



Michigan Technological University
Create the Future Digital Commons @ Michigan Tech

Dissertations, Master's Theses and Master's
Reports - Open

Dissertations, Master's Theses and Master's
Reports

2003

Climatological study of transport to the PICO-NARE site using atmospheric backward trajectories

Robert Christopher Owen
Michigan Technological University

Follow this and additional works at: <https://digitalcommons.mtu.edu/etds>


 Part of the [Civil and Environmental Engineering Commons](#)

Copyright 2003 Robert Christopher Owen

Recommended Citation

Owen, Robert Christopher, "Climatological study of transport to the PICO-NARE site using atmospheric backward trajectories", Master's Thesis, Michigan Technological University, 2003.
<https://doi.org/10.37099/mtu.dc.etds/262>

Follow this and additional works at: <https://digitalcommons.mtu.edu/etds>

 Part of the [Civil and Environmental Engineering Commons](#)

**A Climatological Study of Transport to the PICO-NARE
Site Using Atmospheric Backward Trajectories**

By

Robert Christopher Owen

A THESIS

Submitted in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE IN ENVIRONMENTAL ENGINEERING

MICHIGAN TECHNOLOGICAL UNIVERSITY

May 7, 2003

This thesis, “A Climatological Study of Transport to the PICO-NARE Site Using Atmospheric Backward Trajectories,” is hereby approved in partial fulfillment of the requirements for the degree of MASTER OF SCIENCE IN ENVIRONMENTAL ENGINEERING.

Department of Civil and Environmental Engineering

Dr. Richard E. Honrath, Thesis Advisor

Dr. C. Robert Baillod, Chair, Department of Civil and Environmental Engineering

Date

Abstract

This study developed a transport climatology to the PICO-NARE station, in the central North Atlantic Ocean, using a 40-year set of atmospheric back trajectories. The trajectory set was subjected to a cluster analysis in order to group trajectories into six flow patterns, or clusters. An air flow probability analysis was conducted in conjunction with the cluster analysis in order to determine the source regions for flow to the site. Seasonal differences in the flow patterns were found, which included enhanced westerly flow in the winter, decreased westerly flow in the summer, and spring and fall having moderate westerly flow. The North Atlantic Oscillation had a significant impact on the winter and fall seasons and less significant impacts during spring and summer. The results of the climatology can be used in conjunction with measurements of ozone, CO, NO_x, and NO_y, which are currently being measured at the site, to develop a long-term, seasonal climatology of transport of pollutants to the central North Atlantic.

Acknowledgements

I would like to thank Richard Honrath, my advisor, for allowing a transient Masters International student be a Research Assistant for a year and subsequently being patient enough with that student to allow him to stick around long enough to do this project. I would also like to thank him for his patience and guidance during that process as well as teaching that student to cc ski. Additional thanks to Maria and Chris E. for answering questions that weren't worth bothering Richard over and giving key advice on how to please our advisor.

Thanks to my parents for letting me off the phone when I was hard at work and all the support they have given me in the past that have helped make me who I am. You can never say enough to your parents (good and bad) to explain all your feelings about them. Thanks to Becky for letting me use her computer all the time and getting me green caffeine all the time. Thanks to Dan for helping keeping the apartment liveable while I had to neglect it during my busy periods.

Finally, I would like to thank Roland Draxler for providing the HYSPLIT model.

This project was funded NOAA/Office of Global Programs grant NA16GP1658. Thank you to Richard, Paulo, Mike, Toni, Matt, many people I don't know for making that possible.

Contents

Abstract	iii
Acknowledgments	iv
List of Figures	ix
List of Tables	xxv
1 Introduction	1
1.1 Background	2
1.2 Recent and current research	3
1.3 PICO-NARE	5
1.4 Atmospheric trajectories in a climatology	7
1.5 Study objectives	9
2 Methods	11
2.1 Atmospheric trajectories	11
2.1.1 Trajectory use	11
2.1.2 Description of the HYSPLIT Model	13
2.1.3 Description of the NCEP/NCAR Reanalysis Meteorological Data	16
2.1.4 Trajectory error	18
2.2 Cluster Analysis	20

2.2.1	Clustering algorithms	21
2.2.1.1	K-means	23
2.2.1.2	Steps in the three k-means methods	24
2.2.1.3	Selection of k-means sequence	27
2.2.2	Number of clusters	29
2.2.2.1	Factors in the choice of the number of clusters	32
2.2.2.2	Selection of the number of clusters	35
2.2.3	Units	41
2.2.3.1	Available units	41
2.2.3.2	Criteria for the selection of units	49
2.2.3.3	Final selection of units	51
2.2.3.4	Alternate unit system, GLLH	56
2.2.4	Note on clusters	71
2.3	Airflow Probability Analysis	71
2.3.1	Application and theory	72
2.3.2	Methods	73
2.4	Impacts of the NAO on the climatology of the PICO-NARE Site . . .	78
2.4.1	Climatic variation and the NAO	78
2.4.2	Methods	81
3	Results and Discussion	85
3.1	Climatology results	85
3.1.1	Method of descriptions and discussion	86
3.1.2	Results for 1960-1999	86
3.1.3	Seasons similar to the average period	92
3.1.3.1	Results for spring, 1960-1999	92
3.1.4	NAO seasons similar to the average period	96
3.1.4.1	Results for spring with positive NAOI values, 1960-1999	96

3.1.4.2	Results for spring with negative NAOI values, 1960-1999	100
3.1.4.3	Results for fall with negative NAOI values, 1960-1999	104
3.1.5	Seasons dissimilar to the average period	108
3.1.5.1	Results for winter, 1960-1999	108
3.1.5.2	Results for summer, 1960-1999	113
3.1.5.3	Results for fall, 1960-1999	117
3.1.6	NAO seasons dissimilar to the average period	121
3.1.6.1	Results for winter with positive NAOI values, 1960-1999	121
3.1.6.2	Results for winter with negative NAOI values, 1960-1999	125
3.1.6.3	Results for summer with positive NAOI values, 1960- 1999	129
3.1.6.4	Results for summer with negative NAOI values, 1960- 1999	133
3.1.6.5	Results for fall with positive NAOI values, 1960-1999	137
3.1.7	Summary of cluster analysis	141
3.2	Additional discussion	141
3.2.1	Comments on the vertical component of trajectories	141
3.2.2	Summary of the impacts of the NAO	143
3.3	Future work	144
	References	148
A	Membership plots for whole period	155
A.1	Membership plots for the average period	155
B	Membership plots for winter periods	169
B.1	Membership plots for the average winter	169
B.2	Membership plots for the positive NAO winters	183
B.3	Membership plots for the negative NAO winters	197

C	Membership plots for spring periods	211
C.1	Membership plots for the average spring	211
C.2	Membership plots for the positive NAO springs	225
C.3	Membership plots for the negative NAO springs	239
D	Membership plots for summer periods	253
D.1	Membership plots for the average summer	253
D.2	Membership plots for the positive NAO summers	267
D.3	Membership plots for the negative NAO summers	281
E	Membership plots for fall periods	295
E.1	Membership plots for the average fall	295
E.2	Membership plots for the positive NAO falls	309
E.3	Membership plots for the negative NAO falls	323

List of Figures

1.1	Map of Azores and Pico	8
2.1	Comparison of sequential and classical k-means	30
2.2	Comparison of sequential and classical k-means	31
2.3	Plots of R^2	36
2.4	Plots of variance between clusters	37
2.5	Plots of variance within clusters	38
2.6	Cluster center plot for 4 clusters	42
2.7	Cluster center plot for 5 clusters	43
2.8	Cluster center plot for 6 clusters	44
2.9	Cluster center plot for 7 clusters	45
2.10	Example for discussion of units	52
2.11	DLL 6 cluster plot	57
2.12	DLLH 6 cluster plot	58
2.13	Histogram for DLL and DLLH	59
2.14	KLL 6 cluster plot	60
2.15	KLLH 6 cluster plot	61
2.16	Histogram for KLL and KLLH	62
2.17	SLL 6 cluster plot	63
2.18	SLLH 6 cluster plot	64
2.19	Histogram for SLL and SLLH	65

2.20	NLL 6 cluster plot	66
2.21	NLLH 6 cluster plot	67
2.22	Histogram for NLL and NLLH	68
2.23	GLLH 6 cluster plot	69
2.24	Standard airflow probability analysis plot	75
2.25	Geometrically corrected airflow probability analysis plot	77
2.26	NAOI from 1865 to 2000	83
2.27	NAOI from 1865 to 1999	84
3.1	Cluster plot for 1960-1999	89
3.2	Standard density plot for 1960-1999	90
3.3	Geometrically corrected density plot for 1960-1999	90
3.4	Cluster plot for spring, 1960-1999	93
3.5	Standard density plot for spring, 1960-1999	94
3.6	Geometrically corrected density plot for spring, 1960-1999	94
3.7	Cluster plot for springs with a positive NAOI, 1960-1999	97
3.8	Standard density plot for springs with a positive NAOI, 1960-1999	98
3.9	Geometrically corrected density plot for springs with a positive NAOI, 1960-1999	98
3.10	Cluster plot for springs with a negative NAOI, 1960-1999	101
3.11	Standard density plot for springs with a negative NAOI, 1960-1999	102
3.12	Geometrically corrected density plot for springs with a negative NAOI, 1960-1999	102
3.13	Cluster plot for falls with a negative NAOI, 1960-1999	105
3.14	Standard density plot for falls with a negative NAOI, 1960-1999	106
3.15	Geometrically corrected density plot for falls with a negative NAOI, 1960-1999	106
3.16	Cluster plot for winter, 1960-1999	110

3.17	Standard density plot for winter, 1960-1999	111
3.18	Geometrically corrected density plot for winter, 1960-1999	111
3.19	Cluster plot for summer, 1960-1999	114
3.20	Standard density plot for summer, 1960-1999	115
3.21	Geometrically corrected density plot for summer, 1960-1999	115
3.22	Cluster plot for fall, 1960-1999	118
3.23	Standard density plot for fall, 1960-1999	119
3.24	Geometrically corrected density plot for fall, 1960-1999	119
3.25	Cluster plot for winters with a positive NAOI, 1960-1999	122
3.26	Standard density plot for winters with a positive NAOI, 1960-1999 .	123
3.27	Geometrically corrected density plot for winters with a positive NAOI, 1960-1999	123
3.28	Cluster plot for winters with a negative NAOI, 1960-1999	126
3.29	Standard density plot for winters with a negative NAOI, 1960-1999	127
3.30	Geometrically corrected density plot for negative NAOI winters, 1960- 1999	127
3.31	Cluster plot for summers with a positive NAOI, 1960-1999	130
3.32	Standard density plot for summers with a positive NAOI, 1960-1999	131
3.33	Geometrically corrected density plot for summers with a positive NAOI, 1960-1999	131
3.34	Cluster plot for summers with a negative NAOI, 1960-1999	134
3.35	Standard density plot for summers with a negative NAOI, 1960-1999	135
3.36	Geometrically corrected density plot for summers with a negative NAOI, 1960-1999	135
3.37	Cluster plot for falls with a positive NAOI, 1960-1999	138
3.38	Standard density plot for falls with a positive NAOI, 1960-1999 . . .	139
3.39	Geometrically corrected density plot for falls with a positive NAOI, 1960-1999	139

A.1	Cluster plot for the average period, 1960-1999	156
A.2	Membership plot, 1960-1999	157
A.3	Standard density membership plot, 1960-1999	158
A.4	Geometrically corrected density membership plot, 1960-1999	158
A.5	Membership plot, 1960-1999	159
A.6	Standard density membership plot, 1960-1999	160
A.7	Geometrically corrected density membership plot, 1960-1999	160
A.8	Membership plot, 1960-1999	161
A.9	Standard density membership plot, 1960-1999	162
A.10	Geometrically corrected density membership plot, 1960-1999	162
A.11	Membership plot, 1960-1999	163
A.12	Standard density membership plot, 1960-1999	164
A.13	Geometrically corrected density membership plot, 1960-1999	164
A.14	Membership plot, 1960-1999	165
A.15	Standard density membership plot, 1960-1999	166
A.16	Geometrically corrected density membership plot, 1960-1999	166
A.17	Membership plot, 1960-1999	167
A.18	Standard density membership plot, 1960-1999	168
A.19	Geometrically corrected density membership plot, 1960-1999	168
B.1	Cluster plot for the average winter, 1960-1999	170
B.2	Membership plot for winters, 1960-1999	171
B.3	Standard density membership plot for winters, 1960-1999	172
B.4	Geometrically corrected density membership plot for winters, 1960-1999	172
B.5	Membership plot for winters, 1960-1999	173
B.6	Standard density membership plot for winters, 1960-1999	174
B.7	Geometrically corrected density membership plot for winters, 1960-1999	174
B.8	Membership plot for winters, 1960-1999	175

B.9	Standard density membership plot for winters, 1960-1999	176
B.10	Geometrically corrected density membership plot for winters, 1960-1999	176
B.11	Membership plot for winters, 1960-1999	177
B.12	Standard density membership plot for winters, 1960-1999	178
B.13	Geometrically corrected density membership plot for winters, 1960-1999	178
B.14	Membership plot for winters, 1960-1999	179
B.15	Standard density membership plot for winters, 1960-1999	180
B.16	Geometrically corrected density membership plot for winters, 1960-1999	180
B.17	Membership plot for winters, 1960-1999	181
B.18	Standard density membership plot for winters, 1960-1999	182
B.19	Geometrically corrected density membership plot for winters, 1960-1999	182
B.20	Cluster plot for positive NAO winters, 1960-1999	184
B.21	Membership plot for winters with a positive NAOI value, 1960-1999	185
B.22	Standard density membership plot for winters with a positive NAOI value, 1960-1999	186
B.23	Geometrically corrected density membership plot for winters with a positive NAOI value, 1960-1999	186
B.24	Membership plot for winters with a positive NAOI value, 1960-1999	187
B.25	Standard density membership plot for winters with a positive NAOI value, 1960-1999	188
B.26	Geometrically corrected density membership plot for winters with a positive NAOI value, 1960-1999	188
B.27	Membership plot for winters with a positive NAOI value, 1960-1999	189
B.28	Standard density membership plot for winters with a positive NAOI value, 1960-1999	190
B.29	Geometrically corrected density membership plot for winters with a positive NAOI value, 1960-1999	190
B.30	Membership plot for winters with a positive NAOI value, 1960-1999	191

B.31	Standard density membership plot for winters with a positive NAOI value, 1960-1999	192
B.32	Geometrically corrected density membership plot for winters with a positive NAOI value, 1960-1999	192
B.33	Membership plot for winters with a positive NAOI value, 1960-1999	193
B.34	Standard density membership plot for winters with a positive NAOI value, 1960-1999	194
B.35	Geometrically corrected density membership plot for winters with a positive NAOI value, 1960-1999	194
B.36	Membership plot for winters with a positive NAOI value, 1960-1999	195
B.37	Standard density membership plot for winters with a positive NAOI value, 1960-1999	196
B.38	Geometrically corrected density membership plot for winters with a positive NAOI value, 1960-1999	196
B.39	Cluster plot for negative NAO winters, 1960-1999	198
B.40	Membership plot for winters with a negative NAOI value, 1960-1999	199
B.41	Standard density membership plot for winters with a negative NAOI value, 1960-1999	200
B.42	Geometrically corrected density membership plot for winters with a negative NAOI value, 1960-1999	200
B.43	Membership plot for winters with a negative NAOI value, 1960-1999	201
B.44	Standard density membership plot for winters with a negative NAOI value, 1960-1999	202
B.45	Geometrically corrected density membership plot for winters with a negative NAOI value, 1960-1999	202
B.46	Membership plot for winters with a negative NAOI value, 1960-1999	203
B.47	Standard density membership plot for winters with a negative NAOI value, 1960-1999	204

B.48	Geometrically corrected density membership plot for winters with a negative NAOI value, 1960-1999	204
B.49	Membership plot for winters with a negative NAOI value, 1960-1999	205
B.50	Standard density membership plot for winters with a negative NAOI value, 1960-1999	206
B.51	Geometrically corrected density membership plot for winters with a negative NAOI value, 1960-1999	206
B.52	Membership plot for winters with a negative NAOI value, 1960-1999	207
B.53	Standard density membership plot for winters with a negative NAOI value, 1960-1999	208
B.54	Geometrically corrected density membership plot for winters with a negative NAOI value, 1960-1999	208
B.55	Membership plot for winters with a negative NAOI value, 1960-1999	209
B.56	Standard density membership plot for winters with a negative NAOI value, 1960-1999	210
B.57	Geometrically corrected density membership plot for winters with a negative NAOI value, 1960-1999	210
C.1	Cluster plot for the average spring, 1960-1999	212
C.2	Membership plot for springs, 1960-1999	213
C.3	Standard density membership plot for springs, 1960-1999	214
C.4	Geometrically corrected density membership plot for springs, 1960-1999	214
C.5	Membership plot for springs, 1960-1999	215
C.6	Standard density membership plot for springs, 1960-1999	216
C.7	Geometrically corrected density membership plot for springs, 1960-1999	216
C.8	Membership plot for springs, 1960-1999	217
C.9	Standard density membership plot for springs, 1960-1999	218
C.10	Geometrically corrected density membership plot for springs, 1960-1999	218

C.11	Membership plot for springs, 1960-1999	219
C.12	Standard density membership plot for springs, 1960-1999	220
C.13	Geometrically corrected density membership plot for springs, 1960-1999	220
C.14	Membership plot for springs, 1960-1999	221
C.15	Standard density membership plot for springs, 1960-1999	222
C.16	Geometrically corrected density membership plot for springs, 1960-1999	222
C.17	Membership plot for springs, 1960-1999	223
C.18	Standard density membership plot for springs, 1960-1999	224
C.19	Geometrically corrected density membership plot for springs, 1960-1999	224
C.20	Cluster plot for positive NAO springs, 1960-1999	226
C.21	Membership plot for springs with a positive NAOI value, 1960-1999	227
C.22	Standard density membership plot for springs with a positive NAOI value, 1960-1999	228
C.23	Geometrically corrected density membership plot for springs with a positive NAOI value, 1960-1999	228
C.24	Membership plot for springs with a positive NAOI value, 1960-1999	229
C.25	Standard density membership plot for springs with a positive NAOI value, 1960-1999	230
C.26	Geometrically corrected density membership plot for springs with a positive NAOI value, 1960-1999	230
C.27	Membership plot for springs with a positive NAOI value, 1960-1999	231
C.28	Standard density membership plot for springs with a positive NAOI value, 1960-1999	232
C.29	Geometrically corrected density membership plot for springs with a positive NAOI value, 1960-1999	232
C.30	Membership plot for springs with a positive NAOI value, 1960-1999	233
C.31	Standard density membership plot for springs with a positive NAOI value, 1960-1999	234

C.32	Geometrically corrected density membership plot for springs with a positive NAOI value, 1960-1999	234
C.33	Membership plot for springs with a positive NAOI value, 1960-1999	235
C.34	Standard density membership plot for springs with a positive NAOI value, 1960-1999	236
C.35	Geometrically corrected density membership plot for springs with a positive NAOI value, 1960-1999	236
C.36	Membership plot for springs with a positive NAOI value, 1960-1999	237
C.37	Standard density membership plot for springs with a positive NAOI value, 1960-1999	238
C.38	Geometrically corrected density membership plot for springs with a positive NAOI value, 1960-1999	238
C.39	Cluster plot for negative NAO springs, 1960-1999	240
C.40	Membership plot for springs with a negative NAOI value, 1960-1999	241
C.41	Standard density membership plot for springs with a negative NAOI value, 1960-1999	242
C.42	Geometrically corrected density membership plot for springs with a negative NAOI value, 1960-1999	242
C.43	Membership plot for springs with a negative NAOI value, 1960-1999	243
C.44	Standard density membership plot for springs with a negative NAOI value, 1960-1999	244
C.45	Geometrically corrected density membership plot for springs with a negative NAOI value, 1960-1999	244
C.46	Membership plot for springs with a negative NAOI value, 1960-1999	245
C.47	Standard density membership plot for springs with a negative NAOI value, 1960-1999	246
C.48	Geometrically corrected density membership plot for springs with a negative NAOI value, 1960-1999	246

C.49	Membership plot for springs with a negative NAOI value, 1960-1999	247
C.50	Standard density membership plot for springs with a negative NAOI value, 1960-1999	248
C.51	Geometrically corrected density membership plot for springs with a negative NAOI value, 1960-1999	248
C.52	Membership plot for springs with a negative NAOI value, 1960-1999	249
C.53	Standard density membership plot for springs with a negative NAOI value, 1960-1999	250
C.54	Geometrically corrected density membership plot for springs with a negative NAOI value, 1960-1999	250
C.55	Membership plot for springs with a negative NAOI value, 1960-1999	251
C.56	Standard density membership plot for springs with a negative NAOI value, 1960-1999	252
C.57	Geometrically corrected density membership plot for springs with a negative NAOI value, 1960-1999	252
D.1	Cluster plot for the average summer, 1960-1999)	254
D.2	Membership plot for summers, 1960-1999)	255
D.3	Standard density membership plot for summers, 1960-1999)	256
D.4	Geometrically corrected density membership plot for summers, 1960-1999	256
D.5	Membership plot for summers, 1960-1999	257
D.6	Standard density membership plot for summers, 1960-1999	258
D.7	Geometrically corrected density membership plot for summers, 1960-1999	258
D.8	Membership plot for summers, 1960-1999	259
D.9	Standard density membership plot for summers, 1960-1999	260

D.10	Geometrically corrected density membership plot for summers, 1960-1999	260
D.11	Membership plot for summers, 1960-1999	261
D.12	Standard density membership plot for summers, 1960-1999	262
D.13	Geometrically corrected density membership plot for summers, 1960-1999	262
D.14	Membership plot for summers, 1960-1999	263
D.15	Standard density membership plot for summers, 1960-1999	264
D.16	Geometrically corrected density membership plot for summers, 1960-1999	264
D.17	Membership plot for summers, 1960-1999	265
D.18	Standard density membership plot for summers, 1960-1999	266
D.19	Geometrically corrected density membership plot for summers, 1960-1999	266
D.20	Cluster plot for positive NAO summers, 1960-1999	268
D.21	Membership plot for summers with a positive NAOI value, 1960-1999	269
D.22	Standard density membership plot for summers with a positive NAOI value, 1960-1999	270
D.23	Geometrically corrected density membership plot for summers with a positive NAOI value, 1960-1999	270
D.24	Membership plot for summers with a positive NAOI value, 1960-1999	271
D.25	Standard density membership plot for summers with a positive NAOI value, 1960-1999	272
D.26	Geometrically corrected density membership plot for summers with a positive NAOI value, 1960-1999	272
D.27	Membership plot for summers with a positive NAOI value, 1960-1999	273
D.28	Standard density membership plot for summers with a positive NAOI value, 1960-1999	274

D.29	Geometrically corrected density membership plot for summers with a positive NAOI value, 1960-1999	274
D.30	Membership plot for summers with a positive NAOI value, 1960-1999	275
D.31	Standard density membership plot for summers with a positive NAOI value, 1960-1999	276
D.32	Geometrically corrected density membership plot for summers with a positive NAOI value, 1960-1999	276
D.33	Membership plot for summers with a positive NAOI value, 1960-1999	277
D.34	Standard density membership plot for summers with a positive NAOI value, 1960-1999	278
D.35	Geometrically corrected density membership plot for summers with a positive NAOI value, 1960-1999	278
D.36	Membership plot for summers with a positive NAOI value, 1960-1999	279
D.37	Standard density membership plot for summers with a positive NAOI value, 1960-1999	280
D.38	Geometrically corrected density membership plot for summers with a positive NAOI value, 1960-1999	280
D.39	Cluster plot for negaive NAO summers, 1960-1999	282
D.40	Membership plot for summers with a negative NAOI value, 1960-1999	283
D.41	Standard density membership plot for summers with a negative NAOI value, 1960-1999	284
D.42	Geometrically corrected density membership plot for summers with a negative NAOI value, 1960-1999	284
D.43	Membership plot for summers with a negative NAOI value, 1960-1999	285
D.44	Standard density membership plot for summers with a negative NAOI value, 1960-1999	286
D.45	Geometrically corrected density membership plot for summers with a negative NAOI value, 1960-1999	286

D.46	Membership plot for summers with a negative NAOI value, 1960-1999	287
D.47	Standard density membership plot for summers with a negative NAOI value, 1960-1999	288
D.48	Geometrically corrected density membership plot for summers with a negative NAOI value, 1960-1999	288
D.49	Membership plot for summers with a negative NAOI value, 1960-1999	289
D.50	Standard density membership plot for summers with a negative NAOI value, 1960-1999	290
D.51	Geometrically corrected density membership plot for summers with a negative NAOI value, 1960-1999	290
D.52	Membership plot for summers with a negative NAOI value, 1960-1999	291
D.53	Standard density membership plot for summers with a negative NAOI value, 1960-1999	292
D.54	Geometrically corrected density membership plot for summers with a negative NAOI value, 1960-1999	292
D.55	Membership plot for summers with a negative NAOI value, 1960-1999	293
D.56	Standard density membership plot for summers with a negative NAOI value, 1960-1999	294
D.57	Geometrically corrected density membership plot for summers with a negative NAOI value, 1960-1999	294
E.1	Cluster plot for the averag fall, 1960-1999	296
E.2	Membership plot for falls, 1960-1999	297
E.3	Standard density membership plot for falls, 1960-1999	298
E.4	Geometrically corrected density membership plot for falls, 1960-1999	298
E.5	Membership plot for falls, 1960-1999	299
E.6	Standard density membership plot for falls, 1960-1999	300
E.7	Geometrically corrected density membership plot for falls, 1960-1999	300

E.8	Membership plot for falls, 1960-1999	301
E.9	Standard density membership plot for falls, 1960-1999	302
E.10	Geometrically corrected density membership plot for falls, 1960-1999	302
E.11	Membership plot for falls, 1960-1999	303
E.12	Standard density membership plot for falls, 1960-1999	304
E.13	Geometrically corrected density membership plot for falls, 1960-1999	304
E.14	Membership plot for falls, 1960-1999	305
E.15	Standard density membership plot for falls, 1960-1999	306
E.16	Geometrically corrected density membership plot for falls, 1960-1999	306
E.17	Membership plot for falls, 1960-1999	307
E.18	Standard density membership plot for falls, 1960-1999	308
E.19	Geometrically corrected density membership plot for falls, 1960-1999	308
E.20	Cluster plot for the positive NAO falls, 1960-1999	310
E.21	Membership plot for falls with a positive NAOI value, 1960-1999 . .	311
E.22	Standard density membership plot for falls with a positive NAOI value, 1960-1999	312
E.23	Geometrically corrected density membership plot for falls with a pos- itive NAOI value, 1960-1999	312
E.24	Membership plot for falls with a positive NAOI value, 1960-1999 . .	313
E.25	Standard density membership plot for falls with a positive NAOI value, 1960-1999	314
E.26	Geometrically corrected density membership plot for falls with a pos- itive NAOI value, 1960-1999	314
E.27	Membership plot for falls with a positive NAOI value, 1960-1999 . .	315
E.28	Standard density membership plot for falls with a positive NAOI value, 1960-1999	316
E.29	Geometrically corrected density membership plot for falls with a pos- itive NAOI value, 1960-1999	316

E.30	Membership plot for falls with a positive NAOI value, 1960-1999 . .	317
E.31	Standard density membership plot for falls with a positive NAOI value, 1960-1999	318
E.32	Geometrically corrected density membership plot for falls with a pos- itive NAOI value, 1960-1999	318
E.33	Membership plot for falls with a positive NAOI value, 1960-1999 . .	319
E.34	Standard density membership plot for falls with a positive NAOI value, 1960-1999	320
E.35	Geometrically corrected density membership plot for falls with a pos- itive NAOI value, 1960-1999	320
E.36	Membership plot for falls with a positive NAOI value, 1960-1999 . .	321
E.37	Standard density membership plot for falls with a positive NAOI value, 1960-1999	322
E.38	Geometrically corrected density membership plot for falls with a pos- itive NAOI value, 1960-1999	322
E.39	Cluster plot for the negative NAO falls, 1960-1999	324
E.40	Membership plot for falls with a negative NAOI value, 1960-1999 . .	325
E.41	Standard density membership plot for falls with a negative NAOI value, 1960-1999	326
E.42	Geometrically corrected density membership plot for falls with a neg- ative NAOI value, 1960-1999	326
E.43	Membership plot for falls with a negative NAOI value, 1960-1999 . .	327
E.44	Standard density membership plot for falls with a negative NAOI value, 1960-1999	328
E.45	Geometrically corrected density membership plot for falls with a neg- ative NAOI value, 1960-1999	328
E.46	Membership plot for falls with a negative NAOI value, 1960-1999 . .	329

E.47	Standard density membership plot for falls with a negative NAOI value, 1960-1999	330
E.48	Geometrically corrected density membership plot for falls with a negative NAOI value, 1960-1999	330
E.49	Membership plot for falls with a negative NAOI value, 1960-1999 . .	331
E.50	Standard density membership plot for falls with a negative NAOI value, 1960-1999	332
E.51	Geometrically corrected density membership plot for falls with a negative NAOI value, 1960-1999	332
E.52	Membership plot for falls with a negative NAOI value, 1960-1999 . .	333
E.53	Standard density membership plot for falls with a negative NAOI value, 1960-1999	334
E.54	Geometrically corrected density membership plot for falls with a negative NAOI value, 1960-1999	334
E.55	Membership plot for falls with a negative NAOI value, 1960-1999 . .	335
E.56	Standard density membership plot for falls with a negative NAOI value, 1960-1999	336
E.57	Geometrically corrected density membership plot for falls with a negative NAOI value, 1960-1999	336

List of Tables

2.1	Summary of NCEP/NCAR Reanalysis data	17
2.2	Removal of incomplete trajectories	28
2.3	Summary of NAOI sites	81
2.4	Summary of periods of positive and negative NAO	82
3.1	Summary clusters for the whole study period	91
3.2	Summary clusters for the average springs	95
3.3	Summary clusters for the positive springs	99
3.4	Summary clusters for the negative springs	103
3.5	Summary clusters for the negative falls	107
3.6	Summary clusters for the average winters	112
3.7	Summary clusters for the average summers	116
3.8	Summary clusters for the average fall	120
3.9	Summary clusters for positive NAO winters	124
3.10	Summary clusters for the negative winters	128
3.11	Summary clusters for the positive summers	132
3.12	Summary clusters for the negative summers	136
3.13	Summary clusters for the positive falls	140
3.14	Summary of source regions associated with each cluster	142

Chapter 1

Introduction

Atmospheric measurements are best interpreted when the processes that transport air to a measurement site are understood. Atmospheric trajectories are a common method used to determine some of these processes. Trajectories that are concurrent with measurements can determine the conditions that lead to observed transport events, while a suite of trajectories spanning a long time period can help to determine the climatic conditions that transport air to a location.

The PICO-NARE station, located atop an inactive volcano (Pico mountain) in the Azores Islands (38.47°N, 28.4°W) is currently measuring several important trace gases (CO, O₃, NO, NO_x, and NO_y). A concurrent set of trajectories is being made that will aid the interpretation of those measurements. However, there is not a thorough understanding of the climatic transport to the site. This study will use a 40-year set of trajectories and several statistical techniques to determine the long-term and short-term climatic variation of transport to the site.

This chapter will briefly discuss the several issues relating to the importance the measurements being made at the PICO-NARE site and the use of atmospheric trajectories to characterize the flow of air to the site. This discussion will begin with the importance of air pollution, primarily in the context of O₃, followed by a description

of the mechanisms that transport pollutants to regions downwind of source regions and a review of some of the previous research done to understand the process and the impacts of this transport. It will then highlight the PICO-NARE station and elucidate how measurements made at the site can contribute to these research efforts. Finally, it will clarify the need for an understanding of the climatic patterns of transport to the site and state the objectives of this study, which will fulfill the need for a description of the climatic transport of air to the site.

1.1 Background

Anthropogenic emissions have been found to have profound impacts on the environment on scales ranging from local (*e.g.*, smog episodes) to global (*e.g.*, stratospheric ozone reduction). While the impact and processes governing the impact of some of these emissions are well understood (*e.g.*, the photo-dissociation of CFC's in the stratosphere that leads to stratospheric ozone destruction in the polar regions), others are only partially understood (*e.g.*, the impact of greenhouse gases such as CO₂ on global climate). Extensive research has been and is being conducted worldwide to further the understanding of atmospheric processing of anthropogenic emissions.

One particular compound that has been of interest for over half a century is ozone (O₃). When O₃ was discovered in the 1840's, its ability to destroy gases that were believed to be responsible for epidemic disease led scientists to believe that its presence in the lower atmosphere was positive, despite its noted ability to damage both plant and animal tissue [Meyer, 1996]. In 1952, A. J. Haagen-Smit discovered that O₃ was present in high levels in photochemical smog [O'Brien, 1992; Meyer, 1996]. It was found that ozone was produced by the photochemical reactions between nitrogen oxides (NO+NO₂=NO_x), volatile organic compounds (VOC's), and oxygen (O₂), which were driven by sunlight [USEPA, 2000, 2002; Cooper and Alley, 1999; Finlayson-Pitts and Pitts, 1999]. At relatively low levels, O₃ was found to have numerous harmful

effects on humans, including general discomfort and difficulty breathing, aggravation of asthma, increase susceptibility to respiratory infections, and temporary and permanent lung damage [USEPA, 2000, 2002; Cooper and Alley, 1999]. Low concentrations of O_3 were also found to affect plants (*e.g.*, decreases in agricultural yields induced by O_3) [Meyer, 1996; Krupa and Jager, 1996; Holmes, 1997]. In addition its direct impacts, O_3 is a greenhouse gas and increased production of O_3 has the potential to impact the global climate. Because of the increasing problems of air pollution, the U.S. government took steps to improve air quality. The U.S. passed the first Clean Air Act in 1963, and in 1970 the better known amendments were passed, along with the formation of the U.S. Environmental Protection Agency (EPA) [Nizich *et al.*, 2000]. Since the passage of the 1970 amendments, the general air quality has improved and there has been a decrease in the aggregate emission and production of the six criteria pollutants (-25%) covered by the 1970 amendments, which includes NO_2 and O_3 , despite an increase in the U.S. gross domestic product (161%), energy consumption (42%), and vehicle miles traveled (149%) [USEPA, 2002; Meyer, 1996].

1.2 Recent and current research

Initially it was believed that air pollution was a local problem, as smog events only occurred in and around heavily populated or industrialized regions. However, evidence was found to indicate that O_3 and O_3 precursors could be transported from source regions to regions downwind [USEPA, 2002; Meyer, 1996; Moser *et al.*, 1992; Ottar *et al.*, 1984]. These emissions primarily occur in the Northern Hemisphere (NH), because the majority of the world's population and the majority of the world's industrialized nations are in the NH. Since mixing between hemispheres is significantly less than mixing within hemispheres, research on the transport of the anthropogenic emissions that produce ozone (as well as general research of anthropogenic emissions) are best focused in the NH. In the mid-latitudes, the general motion of the air is

dominated by flow from west to east, which means that pollution is transported from west to east. If regional transport within the mid-latitudes is considered on a continental basis, then intercontinental, long-range transport can occur from Europe to Asia, from Asia to North America, or from North America to Europe. Each of these transport scenarios is unique, as the boundaries between the pairs are unique. Europe and Asia are continuous land masses but they are separated by the Ural Mountains, which can interrupt the flow patterns. The Pacific Ocean separates Asia and North America. Similarly North America and Europe are separated by the Atlantic Ocean.

Transport over oceans presents the potential for removal of pollution from the atmosphere. Air within the marine boundary layer can remove pollution in one of three ways. First, it is typically very humid, which allows scrubbing of pollution by scavenging of pollutants by water followed by wet deposition. Second, the halogenated compounds typically present in the marine boundary layer can undergo photochemically induced reactions that remove ozone. In addition, during the daytime OH is present in the marine boundary layers at relatively high concentrations, which is the primary oxidizing species in the troposphere and able to remove pollution effectively [Finlayson-Pitts and Pitts, 1999]. However, these removal processes do not preclude the possibility of transport within the boundary layer. Transport can also occur above the boundary layer, in the free troposphere (FT), which is generally dry. Therefore, air that travels over the ocean could be clean due to removal in the boundary layer, or could remain polluted due to transport in drier air above the FT.

Of these transport scenarios, the transport of North American pollution across the Atlantic to Europe has had and continues to be the focus of a well coordinated research effort. This effort spans back to at least the late 1980's. In 1987, the Atmosphere/Ocean Chemistry Experiment (AEROCE) project began measurements of aerosols at four locations in the Barbados, Bermuda Islands, the Canary Islands, and Ireland, which are ongoing [Prospero and Jennings]. In 1988, the Global Change Expedition/Coordinated Air-Sea Experiment/Western Atlantic Ocean Experiment

(CCE/CASE/WATOX) was conducted, which made an initial investigation into the interaction of polluted air masses over the Atlantic [Parrish, 2001]. Starting in 1991 with a series of measurements of carbon monoxide (CO) and O₃ made along the eastern coast of Canada (Seal Island, Sable Island, and Cape Race), the North Atlantic Regional Experiment (NARE) began [Parrish *et al.*, 1993; Parrish, 2001]. The NARE program is still ongoing, although it is now part of the Intercontinental Transport and Chemical Transformation (ITCT) program. ITCT has combined the NARE program with the East Asian/North Pacific Regional Experiment (APARE), which was similar to the NARE program, but focused on the North Pacific region. Several intensive studies have been conducted under the direction of the NARE program (*e.g.*, NARE 93 summer intensive), as well as intermittent studies (*e.g.*, NARE 97), and relatively continuous ground based measurements (*e.g.*, Sable Island, Nova Scotia from 1991-1995) in and around the North Atlantic [Fehsenfeld *et al.*, 1996; Roberts *et al.*, 1996; Peterson *et al.*, 2001; Stohl *et al.*, 2002b]. The NARE program is part of a larger research collaboration known as the International Global Atmospheric Chemistry (IGAC) project. IGAC focuses on understanding the biogeochemical processes between air, water, and land and to understand how they affect global air quality. One of the most recent NARE projects is the Pico International Chemical Observatory (PICO-NARE), which is based in the Azores islands in the central North Atlantic.

1.3 PICO-NARE

The Azores islands, located in the central North Atlantic Ocean, are the sole land mass in the mid-latitudes of the North Atlantic ocean (see Figure 1.1). Due to their location, the Azores islands have long been important to human activities. They have been historically important for transportation, as the islands were used as supply stops for ships and refueling stops for aircraft. They have also had historical

scientific importance. Pressure measurements from the Azores dating back to the mid-1800's have been used to help determine climate conditions in the region before standard meteorological measurements were widely made [Jonsson and Miles, 2001; Wanner *et al.*, 2001]. The Azores also provided key meteorological data for use in weather forecasts before the use of satellite data [Honrath and Fialho, 2001]. The Azores have been the base of studies of the atmospheric processes over the Atlantic, including previous NARE studies. For example, ground based measurements, soundings, and aircraft measurements were made in and around the Azores as part of the NARE 93 campaign [Penkett *et al.*, 1998].

The ground based measurements of the NARE 93 campaign primarily sampled air from the marine boundary layer (MBL). However, long transport tends to occur in the FT [Peterson *et al.*, 1998; Parrish *et al.*, 1998; Peterson *et al.*, 2001; Honrath and Fialho, 2001]. The PICO-NARE station is positioned atop Pico mountain, an inactive volcanic island in the Azores (38.47 °N, 28.4 °W, 2225 m ASL). The mountain is the highest point in the Azores, and is believed to be frequently in the FT. Because of the location of the Azores and the placement of the PICO-NARE station on the highest point in the Azores, the station represents a unique location for ground based measurements, as it is likely that it will be able to regularly sample air from the lower FT, and thus sample air affected by long range transport of polluted air.

The PICO-NARE station was installed during the summer of 2001, and in July of 2001, measurements of CO, O₃, aerosol black carbon, wind direction and speed, relative humidity, temperature, and pressure began. As mentioned previously, O₃ is of primary interest in the lower atmosphere because it is a greenhouse gas as well as its deleterious effects on human health and the environment. In the troposphere, O₃ is also important because it is a primary constituent in tropospheric photochemistry [Parrish *et al.*, 1998]. Both CO and black carbon are relatively inert and are useful in determining plume characteristics (*e.g.*, plume age, origin) [Stohl *et al.*, 2002a]. In July 2002 measurements of NO_x and total reactive odd nitrogen (NO_y) also began.

These compounds are ozone precursors and are able to undergo long-range transport. This transport can then result in the formation of O_3 downwind of emission regions [Peterson *et al.*, 1998; Stohl *et al.*, 2002a]. Plans also exist for the installation of measurements of non-methane hydrocarbons (NMHC's) in the spring of 2004 [personal communication, D. Helmig, Univ. Colorado, 2003]. NMHC's are also O_3 precursors, as they are involved in the photochemical reactions with NO_x that produce O_3 . Some of the goals of the PICO-NARE study are to characterize background levels of the measured species in the region, to characterize the effects of transport events from polluted regions on the levels of those species, and to assess the production of O_3 in the region due to the transport of O_3 and O_3 precursors. However, measurements of O_3 and O_3 precursors alone are an insufficient means to accomplish these goals. In order to determine the original constituents of the air and what processes it has undergone during transport, the measurements must be coupled with some means to determine where a measured air mass has been. While the measurements of CO and black carbon are helpful in recognizing polluted air masses and determining some of its history, they are insufficient in fully determining the complete history of the sampled air. Atmospheric back trajectories have been used extensively in conjunction with atmospheric measurements for this purpose, as they are often able to model the motion of the atmosphere and reveal the history of an air mass.

1.4 Atmospheric trajectories in a climatology

Atmospheric trajectories are the Lagrangian view of an infinitesimally small air mass. That is, a trajectory follows an air mass as it travels through the atmosphere. It is easy to see how a history of an air mass could be useful in examining measurements of trace gases. The analysis of measurements of polluted air and a record of an air parcel's history (trajectory) can help determine where the pollution originated (*e.g.*, over the eastern U.S. coast), how long ago the parcel passed over a pollution source region

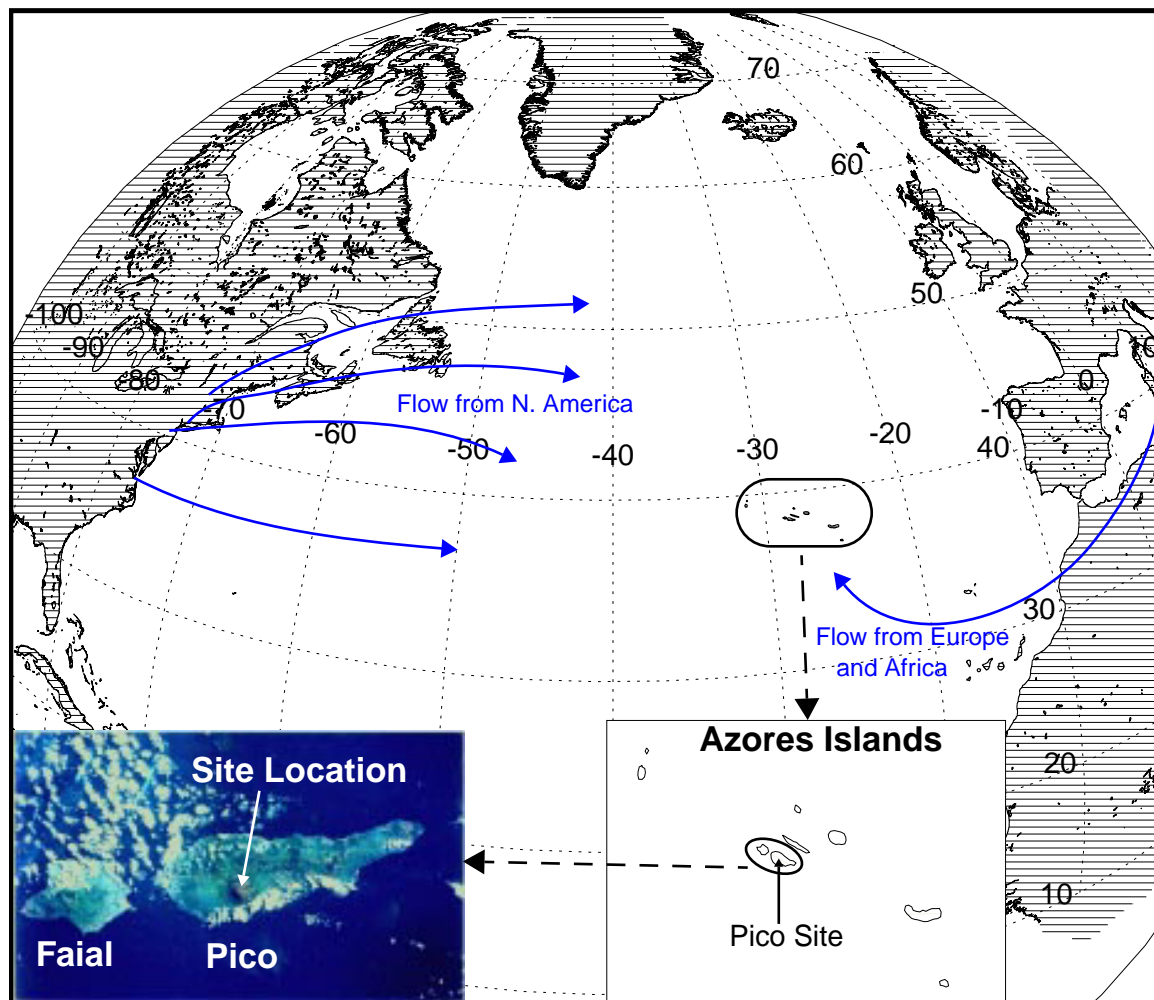


Figure 1.1 Map of the central North Atlantic, showing the location of the Azores and the station on Pico Island.

(*e.g.*, 4 days), and some of the processes the parcel might have undergone during its transport (*e.g.*, scrubbing by water due to time spent in the marine boundary layer). Because trajectories can be used in this manner, they have long been employed to help interpret measurements of atmospheric species. Applications of trajectories of this type are often termed source/receptor studies, as they attempt to determine sources for air received at the measurement site. A few examples of these source/receptor studies for ground based measurements of atmospheric species include those of *Mer-*

rill et al. [1985], *Moody and Galloway* [1988], *Harris et al.* [1992], *Merrill and Moody* [1996], *Draxler* [1996], and *Fast and Berkowitz* [1997]. This method can also be applied to aid in the interpretation of aircraft measurements (*e.g.*, the study of *Vaughan et al.* [2002], which included flights out of the Azores).

Trajectories can also be used to develop average trends of air flow and the transport of important species when a long enough record of measurements and trajectories are available (*e.g.*, *Sirois and Bottenheim*, 1995; *Moody et al.*, 1995]). The resultant trends are termed climatologies. It can also be useful to develop a climatology for measurement sites without coupling the trajectories with a concurrent set of atmospheric measurements. These types of climatologies can be used to understand the characteristics of average flow to a site. An example of a trajectory climatology being used with a non-concurrent set of atmospheric data is *Kahl et al.* [1997], who developed a 44-year climatology to better understand ice cores samples and limited atmospheric measurements taken in Summit, Greenland. Other examples of studies that were made primarily to support a site used for atmospheric measurements are *Miller* [1981], who developed a 5-year climatology for Barrow, AK, and *Harris and Kahl* [1990], who developed 9-year climatology of transport to the the Mauna Loa Observatory. Both locations have hosted a series of atmospheric measurements, though the measurements were not necessarily conducted during a synchronous time with the climatologies. Once measurements are made, they can be compared to the climatologies to assess how often observed events might occur, to assess the impact of inputs from certain regions due to transport, and to extend the analysis measurements with concurrent trajectories to develop a climatology of the measured species.

1.5 Study objectives

The goal of this study is to develop a climatology for use in conjunction with current and future analysis of PICO-NARE measurements. Specifically, the goals are:

- to develop a 40-year set of 10-day back trajectories,
- to identify the primary flow paths of the trajectory set using cluster analysis,
- to identify source regions that contribute to the air that flows to the site using air-flow probability analysis, and
- to determine the impact of short-term and long-term climatic variation on those flow-paths and source regions.

The methods for achieving these goals are presented in the following chapter, and the results of the application of those methods are presented in the final chapter.

Chapter 2

Methods

This chapter explains the methods used in this study to create a climatology for transport to the PICO-NARE site. These methods include the creation of atmospheric trajectories, the clustering of the trajectories, performance of an airflow probability analysis, and the steps taken to assess the impact of the North Atlantic Oscillation.

2.1 Atmospheric trajectories

Atmospheric trajectories have been used extensively in studies relating to atmospheric processes. They are particularly useful in determining possible flow paths and source regions of air flow. This study employs the use of 10-day, 3-D kinematic backward trajectories to develop a climatology for the PICO-NARE site.

2.1.1 Trajectory use

The use of atmospheric trajectories is well established in the scientific community. The applications vary from exploring synoptic meteorology and source-receptor relationships, to determining natural transport phenomena such as dust transport, and to development of a climatology. A few examples of such applications include that

of *Merrill and Moody* [1996], who used both isentropic and kinematic, 3-dimensional (3-D) trajectories in conjunction with synoptic charts to examine the synoptic meteorology and transport during the North Atlantic Regional Experiment (NARE) intensive from August 1 to September 13, 1993. *Draxler* [1996] also examined the synoptic conditions during the NARE intensive by determining transport to Chebogue Point, Nova Scotia using kinematic back trajectories. Another example is the work of *Vaughan et al.* [2002], who conducted aerosol backscatter measurements on a series of flights and employed kinematic, 3-D backward trajectories to explore the origins of air masses in which measurements were taken. Transport phenomena were explored by *Merrill et al.* [1985] who used isentropic trajectories to trace the transport of continental material, mainly dust, from Asia to the Marshall Islands in the Pacific Ocean. Finally, an interesting application of trajectories is given by *Draxler* [1996], who developed a method for planning balloon flights using kinematic trajectories. These are a small sample of the possible applications of trajectories.

An important feature of trajectories is manifested in these examples. Two types of trajectories are mentioned above: isentropic and kinematic, 3-D. These terms refer to the assumption made about vertical motion during the calculation of the trajectory. An isentropic trajectory assumes that the vertical motion of the air follows surfaces of constant entropy, implying adiabatic flow. A kinematic or 3-D trajectory assumes vertical motion follows vertical wind fields. Another common trajectory type is isobaric, which is based on the assumption that motion follows surfaces of constant pressure. The isobaric trajectory is the simplest of the three types. The calculation of pressure surfaces is usually fairly simple compared to the difficulty of calculating calculation of isentropic surfaces. In addition to being the most simple, they are often regarded as the most inaccurate, as the assumption of vertical motion is poor [*Kahl et al.*, 1989]. The isentropic trajectories are more complex than isobaric in that surfaces of constant entropy (isentropes) are more difficult to compute than surfaces of constant pressure (isobars). The calculation of isentropic surfaces requires

more meteorological data than is required for the calculation of pressure surfaces and is computationally more intensive. However, the assumption of vertical motion for isentropic trajectories is much more accurate. The isentropic assumption, however, tends to break down in regions of strong wind gradients, such as fronts [*Kahl et al.*, 1989]. The most accurate but also the most computationally expensive trajectory type is the 3-D kinematic trajectory. These trajectories use horizontal and vertical winds and pressure and temperature gradients [*Kahl et al.*, 1989; *Draxler and Hess*, 1998a,b]. Comparisons have been made among the three trajectory types and it is generally accepted that isentropic trajectories are superior to isobaric and that 3-D kinematic trajectories are superior to the other two types.

2.1.2 Description of the HYSPLIT Model

The trajectory type is dependent upon the model used to calculate it. In general, the calculation of a trajectory requires two things, a trajectory model and a set of meteorological variables. The trajectory model will be described in this section and the meteorological data set will be described in the next sections. The trajectory model used in this study was the Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) model, version 4. The HYSPLIT model is the result of a joint effort between the Atmospheric Research Laboratories (ARL), a division of the National Oceanic and Atmospheric Administration (NOAA), and Australia’s Bureau of Meteorology (BoM). The model was designed to support the BoM’s need to respond to emergencies that have atmospheric implications, such as the release of radiological material [*Draxler and Hess*, 1998b]. The original model used only rawinsonde observations with basic assumptions of mixing and dispersion. The most recent version of the model is able to use a variety of gridded meteorological data sets and employs a more accurate set of equations to model dispersion and mixing [*Draxler and Hess*, 1998a]. This model is described in detail elsewhere [*NOAA*, 1997; *Draxler and Hess*,

1998a,b], but a few technical notes will be provided here.

The HYSPLIT model is capable of computing multiple simultaneous trajectories as well as dispersion and deposition of concentrations of airborne particles. This study made use of only the trajectory component of the model. The trajectories are 3-D kinematic trajectories; that is they are computed with the assumption that an air parcel follows the horizontal and vertical wind fields. HYSPLIT is also capable of computing isotropic, isosigma (terrain-following surfaces), sigma (terrain-following surfaces of constant pressure) isobaric, and isopycnic (surfaces of equal density) trajectories. The wind fields are computed using data from gridded meteorological fields. HYSPLIT is capable of using multiple meteorological fields simultaneously. This can be useful when computing trajectories that travel over areas for which there is sparse data, such as over oceans, and areas for which there is much data, such as over parts of Europe or the U.S. Typically the meteorological data in the areas sparse in data only have courser data fields, while the meteorological data sets over data-rich areas can have finer data fields.

At a minimum, the HYSPLIT model requires U and V (the horizontal wind components), T (temperature), Z (height) or P (pressure), and P_0 (the surface pressure) in the meteorological data to compute trajectories [Draxler and Hess, 1998a]. In order to utilize multiple meteorological data sets, HYSPLIT must transform the meteorological vertical coordinates into compatible forms. To do this, the profiles at each horizontal grid point are re-mapped by linear interpolation to a terrain-following (σ) coordinate system using [Draxler and Hess, 1998a; Draxler, 1996]:

$$\sigma = 1 - \frac{z}{Z_{top}}. \quad (2.1)$$

Where z is the height above ground level and Z_{top} is the top of the model domain. The advection of a trajectory from an initial position P_t to the first-guess position $P'_{t+\Delta t}$ uses the averaged 3-D velocity vectors for the two points. The initial guess is:

$$P'_{t+\Delta t} = P_t + V_{P,t}\Delta t. \quad (2.2)$$

The final position is then:

$$P_{t+\Delta t} = P_t + 0.5 \left(V_{P,t} + V_{P',t+\Delta t} \right) \Delta t. \quad (2.3)$$

The second guess returns a solution that more closely matches the forward velocities at point P_t to the backward velocities at $P_{t+\Delta t}$ [McQueen and Draxler, 1994; Draxler, 1996; Draxler and Hess, 1998a]. A trajectory is terminated if it exits the model top. However, trajectories are allowed to follow the ground if the earth's surface is intersected, an improvement over more basic models. The time step (Δt) can vary in each step, with the constraint that the maximum distance traveled per time step must be less than 0.75 of the grid spacing of the meteorological data set [Draxler and Hess, 1998a,b]. For this study, the time step was set so that 0.01 of the grid spacing of the meteorological data set was traveled, the smallest time step allowed by HYSPLIT.

The trajectories used in this study are 10-day backward trajectories, meaning the trajectory traces the path of an air mass 10 days backward in time from the endpoint at the PICO-NARE site. The results of a trajectory calculation consist of a series of points along the trajectory each hour, referred to as endpoints. Each endpoint includes information about the location of the trajectory, including the latitudinal and longitudinal positions, the altitude in meters, the pressure in hPa, the temperature, the potential temperature, the total column rainfall, and the estimated height of the boundary layer at that location. There is an endpoint for each hour and the trajectories run backward for 10-days, so that an individual trajectory usually consists of 240 endpoints along the travel path, plus the point at the PICO-NARE site, for a total of 241 endpoints. Some trajectories contained fewer endpoints because they were terminated early, either because they exited the model top or due to gaps in the meteorological data set. Trajectories were calculated at the ending time of 00,

06, 12, and 18 coordinated universal time (UTC) for the 40-year period 1960-1999.

2.1.3 Description of the NCEP/NCAR Reanalysis Meteorological Data

The meteorological data used for this study is referred to as the NCEP/NCAR Reanalysis, which is the result of a joint effort by the National Centers for Environmental Prediction (NCEP) and the National Center for Atmospheric Research (NCAR). The goal of the Reanalysis project was to produce set of global meteorological data set using a “frozen” system for transforming raw data into the meteorological variables for use in support of research and climate monitoring communities. The system is referred to as frozen because the methods used for transforming the raw data into the gridded meteorological variables remain unchanged throughout the analysis. This frozen system is used to help reduce changes in the analyzed climate induced by changes in the methods used to analyze the raw data. A modern quality control system was employed to reduce changes in the analyzed climate induced by changes in the climatic observation system. The project is both a historical reanalysis of archived data dating back to 1948 and an on-going project using current data. One of the goals of the project was to incorporate data from as many sources as possible in order to create an analyzed climate that is more complete than previous analysis of the historical data. The sources for the analysis include upper air rawinsonde observations of temperature, horizontal wind and specific humidity, vertical temperature soundings from satellite data, temperature soundings over land, cloud tracked winds from geostationary satellites, aircraft observations of wind and temperature, land surface data of surface pressure and oceanic data of surface pressure and temperature, horizontal wind and specific humidity from a variety of international resources [Kistler *et al.*, 1999]. The use of extensive data and the modern quality control procedures produces gridded meteorological fields that are superior to analyses of the older data. However,

despite the quality control systems, errors in the analysis have been found. Known errors are discussed in the descriptive publications of the reanalysis [*Kalnay et al.*, 1996; *Kistler et al.*, 1999], on the NOAA's Climate Diagnostics Center (CDC) web site (<http://www.cdc.noaa.gov/cdc/reanalysis/problems.shtml>), and in publications of users of the reanalysis [*Trenberth and Stepaniak*, 2002].

The resulting set of meteorological variables are recorded on a 2.5° by 2.5° global grid, every 6 hours, at 00, 06, 12, and 18 UTC. The data include surface values and values on 17 vertical pressure levels, at 1000, 925, 850, 700, 600, 500, 400, 300, 250, 200, 150, 100, 70, 50, 30, 20, and 10 hPa. Numerous variables are available in various formats for download from the web. The data used for this analysis was a pre-packaged set of the reanalysis data developed specifically for use with HYSPLIT. The data included in the package are summarized in Table 2.1. The data packaged for use with HYSPLIT are available for download from NOAA's Atmospheric Research Laboratories (ARL) FTP site (<ftp://www.arl.noaa.gov>). All the reanalysis data are available for download from the CDC web-site (<http://www.cdc.noaa.gov/cdc/data.ncep.reanalysis.html>).

Table 2.1 Summary of the data included in the pre-packaged NCEP/NCAR Reanalysis data used by HYSPLIT for trajectory calculation.

Variable	Level(s)	Units	Least significant digit
Air temperature	Surface and SLP*	$^\circ\text{K}$	0.1
U and V winds	Surface and SLP	m/s	0.1
Geopotential height	SLP	m	1
Omega (vertical velocity)	SLP	Pascal/s	0.001
Pressure	Surface	Pascal	10
Precipitation rate	Total in column	$\text{Kg}/\text{m}^2/\text{s}$	0.000001
Relative humidity	Surface and SLP	%	1.
Standard pressure levels			

2.1.4 Trajectory error

Just as with other data, when using trajectories it is important to take into account possible errors that may result from measurement, calculation, or application methods. Trajectory error is defined as the difference between a calculated point of a trajectory and the real location of the modeled air. Trajectory errors are often referred to in two forms: absolute and relative. Absolute trajectory error is the distance between a calculated point along a trajectory to the actual location of the modeled air, while relative trajectory error is the absolute trajectory error divided by the total distance traveled by the trajectory to reach that point. Trajectories are susceptible to errors in the meteorological data and from the model used in the calculation of the trajectories. Meteorological errors can be the result of errors in the measurements to produce gridded meteorological fields, errors in the interpolation of meteorological variables from measurement sites to output grid-points, and incorrect assumptions made in the calculation of secondary and tertiary meteorological variables from the measured variables [Kahl, 1996; Stohl *et al.*, 1995; Stohl, 1998]. Trajectory models suffer errors from incorrect assumptions made in the calculation of secondary values from the meteorological data, interpolation of data between grid-points, and errors from incorrectly calculated positions of trajectories, and the amplification of the positional errors [Kahl, 1996; McQueen and Draxler, 1994; Stohl, 1998]. It is difficult to assess trajectory error, as it is difficult to compare calculated flow paths with the actual path of modeled flow because it is difficult to measure airflow through the atmosphere. Typical methods of quantification of trajectory error include inter-comparisons of trajectories from different trajectory models and/or meteorological data sets, comparing modeled flow paths using atmospheric tracers such as chemical tracers, smoke or cloud tracts, or balloons, and dynamical tracers, such as potential temperature [Kahl, 1996; McQueen and Draxler, 1994; Stohl *et al.*, 1995; Stohl, 1998].

Calculated trajectory errors range from fairly small values of a few percent of the

travel distance to fairly large values of more than 50% of the total travel distance [Draxler, 1996; McQueen and Draxler, 1994; Stohl, 1998; Vaughan *et al.*, 2002]. Despite the large range of possible trajectory error the commonly accepted error values are around 20% of the total travel distance [Vaughan *et al.*, 2002; Moy *et al.*, 1994; McQueen and Draxler, 1994]. Thus the last calculated point in a trajectory that has traveled 1000 km could be represent the true path of the air or displaced up to 500 km. However, a displacement of 200 km would be the commonly accepted uncertainty. The consequences of trajectory error are that an individual trajectory can only be considered to represent the general motion of the air as opposed to the exact path traveled. There are several ways to minimize the impacts of trajectory error on the conclusion of analyses using trajectories. For shorter study periods, it is common to use an ensemble of trajectories that are initiated at the site of interest and are some initiated simultaneously that are displaced from the site, such as 1° north, south, east, and west of the site [Stohl, 1998; Cape *et al.*, 2000; Stohl *et al.*, 2002a; Lee *et al.*, 1994]. This ensemble of trajectories can show whether the air mass at the site has been transported fairly uniformly (i.e. the trajectories remain close together over their path), or if the air mass has experienced significant shear or mixing (i.e. the trajectories diverge at some point along their path). For a longer study period, it is common to use many years worth of trajectories [Stohl, 1998; Miller, 1981; Brankov *et al.*, 1998; Harris and Kahl, 1994, 1990]. While each individual trajectory may have large uncertainty, a large number of trajectories that show similar features over time can be considered to increase the certainty of the flow pathways. The dominant flow patterns tend to be emerge when large numbers of trajectories as each trajectory portrays parts of these flow patterns. This study makes use of this method to minimize the impact of uncertainty in individual trajectories on flow pathways and source regions for the PICO-NARE site.

2.2 Cluster Analysis

Cluster analysis is a statistical method used to examine data and group it into sets of similar data known as clusters. In general, cluster analysis attempts to minimize the difference between the data within clusters and maximize the difference of data between clusters. It is a useful method for organizing large data sets into smaller, similar groups. These clusters help in examining a data set in two ways. First, the dominant features of the data set can be explored by examining the differences and similarities among the clusters. Second, the individual clusters, which are smaller than the data set, can be examined in detail more easily than the whole data set, due to their reduced size. This ultimately allows the entire data set to be examined more thoroughly by relating each cluster to the data set as a whole. Another advantage of cluster analysis is that it can be automated. The algorithms for performing a cluster analysis can be programmed into a computer, allowing the computer to cluster large data sets.

Cluster analysis was first used with meteorological data by *Kalkstein et al.* [1987] to assess sulfur dioxide concentration measurements. *Moody and Galloway* [1988] were the first to use trajectory coordinates as the clustering variables, in their study of the chemical composition of precipitation [Stohl, 1998]. Since then, the practice of clustering trajectories has become common in the analysis of chemical measurements relating to atmospheric transport ([*Moody and Galloway*, 1988; *Harris and Kahl*, 1990; *Harris et al.*, 1992; *Harris and Kahl*, 1994; *Lee et al.*, 1994; *Moy et al.*, 1994; *Moody et al.*, 1995; *Sirois and Bottenheim*, 1995; *Kahl et al.*, 1997; *Brankov et al.*, 1998; *Cape et al.*, 2000; *Lin et al.*, 2001], etc). Despite its increasingly widespread use, there are no established methods for use when conducting a cluster analysis on atmospheric trajectories, as there is not agreement among the studies listed above in methodology [*Kalkstein et al.*, 1987; *Cape et al.*, 2000].

There are four primary features of cluster analysis that vary among previous

researchers: the algorithm used for clustering, the measure used to compare objects in the data set, the method used determine the number of clusters in the data set, and the units used to express the data. Texts on cluster analysis provide numerous clustering algorithms, many of which also contain several methods for comparing items in the data sets, criteria for choosing the number of clusters, and measures of units for representing the data. These four differences are discussed in the following sections. While clustering may have many applications, the discussion in the following sections will focus on clustering trajectories as it was applied in this study.

2.2.1 Clustering algorithms

There are two primary classes of clustering algorithms used in previous studies: hierarchical and non-hierarchical. There are two classes of hierarchical methods: agglomerative and divisive. In an agglomerative method, all the objects initially belong to separate clusters; that is there are as many clusters as there are objects. In each step of an agglomerative method, the two clusters that are most similar are joined. The clusters from each step are used in the following step for the joining of two more clusters. A divisive method is the opposite of an agglomerative in that all the objects begin in one cluster. At each step, one of the clusters from the previous step is split into a two new clusters until each object belongs to an individual cluster [*Kaufman and Rousseeuw*, 1990].

In contrast, non-hierarchical methods choose a specified number of clusters and then loops through all the objects in the data set. During the loop through the objects, each object is assigned to a cluster depending on one of many measures of similarity or dissimilarity that may be chosen for the procedure. The loop through the objects is repeated (for a loop over loops) until certain criteria are met [*Kaufman and Rousseeuw*, 1990; *Everitt et al.*, 2001].

The most prominent difference between the two methods is that with hierarchical

methods, once a the objects in a cluster has been split or joined with objects in another cluster, those objects are separated (divisive) or grouped (agglomerative) throughout the remainder of the procedure. On the other hand, non-hierarchical methods allow objects to change cluster membership during each loop over the data set, so that an assignment to a cluster is not permanent throughout the procedure.

The second most prominent difference is in the selection of the number of clusters. The selection of the number of clusters is the last step in a hierarchical method, as a set of cluster solutions is created during the progression from many clusters to one cluster (or vice-versa for divisive methods). On the other hand, non-hierarchical methods require a number of clusters be chosen at the onset of the procedure. While the methods used to deduce either the correct number of clusters to retain from a hierarchical cluster solution or the number of clusters to use for a non-hierarchical method are often similar, the fact that the choice for the number of clusters to use for a solution comes as the last step in a hierarchical method and as the first step in a non-hierarchical method is important. This subject will be discussed further in section 2.2.2.

There are two hierarchical agglomerative methods that have been used in previous studies that clustered trajectories: average linkage and Ward’s method. Average linkage selects two clusters to join based on a Euclidean measure of distance between the two. At each step, it calculates the average Euclidean distance between all possible trajectory pairs between each pair of clusters and joins the pair of clusters that is found to have the smallest average distance between them [Moy *et al.*, 1994; Everitt *et al.*, 2001]. Ward’s method selects cluster pairs to join based on an error sum of squares criterion. At each step in the process of joining clusters, Ward’s method minimizes the increase in the total within-cluster variance (denoted as V_w), defined as:

$$V_w = \sum_{k=1}^{n_c} \sum_{j=1}^{n_m} \sum_{i=1}^{n_p} (d_{ijk})^2, \quad (2.4)$$

where n_c is the number of clusters, n_m is the number of members of cluster k , n_p is the number of points in trajectory j , and d_{ijk} is the distance from point i on trajectory j to the average of point i for the members of cluster k [Kaufman and Rousseeuw, 1990; Everitt et al., 2001]. Ward's method has been used to cluster atmospheric trajectories by Moody and Galloway [1988], Kahl et al. [1997], Cape et al. [2000], and Lin et al. [2001]. Two comparisons of average linkage, Ward's method, and a third hierarchical method, known as the centroid method, have been made [Kalkstein et al., 1987; Moy et al., 1994]. Both concluded that the average linkage method was the superior hierarchical method as it tended to more appropriately cluster the generated data used in the studies.

The most popular non-hierarchical method is a procedure known as k-means. K-means is a four-step iterative algorithm developed by Hartigan [1975] that uses a specified number of clusters, usually termed k-clusters, to partition the data by comparing each object to the mean of the members of each of the k-clusters. The k-means algorithm has been employed by many groups to cluster trajectories [Harris and Kahl, 1990; Harris et al., 1992; Lee et al., 1994; Harris and Kahl, 1994; Sirois and Bottenheim, 1995; Moody et al., 1995; Brankov et al., 1998]. It has also been selected for use in this analysis, primarily due to its convenient availability. No studies comparing the use of hierarchical and non-hierarchical methods for clustering trajectories were found. Given the successful use of both methods for studies of atmospheric flow, there is no clear reason presented in these studies to choose one over the other. However, it was felt that the ability of trajectories to change clusters in an iterative procedure such as a non-hierarchical method was the most appropriate.

2.2.1.1 K-means

The k-means algorithm has four basic steps, each of which can be completed in a variety of ways. Three variations of the k-means algorithm were used in this study. These varied slightly from each other in one or more of the steps of the algorithm, as

will be described below. The three variations will be identified as sequential city-block k-means (SC), classical city-block k-means (CC) and, classical Euclidean k-means (CE).

Before a discussion of the three variations can begin, a few concepts that are common among the three must be discussed. The most important concept in the k-means algorithm is the cluster mean, referred to as the cluster center. A cluster center is the arithmetic mean or average of all the members of a cluster. It is defined as:

$$\bar{X}_k = \frac{1}{n_m} \sum_{j=1}^{n_m} X_{j,k}, \quad (2.5)$$

where \bar{X}_k is the cluster center of cluster k , n_m is the number of members in cluster k , and $X_{j,k}$ is trajectory j in cluster k . For this study, \bar{X}_k and $X_{j,k}$ each consist of 241 points, each containing a coordinate of latitude, longitude, and possibly altitude.

A second key concept in the k-means algorithm is the clustering criterion. The clustering criterion is a measure of similarity between two trajectories. Specifically, in the k-means procedures used here, the criterion is a measure of the distance between a trajectory and a cluster center. The method of calculation of the clustering criterion varies among the three k-means procedures and will be defined in the following sections during the discussion of each procedure. The next section will describe the steps of each method and how each is similar or different from the others.

2.2.1.2 Steps in the three k-means methods

Each procedure has four basic steps, which vary among the three procedures.

Step one: Find an initial partitioning of the trajectories. The first step in any k-means algorithm requires the user to define an initial partition [Hartigan, 1975; Jain and Dubes, 1988; Kaufman and Rousseeuw, 1990; Everitt et al., 2001].

While each of the three procedures contains this first step, it can be accomplished in a number of ways. Initial partitions could be created by randomly selecting trajectories for partitions, performing a initial analysis and placing trajectories in partitions manually, or using the results of another clustering procedure.

Step two: calculate the change in clustering criterion that result from changes in membership and reassign trajectories. The second step in a k-means algorithm is to compare each trajectory with the cluster centers of each cluster using the clustering criterion specified for each variation of the algorithm [*Jain and Dubes*, 1988; *Kaufman and Rousseeuw*, 1990]. The SC and CC methods use the same criterion and the CE method uses a different criterion. SC and CC use a criterion known as the city-block distance. It is calculated according to:

$$cd_{ijk} = |x_{ij} - \bar{x}_{ik}| + |y_{ij} - \bar{y}_{ik}| + |z_{ij} - \bar{z}_{ik}|, \quad (2.6)$$

where cd_{ijk} is defined as the city-block distance between point i on trajectory j and point i on cluster center k , $x_{i,j}$ is the distance east/west (or longitude) from the origin to point i on trajectory j , $\bar{x}_{i,k}$ is the corresponding distance to point i on cluster center k , y_{ij} and \bar{y}_{ik} are similarly defined as distances north/south (or latitude), and z_{ij} and \bar{z}_{ik} are similarly defined as the vertical distance (or elevation) [*Kaufman and Rousseeuw*, 1990; *Everitt et al.*, 2001].

The CE method uses Euclidean distance from each point along the trajectory to the corresponding point on the cluster centers as its clustering criterion. The Euclidean distance is given by:

$$Ed_{ijk} = \sqrt{(x_{ij} - \bar{x}_{ik})^2 + (y_{ij} - \bar{y}_{ik})^2 + (z_{ij} - \bar{z}_{ik})^2}, \quad (2.7)$$

where Ed_{ijk} is the Euclidean distance between point i on trajectory j and point i on cluster center k . Using the appropriate measure, city-block or Euclidean depending on the method being used, the distance from a trajectory to each cluster center

is calculated for every point along the trajectory and summed. The result for a single trajectory is k-sums of distances, one for each cluster center. The trajectory is assigned to the cluster with the smallest total distance from its cluster center.

Step three: recalculate the cluster centers. The third step is to recalculate the cluster centers. In the third step the CC and CE methods are the same, while SC has a different third step. The third step for the CC and CE method occurs after all the trajectories have been examined and assigned [Jain and Dubes, 1988; Kaufman and Rousseeuw, 1990; Everitt et al., 2001]. The cluster centers are then calculated using equation 2.5. In contrast, the SC method updates the cluster centers after each trajectory is examined. The cluster center of the cluster to which the trajectory is assigned is updated using:

$$\overline{X}_k = \overline{X}_k + a(X_j - \overline{X}_k), \quad (2.8)$$

where \overline{X}_k is the cluster center of cluster k , X_j is trajectory j , and a is a value between 0.5 and 0.1. This is more easily understood if a single coordinate is considered. The general case would then be:

$$\overline{x}_{ik} = \overline{x}_{i,k} + a * (cd_{ijk})_x. \quad (2.9)$$

This states that $\overline{x}_{i,k}$, which is the x -coordinate of cluster k at point i , is adjusted by a fraction a of the city-block distance $(cd_{ijk})_x$ of the x -coordinate distance between the value of the cluster center of cluster k and trajectory j at point i . The value of a begins at 0.5 and decreases linearly to 0.1 during each pass over all the clusters. The value of a was predetermined by Interactive Data Language (IDL), the software package from which SC originated.

Step four: repeat. Step four is to return to steps two and three. Due to the differences in step three, SC is again different from CE and CC. SC repeats steps

two and three for a specified number of times. There are two reasons for this. The primary one is that the procedure, which was provided by IDL, required that a number of iterations be input into the program. The second is that a predetermined number of iterations must be used for a in order to linearly decrease from 0.5–0.1 in equations 2.8 and 2.9. In the CC and CE methods step four can be completed in two ways. Steps two and three can be repeated a specified number of times, as in SC, or can be repeated until no trajectory (or a specified number of trajectories) changes its cluster assignment. The point at which no trajectories change cluster assignment is referred to as convergence [Hartigan, 1975; Jain and Dubes, 1988; Kaufman and Rousseeuw, 1990; Everitt *et al.*, 2001].

One feature of these procedures that may not be clear from this discussion is the effect of incomplete trajectories on a clustering procedure. It is important for the criteria used that each trajectory be complete. If a trajectory were to terminate early, it would have fewer endpoints to use to find the total distance between pairs of points. This would result in different weighting of trajectories of different lengths. There are two possible ways to resolve this difficulty. The first would be to use an average distance, as opposed to a total, in the distance criterion used. The second would be to remove all trajectories that terminated early from the cluster analysis. The second approach was taken by all the authors cited above and will also be applied here. Table 2.2 summarizes information about trajectories removed from each month’s original set.

2.2.1.3 Selection of k-means sequence

To summarize, the SC method measures the distances from a trajectory to each cluster center using the city-block distance. It assigns the trajectory to a cluster and then updates the cluster center determined to be closest to the trajectory in question by adding a fraction of the measured distance to the cluster center. This causes the cluster centers to act like a moving average of its members. It also means

Table 2.2 Summary of the number of trajectories for each month and the number of incomplete trajectories removed for clustering.

Month	Total trajectories	Trajectories removed	Percent total remaining	Remaining trajectories
January	4,932	142	97.12	4,790
February	4,416	180	95.92	4,236
March	4,937	189	96.17	4,748
April	4,800	61	98.73	4,739
May	4,960	39	99.21	4,921
June	4,800	36	99.25	4,764
July	4,960	17	99.66	4,943
August	4,959	26	99.48	4,933
September	4,799	65	98.65	4,734
October	4,960	92	98.15	4,868
November	4,794	81	98.31	4,713
December	4,954	131	97.36	4,823
Total	58,271	1059	98.18	57,212

that the set of cluster centers used for comparison change from one trajectory to the next. The CC method also uses the city block distance for the measure between a trajectory and a cluster center, but the cluster centers are not changed until all the trajectories have been examined and their membership determined. This has an advantage over the SC method in that during each loop over the trajectories, the same cluster centers are used for comparison for all the trajectories. The CE method uses the Euclidean distance for the measure between a trajectory and each cluster center and the cluster centers are not changed until all the trajectories have been examined and their membership determined. This has an advantage over the CC method in that the Euclidean distance more closely measures the distance between two points than the city-block distance.

These advantages indicate that simply using CE would be the best option. However, CE has a disadvantage in that it takes much more computational time than the other two. Similarly, CC takes much more computational than does SC. Ultimately, the most accurate result is the most important, so CE was used to prepare the final clustering solution. However, it was found that SC and CC could be used as a first

and second step, respectively, for producing initial partitions for use with CE. The resulting procedure for computing clusters begins with the SC method, which creates random initial partitions. The SC method was run for 25 iterations. The clusters from the SC method are used as the initial partitions for the CC method. The CC method was run to convergence. The clusters from the CC method were used as initial partitions for the CE method. The CE method was run until convergence and the resulting clusters are the final clustering solution. This was found to generally be the fastest way to produce the most accurate cluster solution.

It should be noted that different seeds can produce different final solutions. The IDL procedure SC was initiated with the same set of pseudo-random numbers as seeds, so rerunning a clustering procedure on the same data set, with the same parameters (e.g., number of iterations for SC), produced the same result. However, *Hartigan* [1975] noted that k-means may not always produce the same solution with the same data set. Additionally, *Kalkstein et al.* [1987] noted that meteorological data is generally not strongly nucleated, indicating that non-hierarchical procedures might produce slightly varying results. In this study, although somewhat varying results were obtained with different seeds, the differences between the clusters in separate runs were small. An example of two different cluster solutions can be seen in Figure 2.1 and Figure 2.2.

2.2.2 Number of clusters

Selecting the number of clusters used to represent a data set is an important issue. If too few clusters are used, information can be lost. If too many clusters are used, “false” clusters that only contain a small portion of the data may appear, or the resulting set of clusters might be too complex to easily interpret, eliminating the usefulness of clustering [*Kaufman and Rousseeuw*, 1990]. The selection of the number of clusters is an issue in any clustering method. As mentioned previously, for

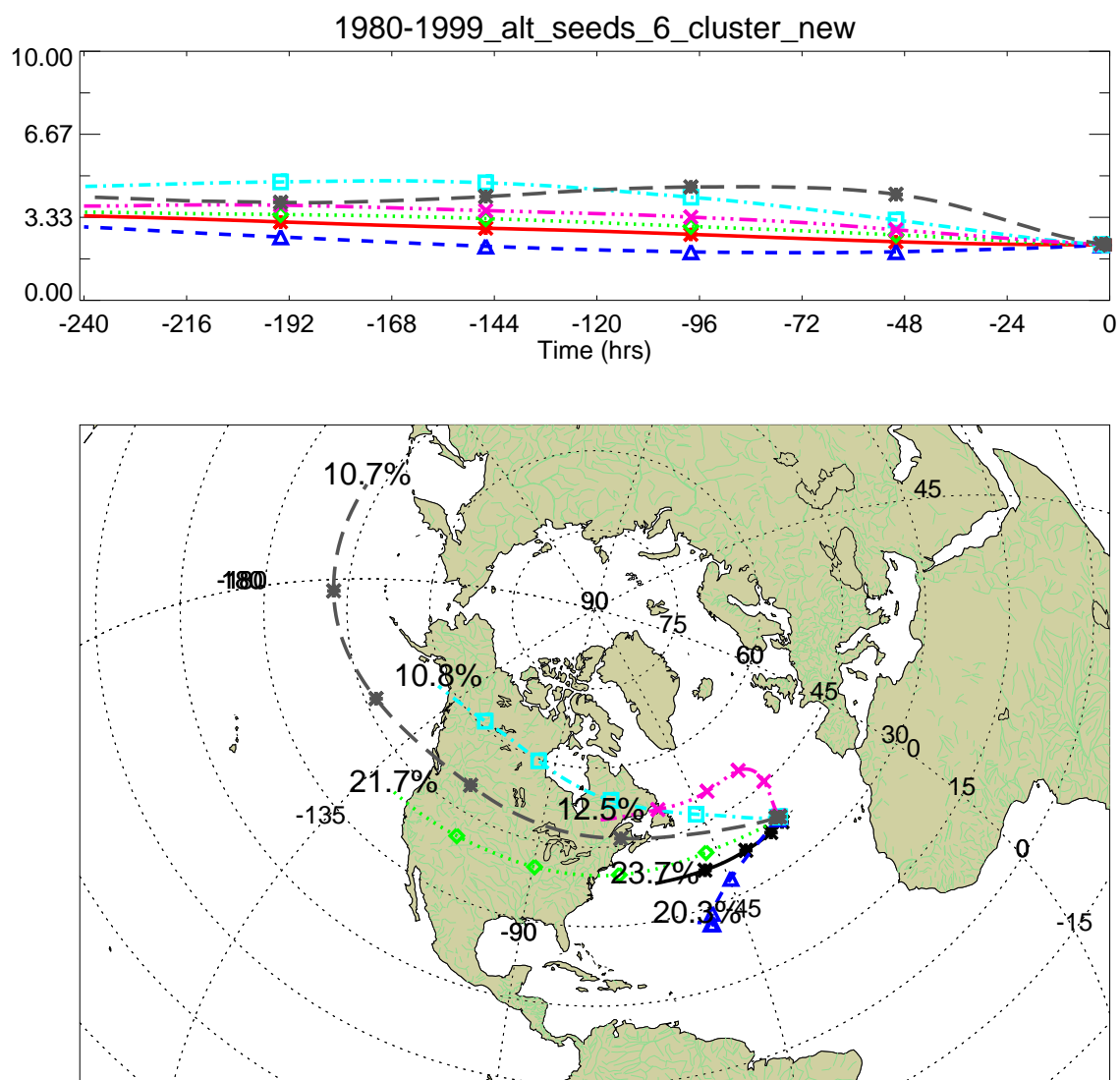


Figure 2.1 The resulting cluster centers from CE using the seeds from the SC method.

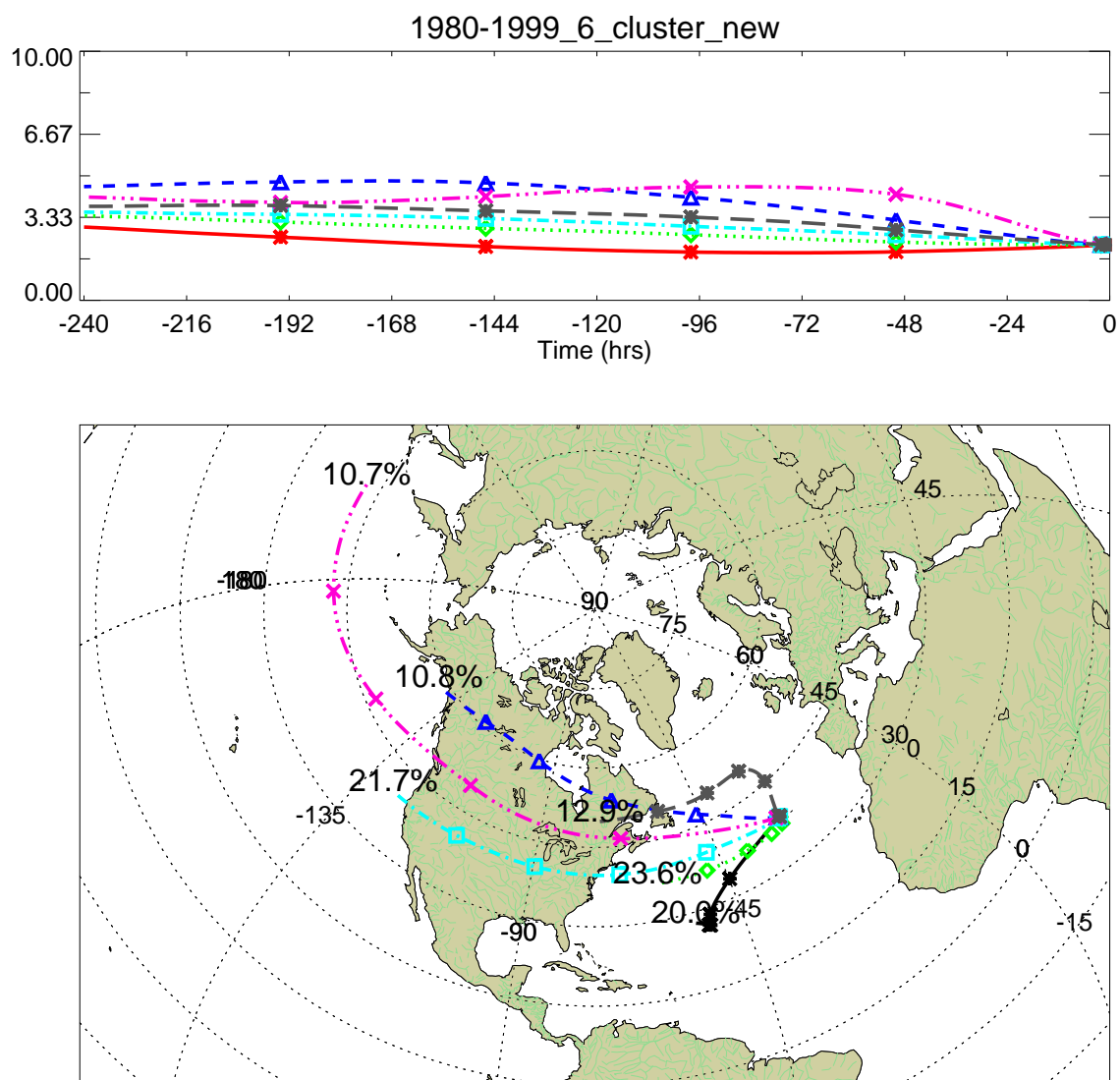


Figure 2.2 The resulting cluster centers from CE using the seeds from the CC method.

hierarchical methods, the selection of the number of clusters is the last step in a clustering procedure. For example, hierarchical, agglomerative methods begin with all objects in the data set in individual clusters, and at each step, clusters are joined until all the objects are in a single cluster. A measure of error is calculated at each step, and after the procedure is completed, the measure of error is inspected for certain characteristics (usually abrupt jumps or dips in the measure of error calculated) to determine the appropriate number of clusters for the data set. Non-hierarchical methods require that the number of clusters be determined before the procedure can be employed, making the determination of the number of clusters a primary decision early in the stages of performing a cluster analysis. Regardless of when the selection of the number of clusters is made, it is an important decision in either method.

2.2.2.1 Factors in the choice of the number of clusters

There are many methods for choosing the number of clusters, but most of those employed by other researchers on trajectory data involve plotting one of several measures of error against the number of clusters. In general, the correct number of clusters tends to coincide with a change in the trend of the values of these measures of error. This is usually accomplished by examining plots of these measures versus the number of clusters. *Moody and Galloway* [1988] used Ward's method to cluster the data and then prepared graphs of the within-cluster variance versus the number of clusters. (The within-cluster variance is as defined in equation 2.4 on page 22 using the Euclidean distance between points as defined in equation 2.7 on page 25.) They chose seven clusters based on “dramatic” changes in the within-cluster variance (100%) or more. Other studies have also used Ward's method to cluster the data, then prepared graphs of the total within-cluster variance [*Harris and Kahl*, 1990; *Harris et al.*, 1992; *Harris and Kahl*, 1994; *Moody et al.*, 1995]. *Harris and Kahl* [1990], *Harris et al.* [1992], and *Harris and Kahl* [1994] chose six and *Moody et al.* [1995] chose seven clusters based on large changes in the slope of the within-cluster variance. They then

used the number of clusters obtained to perform a k-means clustering on the data. *Hamlin* [1995] used a k-means algorithm to cluster trajectories into a range of clusters from two to fifteen and chose six clusters, based on changes in the value of R^2 . The R^2 statistic has been described as the fraction of the ratio of the between-cluster variance to the total-cluster variance that is explained by the current number of clusters [*Moy et al.*, 1994; *Hamlin*, 1995]. It can be calculated using the ratio of the between-cluster variance (V_b) to the total-cluster variance (V_t):

$$R^2 = (V_b/V_t). \quad (2.10)$$

The total-cluster variance is similar to the within-cluster variance, but as opposed to summing the total distance from each trajectory to its cluster center, the total distance from each trajectory to the average of all the trajectories is summed:

$$V_t = \sum_{j=1}^{n_t} \sum_{i=1}^{n_p} (d_{ij})^2, \quad (2.11)$$

where n_t is the number of trajectories in the data set, n_p is the number of points on the trajectories and d_{ij} is the distance from point i on trajectory j to the average of all the points i in the data set. The between-cluster variance (V_b) is the sum of the distances from each cluster center to the average of all the trajectories, given by:

$$V_b = \sum_{k=1}^{n_c} \sum_{i=1}^{n_p} (d_{ik})^2, \quad (2.12)$$

where i, k, n_c , and n_p are the same as in equation 2.4, and d_{ik} is the distance from point i on cluster center k to point i on the average of all trajectories. The values of R^2 can range from 0.0 to 1.0. If all the trajectories are in one cluster, then V_b equals 0.0 and R^2 equals 0.0. Alternately, if all the trajectories are in individual clusters, V_b would equal V_t and R^2 would equal 1.0.

Moy et al. [1994] and *Cape et al.* [2000] used the hierarchical method average linkage to cluster the data, and considered changes in R^2 in determining the correct

number of clusters. *Dorling and Davies* [1995] and *Brankov et al.* [1998] also prepared cluster solutions for a range of clusters but used the percent change in the within-cluster variance as a function of the number of clusters as the indicator of the proper number of clusters to use. *Dorling and Davies* [1995] performed an analysis on a number of sites and found the numbers of clusters ranged from six to ten. *Brankov et al.* [1998] only studied one site and used eight clusters. *Kahl et al.* [1997] also used Ward's method and used within-cluster variance versus the number of clusters and chose seven clusters based on both a change in the within-cluster variance and a visual inspection of aggregate trajectory plots, or membership plots, of each cluster. *Lin et al.* [2000] used Ward's method and based the choice of eight clusters on the expected number of clusters indicated by other studies and a visual inspection of cluster membership plots without examining any graphs of variance. *Lee et al.* [1994] used a k-means algorithm to cluster trajectories and chose to retain the five clusters that each contained at least 10% of the data. *Sirois and Bottenheim* [1995] chose seven clusters using an unspecified method to cluster trajectories with k-means.

Since no common method was used in previous studies several of the methods were considered in this study. The k-means algorithm used here requires that the number of clusters be input at the initial step of the algorithm, so a number of alternative cluster solutions were prepared, for numbers of clusters ranging from two to twenty-five. Initially, graphs of several statistics were examined. The within-cluster variance (equation 2.4) was plotted versus the number of clusters. The between-cluster variance (equation 2.12) was also plotted against the number of clusters. Finally, the total-cluster variance (equation 2.11) was used with the between-cluster variance to calculate the R^2 statistic, which was also plotted against the number of clusters. The Euclidean distance (equation 2.7) was used as the distance between points in each of the calculations as it was the distance used for the final cluster solution.

To give greater insight into the question of the number of clusters to be used, the percent change in each of the statistics (V_w , V_b and R^2) was also plotted against the

number of clusters. The percent change was calculated in descending order, from the greatest number of clusters to the smallest number of clusters:

$$\% \Delta MV_n = \frac{MV_n - MV_{n-1}}{MV_n}, \quad (2.13)$$

where $\% \Delta MV_n$ is the percent change of the measure of variance at number of clusters n and MV_n is the value for the measure of variance at n number of clusters. Plotting the percent change helps display changes in the trend of the change in the particular measure of variation.

2.2.2.2 Selection of the number of clusters

The general use of variance plots is to identify features that would indicate the correct number of clusters to be chosen for the data set. These features are usually abrupt changes in slope, but they could also include either spikes or dips, or simply passing a threshold value. *Brankov et al.* [1998] suggested a threshold value of 5% for their graphs of percent change of the within-cluster variance. An inspection of the plots for R^2 (Figure 2.3 a), the between-cluster variance (Figure 2.4 a), and the within-cluster variance (Figure 2.5 a) do not reveal any outstanding features; the changes in variance are actually quite smooth. One feature displayed is the similarity between the plot of R^2 and the between-cluster variance. This is not surprising as R^2 is the between-cluster variance divided by the total variance, which is constant. This means the two plots should have the same shape, but different magnitudes. As a result, it is not necessary to examine both the within-cluster variance and R^2 . Using the within-cluster variance would suffice when working with one data set. However, the use of R^2 might be useful for the comparison between multiple data sets, as its values must be between 0 and 1.

Although the plots of the within-cluster and between-cluster variance do not show any distinct features that indicate the appropriate number of clusters to use, they

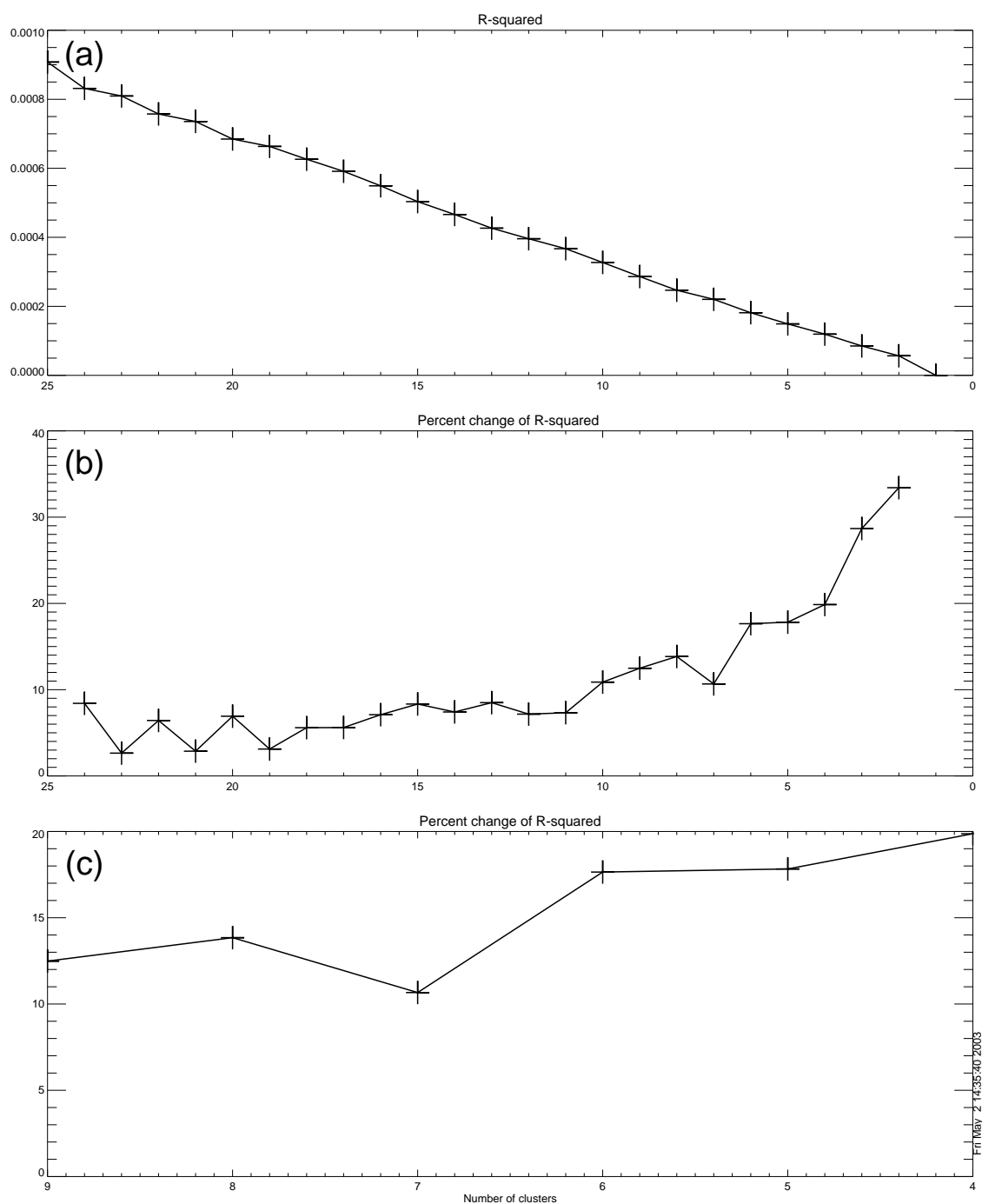


Figure 2.3 Plots of R^2 as a function of the number of clusters. The top plot shows R^2 . The middle plot shows the percent change in R^2 . The bottom plot shows a more detailed view of the middle plot.

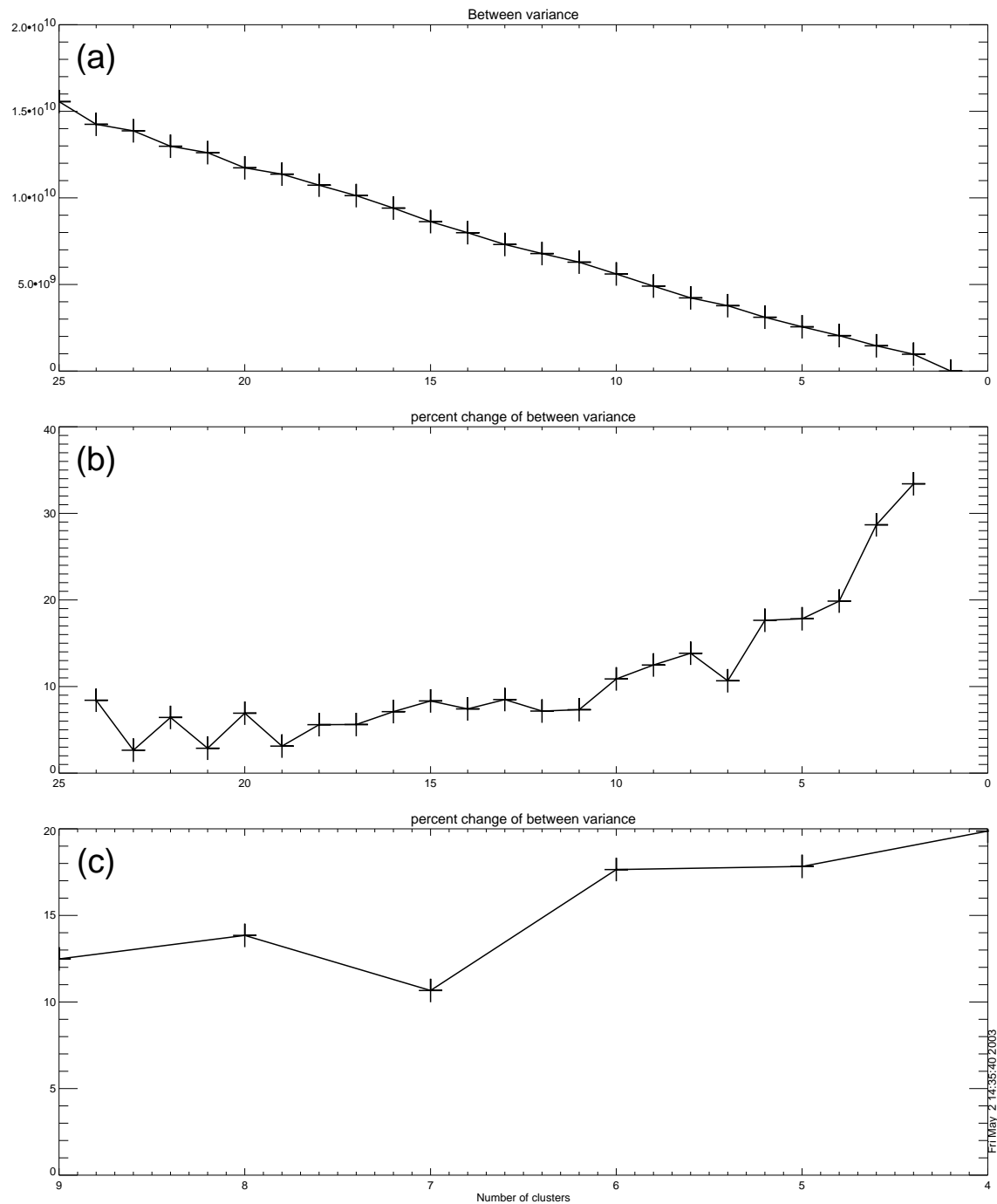


Figure 2.4 Plots of variance between clusters as a function of the number of clusters. The top plot shows the change between clusters. The middle plot shows the percent change in the between variance versus the number of clusters. The bottom plot shows a more detailed view of the middle plot.

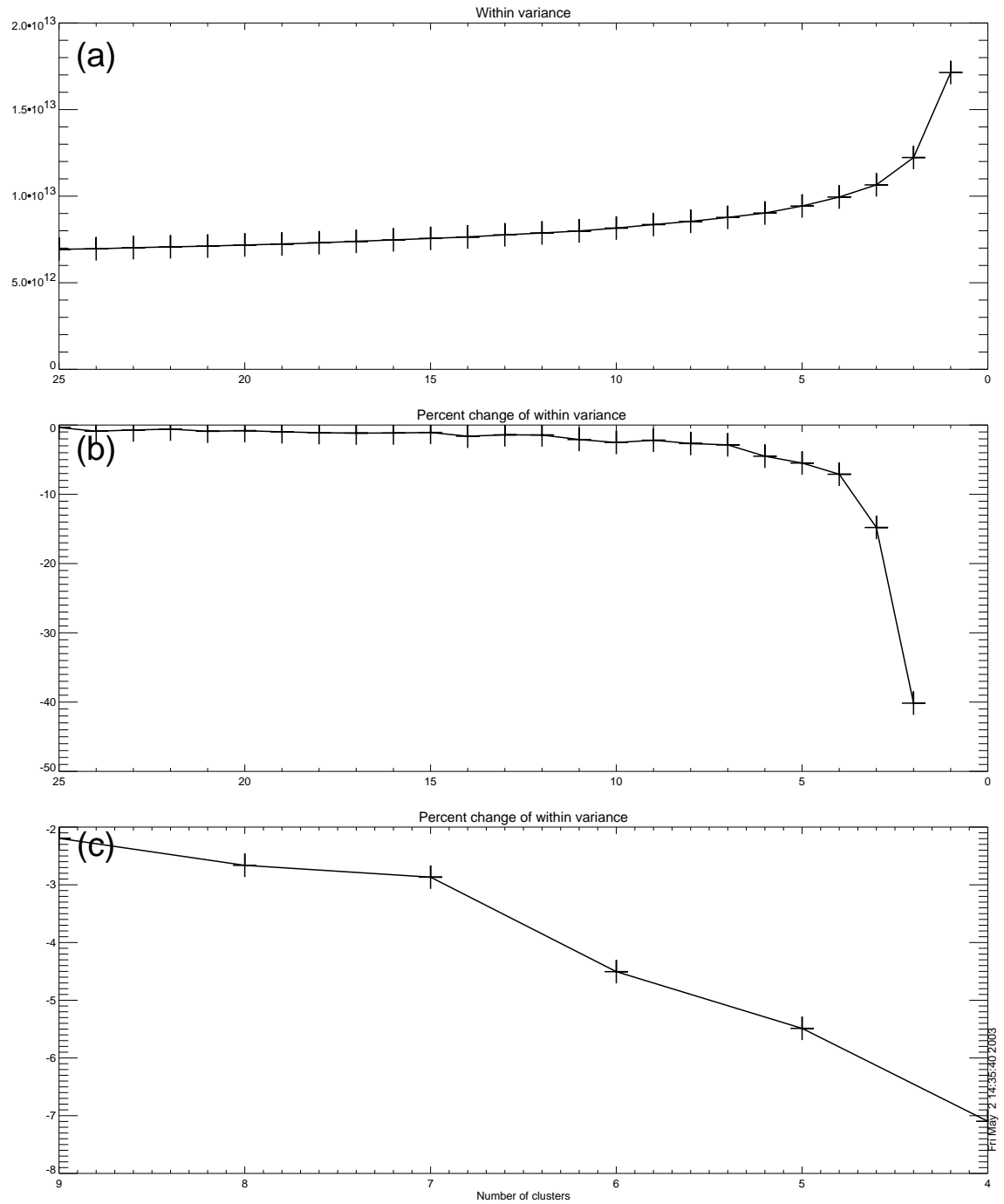


Figure 2.5 Plots of variance within clusters as a function of the number of clusters. The top plot shows the change of the within-clusters variance. The middle plot shows the percent change in the within-cluster variance versus the number of clusters. The bottom plot shows a more detailed view of the middle plot.

provide information about how strongly nucleated the data set is. The between-cluster variance shows a relatively slow and steady decrease as the number of clusters is decreased. This indicates that no clusters with drastically different cluster centers are joined during each step. The within-cluster variance shows a similar effect, but has a few more features than the between-cluster variance. In the lower portion of the graph, of approximately 8 clusters to 25 clusters, there is only a slight increase of the within-cluster variance as the number of clusters is decreased. However, after approximately eight clusters, a more distinct increase can be seen with the removal of each cluster, increasing dramatically with the removal of the last few clusters. While this does not lead to a clear choice of a particular number of clusters, it does give a range to use. Clearly 3 clusters is too few, as there is a notable increase of the within-cluster variance with the reduction from 4 to 3 clusters and 3 to 2 clusters. This means that during these steps, clusters are joined that have notably different members. Nine or more clusters is probably too many, since the increase of the within-cluster variance between each of these steps is small. The small increase indicates that the clusters joined in these steps are not notably different.

The plots of percent change of each of the variance terms can be used to provide further insight in the trends of variance and the resulting number of clusters to use. The plot of the percent change of the within-cluster variance (Figure 2.5b) mirrors the within-cluster variance fairly closely. The magnitude of the percent change on the left side of the within-cluster variance is quite small (less than -2.0%) and relatively constant. On the right side, the magnitude of the percent change can be seen to increase in a similar manner to the within-cluster variance. To explore this figure further, a zoomed view from 4 to 9 clusters is shown as well (Figure 2.5c). The largest change shown in this range is between 7 (the 7 to 6 clusters transition) and 6 (the 6 to 5 clusters transition), from -2.7% to -4.5% . The decrease from 5 (the 5 to 4 clusters transition) to 4 (the 4 to 3 clusters transition) clusters also shows a significant increase as well, from -4.5% to -5.4% and also crosses the threshold value

of 5% proposed by *Brankov et al.* [1998]. Either of these features could indicate 4 or 6 clusters as the correct number of clusters to use.

The plot of the percent change of the between variance also provides additional insight. The left side (25-10 clusters) again shows a fairly small decrease. The decrease is relatively small, on the order of 10% versus a total change of 35% from the most to fewest clusters. The magnitude of the percent change begins to gradually increase at around 11 clusters. A feature that stands out in this figure is the dip that occurs at 7 (the 7 to 6 clusters transition). This feature indicates the clusters joined in the transition from 7 to 6 clusters were more similar than those joined in the previous step (8 to 7) and the next step (6 to 5). This indicates that the use of seven clusters is not significantly better than the use of six, which would indicate the use of 6 clusters. The values of the percent change from 6 down to 4 are not very different, while the change in the last few steps becomes larger. The larger changes in the last few steps are partly due to the smaller values of the between-cluster variance and partly due to the joining of dissimilar clusters.

Clearly, the choice of the number of clusters is somewhat subjective. However, these analyses provide significant guidance for the choice, as the use of the plots of variance have helped narrow the choice of clusters to two options, 4 or 6 clusters. After the examination of the plots of variance, cluster plots were examined to make the final choice of the number of clusters. Plots of the cluster centers for numbers of clusters ranging from 4 to 7 are shown in Figures 2.6–2.9. The plots for 4 through 6 clusters were examined as well as the plot for 7 clusters, which served as an additional reference. The plots of 4 (Figure 2.6), 5 (Figure 2.7) and 6 (Figure 2.8) clusters all have distinct cluster centers, with the distinction being greater in the 4-cluster plot than in the 5-cluster plot, and in the 5-cluster plot than the 6-cluster plot. Since the 6-cluster plot has distinct clusters, the use of 6 rather than 5 or 4 clusters conveys more information, as it contains additional clusters that portray a unique flow path. Additionally, six clusters was the the only number indicated by both the within-cluster

and between-cluster variance. For these two reasons, six clusters were retained for use in this analysis.

2.2.3 Units

The selection of units can have a profound impact on the results of a cluster analysis. There are a number of ways to represent the data. The default is coordinates of latitude, longitude, and altitude, but these can be transformed to other sets of coordinates. Possible transformations include normalizing each coordinate by changing each to the deviation from the average and divided by the standard deviation, converting each coordinate into a distance from the site (*e.g.*, changing latitude into km north or south of the site), and scaling the coordinates by dividing by potential maximum/minimum values to force coordinates into smaller ranges. The choice of methods to represent the data can have a profound impact on the results of a cluster analysis.

2.2.3.1 Available units

When *Moody and Galloway* [1988] introduced the use of cluster analysis on trajectory data sets, they used trajectory endpoints of latitude and longitude as their units. Many other studies have also followed this precedent, using longitude and latitude as the input variables to the cluster analysis [*Harris and Kahl*, 1990; *Harris et al.*, 1992; *Harris and Kahl*, 1994; *Moody et al.*, 1995; *Kahl et al.*, 1997; *Dorling and Davies*, 1995; *Brankov et al.*, 1998; *Sirois and Bottenheim*, 1995]. *Lee et al.* [1994] also used horizontal and vertical coordinates, but standardized their data to reduce the differences in variance among latitude, longitude and pressure level. This was achieved by converting each datum into a normalized deviation from that variable's average value. They then linearly weighted the trajectories to produce variance 20 times larger at the starting points (the points farthest back in time) than at the end points (the points

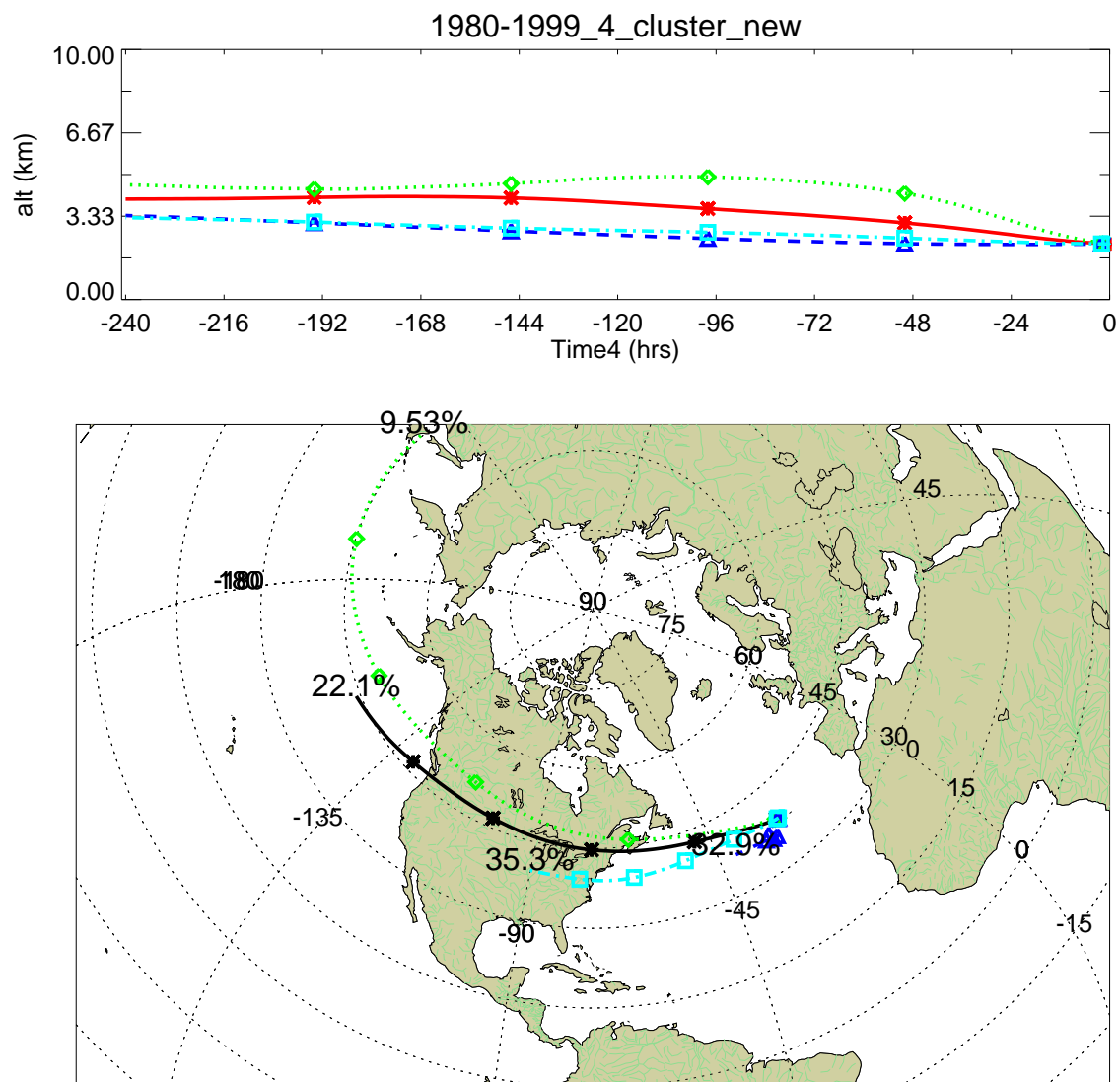


Figure 2.6 Plot of the cluster centers for 4 clusters.

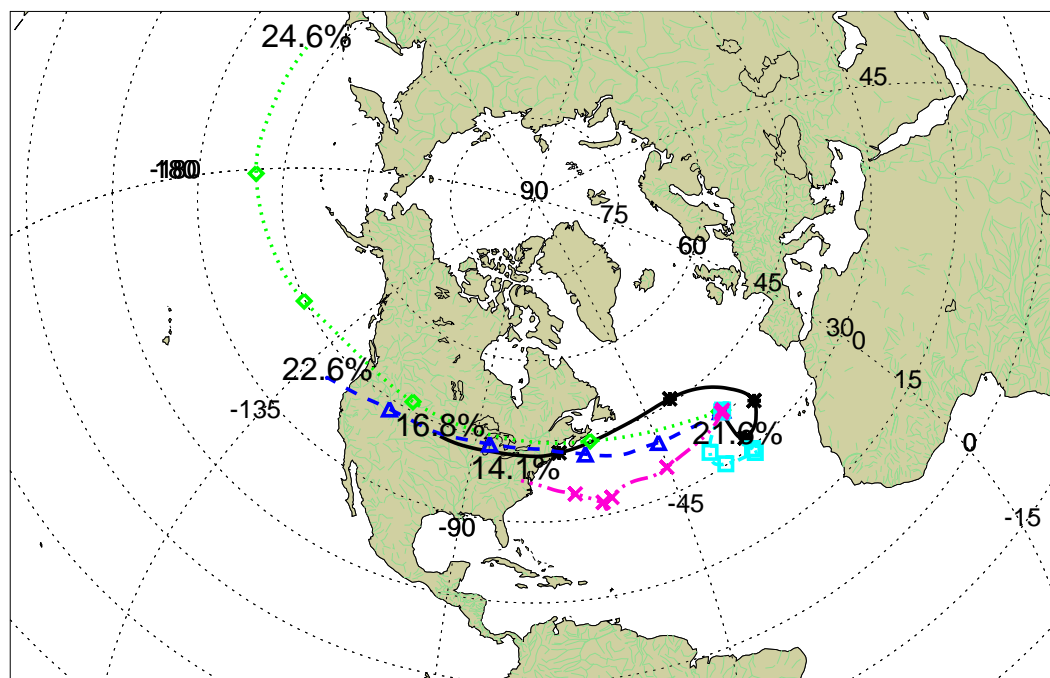
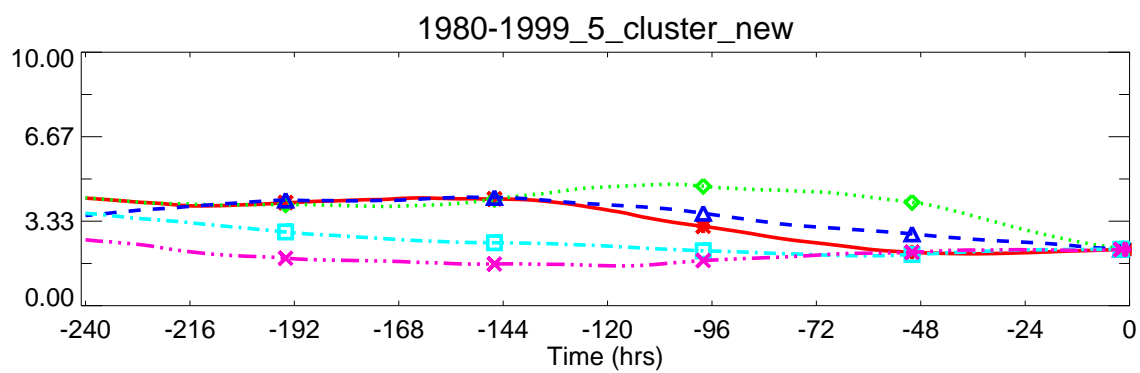


Figure 2.7 Plot of the cluster centers for 5 clusters.

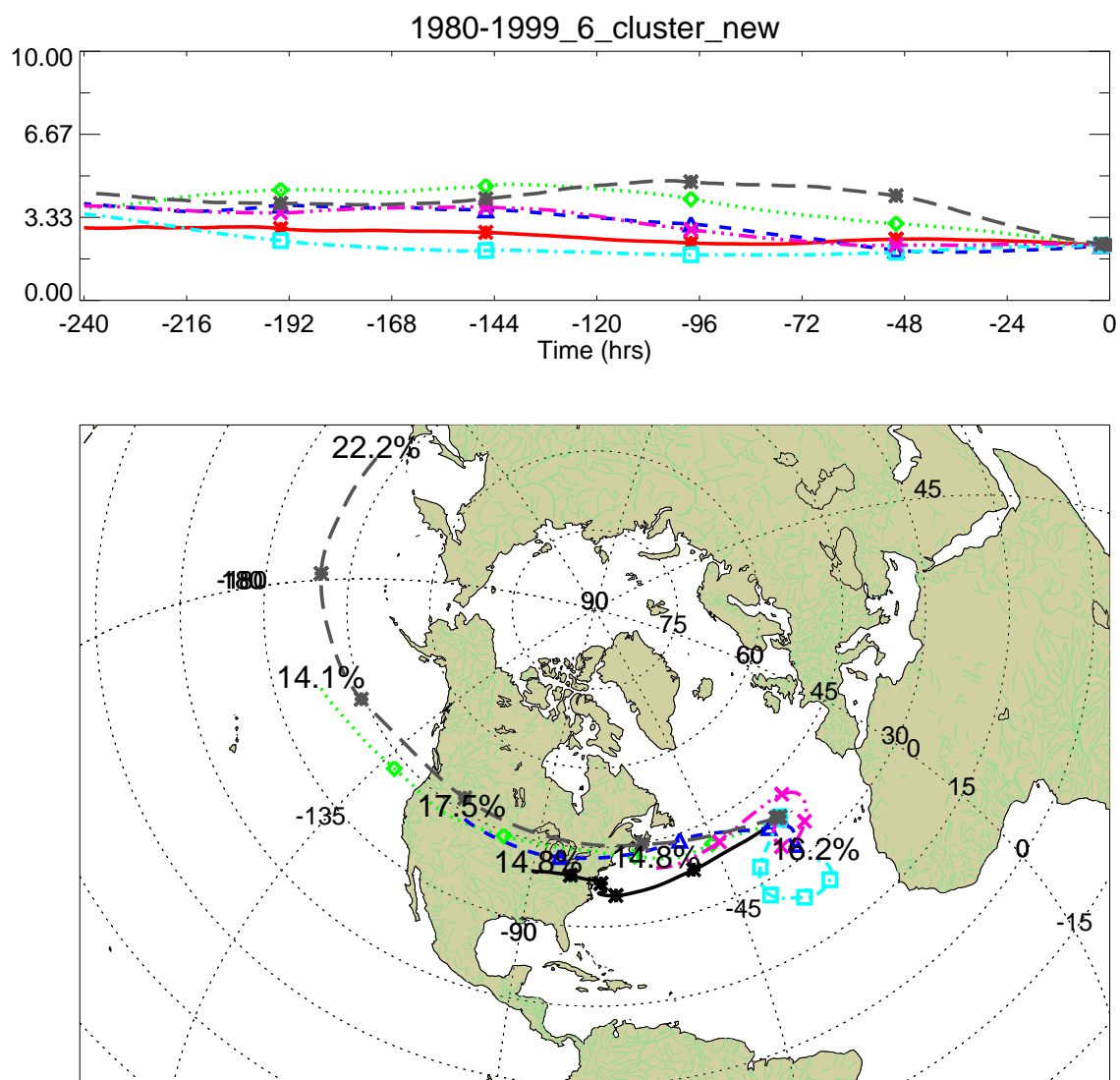


Figure 2.8 Plot of the cluster centers for 6 clusters.

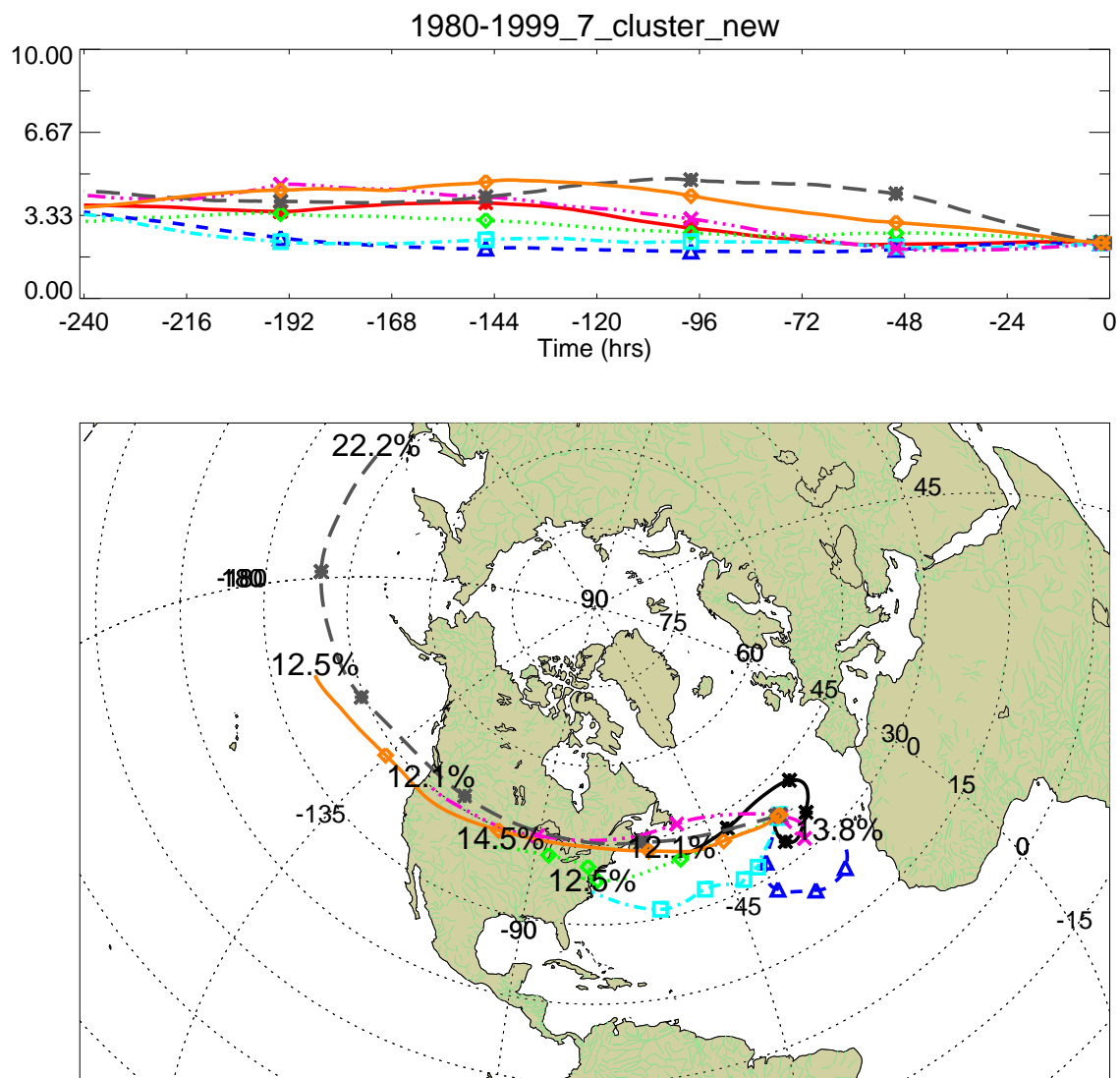


Figure 2.9 Plot of the cluster centers for 7 clusters.

at the arrival site) to emphasize source regions near the site while preserving equal magnitudes of variance for latitudinal, longitudinal and vertical positions. *Cape et al.* [2000] converted coordinates of latitude and longitude to distances north/south and east/west in km and added a vertical pressure coordinate measured in hPa. *Hamlin* [1995] used scaled latitude and longitude coordinates without a vertical coordinate. Each latitude was scaled by converting it to its distance from the equator by the maximum possible distance from the equator. Each longitude was similarly scaled using distances from the prime meridian. This process forced each clustered variable into the range of 0 to 1. Another variation was employed by *Lin et al.* [2000], who used geographic distances, which they defined as the length of the great circle arc formed by two points on the earth, in degrees.

Due to the importance of units, as emphasized in literature on cluster analyses, and the variety of units used in previous studies, several measures of units were explored here. The use of degrees latitude and longitude will serve as a “control” since it was used by many other studies, and will be referred to as DLL in this section. Additionally, latitude and longitude were coupled with altitude in km (referred to as DLLH). Other unit systems include measures of latitude and longitude which were converted to distances in km east/west and north/south from the PICO-NARE site (referred to as KLL), the converted distances along with the vertical coordinate in km (referred to as KLLH), a scaled horizontal and vertical deviation above/below the PICO-NARE site (referred to as SLL and SLLH), and a normalized unit system (referred to as NLL and NLLH). One additional unit system will be commented on, though it was not considered for use in producing a final cluster solution, as it was developed too late to perform the extensive analysis that has been done on the other systems. The additional unit system uses great circle distances and is referred to as GLLH. Each will be described in detail in the following paragraphs, followed by a discussion of the advantages and disadvantages of each.

Latitudinal and longitudinal coordinates (DLL). The DLL system used the trajectory coordinates in latitude and transformed longitude. The longitudinal coordinates were transformed in two ways. First, the arrival site was shifted to the prime meridian by adding 28.4 to all the longitudinal coordinates. Second, the longitudinal coordinates were transformed so that a trajectory traced backward in time would be continuous. When a trajectory changed longitudinal coordinates from -179° to $+179^\circ$, it would be converted to -181° .

Latitude, longitude and altitude coordinates (DLLH). The DLLH system used the same latitudinal and longitudinal coordinates as DLL, but added a third coordinate: the vertical height of each trajectory endpoint measured in km above sea level.

Distance east/west and north/south in km (KLL). The KLL system used coordinates of distance east/west and north/south of the trajectory arrival at the PICO-NARE site expressed in km. Each latitude coordinate was converted to the great circle distance north or south from the arrival point to a point at the same longitude as the PICO-NARE site. To distinguish direction, north was identified as positive and south as negative. Each longitude coordinate was converted to the great circle distance east or west from the arrival point to a point at the same latitude as the PICO-NARE site. To distinguish direction, east was identified as positive and west as negative. The distances east/west were modified to be continuous, similar to the longitudinal coordinates in the DLL method. For a trajectory that originated at $(38.47^\circ, 0^\circ)$ and traveled east to arrive at the PICO-NARE site $(38.47^\circ, -28.4^\circ)$, its converted longitudinal coordinate would be $-20,480$ km. For a trajectory that traveled west to the PICO-NARE site with the same origin would have a longitudinal coordinate of $2,465$ km.

Distance east/west, north/south and above/below in km (KLLH). The KLLH system uses the same horizontal coordinates as KLL, but adds a vertical coordinate. The vertical coordinate was calculated as the difference between the vertical height at any point and the arrival height of 2.223 km above sea level. Heights above the arrival height were positive and heights below the arrival height were negative.

Scaled distances east/west and north/south (SLL). The SLL system uses the distances calculated in KLL and normalizes each by maximum values, as was done by *Hamlin* [1995]. The east/west coordinate values were divided by the circumference of the earth at the destination latitude. The north/south distances were divided by the distance from the destination latitude to the closest pole (the north pole in this case).

Scaled distances east/west, north/south and above/below (SLLH). The SLLH system uses the same horizontal distances as in SLL, but adds a scaled vertical coordinate. The vertical coordinate is the distance above/below the arrival site used in KLLH but divided by the distance from the arrival site to the maximum altitude in the data set. The maximum altitude in the data set is used as the model top, which makes it the maximum distance in the vertical direction that a trajectory can travel.

Normalized latitude and longitude (NLL). The NLL system uses the coordinates of DLL and normalizes them to reduce the effects of outliers and large differences in the variation between latitude and longitude. The method for normalizing the data was taken from *Jain and Dubes* [1988]. Each datum is normalized by:

$$x_{ij} = \frac{x_{ij}^* - m_i}{s_i}, \quad (2.14)$$

where x_{ij} is the normalized value for datum j at point i , x_{ij}^* is the raw value for the datum, m_i is the average value for the trajectories at point i , and s_i is the variance of the data at point i given by:

$$s_i^2 = \frac{1}{n_t} \sum_{j=1}^{n_t} (x_{ij}^* - m_i)^2, \quad (2.15)$$

where n_t is the number of trajectories.

Normalized latitude and longitude and height (NLLH). The NLLH system uses the same horizontal coordinates as NLL but adds a vertical coordinate. The vertical coordinate is the altitude in km, normalized in the same manner as in NLL.

Great circle distance between points (GLLH). The GLLH system combines both the unit system and the measurement criteria used in the k-means methods presented in section 2.2.1. This is done by using the coordinates of latitude and longitude as the input unit system, but the distance calculated between each pair of points begins with calculating the great circle distance between each pair of points. The Euclidean distance between the two points is then calculated using the great circle distance and difference in altitude between the points.

2.2.3.2 Criteria for the selection of units

The examination of the different unit systems had two goals. First, to examine what differences, if any, the unit system had on the results of the clustering process. All the unit systems explored have been used in other studies. However, no comparisons among them were found in literature, indicating that this might be the first examination into this issue. It was possible that each unit system would produce identical clusters, but this was not clearly the case. The second goal was dependent upon the results of the first analysis. If there were differences in cluster solutions among the

unit systems, the second goal was to determine which system was most appropriate for use with the methods developed in this study.

This first goal is fairly straightforward. Plots of the cluster solutions for the nine unit systems are presented in Figure 2.20–Figure 2.23. There are clearly a differences among most of the plots. However, there are a few exceptions. The solutions for the KLL and the KLLH systems appear identical. Similarly, the the solutions for the DLL and the DLLH systems are not notably different. Thus, inclusion of altitude does not impact the results of the cluster analysis using these systems. Because the two sets are similar and because further examination was dependent upon cluster solutions being dissimilar, it was not necessary to use all four of these systems in the examination of the different unit systems. Therefore KLL and DLL were not included in further analysis. Even though the vertical components did not affect the clustering results in this analysis, it is possible that it could affect other clustering solutions within this study. For this reason, the the systems that include the trajectory heights (KLLH and DLLH) were retained over those that do not (KLL and DLL).

The second goal is to determine which unit system is most appropriate with the cluster analysis methods. The selection of the most appropriate unit system is not obvious, as there are advantages and disadvantages to each. The primary goal of using a particular unit system is to accurately portray the distances between trajectories, as the trajectories in a cluster should be physically close to each other. A secondary goal of a unit system is to reduce the effect of outliers on the clustering procedure, as the clustering algorithm (k-means) could be affected by them. It is difficult to weigh how well each unit system accomplishes these goals because there is not a standard by which to compare or method for determining how well a unit system fulfills the goal of using a particular system. Because of the lack of methods for choosing a unit system, the decision must then be made on limited observations and “educated guesses” using what can be deduced from known conditions and procedures.

There are two exercises that can be used to aid the “educated guess” of how well

a unit system fulfills the goals of accurately portraying the closeness of clusters and reducing the impact of clusters. The first is an examination of histograms of each coordinate for each system. This can help one visualize how the conversion from one unit system to another can change differences in magnitude within a particular coordinate and between coordinates as well as change the distribution of the values within a coordinate. The second is an examination of criteria used to measure the distances between trajectories and how it relates to the coordinate system. The criteria is the Euclidean distance between points (equation 2.7). For most unit systems, a quantitative measure can be derived for this examination. However, this examination is limited to a few examples. Therefore this quantitative measure is reduced to a qualitative measure. These two can be used to help examine how well a unit system fulfills its goals, but neither present a broad quantitative measure for comparison among the systems. Instead, they provide some qualitative information about how a unit system fulfills its goals. It is important to emphasize that even the qualitative information provided is limited and the choice of a unit system is somewhat subjective.

2.2.3.3 Final selection of units

With the tools for the analysis of the unit systems established, a discussion of each unit system with respect to these tools can begin. For comparisons of the histograms, the histogram for DLLH can be considered the original set because they are the typical coordinate system used to express locations on the planet. Figure 2.10 shows the PICO-NARE site and three points that will be used for discussion of the ability of a unit system to measure the closeness of trajectories. The discussion will group the SLL and SLLH systems together and the NLL and NLLH systems together. This is done because the systems are similar in how they transform coordinate values between systems.

The histogram for the SLL and SLLH system is shown in Figure 2.19. An observa-

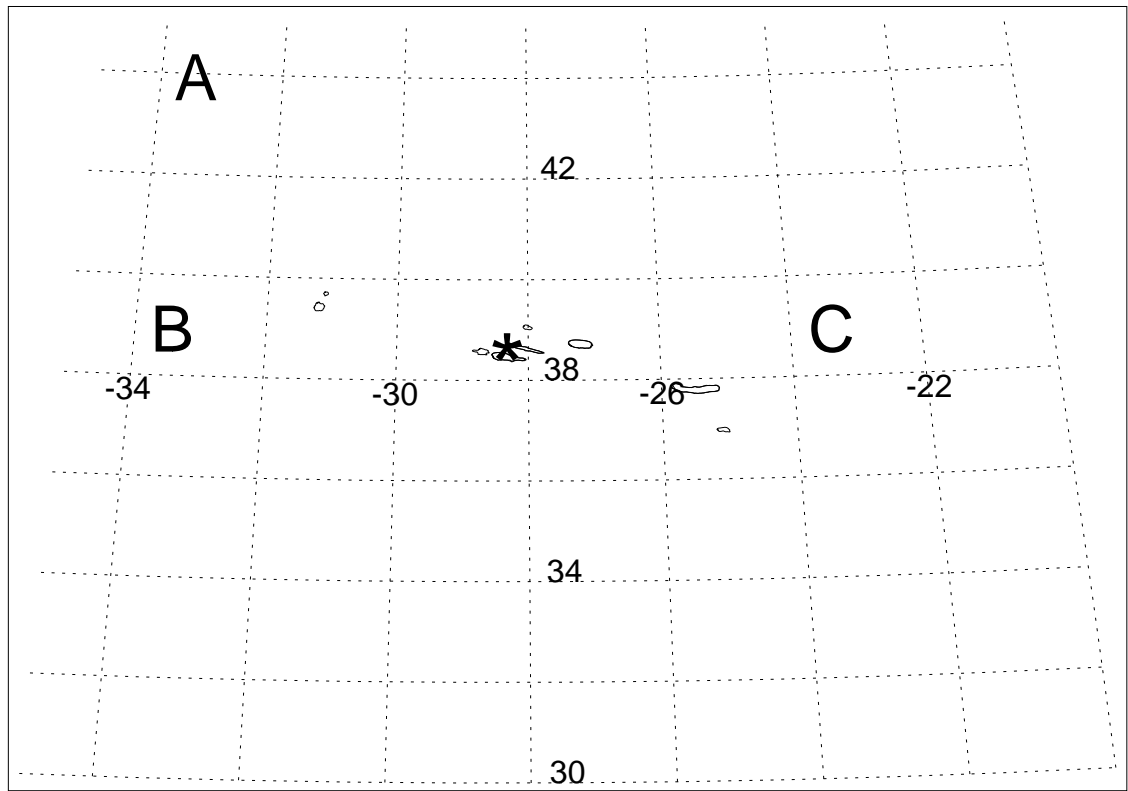


Figure 2.10 An example showing the PICO-NARE site (indicated by asterisk) and three points for use in the discussion of units. The three points are: point A, (43.47°N, -33.4°W, 2000 mASL), point B, (38.47°N, -33.4°W, 5000 mASL), point C, (38.47°N, -23.4°W, 5000 mASL).

tion that can be made from the histogram is that the magnitudes within coordinates are preserved, as the shape of each coordinate is similar to the DLLH histogram (Figure 2.13). Therefore, if a point is an out-lier within a coordinate, it will remain an outlier within that coordinate. Since it is simple to calculate the scaled units, the three points in Figure 2.10 can be used in this discussion. It is clear from the figure that points A and B are closest to each other. According to the scaling process used, the coordinates for A are (0.056, -0.046, -0.029), the coordinates for B are (0, -0.046, 0.357) and for C are (0, 0.046, 0.357). According to the Euclidean distance with the three, points B and C are closest in the SLLH system and points A and B are closest in the SLL system. This is contradictory to the apparent closeness portrayed in Figure 2.10, which demonstrates the inability of these systems to accurately represent the closeness of trajectories. Another uncertainty introduced by scaling is that this result could easily change with the same points and a different scaling unit, which was arbitrarily selected. In addition, it is unclear what the Euclidean distance is between the scaled points, as one unit of distance in each coordinate represents a different unit of distance.

The NLL and NLLH systems have disadvantages that are similar to the SLL and SLLH systems. Figure 2.22 shows that the process of normalizing the data also reduced the differences in magnitude between coordinates, as the majority of the coordinate values are contained in a range from -4 to +4 for all coordinates. The histogram also shows that the process of normalizing the data changes the differences of magnitude within a coordinate, as the shape of the histogram for each coordinate is notably different from those for the other unit systems. This indicates that this unit system may reduce the impact of outliers. However, there is no simple method to determine this. It is not possible to use the points in Figure 2.10 in direct reference to the set of trajectories because the normalization process requires the calculation of a mean and standard deviation of the data set as a whole. However, the average and standard deviation of these three points can be calculated in order to observe how

normalization would affect the apparent closeness of these points. The normalized coordinates are $(-1.84, 2.55, -46.60)$ for A, $(-1.84, -0.85, 44.08)$ for B, and $(3.07, -0.85, 44.08)$ for C. If the altitude is included, then according to the Euclidean distance, points B and C are closest. If the altitude is excluded, then points A and B are closest. The normalized data were latitude, longitude and altitude. If an alternate unit system (i.e. KLLH) were used for normalization, these results would change. If points were added, the results would likely change. For this particular example, NLL appears able to distinguish which points are closest, while NLLH does not. However, this example is very specific to these three points. Qualitatively, this shows that the the NLL and NLLH systems are “volatile”, in that any change in the data set produces changes in the values of all of the coordinates. In addition, it is unclear what the Euclidean distance is between the normalized points, as one unit in each coordinate, at each time represents a different unit of distance.

In terms of preserving magnitudes, the DLLH system has two advantages. It is the “original” coordinate system, so it clearly preserves the magnitudes within coordinates, and when only considering the horizontal components, it also preserves magnitudes between coordinates. However, the units of the horizontal and vertical coordinates are different. The DLLH system makes no attempt to minimize the impact of outliers on the clustering process because it does not reduce the differences of magnitude within coordinates. The DLLH system can easily be applied with the sample points in Figure 2.10. The coordinates for each point remain as they are originally, with the altitude in km. According to the Euclidean distance, points A and B are closest, which agrees with the visual inspection. The vertical coordinate could be converted to any number of units (*e.g.*, meters, yards, etc.), which could easily change the the importance of the vertical component on this solution. Similarly, the horizontal units could be converted (*e.g.*, radians), which would also impact this solution. In addition, the Euclidean distance between two coordinates measured in degrees and one coordinate in km is unclear.

In many ways, the KLLH system is similar to the DLLH system. An inspection of Figure 2.16 shows that KLLH preserves magnitudes within coordinates, as the shape of the histograms are similar to the histogram for the DLLH system. The KLLH system also has the advantage that all coordinates are measured in the same units. The conversion from one unit of length to another (*e.g.*, from km to inches) would result in the same clustering solution. However, the KLLH system makes no attempt to minimize the impact of outliers because it does not attempt to reduce the differences of magnitude within coordinates. When the KLLH system is applied to the sample points in Figure 2.10, the converted coordinates are (-489.6 km, 556.6 km, -0.23 km) for A, (-489.6 km, 0 km, 3.78 km) for B, and (489.6 km, 0 km, 3.78 km) for C. According to the Euclidean distance with these coordinates, points A and B are closest to each other, which agrees with the visual inspection. The KLLH system has two elements that distinguish it from the others. Firstly, that its coordinates cannot change relative to each other, as they do in the other three systems. Secondly, that when the coordinates are used to find a Euclidean distance, the meaning of that distance is clear, as it is simply a measure of length in km.

The selection of a unit system is not straightforward, but there are indicators to aid the decision. None of the systems clearly demonstrated the ability to deal with outliers, although this was not thoroughly tested. However, outliers tend to have a greater impact on smaller data sets, and since this data set is so large, it is likely that outliers have little effect on the results. The most important feature of a unit system is to accurately portray the closeness of clusters. An examination of the sample points and how the unit systems enabled the clustering criterion to determine the closeness of the points provided good insight into unit selection. It was clear that the SLL, SLLH, NLL and NLLH systems were not always able to represent the closeness of points. In addition, the manner in which NLL and NLLH portrayed the nearness of points was sensitive to changes in the data set, and all four of these unit systems were created by an arbitrary selection of units and scaling factors. Because none of these appeared to

consistently represent the closeness of points, there were eliminated. The remaining two systems, KLLH and DLLH, appeared to do a better job of this. However, the DLLH system used different units in the horizontal and vertical coordinates, which means its ability to gauge the closeness of trajectories could change by an arbitrary selection of different units for either coordinate. Therefore, this impaired its ability to accurately represent the closeness of trajectories. On the other hand, the KLLH system used the same units for all three coordinates, and even if different a unit of length were chosen, the conversion should not impact the results of the clustering process. Because of these factors, the KLLH unit system was selected for use in the analysis.

2.2.3.4 Alternate unit system, GLLH

GLLH, an additional unit system/clustering criterion was mentioned that was not considered for use in the final clustering solutions. It was not considered for use because it was not developed early enough to analyze in the same manner the other systems were analyzed. However, it presents a method that could be the best method for measuring the closeness of trajectories, and it eliminates the need to consider alternative unit systems. It should be noted as a possibility for future applications. When determining the closeness of a trajectory to a cluster center, the GLLH system measures the great circle distance between the pair of points expressed in km. It then uses the calculated great circle distance between two horizontal points and the difference in altitude between two points to calculate a Euclidean distance between points. This distance is summed over all the pairs of points to determine the “total distance” from a trajectory to a cluster center. The great circle distance measures the shortest possible distance between two points over the earth’s surface, making it a good method for determining the closeness of trajectories. As with the KLLH and DLLH systems, the addition of height made no discernible difference in the cluster solution. When applied to the three sample points in Figure 2.10, it correctly

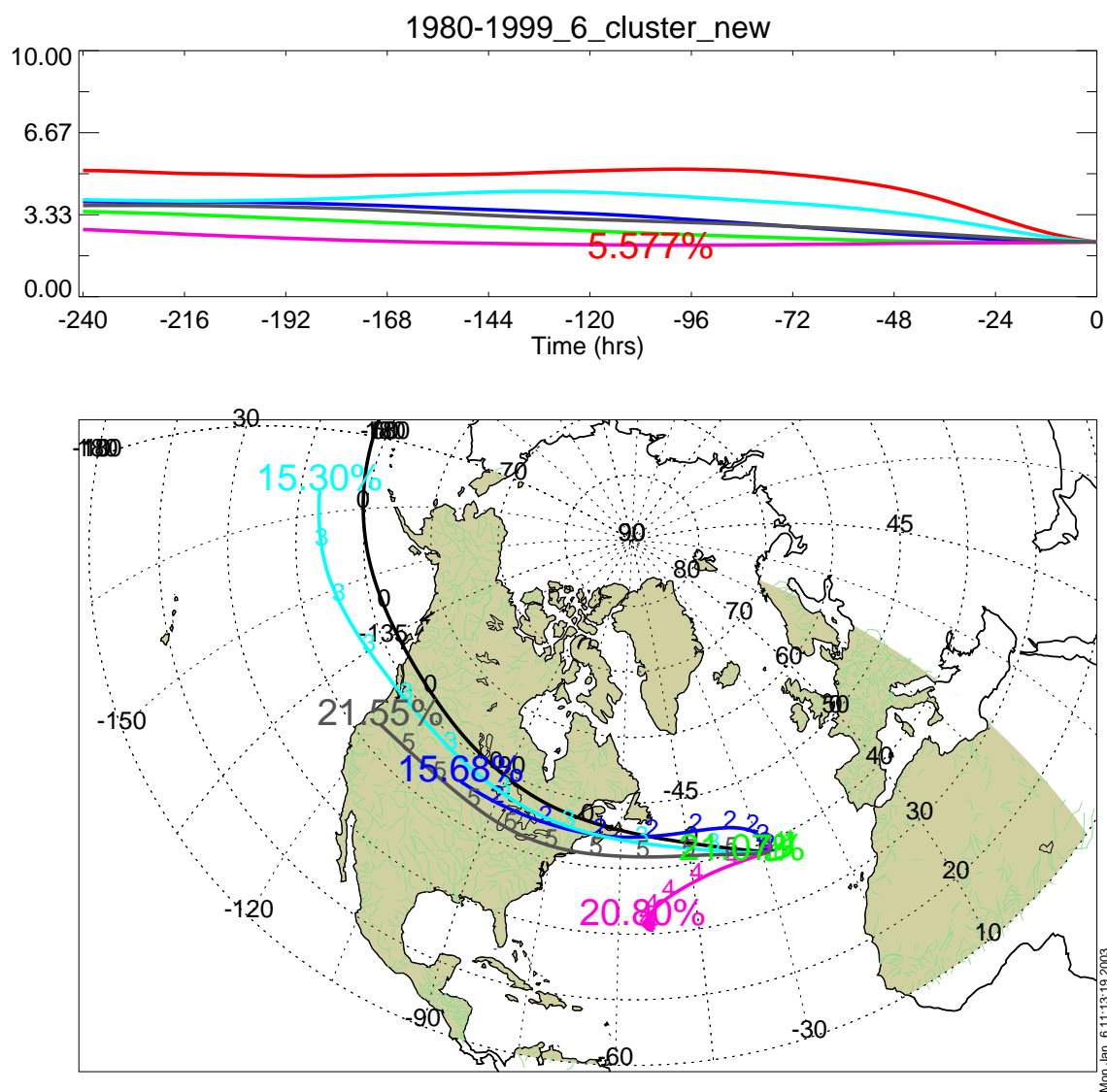


Figure 2.11 Plot of the cluster centers for DLL for 6 clusters.

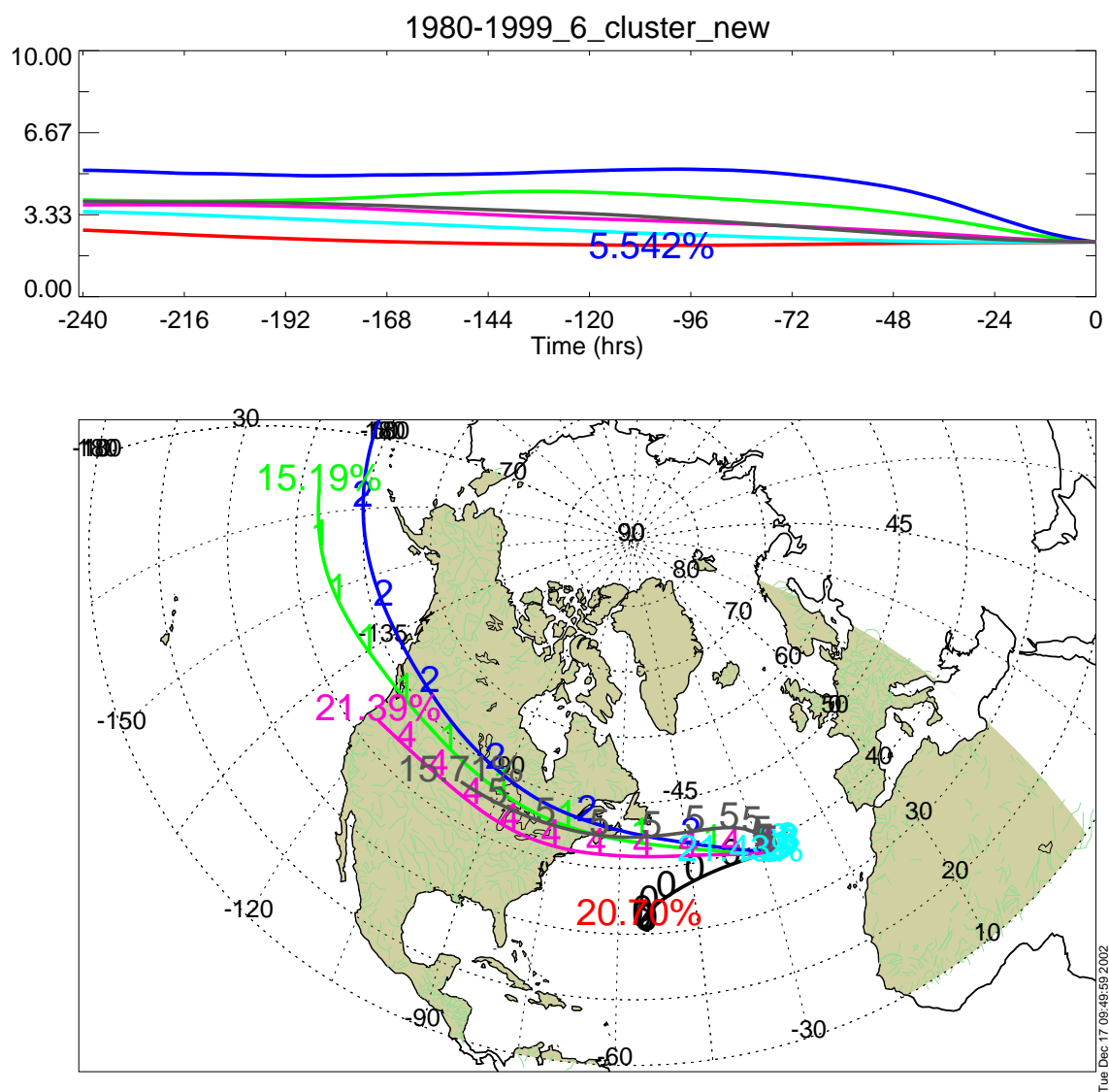


Figure 2.12 Plot of the cluster centers for DLLH for 6 clusters.

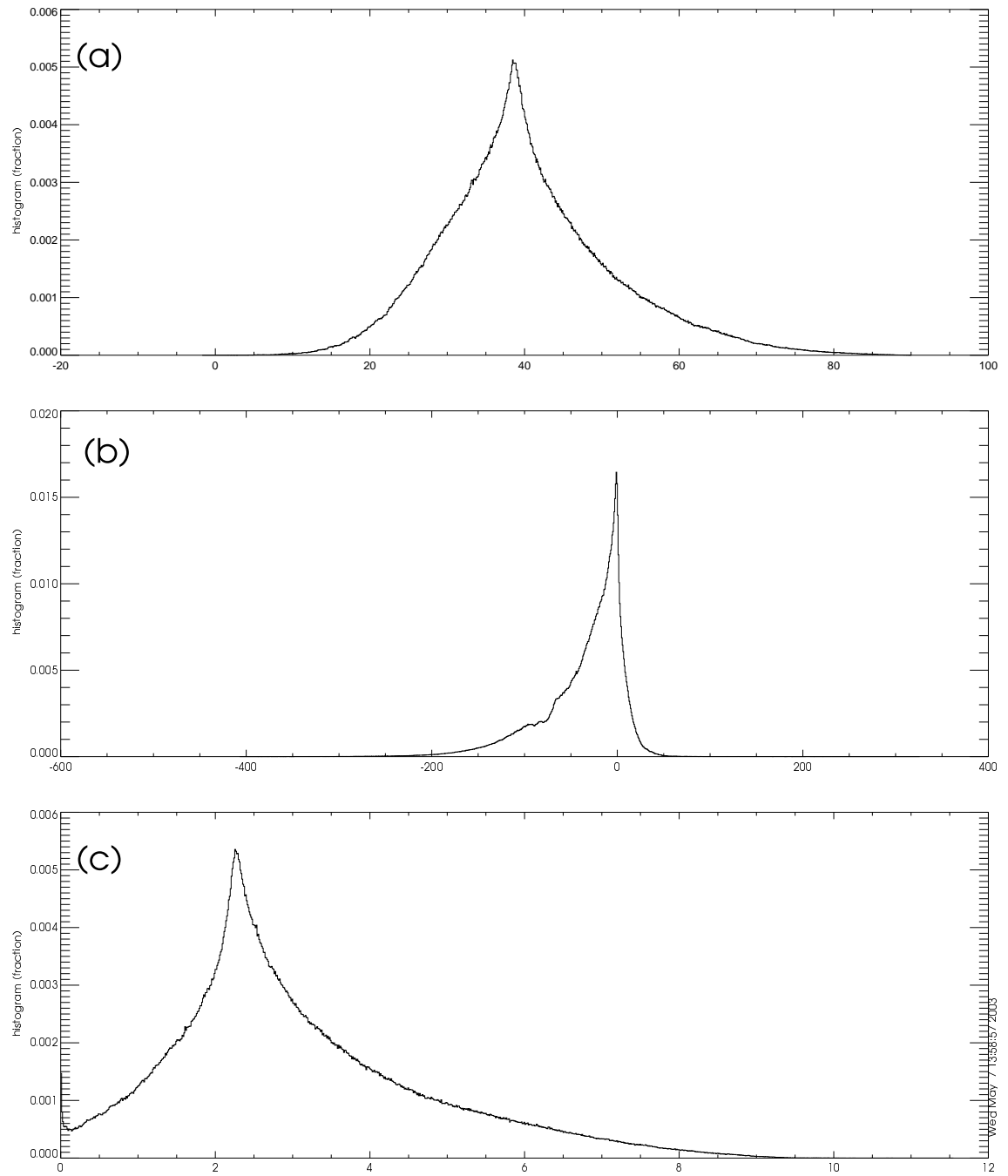


Figure 2.13 Histogram for the coordinates for the DLL and DLLH unit systems. (a) shows the latitudes, (b) shows the longitude, and (c) shows altitude (km).

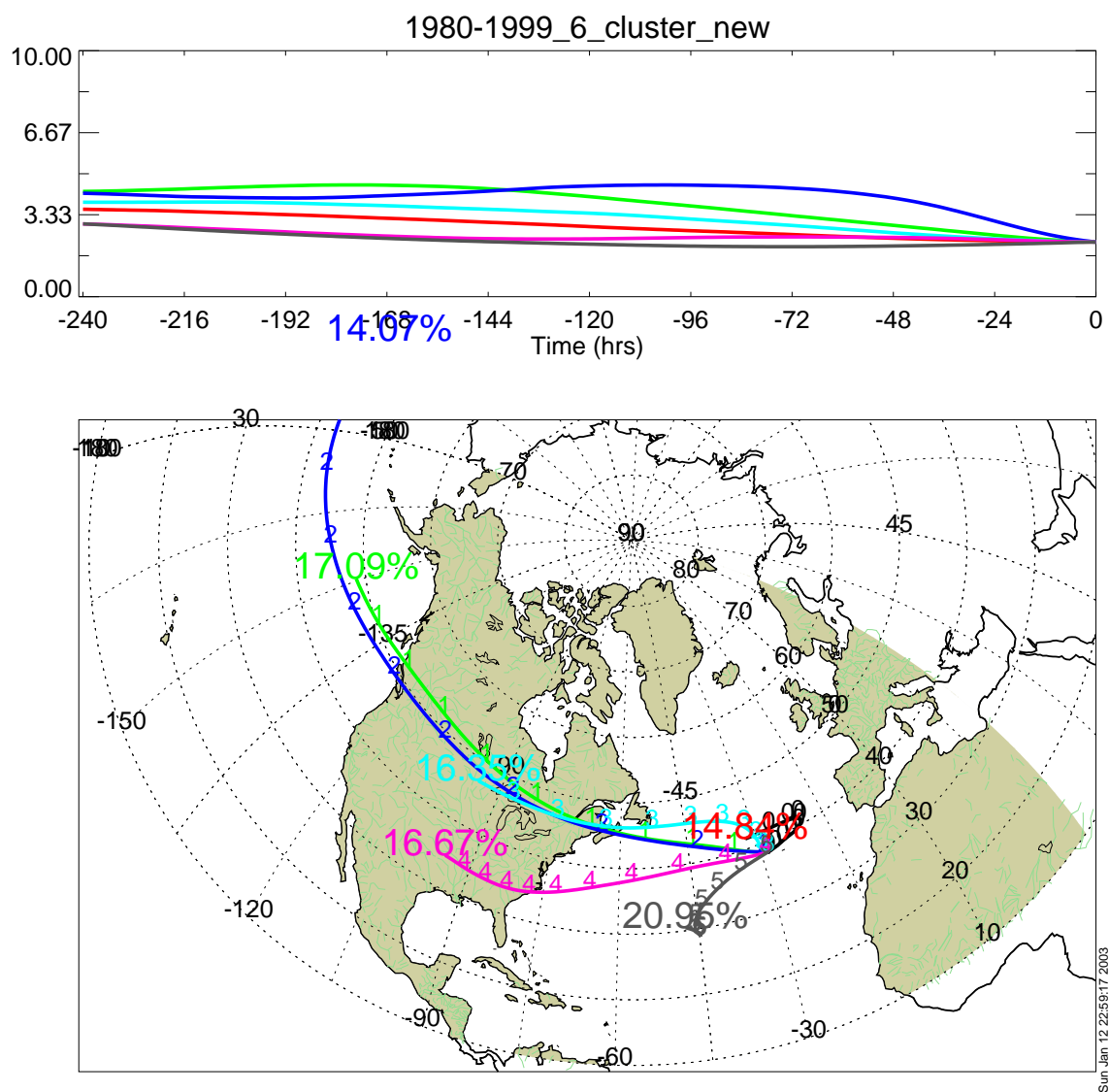


Figure 2.14 Plot of the cluster centers for KLL for 6 clusters.

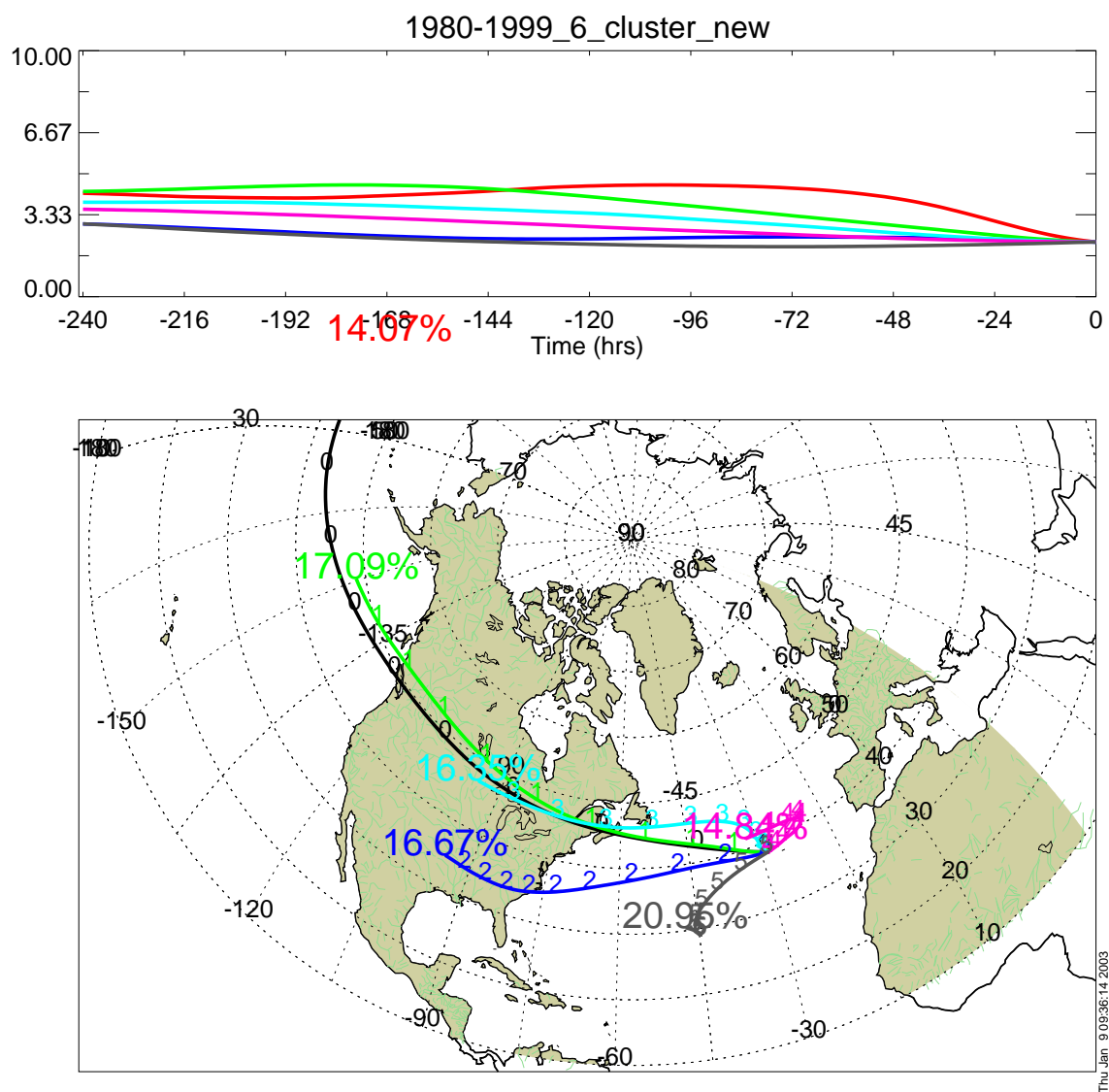


Figure 2.15 Plot of the cluster centers for KLLH for 6 clusters.

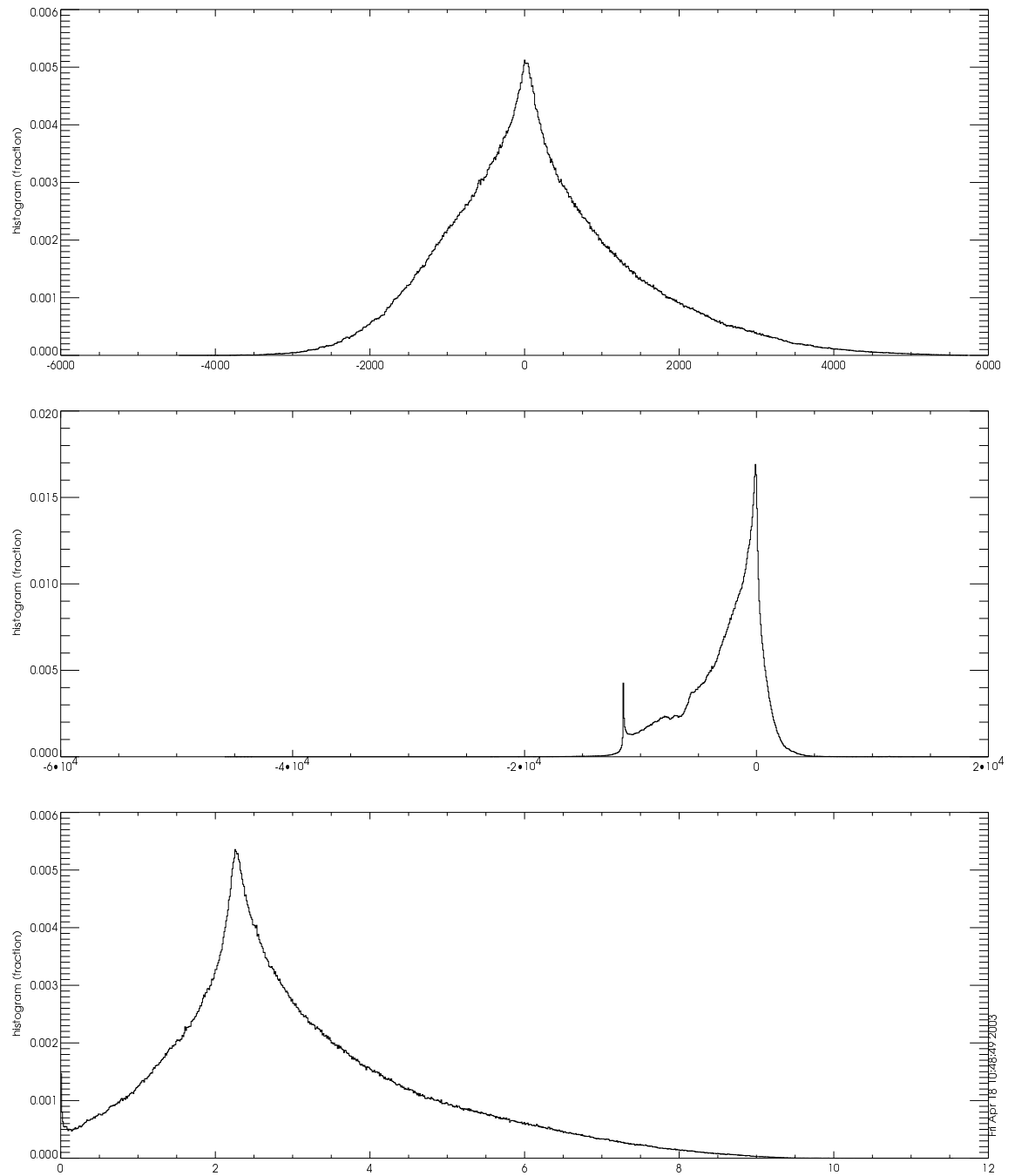


Figure 2.16 Histogram for the coordinates for the KLL and KLLH unit systems. (a) shows distances north/south (km), (b) shows distances east/west (km), and (c) shows altitude (km).

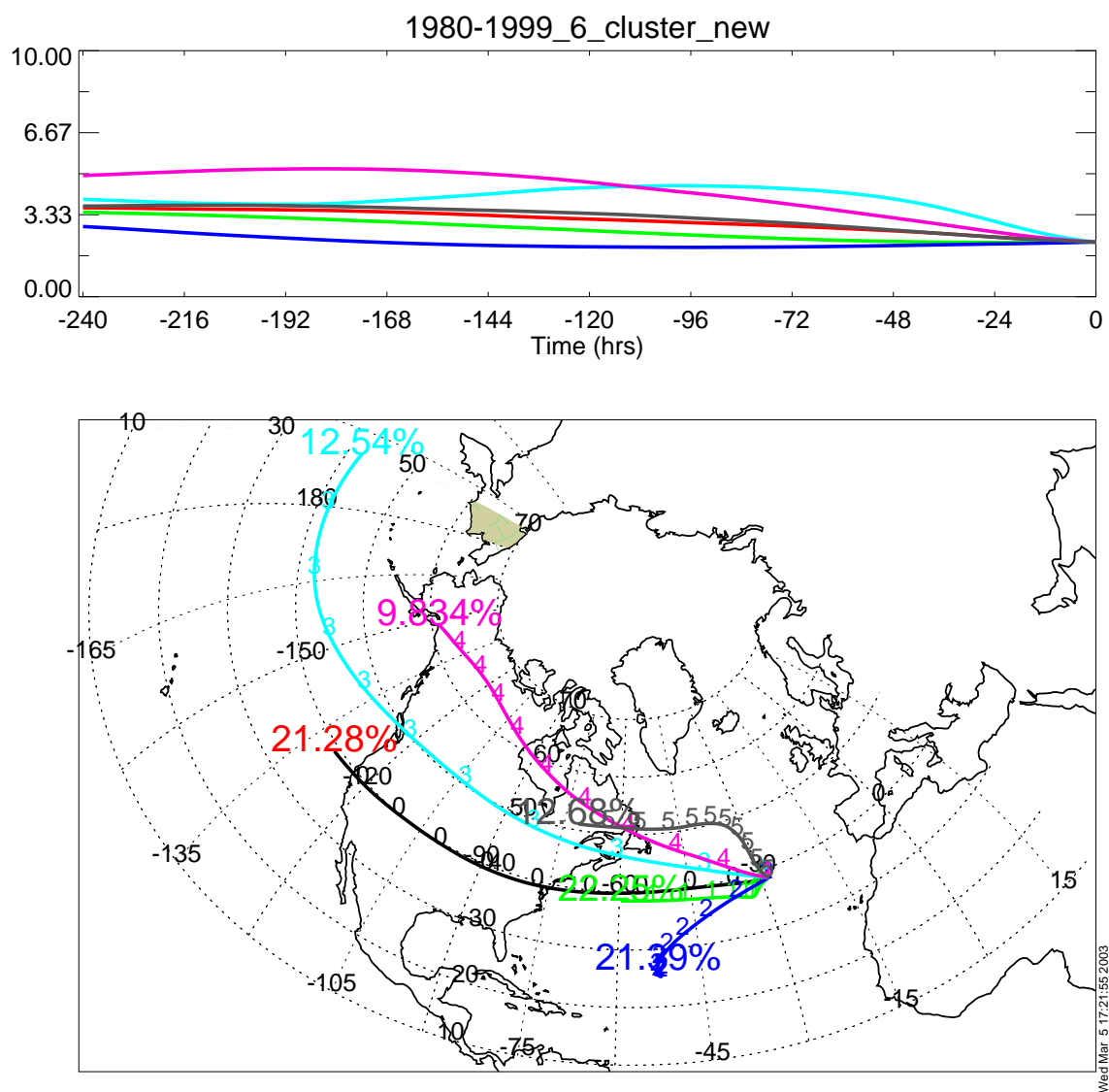


Figure 2.17 Plot of the cluster centers for SLL for 6 clusters.

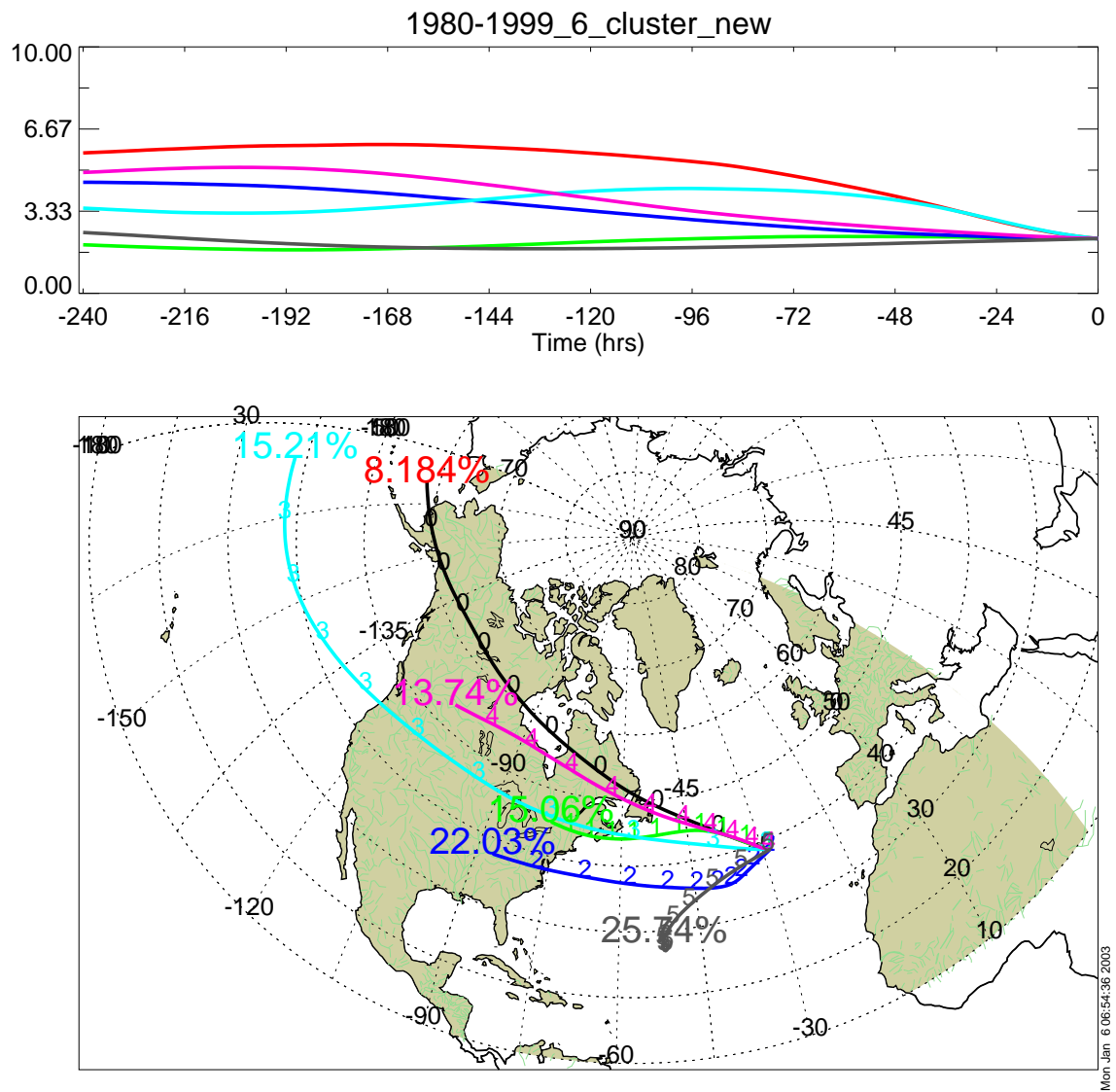


Figure 2.18 Plot of the cluster centers for SLLH for 6 clusters.

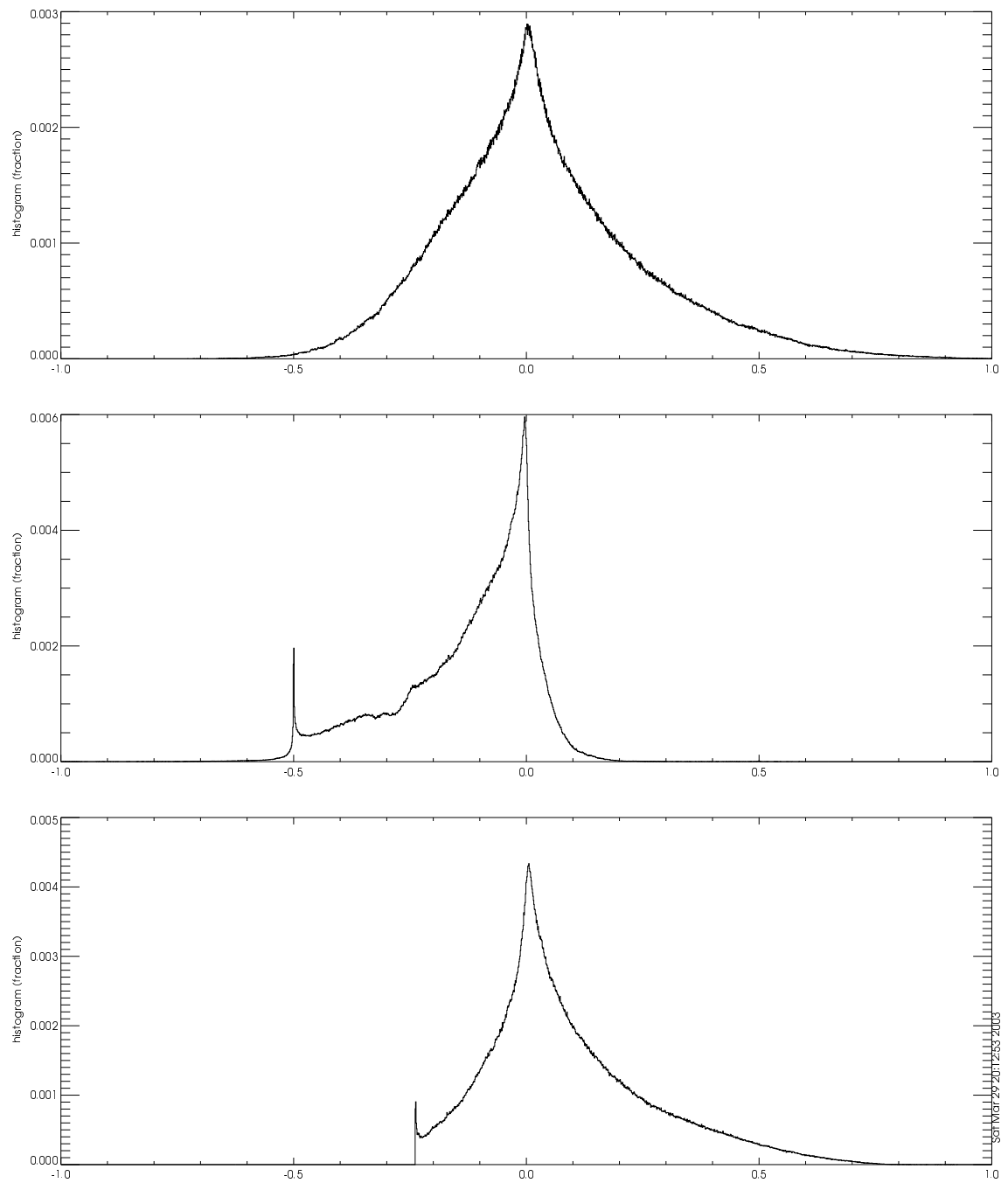


Figure 2.19 Histogram for the coordinates for the SLL and SLLH unit systems. (a) shows the standardized distances north/south (km/km), (b) shows the standardized east/west (km/km), and (c) shows the standardized altitude (km/km).

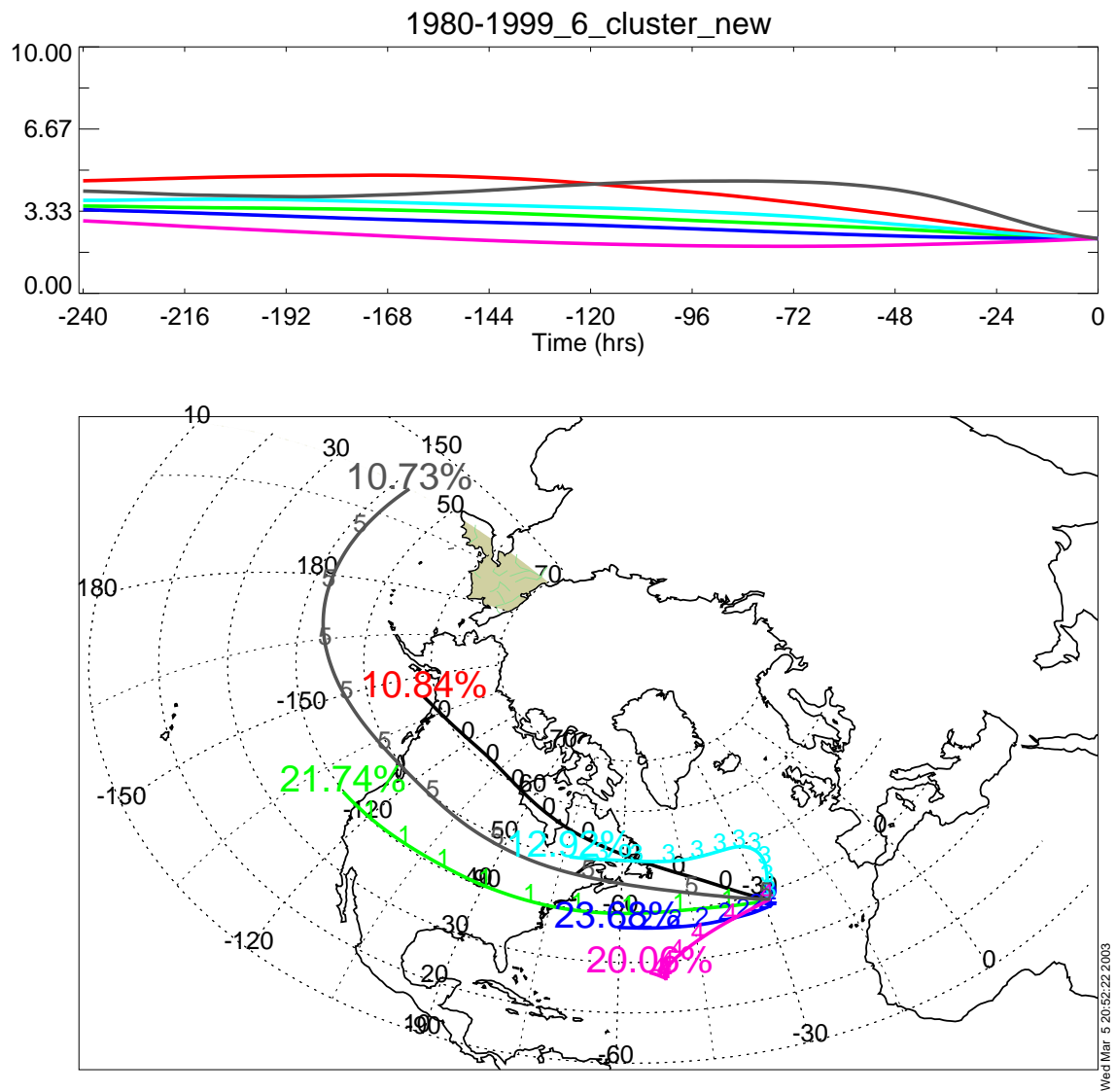


Figure 2.20 Plot of the cluster centers for NLL for 6 clusters.

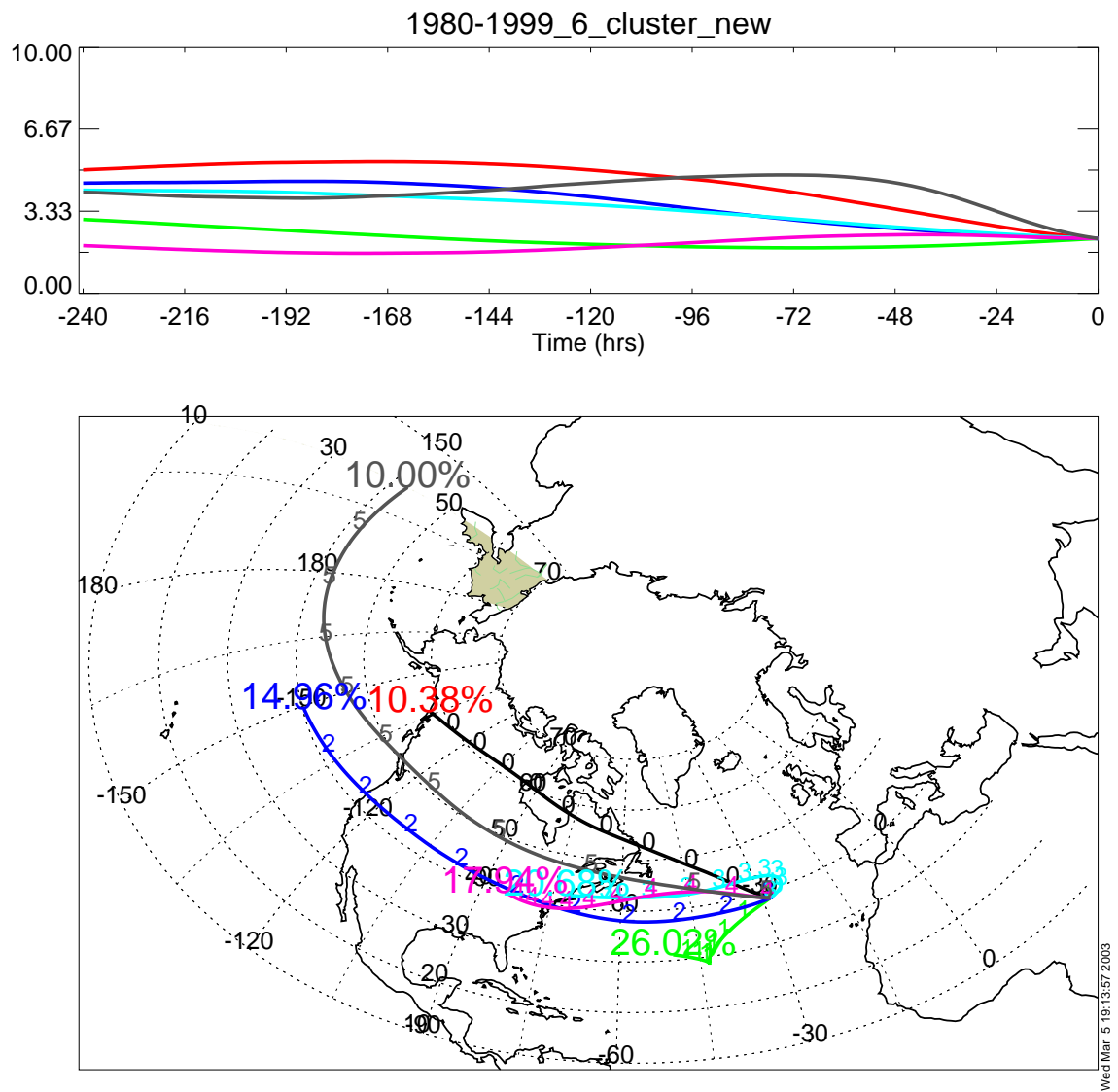


Figure 2.21 Plot of the cluster centers for NLLH for 6 clusters.

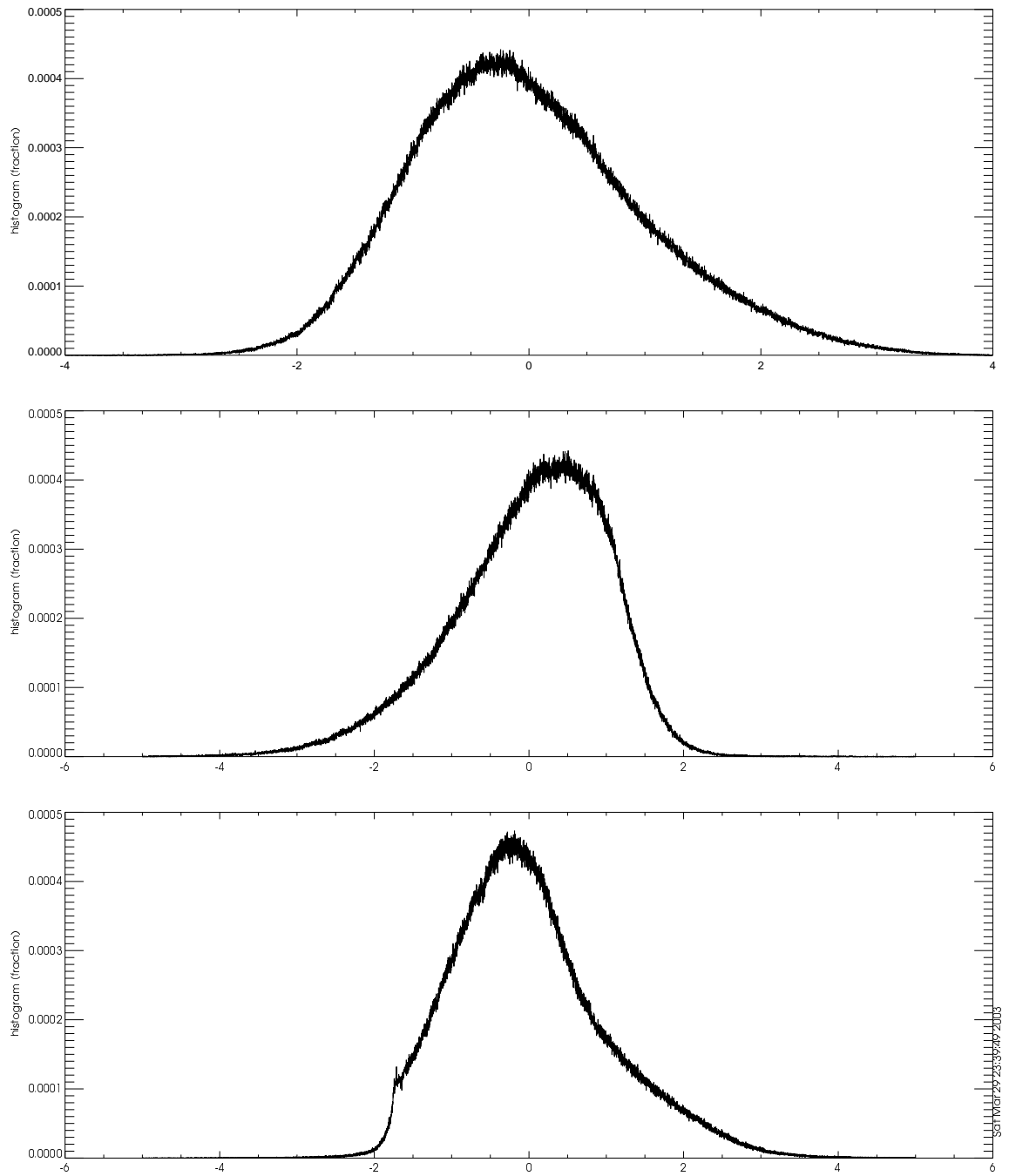


Figure 2.22 Histogram for the coordinates for the NLL and NLLH unit systems. (a) shows the normalized latitudes, (b) shows the normalized longitude, and (c) shows the normalized altitude (unit less).

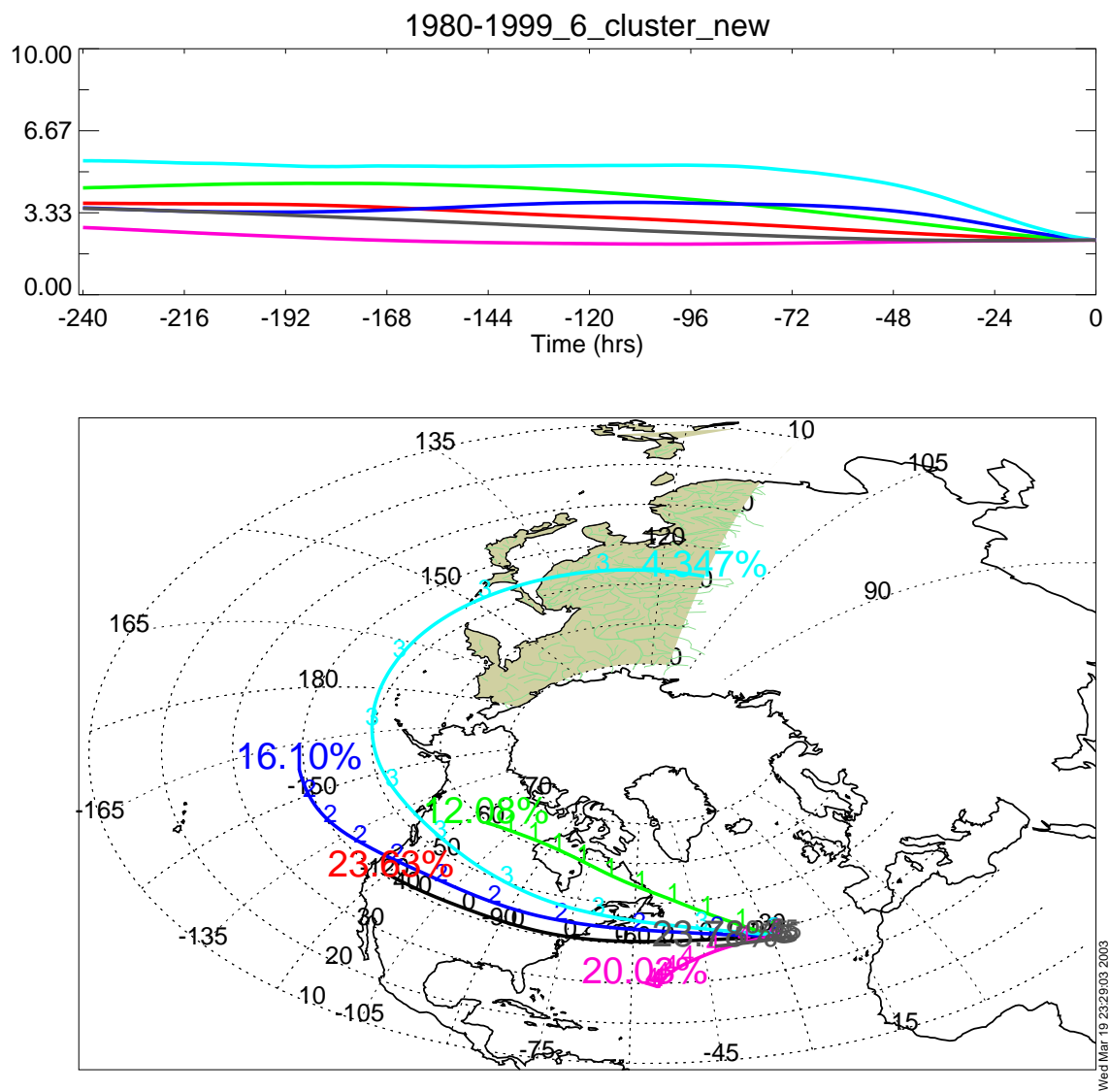


Figure 2.23 Plot of the cluster centers for GLLH for 6 clusters.

recognized points A and B as closest. This method should produce clusters that contain trajectories that are physically closest to each other.

2.2.4 Note on clusters

Before ending the discussion of cluster analysis, one comment should be made for consideration when viewing cluster center plots. A cluster center, which is plotted in the cluster plots, represents the average of many trajectories. This means that there can be considerable deviation of the member trajectories from the average trajectory. The cluster analysis determines the primary features of the trajectories. However, it cannot capture all the features without using a very large number of clusters. A large number of clusters would be able to capture many unique flow patterns, but this would be contrary to goal of clustering in this study, which is to reveal the dominant patterns within the trajectory set. The degree to which cluster members deviate from the average trajectory can be determined from cluster membership plots, and can be viewed in order to see the trajectories that are represented by each cluster. In some cases, because of the deviations from the average cluster, cluster centers can represent flow paths that are not directly shown by the cluster center. This could have important implications for many of the clusters. Some clusters that might seem well defined to represent, for example, North American continental flow, might also transport clean air from remote regions or Arctic air. While all the possibilities of transport for each cluster are not discussed, the appendix includes cluster membership plots for each cluster, as well as a density and corrected density plot for the trajectories belonging to each cluster. These tools can be used for further inspection of specific clusters when questions arise about the possible additional flow paths within a cluster.

2.3 Airflow Probability Analysis

The cluster analysis helps to reduce the size of the data set into manageable clusters and to identify the dominate pathways that air travels to the site. Equally important to the pathways that air travels is the regions the air spends time over. Airflow prob-

ability analysis is an effective method for exploring this aspect of a set of trajectories.

2.3.1 Application and theory

Cluster analysis and airflow probability analysis (APA) are useful methods for extracting information from a large set of trajectories. However, the two are quite different in both the methods used and the types of information they convey. In terms of the method of application, a cluster analysis deals with each trajectory as a whole, considering all 241 endpoints as a unit. It compares each trajectory's 241 endpoints to other trajectory's' 241 endpoints. On the other hand, APA considers each trajectory's endpoint independently of the other endpoints of that trajectory. This is not to say it ignores all but one endpoint, but it does not deal with the endpoints of a trajectory as a whole. So opposed to a set of 28,542 trajectories, each with 241 endpoints, there is simply a set of 7,014,787 endpoints. In terms of information conveyed, cluster analysis reveals the dominant paths air follows as it flows to the site, while an APA shows the regions of the earth that air typically spends time over before it arrives at the site. By dealing with just the trajectory endpoints as opposed the set of endpoints that makes a trajectory, the APA helps regions where air is likely to spend time over as opposed to showing the path that air takes when traveling to a site.

Airflow probability analysis as used in this study was developed by *Poirot and Wishinski* [1986] for an analysis of aerosol particle measurements in northern Vermont. It was used to identify source regions that contributed to increased levels of particulate matter observed at their site. It has been used in several other studies since then. *Merrill et al.* [1989] used APA with the analysis aerosol samples taken over the Pacific ocean during the period 1981-1986. *Merrill* [1994] used APA to identify source regions for Atmosphere/Ocean Chemistry Experiment for the period 1986-1993. *Sirois and Bottenheim* [1995] used APA to explore 5 years of measurements of

PAN and O₃ make in Kejimikujik National Park in Nova Scotia. *Moody et al.* [1995] identified source regions for measurements of O₃ at Bermuda during 1988-1991 using APA. An extended method of APA known as potential source contribution function (PSCF) allows the emphasis of regions traversed by trajectories identified to be associated with pollution events. *Cheng et al.* [1993] used PSCF to identify source regions for aerosol measurements at Alert, Northwest Territories. *Lin et al.* [2001] also used PSCF to identify source regions of gaseous mercury measurements at Alert.

2.3.2 Methods

The application of APA is relatively simple compared to a cluster analysis. It is described in detail in *Poirot and Wishinski* [1986], but the methods employed in this study will be explained below. The basic concept used is based on the amount of time air spends over regions of the earth. The longer the time spent over a region, the more likely that air is going to be affected by that region. A region can affect an air parcel in two ways. First, if a region has pollution sources, then air passing over this region can be mixed with polluted air. Second, if a region does not have pollution sources, then air over this region can be mixed with clean air, diluting polluted air. The time spent over a particular region is known as residence time. The residence time can be calculated by counting the number of hours spent by trajectories over a particular area. In order to do this, the globe was divided into grid cells, 1° on each side. The number of endpoints in each grid cell was then counted. The number of endpoints in a grid cell is used to represent the number of hours spent by trajectories over that grid cell. This is a good estimate because an endpoint is calculated every hour and it is likely that a trajectory will spend approximately an hour entering and leaving the grid cell in order to have an endpoint in it. Even the smaller grid cells have an area of 423 km², which translates to a 20x20 km box. This means an air parcel would have to be traveling 20 km/hr (5.6 m/s) to pass over the smallest grid

box without having an endpoint within that box. While this speed is not uncommon, this is a worst case scenario with most grid boxes being considerably larger. Even in this event, it is likely that an adjoining grid box would have one of these trajectories endpoints and adjoining endpoints would be indiscernible on the resulting graphs.

Each grid cell is not equal in size, with grid cells closer to the equator being larger than those near the poles. This occurs because the grid cells were delineated using degrees as opposed to km, or some other measure of distance. For example, the grid cell bounded by 2°N, 3°N, -28°W, and -29°W has an area of approximately 12,382 km². While the grid cell at the same longitude, but bounded by 87°N and 88°N has an area of approximately 510 km². The area of a grid cell near the equator can be more than 25 times larger than one near the poles. To correct for this fact, each endpoint count was multiplied by a correction factor. The correction factor was the ratio of the area of a grid cell at the equator to the area of each grid cell:

$$C_{x,y}^* = C_{x,y} * \left(\frac{A_{equator}}{A_{x,y}} \right), \quad (2.16)$$

where $C_{x,y}^*$ is the new endpoint count in grid cell x, y , $C_{x,y}$ is the actual count for grid cell x, y , $A_{equator}$ is the area of a grid cell at the equator, and $A_{x,y}$ is the area grid cell x, y . This results in endpoint counts for each grid cell that approximates the endpoint count that a particular grid cell would have received if it were the same size of a grid cell at the equator.

Once the count of all the endpoints in all the grid cells has been completed and the area correction applied, a probability function (equation 2.17) can be applied that will give the likelihood of a trajectory passing over that grid cell. If there are N trajectories and grid cell i, j has $n_{i,j}$ endpoints in it, then the probability that air flows over that grid cell is:

$$P[A_{i,j}] = \frac{n_{i,j}}{N}. \quad (2.17)$$

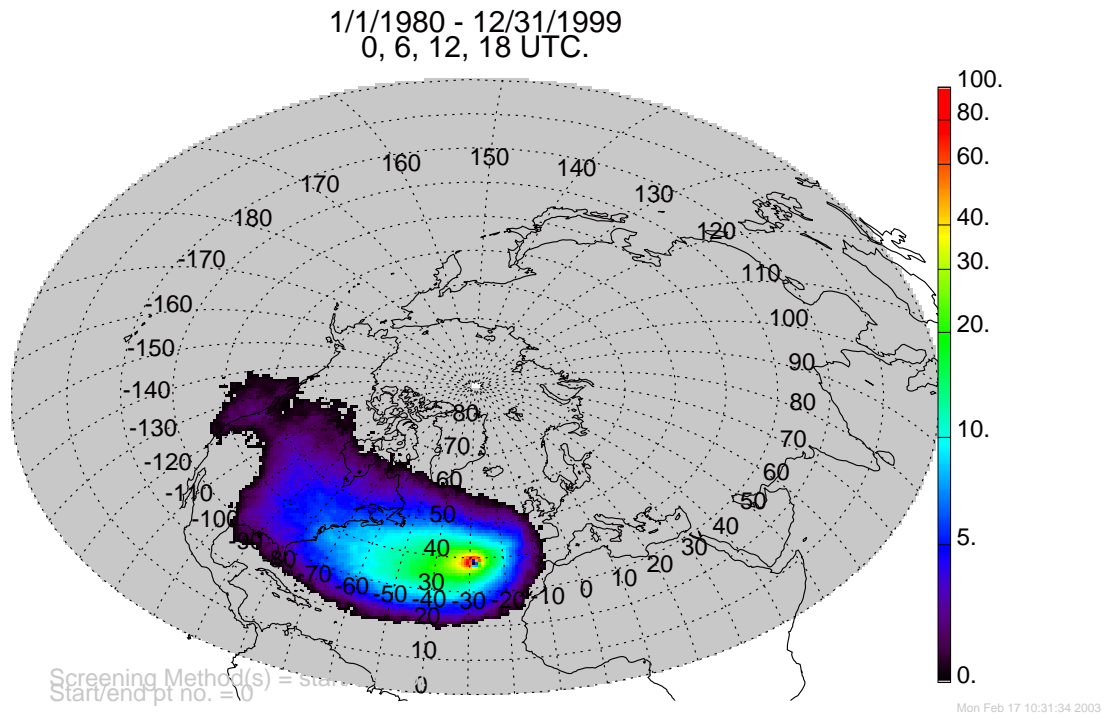


Figure 2.24 The standard airflow probability analysis plot for the period 1960-1999.

The lowest 10% of the probability values, that is $P[A_{i,j}] < 0.1$ are set to 0. This is done for two reasons. First, the regions of lower probabilities are not significant source regions. Second, it helps to reduce noise when plotted, giving the plots a smoother outer region. Each probability value is assigned a color, ranging from red (highest probability) to black (lowest probability), and plotted on a map of the regions covered. The result of this process is shown in Figure 2.24, which shows the results for the APA for a 40-year data set, 1960-1999.

One feature of that is apparent in Figure 2.24 is the higher probability near the site, decreasing rapidly away from the site. This is not surprising, as all trajectories arriving at the site will spend some time in the grid cells near the site. It is intuitive that the probability a grid cell will contain trajectories will decrease the further the grid cell is from the site. Consider two grid cells, (x_1, y_1) and (x_2, y_2) , each with with

an area of 1 km^2 , at a distance of R_1 and R_2 respectively from the site, where R_1 is greater than R_2 . The number of grid cells at the same distance as grid cell (x_1, y_1) can be found by dividing the area of an annular ring containing the grid cell by the area of the grid cell:

$$N_{grid \text{ cells}} = \frac{A_{ring}}{A_{cell}} = \frac{(2 * \pi * R_1) * 1 \text{ km}}{1 \text{ km}^2} = (2 * \pi * R_1). \quad (2.18)$$

Similarly, the number of grid cells that are at the same distance from the origin as grid cell (x_2, y_2) would be $(2 * \pi * R_2)$. Therefore, the likelihood of passing over any cell at a distance of R_1 is R_2/R_1 less than passing over any grid cell at a distance of R_2 . To account for the decreased probability due to the distance from the site, the count in each equal-area grid cell is multiplied by the the great circle distance calculated from the PICO-NARE site to each grid cell. These adjusted values are then assigned colors and plotted as in the previous procedure. The plots that are corrected for the probability induced by geometry are referred to as geometrically corrected. An example can be seen in Figure 2.25, which shows the plot for a 40-year data set, 1960-1999.

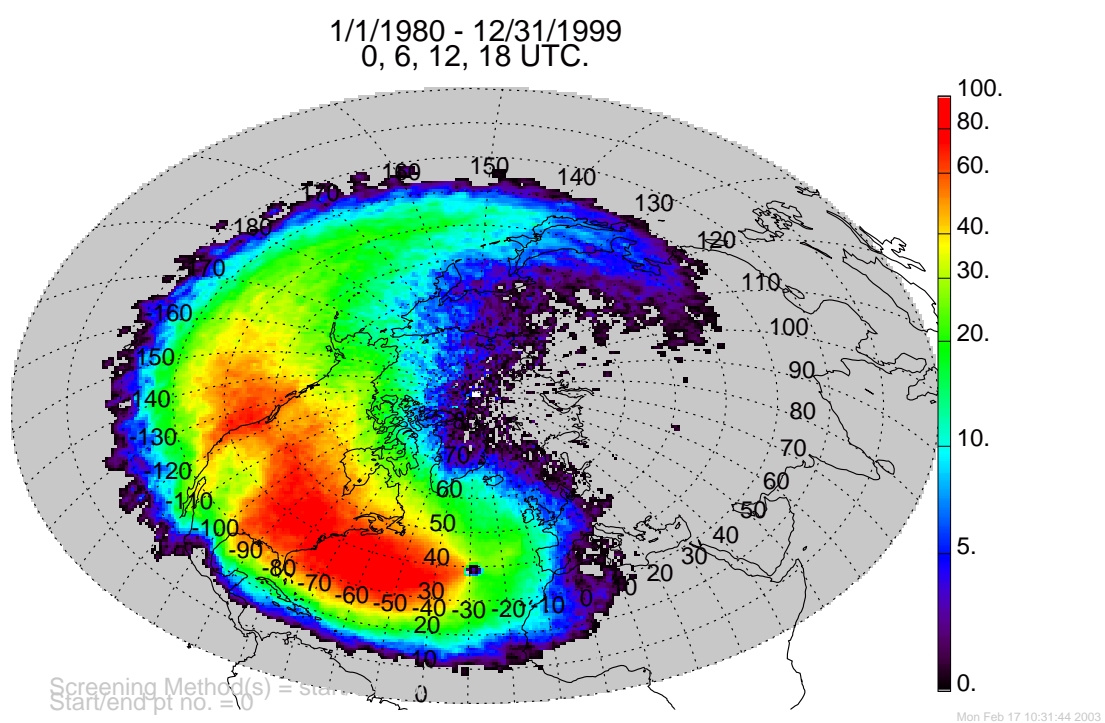


Figure 2.25 The geometrically corrected airflow probability analysis plot for the period 1960-1999.

2.4 Impacts of the NAO on the climatology of the PICO-NARE Site

It is important in the development of a climatology to account for any temporally induced changes in the flow patterns. Short-term variations, such as seasonal variations, are fairly straightforward to deal with. However long-term variation is often a more complicated issue for two reasons. First, historic data often does not extend far back enough to observe long-term climatic change. Second, when the long-term climatic change is known, it is often difficult to discern long-term climate variation from year to year variability. The North Atlantic Oscillation is a long-term climate variation that affects the central North Atlantic, with records dating back to the mid-1800's, which can be used to observe this variation.

Attempts have been made to reconstruct an NAOI as far back as AD 1400 using paleoclimate records, which show that the NAO persists back farther than the available SLP records [Cook, 2003].

2.4.1 Climatic variation and the NAO

Synoptic climatic conditions can have large variations on a temporal scale. The dominant short-term variations that can be accounted for are seasonal changes. Seasonal changes can be assessed by selecting trajectories that begin during each season, usually identified by different months. For the central North Atlantic, the winter season has been identified as being December, January, and February, summer as June, July and August, and spring and fall the three months in between [Marshall *et al.*, 2001; Jonsson and Miles, 2001; Hurrell *et al.*, 2003]. Seasonal variation will be the shortest time scale investigated in this study.

In addition to short-term variations, long-term variations can have an effect on synoptic climatic conditions as well. One of the oldest known weather patterns is

a variation between winter weather in Greenland and eastern Europe [Stephenson *et al.*, 2003]. This variation is driven by differences in pressure centers over the North Atlantic, known as the North Atlantic Oscillation (NAO) [Jonsson and Miles, 2001]. The NAO has received increasing interest over the past few years, with the number of papers published with the NAO in the title or abstract increasing dramatically from only 3 in 1991 to 179 in 2001 [Stephenson *et al.*, 2003]. The NAO is the variation of sea level pressures between two centers of action, the Icelandic Low and the Azores High. The Icelandic Low is a semi-permanent low-pressure system that has its center of action over Iceland. The Azores High is a semi-permanent high pressure that has its center of action spanning from Bermuda over the Azores to eastern Europe.

The NAO is usually examined by calculating differences between the two pressure centers. The values, known as the NAO Index (NAOI), are calculated by using normalized averages of the sea level pressures at one of two sites in Iceland and at one of several sites contained within the Azores High. In Iceland, the sea level pressure record from either Reykjavik or Stykkisólmur may be used, while possible sites for the Azores High sea level pressure include Ponta Delgada in the Azores, Lisbon in Portugal, Gibraltar, Cadiz in Spain and Bermuda [Jonsson and Miles, 2001; Jones *et al.*, 2003]. Because there are a number of sites for measuring sea level pressure as well as a variety temporal scales to use, there are several possible ways to calculate the NAOI [Hurrell *et al.*, 2003]. Since the shorter term values (*e.g.*, hourly or daily) for the NAOI between any pair of locations can have fairly high variability, switching from a large positive value to a large negative value (or the converse), generally longer time scales are used. However, if too long of a time scale is used (*e.g.*, a year or decade) changes in variation in the index can be lost. Most reported indices are monthly or seasonal, since they tend to relate the long-term variability and not to the noise of short-term meteorological events [Jonsson and Miles, 2001; Marshall *et al.*, 2001; Hurrell, 2001]. The sea level pressures are normalized by dividing the difference between the long-term average by the standard deviation for that period (*e.g.*, spring)

at the site (*e.g.*, the Azores). This is done because the absolute variability of the sea level pressure at the Icelandic sites are up to four times greater than the sea level pressure at the sites for the Azores High [Jonsson and Miles, 2001; Hurrell *et al.*, 2003]. Because this study will focus on seasonal variability, seasonal NAOI values will be used.

In addition to the temporal scale to be used when calculating the NAOI, the sites for sea level pressure measurements must be selected as well. The measurements at the two Iceland sites, Reykjavik and Stykkisólmur are similar, as the two sites are about 125 km apart and the selection of one over the other is not critical [Jones *et al.*, 2003]. The selection of the site for sea level pressures around the Azores High has more impact on the resulting NAOI. The location of the center of action for the Azores High can vary with time, so it is difficult to choose which of the sites that are affected by this pressure system to use for calculating the NAOI [Jones *et al.*, 2003]. It is important for the AH site to contain a record dating back as far as possible because time averaged values are used to calculate the NAOI. The site used for the AH are most often Ponta Delgada in the Azores and Lisbon, while Gibraltar, Bermuda and Cadiz are used less frequently. The locations and period of record for each site are listed in Table 2.3. It is most appropriate to use pressure values obtained from Ponta Delgada, in the Azores, since these values are expected to capture how the NAO affects the Azores specifically.

The NAO has a variety of impacts on both the North Atlantic and surrounding regions. The most prevalent results are the zonal winds across the Atlantic. Since air flows counterclockwise around low pressure and clockwise around high pressure in the Northern Hemisphere, westerly flow during high NAO periods is enhanced, up to 8 ms^{-1} stronger than low NAO periods Hurrell [1995b]; Hurrell *et al.* [2003]. The enhanced air flow in turn affects the temperatures and rainfall across the region. During the positive phases of the NAO, the climate over the North Atlantic and Mediterranean regions are colder and drier than usual, while the climate in northern

Table 2.3 The location and period of record for sites recording SLP for use in the calculation of the NAOI [Hurrell, 1995b; Jones *et al.*, 2003].

Site	Latitude (°N)	Longitude (°W)	Period of Record
Reykjavik	64.0	22.0	1821-current
Stykkisholmur	64.0	22.8	1846-1990
Ponta Delgada	37.7	25.7	1865-1997
Lisbon	38.7	9.1	1855-current
Gibraltar	36.2	5.4	1821-current
Bermuda	32.2	64.7	1837-current
Cadiz	36.5	6.3	1786-1998

Europe, the eastern U.S. and Scandinavia are warmer and wetter than normal [Visbeck *et al.*, 2001; Marshall *et al.*, 2001; Stenseth *et al.*, 2002; Hurrell, 2001]. These changes in climatic parameters have been found to correlate with many other natural processes, such as marine fish assemblages and productivity; terrestrial vegetation; patterns of land animals such as predator/prey relationships; and even stratospheric cooling and column losses of ozone northward of 40°N [Visbeck *et al.*, 2001; Attrill and Power, 2002; Stenseth *et al.*, 2002; Hurrell *et al.*, 2003]. The effects of the NAO have been found to be the leading mode of regional circulation variability across the North Atlantic and Europe [Jonsson and Miles, 2001; Marshall *et al.*, 2001; Hurrell *et al.*, 2003; Jones *et al.*, 2003; Cook, 2003]. Considering the profound impact the NAO has on the region, especially in regard to the circulation patterns, it is clear why it is important to consider the effects it has on the climatology of the PICO-NARE site.

2.4.2 Methods

The resulting seasonal NAOI calculated from the SLP between Ponta Delgada, Azores and Stykkisholmur/Reykjavik, Iceland is shown in Figure 2.26 as obtained from [Hurrell, 1995a]. The seasonal (DJF, MAM, JJA, SON) values are the mean of each seasonal pressure divided by the long-term mean standard deviation. There is clearly

Table 2.4 Summary of the seasonal periods of NAOI used to differentiate between the positive and negative phases.

NAO phase	Season	Number of seasons	Number of trajectories
Positive	Winter	18	6531
Negative	Winter	18	5704
Positive	Spring	19	6531
Negative	Spring	12	6508
Positive	Summer	15	6839
Negative	Summer	22	5099
Positive	Fall	18	6810
Negative	Fall	19	4288

considerable variability from year to year and even from season to season. Consider 1888 and 1889 when the NAOI shifted from -2.94 in 1888 to $+2.57$ in 1889. A significant seasonal shift can be seen in 1973 when the winter value of 2.40 drops to -1.40 . Despite the dramatic shifts that can occur, there are periods in the yearly values that remain either positive or negative for extended periods of time. Possibly the most pronounced is the period from 1903 to 1914, where every year except one is positive. Another example is from 1961 to 1971 when every year but one was negative. Figure 2.27 shows the NAOI for the study period.

In order to clearly distinguish the effects of positive and negative NAOI, seasons with positive or negative NAOI were grouped together and clustered. Positive periods were identified by having a NAOI value greater than 0.25 ; likewise, negative periods were identified by having a NAOI less than -0.25 . The selection of values below -0.25 or above 0.25 was done to ensure that the positive and negative phases were distinctly different from each other by eliminating near-zero NAOI seasons. The values of -0.25 and 0.25 were arbitrarily selected.

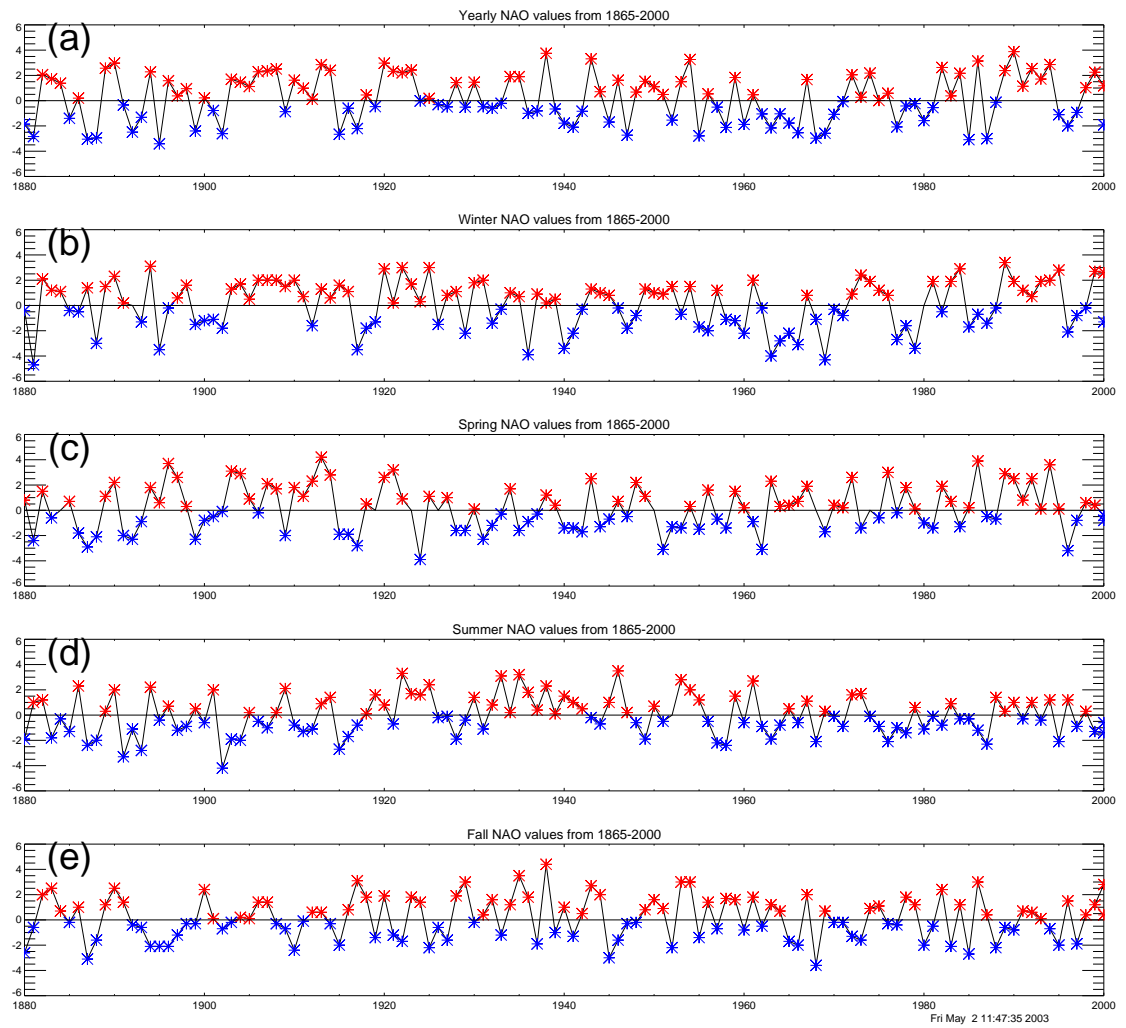


Figure 2.26 The NAOI between Ponta Delgada, Azores and Stykkisholmur/Reykjavik, Iceland for the available time period between the two sites (1865-2000) (a) shows the yearly averaged values, (b) shows the seasonal value for winter; (c), (d), and (e) show spring, summer, and fall respectively.

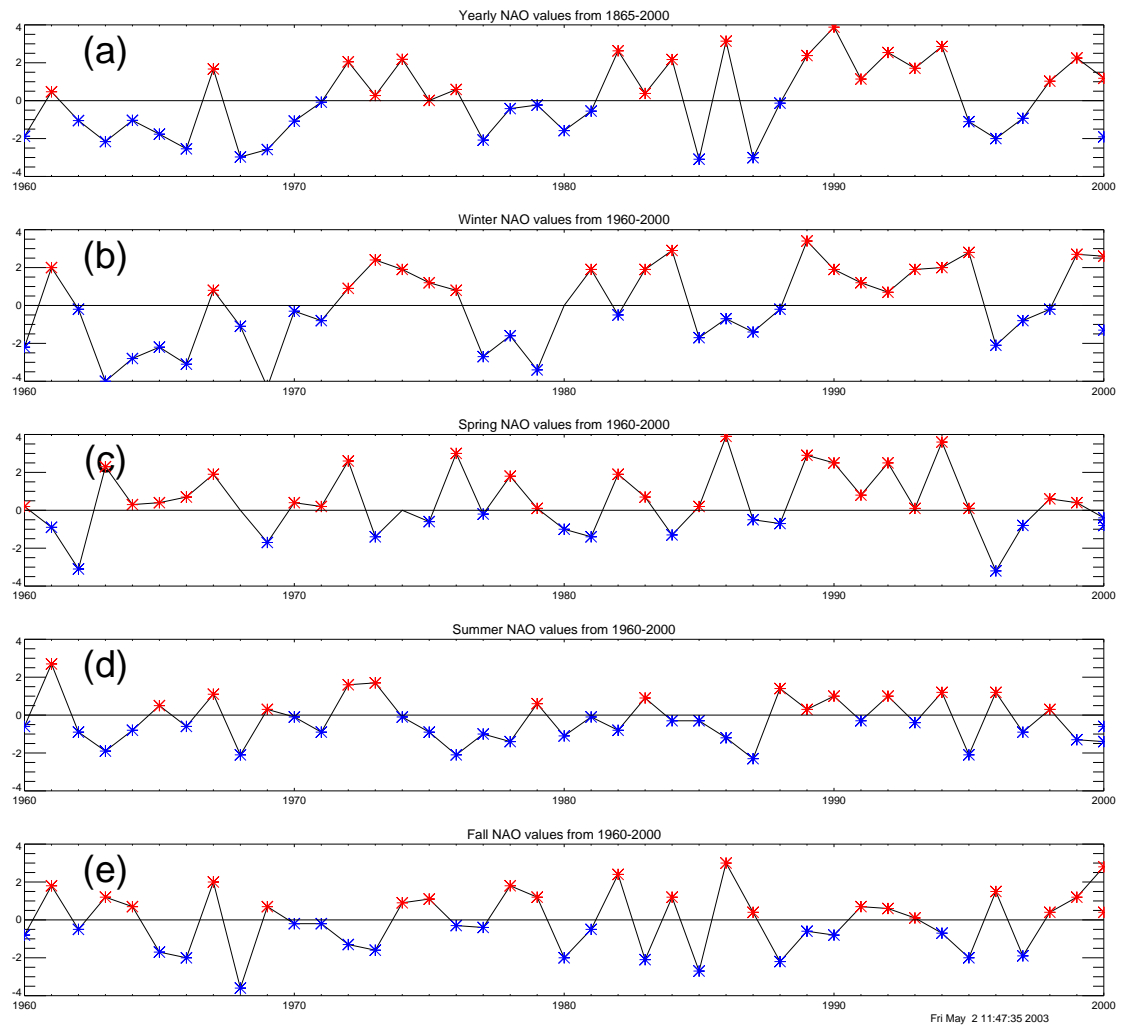


Figure 2.27 The NAOI between Ponta Delgada, Azores and Stykkisholmur/Reykjavik, Iceland for the study period (1960-1999). (a) shows the yearly averaged values, (b) shows the seasonal value for winter; (c), (d), and (e) show spring, summer, and fall respectively.

Chapter 3

Results and Discussion

The previous chapter highlighted the methods used in this study to produce a 40 year (1960–1999) set of 10-day back trajectories to the PICO-NARE site. It also presented the procedures used to analyze those trajectories in order to determine the dominant flow paths for air masses transported to the site as well as source regions that contribute to those air masses. This chapter will present the results of the analysis of the trajectory set and some implications of those results. This chapter will also discuss possible future work using both the results of this analysis and the methods developed.

3.1 Climatology results

Ten-day atmospheric backward-trajectories were calculated in order to help determine source regions and transport pathways to the PICO-NARE site. A cluster analysis was performed in conjunction with an airflow probability analysis (APA) to identify these pathways and regions. Both seasonal and long-term (NAO) climatic variation were taken into account during this process. The results of the analysis are presented in the following section.

3.1.1 Method of descriptions and discussion

In order to facilitate the discussion of the results, the manner in which the results are presented and discussed is outlined in this section. The presentation will begin with the cluster and density plots for the whole period, henceforth referred to as the average period, accompanied with a descriptive analysis. Following the average period, other periods that have clusters do not deviate largely from the average will be presented and discussed. This will begin with seasonal results and will be followed by seasonal results with either positive and negative NAOI value. Following this will be the presentation and discussion that have notable deviation from the average period.

In order to discuss the clusters, several flow patterns that represent flow originating from and passing over important regions will be identified and named. These patterns will be delineated by an examination of the cluster center, as presented in the text, and the membership plots and membership density plots, as presented in the appendix. Viewing the membership plots will aid in the discussion by allowing discussion of flow patterns by name as opposed to using a description each time it is referred to. This will also allow the frequency of the identified flow patterns to be summarized. The flow paths will be identified and named as they occur during the presentation and discussion of the results.

3.1.2 Results for 1960-1999

The results for the average period are shown in Figures 3.1–3.3 and summarized in Table 3.1. This period has two clusters that originate over the Pacific Ocean and travel over the northern U.S./lower Canada. The membership plots for those clusters (Figure A.8 and Figure A.14) show that the trajectories generally travel back to the Pacific and also cover most of North America. The membership density plots (Figure A.9, Figure A.10, Figure A.15, and Figure A.16) show that the trajectories are fairly concentrated around the cluster center, with a lot of time spent over North

America. These clusters have the highest altitudes of all clusters for the entire 10-day period, which means trajectories that are associated with these clusters are likely traveling in the free troposphere. The flow paths portrayed by these clusters have the potential to mix with polluted air while over North America and transport it to the site. However, these clusters are at a fairly high altitude and many trajectories within these clusters travel in the northern regions of North America, which do not have many pollution sources. Therefore these clusters could also transport relatively clean air to the site. These and other clusters with similar sources and flow paths will be identified as Pacific-North American (PaAm) clusters.

There is also a pair of clusters that originates over North America, one in the mid-northern United States, and another in the mid-southern United States. The membership plots (Figure A.5 and Figure A.11) show a large area of coverage. However, the membership density plots (Figure A.6, Figure A.7, Figure A.12, and Figure A.13) show that these clusters are concentrated over the eastern U.S. and western Atlantic. These clusters also have a fairly high altitude, but the membership plots show that there is some concentration at lower altitudes, with a larger degree of variability at higher altitudes. These clusters appear to present the best chance for the transport of polluted air, because they originate over the populated regions of North America and many trajectories in the membership plots exhibit travel at lower altitudes, making it easier for these trajectories to mix with polluted air from North America. However, because of the travel at lower altitudes, while these trajectories are over the ocean, there is potential for the removal of pollution picked up while over North America. Therefore, these clusters could transport polluted or clean air to the site, but the passage over North America makes it likely they will transport some polluted air. Clusters that are likely to be associated with the transport of polluted air from North America and originate over North America, similar to these clusters, will be identified as North American (NoAm) clusters.

The two remaining clusters, according to the cluster centers, originate and spend

time exclusively over the Atlantic Ocean. The membership plots (Figure A.2 and Figure A.17) cover all the North Atlantic and the coasts of the adjacent continents. However, the membership density plots (Figure A.3, Figure A.4, Figure A.18, Figure A.19) primarily show time spent over the ocean, with the outer edges of the density plots showing some time spent over adjoining continents. These clusters generally have a lower altitude, with the membership plots showing some concentration at lower altitudes. It appears that flow in the northern cluster is dominated by the Icelandic Low, while flow in the southern cluster is dominated by the Azores High. The corrected membership density plot for the southern cluster touches the coast of the U.S. and Africa, while the corrected membership density plot for the northern cluster indicates a significant amount of time spent over western Europe and northwestern Africa and it borders Arctic regions over the Atlantic. While both clusters present a strong possibility for transport of clean oceanic air, the northern cluster also presents a good possibility for transport of air from either Africa, Europe, and the Arctic. Clusters that spend time almost exclusively over the ocean, similar to the southern cluster, will be identified as Atlantic Oceanic (AtOc) clusters. Clusters that spend most of their time over the ocean, but also present the possibility of transport from either Europe, Africa, or the Arctic, similar to the northern cluster, will be identified as Eastern Atlantic (EeAt).

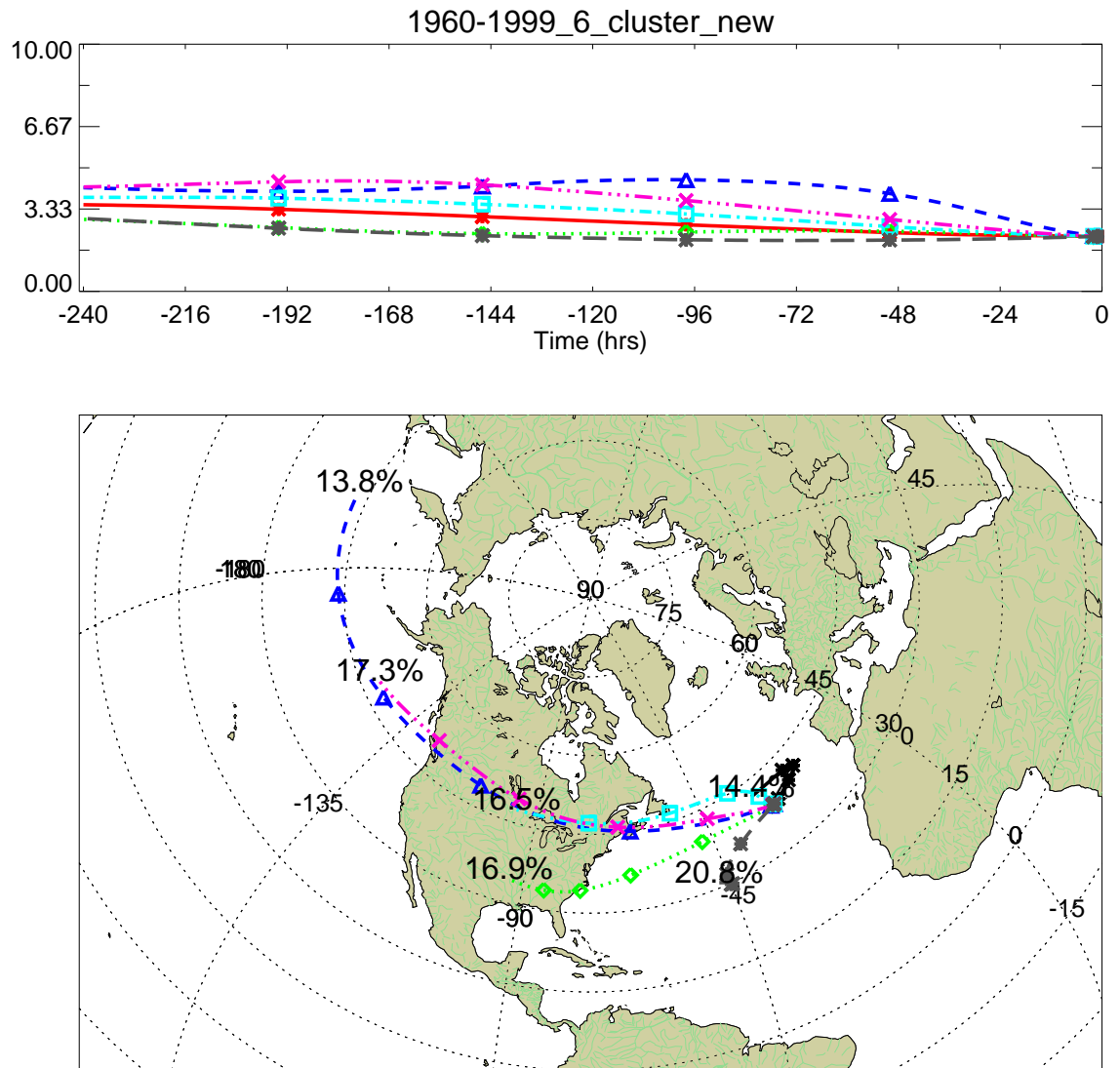


Figure 3.1 The plot of the cluster centers for the extended study period, 1960-1999. Percentages shown are the percentage of trajectories belonging to each cluster, relative to the number of trajectories for this period. Symbols are plotted every 2 days.

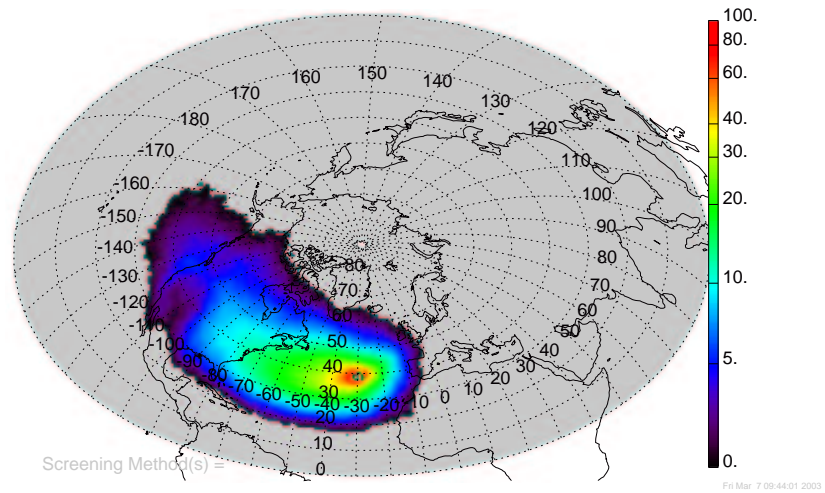


Figure 3.2 The density plot for the average period, 1960-1999.

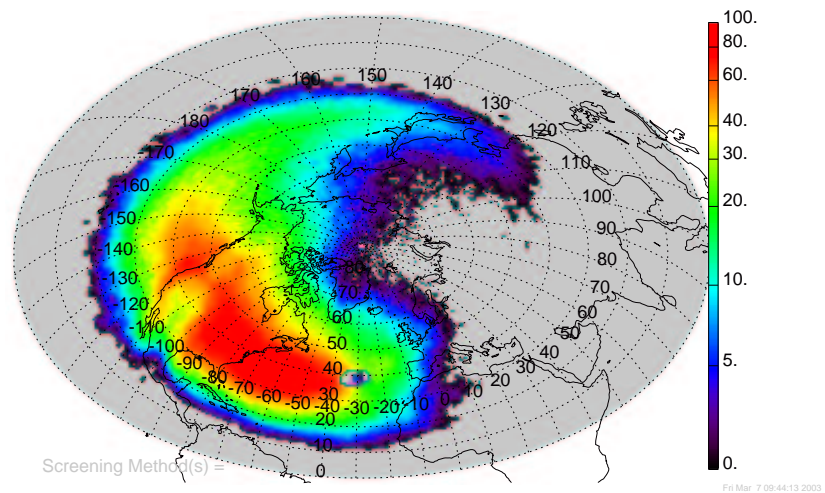


Figure 3.3 The geometrically corrected density plot for the average period, 1960-1999.

Table 3.1 Summary of the clusters for the study period (1960–1999). Information about the cluster center and the trajectories belonging to each cluster (cluster membership).

Variable	AtOc A.17	PaAm A.14	EeAt A.2
Number of Trajectories	11933	9967	8268
Percent total for period: (%)	20.84	17.41	14.44
Cluster origin: (lat,lon)	(30.97,-47.17)	(48.34,-149.39)	(42.17,-31.98)
Cluster travel distance: km	2.32E+03	9.30E+03	2.02E+03
Cluster box: (max lat, max lon)	(38.47,-28.40)	(51.44,-28.40)	(42.27,-19.11)
Cluster box: (min lat, min lon)	(29.57,-47.17)	(38.47,-149.39)	(38.38,-31.98)
Cluster speed: km/hr (m/s)	9.66 (2.68)	38.74 (10.76)	8.41 (2.34)
Mean membership dist.: km	6.56E+03	1.21E+04	7.49E+03
Median membership dist.: km	6.27E+03	1.17E+04	7.21E+03
Standard deviation of dist.: km	2.04E+03	2.83E+03	2.58E+03
Skewness of dist.:	0.864	0.795	0.656
Mean membership speed: km/hr (m/s)	27.32 (7.59)	50.38 (13.99)	31.19 (8.66)
Median membership speed: km/hr (m/s)	26.13 (7.26)	48.90 (13.58)	30.02 (8.34)
Standard deviation of speed: km/hr	8.49 (2.36)	11.81 (3.28)	10.74 (2.98)
Membership box: max(lat,lon)	(89.43,219.20)	(90.00,209.67)	(90.00,415.23)
Membership box: min(lat,lon)	(-2.03,-124.13)	(12.90,-376.47)	(4.25,-150.78)

Variable	NoAm A.11	NoAm A.5	PaAm A.8
Number of Trajectories	9445	9724	7911
Percent total for period: (%)	16.50	16.99	13.82
Cluster origin: (lat,lon)	(47.48,-95.87)	(34.47,-93.30)	(43.02,-197.88)
Cluster travel distance: km	5.60E+03	5.85E+03	1.33E+04
Cluster box: max(lat,lon)	(47.54,-28.40)	(39.48,-28.40)	(49.50,-28.40)
Cluster box: min(lat,lon)	(38.47,-95.87)	(33.73,-93.30)	(38.47,-197.88)
Cluster speed: km/hr (m/s)	23.33 (6.48)	24.36 (6.77)	55.59 (15.44)
Mean membership distance: km	9.63E+03	8.85E+03	1.60E+04
Median membership distance: km	9.20E+03	8.55E+03	1.54E+04
Standard deviation of distance: km	2.82E+03	2.26E+03	3.62E+03
Skewness of distance:	0.819	0.813	0.735
Mean membership speed: km/hr (m/s)	40.13 (11.15)	36.86 (10.24)	66.72 (18.53)
Median membership speed: km/hr (m/s)	38.33 (10.65)	35.64 (9.90)	64.29 (17.86)
Standard deviation of speed: km/hr	11.76 (3.27)	9.42 (2.62)	15.07 (4.19)
Membership box: max(lat,lon)	(90.00,238.65)	(88.68,72.11)	(90.00,215.13)
Membership box: min(lat,lon)	(15.80,-244.91)	(6.07,-168.65)	(5.96,-483.76)

3.1.3 Seasons similar to the average period

This section will present and discuss the seasons that are similar to the average period. The discussions will be fairly brief, as the flow patterns and their implications were discussed in the previous section. The discussion will focus on why these periods are similar to the average period.

3.1.3.1 Results for spring, 1960-1999

The results for spring are presented in Figures 3.4–3.6 and table 3.2. The clusters are very similar to the average period. There are two PaAm clusters, two NoAm clusters, one AtOc and one EeAt cluster. The relative percentage of trajectories belonging to each cluster are also very similar. During the summer, the Icelandic Low tends to break up, which means spring is a transitional period from the stronger Icelandic Low in the winter to the weaker, or non-existent Icelandic Low in the summer. Since spring is a transition from the winter to the summer conditions, it is understandable that it would be similar to the average period, which is the averages of the summer and winter conditions as well as the transitional spring and fall.

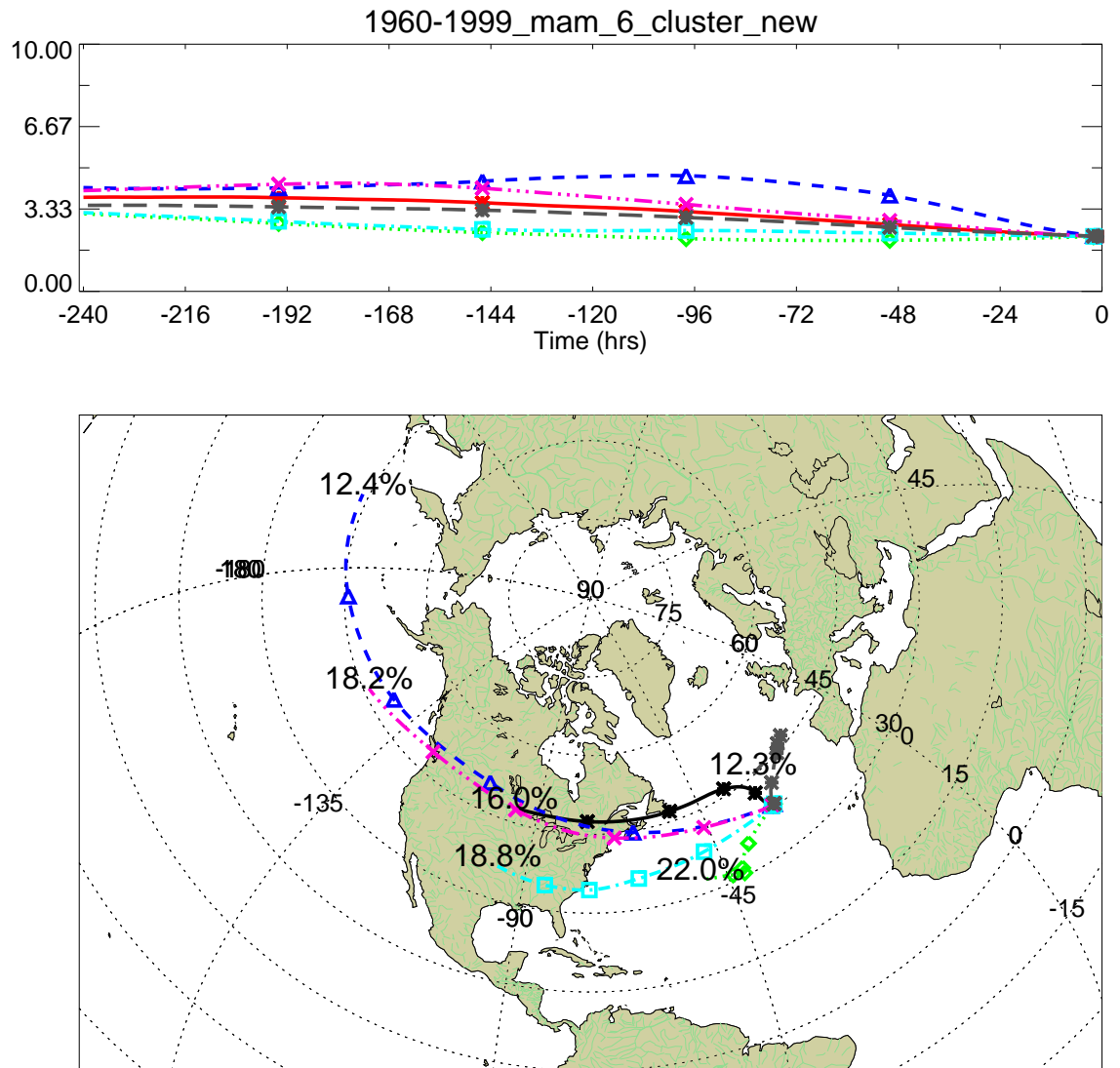


Figure 3.4 The plot of the cluster centers for spring of the extended study period (1960-1999). Percentages shown are the percentage of trajectories belonging to each cluster, relative to the number of trajectories for this period. Symbols are plotted every 2 days.

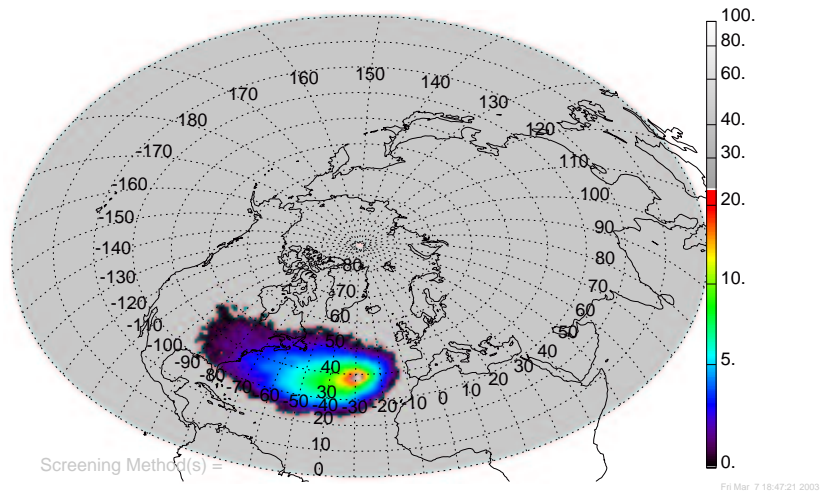


Figure 3.5 The density plot for spring for the extended study period, 1960-1999.

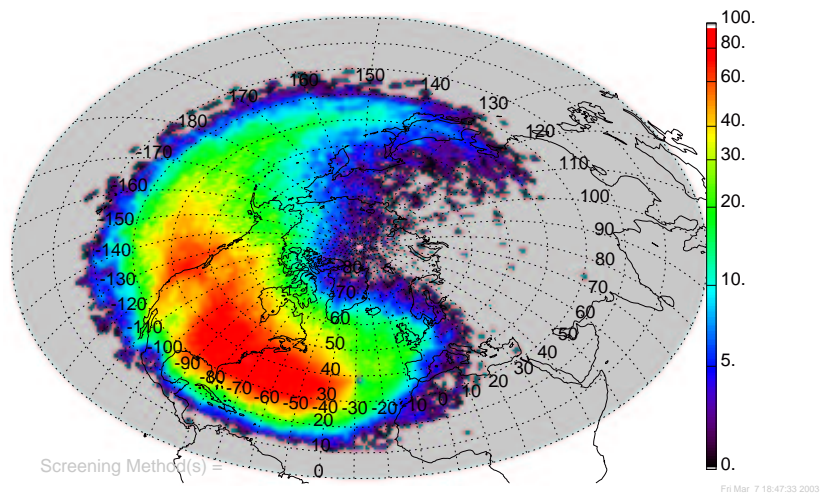


Figure 3.6 The geometrically corrected density plot for spring of the extended study period, 1960-1999.

Table 3.2 Summary of the clusters for spring (1960-1999). Information about the cluster center and the trajectories belonging to each cluster (cluster membership).

Variable	NoAm C.2	AtOc C.5	PaAm C.8
Number of Trajectories	2323	3190	1795
Percent total for period: (%)	16.08	22.09	12.43
Cluster origin: (lat,lon)	(47.65,-96.74)	(32.71,-50.81)	(43.89,-197.48)
Cluster travel distance: km	5.72E+03	2.48E+03	1.31E+04
Cluster box: (max lat, max lon)	(47.81,-28.40)	(38.47,-28.40)	(50.84,-28.40)
Cluster box: (min lat, min lon)	(38.47,-96.74)	(30.62,-50.81)	(38.47,-197.48)
Cluster speed: km/hr (m/s)	23.82 (6.62)	10.35 (2.88)	54.66 (15.18)
Mean membership distance: km	1.05E+04	7.17E+03	1.59E+04
Median membership distance: km	1.00E+04	6.94E+03	1.54E+04
Standard deviation of distance: km	2.72E+03	2.15E+03	3.33E+03
Skewness of distance:	0.775	0.625	0.702
Mean membership speed: km/hr (m/s)	43.54 (12.10)	29.89 (8.30)	66.19 (18.38)
Median membership speed: km/hr (m/s)	41.74 (11.59)	28.93 (8.04)	64.13 (17.82)
Standard deviation of speed: km/hr	11.33 (3.15)	8.95 (2.49)	13.87 (3.85)
Membership box: max(lat,lon)	(90.00,238.65)	(85.20,38.47)	(90.00,178.97)
Membership box: min(lat,lon)	(17.12,-226.12)	(5.06,-113.98)	(7.78,-483.76)

Variable	NoAm C.11	PaAm C.14	EeAt C.17
Number of Trajectories	2723	2634	1779
Percent total for period: (%)	18.85	18.24	12.32
Cluster origin: (lat,lon)	(36.24,-96.55)	(46.03,-149.03)	(45.26,-28.21)
Cluster travel distance: km	6.15E+03	9.54E+03	2.41E+03
Cluster box: max(lat,lon)	(38.47,-28.40)	(49.13,-28.40)	(46.72,-15.91)
Cluster box: min(lat,lon)	(34.38,-96.55)	(38.47,-149.03)	(38.47,-28.48)
Cluster speed: km/hr (m/s)	25.65 (7.12)	39.76 (11.04)	10.02 (2.78)
Mean membership distance: km	9.41E+03	1.25E+04	8.36E+03
Median membership distance: km	9.16E+03	1.22E+04	8.22E+03
Standard deviation of distance: km	2.17E+03	2.73E+03	2.45E+03
Skewness of distance:	0.774	0.814	0.533
Mean membership speed: km/hr (m/s)	39.22 (10.89)	52.25 (14.51)	34.82 (9.67)
Median membership speed: km/hr (m/s)	38.18 (10.60)	50.81 (14.11)	34.24 (9.51)
Standard deviation of speed: km/hr	9.03 (2.51)	11.36 (3.16)	10.22 (2.84)
Membership box: max(lat,lon)	(85.48,28.53)	(89.97,64.58)	(89.31,415.23)
Membership box: min(lat,lon)	(7.32,-168.65)	(12.90,-374.48)	(9.16,-126.63)

3.1.4 NAO seasons similar to the average period

This section will present and discuss the seasons with positive or negative NAOI values that are similar to the average period. The discussions will be fairly brief, as the flow patterns and their implications were discussed in the previous section. The discussion will focus on why these periods are similar to the average period.

3.1.4.1 Results for spring with positive NAOI values, 1960-1999

The results for spring with positive NAOI values are presented in Figures 3.7–3.9 and table 3.3. The clusters are very similar to the average period, as well as the average spring. There are two PaAm clusters, two NoAm clusters, one AtOc and one EeAt cluster. The relative percentage of trajectories belonging to each cluster are also very similar.

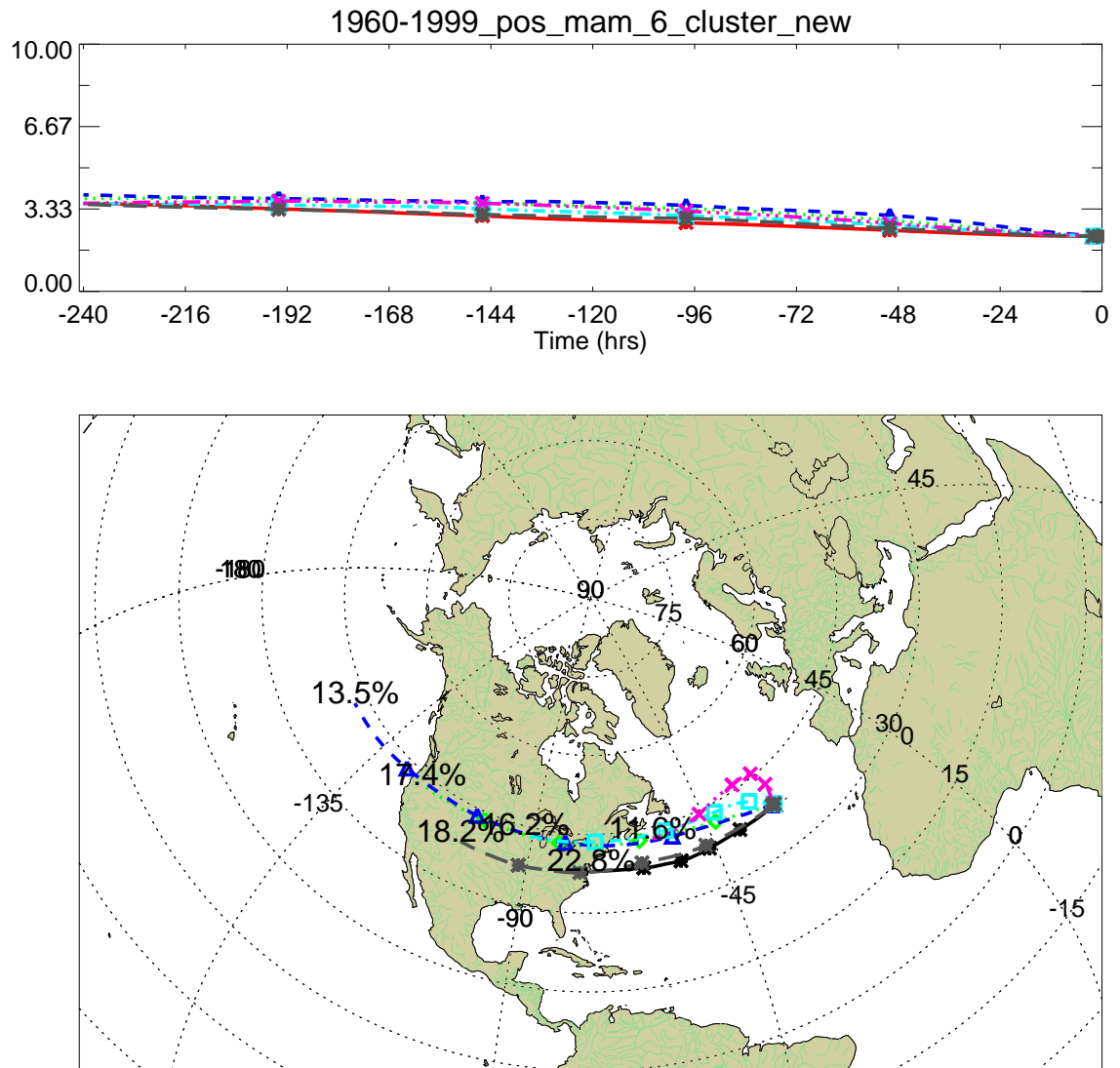


Figure 3.7 The plot of the cluster centers for springs with a positive NAOI, 1960-1999. Percentages shown are the percentage of trajectories belonging to each cluster, relative to the number of trajectories for this period. Symbols are plotted every 2 days.

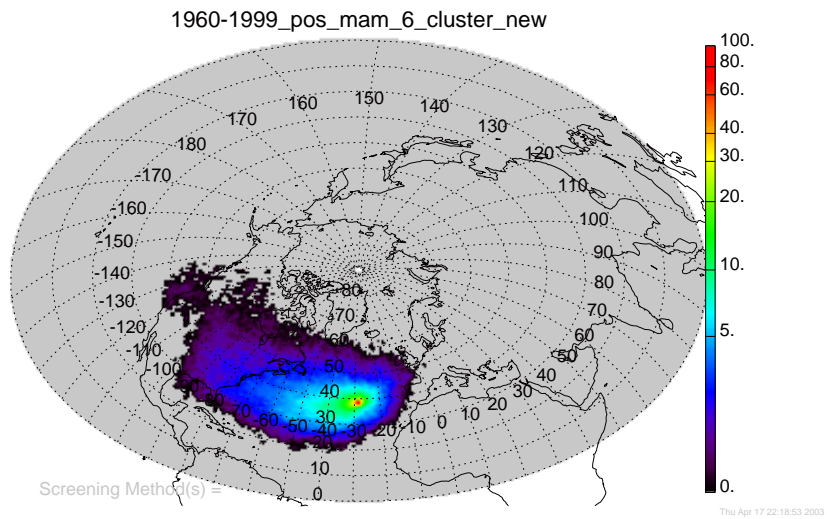


Figure 3.8 The density plot for fall for springs with a positive NAOI, 1960-1999.

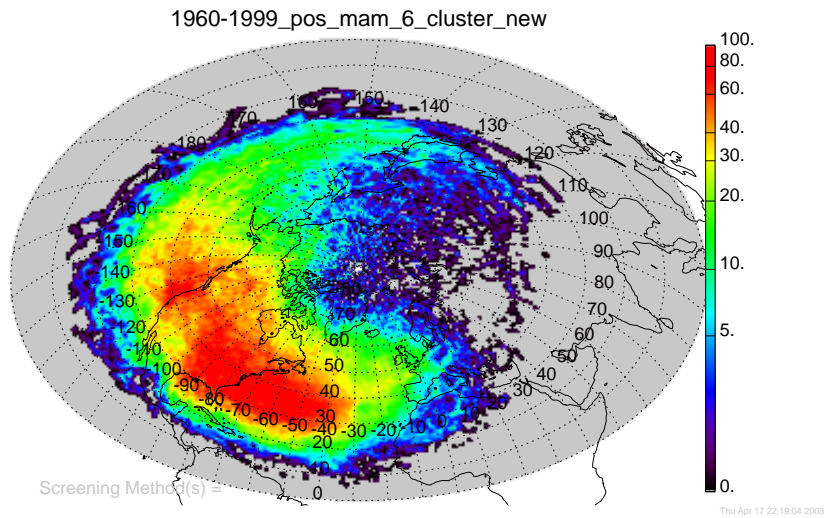


Figure 3.9 The geometrically corrected density plot for springs with a positive NAOI, 1960-1999.

Table 3.3 Summary of the clusters for positive NAO springs (1960-1999). Information about the cluster center and the trajectories belonging to each cluster (cluster membership).

Variable	NoAm C.36	PaAm C.27	EeAt C.33
Number of Trajectories	1192	877	773
Percent total for period: (%)	18.26	13.43	11.84
Cluster origin: (lat,lon)	(53.86,-119.65)	(43.84,-36.87)	(29.74,-75.50)
Cluster travel distance: km	7.06E+03	2.09E+03	4.63E+03
Cluster box: (max lat, max lon)	(54.41,-28.40)	(44.75,-21.84)	(41.28,-28.40)
Cluster box: (min lat, min lon)	(38.47,-119.65)	(38.47,-36.87)	(29.74,-76.20)
Cluster speed: km/hr (m/s)	29.40 (8.17)	8.72 (2.42)	19.31 (5.36)
Mean membership distance: km	1.01E+04	1.42E+04	8.72E+03
Median membership distance: km	9.66E+03	1.40E+04	8.40E+03
Standard deviation of distance: km	3.15E+03	3.66E+03	2.91E+03
Skewness of distance:	1.204	0.399	0.981
Mean membership speed: km/hr (m/s)	42.23 (11.73)	59.25 (16.46)	36.32 (10.09)
Median membership speed: km/hr (m/s)	40.25 (11.18)	58.21 (16.17)	35.01 (9.73)
Standard deviation of speed: km/hr	13.11 (3.64)	15.26 (4.24)	12.11 (3.36)
Membership box: max(lat,lon)	(88.15,325.55)	(89.86,65.89)	(88.74,179.41)
Membership box: min(lat,lon)	(8.02,-366.27)	(14.78,-483.76)	(7.92,-387.82)

Variable	NoAm C.21	AtOc C.30	PaAm C.24
Number of Trajectories	1060	1481	1146
Percent total for period: (%)	16.24	22.68	17.55
Cluster origin: (lat,lon)	(55.07,-168.15)	(45.40,-82.35)	(32.08,-45.63)
Cluster travel distance: km	9.66E+03	4.79E+03	2.63E+03
Cluster box: max(lat,lon)	(60.17,-28.40)	(45.45,-28.40)	(39.46,-28.40)
Cluster box: min(lat,lon)	(38.47,-168.15)	(38.47,-82.35)	(31.09,-48.59)
Cluster speed: km/hr (m/s)	40.25 (11.18)	19.97 (5.55)	10.97 (3.05)
Mean membership distance: km	1.01E+04	7.81E+03	1.18E+04
Median membership distance: km	9.66E+03	7.34E+03	1.15E+04
Standard deviation of distance: km	2.93E+03	2.94E+03	3.17E+03
Skewness of distance:	0.727	1.226	0.707
Mean membership speed: km/hr (m/s)	41.91 (11.64)	32.53 (9.04)	49.31 (13.70)
Median membership speed: km/hr (m/s)	40.26 (11.18)	30.58 (8.50)	47.73 (13.26)
Standard deviation of speed: km/hr	12.22 (3.39)	12.24 (3.40)	13.22 (3.67)
Membership box: max(lat,lon)	(90.00,217.53)	(87.78,94.46)	(90.00,175.48)
Membership box: min(lat,lon)	(8.29,-374.48)	(5.10,-281.57)	(7.64,-391.73)

3.1.4.2 Results for spring with negative NAOI values, 1960-1999

The results for spring with negative NAOI values are presented in Figures 3.10–3.12 and table 3.4. The clusters are very similar to the average period, as well as the average spring and positive NAOI spring. There are two PaAm clusters, two NoAm clusters, one AtOc and one EeAt cluster. The relative percentage of trajectories belonging to most clusters are also similar.

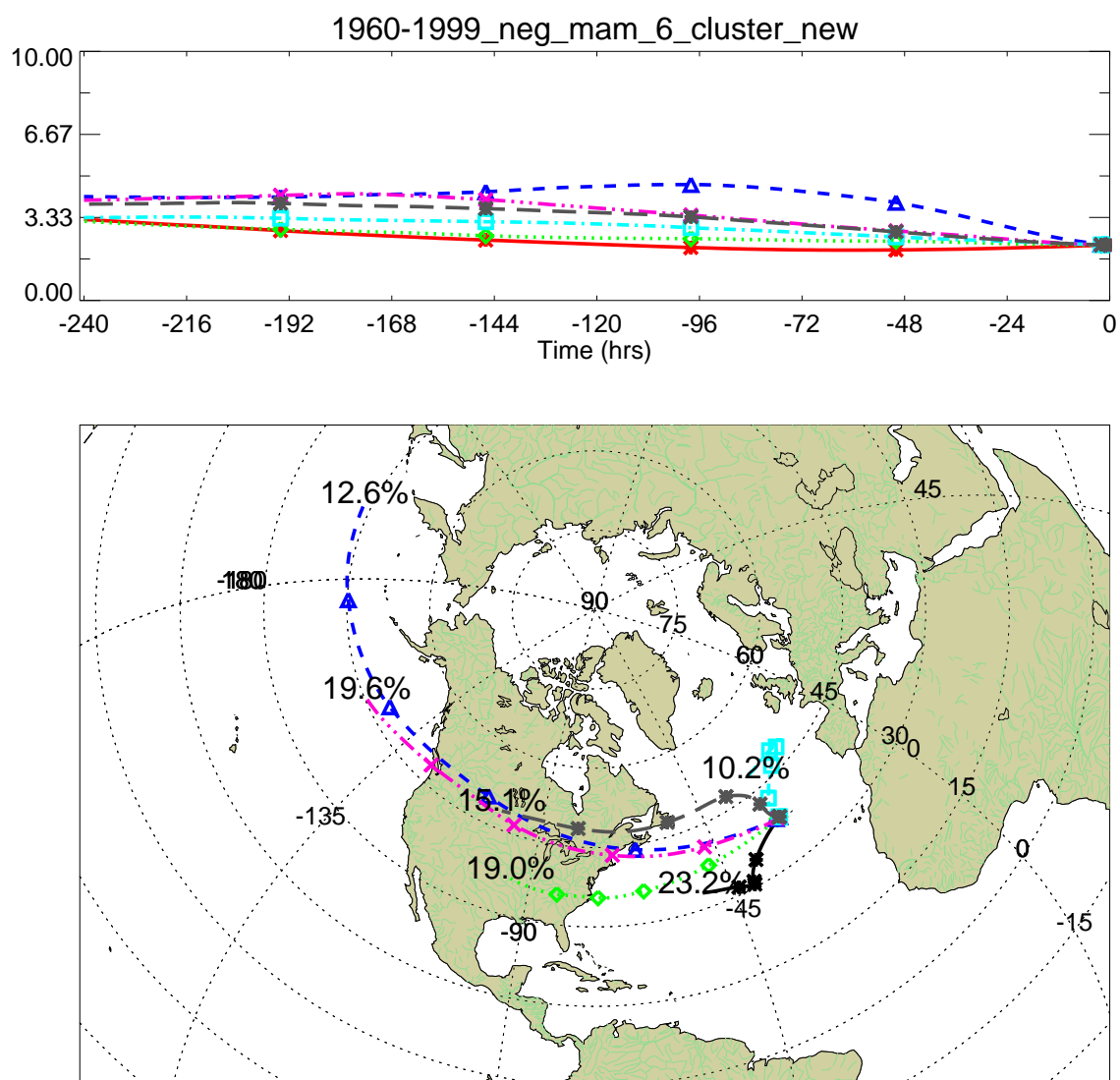


Figure 3.10 The plot of the cluster centers for springs with a negative NAOI, 1960-1999. Percentages shown are the percentage of trajectories belonging to each cluster, relative to the number of trajectories for this period. Symbols are plotted every 2 days.

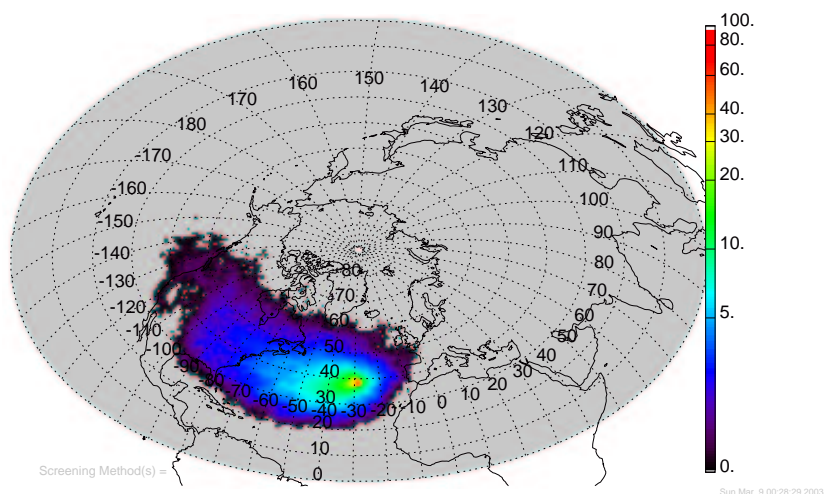


Figure 3.11 The density plot for fall for springs with a negative NAOI, 1960-1999.

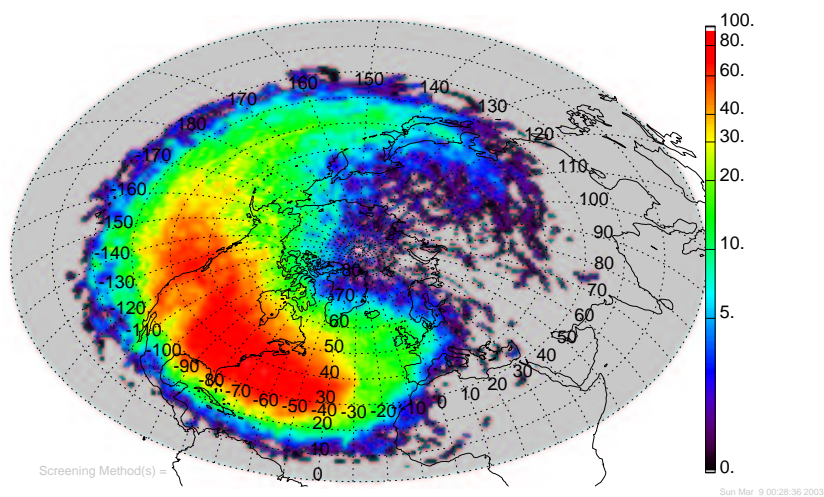


Figure 3.12 The geometrically corrected density plot for springs with a negative NAOI, 1960-1999.

Table 3.4 Summary of the clusters for negative NAO spring (1960-1999). Information about the cluster center and the trajectories belonging to each cluster (cluster membership).

Variable	AtOc C.40	NoAm C.43	PaAm C.46
Number of Trajectories	1510	1243	821
Percent total for period: (%)	23.20	19.10	12.62
Cluster origin: (lat,lon)	(32.99,-51.36)	(36.95,-94.49)	(43.62,-197.70)
Cluster travel distance: km	2.58E+03	5.92E+03	1.32E+04
Cluster box: (max lat, max lon)	(38.47,-28.40)	(38.47,-28.40)	(50.34,-28.40)
Cluster box: (min lat, min lon)	(30.56,-51.36)	(35.46,-94.49)	(38.47,-197.70)
Cluster speed: km/hr (m/s)	10.73 (2.98)	24.68 (6.86)	55.16 (15.32)
Mean membership distance: km	7.12E+03	9.28E+03	1.62E+04
Median membership distance: km	6.89E+03	9.00E+03	1.56E+04
Standard deviation of distance: km	2.18E+03	2.18E+03	3.41E+03
Skewness of distance:	0.726	0.847	0.828
Mean membership speed: km/hr (m/s)	29.67 (8.24)	38.65 (10.74)	67.40 (18.72)
Median membership speed: km/hr (m/s)	28.71 (7.97)	37.50 (10.42)	65.13 (18.09)
Standard deviation of speed: km/hr	9.07 (2.52)	9.10 (2.53)	14.19 (3.94)
Membership box: max(lat,lon)	(85.20,38.47)	(85.48,28.53)	(89.99,178.97)
Membership box: min(lat,lon)	(5.06,-102.50)	(6.85,-155.22)	(15.24,-432.84)

Variable	EeAt C.49	PaAm C.52	NoAm C.55
Number of Trajectories	670	1279	985
Percent total for period: (%)	10.30	19.65	15.14
Cluster origin: (lat,lon)	(48.03,-29.56)	(45.34,-150.06)	(48.38,-101.86)
Cluster travel distance: km	2.48E+03	9.73E+03	6.12E+03
Cluster box: max(lat,lon)	(48.64,-17.49)	(48.56,-28.40)	(48.97,-28.40)
Cluster box: min(lat,lon)	(38.47,-29.56)	(38.47,-150.06)	(38.47,-101.86)
Cluster speed: km/hr (m/s)	10.34 (2.87)	40.54 (11.26)	25.51 (7.09)
Mean membership distance: km	8.62E+03	1.26E+04	1.06E+04
Median membership distance: km	8.55E+03	1.22E+04	1.01E+04
Standard deviation of distance: km	2.35E+03	2.70E+03	2.72E+03
Skewness of distance:	0.368	0.795	0.838
Mean membership speed: km/hr (m/s)	35.91 (9.97)	52.38 (14.55)	44.34 (12.32)
Median membership speed: km/hr (m/s)	35.61 (9.89)	50.99 (14.16)	42.23 (11.73)
Standard deviation of speed: km/hr	9.79 (2.72)	11.24 (3.12)	11.35 (3.15)
Membership box: max(lat,lon)	(88.52,204.47)	(89.30,64.58)	(89.54,50.78)
Membership box: min(lat,lon)	(14.73,-101.34)	(14.00,-249.86)	(18.42,-226.12)

3.1.4.3 Results for fall with negative NAOI values, 1960-1999

The results for fall with negative NAOI values are presented in Figures 3.13–3.15 and Table 3.5. The clusters are very similar to the average period, as well as the average fall. There are two PaAm clusters, two NoAm clusters, one AtOc and one EeAt cluster. The relative percentage of trajectories belonging to most clusters are also similar. Although negative NAO falls are not different from the average period, the average fall and positive NAO fall are. These are discussed in sections 3.1.5.3 and 3.1.6.5.

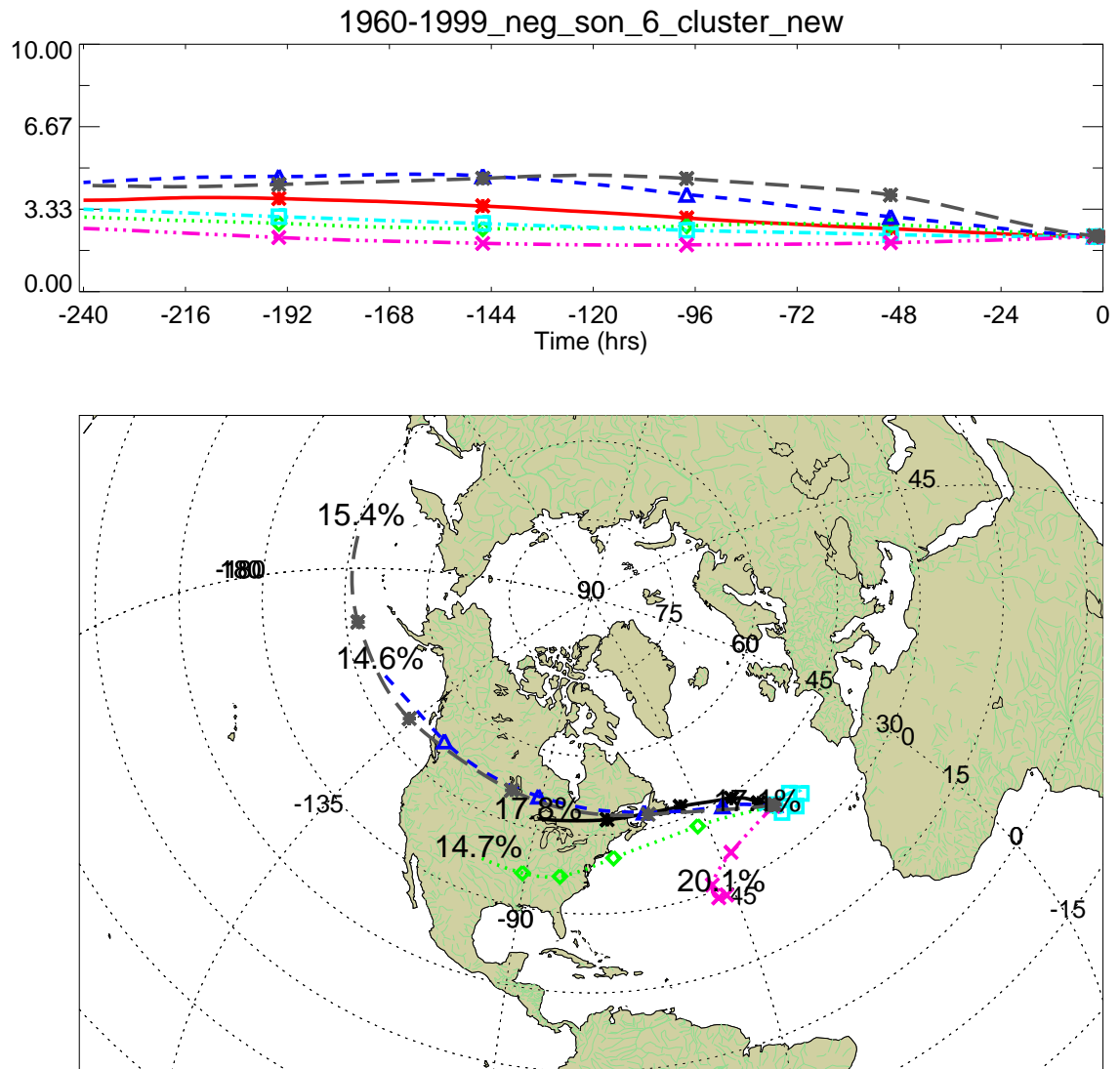


Figure 3.13 The plot of the cluster centers for falls with a negative NAOI, 1960-1999. Percentages shown are the percentage of trajectories belonging to each cluster, relative to the number of trajectories for this period. Symbols are plotted every 2 days.

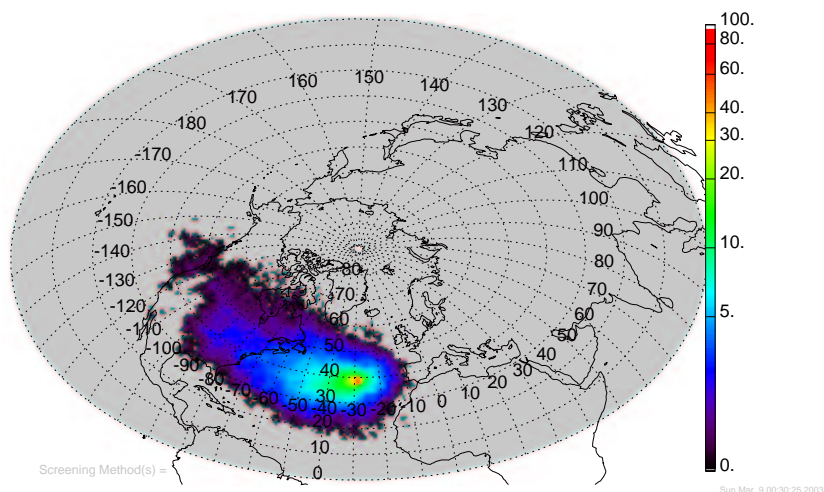


Figure 3.14 The density plot for fall for falls with a negative NAOI, 1960-1999.

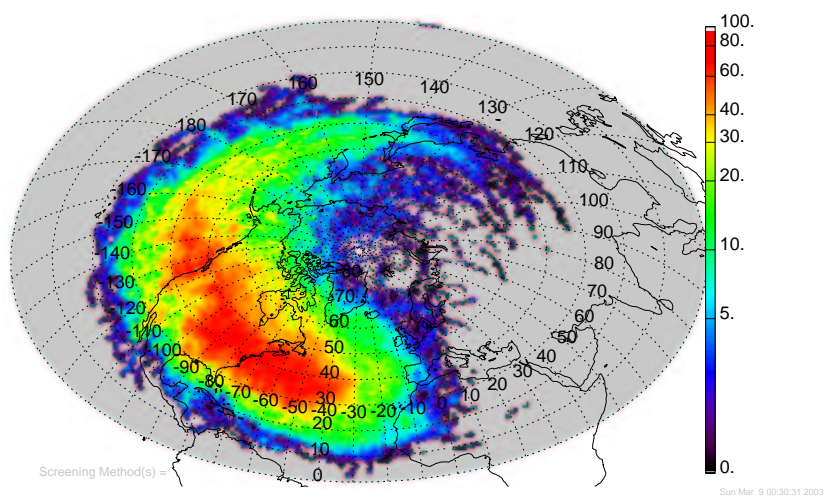


Figure 3.15 The geometrically corrected density plot for falls with a negative NAOI, 1960-1999.

Table 3.5 Summary of the clusters for negative NAO fall (1960-1999). Information about the cluster center and the trajectories belonging to each cluster (cluster membership).

Variable	NoAm E.40	NoAm E.43	PaAm E.46
Number of Trajectories	765	633	630
Percent total for period: (%)	17.84	14.76	14.69
Cluster origin: (lat,lon)	(47.05,-89.75)	(36.76,-101.41)	(49.17,-152.50)
Cluster travel distance: km	5.04E+03	6.42E+03	9.32E+03
Cluster box: (max lat, max lon)	(47.95,-28.40)	(42.65,-28.40)	(51.96,-28.40)
Cluster box: (min lat, min lon)	(38.47,-89.75)	(36.31,-101.41)	(38.47,-152.50)
Cluster speed: km/hr (m/s)	21.01 (5.84)	26.76 (7.43)	38.81 (10.78)
Mean membership distance: km	9.24E+03	9.28E+03	1.22E+04
Median membership distance: km	9.18E+03	9.07E+03	1.19E+04
Standard deviation of distance: km	2.18E+03	2.09E+03	2.72E+03
Skewness of distance:	0.312	0.354	0.652
Mean membership speed: km/hr (m/s)	38.49 (10.69)	38.66 (10.74)	50.71 (14.09)
Median membership speed: km/hr (m/s)	38.26 (10.63)	37.80 (10.50)	49.44 (13.73)
Standard deviation of speed: km/hr	9.09 (2.52)	8.69 (2.41)	11.32 (3.14)
Membership box: max(lat,lon)	(87.05,43.16)	(80.73,17.87)	(90.00,103.70)
Membership box: min(lat,lon)	(17.05,-212.45)	(8.77,-150.10)	(20.59,-310.65)

Variable	EeAt E.49	AtOc E.52	PaAm E.55
Number of Trajectories	736	863	661
Percent total for period: (%)	17.16	20.13	15.42
Cluster origin: (lat,lon)	(39.45,-32.68)	(29.54,-47.47)	(45.68,-190.41)
Cluster travel distance: km	1.90E+03	2.74E+03	1.23E+04
Cluster box: max(lat,lon)	(39.45,-21.97)	(38.47,-28.40)	(51.25,-28.40)
Cluster box: min(lat,lon)	(35.66,-32.68)	(28.30,-49.21)	(38.47,-190.41)
Cluster speed: km/hr (m/s)	7.90 (2.19)	11.40 (3.17)	51.33 (14.26)
Mean membership distance: km	7.00E+03	6.36E+03	1.48E+04
Median membership distance: km	6.74E+03	6.07E+03	1.43E+04
Standard deviation of distance: km	2.32E+03	1.92E+03	3.21E+03
Skewness of distance:	0.766	1.071	0.742
Mean membership speed: km/hr (m/s)	29.16 (8.10)	26.51 (7.36)	61.63 (17.12)
Median membership speed: km/hr (m/s)	28.10 (7.81)	25.30 (7.03)	59.76 (16.60)
Standard deviation of speed: km/hr	9.68 (2.69)	7.98 (2.22)	13.38 (3.72)
Membership box: max(lat,lon)	(87.07,146.17)	(72.01,26.31)	(89.88,18.31)
Membership box: min(lat,lon)	(5.70,-119.79)	(1.61,-99.07)	(18.26,-443.56)

3.1.5 Seasons dissimilar to the average period

This section will present and discuss the seasons that are dissimilar to the average period. The discussions will be more detailed than the previous section, as new flow paths will need to be identified and the causes of the differences discussed.

3.1.5.1 Results for winter, 1960-1999

The results for the average winter are shown in Figures 3.16–3.18 and Table 3.6. There are some features similar to the average period. There are still two PaAm clusters, as well as two NoAm clusters. However, the location of the NoAm clusters are different from the average period. One originates much farther east and the other originates farther south west than the corresponding clusters in the average period. There is an additional cluster that originates in over North America, but it is farther north than the typical NoAm clusters. This motion is likely the result of air being transported at higher latitudes encountering a trough and being forced south, around this low pressure system. A similar flow pattern for Arctic outflow was noted by *Honrath et al.* [1996]. The membership plots (Figure B.8–B.10) show that this cluster spends far less time over populated regions of North America, although some members pass over the eastern coast of the U.S., which is densely populated and has many pollution sources. It is not clear whether this cluster presents more opportunity to transport clean air, because it spends more time in remote or marine regions than the typical NoAm cluster, or if it presents equal opportunity to transport pollution, like the typical NoAm cluster because it does pass over the eastern coast of the U.S. Because this cluster does not clearly present the transport of polluted air from North America, it should not be included in the NoAm classification. Instead, it will be classified as Northern North American (NoNA) cluster.

The final cluster appears to be primarily an Atlantic cluster. However, an inspection of density plots (Figure B.14–B.16) shows that the trajectories in this cluster

spend time south and east of the site, including time over Europe and Africa. Because this cluster does not include the northern influence of the EeAt cluster, it will be classified separately. Instead it will be classified as South-Eastern Atlantic (SEAt) cluster. In general, the winter period shows enhanced westerly flow. The dominant westerly flow of the mid-latitudes is enhanced during the winter months.

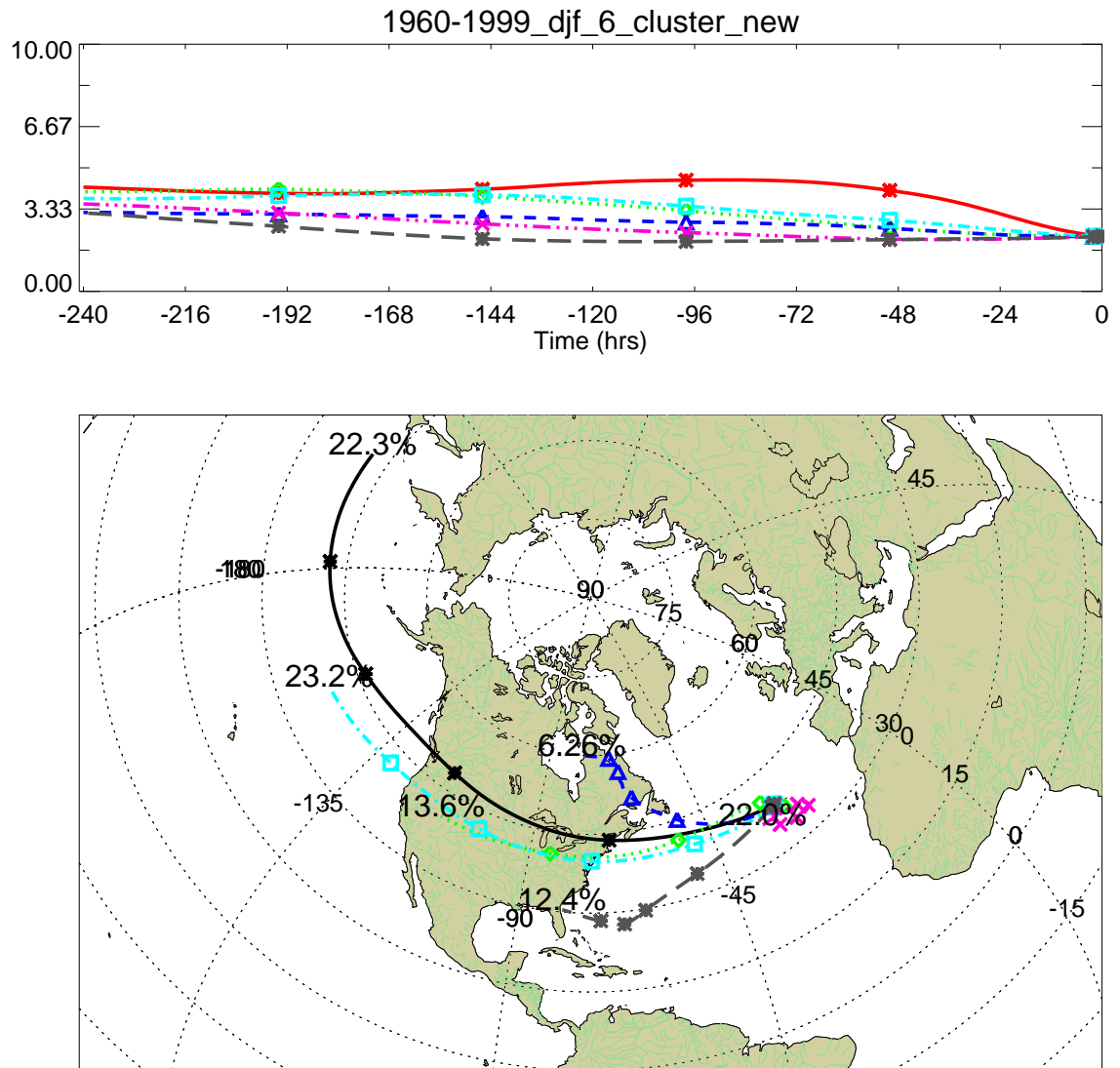


Figure 3.16 The plot of the cluster centers for winter of the extended study period (1960-1999). Percentages shown are the percentage of trajectories belonging to each cluster, relative to the number of trajectories for this period. Symbols are plotted every 2 days.

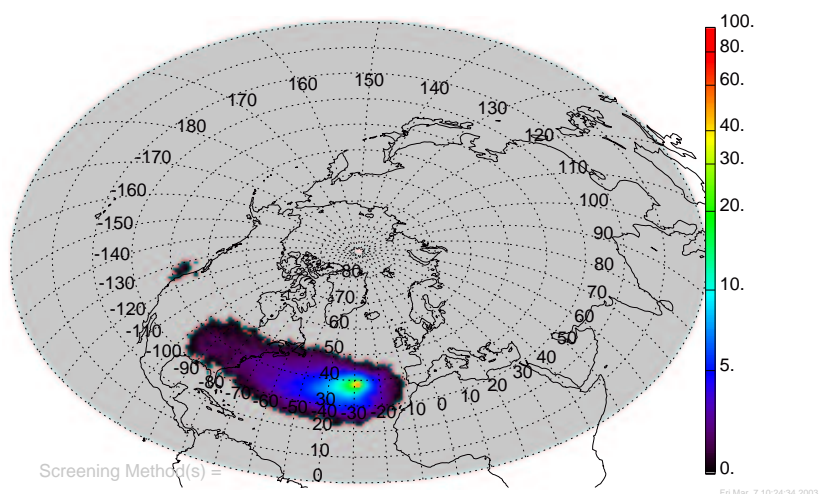


Figure 3.17 The density plot for winter for the extended study period, 1960-1999.

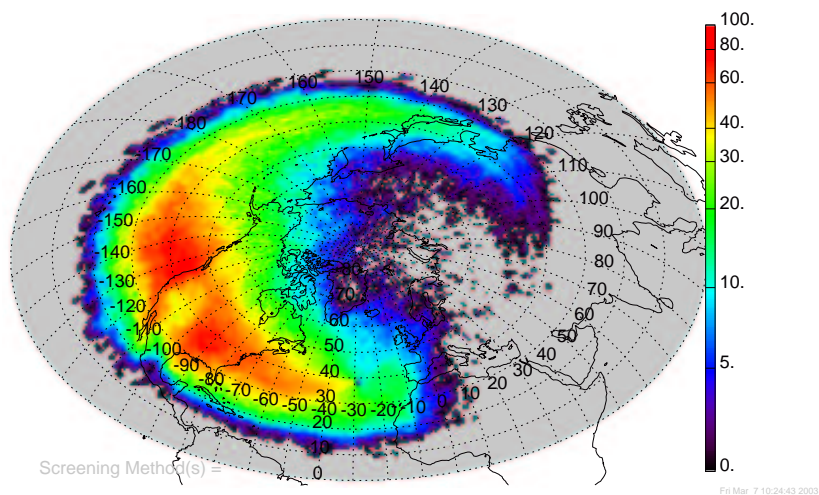


Figure 3.18 The geometrically corrected density plot for winter of the extended study period, 1960-1999.

Table 3.6 Summary of the clusters for winter (1960-1999). Information about the cluster center and the trajectories belonging to each cluster (cluster membership).

Variable	PaAm B.2	NoAm B.5	NoNA B.8
Number of Trajectories	3090	1891	868
Percent total for period: (%)	22.31	13.65	6.27
Cluster origin: (lat,lon)	(41.66,-206.29)	(40.31,-113.30)	(59.98,-78.22)
Cluster travel distance: km	1.43E+04	7.87E+03	4.76E+03
Cluster box: (max lat, max lon)	(49.00,-28.40)	(41.46,-26.63)	(59.98,-28.40)
Cluster box: (min lat, min lon)	(38.47,-206.29)	(36.68,-113.30)	(38.47,-78.22)
Cluster speed: km/hr (m/s)	59.56 (16.55)	32.77 (9.10)	19.85 (5.51)
Mean membership distance: km	1.73E+04	1.20E+04	1.11E+04
Median membership distance: km	1.67E+04	1.18E+04	1.05E+04
Standard deviation of distance: km	3.74E+03	2.93E+03	2.80E+03
Skewness of distance:	0.669	0.553	1.071
Mean membership speed: km/hr (m/s)	71.91 (19.98)	50.20 (13.94)	46.22 (12.84)
Median membership speed: km/hr (m/s)	69.60 (19.33)	49.06 (13.63)	43.89 (12.19)
Standard deviation of speed: km/hr	15.58 (4.33)	12.19 (3.39)	11.67 (3.24)
Membership box: max(lat,lon)	(90.00,99.84)	(87.76,82.72)	(90.00,219.20)
Membership box: min(lat,lon)	(5.96,-403.79)	(14.90,-260.93)	(22.11,-260.08)

Variable	PaAm B.11	SEAt B.14	NoAm B.17
Number of Trajectories	3218	3055	1727
Percent total for period: (%)	23.24	22.06	12.47
Cluster origin: (lat,lon)	(39.43,-154.05)	(36.63,-33.26)	(30.45,-80.97)
Cluster travel distance: km	1.07E+04	2.30E+03	5.17E+03
Cluster box: max(lat,lon)	(42.31,-28.40)	(38.47,-22.35)	(38.47,-28.40)
Cluster box: min(lat,lon)	(38.34,-154.05)	(33.63,-33.26)	(27.58,-80.97)
Cluster speed: km/hr (m/s)	44.48 (12.36)	9.58 (2.66)	21.52 (5.98)
Mean membership distance: km	1.34E+04	7.97E+03	9.12E+03
Median membership distance: km	1.31E+04	7.57E+03	8.93E+03
Standard deviation of distance: km	2.86E+03	2.53E+03	2.17E+03
Skewness of distance:	0.758	0.610	0.599
Mean membership speed: km/hr (m/s)	55.99 (15.55)	33.20 (9.22)	38.01 (10.56)
Median membership speed: km/hr (m/s)	54.65 (15.18)	31.56 (8.77)	37.22 (10.34)
Standard deviation of speed: km/hr	11.92 (3.31)	10.56 (2.93)	9.02 (2.51)
Membership box: max(lat,lon)	(89.86,38.04)	(89.75,262.77)	(81.19,31.34)
Membership box: min(lat,lon)	(6.07,-296.43)	(4.02,-113.55)	(4.26,-134.41)

3.1.5.2 Results for summer, 1960-1999

The results for the average summer are shown in Figures 3.19–3.21 and Table 3.7. The summer period is the least similar to the average period. There is one PaAm cluster which originates in the Eastern edge of the Pacific Ocean. An examination of the membership plots (Figures D.11–D.13) shows that this cluster does not spend as much time over the Pacific as other PaAm clusters, but is sufficient to remain under this classification. The next two clusters belong to the NoAm classification. There appear to be two AtOc clusters, which is correct for the slowest cluster, but an inspection of the other membership plots (Figure D.5–D.7) for the other cluster shows that this cluster should not be classified with AtOc because there is some time spent over North America. However, there is less likelihood that this cluster can transport polluted air from North America than typical NoAm clusters. Because this cluster does not clearly present the transport of clean air from the Atlantic, it should not be included in the AtOc classification. Instead, it will be classified as South-Western Atlantic (SWAt). The final cluster appears to belong to the EeAt, and an inspection of its membership plots confirm this. In general, the summer flow patterns show much weaker westerly flow, which is not surprising as the dominant westerly flow of the mid-latitudes is decreased in the summer months.

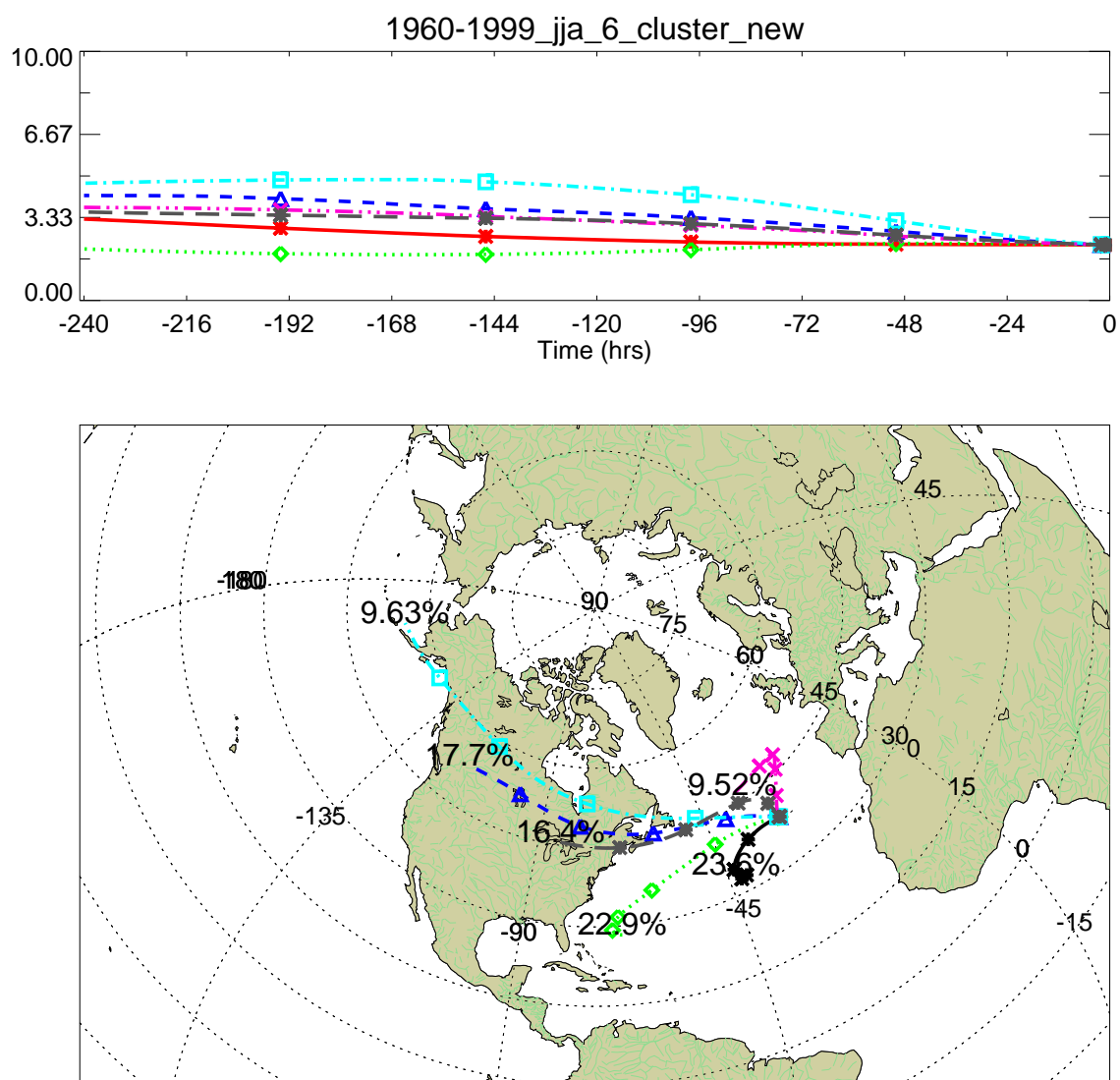


Figure 3.19 The plot of the cluster centers for summer of the extended study period (1960-1999). Percentages shown are the percentage of trajectories belonging to each cluster, relative to the number of trajectories for this period. Symbols are plotted every 2 days.

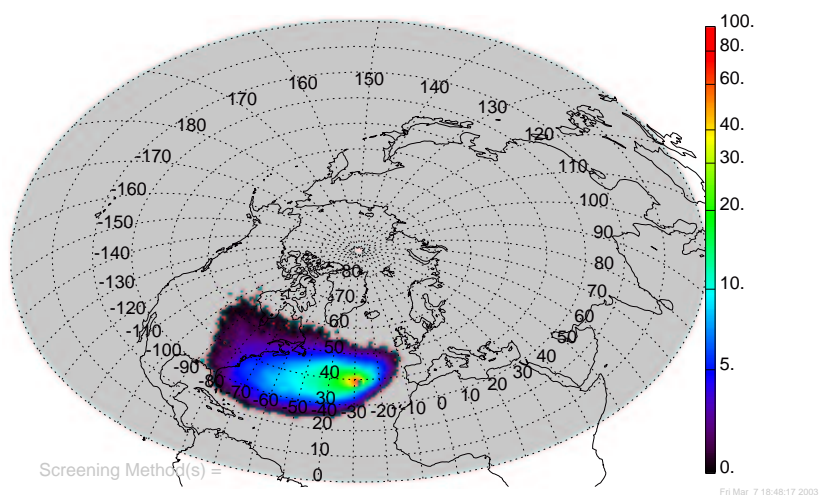


Figure 3.20 The density plot for summer for the extended study period, 1960-1999.

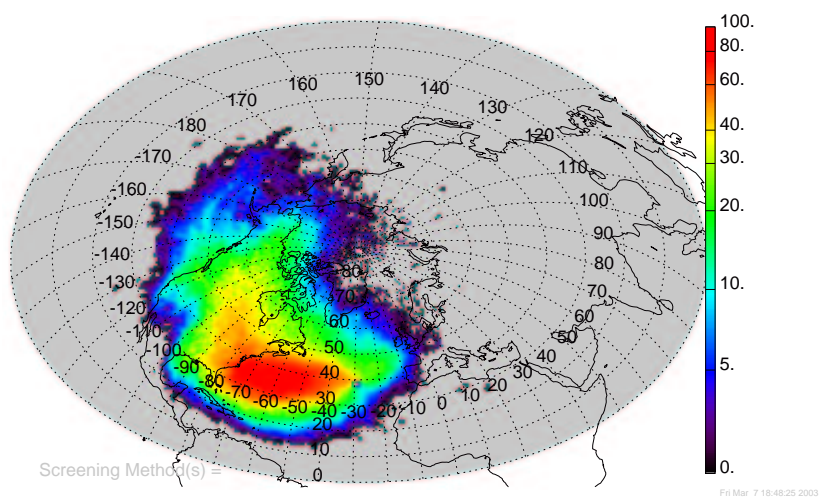


Figure 3.21 The geometrically corrected density plot for summer of the extended study period, 1960-1999.

Table 3.7 Summary of the clusters for summer (1960-1999). Information about the cluster center and the trajectories belonging to each cluster (cluster membership).

Variable	AtOc D.2	SWAt D.5	NoAm D.8
Number of Trajectories	3465	3360	2604
Percent total for period: (%)	23.67	22.95	17.79
Cluster origin: (lat,lon)	(33.82,-43.16)	(28.31,-69.42)	(52.80,-116.56)
Cluster travel distance: km	2.10E+03	4.45E+03	6.84E+03
Cluster box: (max lat, max lon)	(39.01,-28.40)	(40.78,-28.40)	(53.25,-28.40)
Cluster box: (min lat, min lon)	(32.47,-43.39)	(28.31,-71.47)	(38.47,-116.56)
Cluster speed: km/hr (m/s)	8.76 (2.43)	18.54 (5.15)	28.50 (7.92)
Mean membership distance: km	5.58E+03	6.84E+03	8.98E+03
Median membership distance: km	5.39E+03	6.70E+03	8.71E+03
Standard deviation of distance: km	1.62E+03	1.36E+03	1.69E+03
Skewness of distance:	0.660	0.571	0.713
Mean membership speed: km/hr (m/s)	23.25 (6.46)	28.52 (7.92)	37.40 (10.39)
Median membership speed: km/hr (m/s)	22.44 (6.23)	27.92 (7.75)	36.27 (10.08)
Standard deviation of speed: km/hr	6.75 (1.87)	5.66 (1.57)	7.04 (1.96)
Membership box: max(lat,lon)	(89.43,182.23)	(83.30,20.34)	(90.00,176.38)
Membership box: min(lat,lon)	(4.59,-96.62)	(4.97,-100.69)	(18.28,-244.91)
Variable	PaAm D.11	EeAt D.14	NoAm D.17
Number of Trajectories	1411	1394	2406
Percent total for period: (%)	9.64	9.52	16.43
Cluster origin: (lat,lon)	(55.46,-166.60)	(46.98,-35.11)	(45.59,-84.11)
Cluster travel distance: km	9.52E+03	2.44E+03	4.89E+03
Cluster box: max(lat,lon)	(60.00,-28.40)	(48.04,-19.65)	(45.59,-28.40)
Cluster box: min(lat,lon)	(38.47,-166.60)	(38.47,-35.11)	(38.47,-84.11)
Cluster speed: km/hr (m/s)	39.68 (11.02)	10.17 (2.82)	20.37 (5.66)
Mean membership distance: km	1.13E+04	6.67E+03	7.52E+03
Median membership distance: km	1.11E+04	6.58E+03	7.26E+03
Standard deviation of distance: km	2.10E+03	2.13E+03	1.79E+03
Skewness of distance:	0.516	0.660	0.775
Mean membership speed: km/hr (m/s)	47.04 (13.07)	27.78 (7.72)	31.33 (8.70)
Median membership speed: km/hr (m/s)	46.34 (12.87)	27.41 (7.61)	30.23 (8.40)
Standard deviation of speed: km/hr	8.76 (2.43)	8.87 (2.46)	7.48 (2.08)
Membership box: max(lat,lon)	(90.00,185.01)	(89.93,275.14)	(89.65,38.75)
Membership box: min(lat,lon)	(23.19,-403.31)	(14.34,-112.54)	(16.85,-227.93)

3.1.5.3 Results for fall, 1960-1999

The results for fall are presented in Figures 3.22–3.24 and Table 3.8. The clusters are very similar to the average period, as well as the spring period. There are two PaAm clusters, two NoAm clusters, two clusters that appear to belong to the AtOc classification. However, an inspection of the membership plots for the one centered over the Azores (Figure E.2–Figure E.4) shows that this cluster spends a significant amount of time over Africa and southern Europe. For this reason it should not be classified as AtOc. Instead, it will be classified as SEAt. This cluster presents the possibility to transport polluted air from Africa or Europe. During the summer, the Azores High typically covers a larger area and its center of action is located the Azores than the other. This means it has the opportunity pull south-easterly air towards the pressure center. This action can be seen to some degree in the membership plot.

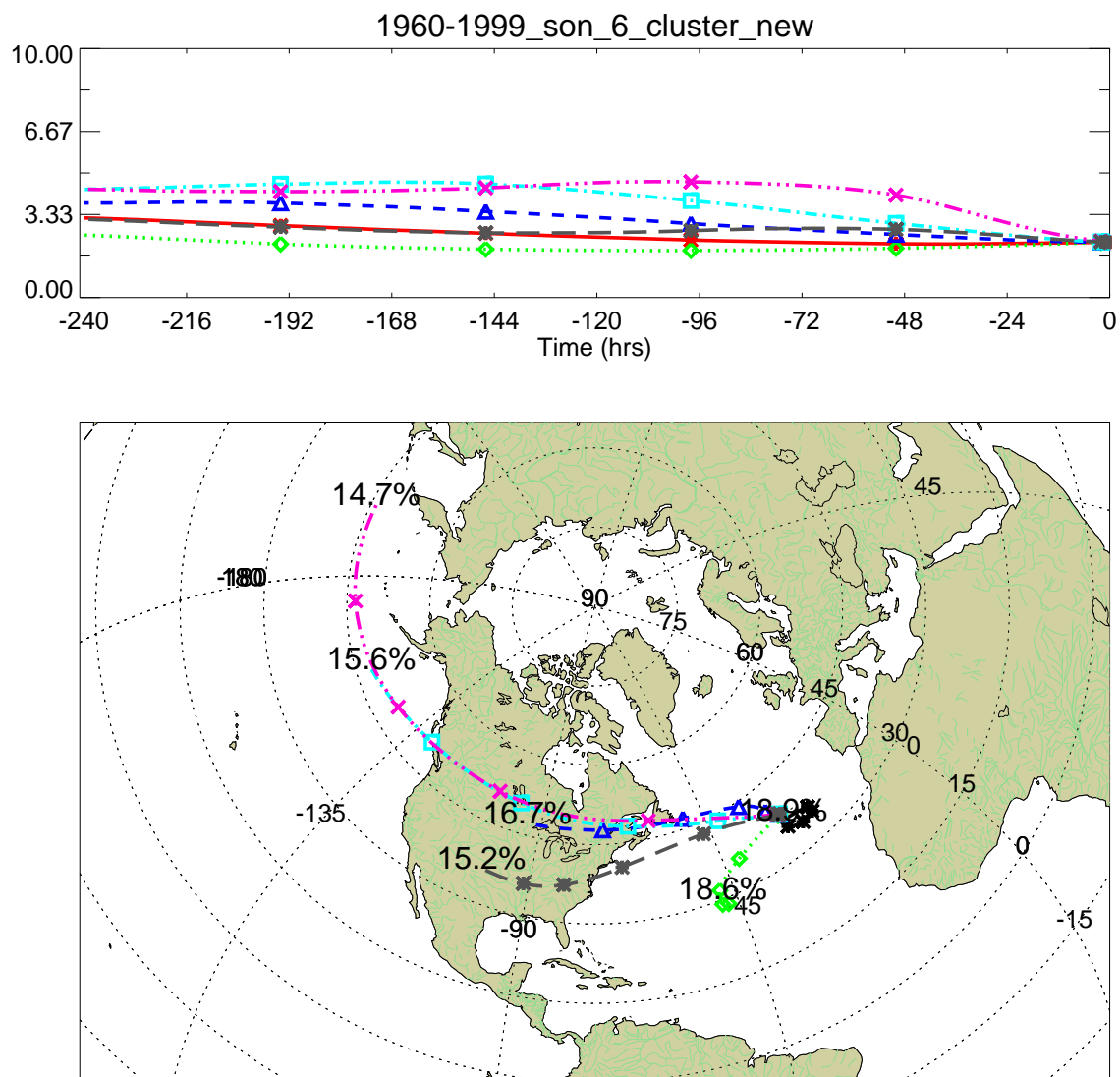


Figure 3.22 The plot of the cluster centers for fall of the extended study period (1960-1999). Percentages shown are the percentage of trajectories belonging to each cluster, relative to the number of trajectories for this period. Symbols are plotted every 2 days.

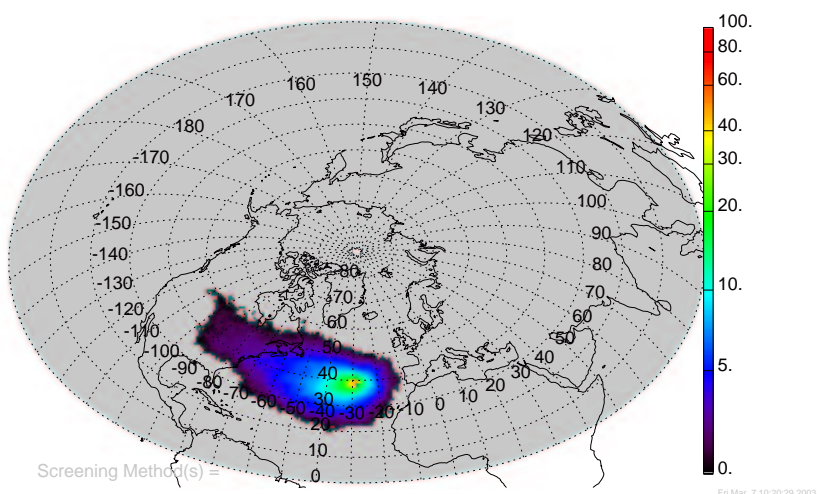


Figure 3.23 The density plot for fall for the extended study period, 1960-1999.

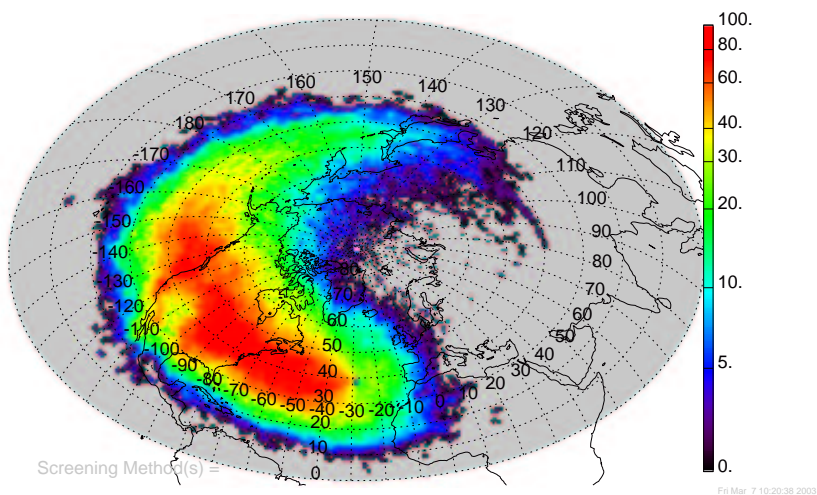


Figure 3.24 The geometrically corrected density plot for fall of the extended study period, 1960-1999.

Table 3.8 Summary of the clusters for fall (1960-1999). Information about the cluster center and the trajectories belonging to each cluster (cluster membership).

Variable	SEAt E.2	AtOc E.5	NoAm E.8
Number of Trajectories	2706	2669	2392
Percent total for period: (%)	18.90	18.64	16.71
Cluster origin: (lat,lon)	(37.10,-28.59)	(30.07,-47.87)	(47.29,-93.71)
Cluster travel distance: km	1.69E+03	2.57E+03	5.37E+03
Cluster box: (max lat, max lon)	(38.47,-21.72)	(38.47,-28.40)	(47.76,-28.40)
Cluster box: (min lat, min lon)	(34.50,-29.16)	(28.72,-48.60)	(38.47,-93.71)
Cluster speed: km/hr (m/s)	7.03 (1.95)	10.71 (2.97)	22.36 (6.21)
Mean membership distance: km	6.60E+03	6.28E+03	9.50E+03
Median membership distance: km	6.32E+03	6.02E+03	9.29E+03
Standard deviation of distance: km	2.19E+03	1.80E+03	2.47E+03
Skewness of distance:	0.839	0.869	0.709
Mean membership speed: km/hr (m/s)	27.51 (7.64)	26.18 (7.27)	39.59 (11.00)
Median membership speed: km/hr (m/s)	26.33 (7.31)	25.07 (6.97)	38.72 (10.76)
Standard deviation of speed: km/hr	9.11 (2.53)	7.48 (2.08)	10.28 (2.86)
Membership box: max(lat,lon)	(90.00,239.92)	(87.07,146.17)	(89.22,124.79)
Membership box: min(lat,lon)	(1.78,-100.16)	(-2.03,-99.07)	(17.05,-214.50)

Variable	PaAm E.11	PaAm E.14	NoAm E.17
Number of Trajectories	2244	2115	2189
Percent total for period: (%)	15.68	14.77	15.29
Cluster origin: (lat,lon)	(47.79,-155.67)	(45.92,-197.49)	(36.47,-101.26)
Cluster travel distance: km	9.66E+03	1.28E+04	6.41E+03
Cluster box: max(lat,lon)	(51.28,-28.40)	(51.26,-28.40)	(42.79,-28.40)
Cluster box: min(lat,lon)	(38.47,-155.67)	(38.47,-197.49)	(36.10,-101.26)
Cluster speed: km/hr (m/s)	40.25 (11.18)	53.42 (14.84)	26.72 (7.42)
Mean membership distance: km	1.22E+04	1.52E+04	9.15E+03
Median membership distance: km	1.19E+04	1.47E+04	8.94E+03
Standard deviation of distance: km	2.56E+03	3.33E+03	1.95E+03
Skewness of distance:	0.759	0.696	0.578
Mean membership speed: km/hr (m/s)	51.01 (14.17)	63.36 (17.60)	38.11 (10.58)
Median membership speed: km/hr (m/s)	49.53 (13.76)	61.45 (17.07)	37.25 (10.35)
Standard deviation of speed: km/hr	10.67 (2.96)	13.86 (3.85)	8.15 (2.26)
Membership box: max(lat,lon)	(90.00,103.70)	(90.00,85.78)	(88.05,73.24)
Membership box: min(lat,lon)	(18.54,-336.40)	(11.53,-443.56)	(8.75,-143.84)

3.1.6 NAO seasons dissimilar to the average period

This section will present and discuss the seasons with positive or negative NAOI values that are dissimilar to the average period. The discussions will be fairly brief, as the flow patterns and their implications were discussed in previous sections.

3.1.6.1 Results for winter with positive NAOI values, 1960-1999

The results for winter with positive NAOI values are presented in Figures 3.25–3.27 and table 3.9. The clusters for this period are very similar to the average winter. It contains two PaAm clusters, as well as two NoAm clusters, one NoNA and one SEAt cluster. All these clusters are in similar locations and have a similar relative number of trajectories.

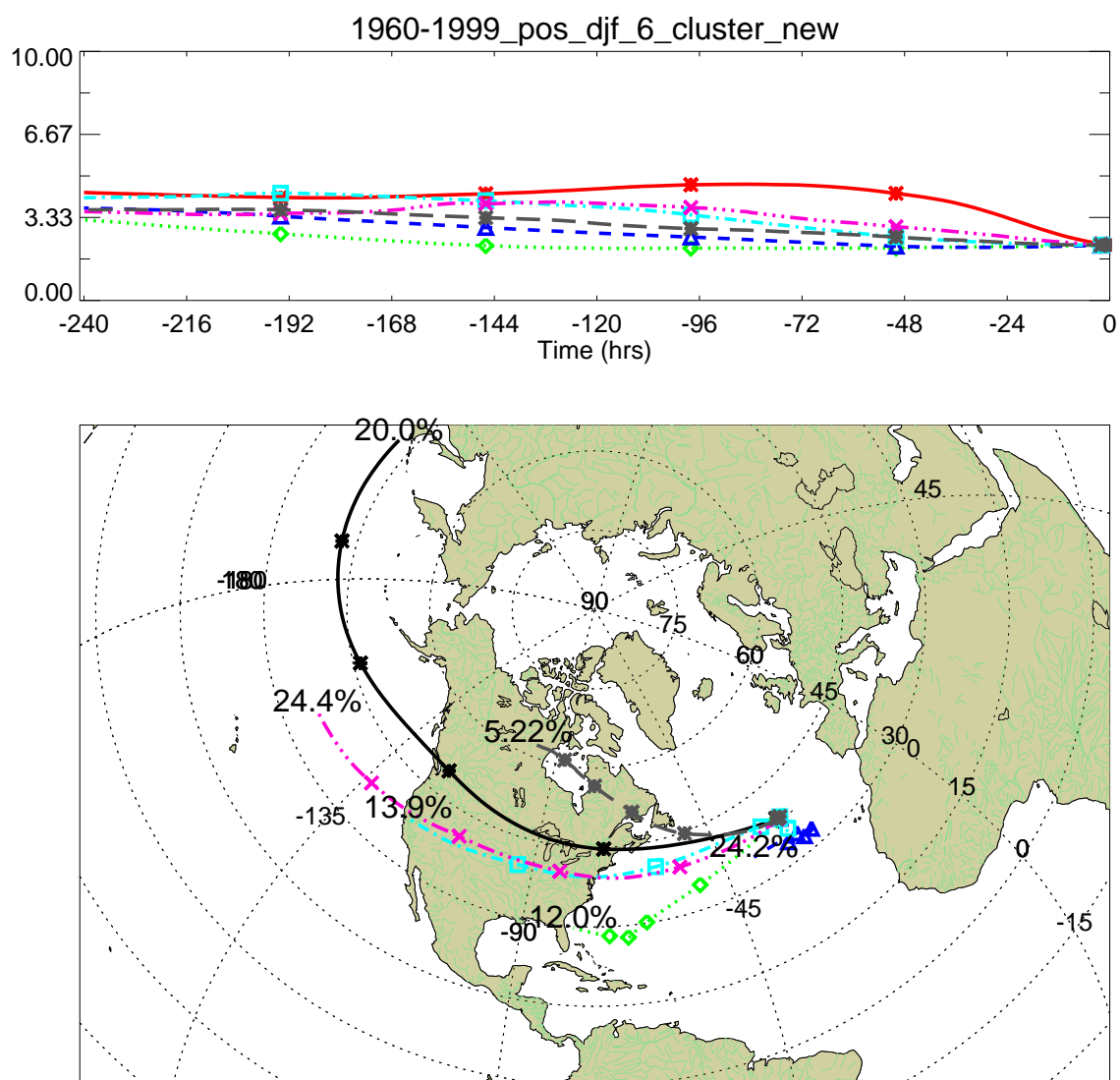


Figure 3.25 The plot of the cluster centers for winters with a positive NAOI, 1960-1999. Percentages shown are the percentage of trajectories belonging to each cluster, relative to the number of trajectories for this period. Symbols are plotted every 2 days.

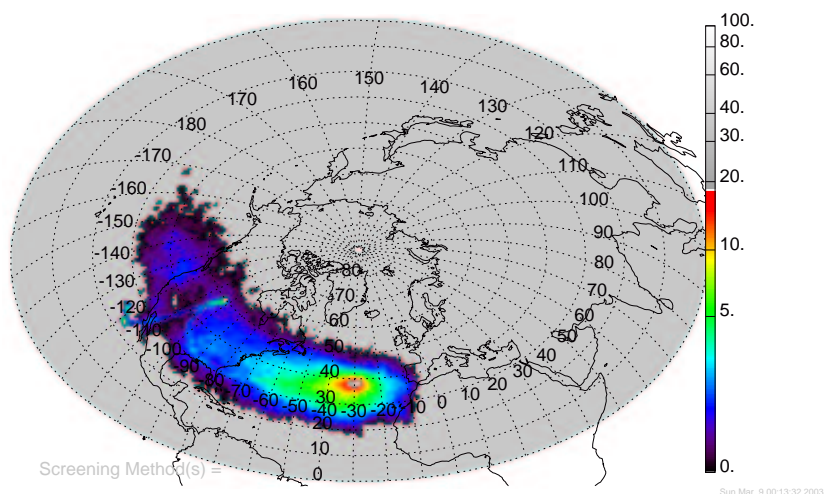


Figure 3.26 The density plot for fall for winters with a positive NAOI, 1960-1999.

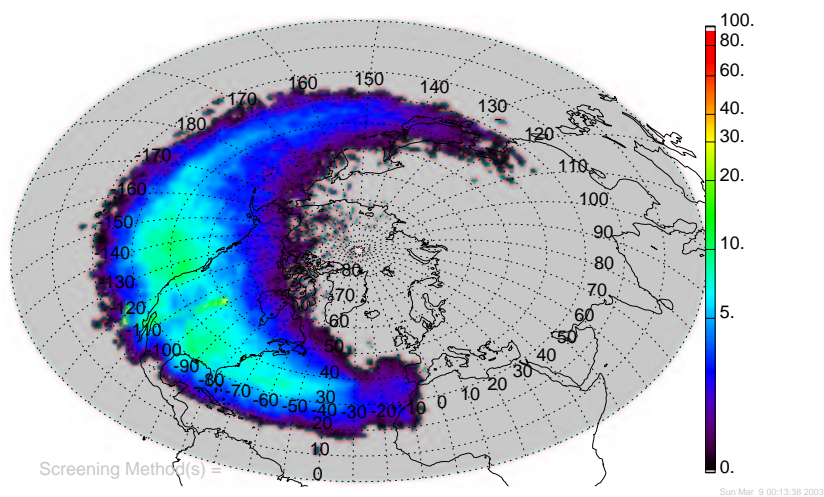


Figure 3.27 The geometrically corrected density plot for winters with a positive NAOI, 1960-1999.

Table 3.9 Summary of the clusters for positive NAO winters(1960-1999). Information about the cluster center and the trajectories belonging to each cluster (cluster membership).

Variable	PaAm B.21	NoAm B.24	SEAt B.27
Number of Trajectories	1308	786	1584
Percent total for period: (%)	20.03	12.03	24.25
Cluster origin: (lat,lon)	(42.18,-212.98)	(29.71,-79.45)	(35.05,-38.06)
Cluster travel distance: km	1.47E+04	5.00E+03	2.46E+03
Cluster box: (max lat, max lon)	(49.93,-28.40)	(38.47,-28.40)	(38.47,-24.29)
Cluster box: (min lat, min lon)	(38.47,-212.98)	(27.66,-79.45)	(32.72,-38.06)
Cluster speed: km/hr (m/s)	61.19 (17.00)	20.81 (5.78)	10.23 (2.84)
Mean membership distance: km	1.77E+04	8.99E+03	8.04E+03
Median membership distance: km	1.71E+04	8.77E+03	7.56E+03
Standard deviation of distance: km	3.63E+03	2.02E+03	2.63E+03
Skewness of distance:	0.578	0.575	0.667
Mean membership speed: km/hr (m/s)	73.59 (20.44)	37.47 (10.41)	33.49 (9.30)
Median membership speed: km/hr (m/s)	71.45 (19.85)	36.56 (10.16)	31.51 (8.75)
Standard deviation of speed: km/hr	15.13 (4.20)	8.43 (2.34)	10.94 (3.04)
Membership box: max(lat,lon)	(89.69,31.16)	(71.22,20.03)	(87.50,151.88)
Membership box: min(lat,lon)	(13.41,-394.90)	(7.14,-134.37)	(4.02,-113.55)

Variable	NoAm B.30	PaAm B.33	NoNA B.36
Number of Trajectories	913	1599	341
Percent total for period: (%)	13.98	24.48	5.22
Cluster origin: (lat,lon)	(38.11,-121.76)	(36.23,-153.09)	(62.19,-102.73)
Cluster travel distance: km	8.68E+03	1.08E+04	5.89E+03
Cluster box: max(lat,lon)	(40.52,-28.10)	(40.79,-28.40)	(62.32,-28.40)
Cluster box: min(lat,lon)	(35.66,-121.76)	(36.23,-153.09)	(38.47,-102.73)
Cluster speed: km/hr (m/s)	36.15 (10.04)	45.08 (12.52)	24.54 (6.82)
Mean membership distance: km	1.23E+04	1.35E+04	1.14E+04
Median membership distance: km	1.21E+04	1.31E+04	1.07E+04
Standard deviation of distance: km	2.83E+03	2.70E+03	2.86E+03
Skewness of distance:	0.514	0.705	1.100
Mean membership speed: km/hr (m/s)	51.42 (14.28)	56.33 (15.65)	47.34 (13.15)
Median membership speed: km/hr (m/s)	50.44 (14.01)	54.38 (15.10)	44.52 (12.37)
Standard deviation of speed: km/hr	11.81 (3.28)	11.23 (3.12)	11.90 (3.31)
Membership box: max(lat,lon)	(82.23,54.37)	(88.78,36.60)	(89.86,207.76)
Membership box: min(lat,lon)	(15.23,-214.73)	(7.65,-257.14)	(24.22,-256.72)

3.1.6.2 Results for winter with negative NAOI values, 1960-1999

The results for winter with negative NAOI values are presented in Figures 3.28–3.30 and table 3.10. The clusters for this period have several similar features to the average winter. It contains two PaAm clusters, as well as two NoAm clusters. However, the two NoAm clusters are more similar to the NoAm clusters for the average period than the NoAm clusters for the average winter. There is also appears to be a AtOc and an EeAt cluster. However, an examination of the density plots the AtOc cluster (Figure B.55–Figure B.57) reveals that some members spends time over Africa, which means it should be classified as in the SEAt. The EeAt cluster appears to extend much farther north, and an inspection of the membership plots (Figure B.40–Figure B.42) shows that this cluster does not spend any time over Africa and extend exceptionally far north. It is uncertain if this cluster needs a new classification or whether it should remain under the EeAt classification. Because of the dramatic difference in the cluster center from other cluster centers under the EeAt classification, it will receive a new classification, North-Eastern Atlantic (NEAt).

Some of the differences between this period and the average winter can be explained by the fact that during negative NAO periods, the speed of flow across the Atlantic is decreased from the average winter. This would mean flow patterns would be more similar to the average period, which they are with the presence of the NEAt cluster, which is a more northerly version of the EeAt cluster, in the average period.

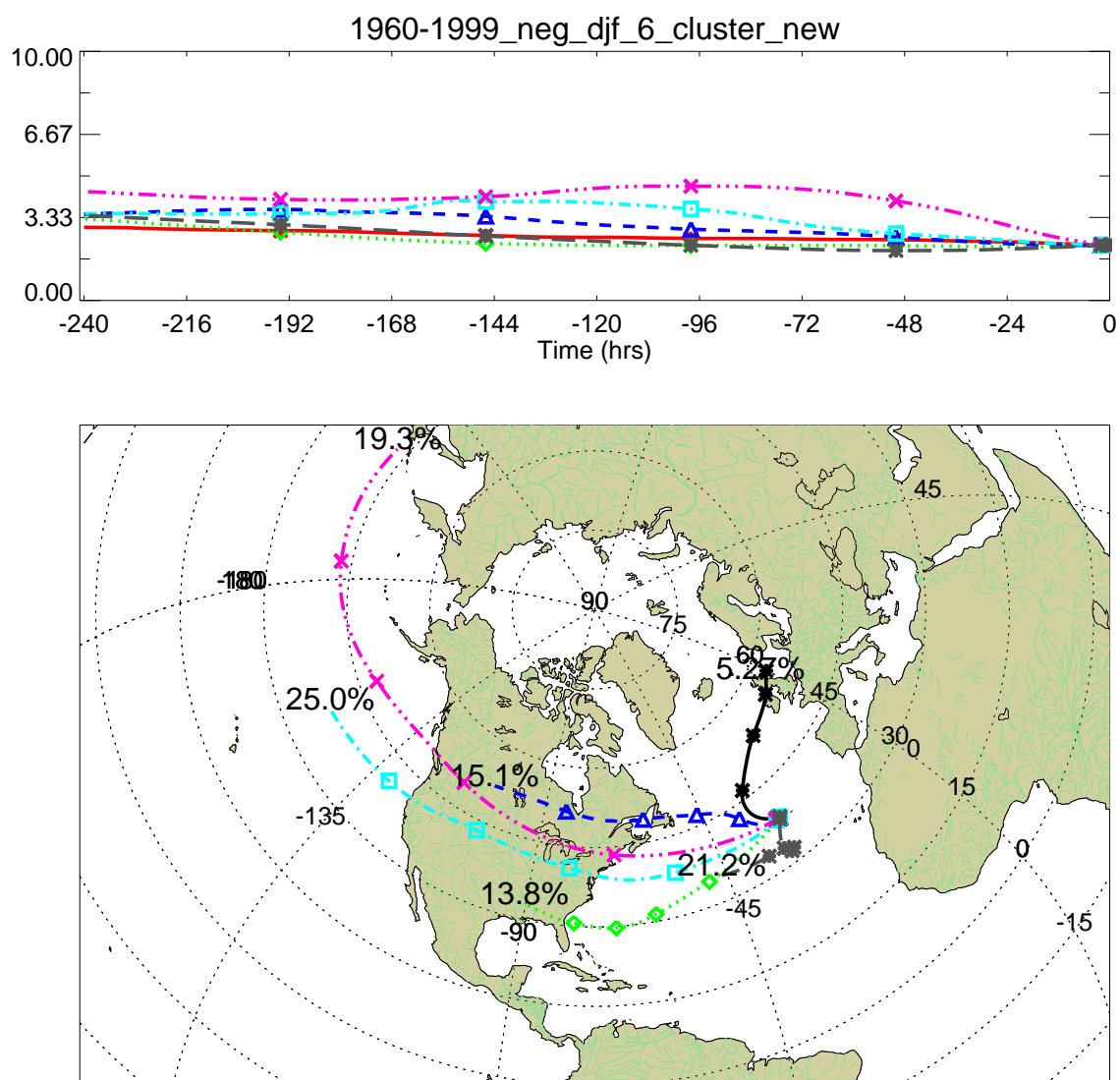


Figure 3.28 The plot of the cluster centers for winters with a negative NAOI, 1960-1999. Percentages shown are the percentage of trajectories belonging to each cluster, relative to the number of trajectories for this period. Symbols are plotted every 2 days.

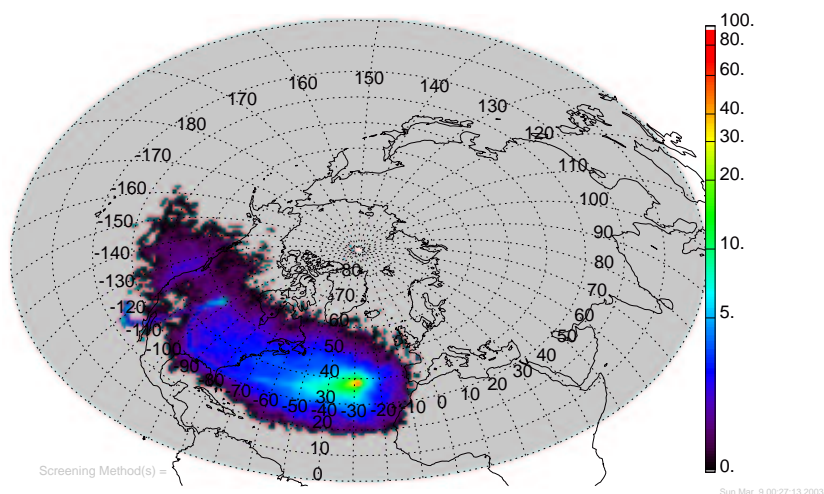


Figure 3.29 The density plot for winters with a negative NAOI, 1960-1999.

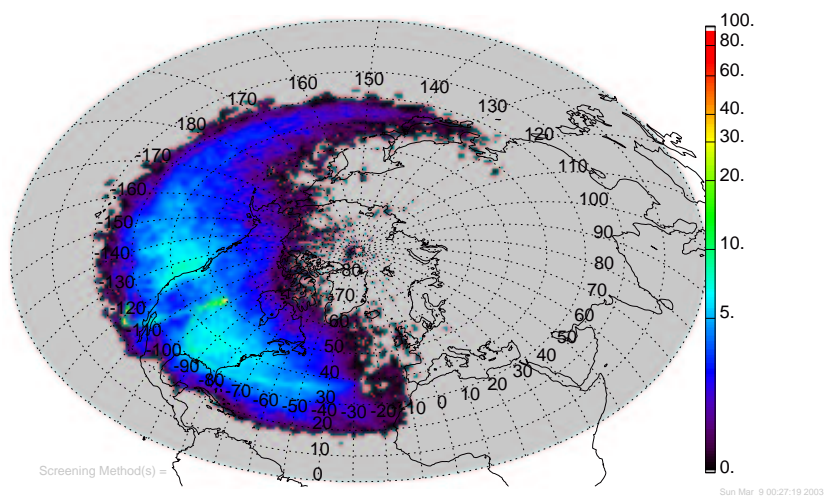


Figure 3.30 The geometrically corrected density plot for winters with a negative NAOI, 1960-1999.

Table 3.10 Summary of the clusters for negative NAO winters (1960-1999). Information about the cluster center and the trajectories belonging to each cluster (cluster membership).

Variable	NEAt B.40	NoAm B.43	NoAm B.46
Number of Trajectories	301	791	862
Percent total for period: (%)	5.28	13.87	15.11
Cluster origin: (lat,lon)	(57.09,-3.54)	(32.98,-90.25)	(52.06,-106.74)
Cluster travel distance: km	3.98E+03	5.93E+03	6.14E+03
Cluster box: (max lat, max lon)	(57.09,0.98)	(38.47,-28.40)	(52.54,-28.40)
Cluster box: (min lat, min lon)	(38.47,-33.95)	(29.68,-90.25)	(38.46,-106.74)
Cluster speed: km/hr (m/s)	16.60 (4.61)	24.69 (6.86)	25.60 (7.11)
Mean membership distance: km	9.88E+03	9.93E+03	1.21E+04
Median membership distance: km	9.61E+03	9.73E+03	1.17E+04
Standard deviation of distance: km	2.59E+03	2.41E+03	3.05E+03
Skewness of distance:	0.855	0.539	0.560
Mean membership speed: km/hr (m/s)	41.16 (11.43)	41.38 (11.49)	50.31 (13.98)
Median membership speed: km/hr (m/s)	40.05 (11.13)	40.56 (11.27)	48.73 (13.54)
Standard deviation of speed: km/hr	10.79 (3.00)	10.04 (2.79)	12.72 (3.53)
Membership box: max(lat,lon)	(89.75,262.77)	(84.84,219.20)	(89.91,151.83)
Membership box: min(lat,lon)	(21.36,-170.15)	(5.04,-155.76)	(17.81,-260.93)

Variable	PaAm B.49	PaAm B.52	SEAt B.55
Number of Trajectories	1431	1105	1214
Percent total for period: (%)	25.09	19.37	21.28
Cluster origin: (lat,lon)	(38.51,-152.04)	(43.05,-211.24)	(34.73,-46.18)
Cluster travel distance: km	1.06E+04	1.45E+04	2.52E+03
Cluster box: max(lat,lon)	(43.15,-28.40)	(50.59,-28.40)	(38.47,-28.40)
Cluster box: min(lat,lon)	(38.08,-152.04)	(38.47,-211.24)	(32.01,-46.18)
Cluster speed: km/hr (m/s)	44.15 (12.26)	60.55 (16.82)	10.49 (2.91)
Mean membership distance: km	1.35E+04	1.76E+04	8.22E+03
Median membership distance: km	1.29E+04	1.69E+04	7.82E+03
Standard deviation of distance: km	2.72E+03	3.83E+03	2.61E+03
Skewness of distance:	0.896	0.735	0.630
Mean membership speed: km/hr (m/s)	56.44 (15.68)	73.16 (20.32)	34.25 (9.51)
Median membership speed: km/hr (m/s)	53.73 (14.92)	70.49 (19.58)	32.60 (9.06)
Standard deviation of speed: km/hr	11.35 (3.15)	15.94 (4.43)	10.87 (3.02)
Membership box: max(lat,lon)	(90.00,209.67)	(90.00,99.84)	(84.57,60.09)
Membership box: min(lat,lon)	(5.96,-296.43)	(10.30,-403.79)	(4.20,-138.86)

3.1.6.3 Results for summer with positive NAOI values, 1960-1999

The results for summer with positive NAOI values are presented in Figures 3.31–3.33 and table 3.11. The clusters for this period are very similar to the average summer. It contains one PaAm cluster, two NoAm clusters, one SWAt cluster, one AtOc cluster and one EeAt cluster.

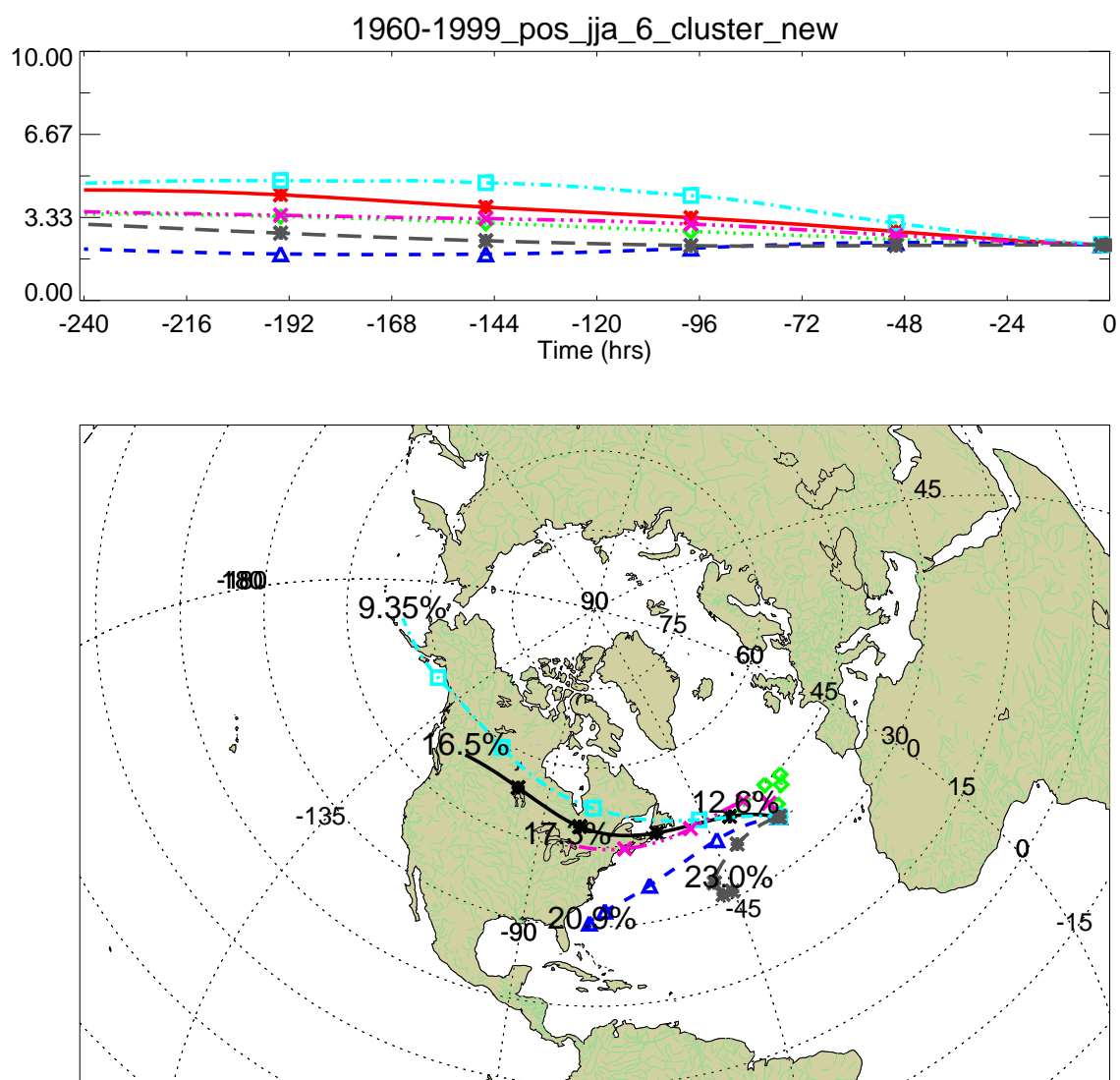


Figure 3.31 The plot of the cluster centers for summers with a positive NAOI, 1960-1999. Percentages shown are the percentage of trajectories belonging to each cluster, relative to the number of trajectories for this period. Symbols are plotted every 2 days.

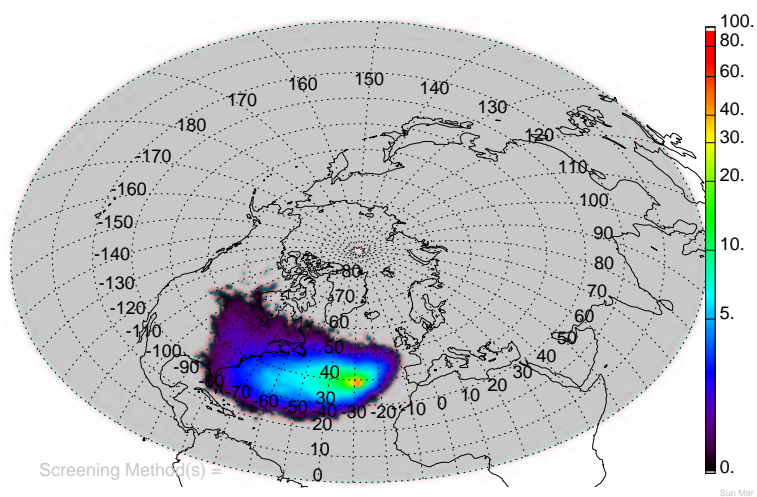


Figure 3.32 The density plot for fall for summers with a positive NAOI, 1960-1999.

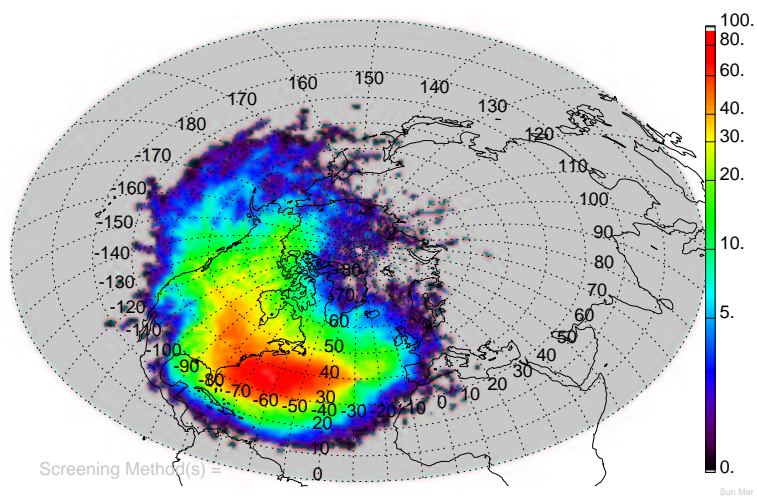


Figure 3.33 The geometrically corrected density plot for summers with a positive NAOI, 1960-1999.

Table 3.11 Summary of the clusters for positive NAO summers (1960-1999). Information about the cluster center and the trajectories belonging to each cluster (cluster membership).

Variable	NoAm D.21	EeAt D.24	SWAt D.27
Number of Trajectories	1135	867	1433
Percent total for period: (%)	16.60	12.68	20.95
Cluster origin: (lat,lon)	(53.86,-119.65)	(43.84,-36.87)	(29.74,-75.50)
Cluster travel distance: km	7.06E+03	2.09E+03	4.63E+03
Cluster box: (max lat, max lon)	(54.41,-28.40)	(44.75,-21.84)	(41.28,-28.40)
Cluster box: (min lat, min lon)	(38.47,-119.65)	(38.47,-36.87)	(29.74,-76.20)
Cluster speed: km/hr (m/s)	29.40 (8.17)	8.72 (2.42)	19.31 (5.36)
Mean membership distance: km	9.04E+03	6.37E+03	6.98E+03
Median membership distance: km	8.77E+03	6.25E+03	6.79E+03
Standard deviation of distance: km	1.72E+03	2.13E+03	1.35E+03
Skewness of distance:	0.733	0.640	0.576
Mean membership speed: km/hr (m/s)	37.68 (10.47)	26.53 (7.37)	29.09 (8.08)
Median membership speed: km/hr (m/s)	36.55 (10.15)	26.02 (7.23)	28.29 (7.86)
Standard deviation of speed: km/hr	7.18 (2.00)	8.86 (2.46)	5.64 (1.57)
Membership box: max(lat,lon)	(90.00,105.88)	(86.92,199.14)	(73.66,18.59)
Membership box: min(lat,lon)	(21.39,-244.91)	(10.11,-85.69)	(7.18,-102.84)

Variable	PaAm D.30	NoAm D.33	AtOc D.36
Number of Trajectories	640	1186	1578
Percent total for period: (%)	9.36	17.34	23.07
Cluster origin: (lat,lon)	(55.07,-168.15)	(45.40,-82.35)	(32.08,-45.63)
Cluster travel distance: km	9.66E+03	4.79E+03	2.63E+03
Cluster box: max(lat,lon)	(60.17,-28.40)	(45.45,-28.40)	(39.46,-28.40)
Cluster box: min(lat,lon)	(38.47,-168.15)	(38.47,-82.35)	(31.09,-48.59)
Cluster speed: km/hr (m/s)	40.25 (11.18)	19.97 (5.55)	10.97 (3.05)
Mean membership distance: km	1.13E+04	7.53E+03	5.83E+03
Median membership distance: km	1.11E+04	7.31E+03	5.68E+03
Standard deviation of distance: km	2.07E+03	1.78E+03	1.64E+03
Skewness of distance:	0.472	0.812	0.694
Mean membership speed: km/hr (m/s)	47.05 (13.07)	31.38 (8.72)	24.31 (6.75)
Median membership speed: km/hr (m/s)	46.37 (12.88)	30.48 (8.47)	23.67 (6.58)
Standard deviation of speed: km/hr	8.62 (2.40)	7.42 (2.06)	6.81 (1.89)
Membership box: max(lat,lon)	(90.00,185.01)	(84.61,120.13)	(64.26,29.70)
Membership box: min(lat,lon)	(24.80,-384.20)	(17.42,-170.80)	(7.77,-106.16)

3.1.6.4 Results for summer with negative NAOI values, 1960-1999

The results for summer with negative NAOI values are presented in Figures 3.34–3.36 and table 3.12. The clusters for this period are very similar to the average summer. It contains one PaAm cluster, two NoAm clusters, one SWAt cluster, one AtOc cluster and one EeAt cluster.

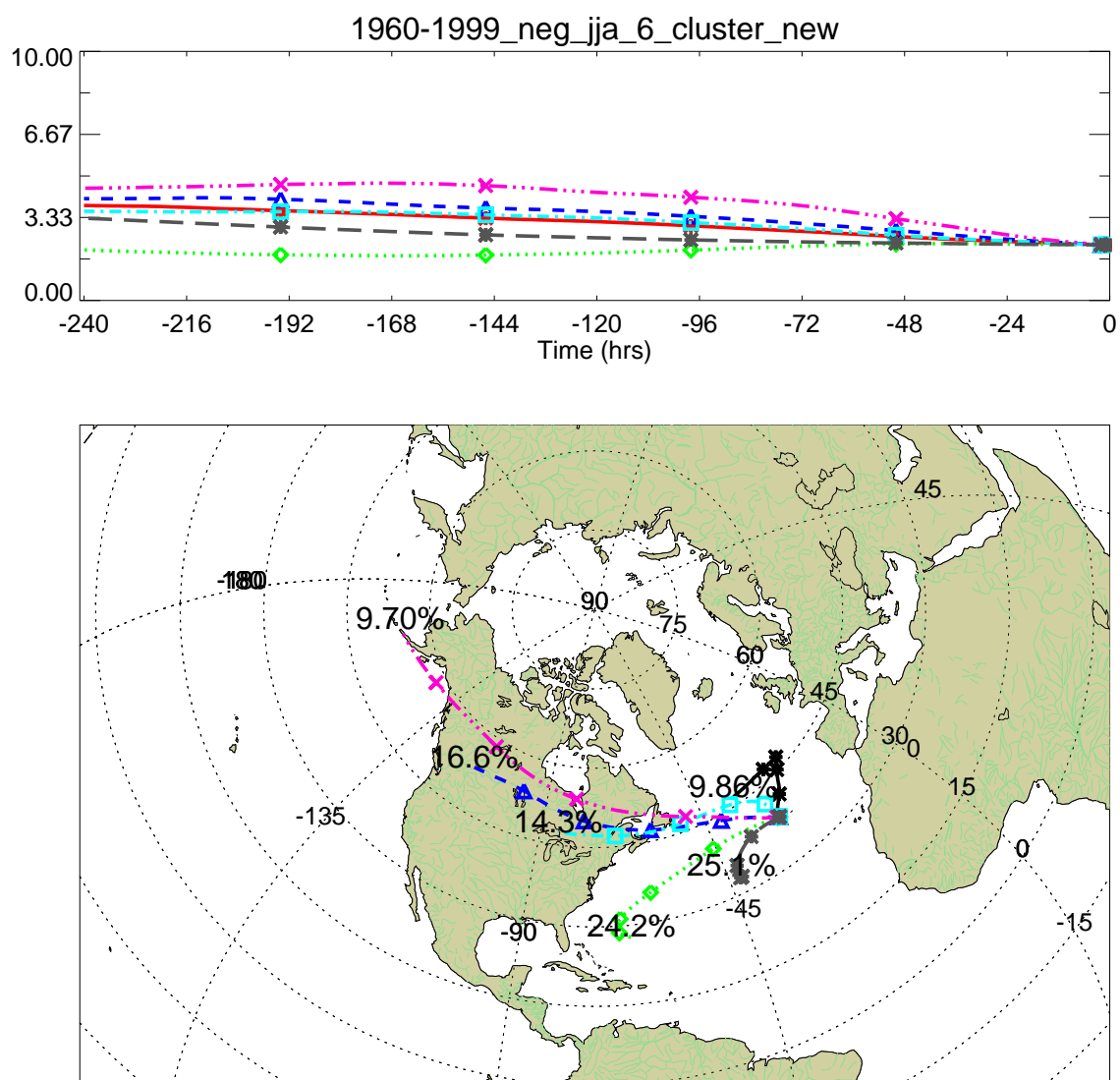


Figure 3.34 The plot of the cluster centers for summers with a negative NAOI, 1960-1999. Percentages shown are the percentage of trajectories belonging to each cluster, relative to the number of trajectories for this period. Symbols are plotted every 2 days.

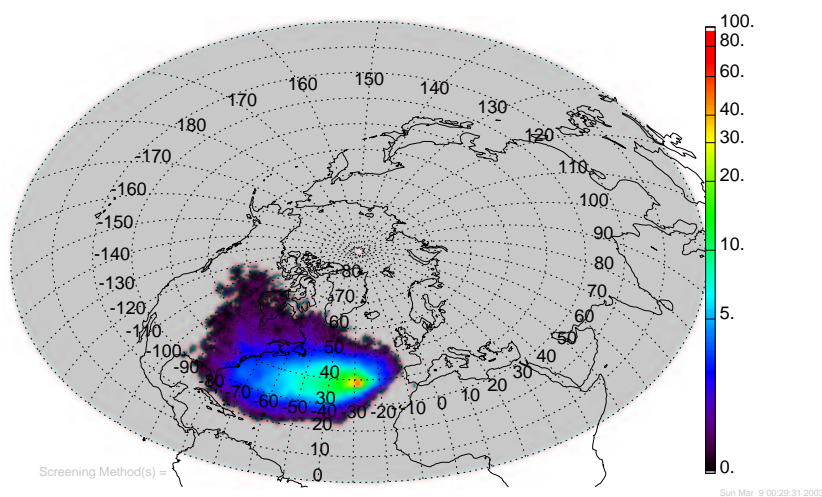


Figure 3.35 The density plot for fall for summers with a negative NAOI, 1960-1999.

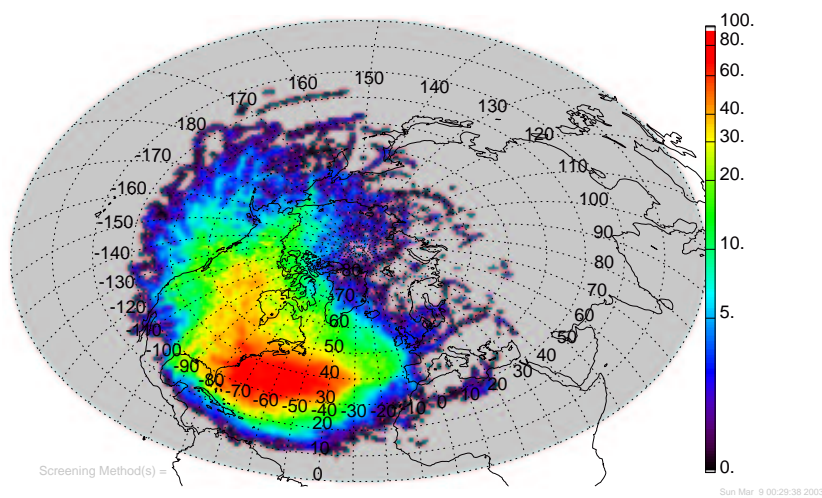


Figure 3.36 The geometrically corrected density plot for summers with a negative NAOI, 1960-1999.

Table 3.12 Summary of the clusters for negative NAO summers (1960-1999). Information about the cluster center and the trajectories belonging to each cluster (cluster membership).

Variable	EeAt D.40	SWAt D.43	NoAm D.46
Number of Trajectories	503	1236	851
Percent total for period: (%)	9.86	24.24	16.69
Cluster origin: (lat,lon)	(46.59,-35.15)	(27.77,-67.63)	(53.04,-115.06)
Cluster travel distance: km	2.41E+03	4.46E+03	6.67E+03
Cluster box: (max lat, max lon)	(47.05,-19.39)	(40.29,-28.40)	(53.71,-28.40)
Cluster box: (min lat, min lon)	(38.47,-35.15)	(27.77,-70.45)	(38.47,-115.06)
Cluster speed: km/hr (m/s)	10.05 (2.79)	18.58 (5.16)	27.78 (7.72)
Mean membership distance: km	6.59E+03	6.90E+03	8.95E+03
Median membership distance: km	6.39E+03	6.80E+03	8.70E+03
Standard deviation of distance: km	2.22E+03	1.32E+03	1.67E+03
Skewness of distance:	0.845	0.654	0.629
Mean membership speed: km/hr (m/s)	27.45 (7.62)	28.73 (7.98)	37.30 (10.36)
Median membership speed: km/hr (m/s)	26.64 (7.40)	28.35 (7.87)	36.26 (10.07)
Standard deviation of speed: km/hr	9.23 (2.56)	5.50 (1.53)	6.96 (1.93)
Membership box: max(lat,lon)	(88.64,275.14)	(89.43,166.58)	(89.85,170.79)
Membership box: min(lat,lon)	(20.48,-112.54)	(4.97,-91.44)	(22.98,-207.24)

Variable	NoAm D.49	PaAm D.52	AtOc D.55
Number of Trajectories	733	495	1281
Percent total for period: (%)	14.38	9.71	25.12
Cluster origin: (lat,lon)	(47.42,-85.10)	(54.53,-165.20)	(34.04,-44.21)
Cluster travel distance: km	4.83E+03	9.49E+03	1.93E+03
Cluster box: max(lat,lon)	(47.42,-28.40)	(59.21,-28.40)	(38.92,-28.40)
Cluster box: min(lat,lon)	(38.47,-85.10)	(38.47,-165.20)	(32.70,-44.21)
Cluster speed: km/hr (m/s)	20.11 (5.59)	39.55 (10.99)	8.04 (2.23)
Mean membership distance: km	7.64E+03	1.12E+04	5.59E+03
Median membership distance: km	7.35E+03	1.11E+04	5.41E+03
Standard deviation of distance: km	1.81E+03	2.14E+03	1.67E+03
Skewness of distance:	0.733	0.541	0.458
Mean membership speed: km/hr (m/s)	31.83 (8.84)	46.82 (13.01)	23.29 (6.47)
Median membership speed: km/hr (m/s)	30.61 (8.50)	46.06 (12.79)	22.54 (6.26)
Standard deviation of speed: km/hr	7.55 (2.10)	8.94 (2.48)	6.95 (1.93)
Membership box: max(lat,lon)	(89.93,152.03)	(89.70,15.68)	(69.72,42.74)
Membership box: min(lat,lon)	(18.35,-214.94)	(23.19,-403.31)	(8.91,-75.94)

3.1.6.5 Results for fall with positive NAOI values, 1960-1999

The results for fall with positive NAOI values are presented in Figures 3.37–3.39 and table 3.13. The clusters for this period are different from the average fall. It contains two PaAm clusters, but only one NoAm cluster. In addition it contains one SWAt, one SEAt, and one EeAt cluster. Enhanced westerly flow would be expected during any positive NAO period. However, that does not appear to be the case here. One possible explanation is that the center of action for the Azores High is more centered over the Azores during positive NAO falls, which means there would be more circulation around the Azores as opposed to increased flow over the Azores. The clusters are consistent with this hypothesis, as the farther reaching clusters have higher altitudes and the clusters that then circulate over the Atlantic have lower altitudes. The clusters with higher altitudes are transported higher in the free troposphere before being drawn down by the high pressure center. On the other hand, the clusters with lower altitudes are already caught in the circulation around the pressure center and thus spend more time over the ocean around the pressure center. However, this is speculation based upon the assumption that the Azores High resides over the Azores during these periods.

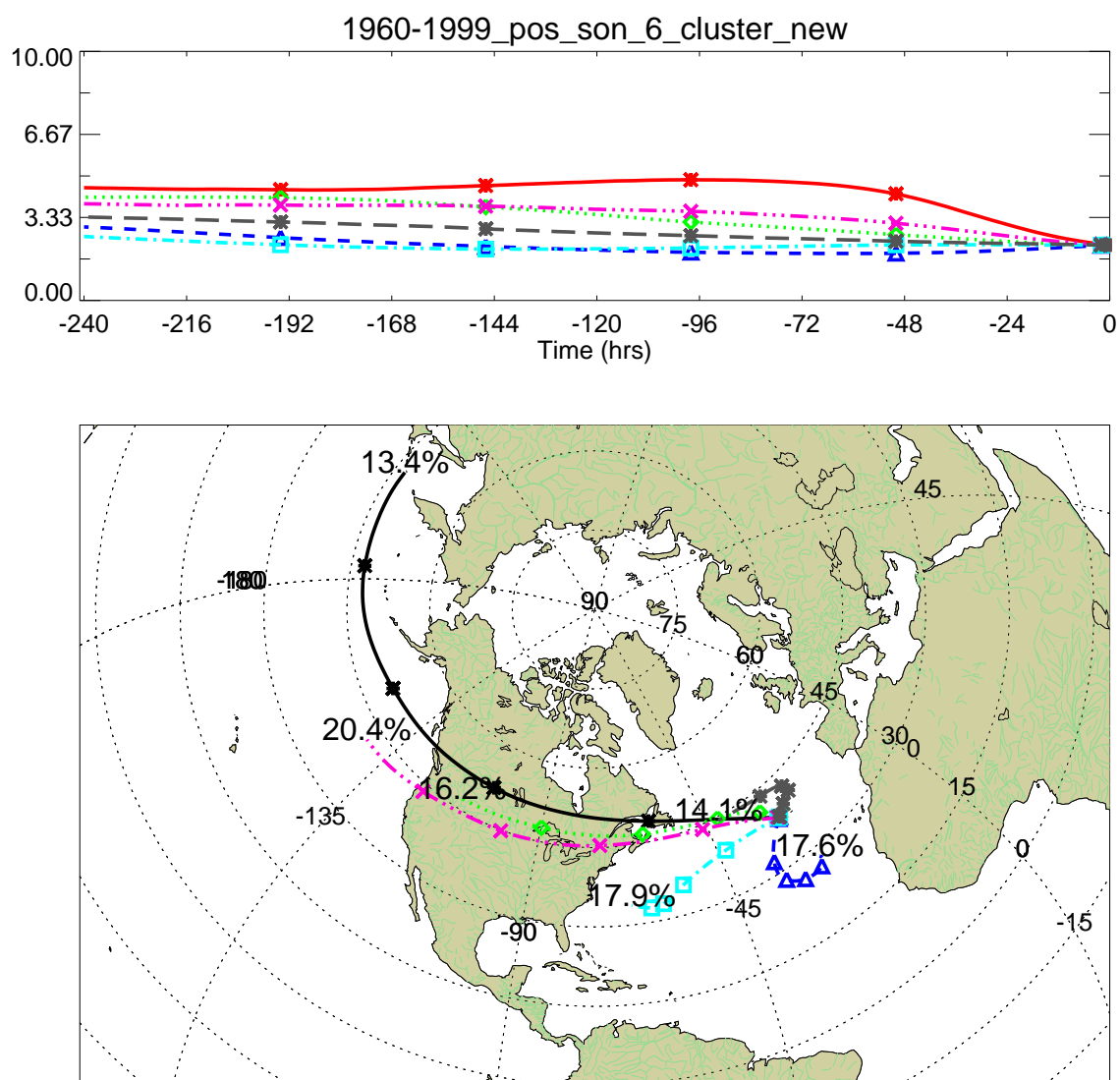


Figure 3.37 The plot of the cluster centers for falls with a positive NAOI, 1960-1999. Percentages shown are the percentage of trajectories belonging to each cluster, relative to the number of trajectories for this period. Symbols are plotted every 2 days.

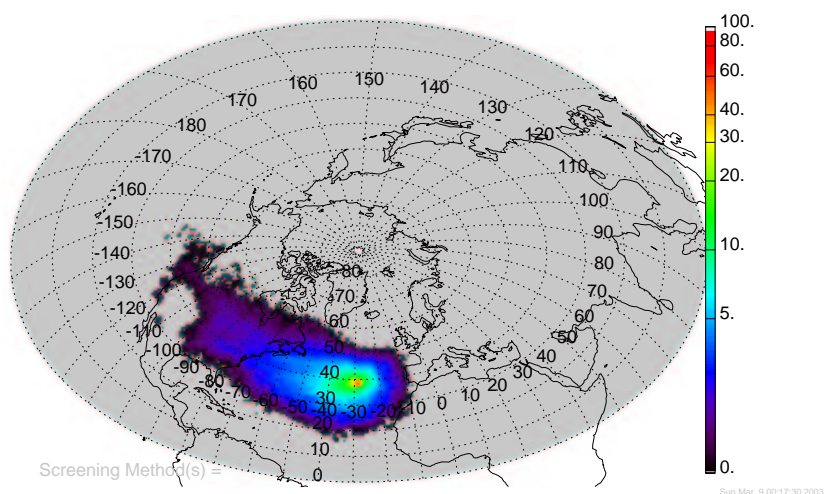


Figure 3.38 The density plot for fall for falls with a positive NAOI, 1960-1999.

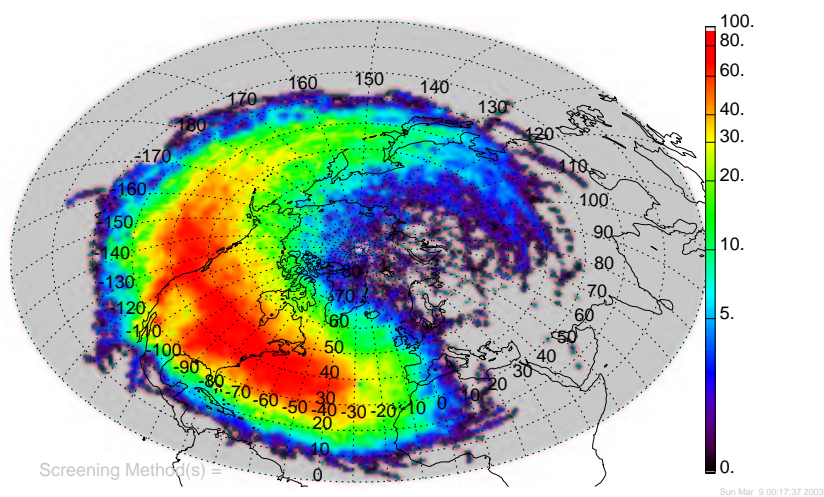


Figure 3.39 The geometrically corrected density plot for falls with a positive NAOI, 1960-1999.

Table 3.13 Summary of the clusters for positive NAO falls (1960-1999). Information about the cluster center and the trajectories belonging to each cluster (cluster membership).

Variable	PaAm E.21	NoAm E.24	SEAt E.27
Number of Trajectories	916	1106	1200
Percent total for period: (%)	13.45	16.24	17.62
Cluster origin: (lat,lon)	(46.85,-207.67)	(46.93,-113.09)	(28.16,-26.30)
Cluster travel distance: km	1.34E+04	6.79E+03	2.42E+03
Cluster box: (max lat, max lon)	(51.69,-28.40)	(47.69,-28.40)	(38.47,-25.95)
Cluster box: (min lat, min lon)	(38.47,-207.67)	(38.47,-113.09)	(26.14,-35.17)
Cluster speed: km/hr (m/s)	55.89 (15.52)	28.29 (7.86)	10.10 (2.81)
Mean membership distance: km	1.57E+04	1.03E+04	5.77E+03
Median membership distance: km	1.53E+04	9.87E+03	5.70E+03
Standard deviation of distance: km	3.28E+03	2.59E+03	1.54E+03
Skewness of distance:	0.587	0.778	0.532
Mean membership speed: km/hr (m/s)	65.52 (18.20)	43.07 (11.96)	24.04 (6.68)
Median membership speed: km/hr (m/s)	63.67 (17.68)	41.15 (11.43)	23.73 (6.59)
Standard deviation of speed: km/hr	13.65 (3.79)	10.81 (3.00)	6.40 (1.78)
Membership box: max(lat,lon)	(90.00,85.78)	(89.49,78.90)	(64.40,50.48)
Membership box: min(lat,lon)	(11.53,-429.55)	(16.45,-214.50)	(-1.74,-76.94)

Variable	SWAt E.30	PaAm E.33	EeAt E.36
Number of Trajectories	1225	1396	967
Percent total for period: (%)	17.99	20.50	14.20
Cluster origin: (lat,lon)	(33.67,-66.96)	(41.95,-141.55)	(44.28,-41.57)
Cluster travel distance: km	3.61E+03	9.22E+03	2.25E+03
Cluster box: max(lat,lon)	(39.15,-28.40)	(45.41,-28.40)	(44.28,-22.66)
Cluster box: min(lat,lon)	(32.54,-66.96)	(38.47,-141.55)	(38.47,-41.57)
Cluster speed: km/hr (m/s)	15.04 (4.18)	38.40 (10.67)	9.39 (2.61)
Mean membership distance: km	7.33E+03	1.14E+04	7.56E+03
Median membership distance: km	7.08E+03	1.11E+04	7.30E+03
Standard deviation of distance: km	2.01E+03	2.59E+03	2.53E+03
Skewness of distance:	0.726	0.771	0.687
Mean membership speed: km/hr (m/s)	30.52 (8.48)	47.63 (13.23)	31.49 (8.75)
Median membership speed: km/hr (m/s)	29.49 (8.19)	46.30 (12.86)	30.40 (8.44)
Standard deviation of speed: km/hr	8.37 (2.33)	10.78 (2.99)	10.55 (2.93)
Membership box: max(lat,lon)	(88.05,73.24)	(86.55,31.67)	(89.58,239.92)
Membership box: min(lat,lon)	(7.39,-125.80)	(8.75,-330.00)	(15.01,-111.15)

3.1.7 Summary of cluster analysis

The cluster analysis shows that flow to the PICO-NARE site is dominated by the westerly motion of the mid-latitudes. All periods, with the exception of summers, have two clusters that originate over the Pacific Ocean and travel over North America on their way to the site (PaAm clusters). The summer seasons only have one PaAm cluster. This can be explained by the decreased westerly flow that generally occurs in the Northern Hemisphere as well as the strengthening of the Azores High in the summer seasons.

The dominance of westerly flow is also apparent by the presence of clusters that originate over North America or travel over North America. The PaAm clusters travel over North America and appear in all periods except summer. All periods have at least two clusters that originate over North America. These clusters include NoAm, NoNA, and NoNA.

The final feature is flow that appears to be dominated by circulation around either the Azores High or the Icelandic Low. These clusters are not necessarily slower than the clusters that originate over North America. These clusters also present the opportunity for transport from Europe, Africa, and the Arctic, as air from these regions appears to be drawn into circulation around the pressure centers and then transported to the site. The SEAt, SWAt, and AtOc usually appear to circulate around the Azores High, while EeAt appears to circulate around the Icelandic Low, then spend time circulating around the Azores High.

3.2 Additional discussion

3.2.1 Comments on the vertical component of trajectories

Thus far, little mention has been made of the vertical component of the trajectories. Certainly the vertical position of an airmass can have profound impacts its compo-

Table 3.14 Summary of source regions associated with each cluster. *Atlantic regions generally exclude PaAm and NoAm clusters. **The Arctic, Europe, and Africa include the clusters associated with flow from these regions, not flow exclusively from these regions. Therefore, these regions are over estimated in all periods.

Period	Pacific	N. America	Atlantic*	Arctic**	Europe**	Africa**
Average Period	31.1%	64.5%	35.3%	14.4%	14.4%	14.4%
Average Winter	45.5%	77.8%	28.5%	28.5%	22.2%	22.2%
Positive NAOI Winter	44.4%	75.5%	36.5%	29.7%	24.5%	24.5%
Negative NAOI Winter	44.3%	73.2%	26.8%	26.4%	26.4%	21.2%
Average Spring	30.6%	65.4%	34.2%	12.2%	12.2%	12.2%
Positive NAOI Spring	27.9%	62.3%	34.4%	11.6%	11.6%	11.6%
Negative NAOI Spring	32.2%	66.3%	33.4%	10.2%	10.2%	10.2%
Average Summer	9.6%	66.6%	56.0%	9.5%	9.5%	9.5%
Positive NAOI Summer	9.4%	64.1%	56.5%	12.6%	12.6%	12.6%
Negative NAOI Summer	9.7%	64.8%	59.2%	9.9%	9.9%	9.9%
Average Fall	30.3%	62.2%	31.5%	12.9%	12.9%	12.9%
Positive NAOI Fall	33.8%	67.9%	49.6%	14.1%	14.1%	31.7%
Negative NAOI Fall	30.0%	62.5%	37.5%	17.4%	17.4%	17.4%

sition and should not be ignored. For example, a change of 10 km in the horizontal direction can have very little impact on the type of air (polluted or clean) that mixes with a parcel. On the other hand, an equal change in the vertical direction can have a profound impact. Air at lower altitudes is able to mix with air near the surface. This could mean an increase in pollution if it is near regions with pollution sources, or relatively clean air if it is in remote regions. On the other hand, air from higher altitudes can contain high levels of O_3 from the stratosphere. Air from the upper boundary layer or lower free troposphere could contain a mix of air from lower and higher regions.

The reason the vertical component has received little attention is that during clustering, the results from unit selections showed that the vertical component plays a minimal role in determining the clusters. It was mentioned in the discussion of units for use in the cluster process that in several cases, the unit systems that included height resulted in identical clusters as unit systems that excluded the height, particularly in the unit system chosen for use in the analysis (KLLH). This implies that the primary component of a cluster is the horizontal coordinates. This is true for

two reasons. First, because the magnitude of the horizontal coordinates are several orders of magnitude larger than those for the vertical can have very little impact on the type of air (polluted or clean) coordinates. Second, because the variability of the horizontal coordinates is larger in the horizontal than the vertical, particularly the east/west coordinate. This can easily be seen in the histogram for the KLLH system (Figure 2.16). The altitude coordinates are almost normally distributed around the starting height, with the exception of the cutoff forced by the lower boundary (*i.e.*, altitude of 0 mASL). Because the altitude plays a minimal role in the formation of the clusters, the vertical component is the primarily the average of the heights that accompany the horizontal components. It is possible that trajectories with similar horizontal path will also have a similar vertical path. When the altitude was commented on during the discussion of clusters, this appeared to be the case at times. However, there was often a large degree of variability in the heights, which is unavoidable with the large number of trajectories belonging to each cluster. An inspection of the cluster membership plots in the appendix can reveal this about the vertical components for any individual cluster, as was done during the discussion of the cluster.

3.2.2 Summary of the impacts of the NAO

During the discussion of clusters, the effect of positive and negative NAOI phases on each season were noted. It was noticed that during two seasons, the high and low NAO datasets did not deviate significantly from the seasonal average. The exception to this was most noticeable in positive NAOI falls and negative NAOI winters. It was noted that this might be caused by the location of the Azores High coupled with a stronger than normal Azores High for the season. Aside from these periods, the NAO did not appear to profoundly impact the seasonal results, despite that the flow across the Atlantic during high NAO periods is enhanced significantly, especially during the winter months. One possible cause for this is that the center of action for the Azores

High is located such that the enhanced flow is north of the Azores.

3.3 Future work

This project has begun research that has several possible extensions in the future. The most obvious expansion is the application of the results. The results from this study can be used in conjunction with any study relating to atmospheric transport to the PICO-NARE site. It is possible that it could also be extended for use in similar studies in the Azores or surrounding North Atlantic. When used in conjunction with measurement of trace gases, the results of this study can help quantify the impact of transport to the North Atlantic. This can be achieved by comparing observed transport events with the results of the climatology. This could help determine how often similar transport events occur and thus quantify their impacts.

In addition to the application of the results, the methods developed during this study have further application. In addition to the refinement of the clustering techniques, during the development of the climatology, many programs were created to automatically create trajectories and perform an analysis on them. With minor adjustments, these programs could be modified to be applied to any study requiring the use of trajectories, a cluster analysis, or an APA. This is limited by the availability of meteorological data, which is readily available from 1948–present. This time period is relatively unrestrictive, which means the methods have wide applicability both spatially and temporally.

There remains a number of ways to develop a climatology, including using shorter trajectories (*e.g.*, 5-day) and the use of meteorological charts or composite meteorology maps for various periods. There is clearly more information that can be obtained about the climatic conditions associated with the various transport patterns portrayed by the clusters. This includes a more in depth investigation into the NAO and other permanent or semi-permanent meteorological features in the North Atlantic,

specifically the Azores High. There could also be an investigation into how the site is affected by the passage of frontal systems or other common weather patterns.

Despite the usefulness of both the results and the methods, additional work can be done investigating the methods in order to produce more refined results. During the discussion of the cluster analysis, several issues of practice and application were raised. Despite a relatively thorough analysis and comparison of some of the elements of a cluster analysis, there remains no clear method for clustering trajectories. These elements of different methods include the methods to represent the coordinates of the trajectories, the method for measuring distance between the cluster variables, and the method for the selection of the number of clusters. However, other aspects of cluster analysis were not investigated in this study which also require further investigation. In the discussion of cluster analysis, two primary techniques were mentioned for clustering: hierarchical and non-hierarchical methods. No in-depth investigation was made during this study into hierarchical methods. However, there is the potential to investigate hierarchical methods and attempt to assess which method is more applicable in relation to atmospheric trajectories. It is likely that a precise method for dealing with a large number of trajectories could be developed for general use.

In addition to investigating the methods for creating clusters, there is room for work to be done in expressing clusters. In the discussion of the results, it was noted that a cluster does not portray all possible flow paths with the cluster, only the average. The use of membership plots and density plots of particular clusters can be useful in assessing some of the variation within a cluster, but it would be useful if this information were portrayed in a more direct, obvious, and quantitative fashion. One possible solution would be to introduce an 'error bar', similar to error bars used when plotting measurement data, that would show the cluster center along with a bounding line which would be plotted at a distance of, for example, one standard deviation of the variance at each point on both sides of the cluster center of the data at that particular point. This would allow an inspection of the cluster plots to portray

not only the average, but some degree of the variance within the cluster. Another possibility would be to cluster the clusters. That is, when a cluster is composed of a large number of trajectories and the cluster center does not clearly portray known (or unknown) flow paths within the cluster, a cluster analysis could be performed on the member trajectories. This would be similar to a membership plot, but would condense some of the noise of outlying trajectories that make the use of a membership plot difficult.

Another area of potential improvement is the manner in which height of the trajectories is dealt with. It was noted that, during a cluster analysis, when the vertical coordinate was used in conjunction with horizontal coordinates, the vertical coordinates had a negligible impact on the resulting clusters. Methods such as normalization and scaling allowed the vertical coordinate to have more weight in the clustering process. However, this process generally changed the weighting of horizontal components relative to each other as well, which raised questions of the validity of such procedures. One possible method to investigate is to cluster the trajectories according to their vertical coordinates first, then cluster each vertical cluster according to horizontal coordinates. This process has several foreseeable problems. The primary problem is that this would greatly increase the number of cluster plots, creating one horizontal cluster plot per vertical cluster, which somewhat defeats the process of clustering, which is to reduce the amount of items to examining to extract information from. Another problem would be the method for determining the number of clusters to choose, which is twofold. While a method for determining the number of clusters exists in terms of horizontal and vertical components combined, it is possible that this method would not be applicable to the vertical cluster. In addition, the problem would arise whether the same number of horizontal clusters would be appropriate for each vertical cluster, or if the number of horizontal cluster needs to be evaluated for each vertical cluster.

The vertical component of trajectories is dealt with poorly in the density plots

as well. The simplest method to improve how density plots deal with differences in altitudes would be to slice the atmosphere into layers (e.g., boundary layer, free troposphere) by estimating altitudes for the pertinent layers and then summing only endpoints that fall within each layer in separate probability analyses. The obvious problem with this method is the difficulty determining what layers are appropriate and estimating the altitude of these layers. It might be more appropriate to determine an elevation for each grid point individually, and perhaps individually for each moment in time according to the meteorological condition at that location. However, this would require extensive work with the meteorological data which is likely to be labor intensive. However, the output from HYSPLIT includes an estimation of the boundary layer height, which means it would at least be possible to dynamically separate endpoints above and below the boundary layer. However, no information was found to indicate how reliable HYSPLIT's estimation of the boundary layer is. Despite its uncertainties, this process might be better than the current calculation of density plots.

References

- Attrill, M. J., and M. Power, Climatic influence on a marine fish assemblage, *Nature*, *417*, 275–278, 2002.
- Brankov, E., S. T. Rao, and P. S. Porter, A trajectory-clustering-correlation methodology for examining the long-range transport of air pollutants, *Atmos. Environ.*, *32*, 1525–1534, 1998.
- Cape, J. N., J. Methven, and L. E. Hudson, The use of trajectory cluster analysis to interpret trace gas measurements at Mace Head, Ireland, *Atmos. Environ.*, *34*, 3651–3663, 2000.
- Cheng, M.-D., P. Hopke, L. Barrie, A. Rippe, M. Olson, and S. Landsberger, Qualitative determination of source regions of aerosol in Canadian high arctic, *Environ. Sci. Technol.*, *27*, 2063–2071, 1993.
- Cook, E., Multi-proxy reconstructions of the North Atlantic Oscillation (NAO) Index: a critical review and a new well-verified winter (NAO) Index reconstruction back to AD 1400, in *The North Atlantic Oscillation: Climatic Significance and Environmental Impact*, edited by J. Hurrell, Y. Kushnir, G. Ottersen, and M. Visbeck, vol. 51–62, American Geophysical Union, Washington, DC, 2003.
- Cooper, C., and F. Alley, *Air pollution control: a design approach*, Wayland Press, Inc., Prospect Heights, IL, 1999.
- Dorling, S. R., and T. D. Davies, Extending cluster analysis–synoptic meteorology links to characterise chemical climates at six northwest European monitoring stations, *Atmos. Environ.*, *29*, 145–167, 1995.
- Draxler, R. R., Trajectory optimization for balloon flight planning, *Weather and Forecasting*, *11*, 111–114, 1996.
- Draxler, R. R., and G. Hess, NOAA technical memorandum ERL ARL-224: Description of the HYSPLIT 4 modeling system, *Tech. rep.*, 1998a.

- Draxler, R. R., and G. Hess, An overview of the HYSPLIT 4 modeling system for trajectories, dispersion, and deposition, *Australian Meteorological Magazine*, 47, 295–308, 1998b.
- Everitt, B. S., S. Landau, and M. Lesse, *Cluster Analysis*, Oxford University Press, Inc., New York, NY, 2001.
- Fast, J., and C. Berkowitz, Evaluation of back trajectories associated with the ozone transport during the 1993 North Atlantic Regional Experiment, *Atmos. Environ.*, 31, 825–837, 1997.
- Fehsenfeld, F., P. Daum, W. Leaitch, D. Parrish, and G. Hubler, Transport and processing of O₃ and O₃ precursors over the North Atlantic: An overview of the 1993 North Atlantic Regional Experiment (NARE) summer intensive, *J. Geophys. Res.*, 101, 28,877–28,891, 1996.
- Finlayson-Pitts, B. J., and J. N. Pitts, Jr., *Chemistry of the Upper and Lower Atmosphere*, John Wiley, New York, 1999.
- Hamlin, A. J., Transport of ozone precursors from the Arctic troposphere to the North Atlantic region, Master's thesis, Michigan Technological University, 1995.
- Harris, J. M., and J. D. Kahl, A descriptive atmospheric transport climatology for the Mauna Loa Observatory, using clustered trajectories, *J. Geophys. Res.*, 95, 13,651–13,667, 1990.
- Harris, J. M., and J. D. W. Kahl, Analysis of 10-day isentropic plow patterns for Barrow, Alaska: 1985–1992, *J. Geophys. Res.*, 99, 25,845–25,844, 1994.
- Harris, J. M., P. P. Tans, E. J. Dlugokencky, K. A. Masarie, P. M. Lang, S. Whittlestone, and L. P. Steele, Variations in atmospheric methane at Mauna Loa Observatory related to long-range transport, *J. Geophys. Res.*, 97, 6003–6010, 1992.
- Hartigan, J. A., *Clustering algorithms*, John Wiley and Sons, Inc., New York, NY, 1975.
- Holmes, J., Estimated crop yield losses from air pollutants in California: 1989–1992, *Tech. Rep. Research Notes 97–1*, California Environmental Protection Agency, 1997.
- Honrath, R. E., and P. Fialho, The Azores Islands: A unique location for ground-based measurements in the MBL and FT of the central North Atlantic, 2001.
- Honrath, R. E., A. J. Hamlin, and J. T. Merrill, Transport of ozone precursors from the Arctic troposphere to the North Atlantic region, *J. Geophys. Res.*, 101, 29,335–29,351, 1996.
- Hurrell, J., NAO Index Data provided by the Climate Analysis Section, NCAR, Boulder, USA, <http://www.cdg.ucar.edu/~jhurrell/nao/>, 1995a.

Hurrell, J., Y. Kushnir, G. Ottersen, and M. Visbeck, An overview of the North Atlantic Oscillation, in *The North Atlantic Oscillation: Climatic Significance and Environmental Impact*, edited by J. Hurrell, Y. Kushnir, G. Ottersen, and M. Visbeck, vol. 1–36, American Geophysical Union, Washington, DC, 2003.

Hurrell, J. W., Decadal trends in the North Atlantic Oscillation: regional temperatures and precipitation, *Science*, *269*, 676–679, 1995b.

Hurrell, J. W., North Atlantic Oscillation, in *Encyclopedia of Ocean Sciences*, edited by S. T. J. Steele and K. Turekian, vol. 4, 1904–1911, Academic Press, London, United Kingdom, 2001.

Jain, A. K., and R. C. Dubes, *Algorithms for clustering data*, Prentice Hall, Englewood Cliffs, NJ, 1988.

Jones, P., T. Osborn, and K. Briffa, Pressure-based measure of the North Atlantic Oscillation (NAO): a comparison and an assessment of changes in the strength of the NAO and its influence on surface climate parameters, in *The North Atlantic Oscillation: Climatic Significance and Environmental Impact*, edited by J. Hurrell, Y. Kushnir, G. Ottersen, and M. Visbeck, vol. 51–62, American Geophysical Union, Washington, DC, 2003.

Jonsson, T., and M. W. Miles, Anomalies in the seasonal cycle of sea level pressure in Iceland and the North Atlantic Oscillation, *Geophys. Res. Lett.*, *28*, 4231–4234, 2001.

Kahl, J. D., On the prediction of trajectory model error, *Atmos. Environ.*, *30*, 2945–2957, 1996.

Kahl, J. D., J. M. Harris, G. A. Herbert, and M. P. Olson, Intercomparison of three long-range trajectory models applied to arctic haze, *Tellus*, *41B*, 524–536, 1989.

Kahl, J. D. W., D. A. Martinez, H. Kuhns, C. I. Davidson, J.-L. Jaffrezo, and J. M. Harris, Air mass trajectories to Summit, Greenland: A 44-year climatology and some episodic events, *J. Geophys. Res.*, *102*, 26,861–26,875, 1997.

Kalkstein, L. S., G. Tan, and J. A. Skindlov, An evaluation of three clustering procedures for use in synoptic climatological classification, *American Meteorological Society*, *26*, 717–730, 1987.

Kalnay, E., et al., The NCEP/NCAR 40-year reanalysis project, *Bulletin of the American Meteorological Society*, *77*, 437–471, 1996.

Kaufman, L., and P. Rousseeuw, *Finding groups in data: An introduction to cluster analysis*, John Wiley and Sons, Inc., New York, NY, 1990.

Kistler, R., et al., The NCEP/NCAR 50-year reanalysis, *Tech. rep.*, 1999.

Krupa, S., and H. Jager, Adverse effects of elevated levels of ultraviolet (UV)-B radiation and ozone (O_3) on crop growth and productivity, in *Global climate change and agricultural production*, edited by M. Treshow, vol. 141–169, John Wiley and Sons, New York, NY, 1996.

Lee, G., J. T. Merrill, and B. J. Huebert, Variation of free tropospheric total nitrate at Mauna Loa Observatory, Hawaii, *J. Geophys. Res.*, *99*, 12,821–12,831, 1994.

Lin, C., M. Cheng, and W. Schroeder, Transport patterns and potential sources of total gaseous mercury measured in Canadian high Arctic in 1995, *Atmos. Environ.*, *35*, 1141–1154, 2001.

Lin, C.-Y. C., D. J. Jacob, J. W. Munger, and A. M. Fiore, Increasing background ozone in surface air over the United States, *Geophys. Res. Lett.*, *27*, 3465–3468, 2000.

Marshall, J., et al., North Atlantic climate variability: Phenomena, impacts and mechanisms, *International Journal of Climatology*, *21*, 1863–1898, 2001.

McQueen, J. T., and R. R. Draxler, Evaluation of model back trajectories of the Kuwaiti oil fires using digital satellite data, *Atmos. Environ.*, *28*, 2159–2174, 1994.

Merrill, J. T., Isentropic air-flow probability analysis, *J. Geophys. Res.*, *99*, 25,881–25,889, 1994.

Merrill, J. T., and J. L. Moody, Synoptic meteorology and transport during the North Atlantic Regional Experiment (NARE) intensive: overview, *J. Geophys. Res.*, *101*, 28,903–28,921, 1996.

Merrill, J. T., R. Bleck, and L. Avila, Modeling atmospheric transport to the Marshall Islands, *J. Geophys. Res.*, *90*, 12,927–12,936, 1985.

Merrill, J. T., M. Uematsu, and R. Bleck, Meteorological analysis of long range transport of mineral aerosols over the North Pacific, *J. Geophys. Res.*, *94*, 8584–8598, 1989.

Meyer, W. B., *Human impact on the earth*, Cambridge University Press, Melbourne, Australia, 1996.

Miller, J., A five-year climatology of five-day back trajectories from Barrow, Alaska, *Atmos. Environ.*, *15*, 1401–1405, 1981.

Moody, J. L., and J. N. Galloway, Quantifying the relationship between atmospheric transport and the chemical composition of precipitation on Bermuda, *Tellus*, *40B*, 463–479, 1988.

Moody, J. L., S. J. Oltmans, H. L. II, and J. T. Merrill, Transport climatology of tropospheric ozone: Bermuda, 1988–1991, *J. Geophys. Res.*, *100*, 7179–7194, 1995.

Moser, T., J. Barker, and D. Tingey, Anthropogenic contaminants: atmospheric transport, deposition, and potential effects on terrestrial ecosystems, in *The science of global change*, edited by D. Dunnette and R. O'Brien, vol. 134–148, American Chemical Society, Salem, MA, 1992.

Moy, L. A., R. R. Dickerson, and W. F. Ryan, Relationship between back trajectories and tropospheric trace gas concentrations in rural Virginia, *Atmos. Environ.*, *28*, 2789–2800, 1994.

Nizich, S. V., A. A. Pope, S. Bromberg, L. Tooly, L. B. Kuykenda, and D. Misenheimerl, National air pollutant emission trends, 1900–1998, *Tech. Rep. EPA-45/R-00-002*, U. S. Environmental Protection Agency, 2000.

NOAA, HYSPLIT4 model (HYbrid Single-Particle Lagrangian Integrated Trajectory) model, Web address: <http://www.arl.noaa.gov/ready/hysplit4.html>, NOAA Air Resources Laboratory, Silver Spring, Maryland, 1997.

O'Brien, R., The atmospheric component, in *The science of global change*, edited by D. Dunnette and R. O'Brien, vol. 21–23, American Chemical Society, Salem, MA, 1992.

Ottar, B., H. Dovland, and A. Semb, Long range transport of air pollutants and acid precipitation, in *Air pollution and plant life*, edited by M. Treshow, vol. 39–71, John Wiley and Sons, New York, NY, 1984.

Parrish, D., The North Atlantic Regional Experiment (NARE), *IGACtivities*, *24*, 2, 2001.

Parrish, D. D., J. S. Holloway, M. Trainer, P. C. Murphy, G. L. Forbes, and F. C. Fehsenfeld, Export of North American ozone pollution to the North Atlantic Ocean, *Science*, *259*, 1436–1439, 1993.

Parrish, D. D., M. Trainer, J. S. Holloway, J. E. Yee, M. S. Warshawsky, F. C. Fehsenfeld, G. Forbes, and J. L. Moody, Relationships between ozone and carbon monoxide at surface sites in the North Atlantic region, *J. Geophys. Res.*, *103*, 13,357–13,376, 1998.

Penkett, S., A. Volz-Thomas, D. Parrish, R. Honrath, and F. Fenssenfeld, Preface: NARE II, *J. Geophys. Res.*, *103*, 13,353–13,355, 1998.

Peterson, M., A. Hamlin, and M. V. Martín, Transport of anthropogenic pollutants within the marine boundary layer of the Central and Nothwestern North Atlantic, *IGACtivities*, *24*, 5–6, 2001.

Peterson, M. C., R. E. Honrath, D. D. Parrish, and S. J. Oltmans, Measurements of nitrogen oxides and a simple model of NO_y fate in the remote North Atlantic marine atmosphere, *J. Geophys. Res.*, *103*, 13,489–13,503, 1998.

Poirot, R. L., and P. R. Wishinski, Visibility, sulfate and air mass history associated with the summertime aerosol in Northern Vermont, *Atmos. Environ.*, *20*, 1457–1469, 1986.

- Prospero, J., and S. Jennings, The Atmosphere/Ocean Chemistry Experiment, *IGACTivities*.
- Roberts, J. M., et al., Episodic removal of NO_y species from the marine boundary layer over the North Atlantic, *J. Geophys. Res.*, *101*, 28,947–28,960, 1996.
- Sirois, A., and J. W. Bottenheim, Use of backward trajectories to interpret the 5-year record of PAN and O_3 ambient air concentrations at Kejimikujik National Park, Nova Scotia, *J. Geophys. Res.*, *100*, 2867–2881, 1995.
- Stenseth, N., A. Mysterand, G. Ottersen, J. W. Hurrell, K. Chan, and M. Lima, Ecological effects of climate fluctuations, *Science*, *297*, 1292–1296, 2002.
- Stephenson, D., H. Wanner, S. Brönnimann, and J. Luterbacher, The history of scientific research on the North Atlantic Oscillation, in *The North Atlantic Oscillation: Climatic Significance and Environmental Impact*, edited by J. Hurrell, Y. Kushnir, G. Ottersen, and M. Visbeck, vol. 37–50, American Geophysical Union, Washington, DC, 2003.
- Stohl, A., Computation, accuracy and applications of trajectories—a review and bibliography, *Atmos. Environ.*, *32*, 947–966, 1998.
- Stohl, A., G. Wotawa, P. Seibert, and H. Kromp-Klob, Interpolation errors in wind fields as a function of spatial and temporal resolution and their impact on different types of kinematic trajectories, *Journal of Applied Meteorology*, *34*, 2149–2165, 1995.
- Stohl, A., S. Eckhardt, C. Foster, P. James, N. Spichtinger, and P. Seibert, A replacement for simple back trajectory calculations in the interpretation of atmospheric trace substance measurements, *Atmos. Environ.*, *36*, 4635–4648, 2002a.
- Stohl, A., M. Trainer, T. B. Ryerson, J. S. Holloway, and D. D. Parrish, Export of NO_y from the North American boundary layer during 1996 and 1997 North Atlantic Regional Experiments, *J. Geophys. Res.*, *107*, 10.1029/2001JD000519, 2002b.
- Trenberth, K. E., and D. P. Stepaniak, A pathological with NCEP reanalysis in the stratosphere, *J. Climate*, *15*, 690–695, 2002.
- USEPA, Air quality index: A guide to air quality and your health, *Tech. Rep. EPA 454/K-02-001*, USEPA, 2000.
- USEPA, Latest findings in national air quality standards: 2001 status and trends, *Tech. Rep. EPA 454/K-02-001*, USEPA, 2002.
- Vaughan, J. M., R. H. Maryon, and N. J. Geddes, Comparison of atmospheric aerosol backscattering and air mass back trajectories, *Meteorology Atmospheric Physics*, *79*, 33–46, 2002.
- Visbeck, M. H., J. W. Hurrell, L. Polvani, and H. Cullen, The North Atlantic Oscillation: past, present, and future, *PNAS*, *98*, 12,876–12,877, 2001.

Wanner, H., S. Bronnimann, C. Casty, D. Gyalistras, J. Luterbacher, C. Schmutz, D. Stephenson, and E. Xoplaki, North Atlantic Oscillation-concepts and studies, *Surveys in Geophysics*, 22, 321–382, 2001.

Appendix A

Membership plots for whole period

A.1 Membership plots for the average period

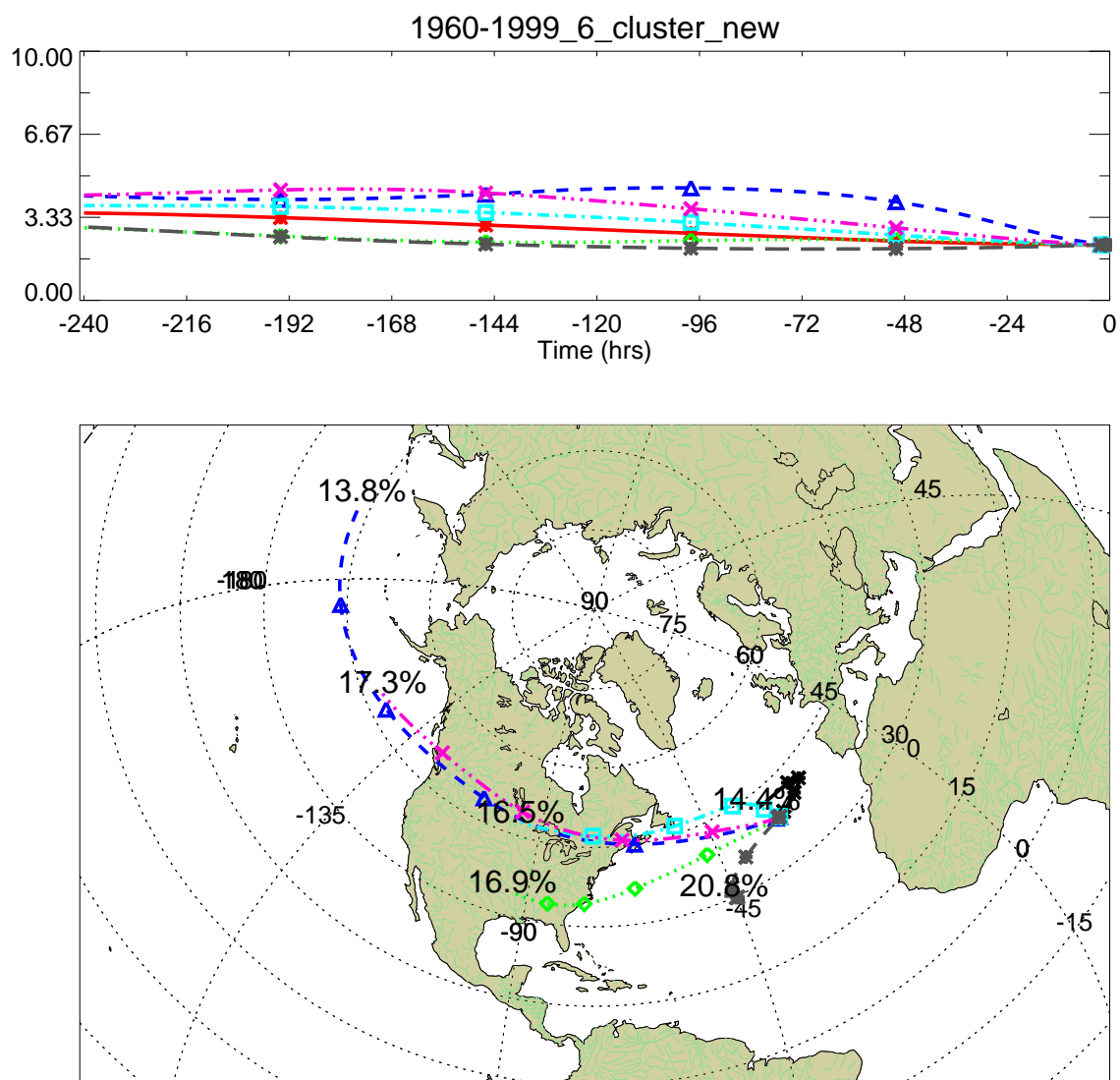


Figure A.1 Cluster plot for the average period, 1960-1999.

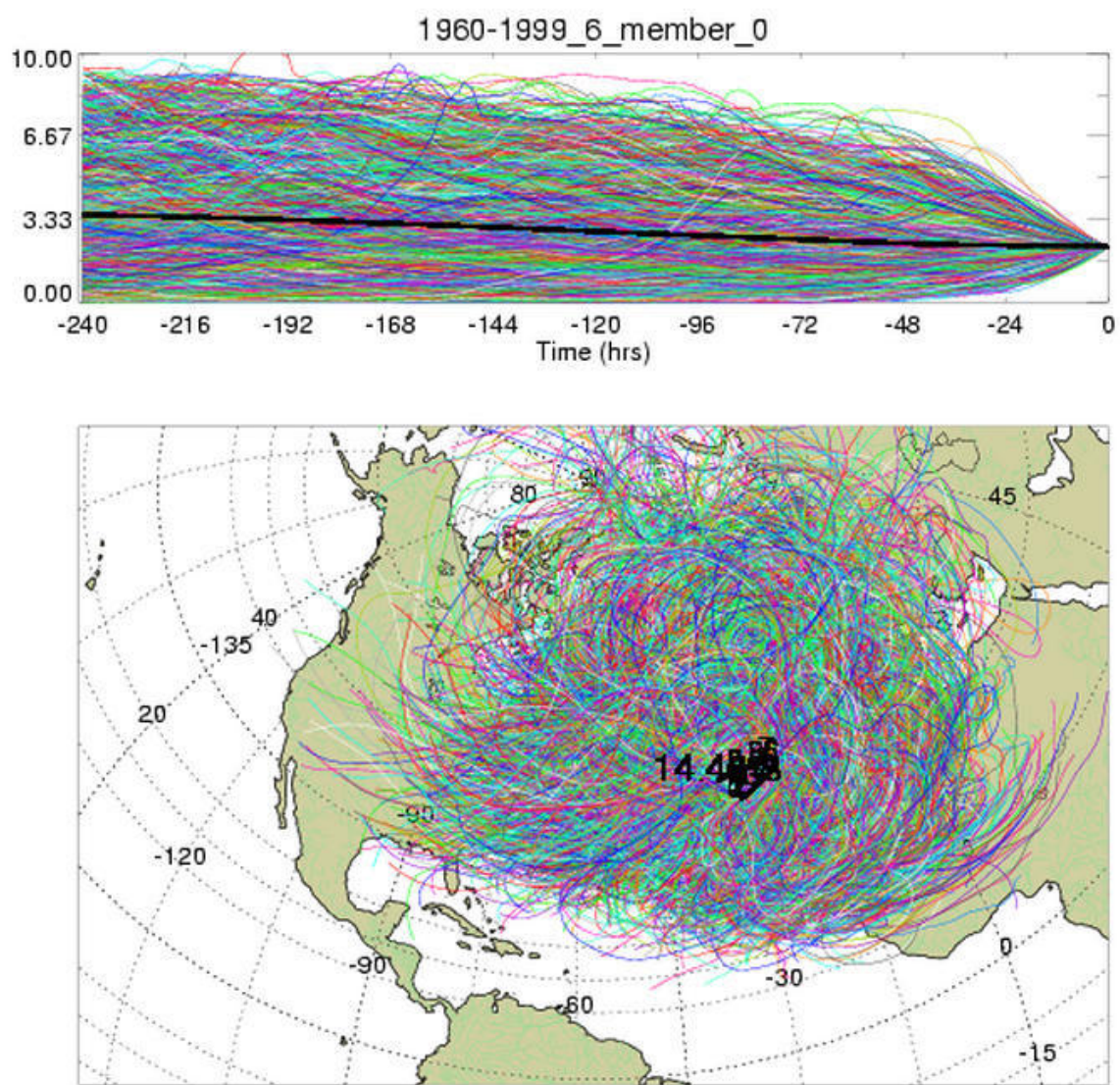


Figure A.2 Membership plot, 1960-1999.

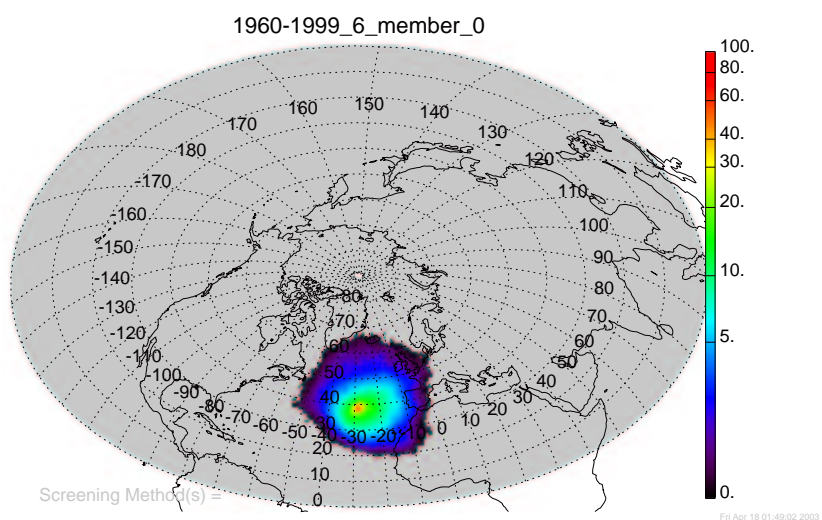


Figure A.3 Standard density membership plot, 1960-1999.

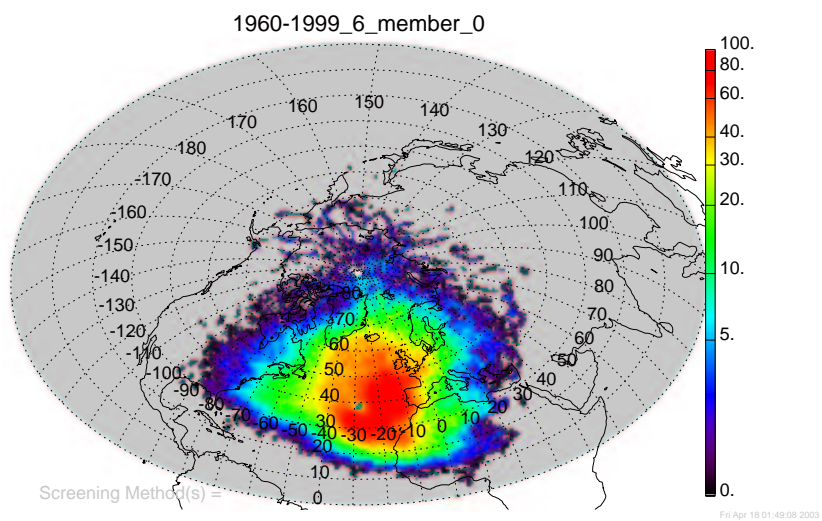


Figure A.4 Geometrically corrected density membership plot, 1960-1999.

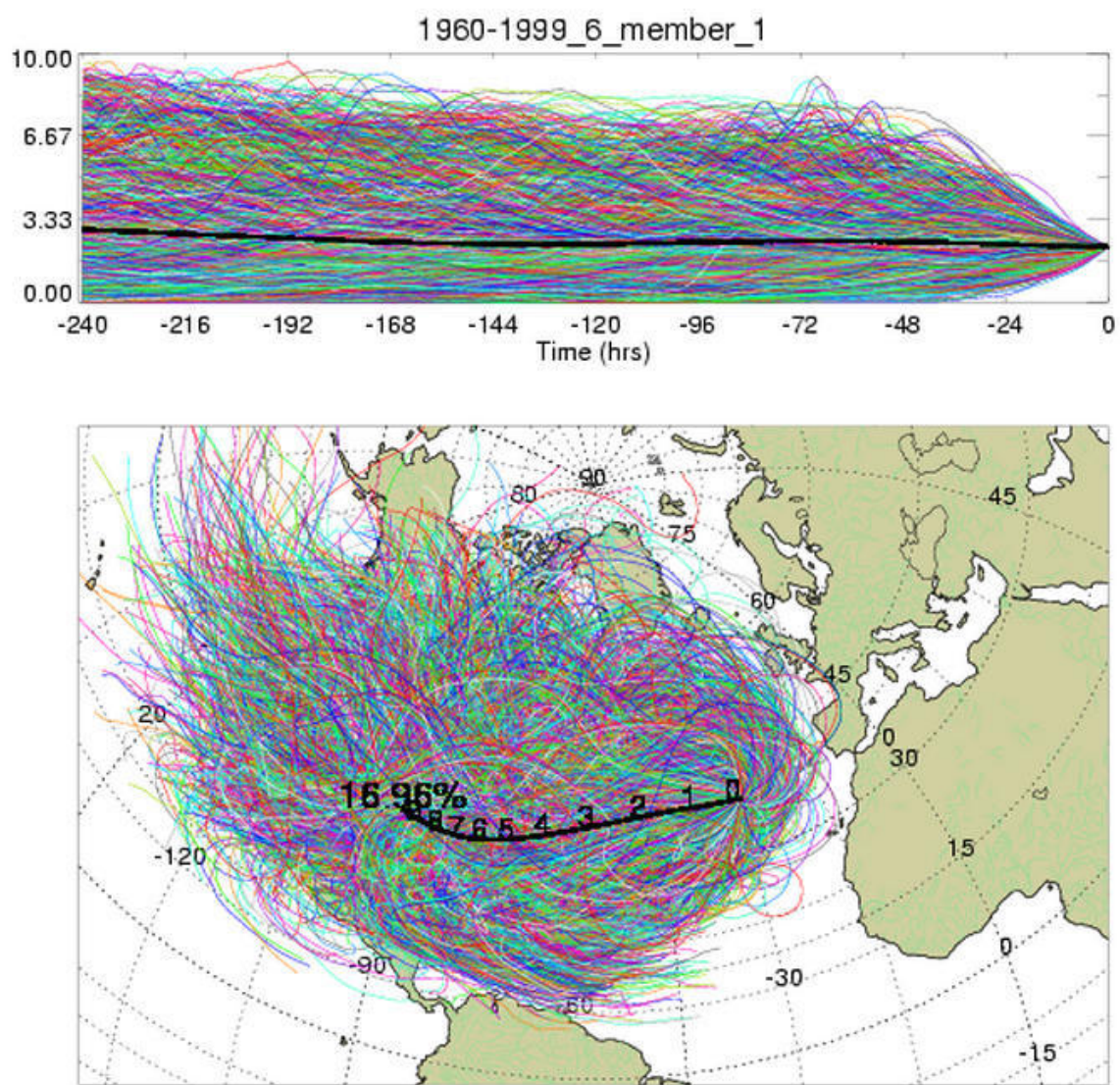


Figure A.5 Membership plot, 1960-1999.

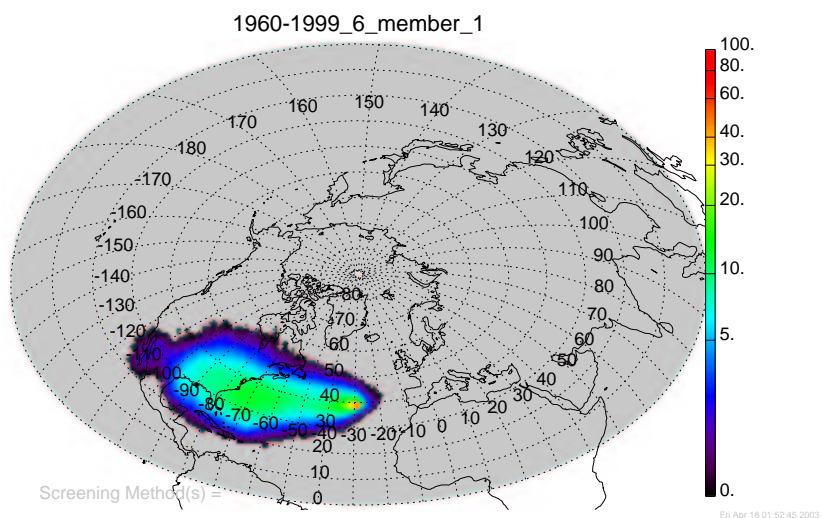


Figure A.6 Standard density membership plot, 1960-1999.

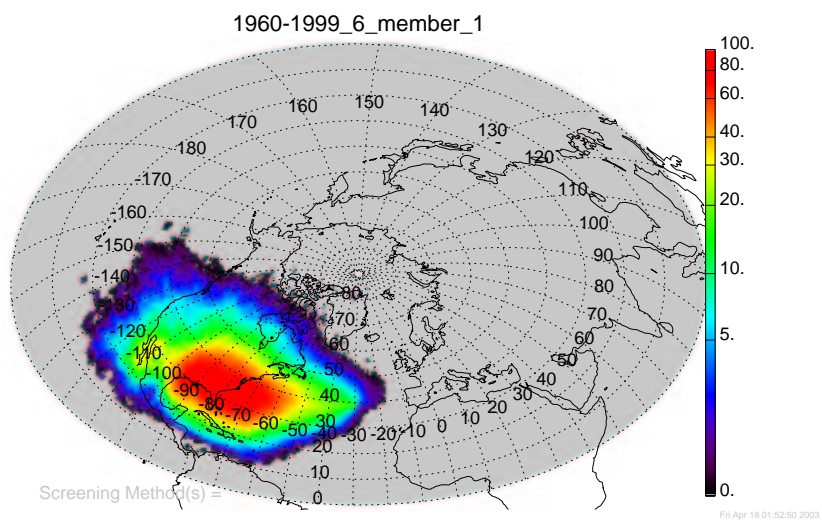


Figure A.7 Geometrically corrected density membership plot, 1960-1999.

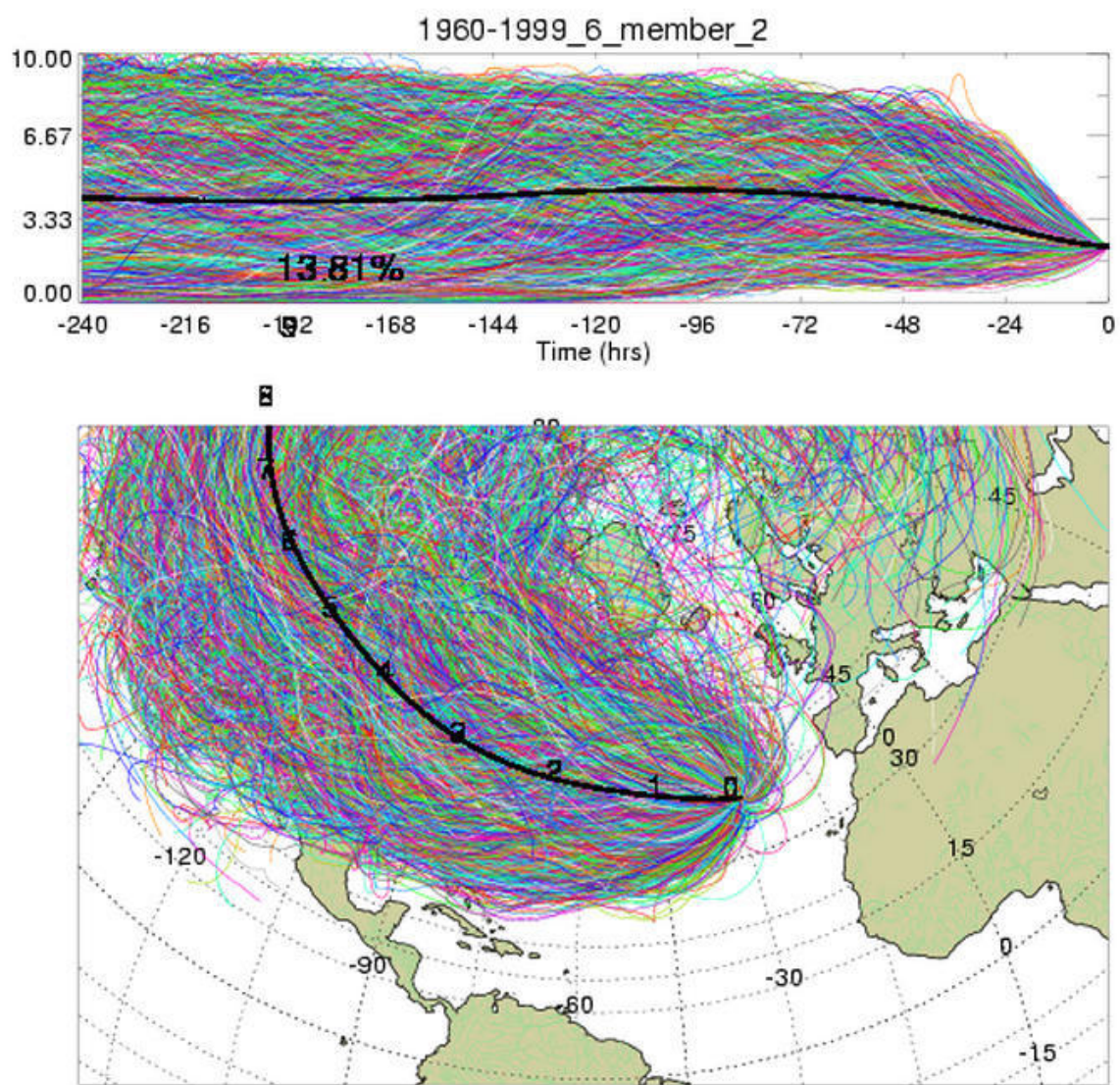


Figure A.8 Membership plot, 1960-1999.

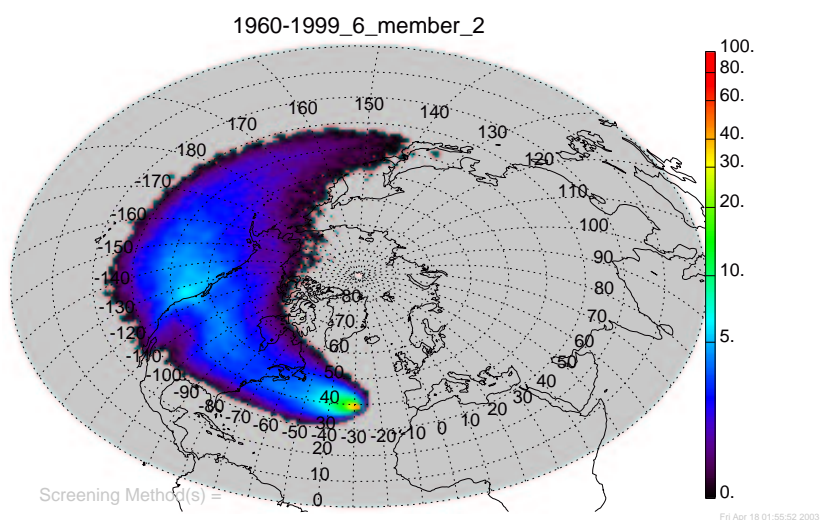


Figure A.9 Standard density membership plot, 1960-1999.

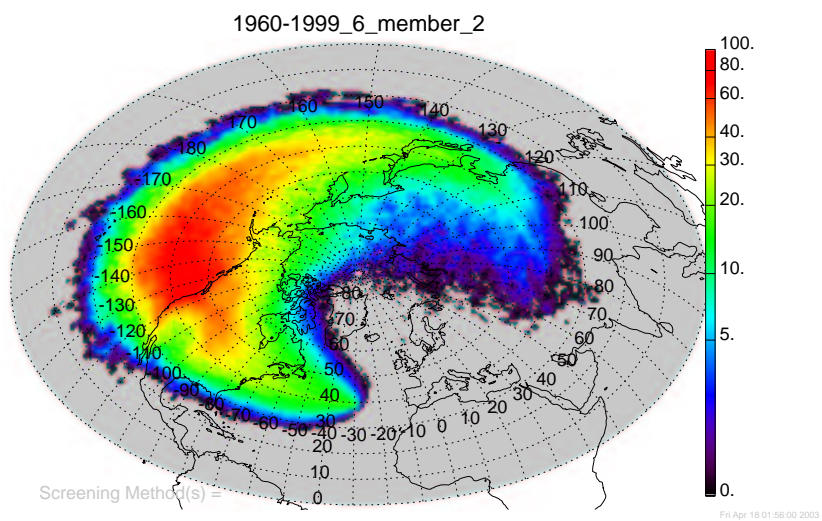


Figure A.10 Geometrically corrected density membership plot, 1960-1999.

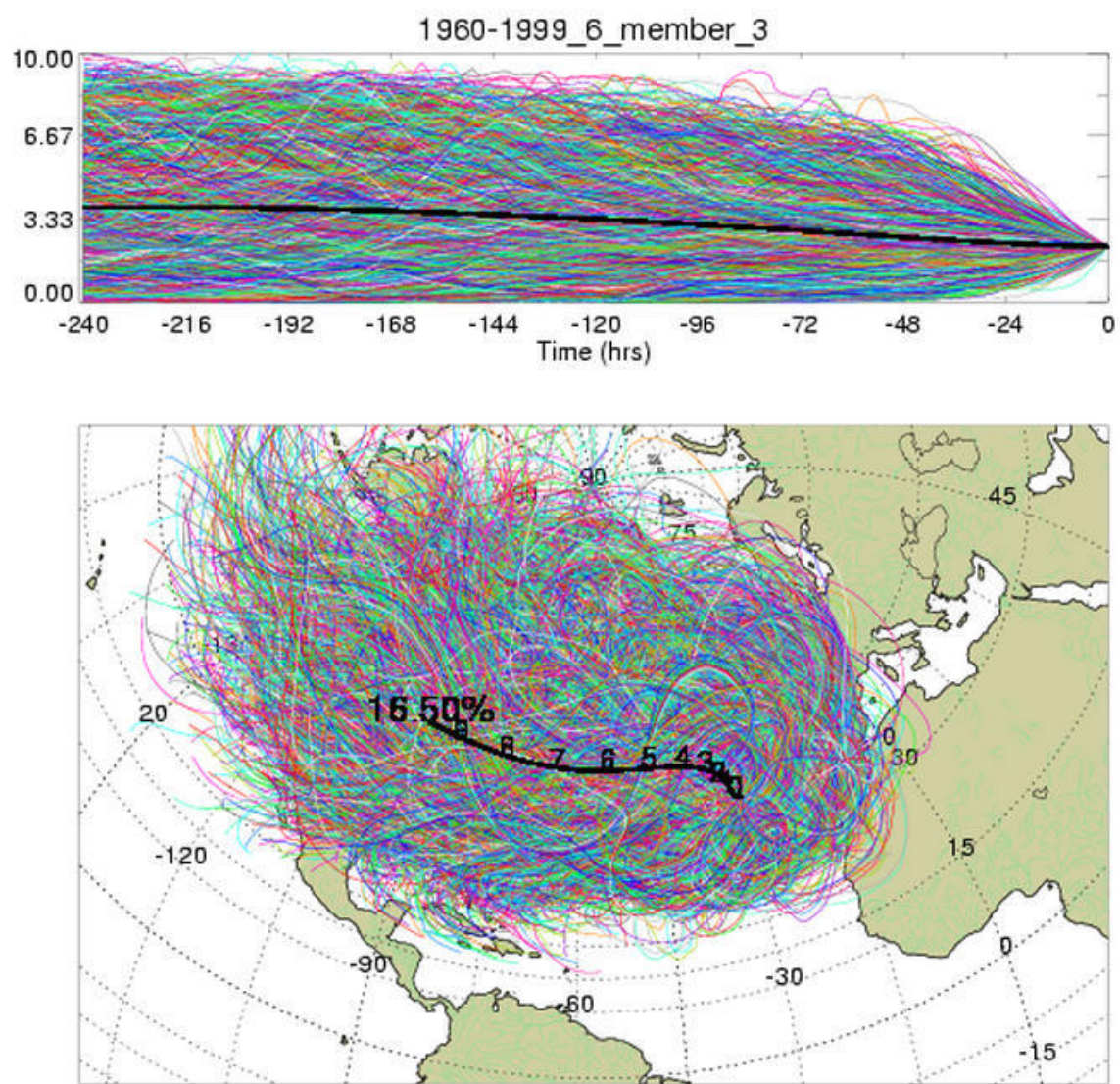


Figure A.11 Membership plot, 1960-1999.

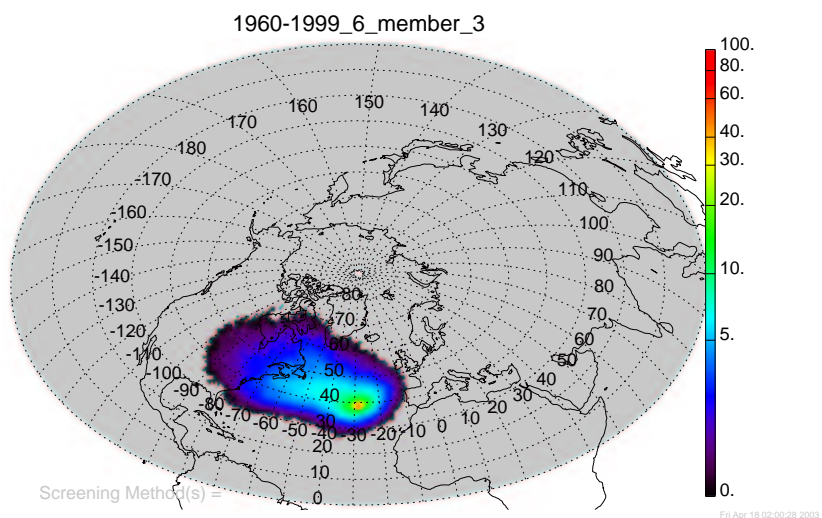


Figure A.12 Standard density membership plot, 1960-1999.

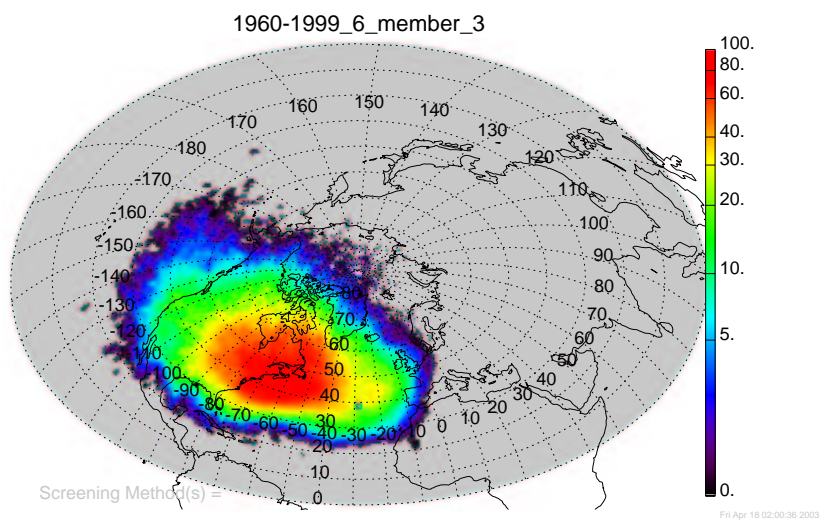


Figure A.13 Geometrically corrected density membership plot, 1960-1999.

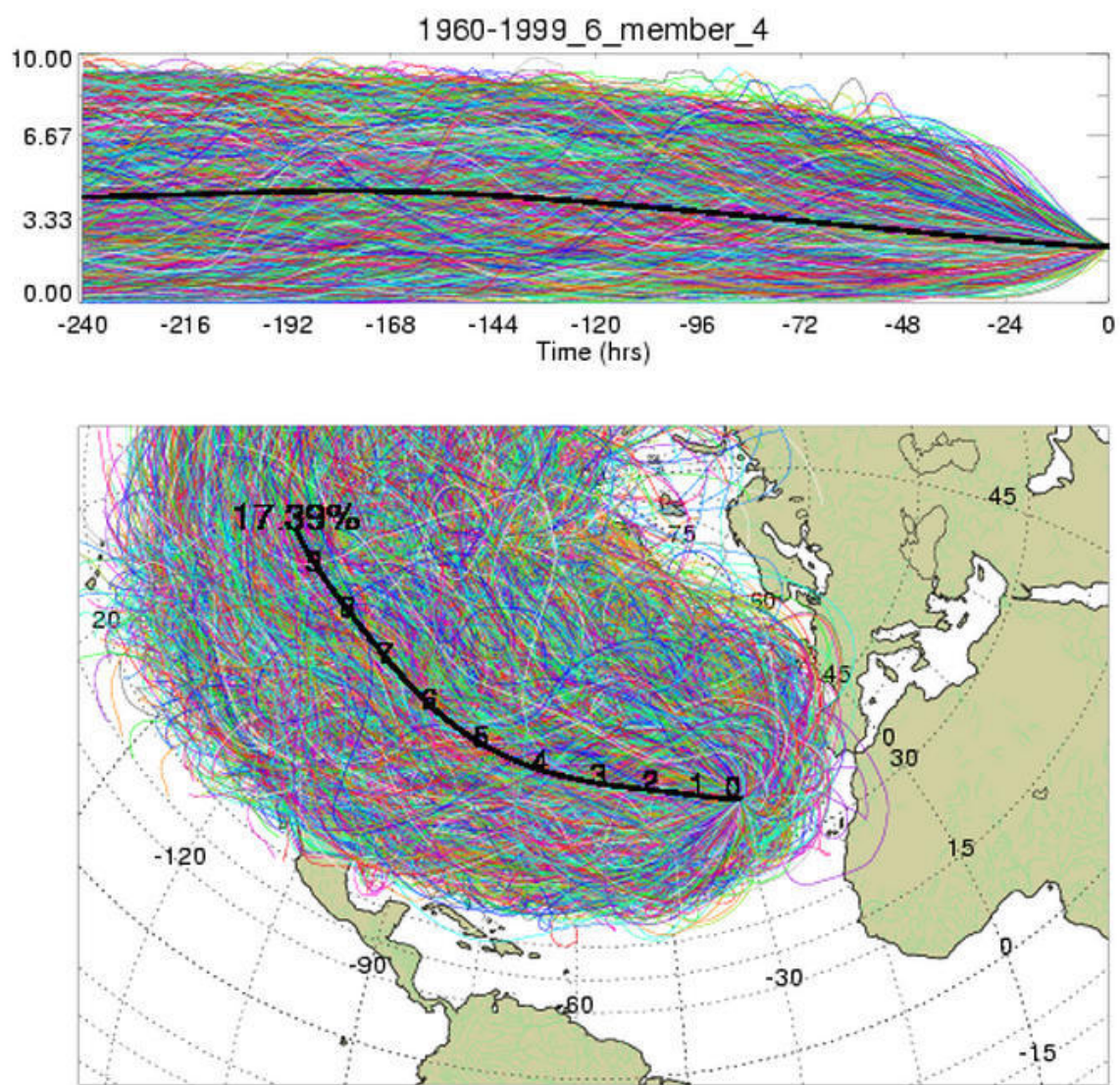


Figure A.14 Membership plot, 1960-1999.

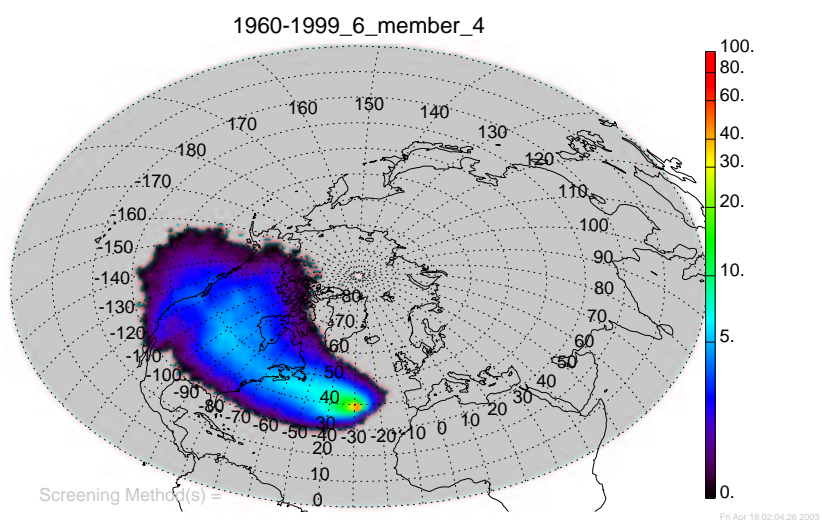


Figure A.15 Standard density membership plot, 1960-1999.

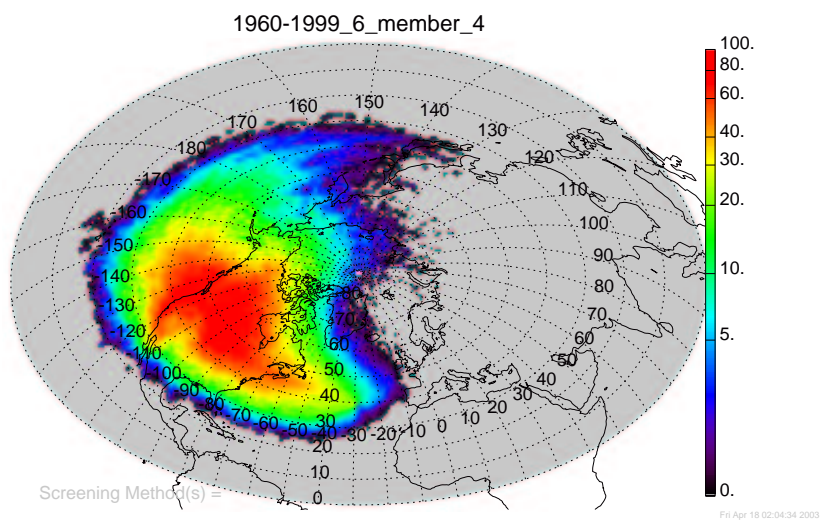


Figure A.16 Geometrically corrected density membership plot, 1960-1999.

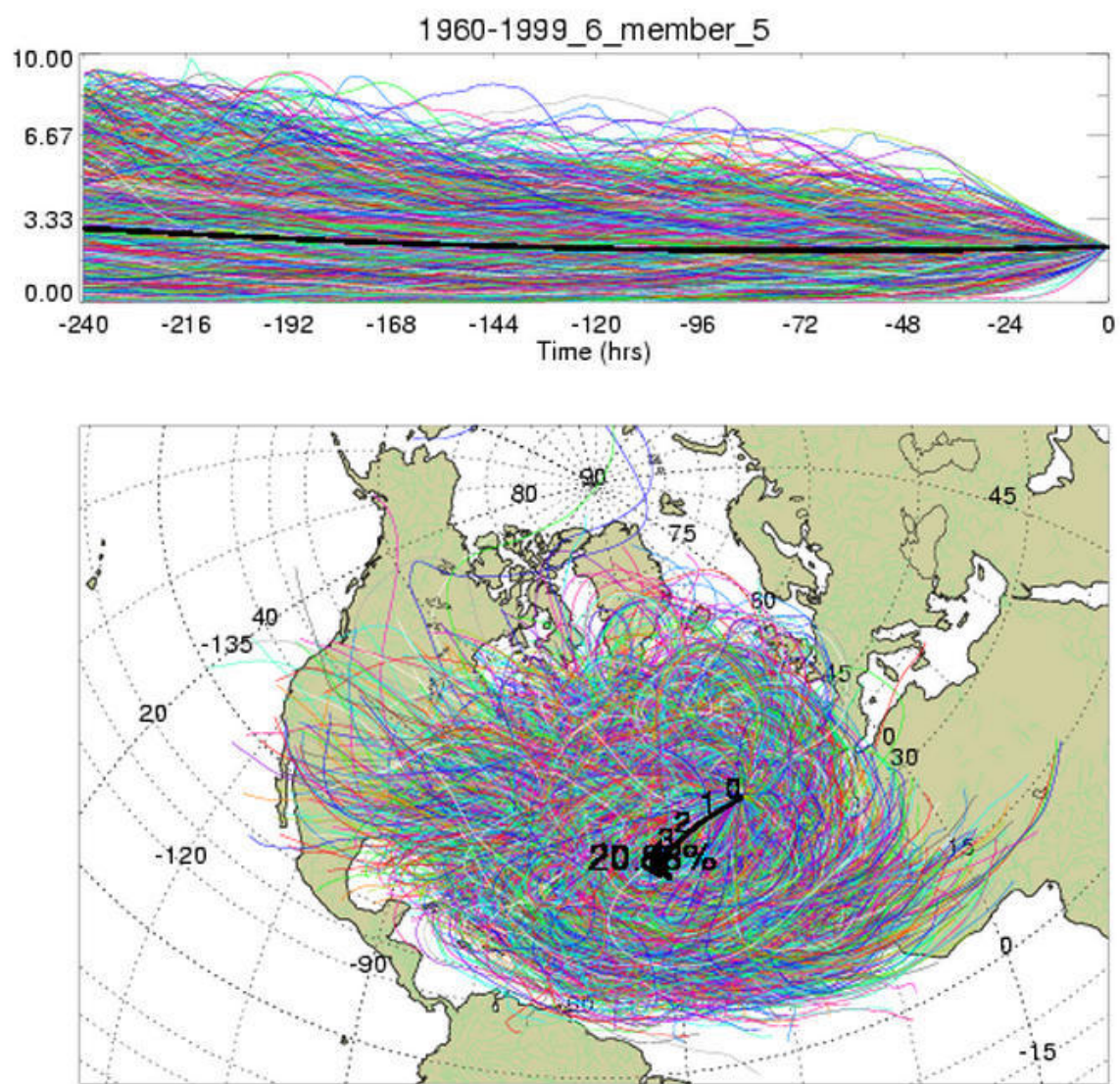


Figure A.17 Membership plot, 1960-1999.

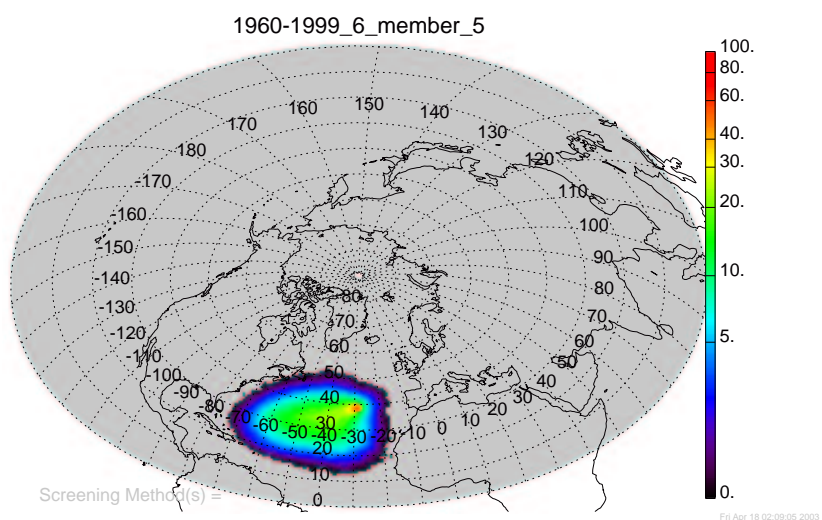


Figure A.18 Standard density membership plot, 1960-1999.

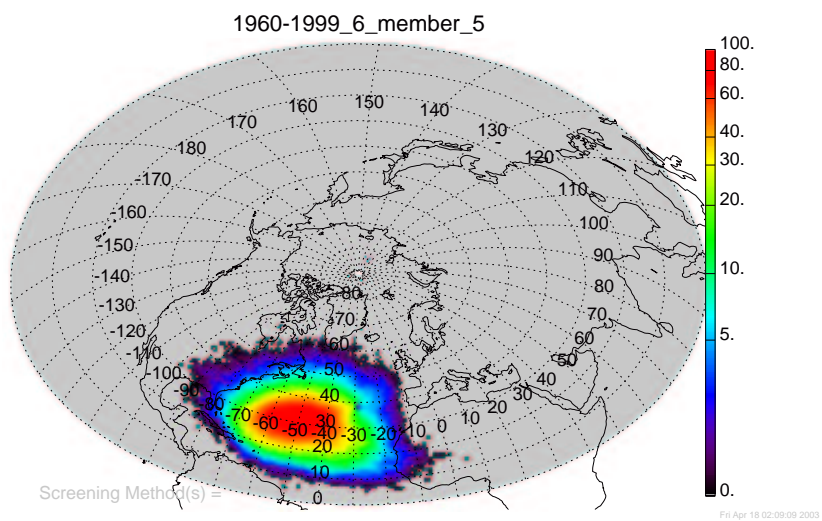


Figure A.19 Geometrically corrected density membership plot, 1960-1999.

Appendix B

Membership plots for winter periods

B.1 Membership plots for the average winter

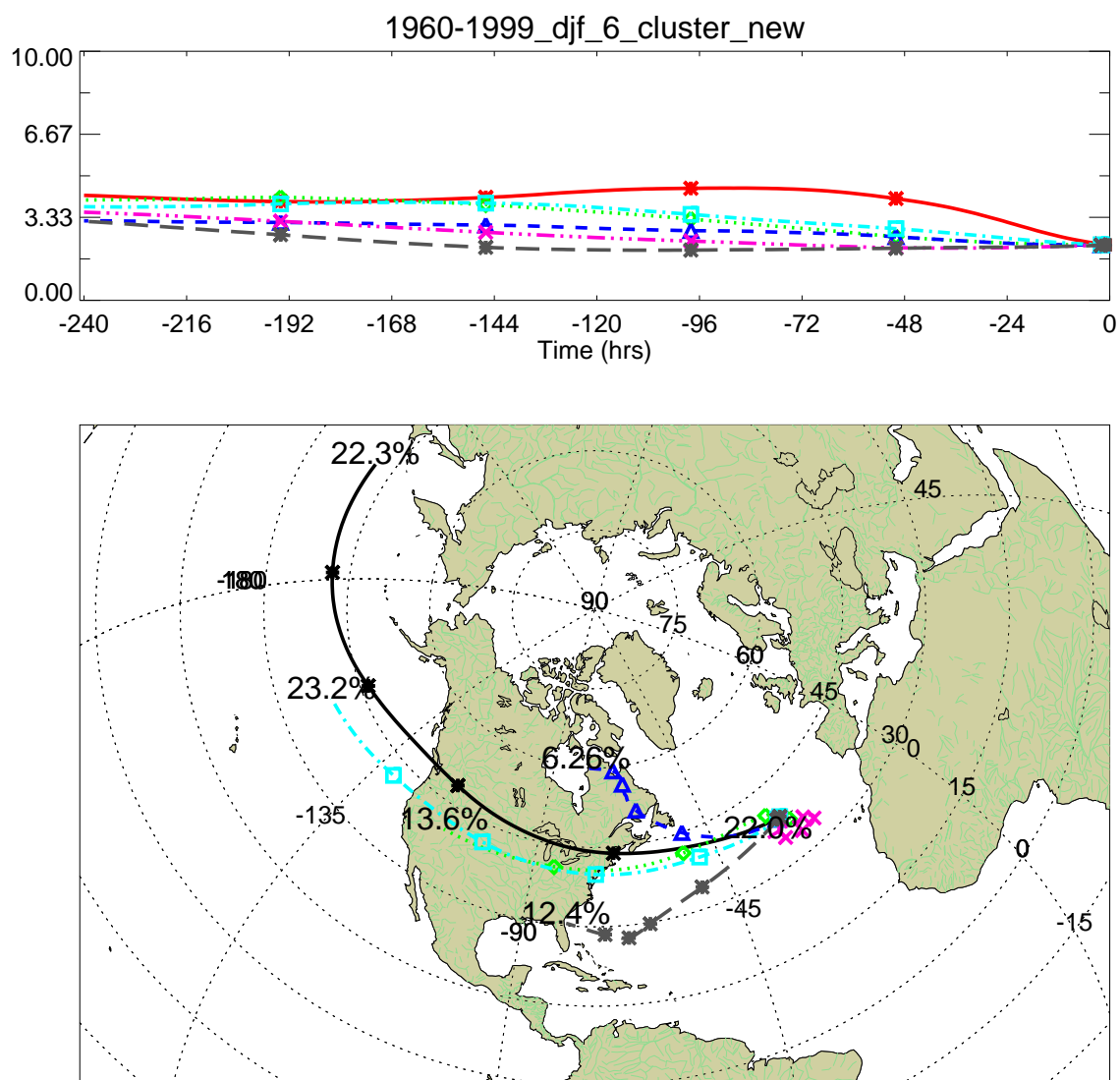


Figure B.1 Cluster plot for the average winters, 1960-1999.

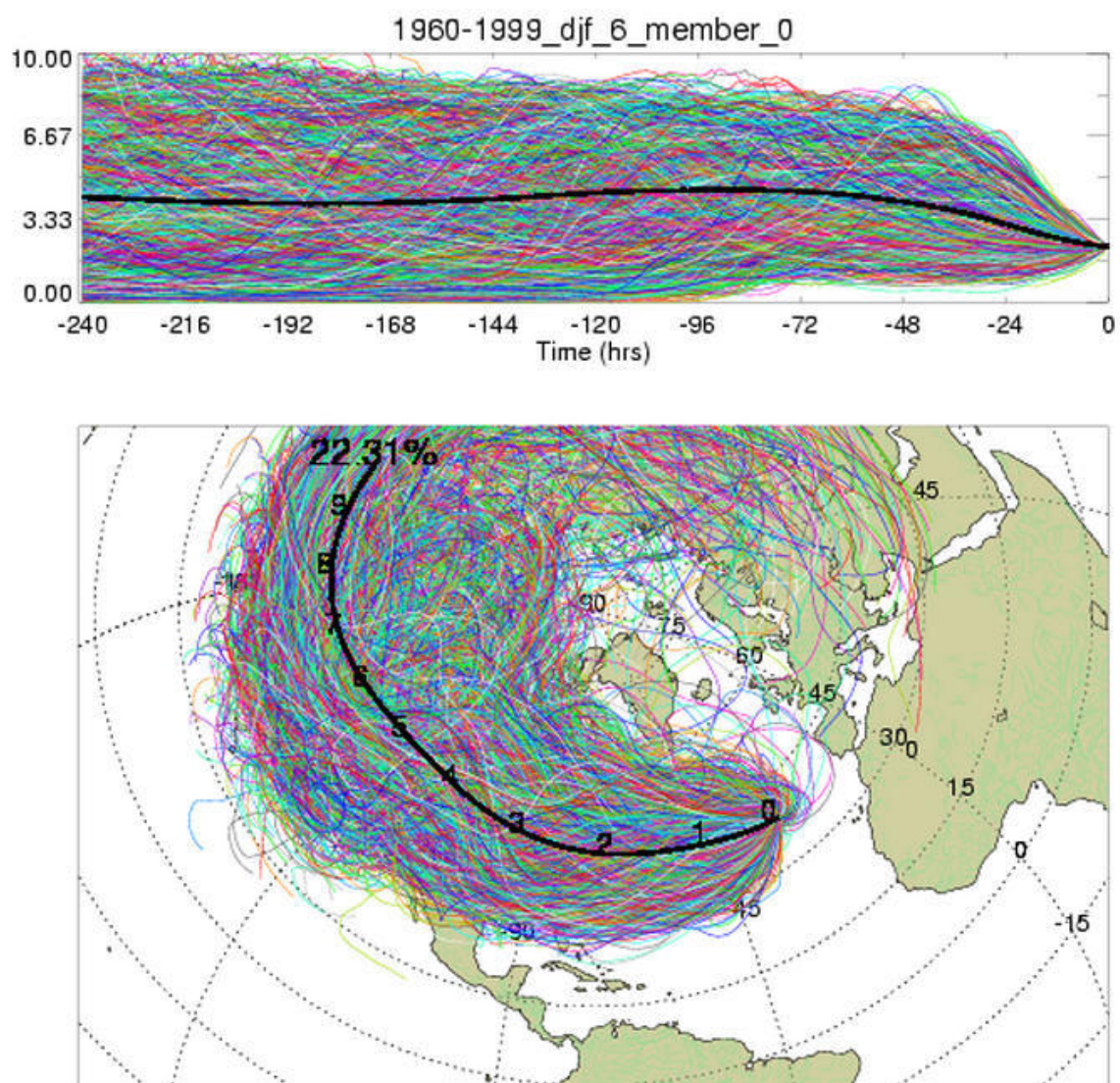


Figure B.2 Membership plot for winters, 1960-1999.

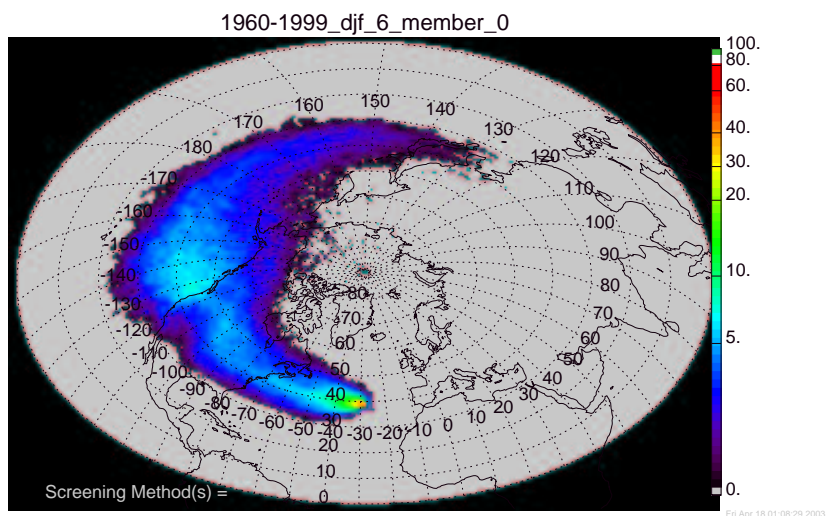


Figure B.3 Standard density membership plot for winters, 1960-1999.

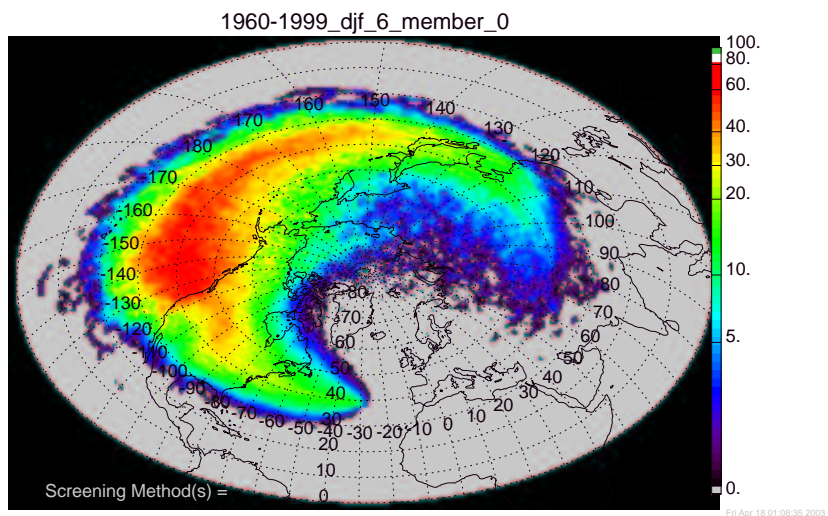


Figure B.4 Geometrically corrected density membership plot for winters, 1960-1999.

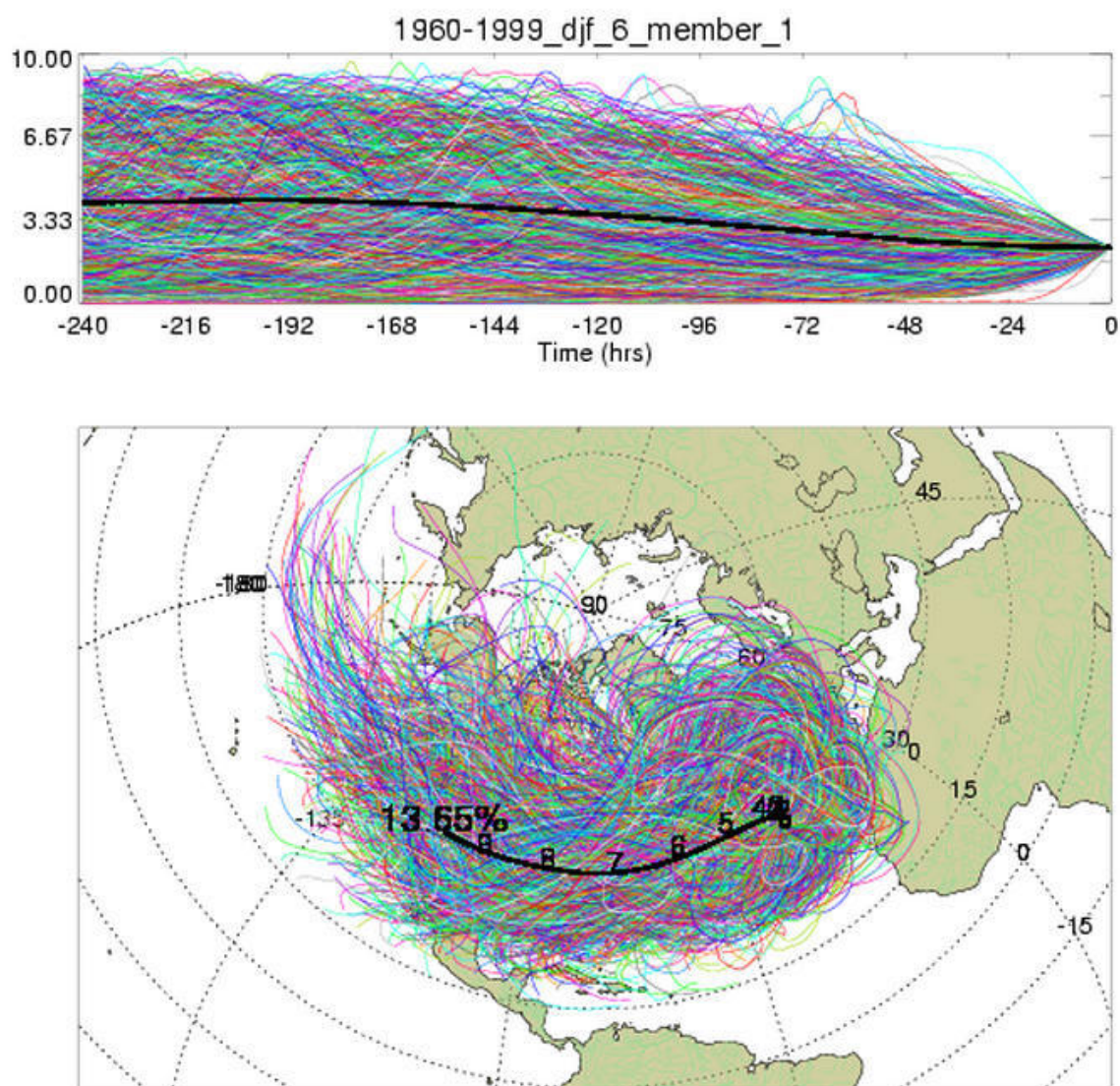


Figure B.5 Membership plot for winters, 1960-1999.

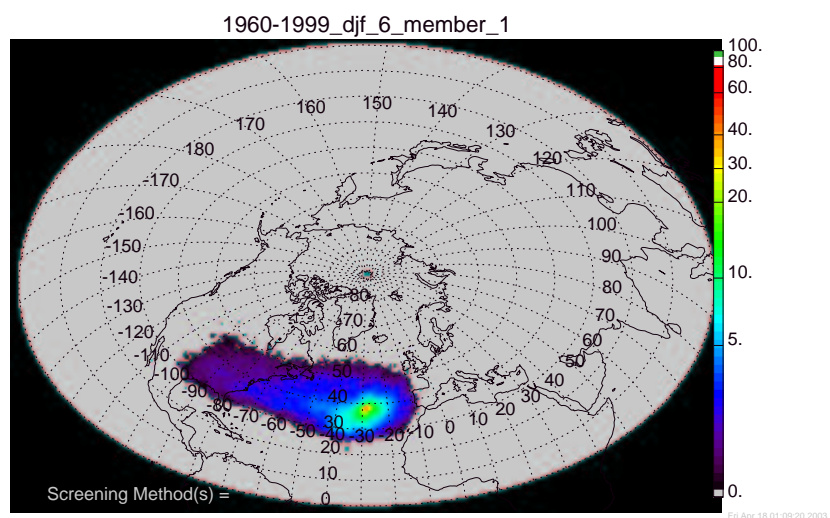


Figure B.6 Standard density membership plot for winters, 1960-1999.

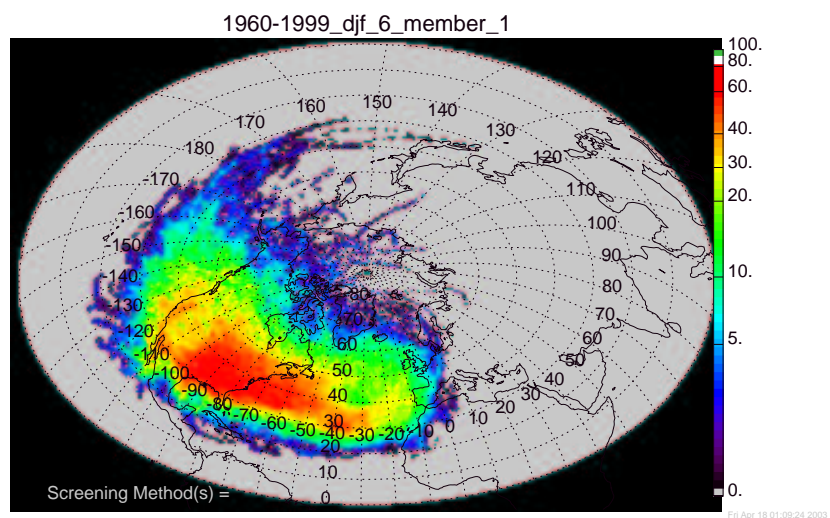


Figure B.7 Geometrically corrected density membership plot for winters, 1960-1999.

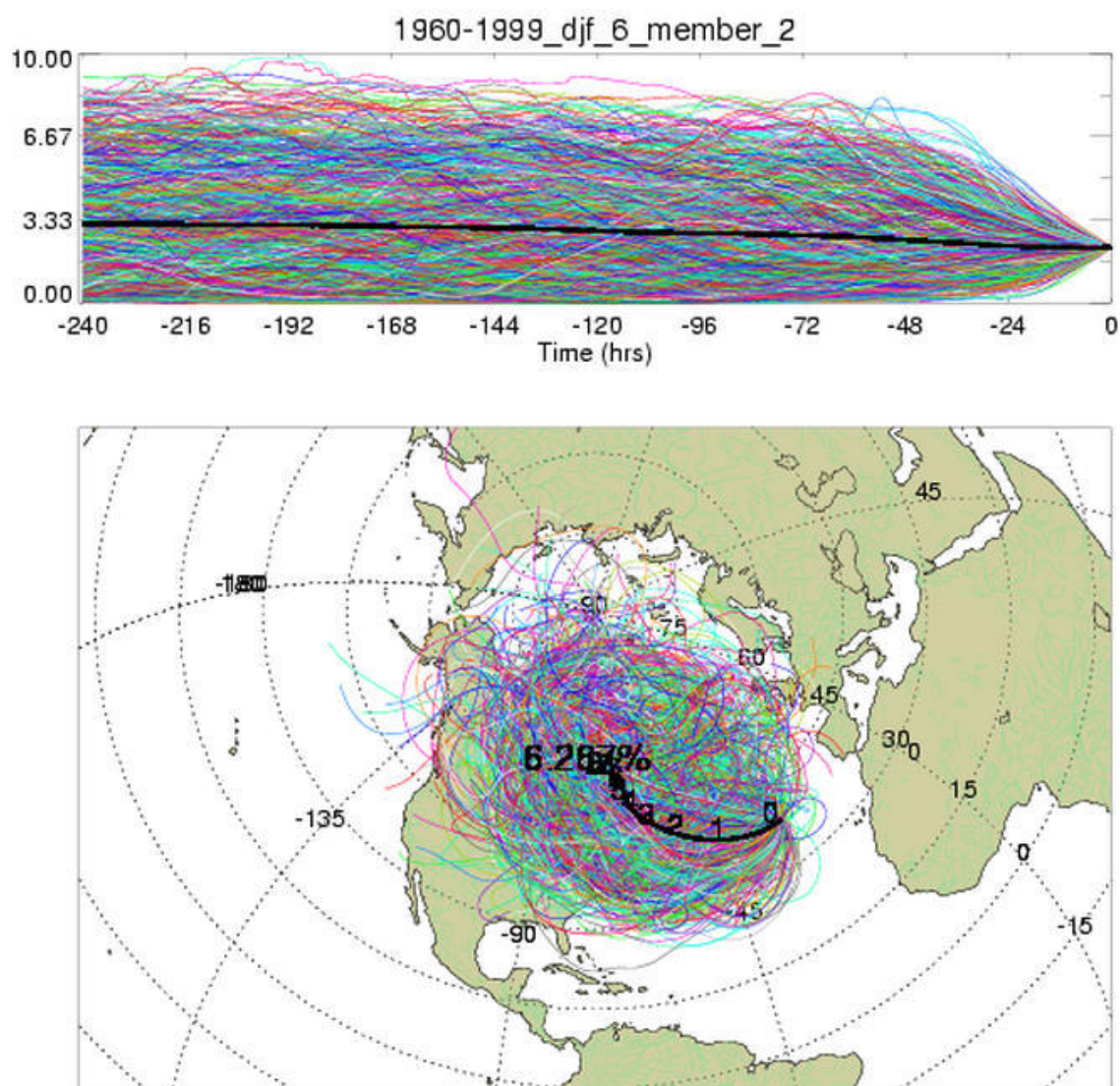


Figure B.8 Membership plot for winters, 1960-1999.

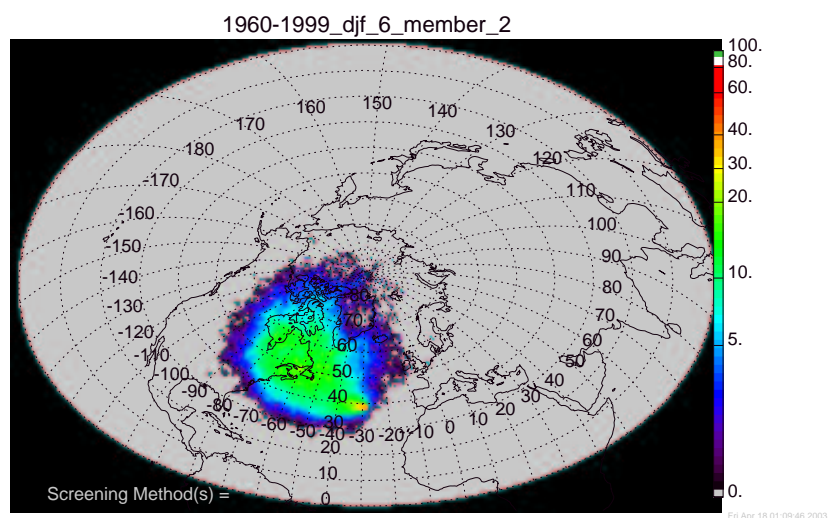


Figure B.9 Standard density membership plot for winters, 1960-1999.

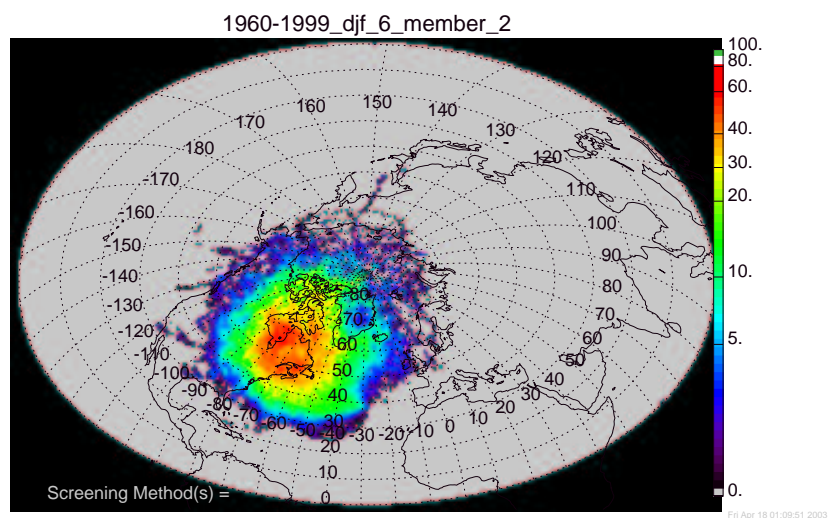


Figure B.10 Geometrically corrected density membership plot for winters, 1960-1999.

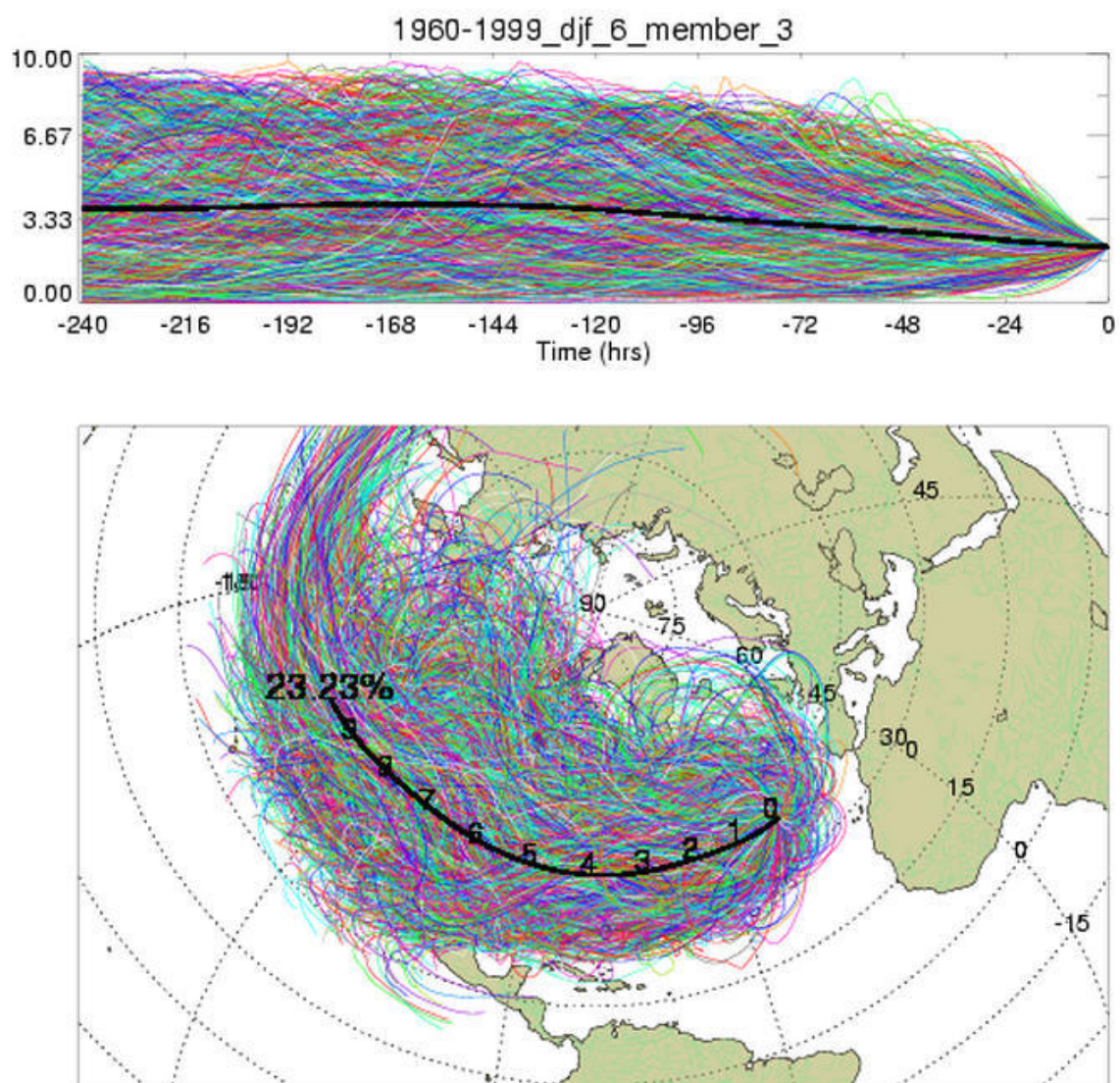


Figure B.11 Membership plot for winters, 1960-1999.

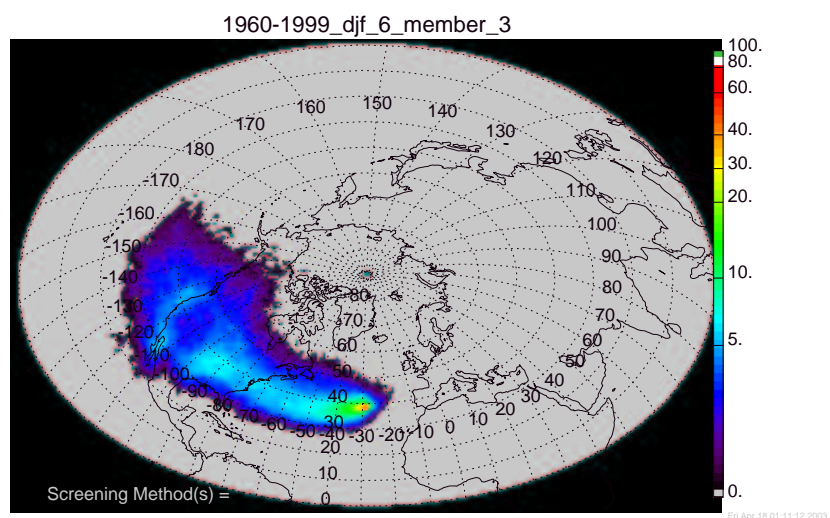


Figure B.12 Standard density membership plot for winters, 1960-1999.

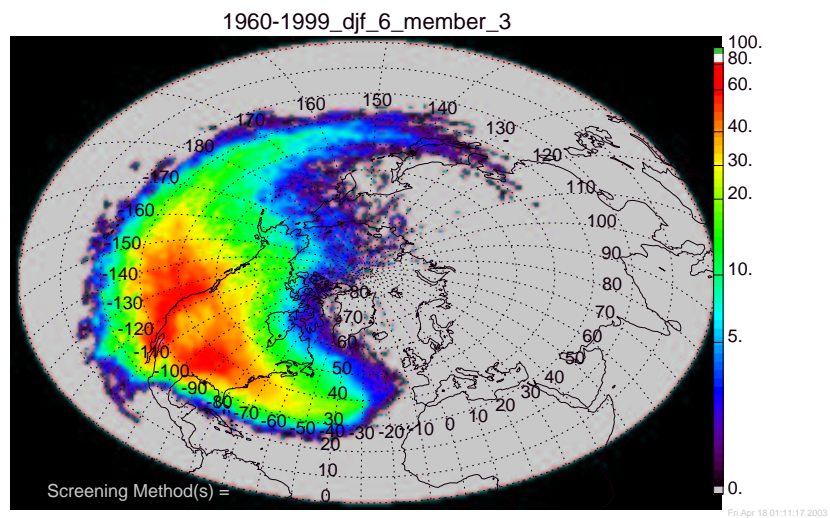


Figure B.13 Geometrically corrected density membership plot for winters, 1960-1999.

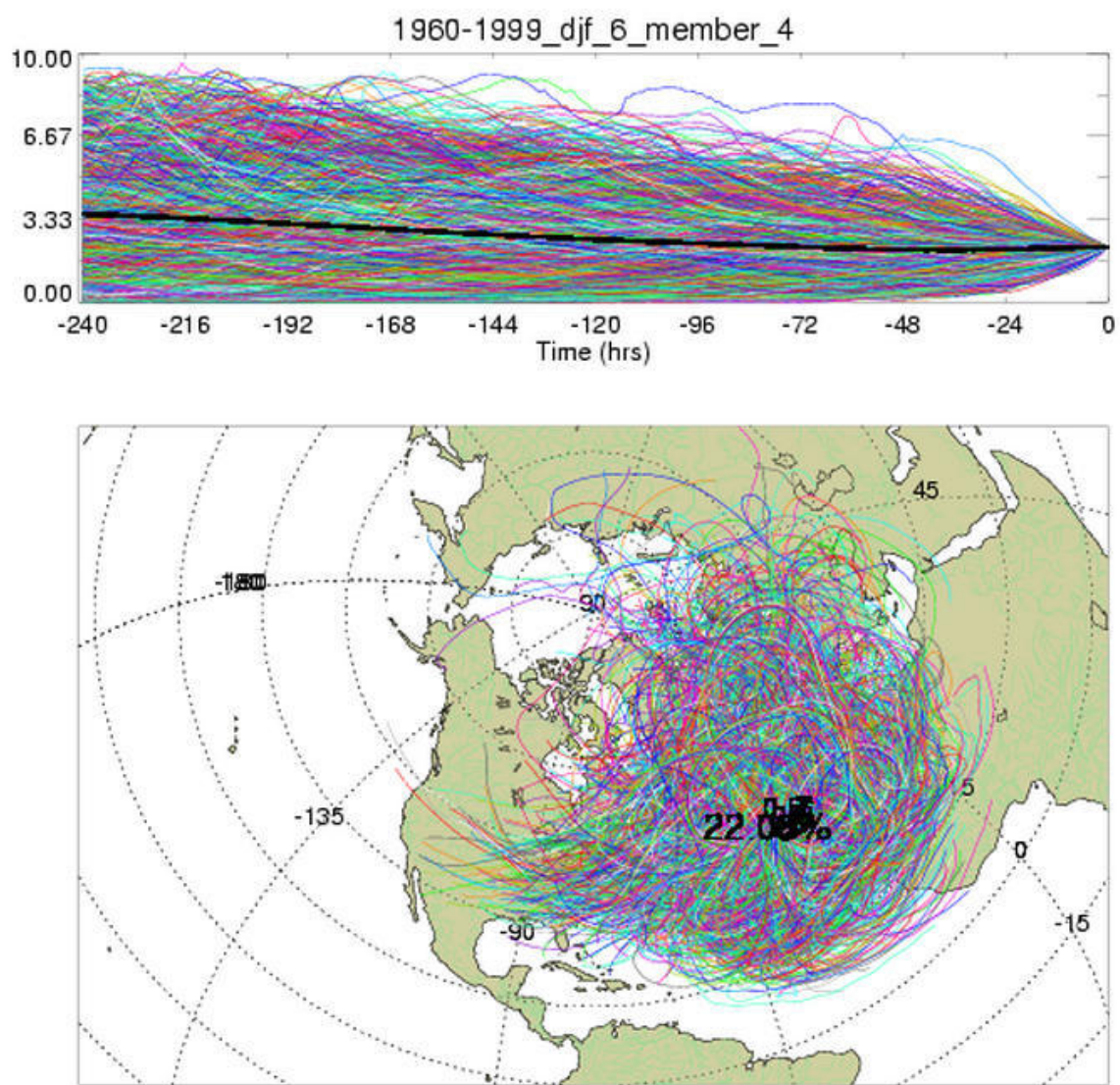


Figure B.14 Membership plot for winters, 1960-1999.

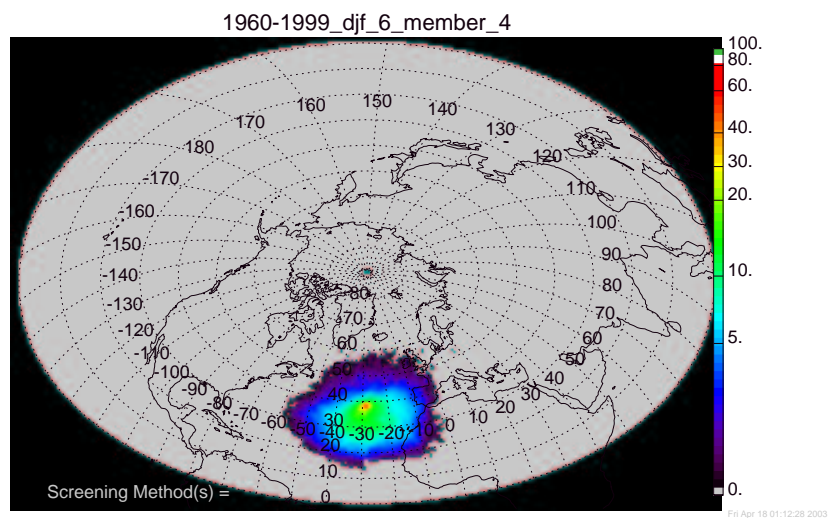


Figure B.15 Standard density membership plot for winters, 1960-1999.

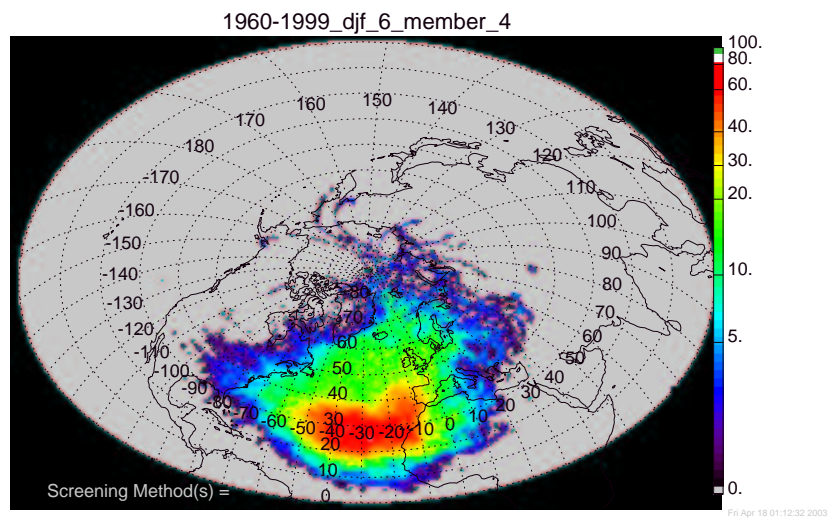


Figure B.16 Geometrically corrected density membership plot for winters, 1960-1999.

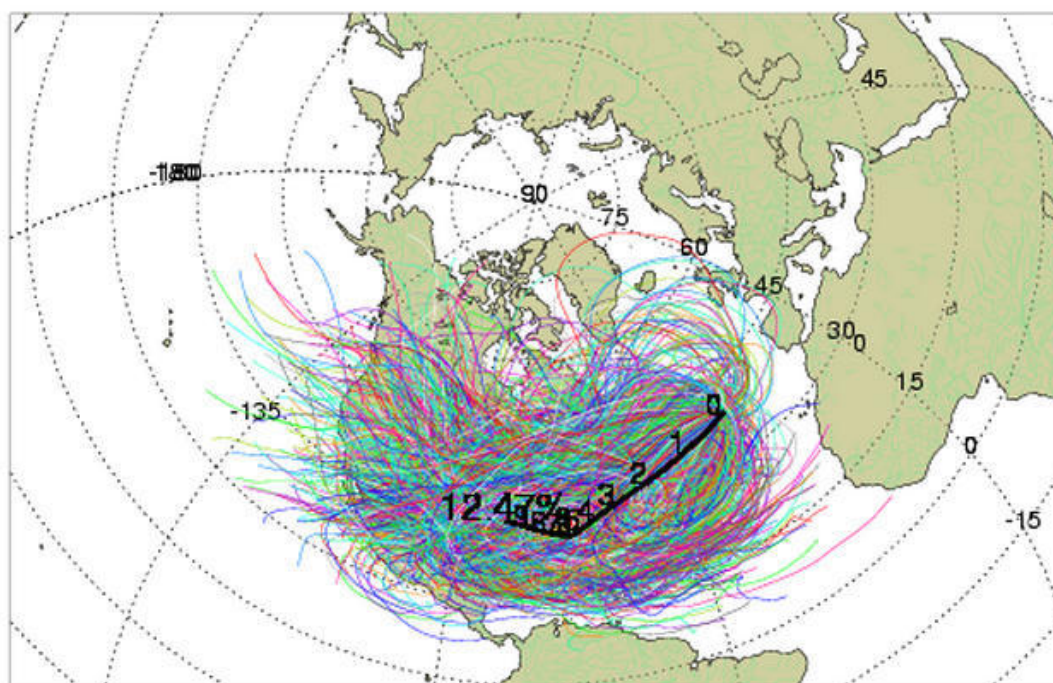


Figure B.17 Membership plot for winters, 1960-1999.

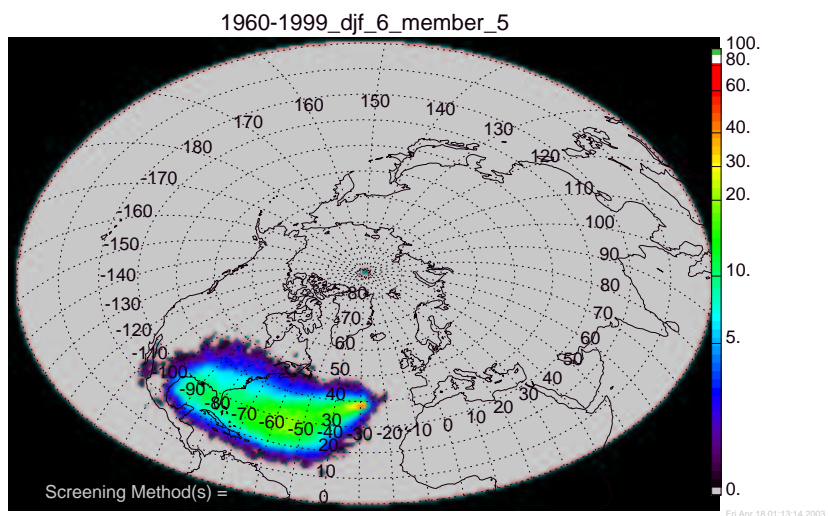


Figure B.18 Standard density membership plot for winters, 1960-1999.

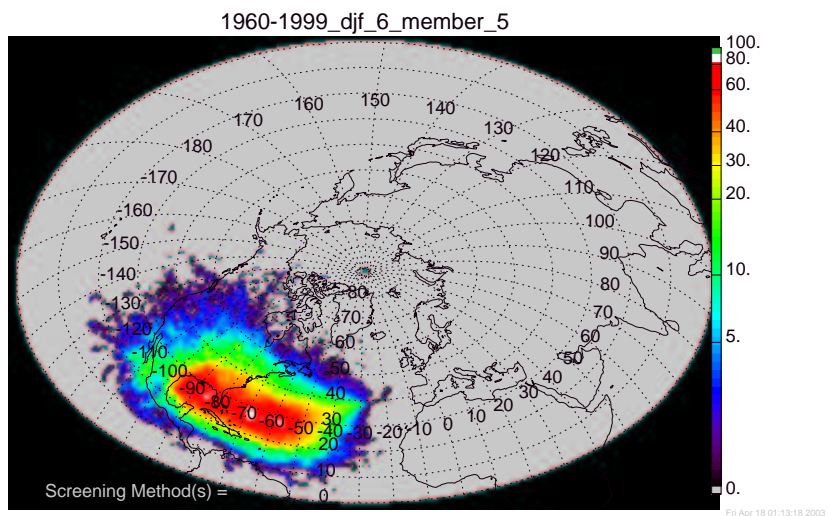


Figure B.19 Geometrically corrected density membership plot for winters, 1960-1999.

B.2 Membership plots for the positive NAO winters

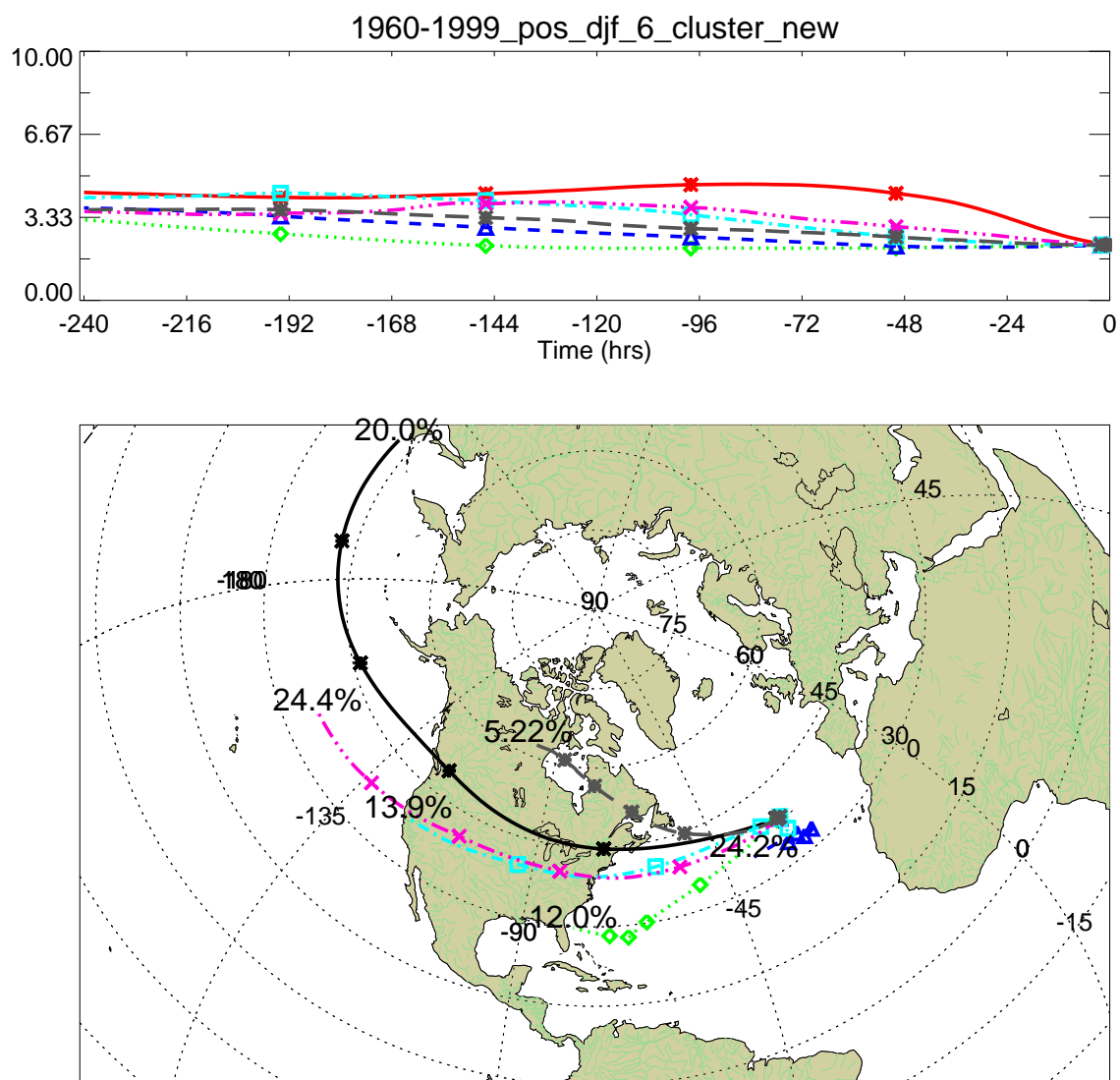


Figure B.20 Cluster plot for positive NAO winters, 1960-1999.

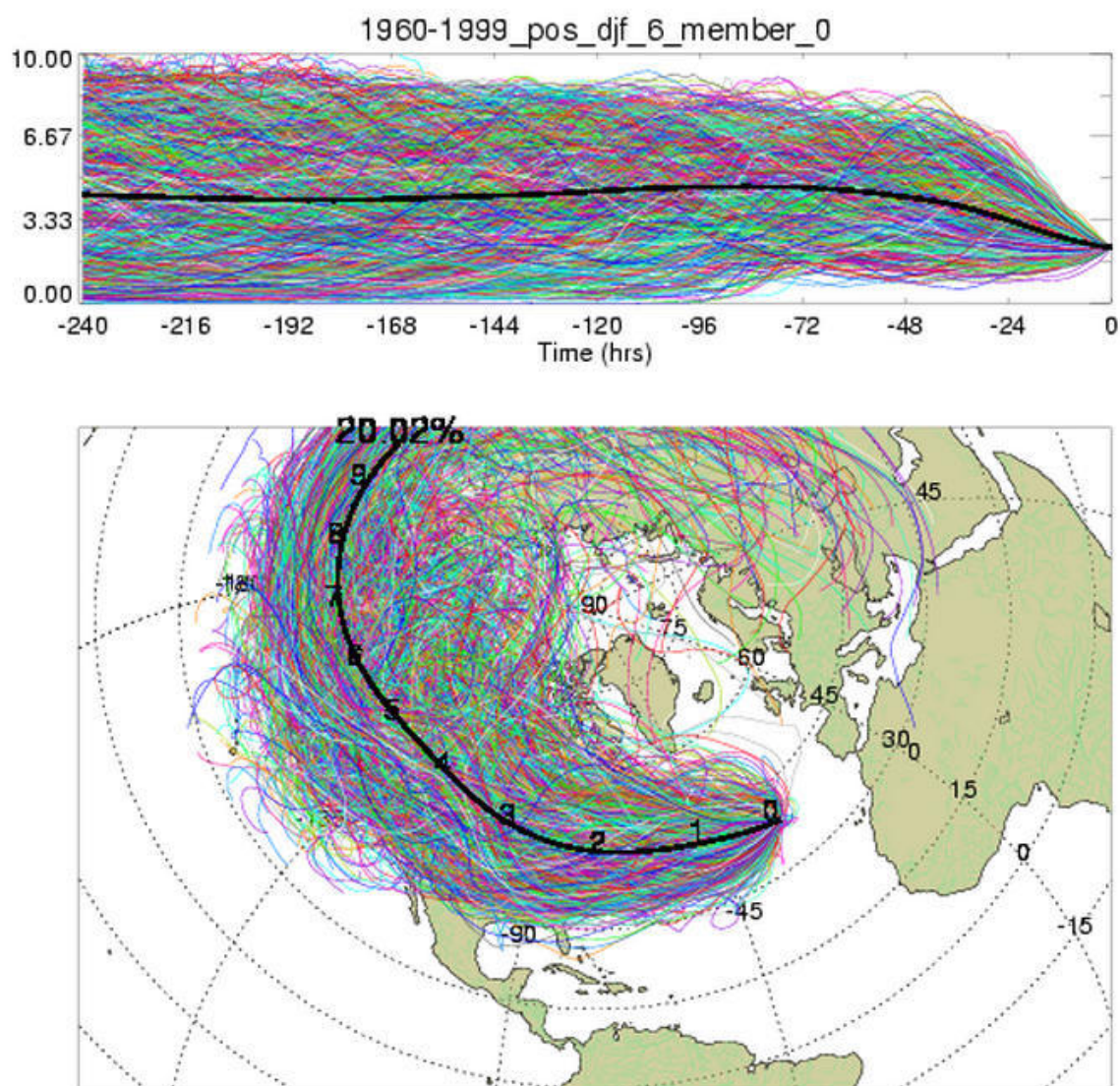


Figure B.21 Membership plot for winters with a positive NAOI value, 1960-1999.

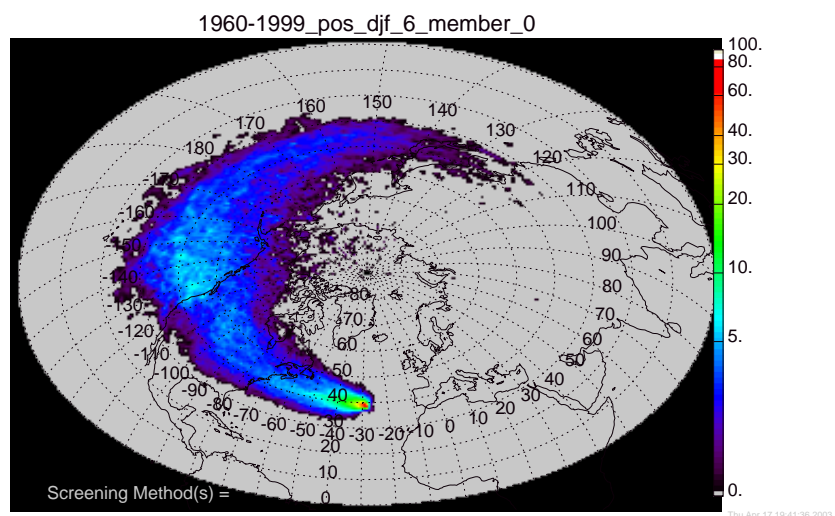


Figure B.22 Standard density membership plot for winters with a positive NAOI value, 1960-1999.

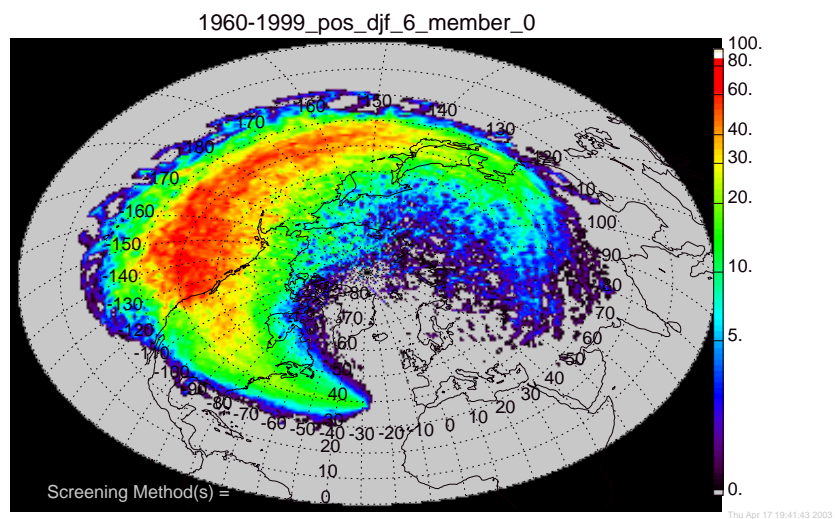


Figure B.23 Geometrically corrected density membership plot for winters with a positive NAOI value, 1960-1999.

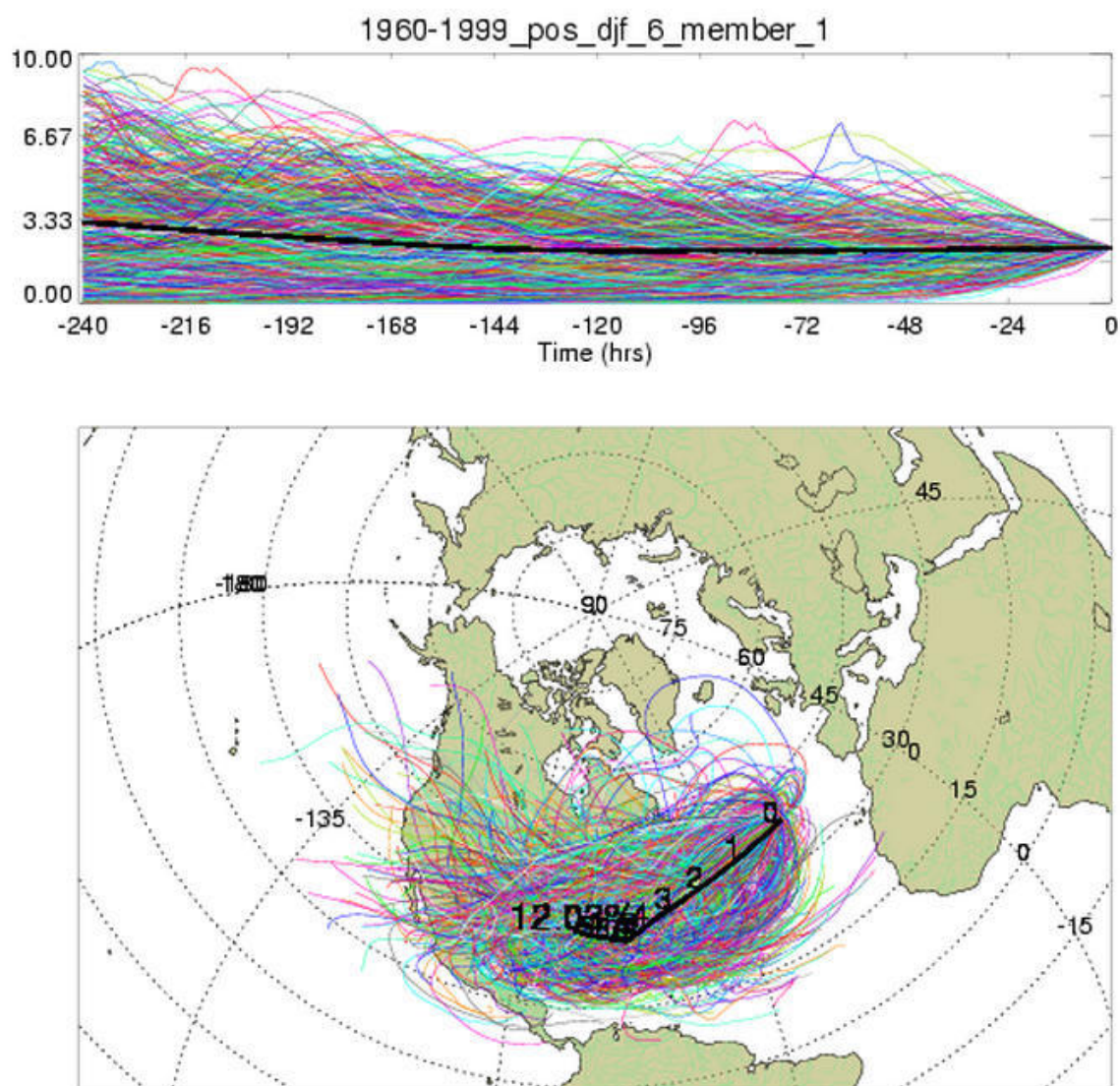


Figure B.24 Membership plot for winters with a positive NAOI value, 1960-1999.

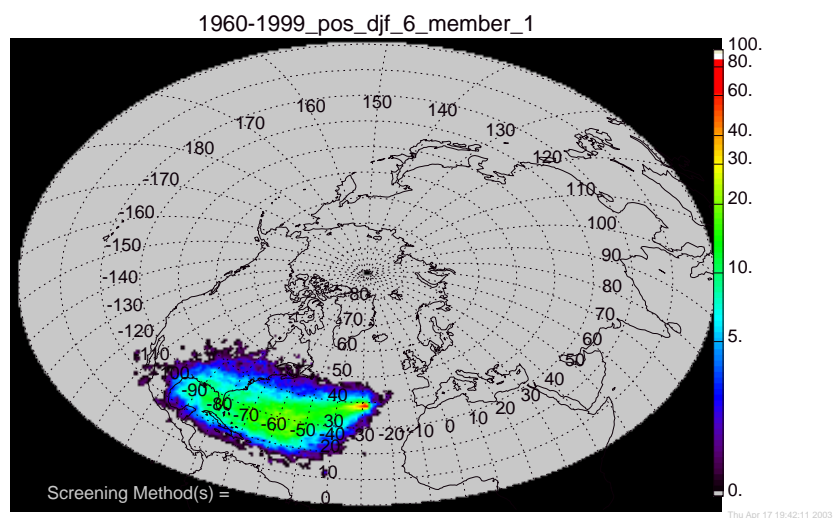


Figure B.25 Standard density membership plot for winters with a positive NAOI value, 1960-1999.

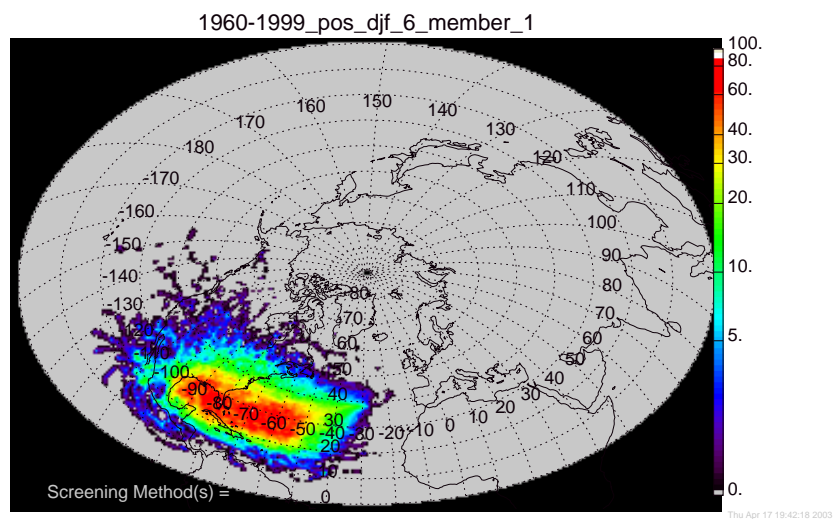


Figure B.26 Geometrically corrected density membership plot for winters with a positive NAOI value, 1960-1999.

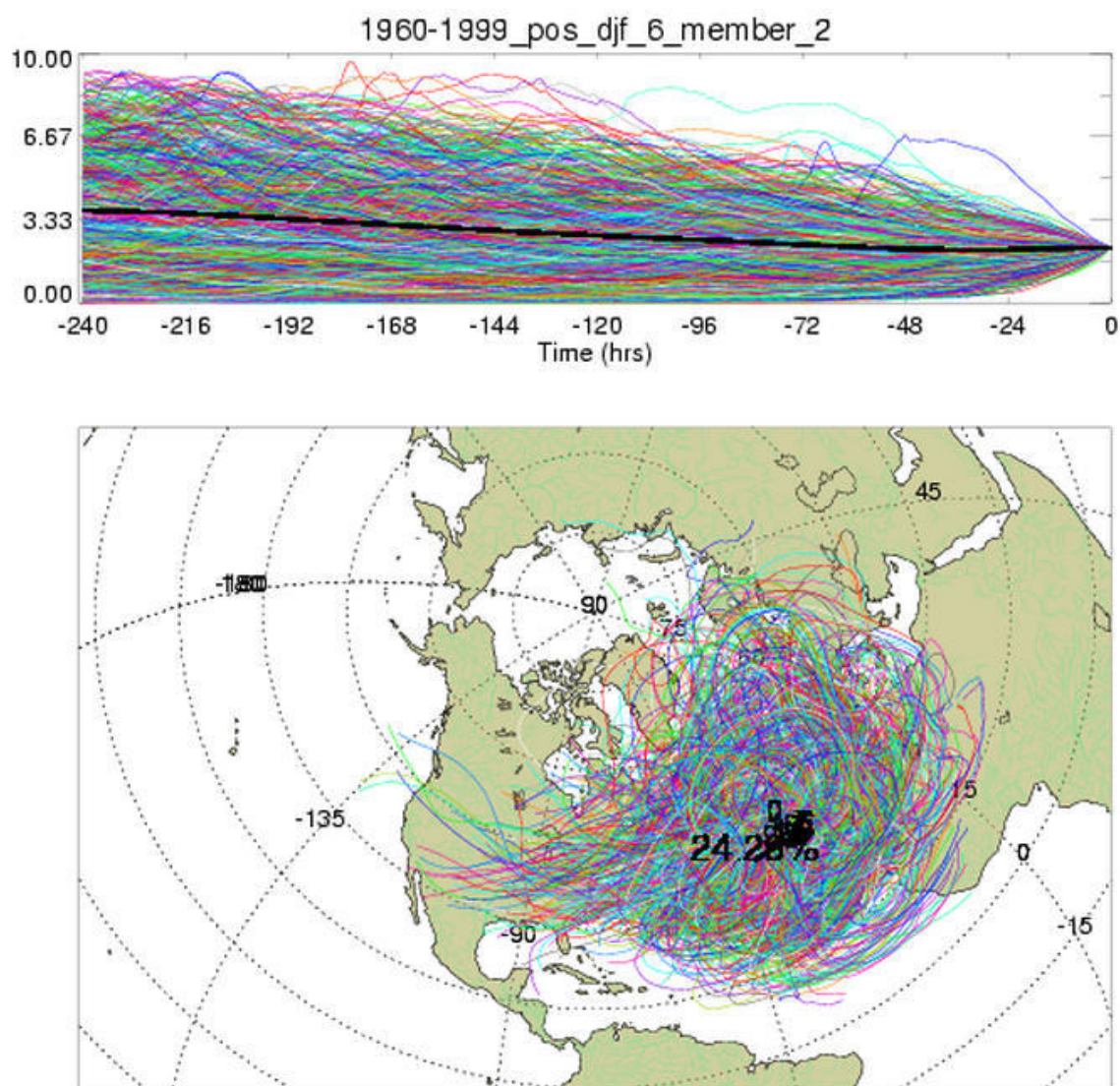


Figure B.27 Membership plot for winters with a positive NAOI value, 1960-1999.

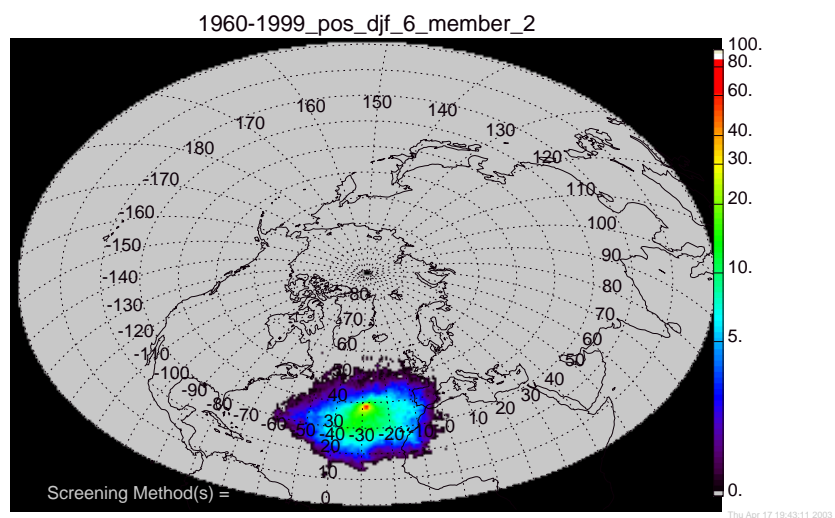


Figure B.28 Standard density membership plot for winters with a positive NAOI value, 1960-1999.

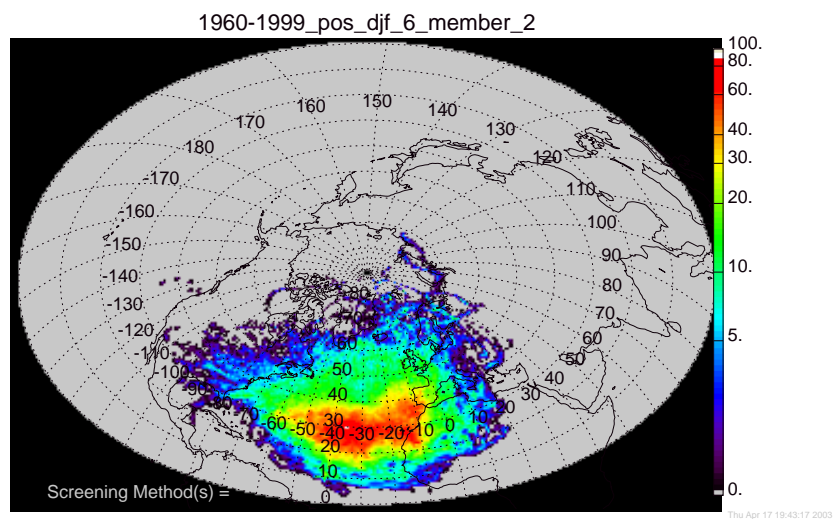


Figure B.29 Geometrically corrected density membership plot for winters with a positive NAOI value, 1960-1999.

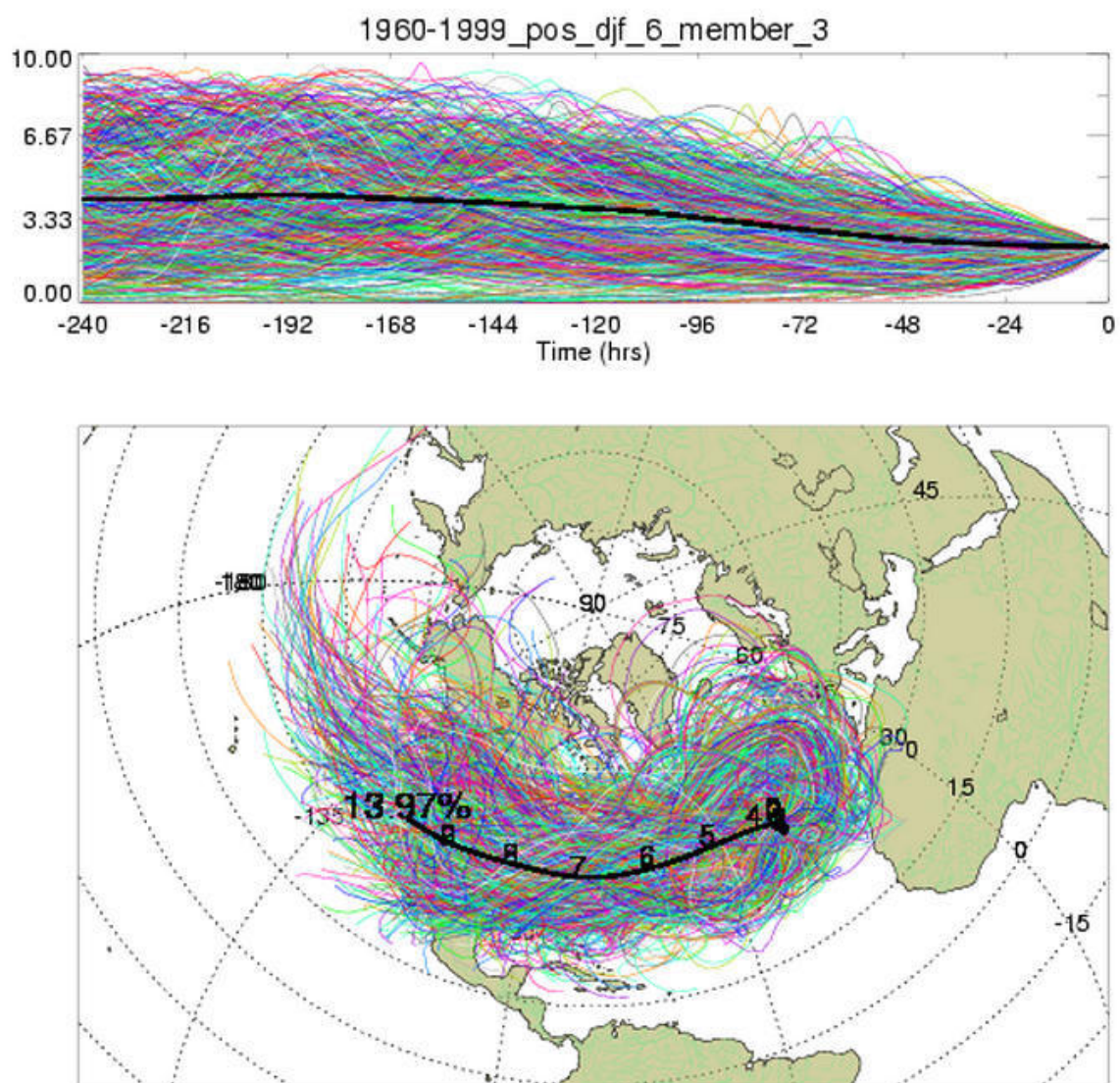


Figure B.30 Membership plot for winters with a positive NAOI value, 1960-1999.

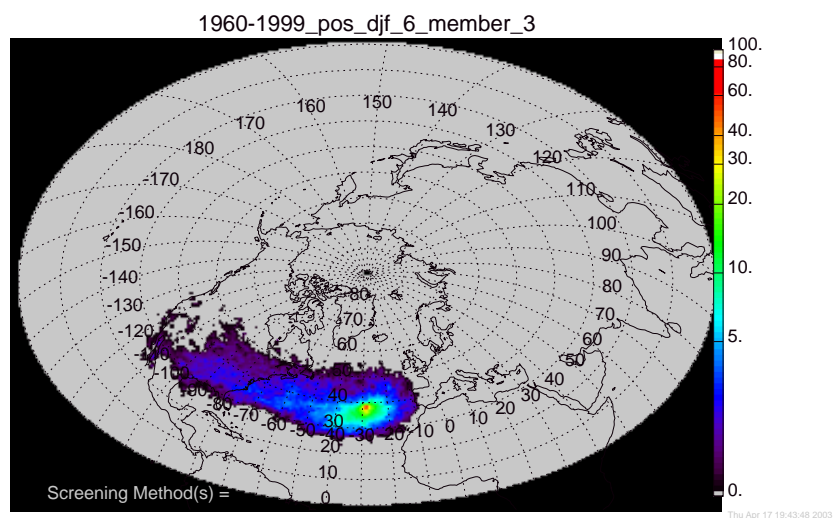


Figure B.31 Standard density membership plot for winters with a positive NAOI value, 1960-1999.

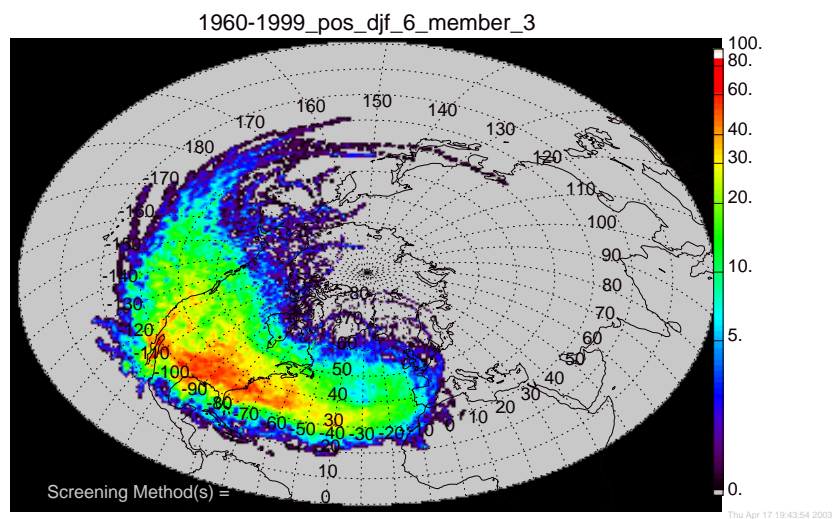


Figure B.32 Geometrically corrected density membership plot for winters with a positive NAOI value, 1960-1999.

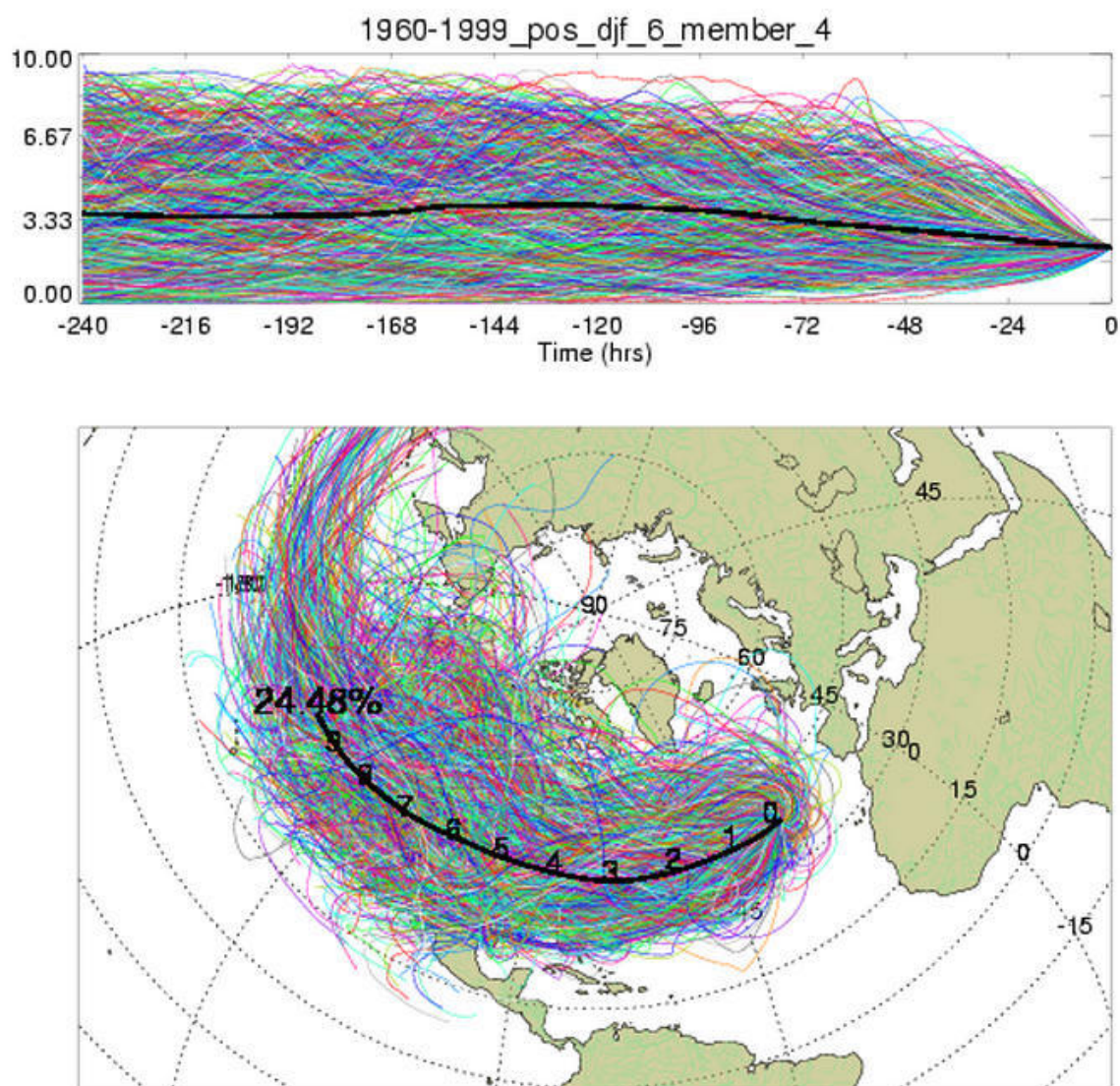


Figure B.33 Membership plot for winters with a positive NAOI value, 1960-1999.

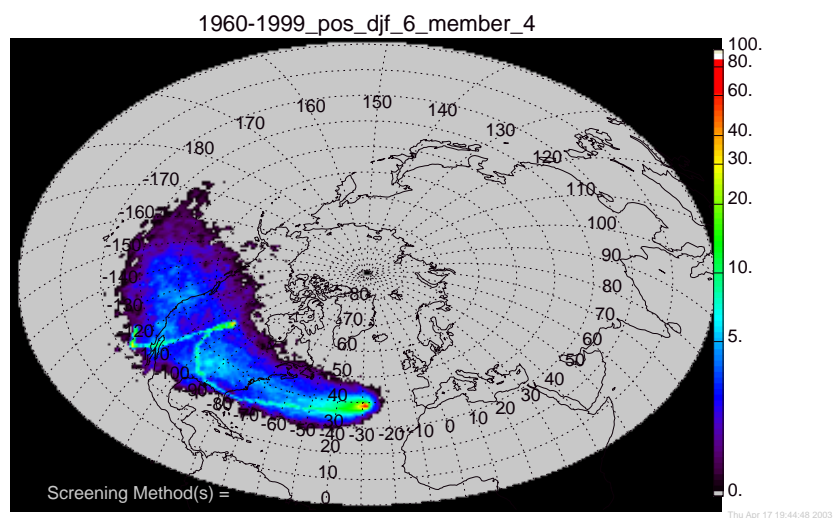


Figure B.34 Standard density membership plot for winters with a positive NAOI value, 1960-1999.

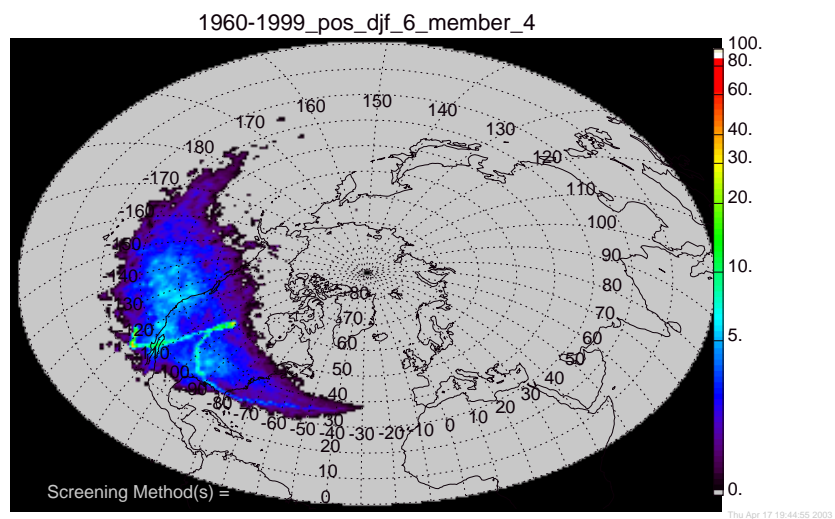


Figure B.35 Geometrically corrected density membership plot for winters with a positive NAOI value, 1960-1999.

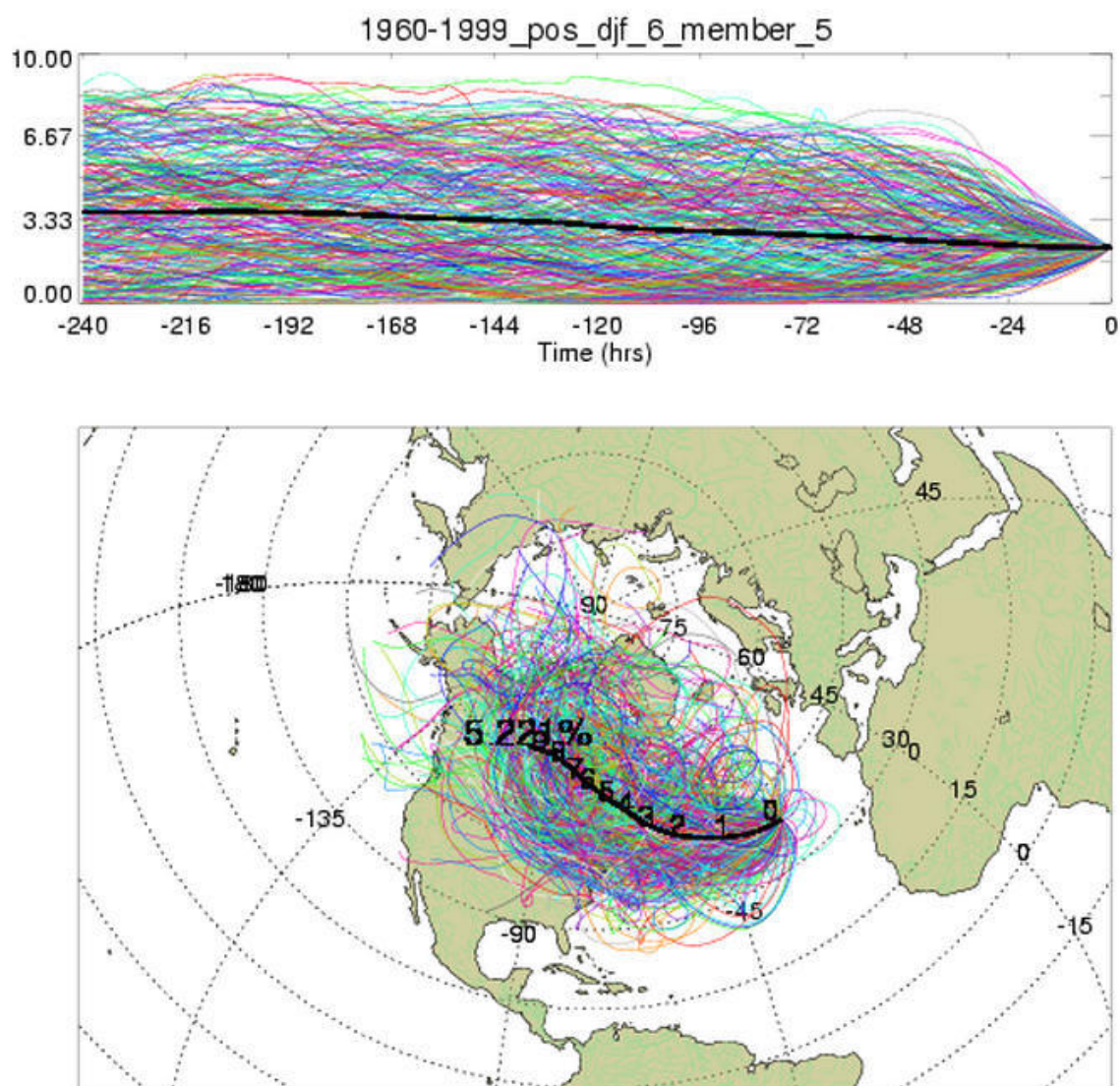


Figure B.36 Membership plot for winters with a positive NAOI value, 1960-1999.

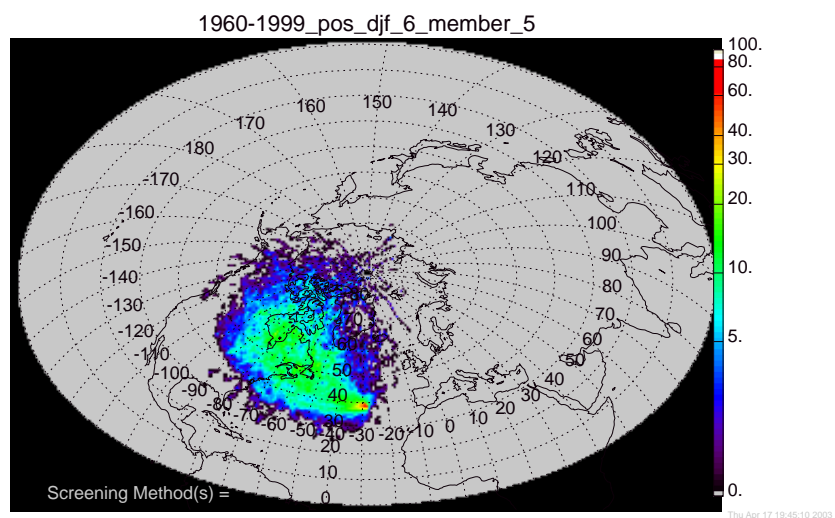


Figure B.37 Standard density membership plot for winters with a positive NAOI value, 1960-1999.

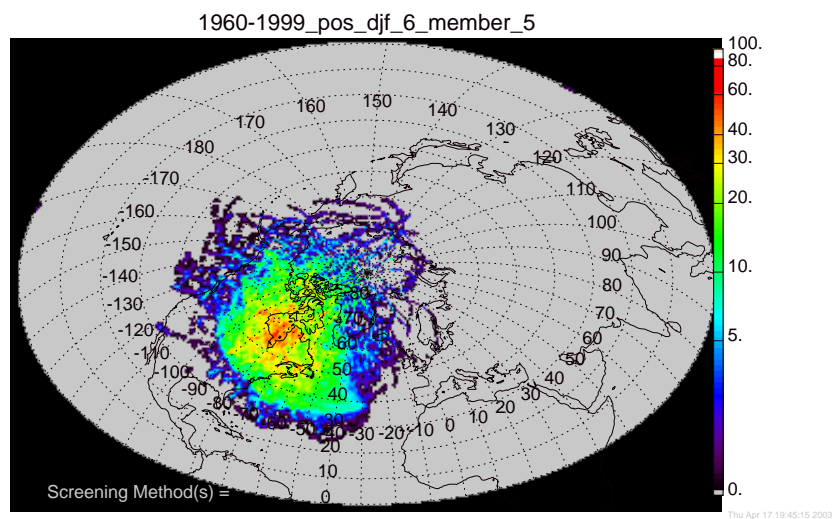


Figure B.38 Geometrically corrected density membership plot for winters with a positive NAOI value, 1960-1999.

B.3 Membership plots for the negative NAO winters

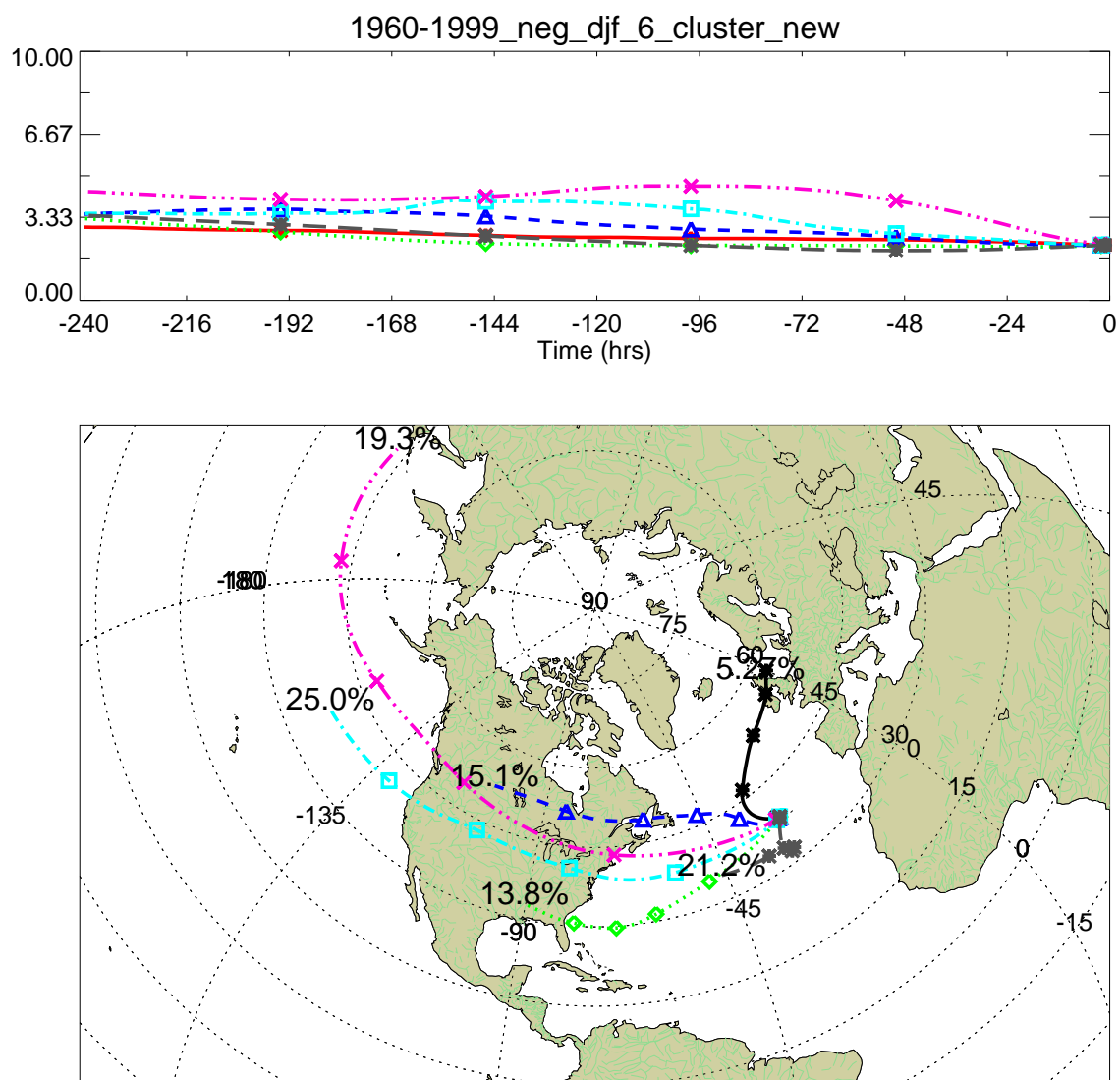


Figure B.39 Cluster plot for negative NAO winters, 1960-1999.

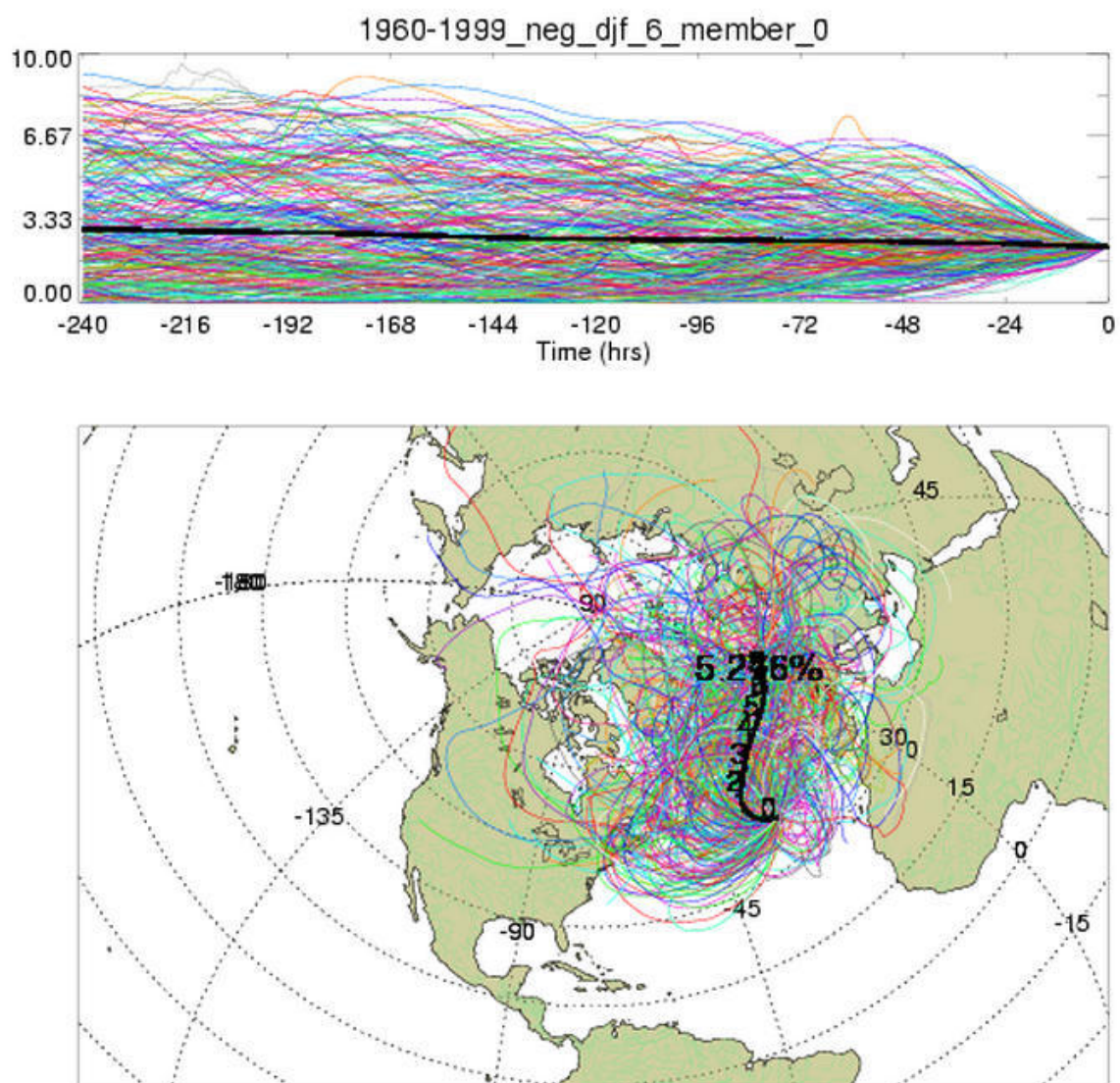


Figure B.40 Membership plot for winters with a negative NAOI value, 1960-1999.

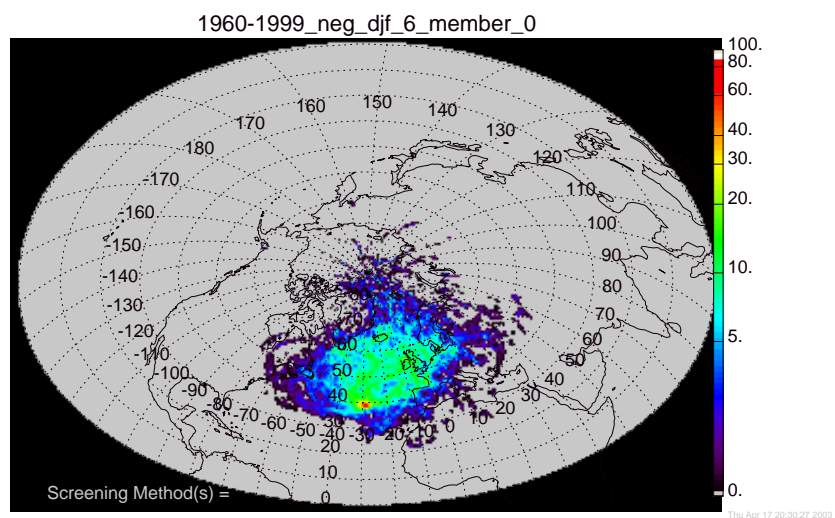


Figure B.41 Standard density membership plot for winters with a negative NAOI value, 1960-1999.

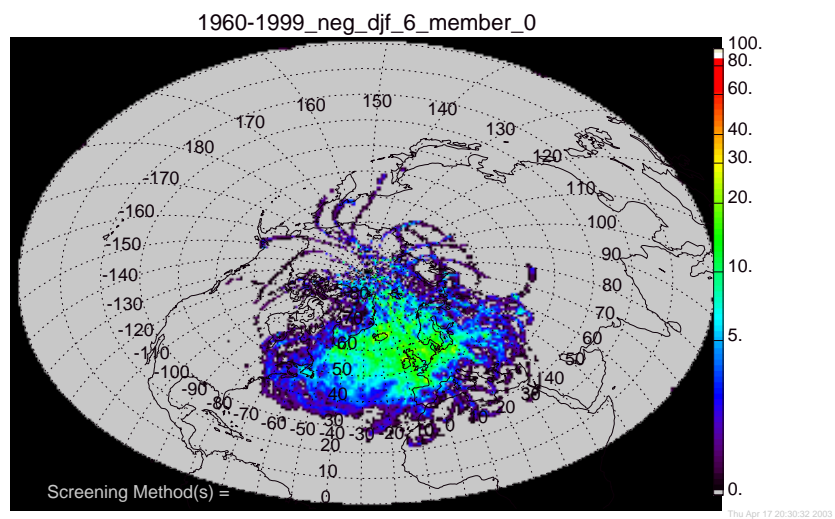


Figure B.42 Geometrically corrected density membership plot for winters with a negative NAOI value, 1960-1999.

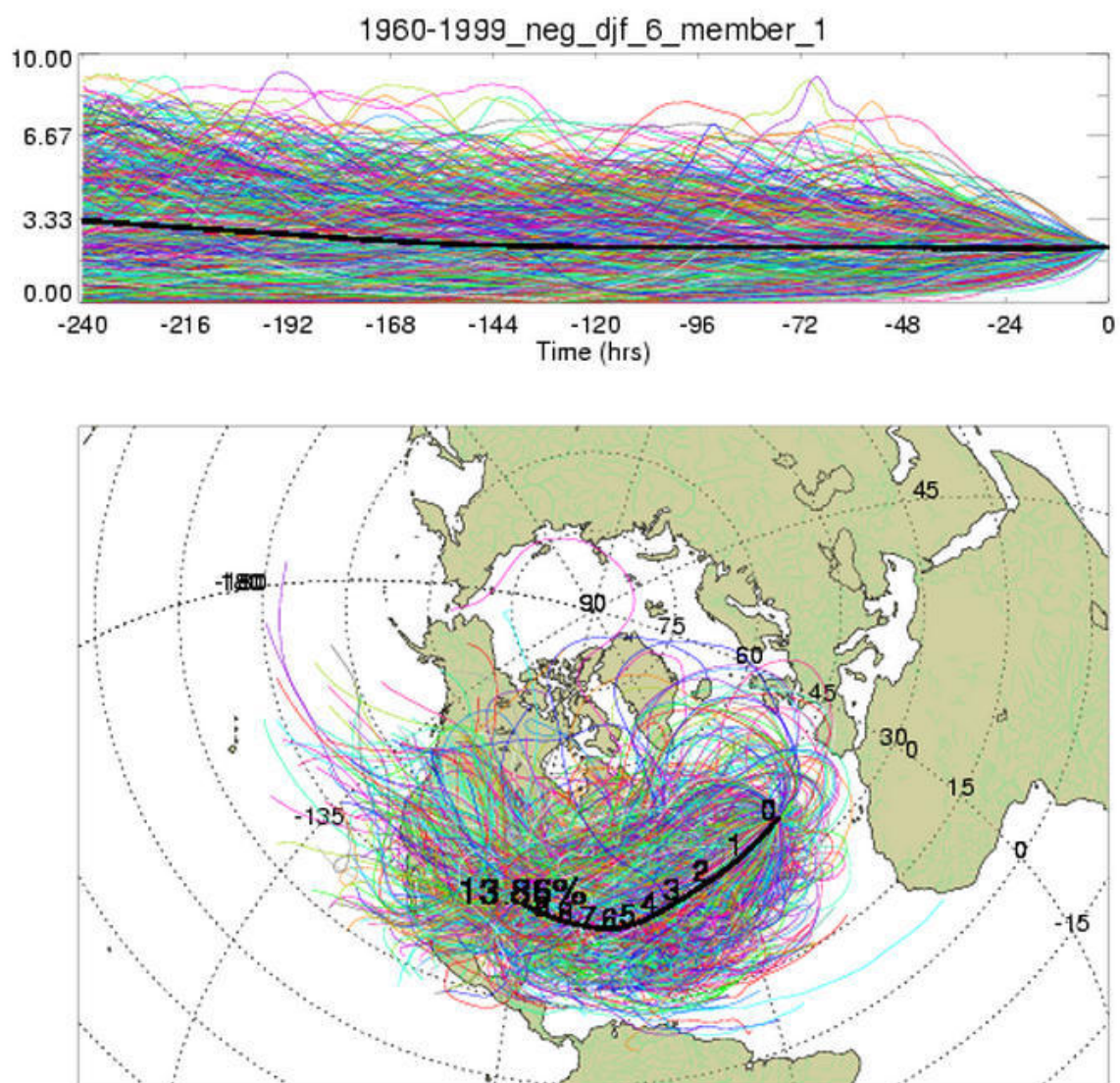


Figure B.43 Membership plot for winters with a negative NAOI value, 1960-1999.

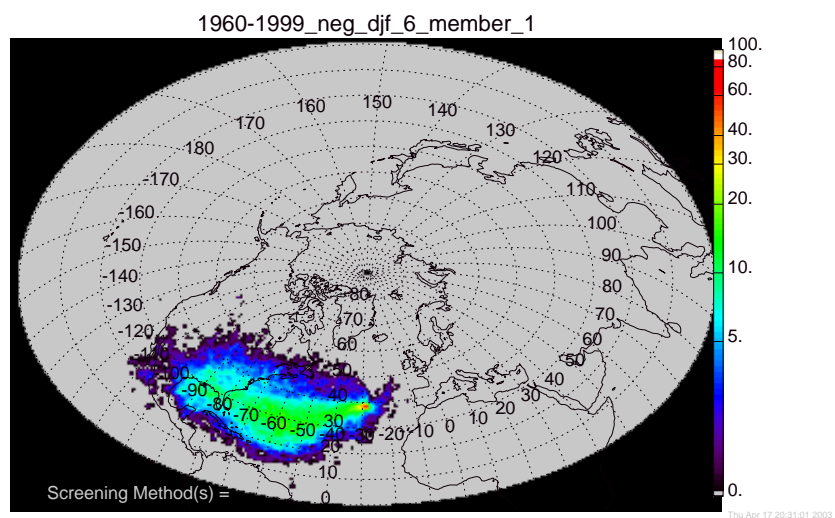


Figure B.44 Standard density membership plot for winters with a negative NAOI value, 1960-1999.

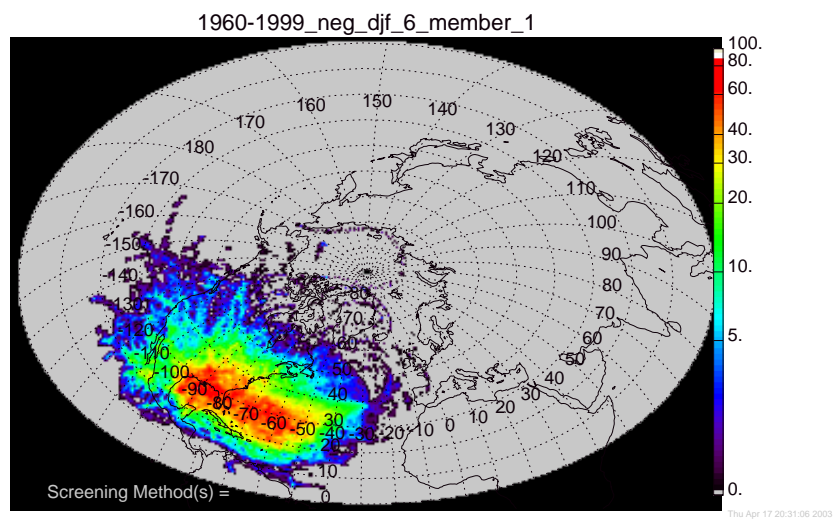


Figure B.45 Geometrically corrected density membership plot for winters with a negative NAOI value, 1960-1999.

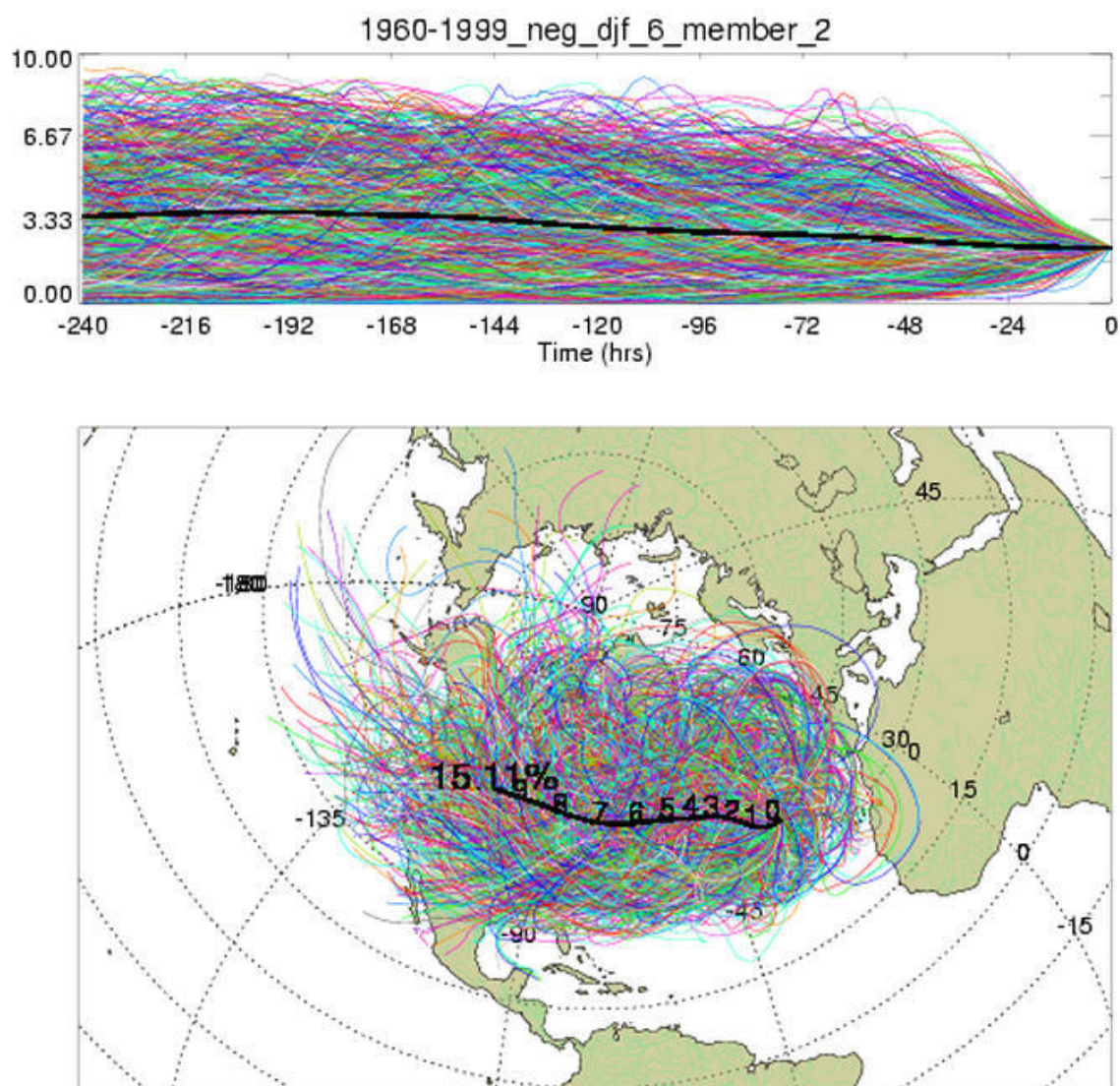


Figure B.46 Membership plot for winters with a negative NAOI value, 1960-1999.

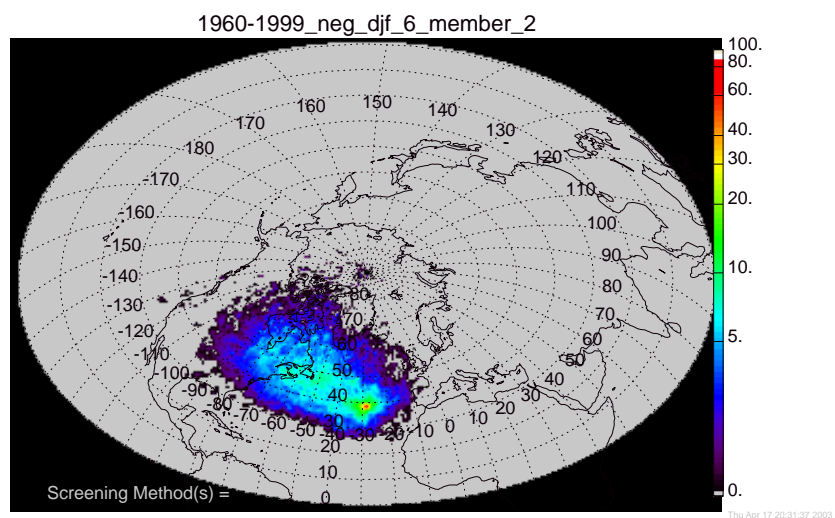


Figure B.47 Standard density membership plot for winters with a negative NAOI value, 1960-1999.

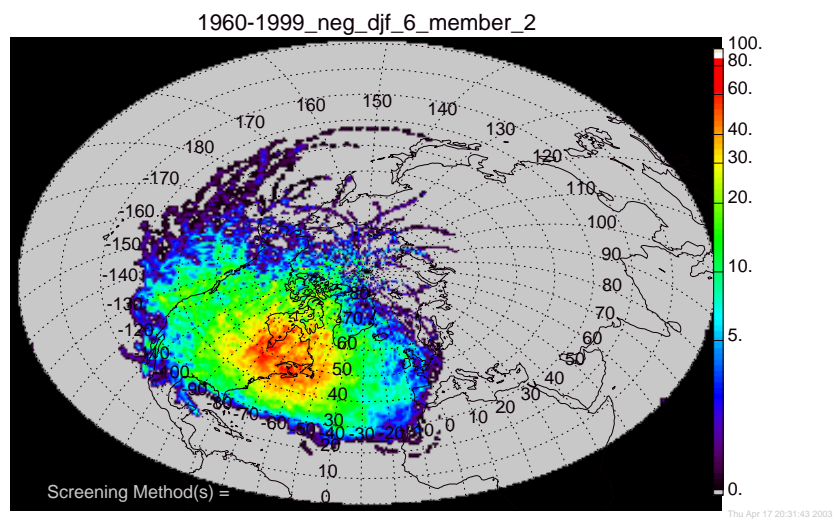


Figure B.48 Geometrically corrected density membership plot for winters with a negative NAOI value, 1960-1999.

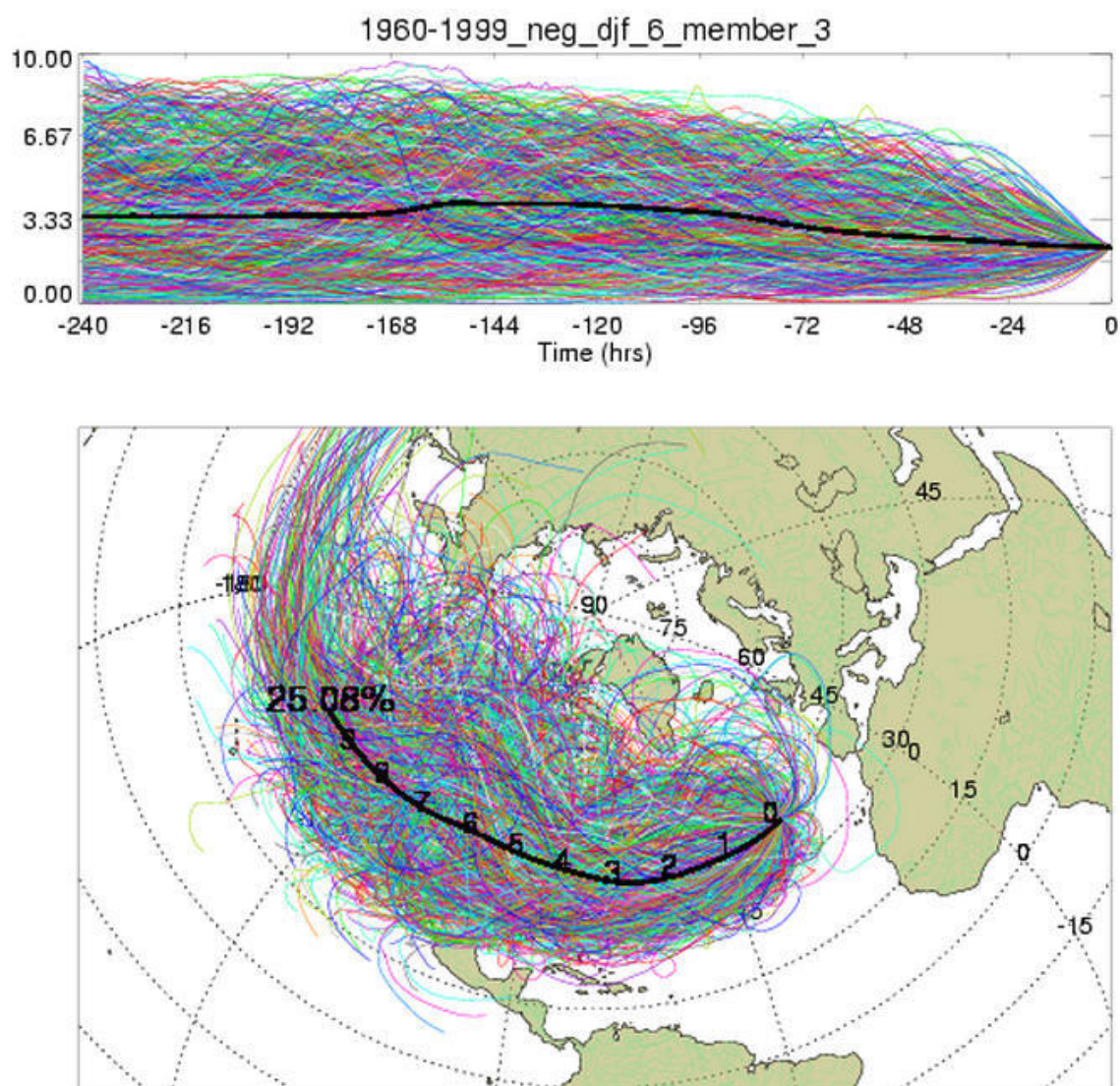


Figure B.49 Membership plot for winters with a negative NAOI value, 1960-1999.

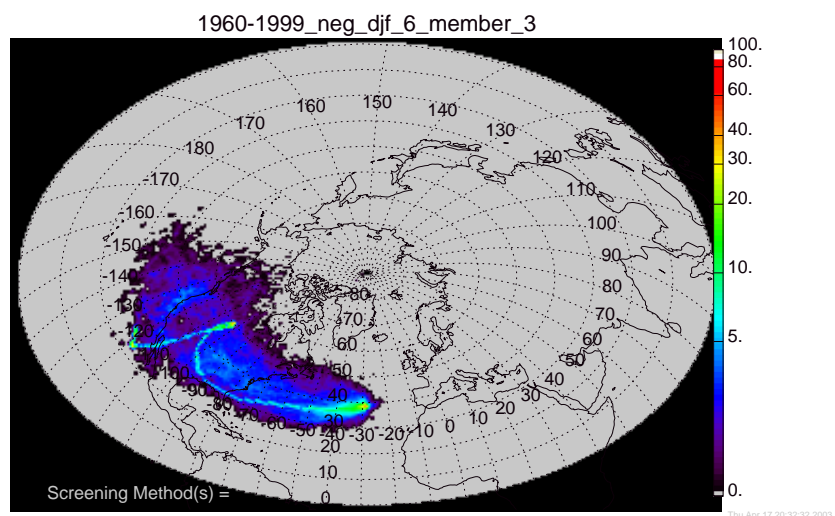


Figure B.50 Standard density membership plot for winters with a negative NAOI value, 1960-1999.

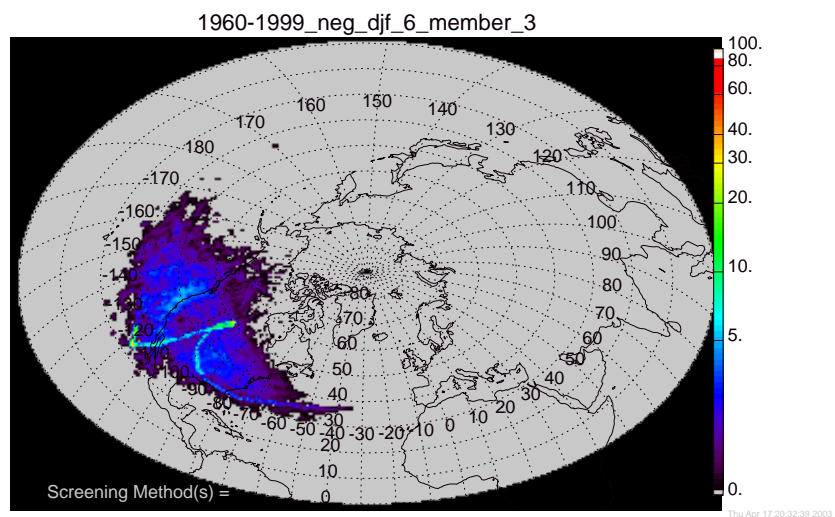


Figure B.51 Geometrically corrected density membership plot for winters with a negative NAOI value, 1960-1999.

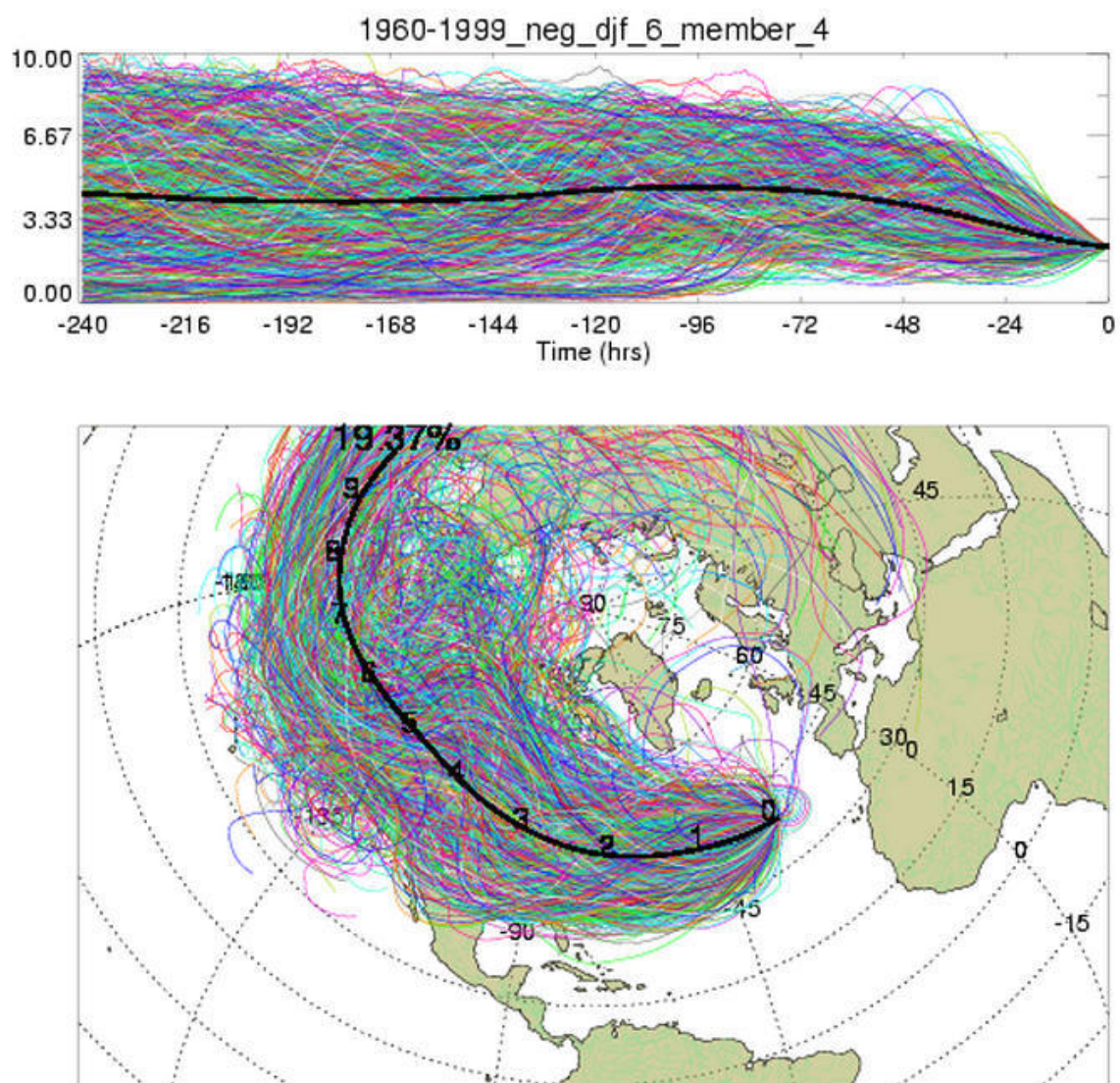


Figure B.52 Membership plot for winters with a negative NAOI value, 1960-1999.

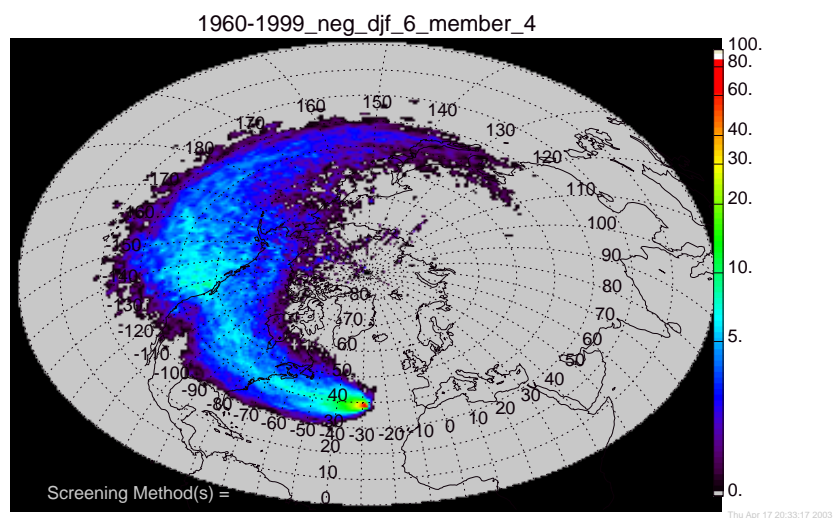


Figure B.53 Standard density membership plot for winters with a negative NAOI value, 1960-1999.

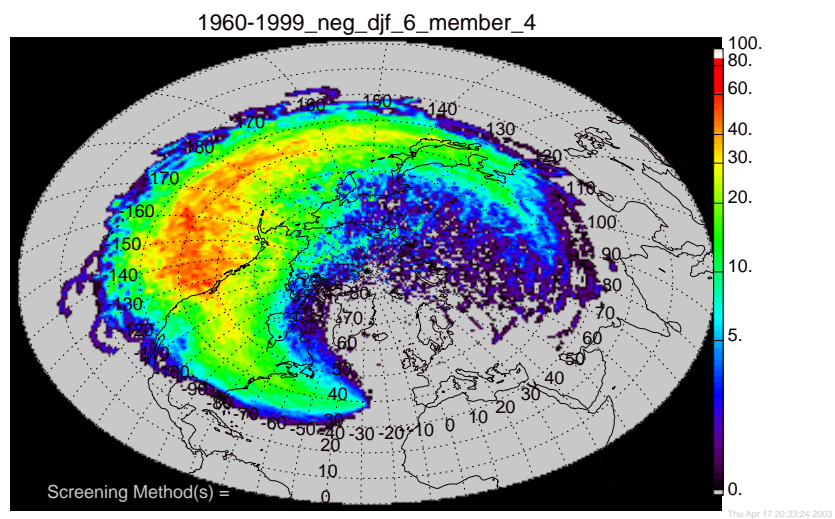


Figure B.54 Geometrically corrected density membership plot for winters with a negative NAOI value, 1960-1999.

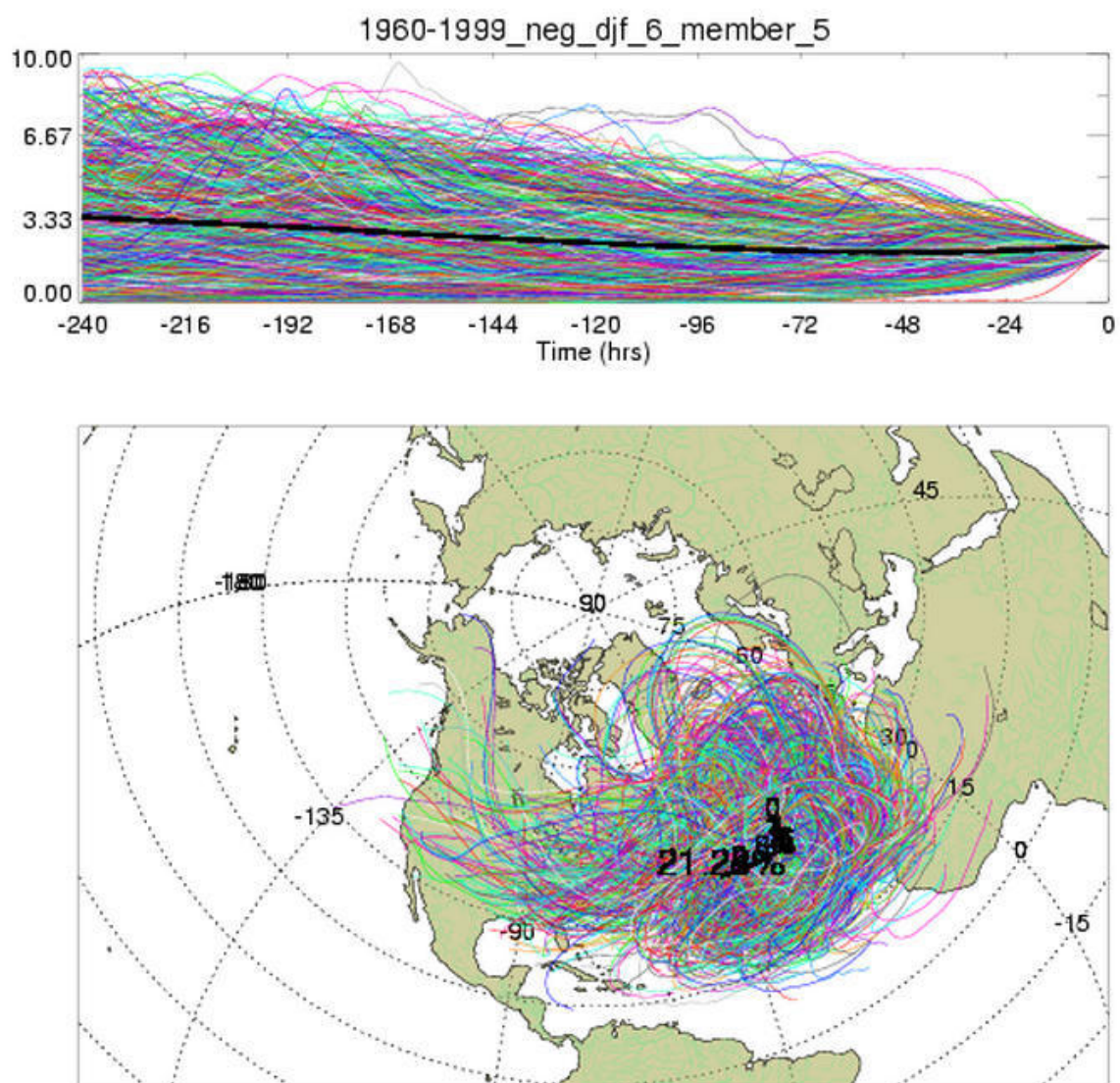


Figure B.55 Membership plot for winters with a negative NAOI value, 1960-1999.

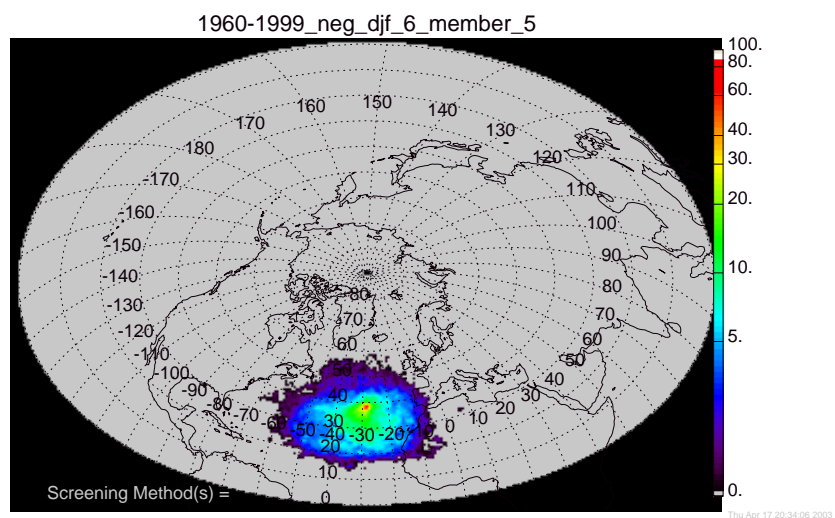


Figure B.56 Standard density membership plot for winters with a negative NAOI value, 1960-1999.

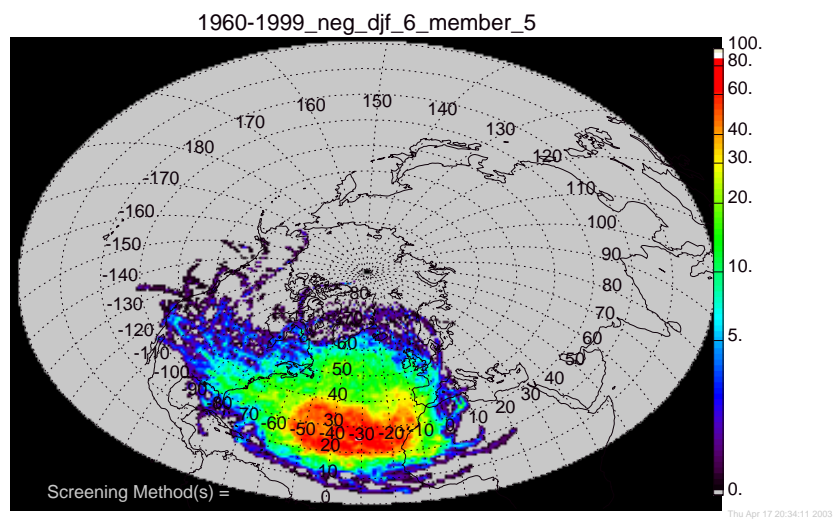


Figure B.57 Geometrically corrected density membership plot for winters with a negative NAOI value, 1960-1999.

Appendix C

Membership plots for spring periods

C.1 Membership plots for the average spring

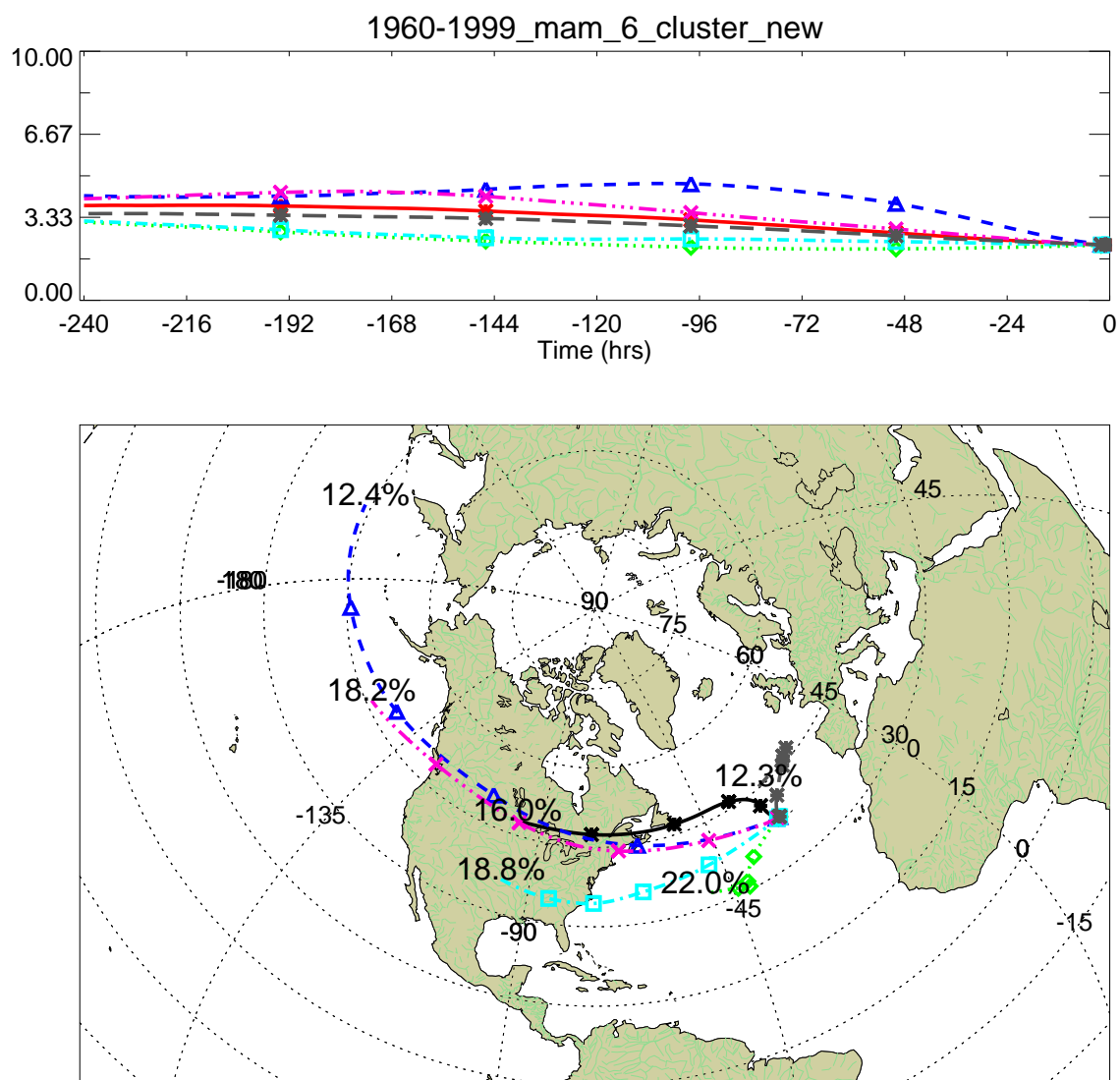


Figure C.1 Cluster plot for the average spring, 1960-1999.

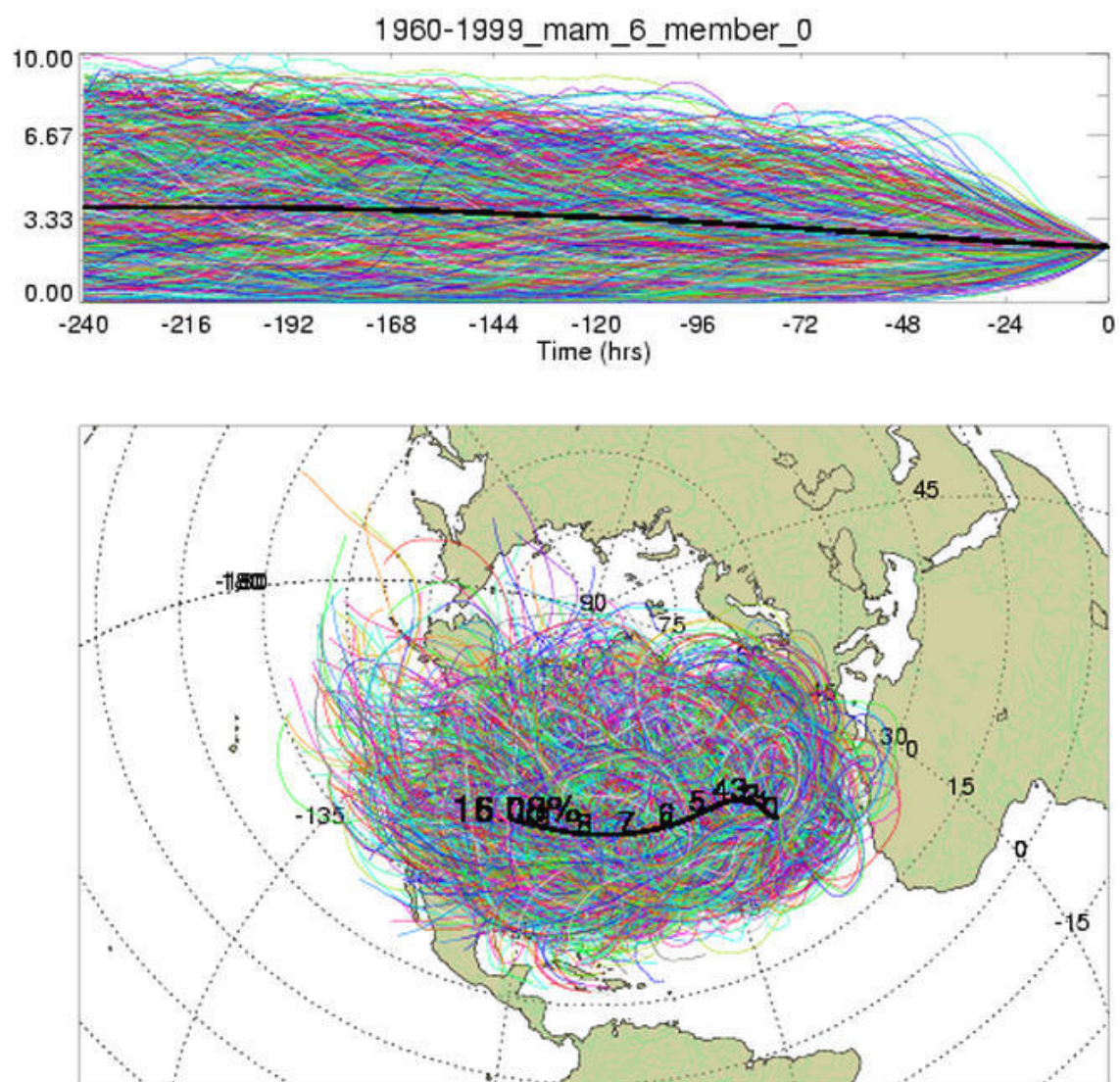


Figure C.2 Membership plot for springs, 1960-1999.

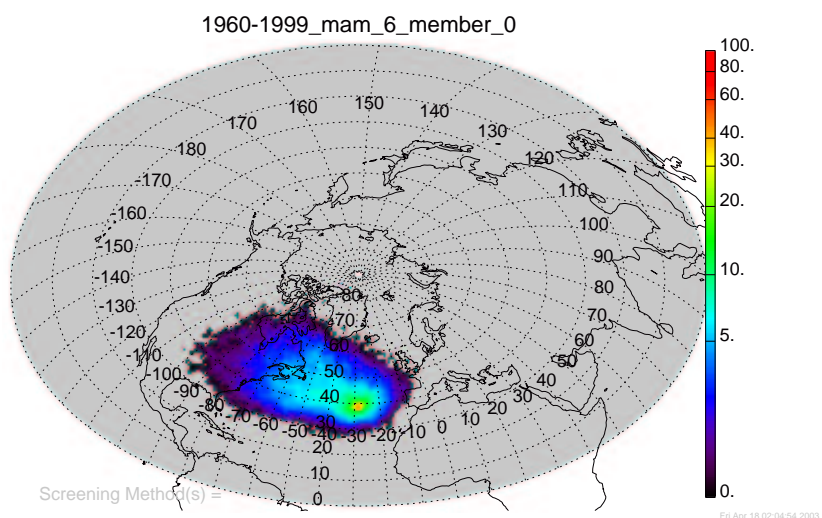


Figure C.3 Standard density membership plot for springs, 1960-1999.

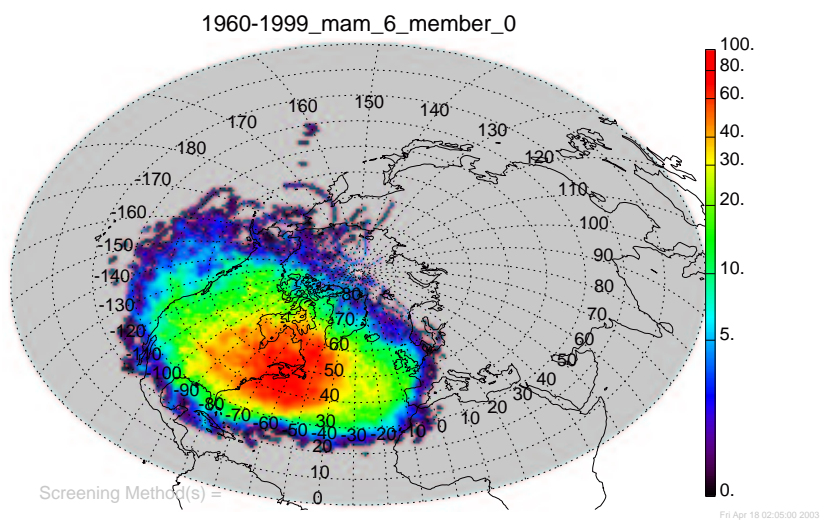


Figure C.4 Geometrically corrected density membership plot for springs, 1960-1999.

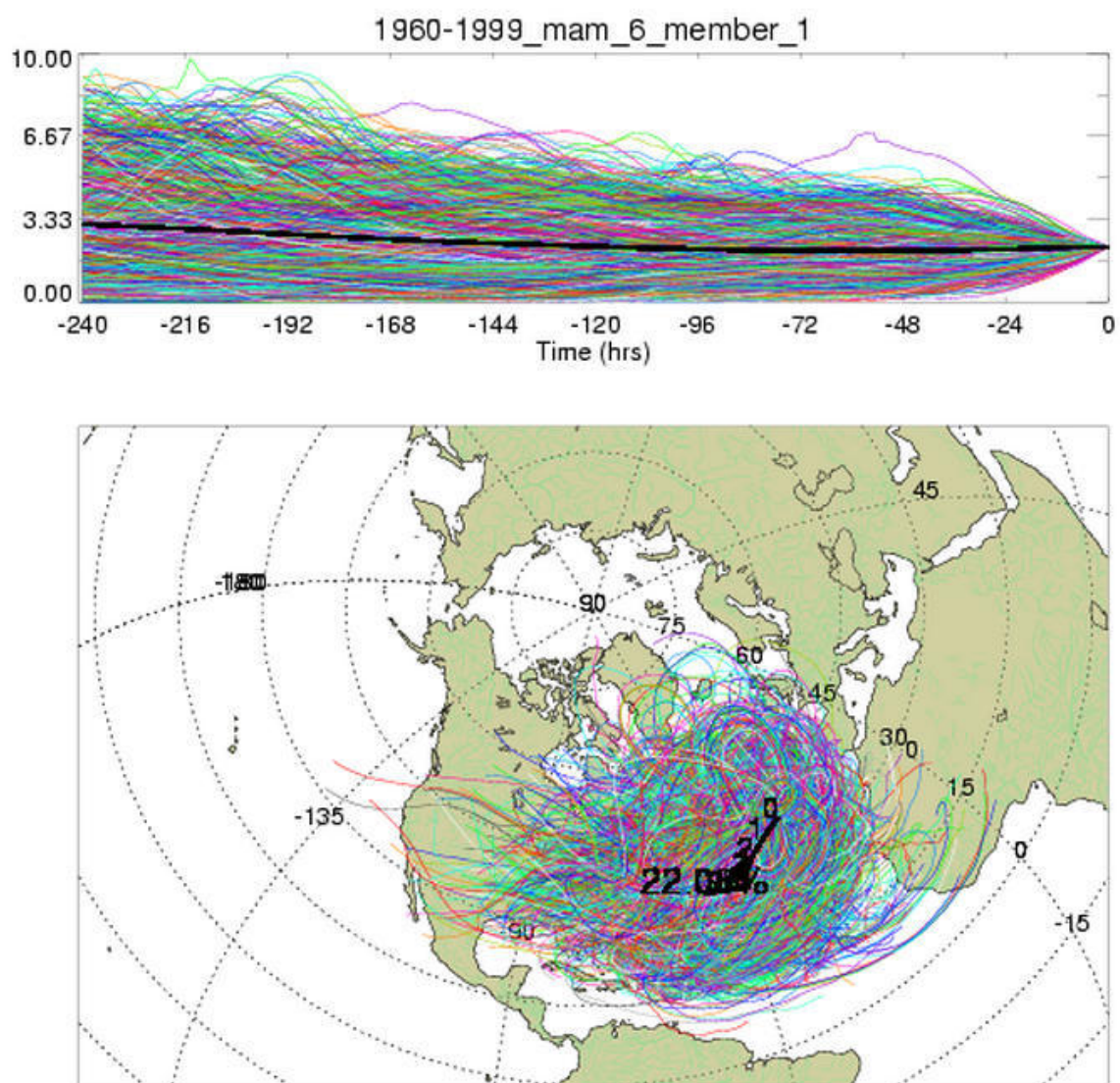


Figure C.5 Membership plot for springs, 1960-1999.

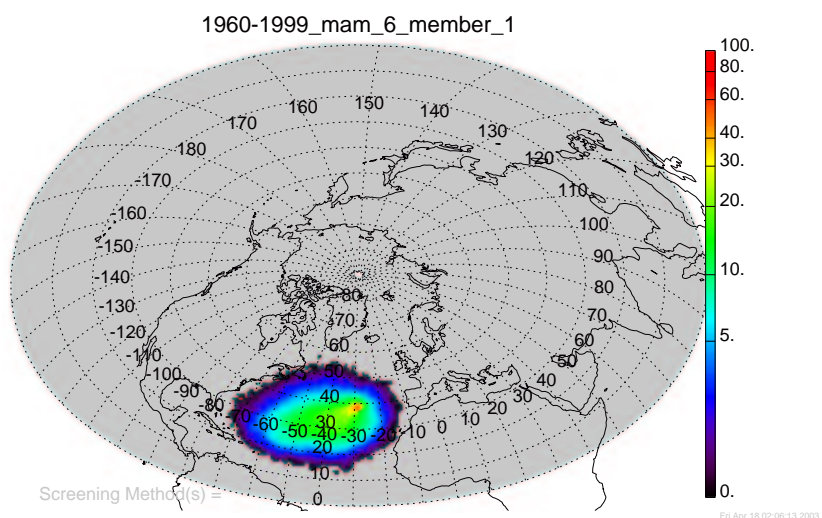


Figure C.6 Standard density membership plot for springs, 1960-1999.

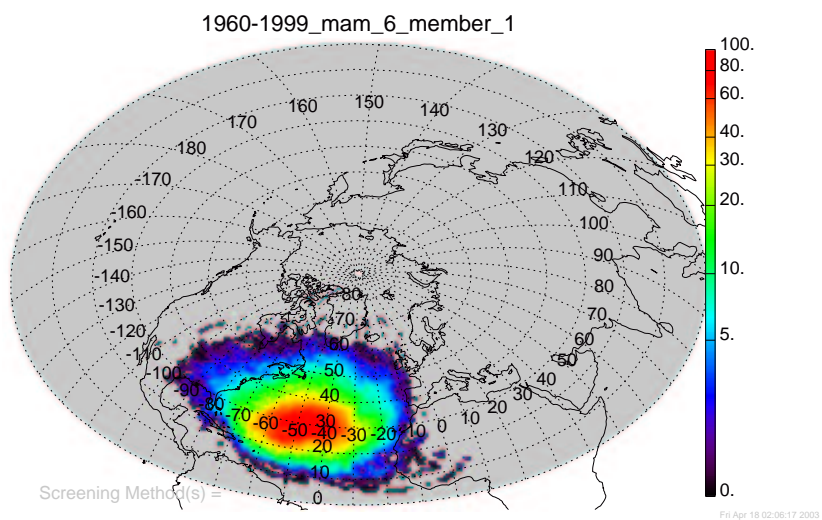


Figure C.7 Geometrically corrected density membership plot for springs, 1960-1999.

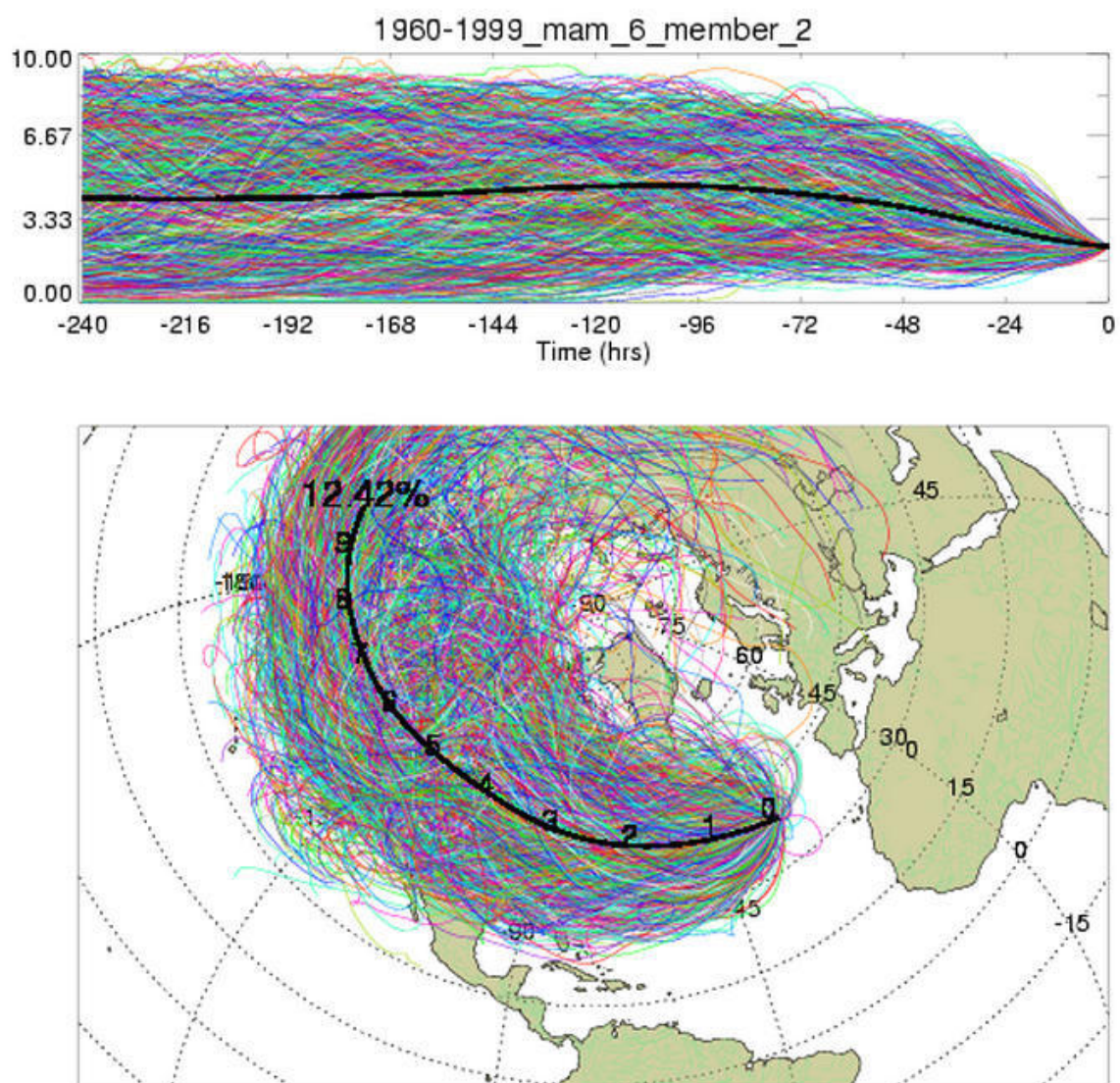


Figure C.8 Membership plot for springs, 1960-1999.

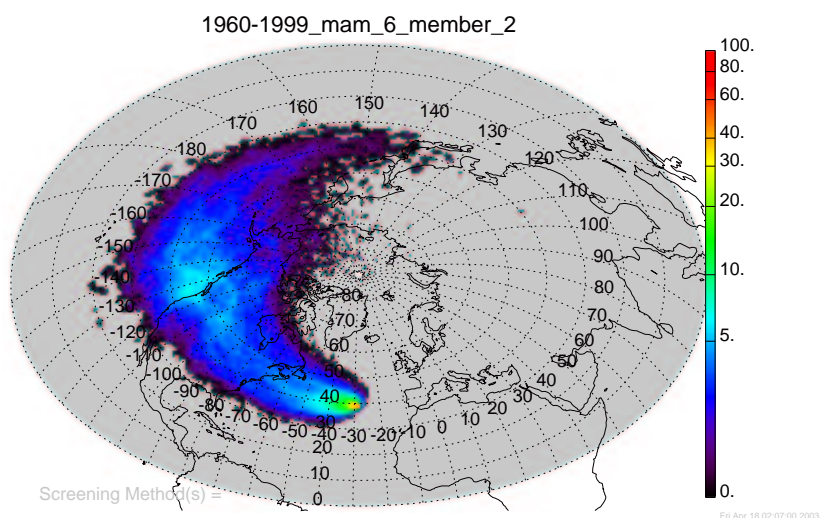


Figure C.9 Standard density membership plot for springs, 1960-1999.

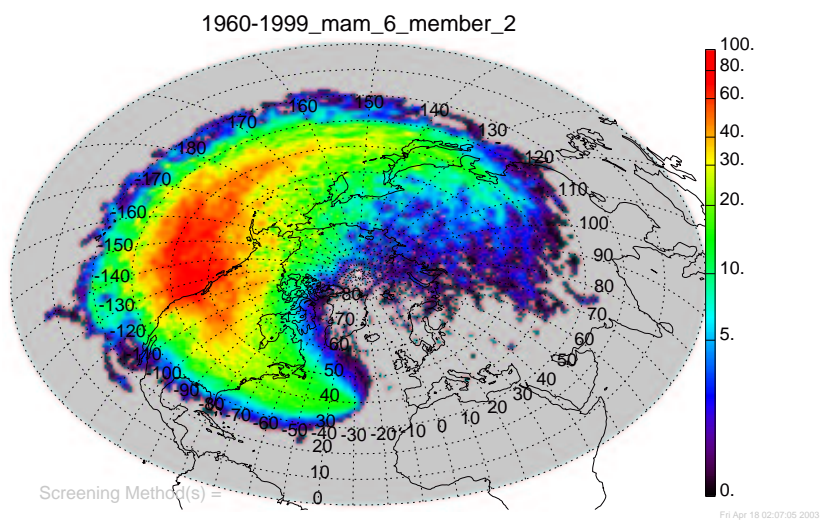


Figure C.10 Geometrically corrected density membership plot for springs, 1960-1999.

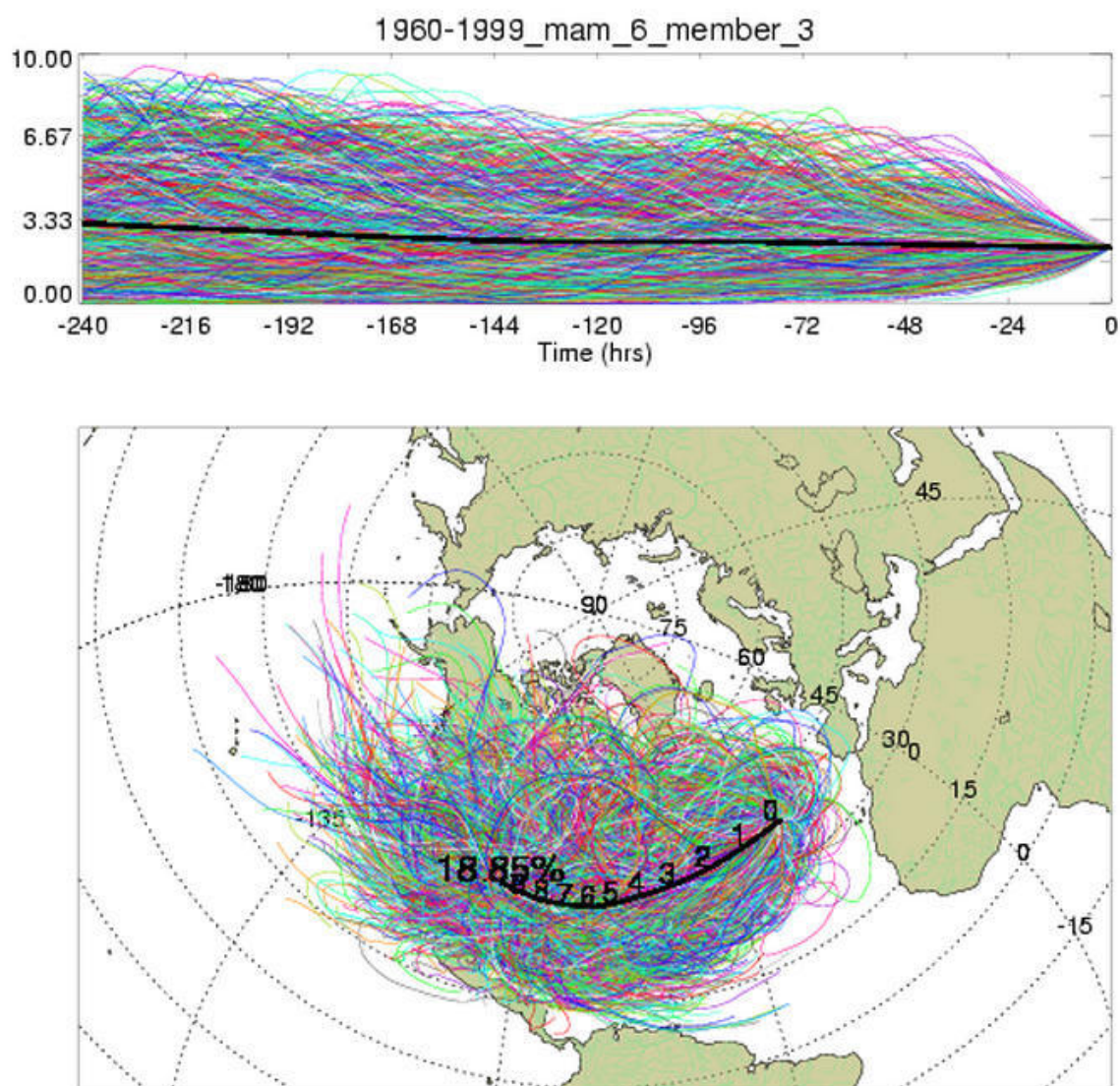


Figure C.11 Membership plot for springs, 1960-1999.

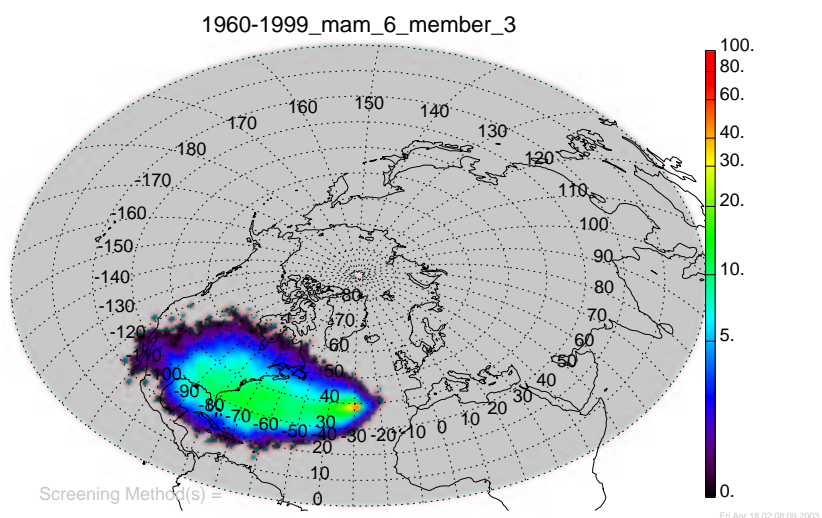


Figure C.12 Standard density membership plot for springs, 1960-1999.

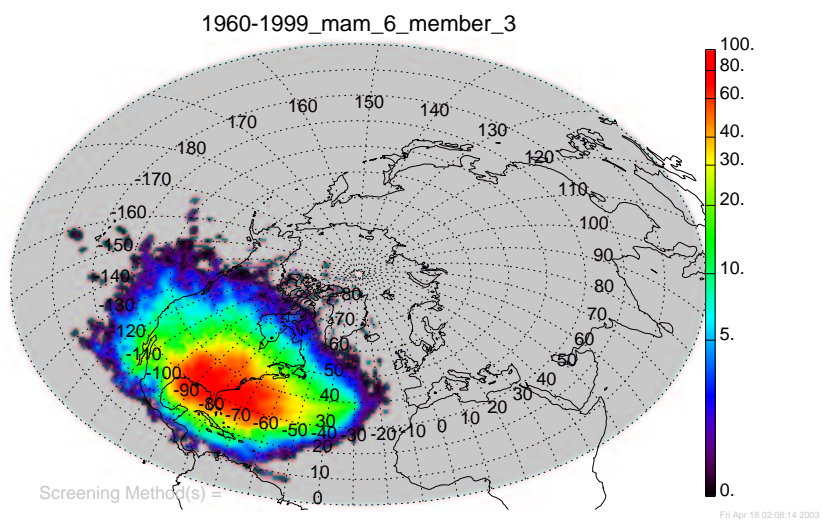


Figure C.13 Geometrically corrected density membership plot for springs, 1960-1999.

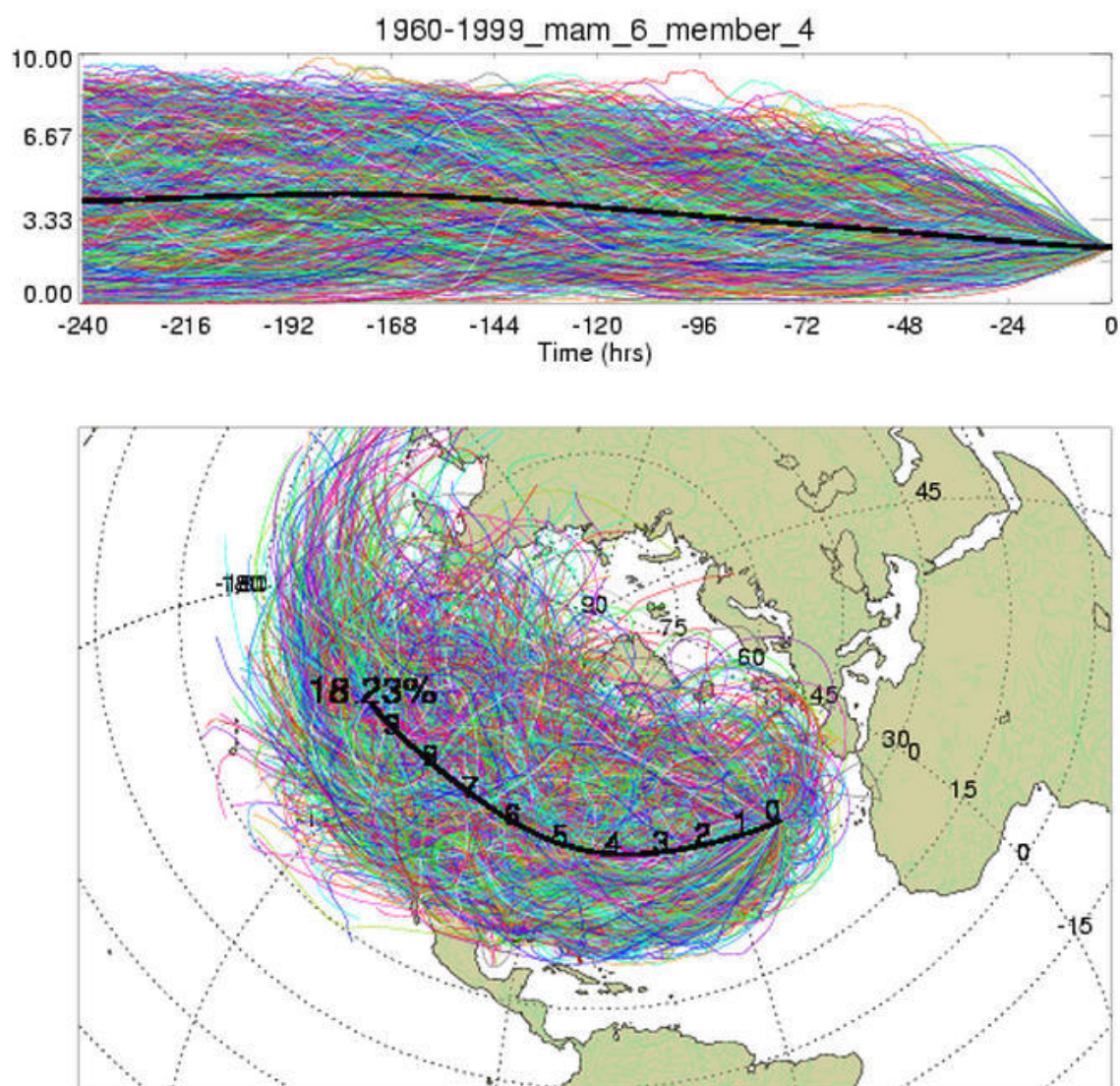


Figure C.14 Membership plot for springs, 1960-1999.

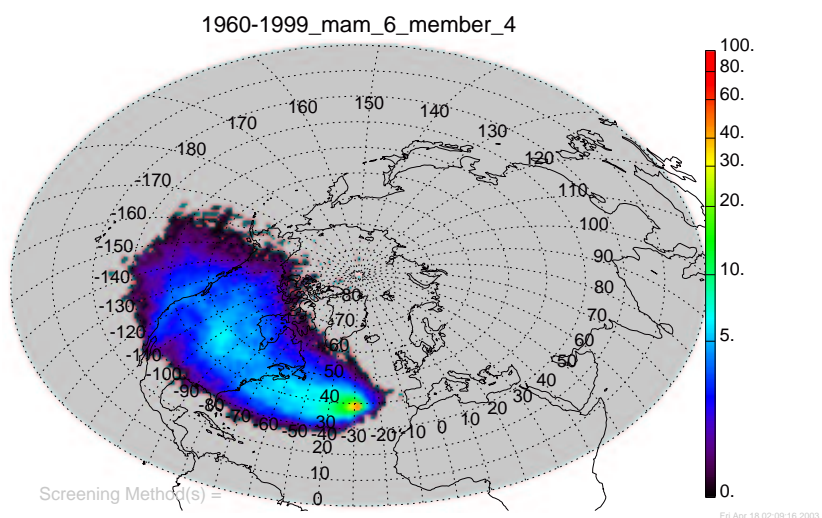


Figure C.15 Standard density membership plot for springs, 1960-1999.

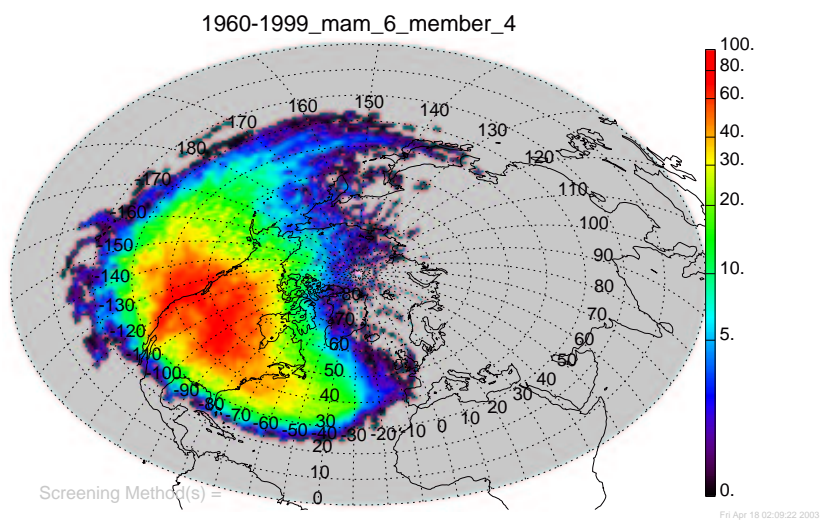


Figure C.16 Geometrically corrected density membership plot for springs, 1960-1999.

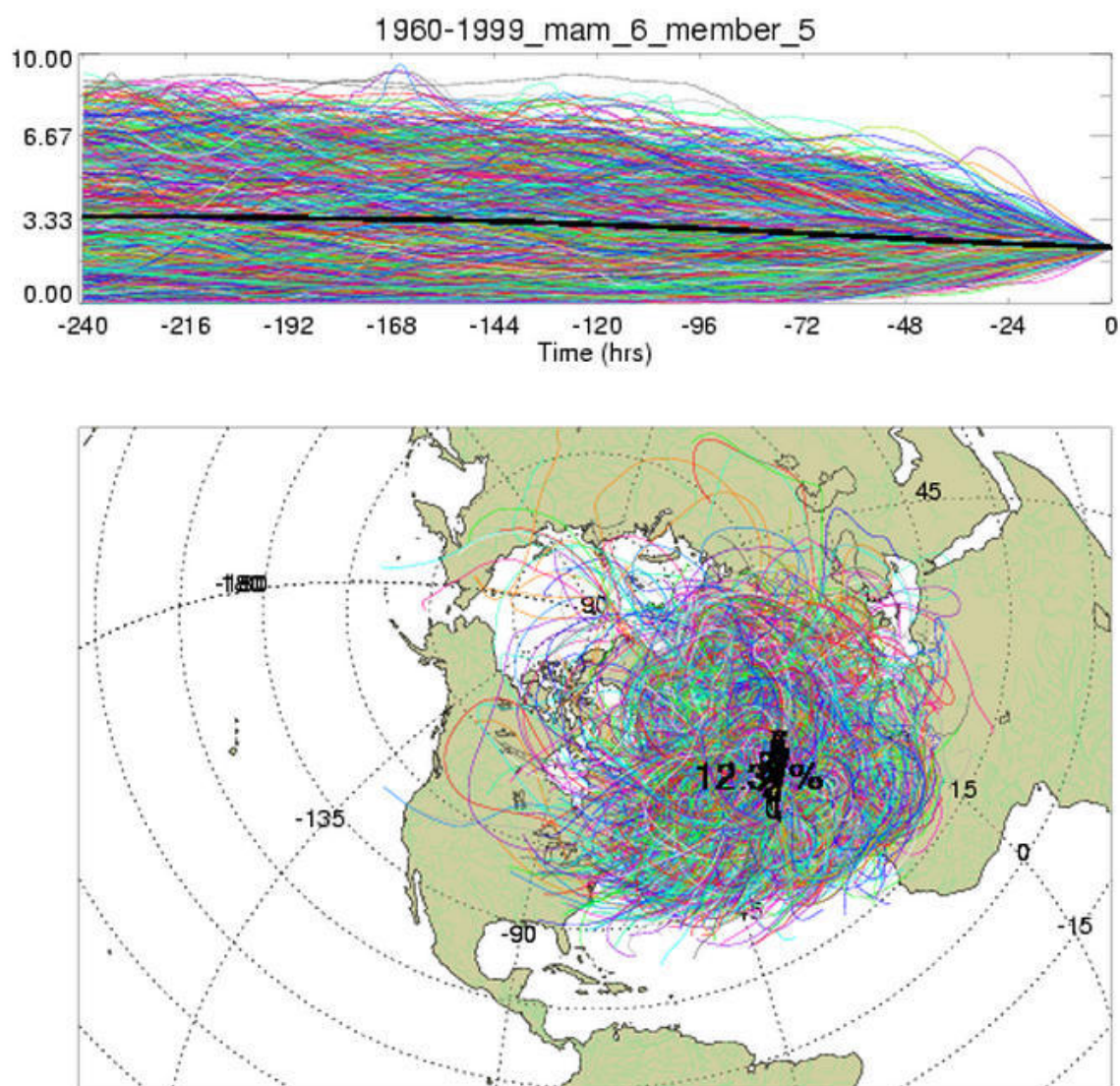


Figure C.17 Membership plot for springs, 1960-1999.

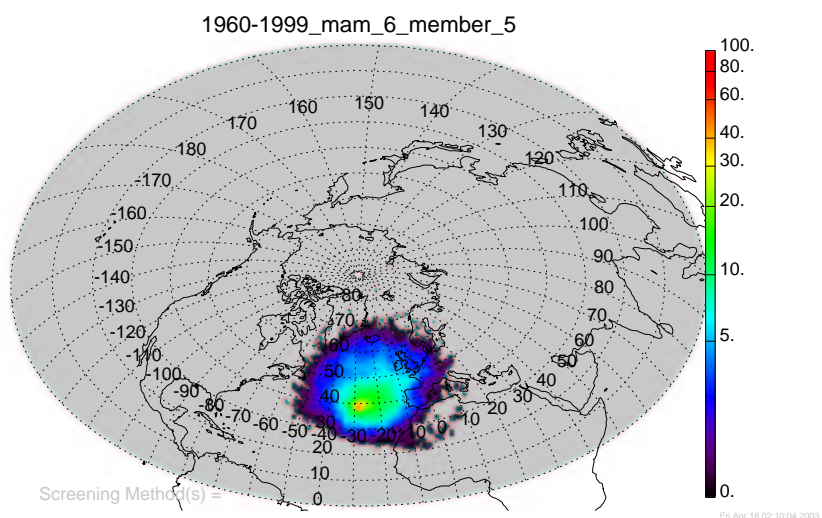


Figure C.18 Standard density membership plot for springs, 1960-1999.

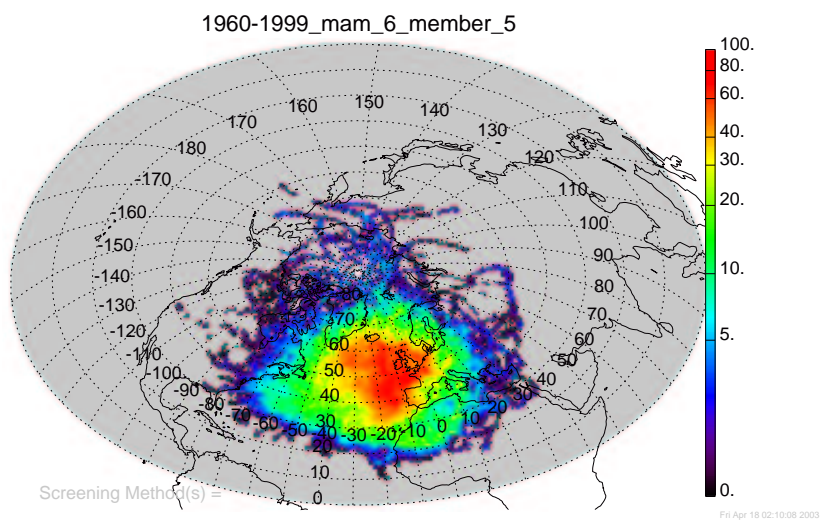


Figure C.19 Geometrically corrected density membership plot for springs, 1960-1999.

C.2 Membership plots for the positive NAO springs

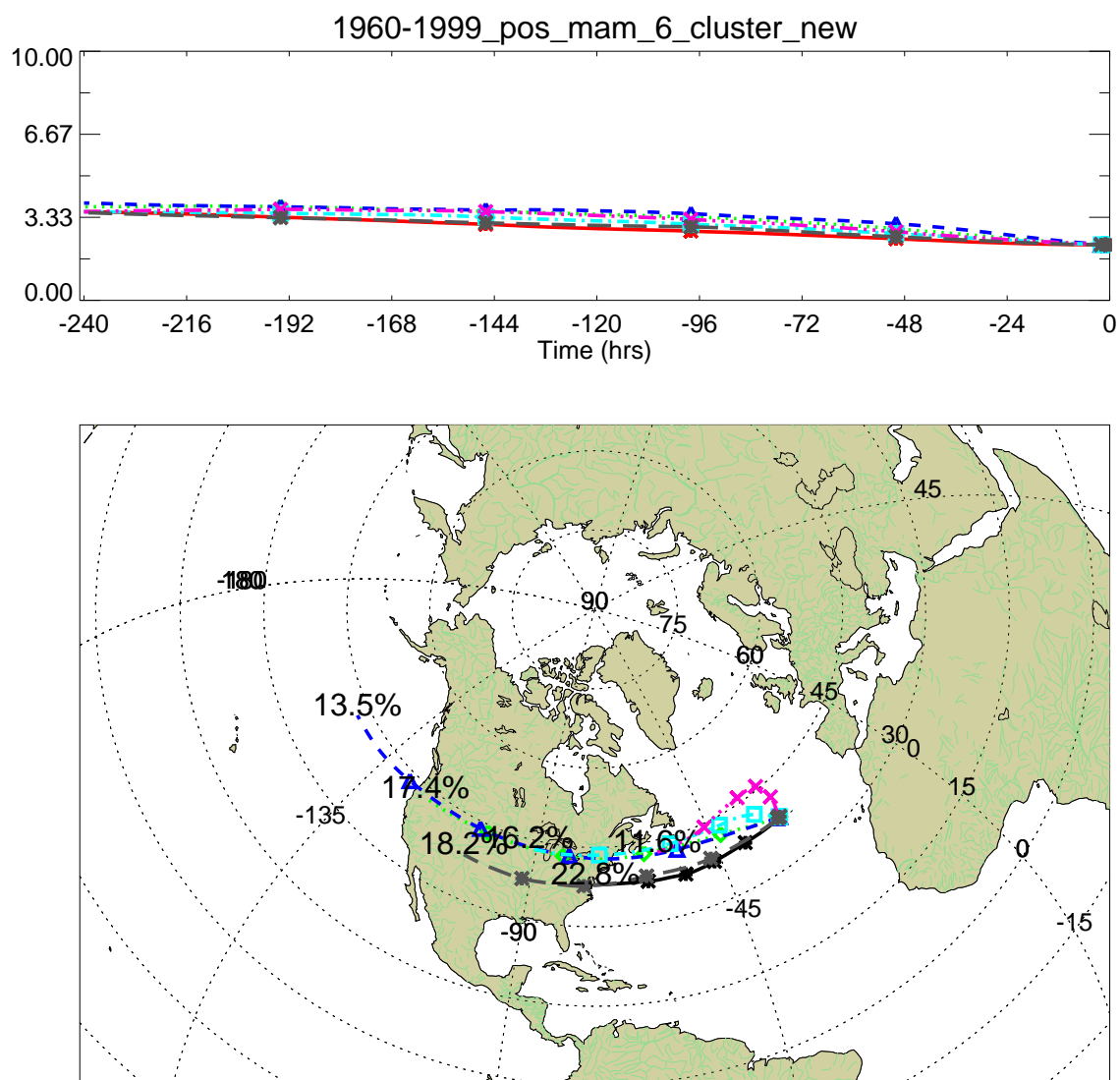


Figure C.20 Cluster plot for positive NAO springs, 1960-1999.

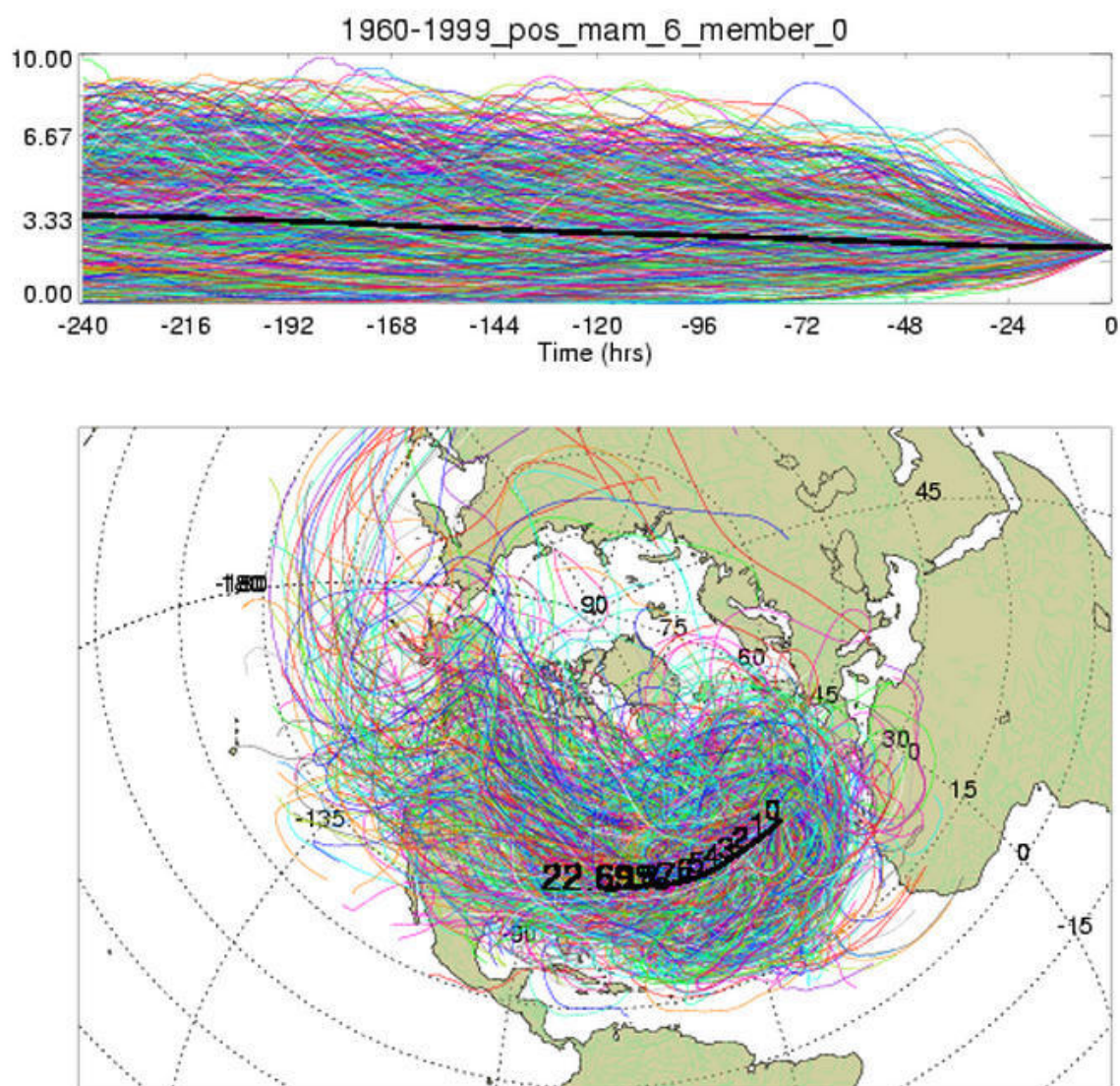


Figure C.21 Membership plot for springs with a positive NAOI value, 1960-1999.

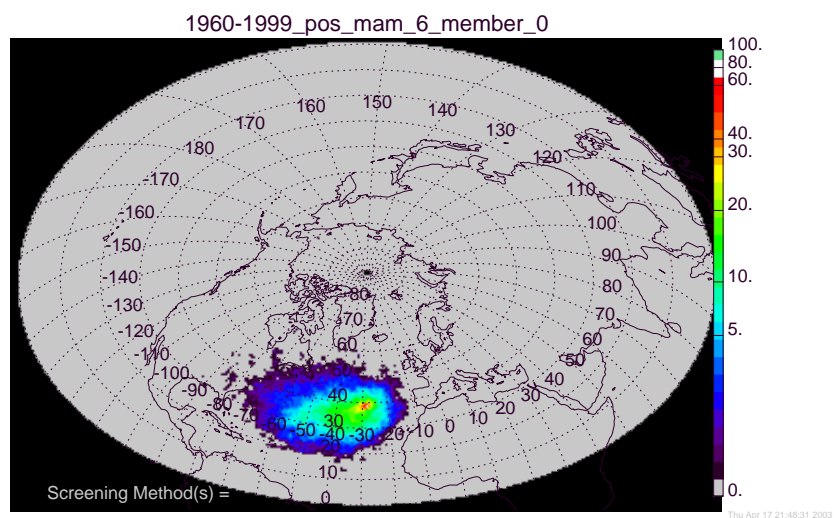


Figure C.22 Standard density membership plot for springs with a positive NAOI value, 1960-1999.

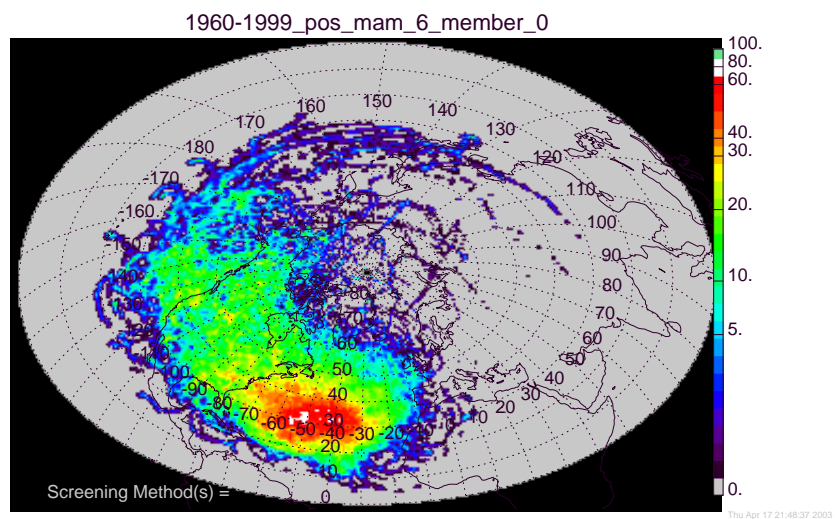


Figure C.23 Geometrically corrected density membership plot for springs with a positive NAOI value, 1960-1999.

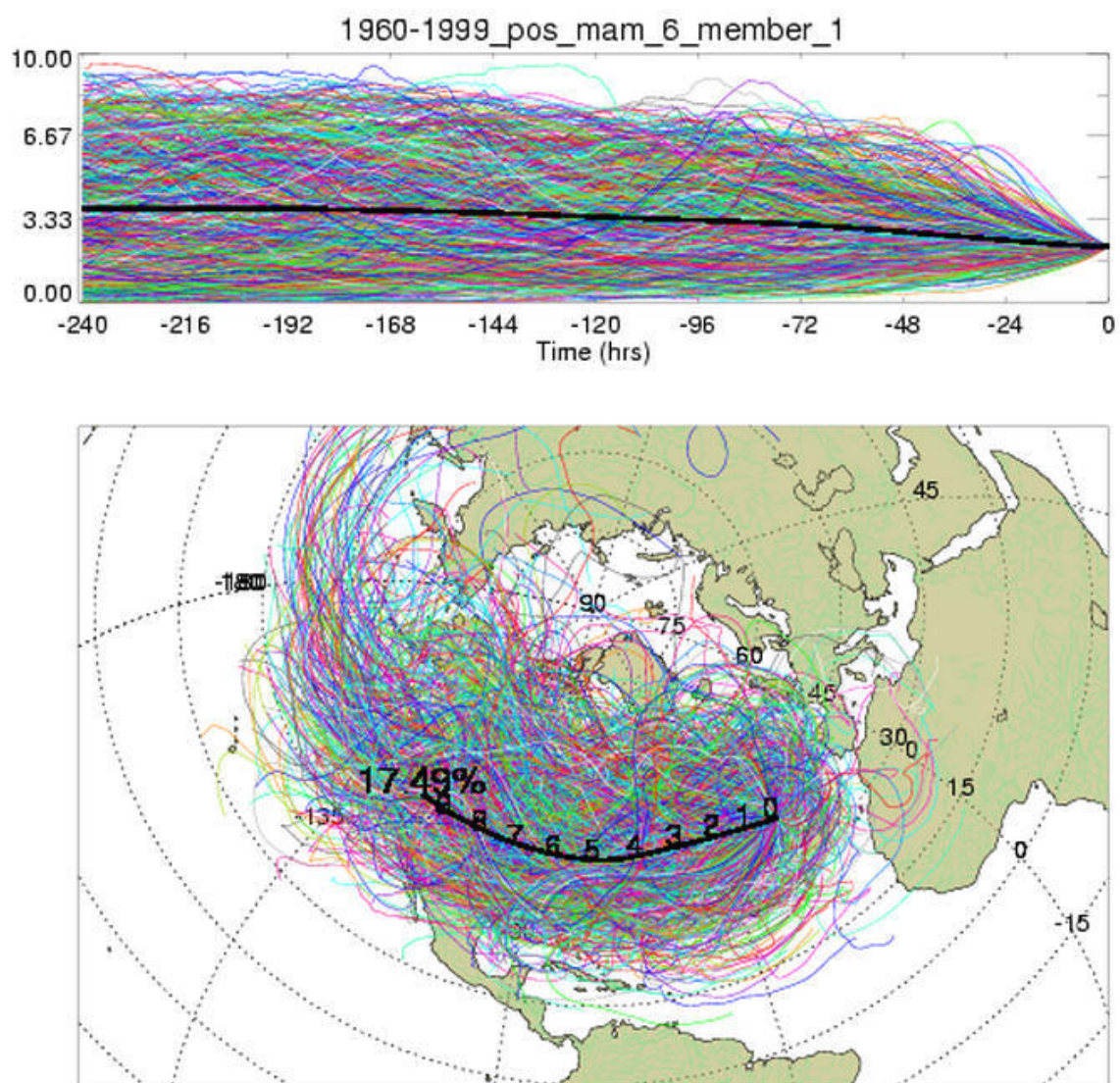


Figure C.24 Membership plot for springs with a positive NAOI value, 1960-1999.

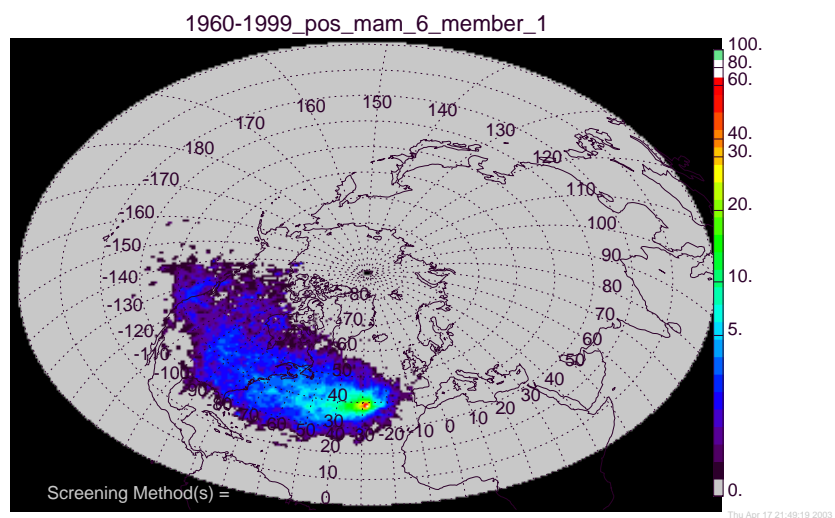


Figure C.25 Standard density membership plot for springs with a positive NAOI value, 1960-1999.

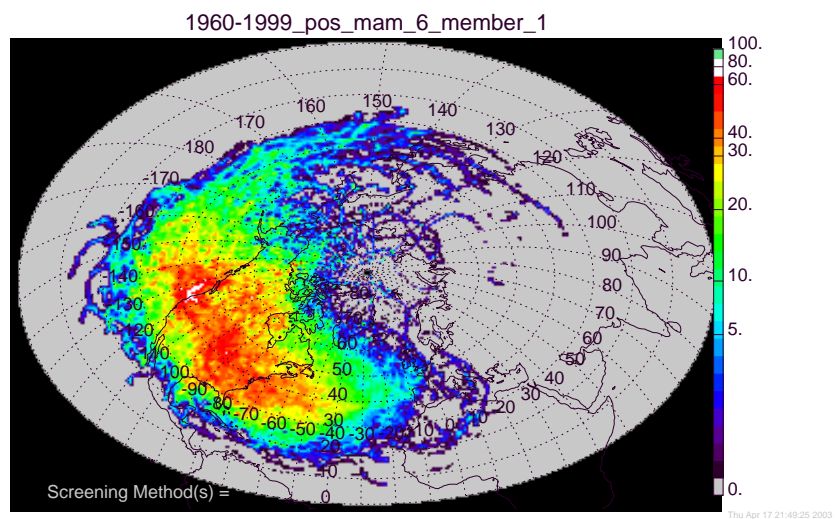


Figure C.26 Geometrically corrected density membership plot for springs with a positive NAOI value, 1960-1999.

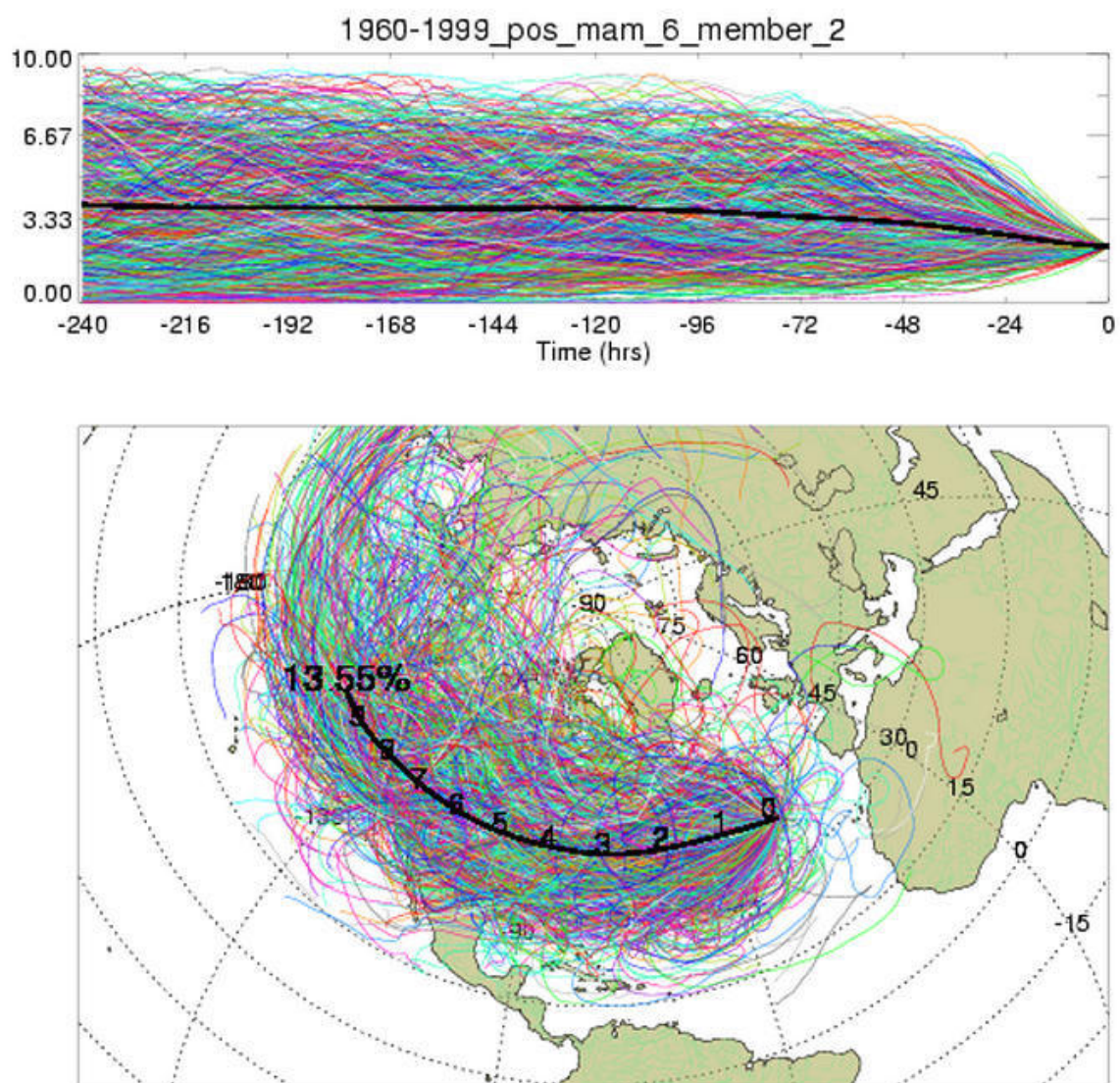


Figure C.27 Membership plot for springs with a positive NAOI value, 1960-1999.

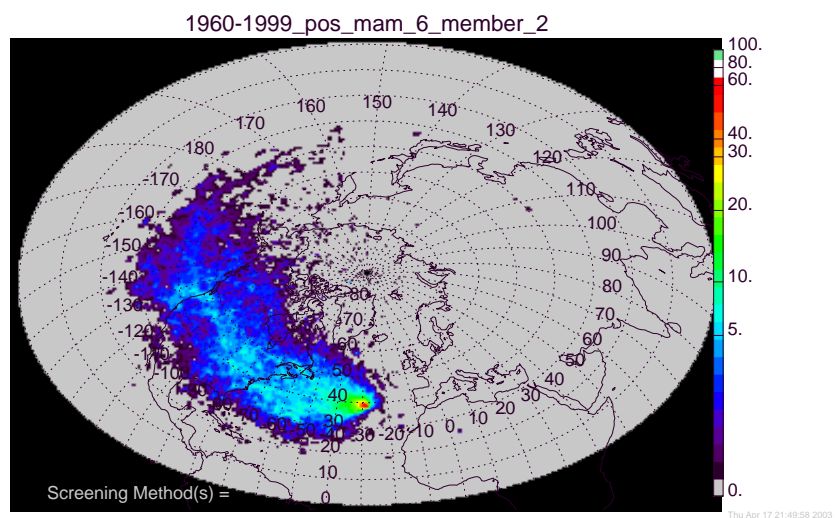


Figure C.28 Standard density membership plot for springs with a positive NAOI value, 1960-1999.

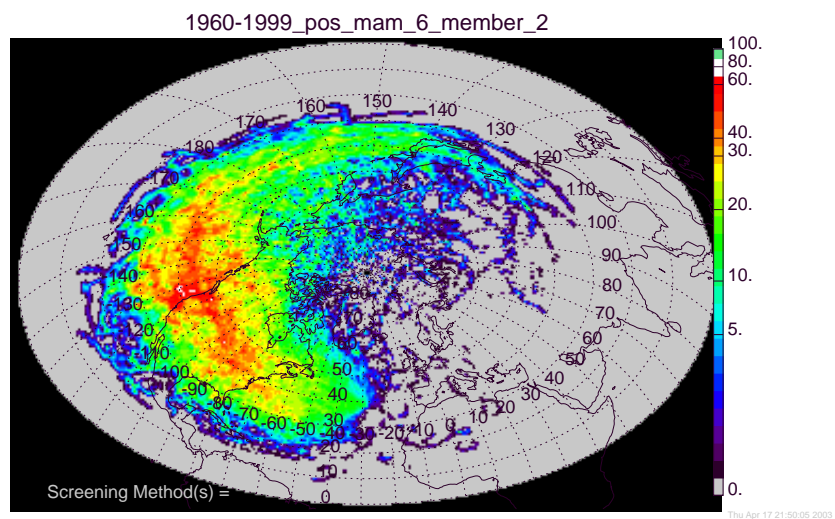


Figure C.29 Geometrically corrected density membership plot for springs with a positive NAOI value, 1960-1999.

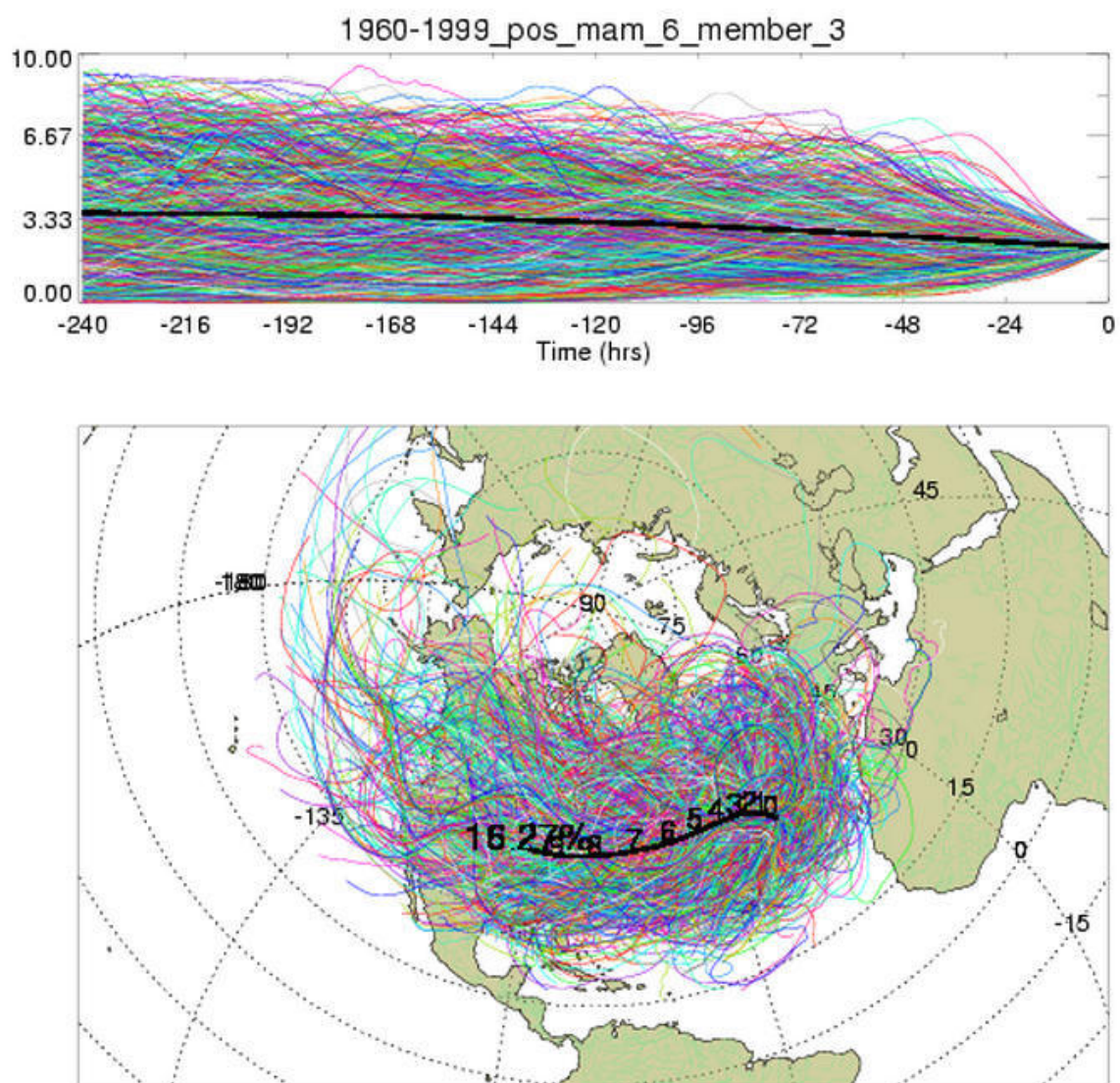


Figure C.30 Membership plot for springs with a positive NAOI value, 1960-1999.

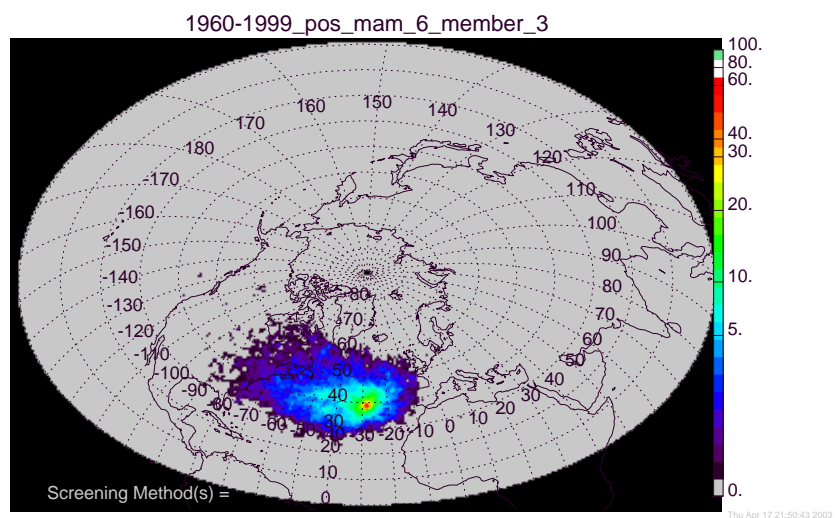


Figure C.31 Standard density membership plot for springs with a positive NAOI value, 1960-1999.

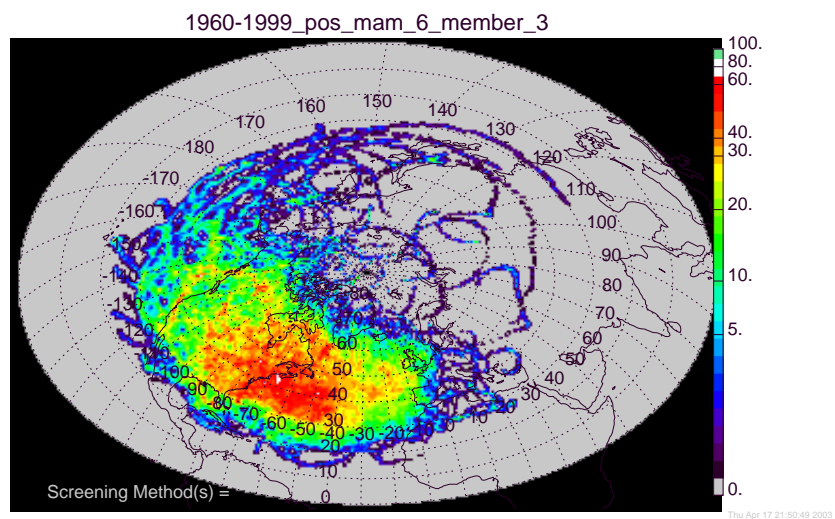


Figure C.32 Geometrically corrected density membership plot for springs with a positive NAOI value, 1960-1999.

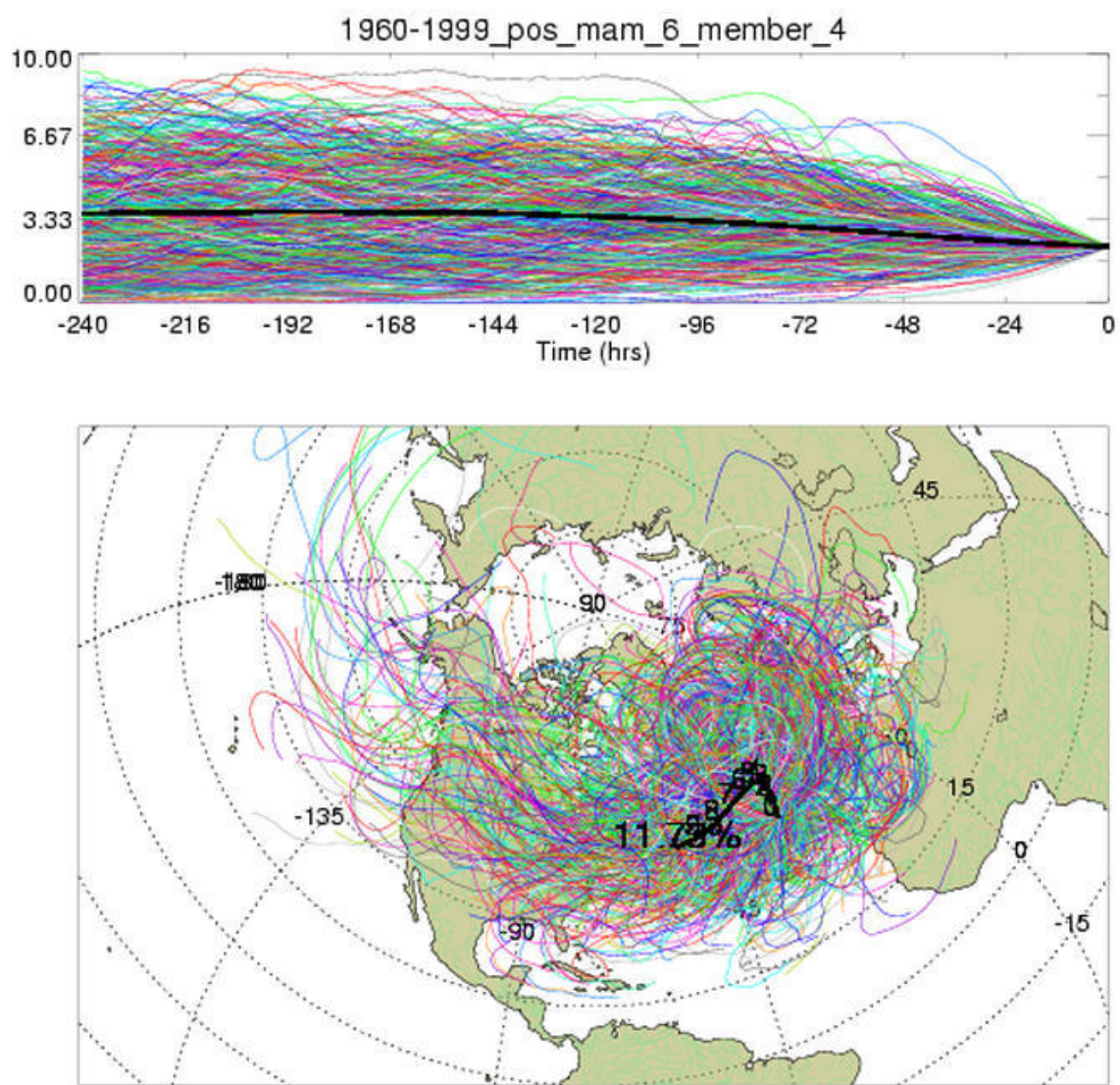


Figure C.33 Membership plot for springs with a positive NAOI value, 1960-1999.

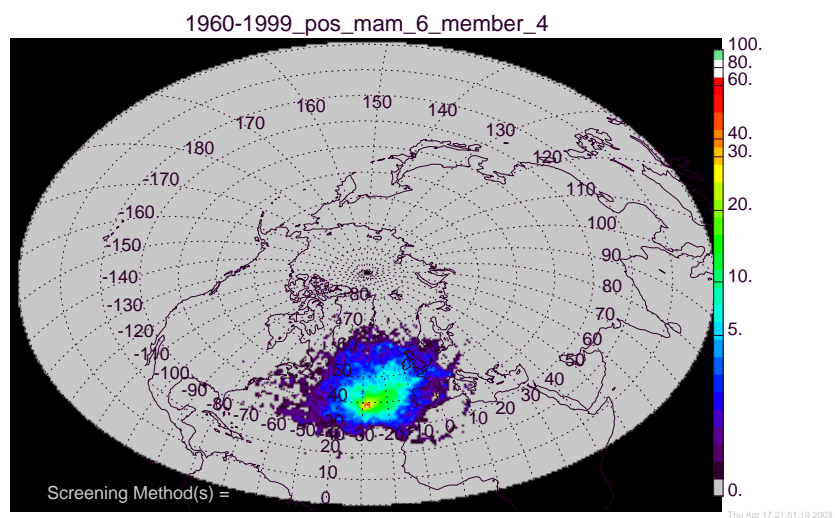


Figure C.34 Standard density membership plot for springs with a positive NAOI value, 1960-1999.

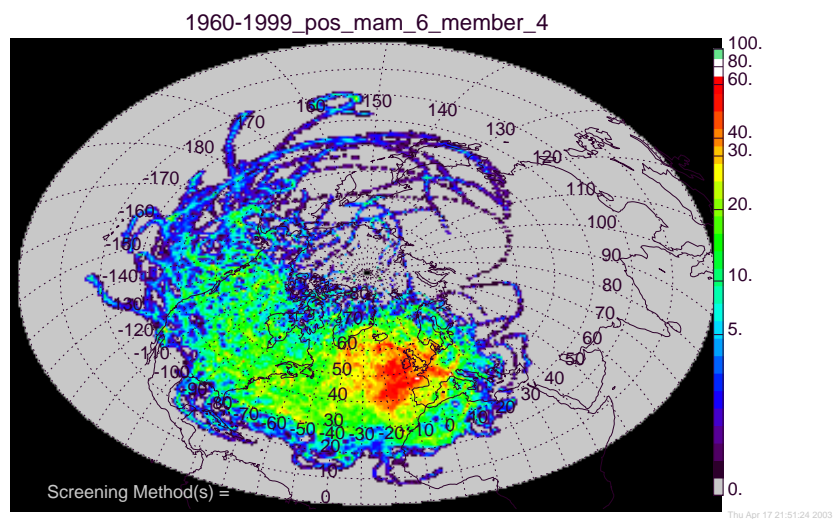


Figure C.35 Geometrically corrected density membership plot for springs with a positive NAOI value, 1960-1999.

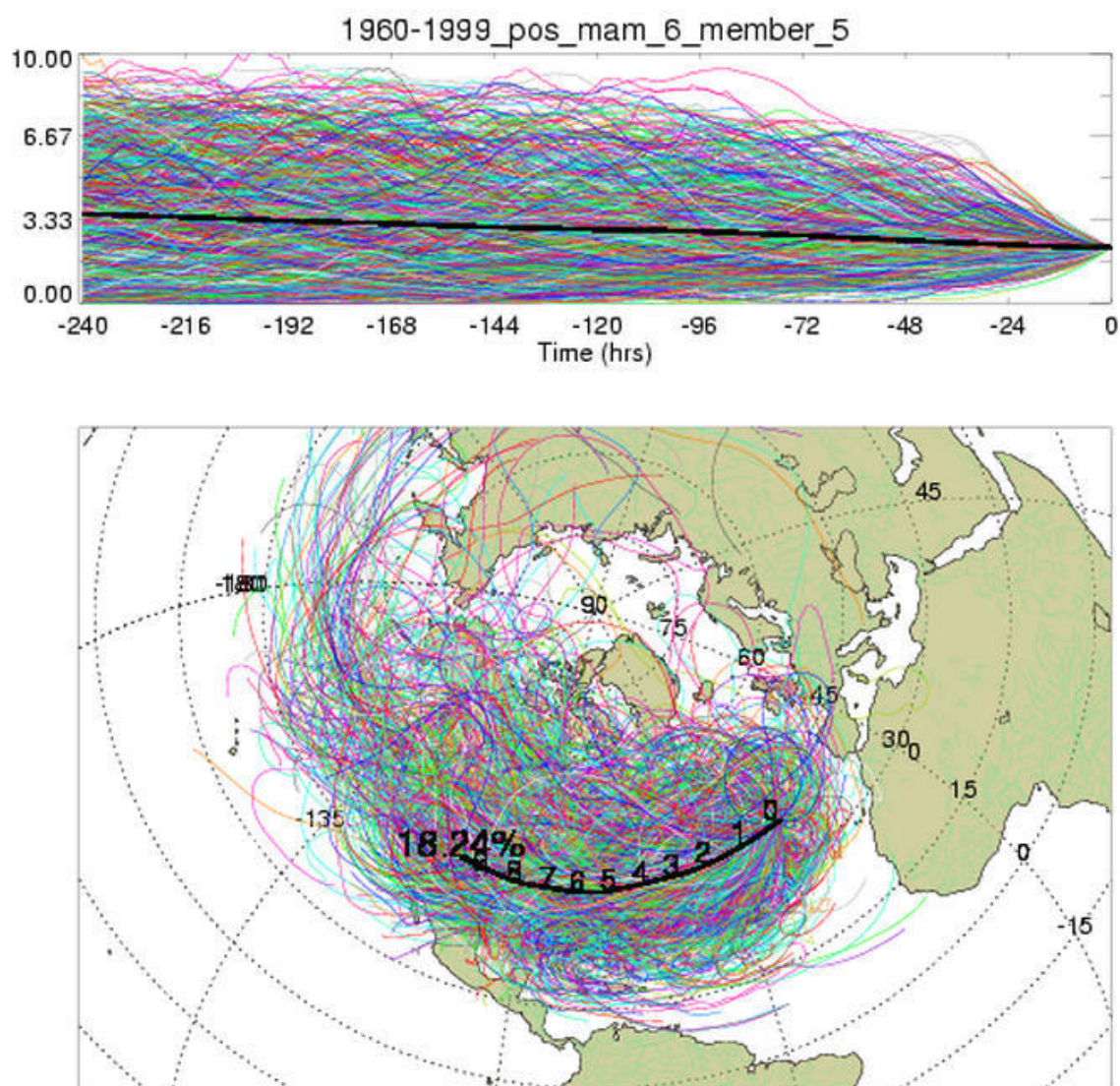


Figure C.36 Membership plot for springs with a positive NAOI value, 1960-1999.

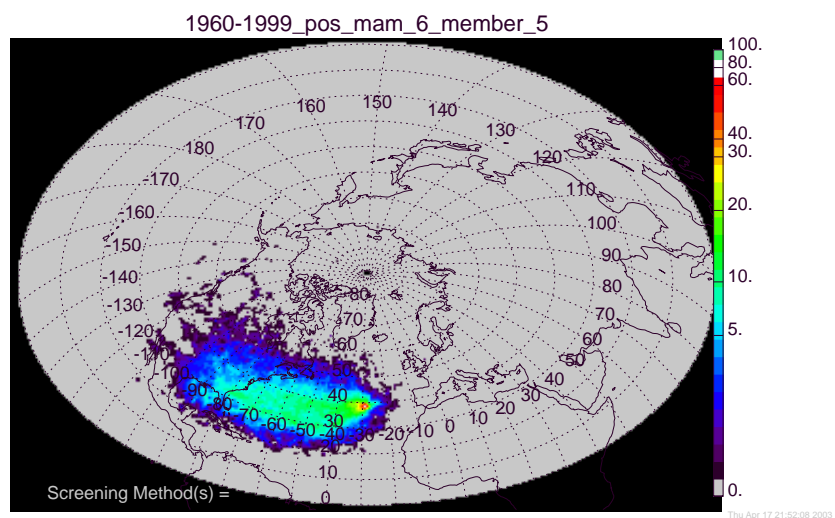


Figure C.37 Standard density membership plot for springs with a positive NAOI value, 1960-1999.

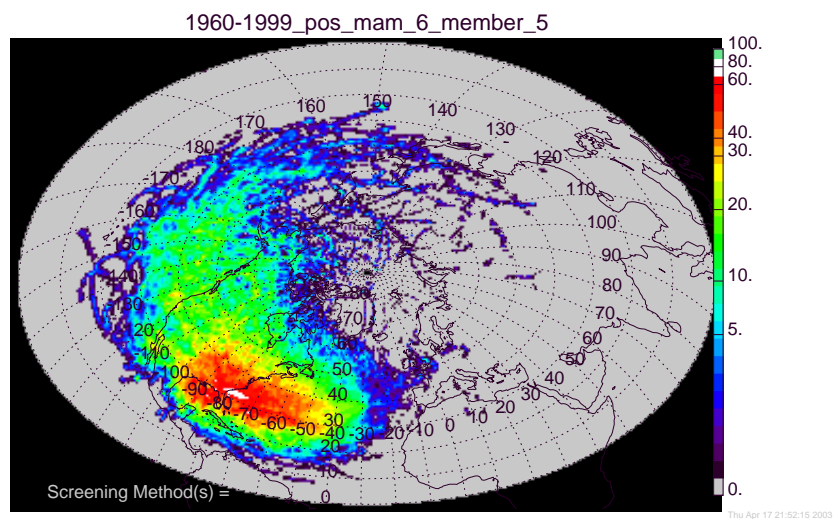


Figure C.38 Geometrically corrected density membership plot for springs with a positive NAOI value, 1960-1999.

C.3 Membership plots for the negative NAO springs

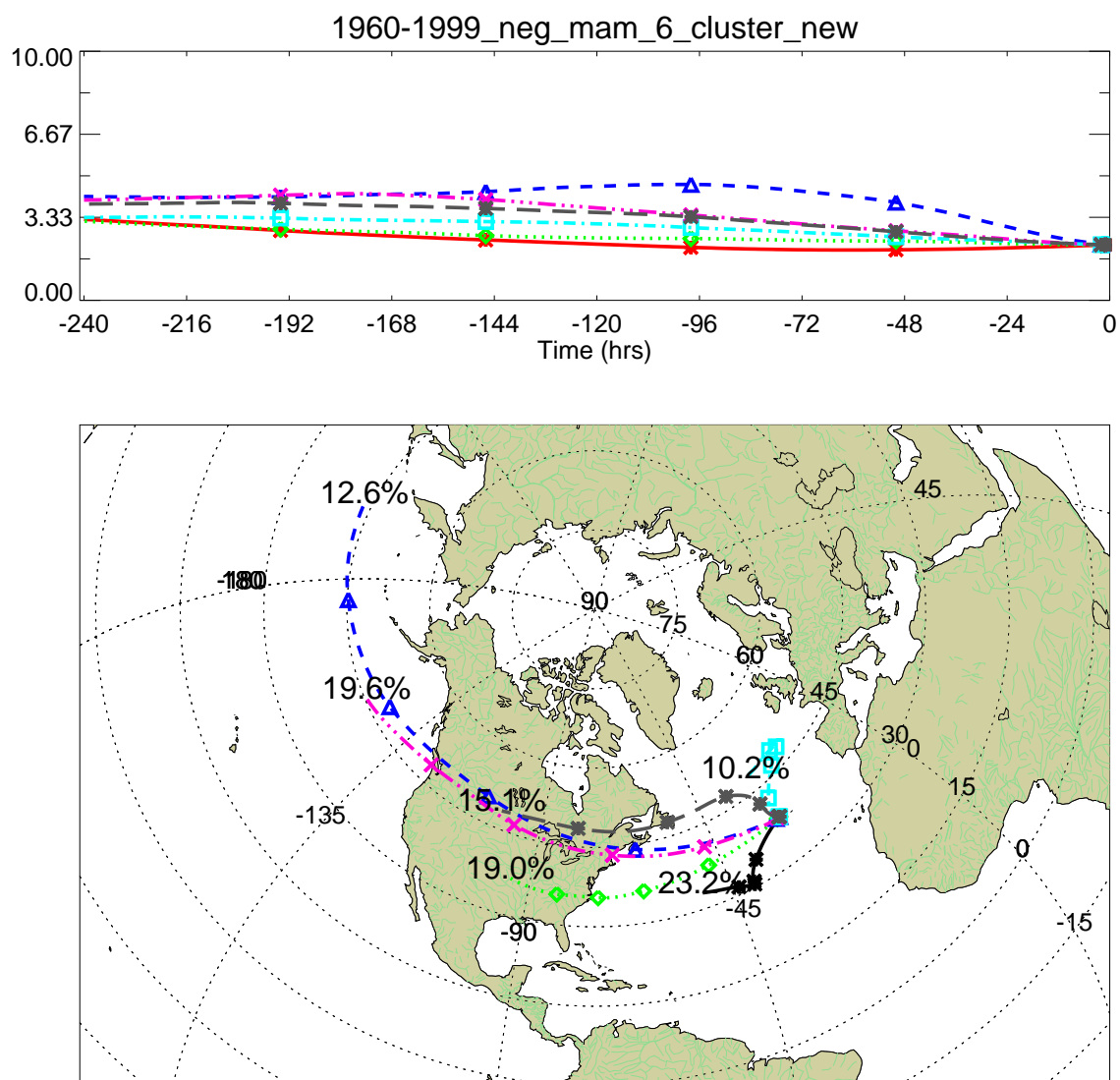


Figure C.39 Cluster plot for negative NAO springs, 1960-1999.

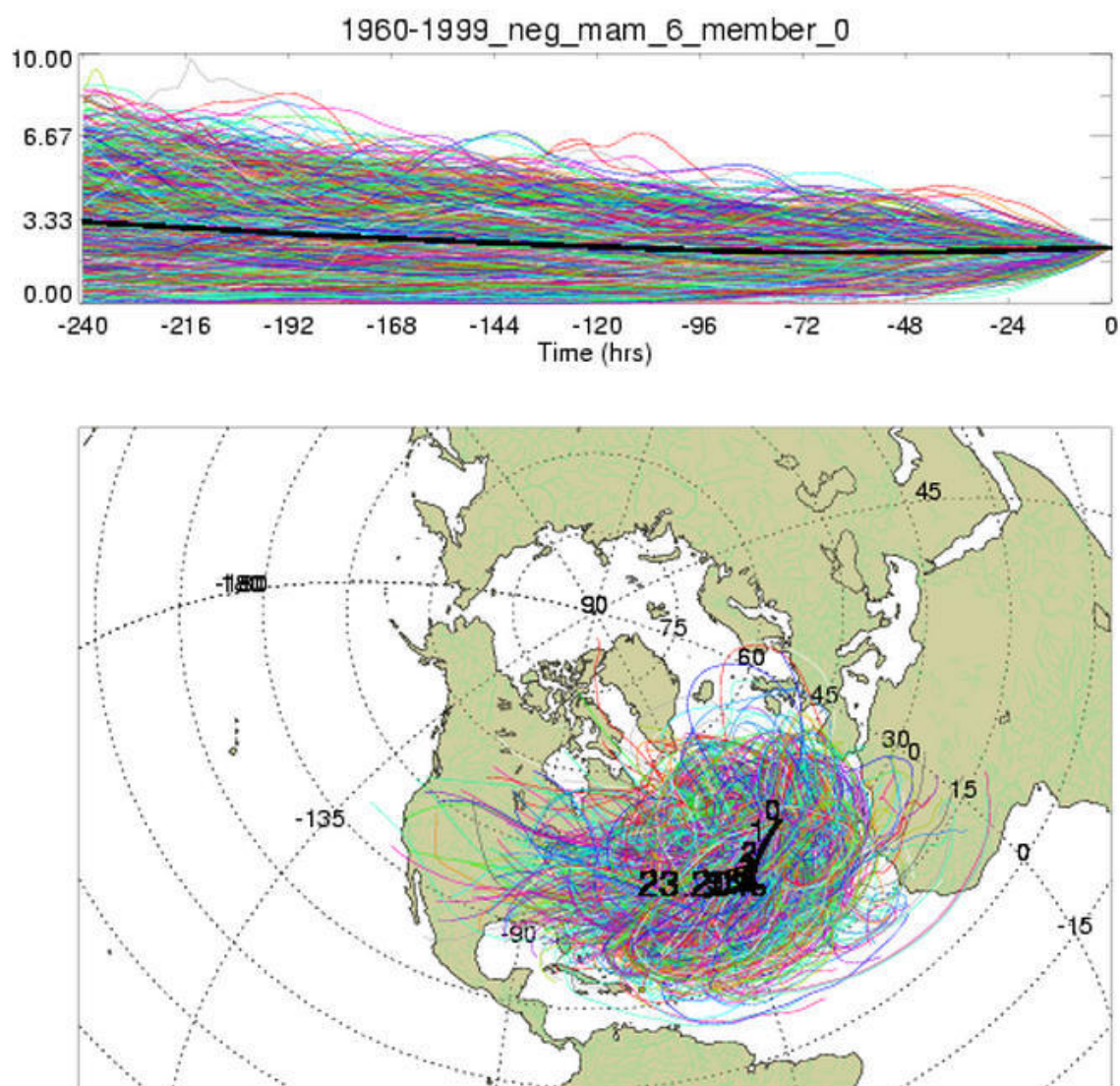


Figure C.40 Membership plot for springs with a negative NAOI value, 1960-1999.

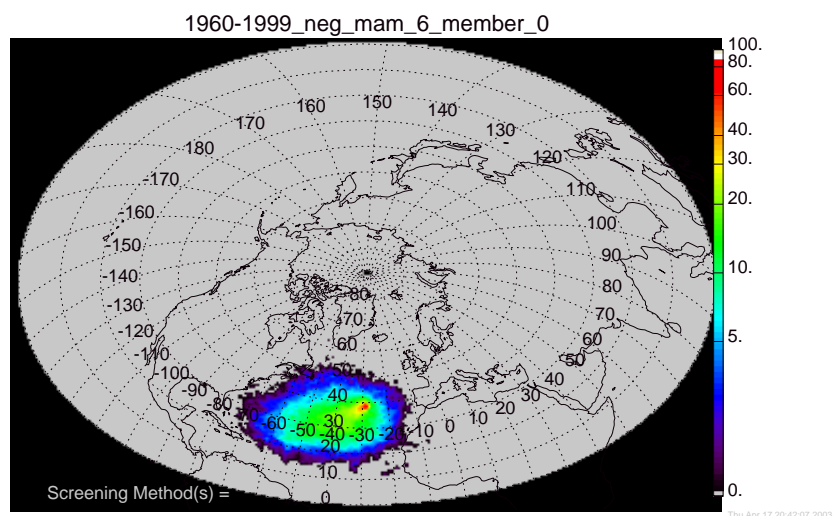


Figure C.41 Standard density membership plot for springs with a negative NAOI value, 1960-1999.

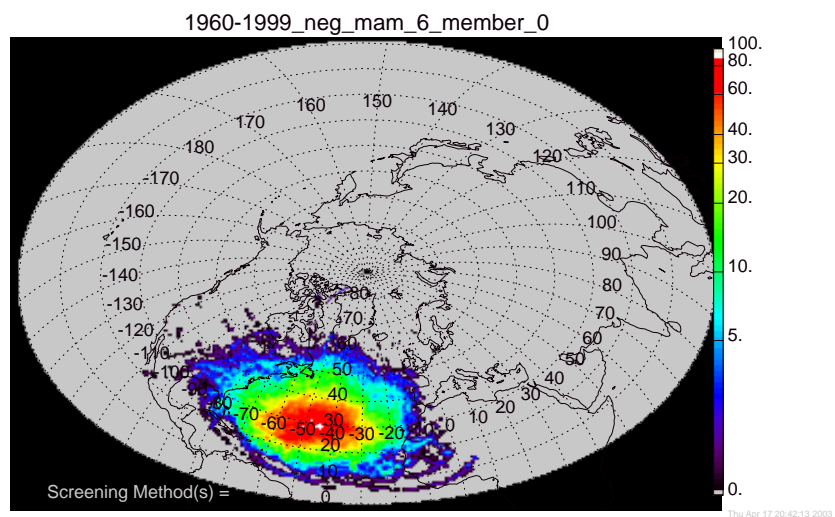


Figure C.42 Geometrically corrected density membership plot for springs with a negative NAOI value, 1960-1999.

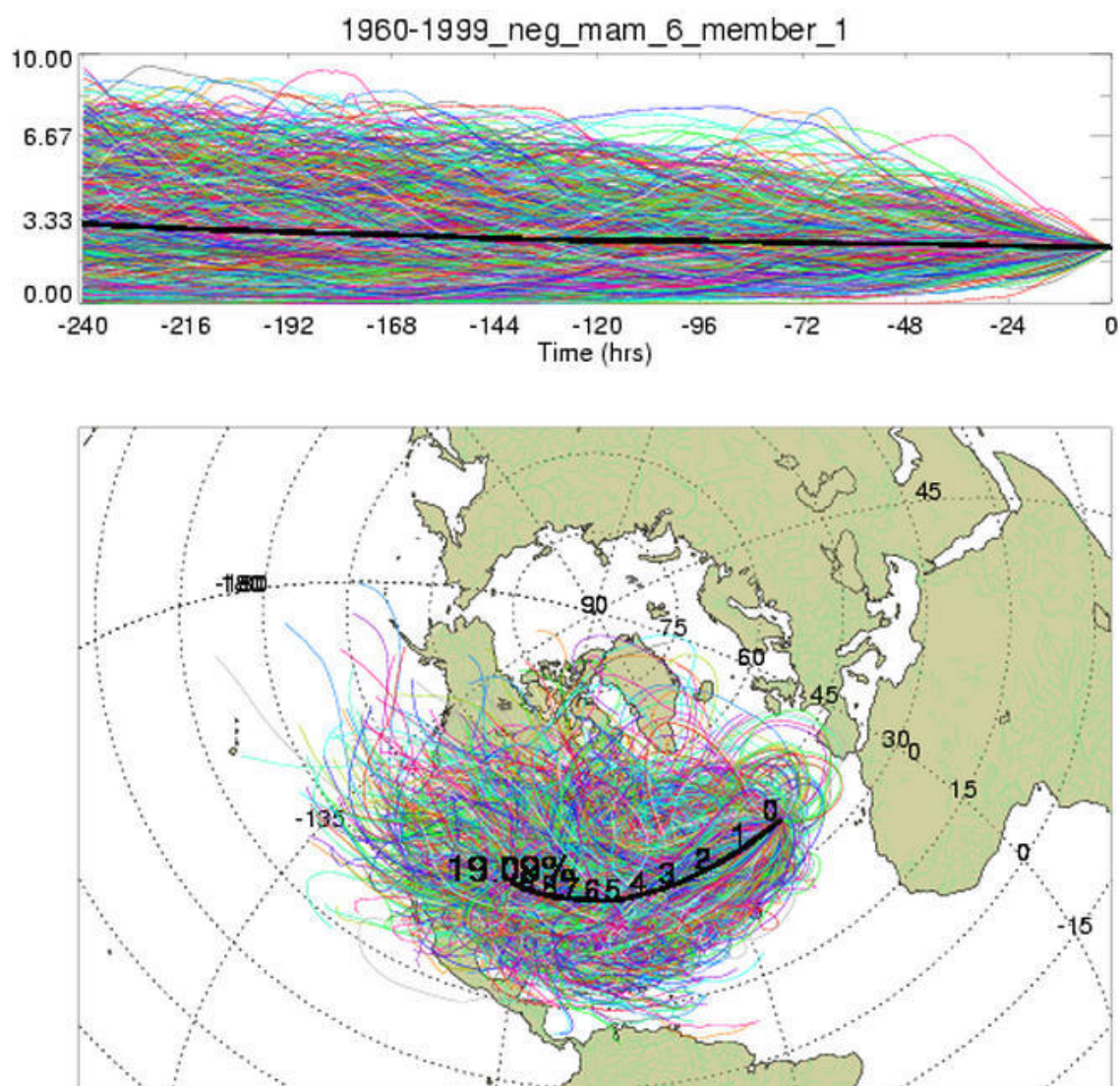


Figure C.43 Membership plot for springs with a negative NAOI value, 1960-1999.

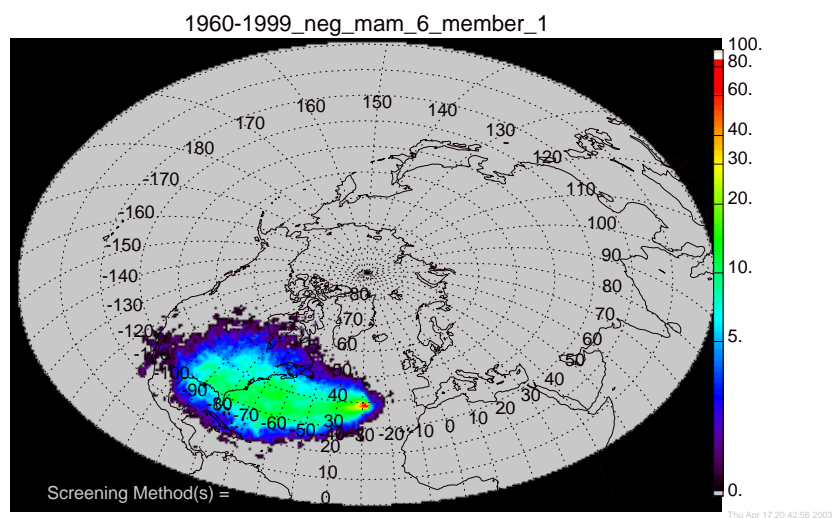


Figure C.44 Standard density membership plot for springs with a negative NAOI value, 1960-1999.

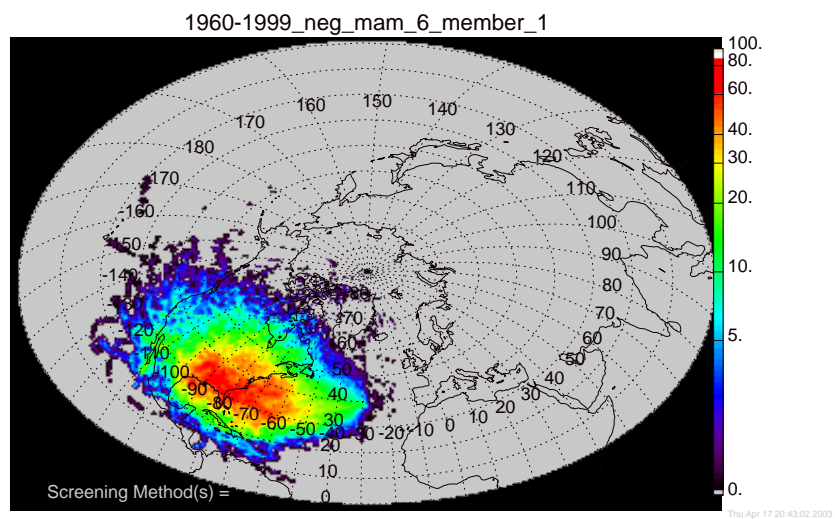


Figure C.45 Geometrically corrected density membership plot for springs with a negative NAOI value, 1960-1999.

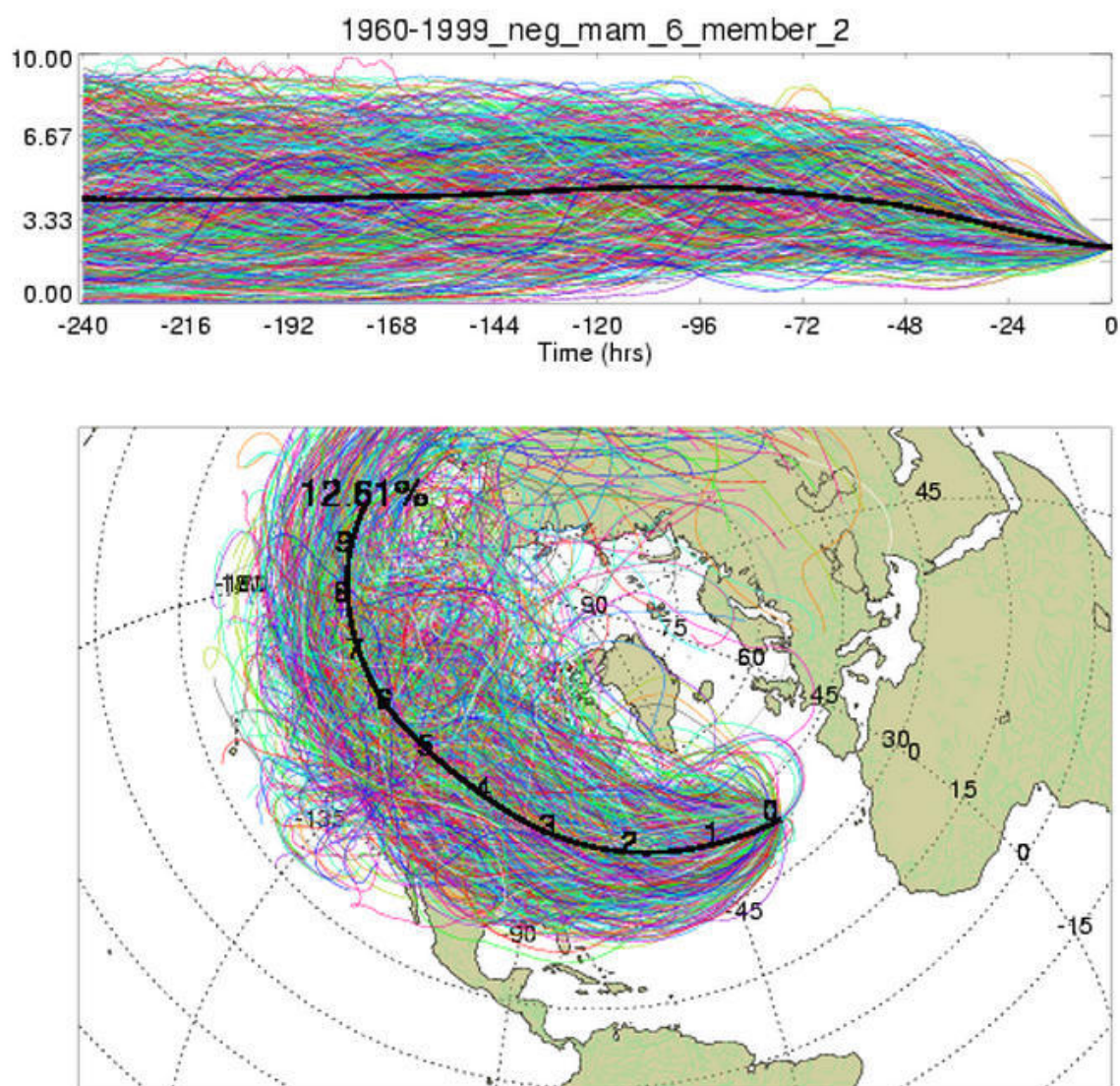


Figure C.46 Membership plot for springs with a negative NAOI value, 1960-1999.

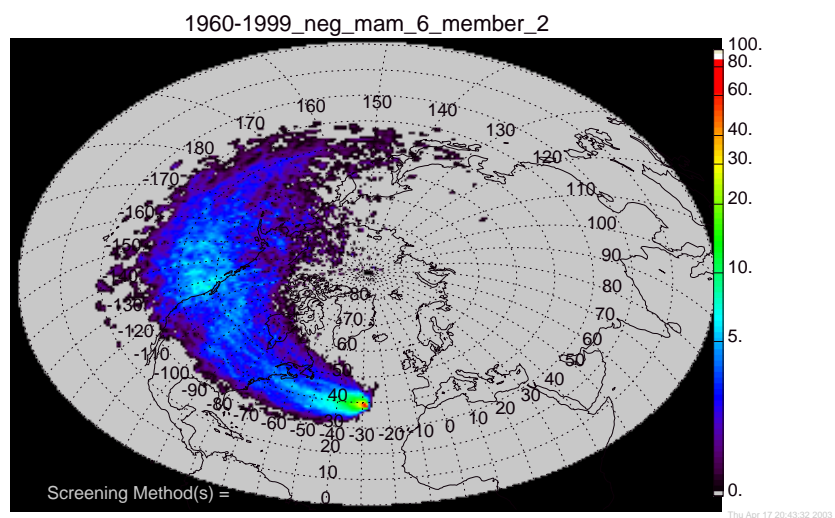


Figure C.47 Standard density membership plot for springs with a negative NAOI value, 1960-1999.

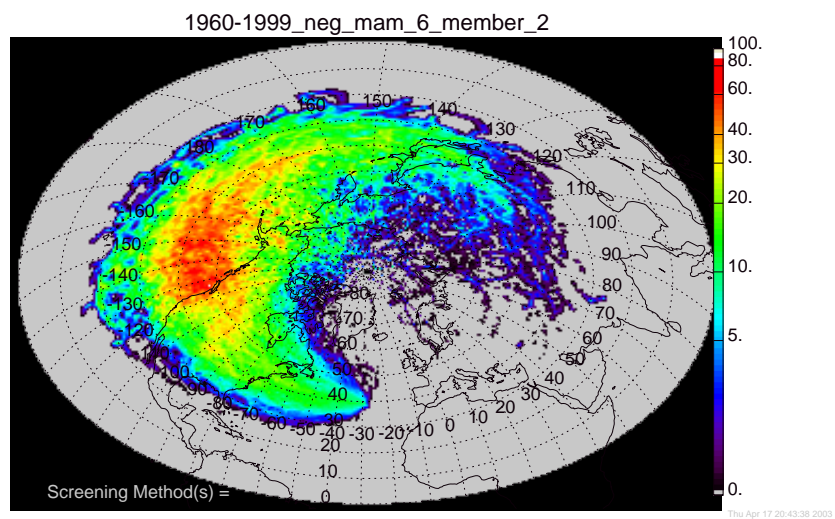


Figure C.48 Geometrically corrected density membership plot for springs with a negative NAOI value, 1960-1999.

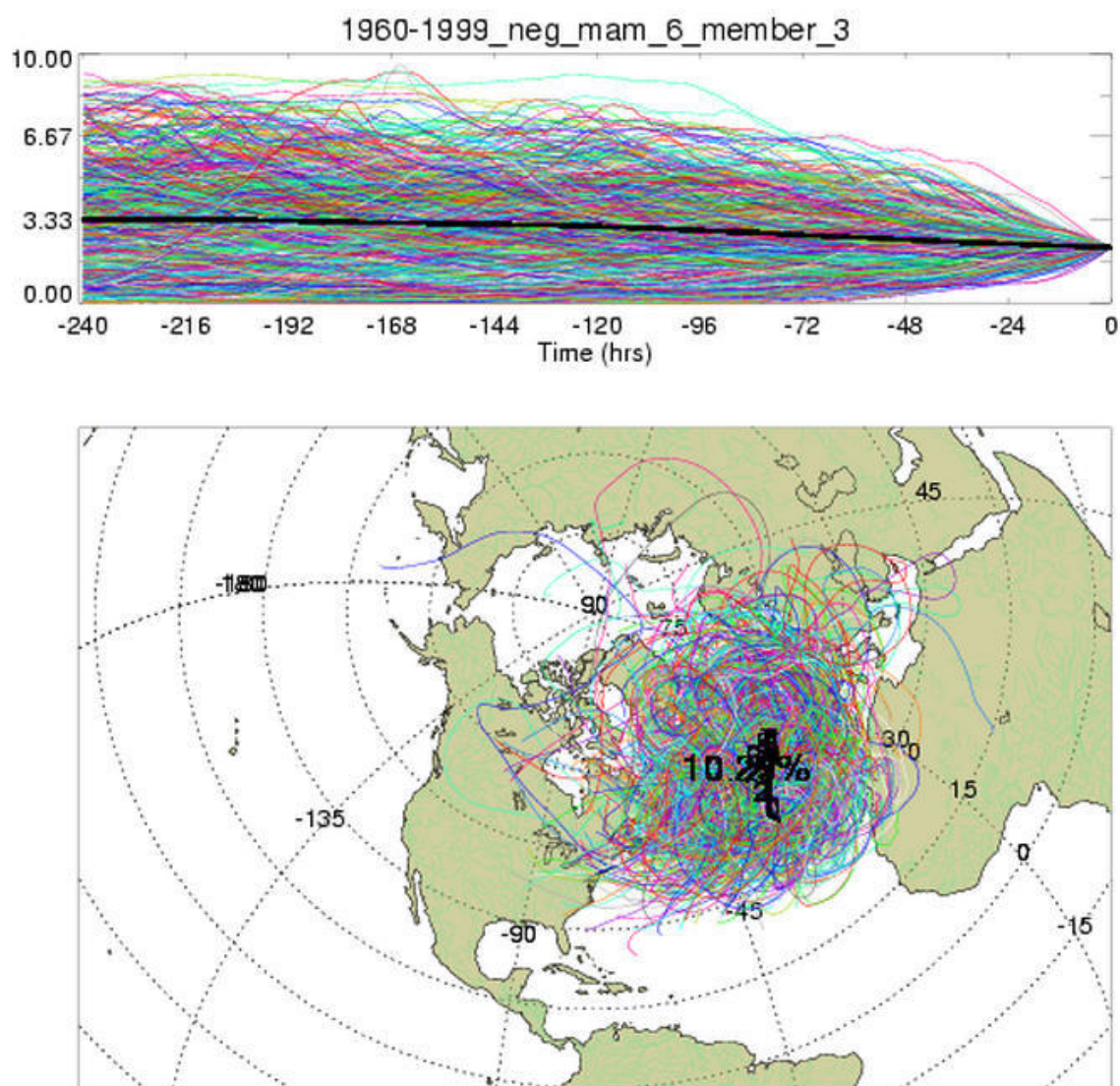


Figure C.49 Membership plot for springs with a negative NAOI value, 1960-1999.

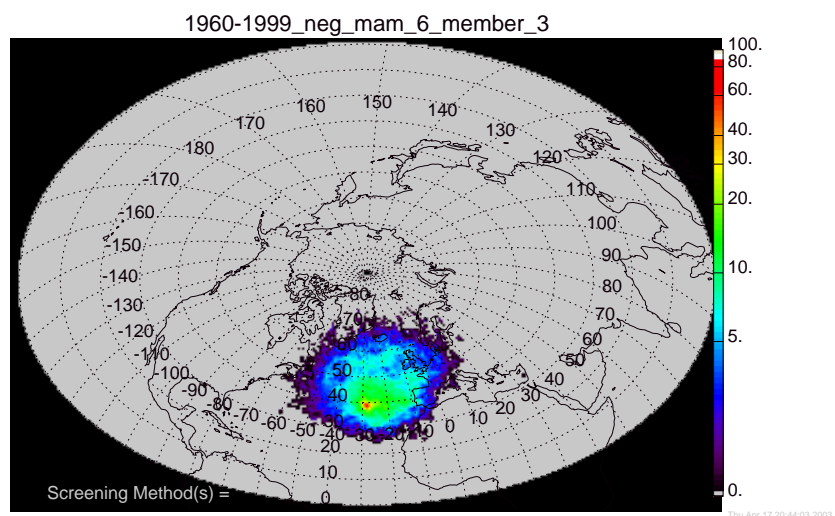


Figure C.50 Standard density membership plot for springs with a negative NAOI value, 1960-1999.

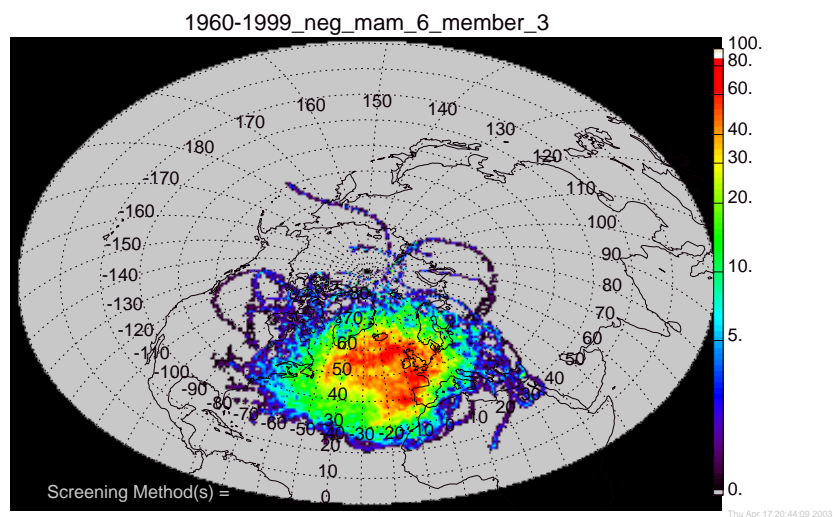


Figure C.51 Geometrically corrected density membership plot for springs with a negative NAOI value, 1960-1999.

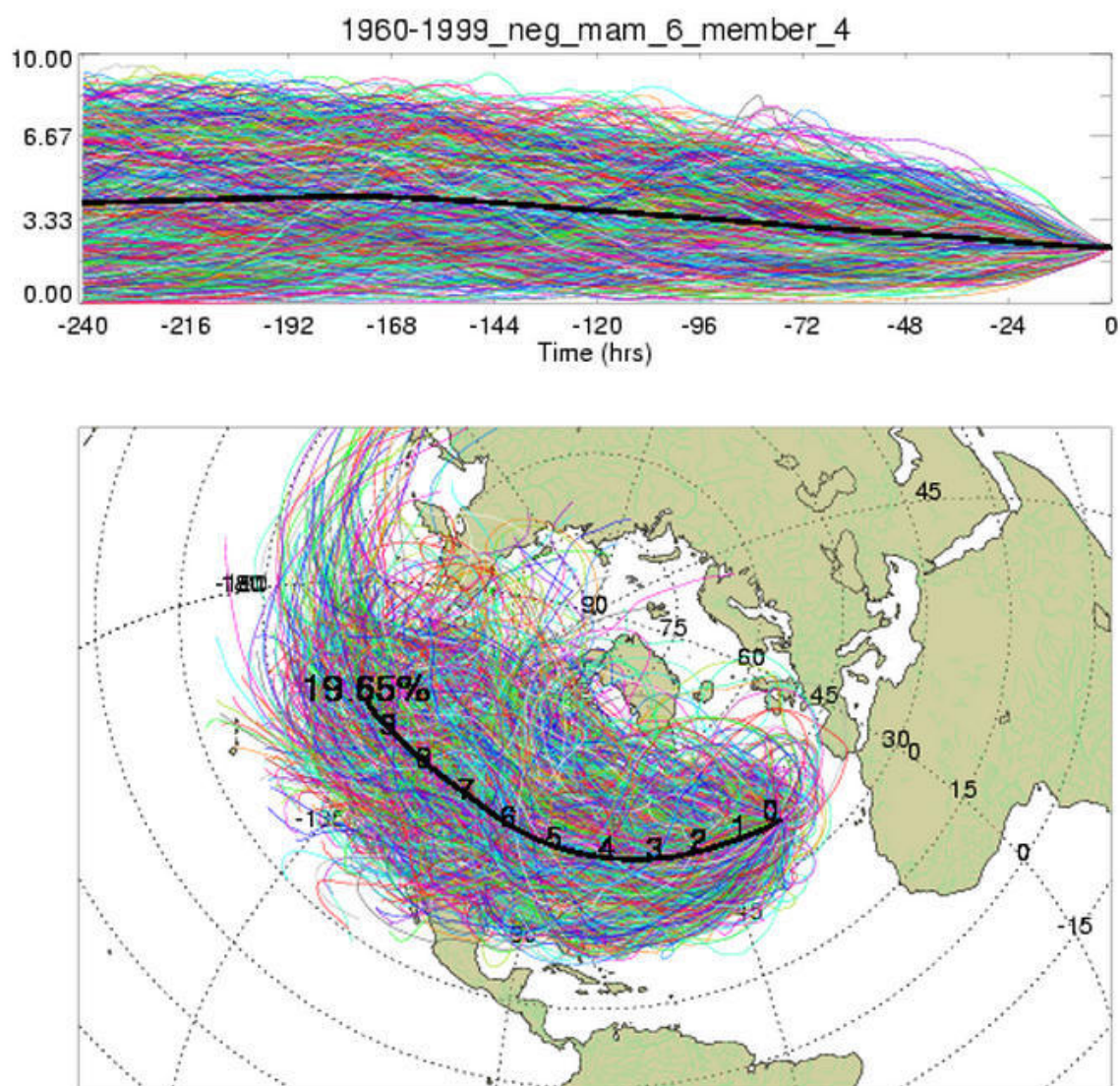


Figure C.52 Membership plot for springs with a negative NAOI value, 1960-1999.

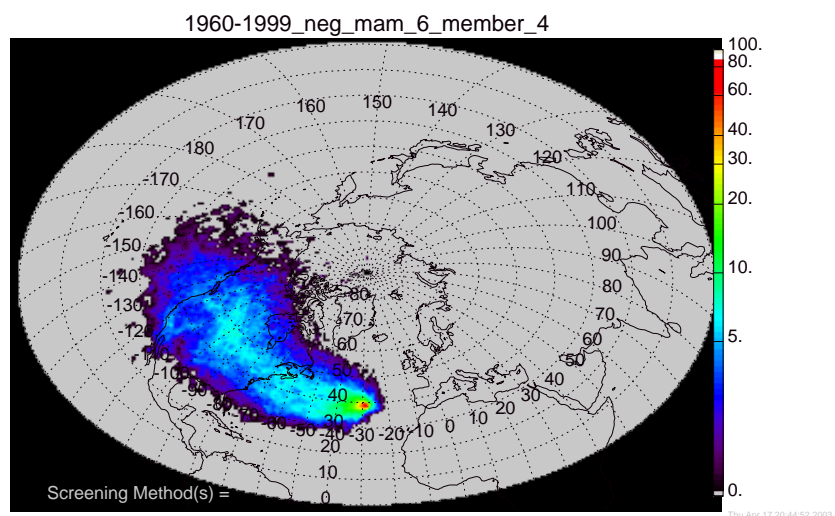


Figure C.53 Standard density membership plot for springs with a negative NAOI value, 1960-1999.

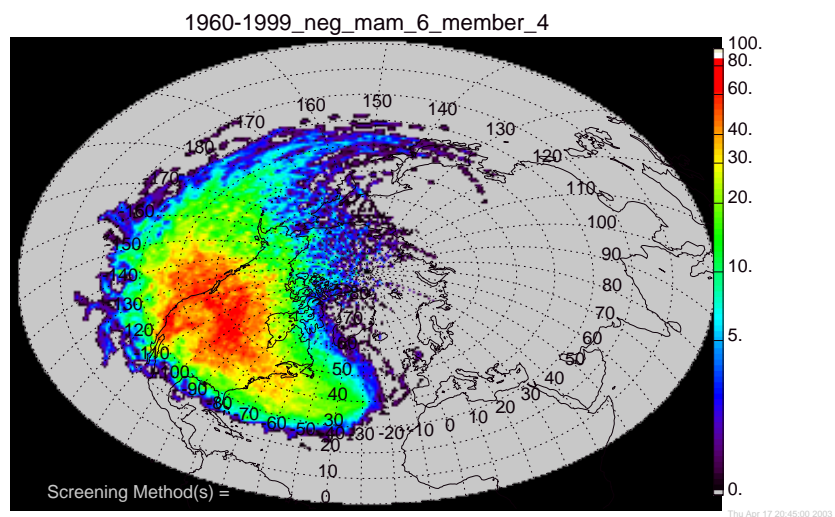


Figure C.54 Geometrically corrected density membership plot for springs with a negative NAOI value, 1960-1999.

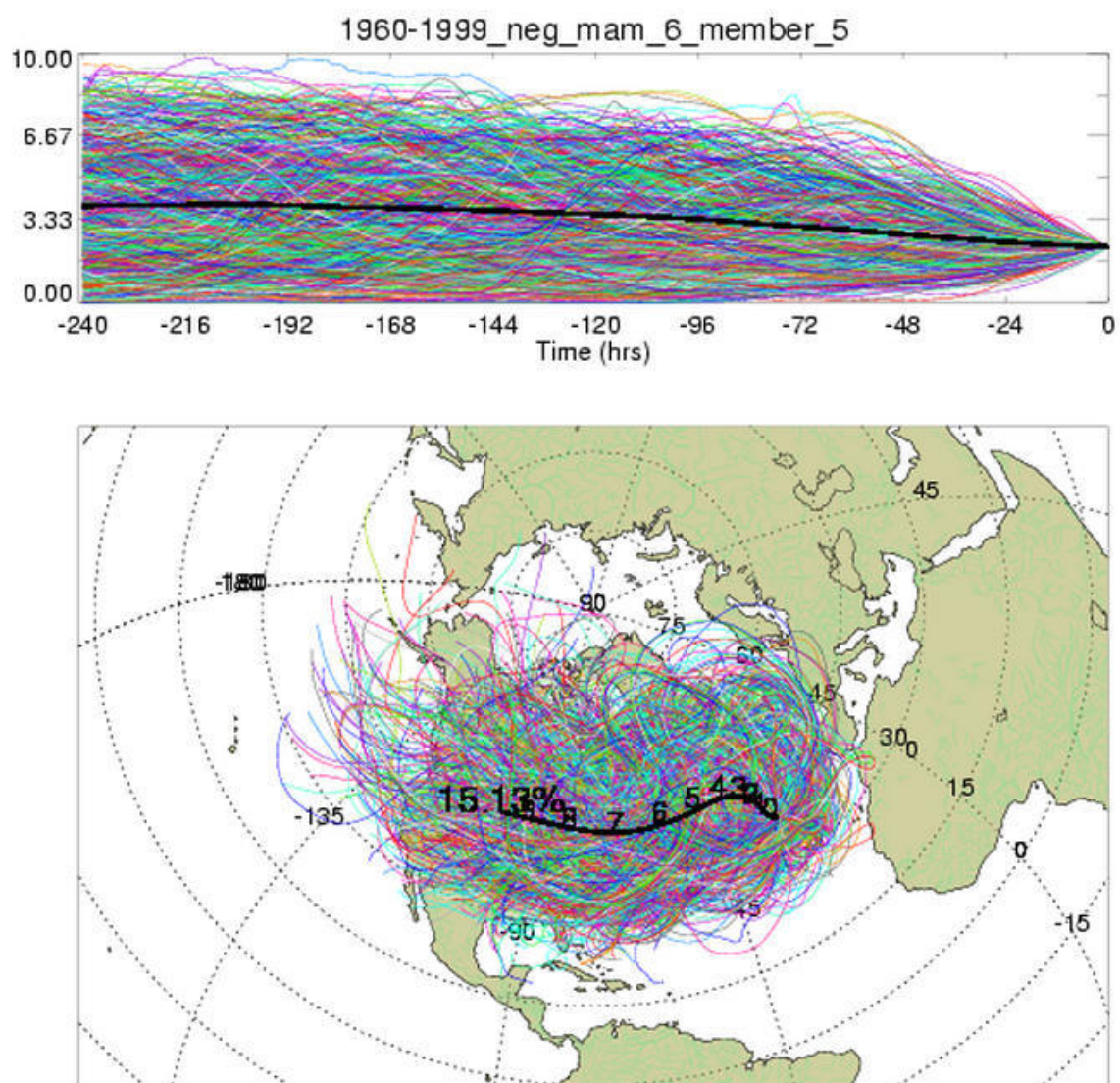


Figure C.55 Membership plot for springs with a negative NAOI value, 1960-1999.

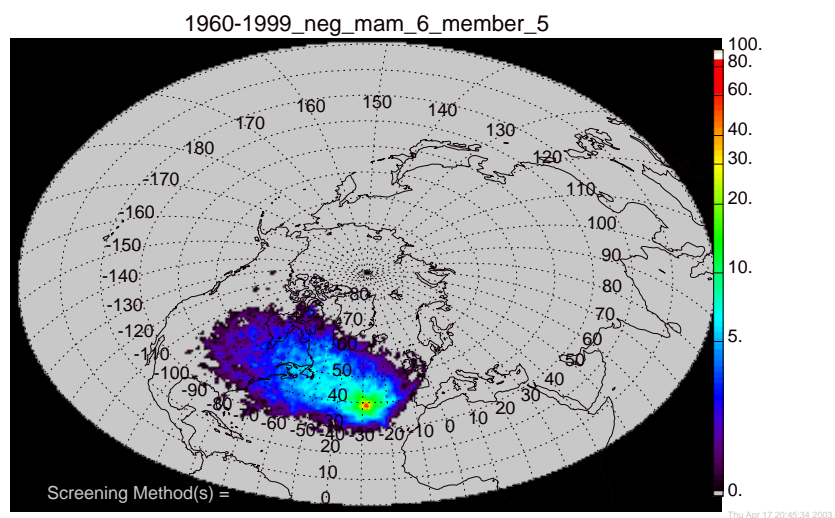


Figure C.56 Standard density membership plot for springs with a negative NAOI value, 1960-1999.

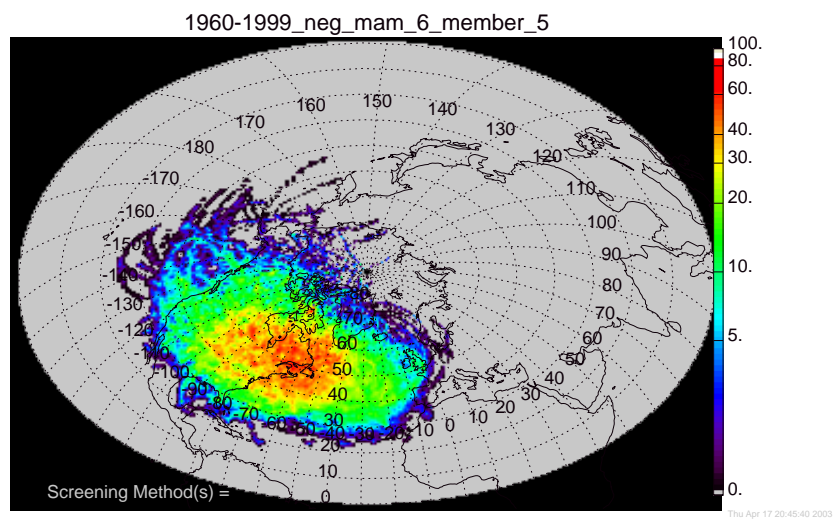


Figure C.57 Geometrically corrected density membership plot for springs with a negative NAOI value, 1960-1999.

Appendix D

Membership plots for summer periods

D.1 Membership plots for the average summer

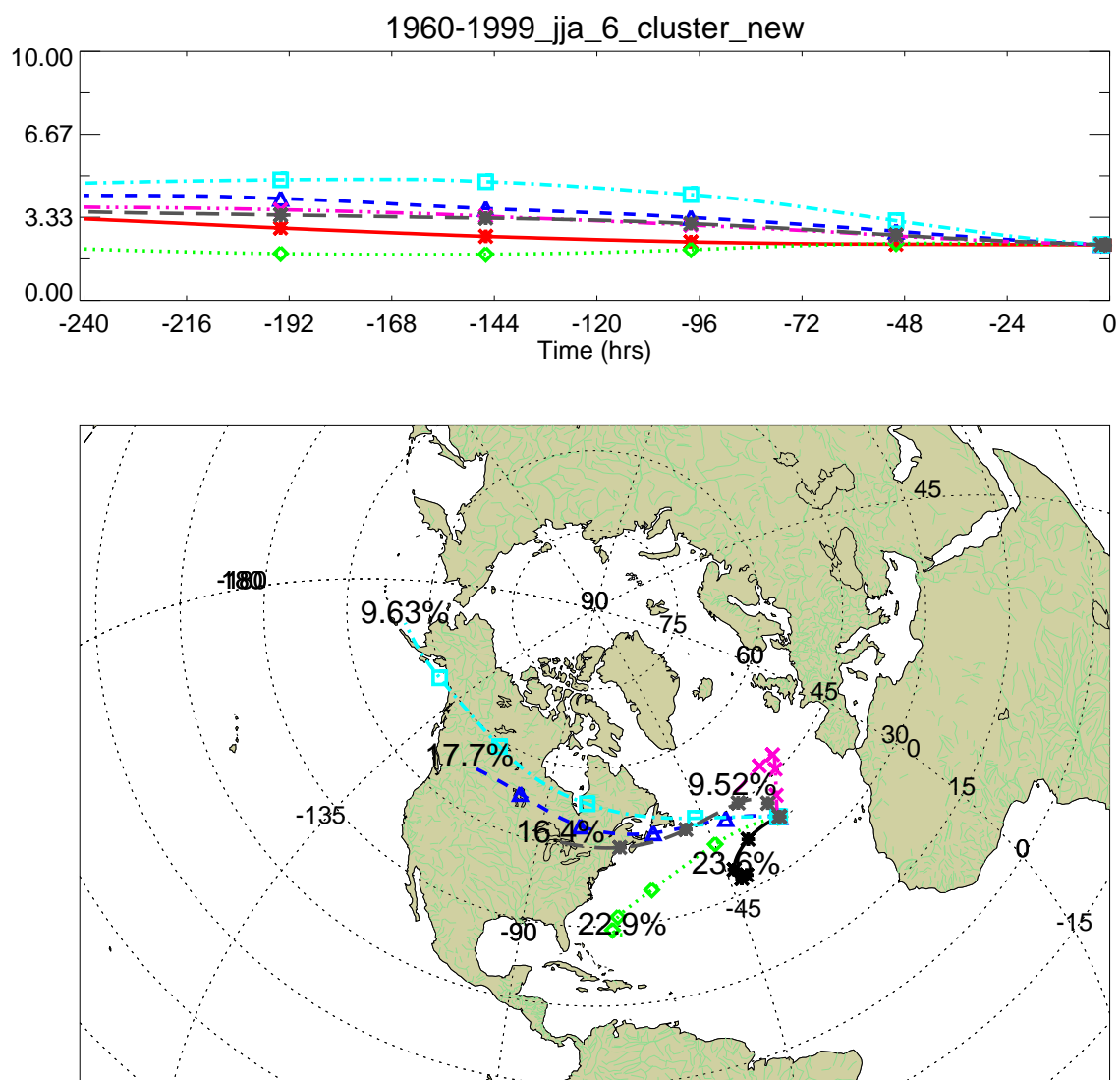


Figure D.1 Cluster plot for the average summer, 1960-1999.

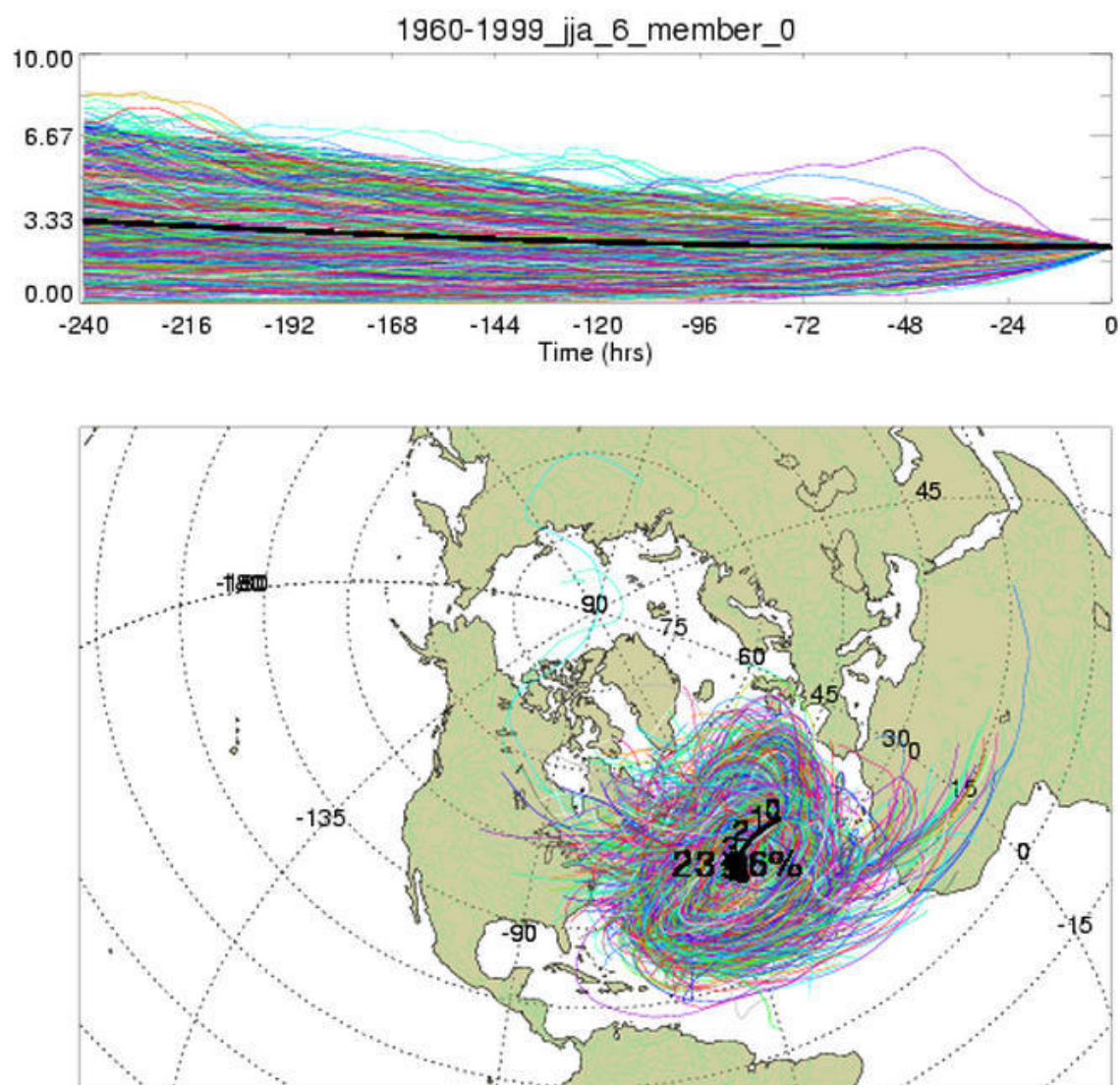


Figure D.2 Membership plot for summers, 1960-1999).

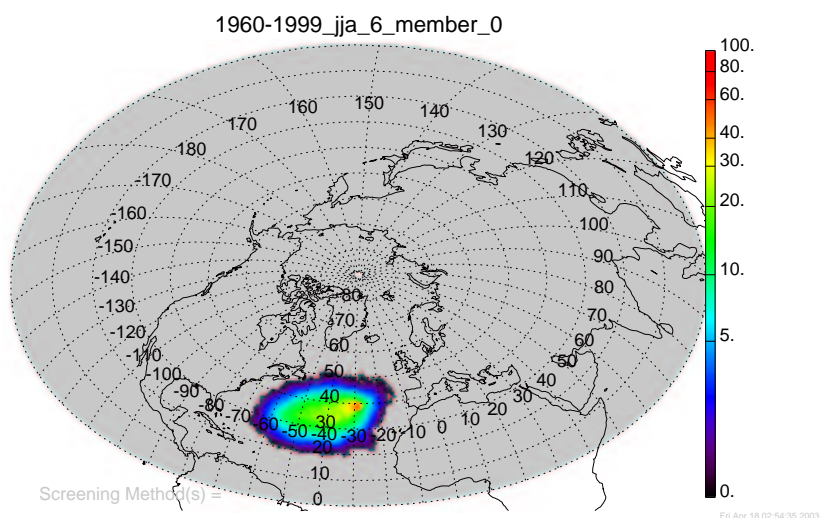


Figure D.3 Standard density membership plot for summers, 1960-1999).

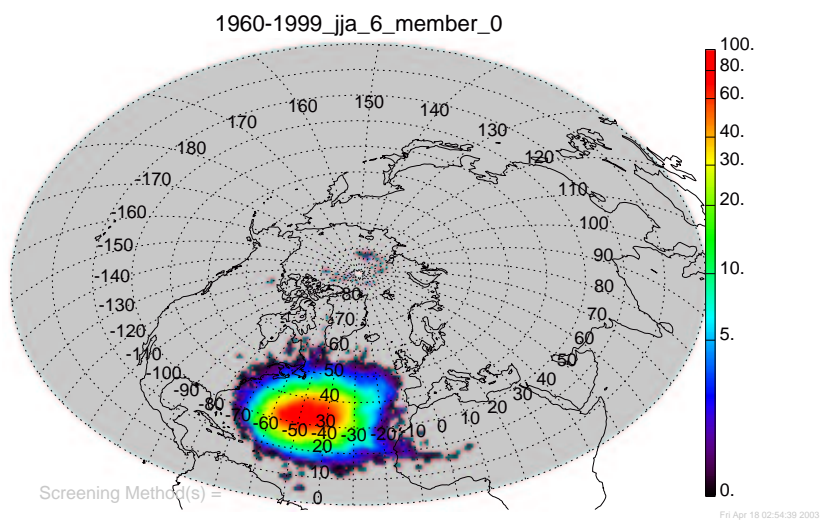


Figure D.4 Geometrically corrected density membership plot for summers, 1960-1999.

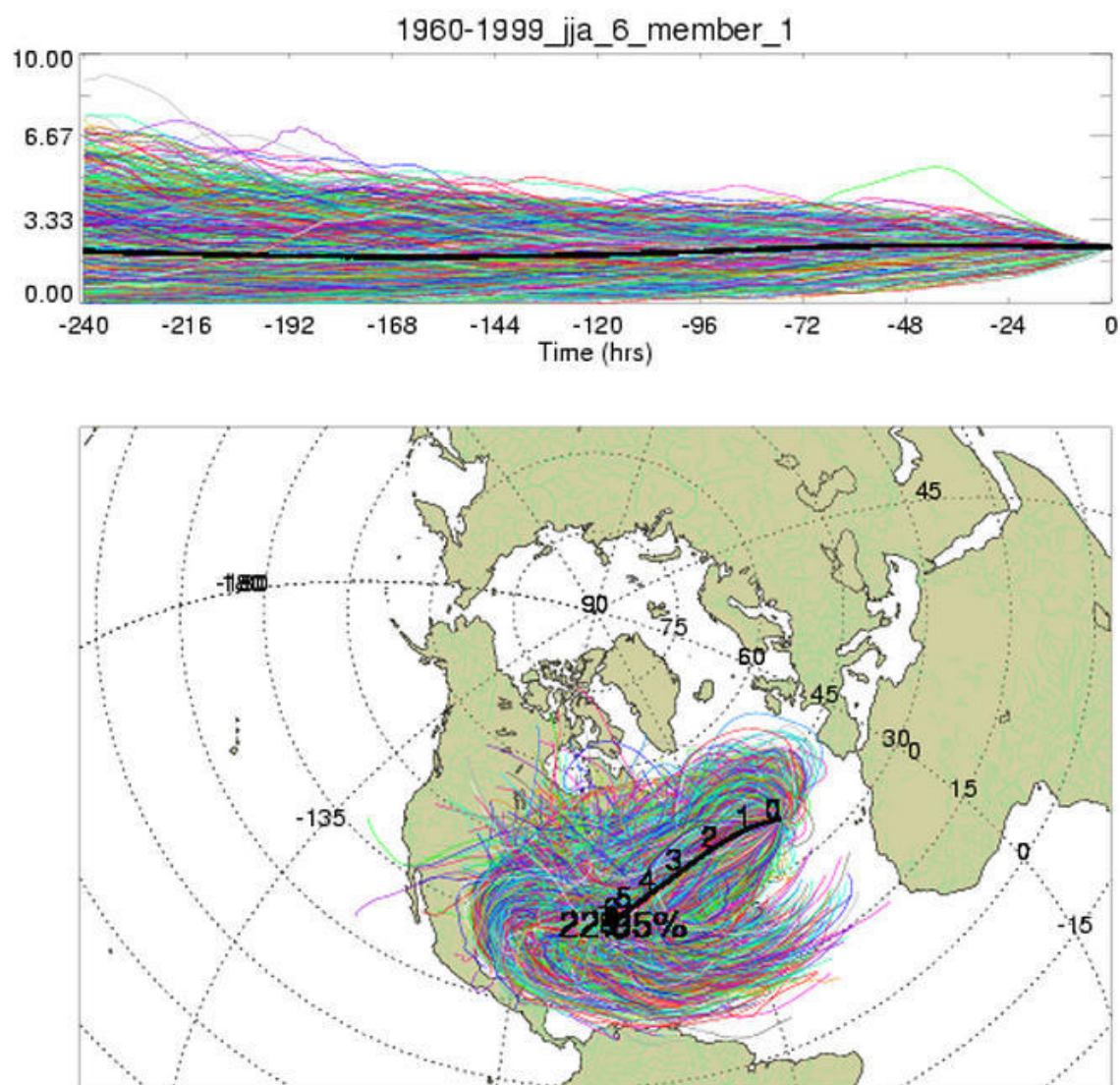


Figure D.5 Membership plot for summers, 1960-1999.

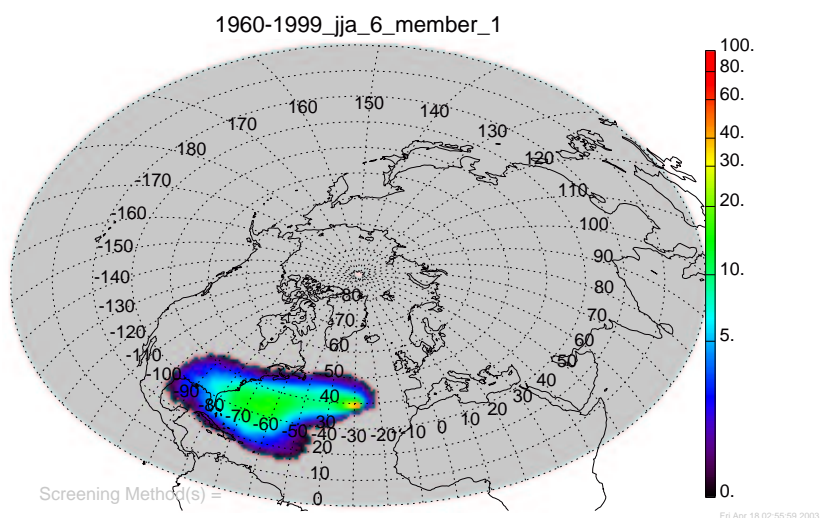


Figure D.6 Standard density membership plot for summers, 1960-1999.

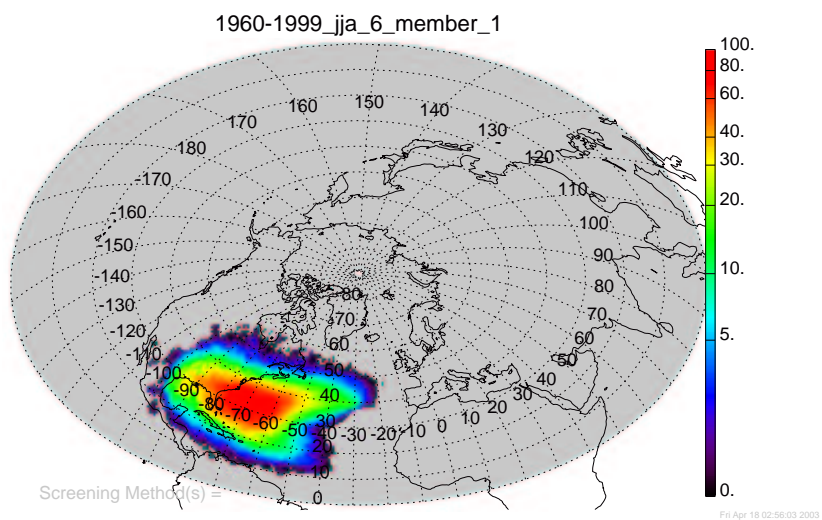


Figure D.7 Geometrically corrected density membership plot for summers, 1960-1999.

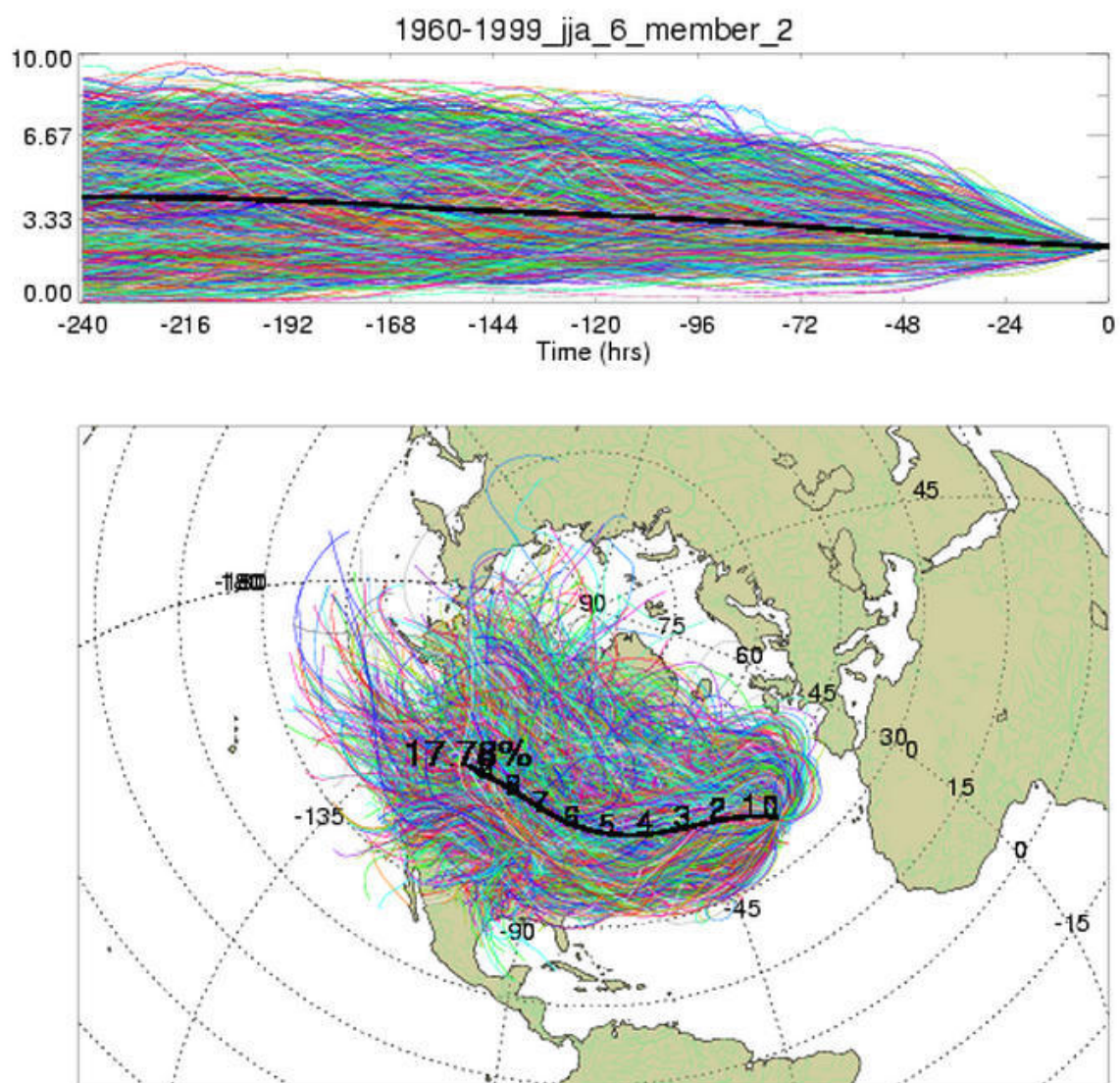


Figure D.8 Membership plot for summers, 1960-1999.

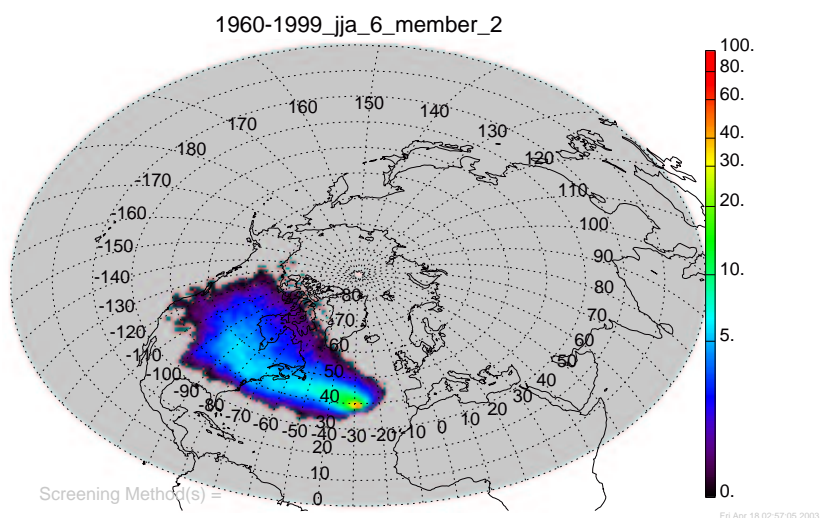


Figure D.9 Standard density membership plot for summers, 1960-1999.

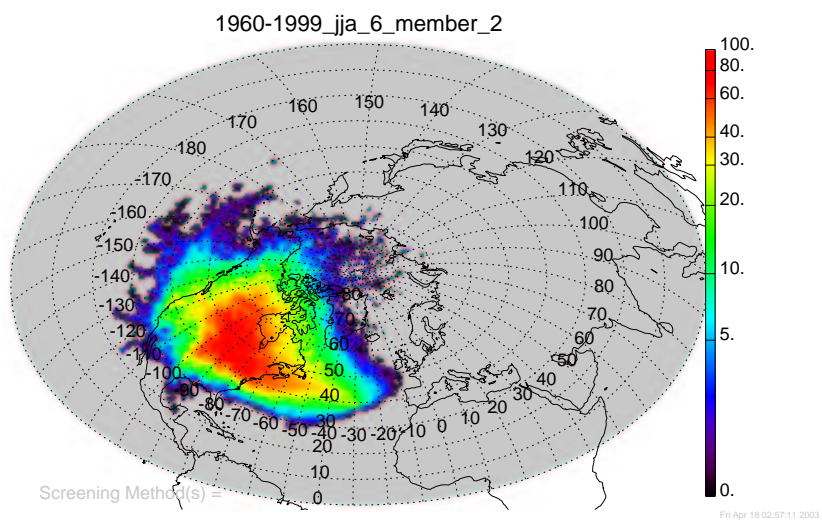


Figure D.10 Geometrically corrected density membership plot for summers, 1960-1999.

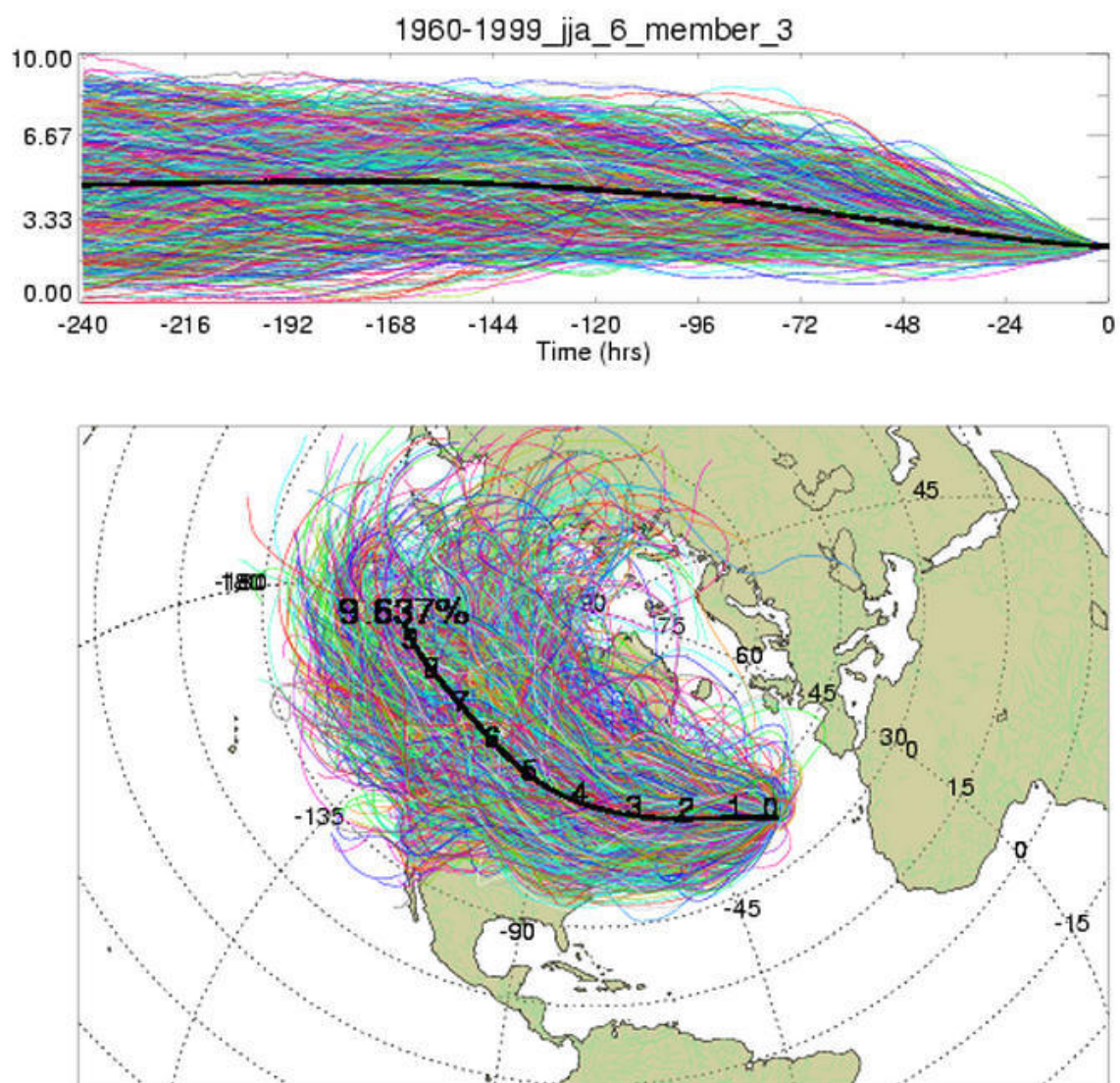


Figure D.11 Membership plot for summers, 1960-1999.

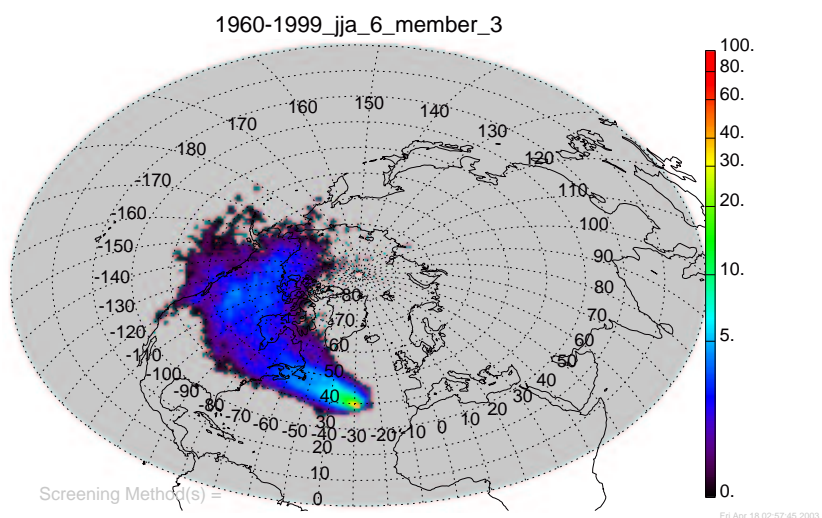


Figure D.12 Standard density membership plot for summers, 1960-1999.

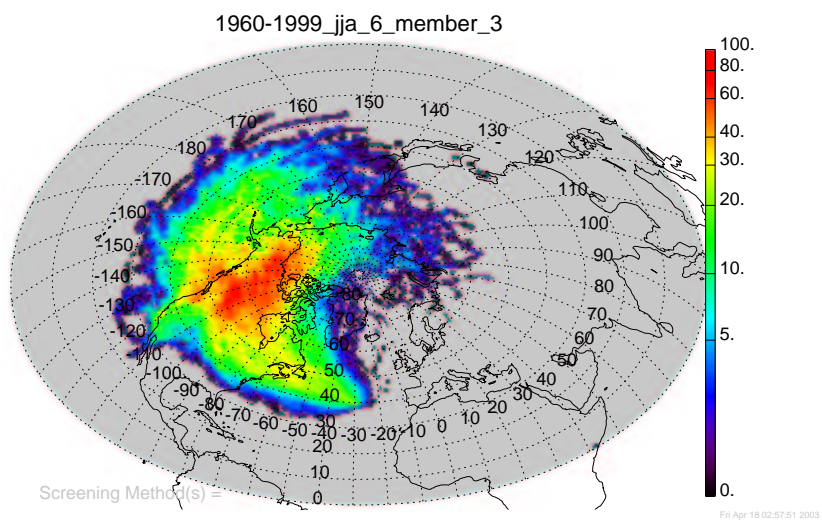


Figure D.13 Geometrically corrected density membership plot for summers, 1960-1999.

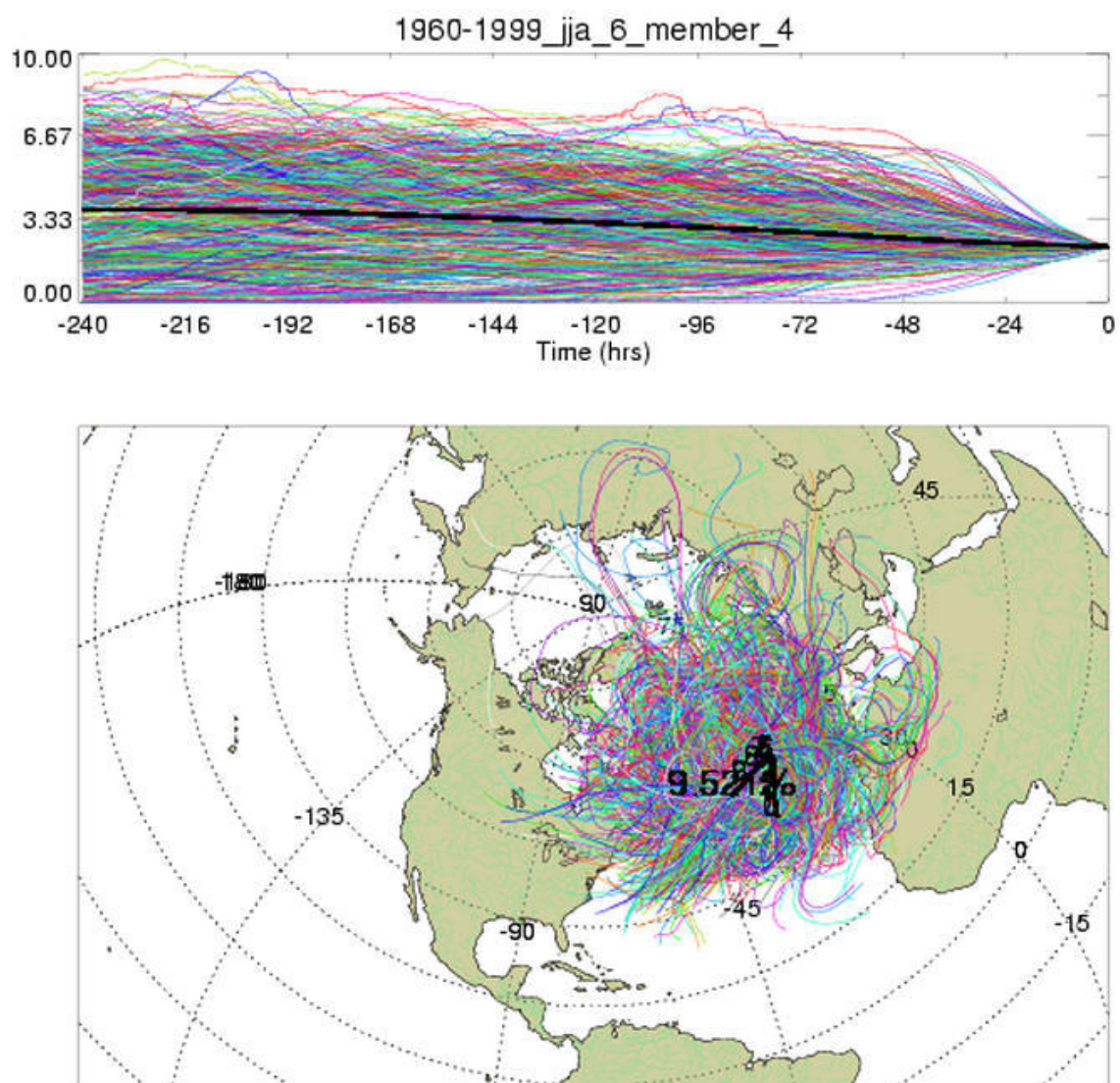


Figure D.14 Membership plot for summers, 1960-1999.

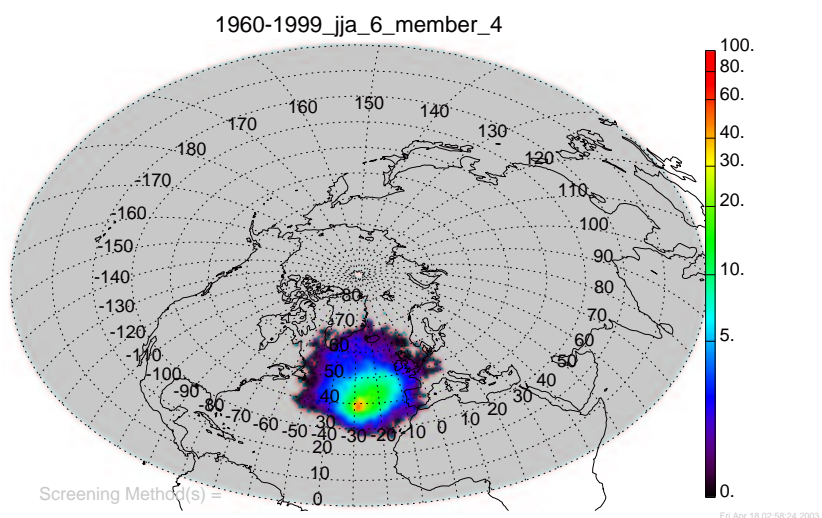


Figure D.15 Standard density membership plot for summers, 1960-1999.

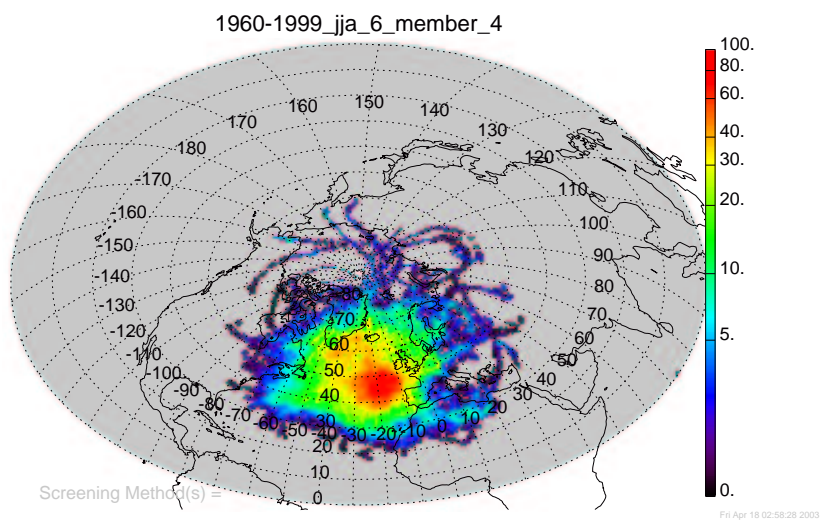


Figure D.16 Geometrically corrected density membership plot for summers, 1960-1999.

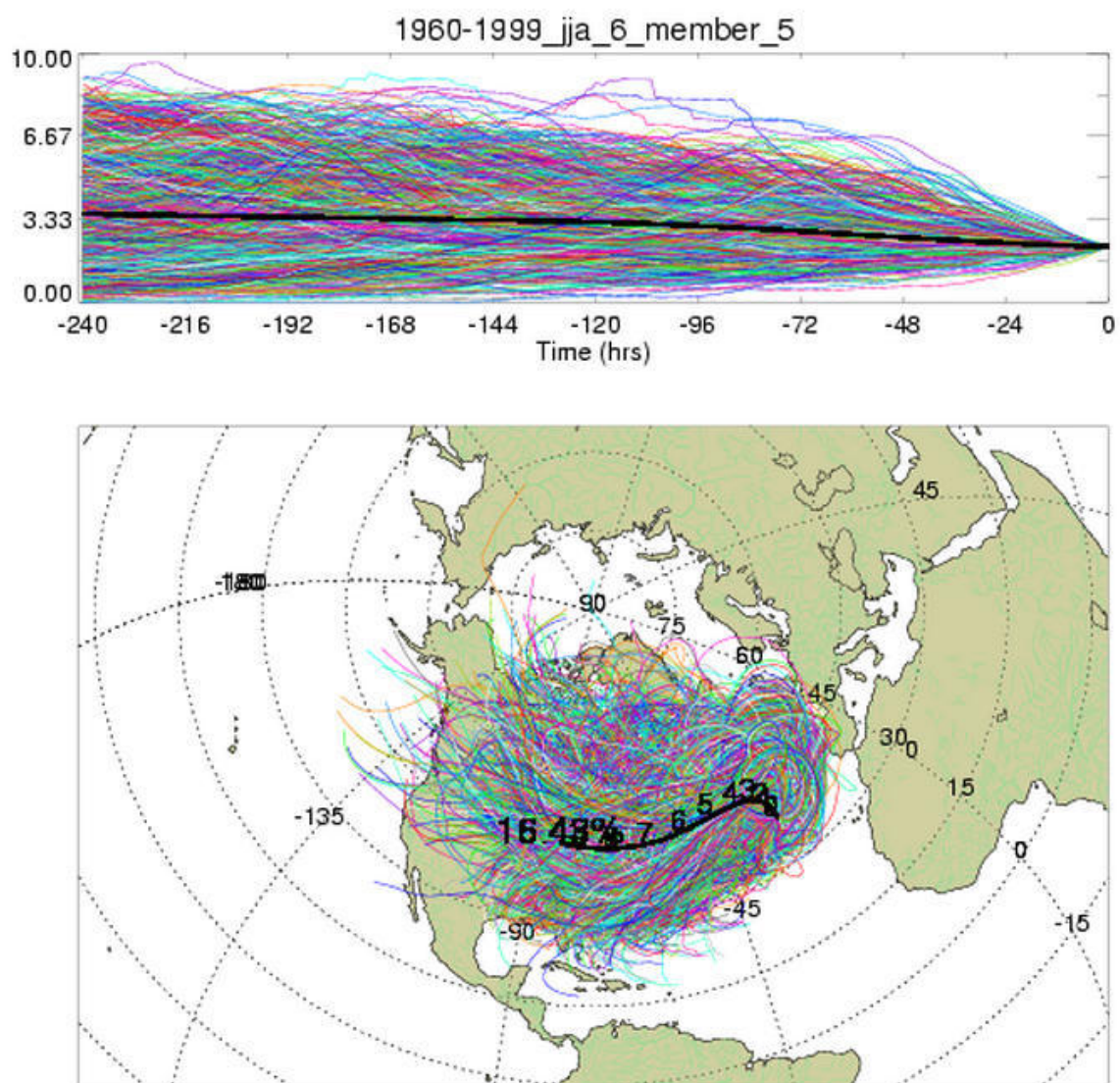


Figure D.17 Membership plot for summers, 1960-1999.

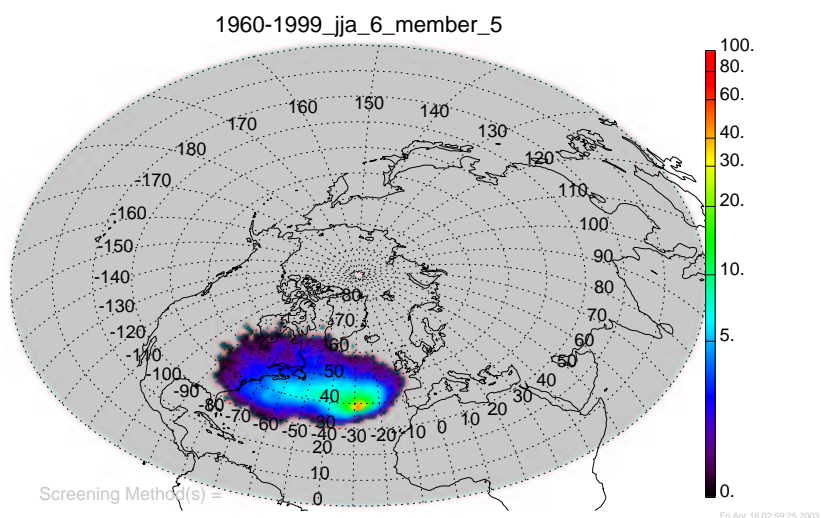


Figure D.18 Standard density membership plot for summers, 1960-1999.

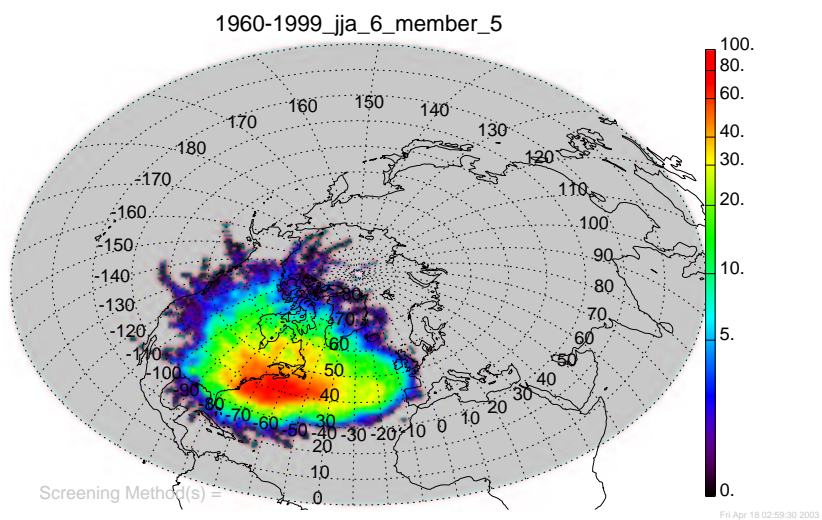


Figure D.19 Geometrically corrected density membership plot for summers, 1960-1999.

D.2 Membership plots for the positive NAO summers

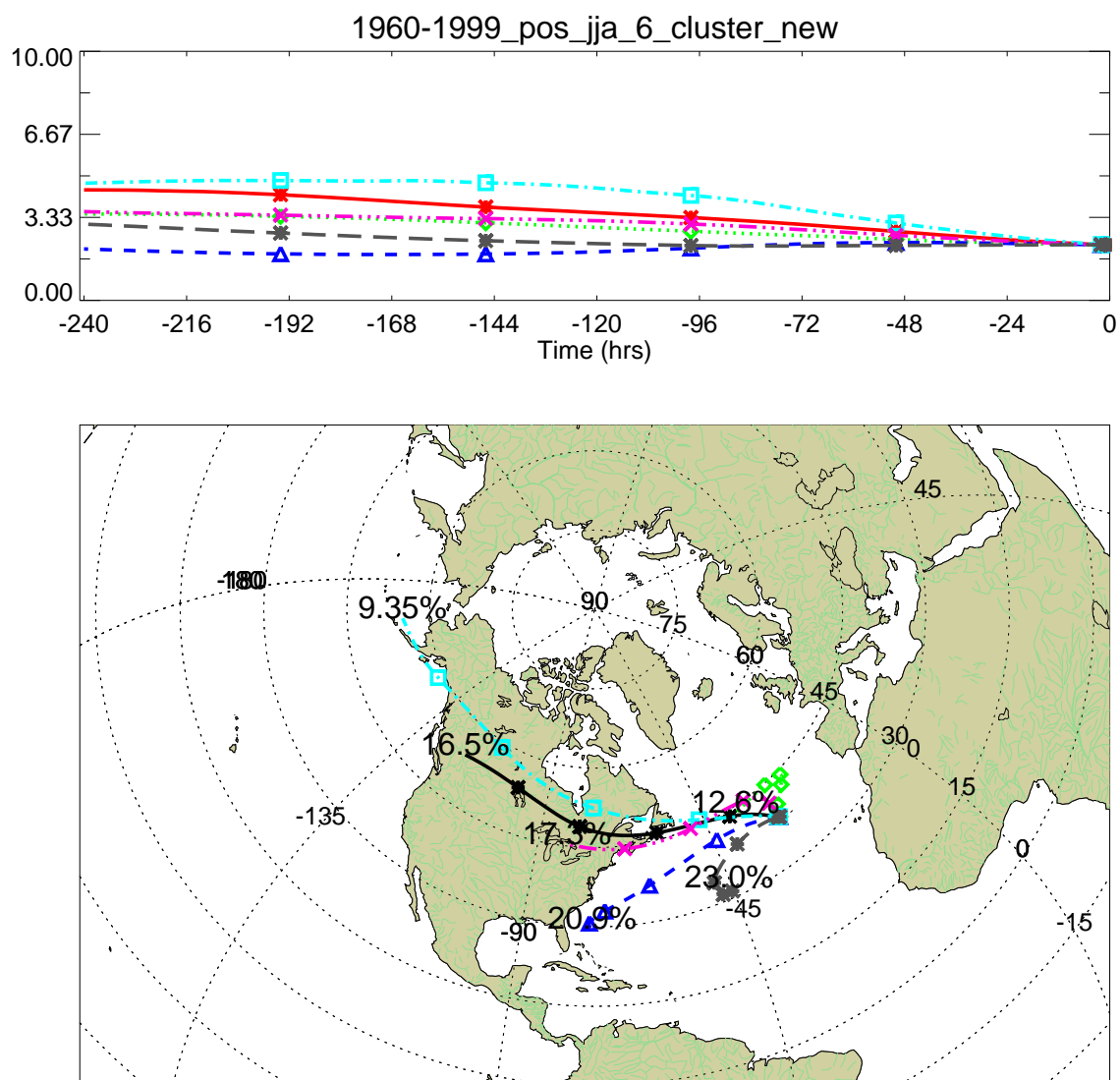


Figure D.20 Cluster plot for positive NAO summer, 1960-1999.

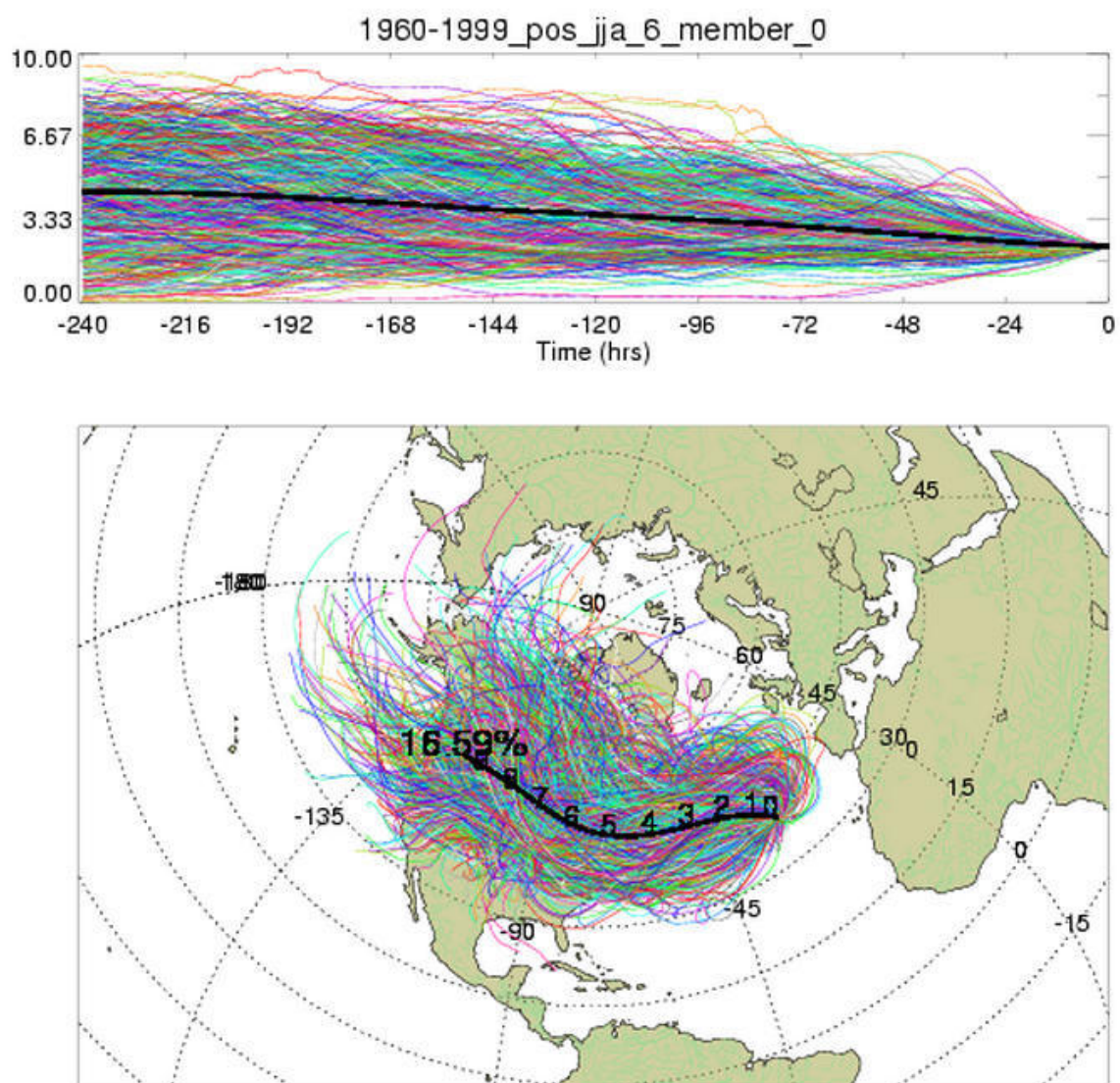


Figure D.21 Membership plot for summers with a positive NAOI value, 1960-1999.

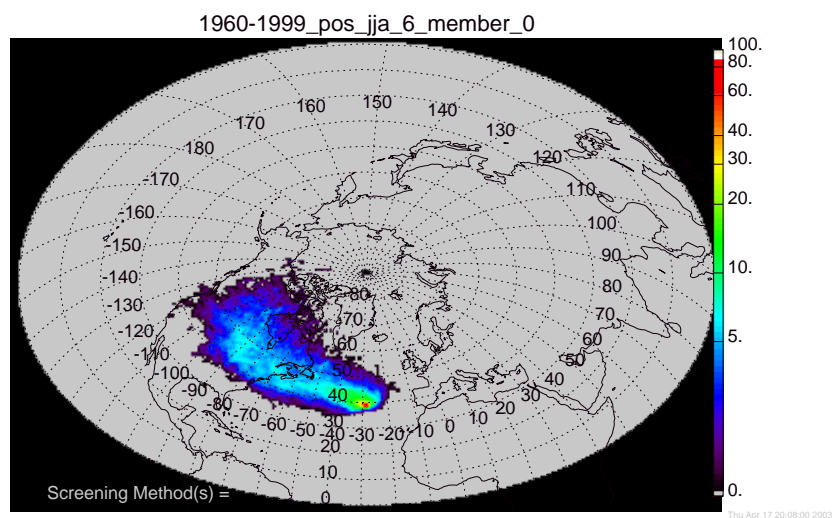


Figure D.22 Standard density membership plot for summers with a positive NAOI value, 1960-1999.

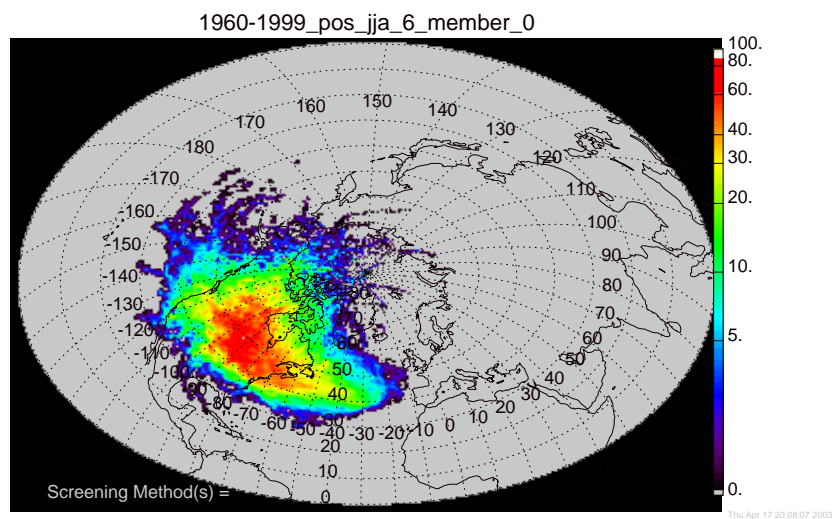


Figure D.23 Geometrically corrected density membership plot for summers with a positive NAOI value, 1960-1999.

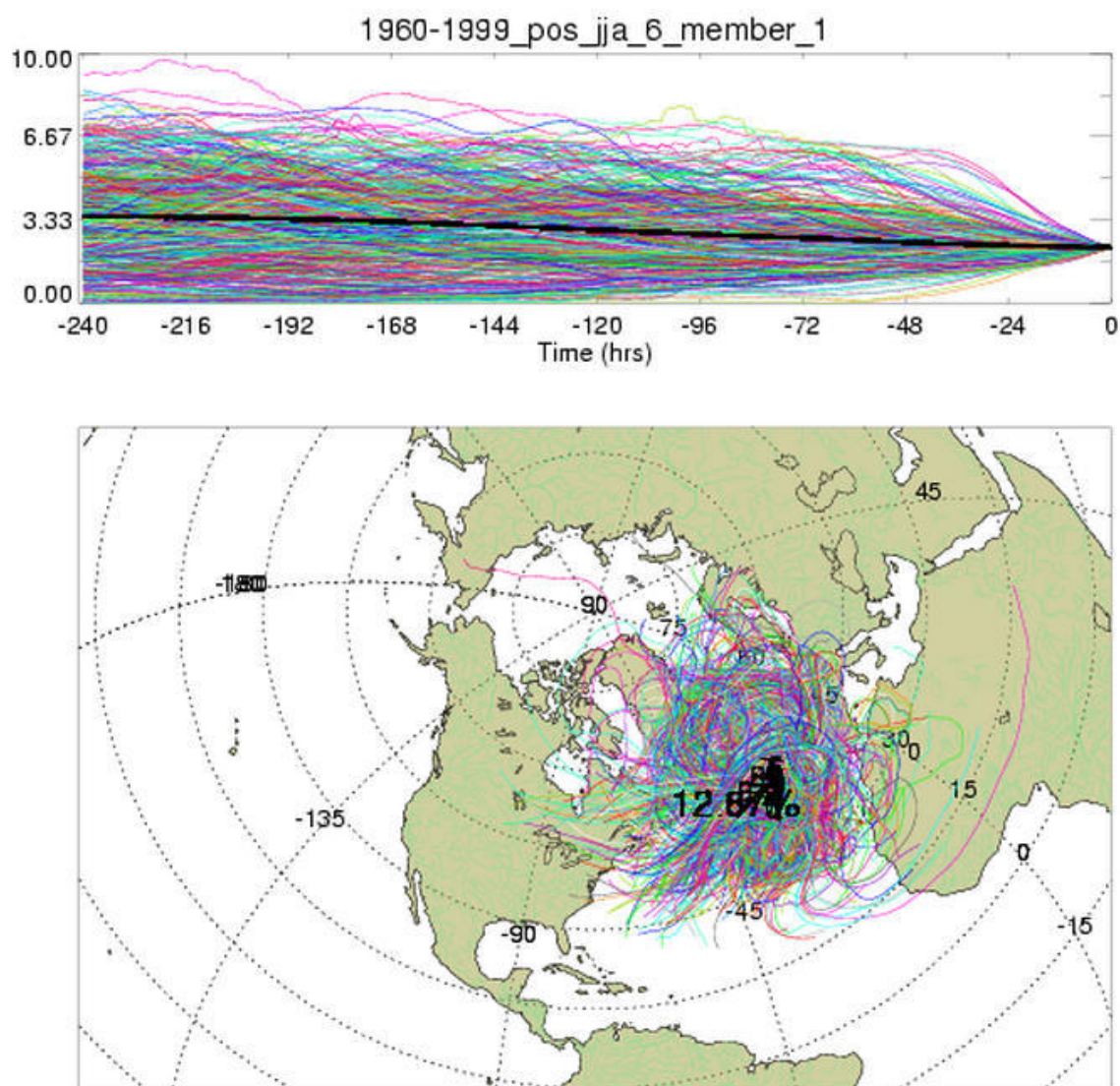


Figure D.24 Membership plot for summers with a positive NAOI value, 1960-1999.

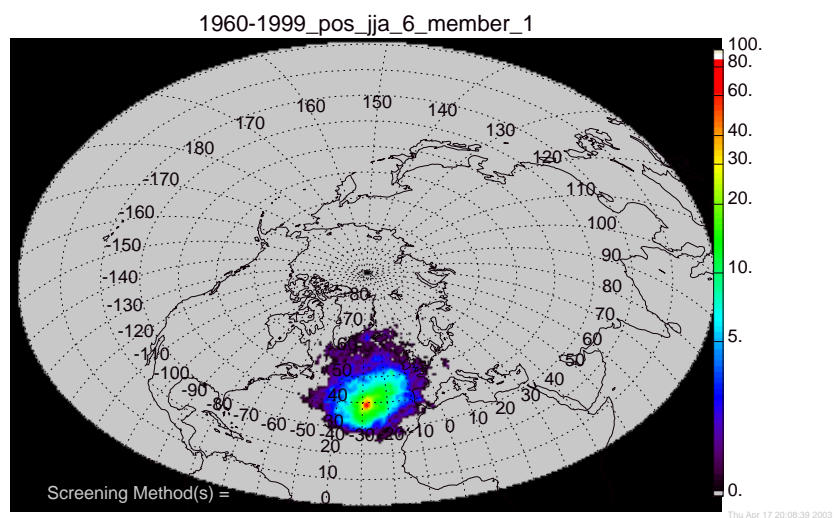


Figure D.25 Standard density membership plot for summers with a positive NAOI value, 1960-1999.

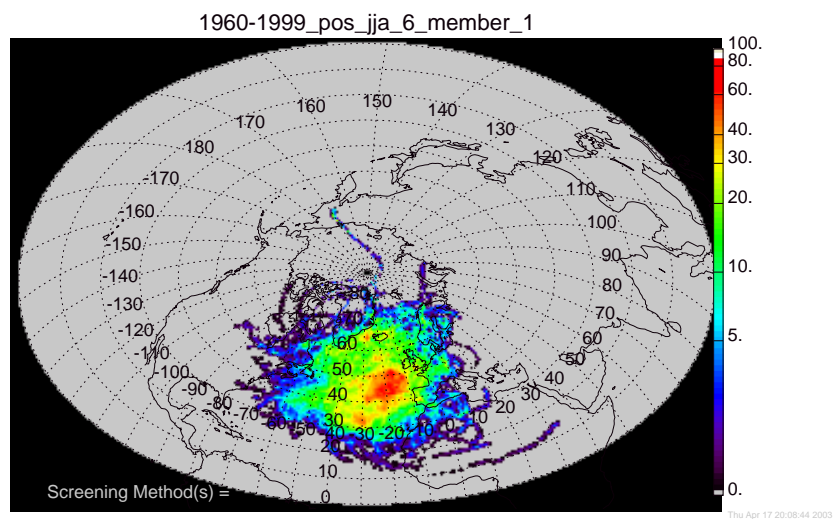


Figure D.26 Geometrically corrected density membership plot for summers with a positive NAOI value, 1960-1999.

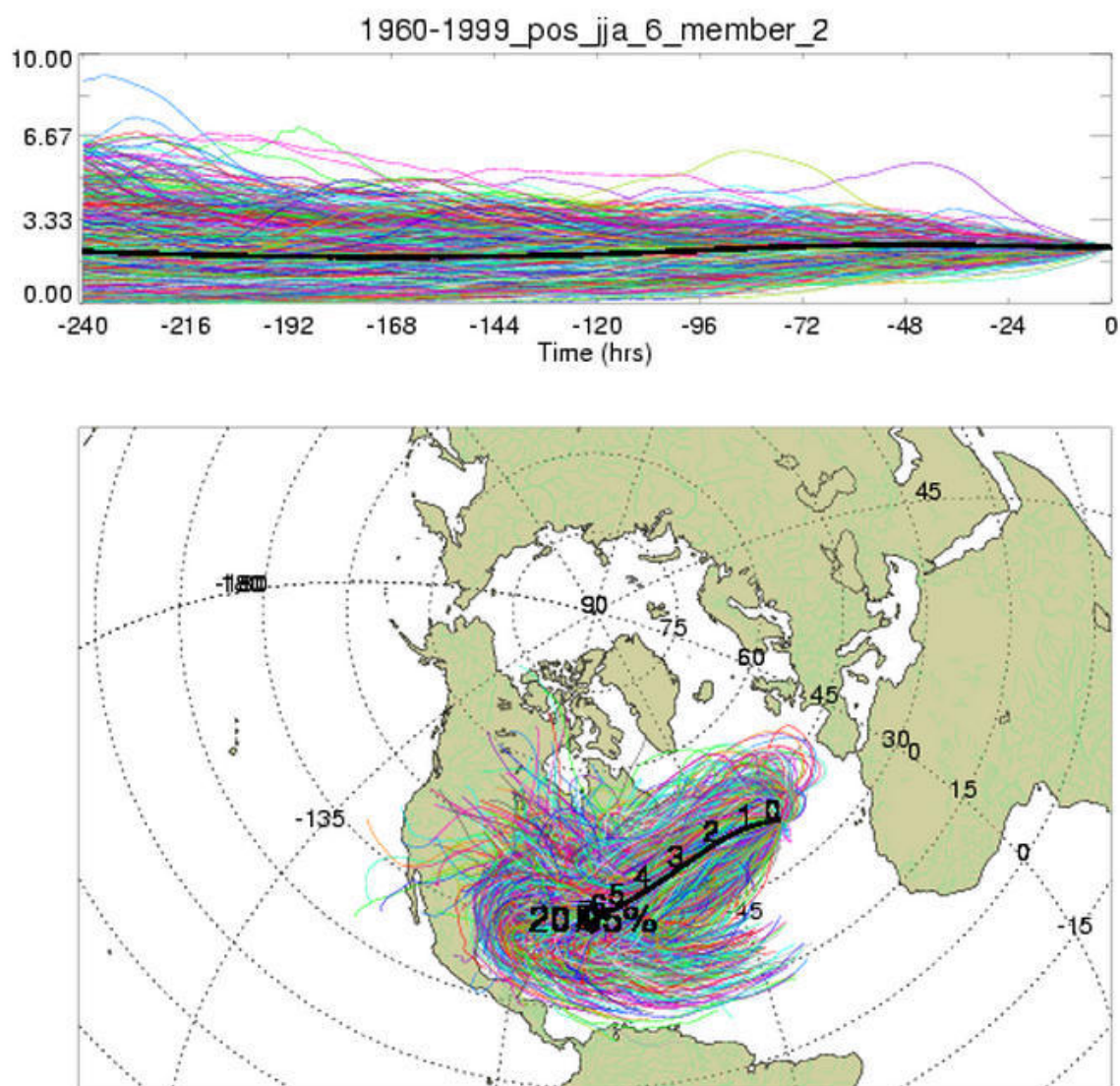


Figure D.27 Membership plot for summers with a positive NAOI value, 1960-1999.

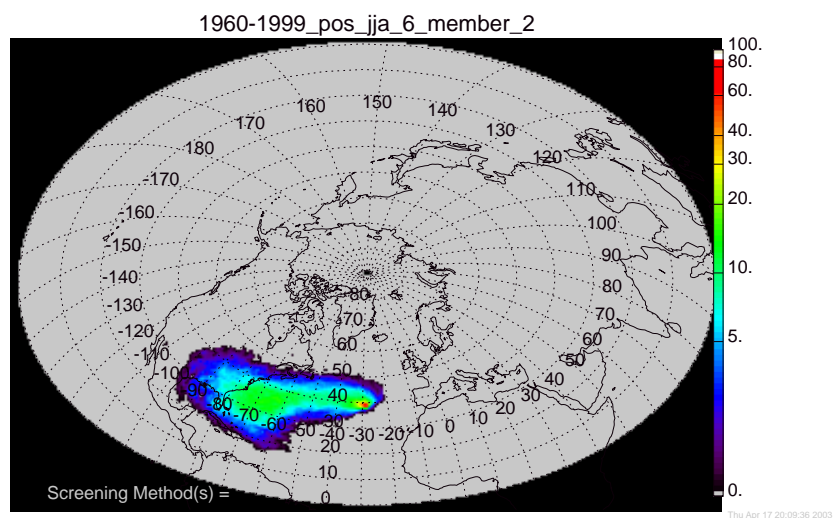


Figure D.28 Standard density membership plot for summers with a positive NAOI value, 1960-1999.

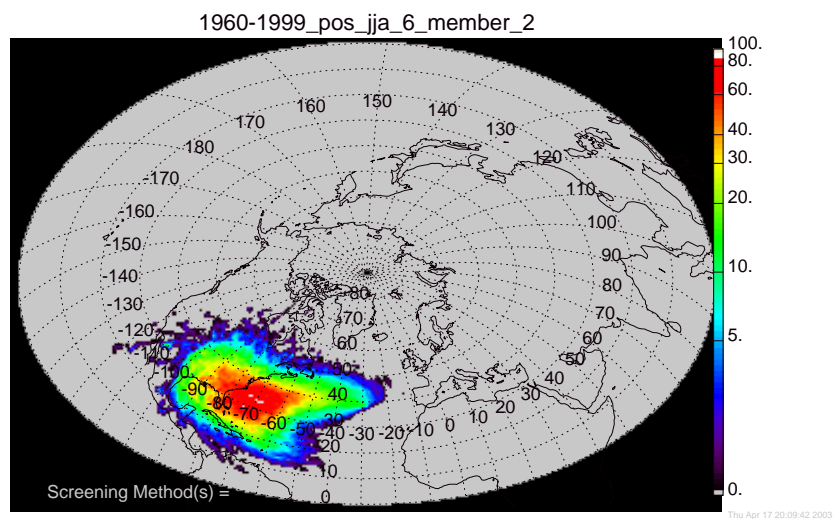


Figure D.29 Geometrically corrected density membership plot for summers with a positive NAOI value, 1960-1999.

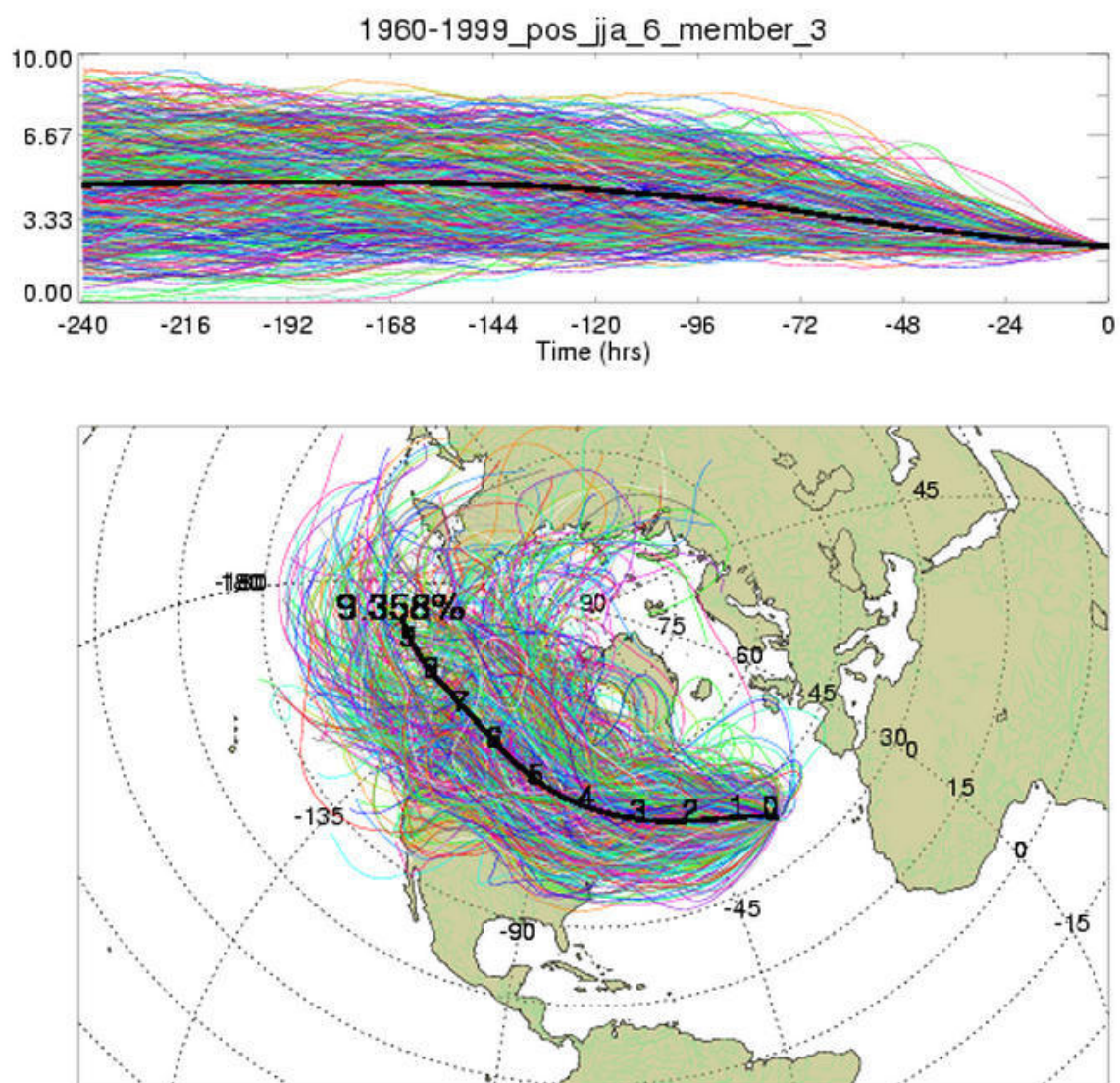


Figure D.30 Membership plot for summers with a positive NAOI value, 1960-1999.

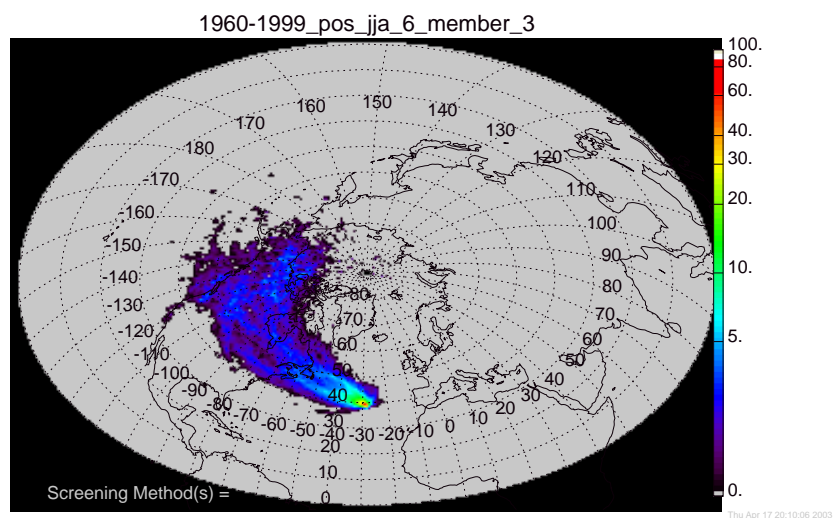


Figure D.31 Standard density membership plot for summers with a positive NAOI value, 1960-1999.

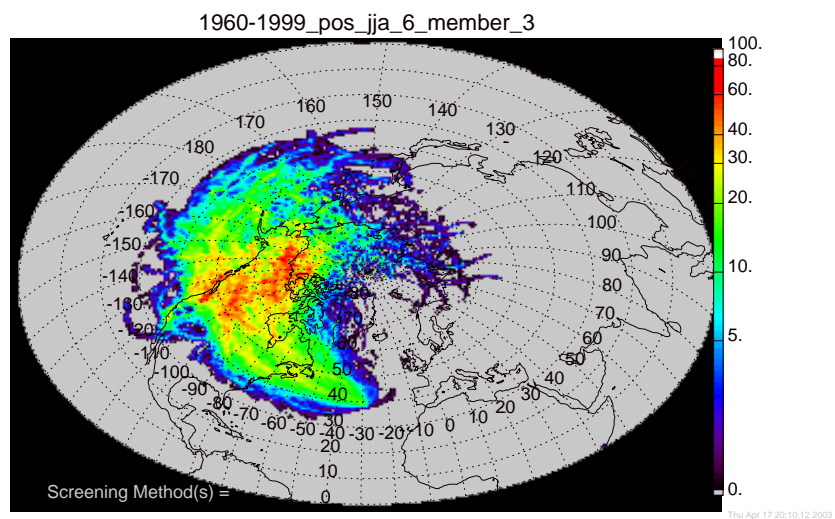


Figure D.32 Geometrically corrected density membership plot for summers with a positive NAOI value, 1960-1999.

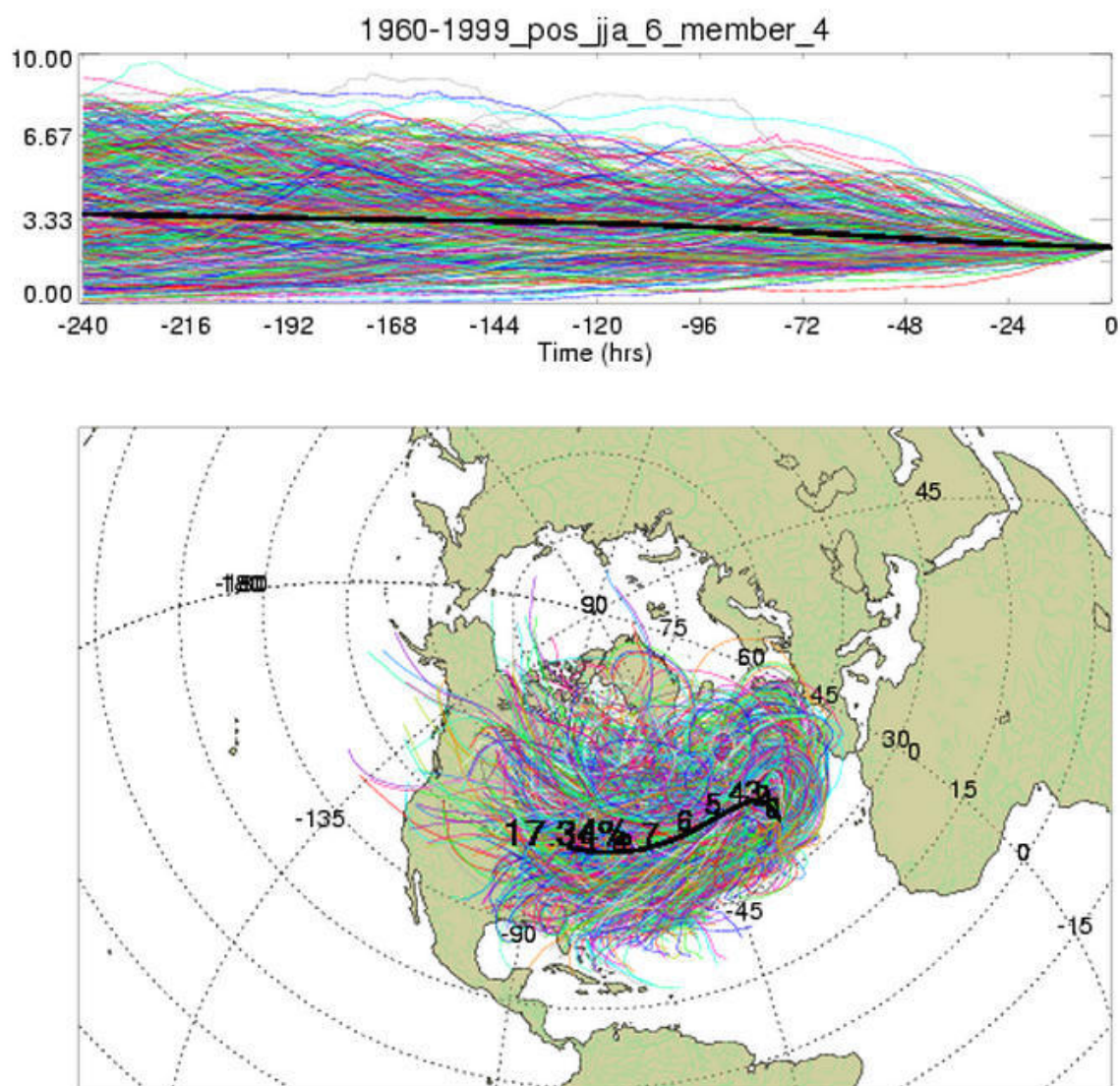


Figure D.33 Membership plot for summers with a positive NAOI value, 1960-1999.

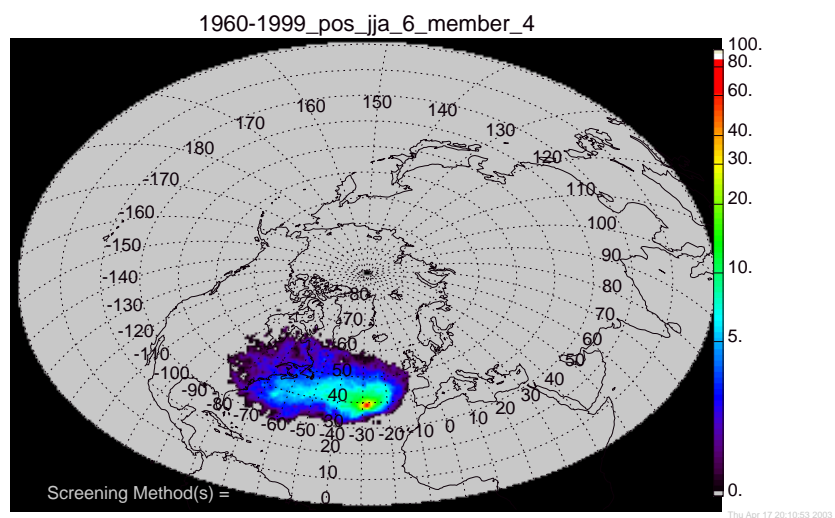


Figure D.34 Standard density membership plot for summers with a positive NAOI value, 1960-1999.

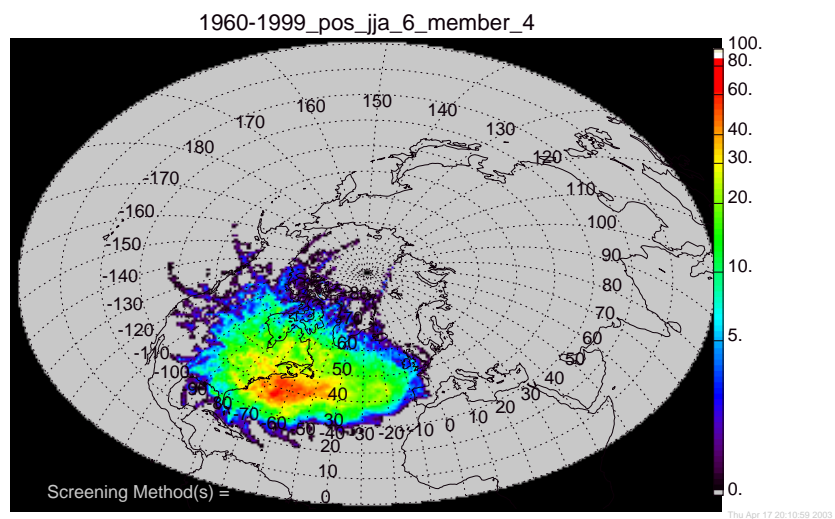


Figure D.35 Geometrically corrected density membership plot for summers with a positive NAOI value, 1960-1999.

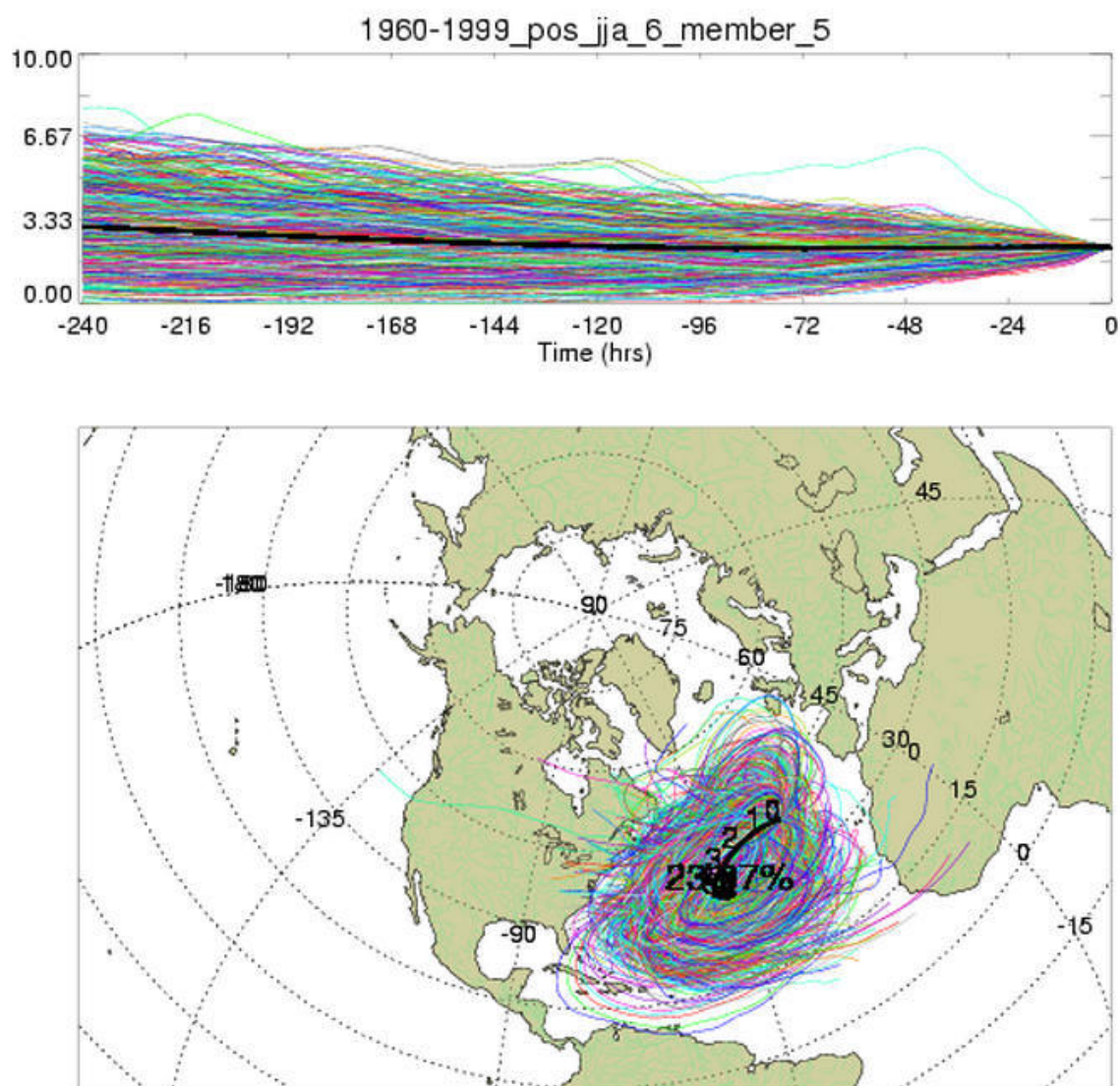


Figure D.36 Membership plot for summers with a positive NAOI value, 1960-1999.

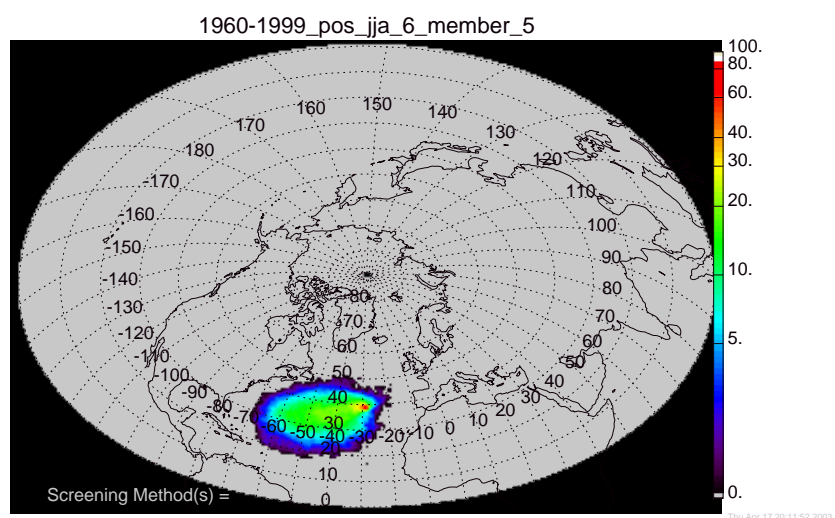


Figure D.37 Standard density membership plot for summers with a positive NAOI value, 1960-1999.

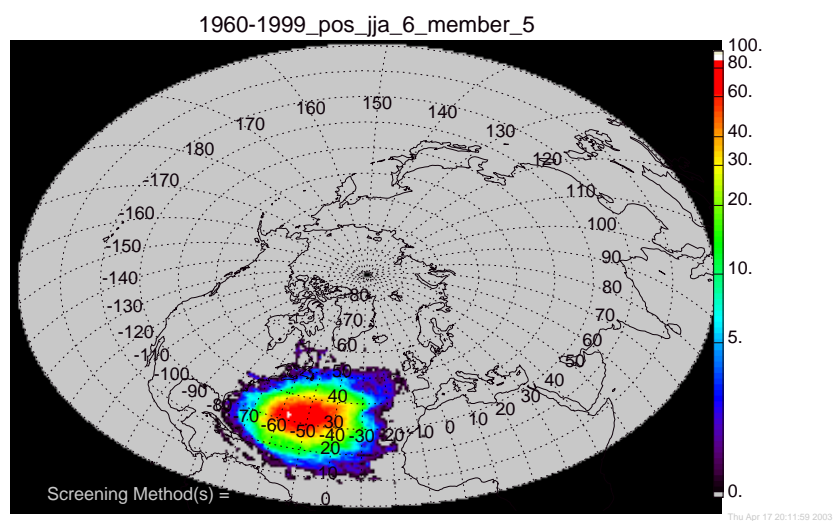


Figure D.38 Geometrically corrected density membership plot for summers with a positive NAOI value, 1960-1999.

D.3 Membership plots for the negative NAO summers

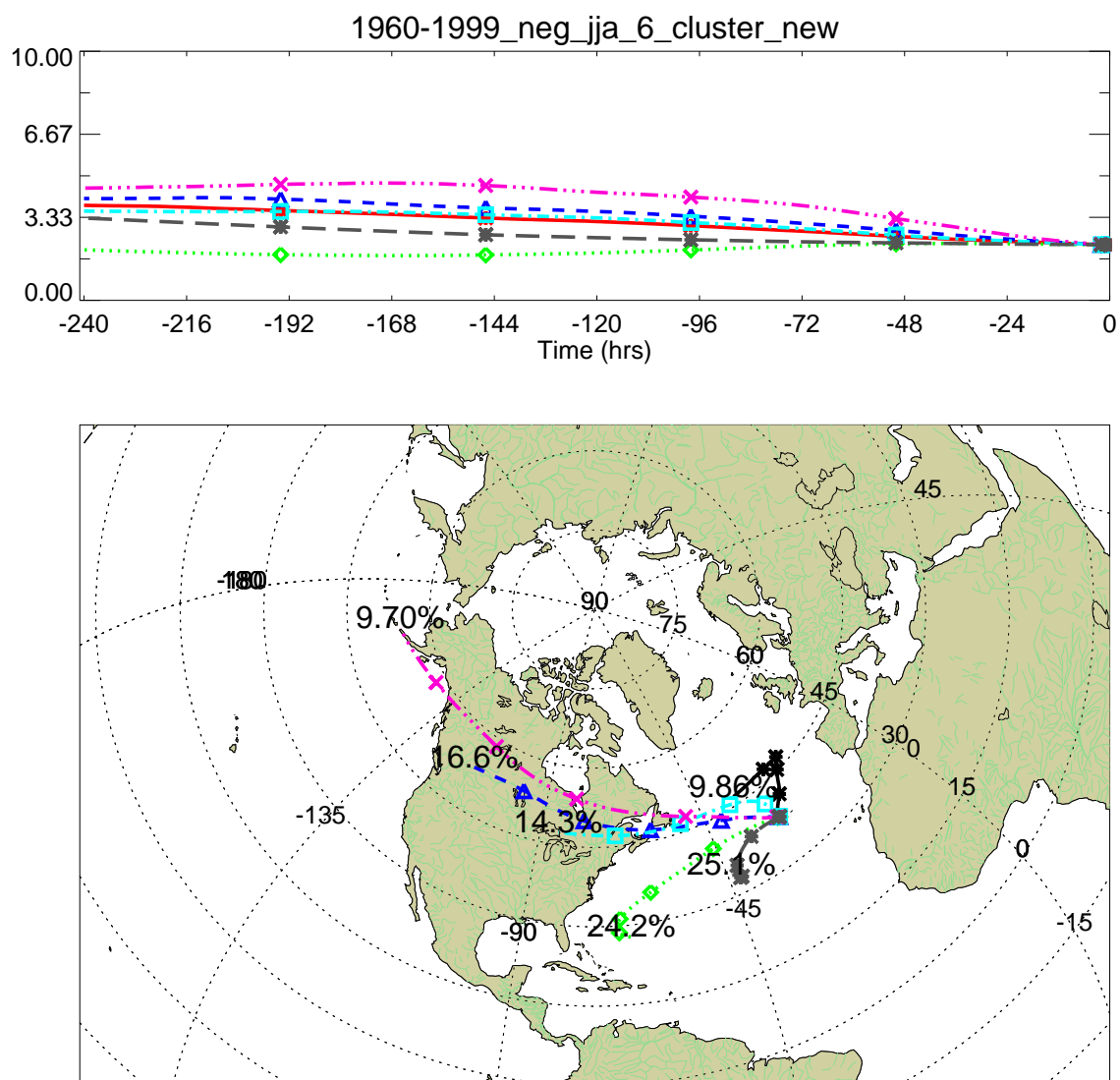


Figure D.39 Cluster plot for negative NAO summer, 1960-1999.

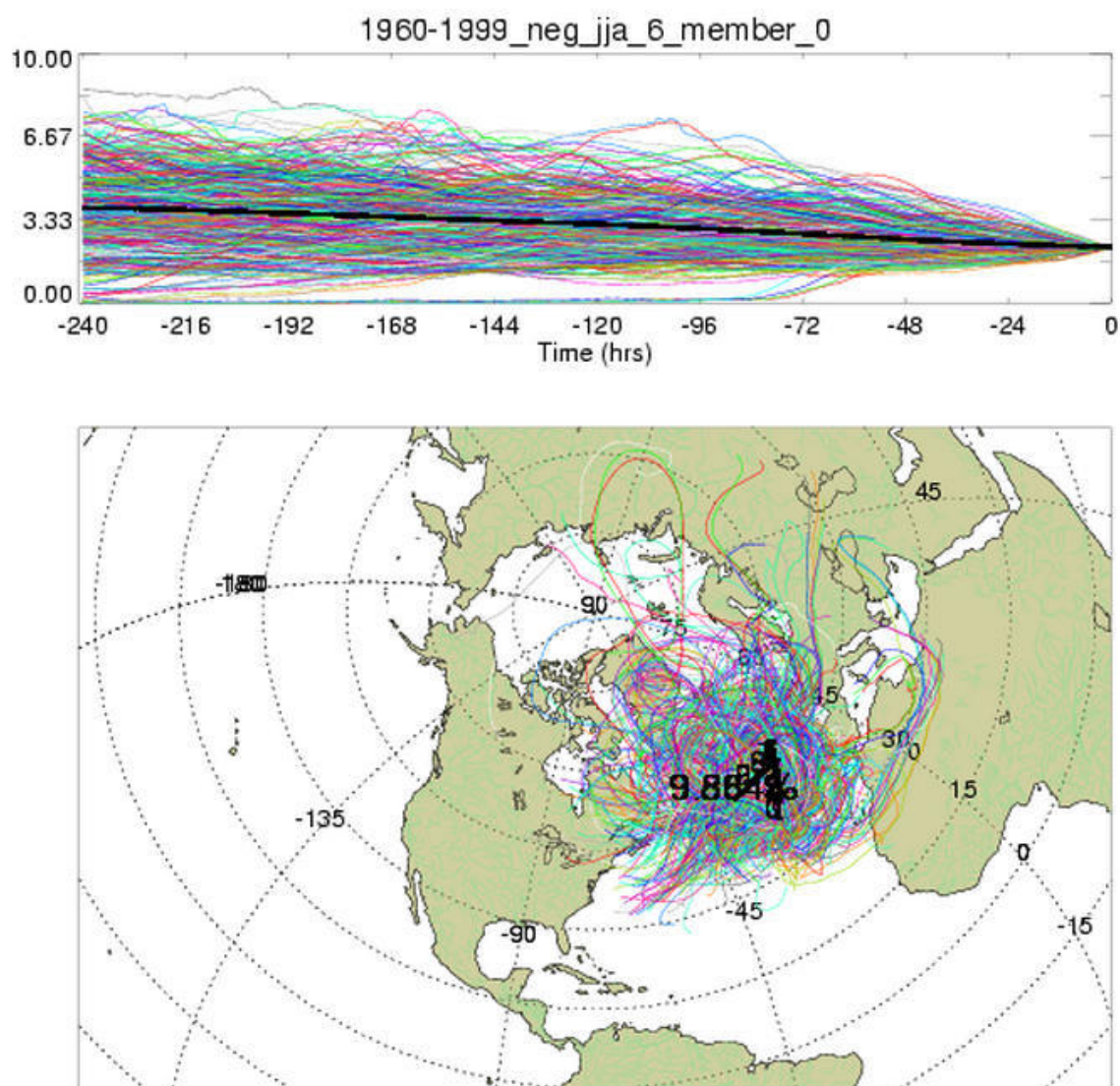


Figure D.40 Membership plot for summers with a negative NAOI value, 1960-1999.

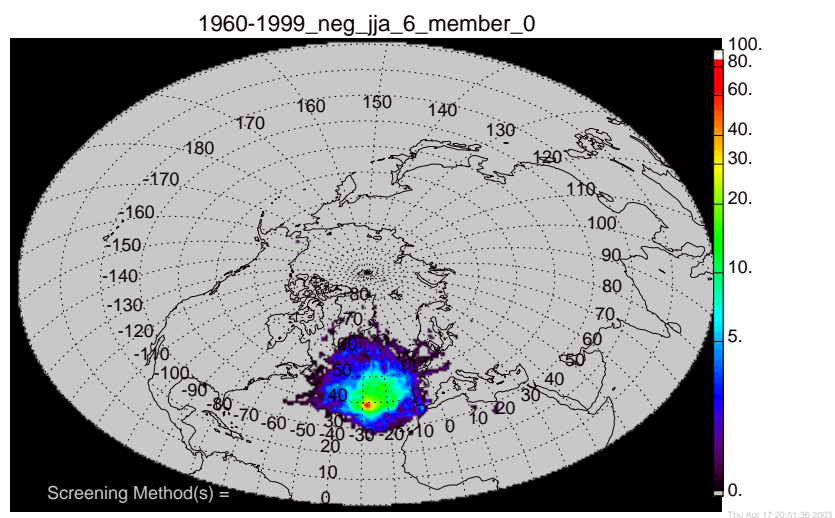


Figure D.41 Standard density membership plot for summers with a negative NAOI value, 1960-1999.

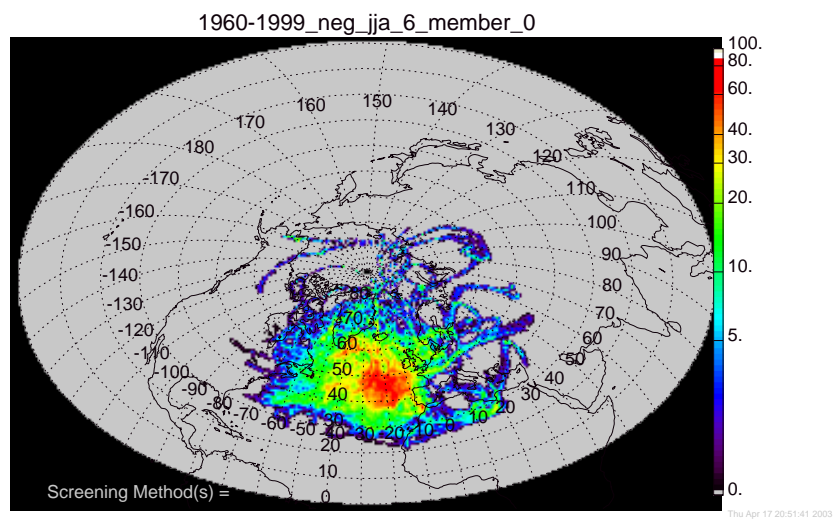


Figure D.42 Geometrically corrected density membership plot for summers with a negative NAOI value, 1960-1999.

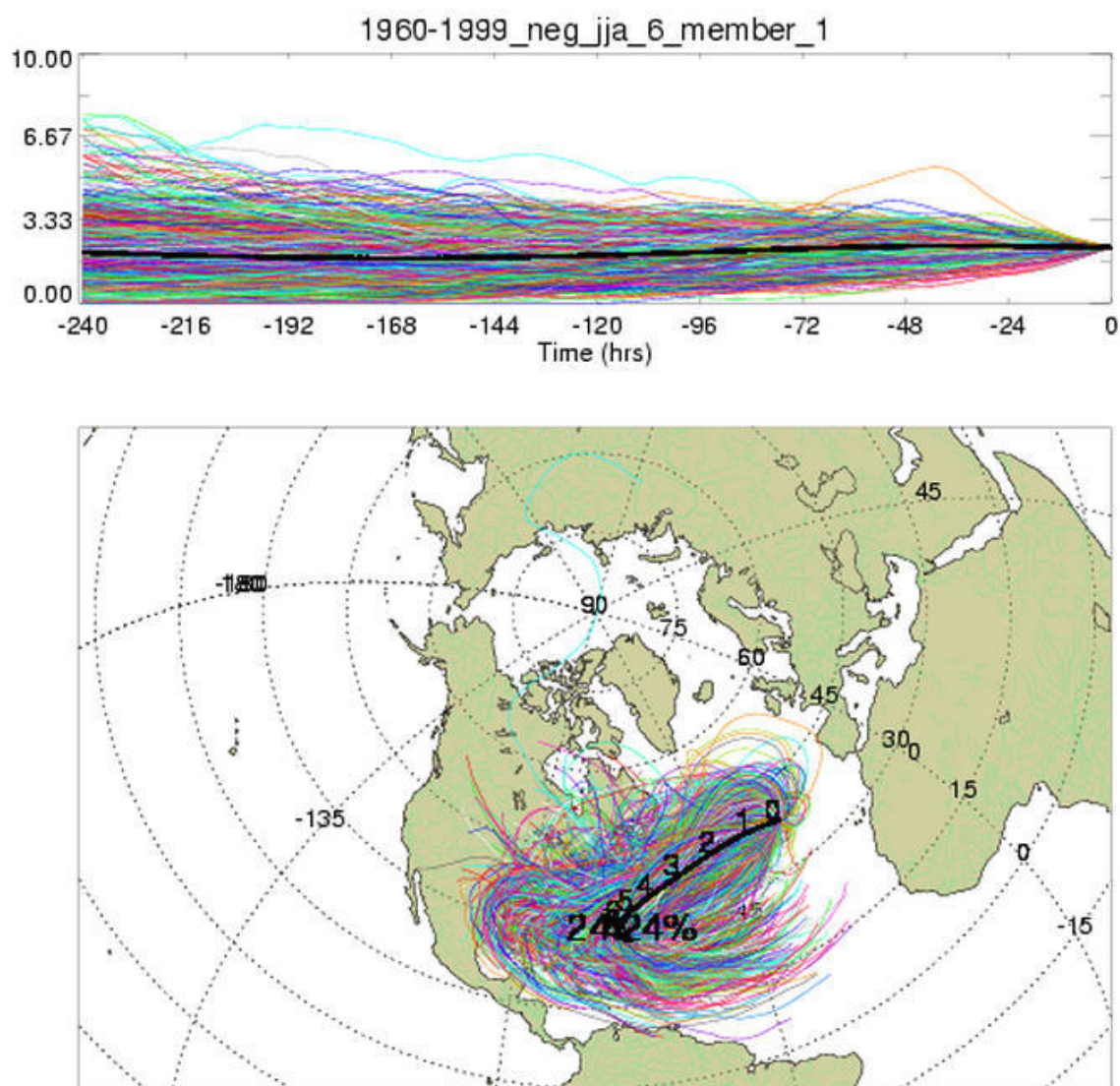


Figure D.43 Membership plot for summers with a negative NAOI value, 1960-1999.

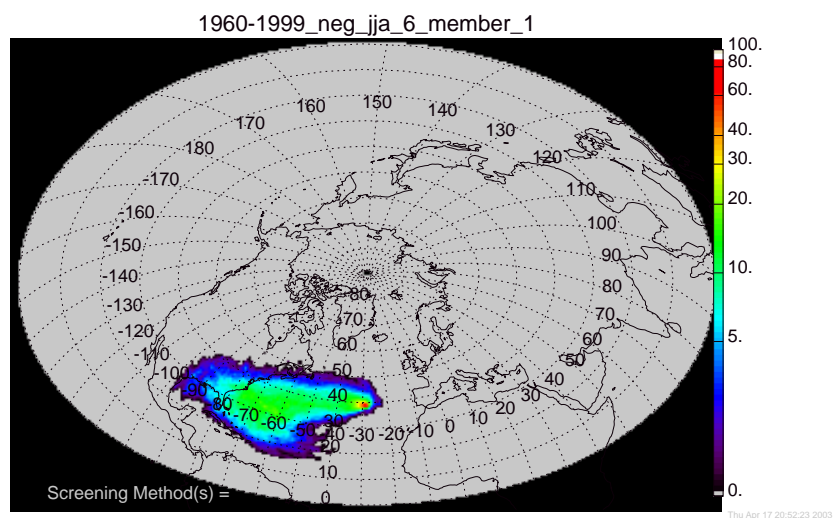


Figure D.44 Standard density membership plot for summers with a negative NAOI value, 1960-1999.

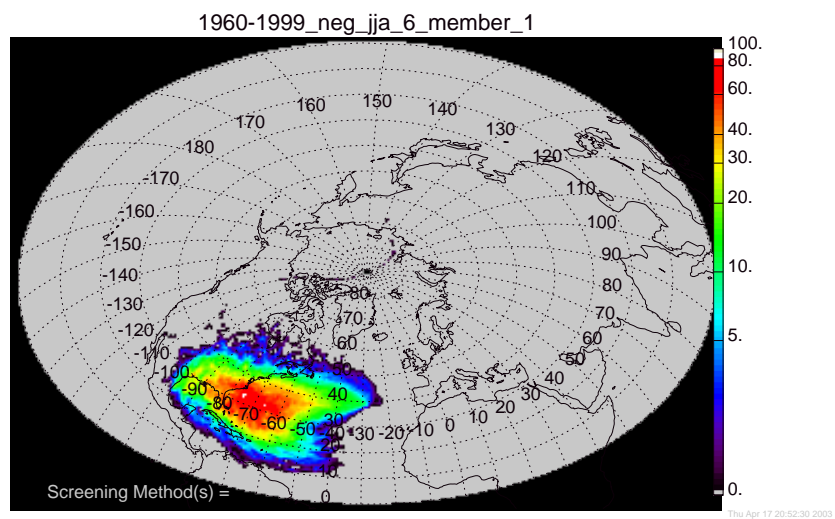


Figure D.45 Geometrically corrected density membership plot for summers with a negative NAOI value, 1960-1999.

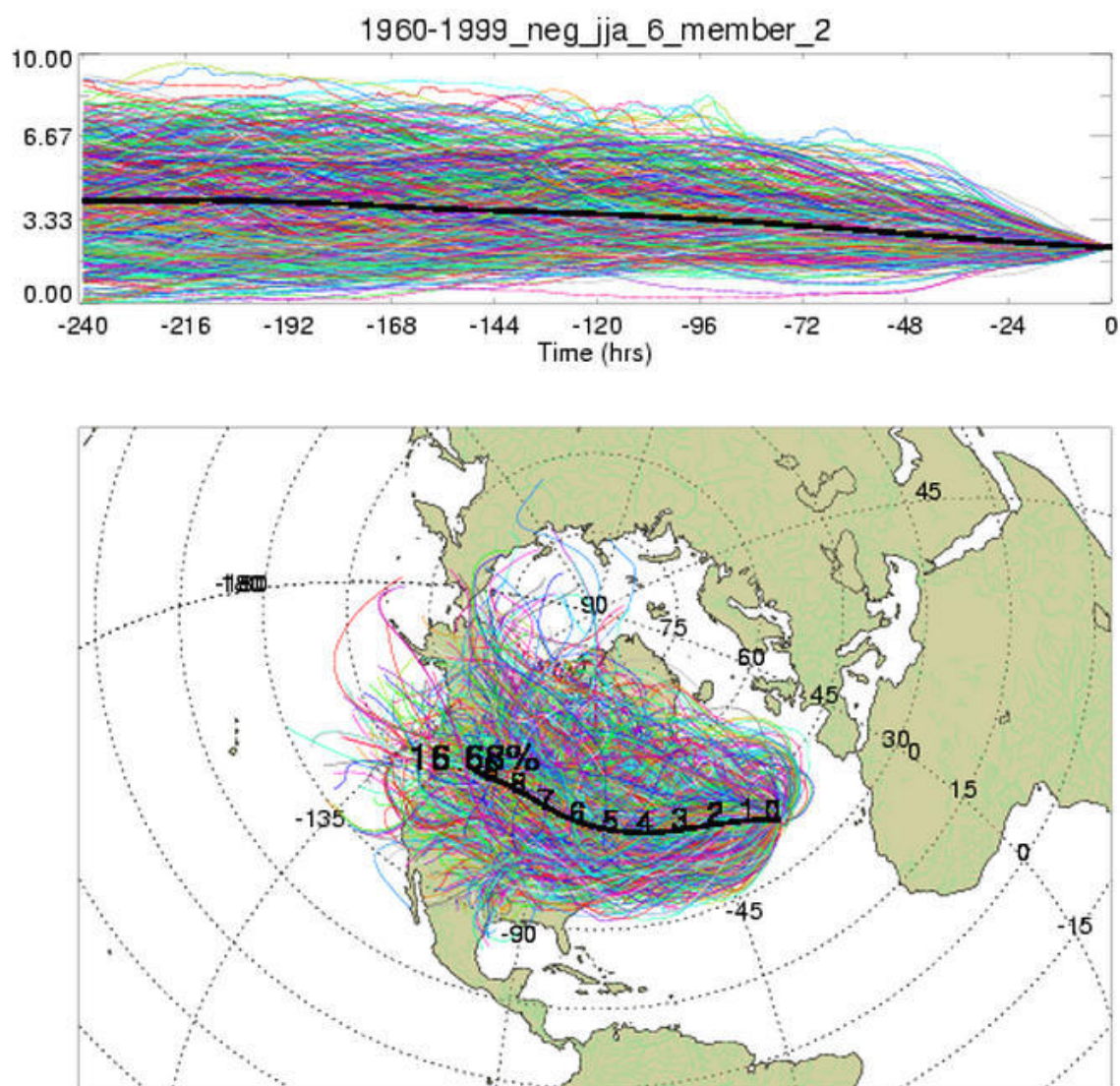


Figure D.46 Membership plot for summers with a negative NAOI value, 1960-1999.

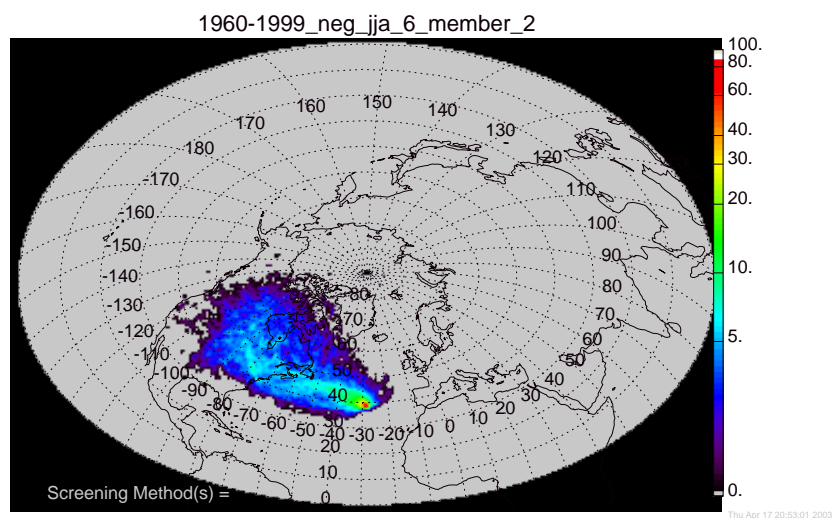


Figure D.47 Standard density membership plot for summers with a negative NAOI value, 1960-1999.

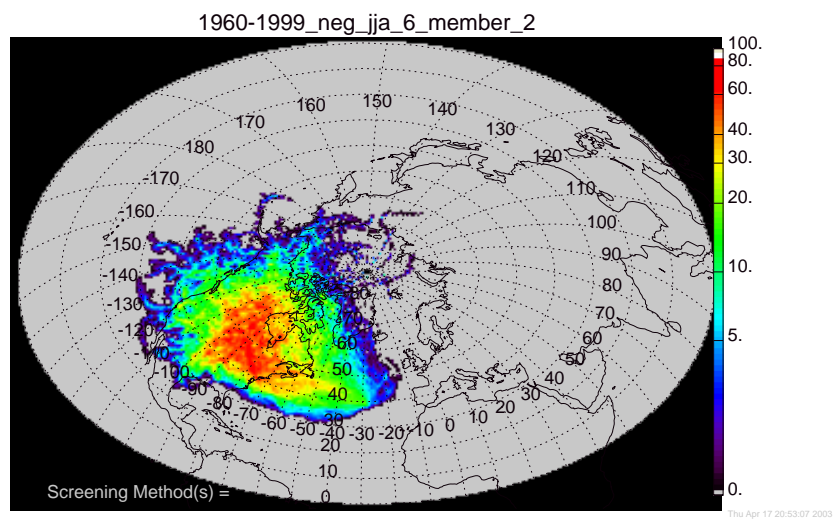


Figure D.48 Geometrically corrected density membership plot for summers with a negative NAOI value, 1960-1999.

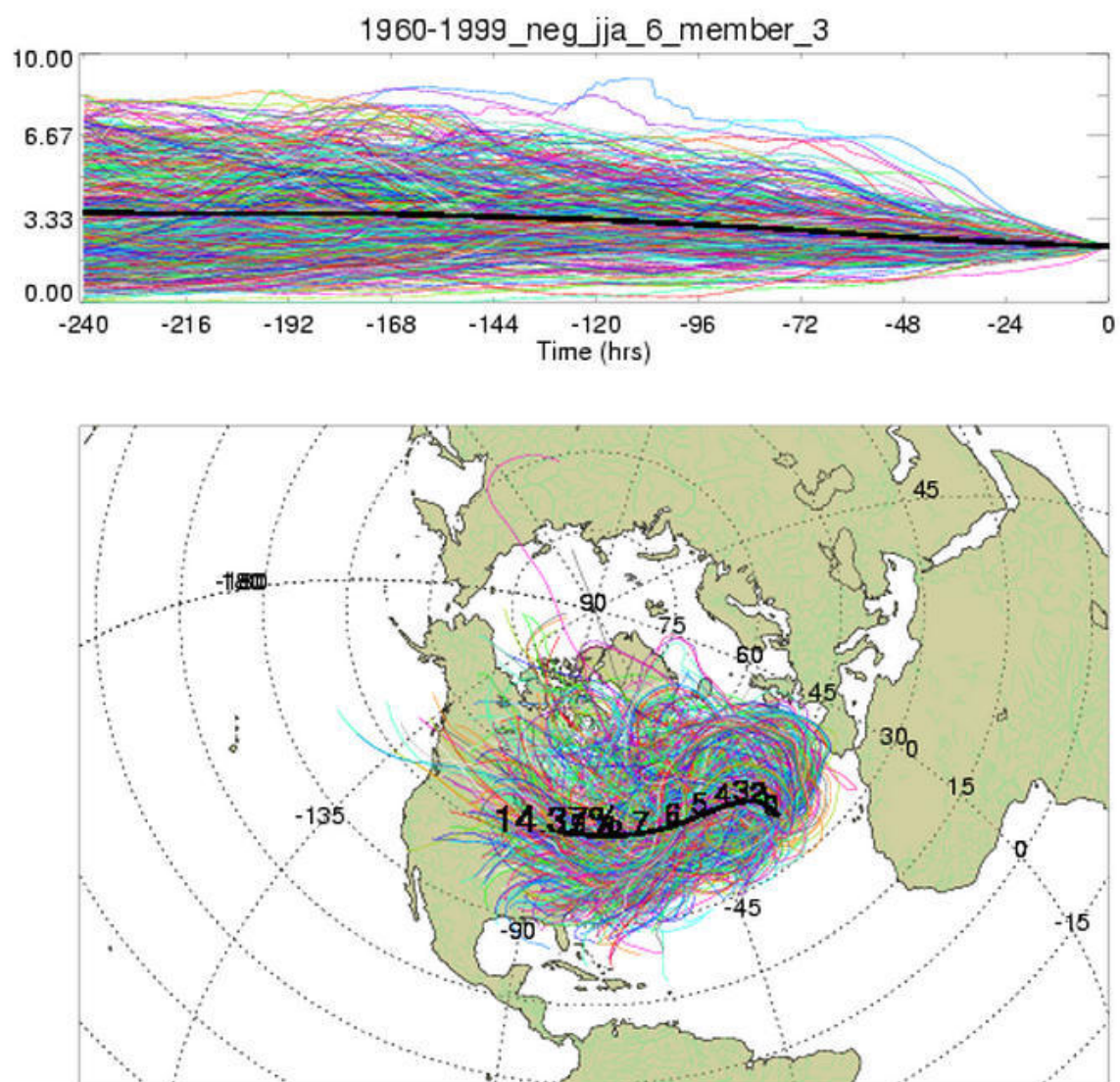


Figure D.49 Membership plot for summers with a negative NAOI value, 1960-1999.

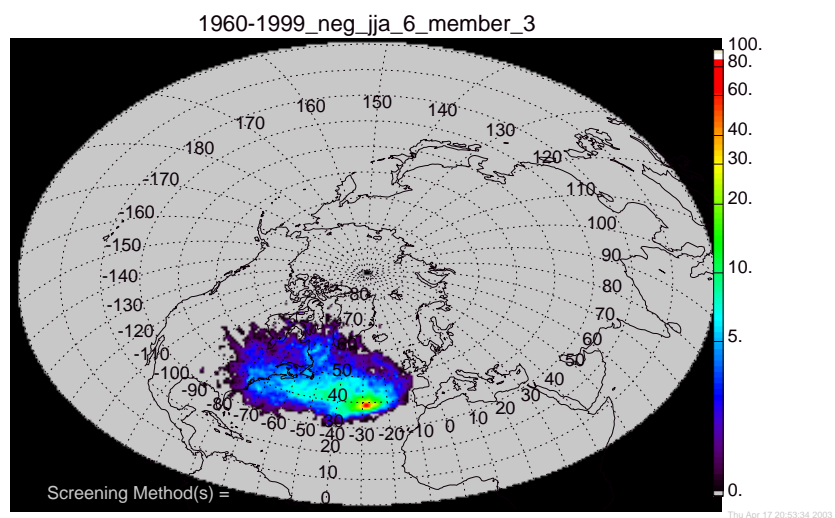


Figure D.50 Standard density membership plot for summers with a negative NAOI value, 1960-1999.

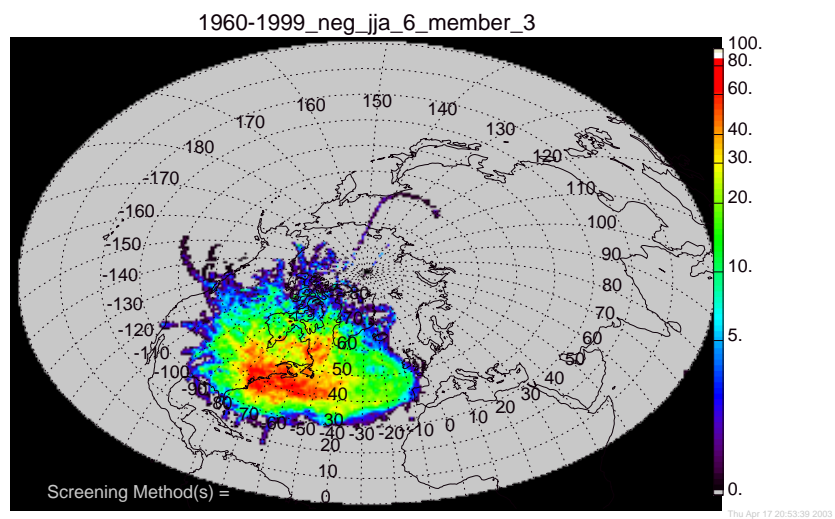


Figure D.51 Geometrically corrected density membership plot for summers with a negative NAOI value, 1960-1999.

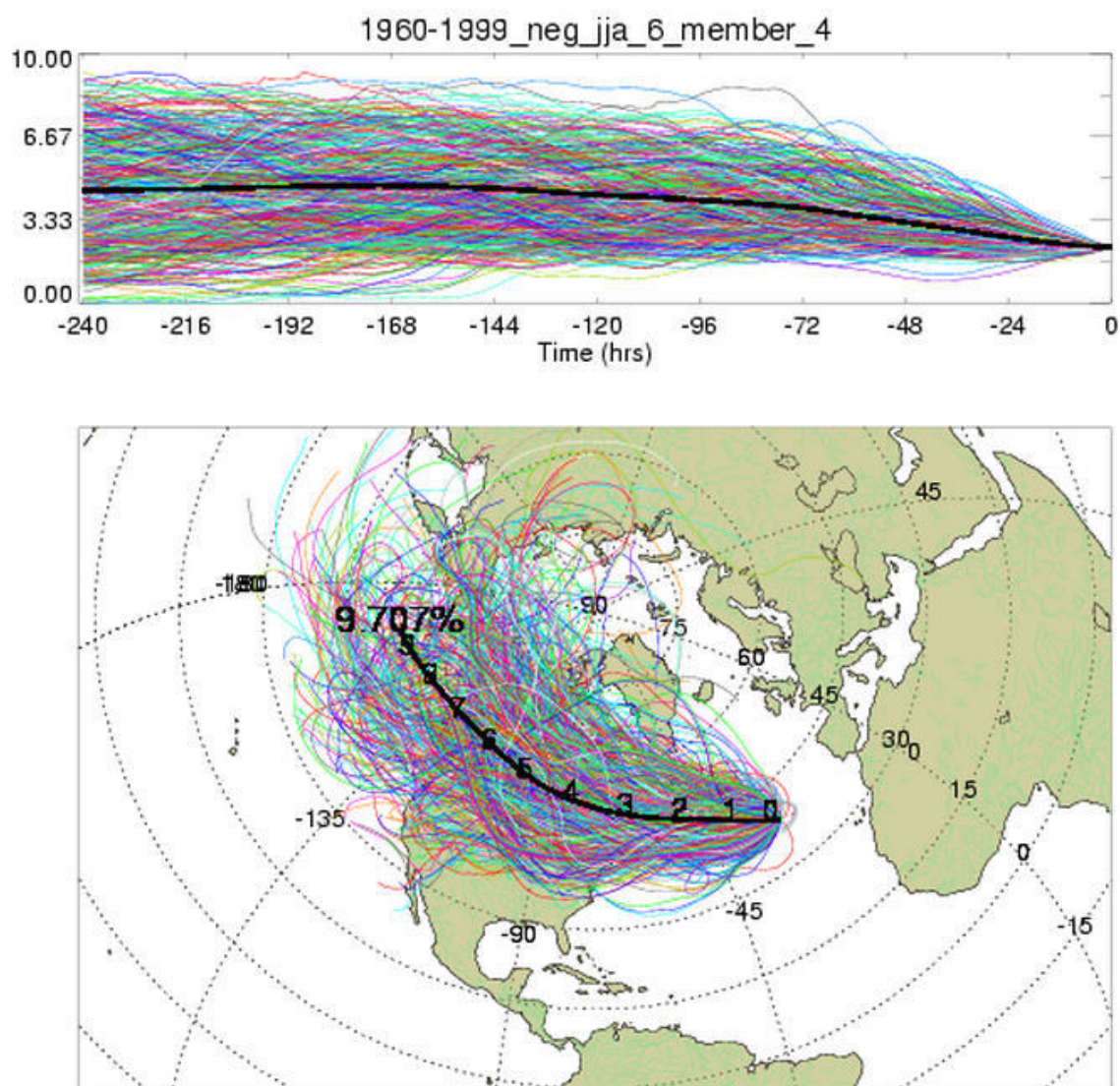


Figure D.52 Membership plot for summers with a negative NAOI value, 1960-1999.

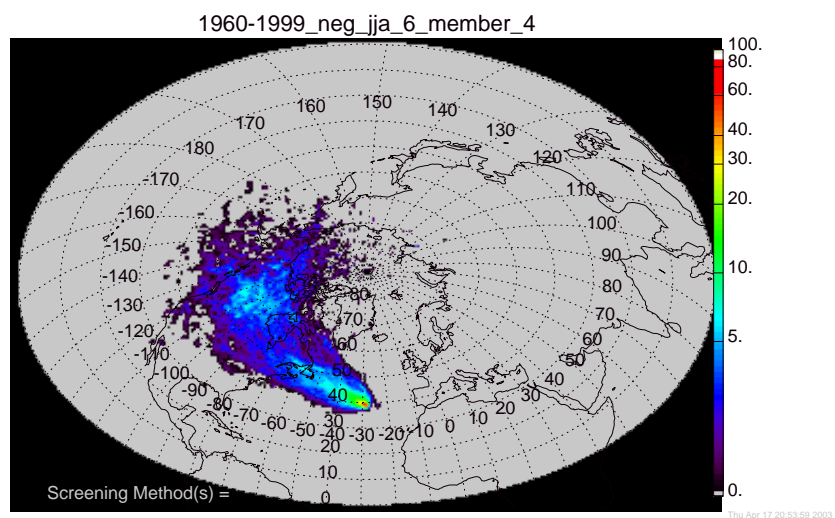


Figure D.53 Standard density membership plot for summers with a negative NAOI value, 1960-1999.

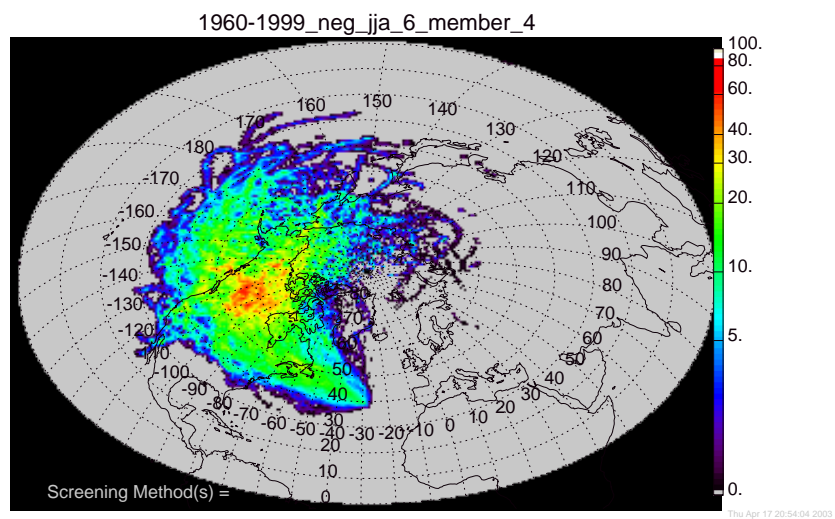


Figure D.54 Geometrically corrected density membership plot for summers with a negative NAOI value, 1960-1999.

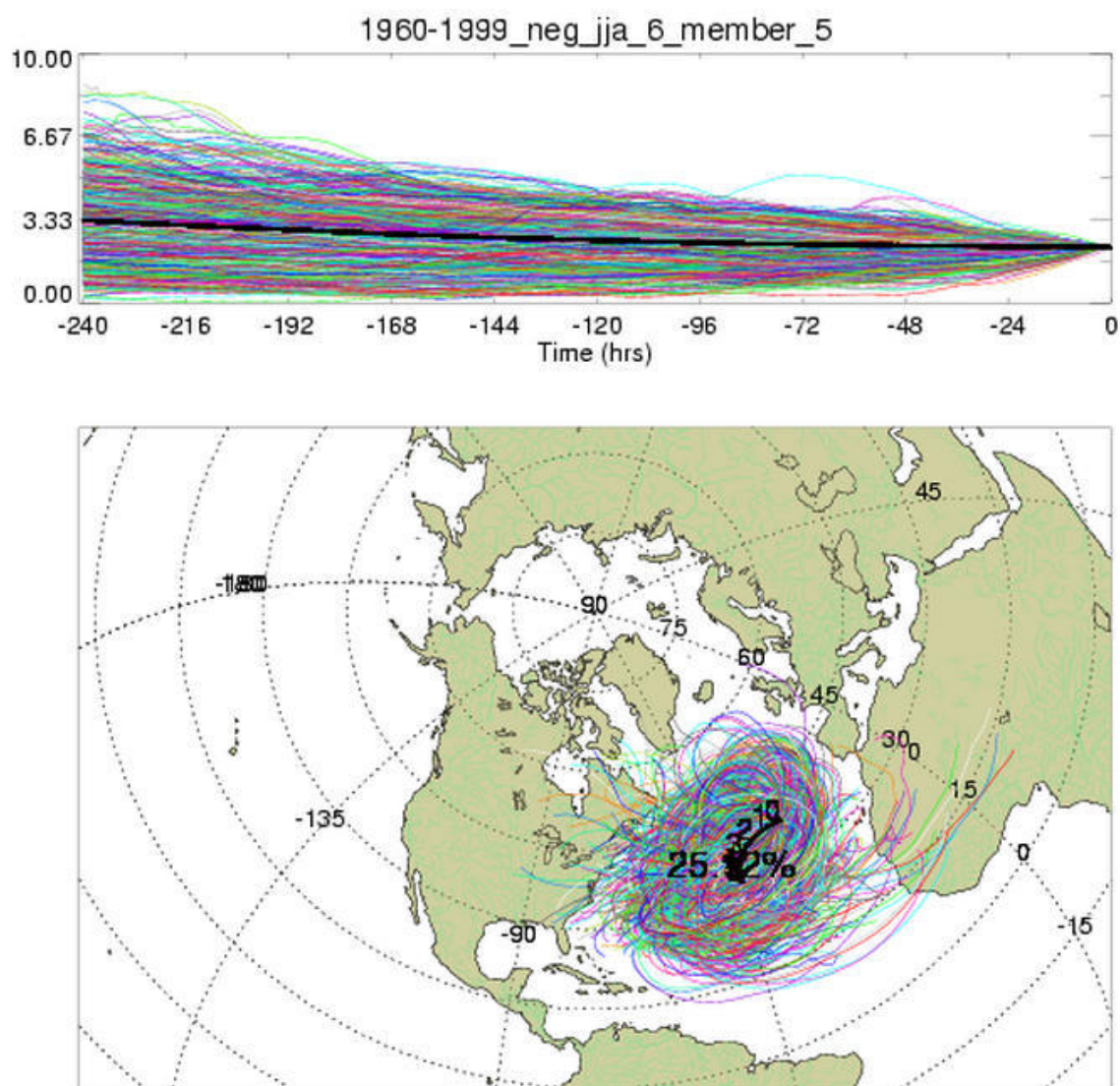


Figure D.55 Membership plot for summers with a negative NAOI value, 1960-1999.

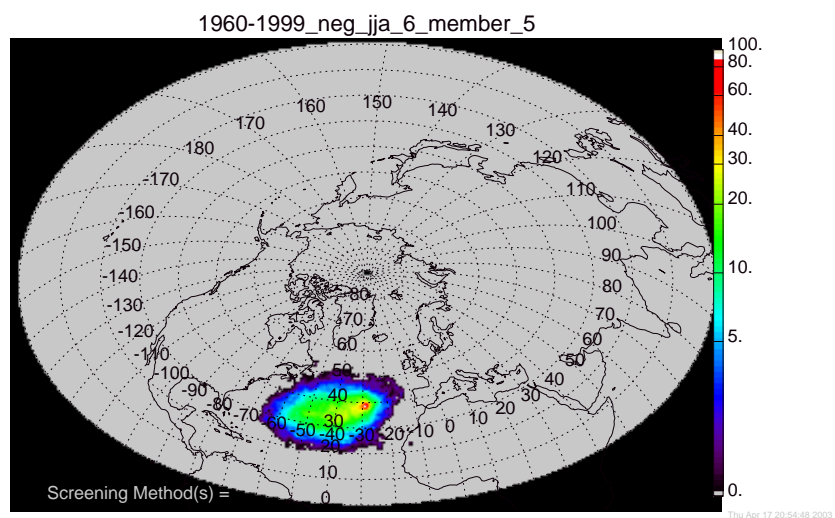


Figure D.56 Standard density membership plot for summers with a negative NAOI value, 1960-1999.

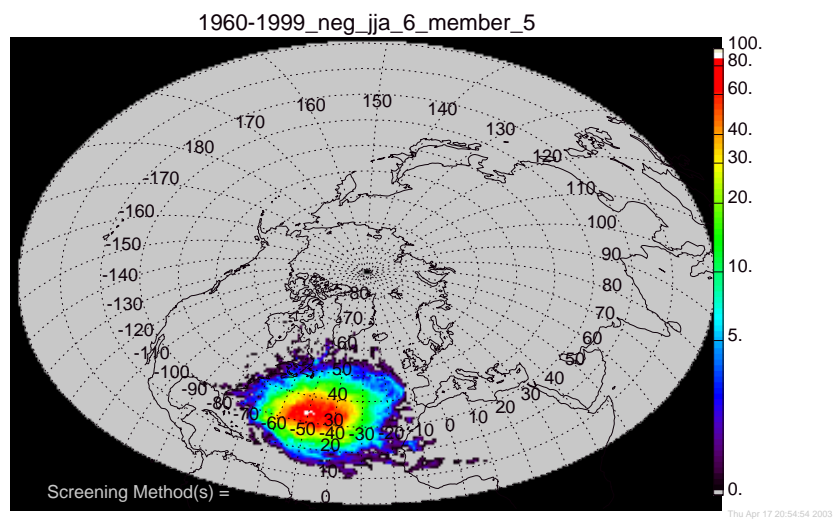


Figure D.57 Geometrically corrected density membership plot for summers with a negative NAOI value, 1960-1999.

Appendix E

Membership plots for fall periods

E.1 Membership plots for the average fall

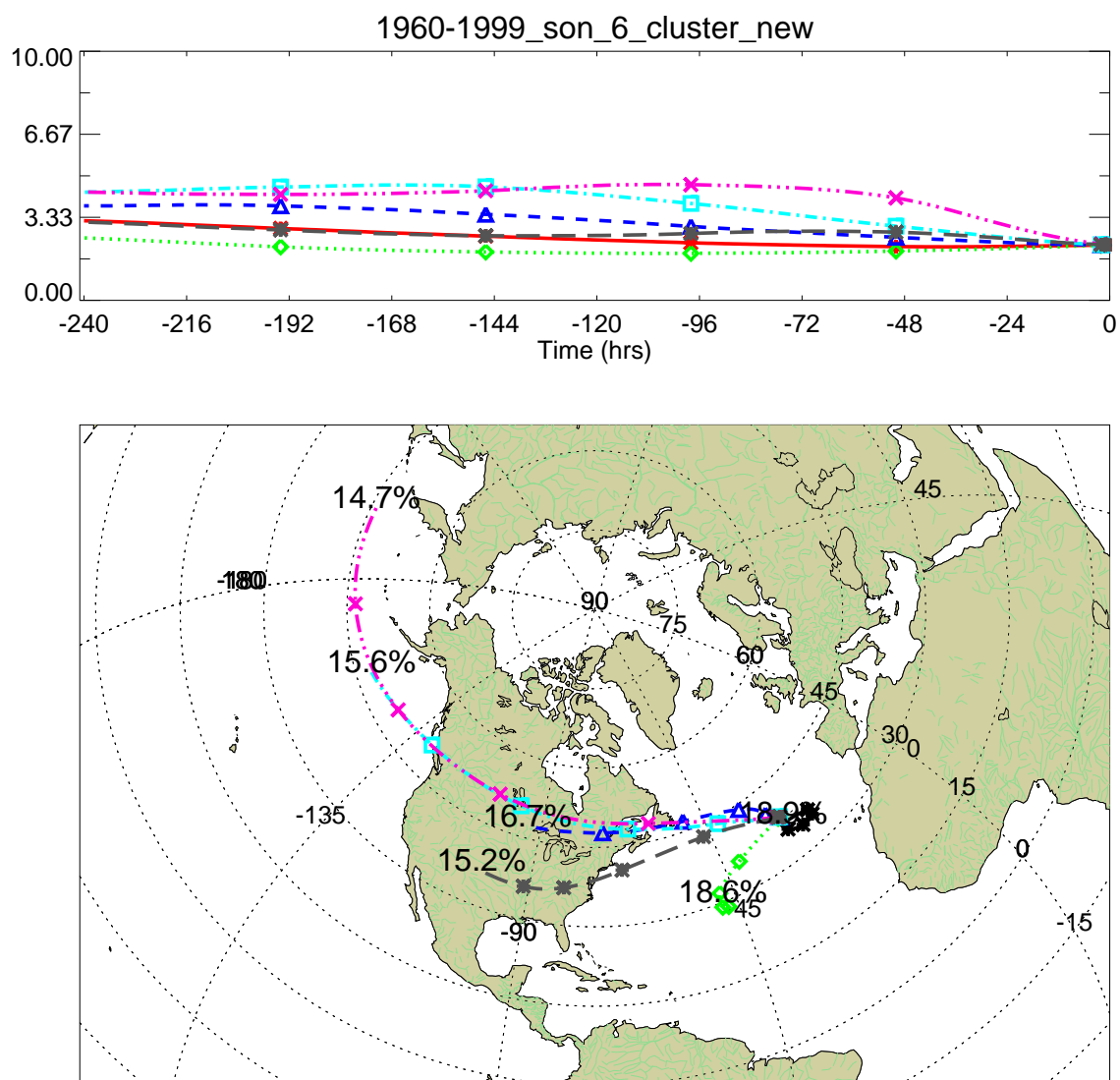


Figure E.1 Cluster plot for the average for fall, 1960-1999.

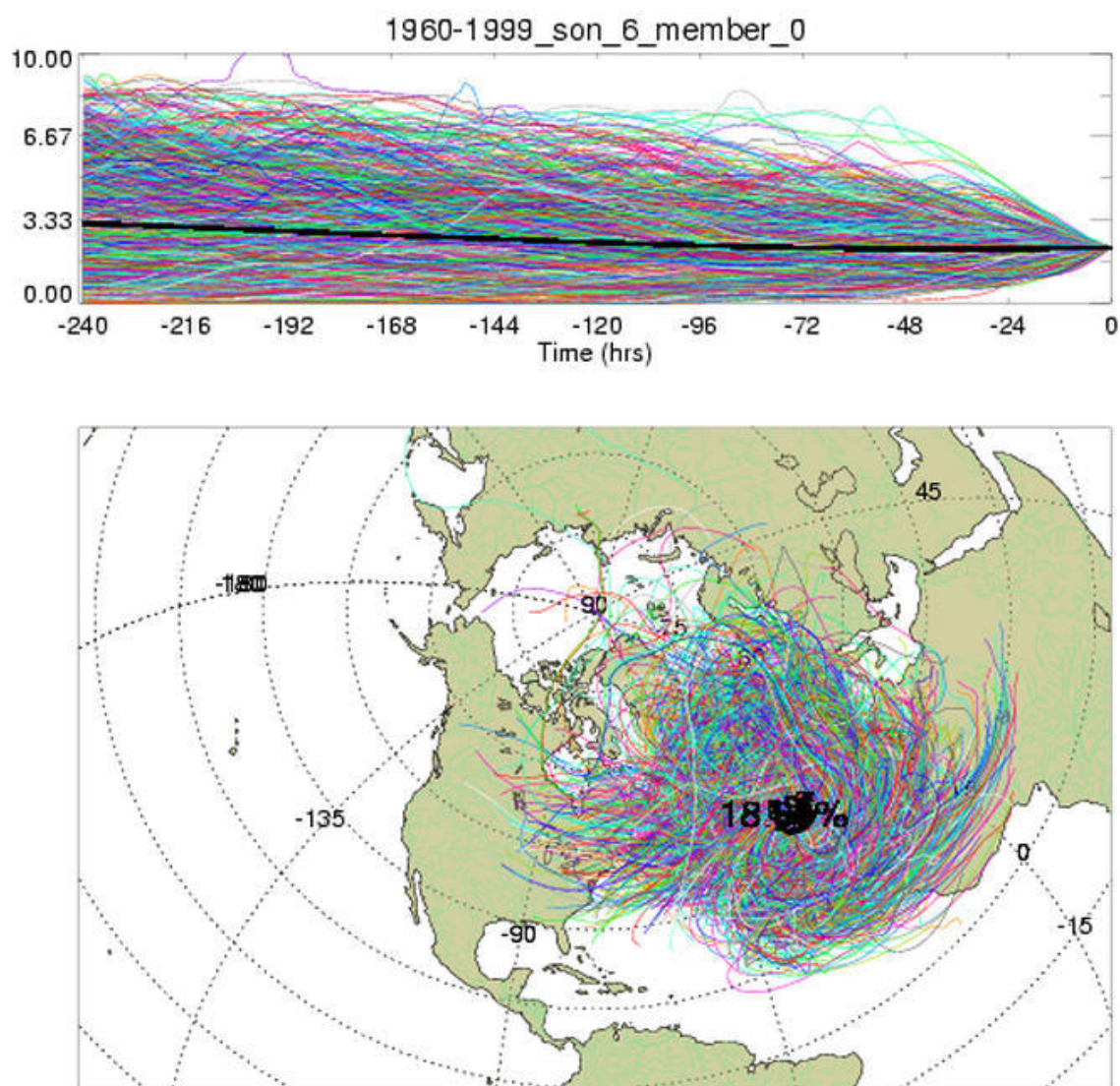


Figure E.2 Membership plot for falls, 1960-1999.

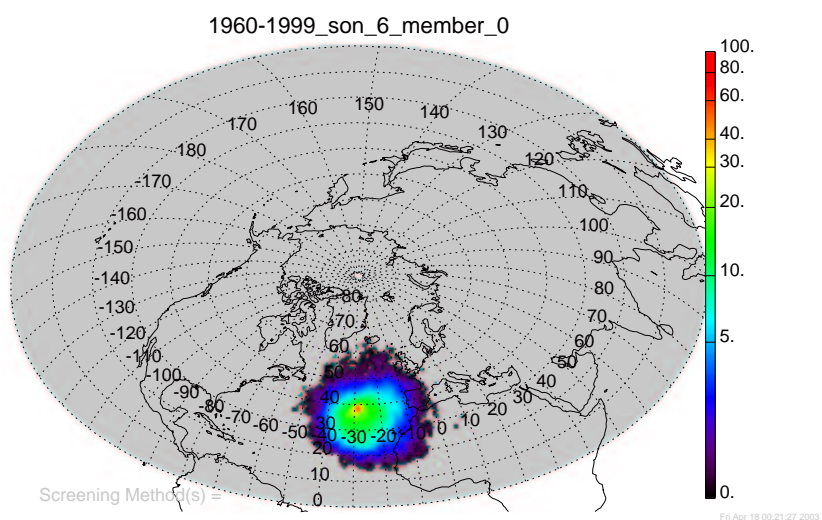


Figure E.3 Standard density membership plot for falls, 1960-1999.

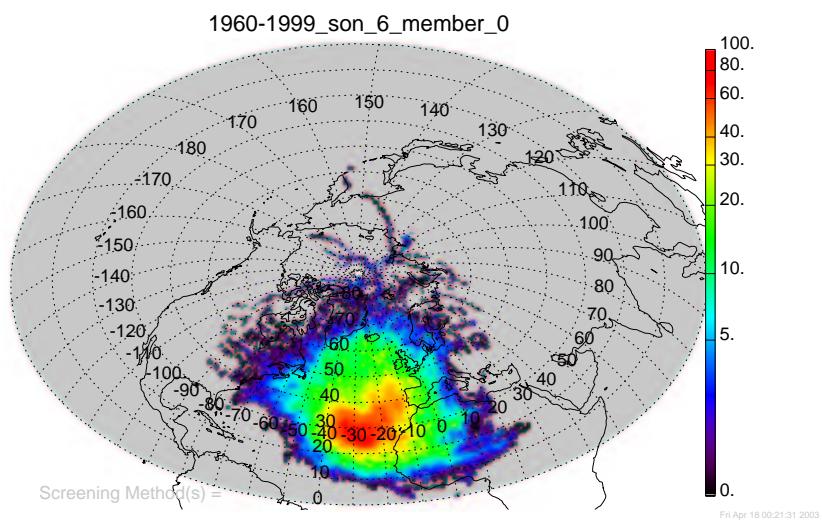


Figure E.4 Geometrically corrected density membership plot for falls, 1960-1999.

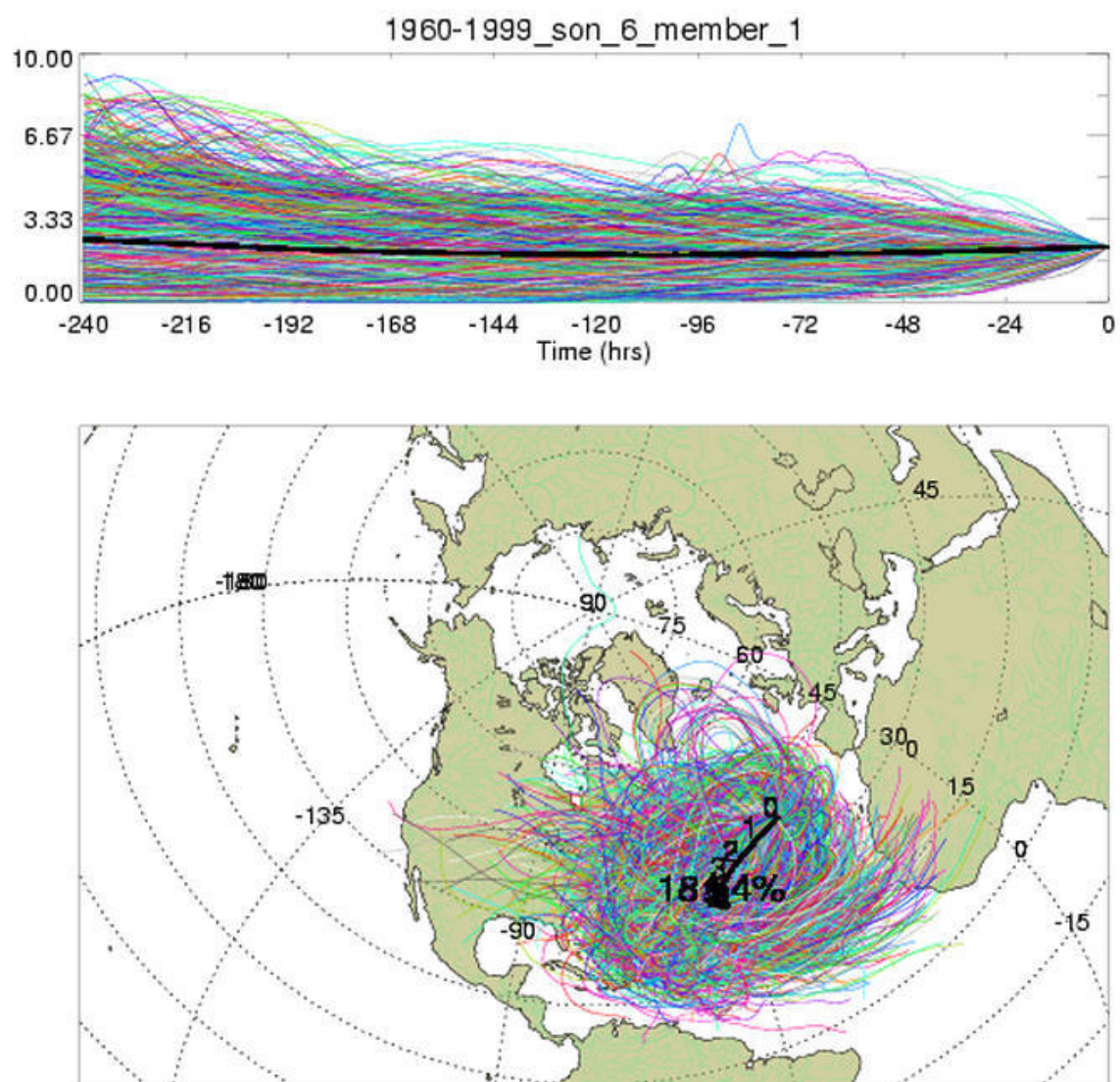


Figure E.5 Membership plot for falls, 1960-1999.

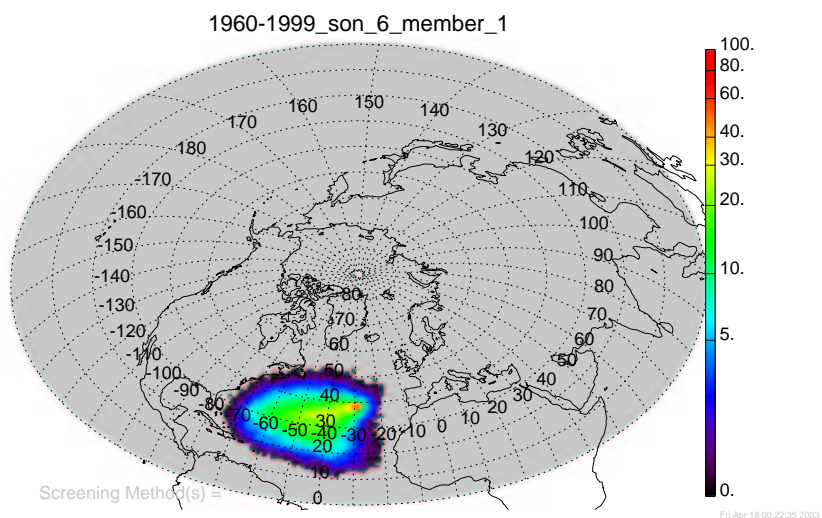


Figure E.6 Standard density membership plot for falls, 1960-1999.

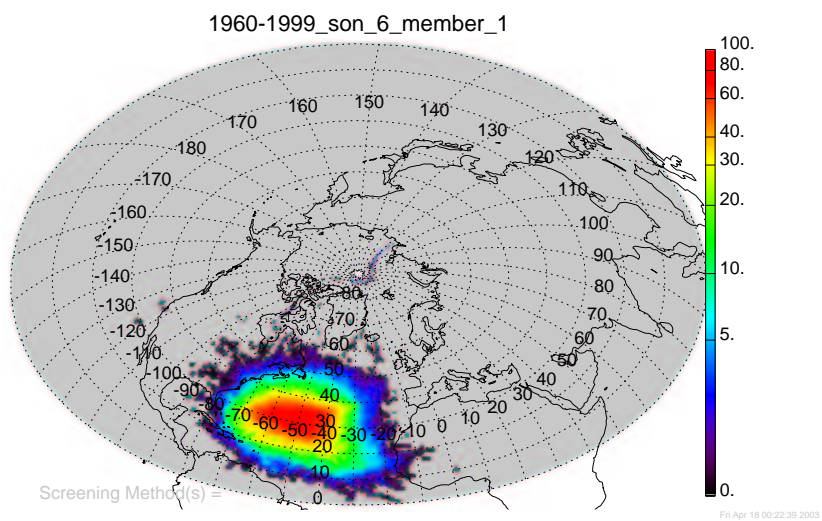


Figure E.7 Geometrically corrected density membership plot for falls, 1960-1999.

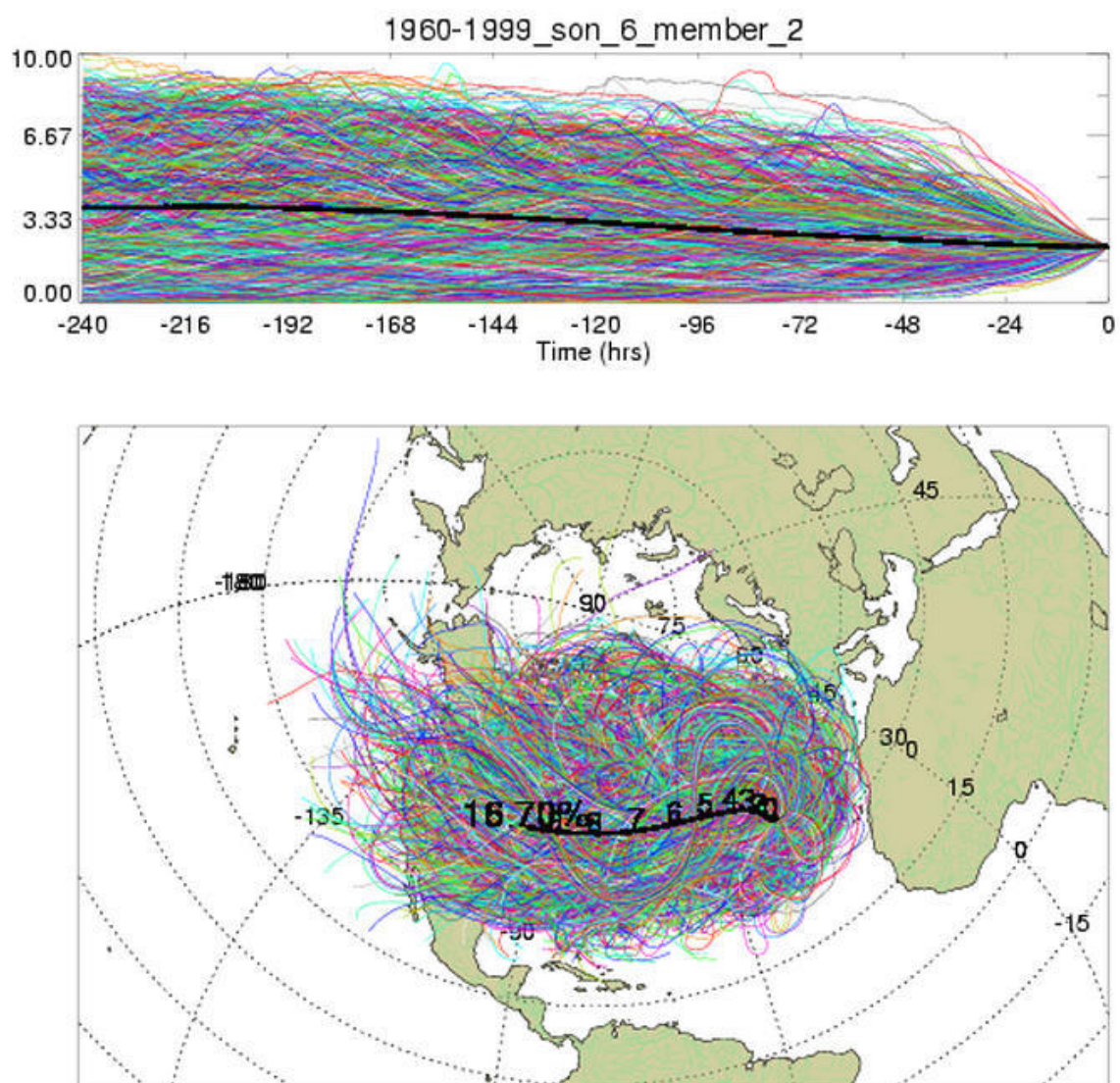


Figure E.8 Membership plot for falls, 1960-1999.

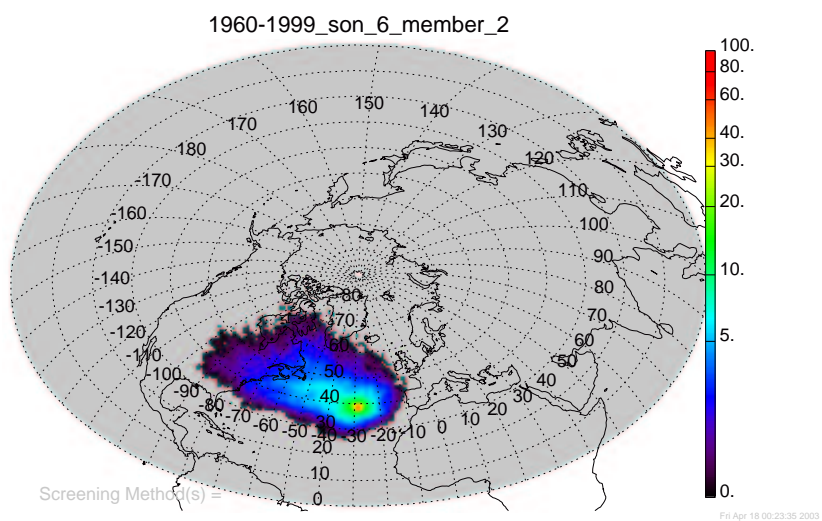


Figure E.9 Standard density membership plot for falls, 1960-1999.

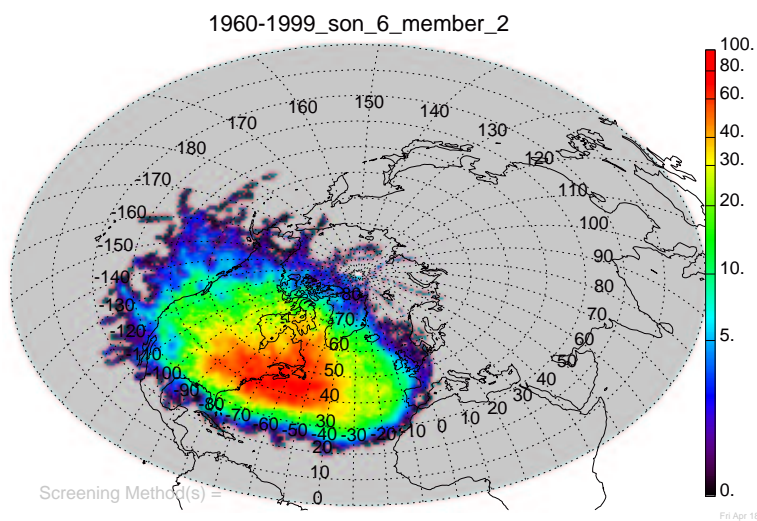


Figure E.10 Geometrically corrected density membership plot for falls, 1960-1999.

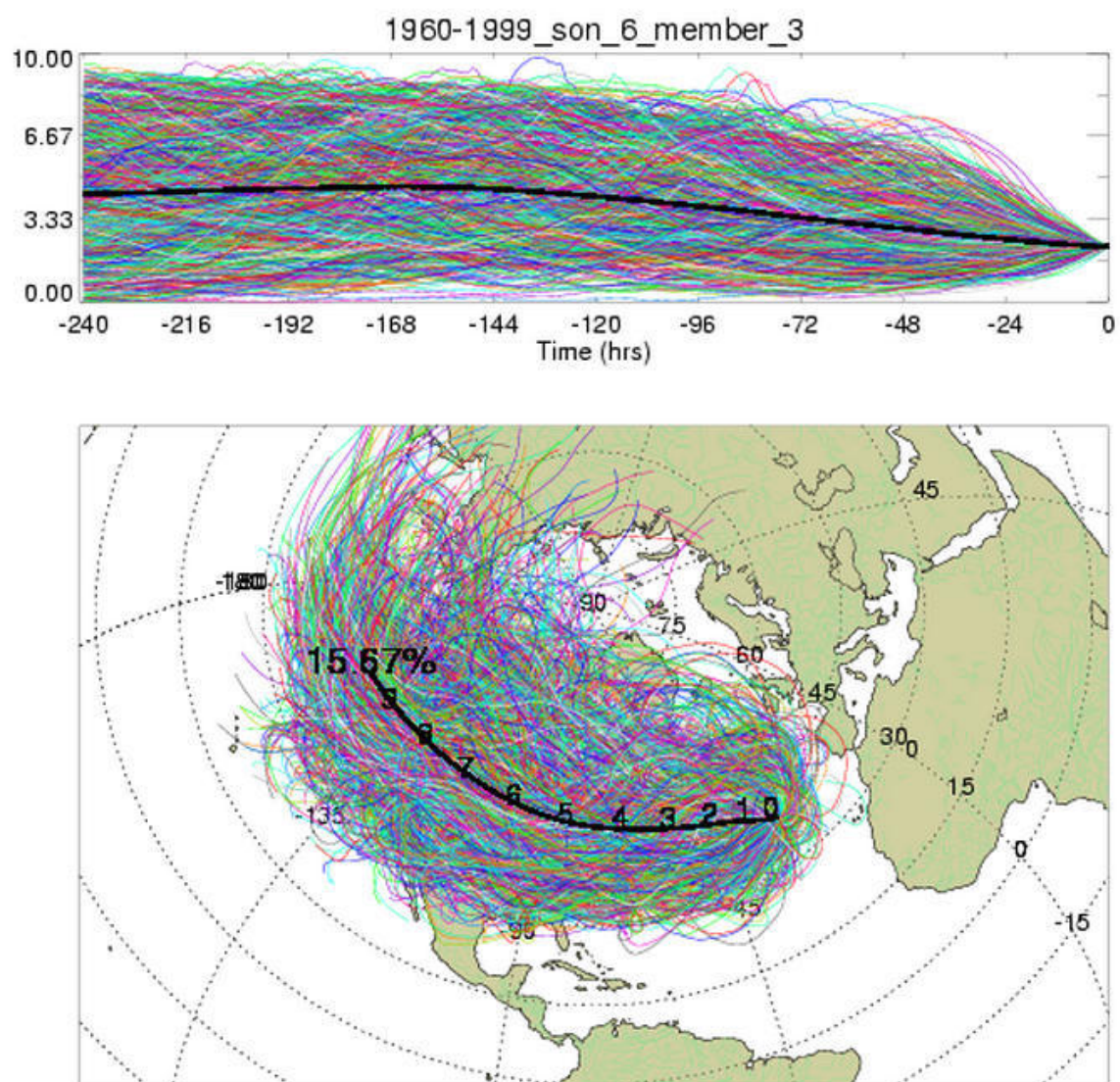


Figure E.11 Membership plot for falls, 1960-1999.

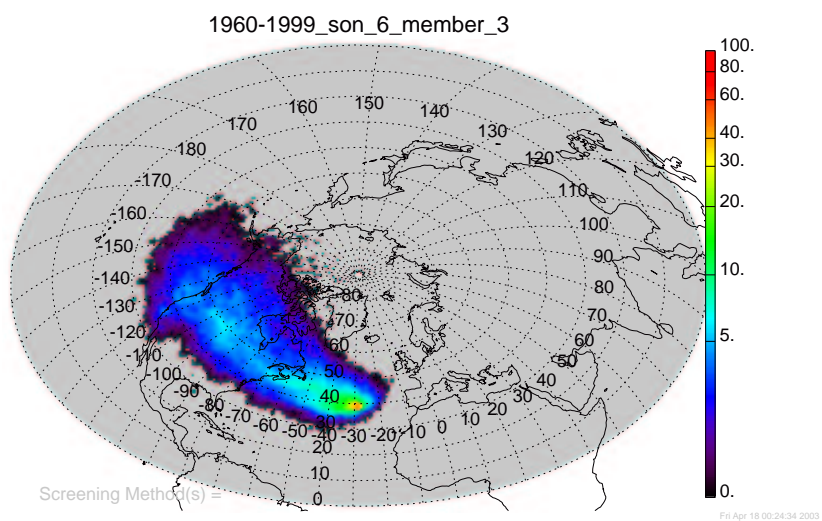


Figure E.12 Standard density membership plot for falls, 1960-1999.

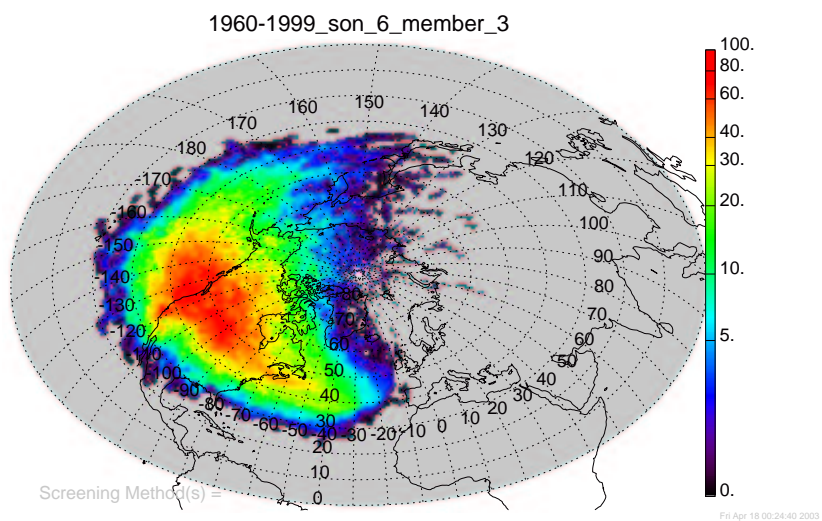


Figure E.13 Geometrically corrected density membership plot for falls, 1960-1999.

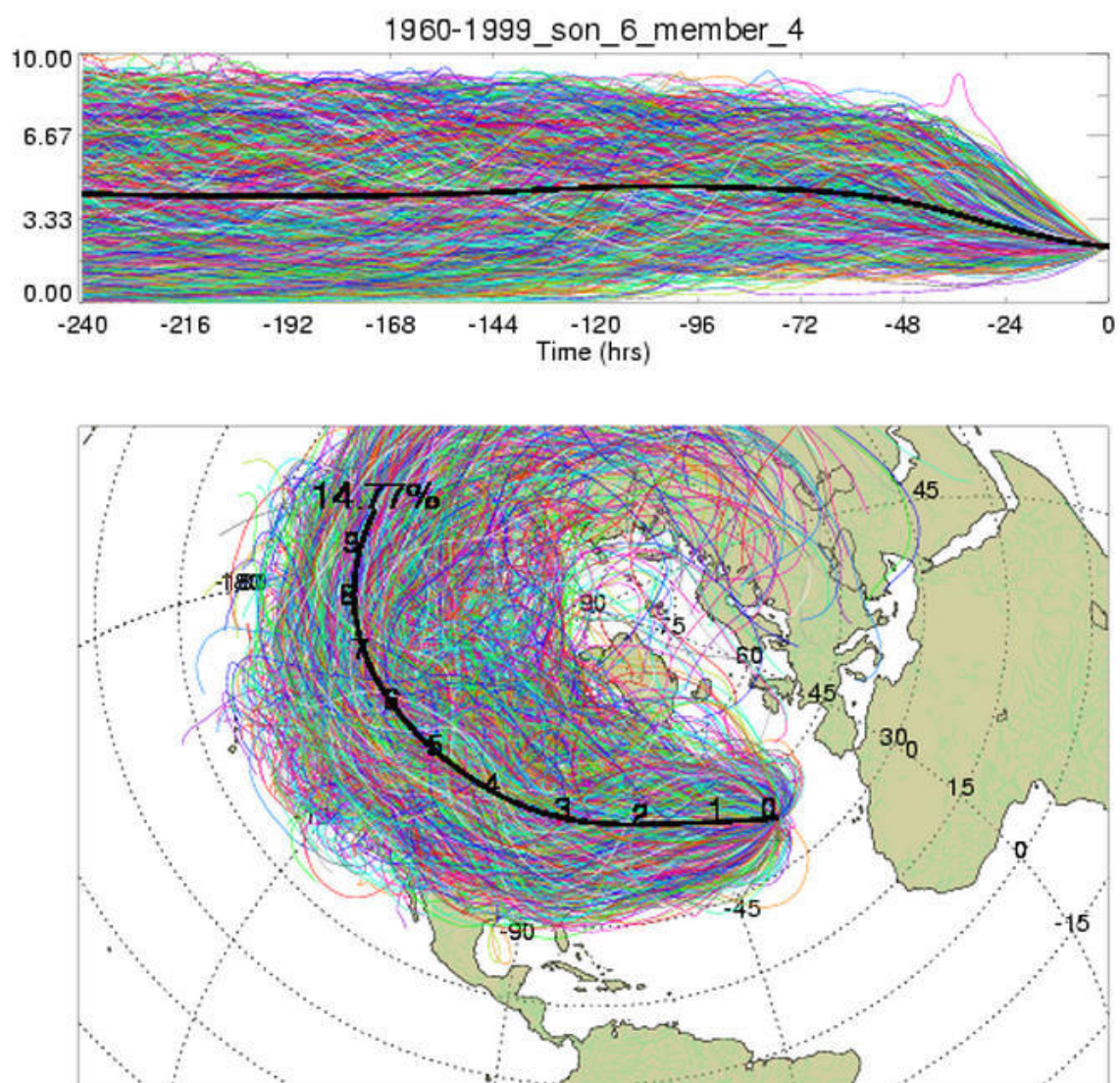


Figure E.14 Membership plot for falls, 1960-1999.

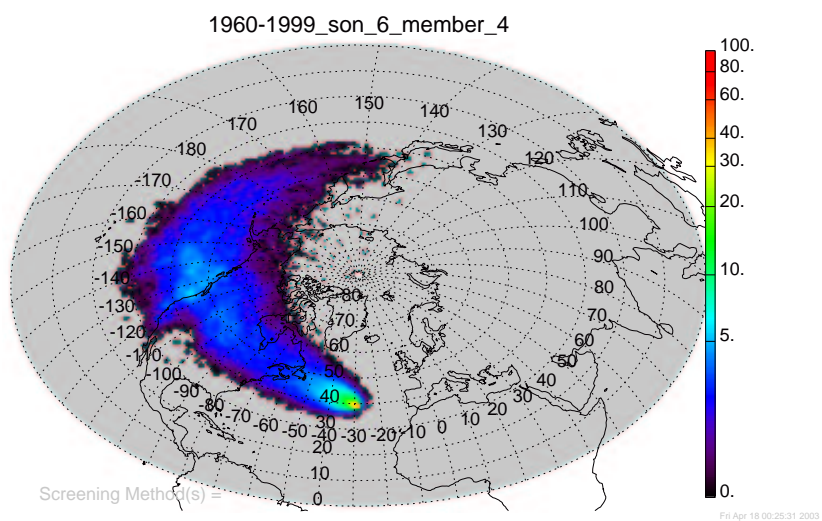


Figure E.15 Standard density membership plot for falls, 1960-1999.

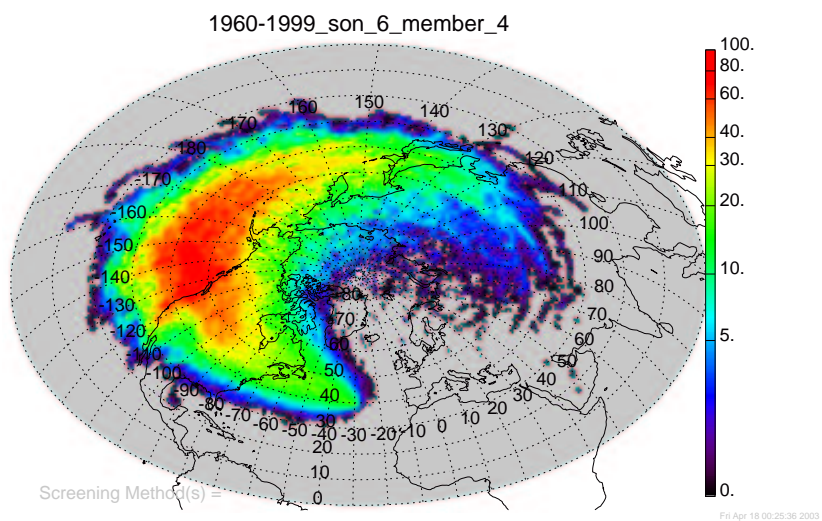


Figure E.16 Geometrically corrected density membership plot for falls, 1960-1999.

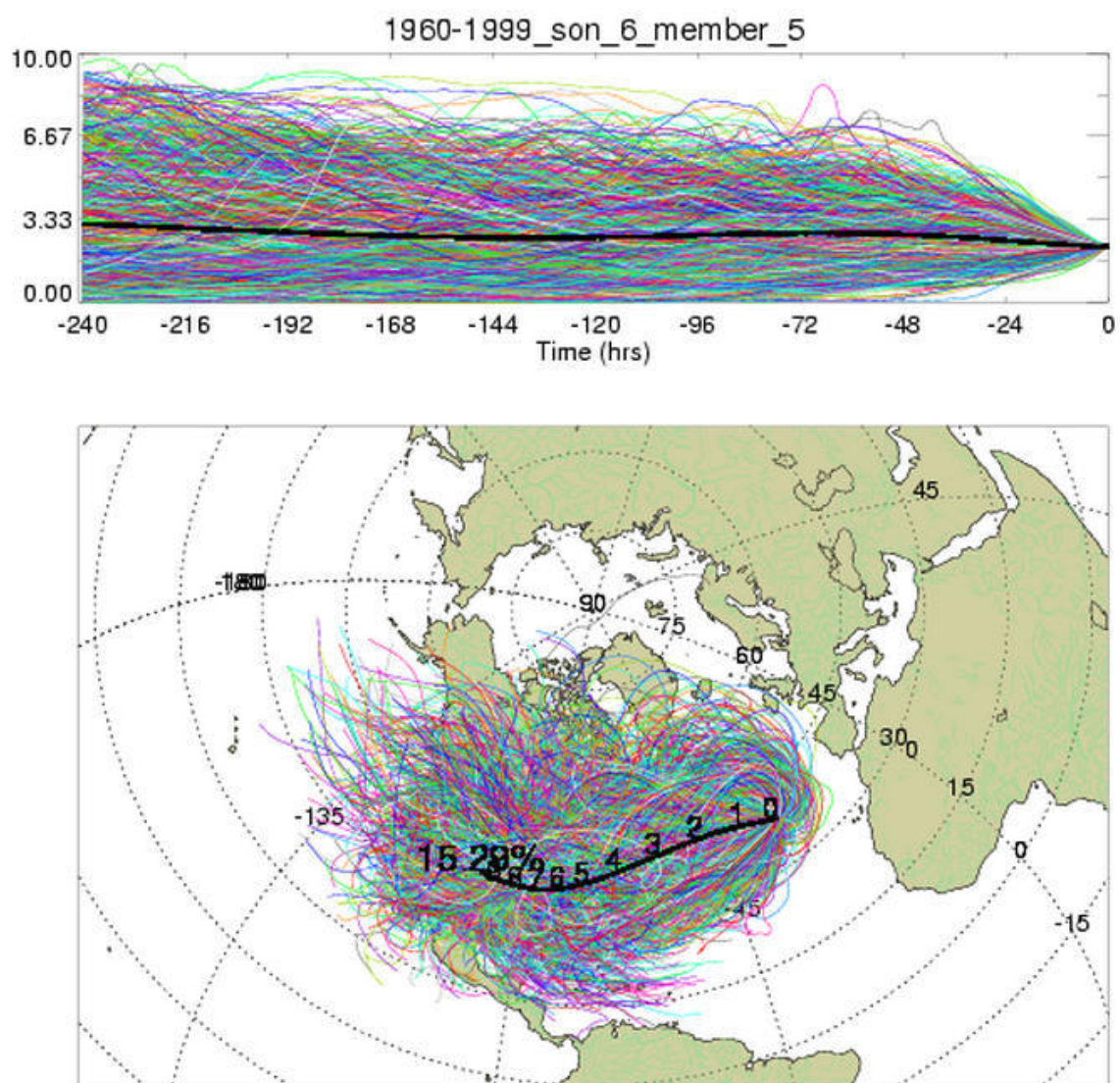


Figure E.17 Membership plot for falls, 1960-1999.

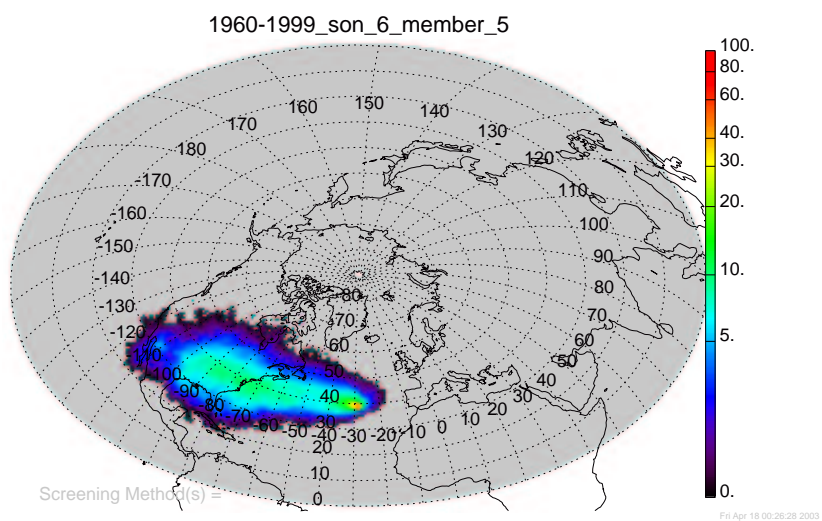


Figure E.18 Standard density membership plot for falls, 1960-1999.

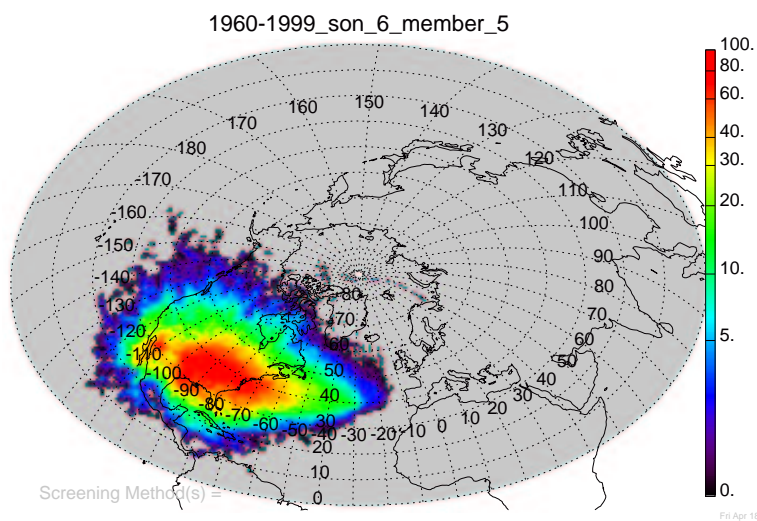


Figure E.19 Geometrically corrected density membership plot for falls, 1960-1999.

E.2 Membership plots for the positive NAO falls

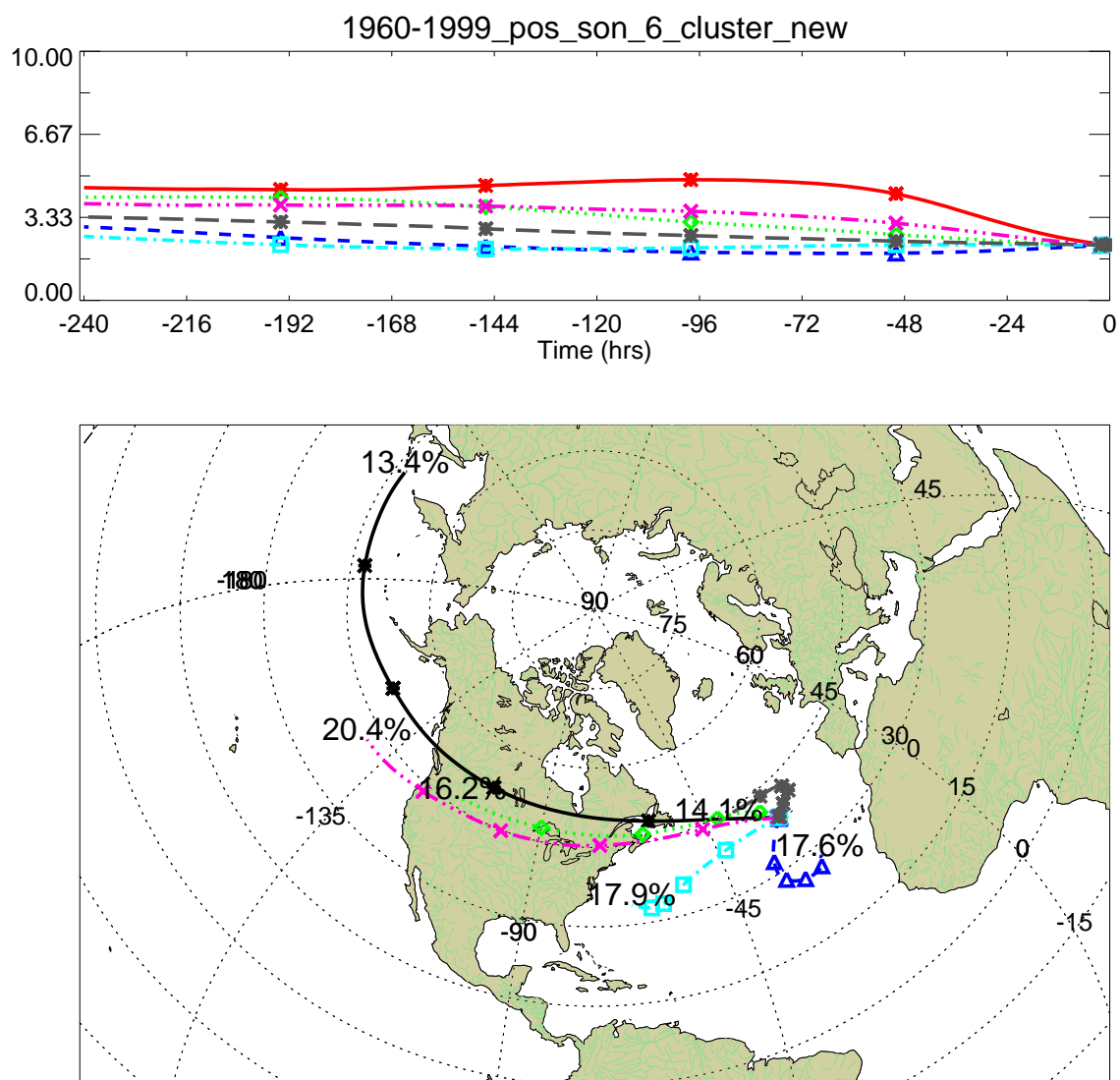


Figure E.20 Cluster plot for the positive NAO for falls, 1960-1999.

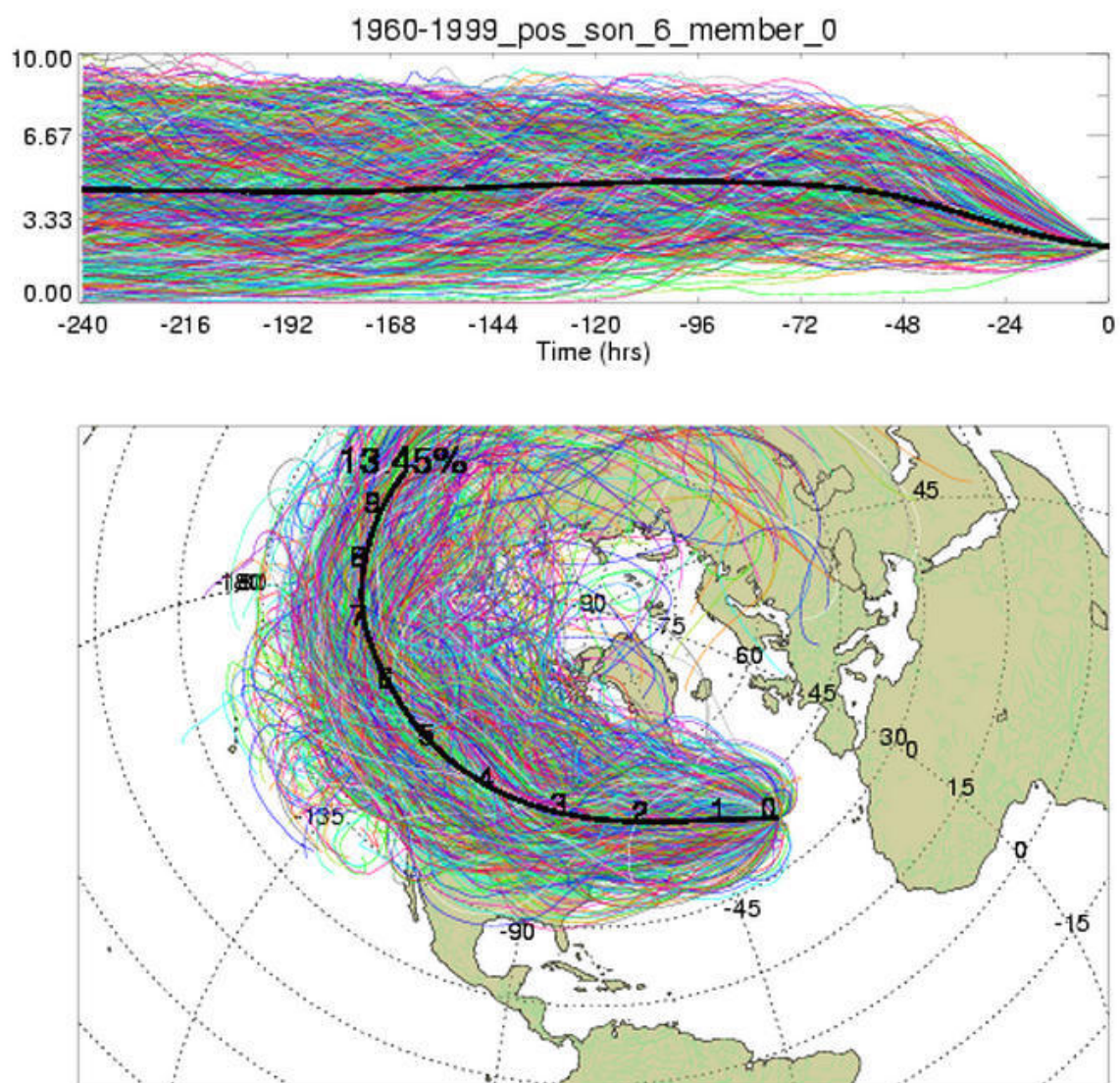


Figure E.21 Membership plot for falls with a positive NAOI value, 1960-1999.

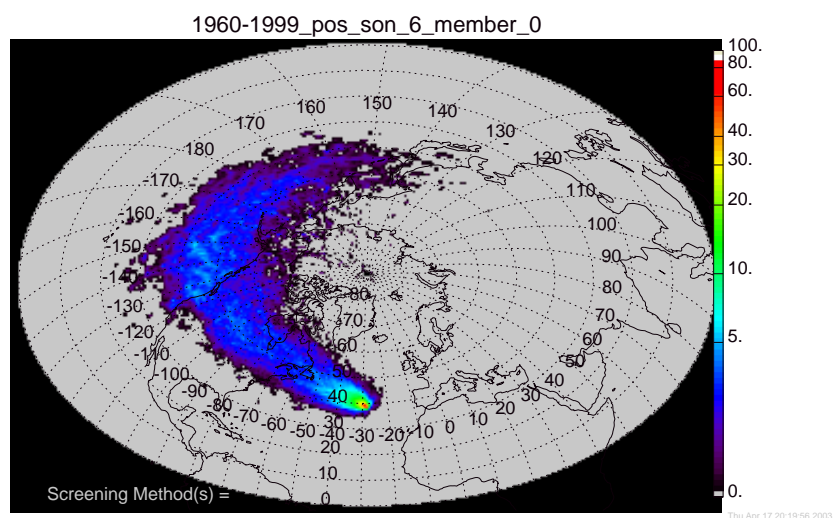


Figure E.22 Standard density membership plot for falls with a positive NAOI value, 1960-1999.

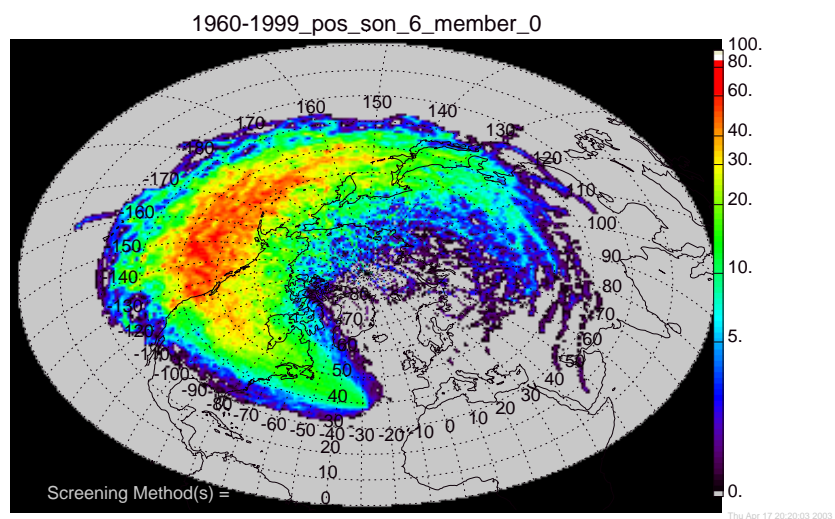


Figure E.23 Geometrically corrected density membership plot for falls with a positive NAOI value, 1960-1999.

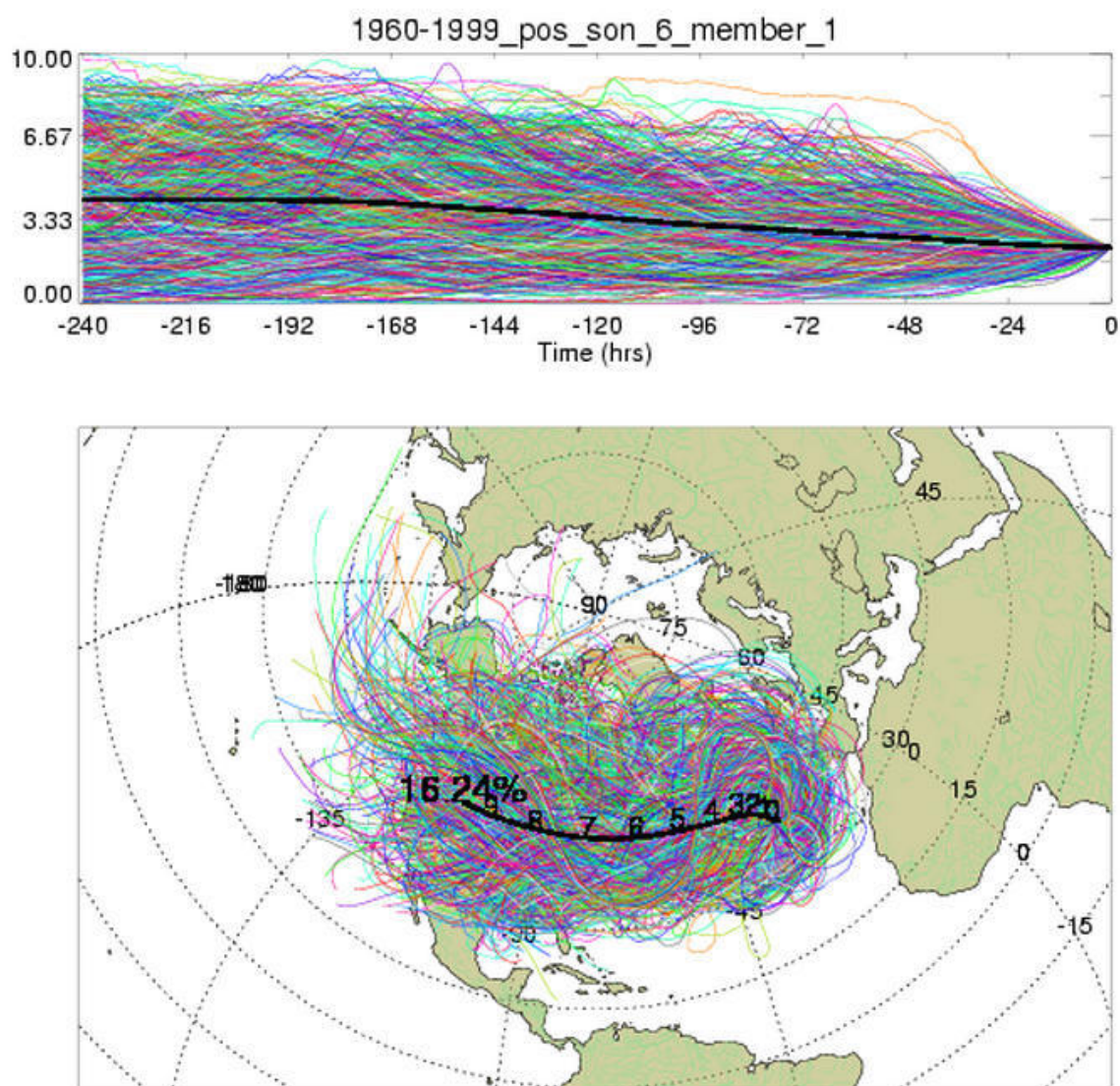


Figure E.24 Membership plot for falls with a positive NAOI value, 1960-1999.

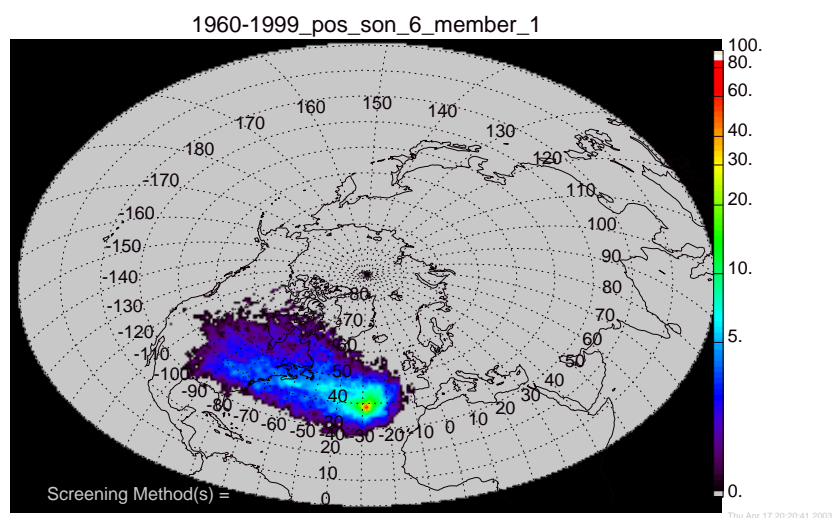


Figure E.25 Standard density membership plot for falls with a positive NAOI value, 1960-1999.

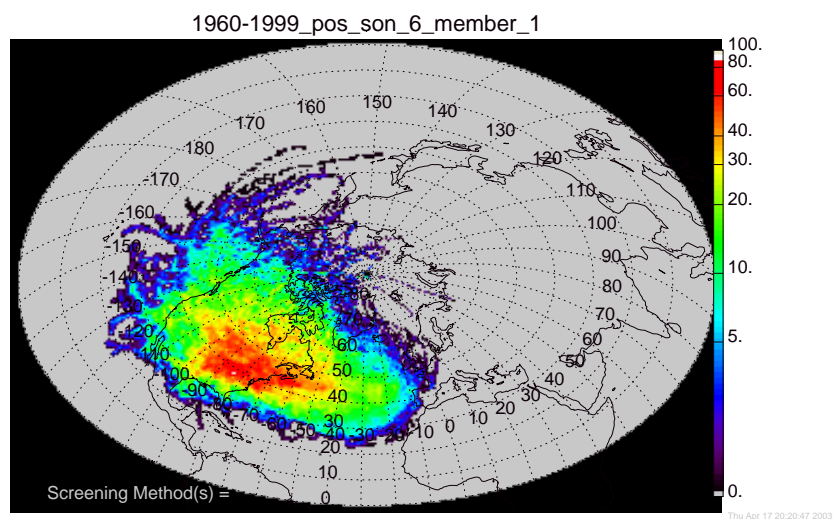


Figure E.26 Geometrically corrected density membership plot for falls with a positive NAOI value, 1960-1999.

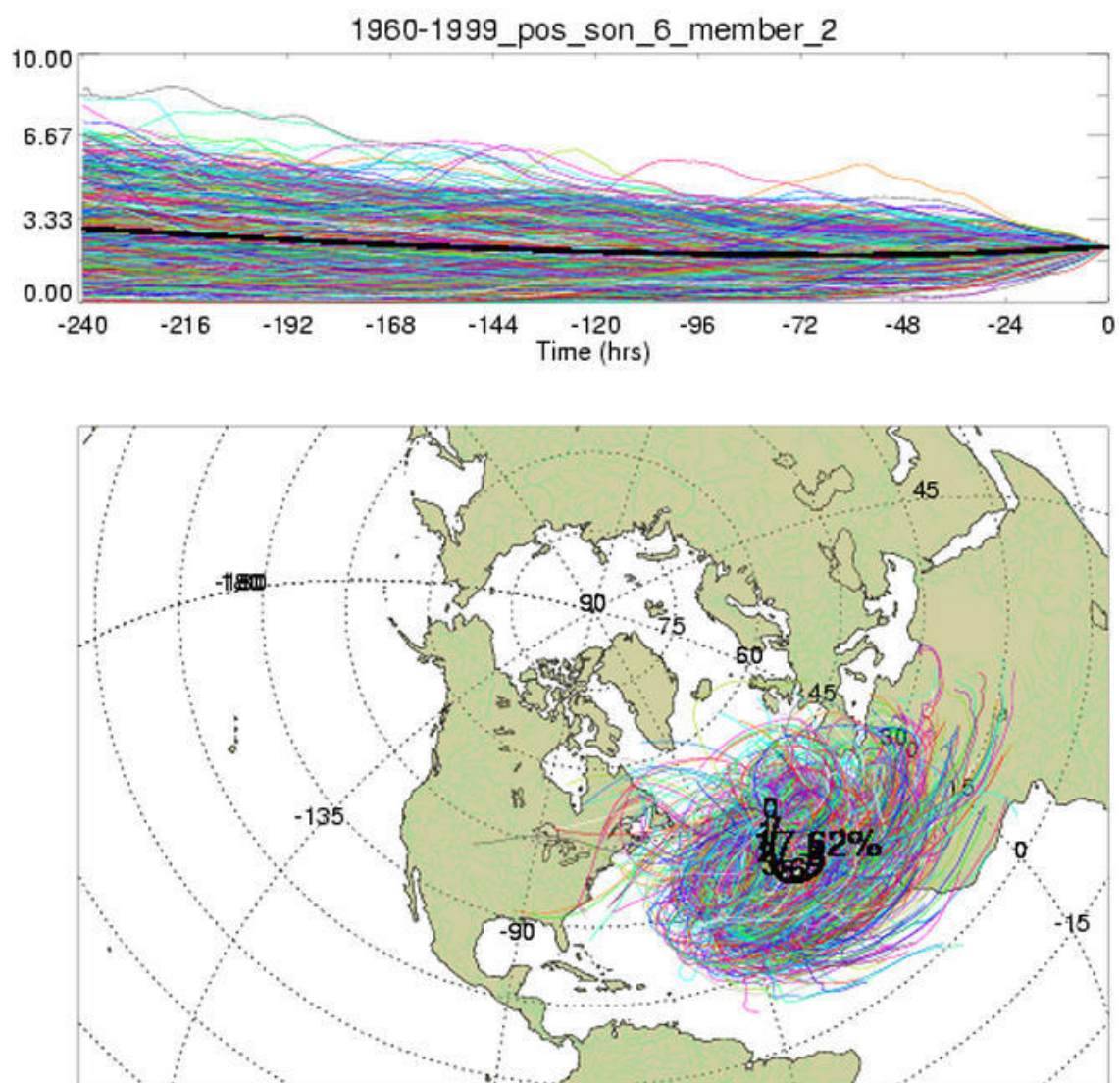


Figure E.27 Membership plot for falls with a positive NAOI value, 1960-1999.

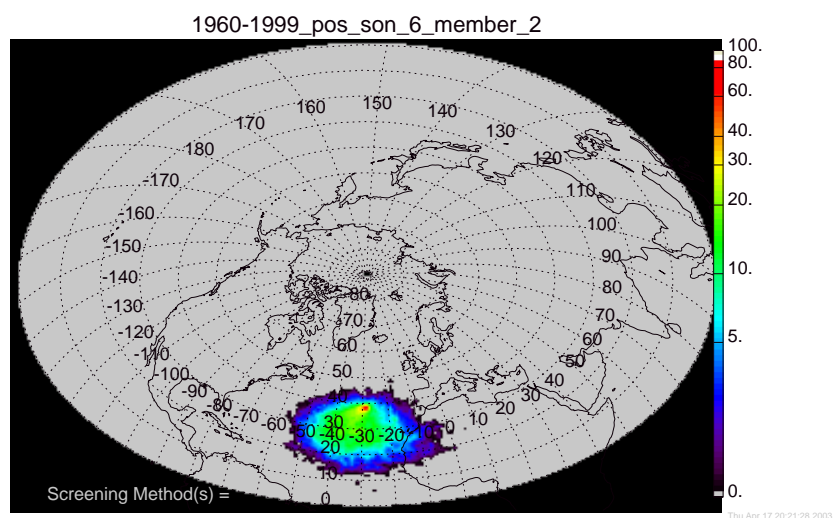


Figure E.28 Standard density membership plot for falls with a positive NAOI value, 1960-1999.

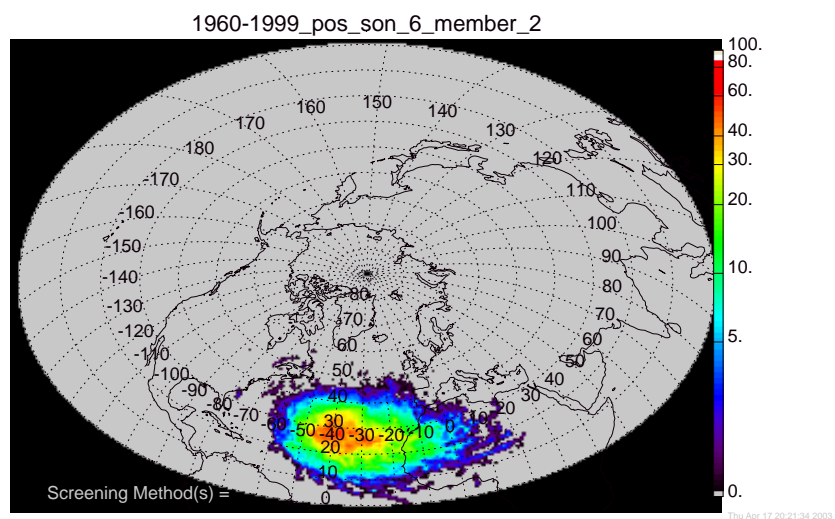


Figure E.29 Geometrically corrected density membership plot for falls with a positive NAOI value, 1960-1999.

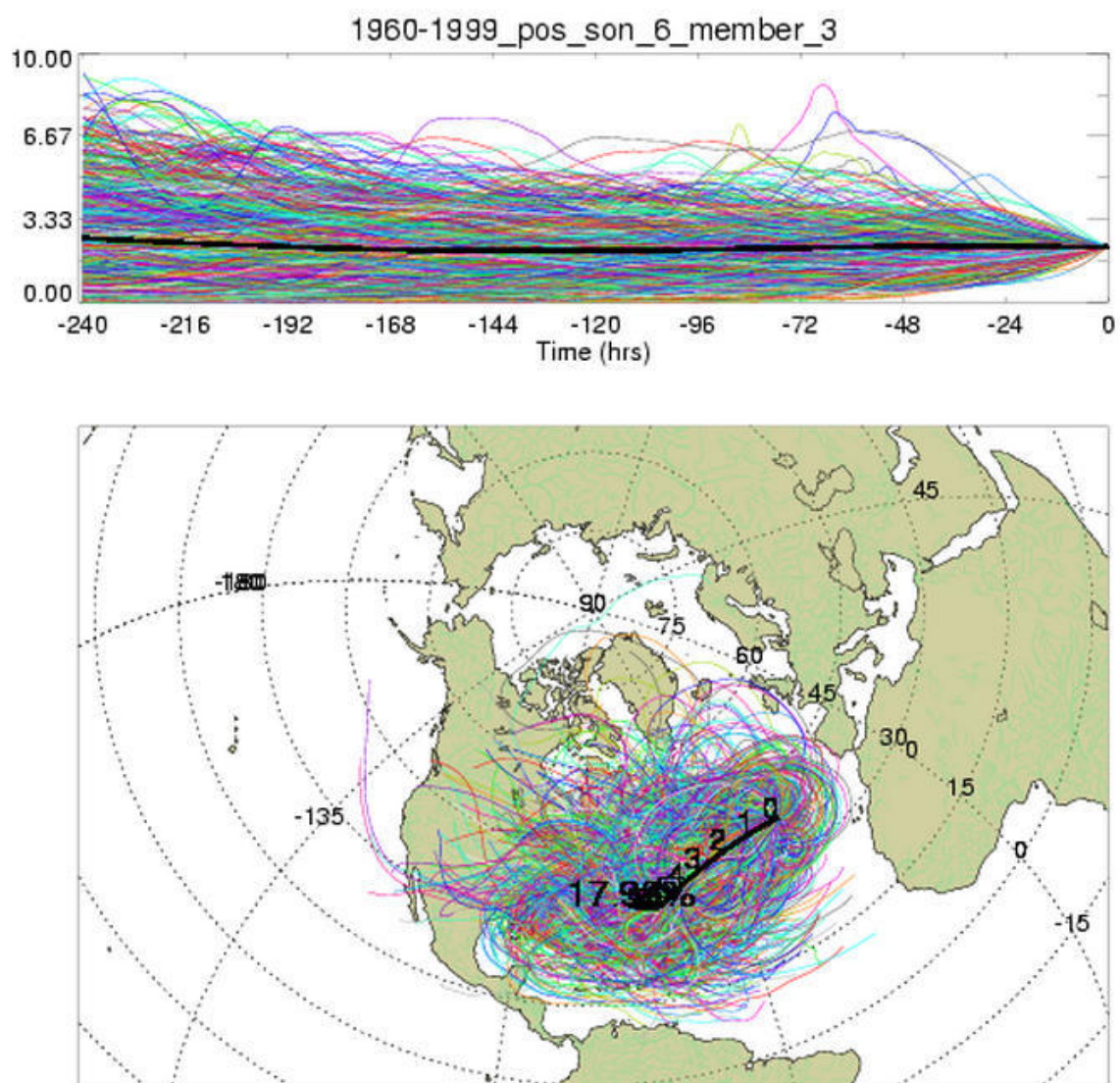


Figure E.30 Membership plot for falls with a positive NAOI value, 1960-1999.

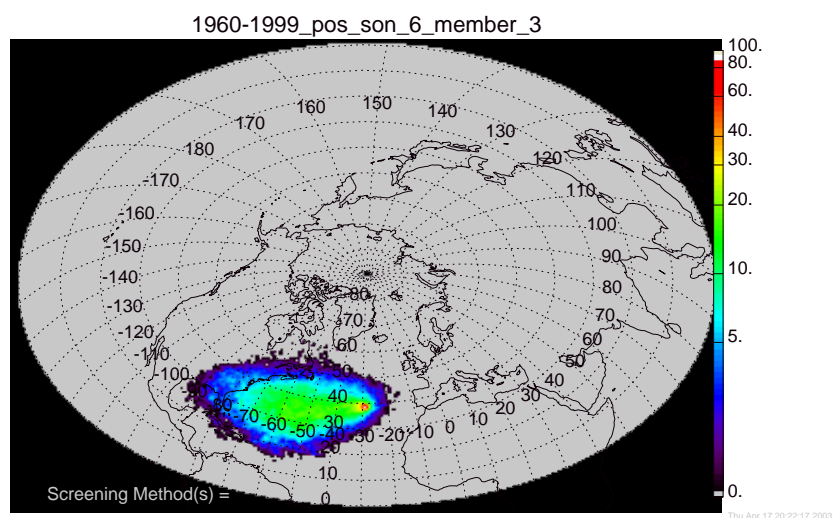


Figure E.31 Standard density membership plot for falls with a positive NAOI value, 1960-1999.

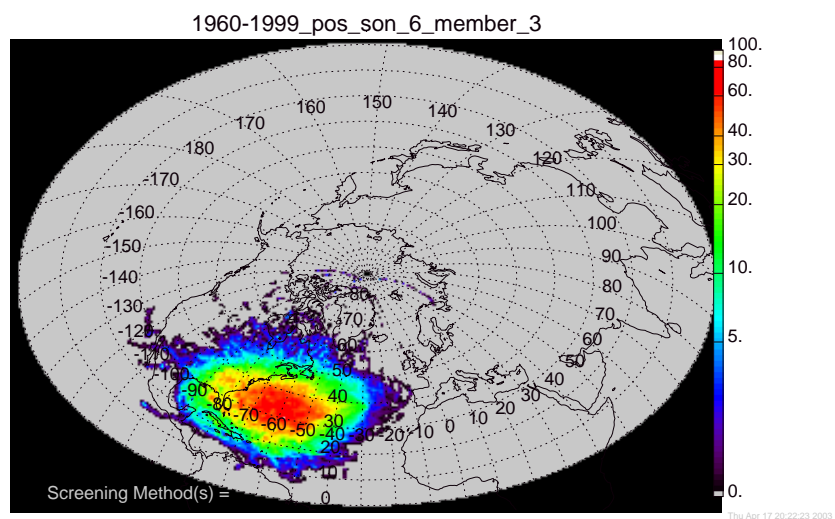


Figure E.32 Geometrically corrected density membership plot for falls with a positive NAOI value, 1960-1999.

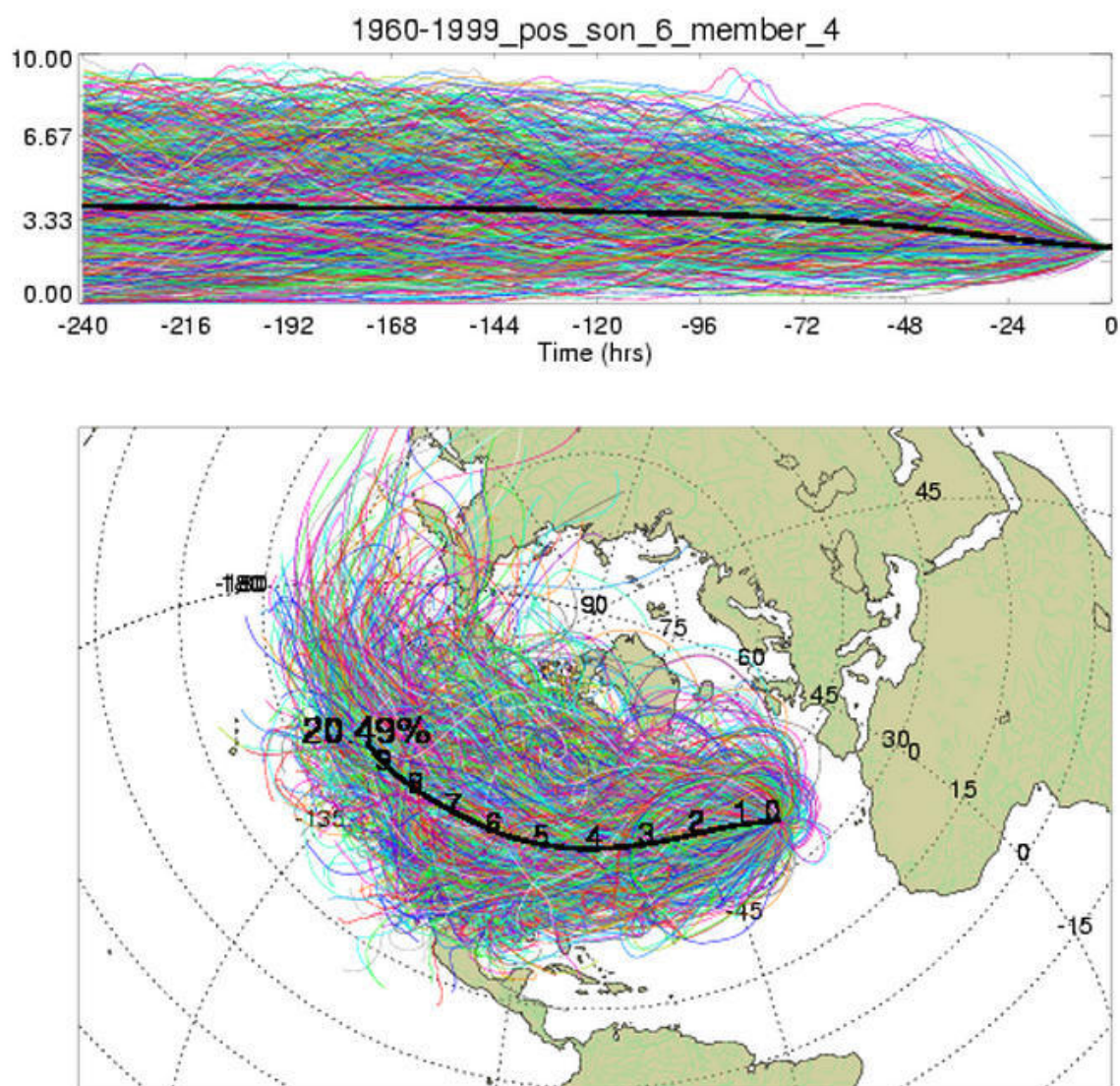


Figure E.33 Membership plot for falls with a positive NAOI value, 1960-1999.

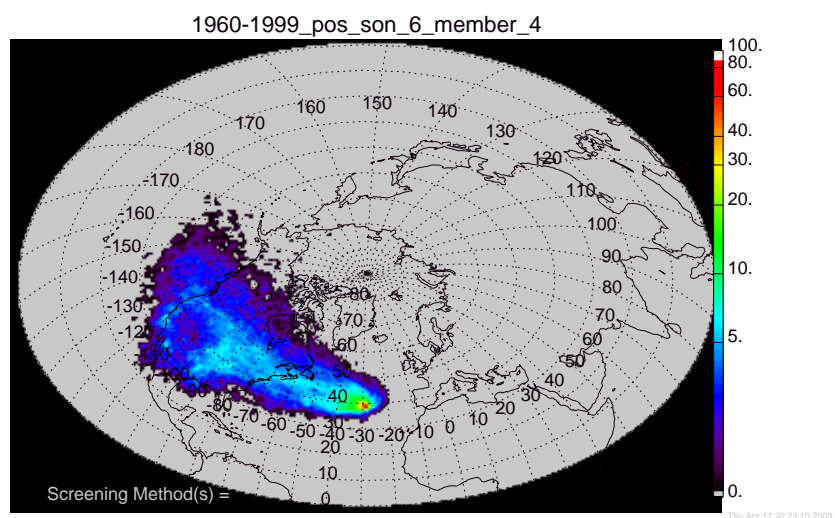


Figure E.34 Standard density membership plot for falls with a positive NAOI value, 1960-1999.

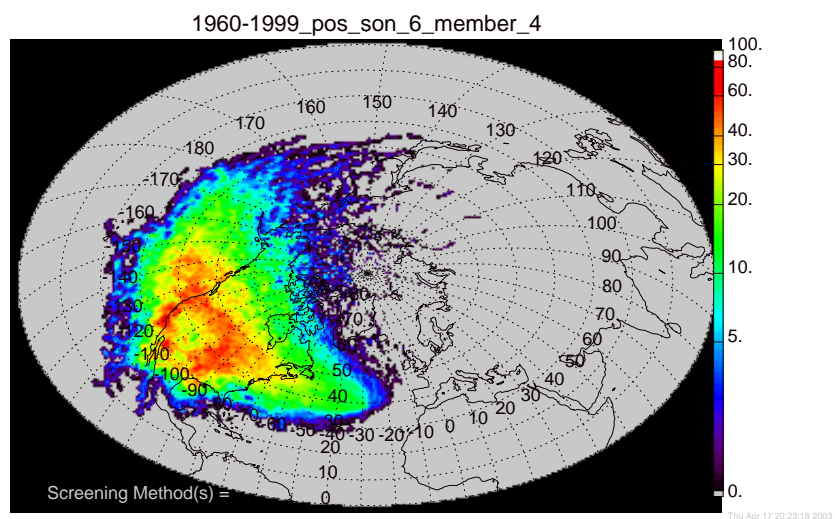


Figure E.35 Geometrically corrected density membership plot for falls with a positive NAOI value, 1960-1999.

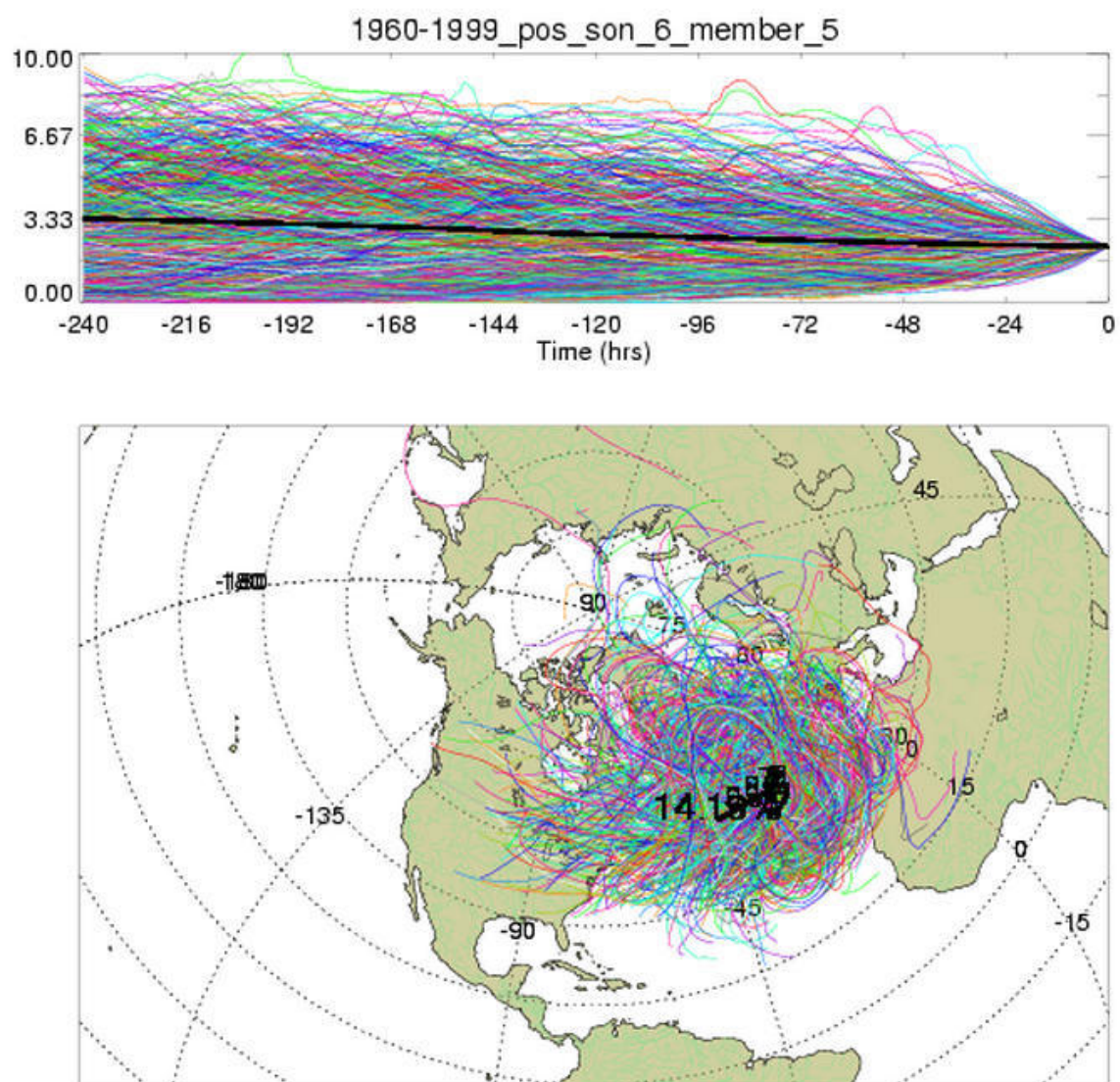


Figure E.36 Membership plot for falls with a positive NAOI value, 1960-1999.

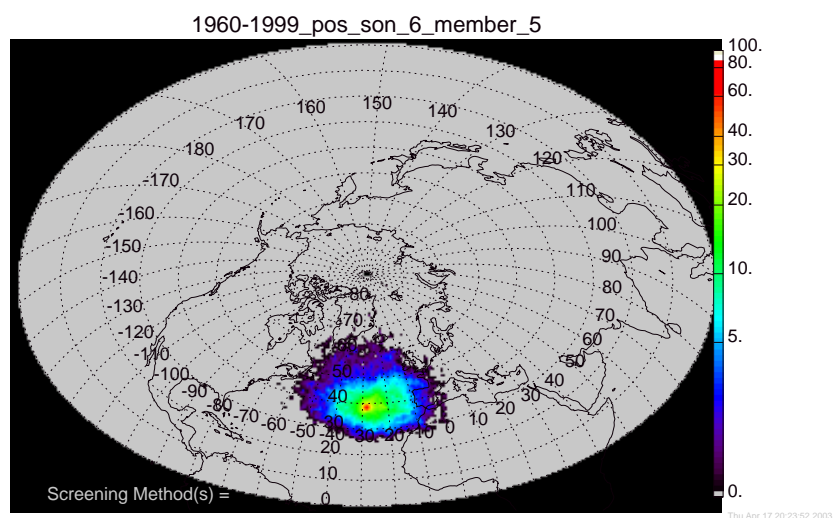


Figure E.37 Standard density membership plot for falls with a positive NAOI value, 1960-1999.

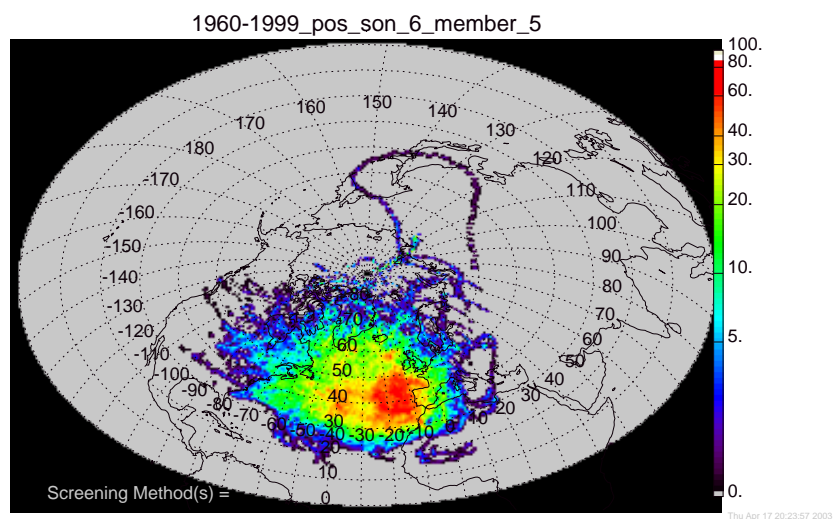


Figure E.38 Geometrically corrected density membership plot for falls with a positive NAOI value, 1960-1999.

E.3 Membership plots for the negative NAO falls

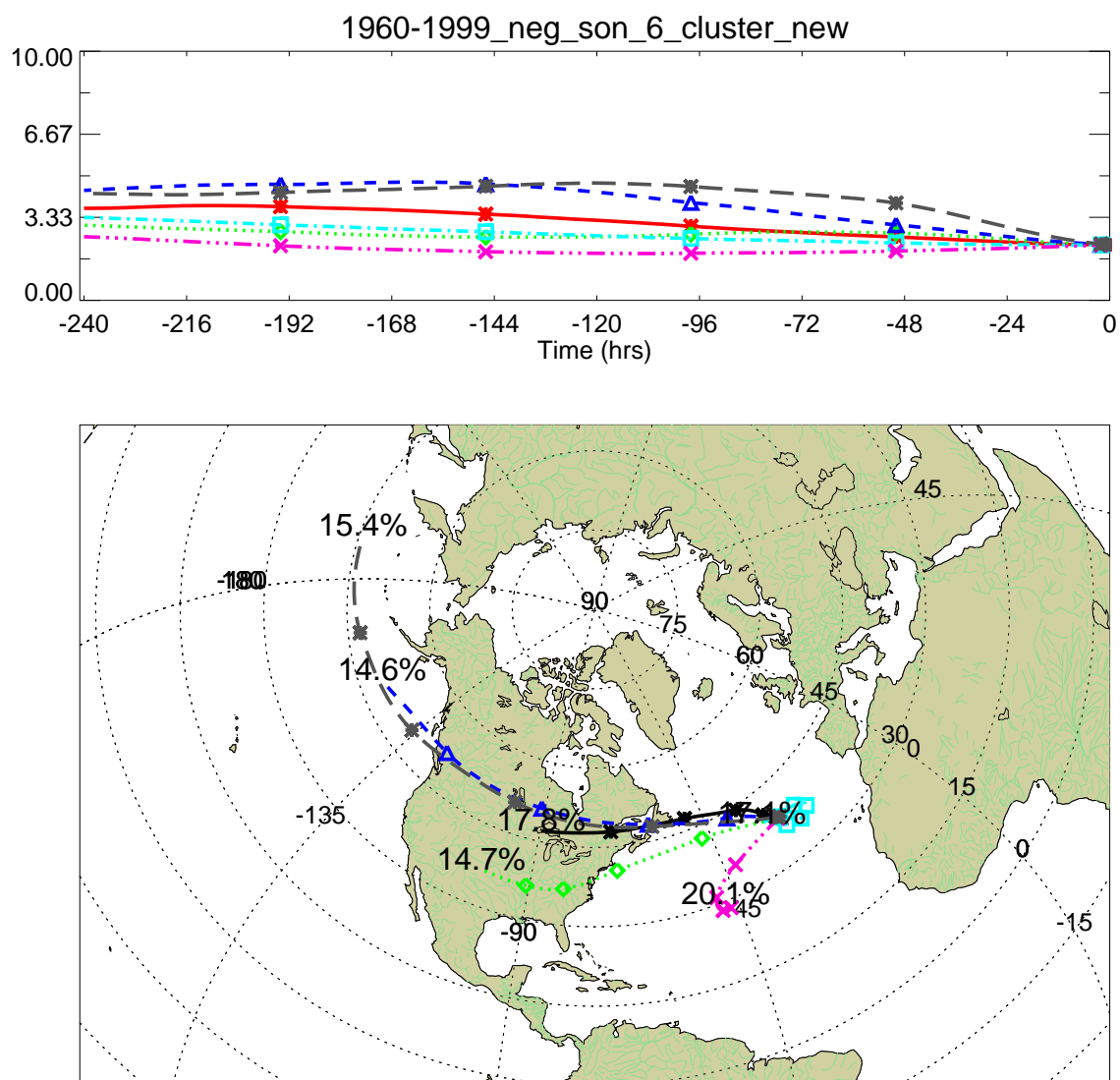


Figure E.39 Cluster plot for the negative NAO for falls, 1960-1999.

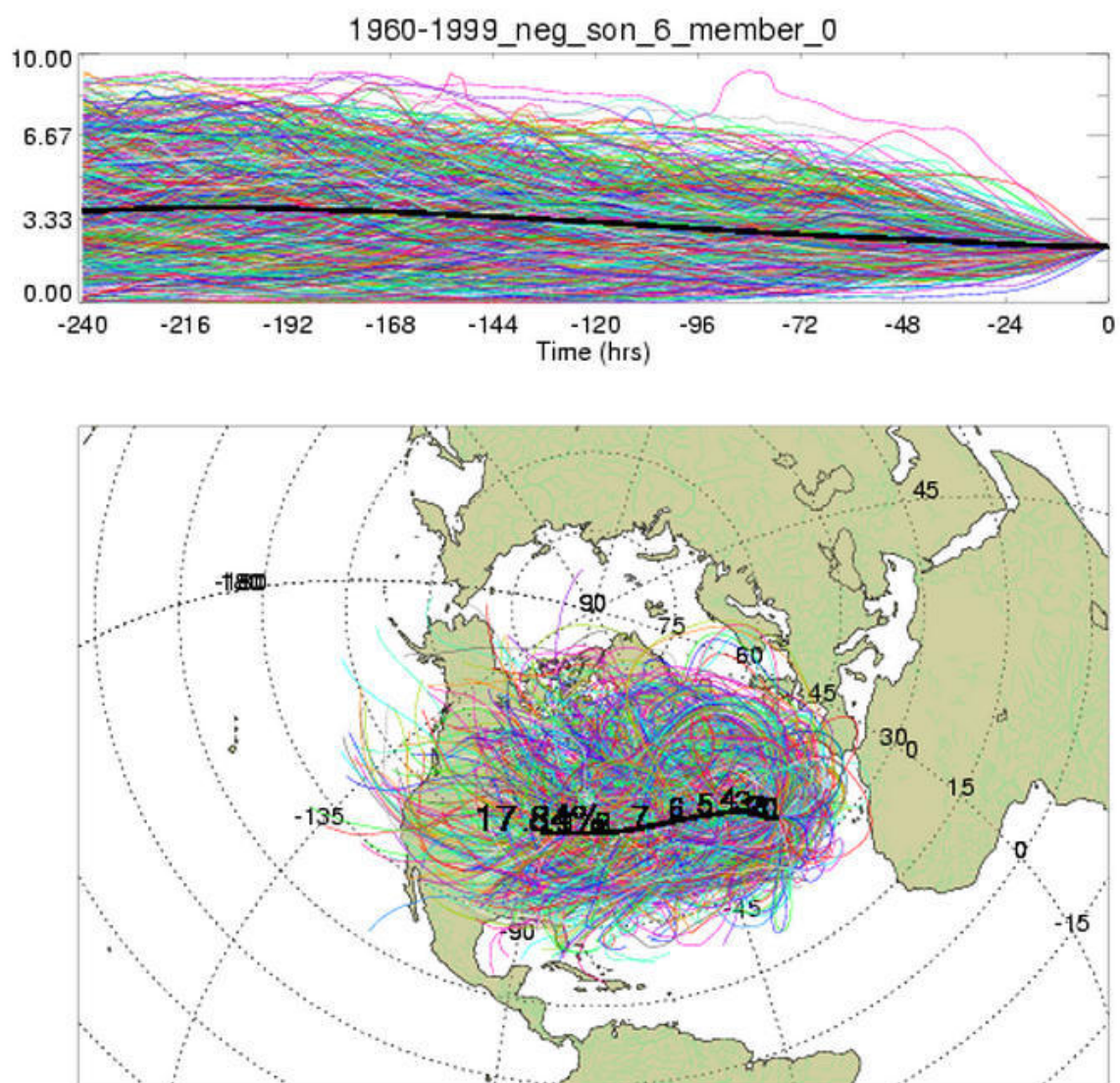


Figure E.40 Membership plot for falls with a negative NAOI value, 1960-1999.

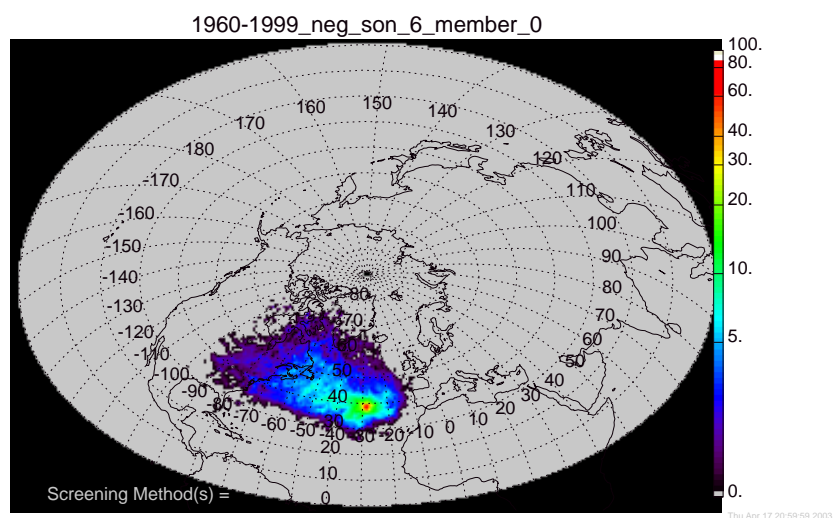


Figure E.41 Standard density membership plot for falls with a negative NAOI value, 1960-1999.

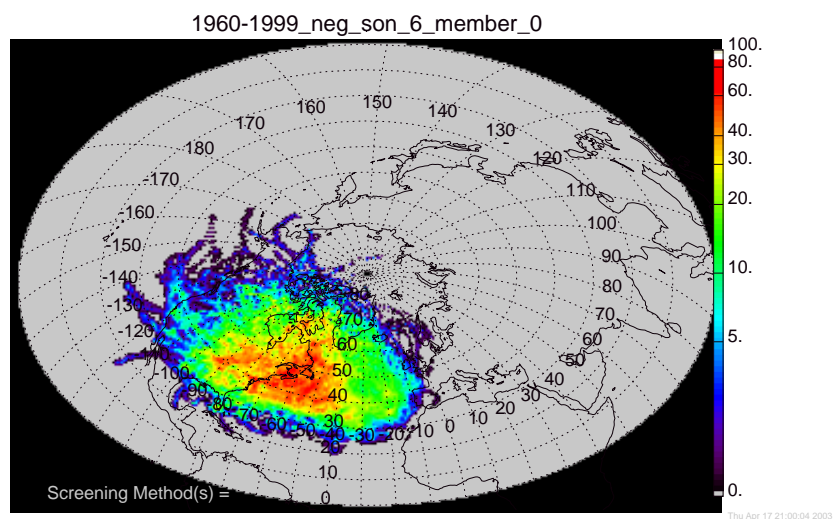


Figure E.42 Geometrically corrected density membership plot for falls with a negative NAOI value, 1960-1999.

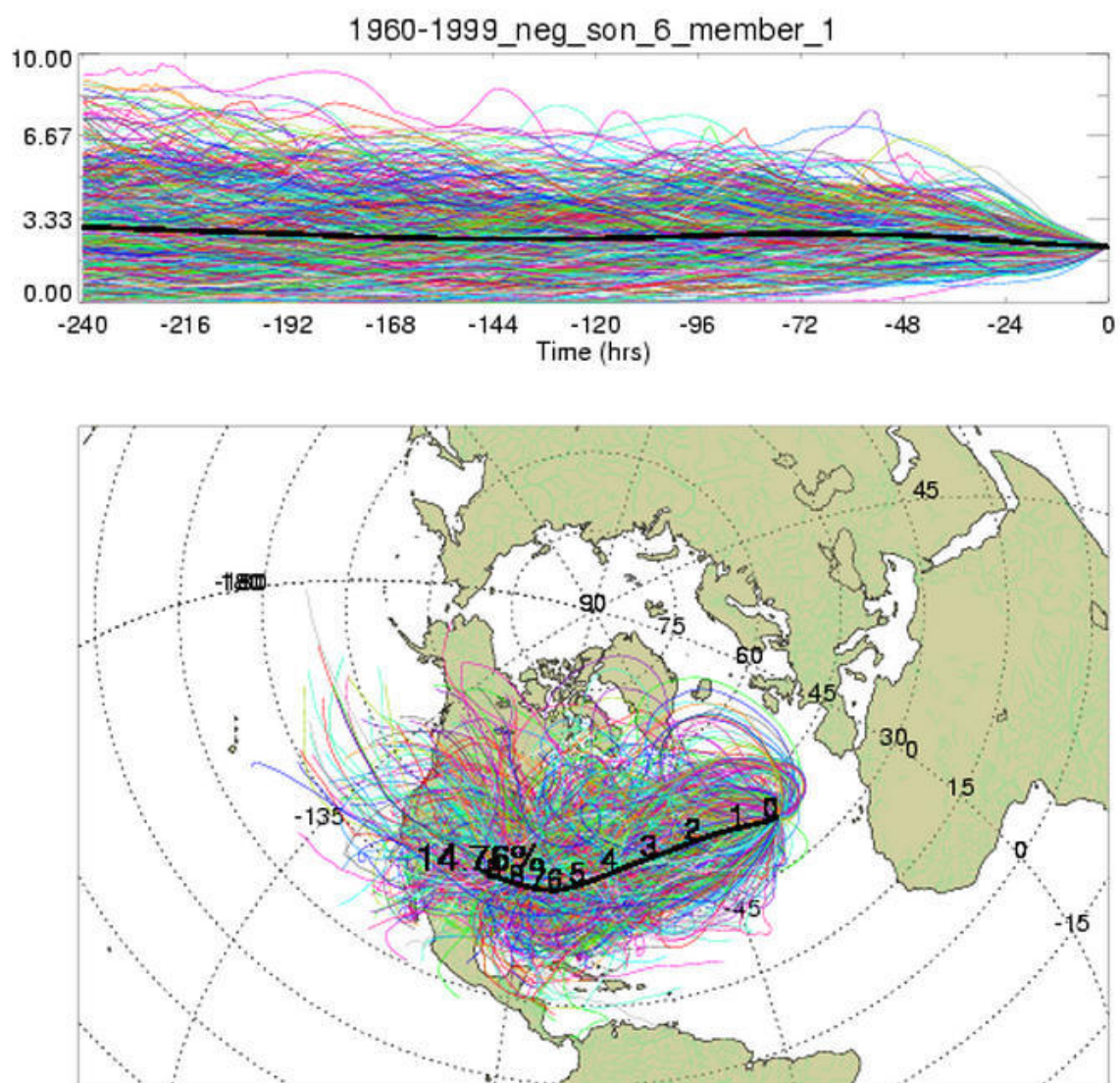


Figure E.43 Membership plot for falls with a negative NAOI value, 1960-1999.

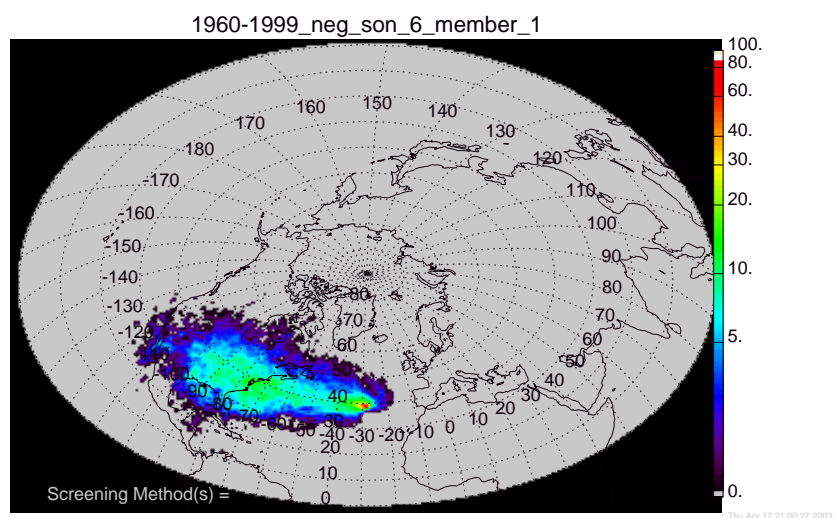


Figure E.44 Standard density membership plot for falls with a negative NAOI value, 1960-1999.

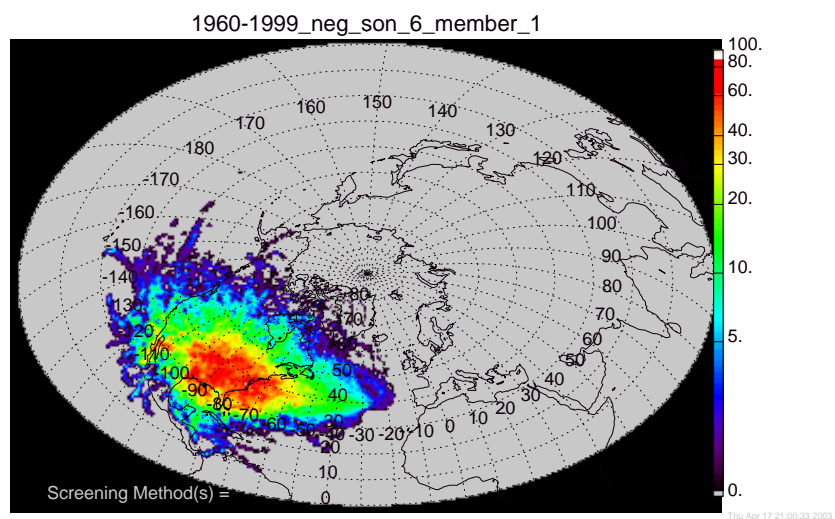


Figure E.45 Geometrically corrected density membership plot for falls with a negative NAOI value, 1960-1999.

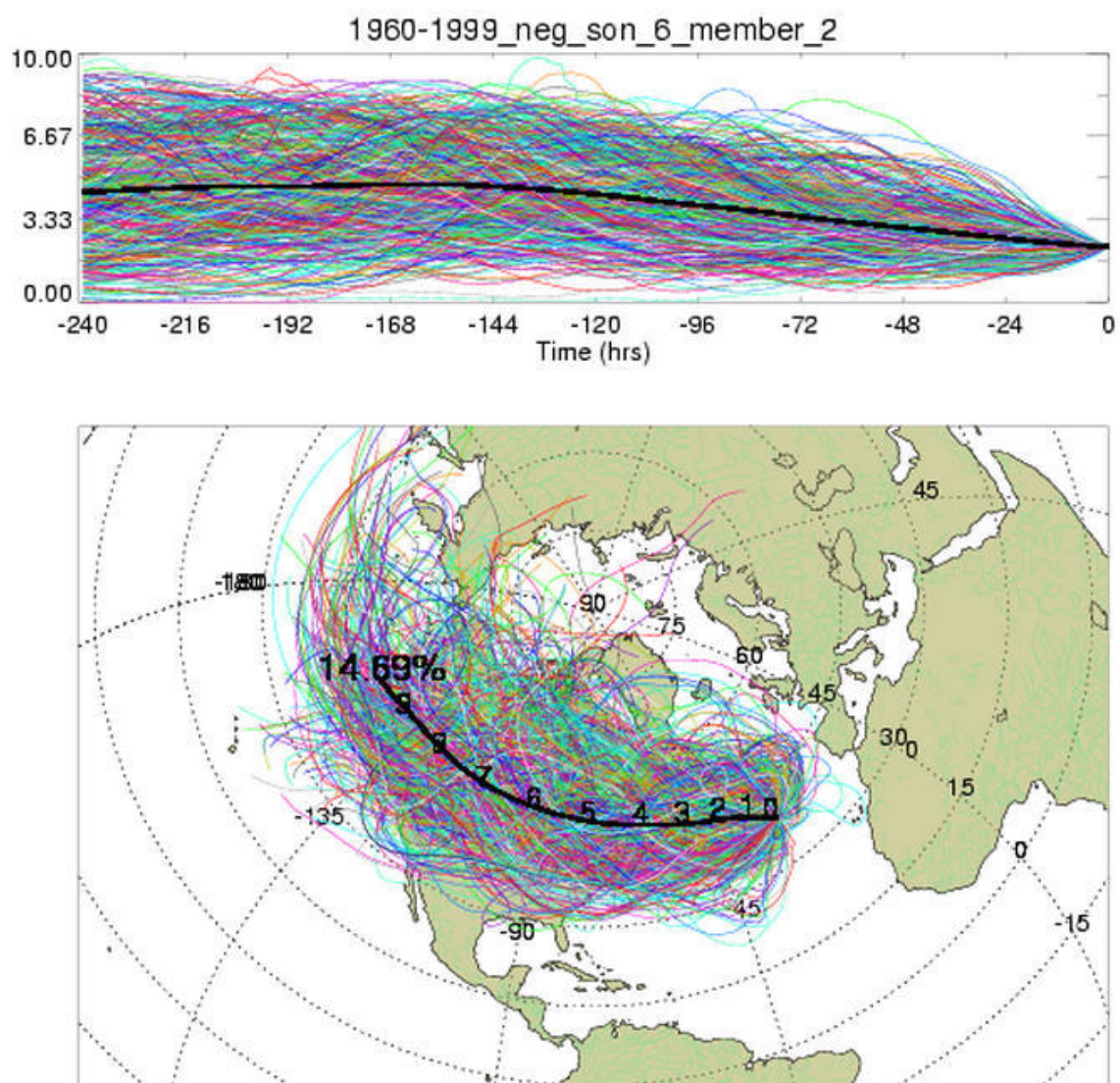


Figure E.46 Membership plot for falls with a negative NAOI value, 1960-1999.

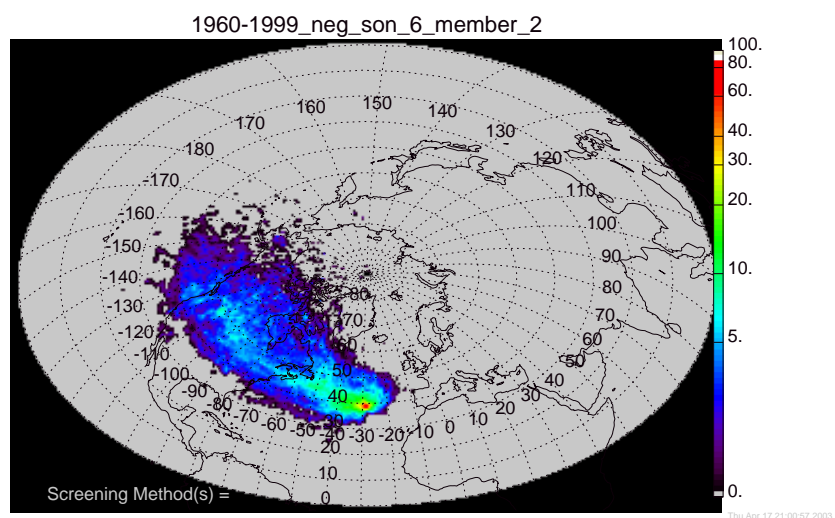


Figure E.47 Standard density membership plot for falls with a negative NAOI value, 1960-1999.

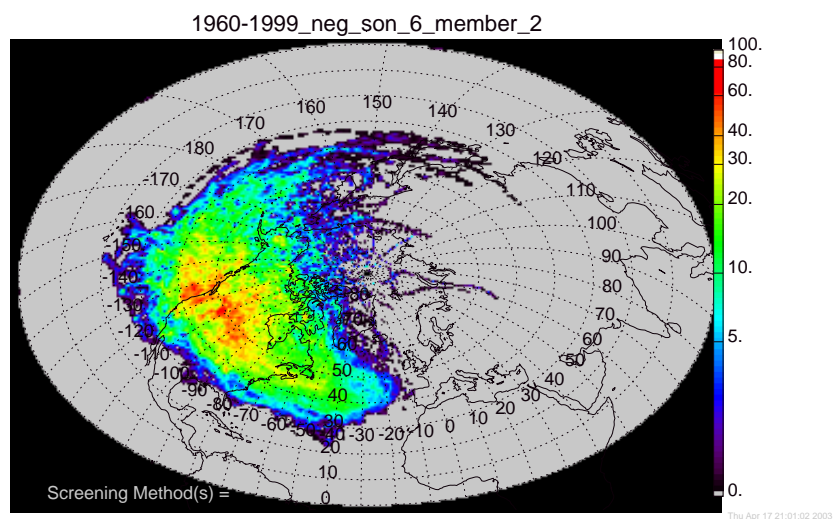


Figure E.48 Geometrically corrected density membership plot for falls with a negative NAOI value, 1960-1999.

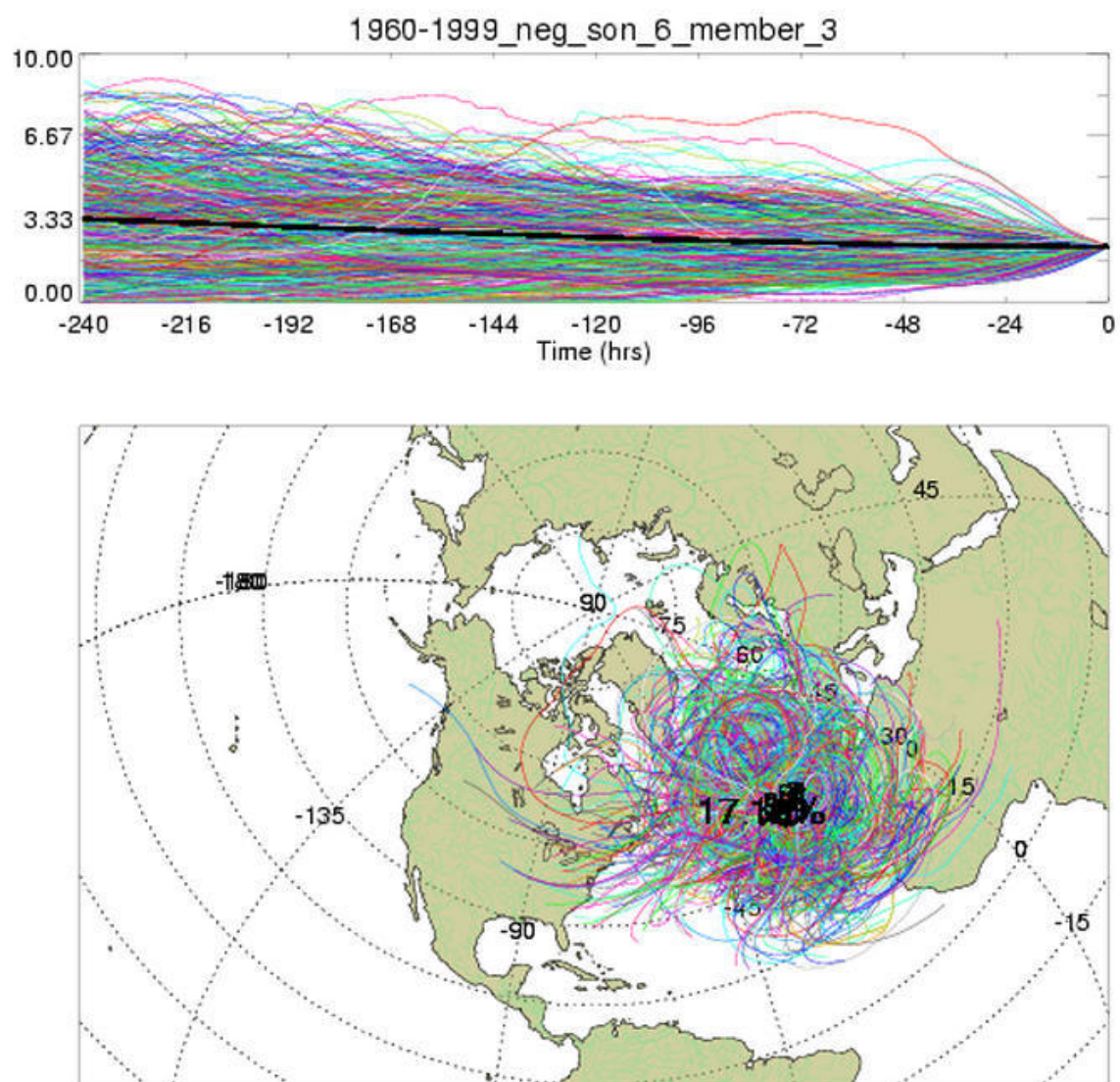


Figure E.49 Membership plot for falls with a negative NAOI value, 1960-1999.

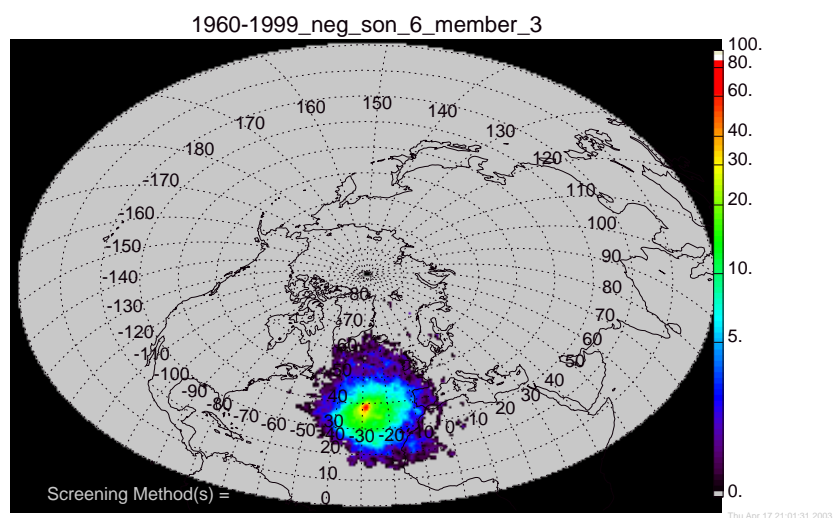


Figure E.50 Standard density membership plot for falls with a negative NAOI value, 1960-1999.

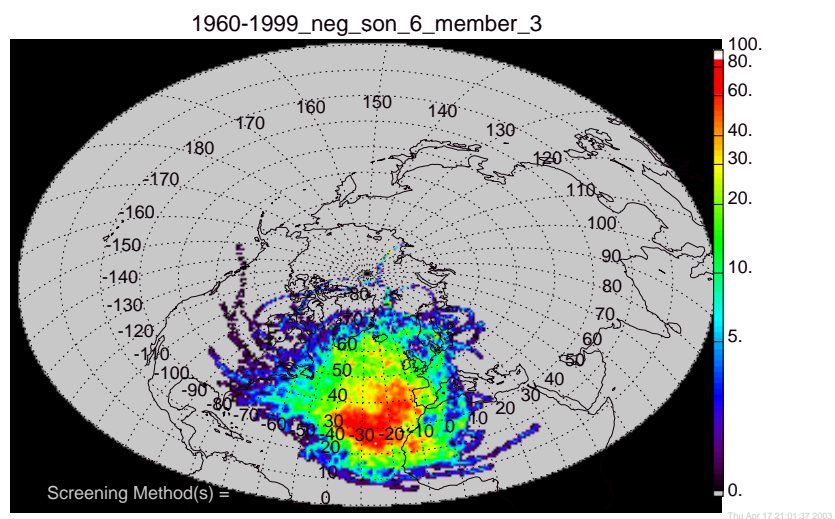


Figure E.51 Geometrically corrected density membership plot for falls with a negative NAOI value, 1960-1999.

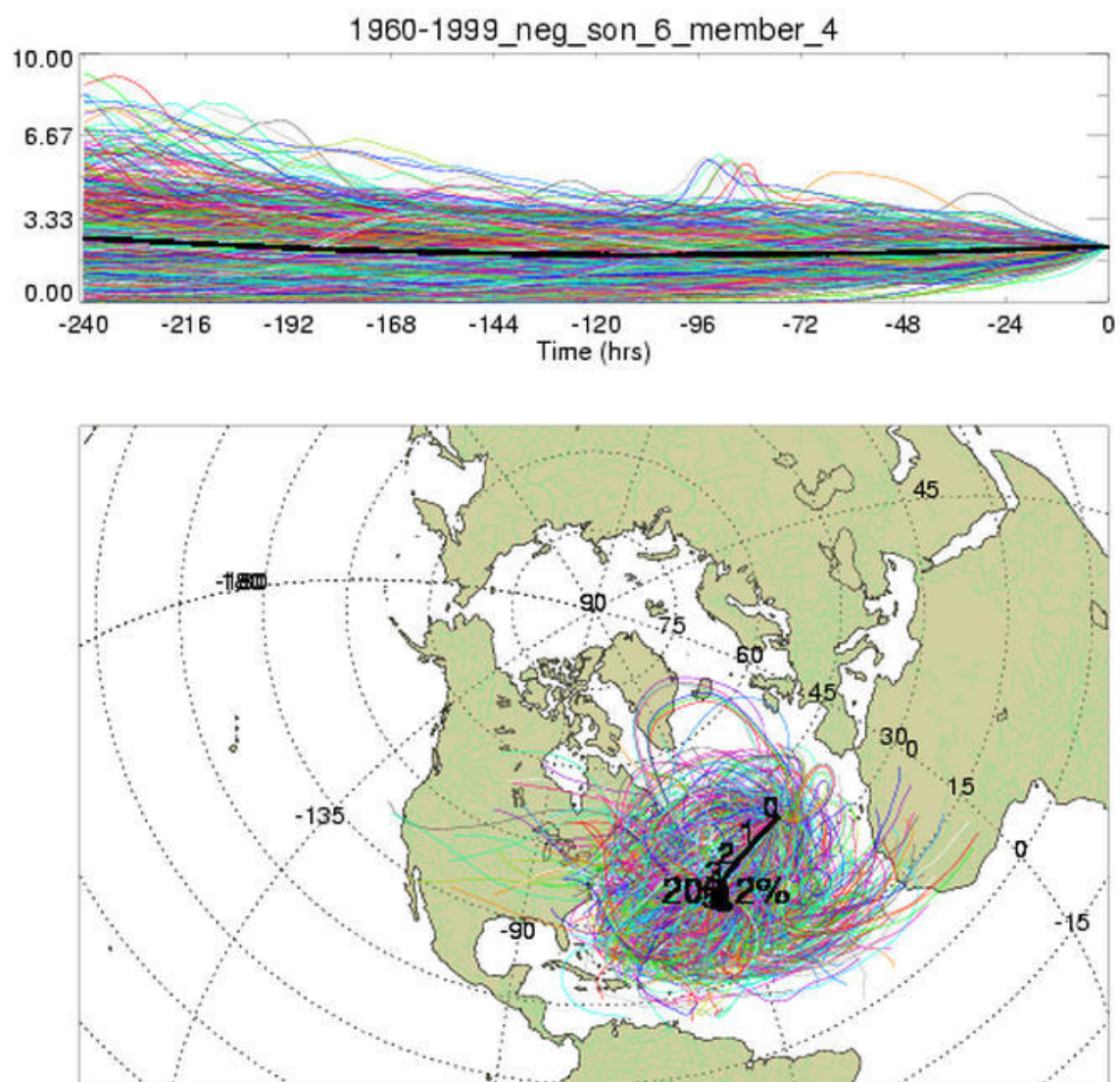


Figure E.52 Membership plot for falls with a negative NAOI value, 1960-1999.

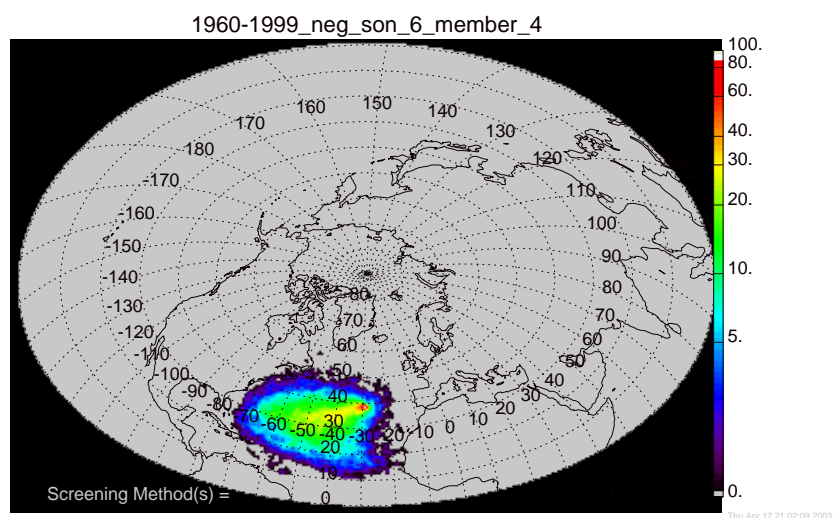


Figure E.53 Standard density membership plot for falls with a negative NAOI value, 1960-1999.

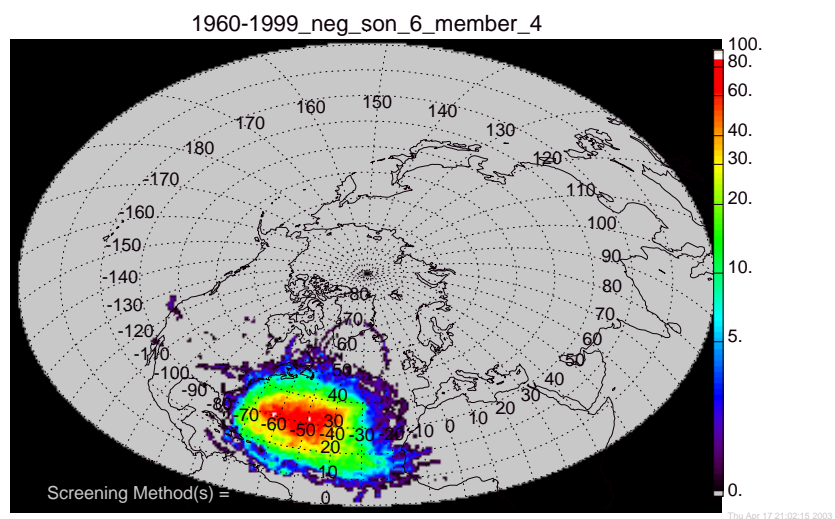


Figure E.54 Geometrically corrected density membership plot for falls with a negative NAOI value, 1960-1999.

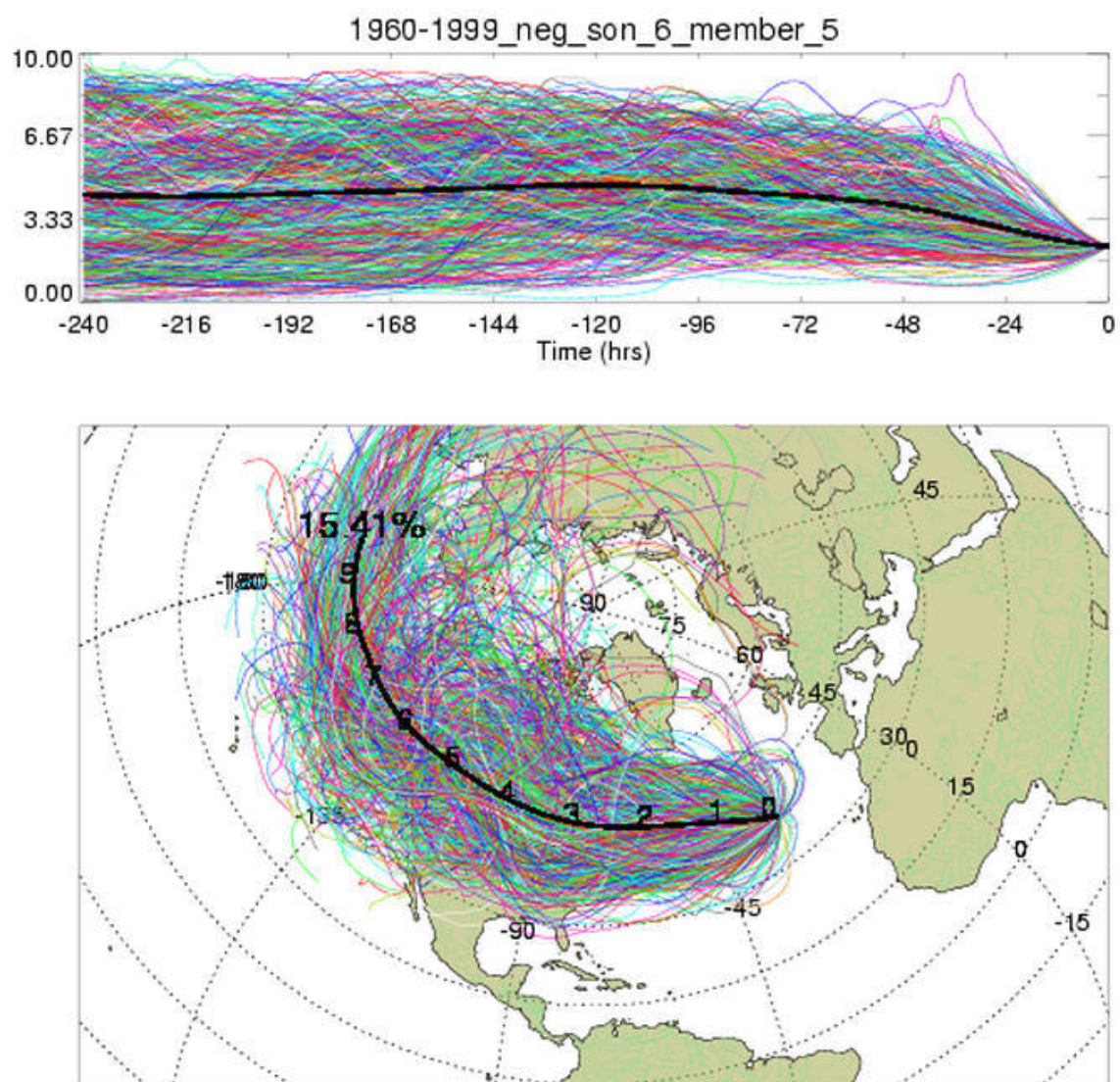


Figure E.55 Membership plot for falls with a negative NAOI value, 1960-1999.

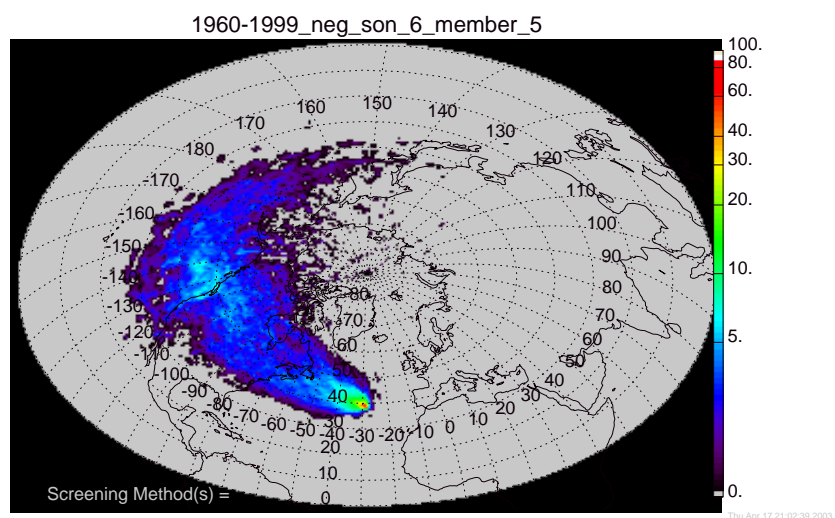


Figure E.56 Standard density membership plot for falls with a negative NAOI value, 1960-1999.

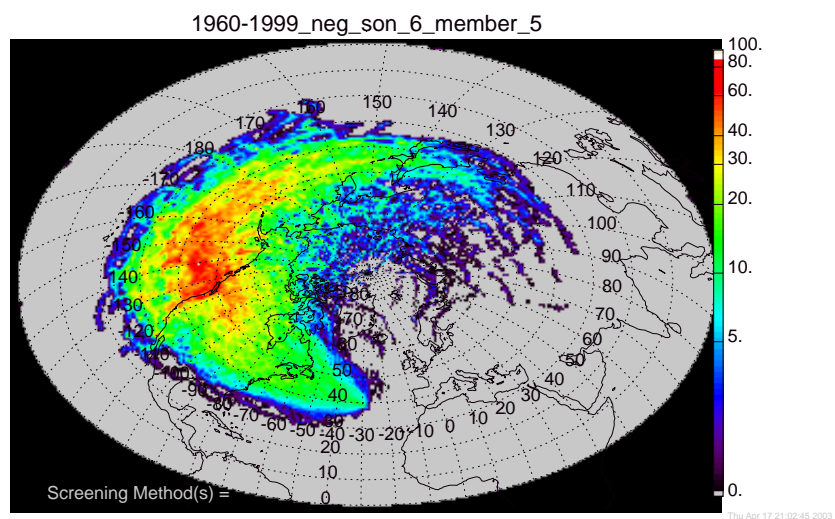


Figure E.57 Geometrically corrected density membership plot for falls with a negative NAOI value, 1960-1999.