air

The design and assessment of the acid deposition monitoring network in Alberta

Government of Alberta

Environment

THE DESIGN AND ASSESSMENT OF THE ACID DEPOSITION MONITORING NETWORK IN ALBERTA

Prepared by

Samuel Shen
Intech Inc
Edmonton, Alberta, Canada

For

Alberta Environment

June 2009

Pub. No:
ISBN (Printed Edition) 978-0-7785-8778-1
ISBN (On-line edition) 978-0-7785-8779-8
Website: www.gov.ab.ca/env/

Disclaimer:
Although prepared with funding from Alberta Environment (AENV), the contents of this report/document do not necessarily reflect the views or policies of AENV, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

Any comments, questions, or suggestions regarding the content of this document may be directed to:

Air Policy Branch
Alberta Environment
10th Floor, Oxbridge Place
9820 – 106th Street
Edmonton, Alberta T5K 2J6

Additional copies of this document may be obtained by contacting:

Information Centre
Alberta Environment
Main Floor, Oxbridge Place
9820 – 106th Street
Edmonton, Alberta T5K 2J6
Phone: (780) 427-2700
Fax: (780) 422-4086
Email: env.infocent@gov.ab.ca

EXECUTIVE SUMMARY

Accurate monitoring of the Alberta acid deposition is a critical step in managing the risks of severe acid rain and maintaining Alberta's currently high standard of air, water, soil and ecosystems. Potential acid input (PAI) is used as a monitoring index, which is calculated from concentrations of several acid ions minus those of several alkaline ions in both wet and dry deposition. Alberta Environment altogether had twelve stations to monitor the PAI field. Among the twelve stations, nine are the continuous long-term stations from 1993 to current: Beaverlodge, Calgary, Cold Lake, Fort Chipewyan, Fort McMurray, Fort Vermilion, Kananaskis, Red Deer, and Suffield. Three stations were closed in the mid 1990s: Drayton Valley (closed in 1995), High Prairie (closed 1996), and Royal Park (closed in 1997). The annual PAI data of the nine stations from 1993-1996 are in Table S1.

Table S1: Annual PAI data from Alberta Environment [keq H^+ $ha^{-1}yr^{-1}$]

Year	Cold Lake	Fort Chipe-wyan	Calgary	Kana-naskis	Beaver-lodge	Fort Vermilion	Red Deer	Suffield	Fort Mc-Murray
1993	0.0575	-0.0020	0.1138	0.1872	0.0497	0.0195	0.1627	0.0969	0.0779
1994	0.0660	-0.0140	0.0706	0.0286	0.0871	0.0276	0.1148	0.0437	0.0581
1995	0.0470	0.0104	0.0664	0.1196	0.0749	-0.0031	0.1432	0.0572	0.1039
1996	0.0881	0.0158	0.1046	0.0983	0.0602	Missing	0.1240	0.0510	0.0922
1997	0.0740	-0.0007	0.0851	0.0726	0.0452	0.0141	0.0736	0.0338	0.0289
1998	0.0578	-0.0004	0.1966	0.1578	0.0778	-0.0127	0.1767	0.1118	0.0244
1999	0.0592	0.0115	0.1591	0.0709	0.0259	0.0065	Missing	0.0401	0.0577
2000	0.0638	0.0061	0.1084	0.0261	0.0558	0.0086	0.1104	0.0489	0.0394
2001	0.0963	0.0069	0.0674	0.0417	0.0567	-0.1045	0.1064	0.0629	0.0726
2002	0.1128	0.0223	0.1526	0.1735	0.0619	-0.1279	0.1375	0.0757	0.0867
2003	0.0908	0.0141	0.0336	0.0783	0.0693	0.0072	0.1404	0.0318	0.0646
2004	0.0891	0.0119	0.1297	0.1097	0.0430	0.0246	0.1623	0.0407	0.0336
2005	0.0650	-0.0035	0.0729	0.0595	0.0437	0.0060	0.1533	0.0173	-0.0010
2006	0.0592	0.0099	0.0392	-0.0140	0.0249	0.0275	0.1380	0.0922	-0.0037
Mean	0.0733	0.0063	0.1000	0.0864	0.0554	-0.0082	0.1341	0.0574	0.0525
StDev	0.0189	0.0095	0.0471	0.0588	0.0183	0.0481	0.0277	0.0274	0.0334

This research assesses Alberta's current acid deposition monitoring network of nine stations and designs future networks with additional stations. The annual PAI data of the current nine stations from 1993-2006 are used in this study. The empirical orthogonal function (EOF) approach is used to calculate the mean square error (MSE) of the station sampling for the Alberta spatial averaging. The minimization of the MSE leads to the results of assessment and design. The following has been found.

1. Inhomogeneous PAI distribution over Alberta: The mean PAI field in the period of 1993-2006 (Figure S1) and the PAI fields of individual years (Figure S2) demonstrate

the strong spatial inhomogeneity of the PAI field. The station PAI data are not highly correlated (See Table S2). The PAI action centre is the Red Deer-Calgary-Kananaskis corridor. This area usually observes high PAI values compared with other areas in Alberta.

2. The current network and PAI patterns: Figure S1 shows the locations of the nine current Alberta Environment's PAI monitoring stations. The 9-station network results in a small RMSE/SD ratio (5.6%). In the period of 14 years (1993-2006), only in three years (1993, 1998, and 2002) the PAI values were higher than the monitoring load of 0.17 [keq H^+ $ha^{-1}yr^{-1}$] at three locations: Red Deer, Calgary and Kananaskis (See Figure S2a, b and c). The highest PAI value, 0.20 [keq H^+ $ha^{-1}yr^{-1}$] observed at Calgary in 1998, is still less than the target load of 0.22 [keq H^+ $ha^{-1}yr^{-1}$] for sensitive soils. Thus, Alberta's acid deposition has not caused major environmental problems. The spatial patterns of the PAI derived from the observations demonstrate consistent high PAI activities in the Red Deer-Calgary-Kananaskis corridor, which is consistent with the RELAD modeling results of Alberta Environment (2006). However, the RELAD model predicted much higher PAI values, which are not supported by the observations contained in this report. The EOF analysis of gridded data, which is displayed in Appendix C, shows a dominant north-south oscillation in the eastern half of Alberta in the first mode, a northeast-midwest oscillation in the second mode, and a strong Red Deer-Calgary-Kananaskis *vs*. Beaverlodge dipole in the third mode. This Red Deer-Calgary-Kananaskis pattern persists in the fourth mode.

3. Ranking of the nine stations of the current network: According to a station's contribution to the reduction of sampling error, the nine stations of the existing PAI network are ranked as follows:
   1) Beaverlodge,
   2) Fort Chipewyan,
   3) Suffield,
   4) Red Deer,
   5) Cold Lake,
   6) Kananaskis,
   7) Calgary,
   8) Fort Vermilion, and
   9) Fort McMurray.

   It is not recommended that the last two stations Fort Vermilion and Fort McMurray be eliminated from the network, because Fort Vermillion is currently the only station in the vast northwest of Alberta and the oil sands exploration in the Fort McMurray area requires the close monitoring of the air quality and water quality risks and hence the station's continuation.

4. Ranking of the additional 10 stations for the future network: The locations of the additional 10 stations are selected under the consideration of monitoring needs,

physical accessibility and improved spatial coverage (see Figure S3). These 10 stations are ranked in the following order:

1) Consort,
2) Peerless Lake,
3) Cardston,
4) Fort Fitzgerald,
5) Edmonton,
6) Jasper,
7) Peace River,
8) Slave Lake,
9) Indian Cabins, and
10) Rainbow Lake.

It is recommended that the first seven stations be established when conditions allow. Consort, Peerless Lake, Cardston, and Fort Fitzgerald stations are helpful in improving the spatial coverage and reducing the sampling error. Edmonton station is useful in monitoring the industrial PAI load. Jasper station is useful in monitoring the air, water and soil quality for the tourism industry. Peace River station is helpful in monitoring the expanded farmland soil for the organic farming industry. The Indian Cabins and Rainbow Lake stations are ranked 9th and 10th, and they are not recommended to be established in the near future. Slave Lake station, ranked 8th, may be useful for some special agricultural monitoring needs.

5. Non-uniqueness of the future networks: Because of the annual PAI’s large spatial scale and hence small sampling error, there is little difference in changing the order of ranking after the first three or four stations. Thus, from the practical need of the PAI’s effective monitoring, the future optimal network is not unique. When the first three or four stations are chosen, the next few stations may be chosen according to the local needs, such as the need at Edmonton due to the rapid population increase or the need at Peace River due to the organic farming.

6. Ongoing effort to assess the monitoring network: When the 2007-2008 PAI data come in, Fort McMurray station’s importance will be likely to advance from the current least important position to a more important position due to the intensive oil sands exploration activities in 2007-2008. Since the EOF assessment and design analysis is an adaptive method, it can take various kinds of observed and modelling data. The other data to be included in the analysis are updated data from the nine stations, the data from the two West Central Airshed Society stations near Edmonton, the RELAD modelling data, and other temporary measurements of the dry PAI. Thus, it is very important for a design to be adaptive and dynamic and for the monitoring errors to be assessed every year with the updated data.

Table S2: Correlation matrix of the annual station PAI data.

	Cold La	Fort Ch	Calgary	Kana	Beav	Fort	Red	Suffield	FortMcM

						Ver	Deer		
Cold La	1.00	0.53	0.02	0.14	0.07	-0.64	-0.25	-0.17	0.27
Fort Ch	0.53	1.00	0.12	0.20	-0.25	-0.41	0.13	0.04	0.37
Calgary	0.02	0.12	1.00	0.61	0.01	-0.21	0.45	0.38	0.05
Kana	0.14	0.20	0.61	1.00	0.28	-0.28	0.49	0.40	0.47
Beav	0.07	-0.25	0.01	0.28	1.00	-0.17	-0.07	0.10	0.46
Fort Ver	-0.64	-0.41	-0.21	-0.28	-0.17	1.00	0.11	-0.22	-0.37
Red Deer	-0.25	0.13	0.45	0.49	-0.07	0.11	1.00	0.37	-0.07
Suff	-0.17	0.04	0.38	0.40	0.10	-0.22	0.37	1.00	0.05
FortMcM	0.27	0.37	0.05	0.47	0.46	-0.37	-0.07	0.05	1.00

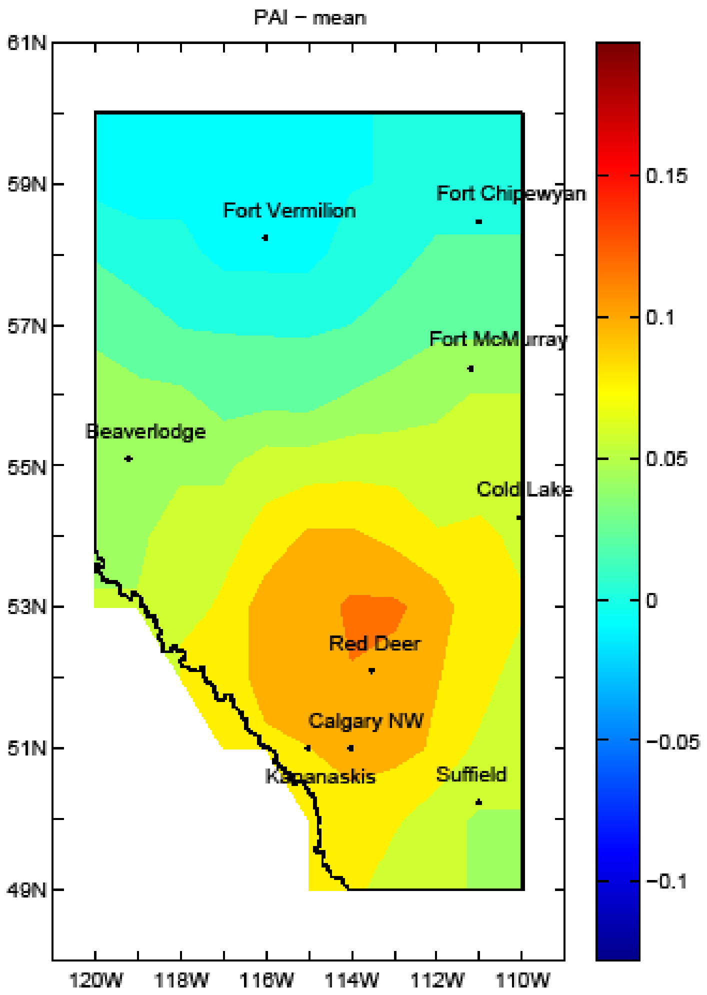
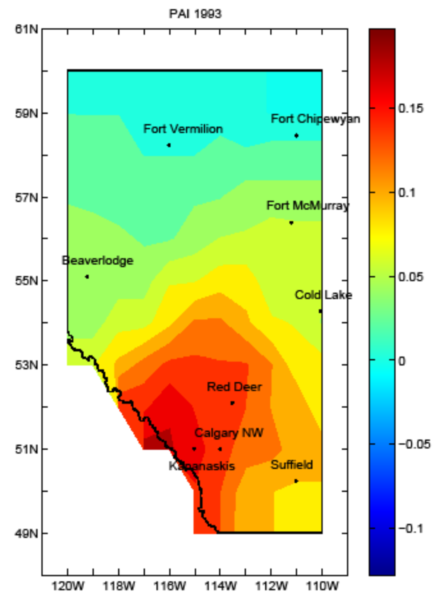
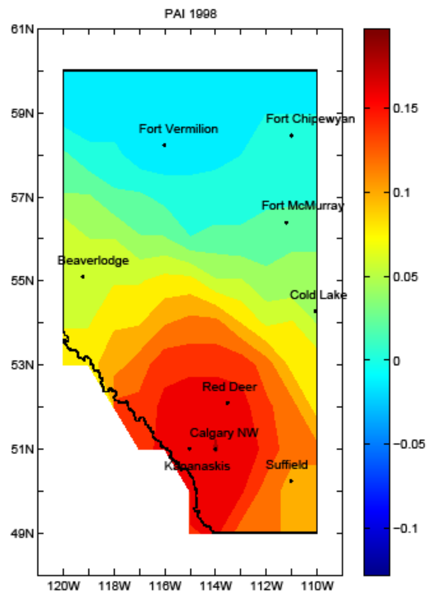


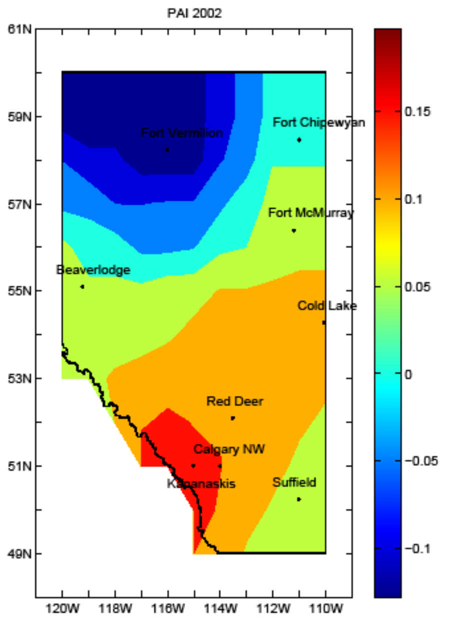
Figure S1: The mean PAI field based on the 1993-2006 annual data. Red Deer-Calgary-Kananaskis corridor shows the highest PAI values [Units: keq H^+ $ha^{-1}yr^{-1}$].



a



b



c

Figure S2: Annual PAI spatial distributions: a. for year 1993, b. for 1998, and c for 2002. Only in these three years, one or more of the three stations (Red Deer, Calgary, and Kananaskis) observed more than 0.17 [keq H^+ $ha^{-1}yr^{-1}$], the monitoring load for the sensitive soil.

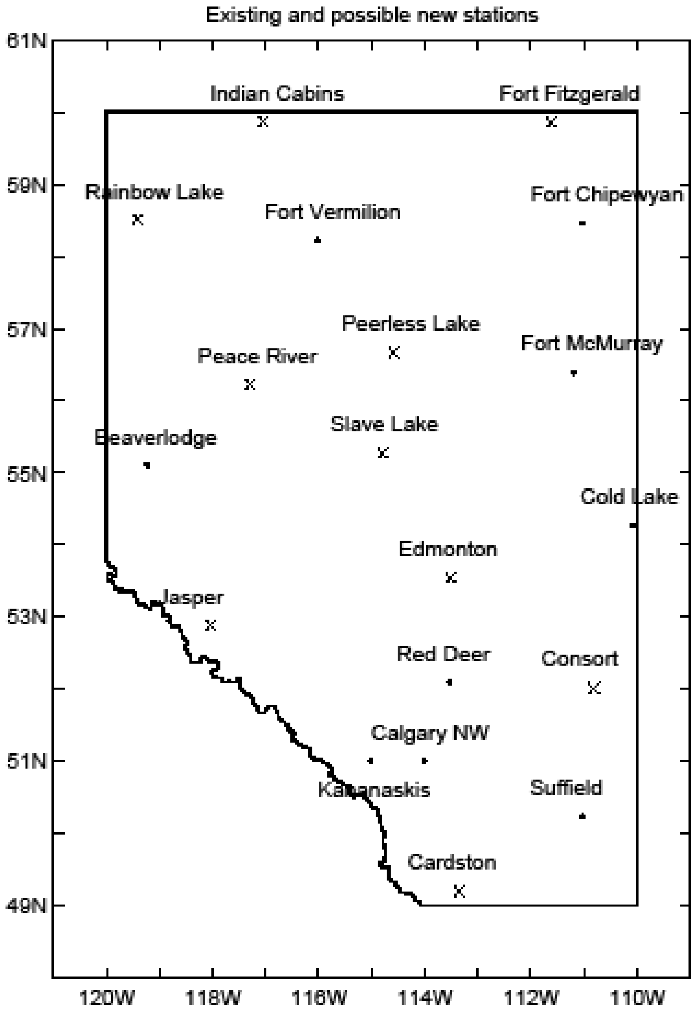


Figure S3: Locations of the additional 10 stations (denoted by crosses x) and the 9 existing stations (denoted by round dots). Rainbow Lake, Indian Cabin, and Slave Lake are not recommended to be established in the near future to monitor the PAI.

TABLE OF CONTENTS

List of Tables

List of Figures

1. Introduction

The purposes of this research are to assess the current acid deposition monitoring network, to rank the importance of the current network's nine stations, and to identify the locations of future optimal monitoring stations for Alberta.

The increased risk of pollution in the air, water and soil over Alberta due to the expansion of the energy industry, agricultural industry and population, as well as the secular climate change requires careful monitoring of various kinds of chemicals. Acid deposition is an important factor for environmental monitoring because of its adverse effects on lakes, rivers, farmland's soil quality, the ecosystem, the tourist industry, and human health. The acidic pollutants can come from various sources, including auto-tail gas, refinery emissions, power station emissions, and other fossil fuel burning or chemical plants. The pollutants may be dispersed from the sources to a vast area covering tens even hundreds of kilometres in the surrounding area, and may be brought down to the earth's surface by rain and snow (wet deposition) or by the pollutant particle deposited from the air (dry deposition) (Metcalfe and Whyatt, 1995).

Alberta Environment has been very diligent in the effort of monitoring the acid deposition and has organized various kinds of studies on the deposition. See Alberta Environment (2006) and references therein. The nationwide acid deposition is also monitored by Environment Canada. The agency has conducted much research on acid deposition (Environment Canada, 2005). Alberta Environment now maintains nine acid deposition monitoring stations that measure potential acid input (PAI). The wet deposition PAI is defined by

$$PAI_{wet} = 2\frac{[SO_4^{2-}]}{96} + \frac{[NO_3^-]}{62} + \frac{[NH_4^+]}{18} - \left(\frac{[K^+]}{39} + \frac{[Na^+]}{23} + 2\frac{[Ca^{2+}]}{40} + 2\frac{[Mg^{2+}]}{24}\right)[keqH^+/ha/yr]$$

and is monitored by nine long-term stations. The dry acid deposition PAI is defined by another formula

$$\begin{aligned} PAI_{dry} = {} & \frac{[SO_2]}{64} + \frac{[NO_2]}{46} + \frac{[NHO_2]}{47} + \frac{[NHO_3]}{63} + 2\frac{[SO_4^{2-}]}{96} + \frac{[NO_3^-]}{62} + \frac{[NH_4^+]}{18} \\ & - \left(\frac{[K^+]}{39} + \frac{[Na^+]}{23} + 2\frac{[Ca^{2+}]}{40} + 2\frac{[Mg^{2+}]}{24}\right)[keqH^+/ha/yr] \end{aligned}$$

and is monitored by four stations. The details of the above two formulas can be found in Appendix A.

The acid pollutant emissions have been increasing in some economic sectors in Alberta. In 1990, oil sand activities released 50 tons of nitrogen oxides gas per day into the atmosphere, the number increased to 150 tons per day by 2003 and reached about 400 tons by 2006, and is continuing to increase. The increased electrical energy demand also leads to an acid deposition increase and imposes higher risks of acid rain. Aherne (2008) already showed that some of the Alberta forest area has started to experience the exceedance of the PAI monitoring load. Close monitoring of the Alberta acid deposition is thus more important than ever.

Accurate monitoring of the acid deposition is a critical procedure to maintain Alberta's superb air quality and to plan for the future reduction of acid pollutants. Experiences in Europe and the US are useful in Alberta's acid deposition monitoring and reduction. Lynch et al. (2000) found reduced acid rain in the eastern US in 1995-1997 compared to 1986-1997. Marin et al. (2001) found reduced acidity in precipitation in terms of pH value increase from 1986-1997 in most parts of Europe except France. Weyhenmeyer (2008) reported the decrease of acidity in both precipitation and lake water over Sweden between 1984 and1996. To ensure Alberta's fine air quality and the quality of soil and water, the expansion of the economy and population call for an assessment of the current monitoring network and a design for the future monitoring network with additional stations. Statistical methods are routinely used in the acid deposition studies (Hales, 1982). The multivariate statistical method named empirical orthogonal function (EOF) analysis, also called principal component analysis (PCA), is used for the assessment and design. The EOFs can effectively take spatial inhomogeneity into account and have been increasingly popular among the climate and environment research community. Yu and Yu (2004) used the method for studying air quality over Taiwan. Here, we use the optimal spectral averaging (OSA) method of Shen et al. (1998) and EOF analysis to estimate the sampling errors of the PAI field, and we assess and design the monitoring network by using these errors.

We considered the total annual PAI which is the sum of the wet and dry depositions. The 1993-2006 annual PAI data were used. Using the EOF method, the nine stations were ranked according to their importance of sampling the PAI. We then selected the sites of ten possible future additional stations, ranked their importance, and designed future monitoring networks.

This report is arranged as follows. Section 2 describes the PAI data. Section 3 explains PAI data interpolation and display. Section 4 contains the correlation and covariance analysis of the PAI data. Section 5 includes the OSA method, station ranking results and the designs of the future monitoring networks. Conclusions and discussions are in Section 6.

2. PAI dataset and its quality assessment

In this study we focus on the total acid deposition measured in terms of PAI. The PAI is an indicator of acid deposition for surface soil and its units are [keq H^+ $ha^{-1}yr^{-1}$]. Alberta Environment is responsible for setting the standard of the PAI index for the Alberta soil and for monitoring the PAI index by a network of stations. The current monitoring standard for the PAI levels are 0.17, 0.35, and 0.7 [keq H^+ $ha^{-1}yr^{-1}$] for sensitive, moderately sensitive, and low sensitive soils (Alberta Environment, 2006). The sensitivity levels are determined depending on the usage of the land, the land types, and the ecosystem over the land. Foster et al. (2001) described the calculation method and the applications of critical, target, and monitoring loads for managing acid deposition in Alberta. The current monitoring loads are based on the work of Foster et al. (2001). Aherne (2008) calculated the critical loads of acid deposition for the forest soil in Alberta. Van Tienhoven et al. (1995) discussed the application of the critical loads of acid deposition in South Africa.

The current Alberta network for PAI monitoring has nine stations whose locations are shown in Figure 1 and their latitude and longitude coordinates are shown in Table 1. Figure 1 shows that the network of nine stations does not cover Alberta well, leaving a vast region in central Alberta uncovered. The northwest has only one station: Fort Vermillion. The Calgary area has three stations: Calgary, Kananaskis, and Red Deer.

Figure 1: Locations of the nine Alberta PAI monitoring stations.

Table 1: Latitude and longitude coordinates of the nine Alberta PAI monitoring stations.

Station Name	Latitude	Longitude	Elevation[m]
Beaverlodge	N55°12′	W119°24′	762
Calgary Northwest	N51°01′	W114°01′	1028
Cold Lake	N54°28′	W110°08′	518
Fort Chipewyan	N58°47′	W111°01′	213
Fort McMurray	N56°40′	W111°20′	370
Fort Vermilion	N58°23′	W116°02′	900
Kananaskis	N51°02′	W115°02′	1403
Red Deer	N52°11′	W113°53′	900
Suffield	N50°24′	W111°02′	778

The high concentration of stations in this region is due to the high PAI values in the region. The Red Deer-Calgary-Kananaskis corridor has consistently high PAI values due to power plants in the region, refineries, heavy traffic on Alberta Highway 2 and Canada Highway 1, and the industries around the Calgary-Red Deer areas. The sparse station distribution in the north is apparently due to the close-to-zero or negative PAI values in the region. See Figure 4 for the 1993-2006 mean PAI field and the figures in Appendix B to see the annual PAI distribution.

The wet PAI is measured from precipitation samples, including both rain and snow. The standard operating procedures of taking precipitation samples for acid deposition are described in Sweet (1993).

The PAI data were obtained from two sources. One is from the Air Evaluation Team of Alberta Environment and the other is the online CASA (Clean Air Strategic Alliance) data warehouse www.casadata.org. We denote the former as ABEN data, and the latter naturally as CASA data. We have identified some inconsistencies between the CASA data and an earlier version of ABEN data. The CASA data was incomplete and missed 2004 data. This motivated us to look for a reliable PAI dataset. We worked closely with the Air Evaluation Team of Alberta Environment to identify a high quality dataset. Finally, a complete dataset spanning the years from 1993 to 2006 was obtained and we used this data to obtain the results to be included in this report.

The annual PAI data of a station is set to be the sum of the multiple PAI observations throughout the year from January 1 to December 31. For example, Suffield station had 16 PAI observations in the year 2004 (Table 2). The annual total 0.0406 [keq H^+ $ha^{-1}yr^{-1}$] is the sum of these 16 measurements. Note that some of the PAI values are negative, which indicates higher alkaline level in the precipitation, according to the PAI calculation formula in Appendix A. The positive PAI values are the results of high acid levels due to sulfuric acid, nitric acid, and ammonium.

Following this example of calculating the annual total, the annual data of the 9 stations are obtained (Table 3).

Table 2: Suffield station's 2004 PAI observations.

Station	Year	Sample Start Date	Sample End Date	Sample Period (days)	PAI [keq H^+ $ha^{-1}yr^{-1}$]
SUFF	2004	18-Jan-04	5-Feb-04	18	0.0019
SUFF	2004	5-Feb-04	8-Mar-04	32	0.0021
SUFF	2004	8-Mar-04	26-Mar-04	18	0.0014
SUFF	2004	26-Mar-04	30-Apr-04	35	-0.0011
SUFF	2004	3-Jun-04	11-Jun-04	8	0.0037
SUFF	2004	11-Jun-04	18-Jun-04	7	0.0046
SUFF	2004	18-Jun-04	25-Jun-04	7	0.0005
SUFF	2004	25-Jun-04	9-Jul-04	14	0.0105
SUFF	2004	9-Jul-04	30-Jul-04	21	0.0067
SUFF	2004	1-Aug-04	16-Aug-04	15	0.0076
SUFF	2004	16-Aug-04	27-Aug-04	11	0.0054
SUFF	2004	15-Sep-04	24-Sep-04	9	-0.0001
SUFF	2004	14-Oct-04	22-Oct-04	8	-0.0009
SUFF	2004	19-Nov-04	26-Nov-04	7	0.0015
SUFF	2004	1-Dec-04	21-Dec-04	20	-0.0031
SUFF	2004	21-Dec-04	21-Jan-05	31	0.0000
				Annual Total	0.0407

Table 3: Annual PAI data from Alberta Environment [keq H^+ $ha^{-1}yr^{-1}$].

Year	Cold Lake	Fort Chipe-wyan	Calgary	Kana-naskis	Beaver-lodge	Fort Vermilion	Red Deer	Suffield	Fort Mc-Murray
1993	0.0575	-0.0020	0.1138	0.1872	0.0497	0.0195	0.1627	0.0969	0.0779
1994	0.0660	-0.0140	0.0706	0.0286	0.0871	0.0276	0.1148	0.0437	0.0581
1995	0.0470	0.0104	0.0664	0.1196	0.0749	-0.0031	0.1432	0.0572	0.1039
1996	0.0881	0.0158	0.1046	0.0983	0.0602	9999	0.1240	0.0510	0.0922
1997	0.0740	-0.0007	0.0851	0.0726	0.0452	0.0141	0.0736	0.0338	0.0289
1998	0.0578	-0.0004	0.1966	0.1578	0.0778	-0.0127	0.1767	0.1118	0.0244
1999	0.0592	0.0115	0.1591	0.0709	0.0259	0.0065	9999	0.0401	0.0577
2000	0.0638	0.0061	0.1084	0.0261	0.0558	0.0086	0.1104	0.0489	0.0394
2001	0.0963	0.0069	0.0674	0.0417	0.0567	-0.1045	0.1064	0.0629	0.0726
2002	0.1128	0.0223	0.1526	0.1735	0.0619	-0.1279	0.1375	0.0757	0.0867
2003	0.0908	0.0141	0.0336	0.0783	0.0693	0.0072	0.1404	0.0318	0.0646
2004	0.0891	0.0119	0.1297	0.1097	0.0430	0.0246	0.1623	0.0407	0.0336
2005	0.0650	-0.0035	0.0729	0.0595	0.0437	0.0060	0.1533	0.0173	-0.0010
2006	0.0592	0.0099	0.0392	-0.0140	0.0249	0.0275	0.1380	0.0922	-0.0037
Mean	0.0733	0.0063	0.1000	0.0864	0.0554	-0.0082	0.1341	0.0574	0.0525
StDev	0.0189	0.0095	0.0471	0.0588	0.0183	0.0481	0.0277	0.0274	0.0334

The maximum PAI value observed during the period of 1993-2006 is 0.20 [keq H^+ $ha^{-1}yr^{-1}$] at Calgary in 1998, which is higher than the government-defined monitoring load 0.17 [keq H^+ $ha^{-1}yr^{-1}$] for the sensitive soil. There are still three other observations in the period with PAI values higher than the monitoring load 0.17 [keq H^+ $ha^{-1}yr^{-1}$] for sensitive soil. These are 0.19 at Kananaskis in 1993, 0.18 at Red Deer in 1998, and 0.17 at Kananaskis in 2002. These three stations are less than 250 km away from each other and are all located along the busiest part of the Alberta Highway #2 that links the two biggest cities of Alberta: Calgary and Edmonton. Each of these has about 1 million people while the whole Alberta has approximately 3 million people. The high PAI values may be attributed to the power plants, auto emissions and other industrial emissions. It is also noticed that the high PAI values at these three locations did not occur at the same year, instead they occurred in three different years: 1993 for Kananaskis, 1998 for Calgary and Red Deer, and 2002 for Kananaskis. This may be related to the output of the power plants and other industrial activities. This non-uniform nature, i.e., inhomogeneity, imposes more difficulty for monitoring.

The last two rows of Table 3 show the mean and standard deviations of the PAI values between 1993 and 2006, with two missing data for Fort Vermilion in 1995 and Red Deer in 1999. The top three PAI mean values are at Red Deer (0.1341 [keq H^+ $ha^{-1}yr^{-1}$]), Calgary (0.1 [keq H^+ $ha^{-1}yr^{-1}$]), and Kananaskis (0.0864 [keq H^+ $ha^{-1}yr^{-1}$]). These are consistent with the individual PAI maxima observed at these three locations aforementioned. It is not surprising that the mean PAI values for the two north most stations are close to zero (-0.0082 [keq H^+ $ha^{-1}yr^{-1}$] for Fort Vermilion at N$58^\circ 22' 58''$ and 0.0063 [keq H^+ $ha^{-1}yr^{-1}$] for Fort Chipewyan at N$58^\circ 46' 36''$). As expected, there is little acid pollution in northern Alberta. Of course, the recent rapid expansion of oil sands exploration in the north such as at Fort McMurray will inevitably alter this situation. Close monitoring of the PAI near the facilities of the oil sands field should be arranged.

It is possible that the alkaline terms in the large bracket in the PAI formula mentioned in the introduction are larger than the sum of the first three acid terms, and hence the PAI value are negative. These negative values were observed in northern stations including Fort Chipewyan, Fort Vermilion, and Fort McMurray. Kananaskis also observed a negative value in 2006: -0.014 [keq H^+ $ha^{-1}yr^{-1}$]. This value seems too big to be reasonable and appears only once in the 14 years period. Further investigation on the accuracy of this datum needs to be made.

The CASA dataset is incomplete, particularly in 2005 and 2006. Within these two years, only Fort Chipewyan station had data, but only 3 observations in 2005 and 2 observations in 2006.

We checked the Alberta Environment dataset ABEN PAI for consistency. We focused on two aspects. First, is there a PAI measurement for every precipitation event? It is understood that some minor events of precipitation of less than 5 mm may not warrant a PAI measurement, but an event of 20 mm or more should warrant a PAI measurement. Second, does every PAI observation have a precipitation event? If not, the data must be either the dry deposition (i.e., the deposition measured from atmospheric deposition) or erroneous. See Appendix A for the definitions of the wet and dry PAI.

The US National Climatic Data Center's GHCN-D (global historical climatology network-daily) dataset was used to make the check. In most cases there was no GHCN-D data for the exact location of the PAI stations. In these cases a suitable GHCN-D station was chosen based on two criteria. First, the station is reasonably close to the PAI station. Second, the GHCN-D station has long and nearly-complete precipitation data from 1993 to 2006. Based upon these two search criteria we were able to select six GHCN-D stations for comparison.

Table 4 shows the names of these six stations as well as the distance from the PAI station in which the comparison was made. The accompanying Excel file to this report shows the details of these comparisons. In the file we use the following notations. 555 is just a continuation of the data from a start date to an end date. 9999 represents missing data (if a station was broken or had missing data for a time period, 9999 is set for each day in that time period.) 6666 represents a date when there was precipitation but no PAI data. This is a flag of possible error, particularly when the precipitation event is severe. 1111 and 2222 are just markers placed in the program to be able to note when a measurement time period began and ended. 1111 is a start date and 2222 is the corresponding end date. Sometimes a data period would start on the same day as another ended. When this has occurred, we see a string of 1111's.

Table 4: The 6 ABEN PAI stations and their corresponding GHCN-D precipitation stations.

Precipitation Station	**Long.**	**Lat.**	**Elev [m]**	**PAI Station**	**Long.**	**Lat.**	**Elev [m]**	**Distance [km]**
Red Deer	52.18	-113.90	905	Red Deer	52.18	-113.88	900	1.225
Beaverlodge CDA	55.2	-119.40	745	Beaverlodge	55.20	-119.39	762	0.490
Calgary Int'l Air*	51.12	-114.02	1084	Calgary NW	51.01	-114.02	1028	12.298
Cold Lake A*	54.42	-110.28	541	Cold Lake	54.46	-110.13	518	10.843
Fort Vermilion CDA#	58.38	-116.03	279	Fort Vermilion	58.38	-116.04	900	0.562
Fort McMurray A*	56.65	-111.22	369	Fort McMurray	56.66	-111.33	370	6.786
Kananaskis	51.03	-115.03	1391	Kananaskis	51.03	-115.03	1403	0.496

Big Difference in Elevation
* Couldn't find a closer station with complete data

We have identified all the days which have PAI data but no precipitation. For example, the Beaverlodge station has the following such days (Table 5).

Table 5: Beaverlodge station dates with no precipitation data but with PAI data.

Start Date			End Date		
1999	12	27	2000	1	3
2000	1	3	2000	1	10
2000	1	10	2000	1	17
2000	2	7	2000	2	14
2000	2	21	2000	2	28
2000	3	6	2000	3	13
2000	3	13	2000	3	20
2000	3	20	2000	3	27
2000	3	27	2000	4	3
2000	4	3	2000	4	10
2000	4	10	2000	4	17
2005	2	28	2005	3	7
2005	3	7	2005	3	14
2005	3	14	2005	3	22
2005	3	22	2005	3	28

During these periods, three possibilities may occur. First, the dry deposition was measured. The sum of the dry deposition and the wet deposition should be used as the acid deposition level for environment assessment. Thus, the dry deposition data is very useful in explaining this kind of data. According to Alberta Environment (2006), Beaverlodge station measured dry acid deposition since 2000. Second, it is a data error of wet deposition measurement. Third, it is uncertain if it is an error of wet deposition due to the spatial discontinuity of rain. In this case, errors may not be certain because the locations of the GHCN-D station and the PAI station are not exactly the same, even though they are less than a kilometer apart. Thus, it is worthwhile to search for precipitation stations even closer to the PAI stations to validate the certainty of the error.

3. Interpolation and display of the annual PAI data

We used two sets of grids for our calculation and data display. One is the 1-by-1 degree grid based at the northwest corner (120^{o}W, 60^{o}N), denoted by G1, and the other is the centres of the grid boxes, denoted by G2. Figure 2's vertices of grid boxes correspond to G1 and the dots correspond to G2. G2 may be regarded as a representation of grid boxes. Thus, although two of the G2 points are outside of Alberta, they are still counted because their corresponding grid boxes partly cover Alberta. We want to ensure that the entire Alberta is covered. For Alberta, G1 has 120 grid points and G2 has 98 grid points. G1 is mainly used for graphics display, and G2 is used for statistical calculations, such as the calculation of a covariance matrix with an area-factor.

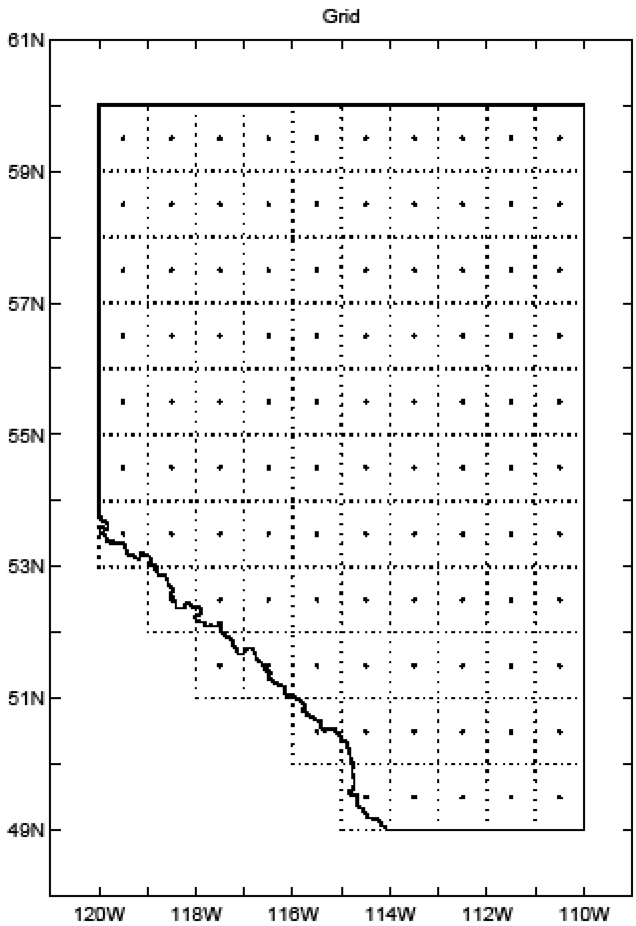


Figure 2: Grid G1 for display has 120 grid points starting from the northwest corner (120^{o}W, 60^{o}N), and grid G2 denotes the centres of the 98 grid boxes that cover the entire Alberta.

The nine stations are irregularly distributed over Alberta. In order to display the data and observe the provincial wide PAI distributions for every year, the station data are interpolated to a 1-by-1 degree grid (namely G1) by the nearest station assignment method (NSA) (Shen et al., 2001). The NSA results are spatially discontinuous. A 5-point moving-average smoothing is applied to the gridded data, as shown below

$$\begin{matrix} 1/5 & 0 & 1/5 \\ 0 & 1/5 & 0 \\ 1/5 & 0 & 1/5 \end{matrix}$$

Namely, the central point and the four corner points are used to obtain the smoothed value of the central point. Each of the five gridded data is weighted 1/5. The original NSA interpolation was made for a domain larger than Alberta so that the boundary points can get the smoothed values. The smoothing was performed using the Matlab built-in function ‘filter2’.

The values outside of Alberta were then deleted using the Matlab function ‘inpolygon’. The gridded data was plotted with the Matlab command ‘contourf’ which creates a filled contour plot of the gridded data. The ‘contourf’ function also interpolates the data, which results in the smoothed data displayed in Figure 3 and those in Appendix B.

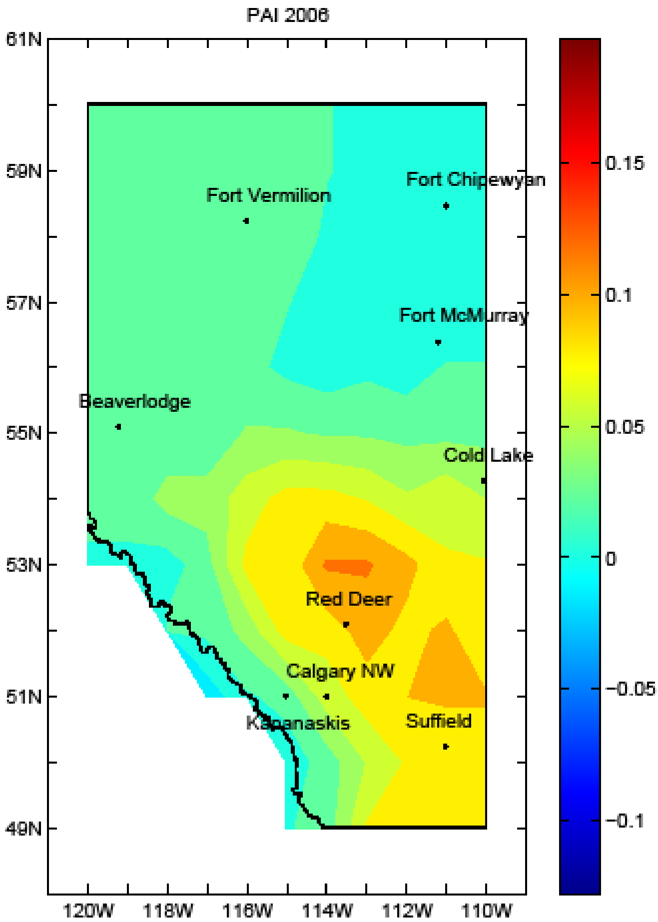


Figure 3: PAI field for Alberta in 2006.

For covariance and EOF analysis, the gridded data was mapped from the grid G1 to the centres of the grid boxes G2. The value for each grid point in G2 is the average of the four surrounding G1 grid points. Thus, the EOF analysis is for the G2 grid of 98 points.

Figure 3 shows the 2006 smoothed gridded PAI data. In 2006, Red Deer observed the largest PAI 0.1380 [keq H^+ $ha^{-1}yr^{-1}$], and Kananaskis and Fort McMurray observed negative PAI values (-0.0140 [keq H^+ $ha^{-1}yr^{-1}$] and -0.0037 [keq H^+ $ha^{-1}yr^{-1}$], respectively). These maximum positive PAI values and the negative one are clearly depicted in Figure 3. Kananaskis' negative value results in high PAI values over a smaller region centralized around Red Deer, rather than the larger corridor area for high values shown in the mean PAI field (Figure 4).

Figure 4 shows the spatial distribution of the 14-year mean of the PAI data over Alberta. The figure shows the high PAI values along the corridor of Red Deer, Calgary and Kananaskis. Most years the PAI follows this pattern. See the complete display of the PAI data in Appendix B with 14 panels ranging from 1993 to 2006. With the recent rapid increase of oil sands exploration activities, rapid increase of population in the north, and additional refineries near Edmonton since 2006, this pattern may be altered. The 2007 and 2008 data at Fort McMurray will be very useful in investigating this possible change.

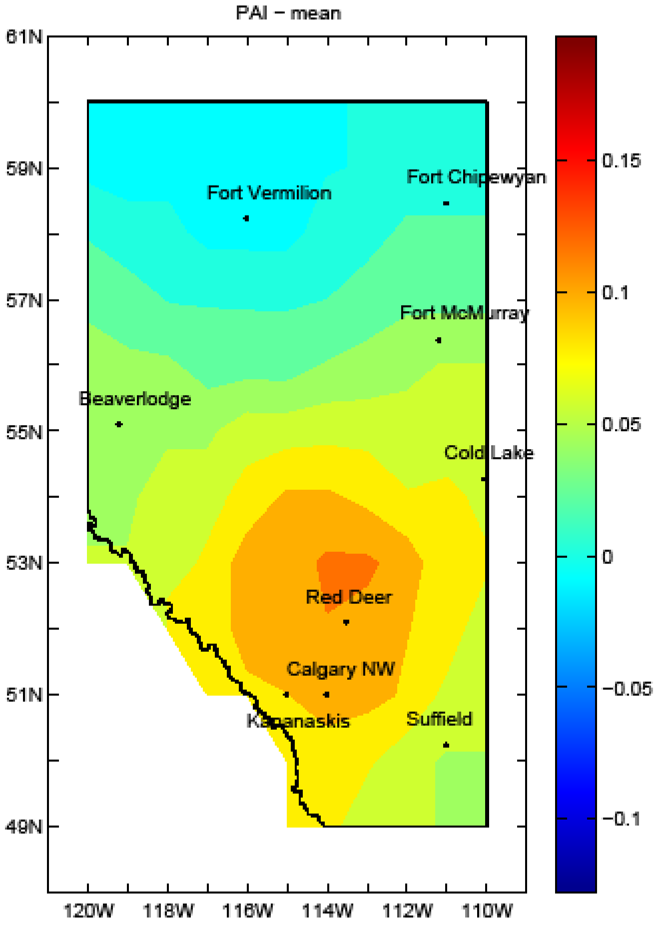


Figure 4: The 1993-2006 mean PAI field over Alberta.

4. Correlation and covariance properties and EOFs

4.1. Correlation and covariance of the station data

The distribution of the existing 9 stations is shown in Figures 1, 3, and 4. Although the annual PAI field has a large spatial scale as indicated by RELAD model (Regional Lagrangian Acid Deposition) (Alberta Environment, 2006; Cheng et al., 1995), the station PAI data are not highly correlated among the 9 stations. See Table 6 for the inter-station correlations. Most correlations fall below 0.3. The main reason for these correlation results is the sparseness of the station distribution. Other reasons may also exist, including that the observational error is large, the data streams are short (less than 30 years), and the acid deposition has strong inter-annual variability. Table 7 shows the covariance of the PAI station data. The standard deviation of each station is the square root of the diagonal elements, which range from 0.01 [keq H^+ $ha^{-1}yr^{-1}$] for Fort Chipewyan to 0.06 [keq H^+ $ha^{-1}yr^{-1}$] for Kananaskis (See Table 3). The large variance of the Kananaskis station PAI may be due to its large measurement error. Another station with large variance is Fort Vermilion. This northern station supposedly observes little acid deposition. However, some large negative values yield large variance (See Table 3). Calgary has a high variance due to its large and fluctuating positive PAI values. The relatively smaller variance of the PAI at Red Deer is because of its consistently large PAI values.

Table 6: Correlation matrix of the annual station PAI data.

	Cold La	Fort Ch	Calgary	Kana	Beav	Fort Ver	Red Deer	Suffield	FortMcM
Cold La	1.00	0.53	0.02	0.14	0.07	-0.64	-0.25	-0.17	0.27
Fort Ch	0.53	1.00	0.12	0.20	-0.25	-0.41	0.13	0.04	0.37
Calgary	0.02	0.12	1.00	0.61	0.01	-0.21	0.45	0.38	0.05
Kana	0.14	0.20	0.61	1.00	0.28	-0.28	0.49	0.40	0.47
Beav	0.07	-0.25	0.01	0.28	1.00	-0.17	-0.07	0.10	0.46
Fort Ver	-0.64	-0.41	-0.21	-0.28	-0.17	1.00	0.11	-0.22	-0.37
Red Deer	-0.25	0.13	0.45	0.49	-0.07	0.11	1.00	0.37	-0.07
Suff	-0.17	0.04	0.38	0.40	0.10	-0.22	0.37	1.00	0.05
FortMcM	0.27	0.37	0.05	0.47	0.46	-0.37	-0.07	0.05	1.00

Table 7: Covariance matrix of the annual station PAI data (covariance times 10,000).

	Cold La	Fort Ch	Calgary	Kana	Beav	Fort Ver	Red Deer	Suffield	Fort McM
Cold La	3.58	0.96	0.19	1.60	0.26	-5.80	-1.30	-0.90	1.71
Fort Ch	0.96	0.90	0.51	1.13	-0.43	-1.88	0.34	0.09	1.17
Calgary	0.19	0.51	22.21	16.78	0.09	-4.81	5.88	4.92	0.86
Kana	1.60	1.13	16.78	34.54	3.04	-7.80	7.95	6.49	9.28
Beav	0.26	-0.43	0.09	3.04	3.37	-1.47	-0.34	0.51	2.82
Fort Ver	-5.80	-1.88	-4.81	-7.80	-1.47	23.09	1.53	-2.85	-5.89
Red Deer	-1.30	0.34	5.88	7.95	-0.34	1.53	7.69	2.78	-0.61
Suffield	-0.90	0.09	4.92	6.49	0.51	-2.85	2.78	7.54	0.50
Fort McM	1.71	1.17	0.86	9.28	2.82	-5.89	-0.61	0.50	11.18

The standard deviations of some stations are quite large compared to their mean. For example, Fort Chipewyan's mean is 0.006 [keq H^+ $ha^{-1}yr^{-1}$] and its standard deviation is 0.009 [keq H^+ $ha^{-1}yr^{-1}$]. Kananaskis, Fort Vermillion, and Fort McMurray also have large standard deviations compared to their means. These data indicate large errors in the PAI observations. Thus, in our optimal weights calculation, we regard that the station observations have errors. In this research, we approximate the station error variance by ½ of the station's PAI variance.

The eigenvalues of the covariance matrix are shown in Table 8 and Figure 5. Since there are only 9 stations with 14 years of data, the covariance matrix in Table 7 is of full rank and all the 9 eigenvalues are positive. The first mode already accounts for 49% of the variance and the first five modes account for 93% of variances.

Table 8. Eigenvalues of the station covariance matrix ($(1/1{,}000)\times$ [keq H^+ $ha^{-1}yr^{-1}]^2$).

n	λ_n	$\sum_{k=1}^{n}\lambda_k$	$\sum_{k=1}^{n}\lambda_k / \sum_{k=1}^{9}\lambda_k$
1	5.54	5.54	0.49
2	2.43	7.98	0.70
3	1.42	9.39	0.82
4	0.70	10.09	0.88
5	0.52	10.61	0.93
6	0.37	10.98	0.96
7	0.25	11.23	0.98
8	0.15	11.38	1.00
9	0.03	11.41	1.00

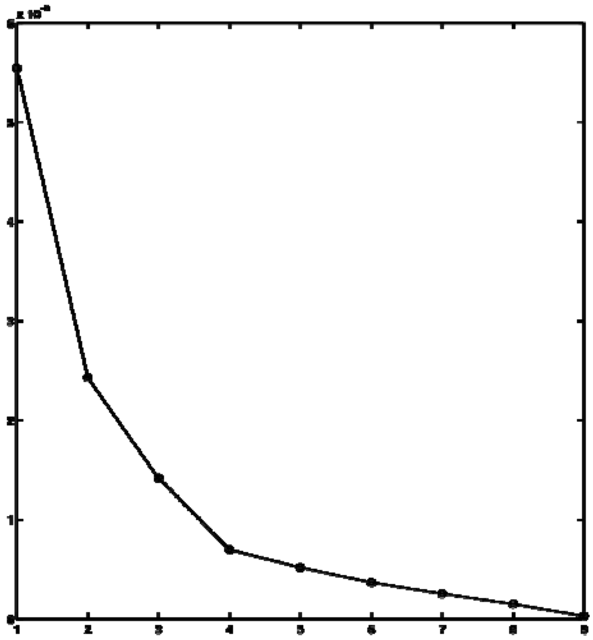


Figure 5: Eigenvalues of the station covariance matrix (units: [keq H^+ $ha^{-1}yr^{-1}]^2$).

The eigenvectors, i.e., the station EOFs, are shown in Appendix D. The first EOF is shown in Figure 6. The first mode demonstrates a dominantly large load over the Kananaskis -Calgary-Red Deer corridor which is the centre of the PAI activities in Alberta as demonstrated in Figures 3 and 4 and Figures in Appendix B. The large negative at Fort Vermilion implies the importance of the station. Hence it is not surprising that this mode accounts for almost 50%

of the total variance. The second mode, shown in Appendix D, accounts for 21% of the total variance and shows a dominantly large load at Fort Vermilion. This further enhances the importance of this station from the point of view of efficient sampling.

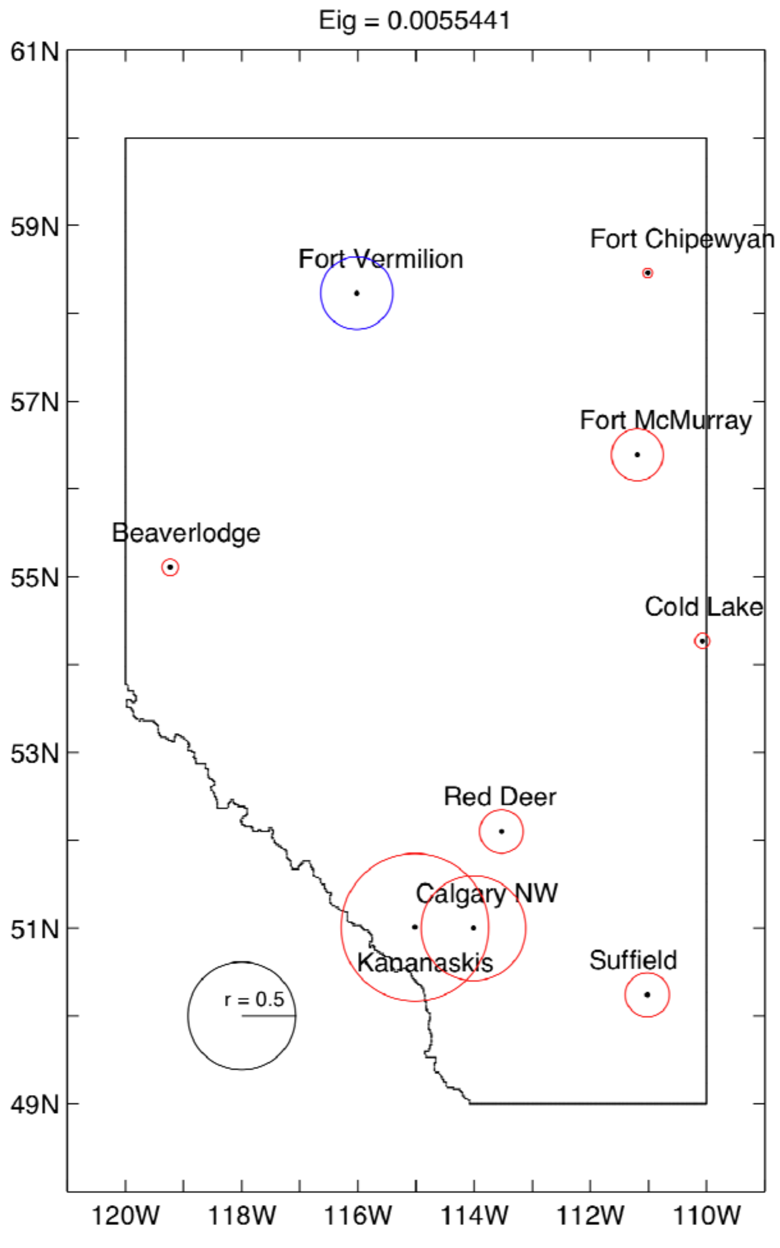


Figure 6: The first EOF mode derived from the station data. The black dots are the station locations, and the round circles denote the spatial patterns of the first empirical orthogonal function (EOF). The red circles indicate positive EOF values while the blue the negative ones. The size of the EOF value at the station location is indicated by the radius of the circle. The scale is given by the black circle which indicates 0.5.

4.2. Covariance properties of the smoothed gridded data

The smoothed gridded data on grid G2 may be regarded as the background data for calculating the covariance matrix. G2 has 98 grid points. The background data do not need to be very accurate and can be the output of a model simulation. However, the data must reflect the essential spatial patterns of the parameter under consideration. The observed data usually have this property. The eigenvalues and the eigenvectors, i.e., mode variance and EOFs, derived from the covariance matrix will be used to calculate the sampling errors, to assess the quality of an observational network, and to rank the importance of each station according to its contribution to the reduction of sampling errors.

The error formula of Shen et al. (1998) requires the use of EOFs calculated from the covariance matrix with an area-factor. With the area-factor, the EOFs may be considered as the eigenfunctions of an integral operator. Because of the inclusion of the area-factor, the units of our eigenvalues are [keq H^+ $ha^{-1}yr^{-1}]^2[km]^2$. The integral operator's kernel is approximated by the covariance matrix with an area-factor

$$C_{ij} = \frac{1}{13}\sum_{t=1993}^{2006}\sqrt{A_i}P_i(t)P_j(t)\sqrt{A_j}\ , \qquad (1)$$

where $P_i(t)$ is the PAI anomaly of grid point *i* and year *t*, and A_i is the area of the grid box associated with grid point *i* and is approximated by

$$A_i = R^2(\pi/180)^2 Cos(\phi_i). \qquad (2)$$

Here, the anomaly is with respect to the 1993-2006 mean PAI field shown in Figure 4, *i* indicates the G2 grid points and runs from 1 to 98. *R* is the radius of the Earth and is approximately 6,356 km, and ϕ_i is the latitude of the grid point *i*. $\sqrt{A_i}$ is called the area-factor.

Because there are only 14 years of data, the 98-by-98 covariance matrix $[C_{ij}]$ has a rank of at most 14. In our case here, the matrix has only 10 non-zero eigenvalues, because the PAI values at the grid points are the interpolation results from only 9 stations and hence 9 non-zero eigenvalues. The smoothing procedure leads to the 10^{th} non-zero eigenvalue, but this one is very close to zero. The eigenvalues are displayed in Table 9 and Figure 7, and the EOFs are presented in Appendix C.

Table 9. Eigenvalues of the covariance matrix with area-factor ([keq H^+ $ha^{-1}yr^{-1}]^2[km]^2$).

n	λ_n	$\sum_{k=1}^{n}\lambda_k$	$\sum_{k=1}^{n}\lambda_k / \sum_{k=1}^{10}\lambda_k$
1	398.91	398.91	0.60
2	178.08	576.99	0.87
3	45.06	622.06	0.94
4	15.54	637.60	0.96
5	14.69	652.29	0.98
6	5.00	657.29	0.99
7	3.54	660.83	1.00
8	1.29	662.12	1.00
9	1.05	663.17	1.00
10	$6.75\cdot10^{-8}$	663.17	1.00

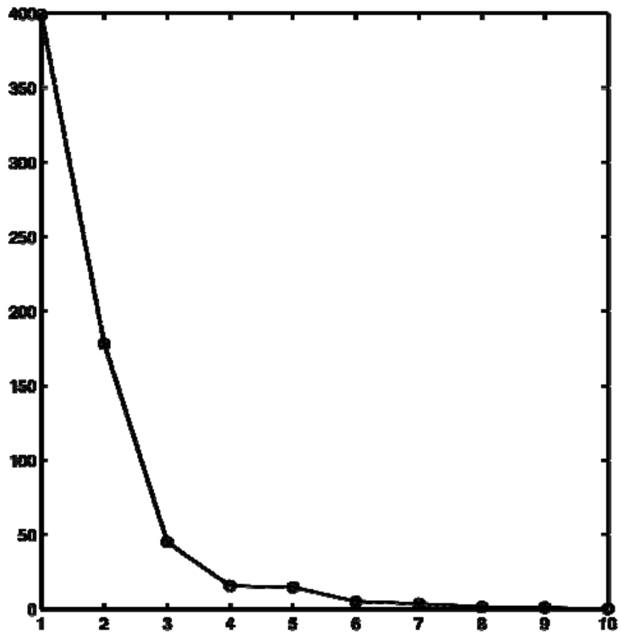


Figure 7: Eigenvalues of the covariance matrix of the smoothed and gridded PAI data.

The calculation of the covariance matrix from the data of the nine stations will be helpful to interpret the spatial patterns of the EOFs. The eigenvalues of the station covariance matrix are shown in Table 8. The eigenvectors are shown in Appendix D.

There is an approximate relationship between the eigenvalues computed from the covariance matrix with an area-factor and those without an area-factor. The eigenvalue problem with an area-factor is

$$\sum_j C_{ij} v_j^{(n)} = \sum_j \frac{1}{14} \sum_{t=1993}^{2006} \sqrt{A_i} P_i(t) P_j(t) \sqrt{A_j} v_j^{(n)} = \lambda_n v_i^{(n)}, \qquad (3)$$

while the one without an area-factor is

$$\sum_j C'_{ij} u_j^{(n)} = \sum_j \frac{1}{14} \sum_{t=1993}^{2006} P_i(t) P_j(t) u_j^{(n)} = \lambda'_n u_i^{(n)}. \qquad (4)$$

Since Alberta's latitude runs only from 49°N to 60°N and the $Cos(\phi_i)$ value runs from $Cos(49^\circ) = 0.6561$ to $Cos(60^\circ) = 0.5$, the area-factor may be regarded as uniform and is approximated by $\sqrt{A} = \sqrt{R^2 (\pi/180)^2 Cos(\phi_i)} \approx 0.7R(\pi/180) = 78$ [km]. Comparing eigenvalues in (3) and (4), we have

$$\lambda'_n \approx \lambda_n / A \approx \lambda_n / 6000 \ \ [\text{keq H}^+ \text{ ha}^{-1}\text{yr}^{-1}]^2 \qquad (5)$$

This relationship can be used to verify the calculations of the EOFs and their associated engenvalues.

5. MSE error, network assessment, and network design

The network assessment and design is made here by assessing the contribution of an individual station to the reduction of sampling error of the Alberta spatial average PAI measurement. The one that has the largest error reduction is the most important station. Based on this principle, we make two rankings. First, we rank the existing 9 stations. Second, we rank the additional 10 stations which are selected for town centres and for improved spatial coverage of Alberta.

The next step is thus to calculate the sampling errors, to make various kinds of numerical simulations to assess the quality of the acid deposition network based on the error results, and to rank the importance of future stations according to their contribution to the reduction of sampling errors.

5.1. MSE calculation method

To calculate the sampling error, we use the Shen et al. (1998) theory, which was developed to calculate the sampling errors for an inhomogeneous field. The PAI field under investigation is such a field.

The EOF $v_i^{(n)}$ in (3) has already had an area-factor in it, and its normalization property is

$$\sum_{j\in G} v_n^2(j) = 1. \qquad (6)$$

The relationship between the EOF $\psi_n(i)$ and the EOF $v_i^{(n)}$ is

$$\psi_n(j) = v_j^{(n)} / \sqrt{A_j} \;. \qquad (7)$$

The Shen et al. (1998) MSE sampling error is

$$\varepsilon^2 = \left\langle \left(\overline{P} - \hat{\overline{P}}\right)^2 \right\rangle \approx \sum_{n=1}^{M} \lambda_n \left[\overline{\psi}_n - \hat{\overline{\psi}}_n\right]^2 \;. \qquad (8)$$

Here M is the number of EOFs used, λ_n is the eigenvalue of the nth EOF mode, $\overline{\psi_n}$ is the "true" spatial average of the nth order EOF mode and is computed by

$$\overline{\psi}_n = \sum_{j\in G} \left(A_j / A_G\right) \psi_n(j) \;, \qquad (9)$$

where $\psi_n(i)$ is the value of the nth EOF mode at box i with station data, A_j is the area of the grid box $j(j \in G)$, G is the grid box network G2, A_G is the total area of Alberta as the sum of the 98 grid boxes' areas, and $\hat{\overline{\psi}}_n$ is the estimated spatial average of the n*th* EOF mode given by

$$\hat{\overline{\psi}}_n = \sum_{i\in N} w_i \psi_n(i). \qquad (10)$$

In this formula, N denotes the station network. The current acid deposition monitoring network has nine stations, thus i runs from 1 to 9. The weights can be chosen in different ways. The commonly used ones are the uniform weights, area-weights, and optimal weights.

Here we choose to use the optimal weights (Shen et al., 1998) for the purpose of optimal design. The optimization is in the sense of minimal MSE and leads to a set of *N+1* linear equations that determine the optimal weights. The linear equations are given below

$$\sum_{j \in G} \rho_{ij} w_j + \left\langle E_i^2 \right\rangle w_i + \Lambda = \overline{\rho}_i \tag{11}$$

$$\sum_{j \in G} w_j = 1 \tag{12}$$

where

$$\rho_{ij} = \sum_{m=1}^{M} \lambda_m \psi_m(i) \psi_m(j) \tag{13}$$

is the smoothed covariance function constructed from the truncated EOFs up to order M. Λ is the Lagrange multiplier and is the (*N+1*)st variable to be solved for in the above system of linear equations, and

$$\overline{\rho}_i = \sum_{m=1}^{M} \lambda_m \overline{\psi}_m \psi_m(i) \tag{14}$$

is the spatial average of the truncated covariance function. $\left\langle E_i^2 \right\rangle$ is the error variance of the station data for the *i*th station and is assumed to be half of the station's PAI variance.

The area weight is proportional to the size of the area a station represents. In general, the weight for a station over a station-sparse region is larger than that over a station-dense region.

The EOF $\psi_n(i)$ is based on a continuous covariance function and is an eigenfunction determined by an integral operator. It has the normalization property

$$\sum_{j \in G} A_j \psi_n^2(j) = 1. \tag{15}$$

Thus, we use (3) to compute the eigenvectors with area-factors, use (7) to find EOF $\psi_n(i)$, and use (11) and (12) to find the optimal weights. Finally, the formulas (6)-(8) yield the MSE sampling error of a network.

5.2. Ranking of the nine existing stations: MSE results and their implications

We calculated the minimum MSE by retaining only 8 stations, equivalently withholding one station at a time. Altogether nine MSE values are obtained for nine designs of the 8-station network. The maximum MSE indicates that without that particular station the sampling network is the worst, then that station is the most important station.

The MSE is compared with the spatially averaged total variance:

$$\sigma^2 = \frac{1}{A_G} \int_{G_1} \left\langle P^2(x,y,t) \right\rangle dxdy = \frac{1}{A_G} \sum_{m=1}^{M} \lambda_m .$$

We choose M here to be 10 since $\lambda_m = 0$ when $n > 10$ in the current case. This variance computed from our smoothed gridded data is $\sigma^2 = 0.00096$ [keq H^+ $ha^{-1}yr^{-1}]^2$, which is equal to 663.17 [keq H^+ $ha^{-1}yr^{-1}]^2$ $[km]^2$ /690,803 $[km]^2$, where $A_G = 690{,}803\,[km]^2$ is the total area of the 98 grid boxes that cover the entire Alberta, and hence is slightly larger than Alberta's actual area of 661,185 [km]^2.

The root mean square error of sampling is compared with the spatially averaged standard deviation σ (sigma). The resulted RMSE/sigma ratios and weights of all the stations are shown in Table 10.

Table 10: Weights, MSEs from 8-station network samplings, and rank of the station importance.

Withheld station	Weights									MSE*1E-5	RMSE/sigma (%)	Rank	Station name
	1	2	3	4	5	6	7	8	9				
1	0	0.18	0.05	0.04	0.33	0.04	0.16	0.11	0.08	0.50	7.2	**5**	Cold Lake
2	0.15	0	0.06	0.04	0.21	0.09	0.17	0.08	0.21	1.42	12.2	**2**	Fort Chip
3	0.16	0.19	0	0.04	0.26	0.04	0.16	0.12	0.03	0.42	6.6	**7**	Calgary
4	0.16	0.19	0.06	0	0.26	0.04	0.15	0.11	0.03	0.43	6.7	**6**	Kananaskis
5	0.32	0.14	0.05	0.04	0	0.10	0.15	0.13	0.06	2.64	16.6	**1**	Beaverlodge
6	0.16	0.19	0.05	0.04	0.28	0	0.14	0.11	0.03	0.39	6.3	**8**	Fort Verm
7	0.18	0.20	0.08	0.05	0.27	0.04	0	0.15	0.03	0.73	8.7	**4**	Red Deer
8	0.15	0.16	0.09	0.04	0.28	0.03	0.22	0	0.03	0.80	9.1	**3**	Suffield
9	0.17	0.20	0.05	0.04	0.26	0.04	0.13	0.11	0	0.29	5.5	**9**	Fort McMur
0	0.16	0.18	0.05	0.04	0.26	0.04	0.13	0.11	0.03	0.30	5.6		All stations

Our numerical results indicate that the most important station is Beaverlodge, a station located in western Alberta with no stations directly north of it. The RMSE/sigma ratio of the 8-station network of missing this station is 17%. Another indictor of this importance is the consistently large weights of the Beaverlodge station (see weight column under 5 in Table 10). These weights range between 0.21 and 0.33, while the uniform weights for 8 stations are 0.125.

The second most important station is Fort Chipewyan, and corresponding RMSE/sigma ratio is 12%. This is the north most station and represents a large area of the north eastern Alberta. The weights of this station are also relatively large, ranging from 0.14 to 0.19 with most weights around 0.18 in the 8-station network MSE calculation runs.

The third most important station is Suffield. The importance of this station is mainly due to its sole representation of the southeast Alberta. The two nearest stations are Calgary and Kananaskis to its northwest and are over 200 km away. The Suffield weights of the 8-station networks are close to the uniform weights of 0.125.

The fourth most important station is Red Deer, which is the action centre of the PAI field, including both dry and wet. The 1993-2006 mean PAI shown in Figure 4 indicates a clear high of PAI at Red Deer. The annual PAI maps in Appendix B also indicate consistently high

PAI values over Red Deer. The Red Deer station's high PAI values spread along the corridor of Red Deer-Calgary-Kananaskis. However, the variance of Red Deer's PAI is not large (only 0.0277 [keq H^+ $ha^{-1}yr^{-1}$]) compared with that of Kananaskis (0.0588 [keq H^+ $ha^{-1}yr^{-1}$]), Fort Vermillion (0.0481 [keq H^+ $ha^{-1}yr^{-1}$]), and Calgary (0.0471 [keq H^+ $ha^{-1}yr^{-1}$]). Further explanation can be found from the following description of the sixth and seventh important stations, which are Kananaskis and Calgary. The sizes of the Red Deer station's weights have little variation, ranging from 0.13 to 0.22 with most weights around 0.16.

The fifth most important station is Cold Lake. Like the Suffield station, this station represents a large area in eastern and central Alberta and is the station in the north nearest to the PAI active corridor Red Deer-Calgary-Kananaskis. This station is less important in reducing the sampling MSE than Suffield, but due to its representation of a larger area than Suffield, Cold Lake's weights are larger than those of Suffield and are around 0.16 most times with an exceptional large one of 0.32 when Beaverlodge station is withheld.

The sixth most important station is Kananaskis. This station together with Calgary (ranked seventh) and Red Deer (ranked forth) often observe high PAI values, but the Kananaskis station has an exception: negative but close to zero PAI in 2006. Kananaskis is very close to Calgary, less than 80 km apart. It is thus not surprising that Calgary station is ranked together with Kananskis. The weights of the two stations are comparable. According to the 1993-2006 PAI data shown in Table 3, the four observed PAI values that were higher than the government-defined monitoring load of 0.17 [keq H^+ $ha^{-1}yr^{-1}$] for the sensitive soil are all in this region and observed by these three stations in the following years: 1993 (0.19 [keq H^+ $ha^{-1}yr^{-1}$] at Kananaskis), 1998 (0.20 [keq H^+ $ha^{-1}yr^{-1}$] at Calgary), 0.18 [keq H^+ $ha^{-1}yr^{-1}$] at Red Deer), and 2002 (0.17 [keq H^+ $ha^{-1}yr^{-1}$] at Kananaskis). Among these three years, Red Deer observed consistent high PAIs: 1993 (0.16 [keq H^+ $ha^{-1}yr^{-1}$]), 1998 (0.18 [keq H^+ $ha^{-1}yr^{-1}$]), and 2002 (0.14 [keq H^+ $ha^{-1}yr^{-1}$]). When PAI 0.16 [keq H^+ $ha^{-1}yr^{-1}$] is very close to the monitoring load for sensitive soils, it is a warning sign that the PAI active regions may be over the threshold. For the particular year 1993, Kananaskis station observed 0.19 [keq H^+ $ha^{-1}yr^{-1}$]. Because Kananaskis station PAI has the largest standard deviation (0.0588 [keq H^+ $ha^{-1}yr^{-1}$]) and is located inside the action region, its high PAI is a very accurate indicator about when the acid deposition is over the threshold values for the sensitive, moderately sensitive, and low sensitive soils according to the PAI value. Among the three years when the PAI values are over the monitoring load for sensitive soils, Kanakaskis station consistently observed high PAI values: 1993 (0.19 [keq H^+ $ha^{-1}yr^{-1}$]), 1998 (0.16 [keq H^+ $ha^{-1}yr^{-1}$]), and 2002 (0.17 [keq H^+ $ha^{-1}yr^{-1}$]). In the years of no PAI being over the monitoring load for sensitive soils, this station consistently observed PAIs less than 0.12 [keq H^+ $ha^{-1}yr^{-1}$]. Therefore, Kananaskis station is a good indicator of possibly high PAI over the PAI action region. Nonetheless, Calgary station is not redundant because Calgary observed the higher than monitoring load PAI in 1998 which was only indicated but not observed by Kanakaskis station. Because the monitoring loads were determined based on a chemical model and not based on the real observation network configuration, the monitoring load value 0.17 [keq H^+ $ha^{-1}yr^{-1}$] for the sensitive soil may be only regarded as a guideline. In this sense, Calgary station might be regarded as redundant. Nonetheless, with 35% of the Alberta population and as the gateway to the Banff and Jasper national parks, Calgary is extremely important in maintaining high quality air, water and soil, and hence the Calgary station should remain.

The eighth important station is Fort Vermillion. Although this station represents a vast area in northwest Alberta, the PAI attaches little importance to this region. Negative PAI values are often observed at Fort Vermillion. The 1993-2006 mean PAI is also negative. The station's weights are small (around 0.04) six times during the eight station withheld tests and are around 0.1 in the other two times.

The least important station is Fort McMurray. The corresponding RMSE ratio is 5.5%. Fort McMurray is in eastern Alberta and along the similar longitude as Cold Lake and Fort Chipewyan stations, but is in the middle of the two stations. Further, for all the years between 1993 and 2006 the PAI has little variation between Cold Lake and Fort Chipewyan stations (see the PAI data display in the Appendix B). Thus Fort McMurray station appears to be redundant for the purpose of observing PAI according to the data in the period of 1993-2006. The optimal weights for Fort McMurray are small most of the time, around 0.03 for 5 times and 0.06, 0.08, and 0.21 once respectively. However, the situation may have changed due to the rapidly expanded exploration of oil sands in the area. When the 2007-2008 PAI come in, the result may be different. Thus, it is very important for a design to be adaptive and dynamic. It is also important to have the WCAS station data incorporated into the design. The RELAD modelling data can also be helpful in the assessment and design of the monitoring networks.

5.3. Ranking of additional stations

The expansion of oil exploration, increase of population, and climate change will enhance both the economic and human activities in the north. The acid deposition, reflected in the PAI values, will be likely to change its existing pattern. Thus better coverage of the PAI monitoring network will be needed. A recent study recommended to deploy more than 30 additional stations in Alberta to monitor wet deposition (Alberta Environment, 2008). Since the RMSE/sigma ratio for the current network of 9 stations is only 5.6%, a few additional stations will be sufficient to have an accurate monitoring of the Alberta PAI from the point of view of sampling error. Therefore, we have identified 10 additional stations which are based in towns or easy-to-access communities and improve the spatial coverage of the current network (see Figure 8 for the locations of both additional stations and the existing stations). The selection of these stations also takes into account population increase, power plants, agricultural industry, ecosystem, and tourism. The names and coordinates of the 10 additional stations are shown in Table 11.

Table 11: Latitude and longitude coordinates of the ten possible additional Alberta PAI monitoring stations.

Station Name	Latitude	Longitude	Elevation[m]
Fort Fitzgerald	N59°87’	W111°60’	751
Indian Cabins	N59°87’	W117°04’	297
Rainbow Lake	N58°51’	W119°40’	547
Peace River	N56°23’	W117°28’	325
Peerless Lake	N56°67’	W114°57’	695
Slave Lake	N55°28’	W114°78’	580
Jasper	N52°88’	W118°03’	1062
Edmonton	N53°55’	W113°50’	668
Consort	N52°20’	W110°80’	740
Cardston	N49°20’	W113°32’	1121

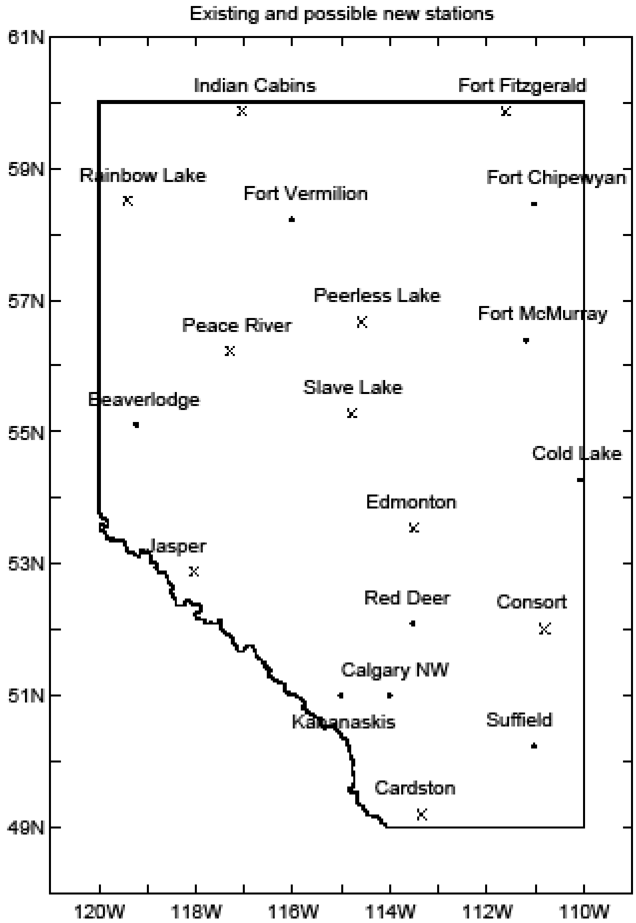


Figure 8: Additional 10 stations (denoted by crosses x) and 9 existing stations (denoted by round dots).

To rank for a station's importance, we test each additional station once at a time. First, we test 10-station networks, which are resulted from the 9 stations plus a station from the 10 additional stations. As a result we calculate the MSE for the 10-station network 10 times. The one that leads to the smallest MSE is the most important station, because this station is most effective. The ranking result is included in Table 12.

Consort station is ranked the first among the 10 additional stations. The new 9+1-station network has a RMSE/sigma ratio 5.81%. Consort is east of Red Deer by about 150 km and is on the east boundary of the PAI action region. It is between the current Cold Lake station in the north (about 250 km away) and Suffield station in the south (about 200 km away). The weight of Consort station in the new 10-station network is 0.09, which is slightly smaller than the uniform weight 0.1. The weight 0.09 is comparable to the weights of the following stations: Fort McMurray (0.09), Red Deer (0.09), Beaverlodge (0.08) and Cold Lake (0.08) (See Table 13). Although Consort is lastly added to the existing network of 9 stations, its weight is only smaller than one station (Fort Vermilion, 0.34) and is larger than six other stations.

Counting Consort as the 10th station, we next find a station from the remaining 9 additional stations to form 11-station networks following the same method. Peerless Lake station is now selected from the 9 additional stations. This station fills in central Alberta's station void. In fact, its weight is higher than six other stations. The new network of 11 stations (Consort + Peerless Lake + 9 existing stations) results in a RMSE/sigma ratio 5% (see Table 14).

Due to agroclimate change, Alberta arable land is further expanded to the north. The Peerless Lake- Peace River-Slave Lake region has now become very important for agriculture production. Organic farming has an increasing importance over the region. Close monitoring of the PAI in this area is very critical in safeguarding the soil quality and proper land use.

Table 12: The MSE for the 10-station network with an additional station on the existing 9 stations.

MSE x 10^6	RMS/sigma (in %)	Station
3.24	5.81	Consort
3.27	5.83	Cardston
3.62	6.14	Peerless Lake
4.14	6.57	Slave Lake
4.48	6.83	Peace River
4.52	6.86	Fort Fitzgerald
5.47	7.55	Indian Cabins
5.56	7.61	Edmonton
5.88	7.82	Jasper
5.89	7.83	Rainbow Lake

Table 13: Weights of the 10-station networks developed from 9+1 stations.

New Stn	Cold La	Fort Ch	Calg	Kana	Beav	Fort Ver	Red Deer	Suff	Fort McM	New Stn	Sum
Fort Fitzgerald	0.12	0.02	0.08	0.07	0.11	0.31	0.10	0.06	0.09	0.02	1.00
Indian Cabins	0.11	0.03	0.08	0.07	0.10	0.31	0.10	0.06	0.10	0.02	1.00
Rainbow Lake	0.11	0.04	0.08	0.07	0.08	0.26	0.10	0.06	0.11	0.09	1.00
Peace River	0.10	0.04	0.08	0.08	0.07	0.30	0.10	0.06	0.11	0.07	1.00
Peerless Lake	0.11	0.02	0.08	0.08	0.09	0.28	0.10	0.06	0.07	0.11	1.00
Slave Lake	0.08	0.03	0.08	0.08	0.06	0.29	0.10	0.06	0.09	0.14	1.00
Jasper	0.10	0.04	0.08	0.06	0.10	0.30	0.10	0.06	0.10	0.05	1.00
Edmonton	0.09	0.03	0.07	0.06	0.10	0.31	0.08	0.06	0.10	0.10	1.00
Consort	0.08	0.04	0.07	0.07	0.08	0.34	0.09	0.05	0.09	0.09	1.00
Cardston	0.11	0.03	0.07	0.06	0.09	0.34	0.09	0.05	0.10	0.05	1.00

Following the same procedure we have ranked the remaining 8 additional stations, and results are shown in Table 14 which has the overall ranking of the additional 10 stations.

The third important station is Cardston. It fills the void of southern Alberta. Both our smoothed observed data and the RELAD model data indicate that PAI values around this location are comparable to that of Cold Lake. The intense agriculture, Waterton Lake National Park, the Alberta's fourth largest city Lethbridge, and the fast population increase in Calgary and in the southern Alberta in general make this station very important.

Table 14: Rank of the importance of the 10 additional stations and their corresponding ratio of RMSE to the spatially averaged standard deviation of the PAI field.

Importance Rank	Station Name	RMSE/sigma (in %)
1	Consort	5.8
2	Peerless Lake	4.9
3	Cardston	4.2
4	Fort Fitzgerald	4.0
5	Edmonton	4.0
6	Jasper	4.3
7	Peace River	4.1
8	Slave Lake	4.3
9	Indian Cabins	4.5
10	Rainbow Lake	

The fourth important additional station is Fort Fitzgerald at the northeast corner of Alberta. The existing data imply that the PAI has more activities over the northeast Alberta than the northwest Alberta. Fort McMurray's recent oil sands exploration may further enhance the PAI activities over the region. However, the PAI values observed at the northeast station Fort Chipewyan have been small (maximum 0.02 [keq H^+ $ha^{-1}yr^{-1}$]), and it is not likely for the PAI to exceed the monitoring load for the sensitive soil (0.17 [keq H^+ $ha^{-1}yr^{-1}$]) anywhere north of Fort Chipewyan in the next 3 or 4 decades despite the rapid expansion of the oil sands exploration, agriculture, and population in the Fort McMurray region. It is not in the foreseeable future that population will be very large over the Fort Fitzgerald region. Thus, Fort Fitzgerald station is not a station of urgent installation, despite its rank at fourth.

The fifth important additional station is Edmonton. This station fills the void in mid-latitude mid-longitude Alberta, and is between Beaverlodge to its west and Cold Lake to its east. Edmonton is also located at the north boundary of the PAI action corridor Red Deer-Calgary-Kananaskis. As the most important engineering and logistical support base to the Alberta oil sands exploration and as the capital of Alberta, Edmonton is in its fast expansion phase including infrastructure, population, industrial facilities, and agriculture. The region is becoming very sensitive to the acid deposition. Close monitoring of the Edmonton PAI is necessary and should be implemented as soon as possible. As a matter of fact, the two WCAS stations at Genesee and Violet Grove, which started to operate from year 2000, are less than 80 km west of Edmonton and are near several power plants.

The sixth important additional station is Jasper, a rocky mountain site situated over the Jasper National Park. This area is sensitive to acid deposition and can be influenced by the future Edmonton's PAI activities. However, it is not likely that this area will be in danger of observing PAI values over 0.17 [keq H^+ $ha^{-1}yr^{-1}$]. Beaverlodge station should be a good representative of Jasper. Notwithstanding the above, if the tourist industry imposes a strict standard of PAI, then a Jasper station may become necessary for monitoring the new standard.

Peace River and Slave Lake are ranked seventh and eighth, respectively. These two stations' contribution to the sampling error reduction is very limited. Peace River station may be needed now for the needs of the organic farming industry.

The ninth station is Indian Cabins, at the northwest boarder with Northwest Territory. The tenth station is Rainbow Lake, at the northwest border with British Columbia. From the point of view of uniform distribution of stations, these two stations would be ranked more important since there is a vast station-void region in the northwest Alberta. However, the PAI activity over this northwest region is negligible, and Fort Vermilion station should be sufficient to monitor the acid deposition in the area.

Note that the addition of the 5^{th}-10^{th} stations does not decrease the sampling error. Usually, the increased number of stations yields smaller RMSE/sigma ratio. When a network has 15 stations, the further increase of the number of stations seems to not result in additional reduction of sampling errors. This is likely true since the annual PAI field has a large spatial scale as indicated by the model and observed data. However, we do not exclude this as a computational artefact due to the short stream of data (only 14 years) and few stations (9 stations) to begin with. The data scarcity makes the covariance matrix, even computed from the smoothed gridded data, have a rank of only 10, with the 10^{th} eigenvalue almost zero. A long time run (over 30 years) of a high-resolution RELAD model (at least 0.5 degrees) should be helpful in improving the covariance matrix and understanding the spatial covariance properties of the PAI field. This will lead to a more accurate design of the networks with additional stations.

Also note that when considering the ranking of the additional stations, the error variance of an additional station cannot be computed from the observed data, and is approximated by a half of the variance of the interpolated data of the grid point nearest to the station. The interpolated data are smoother than the station data and have a smaller variance. Thus, the RMSE/sigma ratio is larger than the case as if the station would have had the data. This explains why the RMSE/sigma ratios for 10 stations in Table 12 are larger than those in Table 10 for fewer than 10 stations.

6. Conclusions and discussions

The EOF method has been applied to the 14 years of Alberta wet and dry deposition PAI data collected at 9 stations to assess the current network, to rank the importance of each station in the PAI sampling, and to design the future PAI monitoring networks. Our results lead to the following conclusions.

1. Inhomogeneous PAI distribution over Alberta: The mean PAI field in the period of 1993-2006 (Figure 4) and the PAI fields of individual years (Figures in Appendix B) demonstrate the strong spatial inhomogeneity of the PAI field. The station PAI data are not highly correlated. The PAI action centre is the Red Deer-Calgary-Kananaskis

corridor. This area usually observes high PAI values compared with other areas in Alberta.

2. The current network and PAI patterns: Figure 1 shows the locations of the nine current Alberta Environment's PAI monitoring stations. The 9-station network results in a small RMSE/SD ratio (5.6%). In the period of 14 years (1993-2006), there were only three years (1993, 1998, and 2002) when the PAI values were higher than the monitoring loads of 0.17 [keq H^+ $ha^{-1}yr^{-1}$] at three locations: Red Deer, Calgary and Kananaskis (See Figure S2a, b and c). The highest PAI value is 0.20 [keq H^+ $ha^{-1}yr^{-1}$] observed at Calgary in 1998 is still less than the target load of 0.22 [keq H^+ $ha^{-1}yr^{-1}$] for sensitive soils. Thus, Alberta acid deposition has not caused major environmental problems. The spatial patterns of the PAI derived from the observations demonstrate consistent high PAI activities in the Red Deer-Calgary-Kananaskis corridor, which is consistent with the RELAD modeling results of Alberta Environment (2006). However, the model results predicted much higher PAI values, which are not supported by the observations contained in this report. The EOF analysis of gridded data shows a dominant north-south oscillation in the eastern half of Alberta in the first mode, a northeast-midwest oscillation in the second mode, and a strong Red Deer-Calgary-Kananaskis *vs*. Beaverlodge dipole in the third mode. This Red Deer-Calgary-Kananaskis pattern persists in the fourth mode.

3. Ranking of the nine stations of the current network: According to a station's contribution to the reduction of sampling error, the nine stations of the existing PAI network are ranked as follows:
    1) Beaverlodge,
    2) Fort Chipewyan,
    3) Suffield,
    4) Red Deer,
    5) Cold Lake,
    6) Kananaskis,
    7) Calgary,
    8) Fort Vermilion, and
    9) Fort McMurray.

    It is not recommended that the last two stations Fort Vermilion and Fort McMurray be eliminated from the network, because Fort Vermillion is currently the only station in the vast area of northwest Alberta and the oil sands exploration in the Fort McMurray area requires the close monitoring of the air quality and water quality risks and hence the station's continuation.

4. Ranking of the additional 10 stations for the future network: The locations of the additional 10 stations are selected under the consideration of monitoring needs, physical accessibility and improved spatial coverage (see Figure 8). These 10 stations are ranked in the following order:
    1) Consort,
    2) Peerless Lake,
    3) Cardston,
    4) Fort Fitzgerald,
    5) Edmonton,
    6) Jasper,
    7) Peace River,

8) Slave Lake,
9) Indian Cabins, and
10) Rainbow Lake.

It is recommended that the first seven station be established when conditions allow. Consort, Peerless Lake, Cardston, and Fort Fitzgerald stations are helpful in improving the spatial coverage and reducing the sampling error. Edmonton station is useful in monitoring the industrial PAI load. Jasper station is useful in monitoring the air, water and soil quality from the tourism industry. Peace River station is helpful in monitoring the expanded farmland soil for the organic farming industry. The Indian Cabins and Rainbow Lake stations are ranked 9th and 10th and are not recommended to be established in the near future. Slave Lake station, ranked 8th, may be useful for some special agricultural monitoring needs.

5. Non-uniqueness of the future networks: Because of the annual PAI's large spatial scale and hence small sampling error, it makes little difference to change the order of ranking after the first three or four stations. Thus, from the practical need of PAI's effective monitoring, the future optimal network is not unique. When the first three or four stations are chosen, the next few stations may be chosen according to the local needs, such as the need at Edmonton due to the rapid population increase or the need at Peace River due to the organic farming.

6. Ongoing effort to assess the monitoring network: When the 2007-2008 PAI come in, Fort McMurray station's importance will be likely to advance from the current least important position to a more important position due to the intensive oil sands exploration activities in 2007-2008. Since the EOF assessment and design analysis is an adaptive method, it can take various kinds of observed and modelling data. The other data to be included in the analysis are updated data from the nine stations, the data from the two West Central Airshed Society stations near Edmonton, the RELAD modelling data, and other temporary measurement of the dry PAI. Thus, it is very important for a design to be adaptive and dynamic and for the monitoring errors to be assessed every year with the updated data.

For the networks with additional stations, one can always ask whether redundancy exists among the stations in the new networks and what is the rank of the additional stations among the existing stations in the new network. We used the same method of ranking the importance of the 9 existing stations. When a network of 10 stations is formed, we deleted one station at a time and saw which station's deletion results in the largest MSE. This leads to the ranking of the 10-station network. The least important station may be considered redundant. This redundant station may not be the one from the 10 additional stations. We ran the tests for ten stations and found the Consort station, as the first additional station, is very important and is ranked third in the 10-station networks. In a similar way, we can rank the stations in the 11-station network.

The accuracy of data needs attention. The Alberta Environment (2006) report shows large concentrations of PAI in the dry deposition, particularly at the two WCAS (West Central Airshed Society) stations located at Genesee and Violet Grove. Both stations are about 100 km west of Edmonton. Within 60 km of each station, there are several power plants, which emit large amounts of acid chemicals into the air. According to the Alberta Environment (2006) report, both stations observed high PAI values (0.41 and 0.21 [keq H^+ $ha^{-1}yr^{-1}$], respectively) in 2005. The monitoring load for moderately sensitive soil is 0.35 [keq H^+ $ha^{-1}yr^{-1}$]. Thus, the Genesee station observed some acid deposition risks. The real situation might be even more serious than this because Genesee station is not right next to the power stations, rather 8-33 km away from the stations. Therefore, it is advisable to further investigate the accuracy of the data and to have a close monitoring of the PAI values near power plants, refinery facilities, and oil sands exploration sites.

In this research, the error variance of the station data for the *i*th station $\left\langle E_i^2 \right\rangle$ is assumed to be half of the station's PAI variance, which is an approximation as was done in Shen et al. (1998). The $\left\langle E_i^2 \right\rangle$ values are an important factor in determining the weights and finally the MSE. Due to this data noise, it may happen that inclusion of an extra station may sometimes result in a larger MSE/sigma ratio as shown in the last two rows of Table 10. Thus, the RMSE/sigma ratios should be regarded as a relative measure of the effectiveness of a network.

The adequacy of the data analysis method also needs attention. A geostatistics method called kriging has often been used for data interpolation in geology and mining. It was also widely used in climate sciences. The method assumes a semi-variagram which describes the spatial covariance of the field under investigation. The interpolation results and errors critically depend on the shapes of the semi-variagram. Because of the semi-varigram's fixed shape, the kriging method is limited to homogeneous fields. The EOF method used in this report is iterative and can take historical data and model data into consideration. The modern computing power make the iterative method's application possible since the EOF method requires much more computing resources than the kriging method. Climate science research has now used the EOF method more often than the kriging method. The Alberta acid deposition data were recently analyzed by the kriging method (Alberta Environment, 2008). It appears that the method has shown limitations and cannot adequately calculate the ranking of the PAI monitoring stations based on sampling error analysis.

7. References

Aherne, J. 2008. *Calculating Critical Loads of Acid Deposition for Forest Soils in Alberta: Critical Load, Exceedance and Limitation.* Final Report. Canadian Council of Ministers of the Environment, 2008, 14pp.

Alberta Environment, 2008. *Air Pollution and Deposition Monitoring Networks in Alberta: A Review and Recommendations for Potential Future Networks*. Technical Report by JN Cape. Alberta Environment.

Alberta Environment, 2006. *2004 Acid Deposition Assessment for Alberta*. Environmental Policy Branch, Alberta Environment.

Cheng, L, Angle, RP, Peake, E, and Sandhu, HS, 1995. Effective acidity modeling to establish acidic deposition objectives and manage emissions. *Atmospheric Environment* **29**, 383-392.

Environment Canada, 2005. *2004 Canadian Acid Deposition Science Assessment: Summary of Key Results.* Meteorological Service of Canada, Ottawa, Ontario.

Foster, KR, McDonald K, Eastlick, K, 2001. Development and application of critical, target and monitoring loads for the management of acid deposition in Alberta, Canada. *Water, Air, and Soil Pollution: Focus* **1**, 135-151.

Hales, J.M., 1982. The MAP3s/RAINE precipitation chemistry network: statistical overview for the period 1976-1980. *Atmospheric Environment* 16, 1603-1631.
Lynch, JA, Bowersox, VC, and Grimm, JW, 2000. Acid rain reduced in eastern United States. Environ. Sci. Technol. **34**, 940-949.
Yu, T-Y, Yu, T-K, 2004. Spatial and temporal features of ambient air-quality over Taiwan. Environ. Sci. and Pollu. Res. **11**, 3-6.
Marin, E., Perez-Amaral, A.R., Hernandez, E., 2001. The evolution of the pH in Europe (1986-1997) using panel data. Chemosphere **45**, 329-337.
Metcalfe, S, Whyatt, D, 1995. Who to blame for acid rain? A regional study of acid deposition in Yorkshire and Humberside. *Trans. Inst. Br. Geogr NS* **20**, 58-67.
Shen, SSP, Dzikowski, P, Li, G, Griffith, D, 2001. Interpolation of 1961-1997 daily climate data onto Alberta polygons of ecodistrict and soil landscape of Canada. *J. Applied Meteo*. **40**, 2162-2177.
Shen, SSP, Smith, TM, Ropelewski, CF, and Livezey, RE, 1998. An optimal regional averaging method with error estimates and a test using tropical Pacific SST data, *J. Climate* **11**, 2340-2350.
Sweet, CW, 1993. *Standard Operating Procedure for Precipitation Sampling Using XAD-2 and MIC Collectors*. Office of Air Quality, Illinois State Water Survey.
Van Tienhoven, AM, Olbrich, KA, Skoroszewski, R, Taljaard, and Zunckel, M, 1995. Application of the critical load approach in South Africa. *Water, Air and Soil Pollution* **85**, 2577-2582.
Weyhenmeyer, CA, 2008. Water chemical changes along a latitudinal gradient in relation to climate and atmospheric deposition. *Climate Change* **88**, 199-208.

Appendix A: Definition of PAI

The PAI is an integrated index calculated from the concentrations of a list of ions measured in the precipitation: NH_4^+, NO_3^-, SO_4^{2-}, Ca^{2+}, K^+, Mg^+, and Na^+ (Cheng, 2004). These concentrations are compared with the standard units of the hydrogen ions in the equilibrium condition. Thus, the PAI units are keq H^+ $ha^{-1}yr^{-1}$, where keq is the chemical reaction equilibrium constant, H^+ stands for hydrogen ions' concentration, ha is hectare, and yr is year.

The acid deposition has two sources: wet and dry. The wet deposition is contained in precipitation, and the dry deposition is contained in air.

The wet PAI is calculated by the following formula (Chen et al., 2004, 2001):

$$PAI_{wet} = 2\frac{[SO_4^{2-}]}{96} + \frac{[NO_3^-]}{62} + \frac{[NH_4^+]}{18} - \left(\frac{[K^+]}{39} + \frac{[Na^+]}{23} + 2\frac{[Ca^{2+}]}{40} + 2\frac{[Mg^{2+}]}{24}\right)[keqH^+/ha/yr]$$

The numerators are the ions' concentrations. The denominators are the corresponding atomic weights, and the coefficient in each term is the number of deficit or surplus electrons in the ions. Most times the PAI values are positive, i.e., precipitations have some acidity. But there are some cases of negative values due to the higher levels of metal ions. The wet acid deposition has been so far observed by 9 stations in Alberta.

The units [keq H^+ $ha^{-1}yr^{-1}$] are related to the deposition's pH value, which is defined by

$$pH = -\log([H+])$$

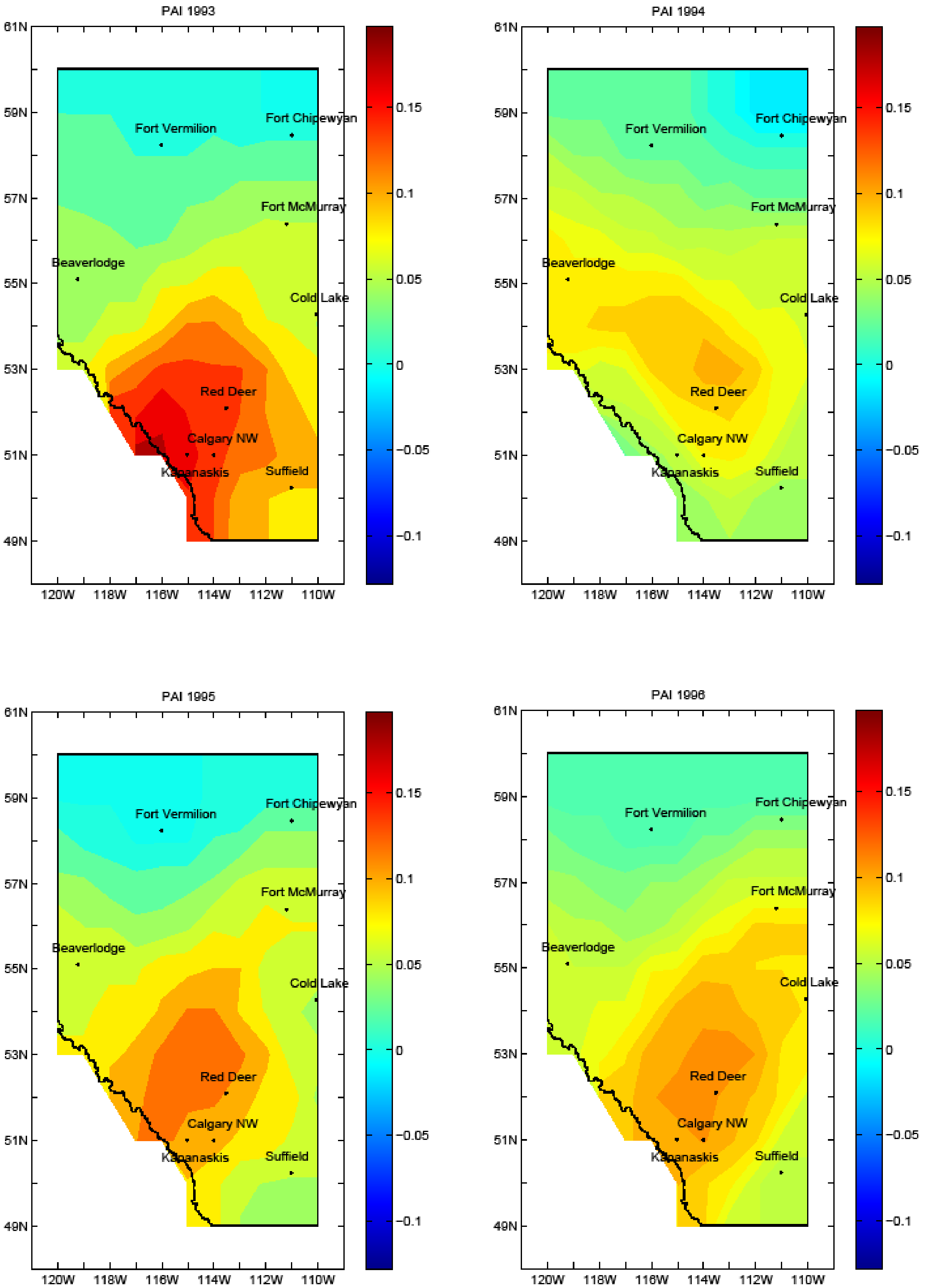
where [H+] is the hydrogen proton concentration in terms of mol/L converted by a dissociation constant *ka*.

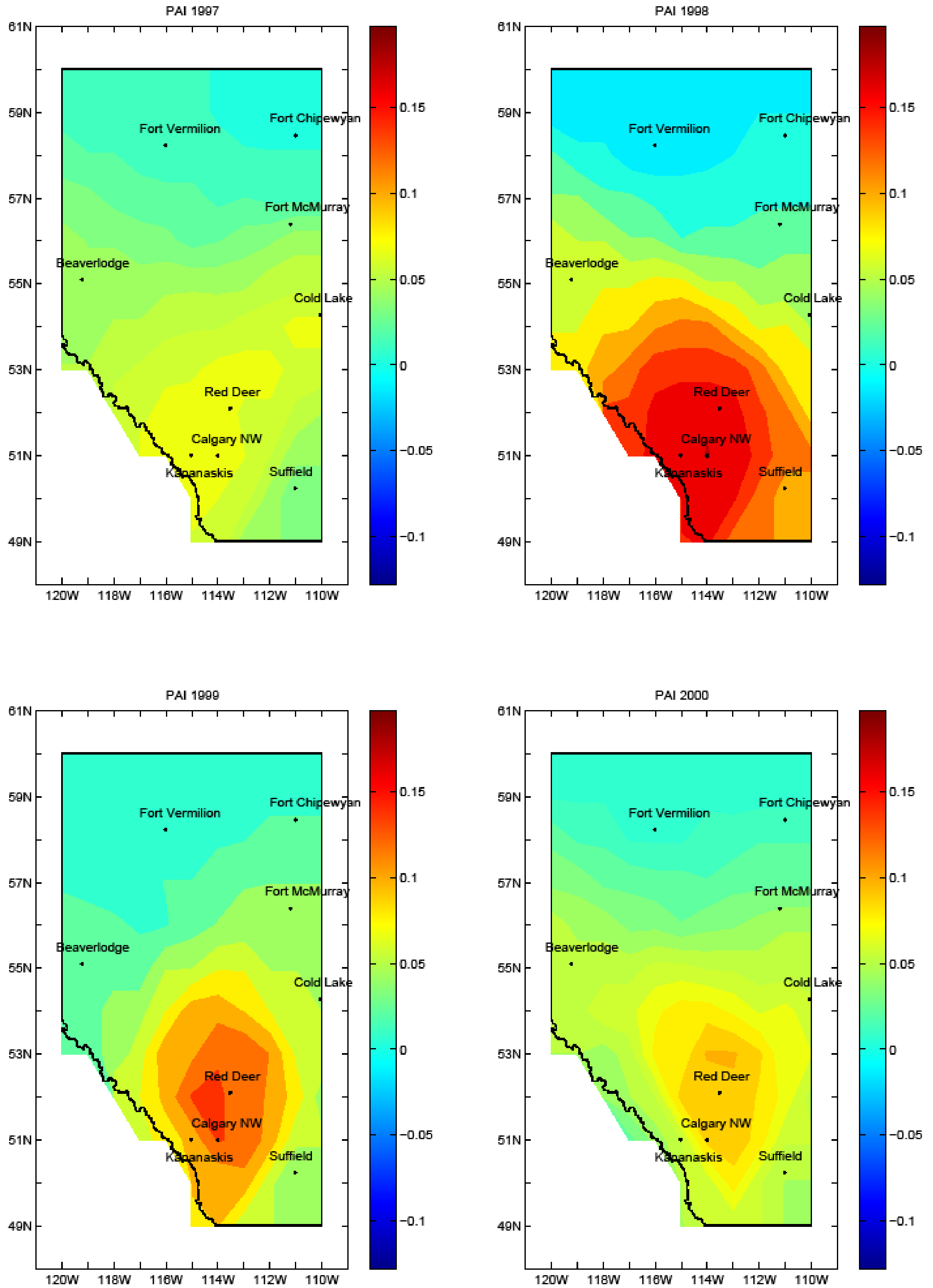
The dry acid deposition has been observed by 4 stations in Alberta. In addition to the chemicals included in the wet deposition, 4 other acidity gases are measured: SO_2, NO_2, HNO_2, and HNO_3. The dry PAI is calculated by the following formula.

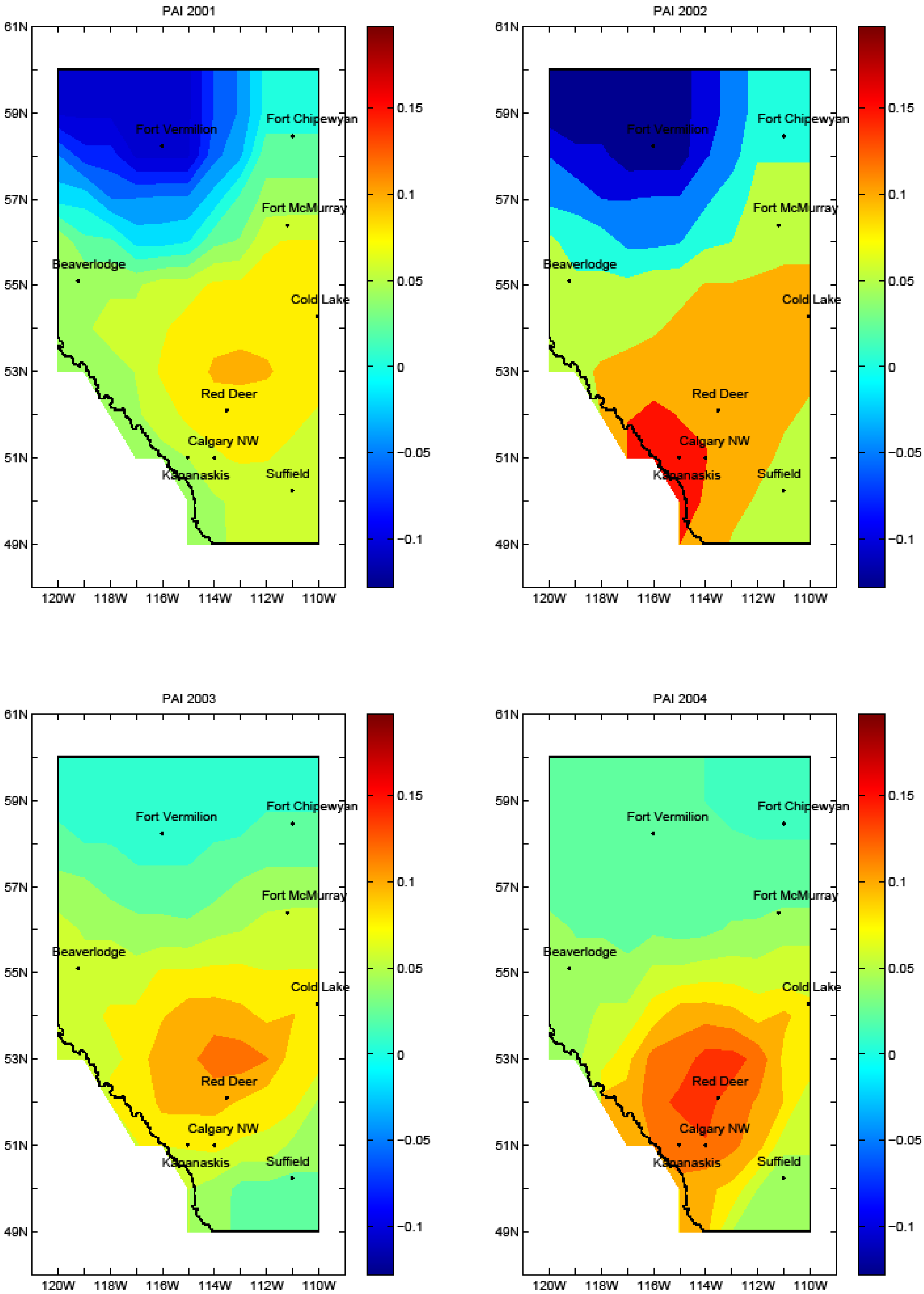
$$PAI_{dry} = \frac{[SO_2]}{64} + \frac{[NO_2]}{46} + \frac{[NHO_2]}{47} + \frac{[NHO_3]}{63} + 2\frac{[SO_4^{2-}]}{96} + \frac{[NO_3^-]}{62} + \frac{[NH_4^+]}{18}$$
$$- \left(\frac{[K^+]}{39} + \frac{[Na^+]}{23} + 2\frac{[Ca^{2+}]}{40} + 2\frac{[Mg^{2+}]}{24}\right)[keqH^+/ha/yr]$$

We will further investigate the locations of the wet and dry PAI stations and check the data validation and quality control.

Appendix B: Display of PAI data


PAI 1993
PAI 1994
PAI 1995
PAI 1996
Fort Vermilion
Fort Chipewyan
Fort McMurray
Beaverlodge
Cold Lake
Red Deer
Calgary NW
Kananaskis
Suffield
61N
59N
57N
55N
53N
51N
49N
120W
118W
116W
114W
112W
110W
0.15
0.1
0.05
0
−0.05
−0.1


PAI 1997
PAI 1998
PAI 1999
PAI 2000
Fort Vermilion
Fort Chipewyan
Fort McMurray
Beaverlodge
Cold Lake
Red Deer
Calgary NW
Kananaskis
Suffield
61N
59N
57N
55N
53N
51N
49N
120W
118W
116W
114W
112W
110W
0.15
0.1
0.05
0
−0.05
−0.1


PAI 2001
PAI 2002
PAI 2003
PAI 2004
61N
59N
57N
55N
53N
51N
49N
120W
118W
116W
114W
112W
110W
Fort Vermilion
Fort Chipewyan
Fort McMurray
Beaverlodge
Cold Lake
Red Deer
Calgary NW
Kananaskis
Suffield
0.15
0.1
0.05
0
−0.05
−0.1

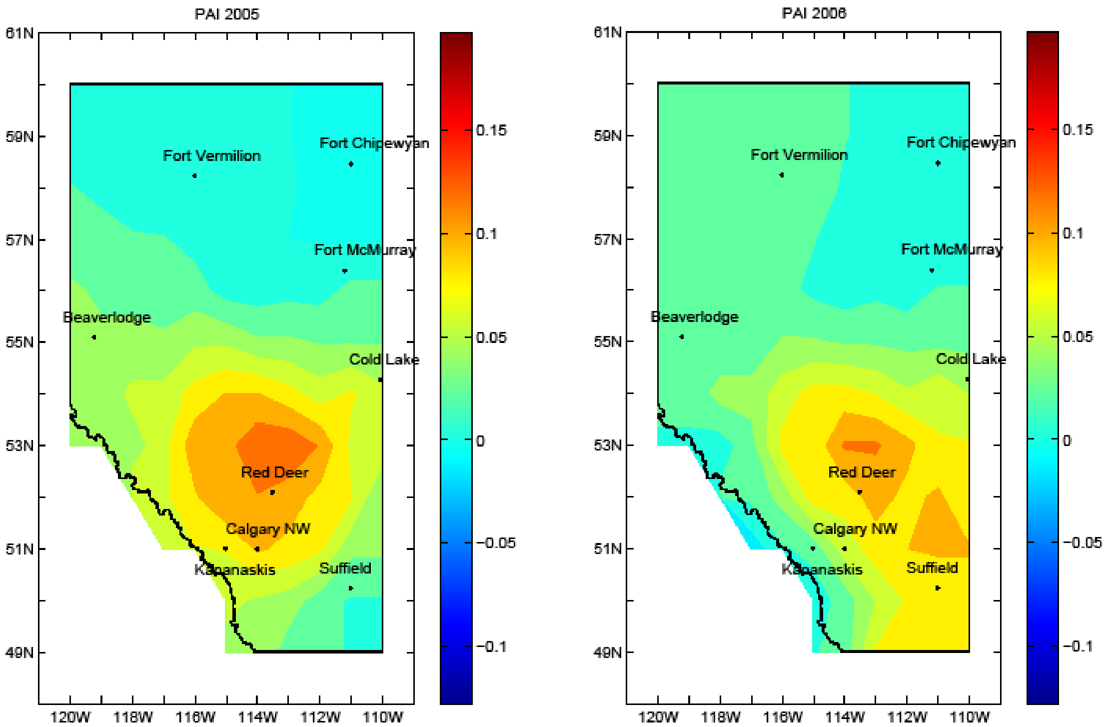
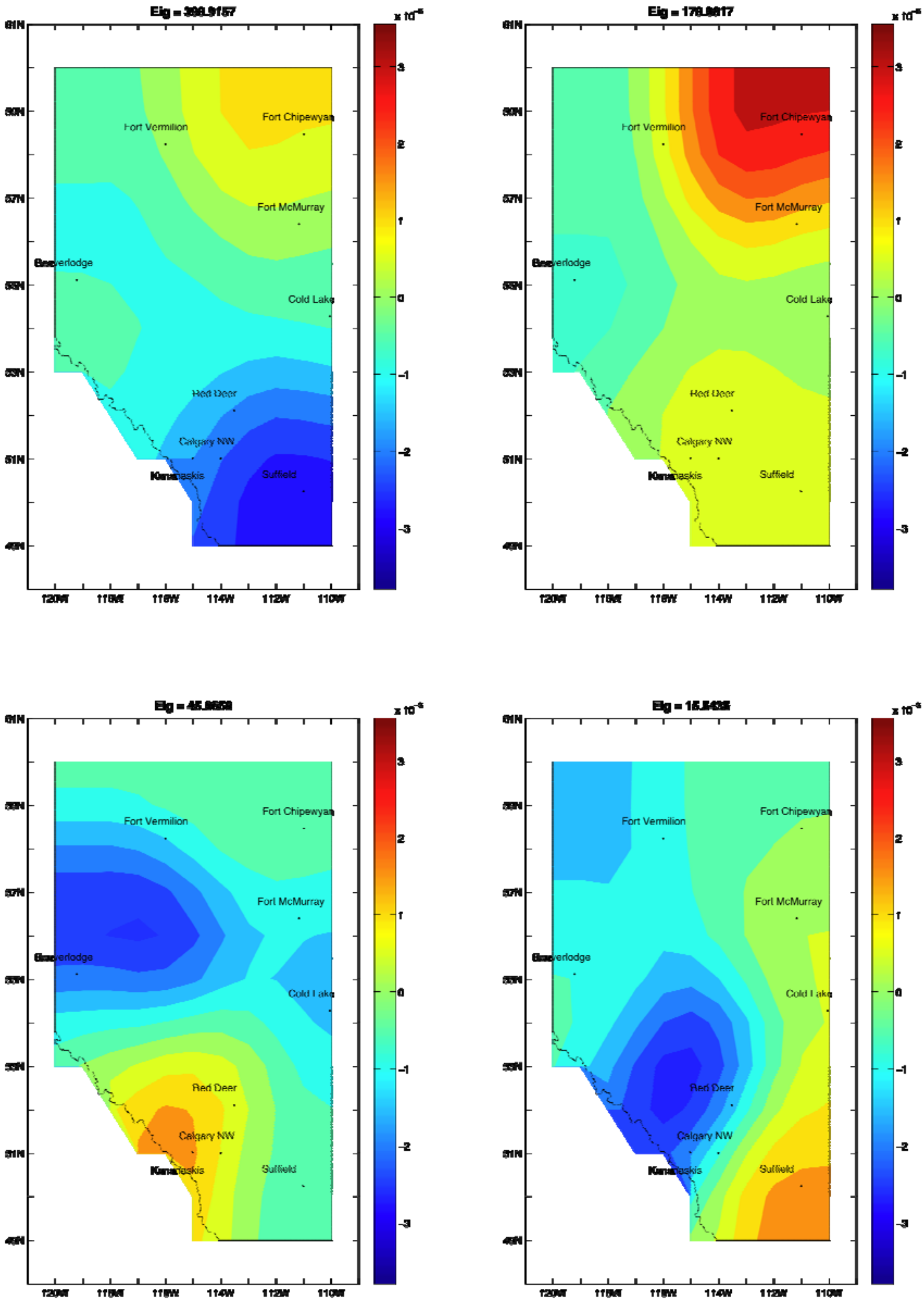
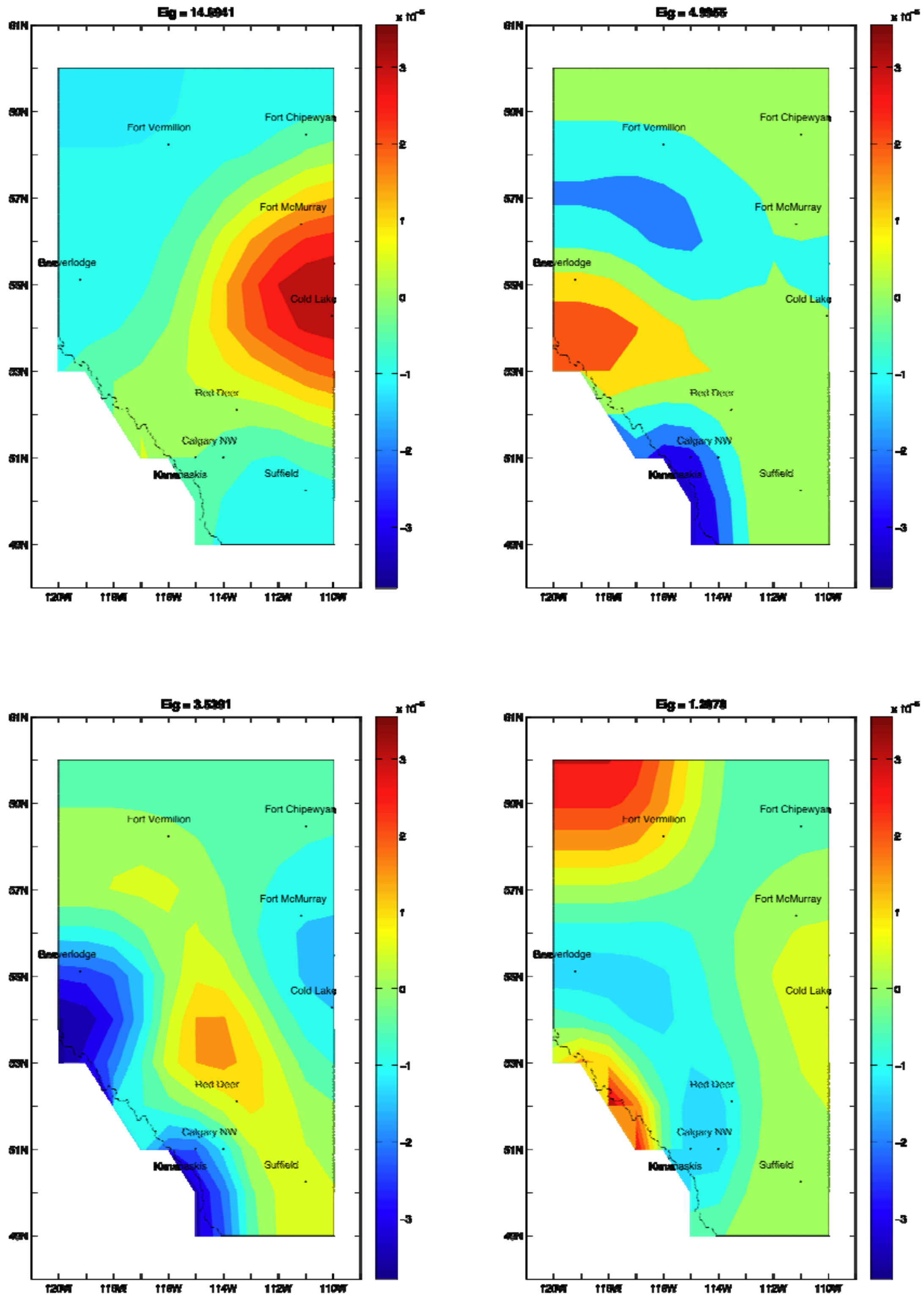


Figure B1: Alberta annual PAI field.

Appendix C: Display of EOFs from gridded data with an area-factor




Fort Vermillion
Fort Chipewyan
Fort McMurray
Beaverlodge
Cold Lake
Red Deer
Calgary NW
Kananaskis
Suffield

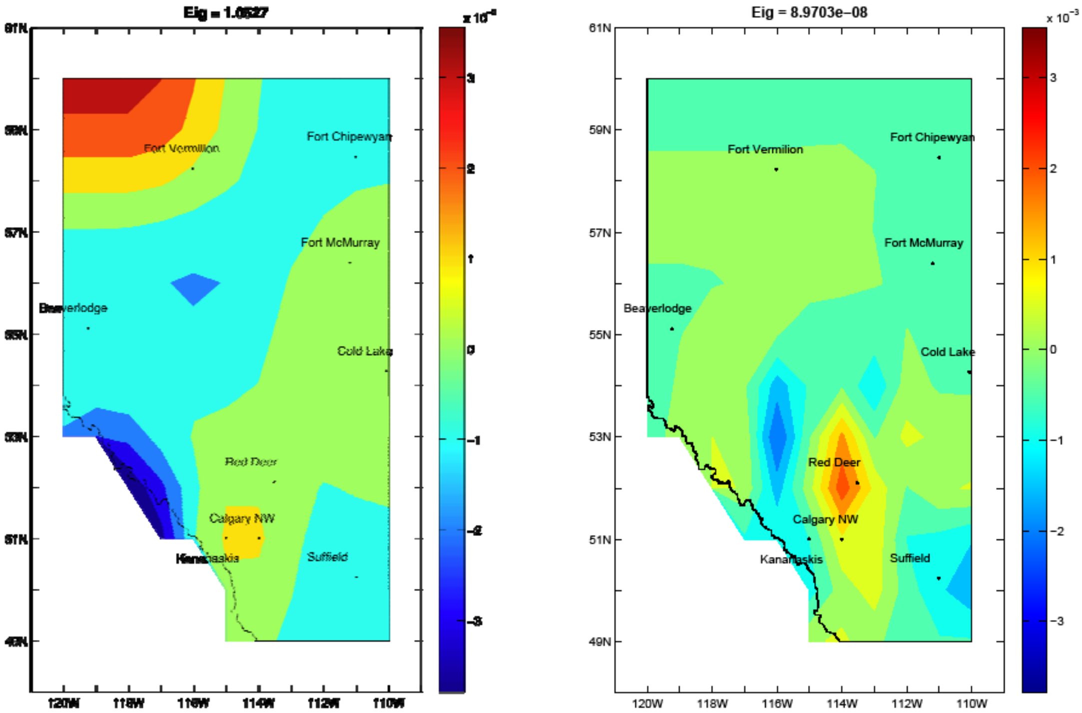
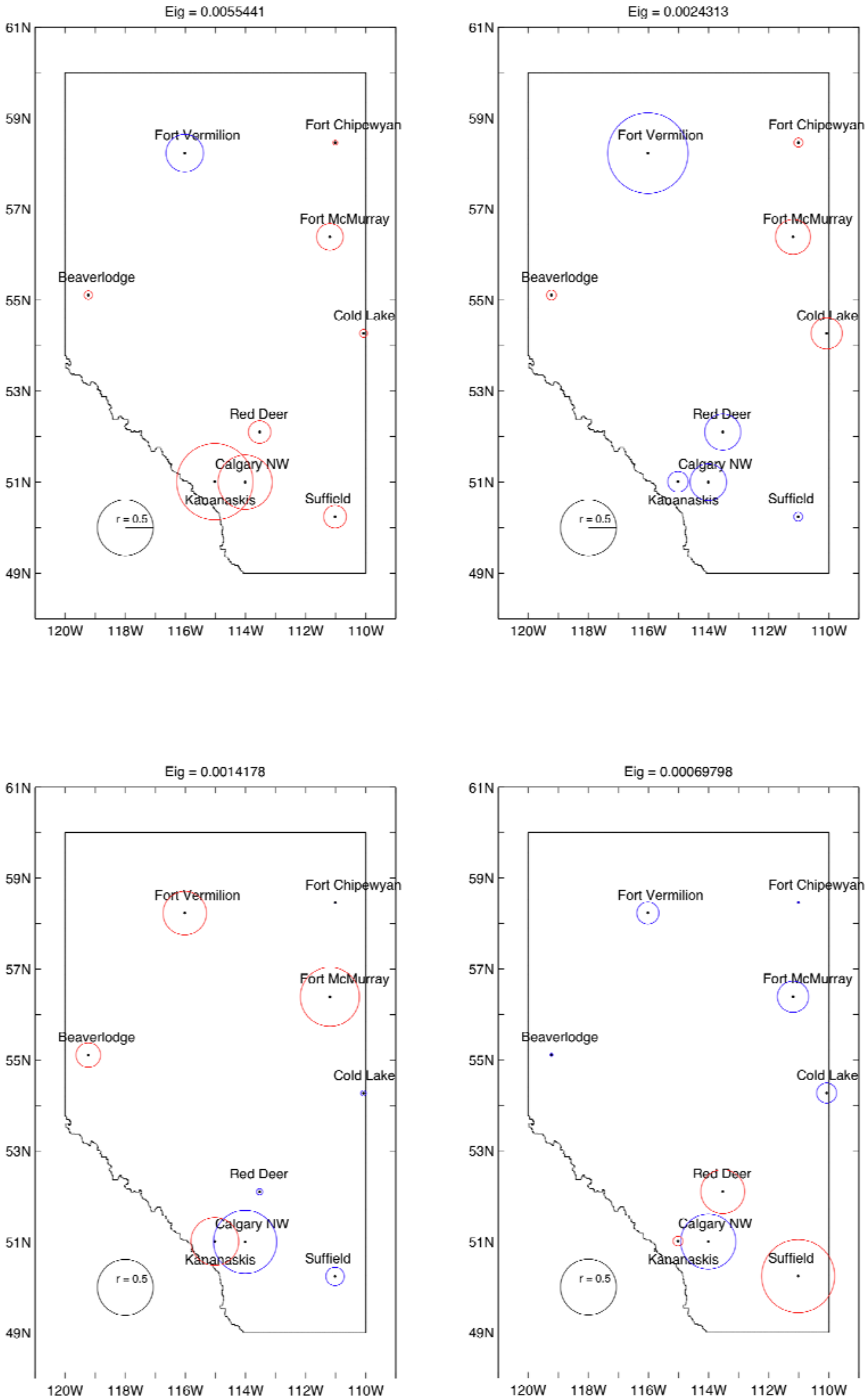
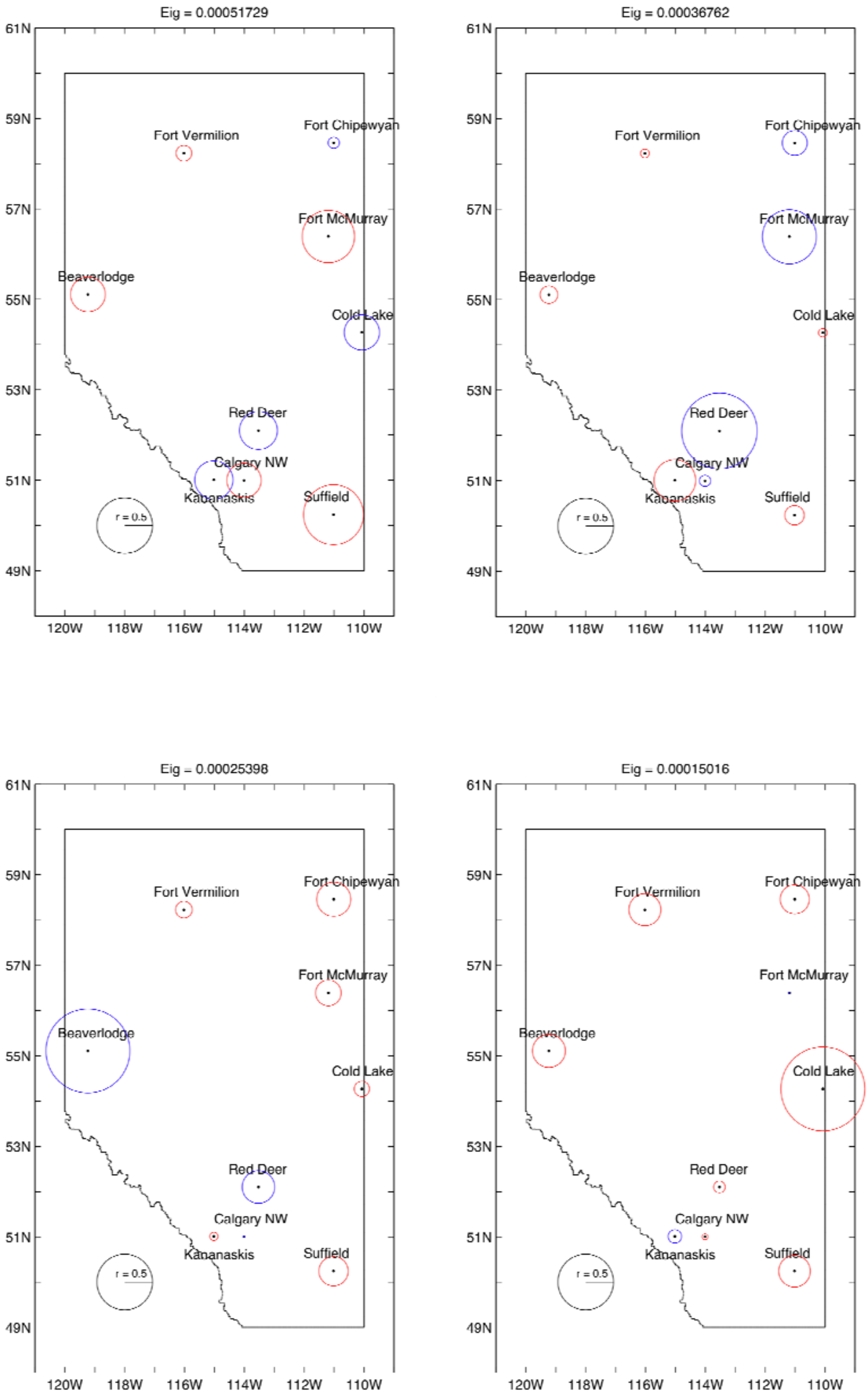


Figure C1: EOFs of Alberta PAI field based on smoothed gridded data.

Appendix D: Display of EOFs calculated from station data




Eig = 0.00051729
Eig = 0.00036762
Eig = 0.00025398
Eig = 0.00015016
61N
59N
57N
55N
53N
51N
49N
120W
118W
116W
114W
112W
110W
Fort Vermilion
Fort Chipewyan
Fort McMurray
Beaverlodge
Cold Lake
Red Deer
Calgary NW
Kananaskis
Suffield
r = 0.5

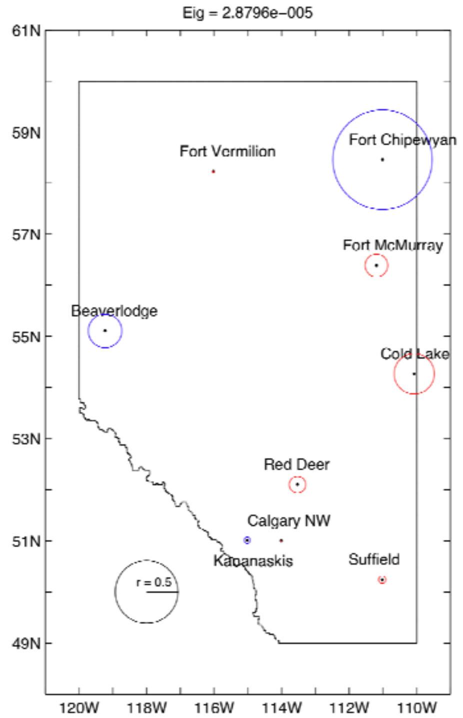


Figure D1: EOFs of the Alberta annual PAI based on station data.