Towards Sustainable Conferences: An Empirical Analysis of ACM SIGPLAN's Carbon Footprint

Cristina V. Lopes   Benjamin C. Pierce   Yannick Zakowski

April 16, 2021

Abstract

In 2017, ACM's Special Interest Group on Programming Languages (SIGPLAN) formed an ad-hoc committee to study issues related to climate change—in particular, how SIGPLAN can contribute to the 45% reduction in carbon emissions that the IPCC says is needed by 2030 to maintain warming under 1.5°C [MDZP+18]. One important part of this group's efforts was to gather data pertaining to SIGPLAN conferences so to gain a better understanding of their present emissions. This paper explains the data we gathered and presents some preliminary analysis of this data. Our main finding is that there is an inherent conflict between SIGPLAN's goals of geographic inclusiveness on one hand and of reducing carbon emissions on the other. Going forward, innovative approaches for how to organize conferences that are both inclusive and carbon efficient will be needed.

Our results are specific to SIGPLAN. However, we hope that other research communities can benefit from similar introspection, drawing their own conclusions from data on their own conferences. To this end, we also describe the open-source Python scripts we developed to conduct our analysis.

Working draft of April 16, 2021

1 Introduction

Given the existential threat of global warming, it is incumbent on individuals and organizations to evaluate the carbon emissions associated with their activities and find ways to reduce them. For many academic researchers, these emissions will overwhelmingly come from air travel, especially to international conferences.

This observation raises a number of questions about how to organize our professional activities so as to maximize progress while minimizing emissions. Should SIGPLAN conference locations be chosen to minimize their carbon impact? If so, how? Should we move toward co-locating conferences? Or, on the contrary, should some conferences be split into regional meetings or held simultaneously at two sites on different continents? Should we continue holding some conferences entirely virtually, post Covid?

To ground discussions about the decisions and compromises that the scientific community may collectively wish to undertake, at least three main sorts of data seem useful.

- The estimated emissions of past conferences.
- The geographical distribution of participants to conferences.
- The overlap in participation between various conferences.

We outline the results of a preliminary analysis of the past several years of registration data for four of the main SIGPLAN conferences. We hope this effort can serve as a basis both for debates about concrete measures and for larger and more comprehensive studies.

After briefly describing our dataset in Section 2, we present estimates of the individual footprints of each conference in Section 3. In Section 4, the core of our analysis, we derive several statistics about the geographical distribution of participants and their habits of cross-participation—across years and across conferences—arguing that these data are correlated to the footprint. We then present in Section ?? a speculative experiment aiming to estimate "ideal locations" for past conferences in order to minimize their footprints. In Section ?? we draw some concrete recommendations for future conference organizers based on these analyses. Finally, in Section 5, we outline the open-source tool we developed to conduct our analyses, in hopes that other communities might piggyback on our efforts to conduct their own similar studies.

2 Dataset

Our dataset consists of 10 years of registration records for the four major SIGPLAN conference series—POPL, PLDI, ICFP, and SPLASH—from the beginning of 2009 until the end of 2018. Data for a few of the conferences in the earlier years is missing. In total, we have data for 33 conferences, corresponding to 8,758 unique participants and 16,374 trips. For each participant, we know all the conferences they attended and from which city they departed to attend the conferences. Before we started working with the data, the names of participants were replaced in the dataset by unique hashes, obscuring each individual's identity while allowing them to be identified across years and across the conferences they attended.

3 Estimating the Footprints of Conferences

Carbon footprint is the key metric that we ultimately seek to reduce and hence also the starting point of our analysis. We introduce in this section the methodology we used and tool we built to conduct all of our analyses, and we describe the first results from our dataset.

3.1 Methodology for Evaluating Carbon Footprint

We conduct all our analyses through a Python 3 script,[1] described in more detail in Section 5. Throughout, we make the following assumptions:

- we assume that participant travel accounts for the entire carbon footprint of a conference;
- we assume that all conference participants travel by plane, in economy class;
- we assume that the airports in the conference city and in each participant's home city are close enough to the actual end points of their travel for their locations to be assimilated;
- we assume that all flights are direct;
- we assume that the geodesic distance is the one taken by planes.

Estimating the errors introduced by these assumptions and refining the analysis to make more realistic assumptions would obviously be very worthwhile. But, for this first effort, we are mainly aiming to get a *relative* evaluation of different potential strategies for reducing footprints; for this purpose, we believe these assumptions are good enough.

The distance traveled by each participant is converted to an amount of emissions expressed in $\mathrm{kg}_{\mathrm{CO_2e}}$. To do this conversion, we use a standard model introduced as part of the `DEFRA 16` report on Greenhouse gas [2] [3] conducted by the British Government.

The model distinguishes three classes of flight, depending on their length (short, medium, or long haul). Each class is associated with a linear coefficient relating the distance of travel to the amount of $\mathrm{kg}_{\mathrm{CO_2e}}$ emitted.

[1] Publicly available at `https://github.com/YaZko/sigplan-carbon-analysis`
[2] `https://www.gov.uk/government/publications/greenhouse-gas-reporting-conversion-factors-2016`
[3] `https://co2calculator.acm.org/methodology.pdf`

A second linear coefficient, identical for all flights, is the so-called *radiative forcing index*; this is used to account for the difference in radiative forcing between the same emissions at ground level compared to high in the atmosphere. We use the value 1.891 for this coefficient, as suggested by R. Sausen et al. [SIH+05]

We thus obtain the following piecewise-linear model of emissions for a flight covering d kms:

$$\text{kg } (\text{CO}_2\text{e}) \text{ per participant} = \begin{array}{ll} 1.891 * 0.14735 * d & \text{if } d < 785 \\ 1.891 * 0.08728 * d & \text{if } 785 \leq d < 3700 \\ 1.891 * 0.077610 * d & \text{if } 3700 \leq d \end{array}$$

3.2 Conference Footprints

We now turn to the estimation of footprints in our dataset. Table 1 depicts the total and average carbon footprint per participant of all conferences analyzed. This footprint is estimated in terms of $\text{t}_{\text{CO}_2\text{e}}$ (metric tons of CO_2-equivalent) of emissions.

The main figure of interest is arguably the last column, depicting the average footprint per participant. The lowest average per-participant footprint of our dataset are tied ICFP'12 and ICFP'14 at $0.88\text{t}_{\text{CO}_2\text{e}}$, while the highest one is ICFP'16 at $1.93\text{t}_{\text{CO}_2\text{e}}$.

Observation 1. *The average per-participant carbon footprint due to air travel varies across conferences in our dataset by around a factor of 2.*

4 Data Analysis: Community

The greenhouse gas emissions from a given event is in direct proportion to the average distance traveled by the participants of this event. To understand emissions, we must therefore estimate the nature of the communities that attend each conference.

The aggregated information we describe below falls into two main categories: first, the demographic distribution of the participants to the conferences conditioned by various factors, and second, the participation habits of the community through recurring participation to a given conference and the overlap in participation between different conferences.

4.1 Demographics: Where Did Participants Come From?

Figure 2[4] and Table 3 show where all participants came from, grouped by "continents": North and South America, Europe, Asia, Africa, and Oceania. For each conference, we depict the distribution of attendance per continent. Table 3 shows the portion of attendants originating from the same continent as the one the event took place in. To a first approximation, maximizing this last metric, i.e. hosting conferences in the continent containing the majority of its community, is a good thing.

Taken as a whole, these conferences attracted 50% of their participants from North America, 36% from Europe, 11% from Asia, 2% from Oceania, 1% from South America, and less than 0.2% from Africa. The data also shows some degree of geographical affinity for the various conferences: PLDI and SPLASH appear to be quite North-America-centric, while ICFP's core community has a strong anchor in Europe as well.

This overall picture, however, hides some interesting facts pertaining to the relationship between the conferences' locations and the origin of the participants. Indeed, aggregating the attendance per conference intrinsically rests upon the assumption of a uniform community that attends every instance of the conference every year. This picture turns out to be quite misleading.

Table 4 and Figure 5 show a more detailed breakdown of the origin of participants for each conference, showing also the geographic region where the conferences were held. These charts make it clear that the

[4] The graphical representations in this preliminary draft are based on a slightly different version of our dataset than the one used by our tool. There may be some minor discrepancies between these representations and the raw tables presented.[BCP: Hopefully we can remove this! Or, if it's just the big red-and-green table, we can at least move this comment there.]

Event	Location	# Participants	Total footprint	Average footprint
ICFP 10	Baltimore	336	399.44	1.19
ICFP 11	Tokyo	336	518.27	1.54
ICFP 12	Copenhagen	481	422.34	0.88
ICFP 13	Boston	505	512.06	1.01
ICFP 14	Gothenburg	483	426.35	0.88
ICFP 15	Vancouver	439	636.14	1.45
ICFP 16	Nara	528	1019.65	1.93
ICFP 17	Oxford	592	610.05	1.03
ICFP 18	St. Louis	487	572.49	1.18
POPL 9	Savannah	331	488.19	1.47
POPL 11	Austin	403	595.97	1.48
POPL 12	Philadelphia	536	586.16	1.09
POPL 13	Rome	540	658.2	1.22
POPL 14	San Diego	533	905.64	1.7
POPL 15	Mumbai	463	748.24	1.62
POPL 16	St. Petersburg	488	695.45	1.43
POPL 17	Paris	719	671.79	0.93
POPL 18	Los Angeles	576	932.93	1.62
PLDI 9	Dublin	255	381.48	1.5
PLDI 13	Seattle	467	595.13	1.27
PLDI 14	Edinburgh	427	545.71	1.28
PLDI 15	Portland	465	599.0	1.29
PLDI 16	Santa Barbara	438	575.25	1.31
PLDI 17	Barcelona	495	784.67	1.59
PLDI 18	Philadelphia	468	421.32	0.9
SPLASH 9	Reno	709	1125.85	1.59
SPLASH 10	Sparks	566	821.57	1.45
SPLASH 12	Tucson	434	665.03	1.53
SPLASH 13	Indianapolis	606	668.87	1.1
SPLASH 14	Portland	491	625.91	1.27
SPLASH 15	Pittsburgh	611	777.82	1.27
SPLASH 16	Amsterdam	584	595.63	1.02
SPLASH 17	Vancouver	582	864.69	1.49

Table 1: For each event: location, number of participants and carbon footprint, total and average per participant, in t_{CO_2e}.

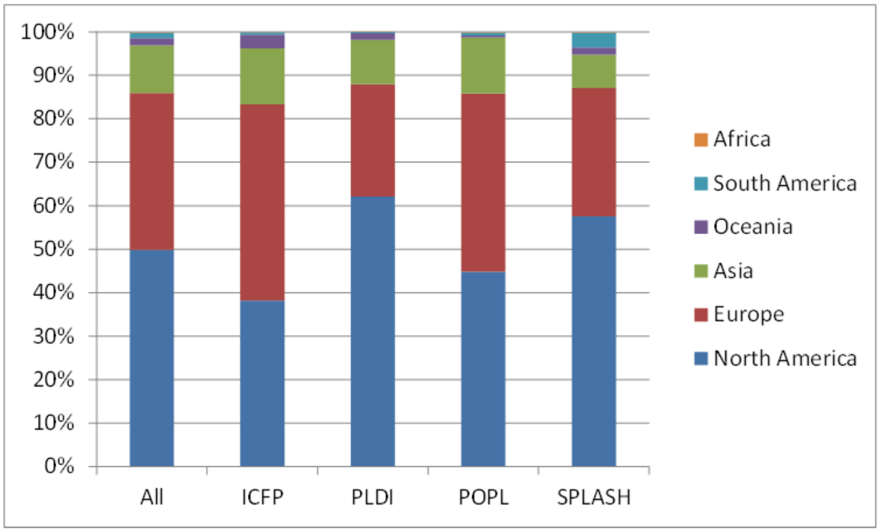

Figure 2: Overall origin of participants per conference.

Conference	**EU (%)**	**NA (%)**	**AS (%)**	**SA (%)**	**AF (%)**	**OC (%)**	**Local (%)**
ICFP	45.09	38.19	12.9	0.62	0.1	3.1	59.18
POPL	41.01	44.82	12.77	0.7	0.17	0.52	56.98
PLDI	25.84	62.16	10.08	0.36	0.03	1.53	60.13
SPLASH	29.61	57.47	7.64	3.43	0.17	1.68	61.75
Any	36.07	49.86	10.87	1.38	0.13	1.69	59.46

Table 3: For each kind of conference, distribution of participants per continent of origin. Among all participants of a category of conferences, displays the percentage of these participants that traveled from the indicated continent. The **Local** column uses for each instance of the conference the same continent as the one the conference took place in. The line *Any* computes the same data, but across all conferences at once.

Event	Location	EU (%)	NA (%)	AS (%)	SA (%)	AF (%)	OC (%)	Local (%)
ICFP 10	NA	34.82	55.95	8.33	0.0	0.0	0.89	55.95
ICFP 11	AS	33.04	17.86	46.43	0.0	0.0	2.68	46.43
ICFP 12	EU	70.06	22.45	6.03	0.21	0.0	1.25	70.06
ICFP 13	NA	31.09	62.18	3.56	0.99	0.0	2.18	62.18
ICFP 14	EU	72.26	19.25	4.55	0.62	0.0	3.31	72.26
ICFP 15	NA	28.93	58.54	5.01	1.59	0.46	5.47	58.54
ICFP 16	AS	32.01	24.43	35.98	0.76	0.19	6.63	35.98
ICFP 17	EU	64.19	24.16	7.26	0.17	0.17	4.05	64.19
ICFP 18	NA	28.95	63.04	6.57	1.03	0.0	0.41	63.04
POPL 9	NA	39.88	49.24	9.37	0.6	0.0	0.91	49.24
POPL 11	NA	35.24	56.08	7.94	0.0	0.0	0.74	56.08
POPL 12	NA	29.48	61.01	8.4	0.19	0.19	0.75	61.01
POPL 13	EU	58.89	29.44	11.3	0.19	0.0	0.19	58.89
POPL 14	NA	36.59	54.22	6.94	1.31	0.19	0.75	54.22
POPL 15	AS	29.37	21.6	48.6	0.0	0.22	0.22	48.6
POPL 16	NA	33.4	57.79	7.79	0.61	0.2	0.2	57.79
POPL 17	EU	63.56	25.45	8.9	1.25	0.42	0.42	63.56
POPL 18	NA	31.42	56.94	9.2	1.56	0.17	0.69	56.94
PLDI 9	EU	29.8	59.61	8.63	0.39	0.39	1.18	29.8
PLDI 13	NA	18.2	69.38	10.28	0.21	0.0	1.93	69.38
PLDI 14	EU	42.86	44.26	10.54	0.7	0.0	1.64	42.86
PLDI 15	NA	20.65	70.75	7.31	0.0	0.0	1.29	70.75
PLDI 16	NA	13.93	73.29	11.87	0.0	0.0	0.91	73.29
PLDI 17	EU	43.23	38.99	14.14	1.01	0.0	2.63	43.23
PLDI 18	NA	13.68	78.21	7.05	0.21	0.0	0.85	78.21
SPLASH 9	NA	25.39	61.35	8.18	3.53	0.28	1.27	61.35
SPLASH 10	NA	23.32	62.9	8.83	2.83	0.18	1.94	62.9
SPLASH 12	NA	22.35	62.21	11.52	2.3	0.23	1.38	62.21
SPLASH 13	NA	24.75	65.02	4.79	3.96	0.17	1.32	65.02
SPLASH 14	NA	19.96	68.84	4.68	4.28	0.2	2.04	68.84
SPLASH 15	NA	30.28	56.14	7.2	4.09	0.0	2.29	56.14
SPLASH 16	EU	60.96	27.4	8.22	2.05	0.34	1.03	60.96
SPLASH 17	NA	27.32	58.08	8.25	4.12	0.0	2.23	58.08

Table 4: For each event, the continent in which it took place and the distribution of attendance per continent of origin of participants. The final column indicates the portion of participants that did not change continent to attend the conference.

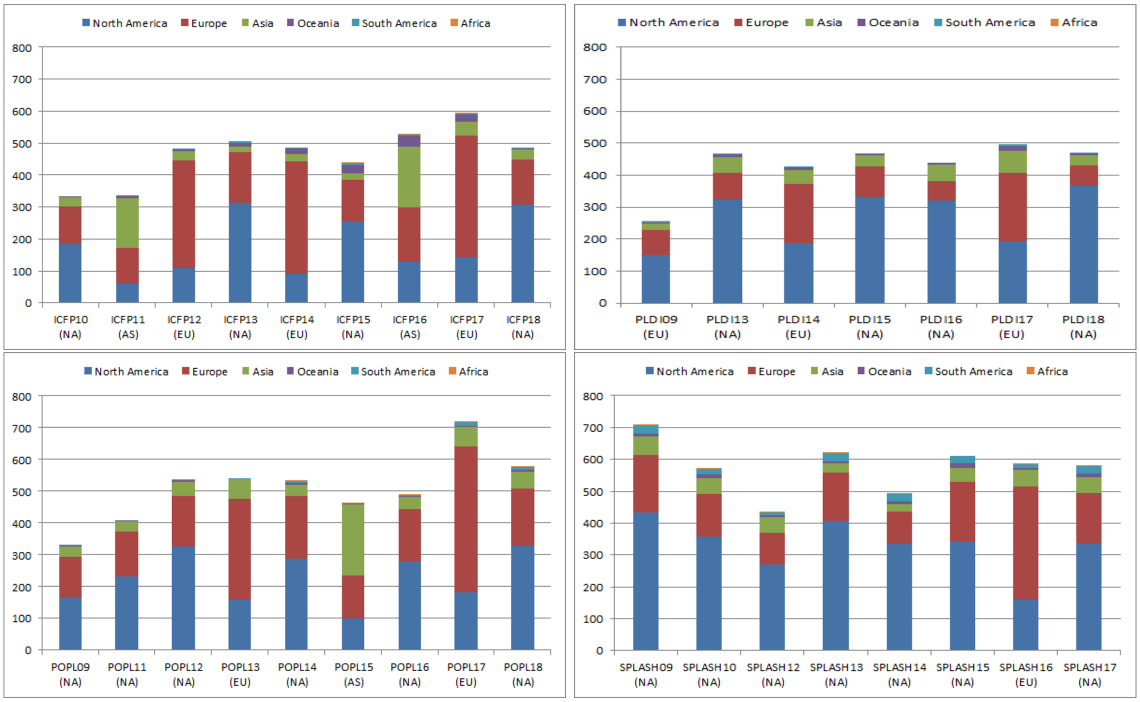

Figure 5: Origin of participants for each conference, detail.

Location	EU (%)	NA (%)	AS (%)	SA (%)	AF (%)	OC (%)	Local (%)
EU	58.35	30.16	8.83	0.79	0.15	1.73	58.35
NA	26.93	62.03	7.69	1.78	0.11	1.46	62.03
AS	31.35	21.78	43.03	0.3	0.15	3.39	43.03
Any	36.07	49.86	10.87	1.38	0.13	1.69	59.46

Table 6: Geographical distribution of participation conditioned by the location of the event: each row indicates the continent in which the conference took place, and each cell of this row depicts the percentage of participants of these conferences that originated from a given continent. The column **Local** corresponds to the same continent as the conference.

location of the conference had a substantial effect on the distribution of attendees, with each conference tending to attract people from the same geographic area. This effect is quite visible for ICFP and POPL, with noticeable ups and downs of the colored bars between North American and European participants when the conferences were located in North America and Europe, respectively. Most strikingly, Asian participation during POPL '15, ICFP '11 and ICFP '16, events that took place on the Asian continent, is significantly higher than the rest: there appears to be a strong locality phenomenon here. Cross-referencing this data with Table 1, one can also notice that the only time SPLASH took place in Europe also turned out to be the least carbon-intensive edition, challenging the previous observation, based on a high-level view of past attendance data, that the conference might appear to be mostly North-America-centric.

Table 6 attempts to measure this locality effect. The table depicts, all conferences being considered at once, the geographical distribution of attendance conditioned by the geographical location of the event. The Asian phenomenon previously hinted at is here extremely apparent: while overall, on average, 10.9% of the participants come from Asia, this number is roughly multiplied by a factor 4 when the event takes place in Asia (without any significant drop in total volume of attendance that could indirectly bump this percentage). Interestingly, this phenomenon also exists in the case of Europe (+22.29% deviation from the average) and

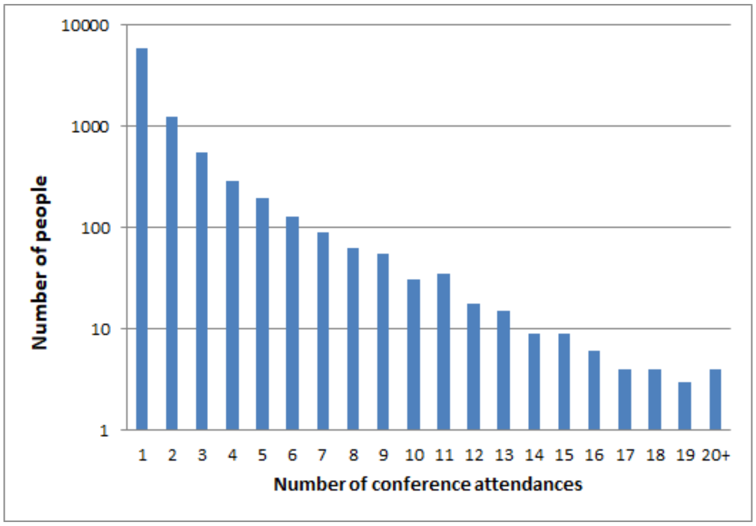

Figure 7: Histogram of attendance.

North America (+12.15% deviation from the average). Thus, despite their name, individual instances of international conferences appear to exhibit a fairly strong local component.

Overall, this data shows that the goal of increasing geographic inclusion was, indeed, accomplished by organizing the conferences in diverse parts of the world. It also places Figure 5 in a broader context: a naive interpretation of that chart might lead us to conclude that North America and Europe are where most of this community is, but it is not that simple. Because of the regional effect on participation, the distribution of participants also reflects the fact that most of these conferences were *held* in North America and Europe (30), only a few were held in Asia (3), and none in South America, Oceania, or Africa.

The situation may be summed up in two elementary observations:

Observation 2. *The vast majority of participants are split between North America and Europe, with the remainder mostly coming from Asia. SPLASH and PLDI are strongly anchored in North America. ICFP and POPL fairly equally split between North America and Europe.*

This distribution, however, is *strongly* dependent on the location of the event.

Observation 3. *There is also a strong locality effect in conference attendance: nearby conferences attract significant numbers of new participants from the area, while longer distances discourage some participants.*

4.2 How Often Did Participants Attend These Conferences?

Section 4.1, through the study of the demographic distribution of attendance, has suggested the existence of local communities that only partake in conferences when they take place close to their place of residency. One can conversely look for groups of "regular attendees" that participate in a given regardless of where it is held.

Figure 7 shows how often the same participants attended multiple conferences. At the extremes, 6,009 people (69%) attended only 1 conference, and just 4 people attended 20 or more conferences. Participation is dominated by single-conference participants, perhaps reflecting a large and transient student population. The pattern is similar for each conference series, shown in Figure 8.

4.3 What Was the Participation Overlap Between These Conferences?

We now take a closer look at the habits of these recurring participants.

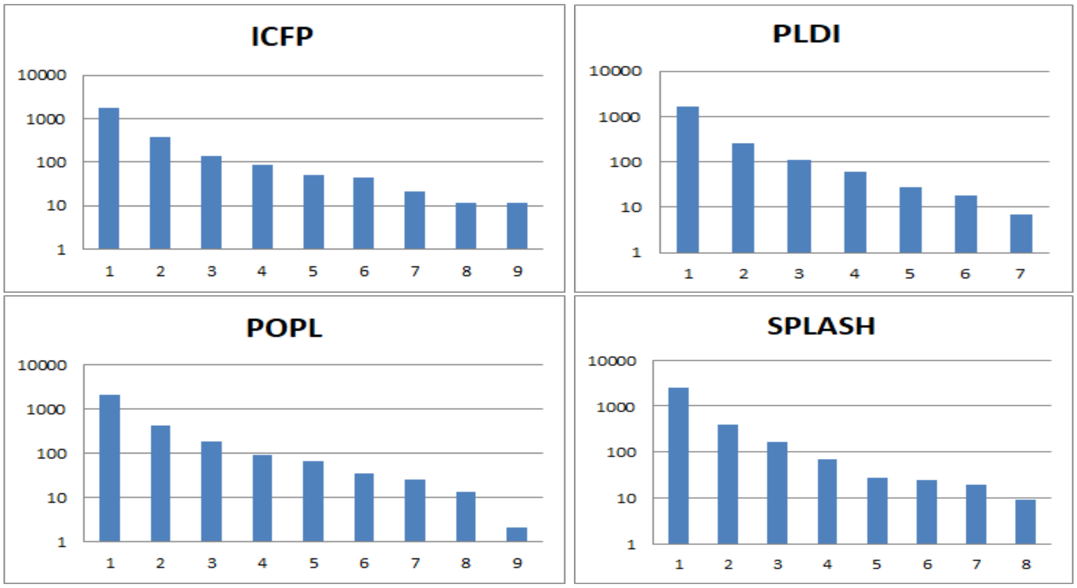

Figure 8: Histogram of attendance for each conference series.

A first natural question is whether there is significant overlap in participation between conferences. Table 9 depicts, for each pairing of the four conferences, the percentage of overlap. This measure is strikingly low for most conferences.

Observation 4. *Cross-conference overlap is low: the tightest pairing sees slightly over 10% of common attendance for a given year. Extending the overlap among any two years, the tightest pairing still sees less than a quarter of unique participants having participated at least once in both conferences. [BCP: This may be the observation that people will be most interested in. Is there any more we can say about it? Are the statistics we're presenting really the most revealing ones? Also: The figure is presented in abvolute nunbers, while the tables use percentages. Maybe we should stick with one or the other?]*

Conversely, one can estimate the overlap for a given conference over time: for a given conference at a time, and for any pair of years, compute the percentage of attendees that participated in both events. This new information, as well as the essential of Table 9, is synthesized graphically on Figure 10. With this bird-eye view of the permanence vs. transience of the participants over time in SIGPLAN conferences, we can make a second observation, temporal this time:

Observation 5. *Temporal overlap is moderate: roughly a quarter of attendees at a given conference were also present the year before at the same conference.*

In principle, it is desirable to have a balance between repeat participants and newcomers. Communities that don't attract new participants tend to stagnate; but communities that don't have a core of repeat participants tend to lose focus.

The existence of a stable community associated with each conference series (i.e., a group that tends to repeat participation) is clearly visible near the diagonal in Figure 10. The highest overlap of all in particular was between ICFP'16 and ICFP'17, with 180 repeaters. The four conference series show what seems to us to be a healthy balance between repeat participation and newcomers.

The weaker overlap between conferences in different series is also apparent. The strongest overlap is between PLDI and POPL, followed by ICFP and POPL and by PLDI and SPLASH. The weakest overlaps are between ICFP and SPLASH, followed by ICFP and PLDI, and by POPL and SPLASH. It is unclear whether the overlap, or lack thereof, between these conference series is due to intellectual reasons or due to their dates. PLDI and POPL is the pair that is most distant in time, typically June and January. Conversely,

Year	Overlap	#ICFP	#POPL
11	40	335	401
12	76	481	536
13	85	505	540
14	60	483	533
15	27	439	463
16	56	528	488
17	89	592	719
18	74	487	575
All	507	3850	4255

(a) ICFP and POPL

Year	Overlap	#POPL	#PLDI
9	23	331	254
13	54	540	467
14	68	533	427
15	31	463	465
16	56	488	438
17	61	719	495
18	68	575	468
All	361	3649	3014

(b) POPL and PLDI

Year	Overlap	#POPL	#SPLASH
9	17	331	709
12	35	536	434
13	31	540	606
14	41	533	491
15	15	463	611
16	32	488	584
17	38	719	582
All	209	3610	4017

(c) POPL and SPLASH

Year	Overlap	#ICFP	#PLDI
13	24	505	467
14	20	483	427
15	21	439	465
16	14	528	438
17	34	592	495
18	35	487	468
All	148	3034	2760

(d) ICFP and PLDI

Year	Overlap	#ICFP	#SPLASH
10	12	335	566
12	9	481	434
13	14	505	606
14	12	483	491
15	20	439	611
16	19	528	584
17	20	592	582
All	106	3363	3874

(e) ICFP and SPLASH

Year	Overlap	#PLDI	#SPLASH
9	23	254	709
13	61	467	606
14	46	427	491
15	70	465	611
16	35	438	584
17	74	495	582
All	309	2546	3583

(f) PLDI and SPLASH

Table 9: For every year, we display the number of participants that attended two given conferences. We also indicate the total attendance of each event for reference. The *All* row depicts the sum over all years.

	ICFP10	ICFP11	ICFP12	ICFP13	ICFP14	ICFP15	ICFP16	ICFP17	ICFP18	PLDI9	PLDI13	PLDI14	PLDI15	PLDI16	PLDI17	PLDI18	POPL9	POPL11	POPL12	POPL13	POPL14	POPL15	POPL16	POPL17	POPL18	SPLASH9	SPLASH10	SPLASH12	SPLASH13	SPLASH14	SPLASH15	SPLASH16	SPLASH17
ICFP10	334	80	101	107	83	69	59	71	52	10	13	13	18	10	20	18	58	55	76	50	54	14	47	43	40	14	12	13	13	13	13	11	12
ICFP11		336	105	82	62	60	88	62	42	5	15	14	16	12	20	7	45	40	45	54	50	20	34	41	36	8	10	5	12	8	10	9	11
ICFP12			481	140	150	104	89	111	71	5	14	21	19	12	22	19	48	50	76	63	68	21	45	57	41	11	13	9	8	17	18	15	16
ICFP13				505	143	138	103	122	101	11	24	26	33	23	39	35	57	56	109	85	100	37	81	75	70	12	14	15	14	22	24	20	25
ICFP14					485	142	108	139	81	7	14	20	20	12	25	17	36	31	52	51	60	20	44	61	36	3	10	6	9	12	14	14	11
ICFP15						439	135	146	108	4	15	17	21	17	32	31	36	29	51	44	62	27	64	69	54	5	12	8	14	17	20	14	26
ICFP16							528	180	112	2	11	16	16	14	39	23	43	26	46	55	55	23	56	71	53	4	13	8	9	13	19	19	26
ICFP17								593	136	6	15	20	21	14	34	28	38	27	55	58	72	25	68	90	79	8	11	10	11	12	21	18	20
ICFP18									487	4	14	19	30	26	45	35	35	26	53	43	65	22	77	66	74	8	16	16	17	20	28	32	40
PLDI9										253	59	39	51	33	30	31	23	38	35	19	18	10	17	14	17	24	24	27	29	29	25	23	32
PLDI13											467	108	127	90	78	72	33	65	74	54	73	37	54	39	54	23	35	49	62	57	51	43	58
PLDI14												427	117	97	72	71	35	47	60	50	68	31	47	35	40	21	23	31	49	46	40	38	58
PLDI15													468	130	90	88	34	51	61	48	76	31	62	47	67	25	38	46	75	72	70	68	67
PLDI16														438	81	106	28	52	46	34	64	29	56	35	63	22	23	35	48	52	49	35	71
PLDI17															497	75	32	45	62	49	62	35	64	61	61	31	32	43	59	43	71	71	74
PLDI18																468	32	46	64	34	62	37	67	60	68	18	21	31	38	43	45	33	53
POPL9																	331	94	101	74	87	29	65	64	58	17	17	15	22	14	12	10	16
POPL11																		404	150	101	103	39	78	74	76	23	21	29	30	31	27	19	32
POPL12																			536	129	147	60	108	101	98	23	27	35	36	39	35	25	42
POPL13																				540	152	83	97	124	91	7	14	23	31	29	33	26	34
POPL14																					533	90	137	132	136	16	13	33	37	41	40	31	42
POPL15																						463	74	75	81	5	2	10	18	18	15	20	17
POPL16																							488	146	156	7	12	24	33	28	38	32	44
POPL17																								719	151	7	14	18	26	24	29	31	38
POPL18																									574	6	11	15	21	23	32	28	35
SPLASH9																										720	145	83	81	66	68	39	46
SPLASH10																											572	101	98	85	89	60	67
SPLASH12																												435	114	94	94	55	83
SPLASH13																													619	123	126	93	97
SPLASH14																														491	127	82	96
SPLASH15																															611	127	128
SPLASH16																																586	116
SPLASH17																																	582

Figure 10: Conference participation overlap.

Conference	Avg	Avg $\geq$ 2	$\geq$ 2	$\geq$ 3	$\geq$ 4	$\geq$ 5
ICFP	1.64	3.22	736 (28.86%)	367 (14.39%)	227 (8.9%)	139 (5.45%)
POPL	1.59	3.11	807 (27.98%)	405 (14.04%)	227 (7.87%)	138 (4.79%)
PLDI	1.43	2.87	485 (22.99%)	223 (10.57%)	112 (5.31%)	53 (2.51%)
SPLASH	1.41	2.88	706 (21.69%)	307 (9.43%)	148 (4.55%)	79 (2.43%)
All	1.88	3.79	2743 (31.43%)	1508 (17.28%)	960 (11.0%)	669 (7.67%)

Table 11: For each conference and overall, the average number of instances a unique individual has taken part of (**Avg**), and the same data omitting individuals that have participated in exactly one conference ever (**Avg $\geq$ 2**). The other columns display the percentage of participants that have attended at least k instances, for $k \in [\![2 \dots 5]\!]$. Note that the means and percentages are computed with respect to the number of *unique* participants.

Year	Avrg	Avrg $\geq$ 2	$\geq$ 2 (%)	$\geq$ 3 (%)	$\geq$ 4 (%)
9	1.05	2.07	56 (4.53%)	4 (0.32%)	0 (0.0%)
10	1.01	2.0	13 (1.46%)	0 (0.0%)	0 (0.0%)
11	1.06	2.0	43 (6.18%)	0 (0.0%)	0 (0.0%)
12	1.09	2.05	110 (8.23%)	5 (0.37%)	0 (0.0%)
13	1.13	2.12	217 (11.57%)	23 (1.23%)	2 (0.11%)
14	1.13	2.14	191 (11.12%)	22 (1.28%)	4 (0.23%)
15	1.09	2.11	152 (8.4%)	16 (0.88%)	0 (0.0%)
16	1.11	2.1	178 (9.66%)	17 (0.92%)	0 (0.0%)
17	1.13	2.13	248 (11.77%)	31 (1.47%)	2 (0.09%)
18	1.12	2.11	146 (10.66%)	16 (1.17%)	0 (0.0%)
All	1.88	3.79	2743 (31.43%)	1508 (17.28%)	960 (11.0%)

Table 12: For each year, the average number of conferences a participant has participated in among POPL, PLDI, ICFP and SPLASH (**Avrg**), and the same data without considering individual that has participated exactly once (**Avrg $\geq$ 2**). The other columns display the percentage of participants that have attended at least k conferences, for $k \in [\![2 \dots 4]\!]$. Note that the means and percentages are computed with respect to the number of *unique* participants.

ICFP and SPLASH is the pair that is the closest in time, typically September and October. One might conjecture that temporal proximity discourages cross-participation.

Finally, Table 11 and 13 offer two different views on recurrent participation. Table 11 represents respectively for the whole dataset (row "ALL") and for each conference individually the average number of editions a participant has been part of, as well as the percentage of participants that have been part of at least a given number of editions of a conference. One striking fact is that no less than 75% of unique participants have been to just a single edition.

Table 13 is a normalization of the information represented in Figure 8: for each instance of each conference, it depicts the percentage of participants that have been part of a previous instance of the conference (in our dataset). Since we take as origin of time the first year for which we have data, the table is naturally overall monotonic as years progress. Exceptions can be noticed, such as POPL'15, that seem to indicate a (proportional) lack of "old timers".

Observation 6. *Over all conferences, the average number of conferences a given participant has attended is just 1.52. Less than 4% of unique participants have been to more than five events among our dataset. Similarly, at any given event, more than half of the participants were experiencing this conference for the first time.*

[BCP: General question about all the pictures and tables: Are they consistent? (I.e., were the pictures generated from the data in the tables, or from earlier versions of that data?)] [YZ: Unfortunately no not

Year	Old timers
9	0 (0.0%)
11	94 (23.33%)
12	190 (35.45%)
13	180 (33.33%)
14	268 (50.28%)
15	141 (30.45%)
16	221 (45.29%)
17	298 (41.45%)
18	311 (53.99%)

(a) Case of POPL

Year	Old timers
10	0 (0.0%)
11	80 (23.81%)
12	154 (32.02%)
13	204 (40.4%)
14	225 (46.58%)
15	209 (47.61%)
16	226 (42.8%)
17	316 (53.38%)
18	221 (45.38%)

(b) Case of ICFP

Year	Old timers
9	0 (0.0%)
13	59 (12.63%)
14	119 (27.87%)
15	178 (38.28%)
16	192 (43.84%)
17	168 (33.94%)
18	188 (40.17%)

(c) Case of PLDI

Year	Old timers
9	0 (0.0%)
10	140 (24.73%)
12	129 (29.72%)
13	174 (28.71%)
14	191 (38.9%)
15	235 (38.46%)
16	206 (35.27%)
17	253 (43.47%)

(d) Case of SPLASH

Table 13: For each conference, percentage of participants that have been part of a previous edition of the same conference.

at the moment: all tables are generated, but the Figures have been manually produced by Crista from her original take on the dataset. It would be great to regenerate them from the generated csv files, but I do not know how they have been generated exactly.]

5 An Open-Source Tool for Analyzing Conference Footprints

We hope that the analysis we have conducted for a few SIGPLAN conferences will offer valuable insights for the organizers of these conferences. Clearly, though, any observations based on our data cannot be taken as universal facts: the situation heavily depends notably on the geographical distribution of the underlying research community and on its cultural habits of attendance. Moreover, the practical conclusions that it should entail may diverge from one community to another. Accordingly, we strongly encourage similar studies to be performed by other groups.

To help with this, we have released an open-source Python 3 script that we have built to be as parameterizable and reusable as possible. All the analyses presented in this paper have been generated using this tool.[5] The script can be found at the following github repository: `https://github.com/YaZko/sigplan-carbon-analysis`. We welcome comments, pull requests, etc., and we would be happy to assist anyone wishing to use the tool for their own analysis.

Detailed documentation is available in the repository. We give here just a high-level overview.

The script takes as an input a dataset described by two `csv` files. The first one describes the list of conferences: each line describes a specific event and the location it took place in, i.e. has the fields `Name, Year, City, State` and `Country`. The second one contains the list of participants of these events: each

[5]The graphical visualizations have been made separately, the script currently only generates tables. Extending it to generate graphical takes on these tables would be an interesting feature.

line describes a unique participation at an event, with the location of origin of the participant, i.e., it has the fields `Identifier, City, State, Country, Conference` and `Year`.

The first pass of the analysis computes the needed raw data. Informal named locations manually provided by participant are mapped to their ISO designation using the `pycountry` library. Once this is done, these named locations are converted to GPS locations using the `geopy` library, which provides a straightforward API to do this. To avoid duplicating requests to online APIs, all of these computations are cached locally.

Distances in kilometers between locations are then computed between GPS locations once again using the `geopy` library. They use the geodesic distance (shortest distance for an ellipsoidal model of the Earth) with a model providing precision that is several orders more precise than we need.

At this point, we know, for each participant in a conference, the distance they traveled. The script then uses a model that computes the carbon footprint of air travel based on this information. For the analysis presented in this paper, we used the `DEFRA 16` model described in Section 3.1, but we are also experimenting with a similar one developed by `CoolEffect`.[6] As long as models are functions of the distance, more can be easily added.

This first pass of the script therefore gives us an estimate of the footprint of our conferences. We have implemented on top of it all the analyses that we described through Section 4, as well as the speculative analysis described in Section **??**. The output of these analyses is encoded into `csv` tables that can be used as-is or as the basis of visualization exercises.

There are room for improvement on pretty much all sides—different footprint models, more complex analyses, and automating the visualization of the data, to cite just a few. But we hope that this preliminary tool will form the basis for fruitful discussion as it grows to address the needs of more research communities.

6 Conclusion

Carbon footprint is becoming a significant consideration for conference organizers. To support effective decision-making, we have conducted an analysis of the participation for several SIGPLAN conferences, drawing both an estimate of their carbon footprint and various correlations between the geographical distribution of its attendees and this footprint.

We believe that the experiment we conducted in this paper should be generalized. To help move toward this goal, as well as to trigger debates over the right way to conduct these analyses, we developed a reusable, open source tool allowing others to easily conduct similar experiments.

Acknowledgements

Michael Hicks and Jens Palsberg were key contributors to early stages of this data analysis exercise. Gregory Bekher wrote the first version of ACM's carbon footprint calculator, on which our current implementation is partly modeled.

References

[MDZP+18] V. Masson-Delmotte, P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P.R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J.B.R. Matthews, Y. Chen, X. Zhou, M.I. Gomis, E. Lonnoy, T. Maycock, M. Tignor, , and T. Waterfield (eds.). *Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty.* Intergovernmental Panel on Climate Change, 2018.

[6] `https://www.cooleffect.org/`

[SIH+05] Robert Sausen, Ivar Isaksen, Didier Hauglustaine, Grewe Volker, David Lee, Gunnar Myhre, Marcus Köhler, Giavanni Pitari, Ulrich Schumann, Frode Stordal, and Christos Zerefos. Aviation radiative forcing in 2000: An update on ipcc (1999). *Meteorologische Zeitschrift*, 14:555–561, 08 2005.