Finding new National Cycle Routes in Kansai

Mathias BALLOT, Yohan LARBRE, TAKUTO Miyata
KUROKI Shota

August 2024

Contents

Chapter 1

Introduction

The National Cycling Route [4] is a system for promoting cycle tourism in which the national government designates routes that meet certain criteria. In Japan, six cycle routes have been promoted to the status of National Cycle Route. These include the flat bicycle route along the lakeshore of Biwaichi, the natural island scenery of Shimanami Kaido, and the Pacific Cycling Road, which runs from Kansai to Kanto along the coast of Japan.

Several requirements have been established for the designation of a National Cycle Route. According to their website, the criteria are based on the beauty of the route, the riding environment, the hosting environment, and the information available along the road. This leads us to the following steps:

1. Extract common features from diverse data sources along the National Cycle Routes, to understand in detail what contributes to a route's quality.

2. Based on the criteria we have selected (which come from the previous step), identify other potential National Cycling Routes.

The **Finding Useful Data** chapter details the process of sourcing relevant data, focusing on Open Street Map and Elevation Data. It explains the methods used to extract and process the data necessary for identifying and characterizing cycle routes.

In the **Characterizing What a National Cycle Route Is** chapter, we analyze the data to identify key features and tags that define a national cycle route, including the incorporation of elevation information.

Finally, the **Finding New National Cycle Routes** chapter describes the methodology for filtering and scoring potential new routes based on the identified characteristics, culminating in the selection of candidate routes.

In **Conclusion**, we nuance our findings and suggest possible implications or future work based on the analysis.

Chapter 2

Finding useful data

2.1 Finding useful data

Obtaining usable data on Japanese roads presents a significant challenge due to the limited availability of publicly accessible information. Unlike some countries, Japan does not widely release detailed data on its roads, despite providing global statistics derived from this data. However, these summaries are insufficient for our specific needs.

Thus, we have turned to other sources and identified several websites with comprehensive datasets. After excluding some options (such as `https://bikesharemap.com`), we have focused our efforts on two primary sources that appear to offer the most promising data:

- Open Street Map: [2] [3]
- Elevation data provided by the Geospatial Information Authority of Japan: [1]

2.1.1 Open Street Map

OpenStreetMap (OSM) is a collaborative project to create a free, editable map of the world. It allows individuals and organizations to contribute geographic data, such as roads, landmarks, and points of interest, which are then compiled into a global database.

For this project we used the Japan map available at [2].

OSM stores its data in `.osm` files using an XML-based format. This format categorizes geographical features into nodes, ways, and relations, each with unique attributes.

Nodes:

Nodes represent individual points on the map defined by latitude and longitude coordinates. For example:

```xml
<node id="12345" lat="51.5074" lon="-0.1278">
  <tag k="name" v="Big Ben" />
  <tag k="amenity" v="clock" />
</node>
```

Listing 2.1: Example of a node

Ways:

Ways are ordered lists of nodes that form polylines or polygons. They are used to represent roads, rivers, and other features:

```xml
<way id="67890">
  <nd ref="12345" />
  <nd ref="67890" />
  <nd ref="98765" />
  <tag k="bridge" v="yes"/>
  <tag k="highway" v="motorway"/>
</way>
```

Listing 2.2: Example of a way

Ways can also have tags describing the whole way: names, road types, etc.

Relations:

Relations describe how elements work together, such as a route composed of multiple ways:

```xml
<relation id="24680">
  <member type="way" ref="67890" role="outer" />
  <member type="way" ref="98765" role="inner" />
</relation>
```

Listing 2.3: Example of a relation

However, the size of OSM data files is too large for direct usage, making it impractical to handle the entire dataset at once. To overcome this challenge, we have to employ Osmosis, a Java-based command-line tool for managing OSM data, that allows us to extract specific subsets of data from large files, based on criteria such as geographic regions, types of features (nodes, ways, relations) or tags.

For this project, we needed to extract OSM data twice:

First Extraction:

We extracted data specifically for national cycle routes using the "network" tag with values "ncn", "rcn", and "icn", which represent national, regional, and international cycle routes respectively.

```
osmosis \
  --read-pbf file="./Data/kansai-latest.osm.pbf" \
  --tf accept-relations network=ncn,rcn,icn \
  --used-way \
  --used-node \
  --write-xml file="./Data/ncn_kansai.xml"
```

Listing 2.4: First Osmosis Extraction

After extracting the OSM data using Osmosis, we utilized JOSM (Java OpenStreetMap Editor) for visualizing the extracted data. It provides a graphical interface for viewing geographic features, editing attributes, and adding new elements to the map:

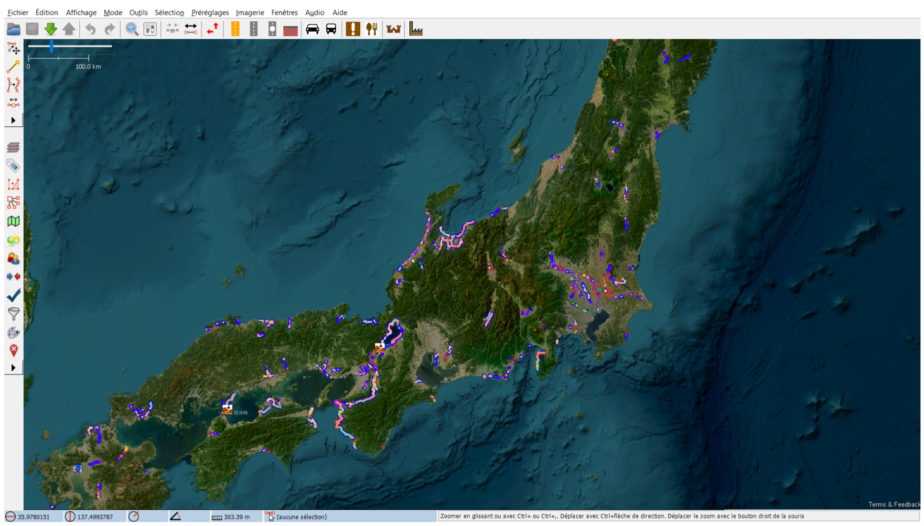


Figure 2.1: National cycle routes visualized with JOSM

This way we could check that the output was correct.

Second Extraction:

The second extraction focused on identifying new national cycle routes across Japan using additional tags identified through our analysis. This part will be explained in due time.

2.1.2 OSM taginfo

The website [6] provided us with comprehensive global statistics for Japan, which significantly saved time during our project. Specifically, we needed tag information for national cycle routes across Japan to compare statistics efficiently. Extracting all the necessary tags using Osmosis commands would have been too time-consuming comparatively.

Page 1 of 708 JSON Displaying 1 to 21 of 14857 items

Tag	Objects		Nodes		Ways		Relations	
building=yes	19 973 194	6.72%	2 166	0.06%	19 951 363	54.86%	19 665	12.31%
source=YahooJapan/ALPSMAP	4 530 481	1.52%	18 446	0.55%	4 512 010	12.41%	25	0.02%
highway=residential	2 898 954	0.98%	2	0.00%	2 898 951	7.97%	1	0.00%
highway=unclassified	2 396 202	0.81%	8	0.00%	2 396 187	6.59%	7	0.00%
source=Bing	1 942 963	0.65%	186 201	5.57%	1 755 832	4.83%	930	0.58%
yh:TOTYUMONO=供用中	1 436 204	0.48%	2	0.00%	1 436 202	3.95%	0	0.00%
building=house	1 398 717	0.47%	95	0.00%	1 394 124	3.83%	4 498	2.81%
yh:STRUCTURE=地上	1 387 865	0.47%	2	0.00%	1 387 863	3.82%	0	0.00%
yh:TYPE=その他一般道	1 367 814	0.46%	2	0.00%	1 367 812	3.76%	0	0.00%
source=GSImaps/std	1 235 507	0.42%	32 283	0.97%	1 202 774	3.31%	450	0.28%
highway=service	952 215	0.32%	2	0.00%	952 172	2.62%	41	0.03%
yh:WIDTH=3.0m〜5.5m	914 844	0.31%	0	0.00%	914 843	2.52%	1	0.00%
landuse=farmland	834 304	0.28%	29	0.00%	831 416	2.29%	2 859	1.79%
source_ref=http://wiki.osm.org/wiki/GSI_KIB…	828 430	0.28%	1 600	0.05%	825 153	2.27%	1 677	1.05%
highway=footway	812 894	0.27%	34	0.00%	812 658	2.23%	202	0.13%
source=Japan␣GSI␣ortho␣Imagery	771 822	0.26%	16 030	0.48%	755 776	2.08%	16	0.01%
source=GSImaps/ort	722 265	0.24%	70 568	2.11%	651 442	1.79%	255	0.16%
yh:WIDTH=1.5m〜3.0m	701 252	0.24%	0	0.00%	701 252	1.93%	0	0.00%
highway=path	699 150	0.24%	11	0.00%	699 138	1.92%	1	0.00%
highway=track	691 228	0.23%	8	0.00%	691 217	1.90%	3	0.00%

Figure 2.2: Tag statistics on Taginfo Japan

Instead of downloading and processing the data manually, we opted to directly integrate the website's API into our code. This API provides us with a `.json` file containing detailed statistics for specific key-value pairs. For instance, to retrieve statistics about the tag "highway" with the value "residential," we simply access the following URL: `https://taginfo.openstreetmap.org/api/4/tag/stats?key=highway&value=residential`


```
url:            "https://taginfo.openstreetmap.org/api/4/tag/stats?key=highway&value=residential"
data_until:     "2024-06-26T00:59:11Z"
total:          4
data:
  0:
    type:            "all"
    count:           66508218
    count_fraction:  0.0065
  1:
    type:            "nodes"
    count:           37
    count_fraction:  0
  2:
    type:            "ways"
    count:           66508153
    count_fraction:  0.0643
  3:
    type:            "relations"
    count:           28
    count_fraction:  0
```


Figure 2.3: JSON information provided by the API

2.2 Data serialization

2.2.1 OSM national cycle routes data

After extracting the OSM data specifically for the national cycle routes using Osmosis commands, our next step was to serialize this data into a more readable JSON format, to process the data efficiently within our Python code.


```
 <node id="335084369" version="16" timestamp="2011-11-14T13:13:06Z" uid="0" user="" lat="34.8136029" lon="135.7146101"/>
-<node id="335841201" version="4" timestamp="2016-11-13T08:17:37Z" uid="0" user="" lat="34.6906984" lon="135.7886877">
   <tag k="KSJ2:LIN" v="奈良線"/>
   <tag k="crossing:barrier" v="yes"/>
   <tag k="crossing:bell" v="yes"/>
   <tag k="crossing:light" v="yes"/>
   <tag k="note" v="National-Land Numerical Information (Railway) 2007, MLIT Japan"/>
   <tag k="note:ja" v="国土数値情報（鉄道データ）平成19年　国土交通省"/>
   <tag k="railway" v="level_crossing"/>
   <tag k="source" v="KSJ2"/>
   <tag k="source_ref" v="http://nlftp.mlit.go.jp/ksj/jpgis/datalist/KsjTmplt-N02-v1_1.html"/>
 </node>
 <node id="335899486" version="4" timestamp="2023-09-04T14:40:17Z" uid="0" user="" lat="35.9391058" lon="140.0122759"/>
 <node id="336207792" version="2" timestamp="2023-01-01T08:55:08Z" uid="0" user="" lat="35.8579242" lon="140.2095186"/>
-<node id="336207793" version="3" timestamp="2022-12-27T13:11:28Z" uid="0" user="" lat="35.8578237" lon="140.2087064">
   <tag k="barrier" v="bollard"/>
   <tag k="bicycle" v="yes"/>
   <tag k="foot" v="yes"/>
 </node>
 <node id="336207794" version="3" timestamp="2022-04-14T03:28:29Z" uid="0" user="" lat="35.8577058" lon="140.2083261"/>
```


Figure 2.4: Data in XML-based OSM format

By doing this we got this kind of data elements:


```
id:          254450746
lat:         34.5885648
lon:         135.7696744
▼ tags:
  highway:   "traffic_signals"
```


(a) JSON node element


```
id:          37901091
▶ nodes:     […]
▼ tags:
  bicycle:   "designated"
  foot:      "designated"
  highway:   "cycleway"
  ref:       "814"
  segregated: "no"
```


(b) JSON way element

Figure 2.5: JSON data examples

Chapter 3

Characterizing what a national cycle route is

3.1 Analysis of tag statistics

In order to know what tags characterize a national cycle route, we needed to find which tags are relevant, or in other words, which tag values are more frequent in the national cycle routes than in average (whole Japan).

So first, we needed to calculate the frequency of each tag value in the OSM data we extracted using Osmosis. To do this, we created a Pandas dataframes from the ways, and sorted them by decreasing frequencies (number of NaN values) using the following command:

```
nan_count_ways = df_ways.isna().sum().sort_values()
nan_count_ways = (df_ways.shape[0] - nan_count_ways )/df_ways.shape[
    0]
```

Listing 3.1: Example JSON

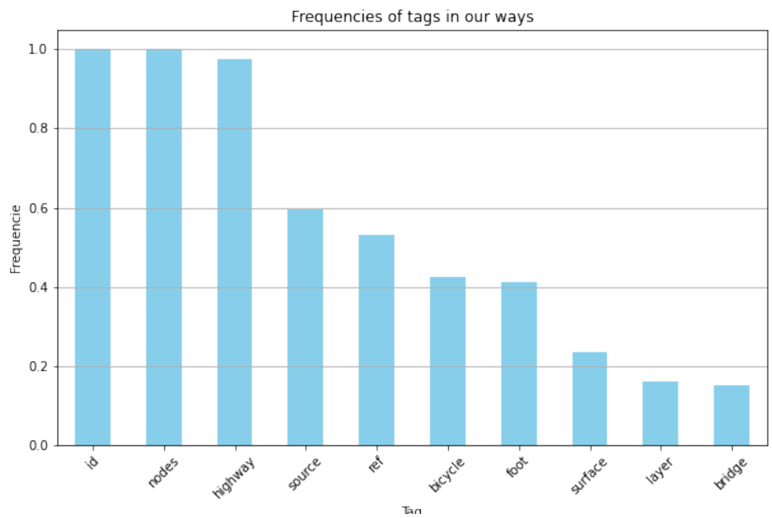


Figure 3.1: Most frequent tags from left to right

This data is interesting, but not relevant without comparing it with the average distribution of tags across Japan. Every ways has an Id and nodes composing itself, that is why we have a frequency of 1.0 .

Prior to comparison, we filtered out tags containing names (like road names or company names) and tags with more than 20 different values. Empirically, these tags typically contained irrelevant information, such as sources of road information.

Previously we have obtained the frequency of tags present in National Cycle Roads. To show which tags are actually relevant, we want to compare those specific frequencies, to the frequencies of tags on every ways in Japan (not just national roads). Then, after retrieving the global frequencies from an API, we applied the following formula to get the relevance of the tag value compared to the average (signed relative error):

$$rel = \frac{f_{ncn}}{f_{Japan}}$$

With:

- f_{ncn}: frequency of the tag value inside the ncn sample
- f_{Japan}: frequency of the tag value inside the whole Japan OSM map

However, there are some tags with a frequency of 0 (slightly higher, but rounded to the fourth digit), and dividing by zero is obviously not possible.

These low frequencies indicate that these tag values are extremely rare, so removing them is not an issue because they would not typically appear when extracting data from OSM with Osmosis. The same applies to frequencies slightly above 0. Therefore, we have implemented a threshold of 0.0001, below which tag values are removed from consideration. The number of tags went from 355 to 132.

By doing this, and applying the previous formula, we get the following order of relevance for the different tag values:

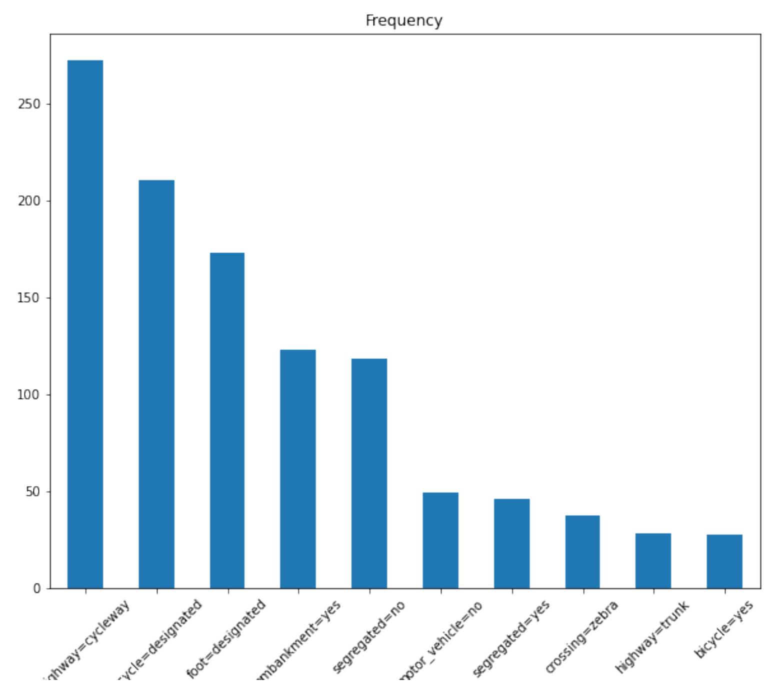


Figure 3.2: 10 most relevant tag values

As you can see, the most relevant tags make sense with what we want to achieve:

- highway=cycleway: a separate way for the use of cyclists in the highway
- bicycle=designated: the road is explicitly designated for use by cyclists
- foot=designated: preferred route for pedestrians (equivalent to foot=yes according to the wiki)
- embankment=yes: the road is placed on a raised bank, with a downward slope on one or two sides

- segregated=no: the cycleway or footway is shared between cyclists and pedestrians
- motor_vehicle=no: access is prohibited for motor vehicles
- segregated=yes: the cycleway or footway is not shared between cyclists and pedestrians*
- crossing=zebra: crossing marked with a striped pattern, but no traffic signals
- highway=trunk: the most important roads in a country's system that aren't motorways
- bicycle=yes: paths where cycling is legal but not designated

Those definitions are taken directly from OSM wiki: [5]

*NB: As you can see, the tag "segregated" includes both values "yes" and "no" among the top 10 relevant tags. This occurrence is because these tags appear exclusively in cycleways and footways, making both frequent compared to the entire Japan map. So, actually, both are relevant. However, "segregated=no" has a score more than twice that of "segregated=yes", and this will be taken into account in the scoring formula later.

Since our aim is to filter the entire map of Japan using the most relevant tags, we have chosen to apply a threshold of 10. This means we will select only those tags whose frequency is above 10. This leaves us with 35 tags.

Nevertheless, all the tags will be included in the scoring formula at the end. Especially, negative tag values such as "building=yes" will be really relevant: the value is almost -1, which means that it does not exist in the national cycle routes sample.

3.2 Incorporating elevation information

However, since elevation information cannot be used as a criterion in the Osmosis search, it will only be integrated into the final scoring algorithm, which will assist us in determining which roads are most suitable to be designated as new national cycle routes.

Chapter 4

Finding new national cycle routes

4.1 Filtering the possible candidates

Now that we had identified the 35 most relevant tags for identifying new national cycle routes, we proceeded to filter all ways in the OSM Japan map using Osmosis. This time, we specified the 35 tags of interest:

```
osmosis \
  --read-pbf file="./Data/osm_map/kansai-latest.osm.pbf" \
  --wkv highway.cycleway,bicycle.designated,foot.designated,
      embankment.yes,
   segregated.no,motor_vehicle.no,segregated.yes,crossing.zebra,
      highway.trunk,
   bicycle.yes,access.no,foot.yes,tunnel.yes,bridge.yes,crossing.
      unmarked...
  --used-way \
  --used-node \
  --write-xml file="./Data/test_wkey.xml"
```

Listing 4.1: Osmosis extraction command

When visualizing the outcome using JOSM, the following is obtained:

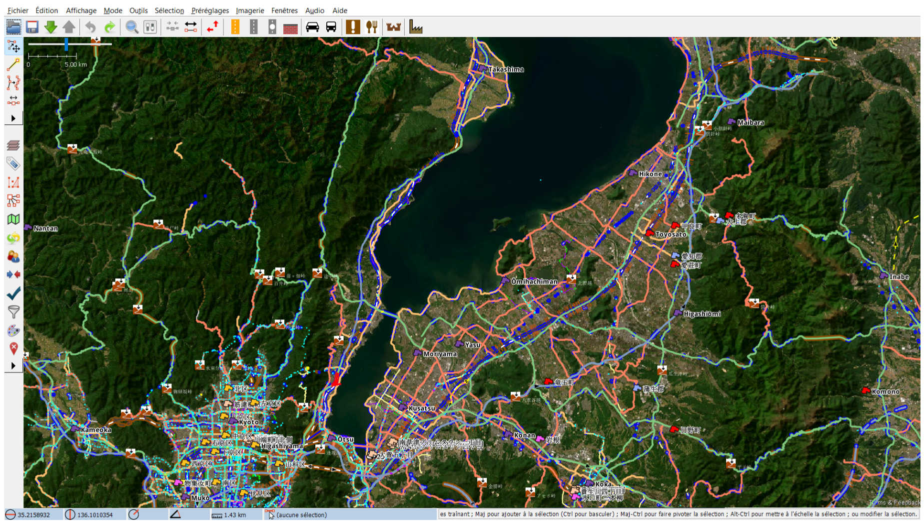


Figure 4.1: Visualization of the filtered tagged ways

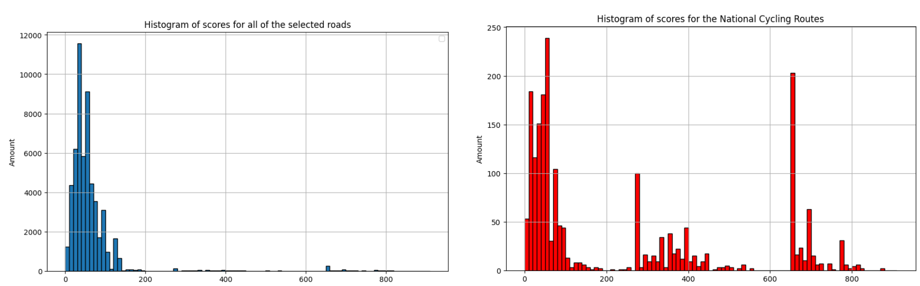
As we can see, there is still a significant number of tagged ways (specifically 55,774). However, many roads have not been retained, indicating that the tags were correctly applied in the filtering process.

And of course, the Biwako line is included in the filtered tagged ways.

4.2 Scoring the ways

Now, we want to append a score to our 55,774 candidates to identify the best ones. To do this, we calculate the weighted sum of the tag value scores obtained previously: having the most relevant tag value will give more points than having the others.

This can also be interpreted as comparing how close the average attribute vector for the NCN roads and the one of the current road are.



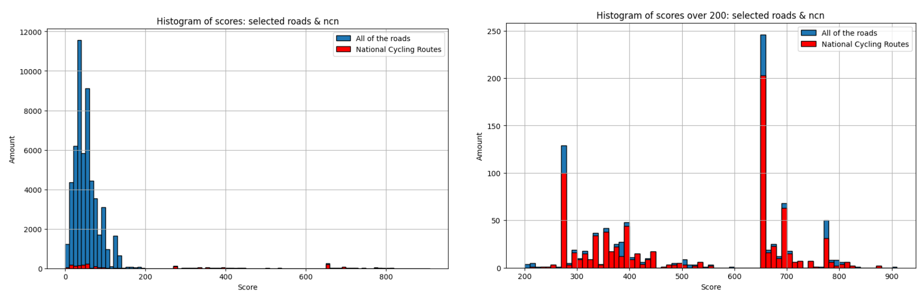
(a) Scores for the selected roads

(b) Scores for the NCN roads

Figure 4.2: Histogram of scores

To visualize the results, we opted to plot histograms depicting the distribution of scores, with each bin representing a range of 10 for clarity. Figure (a) illustrates that the majority of roads have scores below 200. However, Figure (b) reveals a significant contrast when considering only the NCN roads: comparatively, many of these roads have scores above 200.

This disparity suggests that NCN roads generally perform better, validating our analysis. It's important to note that while NCN roads tend to excel overall, there are exceptions where roads may have lower scores: it's logical because we only take into account the average attributes vector. Moreover, they may have only a few tags, which can contribute to these variations.



(a) All of the scores

(b) Lookup for scores > 200

Figure 4.3: Comparison of the histograms

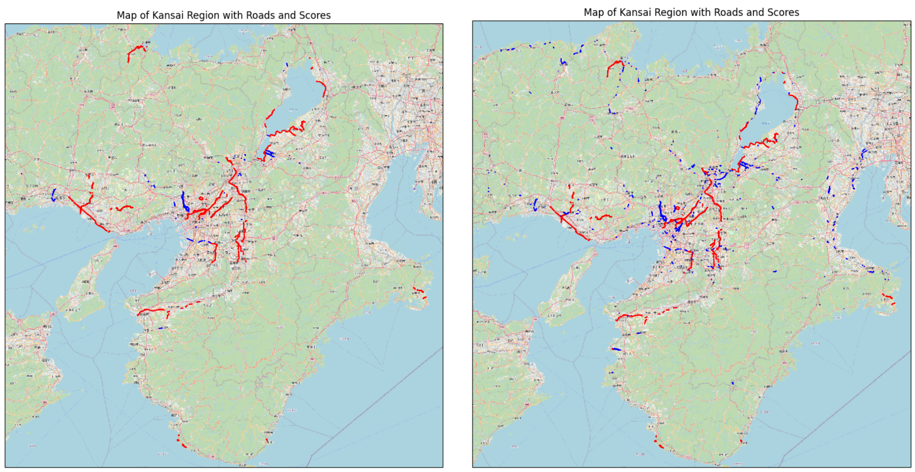
To better visualize the distributions, we plot in blue all the road and in red the NCN roads. In figure (a) we can already see that most of the roads over 200 are red, and by zooming on the range [200-900] in figure (b), we clearly see that it is the case. Even though most of the ways are NCN roads, we still find new roads in blue. Those are really interesting because it corresponds to the non-NCN roads that are the most close to the NCN roads, and are close to be NCN roads.

To enhance the visualization of score distributions, we utilized blue for all roads and red specifically for NCN roads. In Figure (a), we observe that the majority of roads scoring over 200 are depicted in red. Zooming into the range [200-900] in Figure (b) provides a clearer view, confirming this trend.

Despite the predominance of NCN roads in this range, there are still some blue instances. These roads are particularly interesting as they are non-NCN roads that closely resemble NCN roads. They are potential candidates that exhibit characteristics akin to NCN roads.

To validate our analysis and gain a deeper understanding of the distribution of roads scoring over 200, we will overlay them on a map of Kansai using the same color scheme: blue for all roads and red specifically for NCN roads.

This visual representation serves two main purposes. By mapping these roads, we aim to confirm the preservation of NCN roads within our analysis. Additionally, we seek to identify potential convincing new routes.



(a) Score above 200 (b) Score above 125

Figure 4.4: Visualizing the best roads: ncn and new roads

Figure (a) demonstrates that our method successfully preserves the primary roads depicted in Figure 4.1, except for some along the Biwako line. However, it also indicates that selecting only roads with a score above 200 is insufficient for identifying potential candidates for new NCN routes, except for a small blue section north of Osaka, following the Ina River.

To address this limitation, we decided to lower the threshold to 125. Although this represents a significant reduction, it effectively doubles the number of roads considered (from 1000 to 1997). While some of these roads are less optimal compared to the NCN roads, they are still considerably better than the majority of roads, with a median score of 75.71. This adjustment allowed us to identify several promising new candidates:

1. On the right side, a blue route appears to extend the existing red NCN route in Mie Prefecture, stretching from there up to Nagoya along the coast.
2. A small but well-defined route follows the Inagawa River in the western part of Osaka. (even with the theshold set to 200)
3. A minor road connects Kyoto to Biwako.
4. An alternative route to the Biwako National Cycling Route that passes through Takajima City in favor of a lakeside path.

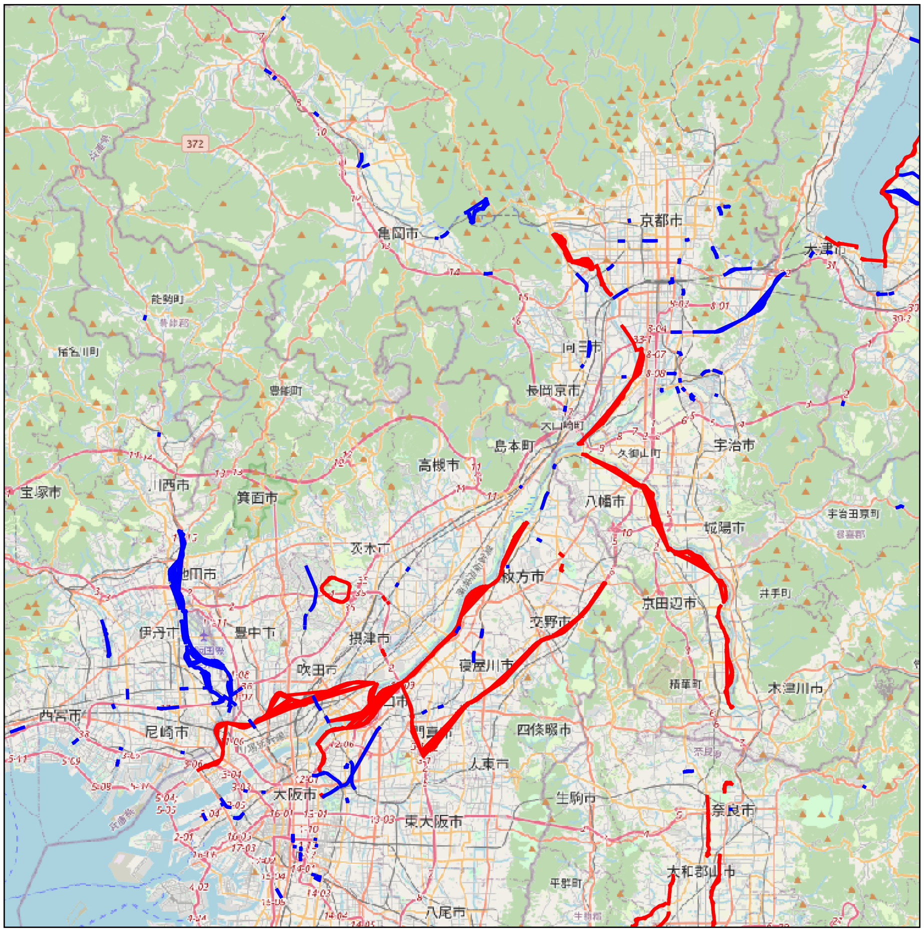


Figure 4.5: Zoom on the Osaka-Kyoto area (T = 125)

Chapter 5

Conclusion

5.1 Areas of improvement

We have to nuance our results because our analysis is biased by the distribution of OSM data: roads in densely populated areas tend to have more tags compared to less frequented roads, simply because there is more information about them. There is also a bias with the length of the ways: since the length can vary a lot, longer ways tend to have more tags simply because they cover more space.

Moreover, until now, we have mostly used criteria related to the state of a way, such as the material of the road, the presence of a cycle lane, and whether it is shared with pedestrians. There are several criteria that we can add to improve our scoring system. One of the recommended criteria for a national cycle road is that the route should avoid sections with continuous steep inclines. We could use the extracted data to measure the steepness of a road and modify the score based on this steepness.

What we could also do is study ways in relation to each other instead of individually. The problem with our previous research is that we only find segments of roads. A good road should continue for at least 100 km. So, we could try to find routes that can connect good segments of the route to make it a long bicycle road. A good road should also have facilities to rest or bicycle repair stations every 50 km. Finding such criteria requires us to analyze the roads more in-depth and find spatial relations between them.

Bibliography

[1] *Elevation Data in Japan*. https://fgd.gsi.go.jp/download/menu.php. (Accessed on 08/01/2024).

[2] *Geofabrik Download Server*. https://download.geofabrik.de/asia/japan.html. (Accessed on 08/01/2024).

[3] *Geofabrik Download Server*. https://download.geofabrik.de/asia/japan/kansai.html. (Accessed on 08/01/2024).

[4] *National Cycle Route — GOOD CYCLE JAPAN — Ministry of Land, Infrastructure, Transport and Tourism*. https://www.mlit.go.jp/road/bicycleuse/good-cycle-japan/national_cycle_route/english.html. (Accessed on 08/01/2024).

[5] *OpenStreetMap Wiki*. https://wiki.openstreetmap.org/wiki/Main_Page. (Accessed on 08/01/2024).

[6] *Tags — OpenStreetMap Taginfo Japan*. https://taginfo.openstreetmap.jp/tags. (Accessed on 08/01/2024).