The Risk Terrain Modeling Diagnostics (RTMDx) Utility is a software application for Risk Terrain Modeling (RTM) and for diagnosing spatial crime vulnerabilities. For more information about RTM, visit www.rutgerscps.org.

The RTMDx Utility helps to identify and communicate environmental attractors of crime incidents at the micro level. Information products can be used to anticipate places that will be most suitable for illegal behavior, identify where new crime incidents will emerge and/or cluster, develop place-based interventions, strategically and tactically allocate resources, and prioritize efforts to mitigate crime risks.

By Joel M. Caplan
Leslie W. Kennedy
Eric L. Piza
SOFTWARE: Copyright © 2013 by Rutgers, The State University of New Jersey

USER MANUAL: Copyright © 2013 by Rutgers Center on Public Security

All rights reserved. No part of this document may be reproduced in any manner without written permission by authorized representatives of the copyright holder.

USER MANUAL Suggested Citation:

SOFTWARE Suggested Citation:

Produced by the Rutgers Center on Public Security:
Based at the Rutgers University School of Criminal Justice, the Rutgers Center on Public Security (RCPS) offers a multidisciplinary approach to the academic study and practical application of ways in which democratic societies can effectively address crime, terrorism and other threats to public security. This involves the prevention of, protection from, and response to natural or human-made events that could endanger the safety or security of people or property in a given area. RCPS engages in innovative data analysis and information dissemination, including the use of GIS, for strategic decision-making and tactical action. Visit RCPS online for current research projects, reports and publications: www.rutgerscps.org
Acknowledgements

The opinions expressed in this document are the views of the authors and do not necessarily reflect the views and opinions of the Rutgers Center on Public Security, Rutgers University, or the School of Criminal Justice.

The Risk Terrain Modeling Diagnostics (RTMDx) Utility graphical user interfaces, icons/buttons, splash screens, dialog boxes, artwork, emblems, and associated materials are the intellectual property of Rutgers and are reproduced herein by permission. Copyright © 2013 Rutgers, The State University of New Jersey. All rights reserved.

Companies, products or websites mentioned herein are trademarks, registered trademarks, or service marks of their respective owners and are the property of their respective owners.
# Table of Contents

Key Concepts .......................................................................................................................... 6
Preface .................................................................................................................................... 7
System Requirements ............................................................................................................. 9
Input Parameter Guidelines .................................................................................................... 9
Installation ............................................................................................................................... 11
Starting the Utility .................................................................................................................. 13
GUI (Graphical User Interface) Overview ................................................................................ 13
Running a Risk Terrain Model ............................................................................................... 15
RTMDx Utility Demonstration ............................................................................................... 26
Statistics of the RTMDx Utility ............................................................................................. 35
Epilogue .................................................................................................................................. 40
Appendix ................................................................................................................................. 40
Key Concepts

The key concepts of risk terrain modeling (RTM) provide a standard framework for discussions about crime problems and crime prevention. As crimes cluster spatially and temporally, RTM visualizes and statistically validates the reasons for these patterns so that police and other stakeholders can prioritize and proactively address elements that drive crime problems.

**SPATIAL INFLUENCE** refers to the way in which features of an environment affect behaviors at or around the feature itself. Operationalizing the spatial influences of crime risk factors to particular landscapes is what makes RTM especially meaningful and practical. For instance: Imagine you are an unfamiliar visitor to an American city. As you stand on the sidewalk, you call a local friend on the cell phone and describe your location as the "bar district" because you observed a high concentration of bars within the area. From a criminological perspective, bars and other liquor establishments are known to correlate with violent crimes. You may be at even greater risk of victimization if you are within one block from a bar as opposed to farther away. So, it could be argued with empirical support that your risk of being the victim of violent crime is not only a function of the physical structure of the bars themselves, but also the distribution of those features throughout the landscape, your proximity to them, and the spatial influences they have on the attraction of potential offenders, suitable victims, and crime. Risk terrain modeling produces maps that visually articulate these environmental and situational contexts.

**RISK** defines the likelihood of a crime event occurring given what is known about the correlates of that event. Risk terrain modeling assumes that all places are risky to some extent (that is, there is rarely zero risk), but because of the spatial influences of certain known correlates of crime, some places are much riskier than others. This is true for many types of crimes in many settings.

**PLACE** refers to a micro-level unit, defined by a raster cell, that represents a real-world location. This provides smaller units of analysis for better precision and permits the modeling of a continuous surface, thereby reducing the need for worries about edge effects or the modifiable area unit problem.
The Risk Terrain Modeling Diagnostics (RTMDx) Utility is a software application for Risk Terrain Modeling (RTM) and for diagnosing spatial crime vulnerabilities. For more information about RTM, visit www.rutgerscps.org.

RTMDx helps to identify and communicate environmental attractors of crime incidents. Information products can be used to anticipate places that will be most suitable for illegal behavior, identify where new crime incidents will emerge and/or cluster, develop place-based interventions, strategically and tactically allocate resources, and prioritize efforts to mitigate crime risks.

RTMDx automates most steps of risk terrain modeling. The algorithm empirically tests a variety of spatial influences and analysis increments for every risk factor input to identify the most empirically- and theoretically-grounded spatial associations with known crime incident locations. Then, it empirically selects only the most appropriate risk factors (with their spatial influences optimally operationalized) to produce a “Best” Risk Terrain Model. The final model articulates the vulnerability for crime with relative risk values at every place throughout the study area. The environmental factors that create specific vulnerabilities at places are listed and weighted according to their relative spatial influence on the outcome event. This aids in the prioritization of risk mitigation efforts.

The diagnostic computation process takes time. In some cases, the “run time” may be several hours (i.e., even double digits) depending on the size of the study area GRID and the total number of variables tested. Note that this software has been rigorously tried and tested to produce valid and reliable results. But, the tradeoff (considering money and time to deployment for public use) is that it currently does not always produce those results quickly for complex models. The speed at which RTMDx operates depends heavily on the processing capabilities of the computer on which it is installed. Therefore, it is highly recommended that you install and operate this software on a desktop computer or server with abundant memory. For large study areas, you may also want to disable “sleep mode” and screen savers to allow your computer to operate for long durations and/or through the night. Also, consider the “Input Parameter Guidelines” discussed on page 9.
System Requirements

RTMDx will run best on a PC (or Intel-based Mac with Boot Camp) with a 64-bit processor, 2 or more CPU cores, and Microsoft Windows installed. It is a Windows-only application. It should not be run on a MAC using Parallels or other virtual Windows application. 16GB or more memory is recommended. RTMDx requires approximately 3.0MB of space on your computer. An internet connection is required for installation.

Input Parameter Guidelines


To improve processing speed and minimize “run time”, it is recommended that you select a cell size that produces a GRID of the study area no greater than 120,000 cells.

The recommended number of risk factors to test for a given outcome event should be made in consideration of “run time” for average computer capabilities, and based on the total “variable count.” The “variable count” should be less than or equal to 100. “Variable Count” = Number of Factors * Number of Analysis Bands * Number of Operationalization Types.

Refer to the chart below to decide how many risk factors your computer can handle.

<table>
<thead>
<tr>
<th>Operationalization Type</th>
<th>E.g., 1 Block, Whole</th>
<th>1 Block, Half</th>
<th>2 Blocks, Whole</th>
<th>2 Blocks, Half</th>
<th>3 Blocks, Whole</th>
<th>3 Blocks, Half</th>
<th>4 Blocks, Half</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: Proximity OR Density</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>25</td>
<td>16</td>
<td>12</td>
<td>8</td>
</tr>
<tr>
<td>2: BOTH Proximity and Density</td>
<td>30</td>
<td>25</td>
<td>16</td>
<td>12</td>
<td>8</td>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>

Recommended Max Number of Risk Factor Inputs

[9]
### Choose Components

Choose which features of Risk Terrain Modeling Diagnostics Utility you want to install.

Check the components you want to install and uncheck the components you don't want to install. Click Next to continue.

<table>
<thead>
<tr>
<th>Select components to install:</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Install GOAL</td>
<td>Position your mouse over a component to see its description.</td>
</tr>
<tr>
<td>Install RTM</td>
<td></td>
</tr>
</tbody>
</table>

Space required: 3.0MB

[Image of installation screen]

[10]
Installation

The setup process will install two programs on your computer: 1) the RTMDx Utility and 2) the Uninstall RTMDx Utility. If you already have a prior version of the RTMDx Utility installed on your computer, uninstall it prior to installing the new version.

Programs (2)

RTM RTMDx Utility
RTM Uninstall RTMDx Utility

RTMDx requires the .NET Framework. If you do not have it installed on your computer you will receive an error message during the setup procedure instructing you to install it first (with directions on how to do so).

To initiate the RTMDx setup procedure, double click on the executable (.exe) installation file. The “Risk Terrain Modeling Diagnostics Utility Setup” dialog box will appear. The setup procedure will automatically install R, GDAL, and RTMDx. You should have administrator privileges for the computer you are using and be connected to the internet. Click the “Next” button to complete the installation process.
Starting the Utility

From the Start Menu: Click on the program name to open the application.

From the Desktop or Taskbar Icon: Click or Double-Click on the program icon to open the application.

GUI (Graphical User Interface) Overview

The RTMDx Utility has three main screens, accessed via tabs. The “Inputs” tab is where you add data and set parameters for the analysis. The “Log” tab displays output for the “Main program log” and the “R script log.” These fields are populated only after the analysis begins. The “About” tab includes copyright and versioning information.
Running a Risk Terrain Model

The manual method of Risk Terrain Modeling involves the following steps:

1. Select an outcome event
2. Choose a study area
3. Choose a time period
4. Obtain base maps
5. Identify all possible risk factors
6. Select model factors
7. Map spatial influence
8. Weight risk map layers
9. Combine risk map layers
10. Finalize maps to communicate information

The RTMDx Utility guides all of these steps. Importantly, it automates steps 6 thru 9, some of the most arduous tasks of manual risk terrain modeling and risk terrain map production. Begin the risk terrain modeling process by giving careful thought to steps 1 through 5 (see also the “RTM Manual” available for free download at http://rutgerscps.org/rtm).

**Step 1: Select an outcome event of interest**

RTM is specifically tailored to the outcome event of interest. Because different crimes have unique and sometimes different underlying causes (i.e., attractors/generators), you should produce separate models for each crime type.

For this tutorial, we will use the crime of **Aggravated Assault**.

**Step 2: Choose a study area for which risk terrain maps will be created**

RTM can be applied to any geographic extent (i.e. local, regional, global; urban, suburban, rural). Select an area for which the information provided by the risk terrain map will be meaningful and actionable (e.g., city-wide jurisdiction, police sector, etc). You should have data that is representative of the entire study area.

For this tutorial, we will use real data from a US jurisdiction. However, we will refer the location generically as “MetroCity.”
Step 3: Choose a time period

The time period should be meaningful for your data and considerate of how the information communicated by the RTMDx output and risk terrain map will be used for decision-making. For crime location diagnostics, use input data that is most representative of the environmental features of the landscape as they were for the time period of the outcome event data being used.

For this tutorial, we will use data representative of calendar year 2012.

Step 4: Obtain base maps of your study area

The study area boundary must be a single feature or multipart polygon Shapefile (.shp). A multipart polygon is a feature that has more than one physical part but only references one set of attributes in the database. For example, a group of islands could be represented as a multipart polygon feature. A collection of police districts could also be represented as a multipart polygon feature. RTMDx will not work properly with other input types, including polyline or multi-feature polygon shapefiles. If your study area boundary is represented by a multi-feature polygon (i.e., there are two or more features in the shapefile, with separate attributes), you must convert it to a multipart polygon prior to inputting it into the RTMDx Utility (See APPENDIX).

The study area boundary must be properly projected. All other inputs (i.e., outcome events, risk factors) must be projected to the exact same coordinate system as the study area boundary.

Step 5: Identify factors that are related to the outcome event

The product of this step should be a comprehensive list of risk factors related to the outcome event. This can be informed by: literature reviews via library databases and/or Google Scholar; reports from reputable research centers (e.g., see rutgerscps.org/pubs.htm, popcenter.org, cops.usdoj.gov, iaca.net, campbellcollaboration.org, etc.); professional experience; and practitioner knowledge. Make a reasonable effort to identify as many factors that you believe to be spatially related to the outcome event in the particular study area. RTMDx can accept up to 30 risk factors as inputs for analysis.
Prior RTM publications have explained that many risk factors do not always produce a better model than a select few factors. Two methods for deciding what risk factors to include in a risk terrain model have been presented. The Ad hoc method is based on insights from various sources and expert knowledge of the study area. It allows for expedient (manual) risk terrain map production and is best used when outcome event data does not exist, when outcome events are under-reported or too few for statistical testing, or if valid and reliable outcome event data is not available. So, the ad hoc method is appropriate for producing risk terrain models for risk assessments that are done prior to the emergence of chronic crime problem places. The Empirical method requires statistical tests of the place-based correlation of each (possible) risk factor on the outcome events. This justifies the use of only the most significantly correlated risk factors in your risk terrain model (from the “pool” of many options). This is the preferred method when outcome event data exists and can be used as input for statistical testing. The RTMDx Utility facilitates the empirical method of risk terrain map production.

**Working with the RTMDx Utility**

Click the “Browse” button to input the study area boundary into the RTMDx Utility. The shapefile path will appear in the “Study Area Boundary (shapefile)” text box. The shapefile name will appear in the “Study Area Name” text box.

![RTMDx Utility Interface](image)

By default the “Block Length” and “Raster Cell Size” are set to 500 and 250, respectively. It is recommended that the block length value be set to the mean length of a block face in your study area. This number has a direct effect on the measures and statistical analyses of spatial influence for your risk factors. For instance, a risk factor that has a proximal spatial influence of 2 blocks would equate to 1,000 feet if the default values were used. Adjust this value to suit your particular study area and analysis needs.

It is recommended that the Raster Cell Size be set to half the “Block Length.” To manually change the raster cell size, uncheck the “Locked” box and enter a value in the “Raster Cell Size” text box. Use the radio buttons to specify whether the Block Length and Raster Cell Size “Units” are in Feet (ft) or Meters (m).
Technically, the Raster Cell Size determines how coarse or smooth the final risk terrain maps will appear; the smaller the cell size, the smoother the map will be. But conceptually, the cell size you select should be meaningful and practical for operational and analytical purposes. It represents a real “place” on a map.

The “**Model Type**” drop down menu allows you to select either “Aggravating” or “Protective.” An aggravating model type assumes that the risk factors input into the RTMDx Utility correlate with the locations of outcome events, and it tests for positive spatial relationships. A protective model type assumes that the risk factors input into the RTMDx Utility correlate with the absence of outcome events, and it tests for negative spatial relationships. Different jurisdictions will likely have different aggravating and protective factors. One example of an aggravating factor might be ATMs for the crime of robbery. An example of a protective factor could be a police sub-station or community garden. The default model type is “Aggravating.”

One should not imply that if a factor is excluded from an aggravating model it is protective instead. These are two very different factor types. Research on “protective” risk terrain models is still in its infancy. Quite frankly, most research to-date on risk terrain modeling has focused on “aggravating” risk terrain models. Therefore, we urge caution in how you interpret the results of a “Protective” model type since the identification of protective factors is not as straightforward as the identification of aggravating factors. There are many statistical complexities when attempting to assess the effect of the presence of one thing (e.g., a protective feature) on the absence of something else (e.g., crime events). Furthermore, intentionally protective factors are often deployed to high crime areas and, so, they will almost always be significantly positively correlated with crime locations (at least in the short term). That is, a geographically-targeted police intervention may be thought of as protective, but because targeted activities focus on high crime areas, the intervention may not be found to be statistically protective in the RTMDx Utility, even if it had a large effect on reducing crimes. As highlighted by this example, the “protective” model type should not be used as an evaluation tool for intervention activities. In the same way that an “aggravating” model type can be used to identify underlying attractors of crime at known hotspots, think of the “protective” modeling function as a means to identify underlying features of an environment that are common to crime “coldspots.”

A conventional method for evaluating the impact of intervention activities focused at high-risk (target) areas is to compare changes in crime counts pre- and post-intervention. Risk-based strategies to combat crime should result in a crime reduction. But another outcome measure of risk-based interventions that are informed by RTMDx should be “mitigated spatial influences” of one or more aggravating risk factors in the risk terrain model. Interventions should aim to reduce crime by changing the situational contexts of places,
rendering them unsuitable illegal behavior settings. Such impacts can be measured via comparisons of pre-and post-runs of the RTMDx Utility. Interventions that aim to mitigate spatial influences of risk factors could be evaluated by re-running the “original” risk factors, using post-intervention crime incident data, under the following hypothesis: “Intervention activities will reduce the relative risk values for risk factors specifically and intentionally addressed by the intervention.” (Ideally, but probably only in rare and exceptional cases, the intervention would have rendered the targeted risk factors insignificant and, thus, excluded from the post-intervention risk terrain model altogether). Continued iterations of this risk mitigation evaluation process may be used to heuristically assess the changing qualities of places and associated criminogenic risks throughout a jurisdiction.

Click the “Browse” button to input the outcome event data into the RTMDx Utility. This data may be in either point shapefile (.shp) format or comma delineated (.csv), with separated XY coordinates, format. The outcome event data must be projected in the exact same coordinate system as the study area boundary. The file path will appear in the “Outcome Event Data” text box. The file name will appear in the “Name” text box by default. Editing the “name” will alter how it appears on the output report.

All outputs will be stored to a folder named according to what is typed in the “Model Name” text box. Use the “Browse” button to specify a location for the output folder. The folder path will appear in the “Destination” text box.

[19]
The “Risk Factors” section allows for the addition and review of risk factors to be considered for the final risk terrain model. The RTMDx Utility allows the input of up to 30 risk factors. Only one risk factor can be added or edited at a time. Click the “Add” button to input a risk factor and set its parameters.

The “Add/Edit Risk Factor” dialog box will appear. Click the “Browse” button to select a risk factor for the analysis. Risk factor data may be in either point shapefile (.shp) format or comma delineated (.csv), with separated XY coordinates, format. File names of risk factor datasets cannot have any special characters, such as dashes, periods, or ampersands (&). Underscores are OK. Rename the files, as needed, before adding them to the Utility. All data must be projected in the exact same coordinate system as the study area boundary. The file path for the risk factor that you selected will appear in the “Risk Factor Data” text box. The file name (without the extension) will appear in the “Risk Name” text box by default. Editing the “risk name” will alter how it appears on the output report.
Select the “**Operationalization**” method. This parameter specifies how to operationalize and then test the risk factor’s spatial influence. “Proximity” (i.e., Euclidean distance) is the default. Proximity assumes that being within a certain distance from a risky feature increases the likelihood of illegal behavior and, ultimately, crime event locations. “Density” assumes that the high concentration area of risky features creates a unique context for illegal behavior and, ultimately, increases the likelihood of crime events at high-density places. Select “Both Proximity and Density” if you are unsure about which operationalization is best and you want the RTMDx Utility to empirically test both and then objectively select the optimal operationalization method. Selecting “Both” could double the analytical run time because there will be twice as many variables tested.

Set the “**Max Spatial Influence**” from the drop-down menu. This parameter specifies the maximum distance for which statistical validity tests will be computed for the particular risk factor. Spatial influence refers to the distances within which risk factor features affect places throughout the environment. Empirical research suggests that spatial influences tend to extend no more than a few blocks. Therefore, the RTMDx Utility is capable of testing up to 4 blocks. Here, a “block” distance is equal to the value entered under “Block Length” on the Utility’s “Input” tab. The default is 3 blocks. So when set to “3 Blocks,” the potential risk posed by the “Proximity” operationalized spatial influence of the risk factor will be assessed at up to 1 block, up to 2 blocks, and up to 3 blocks. Risky places will be defined as cells within these Euclidean distance thresholds. For a “Density” operationalization method, the kernel density bandwidth parameter will be set to values equal to 1 block, 2 blocks, and 3 blocks. Risky places will be defined as cells with density values greater than or equal to 2 standard deviations above the mean density value.
The “Analysis Increments” can be set to “Half” or “Whole”. The default is Whole. “Half” will use half- and whole-block increments when assessing the spatial relationships of outcome events and risk factors at different spatial influences. “Whole” will only use values equal to whole block lengths. So, for example, if “Block Length” equals 500ft, “Analysis Increments” is set to “Half,” and Max Spatial Influence is set to “3 Blocks,” the threshold distances for “Proximity” and the bandwidths for “Density” calculations would be: 250ft, 500ft, 750ft, 1000ft, 1250ft, and 1500ft. Selecting “Half” Analysis Increments could double the analytical run time because there will be twice as many variables tested.

The “Operationalization”, “Max Spatial Influence,” and “Analysis Increments” parameters may be the same among all risk factors or they may differ. It is up to you.

Once all parameters are set for the risk factor, click the “OK” button to add the risk factor as an input. Repeat for each risk factor.

It is highly recommended that you double or triple check all inputs and parameters before clicking the “Run” button. The analysis could take a long time (i.e., up to several hours for complex models in large jurisdictions), so be sure everything is correct before proceeding.

Click the “Run” button to begin the analysis.
A notice will appear reminding you that the analysis may take a long time depending on the model’s complexity. If you want to continue, click the “Yes” button.

The Utility will switch to the “Log” tab to view the main program and R script logs. The “Status” area at the bottom left of the Utility’s window will display the progress of the analysis tasks (it does not represent time remaining, per se). Click the “Cancel” button to terminate the analysis prior to completion.

When the analysis is complete, a dialog box will appear asking if you would like to view the results in your default web browser.

The RTMDx report is in HTML format. It presents a Result Summary, as well as Analysis Input Details, Analysis Parameters, a “Best” Model Specification, an R Text Summary, and instructions for manual Risk Terrain Map Production.
Each individual risk map layer and a (weighted composite) map of the “Best” Risk Terrain Model will be produced as a GeoTiff, which can be imported into GIS software, such as ArcMap®.

**NOTE:** Some versions of the RTMDx Utility do not produce GeoTiffs. Check the “About” tab to see which version of the utility you have. “GeoTiff-producing” versions will say so.

GeoTiffs will be stored in a subfolder located within the output folder (i.e., named according to what you typed in the “Model Name” text box on the Utility’s “Input” tab screen). The HTML report is located within the output folder as well. The HTML report dynamically references some of the other files in the output folder. So keep all files together if you are going to share or archive results.

### Step 9: Combine risk map layers to form a composite map

If your version of the RTMDx Utility produces GeoTiffs, then this step is already done. One of two GeoTiff files of the final risk terrain map may be of interest to you in the output folder: The one that ends with the file name “…output_score” or the one that ends with the file name “…output_prediction.” We recommend working with “…output_score”. “Prediction” values represent the expected count of outcome events in the cells for the same time duration of the input data. For instance, if the model is based on using the last 12 months of crime data, then prediction values represent the expectation of crime counts at each cell for the next 12 months. “Score” values are rescaled to be easier to interpret by dividing the predicted value of each cell by the smallest prediction value among cells. This causes the risk scores to begin at 1 and to represent how many times riskier a cell is compared to the least risky cell. For example, a cell with a risk value of 10 is twice as risky as a cell with the value of 5. For both maps, values are continuous with “real” zeros. Therefore, you can use either of these values as variables in other statistical analyses, such as a control variable for environmental context in a regression model.
If your version of the RTMDx Utility does not produce GeoTiffs, then you will have to produce the final risk terrain map manually using your preferred GIS software. You can produce a risk terrain map by operationalizing the risk factors using the “best” model specifications displayed in the HTML report. Risk factors based upon proximity should be set (e.g., “reclassified” in ArcMap®) to 1 for cells within the distance threshold and 0 for cells elsewhere. For example, in a study area with an average block length of 500ft, a raster map of a risk factor with a spatial influence of ½ block should have all cells within 250ft of the risky feature set to a value of “1” and all other cells set to “0.” Risk factors based upon density should be set (e.g., “reclassified”) to 1 for cells 2 standard deviations above the mean value after applying a kernel density operation of the specified bandwidth and set to 0 for other cells.

Once all risk map layers are manually produced (i.e., only for those risk factors included in the “best” model, not necessarily every inputted factor), combine them through map algebra to produce a risk terrain map and to calculate relative risk scores. For example, using ArcGIS® for Desktop's "Raster Calculator" function you can copy and paste the unique formula that is provided in the HTML report to assign relative risk scores to each cell, updating the risk map layer names as appropriate.

**Example of a formula provided in the HTML report:**

\[
\text{Exp}(-3.7569 + 1.4333 \times \text{DrugMarkeets} + 1.0161 \times \text{PackagedLiquor} + 0.3935 \times \text{Schools}) / \text{Exp}(-3.7569)
\]

**NOTE:** Several types of kernel functions are used by geographic information systems (GIS) to produce density maps. The RTMDx Utility uses the Epanechnikov kernel function which closely matches native ArcMap® kernel density calculations, though not exactly. Tolerances for Join or Clip operations in different software or on different computers may also vary. This means that manually producing (density or proximity) risk map layers in ArcMap® or other GIS software could yield maps and relative risk scores that do not exactly match the RTMDx Utility’s output. The differences, if any, should be slight. This is generally OK and should not be of any concern for most users’ intents and purposes.

**Step 10: Finalize the risk terrain map to communicate meaningful information**

There are many ways to produce actionable intelligence from a risk terrain map. Using your preferred GIS software, adjust the risk terrain map’s symbology, design final map layouts, and/or perform additional analyses based on relative risk scores. It is recommended that you first convert the GeoTiff to GRID format. A GRID is a raster data storage format native to Esri® (See APPENDIX). Work with this new raster GRID layer and/or convert it to a vector shapefile to take advantage of additional GIS tools and functions.

[25]
RTMDx Utility Demonstration

For this demonstration, a risk terrain model will be produced for the crime of aggravated assaults in the study area herein referred to as MetroCity (depicted in the figure below).

The mean length of street segments (i.e., representative of block faces) in MetroCity is 462 feet. So in the RTMDx Utility, we will enter 462 for the “Block Length” and 231 for the “Raster Cell Size.” For simplicity, you may round Block Length to the nearest even number.
Following the steps discussed already, we set up the RTMDx Utility with the following inputs and parameters (see screen capture). Note that this model does not present a comprehensive list of risk factors for testing. It is intentionally simple for demonstration purposes – with only one factor used to show each operationalization method.
The “Density” operationalization method was specified for “Drug Markets” because these areas are defined using drug arrest data. This type of incident data is “fleeting” in the sense that arrests happen at places but do not remain a physical feature of the environment for long duration. For this risk factor, the concentration of data points is the best articulation of the environmental context created by these events and, therefore, the only spatial influence that we want to operationalize for the purpose of defining associated risk at certain places. In other words, “drug markets” are defined by proxy of a high spatial concentration of drug arrests. “High density” of drug arrests is the quality of the environment that we are particularly interested in. We recommend that all similar types of data that have a “fleeting” nature (e.g., calls for service, shots fired, etc.) be tested with only the “Density” operationalization method.

The “Proximity” operationalization method was specified for “Schools” because we know schools to be generally distributed throughout the study area without clustering. So, to save processing (“run”) time, it is reasonable to only test “Proximity” from schools; “Density” spatial influence is not likely to be significant. We recommend that all similar types of data – with spatial patterns that do not cluster, or for which the identification of a “Density” spatial influence would not be practically meaningful for operational purposes, be tested with only the “Proximity” operationalization method.

The “Both Proximity and Density” operationalization method was specified for “Packaged Liquor Stores” because we want to let the RTMDx Utility make an empirical determination as to what spatial influence, if any, these features pose on the locations of outcome events. Packaged Liquor Stores are physical structures of the environment and are not “fleeting” like drug arrests. They also have the potential to be clustered in certain areas, and their clustering could pose different spatial risks than their individual proximities to places in the study area. So we want to test both possibilities.

See the APPENDIX to learn how you can use Nearest Neighbor Analysis to make efforts to reduce your run time by strategically selecting only “Proximity” or “Density” operationalization methods.

After clicking the “Run” button, the “Status” bar will begin to function and the “Log” tab will become active (As shown in the following screen capture).
This analysis took a few minutes to complete. We briefly discuss the results next.

See the APPENDIX regarding an error message that might appear in the “Main program log”: “Error while trying to put something into our datatable”
RTMDx Reports will be “stamped” with the “Model Name” and the Date and Time the run completed. A “Result Summary” box provides a brief overview of the results, including whether a significant risk terrain model was found and how many risk factors it contains. Click on the “Model Specifications” link to jump to the section of the report where details of the risk factors included in the model are located.

The “Analysis Input Details” section of the report provides an overview and some basic descriptive statistics of the original inputs that were entered into the RTMDx Utility for analysis.
The “Analysis Parameters” section of the report provides an overview of the parameters that were set for the risk factors, some statistical information, and data variables that were created and included in the statistical analysis.

The statistical methods for risk factor selection and validity testing that are included in this Utility are far superior to prior manual methods. The statistical methods and procedures were advanced in collaboration with Jeremy Heffner (Azavea). See the following chapter by Heffner that discusses in detail the statistics of the RTMDx Utility. Briefly, the values produced by the Utility from the user inputs are assembled into a table where rows represent cells and columns represent binary variables. In this case, the 3 risk factors generated 24 variables that were tested for significance against the outcome events that occurred within each cell. A Poisson regression model with cross-validation is used to select variables. An iterative process of Bayesian Information Criterion (BIC) score calculations balances model complexity against fit to select significant risk factors and their optimal spatial influence.

Analysis Parameters

<table>
<thead>
<tr>
<th>Name</th>
<th>Feature Count</th>
<th>Operationalization</th>
<th>Spatial Influence</th>
<th>Analysis Increment</th>
</tr>
</thead>
<tbody>
<tr>
<td>DrugMarkets</td>
<td>242</td>
<td>Density</td>
<td>3 Blocks</td>
<td>Half</td>
</tr>
<tr>
<td>Schools</td>
<td>38</td>
<td>Proximity</td>
<td>3 Blocks</td>
<td>Half</td>
</tr>
<tr>
<td>PackagedLiquor</td>
<td>49</td>
<td>Both_Proximity_and_Density</td>
<td>3 Blocks</td>
<td>Half</td>
</tr>
</tbody>
</table>

These 3 risk factors generated 24 variables that were tested for significance. This testing process began by building an elastic net penalized regression model assuming a Poisson distribution of events. Through cross validation, this process selected 17 variables as potentially useful. These variables were then utilized in a bidirectional step-wise regression process starting with a null model to build an optimal model by optimizing the Bayesian Information Criteria (BIC). This score balances how well the model fits the data against the complexity of the model. The stepwise regression process was conducted for both Poisson and Negative Binomial distributions with the best BIC score used to select between the distributions.
The “Best Model Specification” section of the report provides details about the risk factors included in the risk terrain model, their optimal spatial influences and operationalization methods, and their relative weights.

<table>
<thead>
<tr>
<th>Type</th>
<th>Name</th>
<th>Operationalization</th>
<th>Spatial Influence</th>
<th>Coefficient</th>
<th>Relative Risk Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate</td>
<td>DrugMarkets</td>
<td>Density</td>
<td>693</td>
<td>1.4333</td>
<td>4.1925</td>
</tr>
<tr>
<td>Rate</td>
<td>PackagedLiquor</td>
<td>Proximity</td>
<td>1155</td>
<td>1.0161</td>
<td>2.7624</td>
</tr>
<tr>
<td>Rate</td>
<td>Schools</td>
<td>Proximity</td>
<td>1386</td>
<td>0.3935</td>
<td>1.4821</td>
</tr>
<tr>
<td>Rate</td>
<td>Intercept</td>
<td>--</td>
<td>--</td>
<td>-3.7569</td>
<td>--</td>
</tr>
<tr>
<td>Overdispersion</td>
<td>Intercept</td>
<td>--</td>
<td>--</td>
<td>-0.7011</td>
<td>--</td>
</tr>
</tbody>
</table>
Click the “R Text Summary” button to expand the output from the statistical tests. RTMDx utilizes R, an open-source statistical software package (www.r-project.org). Many of the values in the R output are presented in easier-to-read tables within the RTMDx Report. That is why this information is not expanded by default. But, if you like to see “raw” statistical parameters and outputs, this is where they are located.

---

Call: glmulti(formula = crime_count ~ r01d03_DrugMarkets_density_693 + r03p05_PackagedLiquor_proximity_1155 + r02p06_Schools_proximity_1386, sigma.formula = ~1, family = NegativeBinomialII, data = raster.data, method = mixed(3, 10))

Fitting method: mixed(3, 10)

Mu link function: log
Mu Coefficients:

| Term                     | Estimate | Std. Error | t value | Pr(>|t|) |
|--------------------------|----------|------------|---------|----------|
| (Intercept)              | -3.7569  | 0.06164    | -60.945 | 0.000e+00|
| r01d03_DrugMarkets_density_693 | 1.4333   | 0.09800    | 14.825  | 3.585e-48|
| r03p05_PackagedLiquor_proximity_1155 | 1.0161   | 0.08720    | 11.653  | 2.843e-31|
| r02p06_Schools_proximity_1386 | 0.3935   | 0.09071    | 4.338   | 1.448e-05|

Sigma link function: log
Sigma Coefficients:

| Term                     | Estimate | Std. Error | t value | Pr(>|t|) |
|--------------------------|----------|------------|---------|----------|
| -7.011e-01               | 1.129e-01| -6.209e+00 | 5.462e-10|

No. of observations in the fit: 18684
Degrees of Freedom for the fit: 5
Residual Deg. of Freedom: 18599
at cycle: 1

[33]
The “Risk Terrain Map Production” section discusses (in plain language) the values shown in the “Best Model Specifications” section. It also explains how to manually produce the risk terrain map using these values.

For versions of the RTMDx Utility that produce GeoTiffs, this section will also present the file name of the GeoTiff image of the final risk terrain map. View this map by importing it into your preferred GIS software.

![Image](image_url)

Risk Terrain Map Production

The selected risk terrain model was assigned relative risk scores to cells ranging from 1 for the lowest risk cell to 17.2 for the highest risk cell. These scores allow cells to be easily compared. For instance, a cell with a score of 17.2 has an expected rate of crime that is 17.2 times higher than a cell with a score of 1.

You can reproduce these risk scores in common GIS software by operationalizing the risk factors using the "best" model specifications displayed above. Risk factors based upon proximity should be set to 1 for areas within the distance threshold and 0 elsewhere. Risk factors based upon density should be set to 1 for areas 2 standard deviations above the mean value after applying a kernel density operation of the specified bandwidth and set to 0 in other areas.

The 3 manually produced risk map layers can then be combined through map algebra to produce a risk terrain map and to calculate relative risk scores. For example, using ArcGIS for Desktop's "Raster Calculator" function, you can copy and paste the following formula to assign relative risk scores to each cell updating the risk map layer names as needed:

\[
\text{Exp}(3.7569 + 1.4303 \times "\text{DrugMarkets}" + 1.0151 \times "\text{PackagedLiquor}" + 0.3935 \times "\text{Schools}"

) / \text{Exp}(3.7569)
\]

You can also find a GeoTiff of relative risk scores here: geotiff-1234310808/output.csv.1234310808.output-output_score.tif
Statistics of the RTMDx Utility

Chapter Author: Jeremy Heffner, Azavea

Modeling Process
The objective of Risk Terrain Modeling (RTM) is to represent criminal risk by the spatial influences of certain risk factors. The role of statistics in the RTM process is to empirically determine these spatial influences and their significance within a model that balances simplicity with accuracy. User inputs are accepted into the RTMDx Utility in vector format and processed in raster format. Units of analysis are cells within a raster covering that is overlaid onto the study area polygon with enough buffer to account for the spatial influence of risk factors that might fall outside of the study area but which could affect the cell values within the study area. Variables are calculated for each cell within this buffered area. Once all values are calculated, a mask is created to filter only the “study area” cells. This is defined as all cells that have at least some overlap with the study area polygon. The final output is a model of crime risk throughout the study area that is both easy to understand and actionable.

Variable Design
The RTMDx Utility accepts a collection of potential risk factors as inputs -- points in space that represent crime attractors. For each potential risk factor, the Utility determines the geographic reach of its effect by building a series of variables that measure whether each raster cell is within a certain distance of the risk factor (i.e., proximity) or in an area of high concentration of the risk factor (i.e., density). Both processes use a kernel stamping approach to calculate which cells fall within the desired area based upon cell centroids. Proximity calculations use a uniform/step kernel. Proximity variables are represented as 0 (not within distance) or 1/-1 (within distance). Negative values are passed to R when risk factors are considered “Protective” so that the coefficient signs are still positive. Density calculations use the Epanechnikov kernel, which results in an output that is similar to Esri® ArcGIS® for Desktop (ArcMap®). Density variables are reclassified into high-density and other regions. High-density regions are defined by values 2 standard deviations above the mean density value. All values are assembled into a table where rows represent raster cells and columns represent binary variables. The Utility also calculates the number of crimes that are located within each cell to use as the outcome variable.
Culling Variables
Generating such a large number of variables presents potential problems with multiple comparisons, in that, the Utility may uncover spurious correlations simply due to the number of variables it tests. To address this issue, the Utility uses cross-validation to build a penalized Poisson regression model using the ‘penalized’ R package. Penalized regression balances model fit with complexity by pushing variable coefficients towards zero. The Utility selects the optimal amount of coefficient penalization via cross-validation and bypasses the use of statistical significance tests to select variables. This process reduces the large set of variables to a smaller set of variables with non-zero coefficients. It is important to note that using the model resulting from this step, i.e., the penalized model, would be perfectly valid in-and-of-itself. All resulting variables from this process play a useful (“significant”) part within the model. But, since the goal is to build an easy to understand representation of crime risk, the Utility further simplifies the model in subsequent steps (via a bidirectional stepwise regression process).

The penalized regression within the RTMDx Utility is an elastic net model with two fixed L2 penalties and optimized L1 penalties. To optimize the L1 penalty, the Utility forms five stratified folds from the raster cells balancing crime counts between the folds. Since the cells are assigned to folds randomly, different runs of the Utility may result in slightly different variables passing the penalized regression step. In practice, this randomized process rarely results in a different final model because of the subsequent simplification step. The penalized regression also enforces the theoretical form of the model (i.e., Aggravating or Protective) by limiting the fitted coefficient signs appropriately.

So, the process works as follows: Raster cells are bucketed in 5 randomized folds, which are balanced on crime counts (response variable). This balancing process is done to ensure that there is some variance across the folds to aid the numeric stability of the modeling process. The randomization process uses a random seed, which changes on each invocation of the Utility so that cells will be allocated to folds differently for each run of the Utility. Cells are essentially sorted first on the number of events (high to low) then by a random number (to break ties). A sequence of fold allocations is generated such that the first 5 cells are randomly assigned to the 5 folds, then the next 5 cells are assigned, and so forth.

The penalized regression step aims to optimize the amount of penalization. The script uses two fixed L2 penalizations (the ridge components) and optimizes the L1 penalization (the LASSO component) for each of these two fixed values. Since our main objective is variable selection, the L1 component contributes the most to the desired outcome. Providing some fixed L2 penalization increases the chance that two collinear variables are utilized together within the penalized model (one of which will be selected in subsequent steps). For each L2 value, the Utility finds the optimal L1 value by cross validation. This step builds 5
simultaneous models for each particular value of L1 with each model having one fold of the data left out. The accuracy of these models is measured as the average of applying each model to its excluded fold. The two fixed L2 values are arbitrarily selected to provide a small amount of penalization. There is also a command line parameter to the R script that tests more L2 values to potentially include more variables in the penalized model by increasing the L2 penalization (which reduces the L1 penalization). The penalized regression package supports Poisson regression but not overdispersion. This may result in more variables passing through the process than would otherwise be the case if the Utility could model the overdispersion, but since it is mainly using this step to cull variables, having a few extra variables at this stage is not a problem.

The Utility selects the model with the best cross-validated likelihood between the two fixed L2 values. It is possible that numeric instability will cause a particular L2 value to not converge. Fitting two distinct values of L2 reduces the chance that the Utility will error out due to numeric instability.

All variables that have non-zero coefficients in the penalized model are considered potentially useful and are passed through to the subsequent steps.

**Model Building**

To build a more parsimonious model following the penalized regression, the Utility uses a bidirectional stepwise regression process. It begins with a null model (the current candidate) with no variables and measures the Bayesian information criterion (BIC). The BIC score balances model complexity against fit. It individually adds each variable to the null model and measures the resulting BIC scores. The model with the best (lowest) BIC score is selected as the new candidate model. It then repeats the process, attempting to add and remove variables one step at a time to improve the BIC score. It stops when no valid step improves the score. This process results in an optimal model with only the most appropriate risk factors and spatial influences.

Modeling crime as a Poisson process is useful but, technically, crime events are related to one another, which breaks an assumption of the Poisson process. To address this issue the Utility conducts two stepwise regression processes. One model assumes a Poisson distribution. The second model assumes a negative binomial distribution. The negative binomial distribution includes a parameter to represent over-dispersion of counts, which can help to represent dependency between the crime events. At the conclusion of the two stepwise regression processes the Utility selects the model with the best BIC score between the two distributions. This is the final (“Best”) risk terrain model that is delivered in the output report.
The RTMDx Utility uses the ‘gamlls’ R package to conduct the stepwise regressions. By building a custom stepwise regression function within the Utility, rules for how variables enter the model are enforced. For instance, it is required that at most one variable that represents the influence of each risk factor is allowed to enter the model. There is also a check that variables within candidate models are adjusting risk levels in congruence with the theoretical impact of the factors. Finally, there is a check that variables are statistically significant (<=0.05). It is important to note that the Utility is not using significance levels to choose steps (candidate models) within the stepwise regression; it relies on BIC scores for selecting the next step. This additional test is simply another sanity check to make the Utility more robust. In fact, the significance values reported within the final model should not be viewed as regular p-values due to the combination of processes used within the Utility.

The selected model’s fitted values (i.e., predictions) are outputted as the “predictions” column in the output CSV, which is then turned into a GeoTiff. These predictions represent the expected count of outcome events in the cells for the same time duration of the input outcome events. For instance, if the model is based on using the last 3 weeks of crime data, then the predictions represent the expectation for the next 3 weeks for each cell. Expectation values are rescaled to be easier to interpret by dividing the expectation of each cell by the smallest expectation among the cells. This causes the risk scores to begin at 1 and to represent how many times riskier a cell is compared to the least risky cell.

Outputs
The aforementioned statistical process results in an easy to understand model that selects not only the significant risk factors but also their optimal spatial influences. Generating a map of crime risk is simply a matter of applying this model to the variables to generate predicted counts (as attributes) within each raster cell.

Comments
The statistically inclined may raise two points about this approach that should be addressed. First (Many Zeros), even if a user were to model several months of crime data, many cells will likely contain zero events, which may raise a statistical concern. For instance, in aggregating a crime data set of 467 events into 36,752 raster cells in one test jurisdiction, 98.8% of the cells had zero events. At first glance, this seems like a problem. But how many cells should be zero? To answer this question, let’s consider crimes a Poisson process. If crimes were evenly distributed among the 36,752 cells there would be an average of 0.0127 crimes per cell. Simulating draws from a Poisson distribution with this mean value to determine the percent of times it is 0 suggests an answer of 98.7%. Second (Spatial Autocorrelation), the statistical approach does not attempt to incorporate the spatial relationships between cells in the regression process itself. This issue

[38]
often focuses on the fact that regression models assume independent observations and that significance values will be inaccurate when spatial autocorrelation is present among the observations. The Utility’s statistical process deemphasizes the use of significance tests for variable selection in favor of cross-validation. Further, most spatial regression packages do not support Poisson data with low frequencies, which is often the case when analyzing crime counts across a raster.

Conclusion
The novel statistical approach used within the RTMDx Utility allows the user to be confident in the analytic results due to the automatic application of several modeling steps. By using cross validation, the Utility empirically examines the importance and spatial influence of risk factors by testing models against withheld data. By further simplifying models via a stepwise regression, the Utility balances model simplicity against accuracy.
Epilogue

Crime hotspots are symptoms of risky places. A sole focus on crime hotspots is like observing that children frequently play at the same place every day and then calling that place a hotspot for children playing, but without acknowledging the presence of swings, slides and open fields. These features of the place (i.e. suggestive of a playground) attract children there instead of other locations absent such entertaining features. Hotspots of crime serve more as a proxy measure of places where the dynamic interactions of underlying attractors of crime exist or persist over time. The RTMDx Utility diagnoses the specific factors at micro places that affect and enable the seriousness and longevity of crime problems. The RTMDx Utility identifies places where crime is likely to emerge based on relative risk values. This allows public safety agencies and other stakeholders to be proactive in their intervention strategies. Don't chase the hotspots. With RTM, you can identify, validate, communicate and focus on their causes and try to effectively prevent them.

Appendix

Visit www.rutgerscps.org/software for the most current and up-to-date appendices and supplemental resources.