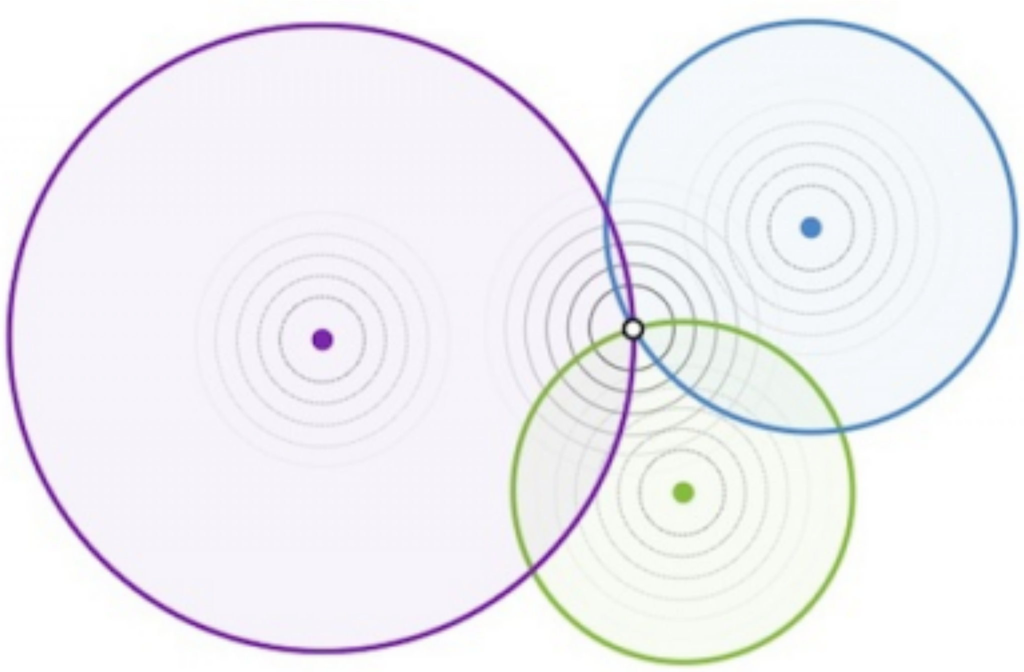
Worldwide Maximum Accuracy of Active IP Geolocation

Massimo Candela
NTT
massimo@ntt.net

Active IP geolocation

- Deriving the geographical location of a connected device by means of latency measurements
  - Devices (landmarks) of known position ping a target device of unknown position
  - Latencies are converted to distances
    The conversion factor (km/ms) is called Speed-Of-Internet (SOI)
  - Multiple distances are used to approximate the target position



Disclaimer

- We do not describe a geolocation system
- It's not an accuracy evaluation of any geolocation method or service
- We will focus only on IP addresses of the Internet Infrastructure
- Results are based (and possibly biased) on RIPE Atlas which currently offers the best characteristics for our intent

We want to…

1. Calculate the maximum theoretical accuracy achievable
   - Worldwide and for each region
2. Study the effects produced by the position of landmarks and by the strategy followed for their enrolment (Infrastructure vs Edge landmarks)
   - Do they produce different SOI?
   - Do they reach the target with a different amount of hops?
3. Quantify the accuracy for various amounts of used landmarks
4. Verify if the accuracy improves by considering only areas with “Internet Infrastructure”
5. Report to the community parameters useful for a possible geolocation system implementation

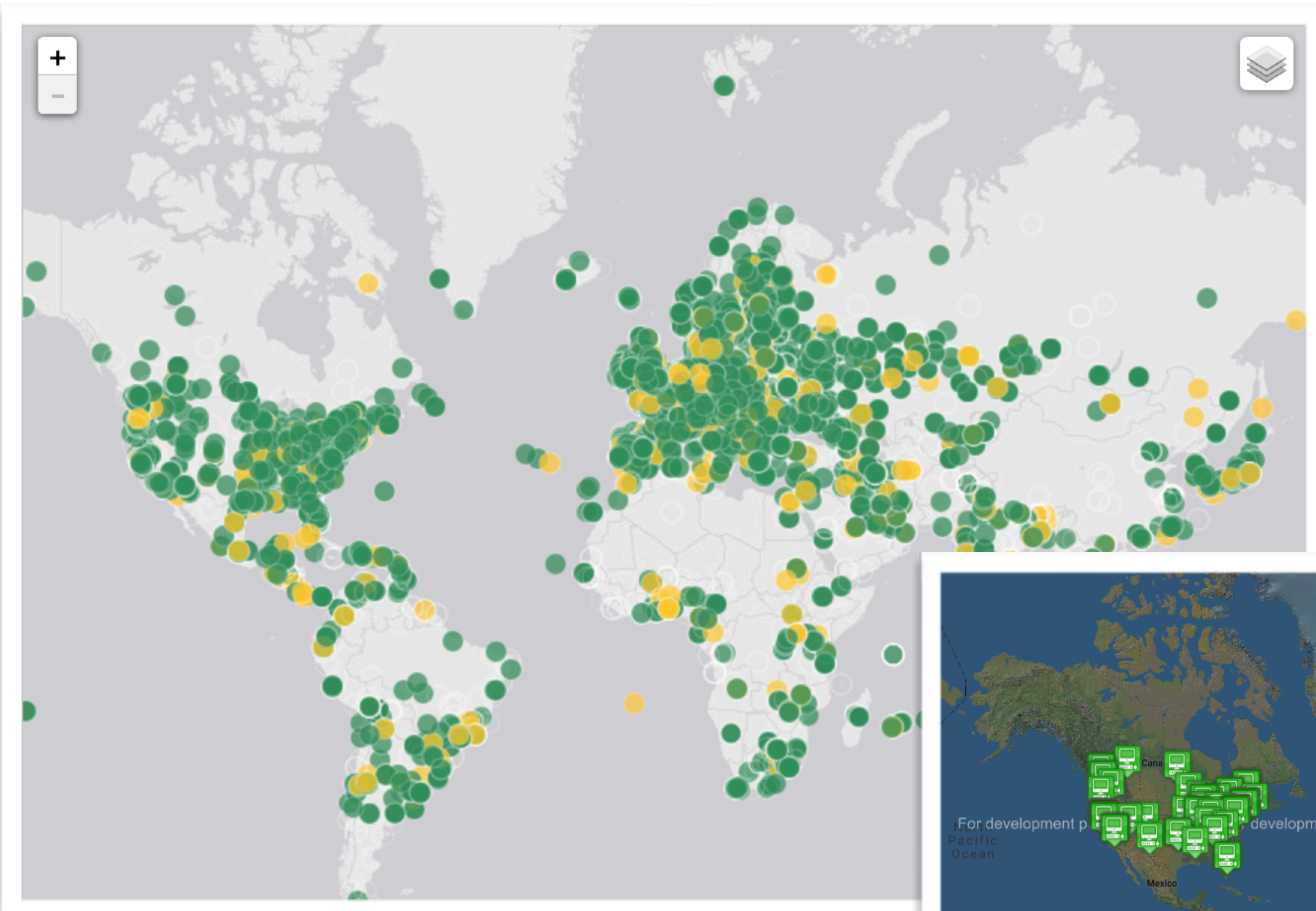
Disclaimer

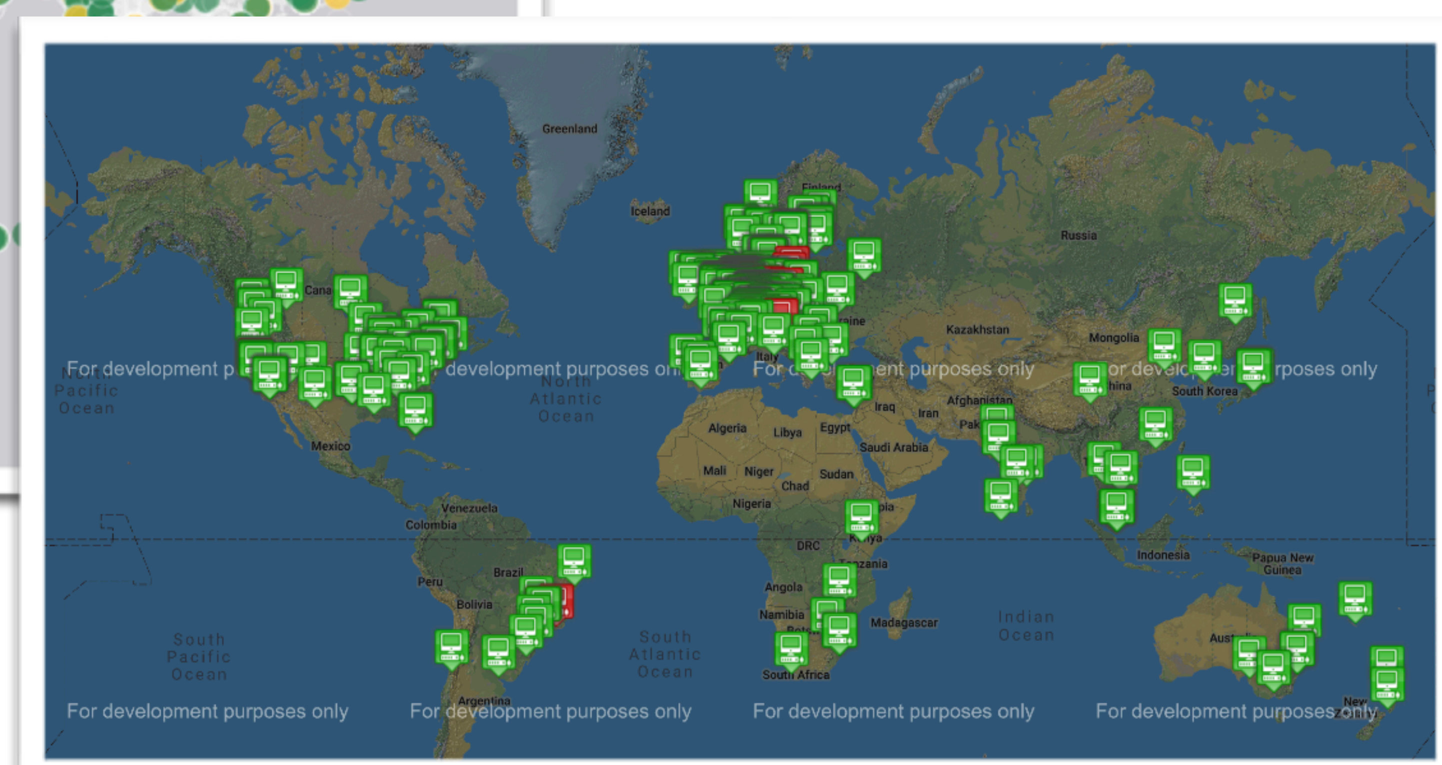

CAUTION
Legacy IP Only
This product does not support the current generation of the Internet Protocol, IPv6.

Datasets

- Targets: Anchors + NLNOG Ring nodes
- Sources: Probes + Anchors
  - *Probes* collect measurements from the edge of the Internet and they are available in a greater amount
  - *Anchors* collect measurements from the infrastructure of the Internet, in more controlled and stable conditions
- We produce **two** ground-truth datasets, **Infrastructure** (Anchors) and **Edge** (Probes), for a total of **23 million** ping measurements
  - 10 pings for each source-target pair, we take the minimum

Coverage





Greenland
Iceland
Russia
Kazakhstan
Mongolia
South Korea
Afghanistan
Iraq
Iran
Saudi Arabia
Algeria
Libya
Egypt
Mali
Niger
Chad
Sudan
Nigeria
DRC
Angola
Namibia
Madagascar
South Africa
Mexico
Venezuela
Colombia
Peru
Brazil
Bolivia
Argentina
Indonesia
Papua New Guinea
North Atlantic Ocean
South Atlantic Ocean
South Pacific Ocean
Indian Ocean
For development purposes only
For development purposes only
For development purposes only
For development purposes only

Model

- To calculate the distance ***d*** between the target and a landmark, we first estimate the Round Trip Time (RTT)
- The observed RTT is then used to compute the One-Way Delay ***o***, calculated as RTT/2
  - This is a (common) simplification/compromise that assumes that the path is traveled in both ways with equal time
- Finally we estimate the distance as ***d = p * o*** where ***p*** is the SOI
  - Both **o** and **p** are affected by errors
- If you know the real distance ***d'***, you can calculate the error ***e*** as ***d = d' + e***

Speed Of Internet of the World

- The SOI of the entire world is:
  - 71.89 km/ms for the Infrastructure dataset
  - 67.11 km/ms for the Edge dataset
    - probes reach the targets more slowly (~7%)
  - Pretty far from 2/3 or 4/9 of the speed of light
- To understand the difference, we analysed the Atlas traceroutes
  - On average Anchors reach the targets with 10.6 hops while probes in 13.2 hops

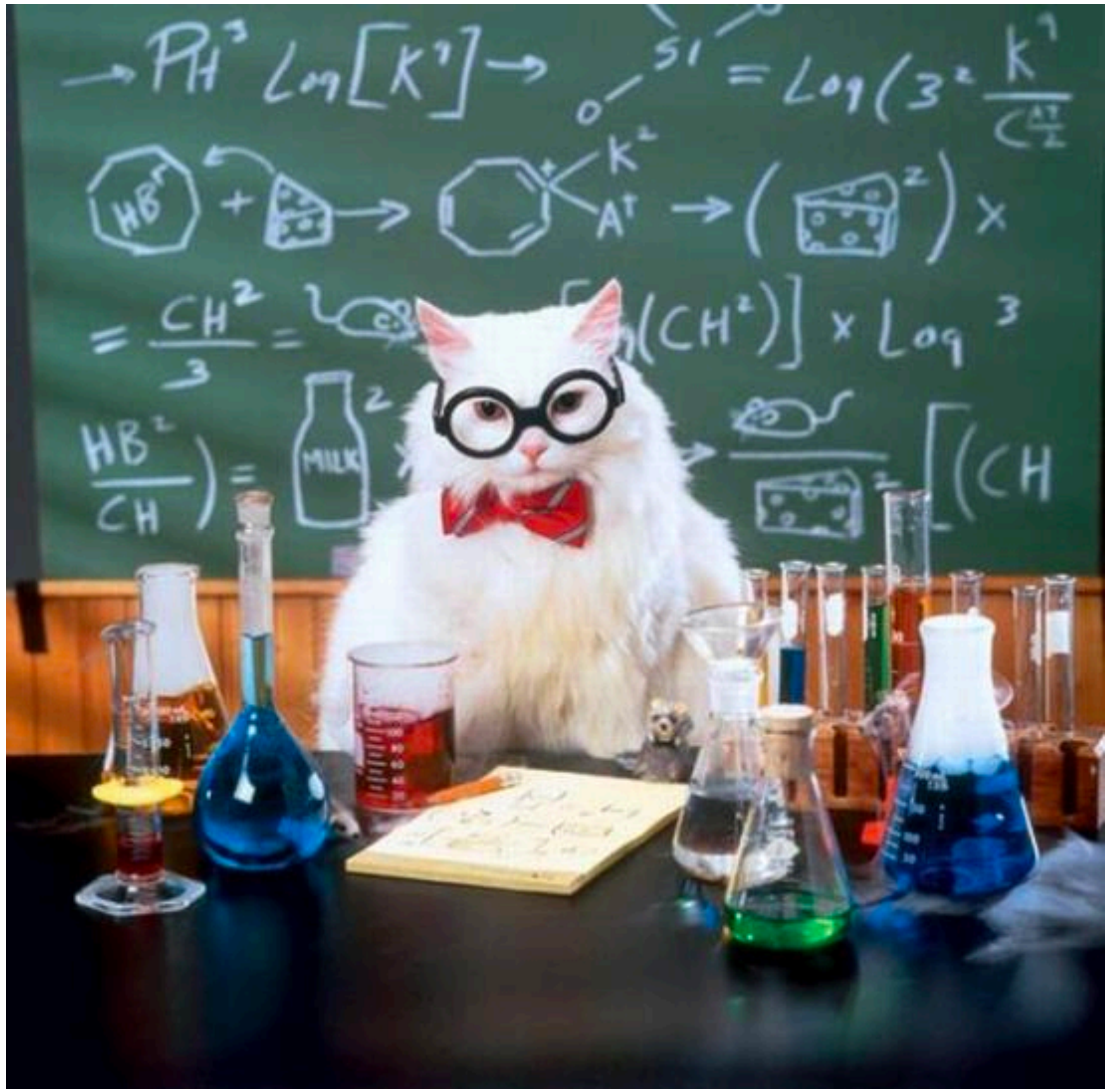
How to calculate the accuracy?

- The method used for calculating the maximum accuracy of a positioning system is usually the Cramér–Rao Lower Bound (CRLB)

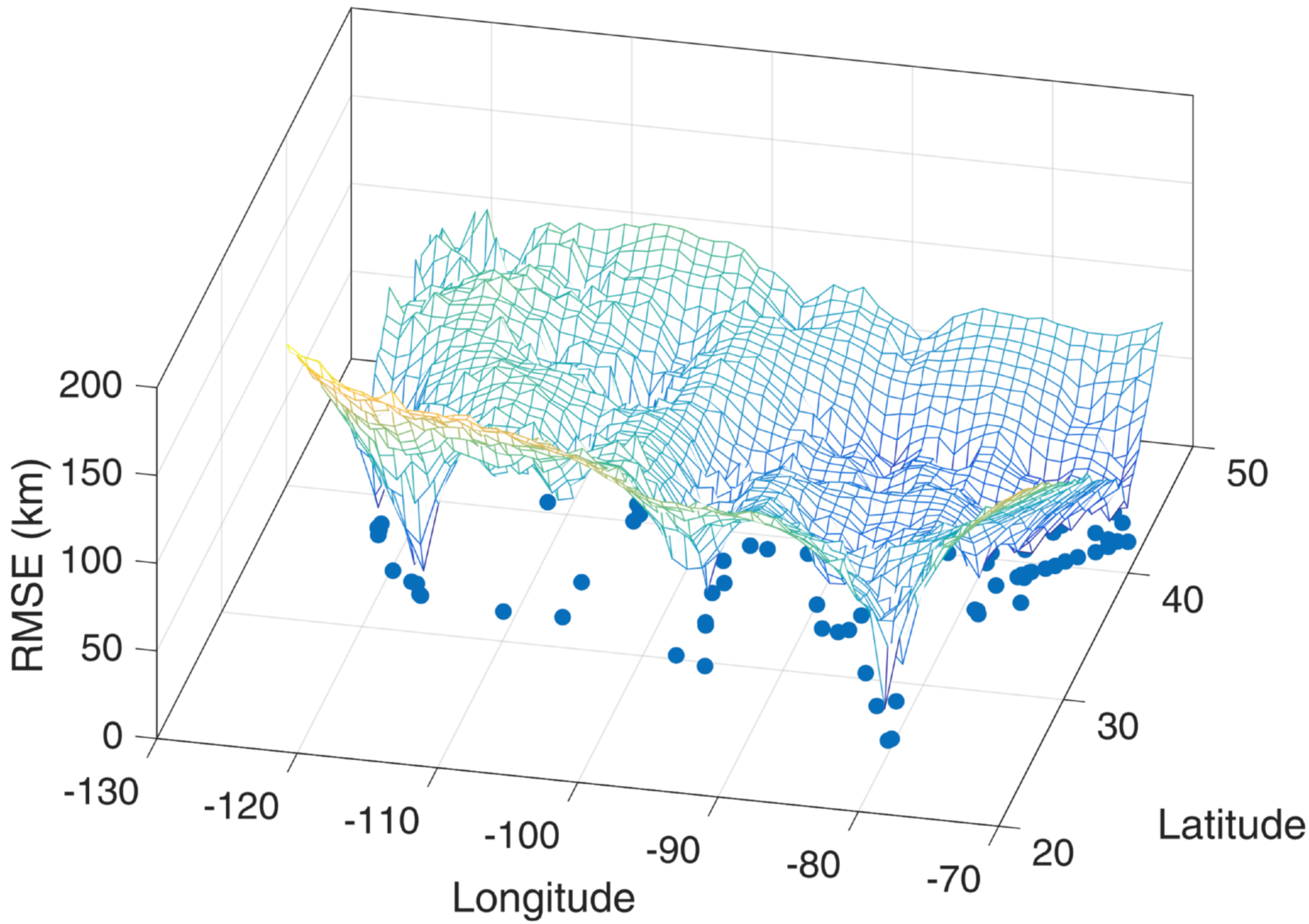
- The CRLB defines a lower bound to the mean square error (MSE) of a position estimate

- Geolocation accuracy can be defined as the square root of the minimum mean square error (RMSE)

???

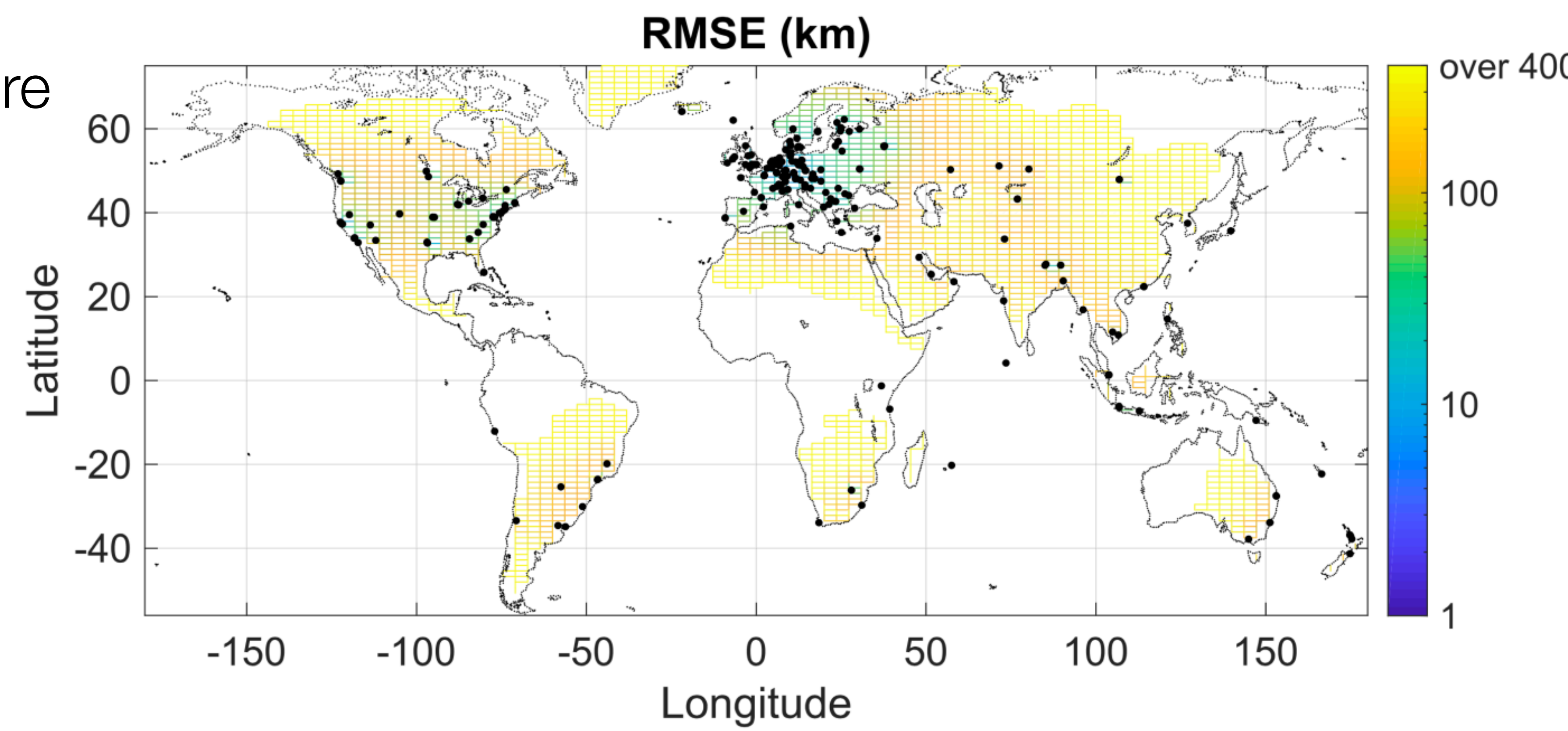

MILK

3D Cramér–Rao Lower Bound (CRLB)



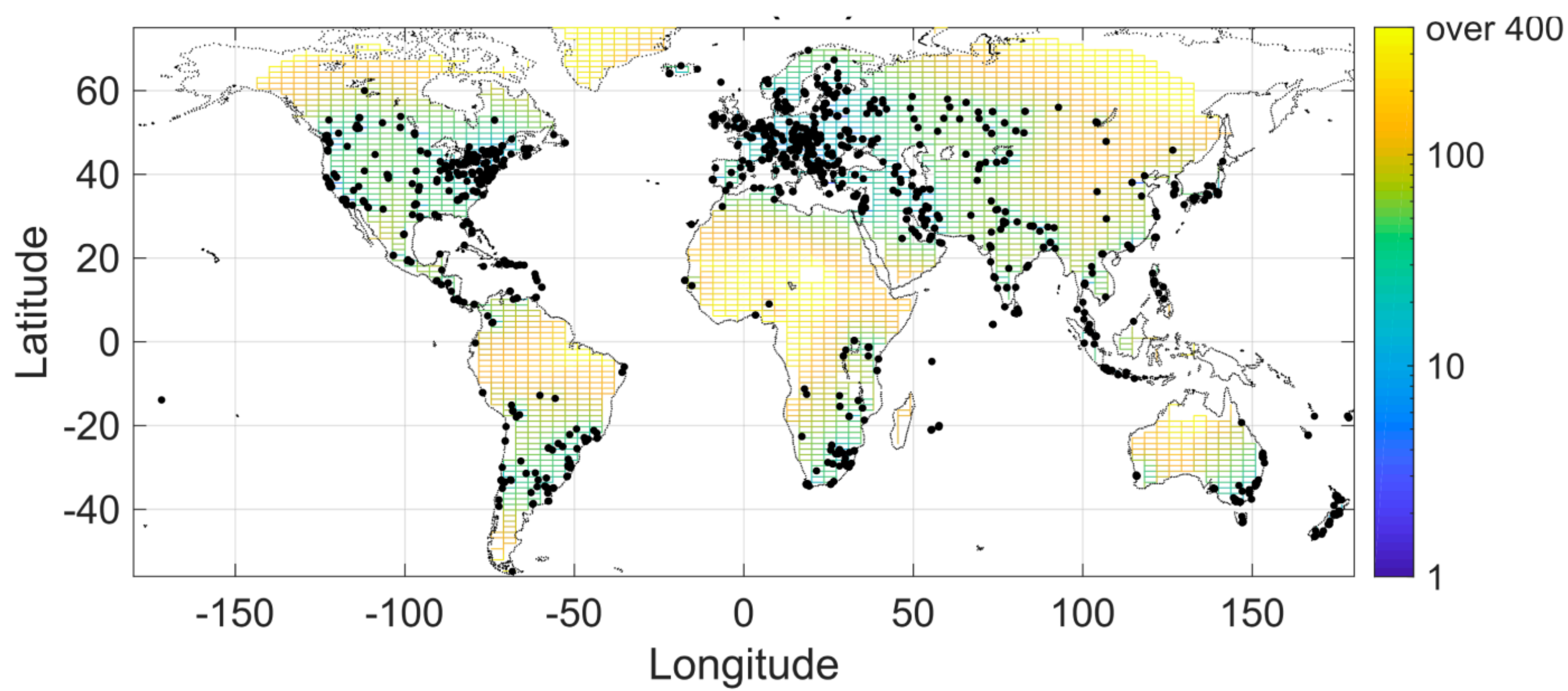
Worldwide accuracy

Infrastructure



One SOI calculated for the entire world

Edge



Worldwide accuracy for a given amount of landmarks

Infrastructure			Edge		
Landmarks	Median RMSE (km)	Coverage	Landmarks	Median RMSE (km)	Coverage
10	615	1.3%	10	651	1.3%
20	501	13.6%	20	512	11.5%
50	390	40.1%	50	399	35.6%
100	314	60.3%	100	320	51.9%
200	256	68.2%	200	249	69.3%
313	254	71.0%	2166	78	91.1%

Regional results

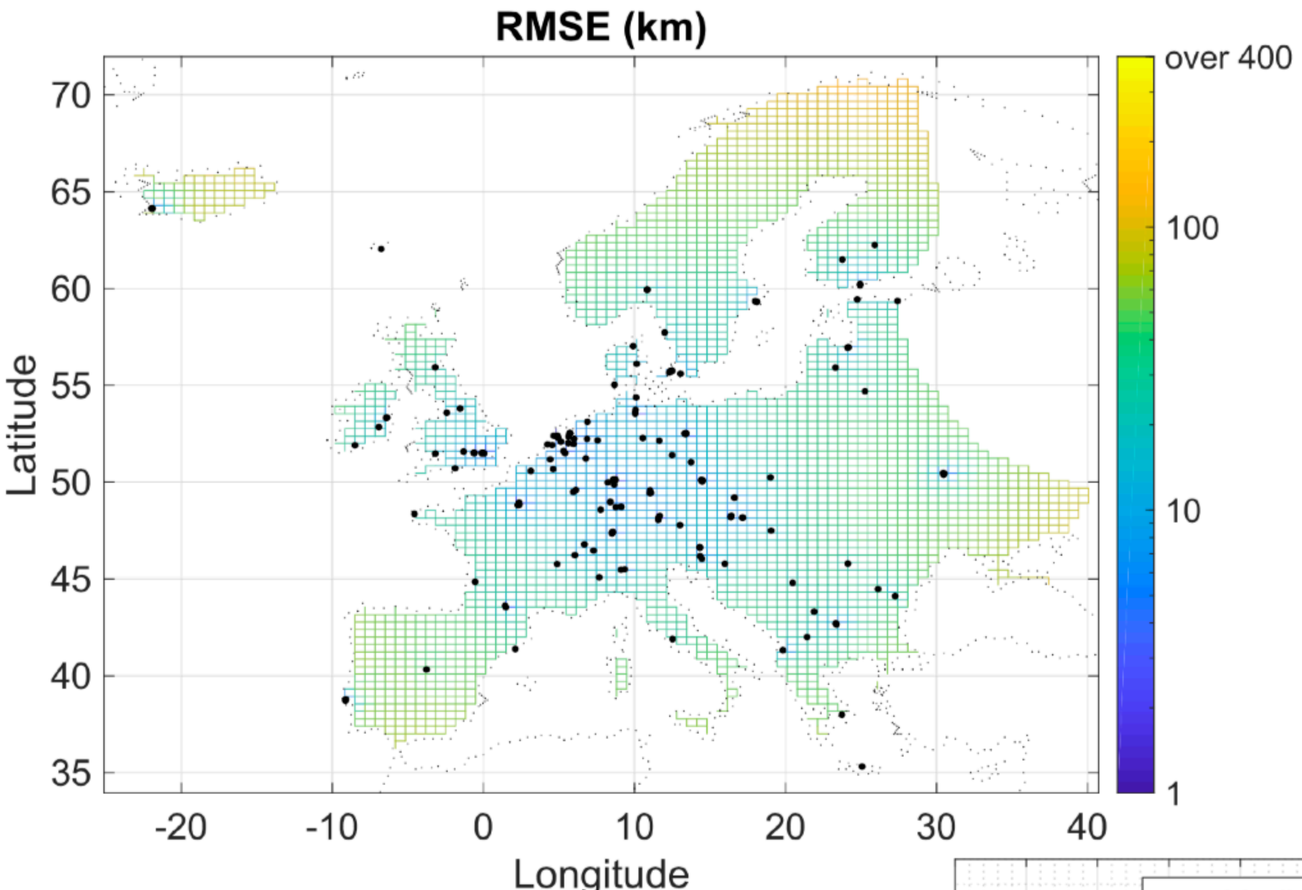
Infrastructure	Asia	Europe	North Am.
Landmarks	37	191	55
p (km/ms)	106.51	68.33	95.63
k	0.39	0.30	0.25
Avg. number of hops	10.5	7.7	8.0
Avg. geographical distance (km)	1561	891	1447
Avg. hop length (km)	149	116	181

Edge	Asia	Europe	North Am.	South Am.	Oceania	Africa	Middle Ea.
Landmarks	484	500	500	210	260	242	238
p (km/ms)	97.13	62.60	73.16	61.40	96.67	80.13	21.12
k	0.36	0.33	0.39	0.42	0.33	0.41	1.03
Avg. number of hops	13.1	10.2	12.4	13.9	9.5	11.2	17.0
Avg. geographical distance (km)	1559	1132	1301	970	1520	1103	193
Avg. hop length (km)	119	111	105	70	160	98	11

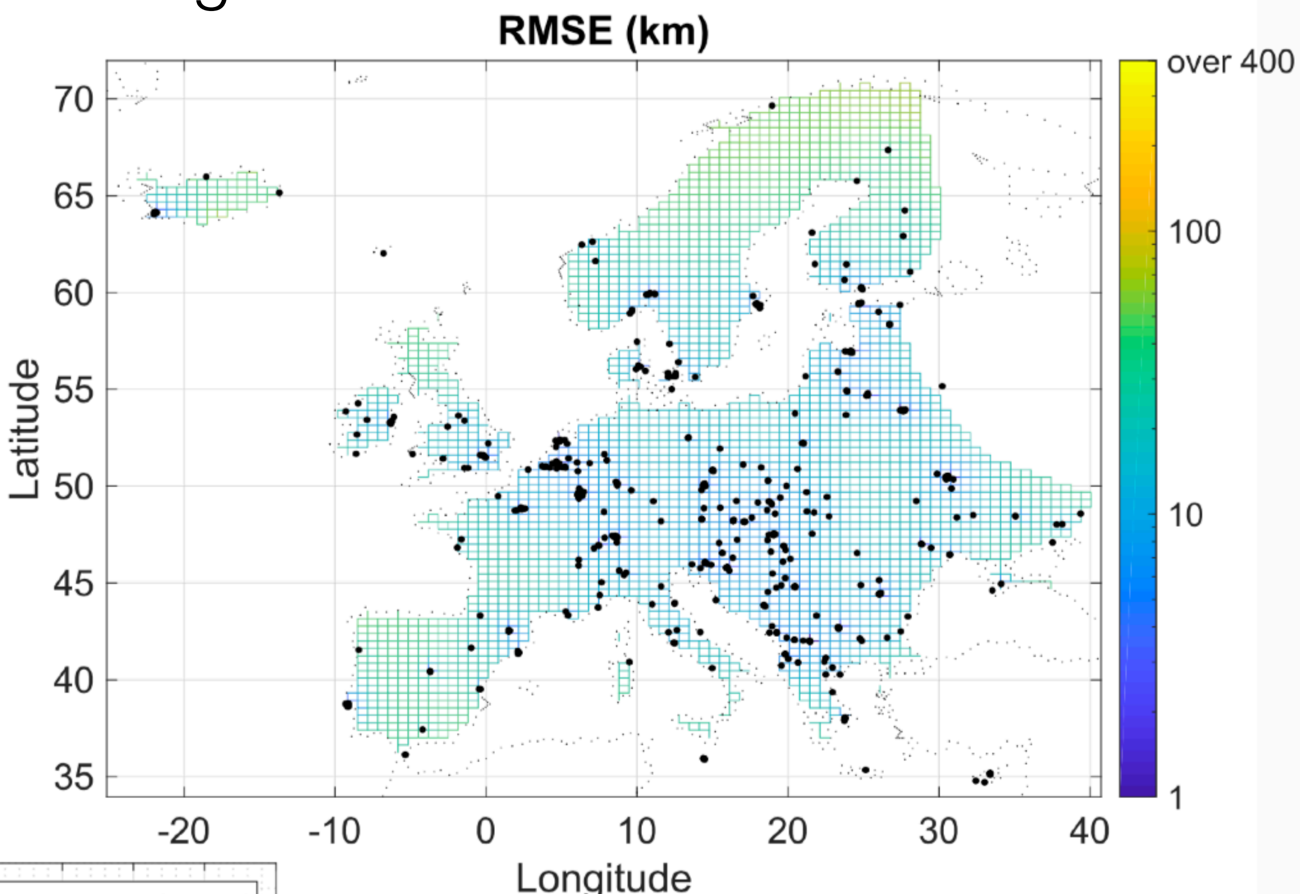
- *P* is the SOI (ms to km conversion factor)
- *K* is the proportionality factor between standard deviation of the error and the distance
    - The distance estimation is more reliable for Anchors

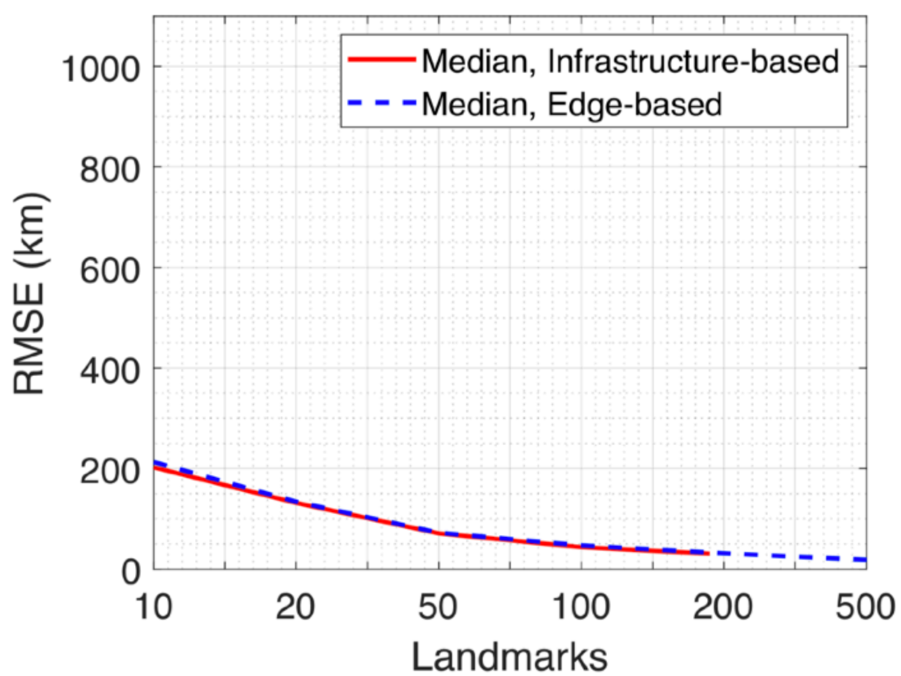
Europe

Infrastructure



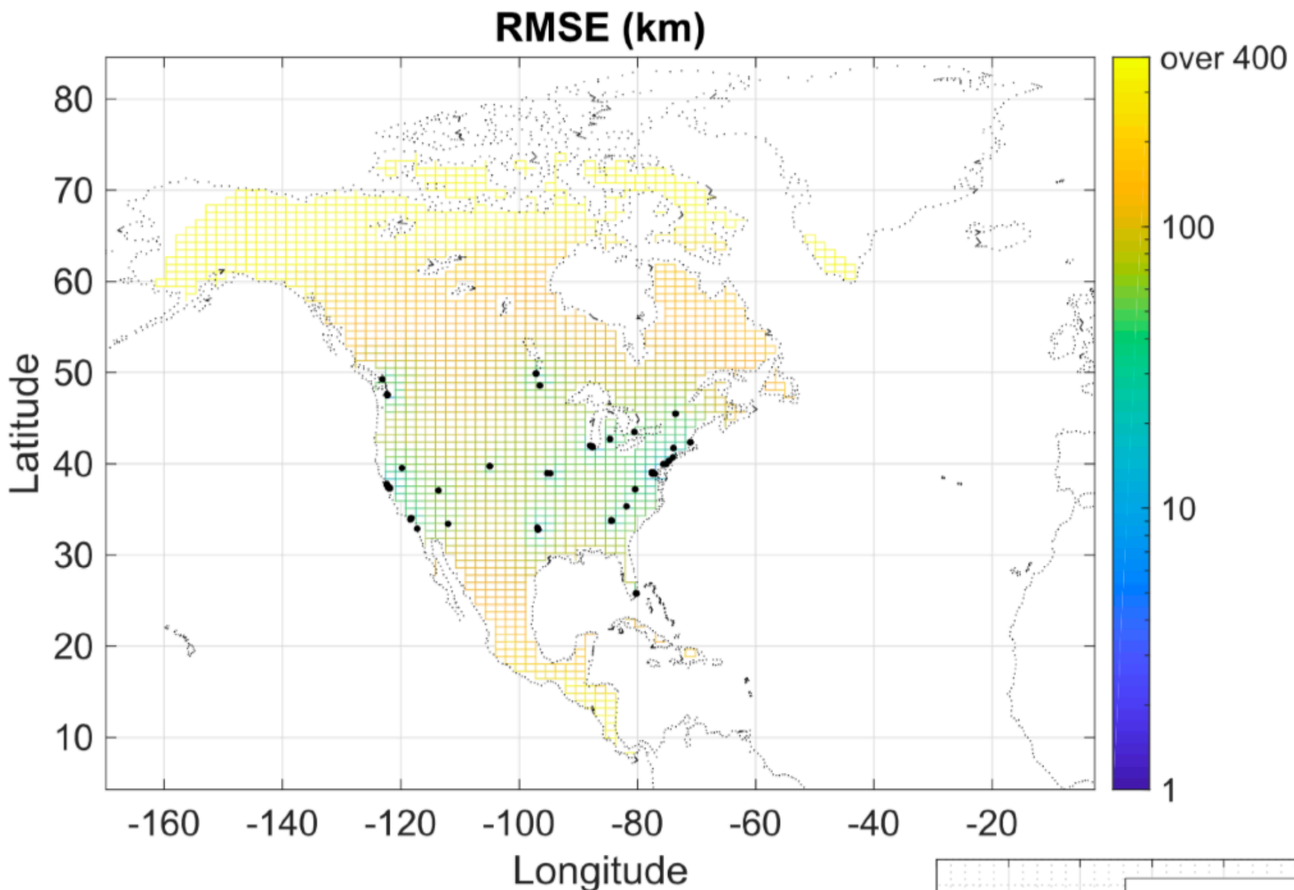
Edge





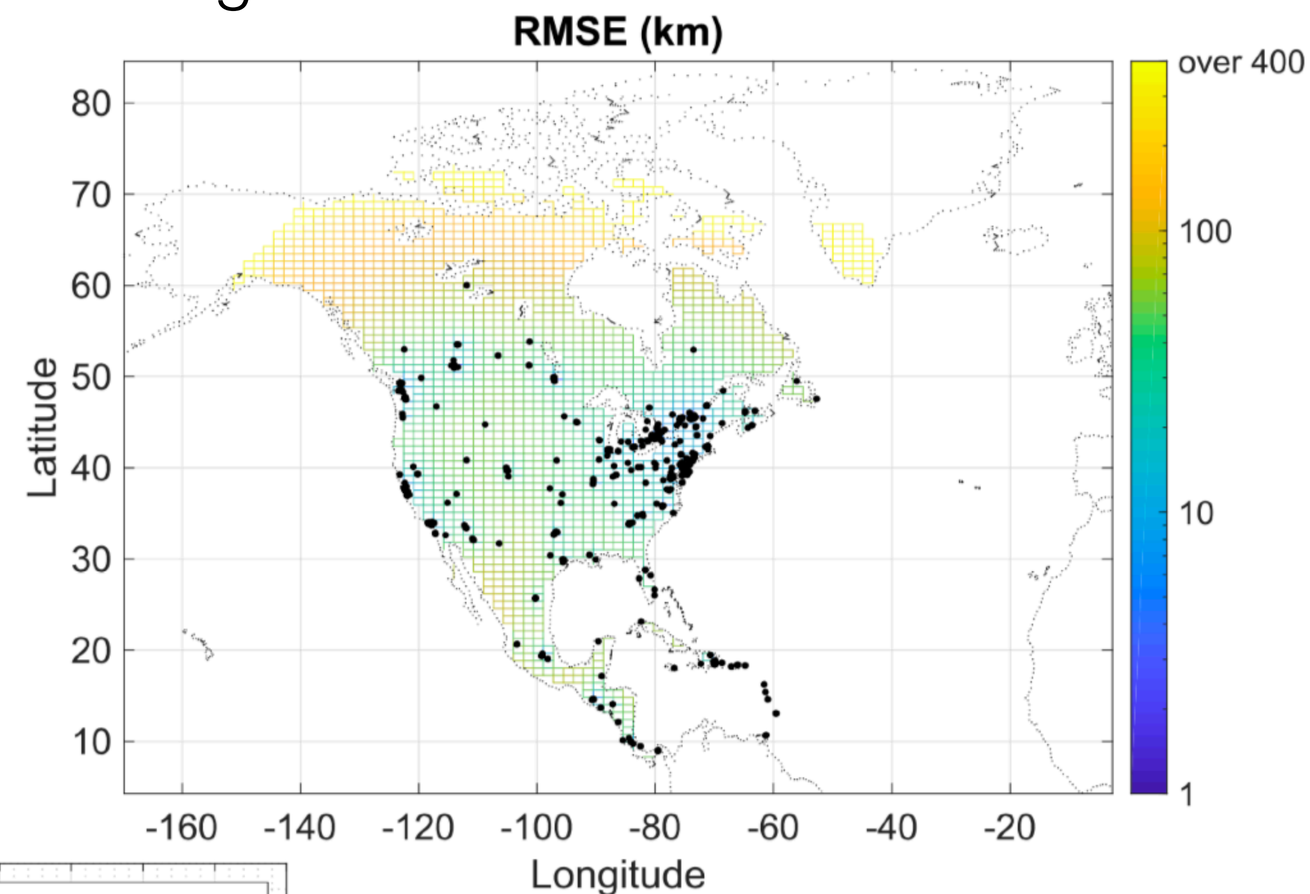
North America

Infrastructure

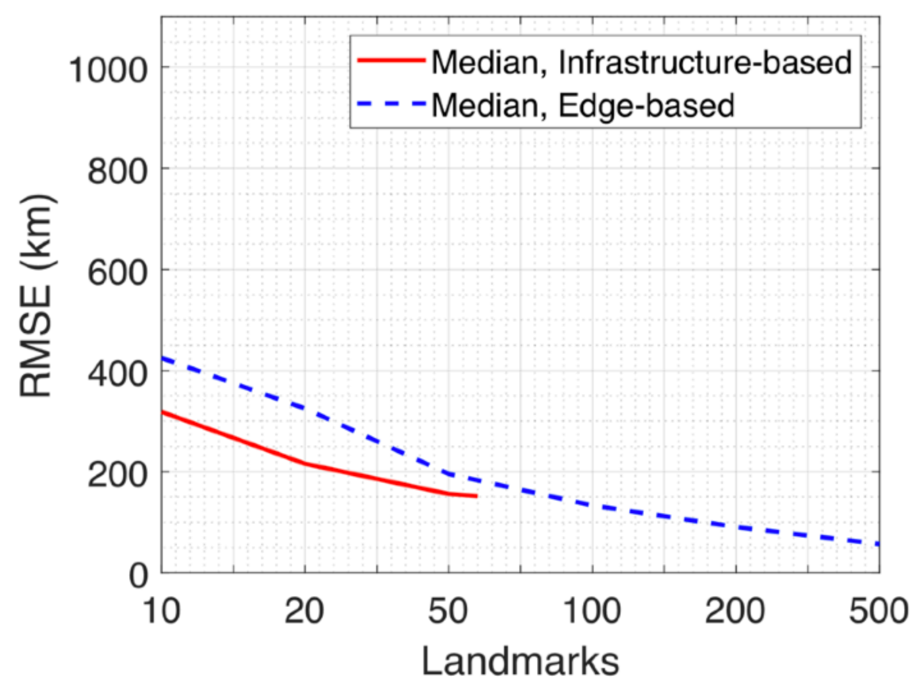


(c)

Edge

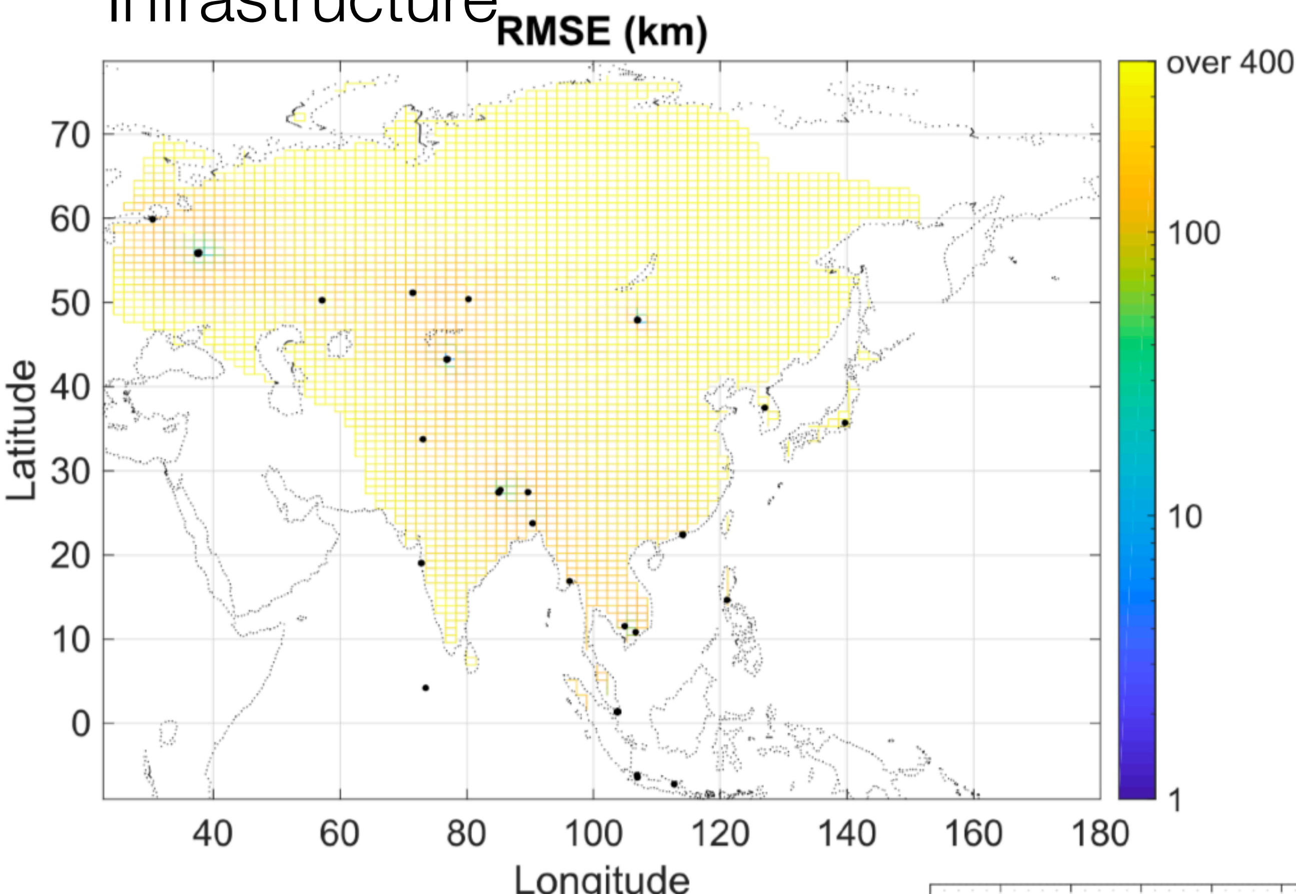


(d)



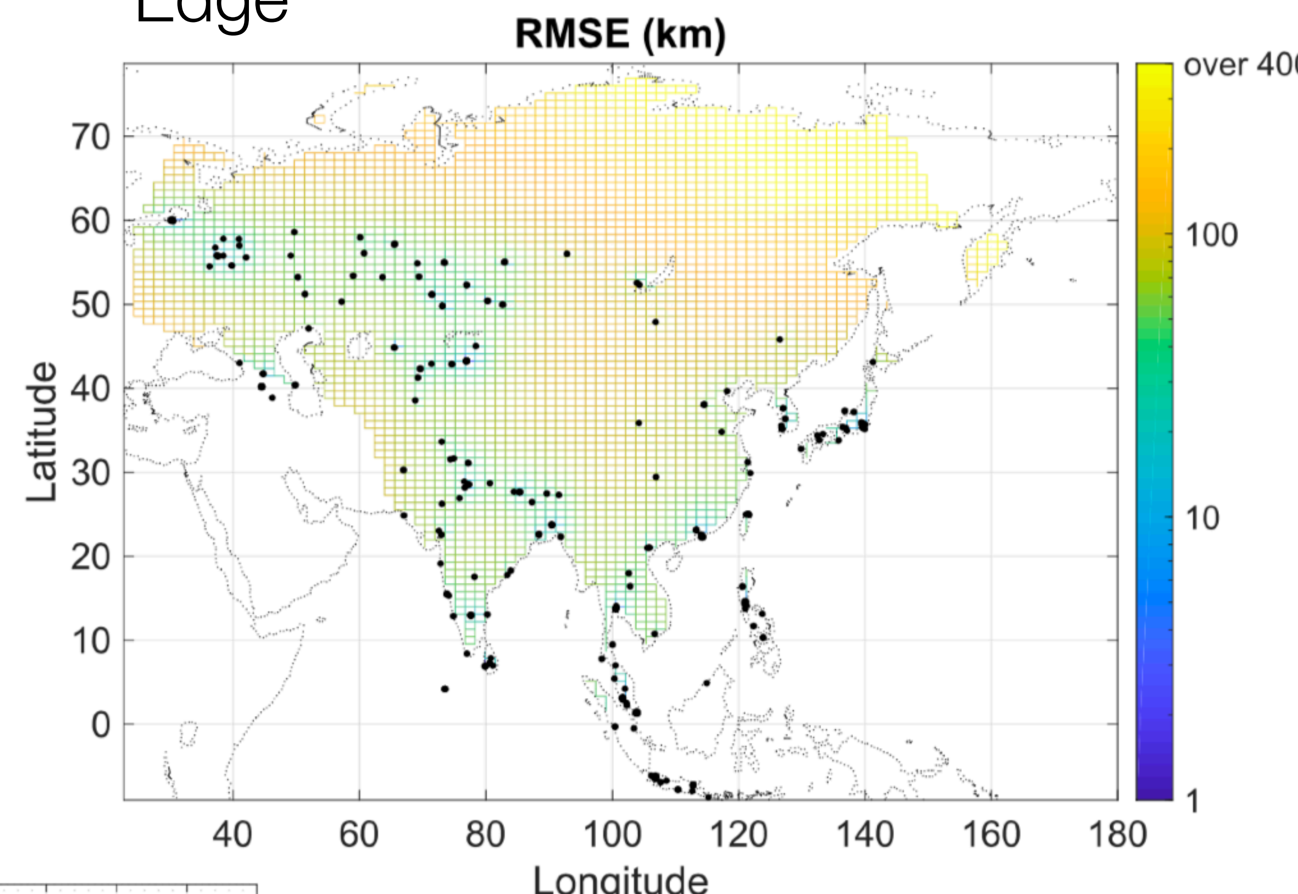
Asia

Infrastructure

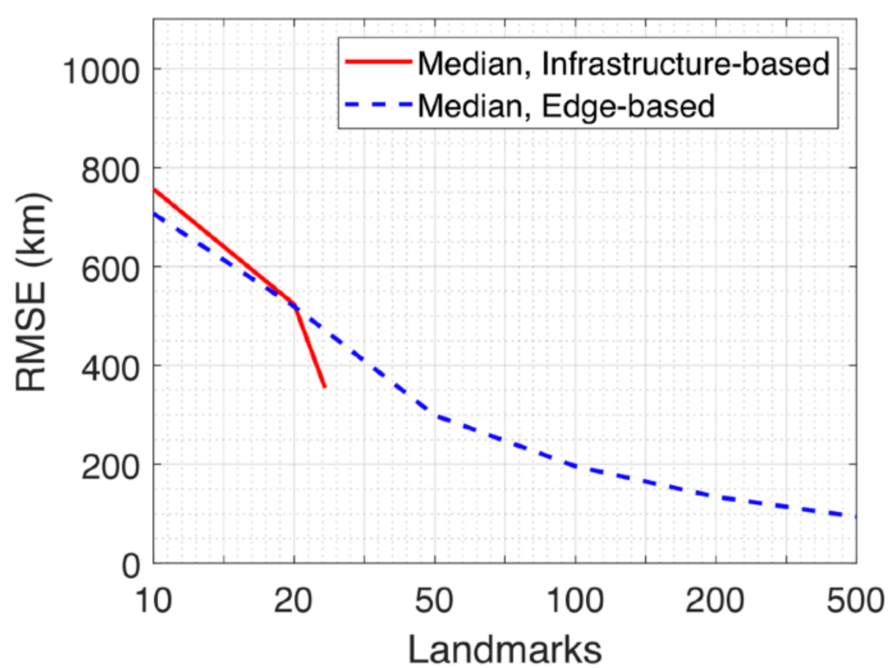


(e)

Edge



(f)



Non-uniform distribution of the targets

- How much the accuracy improves if we consider only areas with “Internet Infrastructure”?

Non-uniform distribution of the targets

- Infrastructure vs. Edge (single SOI)

Infrastructure			Edge		
Landmarks	Median RMSE (km)	Coverage	Landmarks	Median RMSE (km)	Coverage
10	532	9.7%	10	612	5.3%
20	324	68.3%	20	491	52.7%
50	150	91.1%	50	249	86.7%
100	61	97.5%	100	141	94.7%
200	18	97.9%	200	83	98.5%
313	6	97.9%	2166	3	100.00%

- Uniform vs. Non-uniform (a SOI per region)

	Uniform		Non uniform	
Region	Median RMSE (km)	Coverage	Median RMSE (km)	Coverage
Africa	266	88.0%	2	94.3%
Asia	196	89.5%	16	100.0%
Europe	47	100.0%	28	100.0%
Middle-East	118	52.7%	3	99.6%
North Am.	133	65.7%	30	100.0%
Oceania	141	88.5%	1	100.0%
South Am.	130	80.8%	2	100.0%

Conclusions

- IP infrastructure geolocation with active measurements is feasible in terms of precision
- Using a single value of SOI for the entire world is a terrible idea
- The SOI is linearly correlated to the average hop length (source-target distance/hops)
  - Which varies based on the landmark type and region. We provided these values
- Infrastructural landmarks are slightly more precise than edge landmarks
  - When the number of landmarks involved in geolocation increases, the advantage of using infrastructural nodes is less evident
  - The number of landmarks is a more decisive factors BUT this is an extremely important parameter for the scalability of a real geolocation system

Questions?

massimo@ntt.net

@webrobotics

Thanks to my co-authors:

- Enrico Gregori, CNR, Pisa, Italy
- Valerio Luconi, University of Pisa, Italy
- Alessio Vecchio, University of Pisa, Ital