High-latitude eddy covariance temporal network design and optimization.

Martijn Pallandt[1], Martin Jung[2], Kyle A Arndt[3], Susan M. Natali[3], Brendan Rogers[3], Anna- Maria Virkkala[4], and Mathias Göckede[5]

[1]Max Planck Institute for Biogeochemistry
[2]Max-Planck-Institute for Biogeochemistry
[3]Woodwell Climate Research Center
[4]Woodwell Climate Research Center, Falmouth, MA, USA
[5]Max-Planck Institute for Biogeochemistry

October 17, 2023

Abstract

Ecosystems at high latitudes are under increasing stress from climate change. To understand changes in carbon fluxes, in situ measurements from eddy covariance networks are needed. However, there are large spatiotemporal gaps in the high-latitude eddy covariance network. Here we used the relative extrapolation error index in machine learning-based upscaled gross primary production as a measure of network representativeness and as the basis for a network optimization. We show that the relative extrapolation error index has steadily decreased from 2001 to 2020, suggesting diminishing upscaling errors. In experiments where we limit site activity by either setting a maximum duration or by ending measurements at a fixed time those errors increase significantly, in some cases setting the network status back more than a decade. Our experiments also show that with equal site activity across different theoretical network setups, a more spread out design with shorter-term measurements functions better in terms of larger-scale representativeness than a network with fewer long-term towers. We developed a method to select optimized site additions for a network extension, which blends an objective modeling approach with expert knowledge. Using a case study in the Canadian Arctic we show several optimization scenarios and compare these to a random site selection among reasonable choices. This method greatly outperforms an unguided network extension and can compensate for suboptimal human choices. Overall, it is important to keep sites active and where possible make the extra investment to survey new strategic locations.

Hosted file

975437_0_art_file_11438407_s1r2sd.docx available at https://authorea.com/users/670634/articles/670512-high-latitude-eddy-covariance-temporal-network-design-and-optimization

Hosted file

975437_0_supp_11438408_s1r2sd.docx available at https://authorea.com/users/670634/articles/670512-high-latitude-eddy-covariance-temporal-network-design-and-optimization

High-latitude eddy covariance temporal network design and optimization.

Martijn M.T.A. Pallandt[1], Martin Jung[2], Kyle Arndt[3], Susan Natali[3], Brendan M. Rogers[3], Anna-Maria Virkkala[3], Mathias Göckede[1]

[1]Department of Biogeochemical Signals, Max Planck Institute for Biogeochemistry, Jena, 07745, GER

[2]Department of Biogeochemical Integration, Max Planck Institute for Biogeochemistry, Jena, 07745, GER

[3]Woodwell Climate Research Center, Falmouth, MA 02540, USA

Corresponding author: Martijn Pallandt (mpall@bgc-jena.mpg.de)

Key Points:

- The network of high-latitude eddy covariance sites has grown considerably over time, still towers should remain active and new remote locations added.
- Without new measurements our knowledge will degrade at a rate at least equal to which would otherwise be its growth.
- Network optimization methods as shown here are essential for representative network design.

Abstract

Ecosystems at high latitudes are under increasing stress from climate change. To understand changes in carbon fluxes, in situ measurements from eddy covariance networks are needed. However, there are large spatiotemporal gaps in the high-latitude eddy covariance network. Here we used the relative extrapolation error index in machine learning-based upscaled gross primary production as a measure of network representativeness and as the basis for a network optimization. We show that the relative extrapolation error index has steadily decreased from 2001 to 2020, suggesting diminishing upscaling errors. In experiments where we limit site activity by either setting a maximum duration or by ending measurements at a fixed time those errors increase significantly, in some cases setting the network status back more than a decade. Our experiments also show that with equal site activity across different theoretical network setups, a more spread out design with shorter-term measurements functions better in terms of larger-scale representativeness than a network with fewer long-term towers. We developed a method to select optimized site additions for a network extension, which blends an objective modeling approach with expert knowledge. Using a case study in the Canadian Arctic we show several optimization scenarios and compare these to a random site selection among reasonable choices. This method greatly outperforms an unguided network extension and can compensate for suboptimal human choices. Overall, it is important to keep sites active and where possible make the extra investment to survey new strategic locations.

1 Introduction

The Arctic and boreal biomes have been recognized as a domain that is changing rapidly as a result of climate change (Serreze & Barry, 2011; IPCC, 2014; Meredith et al., 2019). These changes may lead to strong positive feedbacks with ongoing climate change, since large stocks of carbon sequestered in soils may become unstable as permafrost thaws (Gustaf Hugelius et al., 2020; E. A. G. Schuur et al., 2015; Edward A. G. Schuur et al., 2008; Serreze & Barry, 2011). For reliable forecasts of future global climate, it is of vital importance to monitor the carbon cycle in these regions and understand the mechanisms that govern it.

Eddy covariance (EC) is a key technique to investigate the carbon cycle. With this method, fluxes of greenhouse gasses (GHG), predominantly carbon dioxide (CO_2) and methane (CH_4, and energy are measured continuously at high temporal resolution above the canopy to quantify their rate of exchange between the atmosphere and biosphere (Baldocchi, 2003; Pastorello et al., 2020; Sulkava et al., 2011). The typical field of view, or footprint area, for EC towers found in most parts of the Arctic is relatively small, usually on the scale of hundreds of meters, (Göckede et al., 2004; Kljun et al., 2002; Rannik et al., 2000; Schmid, 1997; Vesala et al., 2008). To obtain regional carbon budgets, these local measurements need to be upscaled to much larger domains. There are varied methods to upscale fluxes, which have greatly improved over the years (Byrne et al., 2023; Chu et al., 2021; Desai, 2010; Jung et al., 2011; Xiao et al., 2012) with some specifically targeting the Arctic (Birch et al., 2021; Ito et al., 2023; Peltola et al., 2019; Virkkala et al., 2021). Of these methods, machine learning techniques are becoming increasingly important. Still, no matter how advanced the methods, the fluxes used either as input or reference should cover the relevant range of conditions and ecosystem types; otherwise prediction accuracy can neither be guaranteed nor properly assessed. Therefore, location and coverage of the EC towers should be carefully considered in any upscaling endeavor.

Typically, EC towers have been placed to answer specific research questions, while the role of a given tower in the larger observational network plays a minor role in decision making and funding. Moreover, site selection is often strongly constrained by logistical considerations and available infrastructure. This has led to a site distribution in the Arctic that greatly favors Alaska and Europe, often at locations with access to electricity, leaving large areas of northern Canada and Siberia undersampled (Pallandt et al., 2022). When evaluating tower infrastructure for wintertime CO_2 fluxes, or CH_4 fluxes, we see even larger gaps across these regions, with wintertime representativeness values 74% worse and CH_4 48% worse than the summertime CO_2 measurements (Pallandt et al., 2022). The establishment of a long time series of flux measurements is another major challenge: Typically, funding for EC towers is provided on a project basis, which typically guarantees funding only for a couple of years. Researchers cobble together grants to keep towers active for longer, though this is not an ideal basis for a stable monitoring network. Research Infrastructures like ICOS and NEON aim to alleviate this problem by advocating for long-term data collection and flux data standardization, however these are only active in Europe and the USA respectively, and even there not all EC towers fall under their umbrella. Overall, the future of most EC sites is highly uncertain.

Several studies have investigated the representativeness of EC networks, and in some cases, virtually extended these networks by including mechanics to optimize the spatial distribution of the network in case of potential future extension (Chu et al., 2021; Hoffman et al., 2013; Pallandt et al., 2022; Sulkava et al., 2011; Villarreal & Vargas, 2021). Still, no studies have investigated the representativeness of the EC network in relation to long-term temporal data coverage. Pallandt et al. (2022) looked at the differences between the winter- and summertime network representativeness, though only in terms of differences in the spatial component. Still temporal changes are important for the EC network. The longer a monitoring network remains active and expands, the more data it will accumulate, which in turn increases its capabilities to interpolate within its dataspace or extrapolate beyond it (Banko & Brill, 2001; Bosveld & Beljaars, 2001; Loescher et al., 2006; Wisz et al., 2008), though as climate changes, we are entering non-analog climate conditions which past towers may not fully represent. It remains to be quantified how the growing coverage period of an existing network, associated with more accumulated data over time for the same subset of sites, changes our ability to upscale fluxes. This information is crucial to guide us in maintaining and upgrading the network with increased efficiency.

In this paper, we aim to quantify the EC network representativeness potential for upscaling flux data to a larger domain, in relation to temporal factors. As a starting point for our analysis, we update the existing high-latitude EC meta-database used in Pallandt et al. (2022) through further evaluation of meta-data and an updated survey. We then extend the extrapolation index metric first shown in Jung et al. (2020) by including an optimization scheme to investigate network growth and expansion. We use these methods to investigate how choices in the temporal arrangement of the network can affect its representativeness. We do this through several experiments that each test a specific temporal aspect of the network's design and functioning: termination of measurements, limitation of site activity to a few seasons and the tradeoff between few long term and many shorter measurements. Finally we demonstrate a practical application of these techniques in a case study where we combine modeled optimization with expert knowledge in an actual potential network extension.

2 Methods

2.1 Network status

To update our database on high-latitude EC towers to reflect the current status up to 2022, we updated the survey conducted by Pallandt et al. (2022) in 2017 and added more specific questions about a given site's biome, planned future activity and future funding as well as extending the site activity table to 2022. The survey was distributed among the FLUXNET newsletter members and known PIs of high-latitude EC sites. Counting direct correspondence to the survey as well as submissions to the online form we received 37 replies. Combined with our previous results, we now have temporally explicit information for 88 sites from 1993 when the first towers in the Arctic were erected, though not all cover the period from 2018 to 2022. Combining these further with online sources such as the flux databases (e.g., AmeriFlux, AsiaFlux, Fluxnet, ICOS, NEON), personal communication, and collaborating database projects (ABCflux, Virkkala et al. (2022)) we added or updated information on a total of 145 EC sites in comparison to the previous database version. This database is available at the high-latitude carbon flux tool: https://cosima.nceas.ucsb.edu/carbon-flux-sites/, which, besides metadata on EC flux sites, also lists metadata on flux chambers and atmospheric towers.

While Pallandt et al. (2022) limited the study domain to areas above 60 degrees North, in this study we opted for a more natural southern border that follows the extent of Tundra and Boreal biomes (58 ecoregions) as defined by Dinerstein et al. (2017), which is an update of (Olson et al., 2001); details on the domains can be found in figure S2 and table S2.1. By setting the cutoff of the domain based on bioclimatic conditions, we reduce the risk of excluding sites – especially near domain borders – that would be relevant to our representativeness assessment. And through the inclusion of these ecoregions, we can more specifically target and describe regions of interest throughout this work.

2.2 Extrapolation error

The extrapolation error index (EI) metric aims to quantify the relative increase in upscaled flux error as a function of increased distance (in predictor variable space) to the nearest flux measurements used for training, it is conceptually very similar to the Dissimilarity Index from Meyer & Pebesma (2021). For details on the EI method please refer to supplement S2 of Jung et al. (2020), while a short summary follows here for the reader's convenience.

The procedure of estimating EI consists of two steps: 1) Estimating the distance in predictor space between a predicted data point to the nearest training data points, and 2) estimating how the prediction error increases with distance from training data to yield a normalization of this distance. In the first step, weights for predictors variables (to account for different variable importances) and the considered number of nearest training data points is established by an optimization algorithm.

The predictor data space is a set of variables representing the conditions observed at the EC sites, which, in our case, are the nine predictors in the FLUXCOM-RS upscaling model ensemble (Jung et al. 2020, Tramontana et al. 2016, Table 1). The target variable is GPP from FLUXCOM-RS extracted at the locations of available EC sites. The entireThis process of training the model and calculating EI values is repeated 7 times in an ensemble to make the results more robust. Three separate training runs have beenbene performed: one for the temporal experiments, one for optimization runs and one for the comparison with previous work (S5).

Table 1: variables used in the calculation of the EI. All variables are provided in a global grid at 0.0833 degrees spatial resolution. Unless otherwise stated, the temporal range is from 2001 to 2020 with monthly steps, others are either static or a climatology of 12 months. All predictor variables are available at the Max Planck Institute for Biogeochemistry Data Portal file id 260. For a description of the quality flags and gap filling approaches used see (Jung et al., 2020)

Variable	Original Source/ MODIS ID	Temporal resolution

Nadir Reflectance Band7	MCD43B4.006.v4_201905	static
Enhanced Vegetation Index	MOD13A2	static
Day Time Land Surface Temperature	MOD11A2	monthly
Night Time Land Surface Temperature	MOD11A2	monthly
Maximum Day Time Land Surface Temperature	MOD11A2	static
Land Cover Data +C4 fraction croplands	MCD12Q1	static
Fraction of photosynthetically active radiation	MOD15A2	climatology
Normalized Difference Vegetation Index * Rg	MOD13A2	monthly
NDWI Normalized difference water index	MCD43A4	monthly
Gross Primary Productivity - RS	Jung et al. (2020)	monthly

2.3 Temporal effects

We performed several experiments to assess the effect of variations in site activity on the network's EI (see table S3 for an overview). For means of comparison, we first established a base scenario (**Baseline**), which formed the network setup against which all other runs were compared. It represents the full EC network as it has grown from 2001 to 2020, with extra data added to those available previously, just as the dataset of measurements increases over time. Site activity was assigned in several steps: years of sites for which we have explicit monthly activity status required no further steps, while years of sites with known wintertime activity were assumed to be active throughout all the months within the year. For the remaining sites years, we assumed “summertime-only” activity, with data coverage restricted to the months. The following setups differ from Baseline only in aspects listed below.

As a first scenario, to gauge how the network would be affected in the hypothetical absence of new measurements, we performed the **End10** and **End15** runs. These runs progressed exactly as the baseline case, except all measurements were terminated at the start of 2010 and 2015, respectively. From these points onwards the extrapolation could only utilize past data. This experiment not only reflects potential gaps in data acquisition or even termination of sites but could also serve as a measure of the trajectory of uncertainties as temporal distance to the last measurement increases when extrapolating into the future.

In the second scenario, we assessed the effect of limiting site activity to quantify how much the EI increases if sites would only be active for a limited duration, e.g., in the framework of a typical research project. These runs are called **MaxX** where X reflects the maximum number of months sites were allowed to remain active. In these scenarios, for each site we tracked their activity and quit any sampling after the allotted number of active months was completed. Here the **Max12** run represents a full year of measurements, while **Max18** corresponds to three years of summertime measurements. Finally, **Max36** represents three full years of measurements. The three year mark was chosen in correspondence with the survey which indicated this project duration as a period for which most sites had funding.

As a third scenario, we investigated the relative impact of site month distribution over the network, where one site month represents one site being active for one month. For this purpose, we compared the performance of a network with fewer sites with long activity (depth) to that of a network with many sites with shorter activity (breadth). In both cases, the amount of data supplied for the analysis (i.e. the total number of site months considered) was uniform. These depth versus breadth runs **DvB10**, **DvB15** and **DvB20** were based on the networks' total site months in 2010, 2015 and 2020, respectively. The number of sites ranges from 55 (largest depth) to 127 (largest breadth), modifying site number in steps of 12 in between. To keep site months consistent among each of the setups, we had to adjust actual site activity. For example, in the case of a network with 55 sites, all of these sites would typically be active all the time year round. In order to keep a realistic distribution of site activity under these conditions, we developed a pseudorandom data month distribution among the existing site locations as explained in supplement 1.

2.4 Network optimization

To allow the use of our network evaluation tool for the purpose of strategic observation network expansion, we added routines that allow for the optimized addition of sites to an existing network. We test 3 methods here, in all cases starting with the baseline of the current network. In order not to confuse this with the 20-year baseline runs from the temporal effects section, we name this baseline EI_ref. It represents the EI calculated for the network in its 2022 state based on the monthly climatology used in comparison to previous work. Three optimization methods were tested, we eventually used a greedy optimization method which evaluates the EI for all potential candidate sites individually. The algorithm then selects the one site which generates the lowest mean EI over the domain and adds it to the existing network. After updating the baseline for the extended network, the same steps are repeated sequentially, adding one site at a time until none are left in the list of candidate sites. This method is fast, but the independent step-by-step additions cannot guarantee that the optimal site combination is chosen for more than 1 additional site; however, the other two methods (S4) are too computationally expensive to optimize for more than 7 site additions, and this greedy method resulted in the same site selection where we were able to compare.

This method only considers a site's EI impact, though often there are many more considerations that play a role in site selection such as logistic feasibility, a site's history, other research demands etc. Many of these requirements are hard to quantify, and even if quantified, weighing them would be fairly objective making a numerical approach undesirable for these extra considerations. This where an expert would come in such as the PI, they could for example decide between similar sites in regard to network improvement which additional requirements would be a deciding factor in choosing a new location. To facilitate this process, we added further metrics that aid the expert to make informed decisions, where if less than ideal sites are chosen site similarity and loss of improvement can be considered. We compute the similarity between sites as the Euclidean distance between all sites based on the local summertime predictor values. To make the distance metric more intuitive, clusters are created based on these distances following the wards method of hierarchical clustering (Ward, 1963), in which we choose a cutoff that results in 5 clusters that roughly represent a north-south gradient. This information is then combined with the EI metric to show optimal sites and all subsequent less than ideal sites in plots such as figure 5 to create a comprehensive view of all options. In subsequent model runs the preselected sites can be added which the model will then take into account.

2.5 Regional case study for network optimization

In a case study, we used optimization methods described above to guide the improvement of the high-latitude EC network within Canada. As an additional goal, this extension was aiming at the establishment of a north-to-south transect of EC sites that would characterize the transition of forests in warmer climates to the wetlands and treeless tundra in the colder climates. As a first step, a selection of potential sites was made based on proximity to populated places within the target region, and sites in our database that were no longer active. This resulted in a list of 28 potential new sites (listed in Fig. 6 and table S2.2). The EI_ref run showed the EI of the domain based on the network's EC site activity in 2022 , which is used as the basis for further optimizations. Several optimization runs were then performed to gain a better idea of the impact of site selection:

- Free search: This approach considered all potential sites in Canada.
- Fixed search: Using the same subset as the free search, the Iqaluit, Churchill Fen and Reservoir site were selected before starting a ‘free’ optimization. Iqaluit was included at the start because it had the highest positive impact on the EI and it is logistically optimally located. The latter two sites were selected here because their inclusion had been predetermined for other reasons unrelated to network optimization.
- Free exclude search: this approach was similar to ‘free search’ run, except seven sites were excluded prior to network optimization. The Mackenzie river region is fairly well represented thus we focus on Eastern canada in this case. And we removed sites that after further investigation currently lacked the right infrastructure for EC towers.

As a benchmark of the optimization, we evaluated a random allocation instead of an optimized one. For each number of site additions (n = [1,28]), 1000 random site combinations are tested. The highest, mean and lowest domain wide EI means of these randomized trials were calculated. In cases where there were less than 1000 combinations possible ($2 \geq n \geq 26$), we evaluated all combinations.

3 Results

3.1 Network status

The network of high-latitude EC sites has grown significantly over the past 29 years (1993 to 2022) to a total of 213 EC sites being active at least periodically within the boreal and Arctic domain. Of these sites, 119 were active in 2022, and 44 of these remain active throughout the winter months (Figure 1). Sixty-six out of 213 sites feature methane measurements, but only 45 of these sites are active. By the end of 2022 the network has accumulated a total of 15048 site months (Figure 1) assuming unspecified monthly or wintertime activity means they are only active during the summer months. Regarding funding and planned future activity, of the 22 respondents that answered this question in our latest survey, 59% indicated they plan to remain active for 5 years or more, and when only considering sites that are currently active this increased to 76%. When asked how long their funding lasts, PIs that planned to keep their sites active for 5 years or more had funding secured for a mean of 3.1 years.

We evaluated the growth of the network in detail from 2001 to 2020. In this period, the summer activity increased from 34 sites in 2001 to 123 in 2020, while winter site activity underwent a larger relative change from 7 sites to 44 on average. Over these 20 years, on average sites were kept active for 63 months (~5.3 years), with 17 sites active throughout the entire 20-year period though these sites typically have wintertime shutdowns.

For each of the 240 months from 2001 to 2020, we calculated the EI based on the cumulative collected data up until that point. The yearly mean EI dropped from 3.0 to 1.2 (Figure S4.a, Video S6), indicating the mean extrapolation error more than halved during this period. Domain wide pixel based minimum values decreased substantially from 0.12 to 0.0. Maximum values have mostly stayed at a high level, dropping from 20.3 to 15.7 in the first four years and then to 14.7 in the subsequent 16 years. This indicates that while the extension of the network and the longer time series improved our capability for upscaling in most regions, only minor improvements were obtained in some of the most remote or extreme locations.

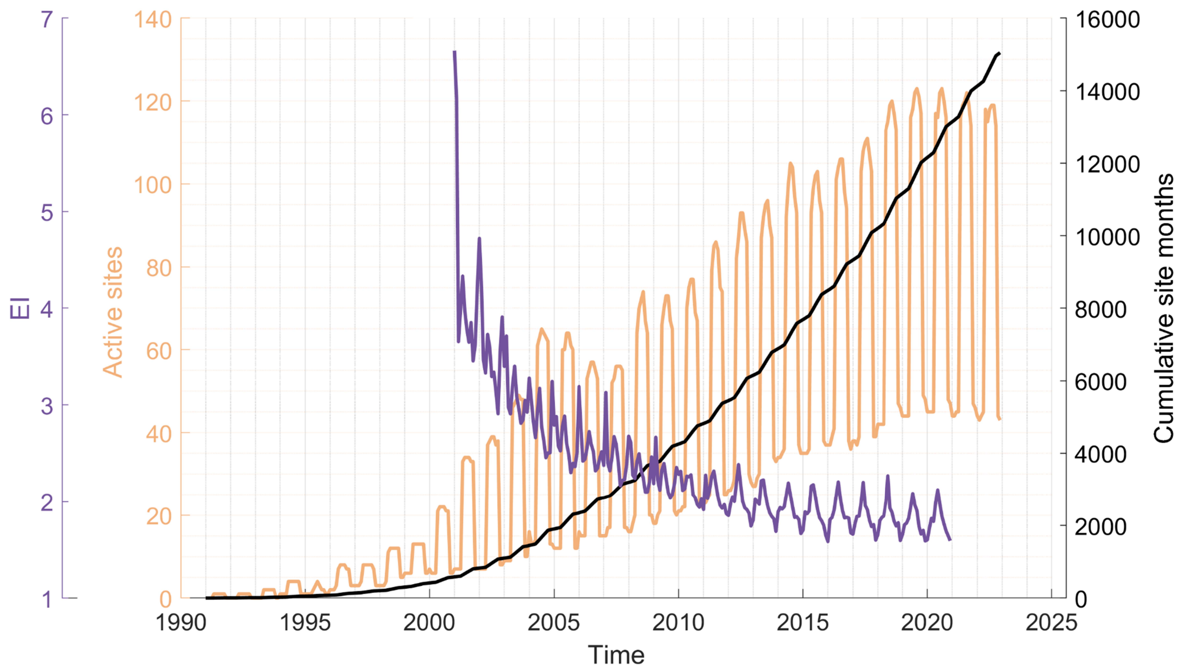

Figure 1. Network growth over time. EI in purple on the leftmost axis shows the mean domain wide extrapolation index per month from 2001 to 2020 which is the timeframe for which FLUXCOM predictor data is available. Active sites in yellow on the left axis graph the site activity per month, a clear annual pattern is visible between winter and summer site activity. Cumulative site months are shown in black, with scale on the right axis in the same color.

The EI, and changes therein over time, are not uniform throughout the domain. In the final year of our assessment, the worst and best represented ecoregion, respectively, are both located in Canada: the *Muskwa-Slave Lake taiga* ecoregion had an average EI of 0.92, whereas the *Canadian High Arctic tundra* had the highest EI rating at 1.99. The greatest improvement over 20 years was observed in the Russian Arctic desert, where EI was reduced by 3.27, i.e., from 5.04 to 1.77. Over the same period, in the *Midwest Canadian Shield forests* an improvement of only 0.60 is detected.

Of note is that there is a difference in summer- and wintertime EI. From 2001 until 2004, summertime EI was lower than wintertime EI, from 2004 until 2011 we calculated similar values for both, whereas from 2011 and onward the situation reversed, and wintertime EI values were lower than summertime EI. Overall, the expansion of the network has resulted in an improved representativeness in all regions and for all seasons. And even though with an ever-growing network improving it becomes more challenging to find new high-impact locations, there is still room for expansion, particularly if new sites are being placed strategically. In response to these observed differences, we investigated differences in spatial and temporal variation between Winter and Summer (table 2). Two of the four temporally explicit variables, NDVIRg and NDWI, as well as GPP, show considerably lower means and standard deviations for winter conditions as compared to summer. Meaning that in wintertime the domain is more spatially homogeneous, and thus with lower variation in the predictors less observation points are required.

Table 1. Median and standard deviations of yearly and monthly explicit predictor variables data over the entire domain. Summer is defined as April through September whereas the remaining months are assigned as Winter. The Overall columns list statistics for all data. The Spatial columns list the mean standard deviation for each time step over the entire domain. The Temporal columns list the mean standard deviation for each location over all time steps. GPP, NDVIRg and NDWI, clearly show in all cases smaller mean and std in winter compared to summer.

	Overall				**Spatial**		**Temporal**	
Variable	Winter mean	Summer mean	Winter std	Summer std	Winter std	Summer std	Winter std	Summer std
Gross Primary Productivity	0.07	2.26	0.14	2.15	0.09	1.48	0.08	1.50
Day Time Land Surface Temperature	253	283	11.1	11.4	8.26	6.80	8.22	9.57
Night Time Land Surface Temperature	250	274	10.2	9.46	7.79	5.05	7.41	8.26
Normalized Difference Vegetation Index * Rg	0.30	6.44	0.78	5.22	0.57	3.69	0.46	3.92
NDWI Normalized difference water index	0.23	0.08	0.10	0.19	0.08	0.13	0.08	0.17

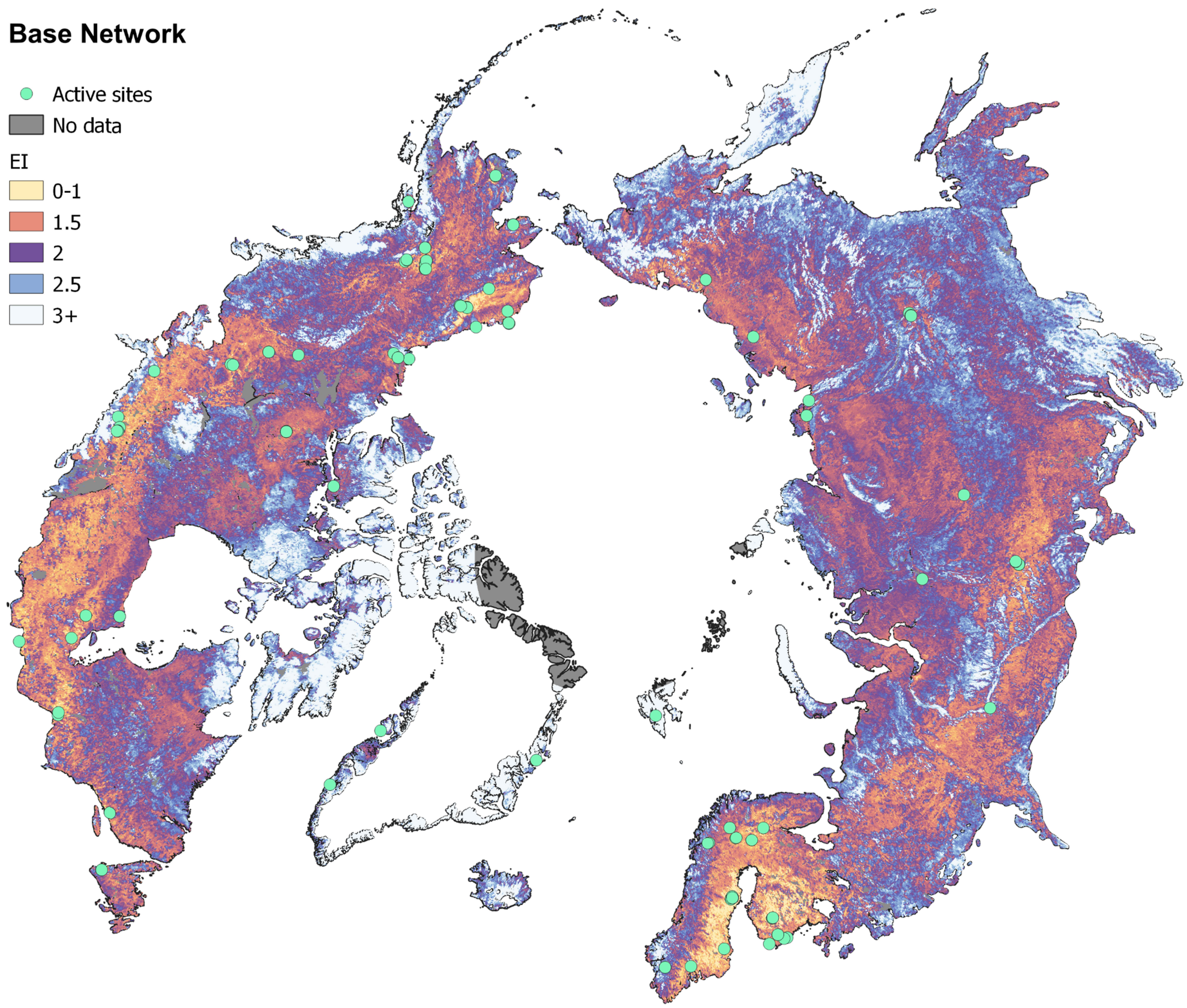

Figure 2. EI of the network in its 2022 state based on one year climatology. Yellow indicates a low EI thus low errors indicating better representativeness whereas colder colors (blue to white) indicate underrepresented areas with high EI ratings. Green dots show the locations of EC sites and grayed out areas indicate no data regions from 80 degrees North and up.

3.2 Temporal effects

In the End10 and End15 scenarios, we investigated how a hypothetical termination of measurements (after in 2010 and 2015 respectively) would affect our capability of upscaling fluxes in the domain (Figure S4 b-c). Compared to the baseline scenario, in the End10 scenario the EI increased on average by 0.005 per year ($p > 0.01$ $n = 11$) for a total increase of EI by 0.08 in 2020. For the End15 scenario, we observed a 0.011 ($p > 0.01$ $n=6$) increase in EI per year for a total increase of EI by 0.07 in 2020. This increase in EI in the End10 scenario is approximately 4 times smaller than the decrease of the baseline in this same period at 0.022 per year, while in the End15 scenario the decrease in the baseline is at the same level as the increase at a 0.011 per year for the End15 scenario. When restricting this evaluation to wintertime, no measurable effect is detected. Regarding spatial variability, these changes are not uniformly distributed throughout the domain: for example, the highest rate of change can be found in the *Yamal-GydanTundra* where a yearly EI increase of 0.019 was observed from 2010, while many other regions did not show any difference at all.

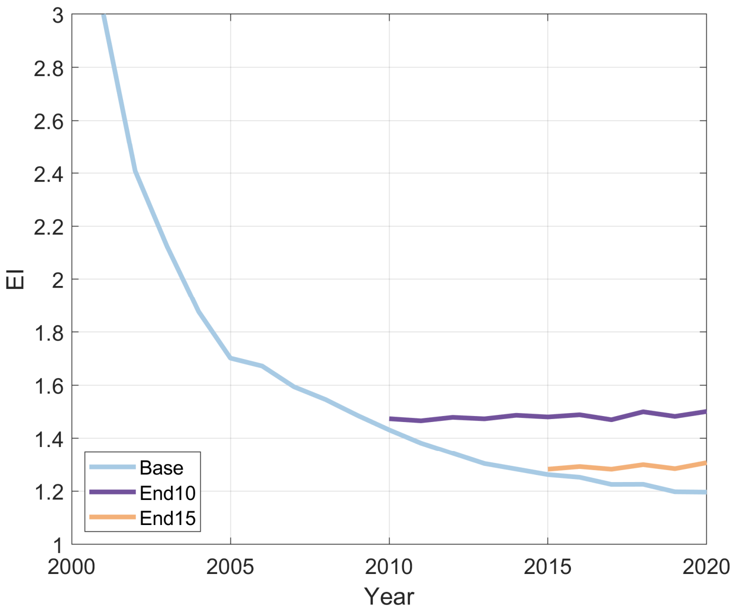

Figure 3. Yearly averages in EI showing the effect of ceasing measurements from 2010 (End10 -purple) and 2015 (End15 - yellow) compared to the baseline (blue) which reflect actual conditions. In both cases the EI became worse over time.

When limiting each site's activity to 12 (Max12), 18 (Max18) and 36 months (Max36), respectively, to investigate the effect of limited tower activity periods we found notable increases in the EI, indicating that extended tower operation periods have a strong beneficial effect on the reduction of upscaling uncertainties (Figure S4 d,e,f). In 2020, the average EI for Max12 was 0.38 higher than the baseline, which is equivalent to setting the network back 14 years to 2006. For the Max18 scenario, EI was 0.27 higher than baseline, equivalent to a 12 year setback to 2008, while for Max36 a reduction in EI of 0.15 was observed, equivalent to rewinding the network to a state 8 years ago to 2012. Across all the experiments, we find a stronger relationship (R^2 0.76, $p > 0.001$, $n=120$) between total mean site activity per year and negative reciprocal transformed EI (-1/EI) (for time passed the relation is weaker R^2 0.44, $p > 0.01$, $n=120$ for total). The strong relation to the negative reciprocal of the data indicates that as more data are added, each addition is relatively less impactful than the ones before.

The depth versus breadth analysis showed that, with the number of active months being exactly equal, there is a slightly better network performance for multiple shorter-lived sites, compared to fewer long active sites. In the BvD10 scenario, the maximum number of sites (127) had a 0.016pp lower EI than the minimum number (55), while in the case of BvD15 and BvD20 these values were 0.017 and 0.024 lower, respectively. In the case of BvD10, the high and low rating fell within each other's estimated standard deviation based on 20 replicate runs. In the BvD15 scenario, the high value fell outside the low value's standard deviation, and in BvD20 both values fell outside each other's standard deviation (Fig. 4). Thus, as the network progresses, differences between the strategies on how to distribute data months across N numbers of sites become more pronounced.

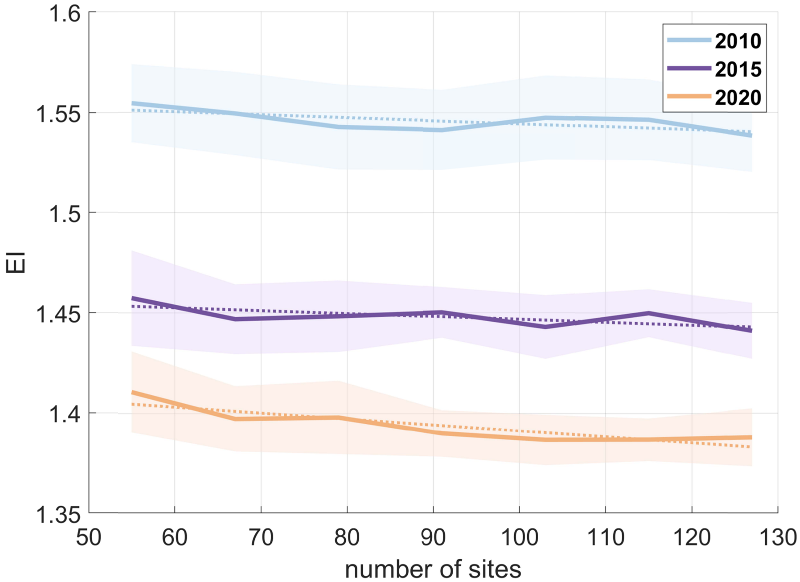

Figure 4. The impact of more shorter-lived sites (left) versus fewer longe- lived sites (right) In this depth versus breadth analysis, site months equal to the combined site months of 2010, 2015 and 2020 (in blue, purple and yellow respectively) were pseudo randomly allocated to sites to answer which performed better: more shorter-lived site or fewer longer lived sites. Lines indicate the mean, dotted lines the trendline, shaded area represent one standard deviation. The slopes for trendlines are 1.5e^{-4}, -1.5e^{-4} and -3.0e^{-4} for the 2010, 2015 and 2020 based allocations respectively. In all experiments network configurations characterized by more shorter-lived sites (n=127) performed better than the few longer active sites (n=55).

3.3 Network optimization

The regional network optimization conducted in the context of this study shows that in the free optimization scenario, a single site can reduce the EI of the entire domain by as much as 0.017, while 8 sites selected from the 28 preselected sites can yield an EI reduction by up to 0.075. In the free exclude setup, where seven sites that were deemed unsuitable for network extension were removed, the EI reduction achieved only 92% of the values obtained in the free optimization scenario (Figure 5). Limiting the degrees of freedom of the algorithm further by prescribing three sites resulted in an initial EI of 67% compared to the free optimization reference, after letting the model choose the best configuration of the remaining five sites, EI increased to 96% of the ideal optimization.

When choosing new sites randomly from the subset of 28 candidate sites, even the best result taken from a subset of 1000 random assignments still lagged the results of the free optimization (91%). With the median random assignment achieving an EI reduction of 74% compared to that of the free optimization, this highlights the clear benefits of the guided optimization. When less than ideal sites are initially chosen, such as tested in the free exclude scenario, the network first drops to EI levels similar to median random assignment however, when the optimization is allowed to choose subsequent sites it fills in the gaps and brings the EI ratings to levels similar to the free exclude scenario, well beyond the random median (Figure 5).

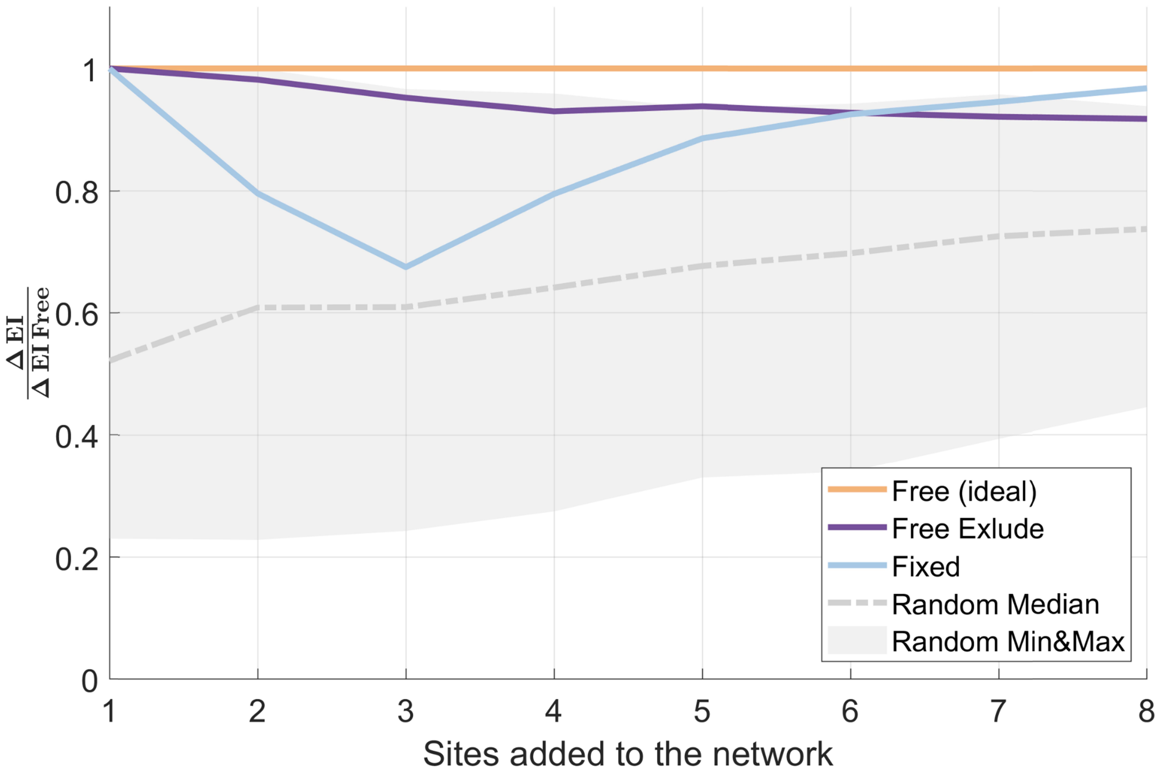

Figure 5. The relative impact of network improvement Network improvements scenarios are shown compared to the ideal case (Free in yellow). Free exclude (purple) with seven less sites for potential selections performed at 91% of the ideal case. The Fixed scenario (blue) with two far less ideal sites in position two and three drops to 67% of ideal at three sites but recovers after optimizing the final 5 sites to 96% of ideal. Gray lines indicate the random allocation, with dashed lines indicating the mean and the dash dotted line the minimum and maximum values.

In Figure 6, the mean improvement to the network is shown for the first step of the network optimization run. Each bar represents its site's improvement to the network if added, as well as their clustering and mean improvement per cluster; the largest gains are present in the three northernmost clusters. Here the optimization algorithm would select Iqaluit as the optimal site. However, if there were practical concerns that would invalidate this location, the graph shows that Ivugvik would be a good alternative location, both in its improvement of the network as well as in environmental conditions, since they share a cluster. The combined use of a similarity metric among potential sites with the EI allowed us therefore to weigh the choices of new sites with knowledge about the location and make informed decisions when choosing between sites without sacrificing much of the potential gains in network performance that would come from a free optimization.

The Free optimization selected the following eight sites in order: Iqaluit, McGill High Arctic Station, Ivugivik, Repulse Bay, Clearwater Lake Station, Bylot Island Field Station, Rankin Inlet and finally Pangnirtung. This corresponds to one site from the Far North cluster, three from the North cluster, three from the Central cluster, and one from the South cluster, which reflects the state of the network (Figure 2) and the distribution of site improvements shown in Figure 2 well. Sites such as Mcgill High Arctic and Clearwater lake are chosen before other sites which performed better in the initial step because the addition of each site impacts the potential benefits of other (especially similar) sites being added. Therefore for every step a new table like Figure 5 is created, reflecting the new rankings of improvement .

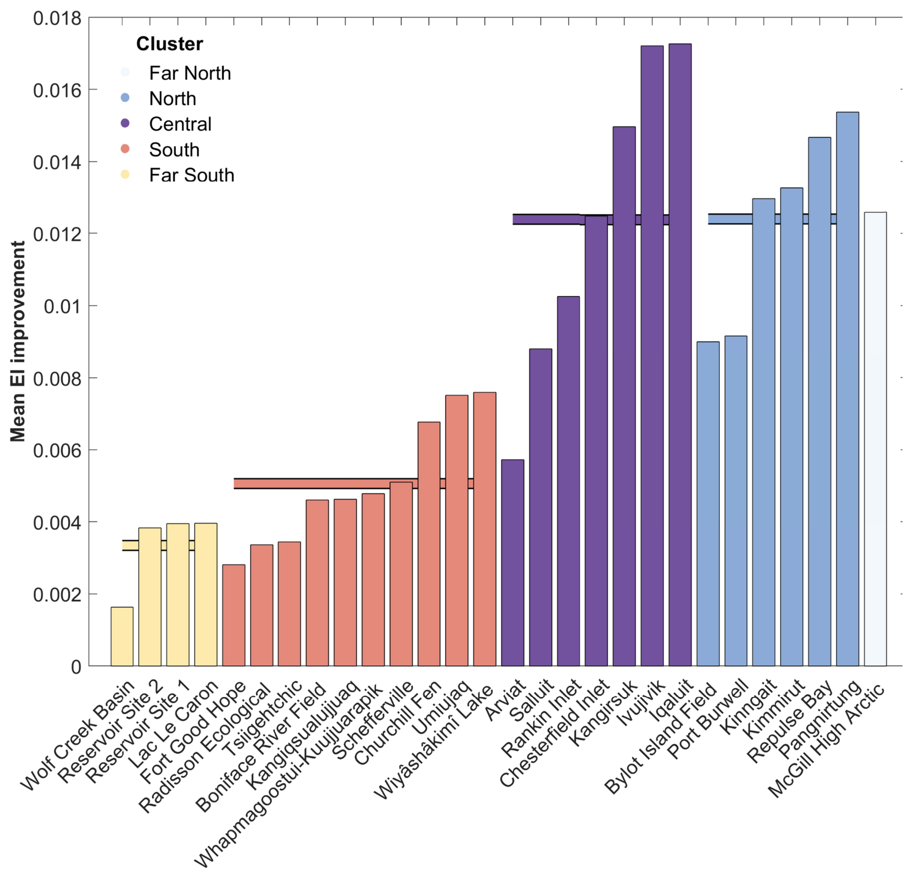

Figure 6. The mean improvement per site for the first network extension step in Canada shown in vertical bars, clustered in 5 groups based on predictor variables to show site similarity, with horizontal bars indicating the clusters mean EI improvement. Graphs like these guide the expert judgment where in one view each individual site's impact can be assessed as well as their similarity to other sites. Here it is clearly shown that the greatest improvement can be found in the Central, North and far North clusters.

4 Discussion

4.1 Network status

Following the expansion of our domain (Figure 4) compared to the assessment presented by Pallandt et al. (2022), the updated representativeness maps indicate that the boreal biome in central Canada is very well represented, on par with Sweden and Finland and selected Alaskan regions. In contrast, the Arctic in Canada lacks representation by the current network of EC towers mainly at high latitudes. Considering the low number of towers in Siberia, the extension of the analysis domain further south in Russia shows that large areas are not well represented in that region. However, even in the case of relatively well represented regions, areas with poor representativeness still exist. Of note is the Aleutian island chain in Western Alaska, which is barely covered by the existing tower network. Having any type of flux measurement here thus appears to be a meaningful upgrade to the network, since, as opposed to many other underrepresented areas, these islands are neither fully mountainous nor Arctic deserts. Furthermore, this rainy region might provide important insights into how northern ecosystems might function in a future wetter climate (Bintanja & Andry, 2017). It should be noted that regions near the southern border may show elevated EI ratings, corresponding to large extrapolation errors, that may not be indicative of their actual status. The reason for this is that we do not consider sites outside the domain that could still influence it, particularly along the southern margins. The overall effect should be minimal though, since in this study we have delineated the domain based on complete ecoregions.

4.2 Extrapolation framework, and uncertainty assessment

The EI approach estimates how the model error increases with distance in predictor space from the training data. The distance considers different predictor importances for the defined target variable. While this yields an objectively defined and interpretable metric it is important to understand caveats of the approach. The choice of the target variable, here GPP, influences the extrapolation assessment because the target variable should determine the set of relevant predictors and associated weights used to calculate the distance to the training data. Transferring the results to other target variables would require that the set of predictors and related importances are similar to the chosen GPP target. Here we used GPP predictions from FLUXCOM extracted at site locations as target variable instead of using real EC data due to a lack of availability. This means that the estimated increase in model error with increased distance to training data in environmental space is larger than if real GPP observations were used and explains the substantially larger EI values compared to Jung et al. 2020. The model error assessed by real observations is much less sensitive to distance to training data because the error is dominated by site-specific peculiarities that are not perfectly captured, for example due to an incomplete predictor set. An incomplete predictor set further implies that we can only assess the 'known unknowns' by our extrapolation assessment (Jung et al. 2020). Essentially, the considerations above imply that (1) the spatial-temporal patterns of estimated EI are qualitatively meaningful but probably optimistic because the chosen predictor set and the FLUXCOM model are not perfect, and that (2) the magnitude of the estimated EI values are likely conservative, i.e. overestimated, because of using model predictions that are more sensitive to distance in predictor space compared to observations.

4.3 Consideration of temporal aspects in representativeness assessments

Since 2011, the network representativeness assessment during winter months performs better, i.e., yields a lower averaged EI, than during summer. This finding appears counter-intuitive, since the wintertime features substantially fewer active towers, with site activity being restricted to the growing season for a large fraction of the EC towers. At the same time, spatial heterogeneity in several of the parameter grids used for upscaling, e.g. NDVIRg or NDWI, is strongly reduced and sometimes zero when snow cover is present, and due to these homogeneous conditions in the wintertime fewer towers are needed to properly reflect conditions within the upscaling domain; note though that our predictors did not include variables describing snow depth and density that might create more spatial variation in the wintertime environmental space. Accordingly, by considering temporal differences in conditions we now can show that this temporal aspect is essential to gain a full picture of the network's performance. However, to fully capture wintertime variability we should utilize actual fluxes as target since wintertime GPP is essentially zero Finally there are increased gaps and errors in wintertime fluxes as a result of adverse measurement conditions (Oechel et al., 2014) which are not present in this dataset. Thus, while these results indicate a reduced need for wintertime monitoring, further research is required to properly account for all nuances in Arctic ecosystems.

From the perspective of managing a continuously operating Arctic observation network, we see a discrepancy between the funding required for proper network performance and the funding that is secured. While our historic data show that many sites stay active for longer than the prevailing three-year funding, the lack of a central, long-term funding source in many regions leads to the discontinuation of EC towers that fill crucial positions within the network. Our results highlight that network representativeness scales with the number of total active months in the dataset, and that continued, long-term measurements are required since our knowledge of the region's fluxes will eventually deteriorate in the absence of new measurements especially with increasing disturbances, ecosystem shifts, and climate change. In addition to upscaling potential of the network, there are other reasons to aim for longer time series such as understanding the ecosystems response to changing conditions, such as (Baldocchi, 2020). In other words, even though it may be sufficient to measure for about three years to constrain a basic carbon budget for a given site, at least if those measurements are done in average site conditions and not during extreme climate or disturbance years, this amount of data is not sufficient to support long-term extrapolation studies. There are programs in place which build long term networks such as the EU based ICOS (Integrated Carbon Observation System (ICOS) Research Infrastructure, 2022) network, and USA based NEON (Schimel et al., 2007). The pan-Arctic network would benefit from having such funding sources for the entire domain.

The depth versus breadth analysis shows that under equal activity there is a slightly better result from the representativeness evaluation for multiple shorter lasting sites over fewer long-term sites. Combined with the results of the case study and previous work, this could lead to the conclusion that raising towers in unique new locations is more impactful than long site activity in a singular space. However, there are other factors to consider beyond regional upscaling. With a focus on breath, we might lose understanding of detailed local processes: as the ecosystem and climate changes, ecosystems respond and new processes and disturbances may happen, which could be missed or only detected after a considerably longer time. Furthermore, the cost of maintaining a tower is one to two orders of magnitude lower than establishing a new tower, where the instrument cost (ICOS ERIC, 2020) and the costs for permits and construction are by far the largest investments. Since the total site months of the network is the most important indicator of EI, the most cost effective method to extend this is by keeping existing towers operational. New towers should then ideally be located in underrepresented regions as selected by this or similar methods, while still answering the project's research questions. In cases where there is no direct experimental need to remain in one location for a long time, from the perspective of the network as a whole, it would be efficient to rotate equipment between several locations. If, at the start of an experiment, power and a tower structure are erected at several locations, then the instruments can be rotated between these sites with relative ease. The results from the Endx experiments show that loss of representativeness represented by increase of EI are relatively slow, therefore gaps should be manageable when considering flux upscaling, and when one would return at regular intervals it allows for any correction of accelerated change in the ecosystem. Furthermore, many remote locations have low expected fluxes (Lafleur et al., 2012; Virkkala et al., 2021); temporary or mobile towers could be ideal to add representativeness of such locations to the network without having to make the investment of a permanent tower. Drone campaigns such as polar Modular Observation Solutions for Earth Systems camping can fulfill a similar purpose. It is clear from these analyses (Figure 3 and 6) that as far as network design is concerned to fill the gaps the EC community has to focus on less accessible locations, even though this comes at increased costs.

The results of the Endx experiments should be considered a conservative estimate with actual EI increase likely higher. Several of the input rasters used are static over the years. And while measurements such as NBAR will not see significant change on these time scales, data such as Enhanced Vegetation Index, Land cover and Maximum Day Time Land Surface Temperature are expected to change and not remain static. If these layers were dynamic, variations over time in these variables would increase and so would the EI when no new measurements are taken. Furthermore, the Arctic is changing at an accelerating rate (Box et al., 2019), in the absence of measurements this leads to an accelerated increase of the EI as known conditions are increasingly exceeded. All of these arguments again speak for the continuation of long-term experiments.

4.4 Network expansion strategies

We have shown here that utilizing a model-guided approach to network extension greatly outperforms a random allocation (to the same feasible locations), and that this holds true even when we include less ideal choices since the model can compensate for this with further selections. As expected from the EI map in Figure 2, there is a clear preference for more northern locations. It should be noted though that this optimization was aimed at reducing the EI, the relative error as a function of distance to closest sites, which does not include the magnitude of the individual fluxes. If flux magnitude were to be considered in the metric, high-latitude sites would be comparatively less likely to be selected owing to typically lower fluxes. However such inclusion would add additional complexity and potential biases as it would either require a model ensemble to establish error magnitudes of the fluxes (Jung et al., 2020) or a weighting of the errors by expected fluxes with ambiguity on the weight the magnitude should have.

When choosing a location for a new site, methods like these where representativeness-based optimization models are used in tandem with expert knowledge combine the best of two worlds.

The modeling component grants objective insights in a potential site's impact to the network and its relation to other sites, and the expert can easily consider aspects that are too unwieldy or impossible to properly model, such as experimental design, infrastructure, and advice and requests from local communities. Quantifying tradeoffs further helps the decision-making process especially with clear visualizations.

Conclusion

We have shown that the high-latitude EC network has grown considerably over time, with significant increases in representativeness. This analysis also shows that the coverage of the EC network still needs to be improved for estimating more robust Arctic-boreal carbon budgets. Large improvements are needed especially in the highest latitudes, mountainous regions and large parts of Russia. However, improving the network requires relatively more effort with each site addition since each additional site will have comparatively less impact than the ones before as the data space is steadily filled. At the same time, we see that the largest gaps are in more remote locations, further adding to the difficulty of expansion.

To further guide the growth of the network we have demonstrated a network optimization method that greatly outperforms a random approach in a case study where we optimize the network by considering future expansions in the Canadian Arctic. We illustrate a way to merge representativeness based optimized network design with expert knowledge in an iterative way that incorporates understanding, local knowledge, and other hard to quantify factors.

Beyond extending the network it has become clear that we cannot be complacent with the existing network, as gaps in data and cessation of measurements will not only freeze our knowledge but deteriorate our ability to understand the carbon cycle. This is especially the case since rapid climate change in the Arctic is bound to move conditions further from past measurements. This is exacerbated by acceleration at which the high latitudes are changing as a result of climate change. And since total site months are central to increasing network representativeness, it is therefore of importance that existing sites should remain active and be funded for as long as possible in addition to efforts to extend the network to underrepresented locations.

Acknowledgements

None of the authors report any affiliations or financial interests that cause conflicts of interest. This work was supported by the Max Planck Society and through funding by the European Commission (INTAROS project, H2020-BG-09-2016, grant agreement no. 727890), and also by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 951288, Q-Arctic). Additional support for this work came from the TED Audacious Project and Woodwell Climate Research Center's Fund For Climate Solutions Project. Kyle A. Arndt is additionally supported by NSF Office of Polar Programs grant no. 2316114. The article processing charges for this open-access publication were covered by the Max Planck Society. We would like to thank Michał Gałkowski for his review of this manuscript.

Open Research

Table 1 lists all original MODIS data codes for raster datasets used in this research, and is available at https://lpdaac.usgs.gov/products, these products have been regridded for use with FLUXCOM those versions can be found in the Max Planck Institute for Biogeochemistry data portal at https://www.bgc-jena.mpg.de/geodb/projects/Home.php file id 260. EC site metadata is available at https://cosima.nceas.ucsb.edu/carbon-flux-sites/. This is an active database that is constantly updated, for transparency a snapshot of the EC component of this database used for this paper is retained and available on request by reviewers. EI code is from Jung et al. (2020) with specific details in supplement 2. All analyses were performed using matlab (The MathWorks Inc, 2022), Figures were produced in matlab, Figure 1 was produced with the addaxis addon to plot an extra axis (Lee, 2023), and Figure 2 utilized the shaded area error bar plot addon for the std shading (Martínez-Cagigal, 2023) , Figure 2 and S2 was created as geotiff in matlab and then finalized using Qgis (QGIS Development Team, 2009).

References

Baldocchi, D. D. (2003). Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: past, present and future. *Global Change Biology*, *9*(4), 479–492. https://doi.org/10.1046/j.1365-2486.2003.00629.x

Baldocchi, D. D. (2020). How eddy covariance flux measurements have contributed to our understanding of Global Change Biology. *Global Change Biology*, *26*(1), 242–260. https://doi.org/10.1111/gcb.14807

Banko, M., & Brill, E. (2001). Mitigating the Paucity-of-Data Problem: Exploring the Effect of Training Corpus Size on Classifier Performance for Natural Language Processing. In *Proceedings of the First International Conference on Human Language Technology Research*. Retrieved from https://aclanthology.org/H01-1052

Bintanja, R., & Andry, O. (2017). Towards a rain-dominated Arctic. *Nature Climate Change*, *7*(4), 263–267. https://doi.org/10.1038/nclimate3240

Birch, L., Schwalm, C. R., Natali, S., Lombardozzi, D., Keppel-Aleks, G., Watts, J., et al. (2021). Addressing biases in Arctic–boreal carbon cycling in the Community Land Model Version 5. *Geoscientific Model Development*, *14*(6), 3361–3382. https://doi.org/10.5194/gmd-14-3361-2021

Bosveld, F. C., & Beljaars, A. C. M. (2001). The impact of sampling rate on eddy-covariance flux estimates. *Agricultural and Forest Meteorology*, *109*(1), 39–45. https://doi.org/10.1016/S0168-1923(01)00257-X

Box, J. E., Colgan, W. T., Christensen, T. R., Schmidt, N. M., Lund, M., Parmentier, F.-J. W., et al. (2019). Key indicators of Arctic climate change: 1971–2017. *Environmental Research Letters*, *14*(4), 045010. https://doi.org/10.1088/1748-9326/aafc1b

Byrne, B., Baker, D. F., Basu, S., Bertolacci, M., Bowman, K. W., Carroll, D., et al. (2023). National CO_2 budgets (2015–2020) inferred from atmospheric CO_2 observations in support of the global stocktake. *Earth System Science Data*, *15*(2), 963–1004. https://doi.org/10.5194/essd-15-963-2023

Chu, H., Luo, X., Ouyang, Z., Chan, W. S., Dengel, S., Biraud, S. C., et al. (2021). Representativeness of Eddy-Covariance flux footprints for areas surrounding AmeriFlux sites. *Agricultural and Forest Meteorology*, *301–302*, 108350. https://doi.org/10.1016/j.agrformet.2021.108350

Desai, A. R. (2010). Climatic and phenological controls on coherent regional interannual variability of carbon dioxide flux in a heterogeneous landscape. *Journal of Geophysical Research: Biogeosciences*, *115*(G3), G00J02. https://doi.org/10.1029/2010JG001423

Göckede, M., Rebmann, C., & Foken, T. (2004). A combination of quality assessment tools for eddy covariance measurements with footprint modelling for the characterisation of complex sites. *Agricultural and Forest Meteorology*, *127*(3), 175–188. https://doi.org/10.1016/j.agrformet.2004.07.012

Hoffman, F. M., Kumar, J., Mills, R. T., & Hargrove, W. W. (2013). Representativeness-based sampling network design for the State of Alaska. *Landscape Ecology*, *28*(8), 1567–1586. https://doi.org/10.1007/s10980-013-9902-0

Hugelius, Gustaf, Loisel, J., Chadburn, S., Jackson, R. B., Jones, M., MacDonald, G., et al. (2020). Large stocks of peatland carbon and nitrogen are vulnerable to permafrost thaw. *Proceedings of the National Academy of Sciences*, *117*(34), 20438–20446. https://doi.org/10.1073/pnas.1916387117

ICOS ERIC. (2020). *ICOS Handbook 2020* (2nd rev. ed.). Helsinki: ICOS ERIC. Retrieved from https://www.icos-cp.eu/resources/reports-and-documents

Integrated Carbon Observation System (ICOS) Research Infrastructure. (2022). *ICOS Handbook 2022*. Helsinki: ICOS ERIC. Retrieved from https://www.icos-cp.eu/resources/brochures

IPCC. (2014). *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (p. 151). Geneva, Switzerland: IPCC. Retrieved from https://www.ipcc.ch/pdf/assessment-report/ar5/syr/SYR_AR5_FINAL_full_wcover.pdf

Ito, A., Li, T., Qin, Z., Melton, J. R., Tian, H., Kleinen, T., et al. (2023). Cold-Season Methane Fluxes Simulated by GCP-CH4 Models. *Geophysical Research Letters*, *50*(14), e2023GL103037.

https://doi.org/10.1029/2023GL103037

Jung, M., Reichstein, M., Margolis, H. A., Cescatti, A., Richardson, A. D., Arain, M. A., et al. (2011). Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations. *Journal of Geophysical Research: Biogeosciences*, *116*(G3), G00J07. https://doi.org/10.1029/2010JG001566

Jung, M., Schwalm, C., Migliavacca, M., Walther, S., Camps-Valls, G., Koirala, S., et al. (2020). Scaling carbon fluxes from eddy covariance sites to globe: synthesis and evaluation of the FLUXCOM approach. *Biogeosciences*, *17*(5), 1343–1365. https://doi.org/10.5194/bg-17-1343-2020

Kljun, N., Rotach, M. W., & Schmid, H. P. (2002). A three-dimensional backward Lagrangian footprint model for a wide range of boundary-layer stratifications. *Boundary-Layer Meteorology*, *103*, 205–226.

Lafleur, P. M., Humphreys, E. R., St. Louis, V. L., Myklebust, M. C., Papakyriakou, T., Poissant, L., et al. (2012). Variation in Peak Growing Season Net Ecosystem Production Across the Canadian Arctic. *Environmental Science & Technology*, *46*(15), 7971–7977. https://doi.org/10.1021/es300500m

Lee, H. (2023). addaxis. Retrieved from https://www.mathworks.com/matlabcentral/fileexchange/9016-addaxis

Loescher, H. W., Law, B. E., Mahrt, L., Hollinger, D. Y., Campbell, J., & Wofsy, S. C. (2006). Uncertainties in, and interpretation of, carbon flux estimates using the eddy covariance technique. *Journal of Geophysical Research: Atmospheres*, *111*(D21). https://doi.org/10.1029/2005JD006932

Martínez-Cagigal, V. (2023). Shaded area error bar plot. Retrieved from https://www.mathworks.com/matlabcentral/fileexchange/58262-shaded-area-error-bar-plot

Meredith, M., Sommerkorn, M., Cassotta, S., Derksen, C., Ekaykin, A., Hollowed, A., et al. (2019). *Polar Regions. Chapter 3, IPCC Special Report on the Ocean and Cryosphere in a Changing Climate*. Geneva, Switzerland: IPCC. Retrieved from https://www.ipcc.ch/srocc/

Meyer, H., & Pebesma, E. (2021). Predicting into unknown space? Estimating the area of applicability of spatial prediction models. *Methods in Ecology and Evolution*, *12*(9), 1620–1633. https://doi.org/10.1111/2041-210X.13650

Natali, S. M., Watts, J. D., Rogers, B. M., Potter, S., Ludwig, S. M., Selbmann, A.-K., et al. (2019). Large loss of CO_2 in winter observed across the northern permafrost region. *Nature Climate Change*, *9*(11), 852–857. https://doi.org/10.1038/s41558-019-0592-8

Oechel, W. C., Laskowski, C. A., Burba, G., Gioli, B., & Kalhori, A. A. M. (2014). Annual patterns and budget of CO2 flux in an Arctic tussock tundra ecosystem. *Journal of Geophysical Research: Biogeosciences*, *119*(3), 323–339. https://doi.org/10.1002/2013JG002431

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., et al. (2001). Terrestrial Ecoregions of the World: A New Map of Life on Earth: A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *BioScience*, *51*(11), 933–938. https://doi.org/10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2

Pallandt, M. M. T. A., Kumar, J., Mauritz, M., Schuur, E. A. G., Virkkala, A.-M., Celis, G., et al. (2022). Representativeness assessment of the pan-Arctic eddy covariance site network and optimized future enhancements. *Biogeosciences*, *19*(3), 559–583. https://doi.org/10.5194/bg-19-559-2022

Pastorello, G., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y.-W., et al. (2020). The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. *Scientific Data*, *7*(1), 225. https://doi.org/10.1038/s41597-020-0534-3

Peltola, O., Vesala, T., Gao, Y., Räty, O., Alekseychik, P., Aurela, M., et al. (2019). Monthly gridded data product of northern wetland methane emissions based on upscaling eddy covariance observations. *Earth System Science Data*, *11*(3), 1263–1289. https://doi.org/10.5194/essd-11-1263-2019

QGIS Development Team. (2009). QGIS Geographic Information System. Open Source Geospatial Foundation. Retrieved from http://qgis.osgeo.org

Rannik, Ü., Aubinet, M., Kurbanmuradov, O., Sabelfeld, K. K., Markkanen, T., & Vesala, T. (2000). Footprint analysis for measurements over a heterogeneous forest. *Boundary-Layer Meteorology*, *97*, 137–166.

Rantanen, M., Kämäräinen, M., Niittynen, P., Phoenix, G. K., Lenoir, J., Maclean, I., et al. (2023). Bioclimatic atlas of the terrestrial Arctic. *Scientific Data*, *10*(1), 40. https://doi.org/10.1038/s41597-023-01959-w

Schimel, D., Hargrove, W., Hoffman, F., & MacMahon, J. (2007). NEON: a hierarchically designed national ecological network. *Frontiers in Ecology and the Environment*, *5*(2), 59–59. https://doi.org/10.1890/1540-9295(2007)5[59:NAHDNE]2.0.CO;2

Schmid, H. P. (1997). Experimental design for flux measurements: matching scales of observations and fluxes. *Agricultural and Forest Meteorology*, *87*(2), 179–200. https://doi.org/10.1016/S0168-1923(97)00011-7

Schuur, E. A. G., McGuire, A. D., Schädel, C., Grosse, G., Harden, J. W., Hayes, D. J., et al. (2015). Climate change and the permafrost carbon feedback. *Nature*, *520*(7546), 171–179. https://doi.org/10.1038/nature14338

Schuur, Edward A. G., Bockheim, J., Canadell, J. G., Euskirchen, E., Field, C. B., Goryachkin, S. V., et al. (2008). Vulnerability of Permafrost Carbon to Climate Change: Implications for the Global Carbon Cycle. *BioScience*, *58*(8), 701–714. https://doi.org/10.1641/B580807

Serreze, M. C., & Barry, R. G. (2011). Processes and impacts of Arctic amplification: A research synthesis. *Global and Planetary Change*, *77*(1), 85–96. https://doi.org/10.1016/j.gloplacha.2011.03.004

Sulkava, M., Luyssaert, S., Zaehle, S., & Papale, D. (2011). Assessing and improving the representativeness of monitoring networks: The European flux tower network example. *Journal of Geophysical Research: Biogeosciences*, *116*(G3), G00J04. https://doi.org/10.1029/2010JG001562

The MathWorks Inc. (2022). MATLAB (Version 9.10.0. (R2021a)). Natick, Massachusetts, United States: The MathWorks Inc. Retrieved from https://www.mathworks.com

Ueyama, M., Iwata, H., Harazono, Y., Euskirchen, E. S., Oechel, W. C., & Zona, D. (2013). Growing season and spatial variations of carbon fluxes of Arctic and boreal ecosystems in Alaska (USA). *Ecological Applications*, *23*(8), 1798–1816. https://doi.org/10.1890/11-0875.1

Vesala, T., Kljun, N., Rannik, U., Rinne, J., Sogachev, A., Markkanen, T., et al. (2008). Flux and concentration footprint modelling: state of the art. *Environmental Pollution (Barking, Essex: 1987)*, *152*(3), 653–666. https://doi.org/10.1016/j.envpol.2007.06.070

Villarreal, S., & Vargas, R. (2021). Representativeness of FLUXNET Sites Across Latin America. *Journal of Geophysical Research: Biogeosciences*, *126*(3), e2020JG006090. https://doi.org/10.1029/2020JG006090

Virkkala, A.-M., Aalto, J., Rogers, B. M., Tagesson, T., Treat, C. C., Natali, S. M., et al. (2021). Statistical upscaling of ecosystem CO2 fluxes across the terrestrial tundra and boreal domain: regional patterns and uncertainties. *Global Change Biology*, *n/a*(n/a). https://doi.org/10.1111/gcb.15659

Ward, J. H. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, *58*(301), 236–244. https://doi.org/10.1080/01621459.1963.10500845

Wisz, M. S., Hijmans, R. J., Li, J., Peterson, A. T., Graham, C. H., Guisan, A., & Group, N. P. S. D. W. (2008). Effects of sample size on the performance of species distribution models. *Diversity and Distributions*, *14*(5), 763–773. https://doi.org/10.1111/j.1472-4642.2008.00482.x

Xiao, J., Chen, J., Davis, K. J., & Reichstein, M. (2012). Advances in upscaling of eddy covariance measurements of carbon and water fluxes. *Journal of Geophysical Research: Biogeosciences*, *117*(G1), G00J01. https://doi.org/10.1029/2011JG001889