Memo: Tests of simplified conductivity stratigraphy by Monte Carlo analysis

Authors: Calvin Whealton, Teresa Jordan, Zach Frone

Executive Summary

Monte Carlo analysis was used to examine the implications of using the COSUNA approximations versus using more detailed information for a set of 77 wells chosen across the region. For each of the wells there were three cases for the stratigraphic columns: detailed stratigraphy with Carter conductivities, COSUNA stratigraphy with Carter et al. (1998) conductivities, and COSUNA stratigraphy with Beardsmore and Cull (2001) conductivities. All stratigraphic assumptions were tested with 50,000 Monte Carlo replicates with most parameters being modeled as triangular distributions.

The results of the analysis are that the differences between the COSUNA stratigraphy with Carter conductivities and the detailed stratigraphy are generally minor when compared over the whole region. When comparing surface heat flow, if there is a systematic difference it is probably around 2-5 mW/m^2, which is typically around 10% of the predicted value. The uncertainty of the two methods for a single well is also close on average, but the actual data shows more noise. When predicting temperature at 3 km, the two methods were typically within about 6 °C of each other when comparing their mean prediciton, which illustrates the robustness of the COSUNA approximation with Carter et al. (1998) conductivities for this region.

Part 1: Acquisition of Detailed Well-Specific Conductivity Stratigraphy Columns

Criteria for inclusion of a given well:

- regional expert geologists (e.g., state geological survey staff or USGS geologists) have made available interpretation of formation tops for these wells
- deep wells (as close to basement as possible in a given county)
- widely and semi-uniformly distributed
- BHT available and judged to be relatively reliable

Data Sources

- state geological survey reports and publications
- Cornell, West Virginia University, and Southern Methodist University prior studies based on a given borehole
- USGS cross sections and specialty reports
- state well information sites (WV Pipeline, NY ESOGIS, PAIRIS-WIS)

Stages of selection work

- from lists of deep wells and county names, looked up which ones have geological reports of depths to formation tops.
- assembled list of >200 candidate wells
- compared the candidates list to wells in NGDS and other sources of BHTs, omitting from the "candidates" list those for which there are not BHT data
- after initially finding no matches of stratigraphically described deep wells and wells with reported BHTs for WV, went into WV Pipeline and well log headers, to add BHT information as another category of data for the WV candidates
- the subset with BHTs plus stratigraphic data available became the adopted well data set

State-specific information sources:

- New York State
  - ESOGIS well files
    - formation ID's that needed de-coding, and their lithologies:

	James Leone, NYSGS suggestion for Rickard identification scheme	Rickard 1964 usage
Irondequoit		not used (seems inappropriate as this name used for a Silurian formation)
DK	Dunkirk shale	Dunkirk shale
PC	Pipe Creek shale	Pipe Creek shale
SB	Scraggy beds, marks Rhinestreet/Angola contact	not used
J	marker bed within Rhinestreet	not used
CO	marker bed within Rhinestreet	not used
BB	marker bed within Rhinestreet	not used
RG	Rhinestreet group?	Roricks Glen
DH	Devonian Hamilton (not sensible given stratigraphic order in which DH occurs)	Dunn Hill
CQ	Cashaqua shale	Cashaqua
SG	Sonyea Group	Sonyea Group (Cashaqua underlain by Middlesex)
M	Middlesex	Middlesex
GG	Genesee Group	Genesee Group
WR	West River shale (upper	West River

	formation in Genesee Group)	
PY	Penn Yan shale	Penn Yan
HP	a marker within the Genesee group	not used
G	Geneseo shale	Geneseo
TULLY-GILBOA		perhaps a lithologically mixed siliciclastic-limestone

  - Sources consulted for lithologic information:
    - Hill, Lombardi, Martin, Fractured shale gas potential in New York: NYSERDA
    - Smith, G., 2002, Conneaut sequence, NYSERDA
    - Lugert et al., NYSERDA report
    - Young, W.H., Jr., and Krediler, W.L., 1957, NYSGA
    - Rickard, 1964, New York State Museum and Science Service Geological Survey, Map and Chart Series, no. 4
    - NY DEC SGEIS (http://www.dec.ny.gov/docs/materials_minerals_pdf/ogdsgeischap4.pdf)

- Pennsylvania wells
  - file of formation tops provided by Michele Cooney of PA geological survey
  - Sources of lithologies:
    - U.S. G.S. Mineral Resources On-Line Spatial Data (for example, http://mrdata.usgs.gov/geology/state/fips-unit.php?code=f42115)

- West Virginia wells
  - Pipeline (online data management system)
    - http://www.wvgs.wvnet.edu/oginfo/pipeline/pipeline2.asp
  - Sources of lithologies:
    - U.S. G.S. Mineral Resources On-Line Spatial Data (for example, http://mrdata.usgs.gov/geology/state/fips-unit.php?code=f54015)

Part 2: Analysis

Well Locations & Sediment Depth

The wells for this analysis are located as shown in Figure 1. In total 77 of the original 78 wells were used because there was not a bottom-hole temperature (BHTs) for one well. One of the wells in West Virginia had two BHTs so these were analyzed separately. Figure 2 compares the sediment depth from the sediment thickness map versus those from the detailed stratigraphy for wells that penetrated the basement. Figure 2 shows that the map sediment depth is generally very close to the true sediment depth.

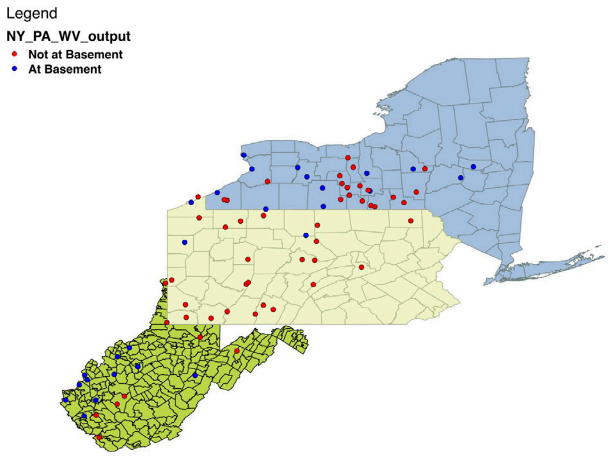


Figure 1: Map of well locations for the sensitivity analysis. The points are color-coded with blue being at basement and red not being at the basement.

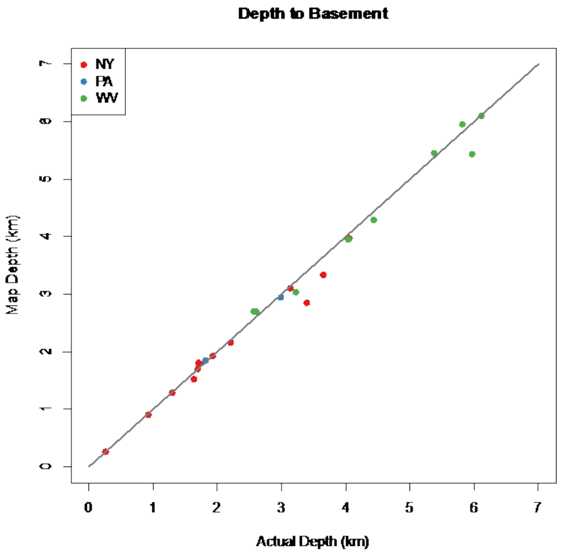


Figure 2: Plot of depth to basement from the sediment thickness map used versus from the depth to sediment for the detailed stratigraphy wells that reached basement. The black line is for perfect prediction (45°). New York wells are in thinner sediments and West Virginia wells are in thicker sediments.

Stratigraphy Sources

The stratigraphic columns can be divided into two types: detailed stratigraphy and COSUNA stratigraphy. More details are provided on both of these types are provided below.

COSUNA Stratigraphy

COSUNA (Correlation of Stratigraphic Units of North America) columns provide an approximate stratigraphy for generally multi-county areas. Jared Smith converted the original COSUNA documents into spreadsheets that contained information for each of the listed units (more details in cu.app-basin-gpfa.us/cu/GIS/COSUNA/COSUNA_Documentation_final.docx). The variables of interest are the thickness and the conductivity variables. The thicknesses used are the "Column Min", "Max", and "Assumed", which are used to define the lower bound, upper bound, and most likely values of a triangular distribution, respectively.

There are multiple values of conductivities depending on which reference values one uses. In this analysis there are two values, the Beardsmore and Cull. and the Carter values. Carter values are for the Anadarko Basin, which is considered a sister basin to the Appalachian Basin. Beardsmore et al. values are essentially standard values for a given rock type, so they will not be as reflective of the burial history in the Appalachian Basin.

Detailed Stratigraphy

For the selected set of wells (part 1), the files contained depth to formation top, conductivity, and conductivity standard deviation. The conductivities are based on the values from Carter (see Memo on conductivities). The thickness of each unit was calculated as the difference of the depth to formation top of the unit below it and its own depth to formation top. In some cases the detailed stratigraphic information was not complete to the basement either because the well did not reach basement or because there were some intermediate layers that with unassigned depths.

When the well did not reach basement, the thickness of the last recorded formation (depth of the formation top recorded) was estimated and the remaining depth to basement was assigned as a single unit. To estimate the thickness of the last recorded formation the depth to the formation top was multiplied by the assumed thickness of the formation in the COSUNA column and divided by the depth to the formation top in the COSUNA column. For instance, if the depth to formation top in the detailed stratigraphy was 2,000 m and the depth to the formation top in the COSUNA column is 3,000 m and the thickness in the COSUNA column is 30 m, then the estimated thickness in the detailed stratigraphy would be (2,000 * 30 / 3,000) = 20 m. The remaining thickness between the filled-in thickness and the basement is assigned a single thickness unit. For instance, if there were 500 m of missing thickness between the last formation top and the basement (evaluated from sediment thickness map), and the last formation was filled-in with 20 m of thickness, then 500 – 20 = 480 m was assigned as a single unit.

There were a few special cases for filling-in the depths. First, in some instances the thickness of the detailed stratigraphic column exceeds the thickness of the sediments from the map (depth to basement map layer). If the estimated thickness of the last formation layer caused the total thickness to be greater than the map sediment thickness, then the missing unit's thickness was set to the difference between the formation top and the map sediment thickness. Secondly, when only the last unit has unknown thickness (for instance the Potsdam in much of NY), then its thickness was set to the difference of the map sediment thicksness and the formation top. Thirdly, if the detailed stratigraphy divided individual groups (several units) into smaller units but the COSUNA column only listed only the group, then the thickness of the group was estimated and the missing unit thickness was the group thickness less the thickness of the other units in the group that were recorded. So if the estimated thickness of the group was 400 m and there were three units in that group, two with known thicknesses from the detailed stratigraphy of 100 m and 125 m, then the estimated thickness of the unknown unit would be 400 – 100 – 125 = 175 m.

In a few cases there were missing thicknesses of intermediate units (defined formation tops above and below the units, but not of the unit or units in question). The first method of addressing the missing intermediate layers was to calculate the total missing thickness of the layers and then multiplying the total thickness by the percentage of thickness for each layer in the assumed COSUNA stratigraphy. When the COSUNA stratigraphy did not provide enough information detailed stratigraphy columns in the same COSUNA section were used to estimate the missing intermediate thicknesses based on the percentage of the thickness.

For each unit we assigned values for the conductivity and the standard deviation of the conductivity. These conductivities are based on the Carter conductivities from the Anadarko Basin, because those strata underwent similar extents of burial and are roughly as old as is the Appalachian Basin.

Distributional Assumptions

This section outlines the distributional assumptions for the parameters in the Monte Carlo experiment. UB, LB, and ML stand for upper bound, lower bound, and most likely (peak), respectively. Most distributions were chosen to be symmetric triangular distributions because they are bounded on reasonable ranges (no negative values) and they reasonably describe a peaked distribution. Most of the COSUNA thicknesses were also symmetric, but they were skewed when the column minimum thickness was not in the same range as the maximum thickness.

Note that the uncorrected bottom-hole temperature (BHT) was used because the BHT corrections are not finalized at the moment and the verification that the model can reproduce the BHT down the borehole does not depend on the BHT measurement.

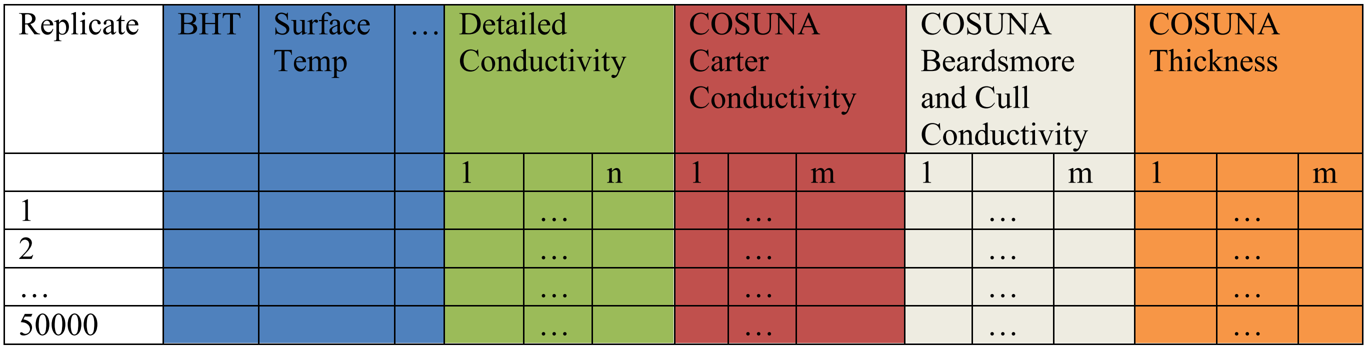
Variable (units)	Distribution	Parameters	Notes
Bottom-Hole Temperature (°C) [uncorrected]	Triangular-Symmetirc	UB = BHT + max(5, 0.1*BHT) LB = BHT - max(5, 0.1*BHT)	Shallow data often has large spread, hence the 5°C minimum; Uncertainty increases with depth because of BHT correction uncertainty
Surface Temperature (°C)	Triangylar-Symmetric	UB = ST + 1 LB = ST - 1	Bounds set as +/- 1°C from the map value
Mantle Heat Flow (mW/m^2)	Triangylar-Symmetirc	UB = 30*1.2 LB = 30*0.8	Mantle heat flow bounds are approximately the expected range
Radiogenic Heat	Triangular-	UB = 1*1.2	Typical value is about 1, used

Production ($\mu W/m^3$)	Symmetirc	LB = 1*0.8	20% as the bounds
Detailed Stratigraphy Conductivity *k* (W/m-°C)	Triangular-Symmetric	UB = k + 2*SD(k) LB = k - 2*SD(k)	Using +/- 2 standard deviations (SD) of the conductivity
COSUNA Stratigraphy Conductivity *k* (W/m-°C)	Triangular-Symmetric	UB = 1.4*k LB = 0.6*k	Using +/- 40% of the conductivity for the bounds because for the Carter values the standard deviation is about 18% of the mean conductivity
COSUNA Thickness (m)	Triangular	UB = Max LB = Column Min ML = Assumed	Used column min and maximum values to defined bounds and the assumed value should be the most likely

Monte Carlo Experiment

The Monte Carlo experiment was designed to test whether there are any systematic differences in the COSUNA approximations and the detailed stratigraphy. This section outlines the generation of the replicates for all of the wells.

For a well the replicates of for the detailed stratigraphy, COSUNA Carter, and COSUNA Beardsmore and Cull were all generated at once. Figure 1 represents the generation of the data. When possible, all of the random inputs were kept the same across the different stratigraphy assumptions. For instance, all of the Monte Carlo replicates for a single well used the same set of BHTs and surface temperature inputs. Additionally, both of the COSUNA variations used the same thickness values. Holding as many parameters the same across the variables allows for a paired test, which should have higher power.



Replicate	BHT	Surface Temp	…	Detailed Conductivity			COSUNA Carter Conductivity			COSUNA Beardsmore and Cull Conductivity			COSUNA Thickness		
				1		n	1		m	1		m	1		m
1					…			…			…			…	
2					…			…			…			…	
…					…			…			…			…	
50000					…			…			…			…	

Figure 3: Representation of the generation of the generation of the replicates for the Monte Carlo experiment. The colors are for different blocks of data. There are n units in the detailed stratigraphy and m units in the COSUNA stratigraphy.

Standard uniform variables were generated for all of the variables. The standard uniform variables were then converted into the distribution (see section "Distributional Assumptions") using the inverse cumulative distribution function for the variable. The seed for each well is separate and based on the depth of the well, which allows reproducibility of the results without causing all of the random variables across the wells to be linked.

All of the output was calculated based on code developed by Jared Smith and Frank Horowitz (bitbucket.org/geothermalcode/jaredthermalconductivity, special functions for sensitivity analysis are in the branch ‘calvinSA’).

Individual Well Statistics

The statistics for an individual well are based on the 50,000 Monte Carlo replicates for the three stratigraphy assumptions of that well. The statistics calculated for the individual wells are:

- Mean (average, measure of location)
- Median (middle of the sorted values, robust measure of location, 50th percentile)
- Standard Deviation (measure of spread)
- Interquartile Range (IQR, robust measure of spread, difference of the 75th and 25th percentiles)

These statistics include both standard and robust measures of location and spread. The units of all of these statistics will be the same as the units of the original output variable.

For the analysis the surface heat flow and the temperature at 3 km are considered. The surface heat flow depends on the BHT, surface temperature, and the “average” conductivity between the BHT and the surface (the conductivity is calculated using a harmonic average accounting for thickness of formation). Temperature at 3 km represents a reasonable estimate of the depth range considered for development.

Figure 2 shows some sample boxplots of the output of the Monte Carlo experiment for different wells. The top two boxplots are for Surface Heat flow and the bottom two boxplots are for temperature at three kilometers.

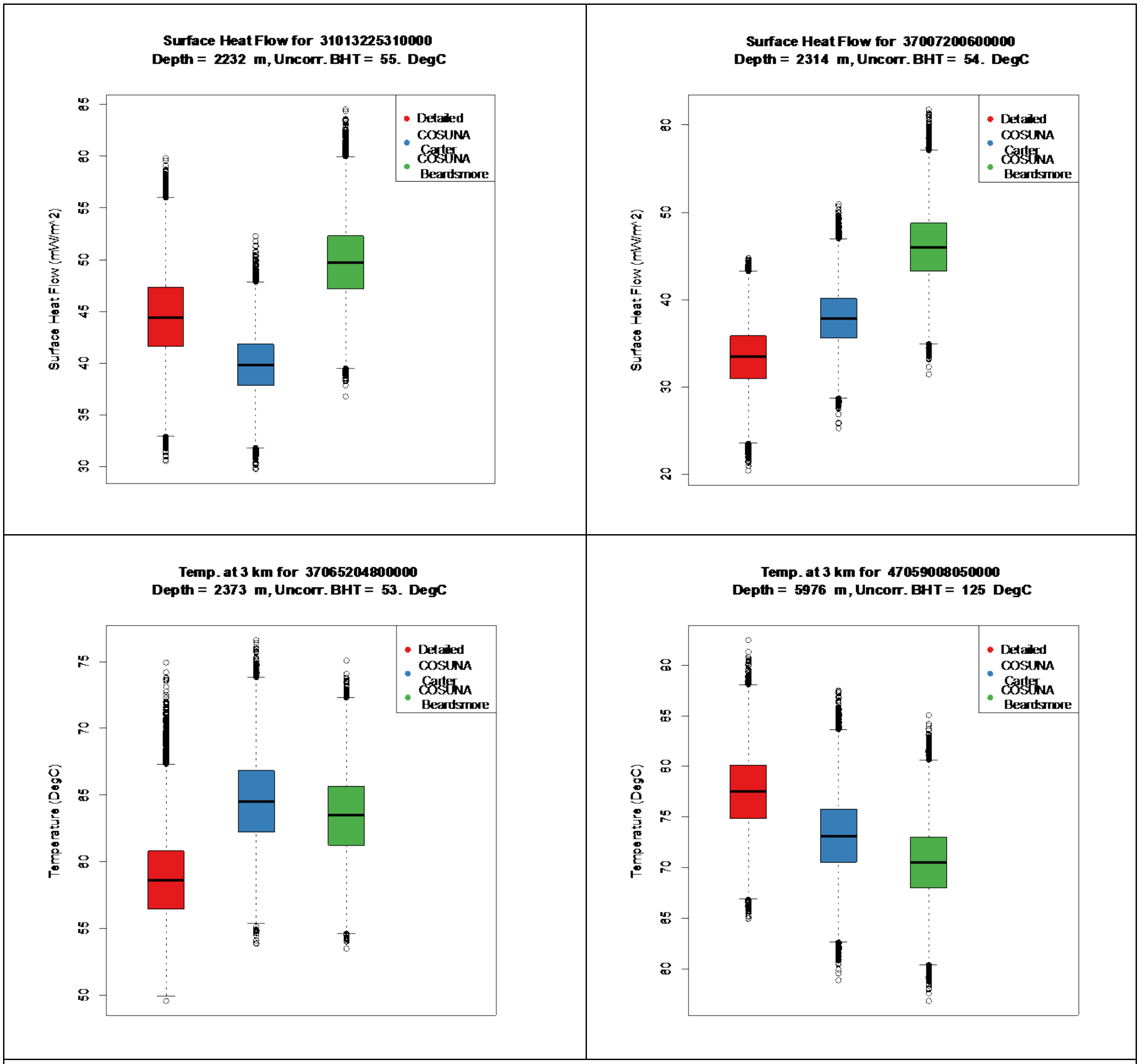


Figure 4: Examples of boxplots showing boxplots of the distribution of the results of 50,000 Monte Carlo replicates for each well stratigraphy/conductivity assumption (each boxplot is 50,000 points). Red is for the detailed stratigraphy, blue is for the COSUNA stratigraphy with Carter conductivities, and green is for the COSUNA stratigraphy with Beardsmore and Cull conductivities. The box is defined from the 25th to 75th percentiles with the middle line at the median (50th percentile). The whiskers extend up to 1.5 times the interquartile range.

Region-Wide Analysis

The goal of this section is to examine whether there were any systematic differences when using the stratigraphic assumptions for surface heat flow or temperature at 3 km. The main questions addressed are:

1. Are the any large systematic biases?
2. Are there any large differences in uncertainty and how does depth impact this?
3. Robustness in predicting temperature at depth?

Systematic Biases

Figure 5 shows plots of the mean and the median surface heat flow for the wells. The mean and the median for each point are calculated based on the Monte Carlo replicates for that point. Generally, the points seem to be clustered around the perfect prediction line (in black). If there is a systematic bias it is probably minor around 2-5 mW/m^2. Comparing the two plots in Figure 5 also shows that the distributions are fairly well behaved because the mean and median plots look very similar, indicating fairly symmetric distributions.

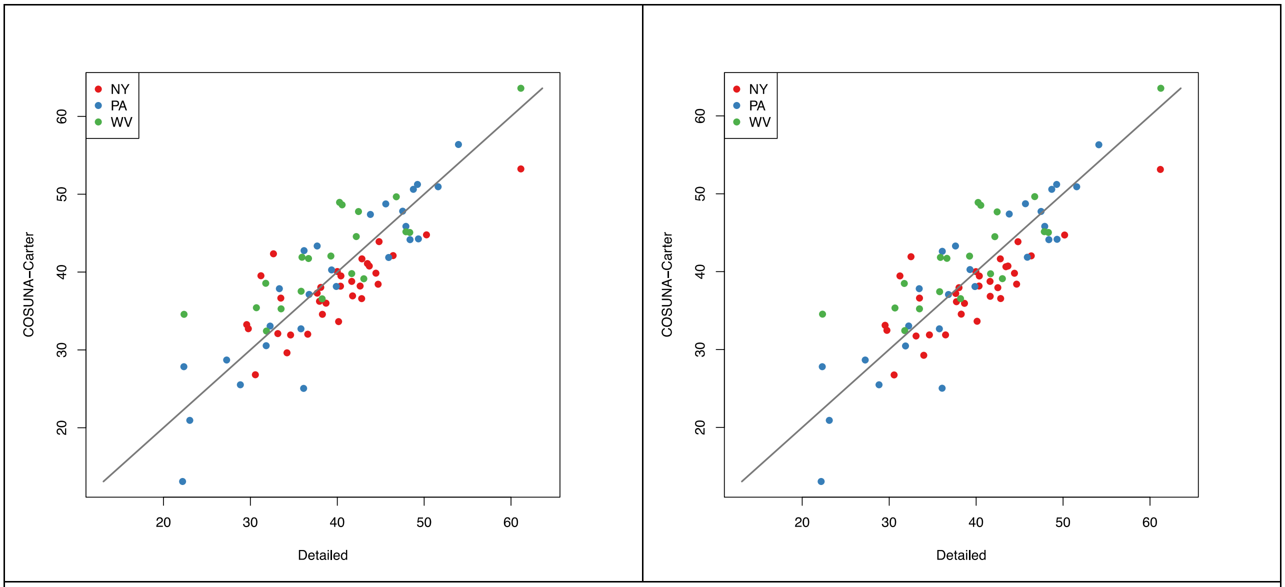


Figure 5: Plots of mean (left) and median (right) surface heat flow when using Detailed Stratigraphy and COSUNA Stratigraphy with Carter Conductivities. Points are color-coded by state (NY=red, PA=blue, WV=green). The black line is the 45° line for perfect matches.

Differences Uncertainty (Spread)

Figure 6 plots the uncertainty (spread) of the distributions of the Detailed Stratigraphy and COSUNA Stratigraphy with Carter conductivities against each other. Generally, the two plots look very similar, which is an indication that the distributions are well behaved and do not have very fat tails. If the distributions were perfectly normal, then the interquartile range would be about 1.35 times the median, which explains the difference in scale of the two figures. The measures of spread seem to be noisier around the prediction line than the mean or median results, but they still seem to be clustered around the line.

Figure 7 shows that as the depth of the bottom-hole temperature (BHT) increases, the uncertainty in the surface heat flow decreases until around 2000 m, at which point it stabilizes. Part of the reason for this behavior is probably due to the assumption of the BHT distribution, which was fairly wide even at shallow depths to reflect that shallow data is often very noisy. As the BHT becomes deeper, the bounds are specified as a percentage of the BHT value, which means the uncertainty is instant relative to the BHT value.

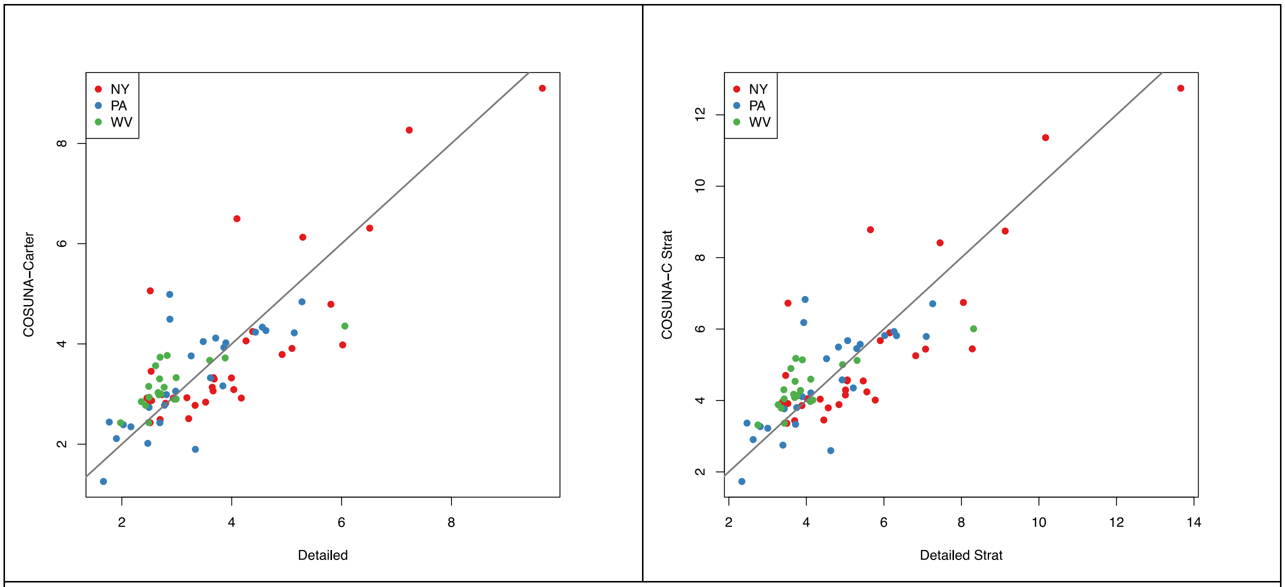


Figure 6: Plots of standard deviation (left) and interquartile range (right) surface heat flow when using Detailed Stratigraphy and COSUNA Stratigraphy with Carter Conductivities. Points are color-coded by state (NY=red, PA=blue, WV=green). The black line is the 45° line for perfect matches.

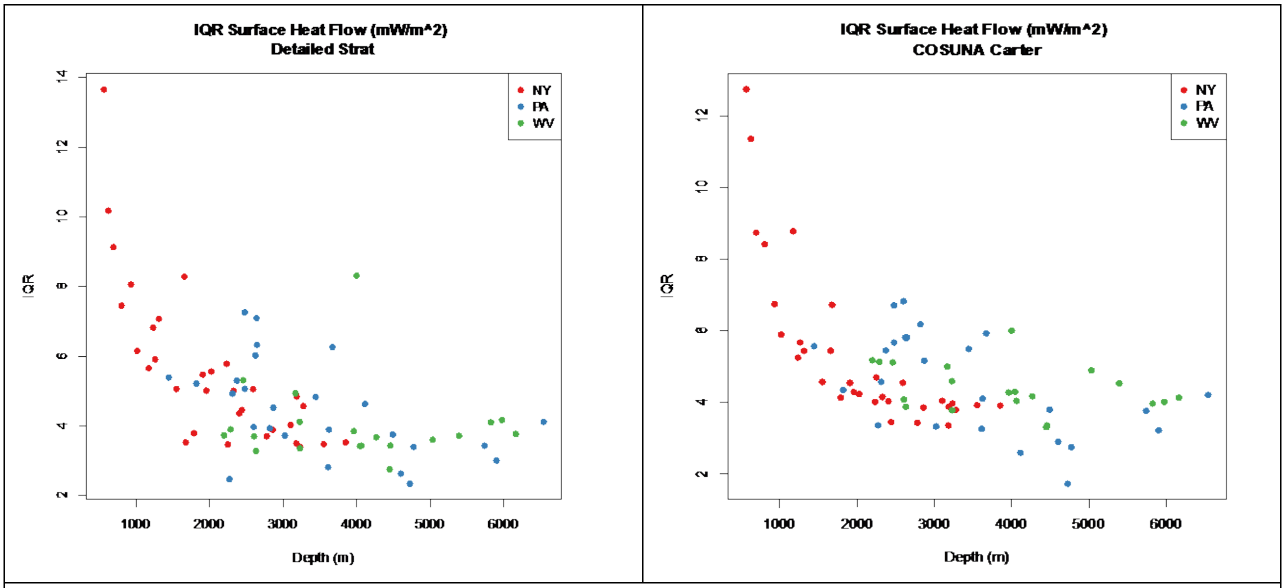


Figure 7: Plots of interquartile range (IQR) of surface heat flow for the Detailed Stratigraphy (left) and the COSUNA Stratigraphy with Carter conductivities (right). The horizontal axis is

the depth of the BHT measurement. Points are color-coded by state (NY=red, PA=blue, WV=green). The black line is the 45° line for perfect matches.

Robustness in predicting temperature at depth

Figure 8 shows the differences in the predicted mean temperature at 3 km for the wells in the Monte Carlo study (mean of 50,000 replicates) for the assumptions of Detailed Stratigraphy and COSUNA stratigraphy with Carter conductivities. Generally, the COSUNA-Carter approximation is very robust in the sense that the estimated temperature at depth is within 6 °C of the Detailed Stratigraphy estimation. It is difficult to determine if there is any difference in spread when the Detailed stratigraphy is known at the depth of estimation (BHTs deeper than 3 km) versus when only the upper portions of the detailed stratigraphy are known and missing units are appended to the detailed stratigraphy (BHTs less than 3 km).

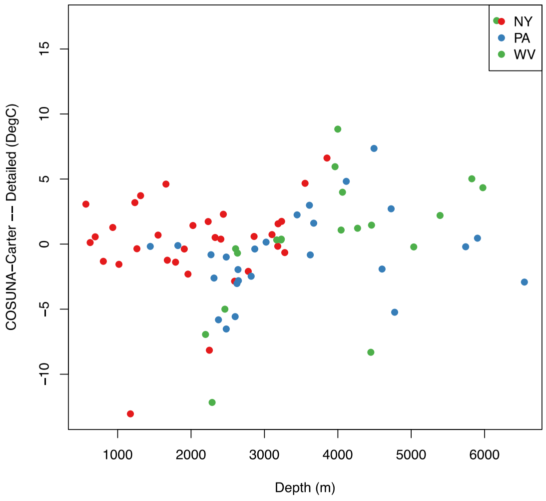


Figure 8: Plots of the difference in the predicted temperature at 3 km based on the depth of the BHT measurement. When BHT depth is greater than 3 km the detailed stratigraphy is know. When the BHT depth is less than 3 km only the upper portions of the detailed stratigraphy are known and the lower portions are assumed.

References

Beardsmore, G.R., and J.P. Cull. (2001). Crustal heat flow: A guide to measurement and modelling. Cambridge University Press.

Carter, L. S., Kelley, S. A., Blackwell, D. D., and Naeser, N. D., 1998, Heat flow and thermal history of the Anadarko Basin, Oklahoma: AAPG Bulletin, v. 82, no. 2, p. 291-316.