Sub-seasonal Forecast Skill: When, Where and How to Find It?

Ángel G. Muñoz
agmunoz@iri.columbia.edu

International Research Institute for Climate and Society
Earth Institute | Columbia University

2020IVMWO, 18 Nov 2020

Sub-seasonal Skill: When, Where and How to Find It?

Ángel G. Muñoz
agmunoz@iri.columbia.edu

Caio Coelho
Andrew W. Robertson
Simon J. Mason
Frédéric Vitart
Kathy Pegion

International Research Institute for Climate and Society
Earth Institute | Columbia University

2020IVMWO, 18 Nov 2020

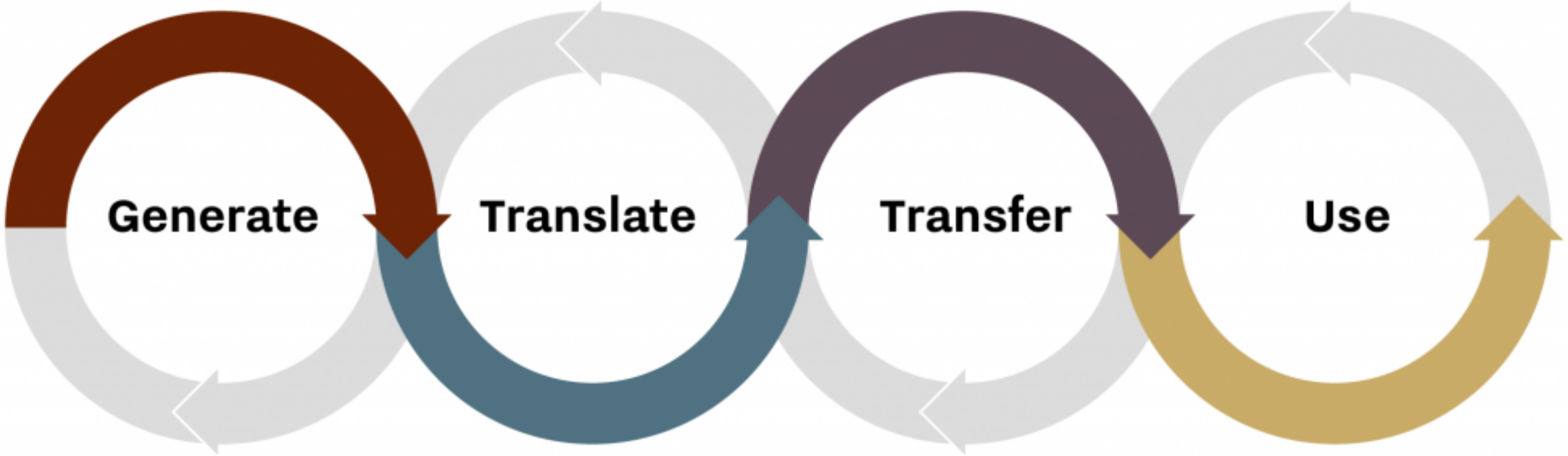
Outline

1. Introduction
2. On the seasonality of sub-seasonal skill
3. Can we improve sub-seasonal skill with calibration?
4. Concluding remarks and recommendations

Introduction

■ Generate climate information and knowledge - learn from the past, monitor the present, forecast the future.

■ Transfer the translated information to the appropriate beneficiaries, in formats and media most useful to their operations



■ Translate the climate knowledge into Information that is relevant to agriculture, public health and other target sectors.

■ Put the translated and transferred climate knowledge to use in operational decision processes, policies and plans. Learn what works and what doesn't.

IRI International Research Institute for Climate and Society
Earth Institute | Columbia University

Vaughan and Dessai (2014); IRI Comms

Ignorance Score

The Ignorance Score (IGN), or negative log-likelihood score, of a probabilistic forecast of *n* categories can be written as (Good, 1952; Roulston & Smith, 2002):

$$IGN = -\log_2(p_k) \qquad\qquad k = 1..n$$

and it can be decomposed into reliability, resolution and uncertainty terms:

$$IGN = \underbrace{REL}_{\text{calibration}} - \underbrace{RES}_{\substack{\text{sharpness} \\ \text{(if reliable)}}} + \underbrace{UNC}_{\text{obs distribution}}$$

(Weijs et al., 2010; Wilks, 2018)

- **It measures the information deficit, or ignorance**, of a person having a probabilistic forecast but not knowing the actual outcome.
- Units are *bits* of information. IGN=0 means perfect forecast (zero ignorance).
- **Each bit of ignorance represents a factor-of-2 increase in uncertainty.**
- Related to expected gambling return if used to place proportional bets on the future (cost-loss scenarios).

Ignorance Score

The Ignorance Score (IGN), or negative log-likelihood score, of a probabilistic forecast of *n* categories can be written as (Good, 1952; Roulston & Smith, 2002):

$$IGN = -\log_2(p_k) \qquad\qquad k = 1..n$$

and it can be decomposed into reliability, resolution and uncertainty terms:

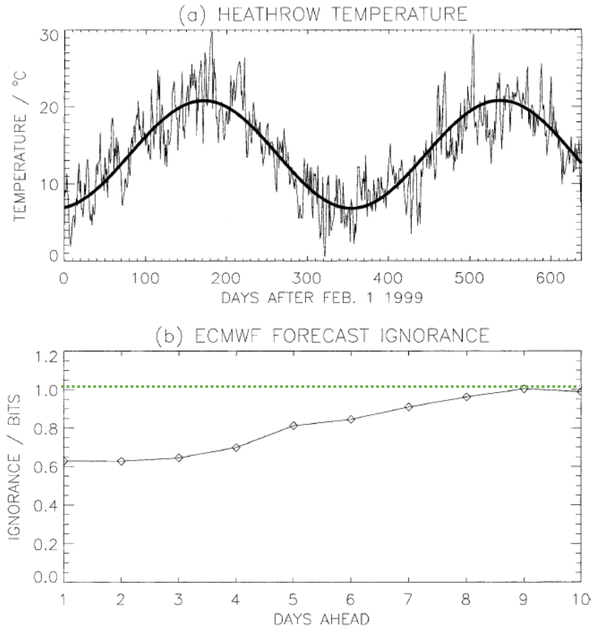
$$IGN = REL - RES + UNC$$

(Weijs et al., 2010; Wilks, 2018)

calibration — sharpness (if reliable) — obs distribution

There are different ways to define a skill score for IGN. Here we use climatology as the reference. For equiprobable climatological categories,

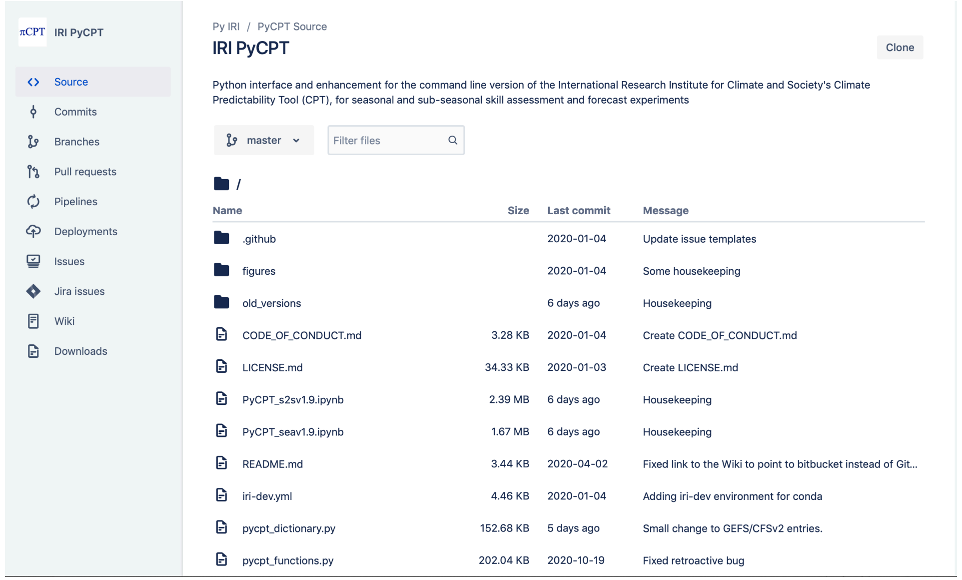
$$ISS = -\frac{\log_2(p_k)}{\log_2(n)} \begin{cases} > 1 & \text{Less info than climatology} \\ = 0 & \text{As good as climatology} \\ < 1 & \text{More info than climatology} \end{cases}$$

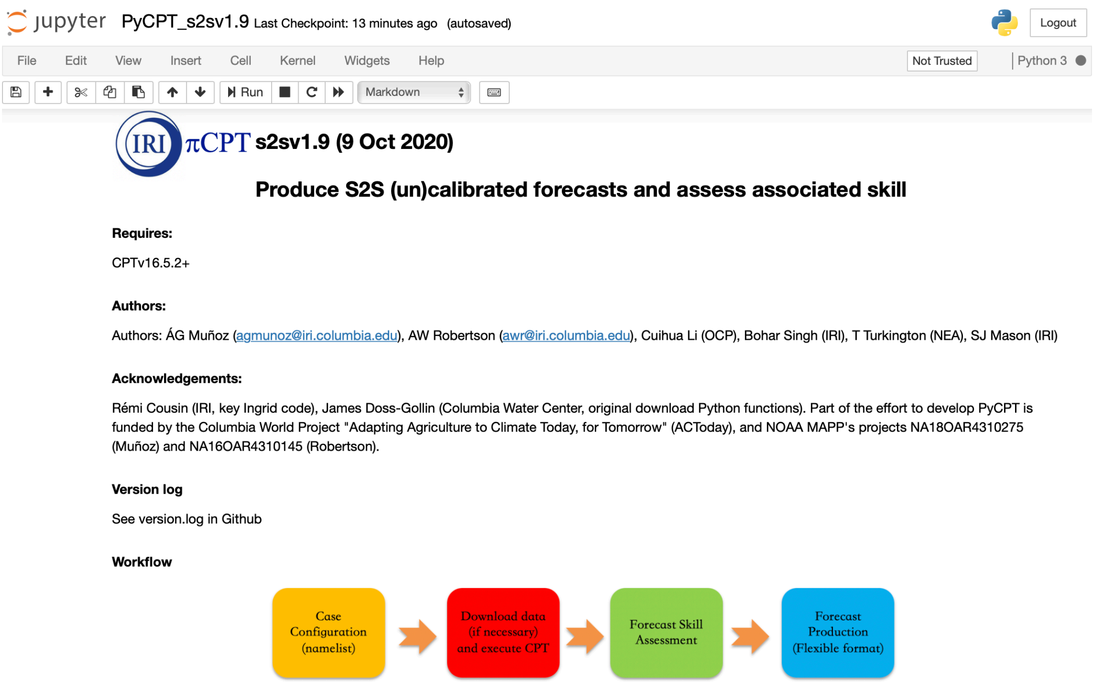


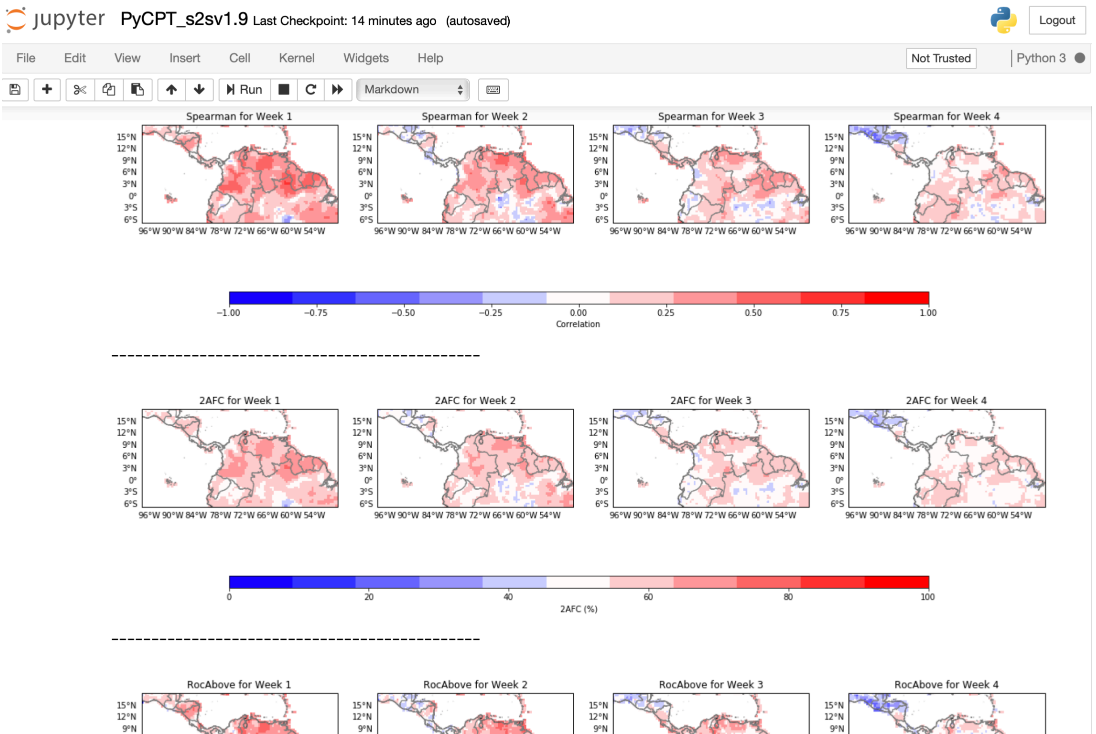
Roulston & Smith (2002)

https://bitbucket.org/py-iri/iri-pycpt/src/master/

PyCPT





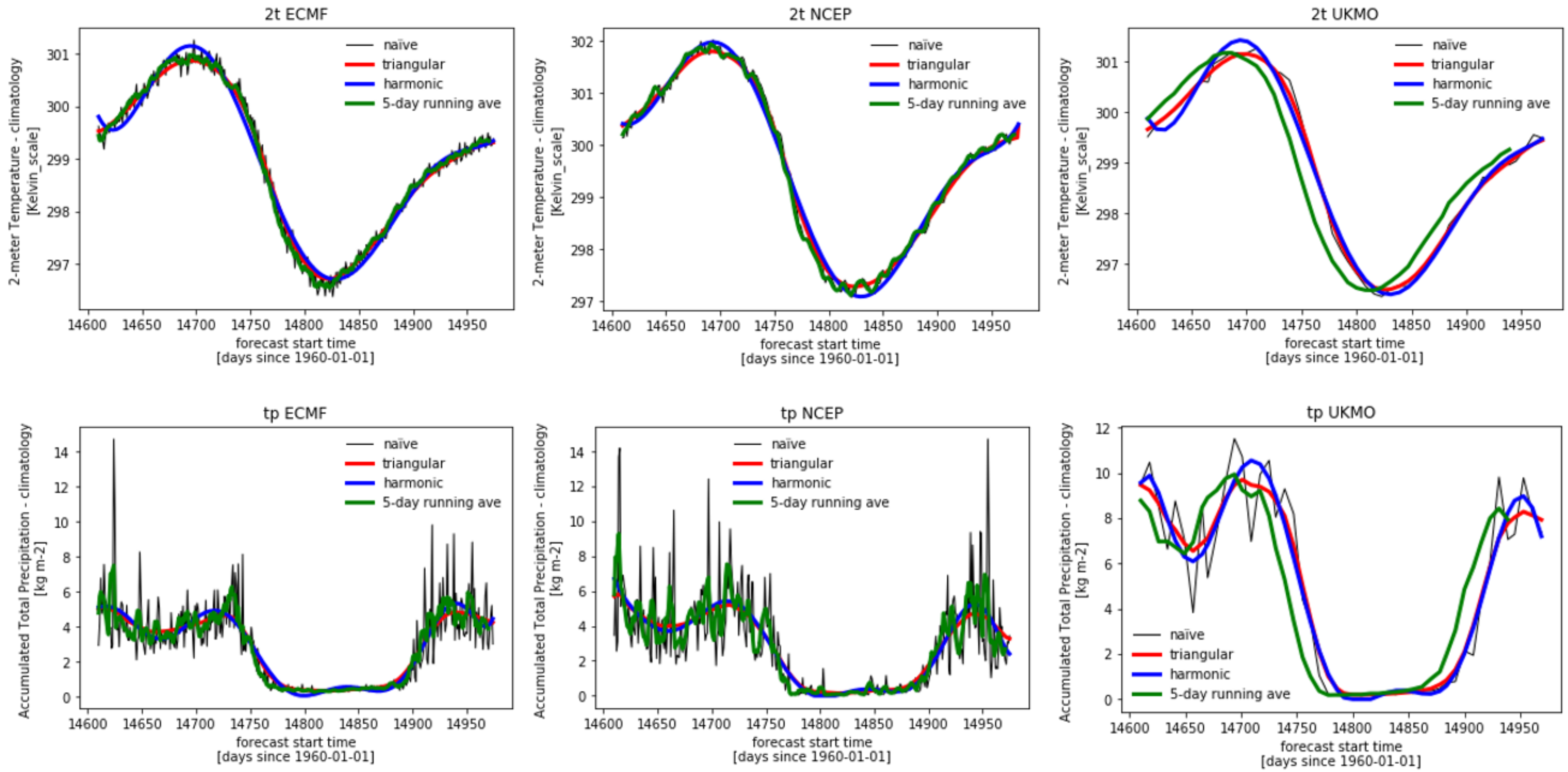


- Python interface for IRI's Climate Predictability Tool (CPT), a widely used research and application Model Output Statistics/Prediction toolbox.
- Publicly available: BitBucket.
- Automatically downloads required observations (CHIRPS, TRMM, CPC Unified) and S2S model data from the IRI Data Library (**S2S Database** and **SubX** – ECMWF, CFSv2, GEFS, others are being included).
- **Computes climatologies, anomalies**, a variety of skill metrics (uncalibrated and calibrated hindcasts) and deterministic and probabilistic sub-seasonal forecasts.

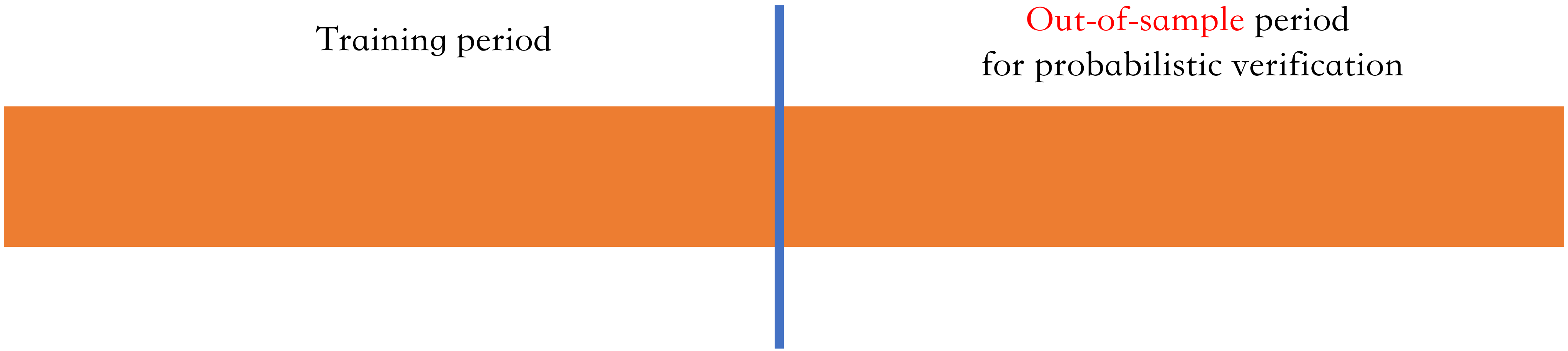
International Research Institute for Climate and Society
EARTH INSTITUTE | COLUMBIA UNIVERSITY

Funded by:
NOAA NA16OAR4310145 (Robertson)
NOAA NA18OAR4310275 (Muñoz)
Columbia World Project "ACToday" (Goddard)

PyCPT: Climatologies and Anomalies on the fly



PyCPT: Approach

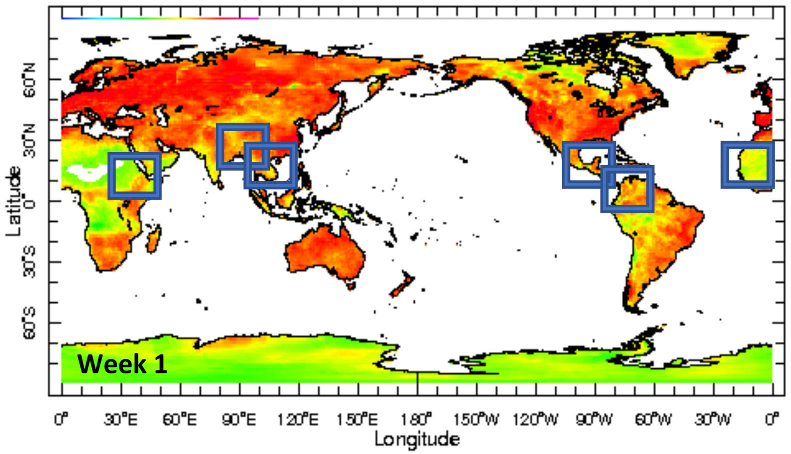


For ECMWF: we use all available initializations in a user-selected period,
e.g., 1 month provides a sample size of 160 years = 2 inits/week x 4 weeks x 20 hindcast years
Half of the period is used for the probabilistic forecast verification

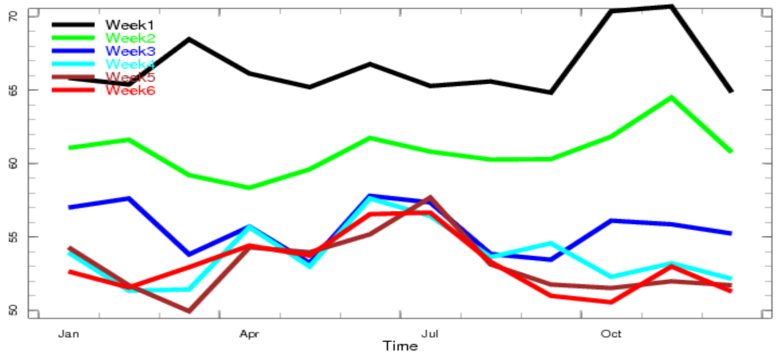
Outline

1. Introduction
2. On the seasonality of sub-seasonal skill
3. Can we improve sub-seasonal skill with calibration?
4. Concluding remarks and recommendations

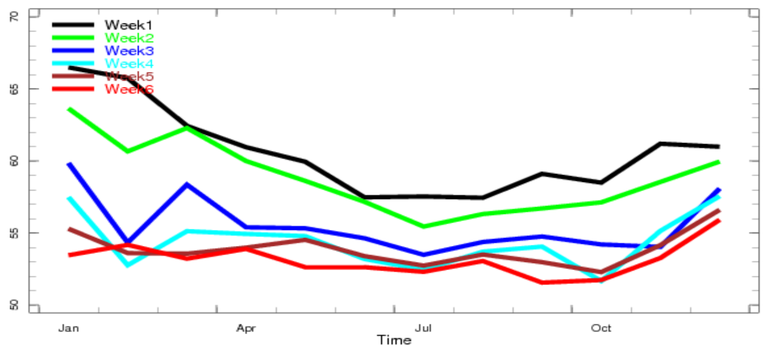
The Seasonality of Sub-Seasonal Skill



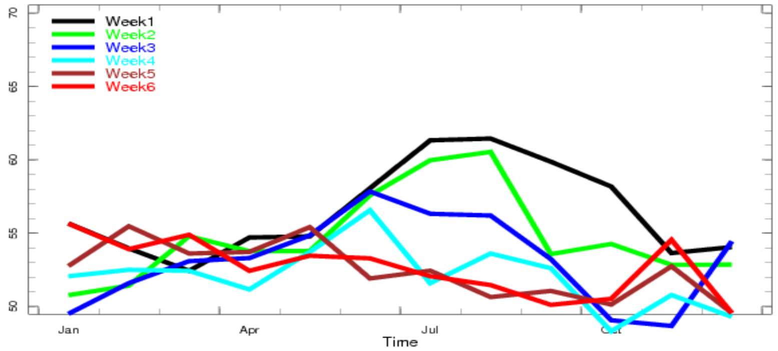
An "Obscurial", or just subseasonal forecast skill?



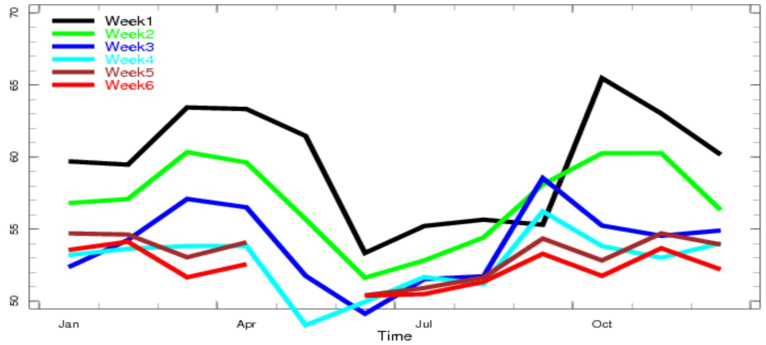
Guatemala - Deterministic 2AFC - Model: ECMWF (uncorrected)



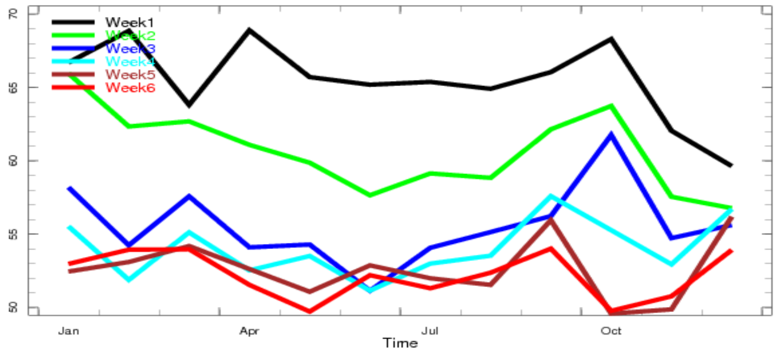
Continental Colombia - Deterministic 2AFC - Model: ECMWF (uncorrected)



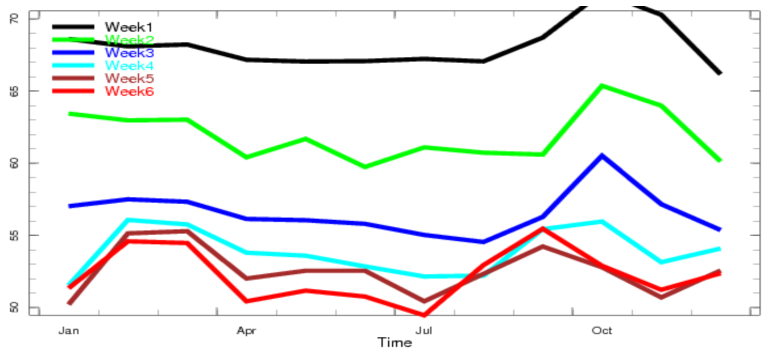
Senegal - Deterministic 2AFC - Model: ECMWF (uncorrected)



Ethiopia - Deterministic 2AFC - Model: ECMWF (uncorrected)



Bangladesh - Deterministic 2AFC - Model: ECMWF (uncorrected)



Vietnam - Deterministic 2AFC - Model: ECMWF (uncorrected)

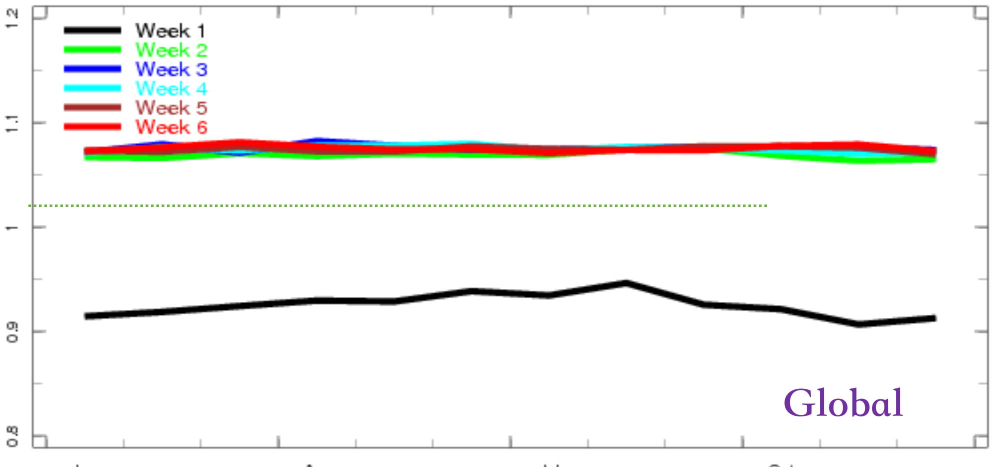
International Research Institute
for Climate and Society
EARTH INSTITUTE | COLUMBIA UNIVERSITY

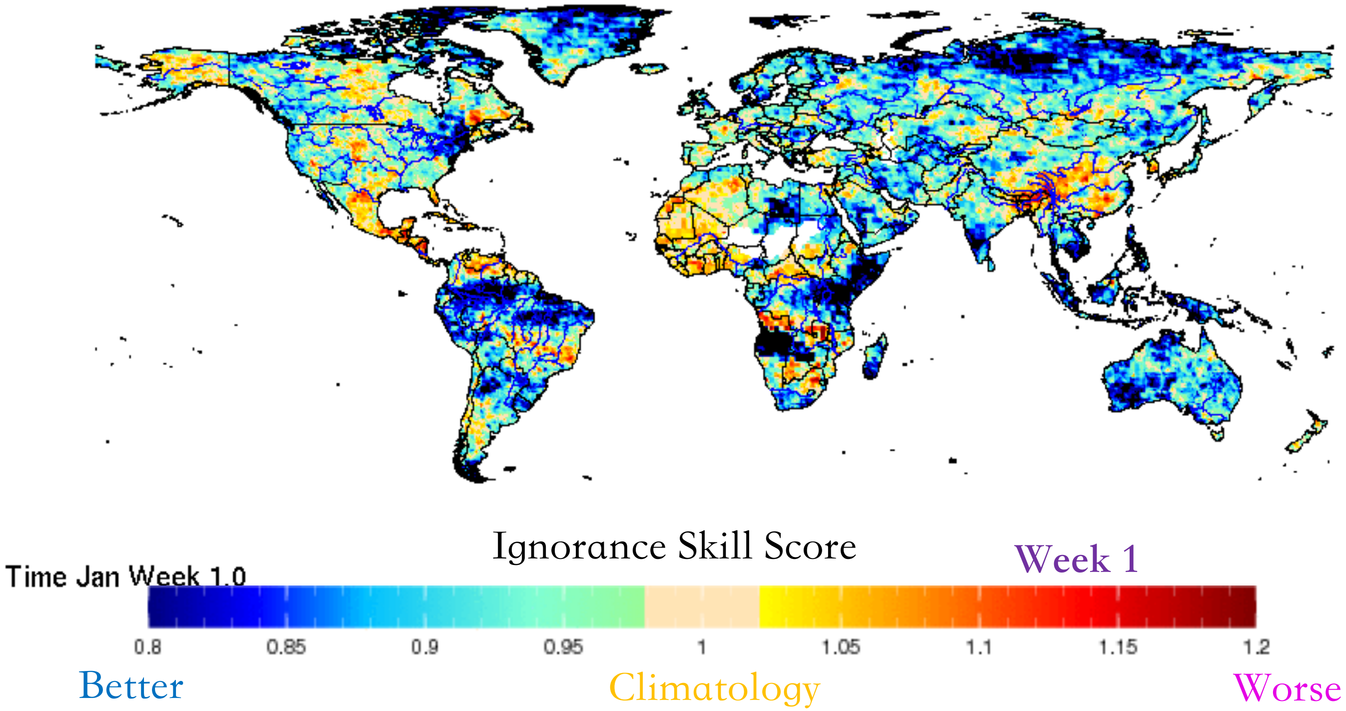
Muñoz *et al.* (2018)

Seasonality of Sub-seasonal Skill

Skill Assessment

- Model: ECMWF
- Rainfall
- Probabilistic Hindcasts
- Obs: CPC Unified
- All initializations available per month **Uncalibrated**
- IGN, RPSS, Brier and decompositions, for Week 1-6

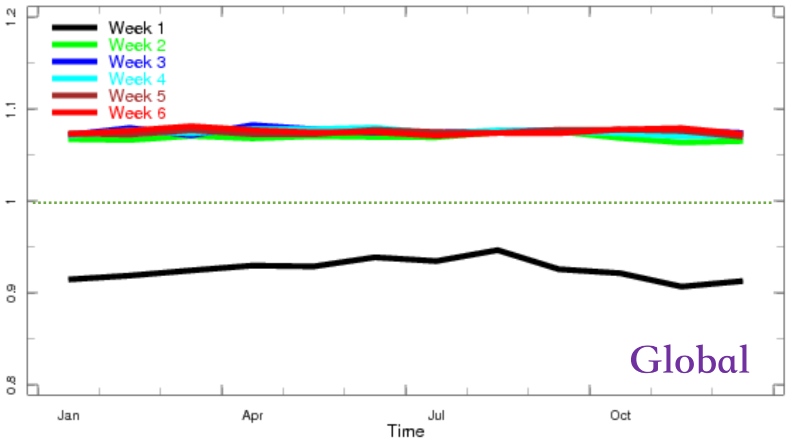


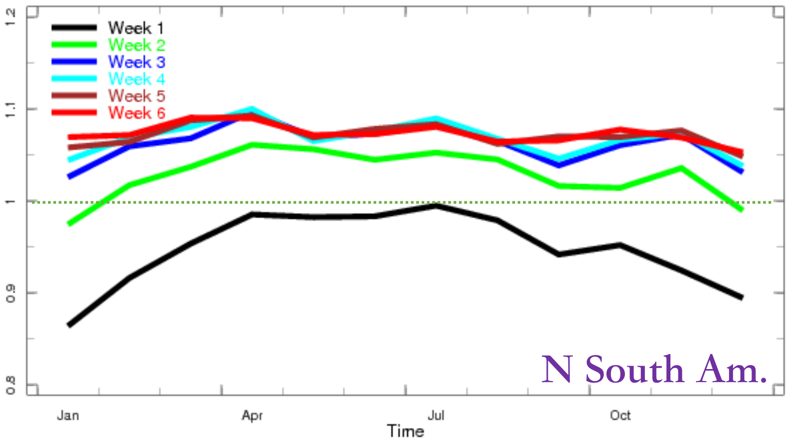


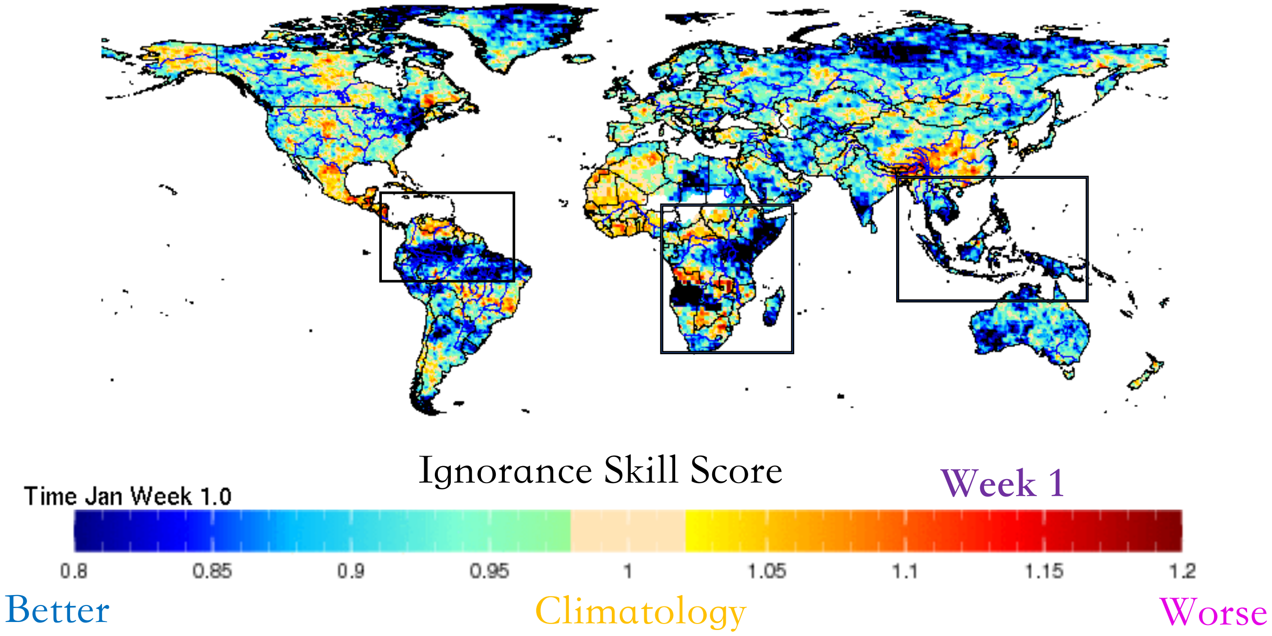
International Research Institute for Climate and Society
EARTH INSTITUTE | COLUMBIA UNIVERSITY

Muñoz et al. (2018)

Seasonality of Sub-seasonal Skill



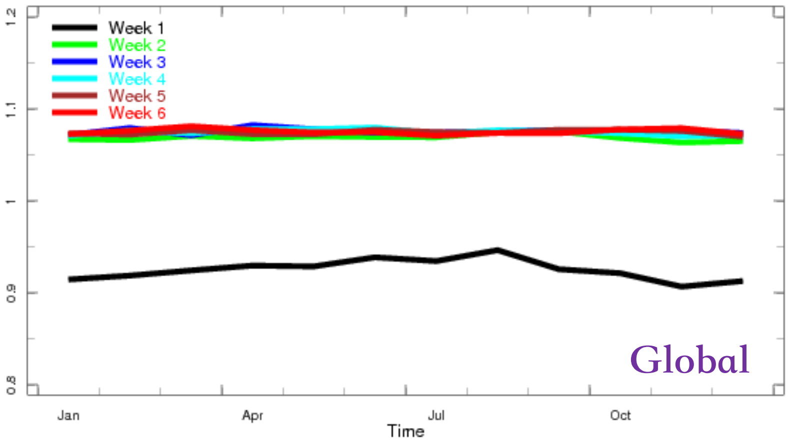


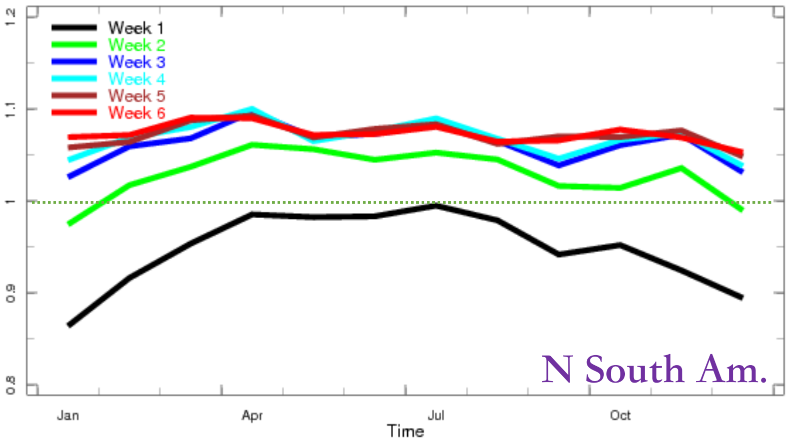


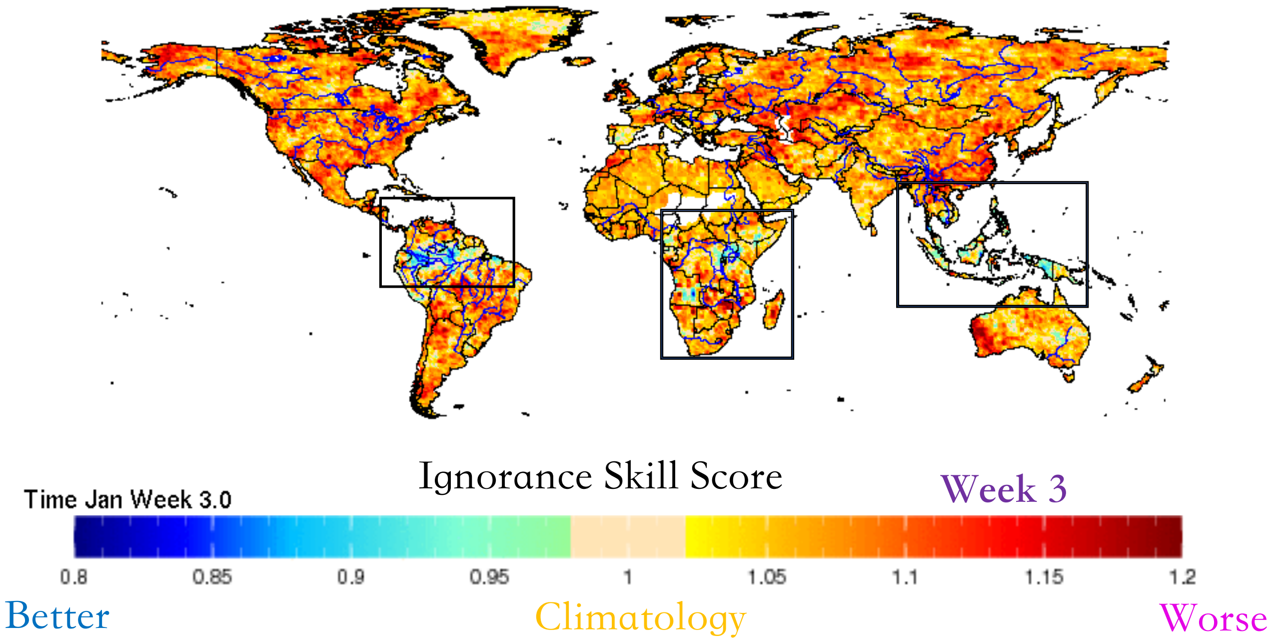
International Research Institute for Climate and Society
EARTH INSTITUTE | COLUMBIA UNIVERSITY

Muñoz et al. (2018)

Seasonality of Sub-seasonal Skill



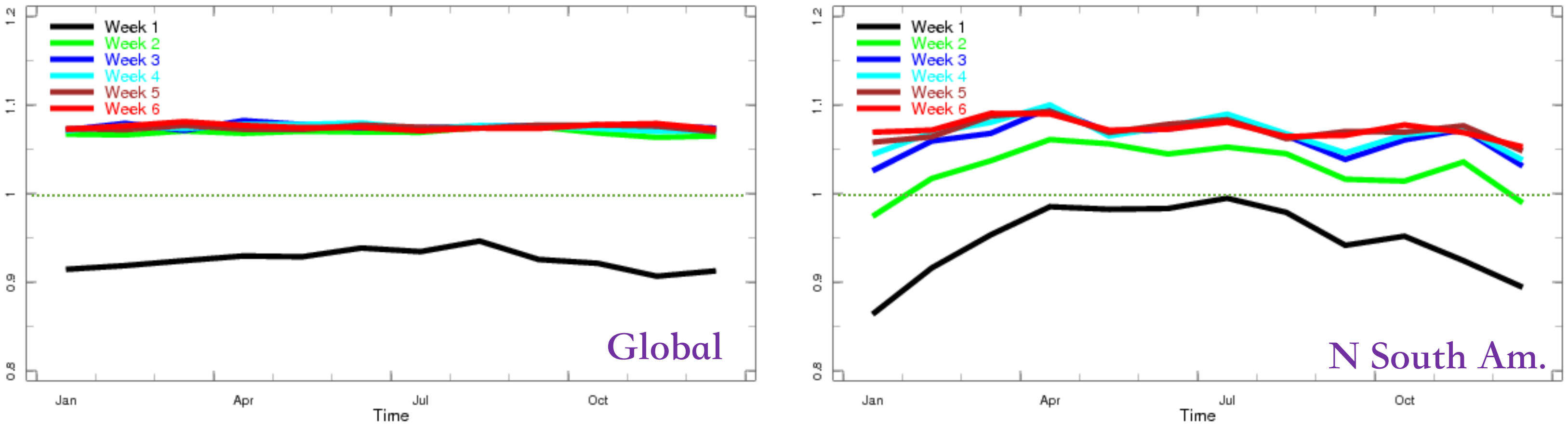


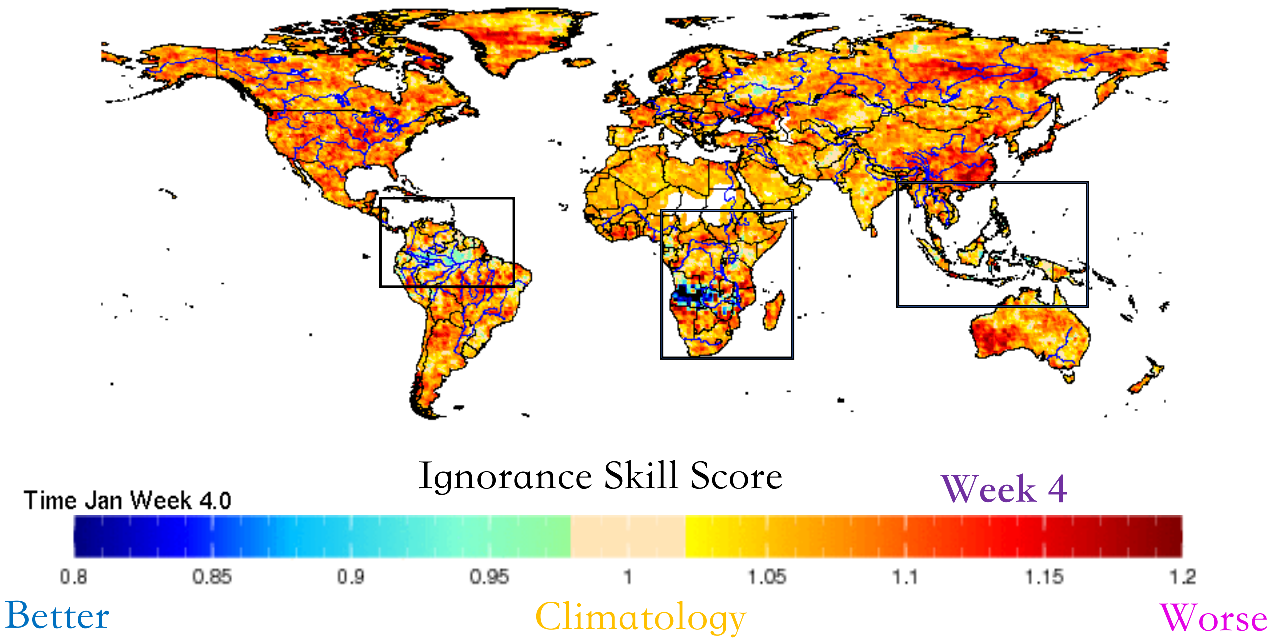


International Research Institute
for Climate and Society
Earth Institute | Columbia University

Muñoz et al. (2018)

Seasonality of Sub-seasonal Skill





International Research Institute for Climate and Society
EARTH INSTITUTE | COLUMBIA UNIVERSITY

Muñoz et al. (2018)

Outline

1. Introduction

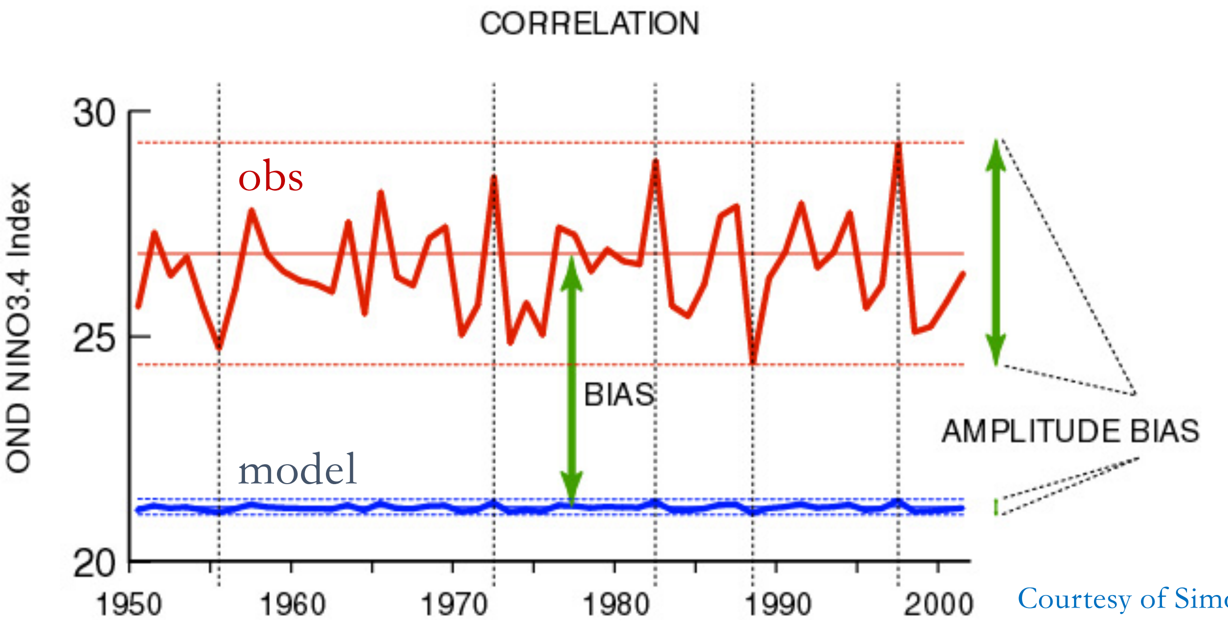
2. On the seasonality of sub-seasonal skill

3. Can we improve sub-seasonal skill with calibration?

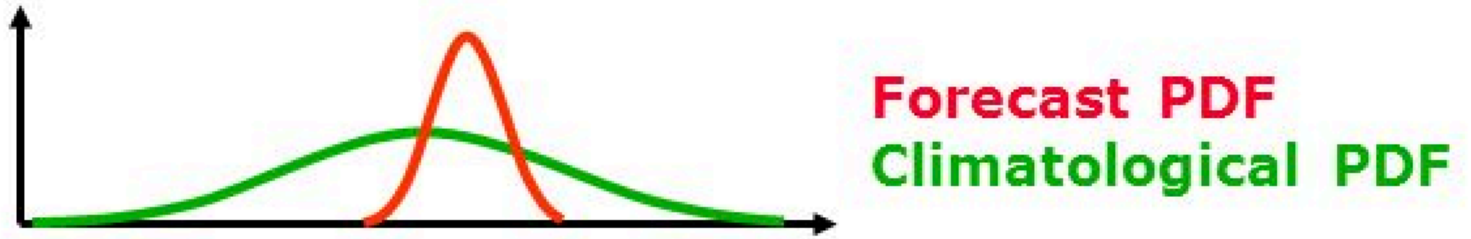
4. Concluding remarks and recommendations

Why do we need to calibrate?

- It is common to use Anomaly Correlation Coefficient to assess forecast skill, but it only measures *association.*
- There are a lot of other forecast attributes of interest!



Courtesy of Simon J. Mason



Courtesy of Renate Hagedorn

Local Calibration

- Grid-box-by-grid-box.
- Example: Extended Logistic Regression (XLR)

$$P(y < q \mid \mathbf{x}) = \frac{\exp(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta})}{1+\exp(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta})} = \Lambda(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}),$$

Logistic Regression

$$P[g(y) < g(Q) \mid \mathbf{x}] = \Lambda\left[\frac{g(Q)-\mu}{\sigma}\right]$$

Extended Logistic Regression

Homoscedastic XLR (Wilks, 2009)

$$\mu = -\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}/\alpha \qquad \sigma = 1/\alpha$$

Heteroscedastic XLR (Messner *et al*, 2014)

$$\mu = \mathbf{x}^{\mathrm{T}}\boldsymbol{\gamma} \qquad \sigma = \exp(\mathbf{z}^{\mathrm{T}}\boldsymbol{\delta}),$$

Particular case: HXLR becomes XLR

$$\mathbf{z} = 1 \qquad \alpha = 1/\exp(\boldsymbol{\delta}) \qquad \boldsymbol{\beta} = -\boldsymbol{\gamma}/\exp(\boldsymbol{\delta}).$$

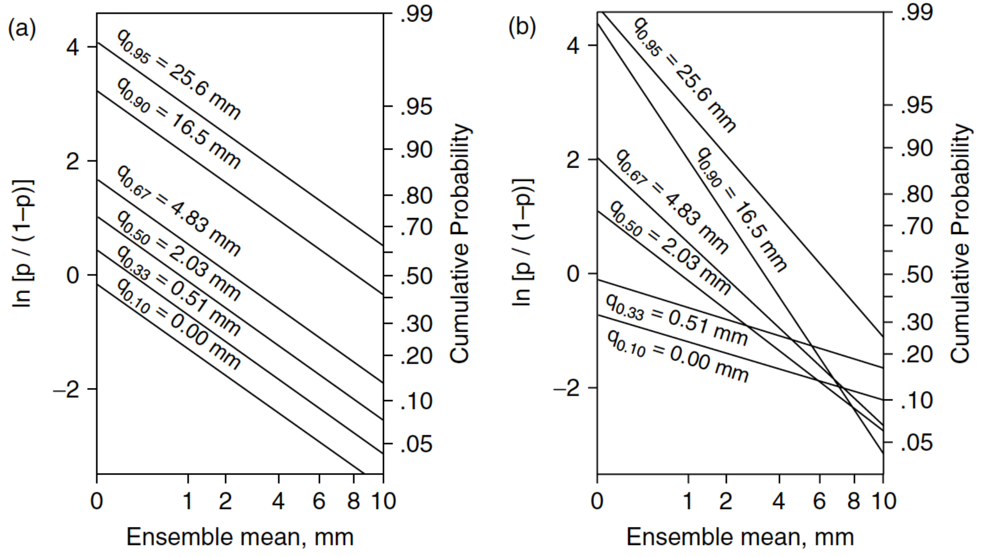


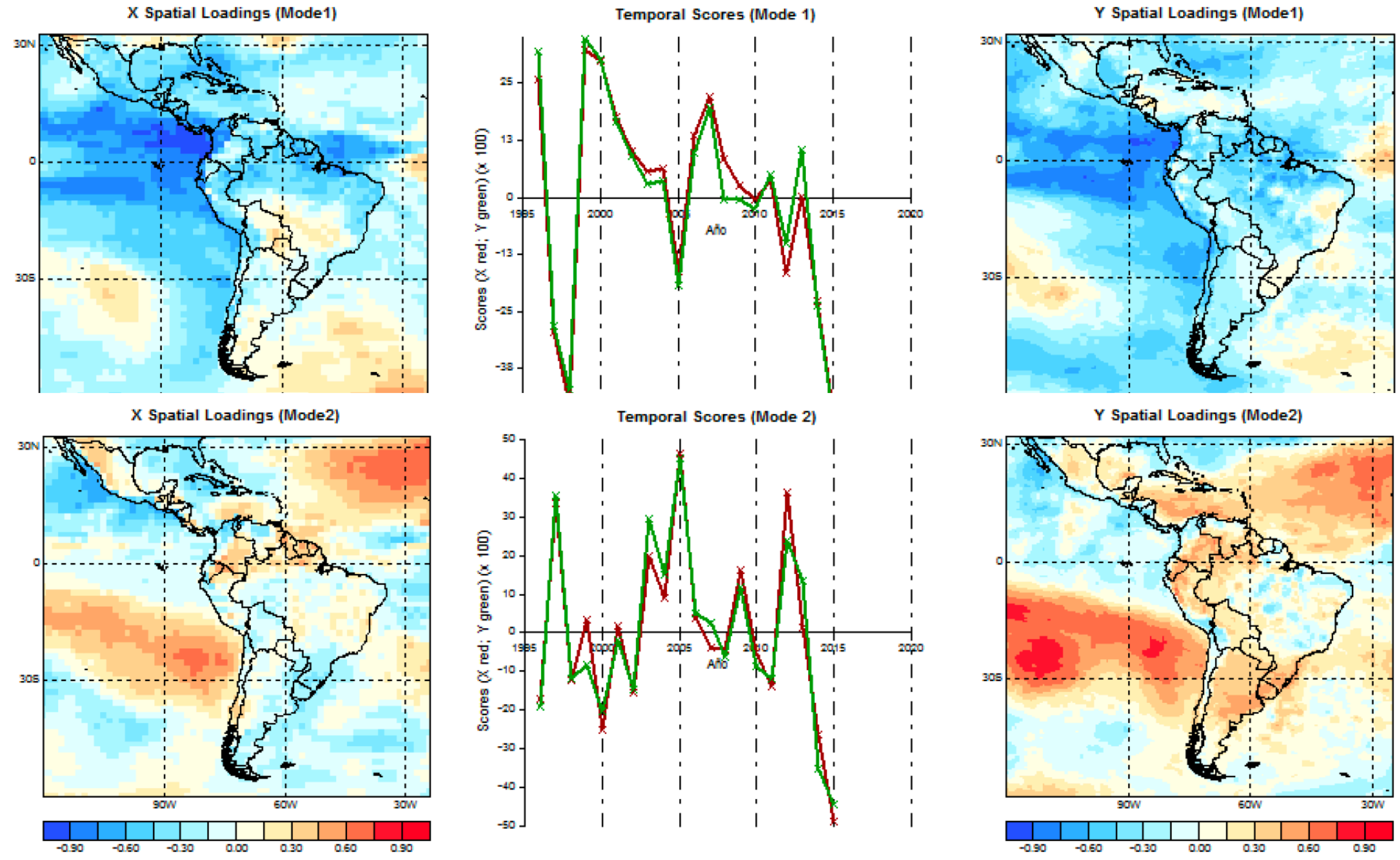
Figure 1. Logistic regressions plotted on the log-odds scale, for 28 November–2 December 2001, fitted using the full 25 year training length, for Minneapolis. Forecasts from Equation (6), evaluated at selected quantiles, are shown by the parallel lines in Figure 1(a), which cannot yield logically inconsistent sets of forecasts. Regressions for the same quantiles, fitted separately using Equation (3), are shown in Figure 1(b). Because these regressions are not constrained to be parallel, logically inconsistent forecasts are inevitable for sufficiently extreme values of the predictor.

Pattern-based Calibration (CCA)

Latin America & Caribbean T2M example

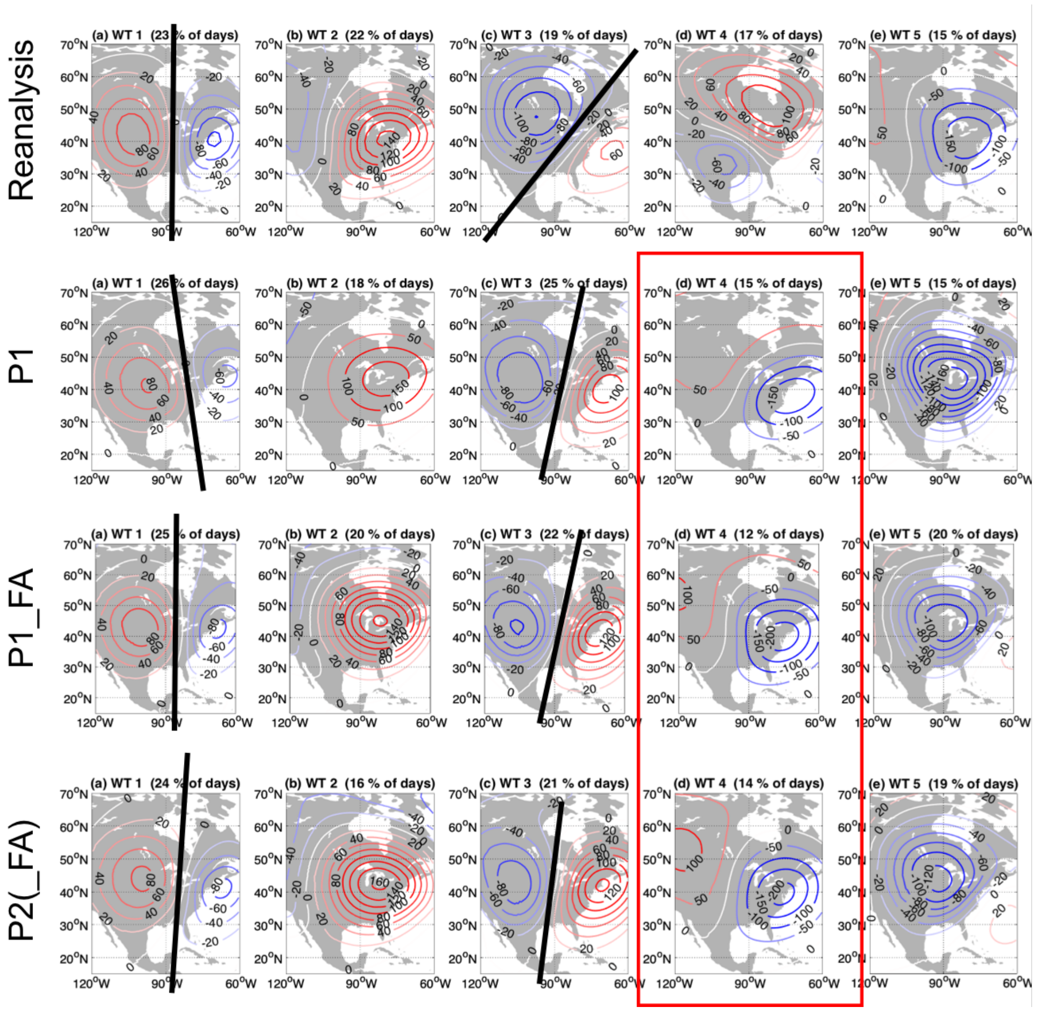
Model

Obs

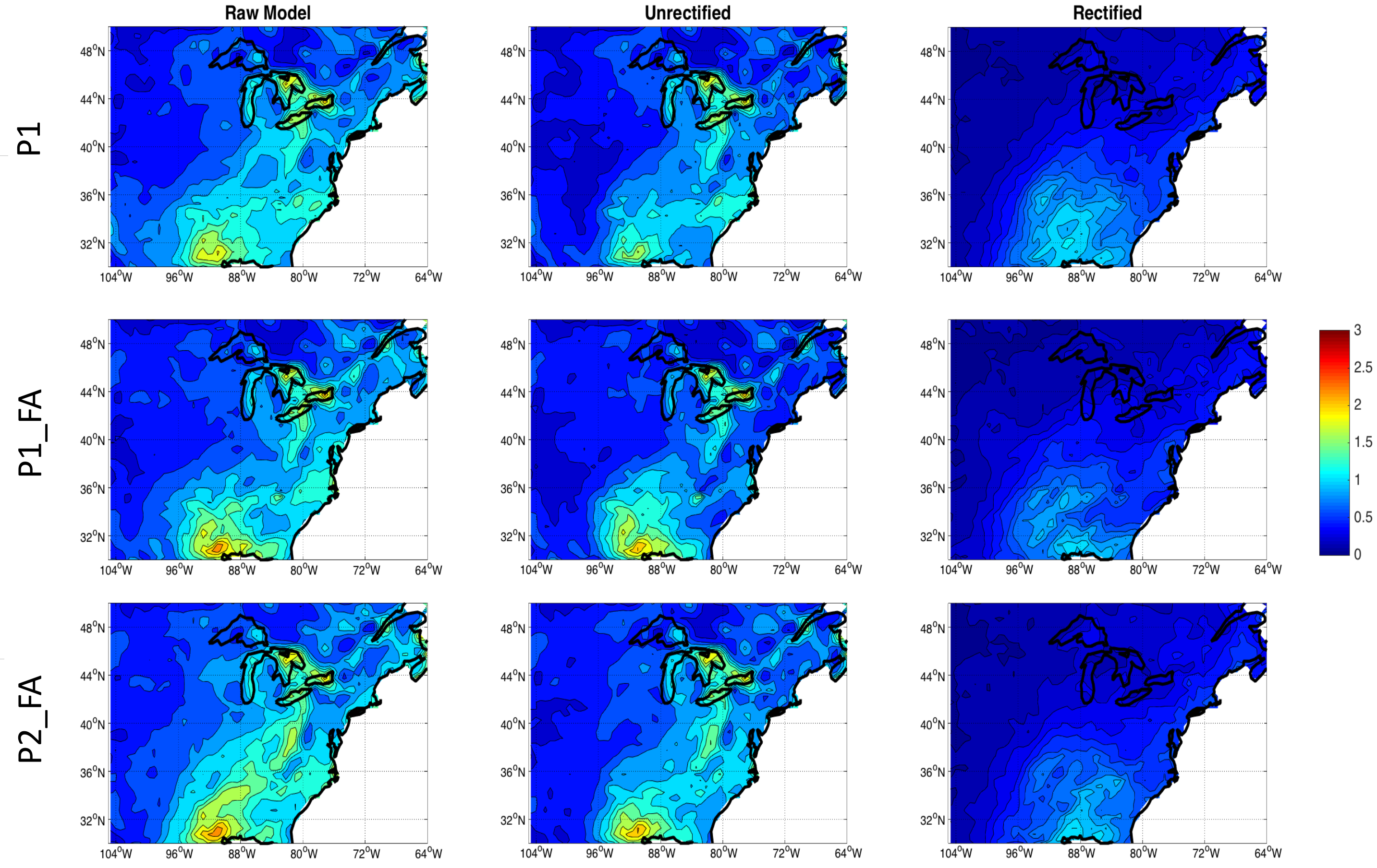


Pattern-based Calibration (WTs)

North America: z500 and rainfall examples (MAM)



RMSE (model rainfall vs obs [CPC unified])



International Research Institute for Climate and Society
Earth Institute | Columbia University

Muñoz et al (2017a,b)
Johnson et al (2018)
Muñoz et al (2020 [AGU])

Comparing calibration methods

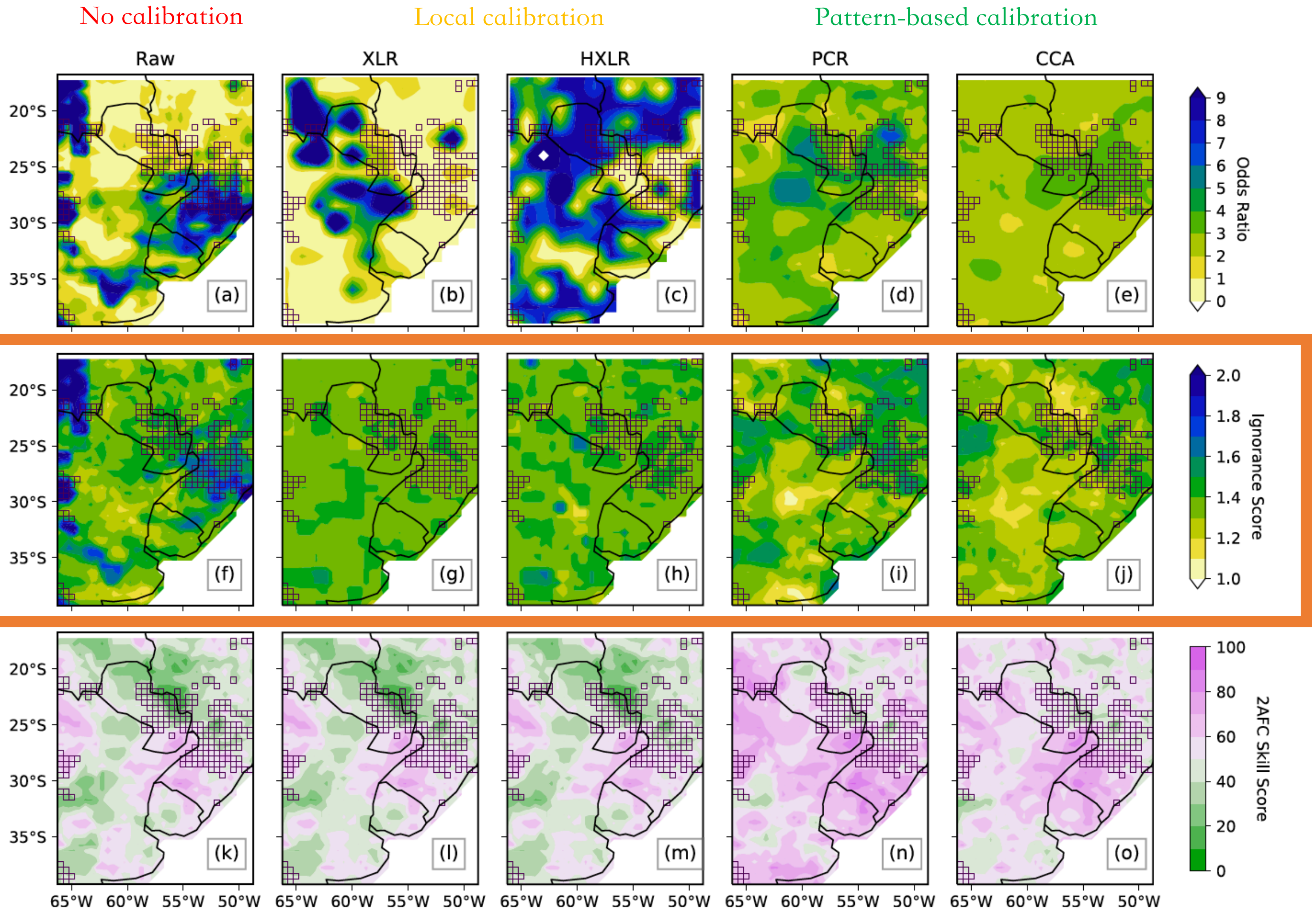


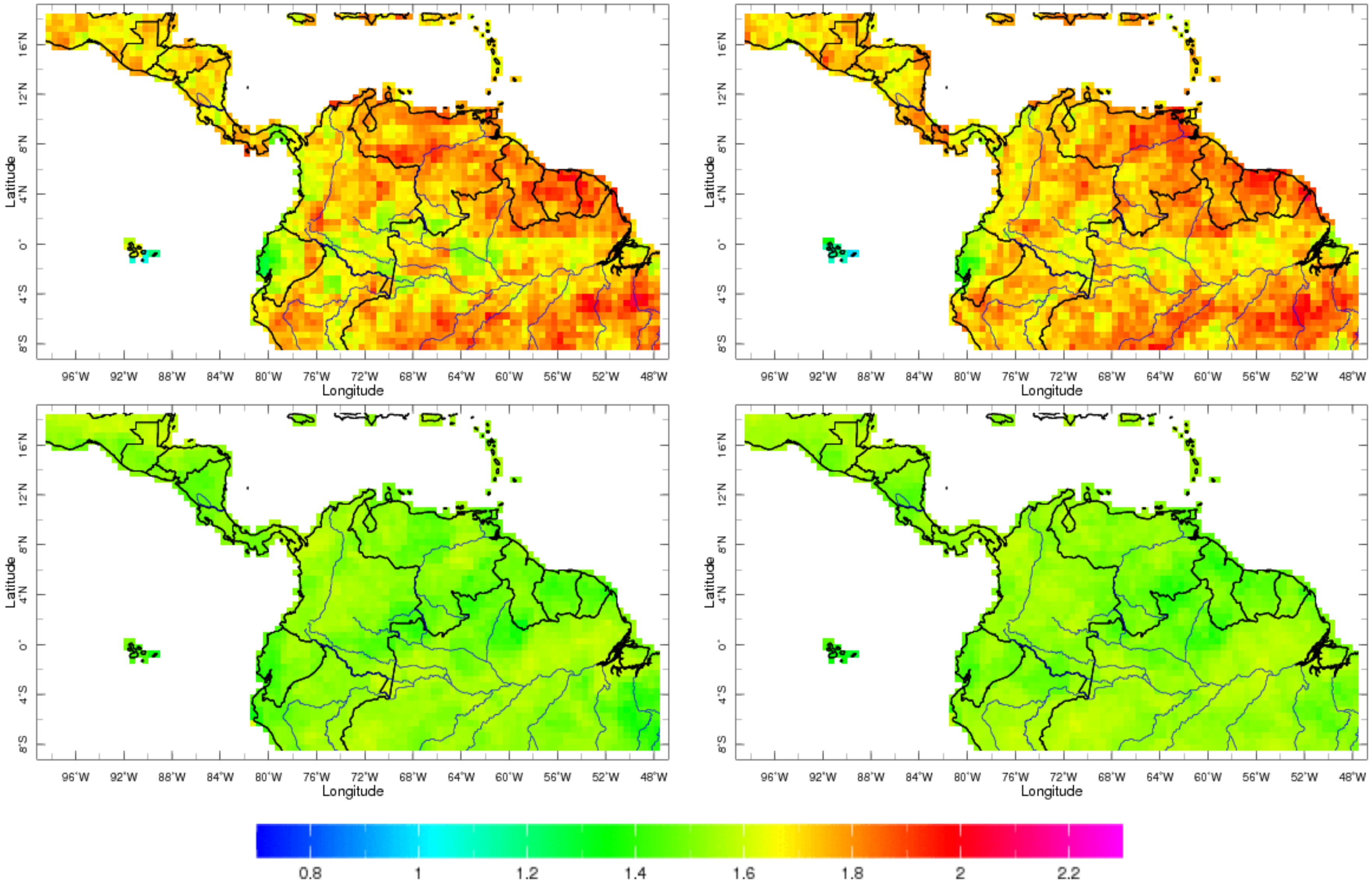
FIG. 10. Raw and MOS-adjusted S2S model forecasts and skill scores for the methods indicated in Table 1. (a)–(e) The heavy rainfall forecast for 1–7 Dec 2015 as odds, defined in Eq. (3) over the target domain. A value greater than 1 indicates that the model forecast greater-than-average odds of rainfall exceeding the 90th percentile. (f)–(j) The IGN defined in Eq. (4), with zero indicating a perfect forecast. (k)–(o) The 2AFC skill score for each grid cell; a value greater than 50 indicates that the model outperforms climatology. Different MOS models except for Raw in (a),(f),(k), which indicates the uncorrected S2S model output. In (top)–(bottom), the grid cells that observed a 90th percentile exceedance for 1–7 Dec 2015 are outlined in black.

Doss-Gollin et al (JClim 2018)

Initializations: April — Week 3 — Week 4

Uncalibrated

Calibrated (CCA)



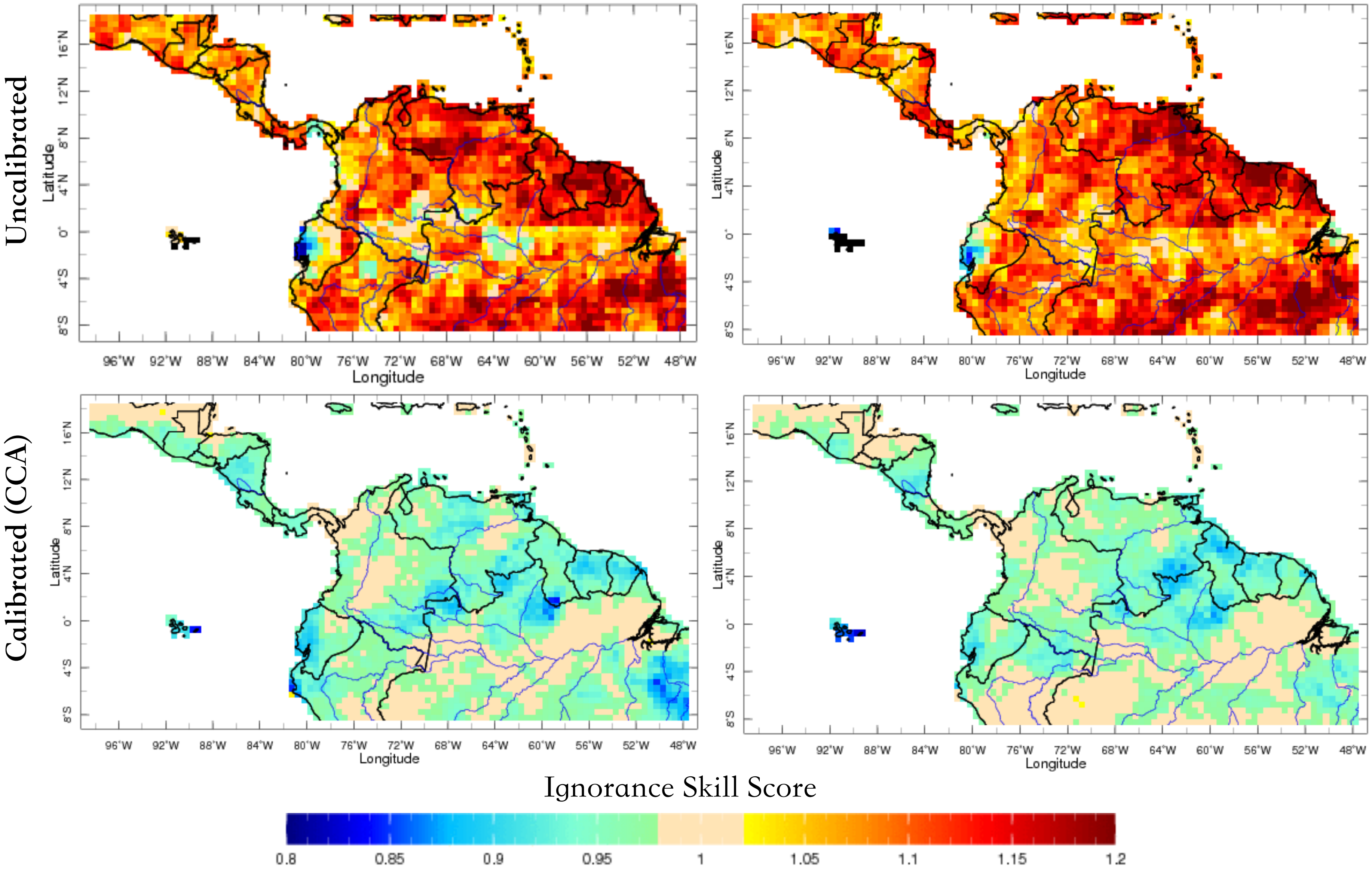
Better — Ignorance Score (bits) — Worse

International Research Institute for Climate and Society
EARTH INSTITUTE | COLUMBIA UNIVERSITY

Coelho and Muñoz (2019)
Muñoz et al. (in prep)

Initializations: April

Week 3 Week 4



Better Climatology Worse

$$IGN = REL - RES + UNC$$

calibration sharpness (if reliable) obs distribution

International Research Institute for Climate and Society
EARTH INSTITUTE | COLUMBIA UNIVERSITY

Coelho and Muñoz (2019 S2S Newsletter)
Muñoz et al. (in prep)

Outline

1. Introduction
2. On the seasonality of sub-seasonal skill
3. Can we improve sub-seasonal skill with calibration?
4. Concluding remarks and recommendations

Concluding Remarks

- Sub-seasonal skill – as measured by the Ignorance Score and 2AFC – at *regional* scale tends to exhibit seasonality. *Global* sub-seasonal skill varies less along the year.

- Generally speaking, ***uncalibrated*** sub-seasonal forecast skill is worse than climatology after Week 2. There are some exceptions: Tropical South America, Eastern and Southern Africa, Maritime Continent.

- Model Output Statistics has the **potential to improve forecast skill at sub-seasonal timescales**. In particular, EOF-based MOS methods like Canonical Correlation Analysis (and Principal Component Regression), available in (Py)CPT, tend to offer skill improvement in different regions of the world, especially in terms of reliability and sharpness.

- The approach discussed in this talk can be used to identify and quantify model biases and their impact in different variables and regions.

- Work in progress – stay tuned.

Sub-seasonal Forecast Skill: When, Where and How to Find It?

Ángel G. Muñoz

agmunoz@iri.columbia.edu

International Research Institute for Climate and Society
Earth Institute | Columbia University

2020IVMWO, 18 Nov 2020

Note that pattern-based calibration (e.g., CCA) does not necessarily increase forecast skill everywhere!

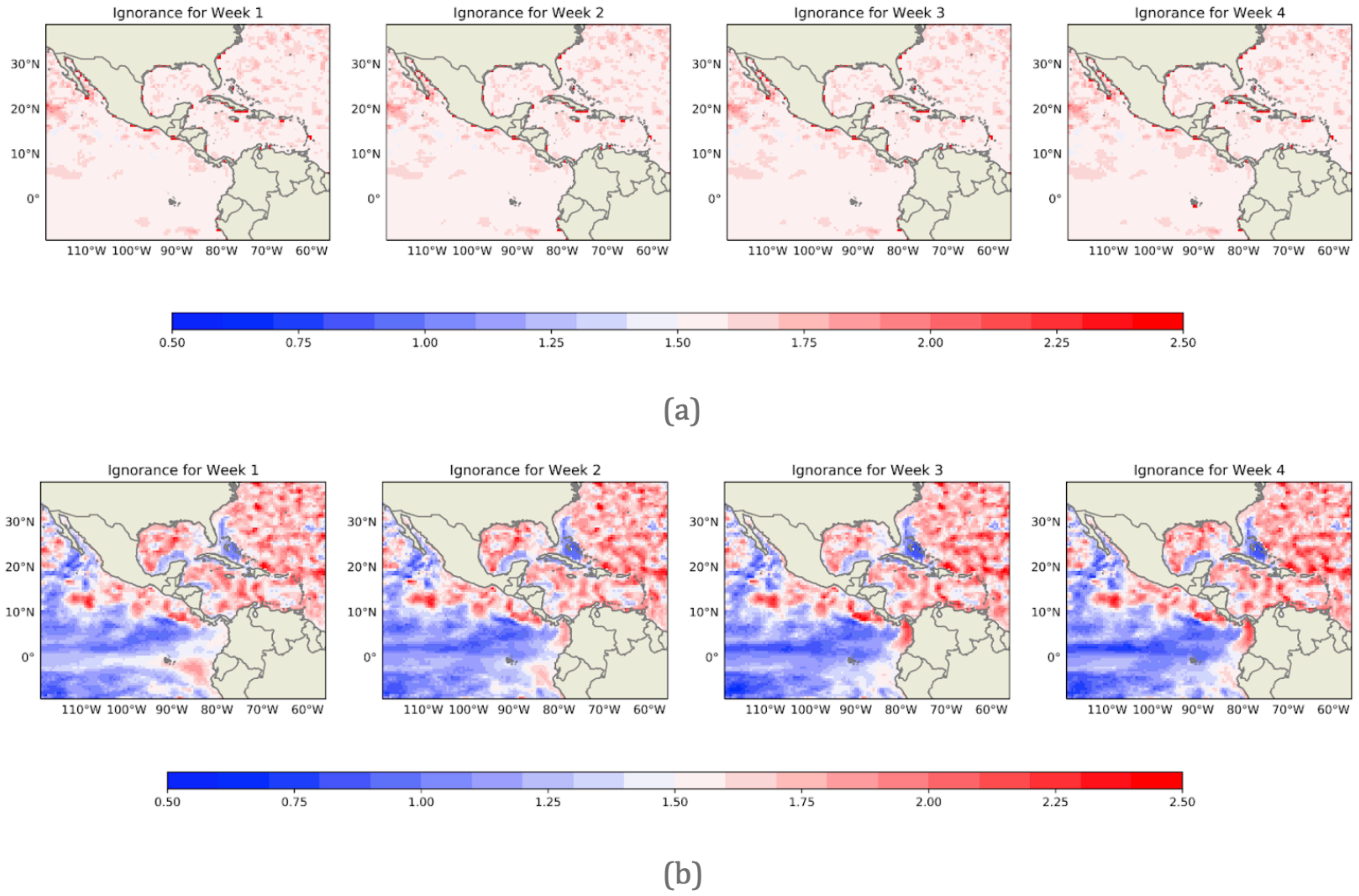


Figure 11. SSH forecast ignorance of (a) uncalibrated and (b) calibrated (CCA) forecasts initialized in March

Campos et al (in prep)