



Combining Mobile Phone Data and Survey Data for the Best Result: Experience from Indonesia

Sarpono^{1}; Titi Kanti Lestari²*

^{1,2} BPS-Statistics Indonesia

Abstract

Data collection is complicated in Indonesia due to vast geographic distances and difficulty of travel. So is it that the inbound tourism survey for measuring tourism is under coverage. Improving the accuracy of inbound tourism statistics with mobile positioning data. Mobile positioning data (MPD) is considered as one of the most promising big data for measuring the mobility of people, including mobility of tourists based. It holds more information, is much faster and more reliable. However, as a new data source, there also challenges and limitation of big data (including mobile positioning data) that have to be taken into account in order to become a valuable and quality statistics. In MPD there is a lack of qualitative data on tourism motivation and the sample does not include non-roamers. One way to overcome the limitations of big data is by combining it with small data obtained from a survey. This paper shows that combining big and small data will provide optimal results. We used mobile positioning data about roaming activities of cross-border tourists in Indonesia at border areas and mobile usage survey conducted at the borders in order to know the motivations for cross-border movements (both roamers and non-roamers). Then, we come up with the formula that combines big and small data to obtain the best result for tourism statistics.

Keywords: Big Data, tourism statistics, remote area

1. Introduction

Indonesia has a border land with Malaysia, Timor Leste, and Papua New Guinea along 3092.8 km. While, the sea area borders with 10 countries, namely India, Malaysia, Singapore, Thailand, Vietnam Philippines, Australia, Timor Leste, Palau, and Papua New Guinea. This sea border covers 92 leading small island, starting from Miangas Island in the north to Dana Island in the south.

During this time foreign tourists were calculated based on the Immigration Office based on the passport swipe, which recorded the traffic of all people entering Indonesian territory. Since the vast condition of the Indonesian territory with diverse border areas (sea and land) and the limitations of the Immigration Office, not all foreign tourists entering Indonesian territory are recorded regularly and on time. There are still many border regions of Indonesia with neighboring countries that are traditional, so there is no recording of people entering and leaving Indonesian territory. Therefore, to obtain a complete and up-to-date data on foreign tourists, data collection was carried out to calculate the number of foreign tourists visiting these border areas. With the addition of these data, the data of foreign tourists visiting have more coverage and can describe the actual conditions of foreign tourists.

To increase the data coverage on the number of foreign tourist visits, especially in border areas that have not been recorded, the BPS Statistics Indonesia and Ministry of Tourism tries to improve the methodology to calculate the number of foreign tourists visiting through the border gates using big data, namely Mobile Positioning Data (MPD) since October 2016. MPD is used in cross border posts in districts where immigration checks are not available and cross-border postal surveys are difficult due to geographical conditions.

This paper shows that combining big and small data will provide optimal results. We used mobile positioning data about roaming activities of cross-border tourists in Indonesia at border areas and mobile usage survey conducted at the borders in order to know the motivations for cross-border

movements (both roamers and non-roamers). Then, we come up with the formula that combines big and small data to obtain the best result for tourism statistics.

2. Methodology

Statistical office, currently, has a challenge to adapt to the rapid technological changes, which produce big data. The main challenges for big data for official statistics is obtaining the data sources. The main and considered as one of the gold standard sources of big data is Mobile Call Data Record. Several information can be obtained from Call Detail Record (CDR) such as the location, duration, the phone type, etc. Having this broad information, the use of CDR now not only limited for mobile transaction purpose. Broad range of the CDR use from Business, politics, education to official statistics. Šćepanović et al (2015) showed how the Mobile Phone Call Data can be used as a Regional Socio-Economic Proxy Indicator. Furthermore CDR can also be used for inferring people migration (Sniowski et.al 2016). Statisticians are stimulated to use big data whether to complement official statistics or to produce indicators. Big data offers great potential for monitoring the sustainable development goals and it has been promoted as a more timely and cheaper alternative to traditional sources of official data (Abdulkadri, Evans, and Ash, 2016). Better decision-making and real-time citizen feedback as the result of more diverse, integrated, timely and trustworthy information which in turn enables everyone to make choices that are good for them and for world they live in (Morales et.al., 2014).

Mobile positioning data (MPD) is considered as one of the most promising big data that can be used for official statistics. Statistics Netherlands studied several uses of mobile phone data for official statistics, it confirmed that mobile phone data may be of use to statistical topics varying from economic activity, tourism, population density to mobility and road use (de Jonge, 2012). The outcomes of the Eurostat study concluded that at present mobile positioning data can be used as a supplement rather than as a replacement source of data for the current official tourism indicators. Furthermore, the study commissioned by Eurostat explained the use of mobile data as a source for tourism indicators, this new source of data can improve timeliness (in some cases up to near-real time), access to statistical information previously not available (new indicators) calibration opportunities for existing data, better resolution, and accuracy in time and space. In Estonia, MPD has been used as an official source of travel statistics since 2008 (Kroon, 2012).

As a new data source, there also challenges and limitations with big data (including mobile positioning data) that have to be taken into account in order to produce valuable and quality statistics, especially regarding its accuracy. In MPD there is a lack of qualitative data on tourism motivation such as the purpose of the trip, expenditure, type of accomodation and means of transport used (Eurostat, 2014) and the sample does not include non-roamers.

One way to overcome the limitations of big data is by combining it with small data obtained from a survey. MPD still has weakness such as related to privacy issues or confidentiality of the costumers and surveillance and also still expensive. In Indonesia, the use of MPD for Tourism Statistics have been initiated since 2016 by collaboration among Indonesia Ministry of Tourism, Statistics Indonesia, and the main MDP Company in Indonesia, Telkomsel. In order to know the ground truth of MPD, BPS Statistics Indonesia, in collaboration with Ministry of Tourism, conducted an extended Cross Border Mobile Usage Survey in 2017 to know the behavior of the border-crosser in using their mobile also the characteristics of the border crosser. The number of entry gates covered was higher than usual. The aim of the survey was also to obtain information that will be used to form ratios for the formula of additional tourism.

BPS-Statistics Indonesia started to use MPD since October 2016 for inbound tourism statistics collaboration with Indonesia Ministry of Tourism and the main Mobile Network Operators in Indonesia, Telkomsel. MPD is a method of tracking the locations of mobile devices in time and

space, collected by MNOs and mobile app developers (Tiru, 2014). Prior to MPD use, Indonesia used administrative data (immigration data) and cross-border (shuttle) survey. The cross border survey were quite expensive due to the borders areas being remote, and the transportation costs to survey locations are high. Also, the survey is only conducted in a month in selected locations, to estimate the number for a year for the entire border. So, there is a coverage problem in the tourism data in Indonesia. If we compare with the neighboring countries, in Indonesia the neighboring countries only constituted about 7 percent in 2015, while for other countries the neighboring countries constitute about 40 percent of tourism.

Mobile positioning data is used to complement tourism data at cross border posts in which Immigration Checkpoint is not available and difficult to conduct Cross-Border Survey. Before releasing the tourism data, BPS compares international visitor arrival data obtained from cross-border survey, immigration checkpoint at cross-border area, and mobile positioning data. If mobile positioning data from Telkomsel (the mobile network operator) is more than data from cross-border survey and immigration checkpoint, the excess of mobile positioning data will be added to the number of international visitor arrival, this is the number of additional international visitor arrivals from mobile positioning data.

However, the use of MPD in the calculation of visits of foreign tourists is not without obstacles. MPD has several shortcomings, among them not all foreign SIM card users are residents abroad. In addition, there is also no information about the number of SIM cards used per person. MPD also cannot provide information about the characteristics of SIM card owners such as the purpose of the visit and expenditure. Other information such as how many strangers passing by who do not carry a mobile phone or use a local sim card also cannot be known with the MPD.

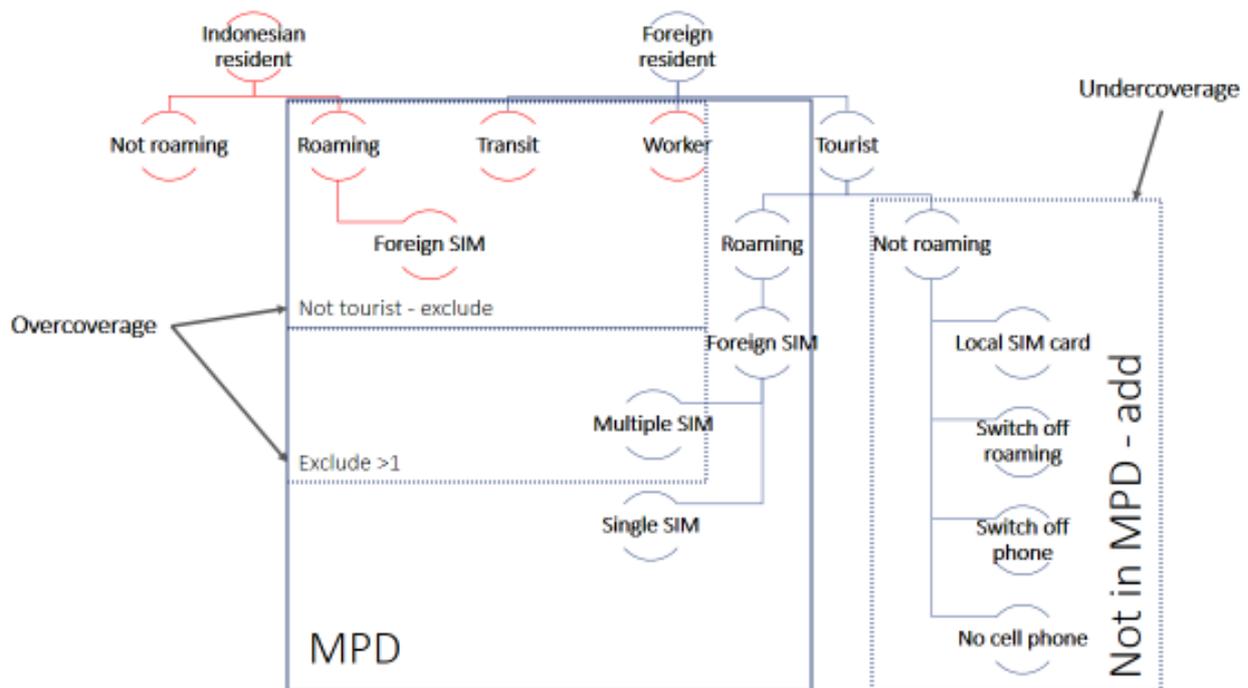


Figure 1. Over and under coverage in MPD

The data collection resulted in cross border ratio as correction factor for MPD data. These values will be used as a basis for calculating the additional number of foreign tourists visiting in the

border areas. The first value is the average foreign SIM card brought by the foreign SIM card holders. This number is calculated from the total number of foreign SIM cards divided by the number of foreign SIM card holders, regardless of whether the card holders are local residents or foreigners. It is also necessary to calculate the ratio of foreign residents who are not recorded in MPD because they do not carry mobile phones or replace their SIM cards with local SIM cards (not roaming).

All the above ratios will be used as a correction factor in calculating the number of visits by foreign tourists by MPD. To facilitate the calculation of the additional number of foreign tourists from the MPD, we then made a formula that includes all the above correction factors:

$$AT = \frac{MPD}{X_{roam}} \times \frac{1}{1 - P_{NR}} \times \frac{1}{MS} - WCI$$

Where:

- AT* = Additional Tourism
- MPD* = Number of SIM cards detected by MNO (Telkomsel) in the border area
- X_{roam}* = The ratio of foreign SIM cards per person that actively roaming;
- P_{NR}* = The ratio of foreign residents with foreign SIM cards who turn off their phone, roaming or switch to local SIM card to total number of foreign residents with foreign SIM cards.
- MS* = Market Sharing
- WCI* = Number of tourists entering through Immigration Post

3. Result

As mentioned above, BPS Statistics Indonesia conducted cross-border mobile usage survey for inbound tourists at the border area. This survey is conducted by BPS Statistics Indonesia, in collaboration with the Ministry of Tourism, to obtain the ground truth of mobile positioning data, to know the behavior of border crosser in using their mobile and to obtain ratio/proportion for the formula.

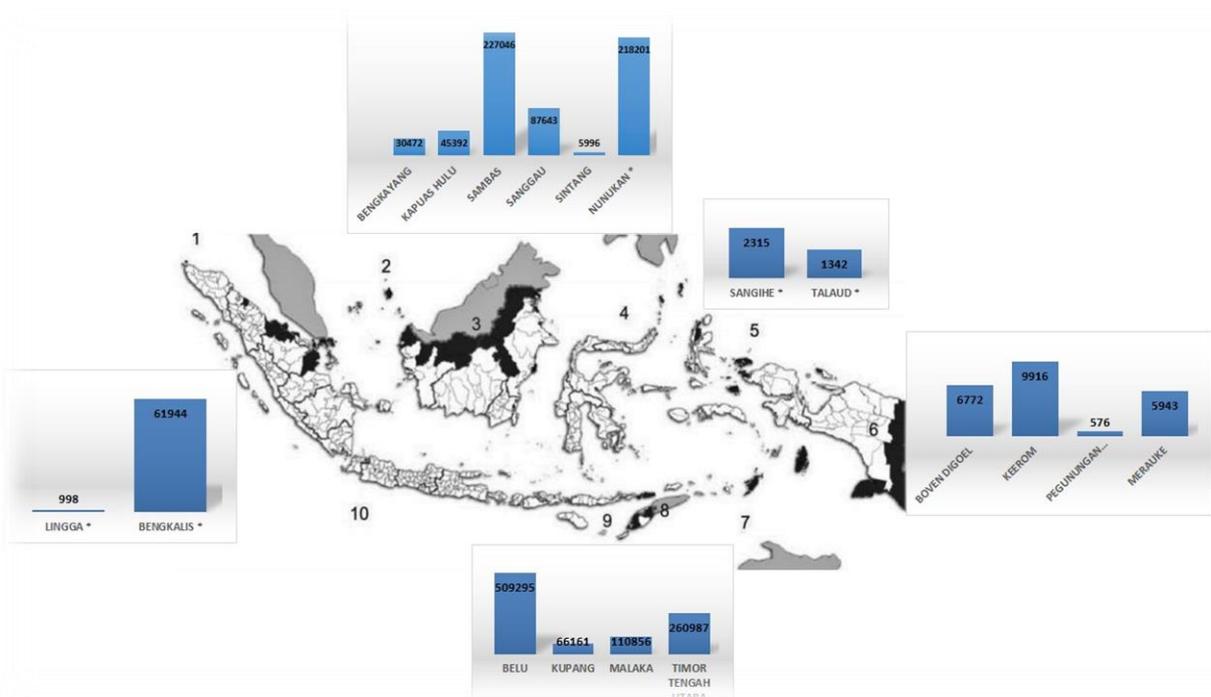


Figure 2. Number of Additional Tourism in Border Areas Jan-July 2018

Figure 2. Shows a total number of additional tourism in border area January to July 2018. The biggest additional number is 509 295 tourists in Belu regency, Nusa Tenggara Timur and the smallest is 576 in Pegunungan Bintan, Papua. Number of additional tourism in border area prove that use the MPD capture significantly and increase the coverage of inbound tourism in Indonesia.

4. Discussion and Conclusion

MPD is useful for BPS Statistics Indonesia as it gives more accurate data on the tourism arrival figures compared to cross-border survey, which can only be conducted during one month and in limited geographical area to estimate the whole year and the entire border. However, there also limitations to MPD. So, there is no data source that is superior compared to other data source. All of data sources could complete each other.

It is important to design a mobile usage survey to accompany the MPD. Moreover, from the cross-validation results, both MPD and survey had their limitations and weaknesses which were apparent once both data sources were compared and used together.

Survey can only be conducted once or twice a year, while MPD is obtained every month with quite high accuracy. After MPD is used, the proportion of cross-border tourism for Indonesia now exceeds 30 percent, which aligns to international benchmarks.

The use of mobile positioning data changes the process of design, build, data collection, data processing, and dissemination. It also changes the process of data collection so that it can enhance the cost efficiency and time efficiency. Real time dissemination also can be achieved by using mobile positioning data. Big data as a part of data revolution needs to be developed, although the verification and its validation need to take into consideration in order to prevent double counting.

This paper showed that combining MPD and survey data will provide optimal results. Mobile positioning data about roaming activities of cross-border tourists in Indonesia at border areas was combined with mobile usage survey at the borders in order to know the motivations and behavior for cross-border movements (both roamers and non-roamers). Then, the formula of additional tourism that combines MPD and mobile usage survey were proposed to obtain the best result for tourism statistics.

References

- Abdulkadri, Abdullahi, Alecia Evans, and Tanisha Ash. (2016). *An Assessment of Big Data for Official Statistics in the Caribbean: Challenges and Opportunity*, ECLAC Subregional Headquarters for the Caribbean. Santiago: United Nations.
- Ahas et.al. (2007). *Mobile Positioning Data in Tourism Studies and Monitoring: Case Study in Tartu, Estonia*, accessed on 25 January 2017, https://www.researchgate.net/publication/221357419_Mobile_Positioning_Data_in_Tourism_Studies_and_Monitoring_Case_Study_in_Tartu_Estonia.
- Ahn, Jeong-Im and Young-Ja Hwang. (2013). *Production of Official Statistics by Using Big Data*, Working Paper, accessed on 16 March 2017, http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.50/2013/Topic_3_Korea.pdf.
- BPS. (2015). *International Visitor Arrival*. Jakarta: Badan Pusat Statistik.
- BPS. (2017). *Emerging Challenges in Data Collection, Survey Methodology and Implications for Official Statistics (NSO Perspectives)*, presented in the Plenary

Session, International Statistical Institute Regional Statistics Conference (ISI RSC) 2017, Bali.

Kroon, J. (2012) Mobile Positioning as a Possible Data Source for International Travel Service Statistics. UNECE Conference of European Statisticians

Landefeld, Steve. (2014). Uses of Big Data for Official Statistics: Privacy, Incentives, Statistical Challenges, and Other Issues, Discussion Paper, accessed on 16 March 2017, <https://unstats.un.org/unsd/trade/events/2014/beijing/Steve%20Landefeld%20-%20Uses%20of%20Big%20Data%20for%20official%20statistics.pdf>.

Morales et.al. (2014). A World that Counts: Mobilising the Data Revolution for Sustainable Development. The United Nations Secretary-General's Independent Expert Advisory Group on a Data Revolution for Sustainable Development (IEAG).

Sanja Šćepanović, Igor Mishkovski, Pan Hui, Jukka K. Nurminen, and Antti Ylä-Jääski. (2015). Mobile Phone Call Data as a Regional Socio-Economic Proxy Indicator.

Seynaeve, Gerdy and Christophe Demunter. (2016). When Mobile Network Operators and Statistical Offices Meet-Integrating Mobile Positioning Data into the Production Process of Tourism Statistics, the 14th Global Forum on Tourism Statistics.

Sniowski, Alessandro, Sorichetta, Ingmar Weber, and Andrew J. Tatem (2016). Inferring Migrations: Traditional Methods and New Approaches based on Mobile Phone, Social Media, and other Big Data. *Technical Report*. European Union.

Rein Ahasa, Anto Aasaa, Antti Roosea, Ülar Markb, Siiri Silma (2008). Evaluating passive mobile positioning data for tourism surveys: An Estonian case study. *Tourism Management*, Vol 29, Issue 3, pp 469–486.

Tiru, Margus. (2014). Overview of the Sources and Challenges of Mobile Positioning Data for Statistics, presented in the International Conference on Big Data for Official Statistics. Beijing: UN Trade Statistics, Positium, and University of Tartu.