A forecasting method using a wavelet-based mode decomposition and application to the ENSO index

Adrien DELIÈGE

University of Liège, Belgium

ISF 2015 - Frontiers in Forecasting
Riverside, California – June 24, 2015

Joint work with S. NICOLAY and X. FETTWEIS

adrien.deliege@ulg.ac.be

- Decomposing time series into several modes has become more and more popular and useful in signal analysis.
- Methods such as EMD or SSA (among others) have been successfully applied in medicine, finance, climatology, ...
- Old but gold: Fourier transform allows to decompose a signal as

$$f(t) \approx \sum_{k=1}^{K} c_k \cos(\omega_k t + \phi_k).$$

- Problem: often too many components in the decomposition.
- Idea: Considering the amplitudes as functions of t to decrease the number of terms.
- Development of a wavelet-based decomposition method which is then used as a forecasting method.

Table of contents

Table of contents

Wavelet transform and spectrum

- The wavelet used in this study is the function

$$\psi(t) = \frac{\exp(i\Omega t)}{2\sqrt{2\pi}} \exp\left(-\frac{(2\Omega t+\pi)^2}{8\Omega^2}\right)\left(\exp\left(\frac{\pi t}{\Omega}\right)+1\right),$$

with $\Omega = \pi\sqrt{2/\ln 2}$, which is similar to the Morlet wavelet ([3]).

- The wavelet transform of the signal is computed as:

$$Wf(a,t) = \int f(x)\bar{\psi}\left(\frac{x-t}{a}\right)\frac{dx}{a},$$

where $\bar{\psi}$ is the complex conjugate of ψ, $t \in \mathbb{R}$ stands for the location/time parameter and $a > 0$ denotes the scale parameter.

- The wavelet spectrum is computed as:

$$\Lambda(a) = E\,|Wf(a,.)|$$

where E denotes the mean over time.

Reconstruction and forecast

- We look for the scales $a_1, \ldots, a_J$ for which the wavelet spectrum Λ reaches a maximum.
- An accurate reconstruction of f is given by

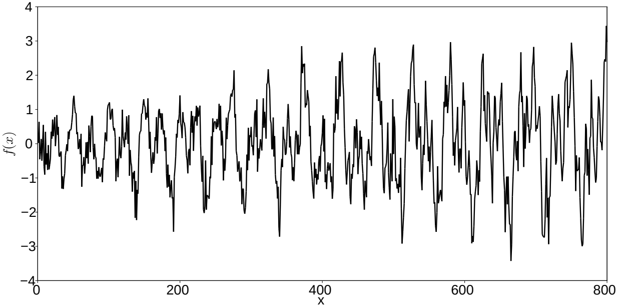
$$f(t) \approx \sum_{j=1}^{J} |Wf(a_j, t)| \cos(\arg Wf(a_j, t)).$$

- Since $\cos(\arg Wf(a_j, t))$ roughly corresponds to a cosine with a period proportional to a_j, forecasts of the reconstructed signal can be obtained with smooth extrapolations (using Lagrange polynomials) of the amplitudes $|Wf(a_j, t)|$.

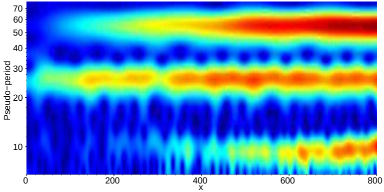
Example

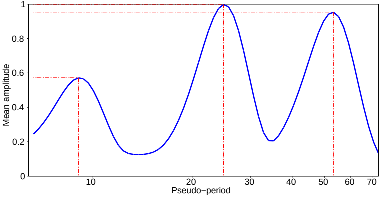
We consider $f(x) = \sum_{i=0}^{3} f_i(x)$ where

- f_0 is a Gaussian (white) noise with mean 0 and standard deviation 0.4,
- $f_1(x) = \frac{x}{700} \cos\left(\frac{2\pi}{\omega(x)} x\right)$ with $\omega(x) = 10 + 0.5 \cos\left(\frac{2\pi}{1600} x\right)$,
- $f_2(x) = \frac{\ln(x)}{6} \cos\left(\frac{2\pi}{25} x\right)$,
- $f_3(x) = \frac{\sqrt{x}}{20} \cos\left(\frac{2\pi}{53} x\right)$.



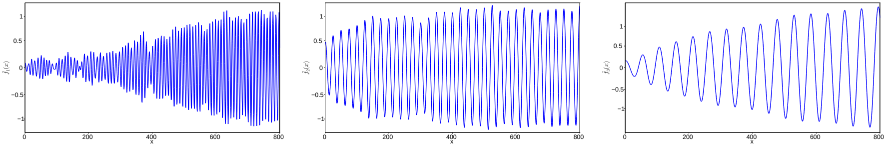
Example





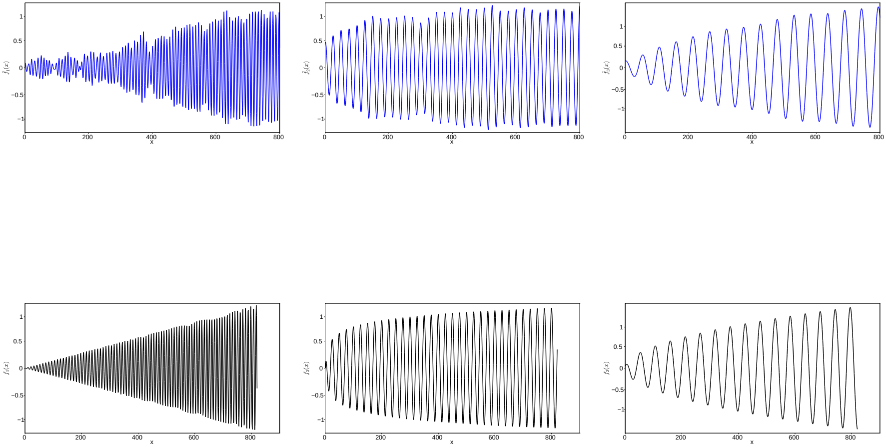
Top: modulus of the wavelet transform of f. Values range from 0 (dark blue) to 1.2 (dark red). Bottom: the associated wavelet spectrum. The three periods are clearly detected.

Example



First row: the reconstructed components $\hat{f}_1$, $\hat{f}_2$ and $\hat{f}_3$.

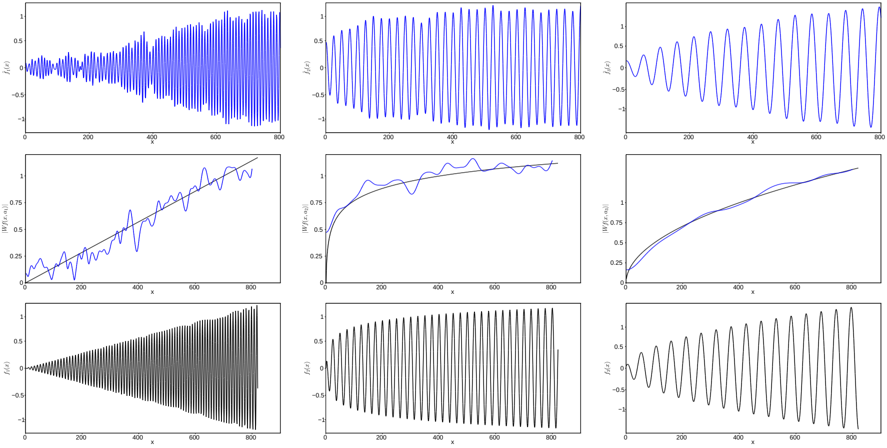
Example



First row: the reconstructed components $\hat{f}_1$, $\hat{f}_2$ and $\hat{f}_3$.

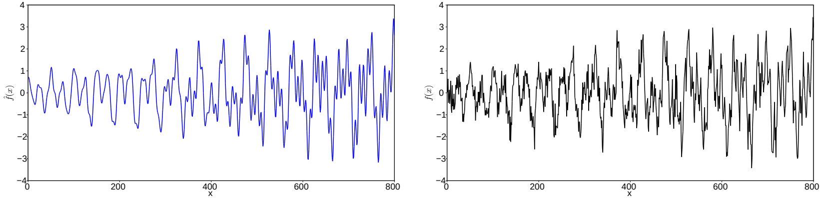
Third row: the original components.

Example



First row: the reconstructed components $\hat{f}_1$, $\hat{f}_2$ and $\hat{f}_3$. Second row: amplitudes of the components $\hat{f}_i$ (blue) and f_i (black). Third row: the original components.

Example

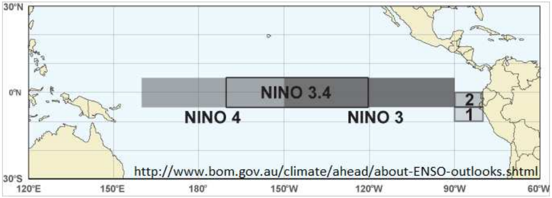


Left: the reconstructed signal. Right: the original signal. The correlation between them is 0.95 and RMSE is 0.37.

Table of contents

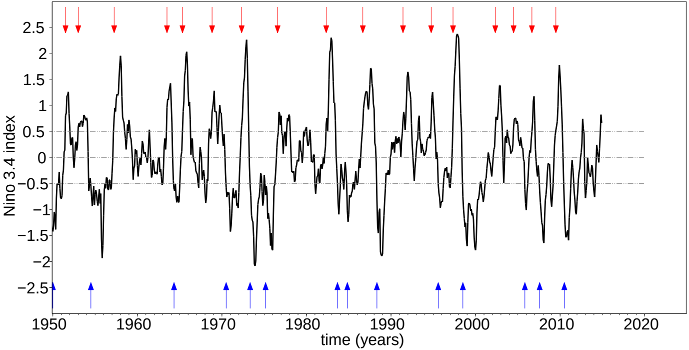
Analyzed data

- Analyzed data: Niño 3.4 time series, i.e. monthly-sampled sea surface temperature anomalies in the Equatorial Pacific Ocean from Jan 1950 to Dec 2014 (`http://www.cpc.ncep.noaa.gov/`).



El Niño/La Niña events

- Niño 3.4 index:



- 17 El Niño events: SST anomaly above $+0.5^{\circ}$C during 5 consecutive months.
- 14 La Niña events: SST anomaly below -0.5°C during 5 consecutive months.

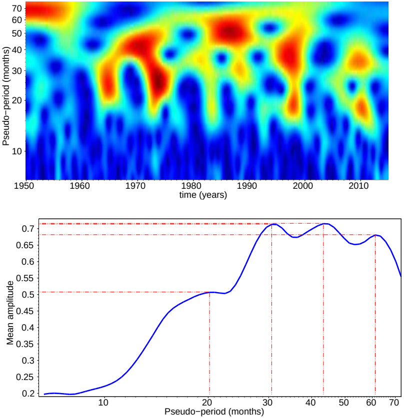
Teleconnections

- Flooding in the West coast of South America
- Droughts in Asia and Australia
- Fish kills or shifts in locations and types of fish, having economic impacts in Peru and Chile
- Impact on snowfalls and monsoons, drier/hotter/wetter/cooler than normal conditions
- Impact on hurricanes/typhoons occurrences
- Links with famines, increase in mosquito-borne diseases (malaria, dengue, ...), civil conflicts
- In Los Angeles, increase in the number of some species of mosquitoes (in 1997 notably, see [5]).
- ...

$\longrightarrow$ Importance of predicting El Niño/La Niña events.

Table of contents

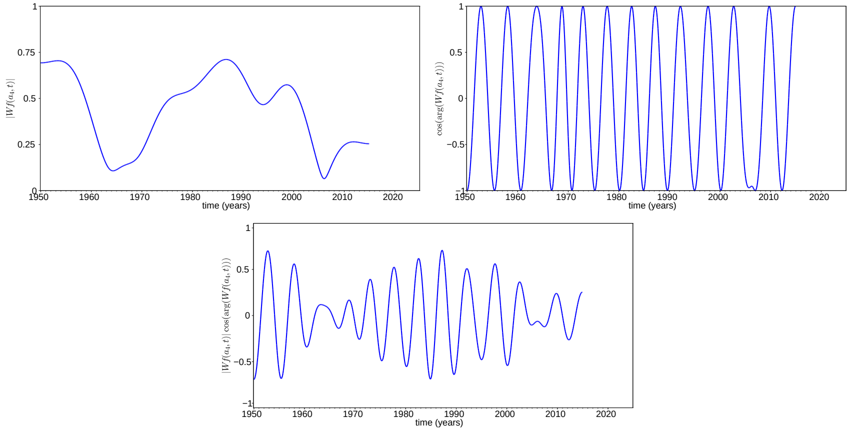
Wavelet transform and spectrum



Top: Modulus of the wavelet transform of the signal (values range from 0 (dark blue) to 1.2 (dark red)). Bottom: Associated wavelet spectrum. Four peaks are detected, corresponding to periods of about 21, 31, 43 and 61 months.

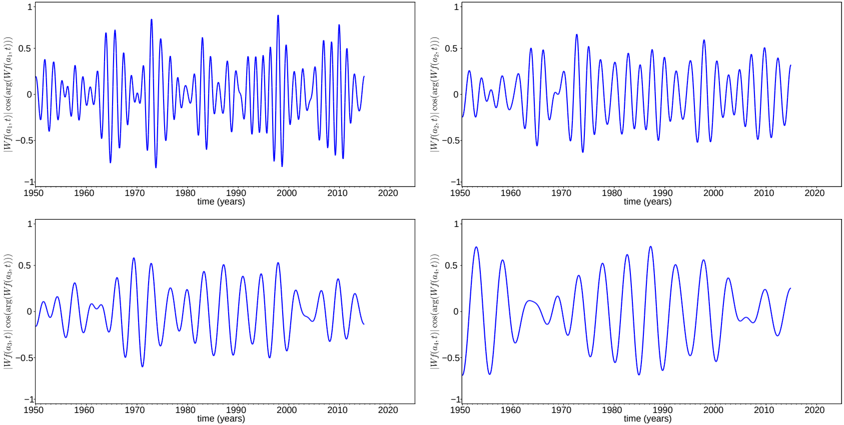
Extraction of the components

Example: extraction of the 4th component (61 months-period).



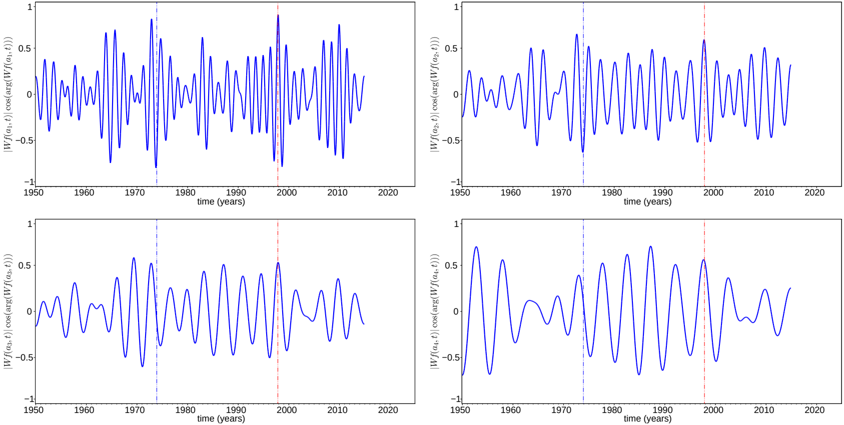
Top: amplitude $|Wf(a_4,t)|$ (left) and oscillatory part $\cos(\arg Wf(a_4,t))$ (right).
Bottom: the 4th component $|Wf(a_4,t)|\cos(\arg Wf(a_4,t))$.

Extraction of the components



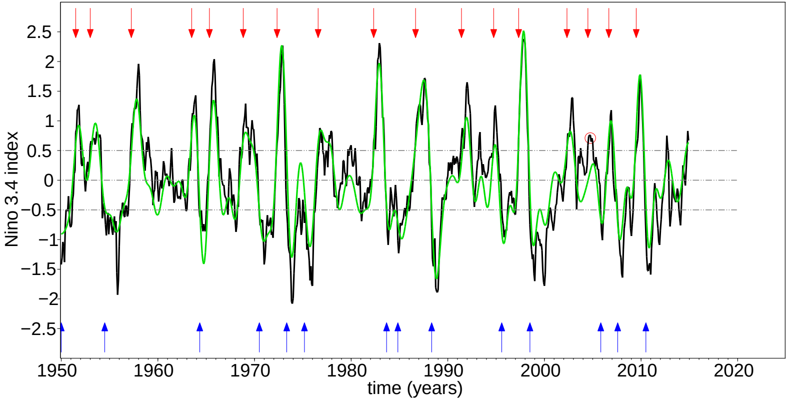
From top to bottom, from left to right: components extracted associated to periods of 21, 31, 43, 61 months.

Extraction of the components



Red: occurrence of the strongest El Niño event. Blue: occurrence of the strongest La Niña event.

Reconstruction of the Niño 3.4 index



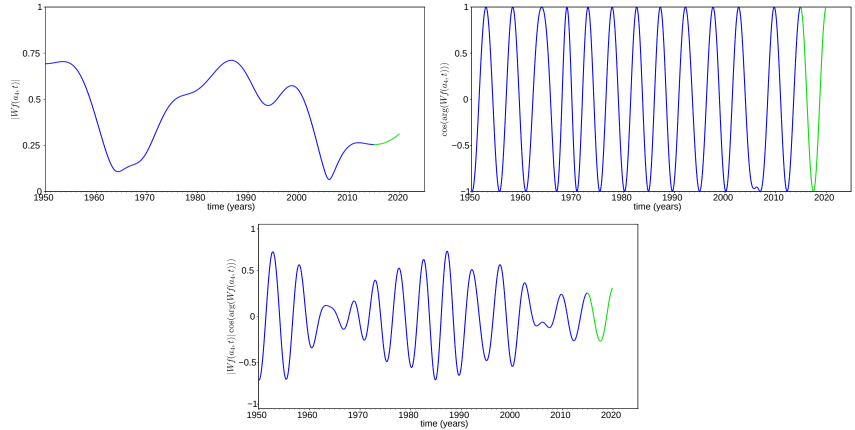
- RMSE $= 0.366^{\circ}$C and correlation $= 0.894$.
- 30/31 (96.7%) of El Niño/La Niña events recovered.

What's the name of the conference again?

Ok, let's forecast now!

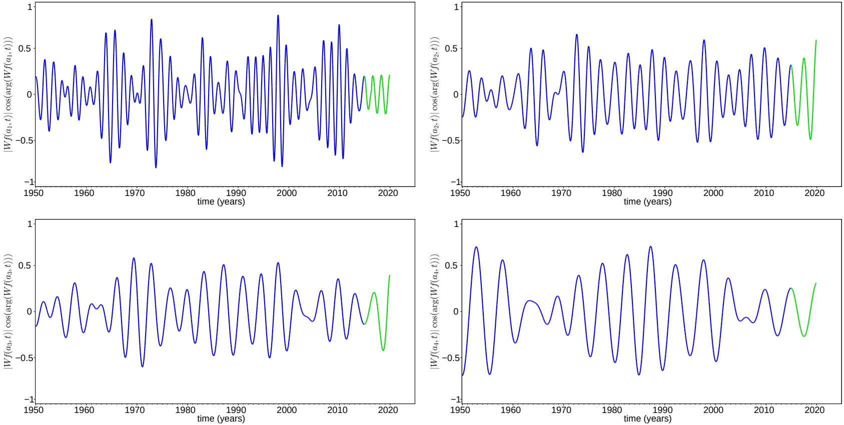
Extrapolation of the components

Example: forecast of the 4th component (61 months-period).



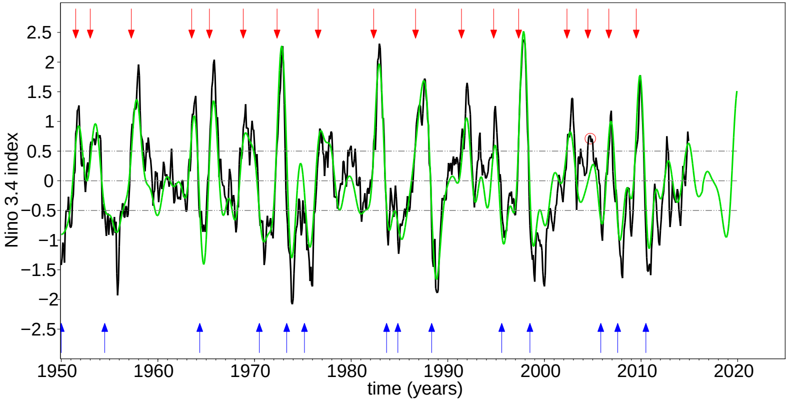
Top: amplitude (left) and oscillatory part (right). Bottom: the 4th component. Green parts: forecasts.

Extrapolation of the components



From top to bottom, from left to right: components extracted associated to periods of 21, 31, 43, 61 months. Green parts: forecasts.

Forecast of the Niño 3.4 index



The next La Niña event should start early in 2018 and should be followed soon after by a strong El Niño event in the second semester of 2019.

Border effects

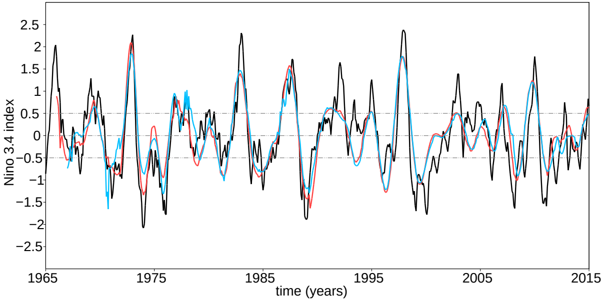
- The last data of the amplitudes of the components are flawed by border effects.
- The oscillating parts are barely affected, thus the possibility that they all reach a maximum around 2018-2019 still holds.
- If we manage to correct the border effects, this forecasting method is efficient for mid-term predictions, as proved by retroactive predictions.

Hindcasts (cheated, without border effects to test the method)

Example: 12 months hindcast starting in 1965.

1) Cut the wavelet transform of the time series at time point t_0 (e.g. $t_0 =$ Dec 1964).
2) Compute the wavelet spectrum and perform the decomposition and reconstruction of this signal.
3) Based on this particular decomposition and reconstruction, make a forecast (at least 12 months).
4) The 12th value of the forecast is the first value of the hindcast.
5) The initial condition t_0 becomes $t_0 + 1$ and steps $1 - 4$ are repeated (the 12th value of the second forecast is the second value of the hindcast, etc.).
6) Each value of the hindcast is predicted 12 months in advance, and no data past the date the value is issued is used.

Examples of hindcasts (12 and 24 months)

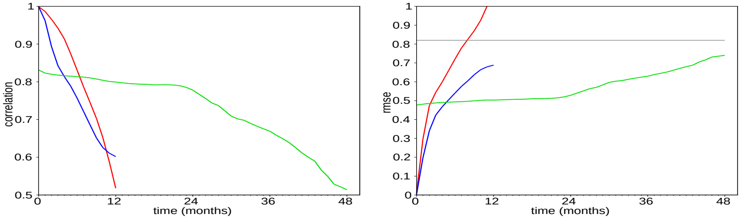


Red (resp. blue): 12 (resp. 24) months hindcast starting in 1965.

Hindcasts: a few numbers

Proportion of El Niño/La Niña events accurately predicted and erroneously predicted (false positive) when using hindcasts (from 1965).

signal	predicted	false positive
12 months hindcast	92%	2
24 months hindcast	87%	2
36 months hindcast	78%	2
48 months hindcast	74%	3



Correlation/RMSE between the *t*-months hindcast and the signal as functions of the lead time *t* (green) and comparison with those of models from [2] (blue) and [7] (red).

Table of contents

Conclusions

- The WMD allows to decompose the Niño 3.4 index into four pseudo-periodic components (21 ,31, 43, 61 months).
- The reconstruction recovers 30/31 El Niño/La Niña events.
- The components can be extrapolated to make a several years forecast (partially affected by border effects).
- If we do not take border effects into account, most of the major events can be predicted several years in advance.
- Our method resolves the large variations of the signal and is particularly competitive for mid-term predictions.

Conclusions

Ideas for future work...

- Developing a method to correct or limit border effects.
- Re-computing the forecast and performing proper cross-validations.
- Understanding the underlying mechanisms governing ENSO variability (i.e. the origin of the “pseudo-periodicities” detected in Niño 3.4).
- Checking if current models take these periods into account; if not, improving current models and forecasting procedures.
- Application to other climate indices (e.g. the North Atlantic Oscillation index where periods of ~ 30, ~ 40 and ~ 60 months have also been found [4]).
- ...

Thanks.

Some references

K. Ashok, S.K. Behera, S.A. Rao, H. Weng, and T. Yamagata (2007)
El Niño Modoki and its possible teleconnection.
J. Geophys. Res, 112(10.1029).

D. Chen, M.A. Cane, A. Kaplan, S.E. Zebiak, and D. Huang (2004)
Predictability of El Niño over the past 148 years.
Nature, 428(6984):733–736.

M.H. Glantz (2001)
Currents of Change: Impacts of El Niño and La Niña on climate society.
Cambridge University Press.

Y.G. Ham, J.S. Kug, and I.S. Kang (2009)
Optimal initial perturbations for El Niño ensemble prediction with ensemble Kalman filter.
Climate dynamics, 33(7):959–973.

D. Heft and W. Walton (2008)
Effects of the El Niño–Southern Oscillation (ENSO) cycle on mosquito populations in southern California.
J. Vector Ecol., 33(1):17–29.

T. Horii and K. Hanawa (2004)
A relationship between timing of El Niño onset and subsequent evolution.
Geophysical research letters, 31(6):L06304.

G. Mabille and S. Nicolay (2009)
Multi-year cycles observed in air temperature data and proxy series.

Some references

S. Mallat (1999)
A Wavelet Tour of Signal Processing.
Academic Press.

S.J. Mason and G.M. Mimmack (2002)
Comparison of some statistical methods of probabilistic forecasting of ENSO.
Journal of Climate, 15(1):8–29.

S. Nicolay (2011)
A wavelet-based mode decomposition.
Eur. Phys. J. B, 80:223–232.

S. Nicolay, G. Mabille, X. Fettweis, and M. Erpicum (2009)
30 and 43 months period cycles found in air temperature time series using the Morlet wavelet method.
Climate dynamics, 33(7):1117–1129.

W. Wang, S. Saha, H.L. Pan, S. Nadiga, and G. White (2005)
Simulation of ENSO in the new NCEP coupled forecast system model (CFS03).
Monthly Weather Review, 133(6):1574–1593.

S.W. Yeh, J.S. Kug, B. Dewitte, M.H. Kwon, B.P. Kirtman, and F.F. Jin (2009)
El Niño in a changing climate.
Nature, 461(7263):511–514.

J. Zhu, G.Q. Zhou, R.H. Zhang, and Z. Sun (2012)
Improving ENSO prediction in a hybrid coupled model with embedded temperature parameterisation.
International Journal of Climatology, online, 2012.