WORKFLOWS FOR DATA INTENSIVE GROUNDWATER ASSESSMENT

Louis-Charles Boutin | Kevin Hayley
Gordon MacMillan | Jens Schumacher

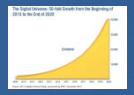
Matrix Solutions Inc.
ENVIRONMENT & ENGINEERING

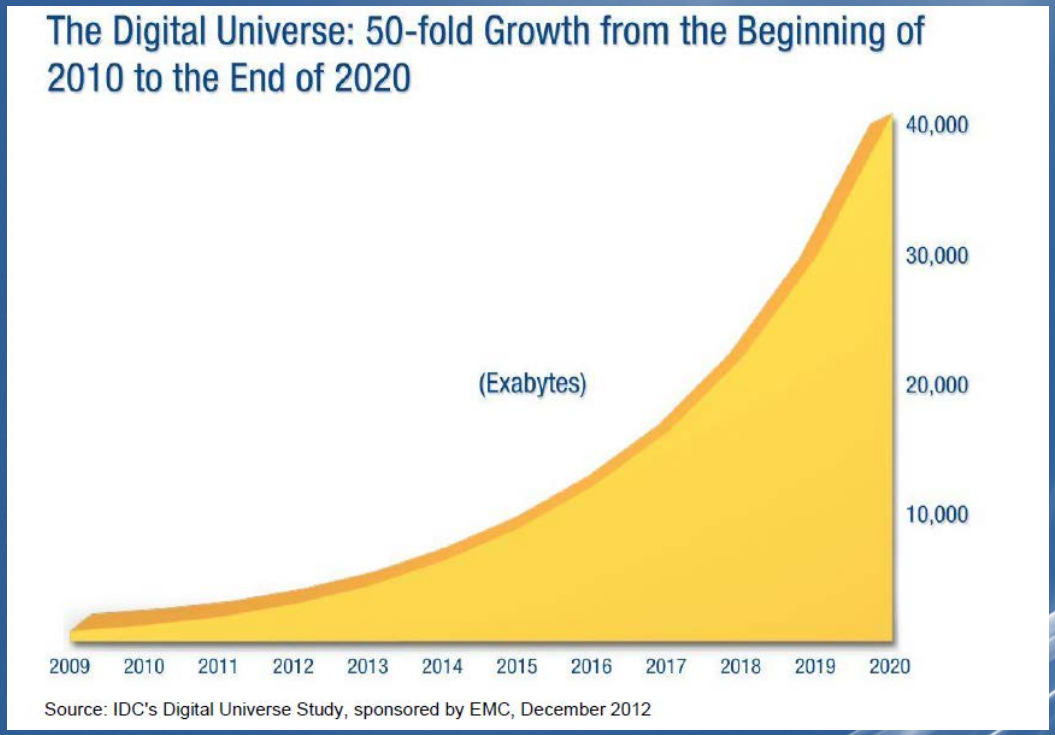
PRESENTATION OUTLINE

- Introduction
- Tools and Workflows
- Conclusion

INTRODUCTION – WHY?

Big Data

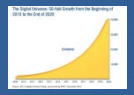




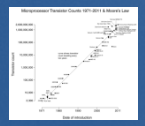
Decimal		
Value	Metric	
1000	kB	kilobyte
1000^2	MB	megabyte
1000^3	GB	gigabyte
1000^4	TB	terabyte
1000^5	PB	petabyte
1000^6	EB	exabyte
1000^7	ZB	zettabyte
1000^8	YB	**yottabyte**

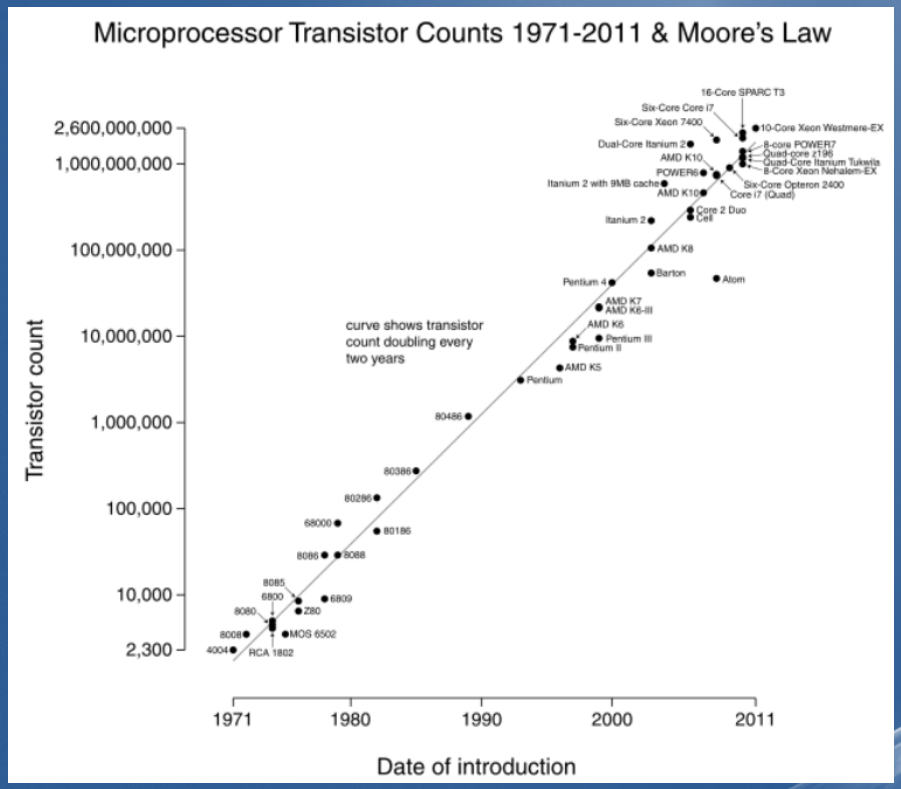
INTRODUCTION – WHY?

Big Data



CPU/Algorithm/Cloud



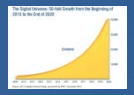


- Simultaneously increases in algorithm efficiency.
- Cloud computing and parallelization of the solver.

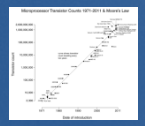
Scientist Skills

INTRODUCTION – WHY?

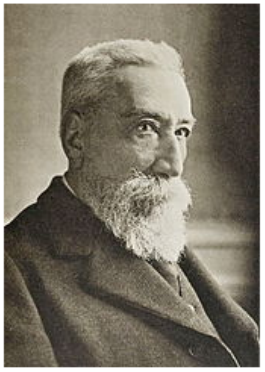
Big Data



CPU/Algorithm/Cloud



Anatole France, Updated



Anatole France (1844-1924)

Then

The law, in its majestic equality, forbids the rich and poor alike to sleep under bridges, to beg in the streets, and to steal bread.

Now

The Internet, in its majestic equality, allows every scientist to analyze massive data sets using web services and cloud computing.

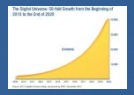
Originally posted 2014-03-07 by Greg Wilson in Opinion.

Greg Wilson
The Mozilla Foundation → open internet
Software Carpentry → http://software –carpentry.org

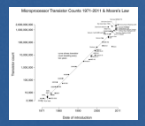
Scientist Skills

INTRODUCTION – WHY?

Big Data



CPU/Algorithm/Cloud



Some of the challenges are:

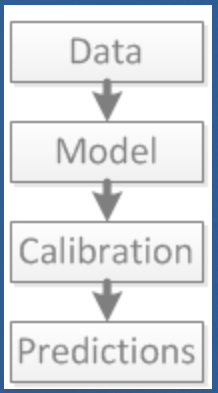
- Capacity to store raw data
- Capacity to search, aggregate, filter, Qa/Qc, and cross-reference data sets
- Repeatability

AND

- Clients, regulators, stakeholders expectations.

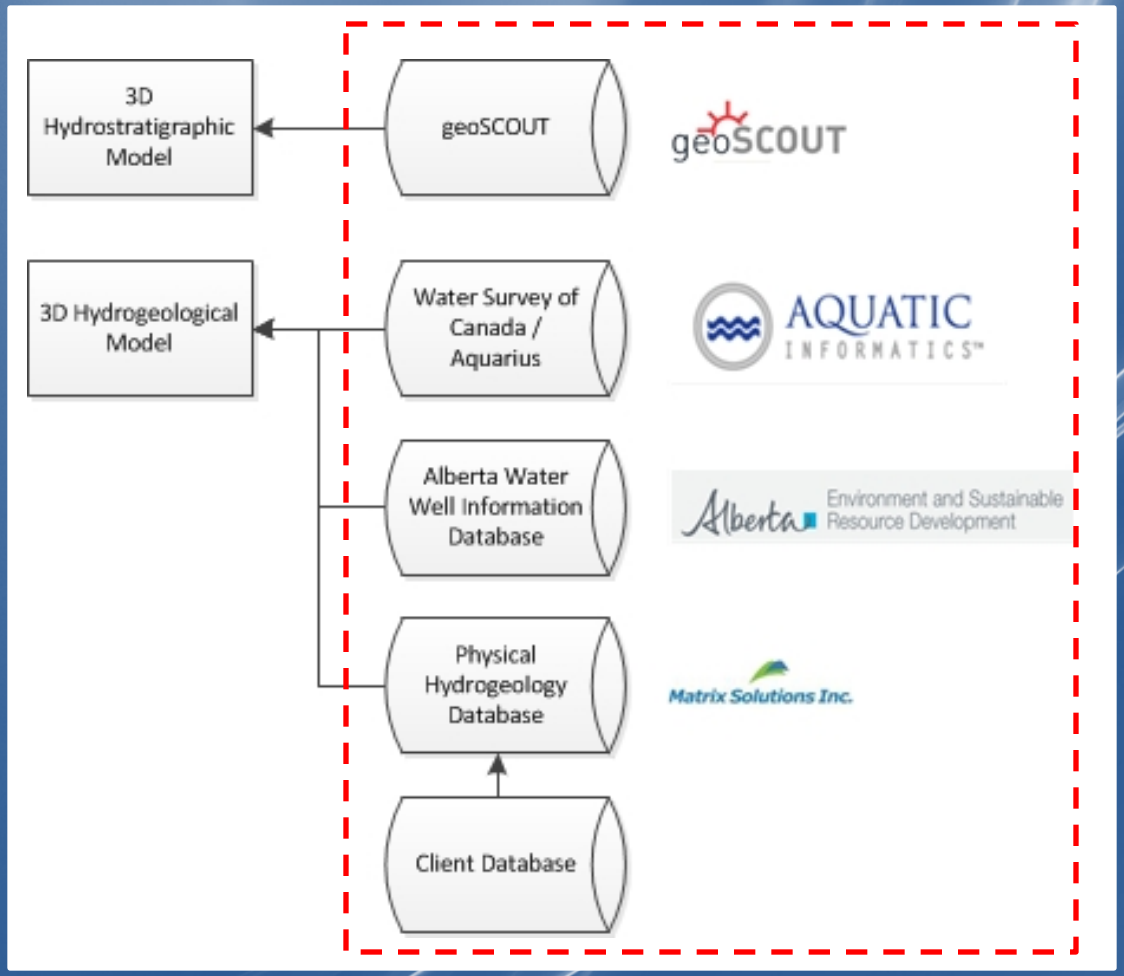
WORKFLOWS FOR DATA INTENSIVE GROUNDWATER ASSESSMENT – TOOLS FROM SOFTWARE ARCHITECTURE AND SOFTWARE ENGINEERING

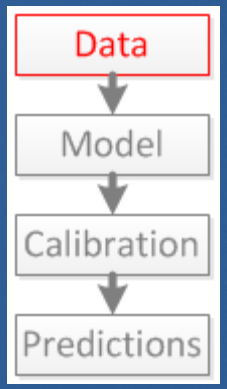
- Databases
- Python Scripts
- Matrix's PEST Graphical User Interface (GUI)
- Cloud Computing
- Version Control System



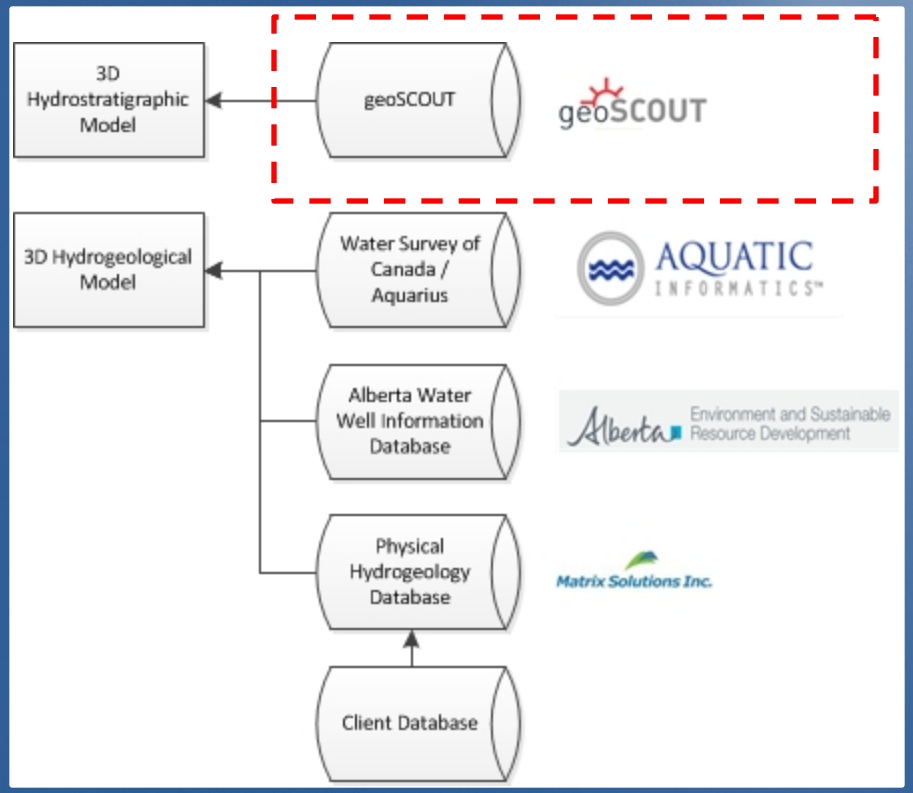
DATABASES

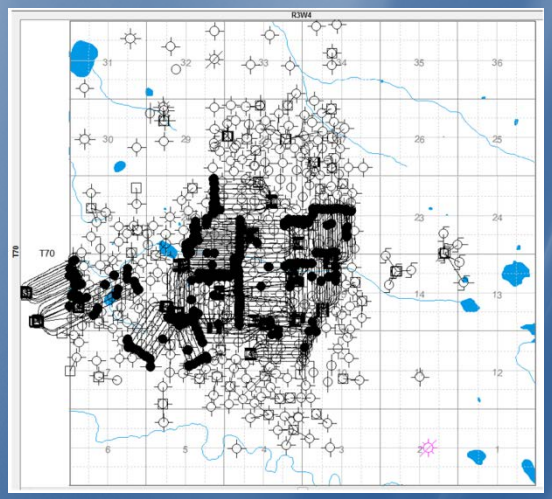
- Geology Dataset
- Hydraulic Heads
- Withdrawal and Disposal Rates
  - Reported
  - Planned
- Surface Water Gauge

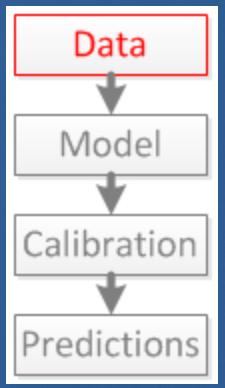


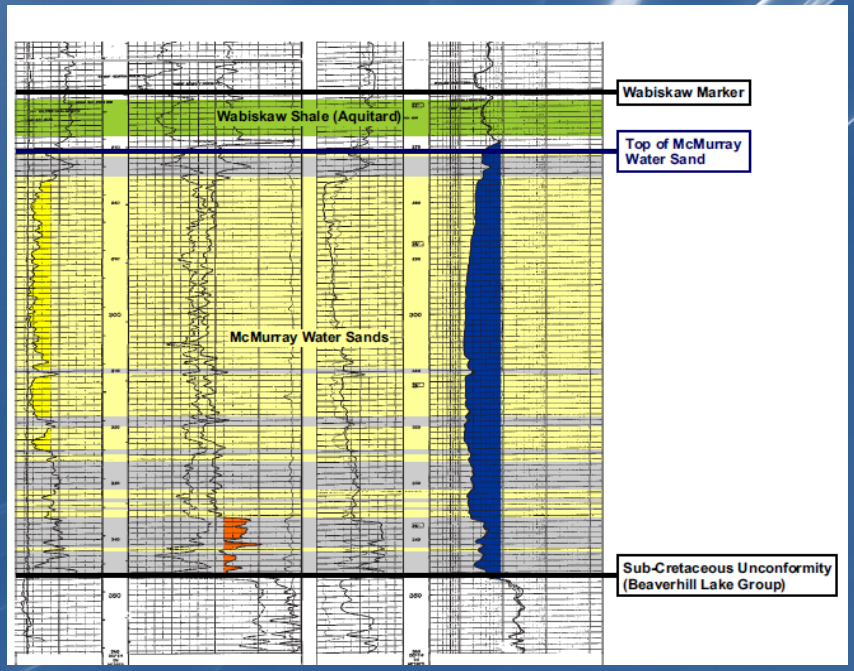


GEOSCOUT

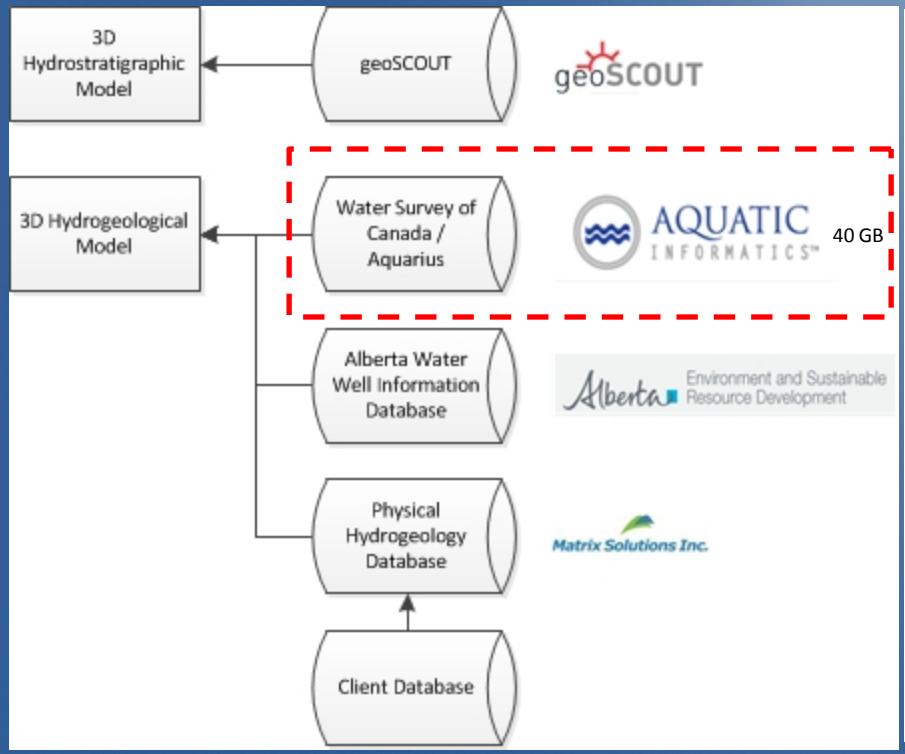

3D Hydrostratigraphic Model
geoSCOUT
geoSCOUT
3D Hydrogeological Model
Water Survey of Canada / Aquarius
AQUATIC INFORMATICS
Alberta Water Well Information Database
Alberta Environment and Sustainable Resource Development
Physical Hydrogeology Database
Matrix Solutions Inc.
Client Database

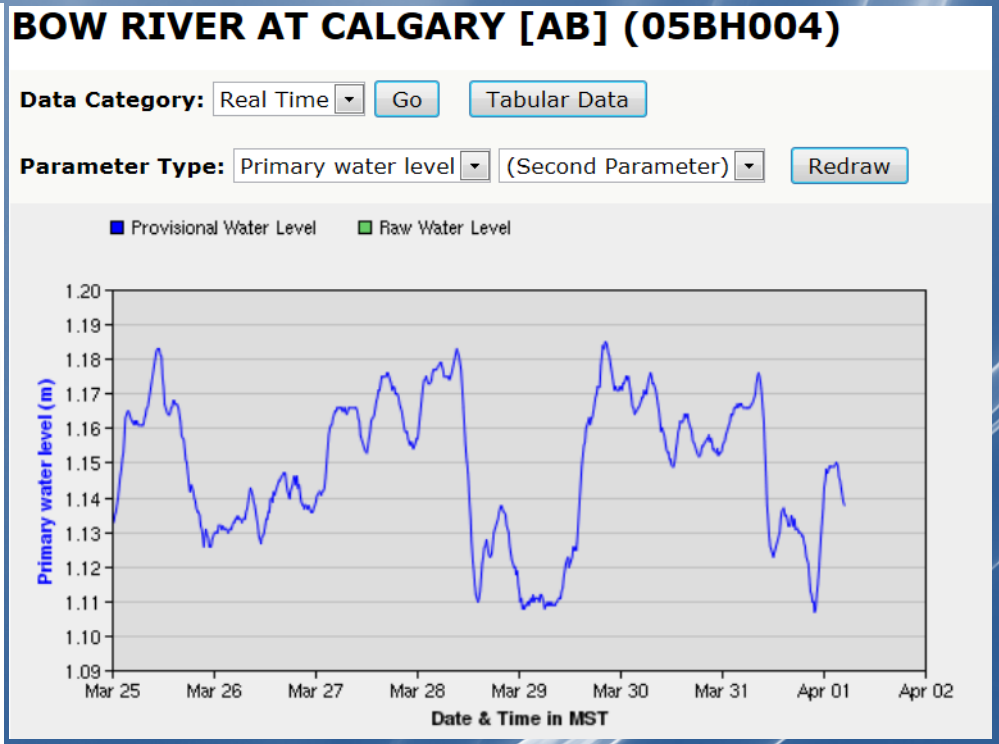


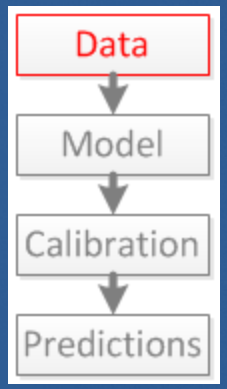

Data
Model
Calibration
Predictions


Wabiskaw Marker
Wabiskaw Shale (Aquitard)
Top of McMurray Water Sand
McMurray Water Sands
Sub-Cretaceous Unconformity (Beaverhill Lake Group)

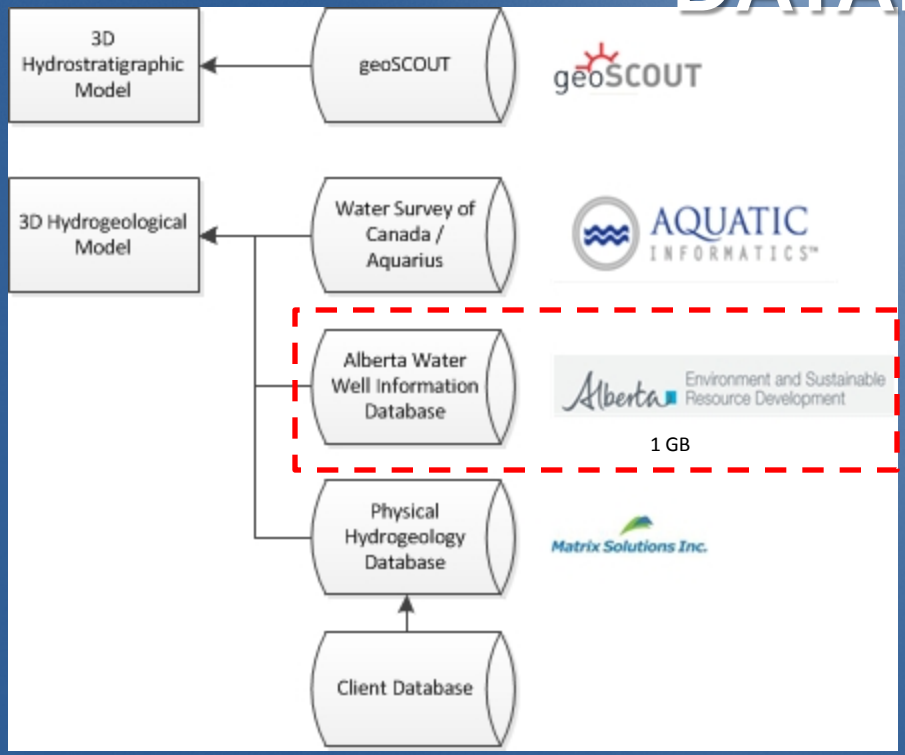
AQUARIUS

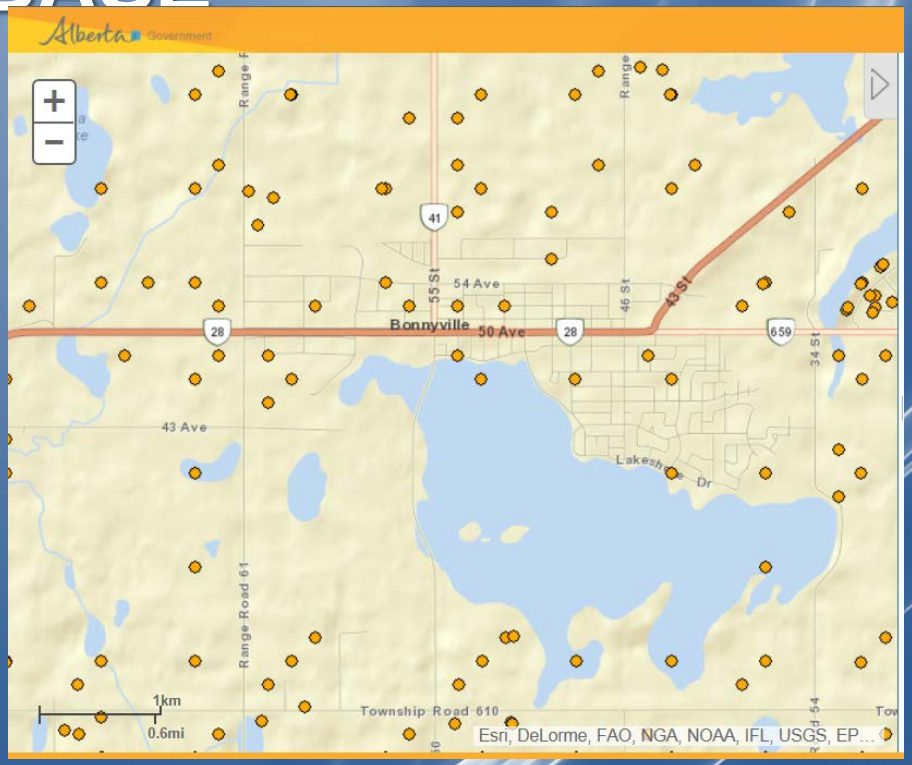

3D Hydrostratigraphic Model
geoSCOUT
geoSCOUT
3D Hydrogeological Model
Water Survey of Canada / Aquarius
AQUATIC INFORMATICS
40 GB
Alberta Water Well Information Database
Alberta Environment and Sustainable Resource Development
Physical Hydrogeology Database
Matrix Solutions Inc.
Client Database

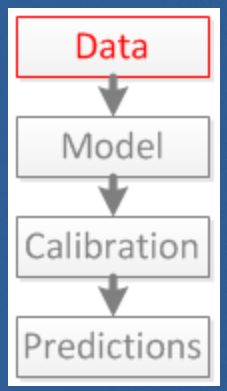

BOW RIVER AT CALGARY [AB] (05BH004)
Data Category: Real Time Go Tabular Data
Parameter Type: Primary water level (Second Parameter) Redraw
Provisional Water Level
Raw Water Level
Primary water level (m)
1.20
1.19
1.18
1.17
1.16
1.15
1.14
1.13
1.12
1.11
1.10
1.09
Mar 25
Mar 26
Mar 27
Mar 28
Mar 29
Mar 30
Mar 31
Apr 01
Apr 02
Date & Time in MST


Data
Model
Calibration
Predictions

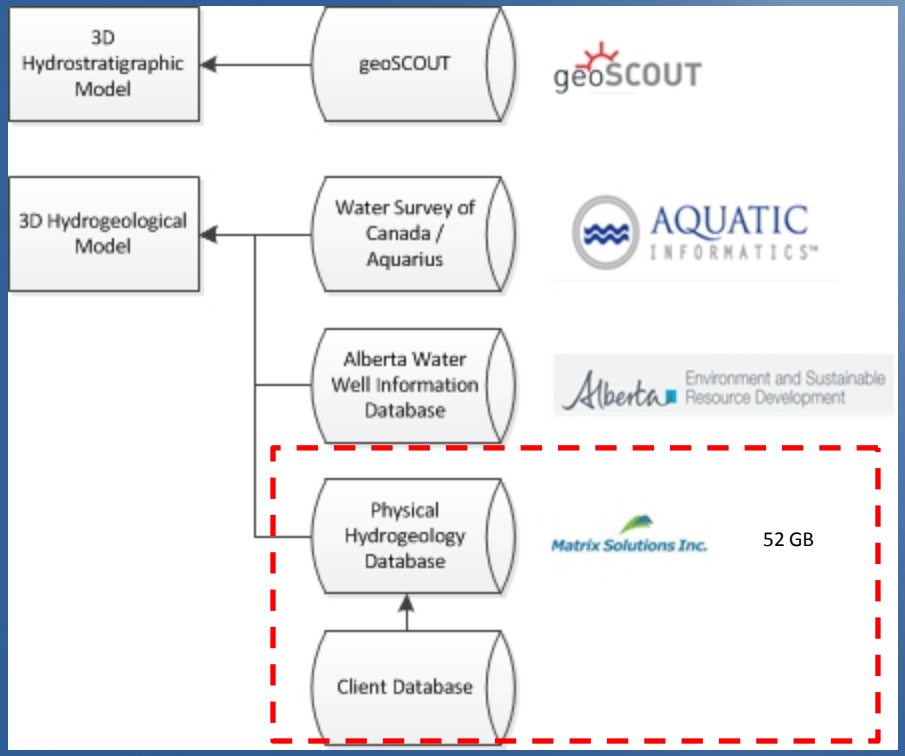
ALBERTA WATER WELL INFORMATION DATABASE

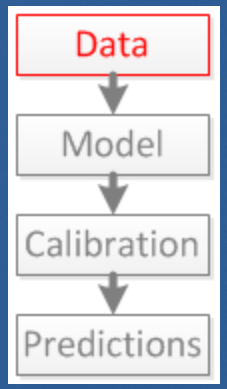






MATRIX'S - PHYSICAL HYDROGEOLOGY DATABASE





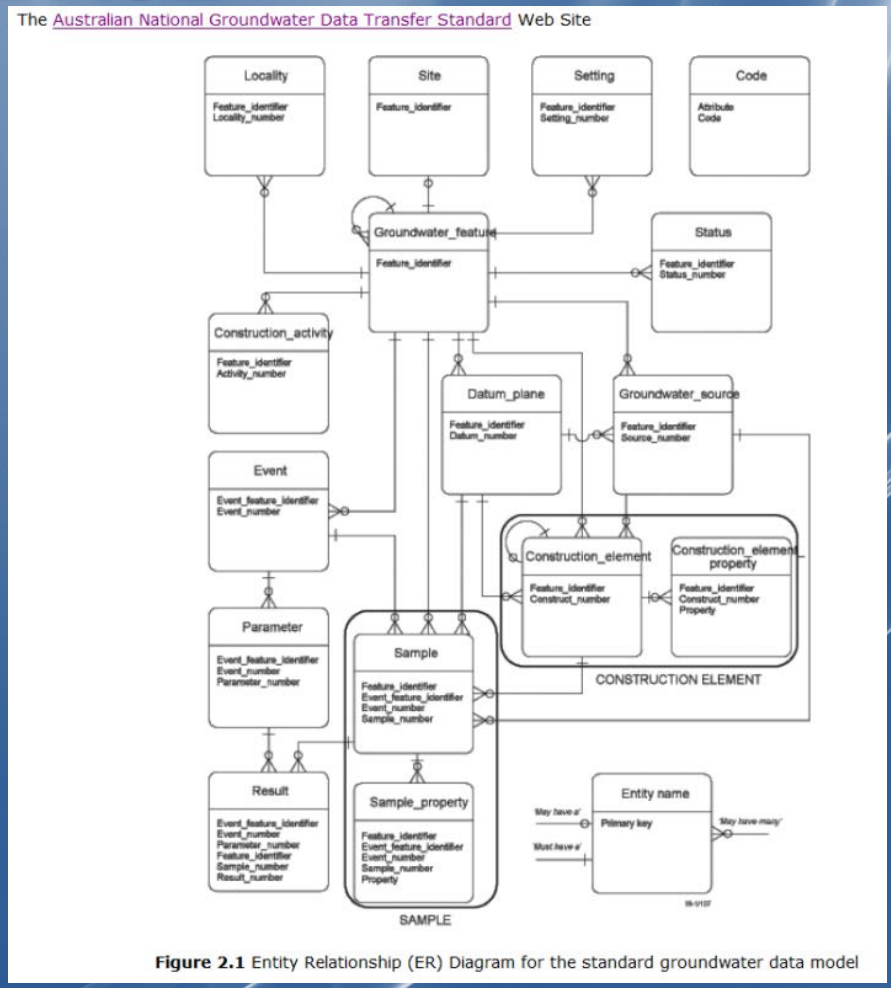
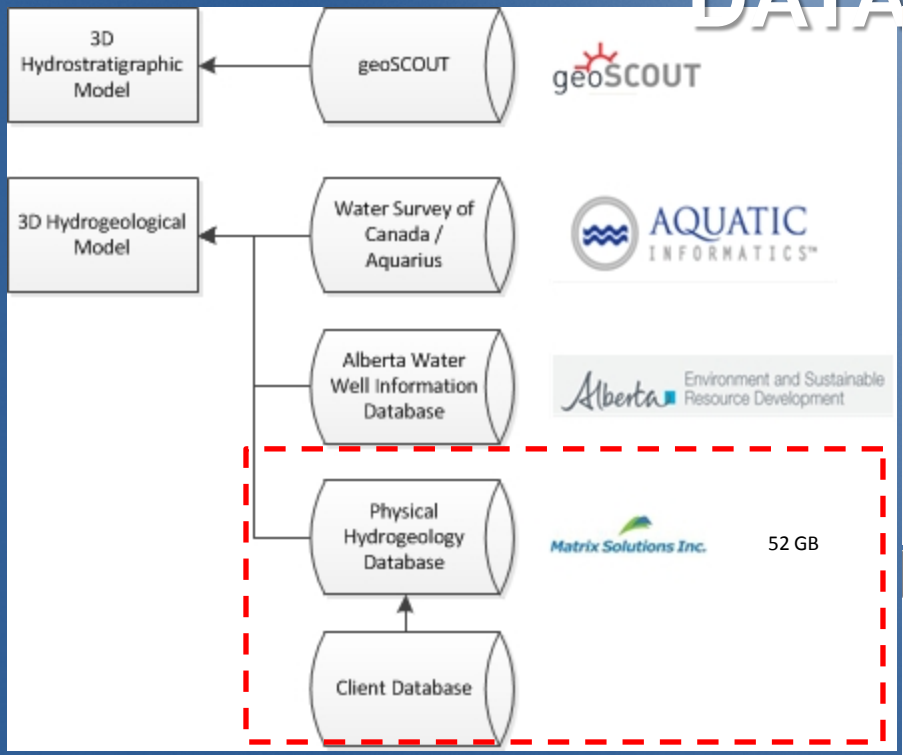
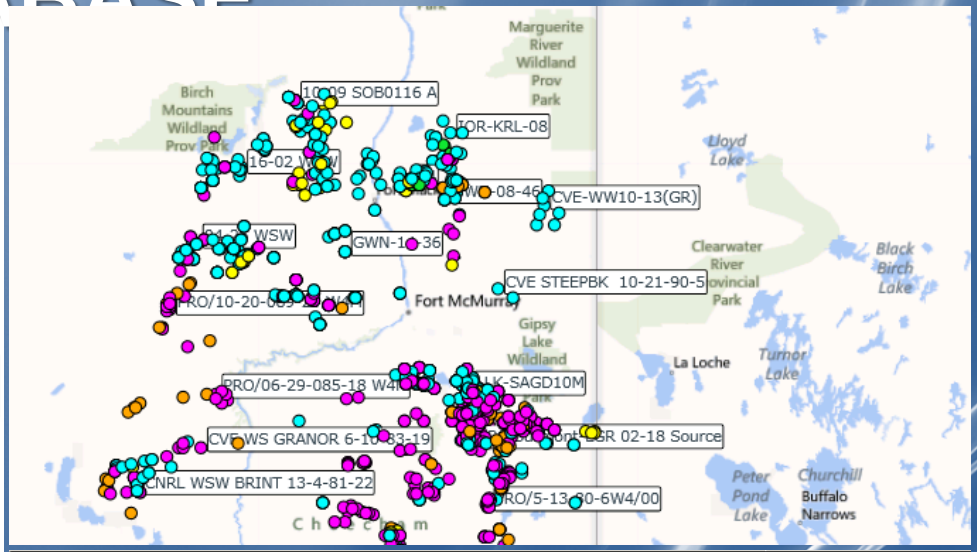
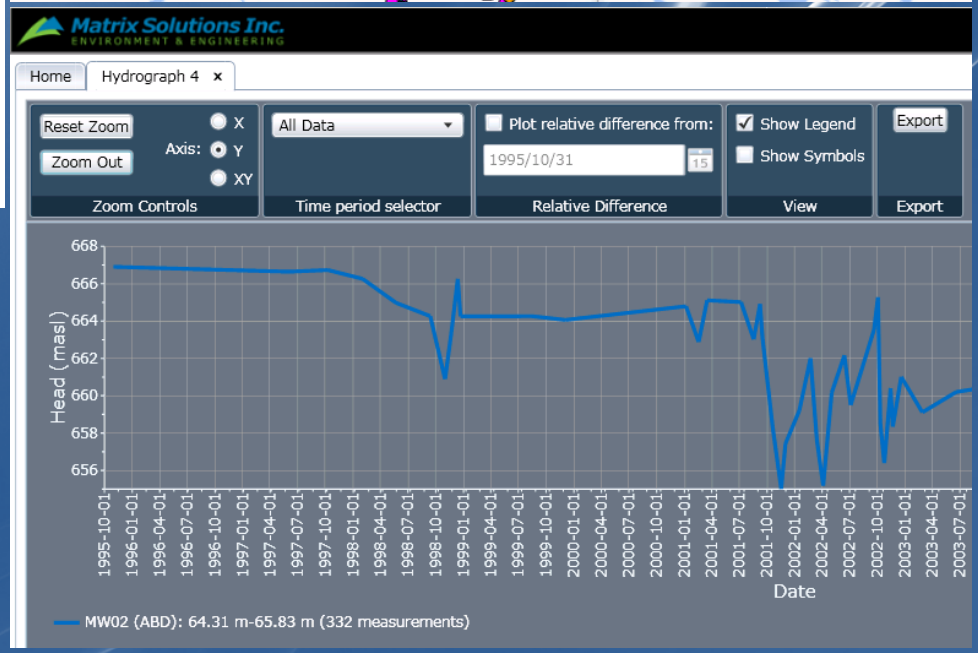


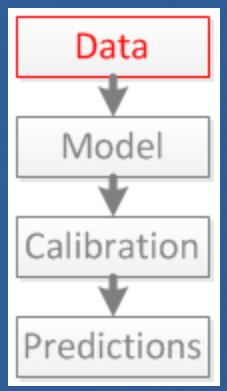
Figure 2.1 Entity Relationship (ER) Diagram for the standard groundwater data model

MATRIX'S - PHYSICAL HYDROGEOLOGY DATABASE

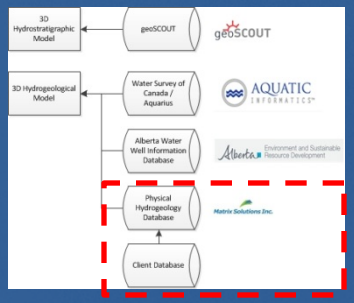


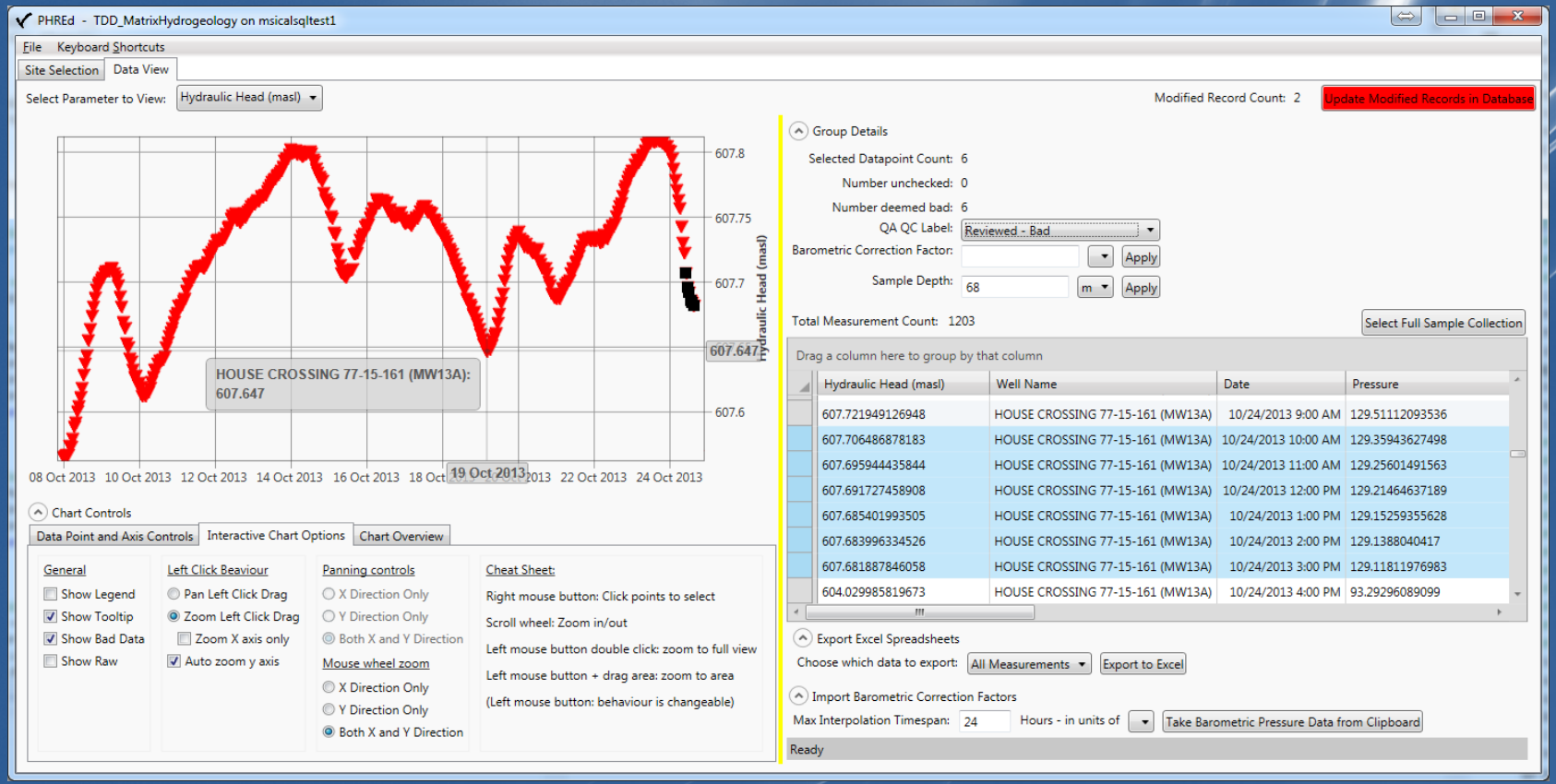


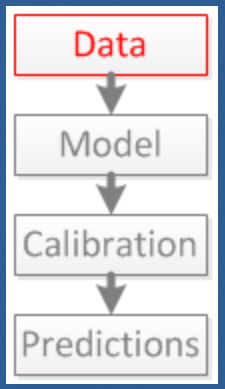




MATRIX'S - PHYSICAL HYDROGEOLOGY DATABASE

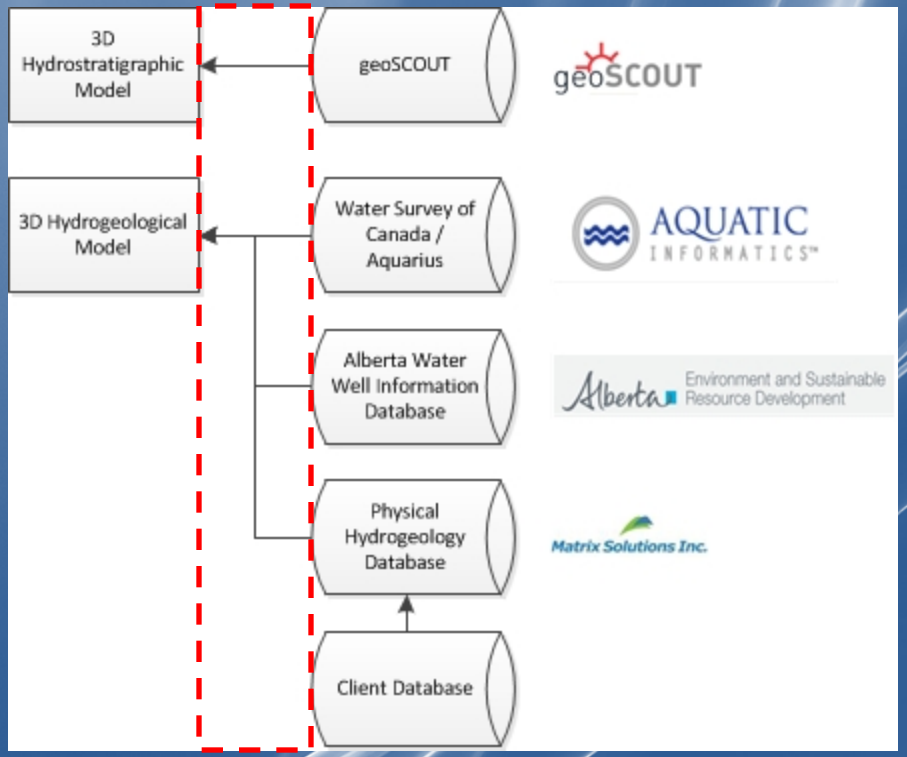


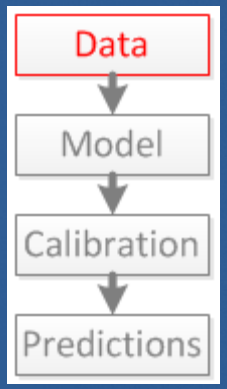




PYTHON SCRIPTS

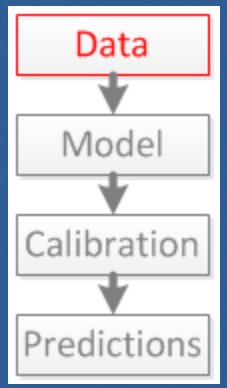
- Used for:
  - Database query
  - Data QA/QC
  - Data filtering
  - Data Manipulation to Input/Output files

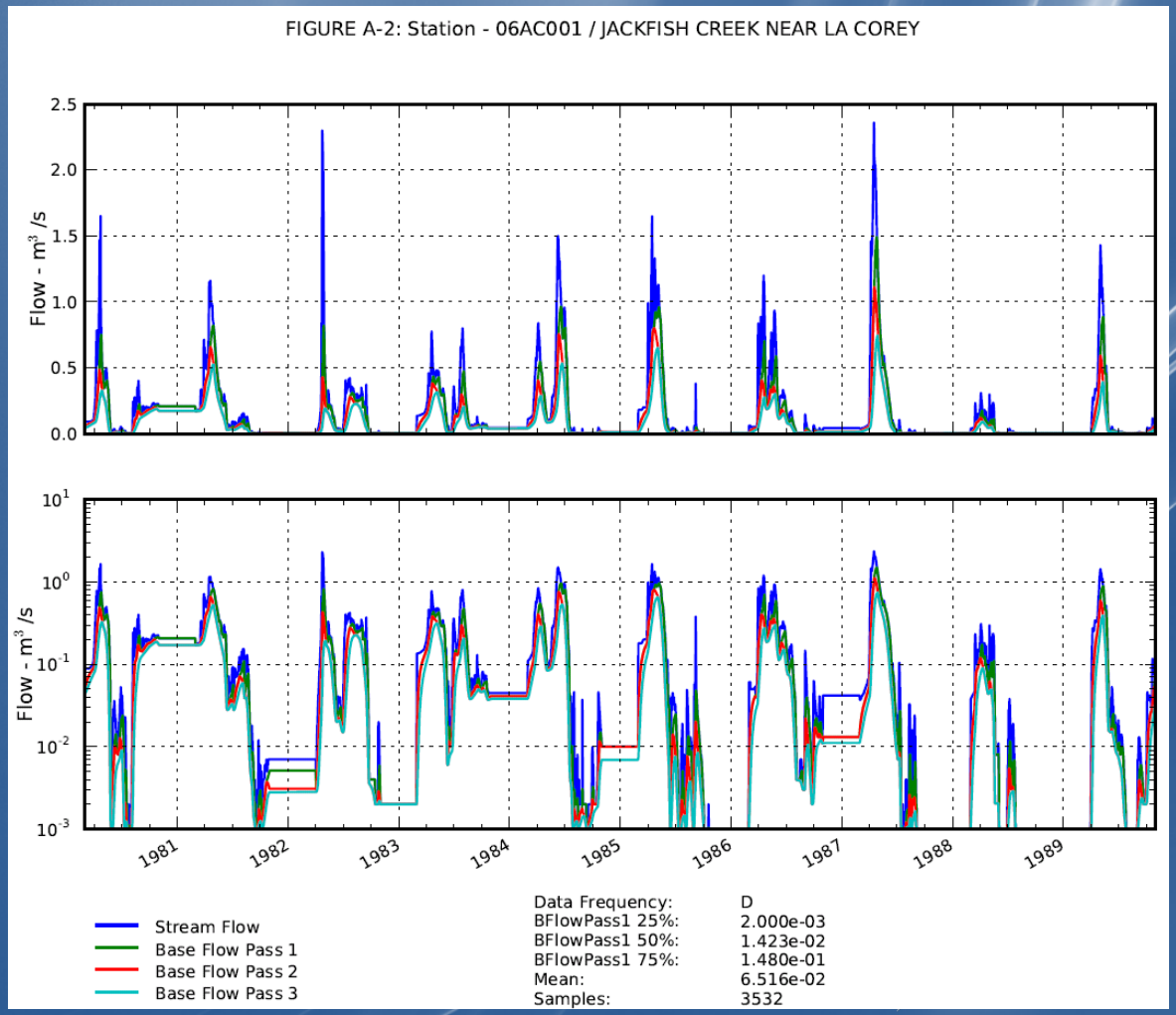




PYTHON SCRIPTS

- Used for:
  - Database query
  - Data QA/QC
  - Data filtering
  - Data Manipulation to Input/Output files



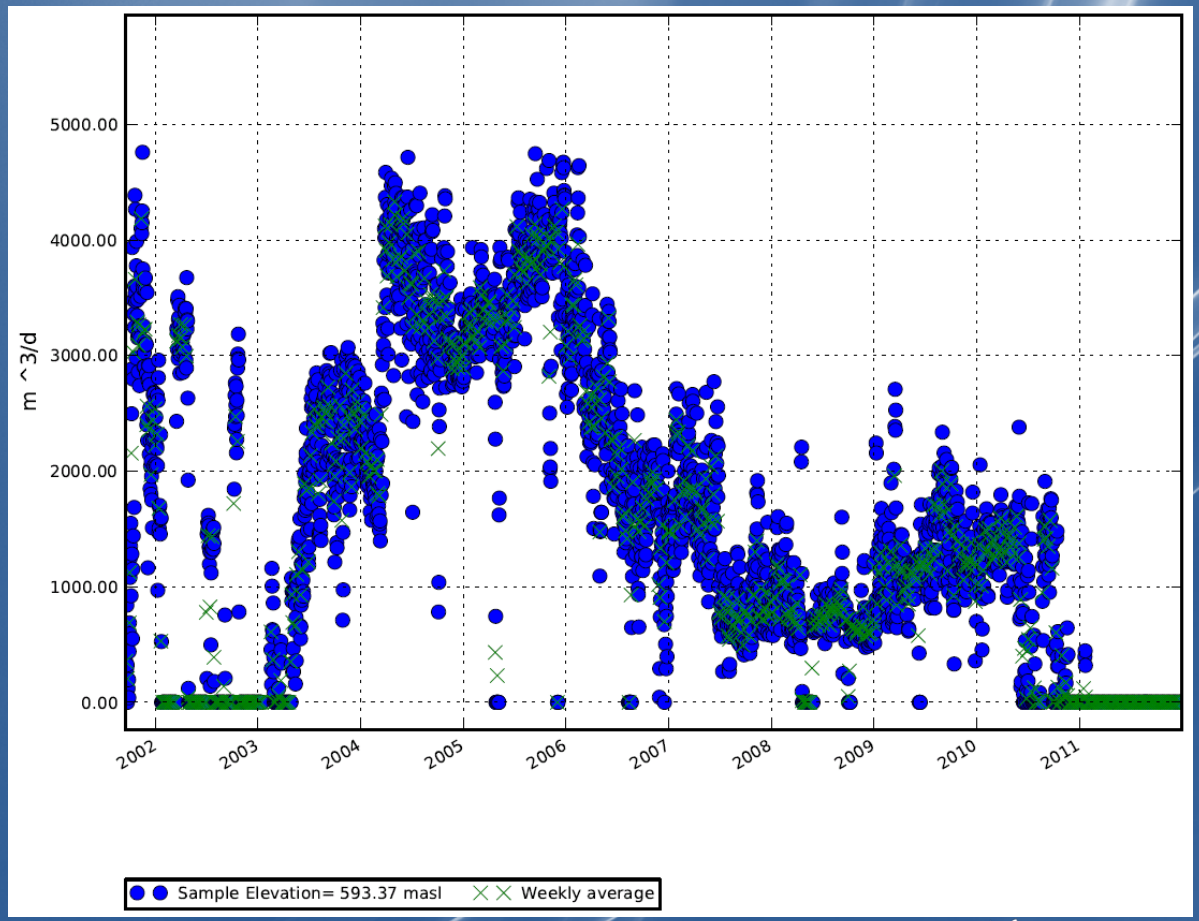


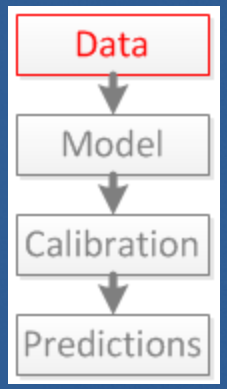
PYTHON SCRIPTS

- Used for:
  - Database query
  - Data QA/QC
  - Data filtering
  - Data Manipulation to Input/Output files

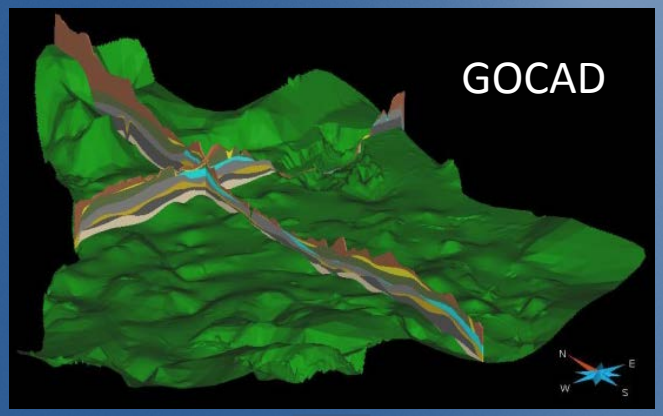
PumpingRates.pow

```
1   #1
2   !Pumping Well1
3   ![type=Constant]
4   0.00 0
5   2450     354.0
6   2456     204.571428571
7   2463     688.0
8   2464     1123.5
9   2470     2155.92857143
10  2477     3028.17142857
11  2484     3658.84285714
12  2491     3287.4
13  2495     3088.76666667
14  2502     0.0
15  END
```



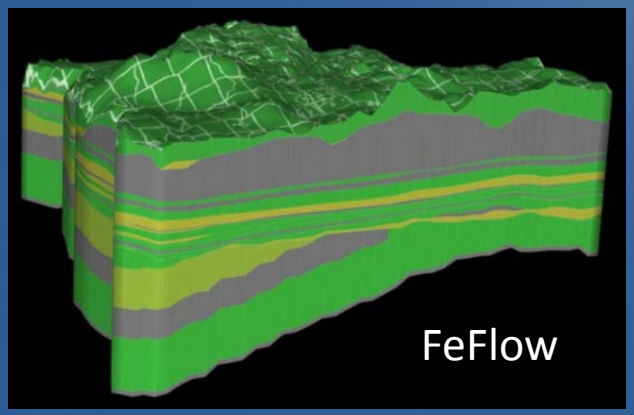


PYTHON SCRIPTS



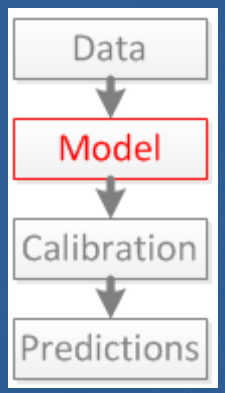
Elevations_Corrected_FEFLOW.dat

	X	Y	Z	Slice
2	570623.632812	6053689.51562	576.592102051	1
3	570623.632812	6053689.51562	555.247802734	2
4	570623.632812	6053689.51562	555.147802734	3
5	570623.632812	6053689.51562	555.047802734	4
6	570623.632812	6053689.51562	554.947802734	5
7	570623.632812	6053689.51562	554.847802734	6
8	570623.632812	6053689.51562	525.287719727	7
9	570623.632812	6053689.51562	525.187719727	8
10	570623.632812	6053689.51562	512.924682617	9



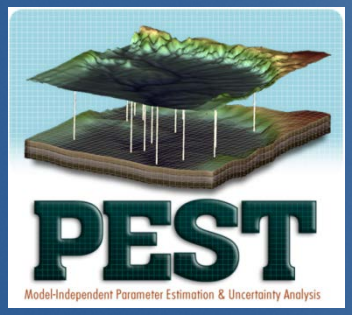
Material_Properties.dat

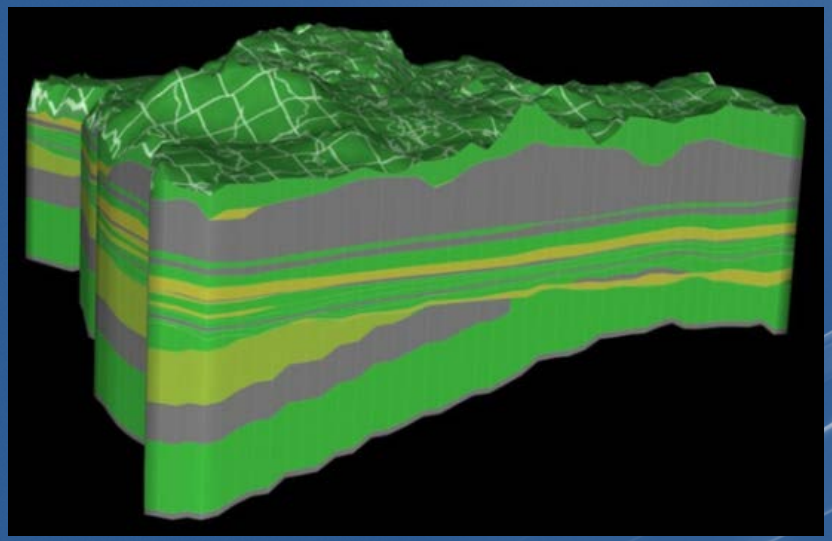
	ELEM	LAYER	X	Y	THICK	Unit
2	2145	1	564440.751668	5986655.60257	32.2664184567	4
3	35104	2	564440.751668	5986655.60257	0.1000000003	17
4	68063	3	564440.751668	5986655.60257	4.982051595	5
5	101022	4	564440.751668	5986655.60257	34.6912027993	6
6	133981	5	564440.751668	5986655.60257	0.1	6
7	166940	6	564440.751668	5986655.60257	63.011022949	10
8	199899	7	564440.751668	5986655.60257	0.1	10
9	232858	8	564440.751668	5986655.60257	24.55512085	8
10	265817	9	564440.751668	5986655.60257	4.976348877	11
11	298776	10	564440.751668	5986655.60257	10.4566853837	12

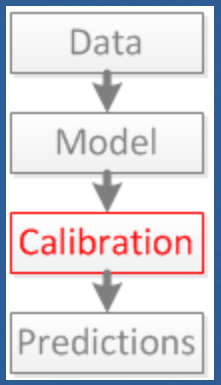


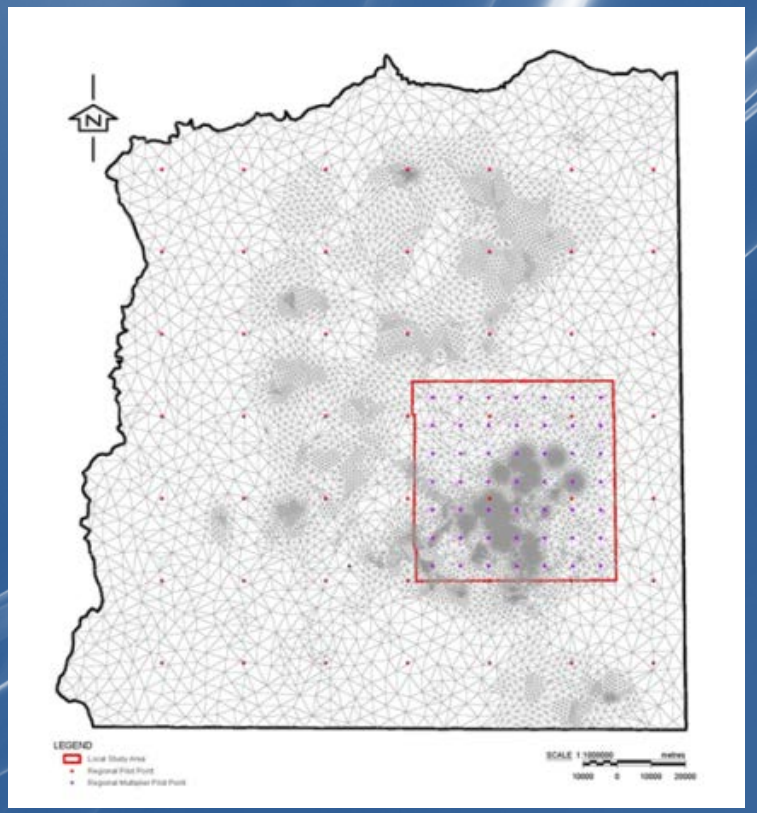
MATRIX'S PEST GUI

- PEST (Model Independent Parameter Estimation)



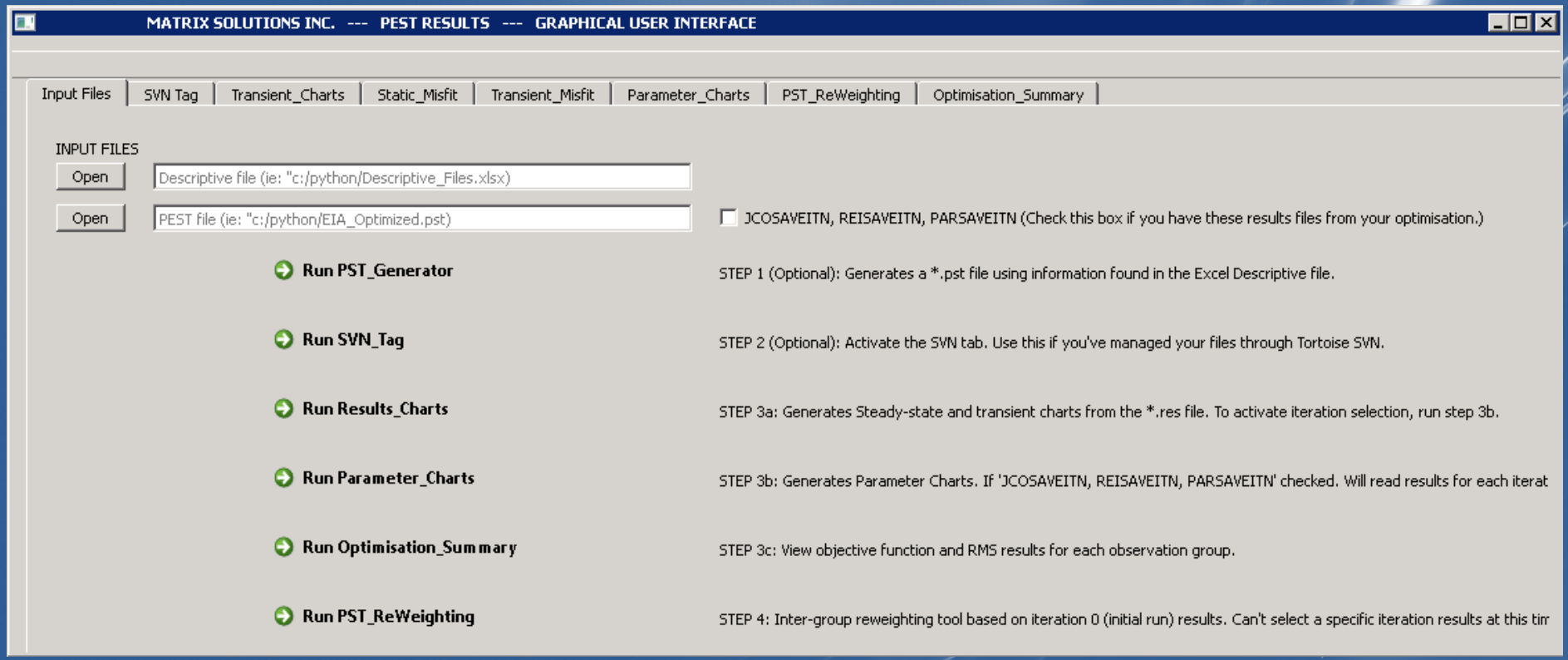






MATRIX'S PEST GUI

- Python / Qt
- Reads Descriptive File (Spatial + Temporal Information)
- Automated generation of PEST files



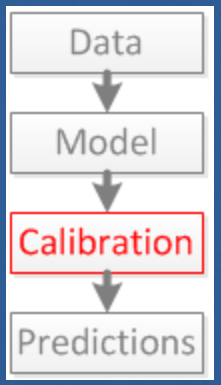
MATRIX'S PEST GUI

- Automated observation group/ regularization reweighting
- Parameter visualization

Obs Group	Cur. Contribution Obj.Func.	Current %	Multiplier Factor
h_aaaa	285208	22.5691	1
eswl	165634	13.107	1
baseflow	160577	12.7068	1
first_head	159534	12.6243	1
grad	130499	10.3267	1
h_aaab	122167	9.66733	1

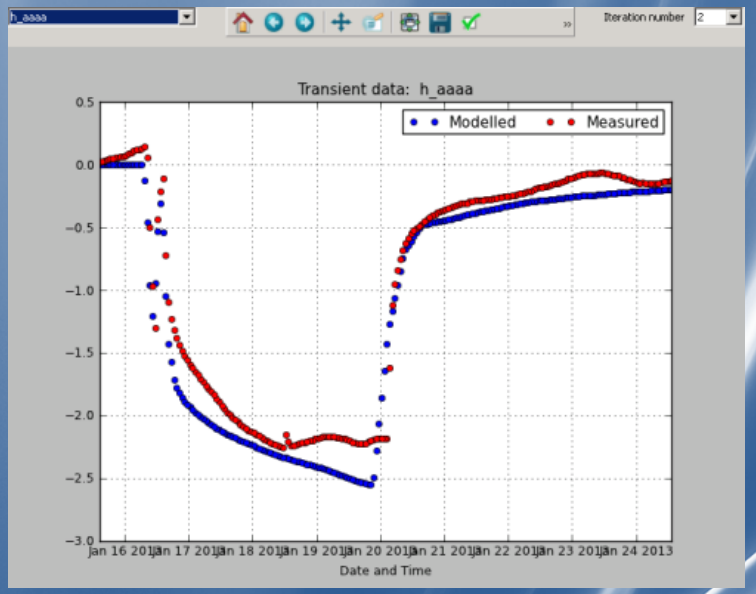
Regul Group	Cur. Contribution Obj.Func.	Current %	Multiplier Factor
regul_pv	14.612	52.927	1
regul_aqt	8.8392	32.017	1
regul_aqf	3.9341	14.2499	1
regul_mlf	0.1319	0.477763	1
regul_ratio	0.0906191	0.328237	1

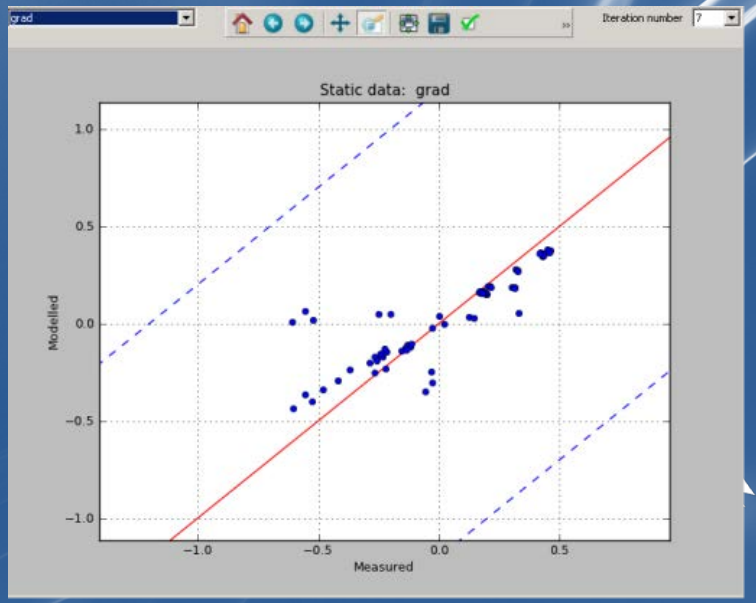
condxy	bf1skx;3.02e-05	bf1tkx;5.00e-08	bf2tkx;4.41e-08	blftkx;5.00e-08	e1skx;7.49e-05
	e2tkx;1.82e-08	e3skx;1.00e-06	elfskx;9.37e-05	etskx;1.00e-06	
	mcftkx;3.48e-08	mlflockx;7.05e-05	mlfskx;3.00e-04	rpgskx;2.42e-05	srfskx;4.22e-05
condz	bf1tkz;1.99e-10	bf2tkz;8.27e-10	blftkz;5.00e-09	e2tkz;1.00e-11	
	mcftkz;7.61e-10				
ratio	ratios;5.00e+00				

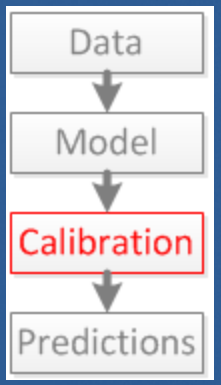


MATRIX'S PEST GUI

- Visualization of data misfit
- Residuals statistics

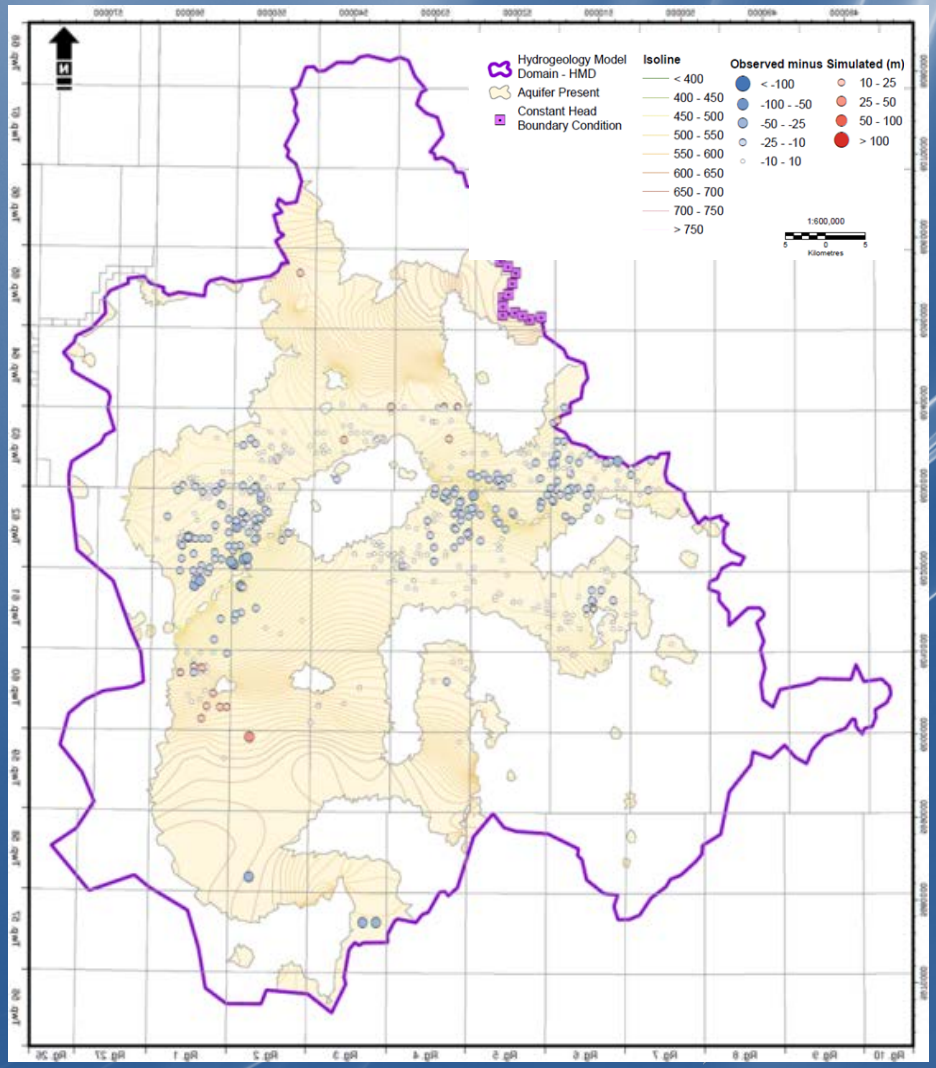


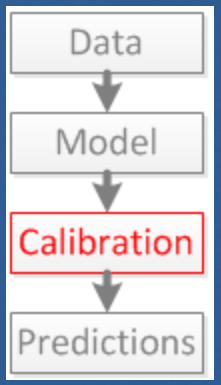




MATRIX'S PEST GUI

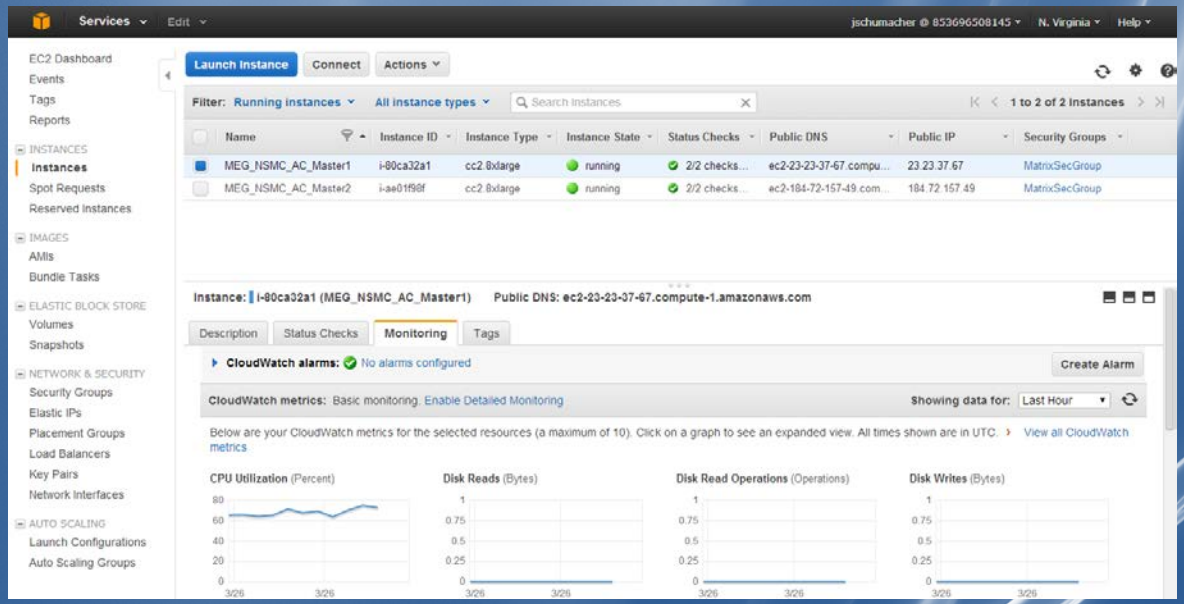
- Plugin for Quantum GIS
- Shapefiles automatically generated

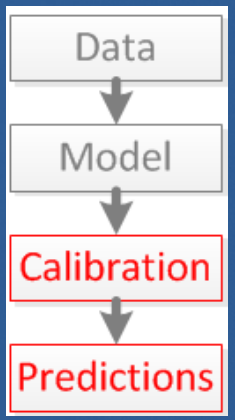


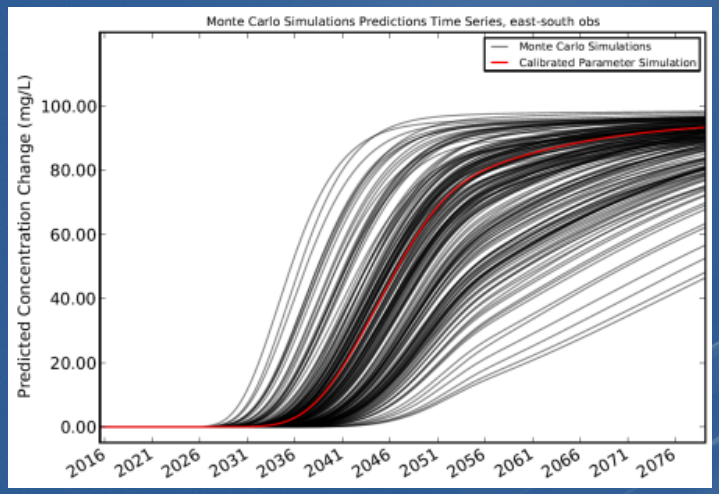


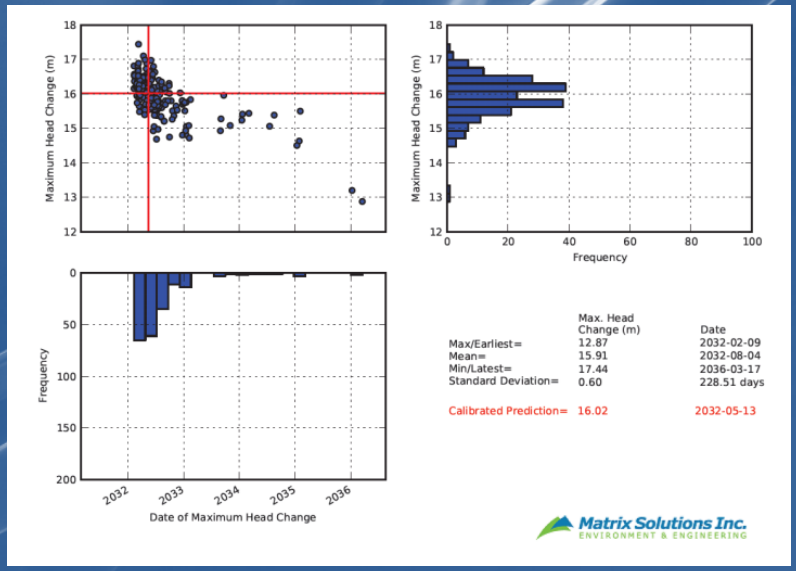
CLOUD COMPUTING

- Amazon EC2 server
  - Solution for higher CPU
  - PEST Optimization
  - Null-Space Monte Carlo Uncertainty Analysis



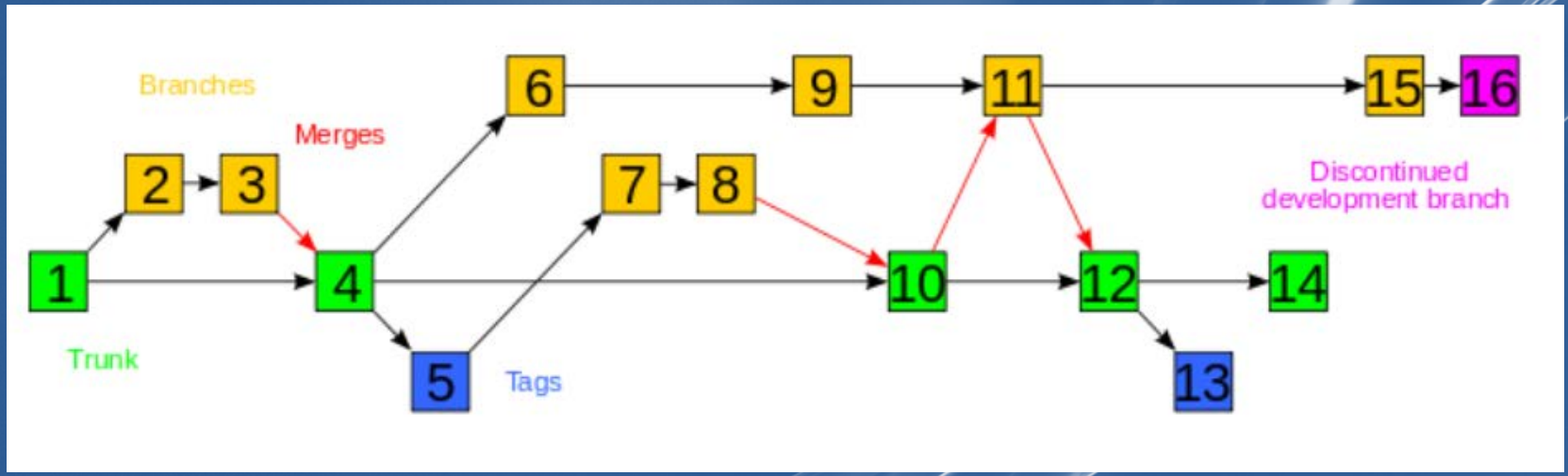


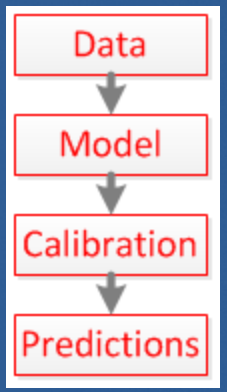




VERSION CONTROL SYSTEM

- Repository with binary Incremental changes
- Keep log of changes and versions
- Extremely useful for collaborative work and script development/deployment/testing (TDD)





SUMMARY

- Data intensive groundwater assessment becoming the norm
- Multiple challenges
- Database development and python scripting has allowed for efficient data management and QA/QC
- Linking PEST files to Python visualization, GIS, and a QT GUI has helped in making subjective weighting and regularization choices when calibrating highly parameterized models with PEST
- Version control provides a efficient workflow management and tracking to allow efficient collaboration and repeatability.

QUESTIONS