Predicting the future? Comparative Forecasting and a test for Persistence in the El Nino Southern Oscillation ... and a few new ideas

Dan Zachary
May 27, 2014

NOAA SEMINAR #2 - NCWCP Conference Center College Park
Brief Introduction

The Bayesian Binary Tree (BBT) model
A path expectation model
Forecasting
Conclusions
Belinda Chiera (University of South Australia)

Jerzy Filar (Flinders University)

Adrian H. Gordon (University of South Australia) – deceased

Daniel Zachary (Johns Hopkins, Tudor)
The Motivation

- **El Nino Southern Oscillation (ENSO):**
  
  “well known source of interannual climatic variability with far-reaching flow-on effects”

- **Economic Impact:** agriculture, commercial fishing, construction & tourism.

- **Environmental Impact:** drought and flooding.

- **Human Impact:** loss of life and livelihood.

**ENSO forecasting now an important part of policy making.**

**Tradeoff?** complex models at the cost of manageability? Does this really translate into forecast quality?

**Alternative?** a simple, probabilistic (Bayesian) forecast that exploits *persistence* in ENSO.
**SOI: The Southern Oscillation Index**

\[
SOI = 10 \cdot \frac{P_d - \bar{P}_d}{\sigma_{P_d}}
\]

- \(P_d\): difference between average Tahiti MSLP & Darwin MSLP
- \(\bar{P}_d\): long-term average of \(P_d\).
- \(\sigma_{P_d}\): long-term standard deviation of \(P_d\).

La Nina

El Nino
SOI Probability distribution vs a Binomial distribution (p = 0.5)

Negative Valued Events

Positive Valued Events

Months

Months
**SOI : Digitised**

Digitised SOI =

\[+1 \text{ if } SOI \geq 0: \quad \text{Up event}\]

\[-1 \text{ if } SOI < 0: \quad \text{Down event}\]

Blue/White bands signal **persistence** in the series.
Up Events:

\[ U_1 = \text{Number of Up episodes of length 1} \]
\[ U_2 = \text{Number of Up episodes of length 2} \]
\[ \vdots \]
\[ U_u = \text{Number of Up episodes of length } u \]
\[ U_T = \sum_{i=1}^{u} U_i \]

Down Events:

\[ D_1 = \text{Number of Down episodes of length 1} \]
\[ D_2 = \text{Number of Down episodes of length 2} \]
\[ \vdots \]
\[ D_d = \text{Number of Down episodes of length } d \]
\[ D_T = \sum_{i=1}^{d} D_i \]
The BBT model

\[
\begin{align*}
\Pr(\text{observing a } (j+2)\text{nd Up event} | \ j+1 \text{ consecutive Up events}) \\
\Pr(\text{observing a } (j+1)\text{st Up event} | \ j \text{ consecutive Up events}) \\
\Pr(\text{not observing a } (j+1)\text{st Up event} | \ j \text{ consecutive Up events}) \\
\Pr(\text{observing a } 2\text{nd Down event} | \ 1 \text{ Down event})
\end{align*}
\]
The Bayesian Binary Tree Model (BBT)
So how do we use a probability model of +1 and -1 events to reconstruct and forecast the SOI?

The BBT model

\[ \hat{\alpha}_j = \frac{\sum_{i=j+1}^{u} \mathcal{U}_i}{\sum_{i=j}^{u} \mathcal{U}_i} \]

\[ \hat{\beta}_j = \frac{\sum_{i=j+1}^{d} \mathcal{D}_i}{\sum_{i=j}^{d} \mathcal{D}_i} \]
The BBT model

- Use the Up/Down episode histories:

  \[ \overline{x}_{m,k} = \text{average SOI value for month } k, \text{ in an Up cluster of length } m \]

  \[ \overline{y}_{n,k} = \text{average SOI value for month } k, \text{ in a Down cluster of length } n \]

- Take into account we may enter a cluster part-way through:

  \[ \mathcal{N}_j = \sum_{k>j} \mathcal{N}_k \text{ and } \mathcal{D}_j = \sum_{k>j} \mathcal{D}_k \]
Forecasting – Part 1/3

- Use the Up/Down episode histories:

\[
x_{m,k} = \text{average SOI value for month } k, \text{ in an } Up \text{ cluster of length } m
\]

\[
y_{n,k} = \text{average SOI value for month } k, \text{ in a } Down \text{ cluster of length } n
\]

- Estimate the positively valued SOI:

\[
\hat{x}_{j+1} = \sum_{k=1}^{u_j} \frac{u_{j+k}}{u_j} \bar{x}_{j+k,j+1}
\]

- Estimate the negatively valued SOI:

\[
\hat{y}_{j+1} = \sum_{k=1}^{d_j} \frac{d_{j+k}}{d_j} \bar{y}_{j+k,j+1}
\]
Construct Forecast Estimators:

\[
\hat{z}(t+k) = E[SOI(t+k)] \\
\hat{z}_U(t+k) = E[SOI(t+k) \mid U] \\
\hat{z}_D(t+k) = E[SOI(t+k) \mid D]
\]

Condition on “Up”:

\[
\hat{\chi}_j \\
\hat{z}_U(t+k) = \sum_{\sigma \in \hat{A}_k^+} \gamma^+(\sigma) \hat{x}_{t+k}(\sigma)
\]

\[
P(\sigma) = \prod_{r=1}^{k} P(\sigma_r)
\]

\[
\sigma = (\sigma_1, \ldots, \sigma_k)
\]

\[
\hat{x}_{t+k}(\sigma) = \hat{z}_U(t+k) + \sum_{\sigma \in \hat{A}_k^+} \gamma^+(\sigma) \hat{x}_{t+k}(\sigma)
\]

\[
P(A_k^+) = \frac{P(\sigma)}{P(A_k^+)}
\]
**Forecasting – Part 3/3**

- **Construct Forecast Estimators:**
  \[
  \hat{z}(t+k) = E[SOI(t+k)]
  
  \hat{z}_U(t+k) = E[SOI(t+k) | U]
  
  \hat{z}_D(t+k) = E[SOI(t+k) | D]
  \]

- **Conditionalise on “Down”:**
  \[
  A_k^- = \{\sigma | \sigma_k = -1\}
  
  P(A_k^-) = \sum_{\sigma \in A_k^-} P(\sigma) \quad \text{and} \quad \gamma^-(\sigma) = \frac{P(\sigma)}{P(A_k^-)}
  
  z_D(t+k) = \sum_{\sigma \in A_k^-} \gamma^-(\sigma) \hat{y}_{t+k}(\sigma)
  \]
BBT model representation – up scenario
Model representation – down scenario

\[ z(t+k) \]

\[ z_{\mathcal{D}}(t+k) \]
Model representation – mixed scenario (2010)
A path expectation model
Building on the expectation of each level

Find the ensemble of expectation values for each ‘level’ of the SOI

Inspired by the concept of most probable path of particles developed in quantum theory – finding the most probable path from A to B.
An alternative strategy

Find the ensemble of expectation values for each ‘level’ of the SOI

\[ X(t+1) = \text{EXP}(X(t+1) \]
\[ X(t+2) = \text{EXP}(X(t+2) \]
\[ X(t+2) = \text{EXP}(X(t+3) \]
\[ . \]
\[ . \]
\[ . \]
PE representation and comparisons – up scenario

Path Expectation model
PE representation and comparisons – down scenario

Path Expectation model

$z(t+k)$

$z_\sigma(t+k)$
PE representation and comparisons – mixed scenario
Mean differences - overview

Projection months
Forecasting
SOI Forecast – compared with IRI/CPC SST forecast

Figure provided by the International Research Institute (IRI) for Climate and Society (updated 13 May 2014).

April 2014
Conclusions
Summary

- The BBT models outperforms the TSF, Random model, and better adapts to the changing trend of the SOI.

- **Worth noting:** BBT (Up/Down) typically do better forecast in periods of sustained Up/Down. BBT (unconditional) does better in mixed periods.

- Methods do reasonably good forecasts for several months beyond the forecast point, although there is generally considered a breakdown in predictability power after this elapse of time.

- **Short-term aim:** achieved. Simple schemes particularly useful in economic SOI-based forecasting (Austria - S. America research fund).

- **Long-term aim:** To develop a generalized the statistical method (e.g. Path Expectation model) and improve the quality of the forecasting tool & for a more realistic (e.g. tri-state) ENSO process, as described by the SOI.
Thank you


Zachary D.S. and Dobsen, S. Does urban space evolve deterministically or entropically? An exploration of urban development models for Sheffield, UK in relation to decentralized energy policies, Energy Policy (In review), 2013


A new strategy
A comparison with NCEP/CPC Markov model

Time evolution of observed and predicted SST anomalies in the Nino 3.4 region (up to 12 lead months) by the NCEP/CPC Markov model (Xue et al. 2000, *J. Climate*, 13, 849-871).
A comparison with NCEP/CPC Markov model
TUDOR Modeling Group – Research snapshot

Integrated assessment: Energy-AQ

- Bayesian risk (atmospheric forecasting)
- Geospatial analysis
- Energy projections

Integrated assessment: Impact

Energy model: ETEM Luxembourg
Optimization: OBOE
Air quality model: AUSTAL2000-AYLTP