United States Department of Agriculture

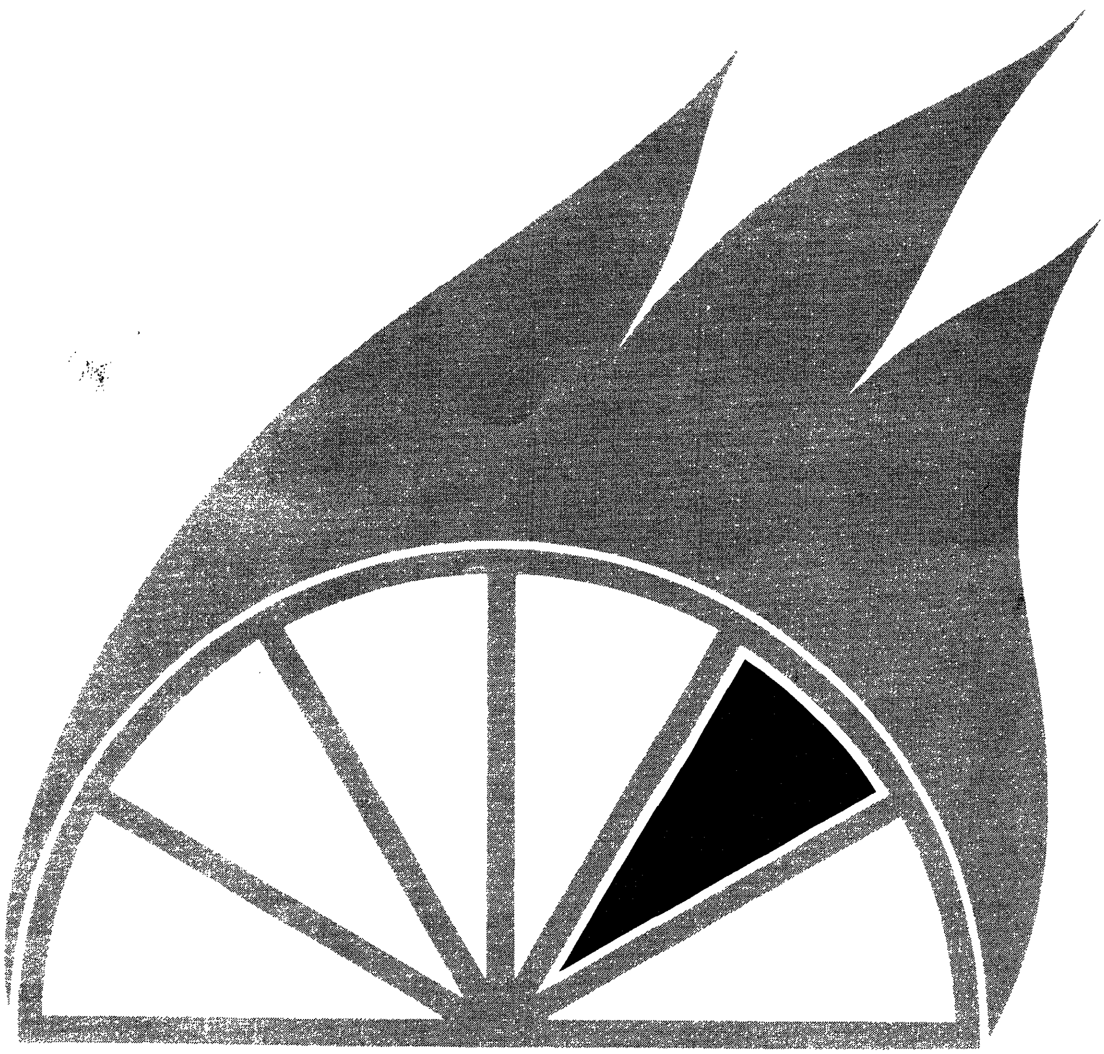
Forest Service

North Central Forest Experiment Station

Research Paper **NC-274**

Fire-Danger Rating and Observed Wildfire Behavior in the Northeastern United States

Donald A. Haines, William A. Main, and Albert J. Simard



North Central Forest Experiment Station
Forest Service—U.S. Department of Agriculture
1992 Folwell Avenue
St. Paul, Minnesota 55108
Manuscript approved for publication April 1, 1986
1986

FIRE-DANGER RATING AND OBSERVED WILDFIRE BEHAVIOR IN THE NORTHEASTERN UNITED STATES

Donald A. Haines, Jr., *Principal Research Meteorologist*,
William A. Main, *Computer Programmer*,
and **Albert J. Simard**, *Supervisory Research Forester*,
East Lansing, Michigan

Fire management expenditures are heavily affected by the success or failure of fire-danger rating systems. Presuppression activities, such as aircraft detection flights and pre-positioning and potential dispatching of firefighters, are guided by these rating systems' numerical values. Fire-danger rating systems, however, are based on simplified models of complex, natural phenomena. Consequently neither the models nor the resultant rating systems completely reflect physical reality. We need to establish the validity of any fire-danger rating system and determine the consequences of its inherent models and assumptions.

The National Fire-Danger Rating System (NFDRS) was established in 1972 and revised in 1978 (Deeming *et al.* 1977). Users of this system include all United States Federal agencies and 35 State agencies charged with forest and rangeland fire protection (Deeming 1983). We have already determined the capability of various fire-danger rating systems (primarily the NFDRS) to predict wildfire occurrence in the Northeastern States (Haines *et al.* 1983). This second phase of the study concentrates on a comparison of predicted values of potential behavior with two measures of observed fire behavior. It emphasizes the use of fire-danger rating as a management planning aid, but not, of course, as a predictor of the behavior of an individual fire.

The NFDRS permits managers to select one or more of 20 possible fuel models; most select those fuel models based on qualitative descriptions given in manuals. One difficulty with this procedure is that each fuel model is defined by a combination of 19 parameters. No one has adequately determined how various combinations of these parameters interact with each other or with the remainder of the system, where 114 equations are used to process 134 variables (Bradshaw *et al.* 1983). Thus, the physical response of each model may not coincide with what might be expected, based on the qualitative description.

As Rothermel (1983) suggests, it is possible to select a fuel model, not only from a description of the physical properties of the vegetation, but also from its known fire behavior characteristics. As Bradshaw *et al.* (1983) state, however, ". . . currently there is no common measurement of any fire phenomenon to correlate with NFDRS ratings." Thus, one objective of this study was to develop procedures for validating fire-danger rating systems. We chose to validate the models from the perspective of the whole system. The intent was to quantify real-world operational variability, in contrast to most published work that has addressed the accuracy of fuel moisture or fire behavior models through carefully controlled experiments (e.g., Andrews 1980, Simard *et al.* 1984, van Wagtendonk and Botti 1984). In our study, standard weather station data were used as system inputs, and wildfire report data were compared with system outputs. Another objective of this study was to determine which of the NFDRS fuel models best fit field observations of fire behavior in the Northeast.

Some investigators have suggested that comparing the NFDRS with observed fire activity is only a partial indicator of system efficacy. The NFDRS, it is argued, is supposed to represent an ambiguous concept termed "fire business." But, we must consider two facts. First, the NFDRS and fire activity are both fuel and weather dependent. Hence, a one-to-one mapping should be possible with a perfect system. Second, considering fire management attributes such as area burned or suppression effort adds a new dimension—suppression efficiency—a concept that the NFDRS makes no attempt to measure. We might expect relations to be stronger between the NFDRS and fire behavior than between the NFDRS and suppression-dependent measures, and tests of those relations indicate that this is indeed the case. Therefore, in NFDRS validation we are largely restricted to tests that use measures of fire behavior.

Validating a fire-danger rating system is a complex undertaking. A poor match between the predicted and observed data could result from inherent variability in the phenomenon studied, measurement errors, sampling error, a poorly devised fire-danger system or, more likely, some combination of these problems. There is no way to identify the dominant problem when validating an isolated system, unless one attempts to redesign the system's physical and mathematical composition and then retest. Including other independently developed systems in the analysis, however, provides an experimental control for an inplace, operational system. Unless there is a bias which favors a specific system, comparison is not affected by data quality. If no one system performs satisfactorily, all systems have failed or, more likely, the problem lies with the observational data. If one or more systems work well relative to the others, we can reasonably assume that the data are adequate and that success and failure lie within the design of the systems. Therefore, in addition to testing the primary system (NFDRS), we also tested comparable components from other systems.

METHODS

Data Collection and Procedures

We established seven study sites—three in Michigan, two in Wisconsin, and two in Pennsylvania. Each study area included a central fire weather station with a 35-mile[1] radius of surveillance. Simard (1972) found that distance did not seriously affect the reliability of most wildfire predictors up to 50 miles from the weather station. Each fire within the circles was reported on an information form we provided. Besides information usually found on standard fire report forms, our form contained the following items that might be used for comparisons with predicted values:

1. Character of the fire on arrival: smoldering, creeping, running, torching, spotting, or crowning;
2. Average rate of spread at the head of the fire at initial attack;
3. Method of spread rate determination: measured, estimated;
4. Average flame length at the head of the fire at initial attack;
5. Time of fire origin and time and size of fire at initial attack;
6. Cover type: grass, brush on well drained sites, brush on poorly drained sites, conifer slash, hardwood slash, old slash, oak-hickory, upland conifer, lowland conifer, upland hardwood-conifer, lowland hardwood-conifer.

The weather stations and all fire control organizations within the surveillance circles were visited twice a year to ensure quality observations. Rothermel and Rinehart (1983) recently compiled detailed field procedures and documentation necessary to verify fire behavior predictions. Although predating their report, most of our procedures paralleled their suggestions.

Our data base included 5,166 daily weather observations and 2,682 wildfires occurring between March 1973 and January 1977. Editing reduced the records because:

1. Suppression forces did not record essential information (777 fires).
2. The fire occurred before 10 a.m. or after 6 p.m., well before or after the 1 p.m. LST weather station observation (623 fires); observed fire behavior measurements were made soon after initial attack.
3. The ratio of the recorded rate of spread to a computed rate of spread based on area increase (Simard 1977) produced an unacceptable difference, i.e., outside the range ($0.33 \leq \text{ratio} \leq 3.0$) (337 fires).
4. The recorded information produced a calculated value of heat per unit area that was physically impossible, given the reported cover type (five fires with obvious recording errors). Heat per unit area can be derived from basic physical relationships if the rate of spread and flame length are known (Andrews and Rothermel 1982, Appendix A). In brief,

$$H = \frac{340\ F^{2.17}}{RS} \qquad (1)$$

[1]*Because all fire-danger rating systems were developed in English units and all fire managers presently do their day-to-day calculations by that means, we used English units in this analysis.*

Where

H = heat per unit area (BTU/ft^2),

F = flame length (ft),

RS = rate of spread (ft/min).

The relation and scaling of NFDRS components to measures of fire behavior (including heat per unit area) are given in our Appendix. After editing, a total of 940 fires remained in the sample (table 1).

Use of Individual Event Versus Grouped Data

We treated the field observations as a collection of individual events rather than as grouped data because the first approach produced significant statistical results, gave reasonable error terms, and made direct comparisons between indices possible. Although grouping the data by increments of index values yielded much higher apparent predictive accuracy, the number of increment groups ranged from 8 to 15, depending upon the predictive index. This caused difficulty in directly comparing statistical measures among indices and produced unreasonably small values for the standard errors of the estimate.

Tests of Fuel Models

The selection of a satisfactory fuel model(s) depends upon managerial needs, and it is unlikely that any one model could fulfill all possible needs. For a discussion of the physical composition of the 20 fuel models, see Deeming *et al.* (1977) and Bradshaw *et al.* (1983), or review the short general descriptions given below:

Fuel model	General description
A	Western annual grasses
B	California mixed chaparral
C	Pine grass savanna
D	Southern rough
E	Hardwoods (winter)
F	Intermediate brush
G	Short-needled pine (heavy dead)
H	Short-needled pine (normal dead)
I	Heavy logging slash
J	Intermediate logging slash
K	Light logging slash
L	Western perennial grass
N	Sawgrass
O	High pocosin

Table 1.--Fires by cover type, season, and location

Cover type	Fires	Total				
	Number	Percent				
Hardwoods	322	34				
Grass	317	34				
Brush	153	16				
Upland conifer	54	6				
Slash	38	4				
Upland mixed	36	4				
Other	20	2				
Total	940	100				
Season						
Spring	377	40				
Summer	399	42				
Fall	164	18				
Location		Hardwoods	Brush	Grass	Upland conifer	Other
		Percent in cover type				
Michigan	199	31	5	30	14	20
Pennsylvania	430	45	32	15	1	7
Wisconsin	311	21	2	61	7	9

P	Southern pine plantation
Q	Alaskan black spruce
R	Hardwoods (summer)
S	Tundra
T	Sagebrush-grass
U	Western long-needled conifer

We conducted a series of tests to determine the capability of several fuel models in the NFDRS to satisfy a number of needs.

1. **Means**—Minimize the absolute difference between the average value predicted by the model and average observed fire behavior. Models that do well in this should be best for long-term planning, because average predicted fire behavior would most nearly match average observed behavior.
2. **Sensitivity**—Minimize the absolute difference between the standard deviation of fire behavior observations and the standard deviation of model predictions. Models that do well in this test would provide about the same level of resolution as the phenomenon being modeled and span about the same range.
3. **Overall Predictive Accuracy**—Explain the variability (R^2) of fire-behavior observations for all fires. Models that do well in this test would provide the most accurate daily fire behavior prediction in management areas that include many different fuel types.
4. **Cover Type Accuracy**—Explain the variability (R^2) of fire-behavior observations within various cover types. Models that do well in this test would provide the most accurate daily fire behavior predictions in management areas that were predominantly of one cover type.
5. **Calibration**—Minimize differences between the regression coefficients and their ideal values (0 and 1). Models that do well in this test would correspond most nearly to observed behavior and would require the least calibration.
6. **Robustness**—Minimize dispersion of fire behavior predictions relative to season, cover type, and location. Models that do well in this test would be well suited to the broadest possible application.

These tests compared the NFDRS Spread Component with the observed rate of spread and the Burning Index with the observed flame length. Five fuel models that performed poorly in preliminary tests (average $R^2 < 0.10$) were deleted from the complete analyses (A, C, L, R, and S). Because responses of models P and U were nearly identical in early tests, only U was tested for all criteria. Models I, J, and K were almost perfectly correlated to each other (although their positions on a fire characteristics chart differed). Models I and J were deleted because they represent heavier fuel loadings than are commonly found in the Northeast. The remaining 12 models (B, D, E, F, G, H, K, N, O, Q, T, and U) were subjected to all tests. Preliminary results from all models are included in this paper. This allows the reader to partially compare the performance of the deleted models with that of the models that underwent complete analysis.

Tests of Various Indices

This part of the study ensured experimental control by including three meteorological elements in the analysis. These weather elements provided baseline results against which we could compare the predictive capability of the more complex (and presumably superior) fire-danger rating systems. We also tested several fire-danger rating indices and their components that might relate to fire spread or energy release. These elements, indices, and components are listed below, and each is described in the Appendix.

Weather elements	
Relative humidity	(RH)
Windspeed	(WS)
Days since 0.10″ precipitation	(PRECIP)
The 1978 National Fire-Danger Rating System	(NFDRS)
1-hour timelag fuel moisture	(FM_1)
10-hour timelag fuel moisture	(FM_{10})
100-hour timelag fuel moisture	(FM_{100})
1,000-hour timelag fuel moisture	(FM_{1000})
Ignition Component	(IC)
Spread Component	(SC)
Energy Release Component	(ERC)
Burning Index	(BI)
The Canadian Forest Fire Weather Index	(Canadian)
Fine Fuel Moisture Code	(FFMC)
Duff Moisture Code	(DMC)
Drought Code	(DC)
Initial Spread Index	(ISI)
Adjusted Duff Moisture Code	(ADMC)
Fire Weather Index	(FWI)
The 1964 National Fire-Danger Rating System	(1964-FDRS)
Fine Fuel Moisture	(FFM)
Fine Fuel Spread Index	(FFSI)
Timber Spread Index	(TSI)
Buildup Index	(BUI)

Fire Load Index (FLI)
Keetch-Byram Drought Index (KBDI)

RESULTS AND ANALYSIS

Statistical Summary of Observed Fire Behavior

The 940 wildfires burned predominantly in hardwoods (34%), grass (34%), and brush (16%) (table 1). They occurred more often in spring (40%) and summer (42%) than in fall (18%). Wisconsin's fires developed mostly in grass (61%), Pennsylvania's mostly in hardwoods (45%) and brush (32%), and Michigan's in hardwoods (31%) and grass (30%). The mean rate of spread reported by observers was much the same for all cover types: all fires (9.6 ft/min), hardwoods (8.9), brush (8.3), grass (10.5), and upland conifer (12.3) (table 2). The highest reported rate of spread was 66 ft/min and the longest reported flame length was 20 feet. The computed mean heat per unit area for all fires (568 BTU/ft^2) was not significantly different from that of hardwoods (632) or upland conifer (775), but was significantly different from that of grass (330) and brush (1,078).

Intercorrelation Among Models

We computed correlation coefficients for the Spread Component and Energy Release Component for all combinations of the 20 fuel models (table 3). Generally, high intercorrelation suggested that the predictive performance of many models will reflect similar performance of many other models. Some models, such as I, J, and K, were almost perfectly intercorrelated in both comparisons, and in this sense they are redundant, although they do differ in terms of absolute values.

On the other hand, many models were highly intercorrelated either in Spread Component or Energy Release Component, but not in both. For example, when we compared predicted rate of spread of model E with that of model F, the correlation was $r = 0.98$; but for heat per unit area the correlation was significantly lower ($r = 0.72$). The difference is difficult to explain solely by examining the values of the critical parameters that determine the composition of the fuel models. The fuel bed depths for models E and F differ by a factor of 10 and the 1-hour fuel loadings differ by 60 percent. Although these parameters affect fire spread, these differences do not appear to affect model intercorrelation. But, factors influencing energy release, such as fuel loadings for the 10-hour, 100-hour, and especially the live fuels, are quite different. This may account for the lowered intercorrelation between fuel models E and F for energy release, but there are many complex relationships among the parameters within each model and simple explanations are likely to be inadequate. See Rothermel (1972) for a review of fuel characteristics that influence fire behavior and Bradshaw *et al*. (1983) for a listing of the physical attributes of each fuel model.

This discussion demonstrates the problem facing fire managers when they attempt to select a fuel model by relying on verbal descriptions or listings of physical attributes. Without more definitive information, the selection process degenerates to a complex guessing game.

Table 2.--Observed and predicted means and standard deviations of rate of spread and heat per unit area

	Observations			
	Rate of spread		Heat per unit area	
	Mean	Standard deviation	Mean	Standard deviation
	Ft/min		BTU/ft^2	
All fires	9.6	9.0	568	1,007
Cover Type				
Hardwoods	8.9[1/]	7.5	632[1/]	1,000
Brush	8.3[1/]	7.0	1,078	1,335
Grass	10.5[1/]	10.3	330	812
Upland conifer	12.3[2/]	14.4	775[1/]	1,767
	Predictions			
Models	Mean	Standard deviation	Mean	Standard deviation
A	53.5	37.9	32	20
B	11.3	6.9	485[2/]	200
C	12.2	11.0	238	105
D	27.3	20.0	905	303
E	9.4[1/]	7.1	458[2/]	103
F	10.0[1/]	5.9	250	113
G	10.9	7.6	580[1/]	303
H	3.1	2.2	310	155
I	26.4	13.9	4,413	1,320
J	17.9	9.4	2,680	743
K	9.3[1/]	5.0	873	260
L	53.5	40.2	63	38
N	62.3	42.5	440	98
O	29.0	26.5	983	223
P	5.7	4.2	548[1/]	118
Q	23.5	16.8	818	228
R	2.5	2.0	300	103
S	7.0	5.7	275	95
T	27.8	28.7	163	105
U	6.1	4.3	530[1/]	150

[1/] No significant difference between this mean and the mean of all fires at $P < 0.001$.

[2/] No significant difference between this mean and the mean of all fires at $P < 0.01$.

Table 3.--Intercorrelation of the 20 NFDRS fuel models for Spread Component and Energy Release Component (sample size = 940)

NFDRS MODEL COMPARISONS
(r x 100)

SPREAD COMPONENT CORRELATIONS

	A	B	C	D	E	F	G	H	I	J	K	L	N	O	P	Q	R	S	T	U	
A		74	79	79	75	74	73	76	51	51	51	98	69	77	78	81	81	80	76	77	A
B	43		93	94	96	98	96	96	86	86	86	80	96	91	95	94	93	93	90	96	B
C	95	51		99	98	95	96	98	74	74	74	85	95	96	99	99	99	99	94	98	C
D	95	36	97		99	96	98	99	77	77	77	86	97	94	99	99	99	99	93	99	D
E	91	71	93	88		98	99	99	85	85	85	81	98	94	99	98	98	98	93	99	E
F	44	95	50	36	72		98	98	88	88	88	81	99	92	97	96	95	94	91	98	F
G	32	52	37	31	49	65		99	88	88	89	79	99	92	99	97	96	96	92	99	G
H	49	67	54	46	68	79	96		84	84	84	82	98	94	99	99	98	98	93	99	H
I	38	75	45	36	63	85	94	97		99	99	57	88	71	81	77	74	73	73	84	I
J	39	79	47	36	66	88	91	96	99		99	57	88	71	81	77	74	73	73	84	J
K	38	75	45	36	64	85	94	97	99	99		57	88	71	81	77	74	73	74	84	K
L	99	40	96	95	90	41	30	47	36	37	36		77	83	84	87	87	86	83	83	L
N	78	83	85	78	95	78	48	65	67	69	67	76		90	97	97	95	95	90	98	N
O	64	78	70	61	83	87	81	92	92	93	92	62	83		95	94	96	96	97	94	O
P	91	66	96	92	98	68	51	69	63	65	63	90	93	84		99	99	99	94	99	P
Q	84	68	89	85	94	74	69	83	78	79	78	83	90	93	97		99	99	92	99	Q
R	95	47	98	97	93	51	44	61	51	52	51	95	82	74	97	93		99	94	98	R
S	85	55	89	87	89	63	73	83	75	74	75	84	81	87	93	97	94		94	98	S
T	94	54	90	86	89	53	24	45	37	39	37	94	80	67	88	79	87	75		93	T
U	74	87	80	70	93	91	62	80	80	83	80	72	94	93	92	93	81	85	77		U
	A	B	C	D	E	F	G	H	I	J	K	L	N	O	P	Q	R	S	T	U	

ENERGY RELEASE COMPONENT CORRELATIONS

The Fire Characteristics Chart

The fire characteristics chart (fig. 1) illustrates three primary characteristics of fire activity—rate of spread, flame length, and heat per unit area (Andrews and Rothermel 1982). The chart allows us to visualize fire behavior in two-dimensional space. Ellipses on this chart show the means and standard deviations of both observed and predicted fire behavior. Because some of the distributions displayed are skewed, the elliptical nature of the joint standard distributions may not be entirely representative in some cases. However, an elliptical approximation appears quite satisfactory for purposes of illustration.

Scaling on the fire characteristics chart can be either logarithmic or linear. With a logarithmic scale (fig. 1), the straight diagonal lines represent flame length and the elliptical standard deviation envelopes are distorted. With a linear scale, the flame length lines are curved and the ellipses appear normal (fig. 2). The logarithmic scale allows the display of a wider range of behavior. Main and Haines (1983) have demonstrated the chart's usefulness in selecting fuel models, showing observed fire behavior, and illustrating means and standard deviations.

Overall fire severity, as well as the character of the fire, may be inferred from a fire's position on the chart. A number of descriptors of fire behavior can locate that point: the NFDRS Spread Component,

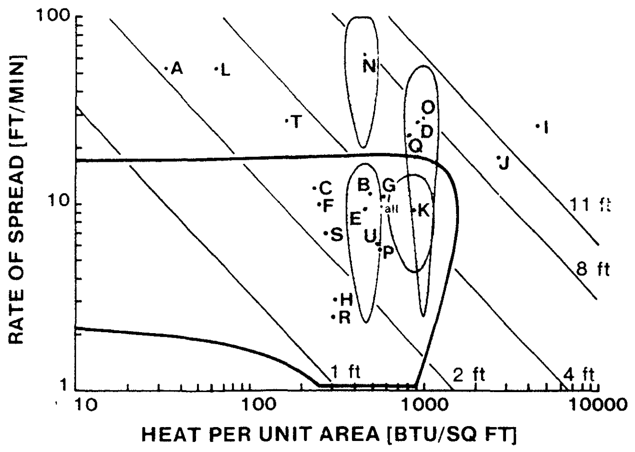


Figure 1.—*Fire characteristics chart displaying the observed behavior mean (*) and a behavior envelope designating one standard deviation (heavy line). Also shown are the means (•) of the predicted potential behavior for each of the 20 NFDRS fuel models, along with envelopes designating one standard deviation for models N, O, E, and K. The diagonal lines represent flame lengths (ft). Logarithmic scale distorts the elliptical envelopes.*

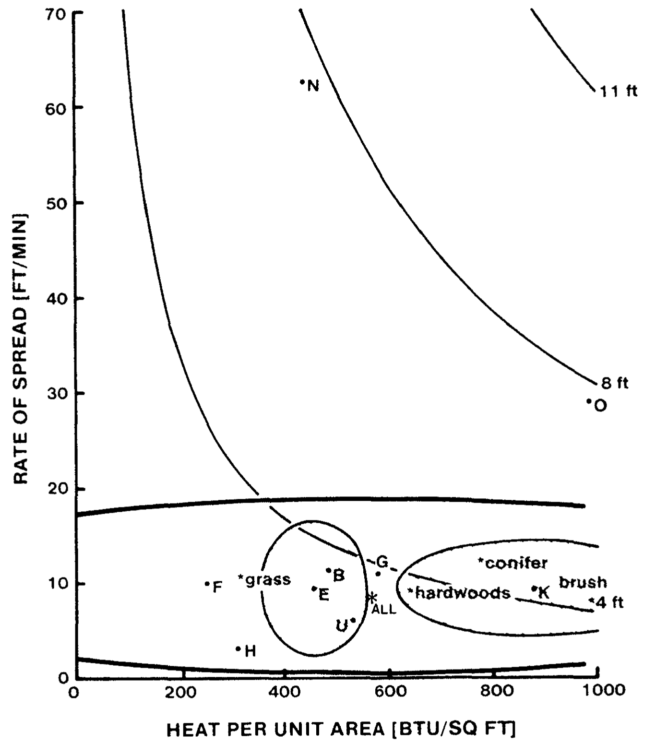


Figure 2.—*Fire characteristics chart displaying observed behavior means (*) for all fires and the following groups: grass, hardwoods, upland conifers, and brush. It also includes means (•) of predicted potential behavior for nine fuel models and behavior envelopes designating one standard deviation for all fires and fuel models E and K. The curved diagonal lines represent flame lengths (ft).*

the NFDRS Energy Release Component, the observed rate of spread, heat per unit area, and flame length. Field measurements of flame length and rate of spread were used in figures 1 and 2.

Means

Figure 1 includes the means of the 20 fuel model predictions and the envelope outlining the standard deviations for models E, K, N, and O; numerical values are listed in table 2. Figure 2 shows the predicted average behavior for nine fuel models and five cover type groups. Elliptical envelopes designate one standard deviation for all fires and for models E and K. Figure 2 shows the isolation of fuel model N from the main body of data. Average fire severity, as well as the character of a model, may be inferred from the position of the mean value on the chart. One interesting feature of figure 1 is the clustering of the means of nine fuel models (B, C, E, F, G, K, U, P, and S) between 238 and 873 BTU/ft^2 and between 5.7 and 12.2 ft/min. This cluster surrounds the observed mean for all fires. Two fuel models, H and R, fell below this cluster due to lower rates of spread. The remaining fuel models were scattered over a broad range of BTU per square foot values, but all average predicted rates of spread for these models are higher than the upper boundary of the envelope outlining one standard deviation of observed behavior.

The mean predicted behavior potentials for fuel models E, F, and K were the only means not significantly different from the all-fires observations ($P < 0.001$). There was no significant difference in mean heat per unit area between models G, P, or U and the mean of all fires ($P < 0.001$); at $P < 0.01$, models B and E are also included (table 2).

Sensitivity

The standard deviation of the observed rate of spread is 9.0 ft/min for the 940 fires (table 2). Six of the 20 fuel models had a notably greater standard deviation than this (A, D, L, N, O, and T); the y-axis elongation was quite evident, for example, in the envelopes for models O and N (fig. 1). These fuel models have a greater response than observed fire behavior. In contrast, 10 fuel models (B, C, E, F, G, I, J, K, Q, and S) had standard deviations in the range $5.0 \le s \le 16.8$ or within a factor of two of that observed. These models provide roughly the same level of resolution as observed fire behavior. Four models (H, P, R, and U) have limited resolution ($s \leq 4.3$) for rate of spread.

The standard deviation for the observed heat per unit area for all fires was 1,007 BTU/ft^2 and only fuel models I and J were within a factor of two. The remaining 18 fuel models had heat per unit area envelopes much smaller than those observed, indicating limited resolution (fig. 1). There are several possible explanations:

1. There are fixed fuel loadings in the models, but this feature is highly variable in nature.
2. Although the model yields a single Burning Index, intensity within individual fires can vary by up to two orders of magnitude (Simard *et al*. 1982).
3. The NFDRS fuel moisture models (which are inputs to the fire behavior models) vary less (have notably lower coefficients of variation) than fuel moisture observed in the field (Simard *et al*. 1984).
4. Flame lengths may be observed less accurately than other information (Johnson 1982).

Predicting Fire Behavior

Although the NFDRS relates to the potential of an initiating fire, our data base contained reports of torching, spotting, or crowning (TSC). With these fires we hoped to identify critical levels of fire danger when such behavior was possible. They were also included in the data base to maintain balance. The fires were identified by adjective classification: smoldering/creeping (22 percent of the total); running, defined as spreading rapidly with a well-defined head (70 percent); and TSC (8 percent). Computations based on the model E Spread Component (see Appendix) indicated that the probability of TSC occurring sometime between initial attack and containment increased at a linear rate as this index scale increased (fig. 3). Even at high index values (SC = 40), however, the probability of TSC behavior was less than 25 percent. In contrast, due to the inherent temporal and spatial variability of fire danger, there was a small but non-zero probability of TSC even at low index values. Running was the most probable behavior, except at very low values of the Spread Component where smoldering or creeping was most probable. On the other hand, there remained a probability that a fire would only smolder or creep at relatively high index values.

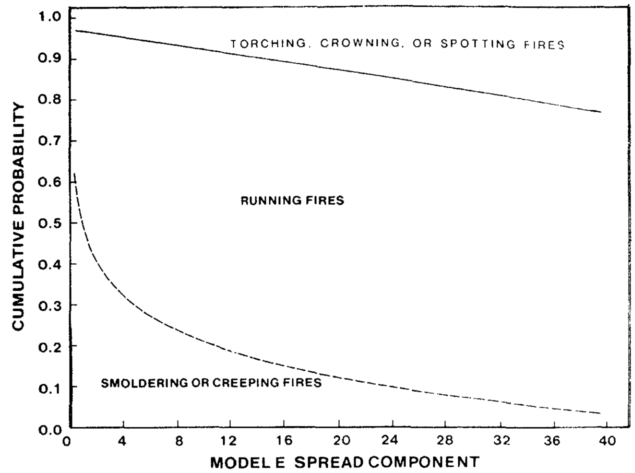


Figure 3.—*Cumulative probability of three classes of worst behavior for 940 fires relative to the model E Spread Component.*

One method of visualizing the relation between observed and predicted fire behavior is with weighted quartile regressions (fig. 4). The lines indicate the values of each index (component) where one-fourth, one-half, and three-fourths of the observations lie below the indicated line. One-half of all observations lie above or below the middle quartile line. An advantage of quartile regression is its resistance to outliers. Note that the ordinary least squares (OLS) solution lies above the middle quartile level due to a few extreme observed values. The divergence of the upper and lower quartile lines indicates that the variance of the observations increases with increasing index values. At the most elementary level, this portrayal shows that the system works—observed values increase with predicted values. It also shows that prediction errors are less

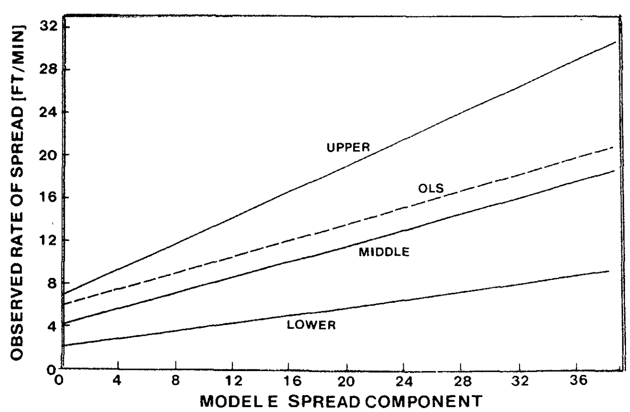


Figure 4.—*Quartile regressions between observed and predicted rates of spread. The OLS line is for the ordinary least squares solution.*

at low index values than at high index values. From a manager's perspective it would, of course, be advantageous if the error situation was reversed.

Figure 5 shows a histogram of standardized residuals for observed rates of spread using the linear regression model. A Gaussian distribution is superimposed to show the departures from normal of the observed data. To test the degree of nonlinearity in the data, we decomposed the error sum of squares into a pure-error component and a lack-of-fit component and then did an F test for linearity (Neter and Wasserman 1974, p. 113). This test showed that the regression function was not linear at the 0.05 level of significance. In addition, we used Bartlett's test for homogeneity of the variance, grouping the observed data by index levels (Neter and Wasserman 1974, p. 509). The observed data groups failed this test, indicating that the increased variance with increased predicted value is significant; therefore, we used a logarithmic transformation on the data. The strongest relations for rate of spread were associated with the form:

$$\hat{Y} = B_0(X^{B_1}) \tag{2}$$

where
Y = estimated rates of spread (ft/min)
X = NFDRS Spread Component.

The strongest relations for flame length and heat per unit area were associated with the form:

$$\hat{Y} = B_0(B_1{}^X) \tag{3}$$

where
Y = flame length (ft)
X = NFDRS Burning Index/10.

There are problems in comparing a logarithmic transformation with an untransformed function because the former generates a multiplicative error term and the latter generates an additive error term. Although some authorities advise caution (e.g., Payandeh 1981), others ignore possible implications (e.g., Lewis-Beck 1980). We believe that comparisons of R^2's and standard errors among equation forms should be interpreted with these error sources in mind.

Overall Predictive Accuracy

Results of the regression analysis for all fires are given in table 4. The most striking feature is the low R^2 values relative to those obtained by other investigators who tested fire behavior models in carefully

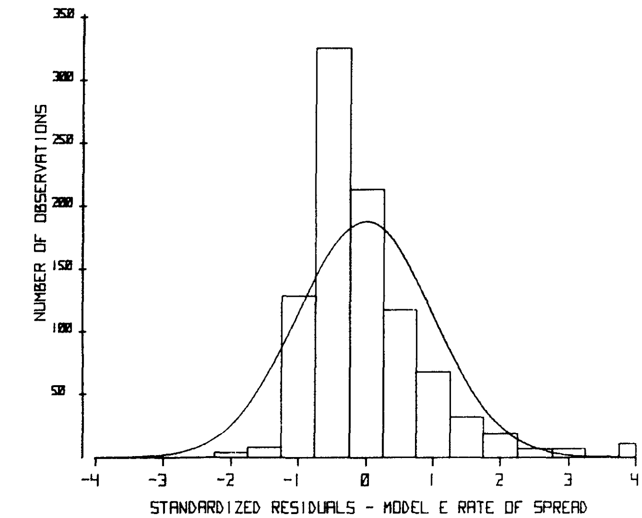


Figure 5.—*Histogram of standardized residuals for rate of spread. A normal distribution is superimposed on the histogram to illustrate the departures from normal.*

controlled experiments (e.g., Andrews 1980, van Wagtendonk and Botti 1984). Values calculated in this study reflect what can be expected when the system is used operationally. Although our study design necessitated treating the observed behavior data as static characteristics, a wildfire is a dynamic process having a wide range of temporal and spatial variability. Although the same applies to prescribed burns, our data generally incorporated only one or two observations of each fire which will, in most cases, differ from the mean value. Low R^2 values will be typical of such data; there are no alternatives if central weather station and wildfire observations are used as the data source.

Table 4 lists the fuel model results for all fires in order of decreasing average R^2. The most accurate combined predictions for all fires were provided by models E and N. Some models scored relatively high for one attribute of fire behavior but low in the other. For example, model G had the second highest R^2 value for rate of spread, but its performance in predicting flame length was unacceptable. Poor performance with the latter could be due to heavy loadings of 100- and 1,000-hour timelag fuels in model G.

For rate of spread, all coefficients differed from the ideal. High intercepts and low slopes indicate that, on the average, all models underpredicted rate of spread at low index values and overpredicted rate of spread at higher values. Model F, for example, has a crossover point at an SC value of about 7. At least two factors are involved in addition to possible model error. The models were designed to predict peak fire

Table 4.--Regression coefficients to transform model-predicted values of potential behavior to observed behavior for all fires

Fuel model[1/]	Rate of spread[2/] R^2	B_0	B_1	Flame lengths[3/] R^2	B_0	B_1	Average R^2
E	0.131	3.12	0.38	0.129	1.12	1.21	0.130
N	.131	1.54	0.38	.129	1.06	1.09	.130
U	.115	3.85	0.35	.129	1.11	1.25	.122
P	.119	3.85	0.37	.123	1.14	1.24	.121
J	.137	1.46	0.55	.097	1.00	1.08	.117
K	.140	2.06	0.55	.092	1.04	1.17	.116
I	.134	1.20	0.54	.091	1.04	1.05	.113
O	.083	3.25	0.24	.133	1.24	1.07	.108
D	.109	2.57	0.32	.103	1.27	1.07	.106
Q	.106	2.87	0.29	.104	1.23	1.08	.105
F	.092	3.79	0.27	.116	1.14	1.27	.104
H	.118	4.87	0.36	.085	1.34	1.32	.102
T	.089	4.05	0.18	.112	1.48	1.10	.101
B	.089	3.73	0.26	.111	1.14	1.18	.100 [4/]
G	.137	2.59	0.43	.059	1.42	1.10	.098
C	.088	4.48	0.20	.105	1.39	1.16	.097
R	.092	5.78	0.25	.102	1.31	1.36	.097
S	.077	1.59	0.16	.099	1.32	1.22	.088
L	.063	4.41	0.12	.074	1.51	1.11	.069
A	.062	4.29	0.13	.066	1.51	1.15	.064

[1/]Listed in order of decreasing combined R^2.

[2/]Rate of spread = $B_0 * (SC)^{B_1}$. Standard errors were grouped between 9.0 and 9.2.

[3/]Flame length = $B_0 * B_1^{(BI/10)}$. Standard errors were grouped between 2.0 and 2.1.

[4/]Models below this line are not recommended for general use in the northeastern United States.

behavior (i.e., they have a built-in bias). Thus, regression equations, which predict average behavior, should have slopes less than one. Conversely, in a perfect world there would be no fires at an index value of zero. To the extent that there are and they spread, it is indicative of the spatial variability of fire danger over one weather station's zone of application. There are no negative spread rate observations to balance the positive rates—hence the positive Y-intercept. Results of the overall regression analysis were combined with other criteria to reduce the number of test fuel models from 20 to 12. Remaining tests concentrated on this reduced set.

Predictive Accuracy by Cover Type

The data were stratified by cover type (hardwoods, brush, grass, upland conifer) and the regression analysis redone. Results for the best five models in each cover type are listed in order of decreasing average R^2 in table 5. The most notable feature is an

Table 5.--Ranking of fuel model performance by major cover groups for all seasons and locations (significant at $P < 0.01$)[1/]

Cover group	Fuel model	Rate of spread[2/] R^2	Rate of spread[2/] $S_{y.x}$	Flame length[2/] R^2	Flame length[2/] $S_{y.x}$	Average R^2
			Ft/min		Feet	
Hardwoods (n = 322)	E	0.196	7.3	0.154	1.8	0.175
	N	.191	7.3	.148	1.8	.170
	K	.209	7.1	.128	1.8	.169
	U	.175	7.3	.159	1.8	.167
	G	.209	7.3	.107	1.8	.158
Brush (n = 153)	K	.249	6.6	.238	2.1	.244
	E	.156	6.8	.248	2.1	.202
	N	.164	6.8	.235	2.1	.200
	U	.150	6.9	.231	2.1	.191
	G	.179	6.8	.188	2.1	.184
Grass (n = 317)	E	.133	10.2	.084	1.9	.109
	N	.127	10.3	.084	1.9	.106
	U	.110	10.3	.083	1.9	.097
	D	.116	10.3	.067	1.9	.092
	B	.109	10.4	.071	1.9	.090
Upland conifer (n = 54)	K	.096	14.9	NS[3/]		
	N	.083	15.1	NS		
	F	.075	15.2	NS		
	B	.073	15.2	NS		

[1/]Listed in order of decreasing combined R^2 values within each group.

[2/]The NFDRS Spread Component was used to predict observed rates of spread, and the Burning Index was used to predict flame lengths.

[3/]The regression equation was statistically insignificant at the 0.01 level.

increased R^2 for hardwoods and brush, indicating, as expected, improved accuracy as the cover types are stratified. Model E was best in two cover types (hardwoods and grass) while K was best in the other two (brush and upland conifer). In contrast, model K ranked low in grass cover (presumably due to its lack of live-fuel moisture) and E ranked low in upland conifer (presumably due to limited inclusion of heavier fuels). Models H, O, Q, and T did not rank in the top five for any cover type. The remaining models (B, D, F, G, N, and U) were ranked in the top five in one to four cover types.

In upland conifers, only four models yielded significant spread predictions at $P < 0.01$. Although no model produced significant results at the 0.01 level for flame length, all models were positively correlated with both rate of spread and flame length. Poor correlation may have resulted from the small sample size relative to the other cover types, coupled with the highest average variability of any cover type (table 2). Further, 26 percent of all upland conifer fires behaved erratically (and hence less predictably) compared with 8 percent for all fires.

Robustness

An important factor in judging fuel model performance is the capability to consistently predict potential fire behavior across seasons, locations, and cover types, thereby simplifying the business of fire-danger rating. We calculated the root mean square (RMS) of differences between the all-fires regression curve for each of the three groups and the individual season, location, or cover-type curves. The test,

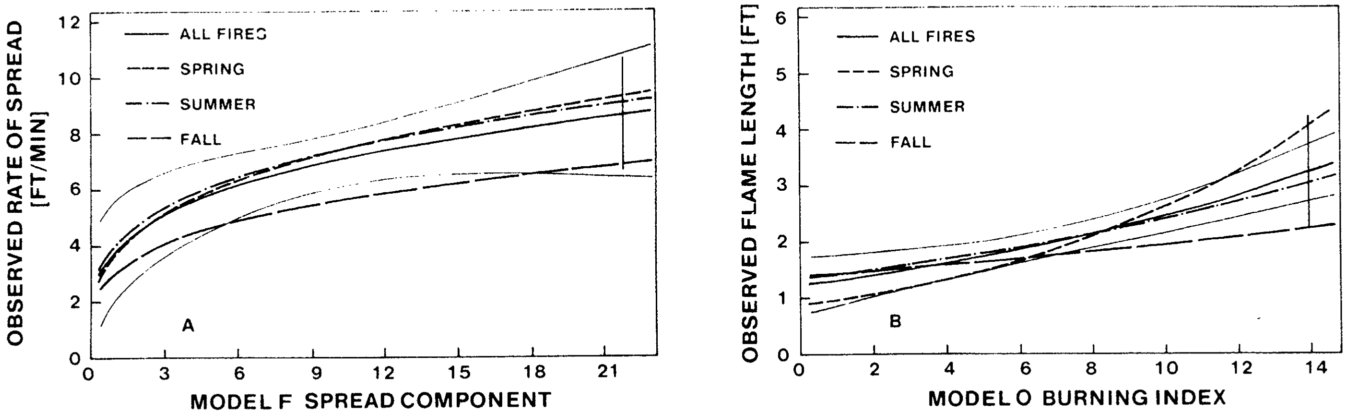


Figure 6.—*(A) Comparison of observed and predicted rates of spread by season using model F Spread Component. The mean plus two standard deviations of the Spread Component is indicated by the vertical line. The heavy solid line denotes the all-fires regression and its confidence limits at (P = 0.01). (B) Comparison of observed and predicted flame lengths by season using model O Burning Index.*

which integrates RMS differences over a range of two standard deviations from the predicted mean (table 2), provides an estimate of each model's robustness.

Models F (fig. 6a) and B produced the lowest RMS (error) value (0.9 ft/min) among seasons for rate of spread, followed closely by model K (1.0) (table 6a). The regression curves for spring and summer were nearly identical, but observed fall values were somewhat lower than those for the other seasons. Model O produced the greatest season-to-season dispersion (RMS = 3.8 ft/min), and models D and E were the next-poorest performers due largely to the departure of the summer season regression (table 6a). Other models (G, H, N, Q, T, and U) performed moderately well from season to season but yielded summer spread rates with steeper slopes than in spring and fall.

Although fuel model O produced the most dispersion with seasonal rates of spread, it, along with B and E, produced the least dispersion (0.3 ft) for seasonal flame lengths (table 6b, fig. 6b). Models G and H failed this test because they did not produce significant fall or summer regression values.

Comparison of variations of the rates of spread by cover type showed that fuel models K (fig. 7a), E, N, and Q yielded the smallest RMS values (0.9 ft/min) (table 6a). Because only models K, N, F, and B produced significant regression equations for upland conifer, data for this cover type were not considered

Table 6a.--A comparative rating of fuel model prediction of rate of spread by various selection criteria[1]

	Criterion	Fuel models											
		B	D	E	F	G	H	K	N	O	Q	T	U
	Means (ft/min & BTU/ft^2) (Table 2)	1.7	17.7	0.2	0.4	1.3	6.5	0.3	52.7[2]	19.4	13.9	18.2	3.5
		+	-	+	+	+	0	+	=	-	0	-	0
(2)	Sensitivity (Table 2)	2.1	11.0	1.9	3.1	1.4	6.8[3]	4.0	33.5	17.5	7.8	12.7	4.7
		+	0	+	0	+	0	0	=	-	0	-	0
(3)	Overall Prediction (R^2) (Table 4)	.089	.109	.131	.092	.137	.118	.140	.131	.083	.106	.089	.115
		-	0	+	-	+	0	+	+	-	0	-	0
(4)	Cover Type Prediction (R^2) (Table 5)	7	19	31	13	32	24	29	30	11	14	6	18
		-	0	+	0	+	0	+	+	0	0	-	0
(5)	Calibration[5]	0	0	0	0	+	-	++	+	-	0	=	0
(6)	Robustness (ft/min & ft)												
	Seasons	.9	2.3[6]	2.4[6]	.9	1.6	1.7	1.0	1.4	3.8[6]	1.9	1.8	1.6
	Cover types[8]	1.2	1.0	.9	1.0	2.4[9]	1.3	.9	.9	1.2	.9	1.1	1.3
	Locations (States)	1.5	1.4	1.4	1.5	1.3	1.4	1.3	1.3	1.5	1.3	1.5	1.4
	Combined robustness	0	-	0	0	-	0	+	+	-	+	-	0
(7)	All-criteria unweighted total	0	-2	+4	0	+4	-1	+6	0	-5	+1	-7	0

Table 6b.--A comparative rating of fuel model prediction of flame length or heat per unit area by various selection criteria

Criterion	Fuel models B	D	E	F	G	H	K	N	O	Q	T	U
(1) Means (ft/min & BTU/ft^2)	83	337	110	318	12	258	305	128	415[2/]	250	405	38
(Table 2)	0	-	0	-	+	0	-	0	-	0	-	+
(2) Sensitivity (Table 2)	807	704	904	894	704	852	747	909	784	779	902	857
	0	+	-	-	+	0	+	-	0	0	-	0
(3) Overall Prediction (R^2)	0.110	0.103	0.129	0.116	0.059[4/]	0.085	0.092	0.129	0.133	0.104	0.112	0.129
(Table 4)	0	0	+	0	=	-	0	+	+	0	0	+
(4) Cover Type Prediction (R^2)	22	10	32	23	4	10	15	26	32	12	20	28
(Table 5)	0	-	+	0	-	-	0	0	+	0	0	+
(5) Calibration[5/]	0	+	0	-	0	=	+	++	+	+	0	-
(6) Robustness (ft/min & ft)												
Seasons	.3	.4	.3	.4	7/	7/	.5	.4	.3	.4	.5	.4
Cover types[8/]	.7	.7	.7	.7	.8	.7	.8	.7	.7	.7	.7	.7
Locations (States)	.5	.5	.5	.5	.6	.5	.5	.5	.5	.5	.5	.5
Combined robustness	+	0	+	0	-	-	-	0	+	0	0	0
(7) All-criteria unweighted total	+1	0	+2	-3	-2	-5	0	+2	+3	+1	-2	+2

1/The methods section gives a detailed explanation of each criterion. The first line in each criterion gives numerical values; the second gives a symbolic ranking relative to the other models (++ above, + marginally above, 0 near average or median, - marginally below, = below).

2/Models N and O are not recommended for general use unless they are first calibrated.
3/Model H had a very limited index range in predicting rate of spread.
4/Model G is not recommended for prediction of flame length.
5/A single mark (+ or -) was used if one coefficient deviated, and double mark if both deviated.
6/Models E and O showed high dispersion for seasonal rates of spread, especially during summer.
7/Models G and H did not produce significant results for either summer or fall flame lengths.
8/Only models B, F, K, and N produced significant results for rates of spread with conifer fires; no models gave significant results for conifer flame lengths.

9/Model G is not recommended for brush fires.

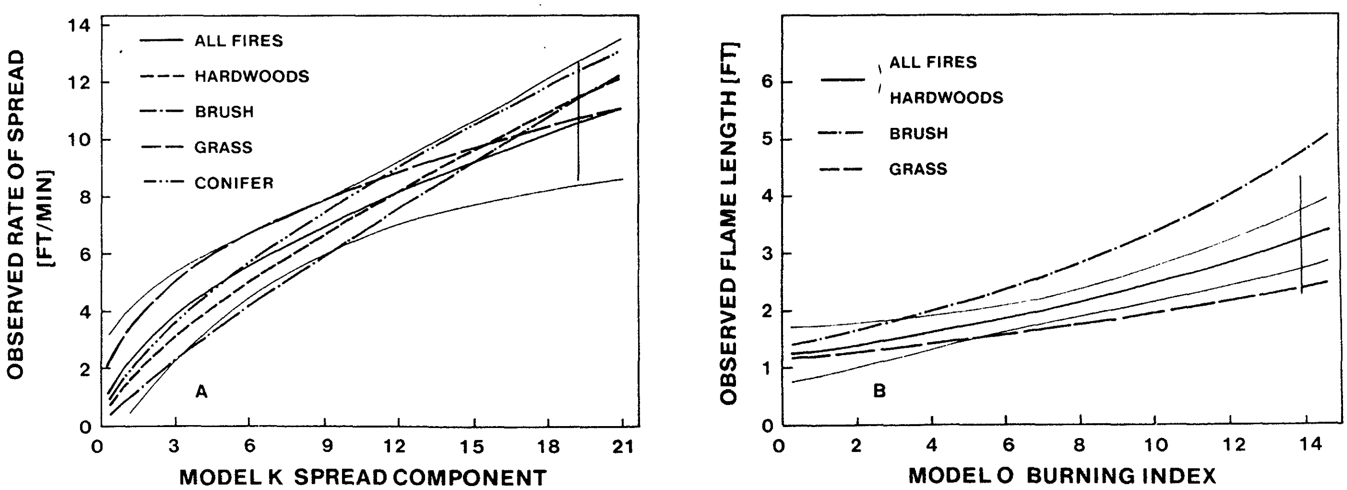


Figure 7.—*(A) Comparison of observed and predicted rates of spread by cover type using model K Spread Component. (B) Comparison of observed and predicted flame lengths by cover types using model O Burning Index. The conifer group did not produce significant results. The regression line representing hardwoods is indistinguishable from all fires.*

in the analysis. The other models yielded RMS values ranging from 1.0 ft/min for models D and F to 2.4 ft/min for model G. For flame length, no fuel models yielded significant regression results for upland conifer. A recomputation of the data using the other cover types showed little difference in RMS values among models (0.7 to 0.8 ft) (table 6b). Regression curves for fuel model O are shown in figure 7b.

Analysis of the data by location showed little dispersion difference for either rate of spread or flame length (RMS = 1.3 to 1.5 ft/min and 0.5 to 0.6 ft, respectively) (fig. 8). The small differences in flame length by location (table 6b) appeared related to differences in cover type and, therefore, the results indicated that location is not a significant factor in evaluating fire-danger rating in much of northeastern United States.

Rating the Models

The 12 models were rated for rate of spread and flame length predictions by six criteria using the following method (table 6). As a first approximation, if the value for a model was within one standard deviation of the mean value for all tested models, it was assigned a 0; if the value was close to or greater than one standard deviation away, it was assigned a + or −, depending on which direction was superior; if close or greater than two standard deviations away, it received a ++ or =.

For criterion (1), we calculated the difference between the model mean and observed mean, and used the standard deviations of the differences to rate the models. The smaller the difference, the better its performance. Thus, for rate of spread, models B, E, F, G, and K ranked higher than the others; D, O, and T ranked lower, and N ranked much lower.

For criterion (2), we calculated the difference between the standard deviation of observed fire behavior and the standard deviation of model predictions. Based on the standard deviation of the differences for rate of spread, B, E, and G ranked notably above average; N, O, and T ranked lower to much lower.

For criterion (3), we calculated the coefficient of determination (R^2) of observed fire behavior for all fires. For rate of spread, four models (E, G, K, and N) were above average and four (B, F, O, and T) were lower.

For criterion (4) we determined the relative rank (first, second, and so forth) of each model's R^2 value for each of three cover types (hardwoods, brush, and grass). Each rank was assigned a point value (first = 12, second = 11, etc.). Points were summed for each model across the three cover types for rate of spread and flame length. For example, model E accumulated 31 points in rate of spread based on the third-highest R^2 value for hardwoods (10 points, table 5), fourth-highest value for brush (9), and highest for grass (12). Upland conifer cover was not considered due to the limited number of significant regressions.

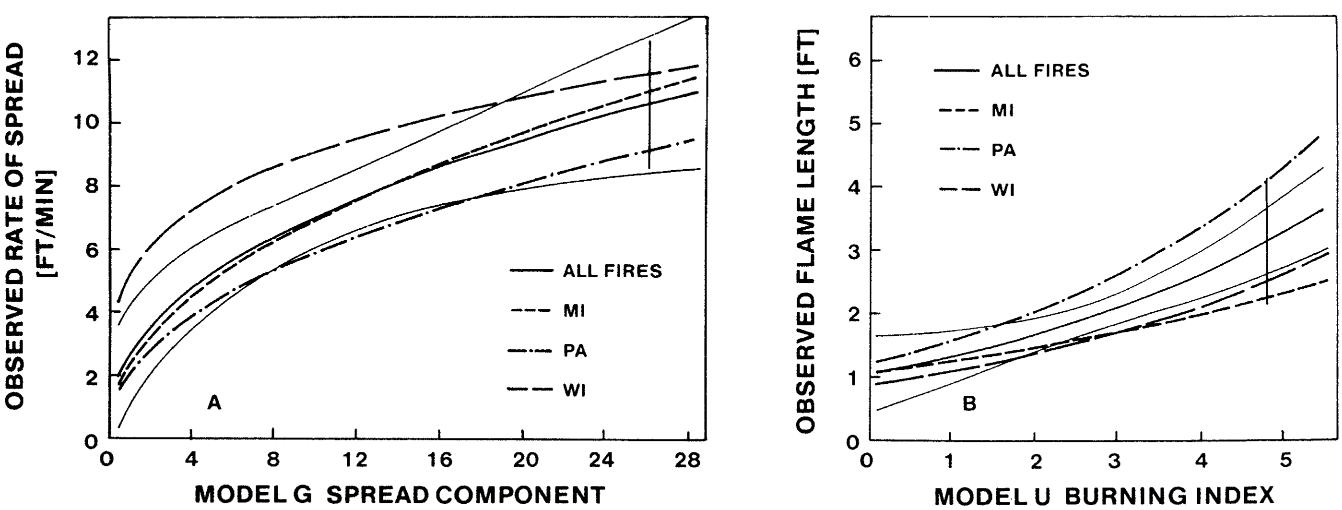


Figure 8.—*(A) Comparison of observed and predicted rates of spread by location using model G Spread Component. (B) Comparison of observed and predicted flame lengths by location using model U Burning Index.*

For criterion (5), the ideal regression coefficients should be $B_0 = 0$ and $B_1 = 1$. We determined the standard deviation of the coefficients and assigned a + or − to each model depending on the closeness of the coefficients to the ideal. A double sign indicates that both coefficients were notably different from average. For rate of spread, both coefficients were better than average for model K. For flame length, both coefficients were poorer than average for model H. Models G and T had compensating differences for flame length.

For criterion (6), we summed the three RMS values for each model and used the standard deviation of the sums to assign a + or −. For rate of spread, three models (K, N, and Q) were superior and four (D, G, O, and T) were inferior.

Various models produced atypical situations for some criteria, and these are noted on the bottom of table 6. At times these situations caused distributions that were decidedly non-normal; consequently, the ratings were adjusted. The best example is the model N value of 52.7 ft/min in criterion (1). In this case model N was assigned a = and deleted from the group. The remaining 11 models were reevaluated without model N.

Composite Rating

If managers have one or two fuel complexes or specific uses, the best NFDRS model choices are indicated by the ratings for individual criteria. If a general-purpose model is needed, however, individual ratings must be combined. One way to combine individual ratings is to simply sum the +'s and −'s. Such a scheme rewards or penalizes only models that were notably better or worse than the average. It emphasizes the distinction between models that passed or failed a test. In contrast, a rank-sum approach ignores the magnitude of differences between test results. For each criterion, we ranked the models from first to last, as we did for the robustness combination, and then calculated the sum of the ranks for all criteria. By using both methods to combine the rankings, we hoped to avoid possible bias associated with the arbitrary choice of one method.

Both of the preceding schemes assumed that all criteria were equally important. In fact, the relative importance placed on each criterion will vary with a manager's needs depending on the final use of the results. We developed a simple weighting scheme for each criterion and used it, along with an unweighted combination, to determine whether or not weights would affect the final ordering of the models. The least weight (w = 1) was assigned to the mean and calibration criteria (1 and 5), because a user can compensate for them. Average weight (w = 2) was assigned to the sensitivity and robustness criteria (2 and 6), and the greatest weight (w = 3) to prediction accuracy criteria (3 and 4).

Each of the four combinations (sign, rank, unweighted, and weighted) was calibrated so that the highest possible score equaled 100 and the lowest equaled zero. Thus, the final rankings indicated not only how well each model performed relative to every other model, but also how well each model performed in an absolute sense. Finally, the four combinations were aggregated into a composite rating by adding the scores and dividing by four (table 7). The calibration ensured that each combination contributed equally to the composite rating. We also considered balance, downgrading models (G, K, and O by 5 points) that did well for one behavior characteristic but poorly for the other. Table 7 also classes the models into five performance groups. In general, models within a higher group did better than models in the next lower group, but differences between closely rated models were slight.

Although the combination schemes were somewhat arbitrarily chosen, the four different procedures yielded remarkably similar results. Eleven of

Table 7.--Relative overall rating of NFDRS models in the northeastern United States[1]

Fuel model	Rating
E	79
K	67
N[2]	64
U	62
B	54
Q	53
G[2]	51
F	46
D	44
O	41
H	33
T	19

[1]Listed in order of estimated overall performance. Differences between models within a group are not considered significant. Ratings are on a 100-point scale.

[2]These models are conditionally acceptable. The major deficiency noted in table 6 (=) must be either compensated for or be unimportant in the use to which the model is put.

the 12 models had a range of three or less between the highest and lowest rank across all four combinations. Thus, the composite rating is not dependent on the characteristics of any single procedure. We therefore conclude that the composite ranking is generic—any reasonable scheme would yield much the same outcome.

A slightly different method of ranking was used by Haines *et al.* 1985 (see their table 5). They calculated a ratio rather than a difference for criterion (2), and also rated only 9 rather than 12 fuel models. These variations in procedures resulted in some changes in final rankings from those shown in table 7. For example, models K and E interchanged first and second positions.

Model E (overall score = 79) is alone in the first group (table 7); it was first in every ranking combination. It was average or better in 11 of the 12 tests (best or second best in 5 tests). Only flame length sensitivity was below average. For most areas in the northeastern United States, model E appears to be the best general-purpose model because it is suited to a wide variety of conditions and applications. Upland conifer forests are a notable exception to model E's overall superiority.

The second group includes three models (K, N, and U). Model K was the best overall spread rate predictor (it was first or second in four of six tests). It was only average at predicting energy release (causing it to be downgraded slightly). Model K was best in brush and upland conifer. The lack of live fuel loading in model K may account for its good performance in the latter case. Its second-highest overall rating, however, appears counterintuitive, but there is a possible explanation. When fuels are green in the Northeast, their moisture content may be so high as to overwhelm other environmental conditions and preclude wildfire occurrence. In fact, in the Northeast observed live fuel moisture typically exceeds that allowed by the NFDRS models by factors ranging from 1.5 to 4 times (Loomis *et al*. 1979, Loomis and Blank 1981). In this situation, living vegetation may be like a switch that turns fires on and off rather than a heat sink that reduces spread rates and intensity. Our behavior data came only from days on which fires occurred and grew to reportable dimensions—days on which high-moisture-content fuels would have to have been cured or nearly so to permit fire spread. Thus, although live-fuel moisture certainly plays a key role in fire behavior, our results suggest that the role is not well defined in this region.

The high intercorrelation between models I, J, and K ($R^2 = 0.99$ in all cases) permits us to extrapolate from model K to I and J. Tables 2 and 4 present the data needed for four of the six tests for models I and J (means, sensitivity, overall predictive accuracy, and calibration). Based on these four criteria, model I would have been marginally poorer than K, and model J would have been marginally better. Greater average spread rates would require calibration prior to use. In essence, if models I or J were desirable for some purpose, they could be substituted for K with confidence.

Although fuel model N ranked within the second group in table 7, it was inconsistent. It was the only model to appear in the top five places in all four fuel groups (table 5). It was best or second best only in the two less important calibration tests, and it was the worst model in three tests. Model N showed significantly substandard performance in the rate-of-spread means and sensitivity tests (table 6a). This problem would have to be either overcome or of little consequence to the user.

Although model U didn't quite reach the rating of models E or K, it is also a good general-purpose model. It appeared in the top five places in three of the four cover groups (table 5). Interestingly, the exception was upland conifer (model U was intended to represent long-needled pine stands—see Methods section). Model U was one of only two models that was below average in only one test (flame length calibration), and it was above average in 3 of 12 tests. In other words, model U is not outstanding, but it appears generally reliable. Because models U and P appear similar in construction and are highly intercorrelated (SC = 0.99 and ERC = 0.92, table 3), P should perform about as well as U.

The next two groups in table 7 include models with lower overall performance. Model B's performance is most like that of U. The former was below model U in five tests, better in four, and tied in three. Like U, model B was second in just one test and it was one of the top fuel models in two fuel groups. Its overall predictive accuracy was next to last in the list of acceptable models (table 4). In general, model B had no outstanding attributes, but few negative ones—in other words, its performance was average. It is most likely that one of the previously recommended models would be superior for most Northeastern applications.

As with model U, Q was a steady performer. It was the only model that was average (10 times) or above (twice) in every test. It never ranked better than fourth nor worse than ninth, and was the best-balanced model tested (spread rate and flame lengths). In summary, model Q can also be classed as average. Given the superior performance of other models, its use should be justified using criteria other than those employed in this study.

Model G's performance might be best classified as a paradox (even more so than with model N). Model G ranked first or second in 5 of 12 tests but 11th or worse four times. It either worked very well or failed miserably! Its predictive accuracy for energy release is considered unacceptable (table 4), yet it was first in the flame length mean and sensitivity tests. It was the second-best spread rate predictor (first or second in three of six tests), but it did poorly (11th) in robustness. Model G was one of the top five models in two fuel types. Although model G cannot be recommended as a general-purpose model, it may be useful as a "specialty" spread rate predictor.

Models D and F both ranked below average more often than above. Both were first through third in just one test; both were never ranked worse than 10th. Both were in the top five models in just one cover type. Both models ranked below average for general use with no notable "specialty" purposes.

Model O was the worst-balanced model tested. It was the best flame length model (first through third in four of six tests) but the worst (of the models considered so far) in rate of spread (10th or worse in every test). Clearly, model O would be a poor general-purpose model but a good energy release "specialty" model.

The last group includes the only two models (H and T) that ranked average or below average in every test. With 10 better models to choose from, it is hard to imagine a general purpose for which these two models would be needed. Models H and T, along with the previously deleted models (A, C, L, R, and S), are not recommended for general use in the Northeast.

Fuel Model Composition-Performance

These data provided an opportunity to postulate causal mechanisms. We considered the relative performance of the 20 fuel models as observational data, and could then conceptually equate fuel model parameters (Bradshaw *et al*. 1983) to experimental treatments. Our task was to learn something about how the NFDRS, with its 114 equations and uncounted assumptions, relates fuel model parameters to fire behavior. The task was complicated by the fact that there are only 20 models to compare when drawing conclusions concerning 12 interrelated treatments with random combinations of treatment levels.[2]

One simple relationship provided some insights. The four lowest ranked models, in terms of overall predictive accuracy (A, L, R, and S—table 4), have 10-hour fuel loadings of 0.5 tons/acre or less (Bradshaw *et al*. 1983, p. 38). Two other models with the same loadings (D and T) finished 9th and 12th, respectively, in our ranking system. Models A, L, R, and S also have extremely light 1-hour loadings (0.5 tons/acre or less), but models D and T have 1-hour loadings of 2.0 and 1.0 tons/acre (within the range of higher ranked models). We therefore suspect that the 1-hour loading is not a limiting factor in poor model performance. We further infer that in the Northeast, light 10-hour loadings (0.5 tons/acre) result in poor-performance fire behavior models. Extrapolation, however, is less certain. Models C and H, for example, yielded the next-poorest predictions and have 10-hour fuel loadings of only 1.0 tons/acre. Model P, on the other hand, also has 1.0 tons/acre of 10-hour fuels and we assume (by association with model U) that it should be a reasonably good predictor.

Four of five models (A, F, L, and T) with a 15 percent moisture of extinction (ME) ranked eighth or lower; B also has a 15 percent ME. It was 14th in terms of predictive accuracy (table 4), but it finished 5th overall based on all tests. Possibly 4 tons/acre of 10-hour fuels helped overcome the 15 percent ME in model B. In general, a 15 percent ME may preclude optimum model performance in the Northeast. Note that the two poorest performance models (A and L) have no 10-hour fuels coupled with a 15 percent ME.

Again, extrapolation is problematic. Models C, H, and U have ME's of 20 percent. Models C and H were not high performers but U was fourth highest. Note, however, that models C and H have 10-hour loadings of 1.0 tons/acre. Perhaps two marginal factors combined to yield poor performance (model P has an ME of 30 percent).

In general, poor model performance appears to be associated with low ME's (20%≤) and low 10-hour fuel loadings (1.0 tons/acre). These results agree with the findings of Simard *et al*. (1984), who found that 10-hour fuel moisture was the best predictor of observed fine fuel moisture. Low 10-hour loadings

[2]*Nineteen parameters are used to define each fuel model, but six are constant and, therefore, do not directly contribute to observed model differences; one only affects a component not tested here.*

reduce the contribution of the 10-hour moisture values to the NFDRS-calculated average fine fuel moisture. They also found that observed moistures were much higher than those calculated by the NFDRS. Hence, a low ME should lead to poor fuel model performance.

In contrast to poor fuel model performance, it is not as easy to identify fuel parameters associated with high model performance. Table 8 lists 10 parameters associated with three of the higher ranked rate of spread predictors (K, E, and G) and energy release predictors (O, E, and U). Insight into a solution requires finding common parameter characteristics within each of the two groups of fuel models. Unique results provide important clues to the adequacy of our methods. For example, model E appears in both groups; thus E should have one or more parameters in common with each group, but there must be at least two groups of parameters—one for rate of spread and one for energy release. Furthermore, model G was a highly ranked rate of spread model but a poor predictor of energy release. Model K was similar but to a lesser extent (best for rate of spread, average for energy release). Model O compliments these results (low for rate of spread, best for energy release). Any favorable characteristics for these unbalanced models on one side of the ledger should be absent on the other.

The model ranking results should provide clear, unambiguous tests, but an analysis of the data in table 8 provides no easy answers. Only one firm conclusion emerges: Resolving the question of what generates good model performance is not likely to be found by simply examining fuel model parameters. Rather, more detailed experiments will have to be made in which the NFDRS is disaggregated below the system level and multiple parameter/model interactions are examined. This is well beyond the scope of the present paper.

Comparative Performances of Fire Activity Predictors

The second step in our validation study compared the performances of various meteorological elements, fuel moisture models, and fire danger indices listed in the Appendix. The predictors of behavior potential that did not yield regression values significant at $P < 0.01$ were eliminated, leaving a shortened list categorized by three groups: meteorological elements, fuel moisture models, and fire-danger indices (table 9). All predictors were compared with observed behavior using functions (2), (3), and:

$$\hat{Y} = B_0 + B_1 X \tag{4}$$

The functional forms listed in table 8 gave the highest R^2 values for each predictor. In all cases standard errors were similar within groups and were higher for rate of spread than for flame length.

Among the meteorological elements, PRECIP and RH produced very low regression values and by themselves appeared to be poor predictors of rate of spread and flame length. Windspeed produced somewhat better results with rate of spread.

The fuel moisture models behaved as expected, yielding comparatively low R^2 values, because they do not incorporate windspeed. The FFM in the 1964-FDRS was the only fuel moisture model that gave statistically significant results when compared with flame length.

The NFDRS values for model E in table 9 are the same as those listed in table 4. Other fuel model values in table 4 can also be compared with fire-danger indices listed in table 9. The fire-danger indices from the Canadian system (ISI, FWI) yielded lower R^2 values than most NFDRS fuel models; the FWI was an especially poor predictor of flame

Table 8.--Fuel model parameters for three higher ranked rate of spread and energy release predictors[1/]

Predictor	Parameter: Fuel load 1-hr	10-hr	100-hr	1,000-hr	Woody	Herbaceous	1-hr SA/V[2/]	Heat content	ME	Fuelbed depth
	- - - Tons/acre						(Ft^{-1})	BTU/lb	Percent	Ft
Rate of spread										
Model K	2.5	2.5	2.0	2.5	--	--	1,500	8,000	25	0.4
Model E	1.5	2.0	.25	--	0.5	0.5	2,000	8,000	25	.6
Model G	2.5	2.0	5.0	12.0	.5	.5	2,000	8,000	25	1.0
Energy release										
Model O	2.0	3.0	3.0	2.0	7.0	--	1,500	9,000	30	4.0
Model E	1.5	2.0	.25	--	.5	.5	2,000	8,000	25	.6
Model U	1.5	1.5	1.0	--	.5	.5	1,750	8,000	20	.5

1/Includes only those parameters that vary among the highest ranked models.
2/Ratio of surface area to volume.

Table 9.--Comparative performance of components[1] of fire-danger rating systems based on predictions of rate of spread and flame length using windspeeds and slopes from fire sites

Fire activity predictor	Rate of spread			Flame length		
	R^2	$S_{y \cdot x}$	Mathematical function	R^2	$S_{y \cdot x}$	Mathematical function
Meteorological elements						
WS	0.069	9.3	(3)	NS	NS	
RH	.032	9.4	(3)	0.022	2.2	(3)
Fuel moisture models						
FFM	.022	9.5	(3)	.052	2.2	(3)
FM_1	.031	9.4	(3)	NS	NS	
FFMC	.033	9.4	(3)	NS	NS	
Fire-danger indices						
SC & BI (Model E)	.131	9.1	(2)	.129	2.1	(3)
TSI	.105	9.2	(2)	.128	2.1	(3)
FFSI	.097	9.2	(2)	.134	2.1	(3)
ISI	.094	9.2	(2)	.105	2.1	(4)
FLI	.083	9.3	(2)	.110	2.1	(4)
FWI	.073	9.3	(3)	.034	2.1	(4)

[1] Only those predictors with significant regression statistics are included.

length. These results may reflect the fact that the Canadian system was designed for a specific fuel type—jack pine and lodgepole pine (Appendix)—and our data base included only 6 percent upland conifers (table 1). The fire-danger indices from the 1964-FDRS, the TSI, and the FFSI produced similar results. The FFSI gave the highest R^2 value for flame length of any index tested, although the index was not significantly better.

No index that measures drought or long-term drying produced significant results. This included the longer timelag fuel moistures in the NFDRS, the DMC, ADMC, and the DC in the Canadian system, the BUI in the 1964-FDRS, and the KBDI.

SUMMARY AND CONCLUSIONS

We compared the performance of 20 NFDRS fuel models in predicting potential rate of spread and flame length for 940 wildfires in the Northeast, and evaluated the system as it is used operationally. Standard weather station data were used as inputs, and supplemental fire behavior data were compared with system outputs. Model E was the best general-purpose model; models K, N, and U ranked slightly lower. Other models, such as G and O, performed well in specific areas, but were substandard in others. In selecting a model, a manager will have to decide which fuels or test criteria are most important and choose the fuel model on that basis.

Our results confirmed the general belief that no single fuel model is best for all purposes. In fact, one model (O) was best at predicting flame length but second worst at predicting rate of spread. We also found that fuel model performance did not necessarily reflect the qualitative descriptions that are often used for model selection. For example, fuel models A and L were among the poorest predictors of fire behavior in grass. The performance of some models was also counterintuitive. For example, the second-best overall fuel model (K) does not include live-fuel moisture. Such anomalies indicate that although we may understand some fire behavior processes, we are weak in understanding how they all interact in the complex situation of fire-danger rating. Our results suggest that poor fuel model performance in the Northeast may be related to low 10-hour fuel loading (1.0 ton/acre ≤) and low moisture of extinction (20%≤).

Since the release of the 1978 revision of the NFDRS, some Northeastern fire managers have questioned the applicability of any of the 20 fuel models to their specific areas and suggested that they need additional models. Given the spatial coverage of fuel models in figure 1, it is difficult to see what other range of activity values might be included. If the heat per unit area falls much below 100 BTU/ft^2, fires will not generate enough energy to maintain themselves. Fires burning at an intensity much above 1,000 BTU/ft^2 usually display erratic characteristics and, by definition, the NFDRS is not applicable in that situation. Nine fuel models predicted higher average rates of spread than occurred in two-thirds (one standard deviation) of observed fires. If anything, the models display more redundancy and overlap than gaps on the fire characteristics chart. We suspect that many managers really need a fire behavior fuel model rather than a fire-danger planning model.

An examination of indices from present and past fire-danger rating systems, fuel moisture models, and drought codes compared their capability to predict the potential of two measures of fire behavior. Eleven NFDRS fuel models yielded comparatively higher R^2's than any other index tested for rate of spread. The Fine Fuel Spread Indices from the 1964-FDRS produced about the same R^2 values for flame length prediction as the three best NFDRS fuel models. Thus, over the past two decades we appear to have improved our ability to predict rate of spread, but have made little progress in predicting energy release rate.

Our results indicate that, although the NFDRS works in the Northeast, there are overall performance problems. In large part, this is due to the inherent temporal and spatial variability of fire behavior. Given such variability, however, one may ask: How good can any practical system be? At what point does system complexity exceed the natural and operational limits of resolution? Only by studying fire-danger rating from the perspective of the system as a whole can such questions be resolved. In taking such an approach, we attempted to shed new light on perhaps the oldest topic in wildland fire research—yet, a topic which remains unresolved after 75 years of study.

ACKNOWLEDGMENT

We thank John E. Deeming, Richard C. Rothermel, and Peter J. Roussopoulos, USDA Forest Service; Carl W. Ramm, Michigan State University; and C. E. Van Wagner, Canadian Forestry Service, for their helpful comments and suggestions.

LITERATURE CITED

Andrews, Patricia L. Testing the fire behavior model. In: Proceedings, 6th conference on fire and forest meteorology; 1980 April 22-24; Seattle, WA. Washington, DC: Society of American Foresters; 1980: 70-77.

Andrews, Patricia L.; Rothermel, Richard C. Charts for interpreting wildland fire behavior characteristics. Gen. Tech. Rep. INT-131. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station; 1982. 21 p.

Bradshaw, Larry S.; Deeming, John E.; Burgan, Robert E.; Cohen, Jack D. The 1978 national fire-danger rating system: technical documentation. Gen. Tech. Rep. INT-169. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station; 1983. 44 p.

Deeming, John E.; Burgan, Robert E.; Cohen, Jack D. The national fire-danger rating system—1978. Gen. Tech. Rep. INT-39. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station; 1977. 63 p.

Deeming, John E. Reflections on the development, application and future of the national fire-danger rating system. In: Proceedings, 7th conference on fire and forest meteorology; 1983 April 25-28; Fort Collins, CO. Boston, MA: American Meteorological Society; 1983: 139-146.

Haines, Donald A.; Johnson, Von J.; Main, William A. An assessment of three measures of long term moisture deficiency before critical fire periods. Res. Pap. NC-131. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station; 1976. 13 p.

Haines, Donald A.; Main, William A.; Frost, John S.; Simard, Albert J. Fire-danger rating and wildfire occurrence in the northeastern United States. For. Sci. 29: 679-696; 1983.

Haines, Donald A.; Main, William A.; Simard, Albert J. Operational validation of the National Fire-Danger Rating System in the Northeast. In: Proceedings, 8th Conference on fire and forest meteorology; 1985 April 29-May 2; Detroit, MI. Boston, MA: American Meteorological Society; 1985: 169-177.

Johnson, Von J. The dilemma of flame length and intensity. Fire Manage. Notes 43(4): 3-7; 1982.

Keetch, John J.; Byram, George M. A drought index for forest fire control. Res. Pap. SE-38. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southeastern Forest Experiment Station; 1968. 32 p.

Lewis-Beck, Michael S. Applied regression—an introduction. Pap. Ser. 07-022. Quantitative applications in the social sciences; Beverly Hills, CA: Sage Publ.; 1980. 79 p.

Loomis, Robert M.; Roussopoulos, Peter J.; Blank, Richard W. Summer moisture contents of understory vegetation in northeastern Minnesota. Res. Pap. NC-179. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station; 1979. 7 p.

Loomis, Robert M.; Blank, Richard W. Summer moisture content of some northern lower Michigan understory plants. Res. Note NC-263. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station; 1981. 4 p.

Main, William A.; Haines, Donald A. Determining appropriate fuel models from field observations and a fire characteristics chart. In: Proceedings, 7th conference on fire and forest meteorology; 1983 April 25-28; Fort Collins, CO. Boston, MA: American Meteorological Society; 1983: 47-52.

Neter, John; Wasserman, William. Applied linear statistical models. Homewood, IL: Richard D. Irwin, Inc.; 1974. 842 p.

Payandeh, Bijan. Choosing regression models for biomass prediction equations. For. Chron. 57(5): 229-232; 1981.

Reifsnyder, William E. Systems for evaluating and predicting the effects of weather and climate on wildland fires. WMO Special Environ. Rep-496. Geneva, Switzerland: World Meteorological Organization; 1978. 40 p.

Rothermel, Richard C. A mathematical model for predicting fire spread in wildland fuels. Res. Pap. INT-115. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station; 1972. 40 p.

Rothermel, Richard C. How to predict the spread and intensity of forest and range fires. Gen. Tech. Rep. INT-143. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station; 1983. 161 p.

Rothermel, Richard C.; Rinehart, George C. Field procedures for verification and adjustment of fire behavior predictions. Gen. Tech. Rep. INT-142. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station; 1983. 25 p.

Simard, Albert J. Forest fire weather index data—reference manual and station catalogue. Inf. Rep. FF-X-32. Ottawa, ON: Canadian Forestry Service, Forest Fire Research Institute; 1972. 28 p.

Simard, Albert J. AIRPRO: An air tanker productivity computer simulation model. Inf. Rep. FF-X-64. Ottawa, ON: Canadian Forestry Service, Forest Fire Research Institute; 1977. 339 p.

Simard, Albert J.; Deacon, Arthur G.; Adams, Kenneth B. Nondirectional sampling of wildland fire spread. Fire Tech. 18: 221-228; 1982.

Simard, Albert J.; Eenigenberg, James E.; Blank, Richard W. Predicting fuel moisture in jack pine slash: a test of two systems. Can. J. For. Res. 14: 68-76; 1984.

U.S. Department of Agriculture, Forest Service. National fire-danger rating system handbook. For. Serv. Handb. 5109.11. Washington, DC: U.S. Department of Agriculture, Forest Service; 1964: 100-300.

Van Wagner, C. E. Structure of the Canadian forest fire weather index. Publ. 1333. Ottawa, ON: Canadian Forestry Service, Petawawa Forest Experiment Station; 1974. 44 p.

van Wagtendonk, Jan W.; Botti, Stephen J. Modeling behavior of prescribed fires in Yosemite National Park. J. For. 82: 479-484; 1984.

APPENDIX

FIRE-DANGER RATING SYSTEMS

The following gives a brief overview of the various systems discussed in this paper and defines the terms used. Although the systems are designed to categorize a number of physical features, the following discussion concentrates only on the components that contribute to the examined predictors of potential fire behavior. Complete discussions are available in the original publications documenting each system.

The 1978 National Fire-Danger Rating System (NFDRS)

Deeming *et al*. (1977) state that this system relates only to an initiating fire—i.e., a fire that does not behave erratically. The NFDRS addresses only those aspects of fire activity involving occurrence and behavior. Aspects of containment such as accessibility, soil condition, and resistance to line construction must be evaluated by other means.

Ratings are relative, not absolute. Indices and their components were designed to be linearly related to the fire problem. This design specified that when the value of the component or index doubled, the rated activity also doubled. The basic observations that implement the system are taken once a day when fire danger is normally the highest, in the open, at midslope, on southerly or westerly exposures. The system, therefore, evaluates the worst conditions of fire danger that might occur on a rating area during the rating period.

The NFDRS begins by estimating four classes of dead and two classes of live fuel moisture (FM): 1-, 10-, 100-, and 1,000-hour timelag, woody, and herbaceous. These moistures are combined with windspeed to produce three fire behavior components: ignition (IC), spread (SC), and energy release (ERC). The SC and the ERC combine to produce a burning index (BI).

The FM_1 (1-hour timelag fuel moisture) estimates the moisture content of the fine fuels. These fuels consist of dead herbaceous plants and roundwood less than 1/4 inch in diameter. This also includes the uppermost layer of litter on the forest floor.

The FM_{10} (10-hour timelag fuel moisture) estimates the moisture content of dead fuels consisting of roundwood 1/4 to 1 inch in diameter and, very roughly, the layer of litter extending from just below the surface to 3/4 inch below the surface.

The FM_{100} (100-hour timelag fuel moisture) estimates the moisture content of dead fuels consisting of roundwood of 1 to 3 inches in diameter and the forest floor from 3/4 inch below the surface to 4 inches below the surface.

The FM_{1000} (1,000-hour timelag fuel moisture) estimates the moisture content of dead fuels consisting of roundwood 3 to 8 inches in diameter and the forest floor more than 4 inches below the surface.

The IC gives a rating of the probability that a firebrand will ignite wildland fuels, causing a fire that requires suppression action.

The SC is a rating of the forward rate of spread of a head fire and is scaled directly to rate of spread in feet per minute.

The ERC yields a rating related to the available energy per unit area within the flaming front at the head of a fire. It is equal to the heat per unit area (BTU/ft^2) divided by 25.

The BI gives a rating related to the contribution of fire behavior to the effort of containing a fire. It is equal to 10 × flame length (ft).

The Canadian Forest Fire Weather Index (FWI)

According to Van Wagner (1974), the designers' major goal was a fire-danger index that would yield uniform results across Canada and be based on once-a-day weather observations. The system is designed for a specific fuel type—jack or lodgepole pine.

This empirically derived system consists of six components. There are three moisture codes: fine fuel moisture (FFMC), duff moisture (DMC), and drought (DC). These moisture codes are combined and windspeed is added to form two intermediate components—the initial spread index (ISI) and the adjusted duff moisture code (ADMC)—and a final product, the fire weather index (FWI). This Canadian system and the system used in Australia appear to be similar (Reifsnyder 1978). The FWI is also used operationally in New Zealand.

The FFMC represents the moisture content of litter and other cured fine fuels in a forest stand in a layer of about 0.05 lb/ft^2 dry weight.

The DMC represents the moisture content of loosely compacted, decomposing organic matter 2 to 4 inches deep and weighing about 1 lb/ft^2 when dry.

The DC represents a deep layer of compact organic matter weighing about 10 lb/ft^2 when dry.

The ISI combines wind and the FFMC to represent rate of spread, without the influence of variable quantities of fuel.

The ADMC combines the DMC and the DC to represent the total fuel available to the spreading fire.

The FWI combines the ISI and the ADMC to represent the intensity of the spreading fire as energy output rate per unit length of fire front.

1964-National Fire-Danger Rating System (1964-FDRS)

The 1964-FDRS was the first attempt at designing a national system, but this system did not progress beyond the spread phase (USDA Forest Service 1964). It estimates fine-fuel moisture (FFM) and combines it with windspeed to yield a Fine Fuel Spread Index (FFSI). Another component of the system estimates a Timber Spread Index (TSI) and a slow-response estimate of drying called a Buildup Index (BUI).

The FFM uses the stage of the herbaceous vegetation along with the ambient temperature and relative humidity to produce an analog of moisture in fine fuels.

The FFSI provides a measure of the relative rate of forward movement of surface fires in light fuels.

The TSI provides a measure of the relative rate of forward movement of surface fires in heavier fuels.

The BUI provides a measure of the progressive drying of fuels and, according to the system designers, is related to the moisture content of 10-day timelag fuels. It appears to be more representative of fuels having a 5-day timelag constant (Haines *et al*. 1976).

The Keetch-Byram Drought Index (KBDI)

Fire-danger rating systems in the United States have never used a slow-response drought index. Keetch and Byram (1968) developed this index to provide fire management with a scale of reference for estimating deep-drying conditions in areas where such information was needed for fire suppression.

The KBDI uses daily maximum temperatures and 24-hour precipitation amounts to estimate the precipitation necessary to return the soil to full field capacity.

☆ U.S. GOVERNMENT PRINTING OFFICE : 1986—654-948/40010

Haines, Donald A.; Main, William A.; Simard, Albert J.
Fire-danger rating and observed wildfire behavior in the Northeastern United States. Res. Pap. NC-274. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station; 1986. 23 p.

Compares the 1978 National Fire-Danger Rating System and its 20 fuel models, along with other danger rating systems, with observed fire behavior and rates the strengths and weaknesses of models and systems.

KEY WORDS: Fuel models, flame length, rate of spread.