Darwin Δ Anomalies Forecast: SVR

dara@lossofgenerality.com

Aug 2014

Data

http://www.cgd.ucar.edu/cas/catalog/climind/tahiti.anom.ascii

http://www.cgd.ucar.edu/cas/catalog/climind/darwin.anom.ascii

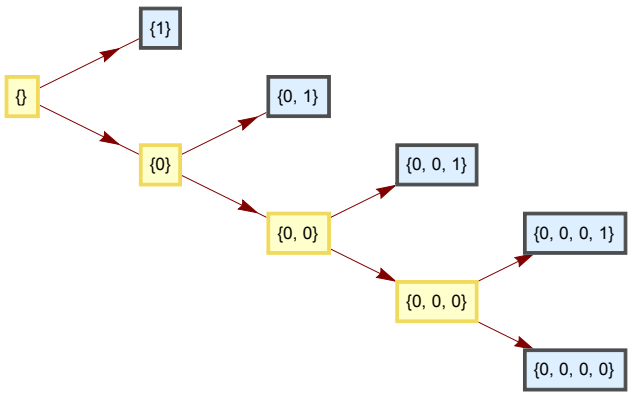
Darwin

1.,-0.1,0.9,0.3,-0.2,0.8,0.3,0.4,-0.1,0.,0.,0.5
...
-0.2,0.8,-1.1,1.3,0.4,1.,0.,1.4,0.3,0.2,0.6,0.4

Wavelet Decomposition

DaubechiesWavelet[4], 4, Padding→"Extrapolated"

Refinement Index



Refinement Index Algebra

{} = Raw Signal

{} = {0} + {1}

{0} = {0, 0} + {0, 1}

{0, 0} = {0, 0, 0} + {0, 0, 1}

{0, 0, 0} = {0, 0, 0, 0} + {0, 0, 0, 1}

Δ

This symbol indicates the signal is time-shifted and subtracted from itself i.e.

Signal(t + 1) – Signal(t)

So “Darwin Δ” means use the Darwin signal and make the time-shifted difference above.

Darwin Energy Fractions

Darwin’s raw {} signal has multiple trends and their energy fractions (contributions to the original signal) is spread out:

```
{{1} → 0.226711, {0, 1} → 0.120706, {0, 0, 1} → 0.105509,
 {0, 0, 0, 1} → 0.152306, {0, 0, 0, 0} → 0.394767}
```

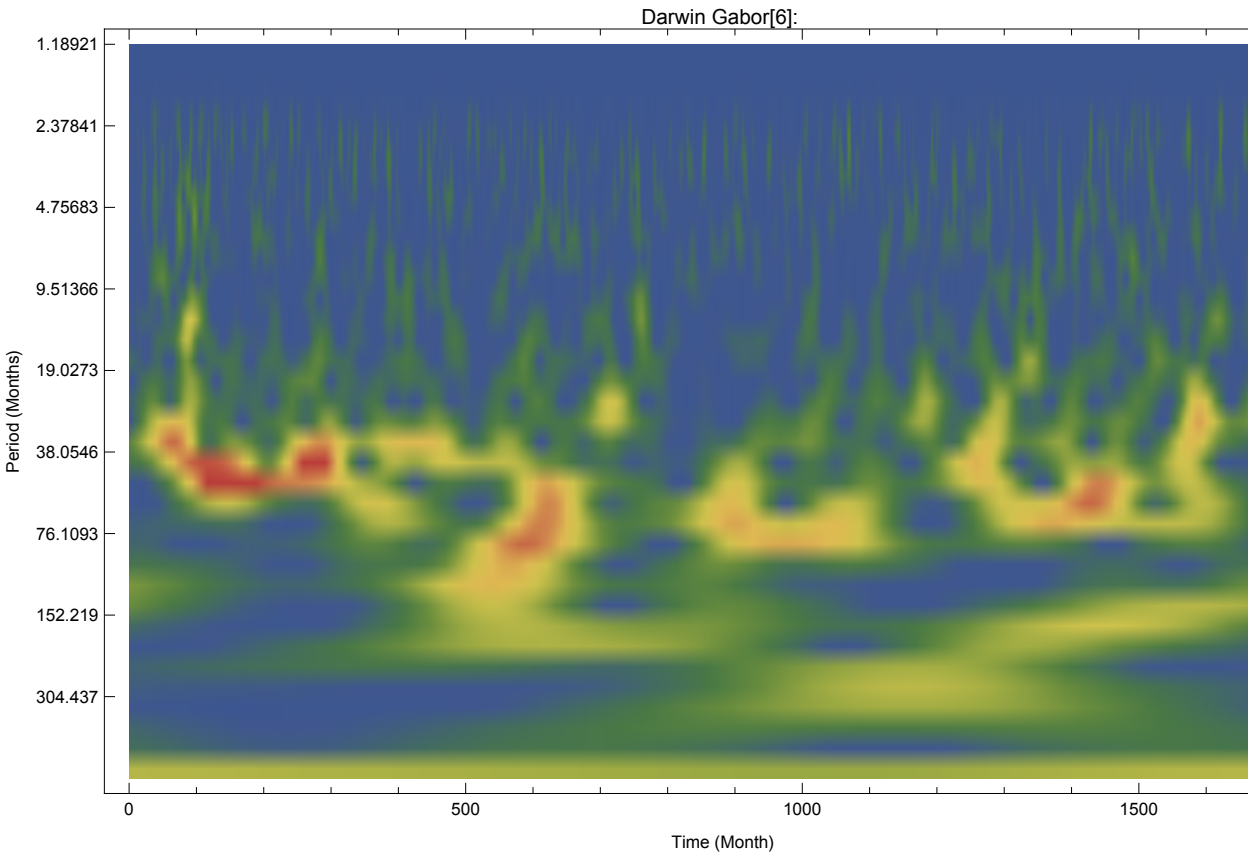
Darwin Δ Energy Fractions

The energy fractions for Darwin Δ change distribution and mostly contributed by {1}:

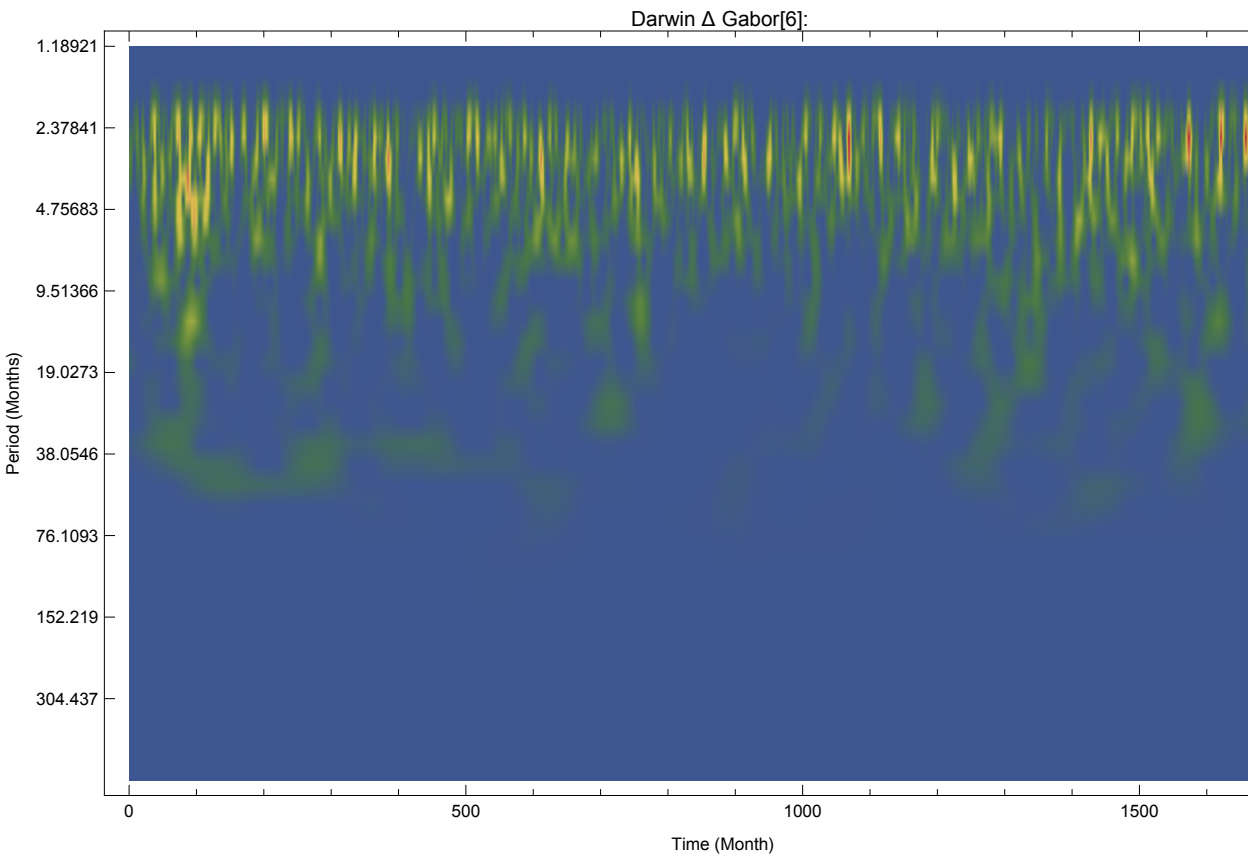
```
{{1} → 0.724091, {0, 1} → 0.17549, {0, 0, 1} → 0.0364491,
 {0, 0, 0, 1} → 0.0120768, {0, 0, 0, 0} → 0.0518935}
```

Scalogram

This is the raw {} Darwin signal, notice most amplitudes are in the longer periods:



Darwin Δ's scalogram shows concentration around 3 period (months):



SVR Forecast for Sign of Darwin Δ

Training Set: A training set of 12+1 past sample data was used to predict the sign of the Darwin Δ i.e. Darwin's Up vs. Down forecast. Algorithm uses the value of this month's Darwin Δ + past 12 months and forecast value for the next month. "this month's Darwin Δ" is Darwin value this month minus Darwin value the previous month.

BackTest: 300 consecutive months ending in Oct 2012 were used, each as THIS MONTH and used the current value and 12 previous values to forecast NEXT MONTH's i.e. 300 times, moving from the past to Oct 2012. The SVR algorithm was computed on each of the 300 months and results were tested for accuracy.

Dual Forecast: On each month two SVR forecast were computed:

i) {1} signal used as the input
ii) {} raw signal used as the input

Statistics: Postmortem statistics are computed for each back testing:

i) μ_+ means the mean for when forecast Darwin Δ has + sign and the actual data showed + sign as well, same setup for μ_-
ii) $\mu_{-\{1\}\{1\}}$ means {1} signal was used as input and its forecast was compared to ACTUAL {1}

In order to get the forecasts produce meaningful values several alterations (fudge) to the parameters used:

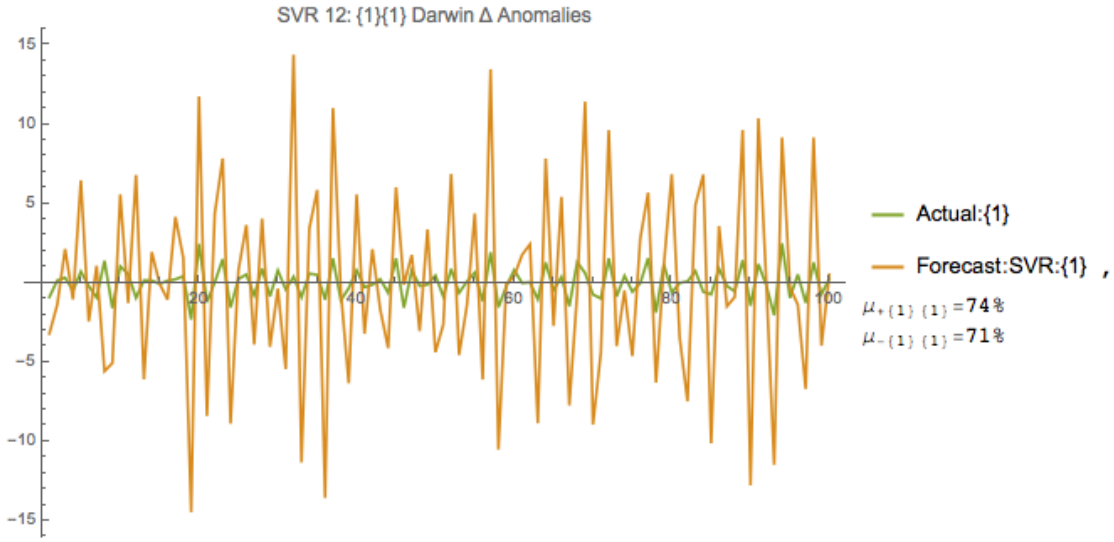
i) **FUDGE1**: the input signal was multiplied by 100 and then, the forecast signal was divided by 100. This was done since SVR needs to have more fluctuating data in order to produce good forecast for the Darwin Δ sign.

ii) **FUDGE2**: There is a drift in the SVR adaptive learning , like a cat chasing a mouse. That drift needs to be used to get a good forecast for the Darwin Δ sign

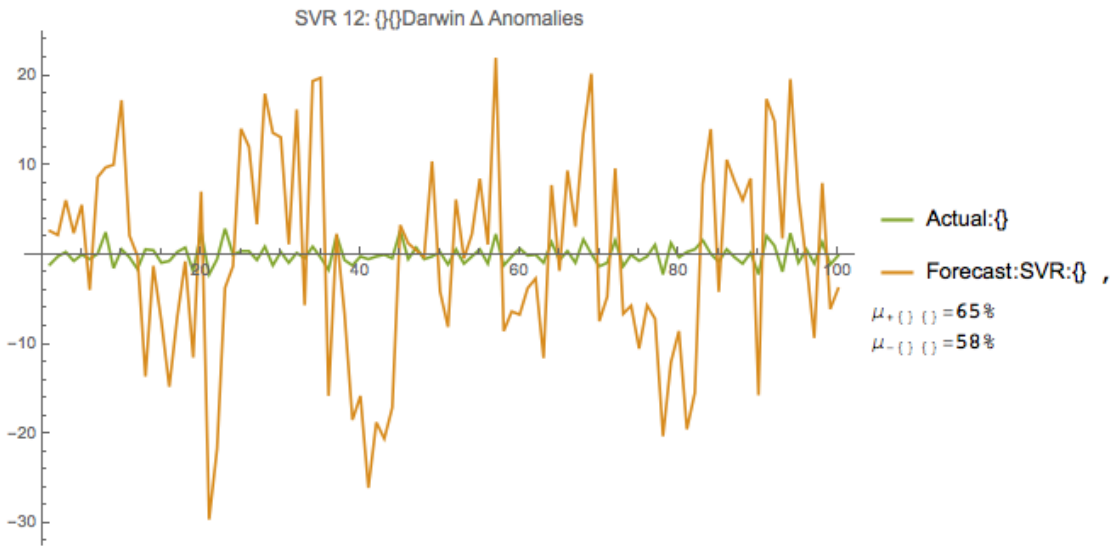
Drift: The Darwin Δ seems to have the form $f(x)\,\mathrm{Sin}(ax)$ with slow varying f, the period is 3 months per scalograms. Therefore time-shift of 1.5 days flips the sign for Darwin Δ. The applied drift of 1-1.5 days i.e. the sigh of forecast value of Darwin Δ and comparison to actual value produces significant accuracy 60% and more.

Remark 1: *One could apply -1 to -1.5 months shift and similar accuracies found again. This could mean the Darwin Δ has a time-reversible component to its signal decomposition that contributes with large amplitude (Energy Fraction).*

{1} Trend forecast comparison to itself (Actual):



Raw signal {} forecast comparison to itself (Actual):



It seems that the Forecast {1} could increase the accuracy to forecast raw signal {} i.e. without Wavelet transforms.

{1} forecast comparison to raw signal {}: