El Nino Return Rates

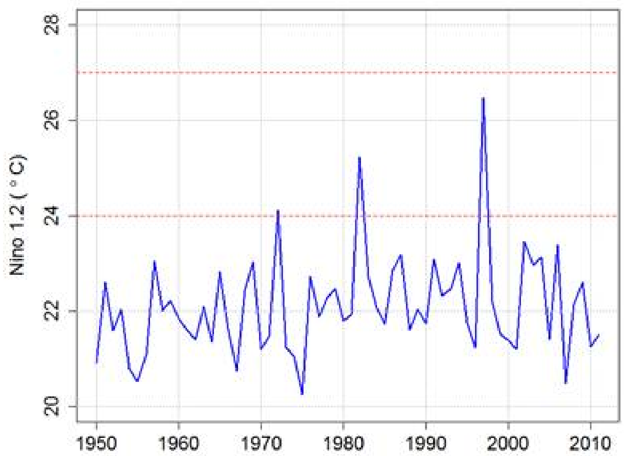
After post-processing some information, I can provide you some background and the related computer program (please see attachment) of the El Nino project. This project was initiated by a client (insurance company) located in Peru. El Nino conditions can affect severely the economy and infrastructure of the country. In 1997 an extreme El Nino event caused multiple problems. To bring some examples, anchovy production (an important economic activity in Peru) was reduced because elevated sea temperatures off the Peruvian coasts reduced spawning and forced the fish population to migrate elsewhere. Agricultural yields on land decreased dramatically as a result of extreme precipitation. Further consequences of extreme precipitation were floods and landslides that destroyed private property and public infrastructure such as roads and bridges. Damage in public infrastructure in turn affected transportation and electricity production/distribution leading to further economic losses. Financial institutions such as banks were also affected since the disruptions in economic activity forced many businesses and private borrowers to suspend loan payments.

As a result, insurance losses from various lines of business were particularly large. Our client asked us to quote on a suggested insurance product that would protect them from extreme El Nino events.

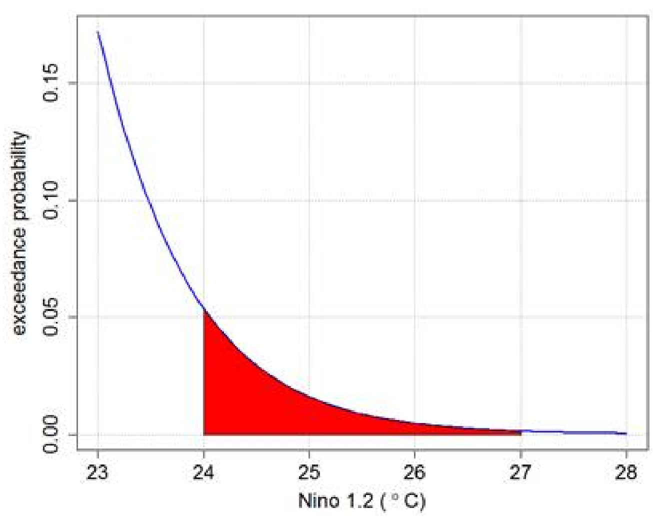
To quantify the severity of El Nino events, the Nino 1+2 index was chosen, since it is most relevant for the area of Peru. The deal was to provide to the client a linear payout that would begin with a certain Nino 1+2 value (entry point) and would reach a maximum at a higher Nino 1+2 value (exit point). The payout values used in the attached R code are fictitious but realistic.

Payout begins at zero for a Nino 1+2 value of 24 °C and increases linearly to the maximum payout of 50M USD for Nino 1+2 = 27 °C. The Nino 1+2 values refer to the November-December average. Most El Nino-related disruptions in Peru start around January each year but the associated elevated sea surface temperatures in the Pacific occur already in November-December. The November-December Nino 1+2 value is known beginning of January each year. This allows the client to receive payout already in January, when losses start occurring. The Nino 1+2 observations are obtained from NOAA as an official, agreed upon agency.

For the pricing of the contract, we need in principle to calculate how frequently payout occurs. We use monthly NOAA Nino 1+2 data since 1950 and we take the November-December average for each year. Data are shown in the plot below. The red dash lines highlight the entry and exit Nino 1+2 values:



A probability distribution function is then fitted to the data. The Generalized Extreme Value (GEV) distribution is chosen as a good choice when focusing on extreme values. The corresponding exceedance probability function (also known as complementary cumulative distribution function) is shown below:

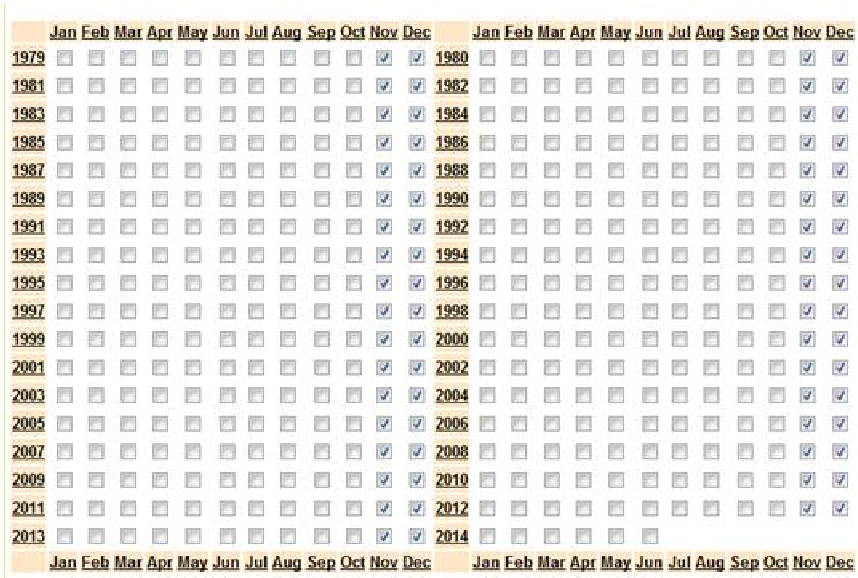


The theoretical exceedance probability for 24 °C is about 0.05, which corresponds to a 20 year return period. However, for the linear payoff structure of the contract in question, we need the average exceedance probability for the temperature range 24 – 27 °C. We obtain this by taking the integral highlighted in the plot and dividing by the corresponding temperature range (3 °C). The resulting annual expected frequency is about 0.014. The annual premium is then calculated by multiplying the payout by this frequency 50M USD * 0.014 = 0.7 M USD.

European Data

Download the data from the ECMWF site, and use the following choices in the data selection screen:

- only Nov-Dec values:



- all four times choices and the zero time-step:

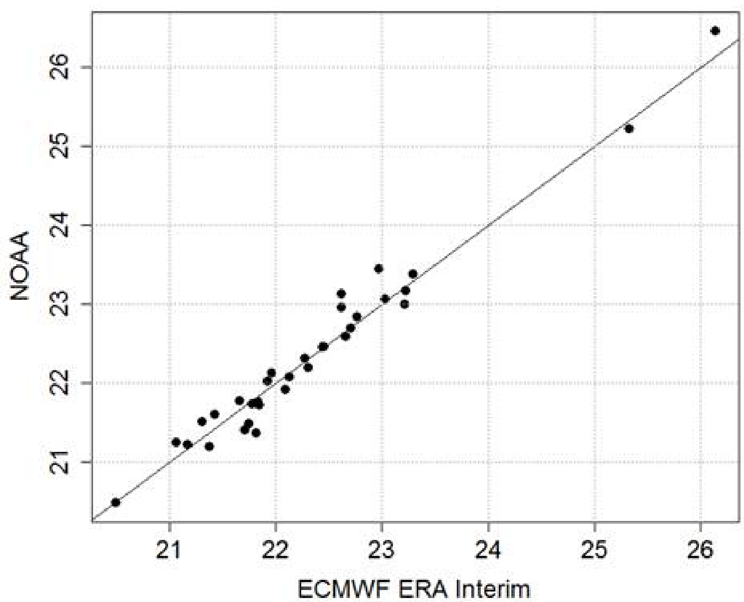
- the SST variable

The estimated return periods are a bit lower and the pricing result is correspondingly higher (~0.9 M USD) compared to the NOAA data. I have not investigated in detail why there are differences but the fact that the ECMWF time series is shorter (1979-2014 instead of 1950-2014) surely plays a role.

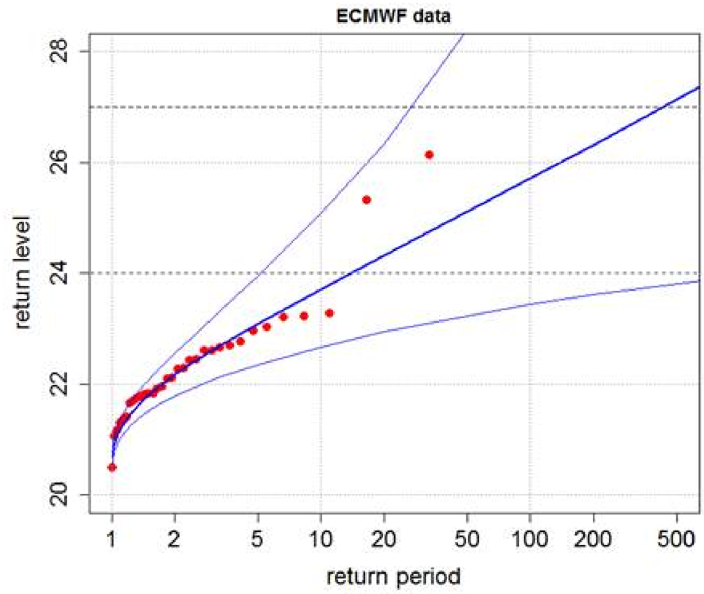
Explanation of Difference

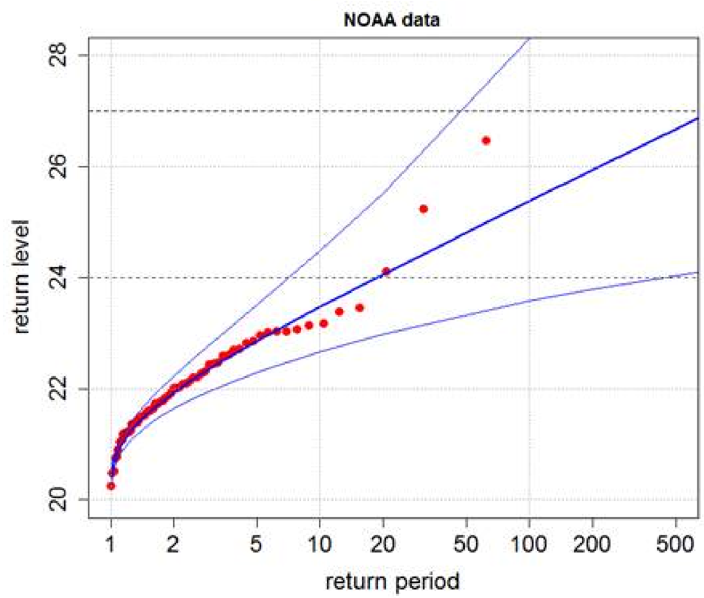
In general a theoretical return period for extreme events has large uncertainty and it is very sensitive to the data being used. If you take the maximum of, say, 50 years of data, your best guess is that the return period is 50 years (empirical return period). However, you cannot really tell if it is a 100-year event that by chance occurred in the last 50 years. It could also be that a 50-year event has not occurred in the investigated period and that the observed maximum is a 40-year event. On the other hand, return periods of moderate/usual events can be estimated accurately since there are enough observations. If we now fit a probability distribution function to the data, we can extrapolate for even more severe events, but the uncertainty becomes even larger. In addition, if we look at extreme events, the sensitivity of the return period to the data becomes larger. If we take for example a quantity measured each year and following the standard normal distribution, the cumulative probability of the value 0 is 0.5, which corresponds to a return period of 2 years. The cumulative probability of the value 0.1 is 0.54, which corresponds to a return period of 1/(1-0.54)=2.2 years. The difference in terms of return periods is 0.2 years for an increase in the quantile by 0.1. If we now go to extreme quantiles, say 3.0 and 3.1 the standard normal cumulative probabilities are 0.9986501 and 0.9990324 respectively, which amounts to a difference in return periods of 293 years i.e. more than 100 times change in return period for the same change in quantile.

The ERA-Interim data and the NOAA data are actually quite similar. This can be seen for example from the scatterplot below that has been plotted for the Nov-Dec averages of El Nino 1+2 in the 1979-2011 period.

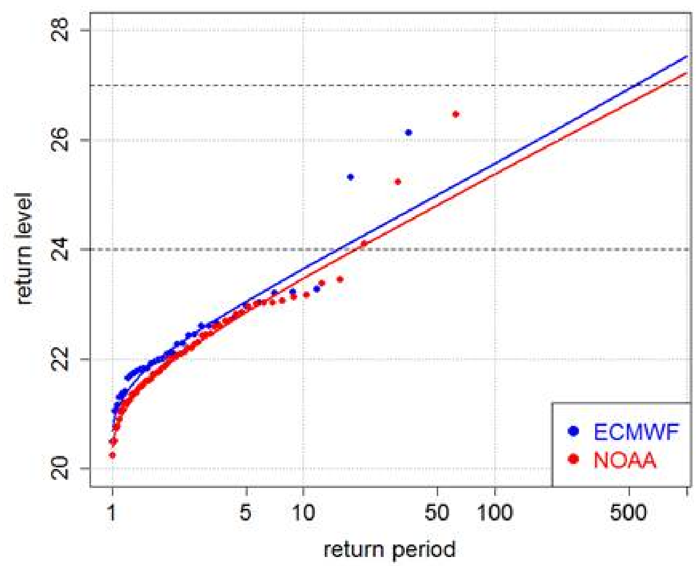


Now I make a return period/return level plot for each dataset:





The red dots show the actual Nino 1+2 Nov-Dec averages with empirical return periods, the blue lines show the GEV fit with the 95% confidence interval. The entry and exit points of the contract (24 and 27 °C) are shown with dash lines. The horizontal axis is logarithmic. There are more data points for the NOAA dataset (1950-2011) compared to ERA Interim (1979-2013). The fits are quite similar. Maybe I would prefer the NOAA data because the correspond to a longer time frame (62 years) and this facilitates investigating extremes. The larger number of data points in the NOAA dataset is also reflected on the narrower confidence intervals.



If I put the two plots together, the two lines have a similar shape but they are shifted by about 0.02 °C. This is the result of a difference in the mean values of the two datasets: the ERA-Interim mean is 22.3 °C and the NOAA mean is 22.1 °C. I guess this is not a big difference in general. However, it becomes big when we go to return periods of extremes as I described above. The corresponding difference in the theoretical return periods starts from 4 years at the entry point and reaches about 200 years at the exit point. What I conclude from this, is that the datasets do not have particularly different features and that the difference in the return periods are the result of the sensitivity of the calculation to small random differences in the underlying data. The difference in the price can obviously be tracked back to those differences in the return periods.