

Astrophysics Analysis with SAS

With a PhD in statistical astrophysics, David Corliss works as a data scientist in the automotive industry while continuing astrophysics research, especially in time series modeling. Dr. Corliss is the founder of Peace-Work, a volunteer cooperative of researchers applying analytics to issue-driven advocacy and writes a monthly column on data for good for the American statistical association.



Astrophysics Analysis with SAS

David J Corliss, PhD

10/22/2020



OUTLINE

Moon Phases and Possible Effects

Solar Activity Using Twitter data

Time Series Cluster Analysis

Sunspot Activity

Summary

Computer Operator Errors and Lunar Phases

Impact of Moon Phases on Human Activity

Overview

Question: do moon phases impact human performance?

Method: longitudinal study of computer operator errors matched to lunar phases

Challenge: lunar months vary slightly in duration, requiring renormalization of the time variable

Impact of Moon Phases on Human Activity

```
**** Set Date for the New Moon before and after each observation ****;
```

```
DATE = CPU_DATE + TIME;
```

```
IF DATE < 14645.54 THEN DO;
```

```
NM_BEFORE = 14615.75; NM_AFTER = 14645.54; END;
```

```
ELSE IF DATE < 14675.22 THEN DO;
```

```
NM_BEFORE = 14645.54; NM_AFTER = 14675.22; END;
```

```
.
```

```
. (Repeat for each month in the data)
```

```
.
```

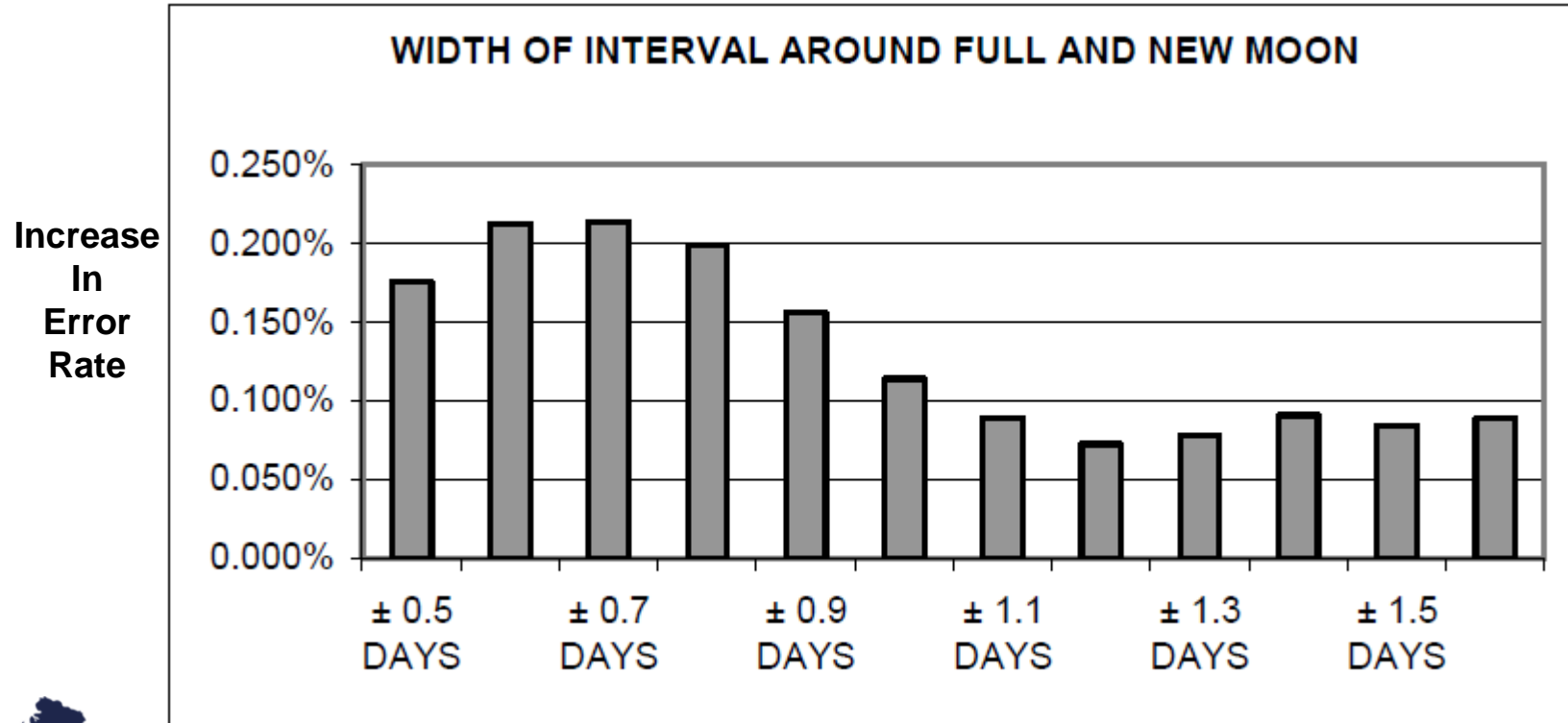
```
**** Interpolate Day of Lunar Month ****;
```

```
L_PCT = ROUND(((DATE - NM_BEFORE) / (NM_AFTER - NM_BEFORE)), .01);
```

```
L_DAY = ( (DATE - NM_BEFORE) / (NM_AFTER - NM_BEFORE) ) * 29.53;
```



Impact of Moon Phases on Human Activity



Impact of Moon Phases on Human Activity

Take-Aways

Tidal forces may have subtle effects of human behavior

Variable event duration can be addressed by transforming the time variable to % completion of the event

Difficult to quantify variables – in this case, the duration of the tidal interval – can be estimated by bootstrapping

Analysis of Solar Activity Using Twitter Data

Analysis of Solar Activity Using Twitter Data Overview

Question: Can social media text reflect physical conditions?

Method: Correlation of electronic equipment failure to solar storm activity

Challenge: estimating the time interval between social media text predicting an action and the timing of the event

Analysis of Solar Activity Using Twitter Data

```
**** Creating lags for the event that happens first ****;
```

```
Data twitter_time_series;  
  set twitter_time_series;  
  solar_lag1=lag1(solar_activity);  
  solar_lag2=lag2(solar_activity);  
  solar_lag3=lag3(solar_activity);  
run;
```

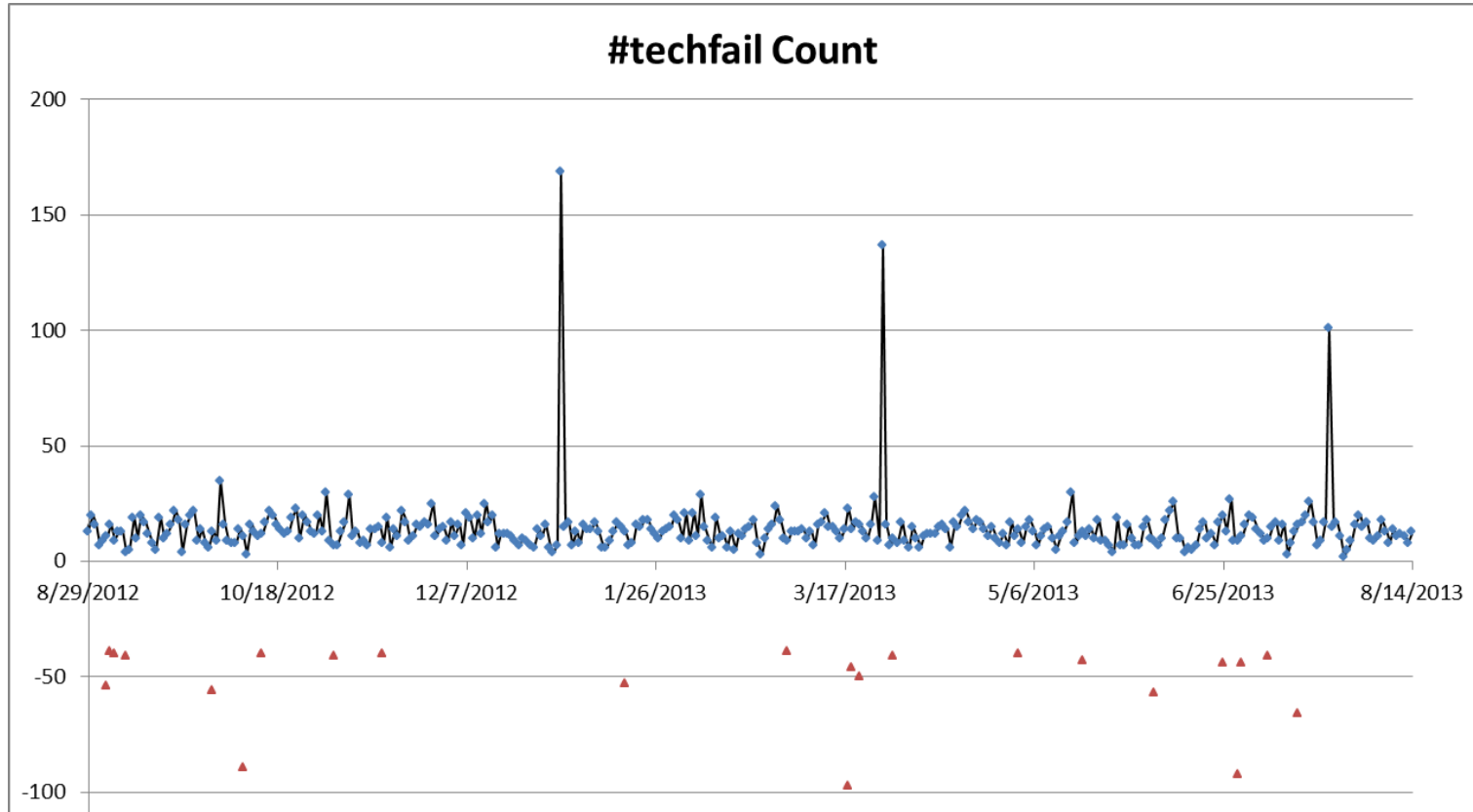
```
**** Time Series Correlation ****;
```

```
proc corr twitter_time_series outp=correlations;  
  var twitter_count solar_activity solar_lag1 solar_lag2 solar_lag3;  
run;
```

```
proc print data=correlations;  
  where _n_ = 1;  
run;
```



Analysis of Solar Activity Using Twitter Data



2 σ + Solar Storms



Analysis of Solar Activity Using Twitter Data

Take-Aways

Social media text can be predictive of latent environmental parameters and can be matched to events for time series analysis

57.9% of top #techfail days follow a solar storm; a burst of #techfail tweets is 3.6 times more likely following a $2\sigma+$ storm

Analysis of social media data at high volume and / or low latency may require special methods optimized for big data

Time Series Cluster Analysis

Time Series Cluster Analysis of Stellar Events

Overview

Question: How to identify stages in development a complex evolving system – high velocity events in hot stars?

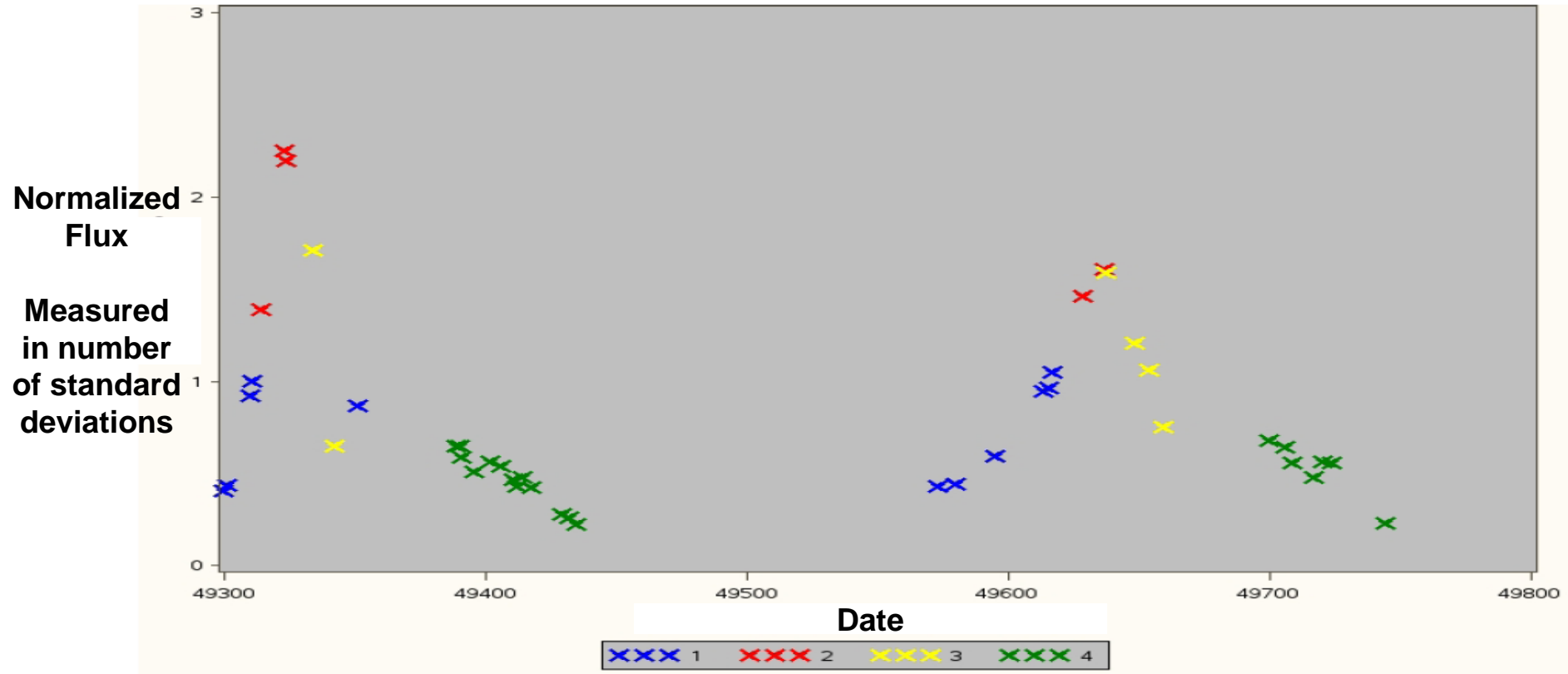
Method: Applying standard clustering methods to a time series, resulting in successive intervals with distinctive behavior

Challenge: Leveraging events following the same pattern over time with different intensities

Time Series Cluster Analysis of Stellar Events

```
**** Standardization of Characteristics Before Clustering ****;  
  
proc standard data=work.star mean=0 std=1  
    out=work.star_stan;  
    var percent_duration flux flux_change absorption absorption_change;  
run;  
  
**** Clustering Algorithm ****;  
  
proc fastclus data=work.star_stan maxc=6 maxiter=20  
    out=work.cluster1;  
    var percent_duration flux flux_change absorption absorption_change;  
run;
```


Time Series Cluster Analysis of Stellar Events



Time Series Cluster Analysis of Stellar Events

Take-Aways

Applying clustering methods to a time series results in clusters that are a succession of intervals within internally consistent behavior

Standardization using PROC STANDARD – transforming an outcome variable into the number of standard deviations above and below a long-term mean – allows comparison of events with similar patterns that differ in intensity

Method can be applied in longitudinal studies to detect changes in seasonal pattern – for example, climate change

Analysis of Sunspot Activity

Sunspot Activity With PROC SPECTRA

Overview

Question: Longitudinal study of sunspots as an indicator of internal solar processes – are there long-term patterns?

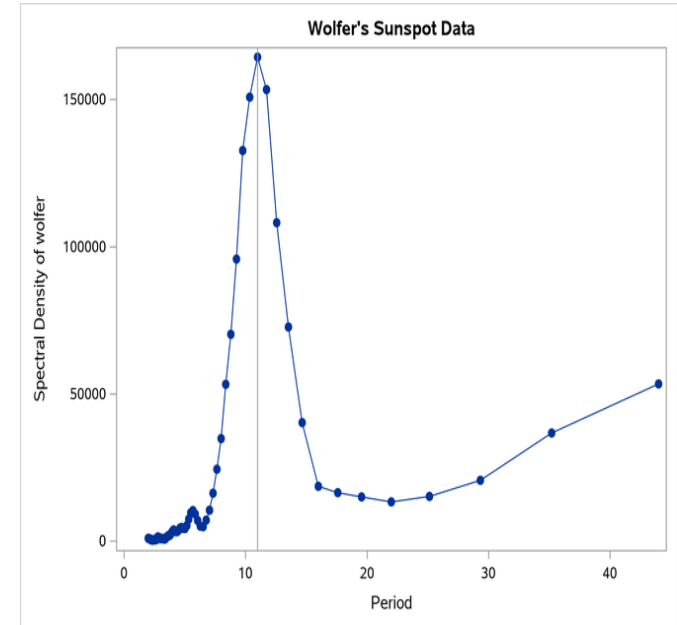
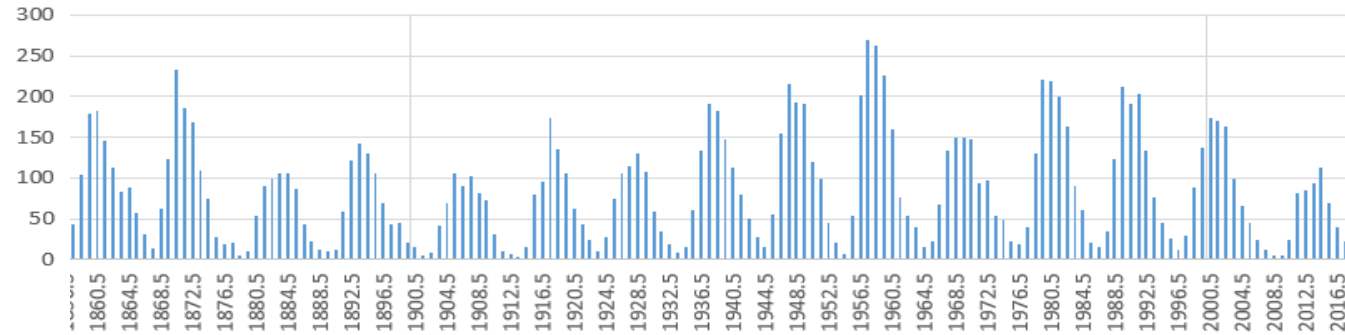
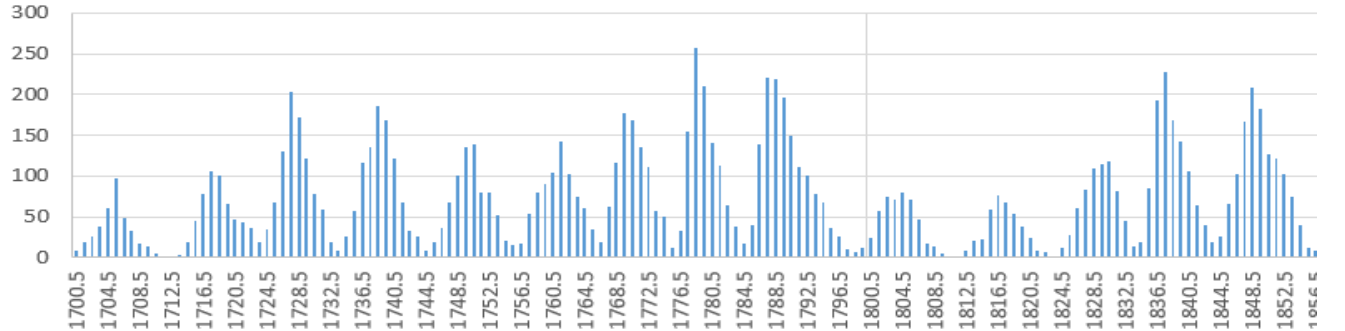
Method: Spectral analysis using SAS ETS procedure SPECTRA

Challenge: Very long time series (320 years) results in long-term variations in the data, with the potential for additional periodicities

Sunspot Activity With PROC SPECTRA

```
**** Analysis of periodicities using PROC SPECTRA ****;  
  
proc spectra data=sunspot out=b p s adjmean whitetest;  
    var annual_mean_count;  
    weights 1 2 3 4 3 2 1;  
run;  
  
proc print data=b(obs=12);  
run;
```

Sunspot Activity With PROC SPECTRA



Sunspot Activity With PROC SPECTRA Take-Aways

Spectra analysis of longitudinal sunspot data identifies a periodicity of about 11 years

Long-term variation is observed, with a low point in the early 1800's and a high in the mid-1900's

Spectral analysis, confirmed using the well-known solar cycle, can be applied to many diverse time series problems

SUMMARY

Summary: Astrophysics with SAS

SAS is a high-performing analytic tool for astrophysics

Researchers in astrostatistics develop new methods that can be applied to many areas

Astrostatistics research is especially helpful for time series analysis, data visualization, and big data applications



Questions?



David J Corliss, PhD
Phone: 734.837.9323
E-mail: dcorliss@gmail.com
Twitter: @dcorliss_astro

