APPENDIX III.B.3 - PREDICTIVE MODEL AND PRIORITIZATION

DRAFT FOR PUBLIC COMMENTS

Version 1.0: July 11, 2019

DENVER WATER

AECOM Imagine it. Delivered.

AECOM
1600 S Quebec Street
Greenwood Village, CO 80111
aecom.com

To: **Alexis Woodrow**
Nicole Poncelet-Johnson

CC: **Steve Ravel**
John Wujek
Quirien Muylwyk

Project name:
Denver Water

Project ref:
LRPP

From:
Constantine Kontos,
Chistopher Pawlowski,
Matthew Harris,
Elizabeth McIlwee,

Date:
July, 2019

DRAFT

Appendix III.B.3 Predictive Model and Prioritization

Table of Contents

List of Tables

List of Figures

Acronyms

p-value	Predictive Model Probability of Lead Score
ACR	American Census Record
ALSLR	Accelerated Lead Service Line Replacement
CDPHE	Colorado Department of Public Health and Environment
COE	Community, Outreach, and Education
EPA	Environmental Protection Agency
LRP	Lead Reduction Plan
LRPP	Lead Reduction Program Plan
LSI	Lead Service Inventory
LSL	Lead Service Line
LSLR	Lead Service Line Replacement
O&M	Operation and Maintenance
PPB	Parts Per Billion
RF	Random Forest
WM	Watermain

References

1. Denver Water Standards / Specifications
2. Denver Water Procurement Process
3. Preliminary Identification of Lead Service Lines (Appendix III.B.2)
4. Filter Program Plan (under development)
5. Accelerated Lead Service Line Replacement Plan (Appendix III.D.1)
6. Communications, Outreach, and Education Plan (Appendix III.A)
7. Spatial Surveillance of Childhood Lead Exposure in a Targeted Screening State: An Application of Generalized Additive Models in Denver, Colorado, 2017

Background and Purpose

The Lead Reduction Program Plan (LRPP) is supported by an Accelerated Lead Service Line Replacement (ALSLR) Program (see Appendix III.D). Denver Water will replace all known lead service lines within its service area (including distributor areas) within 15 years of the approved variance.

The ALSLR Plan details the process and resource estimates to replace the preliminary estimate of 75,000 [final number will be agreed upon by Environmental Protection Agency (EPA), Colorado Department of Public Health and Environment (CDPHE), and Denver Water at August 13th Variance Check-in #5 and inserted into the final Lead Reduction Program Plan (LRPP) submittal due August 20, 2019.] LSLs at a minimum annual replacement rate of 7% per year based on a cumulative annual average. Of interest is the number of total lead services estimated in the Denver Water service area: not only will this serve as the basis for the target for annual replacements, but it also serves as the basis for developing the ALSLR Plan. To efficiently identify the number of lead services that exist in the Denver Water service area, a predictive model will be used with the lead service line inventory to strategically perform explorations.

This technical memorandum (TM) describes the development of the predictive model. The predictive model will be used to generalize the results of explorations completed to date and to guide subsequent explorations in the future, without having to undertake an excavation at every property. Once developed, the predictive model will be applied to the Denver Water Lead Service Line Inventory to prioritize enrollment in the Filter Program and prioritize the replacement of LSLs. The TM defines data sources used to populate the predictive model and its application toward prioritization regarding those efforts.

[This TM will be updated by the August submission, following updates to the Lead Service Line Inventory.]

Predictive Model Implementation

Introduction

Denver Water's LSL Inventory was developed to identify LSLs within Denver Water's service area and surrounding communities (see Appendix III.B.2, Preliminary Identification of Lead Service Lines). A set of logic rules was applied to the data to sort service lines into groups based on the estimated probability that an LSL is present. The probability represents the uncertainty in our knowledge of the service line material and is captured as a "p-value" that is assigned based on known construction practices, historical records, expert judgement, and data interpretation. The inventory assigns a p-value score to each property to guide Filter Program enrollment, service line material exploration, and LSL replacement. The p-value score ranges from 0 to 1, with 0 being a known non-lead service not lead and 1 being a known lead in the service line. The service connections are grouped into classes of likelihood based on p-value. Table 1 (Estimate of Service Materials Based on Probabilities of Lead) shows the estimated number of services in each class. The inventory currently contains a preliminary estimate of approximately 335,457 records and will be updated to incorporate additional information periodically. [To be updated by August 2019 submission.]

Table III.B.3-1. Estimate of Services Based on Probabilities of Lead
DRAFT FOR DISCUSSION/EXAMPLE ONLY

Probability of Finding a Lead Service	p-Value	Estimated Number of Services
Known LSL	p = 1	1,118
Suspected LSL	0.8 ≤ p < 1	63,597
Possible LSL	0.5 ≤ p < 0.8	36,533
Unlikely LSL	0.01 ≤ p ≤ 0.05	83,543
Not LSL	p = 0	150,666
*Preliminary numbers to be updated with the Revised LSL Inventory based on Appendix III.B.2		

The existing inventory (see Appendix III.B.2, Preliminary Identification of Lead Service Lines) was constructed based on data available from several sources. It includes apartments, schools and businesses. To date, some fieldwork has been performed to gain a better understanding of the estimated number of LSLs. As additional data become available for a property, the p-value score for properties with similar characteristics will be adjusted accordingly to reflect the inventory updates. Enhancements to the inventory and predictive model are underway to support enrollment in the Filter Program and implementation of the ALSLR.

Using the Lead Service Inventory to Build the Predictive Model

A predictive model will be used throughout the ALSLR Program to take advantage of results from field verification and service line replacements to better estimate the materials expected. This data-driven approach will permit the estimation of the probable presence of an LSL based on observed property and other common characteristics. The recommended approach involves the use of a machine learning model known as a random forest (RF) (Breiman, 2001).

The RF can be set to include existing rules and has the capacity to generate new rules based on the discovery of relationships between input and output variables. In addition, this approach offers the means to audit and explain the decision-making process. Finally, the model can be used to address data inconsistencies, handling data measurements on a variety of scales, and categorical data.

The model will be used to build on the current LSI inventory based on learning from the results of completed work. The model will make use of the results of field LSL data collection indicating service line composition found, as well as potholing data collected to verify presence/absence of LSLs in areas not participating in the ALSLRP. This data driven approach will permit the estimation of “p-value” scores based on observed direct evidence findings and other common characteristics incorporated into the model. These will include tap data as derived for the initial LSL inventory (year installed, etc.) as well as possible additional variables (sewer age, median income levels, etc.). The model calibration and verification process will identify variables that contribute significantly to the accurate identification of LSLs.

The RF model uses an ensemble of individual decision trees to assign a decision and a probability to observations. A simple decision tree is shown in **Figure 1 Decision Tree Example**.

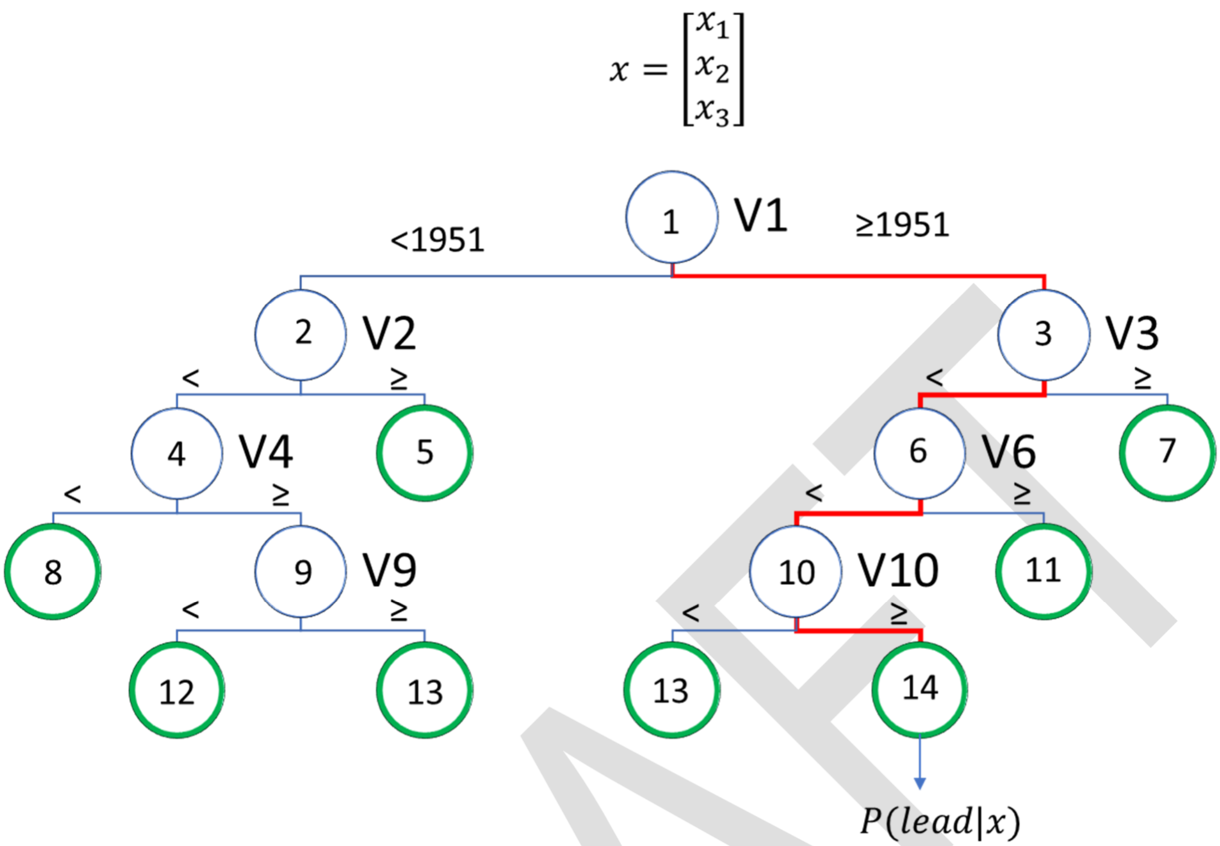


Figure 1 Decision Tree Example

Figure 1 Decision Tree Example has four layers of decision processing. The branching node labeled as 1 is where all properties enter the tree and are split according to a characteristic variable (V1). This variable could represent the tap year with the threshold of 1951 representing the year prior to which service lines are most likely lead (the algorithm uses statistical methods to decide on the variables and thresholds to be used). In this example, properties will be split to node 2 or 3 based on the year and a probability of lead being present assigned based on this variable alone. If this split perfectly distinguished lead services from non-lead services in the data, we could stop there, but this will not be the case. The next layer of decisions at nodes 2 and 3 will use two additional variables (V2 and V3) to further split the property services, such as tap size, and assign probabilities for the presence of lead at the child nodes. This process continues using different thresholds of different variables until the algorithm decides to terminate the branching process. These terminal nodes (known as leaf nodes, in green) contain all the property services. Each leaf node classifies the services that fall into it based on the suite of variables expressed in the rules necessary to reach it. A prediction for a property service based on this tree simply considers all relevant variables starting at node 1 and splits through each node until it lands in a leaf node.

The RF algorithm uses many individual trees (as described above) that are randomized both in terms of the data sampled for training (known as “Bagging” (Breiman 1996)) and the variables used at each split in the decision tree. Each tree provides a prediction that are on average close to the true mean (low bias), but inherently noisy and sensitive to changes in the data (high variance). When the “forest” of many low bias and high variance trees are averaged for the final model, each tree contributes a vote, thereby reducing the variance and retaining a low bias (Hastie et al. 2009:588). Figure 2 shows a schematic representation of how the ensemble method works. This example shows individual trees (1 through b), which can number as computer resources allow, although there is a point of dimensioning returns.

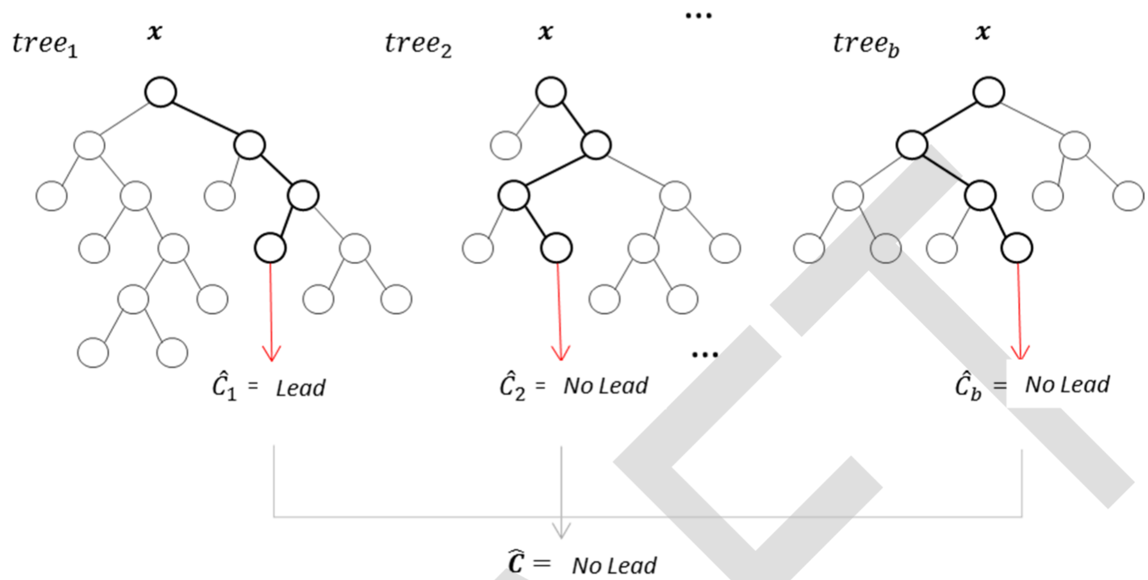


Figure 2 Schematic of Prediction Based on a Decision Tree Ensemble

As a new property observation (x) is sent though each tree, it is split at each node according to the splitting criterion established when the tree was fit to field inspection data that is essential for project delivery. This follows the track of the bolded branches and nodes depicted in **Figure 2 Schematic of Prediction Based on a Decision Tree Ensemble**. Even though the property data x are the same, each tree sends it along a different path because each tree was built with a randomized sample and randomly selected variables at each node. For each tree, the property is split until it reaches a leaf node and its service is then assigned a classification and probability. This can be represented as $\hat{C}_b(x)$ where C-hat is the predicted class of x for the b^{th} tree.

The final prediction, represented as $\hat{C}^B_{rf}(x)$, is simply the class that the majority of trees agree on, in this case two out of three trees predicted that x is not lead. The algorithm also provides a final probability that property x service line is lead, expressed as $P\left(\hat{C}^B_{rf} \mid x\right)$.

An algorithm incorporating the RF model based on decision trees was selected because it is a natural extension of the decision logic developed for the initial LSL inventory. Further, there is a precedent for incorporating RF as part of lead service line models as discussed by Abernethy et al. (2018), Chojnacki et al. (2017), and Goovaerts (2019) for Flint, Michigan; Gurewitsch (2019) for Pittsburgh, Pennsylvania; and Ardila et al. (2016) for Chicago, Illinois. The approach has the capacity to give priority to existing classification rules, to generate new rules based on the discovery of relationships between input and output variables, and weight specific observations. Furthermore, this approach offers the means to audit and explain the decision-making process through machine learning explanation tools (Biecek and Burzykowski 2019). Finally, is robust to data on different scales of measurements as well as categorical data.

The RF algorithm is a non-parametric tree-based estimator focused on reducing prediction variance through the use of randomization (bagging) and the majority-votes principle of an ensemble (Breiman 1996).; as discussed in the text above. The assumptions of this approach are like other parametric and non-parametric classification models. Namely that the input data consist of a set of observed outcomes in specific classes and a series of variables that lend to the discrimination of the observed classes. It is assumed that the observed classes are *exchangeable*, meaning that the reordering of each observation does not change the outcome (i.e. the data do not represent a time series or possess some other inherent ordering). It is also assumed the explanatory variables are not highly correlated, however RF is less affected by this property compared to other models. Finally, for the purposes of model diagnostics and scoring, it is assumed that the model residuals are normally distributed (violations of this assumptions can be verified, and appropriate action taken to control for this). It is acknowledged that spatial correlation will lead to bias in the assumptions of exchangeability and residual distribution. For these reasons' additional steps for spatially valid cross-validation and neighborhood random effects are being explored.

Phase II Model Updates

The model enhancements will change the approach from inventory to prediction, based on field validated results. The predictive model will support decisions regarding the location of future construction activities, will provide support for long-term strategies to maintain the 7% target for LSL replacements, and will be referenced by the Communications, Outreach, and Education (COE) Plan. This will be completed by transforming estimates of the presence of LSLs into actionable items and developing a better understanding of the likelihood (or not) of finding a lead service line.

The predictive model will support Denver Water's annual ALSLR Plan by allowing Denver Water to focus efforts on the areas with a higher likelihood of lead. It will also be used by Denver Water to determine where additional verification activities are needed, particularly at properties enrolled in the Filter Program (i.e., possible lead service). It is anticipated that this model will be updated when field results are available from the previous year's activities. Both the model data inputs, the model itself, and the output property LSL probabilities will be assessed after each update to support the development and prioritization of construction work areas. It is currently projected that the model will be run twice a year to include probability and consequence updates to support enrollment in the Filter Program and construction sequencing of the ALSLR Program.

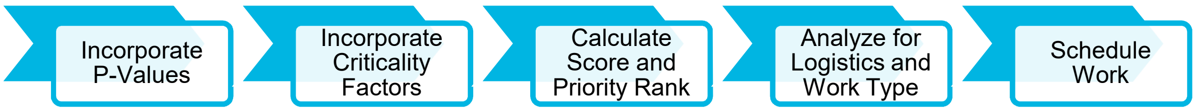
Prioritization

Properties with a known, suspected or possible lead service will be prioritized for i) Filter Program enrollment and ii) ALSLR implementation. The results of the prioritization analysis will be used to identify areas having the greater potential to benefit from the ALSLR Plan while considering logistical needs.

Prioritization involves developing a risk-based approach for long-term construction activity planning that accounts for public impacts to public health as well as capital improvements. The prioritization evaluation will be completed by integrating the p-value from the LSL Inventory with a measure of consequence.

There are three primary factors that are used to develop a prioritization. These include i) a probability of having a lead service, ii) considerations that affect the consequence of lead exposure, and iii) logistical constraints that need evaluation to turn planning into work activities. Each of these individual components are incorporated into an analytical process, the results of which will be evaluated for logistical constraints and then be administered for construction activities as shown in Figure III.B.3- 1 (Prioritization Process).

Figure III.B.3-3. Prioritization Process



Prioritization and Confirming a Lead Service

Under the current Lead Service Line Inventory, properties with known, suspected or possible LSLs will be provided with filter and replacement cartridges and be placed in the ALSLR Program. In order to implement the ALSLR Program prioritized sets of actionable properties must be extracted from the Lead Service Line Inventory based on risk and placed into contracting groups (see below and in Appendix III.D.1). Actions taken on properties within these groups will be based on the group type and their LSL status. These are presented in Table III.B.3-2 (LSL Status Cohorts and Actions).

Table III.B.3-2 LSL Status Cohorts and Actions

Group	LSL Status Cohort	ACTION AND RESPONSE		
		Filter Program	**Lead Inventory**	**ALSLR Program**
A	Known lead service line	Provide filter	Materials known	Add to list for replacement <u>or</u> remove from inventory / Filter Program through replacement
B	Suspected and Possible lead service line	Provide filter	Confirm materials	Add to list for replacement <u>or</u> remove from inventory / Filter Program through replacement
C	Unlikely to have a lead service	Statistically Defensible Select Sampling / COE		
D	Confirmed to be lead-free	Sampling / COE		
E	Other (fire lines, recycled water taps, consecutive system)	No Action / COE		

The preliminary set of cohorts and action groups shown in Table III.B.3-2 (LSL Status Cohorts and Actions) are based on the current Lead Service Line Inventory predictions. Note that the current inventory is such that properties of indeterminate LSL status due to limited or conflicting information are assigned to the possible lead cohort. This approach is conservative in the sense that properties with the most uncertain LSL status are included in the Filter Program. The predictive model that is under development will reduce this uncertainty and refine the inventory LSL status and allow for refinement of the target properties.

Revisions to the inventory LSL status based on future updates will allow action levels to be refined if needed to assess the estimated number of lead services and how this is reported. This review process will allow improvements to:

- Define the activities to most efficiently identify the materials of a service line.
- Identify additional workload needs to support explorations at properties with a known or suspected lead service.

Properties classed as having a known or suspected lead service will be visited and subject to a traditional exploration with the lead service replacement performed as necessary. Properties unlikely to have a lead service will be investigated using a variety of methods, such as records review, customer outreach, visual inspections, water quality sampling, and/or potholing as necessary (in that order) to confirm service line material. Those properties found to have a lead service will have it replaced. Properties confirmed to have no lead will be taken off the Filter Program. The results of these investigations will provide data to verify the results of the model and improve its predictive power.

Criticality Factors

The criticality factor is used to describe the potential impact of lead exposure, based on features unique to a property such as water quality sampling results or the demographics of the occupants.

The consequence of lead will provide a priority categorization separate from the likelihood of lead. It is another aspect to provide justification of selecting work locations on a yearly basis. Each property of the Lead Service Line Inventory will be evaluated to determine the consequence of lead exposure. Consequence factors associated with each property will be identified and weighted. The criticality factors and weights can be defined by analysis tools and/or stakeholder consensus agreement. The process to develop the consequence score is presented in Figure III.B.3- 2 (Criticality Weighting Process).

Figure III.B.3-4 Criticality Weighting Process

Table III.B.3-3A list of initial criteria proposed for consideration for the consequence factors and their weights is presented in Table III.B.3-3 (Proposed Consequence Factors and Criticality Weighting). These factors will be evaluated and combined as shown below. Values for factors will be determined based on the best available data. In some cases estimates will need to be substituted for missing values (e.g. using an average age for a missing age attribute). [Presented for discussion only and will be updated through discussions with Denver Water by August submission.]

Table III.B.3-3 Proposed Consequence Factors and Criticality Weightings
EXAMPLE FOR DISCUSSION ONLY

Consequence Factor* (Cf)	Description	Consequence Co-efficient*	Criticality Weight**(%) (Cw)
Public Health Consideration	Odds Ratio Contours from the Spatial Confounder-Adjusted Spatial Risk Model (Berg, et al, 2017)	odds ratio for childhood lead poisoning	0.3
Filter Adoption Rate	Areas where filter adoption is low.	non-successful filter adoption X 1) / total number of customers per area	0.2
Critical Customers	Day care centers and child care facilities, schools, dialysis centers, nursing facilities, multi-unit dwellings, jails, etc.	(total number of critical customers X 1) + (total number of non-critical customers X 0.5) / total number of customers	0.1
Age (Census Data)	Elderly	probability of adults over 70 years of age	0.2
	Children	probability of children under 5 years of age	
	Females likely within child bearing age range	Probability of Females within the range of 10-54 years of age***	
Social Economic Factors	Probability of being below the Federal Poverty Level	Probability of Residences that fall under the defined federal poverty level	0.2
	Economically depressed areas (identified by others)	Under evaluation	

* Preliminary consideration identified. Denver Water and CDPHE discussion to finalize.
** Preliminary list of criticality factors. Consideration needed from Denver Water and CDPHE to finalize. Details to be provided by August submission
***based on available ACS data that encompasses medically derived age bearing years.

The resultant criticality score that is between 0 and 1 is calculated as follows:

Equation 1 Consequence of Lead

$$F_{LSL_C} = Tap\ Prioritization\ Ranking\ Value$$

$$F_{LSL_C} = (Cw1 * Cf1) + (Cw2 * Cf2) + (Cwx * Cfx) \ldots$$

Probability Factors

The probability of the presence of a lead service is determined primarily through the LSI p-values, or through the predictive model p-values. Subsequent actions can be taken to revise these numbers. These include:

- Digging or potholing
- Water quality sampling [Corona is developing a memo discussing the validity of WQ sampling]
- Visual inspections (by field crews)
- Customer outreach
- Additional and/or more detailed records review

Digging and potholing is considered definitive confirmation of service line material. The validity of the remaining methods in assigning service line status will be evaluated as part of the ALSLR, and where appropriate such information integrated into the predictive model.

Risk Score

Once the individual likelihood (probability of lead, p-value) and criticality scores are generated for each property, a risk score is then calculated for the property to establish the individual risk score (Equation 2 Individual Risk Score).

Equation 2 Individual Risk Score

$$Risk\ =\ Probability\ of\ lead\ x\ Criticality\ of\ lead$$

Individual risk scores are totalized to a common spatial boundary (i.e., the 2010 Census Neighborhood Blocks / American Census Survey records (ACS)) to establish an overall risk score. The result of this is the aggregate risk over an area that is normalized to take into account density using the area of parcels in the census area with a tap (Equation 3 Normalized Risk Score).

Equation 3 Normalized Risk Score

$$Normalized\ Risk\ =\ \frac{\sum(Probability\ of\ lead\ x\ Criticality\ of\ lead)}{\sum Area\ of\ Parcels\ with\ Tap}$$

This controls for areas that have large open parcels with no (or few) taps, assigning a higher normalized risk to areas with higher density of development. The result of this analysis is that individual and accumulated risk scores can be assigned to a spatial feature for further evaluation including logistical considerations and ALSLR contracting work development.

Using Risk Scores to Prioritize Construction and Filter Distribution

The process used for establishing priority ranked activities is based on the results of the probability and consequence evaluations. The goal is to take the risk scores from all the (census) areas and look at replacing lead services in a way that addresses both the (high) risk of lead exposure at a property and the efficiency of working through an area of properties to consider the risk to a broader portion of the community. Additionally, locations that are high risk that are not incorporated in a census area for production are also evaluated for sequence of constructions. As a result, lead service replacements may be completed on an individual basis or as part of a larger grouping of properties.

Both prioritization risk scores are used to produce lists where both the greatest probability and the greatest consequence is considered. An example of how the individual and combined scores (from Equations 2 and 3) will be applied is described in Table III.B.3-4 (Applying Risk Scores for Prioritization).

Table III.B.3-4 Applying Risk Scores for Prioritization

Risk Score Types	Description
Individual	Individual scores are considered for properties defined as high consequence but are geographically isolated.
Geographic Area	Combined scores are considered for properties where the categories of known and possible lead scores define an area.

Evaluation of the two types of risk scores is the basis to prioritize i) enrollment in the Filter Program, ii) sequencing the ALSLR Program contracting needs, and iii) communication efforts. Additionally, the output from this analysis shows where additional investigative efforts are needed to drive the LRP and sustain the year-over-year annual targets for the number of LSL replacements. As described above, all properties in a high-risk contractor group derived using census areas will be investigated.

The Predictive Model and Coordination with Other Capital Programs

The results from the predictive model in terms of prioritizing LSL replacements will be evaluated with other activities within the Denver Water service area for scheduling and coordination of construction. Other considerations (mobilization, street repair, scheduled watermain replacement, etc.) are necessary scheduling components to minimize repeat visits to the same street or block and to efficiently complete the necessary LSL Replacements. The logistical considerations (see Table III.B.3-5, Predictive Model and Coordination with Other Capital Programs) will influence the development of construction activities. Additionally, information related to current customers will be identified to ensure that work is performed at connected services.

Table III.B.3-5 Predictive Model and Coordination with Other Capital Programs

Coordination Item*	Description
Previously Completed Partial Replacements (where some portion of the service line is still lead)	Public to curb box previously completed; follow-up work outside of the full replacements
Watermain Replacement Program Schedule	ALSLR based on scheduled watermain replacements
Long-term Roadway Full Depth Resurfacing Plan	Full depth or resurfacing roadway projects in areas susceptible to lead services
Leak Repairs and Operation and Maintenance Activities	ALSLR based on a response to reported leaks or necessary maintenance
Redevelopment Properties	City of Denver Development in areas susceptible to lead services
Archeological / Cultural / Historic areas / locations	Identification of areas requiring more sensitive construction coordination and approval
Property Type (Single Family Commercial / Industrial / Multi-dwelling units)	Building inventory of data regarding residential, commercial, and industrial units
Active Water Account	Identifies taps that have service agreements
Property Status	Identification of property status (occupied, abandoned, etc.)

Implementing the Predictive Model Outcomes

Upon completion of the analysis phase, information for i) individual and grouped risk scores and ii) logistical considerations that exist within the Denver Water service area will be available to support the annual planning cycle for the LRP. The next step is to apply the results to the Filter Program and ALSLR Program.

Filter Distribution Prioritization

Filter distribution will target the properties with a known, suspected, or probable LSL under the current LSI. The predictive model will be used to iteratively identify candidate properties for enrollment in the Filter Program.

ALSLR Prioritization

The process used to develop the ALSLR Program construction sequence is presented in Figure III.B.3-3 **(**The Role of the Prioritization Analysis for Annual Updates to the ALSLR Plan). This is based on taking the results from the predictive model and prioritization analysis to establish the annual ALSLR work activities.

Figure III.B.3-5 The Role of the Prioritization Analysis for Annual Updates to the ALSLR Plan

The predicative model and consequence data will be used to identify candidate properties for the different ALSLR contracting groups presented in Appendix III.D.1 (see Table III.D.8 Contracting Groups Summary). The contracting groups include:

- Group A – Geographic Area ALSLR Type
- Group A – Individual ALSLR Type
- Group B – Investigation

The risk scores developed from the prioritization analysis will be used to further define the groups and sub-group categories that will comprise the yearly work plan.

Group A - Geographic Area ALSLR Work Type

The list of properties with a known lead service included in Group A – Geographic Area ALSLR will be generated from the output of the prioritization risk analysis. Grouped risk score areas will be reviewed to identify the highest priority areas for inclusion in the annual ALSR Program's scope of work. The properties associated with the identified areas will be collectively issued to contractors for replacement of LSLs on blocks or streets as needed. Figure III.B.3- 4 (Geographic and Individual Area Visual Representation) shows an example of an area selected for the ALSLR Program. In this example, the results of normalized risk (Equation 3) were used to identify the work area. This geographical area shown below would hypothetically receive a high priority ranking and would incorporate all the properties within the boundary for the contract in accordance with the yearly construction goals. As described above, all properties in this group will be investigated using progressively more invasive methods based on p-values.

Figure III.B.3- 6 Geographic Area (left) and Individual Map (right) Visual Representation

Group A – Individual ALSLR Work Type

The properties with a known lead service included in Group A – Individual ALSLR are properties that were individually prioritized as "high" but are not in close enough proximity (geographically) for inclusion in the Geographic LSL Replacement Areas. This ASLR contracting strategy takes into consideration where critical properties would not typically rise to the top of the list from a grouped risk-based analysis. Figure III.B.3- 4 Geographic and Individual Area Visual Representation defines an area where the density of properties is low, but a select group of properties were defined to be critical for prioritized construction activities. In this situation the individual risk score (equation 2) was evaluated and the top ranked properties were identified in accordance with the yearly construction goals.

Group B – Investigation Work Type

The goal of investigation work areas is to gather more information where necessary to produce better predictive model results in areas where available information is limited and to provide a more representative sampling of data. Investigation type activities include detailed records review, non-intrusive inspections, water quality sampling and potholing to support the ALSLR construction and planning. In areas where there are groupings of similar properties with similar p-values, then a sample of the total group population will be investigated to evaluate the composition of taps at these properties.

Another example for where investigation is needed occurs at properties for which risk is high due to a high consequence of having a lead service, but the likelihood of lead is relatively low. In this case, investigations as described above will be performed to determine service line material and support better model prediction outcomes as new iterations of the predictive model are run.