2013

Modeling Of Socio-economic Factors And Adverse Events In An Active War Theater By Using A Cellular Automata Simulation Approach

Halil Bozkurt

University of Central Florida

Showcase of Text, Archives, Research & Scholarship

MODELING OF SOCIO-ECONOMIC FACTORS AND ADVERSE EVENTS IN AN ACTIVE
WAR THEATER USING CELLULAR AUTOMATA SIMULATION APPROACH

by

HALIL BOZKURT
B.S. Electrical and Electronics Engineering, Yeditepe University, 2006
M.S. Engineering Management, Yeditepe University, 2008
M.E. Engineering Management, Rochester Institute of Technology, 2008

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Industrial Engineering and Management Systems
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Summer Term
2013

Major Professor: Waldemar Karwowski

ABSTRACT

Department of Defense (DoD) implemented Human Social Cultural and Behavior (HSCB) program to meet the need to develop capability to understand, predict and shape human behavior among different cultures by developing a knowledge base, building models, and creating training capacity. This capability will allow decision makers to subordinate kinetic operations and promote non-kinetic operations to govern economic programs better in order to initiate efforts and development to address the grievances among the displeased by adverse events. These non-kinetic operations include rebuilding indigenous institutions' bottom-up economic activity and constructing necessary infrastructure since the success in non-kinetic operations depends on understanding and using social and cultural landscape. This study aims to support decision makers by building a computational model to understand economic factors and their effect on adverse events.

In this dissertation, the analysis demonstrates that the use of cellular automata has several significant contributions to support decision makers allocating development funds to stabilize regions with higher adverse event risks, and to better understand the complex socio-economic interactions with adverse events. Thus, this analysis was performed on a set of spatial data representing factors from social and economic data. In studying behavior using cellular automata, cells in the same neighborhood synchronously interact with each other to determine their next states, and small changes in iteration may yield to complex formations of adverse event risk after several iterations of time. The modeling methodology of cellular automata for social and economic analysis in this research was designed in two major implementation levels as follows: macro and micro-level. In the macro-level, the modeling framework integrates

population, social, and economic sub-systems. The macro-level allows the model to use regionalized representations, while the micro-level analyses help to understand why the events have occurred. Macro-level subsystems support cellular automata rules to generate accurate predictions. Prediction capability of cellular automata is used to model the micro-level interactions between individual actors, which are represented by adverse events.

The results of this dissertation demonstrate that cellular automata model is capable of evaluating socio-economic influences that result in changes in adverse events and identify location, time and impact of these events. Secondly, this research indicates that the socio-economic influences have different levels of impact on adverse events, defined by the number of people killed, wounded or hijacked. Thirdly, this research shows that the socio-economic, influences and adverse events that occurred in a given district have impacts on adverse events that occur in neighboring districts. The cellular automata modeling approach can be used to enhance the capability to understand and use human, social and behavioral factors by generating what-if scenarios to determine the impact of different infrastructure development projects to predict adverse events. Lastly, adverse events that could occur in upcoming years can be predicted to allow decision makers to deter these events or plan accordingly if these events do occur.

I dedicate my dissertation work to my loving parents Kadriye and Mete Bozkurt, and my brother Şükrü Bozkurt who encouraged and supported me each step of the way during this effort.

I also dedicate my dissertation work to those who guided me when I was trying to find my way in life.

ACKNOWLEDGMENTS

I wish to thank my committee members who were more than generous with their expertise and precious time. A special thanks to Dr. Waldemar Karwowski, my committee chair, for his guidance, encouragement and, most of all, his patience throughout the entire process. Thank you Dr. Gene Lee, Dr. William Thompson and Dr. Piotr Mikusinski for agreeing to serve in my committee.

Thank you Dr. Tareq Ahram and Dr. Nabin Sapkota who were so generous with their time and expert knowledge. I also wish to thank to Dr. Ahmad Elshennawy for his encouragement and support.

Finally, I would like to thank the faculty and staff in the Industrial Engineering and Management Systems department for assisting me with this research and made my interactions with the department an enjoyable experience.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE: INTRODUCTION

1.1 <u>Background</u>

Department of Defense (DoD) implemented Human Social Cultural and Behavior (HSCB) program to meet the need to develop capability to understand, predict and shape human behavior among different cultures by developing a knowledge base, building models and creating training capacity. This capability will allow decision makers to subordinate kinetic operations and promote non-kinetic operations to govern economic programs better in order to initiate efforts and development to address the grievances among the displeased by adverse events. These non-kinetic operations include rebuilding indigenous institutions' bottom-up economic activity and constructing necessary infrastructure since the success in non-kinetic operations depends on understanding and using social and cultural landscape. This study aims to support decision makers by building a computational model to understand socio-economic factors and their effect on adverse events.

On August 2010, a workshop was organized by the Planning Committee on Unifying Social Frameworks which was convened by the Board on Human-Systems Integration (Frameworks & Council, 2011). At this workshop Major General Michael T. Flynn, U.S. Army, delivered the keynote. MG Flynn was in Afghanistan as the deputy chief of staff, intelligence (CJ2) for the International Security Assistance Force (ISAF). In his keynote, he provided a background by describing how the military dealt with various socio-cultural issues in Afghanistan.

In this keynote MG Flynn highlighted two points by the experience in Afghanistan. The first point was the change in the definition of the warfare. He defined twenty-first century

warfare as information and intelligence, and its subcomponents as precision, perception, and

understanding, which were more than twentieth-century warfare; fire and maneuver, which

components were speed, distance, and lethality. MG Flynn exemplified warfare with the

technological capability of the US military to hit a target with a missile. The process of targeting

the missile process is composed of five components; (1) find, (2) fix, (3) finish, (4) exploit, and

(5) analyze (Frameworks & Council, 2011). He explained the importance of information and

intelligence as the four components of process of targeting the missile process, which are find,

fix, exploit, and analyze. He highlighted the importance of these components as the aim of the

missile. He also noted that the finish component is the only kinetic event in the process.

MG Flynn explained twenty-first-century warfare in a non-kinetic social engineering

construct in which components transform from find, fix, and finish to find, feel, and understand.

This non-kinetic construct aims to change the sense of the community, understand the perception

of the community on the decisions made and the actions taken. This can be exemplified as

understating disturbances in a district or a tribe and identifying their needs. These needs can be

money or basic services such as building a school or constructing a road. In twenty-first-century

warfare exploration and analysis kept their importance due to the large volumes of information.

In addition, warfare exploration provides advantage to understand, or to further the decision

making process

MG Flynn exemplifies the need for a non-kinetic operation with a story about a school

that was built in an Afghanistan village, which had no one willing to teach. This

misunderstanding of the environment caused the five year old school to decay and become a safe

house for the Taliban.

Reward of ONR for the HSCB program by supplying necessary data for this dissertation, and the end result derived from MG Flynn's keynote are the primary motivations for this research. The end result from MG Flynn's keynote can be summarized as the necessity of understanding the effects of socio-economic / cultural factors on adverse events. The aim of HSCB program can be summarized as building capability development by creating a knowledge base, building models to predict and shape human behavior among different cultures. Blending end results of the keynote and the aim of the HSCB program this study intends to support decision makers by building a computational model to understand economic and cultural factors and their effect on adverse events.

In this dissertation, the analysis demonstrates that the use of cellular automata has several significant contributions to support decision makers to allocate development funds to better stabilize regions with higher adverse events risk, and to better understand the nonlinearity and complex socio-economic interactions with adverse events.

1.2   Research Gap

There is currently limited research on modeling the effects of socio-economic factors on population in a war theater. However there are studies for predicting levels of violence, crime and terrorist incidents (Fridman, Zilberstein, & Kaminka, 2011; Russo, Roccato, & Vieno, 2011; Siebeneck, Medina, Yamada, & Hepner, 2009; Spicer, Reid, Ginther, Seifi, & Dabbaghian, 2012): It is observed from the literature that the effects of socio - economic factors in population are mostly investigated for disparities on healthcare and urban growth (Alegría, Pescosolido, Williams, & Canino, 2011; Mago et al., 2012; Pearson & Geronimus, 2011).

3

1.3   Research Objective

The main objective of this dissertation is to investigate the relationship between socio-economic factors and adverse events in an active war theater using the cellular automata modeling approach. This objective is divided into three specific objectives.

The first specific objective of this dissertation is to investigate associations between socio-economic factors and adverse events in active war theater.

The second specific objective of this dissertation is to develop a cellular automata model that is capable of evaluating socio-economic influences that result in change in adverse events, and identifying location, time and impact of these events.

The third specific objective is to apply the cellular automata modeling approach to enhance the capability to understand and use human, socio-economic and behavioral data by generating what-if scenarios to determine the impact of different infrastructure development projects on adverse events in active war theater.

CHAPTER TWO: LITERATURE REVIEW

2.1 <u>Introduction</u>

Cellular automaton (CA) has many military applications including but not limited to the areas of cryptology, intrusion detection, modeling cultural change, policy making. However, there is limited research addressing the effects of socio-economic factors on adverse events. This research focuses on the effects of socio-economic factors on a population using CA via geographic information system (GIS). The following sections describe related literature on cellular automata and its applications for socio-economic factors and GIS.

2.2 <u>Cellular Automata</u>

CA consists of identical computing cells which locally and synchronously interact within each other to determine their next state. However small the change of state of a cell in one iteration, after numerous iterations these small changes create complex patterns at the final condition of the system. This allows CA to be a potential tool for simulating different sophisticated natural phenomena.

CA is the oldest model of natural computing, spanning to over half a century (Kari, 2005). Construction of universal cellular automata was initiated by von Neumann in the 1940s (Von Neumann, Burks, & Burks, 1966).

In the simplest form of CA, space is represented by a tessellation. A tessellation is a uniform bi-dimensional collection of cells, where cells represent the population involved. Each iteration is done based on a set of rules. These rules represent a parametric set of conditions which determines the next state of cells according to its neighboring states in discrete time t. The

change in behavior of the CA could be simply controlled by these rules to enable meaningful results.

2.2.1   History of Cellular Automata

In 1937, Turing introduced a universal machine. Turing machines, with input (m,n), computed the same value as the machine encoded m on input n (Turing, 1937). In 1966 von Neumann introduced a CA model to construct a self-reproducing machine, exemplifying the nature of life. It was a bi-dimensional, 26 state, CA which was capable of simulating similar characteristics with Turing machine(Von Neumann et al., 1966). The simulation of a Turing machine is done with particular sets of Boolean circuits. Following the von Neumann's idea on self-reproduction, Codd designed a CA that reduces the complexity of the automaton to 8 state (Codd, 1968). Banks re-encoded von Nuemans CA and lead to construction of bi-dimensional four state CA model (Banks, 1970).

In 1970, Conway (Conway, 1970) introduced the famous "CA Game of Life", which can be seen as a game played on an infinite grid. At any time some of the cells are alive and the others killed. According to the game, the player chooses the initial configuration of alive cells at initial time. Then the state of each cell at any time follows the previous one according to the rules of the game (Durand, 1999).

In 1977, Toffli showed the existence of computation and construction of reversible CA by embedding any d-dimensional CA in to one having d+1 dimensions (Toffoli, 1977). Margolus build a bi-dimensional reversible CA that has 16 states defined by two by two lattice blocks. He simulated a bouncing billiard ball. However, this simulation was too complex to compute any reversible Boolean function of conservative logic (Margolus, 1984).

6

In 1997, Durand-Lose improved this by constructing a one-dimensional CA simulating any CA (Durand-Lose, 1997) which was first introduced by Albert and Culik (J. Albert, 1987).

2.2.2   Mathematical Definition

Cellular automata can be defined as a cellular state space $\mathcal{L}$ which is composed of a set of Finite State Automata (FSA). Cellular state space is normally generated by a regular n-dimensional grid structure of FSAs. In two dimensions cellular state space can be exemplified as a cellular automaton that consists of a two dimensional grid structure of $\mathcal{L} = m \times n$ where each square is called a cell and each cell is represented by a FSA.

The compositions of all possible states of the cells in a defined neighborhood create a state space of the CA. The state ς of each FSA can be in one of the states that are defined as local value space $\sum$ which represents all possible states for each FSA.

$$\varsigma \in \sum \; \equiv \{0,1,2,3,4,\dots,k-1,k\} \tag{1}$$

A neighborhood $\mathcal{N}$ is generated by a set of $\mathcal{N}$ topologically neighboring cells, which alters a change of state of each updated cell for all iterations of time. Figure 1 shows examples of hexagon, von Neumann, Moore and Margolus neighborhoods. Boundary conditions of the neighborhood can be defined as periodic, fixed or reflecting among others (Hoekstra, Kroc, & Sloot, 2010).
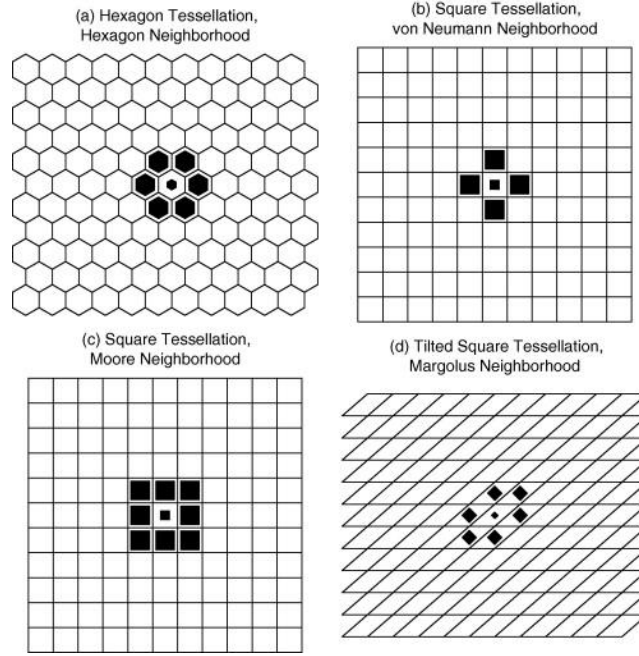
Figure 1 Various types of neighborhoods(White & Kiester, 2008)

The change of the state of each updated cell from its state at time t_0 to the state at t_1 is controlled by a transition rule ϕ. A transition rule applies to all cells of the neighborhood synchronously at an externally provided clock step. A transition rule is defined as;

$$\phi: \overbrace{\Sigma \times \Sigma \times \ldots \Sigma}^{\mathcal{N}} \rightarrow \Sigma \qquad (2)$$

Von Neumann neighborhood with radius of one is defined as below. It covers four neighboring cells and one centering cell which is updated on each iteration of time.(Figure 1 -b).

$$\mathcal{N}_m^1(i,j) = \{\sigma_{k,l}|\ |i-k|+|j-l| \leq 1\} = \{\sigma_{i,j}, \sigma_{i-1,j}, \sigma_{i,j-1}, \sigma_{i+1,j}, \sigma_{i,j+1}\} \qquad (3)$$

Using states of cells in the neighborhood $\varsigma(t)$, transition rule $\phi(t)$ can be expressed as a function of all cells in the neighborhood.

$$\varsigma_{i,j}(t+1) = \phi\big(\varsigma_{k,l}(t)\big|\varsigma_{k,l}(t) \in \mathcal{N}\big) \qquad (4)$$

8

$$\varsigma_{i,j}(t+1) = \phi\big(\varsigma_{i,j}(t), \varsigma_{i-1,j}(t), \varsigma_{i,j-1}(t), \varsigma_{i+1,j}(t), \varsigma_{i,j+1}(t)\big) \qquad (5)$$

Cellular automata can be simply exemplified with an elementary one dimensional CA rule (rule 30) introduced by Wolfram (Wolfram, 1983) is shown in Figure 2 below. Wolfram names rules according to the decimal equivalent of the binary representation of its rule outcomes (30=00011110). In this example rules (ϕ) are defined as function of 3 neighbors including the center cell. States which are first three cells at Figure 2-b involves one black cell in the center and two black cells as neighbors. First rule in Figure 2-c defines that if two neighbors of black cell are black than the next state of center cell is white. And these rules are applied to each cell to obtain the state change given in Figure 2-d.

a- Neighborhood

b- Black And

White

c- Rules

d- Example of

state Change
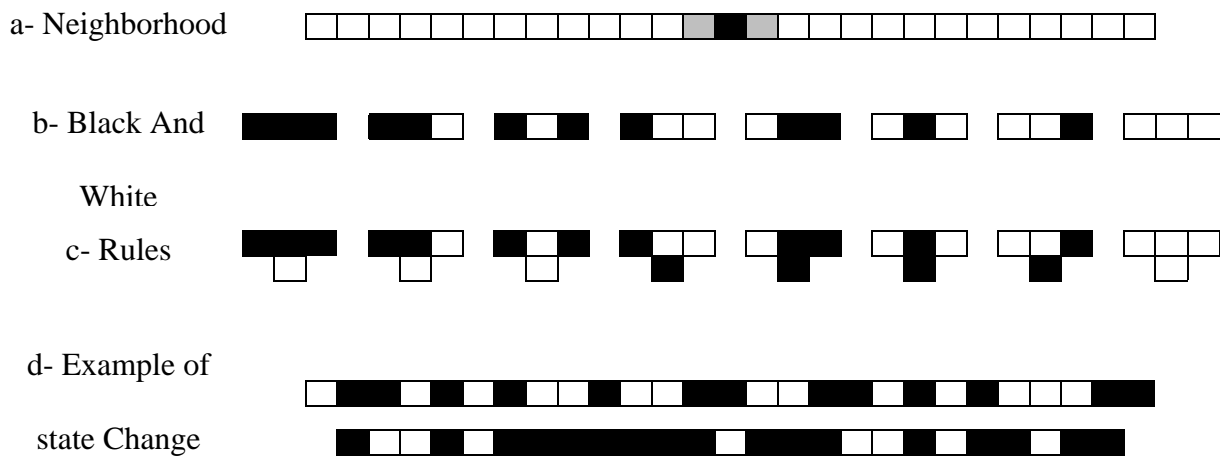
Figure 2 Example of simple cellular automata

2.3   Geographic Information System (GIS)

The geographic information system (GIS) is a capable tool for turning large volumes of spatial data into useful information (Curran, 1984). Many researchers (Jat, Garg, & Khare, 2008; Peccol, Bird, & Brewer, 1996; Yu & Ng, 2007) implement GIS as an advantageous tool to

explore the spatio-temporal characteristics of landscape. They have adopted GIS as a data integrator to conduct spatial data analysis. They also used GIS to visualize the results by taking intermediate results, and facilitating their comparison with the original and other layers of information. They have also used GIS to integrate predictions into the spatial data, to allow planning decisions to be made.

2.4   Studies on Socio-Economic Factors, Cellular Automata and GIS

Vaz (Vaz, Nijkamp, Painho, & Caetano, 2012) proposed a progressive evaluation of growth in densely populated areas in Portugal's Algavre region. This assessment depended upon a sophisticated nonlinear system modeling that used the cellular automata which brought together qualitative data and measurable spatial methodologies. A cellular automaton compares and validates the results, and then draws the trends of future land use through this validation. CA, applied to urban growth in this study, attempts to turn into an urban or a nonurban form in a certain period by the iteration of the predetermined dimensional cell that is determined by social, economic and geographical input. The author suggests that the simplicity of the model accompanied by possible complex and interlacing variables makes this method a focal point to comprehend areas with quick changes in their urban structure. Different versions of the cellular automata associated with Markov transition rules generates 2020 urban growth scenario. The authors claim that cellular automata used within qualitative modeling methods produced never seen before results to model alternative scenarios of future urban design trends.

Sheng (Bing Sheng & Sui, 2008) incorporated remote sensing imagery and social and economic data into a cellular automata based model to examine how global and domestic factors interact with each other in densely populated Asian structures. The complex interaction between

10

two types of dynamics urban and urban spatial is simulated by the cellular automata. This study incorporated the foreign direct investment (FDI) and the change in economic movements to remote sensing data and GIS in an attempt to develop an Asian urban model that is based on cellular automata as it suggested these essential indicators influence globalization on Asian urbanization. Processes with different scales and small scale transformation are two approaches to practically explain the socioeconomic impacts on land use in densely populated and rural areas. These approaches are also used to implement urban dynamics through the CA modeling. The features of Asian urbanization clearly differ from those of Western societies. With the help of combined hierarchical approaches, the study quantitatively draws the distinct characteristics of Asian urbanization processes, and generates the CA-based model to express the significant impact of globalization on Asian urban dynamics with domestic economic/demographic drivers.

Han (Han, Hayashi, Cao, & Imura, 2009) proposed a study that presents an integrated system dynamics and CA model in the analysis of socio-economic drivers and the assessment of urban spatial forms. The authors try to understand and forecast the dynamics of urban development where extensive socioeconomic transition takes place with the integrated spatial design and CA model. The integrated model is proven to be capable of observing and predicting urban growth dynamics. Problems were mainly centered on two aspects: (1) transformation of urban land caused by social and economic attributes depending on geographic location and (2) spatial form of urban spread which is shaped by material factors.

Fuzzy cognitive maps and cellular automata: This is a developmental approach for social systems modeling (Mago et al., 2012). This study used large scale conclusions to describe small scale interactions between individuals by parameters. As a result, the proposed modeling

technique combined the macro-level approach with the micro-level method yielding an adaptable, evolutionary modeling technique. Parameters were recalculated using the statistics feed back to the macro-level model from the micro-level model. The study identified fuzzy cognitive map (FCM) as the macro-level model, and CA as the micro-level. Multiple FCM's and CA's are permitted to interact. Human immunodeficiency virus spread among drug users who share needles during drug injection, the practice claimed to be the most common way of HIV transmission among injection users, was modeled and simulated to prove the relevance of proposed modeling technique.

Wu (Wu, Liu, Wang, & Wang, 2009) coupled an artificial, neural network based stochastic cellular automata method with a number of socioeconomic measures such as; income of the non-urban population per person, income of the urban population per person, population and gross domestic products (GDP) of highly populated areas. The NNSCA model was proven to adequately replicate historical urban expansion, and to produce sensible urban forms. The study also provided practical information to evaluate environmental ramifications of urbanizing.

Lauf (Lauf, Haase, Hostert, Lakes, & Kleinschmit, 2012) introduced an enhancement of the CA model by integrating system dynamics (SD). He affiliated household dynamics and housing decisions as drivers of residential development. CA macro models were used to replicate the change in quantitative land-use (LUC) for densely populated areas that didn't have significant dynamics stimulated by the shift in demographics. This model focused on household dynamics in accordance with the concept of the second demographic transition (SDT). The model applied aging and population contraction to reproduce relevant effects on residential

preference and the proceeding LUC. The author stated that the study benefitted from combined

SD and CA reducing their individual disadvantages.

CHAPTER THREE: RESEARCH DATA

3.1   Introduction

This section provides facts and revisions conducted on datasets used in this study which are provided by HSCB Project management. Datasets provided are grouped in two major groups: 1) Adverse events datasets, which include details of adverse events occurred in Afghanistan, and 2) Aid datasets, including details of development or economic projects that Afghanistan had received.

Provided adverse events datasets cover data collected through news feeds, blogs, reports, and through various databases. Collected data is processed by text processing techniques and analytics to generate suitable set of data to be used in the content of HSCB project. Provided Aid datasets are collected through US and regional databases. Provided datasets are 75 percent accurate.

Provided research data files are shown in Table 1. Each aid and event dataset contains a report date, incident number and a marked geographical location. The structure of data and content suitability for CA analysis enabled us to collate both datasets in a spatial database. However all datasets includes the location content; some of the data is missing this information. This disturbs the consistency of the datasets. Missing location information is mostly seen in events datasets. Additional to geographic location and date content, event datasets contain information on adverse events (killed, wounded and hijacked). Some of the datasets also contain information on incident type (complex attack, kidnapping, threat, arrest ...).

In addition to the supplied datasets, ArcGIS shape files which were needed for geographical location tagging of HSCB data were downloaded from Humanitarian response community open web portal (Humanitarian Response, 2011).

Table 1 Research Data Files

Aid Data	Events Data
USAID-IRD-AISICS	AFGHANISTAN Events
AidData-Projects data Afghanistan	JTIC
CERPData normed	Newly Updated ISGJ
CERP Projects-6 March 2011	Sigacts 2009
USAID Afghan Info Data thru Q1 2011	Sigacts 2010
USAID GeoBase Data	USaid
MRRD Projects Database	Witsgeo

In addition to geographic location and date content, aid datasets contain information on project sector (Table 2), project activity types, project start and end dates and allocated budget. Construction type, population count, and detailed project information is also available in some of the datasets.

Table 2 Sector names in US Aid data

Sector Name	
Agriculture	Environment
Capacity Building	Gender
Commerce & Industry	Governance
Community Development	Health
Education	Security
Emergency Assistance	Transport
Energy	Water & Sanitation

3.2   Visualization of Untouched Research Data

This study aimed to present any possible initial relationship between aid data and adverse events on untouched data. This work has been done by using ArcGIS software which uses Microsoft structured query language (MSSQL) to manipulate the spatial data. During the preliminary study, it was observed that, US aid has influence on adverse events.

Over seven event datasets provided, "witsgeo.csv" dataset is used to present the spread of adverse events. However other data sets have more data for specific time periods, "witsgeo.csv" data has largest spread over 2005 - 2010 time period.

Figure 3 below shows the legend of the maps, which is created using ArcGIS. These maps visualize adverse event density and the amount of aid that has been received on regional base. Using project budget instead of aid disbursement may led to more realistic observations however, the regional budget is not defined for nationwide projects. After collating supplied data in a geospatial database it should be possible to get definitive results.
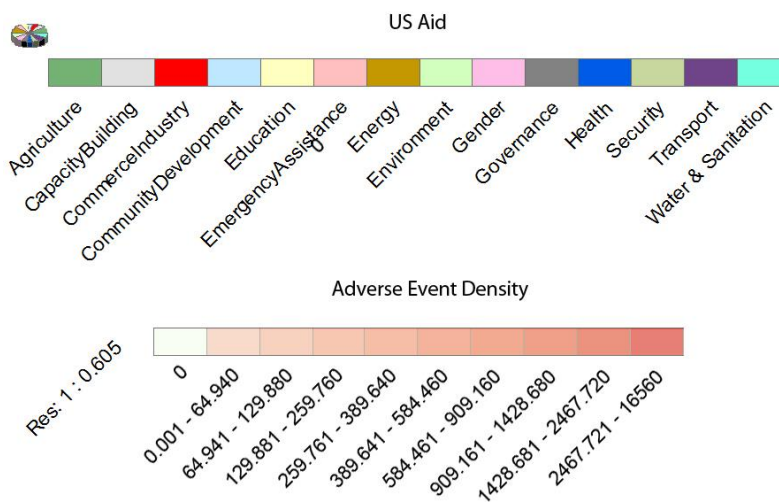
Figure 3 Legend of bare data visulization
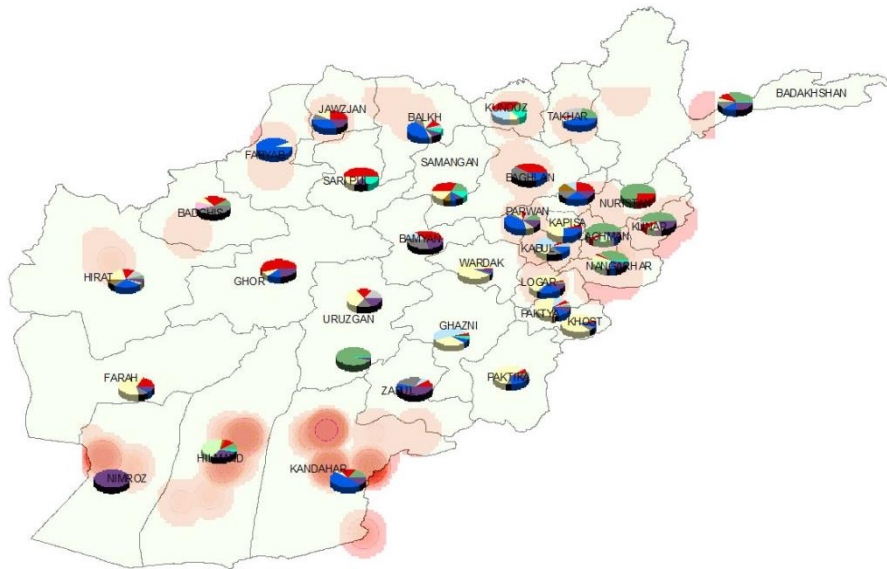
Figure 4 Adverse events and US Aid during 2005 -2006
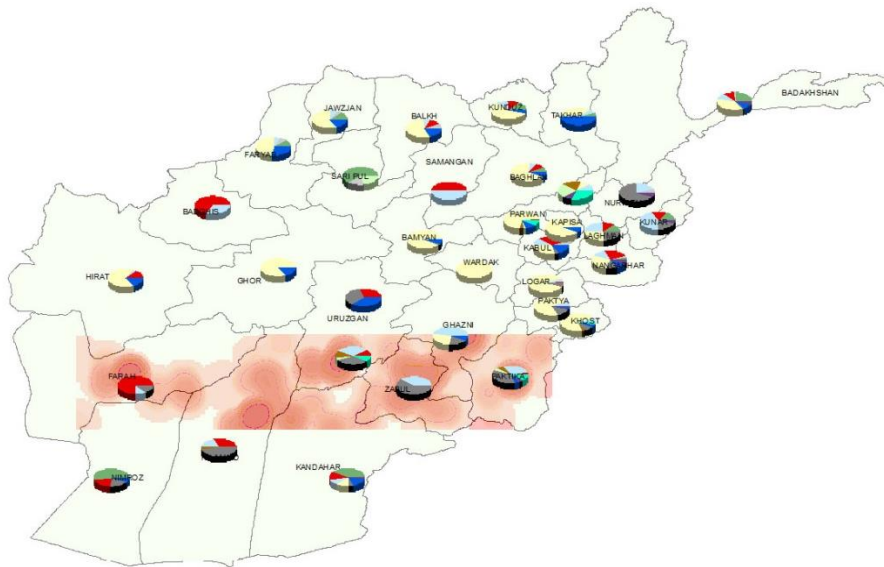
Figure 5 Adverse events and US Aid during 2007 -2008
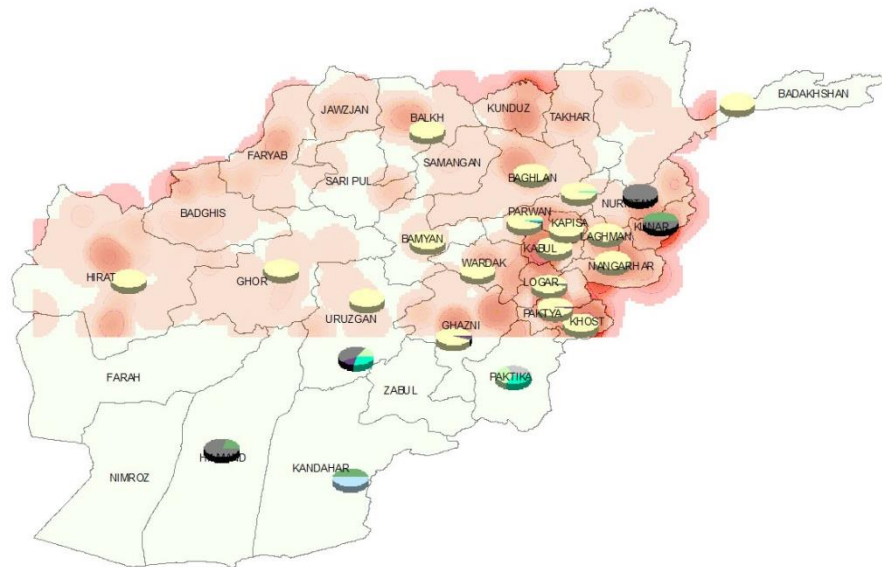
Figure 6 Adverse events and US Aid during 2009

In the USAid data that aid-based projects detailed with project start and end dates. It is not possible to visualize the timeframe for ongoing projects using GIS. This initial visualization of untouched data accepts the midpoint of the aid period, as a time that aid is conducted.

In Figure 4 it is observed that the density of adverse events is not high in regions where aid is received on education sector. It is also observed in Figure 5 and Figure 6 that as the aid received on education decreased, the adverse events are increased.

It is also observed that agriculture aid is being effective over time. It is also observed that provinces which received agriculture aid lowered the density of adverse events happening. Those provinces are kept to encounter with lower amount of adverse events even though they have received little or no amount of agricultural aid in the following years. This can be observed in Kandahar and Nimroz provinces.

3.3   Aid Data

The USAid MS Access dataset is a database file in which most of the relationships among all tables are pre-defined and indexes are presented. In addition, a detail level structure is provided for all data, which is the reason why research team used this data among other files to generate the complete needed economic and development aid dataset. The relationships among various files, needed dataset to perform this research, are represented in Figure 7.
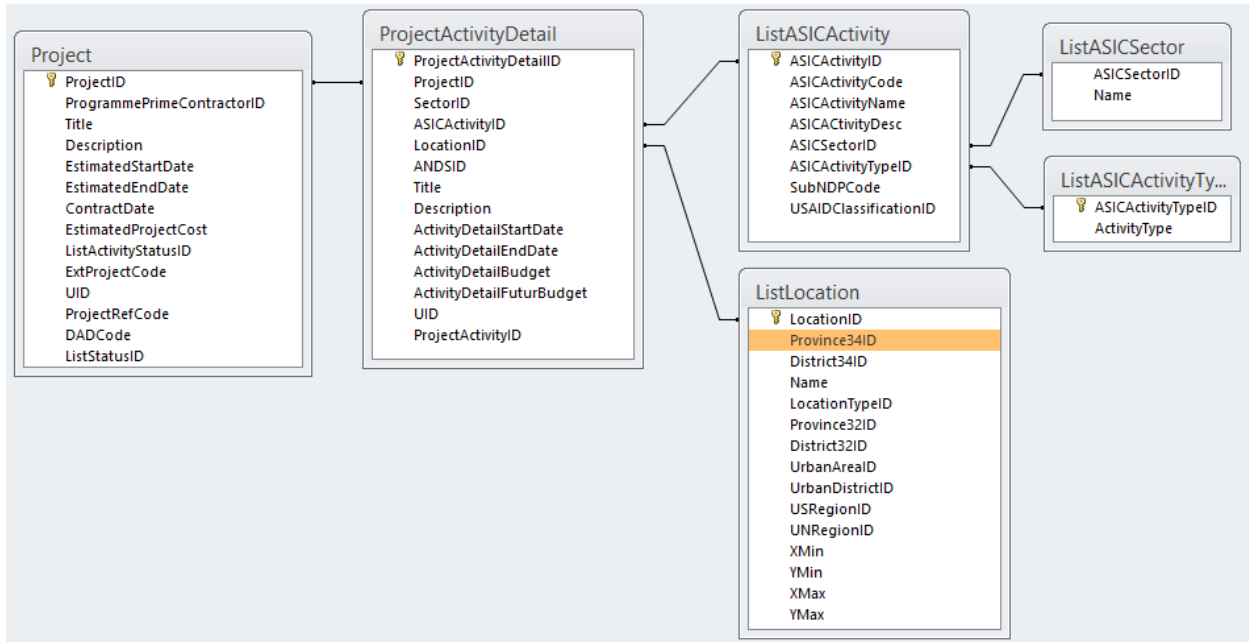
Figure 7 USAid Table relationships used for data set generation

In Figure 7, data file "Project" is the highest level of data that is used for this research. Project data branches to "project activity detail", so that each project has one or more project activity detail. Project activity detail is linked to "location", "sector" and "activity type". For example; the project type "education" can have two different districts with two different activity types such as "construction" and "advisory service".

Although USAid Database is well organized, it is missing some content such as project start and end dates or project budget and activity detail budget. Some data points only have detailed project start and end dates while others only have estimated project start and end dates or some data points only have activity detail budget while others only have estimated project cost.

Another limitation of this database is that data did not branch into province and district level for country wide projects, such as media strategy, community development or tax implementation projects. This is the same issue for province level projects data. A Microsoft Access Query with the relationships showed in Figure 7 returned (22,333) data points. Table 3 explains source tables, table field names and data generated as an output of the query.

Table 3 Structure of USAid query output

Source Table Name	Field Name	Data
Project	ProjectID	Identification of project
Project	Estimated Start Date	Estimated start date in mm/dd/yyyy
Project	Estimated End Date	Estimated end date in mm/dd/yyyy
Project	Estimated Project Cost	Estimated project cost in US Dollars
ProjectActivityDetail	ProjectActivityDetailID	Identification of project activity detail
ProjectActivityDetail	SectorID	Identification of ASIC sector
ProjectActivityDetail	Location ID	Identification of location
ProjectActivityDetail	ActivityDetailStartDate	Detail start date in mm/dd/yyyy
ProjectActivityDetail	ActivityDetailEndDate	Detail end date in mm/dd/yyyy
ProjectActivityDetail	ActivityDetailBudget	Detail project cost in US Dollars
ListLocation	Province34ID	Identification of province
ListLocation	District34ID	Identification of district
ListASICSector	Name	Sector name (1-Agriculture, 2-Capacity Building, 3-Commerce & Industry, 4-Community Development,5- Education, 6-Emergency Assistance, 7-Energy, Emergency Assistance, 8-Energy, Environment, 9-Gender, 10-Governance, 11-Health, 12-Security, 13-Transport, 14-Water & Sanitation)
ListASICActivityType	ActivityType	Activity type(1-Advisory services, 2-Construction, 3-Logistics support, 4-Rehabilitation, 5-Repair, 6-Service delivery)

Database queries have been used to overcome the above mentioned data limitations. The analysis conducted on USAid database are as described below:

1. Country level and province level projects branched into district level.

2. Project activity details, if start and end dates are not available, an estimated project start and end dates are assigned as project activity detail start and end date.

3. Project budget is generated by subtracting sum of available project detail budgets for the same project ID from estimated project cost and, divided by the count of unavailable project detail budget for the same project id.

4. Activity type assigned as sub-category of sector name. For example; "advisory services for education" is represented by "501" and "service delivery" for health sector is represented by "1106".

5. Projects are grouped for each month according to project start and end date. This is done to measure ongoing project budget and count in district for each month. At this level of data processing, a total of "1844519" data points are generated.

USAid projects were grouped by project budget and count of the USAid projects according to the year, month, province ID and district ID and with respect to sector type and activity type.

Generated datasets were represented by "88" columns and "43200" rows. Columns content are; year, month, province ID, district ID and combination of 14 ASIC sector types with 6 ASIC activity types. Rows are grouped by years from (2002) to (2010), based on the 12 months, and a total of (400) districts. Figure 8 and Figure 9 show parts randomly generated data. Table 4 shows structure of generated aid datasets.

Data_Year	Data_Month	PROV_34_ID	DIST_34_ID	101	102	103	104	105	106	201	202	203	204	205	206	301	302
2007	4	6	602	9	7	1	3	7	5	9	10	7	5	2	7	8	1
2007	4	6	603	6	2	3	2	1	2	2	4	5	8	1	1	6	7
2007	4	6	604	9	9	8	2	1	5	1	1	6	10	5	6	4	8
2007	4	6	605	1	3	4	10	4	4	4	2	8	7	7	7	1	9
2007	4	6	606	6	1	9	5	3	3	6	9	8	4	10	1	10	8
2007	4	6	607	2	3	3	4	8	4	10	9	5	4	2	7	8	5
2007	4	6	608	8	8	10	9	6	1	10	5	8	9	10	6	5	3
2007	4	6	609	9	1	8	3	5	1	8	2	2	6	8	3	9	4
2007	4	6	610	3	9	10	5	6	4	9	10	1	5	6	8	2	3
2007	4	6	611	10	5	1	4	8	1	10	4	9	9	5	8	3	7
2007	4	6	612	5	10	8	10	9	4	6	8	6	6	9	4	9	7
2007	4	6	613	8	3	9	6	7	8	8	4	5	4	8	8	3	8
2007	4	6	614	9	4	6	9	4	9	10	5	8	8	2	3	4	6
2007	4	6	615	3	5	6	10	9	8	10	2	3	3	9	1	4	4
2007	4	6	616	1	2	1	5	10	8	10	3	1	2	3	6	5	3
2007	4	6	617	4	3	8	9	10	2	5	3	10	7	7	9	1	8
2007	4	6	618	8	4	3	1	4	8	9	2	8	1	2	4	10	10
2007	4	6	619	3	1	6	9	7	2	6	10	2	7	6	6	1	1
2007	4	6	620	8	10	10	2	8	7	9	7	1	2	2	7	3	6

Figure 8 Example of randomly generated count of aid dataset

Data_Year	Data_Month	PROV_34_ID	DIST_34_ID	101	102	103	104	105
2007	4	5	505	54886.6882429897	10187.2591942538	34295.2709724864	19386.9813923387	65047.8738110898
2007	4	5	506	16938.3917086691	53837.3528251107	88536.9131929032	26885.2766000899	83355.4755340308
2007	4	5	507	75117.2353805206	274.306618634201	43049.3437430338	14458.8629751316	79164.3074987999
2007	4	6	601	87208.446997516	1686.10448010854	13267.54362365	89710.1178529992	56657.1735126449
2007	4	6	602	86429.8076345214	62080.4428477063	7841.02519115087	28727.880534789	62164.6883131137
2007	4	6	603	59934.7465075057	14883.9204198354	26872.2089930395	16322.0697173053	9813.70303864942
2007	4	6	604	80930.7877417557	81347.5540065132	77603.6290990947	14992.0147752652	7887.56983403963
2007	4	6	605	4475.42678726927	23525.0572393871	30553.8556541417	99516.7849763994	33954.7850379341
2007	4	6	606	53019.6284949117	5363.45260649989	80555.2667434238	44849.0676781478	23542.6400709933
2007	4	6	607	11167.1508395756	26968.6734780633	23608.5109778509	38854.2901152005	70073.7108954264
2007	4	6	608	71861.3135044469	79942.3671538164	99364.5087302875	81545.8788726168	56534.3646135004
2007	4	6	609	82778.1655183796	1025.77452007332	72804.3985898153	22287.8890373412	45347.2388161955
2007	4	6	610	25577.0902091019	88923.0512772992	96912.3241168914	44027.7972765337	51164.5434919554
2007	4	6	611	98754.0806413129	48848.6271183311	7918.06240942331	34329.6266855861	77873.555771457
2007	4	6	612	44900.4006977673	93554.5323319388	71851.7381108199	94263.4066648258	89036.247568781
2007	4	6	613	77713.4936833466	22866.9911607056	80116.9507359944	56801.5457674169	67083.192167469
2007	4	6	614	87580.7456873097	37380.9428534125	52336.6069659338	83540.8305646197	31681.7211140947
2007	4	6	615	21720.8581731333	49786.3106066997	53856.5912000906	91432.0697296718	86070.4224661857
2007	4	6	616	5916.24767380463	18969.8841270603	3473.5593349086	43647.6910666924	91747.4581620437

Figure 9 Example of randomly generated sum of aid dataset in US Dollars

Table 4 Structure of generated aid datasets

Column Name	Content	Column Count	Dataset Name	
			Sum of Aid Dataset	Count of Aid Dataset
Data_Year	Year of the data	1	Year (yyyy)	Year (yyyy)
Data_Month	Month of the data	1	Month(mm)	Month(mm)
Prov_34_ID	Province ID of the data	1	Province ID	Province ID
Dist_34_ID	District ID of the data	1	District ID	District ID
101 102 . . . 1406	Sector name and activity type. 101-Advisory services in Agriculture 102-Construction in agriculture . 1406-Service Delivery in Water & Sanitation	84	Sum of amount of ongoing aid(in Us Dollars)	Count of ongoing aid(in Us Dollars)

3.4   Adverse Event Data

WITSGEO dataset was selected to study reported adverse events and incidents. However other datasets have more data points for specific time periods, WITSGEO has the largest spread of data over (2005 – 2010) time period thus offers the most complete data among others files. This dataset contains adverse event date, latitude and longitude information of the event location, and number of people affected from the adverse event in terms of killed, wounded, and hijacked.

Although this dataset had the most complete data among other files, we noticed missing content in province and district ID's. Using ArcMap overlay analysis tool, all adverse event locations were accurately updated with the 34 province and 400 district ID numbers along with the count of adverse event in terms of killed, wounded and hijacked, for each month from (2005-2010) and each district.

Generated adverse event dataset is formed by 8 columns and 43200 rows. Columns contents are year, month, province ID, district ID, number of people affected in terms of killed,

24

wounded, hijacked and count of adverse events. Rows are grouped by year from 2002 to 2010,

12 month, and 400 districts. Figure 10 shows an example of the generated adverse event dataset.

Table 5 shows the content of randomly generated adverse event dataset.

Data_Year	Data_Month	PROV_34_ID	DIST_34_ID	SKT	SWT	SHT	Count
2007	4	5	506	1	14	9	11
2007	4	5	507	19	17	1	20
2007	4	6	601	5	5	8	12
2007	4	6	602	14	19	9	18
2007	4	6	603	2	1	3	16
2007	4	6	604	10	2	16	4
2007	4	6	605	19	6	19	4
2007	4	6	606	5	15	15	14
2007	4	6	607	12	18	11	20
2007	4	6	608	8	16	16	17
2007	4	6	609	5	6	12	20
2007	4	6	610	17	5	20	1
2007	4	6	611	15	2	7	19
2007	4	6	612	1	4	17	17
2007	4	6	613	15	1	12	3
2007	4	6	614	16	20	14	10
2007	4	6	615	18	9	6	20
2007	4	6	616	5	7	16	16
2007	4	6	617	19	13	8	14
2007	4	6	618	17	10	5	8
2007	4	6	619	14	6	10	5
2007	4	6	620	9	1	14	14

Figure 10 Example of randomly generated count of adverse event dataset

Table 5 Structure of generated adverse event dataset

Column Name	Content	Column Count	Dataset Name / Adverse Event Dataset
Data_Year	Year of the data	1	Year (yyyy)
Data_Month	Month of the data	1	Month(mm)
Prov_34_ID	Province ID of the data	1	Province ID
Dist_34_ID	District ID of the data	1	District ID
SKT	Sum of people Killed	1	Number of people
SWT	Sum of people Wounded	1	Number of people
SHT	Sum of People Hijacked	1	Number of People
Count	Adverse event count	1	Amount of Adverse event

3.5 ArcGIS Shape File Processing and Neighborhood Data Analysis

ArcGIS (Geographical Information System) was used to link location information of economic and development projects with adverse event data points and to update district and province ID's of aid data with recent ones. ArcGIS was also used to generate neighboring information for both districts and provinces. Two different shape files were used, which were district level shape file (*af_admbnd_adm3_pol)* and province level shape file set *(af_admbnd_adm2_pol).* Those shape files were obtained from Humanitarian Response Community open source web portal (Humanitarian Response, 2011).

Both shape files contain most recent province and district boundaries of Afghanistan which are 34 provinces and 400 districts. Neighboring districts and provinces data, which were needed for CA model generation were calculated using the neighboring polygon tool in ArcGIS Resource Center(Center, 2011). This tool defines neighboring cell as a cell that has a common border with the center cell. Figure 11 shows an example of the 1st order neighboring information of provinces and districts respectively. The first column of both data sets contains the ID of centering district; rows named $N_{1...n}$ contain n^{th} neighbor of the centering cell.

Center	N1	N2	N3	N4	N5	N6	N7	N8	N9	N10	N11	N12	N13
3207	3206	3208	3209	3210	0	0	0	0	0	0	0	0	0
3206	3207	3216	3210	3205	3201	0	0	0	0	0	0	0	0
3216	3206	3205	3102	3107	0	0	0	0	0	0	0	0	0
3208	3207	3209	0	0	0	0	0	0	0	0	0	0	0
3209	3207	3208	3210	3203	3211	0	0	0	0	0	0	0	0
3210	3207	3206	3209	3203	3201	0	0	0	0	0	0	0	0
3203	3209	3210	3205	3204	3201	3202	3211	0	0	0	0	0	0
3205	3206	3216	3203	3204	3212	3102	3101	3103	3201	3202	0	0	0
3204	3203	3205	3212	3213	3211	0	0	0	0	0	0	0	0
3212	3205	3204	3213	3103	2003	3215	0	0	0	0	0	0	0
3213	3204	3212	3214	3304	2003	2004	3211	0	0	0	0	0	0
3307	3308	3306	3214	3211	0	0	0	0	0	0	0	0	0
3308	3307	3306	3310	3311	3301	0	0	0	0	0	0	0	0
3306	3307	3308	3214	3311	3301	3305	0	0	0	0	0	0	0
3214	3213	3307	3306	3305	3304	3211	0	0	0	0	0	0	0

Figure 11 Example of district level neighborhood dataset.

3.6 **Population Data**

The only provided dataset that includes population information is AISCS data file which

is a part of USAID-IRD-AISICS dataset. This file includes the population of Afghanistan in

2008. This population file consist only 399 districts of total 400 Afghanistan districts. The 400th

district is generated by using the ratios calculated by using forecasted population file. ASICS

data file contains all demographic information for Afghanistan. For this research 5 of these

demographic types are used. They are; total population, male population, female population,

urban population, urban male population, urban female population, rural population, rural male

population, rural female population and population density for each district.

CHAPTER FOUR: METHODOLOGY

4.1   <u>Methodology Outline</u>

This section summarizes the cellular automata implementation for HSCB program. This research is conducted in five phases, visualized in Figure 12. In phase one structure of data and associations between socio-economic factors and adverse events are investigated. In this phase traditional statistical and computational methods are applied to the data however it is observed that these techniques were not sufficient enough to make confirming predictions.

 In phase two possible methods that can serve to the purpose of this research are investigated. This investigation led this research to be conducted using cellular automata modeling approach. Then literature is reviewed focusing on the application of cellular automata modeling approach.

In phase three a provided data is compiled and cellular automata model is developed and in phase four the model is tested. Since the performance of cellular automata model is based on generation of cellular automata rules, definition of states and neighborhood, various models are developed and tested until the conforming results are obtained.

In phase five Results of the model is investigated. In this phase capabilities of generating what-if scenarios and making forecasts for the upcoming years are added to the model.
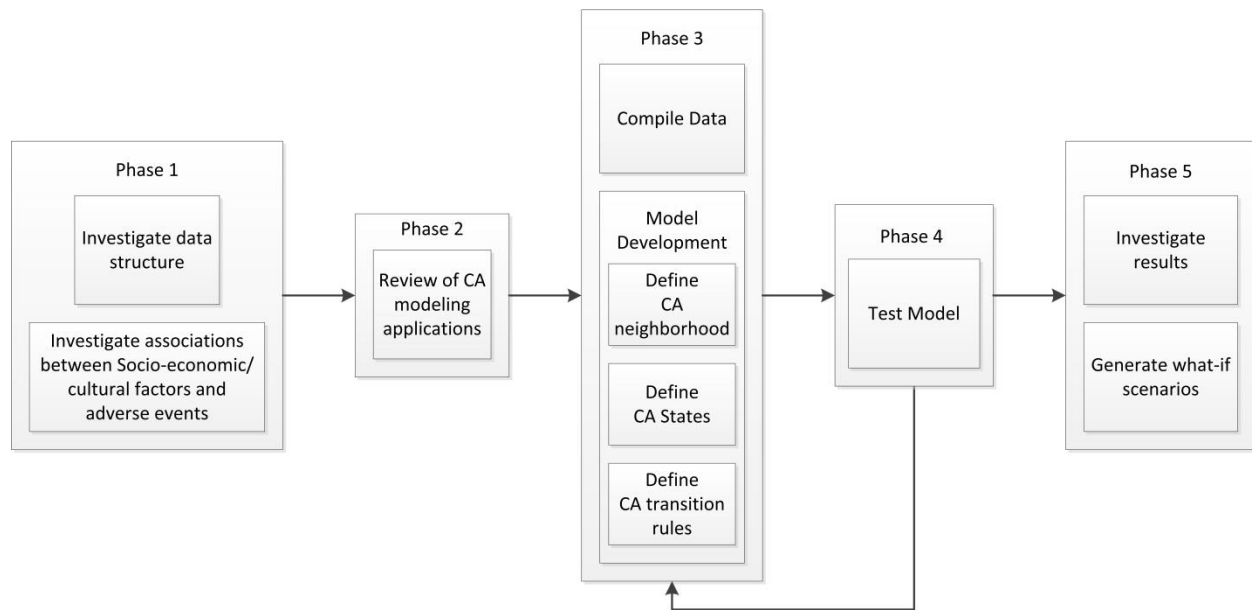
Figure 12 Research methodology

The use of CA has several significant contributions to this research for modeling the

adverse events. This research will be conducted on spatial data, and CA is a computational

model of spatial interactions. Many researchers' (Calvo Garzón, Laakso, & Gomila, 2008;

Chialvo, 2008) approach individual behavior as the final result of simpler elements interacting

through defined rules, leading to the emergence of new and unexpected states. CA consists of

identical computing cells. Cells in the same neighborhood synchronously interact with each other

to determine their next state, and small changes in one iteration yield complex formations after

several iterations. CA could be simply controlled by the rules to enable conforming results.

According to Deadman, (Deadman, Brown, & Gimblett, 1993) GIS techniques have limitations

in modeling changes in the landscape. Pairing GIS with CA can function as an analytical engine

to deliver flexible framework for developing dynamic spatial models so that generated geospatial

data directly links to the spatial structure of cellular automata. Figure 13 illustrates the overall

29

framework of the system. GIS acts as a data integrator in this methodology. A map of

Afghanistan can be displayed using GIS software which acts as a platform to organize and

generate geospatial data.
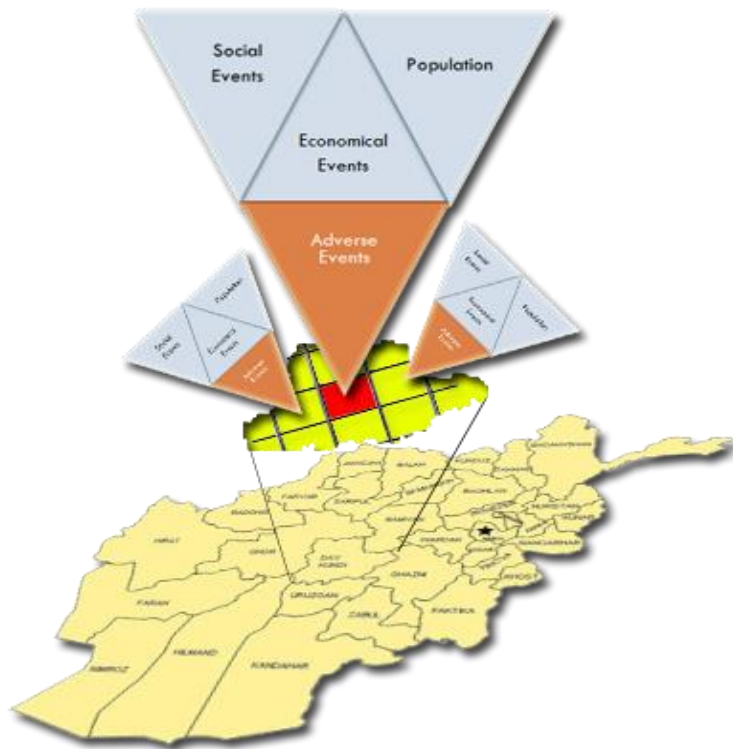
Figure 13 Cellular automata methodology outline

4.2   Cellular Automata Methodology

In the simplest form of cellular automata, the underlying network structure is represented

by a tessellation. As it is described earlier in this report, the tessellation is a uniform bi-

dimensional collection of cells, where cells represent the population involved. These rules

represent parametric set of conditions which determines the next state of cells according to its

neighbor's states, in discrete time "t". Iteration on the state of the cell is done based on a set of

transition rules. The change in behavior of the CA could be simply controlled by transition rules to enable conforming results.

The modeling methodology of CA for social and economic analysis for this research is designed in two major implementation levels, which are macro and micro-level (see Figure 14). In the macro-level, the modeling framework integrates population, social, and economic sub-systems. The macro-level will allow the model to use regionalized representations, and understand why the events have occurred. Macro-level will support cellular automata rules to generate accurate predictions. Prediction capability of CA will be used to model the micro-level interactions between individual actors which are represented by adverse events. Understanding and managing these systems as coherent dynamic entities will help to maintain system integrity. In this methodology, CA iterates in time to generate complex patterns on the overall model by iterating with small changes.
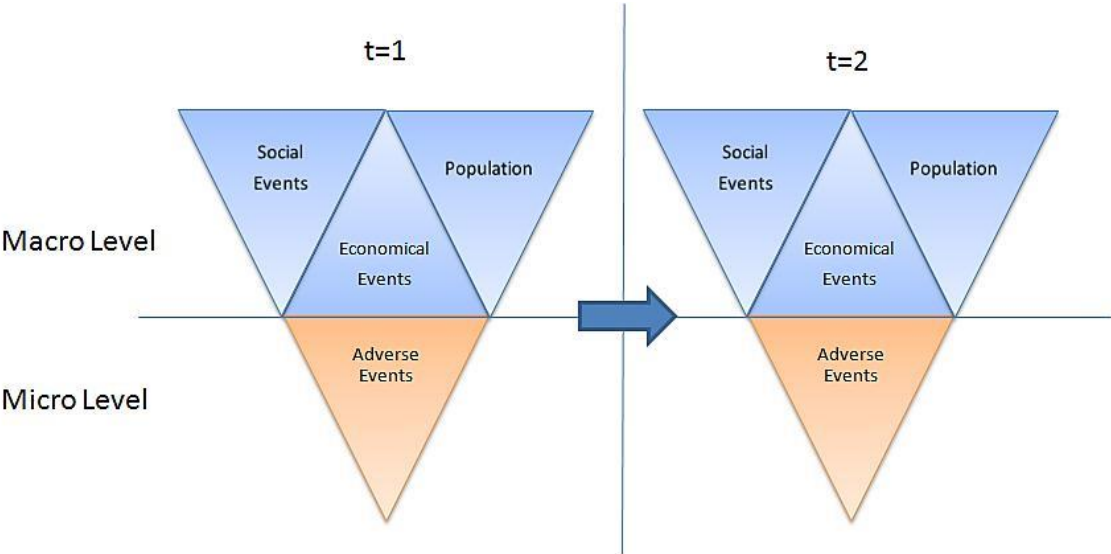
Figure 14 Cellular automata methodology

The CA modeling is performed through two modules, which are transition rule generation and cellular automata prediction. The rule generation module measures, analyzes and modifies the data, and generates rules that will be used on cellular automata prediction. At the level of cellular automata prediction, the model applies cellular automata rules to the predictor data to generate accurate results.

Figure 15 illustrates the modeling framework and input variables of the CA model. A transition rule generator inputs aid data as the count and the sum of the budget of the aid grouped in aid sector and activity types; adverse event the data as count and the impact (population affected by the adverse event) grouped by number of killed, wounded, hijacked people, and the count of adverse events happened; and population data in terms of rural and urban population, female and male population, total population and population density.

In this framework, GIS acts as both location data integrator and neighboring cell identifier using the cellular automata variables. Cellular automata prediction module inputs data for only the time period that prediction will be conducted. Input data is the same in type as the input of transition rule generator module except for adverse event data. Prediction module inputs adverse event data which belongs to prediction start time at (t-1) to apply transition rules and generate prediction results. Next, we describe CA neighborhood presentation, cell states and transition rules performed in this study.

If a user has defined an aid data distribution, the CA module overrides the user specified distribution on the current state of the cell. A CA module also gets landscape data and population specific data from the GIS module to add into the current cell state of the cell. If it is the initial run of the model, the CA module also gets adverse event initial data from the user. Then, the CA

loops in the CA module to create next states of the cells until the defined time span is reached. And then it sends generated data to GIS to visualize and calculate the statistics of the results.
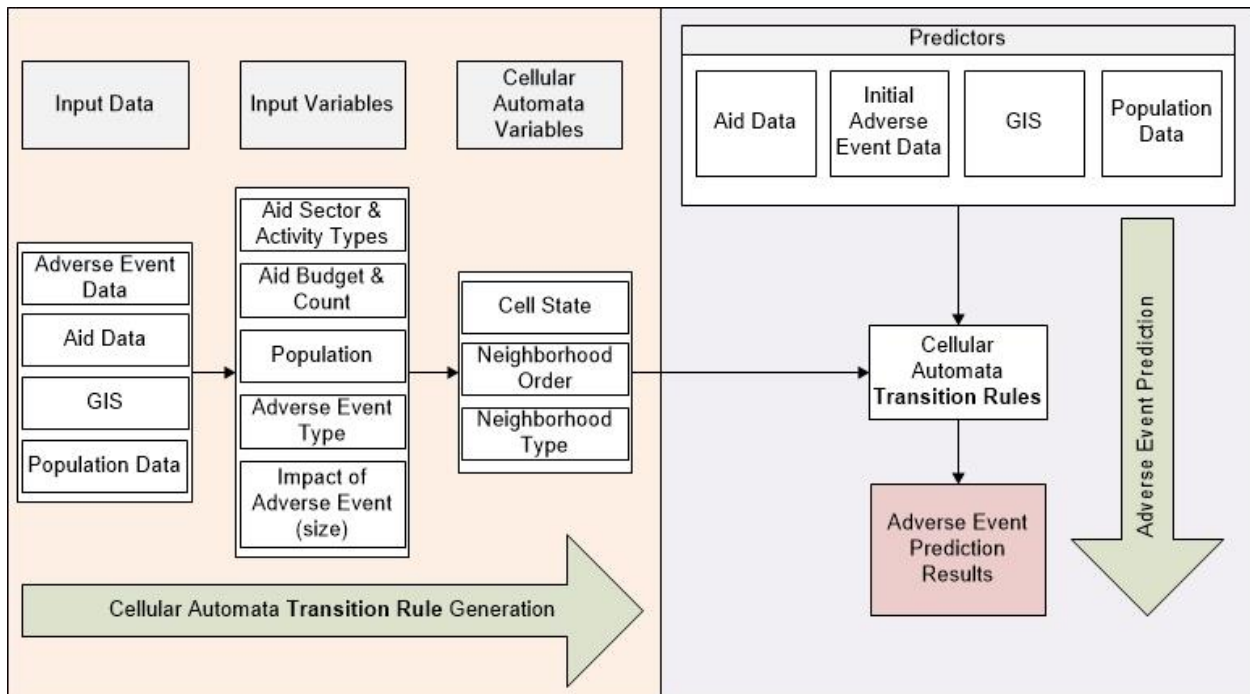

Figure 15 Cellular automata framework

4.2.1   Neighborhood

Practically, the most common social science applications of cellular automata use a rectangular grid as the underlying network structure (Flache & Hegselmann, 2001), such as Moore or Von Neumann Neighborhoods. These neighborhoods have regular grid structures which have the same number of neighbors. The research data features districts as the lowest level of location. Afghanistan districts have a minimum of 1 and a maximum of 12 neighbors. Therefore, using a regular grid as a location is not applicable for this research. However irregular

grid structures are also possible in CA (Hegselmann & Flache, 1998). For this research, an irregular cellular automata neighborhood is generated by using ArcGIS geostatistical analyst tool. Figure 16 represents a Voronoi diagram generated to visually represent neighborhood $\mathcal{N}_d$. It is defined as a set of any 1st order neighbor that has a common border with the center district d. Each cell in the Voronoi diagram represents one of the 400 districts in Afghanistan. To improve visual analyses, size and location of the Voronoi cells are kept similar to the Afghanistan political map.

Figure 16 Visualization of irregular cellular automata neighborhood for Afghanistan districts

4.2.2   State of Cells

State of a cell $\varsigma_d(t)$ is defined as the row vector of size 178 that represents the state of a district d at time t. 168 elements of $\varsigma_d(t)$ is the sum of US Dollars spent $\delta_d(t)$ and the count of aid $\gamma_d(t)$ for each sector and activity type of aid data, 5 elements of it are the number of population in terms of total population, male and female, urban area and rural area populations $\varpi_d(t)$, 1 element of it is population density $\eta_d(t)$ and 4 elements of it are adverse

events measures $\alpha_d^a(t)$. This model involves $a = k, w, h, c$ adverse event measures, which are the number of people killed, the number of people wounded, the number of people hijacked, and the count of adverse events respectively.

$$\varsigma_d(t) = [\delta_d(t) \quad \gamma_d(t) \quad \varpi_d(t) \quad \eta_d(t) \quad \alpha_d^a(t)] \tag{6}$$

4.2.3 Transition Rule

Transition rule ϕ describes the rules of change for each cell from its current state at time t to its next state at time $t + 1$. Figure 17 visualizes the dataflow of cellular automata prediction. Transition rule ϕ inputs predictor transition state $\tau_d^*(t)$, state and adverse event ratio matrices **T** and **R**. In this section first, generation of required inputs are described and then transition rule is introduced.
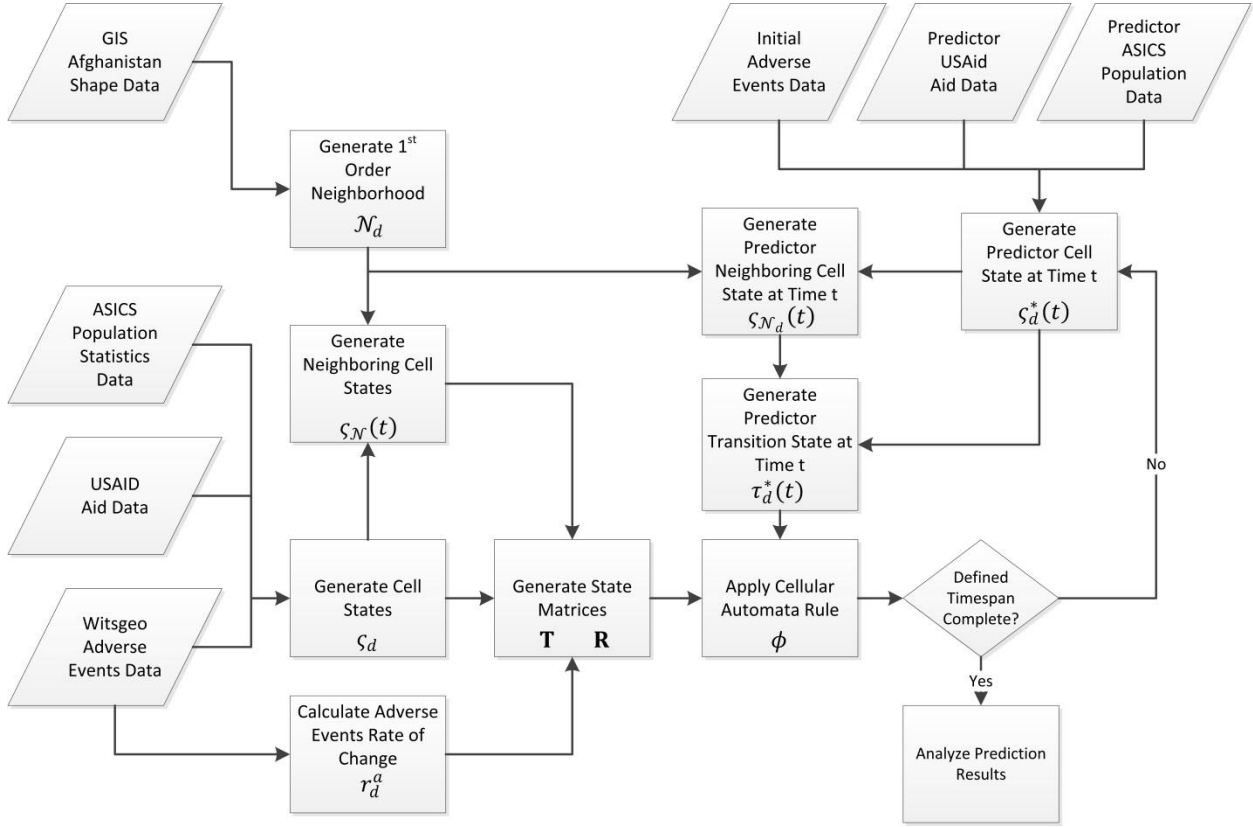
Figure 17 Cellular automata data flow

4.2.3.1 Transition State

State of the neighboring cells $\varsigma_{\mathcal{N}_d}(t)$ is defined as the summation of each neighbor's state for the centering cell d except for the cell state of population density. States of the neighboring cells are defined as;

$$\varsigma_{\mathcal{N}_d}(t) = \sum_{d \in \mathcal{N}_d} [\delta_d(t)\, \gamma_d(t)\, \varpi_d(t)\; \alpha_d^a(t)] \tag{7}$$

Transition state $\tau_d(t)$ is the row vector of size 358 that represents macro and micro-level states of district d at time t. 2 elements of $\tau_d(t)$ is $\lambda_d(t)$ which are year, month of the data

sampled, 356 elements of it are the state of the center cell $\varsigma_d(t)$ and the state of the neighboring

cells $\varsigma_{\mathcal{N}_d}(t)$ respectively.

$$\tau_d(t) = \begin{bmatrix} \lambda_d(t) & \varsigma_d(t) & \varsigma_{\mathcal{N}_d}(t) \end{bmatrix} \tag{8}$$

4.2.3.2 Adverse Event Change Ratio

Micro-level changes in adverse events are defined as the ratio of change from time t

to $t + 1$. It is defined as the ratio of change on adverse event measures $r_d^a(t)$ for all

measures $a = k, w, h, c$ for district d at time t.

$$r_d^a(t) = \begin{cases} \alpha_d^a(t+1) & \text{if, } \alpha_d^a(t) = 0 \\ \dfrac{\alpha_d^a(t+1)}{\alpha_d^a(t)} & \text{otherwise} \end{cases} \tag{9}$$

$\mathbf{T}$ is the state matrix which includes $\tau_{d'}(t')$ for all training data where d' is the district

and t' is the time of the training data. $\mathbf{R}$ is the matrix that includes adverse event rates for each

$\tau_{d'}(t')$ in $\mathbf{T}$. $\mathbf{T}$ represents macro and micro-level states that supports cellular automata rules to

generate predictions with the use of corresponding $\mathbf{R}$.

$$\mathbf{T} = \begin{bmatrix} \lambda_{d'}(t') & \varsigma_{d'}(t') & \varsigma_{\mathcal{N}_{d'}}(t') \\ . & . & . \\ . & . & . \end{bmatrix} \qquad \mathbf{R} = \begin{bmatrix} r_{d'}^k(t') & r_{d'}^w(t') & r_{d'}^h(t') & r_{d'}^c(t') \\ . & . & . & . \\ . & . & . & . \end{bmatrix} \tag{10}$$

4.2.3.3 Weighted Euclidean Distance

Let ω be the row vector of size 356 that represents the weight of impact of the each

variable on the adverse events at time $t + 1$. 172 elements of ω are the weight of impact of sum

of US Dollars spent and count of aid at centering cell $\omega_{\delta\gamma}$, 5 elements of it are the weight of

impact of population of centering cell ω_ϖ, 4 elements of it are the weight of impact of number of

killed, wounded hijacked and total event count $\omega_k, \omega_w, \omega_h, \omega_c$, 172 elements of it are the weight

of impact of sum of US Dollars spent and count of aid at neighboring cells cell $\omega_{\delta\gamma_\mathcal{N}}$, 4

elements of it are the weight of impact of population of neighboring cells $\omega_{\varpi_\mathcal{N}}$, 4 elements of it

are the weight of impact of number of killed, wounded hijacked and the total event count at

neighboring cells $\omega_{k_\mathcal{N}}, \omega_{w_\mathcal{N}}, \omega_{h_\mathcal{N}}, \omega_{c_\mathcal{N}}$.

$$\omega = \begin{bmatrix} \omega_{\delta\gamma} & \omega_{\varpi} & \omega_k & \omega_w & \omega_h & \omega_c & \omega_{\delta\gamma_\mathcal{N}} & \omega_{\varpi_\mathcal{N}} & \omega_{k_\mathcal{N}} & \omega_{w_\mathcal{N}} & \omega_{h_\mathcal{N}} & \omega_{c_\mathcal{N}} \end{bmatrix} \quad (11)$$

Let $\tau_d^*(t)$ be the transition state of the predictor cell for district d at time t. This will be

used to find conforming $\tau_{d'}(t')$ among matrix **T** to gather corresponding $r_{d'}^a(t')$. The decision

making process of selecting the conforming $\tau_{d'}(t')$, and determining the ratio of change can be

described in two steps. The first step is calculating Weighted Euclidean distances e_j in between

$\tau_d^*(t)$ and each row of **T**. Where σ_T is the standard deviation of each column of **T**.

$$E(e_j) = \sqrt{\sum_{i=1}^{\ell} \frac{\omega^2}{\sigma_T^2} \left(T_{di} - \tau_{di}^*(t) \right)^2} \quad (12)$$

4.2.3.4 Prioritization of Minimum Distances

The second step is prioritizing minimum Euclidean distances according to the district,

province and Islamic calendar. Figure 18 is the flowchart of prioritization of the data with

minimum Euclidean distances. First, minimum Euclidean distances e_j are prioritized if $\tau_{d'}(t')$

has same district if not same province with $\tau_d^*(t)$. Second, It is expected that in certain Islamic

months the number of adverse events will happen will follow a different trend than other months

so that, if the Islamic month of transition state predictor cell $\tau_d^*(t)$ is either 9th or 12th month of

Islamic calendar which are Ramadan and Dhul-Hijaah, minimum Euclidean distances are

prioritized after applying the prioritization of district or province. Lastly, the ratio of change in

adverse events r^a is calculated for all $a = k, w, h, c$ as the mean of $r_{d'}^a(t')$ which $\tau_{d'}(t')$ has the

highest priority.
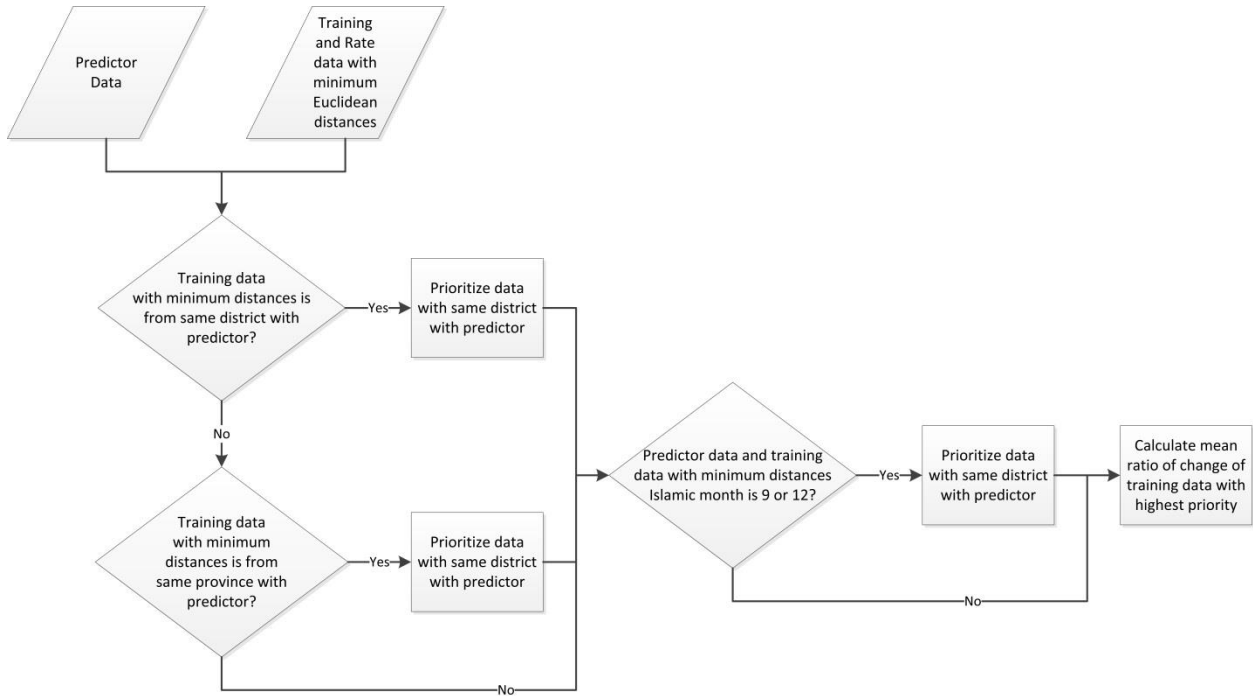
Figure 18 Prioritization of minimum Euclidean distances

4.2.3.5 Transition Rule

Transition rule ϕ is defined as the change of adverse event with the ratio of r^a.

$$\phi : \alpha_d^a(t+1) = \begin{cases} r^a & if, \alpha_d^a(t) = 0 \\ \alpha_d^a(t) * r^a & otherwise \end{cases}$$ (13)

4.3   <u>Software and Computational Tools Used In Modeling and Experimentation</u>

This section describes software and computational tools used to implement cellular automata modeling approach. These models are developed and tested on a high speed personal computer and the need for high performance computing for realizing weight of impact calculations are executed on 256 cores of STROKES Advanced Research Computing Center (ARCC).

The software platforms used for this study are; MSSQL 2010, ArcGIS 10.1, and Matlab 2011a. MSSQL is used to compile, complete and derive necessary information form USAid and adverse event databases. ArcGIS is used to generate necessary neighborhood information for cellular automata modeling. ArcGIS is also used to update location information of both aid and adverse event data and process of creation geo-spatial data. Matlab is used to implement and test cellular automata methodology. Statistical and parallel computing toolboxes in Matlab are used to increase computational time of the model.

Cellular automata simulation model is designed in functions to increase code legibility and to improve parallel computing capability. HSCB_CA is the main function for the cellular automata model. The input variables to run the model through STROKES can be inputted through this function. This function inputs weight of impact, training, simulation initial startup, and GIS neighboring datasets. This main function returns predictions for all weights for each adverse event and performance results of the run. Cellular automata function runs the simulation for given weight of impact. This function inputs training, simulation initial startup, weight of impact, and GIS neighboring datasets and returns the predictions for given weight of impact datasets. Sum of neighbor function, calculates the neighboring states of given district. This

function inputs current district, simulation time, district states and GIS neighboring datasets, and returns sum of neighboring cell states for the given district and simulation time. Rate finder function prioritizes and calculates ratio of change for the given district. This function inputs training data and minimum Euclidean distances and outputs ratio of change in adverse events. Performance measure function calculates performance measures when the simulation run is complete. It inputs prediction and real adverse event datasets. Georgian to Islamic "gtoi" function converts given Georgian month and year to Islamic month. The Compare data function is the developed GIS interface to visualize the output of the simulation model. Necessary functions for running the simulation are presented in Appendix A.

CHAPTER FIVE: RESULTS AND DISCUSSION

5.1   Introduction

This Chapter demonstrates that cellular automata model is capable of evaluating socio-economic influences that result in change in adverse events, and identifies location, time and impact of these events likely to occur. This chapter first explains calculations related to the proposed CA methodology, and then the calculation of weight of impact is explained. Secondly impact of socio-economic influences on adverse events and how these influences affect neighboring districts are discussed. Thirdly, for the scenario examined, developed cellular automata model is used to enhance the capability to understand and use human behavior by generating what-if scenarios to find out the impact of different aid sectors to predict and shape adverse events. Finally, adverse events that will occur in upcoming years are predicted to allow decision makers forecast responses to their actions.

5.2   Performance of the model

The CA model is tested against the actual Afghanistan adverse event data for 2010. This experiment is run to predict adverse events occurred in Afghanistan monthly from January 2010 to December 2010.

Table 6 shows the input variables for the model testing. Aid data, adverse event data, and the population data from January 2005 to November 2009 for 400 districts are inputted to train the model. The aid data from December 2009 to November 2010, the adverse event data of December 2009 for 400 districts and the population data are inputted as the predictor data.

Table 6 Input variables for testing the model

	Aid Data	Population	Adverse Event Data
Transition state data	January 2005 - November 2009	2008	December 2005 - November 2009
Predictor data	December 2008 - November 2009	2008	December 2008

Performance of the model is measured by two performance measures. The first one is the average model performance and the second one is the accuracy. The measure of average model performance involved two performance measures, mean absolute error (MAE) and root mean square error (RMSE). MAE is calculated by summing the absolute values of the errors and then dividing by the number of predictions. RMSE is calculated by summing the squared values of the errors and dividing by the number of predictions and then taking the square root of the mean square error.

$$MAE = \frac{\sum_{i=1}^{n}|e_i|}{n} \qquad\qquad (14)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}|e_i|^2}{n}} \qquad\qquad (15)$$

The final model has shown the highest average model performance among the variations of models that are generated and analyzed, such as (1) model using standardized Euclidean distances and, (2) model using weighted Euclidean distances and.

Table 7 shows the average model performance of two variations of the model. The highest increase on the performance is observed on RMSE of the number of people wounded. High values on RMSE point large errors on predictions. It can be concluded that implanting the weight of impact to the model has increased the average model performance.

Table 7 Average model performance

Model Variation	Performance Measure	Killed	Wounded	Hijacked
1	MAE	0.92	1.8	0.35
	RMSE	3.81	29.32	1.98
2	MAE	0.75	1.33	0.28
	RMSE	3.07	6.88	1.7

The accuracy of this model is measured by the percentage of 2010 adverse event data that is predicted correctly in terms of location, time and impact of these events. Models with high percentage of correct predictions are preferred to those with low percentage. Three different success rates are measured. The first one is the exact success rate which is the rate of correct forecasts over all forecasts having been made. 1-away success rate is the rate of correct forecasts with the error margin of 1 over all forecasts. Third, 3-away success rate is the rate of correct forecasts with the error margin of 3 over all forecasts.

Table 8 Percentage of success rate

	Killed	Wounded	Hijacked
Exact Success (%)	83.29	81.27	93.29
1-Away (%)	88.90	85.22	95.10
3-Away (%)	93.90	91.54	97.33

Table 8 shows the percentage of successful prediction rates of CA model for 2010. The number of hijacked people has resulted in 93.29 percent exactly correct prediction. Both the number of people killed and wounded have resulted in more than 80 percent exactly correct prediction. When the error margin of success rate increases, it is observed that number of people

killed wounded and hijacked is more than 90 percent of all predictions. It is also observed that number of hijacked people is more than 97 percent with error margin of 3.
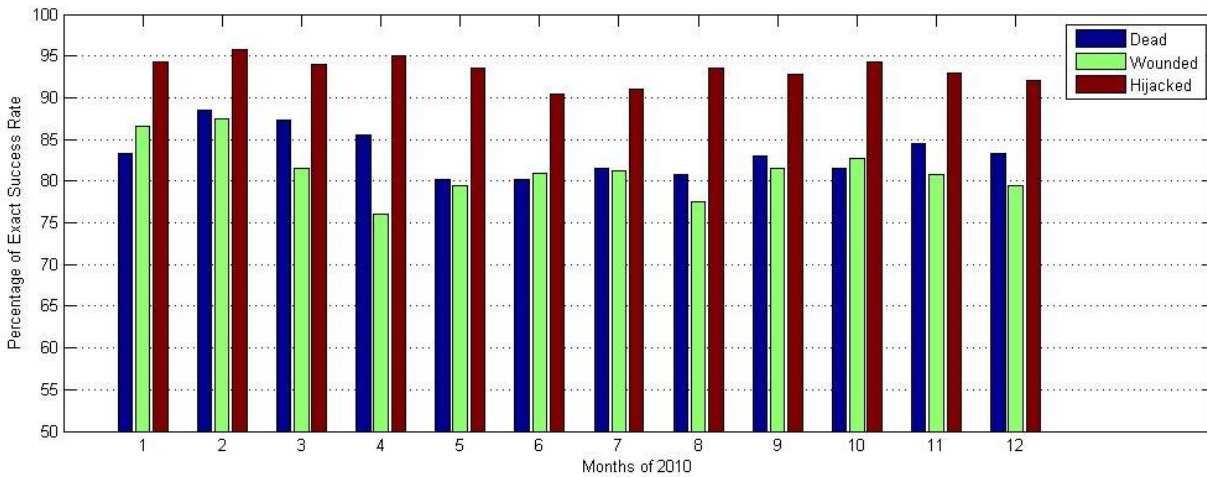
Figure 19 Percentage of exact success rate per month of 2010

Figure 19 visualizes the distribution of percentages of exact success rates for each month of 2010. It is observed that exact success rate of the model for hijacking is close to 95 percent in February, April and October of 2010. Exact success rate of killed is more than 85 percent in February, March, April and November of 2010. The prevalence of exact success rates for killed wounded and hijacked are observed with similar trend. Intuitively, with further iterations this CA model can predict the number of killed wounded and hijacked with similar accuracy.

Figure 20 visualizes the distribution of 1-away success rates for each month of 2010. It is observed that when the error margin of number of people affected form an adverse event is increased by 1; success rate of prediction of number of people hijacked is increased to over 95 percent for most of the months. The success rate of prediction of number of people killed is increased to over 85 percent for each month of 2010.
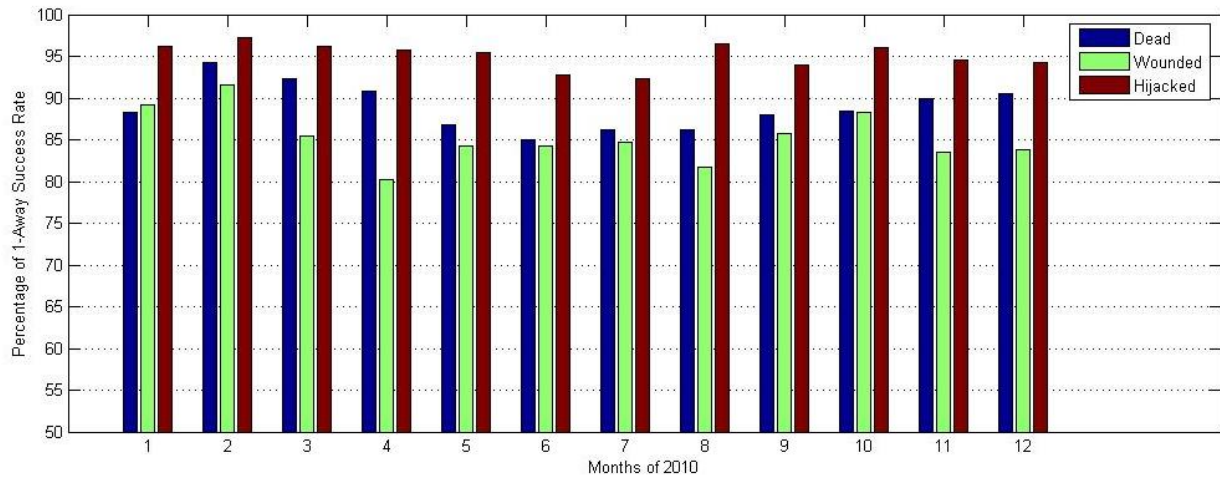
Figure 20 Percentage of 1-away success rate per month of 2010

Figure 21 visualizes the distribution of 3-away success rates for each month of 2010. It is observed that when the error margin of number of people affected form an adverse event is increased by 3; number of people hijacked is predicted by 98 percent accuracy in most of the months of 2010. The success rate of prediction of number of people killed is increased to over 90 percent for each month of 2010. Number of people killed is predicted within 95 percent accuracy for 6 months of 2010. It is also observed that number of people wounded is predicted over 90 percent correctly for 11 months of 2010. The prevalence of 3-away prediction accuracies with a similar trend, increase the confidence to the model accuracy for the further iterations to predict number of people killed wounded and hijacked.

46
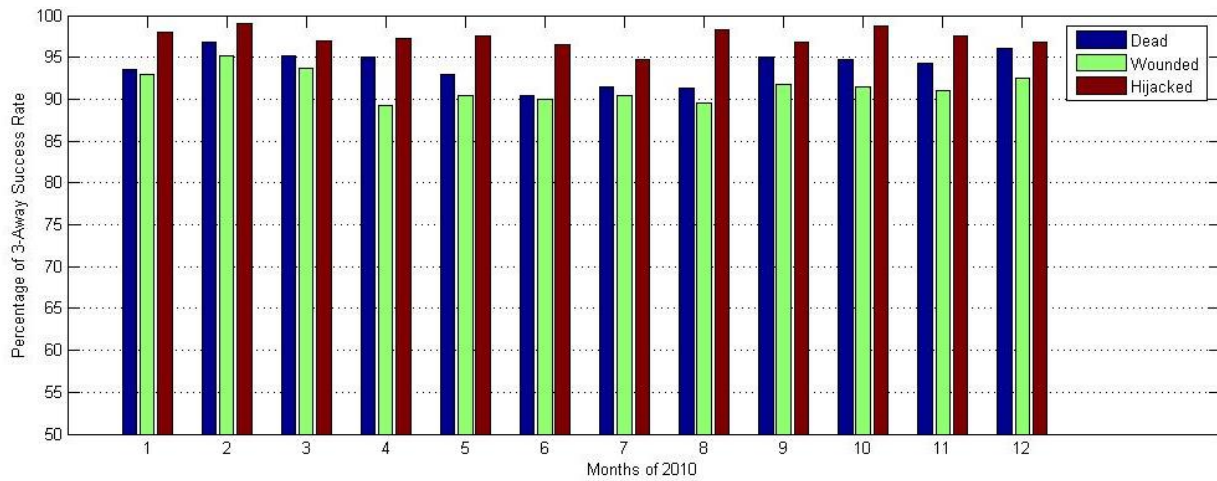
Figure 21 Percentage of 3-away success rate per month of 2010

Figure 22 compares the actual and the predicted data for April 2010 in terms of location, time, and the impact of these events likely to occur. As indicated before, the CA neighborhood for this model consists of 400 districts. Each polygon on the map refers to a district in Afghanistan. The legend for the maps which represents the range of the number of people affected for each adverse event measure.

The first set of Voronoi diagrams of the figure visualizes the actual and the predicted killed people data. For April 2010 More than 85 percent of the killed people are predicted correctly in terms of number of people killed and location. It is observed that on the districts in which people are killed but not predicted correctly by the model, are the 1st or 2nd order neighbors of the districts where killed people are predicted correctly. This also indicates that most of the districts in which people are killed but not predicted correctly by the model, are located in the provinces where killed people are predicted. The second set of Voronoi diagrams of the figure visualizes the actual and the predicted wounded people data. Similar to the killed

47

people data, most of the wrong predictions are located at the neighboring cells of the correct predictions. It is also observed that killed and wounded people data show a similar pattern of distribution among CA neighborhood. The third set of Voronoi diagrams visualizes the actual and the predicted hijacked people data which is predicted 95 percent correctly. Three of the predictions have shown that model is capable of predicting the path, possible location and the impact of the adverse events. Comparisons of the actual and predicted data for each month of 2010 are presented in Appendix B.

Figure 23 to 34 visualize adverse event predictions for the each month of 2010. On the upper left corner of the figure a Voronoi diagram visualizes geographical districts where people were killed. The upper right corner of the figure a Voronoi diagram visualizes districts where people were wounded. The lower left corner of the figure a Voronoi diagram visualizes districts where people were hijacked. And on the lower right corner of the figures bar chart and table visualize exact, 1-away, and 3-away success rates of the predictions. Voronoi diagrams represented the impact of the adverse events with colors of white, green, orange, and red. Districts displayed in white visualize predictions which are resulted with no adverse events. Districts displayed in yellow indicate predictions are resulted with in between 0 and 4 people affected from the adverse event. Districts displayed in orange indicate predictions are resulted with in between 4 and 8 people affected from the adverse event. And districts displayed in red represent predictions that are resulted with more than 8 people affected from the adverse event.
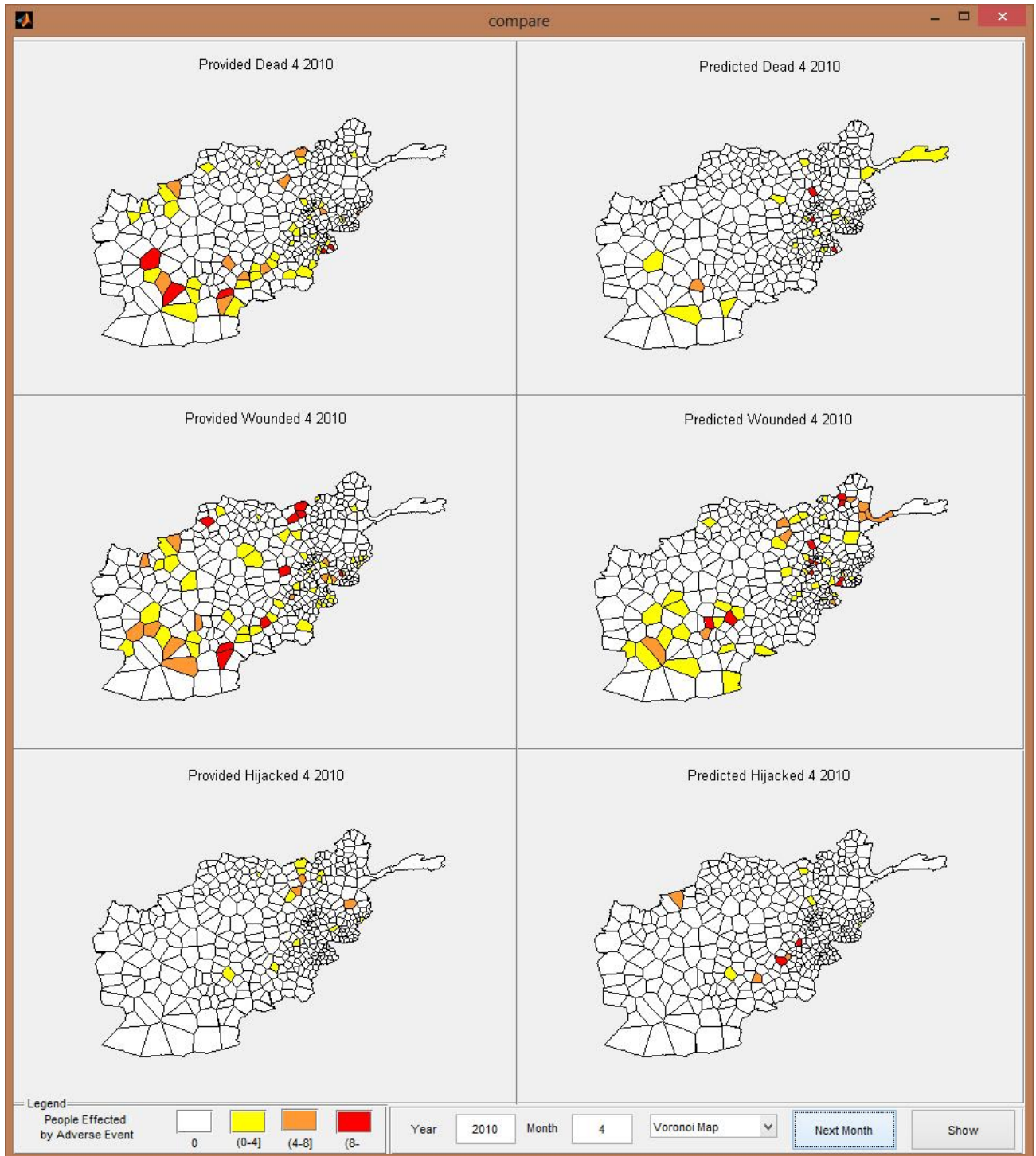
Figure 22 Illustration of actual data and predicted data for April 2010
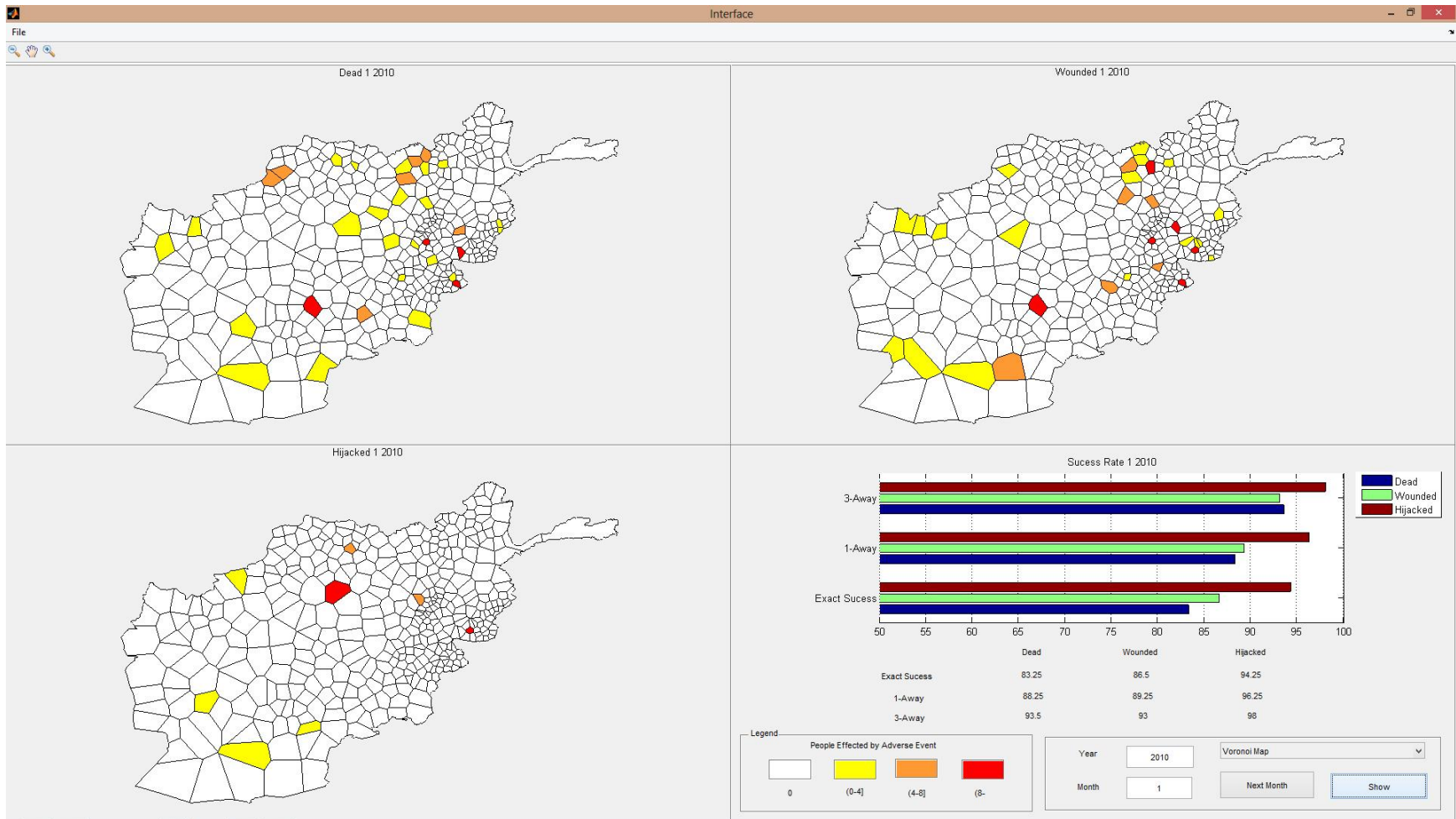
Figure 23 Adverse event predictions for January 2010
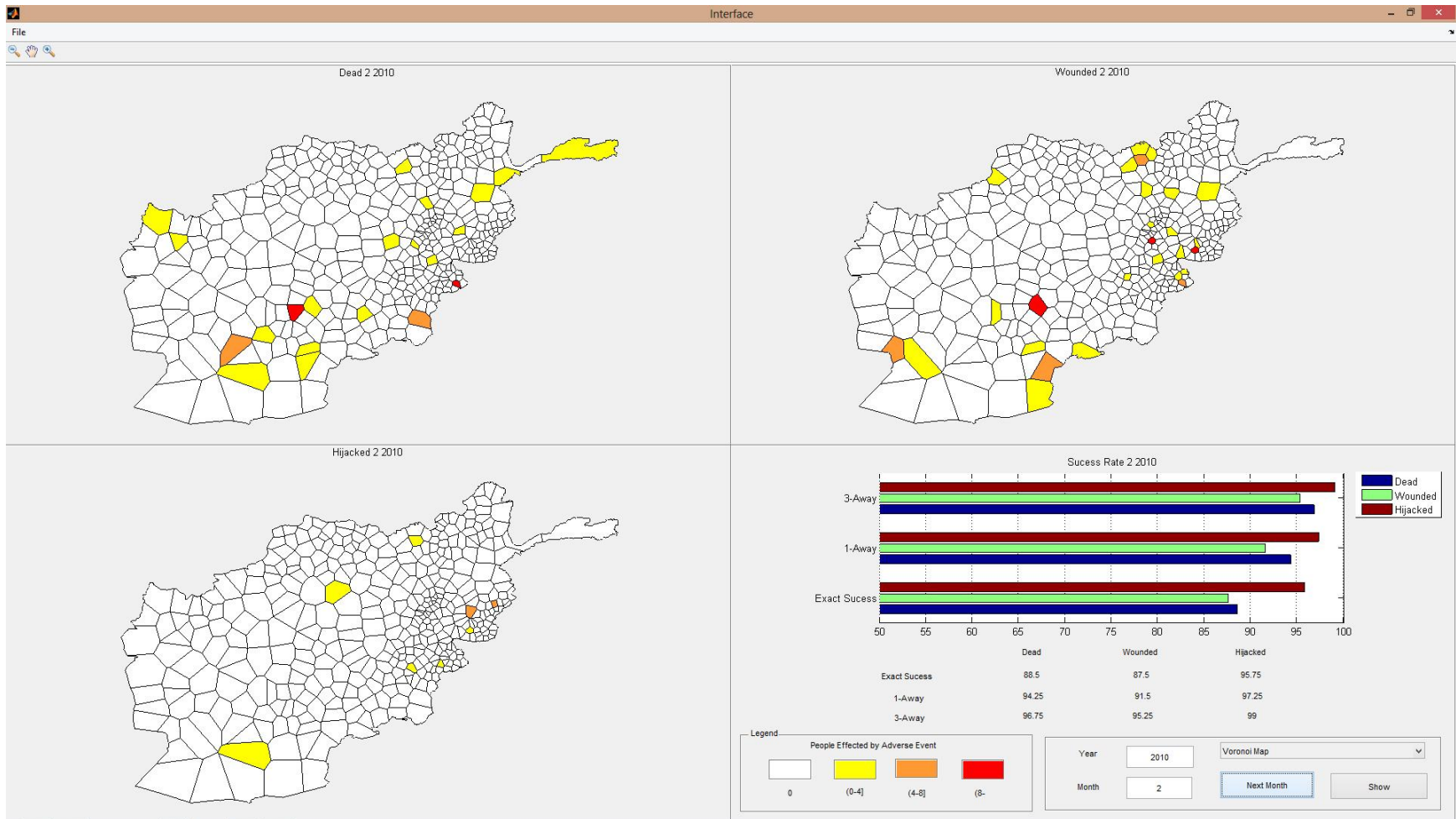
Figure 24 Adverse event predictions for February 2010
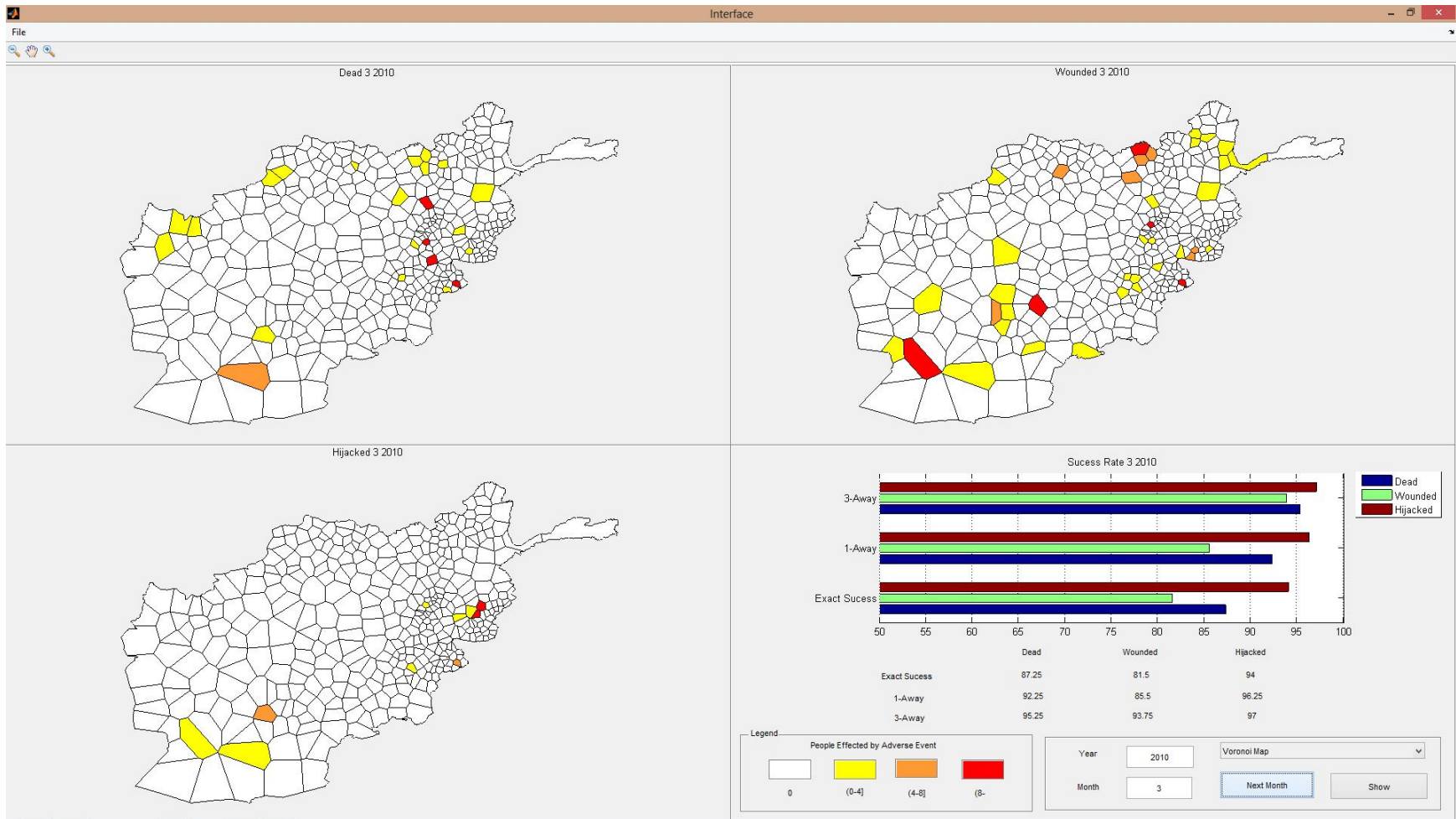
Figure 25 Adverse event predictions for March 2010
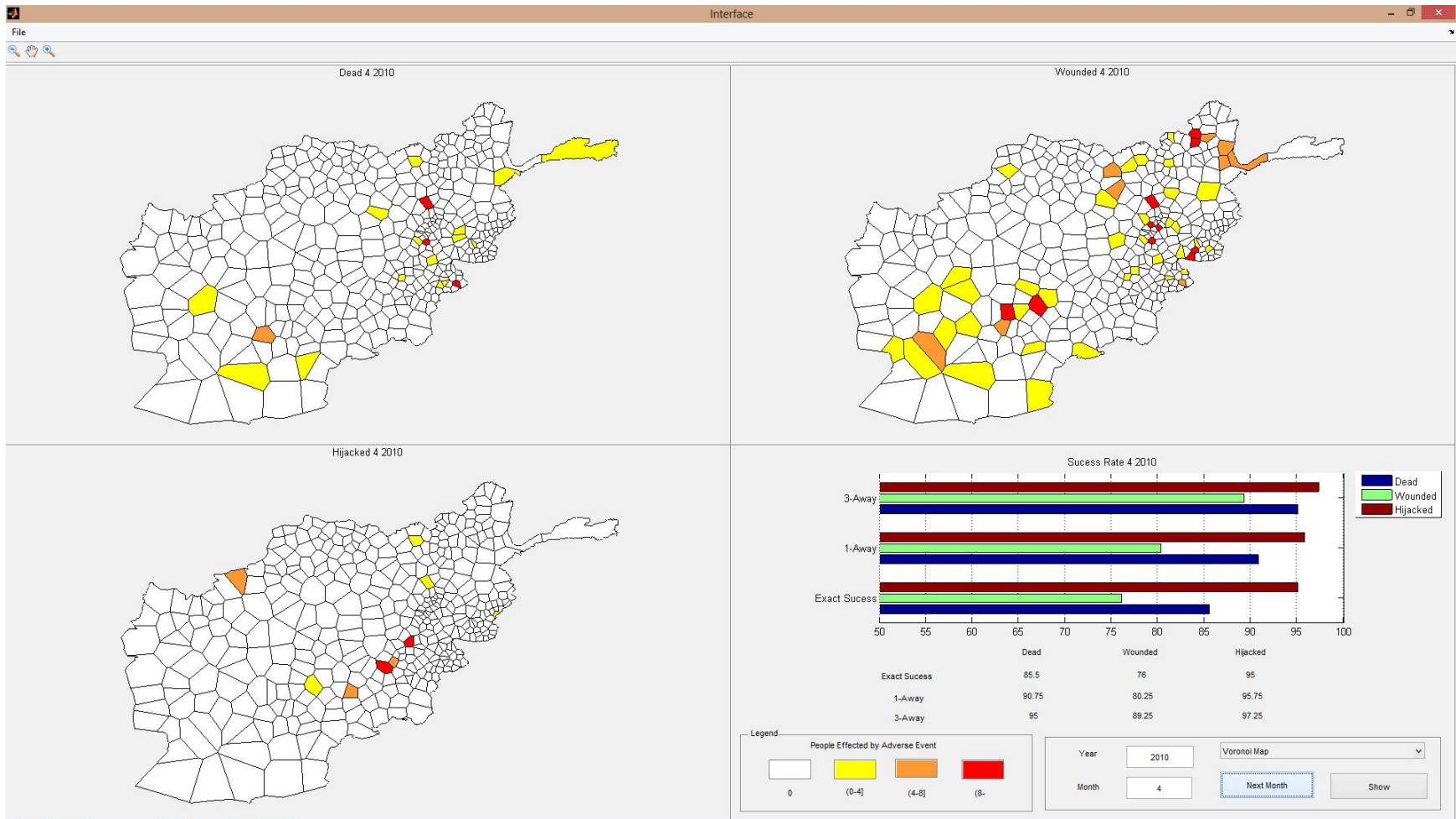
Figure 26 Adverse event predictions for April 2010
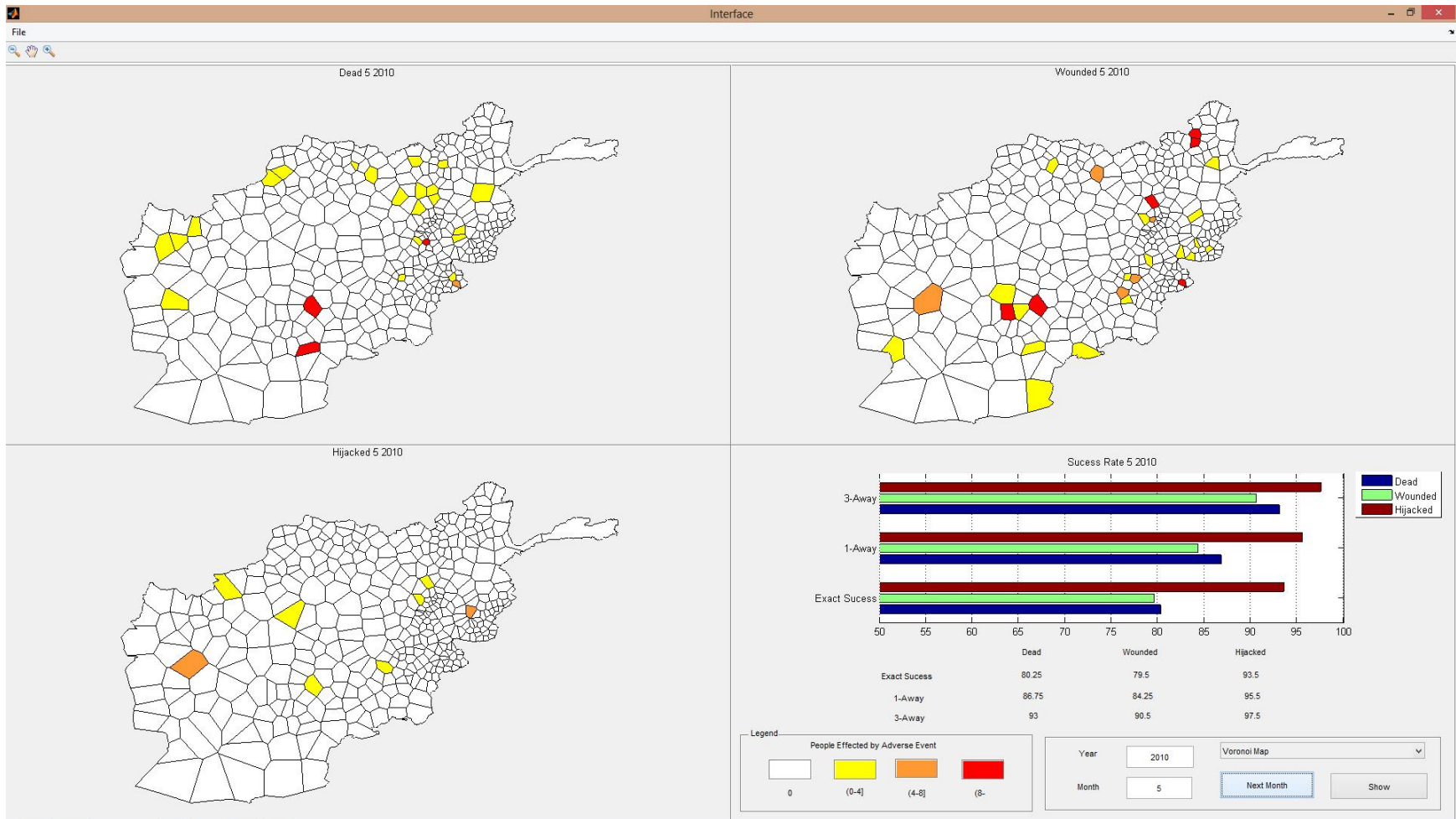
Figure 27 Adverse event predictions for May 2010
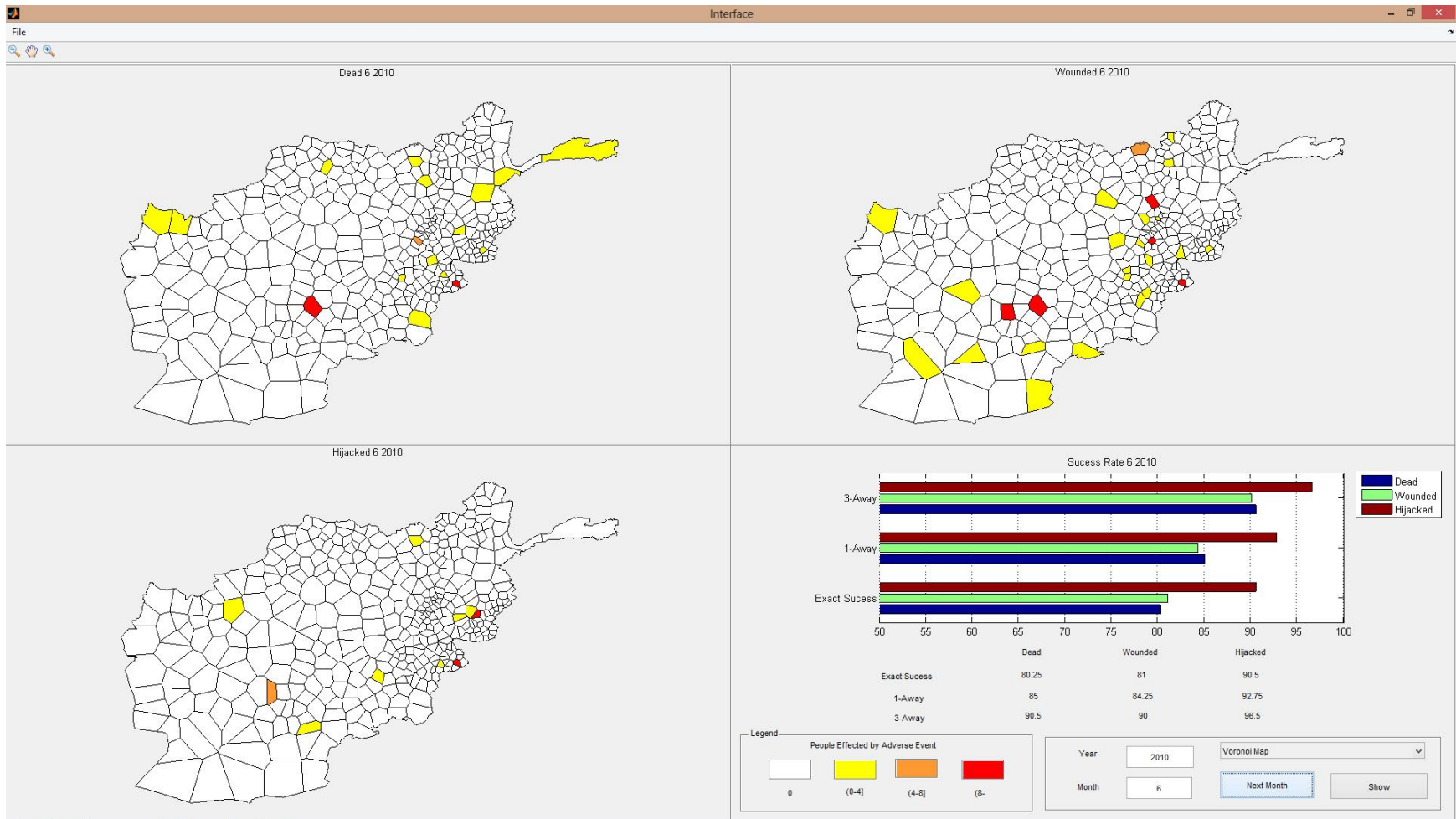
Figure 28 Adverse event predictions for June 2010
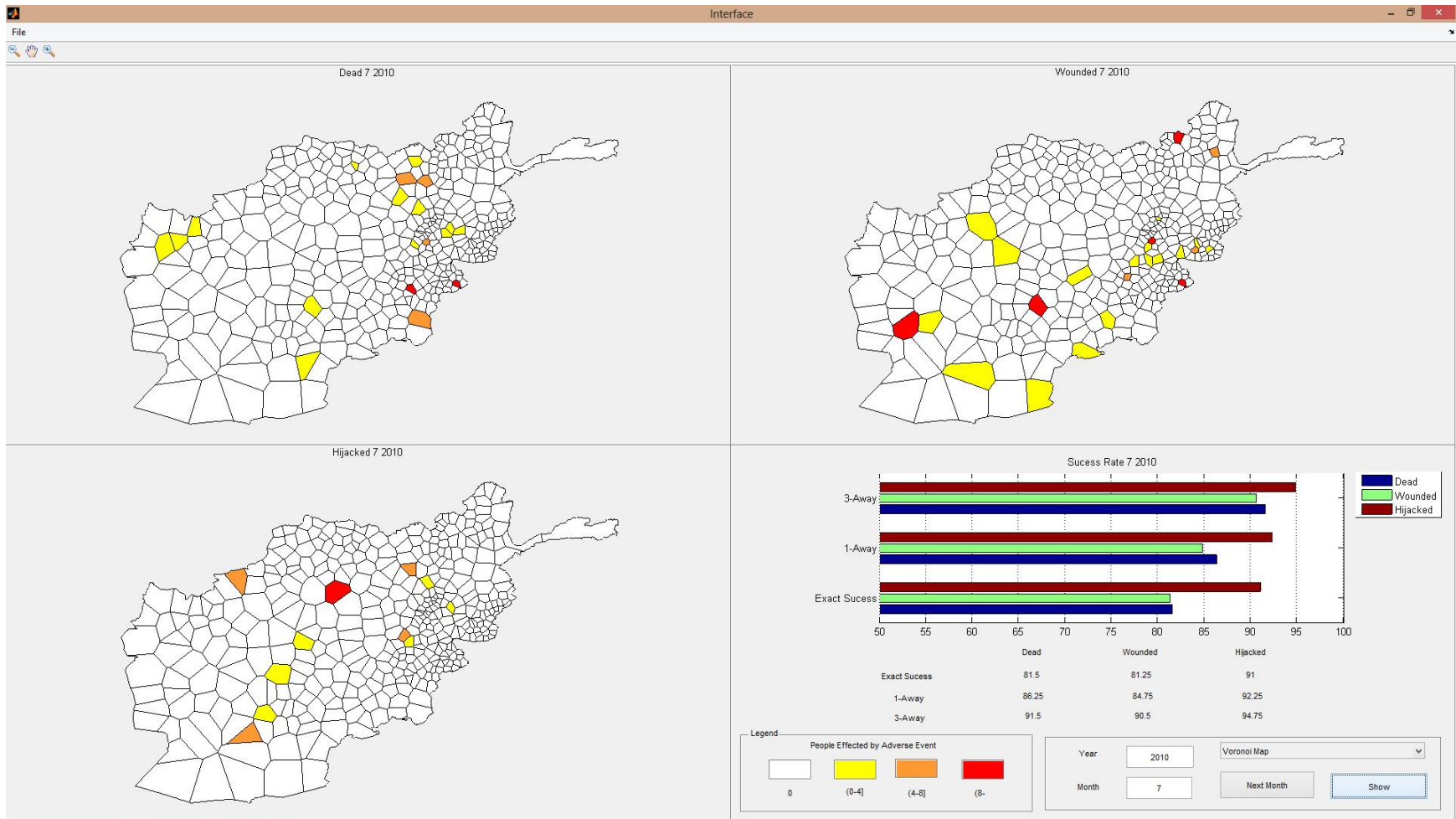
Figure 29 Adverse event predictions for July 2010
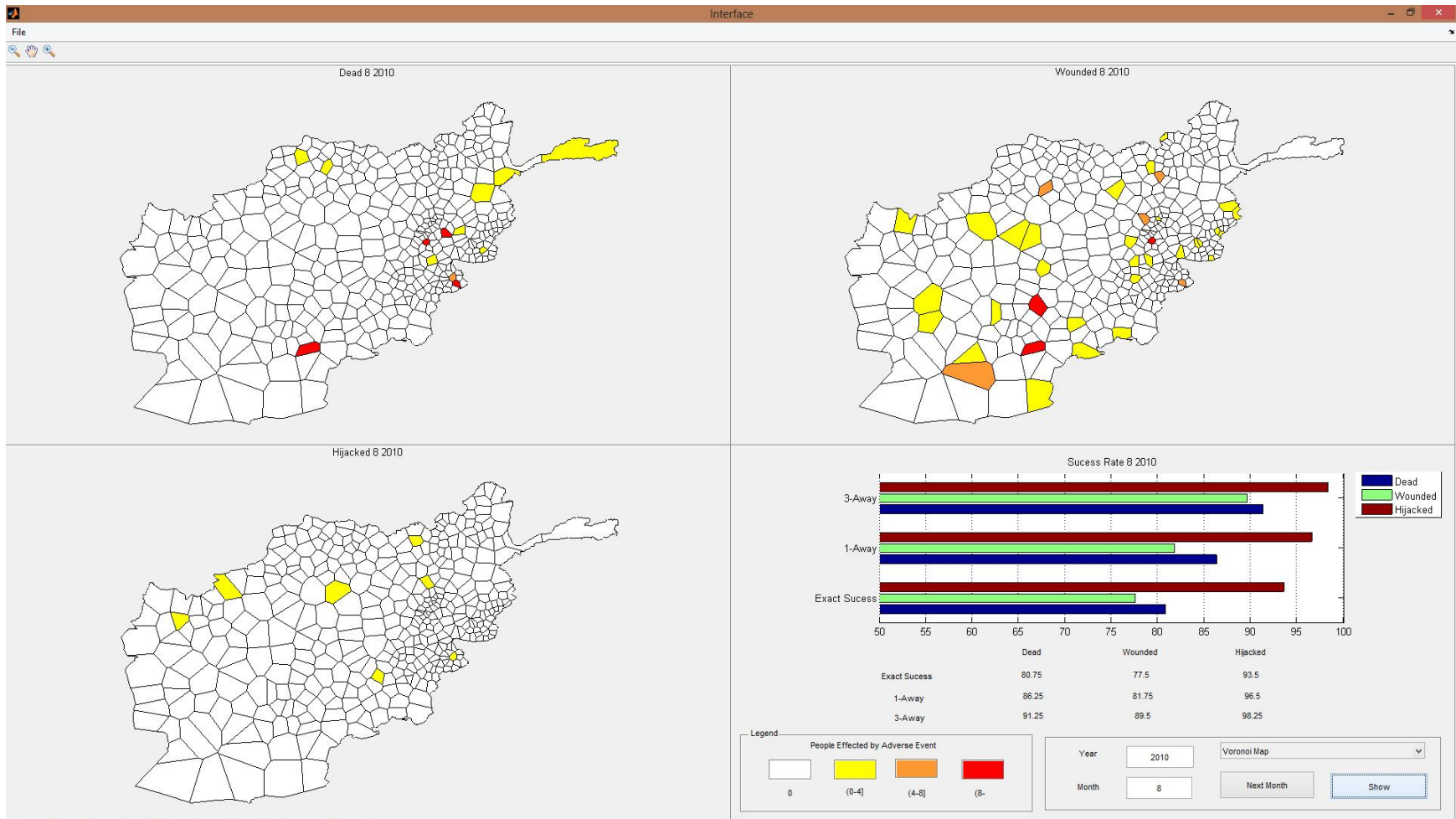
Figure 30 Adverse event predictions for August 2010
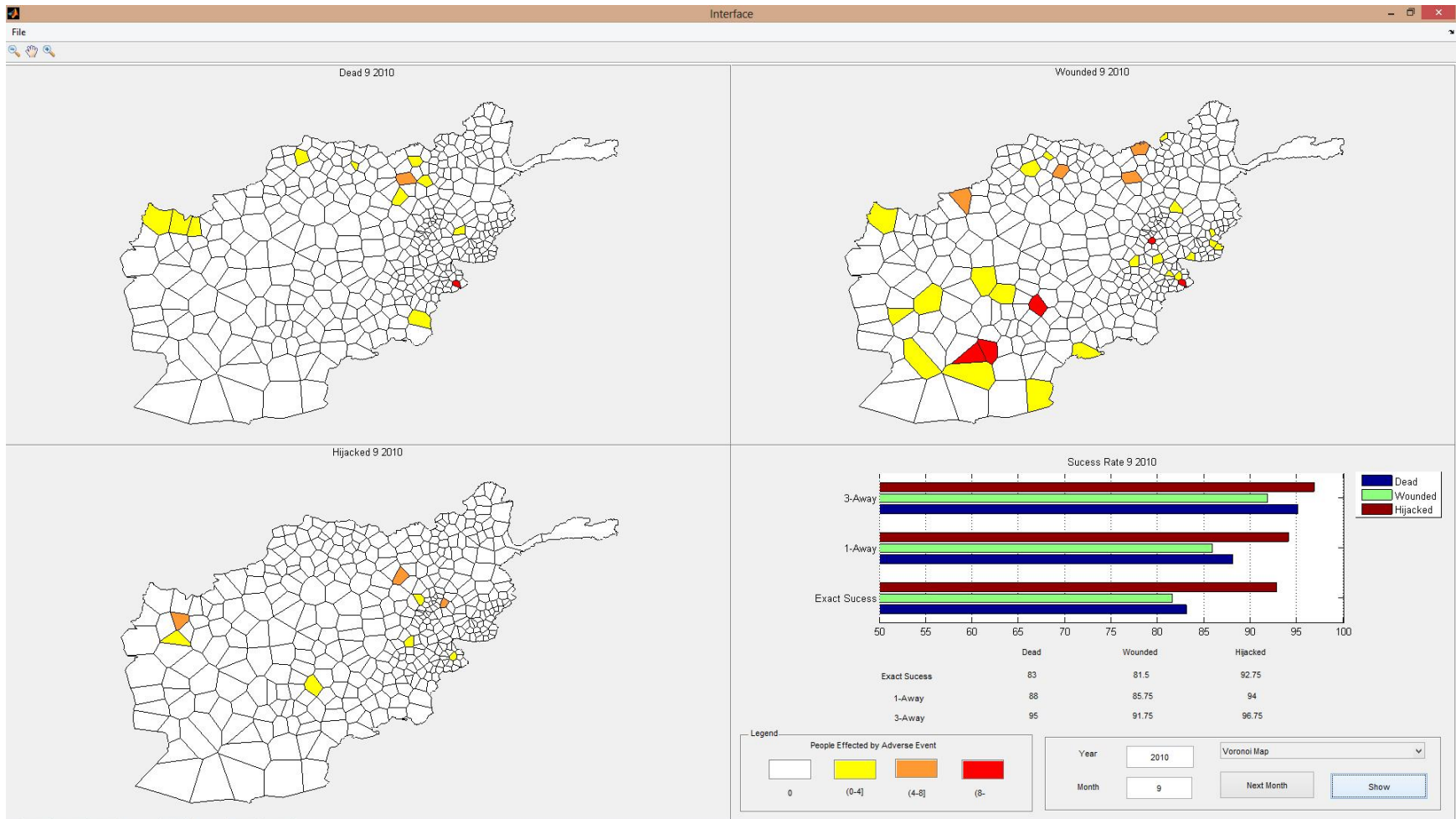
Figure 31 Adverse event predictions for September 2010
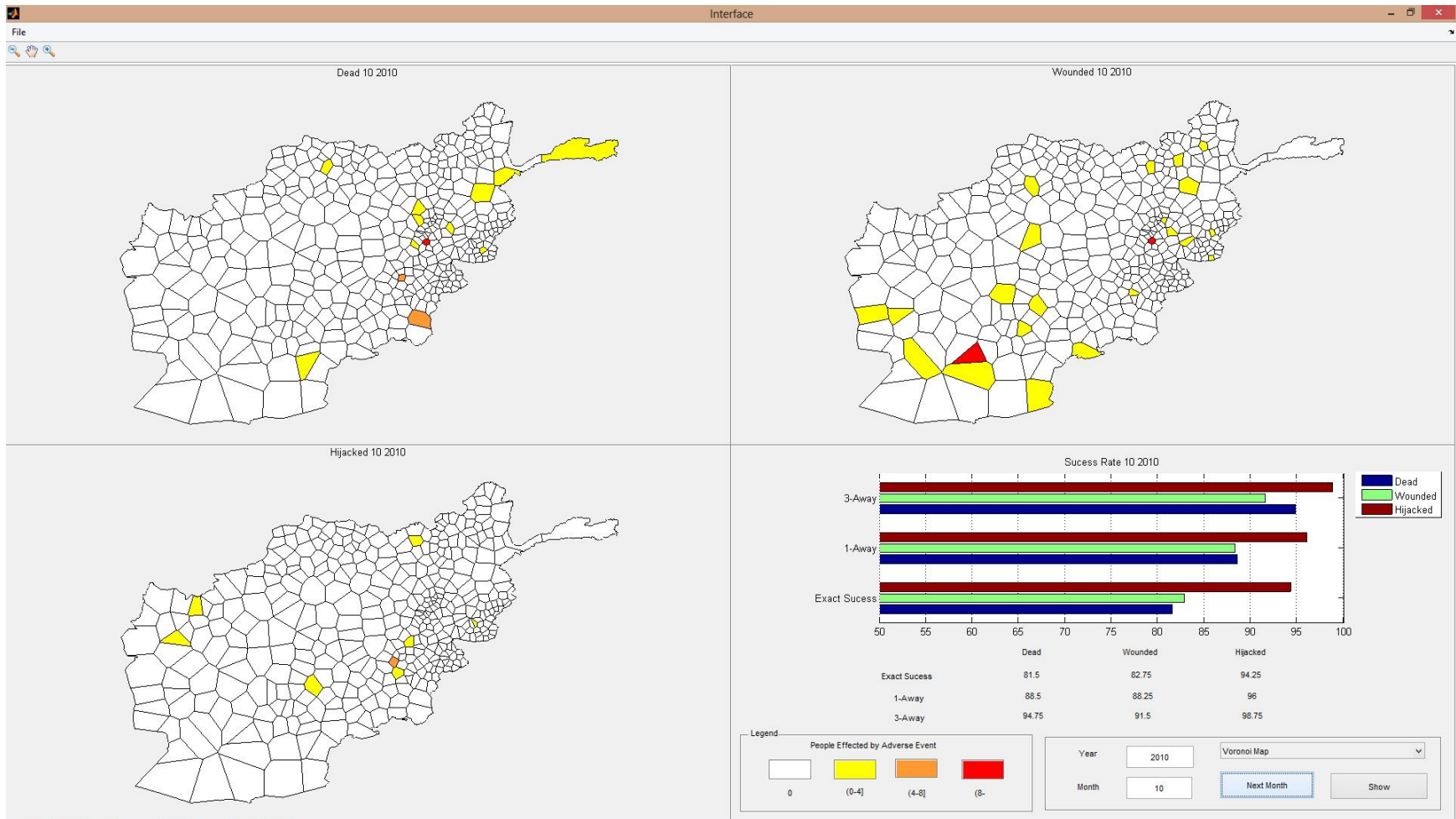
Figure 32 Adverse event predictions for October 2010
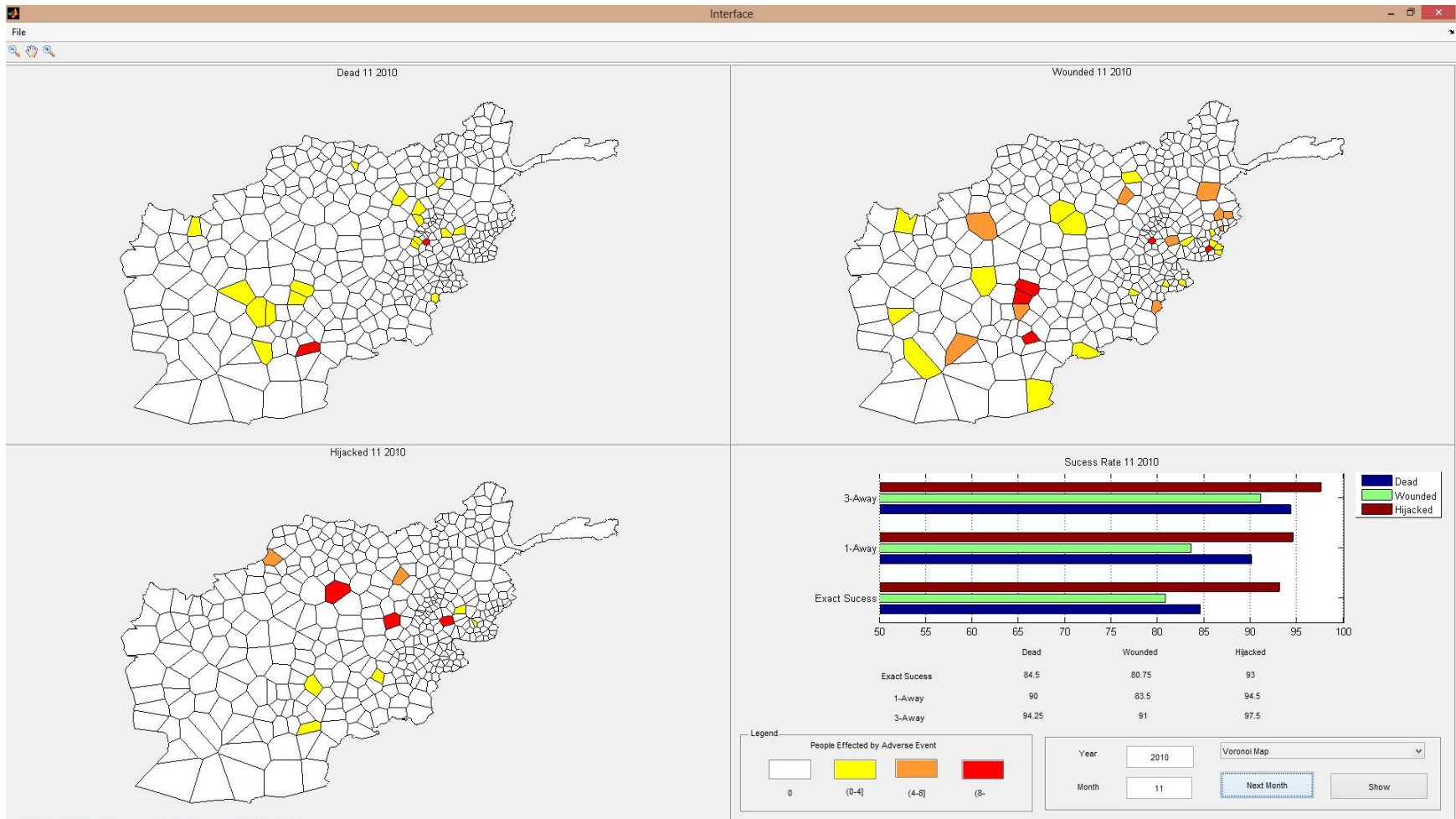
Figure 33 Adverse event predictions for November 2010
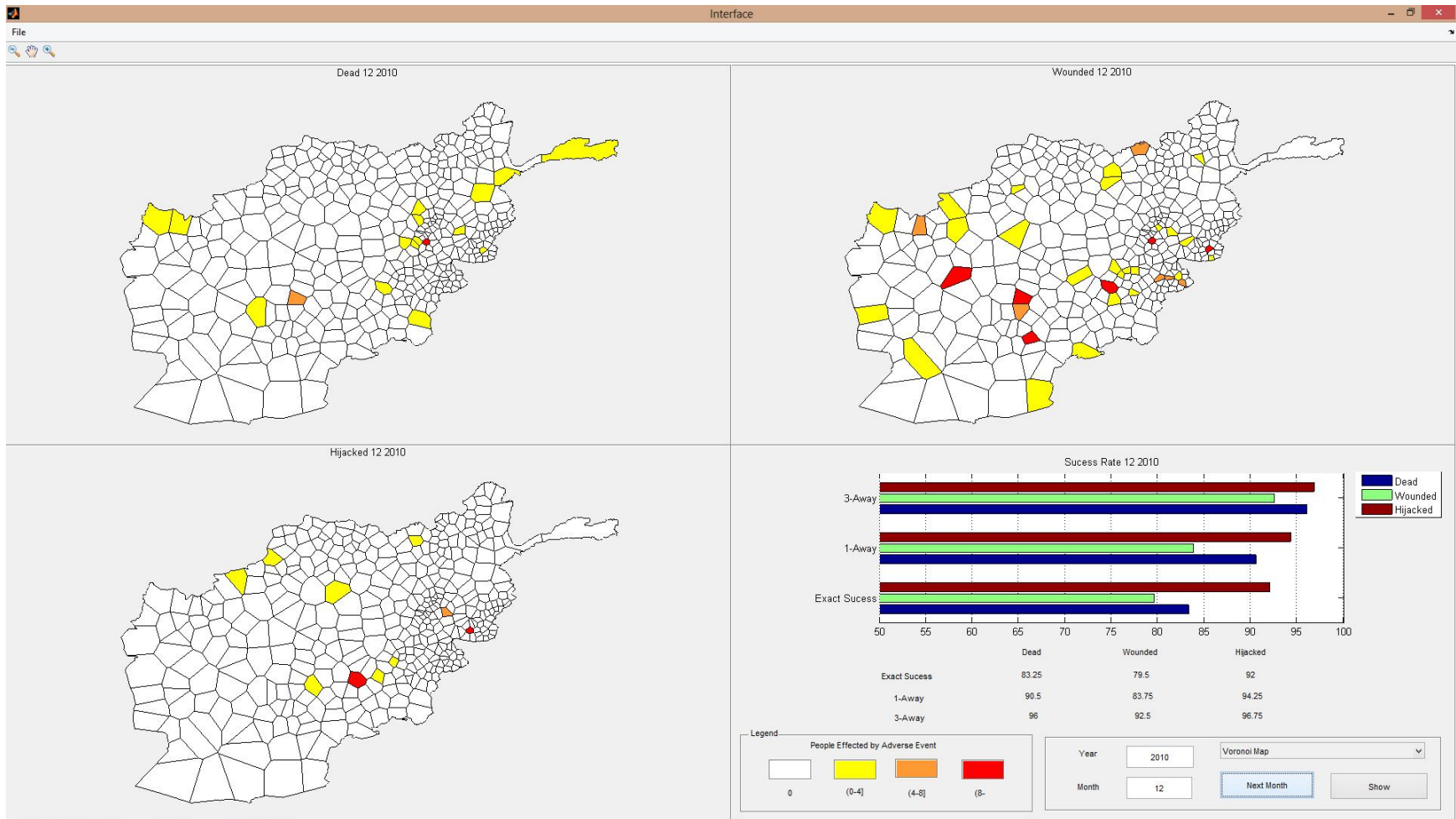
Figure 34 Adverse event predictions for December 2010

5.3   Weight of Impact Calculations

It is intended that implementing and optimizing weighing factors to the elements of transition states will improve model performance. Before implanting the weight of impact, cellular automata model assumed that each element of transition state has the same impact on the output. However, each input might have different level of impact on the output. Implementation of weight of impact to the model has provided model to consider input data with different level of importance. The weight of impact calculations is managed manually. The manual method can be defined as computation of all the inner products of weight vector with every other training and predictor vector. For this calculation, model is set for not to update predictions iteratively in time. This non-iterative process conducted to find weights that result with most accurate model output in each time step.

The need for high performance computing for realizing weight of impact calculations are executed by running the model on STROKES Advanced Research Computing Center (ARCC). A single run of the model takes between 12 to 20 minutes depending on the node specifications of STROKES. It had been intended to calculate the weight of impact for each district and for each input variable to reflect environmental drivers of each district to the model; however long computation times led this research to limit the number of elements of weighting for impact vector and the number of values for each element can take. Values of each component of the weight vector are limited in three levels. For this research total of 8 weights are calculated which is 6561 runs of the model. 6 of the 8 weights are assigned as the weights of centering cells which are limited in three levels; 0.1 0.4 and 0.9. These 6 weights represent a weighting for impact of sum of US Dollars spent and amount of aid at centering district ω_δ, population of centering

district ω_ϖ, and number of killed, wounded hijacked people and total event count ω_k, ω_w, ω_h, ω_c. Weights of neighboring cells are clustered into two groups. First group involves sum of US Dollars spent for aid and count of aid at neighboring district $\omega_{\delta_\mathcal{N}}$, and total population of neighboring district $\omega_{\varpi_\mathcal{N}}$. Second group involves total number of killed, wounded hijacked people and the total count of event happened at neighboring districts $\omega_{k_\mathcal{N}}$, $\omega_{w_\mathcal{N}}$, $\omega_{h_\mathcal{N}}$, $\omega_{c_\mathcal{N}}$. A divider is assigned to each group to calculate the weights of the neighboring district as a function of the centering district. This function is the centering cell weight over denominator. Such as if the divider for group one and two are c_1 and c_2 respectively, then $\omega_{\varpi_\mathcal{N}} = \omega_\varpi/c_1$, $\omega_{k_\mathcal{N}} = \omega_k/c_2$ or $\omega_{w_\mathcal{N}} = \omega_w/c_2$. Possible values of these 2 dividers are limited with 1, 2, and 3.

Table 9 Input variables for weight of impact calculations

	Aid Data	Population	Adverse Event Data
Transition state data	January 2005 - November 2008	2008	December 2005 - November 2008
Predictor data	December 2008 - November 2009	2008	December 2008 - November 2009

Weight of impact is calculated by testing every combination of elements of weight vectors for 2009. Table 9 shows the input variables for weight of impact calculations. Performance of each run is measured by mean absolute error. Aid data and adverse event data from January 2005 to November 2008 and population data for 400 districts are inputted to train the model. Aid data and adverse event data from December 2008 to November 2009 and population data for 400 districts are inputted as a predictor data.

Table 10 Weight of impact of socio-economic factors and adverse events occurred on adverse events that will occurs at a district

| | Impact Factor | | Adverse Event Prediction at time $t_1=t+1$ | | | |
			Killed	Wounded	Hijacked	Count
State of a Centering District at time $t_0=t$	Sum and count of aid	$\omega_{\delta\gamma}$	0.40	0.10	0.10	0.10
	Population	ω_{ϖ}	0.10	0.10	0.40	0.90
	Number of killed people	ω_k	0.90	0.10	0.10	0.10
	Number of wounded people	ω_w	0.40	0.90	0.90	0.10
	Number of hijacked people	ω_h	0.10	0.10	0.90	0.10
	Count of adverse events	ω_c	0.40	0.40	0.10	0.40

Table 10 shows weight of impact calculation results for the centering district. These weights can be interpreted as the impact of socio-economic influences and adverse events happened in a district at time t_0 on the prediction of adverse events such as number of killed, wounded and hijacked people that will happen at time t_1. In this chapter, interpretation of weight of impact is considered as high impact if values are higher than 0.65, as medium impact if values are in between 0.64 and 0.20, and remaining considered as low impact. It is observed that the number of killed, wounded and hijacked people at time t_0 has high impact on the adverse events that will happen on their own category at time t_1. Prediction of the number of people that will be hijacked at time t_1 is highly effected by the number of people hijacked at time t_0. Sum and count of aid, number of wounded people, and count of adverse events happened in a district at time t_0 has medium impact on prediction results of number of people killed at time t_1. It is also observed that population of a district has high impact on the number of adverse events that will occur at time t_1.

Table 11 Weight of impact of socio-economic factors and adverse events occurred in neighboring districts on adverse events that will occur at centering district

			Adverse Event Prediction at time $t_1=t+1$			
	Impact Factor		**Killed**	**Wounded**	**Hijacked**	**Count**
State of a neighboring districts at time $t_0=t$ (Sum of neighbors)	Sum and count of aid	$\omega_{\delta\gamma_{\mathcal{N}}}$	0.13	0.03	0.10	0.05
	Population	$\omega_{\varpi_{\mathcal{N}}}$	0.03	0.03	0.40	0.45
	Number of killed people	$\omega_{k_{\mathcal{N}}}$	0.45	0.03	0.05	0.03
	Number of wounded people	$\omega_{w_{\mathcal{N}}}$	0.20	0.30	0.45	0.03
	Number of hijacked people	$\omega_{h_{\mathcal{N}}}$	0.05	0.03	0.45	0.03
	Count of adverse events	$\omega_{c_{\mathcal{N}}}$	0.20	0.13	0.05	0.13

.

Table 11 shows weight of impact calculation results for the neighboring districts. These weights can be interpreted as the impact of socio-economic influences and adverse events happened in neighboring districts at time t_0 on the prediction of adverse events that will occur at time t_1 in the centering district. It is observed that the number of killed, wounded and hijacked people at neighbors of a district at time t_0 has medium impact on the adverse events that will happen in their own category at time t_1 in the centering district. Addition to this prediction of number of killed people at time t_1 is medially affected by the number of wounded people, sum and count of aid conducted, and the total amount of adverse events occurred in the neighboring districts at time t_0. Prediction of number of hijacked people at time t_1 is medially affected by the number of wounded people and the amount of population in the neighboring districts at time t_0.

5.4   Model Capability

This section exemplifies a what-if scenario and cellular automata model prediction for 2015 to express the capability of the model. The what-if scenario is designed to examine the change in adverse events if the amount of money spent and count of aid are doubled for the year 2010. For the scenario examined, aid distribution among districts has remained the same. Table 12 shows the USAid distribution among sectors for the real case and for the what-if scenario. The amount spent and counts of aid conducted for each sector shown in table are the average for 400 districts of Afghanistan.

Table 12 USAid distribution for the real case and what-if scenario for 2010

USAid Sector	Real Case		What-If Scenario	
	Amount Spent	Count of Aid	Amount Spent	Count of Aid
Agriculture	$220	0.01	$440	0.01
Capacity Building	$31	0.01	$62	0.01
Commerce & Industry	$10,305	0.14	$20,611	0.28
Community Development	$438	0.00	$876	0.01
Education	$705,451	9.56	$1,410,902	19.12
Emergency Assistance	$0	0.00	$0	0.00
Energy	$16,547	0.02	$33,095	0.04
Environment	$10,270	0.02	$20,539	0.05
Gender	$0	0.00	$0	0.00
Governance	$2,132,035	0.24	$4,264,071	0.48
Health	$12,064	0.03	$24,129	0.07
Security	$0	0.00	$0	0.00
Transport	$2,062,176	0.09	$4,124,351	0.17
Water & Sanitation	$20,009	0.04	$40,018	0.08
Grand Total=>	$4,969,547	10.15	$9,939,093	20.30

As the results of the designed what-if scenario, expected changes in the number of adverse events are as follows. If the amount and count of aid is doubled over the year of 2010, it is predicted that at 155 districts the number of killed people, in 147 districts the number of wounded people and in 95 districts the number of hijacked people will be decreased over the months of 2010. However it is predicted that in 57 districts number of killed people, in 76 districts number of wounded people and in 30 districts number of hijacked people will be increased. The results can be interpreted as increasing the USAid does not necessarily reduce the number of people affected by adverse events. Because it is predicted that there are districts in which the number of people affected by adverse events has increased. However, more districts are positively affected by the change in USAid.

In the cellular automata model, districts in the same neighborhood synchronously interact with each other to determine their next state, and changes in iteration in a month yield to complex formations. That is why it is expected that if the distribution of USAid is changed among different districts or months rates of changes on adverse events will be different from the scenario examined. It is also expected that if the amount of money spent for each sector is distributed among different activity types of these sectors, rates of change on adverse events will be different. Figure 35 visualize monthly average output for the real-case and what-if scenario for 2010.
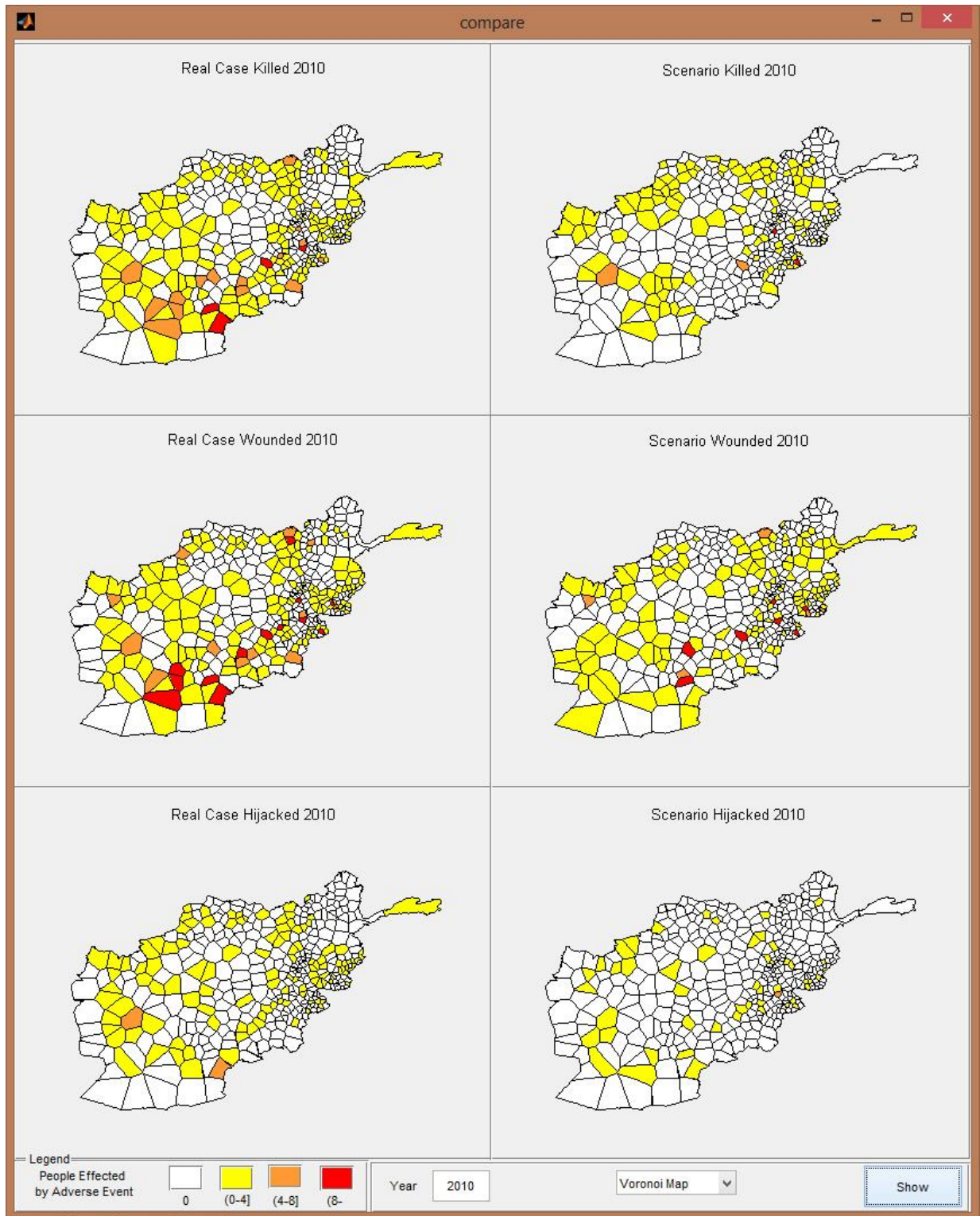
Figure 35 Real case and what-if scenario monthly average output for 2010

According to the news media, U.S. is pulling out in 2014. That is why in this research model forecasting capability is used to predict until 2014. Figure 36, 37, 38 and 39 visualize monthly average of number of people that will be affected by each category of adverse events in years 2011, 2012, 2013 and 2014 respectively. On the lower right corner of the figures bar chart visualize number of districts affected from each adverse event in the boundaries described in the legend.

It is observed among future forecasts that the average number of people affected per adverse event will be reduced. It is also predicted that the spread of adverse events will widen and effect more districts. The increase in the spread of adverse events is predicted on number of districts where people are killed and wounded. However a little or no increase is expected on the number of districts effected from hijacking. Monthly representations of the forecasts made from 2011 to 2014 are presented in Appendix C.
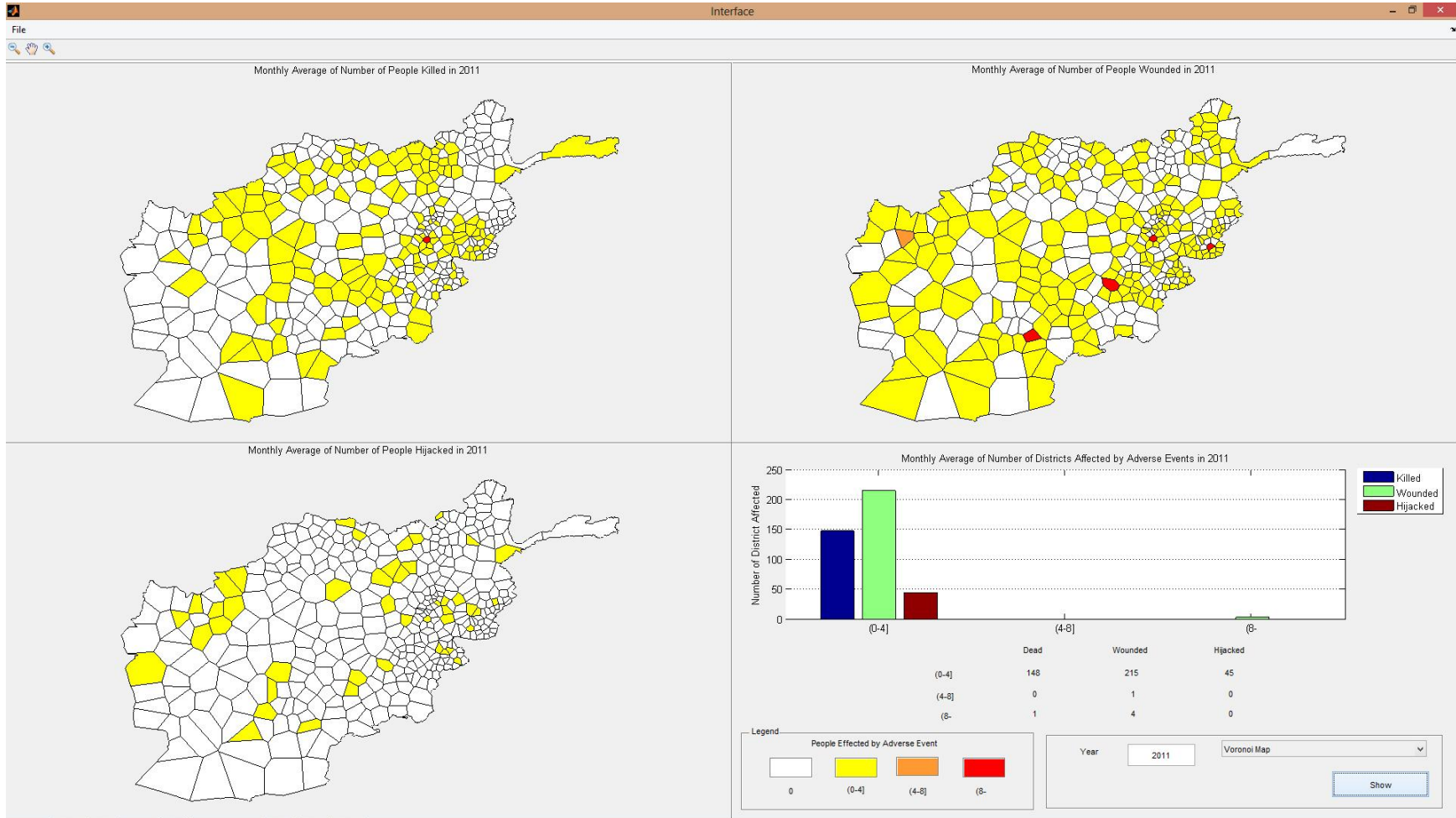
Figure 36 Monthly Average number of people effected by adverse events for 2011
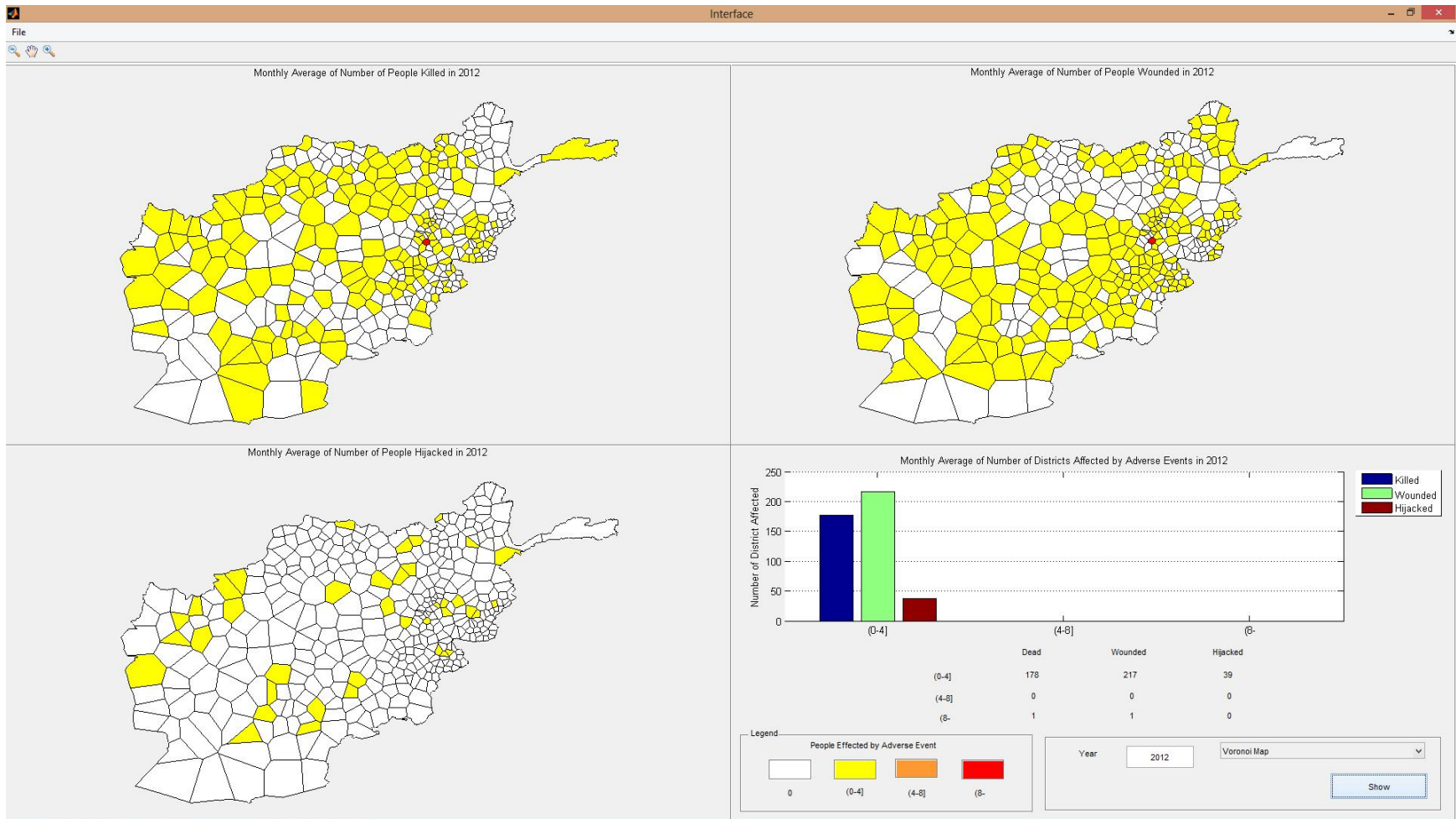
Figure 37 Monthly Average number of people effected by adverse events for 2012
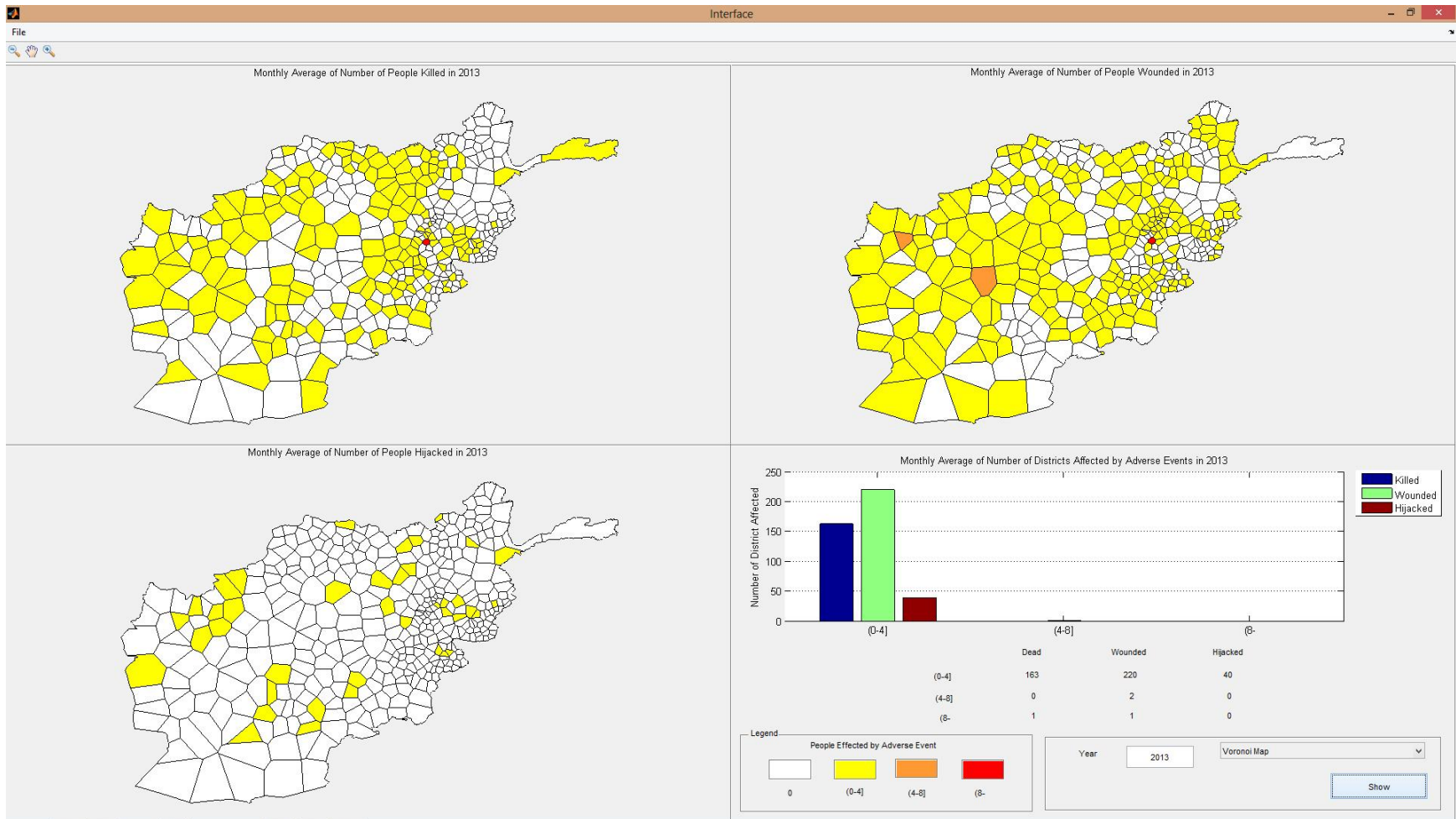
Figure 38 Monthly Average number of people effected by adverse events for 2013
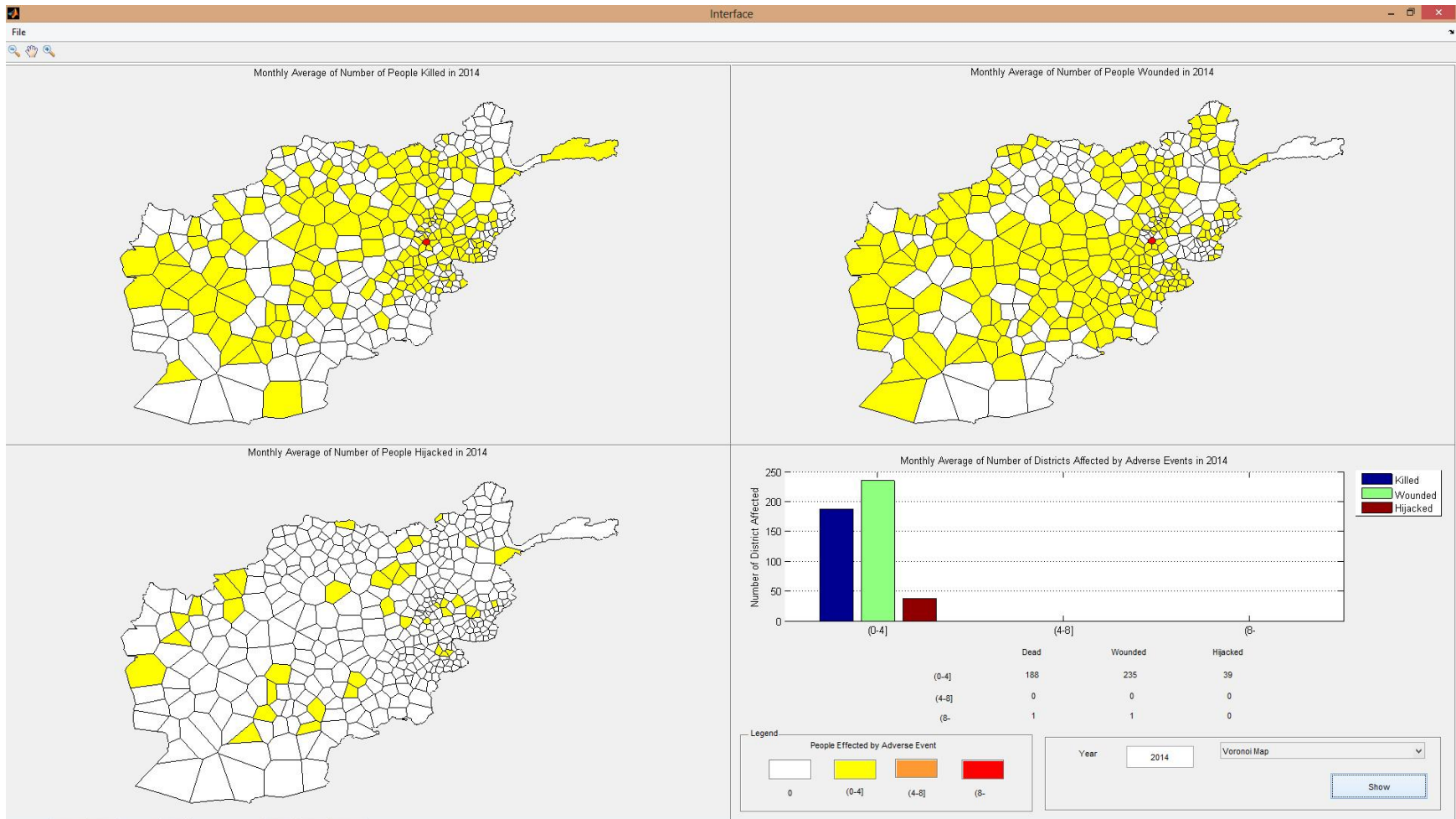
Figure 39 Monthly Average number of people effected by adverse events for 2014

CHAPTER SIX: CONCLUSIONS

This study was set up to support decision makers by building a cellular automata model to understand socio-economic factors and their effect on adverse events. The analysis demonstrated that the use of cellular automata has several significant contributions to support decision makers to allocate development funds to better stabilize regions with higher adverse event risk, and to better understand the nonlinearity and complex socio-economic interactions with adverse events. This capability will allow decision makers to subordinate kinetic operations and promote non-kinetic operations to govern economic programs better in order to initiate efforts and development to address the grievances among the displeased by adverse events.

The analysis performed on a set of spatial data representing factors from socio-economic data and adverse events. This data is used in designed cellular automata model in which the socio-economic factors and adverse event are the states of the districts. Districts in the same neighborhood synchronously interact with each other to determine their next states. Small changes in one iteration of the model yield to complex formations of adverse event risk after several iterations of time. In this study the modeling methodology of cellular automata for social and economic analysis is designed in two major implementation levels; macro and micro-level. In the macro-level, the modeling framework integrates population, social, and economic sub-systems. The macro-level allows model to use regionalized representations, and understand why the events have occurred. Macro-level subsystems are supported by cellular automata rules to generate accurate predictions. Prediction capability of cellular automata is used to model the micro-level interactions between individual actors, which are represented by adverse events.

The results of this study demonstrate that cellular automata model is capable of evaluating socio-economic influences that result in change in adverse events and identifies location, time and impact of these events likely to occur. Performance tests of the model have shown that model is capable of predicting number of people killed, wounded and hijacked with 83, 81 and 93 percent success rates respectively.

Secondly this study demonstrates that the socio-economic influences have different levels of impact on each adverse event, which are number of killed, wounded and hijacked people. While the money spent for USAid and numbers of aids conducted to a district have medium impact on the number of people killed, the population of a district has medium impact on number of people hijacked.

Thirdly, this study demonstrates that the socio-economic influences and adverse events happened in a district have impact on adverse events that will happen in neighboring districts. For instance, total number of wounded people at all neighbors of a district has medium impact on all types of adverse events that will happen in the centering district.

Lastly, to present the capabilities of the developed cellular automata model, a what-if scenario is generated to find out the impact of USAid and population to predict adverse events. For the scenario examined, the amount of money spent and the count of aid conducted on USAid are doubled while keeping the distribution of aid the same. Results of the scenario examined concluded that, doubling USAid does not necessarily reduce the amount of people affected by adverse events in a district. However, more positively affected districts are predicted in comparison to the ones that are affected negatively.

6.1   Contributions to Literature

This study investigates the socio-economic factors and adverse events in an active war theater using the cellular automata modeling approach. There is currently limited research conducted in this area. This study made four significant contributions to the literature.

- The study models the effects of socio-economic factors on adverse events in active war theater.

- A novel cellular automata modeling approach is built. This modeling approach allows for adaptation of changes in the infrastructure aid data to forecast adverse events.

- The developed cellular automata model is the first where the user can create "what-if" scenarios to forecast adverse events.

- The proposed cellular automata modeling approach is capable of identifying location, time, and impact of future adverse events.

6.2   Study Limitations and Future Work

Performance of the model that is introduced in this study can be improved by testing the model on different datasets which have similar spatial data structure used in this study. These data sets include but are not limited to socio-economic factors and adverse event data for different countries.

This study can be improved by broadening the spectrum of the data, including, but not limited to the number of schools, educators or hospitals to create a base infrastructure for districts, and by adding each received USAid as an improvement on to the initially defined base structure.

Future study should include improvements on weight of impact calculations. In this study, weight of impact is calculated by using brute force method which is not efficient in terms of computation time. The high computation time of weight calculations led this study to conduct the calculation of weights within a limited set of variables. With the help of improved weight of impact calculation method, it will be possible to calculate weights for each district, each sector and activity type. Not only is the detailed weight calculation expected to improve the performance of the model but also it can be interpreted as socio-economic drivers of the particular district.

This study can be improved by implementing an optimization method for distributing the optimum amount of USAid through necessary sectors and their activity type. This improvement can support decision makers on deciding on the distribution of aid in terms of amount location and time.

Decision makers can be supported by a simulator which is able to generate forecast on adverse events by the cellular automata model introduced in this study. This simulator will have the ease of use graphical user interface to input different data for different locations or countries. This simulator will support decision makers to allocate development funds to better stabilize regions with higher adverse event risks. With the help of this simulator decision makers will able to generate "what-if" scenarios to forecast adverse events under controlled environments.

The cellular automata model built in this study could be used to predict the impact of U.S. troop reduction and transferring field functions to the Afghani forces. Most likely aid will be reduced but the model could be used to predict some impact. This model can help Taliban to deal with the pull out of U.S. forces

LIST OF REFERENCES

Alegría, M., Pescosolido, B. A., Williams, S., & Canino, G. (2011). Culture, Race/Ethnicity and Disparities: Fleshing Out the Socio-Cultural Framework for Health Services Disparities Handbook of the Sociology of Health, Illness, and Healing. In B. A. Pescosolido, J. K. Martin, J. D. McLeod & A. Rogers (Eds.), (pp. 363-382): Springer New York.

Banks, E. R. (1970, 28-30 Oct. 1970). *Universality in cellular automata.* Paper presented at the Switching and Automata Theory, 1970., IEEE Conference Record of 11th Annual Symposium on.

Bing Sheng, W., & Sui, D. Z. (2008, 7-11 July 2008). *CA-based Simulation of Asian Urban Dynamics: A Case Study of Taipei Metropolitan Area, Taiwan.* Paper presented at the Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International.

Calvo Garzón, P., Laakso, A., & Gomila, T. (2008). Dynamics and Psychology. *New Ideas in Psychology, 26*(2), 143-145. doi: 10.1016/j.newideapsych.2008.02.001

Center, A. R. (2011). *Find Adjacent & Neighboring Polygons* [ArcGIS toolbox]. Retrieved from: http://resources.arcgis.com/gallery/file/geoprocessing/details?entryID=50F58FCF-1422-2418-884B-A053393CEF92

Chialvo, D. R. (2008). Emergent complexity: What uphill analysis or downhill invention cannot do. *New Ideas in Psychology, 26*(2), 158-173. doi: 10.1016/j.newideapsych.2007.07.013

Codd, E. F. (1968). *Cellular automata*. New York: Academic Press.

Conway, J. (1970). The game of life. *Scientific American, 223*(4), 4.

Curran, P. J. (1984). Geographic Information Systems. *Area, 16*(2), 153-158.

Deadman, P., Brown, R. D., & Gimblett, H. R. (1993). Modelling Rural Residential Settlement Patterns with Cellular Automata. *Journal of Environmental Management, 37*(2), 147-160. doi: 10.1006/jema.1993.1011

Durand-Lose, J. (1997). Intrinsic universality of a 1-dimensional reversible Cellular Automaton STACS 97. In R. Reischuk & M. Morvan (Eds.), (Vol. 1200, pp. 439-450): Springer Berlin / Heidelberg.

Durand, B. R., Z. . (1999). The Game of Life: universality revisited. *Matmematics and Its Applications, 460*, 51-76.

Flache, A., & Hegselmann, R. (2001). Do irregular grids make a difference? Relaxing the spatial regularity assumption in cellular models of social dynamics. *Journal of Artificial Societies and Social Simulation, 4*(4).

Frameworks, P. C. o. U. S., & Council, N. R. (2011). *Sociocultural Data to Accomplish Department of Defense Missions:Toward a Unified Social Framework*: The National Academies Press.

Fridman, N., Zilberstein, T., & Kaminka, G. (2011). Predicting Demonstrations' Violence Level Using Qualitative Reasoning

Social Computing, Behavioral-Cultural Modeling and Prediction. In J. Salerno, S. Yang, D. Nau & S.-K. Chai (Eds.), (Vol. 6589, pp. 42-50): Springer Berlin / Heidelberg.

Han, J., Hayashi, Y., Cao, X., & Imura, H. (2009). Application of an integrated system dynamics and cellular automata model for urban growth assessment: A case study of Shanghai, China. *Landscape and Urban Planning, 91*(3), 133-141. doi: 10.1016/j.landurbplan.2008.12.002

Hegselmann, R., & Flache, A. (1998). Understanding complex social dynamics: A plea for cellular automata based modelling. *Journal of Artificial Societies and Social Simulation, 1*(3), 1.

Hoekstra, A., Kroc, J., & Sloot, P. (2010). Introduction to Modeling of Complex Systems Using Cellular Automata Simulating Complex Systems by Cellular Automata. In J. Kroc, P. M. A. Sloot & A. G. Hoekstra (Eds.), (Vol. 0, pp. 1-16): Springer Berlin / Heidelberg.

Humanitarian Response, C. A. F. O. D. R. (2011). *Afghanistan District Boundaries (admin 3)* [Areas (polygons)]. Retrieved from:

http://cod.humanitarianresponse.info/sites/default/files/afg_admbnd_adm3_pol.zip

J. Albert, K. C. I. (1987). A Simple Universal Cellular Automaton and its One-Way and Totalistic Version. *Complex Systems*, 1-16.

Jat, M. K., Garg, P. K., & Khare, D. (2008). Monitoring and modelling of urban sprawl using remote sensing and GIS techniques. *International Journal of Applied Earth Observation and Geoinformation, 10*(1), 26-43. doi: 10.1016/j.jag.2007.04.002

Kari, J. (2005). Theory of cellular automata: A survey. *Theoretical Computer Science, 334*(1–3), 3-33. doi: 10.1016/j.tcs.2004.11.021

Lauf, S., Haase, D., Hostert, P., Lakes, T., & Kleinschmit, B. (2012). Uncovering land-use dynamics driven by human decision-making – A combined model approach using cellular automata and system dynamics. *Environmental Modelling & Software, 27–28*(0), 71-82. doi: 10.1016/j.envsoft.2011.09.005

Mago, V. K., Bakker, L., Papageorgiou, E. I., Alimadad, A., Borwein, P., & Dabbaghian, V. (2012). Fuzzy cognitive maps and cellular automata: An evolutionary approach for social systems modelling. *Applied Soft Computing*(0). doi: 10.1016/j.asoc.2012.02.020

Margolus, N. (1984). Physics-like models of computation. *Physica D: Nonlinear Phenomena, 10*(1–2), 81-95. doi: 10.1016/0167-2789(84)90252-5

Pearson, J. A., & Geronimus, A. T. (2011). Race/Ethnicity, Socioeconomic Characteristics, Coethnic Social Ties, and Health: Evidence From the National Jewish Population Survey. *American Journal of Public Health, 101*(7), 1314-1321. doi: 10.2105/ajph.2009.190462

Peccol, E., Bird, C. A., & Brewer, T. R. (1996). GIS As A Tool For Assessing The Influence Of Countryside Designations and Planning Policies On Landscape Change. *Journal of Environmental Management, 47*(4), 355-367. doi: 10.1006/jema.1996.0059

Russo, S., Roccato, M., & Vieno, A. (2011). Predicting Perceived Risk of Crime: a Multilevel Study. *American Journal of Community Psychology, 48*(3), 384-394. doi: 10.1007/s10464-010-9386-x

Siebeneck, L. K., Medina, R. M., Yamada, I., & Hepner, G. F. (2009). Spatial and Temporal Analyses of Terrorist Incidents in Iraq, 2004–2006. *Studies in Conflict & Terrorism, 32*(7), 591-610. doi: 10.1080/10576100902961789

Spicer, V., Reid, A. A., Ginther, J., Seifi, H., & Dabbaghian, V. (2012). Bars on blocks: A cellular automata model of crime and liquor licensed establishment density. *Computers, Environment and Urban Systems*. doi: 10.1016/j.compenvurbsys.2012.02.004

Toffoli, T. (1977). Computation and construction universality of reversible cellular automata. *Journal of Computer and System Sciences, 15*(2), 213-231. doi: 10.1016/s0022-0000(77)80007-x

Turing, A. M. (1937). *On computable numbers, with an application to the Entscheidungsproblem : a correction*. London: C.F. Hodgson & Son.

Vaz, E. d. N., Nijkamp, P., Painho, M., & Caetano, M. (2012). A multi-scenario forecast of urban change: A study on urban growth in the Algarve. *Landscape and Urban Planning, 104*(2), 201-211. doi: 10.1016/j.landurbplan.2011.10.007

Von Neumann, J., Burks, A. W., & Burks, A. W. e. (1966). *Theory of self-reproducing automata. Edited and completed by Arthur W. Burks*. Urbana: University of Illinois Press.

White, D., & Kiester, A. R. (2008). Topology matters: Network topology affects outcomes from community ecology neutral models. *Computers, Environment and Urban Systems, 32*(2), 165-171. doi: 10.1016/j.compenvurbsys.2007.11.002

Wolfram, S. (1983). Statistical mechanics of cellular automata. *Reviews of Modern Physics, 55*(3), 601-644.

Wu, D., Liu, J., Wang, S., & Wang, R. (2009). Simulating urban expansion by coupling a stochastic cellular automata model and socioeconomic indicators. *Stochastic Environmental Research and Risk Assessment, 24*(2), 235-245. doi: 10.1007/s00477-009-0313-3

Yu, X. J., & Ng, C. N. (2007). Spatial and temporal dynamics of urban sprawl along two urban–rural transects: A case study of Guangzhou, China. *Landscape and Urban Planning, 79*(1), 96-109. doi: 10.1016/j.landurbplan.2006.03.008

APPENDIX A
MATLAB CODE FOR THE CELLULAR AUTOMATA MODEL

Filename: HSCB_CA.m

```matlab
%Project: Cellular Automata Model for HSCB Project
%Author: Halil Bozkurt
%Function: Main function for the HSCB cellular automata model
%-----------------------------------
%Inputs: District states, Neighboring district states, GIS
neighboring
%   data, Ratio of changes for centering districts, Lookup table
for weight factors startup data
load('0509for2020.mat');  %loads
BestID=[3935 6130 2407 5353];
Out=zeros(52800,4);
parfor bid=1:4
    %compose weight of impact
    optindex=BestID(1,bid);
    OptMultiplier=[0.1 0.1 repmat(CPULookup(optindex,1),1,168)
repmat(CPULookup(optindex,2),1,6) CPULookup(optindex,3)
CPULookup(optindex,4) CPULookup(optindex,5)
CPULookup(optindex,6)
repmat(CPULookup(optindex,1)/CPULookup(optindex,7),1,168)
repmat(CPULookup(optindex,2)/CPULookup(optindex,7),1,5)
CPULookup(optindex,3)/CPULookup(optindex,8)
CPULookup(optindex,4)/CPULookup(optindex,8)
CPULookup(optindex,5)/CPULookup(optindex,8)
CPULookup(optindex,6)/CPULookup(optindex,8)];
    %Run cellular automata function
    Predictions=Cellular_Automata( Neigh_Dist, CenterData,
NeighbourData,RateofChange, start, OptMultiplier);
    % Calculate performance measures and write in to file
    OptimizationResults=PerformanceMeasure(Predictions,
RealData);
    fname='outputsummary.txt';
    datafile=fopen(fname,'a');
    fprintf(datafile,'\n');
    fprintf(datafile,'%g\t', [optindex
OptimizationResults(1,:)]);
    fclose (datafile);
    %Compose output for differetn weight factors
    Out(:,bid)=Predictions(:,bid+4);
    Presults(bid,:)=OptimizationResults;
end
```

Filename:Cellular_Automata.m

```matlab
function [ Predictions ] = Cellular_Automata( Neighbourhood,
CenterData, NeighbourData, RateofChange, Start, OptMultiplier )
%Project: Cellular Automata Model for HSCB Project
%Author: Halil Bozkurt
%Function: Runs cellular automata methodology
%------------------------------------
%Inputs: District states, Neighboring district states, GIS
neighboring
%    data, Ratio of changes for centring districts,
%    Lookup table for weight factors, startup data
Compdata=[CenterData(:,5:end) NeighbourData(:,5:end)];
%Standadize training data
[Normalized,comp_mean,comp_dev]=zscore([CenterData(:,1:2)
Compdata]);
Resultset=RateofChange;
Normalizedopt=Normalized.*repmat(OptMultiplier,23600,1);
%Runs model for 400 districts for 12 months since input data is
ordered in years and months iteerations are in months
for i=1:4800
    cur_year=Start(i,1);
    cur_month=Start(i,2);
    cur_prov=Start(i,3);
    cur_dist=Start(i,4);
    % Generates predictor data center data + new neighbours
    %(2"year and month"+178"centering cell"+177"neighboring
cell")
    %Neighboringcellstates calculated by sumofneighbor function
    Testdata=[Start(i,1:2) Start(i,5:end)
sumofneighbour(cur_dist,cur_year,cur_month, Neighbourhood,
Start)];
    Norm_TestData=(Testdata-comp_mean)./comp_dev;
    Norm_TestData(Norm_TestData~=Norm_TestData)=0;
    Norm_TestData=Norm_TestData.*OptMultiplier;
    % Calculates the eucladian distance
    E_Distance=pdist2(Normalizedopt,Norm_TestData, 'euclidean');
    min_indice= find(E_Distance(:,1)==min(E_Distance(:,1)));
    Minimums=Resultset(min_indice,:);
    %Runs ratefinder function for weight of impact
prioritization
rateofchange=ratefinder(Minimums,cur_dist,cur_prov,cur_year,cur_
month);
    %Applies the transition rule
```

```matlab
iteration_indice_t=ind2sub(size(Start),find(Start(:,1)==cur_year
& Start(:,2)==cur_month & Start(:,4)==cur_dist));
    [nxt_year,nxt_month]=nextmonth(cur_year, cur_month);

iteration_indice_t1=ind2sub(size(Start),find(Start(:,1)==nxt_yea
r & Start(:,2)==nxt_month & Start(:,4)==cur_dist)); %iteration
indice atr t=t+1
    for y=179:182
        zerocheck = Start(iteration_indice_t,y);
       if rateofchange (1,(y-178))>0 && zerocheck==0,
           zerocheck=1;
        end
          Start(iteration_indice_t1,y)=zerocheck*rateofchange
(1,(y-178));
    end
end
    %Returns predictions
    Predictions=[Start(401:end,1:4) Start(401:end,179:182)];
end
```

Filename: sumofneighbor.m

```matlab
function [ sum_of_neighbours ] = sumofneighbour(
current_district, current_year, current_month,
Neigh_Dist,N_State)
%Project: Cellular Automata Model for HSCB Project
%Author: Halil Bozkurt
%Function: Sum of neighbor calculates the sum of neighbor
covariates
% Neigh_Dist and N_State datasets are needed
%Inputs current district, current year and current month outputs
the sum
%  of the covairates of the neighbouring cells of the input
cell.
    neig_indice=ind2sub(size(Neigh_Dist),
find(Neigh_Dist(:,2)==current_district));
    cur_neigh_sum=zeros(1,177);
    for Neighbour=3:15
        if Neigh_Dist(neig_indice,Neighbour)==0,
            break;
        end

neigdis_indice=ind2sub(size(N_State),find(N_State(:,1)==current_
```

```matlab
year & N_State(:,2)==current_month &
N_State(:,4)==Neigh_Dist(neig_indice,Neighbour)));
        cur_neigh_sum=cur_neigh_sum +
[N_State(neigdis_indice,5:177) N_State(neigdis_indice,179:end)];
    end
sum_of_neighbours=cur_neigh_sum;
end
```

Filename: ratefinder.m

```matlab
function [ rateofchange ] = ratefinder( MinDistances, cur_dist,
cur_prov, cur_year, cur_month,lt  )
%Project: Cellular Automata Model for HSCB Project
%Author: Halil Bozkurt
%Function: Calculates Appropriate rate of change for minimum
distance
%-----------------------------------
localmindist=find(MinDistances(:,4)==cur_dist);
localminprov=find(MinDistances(:,3)==cur_prov);
% eleminate others If same provience have min distance
if isempty(localmindist)==false,
    MinDistances=MinDistances(localmindist,:);
% eleminate others if same district have min prov
elseif isempty(localminprov)==false,
    MinDistances=MinDistances(localminprov,:);
end
sizeofmindistance=size(MinDistances);
% Prioritize islamic months 9 or 12
if sizeofmindistance(1,1)~=1,
    c_islamicmonth=gtoi(cur_year,cur_month);
    if or(c_islamicmonth == 12, c_islamicmonth ==9),
    md_islamicmonth=zeros(sizeofmindistance(1,1),1);
        for i=1:sizeofmindistance(1,1)

md_islamicmonth(i,1)=gtoi(MinDistances(i,1),MinDistances(i,2));
        end
        monthindice=find(md_islamicmonth(:,1)==c_islamicmonth);
        if isempty(monthindice)==0,
            MinDistances=MinDistances(monthindice,:);
        end
    end
end
%returns ratio of change
    rateofchange=mean(MinDistances(:,5:8),1);
```

```matlab
end
```

Filename: _PerformanceMeasure.m_

```matlab
function [ Optimization_results ] = PerformanceMeasure(
Predictions, rd )
%%Project: Cellular Automata Model for HSCB Project
%Author: Halil Bozkurt
%Function: Calculates Performance Measures in between predicted
and real data
%---------------------------------------
RealData=rd;
SumofRealAdverseEvents=sorae;
SumofAdverseEvents=zeros(12,5);
year=2010;
    for month=1:12
        Monthlytotalindice=ind2sub(size(Predictions),
find(Predictions(:,1)==year & Predictions(:,2)==month));
        SumofAdverseEvents(month,:)=[month
sum(Predictions(Monthlytotalindice,5:end))];
    end
sucessrate=sum((RealData(:,5:end)-
Predictions(:,5:end))==0)./4800;
mae=sum(abs(RealData(:,5:end)-Predictions(:,5:end)))./4800;
mse=sum((RealData(:,5:end)-Predictions(:,5:end)).^2)./4800;
CorrectTarget=sum(or(RealData(:,5:end)>0 &
(Predictions(:,5:end)>0)==1,(RealData(:,5:end)==0 &
Predictions(:,5:end)==0)==1))./4800;
mae12=sum(abs(SumofRealAdverseEvents(:,2:end)-
SumofAdverseEvents(:,2:end)))./12;
mse12=sum((SumofRealAdverseEvents(:,2:end)-
SumofAdverseEvents(:,2:end)).^2)./12;
PearsonChi12=(sum(((SumofRealAdverseEvents(:,2:end) -
SumofAdverseEvents(:,2:end)).^2)./SumofAdverseEvents(:,2:end)));
Optimization_results(1,:)=[ sucessrate mae mse CorrectTarget
mae12 mse12 PearsonChi12];
end
```

APPENDIX B
ADVERSE EVENT PREDICTION COMPARISON FOR 2010
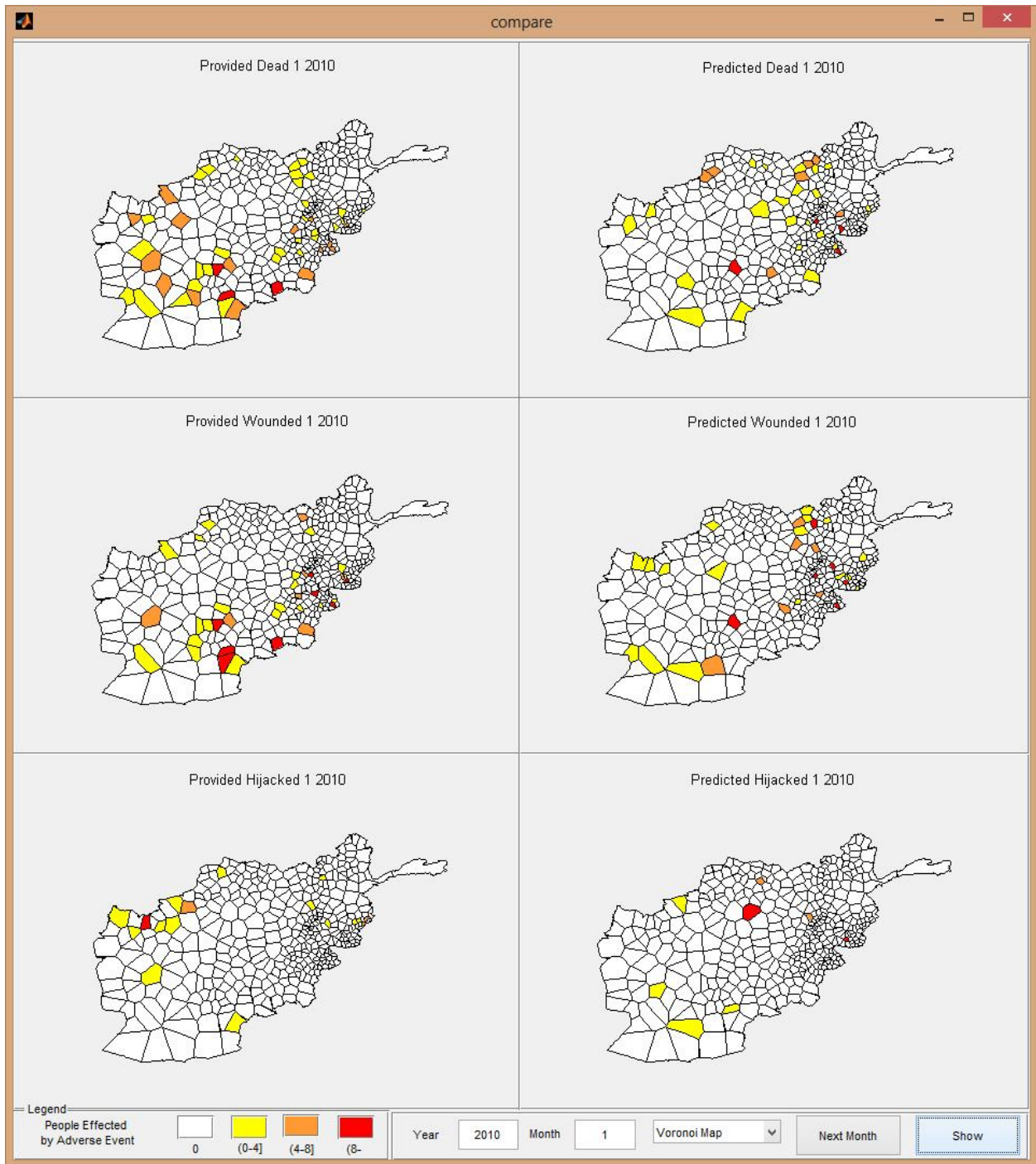
Figure 40 Adverse event comparison for January 2010
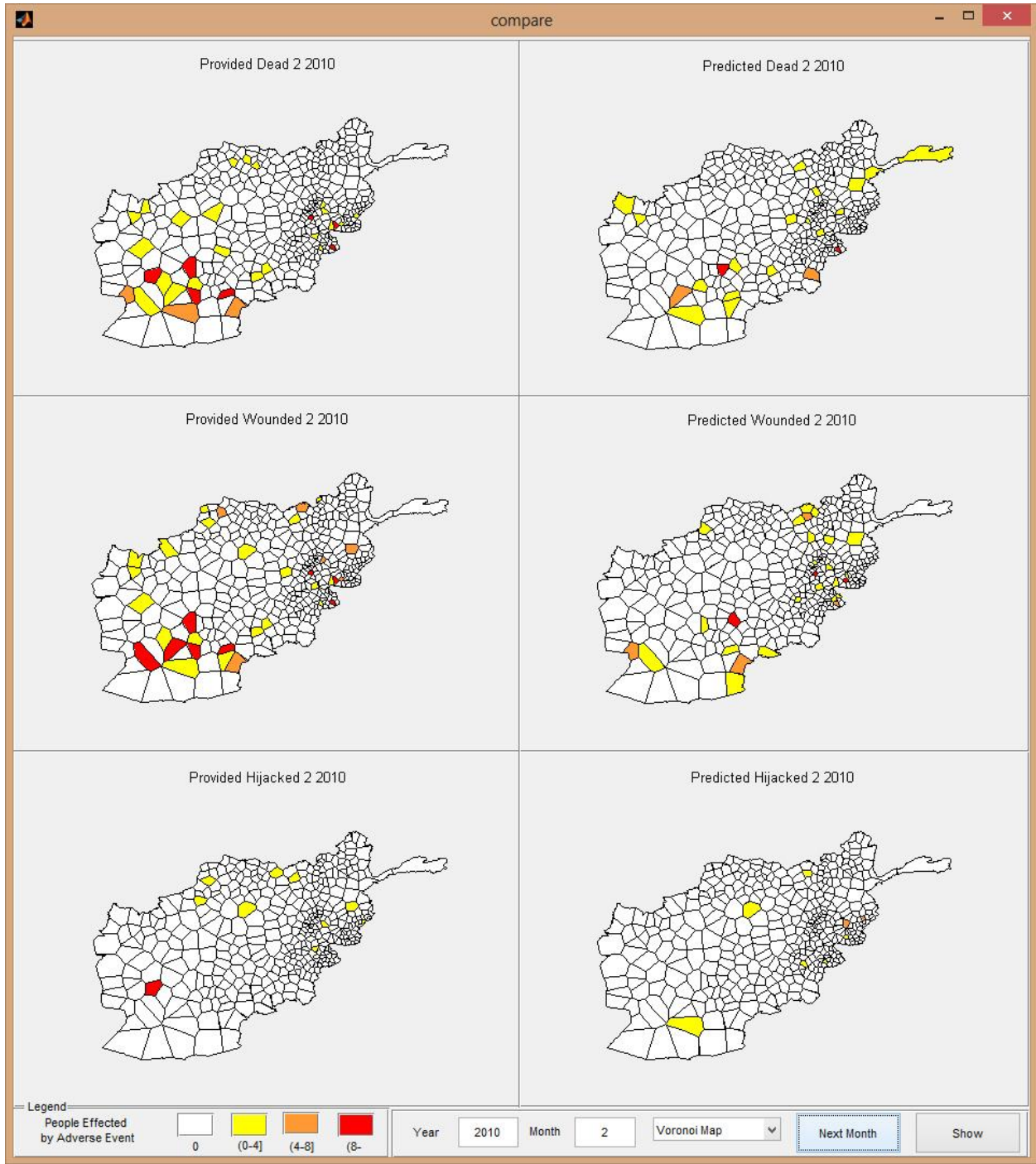
Figure 41 Adverse event comparison for February 2010
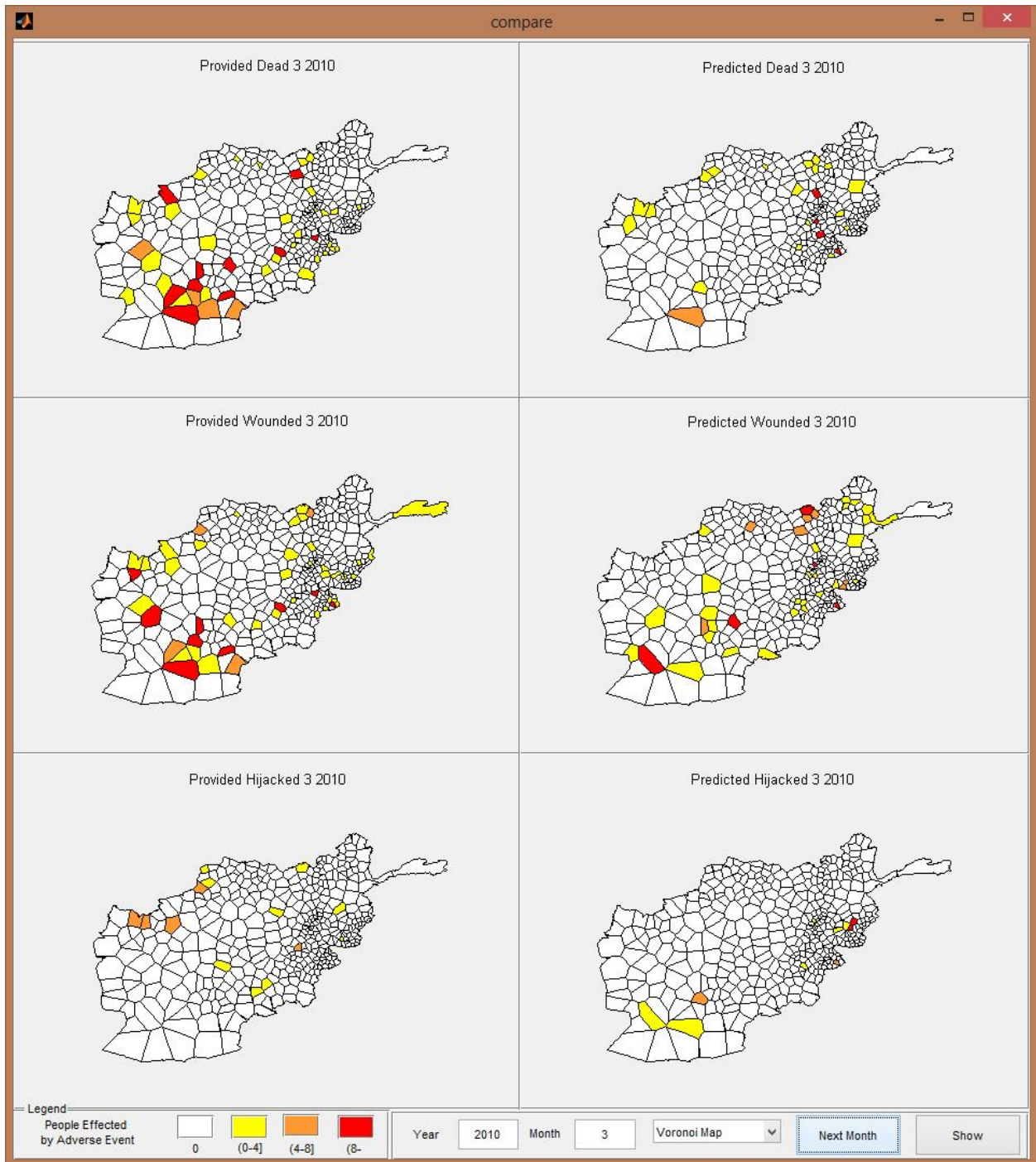
Figure 42 Adverse event comparison for March 2010
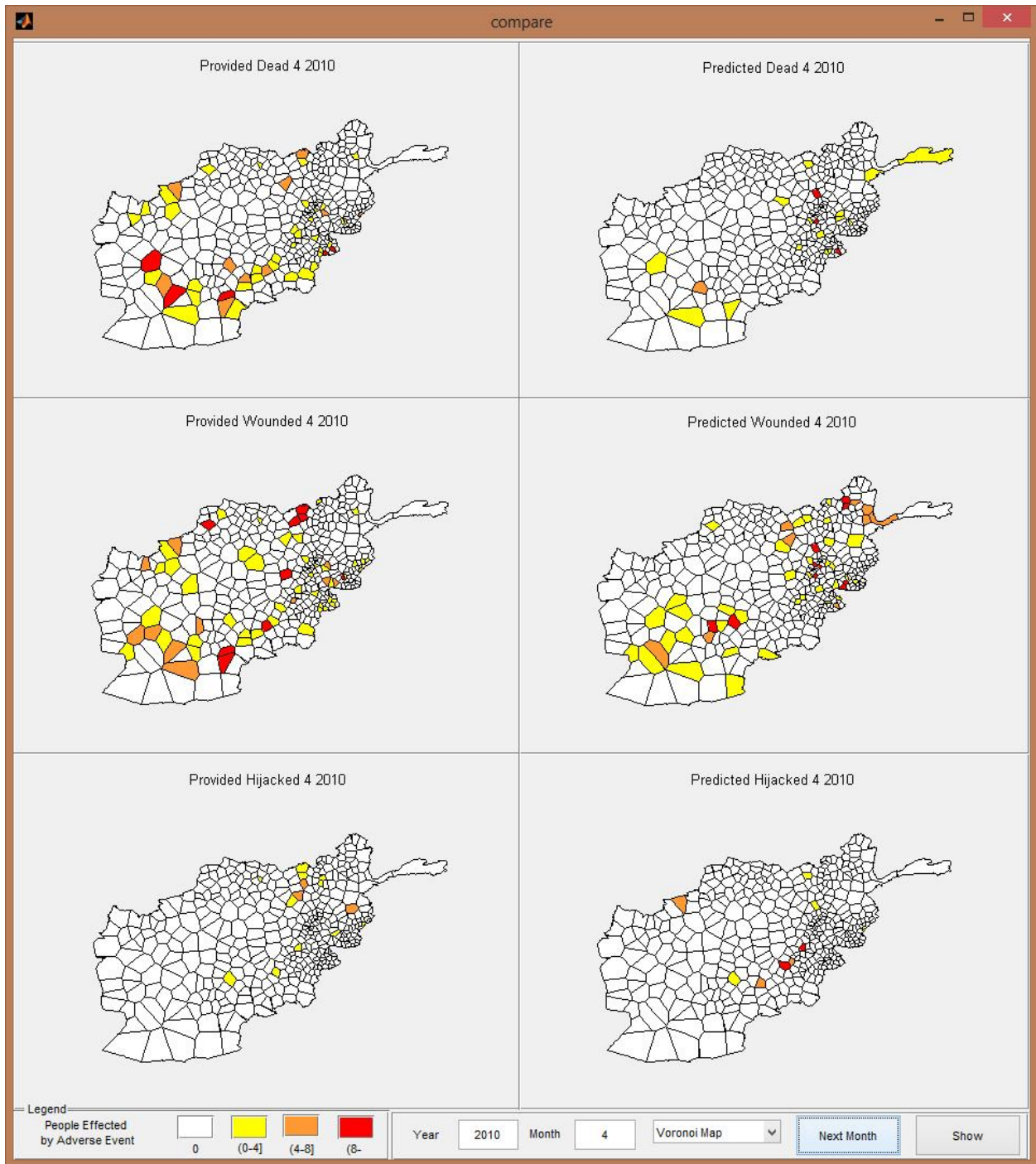
Figure 43 Adverse event comparison for April 2010
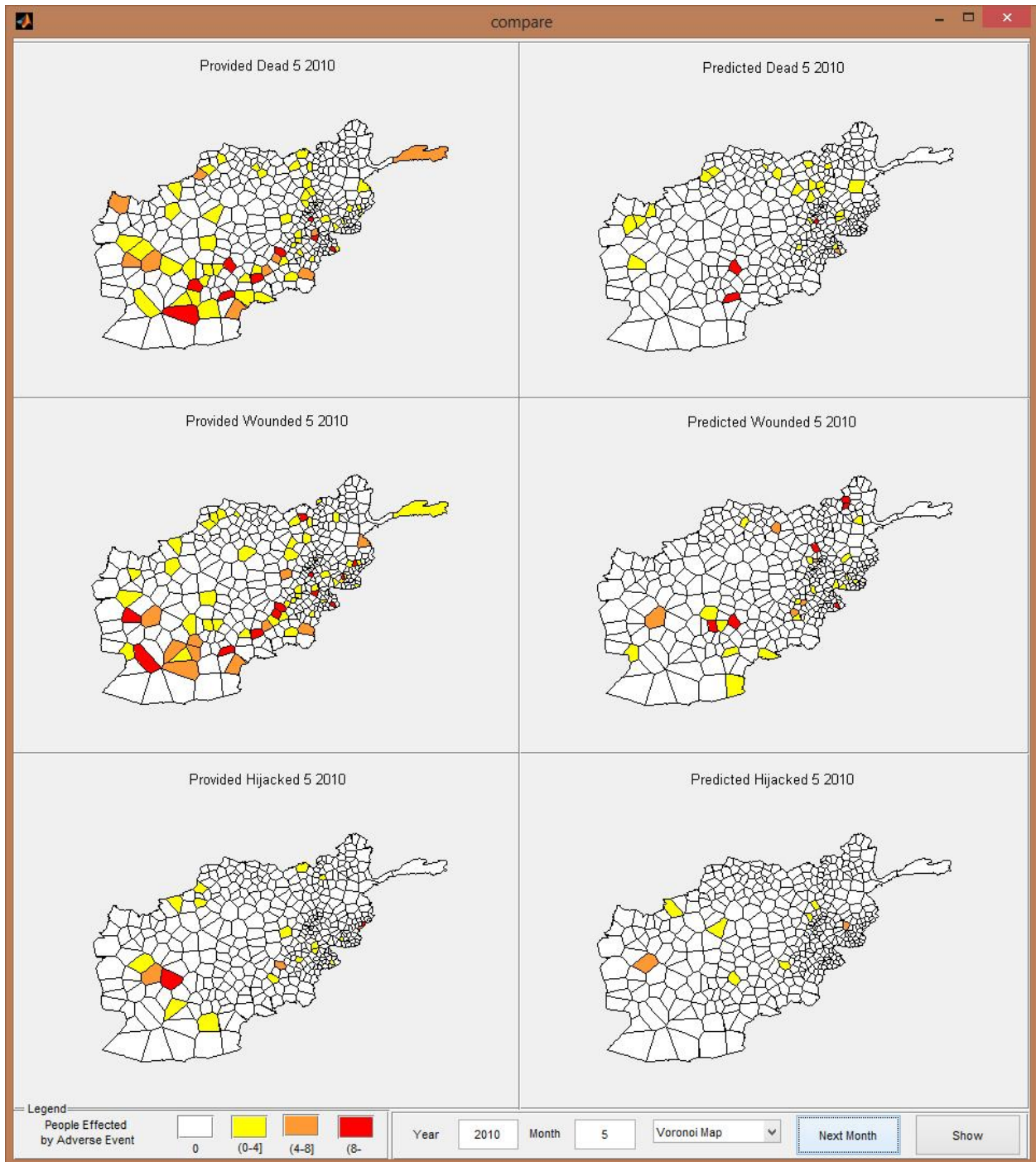
Figure 44 Adverse event comparison for May 2010
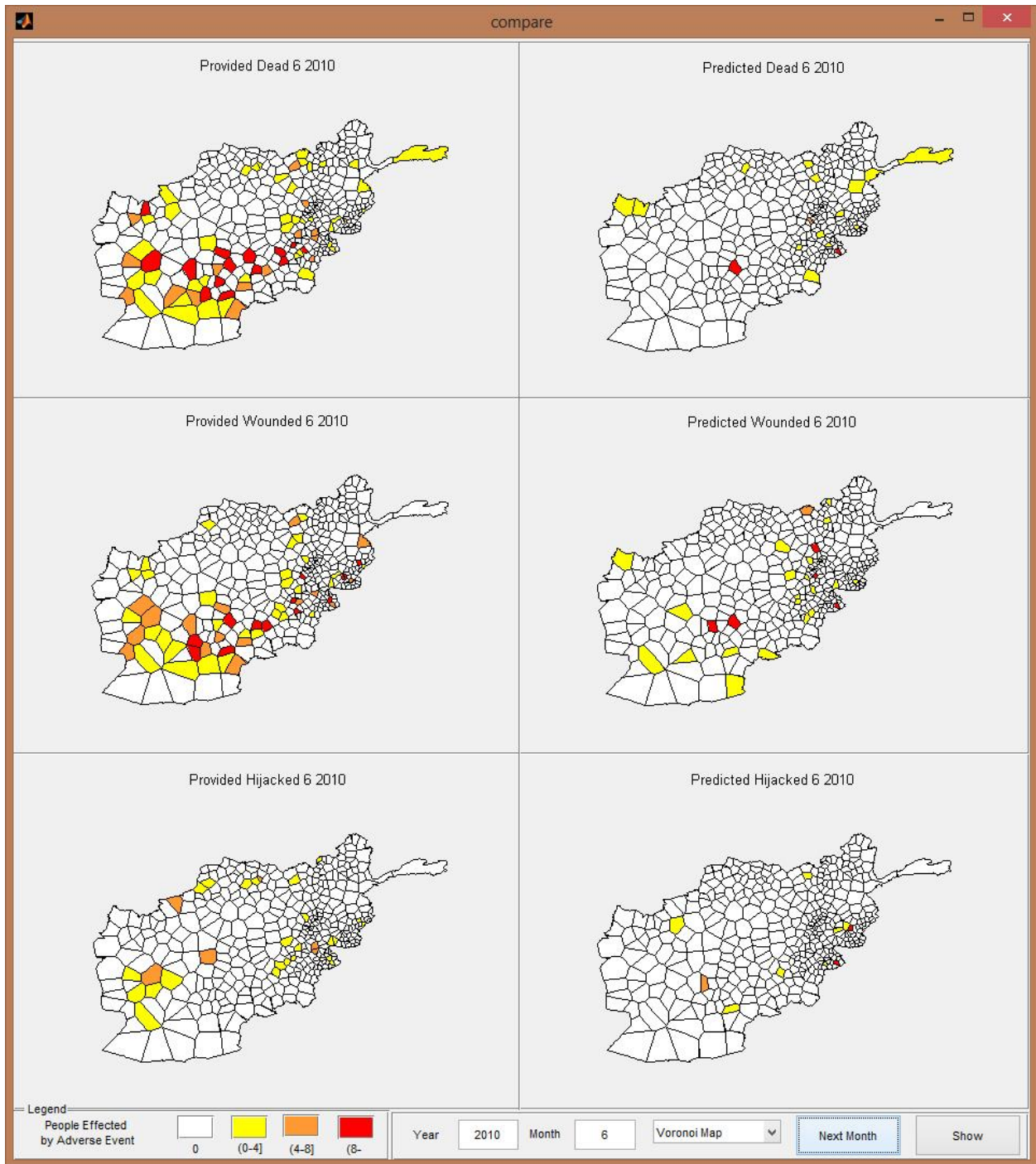
Figure 45 Adverse event comparison for June 2010
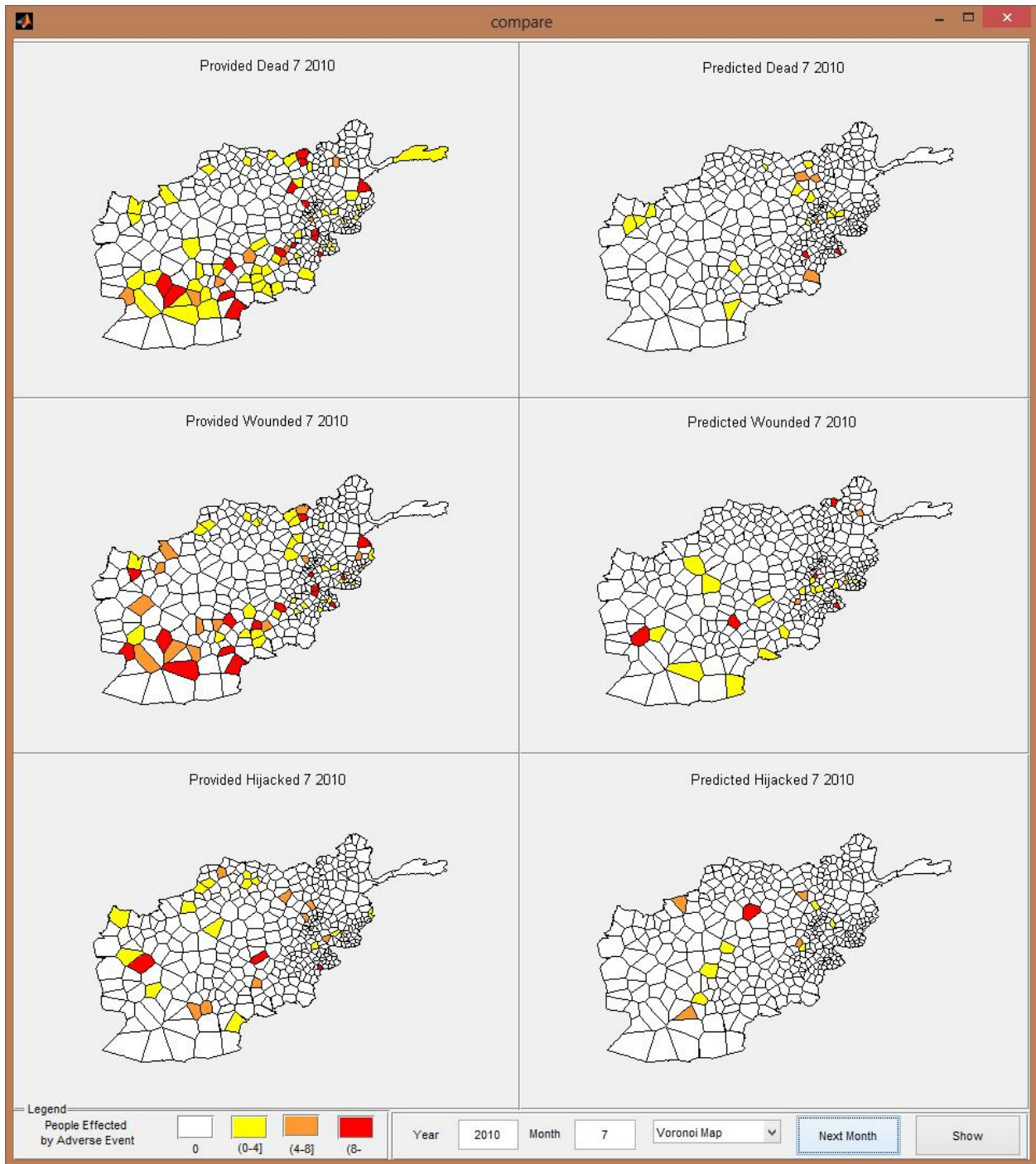
Figure 46 Adverse event comparison for July 2010
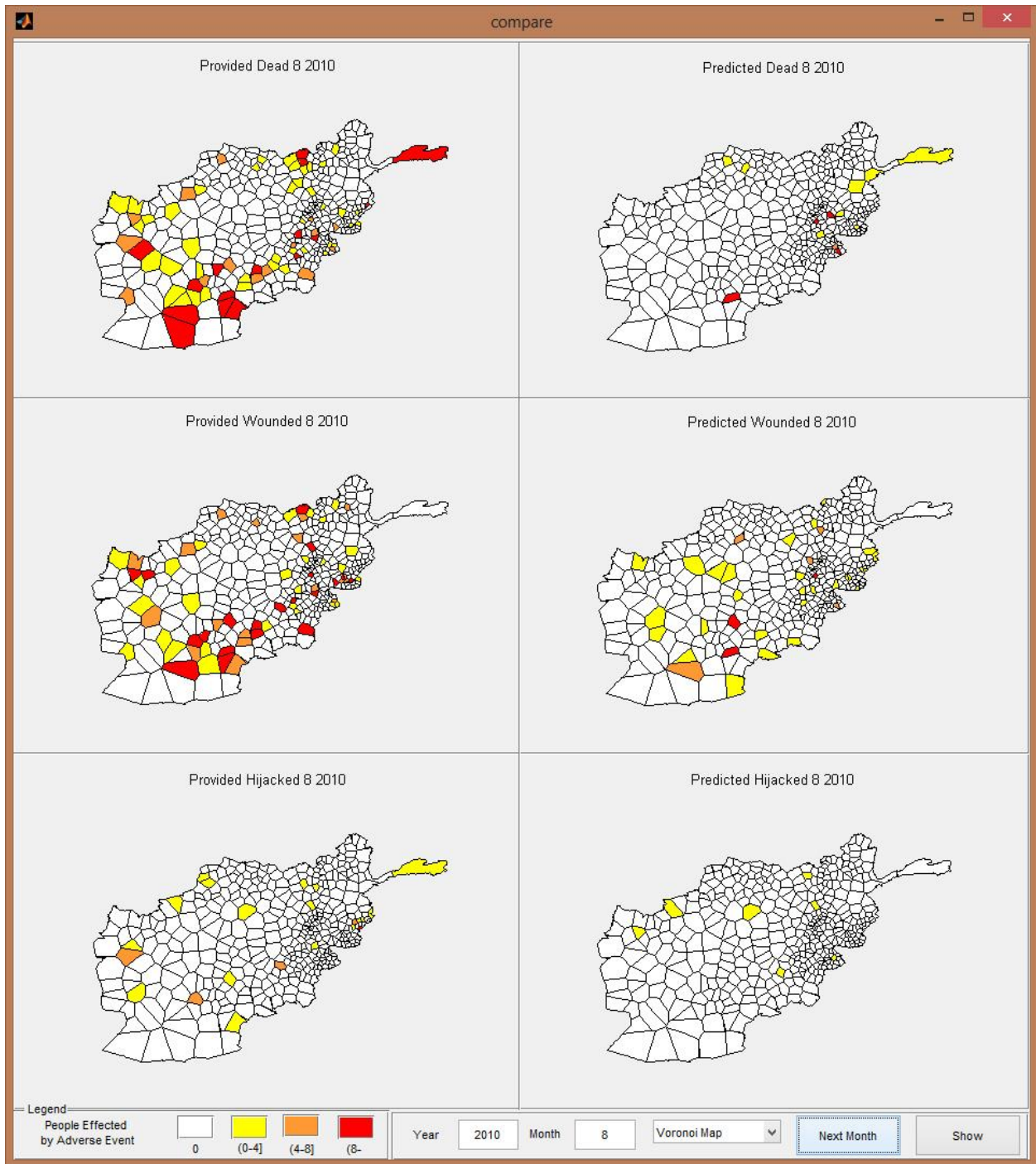
Figure 47 Adverse event comparison for August 2010
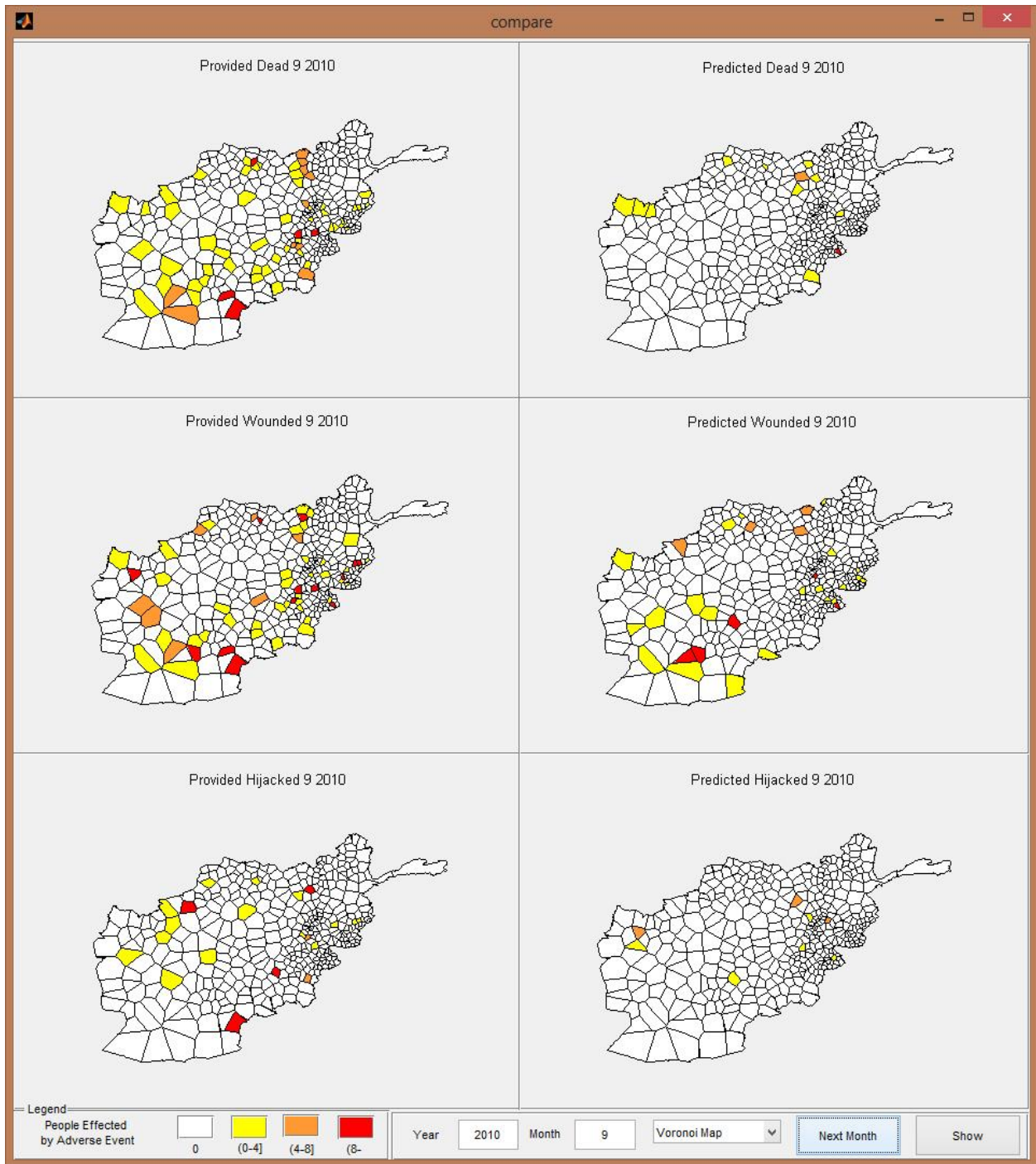
Figure 48 Adverse event comparison for Septembe 2010
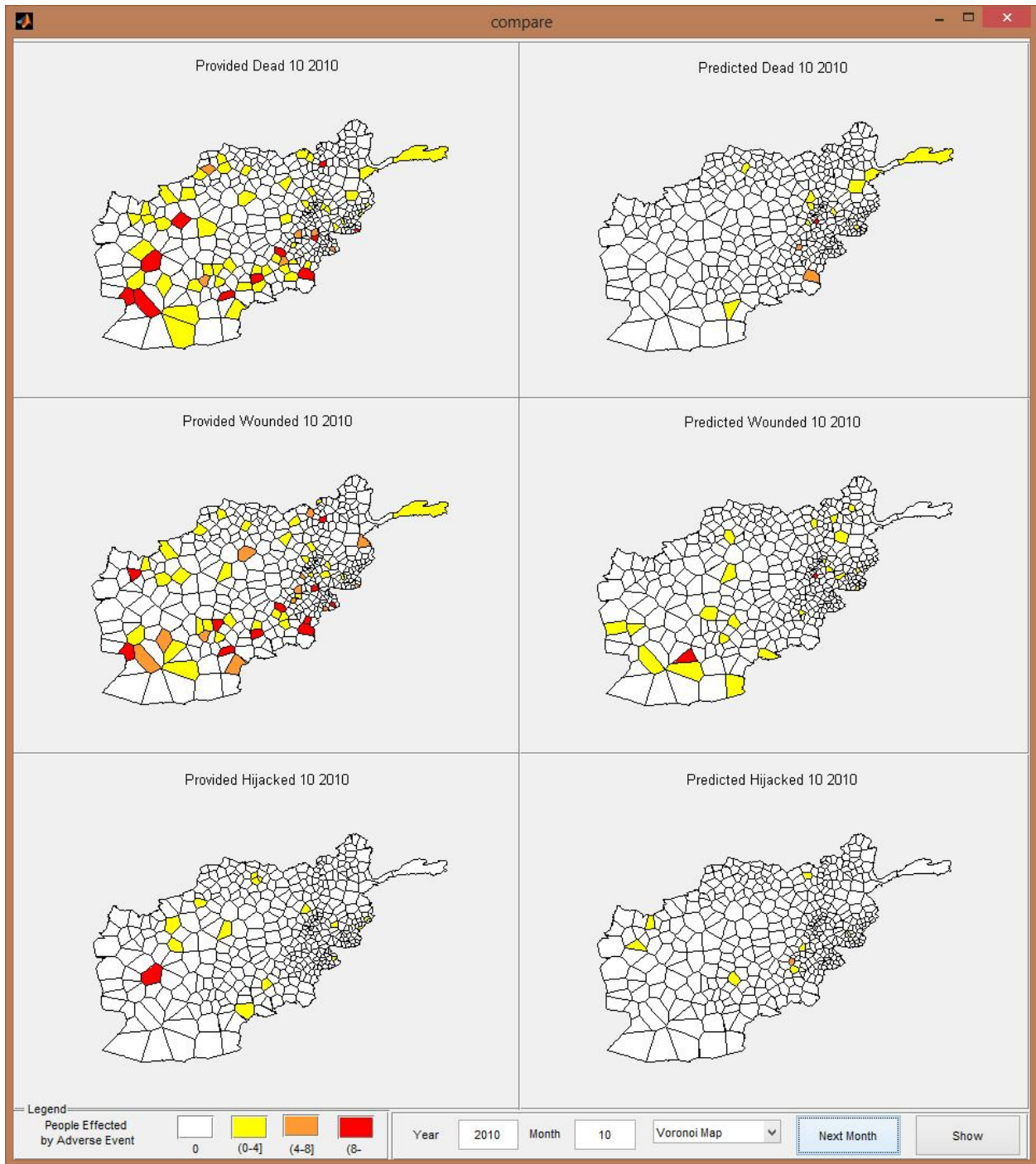
Figure 49 Adverse event comparison for October 2010
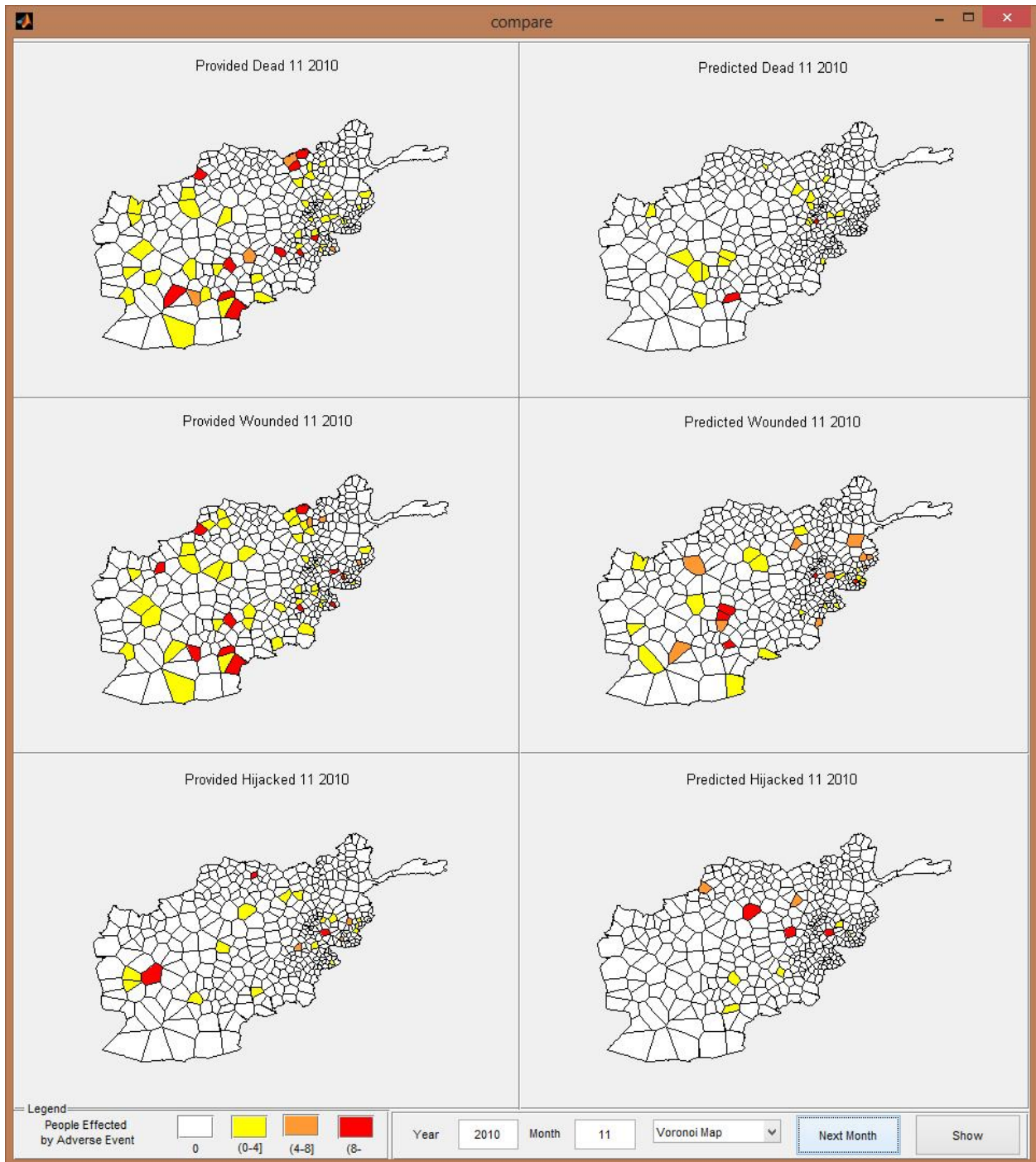
Figure 50 Adverse event comparison for November 2010
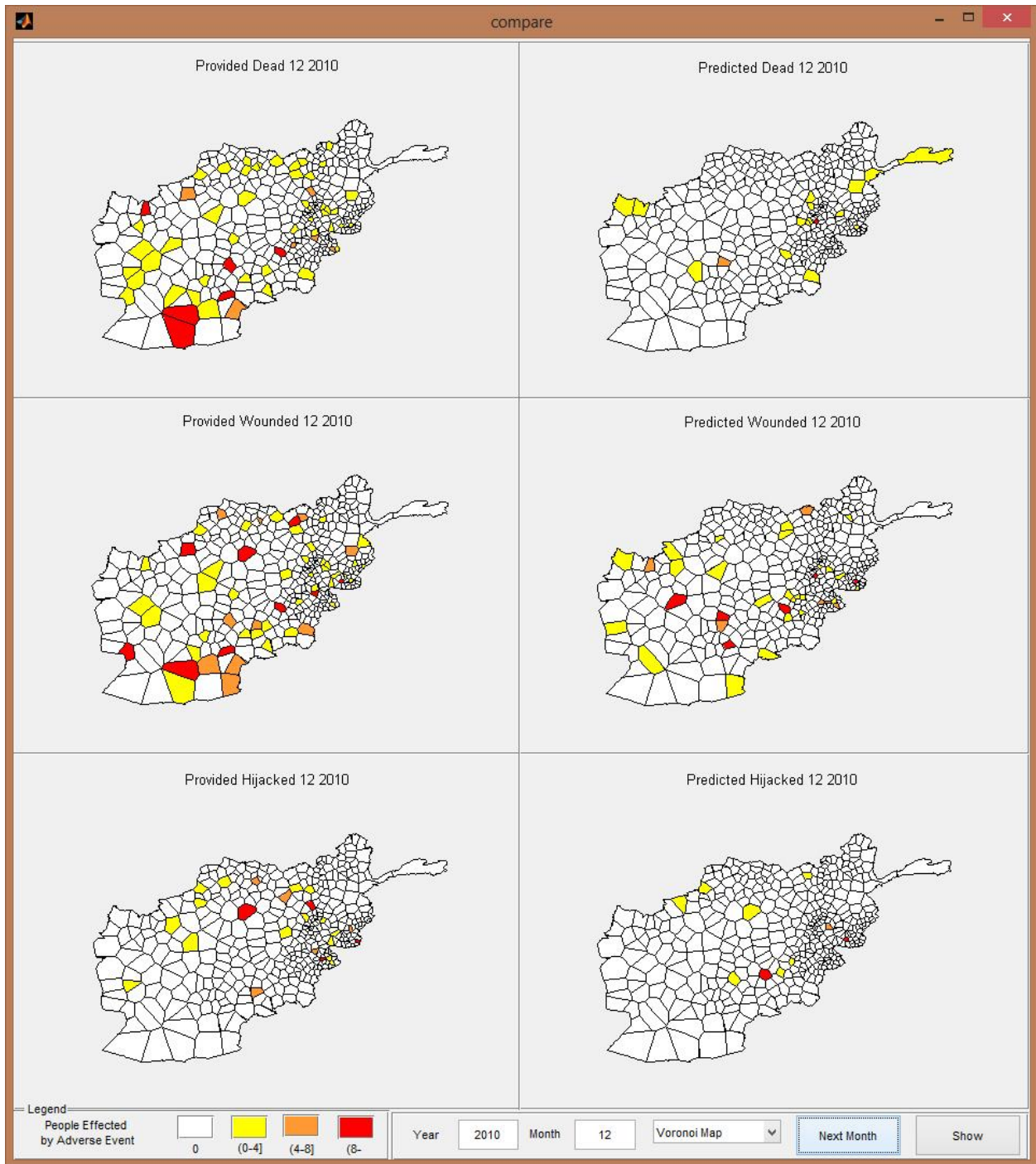
Figure 51 Adverse event comparison for December 2010
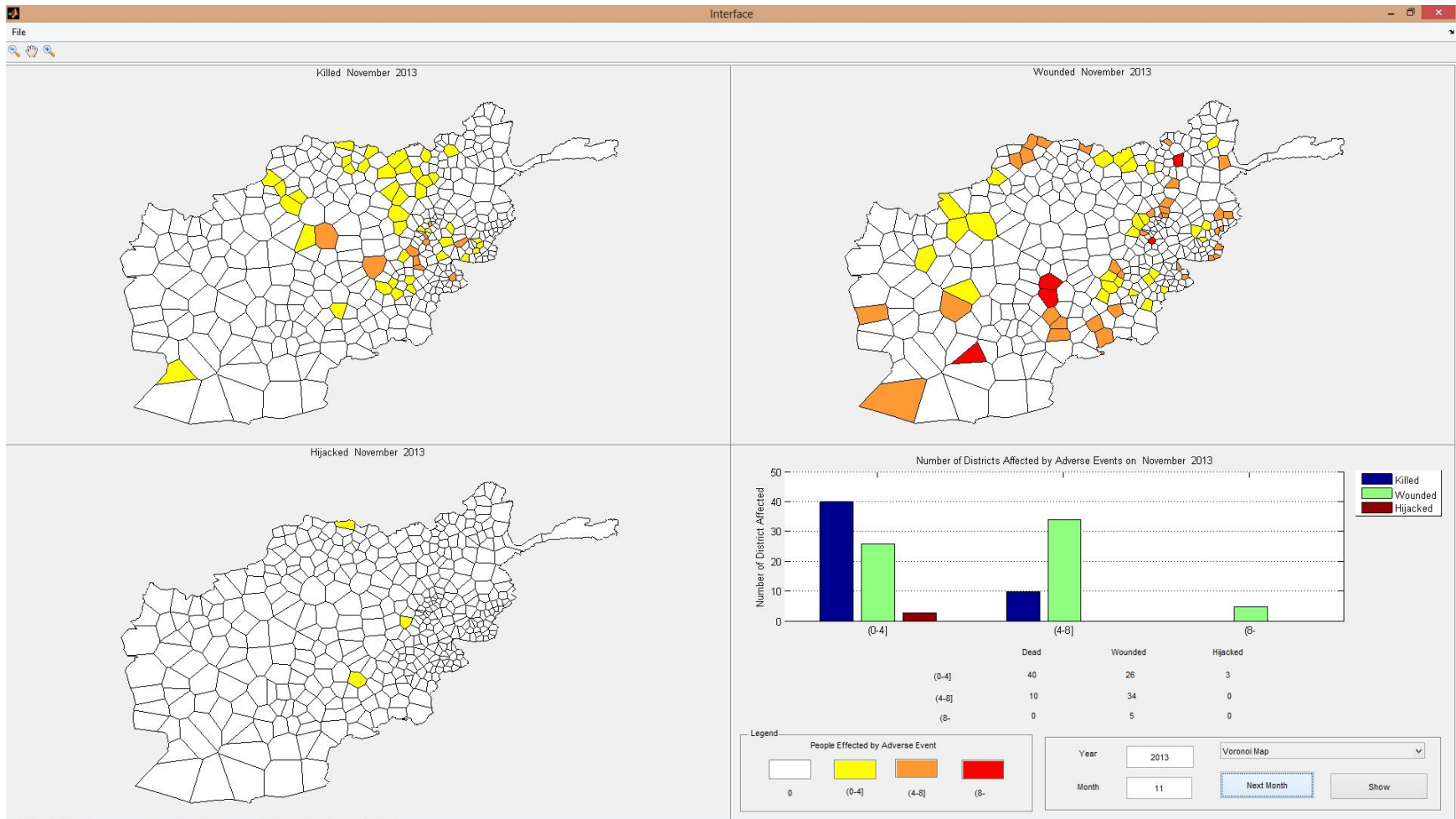
APPENDIX C
ADVERSE EVENT FORECASTS 2011 - 2014
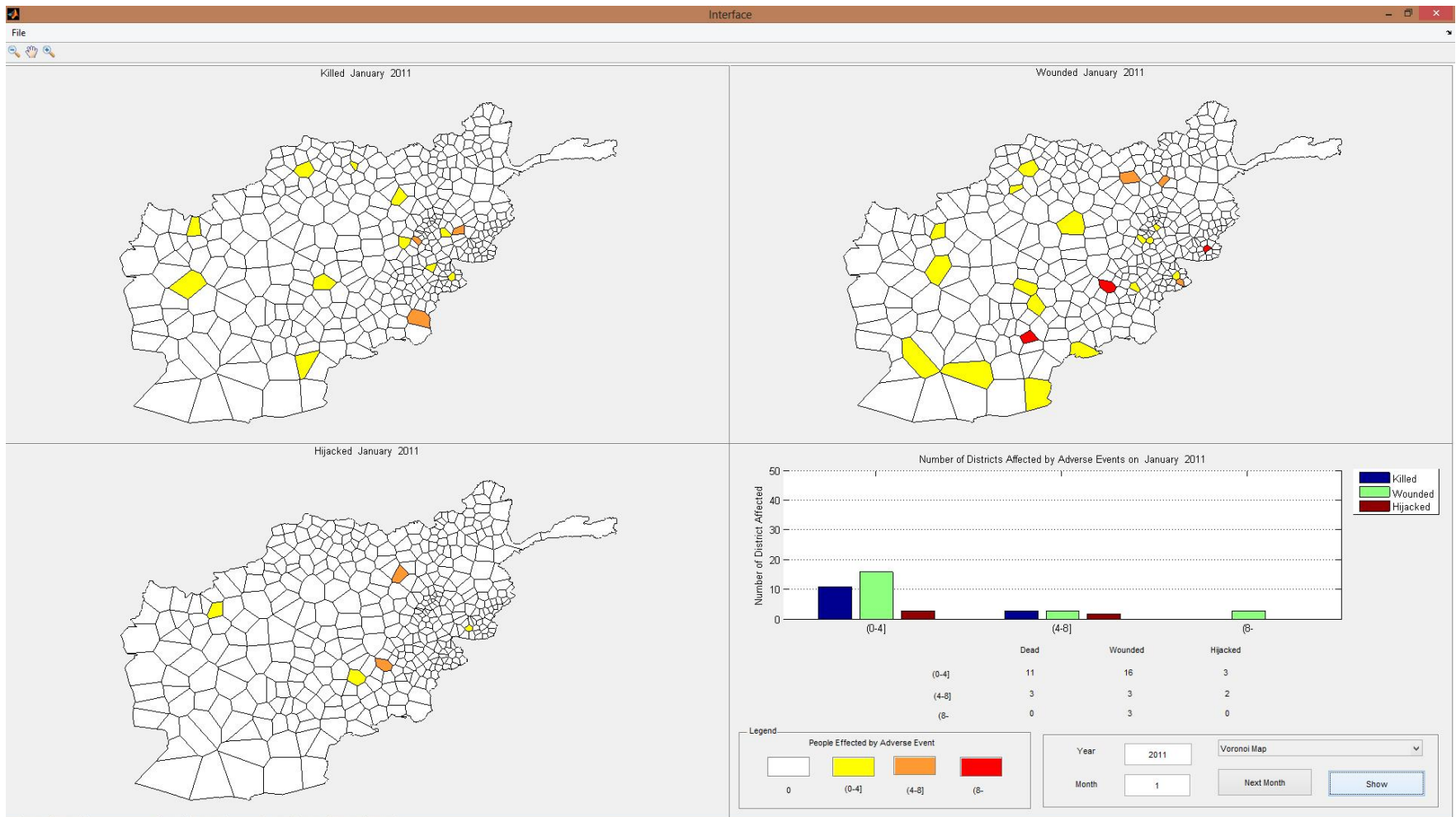
Figure 52 Adverse event forecasts for January 2011
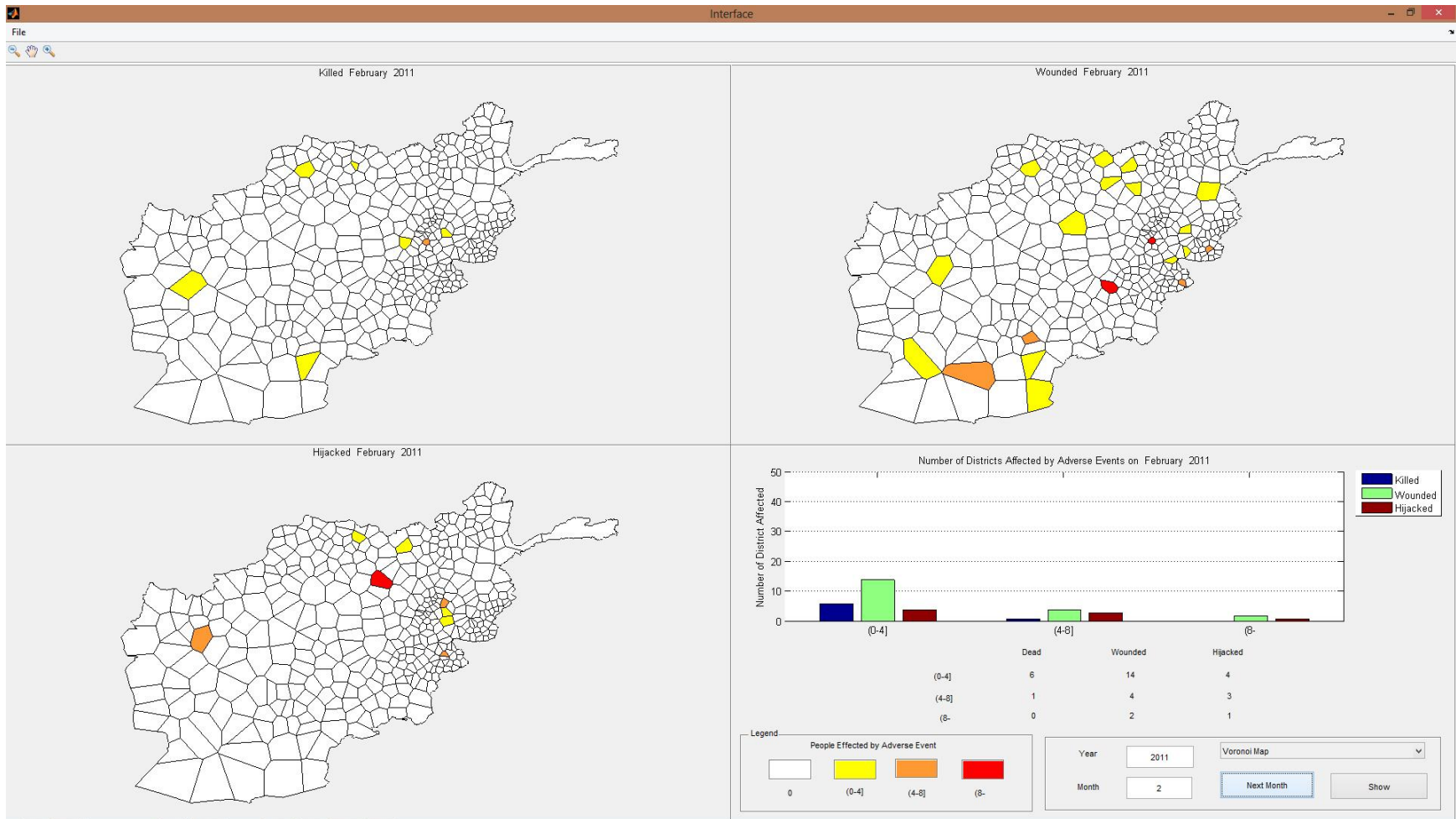
Figure 53 Adverse event forecasts for February 2011
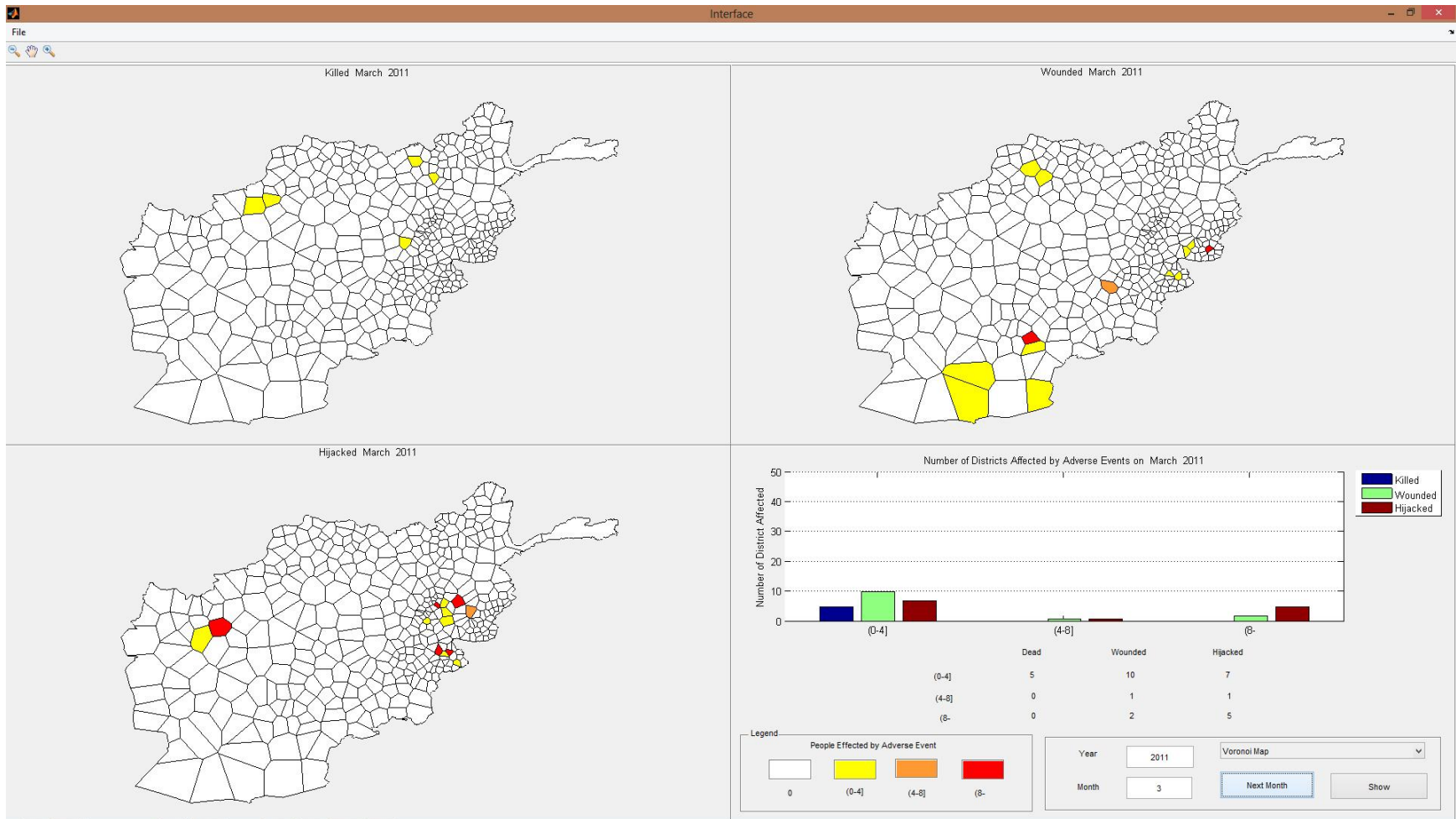
Figure 54 Adverse event forecasts for March 2011
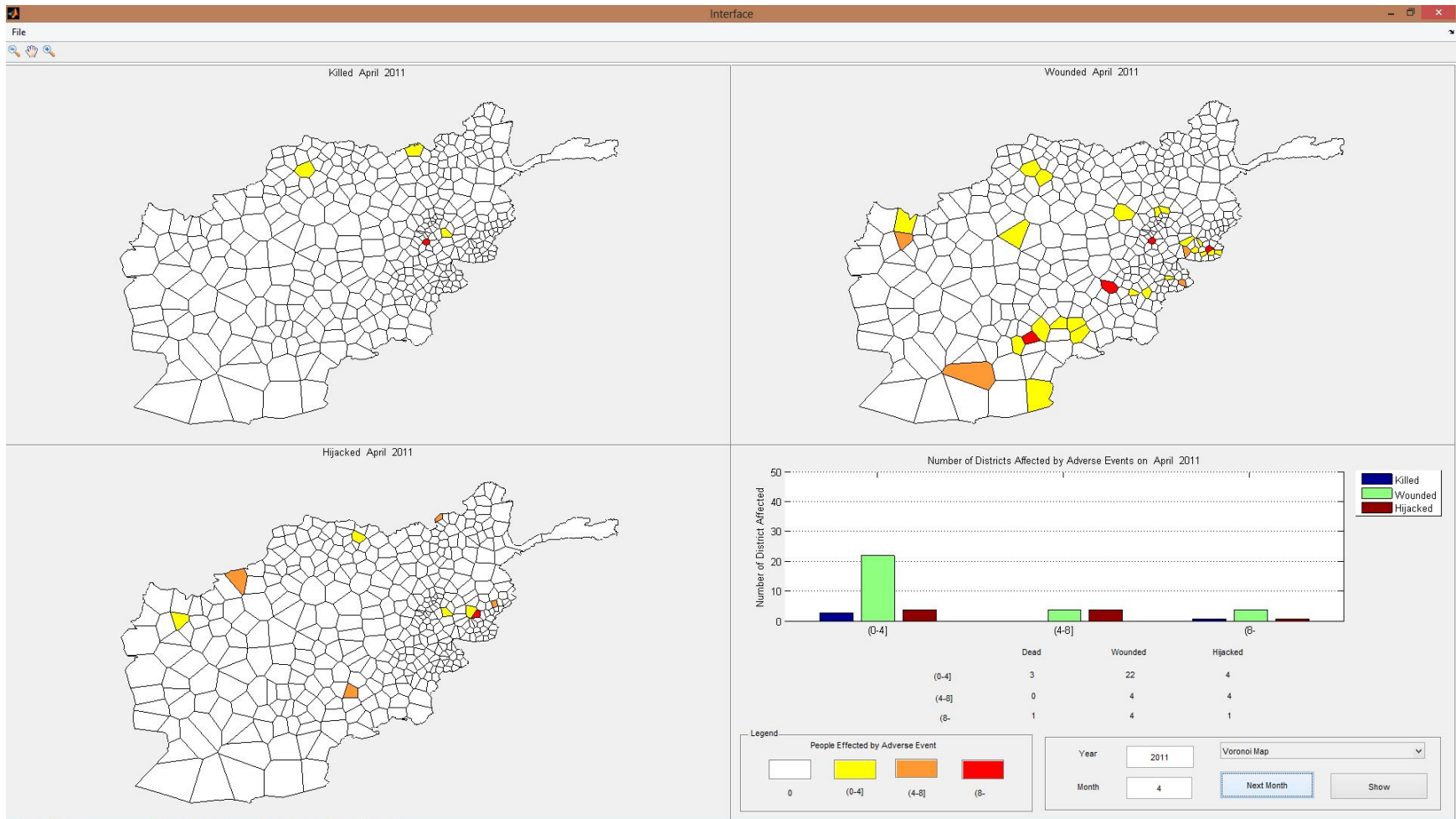
Figure 55 Adverse event forecasts for April 2011
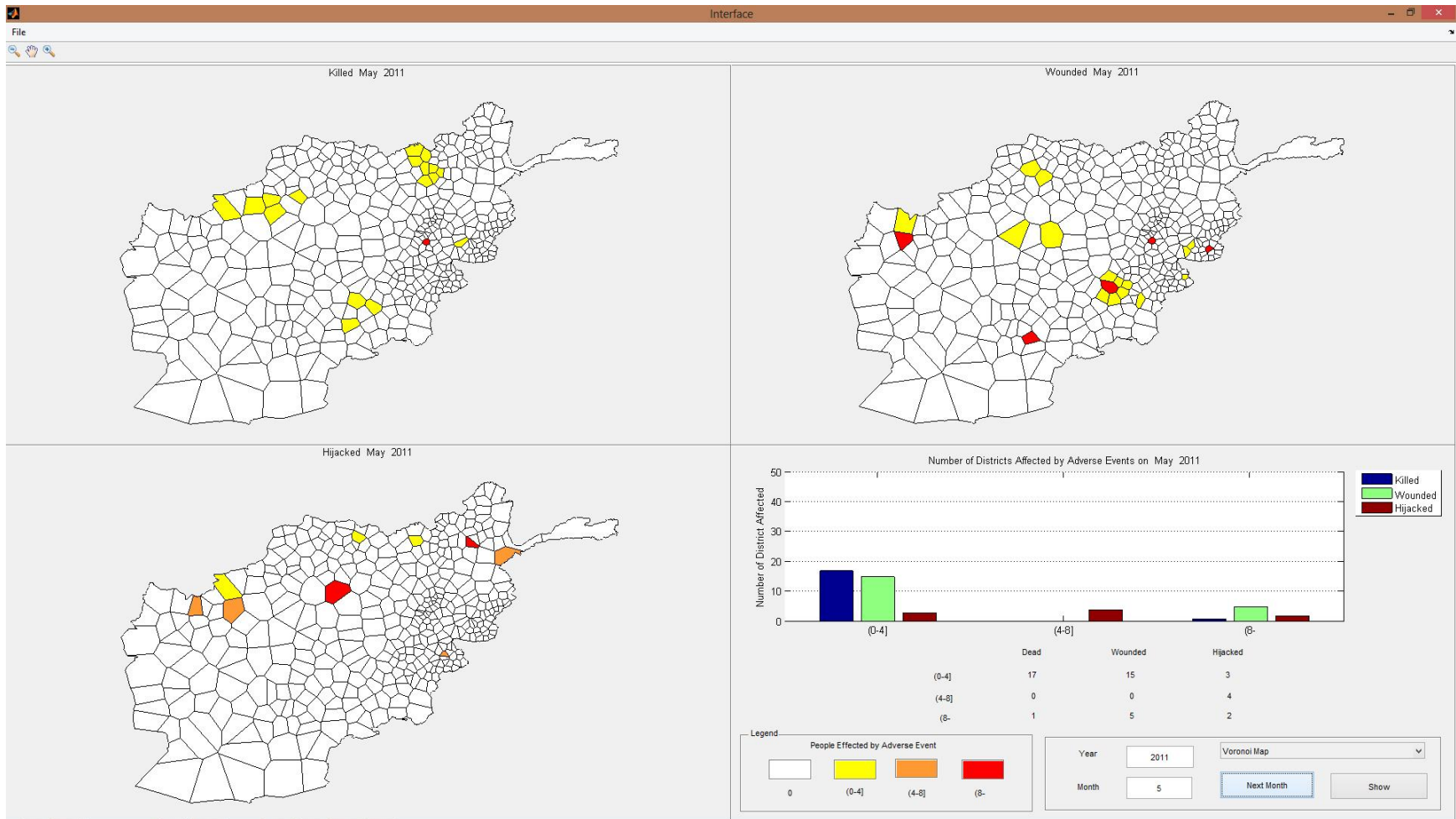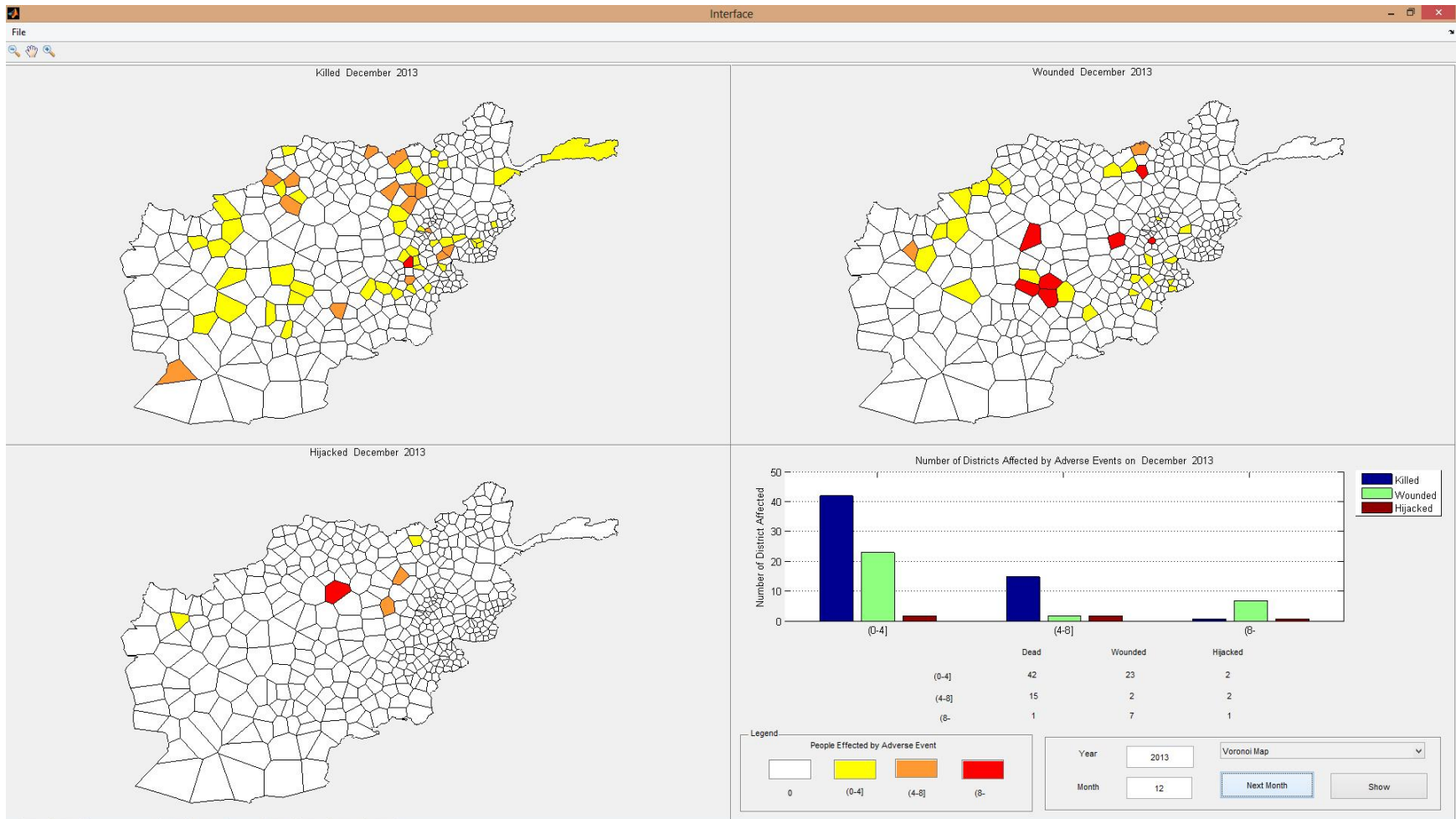
Figure 56 Adverse event forecasts for May 2011
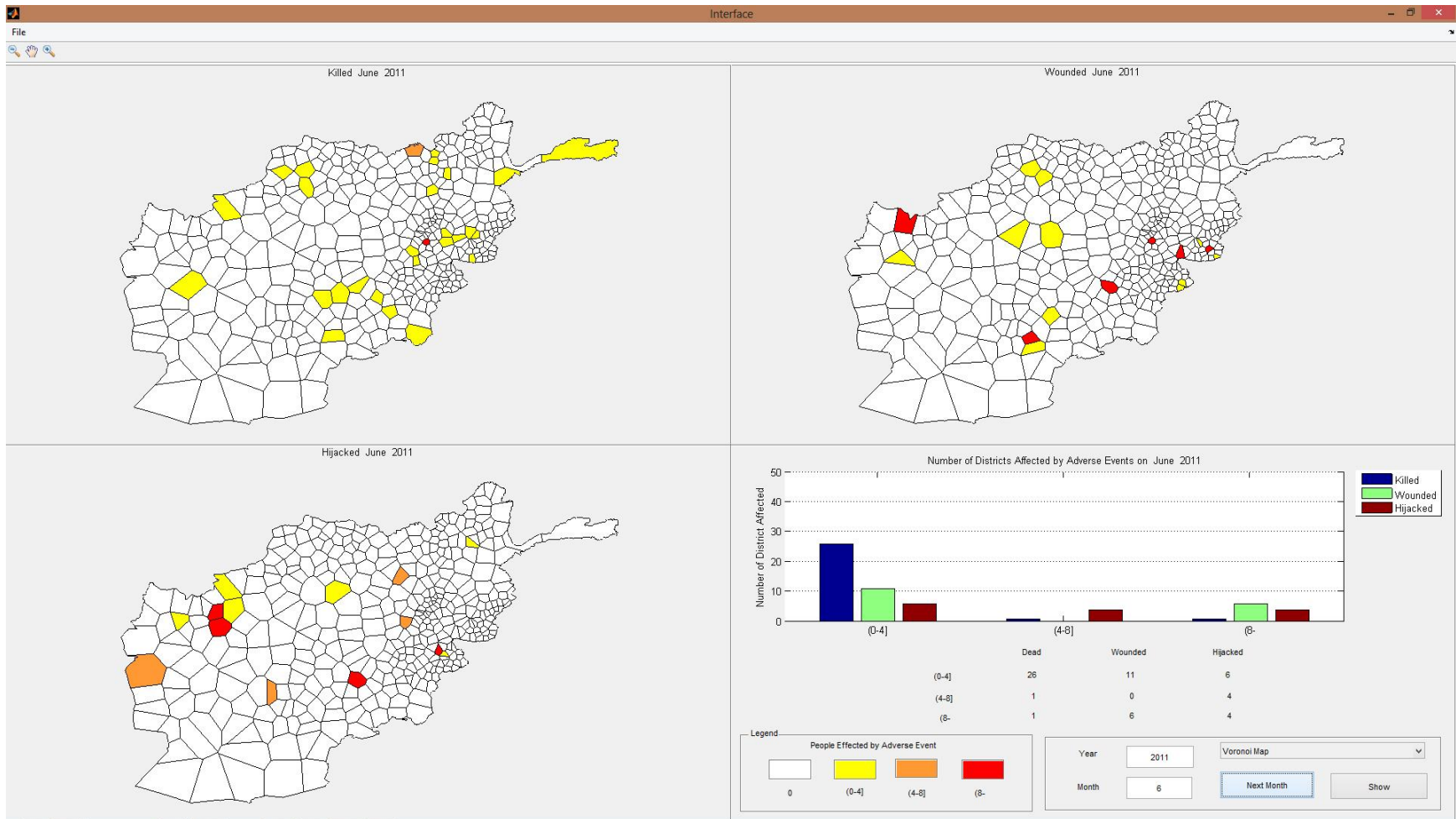
Figure 57 Adverse event forecasts for June 2011
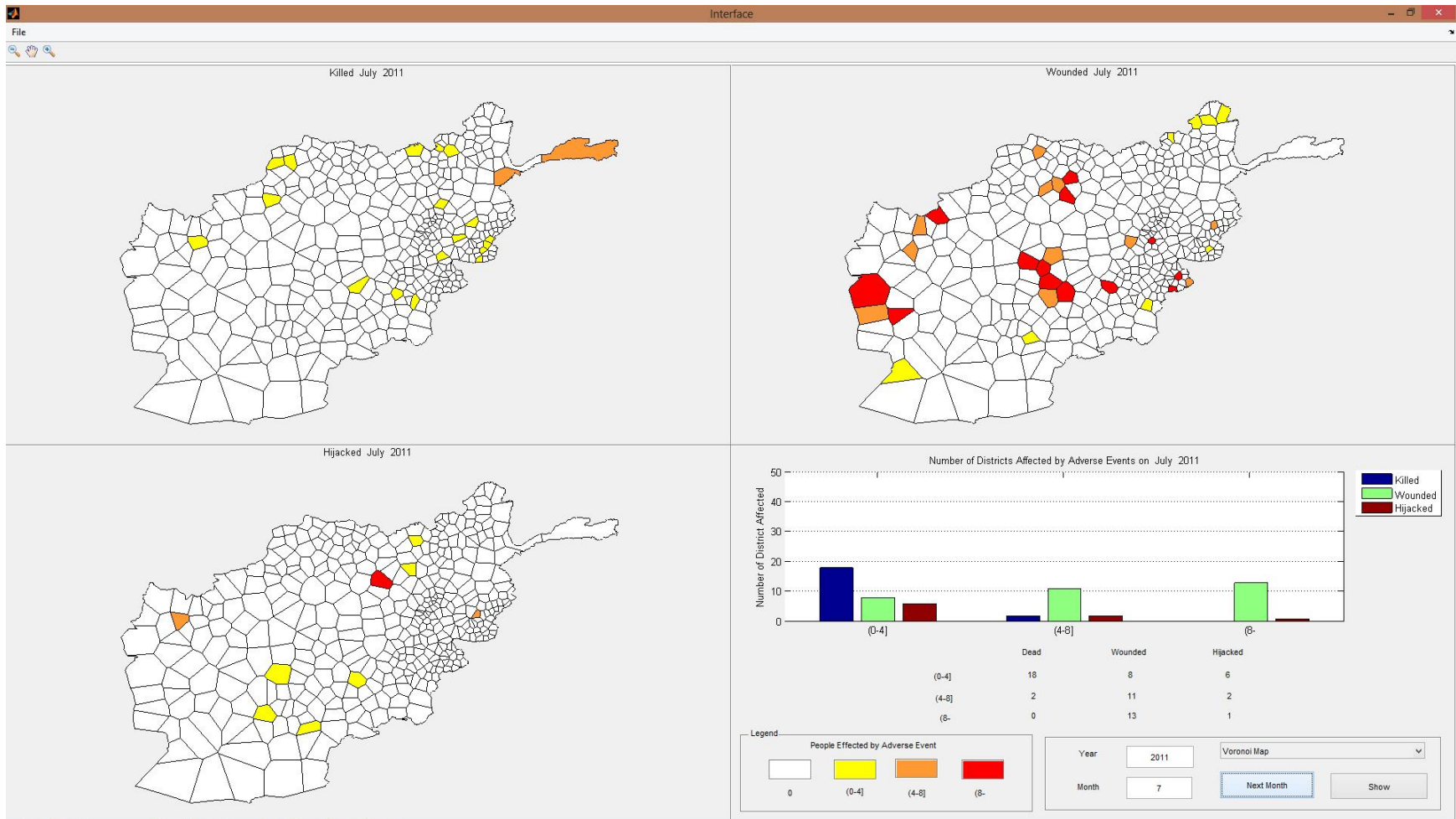
Figure 58 Adverse event forecasts for July 2011
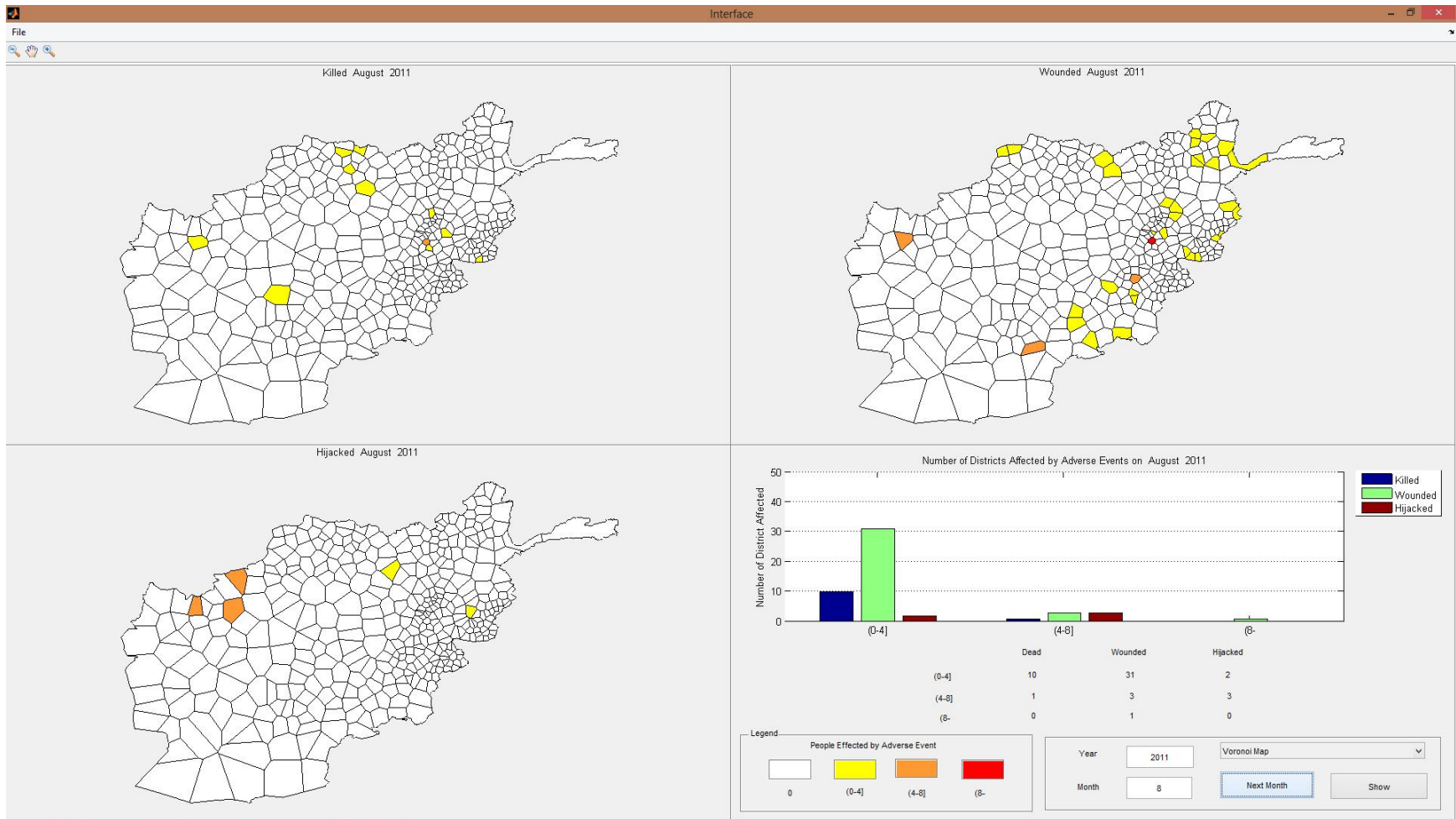
Figure 59 Adverse event forecasts for August 2011
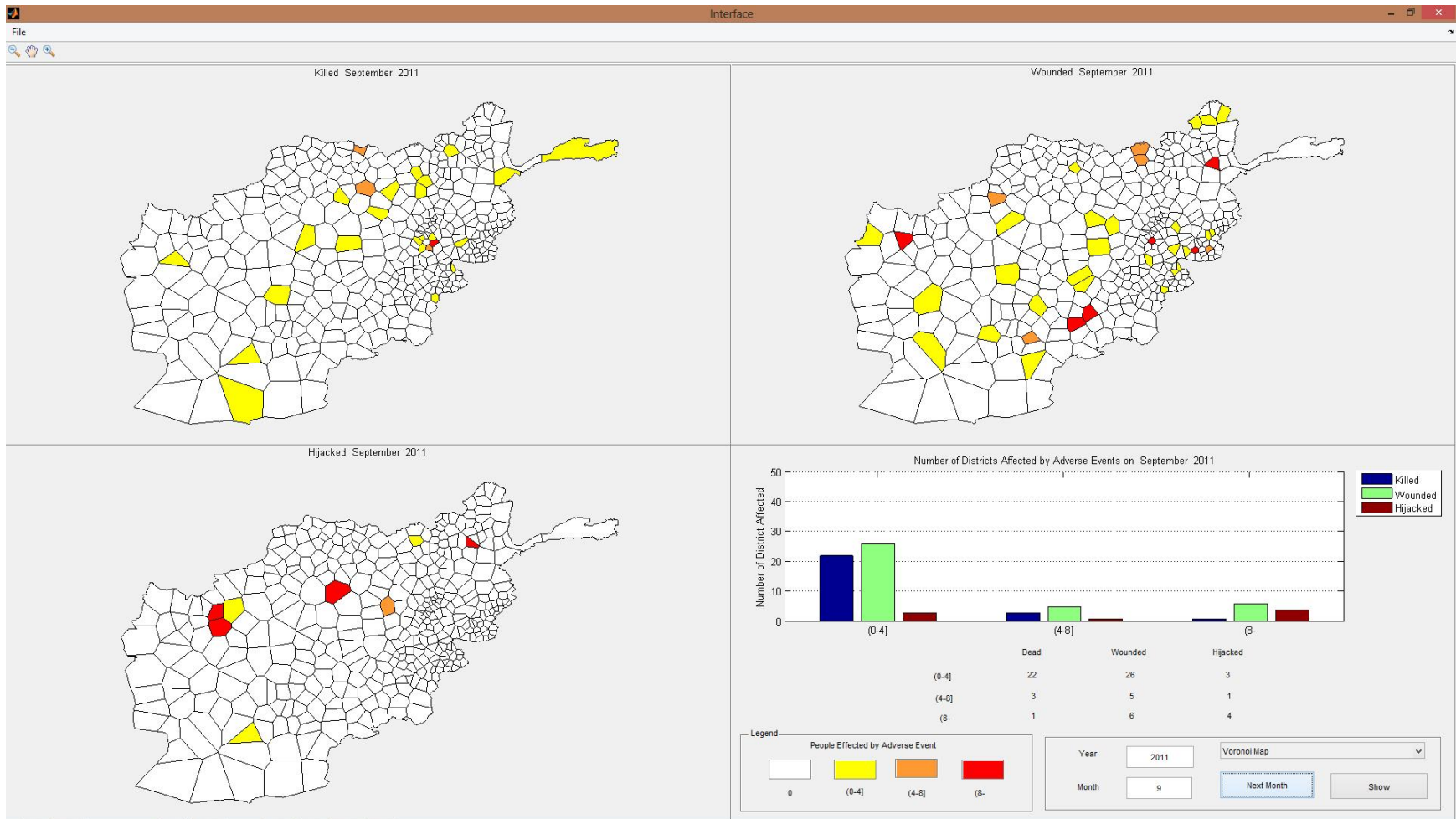
Figure 60 Adverse event forecasts for September 2011
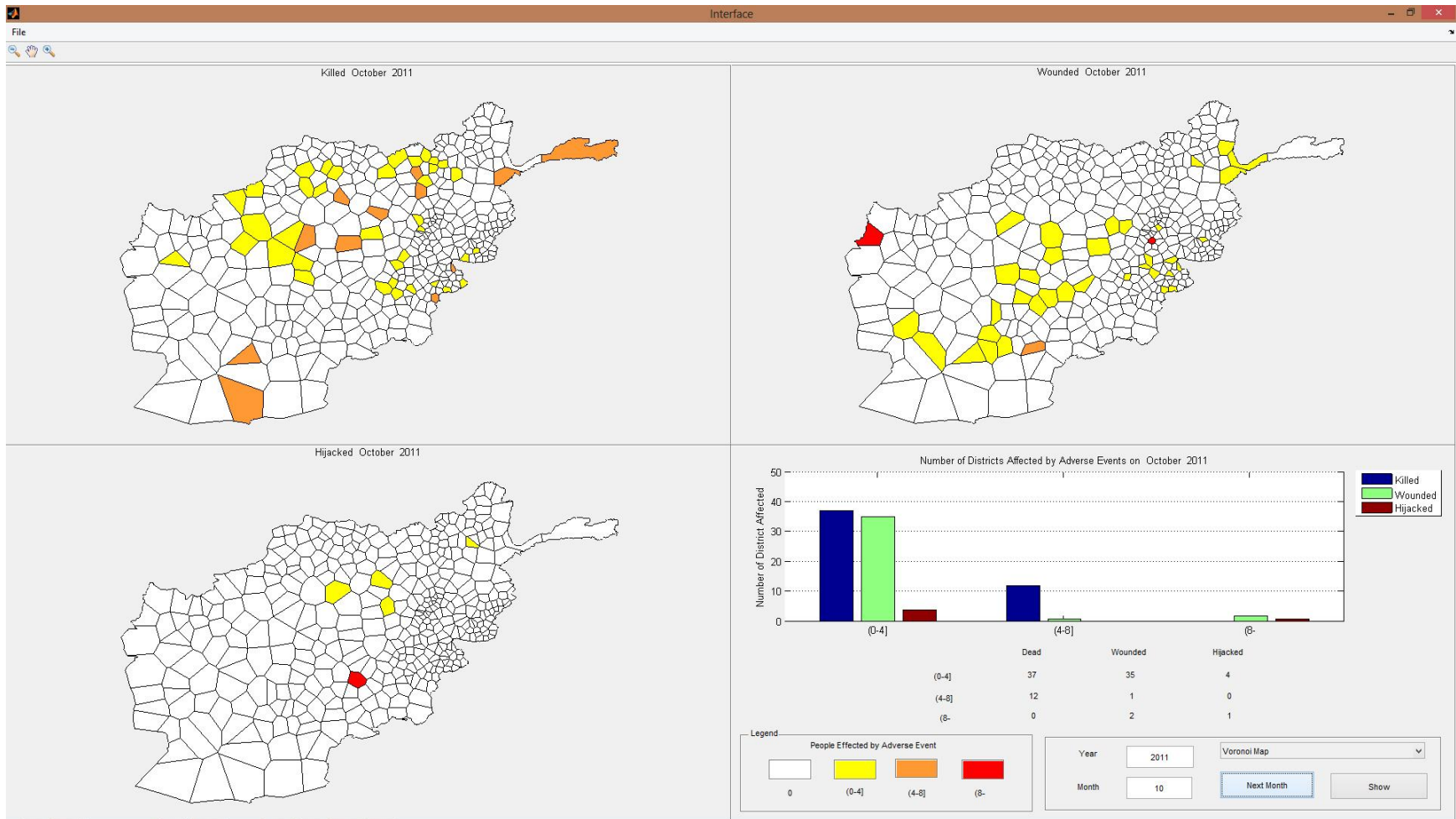
Figure 61 Adverse event forecasts for October 2011
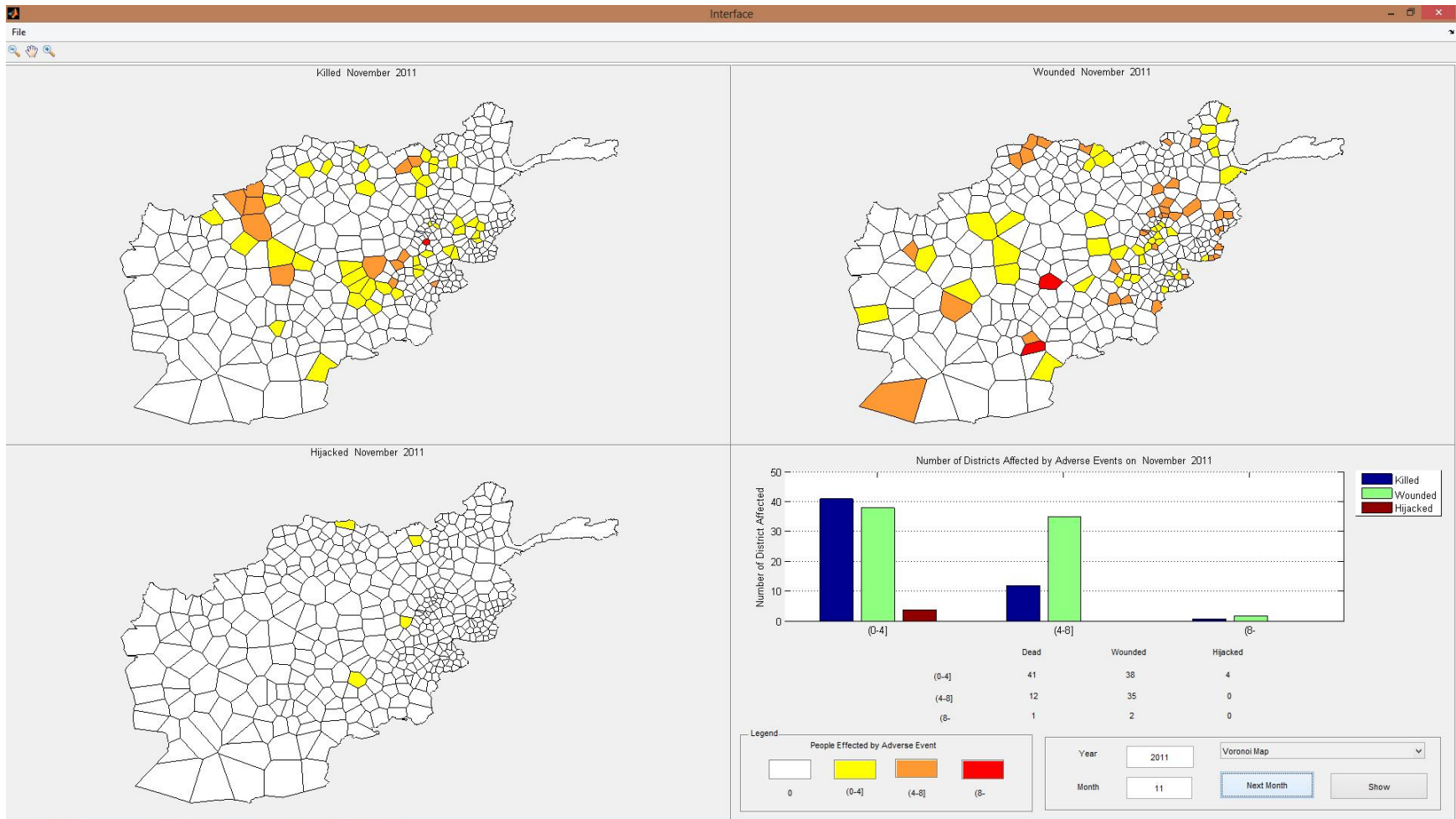
Figure 62 Adverse event forecasts for November 2011
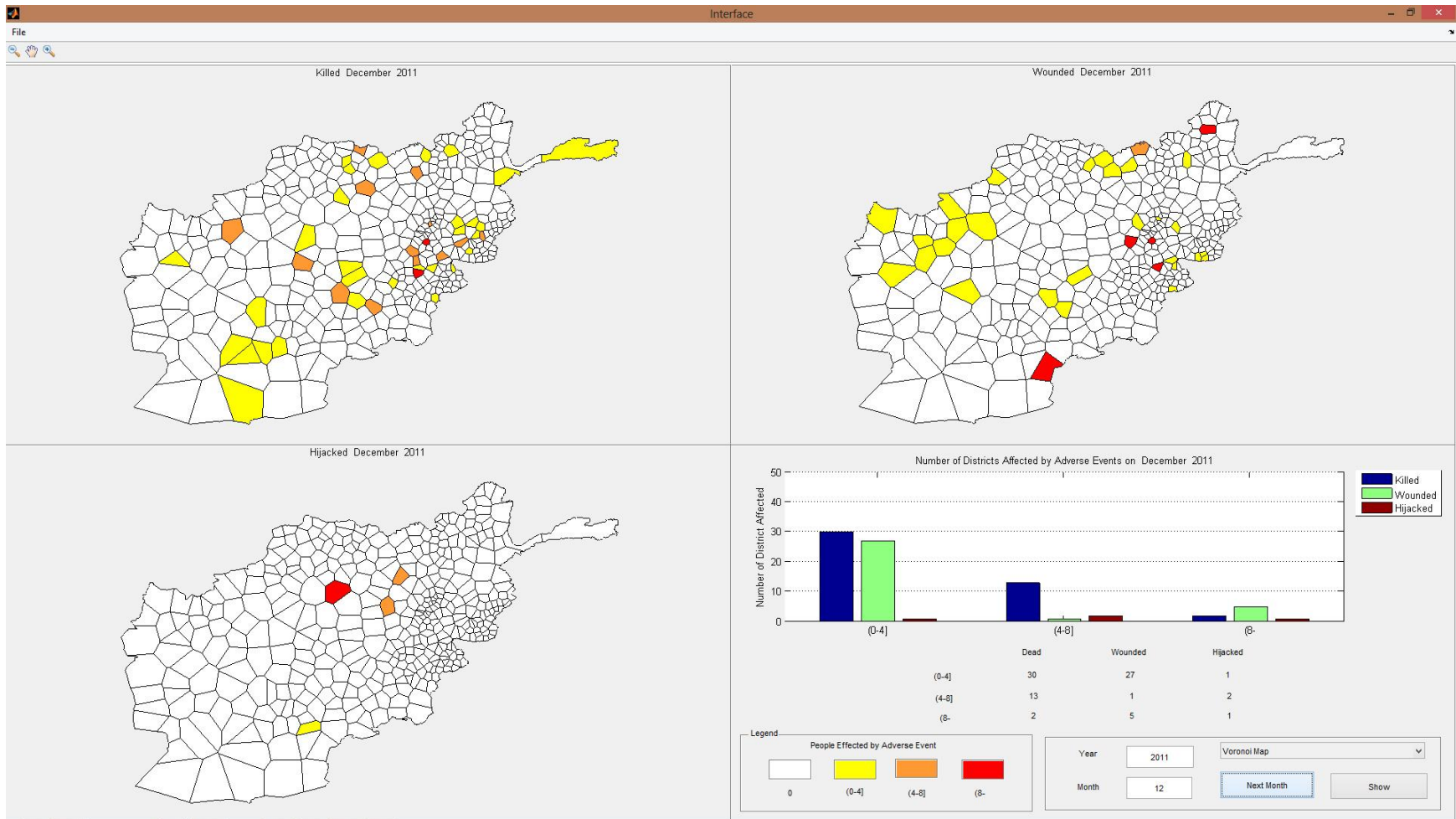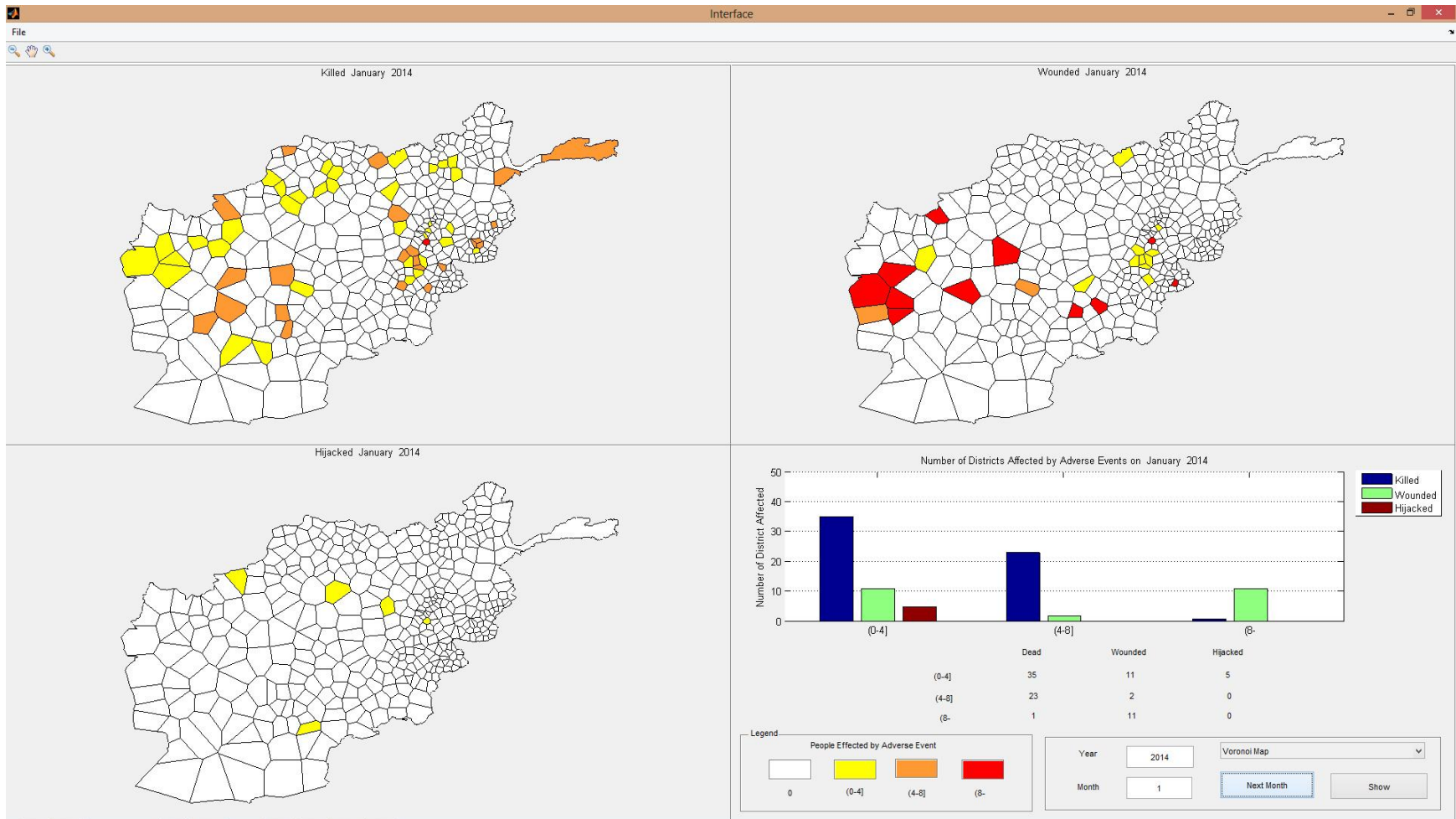
Figure 63 Adverse event forecasts for December 2011
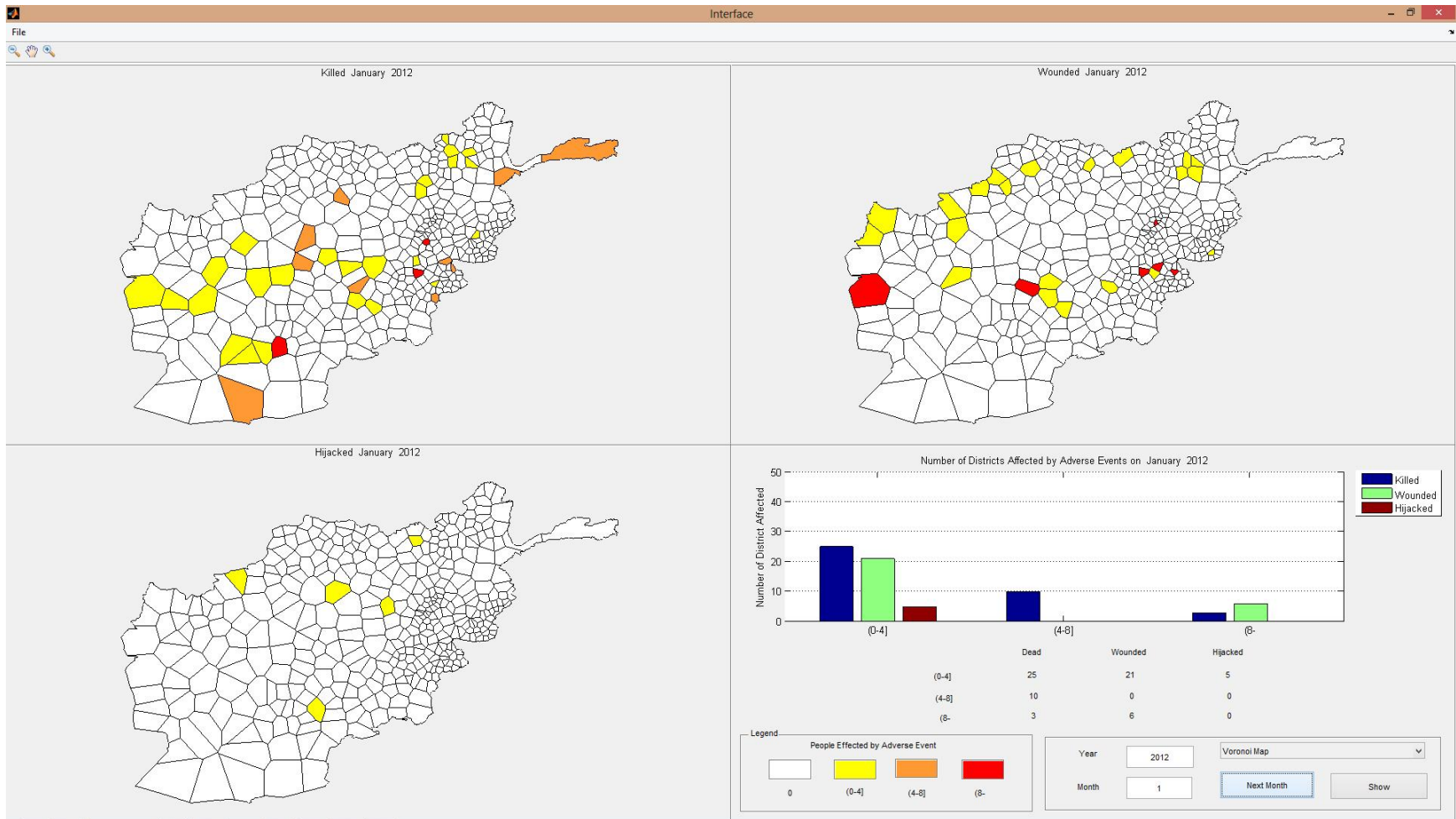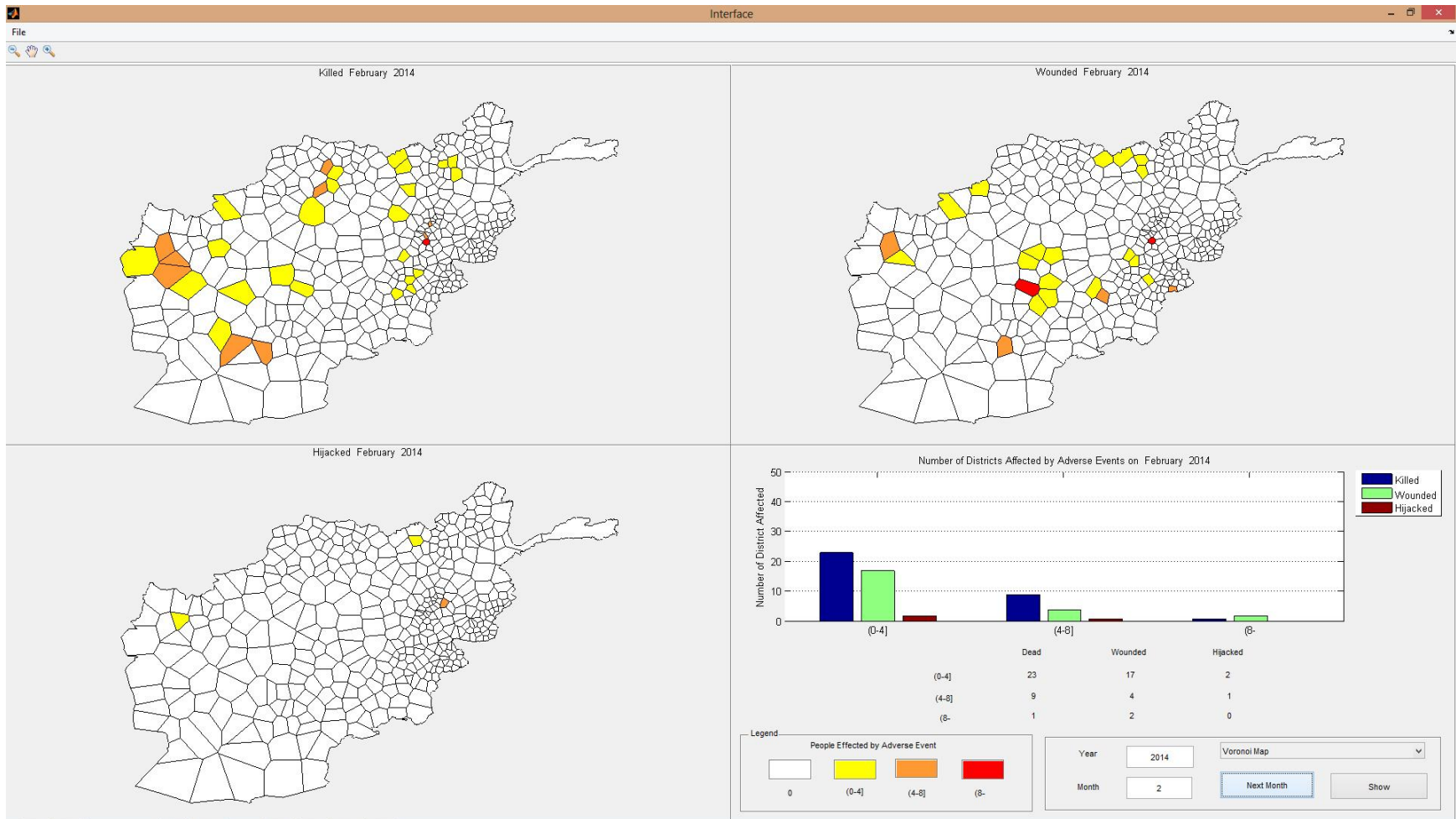
Figure 64 Adverse event forecasts for January 2012
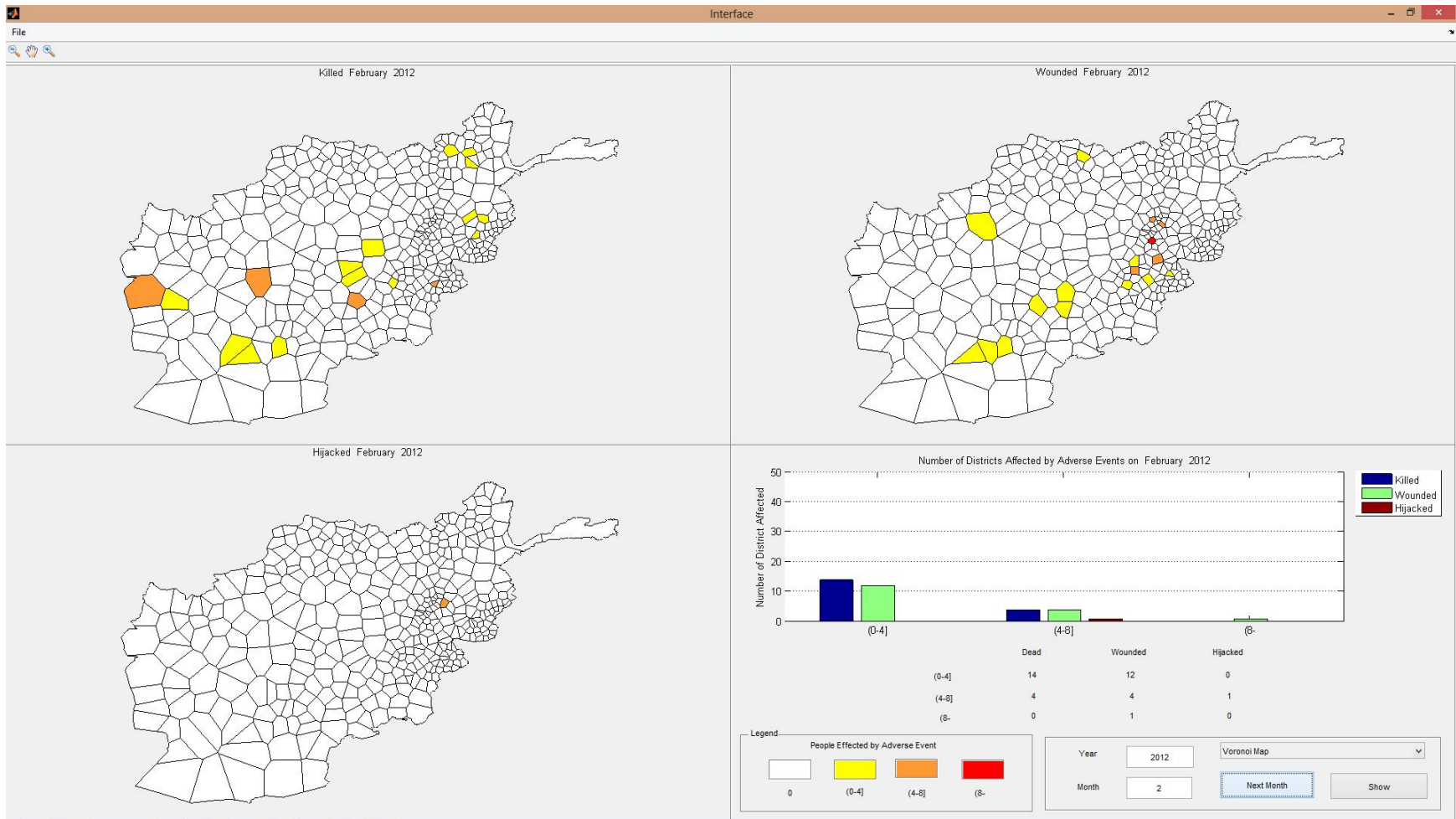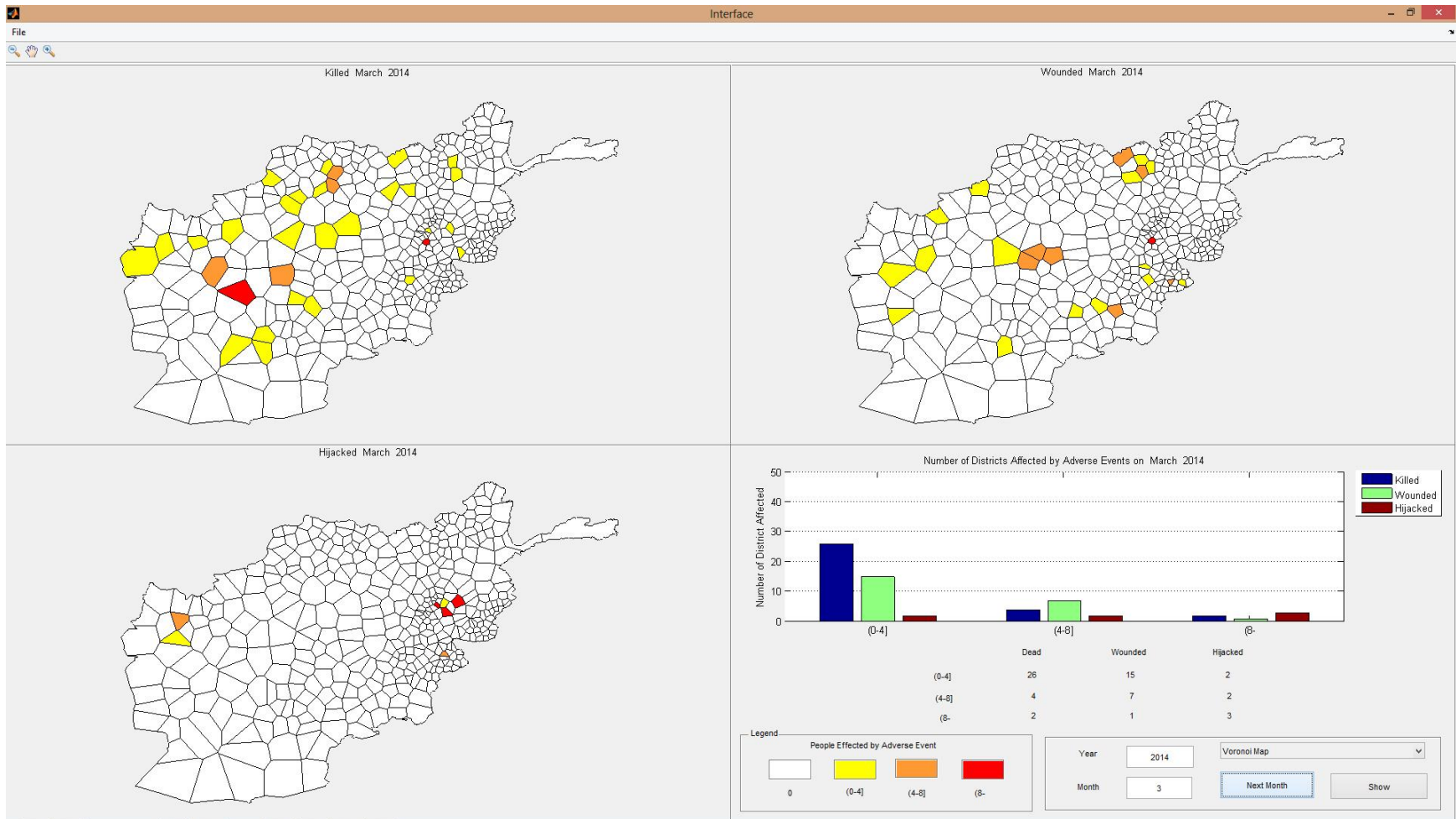
Figure 65 Adverse event forecasts for February 2012
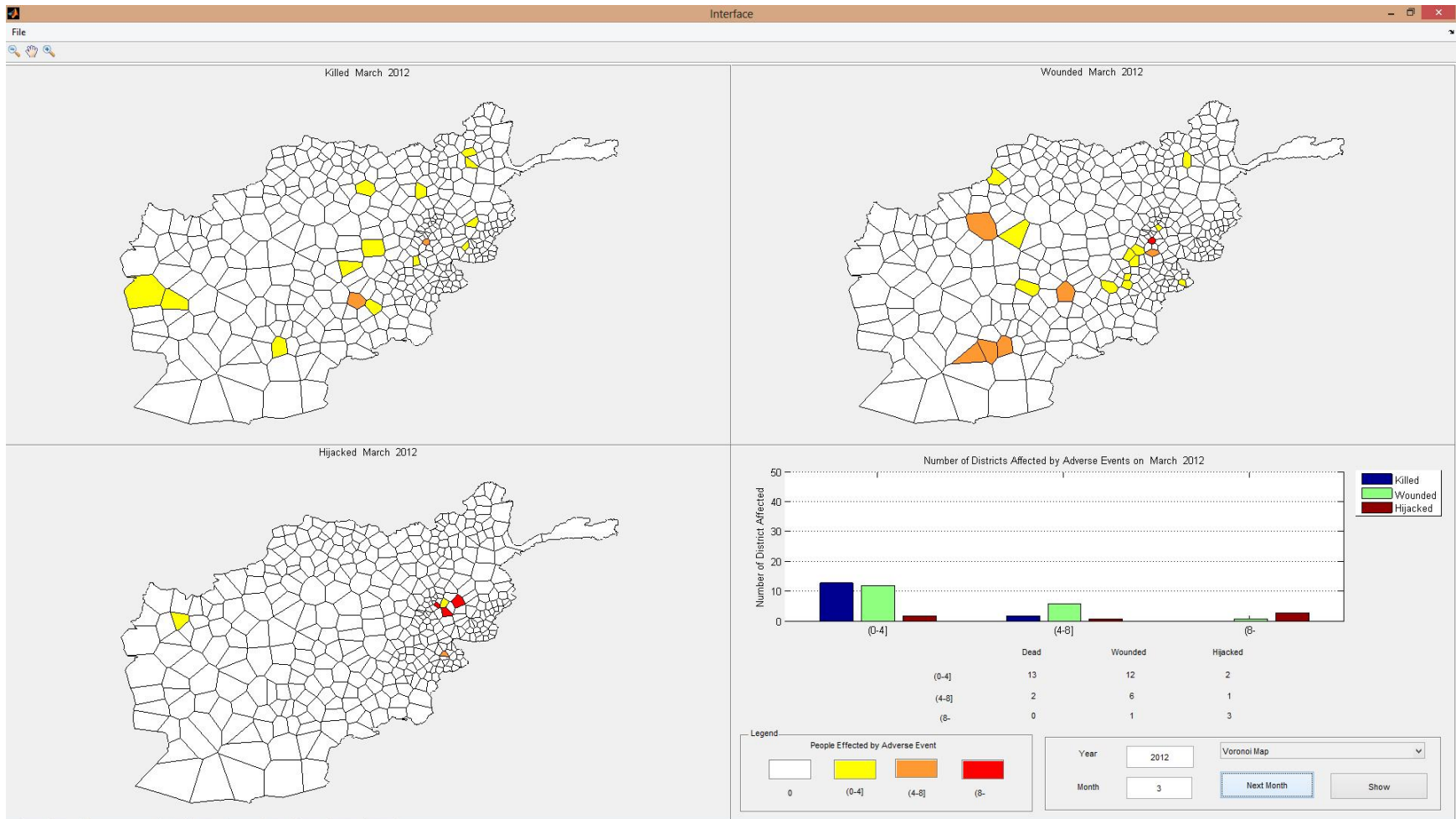
Figure 66 Adverse event forecasts for March 2012
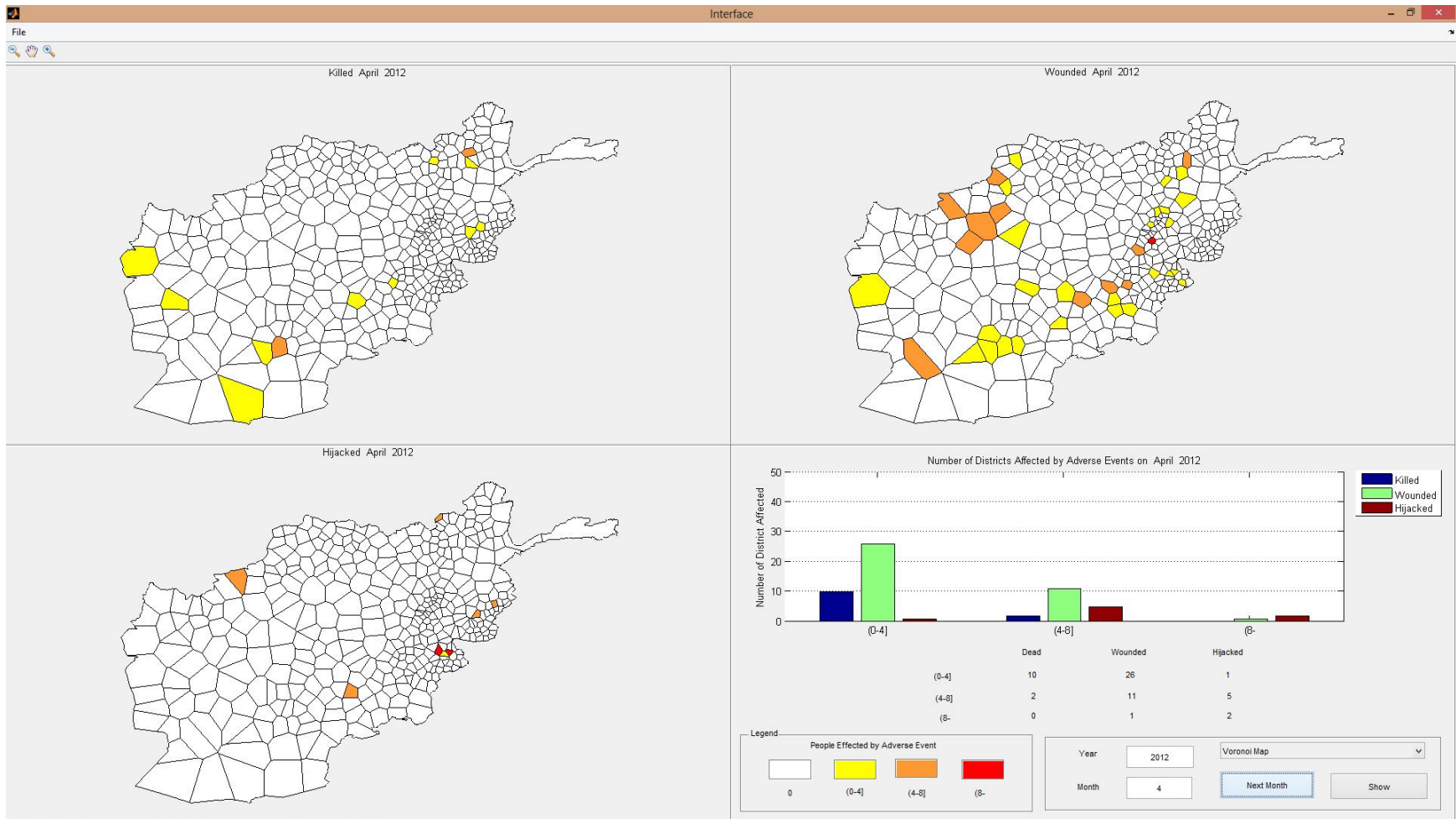
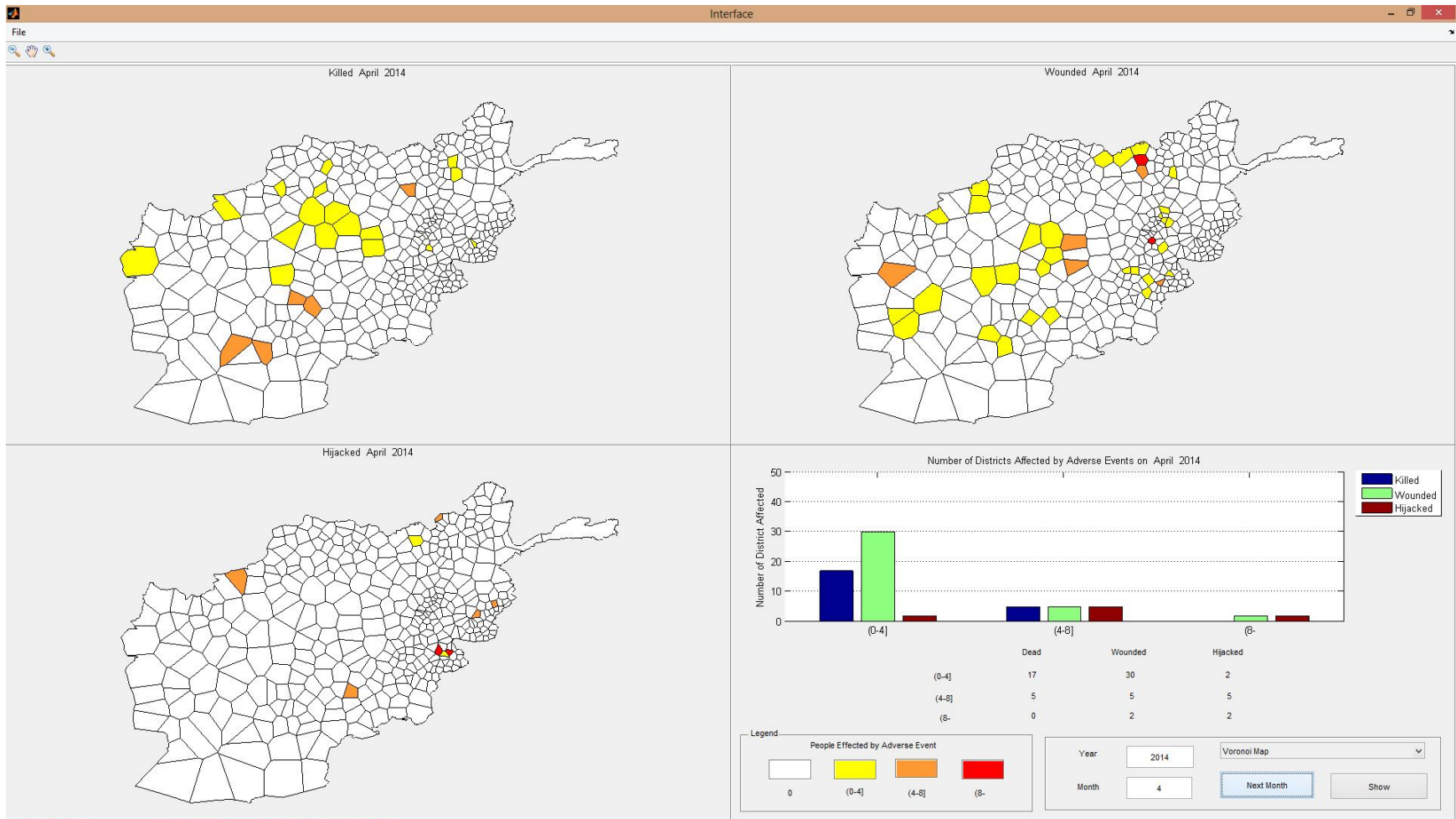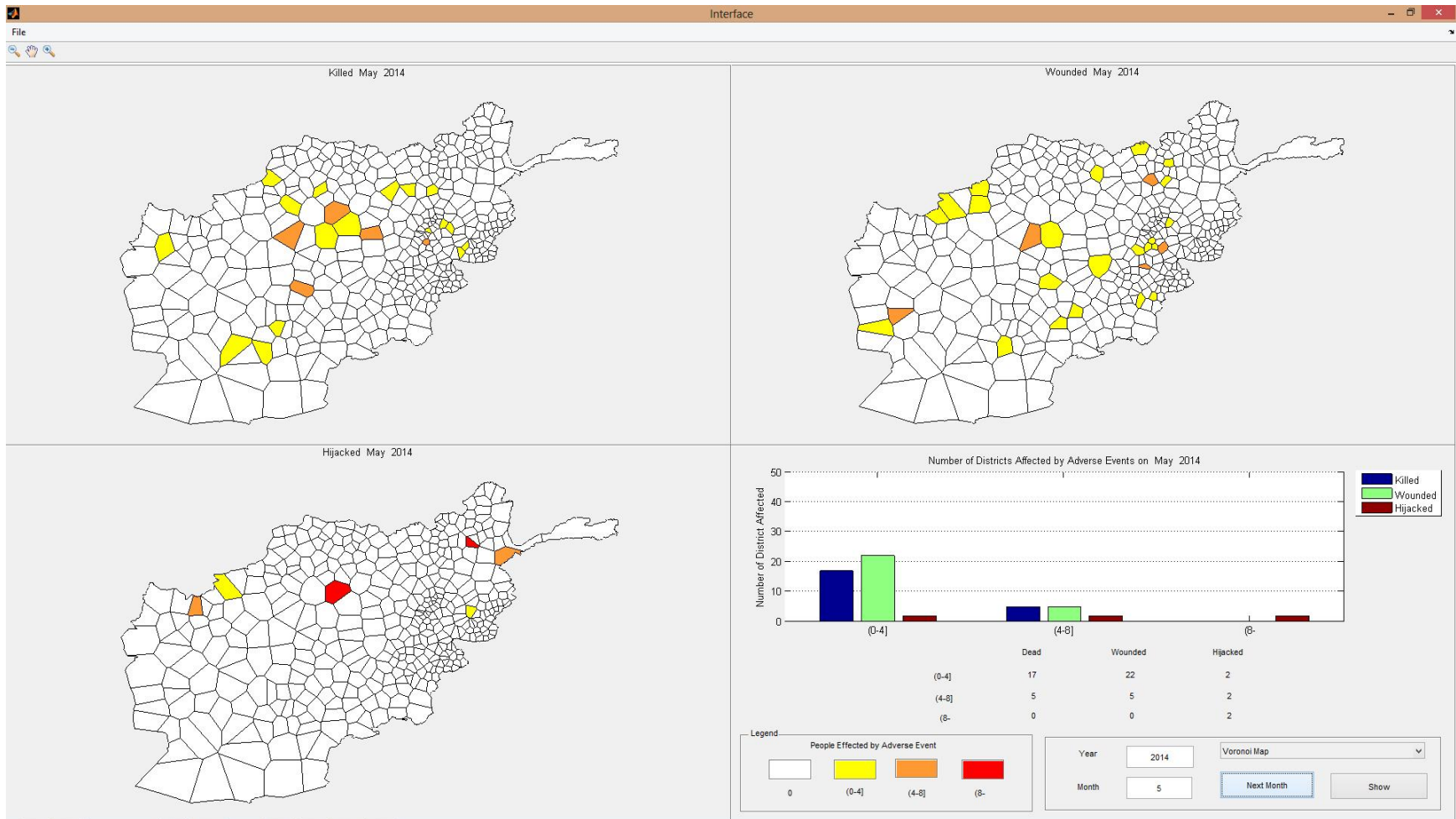Figure 67 Adverse event forecasts for April 2012
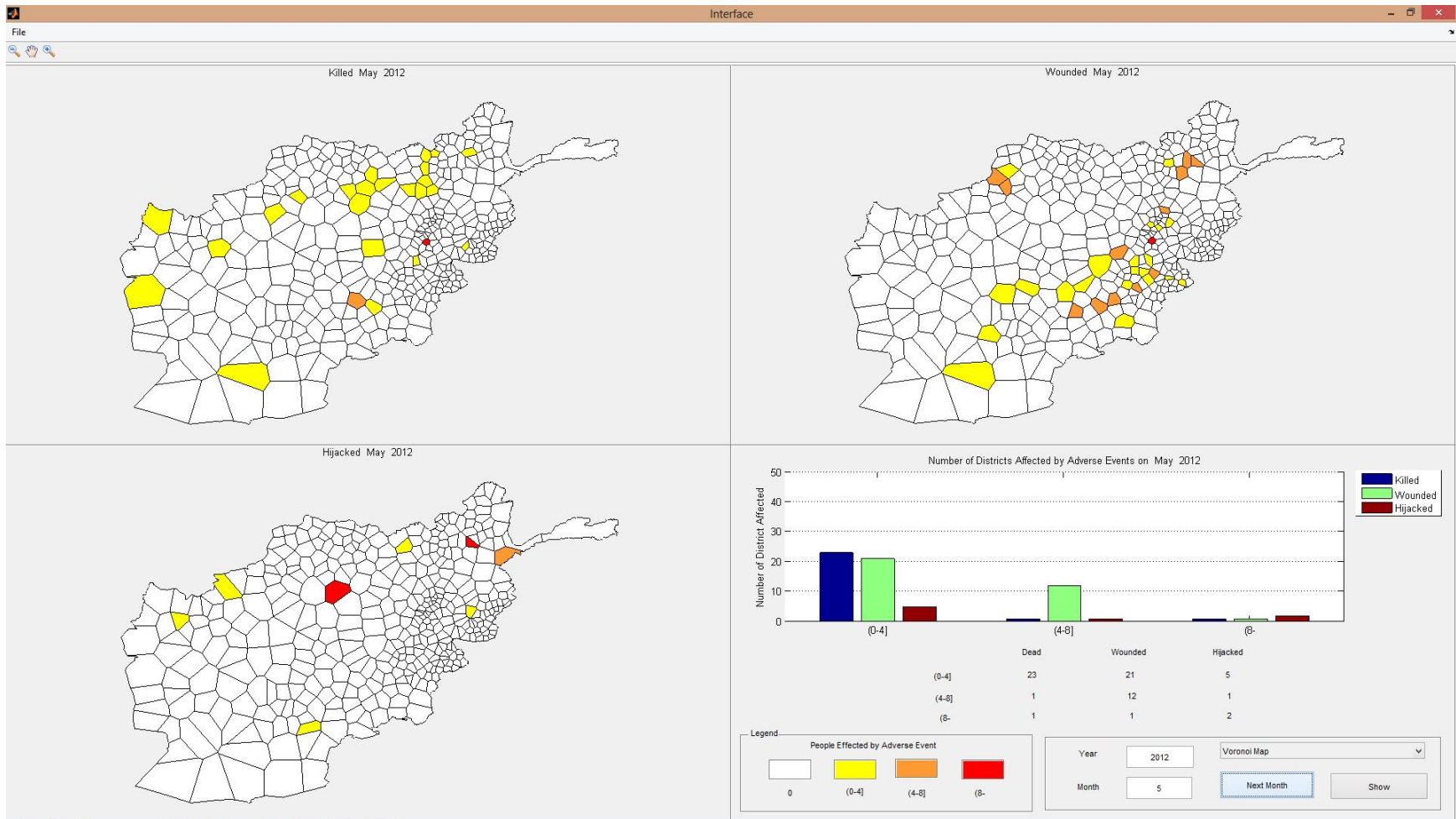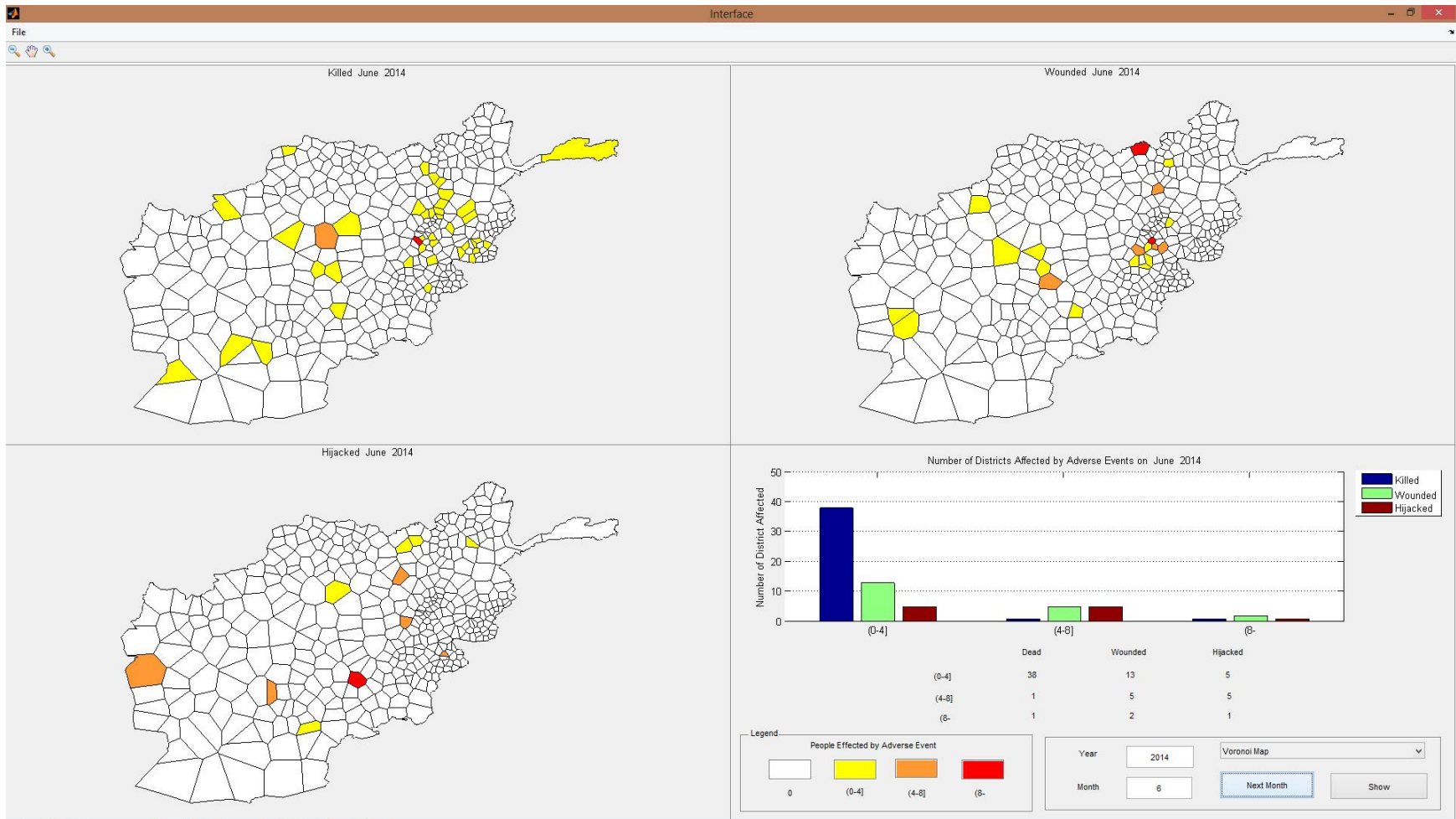
Figure 68 Adverse event forecasts for May 2012
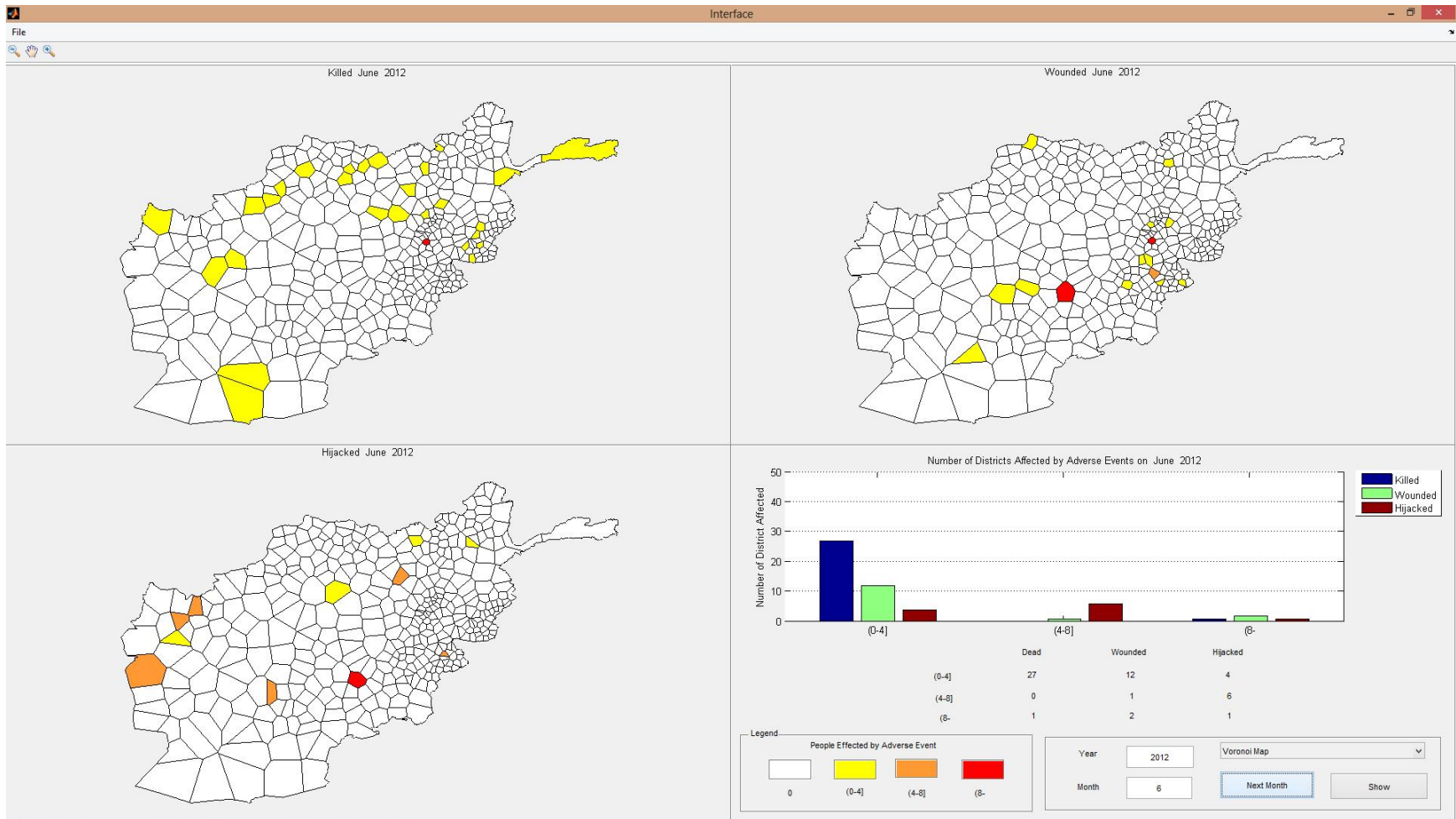
Figure 69 Adverse event forecasts for June 2012
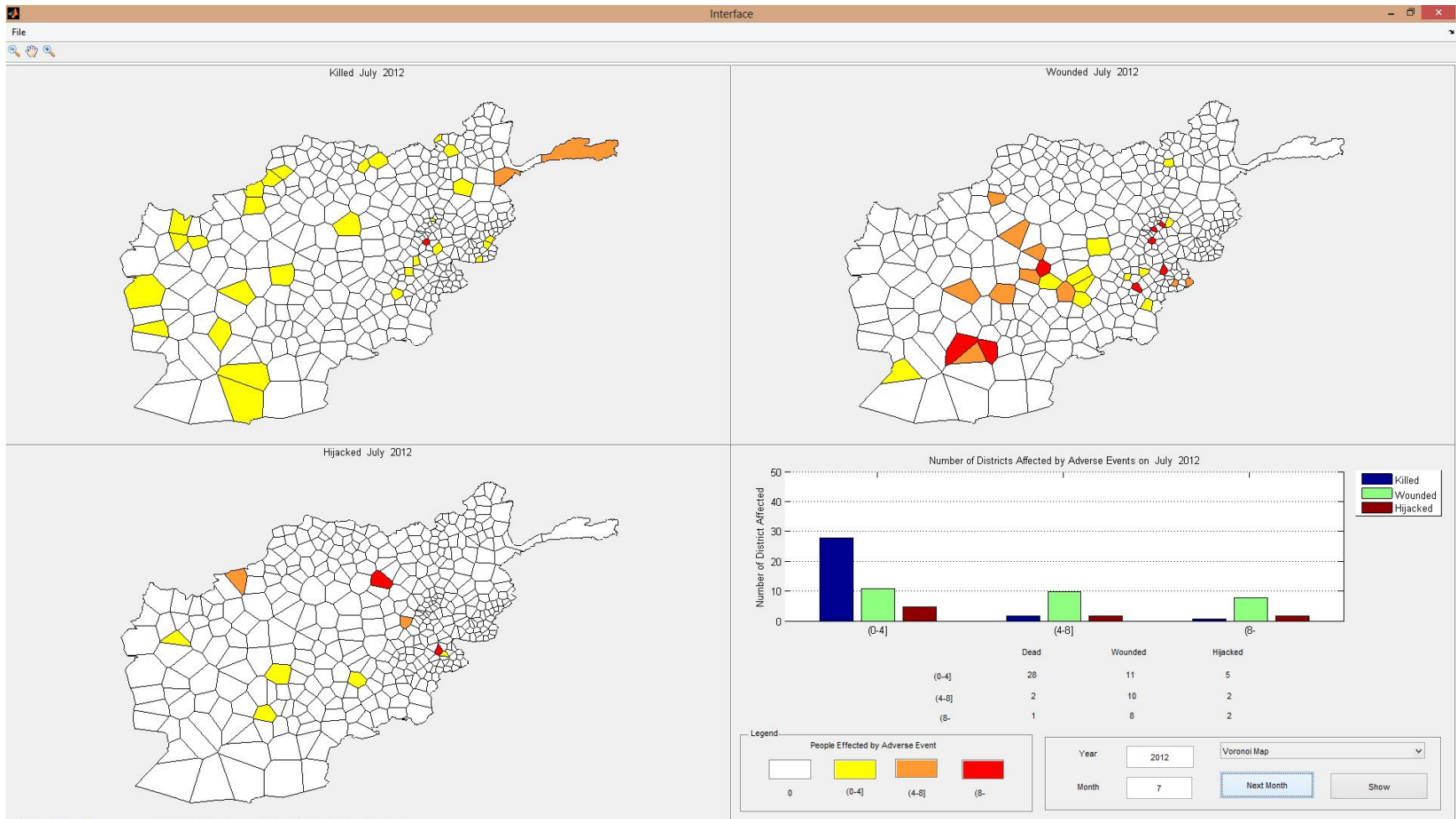
Figure 70 Adverse event forecasts for July 2012
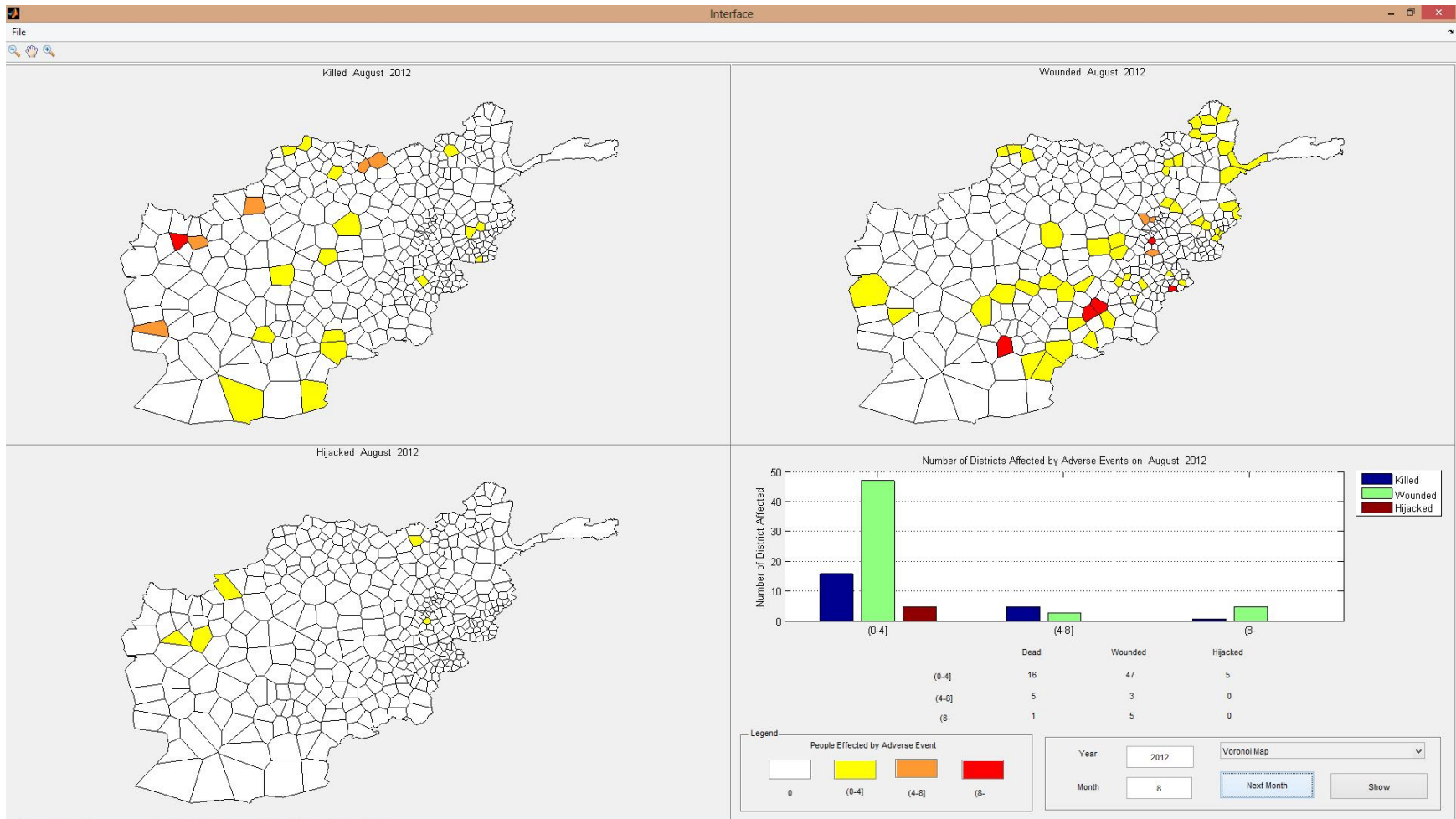
Figure 71 Adverse event forecasts for August 2012
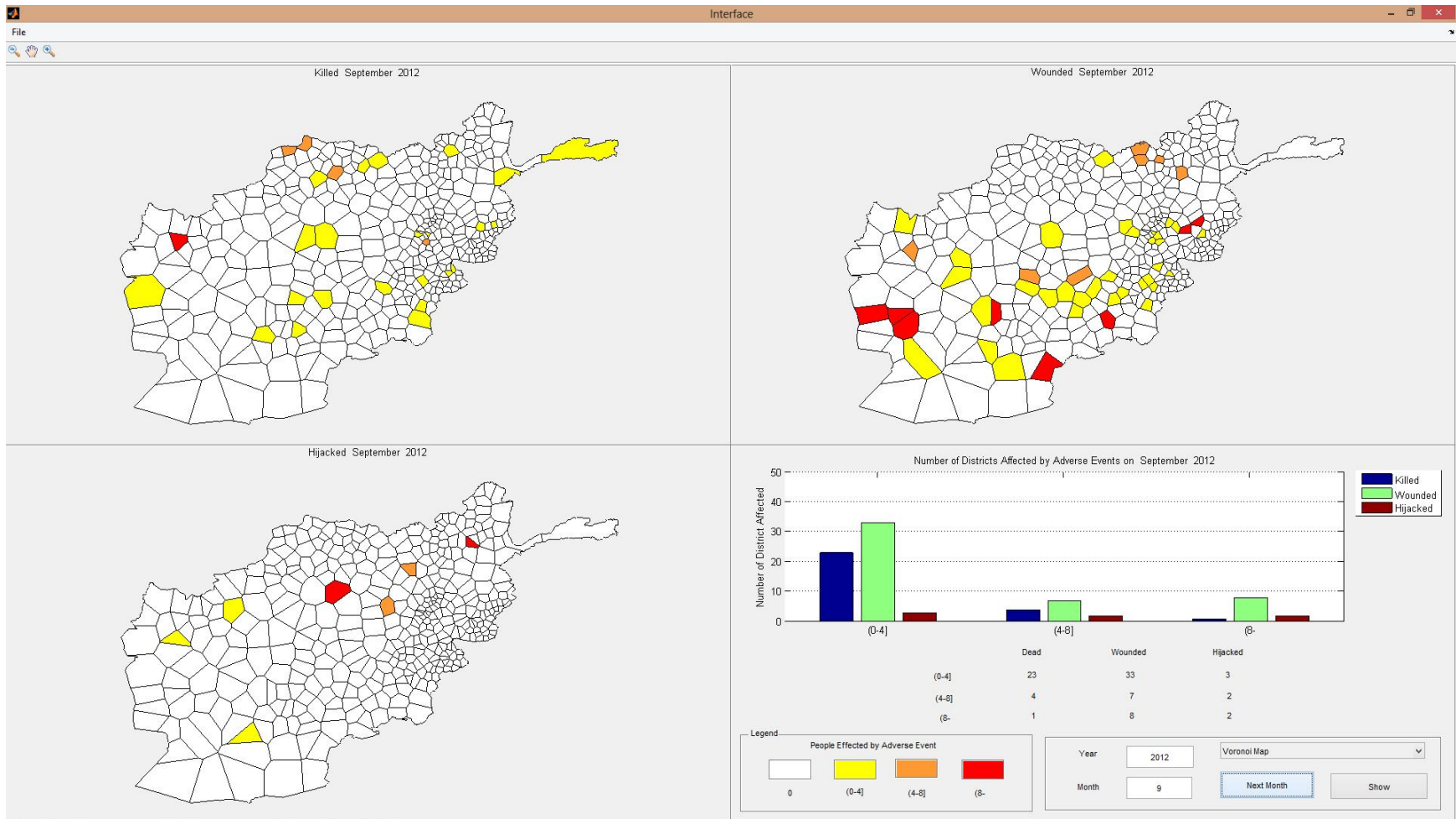
Figure 72 Adverse event forecasts for September 2012
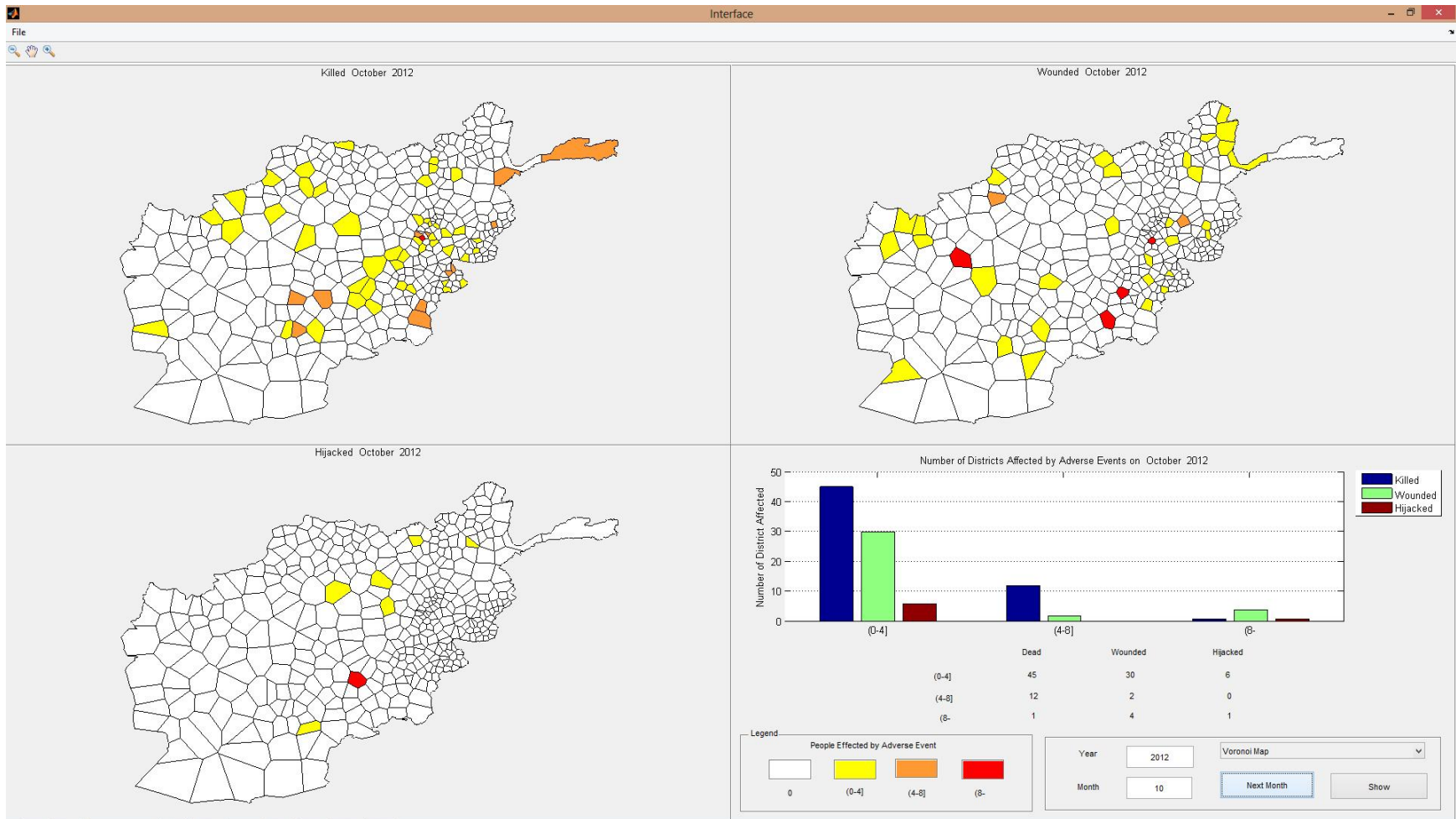
Figure 73 Adverse event forecasts for October 2012
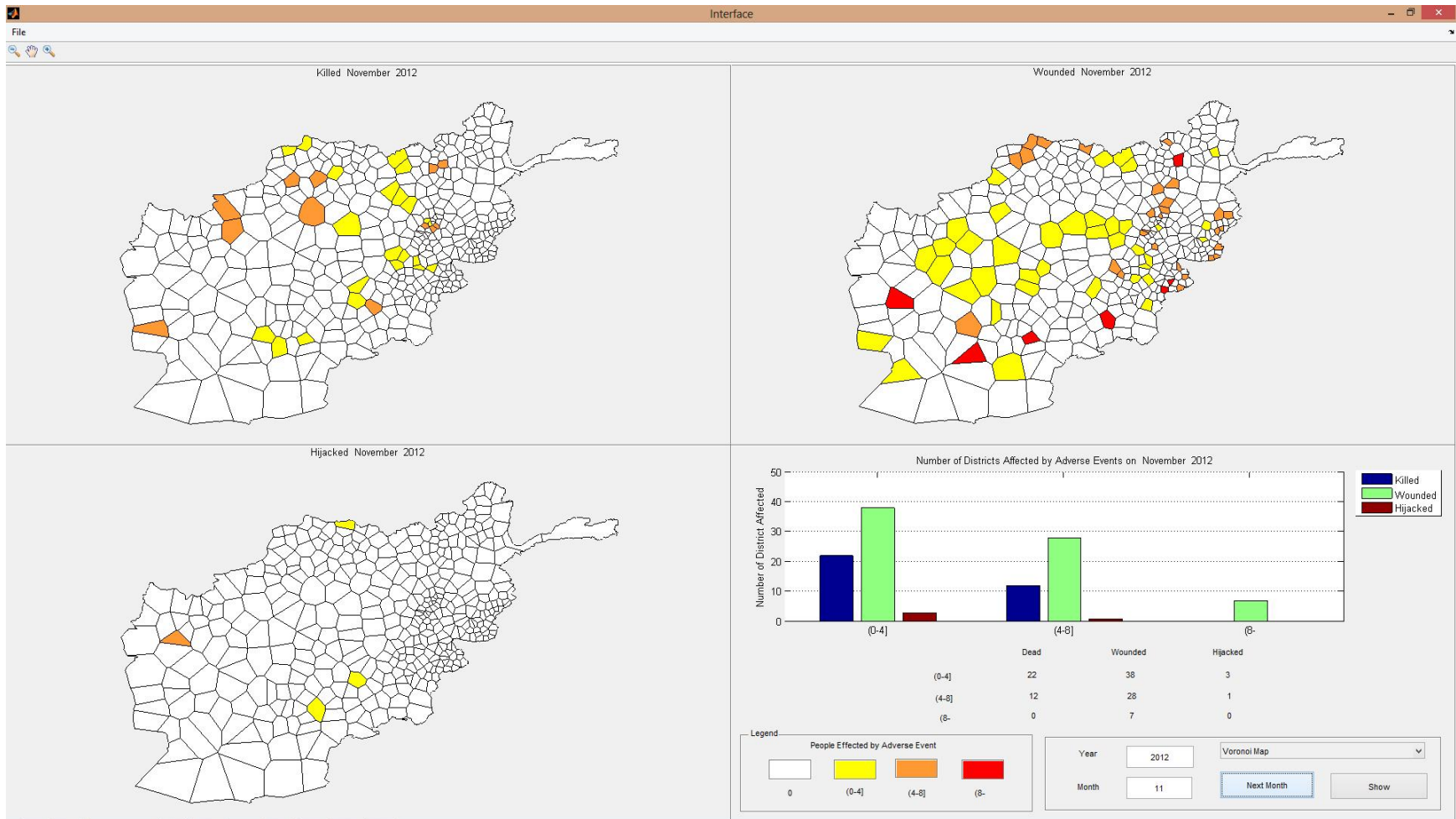
Figure 74 Adverse event forecasts for November 2012
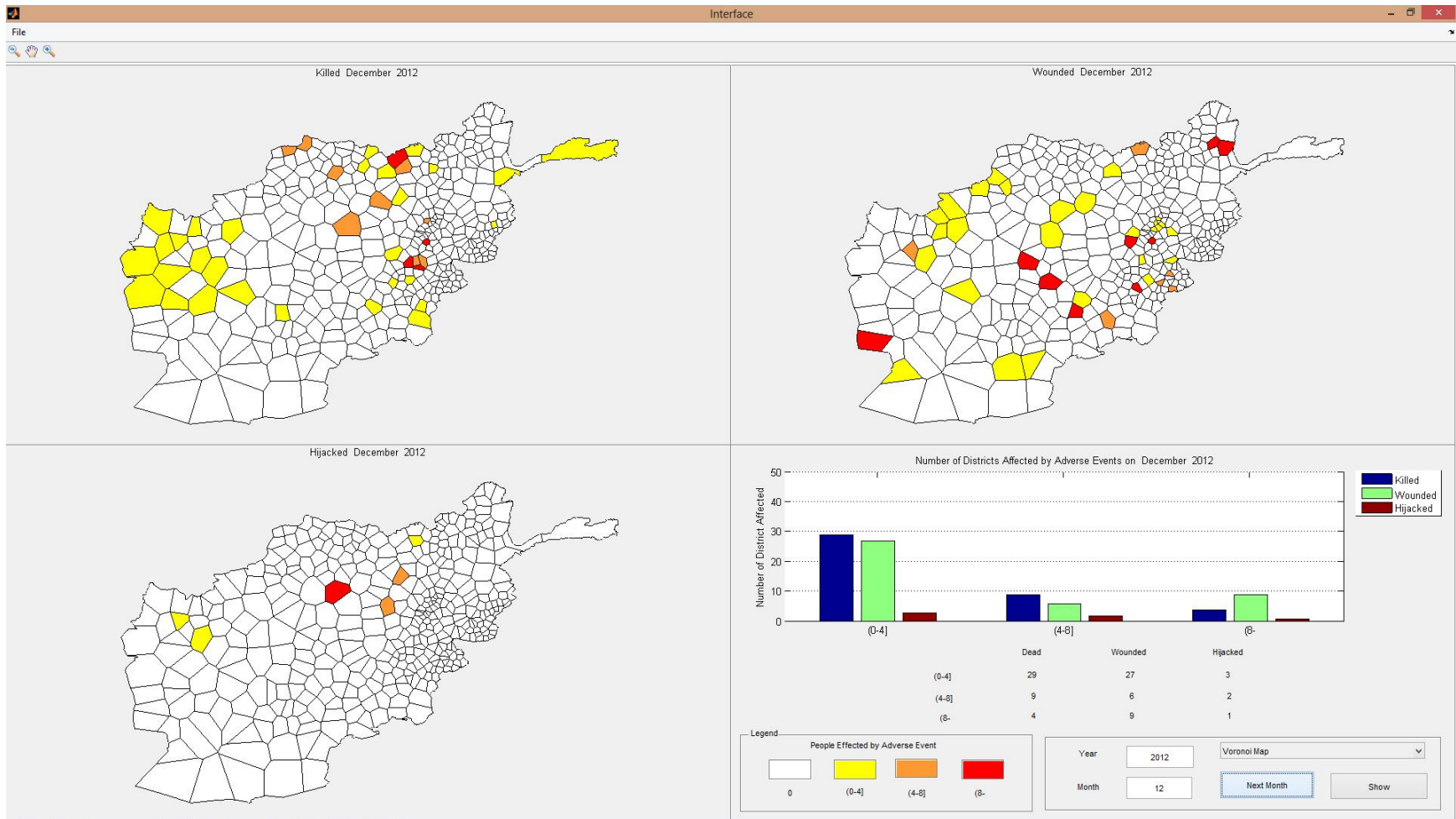
Figure 75 Adverse event forecasts for December 2012
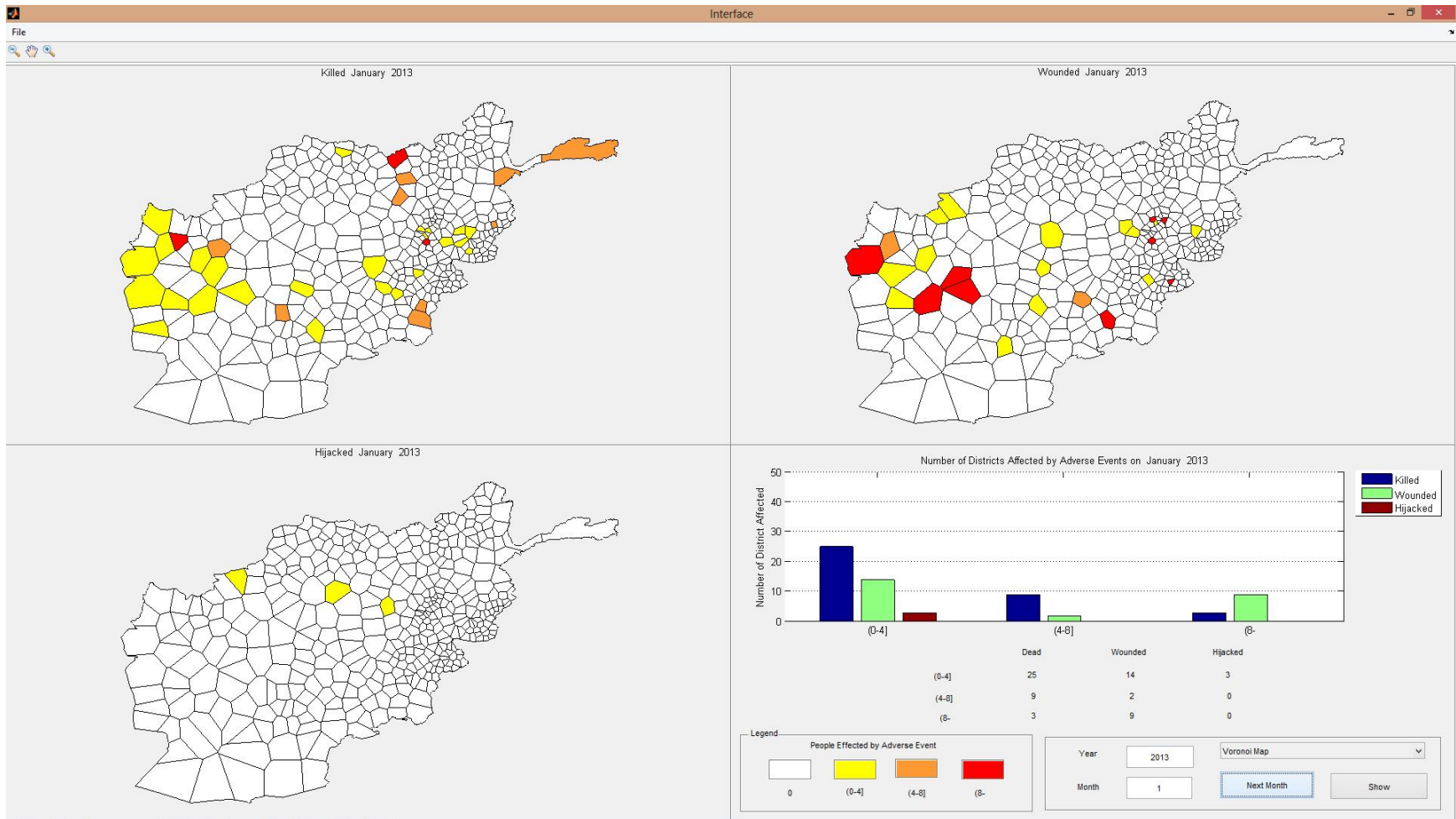
Figure 76 Adverse event forecasts for January 2013
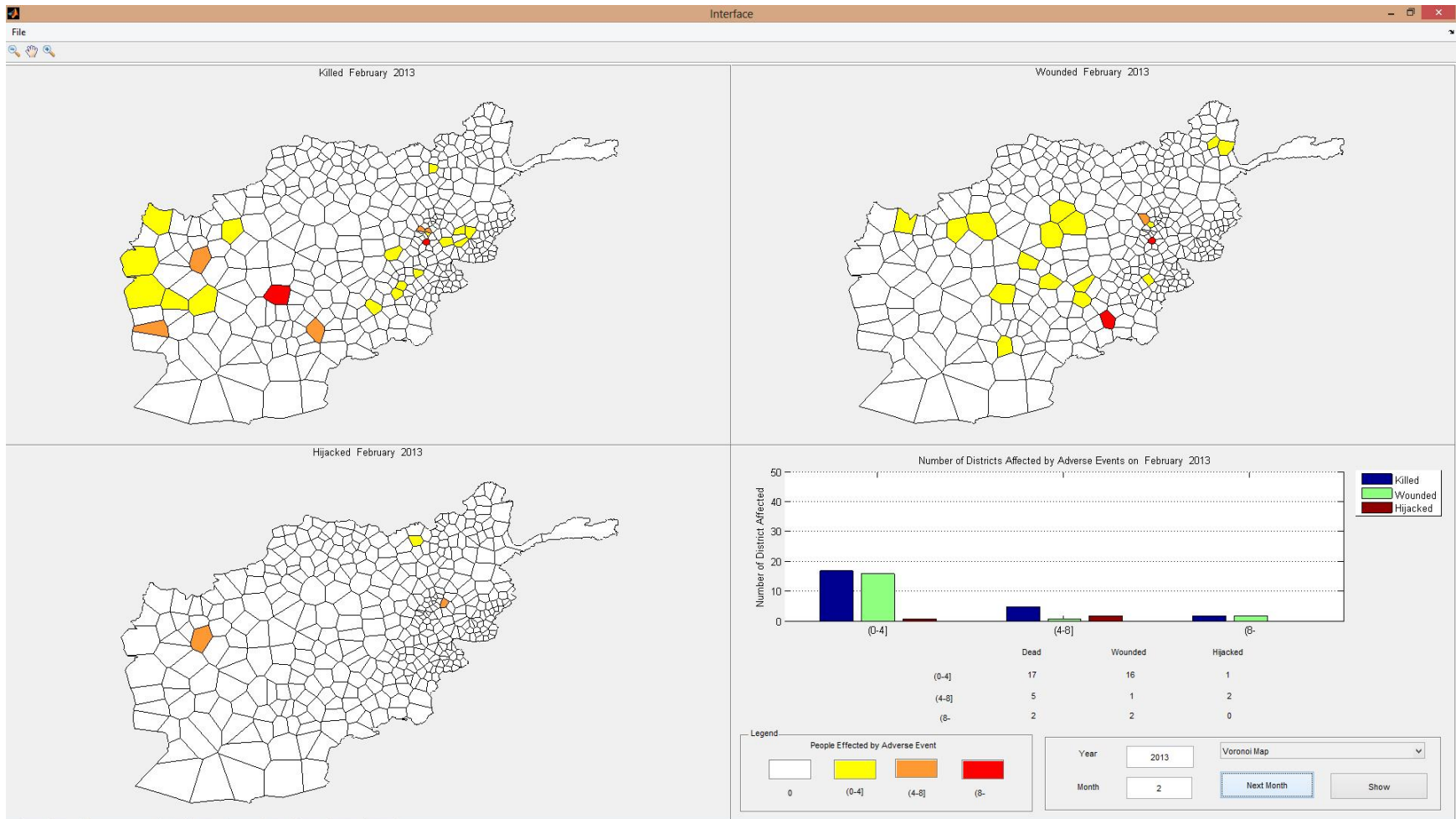
Figure 77 Adverse event forecasts for February 2013
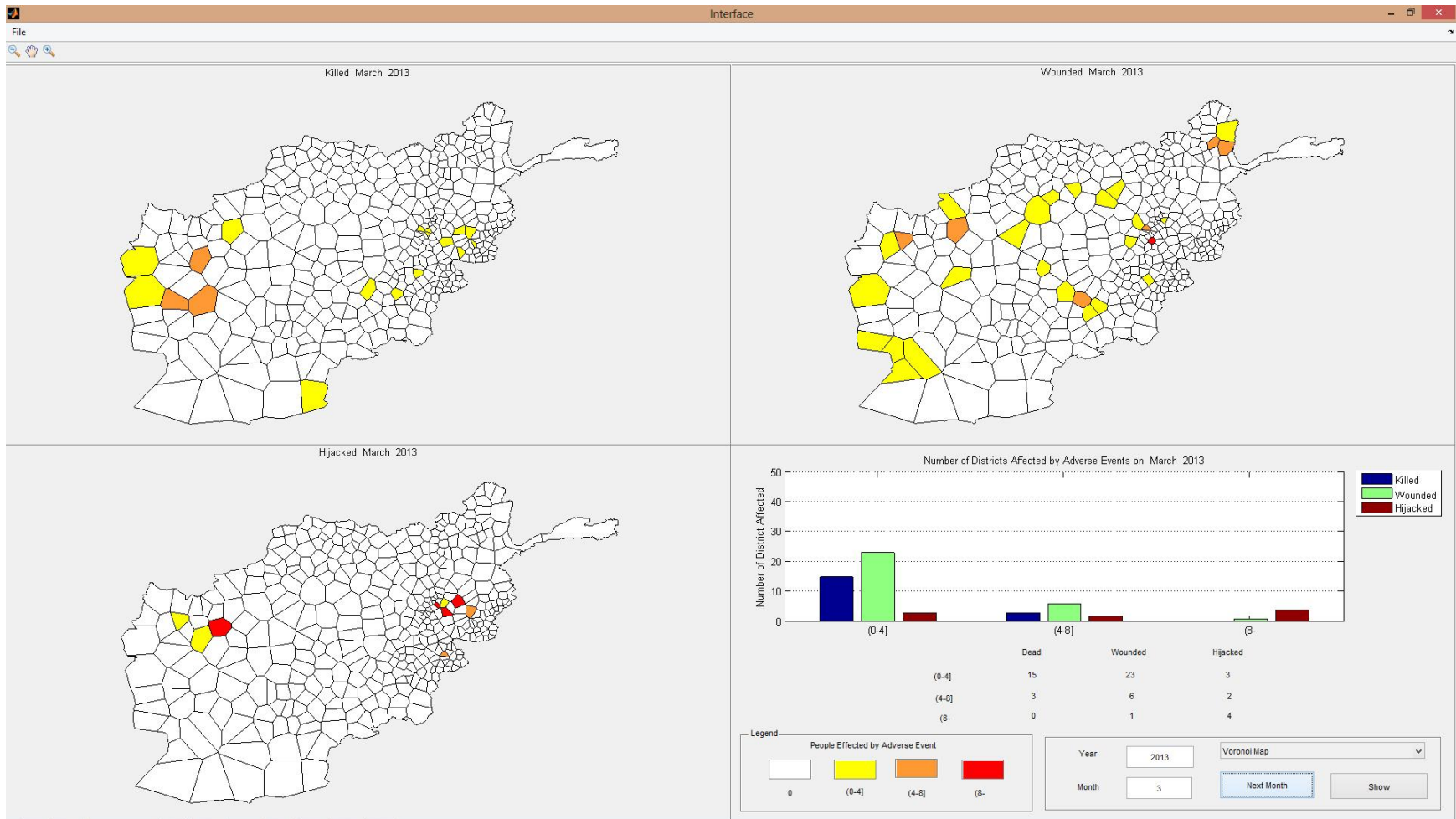
Figure 78 Adverse event forecasts for March 2013
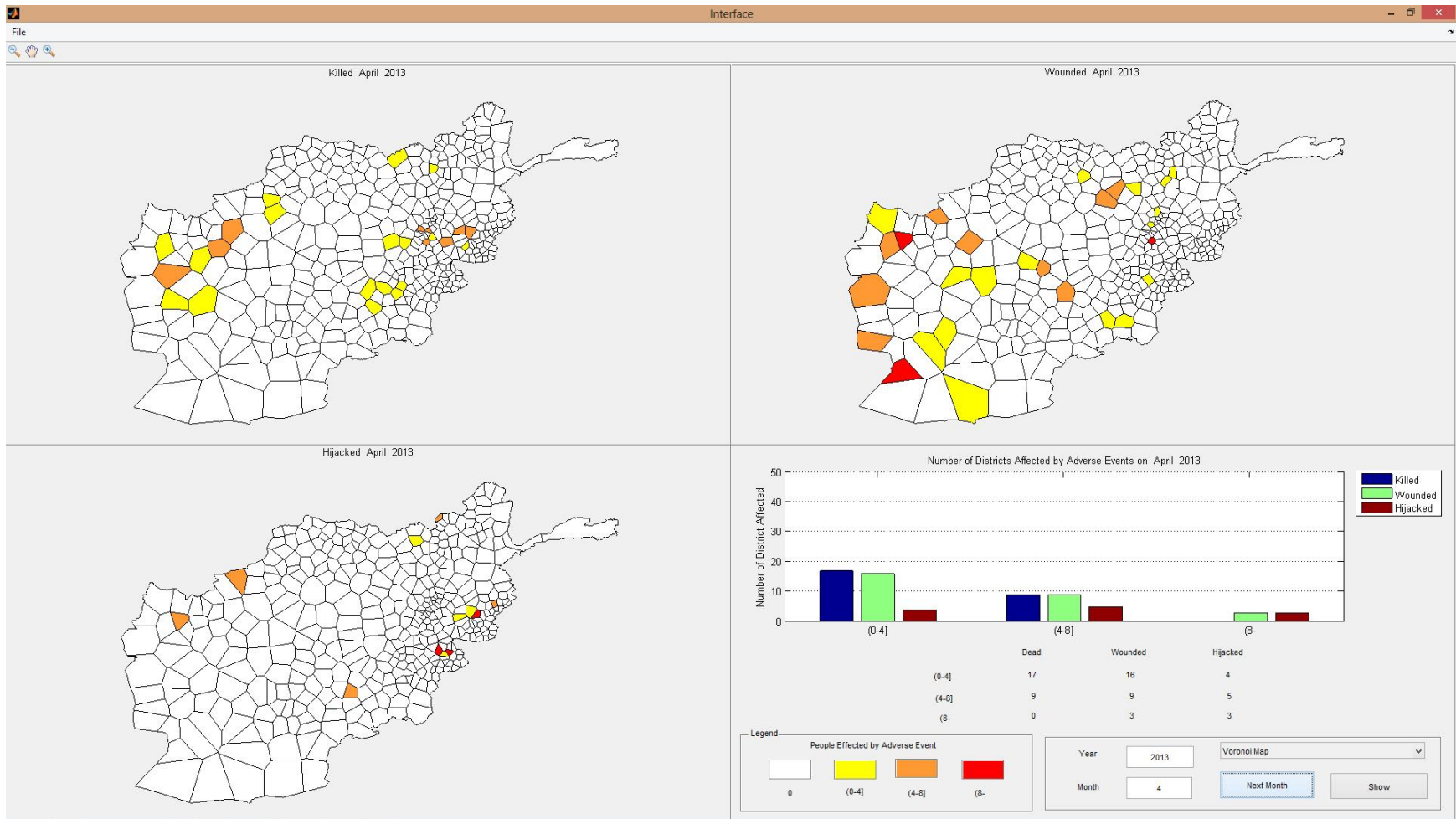
Figure 79 Adverse event forecasts for April 2013
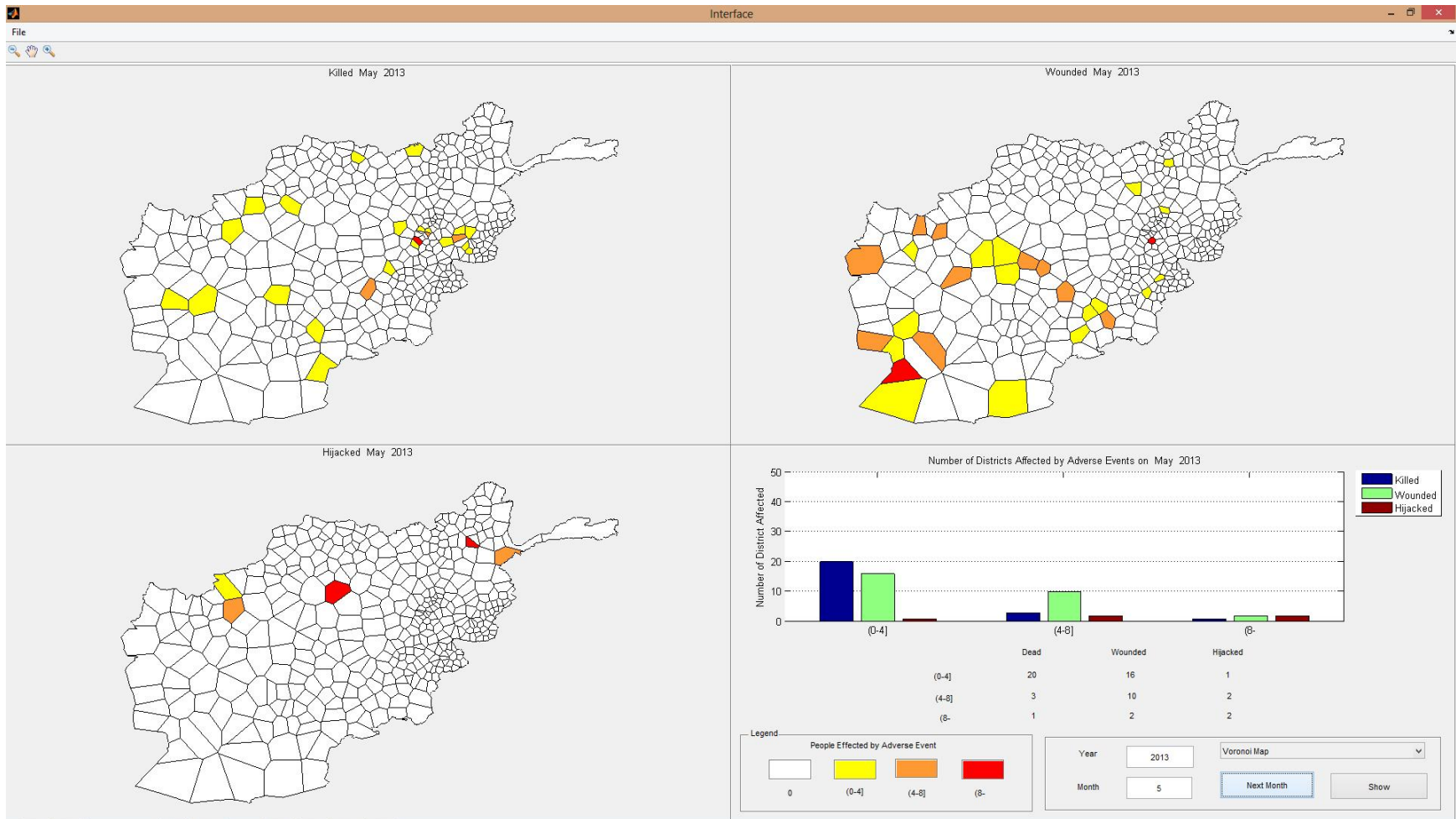
Figure 80 Adverse event forecasts for May 2013
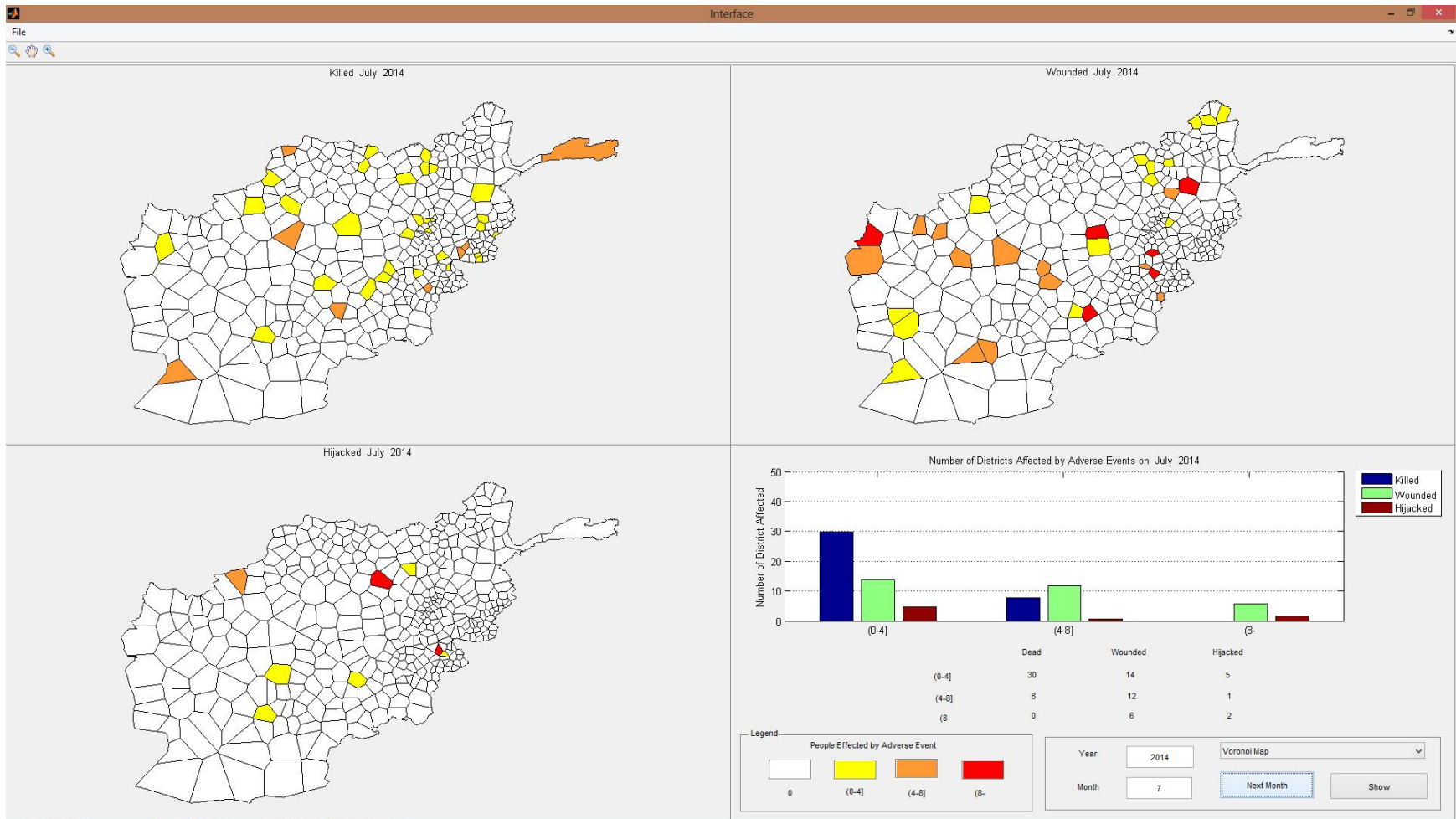
131
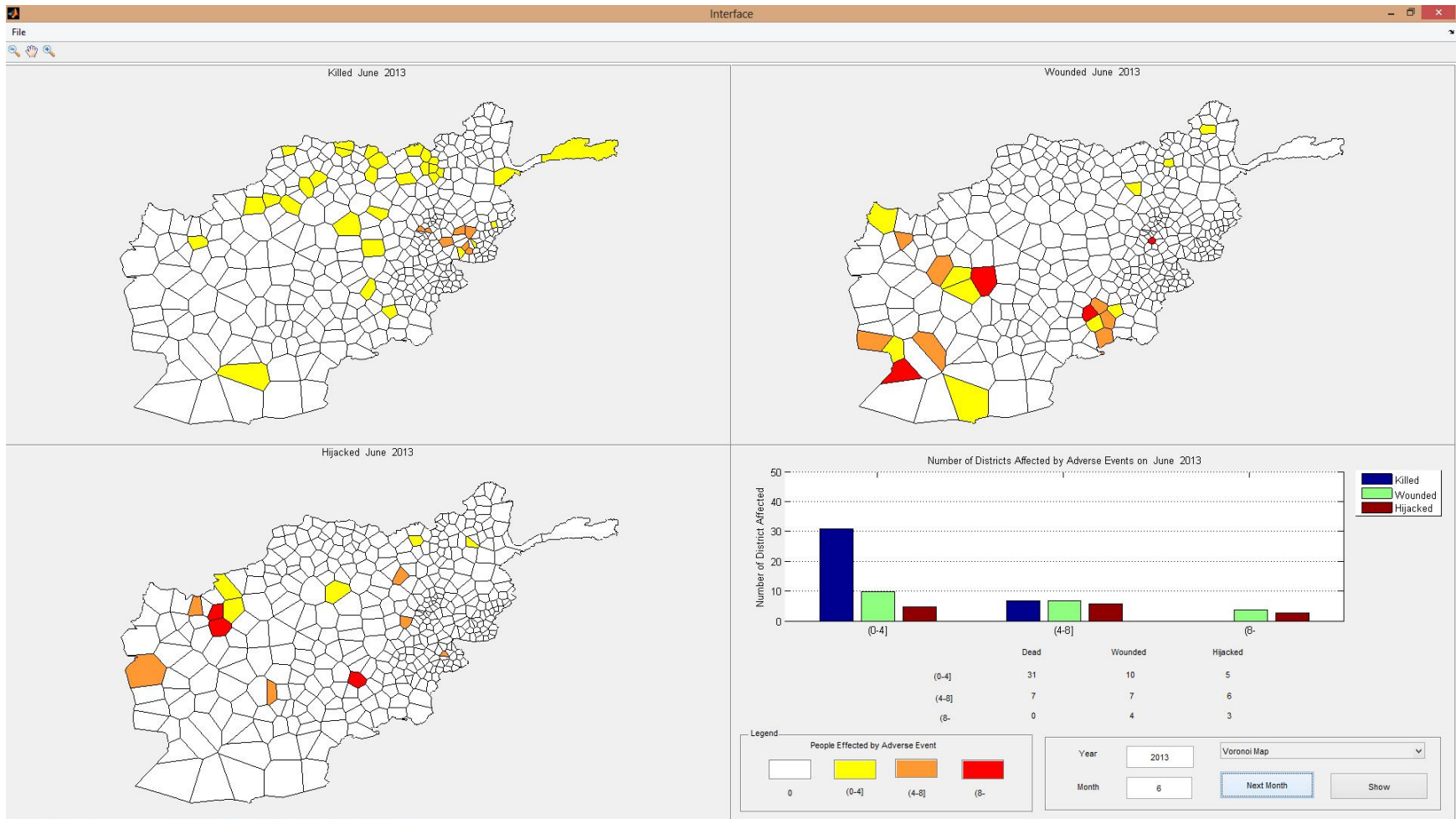
Figure 81 Adverse event forecasts for June 2013
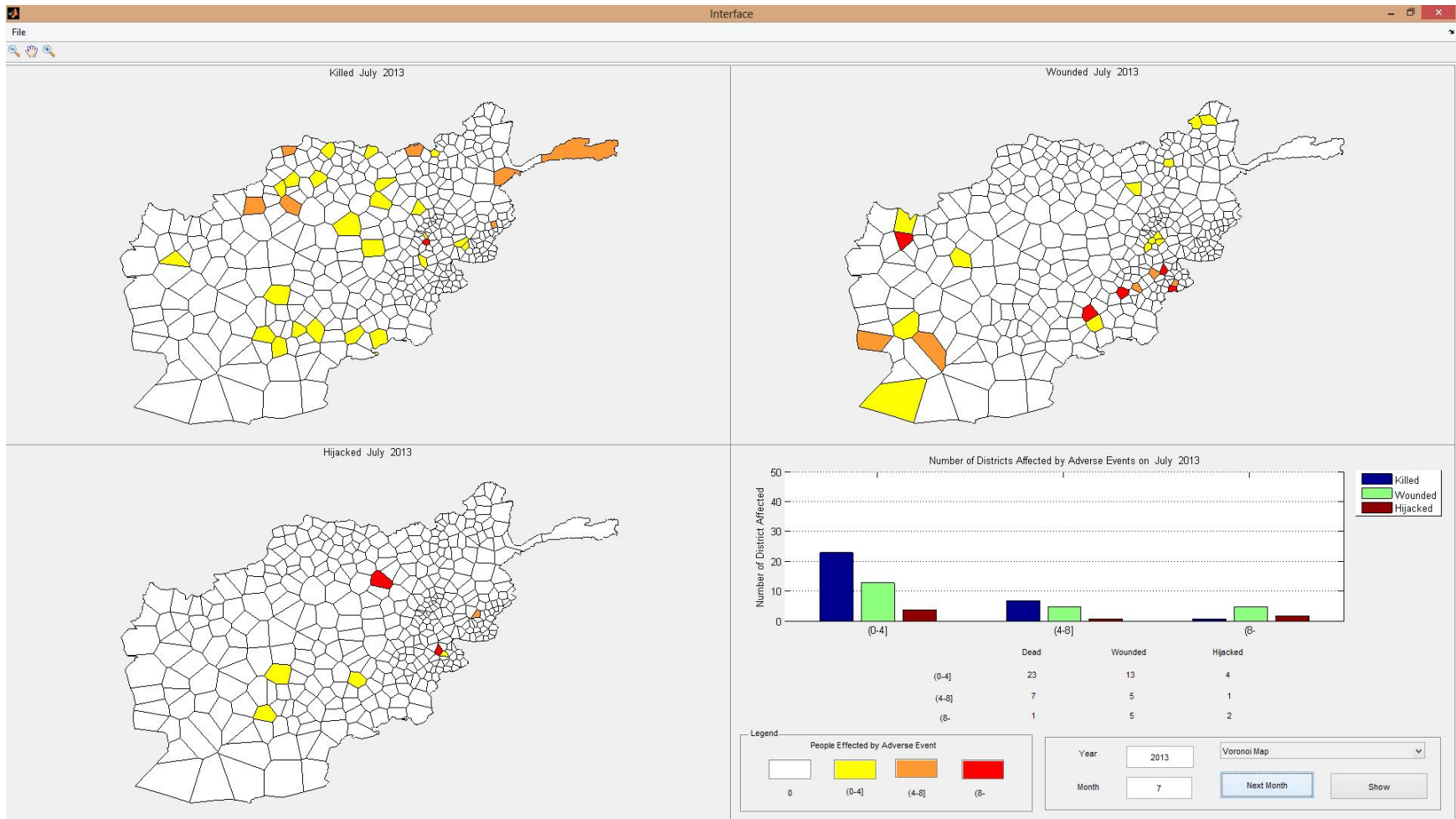
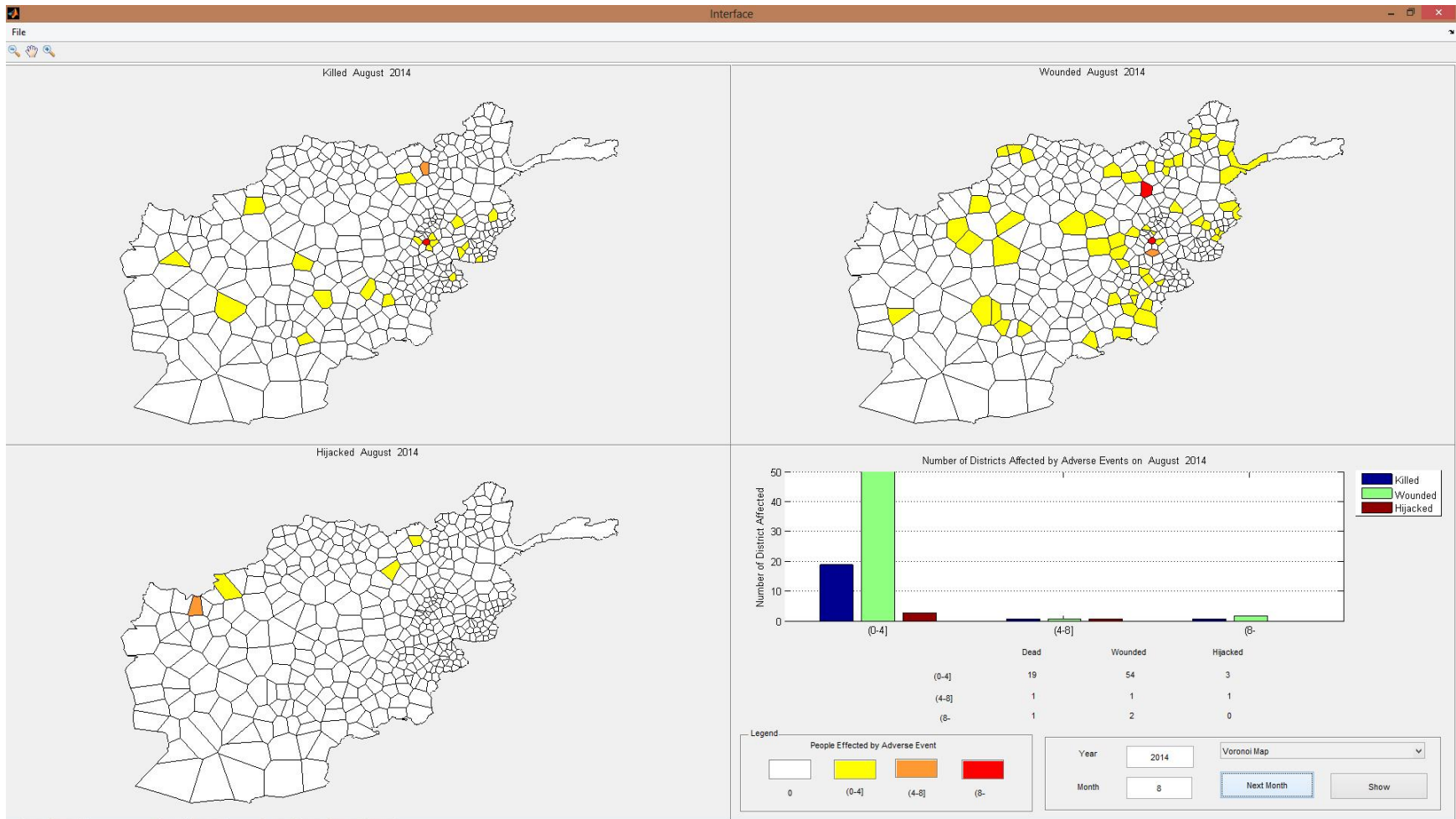Figure 82 Adverse event forecasts for July 2013
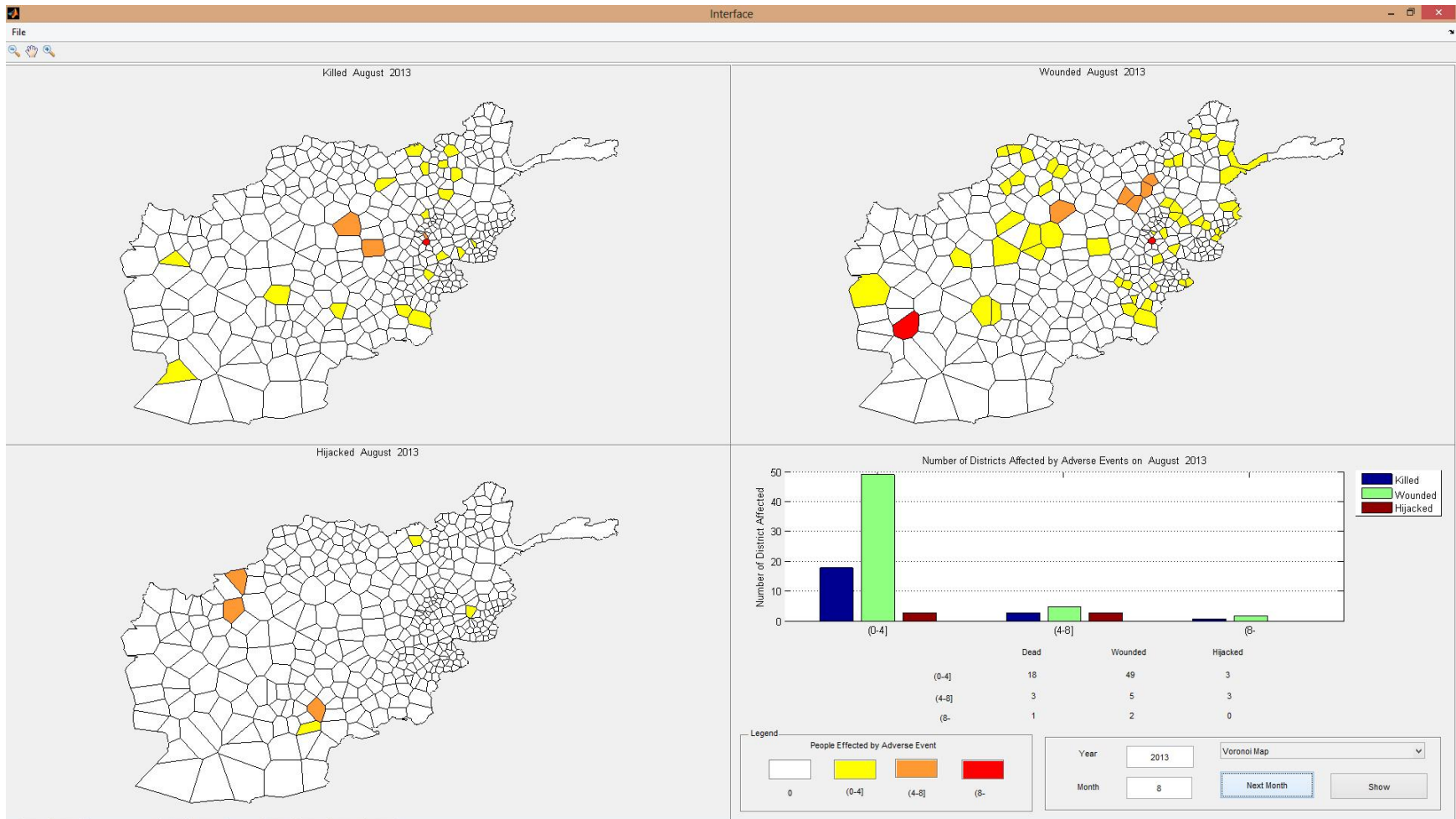
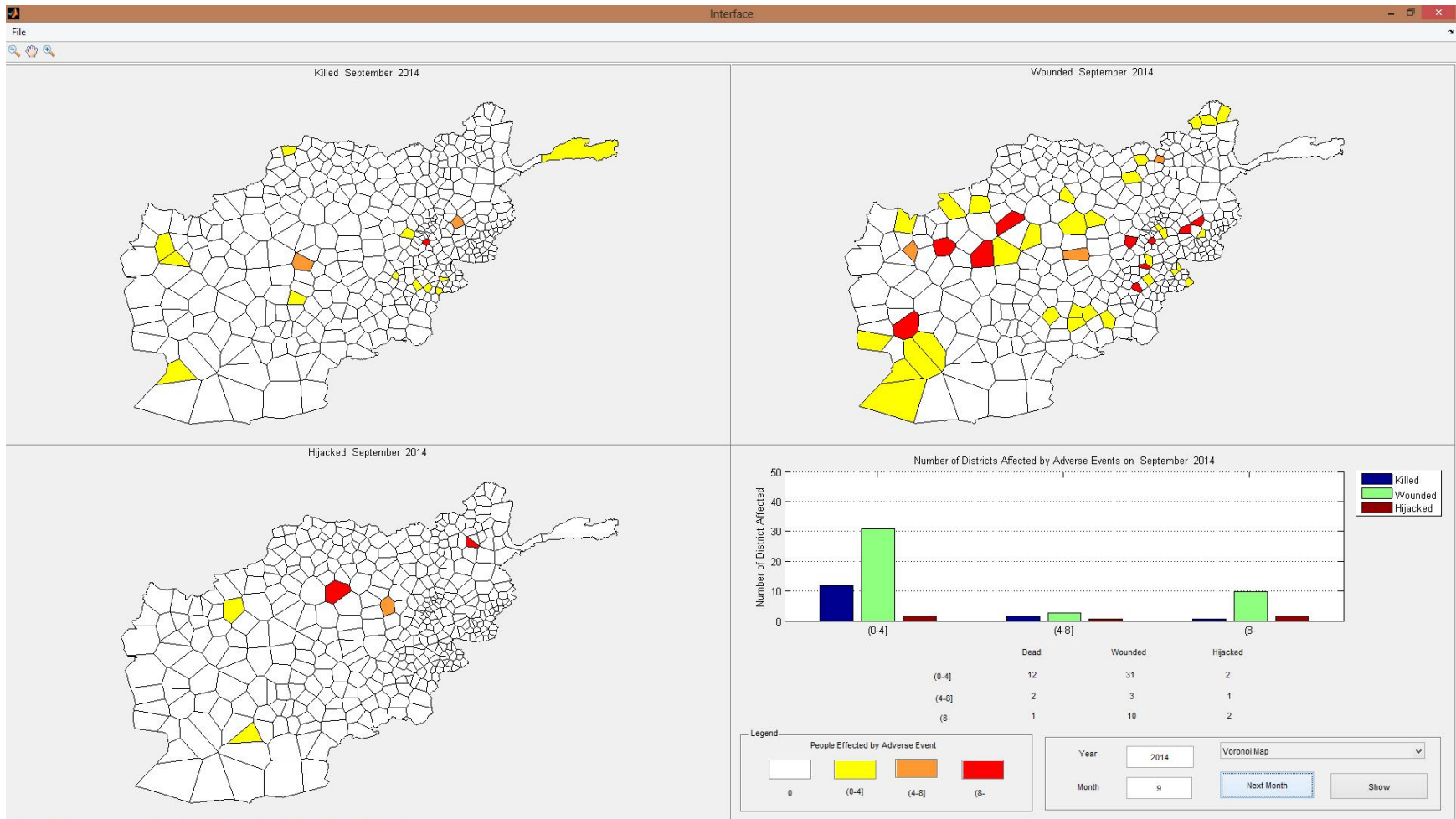Figure 83 Adverse event forecasts for August 2013
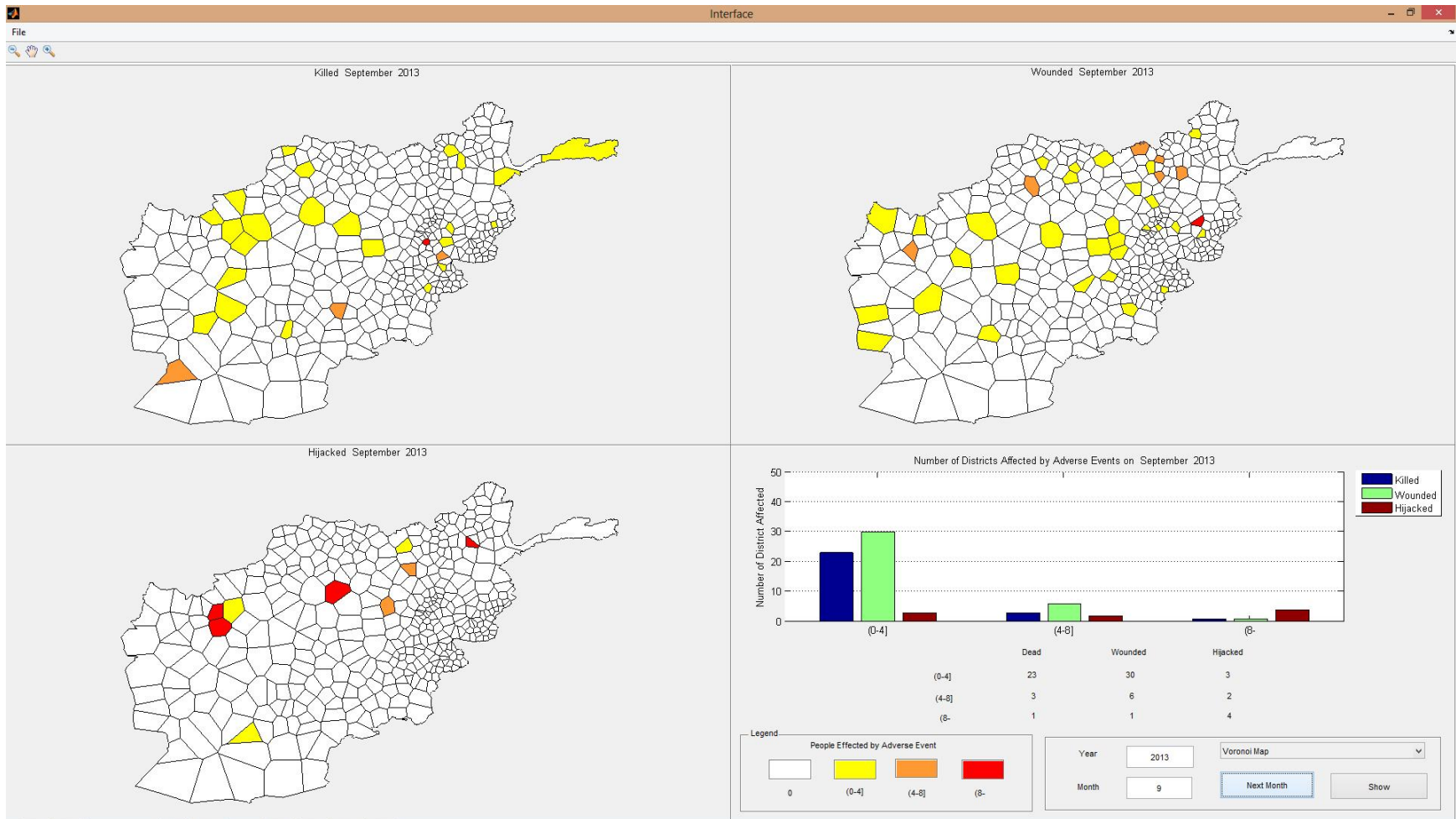
Figure 84 Adverse event forecasts for September 2013
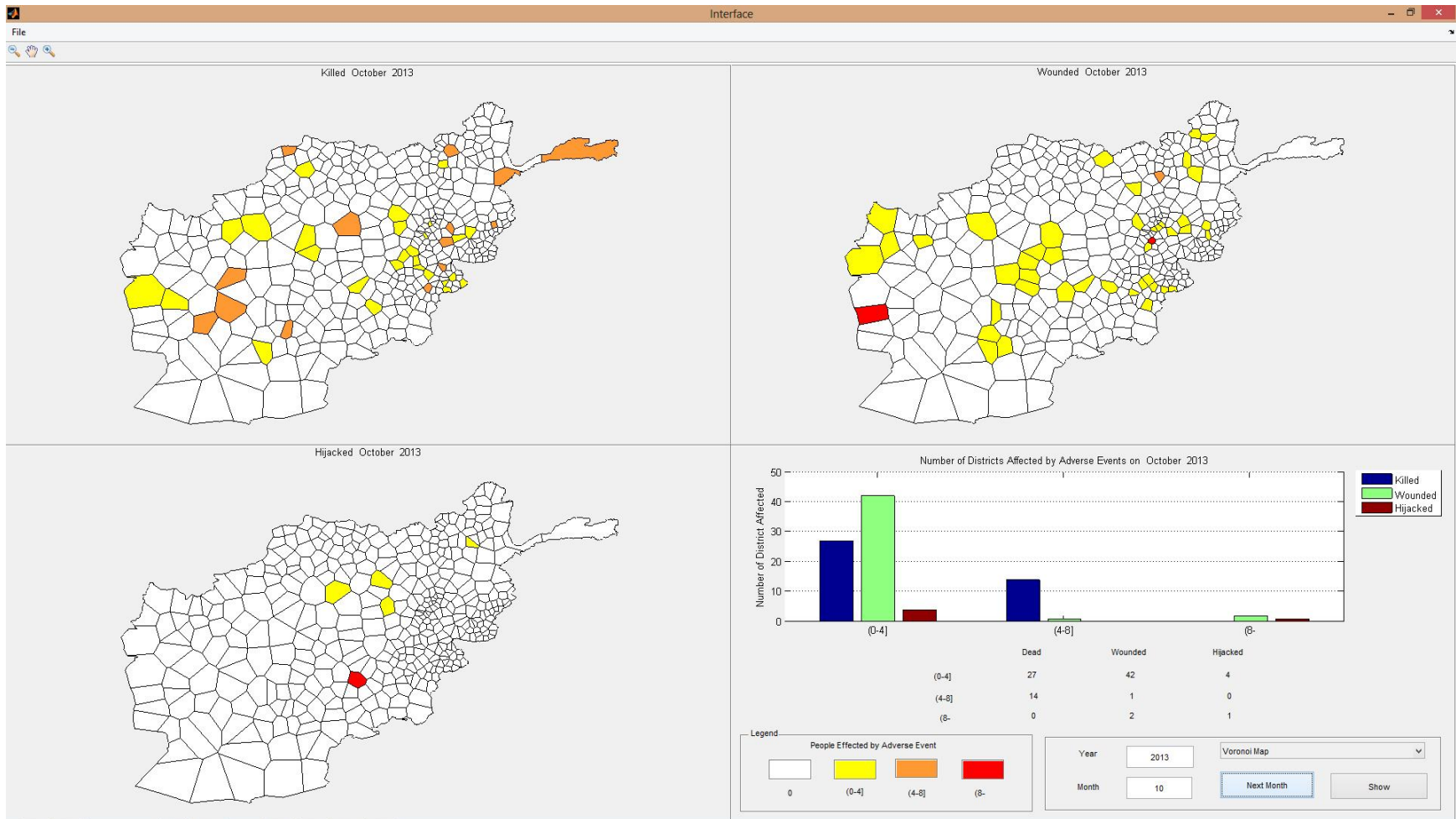
Figure 85 Adverse event forecasts for October 2013

Figure 86 Adverse event forecasts for November 2013

Figure 87 Adverse event forecasts for December 2013

Figure 88 Adverse event forecasts for January 2014

Figure 89 Adverse event forecasts for February 2014

Figure 90 Adverse event forecasts for March 2014

Figure 91 Adverse event forecasts for April 2014

Figure 92 Adverse event forecasts for May 2014

Figure 93 Adverse event forecasts for June 2014

Figure 94 Adverse event forecasts for July 2014

Figure 95 Adverse event forecasts for August 2014

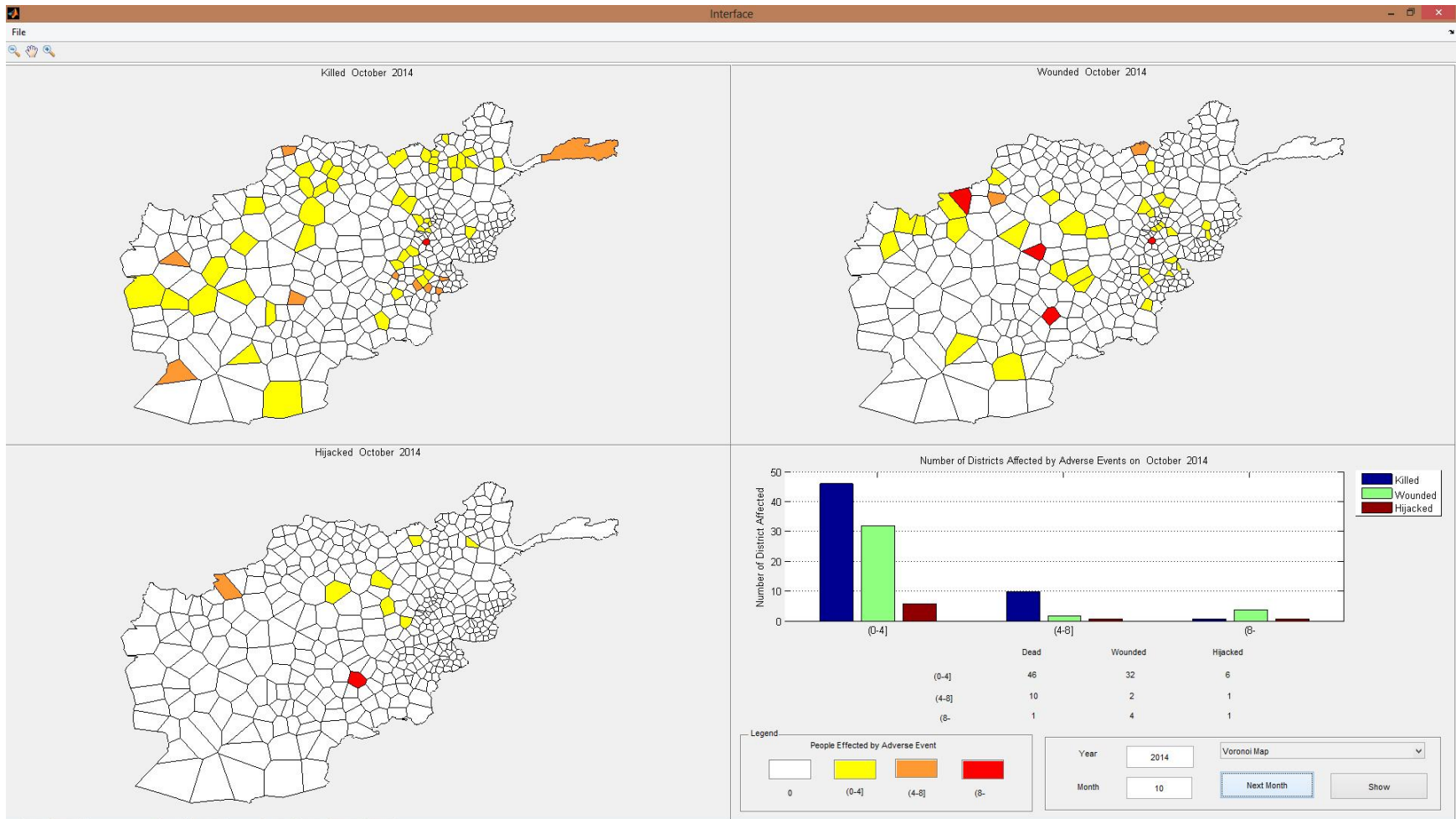Figure 96 Adverse event forecasts for September 2014

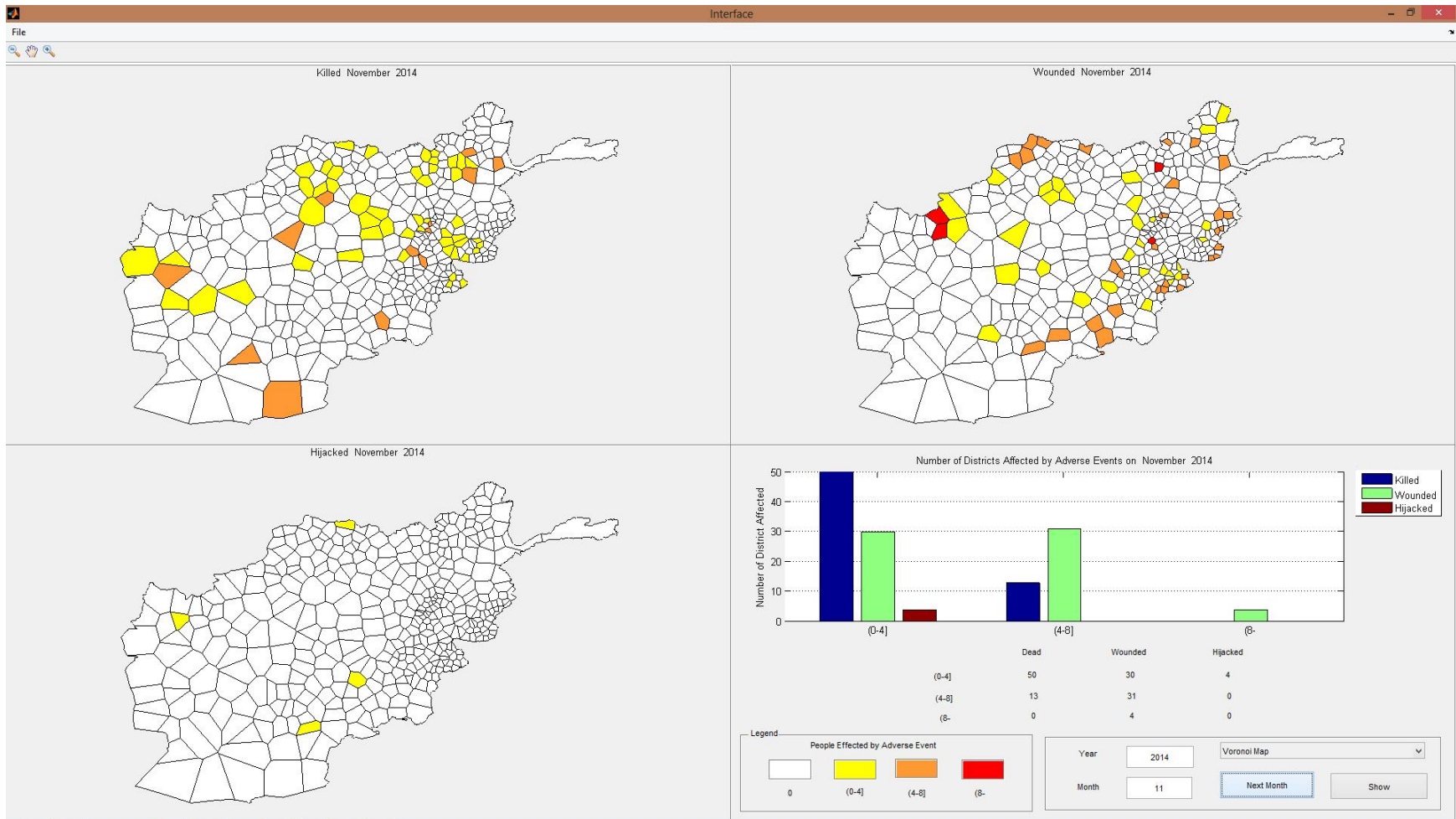Figure 97 Adverse event forecasts for October 2014
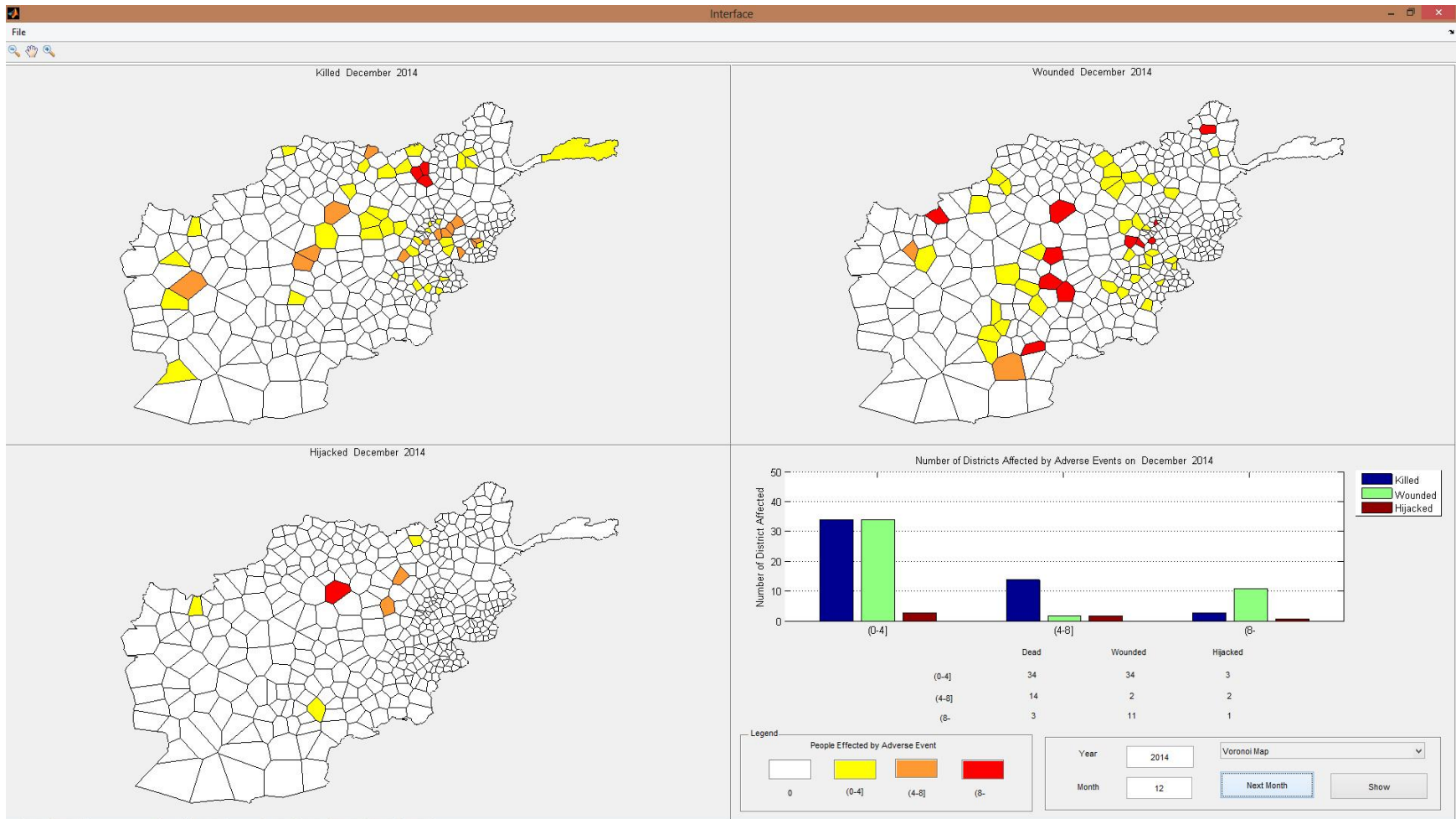
Figure 98 Adverse event forecasts for November 2014

Figure 99 Adverse event forecasts for December 2014