## Exercises - GEO4902-04 - Verification

**The assignment is due by ?, Nov.?, 2021**

***[Save as jupyter notebook or html, with your name in the filename, and upload on canvas]***

### Exercise - Analysis of point “forecast” of two temperature and wind observations

**Use:**

<https://colab.research.google.com/github/maltemuellerm/GEO4902/blob/main/04/Verification_Exercises.ipynb>

The script loads the data of two insitu wind and temperature observations and the corresponding model forecast. Note the model temperatures are in Kelvin. The data is stored as pandas dataframes, for example, for the model temperature and time:

st1\_temp.model

st1\_temp.time

1. Plot as time series the model and observation together in one plot. (for both wind and temperature)
2. Can you find out if the errors are conditional or not? Describe for all four observations/models. (You can look in the lecture slides for more details on conditional errors.)
3. Calculate the bias and root mean square error (RMSE)
4. Remove the bias and compute the RMSE again.
5. Smooth the time series of the de-biased wind observations. e.g., with

from scipy.signal import savgol\_filter

result = savgol\_filter(np.array(st2\_wind.model), 25, 2)

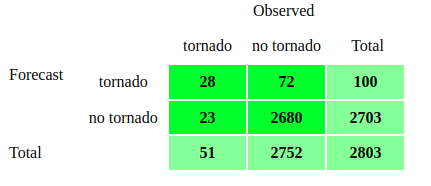
result = savgol\_filter(np.array(st2\_wind.model-bias2\_w), 25, 1)

Plot the time series for observations and smoothed and de-biased models. Recalculate the RMSE values. What changed?

Based on this result, why can a coarser resolution model have a lower RMSE, although it is not necessarily more “accurate”?

### Exercise - 2

**Categorical verification:**  John Finley computed tornado forecasts for the United States for eighteen regions east of the Rocky Mountains. Finley provided the following contingency table for forecasts of tornados (Finley 1884):



1. Compute the *fraction correct*, *frequency bias*, *probability of detection*, *threat score,* and *equitable threat score.* For each score, write a sentence or two how this value can be interpreted.
2. It was pointed out by *Gilbert (1884)* that the *fraction correct* score can be manipulated if just “no tornado” is forecasted. Provide the contingency table for this situation and calculate the fraction correct score to verify Gilberts’s critique.
3. The results of Gilbert led to the development of more meaningful verification scores. And the *Equitable threat score* is also called *the Gilbert skill score*. What is the ETS in case of the manipulated contingency table?

Reflection Notes by students:

1. What did you learn?

How to find the errors between model and observation, finding the unbiased model data with respect to the observation. What the different scores mean and how to calulate it from a table.

1. What was difficult?

Understanding what the different scores means, in comparison to each other.

1. What went well?

Handling the data becomes better, and how to visualize it.

1. What was surprising?

There was no big surprises this time :-)

1. Where would you like to get some more input?

Interpretation of the different scores would be nice, and what the difference between them (in terms of interpretation.) Also more regarding biases and of model data.

What did I learn?

• What a conditional error is.

• How to calculate biases and the RMSE (root mean square error).

• How to use savgol\_filter from scipy.signal to smooth an array.

• Why a coarser resolution model can achieve a lower RMSE, although it is not necessarily more “accurate”.

• How to calculate different verification scores.

• That the fraction correct score doesn’t work well with rare events.

What was difficult?

• To convert the bias free and smoothed arrays back to Series with the same date time index they had prior.

What went well?

• The different calculations went smoothly, and the results were intuitive.

What was surprising?

• That there are so many clever verification scores.

• That the bias can be interpreted as the mean error. I have kind of thought of it solely as a conditional error.

What elements will you take with you and use next time?

• That I must think through which type of verification score to use for different

scenarios.

• That a coarser resolution model can achieve a lower RMSE, although it is not necessarily more “accurate”.

Where would you like to get some more input?

• Can’t think of anything in particular.

This exercise was really clear. For the first part it was really useful that you implemented the script to smooth the time series. I only don’t really know how to explain the last question of part 1. As I can’t attend the lecture on Monday, I hope it will become clear in the script of possible solution.

Part 2 exercise 1 was easy as you just had to fill in formulas. But exercise 2 with Gilbert was a little bit confusing. I missed that lecture due illness. I understand that the contingency table will have too many correct negatives in rare events so that there has to be a change but I don’t really know how to change the contingency table for this situation so I couldn’t do exercise 2 and 3. So I also hope it will

become clear when I see the script of possible solutions.

During this exercise, I think the biggest learning outcome was again, like the other exercises, to visualise the different terms and equations from the lectures. To be forced to work with them in a simplified way helps with the understanding and forces you to think through the different methods once more.

I would have liked to go a bit more into depth on the bias in the models. How it works, and why one can just decompose the values and subtract it afterwards.

What was difficult?

- Not difficult sometimes those assignments have a large volume and require a long time going though, like each questions has subquestions, often it requires comparing notes from script, lectures and my notes, and it all takes time and before submitting I need checking errors possible errors

- Equitable threat score was funky to understand, I read the description a few times and also googled for other definitions, I even found one publication

( https://centaur.reading.ac.uk/16253/1/ets\_is\_not\_equitable.pdf ) that equitable score was not equitable (Equitability Revisited: Why the ‘‘Equitable Threat Score’’ Is Not Equitable ROBIN J. HOGAN Department of Meteorology, University of Reading, Reading, United Kingdom)

- In case of indicators and scores, there was a term “total”. Sometimes it was

confusing what it referred to: (total number of all forecasts?). The contingency table contained 1 total row and total column, in addition to the total of totals... My guess: “total” was referring to the number of issues forecasts, regardless of success.

- in case of conditional errors I was not sure which method we should use - the one presented in script (with different lead times) or the one on slides

- in case of RMSE and smoothing I was not sure how I should calculate RMSE, additionally when I tried to calculate debiased RMSE values there were same as RMSe orignal

What did you learn?

Methods to learn about errors, comparing 2 forecasts. I was surprised and a bit disappointed that those indicators can be so easily manipulated. When I was reading the script - I could easily understand and I appreciated that formulas and pictures were ordered and labeled so it was so much easier to follow the thought.

What went well?

plotting as usual, calculating most eros I believe

What was surprising?

In case of station2 and wind linear correlation between wind and error.

How fast I could apply linear regression (we have data science course this semester)

What elements will you take with you and use next time? Where would you like to get some more input?

Having some hands-on experience on interpreting advantages and disadvantages of different methods of forecast error estimation.