
Miguel G. Cruz¹, Martin E. Alexander², and Ronald H. Wakimoto³

Abstract—Model evaluation should be a component of the model development process, leading to a better understanding of model behavior and an increase in its credibility. In this paper a model evaluation protocol is proposed that encompasses five aspects: (1) model conceptual validity, (2) data requirements for model validation, (3) sensitivity analysis, (4) predictive validation (incorporating statistical tests), and (5) model comparison. The proposed protocol was applied to evaluate fire behavior models that were developed to predict crown fire initiation and spread with potential application in fire management decision support systems. The evaluation protocol highlighted the limitations and the distinct behavior of the specific models and the implications of such limitations when applying those models to support fire management decision-making. The model limitations identified through these results helped the authors to characterize deficiencies in the state-of-knowledge of the determinant processes involved in crown fire behavior, thereby identifying pertinent research needs.

Introduction

Advances in fire behavior science have gradually resulted in the development of fire models to support the decision-making of land managers in a large array of fire management problems (Cohen 1990). The use of such models allows managers to reduce the uncertainties associated with applying fire as a management tool and facilitates proactive management. The complexities associated with wildland fire phenomenology results in a large number of unknowns. These limitations lead researchers to model a specific phenomenon as they rely on distinct simplifying assumptions and include different independent fire environment properties as driving variables. This results in a situation where distinct models attempting to describe the behavior of a particular process respond differently to a given set of conditions. This is noticeable in evaluations of fuel treatment effectiveness in reducing fire potential. Model outputs might be misleading and result in misguided management. Modification of the structure of the fuel complex has been the main, if not the only, method by which fire managers can reduce the fire potential of a given area (Countryman 1974). When the changes that a fuel treatment causes in fire behavior are evaluated, three approaches can be identified: (1) analysis of post fire damage in adjacent treated and untreated stands burned under similar burning conditions (e.g., Weatherspoon and Skinner 1995, Pollet and Omi 2002); (2) monitoring of changes in various fire behavior determinants, such as diurnal and seasonal fuel moisture dynamics, the vertical wind profile, fuel available for combustion, overall fuel complex structure, and experimental fires burns in the various study plots to assess differences in fire behavior.
behavior (e.g., Alexander et al. 1991; Gould et al. 2001; Alexander and Lanoville 2002); and (3) fire behavior simulations integrating fuel descriptors and critical fire weather parameters (e.g., Kalabokidis and Omi 1998, Hirsch and Pengelly 1999).

Due to the nature and complexity of the interactions defining a particular fire environment, it is unlikely that a fire environment monitoring program, unless it was extremely comprehensive, would capture all the variability that would be present in a particular fuel treatment setting. Although this approach would produce the most realistic data, the associated cost limits its applicability. Hence, most evaluations of fuel treatment effectiveness are based on simulations of potential fire behavior that rely on fire behavior models. This approach allows integrating the effects of the determinant input variables through their spectrum of variability. In addition, infrequent combinations of particular fire environment variables and certain fire behavior processes that are not easily measured in the field can accordingly be analyzed. Nevertheless, our understanding of the processes and interactions in the system through the use of models needs to be accompanied with a thorough description of the assumptions and main limitations of the models employed.

Most of the studies that have analyzed the effects of silvicultural/fuel treatment practices on fire hazard at the stand level rely solely on changes in the structure of the fuel complex to infer potential fire behavior (e.g., Fahnestock 1968, Alexander and Yancik 1977, van Wagtendonk 1996, Kalabokidis and Omi 1998, Scott 1998b, Stephens 1998, Hirsch and Pengelly 1999, Fulé et al. 2001, Fiedler et al. 2001). In general, the fire simulations that support these studies have not taken into account the possible effects of changes in fuel moisture gradients and sub-canopy wind flow between treated and untreated situations. Some of these studies produced conflicting results due to degree of comprehensiveness taken in modeling a particular situation and the nature and characteristics of the fire behavior models used to carry out the simulations.

The objective of this paper is to define a model evaluation protocol that results in a better understanding of the use of fire behavior models in these applications, including their limitations and biases, which we hope will eventually increase their credibility when used in evaluating effectiveness of fuel treatments for reducing fire behavior potential.

Defining Criteria for Model Evaluation

We believe that testing and evaluation of models should be a fundamental component of the model development process. These activities assume particular importance in fire behavior science due to the inherent difficulties in measuring and understanding some of fire's determinant processes. In spite of this, model evaluation has not received much emphasis by fire behavior modelers and no comprehensive model evaluation protocol has been applied previously to fire behavior models. Most fire behavior model evaluation that has been done has been limited to two areas: (1) comparisons of model predictions with observed and measured fire behavior data (e.g., Brown 1972, Lawson 1972, Bevins 1976, Sneeuwjagt and Frandsen 1977, Hough and Albini. 1978, Andrews 1980, Brown 1982, Norum 1982, Rothermel and Rinehart 1983, van Wagtendonk and Botti 1984, Albini and Stocks 1986, van Wilgen and Wills 1988, Gould 1991, Hirsch 1989, Marsden-Smedley and Catchpole 1995, Alexander 1998, Grabner et al. 2001); and (2) sensitivity analysis studies.
Critical examination of these studies reveals that they typically use subjective and/or qualitative measures of model performance and often lack validation standards. The process of model evaluation has been approached differently by authors due in part to philosophical interpretations as to what constitutes a model. Models cover a large spectrum of idealizations and complexity, ranging from a model being considered as a mathematical interpretation of a theory/hypothesis bound by certain assumptions to models that are regarded as simple algebraic expressions that reflect a certain dataset. An important aspect to consider in model evaluation is the definition of the criteria that should be followed, which depends on the type of model being evaluated and its potential application. When fire behavior science is considered, theoretical models developed to understand certain physical and chemical phenomena (e.g., Grishin 1997, Linn 1997) should be evaluated in a different manner than models built to support fire management decision-making (e.g., Rothermel 1972, Albini 1976, Alexander 1988, Forestry Canada Fire Danger Group 1992). Theoretical models can be viewed as mathematical descriptions of certain physical fire phenomena and so the emphasis on evaluation should be placed on behavioral validation instead of the numerical precision of the result (Bossel 1991). The unknowns involved in some heat transfer and fluid dynamics processes coupled with our inability to reliably measure certain fundamental fire quantities limit any attempt to evaluate these models by direct comparison with independently derived data. In this paper, we emphasize the latter group of models—i.e., simple empirical or semi-empirical models (Catchpole and de Mestre 1986) that combine physical laws with empirical data to generate important model components describing particular physical fire phenomena. When these models are integrated, they produce workable models or systems that can be used to support operational fire management decision-making.

We consider model evaluation as follows: “The substantiation of a model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model” (Sargent 1984). Based on an examination of the previously referred fire behavior model evaluation studies and other validation studies of engineering and ecological models (e.g., Mayer and Butler 1993, Oderwald and Hans 1993, Bacsí and Zemankovic 1995), a model evaluation protocol based on the approaches of Rykiel (1996) and Sargent (1984) was defined for the purposes of the present work. This protocol includes the following analyses:

**Model conceptual validity**—This involves analysis of the conceptual structure and logic of a model. By taking into account the intended use of the model, we aim to determine the validity of the model’s various theories and simplifying assumptions in capturing the dynamics of the system.

**Data validation**—This includes definition of data quality standards, namely the selection of real world data that represent the phenomena of interest for use in predictive validation, statistical validation, and model comparison. This aspect assumes particular importance when analyzing fire behavior data due to the relative inaccuracy and bias that arise from inherent difficulties in capturing reliable fire behavior data from either experimental fires, operational prescribed fires, or wildfires.

**Sensitivity analysis**—This consists of analyses to reveal the relative influence of model components and input parameters on the behavior of the model overall. This also includes identification of parameters that cause minor or major fluctuations in model outputs.
Predictive validation—This involves comparison of model outputs with an independent dataset of the phenomena under study in order to evaluate model suitability to predict system behavior. This section of the model evaluation incorporates a variety of statistical tests conducted in order to quantitatively evaluate model performance.

Model comparison—This involves a comparison of outputs from several models describing the same phenomena thereby providing an understanding of possible model deficiencies and the limits of applicability. In the present study a comparison of model behavior and validity was done concurrently with the predictive and statistical validations.

In order to better understand the evaluation protocol being proposed, we selected a case study application of the methodology to crown fire initiation and spread models. The models analyzed were Van Wagner (1977), Alexander (1998) and Cruz (1999) for crown fire initiation; and Rothermel (1991), Forestry Canada Fire Danger Group (1992) and Cruz (1999) for crown fire spread rate. The formulation of the two Cruz (1999) models is based on future research needs suggested by Alexander (1998). For the Canadian Forest Fire Behavior Prediction (FBP) System (Forestry Canada Fire Danger Group 1992), we chose fuel type specific fire spread models that can be applied to fuel types in the western United States, namely FBP System Fuel Type C-3 [mature jack (Pinus banksiana) or lodgepole pine (Pinus contorta)], C-6 [Conifer plantation] and C-7 [ponderosa pine (Pinus ponderosa)/Douglas-fir (Pseudotsuga menziesii)]. The emphasis of the present study is on establishing a model evaluation protocol rather than a comprehensive analysis of all available models (e.g., Xanthopoulos 1990, Van Wagner 1993). Some fire behavior prediction systems commonly used to evaluate fuel treatments, such as FARSITE (Finney 1998) and NEXUS (Scott 1999), integrate some of the models mentioned above with their specific interpretations of certain fire behavior processes. As a result, the models in these systems can to a certain extent be viewed as distinct from the original model formulations being analyzed here. Nevertheless, the present study is concerned with the core models driving the systems (i.e., crown fire rate of spread and initiation models). We have assumed that those particular interpretations imbedded in FARSITE and NEXUS do not induce significant changes in the outputs to warrant inclusion in the present comparison. Consequently, the evaluation procedures throughout this study are applied solely to the aforementioned models.

Model Conceptual Validity

The analysis of the validity of model theories and assumptions is of particular importance in models intended to support fire management decision-making. Inappropriate use of the models could lead to detrimental and long lasting effects on the ecosystem. Given the unknowns in fire science and difficulty in correctly measuring certain fundamental fire quantities and processes, any analysis of the theories and simplifying assumptions embodied in a model is limited by the current state-of-knowledge, availability of complete and reliable data sets, and our inability to propose more realistic theories. A detailed examination of the underlying theories of the various models considered here is outside of the scope of this paper. Consequently only a discussion of the most restrictive assumptions of the models will be presented.

A brief introductory description of the various models follows. See table 1 for a listing of the inputs involved in the crown fire initiation models. Both the
Van Wagner (1977) and Alexander (1998) models are based on convective plume theory (cf. Yih 1953) and air temperature (as a surrogate of convective heat flux) decay with height above a linear heat source. The outputs of both the Van Wagner (1977) and Alexander (1998) models are deterministic in nature. The canopy ignition requirements are stated in terms of a critical fireline intensity (as per Byram 1959), as a function of a heat sink (evaluated differently in the two models), and height above the ground. The Alexander (1998) model, although based on the same convective plume theory, relies on a more realistic heat source and heat sink definition and takes into account the interaction between the convective plume and the cross wind with the subsequent tilting of the plume and dilution of the hot gases in the plume. The Cruz (1999) crown fire initiation model is the result of an application of logistic regression analysis to an experimental database (n = 73) and involved both surface and crown fires used in the development of the Canadian FBP System (Forestry Canada Fire Danger Group 1992) in order to predict the onset of crowning. Within this model the vertical stratification of the fuel complex was slightly modified from the commonly accepted definition of a canopy base height (CBH) to the definition of a new fuel complex descriptor, namely fuel strata gap (FSG). This parameter incorporates the effect of dead aerial fuels (canopy base height is associated with live canopy fuels) in reducing the distance between surface fuels and ladder fuels that can support vertical fire propagation. The introduction of this variable was justified by the dependence of canopy base height on just live foliage (Sando and Wick 1972, Kilgore and Sando 1975, McAlpine and Hobbs 1994, Ottmar et al. 1998) and consequently not incorporating the effect of dead fuels in lowering the distance between the surface and canopy fuel layers.

The Rothermel (1991) crown fire rate of spread model is the result of a simple average correction factor relating predicted surface fire behavior by the BEHAVE system (Andrews 1986) using Fuel Model 10 (Anderson 1982) with wind speed adjustment factor set equal to 0.4 and a series of observations of spread rates garnered from wildfires (n = 8). This model does not incorporate any stand structure descriptor in its formulation. The Canadian FBP System models for predicting crown fire rate of spread are based on a sigmoid equation taking into account certain components of the Canadian Forest Fire Weather Index System (Van Wagner 1987) that relate to the potential for fire spread as determined by the moisture content of fine fuels and wind speed. The Cruz (1999) crown fire spread model is based on non-linear regression analysis parameterized using data from high intensity experimental crown fires. This model differentiates crown fires spreading as either continuous or intermittent based on Van Wagner’s (1977) spread rate criterion for active crowning.

**Table 1—Input requirements for the crown fire initiation models being evaluated.**

<table>
<thead>
<tr>
<th>Model</th>
<th>Fireline intensity</th>
<th>FMC</th>
<th>Wind speed</th>
<th>Residence time</th>
<th>CBH/FSG</th>
<th>EFFM</th>
<th>SFC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Van Wagner (1977)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alexander (1998)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cruz (1999)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1 FMC= foliar moisture content; CBH= canopy base height; FSG = fuel strata gap; EFFM = estimated fine dead fuel moisture content; and SFC = surface fuel consumption.
Data Validation

The very nature of crown fires leads to various difficulties, either institutional, social, or experimental, in acquiring reliable outdoor fire behavior data (Alexander and Quintilio 1990) for model development, calibration, and evaluation. Fire behavior data can be gathered from operational prescribed fires or outdoor experimental fires (e.g., Stocks 1987, 1989) and wildfires (e.g., Alexander and Lanoville 1987). Both types of data sources have inherent limitations. The spectrum of fire environment conditions covered by experimental fire data normally does not include extreme fire weather conditions (e.g., high winds or severe drought). Furthermore, data garnered from wildfires is often incomplete or lacks detail and may not be reliable. Two different types of data were used in the present study: (1) fire behavior data in order to perform the predictive validation and then apply statistical tests and (2) weather and fuel complex data for use in the model sensitivity analysis and model behavior comparisons. The use of experimental fire behavior data would be appropriate in the current evaluation exercise given the generally high reliability that characterizes such data. Both the Cruz (1999) crown fire initiation and crown fire rate of spread models were developed using an extensive experimental fire dataset from published and unpublished sources in Canada supplemented with a few observations from Australia. Unfortunately no experimental crown fire behavior datasets are currently available for use in independently model testing. Consequently predictive validation and statistical tests will be based on wildfire data garnered from published cases studies (Alexander et al. 1983, Simard et al. 1983, Lanoville and Schmidt 1985, De Groot and Alexander 1986, Rothermel and Mutch 1986, Alexander and Lanoville 1987, Stocks 1987, Stocks and Flannigan 1987, Hirsch 1989, Alexander 1991). The use of such data will limit evaluation procedures applied to the Canadian FBP System fuel type specific models, as these were originally parameterized with data from some of those same documented wildfires.

The definition of fire environment scenarios used to characterize baseline data for sensitivity analysis and model comparison was based on well documented burning conditions where all relevant input variables were measured or acceptably estimated. This reliable and compatible data constitutes a benchmark dataset for which model behavior can be compared. For the sensitivity analysis

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>10-m open wind speed (km/h)</td>
<td>29</td>
<td>37</td>
<td>15</td>
</tr>
<tr>
<td>Within stand wind speed (km/h)</td>
<td>-</td>
<td>-</td>
<td>5</td>
</tr>
<tr>
<td>1-hr TL FM (%)</td>
<td>8</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>10-hr TL FM (%)</td>
<td>9</td>
<td>6</td>
<td>11</td>
</tr>
<tr>
<td>100-hr TL FM (%)</td>
<td>10</td>
<td>7</td>
<td>12</td>
</tr>
<tr>
<td>Live woody FM (%)</td>
<td>75</td>
<td>75</td>
<td>-</td>
</tr>
<tr>
<td>FMC (%)</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Surface fuel consumption (kg/m²)</td>
<td>-</td>
<td>-</td>
<td>1 - 2</td>
</tr>
</tbody>
</table>

1 The 1-hr time lag (TL) fuel moisture (FM) content values were estimated according to the procedures described in (Rothermel 1983).
2 The 10 and 100-hr TL FM content values were assigned values of plus one and two percent points of the value of the 1-hr TL FM as per Rothermel (1983).
analysis, two distinct fire scenarios were chosen: one conducive to marginal crown fire activity (i.e., Kenshoe Experimental Fire 5 documented by Stocks 1989) and the other conducive to extreme crown fire behavior (i.e., 1979 Lily Lake Fire as described by Rothermel 1983 and Alexander 1991). The use of two fire scenarios (table 2) is justified due to multiple interacting factors within the models being evaluated and possible non-linear effects in fire behavior.

The nature of the distinct modeling approaches that characterize the crown fire initiation models under evaluation with their distinct input requirements (table 1) constrains the type of predictive validation analysis and inter-model comparison that can be applied. The evaluation of the Van Wagner (1977) crown fire initiation model requires the monitoring of the transition period to estimate the critical fireline intensity, which is very difficult to accomplish in an experimental fire (Alexander 1998). The inter-model comparison for crown fire initiation models requires the use of a fire situation where the distinct input variables have been simultaneously measured. There is a scarcity of such data in the published literature, one of the limiting factors being the absence of data describing the vertical wind speed profile. The early experimental crown fires carried out at Petawawa Forest Experiment Station (PFES), Ontario, Canada (Van Wagner 1977) offers a complete description of the fuel complex and fire environment characteristics suiting the various model input requirements. Published data from PFES Experimental Fire R1 (Van Wagner 1968, 1977) suits the present data requirements and will be used in the evaluation of the various crown fire initiation models.

**Sensitivity Analysis**

By quantifying the effect of input variables, sub-models, and model parameters on model output, sensitivity analysis can: (1) expose model components that cause the smallest and largest changes in the model output and (2) assess the degree of uncertainty in the outputs that is associated with inaccurate input estimation. This identifies which input parameters or model components should be most accurately estimated given their influence on the behavior of the system. This is a relevant point in complex model systems, as the interaction between certain variables can induce large changes in the final result. A complete sensitivity analysis scheme should combine the effect of all model components combinations and interactions in a factorial design (Leemans 1991). The complexity of such a process has led to simplified sensitivity analysis schemes (e.g., Bevins and Martin 1978, Scott 1998a). Bartelink's (1998) relative sensitivity (RS) test was chosen for the present study. This parameter can be viewed as an index calculated from the partial derivative of output variables with respect to the perturbation of the input variable. This dimensionless result arises from the following criteria:

\[
RS = \frac{V_{+10\%} - V_{-10\%}}{V_{default}}
\]

where \(V_{+10\%}\) and \(V_{-10\%}\) are the resulting value of the critical parameter when the value of the parameter under analysis is changed by 10 percent and \(V_{default}\) is the resulting value of the critical parameter under default conditions. The value 0.2 is the relative range of the parameter to be analyzed. The 10% intervals were arbitrarily assigned. A RS score indicates the proportional response
of the model to the changes in the perturbed input parameter. A sensitivity scale can be drawn from the results. RS scores less than 1 indicate insensitive (<0.5) or slightly sensitive (0.5 – 1) model responses to inputs; and RS scores larger than 1 indicate model sensitivity, which can be divided into moderate (1 – 2) and high (>2).

The various models were run under the two baseline conditions outlined in table 2 to cover the range of conditions over which crown fire behavior is expected to occur. Although the simplified sensitivity analysis scheme used does not take into account the interactions in a fully comprehensive manner as would be obtained if one were using a full factorial design, the computation of sensitivity scores for the two distinct burning conditions yielded an acceptable range of variability for the relative sensitivity scores. The RS scores computed for the crown fire initiation models (table 3) reflect the distinct modeling approaches that were followed. The non-dynamic nature of the Van Wagner (1977) crown fire initiation model results in no changes in RS scores between the high and very high fire environment severity conditions, with the model results indicating moderate sensitivity to changes in canopy base height (CBH) and foliar moisture content (FMC). Both Alexander (1998) and Van Wagner (1977) show the same sensitivity to canopy base height (CBH) variation, because of their similar formulations for the CBH effect. The sensitivity of the Van Wagner (1977) model to FMC seems to be excessive. This is a result of the pure theoretical formulation of the effect of FMC in increasing the heat sink of a fuel particle. The FMC effect formulation in the Alexander (1998) model is an application of the results of the Xanthopoulos and Wakimoto (1993) laboratory experiments on foliar heating relationships and yields lower sensitivity scores. A more complete analysis of the Van Wagner (1977) and Alexander (1998) crown fire initiation models, both of which are based on convective heat transfer theory, should include a link with a fire spread model for the estimation of fireline intensity, as done by Scott (1999). However, such analysis would include the errors inherent to such models for estimation of surface fire spread (Rothermel 1972, Albini 1976) and fireline intensity (Byram 1959, Andrews 1986), which would confound further analysis. The RS scores from the Cruz (1999) crown fire initiation model should be analyzed considering the shape of the cumulative probability curve that is the outcome of a logistic regression model. The higher magnitude RS scores are relative to the steepest component of the probability curve, which is indicative of transitional behavior, whereas the very low sensitivity values are relative to

<table>
<thead>
<tr>
<th>Input parameters</th>
<th>Fire environment severity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High</td>
</tr>
<tr>
<td>Van Wagner (1977)</td>
<td></td>
</tr>
<tr>
<td>CBH</td>
<td>1.5</td>
</tr>
<tr>
<td>FMC</td>
<td>1.3</td>
</tr>
<tr>
<td>Alexander (1998)</td>
<td></td>
</tr>
<tr>
<td>CBH</td>
<td>1.5</td>
</tr>
<tr>
<td>FMC</td>
<td>1</td>
</tr>
<tr>
<td>Within stand wind speed</td>
<td>0.6</td>
</tr>
<tr>
<td>Flame front residence time</td>
<td>-0.5</td>
</tr>
<tr>
<td>Cruz (1999)</td>
<td></td>
</tr>
<tr>
<td>FSG</td>
<td>-2.8</td>
</tr>
<tr>
<td>Fine dead fuel moisture</td>
<td>-2.4</td>
</tr>
<tr>
<td>10-m open wind speed</td>
<td>2.6</td>
</tr>
</tbody>
</table>
the flatter regions of the curve, which is characteristic of low (<0.15) or high (>0.85) probability scores. This model produces the highest sensitivity scores, making it more prone to amplifying errors due to inaccurate assessment of fire environment input variables.

The crown fire rate of spread models analyzed can be more easily compared due to the commonality of outputs forms and most inputs. Comparisons reveal the models show distinct sensitivities to most of the input variables. The sigmoid equation used in the Canadian FBP System fuel type-specific models result in the same characteristic as the one described for the logistic crown fire initiation model of Cruz (1999). The FBP System fuel type-specific models are extremely sensitive (RS magnitudes between 5.6 to 2.2) to changes in input parameters on the steepest region of the sigmoid curve, which characterizes a transition from surface fire to crown fire spread (table 4). The very high RS scores for the FMC, seen only for FBP System Fuel Type C-6 (3.6 to 1.7), seem unusually high for an unsubstantiated relationship (Van Wagner 1998). Both Alexander (1998) and Cruz (1999) found no evidence for a significant FMC effect on crown fire spread rate. The Rothermel (1991) crown fire spread model responds to environmental changes in the same way as the Rothermel (1972) fire spread model. It shows very low sensitivity to changes in fine dead fuel moisture (RS of -0.2) and moderate sensitivity (RS between 1.3 and 1.4) to wind speed. The Cruz (1999) crown fire rate of spread model,

<table>
<thead>
<tr>
<th>Fire environment severity</th>
<th>Fire environment severity</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>Very high</td>
</tr>
<tr>
<td>Rothermel (1991)</td>
<td></td>
</tr>
<tr>
<td>Fine dead fuel moisture</td>
<td>-0.2</td>
</tr>
<tr>
<td>10-m open wind speed</td>
<td>1.4</td>
</tr>
<tr>
<td>Forestry Canada Fire Danger Group (1992)</td>
<td></td>
</tr>
<tr>
<td>FFMC</td>
<td>-5.6/-2.3</td>
</tr>
<tr>
<td>10-m open wind speed</td>
<td>5.4/2.2</td>
</tr>
<tr>
<td>FMC</td>
<td>3.6/0.3</td>
</tr>
<tr>
<td>Cruz (1999)</td>
<td></td>
</tr>
<tr>
<td>Canopy bulk density</td>
<td>0.2</td>
</tr>
<tr>
<td>10-m open wind speed</td>
<td>0.9</td>
</tr>
<tr>
<td>Fine dead fuel moisture</td>
<td>-1.4</td>
</tr>
</tbody>
</table>

1 10-m open wind speed was converted into 6-m (20 ft) wind speed by the 15% adjustment factor as determined by Turner and Lawson (1978).
2 FFMC = Fine fuel moisture code.

based on a large database of experimental crown fires, is an image of its dataset. This model’s RS scores vary from very low for canopy bulk density (0.2) to moderate for fine dead fuel moisture (-1.4). Wind speed has a proportional effect on rate of spread in this model.

Predictive Validation

Predictive validation is applied here to determine: (1) the model adequacy in capturing the behavior of the real world system under study; and (2) if the

Table 4—Relative sensitivity (RS) values associated with crown fire rate of spread model outputs for the major input parameters.
model accuracy is suitable for its proposed application. Such tests should be done using independent, highly reliable data to decrease the probability of a Type I error (Sargent 1984), rejecting the validity of a valid model. All the high-intensity experimental fire behavior (either near or above the threshold of crowning) in the literature readily available to the authors was used to develop two of the models (Cruz 1999) under analysis. In order to simultaneously compare the various models under a common basis, independent, high-intensity wildfire data was utilized. The characteristics of wildfire derived data generally make it unsuitable for the evaluation of crown fire models (e.g., characteristics lack precise description of fuel complexes and fires spreading under the influence of extreme weather conditions). Consequently, a slightly different approach was followed in the evaluation of crown fire initiation models. The predictive validation of such models was based on the analysis of the model behavior under a well-documented fire situation where all input variables needed to characterize a fire scenario for the various models were either measured and/or acceptably estimated. The crown fire model input requirements from PFES Experimental Fire R1 (Van Wagner 1968, 1977) are presented in Table 2. The comparison between models was based on the 10-m open wind speed requirements given variable vertical stratification in the fuel complex, with CBH values of 2, 4 and 6 m. For these fire environment scenarios, the critical fireline intensities for Van Wagner (1977) and Alexander (1998) models, respectively, are as follows: for CBH = 2 m, 475 and 540 kW/m; for CBH = 4 m, 1344 and 1290 kW/m; and for CBH = 6 m, 2470 and 2180 kW/m. The critical fireline intensities for Alexander (1998) model were based on a residence time of 45 seconds as observed in R1, and a constant of proportionality of 16 as determined for needlebed surface fuel complexes (Alexander 1998). The estimation of fireline intensity for use as input in these two models was based on the output of the two principal fire behavior prediction systems used in North America, the BEHAVE System (Andrews 1986) and the Canadian FBP System (Forestry Canada Fire Danger Group). The use of these two systems to predict a fundamental fire behavior descriptor highlights the potential error propagation problem when using the Byram’s (1959) fireline intensity as an input variable for determining the requirements for crown fire initiation.

As pointed out previously, any comparison between the crown fire initiation models is hindered by differences in model formulations and the dependence of some of the models on fire behavior quantities that must be estimated in advance as inputs in the prediction. The error propagation

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CBH (m)</td>
<td>BEHAVE</td>
<td>FBP C-6¹</td>
<td>BEHAVE</td>
</tr>
<tr>
<td>2</td>
<td>24 (8)²</td>
<td>1</td>
<td>27 (9)²</td>
</tr>
<tr>
<td>4</td>
<td>48 (16)</td>
<td>4</td>
<td>45 (15)</td>
</tr>
<tr>
<td>6</td>
<td>75 (25)</td>
<td>10</td>
<td>69 (23)</td>
</tr>
</tbody>
</table>

¹ FFMC = 92 and BUI = 70 (Alexander 1998).
² Mid-flame wind speeds as required for input in the surface fire spread model embodied in the BEHAVE System (Andrews 1986). Conversion from within stand (1.2-m) to 10-m open wind speed based on linear transformation using a 3.0:1 ratio as measured during R1 experimental fire (Van Wagner 1968, Alexander 1998).
problem becomes especially evident when comparing the 10-m open wind speed required to attain the fireline intensity requirements for crown ignition for the two surface fire spread models tested (table 5). Part of that error is produced by the manner in which fireline intensity is estimated and the distinct models used to estimate surface fire rate of spread. The wind speed requirements for crown ignition using the BEHAVE system seem to be unreasonably high at 48 km/h for a CBH of 4 m and 75 km/h for a CBH of 6 m (table 5), a problem that is also evident in Scott’s (1998b) analysis and Scott and Reinhardt’s (2001) simulations. In contrast, the use of the FBP System Fuel Type C-6 fire spread model yielded what seems to be very low wind speed requirements for crown ignition. The 1 and 4 km/h open wind conditions would result in almost no wind flow within the sub-canopy space, resulting in low intensity surface fires that would hardly induce crown combustion for the fuel complex simulated. These differences in response arise from the fact that the BEHAVE System estimates fireline intensity from the product of reaction intensity (quantity evaluated under no wind, no slope in a laboratory setting) with flame depth (Rothermel 1972, Albini 1976, Andrews 1986). This estimation procedure yields systematically lower values than the original Byram (1959) formulation that was used by Van Wagner (1977) to determine the empirical proportionality constant in his crown fire initiation model. Consequently this inflates the wind speed requirements needed to attain the critical fireline intensities. The Cruz (1999) crown fire initiation model appeared to yield reasonable results for the situations tested, with wind speed requirements varying between 7 and 14 km/h for the CBH range tested. For reference purposes, crowning associated with PFES Experimental Fire (CBH = 7 m) was attained with a wind speed of 15 km/h (Alexander 1998).

The two crown fire initiation models based on convective plume theory (Yih 1953) examined here (i.e., Van Wagner 1977, Alexander 1998) could be regarded as more sound conceptually than the Cruz (1999) model and presumably lead to a greater understanding of crown fire initiation phenomenology. However, some of the limiting assumptions concerning plume theory when applied to free-burning wildland fires and the dependence of Byram’s (1959) fireline intensity to define the heat source potentially limits their use as a robust model. Further focus on convective plume theory in the development of new models of crown fire initiation may in fact stifle innovation. The relationship obtained by Yih (1953) through similarity analysis linking the temperature at a certain height above a linear heat source is technically restricted to still-air conditions, although Alexander’s (1998) model has attempted to account for the cross-wind case. Neither Van Wagner (1977) nor Alexander (1998) attempted to account for the role of radiant heat flux in the onset of crowning from the flames that typically characterize high-intensity surface fires, although the authors readily acknowledged this shortcoming/possible limitation in their models.

Predictive validation applied to the crown fire spread models was based on a wildfire dataset derived from case studies. No Canadian FBP System model was tested here due to the non-independence of these models and the wildfire dataset used here. Although the published wildfire case studies provide general information on the crown fire runs, fuel types and weather conditions, detailed fuel complex descriptions and quantitative data generally do not exist. For evaluation of the Cruz (1999) crown fire rate of spread model, a nominal canopy bulk density value of 0.15 kg/m³ was assigned on the basis of knowledge and experience with experimental fires in generally similar fuel types. All the fires were also assumed to be spreading as active crown fires. This assumption is corroborated by the high spread rates observed on the wildfires.
selected for evaluation purposes. Figure 1 displays a scatterplot of observed versus predicted rate of spread produced by the Cruz (1999) and Rothermel (1991) crown fire rate of spread models. The results suggest strong under-prediction trends for the Rothermel (1991) model and an acceptable agreement, albeit an over-prediction trend for the Cruz (1999) model. The over-prediction trend of the Cruz (1999) model might arise from the worst-case scenario assumed for the crown fire run simulations that extended for several hours. These scenarios use the lower fine fuel moisture content computed for the fire run, whereas a more detailed fire simulation encompassing fine fuel moisture and wind speed variability over the burning period would probably reduce this tendency to over-predict. Hence we are not certain that the over-prediction trend evident in figure 1 is the result of model bias or the inadequacy of the test data/approach used to replicate real-world conditions. A cursory examination of the scatterplot also reveals an inability for the Rothermel (1991) model to predict high rates of spread for many of the situations considered (i.e., for many of the wildfires the model seldom predicted a spread rate greater than 10 m/ min while the observed spread rates varies from 10 to nearly 50 m/ min). In order to quantify the adequacy of the model’s behavior, two deviation measures were sought: the mean absolute error (MAE) and the mean absolute percent error (MA%E) (Mayer and Butler 1993, Cruz 1999). For the dataset utilized in the present analysis, the mean absolute errors computed were 20 m/ min (and MA%E of 62%) for the Rothermel (1991) model and 9.2 m/ min (and MA%E of 34%) for the Cruz (1999) crown fire spread model.

Comparing Two Models

The statistical validation procedures complement some of the quantitative results obtained from the descriptive analyses previously performed on the models. Nevertheless, the definition of any statistical validation criteria is hampered...
by the difficulty in defining adequate tests and appropriate confidence levels for the phenomena under study. Different tests might produce conflicting results, accepting or rejecting the same hypothesis simultaneously. The 0.05 alpha levels commonly accepted for statistical significance in a variety of natural resources studies might not be adequate to analyze phenomena that may vary several orders of magnitude. The previously described limitations in the datasets used here that prevent an independent analysis of the crown fire initiation models also restrict the application of statistical tests to the crown fire rate of spread models. Using the wildfire dataset employed in the Predictive Validation section, the models were analyzed for: (1) their modeling efficiency; (2) linear regression parameters; and (3) simultaneous F-test for slope = 1 and intercept = 0 (Draper and Smith 1981, Mayer et al. 1994). Modeling efficiency, EF, is expressed as follows (from Mayer and Butler 1993):

\[
EF = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}
\]  

where \(y_i\) is the observed and \(\hat{y}_i\) is the predicted value. This measure provides an indication of goodness of fit, with an upper bound of 1 describing a perfect fit and values less than 0 indicating poor model performance; see Mayer and Butler (1993) for further interpretation of EF. The simultaneous F-test for slope = 1 and intercept = 0 evaluates the null hypothesis, \(H_0: (b_0, b_1) = (0,1)\), by the following statistic: \(Q = (\beta - b)'X'X(\beta - b)\), where \(\beta\) is the population parameters to be tested, \(X'X\) the matrix term in the independent variable, and \(b\) is the vector of regression parameters. The null hypothesis is accepted if \(Q \leq ps^2F(p,v,1-a)\), where \(p\) is the regression degrees of freedom, \(s^2\) is the variance, and \(v\) is \(n - p\).

The application of the two crown fire rate of spread models under comparison to the crown fire dataset (see references to the wildfire case studies in Data Validation) produced EF values of -0.14 and 0.68 (table 6) for the Rothermel (1991) and Cruz (1999) models respectively. The slope coefficients (\(b_1\)) resulting from the regression analysis (table 6) reflect the slight over-prediction trend of the Cruz (1999) model and the strong under-prediction trend of the Rothermel (1991) model. For both models the simultaneous F-test for slope and intercept resulted in the rejection of the null hypothesis (both \(Q > ps^2F(p,v,1-a)\)). Given the uncertainty in the input conditions for the wildfire runs and the type of phenomena being analyzed, with rates of spread varying over three orders of magnitude, the results from the F-test should be analyzed with caution, as this test might be too restrictive for the phenomena under study (i.e., the null hypothesis with slope = 1 and intercept = 0, is easily rejected).

<table>
<thead>
<tr>
<th>Model</th>
<th>(\beta_0) (lower/upper 95%)</th>
<th>(\beta_1) (lower/upper 95%)</th>
<th>(Q)</th>
<th>(ps^2F(p,v,1-a))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rothermel (1991)</td>
<td>12.0 (2.52 / 21.5)</td>
<td>1.75 (1.06 / 2.44)</td>
<td>9231</td>
<td>441</td>
</tr>
<tr>
<td>Cruz (1999)</td>
<td>4.5 (-4.45 / 13.46)</td>
<td>0.74 (0.52 / 0.95)</td>
<td>1087</td>
<td>283</td>
</tr>
</tbody>
</table>

(p,v,1-a). Given the uncertainty in the input conditions for the wildfire runs and the type of phenomena being analyzed, with rates of spread varying over three orders of magnitude, the results from the F-test should be analyzed with caution, as this test might be too restrictive for the phenomena under study (i.e., the null hypothesis with slope = 1 and intercept = 0, is easily rejected).
Models are simplified theories developed to approximate the behavior of real world systems. The process of model evaluation acquires particular importance when the models are used to support decision-making. This shifts the emphasis on validation procedures to the potential applications of these models instead of the model itself (Mayer and Butler 1993). There exists a vast list of validation techniques available to evaluate models but hardly a set appropriate for widespread application covering different modeling approaches. The present set of evaluation procedures was developed to provide information about model performance and how well it replicates the behavior of the real-world system. The theoretical basis of the models evaluated and their distinct modeling approaches make them difficult to compare. The problem of model conceptual validity is one of the most challenging issues. The unknowns in certain fundamental fire behavior processes make it difficult to define which theories are most appropriate. The results of the sensitivity analysis showed how the models respond differently to the various input variables dominating some of the processes involved in crown fire initiation and spread. The main differences were found in the way foliar moisture content affects the susceptibility of canopy ignition. The theoretical model from Van Wagner (1977) is the most sensitive to FMC, followed by the Alexander (1998) application of Xanthopoulos and Wakimoto’s (1993) laboratory experiments, which assumes that the duration of heating can be equated to flame front residence time. The Cruz (1999) crown fire initiation model does not incorporate the effect of FMC, reflecting the non-significance of this parameter in the initiation of crown fires within the dataset used in its development. Fuel complex vertical stratification appears to manifest a comparable effect on the various models for crown fire initiation. The various crown fire rate of spread models respond similarly to changes in wind speed, but quite differently to fine fuel moisture content. The Rothermel (1991) model is relatively insensitive to the variation in this parameter whereas the Canadian FBP System models are over-sensitive. This evaluation of model sub-components highlights some areas needing further research and acquisition of new laboratory and field data. Combustion characteristics of live fuels are poorly understood and the effect of foliar moisture content as a heat sink has never been evaluated under heat flux conditions characteristic of wildfire situations. The effect of CBH on heat flux decay in the Van Wagner (1977) and Alexander (1998) crown fire initiation models is solely based on convective heating and no allowance is presently made for a radiative contribution. One would expect that a surface fire burning just below the threshold for crowning will possess an ample flame front depth, leading to a strong effect of the radiative component in heating canopy fuels. Further refinement of existing crown fire initiation models is required in order to accommodate the contribution that radiation plays in the onset of crowning.

The present study has also highlighted the lack of published fire behavior data that can be used in model evaluation. Although the growing complexity of fire management decision-making relies on the extensive use of fire behavior models, there were virtually no studies carried out in the United States that produced suitable data that could be used in the evaluation and calibration of models or systems used for predicting crown fire behavior.
Acknowledgments

This paper benefited from the comments and suggestions of Dr. Bret Butler and Rudy King of the USDA Forest Service and two anonymous reviewers.

References


