuit that is within HFTD. Scope was based upon structures (See "Number of in scope structures"). All asset data and notification data used in this analysis was provided by PG&E

d. The nine independent variables included in the model ranked by importance are as follows:

   a. # of HFTD structures – Number of structures in HFTD area
   b. Conductor type copper – Number of Miles – Number of miles with copper material
   c. Frame type tri-arm – Percent of Structures - % of structures with a Tri-Arm frame type
   d. Bird notifications – # of total notifications - Number of notifications referencing birds
   e. Number of steel pole structures – Percent of structures steel vs non-steel
   f. Frame type horizontal – Percent of Structures - % of structures with a Horizontal frame type
   g. Frame type wishbone – Percent of Structures - % of structures with a Wishbone frame type
   h. Conductor size 4/0 – Percent of miles – Percent of miles with a 4/0 conductor size
   i. Cracked insulator notifications – # of total notifications – Number of notifications referencing a cracked insulator

5.5.2 Likelihood of Failure Calculation – Dependent Variable (Column O)

a. The number of outages on each circuit was selected as the dependent variable (the variable to predict) as a proxy to estimate circuits with a higher likelihood to fail, which could possibly lead to an ignition. Specifically, a circuit was deemed to have a higher likelihood to fail if the circuit higher than an average number of outages per circuit mile of all circuits in PG&E's service territory in Tier 2 and Tier 3 areas.

b. Data used for this analysis was total unplanned outages from 2015-2017 in Tier 2 and Tier 3 areas from ILIS outage data.

c. To define the values for the dependent variable for each circuit, a greater than average outage circuit was calculated as a circuit if the circuit had greater than 6.533 outages per mile. If a circuit scored higher than average, it received a 1 as the dependent variable. If it did not score in higher than average, it received a 0.

d. The dependent variable for the distribution likelihood calculation was the number of unplanned outages between 2015 and 2017. Circuits were given a value of 1 if they had outages greater than average (greater than 6.533 outages) and a zero if not.

5.5.3 Likelihood of Failure Calculation – Calculated Likelihood (Column P)

a. The set of independent variables and dependent variable for each circuit was entered into a logistic regression model. In order to determine the regression type, multiple regressions were investigated to see which best represented the desired outcomes. The regressions investigated included logistic regression, random forest model, Poisson model, negative binomial model, zero inflated variable model, and linear regression. A random forest is used to understand how changes in independent variables are associated with changes in the probability of an event occurring. A logistic regression model was chosen for this methodology because 1) the dependent variable is binary and 2) and the output of a logistic regression is a predicted probability that a circuit or protection zone will be a value of 1, or in this case an outage.

b. In this context, the logistic regression model used a classification model to calculate the likelihood of each circuit having a high likelihood of failure (a dependent variable value of 1) or not (a dependent variable value of 0). See Section 4 for further discussion on the random forest model.

c. The model determines 1) which set of independent variables best predicts the dependent variable 2) the significance level of each independent variable and 3) the likelihood that the each circuit has a high likelihood of failure. The key output is the likelihood of failure for each circuit calculated as a percentage.

d. The likelihood percentage output from the model for each circuit is one of the three components in the Wildfire Risk calculation.

5.6 Likelihood of wildfire spread and consequence (Column Q)

This score was developed by a third-party firm, Reax Engineering. This score was developed following a similar fire modeling methodology to one that influenced the development of the High-Fire Threat Districts included in CPUC's Fire Map 2, Data Source 4. Although this component is shown as one number, this value is derived two separate outputs 1) likelihood of wildfire spread and 2) wildfire consequence score.

5.6.1 Likelihood of wildfire spread

Reax Engineering developed a spread score by using such factors as:

i. Fuel type and density (grass vs. brush)
ii. Topography (slope and natural fire breaks)
iii. Weather / Wind
iv. Distance from fire station / air suppression bases (speed to suppression)

5.6.2 Wildfire Consequence score

Reax Engineering developed a consequence score by using such factors as:

i. Density of population
ii. Density of structures
iii. Potential negative impact to natural resources

5.6.3 Combined Risk Score

An output from the Reax Engineering model was a combined likelihood of wildfire spread and consequence score, which was used for this risk calculation. These scores were calculated as output at 300m resolutions throughout the PG&E's service territory around PG&E assets.

5.6.4 Circuit score calculation

a. Given that circuits can span many miles and these spread scores can change across the span, a mileage-weighted score for each circuit was calculated for each circuit. This was conducted by finding the score from the Reax model for each conductor segment of the circuit, then aggregating the score at the circuit level at on a mileage-weighted basis.

b. For further reading on the calculations, methodology, and output of the Reax Engineering model, see Reax Engineering whitepaper, Data Source 3.

5.7 Egress Score (Column R)

Egress is focused on the potential ease of accessing or exiting a community in case of a mass evacuation during a wildfire. At each census location (town and incorporated community) in California, an assessment was performed that considered the population and the road density, while taking into consideration the type of roads.

5.7.1 Approach

Egress is focused on the potential ease of accessing or exiting a community in case of a mass evacuation during a wildfire and included factors such as:

i. Population of towns and unincorporated communities
ii. The road density for each community by road type
    i. Highways / Interstates
    ii. Country roads
    iii. Residential roads

5.7.2 Data Details

The egress data includes all incorporated and census designated locations in California including roughly 1,500 locations. Egress data was integrated and overlayed with each other using GIS. The road segment lengths were calculated within GIS in miles for all 2 million road records.

5.7.3 Methodology

At each census location (town and incorporated community) in California, an assessment was performed that considered the population and the road density, while taking into consideration the type of roads.

1. **Population and proximity to High Fire Threat Districts (HFTD)**
   i. To assess the population size for each location in proximity to a HFTD, the census population counts were used and superimposed with the HFTDs designated areas on CPUC's Fire Map 2.

2. **Road Density**
   i. The analysis examined the total miles of road in each location using three categories of roads identified from the 2017 National Transportation Dataset.
   ii. Highways
   iii. Country Roads
   iv. Residential Roads

5.7.4 Egress Calculation

1. The egress score for each location is calculated by dividing population from section 2.4.3.1 above and dividing by road density in section 2.4.3.2 above.
2. With these location egress scores, the five nearest locations to each conductor segment of a circuit are identified within a 100 mile radius. The egress score of each location is then divided by the distance of the location to the conductor segment. The average of these five distance-adjusted scores is then calculated to determine an egress score for the conductor segment.
3. A circuit egress score is calculated by taking the mile-weighted average egress scores across each conductor segment. As a final step, the log of this score was calculated. This score results in the third component of the wildfire risk prioritization score calculation.
4. For further information on the Egress score, see Egress Analysis working paper, Data Source 1

5.8 Wildfire risk prioritization score (Column S)

The final score for each circuit was calculated by multiplying the three components: likelihood of failure, the likelihood of wildfire spread and consequence score, and egress score. With this score, circuits were prioritized for transmission inspection work with those with higher scores receiving a higher priority than those with a lower score. In some cases, PG&E made changes to the prioritization order of circuits based on operational factors (e.g. land and environment, safety, planned projects, geographic access, weather, government relations, and customer communications).

5.9 Risk Tiers (Columns T-V)

The last columns in the worksheet indicate the relative rank of each risk component score for the circuit to the total set of scores. If one of the risk components (likelihood of failure, the likelihood of wildfire spread and consequence score, and egress score) has a score in the top third of total scores for all the circuits for that risk component, a value of "High" is shown. The middle third receives a value of "Medium", and the bottom third receives a value of "Low."

6 Understanding HFTD miles

In order to understand the inclusion of miles within scope for each circuit / protection zone, there are some adjustments that were made to arrive at the total miles.

(1) EVM circuit scope only included circuits with trees tagged to them; therefore, any circuit without any tagged trees are considered out of scope.

(2) The mileage scope of System Hardening and Distribution Prioritization are different. This is due to the methodologies used to arrive at the estimated miles. Whereas System Hardening was based upon protection zones with at least some area touching HFTD, Distribution Prioritization miles were based on the true miles within HFTD, potentially not including the entire circuit. This was done intentionally due to the operational nature of the work since it is difficult to complete system hardening intermittently on a line; whereas, it is less difficult to complete Distribution Prioritization intermittently.

7 Data Sources

1. Egress Analysis Working Paper WP_Egress Analysis 1-3-2019 v3_s1st.docx
2. Consolidated Wildfire Risk Prioritization Output Wildfire Risk Assessment Prioritization Output_Consolidated_190223_vF.xlsx
3. Reax Engineering Model 2017 Reax Model Report.pdf
4. CPUC Fire Maps http://www.cpuc.ca.gov/firethreatmaps/
5. The Census data base was used for the population counts for each of the towns and unincorporated communities for egress https://www.census.gov
6. The 2017 National Transportation Dataset (NTD) for California was downloaded in GIS shapefile format from the USGS website: https://www.sciencebase.gov/catalog/item/581d6764e4b0dee4cc8e5b57
7. Included in the egress data is a TNMFRC field which categorizes each road segment by the respective type of road: https://www.usgs.gov/faqs/what-are-code-value-definitions-tnmfrc-attribute
8. California Boundary GIS shapefile was used to define the towns and unincorporated communities and downloaded from the Census website: https://www.census.gov/geo/maps-data/data/cbf/cbf_place.html