Response to OIR Data Request Appendix A.1	Rulemaking:	20-01-007
Subject: Quantitative Risk Assessment Method	Date:	10/21/2022

PURPOSE

To provide a description of SDG&E's medium pressure mains quantitative risk assessment (QRA), including data sources for risk assessment modeling, input variable definitions, description of the QRA model, and output variable definitions.

1. INTRODUCTION

SDG&E, as part of the Distribution Integrity Management Program (DIMP), developed a segment-specific QRA model for medium pressure mains using a combination of internal SDG&E datasets and external publicly available data sources. Medium pressure mains, referred to as mains in the rest of this document, are defined as mains which operate at 60 psig or less. SDG&E uses this QRA model to estimate safety risk, defined as the probability of a serious incident per year, along the mains distribution network.

The remainder of this document is organized as follows: Section 2 describes the scope of the risk assessment, Section 3 describes the input variables and their data sources used in the QRA model, Section 4 provides a summary of the developed QRA model, and Section 5 describes the output variables of the QRA model.

2. SCOPE OF THE RISK ASSESSMENT

Medium pressure mains operating at 60 psig or less with known date of operation are subject to this analysis. Other distribution pipelines are not subject to this analysis, this includes medium pressure mains with unknown date of operation, medium pressure services, and high pressure mains & services (where high pressure is defined as operating above 60 psig but less than 20% SMYS).

3. INPUT VARIABLES AND DATA SOURCES

This section provides an overview of input variables and their data sources that are used in the QRA model development. Input variables and their data sources can be grouped into two categories, which are internal SDG&E and external publicly available categories. The internal and external input variables and their data sources are described in Section 3.1 and Section 3.2, respectively.

3.1. INTERNAL SDG&E INPUT VARIABLES AND DATA SOURCES

This section provides an overview of the internal input variables and their data sources used to develop the QRA model. These variables and their data sources are grouped into six categories, which are pipeline segments, leak records, set pressure, cathodic protection, one-call tickets, and polymer lab.

3.1.1. Pipeline Segments

This data source contains the geospatial location of mains along with tabular information about the attributes of each main segment. [redacted] A business district is an area where distribution facilities are located within [redacted] a potential commercial gathering place, a church, a school, a hospital, or is a location where people have limited mobility. [redacted]

3.1.2. Leak Records

This dataset identifies the geospatial location of individual leaks in the SDG&E distribution network along with tabular information containing the attributes of each leak record. [redacted]

[redacted]

3.1.3. Set Pressure

The SDG&E system of record for set pressure is used as an input variable in the risk assessment model development.

3.1.4. Cathodic Protection

The Cathodic Protection (CP) maintenance and inspection data consists of two main datasets: CP Areas and CP Reads. A CP area is an electrically continuous group of mains that are under a type of cathodic protection (i.e., sacrificial or impressed current systems). The CP Areas dataset contains tabular information on CP areas in the SDG&E network. [redacted]

[redacted]

3.1.5. One-Call Tickets

This dataset contains locations of historical one-call tickets. [redacted] This variable is used as an input to the QRA model.

3.1.6. Polymer Lab

The Polymer Lab dataset contains information related to tests conducted by the Polymer Lab located at SDG&E Engineering Analysis Center (EAC). [redacted]

[redacted]

3.2. External Input Variables and Data Sources

This section provides an overview of the external publicly available input variables and their data sources used to develop the QRA model. These input variables and data sources are grouped into seven categories, which are soil properties, land slide, tree canopy, susceptibility to corrosion, soil temperature, annual rainfall, and building footprints.

3.2.1. Soil Properties

This dataset includes several geospatial raster layers extracted from the SoilWeb application [1]. These raster layers are used to create input variables that characterize soil properties in the SDG&E network.

3.2.2. Land Slide
This dataset is obtained from the California Geological Survey's Landslide Inventory [2] and Susceptibility to Deep-Seated Landslides map [3]. This dataset is used to evaluate the landslide susceptibility in the SDG&E network.

3.2.3. Tree Canopy
A geospatial polygon layer outlining tree canopy cover is obtained from the U.S. Forestry Service's Geospatial Technology and Applications Center [4]. This dataset is used to calculate percent tree canopy estimates across all land covers and types.

3.2.4. Susceptibility to Corrosion
This dataset is a geospatial raster layer that is extracted from the Soil Survey Geographic Database (SSURGO) provided by the Natural Resources Conservation Service [5]. The dataset is used to develop an input variable indicating the dominant susceptibility condition of uncoated steel to corrosion when in contact with the soil.

3.2.5. Soil Temperature
The soil temperature dataset consists of the land surface temperature data and California weather station soil temperature data. The land surface temperature is extracted from the United States Geological Survey website [6]. The California weather station soil temperature data includes surface and soil temperature at the weather stations operated by the California Irrigation Management Information System [7]. These datasets are used to estimate the soil temperature across the SDG&E network.

3.2.6. Annual Rainfall
The annual rainfall dataset is a geospatial raster layer outlining the average annual precipitation (rainfall). This dataset is obtained from the Advanced Hydrologic Prediction Service (AHPS) provided by the National Weather Service [8]. The average annual precipitation is used as an input variable to the QRA model.

3.2.7. Building Footprints
The building footprints dataset contains computer generated building footprints in all 50 U.S. states. The building footprints dataset were generated using a deep Convolutional Neural Network (CNN) algorithm that was applied to optical satellite imagery. The building footprints are obtained from a Microsoft ODbL (Open Data Commons Open Database License) dataset that is published online [11]. This dataset is used to estimate the distance between each leak location and the nearest building as an input to the QRA model.

3.2.8. PHMSA Gas Distribution Annual Data
The *Gas Distribution Annual Data - 2010 to present* [9] dataset contains annual reports for each operator that list the network mileage and count of leak incidents by threat category. This dataset is used to determine the number of hazardous leaks reported to PHMSA for mains.

3.2.9. PHMSA Gas Distribution Incident Data
The *Gas Distribution Incident Data - January 2010 to present* [10] dataset consists of detailed event-specific information for incidents. This dataset is used to determine the number of explosions and serious incidents reported to PHMSA for mains, where a serious incident is defined as an incident causes fatalities or injuries that require in-patient hospitalization.

4. QUANTITATIVE RISK ASSESSMENT MODEL

Risk is a measure of potential loss in terms of both the likelihood of occurrence of an event and the magnitude of the consequences from the event. In the QRA, the event that is modeled is a failure being a loss of containment or a leak. At the highest level, risk is calculated by the following equation:

$$Risk = Likelihood\ of\ Failure \times Consequence\ of\ Failure \tag{1}$$

The developed QRA model is a segment-specific data-driven model which seeks to estimate the safety risk, defined as the probability of a serious incident per year, for each segment on SDG&E's medium pressure main distribution network. The QRA model consists of two main components, which are the likelihood of failure and consequence models. The likelihood of failure model is developed to estimate the annual probability of loss of containment (also referred to as probability of leak in this document) for each main segment. The consequence model is structured such that it estimates the probability of a serious incident for each main segment given a leak has occurred on that segment.

[redacted]

[redacted] [redacted]

[redacted]

4.1. LIKELIHOOD OF FAILURE (LOF) MODEL

The LOF is measured as the annual probability of leak for each main segment in the SDG&E network. The annual probability of leak is estimated using a gradient boosted decision tree model. The trained LOF model is used to predict the probability that a pipe segment will leak on an annual basis.

4.2. CONSEQUENCE OF FAILURE MODEL

To assess the potential for a serious incident after a loss of containment, SDG&E modeled the probability that a leak is hazardous, followed by the probability that the hazardous leak results in a serious incident. [redacted]

4.2.1. Probability of a Hazardous Leak Given Leak

[redacted]

4.2.2. Probability of a Serious Incident Given Hazardous Leak

This probability is estimated by the rate of serious incidents per hazardous leak. Input variables described in Section 3.2.8 and 3.2.9 are used to quantify this rate. [redacted]

5. OUTPUT VARIABLE DEFINITIONS

The developed QRA model has three segment-specific outputs, including the likelihood of failure, consequence of failure, and risk. Below are the descriptions of these output variables:

- Likelihood of Failure: Annual probability of leak for each main segment
- Consequence of Failure: Probability of a serious incident for each main segment given a leak has occurred on that segment
- Risk: Annual probability of a serious incident for each main segment

As mentioned earlier in Section 1 and Section 3.2.9, serious incident is defined as an incident that causes fatalities or injuries that require in-patient hospitalization.

REFERENCES

[1] O'Geen, A., Walkinshaw, M., & Beaudette, D. (2017). SoilWeb: A Multifaceted Interface to Soil Survey Information. Soil Science Society of America Journal, 81(4), 853–862.

[2] California Geological Survey. (2015). California Landslide Inventory. Retrieved December 3, 2019, from https://maps.conservation.ca.gov/cgs/lsi/app/

[3] Wills, C.J. Perez, F.G. and Gutierrez, C.I. (2011). Susceptibility to Deep-Seated Landslides in California. California Geological Survey. Map Sheet 58. https://www.conservation.ca.gov/cgs/Documents/Publications/Map-Sheets/MS_058.pdf

[4] USFS. (2016). USFS Tree Canopy Cover Datasets. Retrieved from: https://data.fs.usda.gov/geodata/rastergateway/treecanopycover/

[5] NRCS. (2020). Geospatial Data Gateway. USDA, https://gdg.sc.egov.usda.gov/GDGHome_DirectDownLoad.aspx

[6] Earth Resources Observation and Science (EROS) Center. (2019). EROS Moderate Resolution Imaging Spectroradiometer (eMODIS) Land Surface Temperature (LST) [Data set]. U.S. Geological Survey. https://www.usgs.gov/centers/eros/science/usgs-eros-archive-vegetation-monitoring-emodis-land-surface-temperature

[7] California Irrigation Management Information System (CIMIS). Available online: https://cimis.water.ca.gov/

[8] National Weather Service. (2021). NOAA. "About the Precipitation Analysis Pages." AHPS Precipitation Analysis. Available online: https://water.weather.gov/precip/about.php

[9] PHMSA. (2020). Gas Distribution, Gas Gathering, Gas Transmission, Hazardous Liquids, Liquefied Natural Gas (LNG), and Underground Natural Gas Storage (UNGS) Annual Report Data. Retrieved from https://www.phmsa.dot.gov/data-and-statistics/pipeline/gas-distributiongas- gathering-gas-transmission-hazardous-liquids

[10] PHMSA. (2020). Distribution, Transmission & Gathering, LNG, and Liquid Accident and Incident Data. Retrieved from https://www.phmsa.dot.gov/data-andstatistics/ pipeline/distribution-transmission-gathering-lng-and-liquid-accident-andincident- data

[11] Microsoft (2018). USBuildingFootprints. https://github.com/Microsoft/USBuildingFootprints