MAP3S/PCN ACID RAIN PRECIPITATION MONITORING DATA BASE

by

R. Gentleman[1]
J.V. Zidek[2]
A.R. Olsen[3]

Technical Report No. 19

August 1985

1. The University of Washington
2. The University of British Columbia
3. Battelle Pacific Northwest Laboratory

MAP3S/PCN ACID RAIN PRECIPITATION MONITORING DATA BASE

R. Gentleman, University of Washington
J.V. Zidek, University of British Columbia
A.R. Olsen, Battelle Pacific Northwest Laboratory

August, 1985

ABSTRACT: This reports surveys the Acid Deposition System (ADS) which is operated by the Battelle Pacific Northwest Laboratory (PNL) and which integrates the data from a number of wet acid deposition monitoring networks. And it presents a detailed description of a data set maintained at the University of British Columbia, MAP3S/PCN data, which has been derived from the PNL database.

1. INTRODUCTION. This report has two aims. First it surveys the Acid Deposition System (ADS), operated by the Battelle Pacific Northwest Laboratory (PNL), which integrates the data from a number of participating wet, acidic deposition monitoring networks. In particular, we review a publication of Watson and Olsen (1984) of the PNL which describes in detail the system and the data accumulating from it. The same document also provides a code manual for users of the ADS data base.

The second goal of this present report is a detailed description of a data set which has been derived by reduction from a parent subset of the ADS data base supplied by the PNL. These data are for a six year period and obtained from the MAP3S/PCN network located in the Northeastern United States. They are currently in a permitted data file maintained by the Statistical Consulting and Research Laboratory of the Department of Statistics at the University of British Columbia (UBC). The file's name is:

SCCA:ACID.COMPS.

The rest of this section consists of a review of ADS and an overview of the MAP3S/PCN network while Section 2 gives a detailed account of the UBC data file. Appendices A and B complete the report by providing background information needed to interpret the data in the file.

1.1 THE ACID DEPOSITION SYSTEM (ADS) FOR STATISTICAL REPORTING. This system has been established at the Battelle Pacific Northwest Laboratory (PNL) for the Environmental Protection Agency (EPA) of the United States. It was established to provide a source of deposition monitoring data for the general

research community and statistics about acidic deposition in North America. The data from several monitoring networks are integrated under a common format to make access convenient. A data base management system has been developed for the ADS at the PNL. Watson and Olsen (1984) have prepared for the EPA, under a Related Services Agreement with Department of Energy of the United States, a description of this system including a user's manual.

The ADS incorporates and updates as new data becomes available, data from a number of distinct monitoring networks. New data are obtained using quality controlled procedures by the participating networks who regularly submit their latest results to the PNL by means of standardized input procedures. Changes in the operation of any network are recorded as part of the data base.

The networks represented in the ADS are together broadly representative of North American deposition areas. The CANSAP and NADP networks, respectively, cover Canada and the whole of the continental United States outside of Alaska. The other networks are more focussed. The APN, APIOS-C and APIOS-D networks are located, except for a single APN station in Northern Saskatchewan, east of Manitoba in Canada; these stations tend to be concentrated in Southern Ontario. The remaining networks, MAP3S/PCN, EPRI/SURE, UAPSP and GLAD are found in the North Eastern United States.

The operation of the various networks differ in detail. The lists of measured chemical components vary to some extent. Collection instruments and the measurement laboratories also vary. Some networks are time based and give weekly or monthly accumulations while others like the MAP3S/PCN give daily

values during and only during periods of precipitation. But the ADS Data Base unifies the data by encoding missing values, by supplying separate data files for such things as site/network characteristics and changes, and so on.

1.2 NEW MAP3S/PCN DATA SET. The PNL has provided a subset of the ADS data base to a group of investigators at the University of British Columbia (UBC). This research group on acid rain is supported by the SIAM Institute for Mathematics and Society (SIMS) through a Co-operative Research Agreement with the EPA. The data derive from the nine stations of the MAP3S/PCN network. Roughly 6 years of record up to the end of 1982 are included. The duration of the record varies from station-to-station but is exceptionally for its duration as a quality daily precipitation network.

This data set has been considerably simplified without loss of information by deleting the redundant entries which were needed to integrate it into the ADS Data Base. The result is about 6 megabytes in length and much less expensive to use in terms of computing costs than the original. Some of the character codes were changed to make the data Fortran readable on UBC's IBM compatible, Amdahl main-frame computer. Finally the data set was reformatted to make it more intelligible (though less compact) to prospective users.

1.3 OVERVIEW OF THE MAP3S/PCN NETWORK.

The MAP3S Precipitation Chemistry Network (MAP3S/PCN) was initiated in 1976 with the objective of creating a long-term high-quality data base for the development of regional transport and deposition models (MAP3S 1982). This nine-station network covers the northeastern United States (see Figure 2) and was chosen to maximize information on regional precipitation chemistry, subject to the constraints of financial limitations on the number of stations and the geographical location of skilled operating groups.

Precipitation samples are obtained on an event basis, "event" being defined as any 24-period during which precipitation occurred. Overall network coordination and central analytical laboratory operation is performed by the Pacific Northwest Laboratory of the Department of Energy. The characteristics and protocol of the network are reported by Rothert and Dana (1983) and in other similar periodic summary reports.

SECTION 2. A DESCRIPTION OF THE DATA

This section consists of two parts. Part 1 describes in detail the form and meaning of the shortened data records. Part 2 explains what components of the original records were deleted and why. This section draws heavily on Watson and Olsen (1984). The latter reference should be consulted for a more complete presentation of the details.

2.1 A DESCRIPTION OF THE OUTPUT RECORD

A list of the fields and the columns they occupy is given in Appendix A, Table 4. Each record corresponds to a single precipitation event. The record consists of two parts. The first part is contained in column 1 through 67 and contains general information about the event. Columns 68 through 202 contain the 15 COMPONENT-RESULTS records. Each of these records contains the value for that chemical component of the sample.

ADS-IDENT

This 6 character field has 2 parts. The first part is a 4 character site identification number given to the site by ADS. See Table 1 of Appendix A for a list of the MAP3S/PCN sites.

The second part is a two digit number. Every time the protocol changes at that site this number is incremented by one. See Table 2 of Appendix A for a list of the protocol changes.

REP-DATE

This is an 8 digit field which indicates the time of the end of the event. A different field in the record gives the hours of rainfall so the time of the start of the event could also be calculated. The first, next, third and last two digits indicate respectively, the year, the month, the day and the hour.

AT-LAB-DATE

This is a 6 digit field which gives the date at which the sample reached the laboratory. This field is formally identical to that of the REP DATE, without the last two digits which indicate the hours (YYMMDD) in the latter case.

PRECIP OCCUR

This one character indicates whether or not a measure of the quantity of precipitation is available. The codes used are:

0: a measure of the quantity of precipitation is available.

1: a measure of the quantity of precipitation is not available.

In the total of 3438 records there were only 40 1's. Typically, no measure of precipitaion indicates that it is not known whether or not a precipitation event occurred.

DEPOSITION-TYPE

This one character field indicates what type of sample was collected:

0: bulk sample. The collection bucket was exposed for the entire sample period.

1: wet deposition. The collection bucket was exposed only during precipitation.

PRECIP-TYPE

This one character field shows what type of precipitation occurred:

0: rain

1: snow or other frozen precipitation

2: mixed rain and snow

3: unknown

4: not available but known

HOURS-OF-RAIN

This is a 4 digit field which indicates how many hours of rain there were during the sampling period.

LID-OPENINGS

This two digit field shows the number of times the lid of the collection device was opened. A value of 00 indicates that no information is available. Any other value is the reported number of lid openings.

SAMPLE-QUANTITY

RAIN-GAUGE-MM

This 8 digit field gives the number of millimeters of rain in the rain gage for that event. If no value is reported the field is set to '99999999'.

SAMPLE-VOLUME-ML

This is an 8 digit field indicating the actual volume of precipitation in the sampling instrument (in milliliters). If no value is reported the field is set to '99999999'.

PREDICTED-VOLUME-ML

This is an 8 digit field that is calculated by multiplying RAIN-GAUGE-MM by INSTRUMENT-AREA-SQ-CM and dividing by 10 to convert to ML. IF RAIN-GAUGE-MM is not available this field is set to '0_ _ _ _ _ _ _'.

PREDICTED-DEPTH-MM

This is an 8 digit field which is calculated by dividing SAMPLE-VOLUME-ML by INSTRUMENT-AREA-SQ-CM and multiplied by 10 to convert it to mm. It is set to '99999999' if SAMPLE-VOLUME-ML is not available.

NOTES

This is a 6 character long field that contains information about the sample. A note consists of a letter followed by two digits. This field is capable of containing two notes which is the maximum number of notes found in any MAP3S/PCN record. See Appendix B for further details and a list of note codes and their interpretations.

COMPONENTS RESULTS

COMPONENT-NUM

This is a two digit field which contains the code for thecomponent. For a list of the codes, their corresponding components, and the units used see Appendix A Table 3.

RESULT-FLAG

A single character field which identifies samples at or below detection limits:

blank: normal

: at or below detection limit; the detection limit will then appear in RESULT-VALUE.

RESULT-VALUE

A six digit field that gives that component's value. One of the spaces is occupied by a decimal point so this field allows for the specification of 5 significant figures.

2.2 CREATING THE NEW MAP3S/PCN DATA SET FROM THAT OF THE ADS DATA BASE.

The following table is quoted from Appendix F of Watson and Olsen (1984) in describing the format of the ADS-OUTPUT-RECORD. Added are notes in parentheses, some explaining why certain fields in the original MAP3S/PCN Data Base were not included in the reduced data set. In particular * means this item is retained in the new data set.

1. ADS-OUTPUT-RECORD
    2. SAMPLE-DEFINITION
        3 SAMPLE-KEY
            4 ADS-IDENT (*)
            4 REF-DATE (*)

3 DATES

4 SAMPLE-START-DATE (MAP3S/PCN is an event network so this information would be of little importance)

4 EVENT-START (generally not given by ADS)

4 EVENT-END (generally not given by ADS)

4 SAMPLE-END (equal to REF DATE in all records and is therefore, redundant).

4 AT-LAB-DATE (*)

3 SAMPLE-DESCRIPTION

4 QC-FLAG (always blank and so deleted)

4 ACTUAL-SAMPLE-PERIOD (always blank and so deleted)

4 PRECIP-OCCUR (*)

4 DEP-TYPE (*)

4 PRECIP-TYPE (*)

4 MET-PROTOCOL (blank in all records and so deleted)

4 DAYS-IN-SAMPLE ('01' in all records and so deleted)

4 HOURS-OF-RAIN (*)

4 LID-OPENINGS (*)

3 SAMPLE-QUANTITY

4 RAIN-GAUGE-MM (*)

4 SAMPLE-VOLUME-ML (*)

4 SAMPLE-VOLUME-ERROR-CODE (blank in all records and so deleted)

4 SAMPLE-VOLUME-ERROR (the error code is a blank in all records so this field is meaningless and deleted)

4 PREDICTED-VOLUME-ML (*)

4 PREDICTED-DEPTH-MM (*)

4 SAMPLING-EFFICIENCY (this is simply a ratio of SAMPLE VOLUME ML to PREDICTED VOLUME ML; it can easily be calculated and so is deleted.)

3 COMPONENT-SUMMARY

4 NUM-MEASURED (the information in these three records

4 NUM-MISSING seems likely to be of little interest

4 REASON-NO-COMPONENTS and so it is deleted for brevity)

3 OBSERVATIONS

4 FIELD-INITIALS (either blank or 'unk' in all records and so deleted)

4 MAX-NOTE (there were at most two notes for any

4 NOTES MAP3S/PCN record so the MAX NOTE was not included and the NOTES field was shortened from 60 to 6 characters.)

2 COMPONENT-PER-RECORD (this is 15 for the MAP3S/PCN records and so is deleted)

2	COMPONENT-RESULTS	
3	COMPONENT-ID	
4	COMPONENT-NUM	(*)
3	RESULTS	
4	TYPE-RESULTS-FLAG	(blank in all records and deleted)
4	DATE-ANALYZED	('999999' in all records and so deleted)
4	ANAL-INITIALS	('unk' in all records and so deleted)
4	RESULT-NOTE	(blank in all records and so deleted)
4	RESULT-FLAG	(*)
4	RESULT-VALUE	(*)
4	ERROR-FLAG	(ERROR-FLAG was blank in all records so
4	ERROR-VALUE	ERROR-VALUE has no meaning; both fields were deleted).

2.3 EXAMPLE.

The records in the new MAP3S/PCN data set each 202 bytes long, are described above. We now examine one record in detail.

The meaning of a record can be more easily perceived if it is read in two parts. The first part consists of the first 67 bytes, or characters in the record and the second part, the remaining 135 bytes. The first part contains the information about the sample, while the second part contains the results for each of the 15 components that MAP3S/PCN analyses. Figure 1a displays line 26 of the new MAP3S/PCN data file along with a few of its neighbours as an illustration.

FIGURE 1 HERE

The first 67 + 36 = 103 bytes appear in Figures 1b and 1c with annotations.

APPENDIX A. TABLES OF BACKGROUND INFORMATION

TABLE A.1 SITE LOCATIONS

ADS SITE IDENTITY	LOCATION	LATITUDE	LONGITUDE	ELEVATION	FIRST ACTIVE DATE
013a	Lewes, Delaware	38 46 00	75 00 00	0	01-Mar-78
020b	Illinois, Illinois	40 03 12	88 22 19	212	20-Nov-77
043a	Whiteface, New York	44 23 26	73 51 34	610	11-Oct-76
044a	Ithaca, New York	42 24 03	76 39 12	509	26-Oct-76
048a	Brookhaven, New York	40 52 00	72 53 00	25	09-Feb-78
057a	Oxford, Ohio	39 31 51	84 43 25	284	01-Oct-78
065a	Penn State, Pennsylvania	40 47 18	77 56 47	393	22-Sept-76
072a	Virginia, Virginia	38 02 23	78 32 31	172	12-Dec-76
171b	Oakridge, Tennessee	35 57 41	84 17 14	341	07-Jan-81

TABLE A.2 SITE REVISIONS BY ADS SITE NUMBER

ADS SITE IDENTITY	SAROAD IDENTITY	ADS REV NUMBER	REV START	REV END	COLLECTION INSTRUMENT
013a	080010101101	00	3/78	11/80	Battelle Northwest
		01	11/80	9/99	Aerochem Metrics model 301
020b	141160101	00	11/77	2/80	Battelle Northwest
		01	2/80	1/82	HASL (standard AEC model)
		02	1/82	9/99	Aerochem Metrics model 301
043a	332020101101	00	10/76	8/77	Battelle Northwest
		01	8/77	9/77	Battelle Northwest
		02	9/77	12/79	Battelle Northwest
		03	12/79	9/99	HASL (standard AEC model)
044a	333300101101	00	10/76	8/77	Battelle Northwest
		01	8/77	9/77	Battelle Northwest
		02	9/77	11/79	Battelle Northwest
		03	11/79	9/99	HASL (standard AEC model)
048a	336870101101	00	2/78	2/81	Battelle Northwest
		01	2/81	9/99	Aerochem Metrics model 301
057a	365300101101	00	10/78	11/80	Battelle Northwest
		01	11/80	9/99	Aerochem Metrics model 301
065a	398620101101	00	9/76	8/77	Battelle Northwest
		01	8/77	9/77	Battelle Northwest
		03	9/77	12/79	Battelle Northwest
		04	12/79	2/81	HASL (standard AEC model)
		05	3/81	9/99	Aerochem Metrics model 301
072a	480680101101	00	12/76	8/77	Battelle Northwest
		01	8/77	9/77	Battelle Northwest
		02	9/77	3/81	Battelle Northwest
		03	4/81	9/99	Aerochem Metrics model 301
171b	440040103103	00	1/81	9/99	Aerochem Metrics model 301

TABLE A.3 ADG COMPONENTS

Component Number	Name of Component	Standard Units
10	Conductivity (Lab)	μ mho/cm
20	pH (Lab)	pH units
21	pH (Field)	pH units
30	Sulfur - IV (Sulfite)	μ mole/L
31	Sulfur - VI (Sulfate)	μ mole/L
33	Nitrite	μ mole/L
34	Nitrate	μ mole/L
36	Chloride	μ mole/L
37	Phosphate (Ortho-)	μ mole/L
40	$H^{\oplus}$ (free)	μ mole/L
41	Ammonium	μ mole/L
42	Sodium	μ mole/L
43	Potassium	μ mole/L
45	Calcium	μ mole/L
46	Magnesium	μ mole/L

TABLE A.4 SHORT RECORD

	LENGTH(BYTES)	COLUMNS
1. STAT-UBC-OUTPUT-RECORD		
2 SAMPLE-DEFINITION		
3 SAMPLE-KEY		
4 ADS-IDENT	6	1-6
4 REF-DATE	8	7-14
3 DATES		
4 AT-LAB-DATE	6	15-20
3 SAMPLE-DESCRIPTION		
4 PRECIP-OCCUR	1	21
4 DEPOSITION-TYPE	1	22
4 PRECIP-TYPE	1	23
4 HOURS-OF-RAIN	4	24-27
4 LID-OPENINGS	2	28-29
3 SAMPLE-QUANTITY		
4 RAIN-GAUGE-MM	8	30-37
4 SAMPLE-VOLUME-ML	8	38-45
4 PREDICTED-VOLUME-ML	8	46-53
4 PREDICTED-DEPTH-MM	8	54-61
3 OBSERVATIONS		
4 NOTES	6	62-67
2. COMPONENT-RESULTS-OCCURS 15 times.		
3 COMPONENT-ID		
4 COMPONENT-NUM	2	r1-r2
3 RESULTS		
4 RESULT-FLAG	1	r3
4 RESULT-VALUE	6	r4-r9

APPENDIX B NOTES & NOTE CODES

ADS has a system of notes grouped into 4 categories, A, B, C and D. The implications of the notes increase in severity from A up to D.

Notes in category B explain why one or more of the components was not reported for an event, which is known to have occurred. Those in B give information about the quality of the sample and field conditions during the sampling period. They may indicate that some or all of the component measurements may be inaccurate due to contamination of the sample. Notes in C explain the sample quantity and in particular, whether the sample is the combination of two or more events. Finally those in D explain why one or more measurements is suspect; they deal with possible equipment malfunctions or improper procedures used during the analysis.

TABLE B.1

NOTE CODES

Code	Note
A21	Equipment failure - all of event missed
A22	Equipment failure - some data loss
A23	Contaminated and discarded in the field
A24	Sample to aged for SO2 analysis
A25	Sample leaked in transit
A26	Sample inadvertently discarded in field
A27	Sample not shipped
A28	Sample spoiled before analysis
A29	Vandalism - some data loss
A30	No explanation from operation
B31	Alternate rain gage used
B32	Alternate collector used
B33	Wind problems
B34	Snowbridge or ice problems
B35	Overflow
B36	Sample Frozen
B37	Late Collection
B38	Organic Debris in sample
B39	Unidentified debris in sample
B40	Flyash in sample
B41	Dust/dirt in sample
B42	Bird droppings in funnel
C21	Two events combined
C22	More than two events combined
C23	Weekly sample
D21	Equipment failure - part of event missed
D22	Possible operator contamination'
D23	Field pH subject due to procedure
D24	Predicted volume suspect due to gage problems

ACKNOWLEDGEMENTS: The work of the first author was supported under a TEPU grant from the Province of British Columbia. The second was supported by SIMS (through a Co-operative Research Agreement with the Environmental Protection Agency of the United States), the National Science Foundation of the United States (Grant #MCS83-01807) and the Natural Science and Engineering Research Council of Canada.

We are indebted to the Battelle Pacific Northwest Laboratory (PNL) for providing the data described in this report. N. Le of the University of British Columbia made the final modifications needed to put the reduced data set into a fully readable form and made valuable comments about the presentation of results.

REFERENCES

MAP3S/RAINE Research Community (1982). The MAP3S/RAINE precipitation chemistry network: statistical overview for the period 1976-1979, Atmos. Environ., 16:1603-1631.

Rothert, J.E. and Dana, M. Terry (1983). The MAP3S precipitation chemistry network: sixth periodic summary report (1982). PNL-4787, Pacific Northwest Laboratory, Richland, WA.

Watson, C.R. and Olsen, A.R. (1984). Acid Deposition System (ADS) for Statistical Reporting: System design and User's Code Manual. Report prepared for the U.S. Environmental Protection Agency, EPA-600/8-84-023.

FIGURE 1. Illustrative Record From Reduced MAPJS/PCN Data Set.

FIGURE 1(a). Record Number 26

```
23   043a0077031415999999010001705l6.43000880.9999805.000017.97959      10  8.4020  4.8221  4.6030-.10  31  6.2033  0.11
     34  10.0036-.40   37-.40   40  15.0041999999943999999943999999945999999946999999
24   043a0077031716999999010001020?.960000232.0000390.00004.734694      10  47.0020  3.9621  4.0530  0.1031  43.0033  0.33
     34  90.0036-.40   37-.40   40 110.0041  36.0042999999943999999945999999946999999
25   043a0077032114999999010001511?.349999130.0000115.00002.653061      10  23.0020  4.2221  4.2330-.10  31  14.0033-.08
     34  44.0036-.40   37-.40   40  60.0041   7.5043999999943999999945999999946999999
26   043a0077032311999999010001902?2.239991087.9991069.99922.20408      10  8.3020  4.6421  4.8530-.10  31  7.2033  0.15
     34   7.8036   1.6037-.40   40  23.0041   5.8042999999943999999945999999946999999
27   043a0077032513999999010001706?.369999326.0000410.00006.653061      10  18.0020  4.2421  4.4030-.10  31  22.0033  0.13
     34  10.0036   1.9037-.40   40  58.0041  11.0042999999943999999945999999946999999
28   043a0077033913999999010000602?.960000436.9999390.00006.918367      10  9.3020  4.5521  4.3730-.10  31  8.2033  0.21
     34   9.1036   3.3037-.40   40  28.0041   3.1043999999943999999945999999946999999
29   043a0077033113999999010000506l0.60999533.0000520.000010.81755      10  36.0020  4.0521  4.0530  0.1931  43.0033  0.28
     34  53.0036   3.9037-.40   40  89.0041  31.0042999999943999999945999999946999999
```

FIGURE 1(b). Background Information In the Record

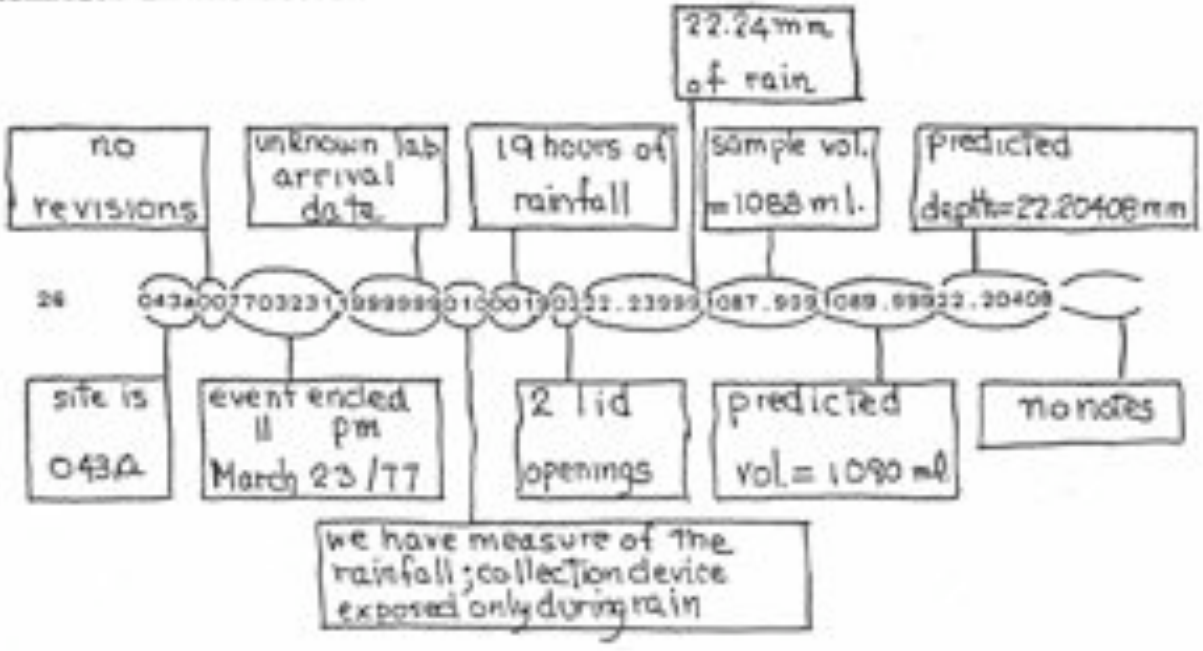

FIGURE 1(c). The first Four Data Components

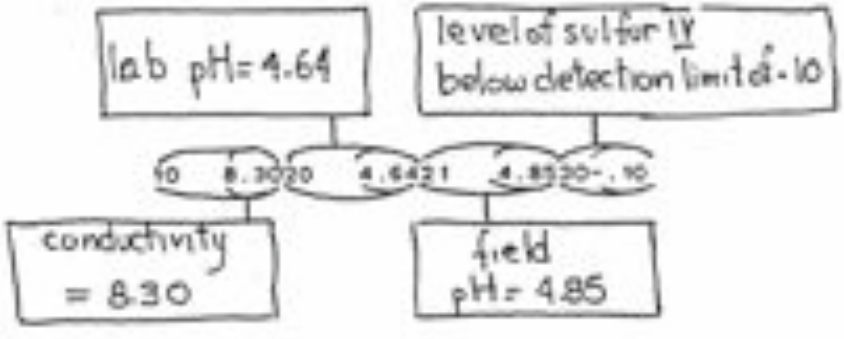

FIGURE 2.

MAP3S/PCN Sites Identified by ADS Code