

# Advective Transport: Importance to Groundwater Remediation and Illustration of Basic Transport Concepts to Introductory Audiences

By

© 2019

Andrew T. Banks

B.S. Mathematics, University of Kansas, 2018

B.S. Geology, University of Kansas 2016

Submitted to the graduate degree program in Geology and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Master of Science.

---

Chair: Mary Hill

---

Rick Devlin

---

Randy Stotler

Date Defended: May 10 2019

The thesis committee for Andrew Banks certifies that this is the approved version of the following thesis:

**Advective Transport: Importance to Groundwater Remediation and Illustration of Basic Transport Concepts to Introductory Audiences**

---

Chair: Mary Hill

Date Approved: May 10, 2019

## Abstract

Developing effective transport models with meaningful results requires a solid understanding of transport concepts and the underlying mechanics the model, and how data can be best integrated into the model. This research makes contribution to each of these three requirements: concept education, understanding simulated mechanics, and integrating data into models.

(1) How can groundwater flow and transport processes be well communicated to introductory audiences, while providing a foundation for complex model development and interpretation? The first part of this work presents *GroundWaterTutor*, a freely available computer module for groundwater education. *GroundWaterTutor* provides a simple, interactive environment for learning how key system characteristics affect hydraulic heads and the flow of tracer particles. The software was developed using *MATLAB* in conjunction with *MODFLOW 2005* and *MODPATH 6*, and thus provides a solid foundation from which students can expand to simulating more complex situations. *GroundWaterTutor* is distributed as a set of freely available standalone executables.

(2) How do simulated advection interact with dispersion in groundwater remediation simulations? This question is addressed in the context of the following research question: How well do advection-based metrics for assessing the effectiveness of active in situ groundwater remediation strategies work? Results are important to developing an efficient optimization framework for in-situ active remediation systems.

(3) Can heteroscedastic data, like concentration data, be integrated into models, such as groundwater models, without log-transformations, which make results hard for many users to interpret? Here the use of error-based weighting methods are investigated, which provide more intuitive regression models than log-transformation in the presence of highly variable (e.g.

heteroscedastic) data. For this problem, log-transformation produced good model fit, while the error-based weighting formulations tested worked poorly.

### **Acknowledgements**

Throughout my time in this department, and at this university, I have had the opportunity to learn from many individuals. As such, it is likely that some acknowledgements may elude me.

Foremost, I would like to thank my advisor, Mary Hill, whose thoughtful criticism, mentorship and guidance facilitated immense academic and personal growth over the last 3 years. She provided me with substantial freedom in my academic pursuits, along with encouragement to challenge myself in ways I would otherwise have not endeavored. For this I cannot thank her enough.

I would like to extend gratitude to the members of my advising committee, Rick Devlin, Erik Van Vleck and Randy Stotler. All three have provided valuable insight and constructive criticism that has undoubtedly improved the quality of this work.

I would also like to thank my family, namely my parents and grandparents. They have provided unconditional support, patience and wisdom, without which this work would not have been possible. Thank you to my office mates and dear friends (of which there are too many to name) who have left me with many fond memories over the years.

I would like to extend a special thanks to Anthony Walton, who provided with valuable mentorship throughout my time at KU. There are many faculty members who have had a substantial impact on my academic pursuits and personal growth. The environment facilitated by the Geology department administrative staff and faculty is exceptional and unique. I would like to extend my final thank you to this group as a whole. Their behind-the-scene efforts are what make this department such a great place.

## Table of Contents

Chapter 1:.....	1
1.1 Introduction.....	1
1.2 Groundwater Flow and Transport Interactive Exercise .....	2
1.2.1 Exercise Prologue .....	2
1.2.2 Installation.....	3
1.2.3 Model Aquifer and Main User Interface.....	5
1.2.4 Model Results .....	12
1.2.5 Example Problem.....	15
1.2.6 Exercise Epilogue .....	17
1.3 References.....	18
Chapter 2:.....	19
2.1 Introduction .....	19
2.1.1 Importance of Mixing .....	19
2.1.2 Mixing, Spreading and Dilution .....	19
2.1.3 Objective .....	20
2.2 Background .....	21
2.2.1 Mixing in Aquifers.....	21
2.2.2 Spreading by Chaotic Advection .....	21
2.2.3 Engineered Injection and Extraction.....	22
2.3 Motivation .....	24
2.4 Modeling Methods .....	25
2.4.1 Model Aquifer and Engineered Injection Extraction System .....	25
2.4.2 Groundwater Flow and Transport .....	26
2.5 Contaminant Degradation Simulations .....	29
2.6 Mathematical Methods for Quantifying Spreading.....	35
2.6.1 Length Stretch.....	36
2.6.2 Principal Stretch.....	37
2.6.3 Affine/Non-affine deformation .....	39
2.6.4 General Methods for Evaluating Spreading Metrics .....	42
2.7 Results .....	44
2.7.1 Spatial Distribution of Spreading.....	44
2.7.2 Time-Dependent Spreading .....	45

2.8	Discussion .....	48
2.8.1	Spatial Distribution of Spreading.....	48
2.8.2	Time-Dependent Spreading .....	50
2.9	Conclusion .....	52
2.10	Appendix A.....	54
2.11	References.....	55
Chapter 3	.....	59
3.1	Introduction.....	59
3.2	Hypothesis to Test.....	60
3.3	Mathematical Methods.....	60
3.3.1	Null Case: No Transformation or Weighting.....	60
3.3.2	Log Transformation .....	61
3.3.3	Error Based Weighting .....	62
3.4	Dataset.....	63
3.5	Results.....	64
3.6	Discussion and Conclusions .....	65
3.7	References.....	66

## List of Figures

Figure 1 GroundWaterTutor loading screen, featuring an image of Henry Darcy. ....	3
Figure 2 Boundary Conditions tab in GroundWaterTutor interface main window. ....	4
Figure 3 Magnified view of the tabs along the top left margin of the GroundWaterTutor interface main window. ....	5
Figure 4 Initial conditions tab in GroundWaterTutor interface main window. ....	6
Figure 5 Parameters tab in GroundWaterTutor interface main window ....	8
Figure 6 Pumping and Recharge tab in GroundWaterTutor interface main window ....	9
Figure 7 Run MODFLOW/MODPATH tab in GroundWaterTutor interface main window ....	10
Figure 8 Command Prompt windows that appear when MODFLOW and MODPATH are running ....	11
Figure 9 Hydraulic Head tab in the GroundWaterTutor results window ....	13
Figure 10 Particle Tracking tab in the GroundWaterTutor results window ....	14
Figure 11 Idealized depiction of plume spreading by chaotic advection ....	22
Figure 12 Map view of the inner part of the model aquifer in its initial configuration ....	25
Figure 13 Particle positions and species based on advective-dispersive and reactive transport simulations. Case A .....	30
Figure 14 Particle positions and species based on advective-dispersive and reactive transport simulations. Case B .....	31
Figure 15 Particle positions and species based on advective-dispersive and reactive transport simulations. Case C .....	32
Figure 16 Spreading metric based on advective transport simulation. Spatially averaged value of Length stretch. .....	33
Figure 17 Simulated contaminant concentration and cumulative pumping for each EIE sequence ....	33
Figure 18 Idealized depiction of the parameterization of a single particle cluster, and its evolution under the flow map .....	35
Figure 19 Idealized depiction of the procedure for computing the local length stretch for a single particle cluster .....	36



Figure 20 Idealized depiction of the procedure for computing the deformation gradient for a single particle cluster.....	37
Figure 21 Geometric interpretation of the principal stretch.....	38
Figure 22 Geometric representation of displacement vectors.....	39
Figure 23 Geometric representation of the linear model .....	40
Figure 24 Geometric representation of the affine/non-affine deformation .....	41
Figure 25 Map view of the inner part of the model aquifer, illustrating the initial arrangement of particle clusters used to evaluate spreading metrics .....	43
Figure 26 Spatial distribution of spreading metrics after step 8 of the Case A EIE sequence .....	44
Figure 27 Spatial distribution of the principal stretch after step 12 of the Case A, Case B and Case C EIE sequences.....	45
Figure 28 Spatially averaged values of the principal stretch.....	46
Figure 29 Spatially averaged values of the affine deformation .....	47
Figure 30 Spatially averaged values of the non-affine deformation .....	47
Figure 31 Spatially averaged values of the contribution ratio .....	47
Figure 32 Histogram of dependent variable .....	63
Figure 33 Residuals and predicted values plotted against observed dependent values for each MLR model .....	64
Figure 34 .Error based observation weights assigned to native values of the dependent variable. ....	65

## Prologue

Predicting contaminant fate and transport through geologic porous media is complex and challenging. Uncertainty is inherent in all models of geologic systems, making the task of simulating groundwater flow and solute transport difficult. This includes education, development, interpretation and stakeholder communication. Practical limitations inhibit the collection of spatially and temporary dense field measurements, so that observations provide only sparse information for model calibration and prediction uncertainty reduction. Sparse information limits attempts to represent the subsurface in detail, while commonly long model execution times and computational limitations determine the methods available for integrating models and data. Modelers end up having to choose from calibration and uncertainty evaluation designs with broad simplifications of transport kinetics and dynamics, either by imposing very limited structures with few defined degrees of freedom, or larger degrees of freedom that are highly regularized (Hill et al., 2015; Doherty and Welter, 2010).

The overwhelming array of subsurface transport processes to account for often leads modelers to develop unnecessarily complex representations of the system, degrading the utility of the model as an effective decision-making tool. Developing transport models with meaningful results requires a solid understanding of transport concepts, underlying model mechanics, and how data can best be integrated into the model. This thesis makes contributions to all three of these fundamental issues.

A substantial body of educational research suggests that students must engage in real-world uses of the theories and concepts presented in the classroom in order to effectively employ them outside of the classroom (*Li and Liu, 2003*). Teaching subsurface transport concepts has

been approached in many ways, but none of the existing methods were found to satisfy both of two important goals: emphasize basic flow and transport concepts, and provide a solid foundation for simulating more complex systems. For example, *Li and Liu (2003)* satisfies the first goal, but used a custom model that does not prepare students well for models they are likely to use in practice. *Vallochi et al (2015)* introduces users to the sophisticated and common models MODFLOW and MODPATH, challenging the user with a very interesting problem, including wetlands, a river, and pumping. However, less emphasis is placed on underlying basic concepts, such as how hydraulic properties, recharge and pumping interact. Understanding the practical manifestations of what are fundamentally mathematical relationships is vital to using models as effective decision making tools. In the first chapter of this work, the challenges of understanding flow and transport concepts, and underlying model mechanics are addressed through the development of *GroundWaterTutor*, a freely available computer module for groundwater education. *GroundWaterTutor* provides a simple, interactive environment for learning how key modeling parameters affect hydraulic heads and the flow of advected tracer particles. The software was developed using *MATLAB* in conjunction with *MODFLOW 2005* and *MODPATH 6*, and thus provides a solid foundation from which students can expand to simulating more complex situations. Chapter 1 describes and demonstrates *GroundWaterTutor*.

Transport in geologic porous media is governed by many processes, including advection, dispersion, diffusion, reaction, sorption and radioactive decay. Although every geologic system is unique, solute transport is typically dominated by advection and dispersion. The interaction of advective and dispersive processes is important in part because advection is fast to calculate, and can be applied meaningfully at a wide range of model discretization levels. Adding dispersion creates a much more computationally demanding problem, for which accurate numerical results

require highly refined discretization. An interesting circumstance in which to study these interactions arises during the remediation of subsurface contaminants using engineered approaches.

The principle of parsimony refers (i.e., Occam's razor) to the idea that things are usually connected or behave in the simplest way. In the second chapter of this work, the challenges associated with developing effective models of complex systems is addressed in the context of the following research question: How well does simulated advection represent what are always actually advective-dispersive systems? Here, several advection-based metrics are investigated in terms of their ability to characterize the effectiveness of active in-situ groundwater treatment systems. These systems act as subsurface "mixers", promoting contaminant dilution and degradation reaction with injections and extractions of water from nearby pumping wells. Practical considerations require the design of active treatment be heavily informed by numerical simulations, and impose unique constraints during model calibration. Typical approaches require simulation of advective-dispersive-reactive transport, which provide a basis for calibrating simulated concentrations to field measurements. The complexity of transport within active treatment systems require broad simplifications of transport kinetics and dynamics, which complicates comparison of model predictions to field measurements. Advection-based metrics capture the most fundamental transport process and provide a computationally frugal alternative to advective-dispersive-reactive based metrics. The importance of these results pertain to the development of an efficient optimization framework for the design of in-situ active remediation systems.

The data available to modelers is heavily influenced by both scientific and practical considerations. Often the relevant components of a systems are hard to measure, or produce

measurements that have undesirable or limiting statistical assumptions. To effectively integrate field data, modelers are forced to introduce transformations which are often misused or results are difficult to understand. For example, log transformations are useful because many types of data have increasing variance as the values get larger. This characteristic is a type of heteroscedasticity, and is common to phenomena studied by virtually all scientific disciplines. In addition to the concentration data of primary interest in this work, streamflow data also tends to exhibit heteroscedasticity. Although log-transformation is the most commonly used remedy for heteroscedasticity, the results of many statistical tests applied to log-transformed data are not relevant or difficult to relate to the native dataset (Feng et al., 2014). The third chapter of this thesis addresses the problem of how data can be best integrated into models by addressing the following research question: Can heteroscedastic data, like concentration data, be integrated into models, such as groundwater models, without log-transformations, which make results hard for many users to interpret? Here the use of log-transformation and error-based weighting methods in the presence heteroscedastic dependent variables are investigated. Error-based weighting encourages modelers to carefully consider sources of error, whether experimental or epistemic, and provide more intuitive regression models than log-transformation. As a set, the three chapters of this dissertation contribute to three primary issues of concern in the simulation of transport in subsurface systems.

Doherty, J., & Welter, D. (2010). A short exploration of structural noise. *Water Resources Research*, 46(5). doi:10.1029/2009wr008377

Hill, M. C., Kavetski, D., Clark, M., Ye, M., Arabi, M., Lu, D., . . . Mehl, S. (2015). Practical Use of Computationally Frugal Model Analysis Methods. *Groundwater*, 54(2), 159-170. doi:10.1111/gwat.12330

Feng, C., Wang, H., Lu, N., Chen, T., He, H., Lu, Y., & Tu, X. M. (2014). Log-transformation and its implications for data analysis. *Shanghai archives of psychiatry*, 26(2), 105–109. doi:10.3969/j.issn.1002-0829.2014.02.009

Li, S. and Liu, Q. (2003), Interactive Groundwater (IGW): An innovative digital laboratory for groundwater education and research. *Comput. Appl. Eng. Educ.*, 11: 179-202. doi:10.1002/cae.10052

## Chapter 1:

### ***GroundWaterTutor* – An Interactive Computer Module for Understanding Groundwater Flow and Transport**

#### **1.1 Introduction**

Communicating the basic principles of groundwater flow and transport to students can be challenging. This work presents *GroundWaterTutor*, a freely available interactive computer module for groundwater education. *GroundWaterTutor* provides a simple, interactive environment for students to learn how key modeling parameters affect hydraulic heads and the flow of tracer particles. Students are presented with options to include the effects of confined and unconfined conditions, heterogeneity, anisotropy, time-discretization, areal recharge, and pumping rates, which allows for a wide range of scenarios to be explored. Interactive visualizations illustrate the resulting hydraulic heads, as well as the transport of tracer particles from three origination sites. The software was developed using MATLAB GUI in conjunction with MODFLOW 2005 and MODPATH 6, and is distributed as a set of standalone executables. A sample exercise to accompany *GroundWaterTutor* is provided, which poses students with several tasks; one of which is to find the largest possible pumping rate without extracting too many “contaminant” particles. This exercise also utilizes a free web applet designed to illustrate the effects of urban and agricultural development on groundwater resources (Valocchi et al., 2015). Classroom trials suggest that these programs complement each other nicely. Feedback from students was overall positive.

## 1.2 Groundwater Flow and Transport Interactive Exercise

### 1.2.1 Exercise Prologue

In this project, students are tasked with managing a subsurface hydrologic system containing contaminants and a single pumping well. Students are be provided some prior knowledge of the hydrologic characteristics of the system, with the objective of estimating the maximum achievable pumping rate such that no more than 50% of the contaminant particles enter the well.

*GroundWaterTutor* graphical interface operates in conjunction the U.S. Geological Survey programs MODFLOW 2005 (Harbaugh et al, 2005), MODPATH 6 (Pollock et al, 2016) and a small library of MATLAB support code. MODFLOW is based on a finite differencing approximation of the groundwater flow equation, and is used by *GroundWaterTutor* to simulate hydraulic heads. MODPATH is a post-processor which is used by *GroundWaterTutor* to simulate advection based particle trajectories based on models of groundwater flow produced by MODFLOW.

The graphical interface was developed with MATLAB 2017, and provides the capability to create input files for MODFLOW and MODPATH, execute the models and visualize model results. A small library of support code was developed to translate user input from the graphical interface into text files readable by MODFLOW and MODPATH.

The program files required to run *GroundWaterTutor*, along with the source code can be acquired for free at <https://github.com/andrewtbanks/GroundWaterTutor>. Users are required to have administrative privileges to a computer with a windows operating system. In theory *GroundWaterTutor* can be used on MacOS or Linux operating systems, however users will be required to manually compile MODFLOW 2005 and MODPATH 6 into standalone executables



suitable for the respective operating system. This will be a challenging process for typical users. GroundWaterTutor has not been tested on MacOS or Linux systems.

### 1.2.2 Installation

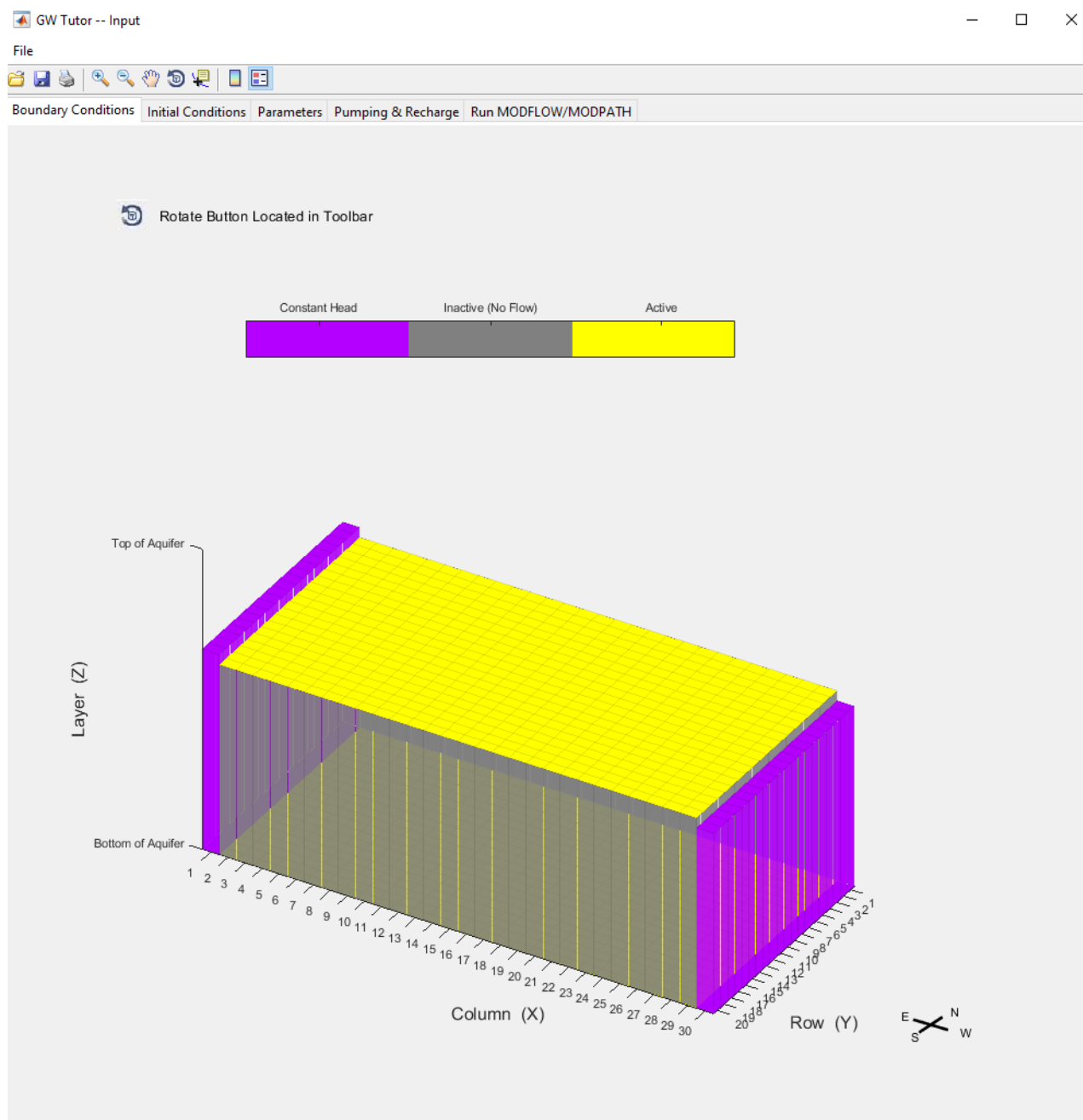
The repository “GWtutor” should be downloaded and saved to a directory containing no spaces. For example “...\One Drive University of Kansas\GWtutor “ would be insufficient, and result in *GroundWaterTutor* failing to locate the directory containing model input files.

Once downloaded, execute “GWtutor.exe”. Although GroundWaterTutor does not require a MATLAB license, installation of freely available support library MATLAB Runtime 2017 is require. The user will be prompted to begin the MATLAB Runtime installation automatically when GWtutor.exe is executed. At this point the user is required to provide credentials for administrative access to the computer.

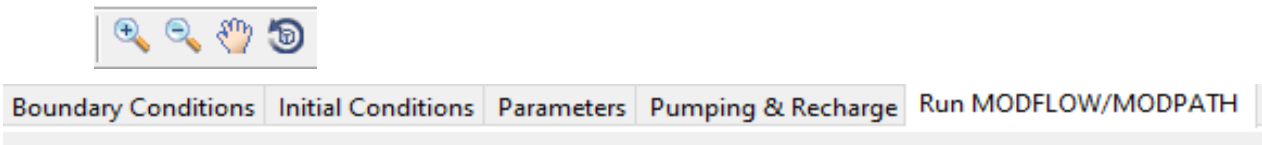
Upon successful installation, a loading screen will appear, followed shortly by the main window for the GroundWaterTutor interface (shown in Figures 1 and 2). The MATLAB Runtime installation will only be prompted if it is not already installed on the machine. The loading screen shown in Figure 1 will appear directly upon execution of “GWtutor.exe” once MATLAB Runtime has been installed.



**Figure 1.** *GroundWaterTutor* loading screen, featuring an image of Henry Darcy.



**Figure 2.** *Boundary Conditions* tab in *GroundWaterTutor* interface main window.



**Figure 3.** Magnified view of the tabs along the top left margin of the *GroundWaterTutor* interface main window.

### 1.2.3 Model Aquifer and Main User Interface

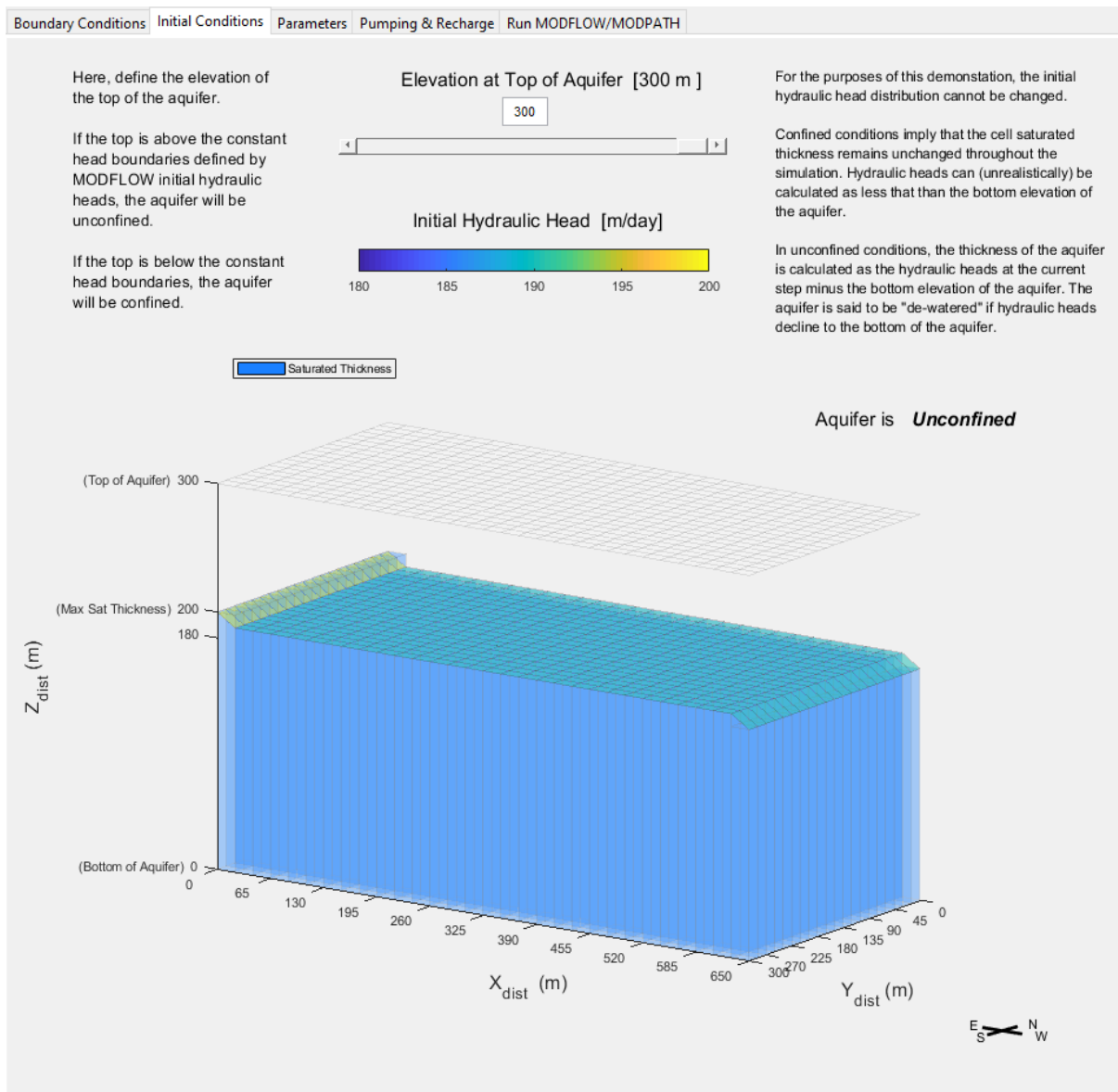
Five tabs along the top left margin of the main interface window can be explored by users (Figure 3 and top left of Figure 2). Each tab contains information about the model under construction. The zoom, pan and rotate toolbar can be used to manipulate how the model input is displayed. A compass showing the four cardinal directions can be seen in the bottom left of all tabs. The compass reorients as the model object is rotated. A brief explanation of each tab and relevant details about the model is provided in sections 1.2.3a-e.

#### 1.2.3a Boundary Conditions

The *Boundary Conditions* tab provides the first view of the model aquifer to users. In this model, neither the boundary conditions nor the spatial discretization of the model domain can be changed. The model spatial discretization consists of 30 columns by 20 rows, uniformly spaced, extending 650 m in the x-direction (E-W), and 300 m in the y-direction (N-S). No flow boundary conditions are enforced along the North and South. Constant head boundary conditions are enforced along the East ( $h = 200m$ ) and West ( $h = 180 m$ ), yielding a regional hydraulic gradient of  $\frac{dh}{dx} = 0.03$ .

### 1.2.3b Initial Conditions

The *Initial Conditions* tab shows information about the initial guess for hydraulic heads input used in the MODFLOW BAS6 package (shown in Figure 4). Here the user may specify the two elevations of the aquifer using the slider bar, or manual input features in the top center of the window. Adjusting the top elevation will affect whether the aquifer is under confined or unconfined conditions. The initial hydraulic heads cannot be changed by the user. It is important



**Figure 4.** *Initial conditions* tab in *GroundWaterTutor* interface main window.

to understand that the initial hydraulic head distribution provided to the user is physically implausible and does not reflect solutions to the groundwater flow equation. This is simply a starting point for MODFLOW to attain a steady state solution. The abrupt head change at the East and West boundaries and the flat head distribution throughout the rest of the domain would violate mass conservation requirements for the groundwater flow equation.

### **1.2.3c Parameters Tab**

The *Parameters* Tab provides the user with visualizations of the parameter fields used in the MODFLOW LPF package, and MODPATH MPSIM package (shown in Figure 5). The dropdown menu located in the top center of the window allows users to select which parameter field is displayed. These include Hydraulic conductivity, specific storage, specific yield and porosity. The slider bar and manual edit box underneath the dropdown menu can be used to scale the selected parameter. The spatial distribution of a given parameter is forced to be homogenous and isotropic. The exception to this is Hydraulic conductivity. Anisotropy can be introduced to the hydraulic conductivities using the manual edit box on the right side of the window. The checkboxes on the left side of the figure allow the user to introduce heterogeneity into the hydraulic conductivity field. The “case 1” checkbox introduces three-cell wide stripe of high hydraulic conductivities running E-W, which are 10 times greater than the baseline values. Similarly, the “case 2” checkbox adds 3-cell wide stripe of low hydraulic conductivities running N-S, which are 10 times smaller than the base line values. If both “case 1” and “case 2” checkboxes are selected, the high and low hydraulic conductivity stripes will be superimposed (shown in Figure 5).

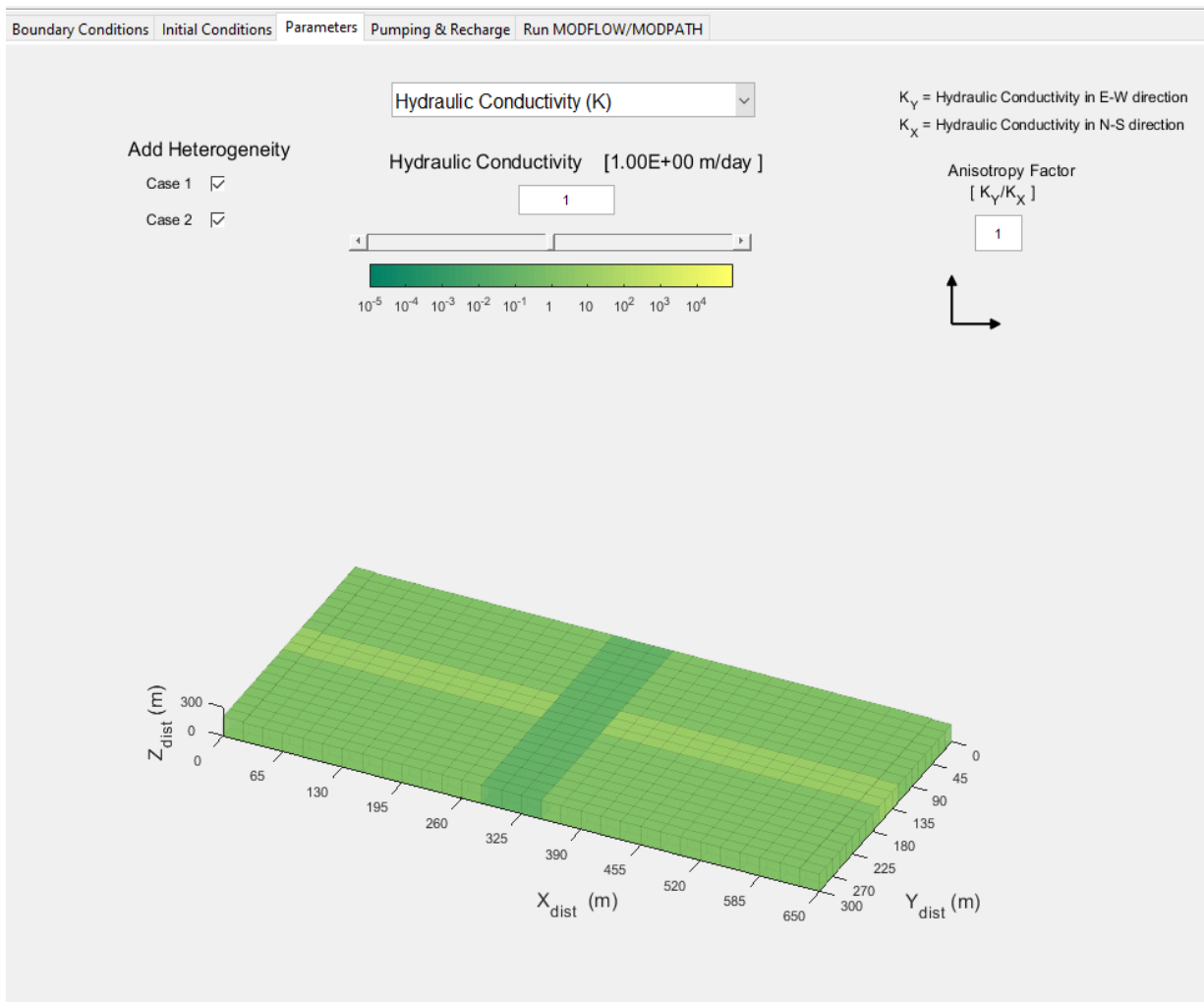
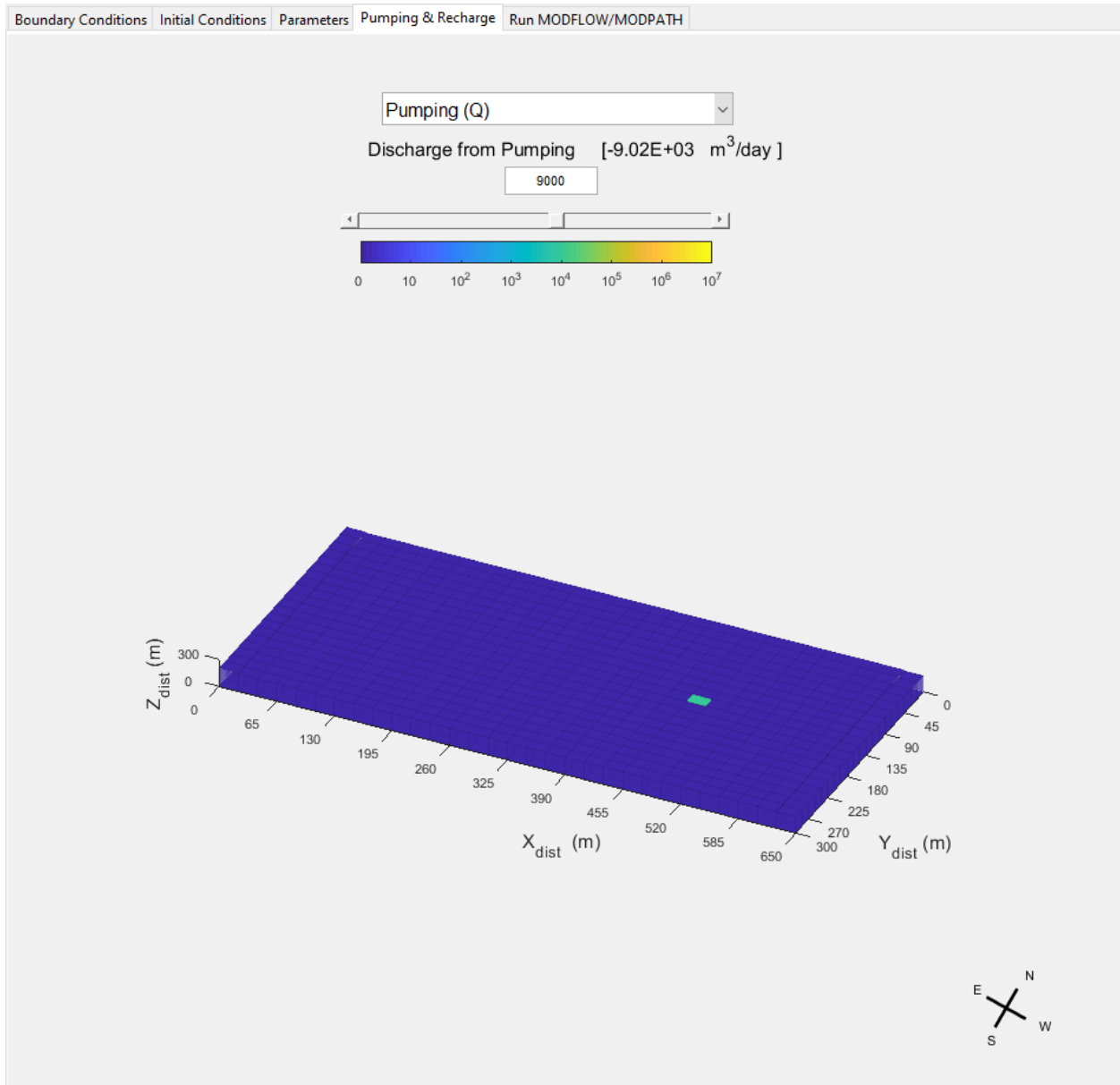


Figure 5. Parameters tab in GroundWaterTutor interface main window.

### 1.2.3d Pumping and Recharge

The *Pumping and Recharge* tab provides the user with visualizations of the source fields used by the MODFLOW WEL and RCH packages (shown in Figure 6). The dropdown menu allows the user to toggle between a display of uniform surface recharge, and discharge from a single pumping well, centered in the western half of the model. The slider bar and manual edit box underneath the drawdown menu can be used to scale the selected source term.



**Figure 6.** Pumping and Recharge tab in GroundWaterTutor interface main window.

### 1.2.3e Run MODFLOW/MODPATH

The *Run MODFLOW/MODPATH* tab provide the user with a summary of the model. The saturated thickness of the aquifer is shown by default (see Figure 7). The checkboxes in the top left of the window superimpose the boundary conditions, the location and discharge rate for the pumping well, and the initial position of three clusters of tracer particles. Three groups of tracer

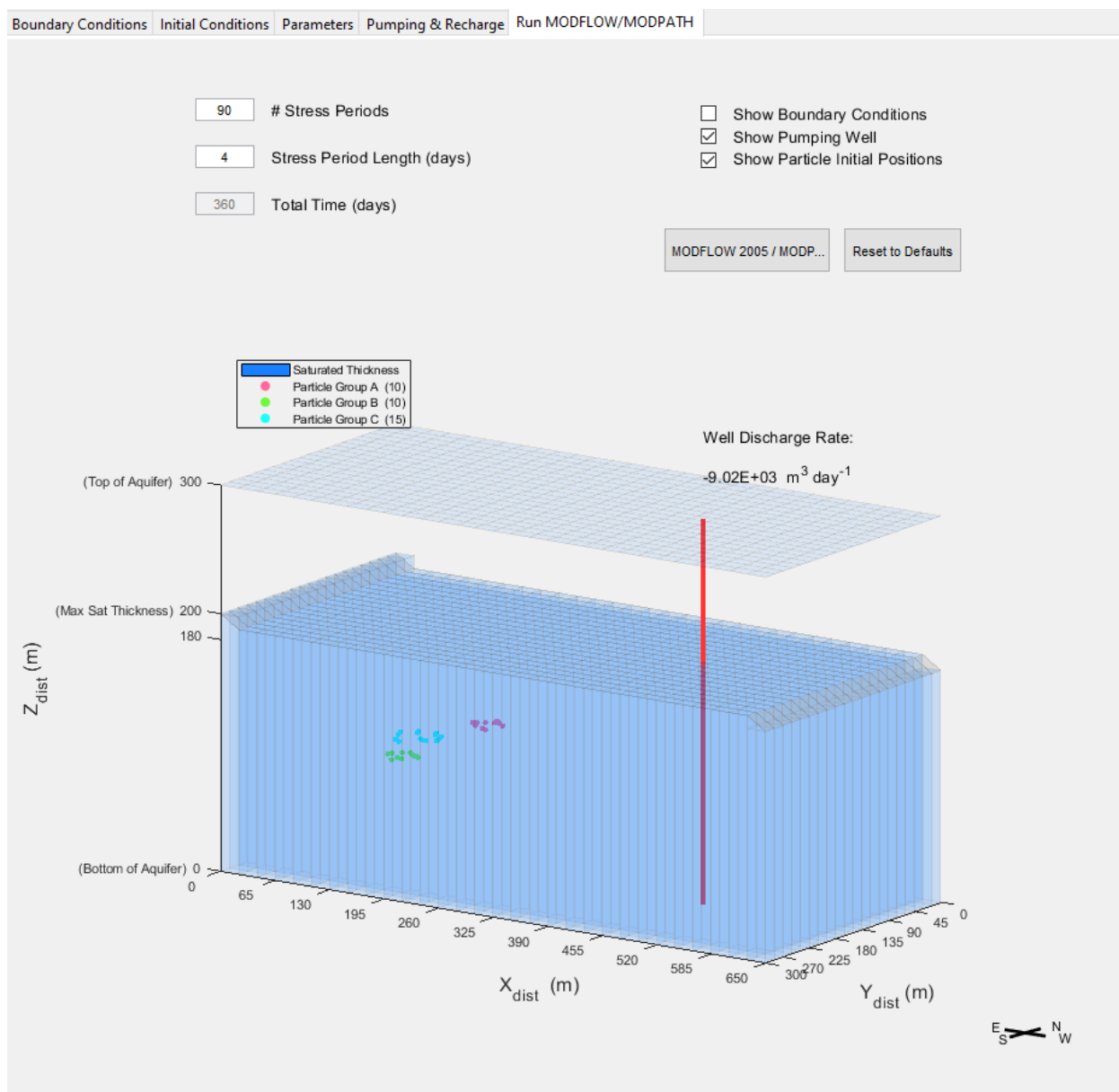


Figure 7. *Run MODFLOW/MODPATH* tab in *GroundWaterTutor* interface main window.



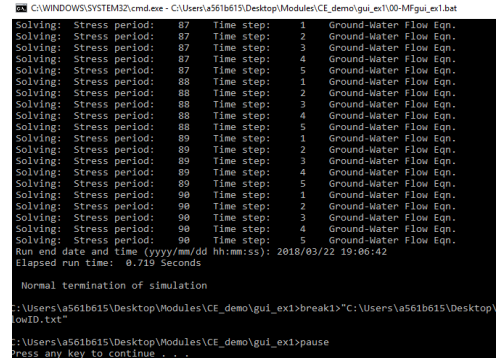
particles (10 green, 10 pink and 15 blue) originate in the East side of the model domain. Here the user can modify the time discretization used in the MODFLOW DIS package. The number of transient stress periods and the length of each stress period (in days) can be specified in the manual edit boxes at the top left of the window. Note that the features provided by *GroundWaterTutor* for visualizing model results work best when the number of stress periods exceeds 20. Specifying less than 20 stress periods causes the sliders to behave erroneously when displaying time dependent hydraulic heads and particle paths in the results window.

The “Reset to Defaults” will reset all inputs to their default values. The same result can be achieved by simply restarting the *GroundWaterTutor* interface.

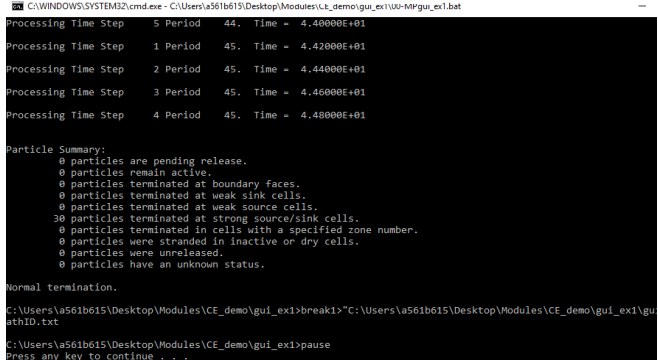
To account for the physically implausible initial hydraulic head distribution provided to users, a steady-state solution is specified automatically by *GroundWaterTutor*. This steady state solution is then used as the initial guess for the first transient stress period.

The MODFLOW 2005/MODPATH button will prompt *GroundWaterTutor* to write input files and execute MODFLOW and MODPATH. All input and output files are stored in the directory

(a)



(b)



**Figure 8.** Command Prompt windows that appear when MODFLOW and MODPATH are running. (a) MODFLOW command prompt window. (b) MODFLOW command prompt window. “...\GWtutor\gui\_ex1”. If the model executed successfully, a command prompt window will

open showing the progress of the MODFLOW simulation, followed shortly by a window showing the progress of the MODPATH simulation (shown in Figure 8). If these windows fail to appear, or return errors, it is likely that the user has saved the GWtutor repository to a directory that contains spaces, or is too long. Try moving the GWtutor repository to a directory with a shorter file path that does not include any spaces.

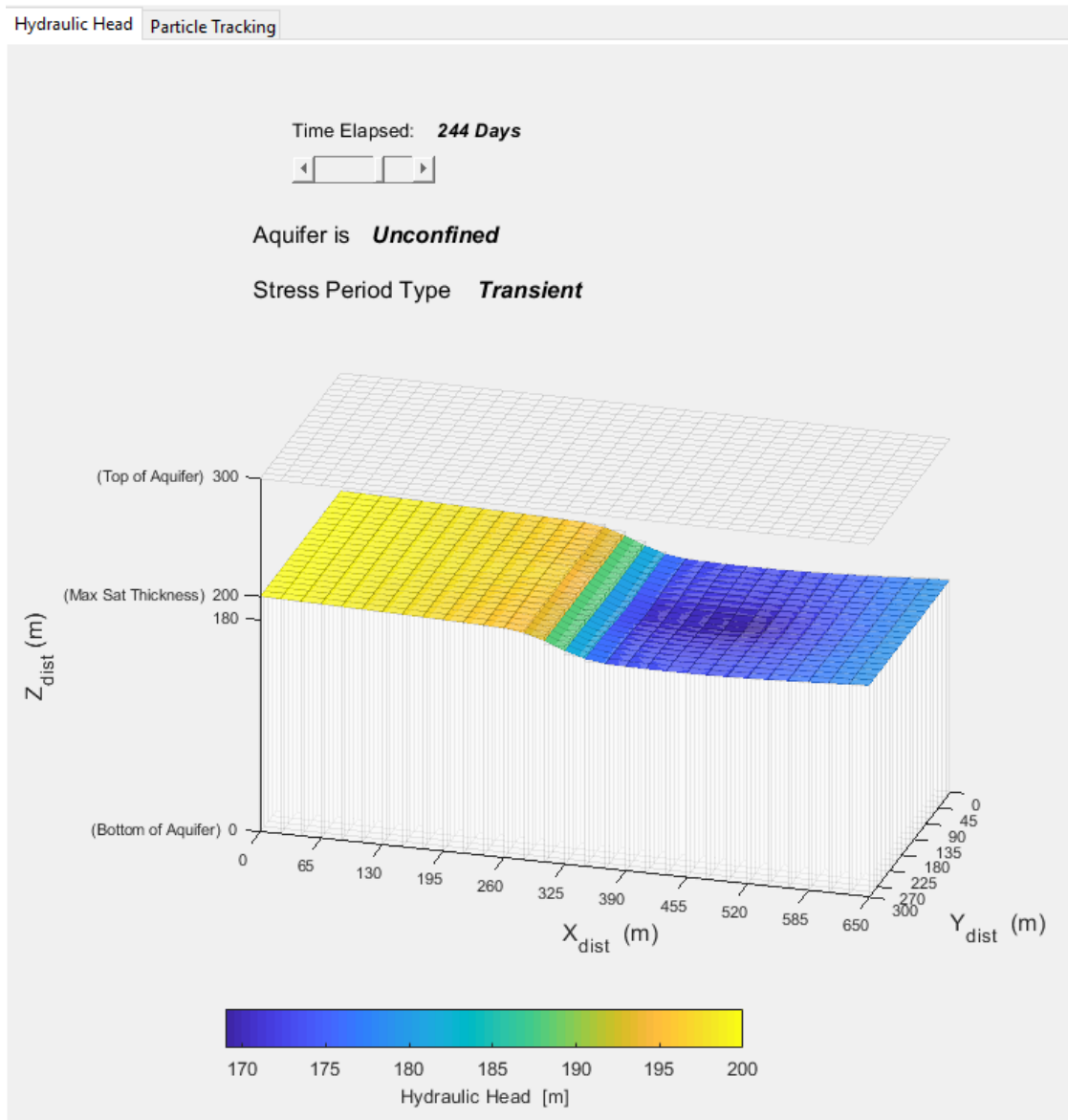
### **1.2.4 Model Results**

After MODFLOW and MODPATH have successfully terminated, a new interface window will appear alongside the main interface window detailed in Section 1.2.3. This process may take several minutes depending on the computer and time discretization specified by the user.

The new window is labeled “output” and features two tabs along the top left. To run a new model formulation (as specified in the main interface window), the “output” window will need to be closed. When trying to execute a new model formulation, *GroundWaterTutor* will prompt the user to close the “output” window with an alarm (short beep). A description of each tab is provided in sections 1.2.4a and 1.2.4b.

#### **2.2.4a Hydraulic Head**

The *Hydraulic Head* tab provides users with a visualization of the simulated hydraulic heads returned by MODFLOW (shown in Figure 9). The slider bar at the top left of the window can be used to specify the time at which hydraulic heads will be displayed. The stress period type is shown below this slider, along with an indication of whether aquifer conditions are confined or unconfined.



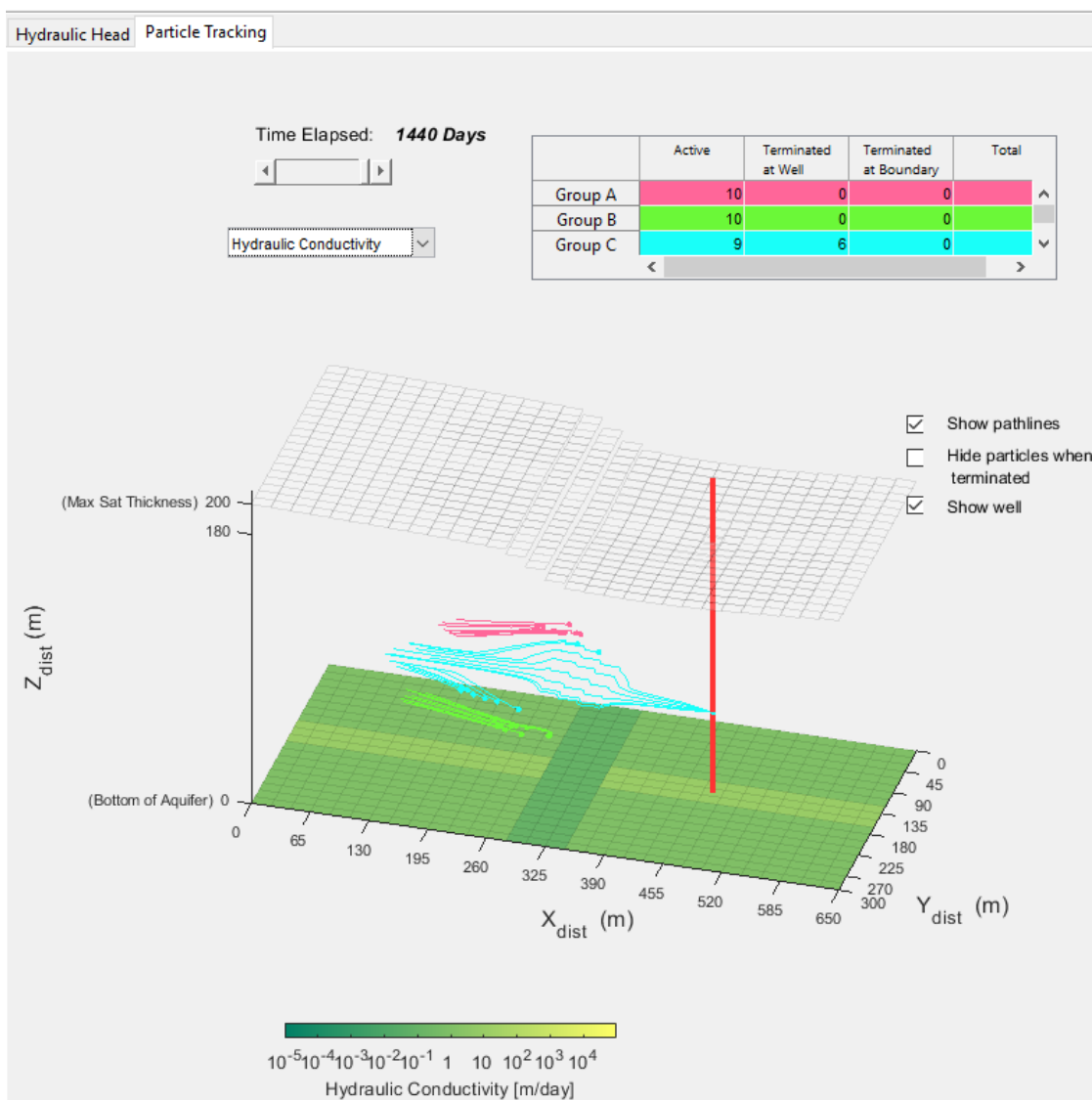
**Figure 9.** *Hydraulic Head* tab in the *GroundWaterTutor* results window.

### 2.2.4b Particle Tracking

The *Particle Tracking* tab provides users with a visualization of the time-dependent position of tracer particles (shown in Figure 10). The slider bar at the top left of the window can be used to specify the time at which particle positions are displayed. The dropdown menu below can be used to toggle whether hydraulic head or hydraulic conductivities are displayed on the bottom

surface of the model object. These features are quite handy for illustrating the effects of heterogenous hydraulic conductivities on hydraulic gradients and resulting particle trajectories.

The table located at the top right corner of the window displays the status of each cluster of particles. The “Active” column indicated how many particles from each group are still eligible to move throughout the model domain. These particles have not been terminated at a model boundary or a strong sink (e.g. pumping well). The “Terminated at Well” column indicates how



**Figure 10.** Particle Tracking tab in the *GroundWaterTutor* results window.

many particles from each group have been captured by the well. The total number of particles in each cluster remains fixed, which is indicated in the “Total” column of the table.

Three checkboxes along the right side of the window provide the ability to: (1) Display the path line taken by each particle up to the specified time, (2) Hide particle and path lines when they are terminated at the well, and to (3) Display the location of the well.

### **1.2.5 Example Problem**

In this problem, users are tasked with consulting a commercial bean farmer named Jillie. With approximately 6000 acres of farm, Jillie's crops require a cumulative total of no less than .0016 m of water per unit area, per day, over the course of 365 days to grow properly. This results in a cumulative total of 0.58 m (1") over the area of the bean crop. In the past, precipitation has been sufficient to keep her beans watered. However, in recent years, drought conditions have significantly decreased crop yields. To compensate, Jillie drilled a water well on her property, which provides water for irrigation. Unfortunately, she failed to account for the proximity of her water well to several nearby zones of contaminated groundwater. Jillie needs to know the highest rate she can pump water without capturing too much contaminated water.

Last year, only 11 inches of precipitation fell. To supply the remaining ~11 acre-inches of water, Jillie will need to extract a minimum of 75,000 cubic meters of water daily, over the course of 365 days (providing ~11 inches of water over the area farmed). Using this value as a constraint on the minimal pumping rate, your task is to use the interface to determine the maximal possible pumping rate while capturing no more than 50% of contaminants in the system.

Data from the geological survey indicates that the area receives a uniform recharge rate of 0.05 (+- 0.01) meters per day. With a uniform specific storage of  $10^{-6}$ , uniform specific yield of 0.2

(or 20%) and uniform porosity of 0.25 (or 25%). The Aquifer in this region is known to have a saturated thickness between 180 and 200 m, with a surface elevation of 300 m. The hydraulic conductivities in the area are not known precisely, however it is known that they vary within an order of magnitude of 10m/day and that the anisotropy factor is 1. It is also unknown whether hydraulic conductivities are homogenous or heterogeneous. For now, the assumption of homogenous hydraulic conductivities will be made.

### **1.2.5a Question 1**

Given the information above, is this a confined or unconfined aquifer? Is the head surface a water table or a potentiometric surface? No modeling is required to answer this question.

### **1.2.5b Question 2**

When first accessed under the Initial Condition Tab, the system shown is unconfined. Why is this obvious? Identify the water table.

### **1.2.5c Question 3**

To ensure that wells do not become too polluted by a substance being transported through the aquifer, change the hydraulic conductivity (*Parameters* tab) and the pumping rate (*Pumping and Recharge* tab) to find the highest rate the well can be pumped while capturing no more than 50% of the particles. The table in the Particle Tracking tab from the “output” window can be used to determine how many particles are captured at the well.

Consider K values ranging between 10 and 100 m/day. The user should observe that larger pumping rates can be achieved using higher K values.

#### 1.2.5d Question 4

Jillie's friend, Billy, owns a farm in a valley with fine floodplain silt deposits that create a 100 m thick aquifer. This system can be represented by lowering the elevation of the top of the aquifer (*Initial Conditions* tab). Repeat Question 2 and Question 3 for Billy's farm.

#### 1.2.5e Question 5

Yet another neighbor, Kelly, has land where there is a high K unit along the length of the valley. In her area, the aquifer is unconfined with a surface elevation of 300 m (same as Jillie's farm). Repeat Question 2 and Question 3 for Kelly's farm.

#### 1.2.6 Exercise Epilogue

Developing a basic understanding of how MODFLOW and MODPATH handle input and output data is vital to developing more complex models. This exercise is intended to introduce users to features provided by *GroundWaterTutor* and provide an example of the types of aquifer scenarios that can be explored. The features provided by *GroundWaterTutor* only scratch the surface of what is possible with MODFLOW and MODPATH. Users are highly encouraged to familiarize themselves with raw input and output files used by MODFLOW and MODPATH, located in the directory "\GWtutor\gui\_ex1". More information about the structure of these files can be obtained through the MODFLOW 2005 Online User Manual

<https://water.usgs.gov/ogw/modflow/MODFLOW-2005-Guide/>

### 1.3 References

Harbaugh, A. W. (2005). MODFLOW-2005 : The U.S. Geological Survey modular ground-water model--the ground-water flow process. Techniques and Methods.

Hu, Y., Valocchi, A. J., Lindgren, S. A., Ramos, E. A. and Byrd, R. A. (2015), Groundwater Modeling with MODFLOW as a Web Application. *Groundwater*, 53: 834-835. doi:10.1111/gwat.1272

Li, S., & Liu, Q. (2003). Interactive Groundwater (IGW): An innovative digital laboratory for groundwater education and research. *Computer Applications in Engineering Education*, 11(4), 179-202. doi:10.1002/cae.10052

Pollock, D. W. (2016). User guide for MODPATH Version 7—A particle-tracking model for MODFLOW. Open-File Report.



## **Chapter 2:**

### **Quantifying Spreading during Enhanced In-situ Remediation of Contaminated Groundwater**

#### **2.1 Introduction**

##### **2.1.1 Importance of Mixing**

Mixing is a fundamental process of importance to many disciplines, including chemical engineering, materials science, biology, geology and hydrology. This chapter considers the problem of mixing during in-situ groundwater remediation, where a treatment solution is injected into the contaminated region of an aquifer to promote degradation reaction. In-situ degradation rates fundamentally depend on the amount of mixing between treatment solution and contaminated groundwater. However, this approach suffers due to the challenges associated with mixing initially segregated solute plumes in porous media. The physics of flow through porous media result in mixing dominantly occurring in a narrow stripe along the plume interface (Dentz et al., 2011).

##### **2.1.2 Mixing, Spreading and Dilution**

Here, mixing is conceptualized as a two-step process. Spatial and temporal variations in flow velocities result in the spreading of plume boundaries, while dilution by dispersion and diffusion softens concentration gradients near plume boundaries, ultimately homogenizing the mixture (Reynolds, 1894; Ottino, 1989). Spreading is dominated by advection, and can be imagined as the stretching and folding of plume boundaries, without changing the volume occupied by the contaminant (Weeks and Sposito, 1998). Stretching reflects the deformation of an initially spherical boundary into an ellipsoid, and folding is the bending of the ellipsoid along its axis. Because flow is incompressible, stretching in one direction means equal contraction across at

least one other direction (Subramanian et al., 2009). As time progresses, stretching (and contraction) elongates and thins plume boundaries, sharpening local concentration gradients and increasing the effective surface where dilution processes can occur (de Anna et al., 2014; Le Borgne et al., 2013; Dentz et al., 2011; Rolle and Chiogna et al., 2009). Folding serves to confine elongated boundaries within a finite region so that stretching may continue (Aref, 2002).

### **2.1.3 Objective**

The goal of this work is to investigate several measures of spreading, one of which is new to the groundwater community, in terms of the information they might provide toward the design of improved active in-situ remediation systems. The objective is to assess whether these measures capture the underlying system dynamics in a way that provides improved insight or computational efficiency than previous approaches. Numerical simulations are used to compare spreading to contaminant degradation resulting from a heuristically determined active treatment system introduced in Mays and Neupauer (2012) and three variations on this system introduced in Piscopo et al (2016).

The first spreading measure under evaluation is the length stretch, which records overall spreading as the elongation of a line of tracer particles as they evolve with the flow. The length stretch was used by Neupauer et al (2014) to measure the spatial distribution of spreading produced by the heuristically determined injection-extraction sequence. The second measure is the principal stretch, a commonly used measure of deformation-strain in continuum mechanics. The principal approximates the deformed state of an initially circular cluster of tracers using a linear combination of basis vectors. The third measure can be referred to as affine/non-affine deformation, and provides explicit quantification of both stretching and folding. Stretching

reflects the ability of a linear model to describe the evolved state of an initially circular cluster. Folding is measured as a function of the linear models residual fit.

## **2.2 Background**

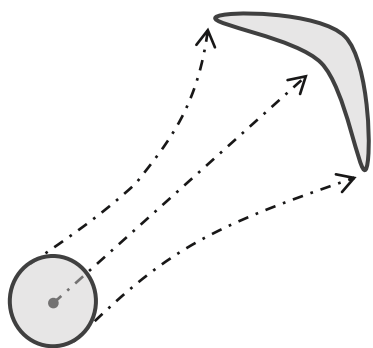
### **2.2.1 Mixing in Aquifers**

Mixing in aquifers is challenging because flow is predominantly laminar, impeding formation of turbulent eddies associated with good mixing in open flows (Dagan, 1989 §4.3.5; Siegrist et al., 2011). The simplest strategies for mixing during in-situ remediation rely on interaction between aquifer heterogeneity and groundwater flow. Flow fields may be passive, produced by background flow, or actively enhanced by pumping (Cirpka and Rolle et al., 2012; Kapoor and Kitanidis, 1998; Weeks and Sposito 1998). Passive strategies suffer because plume spreading may not occur on a practical timescale or may result in contaminants migrating to undesired locations. In contrast, active strategies use pumping wells to generate spatial and temporal fluctuations in flow velocity, promoting plume spreading and contaminant degradation. Active strategies offer the possibility of enhanced plume spreading by chaotic advection, as well as kinematic confinement of contaminants, treatment solution, and degradation byproduct within a limited region (Aref, 2017; Trefry and Lester et al., 2012).

### **2.2.2 Spreading by Chaotic Advection**

Plume spreading can be enhanced by chaotic advection (or deterministic chaos). Chaotic advection is characterized by fluid particle trajectories demonstrating sensitive dependence to an initial condition. Such flows possess distinct regions (in time and space) where the trajectories of initially nearby fluid particles diverge exponentially (Aref, 1984; Liu et al., 1994; Sposito, 2005). Plume boundaries aligned with these regions are stretched and thinned by divergent flow

paths, producing good spreading. A highly idealized illustration of this process is provided in Figure 11.



**Figure 11.** Idealized depiction of plume spreading by chaotic advection. An initially circular plume is elongated (stretched) and bent along its axis (folded) by divergent flow paths.

### 2.2.3 Engineered Injection and Extraction

Building on a long line of study into chaotic advection, mixing, and reaction in porous media (Aref, 1984; Jones and Aref, 1988; Ottino, 1990; Weeks and Sposito, 1998; Sposito, 2006; Bagtzoglou and Oates, 2007; Zhang et al., 2009; Lester and Metcalfe et al., 2009; Rolle et al., 2009; Metcalfe et al., 2010), the work of Mays and Neupauer (2012) introduced Engineered Injection and Extraction (EIE), a novel framework for the design of active spreading systems. These systems use sequences of injection and extraction of clean water from an array of pumping wells to stretch and fold the plume interface between treatment solution and contaminated groundwater, as well as keeping plumes confined within the prescribed remediation area or volume.

Mays and Neupauer (2012) proposed a heuristically determined injection-extraction scheme comprising four wells operating in a twelve step sequence, which is examined in this paper. They used analytical models of purely advective transport to investigate how the plume interface was

stretched and folded during remediation in a 2-D confined, homogenous, isotropic aquifer. They demonstrated that in heterogeneous systems, flow resulting from the injection-extraction scheme produced chaotic advection and belonged to a class of laminar flows known to be optimal for spreading and mixing (Ottino et al., 1994). Subsequent studies (Piscopo et al., 2013, Neupauer et al., 2014) incorporated dispersion and reaction into the system, using numerical models to systematically quantify the effects of aquifer heterogeneity, well spacing, and pumping rates on spreading and contaminant degradation. Under the assumption of instantaneous and irreversible reaction, Piscopo et al (2013) found that contaminant degradation was six to seven times greater with the injection-extraction scheme than with passive remediation, in heterogeneous and homogenous settings. Neupauer et al (2014) showed that introducing mild heterogeneity into the system led to substantial increases in measures of spreading and contaminant degradation, however increases in spreading measures were proportionally greater than increases in contaminant degradation. Additionally, they showed that increased spreading (by aquifer heterogeneity, up-scaled injection-extraction magnitudes or closer well spacing) corresponds to greater probability of capturing treatment solution in an extraction well, which raises practical and regulatory concerns (Mays and Neupauer, 2013). Piscopo et al (2016) posed an injection-extraction scheme within an optimization framework, arriving at several solutions that produce improved contaminant degradation (e.g. Mays and Neupauer, 2012; Piscopo et al., 2013; Neupauer et al., 2014). In their problem formulation, decision variables included the active well and the injection or extraction rate at each of the twelve steps in the sequence. Objective variables included functions of reactant concentration and energy consumption (i.e. injection/extraction rates). Solutions were constrained by thresholds on injection/extraction rates and confinement of the plume interface.

### **2.3 Motivation**

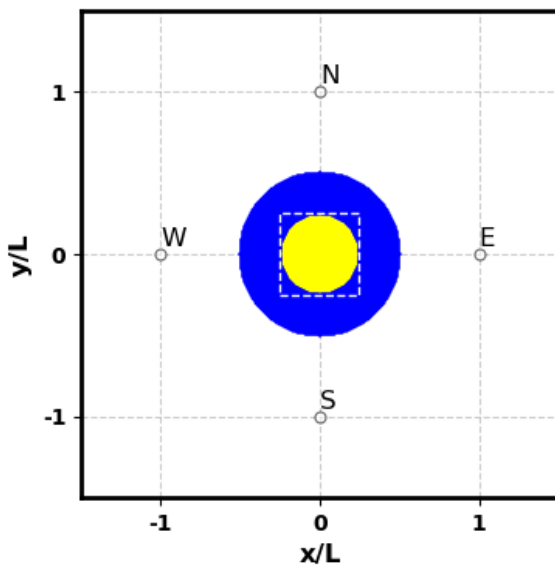
While early investigations into EIE placed emphasis on quantifying spreading properties of the flow (e.g., Mays and Neupauer, 2012; Neupauer et al., 2014), measures of spreading were not directly incorporated into the optimization problem formulation posed by Piscopo et al (2015). Computationally efficient, yet informative metrics are important in optimization problem formulations, which often require hundreds to thousands of iterations to arrive at an optimal solution. As such, reducing the computational effort associated with the optimal design of active spreading systems necessitates use of objective metrics which capture relevant system dynamics in a computationally frugal manner. Ultimately, this reduction in computational effort allows a broader set of candidate system designs to be evaluated during optimization.

Quantifying contaminant degradation can be computationally expensive, requiring advective, dispersive and reactive transport to be simulated. Conversely, spreading can be quantified with less computational effort, requiring only advective transport to be simulated. The observation that substantial rearrangements in plume geometry correspond to significant reaction has been made in previous studies on EIE (e.g. Piscopo et al., 2013; Neupauer et al., 2014), and is the primary motivation for our investigation into the utility of measures of spreading as a computationally efficient proxy for predicted contaminant degradation. Establishing a relationship between spreading and contaminant degradation produced by a given active treatment system design would provide grounds for their use in future optimization efforts.

## 2.4 Modeling Methods

### 2.4.1 Model Aquifer and Engineered Injection Extraction System

In the model aquifer used in this work, four fully penetrating wells are located  $L = 25$  m from the origin, at  $(x, y) = (0, 0)$ , inscribing the contaminated region of a two-dimensional, homogenous, horizontally confined aquifer. The square model domain extends  $> 6L$  from the origin in each cardinal direction. There is no ambient flow in the system, enforced by constant head boundaries of equal value along the east and west sides and no-flow boundaries on the north and south sides. A plume of contaminated groundwater with an outer radius of  $12.5$  m ( $0.5L$ ) encloses a circular



**Figure 12.** Map view of the inner part of the model aquifer in its initial configuration. Treatment solution particles are shown in yellow and contaminant particles in blue. Well locations are indicated by open circles and labeled by cardinal direction. The dashed white line represents the plot region in other figures. The full system is set within a larger system, extends a distance of  $6L$  from the origin in all directions, identified by the normalized axis values  $x/L$  and  $y/L$ .

plume treatment solution with radius  $6.25$  m ( $0.25L$ ), centered about the origin (shown in Figure 12).

During EIE, wells are operated in pre-determined 12 step sequences. Each step consists of one well injecting or extracting clean water for a duration of  $\Delta t = 6.25$  days. Pumping sequences have zero net flux into the aquifer after the final step.

### 2.4.2 Groundwater Flow and Transport

To maintain consistency with previous investigations into EIE, the approach of Piscopo et al (2013) is used to model groundwater flow and transport. The advection-dispersion-reaction equation governs the transport of reacting species, given by (Bear 1979)

$$\frac{\partial C_i}{\partial t} = \nabla \cdot \mathbf{D} \nabla C_i - \nabla \cdot (\mathbf{v} C_i) - R \quad (2.1)$$

Where  $C_i$  is denotes the concentration of the  $i^{\text{th}}$  species ( $i = 1$  for treatment,  $i = 2$  for contaminant and  $i = 3$  for degradation product),  $R$  is the reaction rate,  $t$  is time and  $\mathbf{D}$  is the dispersion tensor.

The velocity from (2.1) is determined from Darcy's law, given by

$$\mathbf{v} = -\frac{1}{\theta} \mathbf{K} \nabla h \quad (2.2)$$

where  $\theta$  is porosity,  $\mathbf{K}$  is the hydraulic conductivity tensor. The hydraulic head is denoted  $h$ , and can be determined by solutions to the groundwater flow equation. For a two dimensional, horizontally confined aquifer, this is given by

$$S_s \frac{\partial h}{\partial t} = \nabla \cdot \mathbf{K} \nabla h + Q_j \quad (2.3)$$

with  $S_s$  as the specific storage and  $Q_j$  as the rate of injection from the active well during the  $j^{\text{th}}$  step of the EIE sequence. The USGS programs MODFLOW (Harbaugh, 2005) and MODPATH (Pollock, 2016) are used in conjunction with the python package FloPy (Bakker et al., 2015) to simulate the advective transport components of (2.1).

Dispersive transport in (2.1) is modeled by a random walk method (Salamon et al., 2006; Uffink, 1989). Random displacements are superimposed to particle positions after each advection step.



Normally distributed random displacements are added in the direction of the local velocity vector, and a direction perpendicular to the local velocity vector, with a mean of zero and respective variances of  $2\alpha_L|\mathbf{v}|\Delta t$  and  $2\alpha_T|\mathbf{v}|\Delta t$ . The USGS python package FloPy was used to develop an algorithm which translates simulated groundwater flows and particle trajectories from MODFLOW and MODPATH, as well as superimpose displacements due to dispersion.

The Reaction component of (2.1) is modeled as an instantaneous, irreversible reaction with a 1:1 stoichiometric ratio, written as



where  $C_1$  is the treatment concentration,  $C_2$  is contaminant concentration and  $C_3$  is an inert reaction byproduct.

Following Piscopo et al (2013) and Neupauer et al (2016), treatment solution and contaminated groundwater are simulated as a collection of conservative tracer particles, initially spaced 0.25 m ( $L$ ) on a uniform grid. Particles are initially assigned a label and mass corresponding to  $C_1$  or  $C_2$ . Each treatment particle is assigned an initial mass of 16 mg and each contaminant particle is assigned an initial mass of 4 mg. A total number of 1961 treatment particles and 5884 contaminant particles are initialized. Thus, the initial concentration of treatment solution is  $25.57 \frac{mg}{m^3}$  and the initial concentration of contaminant is  $6.39 \frac{mg}{m^3}$ .

After each advection step, all particles are spatially grouped into  $0.625 \text{ m} \times 0.625 \text{ m}$  ( $0.025L \times 0.025L$ ) bins. Reaction with a 1:1 stoichiometric ratio is simulated within each bin as follows. The total mass carried by  $C_1$  and  $C_2$  particles is computed and the limiting reactant is determined. The mass of the limiting reactant is completely subtracted from particles of the

limiting reactant. The mass of the excess reactant is reduced by the same amount. To enforce the assumption of a 1:1 stoichiometric ratio, mass is distributed evenly among all remaining treatment or contaminant particles. Mass is conserved, meaning that all reacted mass is converted into reaction product. All particles of the limiting reactant are re-labeled as  $C_3$ , and the total mass reacted (twice mass of limiting reactant) is distributed evenly among them. Modeling parameters (shown in Table 1) are the same as those used by Piscopo et al (2013). An algorithm was developed in python to simulate reaction based on the time-dependent positions of particles based on the advection-dispersion simulations above.

As noted previously by authors (e.g., Piscopo et al., 2013), the coarse time discretization used in this approach tends to underestimate the amount of contaminant degradation. Thus the calculated values of contaminant degradation will be uniformly lower than actual values across all considered EIE sequences.

<b>Table 1.</b> Modeling parameters. Reproduced from Neupauer et al (2016)	
Parameter	Value
Distance from origin to well, $L$	25 m
Storage Coefficient, $S$	$1 \times 10^{-5}$
Mean hydraulic conductivity, $K$	$0.5 \frac{\text{m}}{\text{d}}$
Aquifer top elevation	0 m
Aquifer bottom elevation	-10 m
Aquifer thickness, $b$	10 m
Porosity, $\theta$	0.25
Duration of EIE step	6.25 day
Dimensions of model domain	$-150.125 \text{ m} \leq x \leq 150.125 \text{ m}$ $-150.125 \text{ m} \leq y \leq 150.125 \text{ m}$
Finite difference grid spacing	$0.25 \text{ m} \times 0.25 \text{ m}$
Flow boundary conditions	$h = 10 \text{ m}$ on east and west
Longitudinal dispersivity, $\alpha_L$	0.05 m
Transverse dispersivity, $\alpha_T$	0.005 m
Number of treatment solution particles	1961
Number of contaminant particles	5884
Initial mass per treatment solution particle	4 mg
Initial mass per contaminant particle	1 mg
Ambient flow velocity	0

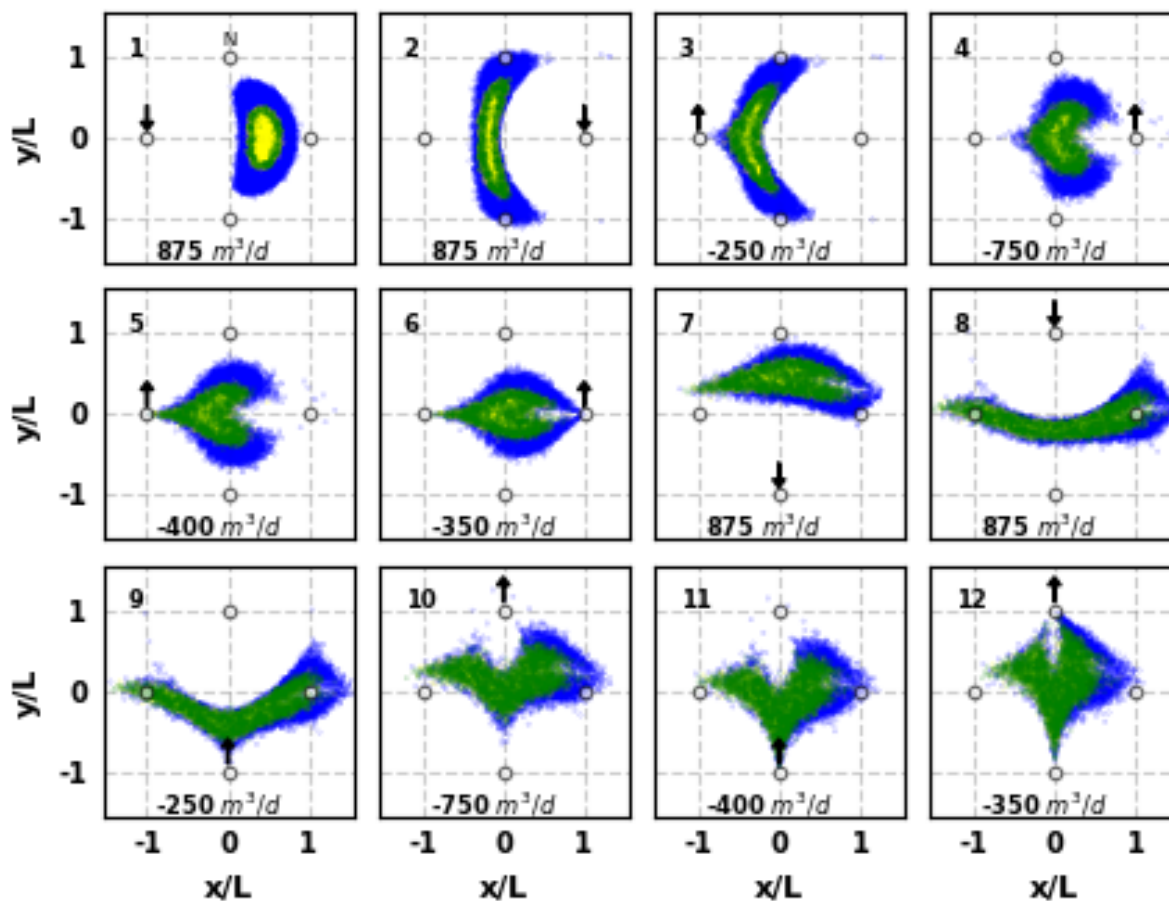
## 2.5 Contaminant Degradation Simulations

In this section simulation results for contaminant degradation produced by three different EIE sequences are replicated from the works of Piscopo et al (2013) and Piscopo et al (2016).

Simulated contaminant degradation will serve as a benchmark to evaluate the performance of spreading measures.

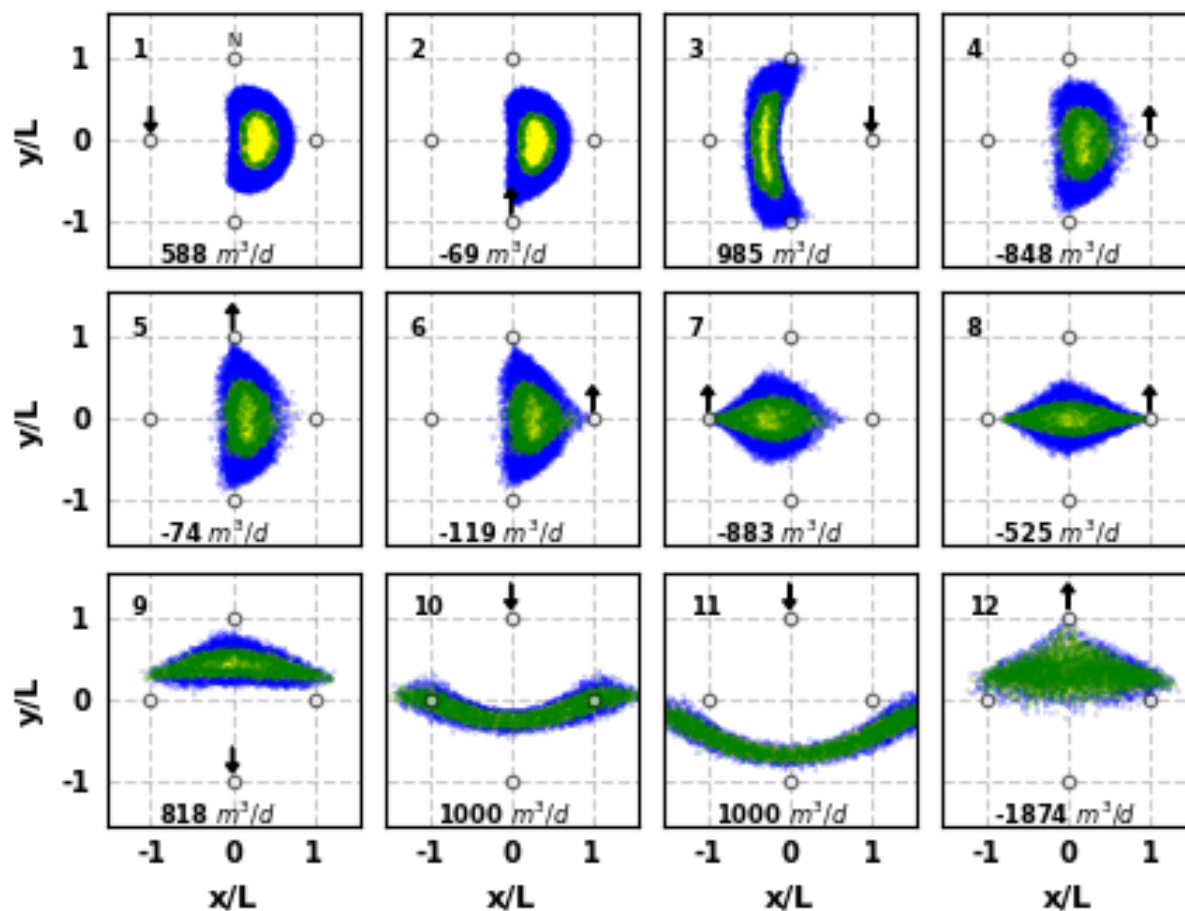
All sequences are simulated in the model aquifer described in Section 2.4.1, over a 75 day time period with  $\Delta t = 6.25$  days. The first EIE sequence (*Case A*) was introduced by Mays and Neupauer (2012), and used in several subsequent studies (e.g., Piscopo et al., 2013; Neupauer et al., 2014; Piscopo et al., 2016). The second and third sequences (*Case B* and *Case C*) are representative optimal solutions from the work of Piscopo et al (2016). The simulated contaminant degradation resulting from the sequences (*Cases A, B* and *C*) reported in Piscopo et al (2016) vary by  $\leq 10\%$  after remediation. To examine whether spreading measures can predict contaminant on a gross scale, *Cases A2, B2* and *C2* are introduced, which are identical to *Cases A, B* and *C*, respectively, except a 50% reduction in pumping magnitude is imposed at each step of the sequence. In these simulations, tracer particles are subject to advection, dispersion, and reaction. The dispersion component is simulated at the beginning of each step in the EIE sequence, and reaction is simulated at the end of each step. Figures 13-15 show the position of treatment solution, contaminant and reaction product particles after each step of *Cases A, B* and *C*, respectively. Results for *Cases A, B* and *C* are shown in Figures A1-A3 of Appendix A.

### Case A



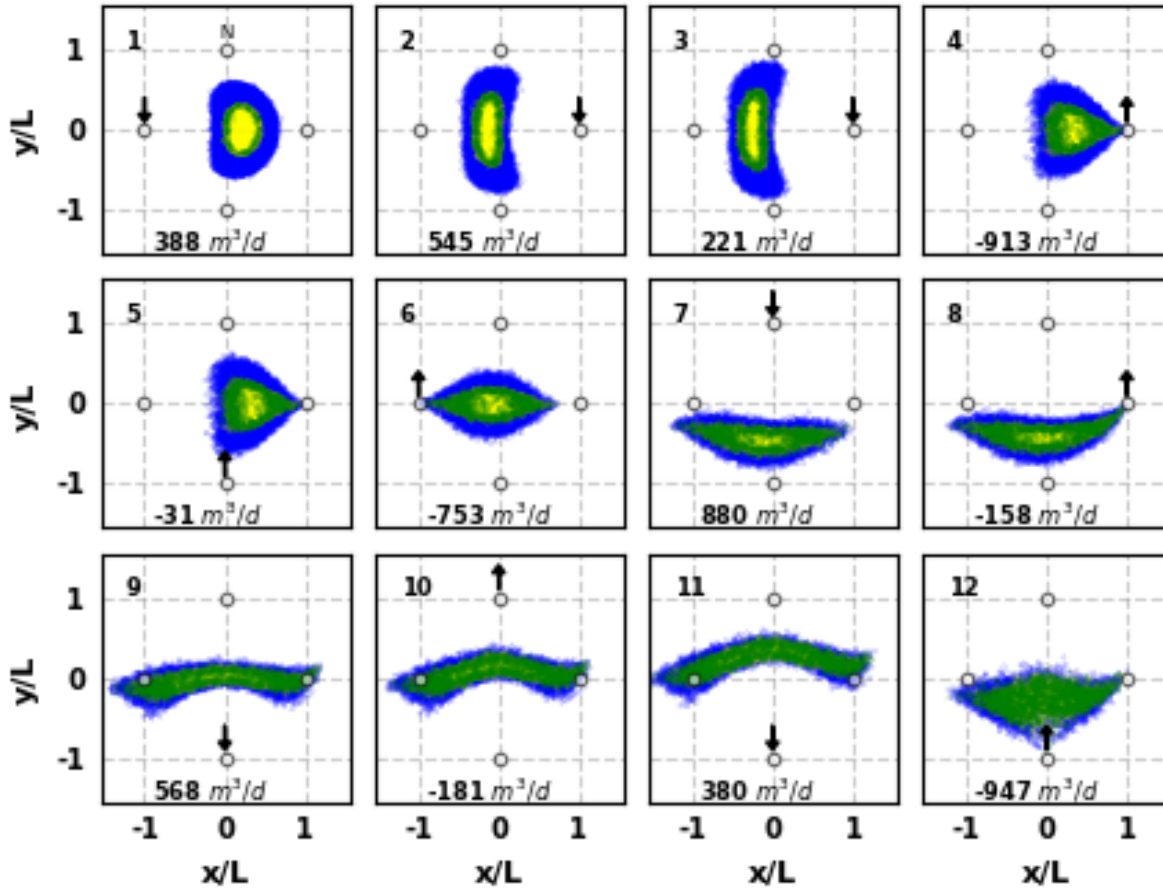
**Figure 13.** Particle positions and species based on advective-dispersive and reactive transport simulations. Locations of treatment solution (yellow), contaminant (blue) and reaction product (green) particles after each step of the *Case A* EIE sequence. The number in the top left of each pane indicates the step in the EIE sequence. The active well during each step is signified by a small black arrow. Downward pointing arrows represent injection and upward pointing arrows represent extraction. Pumping rates for the active well are listed in the bottom of each pane.

### Case B

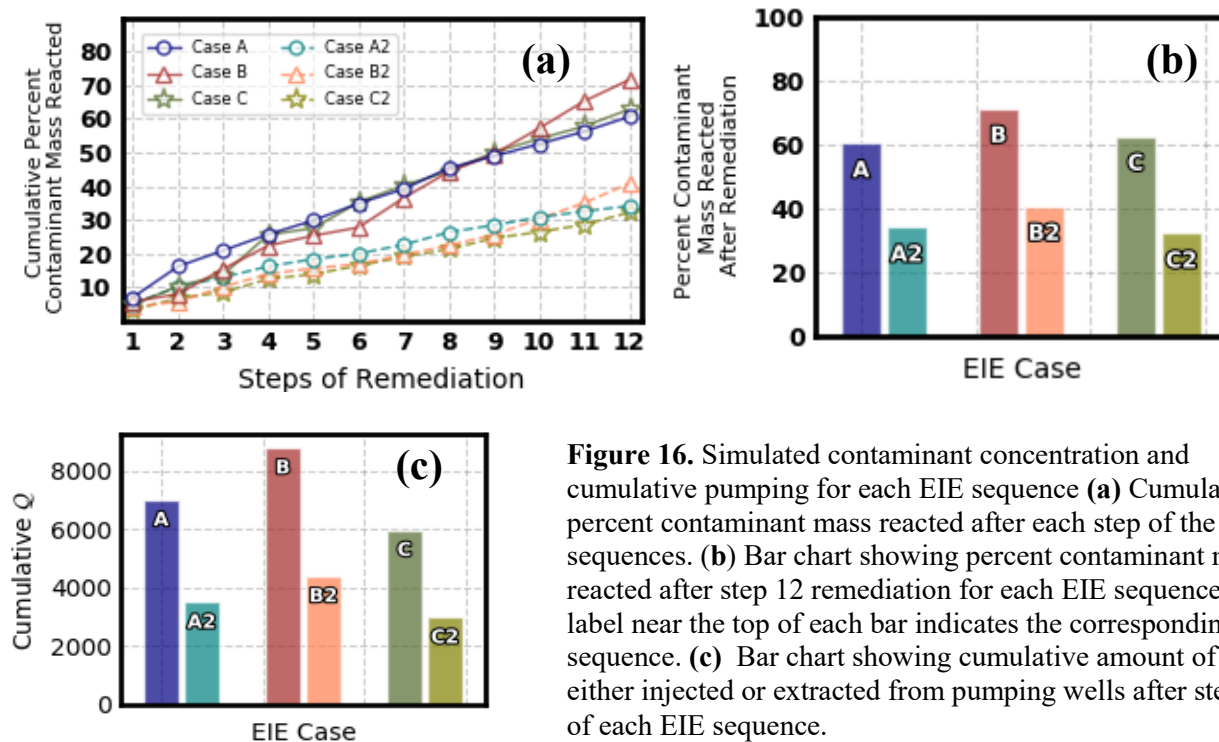


**Figure 14.** Particle positions and species based on advective-dispersive and reactive transport simulations. Locations of treatment solution (yellow), contaminant (blue) and reaction product (green) particles after each step of the *Case B* EIE sequence. The number in the top left of each pane indicates the step in the EIE sequence. The active well during each step is signified by a small black arrow. Downward pointing arrows represent injection and upward pointing arrows represent extraction. Pumping rates for the active well are listed in the bottom of each pane.

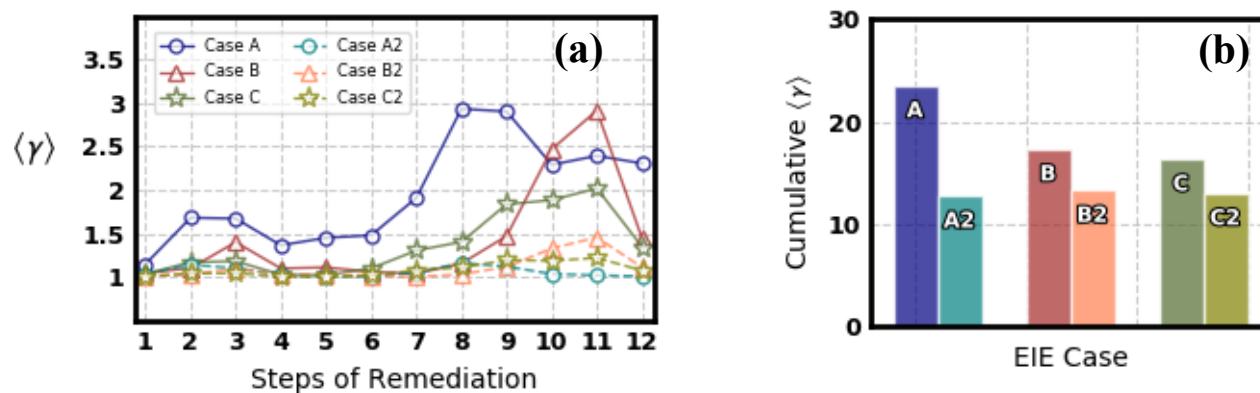
### Case C



**Figure 15.** Particle positions and species based on advective-dispersive and reactive transport simulations. Locations of treatment solution (yellow), contaminant (blue) and reaction product (green) particles after each step of the *Case C* EIE sequence. The number in the top left of each pane indicates the step in the EIE sequence. The active well during each step is signified by a small black arrow. Downward pointing arrows represent injection and upward pointing arrows represent extraction. Pumping rates for the active well are listed in the bottom of each pane.



**Figure 16.** Simulated contaminant concentration and cumulative pumping for each EIE sequence (a) Cumulative percent contaminant mass reacted after each step of the EIE sequences. (b) Bar chart showing percent contaminant mass reacted after step 12 remediation for each EIE sequence. The label near the top of each bar indicates the corresponding EIE sequence. (c) Bar chart showing cumulative amount of water either injected or extracted from pumping wells after step 12 of each EIE sequence.



**Figure 17.** Spreading metric based on advective transport simulation (a) Spatially averaged values of the length stretch  $\langle \gamma \rangle$  after each step of the EIE sequences. (b) Bar chart showing cumulative  $\langle \gamma \rangle$  after step 12 of each EIE sequence.

For each EIE case, the cumulative percent contaminant degradation is computed after every step of the sequence as

$$100 \times \frac{m_2^0 - m_2^t}{m_2^0} \quad (2.5)$$

where  $m_2^0$  is the total initial mass of contaminant and  $m_2^t$  is the total mass of contaminant after time  $t$ . Results are shown in Figure 6. Subplot (a) shows the cumulative percent contaminant mass degraded produced by each EIE sequence, plotted against time. Subplot (b) shows a bar chart of the cumulative percent contaminant degradation for each sequence after remediation (i.e., step 12). Subplot (c) shows the cumulative amount of clean water either injected or extracted from pumping wells after step 12 of each EIE sequences. The optimized sequences (*Cases B* and *C*) produce more slightly more contaminant degradation after remediation than the heuristically determined sequence (*Case A*). The most degradation is produced by *Case B* (72%), followed by *Case C* (63 %) then *Case A* (61%). The cases with 50% pumping magnitude (*A2-C2*) produced less reaction than *Cases A-C*, but by less than 50%. *Case A2* degraded 34% of the initial contaminant mass, *Case B2* degraded 41% and *Case C2* degraded 32%. Figure 17 shows results from Section 2.7.2, and is included for ease of comparison to Figure 16.

For reaction to occur, contaminant and treatment particles need to be in close proximity. Because the treatment and contaminant particles are initially nearby (as shown in Figure 12), significant reaction occurs during the first step of each EIE sequence. In subsequent steps, significant reaction occurs when particle arrangements are substantially reconfigured. For example, during steps 6 and 7 of *Case B*, the eastern and western extremities of the plumes are stretched, while the northern and southern extremities are compressed toward the center (shown in Figure 14),



leading to significant reaction. Similar behavior can be observed in steps 4 and 6 of *Case C* (Figure 15), as well as steps 2, 7 and 8 of *Case A* (Figure 13).

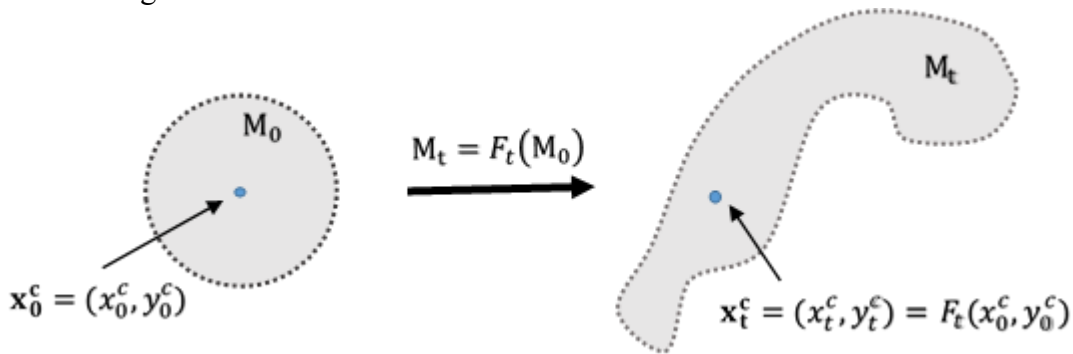
## 2.6 Mathematical Methods for Quantifying Spreading

This section presents a brief mathematical overview of the spreading metrics under evaluation.

To characterize spreading, the rearrangement of tracer particles under purely advective transport is measured. The flow map  $F$  represents the advective component of transport in (2.1),

$$\mathbf{x}_t = F_t(\mathbf{x}_0) \quad (2.6)$$

taking a particle from its initial position  $\mathbf{x}_0 = (x_0, y_0)$  of a particle to its position  $\mathbf{x}_t = (x_t, y_t)$  after time  $t$  has elapsed. Individual clusters of particles are parameterized as sets  $M$  containing the trajectories of  $n$  neighboring particles. The initial configuration of  $M$  is denoted  $M_0$ . In the initial configuration, particles are arranged uniformly in a circle of radius  $r$ , centered about the point  $\mathbf{x}_0^c = (x_0^c, y_0^c)$ . The configuration at time  $t$  is denoted  $M_t$ , given by the mapping of each element in  $M_0$  via (2.6). An illustration of the parameterization of a single particle cluster is provided in Figure 18.



**Figure 18.** Idealized depiction of the parameterization of a single particle cluster, and its evolution under the flow map. The left side of the figure shows the initial orientation of the cluster, labeled  $M_0$ . The gray dotted circle represents the initial location of the  $n$  outer particles of the cluster, which inscribe the central particle represented as the blue dot and labeled  $\mathbf{x}_0^c$ . The right side of the figure shows the orientation of the cluster after time  $t$  has elapsed, labeled  $M_t$ . The gray dotted line depicts the location of the  $n$  outer particles determined by the flow map  $F$ . The location of the central particle is shown as a blue dot and labeled  $\mathbf{x}_t^c$ ,

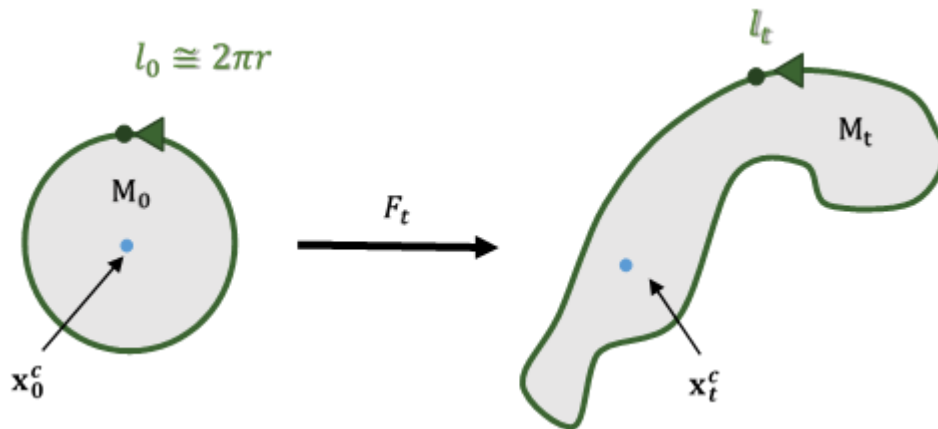
### 2.6.1 Length Stretch

The length stretch is a commonly used measure of topological entropy, chaos and deformation strain in the analysis of dynamical systems (*Ottino, 2004; Boyland et al., 2003; Thiffeault, 2004*).

Defined as the ratio of the length of a deformed line segment to the length of the corresponding un-deformed segment (*Ottino, 1989*), the length stretch  $\gamma$  is given by

$$\gamma(\mathbf{x}_0^c, t) = \frac{l_t}{l_0} \quad (2.7)$$

where  $l_0$  is the length of the line connecting the positions of particles in  $M_0$  and  $l_t$  is the length of the line connecting positions in  $M_t$ . An illustration of the geometric interpretation and procedure for computing the local length stretch is provided in Figure 19.



**Figure 19.** Idealized depiction of the procedure for computing the local length stretch for a single particle cluster. The left side of the figure shows the initial orientation of the cluster, labeled  $M_0$ . The initial perimeter length is labeled  $l_0$  which reflects the length of the green line connecting initial locations of the  $n$  outer particles of the cluster. The right side of the figure shows the orientation of the cluster after time  $t$  has elapsed, labeled  $M_t$ . The perimeter length at time  $t$  is labeled  $l_t$  which reflects the length of the green line connecting locations of the  $n$  outer particles of the cluster. The initial and final locations of the central trajectory of the cluster are labeled and shown as blue dots, but are not used in the computation of the local length stretch.

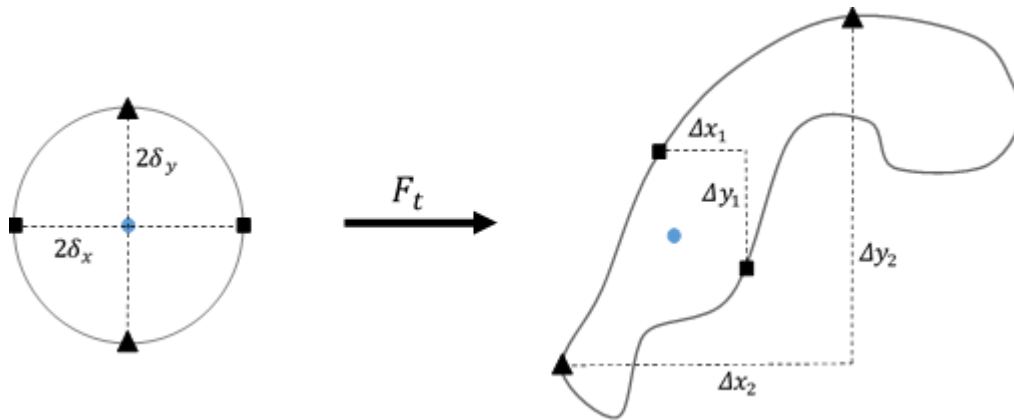
## 2.6.2 Principal Stretch

The principal stretch is another commonly used measure of deformation strain in continuum mechanics, describing local stretching with a linear combination of basis vectors (Haller, 2015).

These are determined by the singular values and right singular vectors of the deformation gradient  $\nabla F$  (of the map  $F$ ). The deformation gradient can be computed using a central differencing approach (e.g. Haller and Yuan, 2000; Haller, 2001) as

$$\nabla F(\mathbf{x}_0^c, t) \approx \begin{bmatrix} \frac{F_t(\mathbf{x}_0^c(1) + \delta_1) - F_t(\mathbf{x}_0^c(1) - \delta_1)}{|2\delta_1|} & \frac{F_t(\mathbf{x}_0^c(1) + \delta_2) - F_t(\mathbf{x}_0^c(1) - \delta_2)}{|2\delta_2|} \\ \frac{F_t(\mathbf{x}_0^c(2) + \delta_1) - F_t(\mathbf{x}_0^c(2) - \delta_1)}{|2\delta_1|} & \frac{F_t(\mathbf{x}_0^c(2) + \delta_2) - F_t(\mathbf{x}_0^c(2) - \delta_2)}{|2\delta_2|} \end{bmatrix} \quad (2.8)$$

where  $\delta_k$  is a small vector denoting perturbation pointing in the  $\mathbf{x}(k)$  spatial direction. Here,  $k=1$  corresponds to the x-direction and  $k=2$  corresponds to the y-direction. For example,  $\mathbf{x}_0^c(1) = x_0^c$  and  $\mathbf{x}_0^c(2) = y_0^c$ . An illustration of the procedure for computing the local deformation gradient is provided in Figure 20.

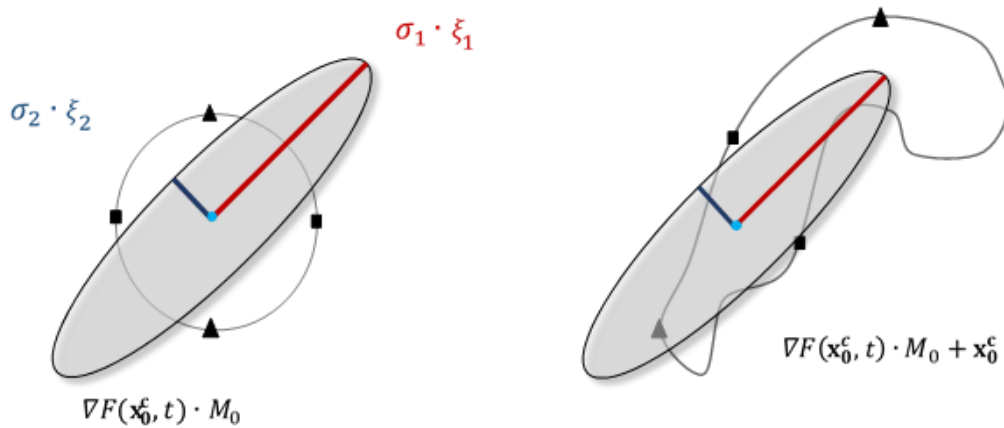


**Figure 20.** Idealized depiction of the procedure for computing the deformation gradient for a single particle cluster. The left side of the figure shows the initial orientation of the cluster, labeled  $M_0$ . Here, only 4 outer particles are used instead of the  $n$  out particles used in the local length stretch and the affine/non-affine deformation. Particles initially separated by the vectors  $2\delta_x$  and  $2\delta_y$  are labeled as black squares and triangles, respectively.

The principal stretch is defined as the eigenvalues and eigenvectors of  $\nabla F(\mathbf{x}_0^c, t)^T \nabla F(\mathbf{x}_0^c, t)$ , which can be obtained using the singular value decomposition of  $\nabla F(\mathbf{x}_0^c, t)$ ,

$$\nabla F(\mathbf{x}_0^c, t) = U\Sigma V := [\overline{u}_1 \quad \overline{u}_2] \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} [\overline{\xi}_1 \quad \overline{\xi}_2] \quad (2.9)$$

The singular values  $\sigma_k$  and corresponding right singular vectors  $\xi_k$  of  $\nabla F(\mathbf{x}_0^c, t)$  form a basis for the principal axes of an ellipse, which represents the linearized deformation (i.e., stretching) experienced by particles originating near  $\mathbf{x}_0^c$ . The principal stretch  $\sigma_1$  is defined as the leading singular value of  $\nabla F(\mathbf{x}_0^c)$ , which describes the magnitude of stretching in the direction  $\xi_1$ , which particles originating nearby  $\mathbf{x}_0^c$  were elongated the most. A geometric interpretation of the principal stretch is provided in Figure 21.



**Figure 21.** Geometric interpretation of the principal stretch. The left side of the figure shows the ellipse (shown in black) formed by projection of the deformation gradient onto  $M_0$ , written as the dot product  $\nabla F(\mathbf{x}_0^c, t) \cdot M_0$ . Although  $M_0$  (shown on left the gray circle underneath the ellipse) is not used in the computation of the principal stretch, this projection approximates the linearized spreading (i.e. stretching) experienced by particles originating nearby  $\mathbf{x}_0^c$  (shown as blue dot on left). The length and direction of major axis of the ellipse is shown in red, given by the scalar product  $\sigma_1 \cdot \xi_1$ . The minor axis of the ellipse is shown in blue, given by the scalar product  $\sigma_2 \cdot \xi_2$ . The right side of the figure shows the ellipse (shown as black line) superimposed over  $M_t$ , written as  $\nabla F(\mathbf{x}_0^c, t) \cdot M_0 + \mathbf{x}_0^c$ .

### 2.6.3 Affine/Non-affine deformation

The distinction between the processes of stretching and folding and their relative contribution to the overall spreading of particles can be determined by classifying the deformation as either linear, affine (primarily stretching) or non-affine (primarily folding). Following the approaches of Kelley and Ouellette (2011), and others (e.g., Faulk and Langer, 1998; Chiogna and Cirpka et al., 2015), the affine (primarily stretching) deformation experienced by  $M_0$  is identified by fitting a least-squares linear model  $\mathbf{d}_{\text{aff}}(\mathbf{x}_0^c, t)$  to the relative displacements of particles in  $M_t$  with respect to their central trajectory. The  $2 \times n$  relative displacement vector  $\mathbf{d}(\mathbf{x}_0^c, t)$  is determined as the distance between the particles in  $M_t$  and their central trajectory  $\mathbf{x}_t^c$  along each coordinate direction. The  $N^{\text{th}}$  element of  $\mathbf{d}(\mathbf{x}_0^c, t)$  is computed as

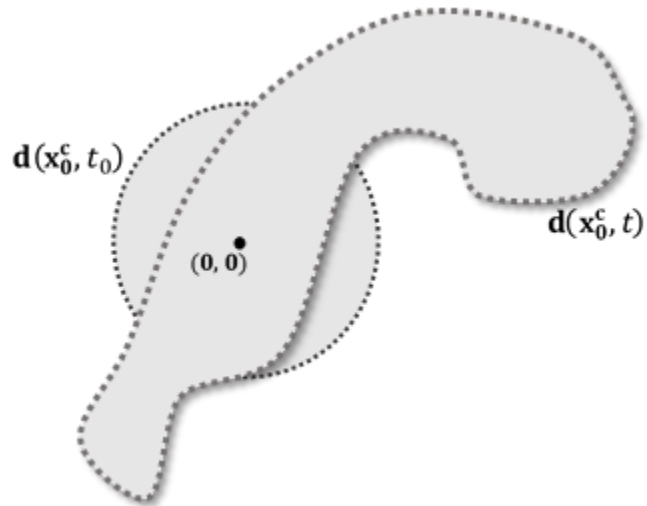
$$\mathbf{d}^{(N)}(\mathbf{x}_0^c, t) = \mathbf{x}_t^c - M_t^{(N)}; \quad N = 1, \dots, n. \quad (2.10)$$

For example, the  $N_{\text{th}}$  element of the displacement vector at  $t = t_0$  would be written as

$$\mathbf{d}^{(N)}(\mathbf{x}_0^c, t_0) = \mathbf{x}_0^c - M_0^{(N)} = r_0; \quad N = 1, \dots, n. \quad (2.11)$$

and would always equal to  $r_0$ . An illustration depicting the geometric interpretation of the displacement vector is provided in Figure 22

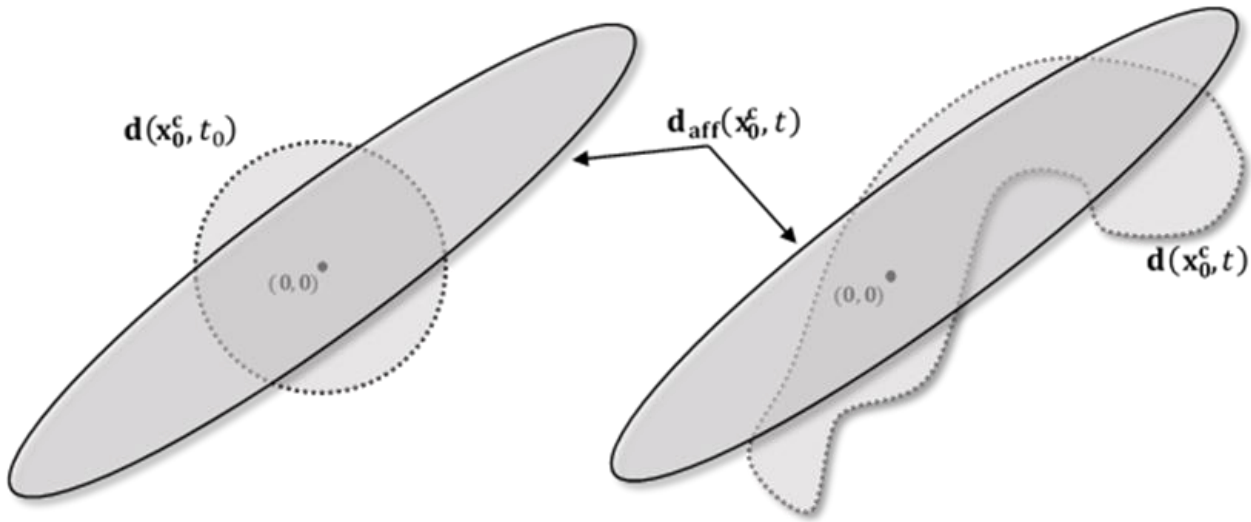
**Figure 22.** Geometric representation of the initial displacement vectors  $\mathbf{d}(\mathbf{x}_0^c, t_0)$ , and the displacement vectors at time  $t$ ,  $\mathbf{d}(\mathbf{x}_0^c, t)$ . Both vectors reflect the transformation of respective particle positions in  $M_t$  and  $M_0$ , such that they are centered about  $(0,0)$ , which is shown as a black dot.



The  $N^{th}$  component of the least squares linear model  $\mathbf{d}_{\text{aff}}(\mathbf{x}_0^c, t)$  is computed as

$$\mathbf{d}_{\text{aff}}^{(N)}(\mathbf{x}_0^c, t) = (\mathbf{A} + \mathbf{I})\mathbf{d}^{(N)}(\mathbf{x}_0^c, t). \quad (2.12)$$

where  $\mathbf{A}$  is the  $2 \times 2$  affine deformation matrix and  $\mathbf{I}$  is the identity matrix. The linear model  $\mathbf{d}_{\text{aff}}(\mathbf{x}_0^c, t)$  essentially describes the ellipse that provides the best fit (in the least squares sense) to the positions of particles contained in  $M_t$ . An illustration depicting the geometric interpretation of the linear model is provided in Figure 23.



**Figure 23.** Geometric representation of the linear model  $\mathbf{d}_{\text{aff}}(\mathbf{x}_0^c, t)$ , shows the black-outlined ellipse on both sides of the figure. The left side shows  $\mathbf{d}_{\text{aff}}(\mathbf{x}_0^c, t)$  superimposed over the initial displacement vectors, represented by the black dotted circle, labeled  $\mathbf{d}(\mathbf{x}_0^c, t_0)$ . The right side of the figure shows  $\mathbf{d}_{\text{aff}}(\mathbf{x}_0^c, t)$  superimposed over the time  $t$  displacement vectors, represented by the black dotted line, labeled  $\mathbf{d}(\mathbf{x}_0^c, t)$ . The position  $(0,0)$  is indicated by black dots on both sides of the figure.

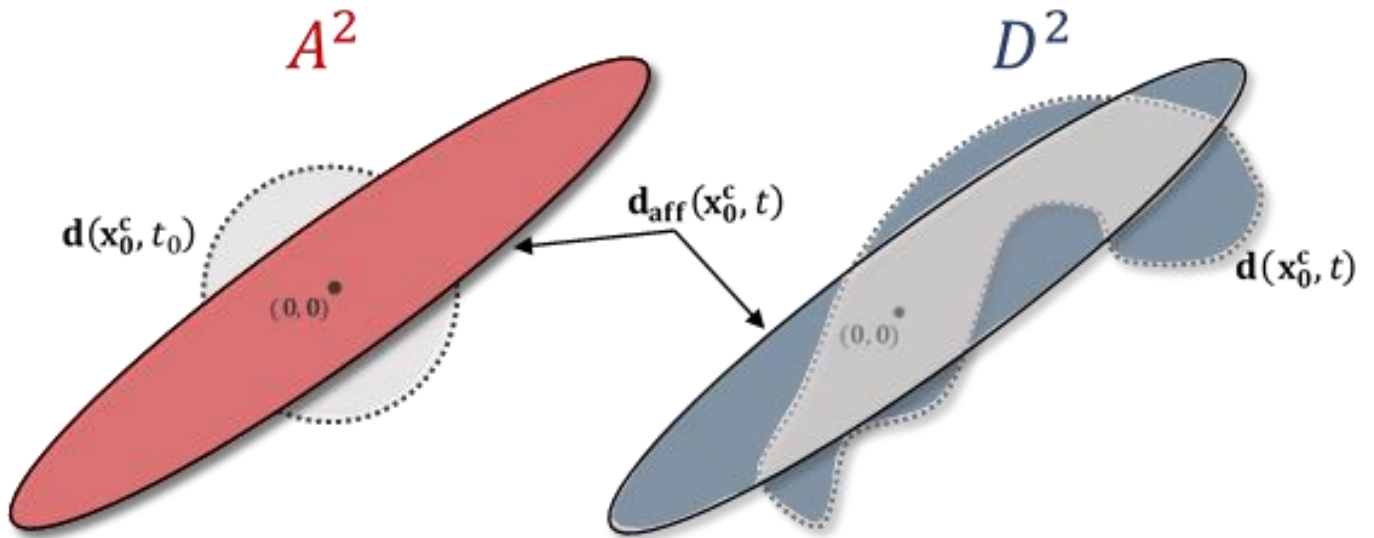
The stretching (affine deformation) experienced by  $M_t$  is described by the dimensionless quantity  $A^2$ , which is computed as

$$A^2(\mathbf{x}_0^c, t) = \frac{1}{r_0^2 n} \sum_{N=1}^n \|\mathbf{A} \cdot \mathbf{d}^{(N)}(\mathbf{x}_0^c, t)\|_2^2 \quad (2.13)$$

where  $\langle \cdot \rangle$  denotes the matrix-vector product,  $r_0$  is the initial radius of the particles in  $M_0$ , and  $n$  is the number of particles in  $M_0$ . Similarly, the folding (non-affine deformation) experienced is described by the dimensionless quantity  $D^2$ , which is computed as

$$D^2(\mathbf{x}_0^c, t) = \frac{1}{r_0^2 n} \sum_{N=1}^n \left\| \mathbf{d}^{(N)}(\mathbf{x}_0^c, t) - \mathbf{d}_{\text{aff}}^{(N)}(\mathbf{x}_0^c, t) \right\|_2^2, \quad (2.14)$$

and represents the inability of a purely linear model to describe the deformation of  $M_t$ . When the linear model fit is poor, values of  $D^2$  are large while values of  $A^2$  are small. If very little, or no deformation (affine or non-affine) is experienced, values of  $D^2$  and  $A^2$  would both be very small. An illustration depicting the geometric interpretation of  $A^2$  and  $D^2$  is provided in Figure 24.



**Figure 24.** Geometric representation of  $A^2$  (left) and  $D^2$  (right). The red shaded area on the left corresponds to  $A^2$  (stretching), representing the cumulative length of all vectors comprising the linear model. The blue shaded area on the right corresponds to  $D^2$  (folding), representing the residual fit between the linear model and the displacement vectors at time  $t$ . The remainder of the figure is identical to Figure 23.

The contribution ratio  $R_{sf}(\mathbf{x}_0^c, t) = \frac{D^2(\mathbf{x}_0^c, t)}{A^2(\mathbf{x}_0^c, t)}$  can be used to measure the relative contributions of stretching and folding to rearrangement of fluid particles (Kelly and Ouellette, 2011). Values of  $R_{sf}(\mathbf{x}_0^c, t) > 1$  indicate that folding is dominant, while values of  $R_{sf}(\mathbf{x}_0^c, t) < 1$  indicate that stretching is dominant. In-situations where  $A^2(\mathbf{x}_0^c, t) \ll D^2(\mathbf{x}_0^c, t) \ll 1$ ,  $R_{sf}(\mathbf{x}_0^c, t)$  can become numerically unstable, and provides little information because the overall amount of spreading is very small. To account for this, a small perturbation  $\varepsilon$  is introduced into the numerator and denominator, yielding

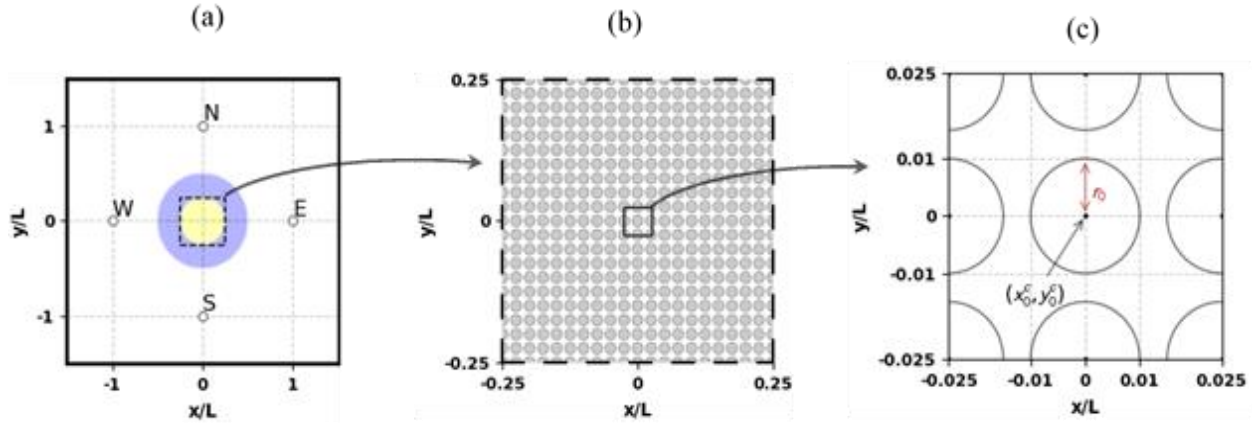
$$R_{sf}(\mathbf{x}_0^c, t) = \frac{D^2(\mathbf{x}_0^c, t) + \varepsilon}{A^2(\mathbf{x}_0^c, t) + \varepsilon}. \quad (2.15)$$

A value of  $\varepsilon = 0.01$  is used throughout the remainder of this work.

#### 2.6.4 General Methods for Evaluating Spreading Metrics

In this section the methods used to produce the results in Section 4 are described. Spreading metrics are computed in space and time for each EIE sequence. The metrics presented in Section 3.1 are computed at every point on a regular  $N_x \times N_y$  grid within the region inscribed by the white dashed line in Figure 12 (also the black dashed line in Figure 25a). At  $t = 0$  a single particle is placed at every point  $(x_0^c, y_0^c)$  on the grid. An additional  $n = 1000$  particles are placed uniformly in a circle about each grid point, with radius  $r_0 = 0.01L$ . Here,  $N_x = N_y = 21$  was used, with equal spacing  $\Delta x = \Delta y = 0.025L$  between grid points in both coordinate directions, ranging from  $(x_{min}, x_{max}) = (y_{min}, y_{max}) = (-0.25L, 0.25L)$ . Figure 15 illustrates the initial configuration of particles used to compute spreading metrics.





**Figure 25.** Map view of the inner part of the model aquifer, illustrating the initial arrangement of the  $N_x \times N_y$  particle clusters used to evaluate spreading metrics. **(a)** Similar to Figure 1. The initial location of contaminant particles (opaque blue) and treatment solution (opaque yellow) plumes used in the simulations from Section 2.4 are shown for reference. The black dashed line is identical to the white dashed line in Figure 12, indicating the region where particles used to compute spreading metrics are initially distributed. **(b)** Magnified view of the region inscribed by the black dashed line in subplot (a). Grey circles represent the area inscribed by particles surrounding each grid point. Grid points are distributed uniformly throughout the region on a . **(c)** Magnified view of the region inscribed by the solid black line in subplot (b), illustrating the initial arrangement of particles around each grid point. The black dot labeled  $(x_0^c, y_0^c)$  denotes the initial location of a particle placed on a given grid point. The solid black line represents the locations of  $n = 1000$  particles placed uniformly on a circle of radius  $r_0 = 0.01L$ . (shown in red) about the gridpoint.

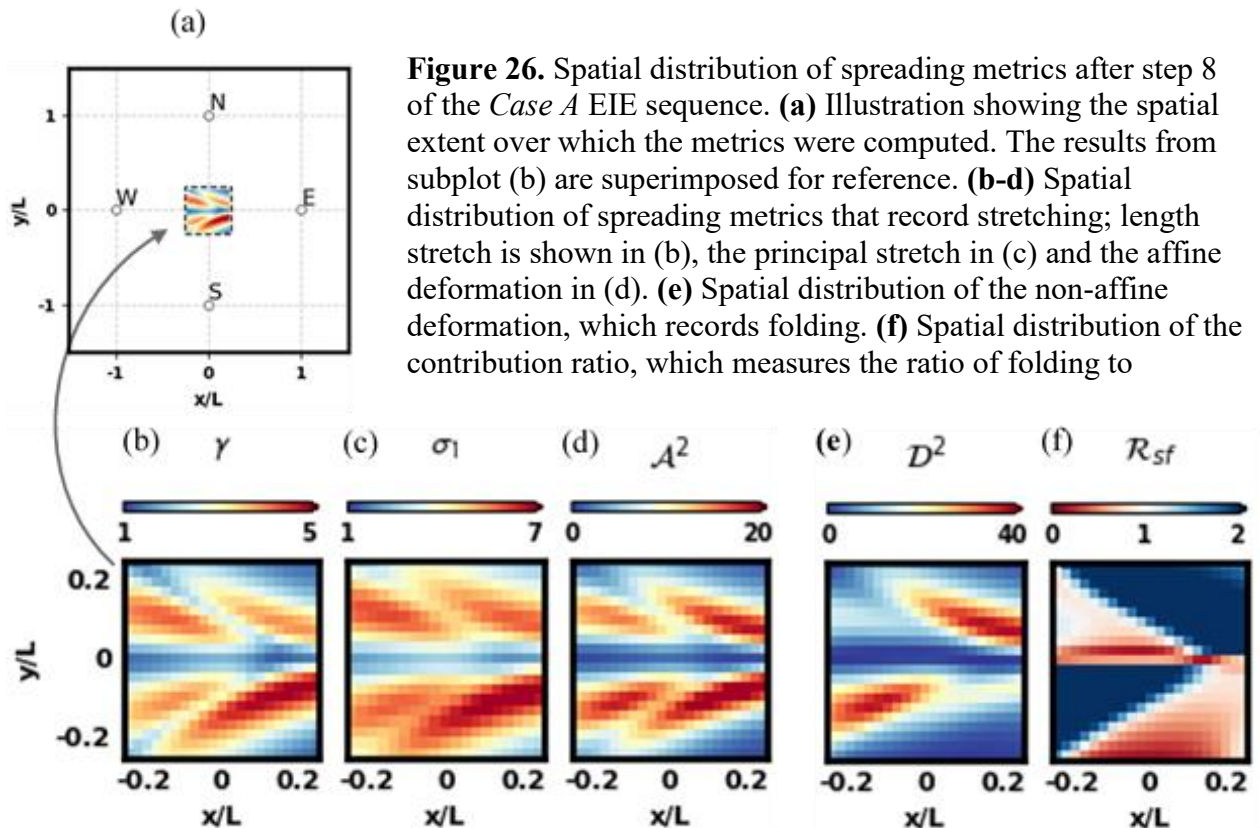
In these simulations, the particles described above are subjected to transport by advection alone. The position of all particles are computed after every step of a given EIE sequence. After each step of the sequence, spreading metrics are computed using the current positions of particles that were initially placed at or around each gridpoint. After time  $t$  has elapsed, for a given initial position  $\mathbf{x}_0^c$ , the length stretch  $\gamma(\mathbf{x}_0^c, t)$  is determined by computing the ratio of the length of the line connecting positions of the outer  $n$  particles at that time to the length of the line connecting them at  $t = 0$ , the latter of which is equal to  $2\pi r_0$  in this configuration. The affine deformation  $A^2(\mathbf{x}_0^c, t)$  and non-affine deformation  $D^2(\mathbf{x}_0^c, t)$  measure the orientation of the  $n$  outer particle positions with respect to their initial orientation, and the contribution ratio

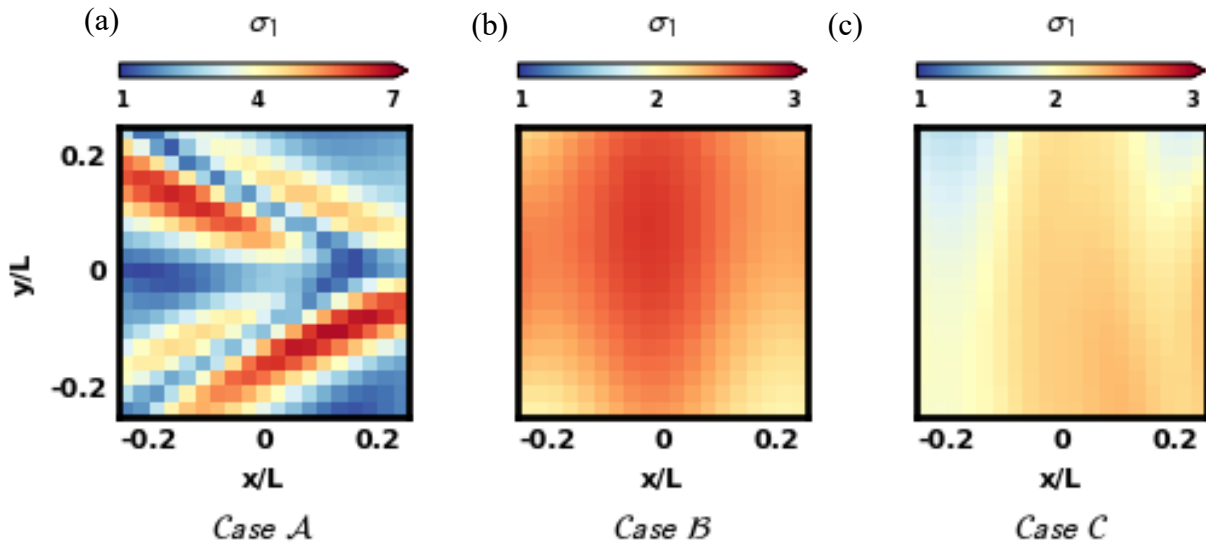
$R_{sf}(\mathbf{x}_0^c, t)$  is computed as the ratio of the two using  $\varepsilon = 0.01$ . The principal stretch  $\sigma_1(\mathbf{x}_0^c, t)$  measures the relative motion of neighboring particles initially placed at each gridpoint. Thus, evaluating  $\sigma_1(\mathbf{x}_0^c, t)$  does not require tracking the additional  $n$  outer particles surrounding each gridpoint, incurring significantly less computational expense.

## 2.7 Results

### 2.7.1 Spatial Distribution of Spreading

At a given time, the spreading metrics can be visualized by showing the metric value throughout space. The metric value at each point in space reflects the amount of spreading experienced over time interval  $[0, t]$  by fluid particles originating nearby that point. An example of this can be seen in Figure 26, which shows the spatial distribution of spreading recorded by each metric, after step 8 of the *Case A* EIE sequence.





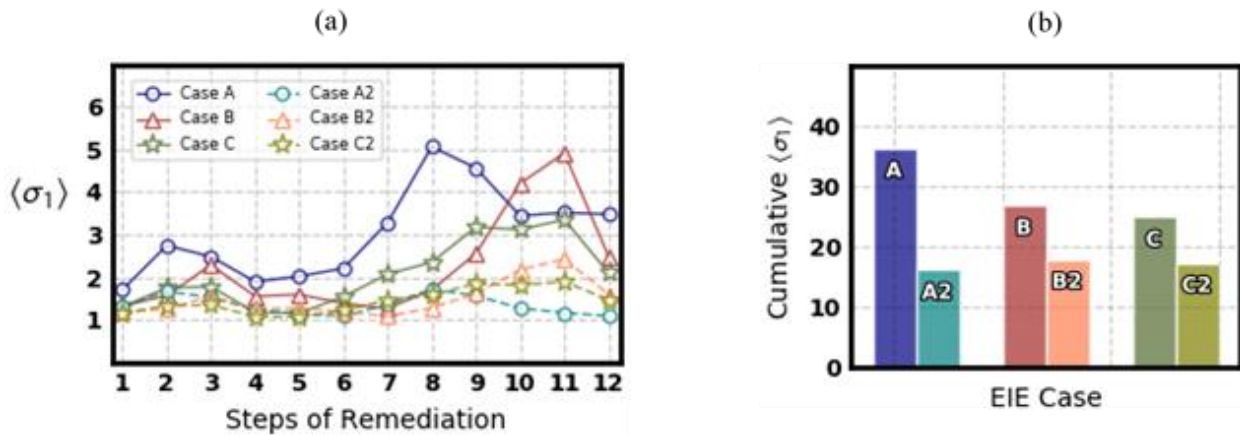
**Figure 27.** Spatial distribution of the principal stretch after step 12 of the *Case A*, *Case B* and *Case C* EIE sequences.

Figure 27 shows the spatial distribution of the principal stretch after step 12 of the *Case A*, *Case B* and *Case C* EIE sequences. Results for *Case A2*, *Case B2* and *Case C2* are shown in Figure A2 of Appendix A. These results show that the *Case A* sequence produces more spatially heterogeneous stretching than *Case B* or *Case C*, however, it is not clear how this information can be used to predict the relative performance of the EIE sequences.

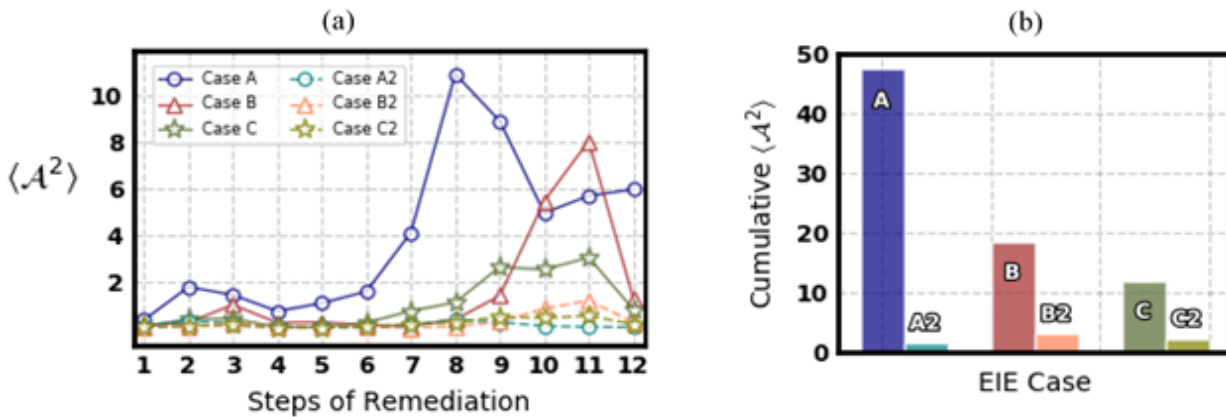
### 2.7.2 Time-Dependent Spreading

To investigate temporal trends in spreading, spatially averaged values  $\langle \cdot \rangle$  of each metric were computed by averaging values across the entire grid after each step of a given EIE sequence. Results are presented in Figure 17 (from Section 2.5) and Figures 28-31. Subplot (a) of each figure shows the spatial averaged values of the corresponding metric for each EIE sequence, plotted against time. Subplot (b) of each show cumulative spatially averaged metric values for each EIE sequence, computed by summing spatially averaged values over time for each metric. Because  $R_{sf}$  is a ratio, cumulative spatially averaged values do not provide meaningful

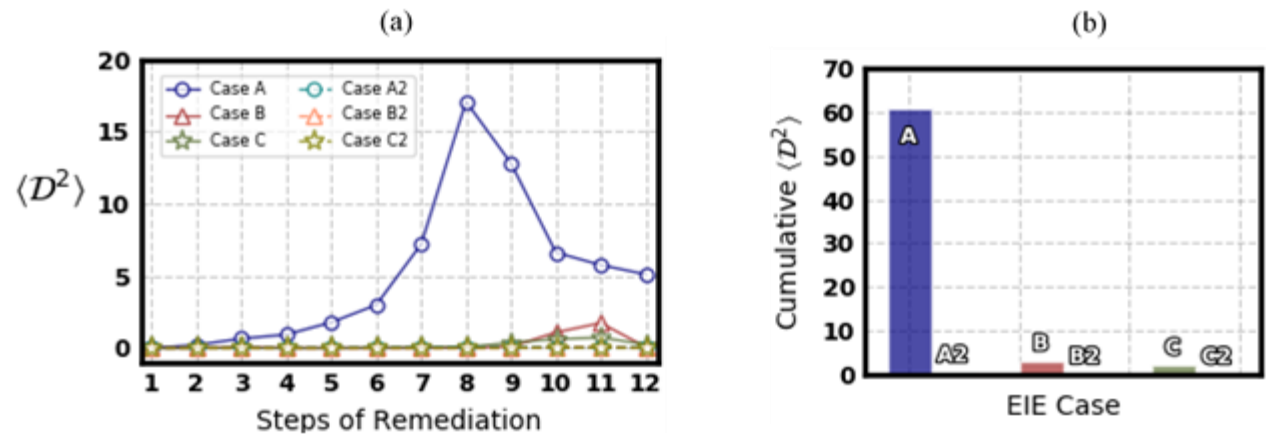
information. Instead the average value of  $R_{sf}$  is computed throughout space and time by averaging  $\langle R_{sf} \rangle$  over all steps of a given EIE sequence, as shown in subplot (b) of Figure 31. The cumulative contaminant degradation and water usage (shown in Figure 16b, 16c) along with the cumulative spreading metrics (shown in Figures 28b-31b) are reported in Table 2 for ease of comparison.



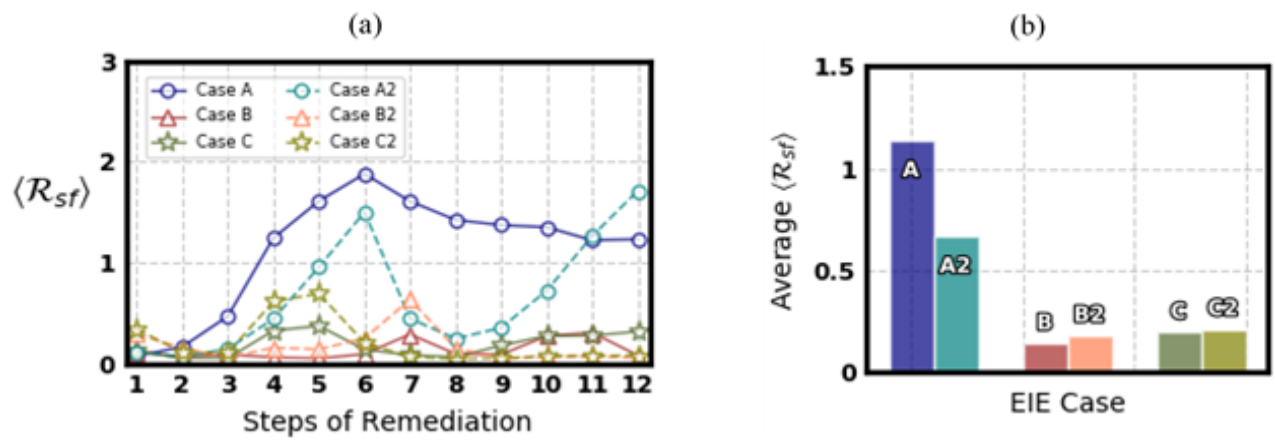
**Figure 28.** (a) Spatially averaged values of the principal stretch  $\langle \sigma_1 \rangle$  after each step of the EIE sequences. (b) Bar chart showing cumulative  $\langle \sigma_1 \rangle$  after Step 12 of each EIE sequence.



**Figure 29.** (a) Spatially averaged values of the affine deformation  $\langle A^2 \rangle$  after each step of the EIE sequences. (b) Bar chart showing cumulative  $\langle A^2 \rangle$  after of each EIE sequence.



**Figure 30.** (a) Spatially averaged values of the non-affine deformation  $\langle D^2 \rangle$  after each step of the EIE sequences. (b) Bar chart showing cumulative  $\langle D^2 \rangle$  after step 12 of each EIE sequence.



**Figure 31.** (a) Spatially averaged values of the contribution ratio  $\langle R_{sf} \rangle$  after each step of the EIE sequences. (b) Bar chart showing the time averaged value of  $\langle R_{sf} \rangle$  after step 12 of each EIE sequence.

**Table 2.** Cumulative values of spatially averaged spreading measures Cumulative Water Use and Cumulative Contaminant Degradation, along with time averaged values of the contribution ratio.

EIE Case	Mass Degraded %	Cum. Water use $m^3/d$	Cum. $\langle \gamma \rangle$	Cum. $\langle \sigma_1 \rangle$	Cum. $\langle A^2 \rangle$	Cum $\langle D^2 \rangle$	Avg. $\langle R_{sf} \rangle$
A	60.6	7000	23.6	34.4	47.5	61.14	1.14
B	71.6	8783	17.3	26.9	18.5	3.12	.14
C	62.8	5965	16.4	24.9	12.0	2.16	.19
A2	34.2	3500	12.8	16.2	1.6	0.48	.66
B2	40.8	4391.5	13.3	17.9	3.0	0.12	.17
C2	32.2	2982.5	13.0	17.3	2.2	0.07	0.2

## 2.8 Discussion

### 2.8.1 Spatial Distribution of Spreading

The spatial distribution of  $\gamma$ ,  $\sigma_1$  and  $A^2$  (Shown in Figure 16b, 16c and 16d) record the same distinct spatial trends, although assume values over different ranges. This is intuitive as they all measure the stretching of initially nearby particle groups. These results can be compared to Figure 9a of Neupauer et al (2014), who used the length stretch ( $\gamma$ ) to measure the spatial distribution of spreading after step 12 of the *Case A* EIE sequence in homogenous and heterogeneous model aquifers.

While there are clear differences in the spatial distribution of spreading produced by each EIE sequence (shown in Figure 17), the utility of this representation of the data is not currently clear.

To a certain extent, a framework for understanding the behavior of chaotic flows is provided by dynamical systems theory, which can be applied in both experimental and numerical settings.

Periodic points and invariant manifolds of the accompanying Poincarè map form the fundamental building blocks of this theory, providing a template for the transport of advected

fluid elements. For a thorough review of these concepts, refer to Ottino (1989, ch.5) and Strogatz (2015, ch.10). This framework was applied extensively in the work of Mays and Neupauer (2012) to identify periodic points, invariant manifolds and heteroclinic orbits in analytical models of the *Case A* EIE sequence used in this work. Neupauer et al (2014) observed that after one iteration of the EIE sequence, the continuous ridges of local maxima in the spatial distribution of the local length stretch aligned with the unstable manifolds of period-one hyperbolic periodic points in the measurement region. Theoretical studies (e.g., Haller and Yuan, 2000; Haller, 2001) have also established a connection between maxima in stretching fields and the invariant manifolds of chaotic flows.

The spatial distribution of  $D^2$  (shown in figure 16e) reveals two isolated regions where initially nearby particles experience a large amount of folding. The spatial distribution of the contribution ratio  $R_{sf}$  (shown in Figure 16f) shows that the dominant mode of spreading produced by the *Case A* EIE sequence (i.e., either stretching or folding) varies spatially. In the top right and bottom right regions of the measurement region, values of  $R_{sf}(\mathbf{x}_0^c, t) > 1$  indicate that folding is dominant. Values of  $R_{sf}(\mathbf{x}_0^c, t) < 1$  indicate that stretching is dominant throughout the remainder of the region.

The work of Chionga and Cirpka (2015) used the affine/non-affine deformation metrics to characterize helical flow in non-stationary anisotropic heterogeneous aquifers. They suggested that the spatial distribution of stretching and folding may be useful for identifying “stagnant” regions, where fluid particles experience relatively low amounts of spreading relative to the rest of the system.

In chaotic flows, exponential stretching causes numerical errors to grow rapidly (Franjione and Ottino, 1987). Errors typically align with the unstable manifolds of hyperbolic periodic points (Soulvaïotis et al., 1995). Consequently, the uncertainty in particle trajectories nearby the unstable manifolds will increase exponentially over time. Because degradation is generally simulated by requiring particles to be in close proximity to react, uncertainty in particle trajectories leads directly to greater uncertainty in predicted reactant concentration. As such, the spatial distribution of spreading for a given EIE sequence might be utilized to quantify the spatial distributions in uncertainty associated with simulation of contaminant degradation.

### 2.8.2 Time-Dependent Spreading

Several observations can be made from the results shown in Figures 28-31. For a given EIE case, the length stretch ( $\gamma$ ), principal stretch ( $\sigma_1$ ) and affine deformation ( $A^2$ ) display similar trends in their spatially averaged values throughout time. For example, all three metrics record a large degree of stretching occurring after Steps 2, 7 and 8 of the *Case A* sequence (shown as the solid blue line, marked with open circles in Figures 17a, 28a, 29a). Similar observations can be made after Step 11 of the *Case B* sequence (shown as a solid red line, marked with open triangles in Figures 17a, 28a-29a), and Steps 9-11 of the *Case C* sequence (shown as a solid green line, marked with open stars in Figures 17a, 28a, 29). This is not the case for the non-affine deformation ( $D^2$ ), which indicates that *Case A* is the only EIE sequence that produces a significant amount of folding (shown as the solid blue line, marked with open circles in Figure 31). Although all other cases produced some degree of stretching (measured by  $\gamma$ ,  $\sigma_1$  and  $A^2$ ), the amount of folding (measured by  $D^2$ ) was small in comparison (compare Figures 17, 28, 29 to Figure 30).



For a given EIE sequence, spatially averaged spreading (i.e. either stretching or folding) can vary considerably over time. For example, the steps 2, 7 and 8 of the *Case A* sequence produce a significant amount of stretching, evidenced by relatively large values of  $\langle \gamma \rangle$ ,  $\langle \sigma_1 \rangle$  and  $\langle A^2 \rangle$  (Figures 17a, 28a, and 29a). Step 8 of *Case A* also results in a large amount of folding, evidenced by relatively large values of  $D^2$  after this step (Figure 30a).

Spatially averaged stretching and folding also varies considerably between EIE sequences. As shown in Figures 17a, 28a-31a, at early times, all EIE sequences produce similar amounts of spatially averaged stretching and folding. At later times, *Cases A-C* produce more stretching than *Cases A2-C2*, evidenced by large values of  $\langle \gamma \rangle$ ,  $\langle \sigma_1 \rangle$  and  $\langle A^2 \rangle$ . *Case A* is the only sequence that produced a significant amount of folding, of which temporal trends in spatially averaged values closely matched spatially averaged stretching metrics.

The 50% reduction in pumping magnitude for *Cases A2-C2* result in less stretching and folding than *Cases A-C*; however, this decrease is not proportional among cases. This can be seen by examination of the cumulative spatially averaged values of each metric, shown in Figures 18b-21b. For all metrics, the difference in cumulative stretching between *Case A* and *Case A2* is much greater than between *Case B* and *Case B2* or *Case C* and *Case C2*. By comparing Figure 7 and Figures 17, 28-30 to Figure 16, it can be seen that *Case A* produces more cumulative stretching than *Cases B* and *C*, yet all three produce similar amounts of contaminant degradation. Spatially averaged stretching and folding metrics were not effective at differentiating between performances of *Case A-C* EIE sequences in terms of contaminant degradation. This is also true between *Cases A2-C2*.

These results suggest that stretching metrics are able to predict contaminant degradation, but only with a mild degree of precision. Furthermore, that metrics quantifying stretching of fluid

parcels ( $\langle \gamma \rangle$ ,  $\langle \sigma_1 \rangle$  and  $\langle A^2 \rangle$ ) are far more correlated with reduction of contaminant mass than metrics quantifying folding ( $\langle D^2 \rangle$  and  $R_{sf}$ ).

## 2.9 Conclusion

The simulations and results presented in this work were designed to illustrate the relationship between spreading and contaminant degradation during several different EIE sequences. Because these simulations were conducted in a homogenous aquifer, all spreading is a result of spatial and temporal variations in groundwater velocity generated by the EIE sequences.

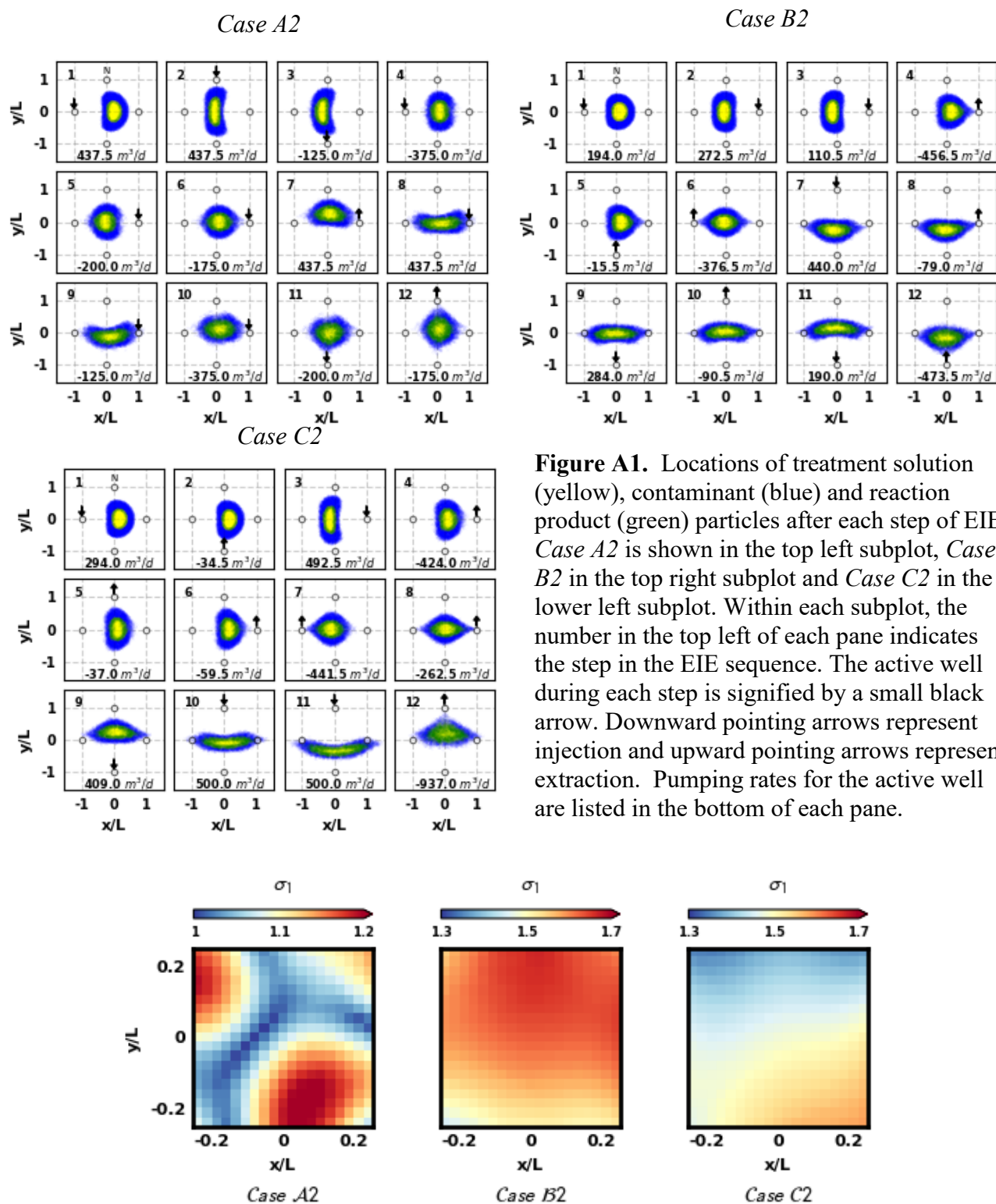
The theory presented in Section 2.1 suggests that increased spreading should correspond to increased contaminant degradation. Optimally determined EIE sequences (*Cases B* and *C*) from Piscopo et al (2016) produced marginally greater contaminant degradation than the heuristic (*Case A*) EIE sequence. However, the amount of spreading produced by the *Case A* sequence is substantially greater than that produced by the *Case B* or *Case C* sequences. The same was true when a 50% reduction in pumping magnitude was imposed (*Case A2-C2*). The amount of degradation produced by *Case A2-C2*

Although spreading metrics were not able to differentiate between EIE cases that produce similar amounts of contaminant degradation (i.e., *Cases A-C* produce ~60-72% and *Cases A2-C2* produce ~32-44%), they were able to differentiate between *Cases A-C* and *Cases A2-C2*, the latter group producing ~26-30% less contaminant degradation than the former. Cumulative spatially averaged stretching metrics were greater for *Cases A-C* than for *Cases A2-C2*.

These results suggest that spatially averaged spreading metrics are an effective proxy for predicted contaminant degradation only at a gross scale. Cumulative spatial averaged spreading can be used to distinguish the contaminant degradation produced by two given EIE sequences,

but only if the amount of contaminant degradation produced by the two sequences differs by >30%. These metrics may be used in optimization to get a computationally efficient but low-precision estimate at how well a candidate sequence may perform. Further, it was found that metrics quantifying folding of fluid parcels displayed poor correlation with reduction of contaminant mass; in contrast, measures of spreading were far more correlated.

## 2.10 Appendix A



## 2.11 References

- Aref, H. (1984). Stirring by chaotic advection. *Journal of Fluid Mechanics*, 143, 1-21. doi:10.1017/S0022112084001233
- Aref, H. (2002). The development of chaotic advection. *Physics of Fluids*, 14(4), 1315-1325.
- Aref, H., Blake, J. R., Budišić, M., Cardoso, S. S., Cartwright, J. H., Clercx, H. J., . . . Tuval, I. (2017). Frontiers of chaotic advection. *Reviews of Modern Physics*, 89(2). doi:10.1103/revmodphys.89.025007
- Arratia, P. E., & Gollub, J. P. (2005). Statistics of Stretching Fields in Experimental Fluid Flows Exhibiting Chaotic Advection. *Journal of Statistical Physics*, 121(5-6), 805-822. doi:10.1007/s10955-005-8664-8
- Bagtzoglou, A. C., & Oates, P. M. (2007). Chaotic Advection and Enhanced Groundwater Remediation. *Journal of Materials in Civil Engineering*, 19(1), 75-83. doi:10.1061/(asce)0899-1561(2007)19:1(75)
- Bakker, M., Post, V., Langevin, C. D., Hughes, J. D., White, J. T., Starn, J. J., & Fienen, M. N. (2016). Scripting MODFLOW Model Development Using Python and FloPy. *Groundwater*, 54(5), 733-739. doi:10.1111/gwat.12413
- Boylund, P., Stremmer, M., & Aref, H. (2003). Topological fluid mechanics of point vortex motions. *Physica D: Nonlinear Phenomena*, 175(1-2), 69-95.
- Cerbelli, S., Zalc, J. M., & Muzzio, F. J. (2000). The evolution of material lines curvature in deterministic chaotic flows. *Chemical Engineering Science*, 55(2), 363-371. doi:10.1016/s0009-2509(99)00331-0
- Chiogna, G., Cirpka, O. A., & Herrera, P. A. (2015). Helical Flow and Transient Solute Dilution in Porous Media. *Transport in Porous Media*, 111(3), 591-603.
- Cirpka, O. A., Rolle, M., Chiogna, G., Barros, F. P., & Nowak, W. (2012). Stochastic evaluation of mixing-controlled steady-state plume lengths in two-dimensional heterogeneous domains. *Journal of Contaminant Hydrology*, 138-139, 22-39. doi:10.1016/j.jconhyd.2012.05.007
- Dagan, G. (1989). *Flow and transport in porous formations*. Berlin: Springer-Verlag.
- De Anna, P., Dentz, M., Tartakovsky, A., & Le Borgne, T. (2014). The filamentary structure of mixing fronts and its control on reaction kinetics in porous media flows. *Geophysical Research Letters*, 41(13), 4586-4593.
- Dentz, M., Borgne, T. L., Englert, A., & Bijeljic, B. (2011). Mixing, spreading and reaction in heterogeneous media: A brief review. *Journal of Contaminant Hydrology*, 120-121, 1-17. doi:10.1016/j.jconhyd.2010.05.002
- Dentz, M., Le Borgne, T., Lester, D. R., & De Barros, F. P. (2017). Mixing in Groundwater. In *The Handbook of Groundwater Engineering* (3rd ed., pp. 383-413). New York, NY: CRC Press.
- Dieter, C. A., Maupin, M. A., Caldwell, R. R., Harris, M. A., Ivahnenko, T. I., Lovelace, J. K., . . . Linsey, K. S. (2018). Estimated use of water in the United States in 2015. Circular. doi:10.3133/cir1441
- Falk, M. L., & Langer, J. S. (1998). Dynamics of viscoplastic deformation in amorphous solids. *Physical Review E*, 57(6), 7192-7205. doi:10.1103/physreve.57.7192
- Franjione, J. G., & Ottino, J. M. (1987). Feasibility of numerical tracking of material lines and surfaces in chaotic flows. *Physics of Fluids*, 30(12), 3641. doi:10.1063/1.866449
- Haller, G. (2001). Distinguished material surfaces and coherent structures in three-dimensional fluid flows. *Physica D: Nonlinear Phenomena*, 149(4), 248-277. doi:10.1016/s0167-2789(00)00199-8
- Haller, G. (2015). Lagrangian Coherent Structures. *Annual Review of Fluid Mechanics*, 47(1), 137-162.
- Haller, G., & Yuan, G. (2000). Lagrangian coherent structures and mixing in two-dimensional turbulence. *Physica D: Nonlinear Phenomena*, 147(3-4), 352-370. doi:10.1016/s0167-2789(00)00142-1

Harbaugh, A. W. (2005). MODFLOW-2005 : The U.S. Geological Survey modular ground-water model--the ground-water flow process. *Techniques and Methods*.

Hernández-García, E., López, C., & Neufeld, Z. (2002). Small-scale structure of nonlinearly interacting species advected by chaotic flows. *Chaos: An Interdisciplinary Journal of Nonlinear Science*,12(2), 470-480. doi:10.1063/1.146824

Jones, S. W., & Aref, H. (1988). Chaotic advection in pulsed source–sink systems. *Physics of Fluids*,31(3), 469. doi:10.1063/1.866828

Jose, S. C., & Cirpka, O. A. (2004). Measurement of Mixing-Controlled Reactive Transport in Homogeneous Porous Media and Its Prediction from Conservative Tracer Test Data. *Environmental Science & Technology*,38(7), 2089-2096.

Kapoor, V., & Kitanidis, P. K. (1998). Concentration fluctuations and dilution in aquifers. *Water Resources Research*,34(5), 1181-1193. doi:10.1029/97wr03608

Kelley, D. H., & Ouellette, N. T. (2011). Separating stretching from folding in fluid mixing. *Nature Physics*,7(6), 477-480.

Le Borgne, T., Dentz, M., & Villermaux, E. (2013). Stretching, Coalescence, and Mixing in Porous Media. *Physical Review Letters*,110.

Lester, D. R., Metcalfe, G., Trefry, M. G., Ord, A., Hobbs, B., & Rudman, M. (2009). Lagrangian topology of a periodically reoriented potential flow: Symmetry, optimization, and mixing. *Physical Review E*,80.

Liu, M., Peskin, R. L., Muzzio, F. J., & Leong, C. W. (1994). Structure of the stretching field in chaotic cavity flows. *AIChE Journal*,40(8), 1273-1286. doi:10.1002/aic.690400802

Mays, D. C., & Neupauer, R. M. (2012). Plume spreading in groundwater by stretching and folding. *Water Resources Research*,48.

Metcalfe, G., Lester, D., Ord, A., Kulkarni, P., Rudman, M., Trefry, M., Morris, J. (2010). An experimental and theoretical study of the mixing characteristics of a periodically reoriented irrotational flow. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*,368(1918), 2147-2162.

Muzzio, F. J., Swanson, P. D., & Ottino, J. M. (1991). The statistics of stretching and stirring in chaotic flows. *Physics of Fluids A: Fluid Dynamics*,3(5), 822-834. doi:10.1063/1.858013

Muzzio, F., & Liu, M. (1996). Chemical reactions in chaotic flows. *The Chemical Engineering Journal and the Biochemical Engineering Journal*,64(1), 117-127. doi:10.1016/s0923-0467(96)03108-9

Neupauer, R. M., Meiss, J. D., & Mays, D. C. (2014). Chaotic advection and reaction during engineered injection and extraction in heterogeneous porous media. *Water Resources Research*,50(2), 1433-1447.

Ottino, J. (1990). Mixing, Chaotic Advection, And Turbulence. *Annual Review of Fluid Mechanics*,22(1), 207-253.

Ottino, J. M. (1989). *The kinematics of mixing: Stretching, chaos, and transport*. Cambridge: Cambridge Univ. Press.

Ottino, J. M., Jana, S. C., & Chakravarthy, V. S. (1994). From Reynolds's stretching and folding to mixing studies using horseshoe maps. *Physics of Fluids*,6(2), 685-699. doi:10.1063/1.868308

Piscopo, A. N., Mays, D. C., & Neupauer, R. M. (2013). Engineered Well Injection-Extraction Schemes to Enhance Reaction for Improved In Situ Remediation of Contaminated Groundwater. *World Environmental and Water Resources Congress 2012*.

Piscopo, A. N., Neupauer, R. M., & Kasprzyk, J. R. (2016a). Optimal design of active spreading systems to remediate sorbing groundwater contaminants in situ. *Journal of Contaminant Hydrology*,190, 29-43. doi:10.1016/j.jconhyd.2016.03.005

Piscopo, A., Kasprzyk, J., & Neupauer, R. (2016b). An iterative approach to multi-objective engineering design: Optimization of engineered injection and extraction for enhanced groundwater remediation. *Environmental Modelling & Software*,69

Pollock, D. W. (2016). User guide for MODPATH Version 7—A particle-tracking model for MODFLOW. Open-File Report.

Reynolds, O. (1894). Study of Fluid Motion by Means of Coloured Bands1. *Nature*,50(1285), 161-164. doi:10.1038/050161a0

Rolle, M., Eberhardt, C., Chiogna, G., Cirpka, O. A., & Grathwohl, P. (2009). Enhancement of dilution and transverse reactive mixing in porous media: Experiments and model-based interpretation. *Journal of Contaminant Hydrology*,110(3-4), 130-142. doi:10.1016/j.jconhyd.2009.10.003

Salamon, P., Fernández-García, D., & Gómez-Hernández, J. J. (2006). A review and numerical assessment of the random walk particle tracking method. *Journal of Contaminant Hydrology*,87(3-4), 277-305. doi:10.1016/j.jconhyd.2006.05.005

Siegrist, R. L., Crimi, M., Broholm, M. M., Mcray, J. E., Illangasekare, T. H., & Bjerg, P. L. (2011). Advances in Groundwater Remediation: Achieving Effective In Situ Delivery of Chemical Oxidants and Amendments. *Clean Soil and Safe Water NATO Science for Peace and Security Series C: Environmental Security*,197-212. doi:10.1007/978-94-007-2240-8\_15

Soulvaioitis, A., Jana, S. C., & Ottino, J. M. (1995). Potentialities and limitations of mixing simulations. *AIChE Journal*,41(7), 1605-1621. doi:10.1002/aic.690410702

Sposito, G. (2006). Chaotic solute advection by unsteady groundwater flow. *Water Resources Research*,42(6). doi:10.1029/2005wr004518

Strogatz, S. (2015). *Nonlinear dynamics and chaos: With applications to physics, biology, chemistry, and engineering*. Boulder, CO: Westview Press.

Subramanian, N., Kellogg, L. H., & Turcotte, D. L. (2009). Statistics of Advective Stretching in Three-dimensional Incompressible Flows. *Journal of Statistical Physics*,136, 926-944.

Swanson, P. D., & Ottino, J. M. (1990). A comparative computational and experimental study of chaotic mixing of viscous fluids. *Journal of Fluid Mechanics*, 213, 227-249. <https://doi.org/10.1017/S0022112090002300>

Thiffeault, J. (2004). Stretching and curvature of material lines in chaotic flows. *Physica D: Nonlinear Phenomena*,198(3-4), 169-181.

Trefry, M. G., Lester, D. R., Metcalfe, G., Ord, A., & Regenauer-Lieb, K. (2012). Toward enhanced subsurface intervention methods using chaotic advection. *Journal of Contaminant Hydrology*,127(1-4), 15-29. doi:10.1016/j.jconhyd.2011.04.006

Truesdell, C., & Noll, W. (1992). *The Non-Linear Field Theories of Mechanics*. *The Non-Linear Field Theories of Mechanics*,1-579.

Tsang, Y. (2009). Predicting the evolution of fast chemical reactions in chaotic flows. *Physical Review E*, 80(2). doi:10.1103/physreve.80.026305

Uffink, G. J. M. (1989), Application of Kolmogorov's backward equation in random walk simulations of groundwater contaminant transport, in *Contaminant Transport in Groundwater*, edited by H. E. Kobus and W. Kinzelbach, pp. 283–289, A. A. Balkema, Brookfield, Vermont.

Vikhansky, A., & Cox, S. M. (2006). Reduced models of chemical reaction in chaotic flows. *Physics of Fluids*,18(3), 037102. doi:10.1063/1.2182373

Voth, G. A., Haller, G., & Gollub, J. P. (2002). Experimental Measurements of Stretching Fields in Fluid Mixing. *Physical Review Letters*,88(25). doi:10.1103/physrevlett.88.254501

Water Science and Technology Board (2013). Alternatives for Managing the Nations Complex Contaminated Groundwater Sites doi:10.17226/14668

Weeks, S. W., & Sposito, G. (1998). Mixing and stretching efficiency in steady and unsteady groundwater flows. *Water Resources Research*,34(12), 3315-3322.

*World fresh water resources" in Peter H. Gleick (editor), 1993, Water in Crisis: A Guide to the World's Fresh Water Resources (Oxford University Press, New York)*

Zhang, P., Devries, S. L., Dathe, A., & Bagtzoglou, A. C. (2009). Enhanced Mixing and Plume Containment in Porous Media under Time-Dependent Oscillatory Flow. *Environmental Science & Technology*,43(16), 6283-6288. doi:10.1021/es900854r



# Chapter 3

## Calibrating Models that Depend on Variable Data

### 3.1 Introduction

Models of human-Earth systems are often developed with the goal of predicting the behavior of one or more dependent variables from multiple independent variables, processes, and parameters. Often dependent variable values range over many orders of magnitude, which complicates evaluation of the fit of the dependent variable values to observations.

Many metrics and optimization methods have been proposed to address dependent variable variability, with little consensus being achieved. In this work, two such methods are evaluated: log transformation (based on the dependent variable being log-normally distributed with a constant variance) and error-based weighting (based on a multi-normal distribution with variances that tend to increase as the dependent variable value increases). Error-based weighting has the advantage of encouraging model users to carefully consider data errors, such as measurement and epistemic errors, while log-transformations can be a black box for typical users. Placing the log-transformation into the statistical perspective of error-based weighting has not formerly been considered, to the best of our knowledge.

Multiple linear regression (MLR) is used to make the evaluation as clear and reproducible as possible. Simulations are conducted with MATLAB. The example represents stream transport of nitrogen with up to eight independent variables. The single dependent variable in this example has values that range over 4 orders of magnitude. Results are applicable to any problem for which individual or multiple data types produce a large range of dependent variable values. This work considers the consequences of these two common ways of addressing variable data.

### 3.2 Hypothesis to Test

The objective of this investigation is to compare the competency of error based weighting to log transformation for Multiple Linear Regression prediction of heteroscedastic dependent variables. Heteroscedasticity refers to the tendency for samples (sub-populations) of random variable to differ in variability. More precisely, the variability of a predicted variable is inconsistent across the range of values used by the explanatory variables to predict it (Carroll and Ruppert, 1988, Ch. 4).

A small dataset will be used to predict the steam transport of nitrogen using eight explanatory variables. Three Multiple Linear Regression models will be constructed. The first model will serve as the benchmark case, predicting the dependent variable with no transformation or weighting. The second model has a log transformed dependent variable, and the third model uses an error based weighting approach outlined in (Carroll and Rupert, 1988, Ch. 4).

### 3.3 Mathematical Methods

#### 3.3.1 Null Case: No Transformation or Weighting

Multiple Linear Regression (MLR) is used to predict a dependent variable  $Y$  using one or more explanatory variables. The governing equation can be written as,

$$Y = \mathbf{X}\beta + \varepsilon \quad (3.1)$$

where  $\mathbf{X}$  is a  $m \times n$  matrix containing  $m$  observations for each of the  $n$  independent variables. The term  $\varepsilon$  denotes the model error and  $\beta$  is a vector containing the true parameter values  $b_i$  associated with each independent variable. The normal equations can be used to approximate the regression coefficients, given by

$$\hat{\beta} = [\mathbf{X}^T \omega \mathbf{X}]^{-1} \mathbf{X}^T \omega \mathbf{Y} \quad (3.2)$$

where  $\hat{\beta}$  is a vector containing the estimated parameter values,  $Y$  contains the observed values of the dependent variable, and  $\omega$  is a  $m \times m$  matrix containing weights for each observation along the diagonal and zeros elsewhere. Predicted values of the dependent variable, denoted  $\hat{Y}$ , can then be computed as

$$\hat{Y} = \mathbf{X} \hat{\beta} + e \quad (3.3)$$

where  $e = |\hat{Y} - Y|$  is the residual model fit.

Equations 3.1-3.3 can be modified to improve the predictive capacity of the model in situations where the dependent variable ranges over several orders of magnitude.

### 3.3.2 Log Transformation

A common and simple approach is to log transform the dependent variable/ substitution into (3.2) yields the log transformed parameter estimates,

$$\hat{\beta} = [\mathbf{X}^T \omega \mathbf{X}]^{-1} \mathbf{X}^T \omega \log_{10}(Y). \quad (3.3)$$

The log transformed linear model can then be expressed as

$$\log_{10}(\hat{Y}) = \mathbf{X} \hat{\beta} + e. \quad (3.4)$$

which yields predicted values of the dependent variable in log space.

### 3.3.3 Error Based Weighting

An alternative approach is to use error based weighting, which assigns weights to observations with large prediction residuals. Weights are determined by fitting a linear model to the residuals, using the dependent variable  $Y$  as an explanatory variable. This auxiliary model can be expressed

$$e = WY + \varepsilon \quad (3.5)$$

With parameter estimates given by

$$\widehat{W} = [Y^T Y]^{-1} Y^T e^T \quad (3.6)$$

The diagonal elements of the weight matrix are determined as

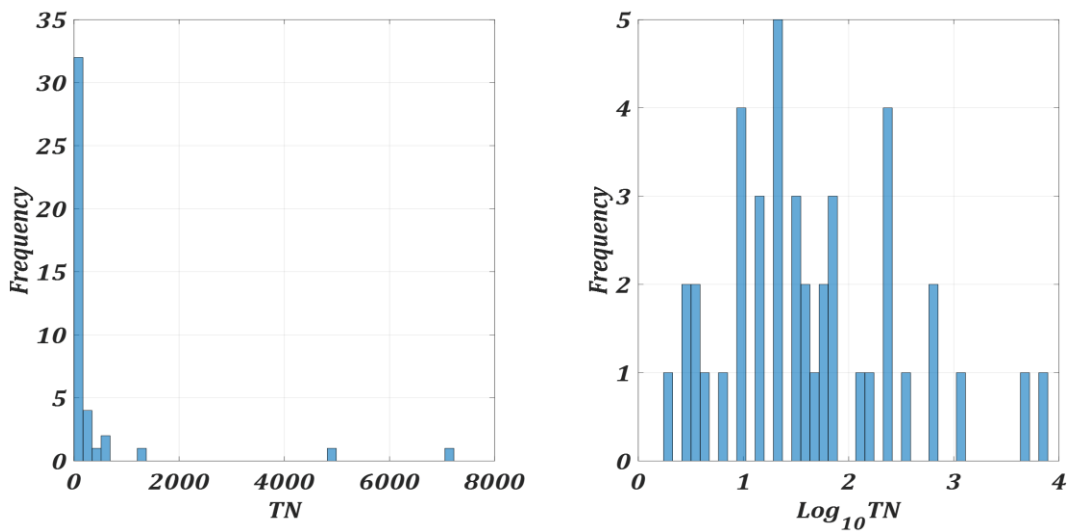
$$\omega_{i,j=k} = \frac{1}{(Y_k \widehat{W}_k)^2}, \quad k = 1, \dots, m \quad (3.7)$$

then substituted back into equation (3.2), yielding new regression coefficients that reflect the error based observation weights.

<b>Table 3.</b> Description of variables in the dataset. Dependent variable TN is indicated by **	
TN**	Total nitrogen load
LOGCA	Log contributing area
LOGIMP	Log impervious area
MMJTEMP	Mean minimum January temperature
MSRAIN	Mean seasonal rainfall
PRES	Percentage of area residential
PNON	Percentage of area non-urban
PCOMM	Percentage of area commercial
PIND	Percentage of area industrial

### 3.4 Dataset

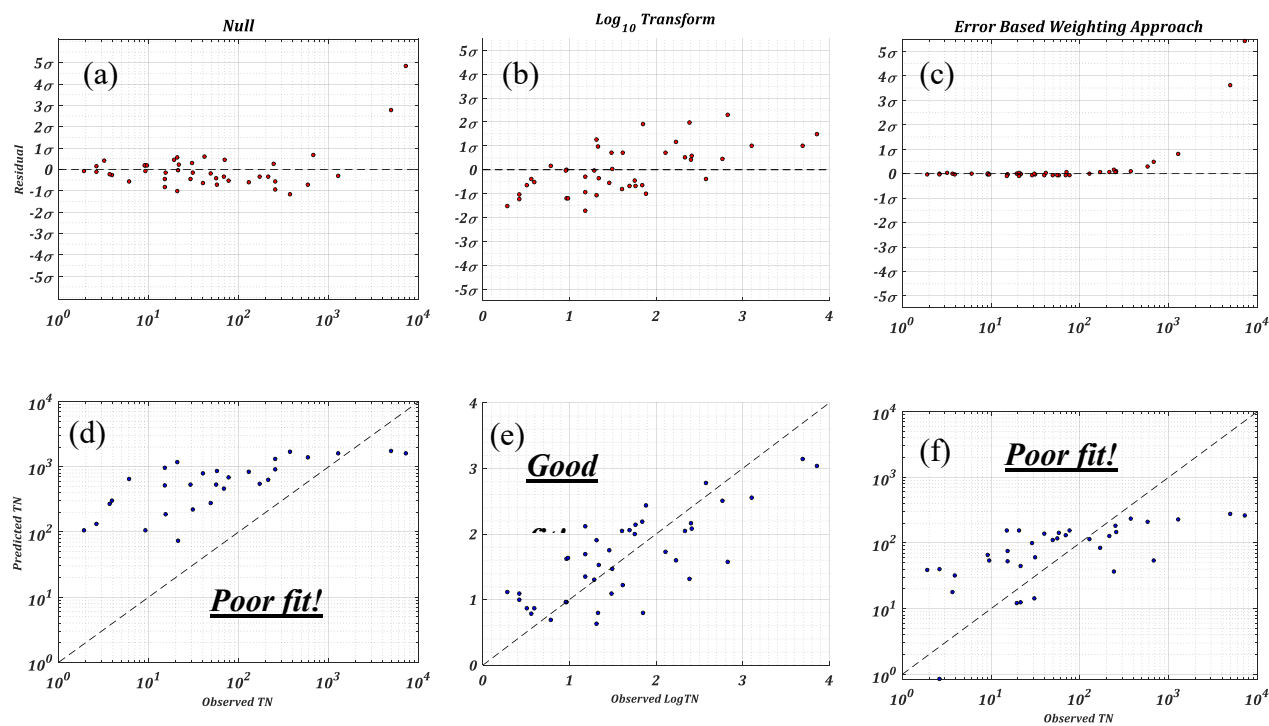
A simple dataset from Helsel and Hirsch (2002, p334) consisting of data from 42 small urban drainage basins located in several cities around the United States. The data consists of 9 variables (shown in Table 3), with 42 observations each. The total nitrogen load (TN) will serve as the dependent variable in this investigation. The remaining 8 variables will be used as explanatory variables in MLR models. Histograms of the native TN values and the LOGTN values are shown in Figure 32. Log transformation normalizes the skewed distribution.



**Figure 32.** Histogram of dependent variable (Left) Histogram of the observed values of total nitrogen load, the dependent variable. (Right) Log transformed total nitrogen.

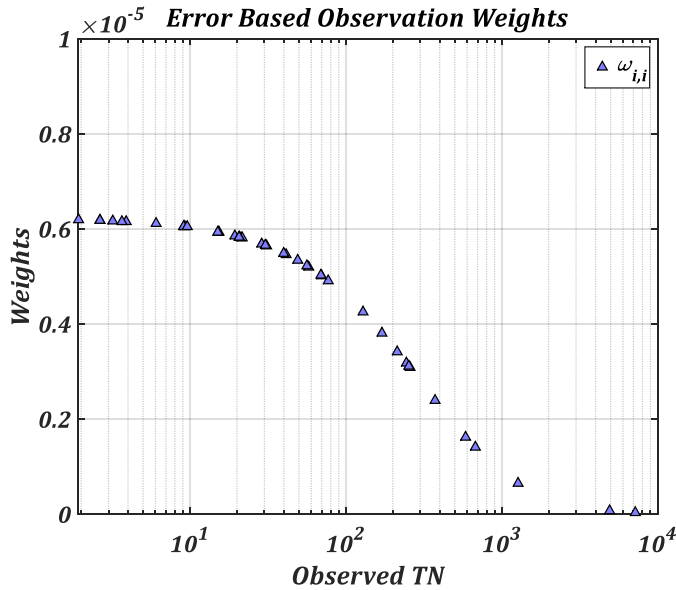
### 3.5 Results

Figure 33 shows the results for each MLR model. Residual quantities for each model are shown on the top subplots (Figure 33a-c). The predicted values of the dependent variable are plotted against observed values in the bottom subplots (Figure 33d-f). The null case, with no log transformation or weighting is shown in the left column. The log transformed model is shown in the middle column and the weighted model, with no log transformation is shown in the right column.



**Figure 33.** Residuals and predicted values plotted against observed dependent values for each MLR model. The Null case (a-c) residuals for each MLR model. (d-e) Prediction results for each MLR model. A 1:1 line is superimposed for qualitative evaluation of model fit.

Error based observation weights determined by equations (3.5), (3.6) and (3.7) are reported in Figure 34. The weight assigned to the native values of the dependent variable decrease in magnitude as the observation values increase in magnitude.



**Figure 34.** Error based observation weights assigned to native values of the dependent variable.

### 3.6 Discussion and Conclusions

The primary motivation for this investigation stems from the difficulties associated with parameter estimation in the presence of heteroscedastic and making meaningful interpretation of model predictions in log space. While log transformation is a simple and effective method for handling heteroscedastic data, converting model predictions from their log transformed value back to their native values is not a simple process. Presenting modeling results in log-space can inhibit project stakeholders from interpreting information provided by the model in a meaningful way.

As shown in Figure 33, the log transformed model has superior model fit compared to the null case and error based weighting model. This supports previous suggestions that error-based weighting derived from a constant coefficient of variation overemphasizes low values and degrades model fit to high values. Applying larger weights to the high values is inconsistent with the log-transformation. Greater consistency is obtained by imposing smaller (by up to a factor of 1/35) weights on the smaller dependent-variable values. From an error-based perspective, the small weights are consistent with large standard deviations. These results show that error based weighting does not provide a viable alternative to log transformation in this circumstance.

### 3.7 References

*Carroll, D. R., & Hirsch, R. M. (1997). Statistical methods in water resources.* Amsterdam: Elsevier.

*Transformation and weighting in regression: By Raymond J. Carroll and David Ruppert* (R. J. Carroll & D. Ruppert, Authors).(1989). Chapter 4. London: Chapman and Hall.



## Epilogue

In this thesis, three fundamental problems related to simulating physical systems were addressed, in the context of solute transport through hydrogeological systems. Results provide a tool for communicating basic, yet vital transport and modeling concepts to introductory audiences, while giving these audiences the fundamental building blocks for practical model applications of groundwater flow and advective transport. In chapter 2, an example of how these basic concepts can be applied in a more complex setting, and expanded to include dispersion processes is provided. Here, it is demonstrated that novel advection-based metrics can be used to capture relevant aspects of a highly complex subsurface system characterized by dispersion and reaction processes. Significantly, it was found that metrics quantifying folding of fluid parcels displayed poor correlation with reduction of contaminant mass; in contrast, measures of spreading were far more correlated. The final chapter of this work addressed the challenges associated with meaningful and effective integration of heteroscedastic data and interpretation of model results. Here, error-based weighting approaches were investigated as alternatives to log-transmutation in the presence of heteroscedastic data. Results suggested that the error-based methods considered produced poor performance of regression-based model calibration relative to log-transforming the data.