RIPE Atlas User Measurement Patterns

by Theodoros Diamantidis and Pavlos Sermpezis,
Data & Web Science Lab (https://datalab.csd.auth.gr/),
Aristotle University of Thessaloniki

1. Introduction

In RIPE Atlas, when a user wants to conduct a measurement, aka “User Defined Measurement” (UDM), they have two options: they can either select specific probes to use in their measurements, or they have the option of not specifying them and letting the RIPE Atlas platform pick the probes for them. RIPE Atlas makes the selection of probes for a measurement based on an algorithm which prioritizes probes with less load over more loaded probes. As a result, the probes that are selected for a measurement are not randomly sampled, which leads to higher bias values.

The set of networks hosting RIPE Atlas probes is not representative of the entire Internet (e.g., biased towards European networks) as shown in [REF-ai4netmon-paper]. The RIPE Atlas allocation algorithm that does not randomly assign probes may introduce further bias in a measurement. In fact, preliminary results [REF] show that by simply choosing a random sample of probes from RIPE Atlas, instead of relying on RIPE Atlas’ selection algorithm, we can have a probe sample that results in measurements with almost half the bias regardless of how many probes we want to select.

Motivated by these studies, we aim to explore the bias in RIPE Atlas measurements, from a complementary perspective: how the users actually select probes. How much bias is there in user defined measurements simply from probe selection? How can the bias sources of such measurements be further explored? Finally, given a set of probes to be used for a user defined measurement, what recommendations can we make to the user so that their measurements have less bias?

In this report we answer these questions and provide an analysis of the User Measurement Patterns for measurements made with RIPE Atlas.

2. Methodology

2.1 Bias Calculation

Before continuing with the description of the different datasets we used and how they were collected, we should briefly mention how bias is calculated for a measurement. As we have already briefly mentioned, each user defined measurement is made using a set of RIPE Atlas probes. Each of these probes belongs to a certain Autonomous System (AS). Each AS contains multiple probes, and it can be the case that two, or more, probes from the same measurement can belong to the same AS.

As a first step in our bias calculation, we first figure out what ASes take part in each measurement. This is important, as the framework we use for the calculation and analysis of bias is the AI4NetMon project, in which data for many different characteristics of IMPs were collected and aggregated at the AS-level [REF Pavlos-Sof paper]. Namely, data for a set of 23 characteristics, henceforth referred to as "bias dimensions" or simply "dimensions", was collected for each AS.

Having now the list of ASNs (Autonomous System Numbers) for each measurement, we compare the distribution of each bias dimension of the sub-sample of ASes in a measurement with the corresponding dimension of the entire population of ASes. If there is a difference between these two distributions, then we say that measurement is biased in that specific bias dimension. Furthermore, bias is quantified via the Kullback–Leibler divergence between each measurement's ASN list, and the list of all the ASNs for each bias dimension. For more details about the definition of bias as well as its calculation please refer to original paper [REF Pavlos-Sof paper] or this blogpost [REF TO FIRST REPORT].

Let us now proceed with a description of the datasets we used for our analysis.

2.2 Datasets

As we briefly mentioned in the introduction, there is a significant difference in measurements' bias depending on the selection method of their probes. We also want to explore what is the bias for user defined measurements, and how it compares with measurements that contain probes selected via random sampling and the RIPE Atlas selection algorithm.

Therefore, three different sets of RIPE Atlas measurements have been used:

1) **Atlas (automated algorithm):** This sample consists of measurements that contain probes which were selected purely via RIPE Atlas' probe selection algorithm. This algorithm takes into account probe availability and network traffic, so it is highly possible that these measurements will be biased. This dataset was created from measurements that we performed ourselves.
2) **Atlas (random sample):** This sample consists of measurements that contain probes which were selected via random sampling. The data for this dataset were collected using an endpoint of the AI4NetMon project. This endpoint, for a given number of probes (which is a parameter of the endpoint's URI), say *n*, creates 10 sets of randomly selected RIPE Atlas probes each with *n* probes, finds their corresponding ASNs, and based on that, it calculates the bias for all bias dimensions and then aggregates them to return a response like the one in the figure below:

https://ai4netmon.csd.auth.gr/api/bias/randomAtlas/50

```json
{
    "Atlas": {
        "50": {
            "RIR region": 0.08948025596673005,
            "Location (country)": 0.3879815184294453,
            "Location (continent)": 0.08747493465738426,
            "Customer cone (#ASNs)": 0.06971766878743633,
            "Customer cone (#prefixes)": 0.12235027050537503,
            "Customer cone (#addresses)": 0.28166537040636747,
            "AS hegemony": 0.12830842877636464,
            "Country influence (CTI origin)": 0.11671082981439955,
            "Country influence (CTI top)": 0.40388835227531417,
            "#neighbors (total)": 0.1341977200564817З,
            "#neighbors (peers)": 0.09294871937428854,
            "#neighbors (customers)": 0.05807974375549955,
            "#neighbors (providers)": 0.07987664223978655,
            "#IXPs (PeeringDB)": 0.06001625521618494,
            "#facilities (PeeringDB)": 0.05585535211061496б,
            "Peering policy (PeeringDB)": 0.0201961264973577,
            "ASDB C1L1": 0.17202576209722625,
            "ASDB C1L2": 0.271764226598211,
            "Network type (PeeringDB)": 0.07694417935816783,
            "Traffic ratio (PeeringDB)": 0.043702935461508816,
            "Traffic volume (PeeringDB)": 0.1654923006192002,
            "Scope (PeeringDB)": 0.09213163111003579,
            "Personal ASN": 0.005695483523583214
        }
    }
}
```


Figure 2.2-1: Example of AI4NetMon's randomAtlas endpoint usage. At the top we see the URI requesting measurements with 50 randomly sampled probes, and at the bottom we see the corresponding response containing the average bias values for the 10 different measurement samples, each of which using 50 randomly sampled probes.

3) **User measurements dataset:** This sample consists of our sampled measurements, i.e. user measurements randomly sampled from the RIPE Atlas API. Since the measurements in this sample were picked at random from the RIPE Atlas API, there was no guarantee that there would be enough measurements with a certain number of probes. For example, it could be the case that we had 60 measurements with 50 probes, but it could also be the case that we only had 1 measurement with 50 probes. In order to make sure that we have a better sample size for measurements with a given number of probes, we decided to keep measurements with the number of probes within a certain range of the corresponding number of probes value. Back to the example of measurements with 50 probes, instead of only keeping measurements in our sample with exactly 50 probes, we kept measurements with 50 +/- 50*20% probes, i.e. number of probes in the range [40, 60], and aggregated their bias values across all bias dimensions. For the rest of this report, when we refer to a subset of measurements with 50 probes from the User measurements dataset, we will actually mean that we kept measurements with number of probes between 40 and 60.

Our goal is to study the level of biases in these three types of measurements as well as how the number of probes affect it. To this end, for a given set of measurements with the same number of probes, we first calculate their bias across all the different bias dimensions, and then, for each measurement, we aggregate the bias values across all bias dimensions, by taking their mean. In the end we have a dataset that looks like this:

num_probes	atlas_rand_avg_bias_10_probes	atlas_rand_avg_bias_50_probes
RIR region	0.23	0.09
Location (country)	0.65	0.39
Location (continent)	0.26	0.09
Customer cone (#ASNs)	0.09	0.07
Customer cone (#prefixes)	0.23	0.12
Customer cone (#addresses)	0.41	0.28
AS hegemony	0.22	0.13
Country influence (CTI origin)	0.33	0.12
Country influence (CTI top)	0.63	0.40
#neighbors (total)	0.28	0.13
#neighbors (peers)	0.13	0.09
#neighbors (customers)	0.09	0.06
#neighbors (providers)	0.18	0.08
#IXPs (PeeringDB)	0.13	0.06
#facilities (PeeringDB)	0.10	0.06
Peering policy (PeeringDB)	0.04	0.02
ASDB C1L1	0.33	0.17
ASDB C1L2	0.39	0.27
Network type (PeeringDB)	0.28	0.08
Traffic ratio (PeeringDB)	0.18	0.04
Traffic volume (PeeringDB)	0.49	0.17
Scope (PeeringDB)	0.33	0.09
Personal ASN	0.01	0.01

Figure 2.2-2: Average bias values for measurements with 10 and 50 probes from the Atlas (random sample) sample.

It should be noted here, that for our analysis we only considered IPv4 measurements except for the Atlas (random sample) sample which took into account both IPv4 and IPv6.

3. Findings

As already mentioned, the way in which the probes are selected for a measurement, can significantly affect the bias of that measurement. To showcase that, we plot the average bias across all bias dimensions for measurements with the same number of probes, against the number of probes for our three different sets:

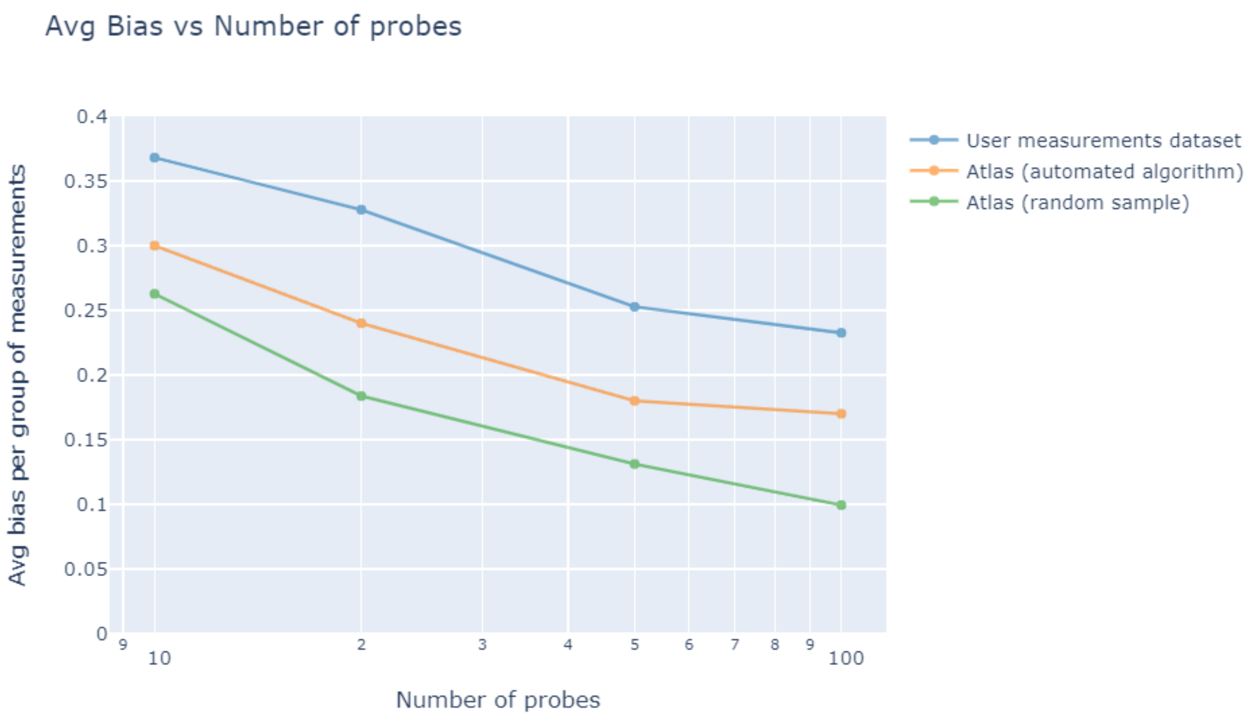

Figure 3.1-1: Average bias per group of measurements with the same number of probes vs the number of probes for our three different samples.

The first thing we see is that the average bias for each sample drops as the number of probes is increased, but for higher numbers of probes, it seems to form a plateau. This suggests that the decrease in bias we can get from simply increasing the number of probes we consider for our measurement is limited; in fact it is bounded by the platform bias itself.

We can understand that better, by taking that argument to the extreme: if we consider *all* RIPE Atlas probes for a measurement, would that measurement have 0 bias? The answer is no, as the distribution of all RIPE Atlas ASes (which probes belong to), has some bias itself as we can see in the figure below:

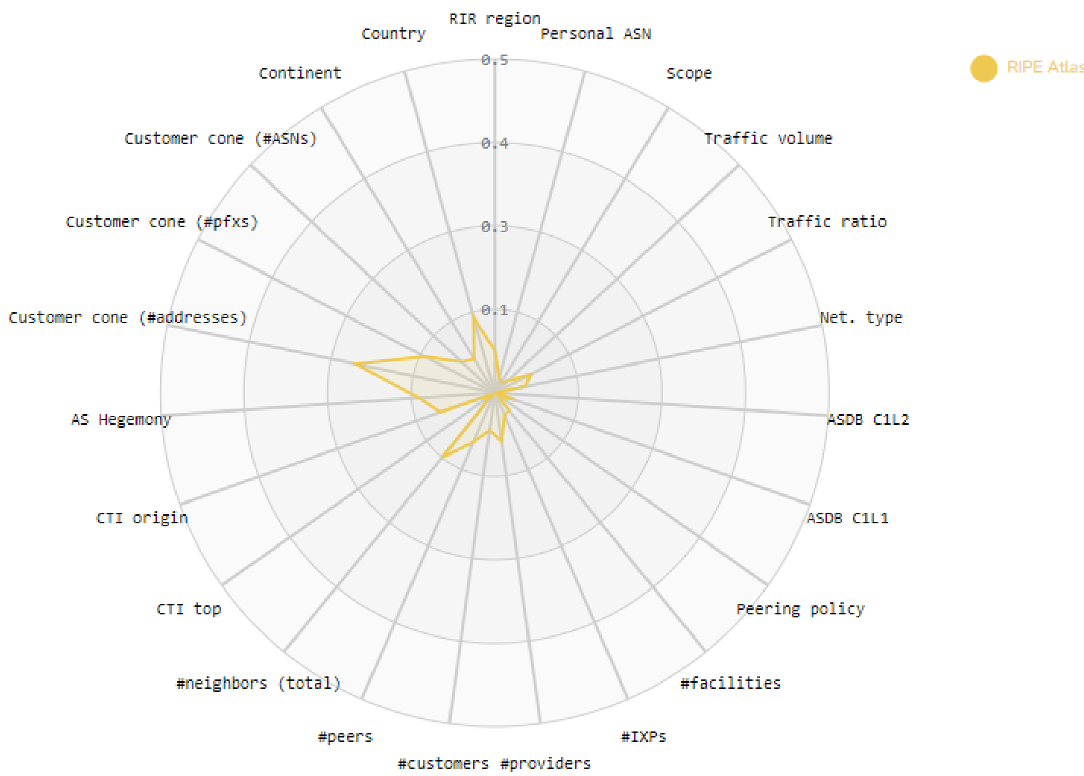

Figure 3.1-2: Bias distribution of all RIPE Atlas ASes.

Therefore, while we can get a significant decrease in bias as we consider more probes, that bias reduction is bounded by the bias of the platform we use.

A more interesting conclusion from Figure 3.1-1 though, is that we see a clear difference between the bias values of the three different sets: the user measurement dataset has the highest bias values, followed by the measurements containing probes selected via the Atlas probe selection algorithm, followed by the measurements containing probes selected via random sampling. Furthermore, we see that this difference in bias values is consistent across different numbers of probes considered.

This tells us that for a given number of probes for our measurement, the best method to select probes is via random sampling, as it would lead to the minimum possible bias values for our measurement.

In the next section, we dive deeper into the bias differences of our three samples: first, we explore the bias distributions for different numbers of probes and finally we study the main bias causes for our user measurement dataset.

4. Comparison of Bias Distributions

In the following plots we see the bias distribution for measurements with 10, 50 and 100 probes from our three samples.

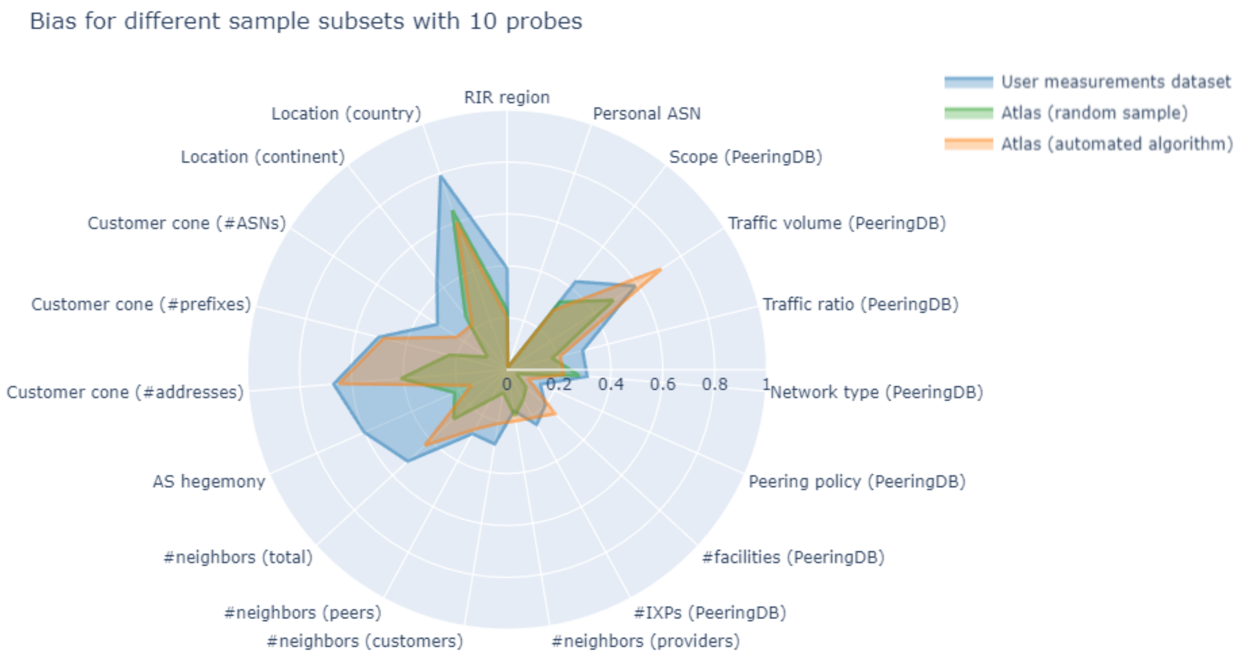

Figure 4-1: Bias distribution for measurements with 10 probes from our three different samples.

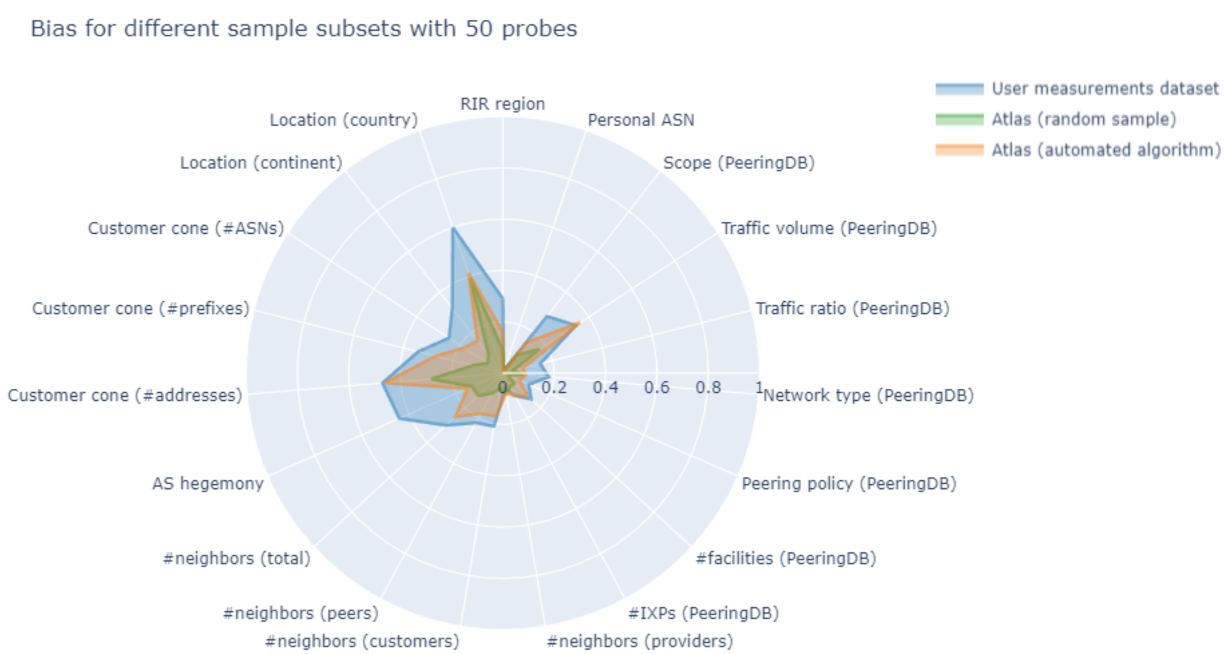

Figure 4-2: Bias distribution for measurements with 50 probes from our three different samples.

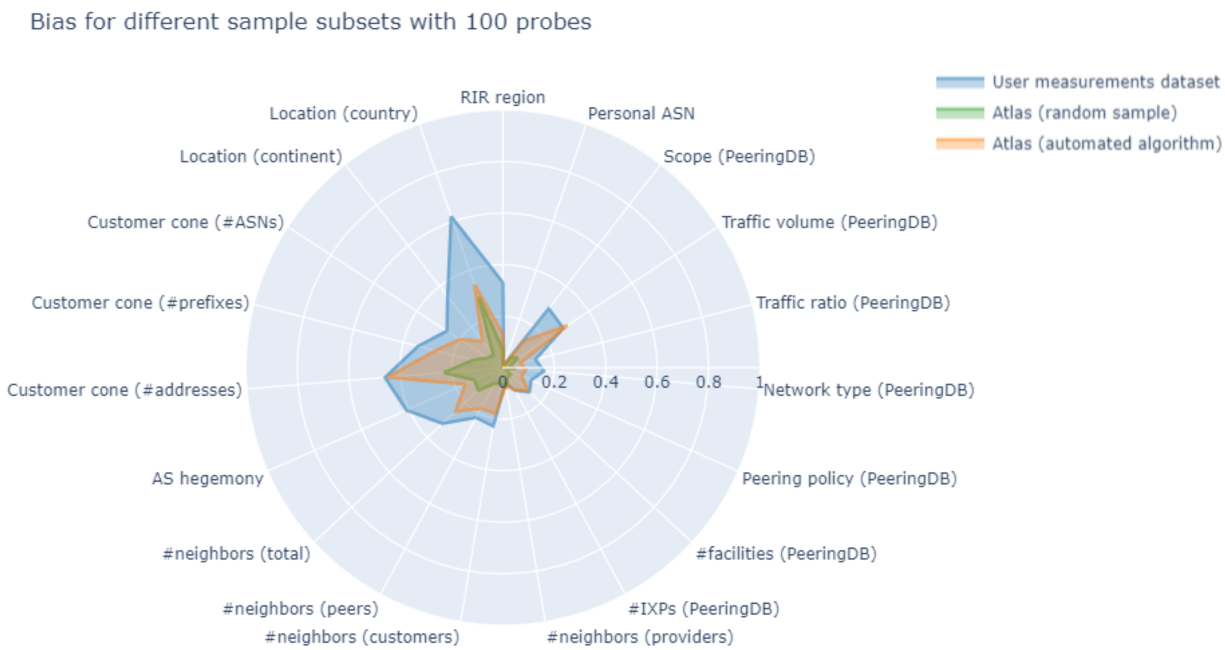

Figure 4-3: Bias distribution for measurements with 100 probes from our three different samples.

4.1 More probes lead to uniformly less biased measurements

The first thing we see is that although the bias values decrease across almost all the bias dimensions between the distributions in the first two figures (measurements with 10 and 50 probes respectively), the same is not true for the distributions between the second and last figure (measurements with 50 and 100 probes). While this may be expected based on the results of the previous section, i.e. after some number of probes there doesn't seem to be any significant bias reduction, what's interesting is that the *decrease happens uniformly across the different bias dimensions*.

In other words, it seems that when we increase the number of probes, the overall reduction in the average bias of our measurements does not stem from a single dimension or a small subset of dimensions, but it seems like all of them contribute to this reduction uniformly. This means that, while more probes might decrease the bias of our measurements, they will not affect their bias distribution. This is the case for all our different samples (User measurements dataset, Atlas (random sample) and Atlas (automated algorithm)).

This can be seen more clearly through Figures 4.1-1, 4.1-2 and 4.1-3 where we can see the reduction in bias values as we go from less to more probes for our three different samples. In these plots we see that there is a significant bias reduction as we go from 10 to 50 probes for almost all dimensions and for all three different samples. This decrease is not as large when we go from 50 to 100 probes.

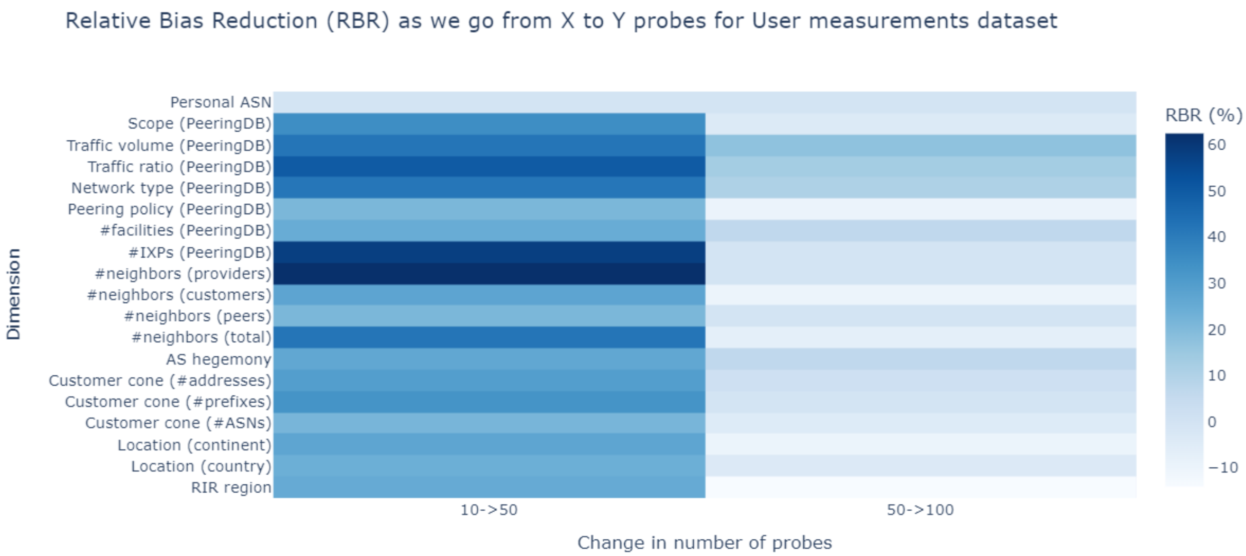

Figure 4.1-1: Relative bias reduction as we go from 10 to 50 and from 50 to 100 probes, for each dimension for the User measurements dataset.

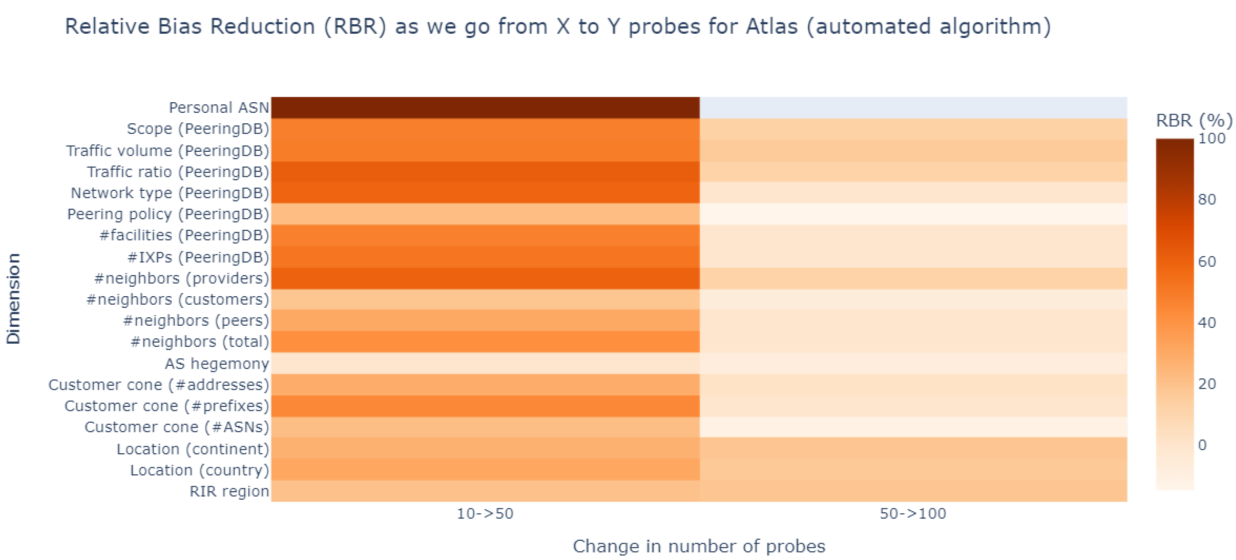

Figure 4.1-2: Relative bias reduction as we go from 10 to 50 and from 50 to 100 probes, for each dimension for the Atlas (automated algorithm).

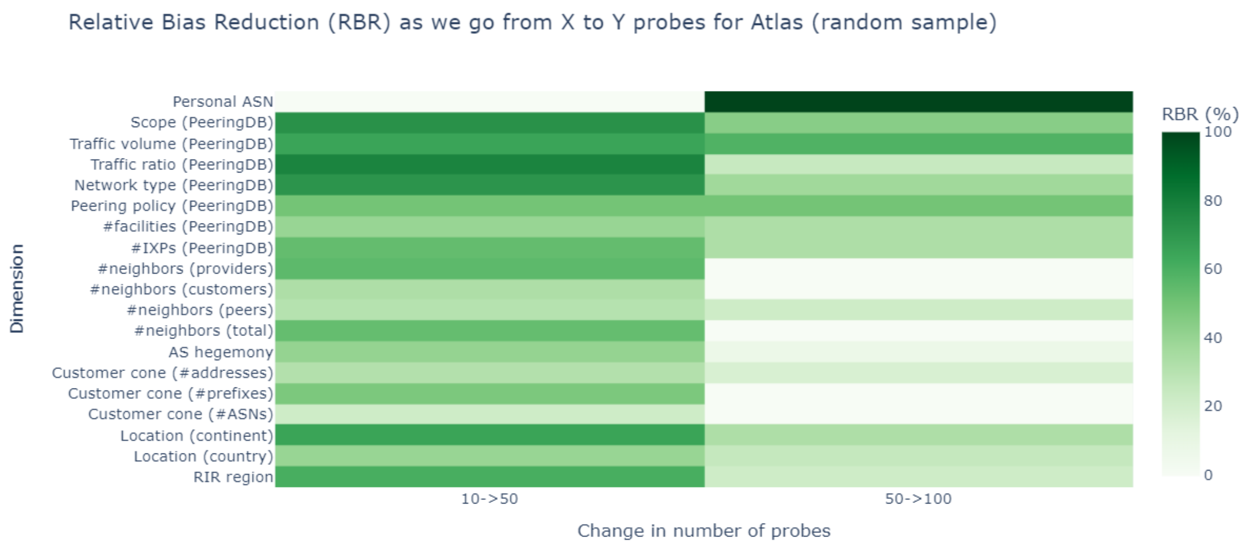

Figure 4.1-3: Relative bias reduction as we go from 10 to 50 and from 50 to 100 probes, for each dimension for the Atlas (random sample).

4.2 Users' selection leads to more bias in specific dimensions

We have seen that bias reduces relatively uniformly across bias dimensions as we increase the number of probes in measurements. What can this decrease in the average bias tell us about RIPE Atlas users?

To answer that question, let us compare the relative bias reductions between the Atlas (random sample) dataset and the User measurements dataset. To do that, we simply divide the relative bias reduction of the User measurement dataset with that of the Atlas (random sample), in other words, we simply divide the values plotted in Figure 4.1-1 with those in Figure 4.1-3. The results are presented in Figure 4.2-1:

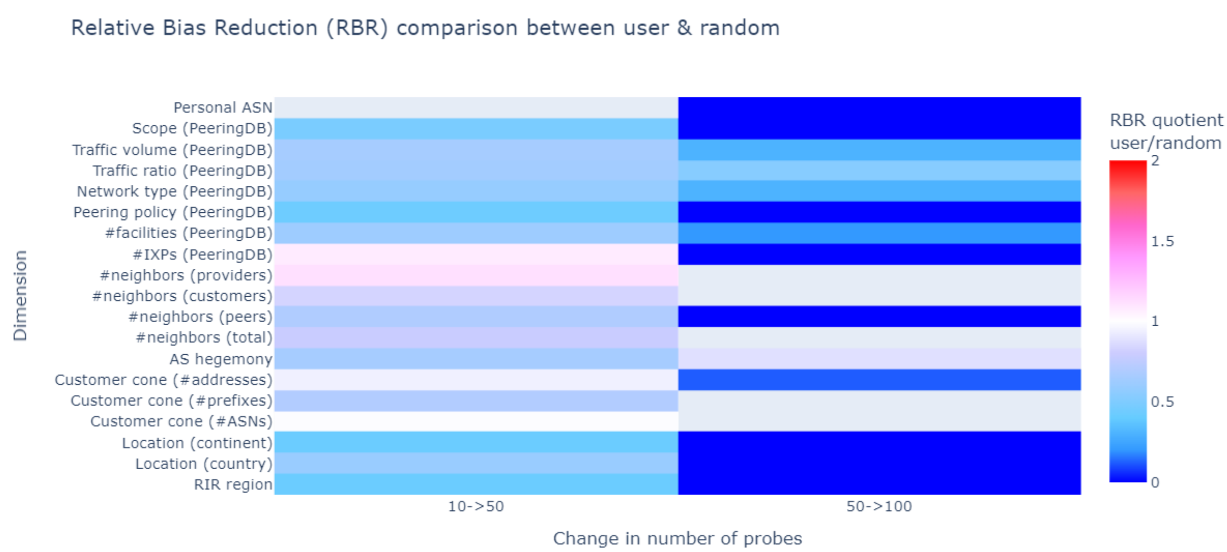

Figure 4.2-1: Comparison of relative bias reduction as more probes are considered, between the User measurements dataset and the Atlas (random sample) dataset.

In Figure 4.2-1, we see that the relative bias reduction as we go from 10 to 50 probes for the UDMs is smaller than the corresponding reduction for the Atlas (random sample) dataset. This means that there is some other factor affecting the reduction of bias for UDMs, other than the number of probes, and this factor is the way in which users select their probes. So how do users select probes?

When examining the bias distributions of the Atlas measurement samples (automated algorithm and random sample, i.e. orange and green in Figures 4-1, 4-2 and 4-3), we notice that their shapes on the radar plots remain quite similar, regardless of the number of probes used for the measurements. The sample created through random sampling (green) consistently shows lower bias than the one generated using the Atlas automated algorithm (orange), however, their shapes remain almost identical, regardless of probe count.

The situation is different when we compare the distribution shapes of the User measurements dataset (blue) with the Atlas samples. In this case, the distribution shapes appear to differ in a few dimensions, specifically AS hegemony, Scope (PeeringDB), and Location (continent). In other words, users seem to select probes in a way that creates more bias in the aforementioned dimensions.

4.3 Summary

Up to this point, we have conducted a comparative analysis between user-defined measurements (UDMs) and measurements employing probes selected through two distinct methods: Atlas' automated algorithm and random sampling. From our observations, we can draw several key conclusions:

1) UDMs exhibit notable differences when compared to measurements utilizing either random probe selection or Atlas' selection algorithm. Specifically, UDMs consistently display higher levels of bias across all dimensions of bias.
2) As we increase the number of probes in our measurements, the average bias experiences a significant decrease, reaching a point of diminishing returns. This pattern holds true regardless of the method employed for probe selection.
3) Measurements with probes selected through random sampling consistently exhibit lower levels of bias when compared to measurements where the Atlas algorithm was used for probe selection. This trend persists across all dimensions of bias.
4) Notably, UDMs distinguish themselves from measurements utilizing probes selected via random sampling primarily in three dimensions in which they consistently have higher bias values: AS hegemony, Scope (PeeringDB), and Location (continent).

In the next section, we will try to better characterize UDMs, in terms of their bias causes.

5. User Measurement Dataset Bias Causes

In this section, we first define what we mean with the term "bias causes" in detail, together with an example. Then we describe how we were able to get the bias causes data for our different samples, and finally, we present the results of our analysis of the bias causes for the User measurements datasets, which we will also refer to as User Defined Measurements (UDMs).

5.1 Bias causes definition

As we have described in Section 2.1, we define the measurement bias across some bias dimension as the difference in the distributions between the measurement's ASes and all ASes for that dimension.

For example, in the Location (continent) dimension, we may have x% of ASes in our measurement sample in Europe, while y% of RIPE Atlas ASes are in Europe. The difference between x and y in each category (which we also refer to as "bin") of the Location (continent) dimension is what causes bias. Other bins of the Location (continent) dimension are, for example, North America, Asia, etc The higher the difference between x and y (negative or positive) the higher the bias. This dimension-bin/category combination is what we refer to as a bias cause.

To quantify and compare the different bias causes, let us introduce some notation first. We use the subscript i to represent dimensions (e.g. Location (continent)), and the subscript j to represent bin/categories (e.g. Europe, N. America, Asia, etc). Let us also call x_{ij} the percentage of ASes that belong to bin j of dimension i, and let us use y_{ij} to represent the percentage RIPE Atlas probes (i.e. probes in UDMs) that belong to category j of dimension i. Then, we calculate:

$$diff_{ij} = y_{ij} - x_{ij}$$

For each dimension (i) and bin (j) combination. That way we know exactly by how much our sample of RIPE Atlas ASes (i.e. those in the UDMs) differ from the entire population of ASes, for each dimension and bin.

We follow that process for each UDM and end up with a set of all the bias causes for each measurement. We then aggregate the results (see Section 5.2 for details), sort them, and keep the top 5 positive and top 5 negative largest differences. These correspond to the top 10 dimension-bin combinations where UDMs are mostly over/underrepresented.

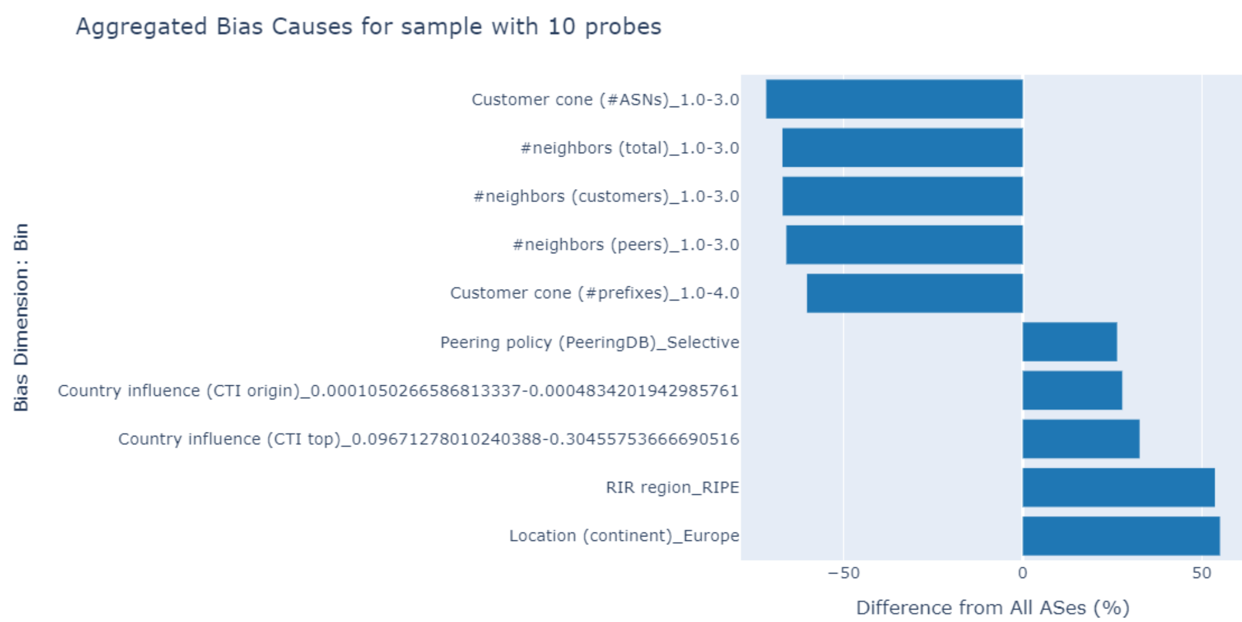

Figure 5.1-1: Top 5 positive and negative bias causes for the measurements with 10 probes from the User measurement dataset.

In Figure 5.1-1, we have plotted the top 5 positive and top 5 negative bias causes for the subset of UDMs containing 10 probes[1]. On the y-axis we have the bias causes for this subset of UDMs, while the x-axis corresponds to the percentage difference each bias cause has compared to all ASes (i.e. the $diff_{ij}$ value from above). Bias causes (y-axis) have the form "Dimension_Bin", giving us very low-level detail about where bias originates from for each sample.

So what does this all mean? We can see, for example, that for the sample of UDMs with 10 probes, the "Location (continent)_Europe" bias cause has a value of around 55%. This means that in the subset of UDMs with 10 probes, we have 55% *more* "Europe" values in the Location (continent) dimension compared to the number of values in that dimension-bin combination for all ASes. Similarly, for the same subset, we see that the "Customer cone (#ASNs)_1.0-3.0" bias cause has a value around -70%, which means that the subset of UDMs with 10 probes has 70% *less* values in that dimension-bin combination.

In other words, if we want to decrease the bias of our sample, we need to include less ASes that belong in Europe, and more ASes with 1-3 ASes in their Customer cone. Therefore, our approach serves a double purpose: it both gives us the specific reason for the presence of bias in our sample, and at the same time it gives us a way to reduce it.

5.2 Bias causes calculation

The calculation of the bias causes was done via the "causes" endpoint of the AI4NetMon API. This endpoint takes as input a set of ASNs (which, in our case correspond to the ASes in a UDM we sampled

[1] We remind the reader here that the number of probes for UDMs is *not* exact as we discussed in Section 2.2 when we described the User measurements dataset.

from the RIPE Atlas API), and returns a dictionary with all the bias causes for the list of ASNs we used as input for the endpoint. Examples of the URI used as well as the corresponding API response can be seen in Figures 5.2-2 and 5.2-3.

Having all the bias causes for all the measurements, we created a dataset with four columns: measurement ID, number of probes, bias causes and value (see Figure 5.2-1). Since we have many different measurements, the same bias cause can appear with different values for different measurements. In order to be able to plot the results for UDMs with different numbers of probes, we aggregated our data.

	meas_id	num_probes	bias_causes	value
0	1007869	10	Customer cone (#ASNs)_1.0-3.0	-43.8360
1	1007869	10	Customer cone (#ASNs)_3.0-9.0	6.2736
2	1007869	10	Customer cone (#ASNs)_9.0-26.0	-1.3839
3	1007869	10	Customer cone (#ASNs)_26.0-76.0	9.4128
4	1007869	10	Customer cone (#ASNs)_76.0-222.0	-0.2738
...	...	...	...	...
151036	1035188	100	Network type (PeeringDB)_Route Collector	-0.0807
151037	1035188	100	Peering policy (PeeringDB)_Open	-45.2921
151038	1035188	100	Peering policy (PeeringDB)_Selective	37.0438
151039	1035188	100	Peering policy (PeeringDB)_Restrictive	9.5581
151040	1035188	100	Peering policy (PeeringDB)_No	-1.3097

Figure 5.2-1: Partial view of the dataframe containing all the bias causes for all UDMs.

https://ai4netmon.csd.auth.gr/api/bias/cause/asn/?asn=8359&asn=6939

Figure 5.2-2: URI used for requesting the bias causes for the measurement with Measurement ID 1022483, which contains the ASNs: 8359, 6939.

The aggregation we did was the following: for each subset of our complete dataframe with a certain number of probes, we grouped by the bias causes column and kept the median value for each dimension-bin combination in the bias causes column. That way, we were left with three dataframes, one for each number of probes we had in our sample (10, 50, 100), in which each bias cause that appears, appears only once. Finally, we kept the top 5 positive and top 5 negative bias causes for each subset. In Figure 5.1-1 we see the results for the UDMs with 10 probes. The plots for the UDMs with 50 and 100 probes can be found in Appendix A1.

Having introduced what we mean by “bias causes” as well as explaining how it is calculated and how we gathered our bias causes data, in the next section we plot and analyze all our findings regarding bias causes for all samples with different numbers of probes.


```
{
    "Custom list": {
        "Location (continent)": {
            "North America": "19.7757%",
            "Europe": "20.523%",
            "Asia": "-24.6456%",
            "South America": "-10.3454%",
            "Oceania": "-3.213%",
            "Africa": "-2.0948%"
```


Figure 5.2-3: Part of the response of the bias causes endpoint for the measurement with ID 1022483.

5.3 Bias causes analysis

In this section we analyze the bias causes of UDMs. We split these UDMs to those with 10, 50 and 100 probes[2] in order to see if there is any correlation between bias causes and number of probes, since we already know that more probes can lead to less biased measurements (Section 3). For this task we utilize the plot in Figure 5.3-1.

In this plot, we see all the top 10 (top 5 positive and top 5 negative) aggregated (see Section 5.2) bias causes that appear in all subsets of UDMs with different numbers of probes. Namely, on the y-axis we have the bias causes, on the x-axis we have the number of probes to denote the different UDM subsets we use, and the fill value is the actual percentage difference between all ASes and our UDMs.

For example, if we take a look at the first column of the heatmap, which corresponds to UDMs with 10 probes, we get all the values we saw in Figure 5.1-1. The same holds for the other columns of the heatmap as well.

In addition, we see that for some columns, there are bias causes (i.e. rows) that have no value, as is the case for the ASDB C1L1_ICT dimension-bin combination for the UDMs with 10 and 50 probes. The reason this happens is because in the y-axis we have included all top 10 bias causes for all subsets. We see that the coloured rows for the UDMs with 10 probes are in total 10, so whichever are not, are simply in the top 10 bias causes of another sample.

Having explained this plot, let us proceed with what we can take away from it.

[2] We remind the reader here that the number of probes for UDMs is *not* exact as we discussed in Section 2.2 when we described the User measurements dataset.

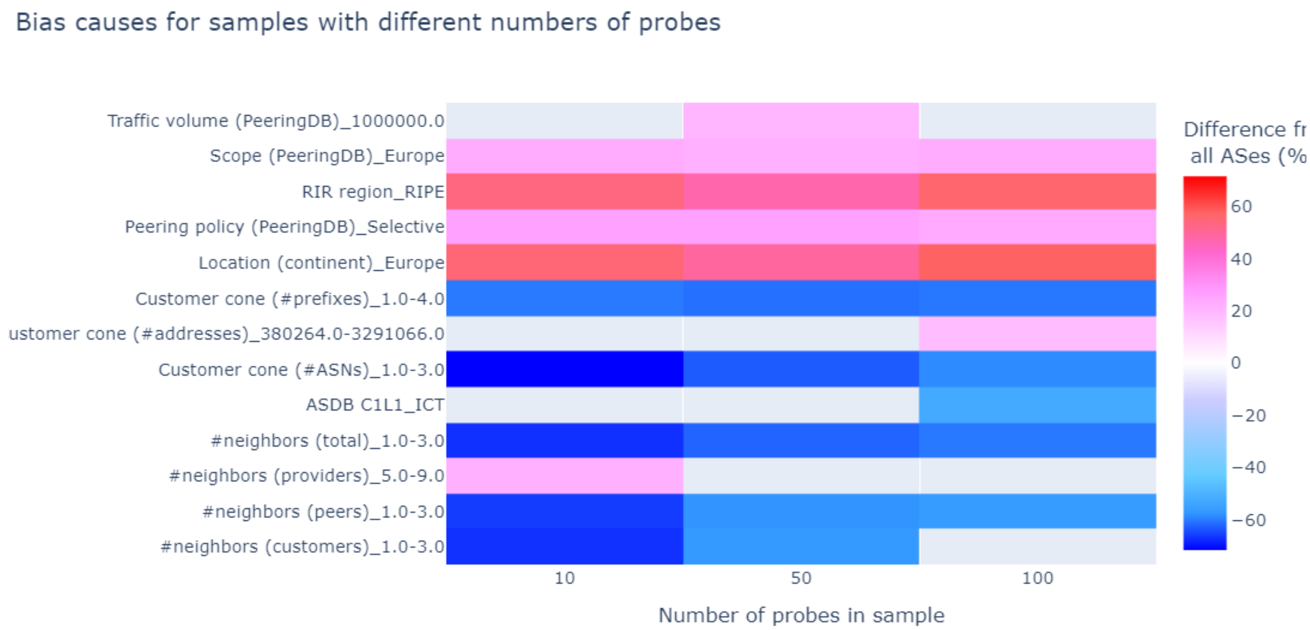

Figure 5.3-1: Bias causes heatmap for UDMs with 10, 50 and 100 probes.

First of all, it seems that the causes of bias are more or less the same regardless of the number of probes we consider. We see that most dimension-bin combinations appear in all samples with different numbers of probes and with, roughly, the same values. This tells us that the number of probes does not seem to significantly affect the reasons why measurements are biased. This was expected though, as we have already seen that the decrease in bias that comes from the increase in the number of probes is quite uniform (Section 4.1).

There are some clear patterns though, namely we see that almost all different UDM subsets have similar values and bias causes. More specifically, we see that all UDMs have values around -50% to -60% in the #neighbors dimensions, and specifically in the 1.0-3.0 bin. This means that UDMs avoid using probes that belong to ASes that have 1-3 neighbors, who are mostly peers and customers. In other words, users prefer ASes that are moderately to well connected. Also since customers and peers are under-represented as neighbors, this means that users probably prefer ASes with mostly providers as neighbors. If we were to decrease the bias of UDMs we would need to use more probes from ASes with more neighbors.

The same is true for the Customer cone bias causes, where we see that users tend to avoid ASes with few ASNs (1-3) and prefixes (1-4) in their Customer cone, rather they seem to prefer medium to larger ASes.

In addition, we see that the bias causes “Location (continent)_Europe” and “RIR region_RIPE” appear with the same values (~50%) for all UDMs, which means that UDMs tend to use probes with ASes that are under the RIPE RIR, and more specifically, in Europe. A similar thing can be said about ASes with Selective Peering policy, as we see that the “Peering policy (PeeringDB)_Selective” bias cause has value around 30%, meaning that users prefer ASes with Selective Peering Policies.

To summarize what we saw so far, we have concluded that measurements made by RIPE Atlas users, tend to:

1) Prefer medium to well connected ASes with mostly providers as neighbors.
2) Prefer medium/larger ASes
3) Mostly use European ASes
4) Prefer ASes with Selective Peering policy.

This gives a nice characterization of user measurements in RIPE Atlas, and seems to be in line with the results of THIS REPORT [REF previous report], in which we saw that the most frequent ASes used in our sample of RIPE Atlas measurements tend to be large ISPs.

Conclusion

In this report, we did two things: first, we analyzed how bias in Internet measurements is affected by the number of probes and the way those probes are selected and then we provided a characterization of UDMs based on the causes of their bias.

Namely, we first compared the bias of measurements with probes selected via random sampling and the Atlas automated algorithm, and compared these with random UDMs from RIPE Atlas. We saw how increasing the number of probes in measurements lowers the average bias of that measurement, and in fact it does so in a uniform way: there seem to be no bias dimensions particularly affected by the increase in the number of probes, rather, they all seem to have lowered bias. This decrease in bias, though, has a lower bound, which is the inherent bias that the IMP, in this case RIPE Atlas, has.

In addition, we saw that UDMs differentiate themselves from measurements with random probes and probes selected via Atlas' algorithm, showing increased bias values in three dimensions: AS hegemony, Scope (PeeringDB), and Location (continent).

Then, we introduced how we define and calculate the bias causes for measurements, and through that, we were able to characterize random RIPE Atlas UDMs. Namely, RIPE Atlas users tend to:

- Prefer medium to well connected ASes with mostly providers as neighbors.
- Prefer medium/larger ASes.
- Mostly use European ASes.
- Prefer ASes with Selective Peering policy.

Therefore, one main takeaway from this work is that users will be able to significantly decrease the bias of their measurements if they:

1) Employ random sampling for the selection of their measurements' probes.
2) User more probes for their measurements.
3) Use probes that lie in smaller, less connected ASes with Open Peering Policies.

Appendix

A1. Bias causes plots for UDMs with different numbers of probes

In this section, we present the bias causes plots for each sample of UDMs we have considered, i.e. UDMs with 10, 50 and 100 probes. In Figure 5.1-1 we have already seen the first one of these (measurements with 10 probes). For each sample we plot the top 5 positive and negative bias causes.

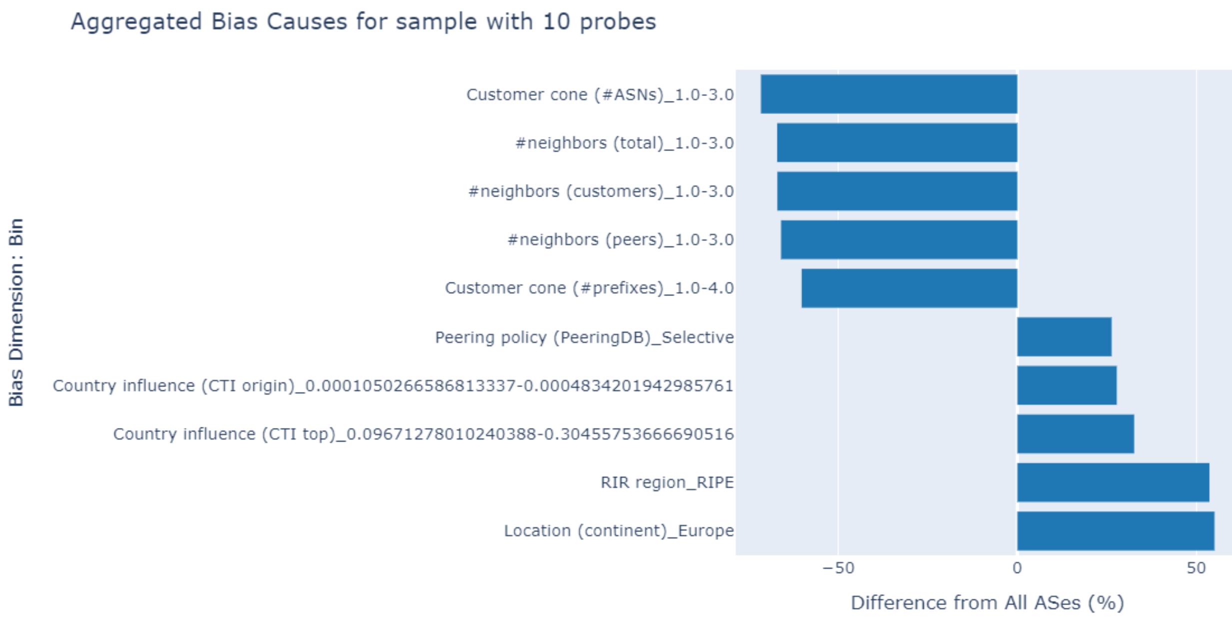

Figure A1-1: Top 5 positive and negative bias causes for the measurements with 10 probes from the User measurement dataset.

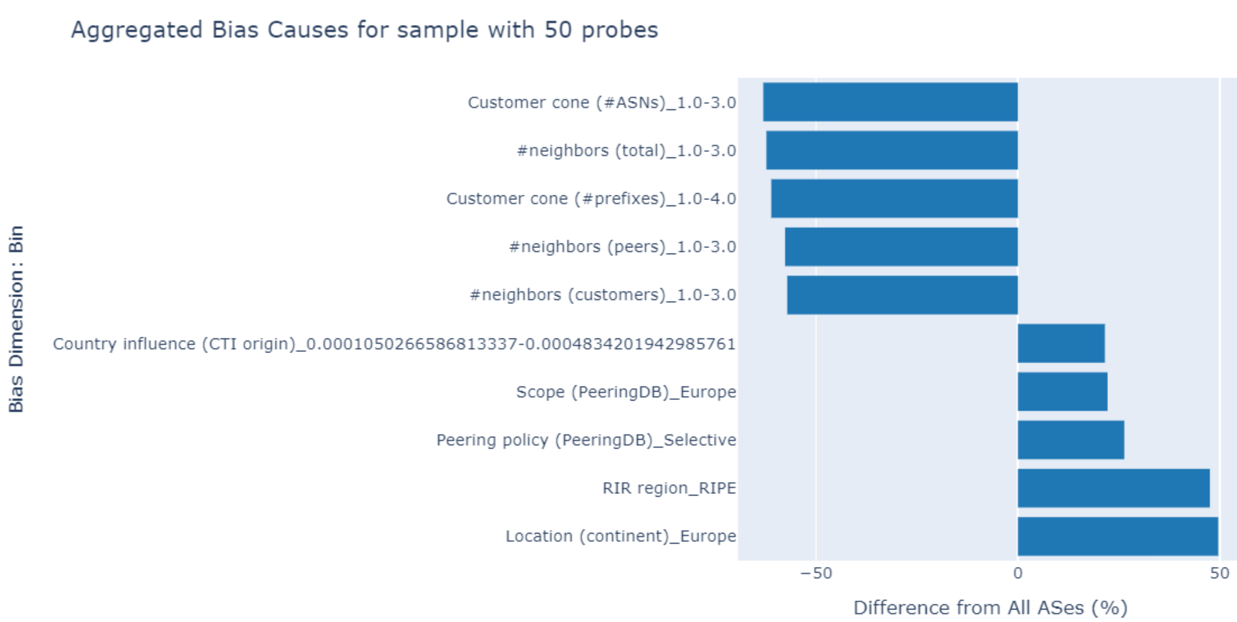

Figure A1-2: Top 5 positive and negative bias causes for the measurements with 50 probes from the User measurement dataset.

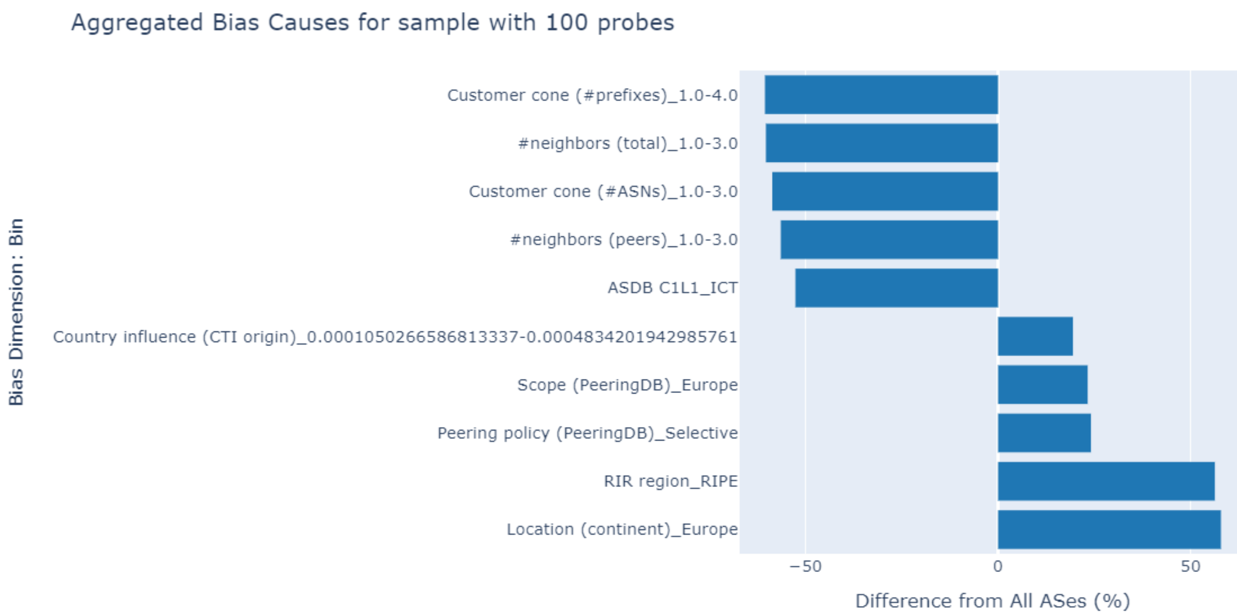

Figure A1-3: Top 5 positive and negative bias causes for the measurements with 100 probes from the User measurement dataset.