Exponential piece wise regression for rainfall data

K Anil Kumar¹, V N Sendil², S Venkatramana Reddy³ and B Sarojamma²*

¹Department of Mathematics, School of Science, GITAM University, Hyderabad -502329,TS, INDIA
²Department of Statistics, S.V.University, Tirupati – 517 502, AP, INDIA
³Department of Physics, S.V.University, Tirupati – 517 502, AP, INDIA

* Author for Correspondence e-mail: saroja14397@gmail.com

Abstract. There are several regression models for the rainfall in the literature. Some of them are linear regression model, multiple regression model, piece wise regression model, Poisson regression model, negative binomial regression model, etc. Various piece wise regression models are linear piece wise regression, polynomial piece wise regression, cubic piece wise regression, quadratic piece wise regression, etc. This paper proposes three piece wise regression models using exponential smoothing models, i.e, simple exponential smoothing spline, holt’s exponential spline, and damped exponential spline for annual rainfall data from 1970 to 2017 with two knots at 1990 and 2005 with three subdivisions. Symmetric mean average percentage error is used as an accuracy measure for choosing the best model among the proposed three exponential spline models.

Keywords: rainfall data, simple exponential smoothing spline, holt’s exponential spline and damped exponential spline, Symmetric mean average percentage error.

1. Introduction:

   In India, the rainy season is generally from June to September, and the annual average rain is recorded between 750 and 1500 millimeters around the region. Usually, rainwater is better than artificial irrigation methods because it does not have added chlorine chemicals. The advantages are that it is good for plants and soil. It reduces runoff pollution, contributes to erosion prevention efforts, and eco-friendly options to keep composts moister. On the other side of the coin are essential created flooding kills thousands around the world every year. Rain causes excessive load on the drainage system.

   W.F.Krejewski etal.[1] explains “Radar hydrology: rainfall estimation”, the authors used Radar observations of rainfall and discussed their use. Methodological advances are needed in several areas of radar-rainfall estimation of particular importance, advancement in rainfall estimatimates using polarimetric radar observations, estimates of the error structure of rainfall rate estimates, and validation of radar rainfall algorithm. J.H.C.Gash[2] studies “An analytical model of rainfall interception by forecasts”, the two major factors which control the evaporation of intercepted rainfall area a) amount of time that the conopy spends saturated during rains and the evaporation rate
applicable under these conditions and b) canopy saturates capacity and the number of times this store is emptied.

Jan G De Gooijer, et al[3] gives 25 years of timeseries forecasting and they explained exponential smoothing models, ARIMA models, Seasonality, State space and structural models, the Kalman filter, Non-linear, long memory ARCH/GARCH, current data forecasting and also explain different accuracy measures for forecasting prediction intervals and densities. M.Sidiq[4], in his article "Forecasting Rainfall with Time Series Model," aims to study the forecast rainfall with the time series model. For computations of ARIMA(1,0,1), ARIMA(0,1,1), ARIMA(1,1,1), AR(1) and MA(1) for months data from 2011 to 2014, i.e. 48 data points of banding in millimeters. They prefer ACF and PACF plots and calculations of AIC, MAE, and MAPE to choose a good model. D.N. Gujarati[5] gives the Essential of econometrics for different econometric models. A. Nogroho et al[6] explains "ARMA model for prediction of rainfall in Regency of semarany-central Java-Republic of Indonesia" and discussed different ARMA models for rainfall data.

2. Methodology:

In the literature, there are various regression models exist. Some of the essential regression models are linear regression, polynomial regression, harmonic regression, Lasso Regression, Net elastic Regression, Binomial regression, Negative binomial regression, etc.

**Piecewise polynomial (Splines):** Generally, fitted data with different polynomials. By fitting higher-order, sometimes it cannot even exhibit a good fit due to downward data. In such situations, it will work for each subdivision of data overcome by using Knots to the data. A Knot means joint points of the piecewise polynomials. By dividing the data into subdivisions and we will fit an appropriate polynomial to each and every subdivision.

### 2.1. Simple exponential smoothing spline:

By dividing the data into subdivisions using Knots. For every subdivision, we fit an appropriate simple exponential smoothing.

| Subdivision | Model |
|--------------|-------|
| $x_1,x_2,\ldots, \ldots, x_n$ | $S_{t_n} = \alpha_1 y_{t_n-1} + (1 - \alpha_1)S_{t_n-1}$ |
| $x_{m+2},x_{m+2},\ldots, x_{2n}$ | $S_{t_{2n}} = \alpha_1 y_{t_{2n}-1} + (1 - \alpha_1)S_{t_{2n-1}}$ |

By combining all the above equations, we will get the equation for the whole data.

$$Y = S_{t_n} + S_{t_{2n}} + S_{t_{3n}} + \cdots \cdots \cdots S_{t_{mn}}$$

### 2.2. Holt’s exponential smoothing splines:

For each subdivision of data, we fit Holt’s winter’s additive method. The Holt’s winter’s additive method is as follows:

$$\hat{y}_{t+h} = l_t + hb_t + S_{t+h-m(k+1)}$$

$$l_t = \alpha(y_t - S_t - m) + (1 - \alpha)(l_{t-1} + b_{t-1})$$

$$b_t = \beta^*(l_t - l_{t-1}) + (1 - \beta^*)b_{t-1}$$

$$S_t = \gamma(y_t - l_{t-1} - b_{t-1}) + (1 - \gamma)S_{t-m}$$

Where $y_t$ = Time series observation at the time $t$.

$S_{t-m}$ = fitted seasonal observation at time point $t-m$
\[ l_t = \text{linear forecast at time point } t. \]
\[ b_t = \text{slope forecast for time point } t. \]

The parameters \( \alpha, \beta \) and \( \gamma \) are restricted to lies between 0 and 1.

2.3. **Damped trend Splines:** For piece wise regression model, we are fitting the Damped trend model for each subdivision. The Damped trend for additive Holt–Winter’s method is as follows:

\[
\hat{y}_{t+h/t} = \left[ l_t + (\varphi + \varphi^2 + \cdots + \varphi^h)b_t \right] S_{t-m+n}\n\]

Level \( l_t = \alpha(y_t/S_{t-m}) + (1-\alpha)(l_{t-1} + \varphi b_{t-1}) \)

Slope \( b_t = \beta^\ast(l_t - l_{t-1} + (1-\beta^\ast)b_{t-1} \)

Seasonal \( S_t = r \frac{y_t}{(l_{t-1} + \varphi b_{t-1})} + (1-\gamma)S_{t-m}\n\]

Where \( \varphi \) is damping variable

\( l_t \) is level equation at time \( t \).

\( b_t \) is slope equation at time \( t \)

\( S_t \) is the seasonal effect equation at time \( t \).

\( \alpha, \beta \) and \( \gamma \) are restricted constants lies between 0 and 1.

2.4. **Holt’s linear trend splines:** By dividing the data into \( n \) subdivisions, for each subdivision, we fit an appropriate Holt’s linear trend. The Holt’s linear trend model is

**Forecast equation** \( \hat{y}_{t+h/t} = l_t + hb_t \)

**Level equation** \( l_t = \alpha(y_t) + (1-\alpha)(l_{t-1} + b_{t-1}) \)

**Trend equation** \( b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1} \)

2.5. **Symmetrical Mean Average Percentage Error (SMAPE):** It is the ratio of the sum of modulus of errors and sum of absolute value and forecast value and is given by

\[
\text{SMAPE} = \frac{\sum_{t=1}^{n}|F_t - Y_t|}{\sum_{t=1}^{n}(|F_t| + |Y_t|)} \times 100
\]

Where \( F_t \) is forecast value at time \( t \) and \( Y_t \) is the time series value at time \( t \).

3. **Empirical investigations:**

The rainfall data of Coastal Andhra Pradesh, Telangana, and Rayalaseema regions from 1976 to 2017 separated into three divisions by taking two Knots between 1990 and 2005[7]. By fitting simple exponential smoothing splines, Damped Holt winter exponential splines, and Holt’s exponential smoothing splines, the best-fitted model for these models' data is identified by employing symmetric mean average percentage error. The time series value and simple exponential smoothing splines estimated (SES) value, Holt’s exponential splines (HES) value, and Damped exponential spline estimated (DES) values for coastal Andhra Pradesh, Telangana, and Rayalaseema are presented in Table-1, Table-2, and Table-3 respectively.
### Table 1: Coastal Andhra Pradesh

| Year | Rainfall | SES   | HES   | DES   |
|------|----------|-------|-------|-------|
| 1970 | 1064.20  | 1046.64 | 1012.88 | 1019.34 |
| 1971 | 1033.20  | 1049.23 | 1026.94 | 1032.17 |
| 1972 | 1028.20  | 1046.86 | 1035.50 | 1040.08 |
| 1973 | 859.10   | 1044.15 | 1042.45 | 1046.58 |
| 1974 | 944.90   | 1016.83 | 1027.89 | 1033.36 |
| 1975 | 1192.40  | 1006.21 | 1025.57 | 1031.24 |
| 1976 | 1090.30  | 1033.70 | 1053.71 | 1057.10 |
| 1977 | 1132.80  | 1042.06 | 1065.98 | 1068.62 |
| 1978 | 1038.30  | 1055.45 | 1081.92 | 1083.61 |
| 1979 | 934.00   | 1052.92 | 1084.40 | 1086.33 |
| 1980 | 1014.80  | 1036.38 | 1073.87 | 1077.06 |
| 1981 | 929.30   | 1032.34 | 1074.47 | 1077.88 |
| 1982 | 888.50   | 1017.13 | 1064.57 | 1069.02 |
| 1983 | 1196.00  | 998.13  | 1050.89 | 1056.58 |
| 1984 | 785.20   | 1027.35 | 1076.39 | 1080.01 |
| 1985 | 967.60   | 991.60  | 1048.68 | 1054.76 |
| 1986 | 1009.70  | 988.05  | 1046.59 | 1052.78 |
| 1987 | 1079.90  | 991.25  | 1049.89 | 1056.75 |
| 1988 | 1164.80  | 1004.34 | 1061.35 | 1066.25 |
| 1989 | 1030.00  | 1028.03 | 1081.77 | 1085.09 |
| 1990 | 1611.10  | 1028.32 | 1053.25 | 1086.71 |

### Table 2: Telangana

| Year | Rainfall | SES   | HES   | DES   |
|------|----------|-------|-------|-------|
| 1970 | 970.30   | 971.29 | 866.37 | 833.78 |
| 1971 | 729.10   | 971.16 | 894.77 | 860.93 |
| 1972 | 670.70   | 939.56 | 897.94 | 882.05 |
| 1973 | 1042.50  | 904.46 | 895.34 | 900.40 |
| 1974 | 830.40   | 922.48 | 927.79 | 923.34 |
| 1975 | 1215.80  | 910.46 | 937.35 | 941.81 |
| 1976 | 1035.60  | 950.32 | 982.08 | 967.09 |
| 1977 | 808.60   | 961.46 | 1005.76 | 990.00 |
| 1978 | 1311.70  | 941.50 | 1005.98 | 1007.27 |
| 1979 | 971.70   | 989.83 | 1053.27 | 1038.10 |
| 1980 | 922.60   | 963.96 | 1047.46 | 1040.10 |
| 1981 | 1143.50  | 958.56 | 1054.44 | 1064.52 |
| 1982 | 933.20   | 982.71 | 1081.45 | 1082.76 |
| 1983 | 1497.50  | 976.25 | 1086.25 | 1096.11 |
| 1984 | 742.80   | 1044.30 | 1143.41 | 1119.99 |
| 1985 | 785.90   | 1004.94 | 1124.59 | 1129.85 |
| 1986 | 1007.30  | 976.34 | 1111.56 | 1135.92 |
| 1987 | 976.90   | 980.38 | 1120.48 | 1142.85 |
| 1988 | 1544.90  | 979.93 | 1125.71 | 1147.26 |
| 1989 | 1161.00  | 1053.89 | 1183.62 | 1162.58 |
| 1990 | 1425.70  | 1069.70 | 1200.17 | 1172.79 |
Table 3: Rayalaseema

| Year | Rainfall | SES   | HES   | DES  |
|------|----------|-------|-------|------|
| 1970 | 776.00   | 798.84| 796.84| 788.90|
| 1971 | 621.10   | 797.03| 794.33| 787.65|
| 1972 | 826.00   | 728.83| 777.52| 770.99|
| 1973 | 689.50   | 786.31| 781.46| 776.49|
| 1974 | 728.50   | 778.50| 772.26| 767.80|
| 1975 | 1041.60  | 774.46| 767.56| 763.87|
| 1976 | 817.00   | 796.03| 792.61| 791.64|
| 1977 | 806.60   | 797.72| 794.30| 794.18|
| 1978 | 869.30   | 803.28| 800.47| 801.42|
| 1979 | 813.40   | 808.61| 806.31| 808.21|
| 1980 | 573.70   | 809.00| 806.38| 808.73|
| 1981 | 785.20   | 790.00| 784.00| 785.22|
| 1982 | 595.60   | 789.61| 783.52| 785.22|
| 1983 | 943.00   | 773.95| 765.33| 766.26|
| 1984 | 733.00   | 787.60| 781.36| 780.93|
| 1985 | 685.80   | 783.19| 776.24| 778.84|
| 1986 | 580.90   | 775.33| 767.18| 769.54|
| 1987 | 769.50   | 759.83| 749.14| 750.67|
| 1988 | 842.38   | 760.42| 750.45| 752.55|
| 1989 | 693.70   | 767.04| 758.45| 761.53|
| 1990 | 930.70   | 761.11| 751.79| 745.34|

The systematic mean average percentage error for coastal Andhra Pradesh, Telangana, and Rayalaseema for simple exponential smoothing spline (SES), Holt’s exponential smoothing spline (HES), and Damped exponential spline (DES) is listed in table-4.

Table 4. Systematic mean average percentage error

| Region            | SES    | HES    | DES    |
|-------------------|--------|--------|--------|
| Coastal Andhra Pradesh | 5.4021 | 5.6038 | 5.5290 |
| Telangana         | 9.6108 | 9.3982 | 9.4039 |
| Rayalaseema       | 6.4355 | 6.2975 | 6.2917 |

Table-5 explains the simple exponential spline estimated value and original values for coastal Andhra Pradesh, Telangana, and Rayalaseema for the time period 1991 to 2005 at the spline II area.

Table 5: The simple exponential spline estimated values (SES) by region wise

| Year | Coastal Andhra Pradesh | Telangana | Rayalaseema |
|------|------------------------|-----------|-------------|
| 1991 | 1128.17                | 986.26    | 889.76      |
| 1992 | 1127.47                | 971.63    | 897.24      |
| 1993 | 1112.99                | 960.10    | 868.24      |
| 1994 | 1085.14                | 952.15    | 870.06      |
| 1995 | 1093.15                | 954.66    | 852.77      |
| 1996 | 1122.79                | 990.00    | 854.85      |
| 1997 | 1132.87                | 989.66    | 895.55      |
| 1998 | 1114.47                | 975.41    | 889.72      |
| 1999 | 1136.95                | 987.95    | 895.34      |
| 2000 | 1101.24                | 974.71    | 860.87      |
Table 6 provides Holt’s exponential smoothing spline (HES), and Damped exponential splines (DES) estimated values for coastal-Andra Pradesh, Telangana, and Rayalaseema.

Table 6. HES and DES estimated values for three regions

| Year | Coastal-Andhra Pradesh | Telangana  | Rayalaseema | Coastal-Andhra Pradesh | Telangana  | Rayalaseema |
|------|------------------------|------------|-------------|------------------------|------------|-------------|
| 1991 | 1097.63                | 997.92     | 899.90      | 1102.63                | 960.83     | 886.84      |
| 1992 | 1105.14                | 981.50     | 904.84      | 1109.49                | 961.73     | 894.90      |
| 1993 | 1098.31                | 968.23     | 872.25      | 1102.64                | 960.21     | 865.01      |
| 1994 | 1077.60                | 958.58     | 871.94      | 1082.46                | 957.50     | 867.22      |
| 1995 | 1091.02                | 959.13     | 852.05      | 1095.02                | 957.78     | 849.55      |
| 1996 | 1124.89                | 991.37     | 852.48      | 1127.32                | 973.61     | 852.03      |
| 1997 | 1139.25                | 989.71     | 893.09      | 1141.05                | 980.28     | 894.59      |
| 1998 | 1125.40                | 974.83     | 885.50      | 1127.72                | 979.56     | 888.63      |
| 1999 | 1150.84                | 985.72     | 889.95      | 1152.11                | 987.28     | 894.58      |
| 2000 | 1119.31                | 972.05     | 852.98      | 1121.85                | 985.24     | 858.84      |
| 2001 | 1109.93                | 981.69     | 853.01      | 1112.75                | 990.83     | 860.09      |
| 2002 | 1103.48                | 974.42     | 858.35      | 1106.44                | 990.12     | 866.53      |
| 2003 | 1063.64                | 951.83     | 832.31      | 1068.01                | 980.09     | 841.40      |
| 2004 | 1072.11                | 951.80     | 819.86      | 1075.87                | 975.84     | 829.81      |
| 2005 | 1054.76                | 921.47     | 810.27      | 1058.98                | 957.55     | 821.00      |

The systematic mean average percentage error in Coastal Andhra Pradesh, Telangana, and Rayalaseema for simple exponential smoothing spline (SES), Holt’s exponential smoothing spline (HES), and Damped exponential spline (DES) is presented in Table 7 for the period 1991-2005.

Table 7. Systematic mean average percentage error for 1991-2005

| S | Coastal Andhra Pradesh | Telangana  | Rayalaseema |
|---|------------------------|------------|-------------|
| SSE | 7.6354                | 7.7101     | 7.6875      |
| HES | 6.8407                | 6.8942     | 6.6724      |
| DES | 8.9044                | 8.8906     | 8.9421      |

The simple exponential smoothing splines estimated (ESS) value, Holt’s exponential splines (HES) value, and Damped exponential spline estimated (DES) values for coastal Andhra Pradesh, Telangana, and Rayalaseema are exhibited in table 8 and table 9.
Table 8. The simple exponential spline estimated values (SES) by region wise for the period 2006 to 2017

| Year | Coastal Andhra pradesh | Telangana | Rayalaseema |
|------|------------------------|-----------|-------------|
| 2006 | 1119.99                | 903.77    | 808.16      |
| 2007 | 1079.28                | 915.42    | 800.04      |
| 2008 | 1117                   | 909.79    | 808.66      |
| 2009 | 1117.55                | 919.58    | 808.15      |
| 2010 | 1079.27                | 899.82    | 802.42      |
| 2011 | 1068.49                | 929.2     | 810.11      |
| 2012 | 1043.31                | 915.46    | 806.67      |
| 2013 | 1019.55                | 922.73    | 802.29      |
| 2014 | 1051.43                | 955.97    | 800.4       |
| 2015 | 1068.49                | 939.62    | 788.53      |
| 2016 | 1073.34                | 933.19    | 800.87      |
| 2017 | 1078.72                | 941.79    | 789.15      |

Table 9. HES and DES estimated values by three regions during 2006-2017

| Year | Coastal Andhra pradesh | Telangana | Rayalaseema | Coastal Andhra pradesh | Telangana | Rayalaseema |
|------|------------------------|-----------|-------------|------------------------|-----------|-------------|
| 2006 | 1016.71                | 943.42    | 822.67      | 948.7                  | 955.44    | 831.62      |
| 2007 | 1010.95                | 949.99    | 800.53      | 974.83                 | 952.48    | 815.4       |
| 2008 | 1083                   | 935.43    | 813.63      | 1021.45                | 942.61    | 822.82      |
| 2009 | 1109.61                | 941.05    | 807.82      | 1059.21                | 940.54    | 817.8       |
| 2010 | 1090.41                | 910.08    | 791.96      | 1079.01                | 923.89    | 805.77      |
| 2011 | 1097.32                | 941.73    | 804.23      | 1097.95                | 931.53    | 812.71      |
| 2012 | 1085.91                | 919.17    | 793.53      | 1105.45                | 922.25    | 804.37      |
| 2013 | 1073.69                | 923.77    | 781.64      | 1104.8                 | 921.95    | 794.96      |
| 2014 | 1118.88                | 961.14    | 775.5       | 1117.41                | 939.64    | 789.23      |
| 2015 | 1146.04                | 936.03    | 749.58      | 1127.86                | 936.99    | 769.79      |
| 2016 | 1158.45                | 924.2     | 774.03      | 1134.26                | 935.18    | 784.34      |
| 2017 | 1170.18                | 931.71    | 748.6       | 1139.2                 | 940.36    | 765.52      |

The systematic mean average percentage error for coastal Andhra Pradesh, Telangana, and Rayalaseema for simple exponential smoothing spline (SES), Holt’s exponential smoothing spline (HES), and Damped exponential spline (DES) is listed in table-10 for the time period 2006-2017.
Table 10. Systematic mean average percentage error for 2006-2017

| Region        | SES     | HES     | DES     |
|---------------|---------|---------|---------|
| Coastal Andhra Pradesh | 8.5219  | 10.1969 | 10.1657 |
| Telangana     | 9.7636  | 9.6574  | 9.3587  |
| Rayalaseema   | 8.6980  | 8.8649  | 8.8948  |

4. Summary and conclusions:

For the data from the year 1970 to 2017 rain fall annual values, we are fitted three exponential smoothing models for coastal Andhra Pradesh, telangana and rayalaseema by taking two Knots at 1990 and 2005. The best fitted model for three regions i.e coastal Andhra Pradesh, telangana and rayalaseema using three models i.e simple exponential smoothing spline, Holt’s exponential smoothing spline and damped exponential smoothing spline by considering accuracy measure symmetric mean average percentage error. Table-11 explains the best model with their symmetric mean average percentage error for region.

Table 11. Choosing the best model by Symmetric mean average percentage error.

| Models | Coastal Andhra Pradesh | Telangana | Rayalaseema |
|--------|------------------------|-----------|-------------|
|        | Subdivision 1 | Subdivision 2 | Subdivision 3 | Subdivision 1 | Subdivision 2 | Subdivision 3 | Subdivision 1 | Subdivision 2 | Subdivision 3 |
| SES    | 5.4021     | 9.6108      | 6.4355      | 7.6354        | 6.8407      | 8.9044      | 8.5219       | 9.7636       | 8.6980       |
| HES    | 5.6038     | 9.3982      | 6.2975      | 7.7101        | 6.8942      | 8.8906      | 8.8906       | 10.1969      | 9.6574       |
| DES    | 5.529      | 9.4039      | 6.2917      | 7.6875        | 6.6724      | 8.9421      | 10.1657      | 9.3587       | 8.8948       |

References:

[1]. Krejewski W F, Smith J A, “Radar hydrology: rainfall estimation”, Advances in Water Resources 25 (2002)1387-1394.
[2]. Gash J H C, “Quarterly Journal of the Royal metrological society”, An analytical model of rainfall interception by forests’, 105(1979)43-45.
[3]. Jan G De Gooijer, “25 years time series forecasting”, International Journal of Forecasting, 22(3)(2006)443-473.
[4]. Sidiq M, “Forecasting rainfall with time series model”, 10p conference series: Mathematical Science and Engineering, 407(2018)012154 doi:10.1.1088/1757-899X/407/1/012154.
[5]. Gujarati D N, “Essential of Econometrics”, Newyork: MC Graw –Hill co. 2009.
[6]. Nugroho A and Simanjuntak B H, “ARMA (Auto Regressive Moving Average) model for prediction of rainfall in the regency of Semarang-central Java-Republic of Indonesia”, International Journal of Computer Science(IJCSI)11(3)(2014)27.
[7]. Data website:URL:http://www.imd.gov.in/Welcome%20To%20IMD/Welcome.php