Indonesia's Experience of using Signaling Mobile Positioning Data for Official Tourism Statistics

Titi Kanti Lestari; BPS Statistics Indonesia, titi@bps.go.id

Siim Esko, Positium LBS, siim.esko@positium.com

Sarpono, BPS Statistics Indonesia

Erki Saluveer, Positium LBS

Rifa Rufiadi, BPS Statistics Indonesia

Abstract

BPS-Statistics Indonesia has used mobile positioning data for official statistics since October 2016. Mobile positioning data from the largest mobile network operator measures cross-border tourism arrivals across over 3000 km of land border and a vast sea border.

Prior to mobile positioning data use, Indonesia used administrative data (immigration data) to measure visitor arrivals in the border areas. Immigration data has its coverage issues and where there is no immigration checkpoint, cross-border surveys aim to fill the gaps. However, surveys in remote cross-border areas are expensive, and the transportation costs to survey locations are high. The survey is only conducted over a month in selected locations, to estimate the numbers for the whole year for the entire border. So, there was a coverage problem in the tourism data in Indonesia, and mobile positioning data aimed to solve that.

The specific type of mobile positioning data used, signaling data, detects on average 3.47 times more roamers at the border areas than call detail records (CDRs). That ratio differs for each particular area. We found that signaling data overcomes some undercoverage issues of CDRs, while it is also adds noise - statistical and non-statistical - created by special types of roamers such as those flying over the country, crossing the country's seas and accidentally roaming across the border.

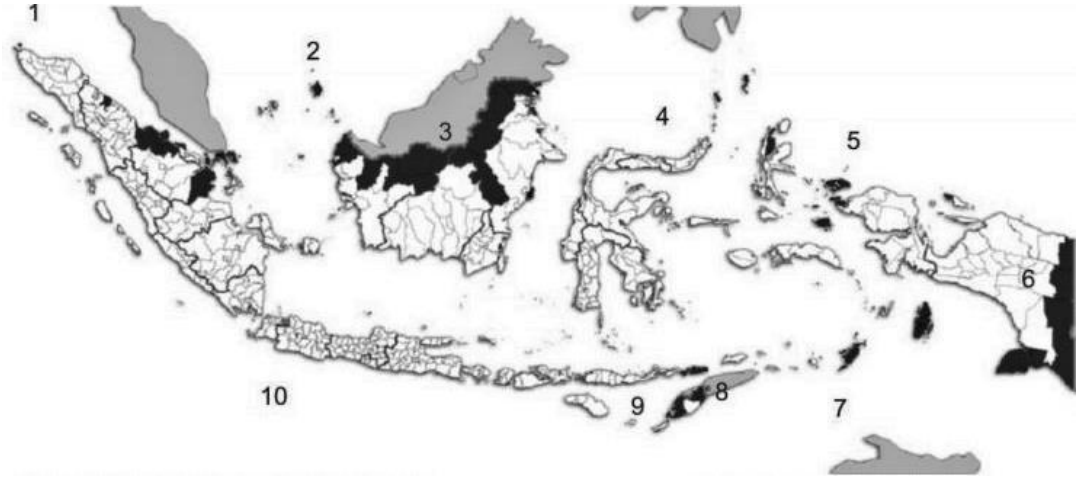
This paper shows how in Indonesia the statistical office measured the error of signaling data and then reduced the error significantly through first creating an estimation formula and then applying appropriate algorithms to reduce the noise in signaling data.

The methods introduced in this paper are now part of regular tourism statistics production in Indonesia released every month. The authors believe the methods can be replicated and adjusted to other countries.

Keywords: Big Data, Mobile Positioning Data, Signaling data, CDR, Error Measurement, Tourism

Motivation for using mobile positioning data for tourism statistics

Indonesia is the largest archipelago in the world with 18,110 islands, an area of 3.1 million km² and territorial waters spanning 5.8 million km². This broad geography means Indonesia borders many countries. Indonesia has a land border with Malaysia, Timor Leste, and Papua New Guinea that stretches along 3092.8 km. At the same time, the sea 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 islands, starting from Miangas Island in the north to Dana Island in the south (Figure 1).



1. Aceh/North Sumatra - Thailand/India/Malaysia
2. Riau/Riau Islands - Malaysia/Vietnam/Singapore
3. East & West Kalimantan - Malaysia
4. Kalimantan/Sulawesi - Malaysia/Philippines
5. North Maluku/West Papua - Palau
6. Papua - Papua New Guinea
7. Papua/Maluku - Australia/Timor Leste
8. East Nusa Tenggara - Timor Leste
9. East Nusa Tenggara - Timor Leste/Australia
10. Outermost Islands - High seas

Figure 1. Border areas in Indonesia (dark shaded)

Prior to mobile positioning data, data on inbound tourists was mainly obtained from the Immigration Office from administrative data (passport swipe) and monthly reports of immigration, which recorded the traffic of all people entering Indonesian territory through official gates. Data is available in good quality, especially on foreign tourists entering through the main gates (19 entry gates), such as Soekarno-Hatta Airport, Ngurah Rai Airport, Kualanamu Airport, Airport and Port in Batam and Juanda Airport. However, because the Indonesian territory is vast with diverse border areas (both sea and land) and the Immigration checkpoints are limited, 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, only guarded by the Indonesian Army or even the head of the local village.

In order to capture inbound tourists that enter through gates without immigration checkpoints, regular cross-border surveys are conducted. However, the surveys in the remote

cross border areas are expensive and the transportation costs to survey locations are high, as they are remote and difficult to reach. Added to this is the fact that the survey is only conducted over a month in selected locations, to estimate the numbers for one year for the entire border.

In summary, there is a coverage problem in the tourism data in Indonesia and the data that exists is not suitable for timely analysis. Benchmarking Indonesia's statistics against other countries, in Indonesia the neighboring countries only contribute about 7 percent of tourism arrivals in 2015, while the figure is between 30 to 60 percent for other nearby countries that compete with Indonesia for tourism arrivals.

Mobile positioning data from the largest mobile network operator has good coverage on the 300,000 km^2 of border areas across over 3000 km of land border and a vast sea border. BPS-Statistics Indonesia, in collaboration with the Ministry of Tourism, started to use mobile positioning data since October 2016 for measuring cross-border international visitor arrivals. It has been implemented to measure international visitor arrival in several regions in Indonesia which are bordering with Malaysia, Singapore, Papua New Guinea, and Timor Leste. There are 20 regencies in cross-border areas in 7 provinces where there is no immigration checkpoint (Sanggau, Natuna, Malaka, Bengkayang, Kapuas Hulu, Kepulauan Anambas, Pelalawan, Kupang, Rokan Hilir, Indragiri Hilir, Sintang, Keerom, Kepulauan Talaud, Kepulauan Sangihe, Lingga, Malinau, Boven Digoel, Pegunungan Bintang, and Mahakam Ulu). In order to measure the international visitor arrivals, visitors should still use the mobile phone number from their country of origin when entering Indonesia and be connected to the mobile operator network while in the zone of observation.

This paper shows Indonesia's experience in using signaling mobile positioning data for official tourism statistics released every month. It also showed that signaling data have more valuable data. It capture more inbound trips than CDR, however it also has drawbacks that is it has more noises (statistical and non-statistical). Furthermore, in this paper we measured the error of signaling data and then reduced the error significantly through first creating an estimation formula and then applying appropriate algorithms to reduce the noise in signaling data. The methods introduced in this paper are now part of regular tourism statistics production in Indonesia released every month. We believe the methods can be replicated and adjusted to other countries that face same problems as Indonesia.

Signalling data captures more inbound trips than CDR

Usually, mobile positioning data refers to call detail records (CDR) as that is most easily accessible and fits the purpose of official tourism statistics (Tiru 2014). In Indonesia, signaling data is used, since the operator Telkomsel is collecting, storing and commercializing signaling data. We found that signalling detects on average 3.47 times more roamers at the border areas compared to CDRs, but that ratio differs for each kabupaten (municipality), as seen on Figure 2. E.g. in Malaka, on the border of Timor Leste, the multiplier is up to 158.75 times.

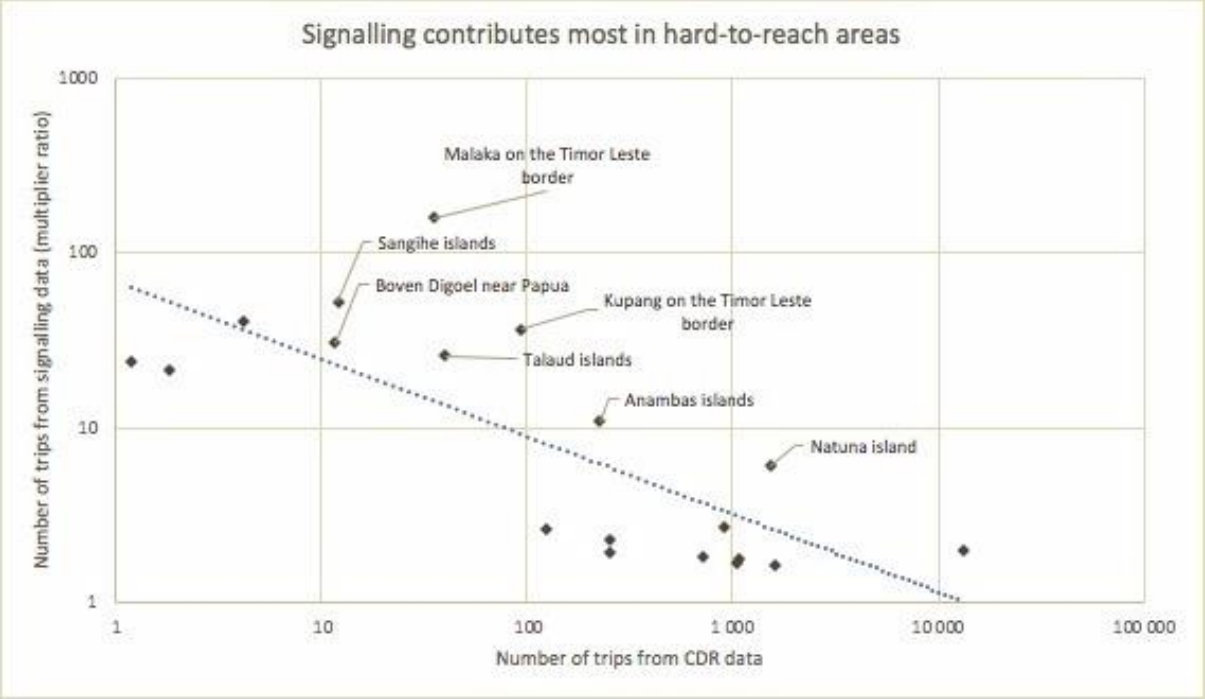


Figure 2. Signalling data is multiple times that of CDR because it catches more roamers entering areas in Indonesia, and the effect is higher in hard-to-reach areas like border islands and Papuan wilderness, rather than the relatively heavily populated borders to Malaysia.

We found that signaling data is indeed more valuable and captures roamers (tourists) that are not using their phone to call or send SMS while roaming to avoid expensive charges. So, signaling data captures more inbound trips than CDR data and overcomes some under-coverage issues of CDRs, especially in areas that are harder to reach. Even with this advantage, questions are still raised as to the accuracy of signaling data.

Signaling data adds non-tourism noise

After studying the aggregated roaming data received from the operator during the first year of cooperation, we then went deeper to analyse the raw data used to create those aggregates. We have found that signaling data adds noise. To measure the effect of noise we compared the signaling MPD data with immigration data in places where immigration has good coverage, and then measured the error.

The method for quantifying people entering Indonesia is quite similar whether you use MPD or immigration data – only the first port of entry is taken into consideration to avoid double-counting. However, the data sources are not equal for counting tourism, as in most cases the coverage differs.

The main differences in immigration data and mobile phone data are (Table 1):

- Immigration data only covers those that cross the border immigration gate, while MPD covers all roaming mobile subscribers.
- Immigration data refers residence from passport and citizenship, while MPD refers residence from foreign SIM card ownership

Immigration data calculates the tourists as their trips. Each entry to country is counted as new tourist. Mobile Positioning data, however, needs a definition for a trip and handling uncertainty. For example, it could be that a particular mobile phone data shows that it disappears from the network for two days and we would not know if it switched off the phone for two days or this is now a new entry into the country.

Table 1. Both data sources for tourism, immigration and MPD, have extreme bias that needs to be countered

Over-estimation in immigration data	Under-estimation in immigration data		Under-estimation in mobile phone data	Over-estimation in mobile phone data
BIAS IN IMMIGRATION DATA		CORRECT MEASUREMENT	BIAS IN MOBILE PHONE DATA	
Foreign passport holders residing/working in the country	Nationals with residence abroad People that do not cross immigration		Non-roaming tourists Tourists that do not roam in the mobile operator in question	Roaming subscribers that are not tourists

In Indonesia, the best data on the quantity of visitor arrivals comes from immigration data in places that have immigration check-points. It is granted that some border areas in Indonesia do not have adequate coverage of immigration check-points on the border. However, there are areas, where the coverage is near perfect – on border islands Batam and Bintan that lie near Singapore.

Filtering out most of the noise helps reduce error

Comparing immigration data and signaling data in Bintan suggests that using the original signaling data the difference between mobile positioning data (signaling) and immigration data were quite large, in this case measured by root mean square error. Figure 1 shows the root mean square error of mobile positioning data for different approaches of MPD data. We found that the difference between those two data sources is due to the noise (statistical and non-statistical) created by special types of roamers such as those flying over the country (fast fliers), crossing the country's seas (seamen) and accidentally roaming across the border. As we removed much of the noise – the fast fliers, seamen and accidental roamers – the root mean square error was also reduced.

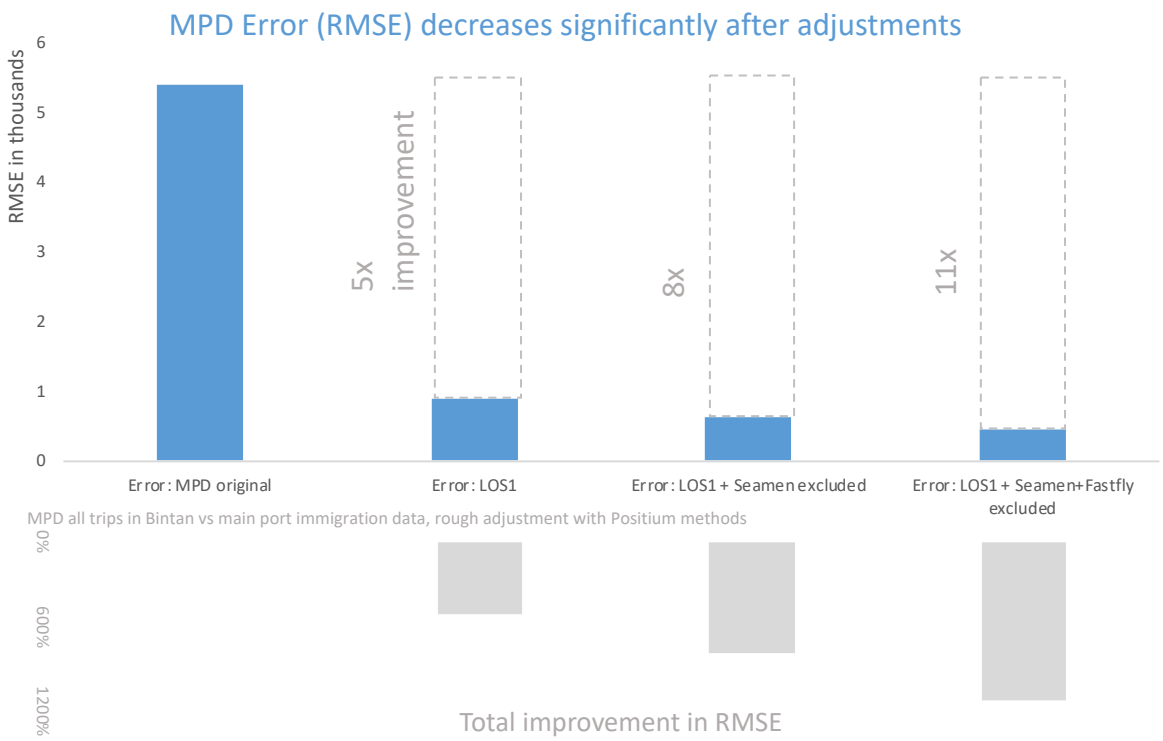


Figure 1. Error is reduced after each step of filtering, the reduction in the root mean square error reaches 1100%

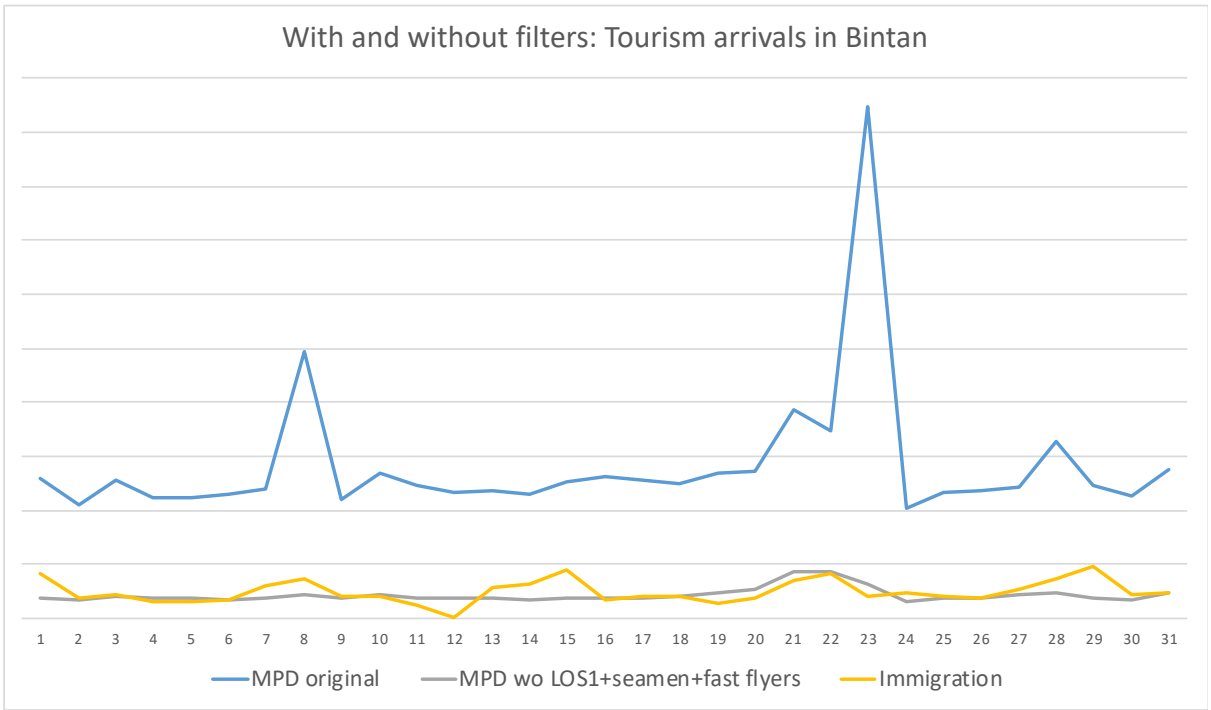


Figure 2. Example of filtering in Bintan island for all tourism arrivals, showing the convergence of MPD and immigration data after filtering.

Cross Border Mobile Usage Survey allows to check the ground truth and counter other coverage issues

After removing some of the noise, some other issues remained that could affect the results drastically. For example, there is 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 pass by without a mobile phone or use a local SIM card also cannot be known through looking at the MPD alone. To overcome these challenges, in 2017 BPS in cooperation with the Ministry of Tourism conducted a Cross Border Mobile Usage Survey in the border area of Indonesia to improve the calculation of foreign tourist visits using MPD.

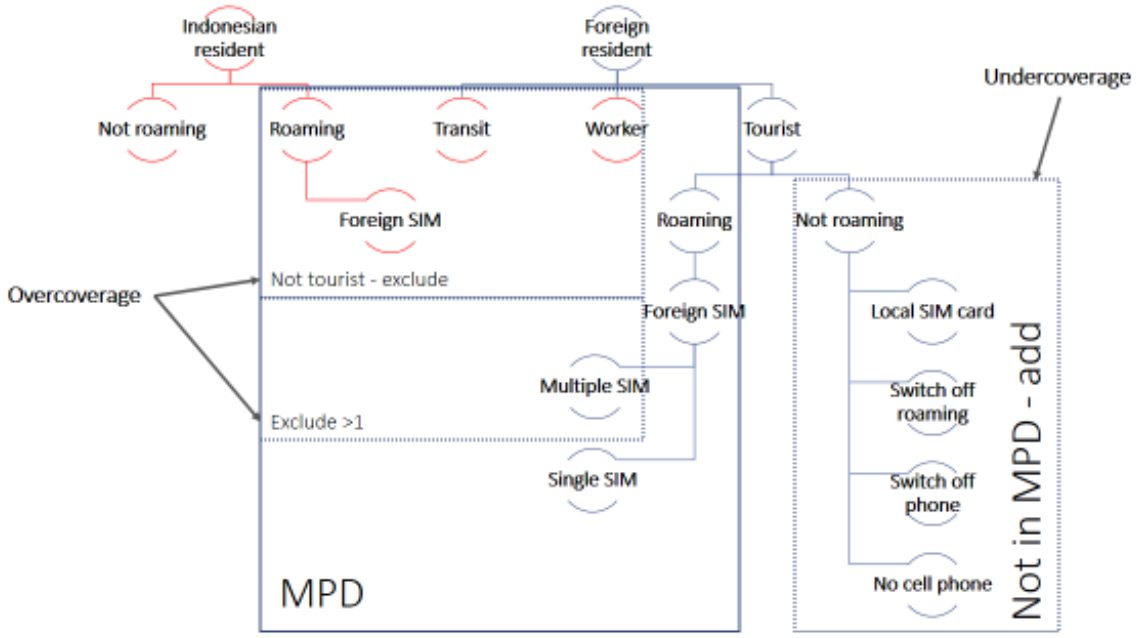


Figure 5. Causes of over and under coverage in MPD and, as a result, questions to be answered with the survey

The survey was conducted in two phases, in June-July and in October-November. The survey is conducted in 23 regency (‘kabupaten’) that have a border to neighboring countries. 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.

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

$$AT = \frac{MPD}{X_{roam}} \times P \times (1 - P_w) \times \frac{1}{1 - P_{NR}} - WCI$$

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	=	The ratio of foreign residents with foreign SIM cards to total number of foreign SIM cards;
P_w	=	The ratio of foreign residents with foreign SIM card that come for work to total number of foreign residents with foreign SIM cards;
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.
WCI	=	Number of tourists entering through Immigration Post

Table 2. Example formula ratios in some of the different areas on the border based on mobile usage survey resultshows the ratio/proportion for the formula which is obtained from the mobile usage survey.

Table 2. Example formula ratios in some of the different areas on the border based on mobile usage survey result

REGENCY	XROAM	P	1-PW	1/(1-PNR)	RESULT
	Proportion of active roaming sim cards	Proportion of foreign residents	Proportion who do not come for the purpose of work	Proportion of non-roamers	Final coefficient as a product of all ratios
BELU	0,83	0,88	0,96	1,12	0,78
BENGKALIS	0,98	1,00	1,00	1,10	1,08
BENGKAYANG	0,91	0,54	0,91	1,08	0,48
BOVEN DIGOEL	0,92	1,00	1,00	2,80	2,58
…	…	…	…	…	…
SINTANG	0,85	0,60	1,00	1,24	0,63
TALAUD	1,00	1,00	1,00	2,00	2,00
TIMOR TENGAH UTARA	0,75	0,81	0,97	1,19	0,71
MERAUKE	0,92	1,00	1,00	2,80	2,58
JAYAPURA	1,00	1,00	1,00	2,01	2,01

The ratios do not display a seasonal character when results of the survey from July and October were compared. Currently flat ratios are used. However, since the survey period was

limited, it is necessary to repeat the survey and continue observe and, if necessary, correct for seasonality with proper algorithm.

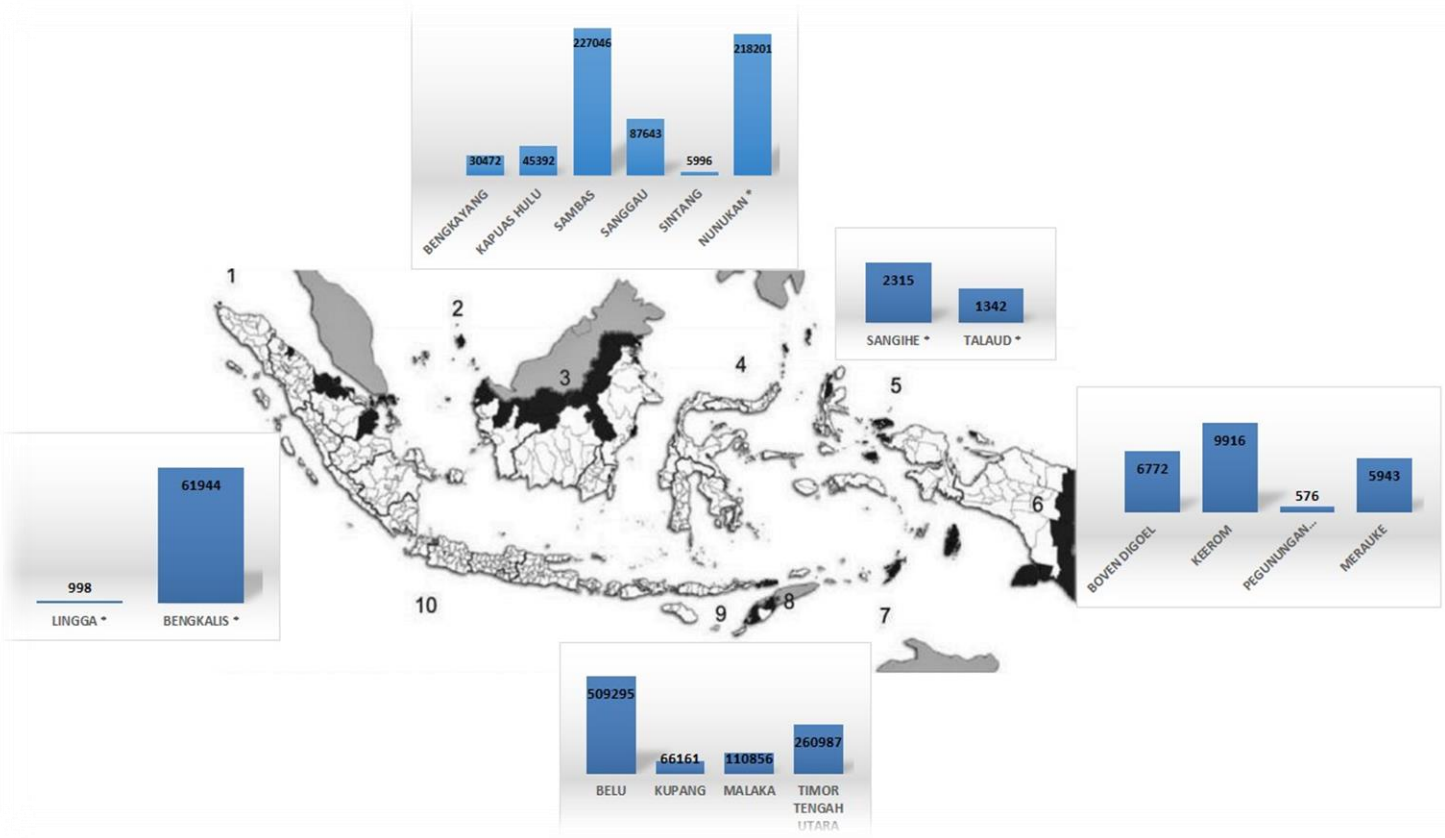


Figure 3. Number of Additional Tourism in Border areas Jan-July 2018

Figure 6 shows the total number of additional tourism in border areas from January to July 2018. Noise filtering and the extrapolation formula are already applied to arrive at these figures and immigration figures are deducted. The biggest additional number is 509 295 tourists in Belu regency, Nusa Tenggara Timur and the smallest is 576 in Pegunungan Bintan, Papua. Timor Leste became one of the biggest contributors to inbound tourism as that is where the previous data captured the least. The number of additional tourism in border areas proves that MPD increases the coverage of inbound tourism arrival statistics in Indonesia.

Lessons Learned

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 – there is no data source that is superior. All data sources could complete each other.

Signaling data is the type of mobile positioning data currently used in the Indonesian case, on average the signaling data detects 3.47 times more roamers at the border areas compared to CDR data. Signaling data overcomes some undercoverage issues of CDRs. It also has the advantage of wide coverage, but the data quality of signaling data is questionable at face value as to the accuracy of what is measured. Certain types of roamers contribute to noise in the data, such as those flying over the country, crossing the country's seas and accidentally roaming across the border. However, if these problems are known, they can be overcame. The simple way is to use correction factors derived from a survey of the behavior of roamers (mobile users). After MPD is used, the proportion of cross-border tourism for Indonesia now exceeds 30 percent, which aligns to international benchmarks.

Although that approach is effective in the short term, and produces good estimates, in the long-term the noise should to be eliminated at the level of raw data – algorithmically. In 2018, the Ministry of Tourism, BPS and Positium will collaboratively create an automated system to process the signaling data at Telkomsel, eliminate noise at the raw data level, and produce tourism statistics with increased timeliness and improved quality.

The next step is to increase the use of mobile positioning data across Indonesia to complement immigration data. MPD also allows new ways of analysing tourism, by looking at the destinations and events that tourists visit. In 2018, BPS, Positium and Telkomsel completed the analysis of international and domestic tourism visits to Asian Games, the second largest multi-sport event in the world, in three weeks.

Furthermore, we are developing MPD as a method for calculating domestic or sub national tourism, while conducting traditional household survey (HS) at the same time for benchmarking purpose, since 2018 is the first year of implementation. The households are surveyed about tourism travel quarterly while MPD is calculated monthly. The data quality of HS depends very much on respondent's memory, while data quality from MPD depends on methodology that can be improved over time and data recalculated as methods improve.

Household survey and MPD both cover 514 districts and cities. However, the HS survey was not designed to estimate regency/city level data, it was to estimate higher administrative unit such as province and national level. Both surveys are still in progress with preliminary results expected by the end of 2018.

Conclusion

The use of mobile positioning data changes the process of design, build, data collection, data processing, and dissemination. Data collection is more efficient in terms of cost and time. Near real-time dissemination can be achieved by using mobile positioning data. Big data is a part of the data revolution and needs to developed as a validated source for statistics.

Signaling data solves undercoverage problems but has its own overcoverage issues. First, mobile usage surveys can be used to reach good estimates of additional tourism. In the meantime, proper algorithms can be developed to make the process of analyzing data automatized and more reliable. Adjustments and applying proper algorithms could reduce the errors and improving the quality of signaling data, making it usable for official tourism statistics.

Next, BPS-Statistics Indonesia are developing MPD as a method for calculating domestic or sub national tourism, while conducting traditional household survey (HS) in the same time for benchmarking purpose. MPD has the potential to increase the level of disaggregation, timeliness, and usefulness for tourism analysis.

The approach described in this article can be replicated for the use of MPD for tourism statistics in other countries.

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.

De Jonge, Edwin, Merijn van Pelt en Marko Roos. (2012). Time Patterns, Geospatial Clustering and Mobility Statistics Based on Mobile Phone Network Data. The Hague: Statistics Netherlands.

Garcia, F. Cortina, et.al. (2016). Quality Implications of the Use of Big Data in Tourism Statistics: Three Exploratory Examples, European Conference on Quality in Official Statistics, Madrid.

Greene, William. H (2003). Econometric Analysis, Prentice Hall, New Jersey.

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 Ahas, 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.