

ENSO and regression to the mean

Elio Campitelli

2021-01-14

On a recent article, Cai et al. (2020) used a perturbed ensemble of 40 members to show that members with relatively low El Niño Southern Oscillation (ENSO) variability at the beginning of the simulated period have a greater increase in ENSO variability towards the end of the simulated period than members with relatively high ENSO variability. They take this as evidence of ENSO self-regulation which would, if real, have important implications for future projections of ENSO variability under global warming. Here, I show that their results are not surprising, are likely product of simple regression to the mean and have no physical interpretation.

1 Introduction

The Community Earth System Model Large Ensemble model (Kay et al. 2015) consists of 40 climate simulations (members) that ran from 1920 to 2100 with identical external forcing and physics. Their only difference is a small, round-off error, perturbation in their atmospheric initial conditions. Cai et al. (2020) computed, for each member, the variability (standard deviation) of an index of Eastern-Pacific ENSO they call “E-index” for the first and last 50 years of the simulation. Owing to the effect of the various initial conditions, the variability in the E-index varied between members. I show their main result in Figure 1, which reproduces their Figure 1b. It shows a negative correlation between the initial E-index variability and its future change, defined as the difference between the final and the initial E-index variability.

```
library(magrittr)
library(data.table)
library(ggplot2)
library(patchwork)
theme_set(theme_minimal(base_size = 10) + theme(panel.grid = element_blank()))
```

```
obs_paper <- fread(here::here("analysis/data/enso-variability.csv")) %>%
  .[, final := initial + change]
moments <- obs_paper[, .(mean_initial = mean(initial),
                        sd_initial = sd(initial),
                        mean_final = mean(final),
                        sd_final = sd(final))]
```

```
set.seed(42)
members <- 40
B <- 10000
sims_normal <- CJ(sim = seq_len(B), member = seq_len(members)) %>%
  .[, `:=`(initial = rnorm(.N, moments$mean_initial, moments$sd_initial),
          final = rnorm(.N, moments$mean_final, moments$sd_final))] %>%
  .[, change := final - initial]
```

```
cor <- obs_paper[, cor.test(initial, change)]

cor_text <- with(cor, paste0("R = ", signif(estimate, 2), "\nP ~ ", signif(p.value, 2)))

obs_paper[, rank := cut_number(initial, 3)]

obs_paper %>%
  ggplot(aes(initial, change)) +
    geom_point() +
    geom_smooth(method = "lm", color = "#0e9a83") +
    annotate("text", label = cor_text, x = .51, y = 0, hjust = 0, size = 4) +
    scale_x_continuous("Initial E-index variability (s.d.)") +
    scale_y_continuous("Change in E-index variability (s.d.)")

## 'geom_smooth()' using formula 'y ~ x'
```

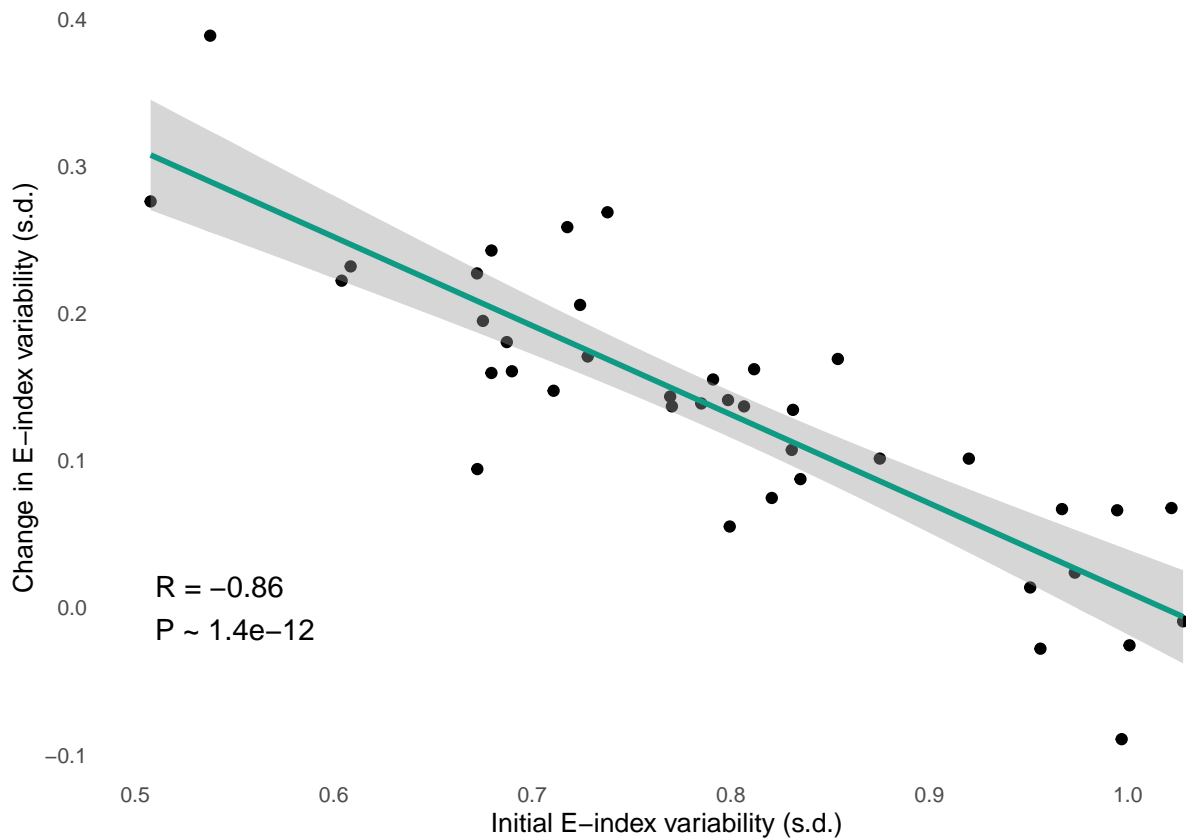


Figure 1: Reproduction of Figure 1b from Cai et al. (2020) based on digitised values. The horizontal axis shows the initial variability of their ENSO index for the first 50 years of simulation (1920 – 1969) and the vertical axis shows the difference between the variability of their ENSO index for the last 50 years of the simulation (2050 – 2099) and the first 50 years.

From this negative correlation, Cai et al. (2020) conclude that “strikingly, in experiments with initially higher ENSO variability (...), [its] amplitude in the future a century later [is] systematically smaller, and vice versa”. They attribute this to a novel mechanism of self-regulation involving non-linear thermal damping through upper-ocean heat exchange and stratification.

2 Simulations

To test how striking this result really is, I created 10^4 simulations of 40 pairs of random numbers representing the initial and final variability of the E-index for 40 hypothetical ensemble members. The random numbers were taken from independent normal distribution with $\mu = 0.8$ and $\sigma = 0.14$ for the initial variability and $\mu = 0.93$, $\sigma = 0.072$ for the final variability. These are the mean and standard deviation from the observed values derived from Figure 1.

As each simulated observation is independent from the rest, this simulations serve as a plausible model for the null hypothesis that the initial variability of ENSO does not influence it's future change.

```
two_plots <- function(sim, obs, n = 1:10) {
  fits <- sim[, metR::FitLm(change, initial, se = TRUE), by = sim] %>%
    .[term != "(Intercept)"] %>%
    .[, p_value := pt(abs(estimate)/std.error, df, lower.tail = FALSE), by = sim]

  obs_fits <- obs[, metR::FitLm(change, initial, se = TRUE)] %>%
    .[term != "(Intercept)"] %>%
    .[, p_value := pt(abs(estimate)/std.error, df, lower.tail = FALSE)]

  p_value <- fits[, mean(r.squared >= obs_fits$r.squared)]

  fit_density <- as.data.table(density(fits[term != "(Intercept)"]$r.squared)[c("x", "y")])

  x_pval <- mean(fits[term != "(Intercept)"][r.squared >= obs_fits$r.squared]$r.squared)
  y_pval <- mean(fit_density$y)

  sim[sim %in% 1:10] %>%
    .[, change := change] %>%
    ggplot(aes(initial, change)) +
    geom_point(size = 0.4, alpha = 0.3) +
    geom_point(data = obs, size = 0.7) +
    geom_smooth(data = obs, method = "lm", se = FALSE, size = 1.4,
               fullrange = TRUE, color = "#0e9a83") +
    geom_smooth(method = "lm", aes(group = sim), se = FALSE, fullrange = TRUE,
               color = "black", size = 0.1, alpha = 0.5) +
    scale_x_continuous("Initial E-index variability (s.d.)") +
    scale_y_continuous("Change in E-index variability (s.d.)") +
    # coord_equal() +

  fit_density %>%
    ggplot(aes(x, y)) +
    geom_area(data = ~.x[x >= obs_fits$r.squared], fill = "#4d4d4d") +
    annotate("text", size = 6, color = "white",
           label = scales::percent(p_value), x = x_pval, y = y_pval) +
    geom_line() +
    geom_vline(xintercept = obs_fits$r.squared, color = "#0e9a83", size = 1.4) +
    scale_x_continuous("r2") +
    scale_y_continuous(NULL) +

  plot_annotation(tag_levels = "a")
}
```

```
compute_pvalue <- function(sim, obs) {
  fits <- sim[, metR::FitLm(change, initial, se = TRUE), by = sim] %>%
    .[term != "(Intercept)"] %>%
    .[, p_value := pt(abs(estimate)/std.error, df, lower.tail = FALSE), by = sim]

  obs_fits <- obs[, metR::FitLm(change, initial, se = TRUE)] %>%
    .[term != "(Intercept)"] %>%
    .[, p_value := pt(abs(estimate)/std.error, df, lower.tail = FALSE)]

  fits[, mean(r.squared >= obs_fits$r.squared)]
}
```

```
two_plots(sims_normal, obs_paper)
```

```
## 'geom_smooth()' using formula 'y ~ x'
## 'geom_smooth()' using formula 'y ~ x'
```

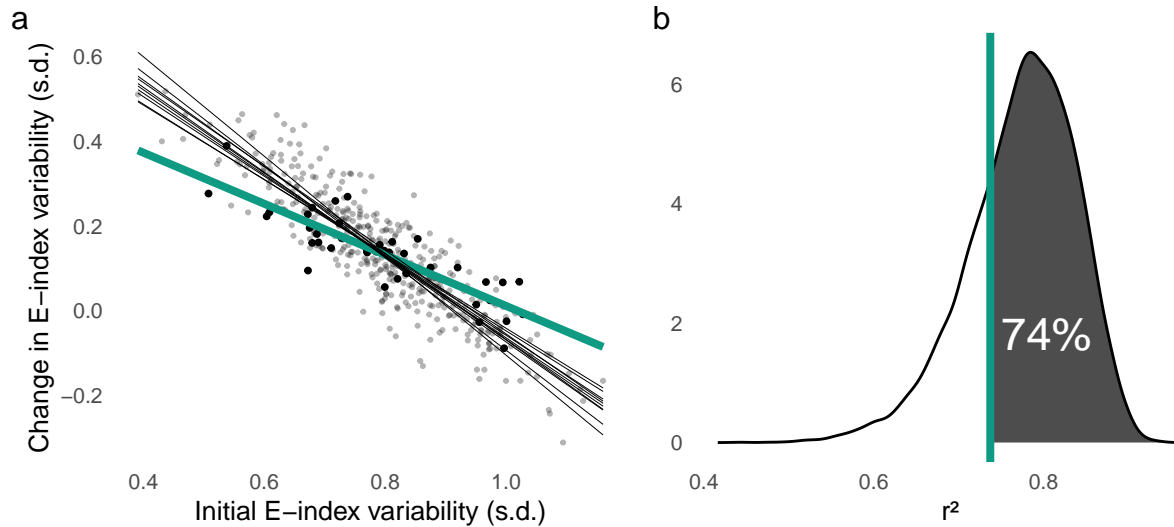


Figure 2: a. Ten simulated ensembles of 40 members drawn from random numbers (see description in text) in small dot and their linear regression in fine lines. The observed values from Cai et al. (2020) in big dots and the linear regression in thick, green. b. Estimated probability density of the coefficient of determination (r^2) under the null hypothesis model. The vertical bar shows the observed r^2 and the shaded area is the proportion of simulated samples from the null hypothesis model with a r^2 equal or greater than the observed one.

Figure 2.a shows a random sample of 10 of these simulations in small dots and their respective linear fit in fine lines. Comparing these random simulations with the observed values taken from Cai et al. (2020) – shown in big dots and thick, green line – puts into perspective the strength of their evidence. The spread of the simulated data is very similar to the real data and all 10 simulation have a steeper regression slope. Both sets are so similar that it would be virtually impossible to distinguish between the simulated and real data if not for the different size of the points.

```
p_value <- compute_pvalue(sims_normal, obs_paper)
```

To quantify how surprising the observed effect actually is, Figure 2.b shows the estimated probability density of the coefficient of determination (r^2) from the linear fit of the 10^4 simulations and the observed value as a

vertical line. 74% of simulations show an r^2 equal or greater than the observed one. This translates to a p-value of 0.74; significantly higher than 1.4×10^{-12} , naively computed from the linear regression.

Then, it's evidently clear that under this model of the null hypothesis the observed negative correlation is not only unsurprising, but completely expected.

The fact that the initial value of a variable is negatively correlated with its “future change” is one of the manifestation of the general principle of regression to the mean first observed by Galton in 1889 (Francis Galton 1889). He noted that very tall parents tended to have relatively shorter children while tall children tended to have relatively shorter parents.

In a perturbed ensemble experiment some ensemble members will have higher initial ENSO variability than average, while other will have lower ENSO variability than average. If this initial ENSO variability is completely independent of further ENSO variability (e.g. the null hypothesis is true), then both groups are equally likely to have average variability further down the simulation. Thus, members which a high initial ENSO variability will tend to show – on average – a negative change in variability, an vice versa.

3 Formal formula

It's straightforward to show this effect mathematically. The correlation between two variables x and $z = y - x$ can be written as:

$$\text{cor}(x, y - x) = \frac{\text{cov}(x, y) - \text{var}(x)}{\sqrt{\text{var}(x)\text{var}(y) + \text{var}(x)^2 - 2\text{var}(x)\text{cov}(x, y)}} \quad (1)$$

If x and y are independent random variables, $\text{cov}(x, y) = 0$ vanish and Equation (1) simplifies to

$$\text{cor}(x, y - x) = \frac{-1}{\sqrt{\text{var}(y)/\text{var}(x) + 1}} \quad (2)$$

This final formula shows that the correlation between x and $y - x$ is bound to be negative and its magnitude depends only on the relationship between their variances. That is, if the variance of y is greater than the variance of x , then the correlation will be small, and vice versa.

Applying Equation (2) to the issue at hand, x becomes the initial ENSO variability and y , the final ENSO variability and $\text{var}(x)$, $\text{var}(y)$ are the ensemble spread in initial and final ENSO variability, respectively. This means that, even in the case of no linear relation between initial and final variability, any process – either driven by physics or models – which reduces ensemble spread will lead to an even stronger negative linear relationship between the variables.

```
alpha_obs <- moments[, sd_final/sd_initial]^2
pred <- -1/sqrt(alpha_obs + 1)
```

In the case of equal variances, Equation (2) predicts a correlation of $-1/\sqrt{2} \sim -0.71$. The observed correlation of -0.86 (Figure 1) is stronger This is explained by the reduced ensemble spread in ENSO variability between the initial and final periods, as the final spread is around 29% of the initial spread. For this value, Equation (2) predicts a correlation of -0.88, which is very close (if a bit stronger) to the observed value.

4 Conclusion

The previous analysis puts the strength of Cai et al. (2020) evidence into perspective. That the change in ENSO variability is greater in ensemble members with initially low ENSO variability is not at all surprising.

By simulating the null hypothesis, I show that a negative correlations as strong or stronger than the one observed by Cai et al. (2020) can be expected 74% of the time.

These strong negative correlation can be explained by regression towards the mean and reduced ensemble spread in the final period compared to the initial period. I don't think it is possible to identify the source of the spread reduction. It might be a real response to the forcing scenario or inability of the ensemble to capture all the sources of variability. In particular, all members share the same physics and boundary conditions, therefore I believe it would not be unexpected if all members slowly converged to a similar state. Kay et al. (2015)'s Figure 2 does appear to show that the ensemble spread is slightly higher before 1970. Of note, recently Bengtsson and Hodges (2019) observed secular reductions of ensemble spread in mean surface temperature in an ensemble whose members where also all forced with the same radiative conditions.

Failing to account for regression to the mean is a fallacy that affects many fields of science – including behavioural science (Kelly and Price 2005) and medicine (Chuang-Stein and Tong 1997) – and features prominently in popular culture, such as the “Sports Illustrated cover jinx” – the perception that an athlete's performance will be “cursed” after appearing in the cover of Sports Illustrated (Goldacre 2008). Cai et al. (2020) is not the first paper nor will be the last to fall for it.

Bengtsson, L., and K. I. Hodges. 2019. “Can an Ensemble Climate Simulation Be Used to Separate Climate Change Signals from Internal Unforced Variability?” *Clim Dyn* 52 (5): 3553–73. <https://doi.org/10.1007/s00382-018-4343-8>.

Cai, Wenju, Benjamin Ng, Tao Geng, Lixin Wu, Agus Santoso, and Michael J. McPhaden. 2020. “Butterfly Effect and a Self-Modulating El Niño Response to Global Warming.” *Nature* 585 (7823): 68–73. <https://doi.org/10.1038/s41586-020-2641-x>.

Chuang-Stein, Christy, and Donald M Tong. 1997. “The Impact and Implication of Regression to the Mean on the Design and Analysis of Medical Investigations.” *Statistical Methods in Medical Research* 6 (2): 115–28. <https://doi.org/10.1177/096228029700600203>.

Francis Galton. 1889. *Natural Inheritance*.

Goldacre, Ben. 2008. *Bad Science*. HarperCollins UK.

Kay, J. E., C. Deser, A. Phillips, A. Mai, C. Hannay, G. Strand, J. M. Arblaster, et al. 2015. “The Community Earth System Model (CESM) Large Ensemble Project: A Community Resource for Studying Climate Change in the Presence of Internal Climate Variability.” *Bull. Amer. Meteor. Soc.* 96 (8): 1333–49. <https://doi.org/10.1175/BAMS-D-13-00255.1>.

Kelly, Colleen, and Trevor D. Price. 2005. “Correcting for Regression to the Mean in Behavior and Ecology.” *Am Nat* 166 (6): 700–707. <https://doi.org/10.1086/497402>.