## **Materials and Methods**

We use a lagrangian transport model HYSPLIT using meteorological fields from National Center for Environmental Prediction (NCEP) (2.50 latitude-longitude grids) to generate 15 days backward trajectories on daily basis at 500m above the ground level [agl] during the period 2000–2019 at 4 locations which are representative of the different Arctic regions: Greenland, Boreal Forest, Coastal Tundra and Central Arctic. Both meteorological dataset has been used widely in a number of studies concerned with the airmass transport, and are able to capture the meteorological variability in Arctic region reasonably well.Generated backward trajectories has been discretised to 10 x 10 latitude-longitude grids. The study of the correlation between the concentration at a measurement site and the backtrajectories gives much better informations about where the source might be located along the path of the trajectory. In this study we use one such method: Potential source contribution function (PSCF).

### Backward Trajectory clustering

A number of studies has proved the superiority of frequecy based trajectory representation (discrete fourier transform (DFT) and discrete wavelet transform (DWT)-based coefficents) over other approximation like least-square based polynomial approximations and Chebyshev approximation.[15](#ref-Naftel2006),[16](#ref-Guo2017) We cluster trajectories with K-Means clustering using non-parametric features extracted from discrete wavelet transform (DWT) based decomposition of projected backward trajectories which encode the time-varying information of the trajectories. We use Haar wavelets to capture local variations in trajectories using the single-level implementation of Mallat’s algorithm for computing the DWT coefficients. We build a feature vector comprising of derived approximation DWT coefficients as follows:[17](#ref-Nawaz2014)

where

fi provides comprehensive description of trajectories as it captures the overall distribution of coefficients non-parametrically. This approach does not require the trajectories to have equal length.

We apply elbow criterion to determine the optimum number of clusters.[18](#ref-Novikov2019) This heuristic method parameterizes the percentage of explained variance as a function of cluster count and chooses a number of clusters in such a way that adding another cluster doesn’t incrase the explained variance. Explained variance (exvar) which is equivalent to the R-squared value in classical regression is given as[19](#ref-Shao2007) :

where SSR is the sum of squares regression (or explained variation) from each cluster (summed over all clusters) and SST is the total sum of squares. We use the mean of trajectories in a cluster to represent that trajectory cluster.

### Potential source contribution function (PSCF)

PSCF is a backward trajectories receptor model which can be used for the determination of potential source areas of ozone at a receptor.[20](#ref-Ashima2017)–[22](#ref-Cristofanelli2018) It is defined as the conditional probability of an air parcel that crosses a cell have ozone concentration higher than a threshold amount given that the trajectory has pass through the cell under consideration. If nij is the number of trajectory endpoints in a cell ij and mij is the number of endpoints which correspond to trajectories with ozone concentration higher than the threshold value at the receptor location, then

where Eij is the event of an end point occurring in a cell ij, CR is the ozone concentration at the receptor and Cth is the threshold ozone concentration for determining the pollution event.

Grid cells with high PSCF values are associated with the arrival of air parcels at the receptor site that have concentrations of the analyte higher than the criterian value. These cells are areas of high potential contributions for the constituent. In this study, the criterion value was set to the third quantile (75%) of the surface ozone concentration at the considered Arctic-based receptor location.

### Weight function for PSCF

We apply an arbitrary weight function to PSCF to reduce the uncertainties arising from small number of trajectory end points in a grid cell. The weight function is given as follows as23 :

Here, nij is the number of end points in a grid cell ij and N is the average number of end points for the whole study region.

Weighted receptor models result in field gradient around the source region; aiding in the identification of the potential sources.

### References

15. Naftel, A. & Khalid, S. Motion Trajectory Learning in the DFT-Coefficient Feature Space. (2006).

16. Guo, Y., Xu, Q., Li, P., Sbert, M. & Yang, Y. Trajectory Shape Analysis and Anomaly Detection Utilizing Information Theory Tools. *Entropy* **19**, 323 (2017).

17. Nawaz, T., Cavallaro, A. & Rinner, B. Trajectory clustering for motion pattern extraction in aerial videos. in *2014 ieee international conference on image processing (icip)* 1016–1020 (IEEE, 2014).

18. Novikov, A. PyClustering: Data Mining Library. *Journal of Open Source Software* **4**, 1230 (2019).

19. Shao, J., Tanner, S. W., Thompson, N. & Cheatham, T. E. Clustering Molecular Dynamics Trajectories: 1. Characterizing the Performance of Different Clustering Algorithms. *Journal of Chemical Theory and Computation* **3**, 2312–2334 (2007).

20. Sharma, A., Mandal, T. K., Sharma, S. K., Shukla, D. K. & Singh, S. Relationships of surface ozone with its precursors, particulate matter and meteorology over Delhi. *Journal of Atmospheric Chemistry* **74**, 451–474 (2017).

21. Giemsa, E., Jacobeit, J., Ries, L. & Hachinger, S. Investigating regional source and sink patterns of Alpine CO 2 and CH 4 concentrations based on a back trajectory receptor model. *Environmental Sciences Europe 2019 31:1* **31**, 1–24 (2019).

22. Cristofanelli, P. *et al.* Analysis of multi-year near-surface ozone observations at the WMO/GAW ‘Concordia’ station (75°06′S, 123°20′E, 3280 m a.s.l. – Antarctica). *Atmospheric Environment* **177**, 54–63 (2018).

23. Stojić, A. & Stanišić Stojić, S. The innovative concept of three-dimensional hybrid receptor modeling. *Atmospheric Environment* **164**, 216–223 (2017).