



Statistical Assessment of Operational Risks for
Induced Seismicity from Hydraulic Fracturing in the
Montney, Northeast BC

Geoscience BC Report 2020-12
(Geoscience BC Project 2019-008)

Final Report

Appendix A – Feature Definitions

by Amy Fox, Scott McKean and Neil Watson

Enlighten Geoscience Ltd.

October 2020

This appendix describes all of the features considered for the model, divided into completion parameters, well survey and production parameters, geological parameters and seismic catalog targets. Each feature is listed with a short description and other relevant information - for example the type of feature (categorical or numerical), the rationale for its inclusion in the initial model or its exclusion from the final model, or the number of missing features.

Completion Parameters

avg_stage_length (numeric)

The average stage length of the completion. There are 1,508 missing entries in the OGC data before filtering for missing proppant and fluid values. 434 are imputed with the mean stage length of the data.

avg_stage_spacing (numeric)

The spacing between stages, center-to-center. There are 1,483 missing entries in the OGC data. The remaining wells with missing data (20) are imputed with the mean stage stage spacing of the data.

base_fluid_cat (categorical)

The base fluid used in the well. Removed from further analysis because a large majority of wells used water as the base fluid.

breakdown_isip_ratio (numeric)

The calculated ratio of breakdown pressure to instantaneous shut-in pressure (ISIP). No further filtering or imputation is necessary once proppant, fluid, ISIP, and breakdown are filtered from the dataset.

calc_fluid_intensity_m3_m (numeric)

The calculated fluid intensity, taken as the total fluid injected divided by the completed well length. No further filtering or imputation is necessary after filtering for missing proppant and fluid values.

calc_proppant_intensity_kg_m (numeric)

The calculated proppant intensity, taken as the total proppant injected divided by the completed well length. No further filtering or imputation is necessary after filtering for missing proppant and fluid values.

calc_total_fluid_m3 (numeric)

The total fluid injected in the well, non-normalized for length. Calculated by summing the fluid injected in stage for each well.

calc_total_proppant_t (numeric)

The total amount of proppant injected in the well, non-normalized for length. Calculated by summing the proppant injected per stage for each well.

cased_well_bool (categorical)

There are two levels to the category: true and false. Roughly 60% of the study wells are cased.

energizer_bool (categorical)

This logical variable encodes whether the completion was energized. There are two levels to the factor: true and false. Roughly 15% of wells in the study are energized.

frac_duration_days (numeric)

The total duration of the completion in days, measured from the start time of the first frac stage to the end time of the final frac stage.

hybrid_frac_bool (boolean)

A Boolean indicator for a hybrid frac – very few wells are indicated as hybrid (true), and this feature is removed from further analysis.

interference_bool (categorical)

There are two levels to the category: true and false. Roughly 20% of the study wells reported interference (true), which is likely an underreported number when compared to industry experience in the Montney and other unconventional plays.

[mean, sd, max]_breakdown_mpa (numeric)

There are 1,432 missing standard deviation values and 1,421 missing mean breakdown measurements in the data. Wells with missing mean breakdown data (28 after filtering proppant, fluid, and isip) are also filtered from further analysis. Wells with missing standard deviation are imputed with the mean standard deviation.

[mean, sd, max]_fluid_per_stage_m3 (numeric)

The fluid injected in each completion stage. There are 1,388 missing entries in the OGC data. We provide three features for analysis in the model - the mean, the standard deviation, and the maximum value. Wells without a mean proppant per stage are excluded since it is considered a primary causal factor. Remaining wells without a standard deviation (17) are imputed to the mean standard deviation of the dataset.

mean_intervals_per_stage (numeric)

There are 1,457 missing values, 70 after filtering for proppant, fluid, breakdown, and ISIP. These missing values are imputed with the mean value of the dataset.

[mean, sd, max]_isip_mpa (numeric)

There are 1,934 missing standard deviation values and 1,504 missing instantaneous shut-in pressure (ISIP) measurements in the data. Wells without a mean ISIP are excluded from the model during feature engineering. The remaining wells with missing ISIP standard deviation are imputed to the mean standard deviation of the data.

[mean, sd, max]_proppant_per_stage_t (numeric)

The proppant for each completion stage. There are 1,388 missing entries in the OGC data. We provide three features for analysis in the model - the mean, the standard deviation, and the maximum value. Wells without a proppant per stage are excluded since it is considered a primary causal category. Remaining wells without a standard deviation (3) are imputed to the mean standard deviation of the dataset.

[mean, max]_rate_m3_min (numeric)

There are 1,425 missing entries for the mean rate in the OGC data. The remaining wells with missing data (40) are imputed with the mean injection rate of the dataset.

[mean, max]_stage_duration_min (numeric)

There are 1,435 missing entries for the mean stage duration in the OGC data. The remaining wells with missing data (45) are imputed with the mean stage duration of the dataset.

[mean, sd, max]_treating_mpa (numeric)

The standard deviation has 1,494 missing values, the mean has 1,439 missing values. The remaining wells (43 for mean and 90 for standard deviation) after filtering are imputed with the mean treating pressure of the dataset.

n_stages (integer)

The total number of completed stages in the well.

number_sand_off_screen_out_stages (integer)

The number of sand off or screened out stages. Assumes zero if not completion information is provided, and therefore has no missing values.

proppant_cat (categorical)

This categorical variable encodes the type of proppant used. There are five levels to the factor: ceramic, hybrid, resin-coated, sand, and unknown. The majority of the wells in the study use a hybrid proppant (60%) or natural sand (35%).

skipped_stages (integer)

Number of skipped stages in a well. We remove this feature from consideration since it may represent a data leak. Wells with induced seismicity often respond by skipping stages.

stim_company (categorical)

There are ten levels to the category, although the majority of the stim company values in the OGC dataset are registered as unknown. This variable is removed from further analysis.

total_gas_injected_m3 (numeric)

The total gas injected in a well - generally considered a low variance parameter since only 136 wells have gas injection in the study area. Remove from further analysis because the parameter had nearly zero variance due to very few wells injecting gas.

viscosity_cat (categorical)

This categorical variable lists the base fluid used in the frac, which is the primary control on viscosity. There are four levels to the factor: slickwater, water, linear, crosslinked. A large majority of completions in the study are slickwater in the study (2,495 of 2,721 wells after the above filtering).

well_completed_length_m (numeric)

The completed length of the well, measured as the largest measured depth of the toe completion minus the smallest measured depth of the heel completion. Wells without a measured completed length (1,383 in the OGC data) are excluded when proppant and fluid are excluded. No further filtering or imputation is necessary after filtering the aforementioned features.

Survey and Production Parameters

horiz_wells_in_[1,5,10,25]_km (numeric)

The number of horizontal wells within 1, 5, 10, and 25 km of the observed well. A proxy for well interference and the density of field development.

mean_ss_tvd (numeric)

The mean true vertical depth (TVD) of the well lateral or deviated section. Surveys with TVDs above zero are filtered out of the study, since they represent a data break.

min_midpoint_dist (numeric)

The minimum distance from the observed completion and the surrounding wells, taken as the midpoint distance of the lateral section. The lateral section is filtered out of the survey by taking all values with a mean inclination above 88 degrees. Wells with midpoint distances of zero (32) are filtered out of the dataset, since they represent a potential refrac that could confound the analysis.

no_prior_wells (numeric)

The number of wells completed before the current well. Removed from further analysis due to sparse data and lack of physical meaning.

no_surrounding_wells (numeric)

The total number of wells surrounding the current well. Removed from further analysis due to sparse data and lack of physical meaning.

Geological Parameters

distance_all_faults_berger_m (numeric)

The distance from each grid point to all faults (listed above and including strike-slip, thrust, normal, listric, and divergent) in the Berger study (Davies et al., 2018).

distance_divergent_faults_berger_m (numeric)

The distance from each grid point to divergent faults in the Berger study (Davies et al., 2018).

distance_listric_faults_berger_m (numeric)

The distance from each grid point to listric faults in the Berger study (Davies et al., 2018).

distance_normal_faults_berger_m (numeric)

The distance from each grid point to normal faults in the Berger study (Davies et al., 2018).

distance_strike_slip_faults_berger_m (numeric)

The distance from each grid point to strike-slip faults in the Berger study (Davies et al., 2018).

distance_thrust_faults_berger_m (numeric)

The distance from each grid point to thrust faults in the Berger study (Davies et al., 2018).

geothermal_gradient_degc_km (numeric)

The geothermal gradient in degrees celcius per km.

paleozoic_structure_m (numeric)

The absolute value of the paleozoic structure in m.

pressure_depth_ratio_kpa_m (numeric)

The reservoir pressure to depth ratio in kPa/m.

shmin_grasby (numeric)

The minimum horizontal stress derived from map by Grasby (2012).

third_order_residual_m (numeric)

The value of the third order residual of the Paleozoic structure.

top_montney_isotherm_degc (numeric)

The absolute isotherm at the top of the Montney structure.

top_montney_structure_mss (numeric)

The elevation of the top of the Montney in meters subsea.

top_montney_tvd_mss (numeric)

The true vertical depth to the top of the Montney in meters subsurface.

Seismic Catalog Targets

mag_mag (numeric)

The maximum magnitude of all the associated seismic events in each well. For the regression analysis, wells without associated seismicity are excluded and the maximum magnitude is encoded as zero.

seismogenic (categorical)

The results of the seismogenic association indicating if a well is associated with seismicity. There are two levels to the factor: true and false. This is the target value for the seismogenic classification.