

Modeling Spatio-Temporal Networks with China Historical Geographic Information System

Merrick Lex Berman
China Historical GIS
Center for Geographic Analysis, Harvard University
September 2007

Abstract: The China Historical GIS¹ [CHGIS] is an infrastructural research project that aims to develop a database of all the known administrative units for the Chinese Dynastic Periods, beginning with the Qin (221 BCE) and ending with the Qing (1911 CE). Changes of placename, administrative unit type, parent jurisdiction, and locations are recorded based on documentation drawn largely from *difangzhi* (local histories) and, when available, historical maps. Each of the CHGIS time series records are annotated with source citations and commentaries by historical geography experts at Fudan University's Center for Historical Geography.²

Since the CHGIS project is specifically focused on developing the core database, we have not had very much time to develop additional applications to make use of the datasets. This paper will briefly demonstrate one experiment that combines geocoding of places (from an historical biography) to CHGIS locations, and creating a spatio-temporal network model to represent the movements of one person during the course of his lifetime.

1.1 Geocoding information with CHGIS

The CHGIS database contains information about historical administrative units, including their names, relationships to higher level jurisdictions, and locations. Developing the database—involving the reconstruction of the administrative hierarchy and its internal changes over the course of 2,000 years—has proved to be extremely painstaking and time-consuming. Owing to the focus of the project, CHGIS has not ventured into the gathering of statistical information or other kinds of data that can be related to the administrative units in the database. Since it is impossible to predict the special interests of each individual researcher, or the extent of data they have already gathered, we rely on researchers themselves to relate their data to the administrative units in CHGIS for their own purposes.

To encourage these efforts, the CHGIS staff have undertaken several experimental projects, such as the geocoding of Buddhist sites from the Qing Dynasty national gazetteer,³ and geocoding of Song Dynasty tax quotas.⁴ Building upon these efforts, this paper will describe both the geocoding of events from a single biography, and a process for creating a digital animation of the sequence of events. The resulting spatio-temporal timeline visualizes the physical location of one person during the course of his life. Although this case study (for lack of data) does not extend the process to include other timelines, the same process could be applied to multiple threads of information, creating a visualization of overlapping spatio-temporal networks.

The first step in geocoding involves matching named locations in the source data with a reference gazetteer (such as CHGIS). For this study, the life of Ouyi Zhixu, a Buddhist Master of the Ming Dynasty, was analyzed in depth.⁵ Of course, for our purposes, the information derived from Zhixu's biography focused on placenames, but we also needed to include the major events of the Master's life and determine the sequence in which they occurred. Interestingly enough, this task was done once before by the Venerable Ch'an Buddhist Master, Chang Shengyan, who compiled his findings into a paper map.⁶ (Figure 1)

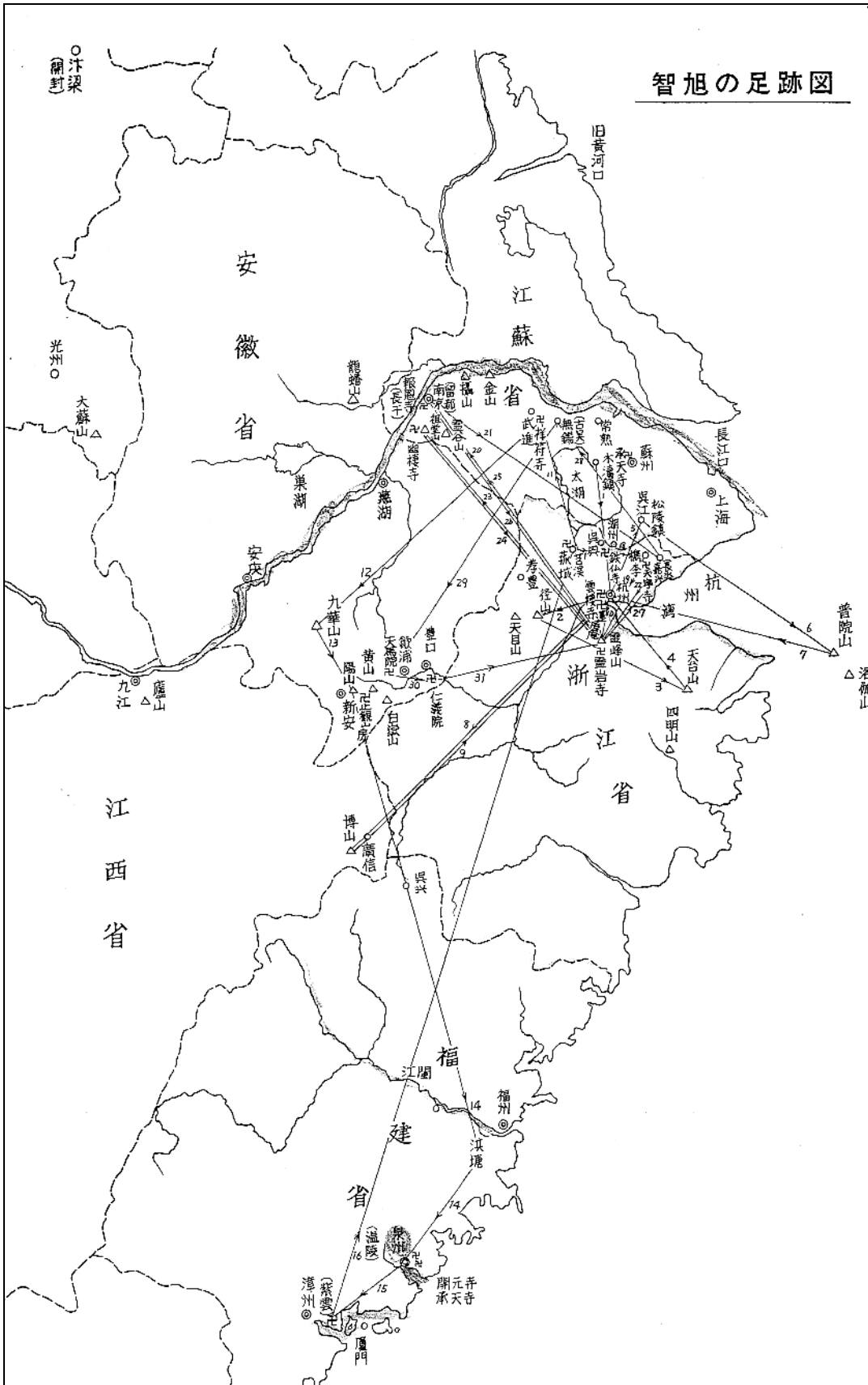


Figure 1: Map of Zhixu's Travels, Master Shengyan (1975)

Shengyan's map features placename annotations and numbered paths indicating the sequence of events in the Zhixu's life. Unique symbols are used to represent towns, mountains, and Buddhist temple sites. One of the challenges in reconstructing a GIS version of this map is to discover what similarities and discrepancies emerge between Shengyan's map and the information derived from Shengyan's thesis and Zhixu's biography.

To begin with, a complete table of the major events of Zhixu's life was compiled. (Figure 2)

place_id	year	season	sequence	source	type_id	type_description	chrono_events_notes
ouyi_1	1599		1	autobio	1	birth	
ouyi_3	1622	summer	2	ZL.8.2.1-8, autobio	10	Yogacara (唯識論)	
ouyi_3	1622	summer	3	ZL.8.2.1-8, autobio	9	received teachings	
ouyi_2	1622	late fall	4	autobio	4	engaged in meditation	
ouyi_6	1622		5	ZL.8.2.1	25	other	first met Xinyi Dazhen 新
ouyi_6	1623		6	選佛譜序	18	divination game	傳燈[選佛圖]隨喜
ouyi_4	1623	spring	7		25	other	謁幽溪傳燈 - pay respe
ouyi_4	1623	fall	8		4	engaged in meditation	

Figure 2: Events Table

Placenames are recorded for each event, and when the table is complete, the contents are grouped by place, so that each place can be given a unique identifier (such as *ouyi_1*). Once this is done the full list of unique places can be normalized into it's own points table and the process of geocoding can begin. In some cases, the historical placename can be related to a particular historical county seat or town record in the CHGIS database. In other cases, further research must be done to find specific mountains, temples, or retreats, using various GIS layers or even satellite imagery to provide contextual information. Finally, the set of unique places must be assigned latitude and longitude coordinates, even if these are only approximate locations. If the points are being matched to CHGIS data, the CHGIS system ID and coordinates can be copied directly to the points table. If derived from other GIS sources, appropriate citations should be included. In some cases, the coordinates were estimated based on locations found in county atlases. All of these sources should be recorded in the points table to reflect the source and accuracy of the spatial data. (Figure 3)

point_id	name	name_ch	province	xian	geo_src_file	cgis_sys_id	x	y	ouyi_copy_notes
ouyi_1	Mu Duzhen	木澧鎮	Jiangsu	Wuxian	v4_1911_twn_pts	138990	120.51	31.253	birthplace
ouyi_2	Jingshan; Qi	徑山	Zhejiang	Xinchang	NIMA	nma_1922389	119.8	30.25	part of Lin'an Xian 临安
ouyi_3	Yunqi si	雲棲寺	Zhejiang	Hangzhou	v4_1911_pref_pts	120998	120.168	30.294	part of Qiantang Xian 乾
ouyi_4	Tiantai shan	天台山	Zhejiang	Tiantai	v4_nima_mtns_pts	nima_31788	121.333	29.5	

Figure 3: Points Table

1.2 Mapping and correcting point locations

After the first pass of geocoding it is always advisable to map the results and see if any obvious anomalies appear that might not be noticed when looking at tables of numeric latitude and longitude coordinates. In this case, the map of all points included some which strayed north of the Yangzi River into the northern parts of Anhui and Jiangsu Provinces. Since Shengyan's map showed all points to the south of the Yangzi River, these points, at least, seemed to be wanting further attention. A second pass over the point locations revealed alternatives that fell within the proximity of Shengyan's map. Further iterations refined the point locations in a way that seemed more reasonable, given the modes and routes of transportation available during Zhixu's lifetime (1599-1655 CE). Finally a map of all the Zhixu points was produced showing a distribution quite similar to Shengyan's interpretation, and for which reasonable matches could be made to either GIS records or points shown on printed maps. (Figure 4)

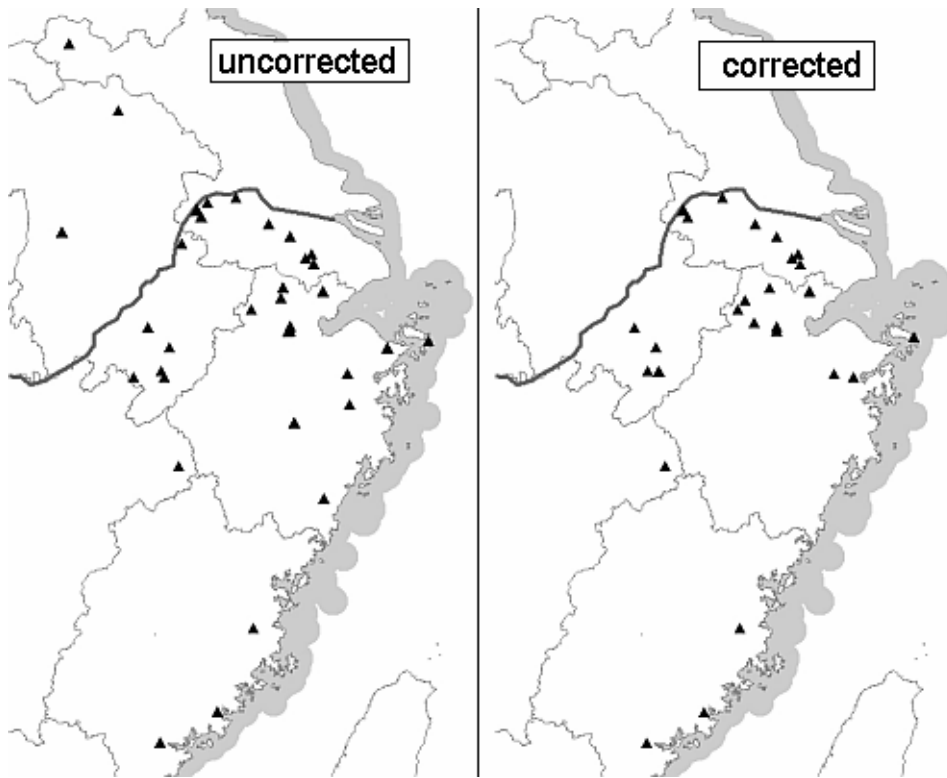


Figure 4: Correction of point locations based on visualization

1.3 Mapping and editing the sequence of points

After the point locations have been fixed the latitude and longitude coordinates can be joined to the sequence of events. In this particular study, each event occurs at a single location. Therefore each record in the events table can simply be joined to its corresponding point using the point_id. Once we have the sequence of events matched to point coordinates we want to develop a means of animating them over time, so that they appear and disappear in the actual order that they occurred.

To animate the sequence of event points we must assign a begin time and end time value to each record. For this purpose it seems natural to use the actual dates of the historical events. This is both possible and highly desirable, if the objective is to model and visualize the relative density of events across time. However, in the case of Zhixu, we find that many events occur clustered together in the same year and we have no exact dates within the year. Therefore, if we were to use YEAR as the unit of temporal granularity, any events that occurred in the same year would appear simultaneously in the animation, rather than in their attributed sequence.

In order to simulate the sequence of events, we therefore used a systematic method of assigning begin times by multiplying the sequence number by a factor of ten. In this way, the point sequence 1, 2, 3 had begin time values of 10, 20, 30, and so on. As for end times, these were created by adding ten to each begin time. Finally the first event (sequence number 1) had its begin time value changed to 1, so that the point appears at the very beginning of the animation. In this way, the event with sequence number 2 was given begin time value = 20 and end time value = 30, while sequence number 3 had begin time value = 30 and end time value = 40, etc. Consequently, the resulting table of points for the animation sequence began with the value = 1 and iterated upwards through the entire sequence without any gaps. (Figure 5)

year	seq	beg_time	end_time	event	point_id	point_x	point_y	placename
1599	1	1	20	birth	ouyi_1	120.51	31.253	Mu Duzhen
1622	2	20	30	Yogacara	ouyi_3	120.168	30.294	Yunqi si
1622	3	30	40	received te	ouyi_3	120.168	30.294	Yunqi si
1622	4	40	50	engaged in	ouyi_2	119.796	30.378	Jing shan
1622	5	50	60	other	ouyi_6	120.16	30.25	Wulin shan
1623	6	60	70	divination c	ouyi_6	120.16	30.25	Wulin shan

Figure 5: Point Sequence Table with Calculated Begin Time and End Time Values

With a continuous time sequence linking the major events being depicted, the Point Sequence Table is now ready for animation. There are two things to keep in mind about the resulting animation: first, the resulting series of points that will appear on the map do not reflect the actual period of time during which they occurred, but are artificially arranged along the timeline to provide an equal amount of time for each event. Second, when a sub-series of events occur at the same location, the resulting point in the animation will appear to linger on the screen for a longer and longer interval, depending on the total number of events in the sub-series. There is certainly room for adjusting the begin and end time values to change the flow of events as they will appear in the animation, but for the purposes of this demonstration we have used the original calculated begin and end time values.

2.1 Creating an animation with TimeMap

Perhaps the only software capable of visualizing GIS data with variable begin and end time increments is TimeMap, which is free for academic use.⁷ With TimeMap, we can simply load up shapefiles and adjust the temporal extent of the spatial objects that are visible by sliding a timebar. (Figure 6) Another convenient feature of TimeMap, which we will use here, is the Generate Animation tool.

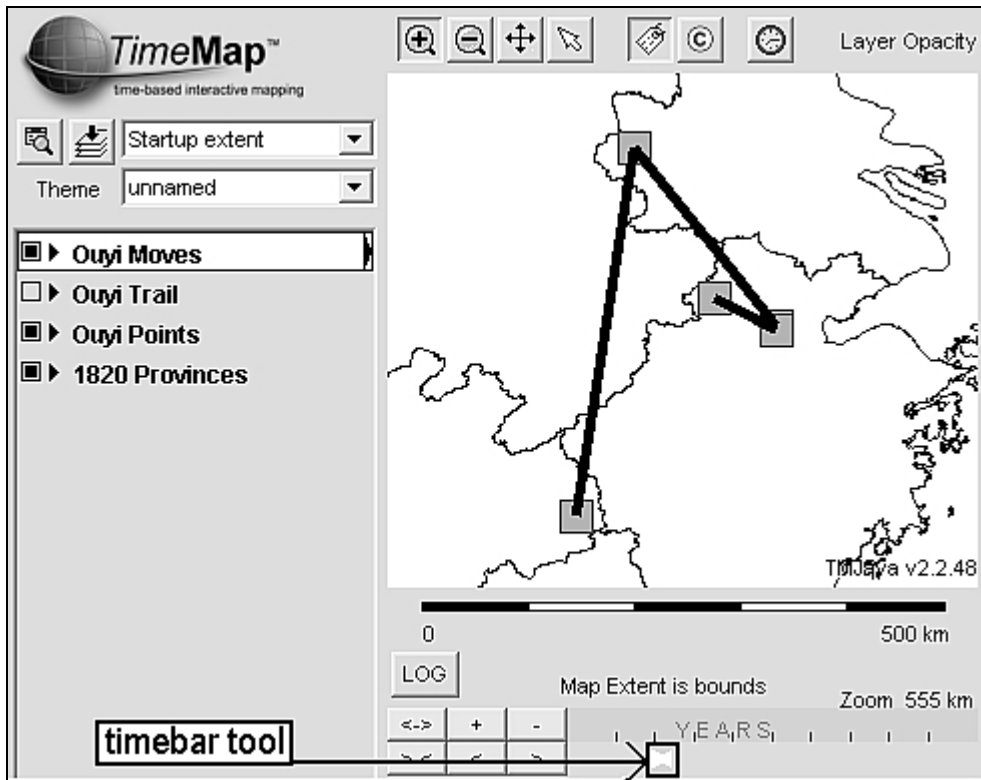


Figure 6: TimeMap Interface Showing Timebar Tool

To begin the animation process, we must first convert the Point Sequence Table to GIS Shapefile.⁸ This is accomplished in ArcView or ArcMap using the Add Event Theme or Add X-Y tools (which we won't describe in detail here). Note that TimeMap can also import a variety of database file formats (including .dbf, .txt, .csv) for plotting points. For our purposes, a standard Shapefile in Geographic Coordinate System, Decimal Degrees of Latitude and Longitude is the easiest format to work with. Keep in mind that a background map of some kind will be useful. For this animation we loaded the CHGIS 1820 Province Polygons as the background map.

Without going into too much detail about TimeMap software,⁹ we must follow these steps: create a MapSpace, Add Data Source to the MapSpace; use the Local File Wizard to Add Vector GIS; load the Point Sequence Shapefile (created from the Point Sequence Table); declare which fields are used for Unique ID, Label Name, Begin Time, and End Time; select Symbol Style and Colors; Map the results.

Once the Map View has been launched in TimeMap, open the Animation Controls (Figure 7) and enter the settings that suit your calculated Time values. The default setting will be for Years with Start = 0 and End = the current year (2007). The Time Span and Step Values will default to one tenth of the current year setting. For this demonstration, we had 134 events in the Points Sequence Table for the life of Zhixu. Therefore our calculated begin and end time values were Start = 1 and End = 1350 (the final end time value being derived as described in the previous section: $134 \times 10 + 10 = 1350$). The Time Span setting is used to control what objects can be viewed on screen at the same time. The Step setting is used to control how many years to jump forward when advancing to the next frame of the animation. Since our begin time and end time values occur at ten year increments we should set both of these to be less than ten. Let us set them both to 5 and run the animation using the standard Play button.

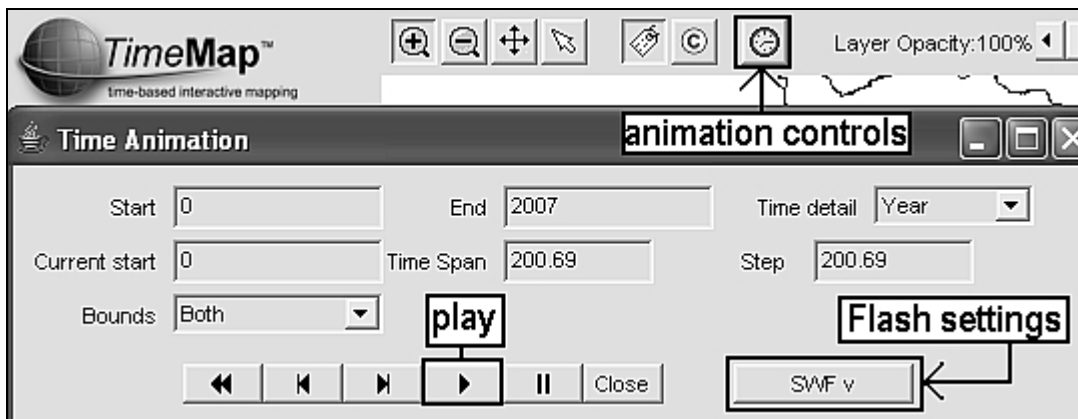


Figure 7: TimeMap Animation Controls

When the animation begins to run, it will play at the default increment of 1 second per Step. This means that a span of 1350 Year increments, divided by a Step value of 5 will play for 270 seconds ($1350/5=270$). By clicking on the Flash Settings button, you can adjust the animation time increment value, which is set at 1000 milliseconds. In this way you can speed up or slow down the final animation, if needed. Note that when the Flash Settings button is clicked, the dialog box will extend downward and you may enter such options as showing the time counter, showing a Title (which will appear on the upper left), and the selection of the filename for the Flash animation (which will be saved in SWF format).

The resulting animation will show a series of points appearing and disappearing according to their begin and end time values. In order to create lines showing the paths between the points we must create a separate Line Sequence Table and Shapefile, as described in the following section.

2.2 Mapping a sequence of lines showing paths of movement

As seen above, mapping a sequence of events as points depends on knowing the latitude and longitude coordinates of each point and assigning a begin and end time value to each point. In order to map a sequence of lines, each of which depicts a path between two events, we need two sets of coordinates for every path, the first being the FROM NODE and the latter being the TO NODE.

We can start by copying the contents of the Point Sequence Table to a Line Sequence Table. Next, beginning with the second row, we copy the x and y coordinate values for all the rows in the table and paste them alongside their immediately preceding records. The result is that the original pair of coordinates represent the FROM NODE and the subsequent pair of coordinates represent the TO NODE. (Figure 8) What we want is for sequence number 1 to have the FROM X and FROM Y values of sequence number 2 in its TO X and TO Y fields, and so on, down to the last record, which will have blank values for TO X and TO Y. Now we need to identify all rows which have exactly the same values in the FROM and TO fields and delete them. In Figure 8 these are marked (sequence numbers 2, 5, and 7).

year	seq	beg_time	end_time	event	point_id	FROM_X	FROM_Y	TO_X	TO_Y
1599	1	1	20	birth	ouyi_1	120.51	31.253	120.168	30.294
1622	2	20	30	Yogacara	ouyi_3	120.168	30.294	120.168	30.294
1622	3	30	40	received te	ouyi_3	120.168	30.294	119.796	30.378
1622	4	40	50	engaged in	ouyi_2	119.796	30.378	120.16	30.25
1622	5	50	60	other	ouyi_6	120.16	30.25	120.16	30.25
1623	6	60	70	divination c	ouyi_6	120.16	30.25	121.3333	29.5
1623	7	70	80	other	ouyi_4	121.3333	29.5	121.3333	29.5

Figure 8: Creating the Line Sequence Table

Why delete these rows? Because the line sequence in the animation is used to show paths of movement between two points. If the origin and destination are the same point, then there is no path of movement to delineate. In addition to identical points, you may wish to delete rows which connect points so close together that a path between them is almost invisible (although you may not be able to determine which points these are until viewing the first draft of the completed animation).

After we have limited the Line Sequence Table to only those rows needed for paths of movement, we should save it as two separate files: Moves Table and Trail Table. These will be converted to two unique Shapefiles and will serve two purposes in the final animation. For the Moves Table we should FIRST copy the end time values and PASTE them onto the begin time values. In other words, the begin time for the line showing the path of MOVEMENT should occur at the end time of the point which is going to disappear.

After we have replaced the begin time values with the end time values, we can now overwrite all the end time values with a new calculated end time value (end time = begin time + 5). The value of 5 can be adjusted, but must be smaller than the increment used to represent the sequence of points. In other words, if the value of 10 is used to show each point sequence increment, use a value less than 10 for the Moves line sequence increment. In the animation this will result in a line which shall appear at the exact time when the preceding event point must disappear, and which shall linger on screen for part of time that the subsequent end point will be seen. In this way, we can use the line sequence (using a THICK line symbol) to show the paths of movement between each event.

Once the Moves Table is done, we copy the begin time values to the corresponding field of the Trail Table. Then we overwrite all the end time values in the Trial Table with the highest end time value of the Points Sequence (in our case 1350). In this way, the line sequence in the

animation can be depicted using the Trail Table (with a thin line symbol) which lasts cumulatively for the entire duration of the animation. In other words, whereas the Moves line sequence appears and disappears whenever a change in locations occurs, the Trail line sequence builds up incrementally move by move and shows the entire path of the sequence of events. (Figure 9)

Moves Table

year	seq	begin_time	end_time	FROM_X	FROM_Y	TO_X	TO_Y
1599	1	20	25	120.51	31.253	120.168	30.294
1622	3	40	45	120.168	30.294	119.796	30.378
1622	4	50	55	119.796	30.378	120.16	30.25
1623	6	70	75	120.16	30.25	121.3333	29.5
1623	8	90	95	121.3333	29.5	119.796	30.378
1623	9	100	105	119.796	30.378	120.168	30.294
1624	12	130	135	120.168	30.294	120.29	31.57

Trail Table

year	seq	begin_time	end_time	FROM_X	FROM_Y	TO_X	TO_Y
1599	1	20	1350	120.51	31.253	120.168	30.294
1622	3	40	1350	120.168	30.294	119.796	30.378
1622	4	50	1350	119.796	30.378	120.16	30.25
1623	6	70	1350	120.16	30.25	121.3333	29.5
1623	8	90	1350	121.3333	29.5	119.796	30.378
1623	9	100	1350	119.796	30.378	120.168	30.294
1624	12	130	1350	120.168	30.294	120.29	31.57

Figure 9: End Times for line sequence differ in Moves Table and Trail Table

2.3 Generating the lines layers for use in animation

The creation of a Shapefile from a table containing pairs of latitude and longitude coordinates is a quite simple in ESRI and other GIS software applications. Creating lines from coordinates is generally a more advanced process. With ESRI software, the basic method is to use the GENERATE command in ArcInfo (which is the command line interface for ArcGIS).¹⁰

Without going into detail on usage of GENERATE command, it is important to mention that the GENERATE function will depend on an input file to create the lines. In our case, we need to eliminate all fields except the sequence ID and the FROM and TO coordinates. The resulting format of the input file should look like this:

```
1
120.51,31.253
120.168,30.294
end
3
120.168,30.294
119.796,30.378
end
4
119.796,30.378
120.16,30.25
end
```


After creating a line coverage for the line sequence, create two identical Shapefiles and save them as Moves and Trail. To the Moves shapefile, we need to rejoin the begin and end time fields which have a short duration. To the Trails shapefile need to rejoin the begin and end time values which all have the maximum end time value.

Once we have successfully joined the begin and end time information to both files, we can load them into the same TimeMap Map Space used previously to animate the Point Sequence. Be sure to define the Moves layer in TimeMap with a thick line symbol that will stand out, and define the Trail layer with a thin line symbol that will not distract too much from the active parts of the animation.

Using the same Animation Control values described above (section 2.1), playing the animation should result in a series of points that appear and disappear over time; thick lines will appear when the moves occur from point to another, and thin lines will remain on the screen showing the cumulative trail of all the points. (Figure 10)

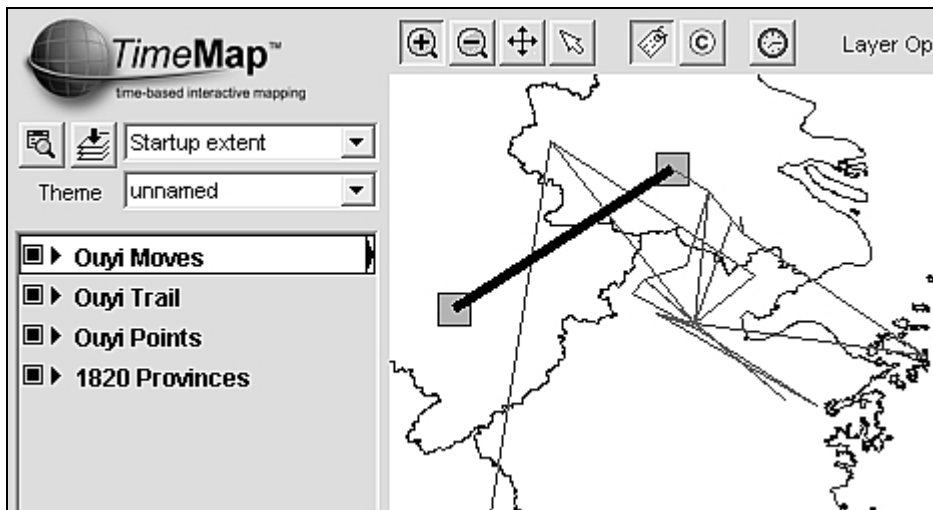


Figure 10: Animation with sequence of points, paths of movement, and cumulative trail

In this example, we have animated the major events from the biography of the Ming Dynasty monk Ouyi Zhixu. The resulting timeline is an unbroken linear depiction of a single path.

A copy of the animation is archived online here:
http://www.fas.harvard.edu/~chgis/work/docs/ouyi_flash_demo.html

Although the results are quite simple in this case, there is no reason why the same techniques cannot be applied to flows of trade goods, or capital, or population, or epidemiological outbreaks. Indeed, the thickness of lines or the size of the points can be set according to a values inherent to the data being visualized. Moreover, there is no reason why the path must be continuous. Breaks can be effectively shown, where data is lacking or where interruptions are known to have occurred. Similarly, there is no reason why multiple paths, or iterations of networks cannot be shown using the same techniques.

Here we have barely scraped the surface of how spatio-temporal data can be modeled. We hope that as the infrastructure of historical locations (such as provided by CHGIS) is completed, and the means of modeling them in spatio-temporal visualizations becomes more widely used, that researchers will discover new and interesting aspects of the underlying data which were not apparent until seen as animations over time.

Notes:

1. China Historical GIS. <http://www.fas.harvard.edu/~chgis>
2. Center for Historical Geography, Fudan University. <http://yugong.fudan.edu.cn/default.asp>
3. Berman, Lex. *Key Issues in Compiling a Digital Gazetteer of China's Historical Religious Sites*. 2004. http://www.fas.harvard.edu/~chgis/work/docs/papers/religious_sites_gazetteer.pdf
4. Bol, Peter Kees. *Liangzhe Tax Quota in 1077 CE* (digital map). 2004. <http://www.fas.harvard.edu/~chgis/maps/thematic.html#liang>
5. This study is entirely based on the PhD research of Beverly Foulks, East Asian Languages and Civilizations Department, Harvard University.
6. Master Shengyan. PhD thesis, Rissho University. Shengyan 聖嚴. *Minmatsu Chūgoku Bukkyō no kenkyū: toku ni Chigyoku o chūshin to shite*
未中國佛教の研究：特に智旭を中心として. Tōkyō : Sankibo busshorin, 1975.
7. TimeMap lead developers: Ian Johnson and Artem Osmakov. <http://www.timemap.net>
8. Shapefile is a standard GIS layer format for ESRI software products (ArcView, ArcMap, ArcInfo, etc). <http://www.esri.com>
9. TimeMap freeware Java Version 2.2.48 was used for this animation.
10. Proficient GIS users will be able to consult the documentation for GENERATE command, such as found at: <http://kb.iu.edu/data/andj.html>