GROUNDWATER MONITORING NETWORK DESIGN USING DEMO METHOD

GROUNDWATER MONITORING NETWORK DESIGN USING ADDITIONAL OBJECTIVES IN DUAL ENTROPY MULTI-OBJECTIVE OPTIMIZATION METHOD

By JAMES LEACH, B.A.Sc

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree Master of Applied Science

McMaster University MASTER OF APPLIED SCIENCE (2015) Hamilton, Ontario (Civil Engineering)

TITLE: Groundwater Monitoring Network Design Using Additional Objectives in Dual Entropy Multi-Objective Optimization Method

AUTHOR: James Leach, B.A.Sc. (University of Regina)

SUPERVISOR: Dr. Paulin Coulibaly (McMaster University)

Dr. Yiping Guo (McMaster University)

NUMBER OF PAGES: x, 77

Abstract

This study explores the applicability of including groundwater recharge and water table variation as additional objective functions in a multi-objective optimization approach to design optimal groundwater monitoring networks. The study was conducted using the Ontario Provincial Groundwater Monitoring Network wells in the Hamilton, Halton, and Credit Valley regions in southern Ontario. The Dual Entropy-Multiobjective Optimization (DEMO) model which has been demonstrated to be sufficiently robust for designing optimum hydrometric networks was used in these analyses. The importance of determining the applicability in using additional design objectives in DEMO, including groundwater recharge and groundwater table seasonal variation, is rooted in the limitations of groundwater data and the time required setting up the models. While recharge allows for the capturing of spatial variability of climate, geomorphology, and geology of the area, the groundwater table series reflect the temporal/seasonal variability. The two set of information are complementary and should provide additional information to the DEMO for optimal network design. Two sources of groundwater recharge data were examined and compared; the recharge provided by the local conservation authorities, calculated using both the Precipitation-Runoff Modeling System (PRMS) and Hydrological Simulation Program--Fortran (HSP--F), and the recharge calculated in situ using only PRMS. The entropy functions are used to identify optimal trade-offs between the maximum possible information content and the minimum shared information between each of the existing and potential monitoring wells. The additional objective functions are used here to quantify the hydrological characteristics of the vadose zone in the aquifer as well as the potential impacts of agricultural, municipal, and industrial uses of groundwater in the area, and thus provide more information for the optimization algorithm to use. Results show that including additional design objectives significantly increases the number of optimal network solutions and provides additional information for potential monitoring well locations. These results suggest that it is worthwhile to include recharge as a design objective if the data is available, and to include groundwater table variation for the design of monitoring wells for shallow groundwater system.

Acknowledgements

I would like to thank my supervisors Dr. Paulin Coulibaly and Dr. Yiping Guo for the knowledge and experience I have gained over that past two years from the opportunity they provided me. I would like to thank Dr. Kurt Kornelsen and Dr. Jongho Keum for being able to help me with any question I had throughout these years, and Dr. Jos Samuel the co-author of the original CRDEMO paper. Finally I would like to thank my wife and family for their support.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ε-hBOA	epsilon-dominance hierarchical Bayesian optimization algorithm
ACO	Ant colony optimization
BB	Branch and bound
BOA	Bayesian optimization algorithm
CR-DEMO	Combined regionalization and dual entropy multi-objective optimization
CVC	Credit Valley conservation authority
DEMO	Dual entropy multi-objective optimization
FEFLOW	Finite Element subsurface FLOW system
GIS	Geographic Information Systems
HHCA	Hamilton-Halton conservation authority
HHCV	Hamilton-Halton-Credit Valley
HSP-F	Hydrological simulation program - Fortran
IDW	Inverse distance weighting
MASL	Meters above sea level
MT3DMS	A modular 3-D multi-species transport model for simulation of advection, dispersion, and chemical reaction of contaminants in groundwater systems
PGMN	Provincial groundwater monitoring network
PRMS	Precipitation-Runoff Modeling System
RT3D	Reactive transport in 3-dimensions
SA	Simulated annealing
SHARCNET	Shared Hierarchical Academic Research Computing Network

List of Symbols

W_u	Potential monitoring well groundwater level
W_i	Existing monitoring well groundwater level
w_i	weighting for existing monitoring wells
h_i	Distance between existing and potential monitoring wells
E	Set of existing monitoring wells
X	Set of potential monitoring wells
$S_{M,N}$	Set of existing and potential monitoring wells
$H(X)$	Entropy (Marginal, Joint, Multivariate)
$C(X)$	Total Correlation
$d(S_{M,N})$	The Euclidean distance values for the chosen optimal networks

1. Introduction

1.1. Purpose of Research

This research was undertaken to fulfill a portion of the HYDRONET project. Under this project, a decision support tool for developing optimal hydrometric networks was to be developed. The design of a network that is able to provide the maximum amount of information for the end user is essential due to the inherent importance of hydrometric data (Mishra & Coulibaly, 2009). Reliable and accurate hydrometric data is fundamental for the effective management of water resources. This includes better optimization of water uses in both the public and private sectors. Under the HYDRONET project several types of hydrometric network design methods are being developed, however, for this research, the groundwater monitoring network design will be the focus. The end goal is to use these hydrometric network design methods to produce optimal networks in areas that currently do not have adequate number of monitoring stations and are potentially data poor areas.

This research will determine the benefits of including informative hydrological/hydrogeological variables as additional objectives in the multi-objective optimization algorithm from (Samuel, Coulibaly, & Kollat, 2013) for the optimal design and/or augmentation of a groundwater quantity monitoring well network. In particular, the research focus is on the optimal spatial locations of the monitoring wells, and how they are influenced by the use of the additional objectives, not the temporal frequency at which each well is monitoring which is another aspect of optimal network design.

1.2. Network Design Background

Hydrometric monitoring networks have an important role in water resources planning and management and have been around for many years. The Canadian Federal Hydrometric Network, which monitors water levels and streamflow, has been around since the 1890s (Environment Canada, 2013b; Mishra & Coulibaly, 2009). In Ontario, the Provincial Groundwater Monitoring Network (PGMN) has been around since 2001, which not only provides groundwater level data but also water quality information (Government of Ontario, 2013b). Groundwater monitoring network design can be divided into two streams: those designed for groundwater quality monitoring and those designed for groundwater quantity monitoring. This does not mean that one network could not perform both tasks, just that the approaches used in the design of each vary based on the intended use of the network. Programs such as these are important to maintain and expand due to the importance of hydrometric data for water resource management practices, as the availability and reliability of hydrometric data is important for planning and management purposes (Mishra & Coulibaly, 2009).

2. Literature Review

Network design approaches which are needed in the design of groundwater quality monitoring networks were reviewed in (Loaiciga et al., 1992); the findings of which suggest that to design a robust monitoring network a combination of hydrogeological and statistical simulation, variance, or probability based methods should be used. There has been a wide range of literature over the past twenty-five years on the design of groundwater quality monitoring networks. These studies used a variety of data generation methods including Inverse Distance Weighting (IDW) (Li & Hilton, 2005, 2007; Reed, Minsker, & Valocchi, 2000; Wu, Zheng, & Chien, 2005), Kriging methods (Babbar-Sebens & Minsker, 2008, 2010, 2012; Ben-Jemaa, Mariño, & Loaiciga, 1994; Dhar & Datta, 2009; Kollat, Reed, & Kasprzyk, 2008; Kollat & Reed, 2006, 2007; Ling, Rifai, & Newell, 2005; Reed, Kollat, & Devireddy, 2007; Reed et al., 2000; Reed & Minsker, 2004; Wu et al., 2005; Wu, Zheng, Chien, & Zheng, 2006), MODFLOW (Alzraiee, Bau, & Garcia, 2013; Asefa, Kemblowski, Urroz, & McKee, 2005; Bashi-Azghadi & Kerachian, 2010; Khader & McKee, 2014; Reed et al., 2000; Wu et al., 2005, 2006), Par Flow (Kollat, Reed, & Maxwell, 2011; Reed & Kollat, 2012, 2013), MT3MDS (Bashi-Azghadi & Kerachian, 2010; Hudak, Loaiciga, & Mariño, 1995; Khader & McKee, 2014; Wu et al., 2005, 2006), and RT3D (Reed et al., 2000). The optimization methods used in the groundwater network design literature include Branch and Bound (BB) (Andricevic, 1990; Ben-Jemaa et al., 1994; Carrera, Usunoff, & Szidarovszky, 1984; Dhar & Datta, 2007; Hudak et al., 1995; Hudak & Loaiciga, 1993; Loaiciga, 1989; Wagner, 1995), Simulated Annealing (SA) (Meyer, Valocchi, & Eheart, 1994; Nunes, Cunha, & Ribeiro, 2004; Nunes, Paralta, Cunha, & Ribeiro, 2004; Storck, Eheart, & Valocchi, 1997), Ant Colony Optimization (ACO) (Li & Hilton, 2005, 2007), entropy theory (Abrishamchi, Owlia, Tajrishy, & Abrishamchi, 2008; Mogheir, de Lima, & Singh, 2009; Mogheir & Singh, 2002; Nunes, Cunha, et al., 2004), various Genetic Algorithms (GA) (Alzraiee et al., 2013; Babbar-Sebens & Minsker, 2008, 2010, 2012; Bashi-Azghadi & Kerachian, 2010; Cieniawski, Eheart, & Ranjithan, 1995; Kollat et al., 2008; Kollat & Reed, 2006, 2007; Reed et al., 2007; Reed, Minsker, & Goldberg, 2001; Reed et al., 2000; Reed & Minsker, 2004; Wagner, 1995; Wu et al., 2006), machine learning (Ammar, Khalil, McKee, & Kaluarachchi, 2008; Asefa et al., 2005; Bashi-Azghadi & Kerachian, 2010; Khader & McKee, 2014), and Bayesian Optimization Algorithms (BOA) (Kollat et al., 2008, 2011; Reed & Kollat, 2013). These methods are ultimately used to design a monitoring network using a minimal amount of wells that have a higher chance of detecting contaminant plumes. Despite the fact that several network design methods have been developed over the years for groundwater quality monitoring networks, few methods have been published on the design of groundwater quantity monitoring networks. Those that have are similar and generally include Kriging variance reduction methods, such as overlaying the Kriging variance map over the study area and placing monitoring wells in high variance locations (Khan, Chen, & Rana, 2008; Prakash & Singh, 2000; Triki, Zairi, & Ben Dhia, 2013; Yang, Cao, Liu, & Yang, 2008; Zhou, Dong, Liu, & Li, 2013).

Several methodological developments in hydrometric network design, particularly those for streamflow and precipitation monitoring networks, were reviewed in (Mishra & Coulibaly, 2009). That review found that the most efficient methods for water monitoring network evaluation and design are entropy-based methods (Caselton, 1980; Husain, 1979, 1987, 1989; Krstanovic & Singh, 1992a, 1992b; Mishra & Coulibaly, 2010) and multi-objective optimization methods (Kollat et al., 2008, 2011). Entropy (information) theory itself has been used in groundwater monitoring network design studies previously (Abrishamchi et al., 2008; Mogheir et al., 2009; Mogheir & Singh, 2002; Nunes, Cunha, et al., 2004). The merit of information theory (Shannon, 1948) is that it is able to directly define information and quantify uncertainty (Harmancioglu & Singh, 1998; Mishra & Coulibaly, 2010; Mogheir, Singh, & de Lima, 2006; Mogheir & Singh, 2002). The use of entropy in hydrology and water resources practices was reviewed by (1997). A fundamental basis of this approach is that the lower the transinformation values between stations, the lower the shared information between these stations and therefore, the more independent the stations are. On the other hand, the larger the transinformation values, the larger the duplicity of the same information and therefore, the more dependent the stations are (Mishra & Coulibaly, 2010). This approach requires exhaustive and repetitive computations to determine the accurate locations of new stations to be added (Husain, 1989; Mishra & Coulibaly, 2010). The most developed network models involve a combination of entropy-based and optimization methods (Alfonso, He, Lobbrecht, & Price, 2012; Alfonso, Lobbrecht, & Price, 2010; Rianna et al., 2012; Samuel et al., 2013). Using binary decision variables in the optimization method, the model has the capability to easily seek the optimal trade-off between several entropy functions by systematically selecting values from within a reliable set of all existing and potential stations. The precise locations of new stations to be added can be well detected and defined.

Accounting for spatial and temporal variability is an important component for designing optimum hydrometric networks, however, there are few studies that do both (Mogheir et al., 2009). In particular, it is critical to provide accurate and qualitative hydrological information of the entire area covered by the networks (Husain, 1989; Mishra &

Coulibaly, 2009). Studies for evaluating and designing optimum networks have been discussed in the literature; however, not many studies have thoroughly evaluated the spatial variability of the optimum networks. This would include using the variation of temporal data resolutions or in the case of this study incorporating informative hydrological/hydrogeological variables in designing optimum hydrometric networks. Optimum hydrometric networks should present the hydrological variables needed, the appropriate time interval of the variables observed, the density of the existing network and the accuracy of the data for end users (Mishra & Coulibaly, 2009). The use of limited data records, the selection of inappropriate sampling intervals, and/or the exclusion of informative variables in designing optimum hydrometric networks may limit the network models in optimizing the space-time trade-off between the locations of the existing and potential new stations, searching for the optimum locations of new additional stations from all available potential locations, and generating and obtaining the most informative spatial distributions of optimal networks. Entropy is a powerful tool that can be used in this respect, as it provides a quantitative measure of the information content within a hydrometric network analysis (Mishra & Coulibaly, 2010; Singh, 1997). The Combined Regionalization and Dual Entropy-Multiobjective Optimization (CR-DEMO) was developed by (2013) as a response to (Coulibaly, Samuel, Pietroniro, & Harvey, 2013), which found current hydrometric monitoring network density in many areas of Canada to be insufficient. In the CR-DEMO, the regionalization approach was used in flow estimation for the potential additional stations and the dual entropy multi-objective optimization (DEMO) approach was used to identify optimal entropy function trade-offs between the maximum possible information content (joint entropy) and the minimum shared information (total correlation) among the stations. Since this study is not focused on the regionalization approaches used, the DEMO portion of CR-DEMO model will be focused on; as it is able to determine the optimum locations for new hydrometric stations to be added in a monitoring network by capturing the information content of the networks (Samuel et al., 2013).

Objectives of groundwater level monitoring include characterizing groundwater systems, analyzing groundwater quantitative status, identifying changes in groundwater recharge, storage and discharge, detecting effects of climate change on groundwater resources, assessing impacts of groundwater development, calibrating groundwater flow models, and assessing the effectiveness of groundwater management and protection measures (Zhou et al., 2013). Additional objective functions that can quantify the hydrologic behaviours of the study area were included in DEMO to determine if the optimal networks found by the model would be enhanced. This research will expand on the DEMO approach of (2013) by evaluating the impact and efficacy of including additional informative hydrological/hydrogeological variables as objective functions in DEMO. Initially, use of an existing metric such as the vulnerability index DRASTIC was considered. DRASTIC has been used before in the evaluation of groundwater quality monitoring networks (Aller, Bennett, Lehr, Petty, & Hackett, 1987; Baalousha, 2010); however it was decided to use values which are more easily quantifiable such as water table variation and groundwater recharge for the design of optimum groundwater monitoring networks.

Changes in groundwater level can have implications for municipal, industrial, and agricultural practices (Lavoie, Joerin, Vansnick, & Rodriguez, 2015). A common cause of changes in groundwater level/availability is overdrafting, although not an issue in Canada, it is known to affect other areas of the world (Environment Canada, 2013a). Overdrafting is an anthropogenic cause of groundwater level reduction/depletion resulting from withdrawal rates that are higher than natural recharge rates (Environment Canada, 2013a). Excessive withdrawal of groundwater not only has a negative impact on the sectors that depend on it through water table reduction and the related increase in pumping cost, it can also negatively affect the land and infrastructure through land subsidence. Land subsidence is the process in which land sinks due to groundwater depletion (Environment Canada, 2013a; United States Geological Survey, 2000). The potential for this enforces the need for adequate monitoring and network design.

Water that reaches groundwater from any direction can be defined as recharge. When choosing a model to estimate recharge in an area, the climate, geomorphology, and geology of that area need to be considered as they control the location and timing of recharge (Scanlon, Healy, & Cook, 2002). Common characteristics of humid regions include shallow water tables and gaining streams (Scanlon et al., 2002). Diffuse recharge, used to describe recharge derived from precipitation or irrigation that occurs over large areas, is dominant in humid regions (Scanlon et al., 2002). Recharge is generally lower in vegetated regions; in these regions however, recharge is higher in areas where annual crops and grasses are present, and lower in areas of trees and shrubs (Scanlon et al., 2002). This gives rise to the need for accurate land use and land cover data. Surface water and saturated-zone models such as Precipitation-Runoff Modeling System (PRMS) and Hydrological Simulation Program – Fortran (HSP-F) are more widely used in humid regions to calculate groundwater recharge (Scanlon et al., 2002; United States Geological Survey, 2013), however other methods do exist to calculate recharge (Healy & Cook, 2002; Scanlon et al., 2002; United States Geological Survey, 2013). Groundwater recharge was chosen as an additional objective in DEMO to help quantify the hydrological characteristics of the vadose zone in the aquifer and thus provide more information for the optimization algorithm. This is important in finding the optimal well locations for the optimal monitoring network. Two models, PRMS and HSP-F, were used to calculate the groundwater recharge used in this study.

3. Study Area and Data

The study area is located in southern Ontario and consists of the watersheds managed by three conservation authorities: Hamilton Conservation Authority (Spencer Creek watershed), Conservation Halton (Bronte Creek and Sixteen Miles Creek watersheds), and Credit Valley Conservation (Credit River watershed), henceforth referred to as the Hamilton-Halton-Credit Valley (HHCV) region. These combined watersheds have a surface area of approximately 2 300 km^2 and contain approximately 80% rural agricultural and forested land as well as 20% urban land (Figure 1) concentrated around the Lake Ontario shoreline. There are 26 groundwater level monitoring wells (Figure 2) which are part of the PGMN of Ontario within the study area; a summary of the wells can be found in Table 1. The soils are predominantly loam types (Figure 3) and the topography is generally flat to rolling hills with the exception of the Niagara Escarpment (Figure 4) (Kornelsen & Coulibaly, 2013; Natural Resources Canada, 2015). Within this study area are portions of several major Ontario aquifers including the Credit River, Oak Ridges Moraine, and Grand River Basin aquifers (Figure 5) (Natural Resources Canada, 2014). Geologic mapping in the area indicates six geological formations, those being the Armabel Formation, Clinton Group, Georgian Bay Formation, Guelph Formation, Lockport Formation, and Queenston Formation (Figure 5); with the major rock types being sandstone, shale, dolostone, siltstone, and limestone (Government of Ontario, 2013a). Hydrostratigraphic analysis of the area indicates the existence of multiple aquifers and aquitards in both the overburden and the bedrock (AquaResources Inc, 2009; Earthfx Inc, 2010a, 2010b). The climate of southern Ontario is characterized by warm summers and mild winters, with the average air temperature varying from 21.2°C to -5.2°C, and the average annual precipitation is 896.4 mm (1981 – 2010 Canadian Climate Normals). Groundwater level data for this study was obtained from PGMN, this data is available as hourly meters above sea level (masl) measurements, but was converted to average daily masl.

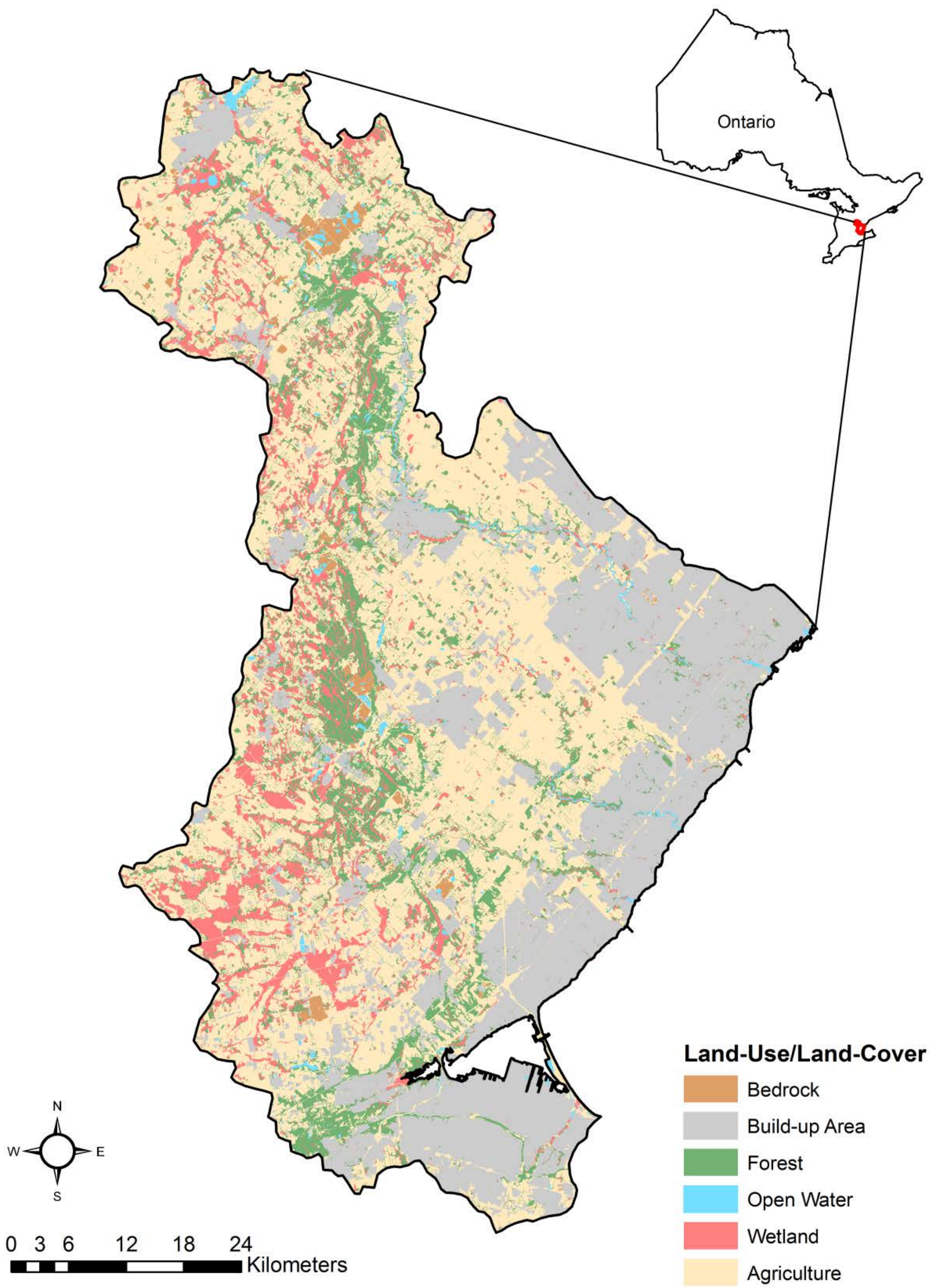


Figure 1: Land use and land cover map of HHCV region of Ontario (Natural Resources Canada, 2009)

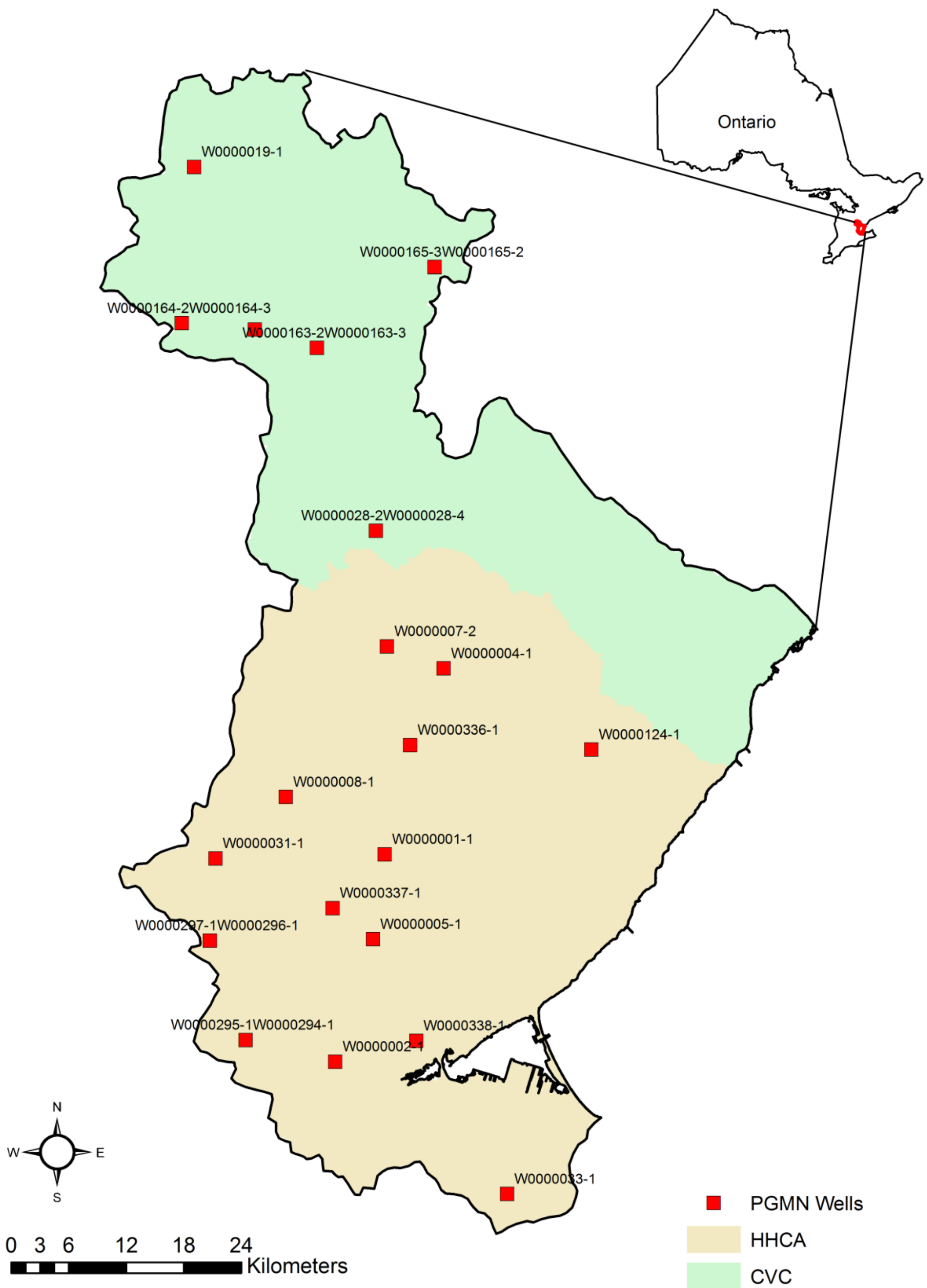


Figure 2: PGMN wells in the Hamilton-Halton Conservation Authority (HHCA) and Credit Valley Conservation Authority (CVC) regions (Government of Ontario, 2013b)

Table 1: PGMN information for wells in the HHCV regions (Government of Ontario, 2013b)

PGMN Well ID	Aquifer Type	Aquifer Lithology	Depth (m)	Latitude	Longitude	Surface Elevation (masl)
W0000001-1	Bedrock	Dolostone	6.30	43.42459	-79.94452	262.78
W0000002-1	Bedrock	Dolostone	23.20	43.28354	-79.99080	241.29
W0000004-1	Overburden	Silt, Sand, Gravel	72.08	43.55075	-79.88927	212.75
W0000005-1	Overburden	Gravel	12.80	43.36691	-79.95546	263.00
W0000007-2	Overburden	Sand	14.33	43.56572	-79.94224	246.33
W0000008-1	Bedrock	Limestone	6.61	43.46354	-80.03700	300.10
W0000019-1	Overburden	Sand, Gravel, Clay	19.20	43.88992	-80.12286	448.85
W0000026-1	Bedrock	Limestone	35.96	43.78018	-80.06601	392.96
W0000028-2	Overburden	Gravel, Sand	6.71	43.64408	-79.95276	279.44
W0000028-4	Overburden	Gravel, Sand	24.38	43.64408	-79.95276	279.44
W0000031-1	Bedrock	Limestone	27.43	43.42179	-80.10300	295.88
W0000033-1	Bedrock	Limestone	16.46	43.19329	-79.82986	192.17
W0000124-1	Overburden	Sand, Silt	14.94	43.49575	-79.75093	195.04
W0000163-2	Overburden	Silty Sand, Sandy Silt	8.14	43.76771	-80.00788	426.84
W0000163-3	Overburden	Gravel	23.16	43.76771	-80.00788	426.84
W0000164-2	Overburden	Silty Sand, Sandy Silt, Sand	8.08	43.78441	-80.13438	427.76
W0000164-3	Overburden	Sand, Gravel	13.26	43.78441	-80.13438	427.76
W0000165-2	Overburden	Clayey Silt	6.10	43.82235	-79.89776	281.72
W0000165-3	Overburden	Sand, Gravel	174.04	43.82235	-79.89776	281.72
W0000294-1	Overburden	Sand	6.10	43.29826	-80.07481	245.88
W0000295-1	Bedrock	Dolostone	11.00	43.29826	-80.07481	245.88
W0000296-1	Bedrock	Dolostone	11.80	43.36596	-80.10821	268.96
W0000297-1	Overburden	Sandy Silt, Clayey silt	6.00	43.36596	-80.10821	268.96
W0000336-1	Bedrock	Limestone, Shale	36.58	43.49881	-79.92068	317.06
W0000337-1	Bedrock	Dolostone	27.60	43.38804	-79.99331	255.74
W0000338-1	Overburden	N/A	3.14	43.29775	-79.91492	148.41

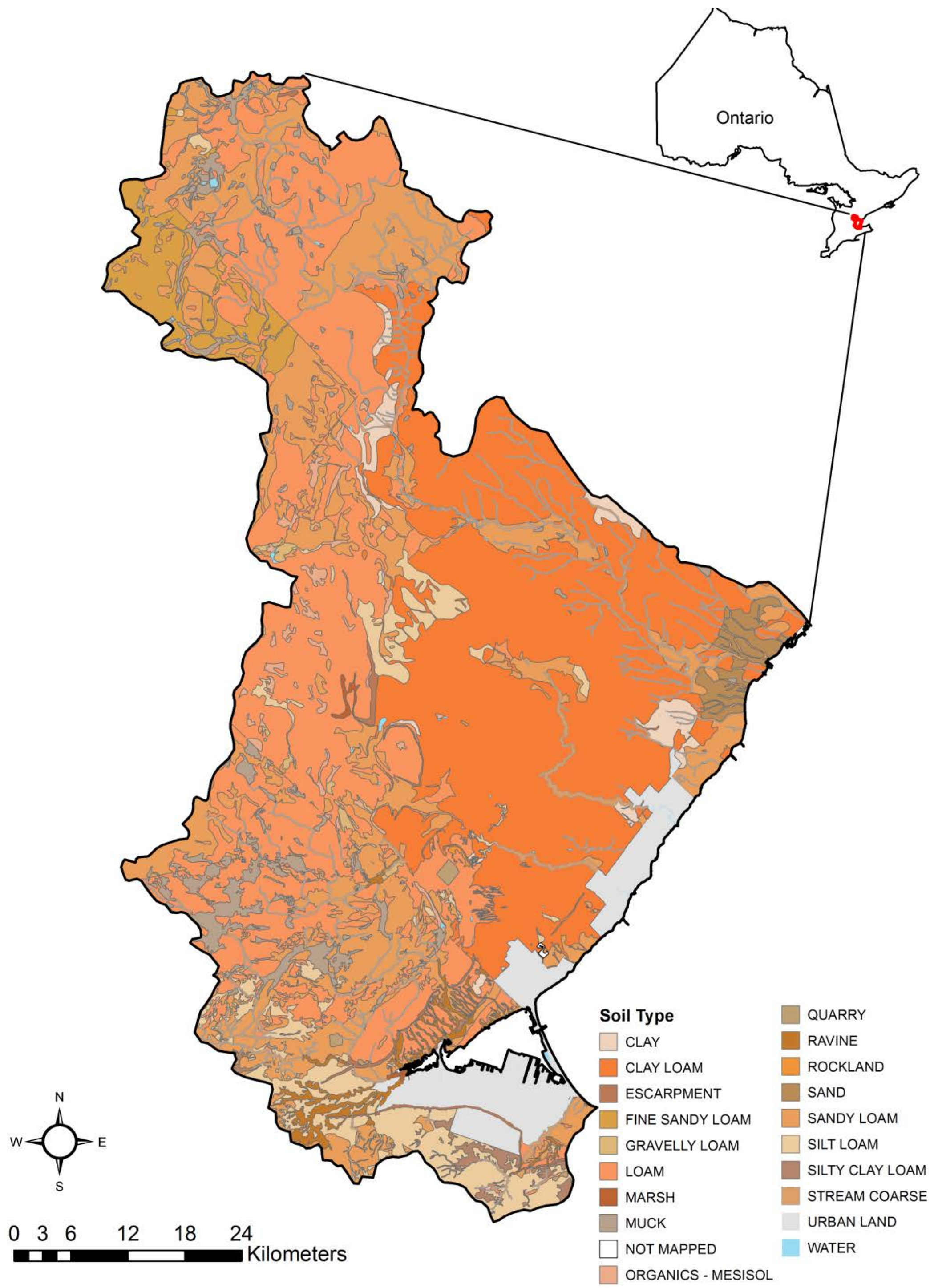


Figure 3: Soil types in the HHCV region of Ontario (Ontario Ministry of Natural Resources, 2012)

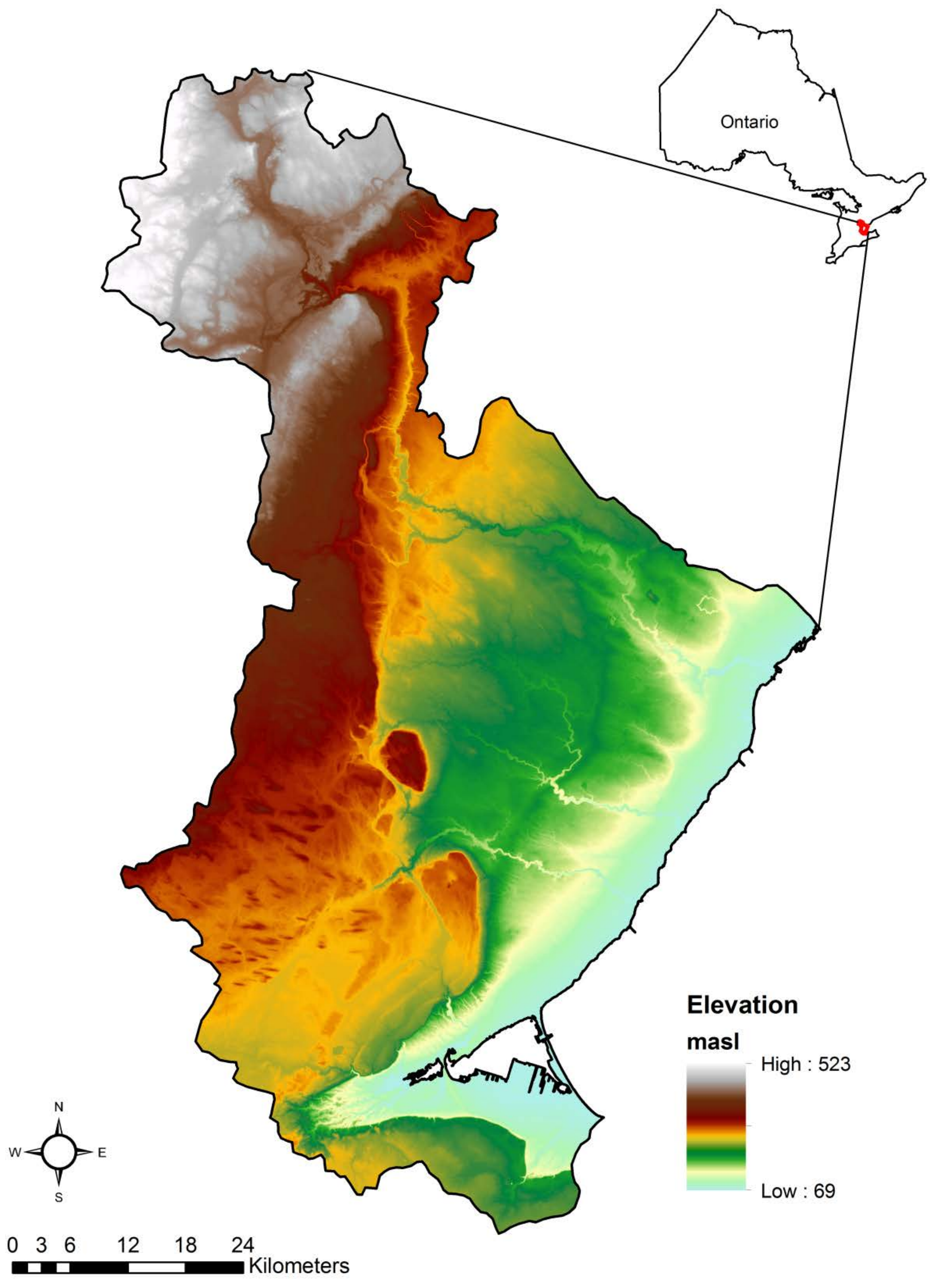


Figure 4: Digital Elevation Model of the HHCV region in Ontario (Natural Resources Canada, 2015)

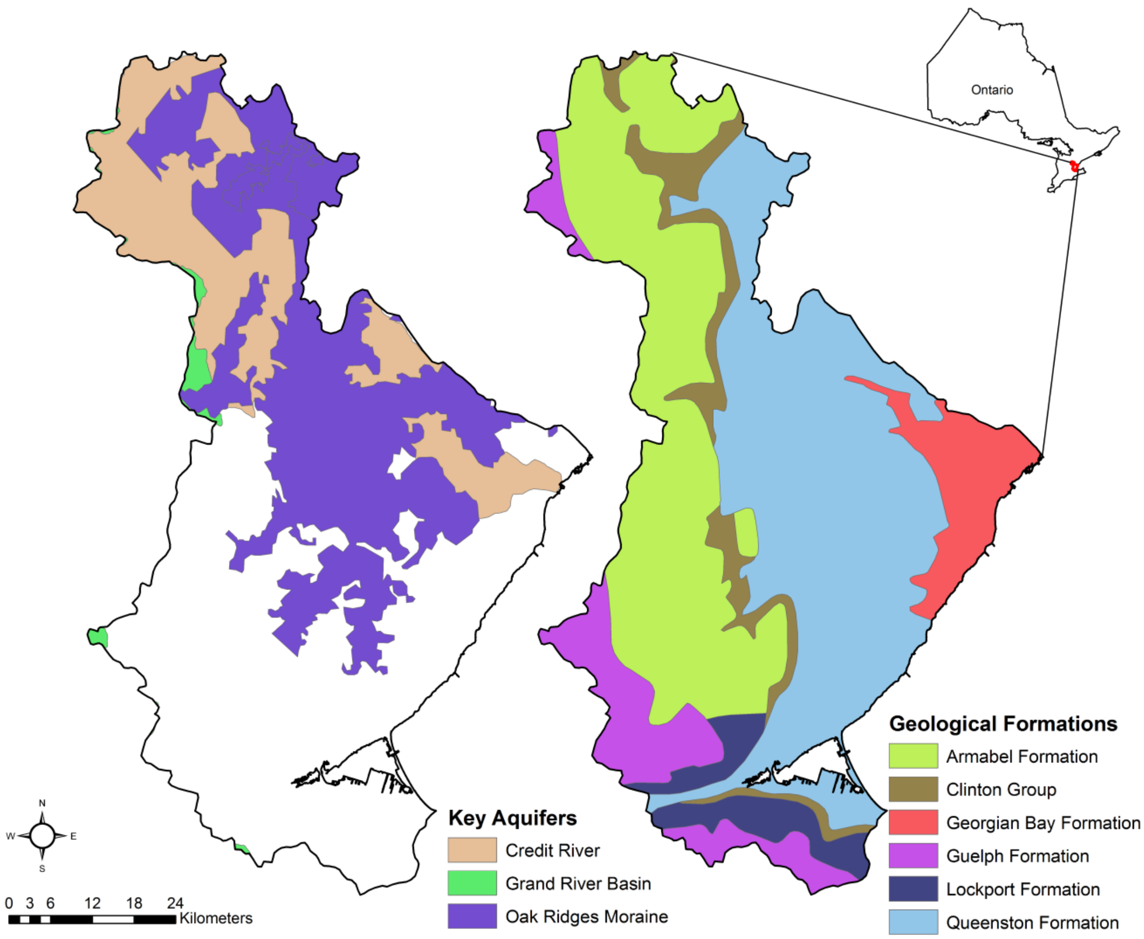


Figure 5: Key Canadian aquifers (left) and geological formations (right) in the HHCV regions of Ontario (Government of Ontario, 2013a; Natural Resources Canada, 2014)

4. Methodology

4.1. Data Processing

There are examples in the literature on network design where a standard grid is overlaid on the study area to give potential locations for new monitoring sites (Cieniawski et al., 1995; Hudak & Loaiciga, 1992; Mahar & Datta, 1997; Meyer et al., 1994; Wagner, 1995); the same method was implemented here. To determine locations for new wells in the study area, a 10 x 30 km grid was generated in ESRI's ArcMAP over the study area; the centroid of each grid cell was then used as the location for each new potential monitoring well. Using this method a total of 144 potential well locations were generated, as shown in Figure 6.

The groundwater level time series data available for use had hourly measurements for dates ranging between 2001 and 2012 however there were significant gaps in the data. This hourly data was averaged into daily values and then the groundwater level data from 2006 to 2010 was used for this study since it has a low percentage of missing data while still being long enough to generate adequate variability in the entropy measurements. A summary of missing data for 2001 to 2012 can be found in Appendix A. This time frame has the additional benefit of being able to represent an area which is data poor and in need of additional wells. If the algorithm is able to find optimal networks using the lower amounts of time series data, it would likely be usable in a study area that is data poor. The missing data in this time frame was filled in using a simple regression method (linear regression). The wells chosen for the linear regression infilling were based on which well within the study area, and up to 10 km around, had the highest time series correlation to the well which had missing data. A summary of the monitoring wells, their missing data, and the correlation with the monitoring well used for infilling can be found in Table 2.

Table 2: Infill summary highlighting percentage of missing data infilled for each monitoring well and the correlation between the data being infilled and the data being used to infill

Monitoring Well	Filled from	Correlation Coefficient	Percentage of Missing Data
W0000001-1	W0000031-1	0.89	9%
W0000002-1	W0000163-2	0.66	8%
W0000004-1	W0000005-1	0.80	6%
W0000005-1	W0000295-1	0.96	10%
W0000007-2	W0000003-1	0.76	4%
W0000008-1	W0000003-1	0.81	5%
W0000019-1	W0000023-1	0.94	27%
W0000026-1	W0000023-1	0.67	25%
W0000028-2	W0000003-1	0.58	30%
W0000028-4	W0000003-1	0.79	48%
W0000031-1	W0000297-1	0.90	31%
W0000033-1	W0000288-1	0.91	10%
W0000124-1	W0000007-2	0.85	23%
W0000163-2	W0000003-1	0.90	24%
W0000163-3	W0000003-1	0.87	31%
W0000164-2	W0000023-1	0.82	41%
W0000164-3	W0000023-1	0.78	35%
W0000165-2	W0000023-1	0.88	33%
W0000165-3	W0000366-1	0.80	31%
W0000294-1	W0000003-1	0.89	1%
W0000295-1	W0000296-1	0.98	12%
W0000296-1	no infill	-	0%
W0000297-1	W0000031-1	0.90	2%
W0000336-1	W0000164-3	0.83	1%
W0000337-1	W0000005-1	0.94	5%
W0000338-1	W0000027-1	0.87	0%

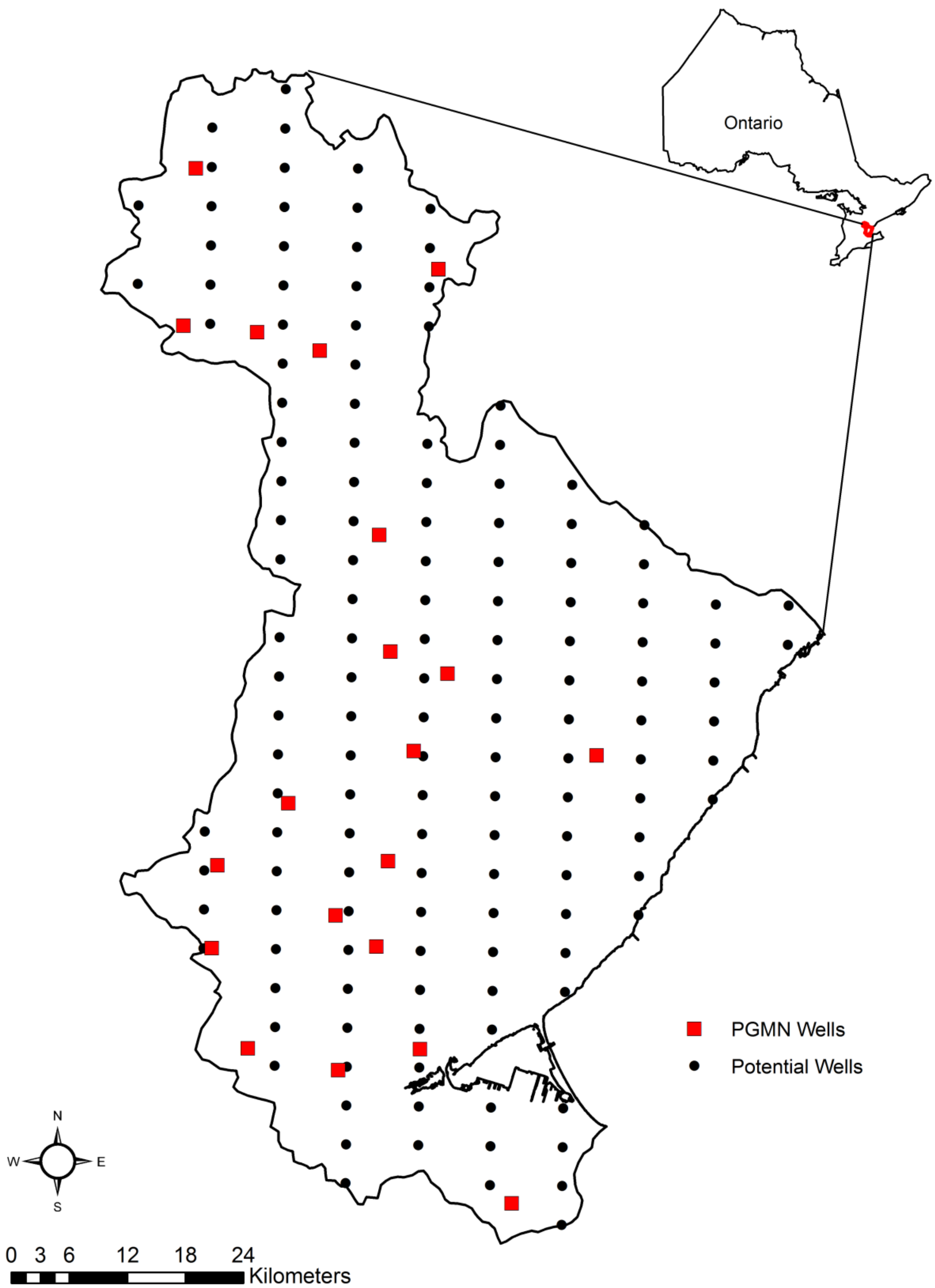


Figure 6: Potential Well locations generated using the centroids of a 10 x 30 km grid in ArcMap

Using the infilled groundwater level data for all the existing wells, the groundwater level data for each of the potential well locations was then generated using Inverse Distance Weighting (IDW), a spatial proximity interpolation method, to generate data at each potential well location (W_u). The IDW method uses the following equations adapted from (Shepard, 1968):

$$W_u = \sum_{i=1}^{n} w_i W_i \tag{1}$$

$$w_i = \frac{h_i^{-2}}{\sum_{i=1}^{n} h_i^{-2}} \tag{2}$$

Where n is the number of existing wells, w_i is the weighting for the i^{th} existing monitoring well, h_i is the distance between the i^{th} existing well and the potential monitoring well location, and W_i is the groundwater level at the i^{th} existing monitoring well.

4.2. DEMO Approach

4.2.1. Epsilon-dominance hierarchical Bayesian optimization algorithm

DEMO uses a multi-objective evolutionary algorithm to evolve a set of network solutions which are determined to be the optimal network solutions based on the objective functions. This set of optimal network solutions is found from the total number of all possible network solutions based on the binomial coefficient, ${}_nC_k$, where n is the total number of potential stations in the network, and k is the number of stations being chosen. In the case of this study, n is 144 and k is 10, thereby the optimal network solutions can be found from a total of 7.67×10^{14} possible network solutions. The evolutionary algorithm used in DEMO is the epsilon-dominance hierarchical Bayesian optimization algorithm (ε-hBOA) from (2008). The ε-hBOA uses Bayesian network models to express and preserve the interdependencies of the decision variables through the evolutionary process (2008). A brief explanation of the steps followed by the ε-hBOA adapted from (Kollat et al., 2008; Samuel et al., 2013) is as follows:

1. Generate an initial (parent) population of random solutions (or using previous generation superior solutions as parents);
2. Construct a Bayesian network based on the parent solutions;
3. Generate child solutions from the Bayesian network's joint probability distribution;
4. Combine parent and child solutions;
5. Sort the population based on Pareto-dominance and assess the fitness of each solution;
6. Apply Pareto ranking and crowded binary tournament selection to parent and child solutions to find a new population of superior solutions which are combined with archived non-dominated solutions;
7. Apply dynamic population sizing to allow the population size to change with problem difficulty;
8. Repeat steps 1 to 7 until termination.

ε-hBOA will search for Pareto-optimal solutions based on either the maximization or minimization of a set of objective functions selected for a particular problem or network. The first two objective functions which are fundamental to DEMO are the joint entropy and total correlation, which provide a measure of the information content of the network (Section 4.2.2 covers information theory as it relates to DEMO).

The multi-objective optimization problem solved using ε-hBOA can be presented mathematically (Alfonso et al., 2010; Samuel et al., 2013) as:

$$\min\{C(S_{M,N}) = C(X_1, X_2, ..., X_M, E_1, E_2, ..., E_N)\} \quad (3)$$

$$\max\{H(S_{M,N}) = H(X_1, X_2, ..., X_M, E_1, E_2, ..., E_N)\} \quad (4)$$

Where $X=\{X_1,X_2,\ldots,X_M\}$ is the set of chosen potential monitoring wells which are set to complement the existing network so that it is optimal, $E=\{E_1,E_2,\ldots,E_N\}$ is the set of existing monitoring wells, $S_{M,N}=\{X_1,X_2,\ldots,X_M,E_1,E_2,\ldots,E_N\}$ is the combination of existing and potential monitoring wells that represent the new optimal network, $C(S_{M,N})$ is the total correlation value of the network (equation 7), and $H(S_{M,N})$ is the joint entropy of the network (equation 6) (Alfonso et al., 2010; Samuel et al., 2013).

More information related to the operation and details of the ε-hBOA can be found in (2008) and (2013), readers who are interested are encouraged to refer to those documents. The selected model parameters used with ε-hBOA in this study are listed in Table 3; they were chosen based on the recommendations of (2008). To allow variability in the results, each model was run with fifty random seeds. The Shared Hierarchical Academic

Research Computing Network (SHARCNET) was used to run the fifty random seeds simultaneously; the results of which were concatenated to form the final Pareto optimal solutions (or optimal networks).

Table 3: Selected ε-hBOA model parameters based on literature values (Kollat et al., 2008; Samuel et al., 2013)

Model Parameter	**All**	**Shallow**	**Deep**
Initial population size	10 000	10 000	10 000
Min. population size	10 000	10 000	10 000
Max. population size	100 000	100 000	100 000
Population sizing scheme	Injection	Injection	Injection
Population scaling factor	0.25	0.25	0.25
Number of decision variables (n*)	170	160	154
Max. generations (2n*)	340	320	308
ε for min. total correlation	0.0001	0.0001	0.0001
ε for max. joint entropy	0.0001	0.0001	0.0001
ε for max. recharge	0.0001	0.0001	0.0001
ε for max. water table variation	0.0001	0.0001	0.0001
Crossover probability for real variables	1	1	1
Mutation probability for real variables (1/n*)	0.00588	0.00625	0.00649
Distribution index for SBX crossover	15	15	15
Distribution index for polynomial mutation	20	20	20
Max. number of function evaluations	1 000 000	1 000 000	1 000 000
Number of random seeds run	50	50	50

*n is the sum of existing and potential wells

4.2.2. Entropy Method Application

Entropy, or uncertainty, has been discussed in the literature for water resources and hydrology previously (Alfonso et al., 2010; Harmancioglu & Singh, 1998; Husain, 1979, 1989; Mishra & Coulibaly, 2009, 2010; Mogheir et al., 2009; Mogheir & Singh, 2002; Samuel et al., 2013; Singh, 1997); it is a nonparametric method with no a priori assumptions about the data being used (Mishra & Coulibaly, 2010; Samuel et al., 2013). In Information Theory, the Shannon entropy (or marginal entropy) $H(X)$ (Alfonso et al., 2010; Samuel et al., 2013; Shannon, 1948) of a discrete random variable (monitoring well) X provides a measure of the information content from a finite sample, in bits, for a single station, shown as follows:

$$H(X) = -\sum_{i=1}^{n} P(x_i)\log_2 P(x_i) \tag{5}$$

Where $P(x_i)$ is the probability of the value x_i in the bin of values n. The number of events (binned values) n was determined by discretizing the groundwater table level time series over a number of bins. The groundwater table data available in this study area, presented in masl, varied by approximately 1 to 7 meters; this variation does not provide enough information to the optimization algorithm to pick out unique values. To help the algorithm account for the variation in each well the values were instead presented as decimeters. To then bin these values, they were rounded up and each decimeter value between the minimum and maximum water levels were used as bins, which gave a variation of binned values between 1 and 72. The entropy can be expanded to provide the joint entropy H($X_1,\ldots,X_N$) (Alfonso et al., 2010; Samuel et al., 2013) for multiple variables (monitoring stations) $X_1,\ldots,X_N$, which is described as follows:

$$H(X_1,\ldots,X_N) = -\sum_{i=1}^{n1}\cdots\sum_{k=1}^{nN} P(x_{1,i},\ldots,x_{N,k})\log_2 P(x_{1,i},\ldots,x_{N,k}) \tag{6}$$

Where $P(x_{1,i},\ldots,x_{N,k})$ is the probability of the values $x_{1,i},\ldots,x_{N,k}$ from the bins of values $n1,\ldots,nN$ for each variable (monitoring well). Maximizing the joint entropy provides the first information based objective function for DEMO. It is a measure of the unique multi-site groundwater level states in the study area that would be measured by a candidate network.

The second information content based objective function is the total correlation (Alfonso et al., 2010; Samuel et al., 2013) which is a measure of the redundant information between each of the variables (monitoring wells), described as:

$$C(X_1,...,X_N) = \left[\sum_{i=1}^{N} H(X_i)\right] - H(X_1,...,X_N) \tag{7}$$

Minimization of the total correlation results in a network with the least amount of shared or redundant information between monitoring wells. Total correlation is the multi-variate extension of the mutual information and is a cumulative measure of how much information would be lost from a network if a particular station/well were removed. These two objectives provide much of the information used by ε-hBOA to search for an optimal network based on the temporal groundwater level information.

4.2.3. Additional Objectives in DEMO

In this study two additional objectives were used in DEMO; water table variation (discussed in section 4.3) and groundwater recharge (discussed in section 4.4). These objectives were chosen to help quantify the different hydrological/hydrogeological characteristics of the vadose zone in the study area aquifers as well as the groundwater use and potential changes in the normal seasonality of the groundwater levels. These additional objectives will help to provide more information for the ε-hBOA to use in determining the optimal monitoring well placements for augmenting or designing a groundwater monitoring network. To use these objective functions in DEMO, the values for each monitoring well, both existing and potential, were first normalized using the feature scaling method as follows:

$$\hat{x}_i = \frac{(x_i - x_{\min})}{(x_{\max} - x_{\min})}, x \in \mathrm{X} \tag{8}$$

Where x is the parameter, and $\hat{x}$ is the normalized parameter which has been normalized by the minimum and maximum values of the parameter set X.

The total Euclidean distance for each additional objective for the chosen network design was then used to measure the amount of information provided by the said objective. The Euclidean distance values for the chosen networks were calculated as follows:

$$d(S_{M,N}) = \sum_{j=1}^{L} \sum_{k=j}^{L} \sqrt{\sum_{i=1}^{n} (\hat{x}_{i,j} - \hat{x}_{i,k})^2} \tag{9}$$

Where $\hat{x}_{i,j}$ represent the normalized objective value i for monitoring well j, L is the number of monitoring wells in the optimal network ($S_{M,N} = X_1, X_2, \ldots, X_M, E_1, E_2, \ldots, E_N$, $L=M+N$), n is the number of objective values in the objective function (Note: if $n = 1$ the objective is found as the difference between values instead of the Euclidean distance between them), and $k = j$ such that a comparison between each monitoring well objective value was made only once. In the ε-hBOA, this objective is used as part of the Pareto optimization process; the maximum Euclidean distance is used in the algorithm by:

$$\max\{d(S_{M,N}) = d(X_1, X_2, \ldots, X_M, E_1, E_2, \ldots, E_N)\} \tag{10}$$

Maximizing the multi-variate Euclidean distance allows for the areas with the most differences (more informative locations) to be highlighted.

4.3. Water Table Variation

Groundwater levels have normal fluctuations which can be attributed to the seasonality of precipitation, irrigation, and evapotranspiration (Healy & Cook, 2002). The water table variation objective is represented by four informative variables that have been extracted from each monitoring well's time series data. They are the annual maximum and minimum 1-day groundwater levels as well as the Julian day in which these annual maximum and minimum values occurred. These values were chosen as the median annual maximum and minimum values for the chosen time period in this study, 2006 to 2010. A cross-correlation analysis of these values was performed to show their differences, as the lower the values are correlated with each other the higher the potential for unique information added, shown in Table 4.

Table 4: Cross-correlation of water table variation values for existing and potential wells in HHCV study area for the chosen study period

	All Well Network				**Shallow Well Network**				**Deep Well Network**			
	Max	JD* max	Min	JD min	Max	JD max	Min	JD min	Max	JD max	Min	JD min
Max	1.00				1.00				1.00			
JD max	-0.22	1.00			-0.34	1.00			-0.01	1.00		
Min	1.00	-0.22	1.00		1.00	-0.34	1.00		1.00	0.00	1.00	
JD min	-0.12	0.05	-0.13	1.00	-0.24	0.09	-0.24	1.00	0.02	0.08	0.02	1.00

*Julian day in which maximum or minimum value occurred

The purpose of this objective function is to capture the seasonal variability and highlight the normal seasonal maximum and minimum groundwater levels with the idea that wells in the study area that do not follow the norm indicate some process which is impacting the groundwater and indicates potential changes that should be monitored. This objective function was used in Scenario 2.

4.4. Groundwater Recharge Estimation Models

4.4.1. Precipitation-Runoff Modeling System

The Precipitation-Runoff Modeling System is a deterministic, distributed parameter modeling system that does not require a groundwater flow model (Cherkauer, 2004; Leavesley, Lichty, Troutman, & Saindon, 1983). Data required to run PRMS includes daily streamflow, precipitation, and maximum and minimum temperatures, as well as land use/land cover, soil maps, and elevation data. The PRMS was used to estimate groundwater recharge distribution in the HHCV study area. In PRMS, the groundwater recharge can be found as a combination of the groundwater sink and the groundwater discharge (Leavesley et al., 1983). PRMS and GIS (Geographic Information Systems) was used by (Cherkauer, 2004) to quantify groundwater recharge at multiple scales and presented a procedure to define most inputs from GIS and hydrological inputs to simplify calibration by reducing the degrees of freedom. A simplification of the method used to estimate groundwater recharge in PRMS is as follows:

1. Delineation of hydrologic response units in study area;
2. Parameterization;
3. Calibration and validation;

4. The sum of groundwater discharge and groundwater sink is equal to recharge value.

Using this model, the simulated annual recharge in millimeters per year was determined for the study area.

Scenario 3 and 4 both used recharge as additional objectives in DEMO; however the methods and models varied for each. For Scenario 3, recharge was calculated using PRMS for a modeling period of 1991 to 2010, where 1991 to 2000 was used for calibration and 2001 to 2010 for validation. For Scenario 4 a combination of groundwater recharge values that were calculated using two models was used; PRMS and HSP-F. The modeling period for PRMS in this case was from 1989 to 1997 (Halton-Hamilton Source Protection, 2010). The PRMS groundwater recharge values used in Scenario 4 were provided by the HHCA.

4.4.2. Hydrological Simulation Program - Fortran

The Hydrological Simulation Program – Fortran is a continuous simulation model, developed to simulate hydrologic and water quality processes (Bicknell, Imhoff, Kittle Jr., Donigan Jr., & Johanson, 1996). Data required to run the model include meteorological time series data, topography, land use/land cover, and soil type. Through calculating the infiltration and net evapotranspiration, HSP-F is able to determine the groundwater recharge (AquaResources Inc, 2009; Bicknell et al., 1996). This method provides the simulated annual recharge in millimeters per year. The modeling period used when calculating groundwater recharge with HSP-F was from 1997 to 2004, where 1997 to 2000 was used for calibration and 2001 to 2004 for validation (AquaResources Inc, 2009). The HSP-F groundwater recharge values used in Scenario 4 were provided by the CVC.

5. Results

Spatial probability plots are used to better represent the results of the DEMO model. These plots are generated using the IDW function of ESRIs ArcMAP. They illustrate the locations in the study area that additional wells are more likely to be placed to produce an optimal monitoring network. The probability of monitoring well selection was determined by the number of occurrences that an individual potential well was selected in a Pareto-optimal solution compared to the total number of times it could have been selected in the final population of solutions. These spatial probability plots do not represent an optimum network, but rather the combined likelihood that each monitoring well is chosen among all the Pareto optimal networks found. When selecting a single network design for development it would be recommended to select one which had many high probability stations. Available in Appendices B, C, and D are some examples of the individual network designs and their corresponding Pareto fronts. A summary of the Scenarios is shown in Table 5.

Table 5: Summary of objectives used for each scenario

Scenario	Total Correlation	Joint Entropy	Water Table Variation	Recharge PRMS	Recharge PRMS+HSPF
1	x	x			
2	x	x	x		
3	x	x		x	
4	x	x			x

Figure 7 highlights the number of optimal solutions found on the Pareto front for each scenario. The scenarios were repeated for three sets of groundwater level data, those that encompass all the PGMN well data, as well as those being classified as either shallow (<15 meters below the surface) or deep wells (>15 meters below the surface). This split allowed for three sets of results for each scenario. Of the 26 existing PGMN wells, 16 were classified as shallow and 10 were classified as deep. For all three network sizes, 10 additional monitoring wells were added using DEMO. Using a range of network sizes will allow for a better understanding of how using the additional objective functions in DEMO can affect the optimal network designs. From Figure 7, it can be seen that the inclusion of additional informative objectives has an impact on the possible number of optimal monitoring networks in the study area, and potentially the spatial locations of the wells in each network.

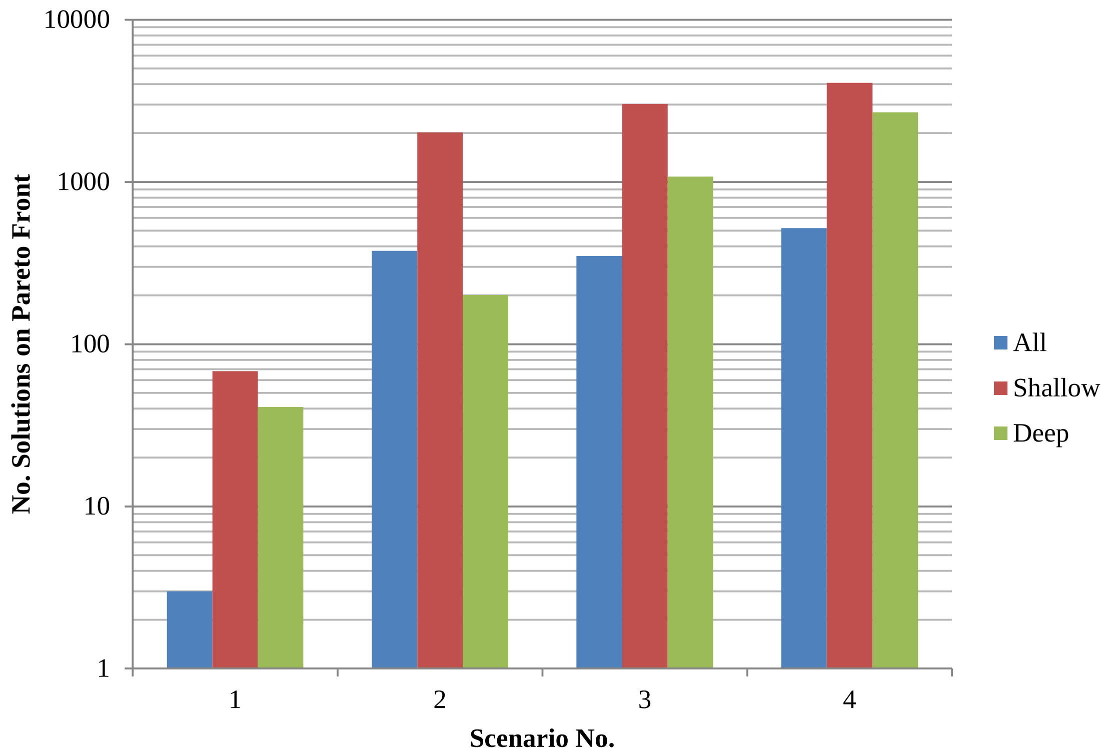


Figure 7: Number of Pareto optimal solutions for each Scenario run in DEMO

Two separate sets of groundwater recharge results were used in this study as additional objective functions. Shown in Figure 8a are the groundwater recharge values (produced by PRMS only) that were used for Scenario 3, and shown in Figure 8b are the groundwater recharge values (produced by both PRMS and HSP-F) used for Scenario 4. The recharge varies greatly between the two results, with values ranging from 31 to 2905 mm/year compared to 44 to 397 mm/year for Figures 8a and 8b respectively. With the PRMS only results of Figure 8a being varied both in spatial distribution and magnitude, and the joint PRMS and HSP-F results of Figure 8b having the higher recharge values corresponding the higher elevations of the study area (above the escarpment) and the lower values corresponding to the lower elevation and more urban areas.

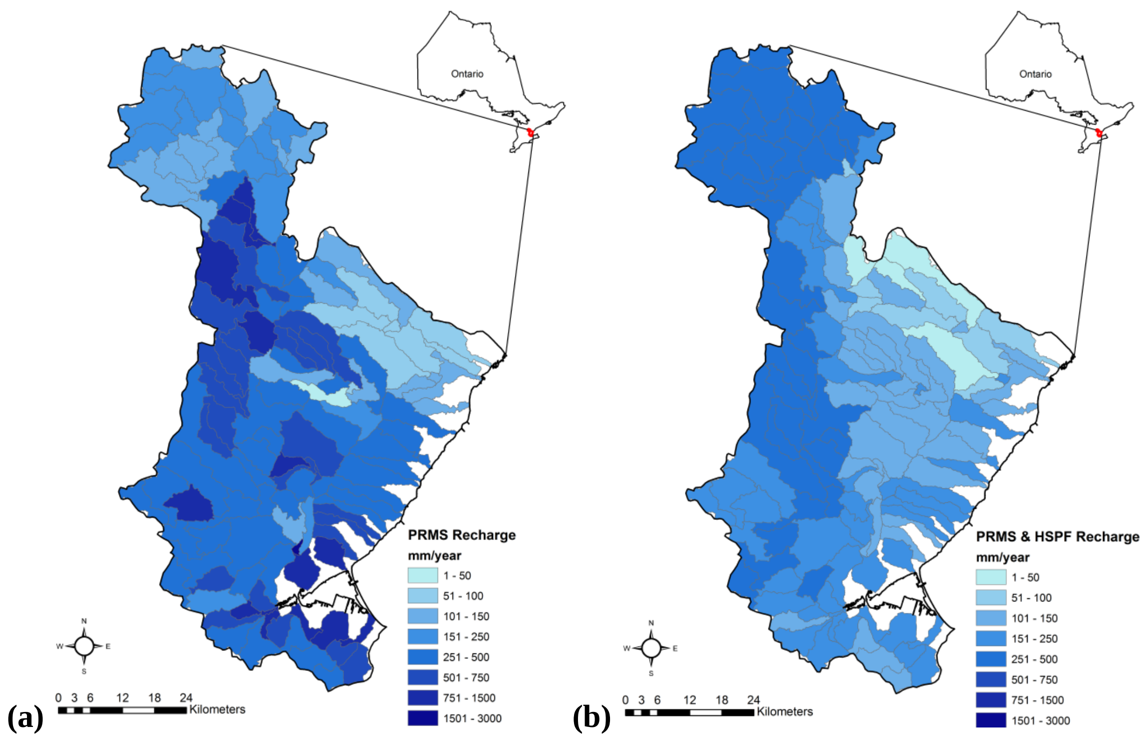


Figure 8: Simulated annual recharge (mm/year) calculated using different groundwater recharge estimation models. (a) Was found using only the PRMS model, and (b) was found using a combination of the PRMS and HSP-F models

5.1. Standard DEMO Results

The results of DEMO using only the standard objective functions, joint entropy and total correlation, are illustrated in Figure 9. Included in this figure are the Scenario 1 results for augmenting the existing PGMN wells (Figure 9a), the wells classified as shallow (Figure 9b), as well as those classified as deep (Figure 9c). Using the results illustrated in

Scenario 1 as a reference point, the impacts that including additional objective functions will have on the design of optimal networks will be more evident.

The standard entropy results of Figure 9a, the probability map for 3 possible optimal network designs, show that little information is added when increasing the network size from 26 to 36, and thus there is little spatial variability of the resultant networks. The additional wells are instead placed in close proximity to the existing wells, which can be accounted for by the limitation of the chosen data generation method. In this case, the standard entropy method is not enough for augmenting the existing PGMN, and additional objective functions are needed to produce more spatially varied monitoring wells. Illustrated in Figure 9b is the probability map for the 68 optimal network designs to accommodate an increase in network size of 10 monitoring wells, for a total of 26 wells. It can be seen that the additional wells are more likely to be placed in the wetland, agricultural, and rural areas in the higher elevations of the study area. Finally, Figure 9c illustrates the probability map for the Scenario 1 run which considered only the deep wells in the study area. It represents the possible 41 optimal network designs of a network that includes 10 additional deep wells for a total of 20. The areas that are more likely to have an additional well are generally in the higher elevations of the study area (above the escarpment). These results show that the entropy functions give a variety of possible optimal networks for both shallow and deep wells; however they are somewhat limited to the proximity around existing wells which is a potential drawback of the IDW method for data generation.

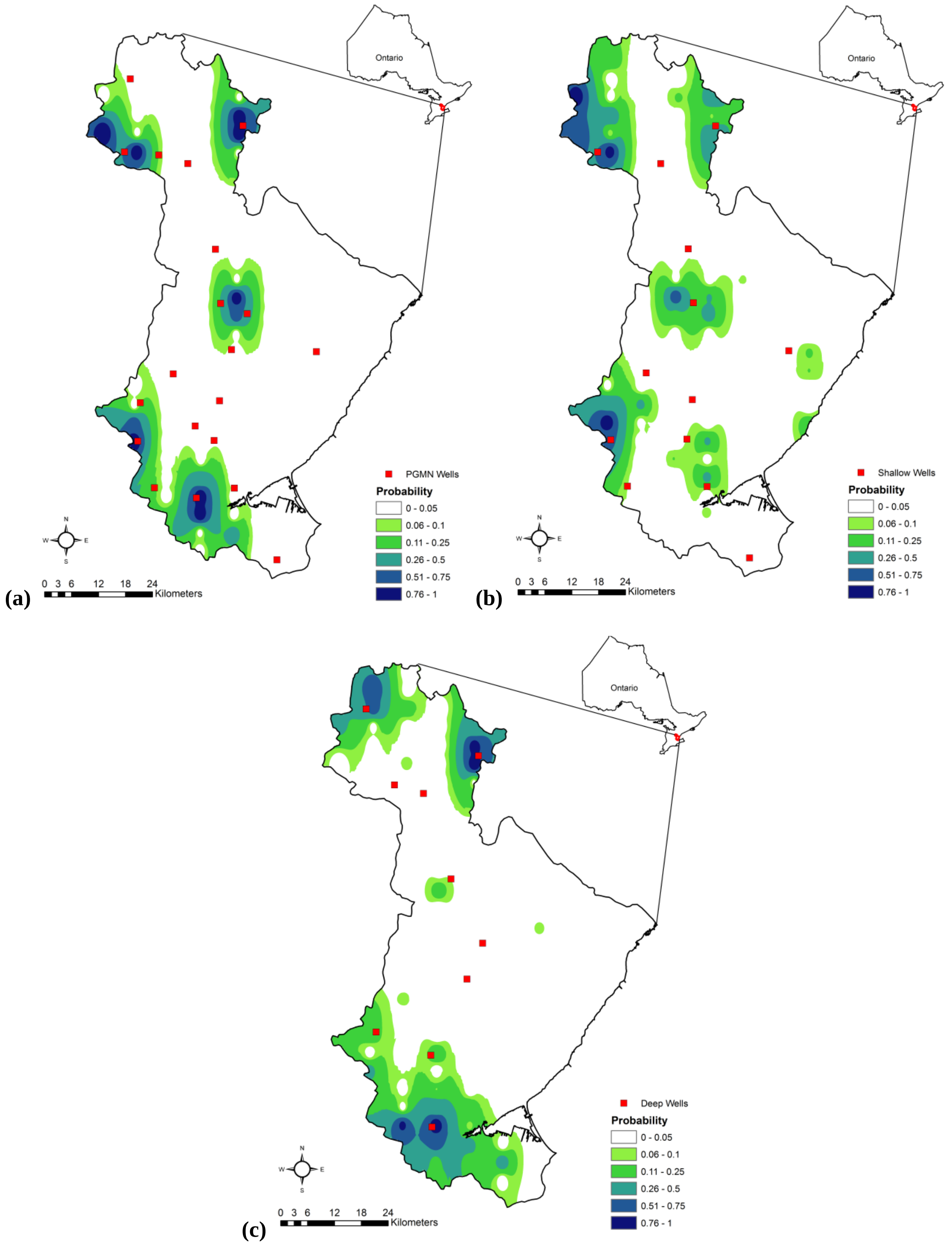


Figure 9: Scenario 1 results for (a) all PGMN wells, (b) PGMN wells classified as shallow, and (c) PGMN wells classified as deep

5.2. DEMO Results with Water Table Variation

Figure 10a illustrates the results of including the water table variation objective function in DEMO to augment the existing PGMN wells in the HHCV study area. The areas in which the additional wells were more likely to be placed, based on the probability plots produced using the 376 Pareto optimal solutions, were mostly in the higher elevation areas (above the escarpment) in the study region as well as around the agricultural and urban built up areas. This is a likely indication that agricultural practices and urban development have had impacts on the water table in the area and these probability hot spots are areas in which additional monitoring wells could be placed to monitor those changes/impacts. These results also show there is a noticeable increase in the spatial variability of new station locations and spatial coverage of the network when compared to the standard DEMO results of Figure 9a. This may further suggest that the IDW method, used to generate the groundwater level at each grid point, did not capture the spatial variability. The use of the additional objective helps to address that issue.

When water table variation is included as an objective function for the shallow well network, illustrated in Figure 10b, the areas where additional wells are placed by DEMO are shifted to the more urban and lower elevation areas in the study area. Inclusion of water table variation as an objective in DEMO helps to account for additional information not picked up by the water table time series data. This information includes surface and vadose zone influence, as well as agriculture/irrigation usage of groundwater. It also helps DEMO to account for the seasonal variation in the lower elevation monitoring wells and highlight their importance which is overshadowed by the naturally higher variation in the higher elevation monitoring well time series data. These results also show that there is a noticeable increase in network variation and spatial coverage of the network when compared to the standard DEMO results of Figure 9b, as the number of Pareto optimal network solutions increased from 68 to 2021. This is similar to the results produced from using the entire well network.

When water table variation is used as an additional objective function to help DEMO to augment the existing deep monitoring well network (results shown in Figure 10c), there is very little change in the locations of additional wells or probability hot spots from those of the standard DEMO results shown in Figure 9c. This is the case despite the increase of Pareto optimal network solutions from 41 to 201. This is likely due to the fact that deeper groundwater is far less seasonally dependant than the shallow groundwater and that deeper groundwater is impacted less by agricultural, municipal, and industrial uses in the study area. Based on these results, including the water table variation objective for the design of deep groundwater monitoring wells is not useful for increasing the spatial coverage of the monitoring network as using the standard DEMO entropy objective functions produces a similar network distribution.

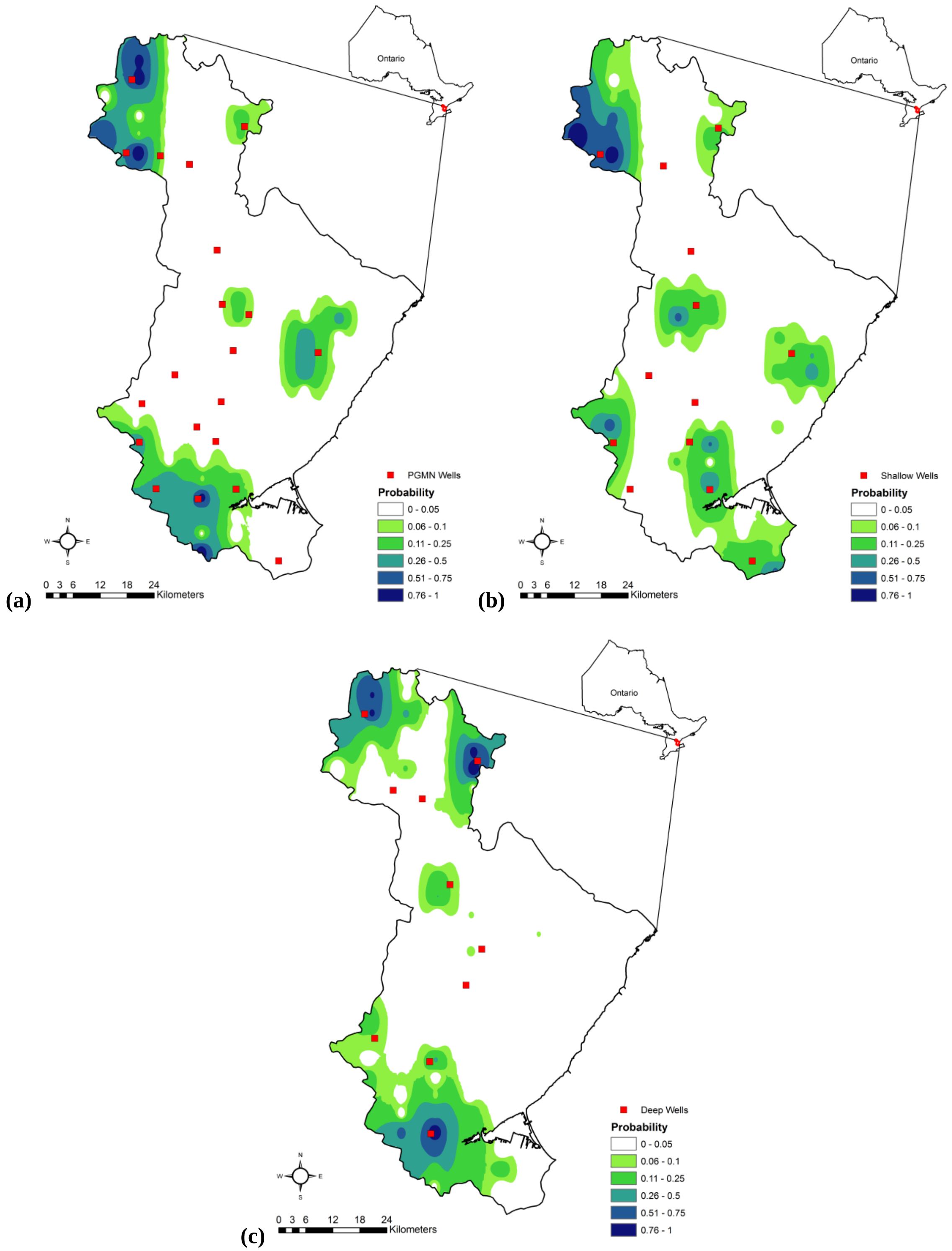


Figure 10: Scenario 2 results for (a) all PGMN wells, (b) PGMN wells classified as shallow, and (c) PGMN wells classified as deep

5.3. DEMO Results with Recharge

For Scenario 3, recharge was used as an additional objective in DEMO. The recharge used for this scenario was calculated using the PRMS results in Figure 8a. The probability plot in Figure 11a continues to include the areas where additional wells are placed in Scenario 1, Figure 9a, while also including the more urban and lower elevation areas of the study area. These additional locations correspond to where the recharge was found to be both the higher and lower recharge areas, but not the mid-range. Inclusion of recharge as an objective in DEMO helps to account for additional information not picked up by the water table time series data and increases the number of Pareto optimal networks from 3 to 350; this information includes surface and vadose zone influence.

The probability plot produced using the DEMO results for augmenting the shallow well network in Scenario 3, using the groundwater recharge from Figure 8a as an additional objective function, is illustrated in Figure 11b. It includes the high probability areas from the entropy only results of Figure 9b as well as expanding the coverage to include much of Hamilton, similar to the DEMO results including all the PGMN wells. Inclusion of recharge as an objective in DEMO helps to account for additional information not picked up by the water table time series data and increases the number of Pareto optimal networks from 68 to 3019 and increase coverage in the Hamilton region.

The probability plot produced using the DEMO results for augmenting the deep well network in Scenario 3 is illustrated in Figure 11c. Using the groundwater recharge from

Figure 8a as an additional objective function in DEMO to produce Pareto optimal networks does not produce networks that are different enough from the entropy only results of Figure 9c, despite the increase in Pareto optimal networks from 41 to 1077. Based on these findings, including the groundwater recharge calculated using the PRMS model (Figure 8a) as an additional objective function for the design of deep groundwater monitoring well networks does not improve the spatial coverage of the network. The standard DEMO results are just as adequate in producing an optimal deep groundwater monitoring well network as those with the additional objective, most likely due to the fact that surface recharge has little influence on the deep groundwater.

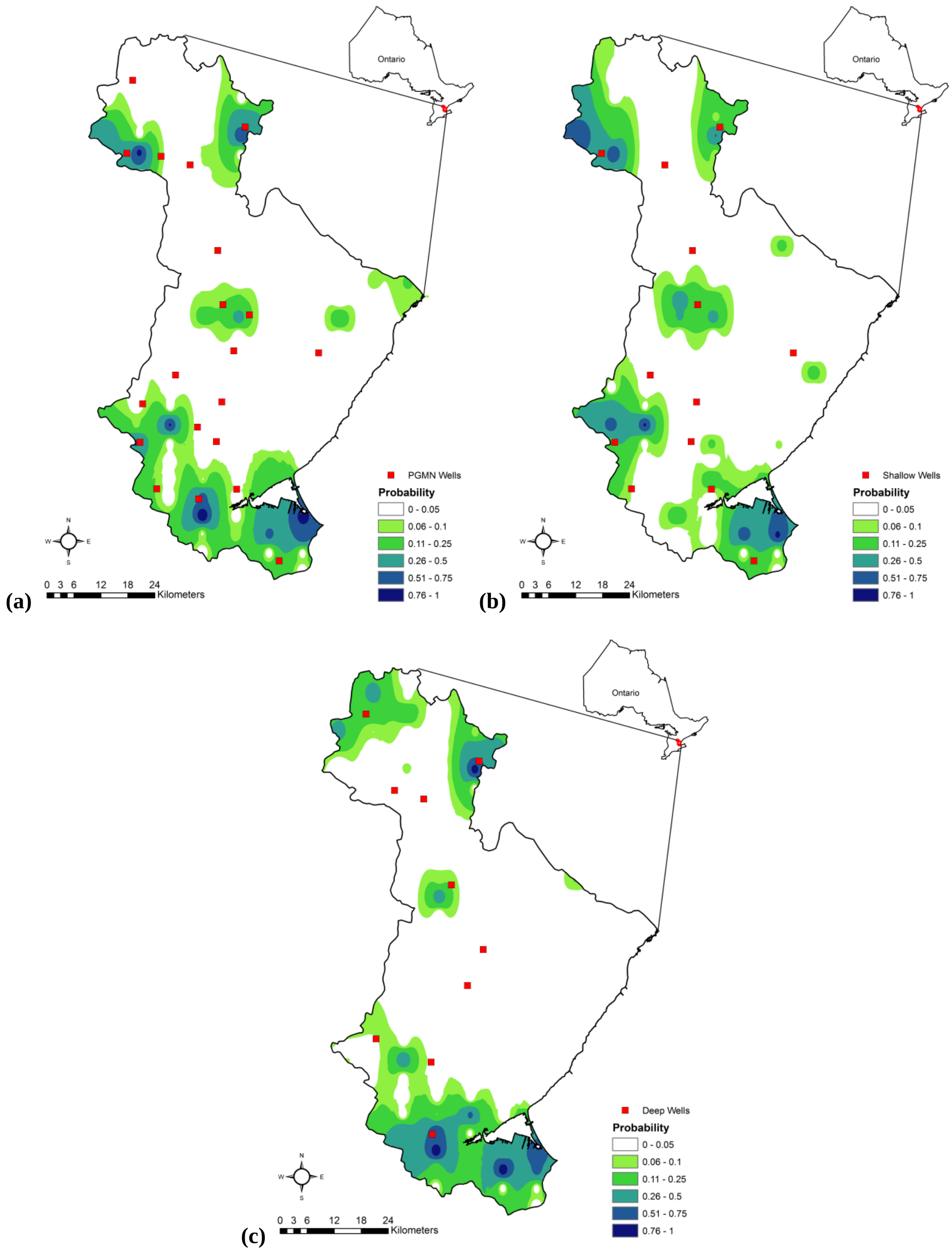


Figure 11: Scenario 3 results for (a) all PGMN wells, (b) PGMN wells classified as shallow, and (c) PGMN wells classified as deep

For Scenario 4 recharge was used as an additional objective in DEMO, the recharge used for this scenario was calculated using the PRMS and HSP-F results in Figure 8b. The probability plot in Figure 12a continues to include the areas where additional wells are placed in Scenario 1, Figure 9a. In addition to the locations found using the entropy only results of DEMO, additional wells were also placed in the more urban and lower elevation areas of the Credit Valley region of the study area, which corresponds to the locations with the lowest recharge values. Inclusion of this recharge increased the number of Pareto optimal networks from 3 to 520.

The probability plot produced using the DEMO results for augmenting the shallow well network in Scenario 4, using the groundwater recharge from Figure 8b as an additional objective function, is illustrated in Figure 12b. It includes most of the high probability areas from the entropy only results of Figure 9b, with less emphasis in the southern areas of the study region, as well as including the more urban and lower elevation areas of the Credit Valley region of the study area, which corresponds to the locations with the lowest recharge values. Inclusion of this recharge increased the number of Pareto optimal networks from 68 to 4086.

The probability plot produced using the DEMO results for augmenting the deep well network in Scenario 4 is illustrated in Figure 12c. It includes most of the high probability areas from the entropy only results of Figure 9c, with less emphasis in the southern areas of the study region, as well as including the more urban and lower elevation areas of the

Credit Valley region of the study area, which corresponds to the locations with the lowest recharge values. Inclusion of this recharge increased the number of Pareto optimal networks from 41 to 2688. From these results it seems that the lower recharge areas, based on the groundwater recharge provided by HHCA and CVC in the HHCV region have a high influence on the network design and monitoring well placements for augmenting the existing PGMN monitoring wells, as well as the well networks classified as shallow and deep.

When comparing the Scenario 4 results to those of Scenario 3, it can be seen that the spatial distribution of optimum networks is fairly different, and the number of possible optimal networks is far higher for those of Scenario 4. The areas in which the existing monitoring wells are less distributed are more highlighted by those results of Scenario 4. This indicates that the optimal networks can be highly influenced by the values used, even if they are values representing the same hydrologic variable. From this it is clear that the quality of the additional objectives used needs to be good as they can largely impact the final designs of the optimal monitoring networks.

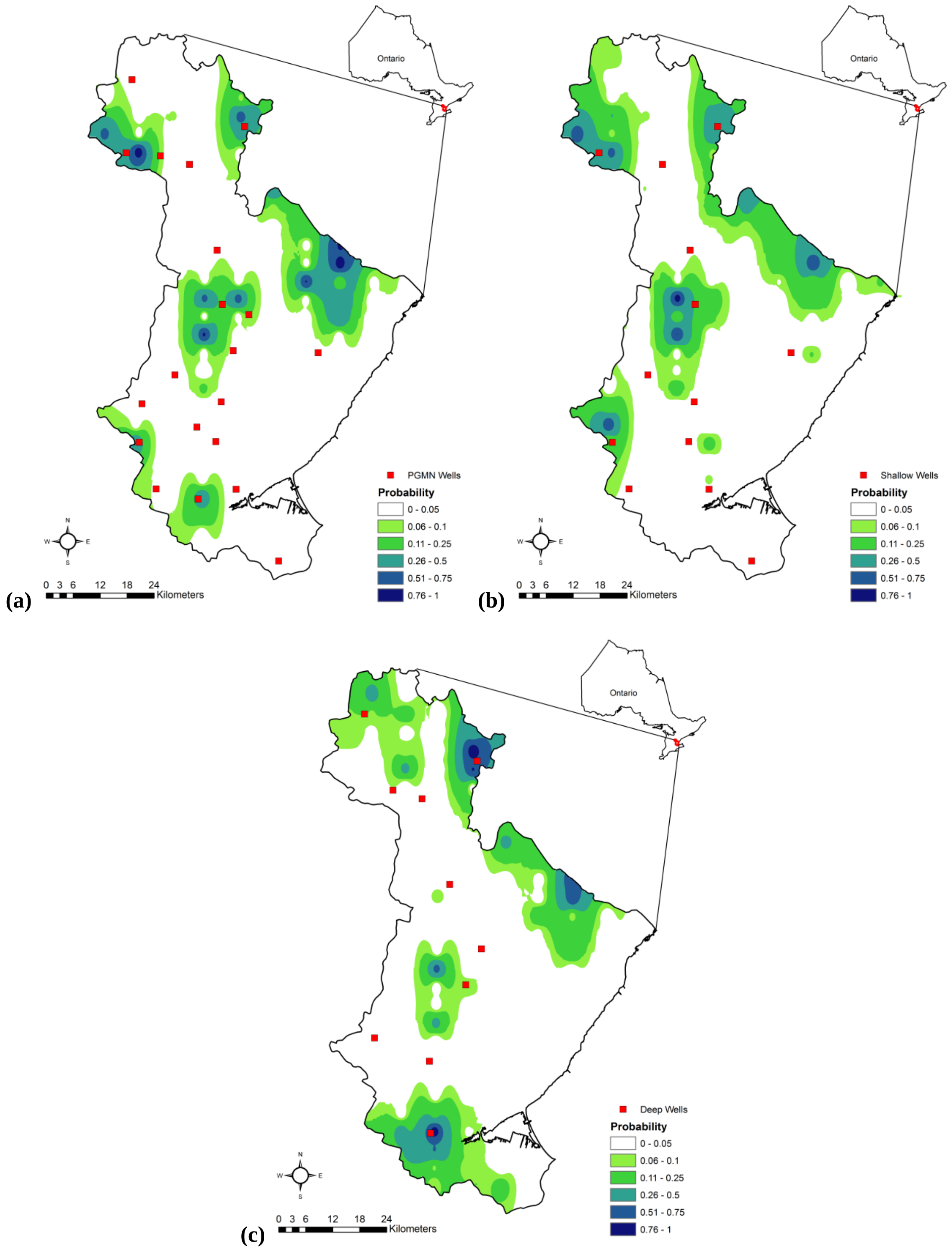


Figure 12: Scenario 4 results for (a) all PGMN wells, (b) PGMN wells classified as shallow, and (c) PGMN wells classified as deep

6. Discussion

When comparing the results, with the land use and land cover map of Figure 1, it becomes evident that the urban development along the coast of Lake Ontario has a low density of PGMN wells. More specifically you see that the lower Credit Valley Region, Halton Region, and Hamilton all lack monitoring from the PGMN. The lack of existing data in these areas reinforces the need for augmenting the existing network, however we are forced to use estimated data not only for the water level time series at the potential well locations, but for the objective functions. In this we see a drawback to the data generation method as it does require existing information to provide time series data for the potential well locations. It may be ideal to include an additional metric which can help the model account for this, or at least when using a data generation method which relies solely on the existing time series information. Despite this drawback, this study does fulfill its goal of providing a groundwater network design tool using DEMO and identifies objective functions which can improve the network spatial coverage.

Scenario 1 results indicate that despite the lack of monitoring in these areas there are few network designs which augment the monitoring network in such a way that well density improves significantly in these areas. As mentioned in the previous section, this is likely a drawback of choosing IDW as a data generation method. Although this can be considered a drawback, it does have the benefit of helping to highlight the spatial coverage gains when including the chosen additional objective functions. Each of these urban areas in the study area has their spatial coverages improved in some way by the chosen objective functions. The water table variation function of Scenario 2 improves coverage mostly in the urban areas of Halton, the groundwater recharge of Scenario 3 improves coverage in Hamilton, and the groundwater recharge of Scenario 4 improves the coverage in lower Credit Valley. This may suggest that including all three additional objective functions along with the standard entropy functions in DEMO would provide optimal network designs which can capture all these areas.

Despite the drawbacks mentioned caused by the data generation method, these results demonstrate that the use of joint entropy and total correlation provide robust measures of the information contained within the study area as suggested by Mishra & Coulibaly (2010). However, the similarities between the four scenarios also show that the network was influenced by the data generation method and for DEMO to capture important characteristics of study area or to increase the network spatial coverage, inclusion of the additional objectives is necessary. The standard DEMO results represented some aspects of the water table variation objective as well as both of the recharge results. However, including either as an objective function explicitly in DEMO had an improvement on the spatial coverage of the network, although not as prominent in the cases which looked at augmenting the deep monitoring well network.

An important consideration in the DEMO approach to network design is that the output of the algorithm contains many Pareto optimal network designs. By considering the probability of station selection, those gauges which are selected most often for Pareto optimal solutions should be regarded as important stations. Considering these wells as important, provides guidance for decision makers on which individual network to build. Furthermore, while it was demonstrated that including additional objectives increased the spatial variability of the network, there was a cost in terms of the number of solutions on the Pareto Front resulting from the additional objectives. This can be viewed as providing more flexibility in an optimal network design, but also gives decision makers many more potential network designs.

7. Conclusions

This study explored the benefits of using water table variation and groundwater recharge as additional objective functions in DEMO for designing optimal groundwater monitoring networks. In particular, this study aimed at determining the added benefits of including or excluding groundwater recharge as an informative hydrological/physical variable, and how sensitive the model is to recharge calculated using different methods on the designed optimum networks. The DEMO model which has been shown to be sufficiently robust for designing optimal minimum hydrometric networks (Samuel et al., 2013) was used to determine the optimum locations for the new additional stations. The particular advantage of DEMO is the combination of the joint entropy and total correlation objectives which optimize the network based on information content. The flexibility of DEMO also allows for the inclusion of water table variation and groundwater recharge as additional objectives.

Including water table variation as an additional design objective in DEMO increased the number of Pareto optimal networks produced by the DEMO model, particularly for the shallow monitoring well network. The inclusion of this objective for the shallow groundwater monitoring well network design allowed for the model to capture more information when finding the optimal networks than it would have using the time series data alone. Using the seasonal water table variation objective will help the network designer produce a more robust network for shallow groundwater level monitoring. For a deep groundwater monitoring network design, the joint entropy and the total correlation commonly used in DEMO appear sufficient. However, appropriate interpolation technique is needed to capture the spatial variability in the time series.

Comparison of networks designed with either PRMS (Scenario 3) or PRMS and HSP-F (Scenario 4) demonstrated that the method used to calculate groundwater recharge has an impact on the distribution of the optimal network. Inclusion of the recharge used for Scenario 3 increased network distribution in the lower portion of the study area around lower Hamilton, whereas inclusion of the recharge used in Scenario 4 increased distribution in the Credit Valley region. The results of Scenario 1 which include only the entropy functions, was able to capture a portion of the network distributions shown in Scenario 3 and 4 results, however these locations are influenced by existing wells. This study shows that inclusion of groundwater recharge as an informative hydrologic/physical variable in designing the hydrometric/groundwater networks increases the number of optimal network solutions and provides additional information for potential locations. These results suggest that it is worth including recharge as a design objective to improve the spatial coverage of the monitoring network.

8. Contributions

The research presented in this thesis expanded on the scope of the DEMO method from Samuel et al. (2013) to include groundwater monitoring networks. In doing this it provided a novel method for groundwater quantity monitoring networks. This method can be used by network designers to design an optimal groundwater quantity monitoring network that provides a high amount of information while reducing the redundant information.

As part of the requirements for the groundwater network design under this project, additional objectives which can be used to aid in network design were to be identified. In this research two objectives were identified which can be used to increase the spatial coverage of the monitoring network, they are the water table variation and the groundwater recharge. Including these objective functions in DEMO allows for additional spatial coverage of the monitoring network when compared to using the standard entropy objectives, joint entropy and total correlation, from DEMO alone.

Additionally this research identified that using these objective functions has far more influences on the shallow groundwater monitoring wells over the deep monitoring wells, which can likely be attributed to them being more influenced by surface activities. This research was also able to show the potential downside to using a data generation method such as IDW in a study area which has a high concentration of monitoring wells in one area and a low concentration in another.

9. Recommendations

- Run DEMO allowing for all wells, both existing PGMN wells and potential wells, to be variable instead of having the existing wells be always chosen in the optimal network.
- Use a groundwater flow model such as GSFLOW (integrated PRMS and MODFLOW) or FEFLOW (Finite Element subsurface FLOW system) to generate input data for DEMO instead of using interpolated values.
- Implement in different study basin to determine if results are similar when using additional objectives in DEMO.
- Explore applicability of other additional objective functions such weighting placement of additional wells on their proximity to water supply wells or other practices that can directly impact the water table as a potential vulnerability metric, or weighting the placement of additional wells based on the cost of operation and maintenance as a potential cost metric.

10. References

Abrishamchi, A., Owlia, R. R., Tajrishy, M., & Abrishamchi, A. (2008). Optimal Design of Groundwater Quality Monitoring Using Entropy Theory. In *International Conference on water scarcity, global changes, and groundwater management responses*.

Alfonso, L., He, L., Lobbrecht, A., & Price, R. (2012). Information theory applied to evaluate the discharge monitoring network of the Magdalena River. *Journal of Hydroinformatics*, *15*(1), 211–228. http://doi.org/10.2166/hydro.2012.066

Alfonso, L., Lobbrecht, A., & Price, R. (2010). Optimization of water level monitoring network in polder systems using information theory. *Water Resources Research*, *46*(1), 1–13. http://doi.org/10.1029/2009WR008953

Aller, L., Bennett, T., Lehr, J. H., Petty, R. J., & Hackett, G. (1987). *DRASTIC : A Standardized Method for Evaluating Ground Water Pollution Potential Using Hydrogeologic Settings*. *Epa-600/2-87-035*.

Alzraiee, A. H., Bau, D. a., & Garcia, L. a. (2013). Multiobjective design of aquifer monitoring networks for optimal spatial prediction and geostatistical parameter estimation. *Water Resources Research*, *49*(6), 3670–3684. http://doi.org/10.1002/wrcr.20300

Ammar, K., Khalil, A., McKee, M., & Kaluarachchi, J. (2008). Bayesian deduction for redundancy detection in groundwater quality monitoring networks. *Water Resources Research*, *44*(8), 1–15. http://doi.org/10.1029/2006WR005616

Andricevic, R. (1990). Cost-effective network design for groundwater flow monitoring. *Stochastic Hydrology and Hydraulics*, *4*(1), 27–41. http://doi.org/10.1007/BF01547730

AquaResources Inc. (2009). *Integrated water budget report – Tier 2, Credit Valley source protection area*.

Asefa, T., Kemblowski, M., Urroz, G., & McKee, M. (2005). Support vector machines (SVMs) for monitoring network design. *Ground Water*, *43*(3), 413–422.

Baalousha, H. (2010). Assessment of a groundwater quality monitoring network using vulnerability mapping and geostatistics: A case study from Heretaunga Plains, New Zealand. *Agricultural Water Management*, *97*(2), 240–246. http://doi.org/10.1016/j.agwat.2009.09.013

Babbar-Sebens, M., & Minsker, B. S. (2008). Standard Interactive Genetic Algorithm—Comprehensive Optimization Framework for Groundwater Monitoring Design. *Journal of Water Resources Planning and Management*, *134*(6), 538–547. http://doi.org/10.1061/(ASCE)0733-9496(2008)134:6(538)

Babbar-Sebens, M., & Minsker, B. S. (2010). A Case-Based Micro Interactive Genetic Algorithm (CBMIGA) for interactive learning and search: Methodology and application to groundwater monitoring design. *Environmental Modelling and Software*, *25*(10), 1176–1187. http://doi.org/10.1016/j.envsoft.2010.03.027

Babbar-Sebens, M., & Minsker, B. S. (2012). Interactive Genetic Algorithm with Mixed Initiative Interaction for multi-criteria ground water monitoring design. *Applied Soft Computing Journal*, *12*(1), 182–195. http://doi.org/10.1016/j.asoc.2011.08.054

Bashi-Azghadi, S. N., & Kerachian, R. (2010). Locating monitoring wells in groundwater systems using embedded optimization and simulation models. *Science of the Total Environment*, *408*(10), 2189–2198. http://doi.org/10.1016/j.scitotenv.2010.02.004

Ben-Jemaa, F., Mariño, M. A., & Loaiciga, H. A. (1994). Multivariate Geostatistical Design of Ground-Water Monitoring Networks. *Journal of Water Resources Planning and Management*, *120*(4), 505–522. http://doi.org/10.1061/(ASCE)0733-9496(1994)120:4(505)

Bicknell, B. R., Imhoff, J. C., Kittle Jr., J. L., Donigan Jr., A. S., & Johanson, R. C. (1996). *Hydrological Simulation Program: Fortran User's Manual for Release 11*. *EPA/600/R-97/080, 755 p.* Athens, Ga.

Carrera, J., Usunoff, E., & Szidarovszky, F. (1984). A method for optimal observation network design for groundwater management. *Journal of Hydrology*, *73*(1-2), 147–163. http://doi.org/10.1016/0022-1694(84)90037-4

Caselton, W. F. (1980). Hydrologic Networks: Information Transmission. *Journal of the Water Resources Planning and Management Division*, *106*(2), 503–520.

Cherkauer, D. S. (2004). Quantifying Ground Water Recharge at Multiple Scales Using PRMS and GIS. *Ground Water*, *42*(1), 97–110. http://doi.org/10.1111/j.1745-6584.2004.tb02455.x

Cieniawski, S. E., Eheart, J. W., & Ranjithan, S. (1995). Using Genetic Algorithms to Solve a Multiobjective Groundwater Monitoring Problem. *Water Resources Research*, *31*(2), 399–409. http://doi.org/10.1029/94WR02039

Coulibaly, P., Samuel, J., Pietroniro, A., & Harvey, D. (2013). Evaluation of Canadian National Hydrometric Network density based on WMO 2008 standards. *Canadian*

Water Resources Journal, *38*(2), 159–167. http://doi.org/10.1080/07011784.2013.787181

Dhar, A., & Datta, B. (2007). Multiobjective Design of Dynamic Monitoring Networks for Detection of Groundwater Pollution. *Journal of Water Resources Planning and Management*, *133*(4), 329–338. http://doi.org/10.1061/(ASCE)0733-9496(2007)133:4(329)

Dhar, A., & Datta, B. (2009). Global optimal design of ground water monitoring network using embedded kriging. *Ground Water*, *47*(6), 806–815. http://doi.org/10.1111/j.1745-6584.2009.00591.x

Earthfx Inc. (2010a). *Vulnerability Analysis for the Milton and Campbellville Wellfields, Regional Municipality of Halton, Ontario*. Toronto, ON.

Earthfx Inc. (2010b). *Vulnerability Assessment and Scoring of Wellhead Protection Areas City of Hamilton, Ontario*. Toronto, ON.

Environment Canada. (2013a). Groundwater - Water - Environment Canada. Retrieved from https://www.ec.gc.ca/eau-water/default.asp?lang=En&n=300688DC-1

Environment Canada. (2013b). Hydrometric Program - Water - Environment Canada. Retrieved from http://www.ec.gc.ca/rhc-wsc/default.asp?lang=En&n=EC164002-1

Government of Ontario. (2013a). 1:250 000 scale bedrock geology of Ontario. Retrieved from https://www.ontario.ca/data/1250-000-scale-bedrock-geology-ontario

Government of Ontario. (2013b). Provincial Groundwater Monitoring Network. Retrieved from https://www.ontario.ca/data/provincial-groundwater-monitoring-network

Halton-Hamilton Source Protection. (2010). *Report on Tier 1 Water Budget and Water Quantity Stress Assessment for Halton-Hamilton Source Protection Region and Report on Tier 2 Water Budget and Water Quantity Stress Assessments for the Upper West Branch of Sixteen Mile Creek and Middle Spencer Creek Subwatersheds*.

Harmancioglu, N. B., & Singh, V. P. (1998). *Entropy in environmental and water resources*. (R. W. Herschy & R. W. Fairbridge, Eds.)*Encyclopedia of Hydrology and Water Resources*. Boston: Kluwer Academic Publishers.

Healy, R. W., & Cook, P. G. (2002). Using groundwater levels to estimate recharge. *Hydrogeology Journal*, *10*(1), 91–109. http://doi.org/10.1007/s10040-001-0178-0

Hudak, P. F., & Loaiciga, H. A. (1992). A location modeling approach for groundwater monitoring network augmentation. *Water Resources Research*, *28*(3), 643–649. http://doi.org/10.1029/91WR02851

Hudak, P. F., & Loaiciga, H. A. (1993). An optimization method for monitoring network design in multilayered groundwater flow systems. *Water Resources Research*, *29*(8), 2835–2845. http://doi.org/10.1029/93WR01042

Hudak, P. F., Loaiciga, H. A., & Mariño, M. A. (1995). Regional-scale ground water quality monitoring via integer programming. *Journal of Hydrology*, *164*(1-4), 153–170.

Husain, T. (1979). *Shannon's Information Theory in Hydrologic Network Design and Estimation*. The University of British Columbia.

Husain, T. (1987). Hydrologic Network Design Formulation. *Canadian Water Resources Journal*, *12*(1), 44–63. http://doi.org/10.4296/cwrj1201044

Husain, T. (1989). Hydrologic uncertainty measure and network design. *Journal of the American Water Resources Association*, *25*(3), 527–534. http://doi.org/10.1111/j.1752-1688.1989.tb03088.x

Khader, A. I., & McKee, M. (2014). Use of a relevance vector machine for groundwater quality monitoring network design under uncertainty. *Environmental Modelling & Software*, *57*, 115–126. http://doi.org/10.1016/j.envsoft.2014.02.015

Khan, S., Chen, H. F., & Rana, T. (2008). Optimizing Ground Water Observation Networks in Irrigation Areas Using Principal Component Analysis. *Ground Water Monitoring and Remediation*, *28*(3), 93–100.

Kollat, J. B., & Reed, P. M. (2006). Comparing state-of-the-art evolutionary multi-objective algorithms for long-term groundwater monitoring design. *Advances in Water Resources*, *29*(6), 792–807. http://doi.org/10.1016/j.advwatres.2005.07.010

Kollat, J. B., & Reed, P. M. (2007). A computational scaling analysis of multiobjective evolutionary algorithms in long-term groundwater monitoring applications. *Advances in Water Resources*, *30*(3), 408–419. http://doi.org/10.1016/j.advwatres.2006.05.009

Kollat, J. B., Reed, P. M., & Kasprzyk, J. R. (2008). A new epsilon-dominance hierarchical Bayesian optimization algorithm for large multiobjective monitoring network design problems. *Advances in Water Resources*, *31*(5), 828–845. http://doi.org/10.1016/j.advwatres.2008.01.017

Kollat, J. B., Reed, P. M., & Maxwell, R. M. (2011). Many-objective groundwater monitoring network design using bias-aware ensemble Kalman filtering, evolutionary optimization, and visual analytics. *Water Resources Research*, *47*(2), 1–18. http://doi.org/10.1029/2010WR009194

Kornelsen, K. C., & Coulibaly, P. (2013). McMaster Mesonet soil moisture dataset: Description and spatio-temporal variability analysis. *Hydrology and Earth System Sciences*, *17*(4), 1589–1606. http://doi.org/10.5194/hess-17-1589-2013

Krstanovic, P. F., & Singh, V. P. (1992a). Evaluation of rainfall networks using entropy: I. Theoretical development. *Water Resources Management*, *6*(4), 279–293. http://doi.org/10.1007/BF00872281

Krstanovic, P. F., & Singh, V. P. (1992b). Evaluation of rainfall networks using entropy: II. Application. *Water Resources Management*, *6*(4), 295–314. http://doi.org/10.1007/BF00872282

Lavoie, R., Joerin, F., Vansnick, J.-C., & Rodriguez, M. J. (2015). Integrating groundwater into land planning: A risk assessment methodology. *Journal of Environmental Management*, *154,* 358–371. http://doi.org/10.1016/j.jenvman.2015.02.020

Leavesley, G. H., Lichty, R. W., Troutman, B. M., & Saindon, L. G. (1983). *Precipitation-Runoff Modeling System: User´s Manual*. *U.S. Geological Survey, Water-Resources Investigations Report 83-4238*.

Li, Y., & Hilton, A. B. C. (2005). Reducing Spatial Sampling in Long-Term Groundwater Monitoring Networks Using Ant Colony Optimization. *International Journal of Computational Intelligence Research*, *1*(1), 19–28.

Li, Y., & Hilton, A. B. C. (2007). Optimal groundwater monitoring design using an ant colony optimization paradigm. *Environmental Modelling & Software*, *22*(1), 110–116. http://doi.org/10.1016/j.envsoft.2006.05.023

Ling, M., Rifai, H. S., & Newell, C. J. (2005). Optimizing groundwater long-term monitoring networks using Delaunay triangulation spatial analysis techniques. *Environmetrics*, *16*(6), 635–657. http://doi.org/10.1002/env.727

Loaiciga, H. A. (1989). An optimization approach for groundwater quality monitoring network design. *Water Resources Research*, *25*(8), 1771–1282. http://doi.org/10.1029/WR025i008p01771

Loaiciga, H. A., Charbeneau, R. J., Everett, L. G., Fogg, G. E., Hobbs, B. F., & Rouhani, S. (1992). Review of Ground Water Quality Monitoring Network Design. *Journal of Hydraulic Engineering*. http://doi.org/10.1061/(ASCE)0733-9429(1992)118:1(11)

Mahar, P. S., & Datta, B. (1997). Optimal Monitoring Network and Ground-Water–Pollution Source Identification. *Journal of Water Resources Planning and Management*, *123*(4), 199–207. http://doi.org/10.1061/(ASCE)0733-9496(1997)123:4(199)

Meyer, P. D., Valocchi, A. J., & Eheart, J. W. (1994). Monitoring network design to provide initial detection of groundwater contamination. *Water Resources Research*, *30*(9), 2647–2659. http://doi.org/10.1029/94WR00872

Mishra, A. K., & Coulibaly, P. (2009). Developments in hydrometric network design: A review. *Reviews of Geophysics*, *47*(2), 1–24. http://doi.org/10.1029/2007RG000243

Mishra, A. K., & Coulibaly, P. (2010). Hydrometric network evaluation for Canadian watersheds. *Journal of Hydrology*, *380*(3-4), 420–437. http://doi.org/10.1016/j.jhydrol.2009.11.015

Mogheir, Y., de Lima, J. L. M. P., & Singh, V. P. (2009). Entropy and multi-objective based approach for groundwater quality monitoring network Assessment and redesign. *Water Resources Management*, *23*(8), 1603–1620. http://doi.org/10.1007/s11269-008-9343-8

Mogheir, Y., & Singh, V. P. (2002). Application of information theory to groundwater quality monitoring networks. *Water Resources Management*, *16*(1), 37–49. http://doi.org/10.1023/A:1015511811686

Mogheir, Y., Singh, V. P., & de Lima, J. L. M. P. (2006). Spatial assessment and redesign of a groundwater quality monitoring network using entropy theory, Gaza Strip, Palestine. *Hydrogeology Journal*, *14*(5), 700–712. http://doi.org/10.1007/s10040-005-0464-3

Natural Resources Canada. (2009). GeoBase - Land Cover, circa 2000-Vector (LCC2000-V). Retrieved from http://geogratis.gc.ca/api/en/nrcan-rncan/

Natural Resources Canada. (2014). Hydrogeological Regions - Canada. Retrieved from http://gin.gw-info.net/service/api_ngwds:gin2/en/wmc/standard.html

Natural Resources Canada. (2015). Canadian Digital Elevation Data (CDED). Retrieved from http://www.geobase.ca/

Nunes, L. M., Cunha, M. C., & Ribeiro, L. (2004). Groundwater Monitoring Network Optimization with Redundancy Reduction. *Journal of Water Resources Planning and Management*, *130*(1), 33–43. http://doi.org/10.1061/(ASCE)0733-9496(2004)130:1(33)

Nunes, L. M., Paralta, E., Cunha, M. C., & Ribeiro, L. (2004). Groundwater nitrate monitoring network optimization with missing data. *Water Resources Research*, *40*(W02406), 1–18. http://doi.org/10.1029/2003WR002469

Ontario Ministry of Natural Resources. (2012). Soil survey complex. Land Information Ontario, data description.

Prakash, M. R., & Singh, V. S. (2000). Network design for groundwater monitoring - a case study. *Environmental Geology*, *39*(6), 628–632. http://doi.org/10.1007/s002540050474

Reed, P. M., & Kollat, J. B. (2012). Save now, pay later? Multi-period many-objective groundwater monitoring design given systematic model errors and uncertainty. *Advances in Water Resources*, *35*, 55–68. http://doi.org/10.1016/j.advwatres.2011.10.011

Reed, P. M., & Kollat, J. B. (2013). Visual analytics clarify the scalability and effectiveness of massively parallel many-objective optimization: A groundwater monitoring design example. *Advances in Water Resources*, *56*, 1–13. http://doi.org/10.1016/j.advwatres.2013.01.011

Reed, P. M., Kollat, J. B., & Devireddy, V. K. (2007). Using interactive archives in evolutionary multiobjective optimization: A case study for long-term groundwater monitoring design. *Environmental Modelling & Software*, *22*(5), 683–692. http://doi.org/10.1016/j.envsoft.2005.12.021

Reed, P. M., & Minsker, B. S. (2004). Striking the Balance: Long-Term Groundwater Monitoring Design for Conflicting Objectives. *Journal of Water Resources Planning and Management*, *130*(2), 140–149. http://doi.org/10.1061/(ASCE)0733-9496(2004)130:2(140)

Reed, P. M., Minsker, B. S., & Goldberg, D. E. (2001). A multiobjective approach to cost effective long-term groundwater monitoring using an elitist nondominated sorted genetic algorithm with historical data. *Journal of Hydroinformatics*, *3*(2), 71–89.

Reed, P. M., Minsker, B. S., & Valocchi, A. J. (2000). Cost-effective long-term groundwater monitoring design using a genetic algorithm and global mass interpolation. *Water Resources Research*, *36*(12), 3731–3741. http://doi.org/10.1029/2000WR900232

Rianna, M., Ridolfi, E., Montesarchio, V., Alfonso, L., Baldassarre, G. D., Russo, F., & Napolitano, F. (2012). Definition of Homogeneous Regions through Entropy Theory. In *The 3rd STAHY International Workshop on Statistical Methods for Hydrolgy and Water Resources Managment*. Tunis, Tunisia.

Samuel, J., Coulibaly, P., & Kollat, J. B. (2013). CRDEMO: Combined regionalization and dual entropy-multiobjective optimization for hydrometric network design. *Water Resources Research*, *49*(12), 8070–8089. http://doi.org/10.1002/2013WR014058

Scanlon, B. R., Healy, R. W., & Cook, P. G. (2002). Choosing appropriate techniques for quantifying groundwater recharge. *Hydrogeology Journal*, *10*(2), 18–39. http://doi.org/10.1007/s10040-0010176-2

Shannon, C. E. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal*, *27*(3-4), 379–423, 623–656.

Shepard, D. (1968). A two-dimensional interpolation function for irregularly-spaced data. In *23rd ACM national conference* (pp. 517–524). http://doi.org/10.1145/800186.810616

Singh, V. P. (1997). The use of entropy in hydrology and water resources. *Hydrological Processes*, *11*(6), 587–626. http://doi.org/10.1002/(SICI)1099-1085(199705)11:6<587::AID-HYP479>3.0.CO;2-P

Storck, P., Eheart, J. W., & Valocchi, A. J. (1997). A method for the optimal location of monitoring wells for detection of groundwater contamination in three-dimensional heterogenous aquifers. *Water Resources Research*, *33*(9), 2081–2088. http://doi.org/10.1029/97WR01704

Triki, I., Zairi, M., & Ben Dhia, H. (2013). A geostatistical approach for groundwater head monitoring network optimisation: case of the Sfax superficial aquifer (Tunisia). *Water and Environment Journal*, *27*, 362–372. http://doi.org/10.1111/j.1747-6593.2012.00352.x

United States Geological Survey. (2000). Ground-Water Resources for the Future: Land Subsidence in the U.S. Retrieved from http://water.usgs.gov/ogw/pubs/fs00165/

United States Geological Survey. (2013). USGS GWRP: Methods for Estimating Ground-Water Recharge In Humid Regions. Retrieved from http://water.usgs.gov/ogw/gwrp/methods/

Wagner, B. J. (1995). Sampling Design Methods For Groundwater Modeling Under Uncertainty. *Water Resources Research*, *31*(10), 2581–2591. http://doi.org/10.1029/95WR02107

Wu, J., Zheng, C., & Chien, C. C. (2005). Cost-effective sampling network design for contaminant plume monitoring under general hydrogeological conditions. *Journal of Contaminant Hydrology*, *77*(1-2), 41–65. http://doi.org/10.1016/j.jconhyd.2004.11.006

Wu, J., Zheng, C., Chien, C. C., & Zheng, L. (2006). A comparative study of Monte Carlo simple genetic algorithm and noisy genetic algorithm for cost-effective sampling network design under uncertainty. *Advances in Water Resources*, *29*(6), 899–911. http://doi.org/10.1016/j.advwatres.2005.08.005

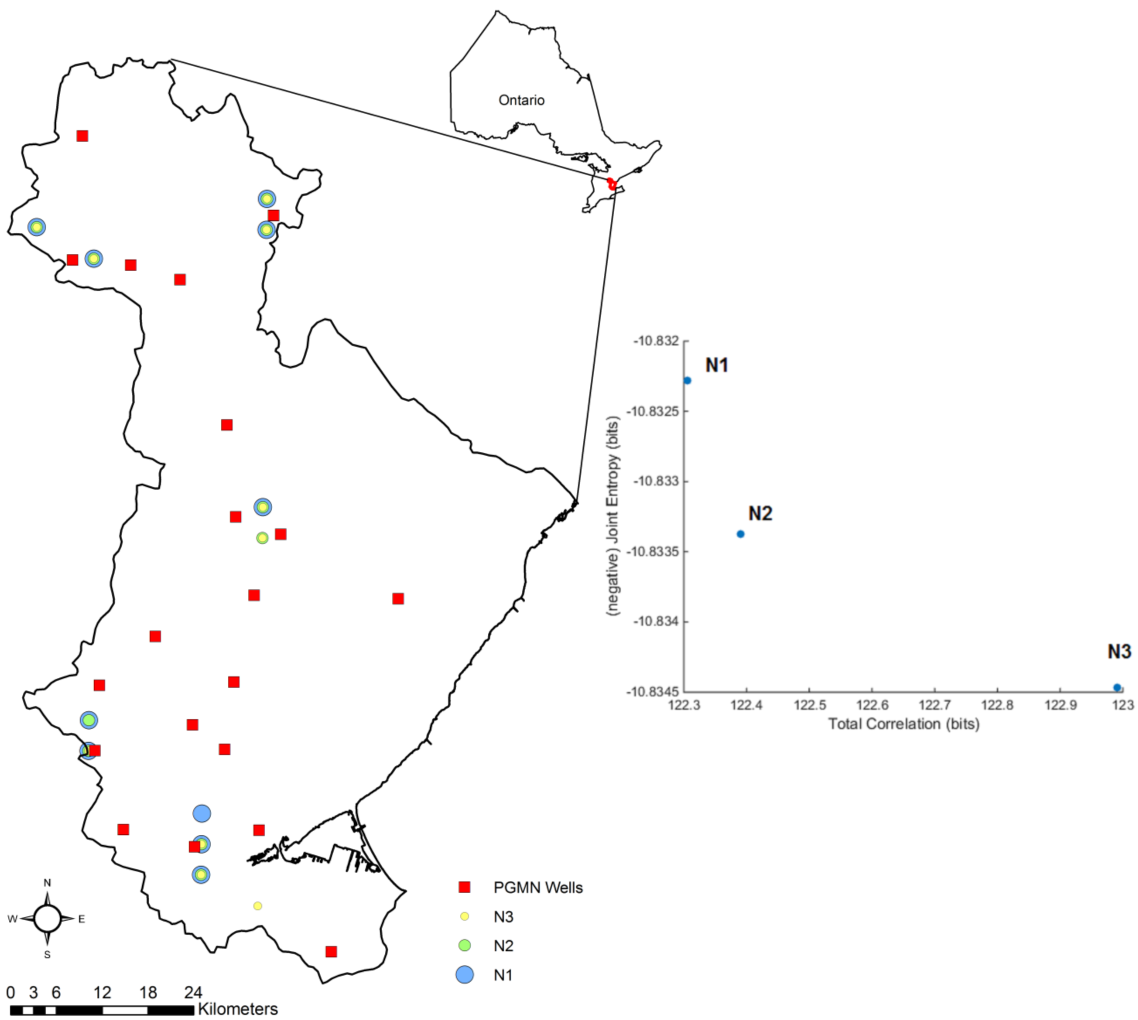
Yang, F., Cao, S., Liu, X., & Yang, K. (2008). Design of groundwater level monitoring network with ordinary kriging. *Journal of Hydrodynamics, Ser. B*, *20*(3), 339–346. http://doi.org/10.1016/S1001-6058(08)60066-9

Zhou, Y., Dong, D., Liu, J., & Li, W. (2013). Upgrading a regional groundwater level monitoring network for Beijing Plain, China. *Geoscience Frontiers*, *4*(1), 127–138. http://doi.org/10.1016/j.gsf.2012.03.008

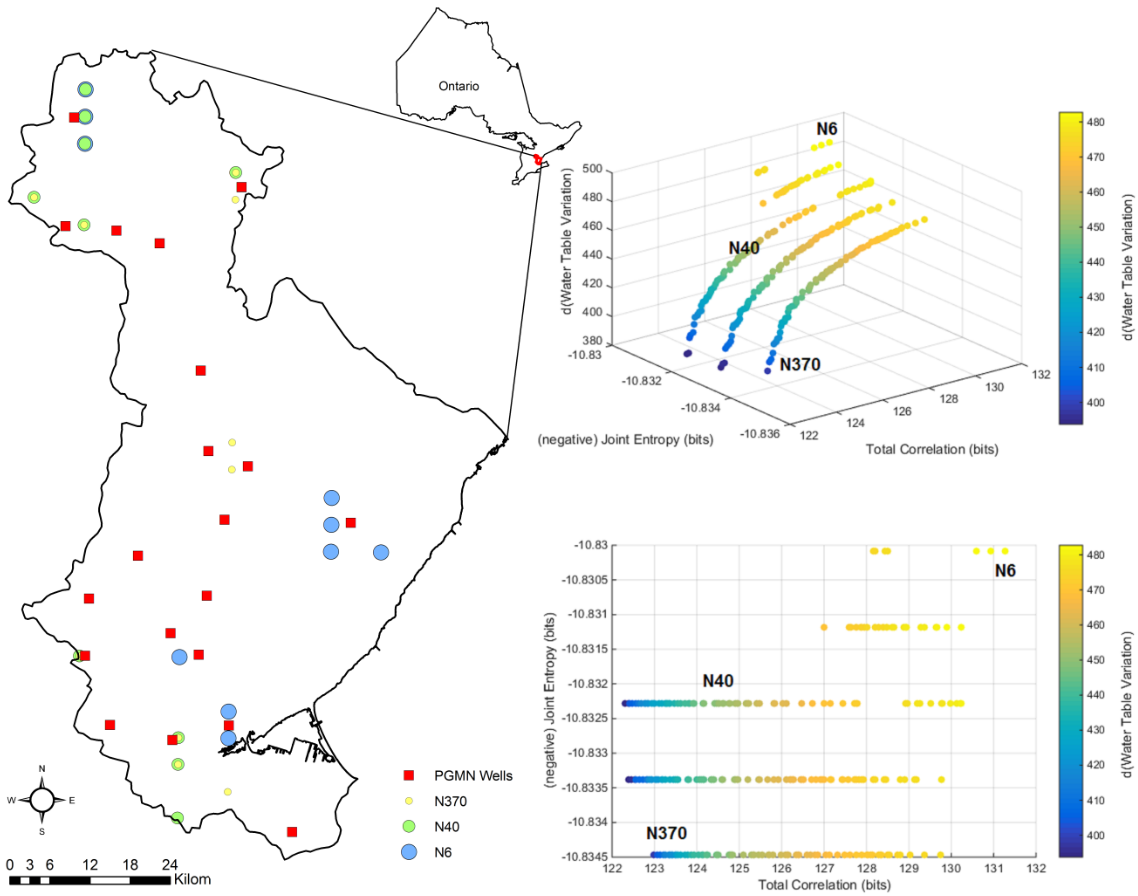
Appendix A: Missing Data Summary Table

	Missing data percentages for each well and time frame								
WELL	**2001-2012**	**2001-2005**	**2002-2006**	**2003-2007**	**2004-2008**	**2005-2009**	**2006-2010**	**2007-2011**	**2008-2012**
W0000001-1	9%	9%	5%	5%	6%	6%	9%	9%	11%
W0000002-1	23%	41%	24%	16%	16%	16%	8%	5%	7%
W0000004-1	14%	9%	5%	5%	5%	5%	6%	21%	24%
W0000005-1	13%	17%	20%	20%	20%	20%	10%	0%	4%
W0000007-2	9%	14%	9%	12%	12%	12%	4%	4%	3%
W0000008-1	15%	29%	24%	19%	20%	20%	5%	0%	3%
W0000019-1	48%	48%	44%	41%	21%	14%	27%	43%	63%
W0000026-1	37%	23%	16%	16%	16%	18%	25%	42%	62%
W0000028-2	37%	20%	13%	13%	13%	20%	30%	50%	66%
W0000028-4	40%	7%	5%	22%	22%	28%	48%	63%	66%
W0000031-1	31%	34%	45%	56%	56%	46%	31%	11%	9%
W0000033-1	22%	38%	31%	31%	31%	30%	10%	5%	11%
W0000124-1	42%	76%	56%	49%	36%	31%	23%	23%	25%
W0000163-2	51%	58%	38%	18%	5%	9%	24%	44%	64%
W0000163-3	44%	35%	15%	7%	7%	11%	31%	51%	64%
W0000164-2	60%	63%	43%	38%	42%	28%	41%	61%	70%
W0000164-3	49%	42%	22%	17%	17%	21%	35%	55%	64%
W0000165-2	47%	39%	23%	11%	11%	17%	33%	49%	66%
W0000165-3	51%	51%	36%	22%	22%	25%	31%	45%	64%
W0000294-1	23%	47%	27%	7%	3%	1%	1%	1%	8%
W0000295-1	26%	44%	24%	13%	9%	12%	12%	12%	10%
W0000296-1	26%	54%	34%	14%	2%	2%	0%	2%	8%
W0000297-1	23%	44%	24%	4%	2%	2%	2%	4%	10%
W0000336-1	20%	46%	27%	7%	1%	1%	1%	0%	3%
W0000337-1	32%	70%	52%	35%	29%	20%	5%	3%	2%
W0000338-1	31%	49%	29%	9%	0%	0%	0%	5%	25%
MIN	9%	7%	5%	4%	0%	0%	0%	0%	2%
AVE	32%	39%	27%	20%	16%	16%	17%	23%	31%
MAX	60%	76%	56%	56%	56%	46%	48%	63%	70%

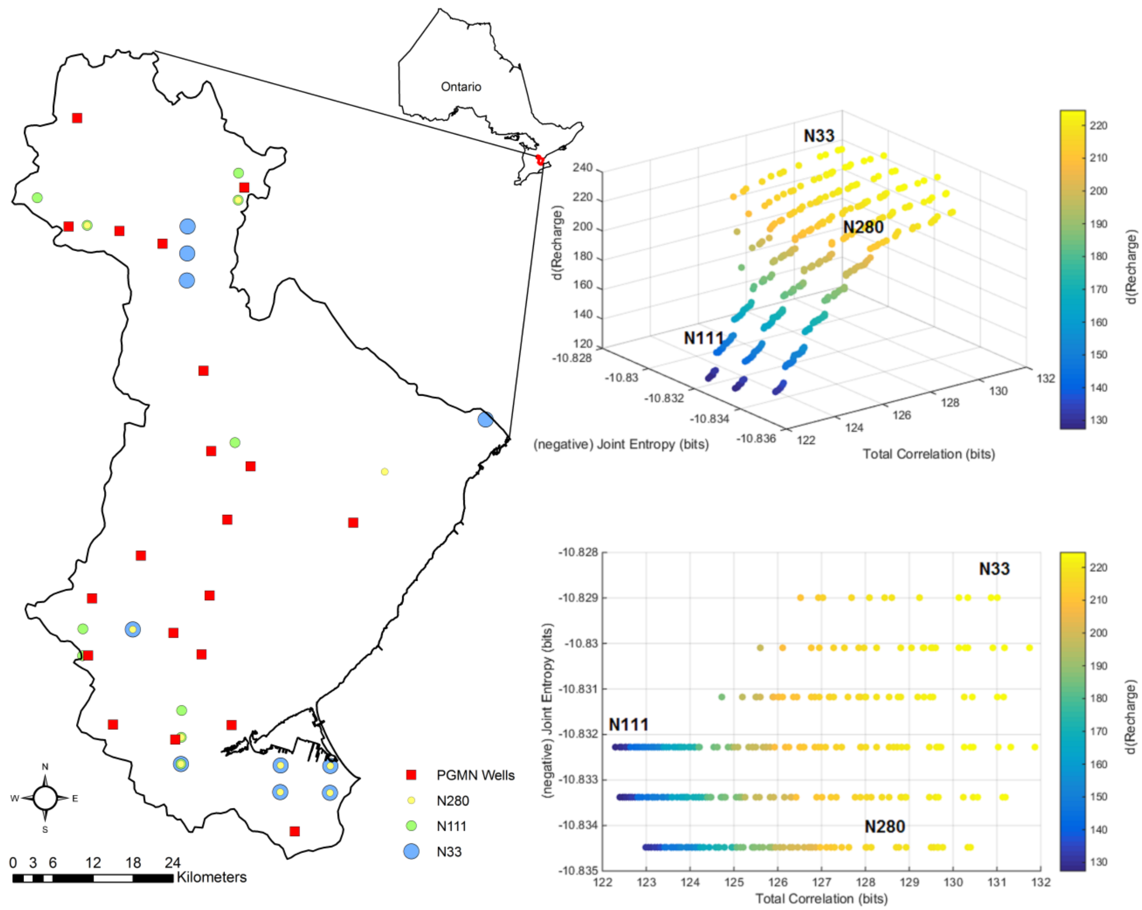
Appendix B: Example network designs and their corresponding Pareto fronts for augmenting the PGMN wells


Ontario
PGMN Wells
N3
N2
N1
N
W
E
S
0 3 6 12 18 24
Kilometers
N1
N2
N3
-10.832
-10.8325
-10.833
-10.8335
-10.834
-10.8345
(negative) Joint Entropy (bits)
122.3
122.4
122.5
122.6
122.7
122.8
122.9
123
Total Correlation (bits)

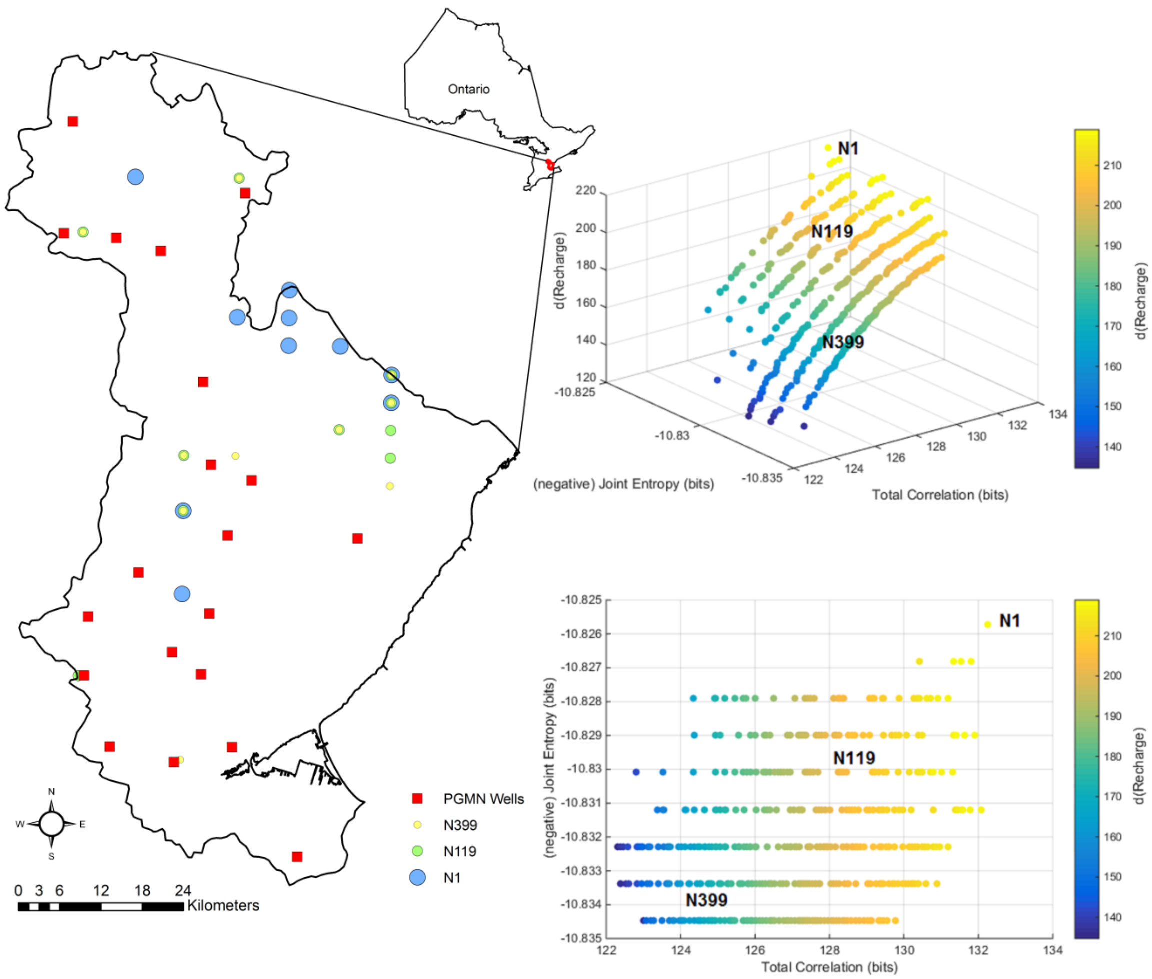
Scenario 1


Ontario
PGMN Wells
N370
N40
N6
0 3 6 12 18 24
Kilom
N6
N40
N370
d(Water Table Variation)
(negative) Joint Entropy (bits)
Total Correlation (bits)
(negative) Joint Entropy (bits)
Total Correlation (bits)
d(Water Table Variation)

Scenario 2

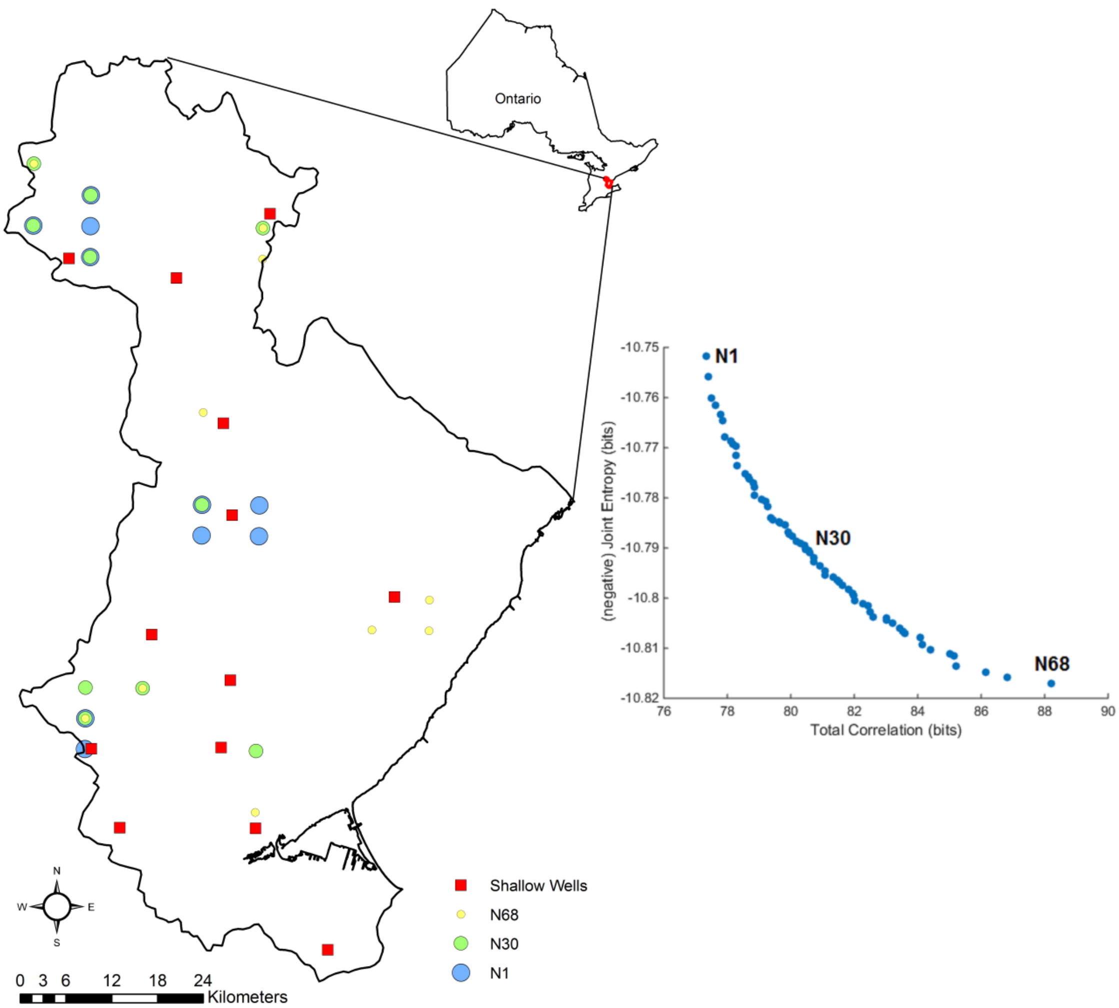

Ontario
PGMN Wells
N280
N111
N33
0 3 6 12 18 24
Kilometers
N33
N280
N111
d(Recharge)
(negative) Joint Entropy (bits)
Total Correlation (bits)
N33
N111
N280
(negative) Joint Entropy (bits)
Total Correlation (bits)
d(Recharge)

Scenario 3

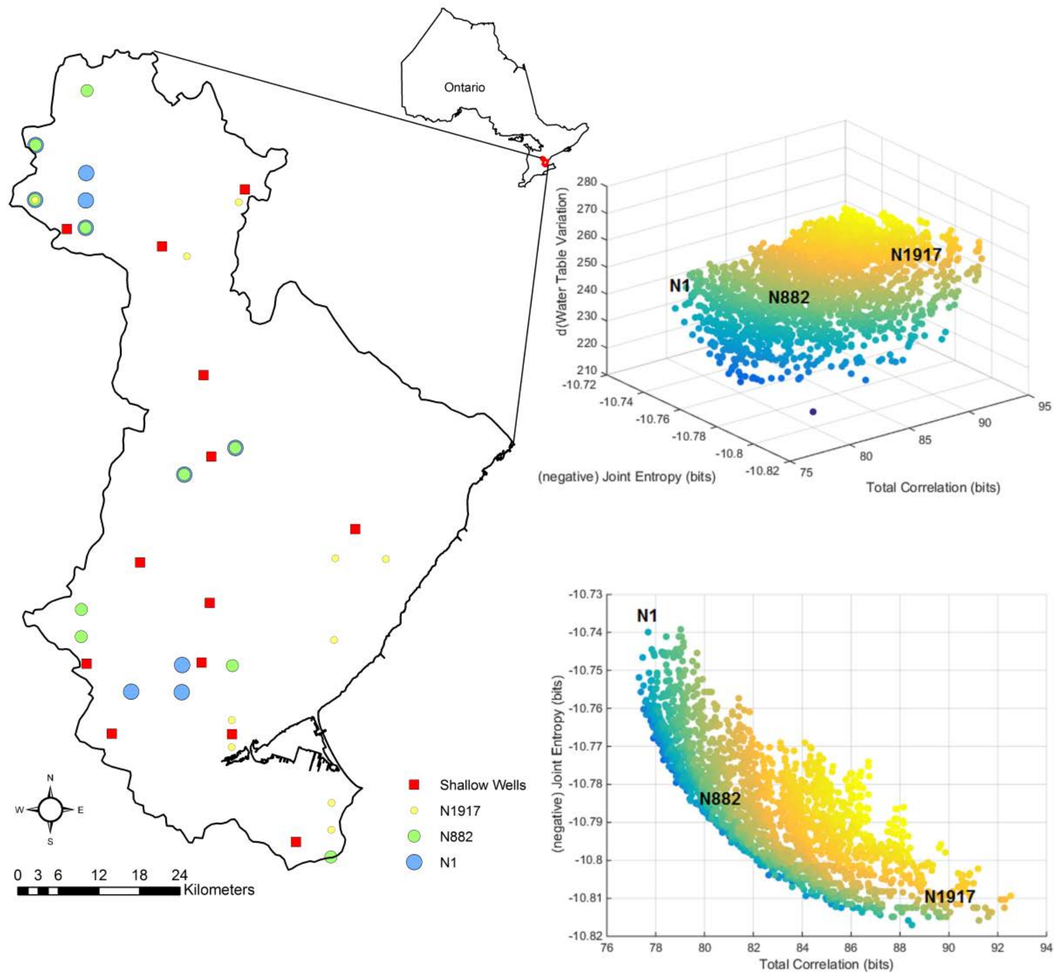

Ontario
PGMN Wells
N399
N119
N1
Kilometers
N1
N119
N399
d(Recharge)
(negative) Joint Entropy (bits)
Total Correlation (bits)

Scenario 4

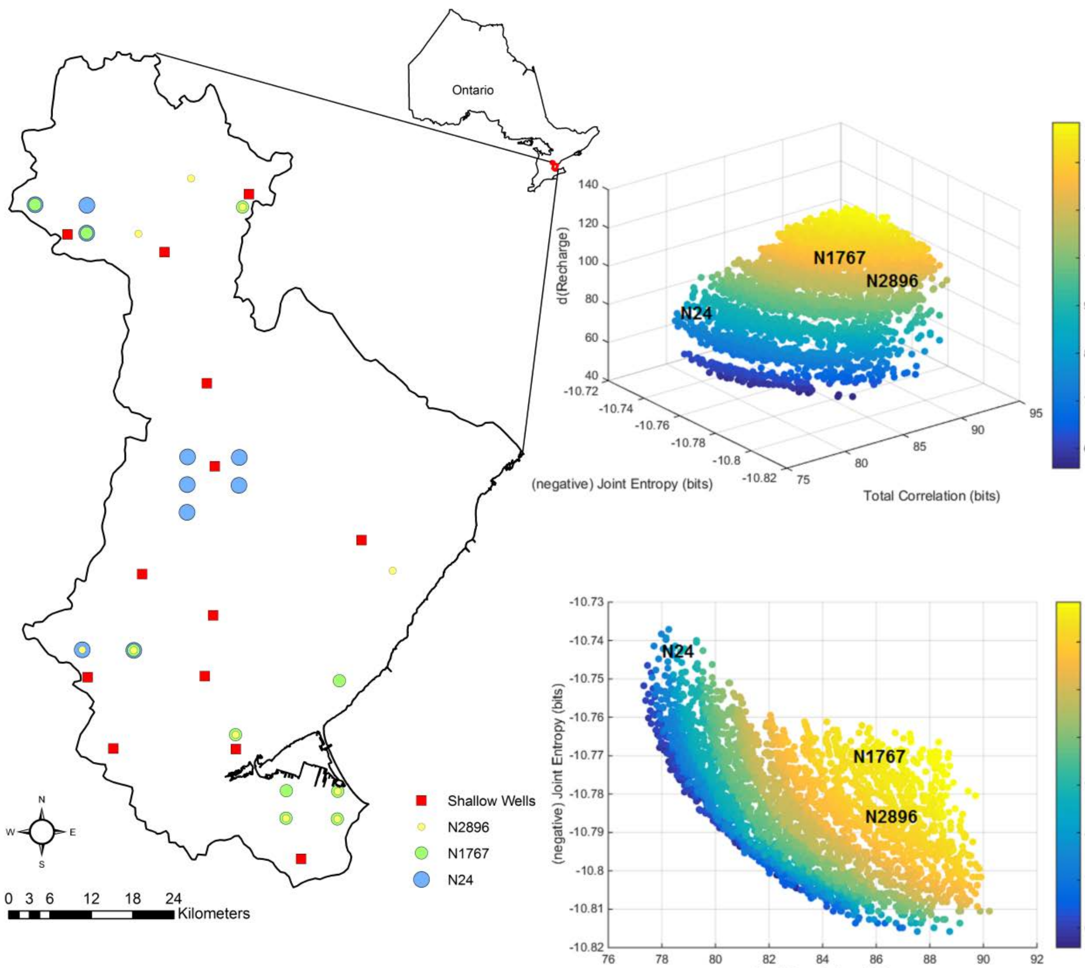
Appendix C: Example network designs and their corresponding Pareto fronts for augmenting the shallow wells


Ontario
Shallow Wells
N68
N30
N1
0 3 6 12 18 24
Kilometers
N
E
S
W
N1
N30
N68
(negative) Joint Entropy (bits)
Total Correlation (bits)
-10.75
-10.76
-10.77
-10.78
-10.79
-10.8
-10.81
-10.82
76
78
80
82
84
86
88
90

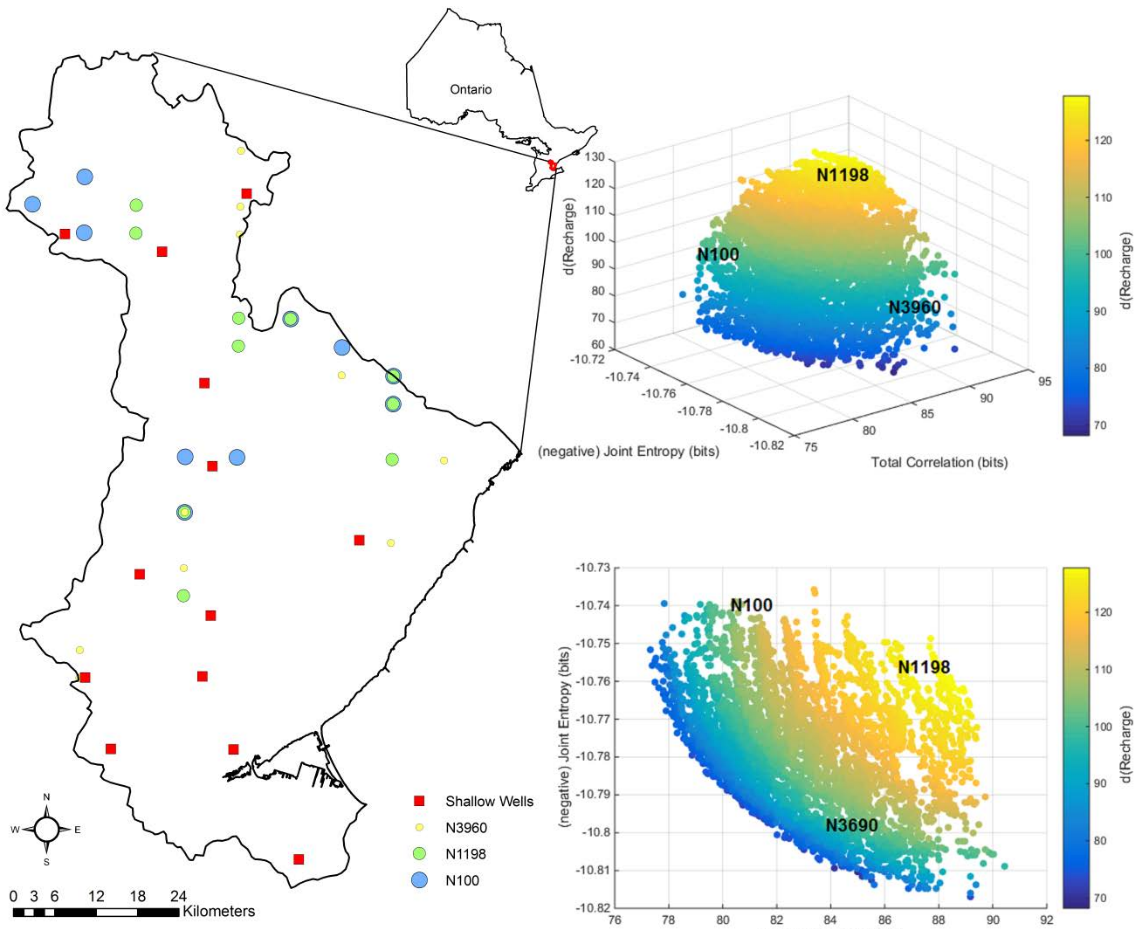
Scenario 1


Ontario
Shallow Wells
N1917
N882
N1
0 3 6 12 18 24 Kilometers
N1
N882
N1917
d(Water Table Variation)
(negative) Joint Entropy (bits)
Total Correlation (bits)
N1
N882
N1917
(negative) Joint Entropy (bits)
Total Correlation (bits)
d(Water Table Variation)

Scenario 2

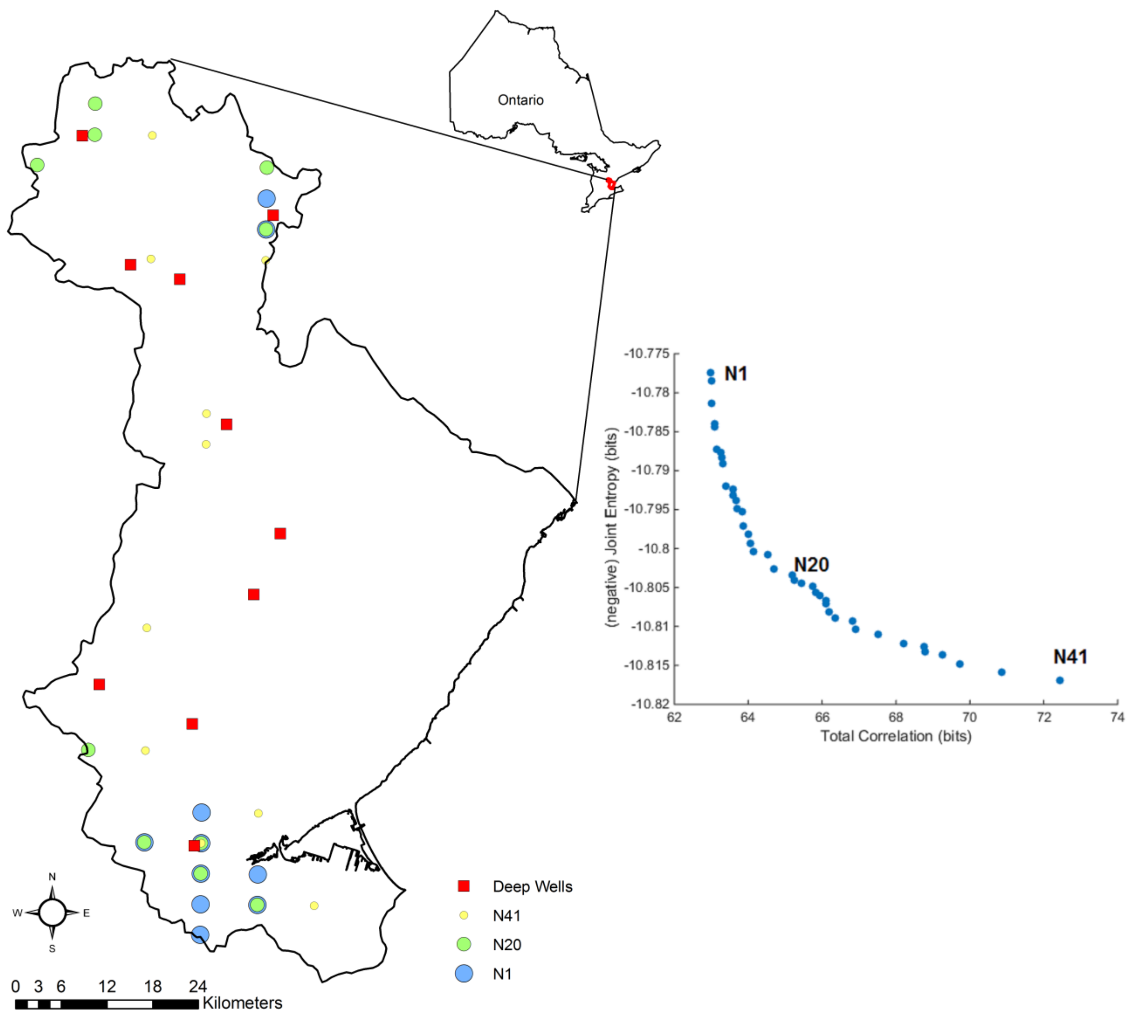

Ontario
Shallow Wells
N2896
N1767
N24
0 3 6 12 18 24 Kilometers
N1767
N2896
N24
d(Recharge)
(negative) Joint Entropy (bits)
Total Correlation (bits)

Scenario 3

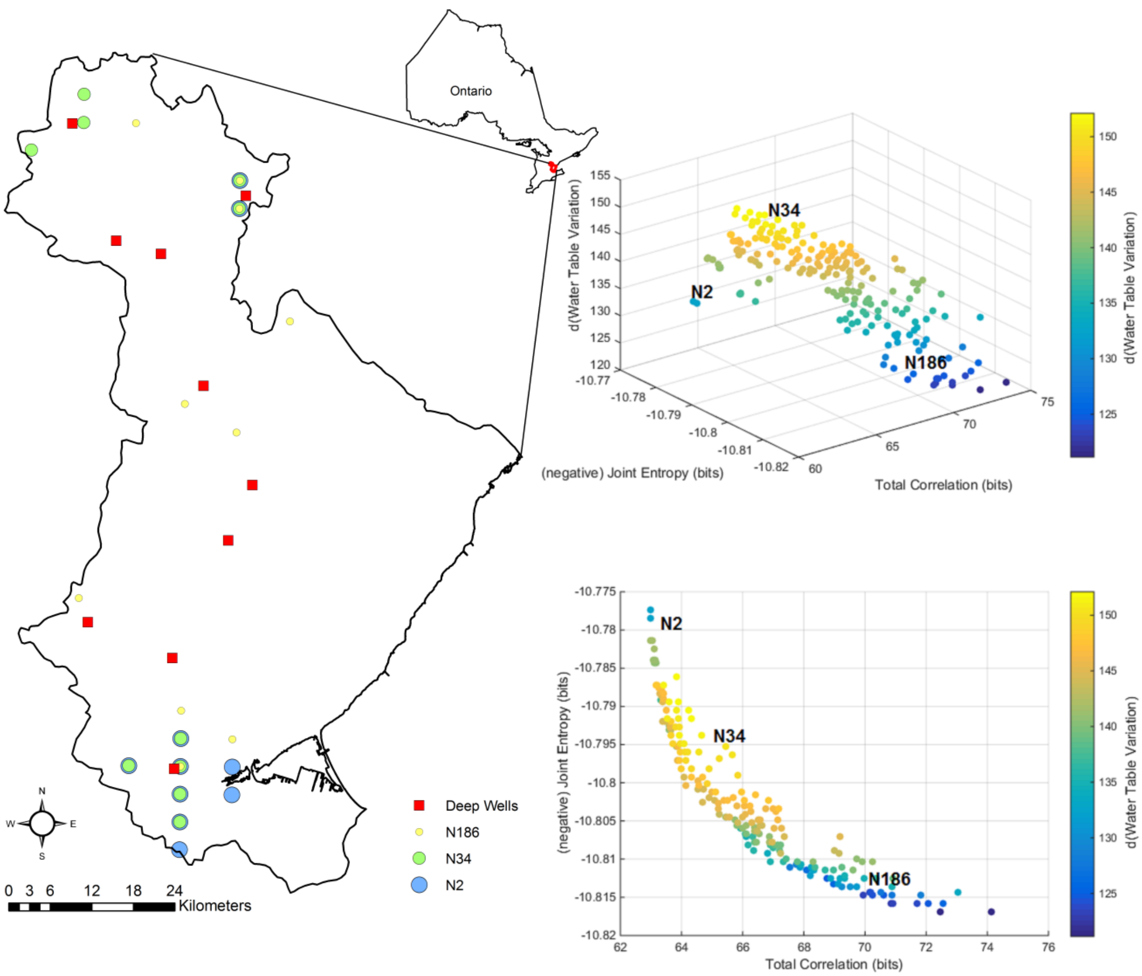

Ontario
Shallow Wells
N3960
N1198
N100
0 3 6 12 18 24
Kilometers
N1198
N100
N3960
d(Recharge)
(negative) Joint Entropy (bits)
Total Correlation (bits)
N100
N1198
N3690
(negative) Joint Entropy (bits)
Total Correlation (bits)
d(Recharge)

Scenario 4

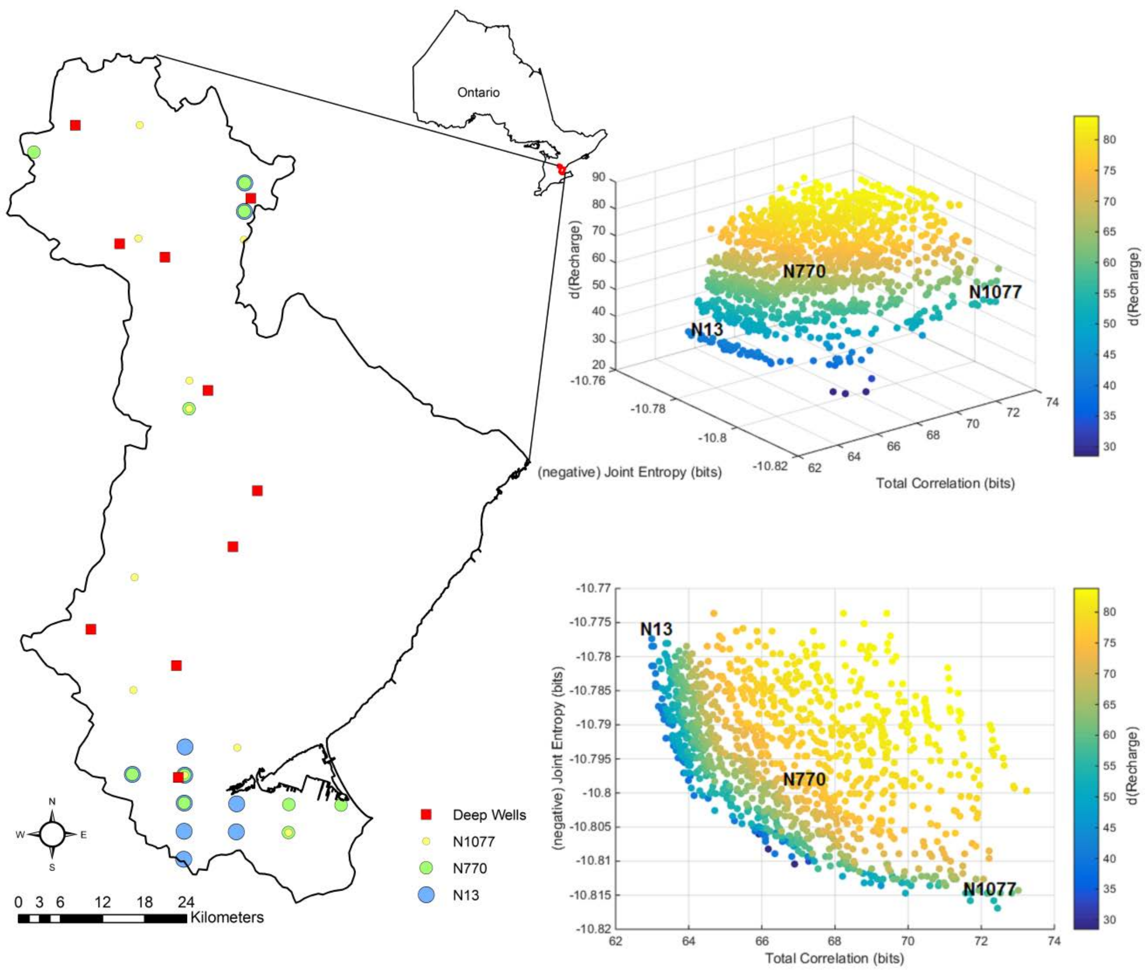
Appendix D: Example network designs and their corresponding Pareto fronts for augmenting the deep wells


Ontario
N
W
E
S
0 3 6 12 18 24
Kilometers
Deep Wells
N41
N20
N1
N1
N20
N41
(negative) Joint Entropy (bits)
Total Correlation (bits)
-10.775
-10.78
-10.785
-10.79
-10.795
-10.8
-10.805
-10.81
-10.815
-10.82
62
64
66
68
70
72
74

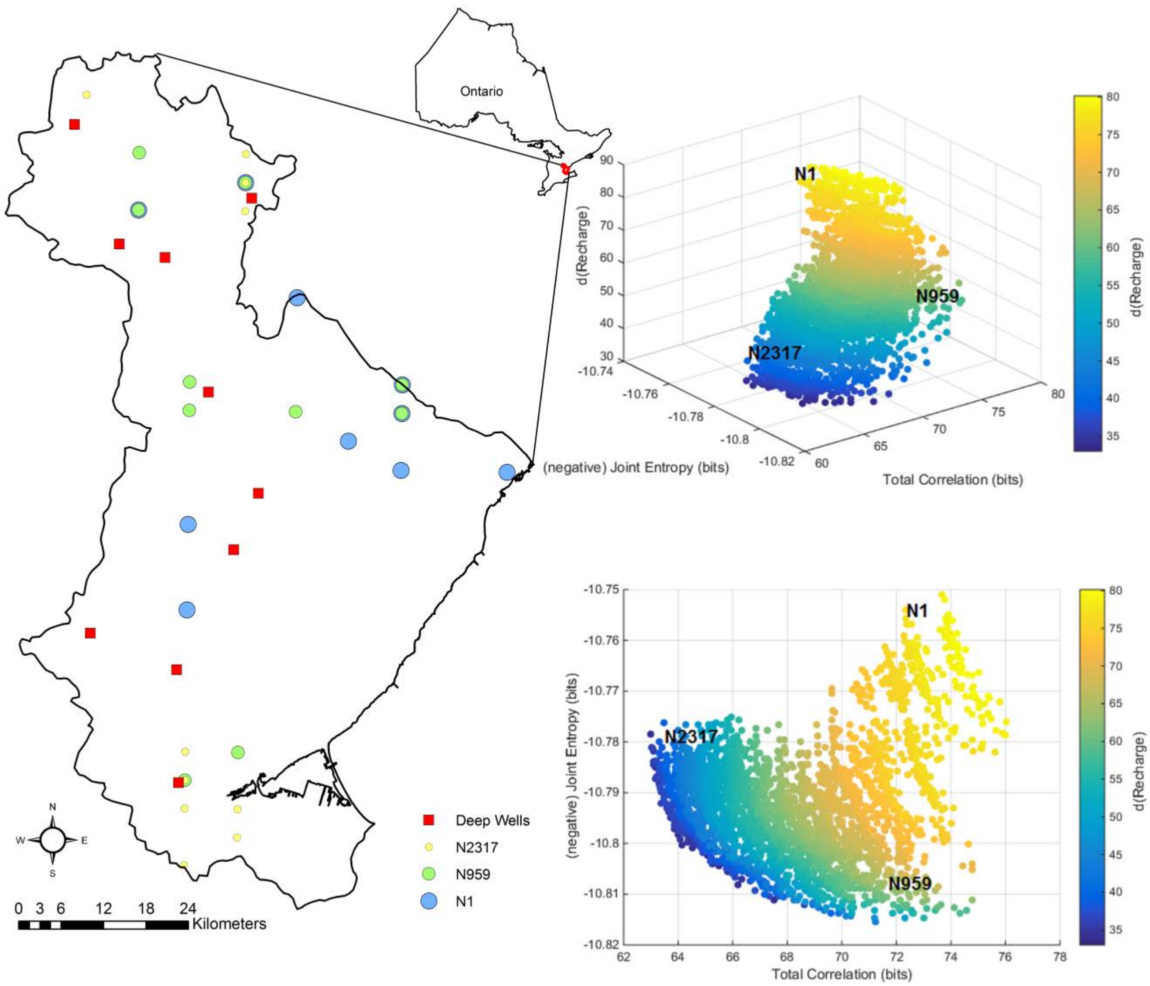
Scenario 1


Ontario
N
W
E
S
0 3 6 12 18 24
Kilometers
Deep Wells
N186
N34
N2
d(Water Table Variation)
(negative) Joint Entropy (bits)
Total Correlation (bits)

Scenario 2


Ontario
Deep Wells
N1077
N770
N13
0 3 6 12 18 24
Kilometers
d(Recharge)
(negative) Joint Entropy (bits)
Total Correlation (bits)
N13
N770
N1077

Scenario 3


Ontario
Deep Wells
N2317
N959
N1
0 3 6 12 18 24
Kilometers
N1
N959
N2317
d(Recharge)
(negative) Joint Entropy (bits)
Total Correlation (bits)
(negative) Joint Entropy (bits)
Total Correlation (bits)

Scenario 4