Time Series Analysis of Wildfires in the US

Final Project of the Coursera course “Specialized models: time series and survival analysis” by IBM

Francesco Tamborra

GitHub: https://github.com/fratambot/wildfires-time-series

Goal of the analysis

Recent studies seems to suggest that rising of frequency of heatwaves due to climate change all around the world are increasing the risk and intensity of wildfires.
Especially in summertime, droughts can increase the incidence of fires and dry vegetation can make the fire more devastating.

The analysis on the wildfire data, described in further details in the next section, aims to put in evidence if there is indeed an increase in wildfire incidence and/or intensity in the last years.

The data

The data comes from the U.S. National Interagency Fire Center and contains the **number of fires**, the **acres burned per fire** and the derived **total acres burned** from January 2000 to July 2022 with monthly frequency in the U.S.

Here is the shape and the tail of the Pandas table once loaded from the CSV file:

```
Shape of the table (rows, columns) =  (271, 4)
```

	Date	Acres Burned	Number of Fires	Acres Burned per Fire
266	202203	368065	7694	47.84
267	202204	602491	7543	79.87
268	202205	799903	5722	139.79
269	202206	2002408	5880	340.55
270	202207	1790284	5300	337.79

Raw data

After using the date as index and setting the correct frequency (the notebook can be found here), the three features as a function of time looks like:

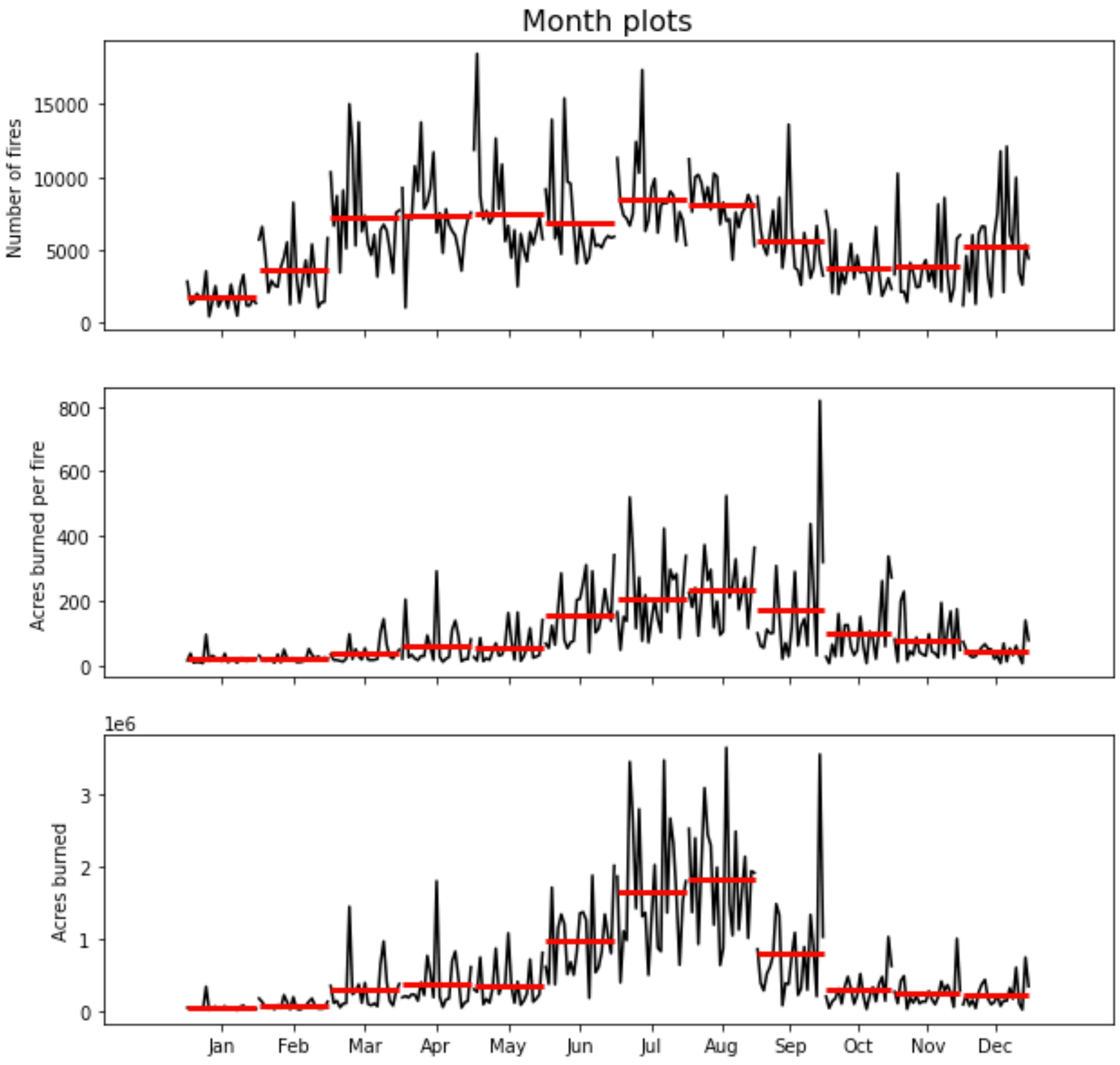
Annual resampling

If we resample the features annually, we can already eyeball a peak in the number of fires in the years 2007/2008 and also a positive trend of acres burned per fire.

However, we expect the data to have a seasonal component (wildfires most likely happen in spring/summer rather than autumn/winter).

Month plots

By using `month_plot` built in the package `statsmodel`, we can clearly see the seasonal dependency which need to be subtracted in order to obtain the correct trend over the years.



Decomposition

In order to extract the trend for the three features we need to perform a decomposition (`seasonal_decompose` in `statsmodel`).

The period for the decomposition is, of course, 12 months.

To make sure that the decomposition is executed correctly, the residuals should be a stationary series; We can use the Augmented Dickey-Fuller (ADF) to test against the null hypothesis that the residuals are not stationary: for stationarity to be true the `adf` should be negative and the `p-value` close to zero.

In the next three slides we show the decomposition of the three features and the `adf` and `p-value` for the ADF test on the residuals

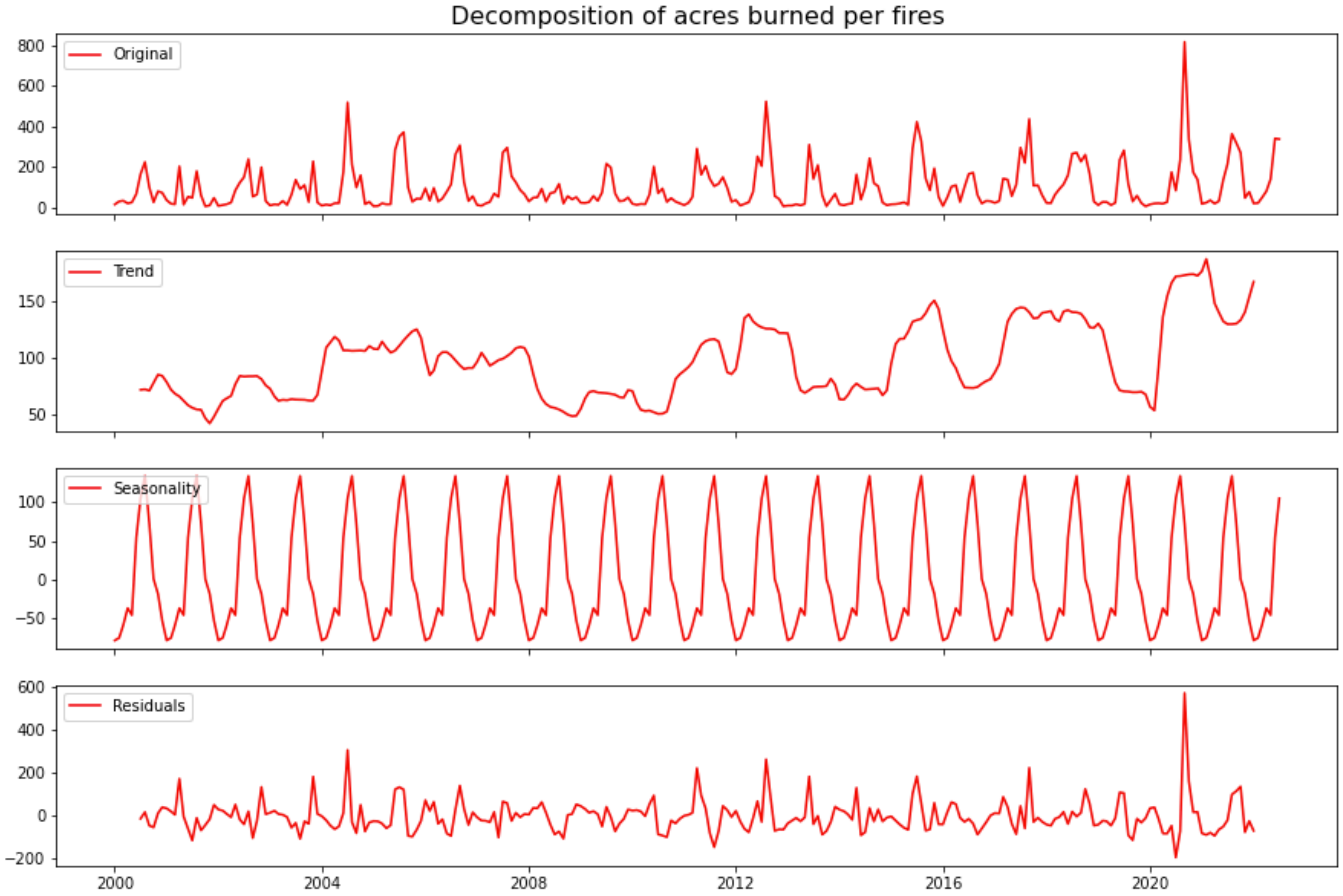
Fires decomposition

We measure a **negative ADF** value and a **p-value very close to 0** for the residuals suggesting that we correctly subtracted the trend and the seasonality from the **number of fires** series


```
ADF =  -9.235056169925187
p-value =  1.6389637626871437e-15
```

Acres burned per fire decomposition

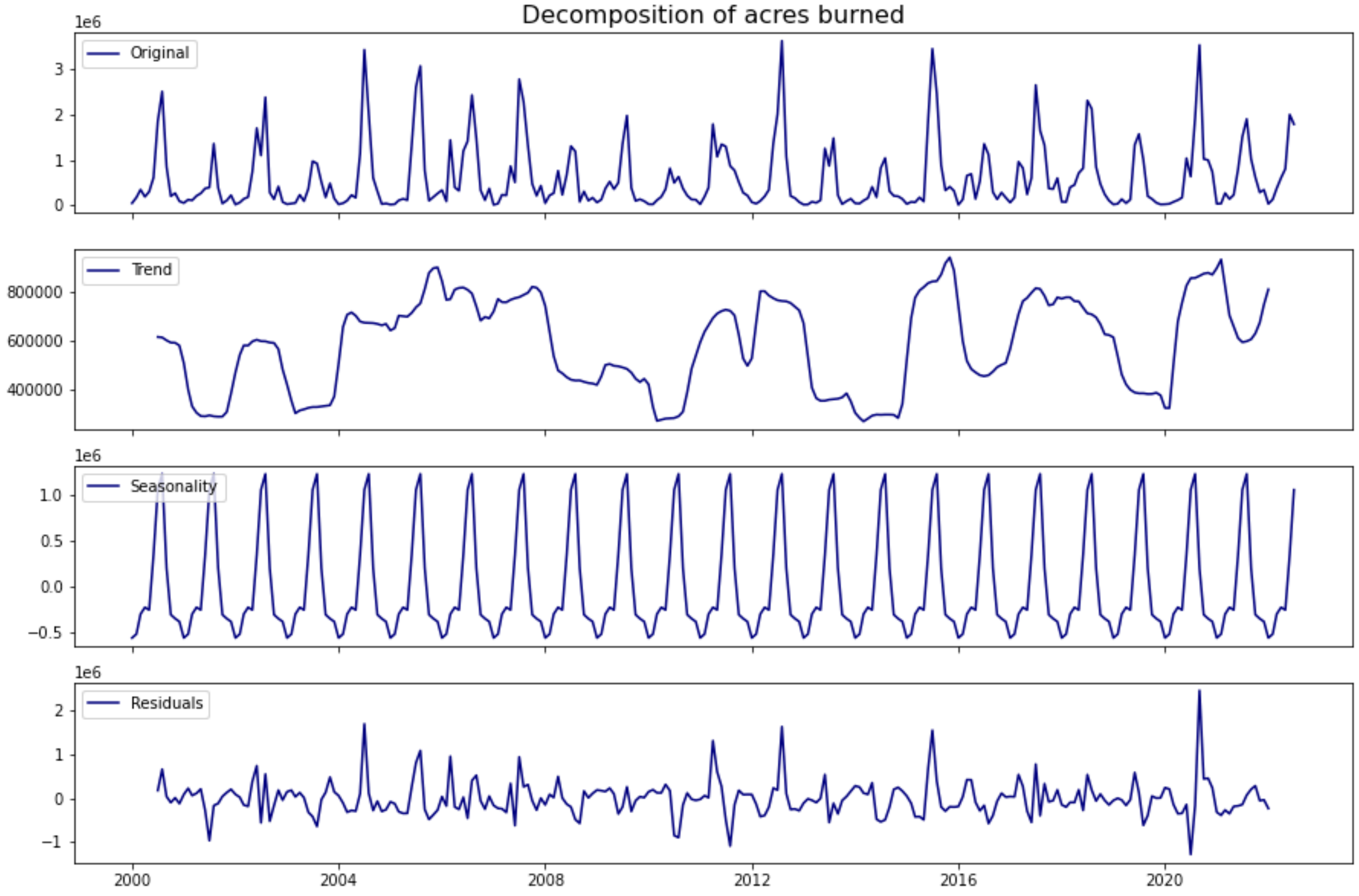
We measure a **negative ADF** value and a **p-value very close to 0** for the residuals suggesting that we correctly subtracted the trend and the seasonality from the **acres burned per fire** series




```
ADF =  -7.537606665563877
p-value =  3.4451499754018575e-11
```

Fires decomposition

We measure a **negative ADF** value and a **p-value very close to 0** for the residuals suggesting that we correctly subtracted the trend and the seasonality from the **total number of acres burned** series




```
ADF =  -7.669011169486848
p-value =  1.6122467703458385e-11
```

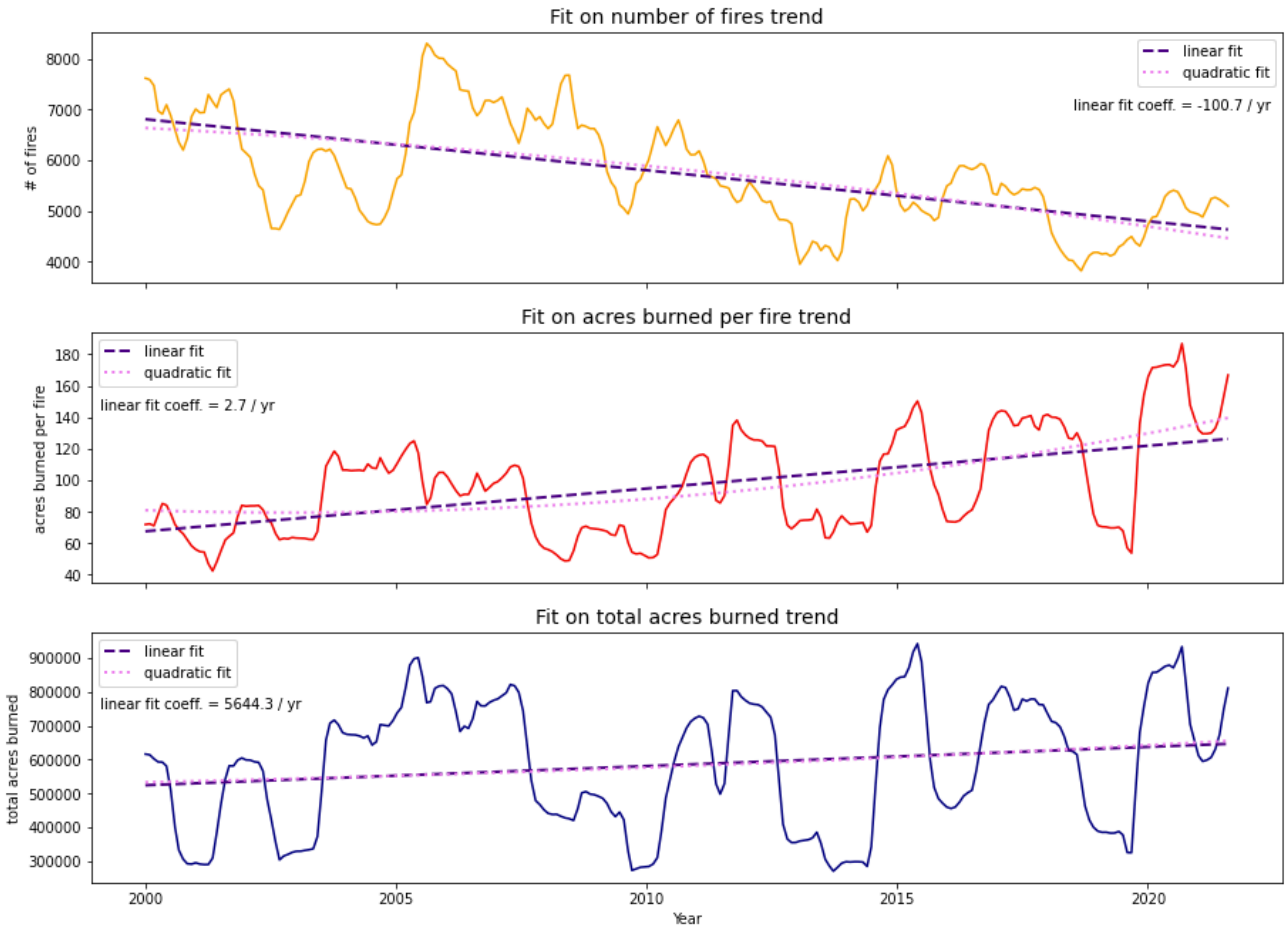
Fits on trends

Now that we have correctly extract the trend over the years for the three features, we can fit them to measure if there is indeed an increasing number of fires and/or force of the fires in term of burned acres over the years and quantify it.

We are interested in a linear fit where the coefficient can be converted into units of interest as a function of time (i.e. "certain amount of number of fires increase/decrease every year"). We performed a quadratic fit as well just to check how far a linear fit is with respect to an eventual non-linear trend (spoiler alert: the trends can be fitted linearly without significant deviations from quadratic fits).

In the next two slides we show the results and the comments on the results, respectively.

Fits on trends: results



Conclusions and future developments

From the previous plot we see that we can safely fit the three trends linearly (quadratic fits almost overlap with linear ones).

In the period of time analyzed (year 2000 to 2020), **the number of fires decreases over time at a rate -100 fires every year**.
However the fires are more violent, burning **+2.7 acres per fire every year**.
The result is an overall increasing of **+5600 more acres burned every year.**

The results of the analysis shows that wildfires in the US became more dangerous over the last 20 years burning more and more acres of vegetation every year :(

In order to assert if the observed increase in wildfire destruction is caused by climate change (heatwaves and droughts in summertime) **we should include historical meteorological data with temperature and humidity at the time and place the wildfire developed**; such geospatial data analysis is out of the scope of this work but it can be carried on for future developments of this project.

Thanks for the peer-review ! I hope you enjoyed the course as I did :)

`Github:` [`https://github.com/fratambot/wildfires-time-series`](https://github.com/fratambot/wildfires-time-series)