

HYDRAULIC MODEL CALIBRATION FOR THE GIRDWOOD,
ALASKA WATER DISTRIBUTION SYSTEM

by

Trevor Ryan Trasky

A thesis submitted in partial fulfillment
of the requirements for the degree

of

Master of Science

in

Environmental Engineering

MONTANA STATE UNIVERSITY
Bozeman, Montana

May 2008

©COPYRIGHT

by

Trevor Ryan Trasky

2008

All Rights Reserved

APPROVAL

of a thesis submitted by

Trevor Ryan Trasky

This thesis has been read by each member of the thesis committee and has been found to be satisfactory regarding content, English usage, format, citation, bibliographic style, and consistency, and is ready for submission to the Division of Graduate Education.

Dr. Anne Camper

Approved for the Department of Civil Engineering

Dr. Brett Gunnink

Approved for the Division of Graduate Education

Dr. Carl A. Fox

STATEMENT OF PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a master's degree at Montana State University, I agree that the Library shall make it available to borrowers under rules of the Library.

If I have indicated my intention to copyright this thesis by including a copyright notice page, copying is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for permission for extended quotation from or reproduction of this thesis in whole or in parts may be granted only by the copyright holder.

Trevor Ryan Trasky

May 2008

TABLE OF CONTENTS

1. INTRODUCTION	1
Statement of Problem.....	1
Flushing Program – EPA Regulation Promulgation.....	3
Need to Understand System	3
Water Quality	4
Develop Recommendations for System Calibration.....	5
State of Model	5
LIDAR - +/-5 to 40ft vs. +/- 2ft, Pump Curves.....	6
State of Calibration – “Calibrated” to Match SS SCADA.....	6
What Needs to be Done	8
2. LITERATURE REVIEW	9
What Has Been Done.....	9
Calibration Methods – Genetic Algorithms, Trial and Error etc.....	9
Data Collection – Fire Flow, SCADA, Pressure Logging	10
Calibration Standards.....	15
3. METHODS AND MEANS.....	17
Calibration Method.....	17
Genetic Algorithm Optimization Calibration.....	17
Fire Flow Test Data Collection	21
Sampling Campaign.....	22
Test Equipment.....	22
Selection of Test Nodes – Number and Location.....	24
Hydrant Functionality	25
Sampling Procedure.....	27
Data Sets - Calibration and Comparison.....	29
SCADA Data Collection	29
Data Analysis	31
Use of Genetic Algorithm Optimization Calibration.....	32
Static Accuracy Comparison vs. Fire Flow Accuracy Comparison	37
High Loss vs. Low Loss Data	38
Model Calibration Accuracy vs. Number of Data Points	38
Pipe Groups: Many vs. Few	39
Data Collection Location	39

TABLE OF CONTENTS - CONTINUED

4. RESULTS	41
Data Sets - Calibration and Comparison.....	41
Three Pressure Zones vs. Four	41
Calibration Accuracy Evaluation (Static vs FF).....	42
Good vs. Useless Data (High Loss vs. Low Loss)	45
Duplicate Data vs. Unique Data	46
Accuracy vs. # of Data Points	47
All Pipe Group Set.....	48
Limited Pipe Group Set.....	51
Comparison	53
Data Collection Location	54
Summary	54
5. DISCUSSION	55
Sampling Campaign Conclusions.....	55
Static Accuracy Check vs. FF Check	55
Good Data vs. Useless Data	57
Duplicate Data vs. Unique Data	59
Effect of the Number of Good Data Points on Accuracy	60
Effect of Number of Pipe Groups.....	61
Optimal Sample Amount?.....	63
Data Collection Location	64
Cost per Sample.....	66
Lessons Learned – Future Research Needs.....	67
6. CONCLUSIONS AND RECOMMENDATIONS.....	68
REFERENCES.....	70

LIST OF TABLES

Table	Page
2.1 ECAC Calibration Guidelines	16
3.1 Statistic Calculation Procedures	36
3.2 Analysis Descriptions and Details	37
4.1 Number of Zones Represented vs. Unique Data Points Used	48

LIST OF FIGURES

Figure	Page
1.1 Map of Distribution System	2
1.2 Model Representation of Booster Station	7
1.3 SCADA Representation of Booster Station	7
3.1 Pipe Roughness Groups.....	18
3.2 Steady State Calibrator.....	19
3.3 Fire Flow Calibrator.....	20
3.4 Extended Period Simulation Calibrator.....	20
3.5 Calibrator Results.....	21
3.6 Hydrant Static Pressure Gauge	23
3.7 Dechlorinating Diffuser with Pitot Gauge Attached.....	23
3.8 Partial Map of Flow and Residual Hydrants	25
3.9 SCADA Interface for Booster Station.....	30
3.10 Proficy Portal Data Logging Interface	31
3.11 Model Verification Roughness Values	34
3.12 GA Calibrator Model Verification Run Process.....	35
3.13 Model Verification Test Comparisons	35
4.1 Mean Squared Error	43
4.2 Standard Deviation of Squared Error.....	43
4.3 Average Percent Difference of Observed and Simulated Values	44
4.4 Mean Squared Error for Static Evaluation	45

LIST OF FIGURES - CONTINUED

Figure	Page
4.5 Mean Squared Error vs. Number of Unique Data Points	49
4.6 Mean Squared Error vs. Number of Unique Data Points	50
4.7 Mean Squared Error vs. Number of Unique Data Points	51
4.8 Mean Squared Error vs. Number of Unique Data Points	52
4.9 Mean Squared Error vs. Number of Unique Data Points for Both Sets.....	53
5.1 Mean Squared Error vs. Number of Unique Data Points for Both Sets.....	62

AC – Asbestos Cement Pipe

AWWU – Anchorage Water and Wastewater Utility

C-Value – Hazen-Williams Coefficient of Friction

CI – Cast Iron Pipe

DI – Ductile Iron Pipe

EPA – United States Environmental Protection Agency

EPS – Extended Period Simulation

GA – Genetic Algorithm

GALV – Galvanized Iron Pipe

GIS – Geospatial Information System

GPM – Gallons per Minute

LIDAR – Light Detection and Ranging

MSE – Mean Squared Error

PRV – Pressure Regulating/Reducing Valve

PSI – Pounds per Square Inch

PVC – Polyvinylchloride Pipe

SCADA – Supervisory Control and Data Acquisition

SE – Squared Error

ABSTRACT

The possible EPA promulgation of regulations requiring flushing programs for water distribution systems to regulate water age is encouraging many utilities to develop a better understanding of their systems. This usually involves the development of a hydraulic model. For the hydraulic model to be of use it must be calibrated using collected field data. Few established guidelines exist for utilities to perform such data collection and calibration. There are many different types of data that can be collected using many different methods. There are also various model calibration methods that can be used. This study sought to develop an optimized sampling plan and calibration method for a small utility that defines the best practice for data type, location, quantity, and collection conditions. A large quantity of many data types were collected at various locations in the system. Different methods were used to calibrate the model with various data sets and the model accuracy was evaluated using a second independent data set. The effect of model input parameter accuracy and pipe grouping during calibration on overall accuracy was also investigated. Data type, location, quantity, and collection conditions had an impact on calibration accuracy. High headloss data from fire flow tests provided better calibrations than low loss data from static pressures. Carefully selected test locations resulted in more efficient calibration than evenly distributed test locations. There was found to be a point of diminishing returns when investigating the amount of data used in calibration versus calibration accuracy. Uncalibrated model input parameters such as elevations and pump characteristics also had a significant impact on model calibration accuracy.

INTRODUCTION

Statement of Problem

The drinking water distribution system that was the focus of this study is located in Girdwood Alaska. Girdwood is a resort community and has approximately 2000 full time residents and many seasonal residents. Girdwood is considered a suburb of Anchorage and the water and wastewater utilities are maintained by the Anchorage Water and Wastewater Utility (AWWU). The system was originally constructed by a private corporation that owned and operated the ski resort. The AWWU took over responsibility of the system in the 1970's.

The water source for the distribution system is a well. The water is chlorinated and pumped to the single reservoir in the system. Downstream of the reservoir there are 6 pressure zones, 4 of which contain fire hydrants. There are 2 booster stations and 3 pressure reducing valve (PRV) vaults. Supervisory control and data acquisition (SCADA) infrastructure is present in the system to log pressures, flows, and chlorine residuals at the well, reservoir, and select PRVs and booster stations. The system consists of 5 pipe types, polyvinylchloride (PVC), cast iron (CI), ductile iron (DI), asbestos cement (AC), and galvanized iron (GALV), of various ages from new to approximately 45 years old. There are approximately 450 pipes and 95 hydrants in the system. A map of the distribution system is shown in Figure 1.1.

The system information available was developed from record drawings of construction and is contained in the AWWU's GIS database. A hydraulic model of the Anchorage and Girdwood system was developed from GIS data by a consultant in 2006 using the program InfoWater[®] by MWHSoft. InfoWater[®] is a hydraulic and water

quality modeling program similar to publicly available EPANet (www.epanet.com) that runs in ArcGIS. The elevations for model components were obtained from a light detection and ranging (LIDAR) database developed for the Anchorage and Girdwood area by a consultant. The hydraulic model developed for the utility would be considered a full pipe model as it contains all infrastructure in the system.

Flushing Program – EPA Regulation Promulgation

As new EPA regulations are proposed and promulgated, utilities must constantly upgrade facilities and practices to be in compliance. The proposed EPA mandates of flushing programs to maintain water quality in distribution systems were the impetus for this study. One goal of a flushing program is to remove water from the system that has a chlorine residual that has degraded to a value lower than the minimum allowable concentration. If it can be demonstrated that such level of degradation does not occur in a system then the expense of developing and maintaining a flushing program for this purpose need not be incurred. It is believed by the AWWU that due to the source water quality and temperature the water quality in the Girdwood system is such that a flushing program to maintain proper chlorine residual concentration will not be required. It is in the best interest of the utility to confirm this prior to promulgation of an EPA mandated flushing program.

Need to Understand System

To determine if the Girdwood distribution system may require this type of flushing program, the system characteristics need to be more thoroughly understood.

One method of achieving an understanding of the system is to collect many water quality samples at every point of interest. Another method would involve the use of a predictive model to obtain the same information dynamically. A hydraulic model of the system such as the one developed for the utility can be used to simulate conditions in the system such as pressures, flows, and water quality parameters such as chlorine residual. To have any confidence in the model's ability to accurately predict these parameters the model must be compared to observed system values and calibrated.

Water Quality

Water quality in the distribution system depends on factors such as source water quality, temperature, residence time, system integrity, and pipe characteristics. The chlorine residual is an indicator of water quality as its presence confirms proper disinfection of the water. Chlorine residual can be determined from grab samples or from an automated analyzer in a static location. Source water quality and temperature can be determined through routine sampling, however, residence time and pipe characteristics are only obtainable by using a dedicated field study or a model. Chlorine residual is currently monitored in the system using a dedicated analyzer at the point of treatment and routine grab sampling from dedicated sampling taps and select residences throughout the system. It is often difficult to obtain chlorine residuals for potential problem areas such as dead end pipes or other low flow areas using such a method, as access to the sampling locations can be limited. To predict chlorine concentrations where grab samples cannot be taken or for proposed system additions, a calibrated water quality model is ideal.

Develop Recommendations for System Calibration

Girdwood is ideal for study as it contains all features of a larger distribution system, while being isolated and having a smaller overall size. The information gained from calibrating the Girdwood system would be beneficial in developing and running a future sampling campaign as well as determining the requirements for obtaining an accurate calibration. Few guidelines exist for utilities attempting to perform calibration of a hydraulic model. The type, quantity, method of collection, and location of data used for calibration varies between different references. The hydraulic model of a system must be accurately calibrated for the water quality model to be useful. One of the goals of this study was to perform a calibration of the Girdwood hydraulic model of sufficient accuracy to be useful in predicting water quality in the system and apply the methods used to the sampling campaign and calibration of other distribution systems.

State of Model

The model parameters for all infrastructure were input into the hydraulic model from the database maintained by the GIS department of AWWU. This includes all model elevations, pipe sizes, materials, ages, lengths, pipe and node locations, PRV sizes and configurations, and pump sizes and configurations. Some pump curves were unknown and input as design head and flow. Demands for all nodes were developed from a previous study by a consultant.

Prior to the calibration exercise, the model accurately predicted the static pressures at the upstream and downstream points of the PRV's in the system. The uncalibrated model was unable to accurately match fire flow tests performed previously to check the model accuracy.

LIDAR - +/-5 to 40ft vs. +/- 2ft, Pump Curves

The model elevations were input using a LIDAR database developed for the Anchorage area. The elevations contained in this database, as discussed further in the methods chapter, are inaccurate in some places, with errors from 5 feet to 40 feet. The AWWU later provided a revised LIDAR database for model correction.

The pump curves in the model are developed from design head and flow points. This does not provide an ideal pump curve with respect to model accuracy.

State of Calibration – “Calibrated” to Match SS SCADA

During initial development the model was calibrated to match the steady state SCADA values recorded by the utility. The PRV elevations were then adjusted such that the upstream pressures matched those available from the logged SCADA data. The reservoir elevation was known. One of the booster stations was modeled as 3 pumps in parallel with a PRV in series in order to match steady state values for the pressure zone downstream (Figure 1.2). However, the physical configuration of the booster station as shown in the utilities SCADA interface consists of 2 pumps in parallel with 2 PRVs in parallel (Figure 1.3). The model configuration allows the model to predict the static pressure in the pressure zone served by the booster station, however, this configuration

does not allow the accurate prediction of fire flows. This is a problem as fire flow prediction and analysis are intended as primary uses for this model.

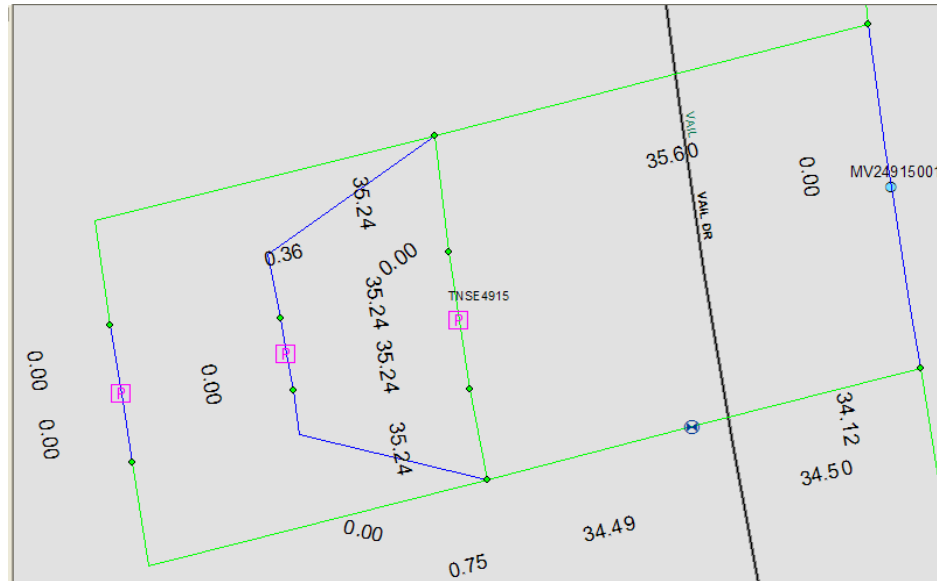


Figure 1.2 Model Representation of Booster Station

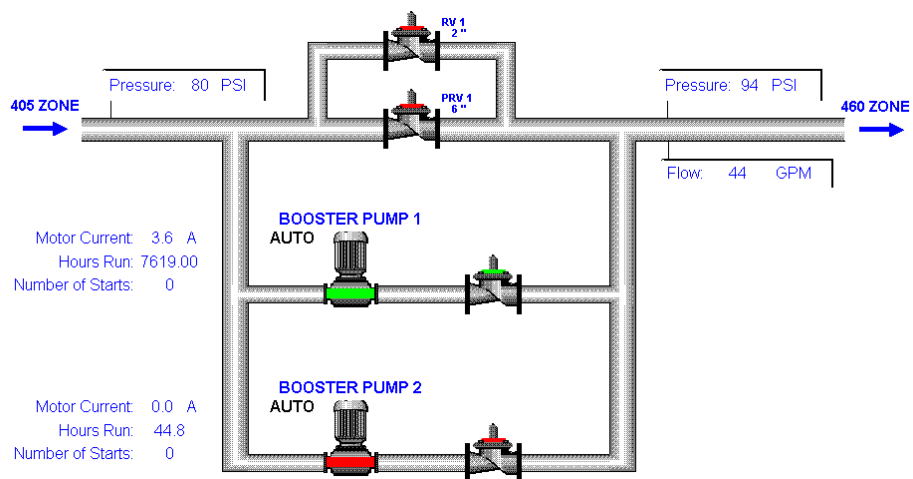


Figure 1.3 SCADA Representation of Booster Station

What Needs to be Done

The goal of this study was to calibrate the hydraulic model in preparation for the calibration of the water quality model.

A calibration method must be selected that will allow for the accuracy of calibration required for the desired model use. For this study, it is necessary that whatever method used for calibration be repeatable such that a unique set of data provides a unique set of corresponding calibration parameters. This is necessary for the method to be appropriate for comparing accuracy of multiple calibrations.

To calibrate the hydraulic model of Girdwood and provide recommendations concerning quality and quantity of data versus calibration accuracy, a sampling campaign must be developed to collect field data to be used in the model. Sufficient data must be collected to allow for an accurate calibration as well as analysis of different quantities and types of data at different locations in the system. The sampling campaign must also take into account the protection of AWWU infrastructure as well as private and public property. Customer service must also be an important aspect of the sampling campaign as the field work will occur in close proximity to customer property and the possibility of water quality complaints exists.

LITERATURE REVIEW

What Has Been Done

There has been some research done in the hydraulic model calibration arena. Different methods of calibration have been evaluated along with data collection methods, types, locations, and quantities. The goal of this study was to evaluate the previously performed work in the area of interest and adapt the results to develop a sampling campaign and calibrate a small scale water distribution system model.

Calibration Methods – Genetic Algorithms, Trial and Error etc

Lansey et al (2001) developed a three step procedure to calibrate a hydraulic model. Field data was collected and screened, and the model's unknown parameters of pipe roughness coefficients were estimated using the screened data. A calibration assessment was then performed by comparing data not used for parameter estimation with the model's results. The results were compared to predetermined assessment criteria. If the model was determined to be inadequate for its designed purpose, a data collection scenario and desired system conditions were developed to provide calibration data that would result in an adequate model. After these data were collected the parameter estimation and assessment process continued. The importance of obtaining quality data was emphasized due to the effort and cost of collection.

Walski (1983) discussed techniques for calibrating a water distribution model by determining pipe roughness and water use. The author does not consider the model to be

calibrated until it can predict, within reasonable agreement, system parameters for a variety of conditions. Both pipe roughness and water demand are factors that can be adjusted in calibration. The author investigates which of the two to adjust and by how much by observing system pressures for different use rates. Fire flow tests are used to achieve the varying use rates in the system.

Ormsbee and Lingireddy (1997) describe a 7 step process for model calibration, including: identifying model use, determining parameter estimates, collection of calibration data, evaluation of model results, macro calibration, sensitivity analysis, and micro calibration. Pipe grouping for roughness calibration was based on pipe type, age, and size. The optimization approach for calibration was compared to the analytical and simulation approaches based on skeletonizing of the system by removing pipes.

Data Collection – Fire Flow, SCADA, Pressure Logging

Walski (2000) studied the effect of field data quality on comparisons between field data and model results. The concept of 3 different qualities of data was presented: good data, bad data, and useless data. Good data was defined as that which is collected when sufficient headloss exists in the system such that the headloss is significantly higher than the potential errors in measurements. Bad data was defined as data containing errors in measurement or lack of good information concerning elevations and boundary conditions. Useless data was defined as data collected when the headloss in the system is similar in magnitude to the potential error in measurements. For a normal day the system demands in a distribution system are generally low enough that there is minimal headloss

present, which results in the collection of useless data. Useless data cannot be used for calibration but can be used to determine the accuracy of boundary conditions in the system including PRV settings, pump curves, and reservoir water levels. When useless data was used by the author for calibration in a particular model the model appeared to be calibrated for C-values ranging from 40 to 120. A model calibrated in such a fashion can predict static pressure satisfactorily; however, when fire flow is predicted the model will likely fail to provide accurate data. To obtain good data, the author attempts to collect data with head loss five times that of the potential error in the measurements by performing fire flow tests. Pressure measurements should be taken as far from the boundary condition as possible in order to maximize the headloss measured. System boundary conditions including pump head, PRV settings, and tank levels must be known during testing as well as any high demand events not part of the testing. Accurate equipment should be used for collection in order to maximize data quality. Gauges should be calibrated and accurate to 1psi. As-found calibrations should also be performed following data collection in order to correct measurements later if necessary. Elevation data should be accurate to 2 ft. According to the author, optimization programs that weight useless and good data equally will produce misleading results, therefore it would be better to use 8 good data points rather than 8 good data points and 6 useless data points. UK standards include references to percentage differences between observed static and predicted model static pressures. The author demonstrated in this paper that this is misleading as the observed static pressures can be very close to predicted pressures when the model is poorly calibrated.

de Schatzen (2000) in the UK used genetic algorithms to calibrate a hydraulic model using data from pressure loggers. The average absolute error of the model was compared to the placement density of the pressure loggers. They showed that the average absolute error in the model decreased as more data loggers were used.

Walski (1986) illustrated that it is possible to accurately calibrate a hydraulic model if a sufficient amount of accurate data is collected. The model was calibrated over a range of demands and patterns for known boundary heads. This ensured that the model calibration yielded the water use and pipe roughness present during collection. The determination of model calibration accuracy is discussed, including methods of comparison and desired accuracy.

Walski et al (2000) detailed the frequently encountered problems associated with calibrating an extended period simulation (EPS) model using SCADA data, specifically relating to the diurnal curve. SCADA data has time stamps associated with each data point. It is important to know whether this time stamp corresponds to the time the data was measured at the sensor or the time at which it was logged. There can be a significant difference in these two times if the polling interval is large. SCADA can provide both average and sampled data, where the average is the average value of the parameter since the last polling, and sampled is the value of the parameter at the exact time of polling. Walski et al (2000) state that the usual way to verify a model is to use it to predict the behavior of another set of conditions, such as calibrating the model for one day and verifying it for a different day.

Howie (1999) discussed the potential problems associated with the use of SCADA data in calibrating a hydraulic model. Field data was collected using pressure loggers and flow meters and used to calibrate a model. Three points in time were used, one for calibration and two for verification. It was shown that SCADA data cannot be taken without review and can be internally inconsistent and different from the simulation results due to the inaccuracy of meters and gauges. It was discovered that when using a large number of simultaneous data points it is much more difficult to determine if one set of system conditions was better than another than when using a small number of field measurements occurring at different times. When a change to pipe roughness or demand allocation was made to improve the match at a certain point, it was more likely to worsen the match at another point when more data points were used.

Kapelan et al (2003, 2005) developed methodologies to find the optimal set of network locations for pressure loggers to collect data for use in the calibration of a hydraulic model. All parameters of the model were assumed to be accurate, with the exception of the pipe roughnesses, with the only error resulting from imprecise measurements. Genetic algorithms (GA) were used to find the optimal locations for the pressure loggers in order to minimize cost and calibrated model uncertainty. The locations selected by the GA were compared to locations selected by an expert and shown by Kapelan et al (2005) to be superior. Kapelan et al (2003) determined that the best locations for data collection in a network are those where pressures are sensitive to changes in the calibration parameters.

Meier and Barkdoll (2000) used genetic algorithms to optimize sampling design, specifically concerning data collection locations, and applied it to the calibration of a network model for a small town in Ohio. Genetic algorithms were used to select collection location as the quality of the calibration data otherwise depends on the experience of the engineer. If the sampled locations yield insufficient data for an accurate calibration, more data must be collected at great expense or a less accurate model must be used. The goal of the GA was to maximize the number of pipes in the system with non-negligible velocities by flowing a limited number of hydrants in certain locations. In general, when the flowed hydrants were on the extreme perimeter of the system, the majority of the pipes in the network had non-negligible velocities. Non-negligible velocities are required to accurately determine pipe characteristics.

de Schaetzen et al (2000) used shortest-path and genetic algorithms to identify the optimal set of monitoring points. The shortest path from the source or boundary condition to every other node in the network is determined and the nodes were ranked by their distance from the source. The node that is furthest away is selected as a monitoring point location. The nodes furthest away from the boundary conditions provide the most information on the system. The locations selected using all methods were compared to locations obtained from an expert's choice. The algorithms selected more locations on small diameter pipes than the expert and also covered a higher percentage of the system. It is likely the selection of a higher number of small pipes is due to the general decrease in pipe size towards the perimeter of a pressure zone.

Calibration Standards

Walski (1995) discussed the standards for model calibration. A simple rule for determining whether a model is “calibrated” or not does not exist and it is not possible to consider a model calibrated in general. According to the author, the intended use of the model should be determined and the accuracy of the model evaluated on the basis of its ability to predict values for its intended use. The individual using the model must decide if the model is sensitive enough to support the decisions to be made. The standard for model calibration should be whether or not the responsible individual can base decisions on the model’s results. The model calibration process is never complete as additional calibration is required as the model is applied to new problems. As there may be a significant amount of time between the initial calibration and subsequent calibrations, the initial calibration must be well documented for reference by others. During hydrant flow tests many residual pressure gauges should be distributed throughout the system to provide several points for data collection. The author analyzed a system in which the data collected during low flow made the model appear calibrated for any C-value used. The model would always work regardless of C-value if a model user were to compare only the observed and predicted pressures during normal flow.

The Water Research Centre in the UK (1989) established standard criteria for the calibration of water distribution models. The type and amount of equipment required for calibration is organized by number of users served by the system. Guidelines established for calibrated model accuracy include the evaluation of the model’s ability to predict flows, static pressures, and reservoir levels.

The AWWA engineering computer applications committee (ECAC) (1999) developed calibration guidelines for water distribution system modeling. Possible sources of error in model calibration are discussed including pipe roughnesses, system demands, system information such as elevations and pump curves, time, and measuring equipment. A table of minimum calibration criteria is presented for the purpose of stimulating the adoption of standard calibration criteria for the US (Table 2.1).

Table 2.1 ECAC Calibration Guidelines

Intended Use	Level of Detail	Type of Time Simulation	Number of Pressure Readings	Accuracy of Pressure Readings	Number of Flow Readings	Accuracy of Flow Readings
Long Range Planning	Low	Steady-State or EPS	10% of Nodes	+/- 5 psi for 100% of Readings	1% of Pipes	+/- 10%
Design	Moderate to High	Steady-State or EPS	5% - 2% of Nodes	+/- 2 psi for 90% of Readings	3% of Pipes	+/- 5%
Operations	Low to High	Steady-State or EPS	10% - 2% of Nodes	+/- 2 psi for 90% of Readings	2% of Pipes	+/- 5%
Water Quality	High	EPS	2% of Nodes	+/- 3 psi for 70% of Readings	5% of Pipes	+/- 2%

METHODS AND MEANS

Calibration Method

A method of calibration performable by a reasonably well equipped utility was to be developed for use in future calibrations. The process was designed to use existing tools that were available to the utility. This calibration method assumes all data in the model is accurate with the exception of pipe roughness values. The pipe roughness values are then adjusted such that the model reflects observed system parameters.

Genetic Algorithm Optimization Calibration

A GA optimization calibrator was selected as the best method for calibration based on literature review (de Schaetzen 2000). The model developed for the AWWU utilized modeling software that contains its own GA optimization calibrator. It was decided to use this calibrator rather than develop one so that the utility would be able to repeat the calibration process in the future using software already purchased and under service from the developer.

The GA calibrator in InfoWater® has three different methods for calibration; steady state, fire flow, and extended period. For all methods the pipes are divided into groups with similar roughness characteristics (Figure 3.1).

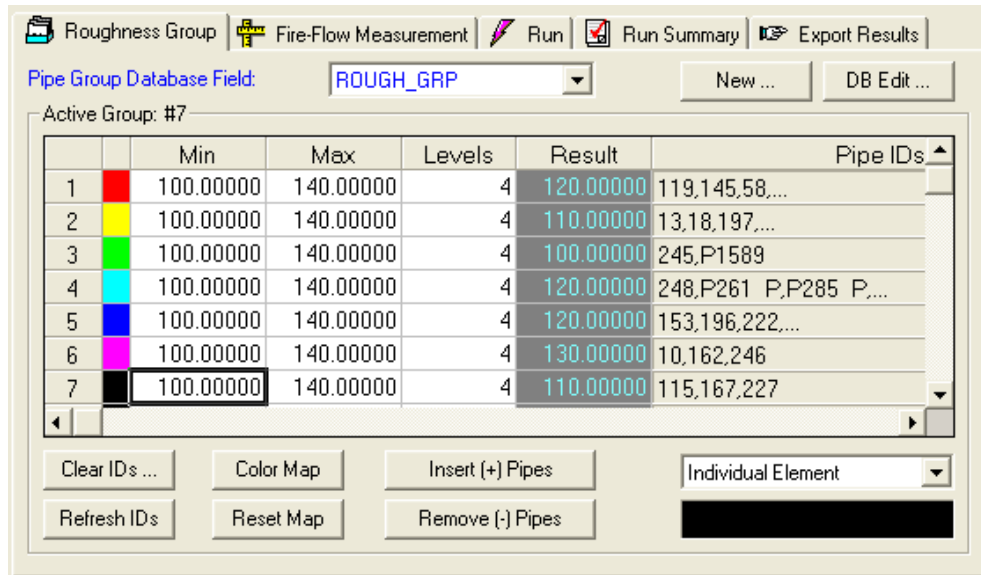


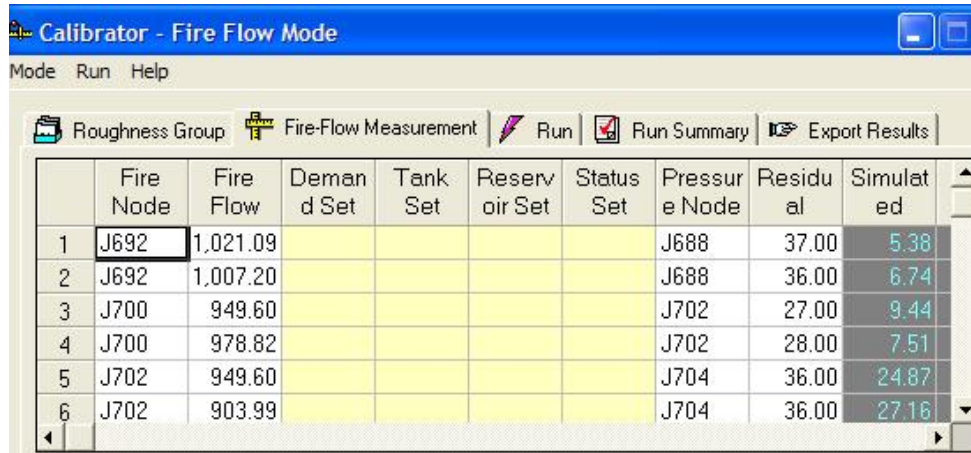
Figure 3.1 Pipe Roughness Groups

Each group is given a minimum and maximum roughness value and a number of levels to divide the range between the minimum and maximum. For example, if the range between the minimum and maximum is 40 and 4 levels are used, the model will adjust the pipe roughnesses in increments of 10. The steady state calibrator assumes that the system demands are constant. Static pressures and pipe flow measurements are used. Roughness values for the pipes are calculated such that the model's simulated flows and pressures match the observed values at corresponding pipes or nodes (Figure 3.2).

	Junction ID	Observed	Simulated	% Diff
1	47	51.00	52.41	2.76
2	38_J	43.00	41.83	-2.73
3	83_J	56.00	56.94	1.67
4	J600	69.00	70.47	2.13
5	J608	50.00	53.66	7.33
6	J612	80.00	84.42	5.52
7	J622	64.00	66.07	3.24
8	J624	65.00	68.61	5.55
9	J630	70.00	72.13	3.05
10	J634	58.00	62.69	8.09
11	J646	64.00	66.87	4.49
12	J660	63.00	66.60	5.71
13	J666	68.00	71.70	5.44
14	J674	48.00	47.14	-1.76

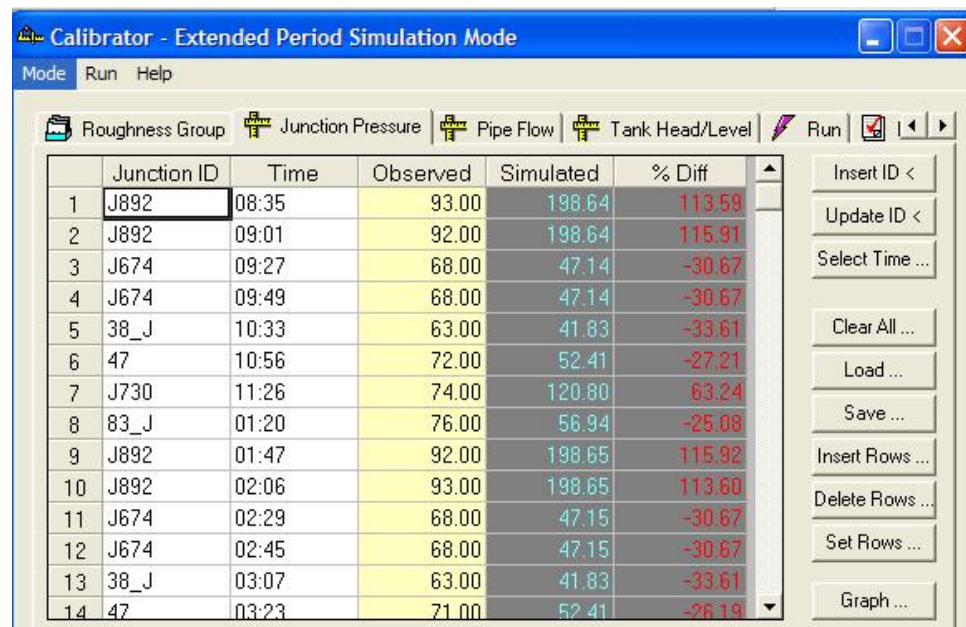
Figure 3.2 Steady State Calibrator

The fire-flow calibrator (Figure 3.3) allows the use of fire-flow tests for calibration. The roughness values for the pipes are adjusted to match the simulated residual pressure at a hydrant with the observed residual pressure for a known flow rate. The extended period simulator (Figure 3.4) uses the input diurnal curve for demands and adjusts pipe roughness to match simulated junction pressures, pipe flows, and tank levels with observed values at given times.



	Fire Node	Fire Flow	Demand Set	Tank Set	Reservoir Set	Status Set	Pressure Node	Residual	Simulated
1	J692	1,021.09					J688	37.00	5.38
2	J692	1,007.20					J688	36.00	6.74
3	J700	949.60					J702	27.00	9.44
4	J700	978.82					J702	28.00	7.51
5	J702	949.60					J704	36.00	24.87
6	J702	903.99					J704	36.00	27.16

Figure 3.3 Fire Flow Calibrator



	Junction ID	Time	Observed	Simulated	% Diff
1	J892	08:35	93.00	198.64	113.59
2	J892	09:01	92.00	198.64	115.91
3	J674	09:27	68.00	47.14	-30.67
4	J674	09:49	68.00	47.14	-30.67
5	38_J	10:33	63.00	41.83	-33.61
6	47	10:56	72.00	52.41	-27.21
7	J730	11:26	74.00	120.80	63.24
8	83_J	01:20	76.00	56.94	-25.08
9	J892	01:47	92.00	198.65	115.92
10	J892	02:06	93.00	198.65	113.60
11	J674	02:29	68.00	47.15	-30.67
12	J674	02:45	68.00	47.15	-30.67
13	38_J	03:07	63.00	41.83	-33.61
14	47	03:23	71.00	52.41	-26.19

Figure 3.4 Extended Period Simulation Calibrator

The GA calibrator (Figure 3.5) finds the roughness value for each pipe group that provides the smallest MSE between simulated and observed values. The resulting roughness values can then be applied in the model to complete the calibration.

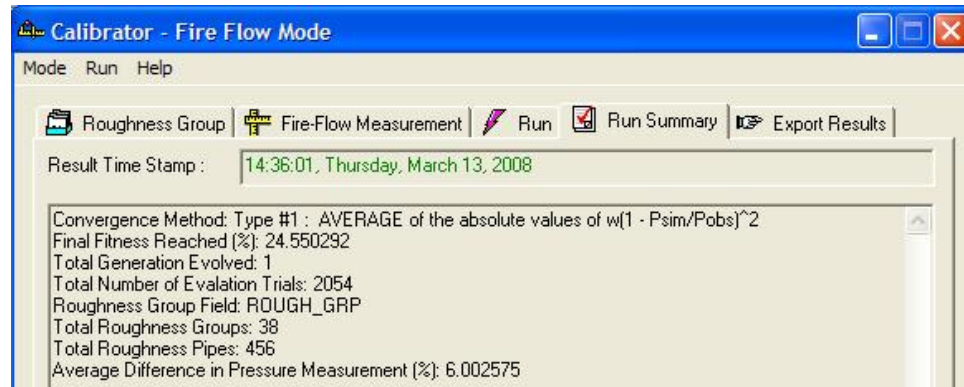


Figure 3.5 Calibrator Results

It was decided that the most efficient way to collect data for all three calibration methods was to perform fire flow tests with an accurate accounting of timing, using the measured static pressures obtained prior to beginning hydrant flow, and also logging the SCADA data for the system. The goal was to develop an extended period model simulation with demands at each node corresponding to the flow rate measured and the start time and duration of the fire flow tests. In this fashion the two most desirable methods of calibration could be combined, allowing the high headlosses from the fire flow tests to obtain accurate pipe roughnesses and the time relationships to develop accurate tank levels and reservoir usages.

Fire Flow Test Data Collection

Fire-flow tests were chosen as a primary method of data collection. From the literature, data collected using fire-flow tests appeared superior to data collected using static tests due to the higher head loss induced in the system from the high flow rates (Walski 2000). The principles of data precision dictate that the higher the values measured, or differences in values measured, the more accurate the measurements are

likely to be. An important reason to use fire-flow tests is to obtain measurements and differences that are higher in magnitude than the error in the measuring instruments. For example, measuring a 2 psi drop with a gauge accurate to 1 psi would likely have a higher percent error measurement than measuring a 20 psi drop using the same gauge.

Sampling Campaign

The requirements for the type of data collected were determined in part by the GA calibrator interface within the model. The data required to use the GA calibrator, including fire-flow tests and static pressures, was collected along with other data, such as logged SCADA data, that would allow for further analysis and verification of the model.

Test Equipment

The test equipment used was already available to the utility. The hydrant gauges were assembled from the AWWU warehouse stock on hand and included a 200 psi Ashcroft® glycerin filled gauge (Figure 3.6). The hydrant gauges were tested to determine their accuracy. The gauges tested accurate within 1 psi. The diffuser was a Pollard Water LPD-250 dechlorinating diffuser that had been previously used for hydrant testing. The associated dechlorinating tablets were also on hand from previous testing. The dechlorinating tablets were composed of sodium sulfite and were intended for this specific diffuser. A pitot gauge was purchased that would attach directly to the diffuser to allow for a higher precision of data collection and the use of a smaller crew size (Figure 3.7).



Figure 2.6 Hydrant Static Pressure Gauge



Figure 3.7 Dechlorinating Diffuser with Pitot Gauge Attached

Selection of Test Nodes – Number and Location

There did not appear to be many accepted guidelines for fire flow test locations and quantities other than rules of thumb. One such rule given to the AWWU by a consultant in preparation for collection in the Anchorage system was to take fire flow tests at 10% of the hydrants in each pressure zone to provide sufficient data for a water quality level calibration. Guidelines for steady state tests developed by the Engineering Computer Applications Committee (ECAC) in 1999 suggest using pressure recorders at 10% of nodes and flow recorders at 5% of pipes. With respect to test location there is some content in the literature about maximizing head loss by selecting points further from boundary conditions such as reservoirs, pumps, and PRVs (Walski 2000).

The goal for this study was to perform fire-flow tests on about 40% of the hydrants in each of the pressure zones. This quantity of data was selected as it was much higher than the rule of thumb for a good calibration and may allow for the determination of an optimal number of tests for calibration. Some guidelines suggest taking fire-flow test data distant from boundary conditions to maximize the headloss between the two points (Walski 2000). The hydrants to be flowed were selected to be evenly distributed throughout the pressure zone to determine optimal sampling locations (Figure 3.8). Residual hydrants were selected as the adjacent upstream hydrant where possible.



Figure 3.8 Partial Map of Flow and Residual Hydrants

Hydrant Functionality

During a fire flow test a high volume of water is being emitted from the hydrant, in excess of 1000 gpm at times. In an area similar to Girdwood that does not have a storm sewer system, the inability to control this volume of water will likely result in damage to property. This can easily occur in the event the hydrant is damaged or becomes damaged once flow is initiated. Each hydrant leg has its own dedicated gate

valve, referred to as a hydrant auxiliary valve, for the purpose of restricting the flow to the hydrant in such an event. However, Girdwood consists mostly of dirt roads and due to the area's snow plowing requirements all valve boxes must be 10 inches below the road surface. To access this valve for manipulation, the valve box must first be located with a metal detector and the road material removed. This process is not feasible in the amount of time necessary to avoid property damage should a hydrant failure occur, requiring the process to be completed prior to beginning the fire flow test.

The AWWU standard fire flow procedure involves having a valve key on the auxiliary valve of the flowed hydrant during testing to expediently shut off flow to the hydrant and minimize the possibility of property damage should the hydrant become inoperable. It was decided that the auxiliary valves for each of the hydrants to be flowed should be exposed and checked for operability prior to testing so this could be performed. Many of the valves were inoperable due to sediment and mud accumulating in the valve box or not locatable using a standard metal detector. This did not prove to be an issue as the functionality of the hydrants was confirmed as explained below.

When flowing hydrants, the Girdwood treatment operators use a fabricated diffuser with a 3 inch gate valve to regulate flow during hydrant operation and to isolate flow in the event of a hydrant failure. This eliminates the need to have the hydrant auxiliary valve keyed for every test. This diffuser was not deemed appropriate for the study as the pitot gauge used to measure flow would not be compatible with its design. It was decided that the best course of action to both minimize the possibility of property damage due to an uncontrollably flowing hydrant and customer complaints concerning

sediment in the water stirred up by the increased pipe flows was to first perform a complete flush of the system through the hydrants. This would remove the accumulated sediment from the pipes and determine which hydrants were inoperable, eliminating the need to have each valve keyed during hydrant flow.

This flushing exercise was implemented and completed in a few days, occurring the weekend before fire flow testing. The Girdwood crew did find a few hydrants that were undesirable for fire flow operation.

Sampling Procedure

The fire flow test sampling procedure was adapted from the procedure that was designed for use in the Anchorage system. As the testing occurred during peak demand during the summer it was desired that the reservoir be completely full at the beginning of testing each day. This was accomplished by running the well pumps in the morning until the reservoir was full. A few modifications to the standard fire flow test procedure were required for use in the Girdwood system. As there is no storm sewer in Girdwood and 3 flowing bodies of water in the vicinity of the township that are inhabited by spawning salmon and other fish, the discharge of the chlorinated water would be unacceptable. The utility had access to a commercially available dechlorinating diffuser that would deactivate the free chlorine in the water discharged from the hydrant during testing. A single individual was able to read the pitot gauge and operate the flowed hydrant as the pitot gauge for flow measurement could be mounted to the diffuser for hands free use. It was decided that a Girdwood treatment operator would operate the flowed hydrant and a

technician and intern would perform the manipulation of the residual hydrant and the data collection. The volume of water discharged in the initially desired full-open hydrant flow time of 5 minutes unnecessarily increased the likelihood of property damage.

Preliminary testing showed that the pitot and hydrant gauge pressure readings were no different after a full open flow of 5 minutes than after 2 minutes. To significantly decrease the amount of water discharged, 2 minutes of full open flow time was used.

With the intention of facilitating the incorporation of the fire flow tests into an extended period model if desired, it was decided that each flow hydrant should be tested at least twice and the times recorded. When possible all tests in a set for a particular pressure zone were collected in a single day for the data to be consistent should there be the need to calibrate a zone separately.

The performance of a fire flow test was as follows:

The identification number on the pitot gauge and hydrant gauge to be used for the day would be recorded for data comparison or adjustment if necessary at a later time. The crew would coordinate using 2-way radios which hydrant would be flowed and which would be residual for the test. The testing crew would travel to the flow hydrant. The intended path of the discharged water would be assessed and physical measures to protect property were utilized if necessary. The technician and intern would then travel to the residual hydrant. The residual hydrant would have a pressure gauge attached and the static pressure read and the time recorded. The hydrant to be flowed would have the dechlorinating diffuser attached. When the diffuser was attached the operator at the flow hydrant would inform the individuals at the residual hydrant by radio. If all individuals

were ready the flow hydrant would be slowly opened and the time recorded. When the hydrant was fully opened the operator at the flow hydrant would inform the other individuals and the time would be recorded. The hydrant would be run full open for at least 2 minutes to let the residual pressure and pitot pressure stabilize. Residual pressure and pitot pressure would be measured and the time recorded. At this point the flow hydrant would be slowly closed. The time at which the hydrant was fully closed would then be recorded. The crew would then continue to the next set of hydrants.

Data Sets - Calibration and Comparison

After reviewing the literature it was decided that the best way to determine the accuracy of calibration was the comparison of the model to a second independent data set (Walski 1986, Lansey et al 2001). Comparing model simulated values to the data that was used for calibration seemed unlikely to determine if the model was capable of accurately predicting model parameters. The two fire flow data sets were taken two weeks apart during similar weather conditions with the expectation of similar flow demands. The second data set duplicated the locations of the first data set with few exceptions based on customer concerns of flow onto private property. The amount of data collected during the second campaign was similar with fewer tests being taken at locations that had little variation and in areas found to have drainage problems.

SCADA Data Collection

The Girdwood system has a SCADA system set up with data logging capabilities. The SCADA interface for one of the booster stations is shown in Figure 3.9. The

SCADA data available for this study included the tank level, well discharge flow, supply pressure, discharge pressure, and flow for one of the PRV's, and the supply pressure, discharge pressure, and flow for the booster station. The data is logged every 5 seconds and was available through the use of the Proficy portal (Figure 3.10). The data was logged for the time period that the fire flow testing took place.

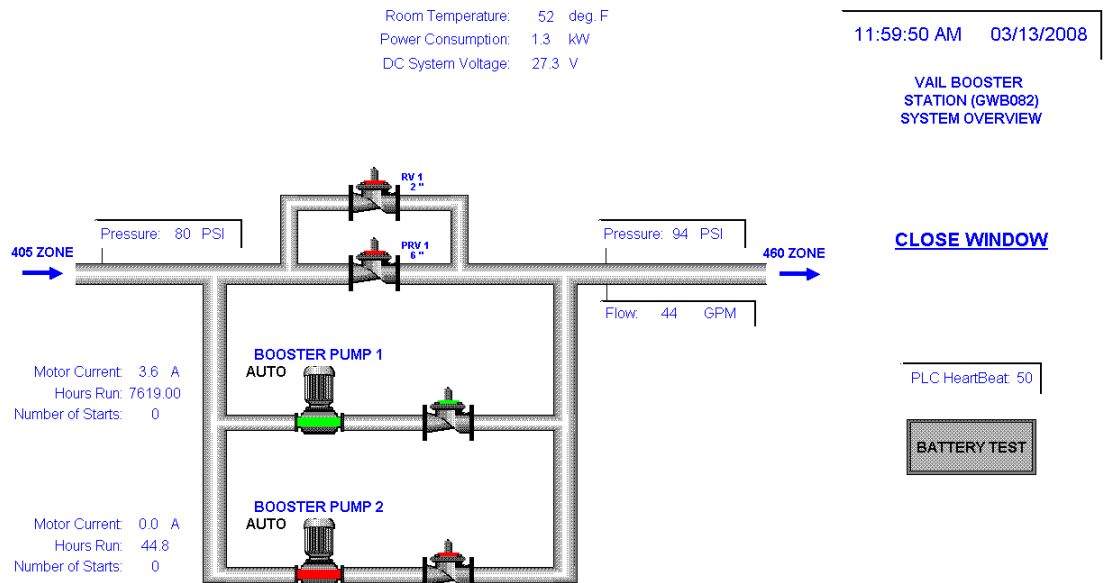


Figure 3.9 SCADA Interface for Booster Station

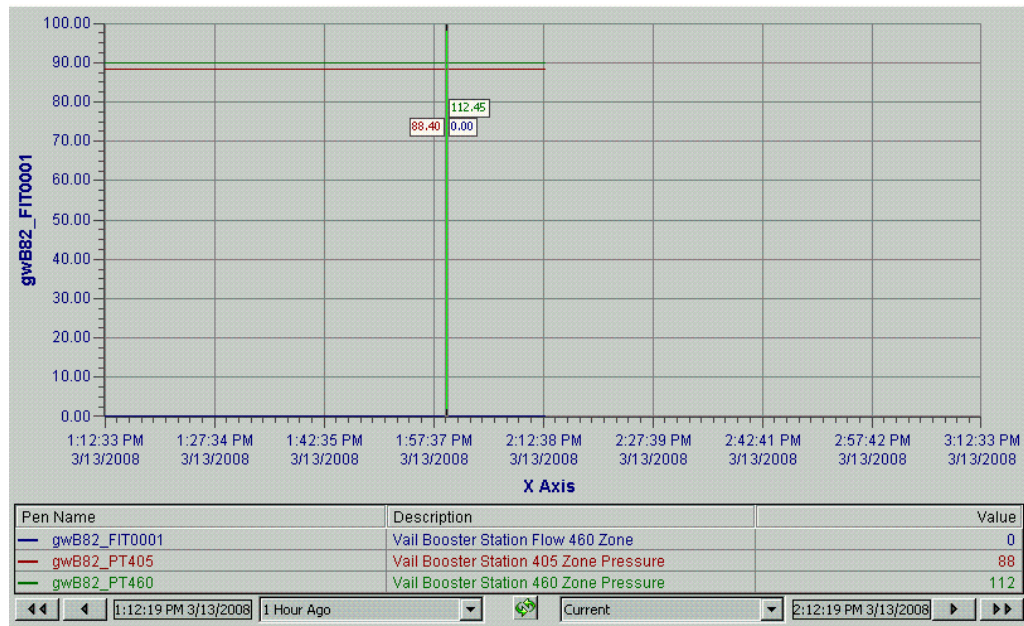


Figure 3.10 Proficy Portal Data Logging Interface

Data Analysis

The data analysis consisted of calibrating the model using various data sets and determining calibration accuracy by comparing model output to a constant independent data set.

The collected model data from the first campaign was grouped into several data sets described later to evaluate the effects of various parameters on calibration. The data from the second sampling campaign was left as a whole for assessment of calibration accuracy. The calibration data set was used in the model's GA calibrator, which returned values of the Hazen-Williams roughness coefficient, C , for each pipe group in the model. The Hazen-Williams equation for headloss is:

$$h_f = 3.02 * L * D^{-1.167} * \left(\frac{V}{C_h} \right)^{1.85}, \quad (1)$$

where h_f is the headloss due to friction, L is the pipe length, D is the pipe diameter, V is the velocity in the pipe, and C_h is the Hazen-Williams roughness coefficient. The C -values were then used to determine the accuracy with which the model could predict the second independent set of data. This process was repeated using different groupings and quantities of data. For this study, a run refers to the act of using a data set to calibrate the model and obtain C -values, then evaluating the calibration accuracy and computing a measure of fit between predicted and observed pressures.

Use of Genetic Algorithm Optimization Calibration

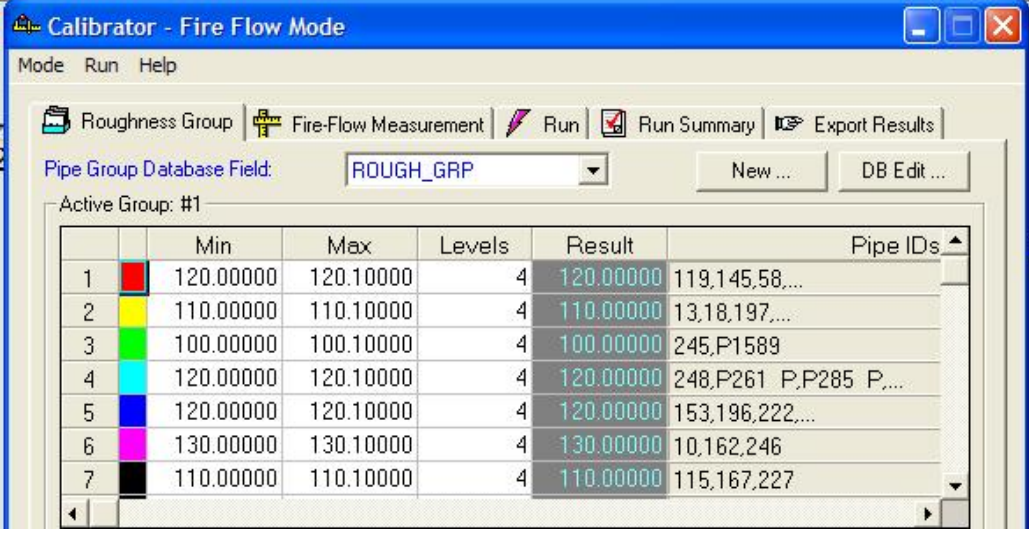
The GA calibrator in the model was used. The pipes were grouped based on parameters affecting roughness including pipe material, age, and diameter. For example, a 6 inch diameter AC pipe installed in 1962 might be group 1, and a similar pipe installed in 1963 might be group 2. There are 5 types of pipe in the Girdwood system, AC, CI, DI, PVC, and GALV. There were a total of 38 pipe groups. The model can use the Darcy-Weisbach, Chezy-Manning, or Hazen-Williams equation for calculating headloss. It was decided to use Hazen-Williams C -values as a basis for calibration as values for aged pipe are available from reference for comparison and there is generally a higher level of familiarity with the equation. The range of possible C -values used for calibration was 100-140 in denominations of 10.

The fire flow data was arranged in spreadsheets such that the node in the model corresponding to the flow hydrant and residual hydrant was read automatically. The pitot pressure was converted to a flow rate using the formula:

$$Q = C_f C_d d^2 \sqrt{P}, \quad (2)$$

where Q is flow rate, C_f is a unit conversion factor, C_d is the discharge coefficient, d is the diameter of the outlet, and P is the pitot pressure. The data were arranged into sets with some containing all tests, and some containing different groups of tests. For example, data was sorted by the magnitude of the pressure drop at the residual hydrant, by pressure zone, or by using all tests or one test per node. The data set was input into the fire-flow measurement tab of the fire flow calibrator. The calibrator was set to run 20000 trials for a calibration run. However, the calibrator would generally stop after fewer runs as model fitness ceased to improve. When evaluating the model's ability to predict observed values, the number of trials was set to 2000 to minimize calibrator run time as 20000 trials yields the same results in a significantly longer amount of time. Only one trial should be required to obtain the comparisons but it was desired to ensure the model was not iterating any of the C-values.

To determine the accuracy of the calibration, the C-values obtained for each pipe group were copied to a spreadsheet. A new range of C-values were input such that the lower bound for the C-value of each pipe group was the corresponding calibrated C-value. The upper bound for the C-value was increased by .1 over the lower bound. This forced the calibrator to use only the calibrated C-value for each respective pipe group (Figure 3.11).



		Min	Max	Levels	Result	Pipe IDs
1		120.00000	120.10000	4	120.00000	119,145,58,...
2		110.00000	110.10000	4	110.00000	13,18,197,...
3		100.00000	100.10000	4	100.00000	245,P1589
4		120.00000	120.10000	4	120.00000	248,P261 P,P285 P,...
5		120.00000	120.10000	4	120.00000	153,196,222,...
6		130.00000	130.10000	4	130.00000	10,162,246
7		110.00000	110.10000	4	110.00000	115,167,227

Figure 3.11 Model Verification Roughness Values

The information for the second set of fire-flow tests would be input into the fire-flow measurement tab of the calibrator. This second set of data included all tests and remained constant for all comparisons. The calibrator was then run for the sole purpose of expediting the comparison of simulated values to observed values as only the calibrated C-values were used (Figures 3.12 and 3.13). As the C-values remained constant and could not be changed by the calibrator to meet or exceed the accuracy threshold that would normally stop the GA, the full number of trials would be run.

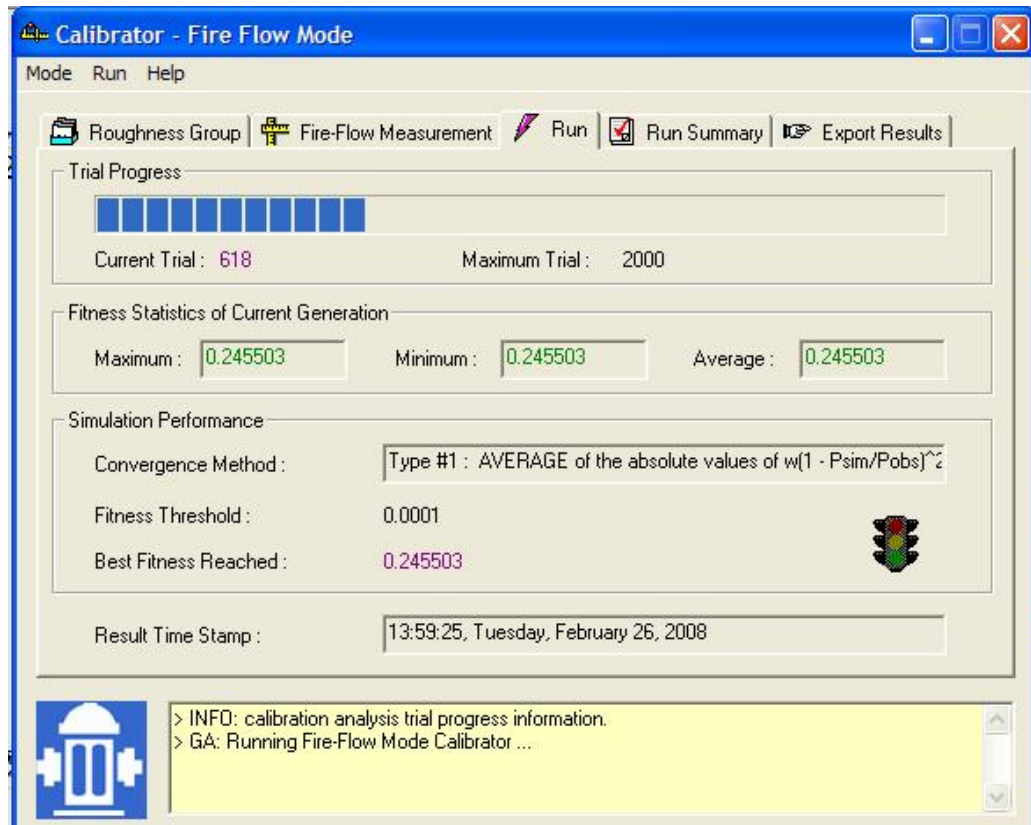


Figure 3.12 GA Calibrator Model Verification Run Process

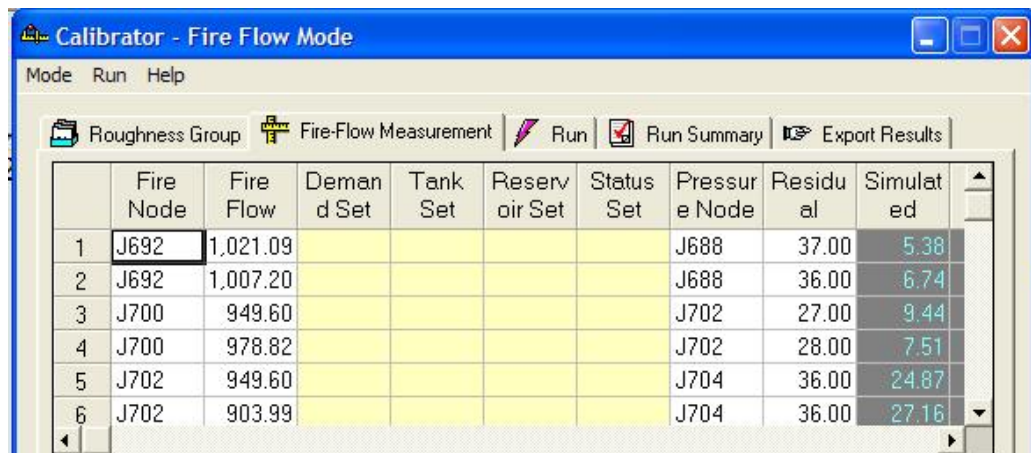


Figure 3.13 Model Verification Test Comparisons

The calibrator yields model simulated values to compare with observed values, percent difference between the simulated and observed values, and error statistics for all

tests. At the completion of a run a spreadsheet was created containing the data set used for calibration and its corresponding simulated residual pressures and percent differences, the calibrated C-value for each pipe group, and the simulated pressures and percent differences for both the static verification and the fire-flow verification. Calculated in the spreadsheet would be the squared relative difference between observed and simulated pressures (SE) for each test comparison, the average squared error (MSE), the standard deviation of the SE, and an average of the percent differences for all tests for both static and fire-flow. Table 3.1 shows the procedure used to calculate each statistic. Table 3.2 shows the description of all analyses performed, their desired outcomes, and the replicates of analysis. Each of these tests will be discussed in more detail below.

Table 3.1 Statistic Calculation Procedures

#	Simulated Pressure	Observed Pressure	Percent Difference	Squared Error
1	$P_{\text{simulated1}}$	$P_{\text{observed1}}$	$\left(\frac{P_{\text{simulated}} - P_{\text{observed}}}{P_{\text{observed}}} \right) * 100$	$\left(\frac{P_{\text{simulated}} - P_{\text{observed}}}{P_{\text{observed}}} \right)^2$
2	$P_{\text{simulated2}}$	$P_{\text{observed2}}$		
3	$P_{\text{simulated3}}$	$P_{\text{observed3}}$		
4	$P_{\text{simulated4}}$	$P_{\text{observed4}}$		

Average percent difference= =average(% diff 1 to 4)

Mean Squared Error =average(Squared Error 1 to 4)

Standard Deviation =stdev(Squared Error 1 to 4)

Table 3.2 Analysis Descriptions and Details

Description	Desired Outcome	Replicates
Static accuracy comparison vs. fire flow accuracy comparison	Determine the best method to evaluate model calibration accuracy	64
High loss data (good) vs. low loss data (useless)	Determine the effect of inducing high flows in the system during data collection on calibration accuracy	4
Duplicate data vs. unique data	Determine the calibration accuracy impacts from collecting more than one data point at a specific node	8
1 All data from all zones		1
2 One test per node from all zones		1
3 All data from 405 zone		1
4 One test per node from 405 zone		1
5 All data from 330 zone		1
6 One test per node from 330 zone		1
7 All data from 280 zone		1
8 One test per node from 280 zone		1
Accuracy vs. number of data points	Determine the optimal number of data points for model calibration	3
1 Elevation variability set		10
2 Good elevation set		10
3 Pipe group set		10
Sample locations	Determine optimal sampling locations for model calibration	2
1 Locations near boundary conditions		1
2 Locations far from boundary conditions		1

Static Accuracy Comparison vs. Fire Flow Accuracy Comparison

For many of the runs, the accuracy of the calibration was evaluated using both the static accuracy comparison and the fire flow accuracy comparison. The static evaluation method would use the calibrated C-values to determine the model's ability to predict measured static pressures at 22 unique nodes. The fire flow evaluation method would use the calibrated C-values to determine the model's ability to predict the 48 observed fire flows. MSE, standard deviation of SE, and average percent difference between predicted and observed values were calculated for both evaluation methods for each run.

High Loss vs. Low Loss Data

A general goal when performing fire-flow tests is to achieve a minimum 10 psi pressure drop between static and residual pressures at the residual hydrant (Walski et al 2001). Others reference minimum pressure drops between 5 and 20 psi (Ormsbee et al 1997). The impetus for this minimum pressure drop is to ensure a significant headloss is induced in the system. For this analysis a minimum of 10 psi drop was selected. Data collected was divided into two types based on this concept. The group with a pressure drop at the residual hydrant of at least 10 psi will be referred to as “good” data according to the convention of Walski (2000). The differences in calibration accuracy between using high loss and low loss data were compared by running calibrations with sets containing different combinations and quantities of both types of data. It is generally thought that high headloss data will provide a more accurate calibration than low loss data from the standpoint of data precision.

Model Calibration Accuracy vs. Number of Data Points

The analysis was continued with the goal of developing a relationship between calibration accuracy and number of tests used. The “good” tests were isolated into a data set and only one test per node was used. The test data for a particular node were similar enough that eliminating duplicates might have little effect. After removing duplicate fire flow tests and tests with less than a 10 psi drop, there were 13 “good” tests at unique nodes. The model calibration and verification using the procedure previously described was performed using this data set. Using a random number generator, one of the tests

was removed and the calibration and verification was performed again. This removal, calibration, and verification process was repeated until only one test remained. This process was repeated 10 times with a different random removal order each time. The average of the MSE for each of the data sets within each of the 10 runs was compared to the number of unique data points in the set.

Pipe Groups: Many vs. Few

The previous analyses were completed using 38 pipe groups. However, in many cases the install date, size, and type of the pipe are not known with confidence making it difficult to accurately place a pipe into a specific group. It is not thoroughly discussed in the literature what impact this may have on accuracy with the exception of the possibility of inaccurate diameters requiring incorrect pipe roughness factors to compensate (Ormsbee and Lingireddy 1997). The 38 pipe groups were reduced to three groups consisting of one group of pipes with “smooth” characteristics such as PVC, those with rougher characteristics such as DI, CI, and GALV, and those with “rough” characteristics such as AC. The accuracy versus number of good data points analysis was performed again using the same data groups and removal orders as before but using only the 3 pipe groups.

Data Collection Location

To evaluate the effect of data collection location on calibration accuracy, data collected from different areas in each pressure zone were used for calibration and the accuracy compared. Three pipe groups were used due to the data set size. One run was

performed using data collected far from the boundary conditions in the zone and another run performed using data collected close to the boundary conditions for each zone to determine the impact of stressing a larger portion of the zone (Walski 2000, Meier and Barkdoll 2000). Each run used 3 data points from 3 unique locations. The accuracy using the static and fire flow evaluation methods were compared.

RESULTS

Data Sets - Calibration and Comparison

The UK standards for model calibration involve determining calibration accuracy based on static pressure and pipe flows. The minimal flow measurement equipment availability in the Girdwood system and the inherent limitations of static pressure data described earlier led to an alternate method of calibration verification. It was decided that the best measure of calibration accuracy would come from a calibration using the first fire flow data set then comparing the residual pressures at corresponding flow rates to the residual pressures measured at the second data set. As the static pressures measured during both data sets were very similar, the comparison should likely yield an indication of calibration accuracy. Table 3.2 summarizes the type, number and desired outcome of runs performed for the analysis.

Three Pressure Zones vs. Four

As the infrastructure in the Girdwood systems 460 pressure zone is served by a booster station, the pump characteristics affect the accuracy of the measurements in this zone much more than pipe roughness. The pump curves in the model for these pumps consist only of a design head and flow. This may or may not accurately represent the actual pump curves. However, the SCADA representation of the pump station shows 2 pumps in parallel with 2 PRVs also in parallel. The model representation has 3 pumps in parallel with one PRV in series. The difference in configuration imparts large differences

in system behavior, the most significant being the PRV placed in series with the pumps. This model is able to predict static pressures in the zone with some accuracy as the PRV is set to the normal downstream pressure of the booster station, however, when fire flow occurs the model becomes unable to predict values for the zone with any reasonable accuracy. SCADA data detailing the upstream and downstream pressures and flow through the pumps is available; however, the data does not differentiate between when one pump is running or both, making the development of a descriptive pump curve difficult. It was decided to eliminate this zone from calibration as it is apparent the model does not represent the actual infrastructure for the booster station.

Calibration Accuracy Evaluation (Static vs FF)

The roughness coefficients from the calibration were used to determine the accuracy of the calibration by evaluating the model's ability to predict independent observed system parameters. These parameters consisted of static pressure measurements and a set of independent fire flow tests including flow and residual pressure. The model's simulated pressure values were compared to the observed values and MSE, standard deviation of SE, and a percent difference for observed and simulated values were calculated for the set as a whole. The static set consisted of 22 static pressures taken at unique nodes. The fire flow test consisted of 48 tests taken at 22 unique nodes.

There are many statistical comparisons that can be used. MSE, standard deviation of SE, and percent difference of the observed and simulated values using the fire flow

evaluation are shown for 38 runs using various data sets in Figure 4.1, Figure 4.2, and Figure 4.3 respectively.

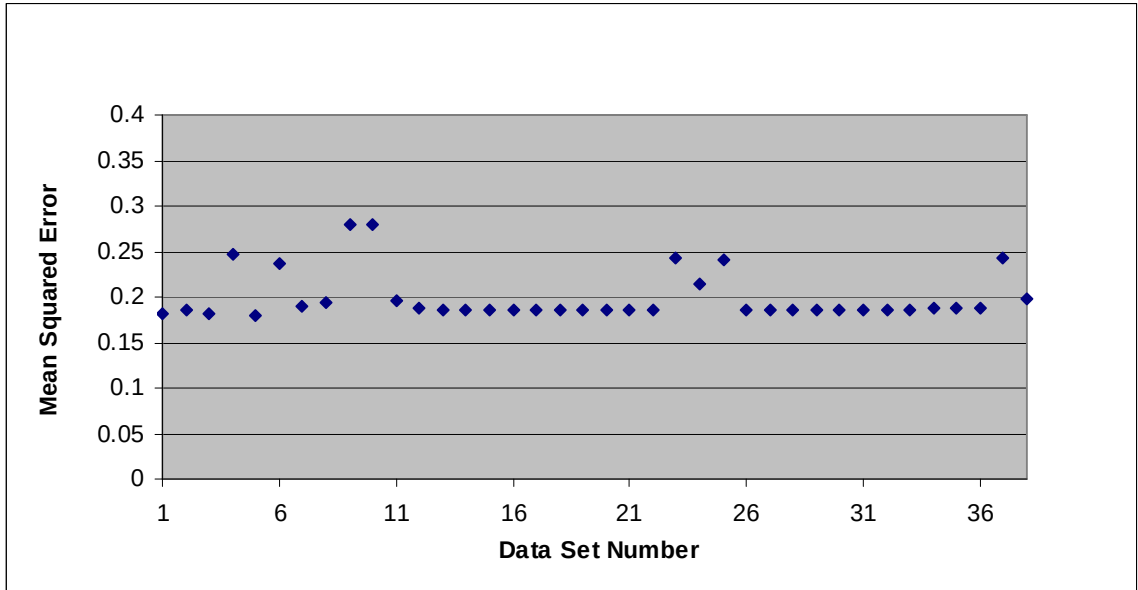


Figure 3.1 Mean Squared Error

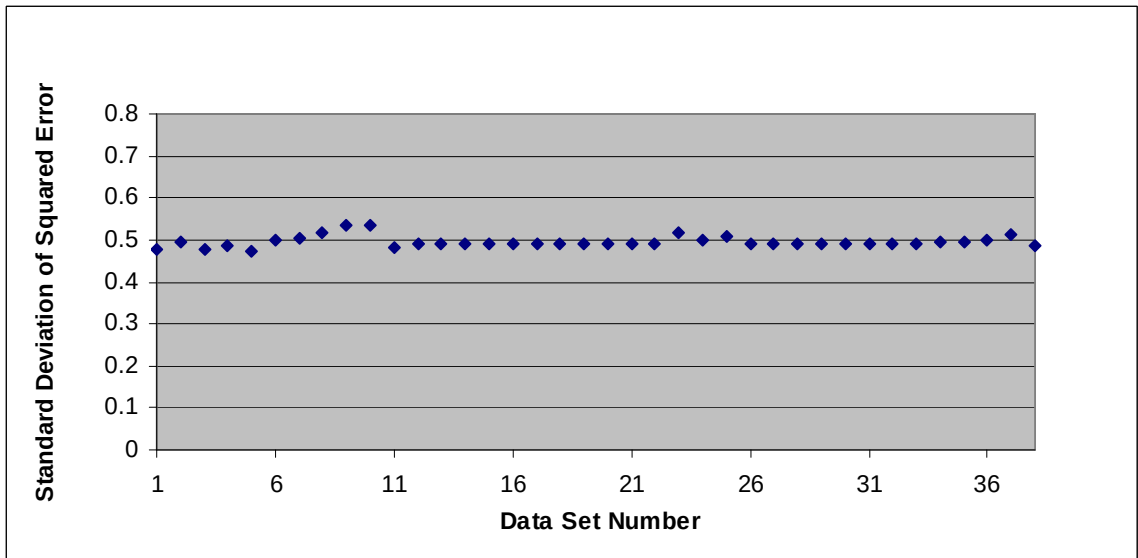


Figure 4.2 Standard Deviation of Squared Error

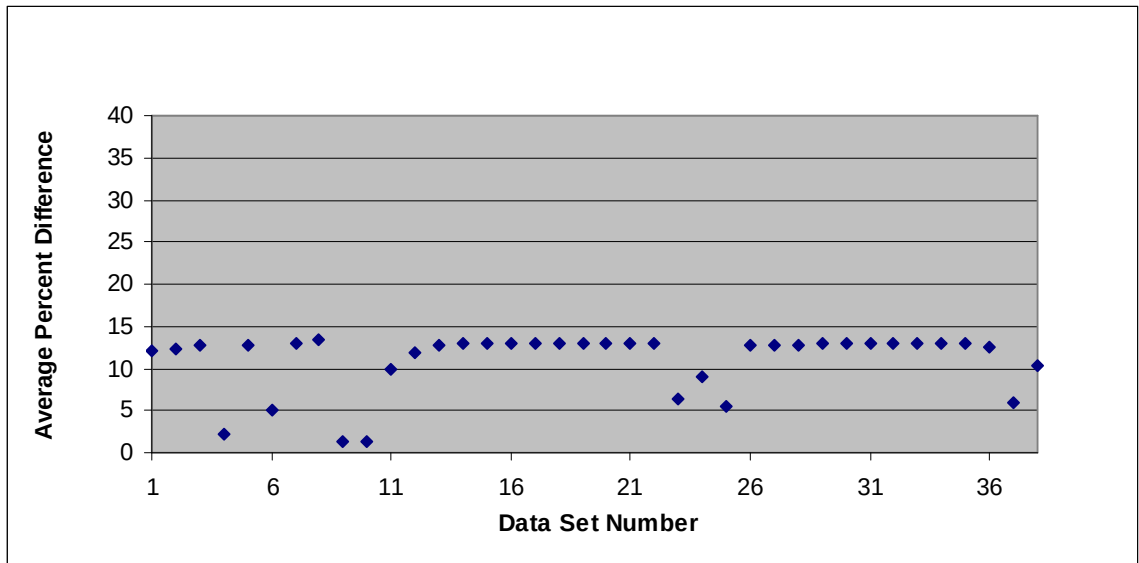


Figure 4.3 Average Percent Difference of Observed and Simulated Values

There was significant variation in the MSE between some data sets, some variation in standard deviation of SE, and some variation in percent difference of the observed and simulated values. The magnitude of percent difference of the observed and simulated values varies inversely with the MSE for given data sets. Therefore, MSE was chosen as a basis for accuracy comparison as it appears to be the most descriptive statistic for determining model accuracy.

Figure 4.4 shows the MSE for the static evaluation for the same data sets.

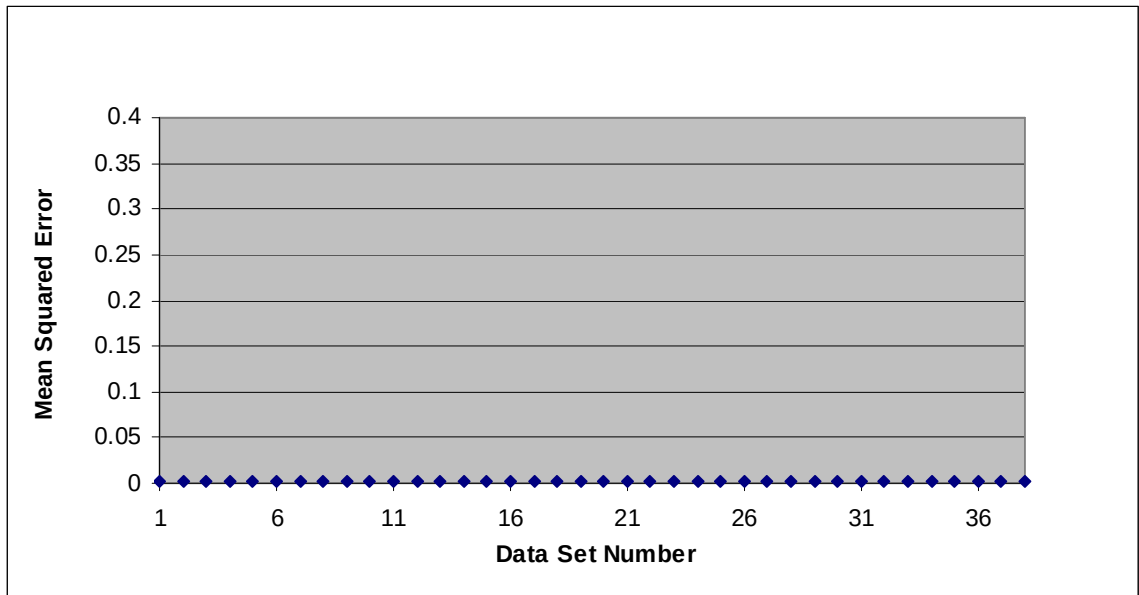


Figure 4.4 Mean Squared Error for Static Evaluation

The static evaluation method shows no variation in the MSE for the different data sets whereas the fire flow comparison shows significant variation for corresponding sets. The static method is illustrated here to be a poor descriptor of the model's ability to predict fire flow event. Thus, the MSE using fire flow evaluation method was chosen to describe model accuracy for this study.

Good vs. Useless Data (High Loss vs. Low Loss)

The accuracy for calibrations using data sets containing only high headloss data were compared to those containing only low headloss data. Low headloss data is considered to be useless as it does not ensure a significant difference between headloss and measurement precision. Therefore, the tests with low headloss have a high potential for measurement error. These sets contained the results of multiple tests at unique nodes.

The calibrations performed using high headloss data had a lower MSE and standard deviation of SE than the calibrations using low headloss data when evaluated using the fire flow comparison. In contrast, the accuracy of these calibrations when evaluated using the static comparison showed the low headloss data to have a lower MSE and standard deviation of SE. The average percent difference between the observed and simulated values was lower in magnitude for the low loss data than the high loss data for both the fire flow and static comparisons.

Duplicate Data vs. Unique Data

The effect of using more than one data point at a unique location on calibration accuracy was investigated. This effect is examined only for fire flow accuracy evaluations and not for static accuracy evaluation as the observed pressures are assumed to be steady state and more than one value cannot exist per node.

When data points from all 3 zones were used for calibration regardless of pressure drop, the difference in calibration accuracy when duplicate tests at each node are not used is small (data not shown). However, it does appear that the accuracy decreases when the duplicate data is dropped. When data from a single zone is used to calibrate the model the difference in accuracy between all data points and using one data point per unique test location is inconsistent. For the 330 zone the accuracy decreases, for the 280 zone accuracy decreases a very small amount, and for the 405 zone the accuracy increases. When using only high headloss data from all 3 zones, the accuracy does decrease by a small margin when duplicate tests are removed (data not shown).

Accuracy vs. # of Data Points

In an effort to understand the relationship between quantity of data and calibration accuracy, the effects of key variables were evaluated.

The first variable to be evaluated was the number of unique data points used. The overall measures of accuracy were generally higher for data sets containing multiple tests for unique nodes, however as the tests at each node were similar it was decided that there would be one test used for each node in the set. Only high headloss data was used, resulting in a data set containing 13 unique nodes and tests from the three pressure zones being used. The procedure described in the methods section was used for removal of a single test and corresponding node from the data set. A different random removal order was used for 10 repetitions of the procedure, resulting in ten sets of data for MSE, standard deviation of SE, and percent difference between observed and simulated pressure versus number of unique data points used in the calibrating data set. This resulted in 130 total data sets for 10 repetitions.

As the same removal order for corresponding repetitions were used for the all pipe group set and limited pipe group set, the number of pressure zones represented by the data used is of interest. To exhaust all combinations of removal orders, 13 factorial (6.2 billion) repetitions would need to be performed for each master set in question. As only 10 were performed, the possibility of the number of zones represented affecting the calibration accuracy must be considered in the event one of the orders removes all data from a pressure zone entirely before any other data. One zone contained 8 data points,

one zone contained 3, and the last zone contained 2. As only the data from high loss tests was used, the zone adjacent to the reservoir that contained the most infrastructure and had the most data initially collected had the least amount of data points representing the zone.

Table 4.1 shows the number of zones represented versus number of unique data points for each repetition.

Table 4.1 Number of Zones Represented vs. Unique Data Points Used

	Removal Order #										Average
	1	2	3	4	5	6	7	8	9	10	
# of Unique Data Points	13	3	3	3	3	3	3	3	3	3	3
	12	3	3	3	3	3	3	3	3	3	3
	11	3	3	3	3	3	3	3	3	3	3
	10	3	3	3	3	3	3	3	3	3	3
	9	3	2	3	3	3	3	3	3	3	2.9
	8	3	2	3	3	3	3	3	3	3	2.9
	7	3	2	3	2	3	3	3	3	3	2.8
	6	3	2	3	2	3	3	3	3	3	2.8
	5	3	2	2	2	3	3	3	3	3	2.7
	4	3	2	2	2	3	3	3	2	3	2.6
	3	3	2	2	2	3	3	2	2	2	2.4
	2	2	2	2	2	2	1	2	2	2	1.9
	1	1	1	1	1	1	1	1	1	1	1
# of Zones Represented											

Most removal orders have all three zones represented at 3 data points, 2 at 2 data points, and 1 at 1 data point. One of the removal orders has only 2 zones represented at 9 data points. Four removal orders have only 2 zones represented at 4 data points. This shows that the majority of the testing has all 3 pressure zones represented and the number of pressure zones represented is not influencing the test accuracy.

All Pipe Group Set

The all pipe group set used 38 pipe groups. The fire flow evaluation method was used. The average MSE from all 10 repetitions and 90% confidence interval was

calculated from the results of the 10 repetitions. The results of this set of 10 removal repetitions can be seen in Figure 4.5. The resulting graph with error bars can be seen in Figure 4.6.

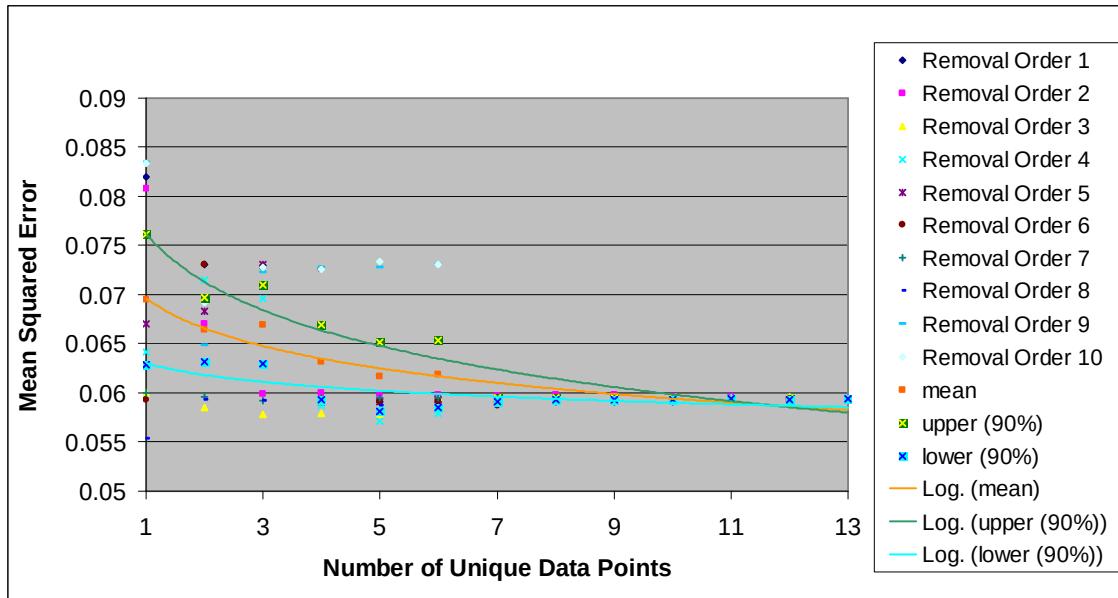


Figure 4.5 Mean Squared Error vs. Number of Unique Data Points

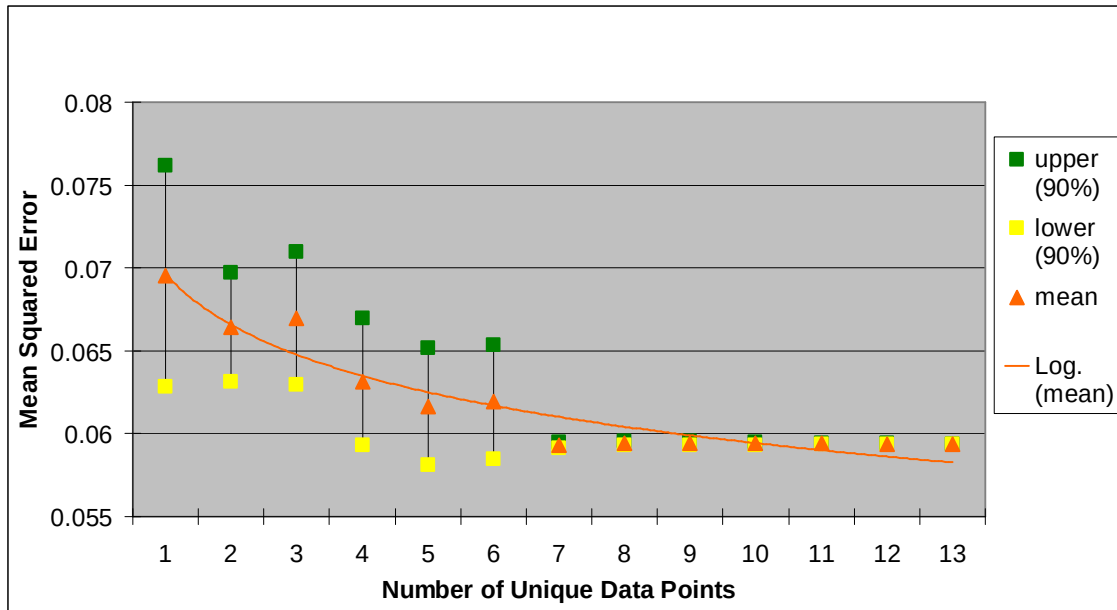


Figure 4.6 Mean Squared Error vs. Number of Unique Data Points

Figure 4.6 shows that the average MSE for all 10 repetitions for the use of all data points, 13, down to 8 data points is almost identical. With the use of 7 data points, the magnitude of the 90% confidence interval starts to increase and the magnitude of the average MSE for all 10 repetitions increases. The average MSE continues to increase for the use of 6 data points down to 2 data points with the 90% confidence interval maintaining a similar magnitude though having increased significantly over the use of 7 data points. The average MSE changes significantly from the use of 2 data points to 1 data point and the magnitude of the 90% confidence interval increases. Figure 4.5 shows a logarithmic curve fit of the upper and lower values of the 90% confidence intervals. This shows an increase in the magnitude of the 90% confidence interval as the number of unique data points used for calibration decreases.

Limited Pipe Group Set

The limited pipe group set used 3 pipe groups. The fire flow evaluation was used. The average MSE from all 10 repetitions and 90% confidence interval was calculated from the results of the 10 repetitions. The results of this set of 10 removal repetitions can be seen in Figure 4.7. The resulting graph with error bars can be seen in Figure 4.8.

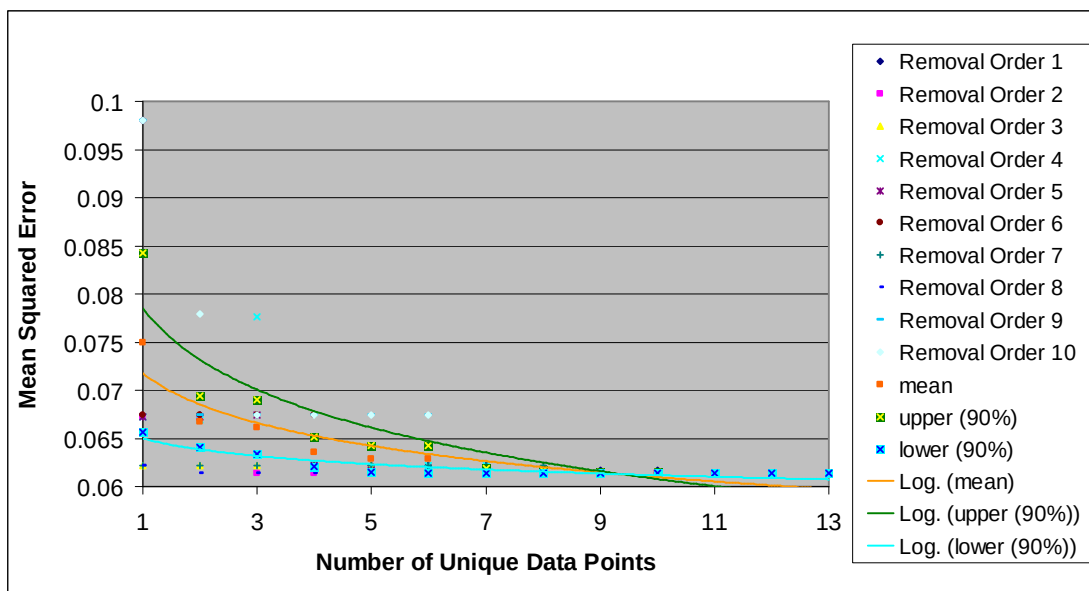


Figure 4.7 Mean Squared Error vs. Number of Unique Data Points

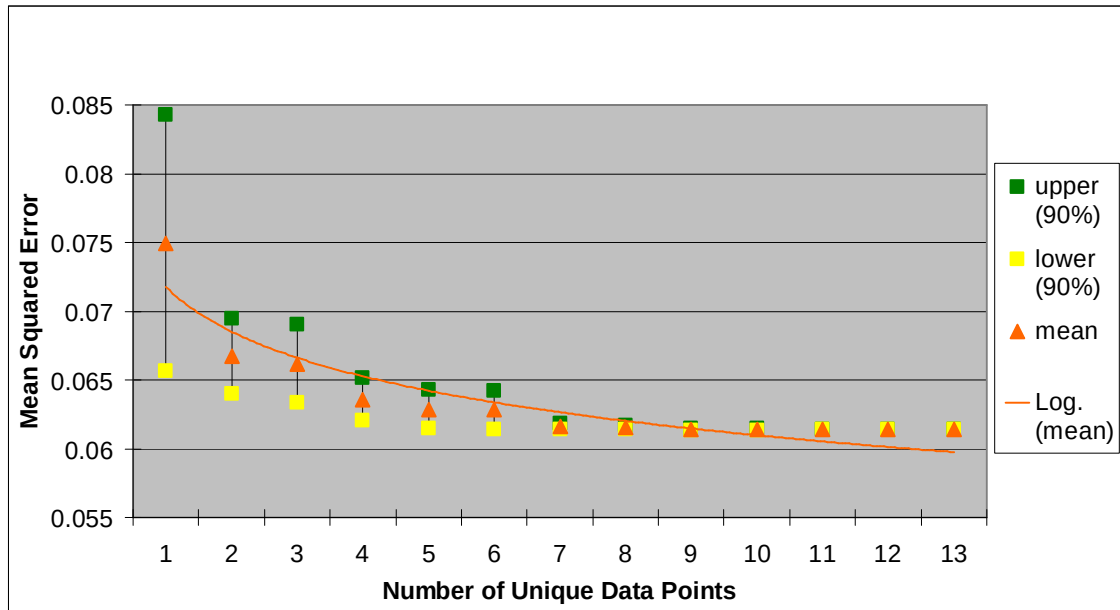


Figure 4.8 Mean Squared Error vs. Number of Unique Data Points

Figure 4.8 shows that the average MSE for all 10 repetitions for the use of all data points, 13, down to 8 data points is almost identical. This is similar to the all pipe group set. With the use of 7 data points, the magnitude of the 90% confidence interval starts to increase and the magnitude of the average MSE for all 10 repetitions increases. The average MSE continues to increase for the use of 6 data points down to 1 data point. The 90% confidence interval increased significantly over the use of 7 data points and continues to increase from 6 data points to 1 data point. Figure 4.7 shows a logarithmic curve fit of the upper and lower values of the 90% confidence intervals. This shows an increase in the magnitude of the 90% confidence interval as the number of unique data points used for calibration decreases.

Comparison

Figure 4.9 shows the all pipe group set and the limited pipe group set graphed on the same set of axes.

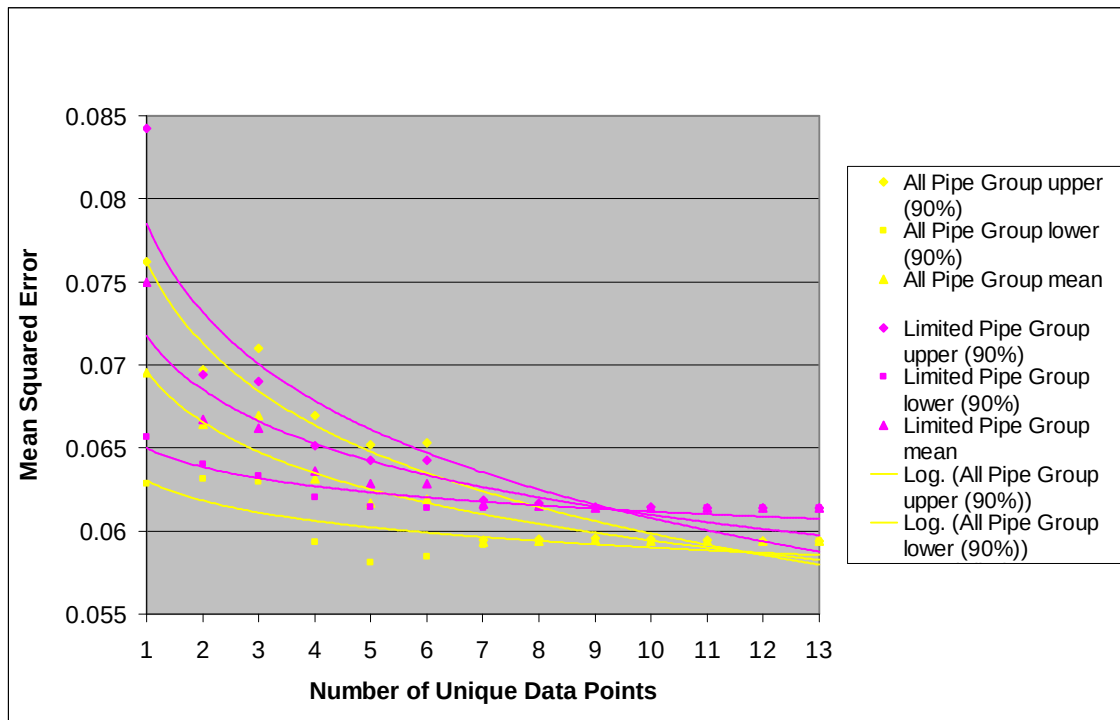


Figure 4.9 Mean Squared Error vs. Number of Unique Data Points for Both Sets

The magnitude of the MSE for the limited pipe group set is somewhat higher than the all pipe group set. However, for 7 data points used to 2 data points the 90% confidence interval for the limited pipe group set is significantly smaller than the all pipe group set. For 1 data point the limited pipe group set has a much higher confidence interval than the all pipe group set.

Data Collection Location

The effect of data collection location on calibration accuracy was evaluated by comparing calibration accuracy using test locations far from boundary conditions and test locations close to boundary conditions. The static evaluation of both runs was compared, and the calibrated C-values using test locations far from the zones boundaries yielded slightly lesser values of MSE, standard deviation of SE, and average percent difference between predicted and observed pressures than the C-values obtained using data collection locations close to zone boundaries. The fire flow evaluation of both runs was compared and the calibrated C-values using test locations far from the zones boundaries yielded a smaller value of MSE than the C-values obtained using data collection locations close to zone boundaries. The standard deviation of SE and average percent difference between predicted and observed pressures were higher than those yielded from C-values obtained using data collection locations close to zone boundaries.

Summary

In summary, the static and fire flow accuracy evaluations were compared, calibration accuracy using high headloss (good) and low headloss (useless) data was compared, the use of multiple tests at a node or unique tests effect on accuracy was compared, the relationship between model calibration accuracy versus number of data points used was examined, and effect of data collection location on calibration accuracy was determined.

DISCUSSION

Sampling Campaign Conclusions

The data collected during fire flow tests did not differ significantly during different times of the day. Approximately 3 tests could be run in an hour with a one day maximum of tests being 28. Every pressure zone had at least two tests run in it within a single day with the exception of the largest zone that was split into two days. As the reservoir started full on each testing day, the well pumps were not often required to fill the reservoir during testing. This means that the pressure to the system for all tests was supplied by the reservoir instead of the well system pumps.

Extreme care was taken to direct flow, and as a result, damage to roadways, driveways, and other property was avoided with the exception of an easily repaired minimal material washout event from a ditch due to high flow. All customer concerns were addressed in the field and no customer service complaints were received during the entire sampling campaign. As a result it can be assumed that no negative or unacceptably low pressures, no water quality issues, and no property damage occurred.

Static Accuracy Check vs. FF Check

The static method of accuracy evaluation does not appear to be a good indicator of the model's ability to predict actual system conditions. The static method may show that the model has an average error between static measurements that is very small, when the model is unable to predict fire flow with a similar accuracy. The static method of

accuracy evaluation also shows that there is no discernable difference between two given data sets that clearly achieve different levels of calibration accuracy when evaluated with the fire flow method. Walski (2000) showed that a model could appear calibrated for a wide range of C-values when a static accuracy evaluation was used.

Many standards for calibration, including those established by the UK Water Research Centre (1989), involve the verification of the model's ability to predict static pressures. The intended use of the model should be taken into consideration; however, it would seem the majority of model users require the prediction of high flow events. The static pressures in a given system at steady state are generally dependent only on the node elevations as head loss due to friction at average flow rates will more than likely prove negligible. Comparing simulated and observed static pressure for a system at steady state serves only to determine if the elevations input to the model are accurate or if there are incorrectly configured valves or pumps present in the model. This was important in this study as the elevations contained in the LIDAR database used to load the model were initially incorrect. Using the static accuracy comparison, the model consistently predicted static pressures 41 feet lower than observed. As a result, the processing of the data used to create the LIDAR elevation database was discovered to be incorrect. The data was reprocessed and a new, corrected LIDAR elevation database was developed and applied in the model. With the new elevation data thought to be accurate to +/- 2 feet, the model was able to predict static pressures very similar to the observed pressures. Thus the use of the static comparison is pertinent to check model input accuracy, but not model calibration accuracy. When a modeler is calibrating a model for such purposes this

should be taken into account and a static comparison between simulated and observed pressures should only be used to determine the adequacy of input to the model database such as elevations and valve conditions. The collection of a second set of independent fire flow data seems to be the best way to determine the models ability to predict high flow events.

Good Data vs. Useless Data

There is a significant difference in model calibration accuracy when high headloss “good” data is used rather than low headloss “useless” data. High headloss data is defined for this study as the data collected from a fire flow test where the difference between the static pressure and the residual pressure at the static hydrant is greater than 10 psi. Calibrations using high headloss data were more accurate than those using low headloss data when the fire flow method of accuracy evaluation was used. It was also shown that accuracy of calibration is relatively unaffected when comparing a calibration using all collected data and a calibration where low headloss data has been removed. This is probably due to the fact that the low headloss data has a higher likelihood of error as the tolerances of the measurement equipment are closer to the magnitude of the measurements being made. This is consistent with Walski’s (2000) and Meier and Barkdoll’s (2000) findings. The measurement on the pitot gauge used in this study consistently varied +/- 5 psi when flow was occurring. This required the operator to read the actual pitot pressure as the value between the two variations. The measurements made using this device have a much higher likelihood of being accurate when the pitot

pressure is of a higher magnitude. It has not been shown in this study that the accuracy of calibration is improved by the removal of the low loss data but it is assumed that the high loss data is driving the calibration and that the low loss data has little effect.

During data collection it has been suggested by some that if the pressure drop measured at the static hydrant is below the threshold designated, in this case 10 psi, another hydrant should be flowed to achieve the desired drop. This would be ideal if the equipment and manpower were available. For this study a 3 person crew was used with one diffuser available. All of the test locations in the 330, 460, and 280 zones did achieve the 10 psi drop with the use of one flow hydrant. Only 2 out of 28 of the test locations in the 405 zone adjacent to the reservoir achieved the desired 10 psi drop. In some cases the flow of another hydrant in the same area at the same time may have resulted in undesirably low system pressures. Thus the expense of another crew member utilizing an independent vehicle for transport and another diffuser would not have been justified for most of the tests. However, as one pressure zone was generally tested in a day, if it could be assumed that a particular zone may require multiple flowing hydrants to obtain high headloss data, a second flow hydrant and operating crew member could be justified for the testing performed in that pressure zone. For this study the majority of the tests taken in the pressure zone adjacent to the reservoir yielded low headloss data. A second crew member to operate flow hydrants would have been justified for tests performed in this zone. However, for this system it would be inadvisable to flow multiple hydrants without first knowing the residual pressure when using one flow hydrant.

A cost analysis comparison specific to a particular system should be performed to determine if it is more economical to use a smaller crew and perform some fire flow tests that will not be used or have more crew on standby in the event more flow hydrants are needed to make a particular test useful. In large systems where it is likely that more than one flow hydrant will have to be used, the same cost analysis should be applied for additional crew members.

Duplicate Data vs. Unique Data

The difference in model calibration accuracy between using all data points and only one data point at each test location leads to the conclusion that the performance of duplicate tests can, in some cases, have a positive impact on calibration accuracy. However, the magnitude of this accuracy improvement may be insignificant when compared to the accuracy improvement that may be obtained with a similar amount of effort applied to improving model input parameters. Any accuracy improvement as a result of duplicate testing is likely a result of reducing the effects of measurement inaccuracies or system variations through averaging of the data. For example, if there are 4 tests taken and 1 has a large measurement difference somewhere, the average of all the tests would yield a more representative test value than any one particular test. If all test equipment is consistent and accurate, duplicate tests for similar system conditions would yield no further useful data. However, taking duplicate tests during different system conditions and averaging the results for use in calibration may result in a model that is

capable of making accurate predictions for a larger number of varying system conditions though it may be less accurate for calibrating any one set of conditions.

The difference in accuracy between duplicate and single tests for this study may be due to the difficulty in reading the particular pitot gauge used or the difference in reservoir level at different times of day. If the conditions during the set used for calibration evaluation are similar to those during a particular test time than that test will appear to yield a more accurate calibration.

Effect of the Number of Good Data Points on Accuracy

For this study the calibration accuracy generally increased as the number of data points used for calibration was increased. This is consistent with de Schaetzen's (2000) findings. However, in this study, a point of diminishing returns for calibration accuracy versus number of unique data points used in calibration does exist. For the two different master sets of analyses, all pipe group and limited pipe group, the same calibration accuracy when using all 13 unique points could be achieved using 8 unique data points (Figures 4.5 to 4.8). Accuracy close to the maximum could be achieved using 7 unique data points.

Only one removal order of the 10 unique removal orders did not have all three pressure zones represented when using 8 unique data points (Table 4.1). However, the data does not consistently show a drop in calibration accuracy when all the data from one pressure zone is removed. This is also the case when there is only one pressure zone represented. However, the drop in accuracy when all the data from one pressure zone is

removed is more consistent when only 3 pipe groups are used when compared to the use of 38 pipe groups. This is likely due to the fact that as more pipes that may have different roughness characteristics are placed in a single group, the removal of the calibration data from a pressure zone that may contain a significant number of the pipes may weight the calibration improperly, resulting in a roughness coefficient less suitable for all pipes in the group. This phenomenon is not as prevalent when 38 pipe groups are used as all pipes with possible differing roughness characteristics have their own unique roughness group.

Effect of Number of Pipe Groups

Figure 4.9 shows the comparison between the all pipe group set and the limited pipe group set. The all pipe group set is able to achieve a higher calibration accuracy overall but the magnitudes of the confidence intervals for the two sets are very similar. Figures 4.6 and 4.8 show the error bars representing the 90% confidence interval of the average MSE for the two sets. The error bars for the use of 3 pipe groups behave much more predictably than those when 38 pipe groups are used as they steadily increase in magnitude as number of data points used decreases. The error bars when using 38 pipe groups increase and decrease slightly as the number of data points used decreases.

When 38 pipe groups were used in calibration rather than 3 pipe groups a higher accuracy was generally achieved when using the same number of unique data points. However, when the magnitude of the error bars is taken into account a higher confidence

in the calibration accuracy was achieved with 3 pipe groups when between 2 and 6 unique data points are used (Figure 5.1).

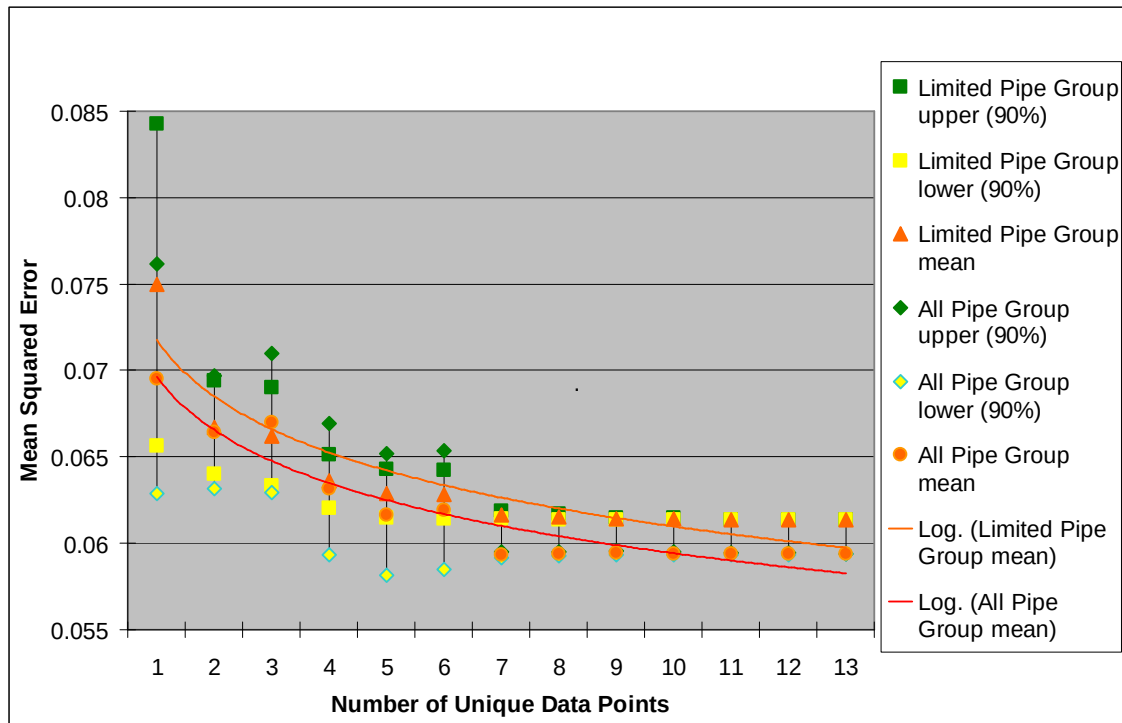


Figure 4.1 Mean Squared Error vs. Number of Unique Data Points for Both Sets

This is possibly due to the fact that 13 unique data points are used as the knowns to solve equations containing 38 pipe group roughnesses as the unknowns. As there are more variables than equations in this case, guesses are made for some of the pipe group roughnesses to best fit the observed data. As the number of data points used decreases, the number of guesses made increases. Some of these guesses may be correct and others incorrect, however, as the number of equations decreases it is more likely that an incorrect guess is used. This does not consider the possibility of differing data point strengths for model calibration as discussed below.

From evaluating the effects of node elevation accuracy and number of pipe groups used on overall calibration accuracy, it can be concluded that when accurate elevations are used, pipes can be grouped by a single criteria, material roughness properties in this case, rather than multiple criteria, resulting in many fewer pipe groups. For this study 38 pipe groups were reduced to 3 based on whether the pipe material displays smooth, intermediate, or rough characteristics. The calibration accuracy achieved using this smaller number of pipe groups containing many more pipes is similar to the accuracy when many pipe groups are used. It can also be concluded that calibration accuracy is much more dependent on input node elevation accuracy than by representing all pipe groups possible. Therefore, it is much more important for an accurate calibration to have accurate elevation data than detailed information for pipe characteristics such as age.

Optimal Sample Amount?

The optimal sample amount for this study was 8 unique fire flow tests where there was a greater than 10 psi drop in pressure at the monitored residual hydrant. This was the minimum number of unique data points used that calibrated the model to the maximum accuracy achievable by using all good data points.

There were 13 unique good data points, 8 from zone 330, 3 from zone 280, and 2 from zone 405. Zone 330 contains 23 hydrants, zone 280 contains 16 hydrants, and zone 405 contains 47 hydrants. The percentages of total good, unique hydrant flow tests compared to hydrants present in a zone were 35%, 19%, and 4% respectively. Using 8 data points with an average of 4.9 points present for zone 330, 1.7 for zone 280, and 1.4

for zone 405, these percentages become 21%, 11%, and 3% respectively. If the rule of thumb of performing fire flow tests on 10% of nodes, and all tests achieved the desired pressure drop at the residual hydrant, the number of tests needed would be 3 for zone 330, 2 for zone 280, and 5 for zone 405, resulting in a total of 10 tests.

The comparison of this makeup with that used for this calibration weights the data more heavily in the zone with the most infrastructure. However, this zone is adjacent to the reservoir and the majority of the tests did not achieve the desired pressure drop. It is likely for this system, based on the numbers shown here, that if fire flow tests were performed at 10% of nodes in each pressure zone and all tests achieved the desired pressure drop, that a calibration as accurate as the one performed would be obtained. It is important to note that, as this study does not examine the effects of all possible removal orders, the conclusions made on data set size are only applicable when data from all zones is included.

Data Collection Location

The comparison of the calibrated model accuracies using differing data collection locations anecdotally confirms the impact of location of data collection on the number of data points required for calibration. As data collected at points far from the boundary conditions of a zone induce non-negligible velocities in more of the pipes in the zone than data collected at points close to boundary conditions, it logically follows that the calibration achieved using such data would yield calibrated C-values that are more representative of all pipes and therefore enable the model to predict observed values with

more accuracy. This supports Kapelan et al's (2003) and Meier and Barkdoll's (2000) findings. In addition, pressure drops at the residual hydrants generally increased as the data collection points became further from the boundary conditions. This is likely due to the combined factors of headloss accumulating through the length of the zone and the increasing headloss in the reduced diameter pipes as the perimeter of the system is reached. This seems to be consistent with Walski's (2000) use of points far from boundary conditions to achieve high pressure drop, and de Schaetzen et al's (2000) use of points farthest from the boundary condition to represent the maximum number of pipes possible.

When 3 pipe groups were used and 3 data points collected at locations far from boundary conditions, accuracy similar to that obtained using 13 evenly distributed data collection locations was achieved. As the 3 data points are also contained in the 13 point data set it is possible that those 3 points drive the calibration for the most part, while some of the remaining points help to accurately represent portions of the system, such as loops and dead ends, that are unable to be stressed by the 3 points.

The selection of data points at the perimeter of the system has been shown in this study to achieve a similar level of accuracy to that obtained by using a larger quantity of evenly distributed data points. It can be assumed that these points have a higher strength for calibration accuracy than points from randomly selected locations.

Cost per Sample

The samples taken for this study involved the use of two vehicles and three crew members. The capital cost of the equipment used is neglected as the majority of it was already in the utility's possession. It was assumed that the vehicles cost \$100 each per day to operate. One of the crew members was an operator qualified to flow hydrants without damaging the system. The other crew members consisted of a staff level engineer and an entry level intern. Labor costs including overhead were assumed to be \$300 per hour. If an average of 23 fire flow tests were taken in a 9 hour day, the cost per fire flow test would be approximately \$125. If another crew member to operate flow hydrants and another vehicle were added the cost would be \$175 per sample, assuming no extra time needed. 123 total fire flow tests were performed to comprise both the calibration set and the accuracy evaluation set. Travel time between tests was negligible as it is a small town with little traffic and testing was done pressure zone by pressure zone so test nodes were fairly close together.

One way to minimize sample costs that has been presented by a company that performs fire flow data collection is to have more than one hydrant flowed in a pressure zone and more than one residual hydrant being monitored. This also applies if one flowed hydrant is sufficient to achieve desired pressure drop. When monitoring multiple residual hydrants many data points can be obtained for a pressure zone simultaneously. This would be the most efficient method for data collection if crew availability allows.

Costs that are not taken into account here as they are difficult to quantify but are necessary for any sampling campaign include determination of sampling locations, developing and organizing a sampling campaign, public notification, customer service, possible water quality issues, and possible damage to infrastructure with associated emergency costs.

Lessons Learned – Future Research Needs

As only 10 repetitions of the removal procedure were used out of the possible 6.2 billion, the effect of pressure zone representation in the calibrating data set has not been fully explored. While it has been shown in the data when using many pipe groups that a more accurate calibration is achieved using data from multiple pressure zones rather than a single zone, the effects of using data from a zone with similar pipe characteristics to a second zone while not using any data from the second zone has not been explored. In a large system when few pipe groups are used, it seems that pipes with similar characteristics would exist in multiple zones. The calibrated roughness coefficient should yield accurate values for such a zone even if no data representing it was used in the calibration.

CONCLUSIONS AND RECOMMENDATIONS

- A well planned sampling campaign will allow the collection of field data required for model calibration at a minimum expense by avoiding property damage, customer service complaints, and collection of unusable data.
- Comparing static pressures is useful for determining the adequacy of model elevation data, however, fire flow tests yield a more accurate measure of calibration accuracy with respect to pressure.
- High headloss data, specifically from fire flow tests, yields a more accurate model calibration than low headloss data.
- Performing multiple tests at a unique location may result in a more accurately calibrated model depending on measurement equipment used and system condition variability.
- The accuracy of calibration increased asymptotically as the number of good, unique, fire flow tests used for calibration increases.
- The increase in calibration accuracy when using as many pipe groups as possible as opposed to a few pipe groups is minimal. In general, the number of pipe groups used should be limited to the number of data points used in calibration.
- For this study a point of diminishing returns was found to exist. After a certain point, the accuracy of the model calibration did not increase as more data points were used in calibration.

- Data collection location should be considered if a minimal amount of data collection is desired. Data collection performed at locations carefully selected to provide the maximum amount of information can yield a similar level of accuracy when compared to a larger quantity of data collected throughout a system.
- Future research would be useful to investigate the possibilities of calibration using data collected from representative zones instead of all zones in an entire system.

REFERENCES

- deSchaetzen, W. (2000), "Optimal Calibration and Sampling Design for Hydraulic Network Models", *PhD Thesis*, School of Engineering and Computer Science, University of Exeter.
- deSchaetzen, W., Walters, G.A., and Savic, D.A. (2000), "Optimal Sampling Design for Model Calibration Using Shortest Path, Genetic and Entropy Algorithms", *Urban Water Journal*, 2, 141-152.
- Engineering Computer Applications Committee (1999), "Calibration Guidelines for Water Distribution System Modelling", *Proc. AWWA 1999 Imtech Conference*.
- Howie, D.C., (1999), "Problems with SCADA Data for Calibration of Hydraulic Models", *Proc. Water Resources Planning and Management Conference*, Tempe.
- Kapelan, Z.S., Savic, D.A., and Walters, G.A. (2003) "Multiobjective Sampling Design for Water Distribution Model Calibration", *Journal of Water Resources Planning and Management*, ASCE, 129(6), 466-479.
- Kapelan, Z.S., Savic, D.A., and Walters, G.A. (2005) "Optimal Sampling Design Methodologies for Water Distribution Model Calibration", *Journal of Hydraulic Engineering*, ASCE, 131(3), 190-200.
- Lansey, K.E., El-Shorbagy, W., Ahmed, I., Araujo, J., and Haan, C.T. (2001), "Calibration Assessment and Data Collection for Water Distribution Networks", *Journal of Hydraulic Engineering*, ASCE, 127(4), 270-279.
- Meier, R.W., and Barkdoll, B.D. (2000) "Sampling Design for Network Model Calibration Using Genetic Algorithms", *Journal of Water Resources Planning and Management*, ASCE, 126(4) 245-250.
- Ormsbee, L.E., and Lingireddy, S. (1997), "Calibrating Hydraulic Network Models", *American Water Works Association, Journal*, 89(2), 42-50.
- Walski, T.M. (1983), "Technique for Calibrating Network Models", *Journal of Water Resources and Planning Management*, ASCE, 109(4), 360-372.
- Walski, T.M. (1986), "Case Study: Pipe Network Model Calibration Issues", *Journal of Water Resources Planning and Management*, ASCE, 112(2), 238-249.

Walski, T.M. (1995), "Standards for Model Calibration", *Proc. AWWA Computer Conference*, Norfolk.

Walski, T.M. (2000) "Model Calibration Data: The Good, The Bad and The Useless," *J AWWA*, 92(1), p. 94.

Walski, T.M. (2000), "Pitfalls in Calibrating an EPS Model", *Proc. Joint Conference on Water Resources and Engineering and Water Resources Planning and Management*, Minneapolis.

Walski, T.M., Chase, D.V. and Savic, D.A. (2001) *Water Distribution Modeling*, Haestad Methods, Inc.

WRc (1989), "Network Analysis - A Code of Practice", Water Research Centre.