Analyzing the Spatio-Temporal Patterns and Impacts of Large-Scale Events in OpenStreetMap

Yair Grinberger, Moritz Schott, Martin Raifer, Rafael Troilo, Alexander Zipf

UNIVERSITÄT HEIDELBERG
ZUKUNFT
SEIT 1386

GIScience

Alexander von Humboldt
Stiftung / Foundation

האוניברסיטה העברית בירושלים
THE HEBREW UNIVERSITY OF JERUSALEM

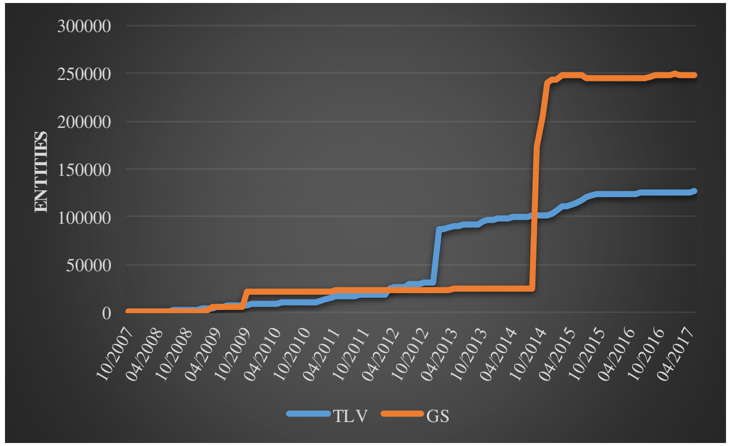
The Vision of Volunteered Geographical Information

- The vision of VGI – democratized and bottom-up geo-data production (Goodchild, 2007)
- The evolution of the vision:
  - Participation and data bias (Haklay, 2016)
  - Considering **process with product** (Sieber & Haklay, 2015)
  - **Contextual effects** on data (Fast & Rinner, 2014)
- OpenStreetMap is rich in contextual effects:
  - Mapping platforms
  - Interaction platforms (wiki, mailing lists, …)
  - Activity of organizations (Anderson et al., 2019; Palen et al., 2015; Poiani et al., 2016)
  - **Data events**

Editathon-Portland by Lxbarth / CC BY-SA 2.0

(Large-Scale Data) Events in OpenStreetMap

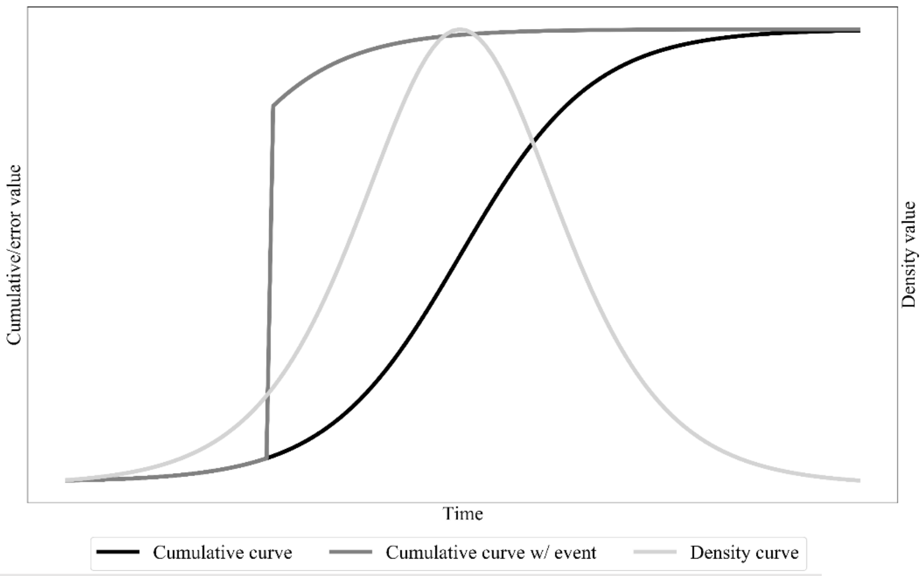
- Defining events in OSM:
  - The social perspective (Juhász & Hochmair, 2018; Mooney et al., 2015)
  - **The data perspective** (Eckle & Albuquerque, 2015; Zielstra et al., 2013)
- Large-scale data events:
  - Can create lasting impacts on data and community
  - **High** volume of contributions over a **short** period
  - **Significantly** affect the data
- The current study:
  - Identifies events which show a **significant change**
  - Analyzes spatio-temporal **patterns**
  - Studies **impacts**



Source: Grinberger, 2018

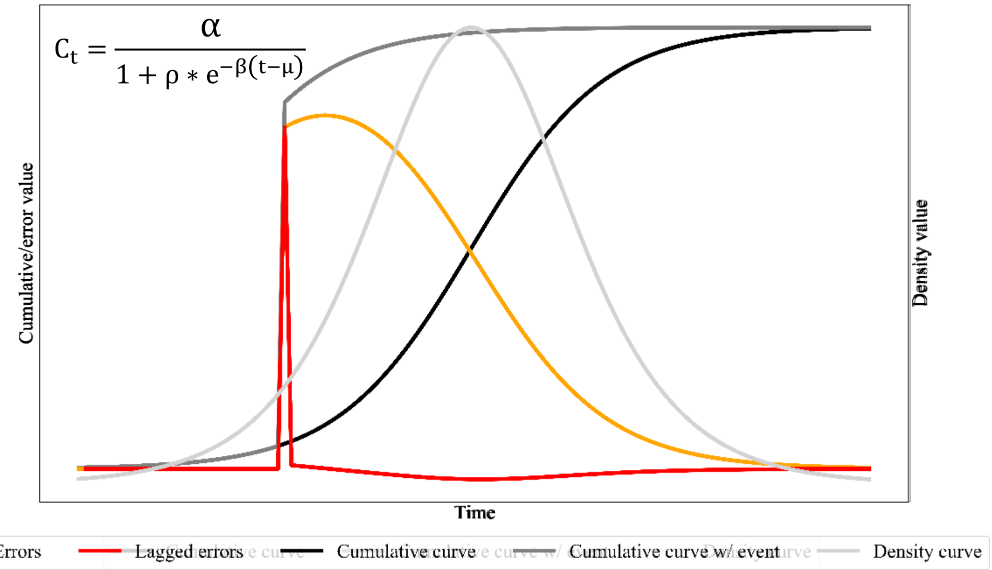
Identifying Large-Scale Events

- Assumption – a 'normative' model of data production (Gröching et al., 2014)
- Definition – events are sharp increases not predicted by the model
- Procedure:
  - Create cumulative series of **contribution actions** over time
  - Fit a logistic curve to the time series



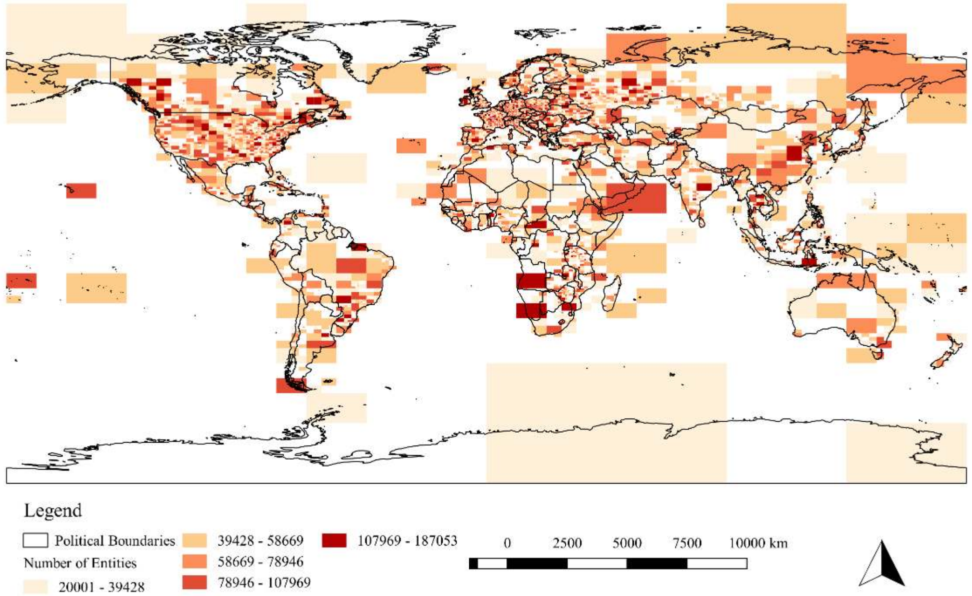
Identifying Large-Scale Events

- Assumption – a 'normative' model of data production (Gröching et al., 2014)
- Definition – events are sharp increases not predicted by the model
- Procedure:
  - Create cumulative series of **contribution actions** over time
  - Fit a logistic curve to the time series
  - Compute lagged residuals
  - Find significant **positive residuals**



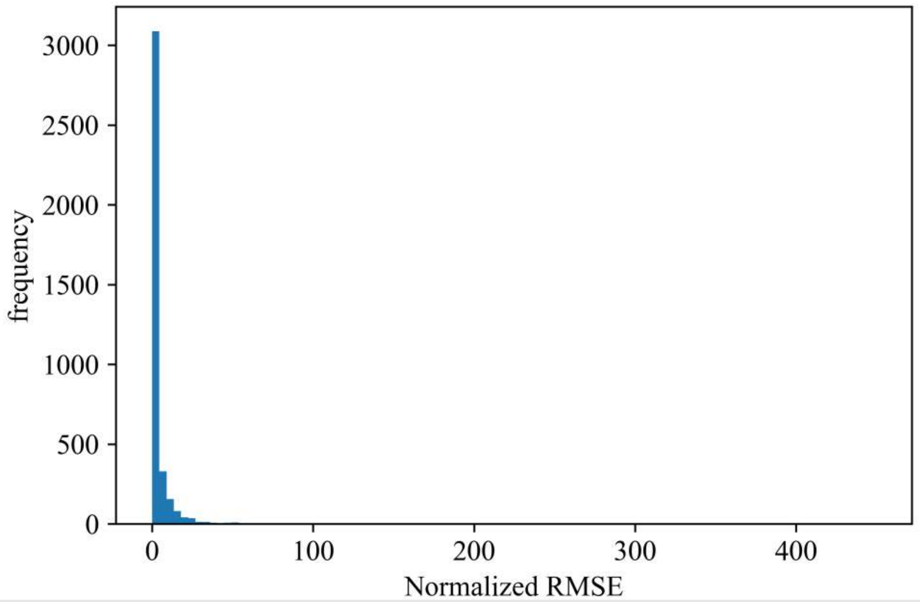
Data Extraction

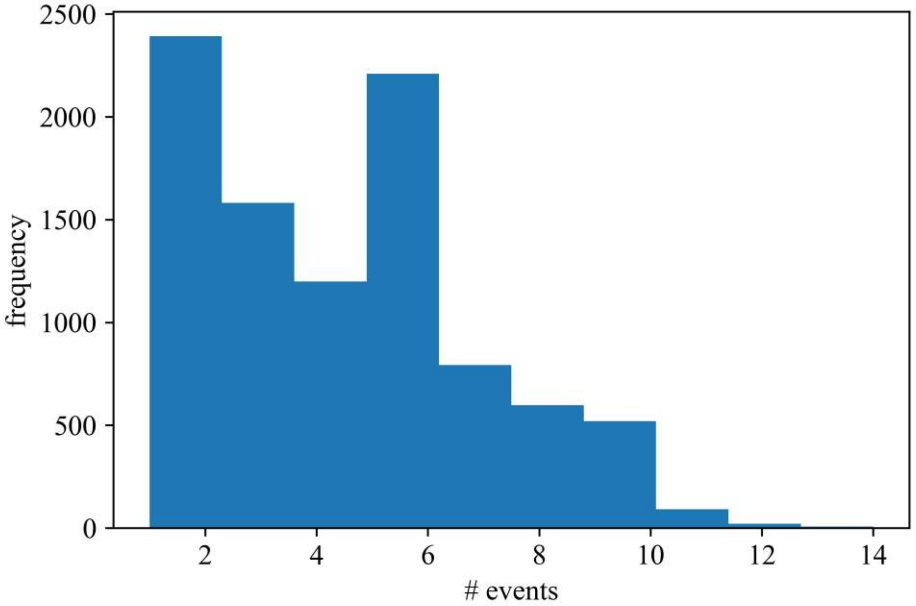
- Quad-tree spatial division by number of OSM entities
- Temporal resolution – one month
- Time period: 11-2007 to 03-2019
- Number of actions extracted using the OSHDB tool
  (Raifer et al., 2019)
- Additional variables:
  - Active users
  - No. of contributions by type
  - Maximal no. of actions by one user
  - Actions per edited entity



Events

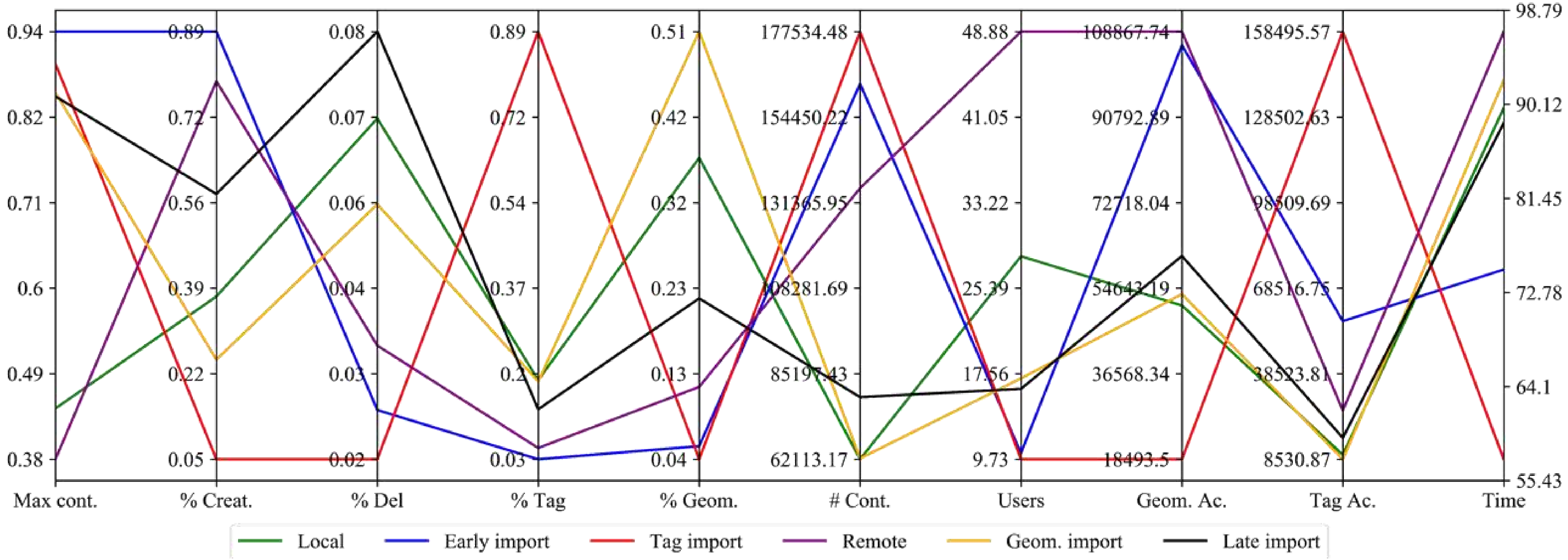
- Convergence errors for 700 cells (6.91%)
- Considered only events with no. of actions > 7,000
- 48,653 events identified
- Median of 5.00 events per cell (average: 5.16, std: 2.72)





Classifying Events

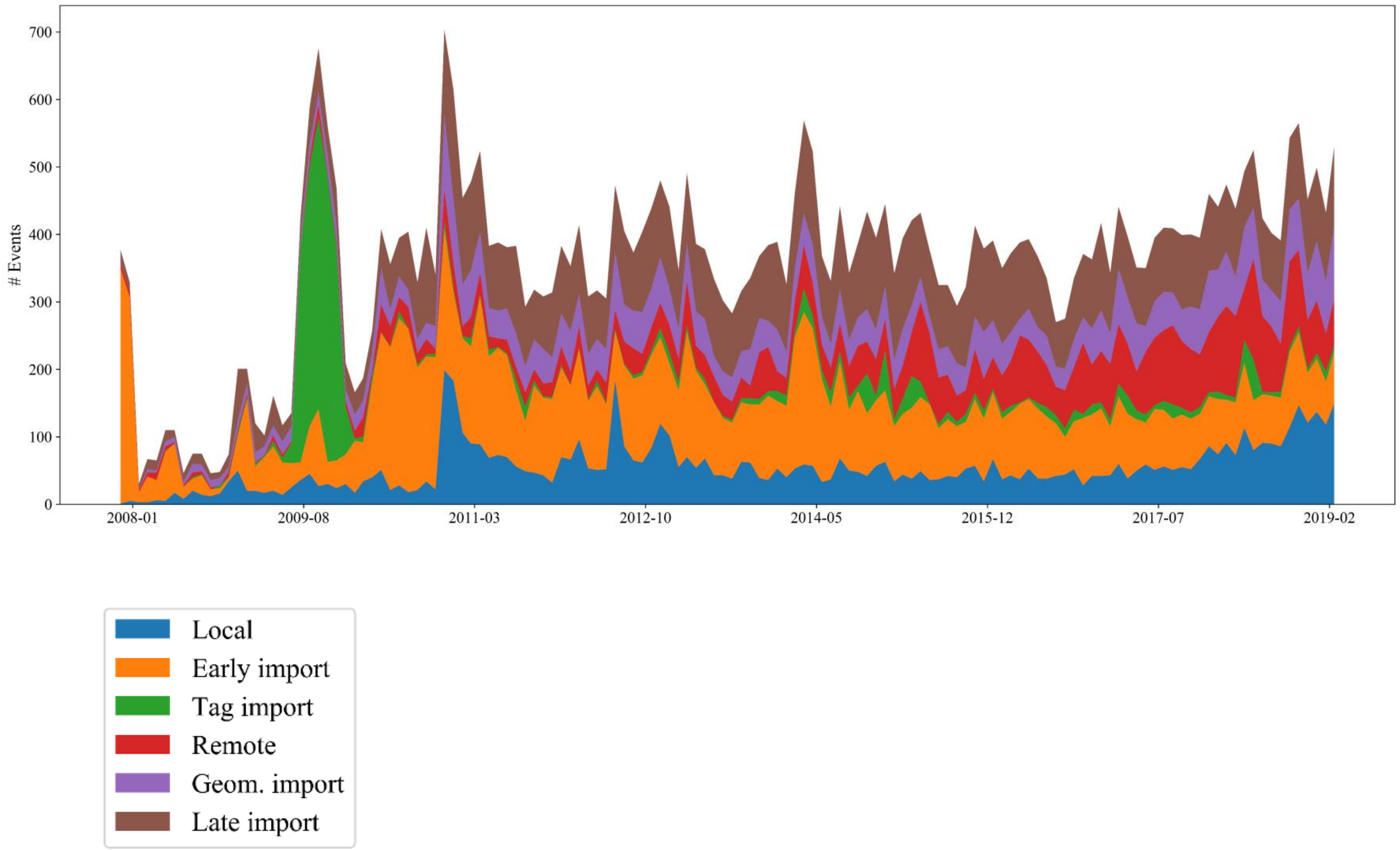
- K-means procedure used to differentiate between events (K=6)
- Variables used:
  - contributions by type (% of all contributions)
  - maximal volume of contribution by one user (% of all contributions)



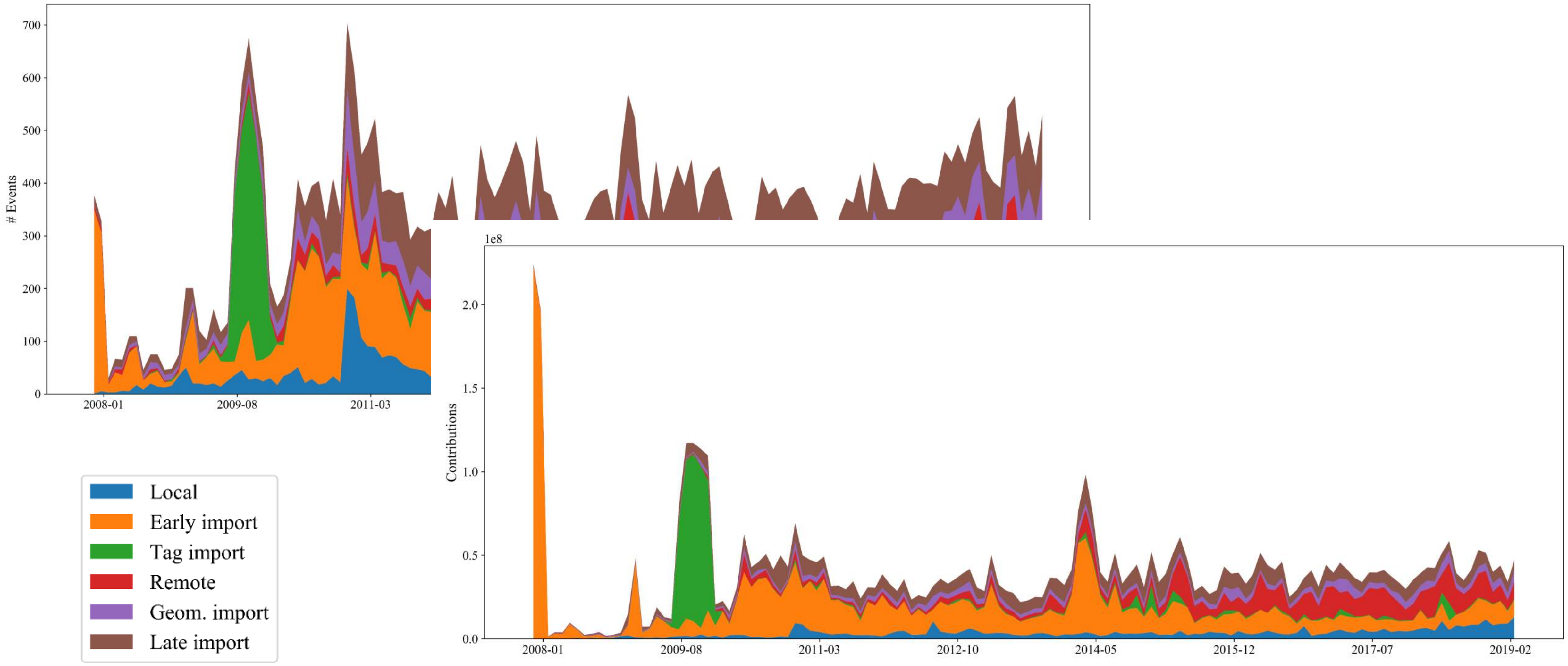
Weight of Events (% of All Actions/Contributions)

	# Events	Act. (Mil.)	Act.	Geom. Ac.	Tag Ac.	Creations	Deletions	Tag Chan.	Geom. Chan.
All	48653.0	5468.89	40.42	38.94	48.62	45.52	30.65	35.98	21.07
Local	7394.0	459.26	3.39	4.03	2.03	2.54	5.79	2.68	5.84
Early import	14080.0	2301.58	17.01	15.94	21.85	25.48	5.65	0.81	1.69
Tag import	3216.0	570.95	4.22	0.64	13.89	0.36	1.34	26.42	0.71
Remote	6145.0	831.19	6.14	7.15	4.29	9.42	3.59	0.82	2.52
Geom. import	6008.0	374.57	2.77	3.43	1.4	1.09	2.0	2.2	6.22
Late import	11810.0	931.34	6.88	7.75	5.16	6.65	12.29	3.05	4.09

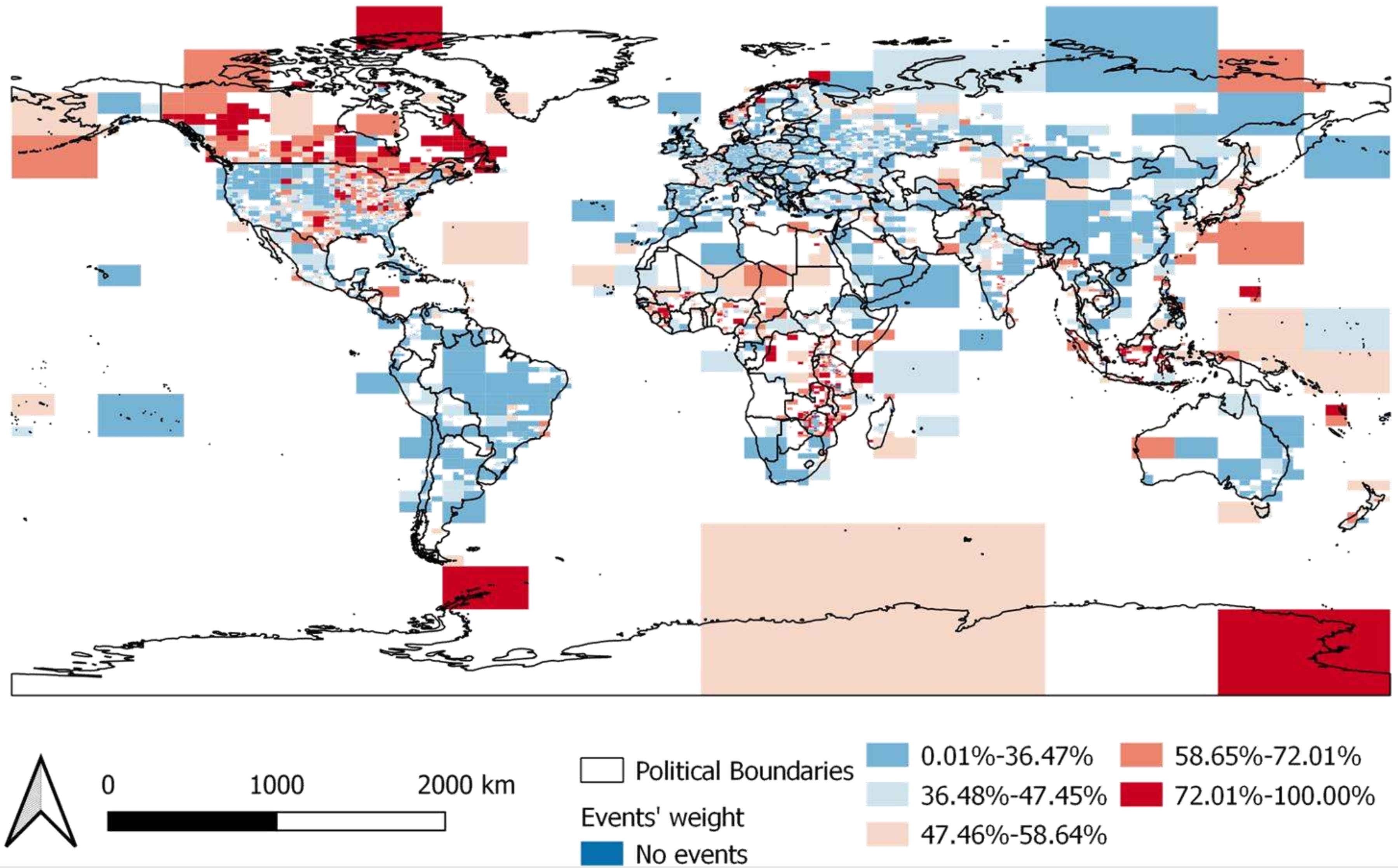
Temporal Patterns



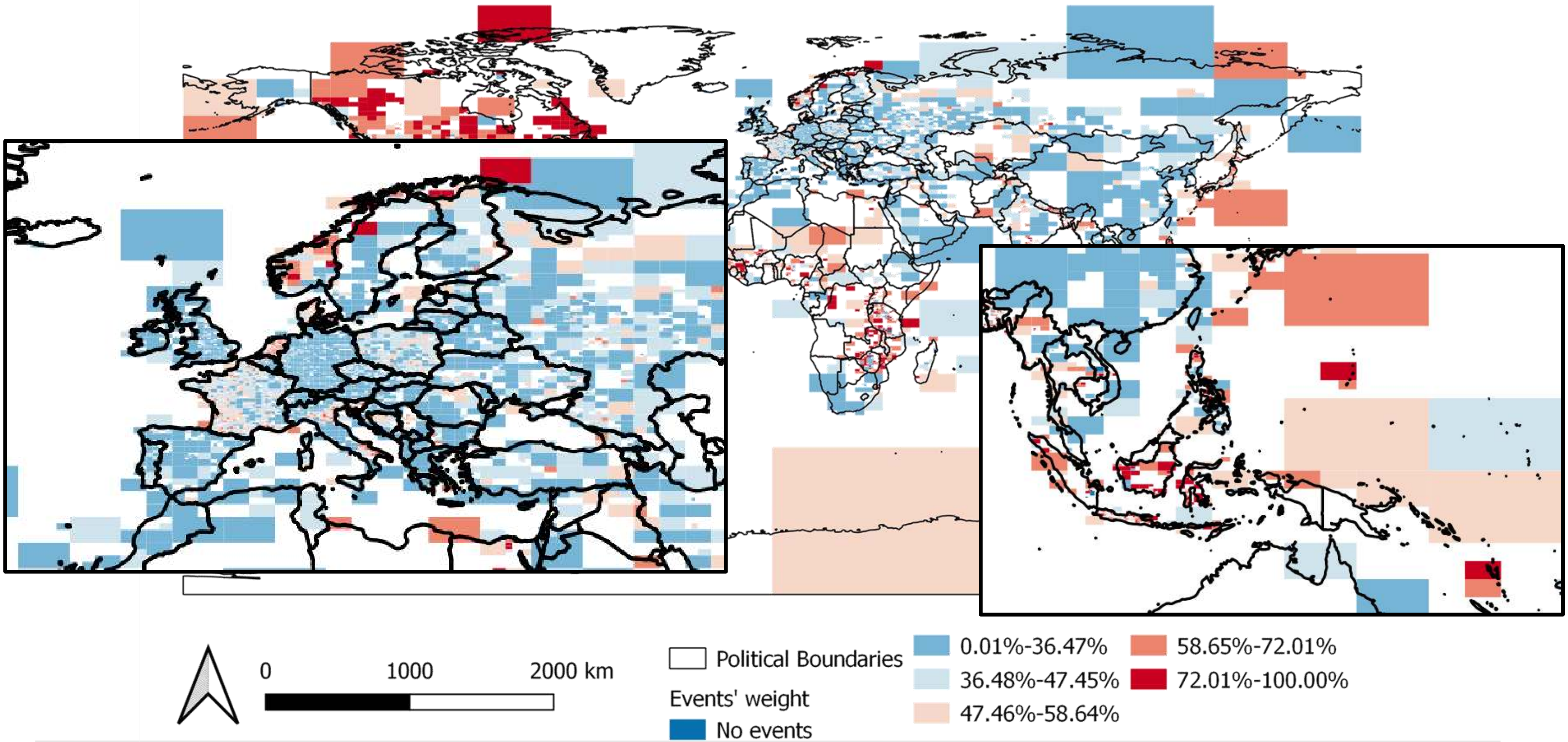
Temporal Patterns



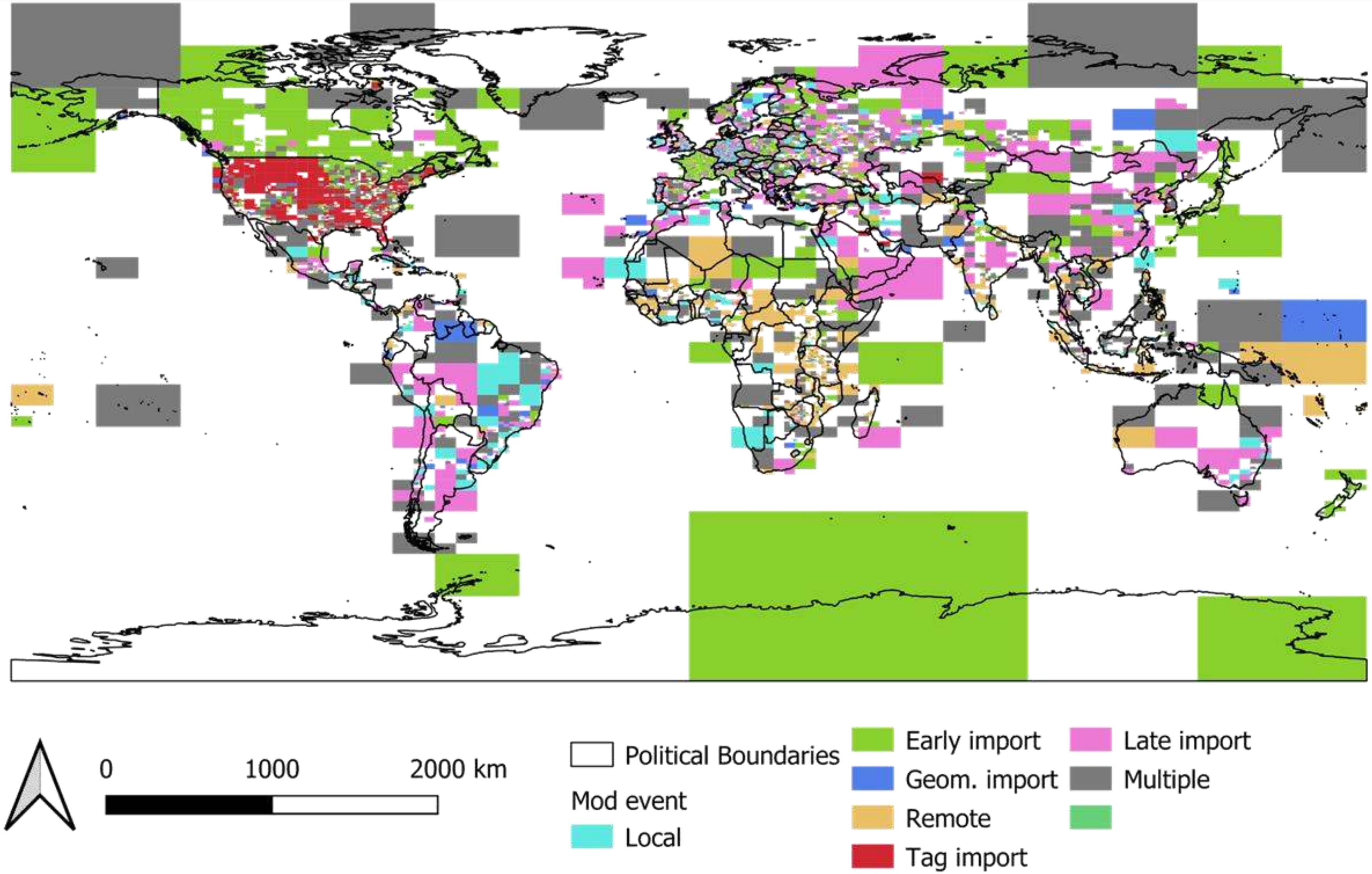
Spatial Patterns – Events' Weights



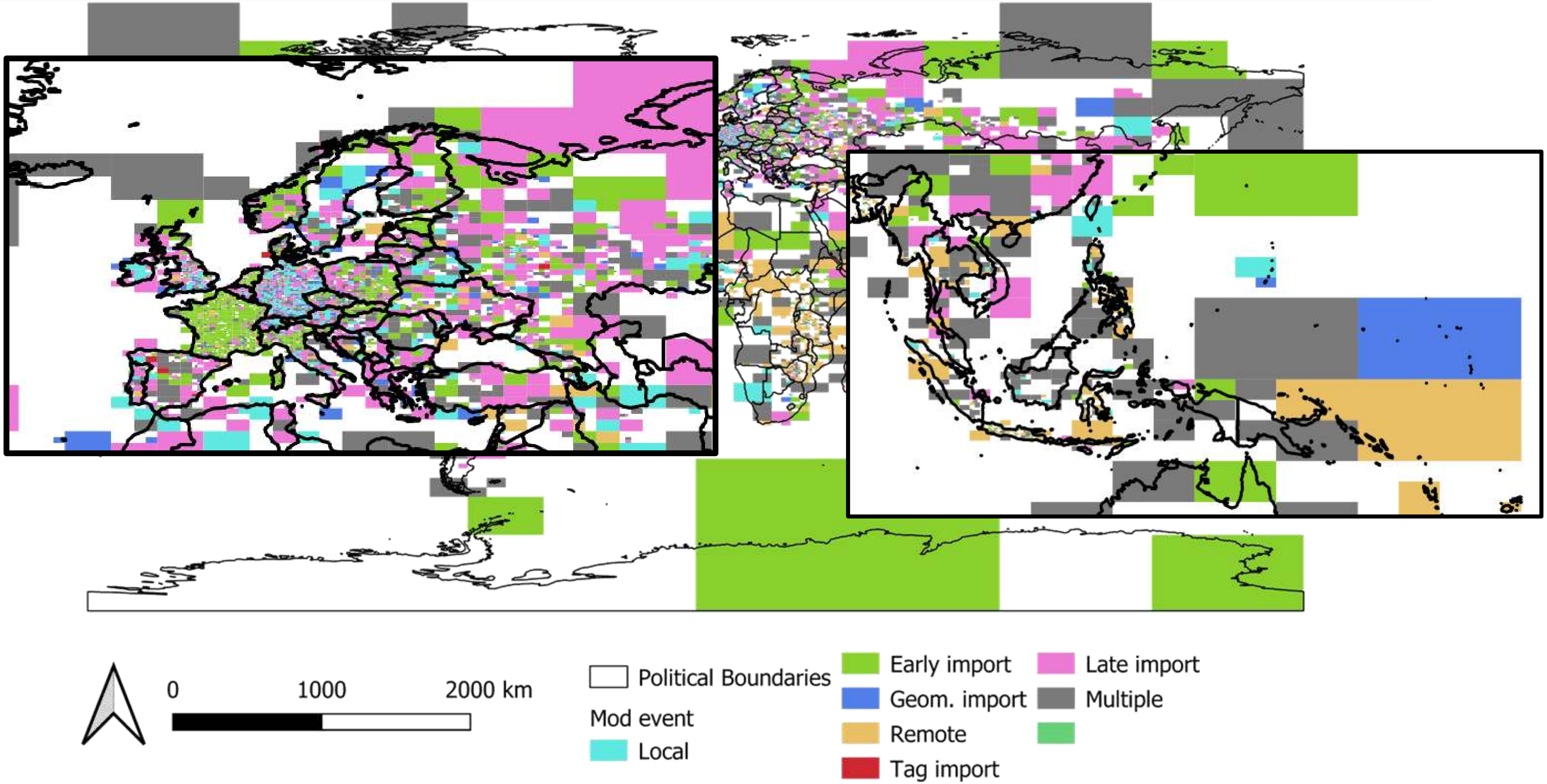
Spatial Patterns – Events' Weights



Spatial Patterns – Most Common Event Type



Spatial Patterns – Most Common Event



Events' Effects on Activity (6 months)

	# events	Actions	Users	Geom. Act.	Tag Act.	Creations	Deletions	Tag Cha.	Geom. Cha.
Control	897441.0	8.34	7.47	9.74	8.4	7.56	2.93	11.63	11.16
All	14623.0	12.96	7.69	10.86	12.31	4.74	20.94	16.84	15.12
Local	1840.0	9.56	6.69	9.33	11.8	5.8	8.46	14.27	10.34
Early import	4683.0	18.07	8.94	12.86	17.61	3.79	34.44	27.75	20.4
Tag import	888.0	-1.04	13.64	-5.51	-2.6	-15.2	25.49	4.08	7.12
Remote	1660.0	31.02	17.44	26.55	36.99	22.33	103.96	53.5	44.76
Geom. import	2041.0	12.2	4.95	11.84	7.14	6.85	1.27	6.06	9.54
Late import	3511.0	7.92	4.62	7.61	8.59	2.0	13.08	8.82	8.8

Events' Effects on Activity (12 months)

	# events	Actions	Users	Geom. Act.	Tag Act.	Creations	Deletions	Tag Cha.	Geom. Cha.
Control	619144.0	15.4	13.79	17.88	14.23	16.67	6.25	15.3	20.46
All	8223.0	16.79	13.08	14.91	16.04	9.37	21.96	23.09	21.24
Local	919.0	15.18	11.68	15.29	14.94	13.33	6.85	15.72	13.72
Early import	2729.0	17.75	13.81	12.82	20.63	6.18	32.67	34.48	31.64
Tag import	606.0	9.75	18.78	9.52	3.29	-13.28	31.44	19.7	19.41
Remote	922.0	41.25	28.5	39.21	58.39	43.8	107.93	93.03	57.62
Geom. import	1135.0	12.54	7.49	13.58	5.89	10.97	-2.57	3.66	10.54
Late import	1912.0	13.44	11.02	14.03	10.35	5.55	16.73	12.38	13.12

First Event and Probability for Future Events

	# events	Local	Early import	Tag import	Remote	Geom. import	Late import
Local	871.0	70.72	33.18	11.48	31.0	51.21	71.3
Early import	4047.0	19.08	66.17	26.19	25.95	24.64	42.43
Tag import	448.0	19.2	48.21	68.75	13.62	35.27	33.26
Remote	1662.0	36.4	29.18	6.26	54.15	16.43	24.79
Geom. import	636.0	60.22	42.92	15.25	28.93	62.11	75.63
Late import	1747.0	44.25	59.24	12.54	33.14	40.98	72.24

Conclusions

- Large-scale data events affect OSM in a meaningful way
- They are contextual **products** with contextual **impacts**:
  - Shifting trends related to the **maturity** of the data/community
  - …but with **socio-geographical variability**
  - They may serve as a means for **exploration**
  - May **adversely** affect activity, but wrongs can make a right!
- Considering context as part of the production of events
- **Further analysis**:
  - Stability of event contributions
  - Tagging schemes during and after events
  - Changes in communities' structures

Thank You!

yair.grinberger@mail.huji.ac.il

UNIVERSITÄT HEIDELBERG ZUKUNFT SEIT 1386

GIScience

Alexander von Humboldt Stiftung/Foundation

האוניברסיטה העברית בירושלים
THE HEBREW UNIVERSITY OF JERUSALEM

References

- Anderson, J., Sarkar, D. and Palen, L. (2019). Corporate mappers in the evolving landscape of OpenStreetMap. *ISPRS International Journal of Geo-Information*, *8*(5), 232.
- Eckle, M. and Albuquerque, J. P. d. (2015). Quality assessment of remote mapping in OpenStreetMap for disaster management perspectives. In: Palen, Büscher, Comens and Hughes (eds.) *Proceedings of the ISCRAM 2015 Conference*. Kristiansand, May 24-27.
- Fast, V. and Rinner, C. (2014). A systems perspective on volunteered geographic information. *ISPRS International Journal of Geo-Information*, *3*(4), 1278-1292.
- Goodchild, M. F. (2007). Citizens as sensors: The world of volunteered geography. *GeoJournal*, *69*(4), 211-221.
- Grinberger, A. Y. (2018). Identifying the effects of mobility domains on VGI: Towards an analytical approach. *Short Paper Presented at the VGI-ALIVE Agile 2018 Workshop,* Lund, 12-15 June 2018.
- Gröching, S., Brunauer, R. and Rehrl K. (2014). Digging into the history of VGI data-sets: Results from a worldwide study on OpenStreetMap mapping activity. *Journal of Location Bassed Services*, *8*(3), 198-210.
- Haklay, M. (2016). Why is participation inequality important? In: Capineri, C., Haklay, M., Huang, H., Antoniou, V. Kettunen, J., Ostermann, F. & Purves, R. (eds.) *European Handbook of Crowdsourced Geographic Information*. London, Ubiquity Press, pp. 35-44.
- Juhász, L., and Hochmair, H. (2018). OSM data import as an outreach tool to trigger community growth? A case study in Miami. *ISPRS International Journal of Geo-Information,* *7*(3), 113.
- Mooney, P., Minghini, M., & Stanley-Jones, F. (2015). Observations on an OpenStreetMap party organized as a social event during an open source GIS conference. *International Journal of Spatial Data Infrastructure Research, 10*, 138-150.
- Palen, L., Soden, R., Anderson, T. J. and Barrenechea, M. (2015). Success & scale in a data-producing organization: The socio-technical evolution of OpenStreetMap in response to humanitarian events. In: Mayer, T. & Do, E. Y.-L. (eds.) *Proceedings of the 33rd Annual CHI Conference on Human Factors in Computing Systems.* New York, The Association for Computing Machinery, 4113-4122.
- Poiani, T. H., Rocha, R. d. S., Degrossi, L. C. and Albuquerque, J. P. d. (2016). Potential of collaborative mapping for disaster relief: A case study of OpenStreetMap in the Nepal earthquake 2015, *2016 49th Hawaii International Conference on System Sciences (HICSS)*, Koloa, HI, pp. 188-197.
- Raifer, M., Troilo, R., Kowatsch, F., Auer, M., Loos, L., Marx, S., Przybill, K., Fendrich, S., Mocnik, F.-B. and Zipf. A. (2019). OSHDB: A framework for spatio-temporal analysis of OpenStreetMap history data. *Open Geospatial Data, Software and Standards, 4*(3), 1-12.
- Sieber, R. E. and Haklay, M. (2012). The epistemology(s) of volunteered geographic information: A critique. *Geo: Geography and Environment*, *2*(2), 122-136.
- Zielstra, D., Hochmair, H., H. and Neis, P. (2013). Assessing the effect of data imports on the completeness of OpenStreetMap – A United States case study. *Transactions in GIS*, *17*(3), 315-334.