6512 Rutherford Road
Plano, TX 75203
(512) 693-8280
contact@simcon-solutions.com
www.simcon-solutions.com

SUPPLEMENTARY DOCUMENT

Title:	**SIMCON Statistics Case Study for Williams Landslide Risk Prediction**		
Document ID:	SIMCON-SPL-010		
To:	Prospective and Existing Clients	**Date:**	04/30/2020
From:	SIMCON Solutions LLC	**Revision #:**	2
Description:	This document is a case study of a statistical analysis project where SIMCON developed a landslide risk analysis and prediction model for Williams.		

Table of Contents

1. Purpose

This case study describes a statistical analysis project where SIMCON developed a landslide risk analysis and prediction model for Williams. This project was undertaken between October 2015 and June 2016. This case study was prepared by SIMCON in conjunction with our client and outlines the client challenges, the SIMCON solution approach, and the key takeaways from the project. It both reflects our experience working on the project and our client's perspective of working with SIMCON on this particular risk modeling project.

2. Client Challenges

Williams Strategic Sourcing Company (Williams) communicated a need for assistance in predicting and mitigating landslide-initiated pipeline ruptures in the Ohio Valley Midstream (OVM) region. The methods they were employing to assess landslide risk, such as LIDAR and manual overhead land surveys via helicopter, were proving inadequate for accurately predicting landslide events across the pipeline region. Pipeline ruptures are extremely costly, both to Williams and to the surrounding environment. In order to support their objectives, Williams contracted SIMCON in October of 2015 to develop a quantitative methodology for determining the sections of pipeline most susceptible to landslides and, consequently, pipeline ruptures. The risk model was to be used to determine areas of existing pipeline where mitigations should be implemented and areas of land that should be avoided for new pipelines. The primary objectives of this project were to:

- Accurately predict past, present, and future landslide events as a function of available data
- Generate landslide risk estimates for existing pipelines
- Generate landslide risk estimates for potential future pipelines
- Provide a means for visualizing landslide risk across the pipeline using heat mapping in Williams GIS software

3. SIMCON Solution Strategy

SIMCON leveraged statistical methods and machine learning algorithms to address the client's challenges. Williams provided our engineering team with historical landslide data for the pipeline region, which divided the pipeline region into 10-m by 10-m grid-cells. A collection of environmental factors (e.g., soil properties) was provided for each grid-cell along with a binary response indicating whether a landslide had been observed in the grid-cell. The SIMCON engineering team then began conducting preliminary analyses on the data to:

- Determine if parametric statistical regression methods could be used (e.g., evaluate linearity assumptions, identify presence and degree of spatial correlations, etc.)
- Quantify the relationship between the environmental factors (i.e., predictors) provided and the actual landslides that had been observed in the region

As a result of these preliminary analyses, we determined that the underlying data was nonlinear and exhibited a high degree of spatial correlation between neighboring grid-cells. These observations indicated that parametric regression methods were not well suited for predicting landslides as a function of the underlying data. Instead, the SIMCON team began experimenting with an array of machine learning algorithms to conduct the landslide risk analysis. This consisted of:

- Dividing the historical landslide data into complementary "training" and "validation" datasets
- Training each of several machine learning algorithms on the "training" datasets
- Using the "trained" machine learning algorithms to predict landslides in the validation datasets
    - Experimenting with alternative parameter specifications for each machine learning algorithm
- Cross-validating predictive models against validation datasets
- Validating the predictive models on new landslide data collected for the same pipeline region
- Applying these models to generate landslide risk heat maps across the pipeline region

Our engineering team then used the most effective machine learning algorithms to generate landslide risk heat maps and compared those maps to new landslides delineated by Williams to further measure their efficacy. Finally, SIMCON documented all analysis methodologies and results and submitted the final report and associated models to Williams.

4. Results and Key Takeaways

The k Nearest Neighbors (KNN) machine learning algorithm was found to be the most effective model for predicting landslide risk in the Williams pipeline region. The final KNN model:

- Demonstrated excellent accuracy for predicting past and present landslide events
- Revealed several new (i.e., non-delineated) areas at high risk of a future landslide event.
- Revealed areas that should be avoided for proposed pipelines
- Provided Williams with a reusable tool to supplement existing landslide prediction methods

Figure 1 and Figure 2 compare two of the heat maps generated by the final KNN algorithm to actual landslides observed in the Williams pipeline region.

In summary, SIMCON was able to provide Williams with an effective tool that significantly improved their ability to predict past, present, and future landslides in the OVM pipeline region as a function of readily available environmental factors. As a result, Williams is now better equipped to identify and reinforce segments of pipeline in areas at high risk of a future landslide event. Circumventing even one landslide-related pipeline rupture as a result of using this tool has the potential to save Williams millions of dollars in revenue and reputation, as well as immeasurable human and environmental costs.

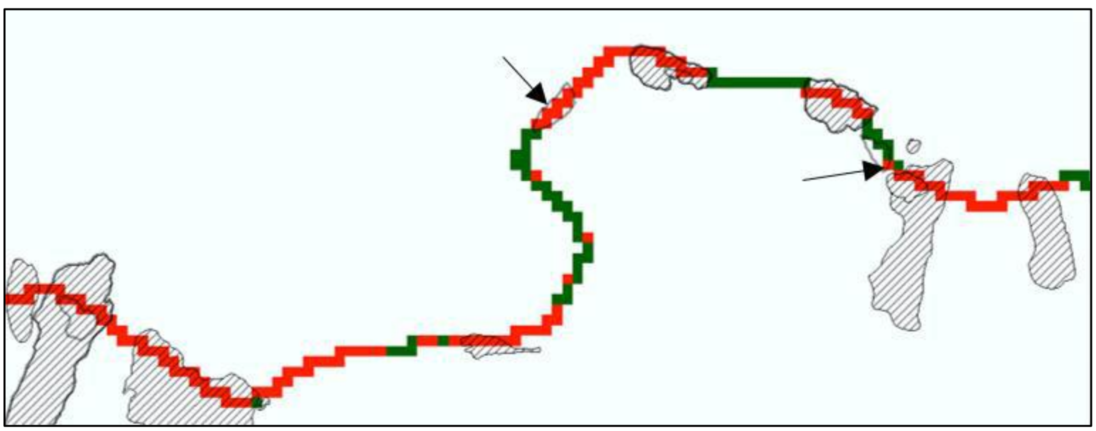

Figure 1. Two-Level Heat Map - KNN Model Results vs. Actual Landslides (Shaded Regions)

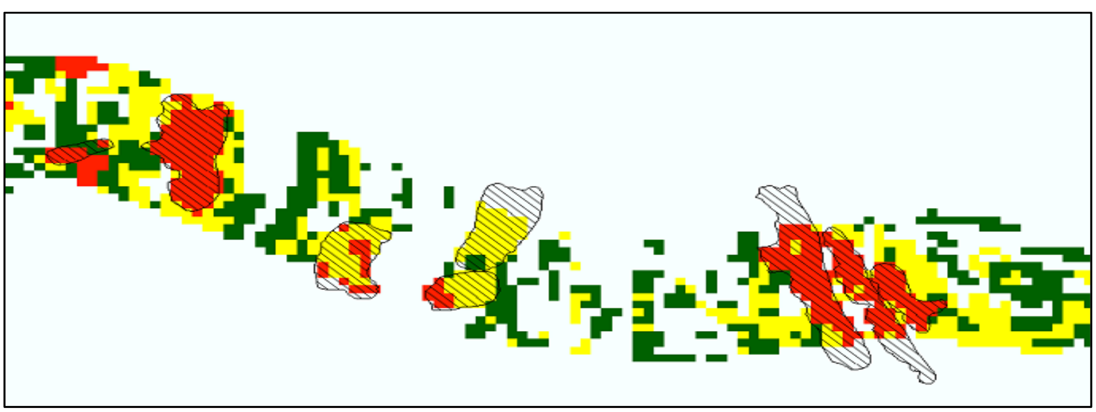

Figure 2. Three-Level Heat Map - KNN Model Results vs. Actual Landslides (Shaded Regions)