MULTI-SOURCED GEOLOGIC DATA INTEGRATION:

A TIME-BASED APPROACH

by

Elliott Andelman

A thesis submitted in partial fulfillment of

the requirements for the degree of

Master of Geological Engineering

at the

UNIVERSITY OF WISCONSIN-MADISON

2017

APPROVAL Page

Advisor Signature: 

Advisor Name (print): Dr. Michael Cardiff

Date: 14 June 2017

Abstract

Complications in geological modeling are often made worse by the wide range of formats in which geologic data are stored. When crossing state lines, the names and interpretations of geologic formations may vary, individual formations can be grouped into one undifferentiated member, and the very existence of a formation may disappear when crossing a neighboring state border. The next generation of geologic mapping requires higher coordination in dealing with multiple datatypes. In order to ameliorate the complications associated with geologic data integration and modeling, I used a 4D time-based approach to integrate multiple different data sources.

The format for data integration contained four columns, having an x and y coordinate in UTM, a z-dimension elevation value in meters and geologic contact age for each data point. The data were extracted from the Geological Map of North America (GMNA) for surface-bedrock records, while subsurface formation thicknesses were provided by drillers logs from at least 150 boreholes and wire-logs across several counties, companies and subsurface exploration projects in Michigan. The size, lack of major faulting and nearly ideal basin shape of the Michigan Basin allowed it to be a reasonable sample site to test this integration method and its ability to produce plausible and highly interactive 3D images of the subsurface.

Post-integration, two methods of geostatistical kriging were performed in the SGeMS software: one in 3D with the dependent variable of interest being geologic time, the other in 2D with respect to the elevation. Further statistical analyses were performed in MATLAB to compare the plausibility of each method. This work will demonstrate approaches to large-

scale 3D data integration and mapping that can be used as a framework for the next generation of computer-based geologic modeling.

Acknowledgements

- Michael Cardiff for supporting me as a graduate student and showing me a side of geological science (and data science) I had never considered and have come to enjoy. He is in my opinion the best advisor that a student could have.
- Jean and Hiroki, the other members of my advising committee, for their support.
- My parents for supporting me through 6 years of geological studies, despite never having a clue of what I was ever talking about.
- Ben Heinle for being my personal MATLAB support unit, saving me countless hours and effort by showing me simpler ways to get to the end result.
- Friends and family at home in Baltimore.
- The geoscience community at UW-Madison.

Contents

List of Tables

List of Figures

1 Introduction

Traditionally data in geology are represented as 1D or 2D. Borehole profiles and stratigraphic strip-logs represent 1D data under the assumption that everything continues outward (in x/y) in the same fashion. Then 2-D data, including map-view (bedrock data for example) and stratigraphic/structural cross-sections are collected and interpreted by geoscientists, but typically only provide a single slice of the subsurface. The development and continuing growth in 3D data storage and visualization is a progression that is needed for geoscience in the next century. Most geologic data are currently stored on paper, with digital aspects existing only in the form of PDF images of maps, cross sections, driller's logs and so on. This makes the process of data integration extremely time-consuming and hinders the level of interactivity among geoscientists looking to make use of the data. Accelerating geologic data storage to 3D and digital formats facilitates interpretation and ultimately makes conceptual models of the subsurface in geoscience more democratic, since more people are able to manipulate the data than just the map author or consulting geologist. This endeavor requires trans-discipline, trans-community, and trans-agency collaboration in a sustained effort (Fan et al., 2015). In order to show the benefits of digitally integrated 3D data, this thesis aims to use geological age as a parameter to connect geological data from various sources (digital and non-digital) to produce a plausible 3D characterization of the subsurface for a region in which 2D representations would not suffice.

1.1 Importance of Three-Dimensional Modeling in Geology

Geoscientists have many reasons to explore the subsurface, whether the end objective is waste site investigation, groundwater flow simulation, mineral resource evaluation, or reservoir and tunnel design. There are many difficulties in integrating different types of geologic data as different states, companies and individual geologists interpret characteristics of the subsurface differently and may pay attention to particular features of the rock while ignoring others.

Use of time as a parameter to integrate datasets is logical because every rock fits into a formation, and every formation has a depositional age. Temporal stratigraphy, or chronostratigraphy, is reliable because geological formations, cover a large lateral area. Even if formation names change (i.e. across some state borders), the time ranges of their deposition remain constant. This allows the data to remain continuous and dependable when integrated from large areas and multiple databases.

1.1.1 Complications in Geological Modeling

The Principle of Original Horizontality states that sediment was deposited in layers that were originally roughly horizontal due to the effect of gravity, causing sediment particles to settle straight down out of water or air (Levin, 2010). If we see flat lying strata today, they are probably in their original positions. Similarly, the Principle of Superposition states that the oldest formation in a sequence of flat-lying strata will be found at the bottom, while the younger formations will be stacked successively on top (Levin, 2010). The combination of these two principles is shown in the simplified case of Figure 1, where for the purpose of modeling formation boundaries, the contact between two formations could be interpreted as a surface. As

long as the coordinates of the contact surface, or data points making up the surface, are known, it is a simple process to interpolate between those points and recreate the contact surface.

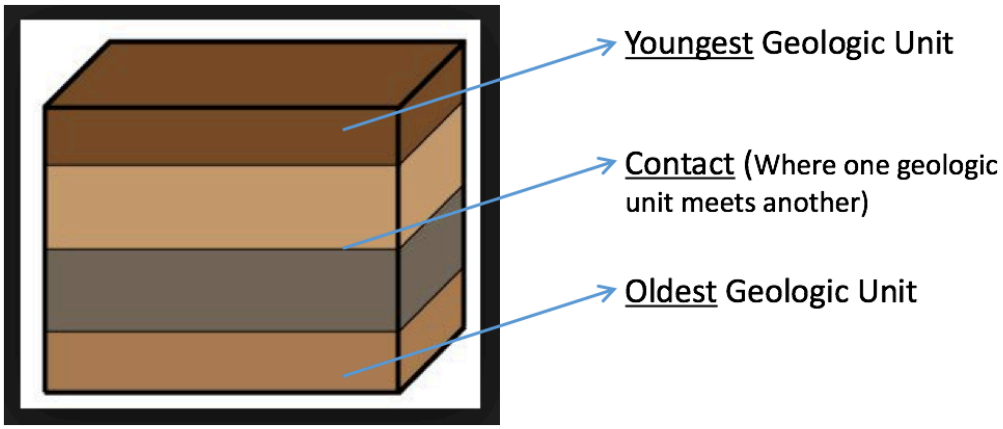


Figure 1: Ideal bedrock stratigraphy sequence.

In an ideal case, the Principles of Original Horizontality and Superposition would sustain through time, but events during and post-deposition contribute to ever evolving complexity. For example, a single formation could contain sediments with significant variability in chemical composition if environmental conditions changed during its depositional phase. Similarly, boundaries between units are further complicated by post-depositional events such as faulting and erosion (Houlding, 1996). Take, for example, a normal fault causing a shift in the formations (Figure 2). Figure 2 a. shows how two surfaces became five with the addition of the fault plane. The fault itself created a new surface, while cutting surface S1 into two separate surfaces and also S2 into two separate surfaces.

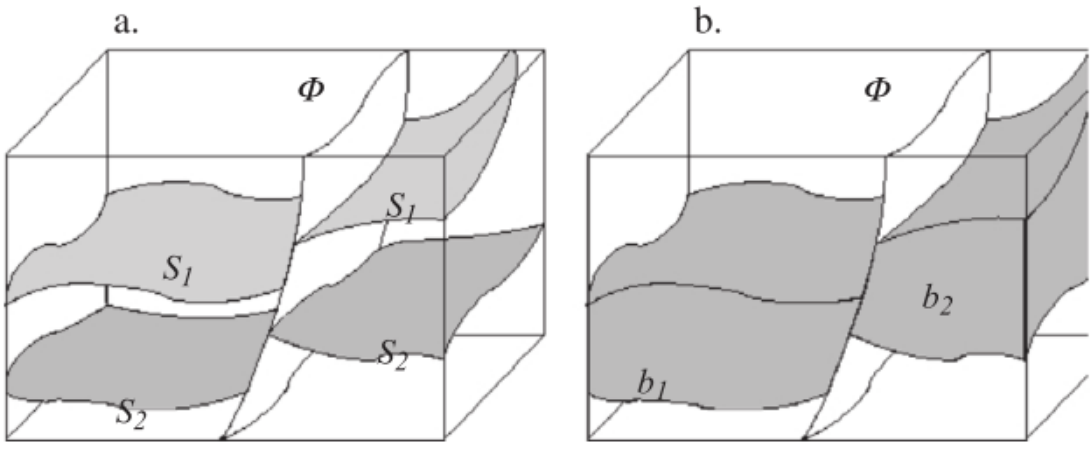


Figure 2: A) Subsurface described as surfaces with fault plane. S1 and S2 represent upper and lower boundaries of same lithological unit offset by normal fault. B) Lithological units described as blocks with fault plane. b1 and b2 represent same lithological unit, offset by normal fault (Perrin and Zhu, 2005).

1.2 Examples of Current Data Integration Efforts

Several efforts for large-scale geological data integration already exist, although most of these are specific to a single category of geologic data (such as soil classifications, tectonic stresses, aquifer maps, etc.). A few of these integrated databases (Global Harmonized Soil Database and Macrostrat) are highlighted as examples of ongoing integration efforts to show their impact on the geologic data community and the benefits of large-scale data integration.

1.2.1 UN-FAO Global Harmonized Soil Database

The Global Harmonized Soil Database unified large amounts of soil-survey data from multiple continents. The database was built "often using different soil taxonomies, horizon definitions and attributes, and then compiled at different scales of resolution and with different formats, becoming harmonized through an international partnership, which defined a new set of soil attributes critical to agriculture and recommended methodologies for developing taxo-

transfer rules" (Fan et al., 2015). The result was 30 arc-sec grids containing 20 physical, chemical and biological soil properties. As a result of this effort, the Global Harmonized Soil Database has become the primary resource for constraining global soil organic carbon stocks and fluxes (Batjes, 1996).

1.2.2 **Macrostrat**

The Macrostrat database integrates existing stratigraphic information by using surface polygons that extend vertically into the earth as stacks of lithostratigraphic and chronostratigraphic layers. More than 36,000 rock units have been added to the Macrostrat database from North America and New Zealand (Peters et al., 2014). Macrostrat puts signatures left by the interactions between biotic and abiotic processes into stratigraphic context, allowing them to be quantified in a space-time framework (Fan et al., 2015). These are just a small portion of the data that Macrostrat has integrated in its objective to produce a 4D model of the evolving Earth.

1.3 **Purpose**

There is an urgent need for large-scale, synthetic data to support the development of integrated Earth system models (ESMs) (Fan et al., 2015). The previously explained complications in geologic modeling are often made worse by the wide range of formats in which geologic data are stored. These data are dispersed and unstructured in the scientific literature, government archives, and myriad online web pages and repositories. Scales, standards and formats also vary (Fan et al., 2015). An example of this is standards employed by drilling companies; in Wisconsin, the practice is to implement the latitude/longitude coordinate system

for the position of the well on the earth's surface, while in Michigan, the township-range coordinate system is used. Complications such as these lead to inconsistencies among parameters in geologic data, which promote difficulties in both small and large-scale geologic modeling. The next generation of geologic mapping requires higher coordination in dealing with multiple datatypes, and more quantitative treatment of geologic subdivisions. In order to ameliorate the complications associated with geological data integration and modeling, this thesis used a time-based approach to integrate multiple different data sources at a sub-continental scale. This time-based structure provides a large-scale framework into which sub-units and depositional sequences can be constructed. In comparison with the DigitalCrust Vision, I envision a 4D space-time (xyz-t) data infrastructure designed to accommodate the structure and properties of the upper crust with use of age deposition as a central organizing concept in order to integrate disparate datasets (Fan et al., 2015).

1.4 Sample Location

In many geologic settings, post-depositional deformation takes place along a single dimension of extension or shortening. In these cases, a model for the subsurface can often be accurately represented by one or a few cross-sections striking in the direction of deformation. That said, as computer-based mapping gains prominence, developing truly 3-D geologic pictures of the subsurface has become more feasible. To demonstrate a 3-D mapping approach on an area that is not easily amenable to cross-section based mapping, the Michigan Basin was chosen as a sample location because of its 3-D geologic basin shape. "The Michigan basin is a nearly circular, intracratonic basin 400 km in diameter and 5 km deep with only minor structural disruption" (Howell, 1999). The benefit of an area this large is that this method of data

integration can be tested simultaneously on a large scale by spreading out data points as well as on a small scale by data clustering. For example, the Geological Map of North America provided locations of surface bedrock formations across the entire area of the Michigan Basin, allowing the integration quality to be tested over large distances. However, by implementing other databases such as driller's logs, boreholes nearby one another could be included in the integration process. This allowed integration quality to be tested among data points on a small scale.

2 Available Data Sources

Several geological databases provided the data integrated in the construction of this model. One database was the Geological Map of North America (GMNA), which is a collection of geospatial files, map images and publication documentation that support the geological map of North America (USGS n.d.). From this database, several thousand geological data points were acquired, which were located on the surface of North America and parts of Canada. Each data point had a latitude and longitude coordinate as well as a depositional age range depending on the formation from which that point was extracted. The GMNA database provided a majority of the data points found at the Earth's surface.

An analysis of drillers logs from the Michigan Core Repository provided data points with depth. Each core had a township and range coordinate with measured thicknesses of each geological formation spanning the length of the core. Similar data were collected from the Wisconsin Core Repository. Two cores from the Milwaukee area were analyzed in person for formation thicknesses and fracture characterization, and several Optical Borehole Image (OBI) logs from Wisconsin were also studied for the same purposes. A third dataset provided by Dr. Jean Bahr included subsurface data of formations extending stratigraphically upward from the Prairie du Chien to the Utica Shale. An analysis of 30 cores and 74 well logs across the Michigan basin provided a series of data points with an x and y coordinate in an arbitrary coordinate system and a depth in feet (Nadon et al., 2000).

2.1 GMNA Data

The Geological Map of North America (GMNA) provided the locations of contacts between different geologic units (at the bedrock surface) across North America, providing an estimate of where geologic contacts roughly intersect the land surface. The GMNA database contained thousands of data points, each with a latitude and longitude coordinate and a depositional age. The age was given as a range in millions of years, therefore, each point had two ages associated with it, one for the beginning of deposition of the overlying unit and one for the end of deposition of the underlying unit. The depositional age associated with each point represented the age of the first bedrock formation one would find if a borehole was drilled straight downward at that latitude/longitude position (i.e. through any soil, glacial till or other surficial deposits). As a result, the exact elevation of the bedrock was unknown. An initial assumption made in data processing was that the location of each data point provided by GMNA was located at the Earth's surface. This assumption would be altered during the integration process using the "Map Showing the Thickness and Character of Quaternary Sediments in the Glaciated United States East of the Rocky Mountains" by David R. Soller and an average of glacial drift thicknesses from driller's logs in order to improve model accuracy.

2.2 Driller's Logs

A spreadsheet containing the driller's logs of roughly 1,500 cores from several counties in Michigan was provided by the Michigan Core Repository. The driller's logs provided a location of the well at the earth's surface in the form of township-range coordinates and exact thicknesses of the lithologic formations the core intersected (Figure 3 & Figure 4). The thickness of each formation (or depth to each contact) was given in units of feet. The importance of the

driller's log data in Michigan was that they provided contact data with depth into the earth, while the GMNA provided contact data at the surface.

Many of the driller's logs used in this project separated the formations by geologic period. For example, all of the formations that the core intersected in the Mississippian period were listed under the heading of "Mississippian", followed by all of those in the Devonian period and so on. An issue that arose from this system is that formations were occasionally placed in the wrong period, thus giving an inaccurate description of the subsurface at that location. For example, the Antrim formation is part of the Devonian period, but in some cases, it was placed at the end of the Mississippian in the log. This could have been due to different drilling company's geologic history standards, or due to the individual examining the lithologic thicknesses. In any case, this lack of uniformity further shows a need for more cooperatively integrated geologic datatypes. An example of two variations of drillers log formats from Michigan can be viewed in Figure 3 and Figure 4.

Page 1-32 7-17N-4W Exploratory (00)
Grant Twp. (Clare Co.) Dry Hole
TD 11864 in Trempealeau (61)

JEM Petroleum Corporation

Permit No. 34611

Weingartz 1-7

Location: NE¼ NE¼ NE¼ Section 7, T17N, R4W
664 feet from north and 330 feet from east line of quarter section.

This well record assembly is a compilation of well data filed by or on behalf of the operator, as required by Act 61, P.A. 1939, as amended; formation tops and stratigraphic intervals determined or confirmed where possible by Geological Survey Division personnel from mechanicial logs and preliminary information collected while drilling was in progress. Records compiled by Randall L. Milstein, Geologist.

ELEVATIONS: 1003.9 Kelly Bushing (use for datum), 1002.6 Rig Floor, 990.8 Ground

Formation tops and stratigraphic intervals determined or confirmed by Geological Survey Division personnel from Schlumberger Compensated Neutron/Formation Density logs and records filed by operator.

PLEISTOCENE:		SILURIAN:	
Drift to	496	Bass Islands	5902
PENNSYLVANIAN:		Salina G Unit	6290
Saginaw Formation	496	F Evaporites	6322
MISSISSIPPIAN:		E Unit	7134
Bayport Limestone	1003	D Evaporite	7270
Michigan Formation	1143	C Unit	7319
Triple Gypsum	1241	B Unit	7432
Brown Lime	1318	B Evaporite	7486
Stray Sandstone	1465	A2 Carbonate	7852
Marshall	1528	A2 Evaporite	7991
Coldwater	1690	A1 Carbonate	8385
Coldwater Red Rock	2643	A1 Evaporite	8439
Sunbury	2651	Niagaran	8835
Bedford	2672	Clinton	8954
DEVONIAN:		Cabot Head Shale	9076
Antrim	2790	Manitoulin Dolomite	9160
Traverse Formation	3201	ORDOVICIAN:	
Traverse Limestone	3262	Cincinnatian	9201
Bell Shale	3860	Utica Shale Group	9691
Dundee	3919	Trenton Group	9966
Detroit River	4170	Black River Group	10270
Amherstburg	5150	Glenwood	10590
Bois Blanc	5417	Prairie Du Chien Group	10629
		CAMBRIAN:	
		Trempealeau	11772

Preliminary or pre-reported formation tops and other well information obtained from Geological Survey Division field reports compiled while drilling was in progress.

Triple Gyspum	1238 Sj.	Traverse Limestone	3255 Sj.
Stray Sandstone	1480 Sj.	Bell	3850 Sj.
Coldwater	1638 Sj.	Dundee	3913 Sj.
Sunbury	2646 Sj.	Detroit River Anhydrite	4170 Sj.
Antrim	2991 Sj.	Amherstburg	5203 Sj.
Traverse Formation	3222 Sj.	Bois Blanc	5645 Sj.
		Bass Islands	5902 Sj.

Figure 3: Example of Michigan driller's log.

18-17N-6W TD 5124 in Detroit River (34)
Garfield Twp., (Clare County) (Dry)

The Pure Oil Co.
A. E. Mulder No. 1 Permit #12376

Drilling Contractor: Gordon Oil Co. (Rotary)

Location: C N½ SE¼ SW¼ section 18, T.17N.,R.6W.
990' from south and 660' from east line of quarter section.

Elevation: 1059.5 feet above sea level.

Record by: Rex Grant from sample log, submitted by Pure Oil Co.

	Thickness (feet)	Depth (feet)
PLEISTOCENE:		
Drift:		
Sand, white	76	76
Sand and gravel	554	630
PENNSYLVANIAN:		
Saginaw:		
Sand, gray, coarse angular	57	687
Shale, black	11	698
Sand, gray, coarse, angular	27	725
Shale, black	50	775
Dolomite, dark brown, shaly	10	785
Sand, gray, medium fine, angular	35	820
Shale, black and dark gray	10	830
Sand, gray medium to angular coarse	170	1000
MISSISSIPPIAN:		
Bayport-Michigan:		
Limestone, light brown, dense	10	1010
Dolomite, brown, dense	5	1015
Shale, gray	5	1020
Shale, gray, gray-green - sandy	25	1045
Sand, white and gray, medium grained	65	1110
Shale, black	6	1116
Dolomite, brown, argillaceous	4	1120
Sand, gray, medium fine	5	1125
Shale, gray-green; shale gray	17	1142
Sand, white to gray, coarse	13	1155
Dolomite, brown, dense, shaly	10	1165
Shale, gray, trace of anhydrite	25	1190
Sand, gray, medium coarse	25	1215
Dolomite, brown, dense; anhydrite	10	1225
Shale, gray to dark gray; anhydrite	15	1240
Sand, gray, medium fine	10	1250
Shale, gray to dark gray; anhydrite	75	1325
Dolomite, brown dense; anhydrite	10	1335
Shale, gray to dark gray; anhydrite	35	1370
Limestone, dark gray, cherty; anhydrite	5	1375

Figure 4: Example of Michigan driller's log.

2.3 Michigan Basin Oil Exploratory Data

An exploration of the Ordovician aged formations for gas reservoirs in the Michigan Basin provided a third dataset for integration (Nadon et al., 2000). The stratigraphic order of these formations is described The contact at the top of the Prairie du Chien and base of the St. Peter marks the end of the Early Ordovician and beginning of the Middle Ordovician period with an age of 470 Ma. The data acquired from this exploration provided a detailed stratigraphic analysis based on 30 cores and 74 well logs distributed throughout Michigan (Figure 5). The paper analyzing these data differentiated 9 formation contact surfaces based off of their typical well log responses to gamma-ray, Photoelectric Factor (PEF), density and porosity (Nadon et al., 2000). Similarly, I was given 9 DAT files, which corresponded to these 9 contact horizons. Each of the 9 DAT files contained an arbitrary x and y coordinate for the well being observed and a depth in feet to the interface of the lithofacies. If each point from a single DAT file was plotted in MATLAB, the result would be a 3D scatter plot of the top surface (end of deposition) of that file's formation. Thus, this dataset provided a surface for 9 differentiated formations from within the range of the Prairie du Chien formation to the Utica Shale in the Michigan Basin.

Table 1: Michigan Basin Exploratory Data DAT Files and Corresponding Formations

DAT File Name	Corresponding Formation
UTICAS	Utica
TRENS	Trenton
BLRVS	Black River
GLWDS	Glenwood
TRANSS	Transition Shale btw St. Peter & Glenwood
SPS	St. Peter
BRAZOS	Brazos
ONEOTAS	Oneota
SHAKS	Shakopee

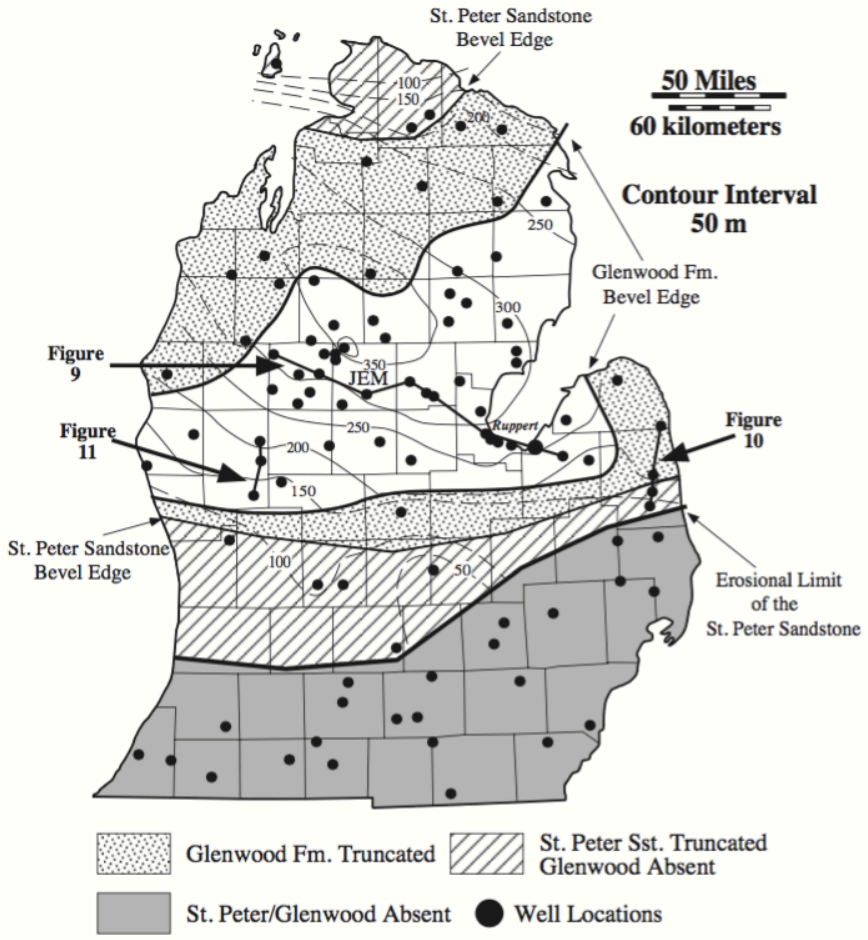


Figure 5: Distribution of wells used in exploration of the St. Peter/Glenwood (Nadon et al., 2000).

3 Interpretation of Data

3.1 Getting data into Integrated Format

3.1.1 Integration Format

Data for this project came from several different sources, each implementing differing coordinate systems and units. Two important goals were established for integrating the data. One was to find the most effective parameter by which these varying data could be connected, while the second was to organize the data into a form such that they could easily be input into software such as SGeMS and MATLAB for geostatistical analysis. This format would be organized in such a way that each data point would have an x, y and z-coordinate with consistent units and a single age value. To keep units consistent across all dimensions the UTM coordinate system was implemented for the x and y dimensions, while the z dimension was expressed in terms of meters (positive above mean sea level, negative below).

3.1.2 Integration Scale

The most effective parameter for integrating the various data was geological age. However, in geology, the concept of time applies to several different sub-categories, from eras and periods to specific formations and lithologies. An attempt was made to connect these data via geologic formations, but formation names can vary over state lines and individually by geologists depending on their own interpretations. The depositional ages of formations vary under these same circumstances, where one geologist may differentiate between two lithologies, another may consider them the same and group them together. The concept of time at the formation scale can vary significantly, propagating error to further phases of analysis. Differentiating time at a larger

scale, such as geological period, was determined to be the most effective gauge while integrating the data from the various sources.

3.1.3 **Integrating GMNA**

The GMNA data were imported into MATLAB, where the function deg2utm was used to convert all of the latitudes to northings and longitudes to eastings. The function takes an input matrix of latitudes in one column and longitudes in another and converts them into UTM northings and eastings respectively. The elevations were calculated for each point by inputting the latitudes and longitudes into GPSVisualizer (GPSvisualizer.com), which reports the elevation in feet of each location in Google Earth relative to mean sea level (MSL). The received elevations were then converted to meters via MATLAB by dividing their elevations in feet by 3.28. Each data point from the GMNA database now had its x, y and z-coordinates in terms of meters, but still had two age values associated with it instead of one.

In order to get each point in terms of one age value, all of the points corresponding to a chosen geologic contact were extracted. This was done in MATLAB by iterating through every point's parameters: easting, northing, and ages. A single point was determined to be on a contact if its coordinates were associated with two different depositional age ranges. The youngest age value of the lower formation would be equal to the oldest age value of the upper formation. The data point on that contact was set equal to the shared value between the two formation's depositional ages. For example, in Figure 6 "contact 1" lies on the contact between Lithological Unit 1 and Lithological Unit 2. Lithological Unit 1 began its deposition at 200 Ma. and ended its deposition at 100 Ma., while Lithological Unit 2 began its deposition at 300 Ma. and ended at 200 Ma. Therefore, if "contact 1" were a data point in this study, it would have an x, y and z

coordinate and its time parameter would be equal to 200 Ma. All data points that were not associated with a geological contact were erased from the dataset. All of the data points from the GMNA were now in the correct format, having an x, y and z coordinate location in meters and a geological age associated with the contact on which the point to which the point corresponds.

The final step taken to try to make the model as accurate as possible was to revise the assumption that the location of each data point provided by GMNA was at the earth's surface. In reality, a layer of Pleistocene unconsolidated material exists across the entire Michigan Basin that ranges from a few meters to several hundred meters depending on location. Therefore, it would lead to an incorrect interpretation if it was assumed the bedrock extended straight to the land surface. The thickness of the Pleistocene layer was recorded in the driller's logs, but a Pleistocene age thickness had to be applied to the GMNA data to maintain consistency in the depth to bedrock. The "Map Showing the Thickness and Character of Quaternary Sediments in the Glaciated United States East of the Rocky Mountains" by David R. Soller. (Soller 1998) gave a general depth to bedrock for the area of Michigan, but it was not digital and difficult to interpret. Reference to the map showed an average Quaternary sediment thickness of 244 m (800 ft) across most of the Michigan Basin area. However, the driller's logs showed only a few locations had glacial drift this thick. A more accurate approach seemed to be to take the average of the unit thicknesses produced by the driller's logs and apply that thickness to all the GMNA data points. The value given by Soller (244 m) was cut nearly in half, as the driller's logs showed an average thickness of 112 m. As a result, all of the GMNA data points in the basin area had 112 m subtracted from their ground elevation to represent a thickness of Pleistocene drift. When

the interpolation phase was performed, the value of 112 m provided smoother surfaces for each of the lithological contact ages.

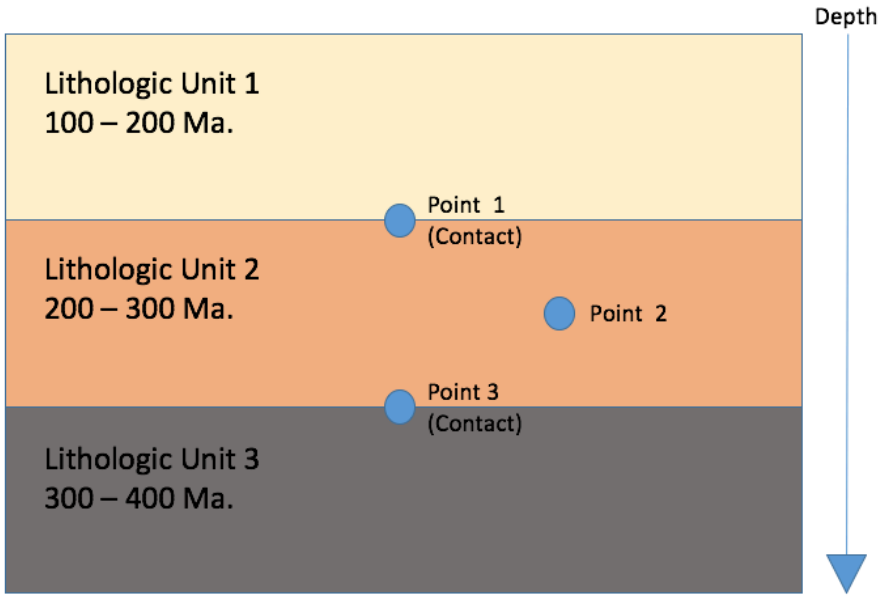


Figure 6: Ideal subsurface lithological sequence. Point 2 lies within Lithological Unit 2, rendering its information redundant since the bounds of the unit are known from Points 1 and 3.

3.1.4 **Integrating Driller's Logs**

The integration of the driller's logs required more labor than the integration of the GMNA data. The first step was to convert the coordinates of the well locations from the Michigan standard of township-range to UTM. The Public Land Survey System (PLSS) is a way of subdividing and describing land in the U.S. The PLSS typically divides land into 6-mile-square townships, which is the level of information included in the National Atlas. Townships are subdivided into 36 one-mile- square sections (USGS 2016). Sections can be further subdivided into quarter sections, quarter-quarter sections, or irregular government lots (Figure 7). The conversion to UTM was completed using the EarthPoint converter

(http://www.earthpoint.us/Townships.aspx), which allowed the user to input the state and county in which the well was located. This information was followed by inputting the township, range and quadrants (to the most well-known degree). The latitude/longitude and the UTM coordinates of the well were then provided by the converter. While some of the driller's logs gave the position of the wells down to the distance in feet from the quadrant boundaries, the most specific degree to which a location could be entered into the converter was up to the 3rd ¼ quadrant. As a result, the exact location of the wells was estimated, but was considered an acceptable degree of estimation for the purpose of contact surface modeling of large-scale integrated data.

Now that each well had its x and y coordinates in UTM, the next step was to give each point its appropriate depth and age value. Since driller's logs were presented in a variety of formats and the associated PDF files were simply scanned documents, each elevation contact had to be manually read from the associated driller's log file. Each individual well provided several data points since one core with a specified easting and northing intersected several contact surfaces with depth. The driller's log for each well provided its elevation relative to mean sea level in feet, which were converted to meters. This elevation (where the earth's surface began) was the first contact, and was given an age value equivalent to the age of the rock formation at the surface. Since there was a layer of Pleistocene drift recorded at the top of each borehole, the age given to the data points located at the earth's surface from each driller's log was the end of deposition of the Pleistocene (0.126 Ma.). The remaining data points were given the eastings and northings of their corresponding boreholes, and the elevation (converted to meters) and age of the contacts that the borehole intersected in the subsurface. As a result, 170 data points were

recovered from 41 sampled driller's logs. Since the data were now in integration-ready format, they could be appended to the GMNA data.

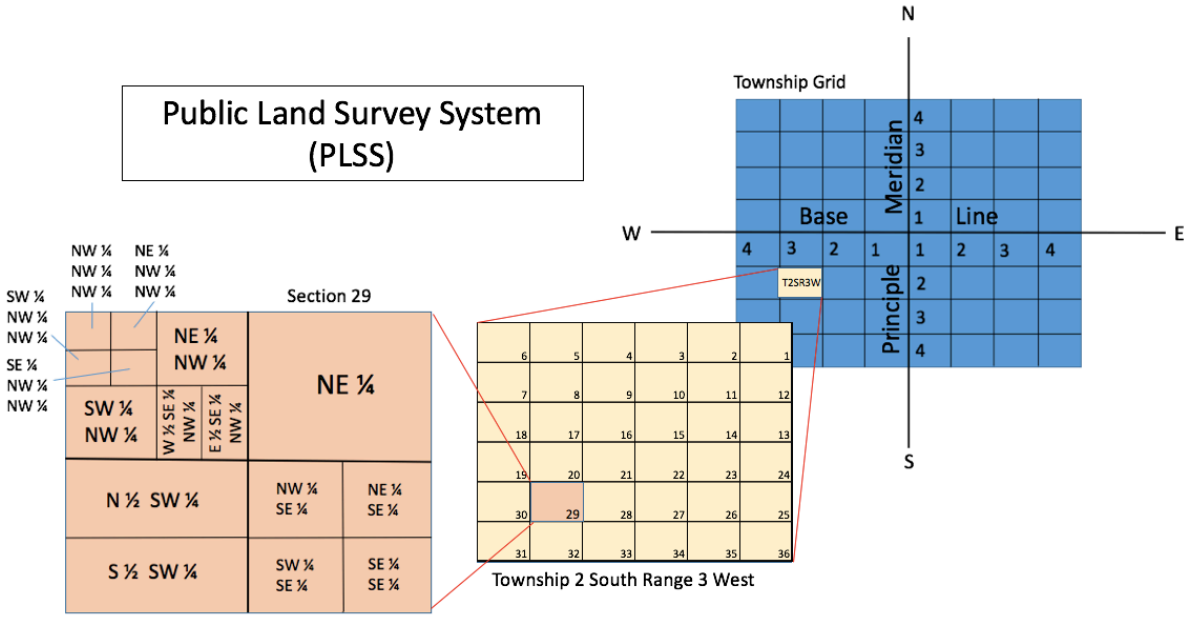


Figure 7: Division of townships via the Public Land Survey System (PLSS)

3.1.5 **Integrating Michigan Basin Oil Exploratory Data**

The data provided by the exploration of the Ordovician aged formations were recovered from 30 cores and 74 well logs. The St. Peter marks the first formation of the Middle Ordovician, a date widely agreed upon as being 470 Ma. The best way to make use of these data was to treat the contact between the Early and Middle Ordovician as its own surface, since none of the formations laid on a major temporal contact. To maintain consistency, this implementation was applied to the GMNA and drillers log data as well. Any known data points along the contact of the Praire du Chien and St. Peter formations (the Praire du Chien being the last formation at the end of the Early Ordovician and the St. Peter being the first of the Middle Ordovician) were given an age of 470 Ma and added to the dataset.

The points from the DAT file SHAKSS.DAT (representing the Shakopee formation) were saved for integration while the rest were deemed unusable. This is because the Shakopee formation is the last formation in the Prairie du Chien group and so the data points from this DAT file represent the contact between the Early and Middle Ordovician. This removed 1178 data points from this dataset, reducing the number of relevant data from 1220 points to 42. Though it was unfortunate to eliminate so many data points, their information was redundant by not defining the bounds of a lithologic contact.

Now that the relevant data were narrowed, they had to have their x and y coordinates converted into UTM and their elevation units from feet to meters. The former was a complicated task, since the original locations of the wells had arbitrary x and y coordinates when they were imported and plotted in MATLAB, ranging from -60 to 80 in the x-direction and -100 to 80 in the y-direction (Figure 8). Several steps were taken in order to find the correct eastings and northings for each well. First, Figure 5 was overlain on Google Earth as close as possible overtop of Michigan. Next, Figure 8 was overlain on top of Figure 5 with increased transparency so that the plotted wells from MATLAB aligned with the actual locations of the wells given by the Figure 5 overlay.

Once the overlays were in place, units of meters were applied to the arbitrary coordinate system. This was done first in the x-direction by finding a well located at the x-position 0 on the arbitrary coordinate system P1(x) (Figure 9). A well was then found at x=20 P2(x) (Figure 9). Then, using the "path" tab in the Google Earth ruler tool, the distance between the two points in meters and the azimuth of the line were calculated (Table 2). A 90-degree triangle was drawn with the height of the triangle vertically parallel to the y-axis of the arbitrary unit overlay (at

x=0) and the base of the triangle horizontally parallel to the x-axis (beginning at the well located at x=20 and moving horizontally back to x=0). With this information, simple trigonometry was used to find the length of the base of the triangle, which converted the length of 20 arbitrary units into meters in the x-direction (Equation 1). Thus, 20 units on the arbitrary scale equaled 54734.09 m using the well coordinates described in Table 2. This value was rounded up to 55000 m because P1(x) and P2(x) were not directly located at x=0 and x=20 respectively and overlaying the two images onto Google Earth could not provide a perfect reciprocation of the location of the chosen wells. Because of this potential for error, rounding the value to 55000 for simplicity was justified.

This same process was performed again, but to convert 20 arbitrary units in the y-direction into meters. P1(y) and P2(y) were two wells closest to y=0 and y=20 respectively on the arbitrary coordinate system (Figure 9). The distance between the two points was 55103 m with an azimuth of 23.95 degrees (Table 2). In order to convert the units in the y-direction into meters, these values were plugged into Equation 2, which revealed that 20 arbitrary units in the y-direction were equivalent to 55103 m. For the same potential reasons for error that persisted in converting the x-coordinates, this value was also assumed to be 55000 m.

$$x = \sin(\theta) * c$$

Equation 1

$$y = \cos(\theta) * c$$

Equation 2

Now that a conversion was calculated, the next step was to convert the x-coordinate of each well into real-world UTM eastings and the y-coordinates into northings so that the data could be integrated with the GMNA data and driller's logs (already in xyz-t format). This was done by locating the actual UTM coordinates, via Google Earth, of the origin (0,0) of the overlain arbitrary coordinate data (Figure 8) and then applying the conversion to each well with respect to this location. The origin in UTM coordinates was (659748.3, 4871950.2). For the x-coordinates, this was done by dividing the x-value of each well by 20 and then multiplying it by 55000, which converted their x-distances from the origin into meters. This distance was then added to the x-value of the origin's UTM coordinates, which gave each of the 42 points their rightful UTM easting (Equation 3). The same process was completed for the y-values, except their distances in meters were added to the y-value of the origin's UTM coordinates, giving each point its UTM northing (Equation 4). Any points which had a negative x-value in the arbitrary coordinate system were correctly placed west of the origin point, while positive values were to the east. Similarly, any points with negative y-values in the arbitrary system were placed south of the origin, while positive valued y-values were north. Although lengthy, this process successfully converted each of the data points from the Michigan Basin oil exploratory data sources into UTM coordinates.

$$Easting = 659748.3 + \left(\frac{\text{arbitrary x}-\text{coord}}{20}\right) * 55000$$

Equation 3

$$Northing = 4871950.2 + \left(\frac{\text{arbitrary y}-\text{coord}}{20}\right) * 55000$$

Equation 4

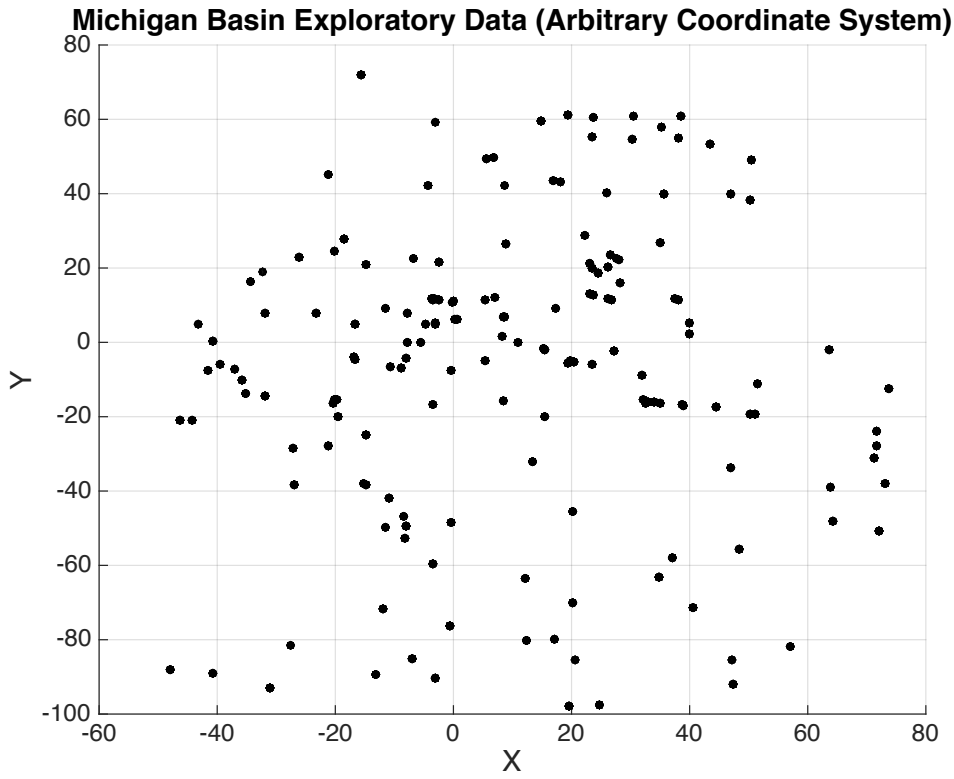


Figure 8: MATLAB scatter plot of well locations from St. Peter/Glenwood exploration emphasizing arbitrary coordinate system in x and y directions.

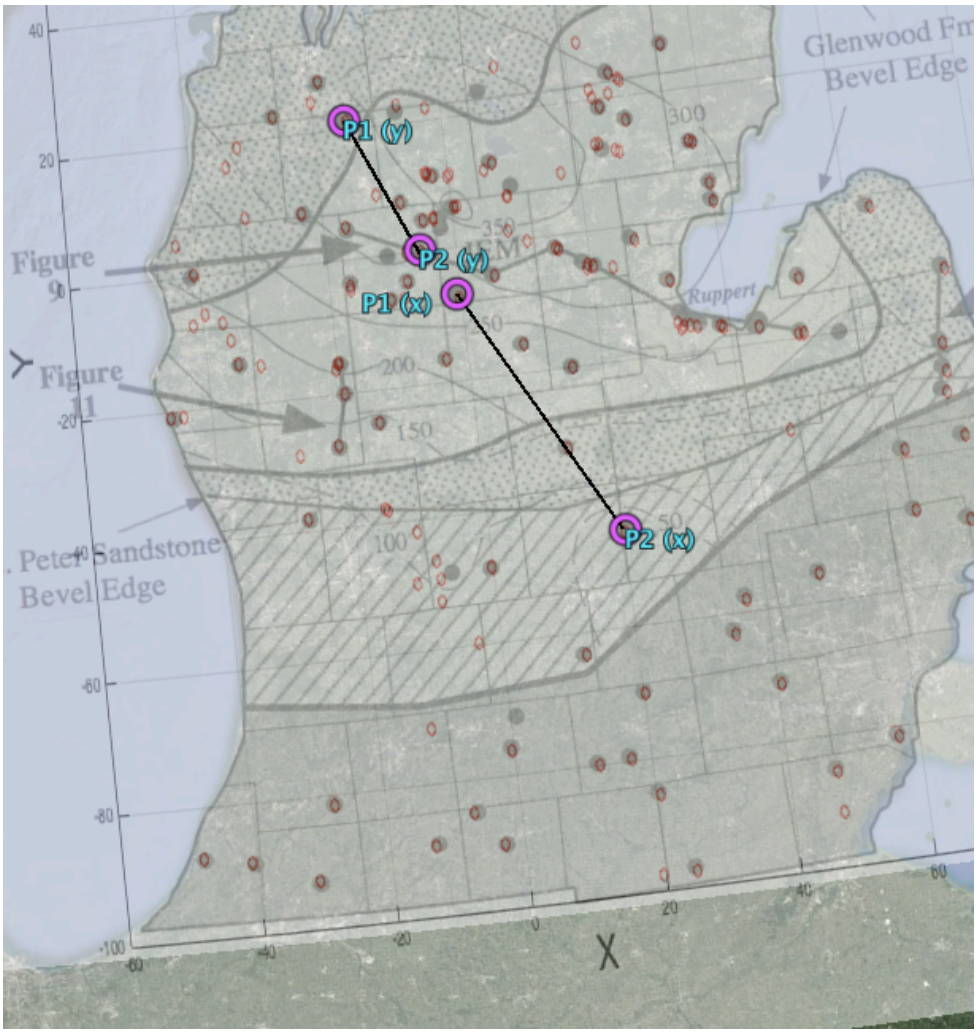


Figure 9: Overlay of Figure 5 (black) and Figure 8 (orange) onto Google Earth. Used to convert arbitrary coordinate system of St. Peter/Glenwood exploration into UTM. P1(x): well at arbitrary coordinate system x=0. P2(x): well at arbitrary coordinate system x=20. P1(y): well at arbitrary coordinate system y=20. P2(y): well at arbitrary coordinate system y=0.

Table 2: Values used to convert St. Peter & Glenwood data to meters.

Converting Arbitrary Units to Meters				
	P1(x)	P2(x)	P1(y)	P2(y)
X-coordinate (arbitrary)	-0.48	20.18	-14.74	-5.52
Y-coordinate (arbitrary)	-7.59	-45.65	21.08	-0.08
Distance btw Points (m)	111496.59		60294.37	
Azimuth (degrees)	29.4		23.95	
Equivalent of 20 arbitrary units in X-direction (m)	54734.59		N/A	
Equivalent of 20 arbitrary units in Y-direction (m)	N/A		55103	

The z-dimension of the Michigan Basin exporatory data were converted from feet to meters by dividing the depth in feet of each data point by 3.28 via MATLAB. Each of the 42 points also received an age value of 470 Ma. to represent the contact between the Early Ordovician and Middle Ordovician. The data from this source were now in the space-time (xyz-t) format and could be appended to the GMNA and driller's logs data.

3.2 Interpreting Surfaces

3.2.1 Why Contacts Make Useful Data

The space-time (xyz-t) format of the data integration was designed with the understanding that only data points lying on geological contacts were relevant for creating a 3D geologic model of the subsurface. Referring to Figure 6, for example, Point 1 represents a data point on the contact between the bottom of Lithological Unit 1 and the top of Lithological Unit 2.

Point 3 represents a contact on the bottom of Lithological Unit 2 and top of Lithological Unit 3. As a result, Points 1 and 3 provide top and bottom boundaries for Lithological Unit 2, thus encompassing Point 2, rendering the information about the location of Lithological Unit 2 redundant. Data points that lie within a geologic unit (Point 2) would be of use in modeling situations where a lack of contact data points existed. They would not perfectly define the surface boundaries of the unit they were collected from, which would decrease model's accuracy, but would at least provide a known location of the lithological unit.

Another issue with the implementation of non-contact data points while using the xyz-t integration format is the ambiguity in age assignment. The argument could be made to give the point the age of the formation it lies within, however, deposition of lithological units can occur over the span of millions of years. A point located at the bottom of a formation (beginning of deposition) may vary severely by age compared to a point located at the top of the same formation (end of deposition). Conversely, a contact point has a single age that marks the definite beginning and end of a lithological unit (assuming no unconformities, faults etc. Refer to 4.3 for descriptions on how to handle implementation issues). Therefore, contact points provide more assurance in both the space and time parameters when defining the boundaries of geological units.

4 Kriging with SGeMS

The Stanford Geostatistical Modeling Software (SGeMS) is an open-source computer package for solving problems involving spatially related variables. It provides geostatistics practitioners with a user-friendly interface, an interactive 3-D visualization, and a wide selection of algorithms (http://sgems.sourceforge.net/). The kriging estimation algorithm from SGeMS was used to provide a statistical analysis on the integrated geological data. Before the kriging could take place, SGeMS required the calculation of a variogram to determine the spatial autocorrelation between data points and the creation of a simple xyz grid within which the kriging algorithm would be implemented. A variogram is a description of the correlated spatial variability of the data (Figure 11). The empirical semivariogram is a set of points calculated using a measure of variability between pairs of points at various distances (Deutsch and Journal 1998) (Equation 5). The empirical semivariogram is created by searching the data in the domain for spatial autocorrelation by applying the variogram parameters input by the user to the empirical semivariogram equation (Figure 10).

In the process of kriging, the empirical semivariogram must be replaced by a semivariogram model, developed based on a family of simple mathematical correlation functions. The reason for this is partly because the kriging algorithm will need to access semivariogram values for lag distances other than those used in the empirical semivariogram (Bohling 2005). More importantly, the semivariogram models used in the kriging process need to obey certain numerical properties in order for the kriging equations to be solvable (Bohling 2005). As a result, there are several semivariogram models to choose from to best fit the empirical semivariogram, such as the spherical model (Equation 6) and the exponential model

(Equation 7). The spherical model reaches the specified sill value, *c*, at the specified range, *a*. The exponential approaches the sill asymptotically, with *a* representing a "practical range," which is the distance at which the semivariance reaches 95% of the sill value. Since empirical variograms can often be quite variable, especially when only a few data points fall within a given distance or "lag," quite a bit of judgement goes into selecting a good model.

Kriging itself is a special case of optimal linear prediction applied to random processes in space or random fields (Stein 1999). It implements the semivariance acquired from the variogram within the created grid and gives the best linear unbiased prediction of the intermediate values. Kriging relies on a neighborhood of points, moving further away from the prediction location, the measured values have less autocorrelation with the prediction value. These values can be eliminated from the calculation of that particular prediction by defining a search neighborhood. The specified shape of the neighborhood restricts how far and where to look for the measured values to be used in the prediction. The values defining the size and shape of the neighborhood are acquired from the semivariogram model ranges used to fit the empirical semivariogram (Figure 11).

Two different kriging methods were applied to the data to model the subsurface: an interpretation based off geologic age and an interpretation from contact elevation. In the first case, the data were imported all at once three dimensionally into SGeMS with the geologic age value of each point being the dependent parameter of interest. A benefit of this method was that the large amount of data points allowed for the calculation of a stable variogram, as each lag had sufficient data points for comparison. This method also produced a colored 3D kriged image of the subsurface in SGeMS (red older, blue younger), allowing for a more interactive and complete

visualization of the subsurface. A drawback to this method was that since the sample area was a basin and the grid surrounding the data was rectangular, the deeper corners of the grid had insufficient data points to be fully interpolated, which may have resulted in some inaccuracy in those regions. For the second method, the data were split into individual groups by age and then imported into SGeMS two-dimensionally (eastings and northings) one age at a time with the parameter of interest being elevation. The drawbacks to this method were that a new variogram had to be calculated for each dataset and the resulting image was only a 2D colored plane describing the elevation changes of the data within each age (red higher, blue lower). This process then had to be repeated for several different age surfaces. Further data post-processing was required in MATLAB to create a 3D mesh-grid surface of the elevation data.

For Equations 5-7:
u: vector of spatial coordinates
Z(u): variable under consideration as a function of spatial location
h: lag distance
N(h) = number of pairs separated by lag h
a: range
c: sill

Empirical: $$2\gamma(h) = \frac{1}{N(h)}\Sigma[Z(u) - Z(u+h)]^2$$

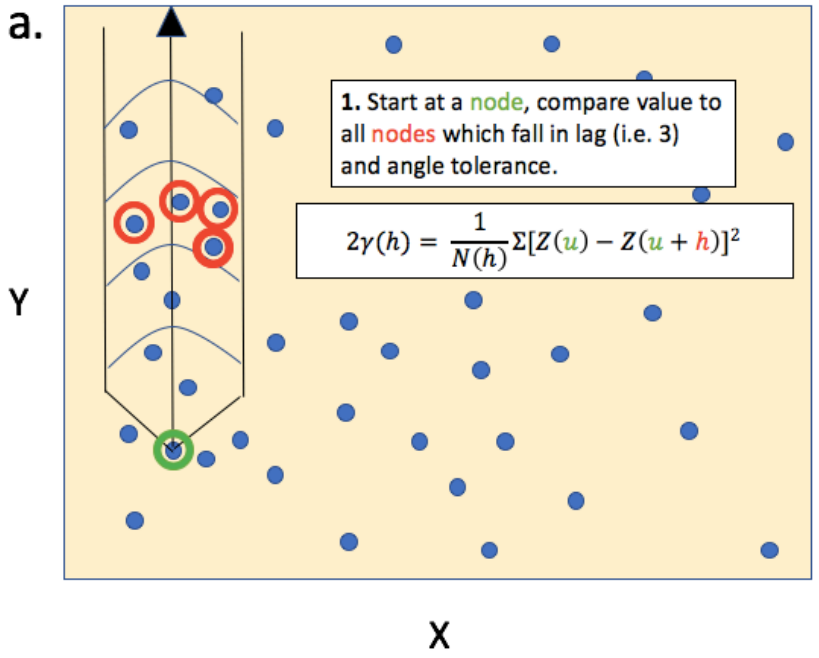
Equation 5

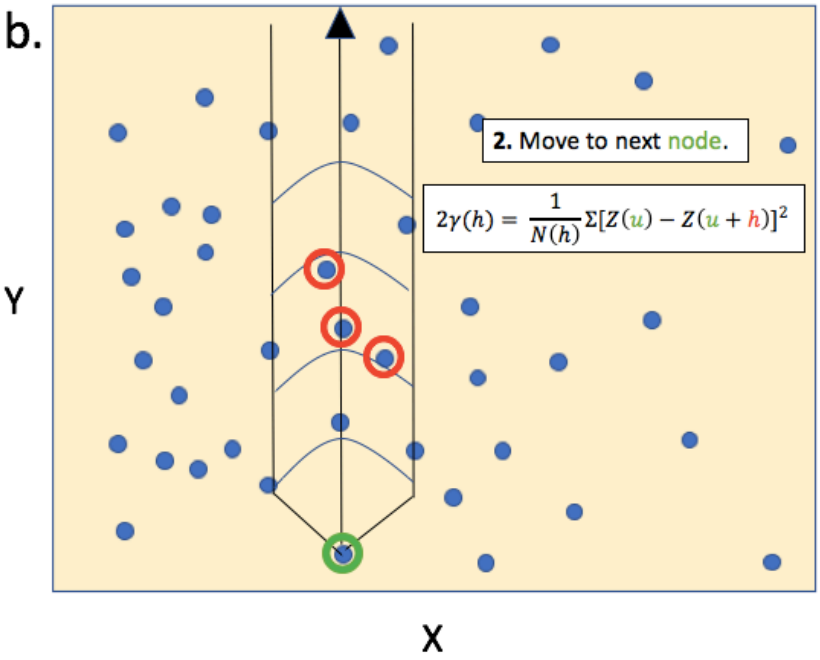
Spherical: $$\gamma(h) = \begin{cases} c * \left(1.5\left(\frac{h}{a}\right) - 0.5\left(\frac{h}{a}\right)^3\right) \; if \; h \leq a \\ c \; otherwise \end{cases}$$

Equation 6

Exponential: $$\gamma(h) = c * \left(1 - \exp\left(\frac{-3h}{a}\right)\right)$$

Equation 7


a.
1. Start at a node, compare value to all nodes which fall in lag (i.e. 3) and angle tolerance.
$2\gamma(h) = \frac{1}{N(h)}\Sigma[Z(u) - Z(u+h)]^2$
Y
X


b.
2. Move to next node.
$2\gamma(h) = \frac{1}{N(h)}\Sigma[Z(u) - Z(u+h)]^2$
Y
X

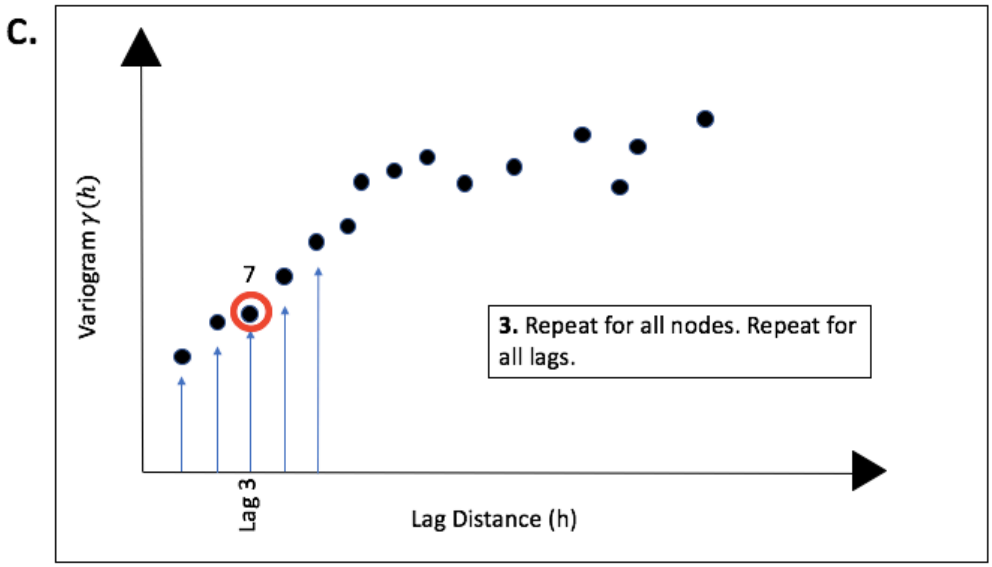


Figure 10: Schematic diagram of the variogram calculation. a) Example of variogram search parameters beginning at arbitrary node within example domain of data points. b) Moving to another point and performing search with same parameters. c) Plot of variogram vs distance adding together points captured within lag 3 from each node.

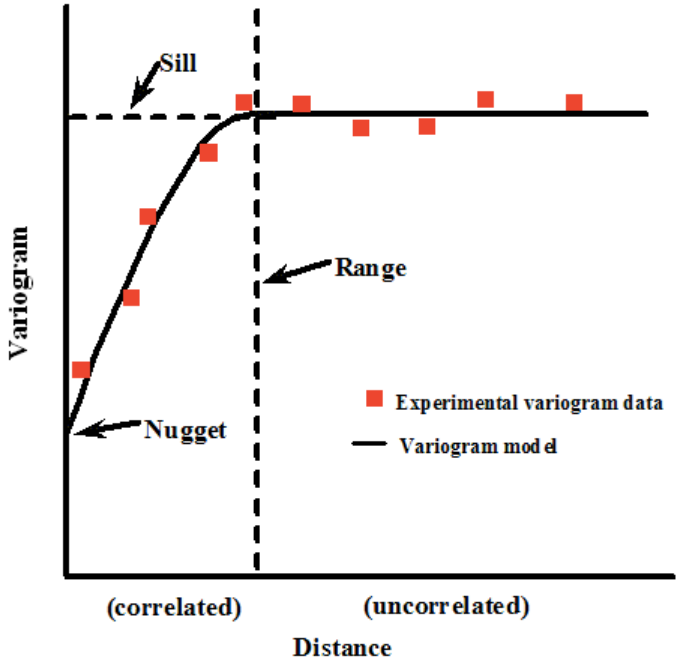


Figure 11: Example variogram with graphical explanation of parameters.
(http://vsp.pnnl.gov/help/index.htm#Vsample/Kriging_Variogram.htm)

4.1 Kriging with Respect to Age

While kriging the data with respect to age, the entirety of the integrated data were imported at once as a point system in the xyz-t format. A separate variogram was calculated in the x, y and z dimensions for spatial autocorrelation comparison among all of the data with respect to age and a 3-dimensional grid was created for the kriging process to occur within.

4.1.1 SGeMS Parameters (Age)

Several iterations of parameter value alterations were performed to acquire the best model possible from the kriging process in SGeMS. The total coverage of the plotted data in the x, y and z-directions was about 550,000, 700,000 and 4,600 meters respectively. In order to calculate a variogram that would capture an acceptable amount of data points per lag (at least 30), the number of lags chosen was 40, while the lag separation was 6,250 (Table 3). This provided a variogram coverage of 250,000 (the number of lags multiplied by lag separation), which fell in the ideal range of being about half the size of the total area of investigation. A variogram was then calculated with a 0-degree azimuth (comparing autocorrelation of points in the northern direction), a 90-degree azimuth (eastern direction) (Table 3) and a third variogram for the z dimension having a 0-degree azimuth but 90-degree dip (Table 4). The spatial correlation of the data points was expected to be more similar in the x and y dimensions than in the z dimension because changes in age could occur over the span of hundreds of thousands of meters laterally, whereas in the z-dimension the same changes in age had occurred over only about 4,000 m of depth. As a result, the variograms were expected to have similar shapes for the northern and eastern azimuths, but a different variogram with a shorter correlation length for the z-dimension variogram.

The variograms produced using these parameters were similar to the expected outcome. The spherical semivariogram model provided the best fit to the data, with x and y correlation lengths estimated as L_x = 100,000 m and L_y = 100,000 m, meaning points within the distance of L_x and L_y were spatially autocorrelated with respect to their geologic age (Figure 12). The variogram in the z-dimension had no recognizable correlation. An omnidirectional approach was attempted in the z-dimension (setting the tolerance within the direction parameters of the variogram to 90), but SGeMS produced the same plot as the single-directional approach.

The algorithm chosen to perform the interpolation was kriging with trend (KT). The dimensions used for the search ellipsoid are provided in Table 6. While kriging with a trend, a locally varying unknown mean could be modeled by a function of the coordinates of the points. This method was chosen in comparison to ordinary kriging because of its superior representation of the geologic subsurface. While performing ordinary kriging, value estimation at locations with fewer data points (near the perimeter of the basin) were given the value of the mean of the entire dataset. This produced an incorrect result, since the mean value was applied to locations stratigraphically below the oldest aged data points. Kriging with a trend allowed a local mean to be applied to these same zones, and thus produced an image more likely to be representative of the subsurface (Figure 13).

Table 3: Variogram parameters (x and y) for interpretation of geologic age.

Lag Parameters	Number of Lags	40
	Lag Seperation	6250
	Lag Tolerance	3000
Direction Parameters (1)	Azimuth	0
	Dip	0
	Tolerance	10
	Bandwidth	100000
Direction Parameters (2)	Azimuth	90
	Dip	0
	Tolerance	10
	Bandwidth	100000
Semivariogram Parameters	Sill	17514
	Max Range	107500
	Med Range	107500
	Min Range	0

Table 4: Variogram parameters (z) for interpretation of geologic age.

Lag Parameters	Number of Lags	22
	Lag Seperation	10
	Lag Tolerance	5
Direction Parameters	Azimuth	0
	Dip	90
	Tolerance	10
	Bandwidth	10000

Table 5: Grid dimensions for kriged interpretation of geologic age.

Number of Cells (x)	110
Number of Cells (y)	140
Number of Cells (z)	46
Size of Cells (x)	5000
Size of Cells (y)	5000
Size of Cells (z)	100
Origin (x) (Lower left corner)	400000
Origin (y) (Lower left corner)	4450000
Origin (z) (Lower left corner)	-4020

Table 6: Search ellipsoid dimensions for kriged interpretation of geologic age.

Directions	Maximum	400000
	Medium	400000
	Minimum	5000
Angles	Maximum	0
	Medium	0
	Minimum	0

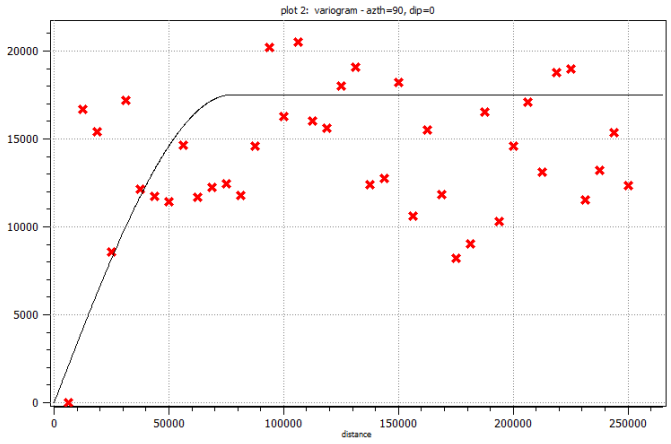


Figure 12: Variogram calculated while kriging via age method in SGeMS. Azimuth=90 (east), dip=0. X-axis: distance in meters. Y-axis: semivariance.

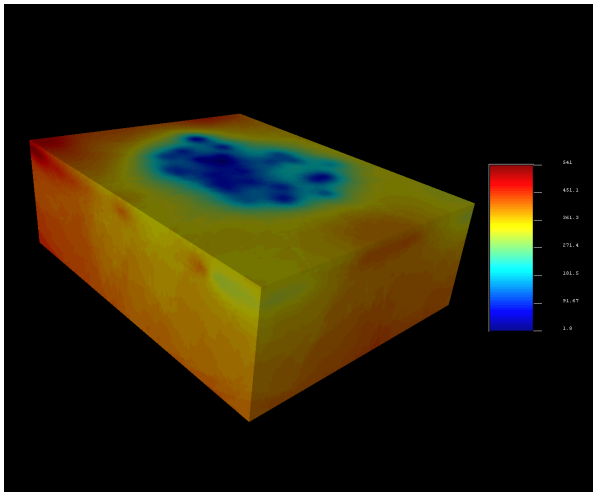


Figure 13: Kriging with Trend (KT) results of age parameter via SGeMS. Local mean applied to zones lacking data points. Color represents time (blue younger, red older).

4.2 Kriging with Respect to Elevation

Kriging with respect to elevation required a different process than kriging with respect to age. Instead of being imported in the xyz-t format, the data were imported as an xyz format. In MATLAB, the data were split up into different tables by age. The tables were organized with three columns: the x, y and z coordinate of each data point from that age (i.e. all points from the Silurian/Devonian were grouped and saved in one table). Each of the datasets were then imported individually into SGeMS, with the conditioning variable for kriging being elevation (z-value). This meant that the data were entered two-dimensionally as a point set (x and y values), variograms were calculated only in the x and y dimensions and the kriging was performed on a 2D grid, which estimated the elevation at unsampled locations for that age.

The downside to this method was that each dataset required its own semivariogram model so that it could be kriged accurately. The search parameters for the variograms were left the same for each dataset, but they each had different variances (sill values) and ranges to create the semivariogram (Table 7). A spherical semivariogram model produced the best fit to the data for each case, while still producing a reasonable kriged outcome. The Devonian/Silurian contact dataset contained the most data points, with about 130. This reduction in data points (compared the method of kriging with respect to age, where all the data was assessed at once) made the kriging process unstable in areas without sufficient data points, but produced a decent output in areas with higher data density (typically the center of the basin).

Table 7: Variogram Parameters for Kriging with Respect to Elevation

	Earth Surface/ Pleistocene	Jurassic/ Pennsylvanian	Pennsylvanian/ Mississippian	Mississippian/ Devonian	Devonian/ Silurian	Silurian/ Middle-Late Ordovician	Middle-Late/ Early Ordovician
# Lags	40						
Lag Separation	12000						
Lag Tolerance	6000						
Azimuth	0						
Dip	0						
Tolerance	90						
Bandwidth	100000						
Sill	2737.57	5993.71	7571.96	167676	31729.3	559341	649658
Max Range	139200	182400	187200	422400	268800	456000	336000
Med Range	139200	182400	187200	422400	268800	456000	336000
Min Range	0	0	0	0	0	0	0

4.3 Implementation Issues

This section aims to explain how some of the complications in geological modeling (described in 2.1.1) could be addressed.

4.3.1 Faults

Several methods could be implemented to handle faults within the model domain. One way is to think about the artificial insertion of a fault simply because the result of the fault is "blocks" of lithofacies not correlated from one block to another (Chilès et al., 2015). Figure 2b is an example of what this may look like. In this case, b1 and b2 are the same geologic unit, offset from one another by a normal fault. If there were enough data points to bound the dimensions of b1 and b2, the offset would be clear without the need to physically input a fault. A second technique is to consider faults as screens, which requires knowledge of the fault planes and also the zones of influence of the faults (Chilès et al., 2015). This method too could be implemented into the xyz-t format and applied to this model. With knowledge of the fault plane's location

(strike and dip) a series of data points could have their xyz coordinates placed along that plane and be given the age of the fault. Since the age of the fault must be younger than the units it offset, the plane would be differentiated from the rest of the data, effectively putting a fault into the model.

4.3.2 **Borehole Ends**

As previously stated, when the driller's logs from Michigan were analyzed, only the intersections of the boreholes with the lithological interfaces were entered as data, despite the fact that the borehole also carries the information that all the points between two successive interfaces belong to the same horizon (Chilès et al., 2015). That additional information was redundant and could be disregarded. An exception arose with the end of a borehole, which did not coincide with an interface and gave information that the next interface was deeper than the borehole (Chilès et al., 2015). In order to maintain accuracy of lithological boundaries within the model, only units up to the oldest (deepest) interface were used. Units extending beyond the depth of the borehole were noted as being present, but unused in the analysis.

4.3.3 **Unconformities**

At the large time scale (mainly geologic periods) with which the geological units were integrated, the Michigan basin had no unconformities. However, if a model was created using the same xyz-t format to integrate data, an unconformity could be entered rather easily. With this format, no single point lying on the contact of an unconformity could be assigned two ages. To work around this, more data points could be added along the unconformity surface to account for the unconformity. For example, if the upper bound of a unit was known to be Ordovician in age, the Silurian eroded, and a unit Devonian in age laid in contact with the Ordovician, then a set of

data points could be entered with coordinates directly on top of the known Ordovician age data points, but given the Devonian age. This way, the xyz-t format would be unchanged and the unconformity would be preserved in the model.

5 MATLAB Processing

Once the kriging was completed in SGeMS, further statistical analysis was performed in MATLAB to test the comparison between both kriging methods. This was done by exporting the kriged data from SGeMS and importing them into MATLAB to create isosurfaces within the model domain at specified contact age values. The isosurface of the 3D kriging method with respect to age was then compared to the 2D kriging method with respect to elevation to see how similar or dissimilar the surfaces would be in space.

Exporting the data from SGeMS was best done by means of a GSLIB (Geostatistical Software Library) file, since SGeMS is most familiar with this file format. When the data were exported for the age kriging method, the file contained a single column of age values 708400 rows long, representing the age of each individual cell from the 3D grid of dimensions 110 x 140 x 46. Each kriged age value in the file corresponded with its cell position within the grid. After the data were imported into MATLAB, the function "reshape" was used to reorient the age values back into their 3D matrix form.

Contours and isosurfaces could then be constructed based on specified age values within the matrix. The MATLAB isosurface function iterates through the selected matrix and finds every value equal to or closely equal to the input value and draws a surface along those values. For example, an isosurface could be drawn for all values within the matrix equal to 323.2 Ma. (Pennsylvanian/Mississippian contact) (Figure 14). Multiple isosurfaces could be added to the image at once, consider the addition of the Jurassic/Pennsylvanian contact isosurface (298.9 Ma.)

above the Pennsylvanian/Mississippian and the Mississippian/Devonian (358.9 Ma.) contact surface below (Figure 15).

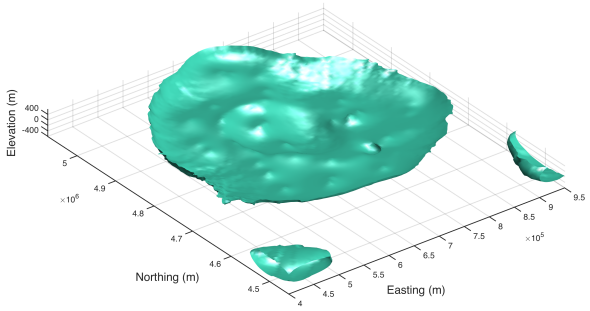


Figure 14: Isosurface of Pennsylvanian/Mississippian contact, extrapolated from the kriging via age method.

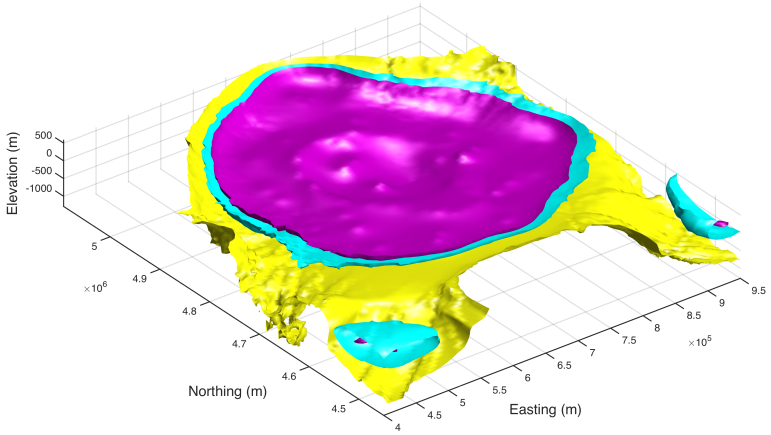


Figure 15: Isosurface of Jurassic/Pennsylvanian contact (purple), Pennsylvanian/Mississippian contact (light blue) and Mississippian/Devonian (yellow) extrapolated from the kriging via age method.

The process of importing the data from the kriging method performed with respect to elevation was similar, yet more time consuming than importing the data from the method of kriging with respect to age. This is because instead of having one matrix with all of the data, a separate GSLIB file had to be exported for each contact age that was kriged. For example, the kriged 2D plane of the Jurassic/Pennsylvanian contact required its own file, and the Pennsylvanian/Mississippian contact surface required an individual file and so on. The MATLAB reshape function was used to reorient the data from the GSLIB file, in two-dimensions. The data were now in matrix format, with an elevation value at each xy location. The MATLAB function "mesh" was used to create a 3D surface from the 2D matrix. Since each contact surface was kriged individually in this method, multiple kriged elevation surfaces of different ages were plotted against one another to check for inconsistencies in the data such as overlapping or intersecting portions (Figure 16).

The surfaces drawn using the two different method approaches were then plotted against one another on the same figure for comparison. A specific contact age was chosen (the Mississippian/Devonian) and the surface at that age was plotted from the age kriging method, then overlaid with the same surface drawn from the elevation kriging method (Figure 17 & Figure 18).

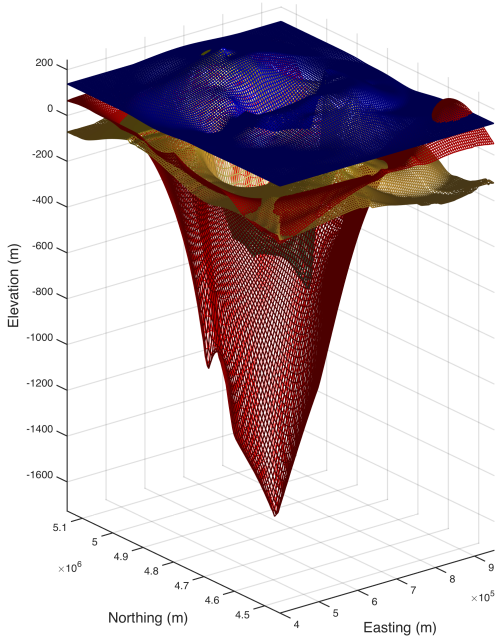


Figure 16: Contact surfaces created from kriging with respect to elevation plotted for comparison. Pennsylvanian/Mississippian contact surface (blue). Mississippian/Devonian contact surface (tan). Devonian/Silurian contact surface (red). Vertical exaggeration 500/1.

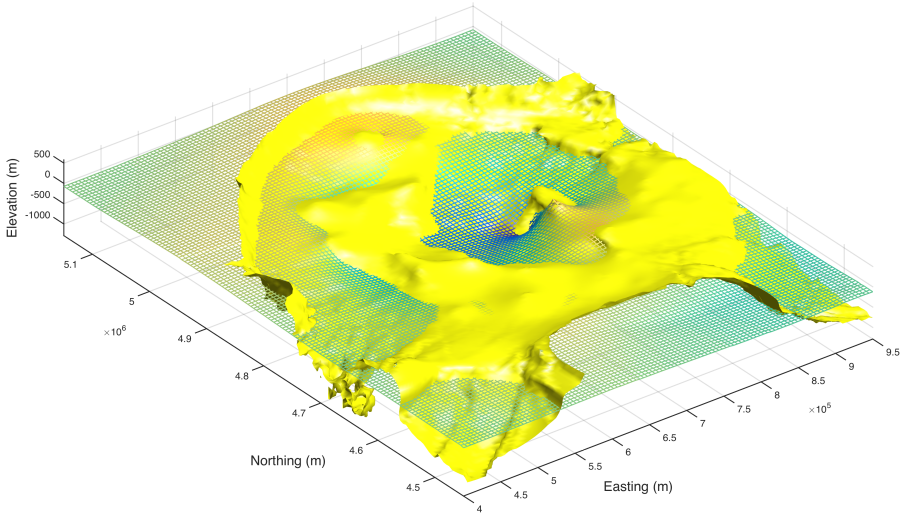


Figure 17: Mississippian/Devonian contact surface. Isosurface from kriging with age (yellow). Mesh surface from kriging with elevation (mesh). Vertical exaggeration 50/1.

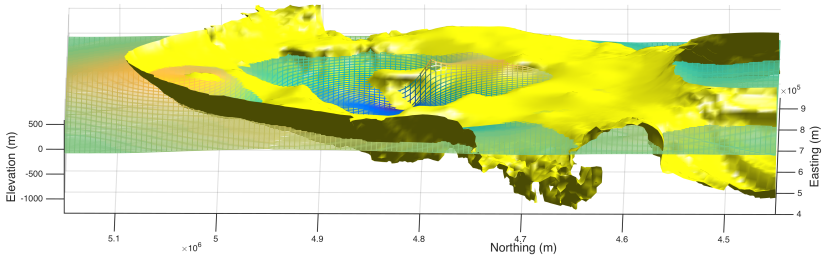


Figure 18: Mississippian/Devonian contact surface. Isosurface from kriging with age (yellow). Mesh surface from kriging with elevation (mesh). Facing east. Vertical exaggeration 50/1.

5.1 Comparison to Macrostrat Database

After the surfaces were extracted from the 3D grid in the method of kriging with respect to geologic age, the unit thicknesses were compared to those given in the Macrostrat database. Only the surfaces extracted from the age kriging method were compared since the method of kriging with respect to elevation showed zones of intersecting contacts, meaning it was a more unstable method. The Macrostrat database implements a system of columns, which could be thought of as being similar to boreholes, that provide an estimated minimum and maximum thickness for geologic formations at specific locations. Since a range of thicknesses were given for each formation in Macrostrat, the average between the minimum and maximum were used. Also, Macrostrat column 429 was used for comparison, as it is located nearly in the center of the Michigan Basin. The elevations of the some of the contact horizons that were compared are summarized in Table 8.

Table 8: Contact Surface Elevation Comparison: Macrostrat Database vs. Age Kriging

Contact Surface Depth				
Contact Surface	Kriging Age Method (m)	Macrostrat Average (m)	Macrostrat Min (m)	Macrostrat Max (m)
Earths Surface/ Pleistocene	334.9	334.9	334.9	334.9
Pennsylvanian/ Mississippian	-33.3	188.3	246.1418	100.516
Mississippian/ Devonian	-559.6	26.9	138.5023	-84.659
Devonian/ Silurian	-1500	-270.4	-16.3582	-524.5

6 Small Scale Data Integration

Fluid circulation in the Earth's crust plays an essential role in surface, near surface, and deep crustal processes. Flow pathways are driven by hydraulic gradients but controlled by material permeability and interconnectivity, which vary and change over time. Although millions of measurements of crustal properties have been made, from minuscule to macroscopic scales, this vast amount of data and information has not been integrated into a comprehensive knowledge system (Fan et al., 2015). The time-based method of data integration proposed in this thesis can provide a framework into which sub-units, depositional sequences, fracture characterizations and general geological interpretations can be structured.

While a majority of this thesis was spent on data integration and formatting to produce a plausible model of the subsurface, significant time was spent at the Wisconsin Core Repository performing fracture analysis on selected core samples. This analysis was performed to explore how fracture continuity can persist across an area and how within-unit analysis, such as fracture characterization could be inserted into a model with the xyz-t integration format.

6.1 Traditional vs. Digitized Data Analysis

Traditional geologic interpretations of the subsurface occur between boreholes, with a geologist connecting similar units between two boreholes at either end of a diagram (Figure 19). In some situations, more boreholes are added to create a fence diagram, consisting of several cross sections resulting in a quasi-3D description of the subsurface. If the information stored within these boreholes was digitized, it could be shared, edited and refined by the scientific community. Moreover, borehole data stored in a 4D (xyz-t) format could be stored and integrated

through the exact same process used to integrate the Michigan Basin data. Then, further parameterization could be supplied to specific lithologic units 3-dimensionally with information such as porosity values, fracture density or any other parameter. The issue is that most of these data are not digitized; even well-understood groundwater-surface water interactions in the top tens of meters of the crust are poorly represented in current Earth System Models (ESMs), and most processes at depths greater than 2-3 m within the subsurface are not included (Fan et al., 2015). Efforts to extend ESMs deeper into the crust have been hindered by deficiencies in subsurface data.

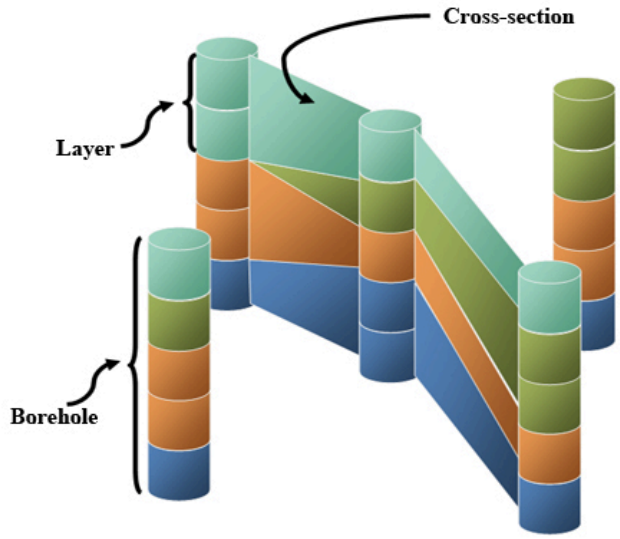


Figure 19: Example of cross-borehole geologic interpretation.
http://blog.simmakers.com/BLOG/zametki/zametka_7/Picture%201a.jpg

6.2 Wisconsin Core Repository

The Wisconsin Core Repository is an abundant source for geological data. However, despite the thousands of cores from a state-wide area available for hands-on analysis, much of the data are not digitized. For example, the closest the associated driller's logs come to digitization is a

PDF form (much like the driller's logs compiled from Michigan). While these data are useful, it is cumbersome and time-consuming to analyze. During the work in this thesis, two cores were analyzed carefully by hand over the span of two days.

The cores were chosen based primarily on their location (so as to contain formations and age ranges found in the Michigan Basin) and secondarily on continuity. Both cores chosen were from the Milwaukee area because they contained some Devonian and Silurian aged strata that extended from the Michigan Basin into eastern Wisconsin. Core 41000751 was 806 ft in total. To the southeast 4.7 miles was the location of core 41000757, which was 665 ft (Figure 20).

Figure 20: Locations of boreholes analyzed from Wisconsin Core Repository.

At the Wisconsin Core Repository in Mt. Horeb, WI, each core was stored in boxes 4 ft long and roughly 1.5 ft wide, with 4 columns of core in each box (Figure 21). The boxes were carried from the pallet to the observation table and analyzed box by box. In most cases, one

representative column was taken out of the box for deeper analysis, while any notable features seen in the other columns were noted (Figure 22). This was the most efficient way to gain the necessary data, given our limited reservation time to observe each core. The main objective in studying the cores was for fracture density analysis. Any noticeable features in the core, whether open or filled, sealed or broken, horizontal or angled, were noted based on their distance in feet from the base of the column that they were in. The bottom of each column was assumed to be foot zero, and the tops averaged 3.7 ft. Some angled or vertical fractures were given a length range based on how far along the column they extended. In determining whether a fracture was induced, by drilling, handling etc., a numerical system ranging from 1 to 5 was used. A fracture determined to be induced was given a 1, while those determined to be natural were given a 5. Numbers 2 through 4 were used if the original cause of the fracture was unknown. For example, if a fracture was thought to be natural, but was not absolutely sure, it was given a 4.

Figure 21: Core ID 41000751, core boxes stored on pallets.

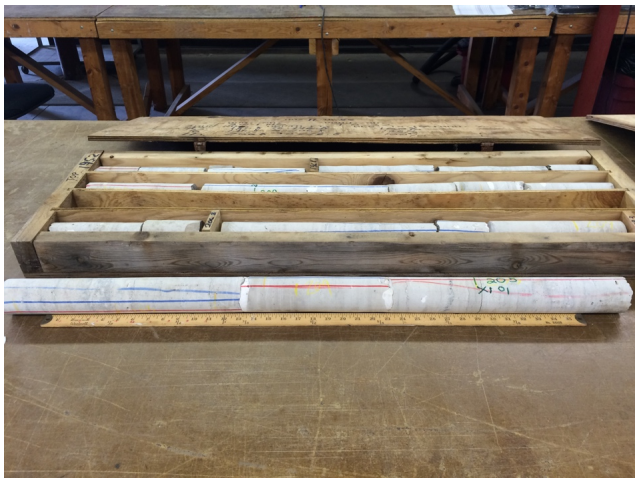


Figure 22: Core ID 41000751, example of analysis process choosing a representative column from each box for deeper analysis.

The result of this analysis was a detailed table for each core in stratigraphic order noting the location of each fracture (natural and induced) with respect to the core, the aperture width and fracture angle. From there, a more condensed table was created giving the number of fractures per geologic age, and more specifically by individual formation (Table 9 & Table **10**). Condensed tables containing relevant fracture characterization data were also created from the driller's logs of the two boreholes for comparison to our hands-on analysis (Table 11 & Table 12). Unfortunately, only two formations overlapped between the two cores, the Racine and the Waukesha from the Silurian, which left limited room for comparison, but still produced interesting results about fracture continuity within formations over a multi-mile scale.

Based off our analysis at the repository, the Racine formation at the location of borehole 41000751 had 5 natural joints, but 64 at the location of borehole 41000757. This may seem lopsided, but the top of core 41000751 began within the Racine formation and retained only about 40 ft of the unit, from 181.3 ft to 223.6 ft. Core 41000757, however, contained the entire

Racine formation from 330 ft to 496.6 ft, totaling 166.6 ft. It makes sense that more fractures would be found in the latter core, but the difference is still quite large. The high number of fractures noted in the Racine formation of core 41000757 could be because it was the very first portion of core analyzed and may have been over-scrutinized as our method of analysis was developing. If we observe the number of natural joints found within the Waukesha formation, a formation entirely covered by both cores, the number is exactly the same at 18. Since the thickness of the Waukesha formation was 46.5 ft in core 41000751 and 53.2 ft in core 41000757 and their fracture densities were the same, it proved that the continuity of fracture density can persist (within the same formation) on a scale up to 5 miles.

Information provided by the driller's logs also shows similarities among the cores. According to the driller's logs, the Racine formation share a very similar number of fractures between the two cores, with core 41000751 and 41000757 having 27 and 24 respectively. On a larger scale, the Silurian age in total had a similar number of joints recorded within each core, although it should be noted that different formations intersected each borehole. These data unfortunately barely fell within the range of the Michigan Basin domain and the driller's logs accumulated from Michigan did not supply a fracture density analysis. This further shows the need for trans-discipline, trans-community, and trans-agency collaboration for digitally stored geologic data. It took two full days of hands-on core analysis to digitize the data from cores 41000751 and 41000757, and with greater collaboration these data could be readily available for the next geologist to use. With more digitized data, the 4D xyz-t format of the Michigan Basin data could host the inclusion of finer-scale sub-units into the model.

Table 9: Core 41000751 Fracture Analysis

Age	Formation	Joints (No.) Natural	Joints (No.) Induced	Total Joints by Fmtn (#)	Total Joints by Age (#)
Silurian	Racine	5	16	21	116
	Waukesha	18	12	30	
	Undif Silurian Strata	25	40	65	
Ordovician	Maquoketa Shale	5	9	14	54
	Sinnipee	24	16	40	
Total:		77	93	170	170

Table 10: Core 41000757 Fracture Analysis

Age	Formation	Joints (No.) Natural	Joints (No.) Induced	Total Joints by Fmtn (#)	Total Joints by Age (#)
Devonian	Milwaukee	3	0	3	12
	Thiensville	9	0	9	
Silurian	Waubakee	27	1	28	155
	Racine	64	5	69	
	Waukesha	18	0	18	
	Undif Silurian Strata	39	1	40	
Total:		160	7	167	167

Table 11: Core 41000751 Driller's Log Fracture Data

Age	Formation	Total Joints by Fmtn (#)	Total Joints by Age (#)
Silurian	Racine	27	39
	Waukesha	7	
	Mayville	5	
Ordovician	Maquoketa	4	6
	Sinnipee Group	2	
Total:		45	45

Table 12: Core 41000757 Driller's Log Fracture Data

Age	Formation	Total Joints by Fmtn (#)	Total Joints by Age (#)
Devonian	Milwaukee	7	
	Thiensville	10	17
Silurian	Waubakee	16	
	Racine	24	40
Total:		57	57

7 Conclusions and Future Direction in Geologic Data Integration

The result from kriging three-dimensionally with respect to geologic age produced a more detailed, and presumably accurate, depiction of the geologic subsurface. Reference to Figure 18 shows a much more realistic contact surface for the case of kriging via age when compared to the mostly flat mesh grid portraying the same contact surface by the method of kriging via elevation. A major cause for the variation between the two images is the amount of data points that could be included in each method. The method of kriging the elevation values of each contact age could only use a fraction of the total data points, but the area over which the kriging was performed was the same as the method of kriging the age values. This can cause a less accurate experimental variogram model, which can then cause instability in the kriging process and likewise cause many kriged locations to be close to the overall mean of the data. Also, when the surfaces extrapolated from each method were compared to each other, the first method of kriging with respect to age produced no overlapping or intersecting surfaces (Figure 15). On the other hand, when the same contact surfaces were plotted against one another for the method of kriging with respect to elevation, the surfaces intersected at several locations, portraying an incorrect depiction of the Michigan Basin (Figure 16). For this reason, the first method of kriging performed (3D, with respect to contact age) produced a more detailed and accurate image of the subsurface.

There was more discrepancy with the elevations of contact horizons when deeper contacts were compared between the age kriging method and the Macrostrat database. This could be

because the average thickness of each lithological unit provided by Macrostrat was used instead of the maximum value. The difference between the minimum and maximum thickness estimates sometimes varied by hundreds of meters. As a result, the averaged values could have had significant error. Also, errors in estimating unit thicknesses closer to the surface would propagate downward into older unit thickness estimates adding greater error with depth. Also, the estimated unit thicknesses provided by Macrostrat are representative of data collected within the area around the column's location. The location of Macrostrat column 429 (used to compare to the kriged surfaces) is almost geographically in the center of the basin, where the contact surfaces are deepest. If the depths of the contact surfaces from the Macrostrat database were compared to the surfaces extracted from the age kriging method in locations outward from the center of the basin, where the contacts are shallower, the depths may have matched better.

The future direction of this research would include the analysis of more cores from within the Michigan Basin, in line with the methods performed in section 6.2. A hands-on analysis of cores from Michigan would add substantial data to the small-scale integration portion of this project, allowing geologic parameters from the analysis to be added as sub-units into the model. Cores would be chosen with a method similar to how the driller's logs were chosen in section 2.2. Ideally, they would be clustered within an area (the center of the basin, for example), with a few outlying locations for large-scale analysis. I would be interested to see how the fracture density of different formations extended across a wide area from borehole to borehole, as well the comparison of fracture persistence across Lake Michigan with boreholes in eastern Wisconsin. With the completion of the fracture analysis of several boreholes, data could begin to be added to the large-scale xyz-t format. At this point in the research, the contacts bounding

several geologic age ranges have been stored and plotted (although could be refined with the addition of more driller log information). If a fracture analysis were performed on an entire unit with a given age range, the value (or amount of fractures) could be stored as a sub-unit within a Silurian section, bound by the units contact surfaces.

A level of uncertainty also comes with the use of borehole data. Although they are some of the most reliable means of extracting geologic data from the earth, wellbore data can produce an element of "compressed" profiles of time. The extraction of cores most likely does not result in a full recovery and this is not accounted for in the driller's logs. This could mean, for example that 1 ft through the earth may be equal to only 0.9 ft of core recovery. As a result, the thickness of each formation may be larger in its location in the earth than the driller's logs present. This could have affected the accuracy of the methods used in this model by placing contact surfaces at a higher elevation in the subsurface than they actually are.

The creation of a global borehole information database would be a tremendous addition to the progression of geoscience data. I imagine a framework in which each core has an ID number, much like a system many state repositories already contain, where a geologist could add information to an already stored core, or add a new core entirely. The analysis of a core can take several hours, days, or even weeks to analyze. If a geologist could add the data he collected while analyzing a core (such as porosity, mineral content etc.) into a database, that could be added to by other geoscientists, the retention of data would be greatly amplified. If such a database existed, the data collection process for this thesis project could have been marginally reduced.

The results produced from the level of integration that this thesis presents show the benefits of collaborative digital data integration. Once the data were imported into MATLAB (and even when it was still in SGeMS), an infinite number of cross-sections, contours and surfaces could be extracted from the dataset. The level of interactivity and understanding increases dramatically compared to the traditional style of paper mapping, where only a few cross-sections may be used to represent an area, giving only a fraction of subsurface depiction. If more data could be added democratically by geoscientists globally into a uniform database, the accuracy of models such as the one created in this thesis would only rise.

8 References

Batjes, N.H., 1996. Total carbon and nitrogen in the soils of the world. Eur. J. Soil Sci. 47, 151–163. doi:10.1111/j.1365-2389.1996.tb01386.x

Bohling, Geoff. 2005. "INTRODUCTION TO GEOSTATISTICS And VARIOGRAM ANALYSIS." Kansas: Kansas Geological Survey , October 17.

Chilès, J.P., Guillen, A., Aug, C., Guillen, A., Lees, T., 2015. Modelling the Geometry of Geological Units and its Uncertainty in 3D From Structural Data : The Potential-Field Method Modelling the Geometry of Geological Units and its Uncertainty in 3D From Structural Data : The Potential-Field Method.

Deutsch, C.V., and A.G. Journal. 1998. *GSLIB Geostatistical Software Library and User's Guide, 2nd Edition*. New York, New York: Oxford University Press.

Fan, Y., Richard, S., Bristol, R.S., Peters, S.E., Ingebritsen, S.E., Moosdorf, N., Packman, A., Gleeson, T., Zaslavsky, I., Peckham, S., Murdoch, L., Fienen, M., Cardiff, M., Tarboton, D., Jones, N., Hooper, R., Arrigo, J., Gochis, D., Olson, J., Wolock, D., 2015. DigitalCrust - a 4D data system of material properties for transforming research on crustal fluid flow. Geofluids 15, 372–379. doi:10.1111/gfl.12114

Houlding, S., 1996. 3D geoscience modeling: Computer techniques for geological characterization. Earth-Science Rev. 40, 299–301. doi:10.1016/0012-8252(95)00067-4

Howell, P.D., 1999. Structural sequences and styles of subsidence in the Michigan basin 974–991.

Levin, H.L., 2010. The Earth Through Time - Harold L. Levin - Google Books [WWW Document]. John Wiley Sons, Inc. URL https://books.google.com/books?id=D0yl7Cqsu78C&pg=PA15&dq=principle+of+original+horizontality&hl=en&ei=Vq_yTIrgO4TDhAfw9fSxDQ&sa=X&oi=book_result&ct=result&resnum=5&ved=0CDsQ6AEwBA#v=onepage&q=principle of original horizontality&f=false (accessed 3.1.17).

Nadon, G.C., Simo, J.A.T., Dott, R.H., Byers, C.W., 2000. High-Resolution Sequence Stratigraphic Analysis of the St . Peter Sandstone and Glenwood Formation (Middle Ordovician), Michigan Basin , U . S . A . 1 7, 975–996.

Perrin, M., Zhu, B., 2005. Knowledge-driven applications for geological modeling. J. Pet. Sci. Eng. 47, 89–104. doi:10.1016/j.petrol.2004.11.010

Peters, S.E., Zhang, C., Livny, M., Ré, C., 2014. A machine-compiled macroevolutionary history of Phanerozoic life.

Remy, Nicholas. 2009. *Applied Geostatistics with SGeMS.* Cambridge, New York: Cambridge University Press.

Soller, David R. 1998. "Map Showing the Thickness and Character of Quaternary Sediments in the Glaciated United States East of the Rocky Mountains." U.S. Department of the Interior. U.S. Geological Survey.

Stein, Michael. 1999. *Interpolation of Spatial Data: Some Theory for Kriging.* New York, New York: Springer Science+Business Media.

USGS. n.d. *GMNA Resources.* Accessed 1 12, 2017. https://ngmdb.usgs.gov/gmna/.

2016. "USGS." *The Public Land Survey System (PLSS).* 12 16. Accessed 4 4, 2017. https://nationalmap.gov/small_scale/a_plss.html.