



Solar activity & UFOs

Is there a relationship between sunspot activity and UFO sightings?

Amanda Parker

Designer

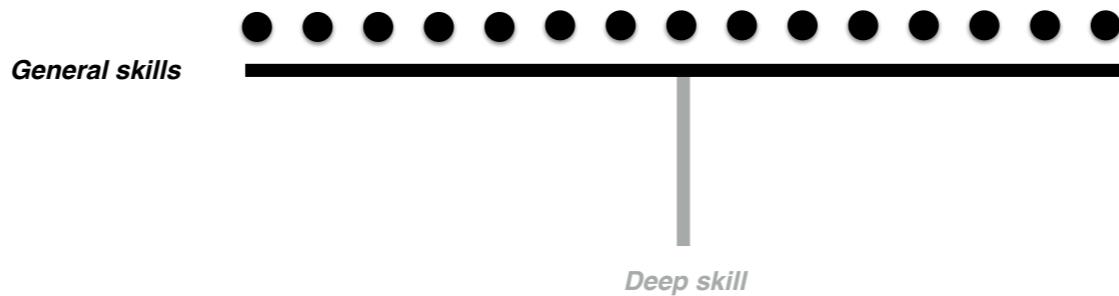
Ph.D Psychology (Perceptual Systems)
Teaching Interaction Design as well

This project explores whether solar activity (sunspot observations) can predict UFO reports or other light/percpetual phenomena as it does the aurora.

Career tips + presentation tips at the end...

First, introduce myself

T-model

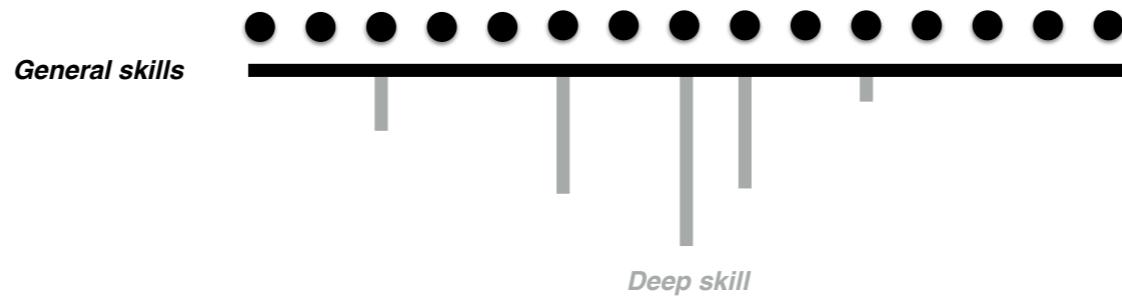


Career advice bit

Can talk about myself:

Ask audience if (or how) t-model applies to them?

T-model



Career advice bit

Can talk about myself:

Ask audience if (or how) t-model applies to them?

MODERN DATA SCIENTIST

Data Scientist, the sexiest job of the 21th century, requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.



MATH & STATISTICS

- ★ Machine learning
- ★ Statistical modeling
- ★ Experiment design
- ★ Bayesian inference
- ★ Supervised learning: decision trees, random forests, logistic regression
- ★ Unsupervised learning: clustering, dimensionality reduction
- ★ Optimization: gradient descent and variants

PROGRAMMING & DATABASE

- ★ Computer science fundamentals
- ★ Scripting language e.g. Python
- ★ Statistical computing packages, e.g., R
- ★ Databases: SQL and NoSQL
- ★ Relational algebra
- ★ Parallel databases and parallel query processing
- ★ MapReduce concepts
- ★ Hadoop and Hive/Pig
- ★ Custom reducers
- ★ Experience withaaS like AWS

DOMAIN KNOWLEDGE & SOFT SKILLS

- ★ Passionate about the business
- ★ Curious about data
- ★ Influence without authority
- ★ Hacker mindset
- ★ Problem solver
- ★ Strategic, proactive, creative, innovative and collaborative

COMMUNICATION & VISUALIZATION

- ★ Able to engage with senior management
- ★ Story telling skills
- ★ Translate data-driven insights into decisions and actions
- ★ Visual art design
- ★ R packages like ggplot or lattice
- ★ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau

MarketingDistillery.com is a group of practitioners in the area of e-commerce marketing. Our fields of expertise include: marketing strategy and optimization: customer tracking and on-site analytics; predictive analytics and econometrics; data warehousing and big data systems; marketing channel insights in Paid Search, SEO, Social, CRM and brand.

Marketing
DISTILLERY
(c) Krzysztof Zawadzki

Data scientist infographic

Need to be competent across all areas, but specialise

Importance of team work

Back to UFOs



<http://midnightinthedesert.com/ufo-sightings-peak-summer/>

Given my perceptual systems focus, I'm interested in perceptual phenomena, like illusions, hallucinations and things like UFO reports and reports of other light phenomena...

There is reason to believe that UFO reports can be attributed to natural phenomena.

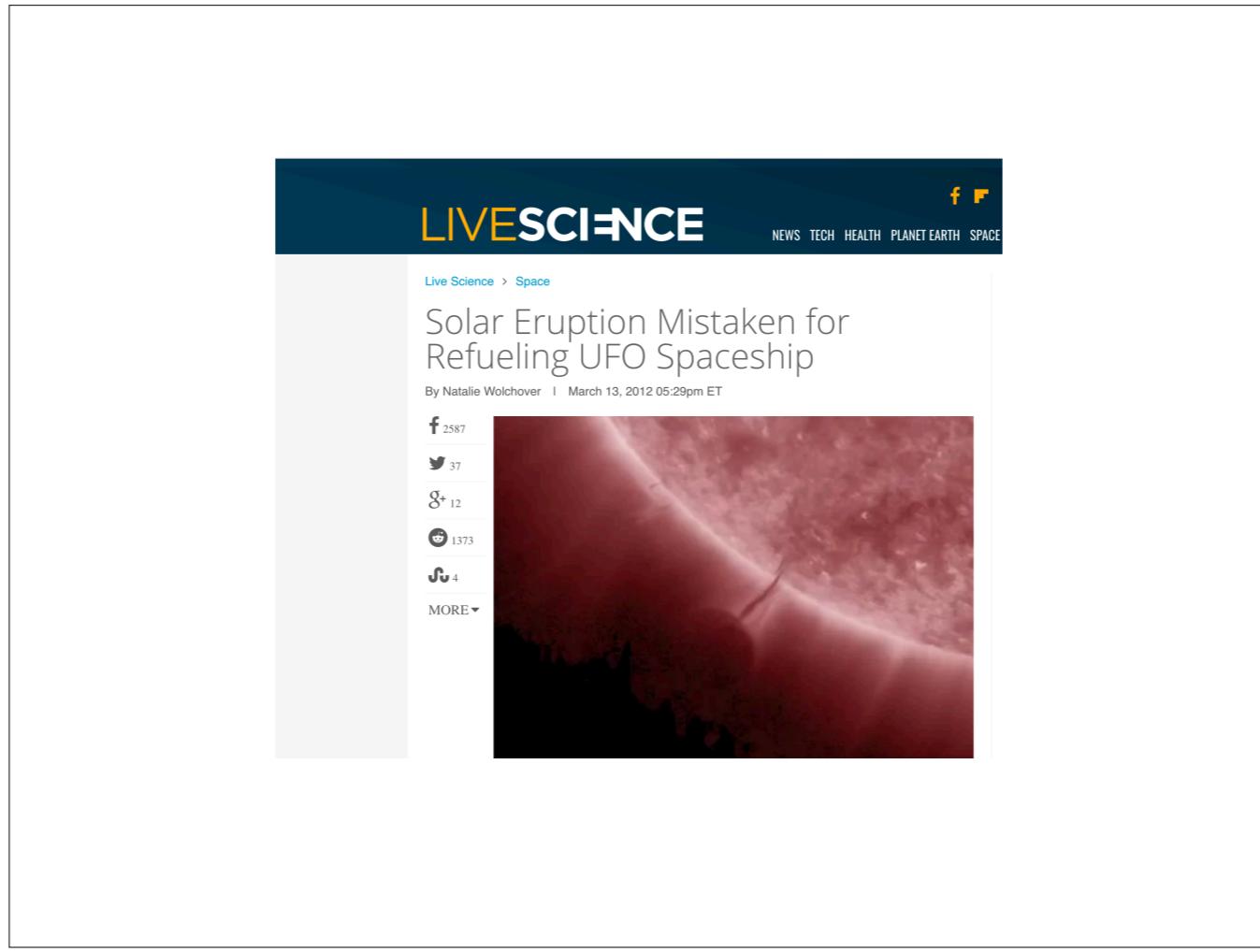
UFO reports peak in the summer, when the sky is clearer : <http://midnightinthedesert.com/ufo-sightings-peak-summer/>

The screenshot shows the Daily Mail Australia website's Science & Tech section. At the top left is the 'Daily Mail AUSTRALIA' logo. To its right is a decorative graphic of molecular structures and the text 'Science & Tech'. The navigation bar includes links for Home, U.K., U.S., News, World News, Sport, TV&Showbiz, Femail, Health, Science (which is highlighted in orange), Weather, Video, Travel, and Fashion Finder. Below the navigation bar are links for Latest Headlines, Facebook, YouTube, Google, and eBay. On the far right of the header is a 'Login' link. The main content area features a bold headline: "'Aliens are controlling the SUN': UFO hunters say they have spotted strange ships near the solar surface in latest bizarre claim". Below the headline is a bulleted list of three items. At the bottom right of the content area is a search bar with fields for 'Site' and 'Web', a search input field with placeholder text 'Enter your search', and a magnifying glass icon labeled 'Search'.

'Aliens are controlling the SUN': UFO hunters say they have spotted strange ships near the solar surface in latest bizarre claim

- 6-7 years ago Russian scientist had claimed to see UFOs by sun everyday
- Images originally capture by Solar Heliospheric Observatory
- Shows glowing orbs floating around the sun in different spots

Numerous pop culture articles on the internet link UFO reports to activity from the sun. The visual appearance of sunspots shares similarities with sun-linked UFO reports (for example seeing a number of black disks 'fly' across the sun, or this example).



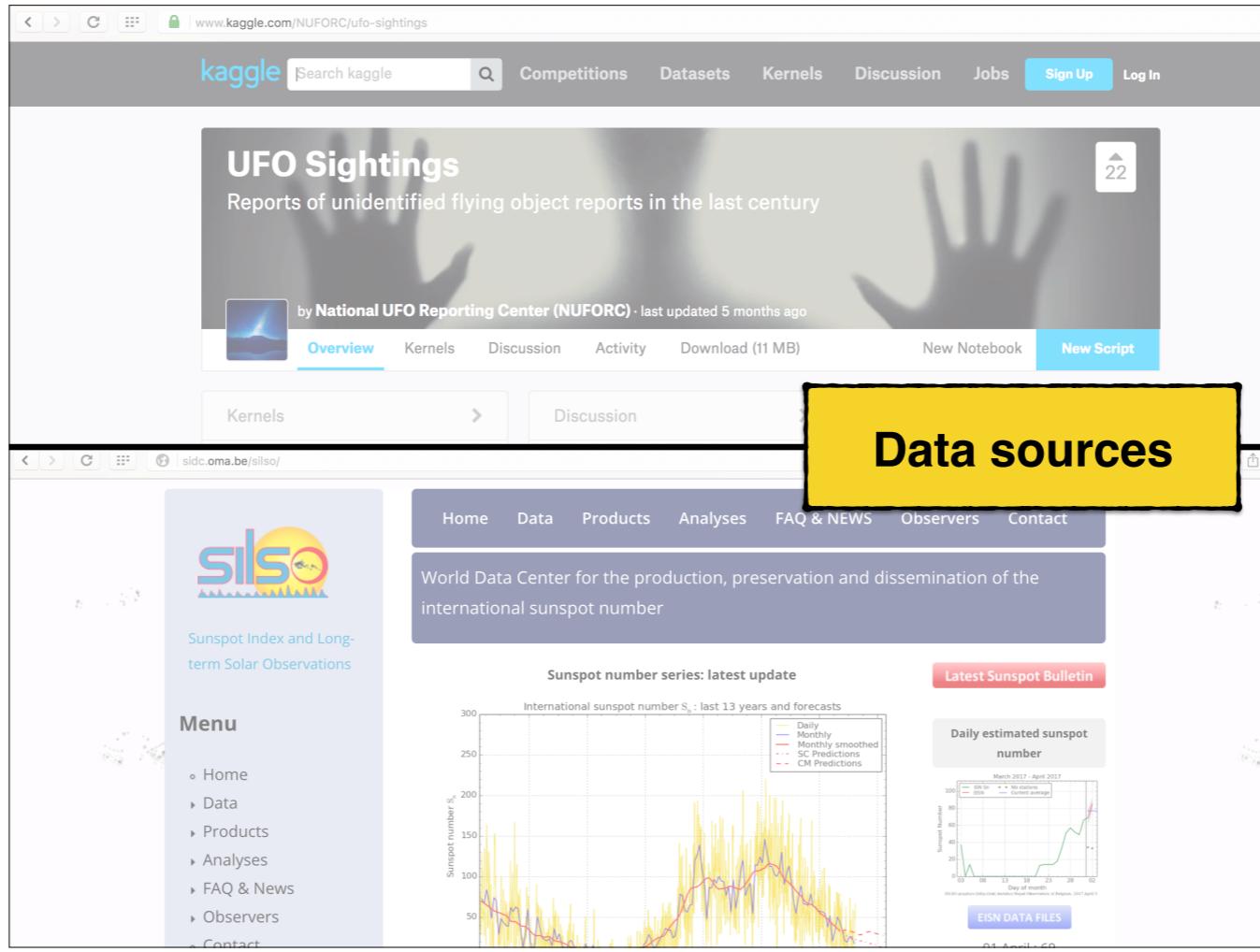
, or this example

The academic literature contains some references to a correlation between solar activity and UFO reports, but there is sparse recent research.

We already know: Atmospheric and solar conditions are linked to the aurora

Inductive reasoning has its place in the scientific method. (from the singular observation to the general). Scientists use it to form hypotheses and theories. **Deductive** reasoning (then) allows them to apply the theories to specific situations (from the general to the singular)

This part is what I call ‘Exploratory Data Analysis’ or simply exploring data. You look at the data and the wider context (domain) to form a *hypothesis*. Once you have an *idea or theory of what might be going on (in the data)*, you then test it, or in DS try to model that theory and see if it can predict the observed pattern with high accuracy.



This analysis revisits the link between solar activity and UFO reports with more recent (and larger) UFO data set than used previously along with sunspot observation data from the Royal Observatory of Belgium website <http://sidc.oma.be/silso/>. The UFO data analysed was mostly collected through an online reporting database NUFORC.org and covers the period of the internet (1996-onwards) increasing report rates. Data older than the nuforc.org website registration in 2001 was presumably added manually.

The value of intersecting data sets

Hypothesis

The theory - UFO reports are actually mistaken sunspot observations

If: UFO reports are mistaken sunspot observations

Then: Sunspot observation time series will predict UFO report time series

If our theory ____ is true,
we predict ____ will happen

Methods

1. Data formatting
2. Signal coherence
3. Time series decomposition
4. Granger causality test
5. ARIMA
-Can one series forecast the other?

Knowing which statistical tests to use and when,
as important as knowing how to perform them

Libraries

```
In [1]: #Compares UFO and sunspot observations over time
import numpy as np
import pandas as pd
from scipy import signal
import statsmodels.tsa.stattools
import statsmodels.api as sm
import matplotlib.pyplot as plt
%matplotlib inline
plt.style.use('seaborn-darkgrid')
```

```
In [2]: # load UFO data
ufo_data = pd.read_csv('scrubbed.csv', usecols=[0, 4, 5], low_memory=False)
ufo_data['datetime'] = pd.to_datetime(ufo_data['datetime'], errors='coerce')
```

Load data

```
In [21]: # load sunspot data M
s2 = pd.read_csv('data_sunspots_m.csv')
s2['datetime'] = pd.to_datetime((s2.YEAR*10000+s2.MONTH*100+1).apply(str),format='%Y%m%d')
```

```
In [31]: # load sunspot data
sunspots = pd.read_csv('data_sunspots.csv', usecols=[0, 1, 2, 5])# http://sidc.oma.be/silso/newdataset
sunspots['datetime'] = pd.to_datetime((sunspots.YEAR*10000+sunspots.MONTH*100+sunspots.DAY).apply(str),format='%Y%m%d')
```

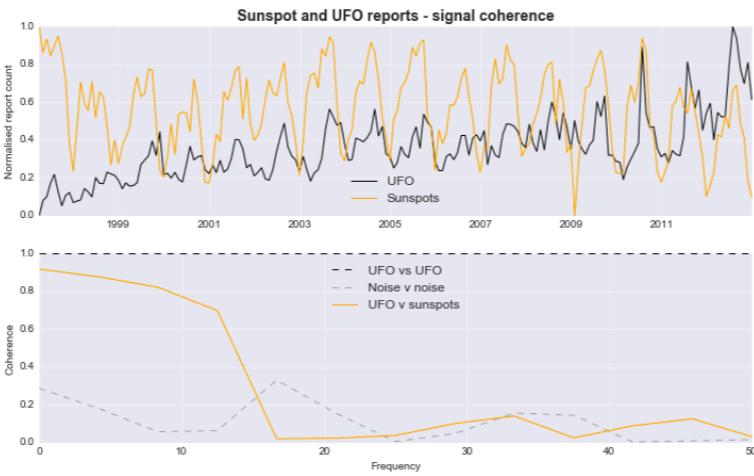
Format data

```
In [32]: # Format and group sunspot data by month
df2 = sunspots
df2 = df2[(df2['datetime'] > '1997-03-30') & (df2['datetime'] <= '2012-12-31')]
df2m = df2.set_index('datetime').groupby(pd.TimeGrouper('1M')).mean()
```

```
In [33]: # Format and group UFO rows by month
df1 = ufo_data
df1 = df1[(df1['datetime'] > '1997-03-30') & (df1['datetime'] <= '2012-12-31')]
df1m = df1.set_index('datetime').groupby(pd.TimeGrouper('1M')).agg(['count'])
```

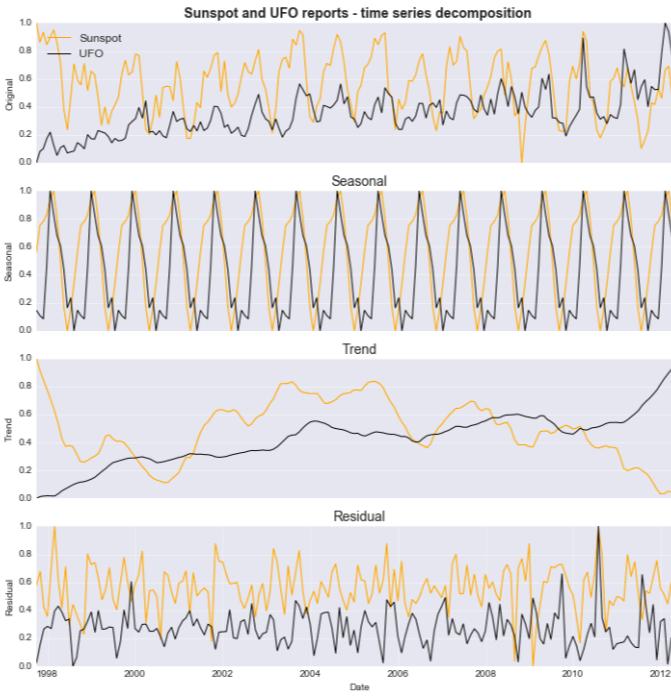
```
In [34]: # # Add month index to join on
df1m['YEAR'] = df2m['YEAR']
df1m['MONTH'] = df2m['MONTH']
```

```
In [35]: # Merge UFO and sunspot data
merged = pd.merge(df2m,df1m, on=['YEAR','MONTH'], left_index=True)
idx = merged.index
```



```
In [19]: # UFO - test signal coherence with sunspot counts
nfft = 24
dt = 0.01
n = 1./dt
plt.figure(figsize=(12,7))
plt.subplot(211)
plt.title('Sunspot and UFO reports - signal coherence', fontsize=14, fontweight='bold')
plt.plot(Y1n,'k',label="UFO")
plt.plot(Y2n,color='orange',label="Sunspots")
plt.ylabel('Normalised report count')
plt.legend(loc='lower center')
plt.subplot(212)
cxy, f = plt.coher(Y1n, Y1n, nfft,n , color='k', alpha=1, linestyle='--', label='UFO vs UFO')
cxy, f = plt.coher(np.random.randn(len(Y1)), np.random.randn(len(Y1)), nfft,n , color='k', alpha=.3, linestyle='--', label='Noise v noise')
cxy, f = plt.coher(Y1n, Y2n, nfft, n, color='orange', alpha=1, label='UFO v sunspots')
plt.ylabel('Coherence')
plt.legend(loc='upper center')
plt.ylim(0,1)
plt.savefig('./imgs/UFOsunspots_freqCoherence.png')
plt.show()
```

Signal coherence



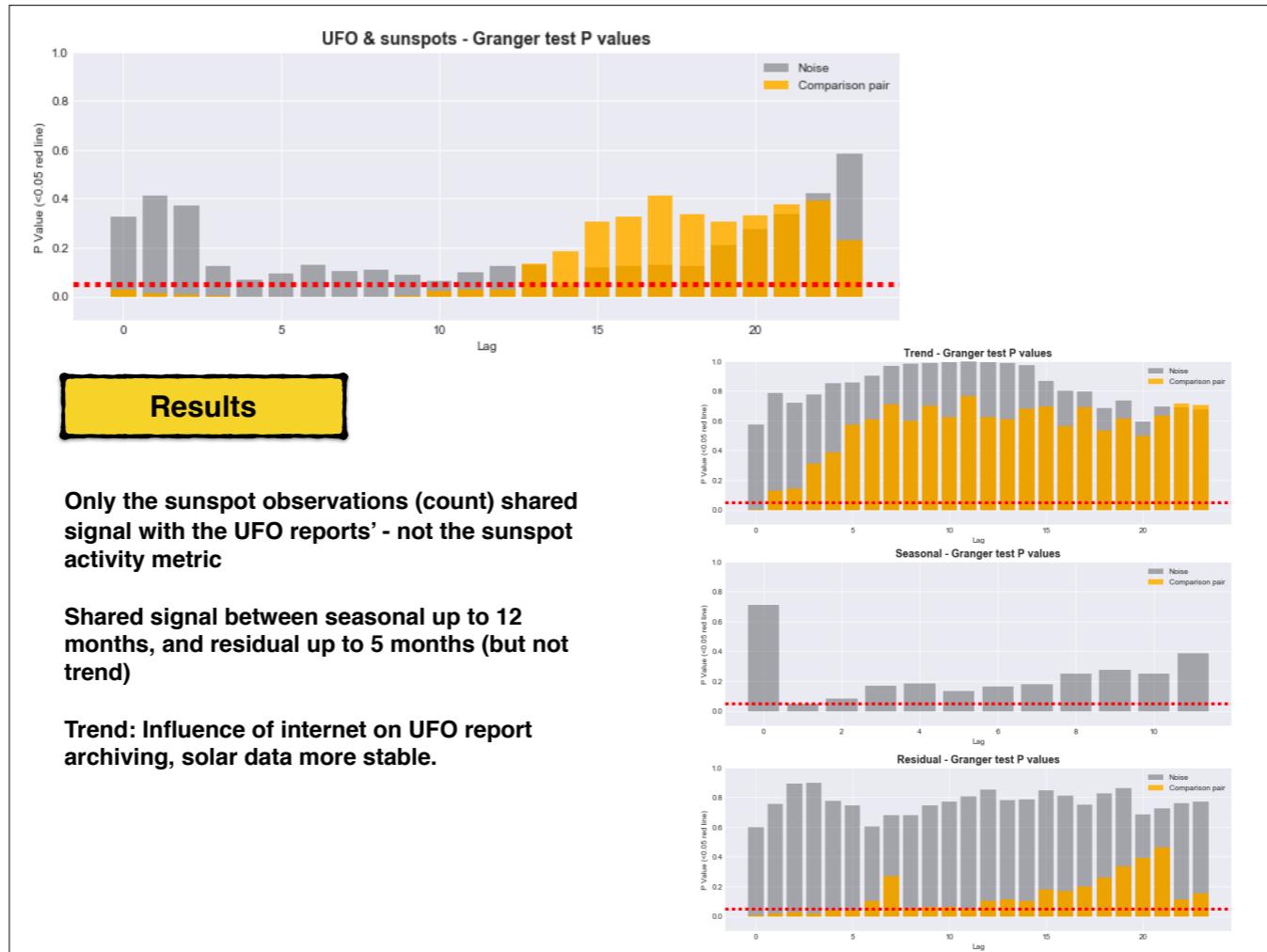
```
In [44]: # Time series decomposition
dta = Y1
# deal with missing values. see issue
dta.interpolate(inplace=True)
res = sm.tsa.seasonal_decompose(dta)
# graph = res.plot()
# graph.suptitle('UFO reports - time series decomposition', fontsize=14, fontweight='bold')
# graph.set_size_inches(10,8)
```

Timeseries decomposition

Granger test

```
In [23]: # Granger test for Y1 x Y2
sig_noise = Y1 + np.random.randn(len(Y1))
x1 = pd.DataFrame([Y1,Y2])
x1 = x1.transpose().copy()
x2 = pd.DataFrame([Y1,sig_noise])
x2 = x2.transpose().copy()
x3 = pd.DataFrame([np.random.randn(len(Y1)),np.random.randn(len(Y1))])
x3 = x3.transpose().copy()
maxlag = 24
gr1 = statsmodels.tsa.stattools.grangercausalitytests(x1, maxlag, verbose=False)
gr2 = statsmodels.tsa.stattools.grangercausalitytests(x2, maxlag, verbose=False)
gr3 = statsmodels.tsa.stattools.grangercausalitytests(x3, maxlag, verbose=False)
```

```
In [24]: #Plot the granger p values
resd1 = {}
for i in range(0,maxlag):
    resd1[i] = gr1[i+1][0]['params_ftest'][1]
resd2 = {}
for i in range(0,maxlag):
    resd2[i] = gr2[i+1][0]['params_ftest'][1]
resd3 = {}
for i in range(0,maxlag):
    resd3[i] = gr3[i+1][0]['params_ftest'][1]
pvals1 = pd.Series(resd1)
pvals2 = pd.Series(resd2)
pvals3 = pd.Series(resd3)
plt.figure(figsize=(12,4))
plt.ylim(-0.1,1)
plt.bar(pvals2.index,pvals2,color='k',alpha=.1, label='UFO+noise')
plt.bar(pvals3.index,pvals3,color='k',alpha=.3, label="Noise")
plt.bar(pvals1.index,pvals1, color='orange', label='UFO+sunspot')
plt.axhline(0.05, ls=':', c='r', linewidth=4)
plt.legend(loc='upper center')
plt.title('Sunspot and UFO reports - Granger test P values', fontsize=14, fontweight='bold')
plt.xlabel('Lag')
plt.ylabel('P Value (<0.05 red line)')
plt.savefig('./imgs/UFOsunspots_granger_pvals.png')
```





Atmospheric visibility

Strong seasonal component like previously reported (sig coh lag up to 12 mths)

What about the residual test? - sunspot observations are predictive separately around 5 months

UFO sighting trend appears determined by growth of internet

Sunspot trend is more cyclical

Solar activity

Maybe the solar activity is not really important if poor visibility means you can't see these fluctuations at all under the seasonal pattern - solar cycle is quite long - may need to test with a longer time series.

Can't rule out other sources of UFO sightings (there are lots of other natural light effects too)

New theory: All natural light phenomena predicts UFO sightings

Next steps: UFO sighting shape (type) linked to light phenomenon type?

Audience: Do you have any theories?