Forest Fires Initial Spread Index Prediction: A Training and Validation Study

MA 575 Fall 2021: Final Project

C2 Team 4 - Zhiquan Shen, Katherine Albrecht, Jiachen Chen, Jiahe Zhang

December 3rd, 2021

Introduction and Background

- Forest fires are serious environmental problems, which threaten both forest preservation and human life, and have become extremely crucial and challenging [Cortez and Morais, 2007]. Historically, meteorological data, including temperature, relative humidity, rain, and wind, have been used for fire prevention and management.
- In this project, the outcome we are interested in predicting is the Initial Spread Index (ISI). ISI is a component included in Canadian forest Fire Weather Index (FWI) that helps rating fire danger, and ISI itself relates to fire velocity spread.
- We expect to help detect scenarios with high risk of rapid fire spread so that forest fire management can be conducted in a more efficient manner.
- The data includes 517 observations, of which each represents a fire occurrence at Montesinho Natural Park. The variables recorded include the time, date (month and day of the week), spatial location, weather observations (temperature, wind, rain, and relative humidity), FWI components, and area burned.

Introduction and Background

- The aim of the study is to develop a linear model for the ISI fire score based on the meteorological observations in the forest fires dataset.
- Since we are predicting ISI, we are not able to use other FWI components (Figure 1).

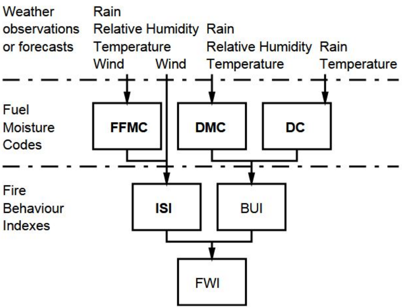

Figure 1: FWI flow chart

Modeling and Analysis

- Approach: training and validating the model [James et al., 2013].
- Predictors: wind, temp, summer (binary), interaction terms (for seasonal differences).

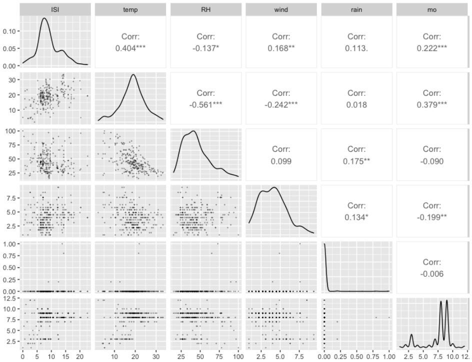

Figure 2: Training set scatterplots

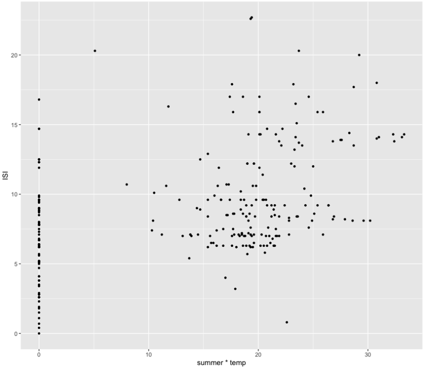

Figure 3: ISI by summer*temperature

Modeling and Analysis

- Final model:

$$\text{ISI} = \beta_0 + \beta_1 * summer + \beta_2 * temp + \beta_3 * wind + \beta_4 * I(temp^2) + \beta_5 * I(summer * temp) + \beta_6 * I(summer * temp^2).$$

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-2.515308	1.946176	-1.292	0.197399
summer	14.403554	3.507131	4.107	5.44e-05 ***
temp	0.623960	0.245916	2.537	0.011780 *
wind	0.670241	0.121738	5.506	9.11e-08 ***
I(temp^2)	-0.011853	0.007718	-1.536	0.125843
I(summer*temp)	-1.311434	0.377229	-3.476	0.000599 ***
I(summer*temp^2)	0.033897	0.010240	3.310	0.001069 **
Residual standard error: 3.338 on 250 degrees of freedom				
Multiple R-squared: 0.3364, Adjusted R-squared: 0.3204				
F-statistic: 21.12 on 6 and 250 DF, p-value: < 2.2e-16				

Table 1: R output for the summary of the proposed linear model

Modeling and Analysis

- Diagnostics:

$$\text{ISI} = 0.670^*\text{wind} + 0.624^*\text{temp} + 14.404^*\text{summer} - 1.311^*\text{I(summer}^*\text{temp)} + 0.034^*\text{I(summer}^*\text{temp}^2)$$

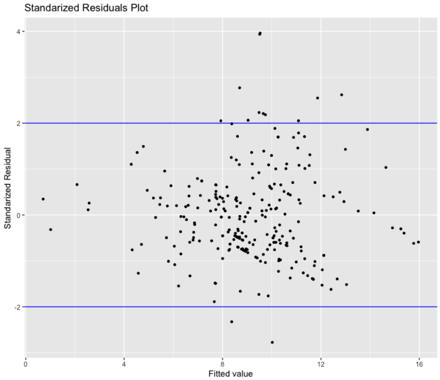

Figure 4: Standardized residuals vs fitted values

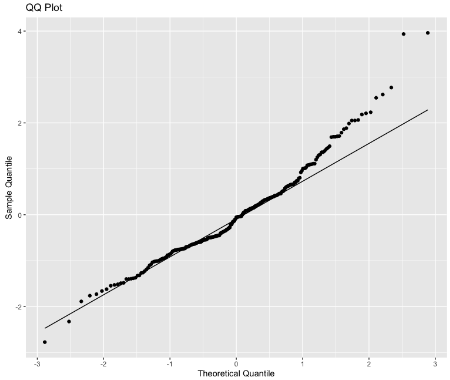

Figure 5: Q-Q Plot

Validation and Prediction

- Acceptable generalizability to the validation data: use the proposed model to predict ISI for the validation set: MSE of ISI for validation data: 12.94, MSE for training data: 10.84 (close), relative MSE for validation data: 0.1385 (not bad).

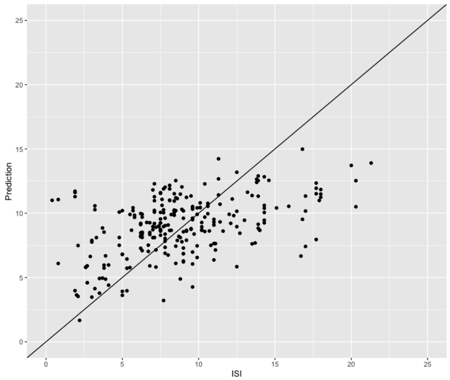

Figure 6: Model fit for ISI on validation set

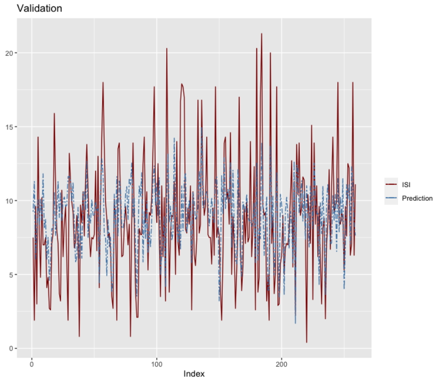

Figure 7: Predicted vs. actual ISI

Validation and Prediction

- Reasonable predictive accuracy for moderate ISI and underestimate higher ISI (fail in extreme cases critically important for fire management).

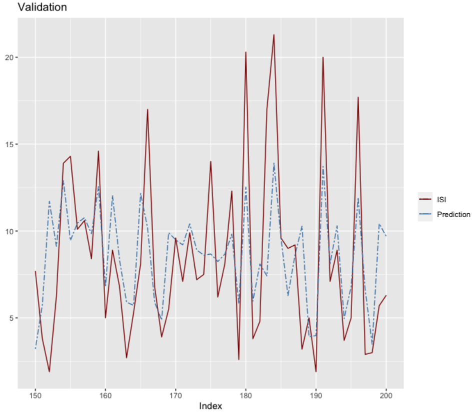

Figure 8: ISI vs. Prediction for Validation Data Set

Discussion

- Relatively good performance when ISI is moderate, but poor prediction of high ISI in both training and validation datasets.
- Unfortunately, the most urgent need is prediction of high ISI!
- Linear regression predicts only the mean ISI for any potential combination of predictors.
- Observations are both temporally and spatially correlated, violating the fundamental assumption of independence.
- A time series approach could be useful in this context, but the variables in the dataset are inadequate.

	X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind	rain	area	summer
504	2	4	aug	wed	94.5	139.4	689.1	20.0	29.2	30	4.9	0	1.95	TRUE
505	4	3	aug	wed	94.5	139.4	689.1	20.0	28.9	29	4.9	0	49.59	TRUE
25	7	4	aug	sat	93.5	139.4	594.2	20.3	23.7	32	5.8	0	0.00	TRUE
212	7	4	aug	sat	93.5	139.4	594.2	20.3	5.1	96	5.8	0	26.00	TRUE
207	2	2	aug	sat	93.5	139.4	594.2	20.3	22.9	31	7.2	0	15.45	TRUE
136	3	5	aug	sat	93.5	139.4	594.2	20.3	17.6	52	5.8	0	0.00	TRUE
486	2	4	aug	mon	95.0	135.5	596.3	21.3	30.6	28	3.6	0	2.07	TRUE
12	7	5	sep	sat	92.8	73.2	713.0	22.6	19.3	38	4.0	0	0.00	TRUE
267	6	5	aug	tue	94.3	131.7	607.1	22.7	19.4	55	4.0	0	0.17	TRUE

Figure 9: Observations of ISI >= 20

Reference I

[Cortez and Morais, 2007] Cortez, P. and Morais, A. d. J. R. (2007).
A data mining approach to predict forest fires using meteorological data.

[James et al., 2013] James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013).
An introduction to statistical learning, volume 112.
Springer.