Uncertainty in Seasonal Forecasting

Nir Krakauer

Department of Civil Engineering,
The City College of New York
nkrakauer@ccny.cuny.edu

In this talk

- What do we mean by a forecast?
- An information-theory approach to quantifying uncertainty
- Application to a simple seasonal prediction framework

Forecasting: part of our job description

Peter von Cornelius (1817), *Joseph Interpreting Pharaoh's Dream*

Good forecasting is an exercise in modesty

- A forecast can be as definite as we like – the proof is in the outcome
- We should know when we don't know – never omit uncertainty
- A good forecaster may not predict “black swan” events, but at least hasn't said they are impossible

What we don't know

- How much would we need to be told so that we're no longer ignorant?
- Information theory (Shannon 1948):
  - Suppose one of *n* outcomes must happen, for which we assign probability p_i
  - If we learn that this outcome did happen, we've learned $\log(p_i)$ bits
  - Summed over possible outcomes, our *expected* missing information is $\sum_{i=1}^{n} p_i \log(p_i)$

How useful is a forecast?

- Suppose that we learn that outcome *i* took place
- Under our baseline ignorance (e.g. climatology), the probability of *i* was p_i
- Suppose a forecaster had charged us to give a probability q_i instead. Intuitively, the forecast proved useful if $q_i > p_i$.
- The information gain from the forecast is $\log(q_i / p_i)$

A forecaster's track record

- Across multiple forecast verifications, the average information content of the forecasts is given by the average $\log(q_i / p_i)$
- Best case is to assign probability 1 to something that does happen: $\log(1 / p_i)$ bits gained
- Assigning zero probability to something that does happen is infinitely bad [log(0)]

Generalization to continuous variables

- Information gain is $\log(q(i)/p(i))$
- If the actual outcome was x, the forecast was Gaussian with mean m and SD σ, and the background had mean m_0 and SD σ_0, the information gain is $(z^2 - z_0^2) - \log(\sigma/\sigma_0)$, where $z = (x - m)/\sigma$

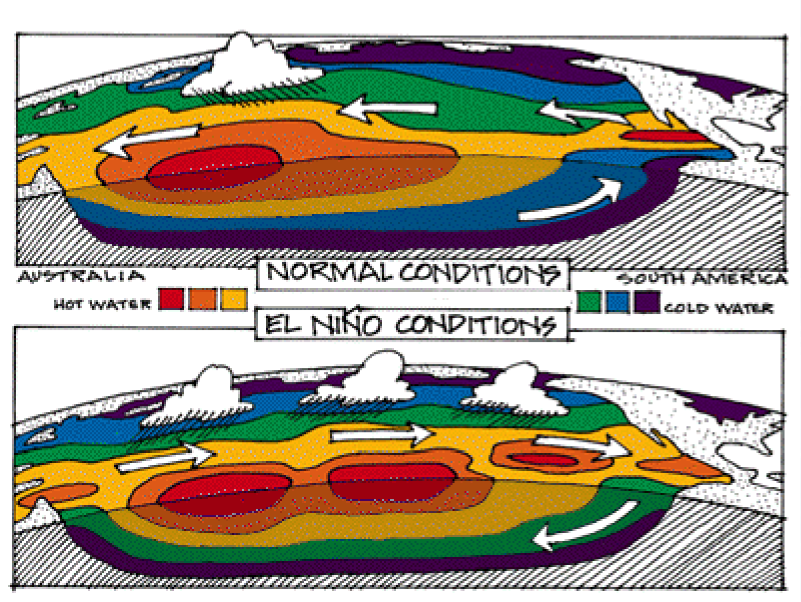
Example problem

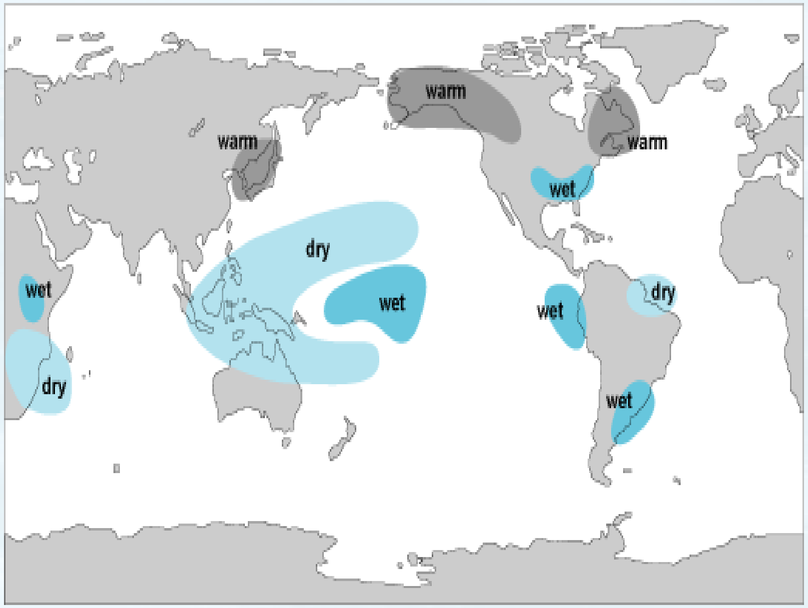
- Seasonal forecasting of temperature – PDF can be used as input to model lake level, streamflow, etc.
- Outside of synoptic forecast range, skill of existing products usually low
- Hindcast experiment: Data from NOAA (GHCN, 5° grid, monthly since 1880)
- Forecast July T from May observations

Approach

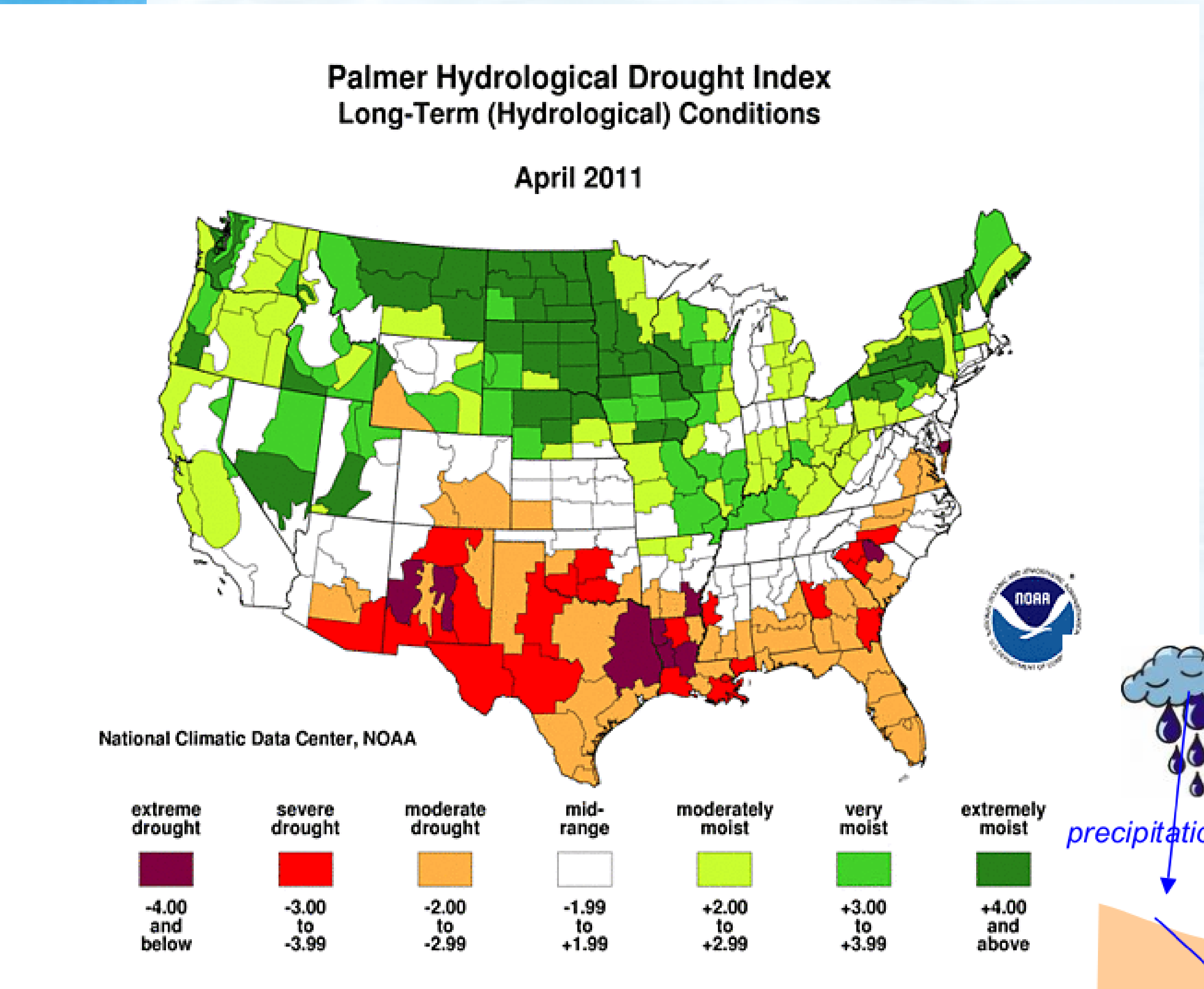
- Background mean and SD: from July past temperature distribution
- Forecast method:
  - Mean from nonlinear regression on quantiles of May SOI or PDSI
  - SD same as background (modesty)

Southern Oscillation Index

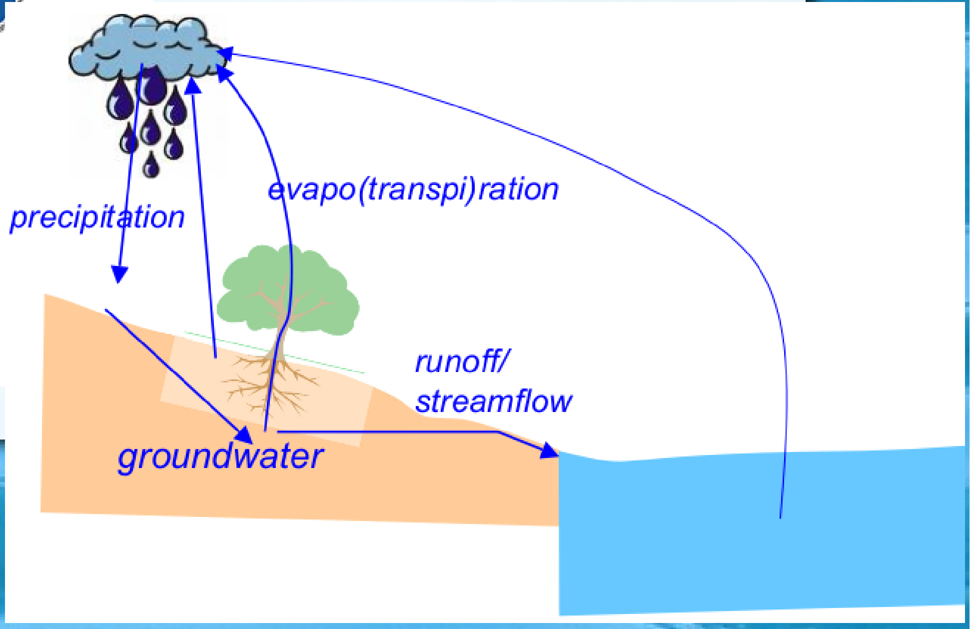




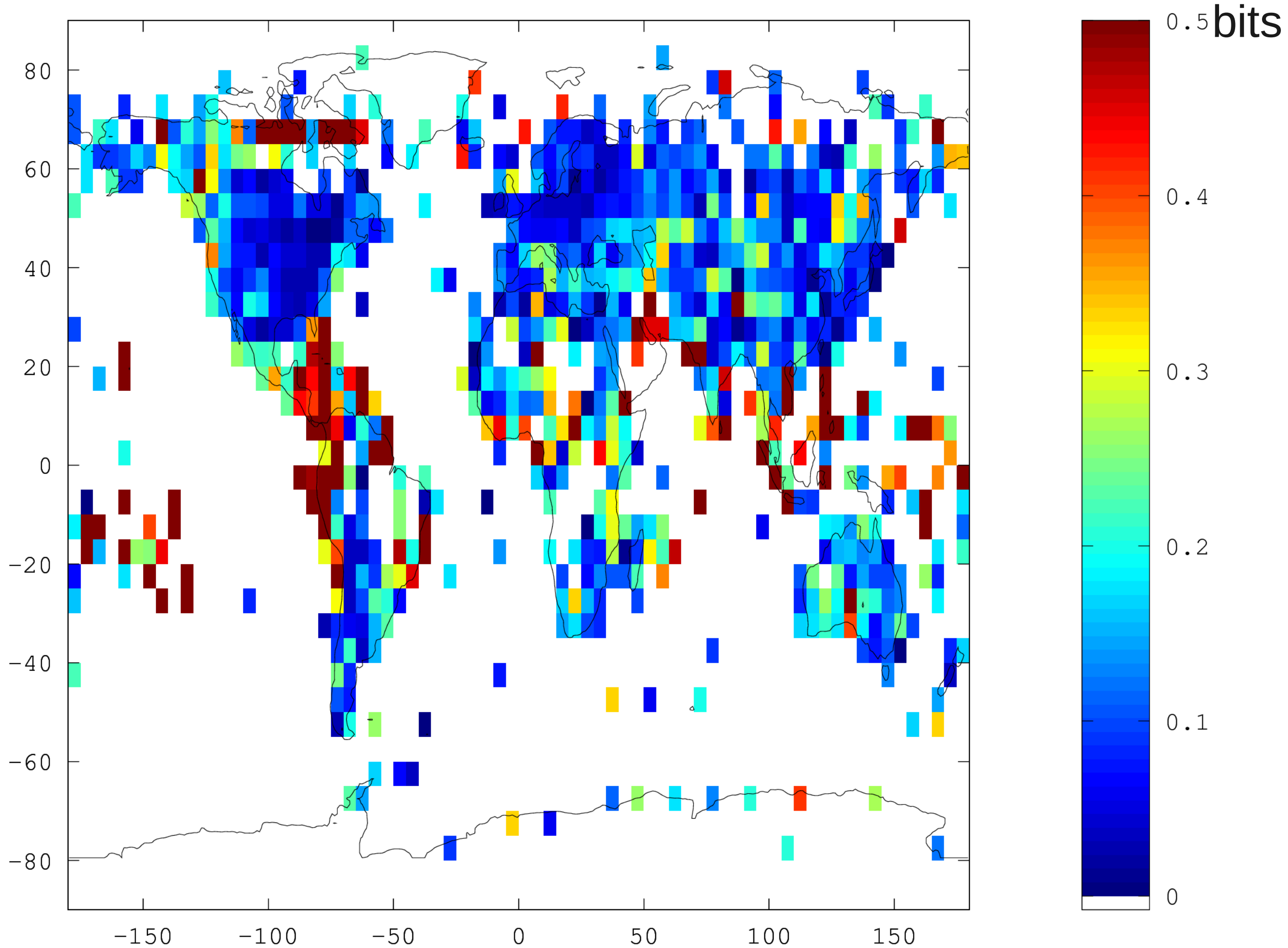
Palmer Drought Severity Index



Low PDSI (drought) leads to hotter summers because there's less water to evaporate

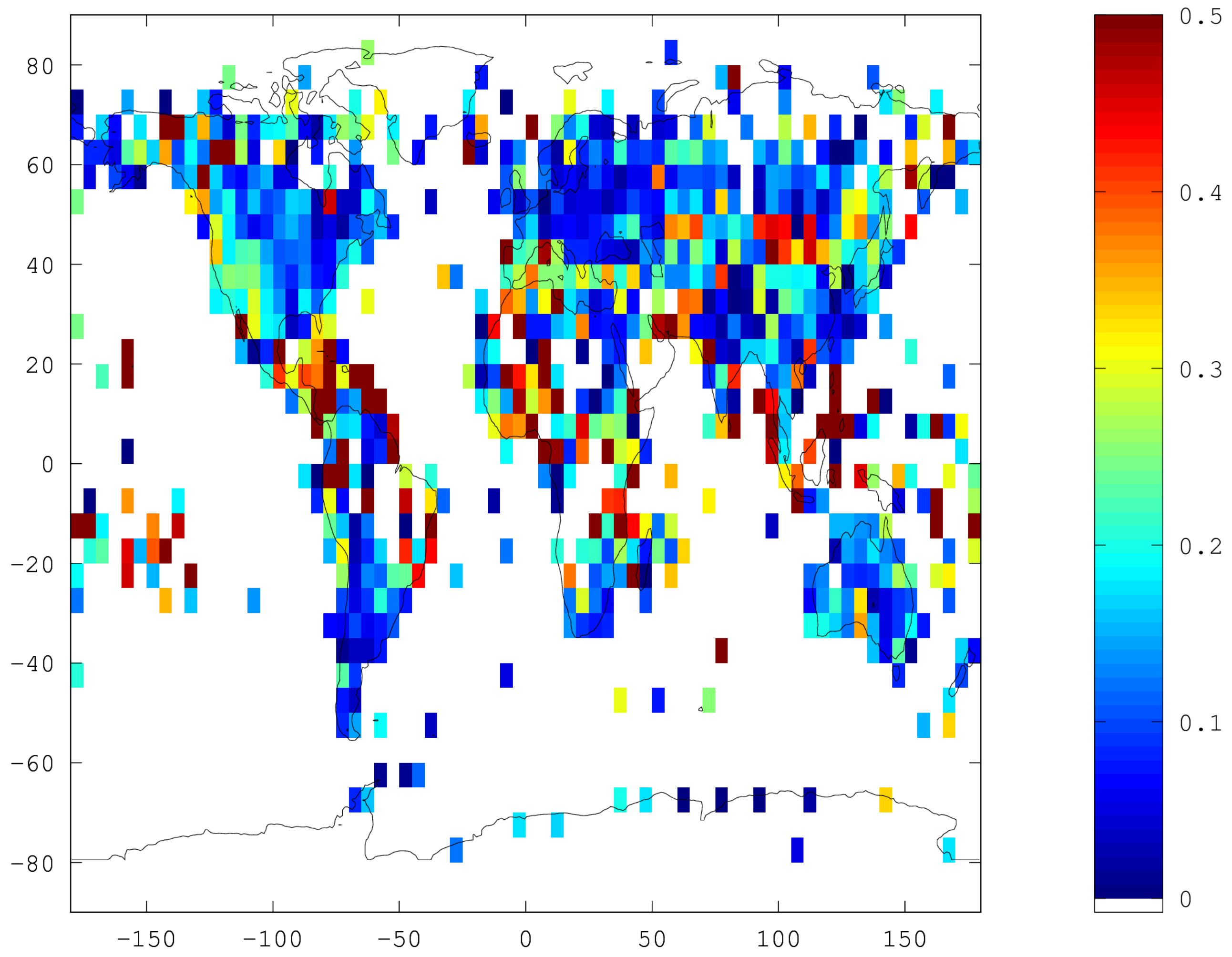


Information gain from SOI



Useful for forecasting land temperatures mainly in the tropics

Information gain from PDSI

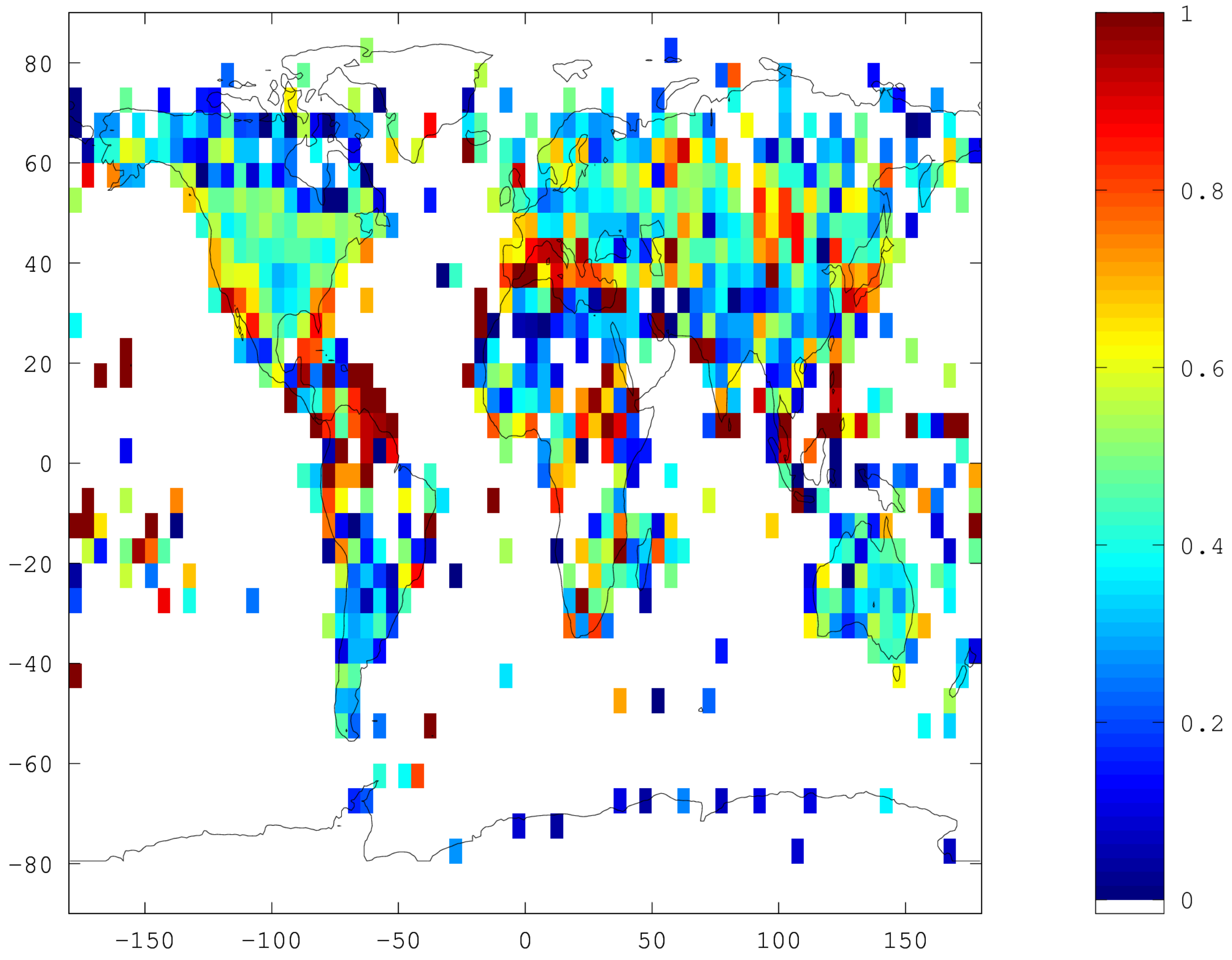


Useful in semiarid tropics and midlatitudes

Combining multiple forecasts

- Weight SOI and PDSI -based predictions for mean based on information gain of past forecasts (spatially variable)
- Add as predictors average Jul T from last few years (brings in temporal trend – single best predictor tested) and mean T in last 12 months (SOI timescale – some additional skill in tropics)

Information gain: combined forecast



Summary

- Thinking of a forecast as having an inherent, preferably explicit, uncertainty enables its usefulness to be evaluated
- An informative forecast must be modest
- Paying attention to uncertainty can yield additional information even for a difficult problem like seasonal prediction