ST-Cokriging ArcGIS extension for crime prediction

For Dr. Lin Liu's crime prediction group.
Please don't share outside the group without explicit permission of Dr. Lin Liu.

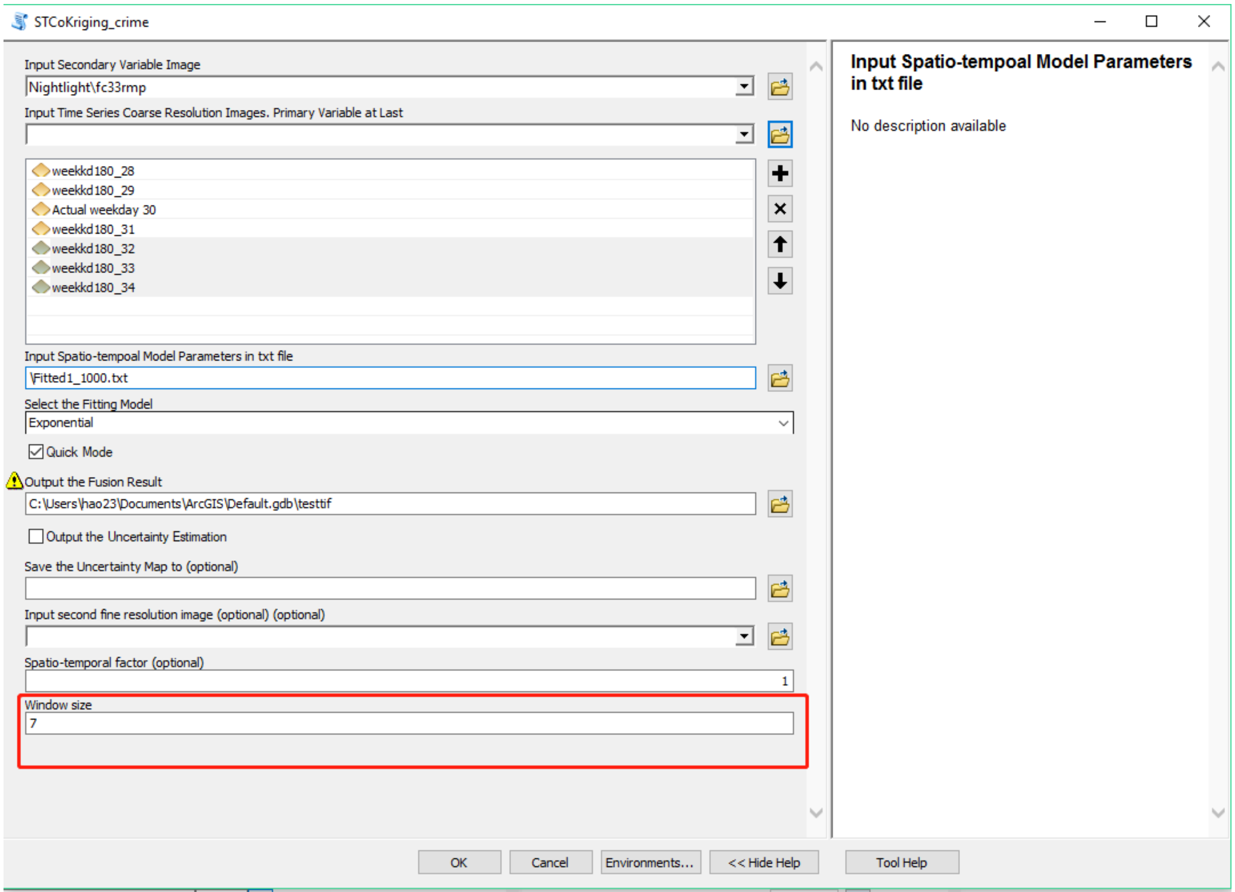
Bo Yang
Department of Geography & GIS
University of Cincinnati

Version 1.1 -Feb 09, 2018

Download:
ST-Cokriging codes v1.1 (here)
ArcToolbox for crime predictionv1.1(here)

Updates:

- Fixed some bugs and issue.
- Added module to calculate PAI and PEI (testing).
- Improved the temporal structure to more flexible.
- Added the window size parameter to fit smaller cell size (see figure below for new interface).



Note:

- The value of window size must be odd number.
- If using the cell size of 50 m by 50m, it is better to use window size from 5-9.
- The larger value of window size lead to longer computation time, so it is not recommended to use window size larger than 15.

Version 1.0 -Nov 12, 2017

I. Introduction

1. Geostatistics and Cokriging

In geostatistics, Cokriging is a multivariate variant of Kriging technique and makes spatial predictions for a sparsely sampled variable (the primary variable) of interest, with help of one or more well-sampled ancillary variables (the secondary co-variables). Cokriging method usually results in more accurate predictions of the target primary variable than Kriging. This is because Cokriging method exploits cross-correlations between the primary variable and the secondary co-variables in addition to the spatial autocorrelation of the primary variable.

In conventional Cokriging method, both the primary variable and co-variables are in spatial domain, and time dimension is not taken into consideration. By extending it from sole spatial domain to the spatio-temporal domain, this algorithm formulated a ST-Cokriging method that is capable of taking advantage of both spatial and temporal correlation within and between primary and co-variables to produce temporally frequent predictions for the primary variable at a high spatial resolution as the co-variables.

2. ST-Cokriging work flow

In ST-cokriging formulation, we assume that the primary variable of interest is coarse spatial resolution images that are sampled at a high temporal frequency (high temporal resolution), and the secondary variable (co-variable) are fine spatial resolution images that are sparsely sampled over time (low temporal resolution), as shown in Figure 1. Without loss of generality, we only consider the case with only one co-variable observed at multiple time points in the mathematical formulation of ST-Cokriging method. The extension to two or more co-variables observed at multiple time points is straightforward.

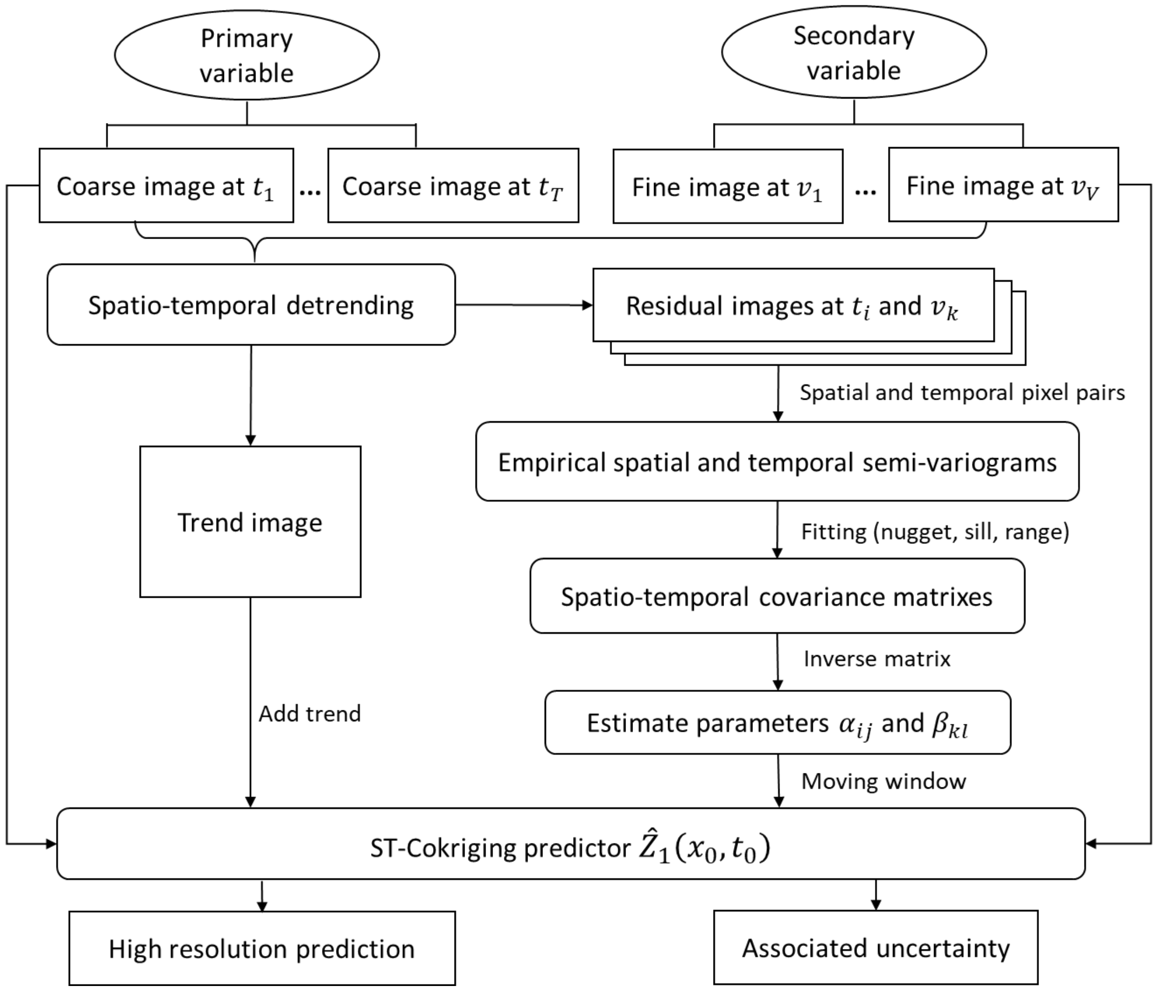


Figure 1. Data processing flowchart of ST-Cokriging method

3. Extension implementation

In the Cokriging linear system, the covariance matrix C is of size $(\sum_{i=1}^{T} N_i + \sum_{j=1}^{V} M_j + 2) \times (\sum_{i=1}^{T} N_i + \sum_{j=1}^{V} M_j + 2)$. Since the co-variable is observed at high spatial resolution, it is very likely that M_j's are very large. Similarly, even with smaller N_i's for the primary variable observed at coarse spatial resolution, $\sum_{i=1}^{T} N_i$ can still be large, since the primary variable is observed at high temporal frequency. Therefore, the matrix C can be of high dimension. Solving such a high-dimensional linear system can be computationally infeasible. One popular method to alleviate this difficulty is to force small numbers in the matrix (vector) to be zero, known as thresholding or tapering. Meanwhile, we take advantage of the feature of regularly gridded data in the application presented in Section 3, which facilitates efficient parallel computing of the Cokriging predictor and variance.

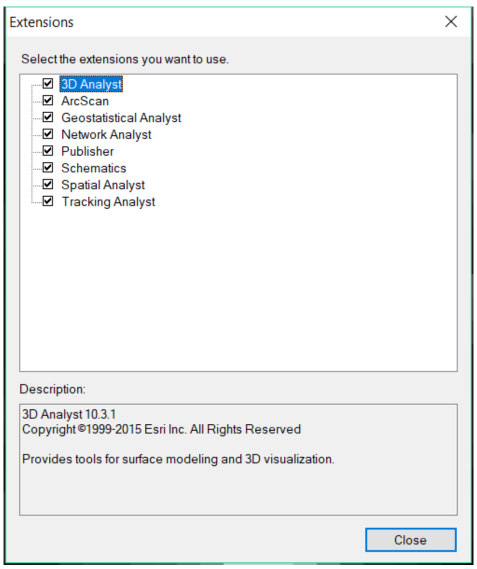
II. Source and Environmental

1. Source code download

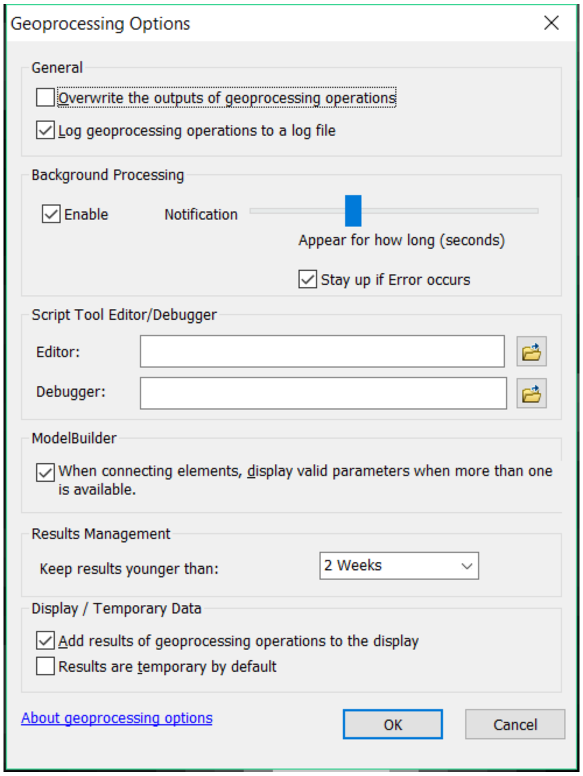
- Download the toolbox in ArcGIS from here.
- Download the script for spatio-temporal semi-variogram here.
- Download the script for ST-Cokriging here.

2. Environmental setup

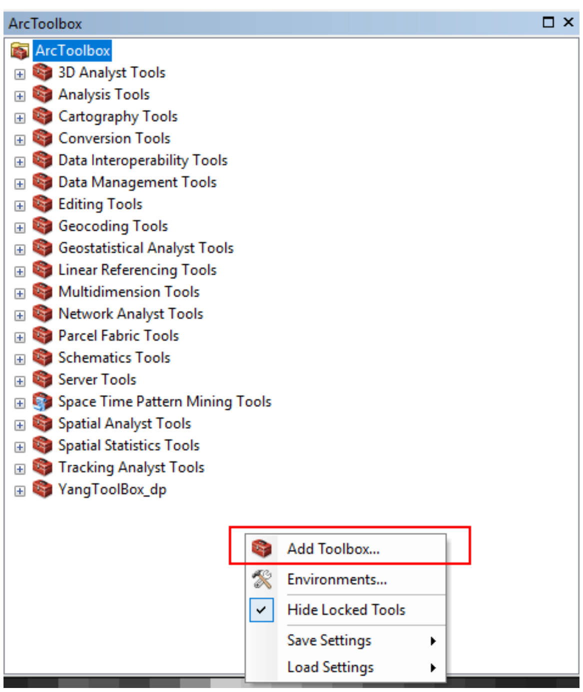
- Enable extensions in ArcMap



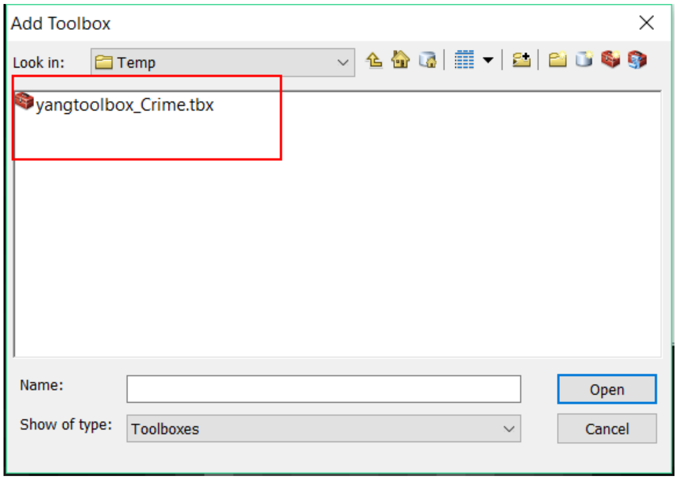
- Setup the Geo-processing option



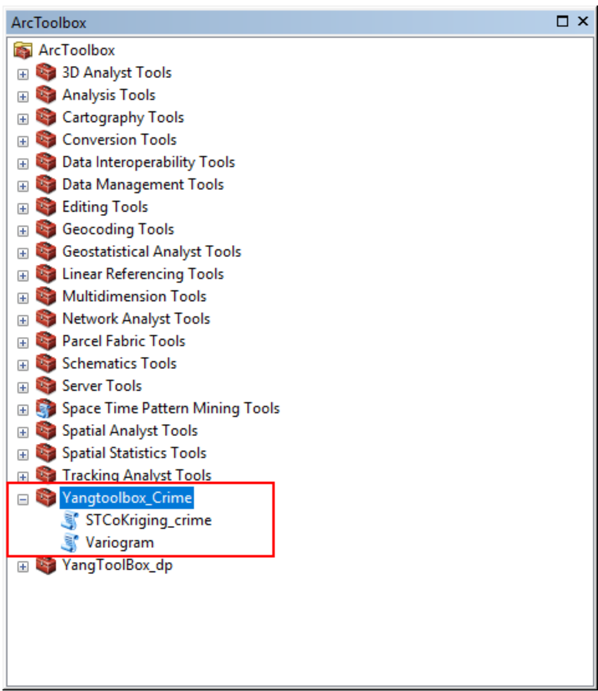
- Open Arctoolbox window, right click and add a toolbox



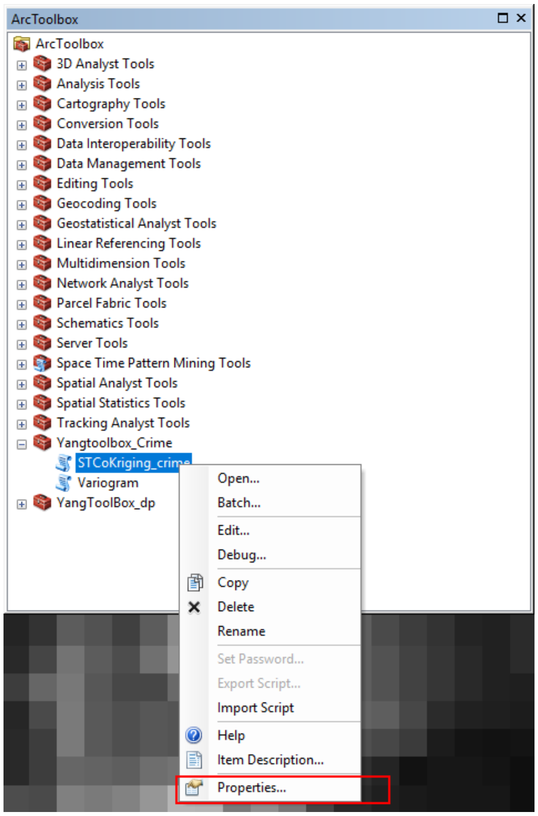
- Navigate to the toolbox just downloaded and select.



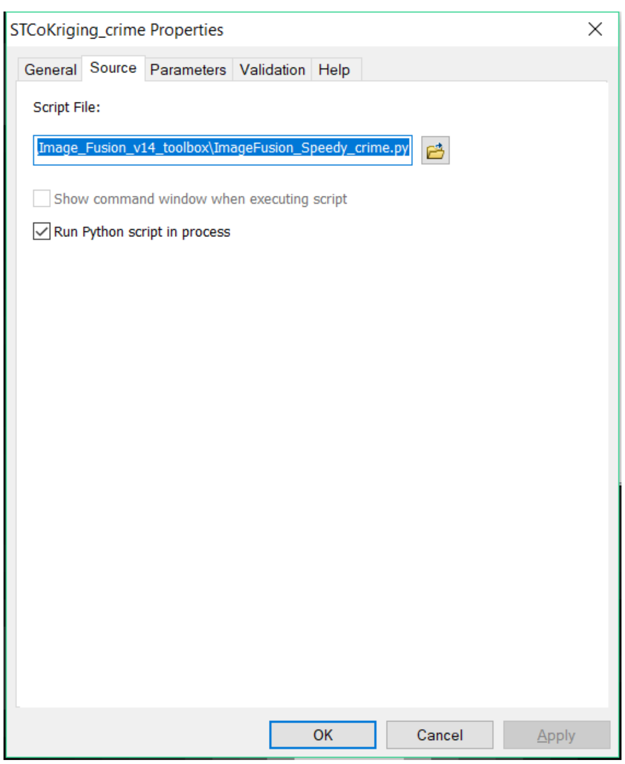
- Unfold the toolbox and the scripts should be appeared in the toolbox



- Right click the ST-Cokriging script and select properties



- Click source tab and link the ST-Cokriging script to the toolbox

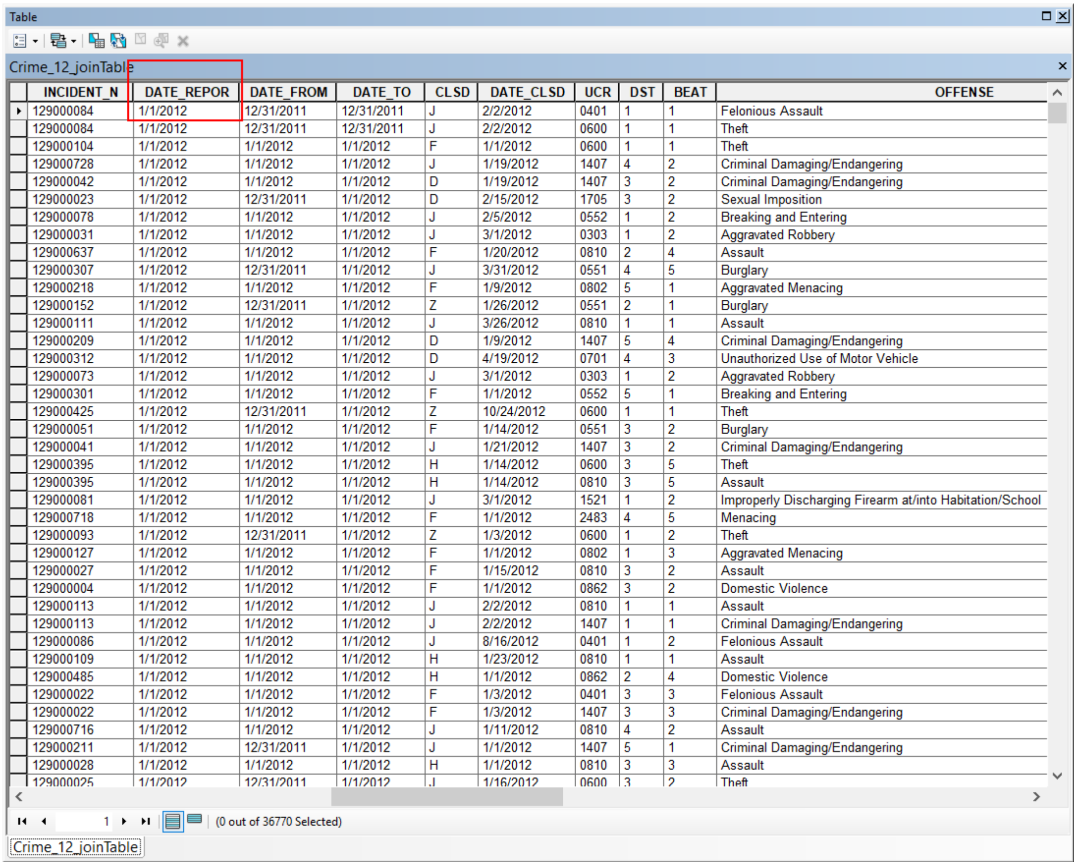


- Do the same procedure to the semi-variogram script and link the downloaded code to the arctoolbox script.

III. Data and Preprocessing

1. Prepare the primary data

- Select one year of crime data, this tutorial uses 2012 crime data in Cincinnati as an example.



INCIDENT_N	DATE_REPOR	DATE_FROM	DATE_TO	CLSD	DATE_CLSD	UCR	DST	BEAT	OFFENSE
129000084	1/1/2012	12/31/2011	12/31/2011	J	2/2/2012	0401	1	1	Felonious Assault
129000084	1/1/2012	12/31/2011	12/31/2011	J	2/2/2012	0600	1	1	Theft
129000104	1/1/2012	1/1/2012	1/1/2012	F	1/1/2012	0600	1	1	Theft
129000728	1/1/2012	1/1/2012	1/1/2012	J	1/19/2012	1407	4	2	Criminal Damaging/Endangering
129000042	1/1/2012	1/1/2012	1/1/2012	D	1/19/2012	1407	3	2	Criminal Damaging/Endangering
129000023	1/1/2012	12/31/2011	1/1/2012	D	2/15/2012	1705	3	2	Sexual Imposition
129000078	1/1/2012	1/1/2012	1/1/2012	J	2/5/2012	0552	1	2	Breaking and Entering
129000031	1/1/2012	1/1/2012	1/1/2012	J	3/1/2012	0303	1	2	Aggravated Robbery
129000637	1/1/2012	1/1/2012	1/1/2012	F	1/20/2012	0810	2	4	Assault
129000307	1/1/2012	12/31/2011	1/1/2012	J	3/31/2012	0551	4	5	Burglary
129000218	1/1/2012	1/1/2012	1/1/2012	F	1/9/2012	0802	5	1	Aggravated Menacing
129000152	1/1/2012	12/31/2011	1/1/2012	Z	1/26/2012	0551	2	1	Burglary
129000111	1/1/2012	1/1/2012	1/1/2012	J	3/26/2012	0810	1	1	Assault
129000209	1/1/2012	1/1/2012	1/1/2012	D	1/9/2012	1407	5	4	Criminal Damaging/Endangering
129000312	1/1/2012	1/1/2012	1/1/2012	D	4/19/2012	0701	4	3	Unauthorized Use of Motor Vehicle
129000073	1/1/2012	1/1/2012	1/1/2012	J	3/1/2012	0303	1	2	Aggravated Robbery
129000301	1/1/2012	1/1/2012	1/1/2012	F	1/1/2012	0552	5	1	Breaking and Entering
129000425	1/1/2012	12/31/2011	1/1/2012	Z	10/24/2012	0600	1	1	Theft
129000051	1/1/2012	1/1/2012	1/1/2012	F	1/14/2012	0551	3	2	Burglary
129000041	1/1/2012	1/1/2012	1/1/2012	J	1/21/2012	1407	3	2	Criminal Damaging/Endangering
129000395	1/1/2012	1/1/2012	1/1/2012	H	1/14/2012	0600	3	5	Theft
129000395	1/1/2012	1/1/2012	1/1/2012	H	1/14/2012	0810	3	5	Assault
129000081	1/1/2012	1/1/2012	1/1/2012	J	3/1/2012	1521	1	2	Improperly Discharging Firearm at/into Habitation/School
129000718	1/1/2012	1/1/2012	1/1/2012	F	1/1/2012	2483	4	5	Menacing
129000093	1/1/2012	12/31/2011	1/1/2012	Z	1/3/2012	0600	1	2	Theft
129000127	1/1/2012	1/1/2012	1/1/2012	F	1/1/2012	0802	1	3	Aggravated Menacing
129000027	1/1/2012	1/1/2012	1/1/2012	F	1/15/2012	0810	3	2	Assault
129000004	1/1/2012	1/1/2012	1/1/2012	F	1/1/2012	0862	3	2	Domestic Violence
129000113	1/1/2012	1/1/2012	1/1/2012	J	2/2/2012	0810	1	1	Assault
129000113	1/1/2012	1/1/2012	1/1/2012	J	2/2/2012	1407	1	1	Criminal Damaging/Endangering
129000086	1/1/2012	1/1/2012	1/1/2012	J	8/16/2012	0401	1	2	Felonious Assault
129000109	1/1/2012	1/1/2012	1/1/2012	H	1/23/2012	0810	1	1	Assault
129000485	1/1/2012	1/1/2012	1/1/2012	H	1/1/2012	0862	2	4	Domestic Violence
129000022	1/1/2012	1/1/2012	1/1/2012	F	1/3/2012	0401	3	3	Felonious Assault
129000022	1/1/2012	1/1/2012	1/1/2012	F	1/3/2012	1407	3	3	Criminal Damaging/Endangering
129000716	1/1/2012	1/1/2012	1/1/2012	J	1/11/2012	0810	4	2	Assault
129000211	1/1/2012	12/31/2011	1/1/2012	J	1/1/2012	1407	5	1	Criminal Damaging/Endangering
129000028	1/1/2012	1/1/2012	1/1/2012	H	1/1/2012	0810	3	3	Assault
129000025	1/1/2012	12/31/2011	1/1/2012	J	1/16/2012	0600	3	2	Theft

- Select all crime records in the categories of all kinds of robberies and assaults. In other case study, please select the target crime category accordingly.

Table

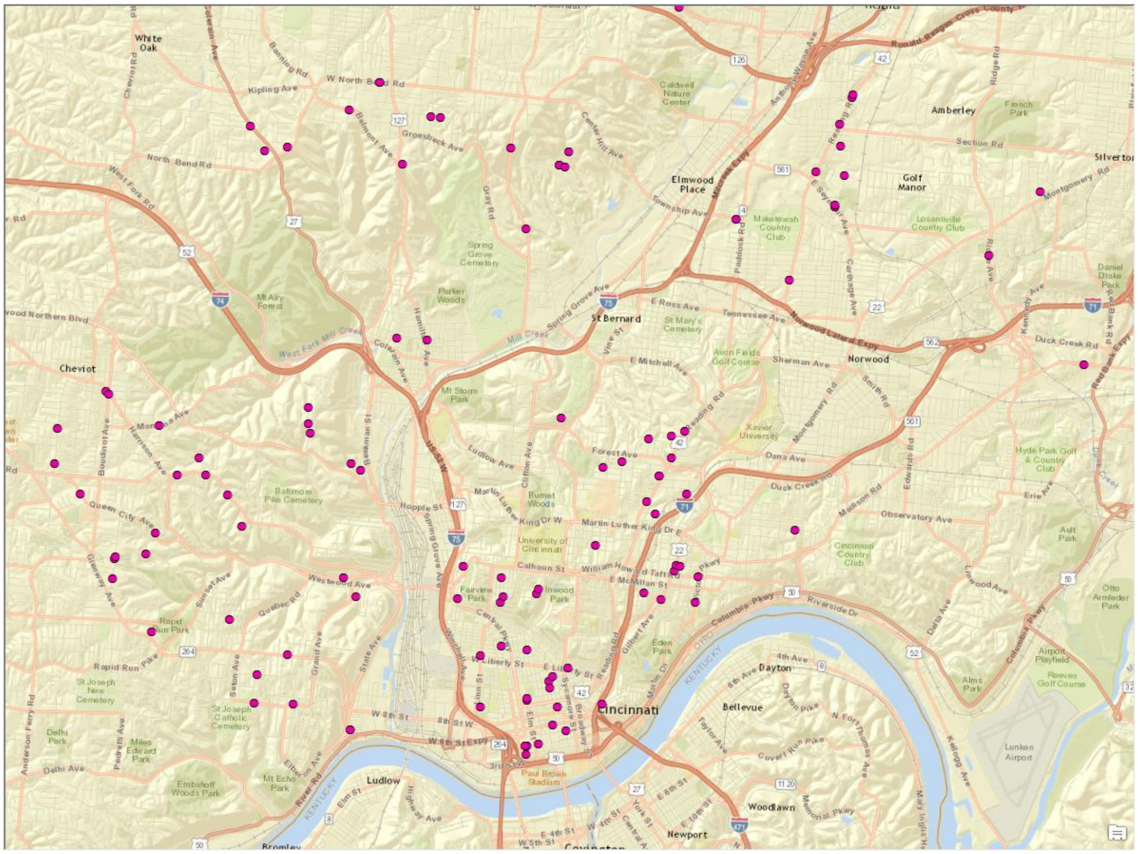
Crime_12_ass_rob

DATE_TO	CLSD	DATE_CLSD	UCR	DST	BEAT	OFFENSE	LOCATION	THEFT_CODE	WEAPON
1/8/2012	D	1/19/2012	0810	4	3	Assault	47		40
1/8/2012	I	<Null>	0810	4	3	Assault	47		40
1/8/2012	D	1/31/2012	0810	5	3	Assault	02		40
1/8/2012	F	1/8/2012	0810	3	3	Assault	02		40
1/8/2012	F	1/15/2012	0810	3	3	Assault	02		40
1/8/2012	J	2/2/2012	0301	1	1	Robbery	47		40
1/8/2012	F	1/15/2012	0810	3	1	Assault	47		40
1/8/2012	J	2/12/2012	0303	2	3	Agg Robbery - Armed, Deadly Weapon	47		11
1/8/2012	J	3/6/2012	0303	4	2	Aggravated Robbery	29		12
1/8/2012	F	1/8/2012	0810	3	3	Assault	48		40
1/8/2012	J	2/17/2012	0303	4	5	Aggravated Robbery	47		11
1/8/2012	J	2/1/2012	0303	5	1	Aggravated Robbery	47		12
1/8/2012	J	1/12/2012	0810	4	4	Assault	02		40
1/9/2012	H	1/25/2012	0810	4	5	Assault	25		40
1/9/2012	D	1/30/2012	0810	1	1	Assault	02		40
1/6/2012	G	1/9/2012	0810	3	2	Assault	47		99
1/9/2012	J	2/18/2012	0810	3	3	Assault	02		40
1/9/2012	I	<Null>	0303	4	2	Aggravated Robbery	16		12
1/9/2012	G	1/9/2012	0810	2	3	Assault	43		40
12/31/2011	D	1/21/2012	0810	5	3	Assault	02		99
1/9/2012	J	1/30/2012	0810	1	2	Assault	01		40
1/9/2012	D	2/21/2012	0810	1	2	Assault	01		40
1/2/2012	J	2/21/2012	0810	1	1	Assault	26		99
1/9/2012	J	2/3/2012	0303	3	3	Aggravated Robbery	47		11
12/27/2011	D	1/19/2012	0810	4	2	Assault	22		40
1/9/2012	J	2/18/2012	0810	3	3	Assault	01		99
1/7/2012	J	2/6/2012	0303	5	1	Aggravated Robbery	47		40
12/30/2011	J	1/14/2012	0810	4	5	Assault	26		40
1/9/2012	H	1/9/2012	0810	4	5	Assault	02		40
1/9/2012	G	1/11/2012	0303	3	3	Aggravated Robbery	47		11
1/9/2012	J	2/8/2012	0810	4	3	Assault	08		40
1/9/2012	G	2/21/2012	0303	5	4	Aggravated Robbery	47		11
1/9/2012	Z	1/9/2012	0810	1	1	Assault	52		40
1/9/2012	J	3/19/2012	0303	1	2	Aggravated Robbery	47		12
1/9/2012	J	2/17/2012	0810	4	3	Assault	08		30
1/9/2012	J	1/17/2012	0810	4	2	Assault	47		40
1/9/2012	J	2/7/2012	0303	1	1	Aggravated Robbery	48		20
1/9/2012	J	2/29/2012	0301	5	4	Robbery	47		40
1/9/2012	J	2/18/2012	0810	3	5	Assault	02		40

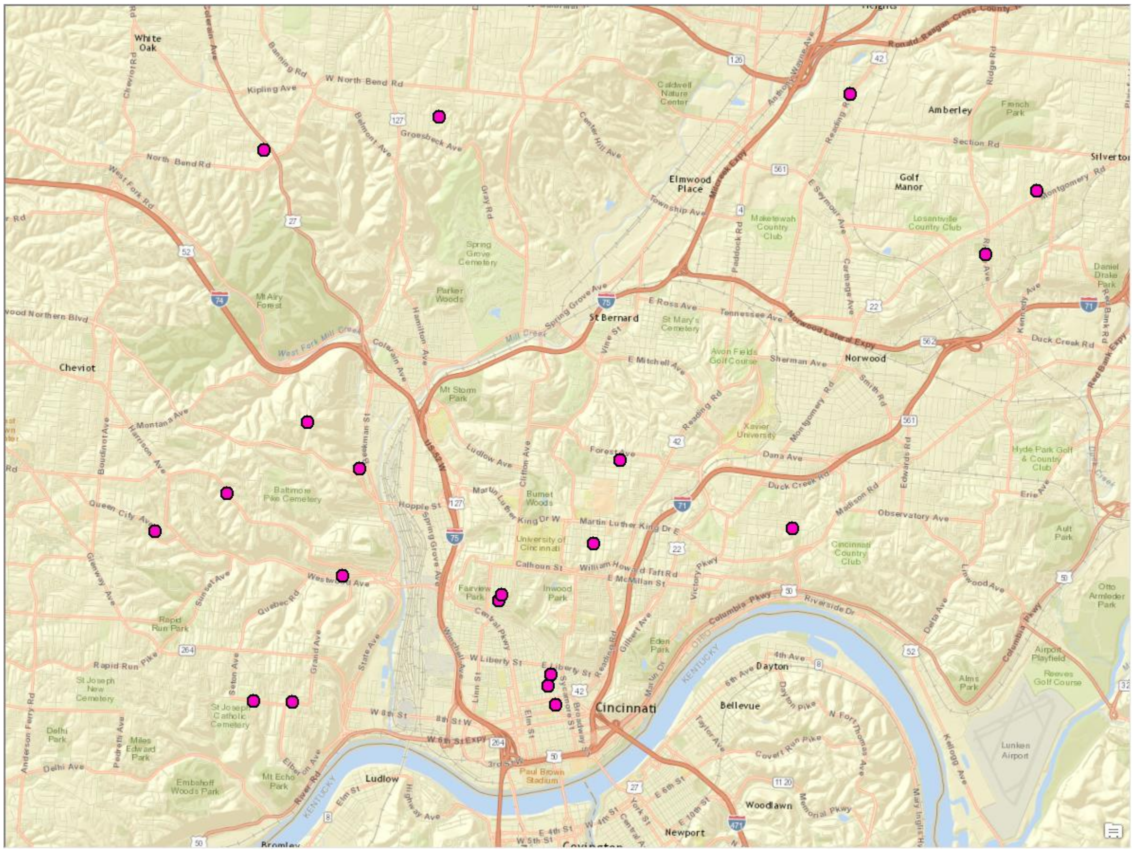
1 (0 out of 5491 Selected)

Crime_12_ass_rob

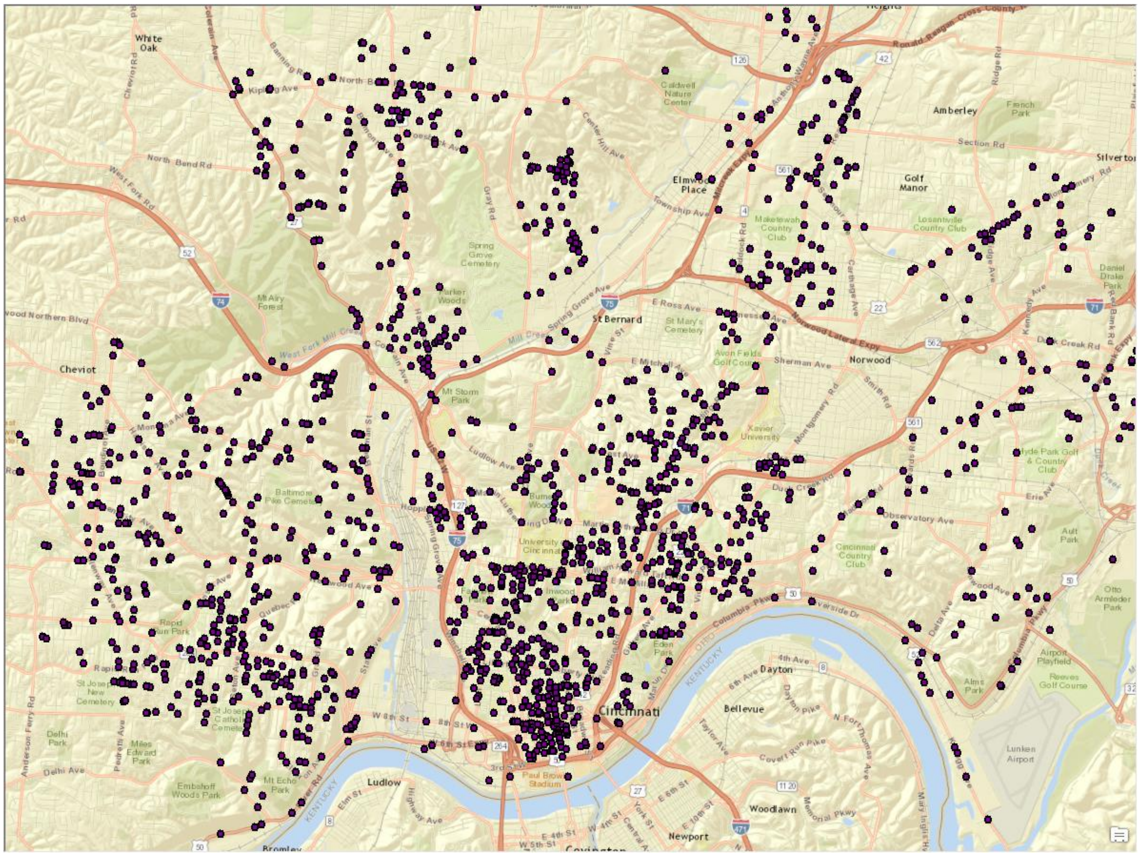
- Create a new feature class of selected crime. The further analysis will be performed on it.
- Based on the crime-reported date, divide the records into weekly based date. There are approximately 53 weeks in a year. Divide the data to 53 separate feature classes which corresponding to each week. Each week should start on Monday 0:00 am and end on Sunday 11:59 pm.
- Separate the weekly data again based on weekday and weekends. That is, further divide each product of last step to two feature classes: week day (five days) and weekend (Saturday and Sunday) based on the crime-reported data.
- The weekday data example (Week 2, Cincinnati 2012):



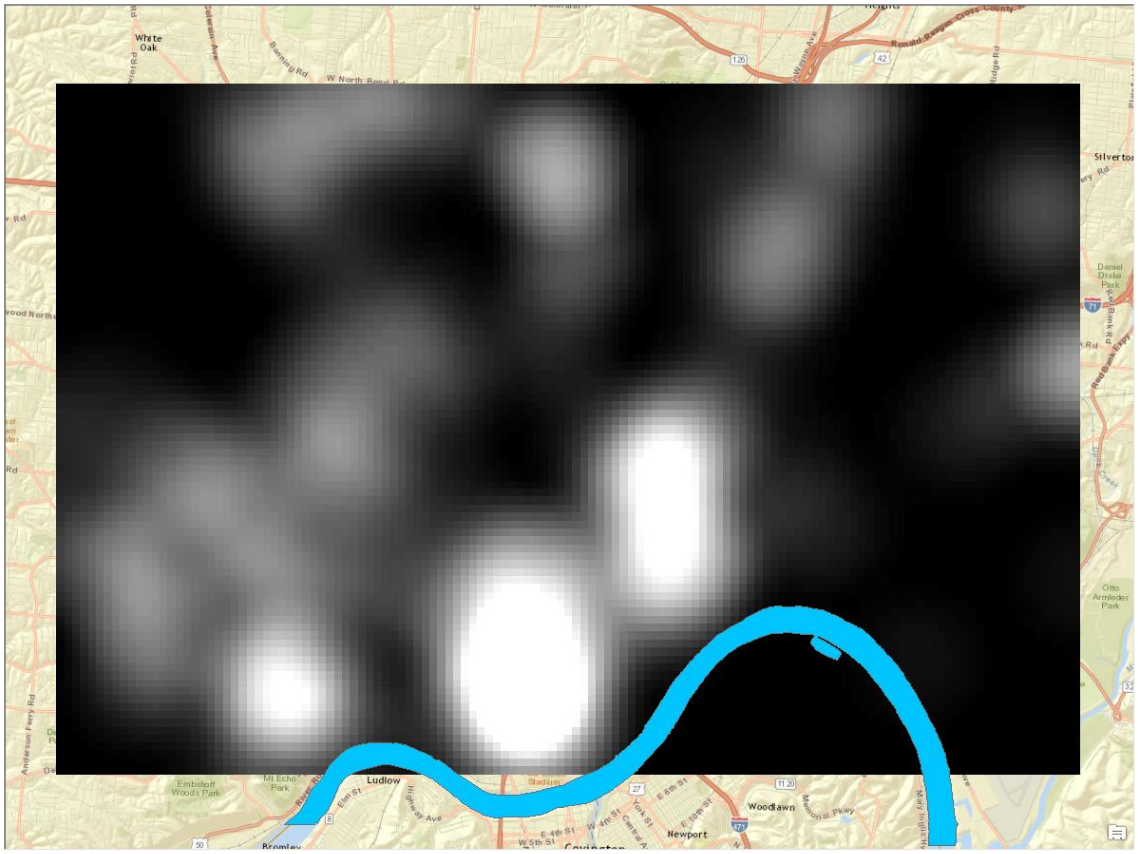
- The weekend data example (Week 2, Cincinnati 2012):



- Aggregate weekly data to quad-week data (union every 4 weeks) for weekday and weekend, respectively.
- The aggregated data example (weekday)



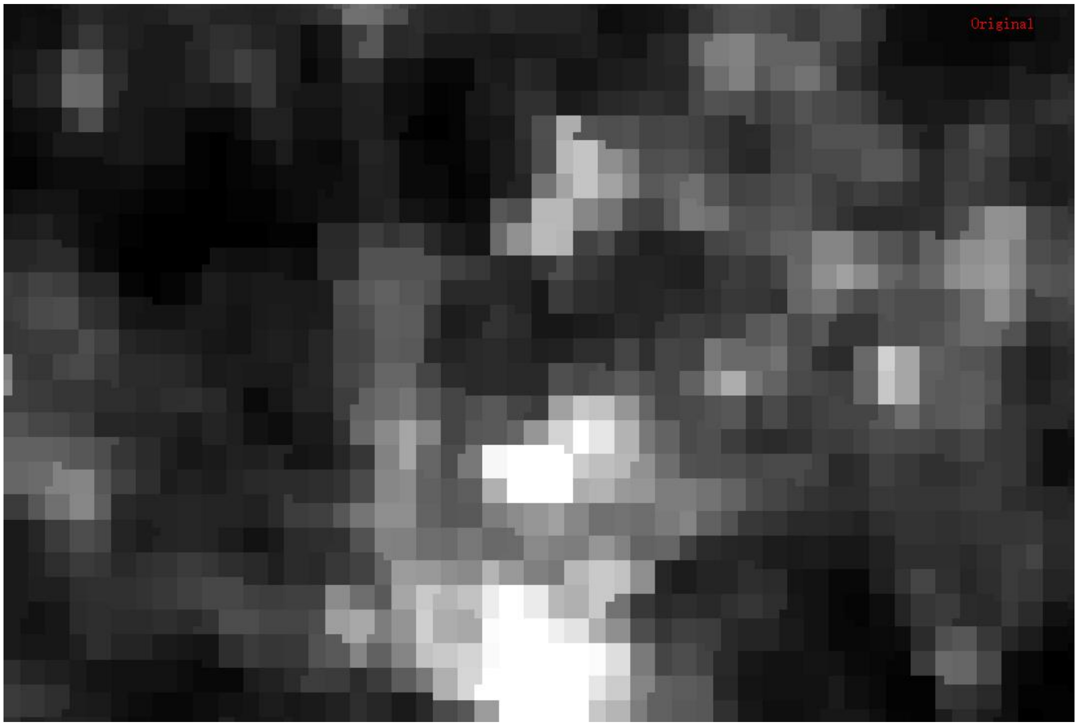
- Generate kernel density maps for processed quad-week data, the kernel density process is performed on each weekday and weekend data. The Output cell size is 450. Output values are EXPECTED_COUNTS.
- The final kernel density map example:

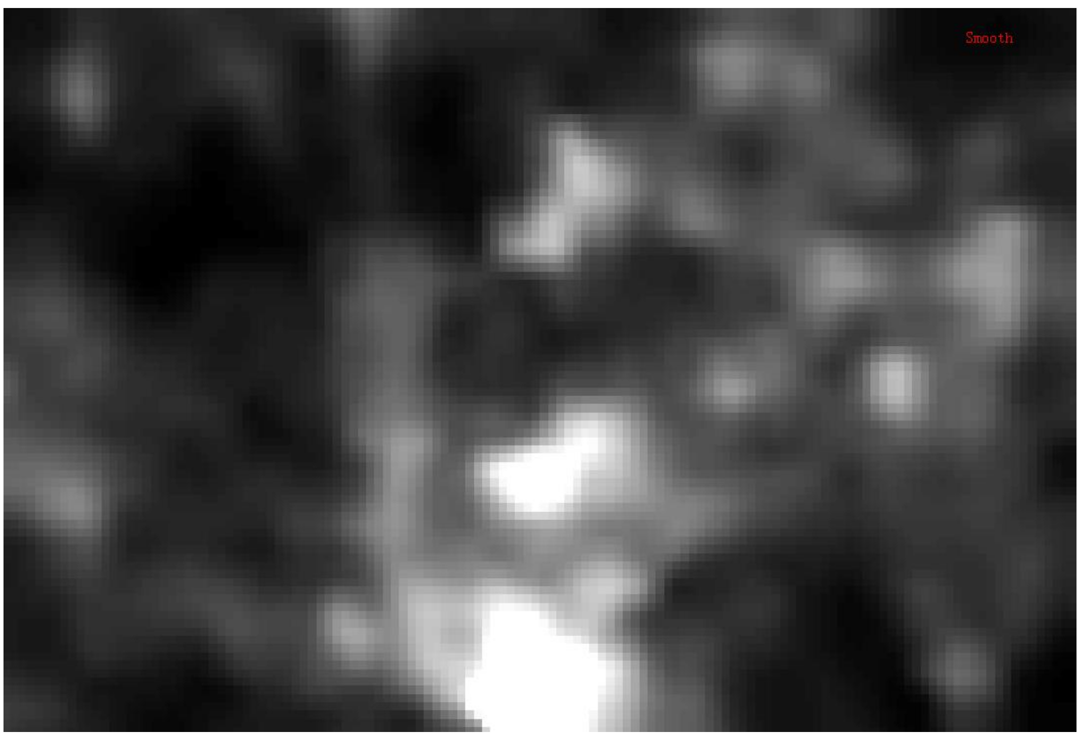


- Inspect all kernel density maps, decide the general area, and clip them to the study area, all kernel density maps are in same coverage and at same resolution.

2. Prepare the covariable

- The covariable could be any data that is correlated with the predicting variable. This tutorial uses nightlight as an example.
- The covariable should be clip to same coverage of the primary variable at same spatial resolution, Covariable should be normalized to 0-1 probability map using equation: New_Value = DN/DNmax
- It’s highly recommended to smooth the covariable if using the night light. Otherwise the final prediction will be influenced by the edge of coarse pixels
- An example of night light (Cincinnati, Nov 2012) product before and after smooth filtering:

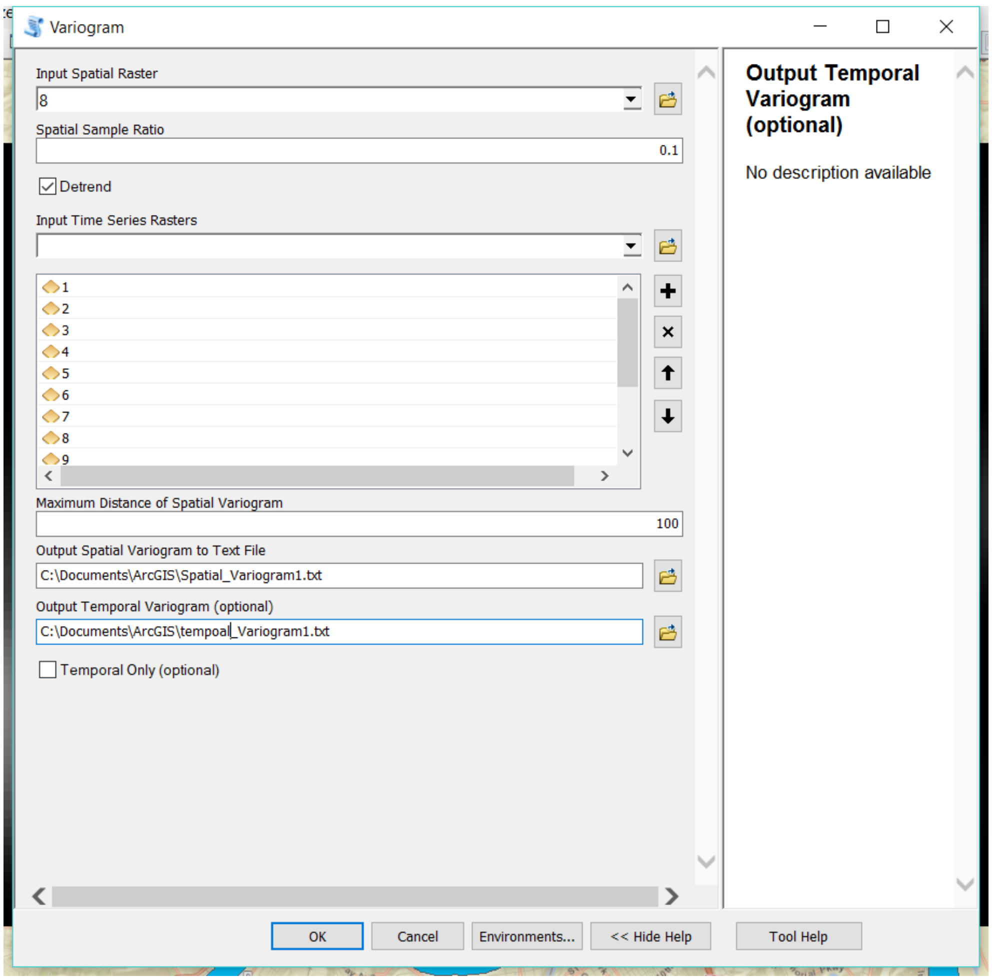




IV. Prediction via Extension

1. Estimate the spatio-temporal semi-variograms

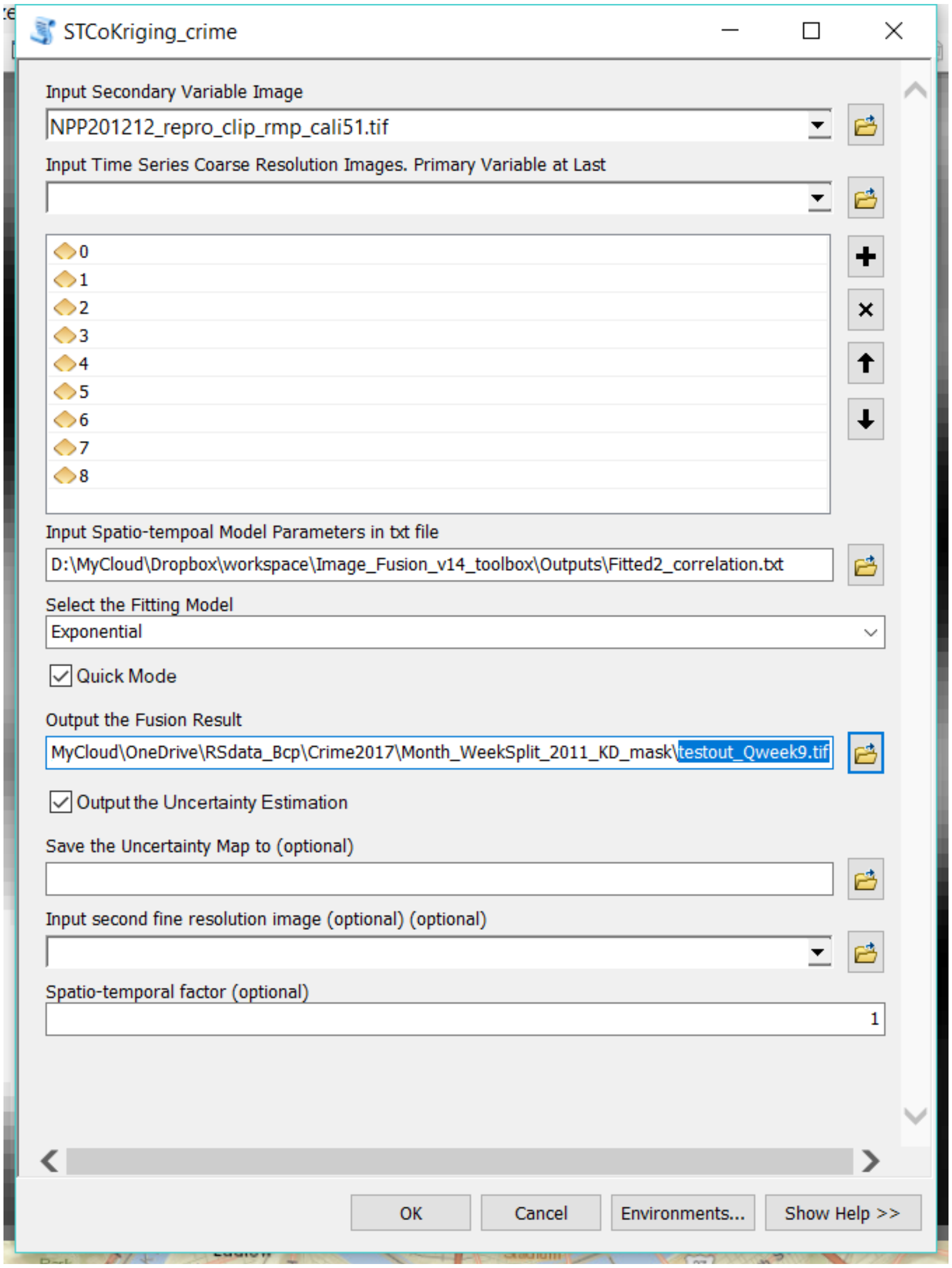
- Input the parameters as shown in the figure below. All 13 quad-week images should be included in the calculation and arranged in chronological order.
- The input spatial raster should be one of the quad-week period with largest number of crime.
- The spatial sample ratio is the percentage of the subset of spatial samples, depends on the resolution and total pixels of the study ration, the subset population should be around 3,000 – 10, 000.
- At last, select the path to store the txt files of spatial and temporal semi-variogram.



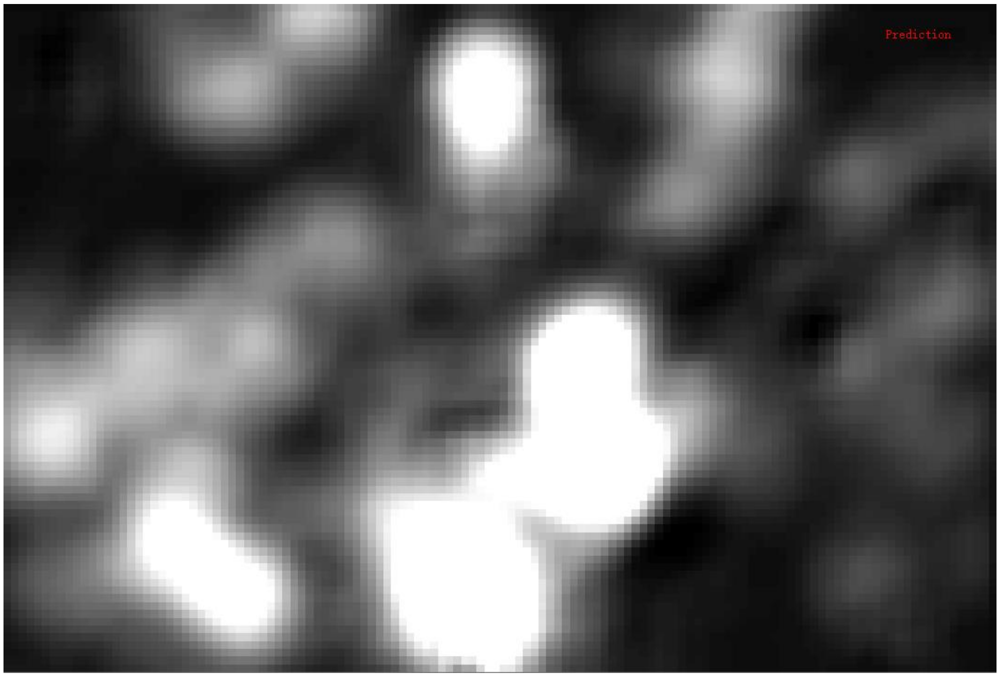
2. Prediction using ST-Cokriging

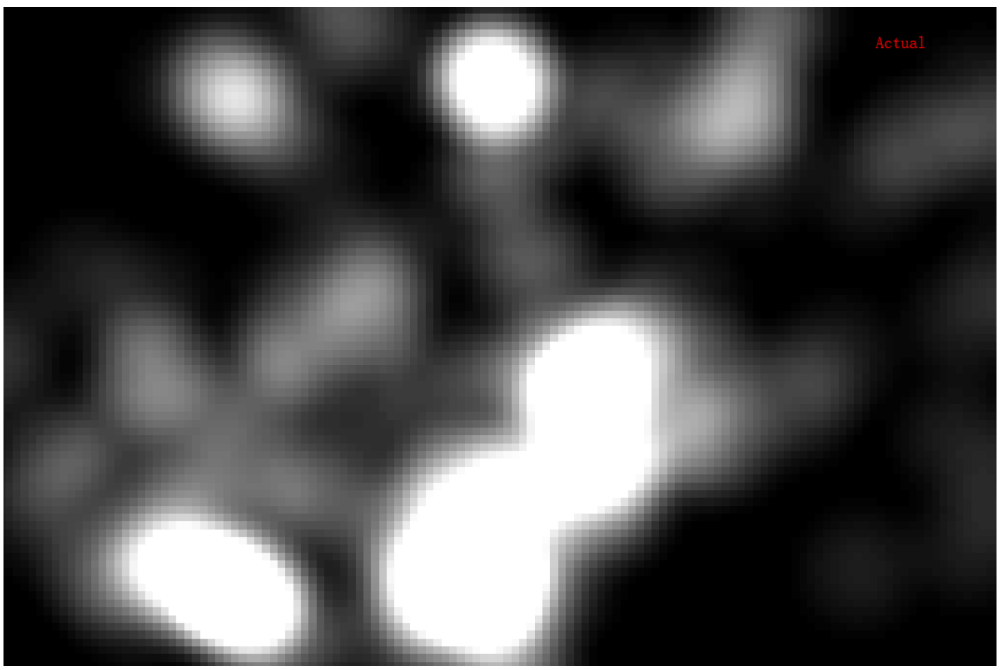
- In put the secondary covariable image, and time-series primary variable, the primary variable should be input in the time-series order. The number of time-series should be no less than 3 for using of spatio-temporal prediction. Other parameters should be included as the example shown below.

- The spatio-temporal model parameters txt file is generated using the first step. Please send the spatial and temporal txt file generated in last step to me for estimating the spatio-temporal model. I also generated a general spatio-temporal model for testing, which can be downloaded here.



- Here is an example of prediction result (Cincinnati September,2012)





V. Troubleshooting

- Both time-series primary variable and co-variable should be re-project to same coordinates system as well as same datum.
- Co-variable should be at finer spatial resolution or same. In this version, the coverage of

the both co-variable and time-series primary variable should be same, and at same spatial resolution.

- If using 32-bit OS software computer, the maximum number of row and column are 6000 by 6000.
- The fusion process may take a while (about 1 hour depends on the CPU and RAM of the computer). Once it started running, please don't use other program at same time. Otherwise window 10 might report ArcGIS no response. When no response appeared, please choose to wait rather than kill the process.
- The final product will have 1-pixel-width edge at constant value (or very smooth). It is normal because current version did not model the edge effects, and leave the edge for showing the trend value.

VI. Reference and contacts

- Related publications to be added.
- Please contact Bo Yang (yangb2@mail.uc.edu) if have any questions.