How much data is needed to make an efficient decision: Crowd Sourcing and Disaster Response

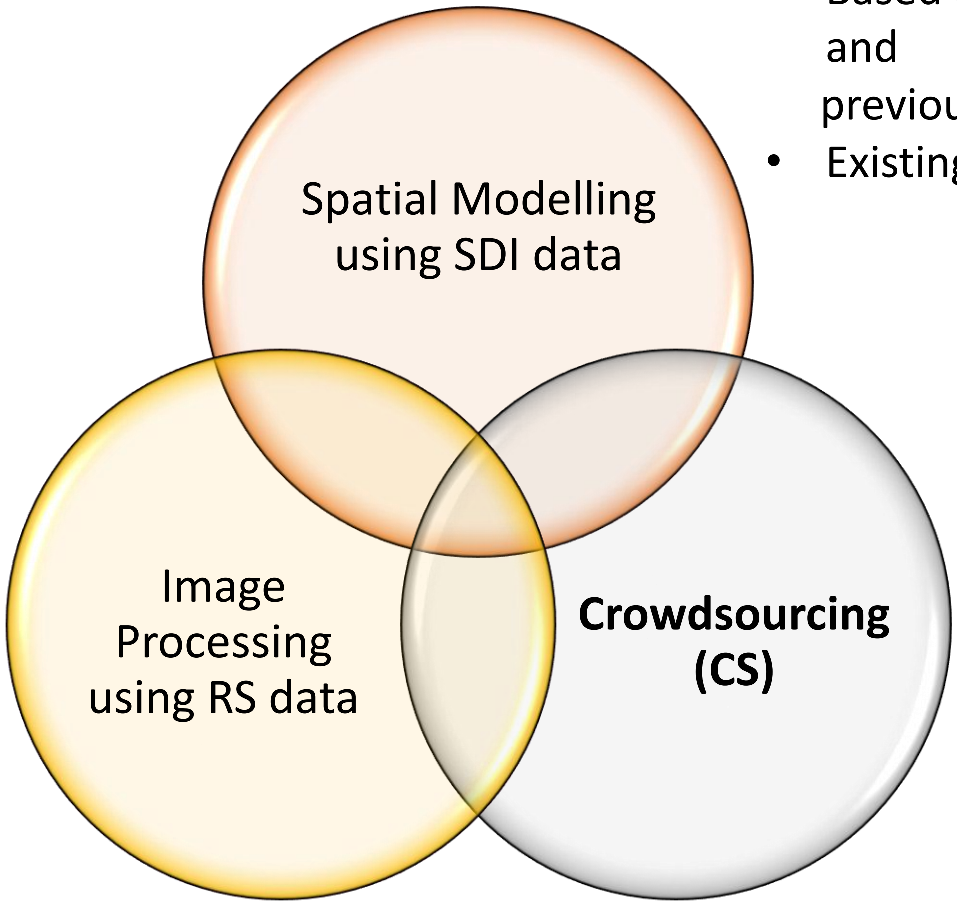
Reza Hassanzadeh

Kerman Disaster Management Centre,

Graduate University of Advanced Technology, Kerman, Iran

UN/India Workshop, 9 Mar 2016 (Hassanzadeh22@yahoo.com)

- Based on expert’s knowledge and
previous experiences
- Existing institutional data



Challenges:
- RS data availability

Challenges:
- CS data availability, quality etc.

It is important to examine the effectiveness of CS data in disaster response

Research Objectives

The overall aim was to improve situational awareness: assessment of damage and losses, and prioritization of damaged areas using Crowd Sourced data (CS data).

Objective:

To investigate the effect of CS data (using hot spot analysis) on prioritization of damaged areas, and to explore how much data is needed to make an efficient decision.

Data collection and preparation

- **The CS data**

  Questionnaire Survey **with people who experienced the earthquake in Bam city, Iran, in 2003**.

  396; two stages cluster sampling method

  Data:

  - the number of injured people in buildings,
  - the number of fatalities,
  - the destruction level of each building,
  - when it was possible for them to submit data.

Post Earthquake Crowd Sourced Data Collection

A)Basic information:

1. Information on sex, age and level of education :
   - Sex: (1) Male □ (2) Female □
   - What is your age: ....
   - What is your level of education ??

(1) Primary □ (2) Lower Secondary □ (3) Higher Secondary □
(4) Diploma □ (5) Degree □ (6) Postgraduate □

B) Household Demographic Information:

2. Were you a resident of Bam city, when earthquake occurred in 2003?
   Yes (1) □ No (2) □
3. How long had you lived in Bam city? .... Years
4. Do you still live in the same residence? Yes (1) □
   No (2) □ , why did you move...........................
5. How many people were living in the building?......... person / persons
6. What age were they?
   .... (1) Under 18 .... (2) 19-34 .... (3) 35-49 .... (4) Over 65
7. Do you know what material the building was?
   (1) Adobe Building □ (2) Unreinforced masonry Building □
   (3) Unreinforced masonry building with Reinforced floor □ (4) Steel building □
   (5) Reinforced masonry building □ (6) Reinforced concrete building □
   Others □
8. How many Storeys had the building at the time of earthquake?........
9. What was the impact of the earthquake on the building?
   (1) No destruction □ (2) Minor cracks in the walls □
   (3) Major cracks in the walls (1 cm) □ (4) One wall was collapsed □
   (5) One wall and some part of roof were collapsed □
   (6) Completely collapsed □
10. Were any people in the building when earthquake happened?
    Yes □, How many?............................ No □, Why?.........................
11. Were there any people trapped under the debris?
    Yes □, How many?............................ No □, Why?.........................
12. Were there any people injured?
    Yes □, How many in total?.........Hospitalized people...... No □, Why?..............
13. Were there any fatalities?
    Yes □, How many?............................ No □, Why?.........................
14. .After the earthquake what happened for you? When were you able to submit this data to the Disaster Management Center's database? (The researcher explain the assumptions and record the submission time in hour or minute)

Only office use:

-Survey code:

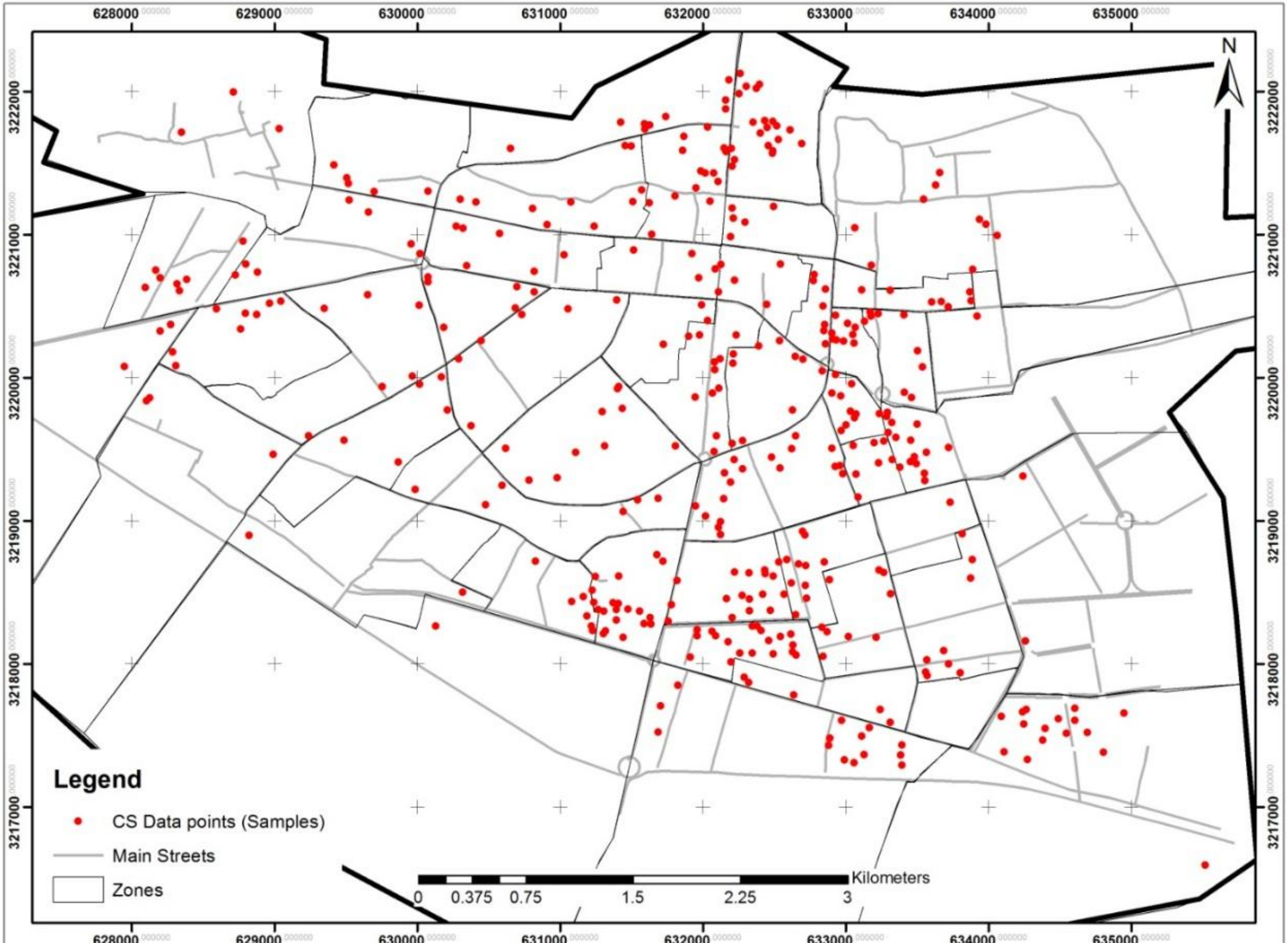
Building coordinate: X: Y:

Address: No:.................. Alley:...........................
Street:.............................

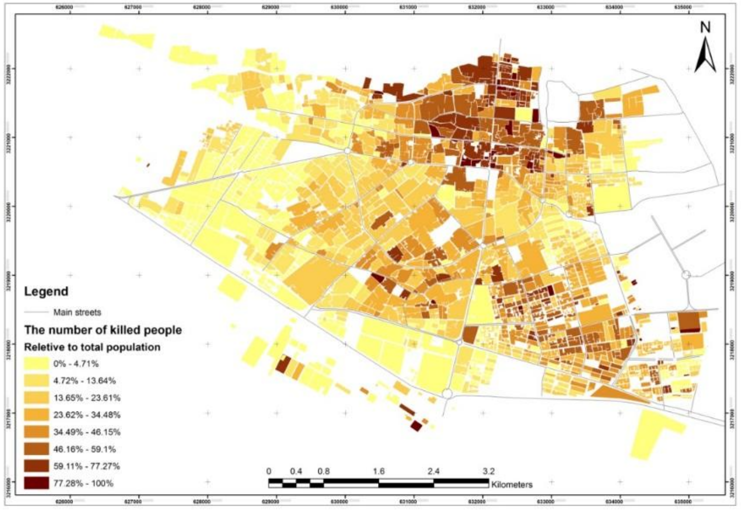
Extra explanation:

Questionnaire form

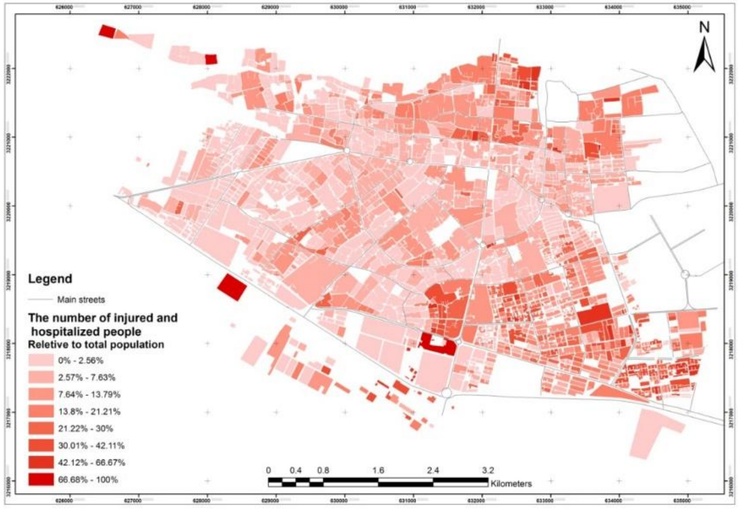
Distribution of respondents (CS data) in the study area



Actual earthquake data (AE data), the Bam earthquake in 26th December 2003



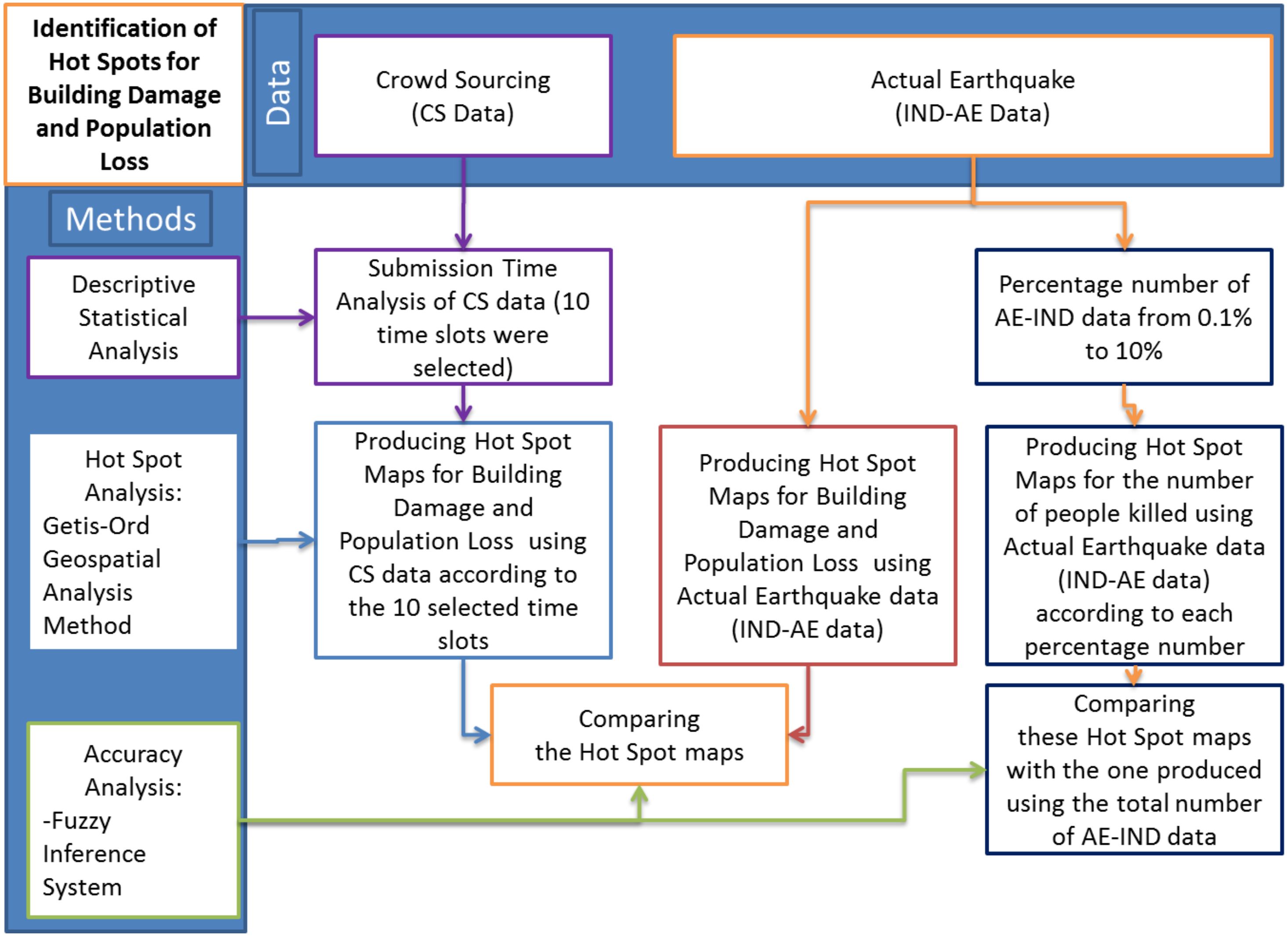
The distribution of people killed (SCI, 2004)



The distribution of people injured and hospitalized (SCI, 2004)

The distribution of completely collapsed buildings (SCI, 2004)

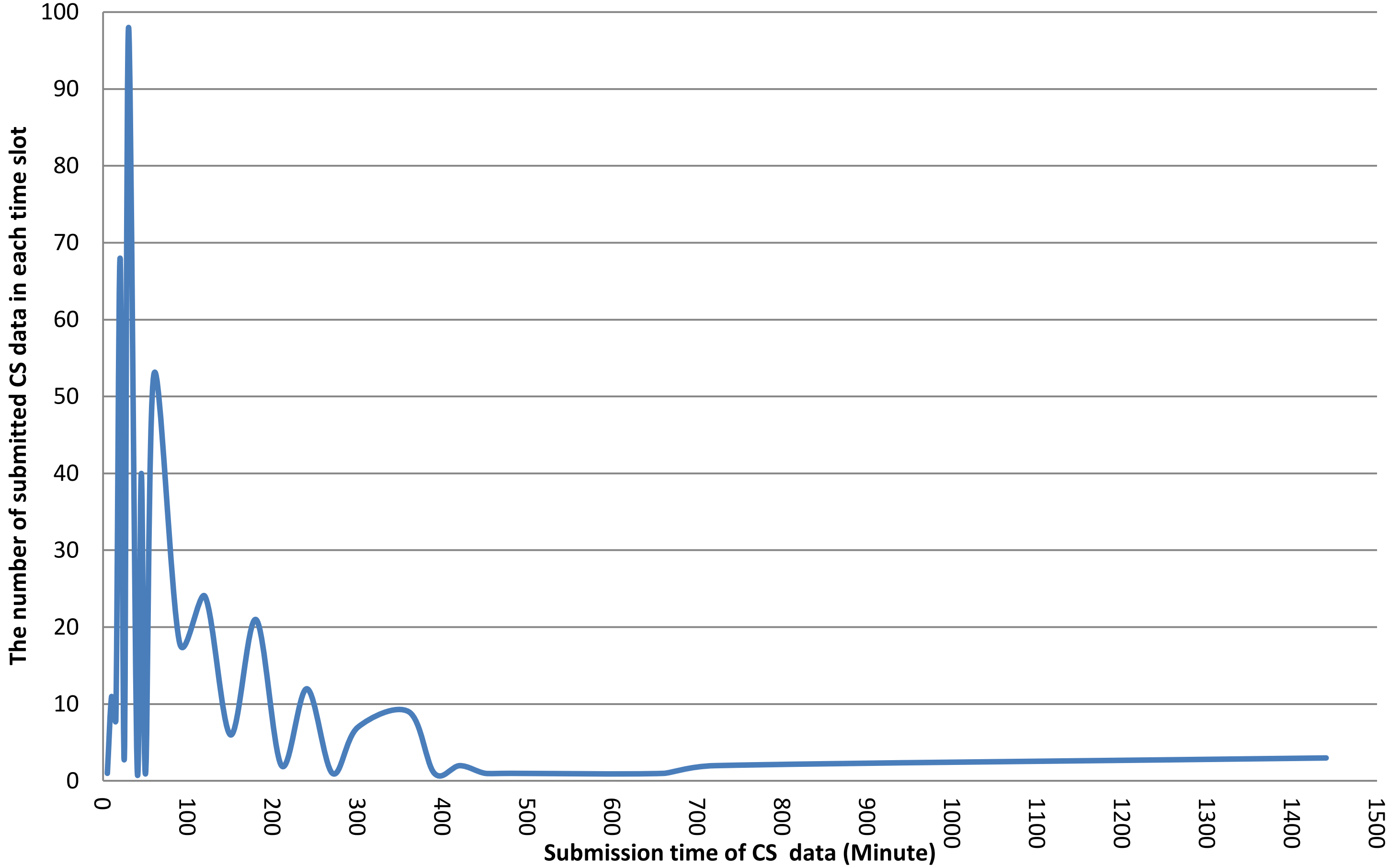
The distribution of people non-injured (SCI, 2004)



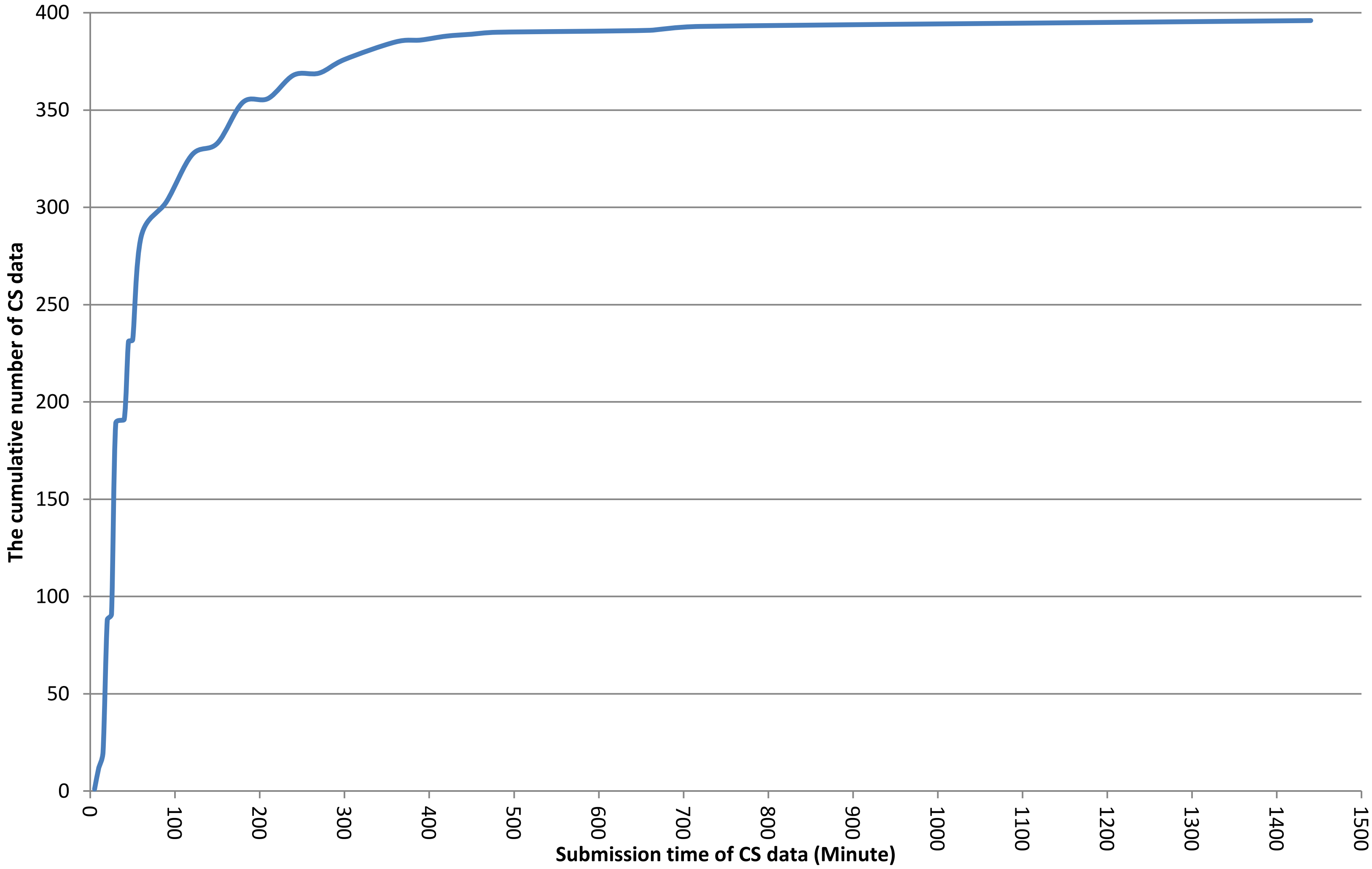
The workflow for hot spot analysis

Analysing submission time of CS data and identifying key time slots

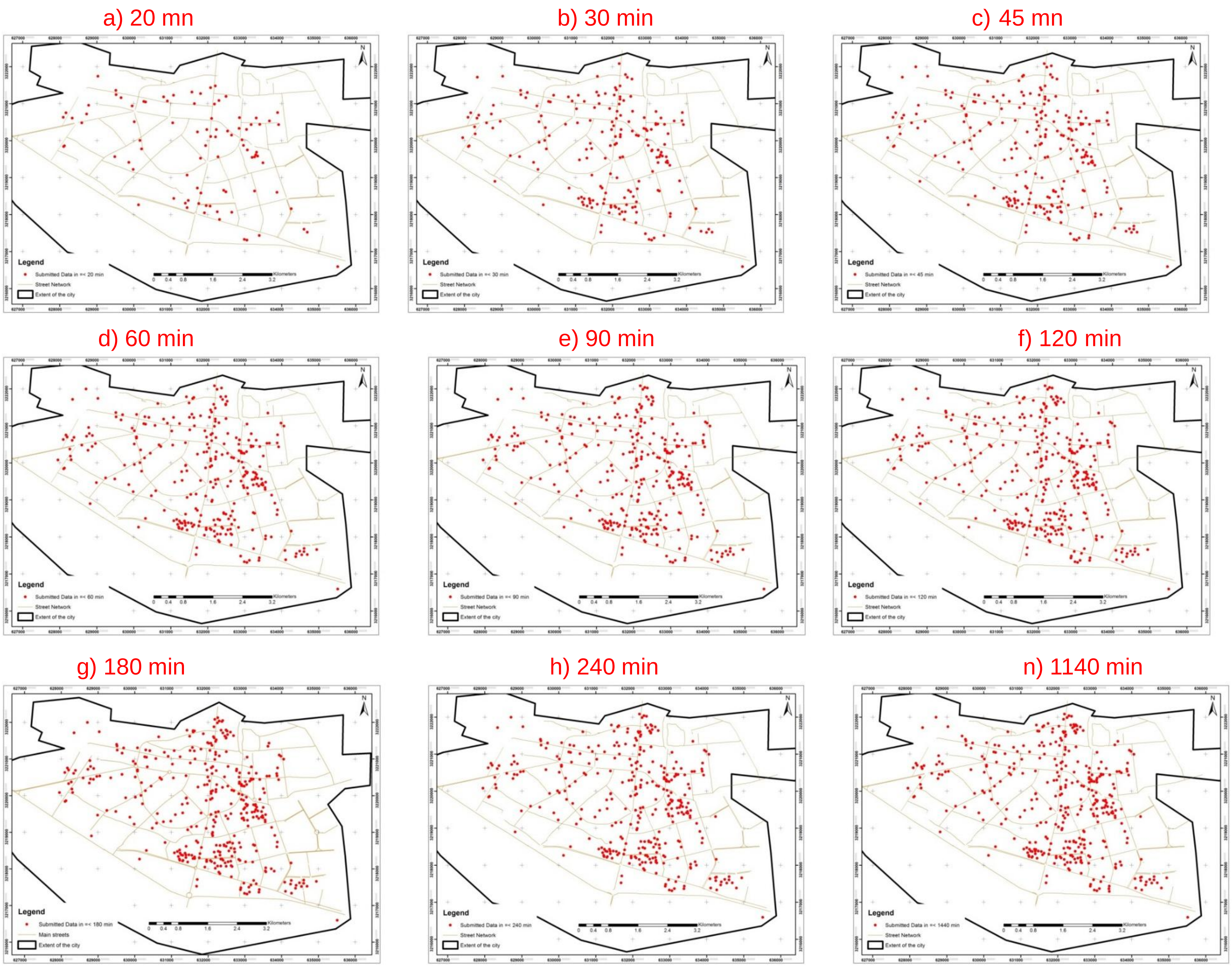
Submission time of CS data (minute)	5	10	15	20	25	30	40	45	50	60	90	120	150
Cumulative number of CS submission	1	12	20	88	91	189	191	231	232	285	303	327	333
Differences between time slots		11	8	68	3	98	2	40	1	53	18	24	6
Large differences (4th)				87		101		42		54	18	24	
Selected time slot		-	-	(1)	-	(2)	-	(3)	-	(4)	(5)	(6)	-
Submission time of CS data (minute)	**180**	**210**	**240**	**270**	**300**	**360**	**390**	**420**	**450**	**480**	**660**	**720**	**1440**
Cumulative number of CS submission	354	356	368	369	376	385	386	388	389	390	391	393	396
Differences between time slots	21	2	12	1	7	9	1	2	1	1	1	2	3
Large differences (9th)	27		14			17							11
Selected time slot	(7)	-	(8)	-	-	(9)	-	-	-	-	-	-	(10)



Continuous CS data submission after the Bam earthquake struck in 203



The cumulative number of CS data submission after the Bam earthquake struck in 2003



CS data points' entry and its distribution in Bam city after the earthquake struck based on submission time slot

How to use CS data?

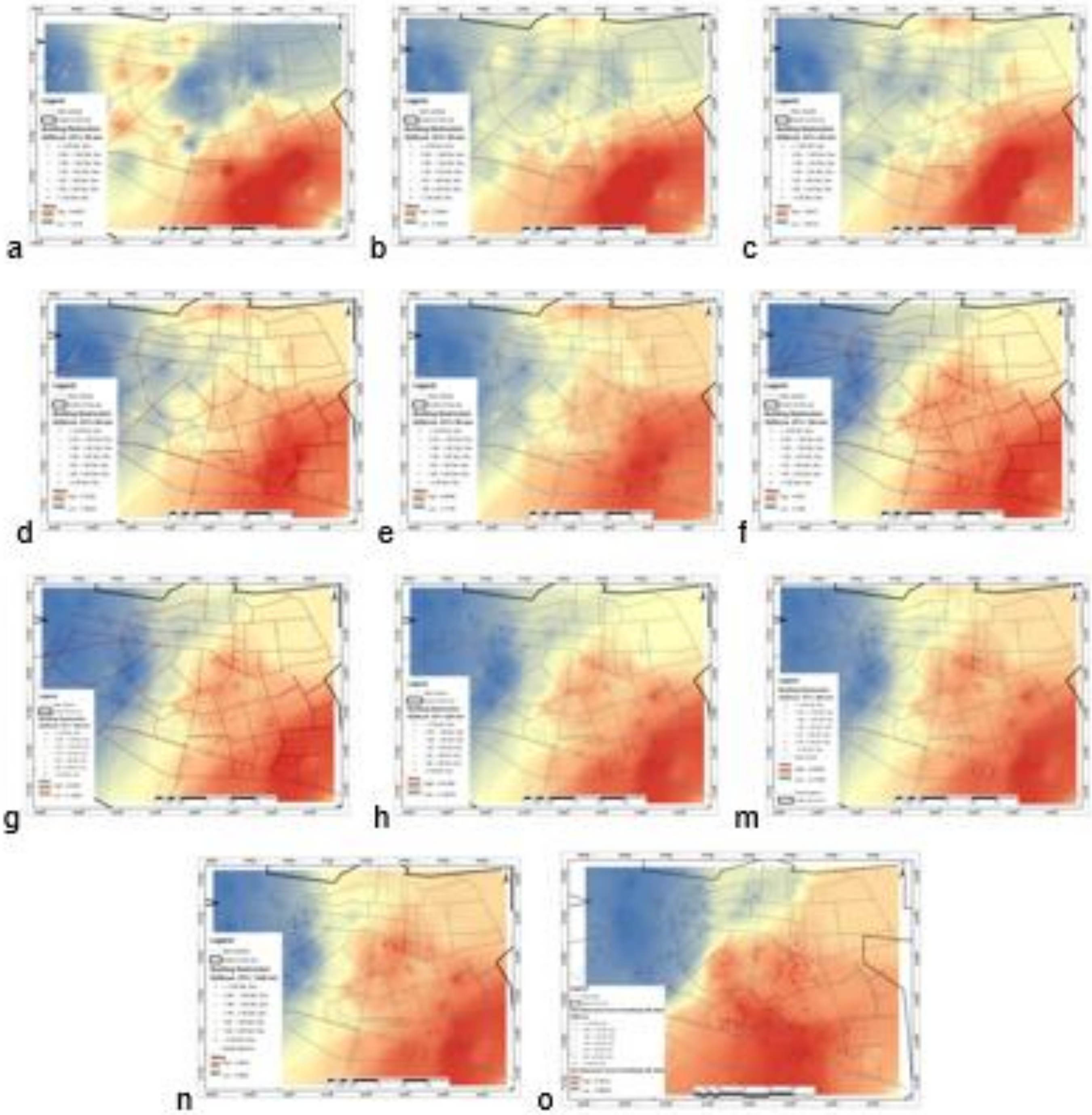
Hot spot identification based on CS data submission time

Hot spot analysis

Getis-Ord geospatial analysis (Ord and Getis, 1995)

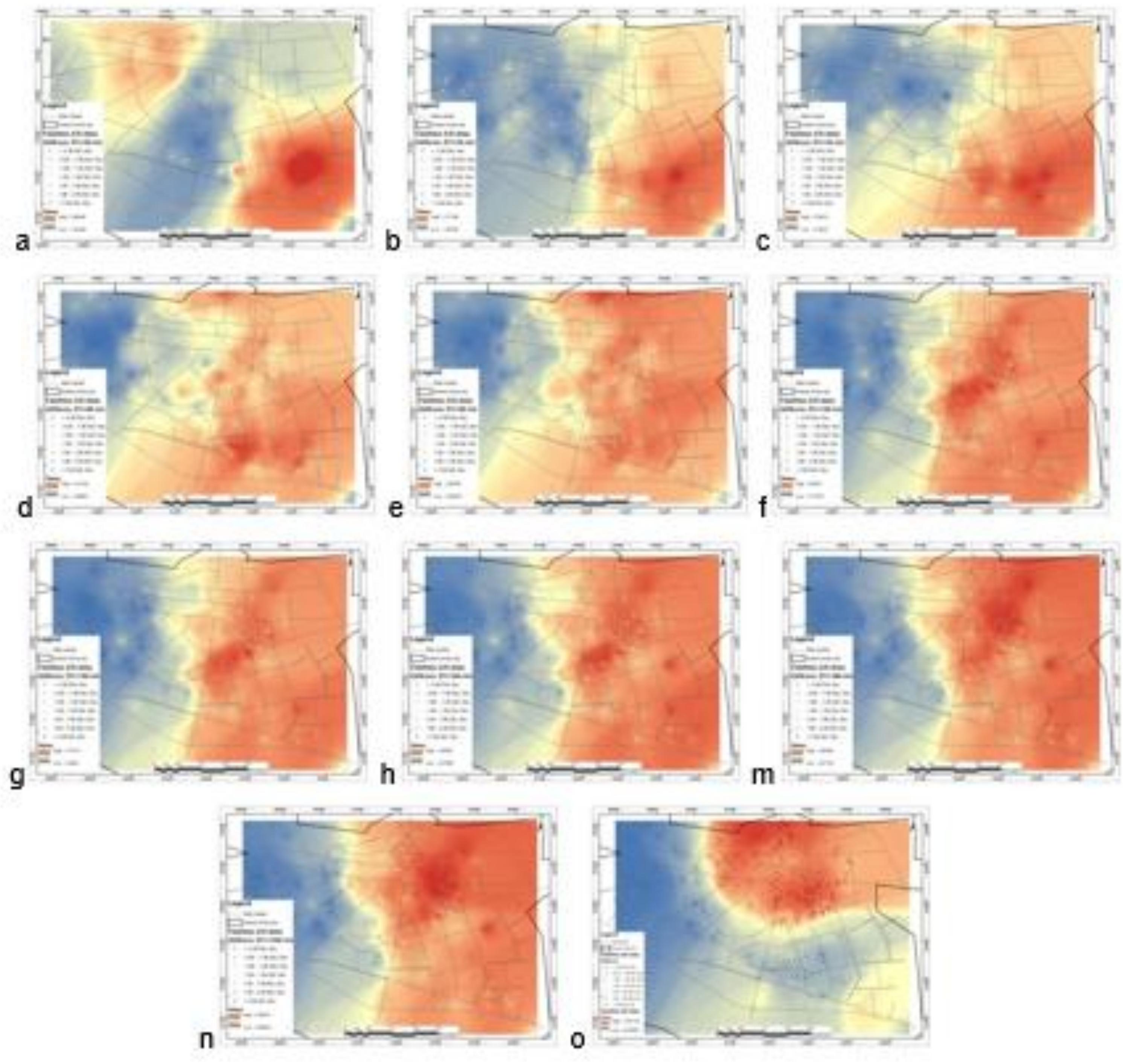
$$G_i^* = \frac{\sum_{j=1}^{n} w_{i,j}\, x_j - \dot{X} \sum_{j=1}^{n} w_{i,j}}{S \sqrt{\frac{\left[n \sum_{j=1}^{n} w_{i,j}^2 - \left(\sum_{j=1}^{n} w_{i,j}\right)^2\right]}{n-1}}}$$

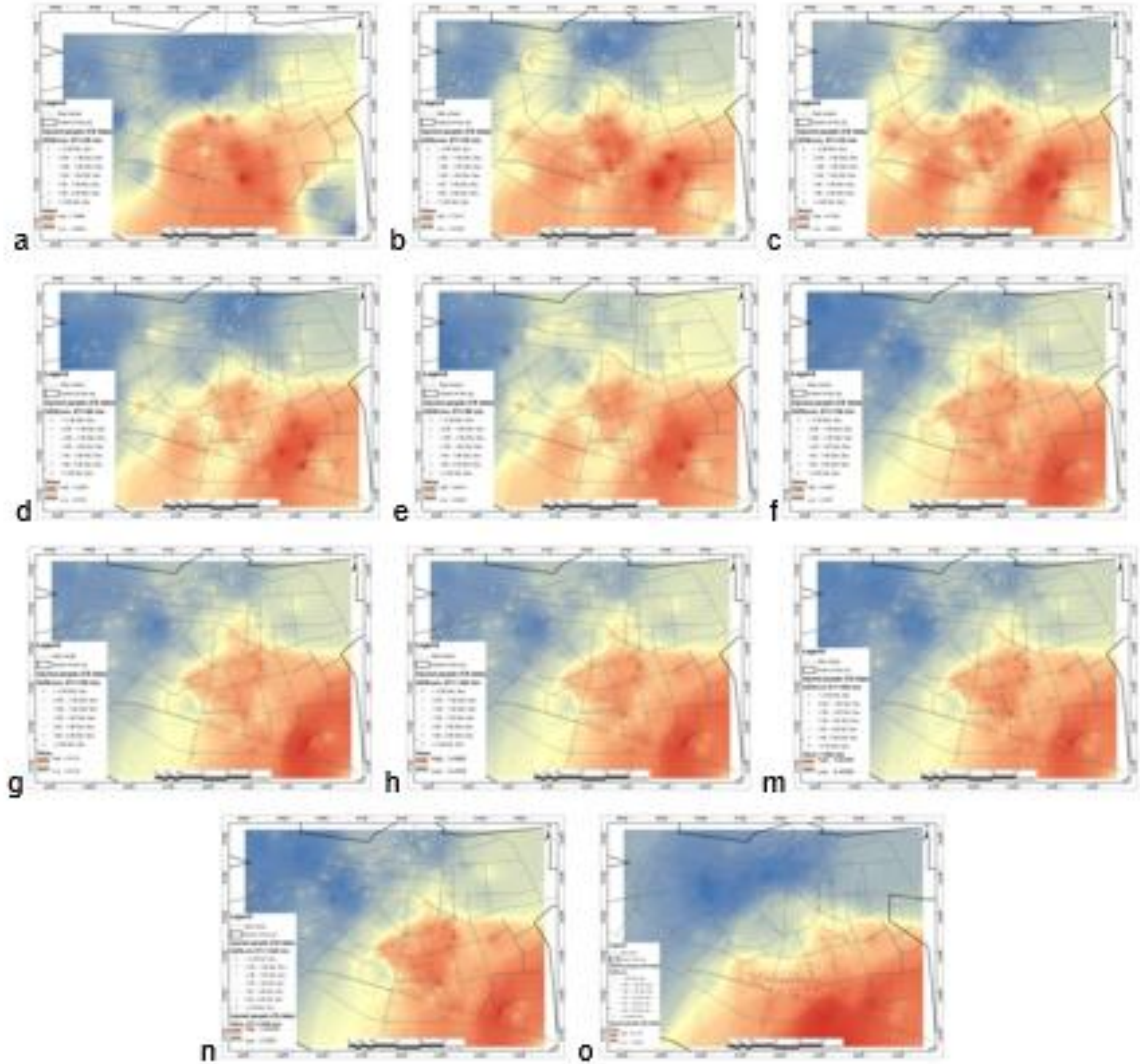
Where:
xj = the attribute value for feature j;
w i,j = the spatial weight between feature i and
j; Ẋ= the average of the attribute value for feature j;
S= standard deviation of the attribute value for feature j; and
Gi= Z score.*



Hot spot maps for the destruction level of buildings based on 10 selected time slots (submission time of CS data after the Bam earthquake) in minutes:
a) 20,
b) 30,
c) 45,
d) 60,
e) 90,
f) 120,
g) 180,
h) 240,
m) 360,
n) 1,440, and
o) based on IND-AE data.

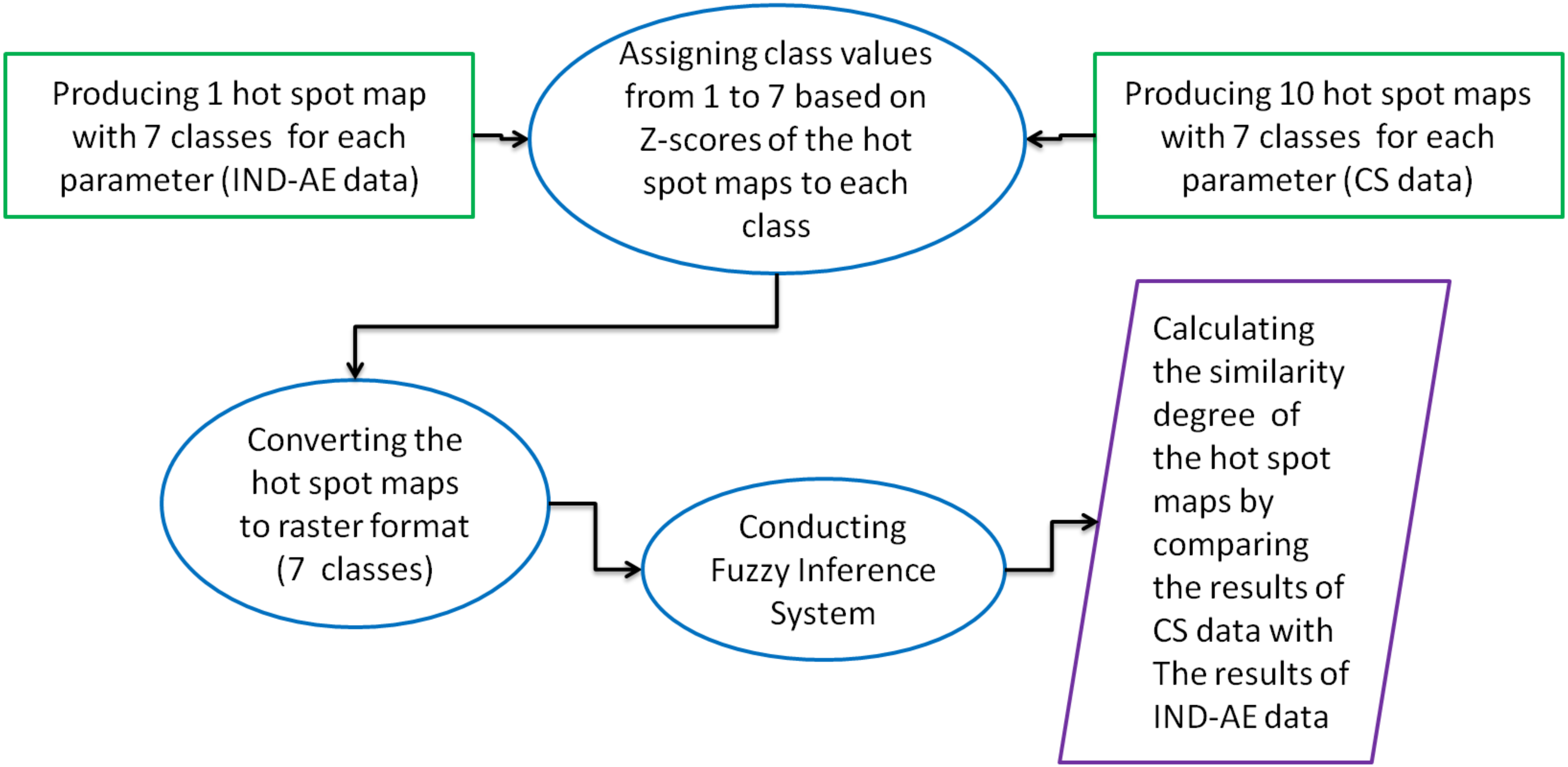
Hot spot maps on the number of people killed (fatalities) based on 10 selected time slots (submission time of CS data after the Bam earthquake) in minutes:
a) 20,
b) 30,
c) 45,
d) 60,
e) 90,
f) 120,
g) 180,
h) 240,
m) 360,
n) 1,440, and
o) based on

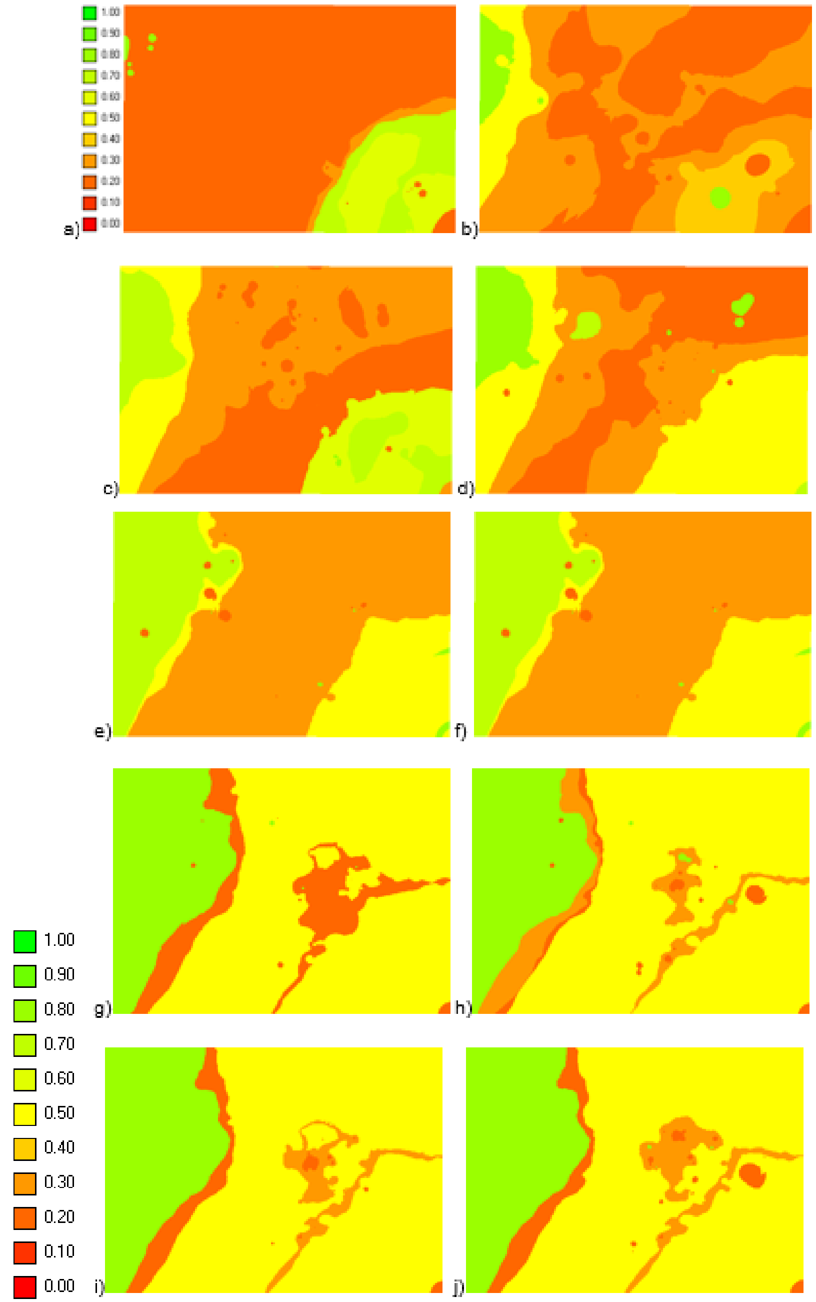




Hot spot maps on the number of injured and hospitalized people based on 10 selected time slots (submission time of CS data after the Bam earthquake) in minutes: a) 20, b) 30, c) 45, d) 60, e) 90, f) 120, g) 180, h) 240, m) 360, n) 1,440, and o) based on IND-AE data.

Accuracy assessment: Fuzzy Inference System





Map comparison based on the Fuzzy Inference System for the destruction level of buildings for 10 time slots in minutes: a) 20, b) 30, c) 45, d) 60, e) 90, f) 120, g) 180, h), 240, i) 360, and j)1440. Similarity index: Red colour- dissimilar = 0 and Green colour- similar = 1

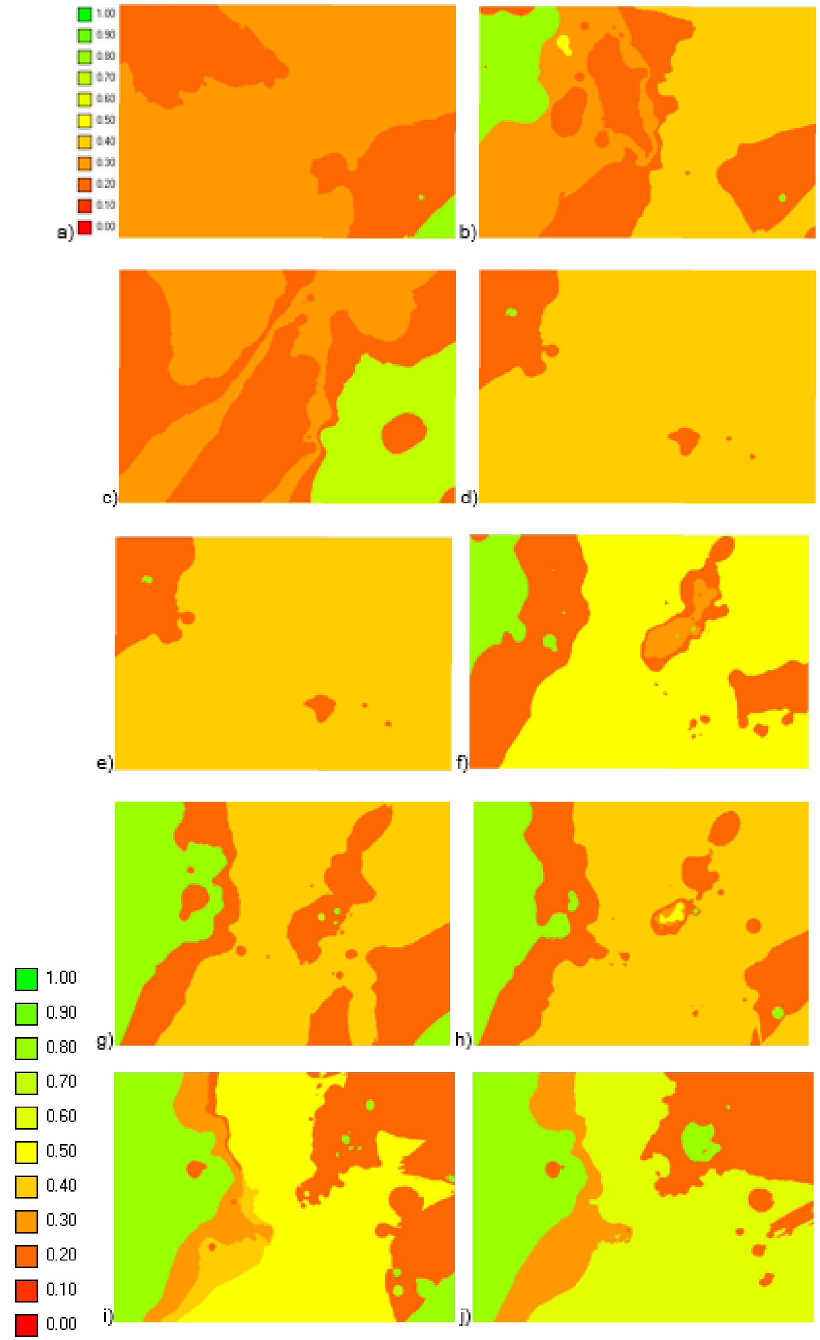
Map comparison based on the Fuzzy Inference System for the number of people killed (fatalities) for 10 time slots in minutes

a) 20, b) 30, c) 45, d) 60, e) 90, f) 120,
g) 180,

h), 240, i) 360, and j)1440. Similarity index:

Red colour- dissimilar = 0 and

Green colour- similar = 1

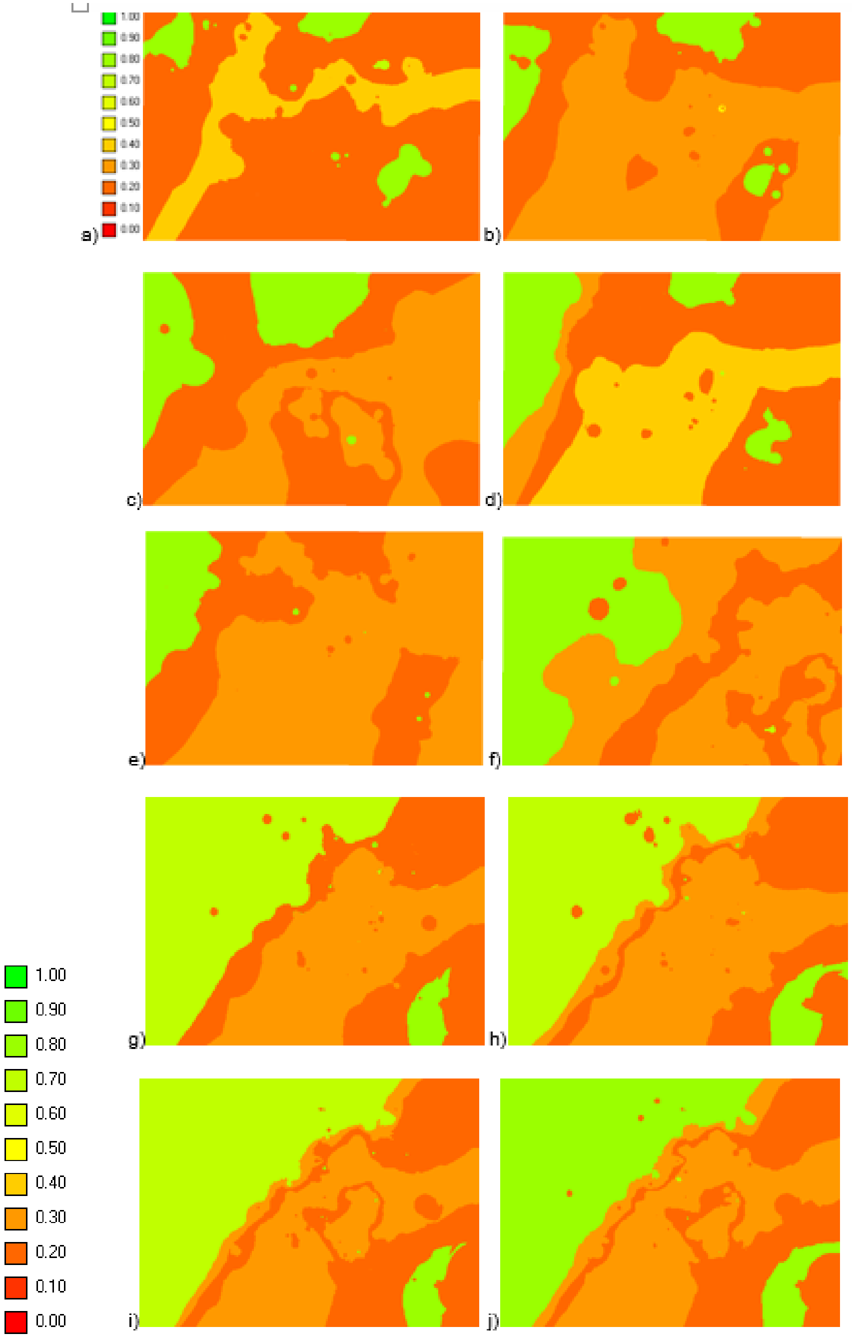


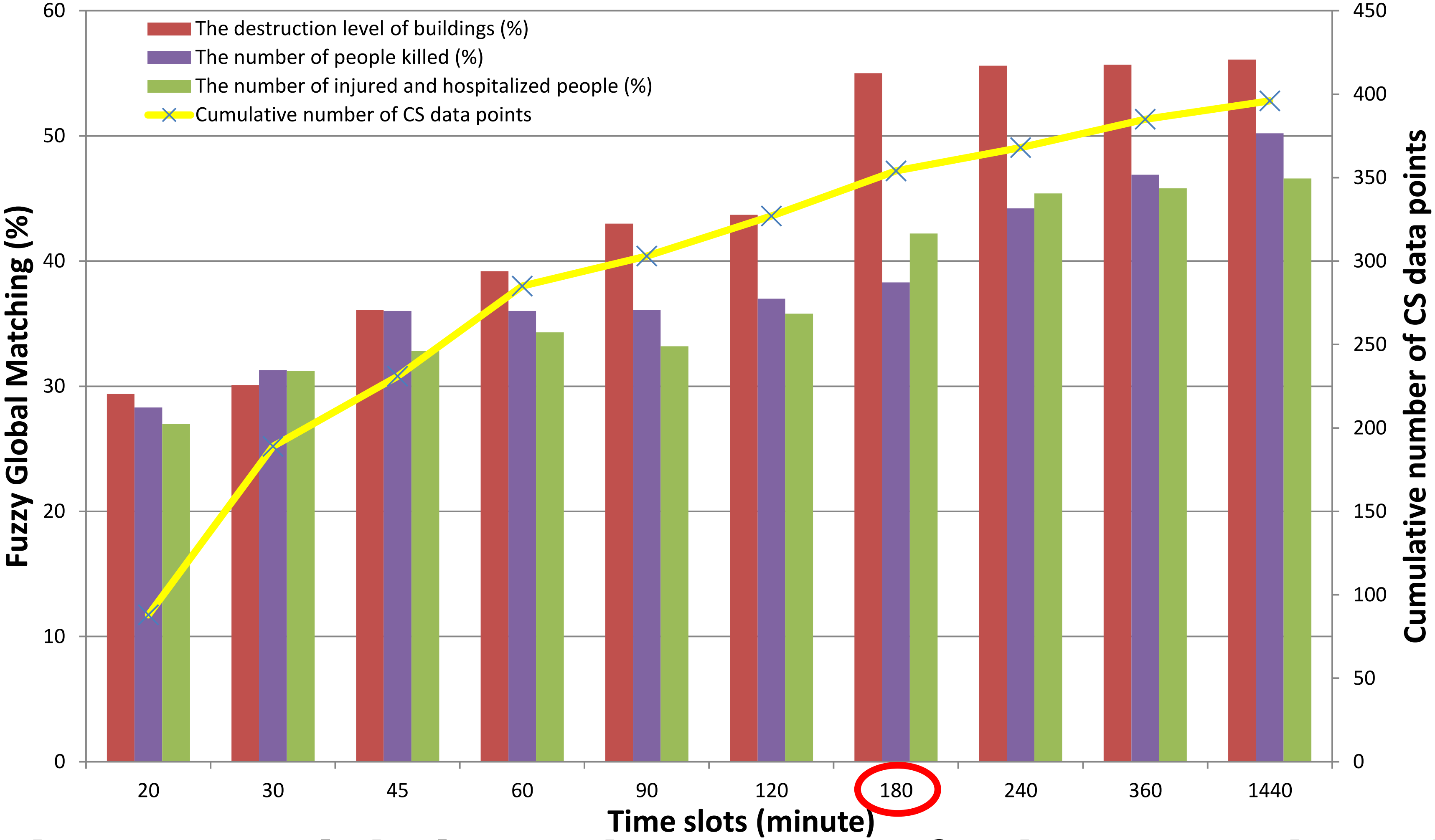
Map comparison based on the Fuzzy Inference System for the number of people injured and hospitalized for 10 time slots in minutes a) 20, b) 30, c) 45, d) 60, e) 90, f) 120, g) 180, h), 240, i) 360, and j)1440.

Similarity index:

Red colour-dissimilar = 0 and

Green colour-similar = 1



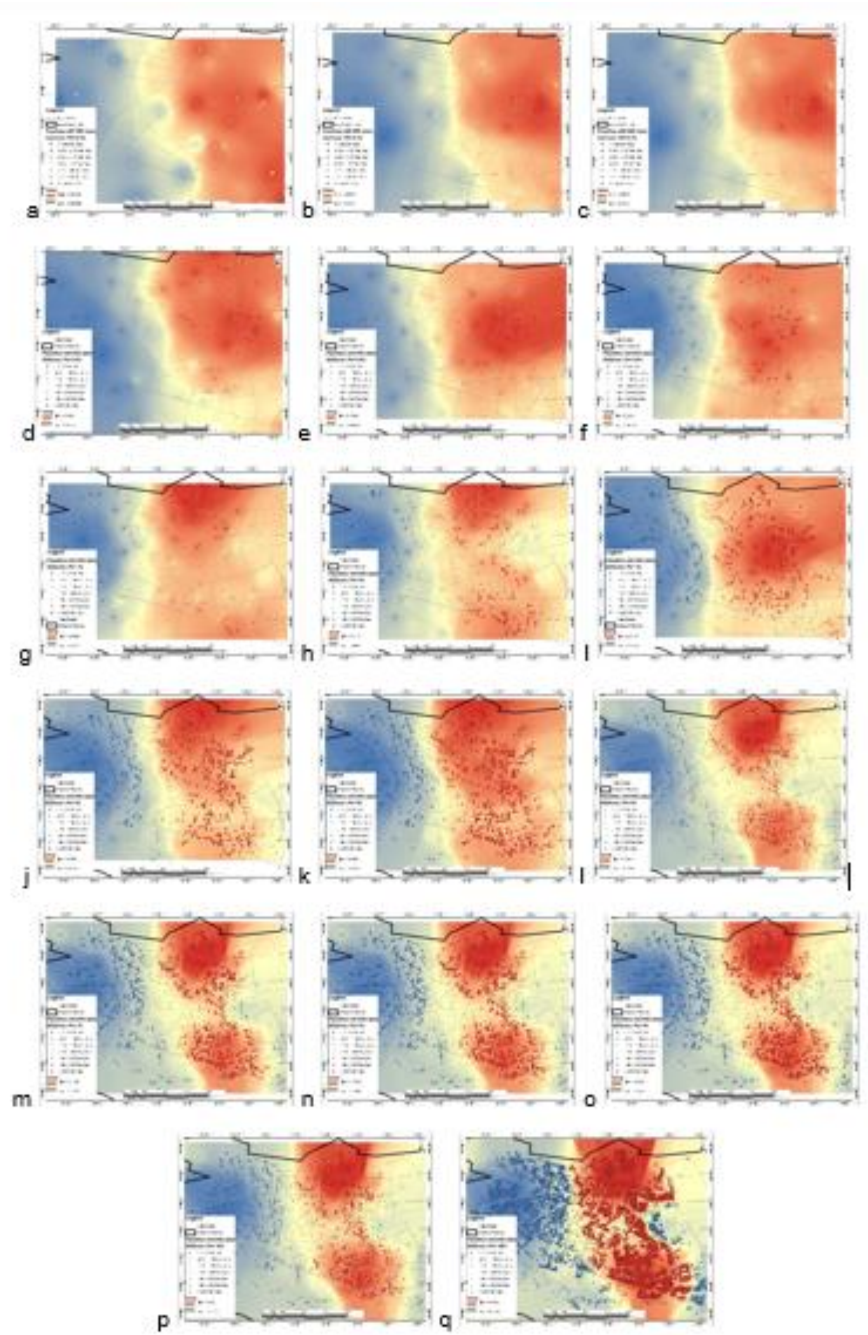


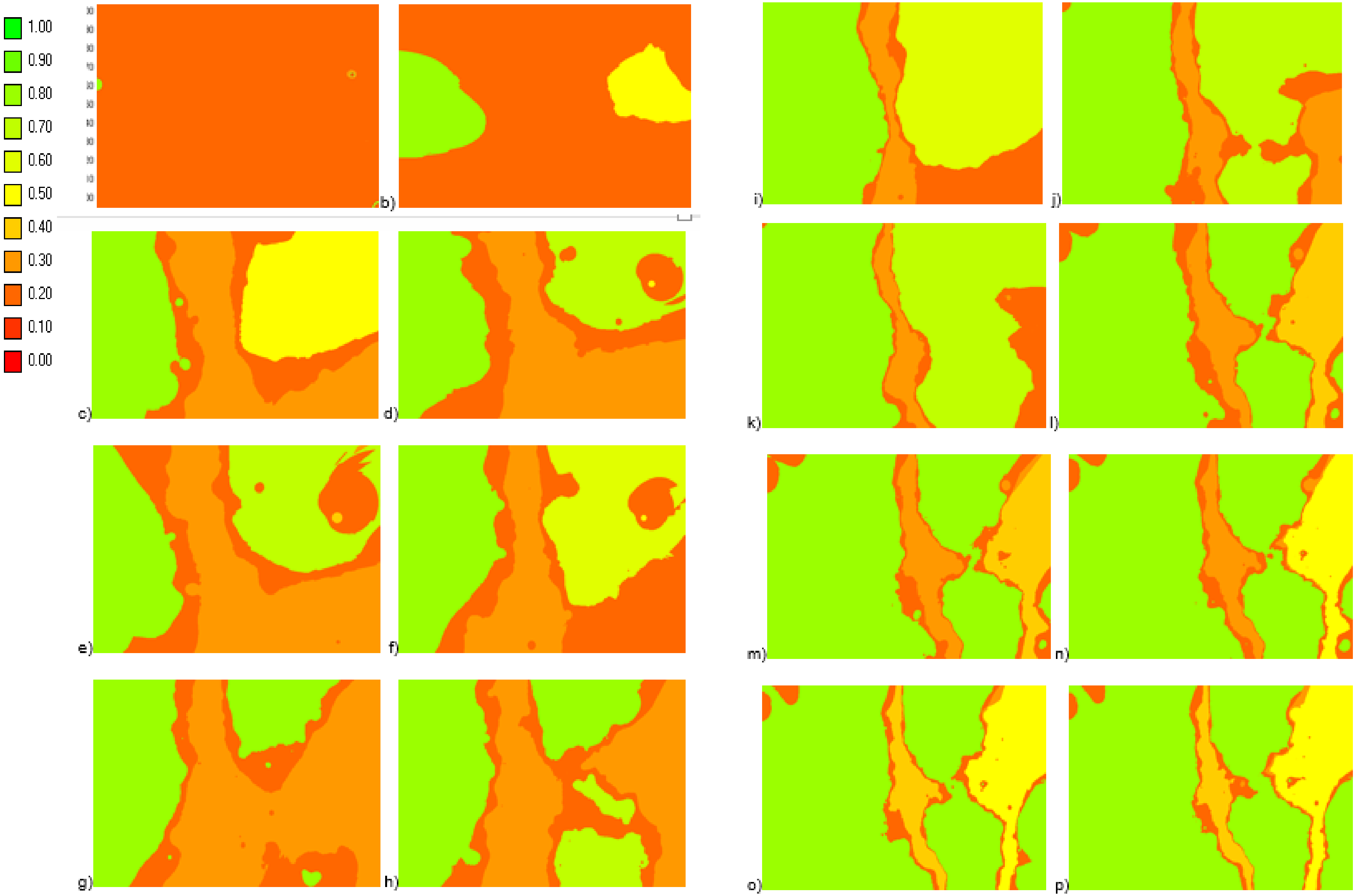
The Fuzzy Global Matching (FGM) for hot spots based on CS data Vs. IND-AE data

How much data we need to make an efficient decision?

Hot spot identification based on the percentage number of IND-AE data

Hot spot maps on the number of people killed based on the percentage number of IND-AE data: a) 0.2%, b) 0.3%, c) 0.4%, d) 0.5%, e) 0.6%, f) 0.8%, g) 1%, h) 2%, i) 3%, j) 4%, k) 5%, l) 6%, m) 7%, n) 8%, o) 9%, p) 10% and q) total number of IND-AE data.



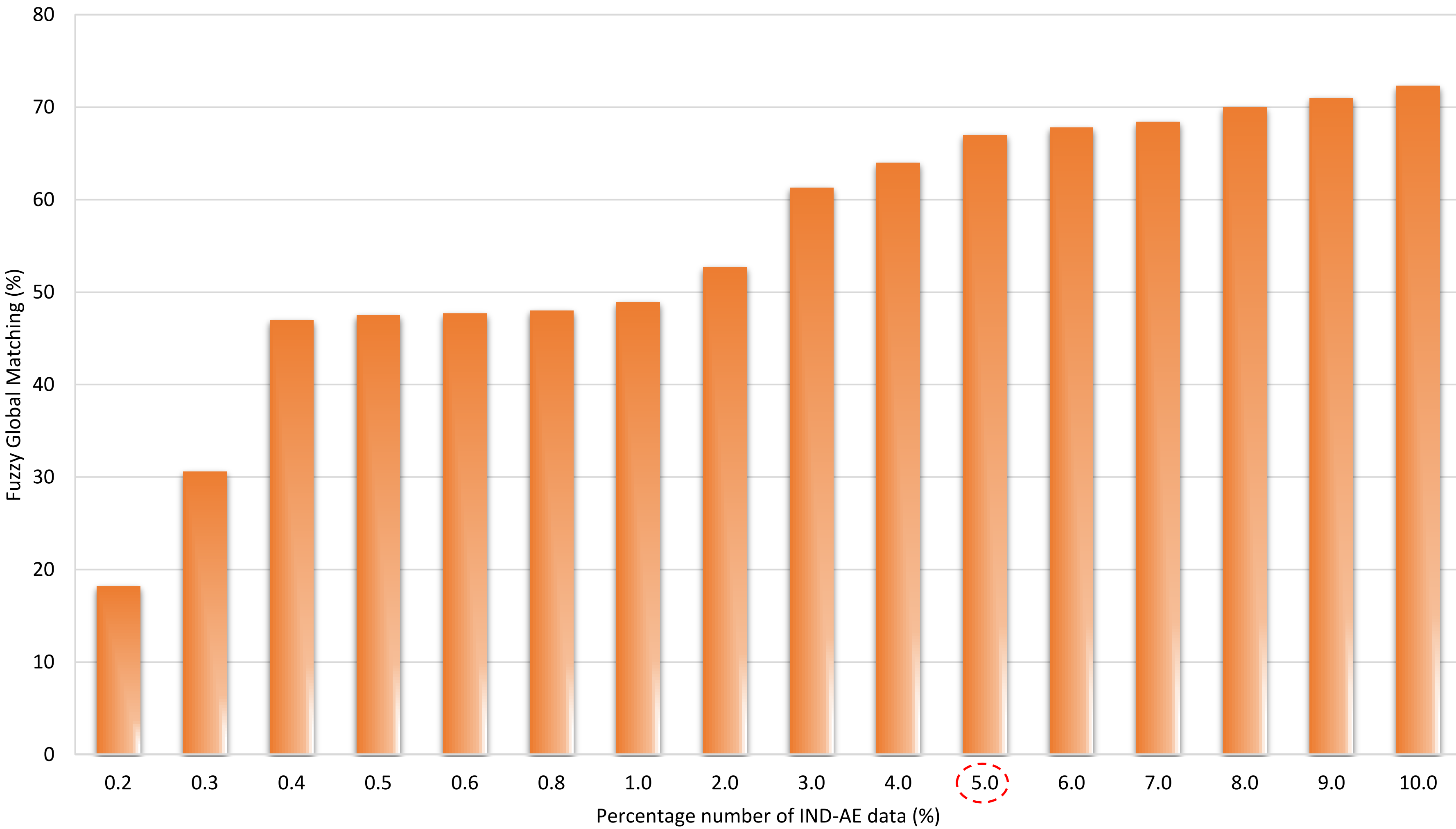


Map comparison based on the Fuzzy Inference System for the number of people killed (fatalities) based on the percentage number of IND-AE data: a) 0.2%, b) 0.3%, c) 0.4%, d) 0.5%, e) 0.6%, f) 0.8%, g) 1%, h) 2%, i) 3%, j)

The Fuzzy Global Matching (FGM) for the number of people killed for 16 rates including 0.2, 0.3, 0.4, 0.5, .06, 0.8, 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10% number of IND-AE data

Percentage	0.10	0.20	0.30	0.40	0.50	0.60	0.80	1.00	2.00
IND-AE data	19	37	56	74	93	111	148	185	370
Fuzzy Global Matching (%)	-	18.2	30.6	47	47.5	47.7	48	48.9	52.7
Percentage	3.00	4.00	5.00	6.00	7.00	8.00	9.00	10.00	
IND-AE data	555	740	926	1111	1296	1481	1666	1851	
Fuzzy Global Matching (%)	61.3	64	67	67.8	68.4	70	71	72.3	

Fuzzy global matching for 10 time slots according to the number of people killed



Discussion and conclusion

- These results demonstrated a sharp increase in the FGM percentage at the time slot of 180 minutes, identifying this time slot as an appropriate cut-off point from which disaster managers could make an efficient decision on the location of hot and cold spot areas in the damaged area.
- The results suggested that 5 to 6 percentage of the total number of data on the households in Bam city was the amount of required information that can help disaster managers in making an efficient decision on the exact locations of hot and cold spots in the damaged area.

Future research

- The main issues for further studies are outlined below:
- The research on CS data application in disaster-response is still at an early stage. Therefore, more studies are needed on the utilization of CS data in earthquake disaster-response activities, based on location and submission time, in order to explore how to apply these data in different settings.
- The design and structure of web-based and mobile applications in facilitating critical CS data collection from the origin point needs further investigation. Such web-based and mobile applications need a structured frame, with defined, pictorial and multiple choice questions in order to improve user-friendliness and the quality of the information provided.
- The integration of formal and informal data are a challenging task in the field of disaster-response. This issue should be addressed in more detail.
- There is a lack of research on the inclusion of CS data reporting into the Community Based Disaster Risk Management (CBDRM) initiative. According to the CBDRM initiative, people learn ways to prepare themselves in order to cope with disasters. Under this initiative, the issue of CS data reporting can be discussed with the community, in terms of what web-based and mobile applications are available, what data to report, and how to report situations that people witness after a disaster. This type of CS data can be called Perceived Crowd Sourced (PCS) data.

The Sendai Framework 2015-2030

Priority 4. Enhancing disaster preparedness for effective response and to "Build Back Better" in recovery, rehabilitation and reconstruction.

- 33(b) Invest in, develop, maintain and strengthen people-centred multi-hazard, multi-sectoral forecasting and early warning systems, disaster risk and emergency communications mechanisms, social technologies and hazard-monitoring telecommunications systems. Develop such systems through a participatory process. Tailor them to the needs of users, including social and cultural requirements, in particular gender. Promote the application of simple and low-cost early warning equipment and facilities and broaden release channels for natural disaster early warning information;

- 33 (f) Train the existing workforce and voluntary workers in disaster response and strengthen technical and logistical capacities to ensure better response in emergencies;

Thank you