Applications and Applied Mathematics: An International Journal (AAM)

Volume 8 | Issue 2 Article 21

12-2013

Using Fuzzy Linear Regression to Estimate Relationship between Forest Fires and Meteorological Conditions

Hande G. Akdemir
Giresun University

Fatma Tiryaki
Yildiz Technical University

Available at
http://pvamu.edu/aam
Appl. Appl. Math.
ISSN: 1932-9466

Applications and Applied Mathematics:
An International Journal
(AAM)

Vol. 8, Issue 2 (December 2013), pp. 673 – 683

Using Fuzzy Linear Regression to Estimate Relationship between Forest Fires and Meteorological Conditions

Hande Günay Akdemir
Department of Mathematics
Faculty of Arts and Science
Giresun University
Giresun, Turkey
hande.akdemir@giresun.edu.tr

Fatma Tiryaki
Department of Mathematics
Faculty of Arts and Science
Yildiz Technical University
Istanbul, Turkey
ftiryaki@yildiz.edu.tr

Received: July 23, 2013; Accepted: July 23, 2013

Abstract

Each year, millions of hectares of forest land are destroyed by fires causing great financial loss and ecological damage. In this paper, our aim is to study the effect of the variation of meteorological conditions on the total burned area in hectares, by using fuzzy linear regression analysis based on Tanaka's approaches. The total burned area is considered a dependent variable. Air temperature (in ºC), relative humidity (in %), wind speed (in km/h) and rainfall (in mm/m^2) are considered to be independent variables. The relationship between input and output data is estimated using data provided in data mining literature. In our study, we apply fuzzy regression, using crisp/fuzzy input data and fuzzy output data expressed in linguistic terms.

Keywords: Fuzzy Set Theory; Fuzzy Linear Regression; Linear Programming; Forest Fires

AMS-MSC 2010 No.: 62J86, 65K05

1. Introduction

Each year, millions of hectares of forest lands are destroyed by fires which cause financial loss and ecological damage which can be excessive when vegetation is dry. Prediction of the total burned area in forest fires can be helpful to prevent, reduce or minimize the damage cause by fire. Fire danger rating systems provide indications and fire management approaches.

The Canadian Forest Fire Danger Rating System includes three moisture indices, namely, fine fuel moisture, duff moisture, and drought codes, which represent the moisture of different layers of forest litter and are used in the estimation of the Fire Weather Index (FWI). The Canadian FWI has six components; the first three are moisture codes; Fine Fuel Moisture Code (FFMC), Duff Moisture Code (DMC), Drought Code (DC), and the other components are Initial Spread Index (ISI), Buildup Index (BUI) and FWI [see Kaloudis et al. (2005), Cortez and Morais (2007)].

The FWI is based on weather readings taken at noon standard time. Weather readings required are

- Air temperature (in the shade),
- Relative Humidity (in the shade),
- Wind speed (at 10 metres above ground level for an average over 10 minutes),
- Rainfall (for the previous 24 hours).

The FFMC is a relative indicator of the ease of ignition and flammability of the fine fuel, ranging from 0 (lowest risk of flammability) to 100 (highest risk of flammability). The FFMC is computed from daily weather observations; temperature, relative humidity, wind speed and rainfall data. It is the fastest changing component of the Canadian FWI. The DMC is computed from temperature, relative humidity and rainfall data, in addition to previous FFMC value. On the other hand, the DC is calculated from only temperature and rainfall data. The DMC and DC are always positive, but have no maximum value. The ISI is a numerical rating of expected velocity of fire spread and combines FFMC and wind speed. The BUI represents the amount of available fuel and combines DMC and DC. The FWI is an indicator of the fire intensity and combines ISI and BUI (Figure 1).

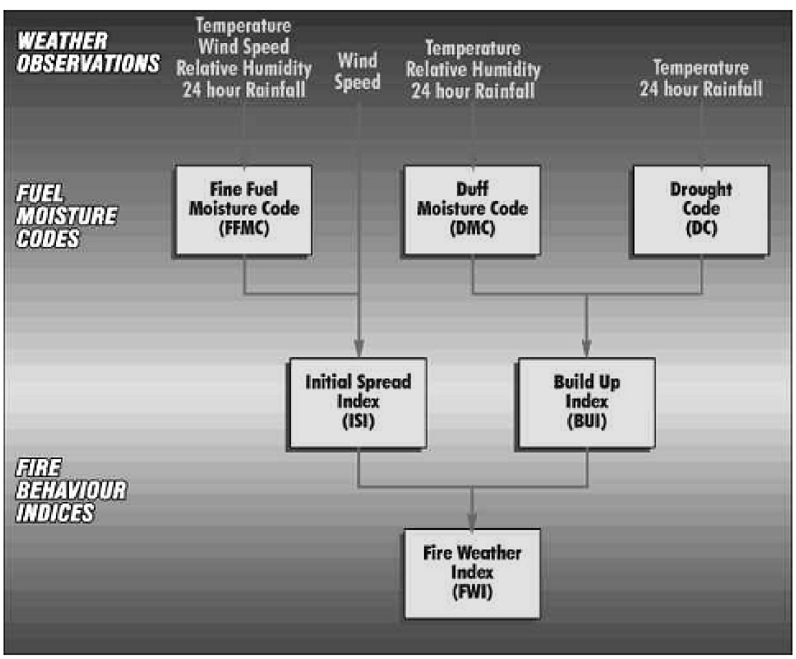

Figure 1. FWI calculation structure[1]

To calculate FFMC, drying and wetting by rain and atmospheric humidity of the litter and fine fuels are considered. However, sunlight intensity, wind speed and atmospheric humidity fluctuate during the day, so weather readings cannot represent accurate values for an entire day. Data cannot be recorded correctly since the conditions vary temporally that causes measurement errors. Furthermore, in practice, when human judgments are involved, data can be described by linguistic terms such as "fine fuels will ignite easily", "extreme fire behavior is likely" and "potentially high level of fire intensity".

Fuzzy set theory allows us to handle problems with imprecise and linguistic information. Fuzzy sets were introduced as generalization of the classical crisp sets to represent inexact data. A linguistic variable takes values as words which can be characterized as fuzzy sets.

Regression analysis is the most widely used statistical technique to model the functional relationship among given correlated variables. Based on the relation, the value of the dependent variable can be predicted or estimated from independent variable(s) as close to the observed data as possible.

2. Fuzzy Linear Regression

Fuzzy linear regression, which is developed by Tanaka et al. (1982), seeks to model vague and imprecise relation between a dependent variable and independent variables using fuzzy parameters. In their study, a linear programming problem was formulated for a fuzzy dependent variable and crisp independent variable. The objective was to minimize the total spread/vagueness of the fuzzy regression coefficients.

Fuzzy linear regression model is expressed as;

$$\tilde{y}_i = \tilde{A}_0 + \tilde{A}_1 x_{i1} + \tilde{A}_2 x_{i2} + \ldots + \tilde{A}_k x_{ik} \quad i = 1, \ldots, n$$

where $\tilde{y}_i = (y_i, e_i)$ is the given symmetric triangular fuzzy output i with center y_i and left/right spread $e_i (e_i \geq 0)$ (Figure 2), $x_i = (x_{i1}, x_{i2}, \ldots, x_{ik})^T$ is the given real-valued input vector i for $i = 1, \ldots, n$, and $\tilde{A}_j$ is the symmetric triangular fuzzy number coefficient with center α_j and left/right spread $c_j (c_j \geq 0)$ for $j = 1, \ldots, k$.

Table 1. Observations

Dependent Variable	Independent Variables			
$\tilde{y} = (y, e)$	x_1	x_2	$\ldots$	x_k
$\tilde{y}_1 = (y_1, e_1)$	x_{11}	x_{12}	$\ldots$	x_{1k}
$\tilde{y}_2 = (y_2, e_2)$	x_{21}	x_{22}	$\ldots$	x_{2k}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$\tilde{y}_n = (y_n, e_n)$	x_{n1}	x_{n2}	$\ldots$	x_{nk}

[1] http://jcweather.us/fwiCalc.php, see Definitions.

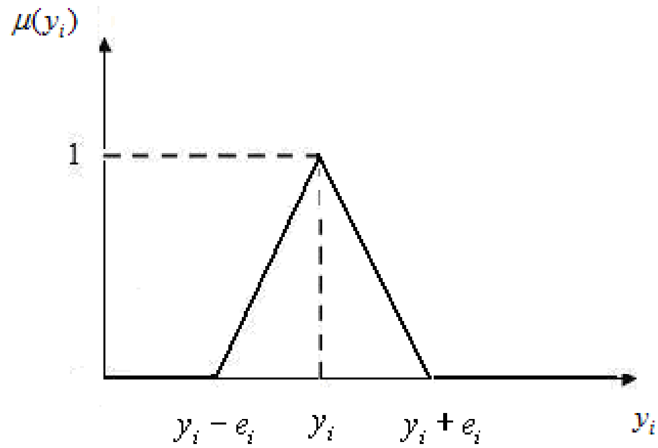

Figure 2. Membership function of the symmetric triangular fuzzy output $\tilde{y}_i$

Linear programming formulation of the fuzzy regression problem follows as;

$$\begin{aligned}
&\text{minimize } \sum_{j=0}^{k} c_j \\
&\text{subject to} \\
&\sum_{j=0}^{k} \alpha_j x_{ij} + (1-h) c_j \left| x_{ij} \right| \geq y_i + (1-h) e_i, \qquad i = 1, 2, \ldots, n \\
&\sum_{j=0}^{k} \alpha_j x_{ij} - (1-h) c_j \left| x_{ij} \right| \leq y_i - (1-h) e_i, \qquad i = 1, 2, \ldots, n \\
&\alpha_j \text{ unrestricted in sign}, \quad c_j \geq 0 \qquad\qquad j = 0, 1, \ldots, k \\
&x_{i0} = 1, \qquad\qquad\qquad\qquad\qquad\qquad i = 1, 2, \ldots, n.
\end{aligned}$$

The degree of confidence h is selected by the decision maker, and is between 0 and 1. The parameter h can be viewed as degree of decision maker's optimism or flexibility. If h is greater (less) than 0.5, then decision maker is optimistic (pessimistic) about the reliability of the data. If h is equal to 0.5, then decision maker is risk neutral. The fuzziness of the output increases when h decreases. In this study, h is taken equal to 0 in order to obtain a minimum spread. The notation

$$\tilde{a}^h = [a_l{}^h, a_u{}^h], \ \ h \in [0,1]$$

denotes h-cuts of fuzzy number $\tilde{a}$. We refer to $a_l{}^h$ and $a_u{}^h$ as lower and upper bounds of $\tilde{a}$, respectively. Scalar multiplication is defined based on interval arithmetic;

$$k\tilde{a} = [\min\left\{ka_l{}^h, ka_u{}^h\right\}, \max\left\{ka_l{}^h, ka_u{}^h\right\}]$$

for given $k \in$ □ .

The advantage of the Tanaka model is its simplicity in computation. However, the model is extremely sensitive to outliers (observations lieing outside the predetermined bounds [D'Urso et al. (2011)]), so, it sometimes gives very wide estimated ranges which is not helpful in application. In Tanaka and Watada (1988), the objective function is modified as;

$$\text{minimize} \quad \sum_{i=1}^{n} \sum_{j=0}^{k} c_j \left| x_{ij} \right|. \tag{1}$$

Remark 1.

The modified Tanaka model seeks to minimize the total fuzziness of the estimated values, i.e., Equation (1).

Remark 2.

Both Tanaka models force the *h*-certain estimated intervals to include the *h*-certain observed intervals. The constraints guarantee that the support of the estimated y value contains the support of the observed y value. The upper bound of the estimated y value is greater than the upper bound of the observed y value. Similarly, the lower bound of the estimated y value is lower than the lower bound of the observed y value (Figure 3).

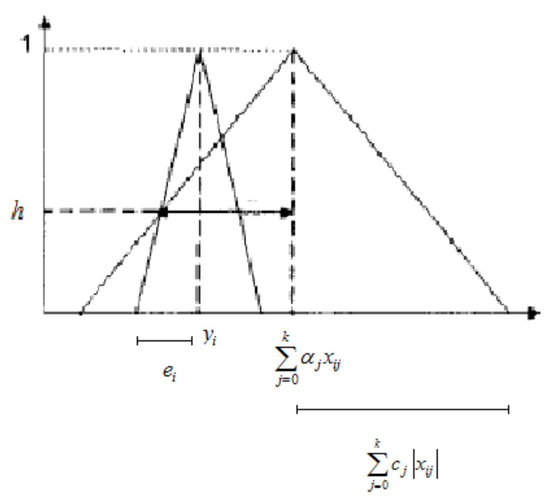

Figure 3. The supports of h -certain predicted and observed intervals

3. Numerical Example

In this example, meteorological data are taken from UCI machine learning repository[2] and are adapted to our fuzzy model, assuming that fluctuations exist. The size of the problem is decreased to avoid computational complications, for example the same samples produce the same linear constraints to Tanaka's linear programming models. The aim is to demonstrate

[2] http://archive.ics.uci.edu/ml/datasets/Forest+Fires

how to predict the burned area of forest fires by using fuzzy meteorological data within a small sized example.

First, crisp FFMC data are transformed into fuzzy sets with non-overlapping membership functions which almost never allow a value to be a member of more than one fuzzy set. In Table 2, a linguistic scale is defined with linguistic terms and fuzzy numbers corresponding to FFMCs (Figure 4).

Table 2. Linguistic terms and fuzzy numbers corresponding to FFMCs[3]

Interval	Linguistic Term	(Center, Spread)
0-33	Low	(16.5,16.5)
33-66	Moderate	(49.5, 16.5)
66-81	High	(73.5, 7.5)
81-84	Very High	(82.5, 1.5)
84-86	Severe	(85, 1)
86-88	Extreme	(87, 1)
88-100	Catastrophic	(94, 6)

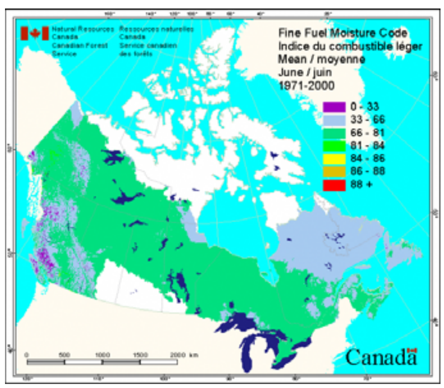

Figure 4. Historical analysis map for FFMCs in Canada (2011)

The total burned area is used as the dependent variable. Air temperature (in ºC), relative humidity (in %), wind speed (in km/h) and rainfall (in mm/m^2) are considered as crisp independent variables, FFMC is also a fuzzy independent variable. In samples, we consider observed total burned areas (except zero values) as symmetrical triangular fuzzy numbers with left and right spread 0.01 in case of measurement errors. In samples, zero value means that areas lower than 0.01 was burned. However, the zero value can be fuzzified as a symmetrical triangular fuzzy number with center 0.005 and left/right spread 0.005. Calculations have been made from 50 samples. All variables and corresponding fuzzy coefficients are summarized in Table 3.

We aim to find symmetrical triangular fuzzy coefficients (α_j, c_j), $j = 1, 2, 3, 4, 5$ of the function

$$(y,e) = (\alpha_0, c_0) + (\alpha_1, 0)(x_1, f_1) + (\alpha_2, c_2)x_2 + (\alpha_3, c_3)x_3 + (\alpha_4, c_4)x_4 + (\alpha_5, c_5)x_5$$

[3] http://cwfis.cfs.nrcan.gc.ca/en_CA/hamaps/fwnormals/ffmc/06

in which

$$\alpha_1(x_1, f_1) = \begin{cases} [\alpha_1(x_1 - f_1), \alpha_1(x_1 + f_1)] & \text{if } \alpha_1 \geq 0 \\ [\alpha_1(x_1 + f_1), \alpha_1(x_1 - f_1)], & \text{otherwise} \end{cases}$$

since $x_1 - f_1 \geq 0$. So, the center and the spread of $\alpha_1(x_1, f_1)$ are $\alpha_1 x_1$ and $|\alpha_1 f_1|$, respectively.

Based on Remark 1, the total fuzziness of the estimated value

$$\sum_{i=1}^{50} \left(c_0 + |\alpha_1 f_{i1}| + c_2 x_{i2} + c_3 x_{i3} + c_4 x_{i4} + c_5 x_{i5} \right)$$

is minimized.

By Remark 2, the upper and lower bounds of the estimated value

$$\begin{aligned} y_{iu}^{estimated} &= \alpha_0 + c_0 + \alpha_1 x_{i1} + |\alpha_1 f_{i1}| + \alpha_2 x_{i2} + c_2 x_{i2} \\ &\quad + \alpha_3 x_{i3} + c_3 x_{i3} + \alpha_4 x_{i4} + c_4 x_{i4} + \alpha_5 x_{i5} + c_5 x_{i5} \end{aligned}$$

and

$$\begin{aligned} y_{il}^{estimated} &= \alpha_0 - c_0 + \alpha_1 x_{i1} - |\alpha_1 f_{i1}| + \alpha_2 x_{i2} - c_2 x_{i2} \\ &\quad + \alpha_3 x_{i3} - c_3 x_{i3} + \alpha_4 x_{i4} - c_4 x_{i4} + \alpha_5 x_{i5} - c_5 x_{i5}, \end{aligned}$$

are used in the constraints, respectively.

Table 3. Variables and fuzzy coefficients

Independent Variables	Fuzzy Coefficients	
	Center	Spread
Constant $x_0 = 1$	α_0	c_0
FFMC (linguistic variable) (x_1, f_1)	α_1	$c_1 = 0$
Temperature x_2	α_2	c_2
Relative humidity x_3	α_3	c_3
Wind speed x_4	α_4	c_4
Rainfall x_5	α_4	c_5
Dependent Variable	**Fuzzy Numbers**	
	Center	Spread
Total burned area $(y \neq 0)$	y	0.01
Total burned area $(y = 0)$	0.005	0.005

Two separate problems; Problem 1 and Problem 2 are solved with an additional constraint $\alpha_1 \geq 0$ or $\alpha_1 \leq 0$, respectively. However, the models give too wide estimated ranges due to outliers. To detect outliers, namely, to identify, to eliminate and/or possibly substitute the anomalous data, we consider the effect of each observation on total fuzziness. Let z^* be the optimal value of the objective function and z^*_i be the corresponding value obtained by eliminating the i^{th} observation. The larger the ratio $\frac{|z^* - z^*_i|}{z^*}$, the more the i^{th} observation is likely to be an outlier [Chen (2001)]. After detecting outliers, the corresponding observations are excluded. In our example, only one sample is found as an outlier and is excluded from the regression dataset. The solutions of the problems and their optimal total fuzziness values are shown in Table 4 and Table 5. In the Appendix, samples and estimated values are given in Table 6.

Table 4. Results before the outliers are eliminated

	Problem 1 ($\alpha_1 \geq 0$)		Problem 2 ($\alpha_1 \leq 0$)	
Independent variables (index)	α_j	c_j	α_j	c_j
Constant ($j=0$)	-1.9339	0.6344	0	0
FFMC ($j=1$)	0.016	0	0	0
Temperature ($j=2$)	2.0546	2.0472	2.0503	2.0503
Relative humidity ($j=3$)	0.0961	0.0706	0.0837	0.0837
Wind speed ($j=4$)	-0.0251	0	0	0
Rainfall ($j=5$)	-33.8067	0	-33.5624	0
Total Fuzziness	1532.0787		1532.8326	

Table 5. Results after the outliers are eliminated

	Problem 1 ($\alpha_1 \geq 0$)		Problem 2 ($\alpha_1 \leq 0$)	
Independent variables (index)	α_j	c_j	α_j	c_j
Constant ($j=0$)	7.3305	3.025	7.6687	2.815
FFMC ($j=1$)	0.0075	0	0	0
Temperature ($j=2$)	-0.2493	0	-0.2288	0
Relative humidity ($j=3$)	0.3384	0.3665	0.3363	0.3716
Wind speed ($j=4$)	-0.2194	0	-0.183	0
Rainfall ($j=5$)	-16.5392	0	-16.5487	0
Total Fuzziness	1076.8085 *		1077.8552	

* The best optimal (minimal) value

4. Conclusion

It is already known that meteorological conditions affect forest fires. One of the explanations of the spread $c_j = 0$ means that the corresponding variable x_j directly influences y. So, based on our model, the temperature, wind speed and rainfall all directly influence the total burned area, but, the relative humidity indirectly influences y. In other words, the relationship between relative humidity and the total burned area cannot be expressed with a simple linear function since many other factors may be included in the relation.

REFERENCES

Chen, Yun-Shiow (2001). Outliers detection and confidence interval modification in fuzzy regression. Fuzzy Sets and Systems, Vol. 119 (2), pp. 259-272.

Cortez, Paulo, and Morais, Anibal (2007). A Data Mining Approach to Predict Forest Fires using Meteorological Data. In J. Neves, M. F. Santos and J. Machado Eds., New Trends in Artificial Intelligence, Proceedings of the 13th EPIA 2007 - Portuguese Conference on Artificial Intelligence, December, Guimaraes, Portugal, pp. 512-523, APPIA.

D'Urso, Pierpaolo, Massari, Riccardo, and Santoro, Adriana (2011). Robust fuzzy regression analysis. Information Science, Vol. 181 (19), pp. 4154-4174.

Kaloudis, S., Tocatlidou, A., Lorentzos, N.A., Sideridis A.B., and Karteris, M. (2005). Assessing Wildfire Destruction Danger: a Decision Support System Incorporating Uncertainty. Ecological Modelling, Vol. 181 (1), pp. 25-38.

Tanaka, H., Uejima, S., and Asai K. (1982). Linear regression analysis with fuzzy model. IEEE Transactions on Systems, Man, and Cybernetics, Vol. 12, pp. 903–907.

Tanaka, H., Watada, J., (1988). Possibilistic linear system and their application to the linear regression model. Fuzzy Sets and Systems, Vol. 27, pp. 275–289.

http://jcweather.us/fwiCalc.php, Retrieved August 3, 2011.

http://cwfis.cfs.nrcan.gc.ca/en_CA/hamaps/fwnormals/ffmc/06, Retrieved August 3, 2011.

http://archive.ics.uci.edu/ml/datasets/Forest+Fires, Retrieved August 3, 2011.

Appendix

Table 6. Samples and estimated values

Observed Values											Estimated Values	
FFMC	Linguistic FFMC	x	f	Temperature	Relative Humidity	Wind	Rain	Total Area	y	e	y	e
86.2	extreme	87	1	8.2	51	6.7	0	0.00	0.005	0.005	21.72716	21.724
63.5	moderate	49.5	16.5	17.0	72	6.7	0	0.00	0.005	0.005	26.35847	29.53675
84.9	severe	85	1	16.7	47	4.9	0	0.00	0.005	0.005	18.63443	20.258
79.5	high	73.5	7.5	23.3	37	3.1	0	0.00	0.005	0.005	13.91372	16.64175
83.9	very high	82.5	1.5	12.7	48	1.8	0	0.00	0.005	0.005	20.63142	20.62825
85.8	severe	85	1	18.0	42	2.7	0	0.36	0.360	0.010	17.10102	18.4255
84.4	severe	85	1	24.2	28	3.6	0	0.96	0.960	0.010	10.6203	13.2945
84.9	severe	85	1	5.3	70	4.5	0	2.14	2.140	0.010	29.34741	28.6875
85.6	severe	85	1	17.4	50	4.0	0	2.69	2.690	0.010	19.67258	21.3575
81.5	very high	82.5	1.5	5.8	54	5.8	0	4.61	4.610	0.010	23.50439	22.82725
84.9	severe	85	1	19.1	32	4.0	0	5.44	5.440	0.010	13.15757	14.7605
86.8	extreme	87	1	12.4	53	2.2	0	6.38	6.380	0.010	22.3442	22.457
81.5	very high	82.5	1.5	5.8	54	5.8	0	10.93	10.930	0.010	23.50439	22.82725
63.5	moderate	49.5	16.5	22.6	38	3.6	0	11.32	11.320	0.010	14.13693	17.07575
83.9	very high	82.5	1.5	8.8	68	2.2	0	13.05	13.050	0.010	28.28393	27.95825
87.6	extreme	87	1	11.0	46	5.8	0	27.35	27.350	0.010	19.53458	19.8915
87.6	extreme	87	1	11.0	46	5.8	0	36.85	36.850	0.010	19.53458	19.8915
84.4	severe	85	1	24.3	36	3.1	0	105.66	105.660	0.010	13.41227	16.2265
84.4	severe	85	1	4.8	57	8.5	0	8.98	8.980	0.010	24.19526	23.923
84	severe	85	1	5.1	61	8.0	0	11.19	11.190	0.010	25.58377	25.389
84.6	severe	85	1	5.1	61	4.9	0	5.38	5.380	0.010	26.26391	25.389
85.4	severe	85	1	4.6	21	8.5	0	17.85	17.850	0.010	12.06272	10.729
85.4	severe	85	1	4.6	21	8.5	0	10.73	10.730	0.010	12.06272	10.729
85.4	severe	85	1	4.6	21	8.5	0	22.03	22.030	0.010	12.06272	10.729
85.4	severe	85	1	4.6	21	8.5	0	9.77	9.770	0.010	12.06272	10.729

Table 7. Samples and estimated values, Continued

Observed Values											Estimated Values	
FFMC	Linguistic FFMC	x	f	Temperature	Relative Humidity	Wind	Rain	Total Area	y	e	y	e
84.7	severe	85	1	2.2	59	4.9	0	9.27	9.270	0.010	26.31008	24.656
85.4	severe	85	1	5.1	24	8.5	0	24.77	24.770	0.010	12.95327	11.8285
86.9	extreme	87	1	8.8	35	3.1	0	1.10	1.100	0.010	16.95302	15.86
85.2	severe	85	1	7.5	46	8.0	0	24.24	24.240	0.010	19.90945	19.8915
18.7	low	16.5	16.5	5.2	100	0.9	0	0.00	0.005	0.005	39.80043	39.79875
84.7	severe	85	1	7.5	71	6.3	0	9.96	9.960	0.010	28.74243	29.054
85	severe	85	1	10.1	62	1.8	0	51.78	51.780	0.010	26.03595	25.7555
87.9	extreme	87	1	21.8	34	2.2	0	6.04	6.040	0.010	13.57118	15.4935
87.1	extreme	87	1	17.0	67	4.9	0	3.95	3.950	0.010	25.34264	27.588
84.6	severe	85	1	8.2	53	9.4	0	4.62	4.620	0.010	21.79658	22.457
87.9	extreme	87	1	10.9	64	3.1	0	3.35	3.350	0.010	26.24309	26.4885
75.1	high	73.5	7.5	4.6	82	6.3	0	5.39	5.390	0.010	33.10155	33.13425
75.1	high	73.5	7.5	5.1	77	5.4	0	2.14	2.140	0.010	31.48236	31.30175
79.5	high	73.5	7.5	4.6	59	0.9	0	6.84	6.840	0.010	26.50311	24.70475
87.2	extreme	87	1	10.2	45	5.8	0	3.18	3.180	0.010	19.39562	19.525
81.6	very high	82.5	1.5	27.8	32	2.7	0	6.44	6.440	0.010	11.25513	14.76425
81.6	very high	82.5	1.5	21.9	71	5.8	0	54.29	54.290	0.010	25.24346	29.05775
81.6	very high	82.5	1.5	21.2	70	6.7	0	11.16	11.160	0.010	24.88211	28.69125
94.4	catastrophic	94	6	25.6	42	4.0	0	0.00	0.005	0.005	14.98862	18.463
91	catastrophic	94	6	18.2	62	5.4	0	0.43	0.430	0.010	23.29428	25.793
95.5	catastrophic	94	6	23.8	32	5.4	0	0.55	0.550	0.010	11.7462	14.798
90.1	catastrophic	94	6	21.2	51	8.9	0	0.61	0.610	0.010	18.05608	21.7615
95.5	catastrophic	94	6	23.8	32	5.4	0	0.77	0.770	0.010	11.7462	14.798
90.4	catastrophic	94	6	14.3	46	1.8	0	0.90	0.900	0.010	19.64199	19.929
91	catastrophic	94	6	21.1	71	7.6	1.4	2.17	2.170	0.010	1.97935	29.0915