The Spatiotemporal Visualization of Historical Earthquake Data in Yellowstone National Park Using ArcGIS

> GIS and GPS Applications in Earth Science Paul Stevenson, pts453 December 2013

Problem Formulation

There exists a tremendous amount of seismic data in the area of Yellowstone National Park. I desire to display this data using ArcGIS software in a way that clearly displays the spatial and temporal distribution of the data. I think that this will allow the viewer to notice several important patterns. First, the lack of periodicity and therefore lack of predictability of earthquakes. Second, temporal groupings, including earthquake swarms that are extensively discussed in literature regarding Yellowstone. These temporal groupings should reflect several laws of earthquakes including Omori's time decay law, Bath's empirical law, the Gutenberg-Richter law, and the after-shock geographic density with distance law.



Source: http://www.gccapitalideas.com/2012/04/25/spatial-and-temporal-earthquake-clustering-part-2-%E2%80%93-earthquake-aftershocks-probabilistic-hazard-analysis-temporal-and-spatial-properties/

Finally, the viewer should be able to observe spatial relationships between the caldera, the known Yellowstone fault locations, and earthquake epicenters. Using several tools available in ArcGIS it should also be possible to construct some sort of earthquake density map that should be able to reveal to some degree the locations of the most active fault zones. If the data is partitioned temporally, changes in the most active portions of the park should be clearly visible.

Data Collection

Quite a lot of seismic data is available for the Yellowstone area. I chose to use the data from <u>http://www.yellowstonegis.utah.edu/data/seismicity3d/index.html</u>. This choice was made after reading <u>http://www.yellowstonegis.utah.edu/data/seismicity3d/BSSA_Husen_Smith_2004.pdf</u> which explains the methods used to create the extensive database of some 25,267 earthquakes in the date range from 1972 to

2003. This data set takes into account the highly unique seismic velocity properties of the Yellowstone caldera area, a variable that has been previously ignored according to Husen and Smith. The result their analysis is tighter clustering of epicenters and focal depths as compared to the original data. This is important for my analysis of spatial location.

Other data used in the project is located at the Wyoming State Geological Survey GIS Data Center, <u>http://www.wsgs.uwyo.edu/Research/Yellowstone/GIS.aspx</u>. From this data set I used the caldera shapefile, the boundaries shapefile, the faults shapefile, and the geologic units data. The United States map for the inset map was obtained from lab 2 data file Mapproj.mdb, USA, states.lyr.

Data Preprocessing

In order to use the Tracking Analyst tool of the ArcGIS software package a date field must be generated with the format YYYYMMDDHHMMSS. This proved to be quite a stumbling block as the dates and times in the attribute table of the Yell_seis_3d_7203.shp are in separate fields and lack leading zeros, 1 instead of 01 and so on.

	A	В	С	D	E	F	G	Н	Ι	J	Kι	M	NC	P	QF	S
1	LONGITUD	LAT,N,19	YR,N,▶	MON	DAY,	MAG,	DEPTH,▶	HR,▶	MIN	SEC,N	P D	0	ÞD	•	ÞÞ	• 0
2	-110.6966	44.5846	1972	11	8	2.9	-2.94	9	54	27.3	0#	##	##	##	##	0
3	-110.7885	44.2406	1972	11	30	2	-3.95	13	19	47.7	0#	##	##	##	##	0
4	-110.7817	44.9086	1973	1	4	1.1	7.29	1	33	7.35	0#	##	#0)#	##	₽
5	-110.7959	44.5888	1973	1	5	1.3	7.7	17	43	7.83	##	##	#0)#	##	₽
6	-110.4516	44.8029	1973	1	6	1.4	10.13	13	33	5.34	0#	##	##	##	##	0
7	-111.1641	44.5317	1973	1	7	0.2	-1.78	2	10	31.43	0#	##	##	##	##	0
8	-111.5975	44.7915	1973	1	7	1.7	-3.9	6	55	50.82	0#	##	##	##	##	0
9	-110.8323	44.7544	1973	1	8	0.3	6.05	12	9	57.64	0#	##	#1	#	##	0
10	-110.8052	44.7929	1973	1	9	1.7	-3.97	14	42	57.3	0#	##	##	##	##	0
11	-110.4847	44.6859	1973	1	11	1.4	-3.85	9	59	31.26	0#	##	##	##	##	0
12	-111.5251	44.8029	1973	1	12	2.3	10.54	6	46	29.99	0#	##	##	##	##	0
13	-111.064	44.703	1973	1	13	1	5.95	2	21	37.35	0#	##	#0)#	##	₽
14	-111.0658	44.7044	1973	1	13	1.4	7.42	3	5	29.75	0#	##	#1	#	##	₽
15	-110.8378	44.7673	1973	1	13	0.8	3.97	18	42	15.47	0#	##	#1	#	##	0
16	-111.3179	44.7544	1973	1	14	1.5	11	3	41	55.71	0#	##	##	##	##	0
17	-111 0404	44 6987	1973	1	14	2	8 19	16	13	43 39	0#	##	##	##	##	0

Original state of the data

I attempted a number of different methods within ArcGIS, utilizing the field calculator along with python code to modify the data then combine the data fields in the proper manner. After many failed attempts I was forced to use Open Office Calc to modify the data in the appropriate manner. I first moved each column individually to its own sheet, I then converted all numeric fields to text.



I then used the find and replace function to replace all single digit numbers from 0-9 with their respective two digit formats 00-09. I also used Open Office Calc to round all of the seconds data to integers. Modifying or even worrying about the seconds data proved to be worthless for this particular project, but for future data applications it might be nice to have the whole date time field. Each column was then copied and pasted back into an empty column in the original database (this performs essentially the act of "Add a Field" in ArcGIS). With the data marked as text I was able to use the concatenate function of Open Office Calc to combine the data in a new YYYYMMDDHHMMSS column.

𝑘 𝔅 = =CONCATENATE(U2; V2; W2; X2; Y2; Z2)

Note that this function, though available in ArcGIS python code, does not seem to work with field types designated as numbers or dates. Before saving the newly modified database make sure to format the cells of the new YYYYMMDDHHMMSS column to number format, otherwise ArcGIS will have a problem using them as dates.

	Α	В	C	D	E	F	G	Н	Ι	J	KI	MNO	PQR	ST	U	V	W	Х	γ	ZA	A		AB		
1	LONGITUD	LAT,N,19	YR,N,▶	MON	DAY,	MAG,	DEPTH,	HR,▶	MIN▶	SEC,N	► P D	• • •		• •	YRTX	Mo	DA	H0	MI▶	SP'	Y	(YYM	MDD	HHM	MSS
2	-110.6966	44.5846	1972	11	8	2.9	-2.94	9	54	27.3	0#	####	###	•#	1972	11	80	09	54	27	19	72110	8095	427	
3	-110.7885	44.2406	1972	11	30	2	-3.95	13	19	47.7	0#	###	###	•#	1972	11	30	13	19	48	19	72113	0131	948	
- 4	-110.7817	44.9086	1973	1	4	1.1	7.29	1	33	7.35	0#	###0	###	₽#	1973	01	01	01	33	00	19	73010	1013	300	
5	-110.7959	44.5888	1973	1	5	1.3	7.7	17	43	7.83	3##	###0	###	₽#	1973	01	01	17	43	00	19	73010	1174	300	
6	-110.4516	44.8029	1973	1	6	1.4	10.13	13	33	5.34	0#	###	###	•#	1973	01	01	13	33	00	19	73010	1133	300	
7	-111.1641	44.5317	1973	1	7	0.2	-1.78	2	10	31.43	80#	###	###	•#	1973	01	01	02	10	31	19	73010	1021	031	
8	-111.5975	44.7915	1973	1	7	1.7	-3.9	6	55	50.82	20#	###	###	•#	1973	01	01	06	55	51	19	73010	1065	551	
9	-110.8323	44.7544	1973	1	8	0.3	6.05	12	9	57.64	0#	###1	###	•#	1973	01	01	12	09	58	19	73010	1120	958	
10	-110.8052	44.7929	1973	1	9	1.7	-3.97	14	42	57.3	80#	###	###	•#	1973	01	01	14	42	57	19	73010	1144	257	
11	-110.4847	44.6859	1973	1	11	1.4	-3.85	9	59	31.26	60#	###	###	•#	1973	01	01	09	59	31	19	73010	1095	931	
12	-111.5251	44.8029	1973	1	12	2.3	10.54	6	46	29.99	0#	###	###	•#	1973	01	01	06	46	30	19	73010	1064	630	
13	-111.064	44.703	1973	1	13	1	5.95	2	21	37.35	0#	###0	###	₽#	1973	01	02	02	21	37	19	73010	2022	137	
14	-111.0658	44.7044	1973	1	13	1.4	7.42	3	5	29.75	0#	###1	###	₽#	1973	01	02	03	05	30	19	73010	2030	530	
15	-110.8378	44.7673	1973	1	13	0.8	3.97	18	42	15.47	0#	###1	###	•#	1973	01	02	18	42	15	19	73010	2184	215	
16	-111.3179	44.7544	1973	1	14	1.5	11	3	41	55.71	0#	####	###	•#	1973	01	02	03	41	56	19	73010	2034	156	
17	-111.0404	44.6987	1973	1	14	2	8.19	16	13	43.39	0#	####	###	•#	1973	01	02	16	13	43	19	73010	2161	343	
18	-111.0404	44.6987	1973	1	14	1.8	8.25	19	1	18.43	0#	###	###	•#	1973	01	02	19	01	18	19	73010	2190	118	

Fully Modified Database

Between the trial and error of trying to get ArcGIS to perform these steps and the actual work in Open Office Calc this portion of the project took quite a lot of time to accomplish with such a large data set. **ArcGIS Processing**

Yellowstone National Park is contained in its entirety within UTM zone 12. This was therefore chosen as the coordinate system for this project. Projections, using the Project (Data Management) tool were performed on the seismic data and the states.lyr file. The rest of the data was already projected with this coordinate system.

An initial map designed to display the entire data set was constructed with the projected Yellowstone boundary shapefile, the projected seismic shapefile, the geologic units shapefile, the caldera shapefile, and the faults shapefile. Symbolization of each layer was performed, highlighting the faults in yellow and the caldera boundaries in red. The geologic shapefile was symbolized using the layer file contained in the yellowstone.zip file procured from the Wyoming State Geological Service. A separate full page legend was constructed for the geologic layer. DEM and hillshade layers were not used as they muted the colors of the geologic layer. Most of the details of the surficial geology will not be visible in this project, but for those interested the geological units layer is very well done.





This map illustrates only the spatial distribution of the entire seismic data set. It is useful in visualizing the cumulative grouping of the data over a time span of approximately 30 years. From this view we are able to see that the most seismically active areas over this time period are in the Northwest areas of the park. Interestingly, we can also see that this area lacks surficial faulting to a large degree. From this view one might surmise that a more extensive set of subsurface faults exists in that area. Another interesting observation to be gained from this view is the pattern wherein the grouping of the seismic data follows the contours of the caldera, in some locations grouping quite significantly near the caldera edges and showing a distinct break at the edge. This could provide insight into the nature of the stack of concentric "columns" of collapsed roof.

One of the strange things revealed by this visualization is the apparent linear nature of some of the data, particularly a distinct line to the east, two distinct lines to the west and one to the north. I believe this to be an artifact of the temporal and spatial analysis bounds placed during the re-analysis of the data performed by Husen and Smith. I believe that it stems from the reorganization of the seismic instrumentation taking place in 1981, then 1984, and again in 1995. This may point to a serious inconsistency in the analysis or merely to the need for the analysis to be run with a larger spatial extent or different time bins. To double check that this was not an error in conversion to shape files I used the text file available from the website, then converted it to a database file with latitude and longitude coordinates, upon adding the XY data to a map in Arcmap (Using; File, Add Data, Add XY Data) I found that the oddity exists within the original data set. Being as far invested into the project at this point I chose to use the data set but limit the spatial extent of the project to exclude the inconsistencies, just as the original authors chose to do on their display.



http://www.yellowstonegis.utah.edu/data/seismicity3d/index.html

We can further explore the spatial aspects of the epicenters, while adding a partial temporal view to the data, by utilizing the point density tools that ArcGIS provides. I first created a new map in ArcMap, added the seismic data shapefile, the faults shapefile, and the caldera shapefile. I then used the Point Density tool to create a raster representing the number of earthquakes per 10 by 10 meter area.

input point reat	tures		^_	Annulus—A torus
EQs90_94_UT	MZ12			(donut shaped)
Population field				defined by an inner
NONE			· · · · · · · · · · · · · · · · · · ·	and outer radius.
Output raster				Circle—A circular
C:\Users\Comp	puterAdmin\ArcGISD	ata\90-4pntdns		neighborhood with
Output cell size	(optional)			the given radius.
10				This is default where
Neighborhood (optional)			shortest of the width
Circle	•			or height of the
Neighborhood	Settings			extent of the input
Radius:	4236.228123			point features, in the output spatial reference, divided by
Units:	Cell	Map		30.
				Rectangle—A

This screen shot was taken using a time partitioned version of the main seismic shapefile (the reason for which follows) but the general characteristics used are the same. Input a point file, leave the population field on "NONE" unless you want to use one of the numeric fields recorded with each point as the additive characteristic. Leaving it on "NONE" adds 1 to the raster cell for each point within a certain cell size. I chose an output cell size of 10 (Meters for the UTM) as it outputs a nice relatively smooth looking raster. Beware that the files produced when using such a small cell size are quite large. Don't attempt to use such small cells while saving to a flash drive as the tool will simply fill it then abruptly stop with no warning telling you that you are out of space, the only hint to that fact is the output of a partial raster! I used this tool with and without the Neighborhood, which can be disabled by setting radius to zero. For the default radius the neighborhood function makes no difference in the output because the default is literally the point size on the map.

After seeing the output of the density function, I concluded that the tool could be used to visualize the data to some degree temporally, showcasing the areas with the highest density of earthquakes within temporal bins. I proceeded to take the original seismic data and partition it into 7 bins using the select by attribute function of ArcGIS, exporting each selection to an independent shapefile.



I added each of these temporally binned shapefiles, within their own labeled data frame, into the original map, each with their own faults and caldera shapefiles. I used the point density function to create density rasters for each set of data. By modifying the symbology, and making both the density raster and the points partially transparent I was able to construct a set of figures that represent the five year binned densities of earthquakes in the Yellowstone area. In the process of this analysis I discovered the reason for the odd lines in the data. When I created the 1985-89 density raster I discovered that it covered far less of a spatial extent. This is a direct result of the 1984 reconfiguration/reduction of the seismic array used in the area. Image Source: Husen and Smith



Figure 2. Seismic network layouts for three-different time periods in the Yellowstone National Park region. Seismic stations are shown by different symbols for different time periods. Black squares denote three-component seismometers; all other symbols denote one-component seismometers.

The required variance in the number of years per bin in these images distorts the impact of their side by side portrayal to some degree. In order to help combat this distortion the standard deviation classification scheme was used to symbolize the density rasters. This maintains the purpose of showing maximum density areas in dark red while negating the numerical impact of adding uneven numbers of years (improper binning) resulting, theoretically, in skewed high numbers of concentrated earthquakes per cell in the larger bins and lower numbers per cell in the smaller bins. The changes made to the instrumentation over the years is much more difficult to constrain statistically but should only affect the western portions of the images as that is where the majority of the stations were removed from in 1984.

The resulting map, which is available on the next page, displays the earthquake "hot spots" per binned time period. A couple of interesting observations include the extremely active period of time from 1975 to 1979 with large portions of the park showing spatially concentrated groups of earthquake epicenters. This period of park wide high concentration activity tapers off in the 1980 to 1984 period. The period from 1985 to 1989 is very quiet. When the activity picks back up its pace from 1990 to 1999 we see that most of the density is concentrated solely on the northwest fault zone. Finally, we see a rather pronounced concentration of earthquakes from 2000 to 2003, also generally located on the northern fault zone. Many conclusions could be drawn from this portrayal of the data.

My original intent was to overlay temporally defined interferograms on top of some sort of spatiotemporal portrayal of the seismic activity. If the data could be found it might be interesting to overlay it onto a similar visualization as this, comparing times of uplift to times of subsidence with which fault zones are active as well as levels of activity. Unfortunately much of the data is not available to the public so far as I have been able to find.

I have also included a legend of sorts in order to show the individual raster cell value ranges which are of interest to the viewer but far too large to fit on the original display page.





As a final portrayal of the data I decided to use the Tracking Analyst tool of ArcGIS to create an animation or movie of the earthquake data. After the processing of the data is finished, adding a useable time field, the process of creating an animation is actually very simple. The base map created at the beginning of this document was used as a background. Under the Customize-Toolbar menu in Arcmap turn on the Tracking Analyst toolbar.

Tracking Analyst 🕶 | 🕒 | 🕟 | 🥎 🖕

Press the yellow circle with the plus sign to Add Temporal Data. In the wizard that comes up leave the "Copy all data to memory," also leave the "A feature class or shapefile containing temporal data" checked. Set the "Data Source" to your seismic layer file. Set the "field containing the date/time" to the field that was created in Open Office Calc. The software claims that this is a text field, but without making sure to designate the field as numbers using the format cells dialog in Open Office Calc I was unable to get this to work. For the purposes of earthquake analysis, tracks was left on the <none> setting. Press Next.



The next wizard window allows you to specify the exact time format, using the "Time field format:" select the yyyMMddHHmmss format, and note that many other formats would have worked but this one seemed to be the most complete and simplest to compose in Open Office Calc. Press Next.



After a little processing is done by the program you will notice that the play button on the tracking analyst toolbar is no longer grayed out. Press the play button. A window called the "Playback Manager" appears.

Playbac	ck Manager							23
								
Start:	11/ 8/1972 9:54:27 AM	Current:	11/ 8/1972	9:54:27 AM	End:	12/31/2003 1	1:41:33 PM	
[₩][<)[■][→][→] □ Loop	S	lower		 1 1 1	Faster	Options	>>

This playback manager graph contains quite a bit of useful information. In particular the high points represent points in time during which there are a large number of temporally close events. These are commonly called earthquake swarms. Earthquake swarms are a relatively common occurrence in the Yellowstone area. A quick search on Google reveals that one happened just two months ago. Several settings within the playback manager window are useful for visualization purposes, clicking the options button reveals the following.

Playback Manager					23
and and the Marther ster of		adda - A		here have been been been been been been been be	> <
Start: 11/ 8/1972 9:54:27 AM	Current: 1/31/	1977 4:53:11 PM	End: 12	2/31/2003 11:41:33 PM	-
[₩4][4][=][▶][▶]]Loop Slower		<u>]</u>	Faster Options <	<
Set playback window to temporal e	xtent of: Set the p	playback rate:			
Contract of the second		- Current	0.07	Hide Histor	100.01

The setting of playback rate is particularly useful. You can also set the temporal playback rate of multiple temporal layers, this would be useful, for instance, if you had an interferogram that "played" at a different temporal scale than the earthquakes. I have also used this last figure to showcase the current time slider and its corresponding time display. Note that behind the red line is the first significant earthquake swarm. It occurred during the same time period 1975-1979 that we saw such an extreme concentration of activity over the entire park. A graduated symbology based upon the magnitude of the earthquakes has been made, labeling has been set to show the magnitude of the earthquakes with a mask for visibility, and labeling has been activated for the seismic layer. Note that there are several new options available under the Symbology tab of the time enable layer.



Time Window	Only draw even	ts that fa	all into this time wir	ndow	81.		
Events Time Window Advanced Event Opt Most Current Events	Time Window						
Advanced Most Curre	Period: 1 With this opt only those ex	ion, there vents falli	• is no modification ing into this time wi	Units: to how th ndow will	Days e event be drawn	symbols are drav n.	vn, but

The time window is important when you need to control the window or timescale of the display. Using a period of one day the data will be displayed in single barrages of a full 24 hours of earthquakes, all at once. For the animation and temporal distribution resolution that I am using, this is a reasonable setting. However, somebody attempting to show much smaller time periods would need to reduce this value to hours, minutes, or even seconds. Below is one last very important setting for this analysis.

General	Source	Selection	Display	Symbology	Fields	Definition Query	Labels	Time	HTML Popup	Actions	
Choose	temporal :	settings:	📝 Enab	ole time on th	is layer						
Propert	ies		This settin	g controls wh	iich even	ts in this layer are	displayed	1			
Tempor	al Offset		Displa	events in	n the lay	er Event	3000 s and Tra	cks			
			Oispla	y only the mo	ost curre	nt events in the la	yer	_			
				0		/	-0				
			Mos	st Current Eve ay Future Eve	ents ents	Most Ci an	urrent Evo d Tracks	ents			

For the purposes of earthquake temporal spatial relationships I have already shown the cumulative results. For this particular animation I want to show essentially the temporal relationships only, illustrating the sudden onset of an earthquake swarm followed by a relatively slowly decaying set of aftershocks relieving built up stresses throughout the area. I therefore selected "show only the most current events" in the box above.

Saving the map after having all of this completed, and running the map on the same computer seems to be important for ArcMap to realize that this is a time activated layer and allow playback. Otherwise stated these maps do not seem to be able to open and automatically work on just any computer, even copies of the map lose their Tracking Analyst properties.

Supposedly, it is possible to use the animation toolbar to create an animation of this time enabled layer. I consistently receive an error message stating that the map does not contain a time enabled layer. This, despite the fact that the check box for time enabled located under the layer properties, time tab, of the seismic layer, clearly indicates that it is time enabled. Regardless, I prefer to have the current time display and the graph of the playback manager visible during playback. Therefore, I have used the free CamStudio software to record a short demo of the capabilities of ArcGIS Tracking Analyst. Included with this project are two short videos of the earthquake swarms of 1977 and 1999 (The two highest peaks on the playback manager graph). The frame rate required to catch very rapidly passing events yet show any reasonable portion of time requires a lot of disk space. I cannot include a video of the entire sequence of 30 years which totals over 12 gigabytes of video in avi format.

Summary

This project showcases the ability of ArcGIS to visualize both the spatial and temporal characteristics of earthquake data, as well as its ability to reveal inconsistencies or at least oddities in a data set. Using ArcGIS I have illustrated the spatial characteristics of the seismic data collected over a period of thirty years, in particular showing how the data can be used to show seismically active areas of Yellowstone National Park, possibly showing the location of subsurface fault zones. I have also used the software to showcase the change in temporally active areas of the park. Finally, I have used the software to show the purely temporal nature of earthquake swarms. Though not precisely constrained quantitatively, the general relationships Omori's time decay law, Bath's empirical law, the Gutenberg-Richter law, and the after-shock geographic density with distance law, should be visible to the careful observer in the videos presented. ArcGIS is a powerful tool allowing visualization of the vast amounts of data collected in a way that all can understand, I look forward to continued application and use of the general skills that I have learned in this class.