AGGREGATED EXPERTISE OF USER CONTRIBUTIONS TO ASSESS THE CREDIBILITY OF OPENSTREETMAP

Bani Idham Muttaqien

ESRI USER CONFERENCE

MOTIVATION

- **Increasing trend to use VGI**
- **Quality and fitness-for-purpose of VGI : ongoing area of research**
- **Verify using external (authoritative) data sources?**
  - Absence of global authoritative data set
  - Should have the same geographical coverage and a similar data model
  - Prohibitive procurement costs and licensing restrictions
  - OSM has a better coverage than authoritative datasets in some areas

RESEACH OBJECTIVES

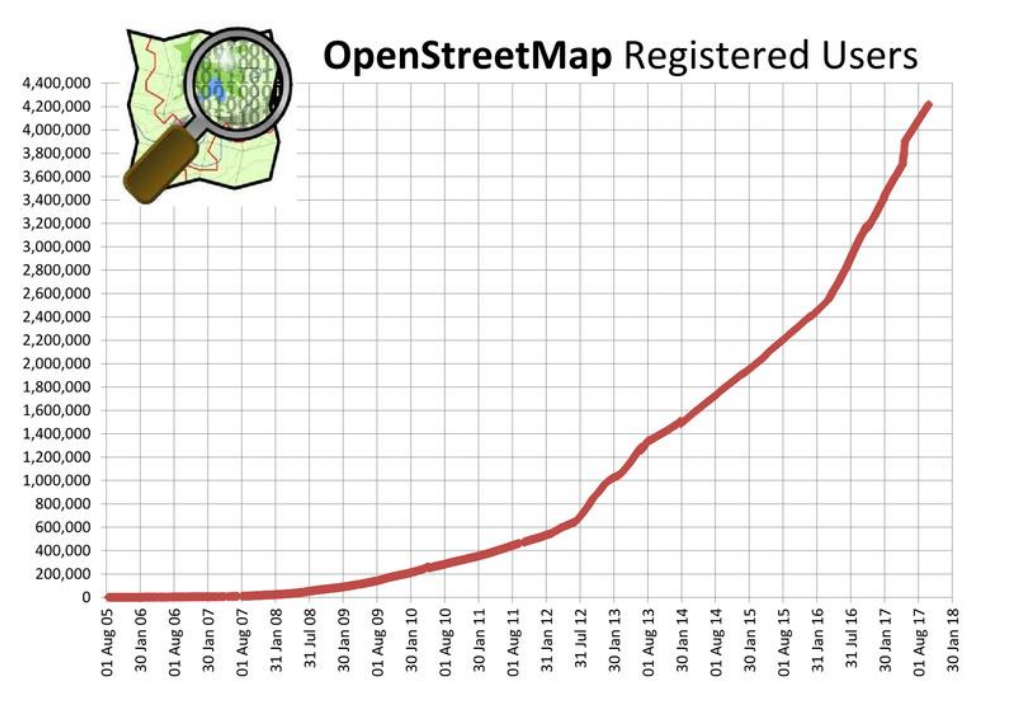
“To develop and implement a model based on aggregated expertise to assess the credibility of OSM features”

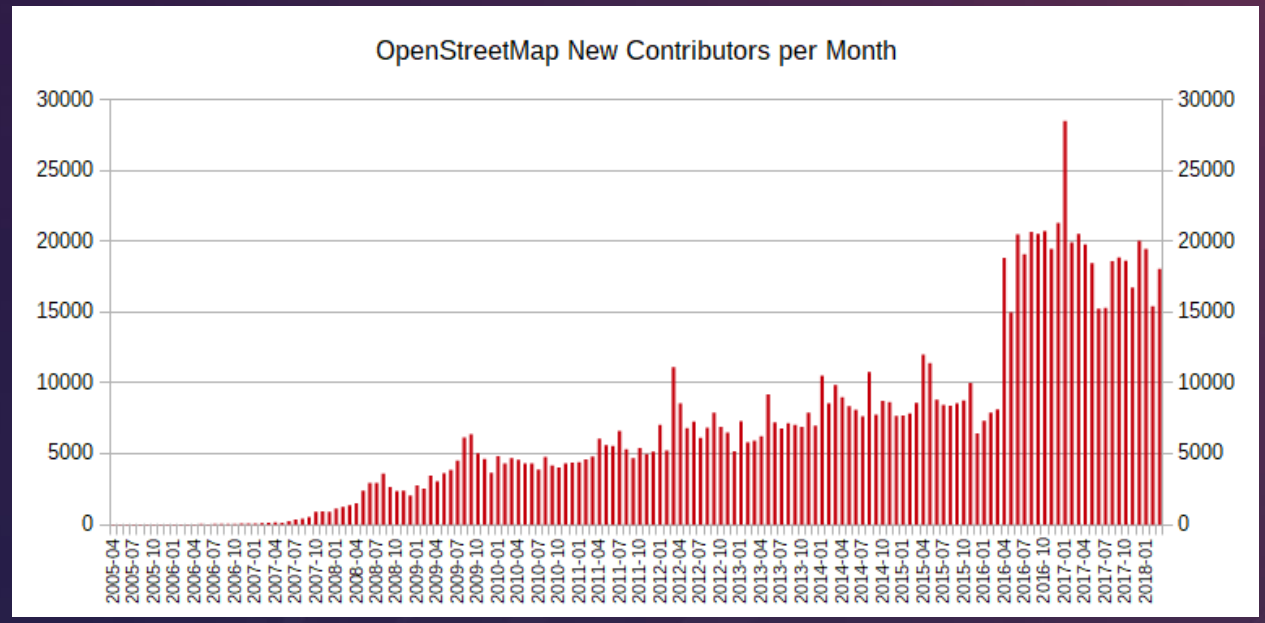
RESEARCH QUESTIONS

- How can we operationalize the concept of aggregated expertise for the purpose of assessing credibility of collaborative contributions?
- How well can aggregated expertise predict the quality of OSM features in our case study area?

OPENSTREETMAP

"…a free, editable map of the whole world that is being built by volunteers largely from scratch and released with an open-content license"





https://wiki.openstreetmap.org/wiki/About_OpenStreetMap
https://wiki.openstreetmap.org/wiki/Stats

MEASURES AND METHODS

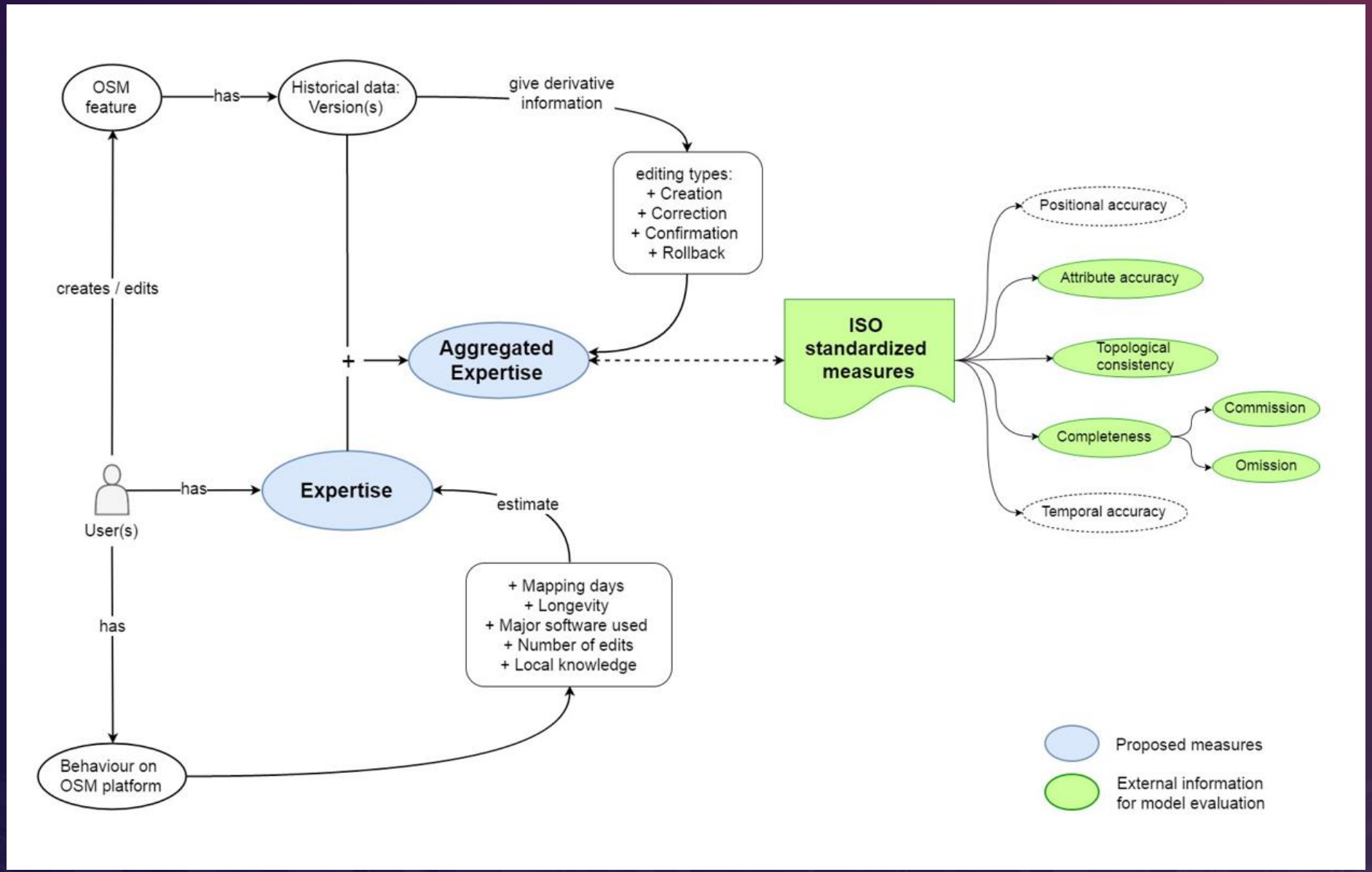


Figure 1. Aggregated expertise to assess the credibility of VGI

AGGREGATED EXPERTISE

Expertise score $E(u)$ of the VGI contributors;

$$E(u) = \sum_{i=1}^{N} w(i).p(i)$$

where $p(i)$ is expertise parameters with $w(i)$ being weight for expertise parameters

Aggregated expertise $AE\ (fi)$ is defined as follows:

$$AE\ (fi) = \sum_{i=1}^{v} et(u)\, E(u)$$

where $et(u)$ is the weight for what type of editing was contributed by user u, and v is number of versions for feature fi.

CHECKING THE GROUND TRUTH

- **The field study is a 1 km^2 area of a touristic neighborhood in Jakarta**
- **Overall, it took 3 days to collect information on 398 OSM features in the field.**
- **Three ISO quality measures are suitable and feasible to measure in the field;**
  - Attribute Accuracy
  - Completeness
  - Topological consistency

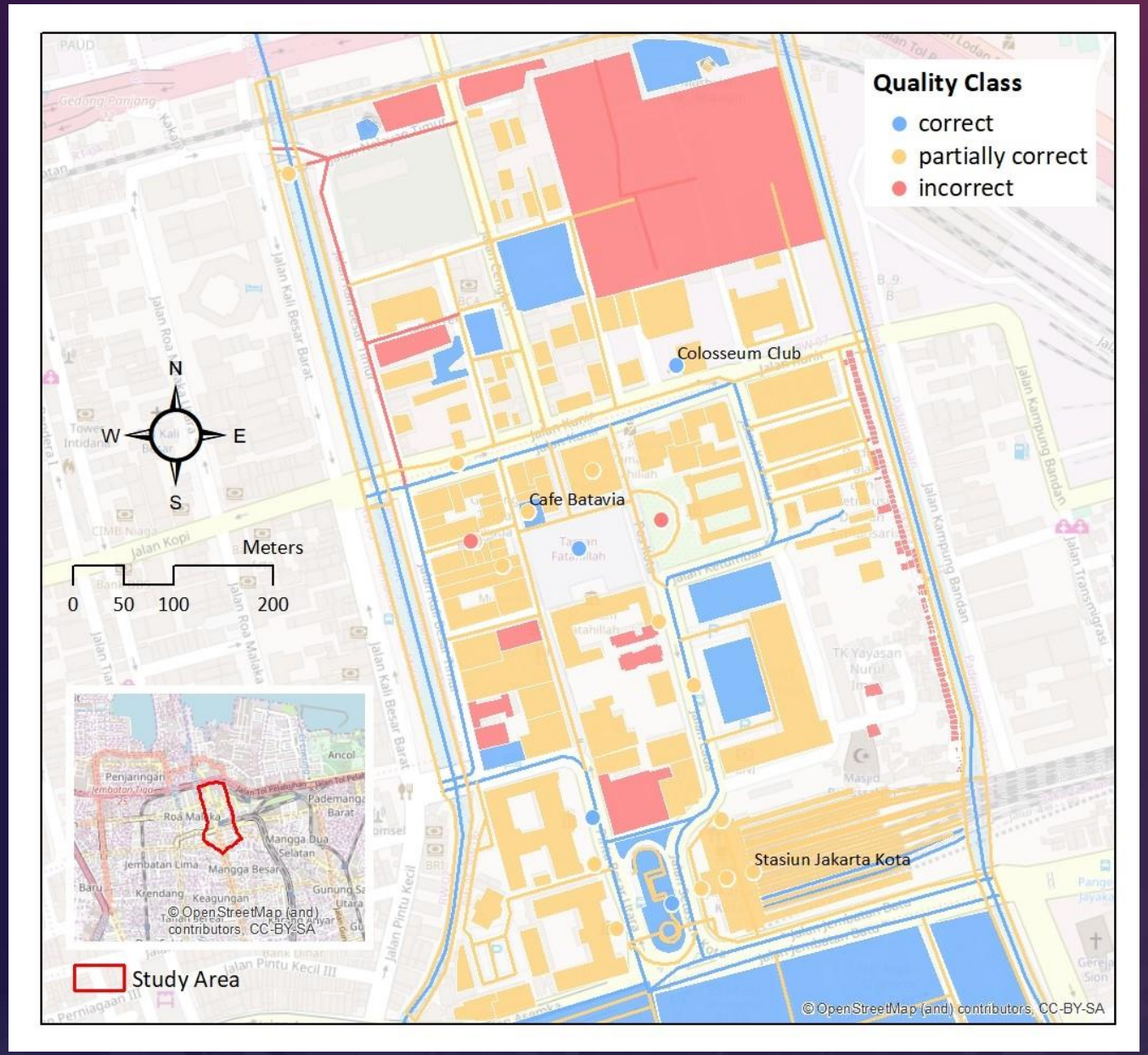


Figure 2. Map of the study area showing feature quality class of OSM features

STATISTICAL ANALYSIS

- Examine the association between the variables that contribute to the aggregated expertise and the feature quality score (ground truth) using **Kendall’s Tau (τ) correlation**

- Evaluate feature credibility by using a **multinomial regression model** to predict the quality class from the Aggregated Expertise .

RESULTS

DATA ANALYSIS

- The collaborative edits on OSM platform in the study area resulted in **737 unique versions** for 398 OSM features.
- For features with more than one version, the number of users participating in the editing process often differs from the number of versions created.
- Feature type were used for further correlation analysis to expose the specific behavior about the feature type and the persistence conditions over time.

CORRELATION ANALYSIS RESULT

Correlation test results for expertise parameters against feature quality scores

Expertise Parameters	Kendall's Tau correlation coefficient (τ)
Number of edits	0.304
Number of mapping days	0.299
Longevity of activity	0.332
Local knowledge	0.145
Software skill	0.533

CORRELATION ANALYSIS RESULT

Correlation test results for recency of last version time effect parameter of trustworthiness against feature quality scores

Map Feature Type	Kendall's Tau correlation coefficient (τ)
Highway, Railway, Waterway	-0.136
General building and Landuse	0.029
Commercial and informal settlement	-0.383
Long living object	-0.169

MULTINOMIAL LOGISTIC REGRESSION

The final probability model:

$$\ln\left(\frac{P\,(correct)}{P\,(incorrect)}\right) = -3.407 + 0.850(AE)$$

$$\ln\left(\frac{P\,(partially\ correct)}{P\,(incorrect)}\right) = -1.310 + 0.754(AE)$$

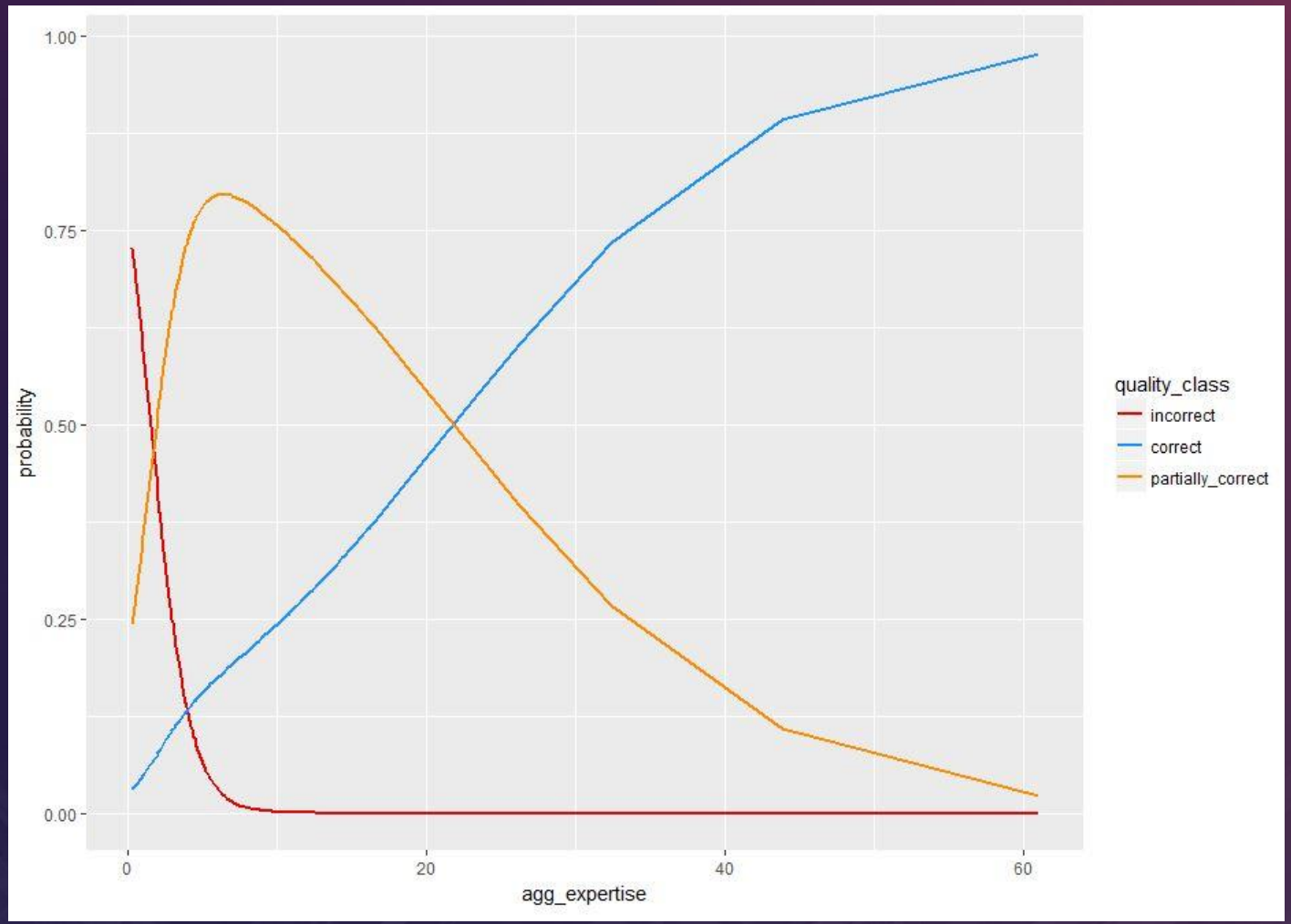


Figure 3. Plot of predicted probabilities across aggregated expertise scores for each quality class

CONCLUSION AND FUTURE WORK

CONCLUSION

- The concept of aggregated experience based on feature edit history and contributor characteristics can be operationalized to assess the credibility of VGI features
- The aggregated expertise allows to predict the credibility of OSM features, and by extension the likelihood of a feature to be of sufficient high or low quality.
- By helping to discover areas of low-quality scores, it can help to prioritize mapping efforts.
- The exact implementation depends on the available meta-data from the VGI platform. Any VGI platform that supports feature versioning and records user profile information can adapt the approach.

FUTURE WORK

- The use of Linked Data in relationship with a domain ontology to enrich the semantics of the contributors' profiles and the features they edit.
- Larger case studies in different geographic areas to demonstrate geographic transferability of our approach.
- Download and use the OSM Planet History files to reduce load of OSM servers for larger case studies. Cloud computing infrastructure offers a flexible approach to host larger datasets to scale up the automated processing.
- Explore additional combinations of variables and their weights for an aggregated expertise scores and test the validity of the approach for specific feature types.

THANK YOU