Project Title: **CBET-EPSRC Efficient Surrogate Modeling for Sustainable Management of Complex Seawater Intrusion-Impacted Aquifers**

Investigators:

Alex S Mayer, University of Texas at El Paso, United States (Project Grant Period: 01/01/2020 - 31/8/2022)

Domenico Bau, University of Sheffield, United Kingdom (Project Grant Period: 01/04/2020 - 3/31/2023)

**Major goals of the project**

The overarching goal of this research is the sustainable management of water resources in coastal regions with diverse geological, hydro-technical and governance settings. Pressures on water resources in coastal regions are already great and are expected to intensify due to increasing populations, standards of living and impacts from climate change and sea level rise (SLR). We will focus on coastal areas where aquifer over-drafting has caused seawater intrusion (SWI), thus deteriorating groundwater quality, and where SLR is expected to further reduce availability of fresh groundwater. Solutions to these problems will involve combinations of more efficient pumping schemes, demand reduction, and technological interventions such as desalination. However, determining optimal solutions for these problems poses extreme computational demands. This project will greatly advance the development and application of simulation-optimization (SO) by developing computationally efficient, robust, and accurate surrogate models for coastal groundwater systems.

**Accomplishments under these goals and objectives for 2021**

1. **Modified Henry’s Problem**. As one of the classic seawater intrusion problem set-ups, Henry’s problem (Henry, 1964) has been widely benchmarked and modified for further analysis (e.g., Javadi et al., 2012; other references). We adopted the original model set up from Henry (1964) in our investigation by enlarging the model domain to a quasi-2D domain with a dimension of 200-m long by 100-m deep by 1-m wide (into the page) (like the model dimension used in Javadi et al., 2011), and adding one single pumping well located in the central part of the aquifer (Figure 1). We developed a variable density groundwater model using SEAWAT that couples fluid flow and solute transport to conduct our analysis. The numerical model is discretized into 100 columns and 50 rows, with each cell having a dimension of 2m by 2m. The landward recharge is represented by a constant (1140.4 m3/day) fresh-water influx coming from the left side boundary. The coastal boundary is represented by a constant seawater head boundary for groundwater flow and point source (35g/L) boundary for the solute transport. The top and bottom boundaries are no-flux boundaries for both flow and transport.
   1. Homogeneous Cases: We chose three values (8.64, 86.4, 864 m/day or 0.0001, 0.001, 0.01m/s) of hydraulic conductivity (Hk) with variation up to 2 orders of magnitude for the homogeneous cases.
   2. Layered Cases: We chose two hydraulic conductivity values (86.4, 864 m/day or 0.001, 0.01 m/s) for a low-K and high-K layer, respectively and we compared two layering orders: i) low-K over high-K (e.g., Atoll Island); ii) high-K over low-K, with each layer spanning half of the domain depth (50m).
   3. Heterogeneous Cases: 2D heterogeneous spatially correlated hydraulic conductivity fields were generated using the Cholesky decomposition simulation method of the geostatistical package SGeMS (Remy et al., 2009). The variogram is set to be spherical with a fixed vertical correlation length of 10 m. We varied the horizontal correlation length among five different values (10, 40, 70, 100, 200m) that increases incrementally with the large values (100, 200m) to examine the impacts of aquifer connectivity to the ocean (Geng and Michael, 2020). For the log-normal reference distribution, we kept the mean constant, and varied the variance among 0.5, 1, and 2.

A screenshot of a computer

Description automatically generated with low confidence

**Figure 1.** Schematic of the modified Henry’s problem

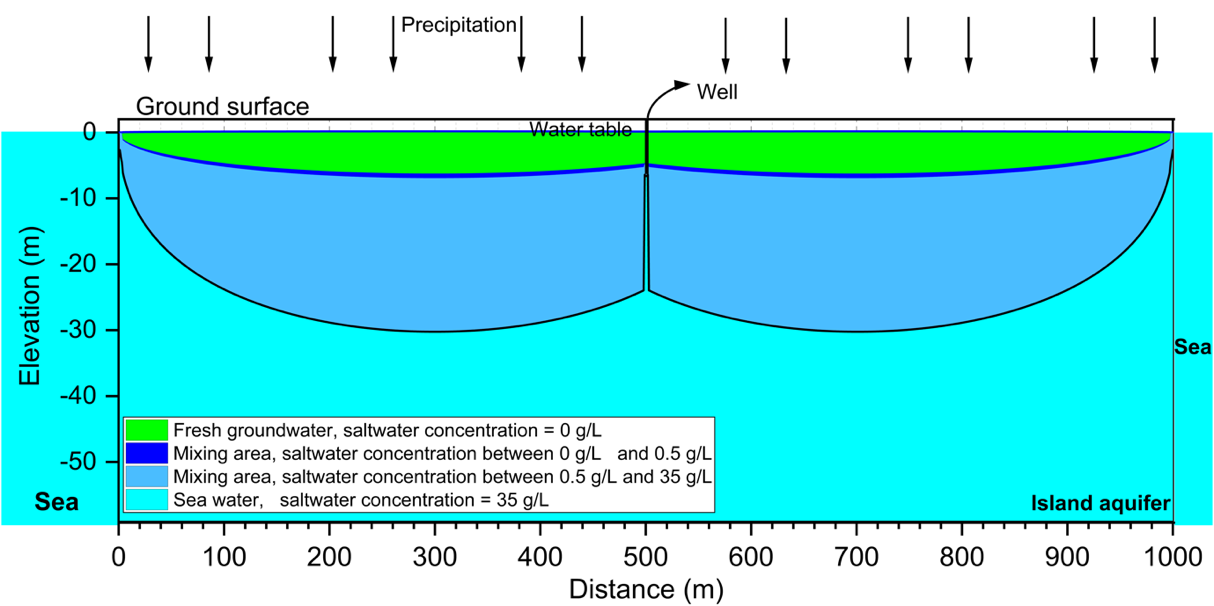
1. **Biscayne Aquifer.** The fully calibrated SEAWAT model of the Biscayne Aquifer in the southeast Florida has been developed by our USGS collaborators (Hughes et al., 2016). The active flow model area (denoted by the yellow shades in Figure 2) covers approximately 450 square miles (mi2) in the eastern part of Broward County and the north-eastern part of Miami-Dade County. It is bounded by the Everglades and several other water conservation areas to the west, and by the Atlantic Ocean to the east. The northern and the southern boundary of the model are bounded by the Hillsboro canal basin and the C-8 canal basin, respectively. A subset of the domain (190 mi2) simulating active solute transport extends from the Atlantic Ocean to the approximate inland location (4.5 mi) of the seawater/freshwater interface in the Broward County (denoted by the brown shades in Figure 2). The prominent management features or decision variables (DVs) of the Biscayne Aquifer are i) stages of the extensive canal network; ii) locations of the salinity control structures; iii) pumping rates of the extraction wells; iv) injection rates of the recharge wells.

Diagram, map

Description automatically generated

**Figure 2.** Map view of the active model domain for flow and transport. Surface water basins, canals, and salinity control structures are also denoted in the figure.

1. **Island Aquifer.** We consider the case of an island aquifer, as represented by the conceptual model shown in Figure 3. The aquifer consists of a thin freshwater lens recharged by the infiltration from precipitation. The lens thickness depends on the recharge intensity and the conductivity properties of the formation. Underneath this lens, salt concentration in water transitions gradually increases to seawater levels (35 g/L).



**Figure 3.** Schematic of island aquifer domain.

**Significant Results**

1. Modified Henry’s Problem - Homogeneous Cases. The salinity contours of each hydraulic conductivity with different pumping rates at the end of the simulation are shown in Figure 4. We observed that the most significant differences of the salinity contours occur at pumping rates of 0 and 1 unit of the freshwater influx (Qin). The response surfaces for three state variables (SVs) are shown in Figure 5. We found that all three SVs roughly follow the same trend as the pumping rate increases, but the toe position and the total dissolved mass seem to spread out more among different Hk values compared to the concentration at the well.

Shape

Description automatically generated

**Figure 4.** Salinity contours of the homogeneous cases of modified Henry’s problem

Chart, line chart

Description automatically generated

**Figure 5**. Response surfaces of three state variables (from the left to the right: toe position, concentration at the well, and the total dissolved salt) of homogeneous cases of Henry’s problem

1. Modified Henry’s Problem - Layered Cases. The salinity contours of each hydraulic conductivity with different pumping rates at the end of the simulation are shown in Figure 6. Like in homogeneous cases, we observed that the most significant differences of the salinity contours occur at pumping rates of 0 and 1 unit of the freshwater influx (Qin). And the fact that the second layering case (high K over low K) has a high K area near the top right corner is consistent with the observation that more saltwater is drawn towards the well at Qw equals 1 unit of Qin. The response surfaces of the three SVs are shown in Figure 7. We found that all three SVs tend to cluster closer to each other for the two layering orders compared to the corresponding homogeneous cases (Hk = 864 and 86.4 m/day).

Shape, polygon

Description automatically generated

**Figure 6.** Salinity contours of cases with layered hydraulic conductivity of modified Henry’s problem

**Chart, line chart

Description automatically generated**

**Figure 7**. Response surfaces of three state variables (from the left to the right: toe position, concentration at the well, and the total dissolved salt) of cases of layered hydraulic conductivity of Henry’s problem.

1. Modified Henry’s Problem - Heterogeneous Cases. The response surfaces of the three SVs (toe position, well concentration, and total dissolved mass) are shown in Figure 8, Figure 9 and Figure 10, respectively. The corresponding homogeneous case is denoted as the black connected dots. We found that although all three SVs tend to follow the general trend with increasing pumping rate, the toe position (Figure 8) is the SV that each realization of the heterogeneous case deviates the most from the homogeneous case as well as from the other realizations within the same geostatistical parameters, as the horizontal correlation length and the variance increase. We also found that the heterogeneity doesn’t have a significant impact on the response surfaces of the well concentration as they fall on top of each other (Figure 9), especially with large pumping rates (Qw > 2Qin). It is possibly because that at large pumping rates, most of the freshwater influx is extracted by the pumping well, and the well is so close to the intruding saltwater front (see Figure 4 and Figure 5 at Qw > 2Qin for reference) that the water withdrawn by the well is essentially a mixture of freshwater and saltwater, rendering the heterogeneity away from the well less important.

**Shape, polygon

Description automatically generated**

**Figure 8**. Response surfaces of the toe position for 10 realizations of hydraulic conductivity fields of heterogeneous cases of Henry’s problem (from the left to the right: sigma^2 is 0.5, 1, and 2; from the top to the bottom, lx = 10, 40, 70, 100, and 200m). The corresponding homogeneous case is denoted as the black connected dots.

**Shape, polygon

Description automatically generated**

**Figure 9**. Response surfaces of the well concentration for 10 realizations of hydraulic conductivity fields of heterogeneous cases of Henry’s problem (from the left to the right: sigma^2 is 0.5, 1, and 2; from the top to the bottom, lx = 10, 40, 70, 100, and 200m). The corresponding homogeneous case is denoted as the black connected dots.

**Chart, shape, polygon

Description automatically generated**

**Figure 10**. Response surfaces of the total dissolved mass for 10 realizations of hydraulic conductivity fields of heterogeneous cases of Henry’s problem (from the left to the right: sigma^2 is 0.5, 1, and 2; from the top to the bottom, lx = 10, 40, 70, 100, and 200m). The corresponding homogeneous case is denoted as the black connected dots.

1. Biscayne Aquifer. We started our sensitivity analysis of the Biscayne Aquifer by varying the pumping rates of the extraction wells. There are 26 composite pumping well fields across the actively simulated transport area. As a first-cut analysis, we uniformly varied the pumping rates of all the wells by the same percentage to see how the salinity distribution will respond to different pumping schemes. Four scenarios (20%, 50%, 200%, and 500% of the original pumping rate) of pumping scheme have been established in the sensitivity analysis, and we keep the original pumping scheme (100%) as the base case for comparing purposes. The results are displayed in Figure 11 with the salinity denoted as fraction of seawater.

Chart, histogram

Description automatically generatedA picture containing map

Description automatically generated

**Figure 11**. Salinity (in fraction of seawater) contours of Layer 10 (base of the Biscayne Aquifer) of the sensitivity analysis with uniformly-varying pumping rates (top: the whole map view and the region where the enlarged view of the bottom figures are located; bottom: from the left to the right are (A) base case, (B) 50% of the original pumping rate, (C) 20% of the original pumping rate, (D) 2 times of the original pumping rate, (E) 5 times of the original pumping rate.

1. Island Aquifer. To avoid the direct inclusion of such a model into the optimization loop, which is computationally impractical, a surrogate of the operation cost objective function has been developed. This is based on a Gaussian process regression, which assumes the cost function as the expected value of a multivariate stochastic process in the decision variable space, and thus allows for quantifying the confidence interval of the estimation. Figure 12 provides an example of the surrogate-cost optimization, where the decision variable of interest is the depth of the pumping well screen. A machine learning technique (ML) is developed to identify a sequence of model runs, that ultimately form a set of training data leading to a Gaussian process having a point of minimum that is close to that of the actual cost function. The ML approach uses the points of minimum of the surrogate function to extend the training dataset and thus updates the Gauss process iteratively until this converges to the actual cost function where its minimum is situated.

Chart

Description automatically generated

**Figure 12**. Gaussian Process surrogate model for optimized island aquifer.

**Opportunities for training and professional development**

* The US postdoc, Yipeng Zhang started 9/1/2020. He has learned how to simulate density-dependent subsurface flow and transport, use Gaussian process modeling to develop surrogate optimization models, and to use FlowPy software for setting up and processing results from the MODFLOW family of codes. He is a co-author on an upcoming paper on seasonal recharge impacts on seawater intrusion and will be first author on an upcoming paper on using Gaussian process modelling for surrogate models of optimizing Henry's problem with pumping.
* The UK postdoc, Mohammadali Geranmehr started 25/10/2021. He has learned how to simulate density-dependent subsurface flow and transport, as well as use FlowPy software for setting up and processing results from the MODFLOW family of codes. He has also started developing reduced-order modelling methods to apply to the develop of surrogate optimization models for the Santa Barbara aquifer.
* A US PhD student, Lauren Mancewicz, who works on another NSF seawater intrusion project, has been participating in our surrogate modelling group discussions and has been learning the principles of surrogate modelling. She has also trained the UK PhD student how to use the USGS SEAWAT mode l and FlowPy and how to set up an island-freshwater lens model. This has given her valuable experience in working on group projects.
* A UK PhD student, Weijang Yu, not directly involved in the project, has been participating in our surrogate modelling group discussions and has been developing machine learning based Gauss process surrogate models of Island aquifers, which has given him valuable experience in working on group projects.

**Results Dissemination**

* Paper on island aquifer simulations presented at the 2020 American Geophysical Union meeting.
* Paper on island aquifer surrogate modelling submitted for presentation at the 2022 American Geophysical Union meeting.

**Planned Activities for 2022**

The major activities in the current project year will include the following:

* Modified Henry’s problem: Complete modified Henry’s problem simulations and develop Gaussian process surrogate models for the response surfaces. Submit paper on this work. Present results at conference.
* Biscayne aquifer: Develop optimization problems for Biscayne aquifer in coordination with water resources agencies and USGS scientists. Use the simulation results to develop response surfaces for a range of state variables. Replace response surfaces with Gaussian process surrogate models. Submit paper on this work. Present results at conference. Develop reduced order model for the Biscayne aquifer.
* Santa Barbara aquifer: Begin simulations of the Santa Barbara aquifer. Develop optimization problems for Santa Barbara aquifer in coordination with water resources agencies and USGS scientists. Use the simulation results to develop response surfaces for a range of state variables.
* Island aquifer: Complete Gaussian process surrogate models. Submit paper on this work.

**References**

1. Dausman, A., & Langevin, C. D. (2005). Movement of the saltwater interface in the surficial aquifer system in response to hydrologic stresses and water-management practices, Broward County, Florida. US Department of the Interior, US Geological Survey.
2. Geng, X., & Michael, H. A. (2020). Preferential flow enhances pumping‐induced saltwater intrusion in volcanic aquifers. Water Resources Research, 56(5), e2019WR026390.
3. Henry, H. R. (1964). Effects of dispersion on salt encroachment in coastal aquifers, in"Seawater in Coastal Aquifers". US Geological Survey, Water Supply Paper, 1613, C70-C80.
4. Hughes, J. D., Sifuentes, D. F., & White, J. T. (2016). Potential effects of alterations to the hydrologic system on the distribution of salinity in the Biscayne aquifer in Broward County, Florida (No. 2016-5022). US Geological Survey.
5. Javadi, A. A., Abd‐Elhamid, H. F., & Farmani, R. (2012). A simulation‐optimization model to control seawater intrusion in coastal aquifers using abstraction/recharge wells. International Journal for Numerical and Analytical Methods in Geomechanics, 36(16), 1757-1779.
6. Langevin, C.D., D.T. Thorne Jr., A.M. Dausman, M.C. Sukop and Weixing G. (2008). SEAWAT Version 4: A Computer Program for Simulation of Multi-Species Solute and Heat Transport: U.S. Geological Survey Techniques and Methods Book 6, Chapter A22, 39 p.
7. Rasmussen C.E. and C.K.I. Williams (2006). Gaussian Processes for Machine Learning, The MIT Press, ISBN 0-262-18253-X.
8. Remy, N., Boucher, A., & Wu, J. (2009). Applied geostatistics with SGeMS: A user's guide. Cambridge University Press.