Supplement of Atmos. Chem. Phys., 18, 6543–6566, 2018
https://doi.org/10.5194/acp-18-6543-2018-supplement

Atmospheric Chemistry and Physics
Open Access
EGU

Supplement of

The use of hierarchical clustering for the design of optimized monitoring networks

Joana Soares et al.

Correspondence to: Joana Soares (joana.soares@canada.ca)

Supplement 1

Table S1. Continuous NO_2 monitors included for the continuous network optimization analysis. Airshed acronyms: for the Airsheds West Central Airshed Society (WCAS), Wood Buffalo Environmental Association (WBEA), Fort Air Partnership (FAP), Alberta Capital Airshed Alliance (ACAA), Calgary Regional Airshed Zone (CRAZ), Peace Airshed Zone Association (PAZA), Palliser Airshed Society (PAS), Parkland Airshed Management Zone (PAMZ) and Lakeland Industrial Community Association (LICA).

Name	ID	latitude	longitude	Aished
Edmonton Central	1028	53.5444	-113.4988	ACAA
Fort McKay	1032	57.1894	-111.6406	WBEA
Edmonton South	1036	53.5001	-113.5260	ACAA
Calgary Northwest	1039	51.0792	-114.1418	CRAZ
Lethbridge	1049	49.7162	-112.8005	AEP
Violet Grove	1052	53.1423	-115.1381	WCAS
Tomahawk	1053	53.3724	-114.7686	WCAS
Carrot Creek	1054	53.6211	-115.8692	WCAS
Steeper	1055	53.1324	-117.0912	WCAS
Genesee	1057	53.3016	-114.2211	WCAS
Meadows	1058	53.5300	-114.6361	WCAS
Power	1059	53.6330	-114.4199	WCAS
Edson	1062	53.5936	-116.3928	WCAS
Athabasca Valley	1064	56.7336	-111.3904	WBEA
Patricia McInnes	1070	56.7514	-111.4767	WBEA
Millennium Mine	1075	56.9693	-111.4007	WBEA
Fort McKay South	1076	57.1492	-111.6423	WBEA
Caroline	1092	51.9472	-114.6976	PAMZ
Red Deer-Riverside	1142	52.2986	-113.7934	PAMZ
Redwater Industrial	1156	53.8437	-113.0992	FAP
Elk Island	1157	53.6824	-112.8681	FAP
Ross Creek	1159	53.7162	-113.1999	FAP
Range Road 220	1161	53.7524	-113.1259	FAP
Lamont County	1162	53.7604	-112.8802	FAP
Grande Prairie	1165	55.1767	-118.8075	PAZA
Beaverlodge	1168	55.1963	-119.3968	PAZA
Crescent Heights	1172	50.0490	-110.6813	PAS
Calgary Central 2	1221	51.0469	-114.0747	CRAZ
Anzac	1225	56.4493	-111.0372	WBEA
CNRL Horizon	1226	57.3037	-111.7395	WBEA
Wagner2	1241	53.4294	-114.3806	WCAS
Shell Muskeg River	1244	57.2491	-111.5087	WBEA
Maskwa	1248	54.6051	-110.4527	LICA
ST.LINA	1250	54.2164	-111.5023	LICA

Bruderheim	2000	53.8001	-112.9278	FAP
Fort Saskatchewan-92 St and 96 Ave	2001	53.6988	-113.2232	FAP
Woodcroft	2002	53.5644	-113.5626	ACAA

Table S2. Continuous SO_2 monitors included for the continuous network optimization analysis. Airshed acronyms as described in Table S1.

Name	ID	latitude	longitude	Aished
Edmonton East	1029	53.5482	-113.3681	ACAA
Fort McKay	1032	57.1894	-111.6406	WBEA
Lethbridge	1049	49.7162	-112.8005	AEP
Violet Grove	1052	53.1423	-115.1381	WCAS
Tomahawk	1053	53.3724	-114.7686	WCAS
Carrot Creek	1054	53.6211	-115.8692	WCAS
Steeper	1055	53.1324	-117.0912	WCAS
Genesee	1057	53.3016	-114.2211	WCAS
Meadows	1058	53.5300	-114.6361	WCAS
Power	1059	53.6330	-114.4199	WCAS
Edson	1062	53.5936	-116.3928	WCAS
Breton	1063	53.0903	-114.4606	WCAS
Athabasca Valley	1064	56.7336	-111.3904	WBEA
Mildred Lake	1066	57.0500	-111.5641	WBEA
Mannix	1069	56.9680	-111.4821	WBEA
Patricia McInnes	1070	56.7514	-111.4767	WBEA
Lower Camp	1074	57.0325	-111.5064	WBEA
Millennium Mine	1075	56.9693	-111.4007	WBEA
Fort McKay South	1076	57.1492	-111.6423	WBEA
Caroline	1092	51.9472	-114.6976	PAMZ
Red Deer-Riverside	1142	52.2986	-113.7934	PAMZ
Redwater Industrial	1156	53.8437	-113.0992	FAP
Elk Island	1157	53.6824	-112.8681	FAP
Ross Creek	1159	53.7162	-113.1999	FAP
Range Road 220	1161	53.7524	-113.1259	FAP
Lamont County	1162	53.7604	-112.8802	FAP
Grande Prairie	1165	55.1767	-118.8075	PAZA
Evergreen Park	1166	55.1175	-118.7647	PAZA
Smoky Heights	1167	55.4028	-118.2814	PAZA
Beaverlodge	1168	55.1963	-119.3968	PAZA
Valleyview	1170	54.9405	-117.2153	PAZA
Crescent Heights	1172	50.0490	-110.6813	PAS
Anzac	1225	56.4493	-111.0372	WBEA
CNRL Horizon	1226	57.3037	-111.7395	WBEA
Wagner2	1241	53.4294	-114.3806	WCAS

Shell Muskeg River	1244	57.2491	-111.5087	WBEA
Maskwa	1248	54.6051	-110.4527	LICA
ST.LINA	1250	54.2164	-111.5023	N_LICA
Bruderheim	2000	53.8001	-112.9278	FAP
Fort Saskatchewan-92 St and 96 Ave	2001	53.6988	-113.2232	FAP
Woodcroft	2002	53.5644	-113.5626	ACAA

Table S3. Passive and continuous NO_2 monitors included for the passive network optimization analysis, ID ending in a "(C)" is a continuous monitor and with a "(P)" a passive monitor. Airshed acronyms as described in Table S1.

Name	ID	Latitude	longitude	Airshed
Edmonton Central	1028C	53.5444	-113.4988	ACAA
Edmonton East	1029C	53.5482	-113.3681	ACAA
Edmonton South	1036C	53.5001	-113.5260	ACAA
Lethbridge	1049C	49.7162	-112.8005	AEP
Calgary Northwest	1039C	51.0792	-114.1418	CRAZ
Redwater Industrial	1156C	53.8437	-113.0992	FAP
Elk Island	1157C	53.6824	-112.8681	FAP
Ross Creek	1159C	53.7162	-113.1999	FAP
Range Road 220	1161C	53.7524	-113.1259	FAP
Lamont County	1162C	53.7604	-112.8802	FAP
Fort Saskatchewan-92 St and 96 Ave	2001C	53.6988	-113.2232	FAP
Redwater Industrial	1156C	53.8437	-113.0992	FAP
Therien	1176P	54.3108	-111.2261	LICA
Flat Lake	1177P	54.0726	-111.2051	LICA
Lake Eliza	1178P	53.8242	-111.1661	LICA
Muriel-Kehiwin	1181P	54.0934	-110.7444	LICA
Dupre	1182P	54.3346	-110.7796	LICA
La Corey	1183P	54.4997	-110.8179	LICA
Primrose	1186P	54.7585	-110.4522	LICA
Maskwa	1248C,1187P	54.6052	-110.4526	LICA
Frog Lake	1189P	53.8907	-110.3842	LICA
Clear Range	1190P	53.5565	-110.1542	LICA
Fishing Lake	1191P	53.9030	-110.0762	LICA
Beaverdam	1192P	54.1693	-110.2328	LICA
Cold Lake South	1174C, 1193P, 1227P	54.4137	-110.2328	LICA
Fort George	1195P	53.8783	-110.7481	LICA
Town of Bonnyville	1199P	54.2709	-110.7435	LICA
St. Lina	1250C,1252P	54.2160	-111.5033	LICA
PAS 18	9936P	49.9949	-110.6292	PAMZ
Bow Summit	9938P	51.7083	-116.4792	PAMZ
Parker Ridge	9939P	52.1173	-115.4069	PAMZ
Bighorn	9940P	52.4300	-116.2000	PAMZ

Sunchild	9941P	52.6471	-115.4269	PAMZ
Baseline Mountain	9942P	52.1704	-115.4183	PAMZ
Limestone Mountain	9943P	51.8755	-115.4528	PAMZ
Fallen Timber	9945P	51.6126	-115.0227	PAMZ
Bearberry	9946P	51.8820	-115.0010	PAMZ
Ricinus	9947P	52.1508	-114.9777	PAMZ
Twin Lakes	9948P	52.4054	-115.0237	PAMZ
Rimbey	9949P	52.6465	-114.3729	PAMZ
Leslieville	9950P	52.4101	-114.5769	PAMZ
Raven River	9951P	52.1421	-114.5799	PAMZ
Sundre	9952P	51.8825	-114.5697	PAMZ
South Elkton	9953P	51.6205	-114.5687	PAMZ
Bottrel	9954P	51.4093	-114.3944	PAMZ
Crossfield-Carstairs	9955P	51.4819	-113.9986	PAMZ
Netook-Olds	9956P	51.8751	-114.1437	PAMZ
Sylvan Lake	9958P	52.4068	-111.1534	PAMZ
Range Rd	9959P	52.5627	-113.6261	PAMZ
Mayton	9960P	51.8863	-113.6913	PAMZ
Sunnyslope	9961P	51.6204	-113.6661	PAMZ
Grainger	9962P	51.4379	-113.3338	PAMZ
Elnora	9963P	52.0339	-113.3379	PAMZ
Alix	9964P	52.4118	-113.3275	PAMZ
Samson	9965P	52.7268	-113.1775	PAMZ
Ferrybank	9966P	52.7627	-113.9942	PAMZ
Pakkwaw	9967P	52.1430	-113.7187	PAMZ
Kersey	9968P	51.3580	-113.8141	PAMZ
Panther River	9970P	51.6075	-115.3893	PAMZ
Markerville	9971P	42.1446	-114.1270	PAMZ
Caroline	1092C,P	51.9469	114.6974	PAMZ
Red Deer Riverside	1142C, P	52.2987	113.8103	PAMZ
Crescent Heights, PAS 19	1172C,P	50.0491	-110.6814	PAS
SM7	9921P	55.6857	-111.8154	PAS
PAS 1	9922P	50.3970	-110.1216	PAS
PAS 2	9923P	50.6951	-110.1299	PAS
PAS 3	9924P	51.2288	-110.5160	PAS
PAS 4	9925P	51.6044	-110.2091	PAS
PAS 6	9926P	51.2114	-111.9586	PAS
PAS 7	9927P	50.7315	-112.4055	PAS
PAS 8	9928P	50.7483	-111.0208	PAS
PAS 11	9929P	50.5897	-111.8508	PAS
PAS 12	9930P	50.4347	-111.7000	PAS
PAS 13	9931P	49.8422	-111.4968	PAS
PAS 14	9932P	49.8432	-111.0945	PAS

PAS 15	9933P	49.5233	-111.4722	PAS
PAS 16	9934P	49.1753	-110.9475	PAS
PAS 17	9935P	49.2208	-110.2081	PAS
Beaverlodge	1168C	55.1963	-119.3968	PAZA
Grande Prairie	1165C	55.1767	-118.8075	PAZA
Fort McKay-Bertha Ganter	1032C,P	57.1894	-111.6406	WBEA
Fort McMurray-Athabasca Valley	1064C	56.7336	-111.3904	WBEA
Fort McMurray-Patricia McInnes	1070C	56.7514	-111.4767	WBEA
Fort Chipewyan	1071C	58.7092	-111.1750	WBEA
Millennium Mine	1075C	56.9693	-111.4007	WBEA
Fort McKay South	1076C	57.1492	-111.6423	WBEA
Anzac	1225C	56.4493	-111.0372	WBEA
CNRL Horizon	1226C	57.3037	-111.7395	WBEA
Shell Muskeg River	1244C	57.2491	-111.5087	WBEA
AH3	9901P	56.6964	-111.1223	WBEA
AH8	9902P	57.1019	-112.0754	WBEA
BM10	9903P	57.3201	-112.3970	WBEA
BM11	9904P	57.6913	-111.9095	WBEA
BM7	9905P	58.0582	-112.2814	WBEA
JP101	9906P	56.5395	-112.2756	WBEA
JP102	9907P	56.9096	-111.5407	WBEA
JP104	9908P	57.1190	-111.4254	WBEA
JP107	9909P	57.8903	-111.4352	WBEA
JP205	9910P	57.8378	-110.4490	WBEA
JP210	9911P	56.2728	-110.4487	WBEA
JP212	9912P	57.0537	-111.4092	WBEA
JP213	9913P	57.0473	-109.7497	WBEA
NE10	9914P	56.6085	-110.1926	WBEA
NE11	9915P	57.2879	-111.2171	WBEA
NE7	9916P	57.1467	-110.8663	WBEA
SM8	9917P	56.2016	-111.1753	WBEA
WF4	9918P	57.1479	-111.9840	WBEA
AH7	9919P	56.8298	-111.7678	WBEA
R2	9920P	56.8298	-111.7678	WBEA
Violet Grove	1052C	53.1423	-115.1381	WCAS
Tomahawk	1053C	53.3724	-114.7686	WCAS
Carrot Creek	1054C	53.6211	-115.8692	WCAS
Steeper	1055C	53.1324	-117.0912	WCAS
Genesee	1057C	53.3016	-114.2211	WCAS
Meadows	1058C	53.5300	-114.6361	WCAS
Power	1059C	53.6330	-114.4199	WCAS
Breton	1063C	53.0903	-114.4606	WCAS
Wagner2	1241C	53.4294	-114.3806	WCAS

Table S4. Passive and continuous SO_2 monitors included for the passive network optimization analysis. ID ending in a "(C)" is a continuous monitor and with a "(P)" a passive monitor. Airshed acronyms are as described in Table S1.

Name	ID	Latitude	longitude	Airshed
Edmonton East	1029C	53.5482	-113.3681	ACAA
Edmonton South	1036C	53.5001	-113.5260	ACAA
Lethbridge	1049C	49.7162	-112.8005	AEP
Redwater Industrial	1156C	53.8437	-113.0992	FAP
Elk Island	1157C	53.6824	-112.8681	FAP
Ross Creek	1159C	53.7162	-113.1999	FAP
Range Road 220	1161C	53.7524	-113.1259	FAP
Lamont County	1162C	53.7604	-112.8802	FAP
Fort Saskatchewan-92 St and 96 Ave	2001C	53.6988	-113.2232	FAP
FAP-01	1258P	113.2489	53.5952	FAP
FAP-02	1259P	113.1019	53.5885	FAP
FAP-03	1260P	53.8667	-112.8270	FAP
FAP-04	1261P	112.7762	54.0988	FAP
FAP-05	1262P	113.1331	54.1523	FAP
FAP-07	1264P	113.4242	53.8338	FAP
FAP-08	1265P	113.4242	53.8338	FAP
FAP-10	1267P	53.7512	-113.1883	FAP
FAP-11	1268P	53.8320	-113.0489	FAP
FAP-12	1269P	53.8658	-113.0254	FAP
FAP-15	1272P	53.8044	-112.9765	FAP
FAP-18	1275P	53.7525	-113.0005	FAP
FAP-20	1277P	53.7594	-112.8767	FAP
FAP-21	1278P	53.7085	-112.9754	FAP
FAP-22	1279P	53.6876	-112.8769	FAP
FAP-23	1280P	53.6633	-112.9265	FAP
FAP-24	1281P	53.8187	-113.0857	FAP
FAP-26	1283P	53.8035	-113.1495	FAP
FAP-27	1284P	53.8813	-113.0004	FAP
FAP-28	1285P	53.9045	-112.9508	FAP
FAP-29	1286P	53.9491	-112.9483	FAP
FAP-30	1287P	53.9343	-113.1001	FAP
FAP-31	1288P	53.8107	-113.1084	FAP
FAP-32	1289P	53.8333	-113.1322	FAP
FAP-33	1290P	53.7451	-113.2482	FAP
FAP-34	1291P	53.7454	-113.4836	FAP
FAP-35	1292P	53.8251	-113.4789	FAP
FAP-37	1294P	53.8631	-113.2236	FAP
FAP-38	1295P	53.9218	-112.6787	FAP

FAP-39	1296P	53.8325	-112.6783	FAP
FAP-40	1297P	53.7452	-112.7029	FAP
FAP-42	1299P	54.0070	-113.0245	FAP
FAP-43	1300P	53.7452	-113.1671	FAP
FAP-45	1301P	53.7745	-113.0639	FAP
FAP-46	1302P	53.7128	-113.0693	FAP
FAP-47	1303P	53.5414	-112.7178	FAP
FAP-48	1304P	54.0935	-113.0287	FAP
FAP-49	1305P	54.0071	-112.8640	FAP
FAP-50	1306P	54.1805	-112.8479	FAP
FAP-51	1307P	54.2388	-112.7258	FAP
FAP-52	1308P	54.2682	-113.0519	FAP
FAP-53	1309P	54.2681	-113.4004	FAP
FAP-54	1310P	54.1815	-113.2237	FAP
FAP-55	1311P	54.1019	-113.3748	FAP
FAP-56	1312P	54.0071	-113.2248	FAP
FAP-57	1313P	53.8012	-113.0509	FAP
Maskwa	1248C,1187P	54.6052	-110.4526	LICA
Therien	1176P	54.3108	-111.2261	LICA
Flat Lake	1177P	54.0726	-111.2051	LICA
Lake Eliza	1178P	53.8242	-111.1661	LICA
Telegraph Creek	1179P	53.7520	-110.5740	LICA
Muriel-Kehiwin	1181P	54.0934	-110.7444	LICA
Dupre	1182P	54.3346	-110.7796	LICA
La Corey	1183P	54.4997	-110.8179	LICA
Wolf Lake	1184P	54.6954	-110.8425	LICA
Foster Creek	1185P	55.0334	-110.5045	LICA
Primrose	1186P	54.7585	-110.4522	LICA
Frog Lake	1189P	53.8907	-110.3842	LICA
Clear Range	1190P	53.5565	-110.1542	LICA
Fishing Lake	1191P	53.9030	-110.0762	LICA
Beaverdam	1192P	54.1693	-110.2328	LICA
Cold Lake South	1174C, 1193P, 1227P	54.4137	-110.2328	LICA
Medley	1194P	54.7243	-110.0662	LICA
Fort George	1195P	53.8783	-110.7481	LICA
Burnt Lake	1196P	54.7946	-110.3444	LICA
Mahihkan	1197P	54.6230	-110.5089	LICA
Hilda Lake	1198P	54.5769	-110.3690	LICA
Town of Bonnyville	1199P	54.2709	-110.7435	LICA
St. Lina	1250C,1252P	54.2160	-111.5033	LICA
Grande Prairie	1165C	55.1767	-118.8075	PAZA
Evergreen Park	1166C	55.1175	-118.7647	PAZA
Smoky Heights	1167C	55.4028	-118.2814	PAZA

Beaverlodge	1168C	55.1963	-119.3968	PAZA
Valleyview	1170C	54.9405	-117.2153	PAZA
Fort McKay-Bertha Ganter	1032C,P	57.1894	-111.6406	WBEA
Fort McMurray-Athabasca Valley	1064C	56.7336	-111.3904	WBEA
Fort McMurray-Patricia McInnes	1070C	56.7514	-111.4767	WBEA
Fort Chipewyan	1071C	58.7092	-111.1750	WBEA
Millennium Mine	1075C	56.9693	-111.4007	WBEA
Fort McKay South	1076C	57.1492	-111.6423	WBEA
Anzac	1225C	56.4493	-111.0372	WBEA
CNRL Horizon	1226C	57.3037	-111.7395	WBEA
Shell Muskeg River	1244C	57.2491	-111.5087	WBEA
Mildred Lake	1066C	57.0500	-111.5641	WBEA
Mannix	1069C	56.9680	-111.4821	WBEA
Lower Camp	1074C	57.0325	-111.5064	WBEA
AH3	9901P	56.6964	-111.1223	WBEA
AH8	9902P	57.1019	-112.0754	WBEA
BM10	9903P	57.3201	-112.3970	WBEA
BM11	9904P	57.6913	-111.9095	WBEA
BM7	9905P	58.0582	-112.2814	WBEA
JP101	9906P	56.5395	-112.2756	WBEA
JP102	9907P	56.9096	-111.5407	WBEA
JP104	9908P	57.1190	-111.4254	WBEA
JP107	9909P	57.8903	-111.4352	WBEA
JP205	9910P	57.8378	-110.4490	WBEA
JP210	9911P	56.2728	-110.4487	WBEA
JP212	9912P	57.0537	-111.4092	WBEA
JP213	9913P	57.0473	-109.7497	WBEA
NE10	9914P	56.6085	-110.1926	WBEA
NE11	9915P	57.2879	-111.2171	WBEA
NE7	9916P	57.1467	-110.8663	WBEA
SM8	9917P	56.2016	-111.1753	WBEA
WF4	9918P	57.1479	-111.9840	WBEA
Violet Grove	1052C	53.1423	-115.1381	WCAS
Tomahawk	1053C	53.3724	-114.7686	WCAS
Carrot Creek	1054C	53.6211	-115.8692	WCAS
Steeper	1055C	53.1324	-117.0912	WCAS
Genesee	1057C	53.3016	-114.2211	WCAS
Meadows	1058C	53.5300	-114.6361	WCAS
Power	1059C	53.6330	-114.4199	WCAS
Edson	1062C	53.5936	-116.3928	WCAS
Breton	1063C	53.0903	-114.4606	WCAS
Wagner2	1241C	53.4294	-114.3806	WCAS

Table S5 Hourly NO_2 Similarity Ranking for the 1-R and Euclidean Distance (EuD) metrics. Note that stations at the *bottom* of the two columns are the most similar (hence one measure of their level of redundancy) with respect to each metric of dissimilarity. Acronyms are as in Table S1.

1-R	Name	ID	Aished	EuD (ppbv)	Name	ID	Aished
0.72	Maskwa	1248	LICA	1009	Shell Muskeg River	1244	WBEA
0.61	Anzac	1225	WBEA	950	Millennium Mine	1075	WBEA
0.60	ST.LINA	1250	LICA	950	Fort McMurray-Athabasca Valley	1064	WBEA
0.56	Steeper	1055	WCAS	923	Grande Prairie (Henry Pirker)	1165	PAZA
0.56	Caroline	1092	PAMZ	839	Calgary Northwest	1039	CRAZ
0.55	Lethbridge	1049	AEP	839	Calgary Central 2	1221	CRAZ
0.55	Crescent Heights	1172	PAS	807	Redwater Industrial	1156	FAP
0.54	Wagner2	1241	WCAS	769	Red Deer-Riverside	1142	PAMZ
0.54	Genesee	1057	WCAS	735	Edson	1062	WCAS
0.51	Shell Muskeg River	1244	WBEA	722	Meadows	1058	WCAS
0.51	CNRL Horizon	1226	WBEA	720	CNRL Horizon	1226	WBEA
0.46	Fort McMurray-Athabasca Valley	1064	WBEA	693	Edmonton South	1036	ACCA
0.46	Fort McMurray-Patricia McInnes	1070	WBEA	638	Fort McMurray-Patricia McInnes	1070	WBEA
0.46	Edson	1062	WCAS	566	Lethbridge	1049	AEP
0.46	Carrot Creek	1054	WCAS	566	Crescent Heights	1172	PAS
0.46	Meadows	1058	WCAS	528	Wagner2	1241	WCAS
0.42	Grande Prairie (Henry Pirker)	1165	PAZA	511	Woodcroft	2002	ACCA
0.42	Beaverlodge	1168	PAZA	511	Edmonton Central	1028	ACCA
0.39	Millennium Mine	1075	WBEA	439	Genesee	1057	WCAS
0.37	Redwater Industrial	1156	FAP	430	Beaverlodge	1168	PAZA
0.35	Power	1059	WCAS	428	Range Road 220	1161	FAP
0.32	Red Deer-Riverside	1142	PAMZ	428	Bruderheim	2000	FAP
0.28	Violet Grove	1052	WCAS	415	Carrot Creek	1054	WCAS
0.28	Tomahawk	1053	WCAS	411	Fort McKay South (Syncrude UE1)	1076	WBEA
0.21	Bruderheim	2000	FAP	411	Fort McKay-Bertha Ganter	1032	WBEA
0.19	Calgary Northwest	1039	CRAZ	403	Maskwa	1248	LICA
0.19	Calgary Central 2	1221	CRAZ	400	Ross Creek	1159	FAP
0.18	Range Road 220	1161	FAP	400	Fort Saskatchewan-92 St and 96 Ave	2001	FAP
0.16	Lamont County	1162	FAP	387	Anzac	1225	WBEA
0.16	Elk Island	1157	FAP	350	Violet Grove	1052	WCAS
0.16	Fort McKay South (Syncrude UE1)	1076	WBEA	350	Tomahawk	1053	WCAS
0.16	Fort McKay-Bertha Ganter	1032	WBEA	348	Power	1059	WCAS
0.15	Edmonton Central	1028	ACCA	301	Caroline	1092	PAMZ

0.14	Woodcroft	2002	ACCA	280	Steeper	1055	WCAS
0.14	Edmonton South	1036	ACCA	280	ST.LINA	1250	LICA
0.11	Ross Creek	1159	FAP	263	Lamont County	1162	FAP
0.11	Fort Saskatchewan-92 St and 96 Ave	2001	FAP	263	Elk Island	1157	FAP

Table S6 Hourly SO_2 Similarity Ranking for the 1-R and EuD metrics. Note that stations at the *bottom* of the two columns are the most similar (hence one measure of their level of redundancy) with respect to each metric of dissimilarity. Acronyms are as in Table S1.

1-R	Name	ID	Aished	EuD (ppbv)	Name	ID	Aished
1.01	Redwater Industrial	1156	FAP	1594	Redwater Industrial	1156	FAP
0.95	Caroline	1092	PAMZ	709	Mannix	1069	WBEA
0.88	Valleyview	1170	PAZA	532	Mildred Lake	1066	WBEA
0.88	Smoky Heights	1167	PAZA	470	Millennium Mine	1075	WBEA
0.85	Maskwa	1248	LICA	412	Shell Muskeg River	1244	WBEA
0.85	Mannix	1069	WBEA	372	Lower Camp	1074	WBEA
0.83	Red Deer-Riverside	1142	PAMZ	269	CNRL Horizon	1226	WBEA
0.81	Steeper	1055	WCAS	231	Wagner2	1241	WCAS
0.81	Power	1059	WCAS	231	Genesee	1057	WCAS
0.81	Meadows	1058	WCAS	220	Edmonton East	1029	ACCA
0.77	Edmonton East	1029	CRAZ	215	Maskwa	1248	LICA
0.76	Mildred Lake	1066	WBEA	209	Ross Creek	1159	FAP
0.74	ST.LINA	1250	LICA	195	Meadows	1058	WCAS
0.74	Anzac	1225	WBEA	193	Fort McMurray-Patricia McInnes	1070	WBEA
0.73	Woodcroft	2002	ACCA	193	Fort McMurray-Athabasca Valley	1064	WBEA
0.71	Violet Grove	1052	WCAS	192	ST.LINA	1250	LICA
0.70	Millennium Mine	1075	WBEA	177	Fort McKay South (Syncrude UE1)	1076	WBEA
0.70	Lower Camp	1074	WBEA	177	Fort McKay-Bertha Ganter	1032	WBEA
0.68	Beaverlodge	1168	PAZA	167	Valleyview	1170	PAZA
0.67	Shell Muskeg River	1244	WBEA	144	Anzac	1225	WBEA
0.65	Tomahawk	1053	WCAS	142	Lamont County	1162	FAP
0.65	Breton	1063	WCAS	142	Bruderheim	2000	FAP
0.62	Edson	1062	WCAS	138	Violet Grove	1052	WCAS
0.62	Carrot Creek	1054	WCAS	130	Tomahawk	1053	WCAS
0.61	Grande Prairie (Henry Pirker)	1165	PAZA	130	Breton	1063	WCAS
0.61	Evergreen Park	1166	PAZA	130	Elk Island	1157	FAP
0.59	Ross Creek	1159	FAP	124	Power	1059	WCAS
0.57	Lethbridge	1049	CRAZ	118	Woodcroft	2002	ACCA
0.57	Crescent Heights	1172	PAS	108	Edson	1062	WCAS
0.53	CNRL Horizon	1226	WBEA	106	Range Road 220	1161	FAP

0.53	Elk Island	1157	FAP	106	Fort Saskatchewan-92 St and 96 Ave	2001	FAP
	Wagner2	1241	WCAS	102	Caroline	1092	PAMZ
0.48	Genesee	1057	WCAS	91	Smoky Heights	1167	PAZA
0.45	Range Road 220	1161	FAP	79	Carrot Creek	1054	WCAS
0.45	Fort Saskatchewan-92 St and 96 Ave	2001	FAP	70	Lethbridge	1049	CRAZ
0.39	Lamont County	1162	FAP	58	Beaverlodge	1168	PAZA
0.39	Bruderheim	2000	FAP	55	Grande Prairie (Henry Pirker)	1165	PAZA
0.35	Fort McMurray-Patricia McInnes	1070	WBEA	50	Crescent Heights	1172	PAS
0.35	Fort McMurray-Athabasca Valley	1064	WBEA	42	Evergreen Park	1166	PAZA
0.19	Fort McKay South (Syncrude UE1)	1076	WBEA	24	Steeper	1055	WCAS
0.19	Fort McKay-Bertha Ganter	1032	WBEA	24	Red Deer-Riverside	1142	PAMZ

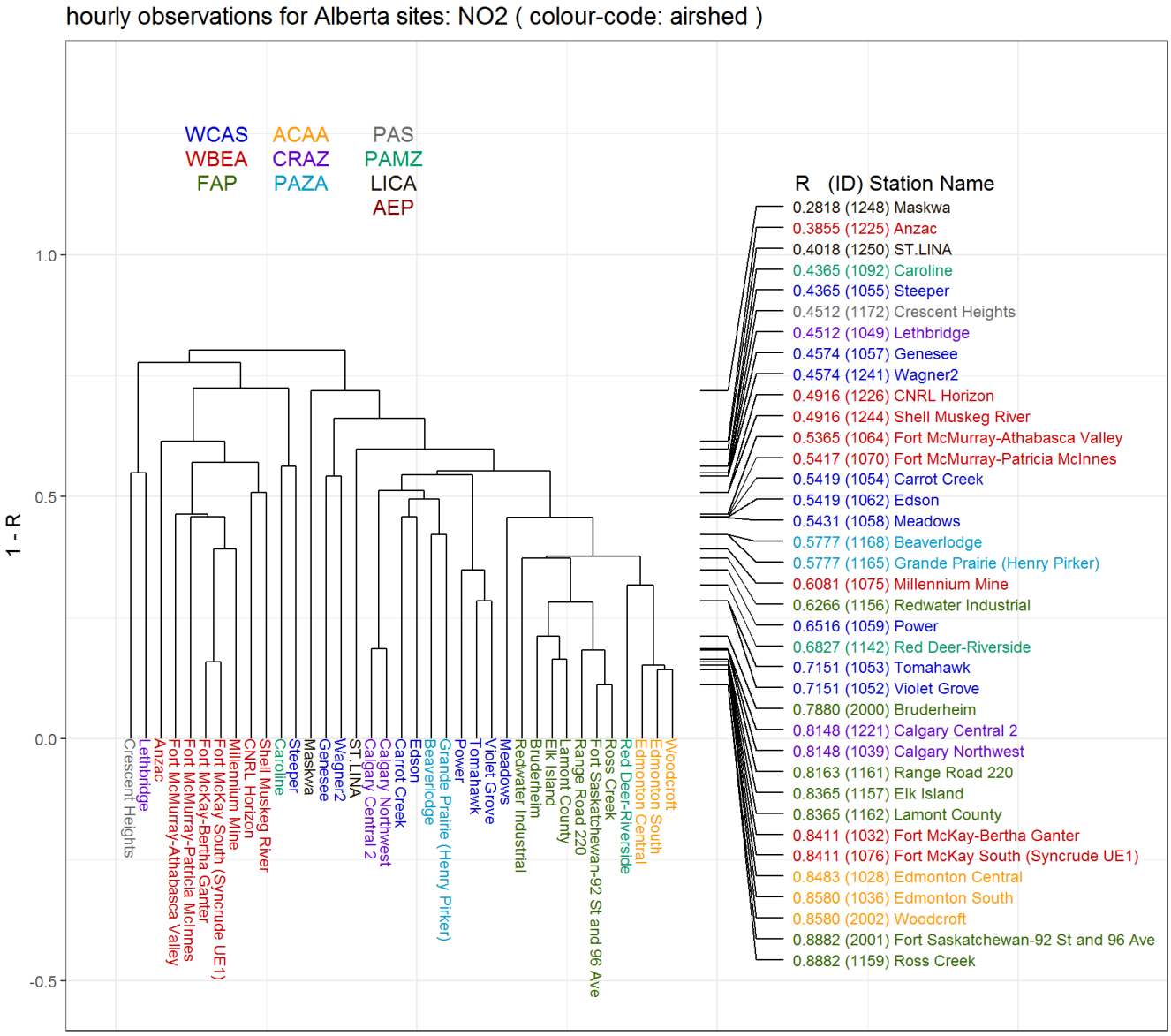

Figure S1: Dendrogram analysis for continuous hourly NO_2 averages using 1-R as the metric to compute the dissimilarity matrix. Airshed acronyms: West Central Airshed Society (WCAS), Wood Buffalo Environmental Association (WBEA), Fort Air Partnership (FAP), Alberta Capital Airshed Alliance (ACAA), Calgary Regional Airshed Zone (CRAZ), Peace Airshed Zone Association (PAZA), Palliser Airshed Society (PAS), Parkland Airshed Management Zone (PAMZ) and Lakeland Industrial Community Association (LICA). The dendrongram is colour-coded according to Airshed. Right side:stations ranked from low (top) to high (bottom) correlation level.

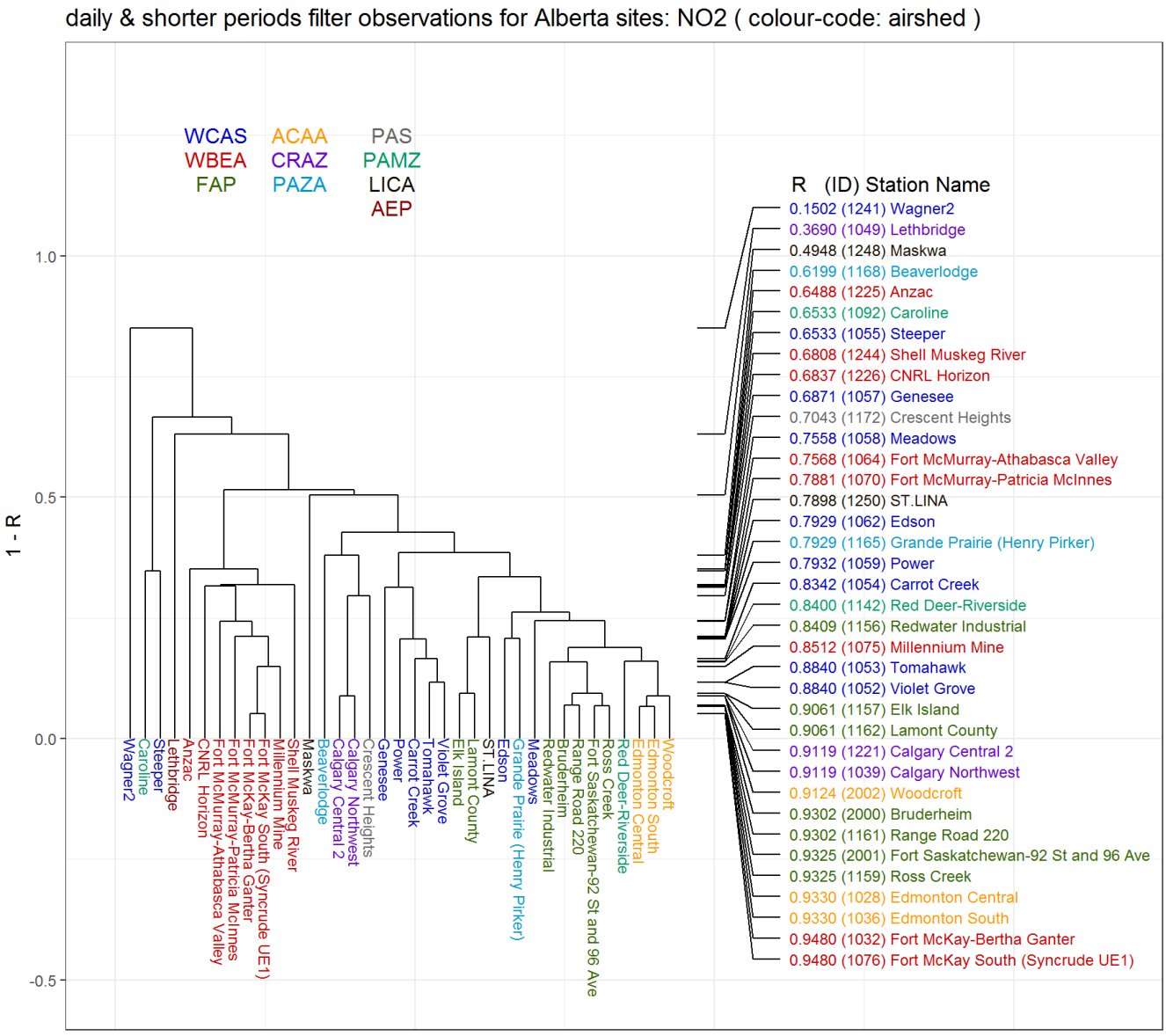

Figure S2: Dendrogram analysis for NO_2 daily or shorter time scales time series using 1-R as the metric to compute the dissimilarity matrix, for the Airsheds decribed in Figure S1. The dendrongram is colour-coded according to Airshed. Right side:stations ranked from low to high correlation level.

5

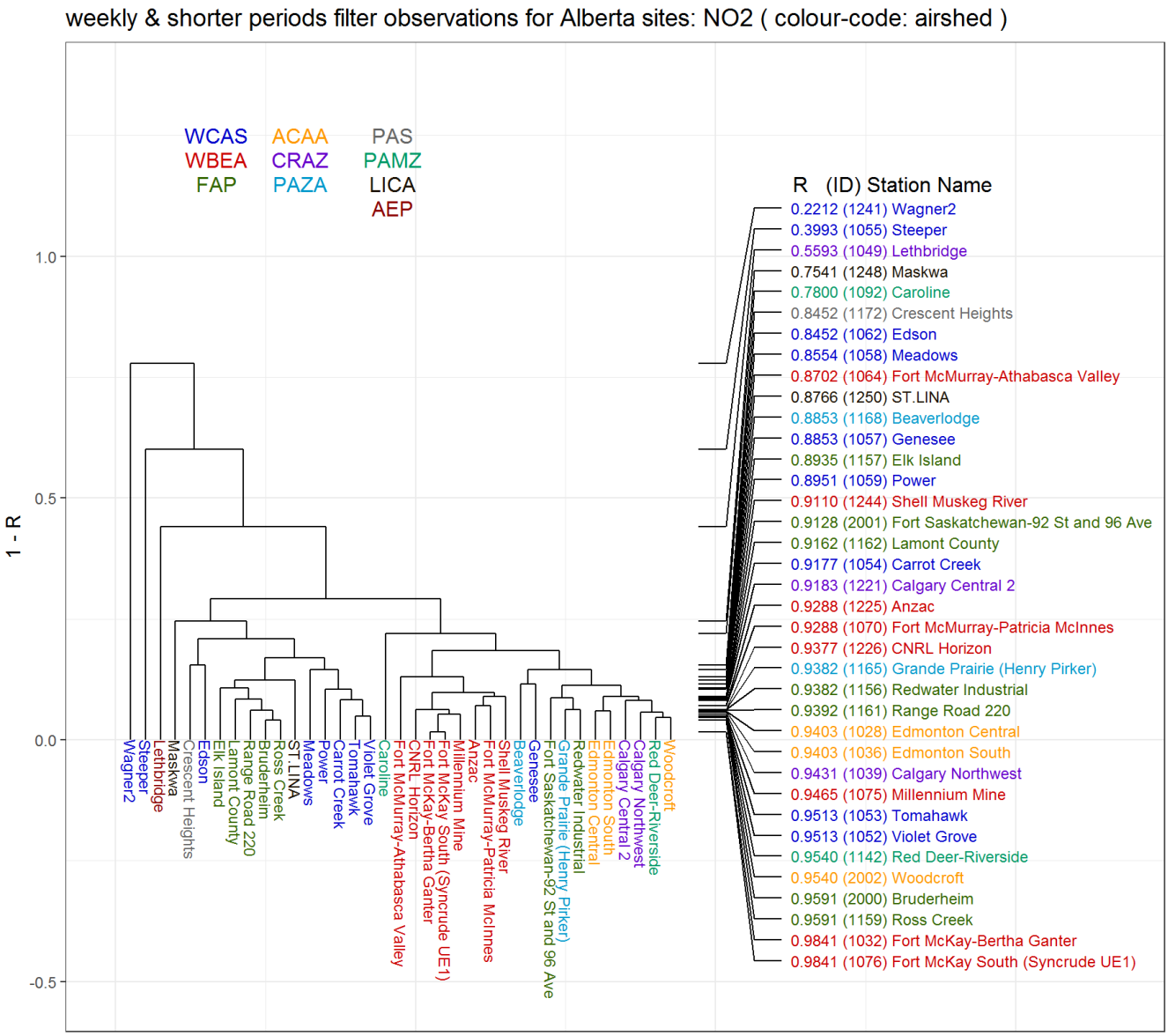

Figure S3: Dendrogram analysis for NO_2 weekly or shorter time scales time series using 1-R as the metric to compute the dissimilarity matrix, for the Airsheds decribed in Figure S1. The dendrongram is colour-coded according to Airshed. Right side:stations ranked from low to high correlation level.

5

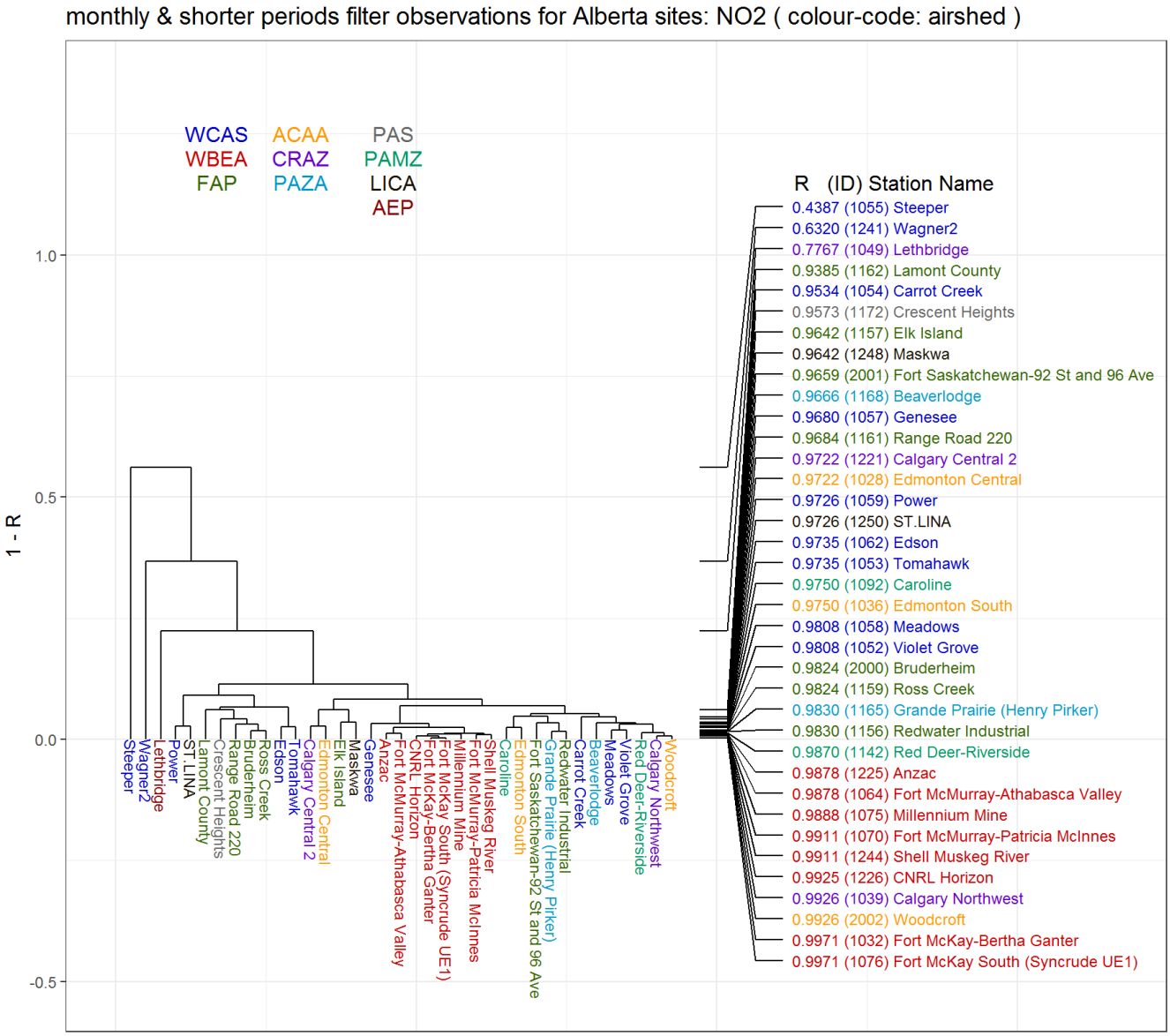

Figure S4: Dendrogram analysis for NO_2 monthly or shorter time scales time series using 1-R as the metric to compute the dissimilarity matrix, for the Airsheds decribed in Figure S1. The dendrongram is colour-coded according to Airshed. Right side:stations ranked from low to high correlation level.

5

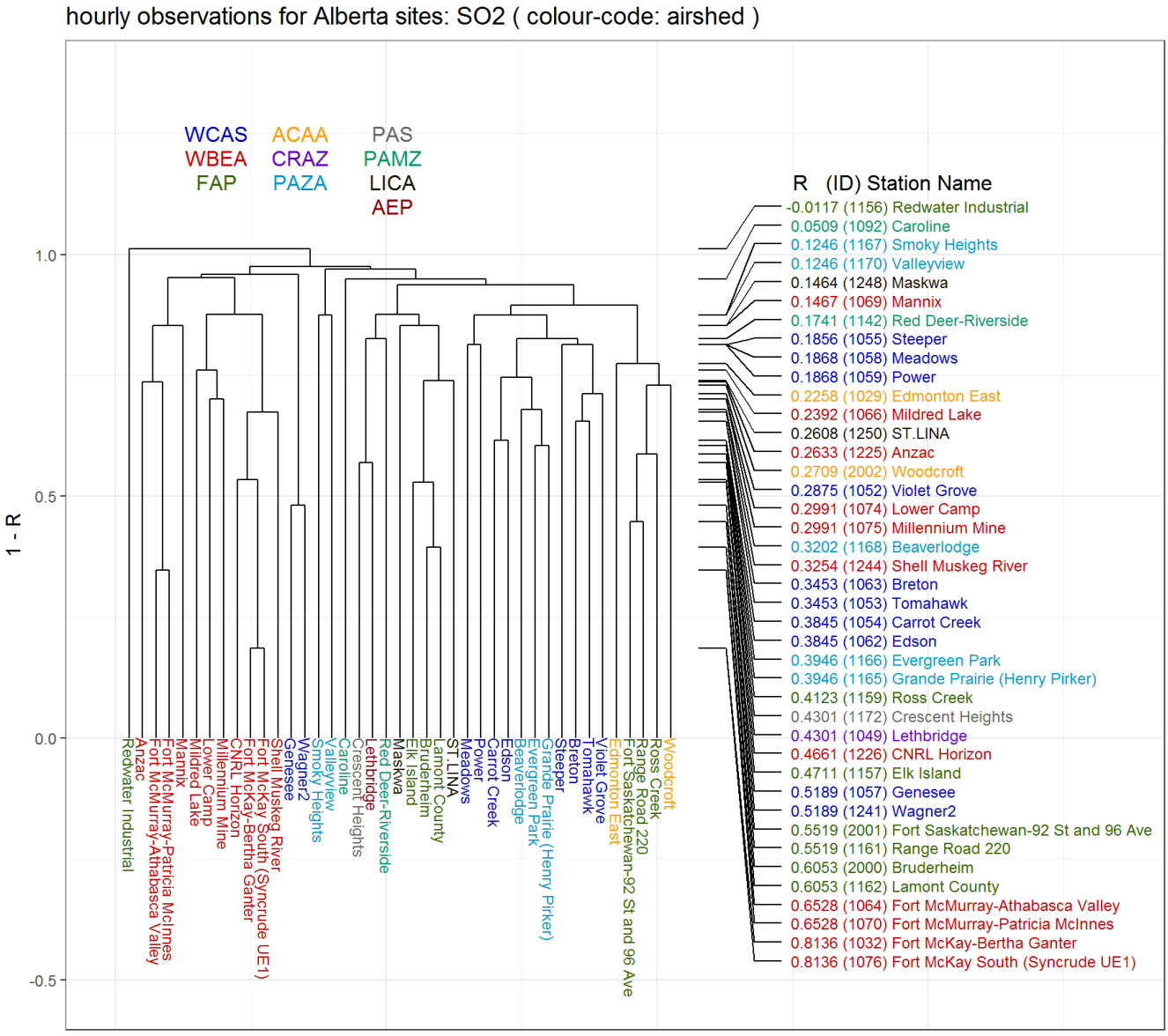

Figure S5: Dendrogram analysis for continuous hourly SO_2 averages using 1-R as the metric to compute the dissimilarity matrix, for the Airsheds decribed in Figure S1. The dendrongram is colour-coded according to Airshed. Right side:stations ranked from low to high correlation level.

5

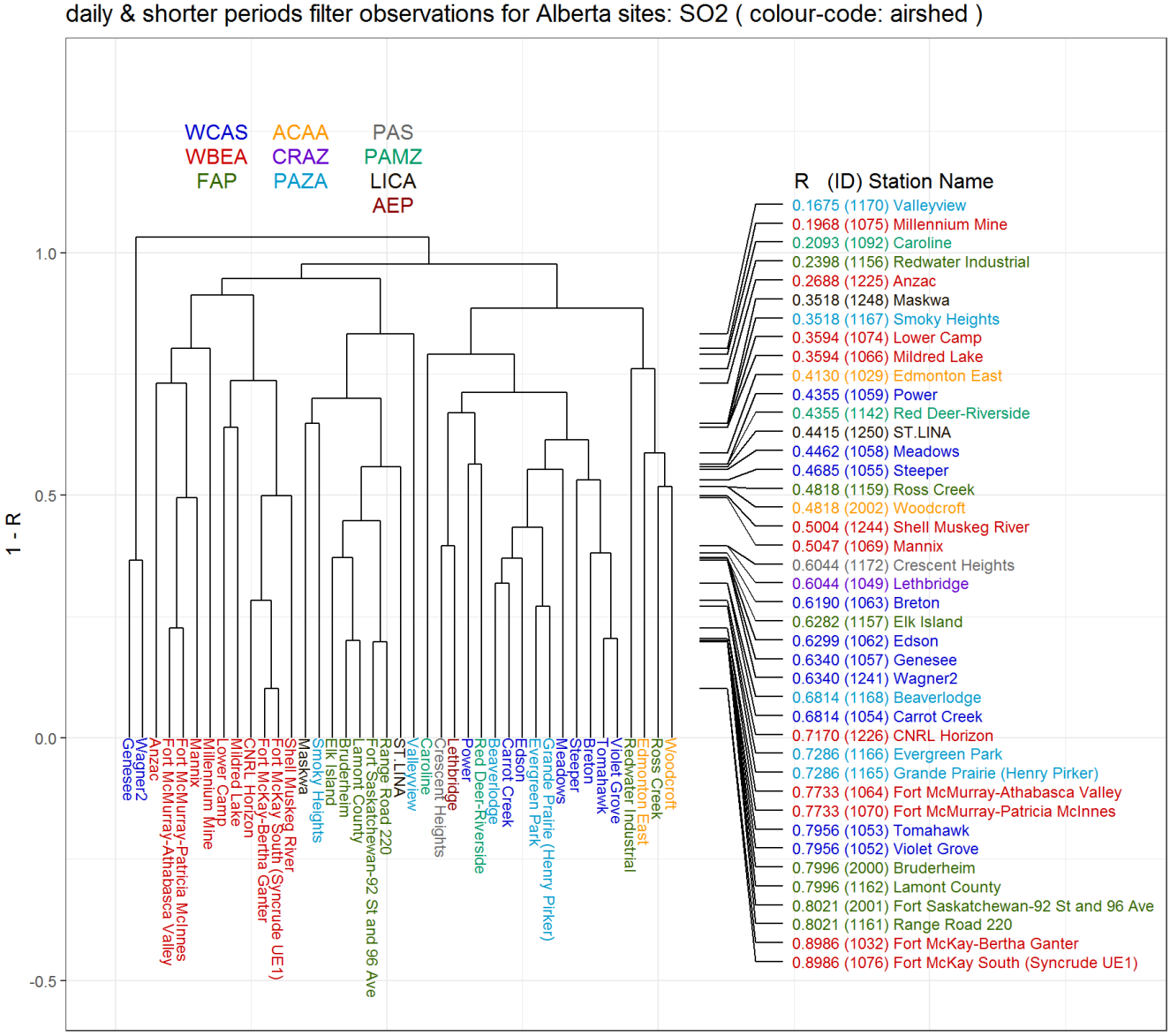

Figure S6: Dendrogram analysis for SO_2 daily or shorter time scales time series using 1-R as the metric to compute the dissimilarity matrix, for the Airsheds decribed in Figure S1. The dendrongram is colour-coded according to Airshed. Right side:stations ranked from low to high correlation level.

5

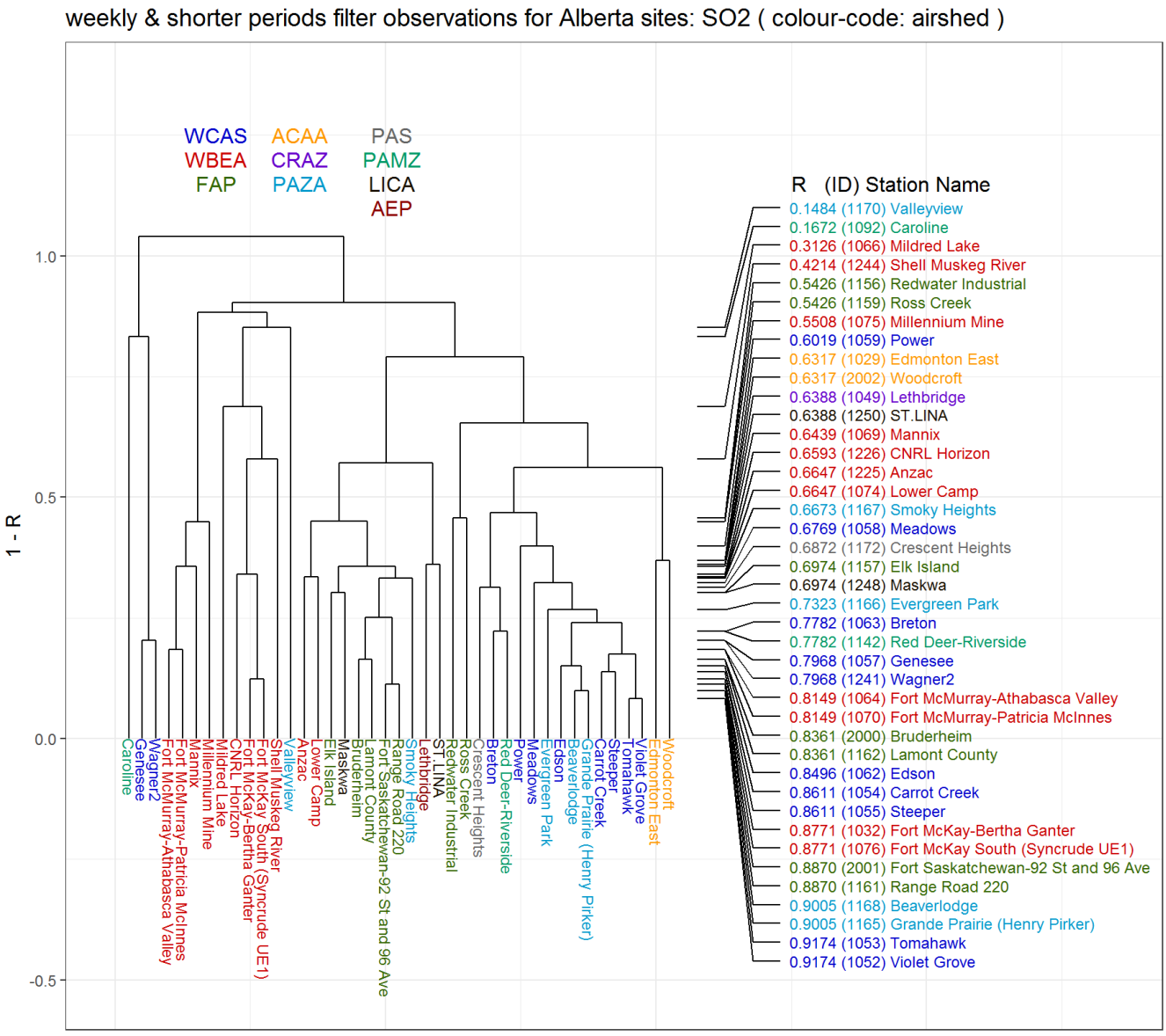

Figure S7: Dendrogram analysis for SO_2 weekly or shorter time scales time series using 1-R as the metric to compute the dissimilarity matrix, for the Airsheds decribed in Figure S1. The dendrongram is colour-coded according to Airshed. Right side:stations ranked from low to high correlation level.

5

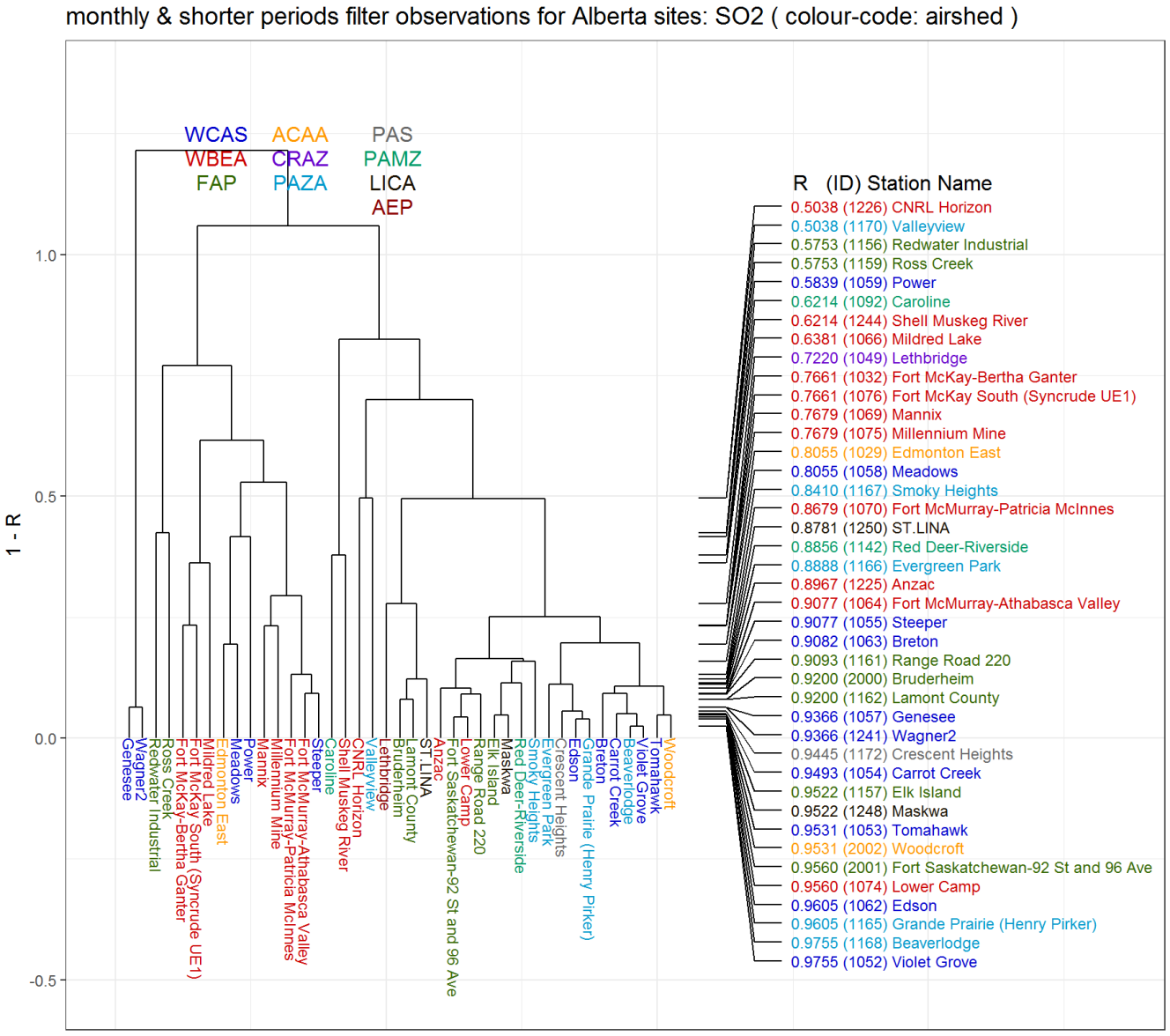

Figure S8: Dendrogram analysis for SO_2 monthly or shorter time scales time series using 1-R as the metric to compute the dissimilarity matrix, for the Airsheds decribed in Figure S1. The dendrongram is colour-coded according to Airshed. Right side:stations ranked from low to high correlation level.

5

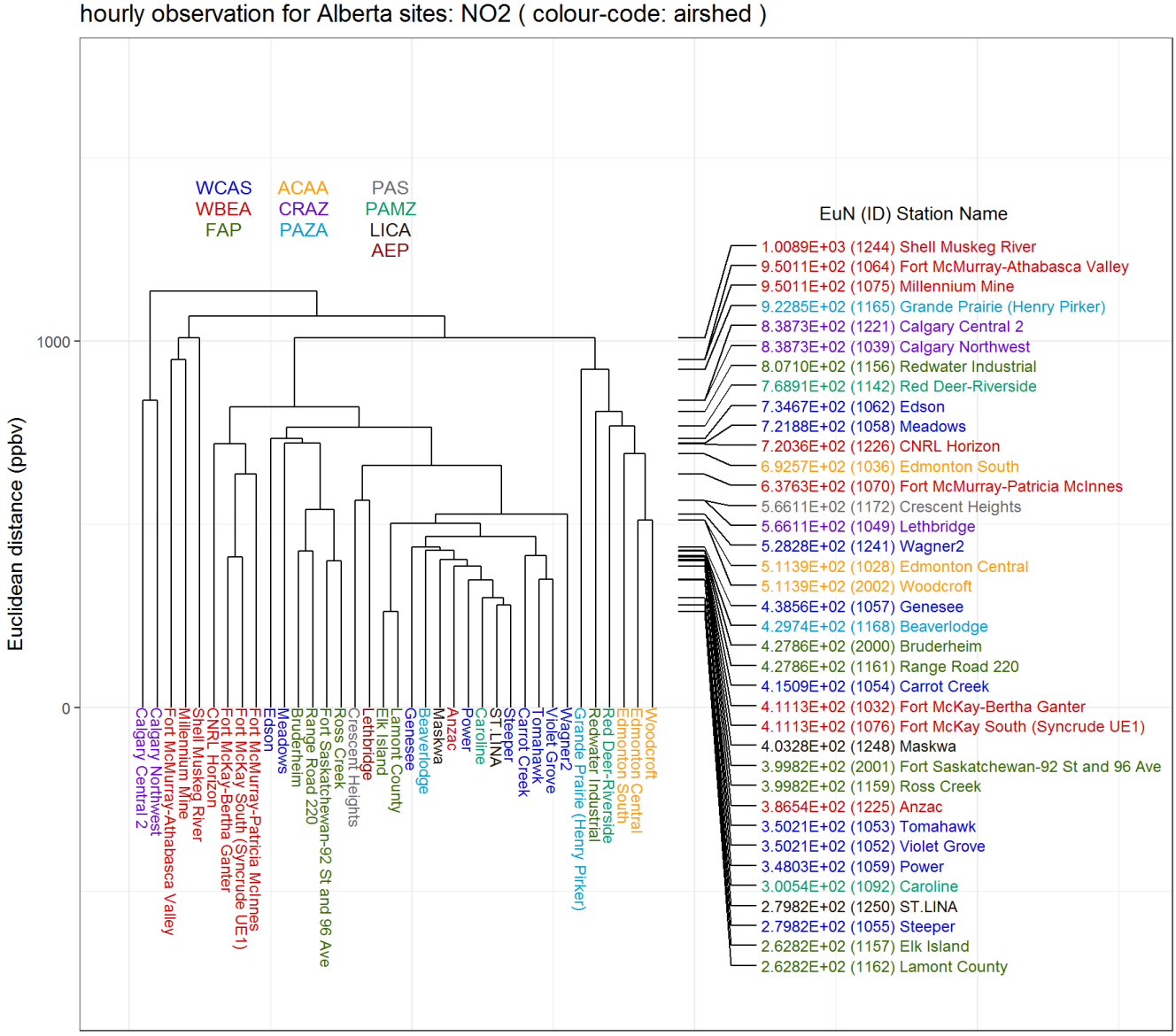

Figure S9: Dendrogram analysis for continuous hourly NO_2 averages using Euclidean distance as the metric to compute the dissimilarity matrix, for the Airsheds decribed in Figure S1. The dendrongram is colour-coded according to Airshed. Right side: stations ranked from high to low Euclidean distance level.

5

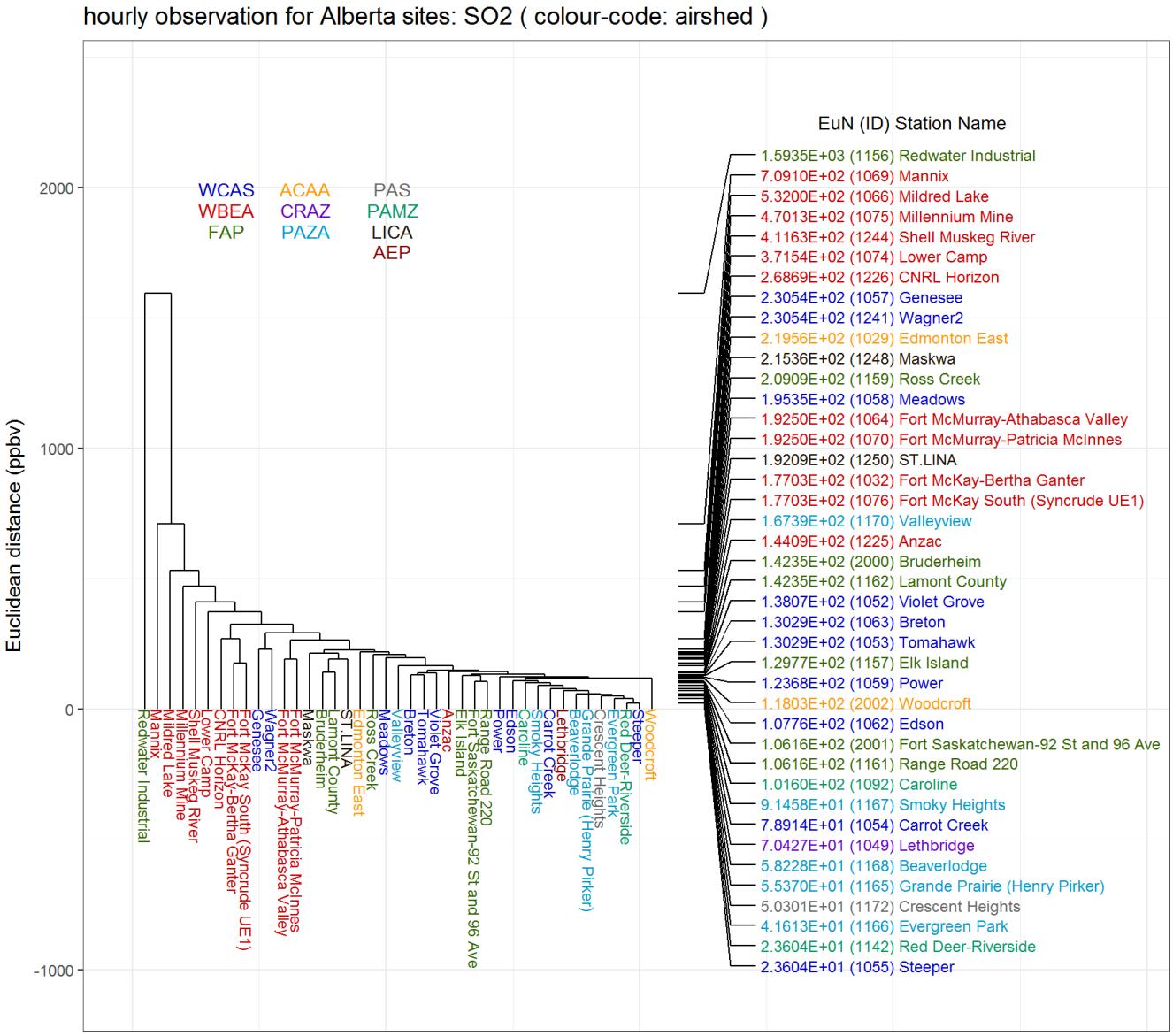

Figure S10: Dendrogram analysis for continuous hourly SO_2 averages using Euclidean distance as the metric to compute the dissimilarity matrix, for the Airsheds decribed in Figure S1. The dendrongram is colour-coded according to Airshed. Right side: stations ranked from high to low Euclidean distance level.

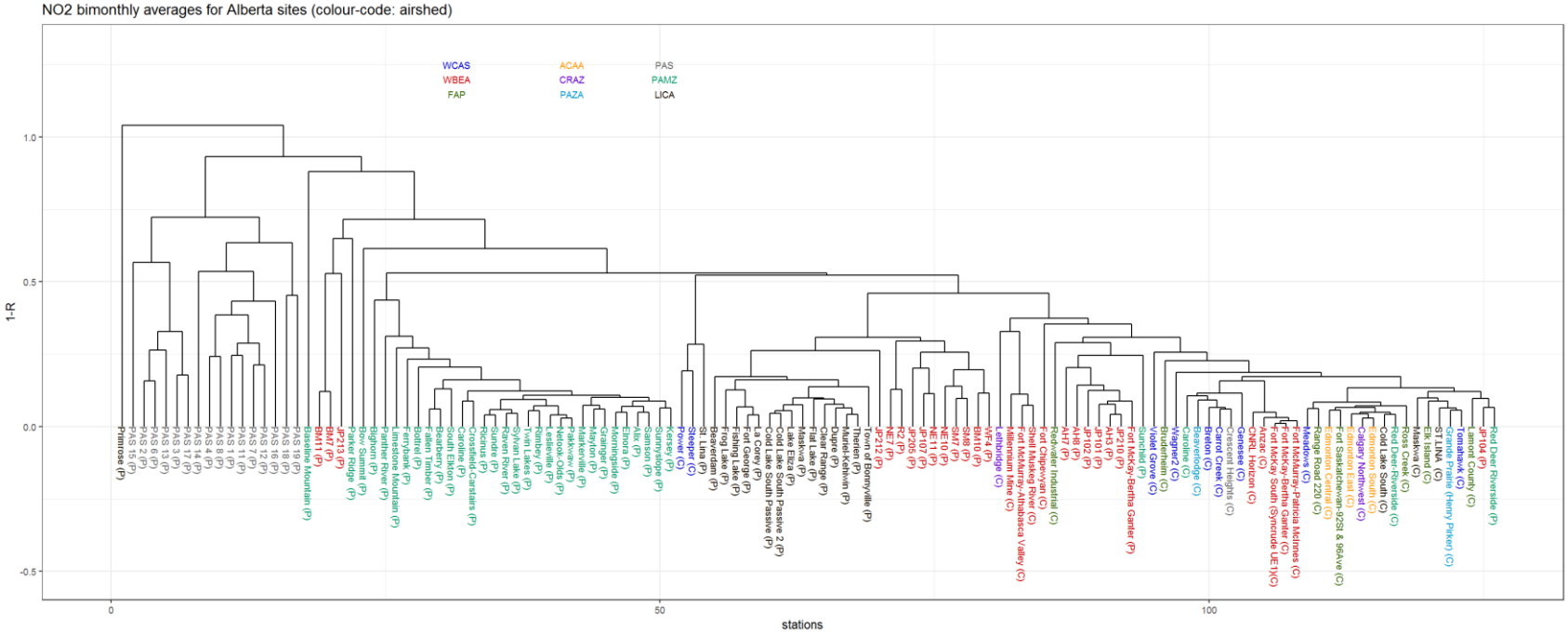

Figure S11 Dendrogram analysis for passive and continuous bimonthly NO_2 averages using 1-R as the metric to compute the dissimilarity matrix, for the Airsheds decribed in Figure S1. Stations are colour-coded according to Airshed. Station names which are continuous end in a "(C)", and stations which are passive end in a "".

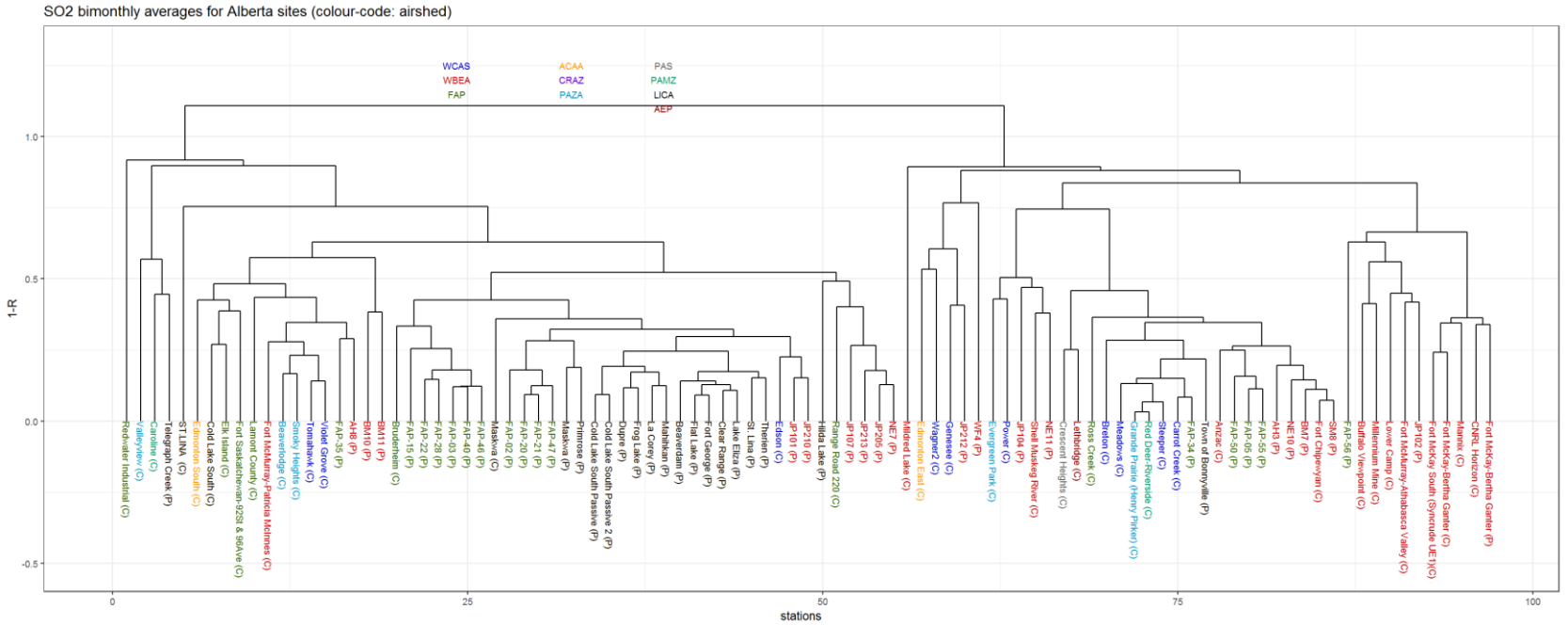

Figure S12 Dendrogram analysis for passive and continuous bimonthly SO_2 averages using 1-R as the metric to compute the dissimilarity matrix, matrix, for the Airsheds decribed in Figure S1. Stations are colour-coded according to Airshed. Station names which are continuous end in a "(C)", and stations which are passive end in a "(P)".

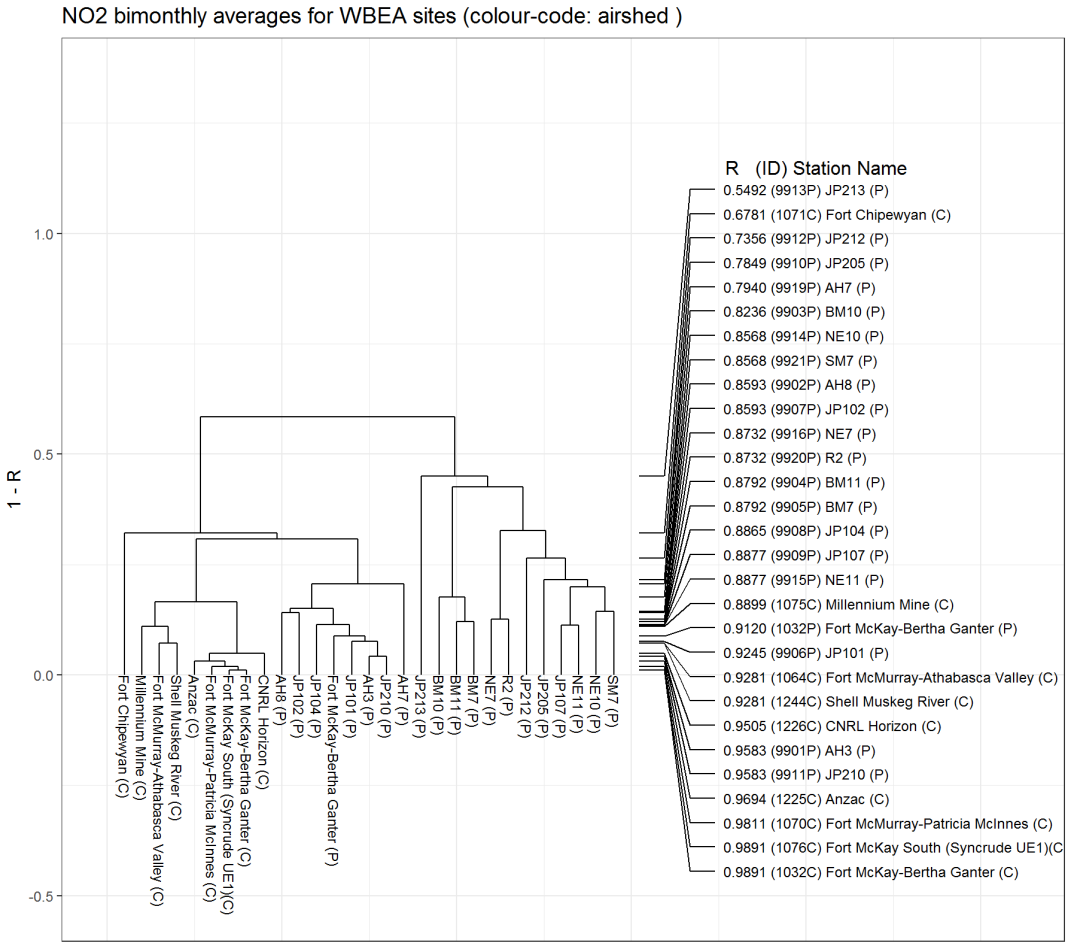

Figure S13 Dendrogram analysis for passive and continuous bimonthly NO_2 averages using 1-R as the metric to compute the dissimilarity matrix, for the Wood Buffalo Environemental Association (WBEA). Station names which are continuous end in a "(C)", and stations which are passive end in a "(P)". Right side:stations ranked from low to high correlation level.

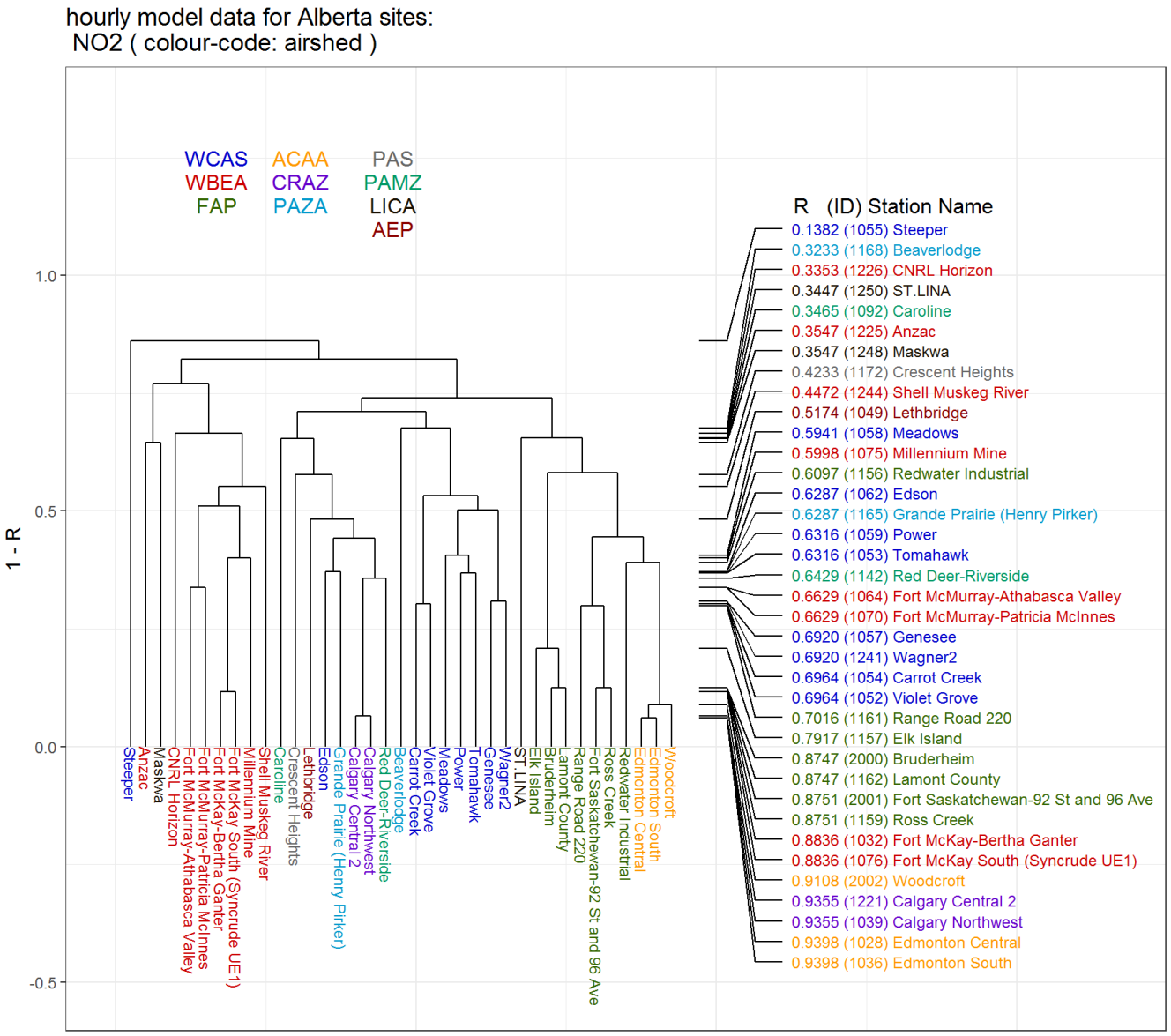

Figure S14: Dendrogram analysis for modelled hourly NO_2 averages using 1-R as the metric to compute the dissimilarity matrix, for the Airsheds described in Figure S1.

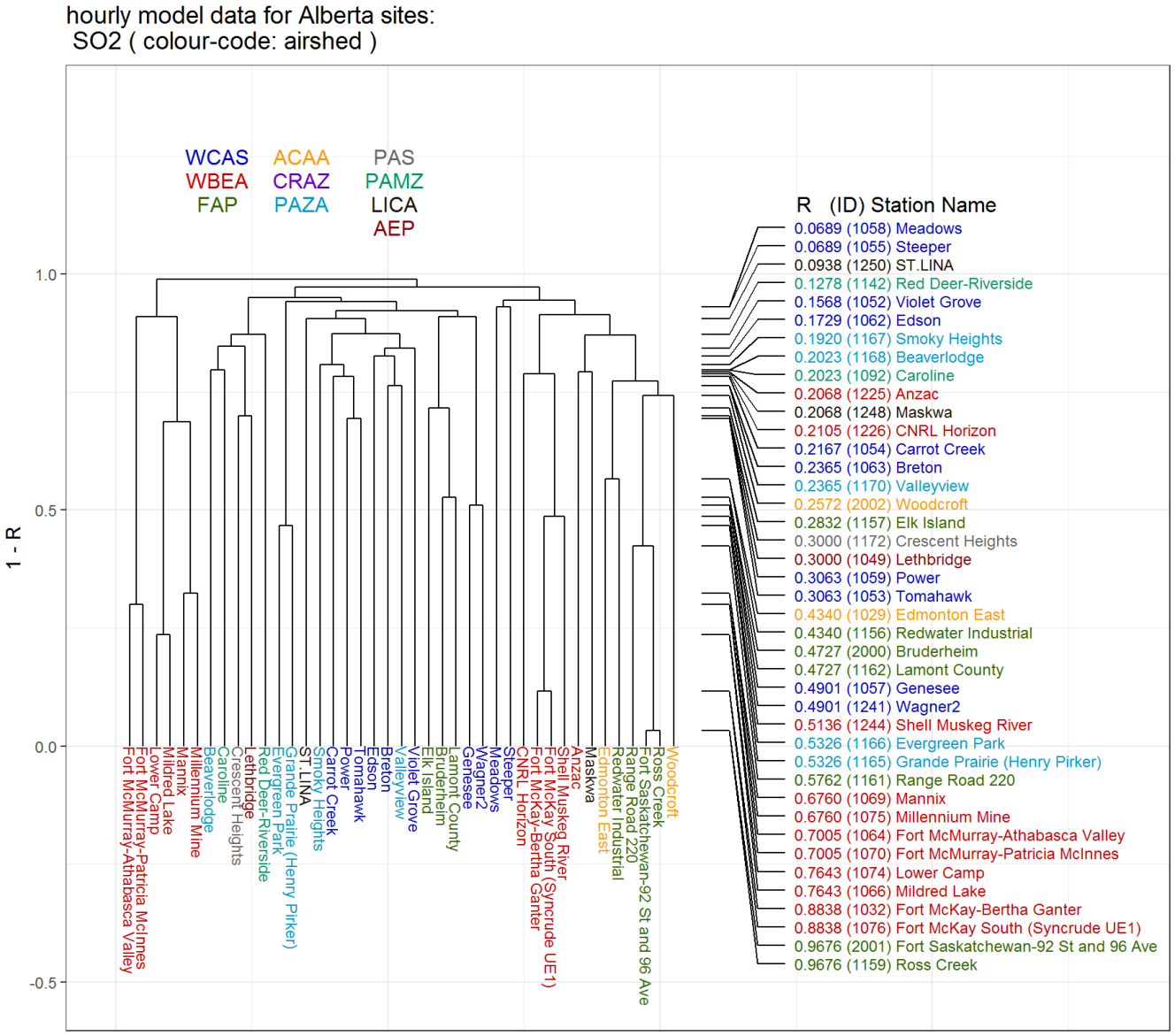

Figure S15: Dendrogram analysis for modelled hourly SO_2 averages using 1-R as the metric to compute the dissimilarity matrix, for the Airsheds described in Figure S1.

Supplement 2

The KZ Filter, Low-Pass versus Band Pass Filtering

The KZ-filter is defined as an iteration of a moving average filter applied on a time-series $S(t_i)$ (Zurbenko, 1986):

$$KZ_{m,p} = R_{i=1}^{p}\left\{J_{k=1}^{W_i}\left[\frac{1}{m}\sum_{j=\frac{m-1}{2}}^{\frac{m-1}{2}} S(t_i)_{k,j}\right]\right\}\{\, W_i = L_i - m + 1 \qquad \text{(A1)}$$

Where R is the iteration, m is the window size, p is the number of iterations, J is the running window, $S(t_i)$ the time series, and L_i is the length of the time series $S(t_i)$. Equation (A1) may be interpreted as p successive applications of a moving average of length m to the time series S, with the updated S being used as the starting time series for the subsequent moving average. The initial time series must thus have additional entries before and after the period L of interest, in order to result in a filtered time series of length L following the last application of the moving average. The first moving average is computed with a running window J and becomes the input for the second pass, and so on. The KZ filter controlling parameters m and p allow different time scales to be removed and filtered, as is described below.

The KZ belongs to the class of low-pass filters (since it filters periods smaller than a selected cut-off represented by a specific pair of m and p. The filter removes high frequency variations from the data (with respect to the window size) and belongs to the class of low-pass filters (since it filters periods smaller than the selected cut-off period). The KZ filter's original intent was a low-frequency pass filter but has been used as a band-pass in several air quality applications (e.g. Kang *et al* (2013), Galmarini *et al* (2013), Hogrefe *et al* (2000), Rao *et al*, (1997)), through taking the differences of time series pre-filtered for different time scales. However, the application of the difference in KZ filters for band-pass purposes does not separate the spectral components completely, with the energy spectrum overlapping on between the neighbour components (Hogrefe *et al*., 2000, 2003). The band-pass applications of the KZ filter suggested by Solazzo and Galamarini (2015) were tested by the authors of the current report. Artificial time series were constructed to examine the band-pass application's ability to separate known time-scales in those time series; the results were mixed, with intermediate time scales known to be in the input data failing to be resolved in subsequent clustering analysis. The band-pass approach's inability to completely separate adjacent time scales is the likely cause of this problem; too much energy leakage occurred, reducing correlations in clustering, and adding "noise" to the analysis.

However, the KZ filter in its original low-pass form was found to be able to separate the time scales in the test data accurately, with clustering showing the influence of the different time scales, given an appropriate choice of the filtering parameters m and p. That is, the analysis used here removes all of the energy below each of the time scales of interest (or, equivalently, above specific frequency thresholds), rather than attempting a band-pass approach. In addition to the hourly

QA/QC and gap-filled data, the KZ filter was thus used to remove the energy for periods less than 1 day ($KZ_{17,3}$), less than 7 days ($KZ_{95,5}$) and less than 30 days is removed ($KZ_{523,3}$).

The choice of the values of *m* and *p* for these filters follows from the energy characteristics of the filter system. These can be derived from the transfer function of the KZ filter (see Eskridge *et al*. (1997) and Zurbenko, (1986) for details on the transfer function), given by

$$\left|\emptyset_{m,p}(\omega)\right|^2 = \left[\frac{1}{m}\frac{\sin(\pi m\omega)}{\sin(\pi\omega)}\right]^{2p} \qquad \text{(A2)}$$

where ω has units of cycles per hour (frequency), for hourly observation data. The transfer function defines the energy passed or removed by the filter as a function of frequency. Figure A1 shows the lines defining the low-pass filters for (m,p) = (17,3), (95,5), and (523,3) used in the current analysis. The frequencies to the left of the lines are "passed" by the filter for the given value of (m,p), those to the right are removed.

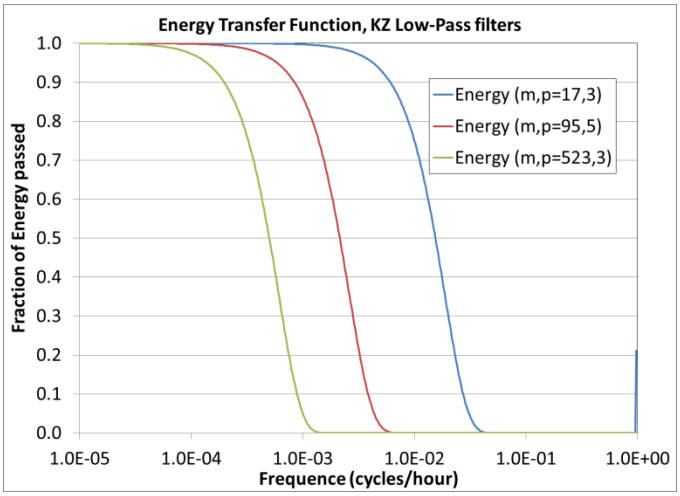

Figure A1 Energy transfer functions for $KZ_{17,3}$, $KZ_{95,5}$, $KZ_{523,3}$

It can be seen from inspection of Figure A1 that the lines forming the boundaries between frequencies which are passed and those which are removed are not step-functions, but have a gradual change – for example, the (523,3) KZ filter passes 99.75% of the energy for frequencies less than 3.0×10^{-5} cycles/hour (period greater than 3.75 years), 50% of the energy at 5.01×10^{-4} cycles per hour (periods of 83.2 days) and less than 0.25% of the energy at frequencies greater than 1.36×10^{-3} cycles/hour (periods less than 31.1 days). The filter characteristics of the three low-pass filters used here are given in Table A1 below.

Table A1 Frequency and period pass characteristics of the three KZ filters used here.

M	P	Frequency 99.75 %	Period 99.75 %	Frequency 50 %	Period 50 %	Frequency 0.25 %	Period 0.25 %
17	3	9.38332×10^{-4}	44.4 days	1.54338×10^{-2}	2.70 days	4.12341×10^{-2}	1.01 days
95	5	1.26846×10^{-4}	328 days	2.14594×10^{-3}	19.4 days	6.06578×10^{-3}	6.97 days
523	3	3.04471×10^{-5}	3.75 years	5.00840×10^{-4}	83.2 days	1.35751×10^{-3}	31.1 days

Table A1 shows that the three different (m,p) pairs selected for our work remove 99.75% of the energy for periods less than 1 day (17,3), 7 days (95,5), and 31 days (523,3), respectively. In subsequent figures and drawings, time series subjected to these filters will be referred to as having removed periods less than 1 day, less than 1 week, and less than one month. It should also be noted that despite the gradual slope of each band-pass filter, the near-complete removal of energy for periods less than those in the final column of Table A1 is a well-defined quantity. One can say with good confidence then that the resulting filtered time series will have less than 0.25% of the energy remaining for periods less than the limits shown in the table. One caveat on that is the 17,3 "daily" filter, which shows some energy leakage at periods specifically of the original time series (one cycle per hour); this daily filter will contain about 20% of the energy of periods equal to the original hourly time series interval.

The gradual slope, rather than a square-wave cut-off, for the KZ *low-pass* filter, highlights a potential difficulty with the use of the past use of that filter for *band-pass* purposes (e.g. Hogrefe *et al*, 2000, Solazzo and Galmarini, 2015). The use of the KZ filter as a band-pass filter involves two steps. In the first step, the KZ filter is applied on the original data for two different sets of (m,p) pairs, resulting in two different filtered time series. In the second step, the difference between these time series at each time point is constructed (lower order pair filtered time series – higher order pair filtered time series, at each time step). The KZ filters used in these past applications are shown in Figure A2: $KZ_{3,3}$, $KZ_{13,5}$, $KZ_{103,5}$, and $KZ_{310,7}$. The time series resulting from the difference between the original time series and $KZ_{3,3}$ is referred to as "intra-day" (periods less than 12 hours), whereas "diurnal" (periods between 12 hours and 2.5 days), "synoptic" (periods between 2 days and 21 days) and "long-term" (periods between 21 and 90 days) time series are formed from the differences $KZ_{3,3}$-$KZ_{13,5}$, $KZ_{13,5}$-$KZ_{103,5}$, and $KZ_{103,5}$-$KZ_{310,7}$, respectively. $KZ_{310,7}$ is said to form the "seasonal" (periods over 90 days) component of the time series.

The energy transfer functions for these "standard" filters are applied two ways in Figure A2. Figure A2 (a) shows the low-pass filters for the regions bounded by $KZ_{3,3}$, $KZ_{13,5}$, $KZ_{103,5}$, and $KZ_{310,7}$. The energy response of the filters has a similar shape to those in Figure A1, though the energy response of the $KZ_{3,3}$ filter can be seen to have a more significant

contribution near frequencies of 1 cycle per hour, suggesting a significant "leakage" of energy from short time scales with this filter. The regions used in previous work to describe different filter bands are labelled as noted above. While it can be seen by inspection of Figure A2 (a) that the energy associated with the difference between any two KZ filters will vary depending on frequency, the implications of that variation are more clearly displayed in Figure A2(b), in which the differences between low-pass filters are used to define the band-pass filters used in previous work. Figure A2 (b) shows significant overlap in filtered energy between the "diurnal" (green), "synoptic" (purple), "long term" (light blue), and "seasonal" (red) filters. For example, the seasonal and synoptic filters both pass 47% of the energy at a frequency of 5.75×10^{-4} cycles/hour (72 days), and the seasonal and synoptic filters both pass 49% of the energy at 2.0×10^{-3} cycles/hour (21 days); the diurnal and synoptic filters both pass 49% of the energy at 1.58×10^{-2} cycles/hour (2.6 days) and the diurnal and intraday filters share the same boundary for frequencies greater than 4.17×10^{-2} (1 day), including the region near frequencies of 1 cycle/hour. Some energy leakage occurs between the diurnal and synoptic filters as well, at the less than 0.05 level. The filters are not the ideal "square wave" associated with a band pass, but are subject to considerable overlap.

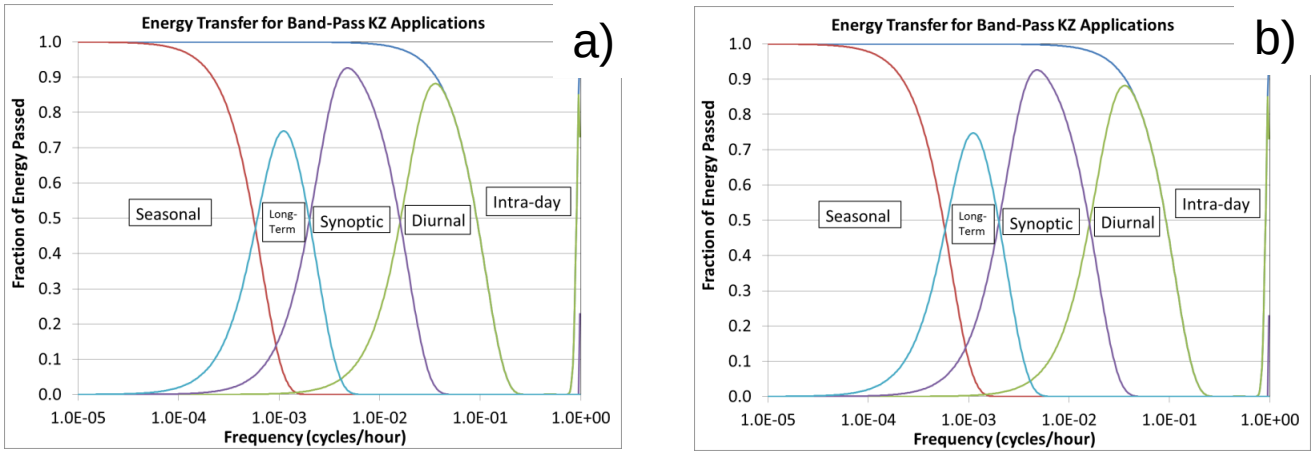

Figure A2 Energy transfer for previous applications of KZ filtering as a "band-pass" filter. (a Transfer functions for low-pass filters. (b Transfer functions for band-pass difference filters, as well as intra-day band-pass and seasonal low-pass filters

This degree of overlap has significant implications for the "bandpass" use of the KZ filter in the manner described in previous work (Eskridge *et al*, 1997, Hogrefe *et al*, 2000, Solazzo and Galmarini, 2015). The time labels for these filters are based on the 50% energy transfer levels of the differences to define a range in time represented by the filters. Figure A2(b) shows that these boundaries are not unique in energy – that is, a significant fraction of "seasonal" energy will be present in the "long-term" signal, a significant fraction of the "long-term" energy will be present in the "seasonal" signal, and so on.

In order to determine the potential impact of the overlap in band-pass on hierarchical clustering (described in more detail below), three time series were constructed for testing both low-pass and band-pass filtering combined with correlation

analysis. The three time series are intended to represent hypothetical observations from three observation sites (A,B,C), and the time signals going into their construction are shown in Figure A3, with the formulae describing the components of each time series and the net time series shown in Table A2.

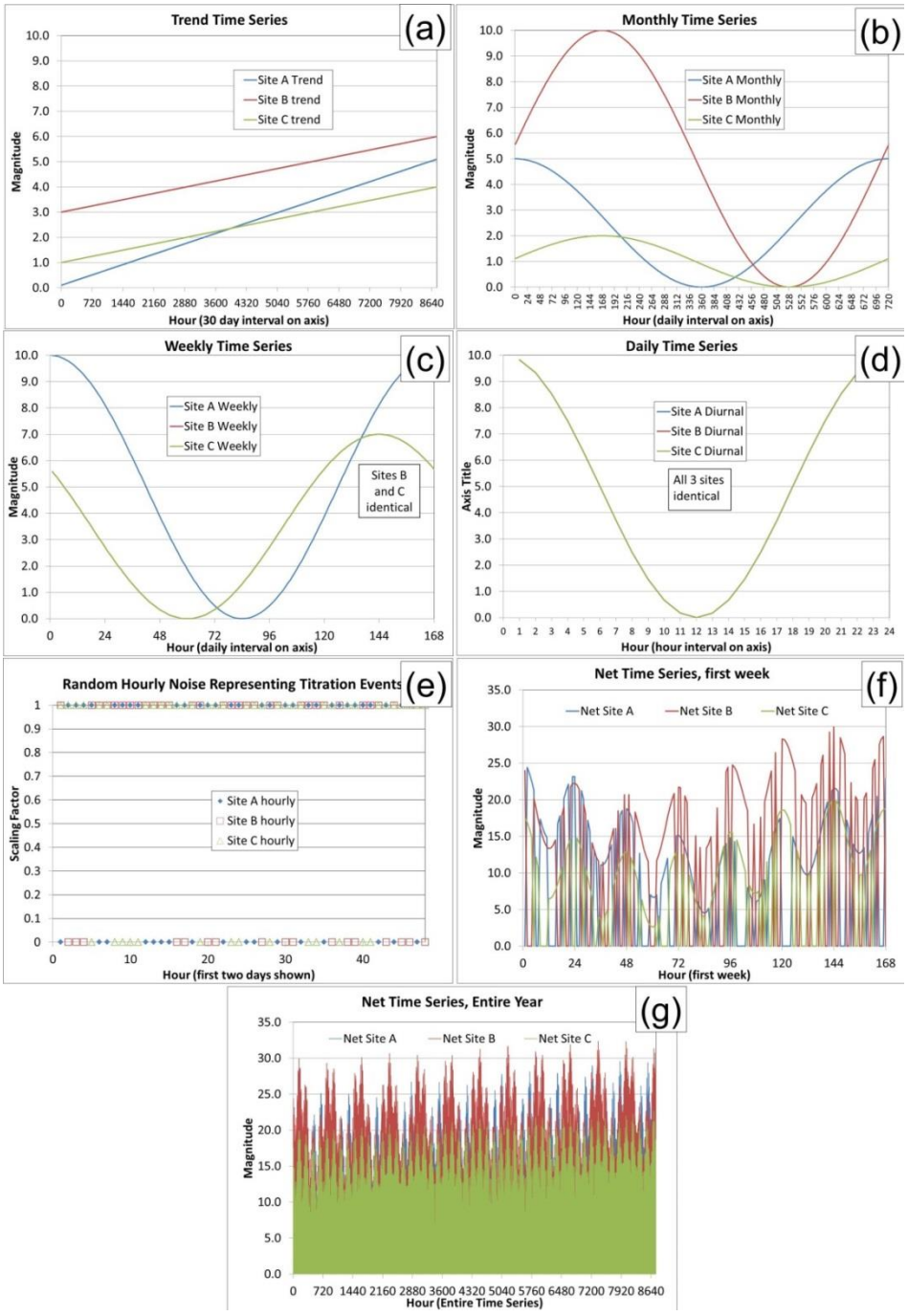

Figure A3 Construction of the test time series for three hypothetical stations. (a) Annual trend. (b) Monthly variation. (c) Weekly variation. (d) Diurnal variation. (e) Random noise (0 or 1) used to represent plume titration. (f) Resulting net signal. (g) Net signal for an entire year of hourly values.

Table A2 Components of Time Series for Testing

Site	Time Component Formula (h=hour of year)			
	Trend	**Monthly**	**Weekly**	**Daily**
A	$h\left(\frac{5}{8760}\right)+0.1$	$2.5\left[cos\left(\frac{\pi h}{360}\right)+1\right]$	$5\left[cos\left(\frac{\pi h}{84}\right)+1\right]$	$5\left[cos\left(\frac{\pi h}{12}\right)+1\right]$
B	$h\left(\frac{3}{8760}\right)+3$	$5\left[cos\left(\frac{\pi(h-168)}{360}\right)+1\right]$	$3.5\left[cos\left(\frac{\pi(h+24)}{84}\right)+1\right]$	$5\left[cos\left(\frac{\pi h}{12}\right)+1\right]$
C	$h\left(\frac{3}{8760}\right)+1$	$\left[cos\left(\frac{\pi(h-168)}{360}\right)+1\right]$	$3.5\left[cos\left(\frac{\pi(h+24)}{84}\right)+1\right]$	$5\left[cos\left(\frac{\pi h}{12}\right)+1\right]$

In order to create the time series used for testing and displayed in Figure A3 (f,g), the four components for each site described in Table A2 were added. The resulting time series where then multiplied by random numbers whereby at any given hour, two of the three time series values were multiplied by unity, with the remaining site value by zero, the choice of which time series to locally zero being chosen at random. This addition of random zeroing was added to mimic plumes which may reach only one station at a time (e.g. the summed time series representing ozone, and the zeroing representing a plume of NOx titrating ozone at one station and not the others). Table A2 and Figure A3 show that all three stations are identical in terms of their diurnal variation, stations B and C are identical for the weekly variation, stations B and C have the same monthly variation and a magnitude offset, and the stations all have different long-term trends. Having constructed this test dataset of three stations, it may be used for different KZ filtering approaches in order to determine whether those approaches may discern the timescales known to exist within the constructed time series.

The aim of the dissimilarity analysis (described in more detail in the following section) is to compare station time series based on a metric such as 1-R, where R is the Pearson correlation coefficient, in order to group stations based on the lowest level of dissimilarity (or highest correlation). For a simple set of only 3 stations such as has been constructed here for testing, the correlation between their time series need only be calculated three times for the three unique pairs of the stations ((A,B), (B,C) and (A,C)). The different methodologies for KZ filtering are first applied to the time series, and then correlations are calculated for the three resulting pairs of filtered time series; these may be used to determine whether the methodology used recovers information about the time scale used. Table A3 shows this analysis using the 1-R metric, for the band-pass filters, starting from the time series constructed from Table A2 and Figure A3.

Table A3 1-R values between pairs of test time series, for original time series and KZ band-pass filters.

Original Hourly Time Series				**Interpretation**
	A	**B**	**C**	The dissimilarity (1-R) is greater than unity for all pairs – the addition of the random zeroing has created sufficient noise that the original time series are anticorrelated.
A	0.0000	1.298	1.260	
B	1.298	0.0000	1.233	
C	1.260	1.233	0.0000	

Intra-day dissimilarity (original time series - $KZ_{3,3}$)				**Interpretation**
	A	**B**	**C**	The intraday dissimilarity includes most of the random noise: since a different station is being zeroed at every hour, most of the noise appears in this time scale – the dissimilarity values are all greater than unity, indicating that most of the noise occurs at this time scale.
A	0.0000	1.489	1.503	
B	1.489	0.0000	1.466	
C	1.503	1.466	0.0000	

Diurnal dissimilarity ($KZ_{3,3} - KZ_{13,5}$)				**Interpretation**
	A	**B**	**C**	The diurnal component is usually assumed to retrieve signals between 0.5 to 2.5 days. However, despite the *identical* diurnal signal present in all three original time series, the dissimilarity between all three time series remains high (the correlation remains low). The dissimilarity pairs ordered from lowest to highest are (A,C), (B,C), (A,B): the conclusion from this analysis would be that (A,C) are the most similar stations at this time scale, followed by (B,C) then (A,B). However, the temporal variation used to construct the time series is identical at this time scale – the band-pass methodology would lead to an erroneous conclusion. The low correlation is likely due to the lower frequency end of the band-pass including the time scale incorporating most of the noise.
A	0.0000	1.032	0.8968	
B	1.032	0.0000	0.9458	
C	0.8968	0.9458	0.0000	

Synoptic dissimilarity ($KZ_{13,5} - KZ_{103,5}$)				**Interpretation**
	A	B	C	The synoptic component is usually assumed to retrieve signals between 2.5 and 21 days. The pairings here from lowest to highest dissimilarity are (B,C) < (C,A) < (A,B). The
A	0.0000	0.7909	0.7222	

B	0.7909	0.0000	0.5286	methodology has successfully identified the (B,C) pair as the most similar; from
C	0.7222	0.5286	0.0000	Table A1, this is correct – the weekly signal is identical for this pair. The other two pairs should be *equally* dissimilar based on Table A1, but this is only true to the first digit in the band-pass analysis.

Long-term dissimilarity ($KZ_{103,5} - KZ_{310,7}$)				**Interpretation**
	A	B	C	The long term dissimilarity is intended to isolate signals between 21 and 90 days. The Monthly signal from Table A1 should therefore be resolved by this analysis. From Table A1, the (B,C) pair should have the greatest degree of similarity – and this is reflected in the analysis. (A,C) is shown to have a greater degree of similarity than (A,B). The periodicity differences between (A,C) and (A,B) should be identical, but the average value of the signals (A: 2.5, B: 5.0, C: 1.0) are likely why (A,C) has been identified as being more similar than (A,B). So at this time scale the results are reasonable.
A	0.0000	0.8408	0.6003	
B	0.8408	0.0000	0.4364	
C	0.6003	0.4364	0.0000	

Table A3 suggests that the diurnal filter may be severely affected by energy leakage from other parts of the frequency spectrum, failing to identify the identical similarity in the diurnal signal constructed here (and indicating a low degree of similarity at that time scale in general). The synoptic dissimilarity correctly identified the most similar pair, but failed to give the remaining two pairs identical similarities, indicating that energy leakage from the adjacent bands are also present at this time scale. The long-term dissimilarity seems to have captured the main features of that signal; a combination of similarities in magnitude or phase lag of the monthly time series.

Table A4 provides the low-pass filter results for the filters described in Table A1 and Figure A1.

Table A 4 1-R values between pairs of test time series, for KZ low-pass filters.

Filters out time scales less than 1 day ($KZ_{17,3}$)				**Interpretation**
	A	B	C	The daily time (and shorter) variation has been removed. (B,C) are the most similar due to their identical weekly time series and time variation for the monthly time series. (A,B) are the least similar due to their difference in magnitude and period at both monthly and weekly time scales. (A,C) are intermediate due to the similar period at monthly time scales and the relatively small
A	0.0000	0.8236	0.7208	
B	0.8236	0.0000	0.5282	
C	0.7208	0.5282	0.0000	

				size of the offset at weekly time scales.
Filters out time scales less than one week ($K_{95,5}$)				**Interpretation**
	A	B	C	The highest similarity is between (B,C), suggesting that the trend and the magnitude of the monthly signal dominates the similarity. (A,B) are the least similar, indicating that the monthly period offset between the signals and the difference in slope in the trend between these stations results in lower similarity than between (A,C). The intermediate values of the latter reflect the identical periods of the monthly signal and the identical slope in the trend.
A	0.0000	0.8455	0.6172	
B	0.8455	0.0000	0.4377	
C	0.6172	0.4377	0.0000	
Filters out time scales less than 1 month ($KZ_{532,3}$)				**Interpretation**
	A	B	C	The dissimilarities are low (and hence the similarities are high) for all variable pairs. (A,C) are the most similar, reflecting the similarity in the magnitudes of these lines during the year. (B,C) are the next most similar pair, reflecting the similarity in the slopes. (A,B) are the least similar pair, reflecting the similarity in the slopes but the constant offset between this last pair.
A	0.0000	0.2855	0.1083	
B	0.2855	0.0000	0.2146	
C	0.1083	0.2146	0.0000	

The use of the KZ as a low-pass filter as in Table A4 has some advantages for the shorter time scales compared to the band-pass filters of Table A3 – the noise leakage from the short term variations has contaminated the band-pass filters for the diurnal signal, creating negative correlation coefficients, reducing correlations and obscuring the identical variation at that time scale. The synoptic band-bass similarity also shows some energy leakage. The low-pass filters have removed the high frequency noise due to the choice of *m* and *p* values. The interpretation between band-pass and low-pass filters of course differs – the low-pass includes all frequencies less than the cut-off frequency (or all periods greater than the cut-off period), and must be interpreted in that context. Here we choose to use the low pass filters for our subsequent analysis, largely to avoid the high frequency noise and energy overlap issues shown above.

A McLauren series expansion of the sinusoids in Equation (A2), to the first two terms in the expansion for the numerator and denominator functions may be used to approximate the frequency energy cut-off- curves. If A is the fractional energy passed at a given frequency in the line, then that frequency may be approximated by:

$$\omega_0 \approx \frac{\sqrt{6}}{\pi} \sqrt{\frac{1-(A)^{\left(\frac{1}{2p}\right)}}{m^2-(A)^{\left(\frac{1}{2p}\right)}}} \tag{A3}$$

where ω_0 is the desired separating frequency and the approximate solution to the equation

$$\left|\emptyset_{m,p}(\omega)\right|^2 = A \tag{A4}$$

A value of A = ½ has been used in band-pass applications in the past to indicate the "boundaries" of these filters, though it can be seen from Figure A2(b) and Table A2 that the band-pass applications have significant energy leakage beyond these bounds.

References

Galmarini, S., Kioutsioukis, I., Solazzo, E.: E pluribus unum: ensemble air quality predictions, Atmos. Chem. Phys., 13, 7153-7182, 2013.

Kang, D., Hogrefe, C., Foley, K.L., Napelenok, S.L., Mathur, R., Rao, S.T.: Application of the Kolmogorov-Zurbenko filter and the decoupled direct 3D method for the dynamic evaluation of a regional air quality model, Atmos. Environ, 80, 58-69, 2013.

Hogrefe, C., Rao, S.T., Zurbenko, I.G., Porter, P.S.: Interpreting information in time series of ozone observations and model predictions relevant to regulatory policies in the eastern United States, Bulletin of the American Meteorological Society 81, 2083–2106, 2000.

Hogrefe, C., Vempaty, S., Rao, S.T., Porter, P.T.: A comparison of four techniques for separating different time scales in atmospheric variables, Atmos. Environ., 37, 3, 313-325, doi: 10.1016/S1352-2310(02)00897-X, 2003.

Rao, S.T., Zurbenko, I.G., Neagu, R., Porter, P.S., Ku, J.Y., Henry, R.F.: Space and time scales in ambient ozone data, Bull. Am. Meteorol. Soc. 78, 2153-2166, 1997.

Solazzo, E. and Galmarini, S.: Comparing apples with apples: Using spatially distributed time series of monitoring data for model evaluation, Atmos. Environ., 112, 234–245, 2015.