PREDICTING LIGHTNING OCCURRENCE AND FREQUENCY FROM UPPER AIR SOUNDINGS OVER STONY PLAIN, ALBERTA

by

Kerry Anderson

Forestry Canada
Edmonton, Alberta

and

Robert Charlton

University of Alberta
Edmonton, Alberta

1. INTRODUCTION

Over the last decade, the deployment of lightning detection systems has been prolific. Lightning location data is now a standard source of data for most weather services in the US and Canada. Yet, meteorologists confronted with the problem of forecasting lightning can attest to the elusiveness of this phenomena. Given the standard tools of the forecaster, are there reliable procedures to forecast lightning flash occurrence and frequency?

One of these standard tools is the upper air sounding. Released every twelve hours, soundings provide altitude, pressure, temperature, and humidity measurements to which cloud physics principles can be applied. By nature, lightning is a convective phenomena, but to what degree can lightning be related to these indicators?

To assess this problem, a statistical analysis was conducted using information from the Alberta Forest Service's lightning detection system (Nimchuk 1985) and sounding data from the Stony Plain upper air station.

2. DATA SOURCES

In 1982, the Alberta Forest Service installed an LLP lightning detection system (Krider et al 1976, 1980). For the past few years, the system has operated with 12 direction finders located to provide maximum coverage over the forested area in Northern and Western Alberta.

Located 30 kilometres west of Edmonton, the Stony Plain upper air station is situated in the area of maximum lightning detection coverage. Soundings at 12:00 and 00:00 UTC correspond to 5:00 and 17:00 LST, respectively.

3. METHODOLOGY

The statistical analysis included t tests and logistic regressions conducted on lightning occurrence and linear regressions and multiple linear regressions conducted on lightning frequency. As the goal of this paper was to assess the practicality of predicting lighting, details such as actual regressions equations have been left out of the text.

The predictands in the study were lightning flash occurrence and lightning flash frequency. The two lightning flash polarities -- positive and negative -- were studied separately to determine possible trends in each. Lightning data was restricted to an area two degrees latitude and two degrees longitude centered at Stony Plain (222 km by 132 km). Daily flash frequencies were tabulated within this area from 12:00 LST to 23:59 LST. Lightning occurrence was defined as the occurrence of one or more flashes within the defined area and time period.

Predictors were based on sounding measurements at the mandatory pressure levels. These include pressure heights, dry-bulb and dew-point temperatures. Derived fields include the MSL pressure, the wet-bulb zero height, wet-bulb temperatures, temperature advections between 850, 700 and 500 mb (calculated from observed thermal winds), the 1000-500 mb thickness, 24 hour temperature, height and thickness changes, and convective indices. Convective indices include George's K, a modified K (T 500 mb - Td 850 mb + Td 850 mb), Showalter, vertical totals, cross totals, and total totals.

Additional predictors included the previous day's positive and negative flash totals, and a solar radiation function, Q. The radiation term is a sine function of the Julian date peaking at June 22.

Table 1 lists the predictors and the abbreviations used in the study.

Table 1. Predictors included in study.

	Predictor	Description
1.	PTOT2	Previous day's positive flash total
2.	NTOT2	Previous day's negative flash total
3.	WBZ	Wet-bulb zero height
4.	MSLPR	MSL pressure
5.	ZS	1000 mb height
6.	Z8	850 mb height
7.	Z7	700 mb height
8.	Z5	500 mb height
9.	TS	Surface temperature
10.	T8	850 mb temperature
11.	T7	700 mb temperature
12.	T5	500 mb temperature
13.	TdS	Surface dew-point temperature
14.	Td8	850 mb dew-point temperature
15.	Td7	700 mb dew-point temperature
16.	Td5	500 mb dew-point temperature
17.	dPR	24 hr surface pressure change
18.	dZS	24 hr 1000 mb height change
19.	dZ8	24 hr 850 mb height change
20.	dZ7	24 hr 700 mb height change
21.	dZ5	24 hr 500 mb height change
22.	dTS	24 hr surface temperature change
23.	dT8	24 hr 850 mb temperature change
24.	dT7	24 hr 700 mb temperature change
25.	dT5	24 hr 500 mb temperature change
26.	TwS	Surface wet-bulb temperature
27.	Tw8	850 mb wet-bulb temperature
28.	T87adv	850-700 mb temperature advection
29.	T75adv	700-500 mb temperature advection
30.	TH	1000-500 mb thickness
31.	dTH	24 hr thickness change
32.	K	George's K index
33.	K2	Modified K index
34.	VT	Vertical totals
35.	CT	Cross totals
36.	TT	Total totals
37.	SI	Showalter index
38.	LI	Lifted index
39.	Q	Radiation

Morning and afternoon soundings, 12:00 and 00:00 UTC, were studied separately. Models developed from the morning sounding would provide users with the ability to forecast lightning from actual weather. Models developed from the afternoon sounding -- coinciding with the period of peek lightning activity -- would require forecasted inputs to be useful as a prognostic tool.

244 days of data from May, June, July and August for 1986 and 1987 were used to build the statistical models. During this period, positive lightning flashes occurred on 96 of days and negative flashes occurred on 121. The average number of lightning flashes per day were 9 positive and 134 negative with peak daily totals of 357 positive and 4379 negative. 122 days of data for the same months of 1988 were used for verification.

Statistical tests were run using the 1988 edition of BMDP, a statistical software package provided by UCLA.

4. DISCUSSION

4.1 t tests

A t test was conducted on each predictor to test the hypothesis that the means of the two populations -- days with lightning occurrence and days with no lightning occurrence -- are different. Though the test does not measure the degree of uniqueness of each population, the results are useful as guidelines for lightning occurrence prediction.

Table 2 highlights t test results. For 244 records, |t| values exceeding 4.00 would correspond to a greater than 99.99% confidence in the hypothesis.

Table 2. |t| test results for days with lightning versus days without lightning.

Parameter	12:00 UTC		00:00 UTC	
	Positive	Negative	Positive	Negative
PTOT2	1.32	1.03	1.35	1.06
NTOT2	1.52	1.49	1.55	1.53
WBZ	5.92*	6.46*	4.33*	5.15*
MSLPR	3.02	3.20	3.17	3.30
ZS	2.95	3.11	3.23	3.36
Z8	1.06	0.97	2.44	2.50
Z7	1.04	1.42	0.88	0.65
Z5	2.07	2.45	0.07	0.32
TS	5.18*	5.71*	0.75	0.99
T8	4.78*	5.22*	2.14	2.26
T7	3.95	4.66*	2.50	2.93
T5	2.23	2.29	0.09	0.61
TdS	5.63*	5.88*	6.79*	6.73*
Td8	5.34*	5.78*	6.39*	7.34*
Td7	5.24*	5.71*	6.17*	6.66*
Td5	0.67	1.18	0.79	1.19
dPR	3.53	4.15*	2.95	3.31
dZS	3.73	4.24*	2.23	2.40
dZ8	3.15	3.95	3.52	4.09*
dZ7	1.79	2.69	4.00*	4.78*
dZ5	0.86	1.56	4.48*	4.99*
dTS	3.04	2.45	2.78	3.44
dT8	2.73	1.96	1.48	2.12
dT7	2.60	2.25	1.22	1.22
dT5	0.12	0.04	3.68	3.59
TwS	6.05*	6.40*	4.31*	4.59*
Tw8	6.01*	6.49*	4.82*	5.54*
T87adv	0.69	0.83	1.07	0.83
T75adv	0.07	0.42	0.05	0.15
TH	4.19*	4.68*	2.00	2.35
dTH	2.33	1.91	2.24	2.44
K	6.46*	6.84*	9.55*	9.55*
K2	7.29*	8.00*	7.83*	8.15*
VT	4.28*	4.84*	3.09	2.47
CT	3.92	4.31*	8.42*	8.45*
TT	5.60*	6.20*	8.62*	8.23*
SI	6.86*	7.46*	8.67*	8.77*
LI	6.62*	6.98*	7.54*	7.02*
Q	1.77	2.02*	1.88	2.08

* indicates a very significant |t| value with a greater than 99.99% confidence that means are different.

The most significant results were from the convective indices (K, K2, VT, CT, TT, SI, and LI). The 00:00 UTC George's K scored the highest $|t|$ value at 9.55 for both positive and negative flash occurrence.

All moisture terms (WBZ, TdS, Td8, Td7, TwS, and Tw8) below 500 mb were significant.

Dry-bulb temperatures showed significance for the 12:00 UTC soundings, apparently because of the relationship between high morning temperatures and high moisture content. The significance of temperature dropped off with height.

The 24 hour height changes at 00:00 UTC showed significance, with this significance increasing somewhat with height. This supports the importance of the instability associated with upper troughs (Nimchuk 1983). The lesser significance of the 24 hour height changes at 12:00 UTC can be attributed to building ridges likely to follow the trough prior to the period of peek lightning activity.

4.2 Logistic regression

Logistic regression is a method of regression analysis used on logical (true or false) data (Cox 1970). The analysis predicts the occurrence of one of the events in terms of a probability. In this study, the technique was used to predict the probability of lightning occurrence.

Two logistic regression models were build. Model 1 used predictors that were added and removed in a stepwise manner (Neter et all 1985). Model 2 used pre-chosen parameters based on meteorological principles and the results of the t tests. Table 3 lists the predictors entered into the models in the order of their entry.

Table 3. Predictors chosen for the logistic regression models used to predict lightning occurrence.

Model	Time (UTC)	Polarity	Predictors*
1	12:00	Pos	K2, LI, dT8, Q
1	12:00	Neg	K2, Q
1	00:00	Pos	K, dTS, dT8, LI, (SI)
1	00:00	Neg	K, dZ5, K2, TS
2	12:00	Pos	K2, Q, T87adv, TdS, T75adv, PTOT2, WBZ
2	12:00	Neg	K2, Q, T87adv, T75adv, TdS, WBZ, NTOT2
2	00:00	Pos	K, dZ5, TdS, PTOT2, WBZ
2	00:00	Neg	K, dZ5, TdS, NTOT2, WBZ

* Predictors in brackets were removed.

To analyse the results from the models, skill scores were used (Andersson et all 1989; Donaldson et all 1975). These included the detection rate (Pd), the false alarm rate (Pf), the critical success rate (Psi), and the total percent correct (Pcor). These are defined as follows.

$$Pd = \frac{\text{correct lightning occurrence predictions}}{\text{lightning occurrence observations}}$$

$$Pf = \frac{\text{incorrect lightning occurrence predictions}}{\text{lightning occurrence predictions}}$$

$$Psi = \frac{\text{correct lightning occurrence predictions}}{\text{lightning occurrence observations + incorrect lightning occurrence predictions}}$$

$$Pcor = \frac{\text{total correct predictions}}{\text{total number of observations}}$$

Table 4 summarizes the skill scores as percentages for the prediction of lightning occurrence using the 1986 and 1987 data. Cut-off probabilities used to define the prediction of lightning occurrence were chosen for the maximum Psi results. These ranged from 27.5 % to 55.8%.

Table 4. Skill scores for logistic regression models.

Model	Time (UTC)	Polarity	Pd (%)	Pf (%)	Psi (%)	Pcor (%)
1	12:00	Pos	85.11	40.74	53.69	71.25
1	12:00	Neg	81.51	26.52	62.99	76.45
1	00:00	Pos	82.80	28.70	62.10	80.33
1	00:00	Neg	78.81	17.70	67.39	81.09
2	12:00	Pos	90.32	43.24	53.50	69.33
2	12:00	Neg	87.29	31.33	62.42	73.95
2	00:00	Pos	78.02	31.07	57.72	77.97
2	00:00	Neg	97.44	32.94	65.90	75.00

On this table, the critical success rate (Psi) ranges from 53.50% to 67.39%, the detection rate (Pd) from 78.02% to 97.44%, and the false alarm rate (Pf) from 17.70% to 40.74%. These can be considered very good prediction results though the high false alarm rate in some cases is a concern.

The results show small differences between the models. Model 1 has the best Psi scores overall, though in most cases, the differences between them are only a percentage point. Model 2 has better detection rates, but also has high false alarm rates resulting in poorer Psi scores.

4.3 Linear regression

Linear regressions were conducted on all predictors individually to study the correlations each had with lightning flash frequency. Table 5 lists the results.

The table shows poor correlations from a forecaster's point of view. Excluding results using the previous day's flash totals (which have a large number of points at the origin), the highest value of 0.382 gives an r^2 value of 0.146 indicating that only 14.6% of the variance is explained by the regression line.

For 244 cases, an r value greater than 0.220 or

Table 5. Correlation coefficients (r) for the linear regressions of parameters against the number of lightning flashes.

Parameter	12:00 UTC Positive	12:00 UTC Negative	00:00 UTC Positive	00:00 UTC Negative
PTOT2	0.313*	0.215	0.313*	0.215
NTOT2	0.451*	0.300*	0.451*	0.300*
WBZ	0.269*	0.273*	0.278*	0.320*
MSLPR	-0.107	-0.061	-0.190	-0.153
ZS	-0.106	-0.058	-0.194	-0.155
Z8	-0.021	0.038	-0.135	-0.071
Z7	0.101	0.156	-0.013	0.057
Z5	0.169	0.211	0.066	0.127
TS	0.197	0.222*	0.068	0.129
T8	0.231*	0.245*	0.199	0.239*
T7	0.271*	0.284*	0.214	0.241*
T5	0.175	0.178	0.131	0.144
TdS	0.270*	0.266*	0.265*	0.282*
Td8	0.289*	0.268*	0.288*	0.321*
Td7	0.176	0.166	0.264*	0.275*
Td5	-0.037	-0.061	0.023	0.084
dPR	-0.098	-0.046	-0.203	-0.171
dZS	-0.096	-0.049	-0.180	-0.152
dZ8	-0.094	-0.041	-0.216	-0.184
dZ7	-0.074	-0.036	-0.193	-0.159
dZ5	-0.065	-0.033	-0.170	-0.122
dTS	0.037	0.049	-0.069	-0.064
dT8	0.035	0.000	0.017	0.016
dT7	0.015	0.012	0.032	0.059
dT5	-0.027	-0.029	-0.117	-0.069
TwS	0.261*	0.271*	0.222*	0.262*
Tw8	0.323*	0.315*	0.291*	0.333*
T87adv	0.085	0.093	0.029	0.030
T75adv	0.035	0.035	0.194	0.144
TH	0.245*	0.259*	0.188	0.224*
dTH	0.018	0.010	-0.009	0.011
K	0.203	0.184	0.313*	0.345*
K2	0.318*	0.310*	0.330*	0.382*
VT	0.122	0.139	0.132	0.171
CT	0.170	0.142	0.224*	0.250*
TT	0.201	0.190	0.256*	0.301*
SI	-0.300*	-0.282*	-0.313*	-0.362*
LI	-0.194	-0.206*	-0.215	-0.256*
Q	0.048	0.061	0.045	0.059

* r values greater than 0.220 or less than -0.220 indicate the variance is significant with greater 99.9% confidence.

less than -0.220 (r^2 of 0.0484) indicates a greater than 99.9% confidence that a correlation is not due to chance. Parameters that met this criteria include the wet-bulb zero height, the dry-bulb, wet-bulb and dew-point temperatures below 500 mb, the 1000-500 mb thickness, and all the convective indices except the vertical totals.

There appears to be no significant differences in the correlations for the two time periods or for the two polarities, with the exception of the convective indices. These showed better correlations for afternoon models in all cases.

Because of the low r values, it would be a mistake to make any serious conclusions from these results. One could conclude, however, that the results support the convective nature of lightning. In both positive and negative flash totals, the significant predictors are those that measure convective instability, available low level moisture, and surface heating.

4.4 Multiple linear regression

Three multiple linear regression models were built to predict lightning flash frequency. Model 1 used all days of data. Model 2 used data for days when lightning occurred. Model 3 used the logarithm of lightning flash frequency as the predictand, also restricting its data to days when lightning occurred.

Predictors were added and removed individually in a stepwise manor to produce the best correlation. Table 6 lists the variables entered and removed by the stepwise regressions for each model.

Table 6. Predictors chosen for the multiple linear regression models to predict lightning flash frequency.

Model	Time (UTC)	Polarity	Predictors*
1	12:00	Pos	K2, Q, T87adv, K, dT5,TdS, NTOT2, TS, CT, (K), dT8
1	12:00	Neg	K2, T87adv, Q, K, dT8, Z8
1	00:00	Pos	K2, TS, CT, PTOT2, Q
1	00:00	Neg	K2, dTS, dT8, Q, SI, LI, WBZ, (K2), Z5
2	12:00	Pos	K2, dTS, T87adv, T8, SI, (K2), Q
2	12:00	Neg	K2, T87adv, K, dT8, Z8, Q, NTOT2
2	00:00	Pos	K2, LI, Z8, T87adv, PTOT2, TS
2	00:00	Neg	K2, dTS, dT8, TdS, K, SI, TT
3	12:00	Pos	K2, Q, PTOT2, TdS, dZ5, TS
3	12:00	Neg	TdS, Q, SI, TS, CT, dT5, dTS, TT, (CT)
3	00:00	Pos	Td8, PTOT2, Q, dZ5, T87adv, K
3	00:00	Neg	K, T75adv, Q, TT, dTS, dT8, PTOT2, WBZ, SI, T87adv

* Predictors in brackets were removed.

In ten out of the twelve models, the first predictor entered was the convective index. In nine of the models, the second predictor entered showed surface heating. After these, a variety of other predictors were entered with less significant contributions to the overall correlation.

Table 7 summarizes the correlation results from the regression models. Correlation coefficients now range from 0.3339 to 0.6998, showing a marked improvement over the linear regressions performed on individual parameters. Model 2 was the best at predicting the positive flash frequency while

Table 7. Correlation coefficients (r) for the multiple linear regression models to predict lightning flash frequency.

Model	Time (UTC)	Polarity	r	r^2
1	12:00	Positive	0.3958	0.1567
1	12:00	Negative	0.3339	0.1115
1	00:00	Positive	0.5497	0.3022
1	00:00	Negative	0.4396	0.1933
2	12:00	Positive	0.4413	0.1948
2	12:00	Negative	0.4016	0.1613
2	00:00	Positive	0.6998	0.4897
2	00:00	Negative	0.4888	0.2389
3	12:00	Positive	0.4200	0.1764
3	12:00	Negative	0.4575	0.2093
3	00:00	Positive	0.4262	0.1817
3	00:00	Negative	0.5678	0.3224

model 3 was the best at predicting the logarithm of the negative flash frequency. The poorest results are from the first model using all lightning data.

When comparing results between the two polarities and the two time periods, there are variations. In four out of the six cases, predictions of positive flash frequencies were better than those for negative flash frequencies. In all three models, afternoon soundings predicted lightning frequency better than the morning soundings.

Still, the results are not very satisfactory. Models 2 and 3 only explain between 16% and 49% of the variation.

5. VERIFICATION

5.1 Logistic Regression

To verify the logistic regression models, the independent 1988 data was fed into the model equations to predict lightning occurrence. Predicted and observed results were tabulated as skill scores shown in table 8.

Table 8. Skill scores for the prediction of 1988 lightning occurrence.

Model	Time (UTC)	Polarity	Pd (%)	Pf (%)	Psi (%)	Pcor (%)
1	12:00	Pos	68.18	53.13	38.46	58.62
1	12:00	Neg	66.13	37.88	47.13	61.34
1	00:00	Pos	63.64	46.15	41.18	66.39
1	00:00	Neg	62.90	32.76	48.15	64.71
2	12:00	Pos	75.61	55.07	39.24	56.36
2	12:00	Neg	68.97	38.46	48.19	60.91
2	00:00	Pos	72.73	42.86	47.06	69.49
2	00:00	Neg	82.26	39.29	53.68	62.71

Results from the verification conform well with those of the model. The skill scores for the verification are still good with Psi (critical success rate) values ranging from 38.46% to 53.68%. Detection rate (Pd) is high, the best results predicting more than 70% of the days with lightning. Yet, the high false alarm rates, as mentioned before, indicate the models have over predicted lightning, such that as much as 55% of the lightning forecasts turned out to be for non-lightning days.

5.2 Multiple linear regression

To verify the multiple linear regression models, lightning frequencies were predicted using the independent data set. Predicted frequencies were compared with observed frequencies using a paired t test. The hypothesis is that there is no significant difference between pairs of observations; in other words, the total of the differences between predicted and observed lightning flash frequencies is zero. P values measure the confidence in the hypothesis. Results are shown on table 9.

Table 9. Paired t test results of predicted lightning frequency vs observed for 1988.

Model	Time (UTC)	Polarity	t	P
1	12:00	Pos	1.44	0.1521
1	12:00	Neg	2.86	0.0050
1	00:00	Pos	1.57	0.1198
1	00:00	Neg	3.42	0.0009
2	12:00	Pos	0.49	0.6251
2	12:00	Neg	2.13	0.0377
2	00:00	Pos	1.24	0.2224
2	00:00	Neg	1.36	0.1775
3	12:00	Pos	-0.52	0.6064
3	12:00	Neg	2.38	0.0203
3	00:00	Pos	-0.54	0.5932
3	00:00	Neg	-0.62	0.5377

The positive t values for models 1 and 2 indicate on average an overprediction of the number of lightning flashes. The P values, ranging from 0.005 to 0.6251, fall into two distinct groups: the first group with values below 0.23 and the other with values above 0.53. Considering the high degree of scatter in the original models, the later group could be considered an acceptable verification. Those models that provided verification included the 12:00 UTC model 2 equation, predicting the positive flash frequency, and all but one of the model 3 equations, predicting the logarithm of the flash frequency.

Curiously, the 12:00 UTC predictions of negative lightning flash frequency were poor in all three models, with a less than 5% confidence in the hypothesis.

6. CONCLUSIONS

The results of the logistic regression models show that the potential for the predictability of lightning occurrence (the detection rate) is above 80%, though high false alarm rates, 30% on average, reduce the value of these predictions.

The multiple linear regression models built to predict lightning frequency are poor at best. The models explain 16% to 49% of the variation.

It is the authors' opinion that an experienced meteorologist could predict lightning with as much accuracy as the models -- if not better. However, this study has emphasized the importance of certain parameters and could serve as a guideline to the forecaster.

7. BIBLIOGRAPHY

Andersson, T., M. Andersson, C. Jacobsson and S. Nilsson, 1989: Thermodynamic indices for forecasting thunderstorms in southern Sweden. Meteorological Magazine, 118. pp 141-146.

Cox, D. R., 1970: Analysis of binary data. Chapman and Hall, London.

Donaldson, T.J., R.M. Dyer and M.K. Kraus, 1975: An objective evaluator of techniques for predicting severe weather events. Proc. 9th Conf. Severe Local Storms, Morman. AMS, Boston. pp 321-326.

Krider, E.P., R.C. Noggle, and M.A. Uman, 1976: A gated, wideband magnetic direction finder for lightning return strokes. J. Appl. Meteo., 15. pp 301-306.

Krider, E.P., R.C. Noggle, A.E. Pifer, and D.L. Vance, 1980: Lightning direction-finding systems for forest fire detection. Bull. Amer. Meteo. Soc., 61. pp 980-986.

Neter, J., W. Wasserman, M.H. Kutner, 1985: Applied linear statistical models. Irwin. Homewood, Ill. 1127 pp.

Nimchuk, N. 1983: Wildfire behaviour associated with upper ridge breakdown. Alberta Energy and Natural Resources, ENR Rep No. T/50. Edmonton. 45 pp.

Nimchuk, N. 1985: The Lightning Location and Protection (LLP) system: Alberta's operational experience. Proceedings of the Second Central Region Fire Weather Committee Scientific and Technical Seminar. Canadian Forestry Service, Edmonton. pp 11-17.

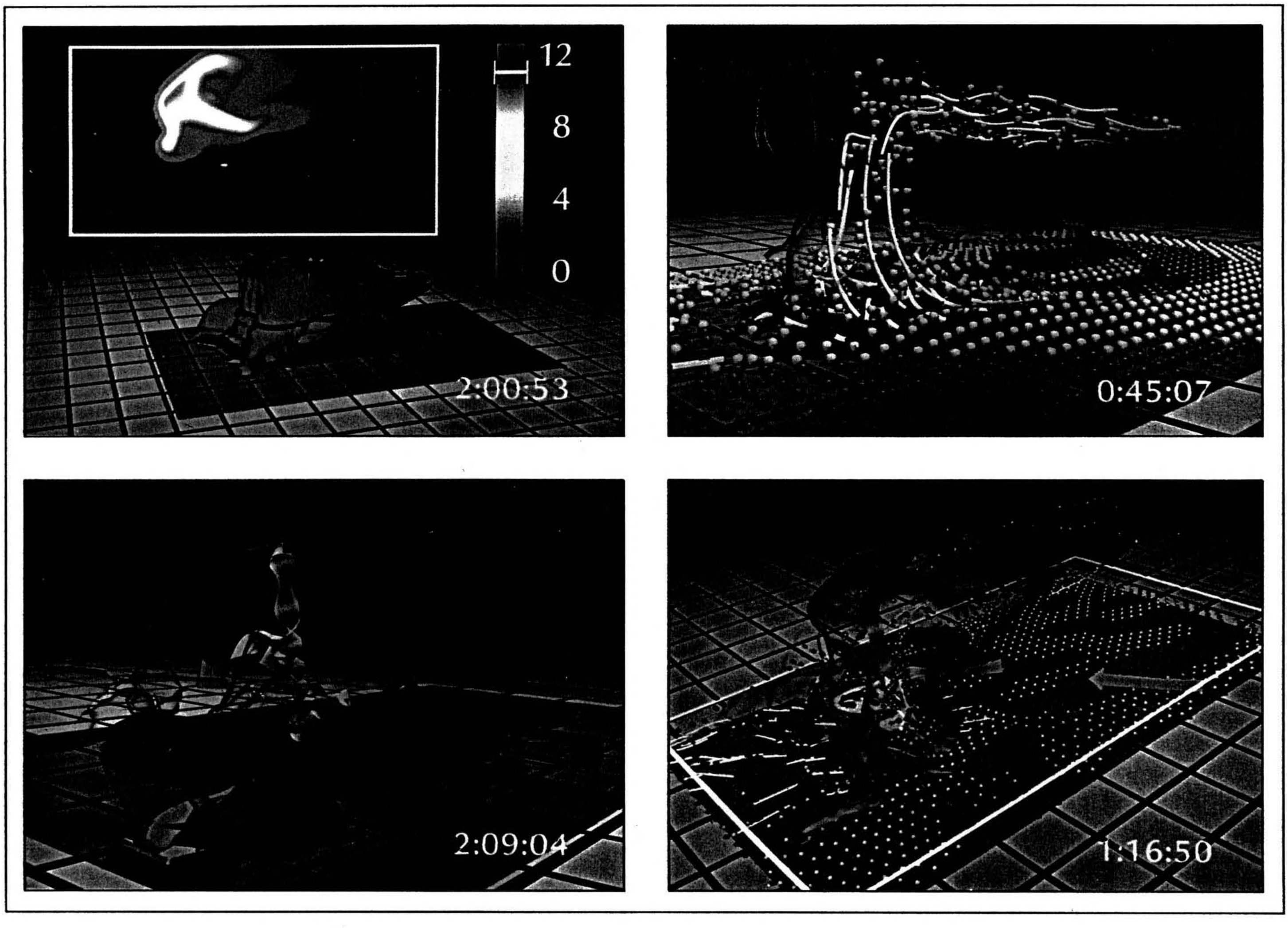
Preprints

16TH CONFERENCE ON SEVERE LOCAL STORMS

CONFERENCE ON ATMOSPHERIC ELECTRICITY

October 22 - 26, 1990

Kananaskis Park, Alta., Canada



AMERICAN METEOROLOGICAL SOCIETY

16TH CONFERENCE ON SEVERE LOCAL STORMS

CONFERENCE ON ATMOSPHERIC ELECTRICITY

October 22 - 26, 1990 Kananaskis Park, Alta., Canada

Sponsored By
American Meteorological Society

Cosponsored by
Alberta Research Council
American Geophysical Union
Canadian Meteorological and Oceanographic Society

See reverse for Front Cover captions.

AMERICAN METEOROLOGICAL SOCIETY

45 Beacon Street, Boston, Massachusetts USA 02108