machineTRUST: Knn, Trend, Noise and Correlation

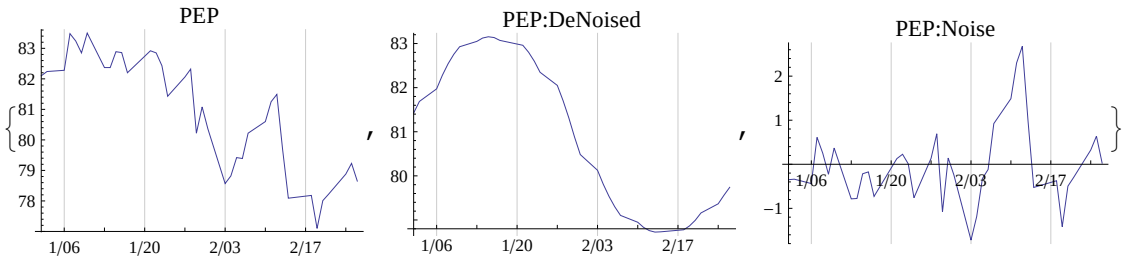
Feb 24-28 2014

Analyst: Dara O Shayda
Principle Engineer: Roshan Rush
Visualization: Lee Matthew
Systems: Wayne Allen

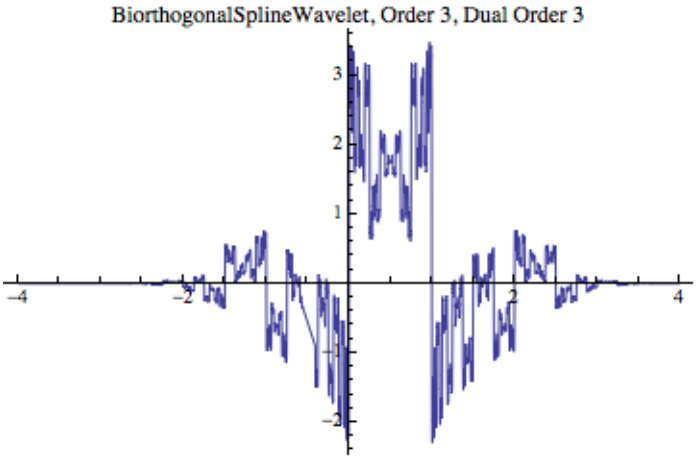
Portfolio

“IBM”,“AAPL”,”AMC”,”GE”,”C”,”XOM”,”BA”,”IP”,”HPQ”,“MRK”,“KO”,”MMM”,”DD”,”DIS”,”PG”,”MCD”,”JNJ”,”MO”,”CVX”,”JPM”,”AA”,”AEP”,”ED”,”CAT”,”UTX”,”HON”,”DTE”,”CNP”,”MRO”,”WMT”,”AXP”,”DAL”,”DOW”,”PEP”,”BMY”,”LMT”,”XRX”,”GD”,”KR”,”WY”,”CL”,”F”,”AET”,”UNP”,”FDX”

Choose a stock e.g. PEP and decompose its closing-price signal into **Trend** (DeNoised) and **Noise**, Jan 1 to Feb 26 2014:



Bi-Orthogonal Spline Wavelet [3, 3], a **Discrete Transform** used for decomposition of the signals:



Once such decomposition happens then we have 2 **Normed Vector Spaces**, by normed we means vectors have length, by vector we mean arrays of number (in this case) which could added or subtracted or scaled:

1. **Trend** space which has a metric for measuring distances
2. **Noise** space with metric as well

We choose for both spaces, the metric **Correlation Distance**:

$$1 - \frac{(u-\mu_u)\cdot(v-\mu_v)}{\|u-\mu_u\|\,\|v-\mu_v\|}$$ '·' is scalar product

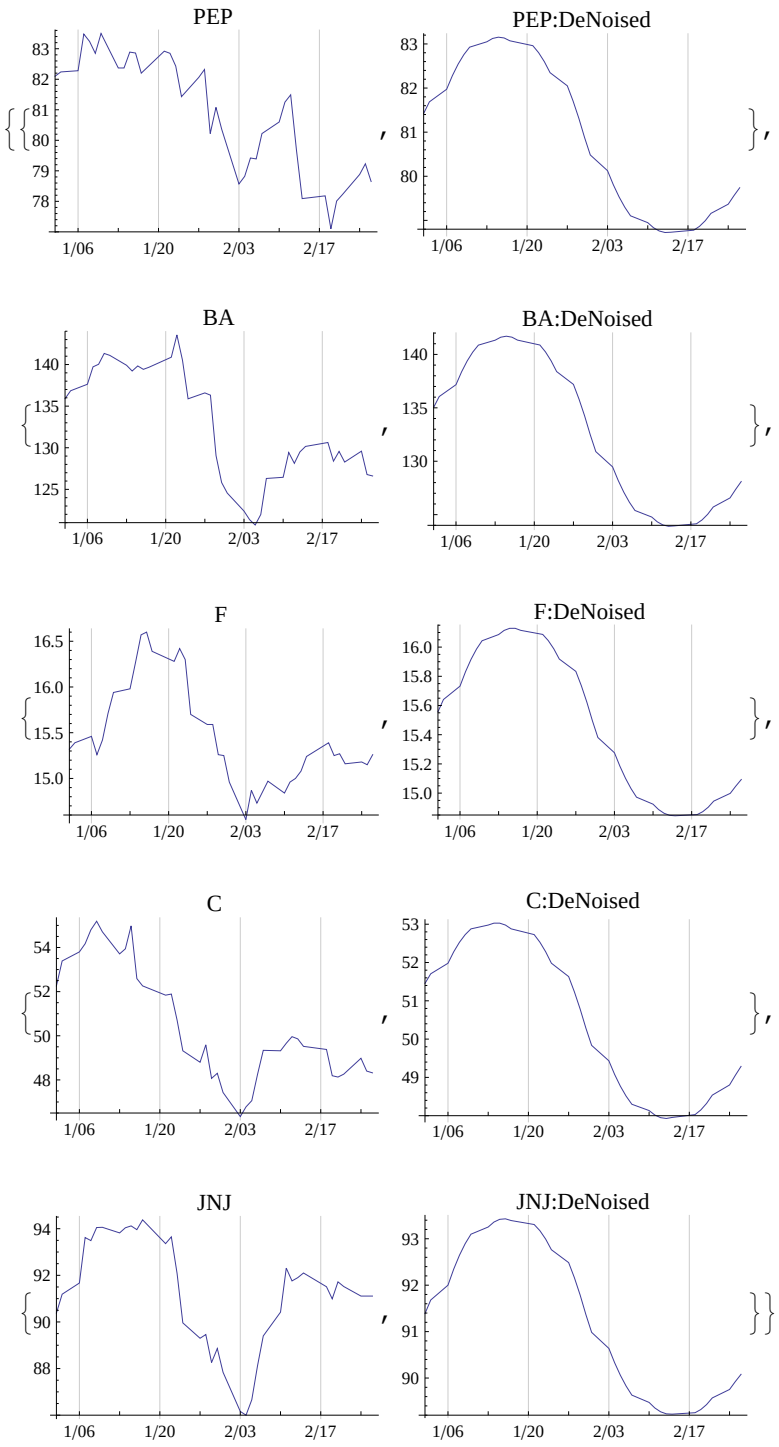
Therefore we could then investigate the nearness and farness between different signals as well as within the signals' decompositions.

Nearest Neighbor Trend

Then find the 5 Nearest portfolio stocks by Correlation Distance, including itself:

"PEP", "BA", "F", "C", "JNJ"

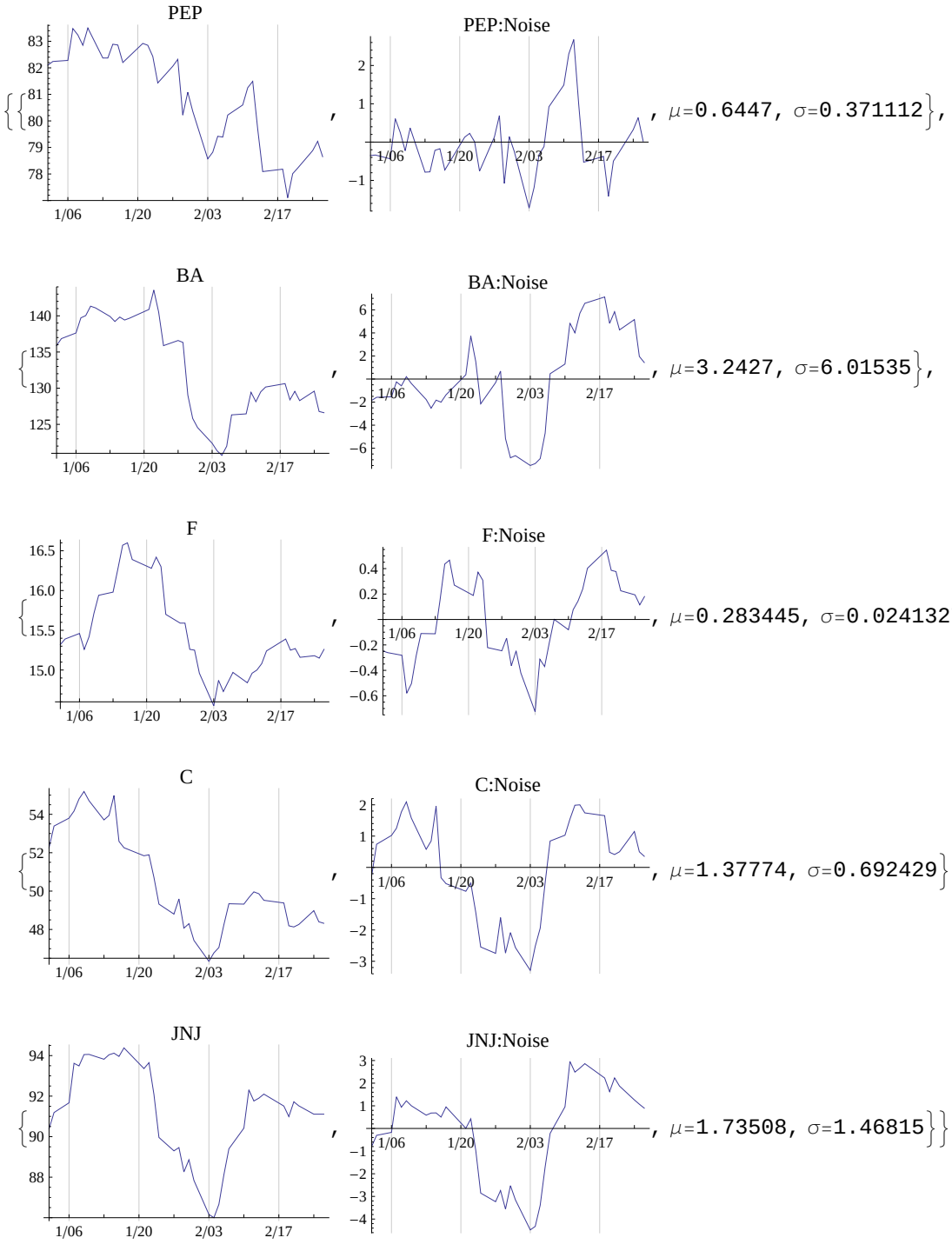
Let's look at the nearest neighbors and their Trend:



It is astonishing that although the closing-price signals look not even vaguely similar, their Trend curves are indeed amazingly similar in shape!

Remark 1: *The DeNoised curves although looking the same are numerically different than each other.*
Remark 2*: The Nearest Neighbors are not in the same sectors e.g. PEP is a soda company and BA is an aeronautics company!*

If we look at the corresponding Noise signal for the above Nearest Neighbors, surprisingly we mismatches e.g. BA:



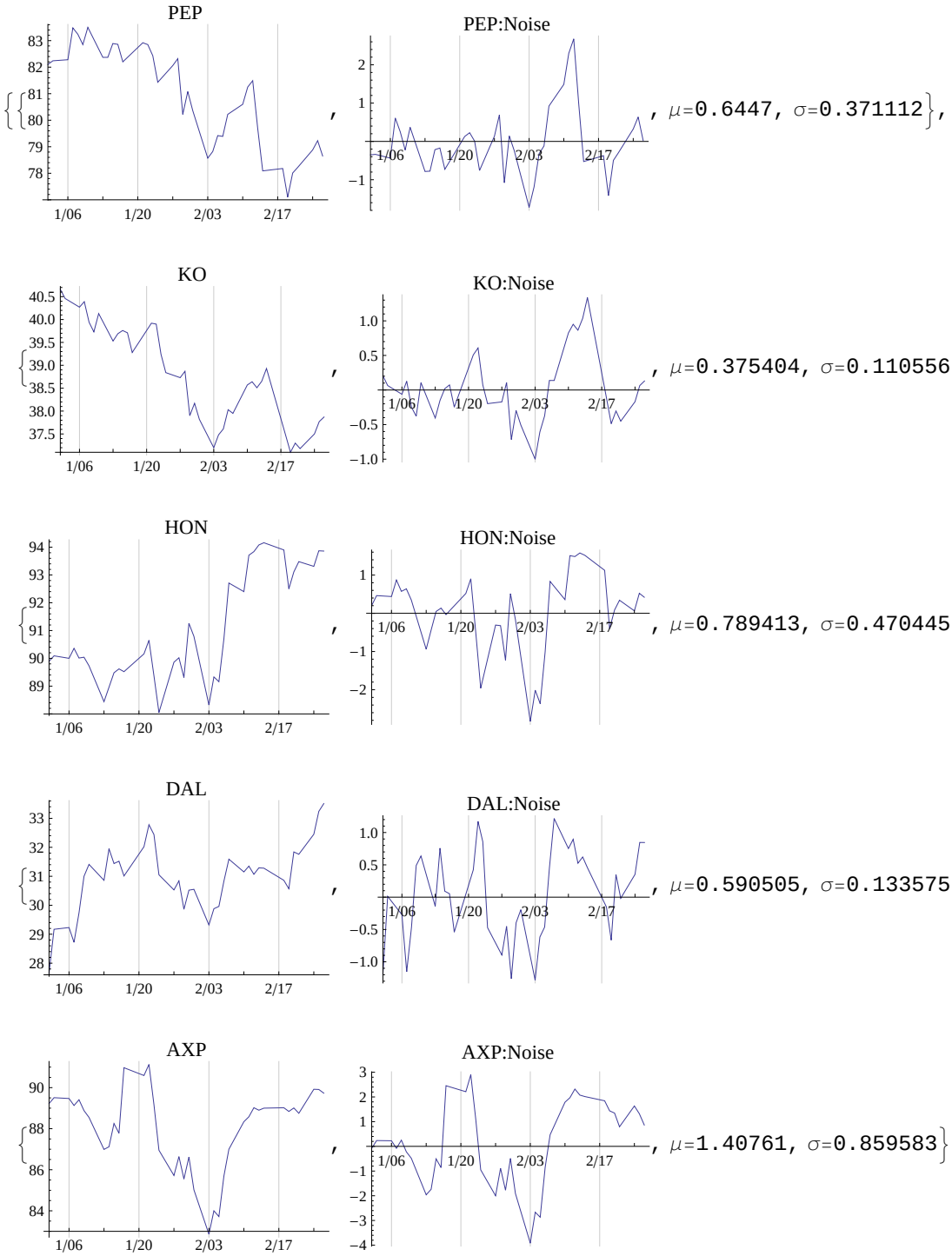
First number stands for Mean of absolute value of Noise and second number is the variance of the absolute value of the Noise. As in the case of BA there is a gross mismatch in spite of close correlation

of BA’s Trend to PEP.

Nearest Neighbor Noise

Compute the **Nearest Neighbor** algorithm on Noise part of the closing-price signal, not so surprisingly we find much closer matches, a different list:

“PEP”, ”KO”, ”HON”, ”DAL”, ”AXP”



Therefore we easily could conclude:

Nearest Neighbor algorithm with Correlation Distance should run on Trend and separately on the Noise.

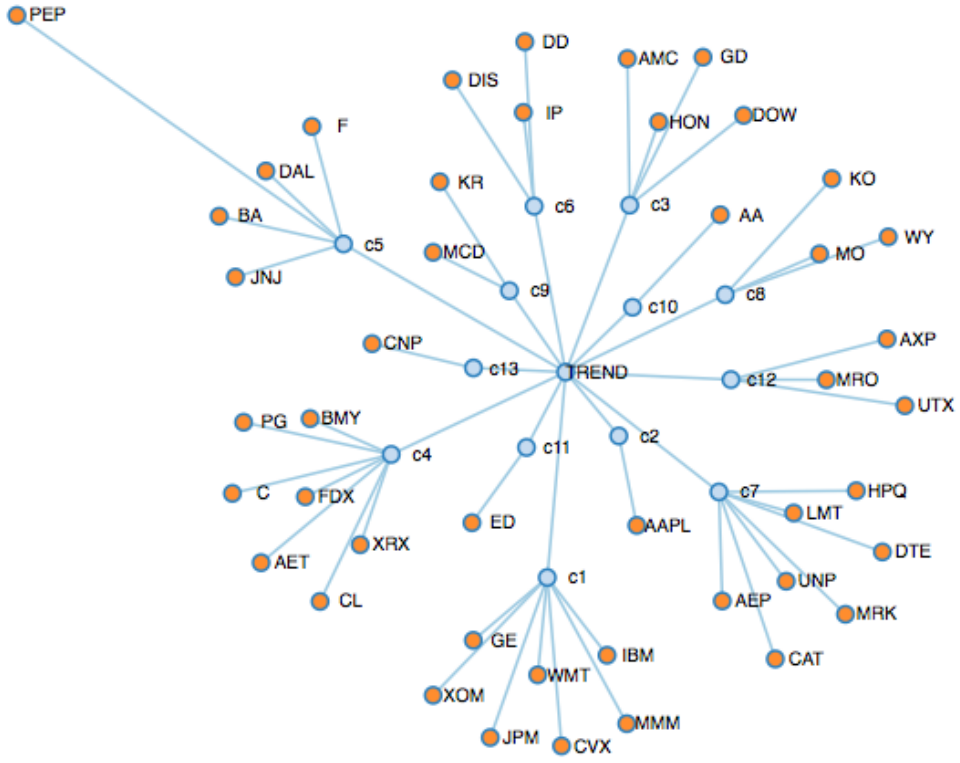
We call this **decoupling of Trend from Noise**.

Remark 3: *PEP's (Pepsi) Nearest Neighbor for Noise is KO (Coca Cola) but not for Trend!*

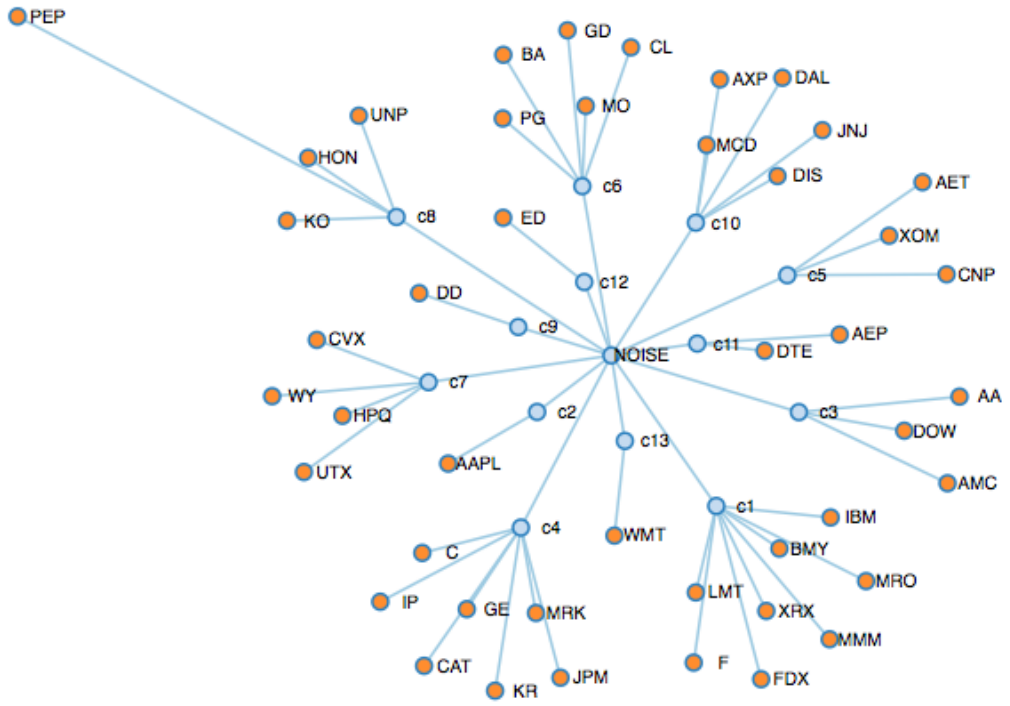
Signal Clusters

Both Trend and Noise spaces form clusters, **K-Medoids** algorithm, with the Correlation Distance Metric. Note that the Trend clusters differently than Noise!

Trend Clusters
Note PEP BA F and JNJ are in the same cluster 'c5'.
Source: http://thorek01.lossofgenerality.com/visualize/cluster_collapsible_mTRUST2.html



Noise Clusters
Note PEP KO and HON belong the same cluster 'c8'.
Source: http://thorek01.lossofgenerality.com/visualize/cluster_collapsible_mTRUST2NOISE2.html



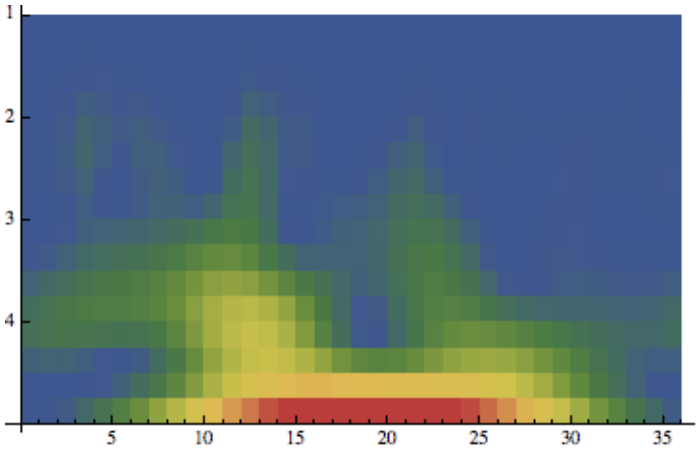
Sub-Band Noise

The closing-price was decomposed into 2 signals Trend and Noise. But while Trend, in this case, is quite atomic and indivisible, the Noise could be further decomposed into **Sub-Bands**:

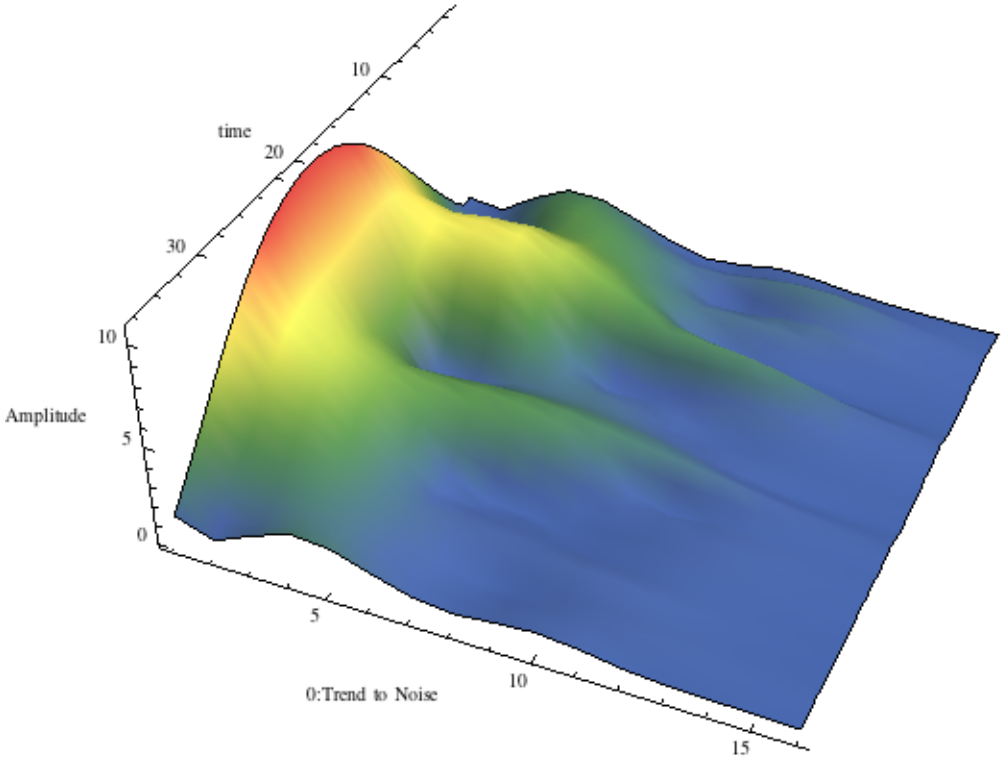
1. Sub-Bands are indexed from 0 to 16, 0 is Trend and 16 highest frequency noise
2. Red indicates highest amplitude for that frequency range, Yellow lesser amplitude, Green even lesser and Blue the least
3. **2D Scalogram**'s vertical axis is frequency ranges (octaves) with the lowest being the Trend around 0. The horizontal axis is time or periods of trading
4. **3D Scalogram** has one more z-axis for Amplitude to visualize the sub-bands in 3D

Remark 4: *Image the vertical axis as keyboard on a piano, 0 is the leftmost lowest frequency key, 1 is the rightmost the highest frequency key.*

2D Scalogram



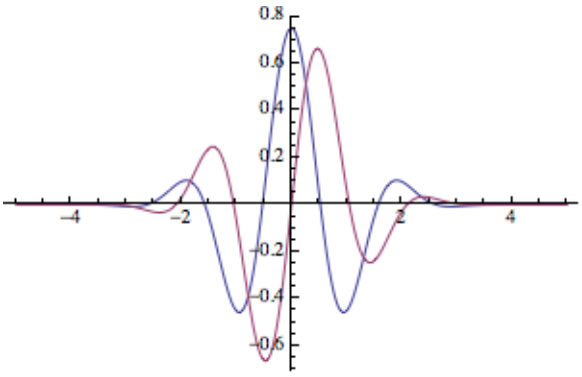
3D Scalogram



Based upon the above Scalograms:

1. PEP signal around 20th trading day i.e. Jan 20th, 5 days before and after, was mostly Trend with little Noise
2. Jan 1 to 12 sub-band noises in 3rd and 4th octaves appeared
3. Jan 23rd shows another peak in sub-band noise in 2nd and 3rd octaves

Gabor Wavelet of frequency 3 was chosen to compute the Continuous Transforms above. Blue curve is the Real and Red the Imaginary part of the wavelet.



Knn Regression

5 Nearest Neighbors are computed from the portfolio for the day before the forecasting date, once for Trend and once for Noise. The list includes the stock itself.

Trend: The last two prices of each Nearest Neighbors are linearly interpolated to predict the next day result. Then the average for the slope of the linear interpolation is found and applied to the forecast stock.

Noise: Average for the absolute value of Noise for each Nearest Neighbor is computed. Again the average taken and that is reported as the forecast Noise for the next day.

Knn Forecast for Feb 27 2014

Two closing-price forecasts computed i.e. Knn Regression applied to both Trend and Noise:

Trend = $80.0206
Noise = ±$0.6447

Post-Forecast: *Actual closing-price on Fri Feb 27 was $79.07.*

Knn Forecast for Feb 28 2014

Two closing-price forecasts computed i.e. Knn Regression applied to both Trend and Noise:

Trend = $80.6522
Noise = ±$0.63591

Post-Forecast: *Actual closing-price on Fri Feb 27 was $80.07.*

Trend curve for all Nearest Neighbours indicate upwards motility and as such the PEP stock moved to slightly higher prices, just about the Noise amount.