# Abstract

Representativeness and expansion potential of the northern high latitude eddy covariance network.

Representatives, network optimization, high latitude, Arctic, carbon cycle, eddy covariance

# Introduction

Because of the vastness, inaccessibility and extreme climate of the Arctic climate zone, research in this region is a complex endeavor. There are vast stocks of carbon rich permafrost that have accumulated over the ages, which are at increased risk of thawing linked to climate change and its associated Arctic amplification (Serreze & Barry, 2011). With limited insights into current carbon processes, it is complicated to determine trends in Arctic carbon budgets. Therefore, our ability to establish quantitative links between climate change and carbon processes, and to forecast future carbon exchange processes, is severely compromised.

The eddy covariance (EC) method is a widely used method to measure ecosystem scale greenhouse gas fluxes. The method is nondestructive, and allows to continuously monitor fluxes at high temporal resolution. Despite the difficulties listed above, many EC sites that measure these greenhouse gases fluxes have also been established in the Arctic, and most of them are listed in global and regional EC flux networks (e.g. Fluxnet, AmeriFlux, Asiaflux, ICOS) facilitating multi-site syntheses. Unfortunately, there is no such network yet that specifically lists all the sites in the Arctic. Moreover, the metadata in these networks is sometimes outdated or incomplete, and some sites are simply not listed, which makes it difficult for scientists working in this domain to gain a clear overview of all available EC data.

Knowing the current and past spatiotemporal distribution of EC sites is not enough to fully understand to which degree this network represents the Arctic domain. The reason for this is that EC towers have footprints (i.e. the area covered by its measurements) that typically do not extend further than a kilometer from the tower, often less. Accordingly, with about a hundred EC towers in the Arctic only a very small fraction of the region gets directly observed, while most of its expanse remains unsampled. For upscaling purposes, typically a tower is held as representative for the ecosystem and the region where it is stationed; however, except when using a very coarse classification of ecosystem types, the existing EC network still cannot cover all ecosystems across the Arctic. There have been several studies that aim at evaluating the spatial coverage of EC networks (Hoffman, Kumar, Mills, & Hargrove, 2013; Shiga, Michalak, Randolph Kawa, & Engelen, 2013), though so far none focused on the Arctic. Especially the pronounced spatial variability in Arctic ecosystem characteristics across scales make this evaluation more difficult, but at the same time also highly important.

Building on a study by Hoffman et al. (2013) that presented an analysis of the Alaskan EC network, in this study we will provide a first in-depth evaluation of the current and past pan-Arctic EC flux observation infrastructure. This analysis aims at evaluating results from the representativeness analysis of this network regarding its coverage of the pan-Arctic ecosystem distribution. Moreover, we use the results from the representativeness analyses to identify the most suitable locations for new observation sites, and upgrades to existing infrastructure, that would optimally enhance the performance of the Arctic EC network as a whole. Finally, this manuscript and its corresponding online tool aim at providing an easily accessible source of information on Arctic flux monitoring infrastructure for scientists working on the carbon cycle.

# Methods

## Survey on flux site infrastructure

To properly model the extent of the Arctic EC network, a comprehensive inventory is required of all sites within the domain. As a first step towards creating an inventory of the circum-Arctic network of eddy-covariance flux sites, we combined those sites listed within the established flux networks websites: Fluxnet (<https://fluxnet.fluxdata.org>), Ameriflux (<https://ameriflux.lbl.gov/sites/site-search>), the European Fluxes Database Cluster (<http://www.europe-fluxdata.eu/home/sites-list>), ICOS (<https://www.icos-cp.eu>), and Asiaflux (<https://db.cger.nies.go.jp/asiafluxdb>). The initial search was restricted to sites located above 60 degrees Northern latitude. Even though this publicly available information already covered a large part of the final site list, we discovered a few limitations with these datasets: First, in some cases when a site would appear in several databases, metadata was not consistent between them. Second, often some part of the metadata fields was missing, especially detailed information on temporal coverage. Here, generally only start and, if applicable, end times were mentioned, while no information was provided on the seasonal discontinuation of operation that is prevalent at many Arctic sites. Third, we were aware that a considerable number of sites were not listed in any of the flux networks listed above.

To improve this situation, in early 2018 we conducted an online survey among PIs of flux sites in the Arctic and surrounding boreal regions. Besides confirming basic information such as e.g. exact location, contact information, and, where applicable, references that describe site operations in detail, we specifically asked for the following items:

* Detailed times of operation (on a monthly scale), broken up by CO2 and CH4 fluxes
* List of gas species measured
* Details on eddy-covariance instrumentation (e.g. types of sonic anemometer and gas analyzer)
* Details on auxiliary measurements, for example snow depth and precipitation, including power supply
* Mode of data availability (e.g. open, password restricted, upon request).

At the time of writing, we have received 66 responses to our metadata request from site PIs. For all those sites for which new data was provided by PIs, in our final site list we used the more recent information from our survey to replace existing information from the databases. We contacted PIs and flux networks In case of conflicting information.

An overview of the eddy covariance flux network that our list comprises will be given in the results section below. To make this information accessible to the Arctic research community, we created an online mapping tool hosted by the Arctic Data Center of the National Center for Ecological Analysis and Synthesis (https://www.nceas.ucsb.edu/arctic-data-center). This tool, which is accessible at [http://cosima.nceas.ucsb.edu/carbon-flux-sites](http://cosima.nceas.ucsb.edu/carbon-flux-sites/), provides an easy to use web interface that allows the user to identify data availability within certain regions, timeframes, or biome types. The main tool consists of three elements: The central interface holds maps in several layers where the location of the sites is shown, and basic information can be retrieved in popup windows. Furthermore, a panel allows to make selections of sites based on type, location, activity and duration of observations, while a table on the bottom contains detailed information on all selected sites and, if available, direct links to the actual date are provided. Lists of selected sites for a given search can be downloaded as csv files.

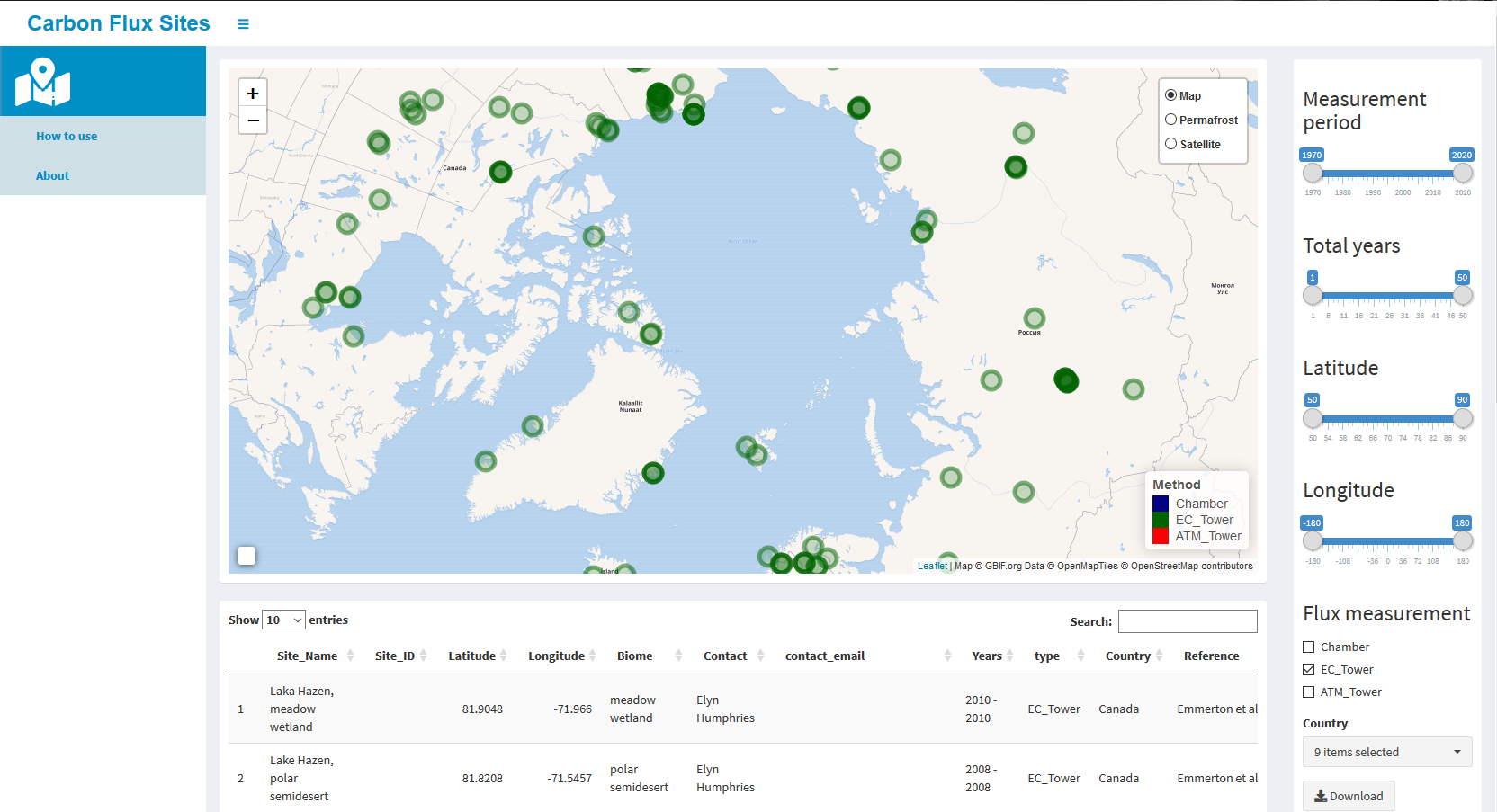


Figure : Screenshot of the online flux site mapping tool that was developed to facilitate access to our pan-Arctic meta-database. Please refer to the main text for detailed information on the functionality.

## representativness ASSESSMENT

A short description of his method, referring mostly to earlier work (e,g their Alaska paper) for validation probably by Jitu as he knows the method best.

We evaluated the representativeness of the EC network in the Arctic based on all sites, active and inactive, summarized in our database, and used four specific subsets of the eddy-covariance network of sites:

1. All active sites: This first set of sites includes those that reflect the current network coverage. Here, we chose those sites that were listed as active at the end of 2017. This subset serves as the starting point for any recommendations on network extension.
2. Long-term active sites: The second subset comprises sites that have been active within at least 5 data years since 1993. Data coverage does not necessarily need to be continuous in this context. We included this subset based on the assumption that multiple years of data allow to account for interannual variability, and therefore provide improved insight into functional relationships between fluxes and environmental conditions.
3. Wintertime network coverage: In this third subset, we selected sites that provide data coverage during the Arctic wintertime. With recent studies demonstrating the importance of wintertime fluxes for year-round flux budgets in the Arctic, information on how well our observational infrastructure can capture these signals across the Arctic domain is crucial.
4. Sites with methane fluxes: The final set consists of all sites that measure methane fluxes. Even though the total carbon release of methane is much lower compared to CO2 fluxes, due to its high global warming potential methane needs to be accounted for when constraining carbon cycle feedbacks with global climate change. This is particularly the case for the large fraction of waterlogged areas throughout the Arctic. Since methane fluxes are far more dependent on micro topography than CO2 fluxes (Peltola et al., 2019), and therefore display an elevated spatial variability, extrapolating methane flux results is associated with elevated uncertainties.

UPGRADES TO OBSERVATIONAL NETWORK

The core question we aim to answer for each of these subsets of sites is how well the existing network is capable of capturing spatio-temporal variability in surface-atmosphere fluxes. One closely related question we set out to answer is what would be the optimal locations of network expansion, in case our analysis reveals substantial gap in network coverage. Since testing each cell and each combination of a number of expansion locations would come at excessive computational costs, we developed the following approach to reduce the number of permutations. First, the domain was clustered into ecoregions.

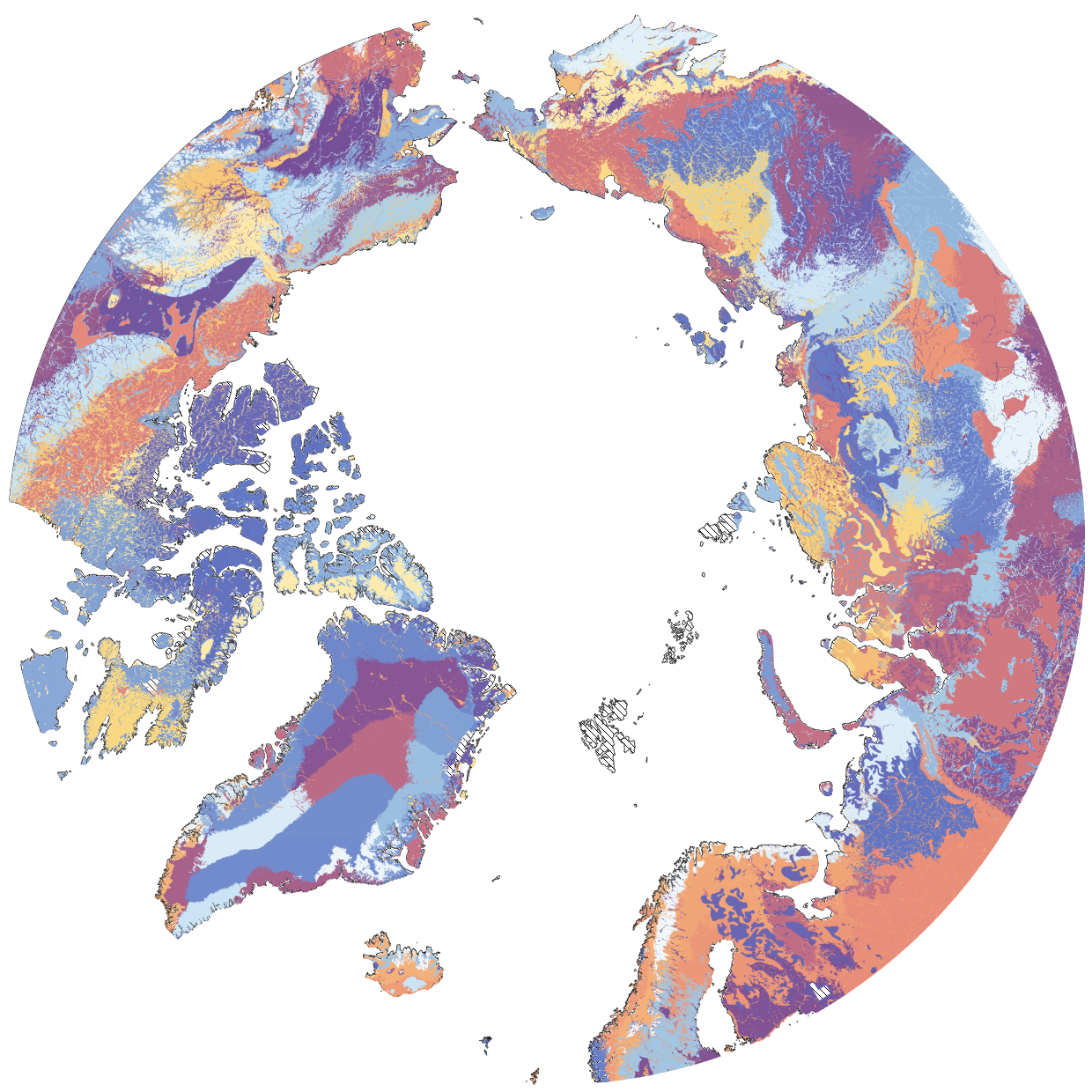
Based on these ecoregions we can then optimize the network by focusing on the least represented ecoregions that had the largest negative impact on the overall network performance, we. We calculated the expected gain in representativeness to the network for each unrepresented ecoregion as follows:

Here *Ei* is the expected gain in representativeness for the ith ecoregion, *Uim* the mean representativeness of the ith unrepresented ecoregion, *Rm* the mean representativeness of all represented regions, i.e. those ecoregions with a site in their domain, and *Uia* the ith unrepresented ecoregions area. Accordingly, *Ei* indicates the potential gain of information to the network as a whole if a new EC site would be placed in this region. As an alternative version (*ES*) to this metric, we also multiplied these *Ei* values by normalized soil organic carbon (FAO GSOC), in order to emphasize those regions with carbon-rich soils that may particularly experience enhanced carbon releases under future climate change.

The Ei metric identifies those ecoregions that would benefit most from a new monitoring site, but it does not find the most suitable location for a new site within this domain. Mathematically, the ideal location would be the centroid of these selected ecoregions, i.e. the one coordinate where local environmental conditions are closest to the mean conditions within the ecoregion. While these locations are already computed as part of defining the ecoregions, these ideal locations would rarely be practicable, since infrastructure availability and logistics pose considerable limitations in the Arctic. Therefore, instead of working with those centroid coordinates we opted to look for villages in these regions, and placed new stations near them as realistic network expansion locations. As a mapping resource, here we used the natural earth populated places dataset (https://www.naturalearthdata.com/downloads/10m-cultural-vectors/10m-populated-places/).

For the data subsets ‘wintertime’ and ‘methane measurements’, we used a slightly different method to identify ideal locations for network expansion. Compared to the cost of building a new site, upgrading an existing one with a methane analyzer, or keeping an existing site active throughout the winter, it is far more economically viable. Therefore, for this step we looked at ecoregions that contain one or more sites, but currently still lack a methane analyzer and/or wintertime activity. Thus, the same metric as above was used switching out unrepresented regions for those that are represented but lack a methane analyzer or wintertime activity. This yielded a list of sites where an upgrade of the existing instrumentation would best increase the respective methane or wintertime representativeness.

Example of 100 ecoregions



# Results

## Arctic flux data

Through merging information from existing databases, and adding details from the online survey among site PIs described above, we identified 121 sites situated within the domain North of 60 degrees latitude. 83 of these sites (69 %) were listed as active at the end of 2017, while the remaining 38 sites had been either permanently or temporarily discontinued at that time. The distribution of these sites across the study domain is uneven, with the majority located in Europe and Alaska (61 % of all active sites), i.e. regions that only account for about 12 % of the total surface area. This imbalanced distribution of sites (see also Figure 2) leaves large regions of the Arctic with comparatively sparse network coverage, particularly regarding Central and Eastern Siberia, and Eastern Canada.

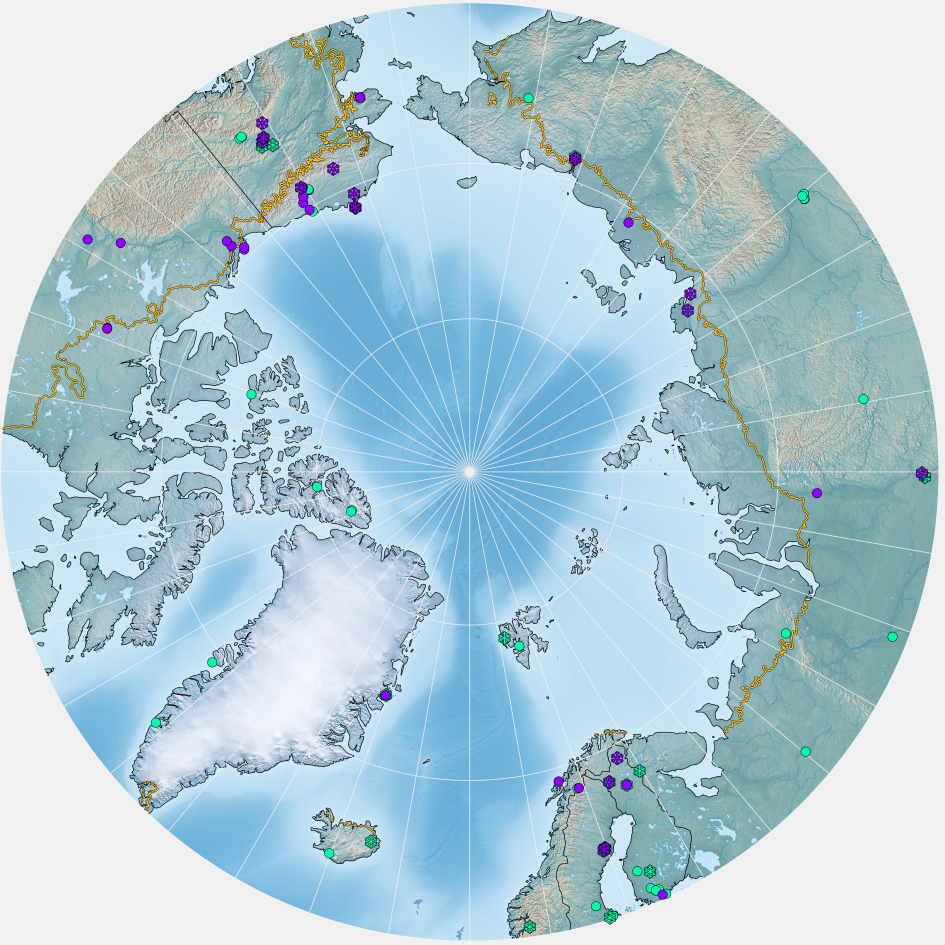


Figure : All sites in the database. Green indicates sites with CO2 fluxes only whereas purple indicates CO2 and CH4 flux measurements. Snowflakes show sites with reported wintertime measurements. The yellow line indicates the Arctic treeline https://www.geobotany.uaf.edu/cavm/data/ (last accessed 28-02-2020), as the background map Natural Earth II (Šavrič, Patterson, & Jenny, 2015).

The number of sites within the Arctic eddy-covariance network has steadily grown since the establishment of the first site in 1993 in Barrow, Alaska. Figure 3 indicates that the installation of new sites gained momentum in the late 1990, and the network steadily grew until reaching its current level of 84 active sites around 2011. Since that time, the size of the network has remained more or less stable, i.e. newly established sites largely balance site shutdowns. Owing to the harsh Arctic climate conditions, wintertime site activity is clearly lagging behind summertime data coverage. In its current state, the Arctic network features only 25 sites with year-round activity, i.e. cold season data coverage today is at about the same level as the summertime measurements were 15 years ago. Moreover, of these wintertime measurements 81% took place in Europe and Alaska, leaving most parts of Canada and Russia with very low data coverage outside the growing seasons.

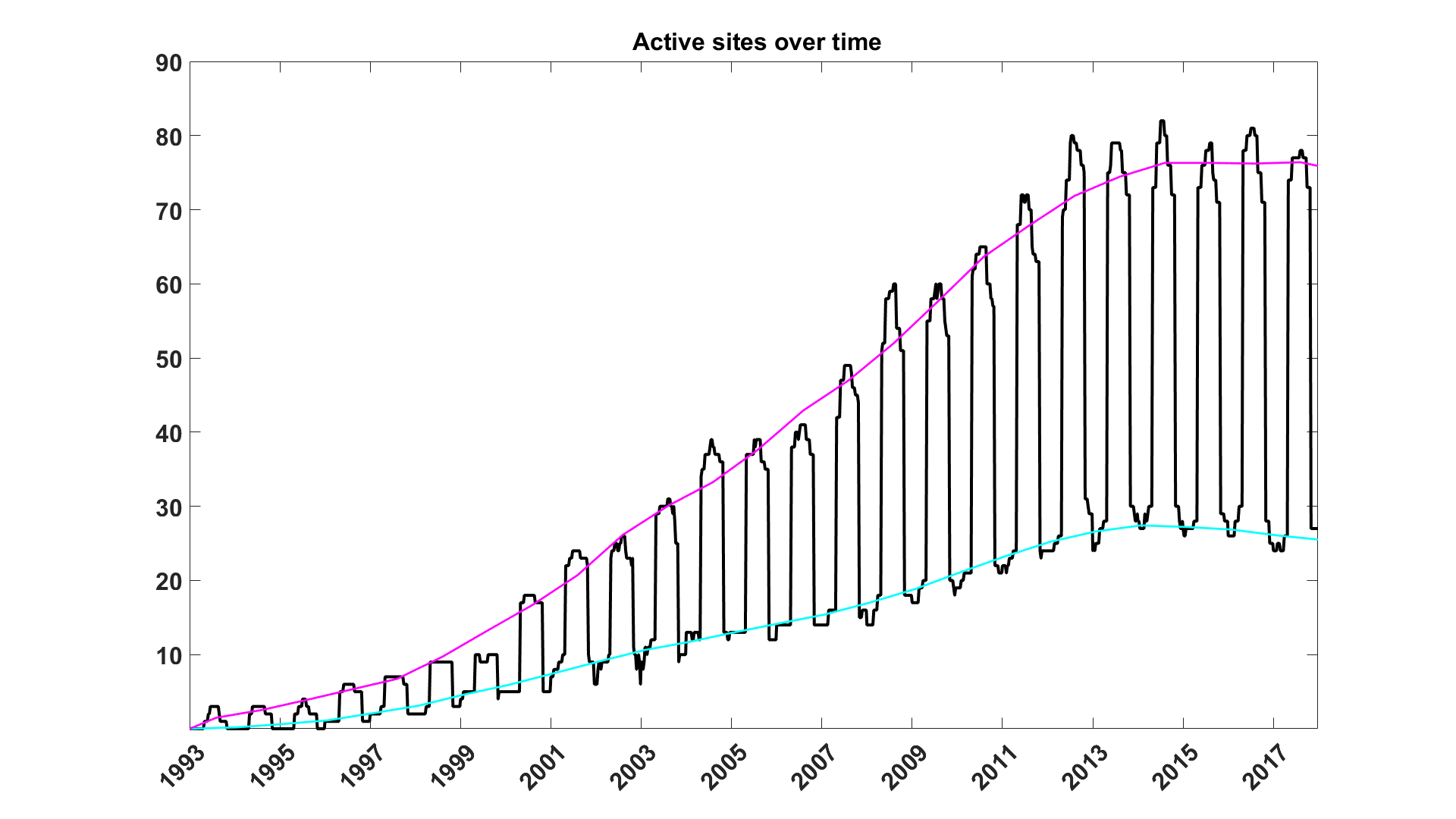


Figure : Development of eddy-covariance network data coverage at monthly timesteps, starting with the establishment of the first Arctic site near Barrow, Alaska. The fluctuating line gives the total number of active sites per month, magenta and cyan lines indicating the long-term development of data coverage during summer and winter, respectively. For sites where only yearly activity data was available, summertime-only data coverage was assumed.

Table : Overview on active and inactive eddy-covariance sites located in the high northern latitudes (>60deg N). Since existing networks in Europe and Alaska are dominating these site lists, they are listed separately to document the imbalanced distribution of observational infrastructure in the Arctic. For wintertime when monthly explicit data was not available, we assumed no activity.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | All sites >60deg N | | Sites in Europe & Alaska (>60deg N) | |
| Site status | Active | Inactive (2018) | Active | Inactive (2018) |
| CO2 | 83 | 38 | 52 | 21 |
| CO2 long term | 73 | 16 |  |  |
| CO2 , wintertime | 33 | 8 | 25 | 6 |
| CH4 | 32 | 13 | 19 | 9 |
| CH4 , wintertime | 16 | 3 | 11 | 3 |

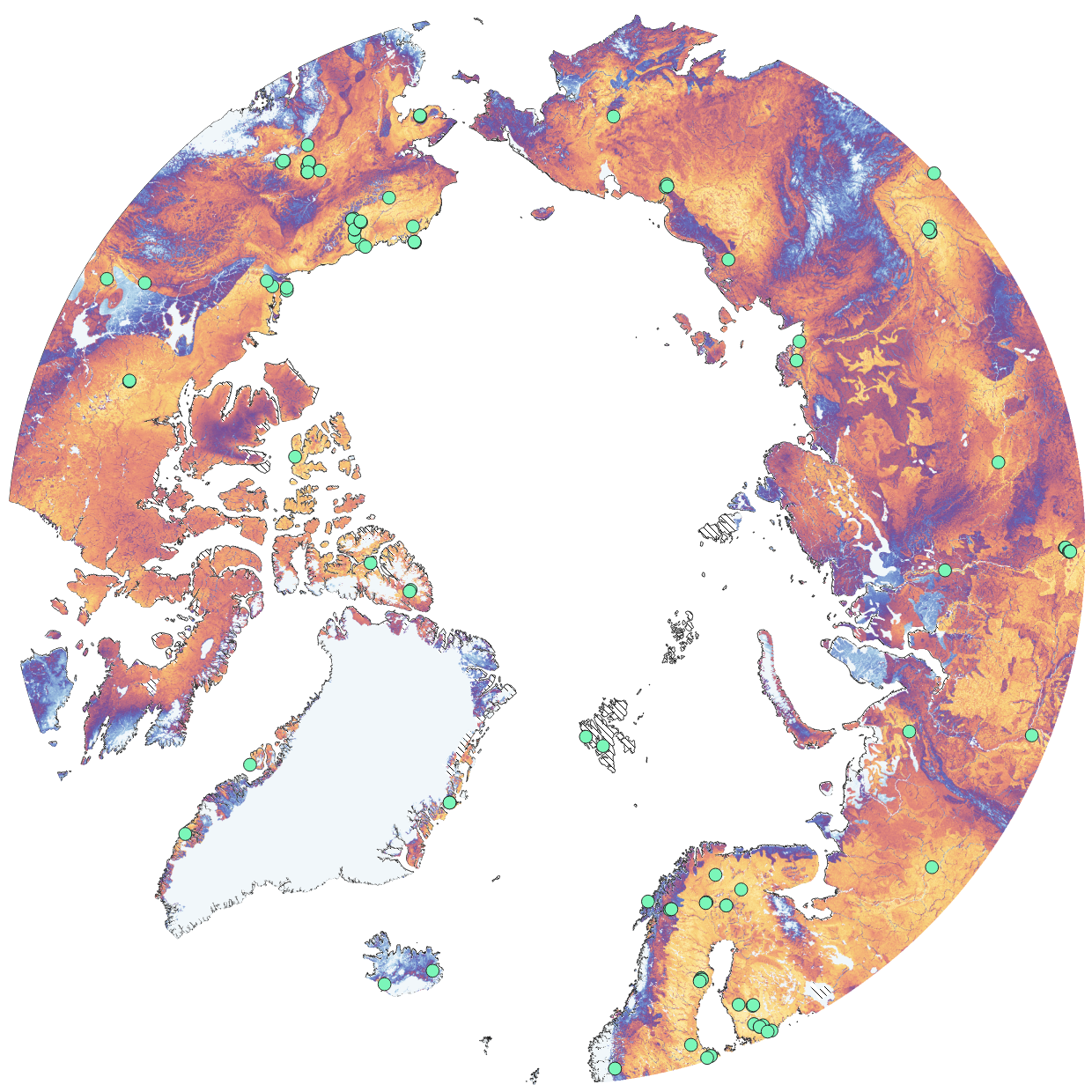
Regarding the length of the time series covered by the eddy flux sites, naturally there is a pronounced variability across the network. The longest running site (Barrow CMDL) has been active for >25 years at the time of writing, and due to the substantial extension of the network in the 2000s, today the median activity among all sites is 8 years. The steady increase in the length of the data records over time implies that a growing number of sites is suitable to monitor the effects of an ongoing climate change at site level across the Arctic, providing data-based evidence how surface-atmosphere exchange processes of carbon and energy react to the altered environmental conditions.

Regarding the measurement of non-CO2 fluxes, only for methane a considerable number of observation sites could be identified that provides longer-term flux data coverage. Even though this network has been growing steadily over the past years owing to the availability of a new generation of gas analyzers, the number of sites at which CH4 fluxes are monitored is lagging far behind the CO2 data coverage: In 2018, only 32 sites were identified, 13 of which (28 %) were inactive. This is similar to the wintertime data coverage: during the cold season, even though the situation is improving there are still large gaps in the network, and data coverage is at about the level the CO2 summertime data featured in the early 2000s. For other non-CO2 gases, such as e.g. N2O, no observational infrastructure could be identified in the context of our data survey.

Data availability is a crucial factor when it comes to the usefulness of the eddy-covariance observations for community-wide research efforts in the context of climate change. PI responses to our survey indicated that the majority of the eddy-covariance datasets is currently available to interested users: 18% of the datasets were reported as open access, and a further 44% will be made available on request. 36% of the datasets comprised in our database is still being processed and/or reviewed by the site PIs, but will be made available in the future. Only a small fraction (2%) is not intended to be shared publicly.

## NETWORK Representativeness

The representativeness analysis of the network shows large areas of Siberia and Canada underrepresented (Figure 4). Moreover, mountain ranges in general experience a low EC flux data coverage across the Arctic. The current network status and the 5 years active networks status differ only slightly in overall coverage and pattern, whereas in comparison wintertime activity and methane measurements show a drastic reduction in network coverage. Dense methane site coverage is mostly limited to the Alaska North Slope, the Fairbanks region, Sweden and Finland. Outside these regions, coverage is sporadic. Regarding wintertime measurements, a similar picture emerges as described for methane fluxes, but here some extra sites in Canada enhance network coverage in this domain.



All active sites yellow good, light blue bad, green dots sites, dashed areas no data (need to make a scalebar by hand)

|  |  |
| --- | --- |
| Active | 5 year active |
| CH4 | Winter |

Fig x 4 ecoregionsshort explanation, not there was no data for Svalbard insome of the imput data.

Some text with the stats:

## NETWORK OPTIMIZATION

Ecoregions identified by their potential impact on network representativeness are shown in table x interested ones for network expansion see table x with the regions number in figure 5 region xyz come up in several of these metrics and are those selected as prime regions for network expansions

With the result of adding their centroid to the network visible in figure 6

Impovements by the method are

for the active network:

winter:

methane:

this method is clearly able to target specific ecoregions for improvement:

with a 24.7 % improvement within targeted ecoregions compared to untargeted regions.

For the winter and methane method these values are lower though still clear differences are visibe:

With 2.4 and 3.6 times higher improvement compared to untargeted regions.

The extension method for active sites performs will not only in improving the network but also in targeting the areas where most gain is expected as there is a clear correlation between expected gain and actual gain (rho 0.87, p 0.008 (n=6 Kendall's Tau Coefficient right tailed) for the winter and methane scenraio’s no significant relation was found.

15% improvement of representativeness compared to regions without a new tower. Which is a clear indicator

# SECTION 4: DISCUSSION

4.1 ARCTIC EC FLUX DATA ASSESSMENT

While the greatest care has been taken in creating the database of Arctic flux sites, it is by its very nature a work in progress. Accordingly, the state of the online database will deviate slightly from the version used in this paper since we continuously work in data updates provided by site PIs. For reference, a version of the database reflecting the state when writing this manuscript has been retained. Since we rely on PI feedback to ensure correctness of the collected information, despite our best efforts gaps and outdated data in the database are possible in cases.

Since we did not receive PI feedback on our own database survey for some sites, we do not have information on site activity in monthly (or seasonal) time steps for the entire Arctic network. For sites where this information is missing, we decided to assume summertime activity only, i.e. no non-growing season flux data is available. This assumption is based on a general workflow where in spring, once sites become accessible again, the equipment is serviced and activated for operation during the growing season, and then is kept running into autumn and winter until instrument failures and/or loss of energy supply terminate data acquisition. As a consequence, the site lists used for the non-growing season represent a conservative picture of the network coverage, and we anticipate to refine this assessment with additional PI entries into our database survey.

In Figure 2, the network growth seems to level off around 2012. However, it is not clear if there is an actual slow-down in growth, or if this is an artefact reflecting delays in updating sites and studies in the online depositories. Since it is not uncommon to restrict data access until first results have been published by the data owners, data availability for the most recent years may in fact be higher than reflected by our database.

4.2 NETWORK REPRESENTATIVENESS ANALYSIS

The method utilized here to indicate the network representativeness is powerful in showing the patterns associated with the networks coverage, as these are based on key ecosystem variables, and especially taking into account Arctic-specific controls such as e.g. permafrost extent. However, the network assessment remains mostly on a qualitative level, since no clear cutoff value between good and bad representation is directly apparent from this method. To add a level further level of quantification we take the represented ecoregions mean *Rm + two times MAD* as an indicator of a region that is well enough represented for current practices and scale the colouring of the images so.

* This paper functions on the premise that an EC site represents its ecosystem as is generally practice when working with EC data, which is often up-scaled to the ecosystem scale. However a paper by (Hill, Chocholek, & Clement, 2017) indicates that to represent an ecosystem with more certainty often more than one EC site needs to be present. This is good to keep in mind with the high density of towers in Europe and Alaska, that even though some might not sample a new or unique region they are certainly valuable in their ability to improve quality and certainty of the flux budget in these regions.
* The use of multiple towers might be even more important when we look at methane fluxes which are very dependent on local topography (Peltola et al., 2019) and especially water levels. Ideally for each tower an analysis would be done of how representative the tower is for their listed ecosystem. As the complexity of a landscape increase so would the importance of an analysis like this. Where a boreal forest might be quite similar over large areas, the polygon nature of some Arctic tundras make it exceptionally difficult to arrive at one value for the entire region with just one tower. Mobile towers or the qualitatively lower towers mentioned in Hill, et all would be a solution to this. Therefore in this study the representatives of the methane scenario should be considered a best case scenario with actual representation lower.
* Discuss selection of data layers for Jitu’s method – is the selection comprehensive, how well do data layers reflect the key factors that will most likely affect flux variability across the Arctic? Also applies to the network upgrade discussion below.

4.3 NETWORK UPGRADES

Concerning the definition of ecoregions, our choice to separate the Arctic study domain into k=100 ecoregions is based on the following considerations: First, for a smaller number, ecoregions would become excessively large, and also less and less homogeneous, therefore they would hardly represent coherent units. Second, separating the domain into much larger numbers would result in ecoregions so small they would not grant much improvements over using the raw distance. Deviating too far from an optimum in both directions would compromise the network optimization, since either no real gaps in spatial representativeness could be detected, or the gaps would become so small that placing a new site there would not affect the overall network performance substantially. Accordingly, our choice of k=100 reflects a compromise that we found to agree well with our study objectives.

* Statistics: depends on the method that we’ll use in the end. If you manage to evaluate all possible combinations of 5 sites, then there’s little to discuss. If you simplify, it needs to be mentioned
* Different selection methods for wintertime/methane sites: what effect on optimization performance?

# Conclusion

current state of network

Overall we see the network of towers in the Arctic grow to a cover an impressive area despite the difficulty of these operations. Main gaps exists high latitude regions of Russia the Canadian peninsula. While also most mountainous regions are not sampled , even in regions with otherwise high coverage (e.g. the Norwegian mountains and Alaskan mountain ranges). This is a logical result of low expected fluxes because the extreme climate caused by high latitude and high elevation and at the same time the increased difficulties of operating a site in such locations. Similarly, we see large gaps in wintertime measurements again lower expected fluxes and increased difficulty are the main reasons to forgo such measurements.

Strategy:

keep existing towers running, it requires multiple years to say anything about the climate and many more to make informed statements about any changes in climate. After that methane measurements are a relative affordable extension to any site to capture an important greenhouse gas. This counts even more so in the Arctic were the complex wetland, permafrost, and thawing dynamics make any prediction about the release of carbon from the permafrost and in which form CO2 or CH4 difficult. Keeping sites active during the winter has been shown to capture a relatively high part of the sites carbon budget. Even if 100% uptime cant be guaranteed having enough snapshots of the winter can make it possible to capture the winter trends. As the we look for regions to expand the network to these and these regions show the greatest promise. And again there are no N2O fluxes which is a great los as its both a strong greenhouse gas as a powerful indicator of biological processes.

Acknowledgement: flux sites all colleagues that filled in the survey, Intaros. IMPRS

# Appendix

**Bits and pieces**: appendix

e.g. IRGA type, power type, maybe correlated to latitude/winter measurements/methane measurements.

For full site list see online tool (which might be a work in progress at that time but should probably be online)

Tool description.