

KDE+ method

Clustering of traffic accidents

Richard Andrášik

19th June 2015



Department in Olomouc



<http://tourism.olomouc.eu>

Outline

- **Statement of the problem**
- **KDE+ method**
 - description of the method
 - further modification
 - software realization
- **Results**



Statement of the problem

Research questions

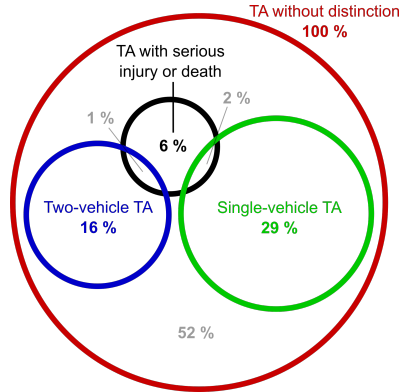
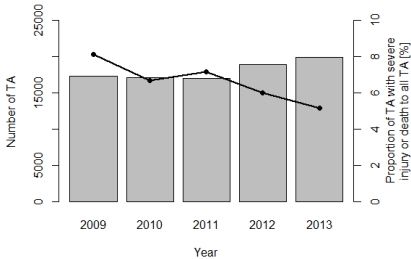
- Are traffic accidents distributed randomly or not?
- Which locations are the most dangerous?
- Which method is suitable for this analysis?



H_0 : Traffic accidents follow random uniform distribution.

Traffic accidents

- Police database
- 2009 – 2013



Methods

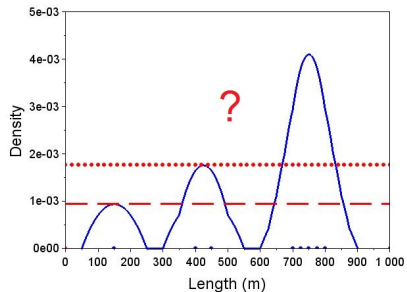
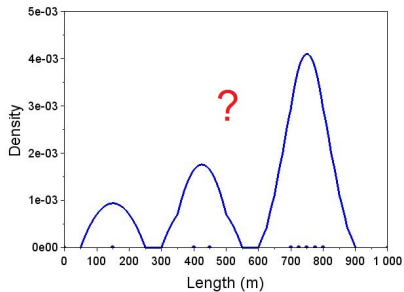
Lord D., Mannering F.: *The Statistical Analysis of Crash-Frequency Data: A Review and Assessment of Methodological Alternatives*, 2010

Geurts K., Wets G.: *Black Spot Analysis Methods: Literature Review*, 2003

- Aggregated counts of traffic accidents
- Empirical Bayes method
- Cluster analysis
 - tendency to clustering (K-function, Nearest neighbor)
 - localization of clusters (KDE, Clumping method)

KDE method

How to define a threshold?



KDE+ method

Accident Analysis and Prevention 55 (2013) 265–273



Contents lists available at SciVerse ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap



Identification of hazardous road locations of traffic accidents by means of kernel density estimation and cluster significance evaluation



Michal Bíl*, Richard Andrášik, Zbyněk Janoška

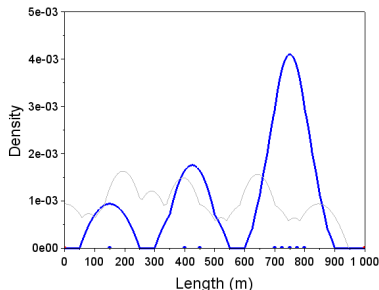
Transport Research Centre, Lžeňská 33a, Brno, 636 00, Czech Republic

Testing significance

H_0 : Traffic accidents follow random uniform distribution.

- Monte Carlo method
- Repeated simulations
- Threshold = 95th percentile

One significant cluster

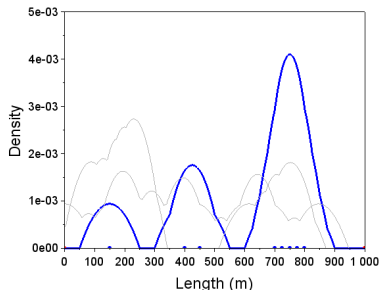


Testing significance

H_0 : Traffic accidents follow random uniform distribution.

- Monte Carlo method
- Repeated simulations
- Threshold = 95th percentile

One significant cluster

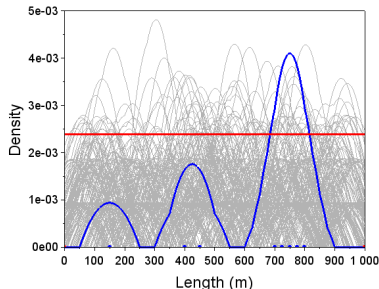


Testing significance

H_0 : Traffic accidents follow random uniform distribution.

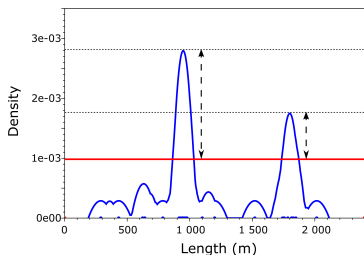
One significant cluster

- Monte Carlo method
- Repeated simulations
- Threshold = 95th percentile



Cluster strength

cluster strength : significant cluster $\mapsto (0; 1)$

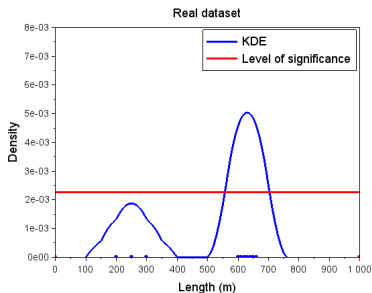


- relative measure
- degree of H_0 violation

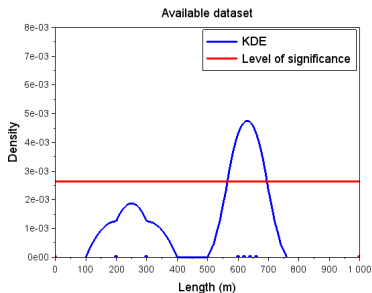
→ ordering of clusters

Stability

- missing data
- highly inaccurate data

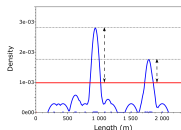
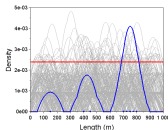


10 traffic accidents

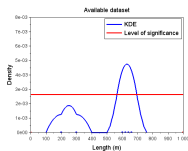
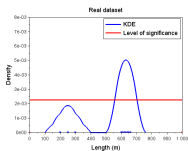


6 traffic accidents

Comparisons



	Localization	Significance test	Ordering	Stability
KDE	+	-	-	+
Dangerousness index	+	+	-	+
K-function	-	-	-	+
Hierarchical clustering	+	-	-	+
Clumping method	+	+	-	-
KDE+	+	+	+	+

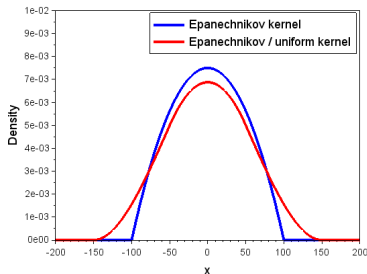


Further modification

Error in the data

- linear referencing system (Catalonia)
- snapping error (Sweden)

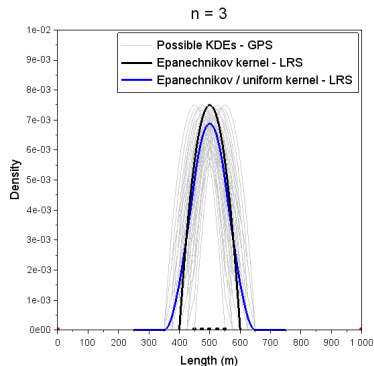
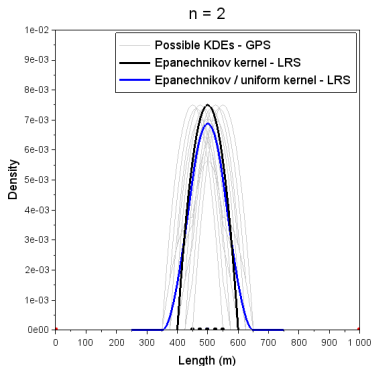
⇒ change your kernel



bandwidth = 100 m
error = 50 m

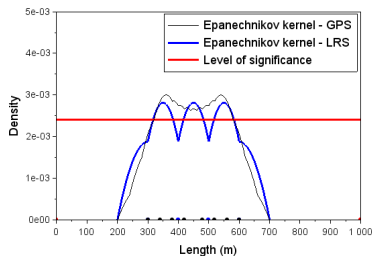
Modified kernel

Epanechnikov / uniform kernel
takes into account all possible outcomes!

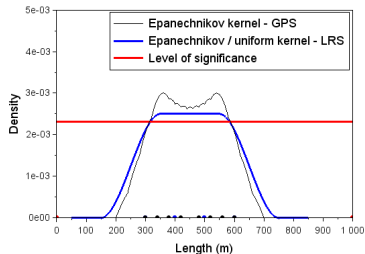


Modified kernel

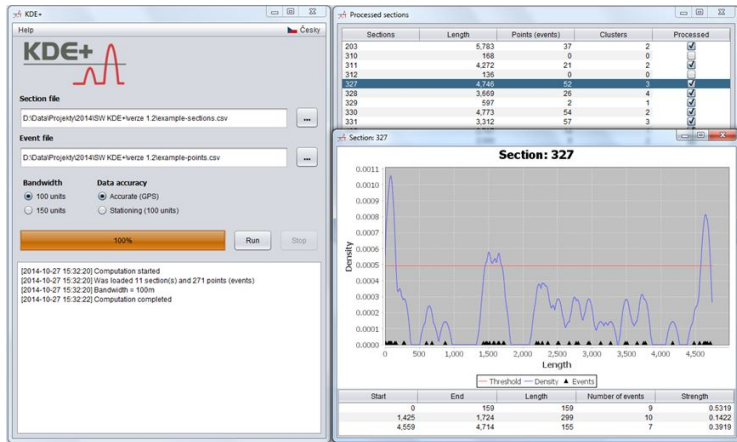
Epanechnikov kernel



Epanechnikov / uniform kernel

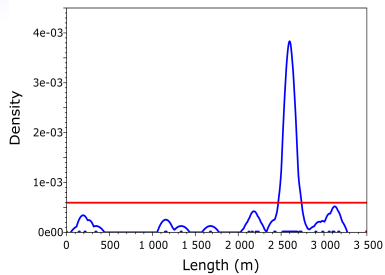
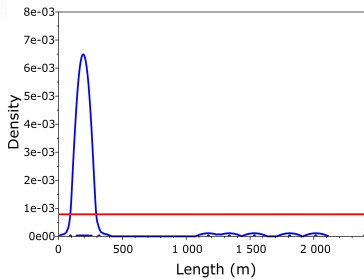


Software



www.kdeplus.cz/en

Results

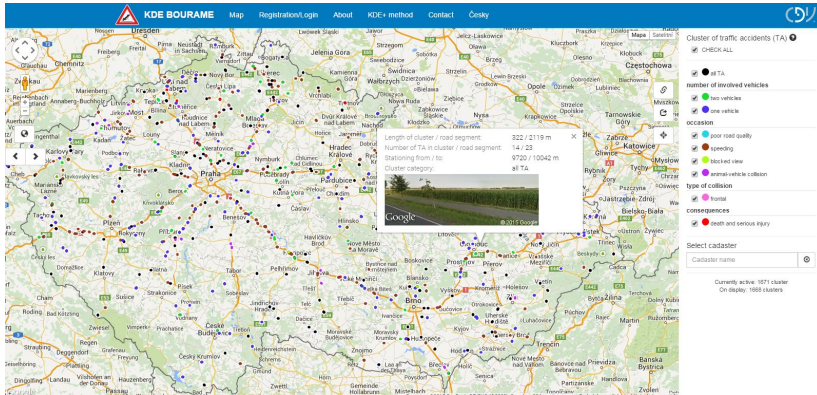


Random or pattern occurrence?



Type of TA	Single-vehicle	Severe injury or death
Number of TA	59811	5953
Number of clusters	6555	406
Number of TA in clusters [%]	39.9	15.3
Total length of clusters [km]	740	29
Total length of clusters [%]	1.98	0.08
Mean length of clusters [m]	113	70

Where we crash?



www.kdebourame.cz/en

Conclusion

Research questions

- Are traffic accidents distributed randomly or not?
- Which locations are the most dangerous?
- Which method is suitable for this analysis?

Answers

- global factors x local factors
- we found them
- KDE+ method

Conclusion

Research questions

- Are traffic accidents distributed randomly or not?
- Which locations are the most dangerous?
- Which method is suitable for this analysis?

Answers

- **global factors x local factors**
- we found them
- KDE+ method

Conclusion

Research questions

- Are traffic accidents distributed randomly or not?
- Which locations are the most dangerous?
- Which method is suitable for this analysis?

Answers

- global factors x local factors
- **we found them**
- KDE+ method

Conclusion

Research questions

- Are traffic accidents distributed randomly or not?
- Which locations are the most dangerous?
- Which method is suitable for this analysis?

Answers

- global factors x local factors
- we found them
- **KDE+ method**

Thanks for your attention!

e-mail: andrasik.richard@gmail.com

