THE ARCHIPELAGO BAROMETER, SOURCE OF DATA AND PROCESSING

PM by Joakim Franklin, WSP. Joakim.franklin@wsp.com

About the Archipelago barometer

The main objective of the Archipelago barometer is to answer the question of how many people visit the Stockholm archipelago. The definition of the Stockholm archipelago used in the study is islands and marine areas along the Baltic coast that lack permanent mainland connections and are part of the Stockholm County. This is an area spanning from Arholma in the north to Landsort in the south. This area matches the archipelago definition used by Region Stockholm for regional planning and development.
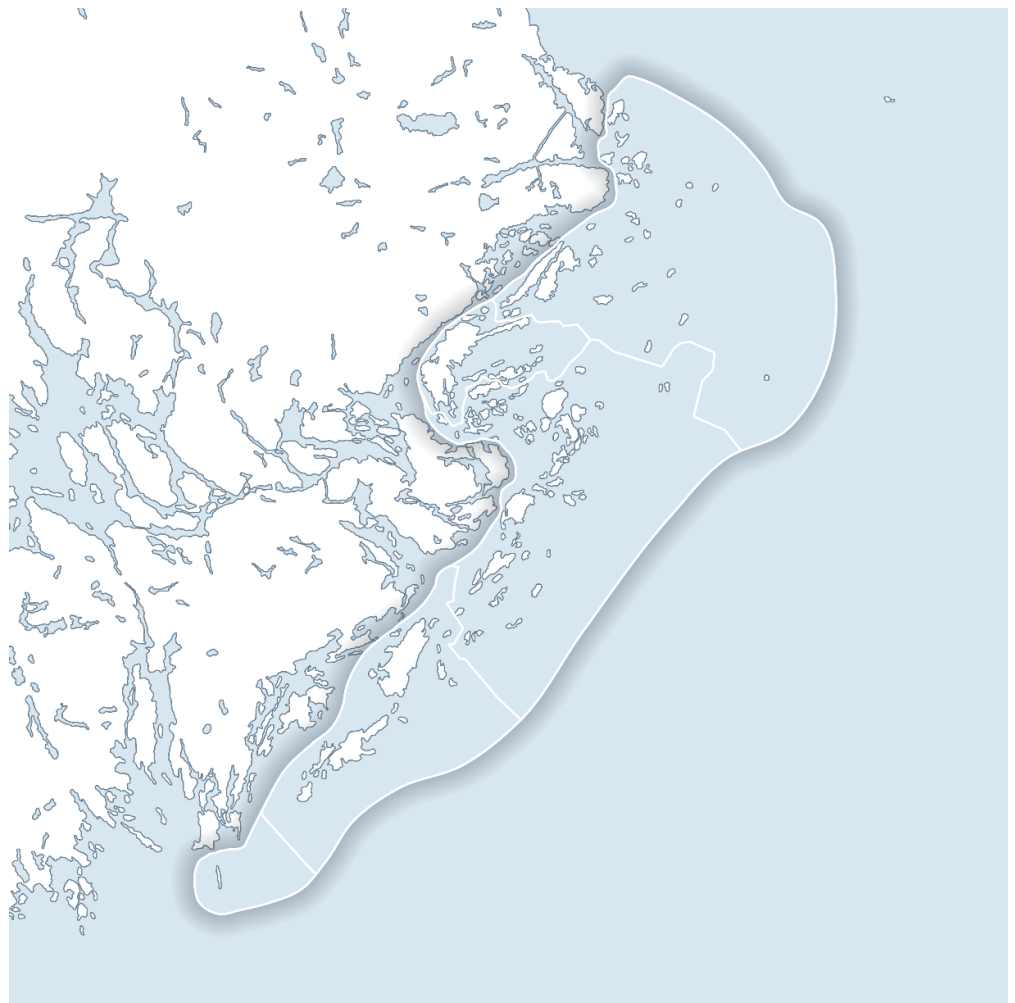

Fig. 1. The defined area of the Stockholm archipelago with municipality borders.

The arcipelago area is analyzed as a whole and per municipality, using the part of each municipality that falls within the archipelago area. The archipelago barometer also estimates the number of visitors in all areas administrated by the Archipelago Foundation (Skärgårdsstiftelsen) and designated transit nodes for the public archipelago traffic in the Stockholm Region (replipunkter).

About cellular positioning data and TCI

All cellphones connect to communication antennas and repeatedly ping them to keep the connection. By knowing what antenna a cellphone is connected to, when and for how long to it is possible to spatially track the movement of cellphones and their users. For this, Telia Company has developed Telia Crowd Insight, TCI. TCI provides an estimated number of people in an area. This is achieved using two steps:

TCI first sets an estimated position of every cellphone in the Telia network. This is done using information about what antenna the cell phone is connected to. TCI also utilizes detailed geographical information about land usage in this process. As an example, if TCI can choose to position a cellphone user on land or in water, it prioritizes land. By default, it also prioritizes residential, commercial or industrial areas over agricultural or forest covered areas[1].

After deciding estimated positions of cellphones TCI extrapolates the count of Telia cellphones in the area of interest to an estimated total number of cell phones in the same area, regardless of service provider. This extrapolation utilizes detailed knowledge about Telia's share of the cell phone market on a zip-code level together with knowledge about where the cellphone spends the majority of its time before 9 am. This position sets the "home" of the cellphone. The extrapolation method in use was developed by Telia in collaboration with Sweden's national statistics authority, SCB.

TCI keeps track of all cellphones connected to the Telia Network 24/7. The identity of users is kept safe in several ways. Every night all cellphones are re-anonymized and given a random ID for the coming 24 hours. It's not possible to track a cellphone for longer than this period. Before presenting results TCI also checks for results of less than five people per cell in the result table. Cells with 0-4 people are treated with secrecy and blanked. TCI also excludes cellphones belonging to users under 18 years of age. Last, as of today TCI does not include observations of foreign cellphones connected to the Telia network.

The result from TCI is thus an estimated number of people in an area, representing domestic cellphone users 18+ years. In the result it's possible to read the following:

- Date
- Area of interest
- Home – Says where the cellphone's first ping of the day was made. In Skärgårdsbarometern this is presented per municipality.
- Overnight / Not overnight – Says whether the cellphone's last ping of the day was made in the area of interest or if the cellphone left the area before that.
- Number of people

---

[1] If needed, these settings can be changed.

Separating visitors from residents

The scope of the Archipelago barometer is to give an estimated number of visitors in the Stockholm archipelago. The results from TCI must therefore be processed to exclude residents in the area of interest. This is achieved using population data on a grid level together with the TCI data on a day-per-day basis.

The official population of the archipelago area of interest is around 4700 individuals according to the national statistics authority of Sweden, SCB, with some difference between years. This number is subtracted on a daily basis from the number of overnight people in the archipelago area. When the number of overnight people is less than this, the number of overnight visitors is set to 0. The number of not overnight visitors is not changed. The same method is used to isolate visitors from residents at transit nodes (replipunkter). In this case both TCI data and population data is based on matching geographical grids.

For areas administrated by the Archipelago Foundation (Skärgårdsstiftelsen) the method described above would impose challenges because the population grid doesn't match the polygons from TCI. Many areas are sparsely populated which makes secrecy in population data an issue. Instead, a different approach is used that is based on the TCI data itself.

The seasonal differences in number of visitors to the Stockholm archipelago are vast. This can be observed both in number of accommodation guests and in number of open accommodation facilities on a monthly basis. Many accommodation facilities are closed during winter and the number of guests during June to August can often exceed the number of guests during the rest of the year. This seasonal pattern is especially present in the parts of the Stockholm archipelago that lack permanent mainland connections, to which a majority of areas administrated by the Archipelago Foundation belong. This known seasonal pattern of visitors is therefore used to estimate the number of residents in close vicinity to each area administrated by the Archipelago foundation. This is made by examining the number of overnight people according to TCI during weekdays in January and February, when the number of overnight visitors in the areas of interest is believed to be negligible. What we are looking for is a recurring minimum level of people, not the absolute minimum. This number is then used as an estimated number of residents in close vicinity to the area of interest and subtracted from the TCI results on a daily basis. When the number of overnight people is less than this recurring minimum, the number of overnight visitors is set to 0. The number of not overnight visitors is not changed.

Compensating for data loss

The confidence of the estimates made by TCI varies on a daily basis. Atmospheric conditions, updates as well as technical errors affect the confidence of the data the TCI uses to position cellphones and extrapolate them to represent a total population. TCI rates every day based on the signal quality, which is reported on municipality level for each day based on the percentage drop from the long-term baseline calculated as the 6-week moving median. If the drop is more than -15% the data is considered inconsistent. During the period of 2019-2021 this happened 19 out of 1096 days (1,7%). Because of the vast seasonal variations in visits a few low rated

days occurring in the busy summer period can affect the results even on a yearly basis, even though the share of low rated days is less than 2%. Before presenting results the low rated days are processed to compensate for this data loss. Low rated days without results are given estimated number of visitors based on the weekday and season. This is made by examining visitor patterns during the weeks leading up to and following each day that lacks results. This is the last step of processing the data from TCI, after which the pattern of visitors can be analyzed and presented.