Matrices and Algorithms: Applying Mathematical Models in Prioritizing Wastewater System Needs

David E. Koberlein, P.E., GSWW, Inc.
Jerome A. Iltis, P.E., San Antonio Water System
Steven D. Sanders, P.E., D.Eng, GSWW, Inc.
Joseph C. Molis, GSWW, Inc.
1016 Mopac Circle, Suite 201
Austin, Texas 78746

Abstract

Managing a wastewater collection system of over 4,400 miles of infrastructure is a challenge for any entity, but prioritizing the needs of the system on a budget requires the application of innovation. The San Antonio Water System (SAWS) with GSWW, Inc. has developed a means for logical prioritization of their 4,400 miles of sanitary sewer collection system by applying system data in a mathematical model that generates a prioritization score for inspection. The "Decision Tree Model" for large diameter mains and the "Collection System Algorithm" for small diameter mains was created specially for SAWS to address growing system needs through the application of existing collected system information.

Like many collection system owners, the SAWS wastewater collection system is constantly facing the challenge of handling new growth combined with existing aging infrastructure. The condition and age of large diameter mains had become such a concern that SAWS officials needed a way to prioritize the inspection and subsequent remediation of these critical mains. With environmental issues such as watershed protection and management, the prioritization methodology applied needed to addresses SAWS specific issues and not just general wastewater infrastructure evaluation criteria. The Decision Tree Model was developed to include categories such as age, while including area specific issues such as proximity to creeks and waterways. Running as a dynamic computer model and applying appropriate engineering logic, the risk of failure, and thus the need for inspection associated with a particular pipe segment could be computed. The accuracy of the model increases as more system data is collected and added to the model.

For small diameter sewer mains, another set of logic applies. While similar materials and construction methods are applied in small diameter wastewater infrastructure, the criterion associated with prioritizing needs is different. Small diameter mains are usually constructed at the same time, using similar materials and methods as in a residential community or subdivision, and thus exhibit similar properties. Unlike the large diameter infrastructure, previous system evaluations of the smaller diameter pipes in the system had been performed and volumes of data collected. Again, SAWS worked with GSWW, Inc. to develop a statistical method to prioritize the needs of the small diameter pipes in their system through the development of the Collection System Algorithm. Looking at

the system in clusters according to date of construction and applying results of the previous system studies; GSWW, Inc. used spatial tools to apply a mathematical algorithm to the SAWS small diameter wastewater infrastructure to determine the most advantageous order for inspection and renewal.

Keywords

Wastewater Collection Systems, Sanitary Sewer Overflows, Inflow and Infiltration, Collection System Management, Decision Models

Introduction

The San Antonio Water System is charged with a number of tasks relative to the management of the environment and the public infrastructure. Management of the environment includes activities such as prevention of sanitary sewer overflows and ensuring that the entire SAWS infrastructure system exists in harmony with the surrounding environment. Maintaining and managing the wastewater collection system serving the City of San Antonio in a specified operational condition is a major portion of this effort.

The management of a public utility system involves making decisions that can have long lasting economic, social, technical, political, and environmental impacts. For the significant and complicated decisions, tools, methods, policies, and information are available to aid in the process. Decision trees are utilized everyday for all types of decisions, ranging from as basic as “Do I want a cup of coffee?” to disaster evacuation planning. Decision Trees are efficient tools for providing a logical and defensible method for choosing between competing alternatives. A properly constructed decision tree allows for the investigation of several alternatives based upon available or predictable data and before committing to an expensive or irreversible course of action.

Matrices and algorithms are mathematical models that are similar in function to a decision tree in that they provide a structure to large volumes of data for the purpose of analysis and provide a logical and defensible path to an end result. Large diameter interceptor systems and small diameter collection mains have different failure modes, different impacts on the environment when they fail, and different operational characteristics. Because of these differences, the San Antonio Water System has chosen to participate in the development of both a Large Diameter Decision Tree Model and a Small Diameter Algorithm to help manage the wastewater collection system.

Daily, SAWS is involved in thousands of decisions relative to the wastewater collection system in order to provide the citizens of San Antonio with the best possible wastewater service. Each decision, no matter how small, has an impact on the management of the collection system. In support of this effort, SAWS has participated with GSWW, Inc. in the development of two modeling tools to assist with the decisions required for prioritization of inspection activities and risk of failure assessment of components of the wastewater collection system. One is a decision tree model for large diameter wastewater interceptors (36” and larger); the second tool is a prioritization algorithm for inspection of small diameter collector mains (15” and smaller). The purpose of developing both the large diameter decision tree model and the small diameter algorithm is to reduce the overall cost of ownership of the subject sewer mains by predicting areas where sewer main inspections should be conducted in an organized and prioritized manner. Rehabilitation and renewal of infrastructure before failure occurs is generally considered to be less costly than repairing the same infrastructure after it has failed.

The Large Diameter Pipeline Decision Tree Model

A generic decision tree model can be a number of things; artificial intelligence, neural networks, classification and regression trees. In its basic form, a decision tree model is simply the logical and graphical representation of the possible outcomes from a series of choices among a system of variables. The model is represented graphically as a set of nodes, which contain the original dataset and links that relate the nodes in a logical fashion along with data that considers the probability of the particular related event occurring. The individual parent nodes are connected to child nodes, which are connected to additional child nodes, and so on until the system of variables is fully represented. The dissolution of parent nodes into child nodes is called recursive partitioning. One of the assumptions of a decision tree model is that the user facing a pure uncertain decision must select at least and at most one final option from all of the possible options.

The purpose of the implementation of the SAWS Large Diameter Decision Tree Model is to determine, based on available information and experience, whether or not SAWS should expend the effort and funding to inspect particular large diameter interceptors and to provide an assessment of the risk associated with a major failure of individual pipe segments. The model developed for SAWS’ large diameter wastewater interceptors follows both an innovative and classic approach to development with the added feature of the application of engineering experience and logic associated with the relation of the nodes. The recursive partitioning exercise for the model resulted in the identification of more than 55 factors that could be related to the condition and operation of SAWS’ interceptors and risk associated with catastrophic pipeline failure. These variables represent all of the potential influences that could have an impact on the SAWS large diameter wastewater collection system as determined by SAWS and GSWW. It should be noted that this listing is not intended to be an all inclusive list of factors that could affect the performance of a wastewater collection system. A partial list of these factors is shown below.

Variable	Comment	Variable	Comment
Location relative to the Edwards Aquifer	Drinking Water Source	Soil Type	None
Majority Surface Cover over the Pipeline	None	Proximity to Drinking Water Sources	Used to Assess Consequences of Pipe Failure

Variable	Comment	Variable	Comment
Location of the Pipeline relative to High Impact Features in the City	Features such as historic landmarks, public areas, etc	Depth of Cover over Pipeline	None
Proximity to Lift Stations and Siphons	Used to Assess the Probability of H2S Concentrations	Approximate Surface Loading over Pipeline	None
City of San Antonio Land Use over the Pipeline	Used to Assess Consequences of Pipe Failure	Backfill and Embedment Characteristics	None
Pipe Capacity and Capacity in Use	From SAWS Hydraulic Models	Pipe Age	Used to Assess the Remaining Design Life of the Pipeline
Pipe Slope	None	Pipe Material	Used to Assess the Longevity of the Pipeline
Pipeline Criticality Rating	None	Pipe Length	None
Sediment Build-up	Used to Populate Condition Variables in the Decision Tree	Inspection History	None
Grease Build-up	Used to Populate Condition Variables in the Decision Tree	Pipe Diameter	None
Debris Build-up	Used to Populate Condition Variables in the Decision Tree	Manhole Surcharge	Used to Populate Condition Variables in the Decision Tree
Failure History	None	Manhole/Pipe Transition	None
SSO History	Used to Populate Condition Variables in the Decision Tree	Manhole Status	Active/Inactive Manhole
Odor	Used to Populate Condition Variables in the Decision Tree	Pipe Encasement	None
Manhole Debris/Sediment	None	H2S Concentrations	Indicator of Pipeline Deterioration
Flow Velocity	None	Cleaning History	None

Variable	Comment	Variable	Comment
Repair History	None	Pipe Wall Erosion/Corrosion	None
Roots	None	Offset Pipe Joints	None
Sags	None	Manhole Grease	None

In order to build and use a computer model, data is required. For the Large Diameter Decision Tree Model, data was initially available for 19 of the original 57 identified variables. These were categorized into three broad groups; Location, Physical, and Condition. These variables are listed in the following table.

Variable Name	Group	Variable Name	Group
Pipe Age	Physical	Floodplain	Location
Pipe Material	Physical	Pipe Impact	Location
Pipe Diameter	Physical	Surcharging	Condition
Horizontal Alignment	Physical	SSO Proximity	Condition
Design Capacity	Physical	Capacity Used	Condition
Edward's Aquifer	Location	Maintenance History	Condition
Surface Cover	Location	Odor	Condition
Land Use	Location	General Condition	Condition
Siphon Proximity and Type	Location	PACP Structural Score	Condition
Pipeline Crossing	Location	-	-

The three major groups of variables were then used as the basis for individual decision trees. These models were then combined into a single, larger model. This collection of inter-related models is called a Decision Tree Forest. This method of development also allows for a modular and expandable approach to model building.

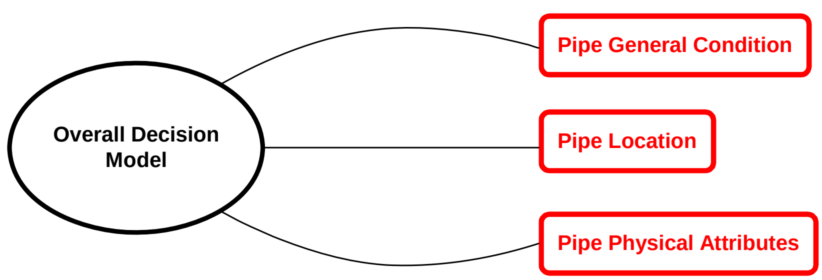

Figure 1 – Decision Tree Forest – Large Diameter Model

Features required for development (SAWS):

1. Modularity or the ability to add to or take away from the Large Diameter Model
2. Use of open architecture software in the development of both the Large and Small Diameter Models.
3. Models must be compatible with SAWS' geographic information systems (GIS) and to be able to display the results of the models as a visualization of the wastewater collection system.

Each of the three broad groups was then fashioned into an individual DTM based on the availability of data for each of the models. Identifying the variables is the first step in the model building process. The next step is to relate the variables and develop a pathway or link between each of the nodes in the model. The SAWS Large Diameter Model uses a "Go/No Go" inspection scenario as the final outcome of the model. The risk of pipeline failure is determined from the model inputs as they are processed through the model. A schematic view of the Decision Tree is shown below.

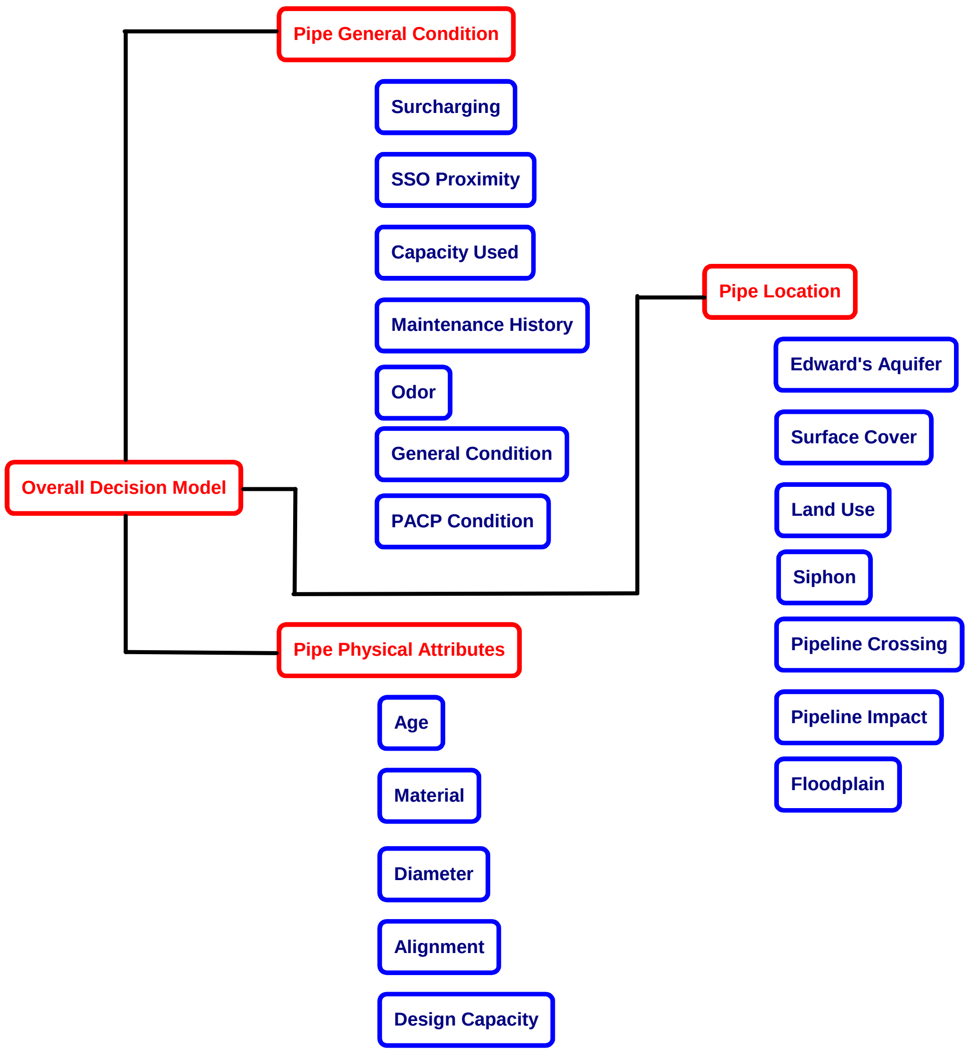

Figure 2 – Large Diameter Decision Tree Model Schematic

As shown above, each of modules in the Large Diameter Model focuses on a particular aspect of the overall model. In the PHYSICAL node of the model, the focus is on the physical attributes of the collection system, such as pipe diameter, material, length, and age. Based on data that is specific to the SAWS interceptor system and engineering experience, a score is assigned to each of the possible values of each of the physical variables for purposes of computation. As an example, the size of any pipeline is usually defined by a specific diameter such as 36”, 38”, 39”, and so on. Each of the valid values of pipeline diameter was assigned a specific score, with the scores normalized to a top value of 100.

Pipe Diameter
36" - 42" Score: 22.7
43" - 60" Score: 45.5
61" - 72" Score: 68.2
72" and larger Score: 100

The entire listing of variables and values used in the physical node is shown below.

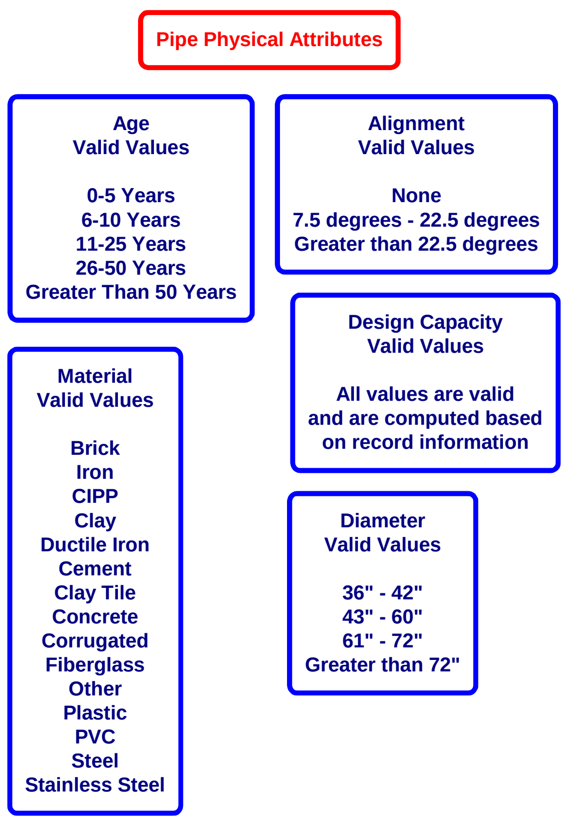

Figure 3 – Physical Model Node Attributes

The **LOCATION** node is where the model scores and ranks the influence of the pipeline location on the decision to perform inspections. The location of an interceptor plays a large part in the determination of the risk of pipeline failure.

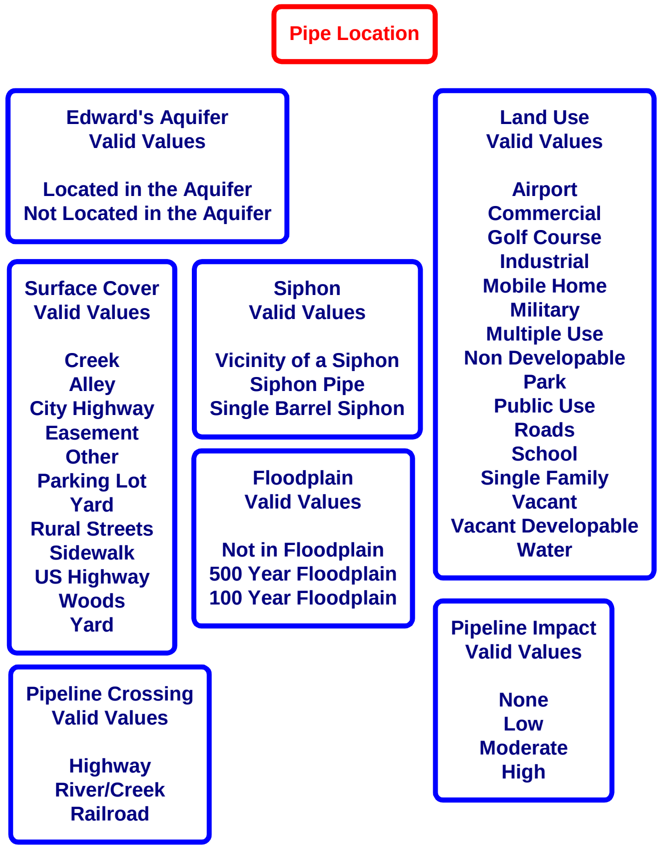

Figure 4 – Location Model Node Attributes

The **CONDITION** Node is perhaps the most complicated and most important of the three individual modules. Until recently, standardized assessments of the condition of a particular pipeline segment were not available. This has a bearing on the methods used to compute the risk associated with pipeline failure and in the computations associated with the model. As development of the program continued, SAWS adopted the National Association of Sewer Service Companies' (NASSCO) Pipeline Assessment Certification Program (PACP) for all future assessments of pipelines. The standardized assessment activity streamlines the relationship between condition nodes in the model. Existing condition assessments were applied to individual pipe segments using a spatial nearest neighbor analysis in the SAWS Geographic Information System. The condition module variables are used to assess the vulnerability of particular pipelines to failure and to apply condition information to neighboring pipelines for computation in the Decision Tree Model.

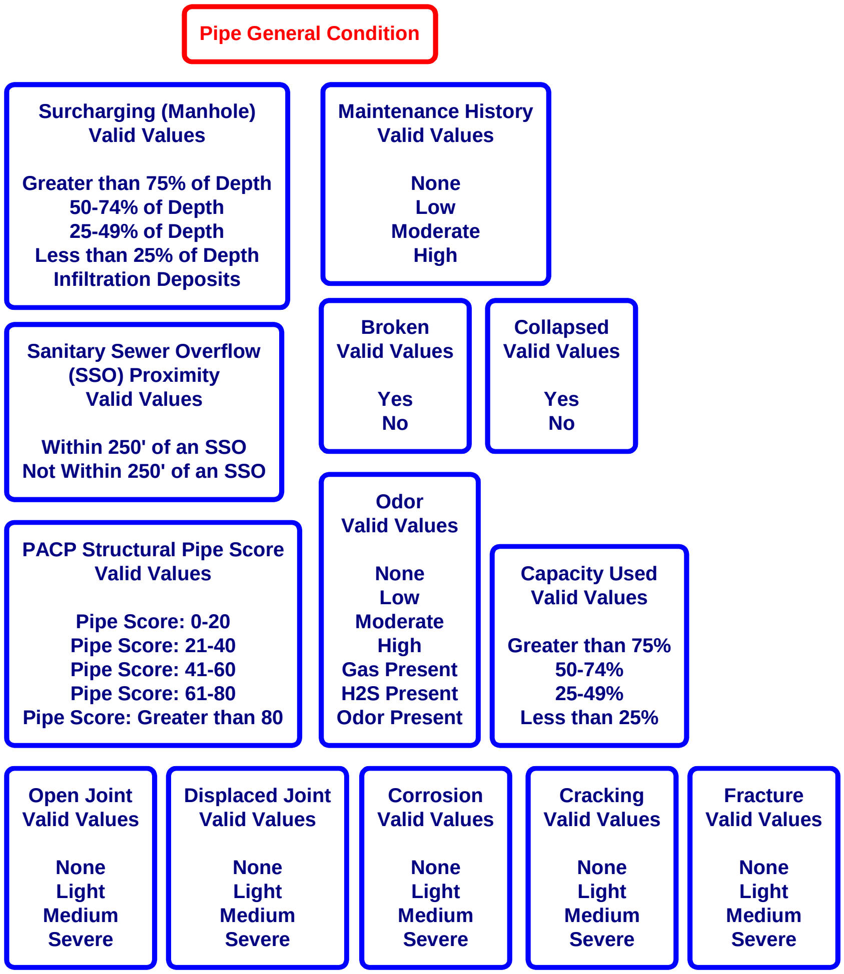

Figure 5 – Condition Model Node Attributes

The Large Diameter Decision Tree has two additional special modules that allow the end user to elevate selected outcomes based on data not considered in the model. These were added to account for issues that are outside of the scope of the operation and maintenance of the collection system, such as the location of particular pipelines in certain drainage areas and pipelines located in certain political subdivisions. A schematic diagram of the entire Decision Tree Model is shown on the next page.

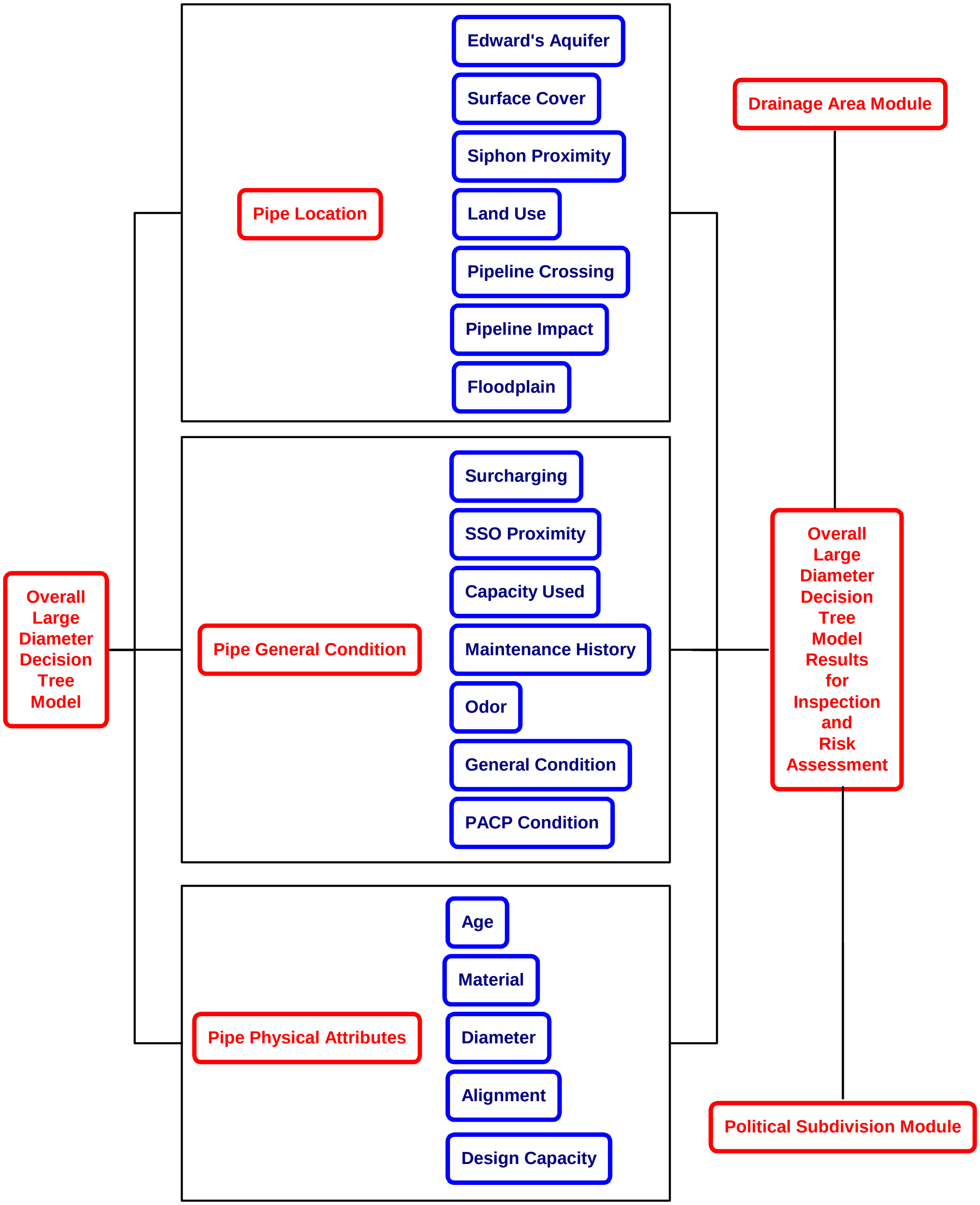

Figure 6 – Schematic Decision Tree Model Node

The definition of risk is generally associated with events or outcomes that are to be avoided. In SAWS' case, the computation of risk relative to the large diameter interceptor system is based on avoidance of a catastrophic failure of a pipe segment, protection of the public and the environment, and providing the best wastewater service possible at the lowest cost. Risk is a difficult thing to quantify. All of the data developed during the modeling exercise was integrated into GIS, per the SAWS requirements and the valid values of every component of the model is compatible and linkable with the overall SAWS Enterprise Data Management System without conversion. This has allowed SAWS and GSWW, Inc. to view the results of the model in a spatial format which facilitates the location of high risk "hot spots" and to utilize the data from the model in conjunction with other data and information, and to assist with other infrastructure management efforts undertaken by SAWS.

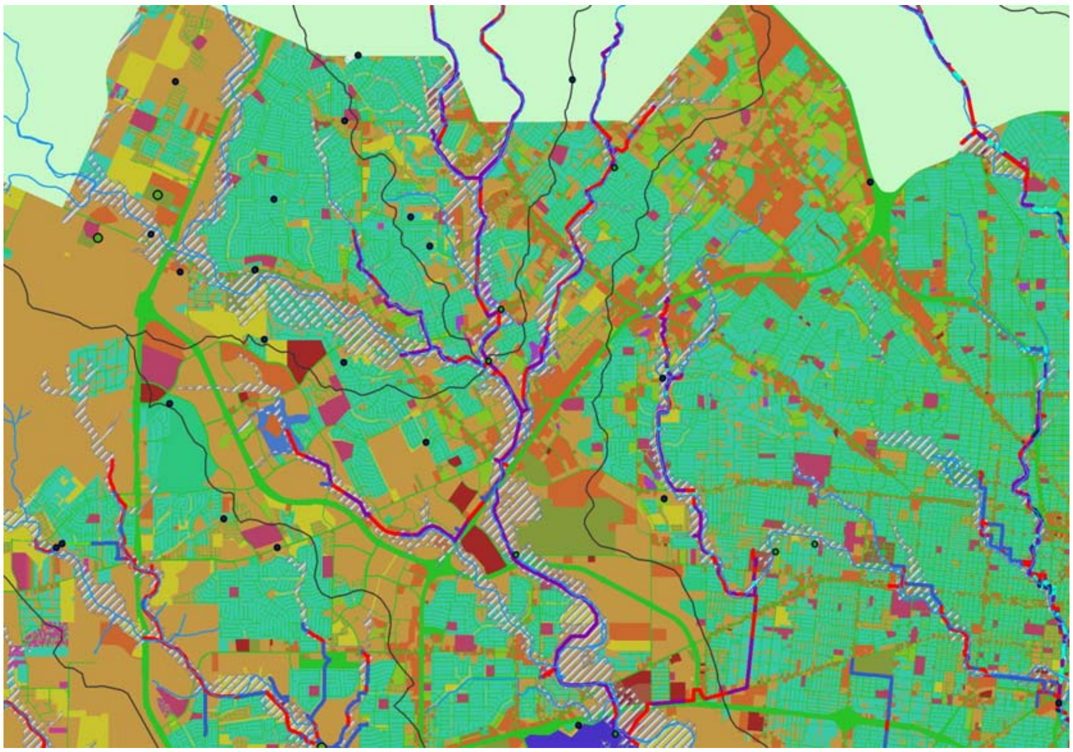

Figure 7 – Decision Model Results in GIS

The Small Diameter Pipeline Prioritization Algorithm

The Small Diameter Pipeline Prioritization Algorithm was developed as a companion to the Large Diameter Decision Tree Model. The San Antonio Water System had identified a need for a program to prioritize small diameter sewer mains for inspection based on existing information. SAWS currently owns and maintains approximately 90,000 individual small diameter sewer mains, representing over 20 million feet of pipeline. Based on the different operational and maintenance characteristics of the individual small diameter mains, they can be considered to pose a smaller risk to the public and the environment. This is due in part, to the fact that small diameter sewer mains have different characteristics regarding pipeline criticality, volume of waste carried, and deterioration. Although the failure of this portion of the collection system may pose a smaller threat to life and limb, small diameter sewer main maintenance and operations are responsible for the majority of maintenance activities and rehabilitation/renewal budgets.

During the development of the Small Diameter Sewer Main Algorithm (SDA), a ranking system was established to give priority to each of the possible outcomes or scores for each pipe segment.

Overall Cluster Score	Classification
<6.0	Class 5 – Lowest Priority
6.0-8.0	Class 4 – Low Priority
8.0-10.0	Class 3 – Medium Priority
10.0-12.0	Class 2 – High Priority
>12.0	Class 1 – Highest Priority

The criteria for selection of variables for use in the SDA was limited to ensure that the SDA could be quickly and simply utilized with a minimum amount of information. The criterion utilized in the development of the algorithm is shown below.

- SAWS Project/Job Number
- Wet to Dry Weather Peaking Factor
- Legacy Evaluation Data
- Pipe Age

The inclusion of the SAWS Project/Job Number in the criteria listing requires further explanation. Rather than examine each pipe segment in the set of sewer mains under study individually, it was determined that a clustered or grouped approach would be used. Since SAWS issues a single project/job number to groups of sewer mains that in a particular construction project, it is possible to consider all sewer mains that have a common project/job number as one unit. The clustering approach provides an important advantage to the computations considered in the Small Diameter Algorithm in that it allows all of the small diameter sewer mains to be grouped based on construction features such as where the main was constructed, how it was constructed, by whom it was constructed, the materials used in the construction, etc. This is the underlying tenet of the assumption that these clusters of sewer mains would deteriorate at approximately the same rate.

Another advantage to the clustering approach is that SAWS prefers to prepare work orders for in-house inspection teams that have geographic commonality. This gives preference to inspection of groups of sewer mains in a particular geographic area, rather than expending time and effort mobilizing from one area of the collection system to another. SAWS also utilizes this approach when using consultants and contractors to perform the inspection work. Recall that the purpose of the Small Diameter Algorithm is to prioritize the inspection of the groups of sewer lines, thus the clustering fits well within the current operational preferences of SAWS. As the SAWS Project/Job Number provides a common link between collections of sewer mains, using the clusters allows for certain manipulations of the dataset, such as:

- Setting a minimum age of a set of pipelines for consideration in the algorithm (based on the predominant age of pipes in the cluster)

- Development of geospatial areas where inspections should be conducted
- Validation of the assumption that sewer infrastructure that was constructed under the same Project/Job Number was constructed in similar regions.

An influence diagram for the Small Diameter Algorithm is shown below:

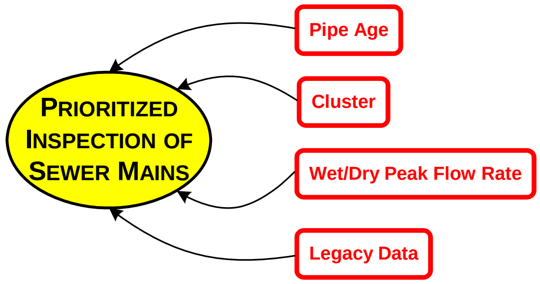

Figure 8 – Small Diameter Algorithm Influence Diagram

Another important feature of the Small Diameter Algorithm is that the entire process, from conception to completion was accomplished through the use of the SAWS Geographic Information System and existing data. Utilization of the GIS allowed SAWS and GSWW, Inc. to further analyze the results of the algorithm with other geospatial data such as proximity to floodplains, correspondence with City of San Antonio street projects, relative location to lift stations, and new development. This was accomplished in order to maximize the benefit of determining the inspection priority and the condition of the clusters of pipelines. An example of this synergy is shown below:

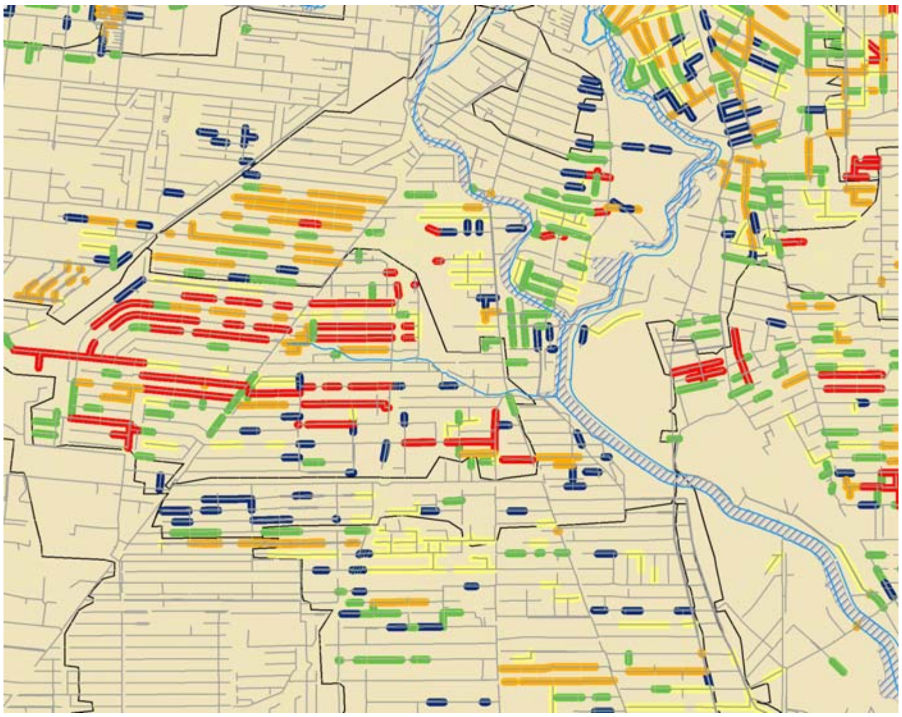

Figure 9 – Small Diameter Algorithm Pipe Clusters

The algorithm provides a normalization feature to ensure that the data for each cluster such as the differences in total pipe length contained in a particular cluster does not inappropriately slant the results from the data. The number of observed defects contained in the cluster was recorded and utilized to score each of the clusters.

The Wet to Dry Wastewater Peaking Factor data was developed during studies conducted by SAWS and is used to represent the rate of inflow and infiltration in each of the flow meter sub areas. The assumption used in the algorithm is that defective pipelines or clusters of pipelines will allow more inflow and infiltration into the collection system and thus are in need of inspection.

The Legacy Evaluation Data was taken from studies conducted by SAWS in the late 1990's. The data is used to reflect the severity of the defects observed during the studies. For example, a cluster containing one pipeline that was observed to be in need of complete replacement receives a higher score than a similar pipeline that only required a minor repair at that time. A cluster containing a pipeline that was observed to be defective during the study is assumed to be representative of the severity of defects possible in other pipelines in that cluster. The predominant age of the pipelines in the cluster is used as a multiplier in the algorithm to give preference to older pipelines.

To review, the Small Diameter Algorithm was built to prioritize small diameter sewer mains into five separate classifications for further investigation prior to renewal or rehabilitation. In order to accommodate external factors associated with conducting the prioritized inspections, the individual sewer mains were dissolved into clusters based on the SAWS project number, which is an indicator of when and how the sewer mains were constructed. Variables used in the algorithm include:

- Wet to Dry Wastewater Peaking Factors
- Legacy Evaluation Data
- Cluster Score
- Pipe Age

The equation used to organize the clusters into prioritized lists for inspection follows the form of:

[(Cluster Score) + (Peaking Factor) + (Legacy Evaluation)] * Pipe Age =
Overall Cluster Score

After population of the dataset and running the algorithm, the overall cluster scores were used to organize the clusters into the five tier matrix discussed earlier, based on the normalized scores from each cluster.

An example of the use of the algorithm is shown on the following page for the following data (actual computations are not shown):

A **Cluster** is observed to contain **15 pipeline defects**, ranging from Total Line Replacements to Mechanical Root Removal. The sum of the **SSES Main Defect Scores for these defects is 28**. The Cluster also contains 100 individual pipeline segments that share a common SAWS Job Number. The total length of pipeline in the cluster is **32,491 linear feet.** There are **8,650 linear feet** of defective pipe. The line segment is located in a drainage sub area with a **Peaking Factor of 6.4**. The pipe was originally installed in **1996** and has no recorded maintenance history. Application of the algorithm shows:

Cluster Score: 1

Peaking Factor Score: 5

SSES Main Defect Score: 3

Pipe Age Score: 1.2

Therefore, we have the final **Overall Cluster Score** computation as:

$$\{[(1) + (5) + (3)] * 1.2\} = 10.8$$

This cluster score would be assigned a **Class 1 – Highest Priority** for inspection.

Results

In both the Large Diameter Decision Tree Model and the Small Diameter Pipeline Prioritization Algorithm, the results of the modeling efforts have produced positive results and have been validated through inspection. The joint effort of SAWS and GSWW, Inc. to develop the models have resulted in a reduction in the total cost of ownership, an allocation of risk relative to large diameter interceptor failures, and provided a platform for future modeling efforts. The Large Diameter Decision Tree Model resulted in the identification of a set of large diameter interceptors for cleaning and inspection. The Large Diameter Decision Tree Model forms a part of SAWS' Best Management Practices for wastewater collection system operation and maintenance.

References and Credits

The development of both of the tools presented in this paper would not have been possible without the contributions of the authors and following persons and sources of data:

Mr. Jeff Haby, P.E. – San Antonio Water System
Mr. Bernard Dsouza – GSWW, Inc.
www.mindtools.com
Decision Support Systems for Wastewater Facilities Management – Final Report 2004
WERF - Z. Cello Vitasovic, P.E.