

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

Ángel G. Muñoz

agmunoz@iri.columbia.edu

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

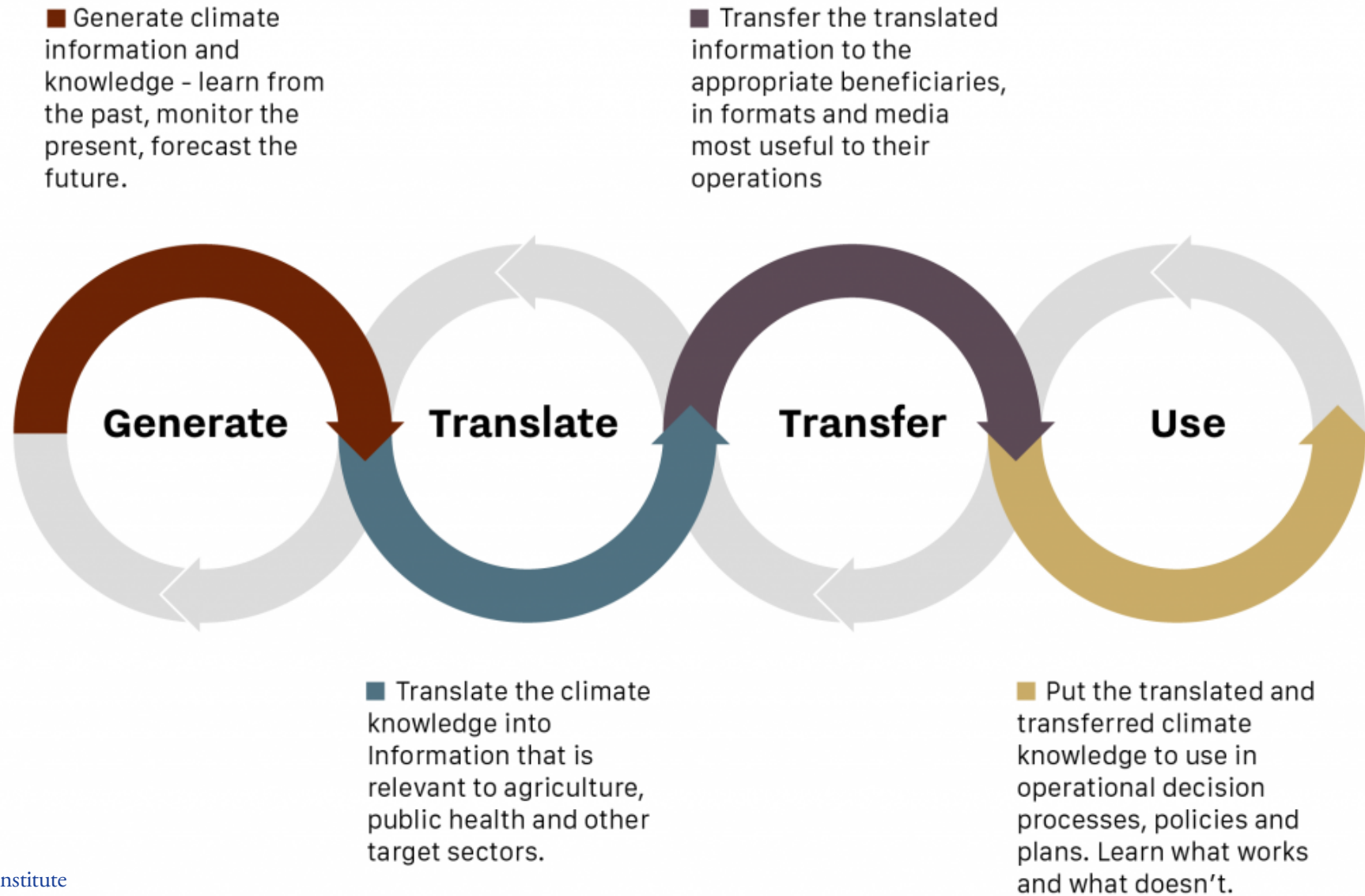


Outline



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



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):

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

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

$$IGN = \underbrace{REL}_{\text{calibration}} - \underbrace{RES}_{\text{sharpness (if reliable)}} + \underbrace{UNC}_{\text{obs distribution}} \quad (\text{Weijjs et al., 2010; Wilks, 2018})$$

- 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):

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

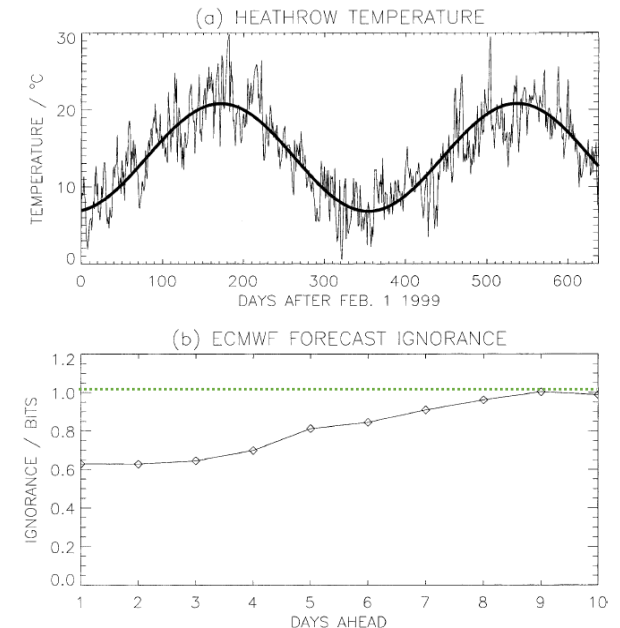
and it can be decomposed into reliability, resolution and uncertainty terms:

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

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

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

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



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

PyCPT

Py IRI / PyCPT Source
Clone

IRI PyCPT

Python interface and enhancement for the command line version of the International Research Institute for Climate and Society's Climate Predictability Tool (CPT), for seasonal and sub-seasonal skill assessment and forecast experiments

master ▼
 Filter files 🔍

Name	Size	Last commit	Message
.github		2020-01-04	Update issue templates
figures		2020-01-04	Some housekeeping
old_versions		6 days ago	Housekeeping
CODE_OF_CONDUCT.md	3.28 KB	2020-01-04	Create CODE_OF_CONDUCT.md
LICENSE.md	34.33 KB	2020-01-03	Create LICENSE.md
PyCPT_s2sv1.9.ipynb	2.39 MB	6 days ago	Housekeeping
PyCPT_seav1.9.ipynb	1.67 MB	6 days ago	Housekeeping
README.md	3.44 KB	2020-04-02	Fixed link to the Wiki to point to bitbucket instead of Git...
iri-dev.yml	4.46 KB	2020-01-04	Adding iri-dev environment for conda
pycpt_dictionary.py	152.68 KB	5 days ago	Small change to GEFS/CFSv2 entries.
pycpt_functions.py	202.04 KB	2020-10-19	Fixed retroactive bug

- **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**.



Funded by:

NOAA NA16OAR4310145 (Robertson)

NOAA NA18OAR4310275 (Muñoz)

Columbia World Project “ACToday” (Goddard)

Jupyter

PyCPT_s2sv1.9

Last Checkpoint: 13 minutes ago (autosaved)

Logout

FileEditViewInsertCellKernelWidgetsHelp

Not TrustedPython 3

+

⌕

📄

⬆️

⬇️

▶️ Run

⏏


↺

▶️▶️

Markdown

⌵

🗨️



πCPT

s2sv1.9 (9 Oct 2020)

Produce S2S (un)calibrated forecasts and assess associated skill

Requires:

CPTv16.5.2+

Authors:

Authors: ÁG Muñoz (agmunoz@iri.columbia.edu), AW Robertson (awr@iri.columbia.edu), Cuihua Li (OC), Bohar Singh (IRI), T Turkington (NEA), SJ Mason (IRI)

Acknowledgements:

Rémi Cousin (IRI, key Ingrid code), James Doss-Gollin (Columbia Water Center, original download Python functions). Part of the effort to develop PyCPT is funded by the Columbia World Project "Adapting Agriculture to Climate Today, for Tomorrow" (ACToday), and NOAA MAPP's projects NA18OAR4310275 (Muñoz) and NA16OAR4310145 (Robertson).

Version log

See version.log in Github

Workflow

Case Configuration (namelist)

Download data (if necessary) and execute CPT

Forecast Skill Assessment

Forecast Production (Flexible format)

JupyterLab interface showing the execution of a Python script (PyCPT_s2sv1.9) with a last checkpoint 14 minutes ago. The interface includes a menu bar (File, Edit, View, Insert, Cell, Kernel, Widgets, Help), a toolbar, and a status bar (Not Trusted, Python 3).

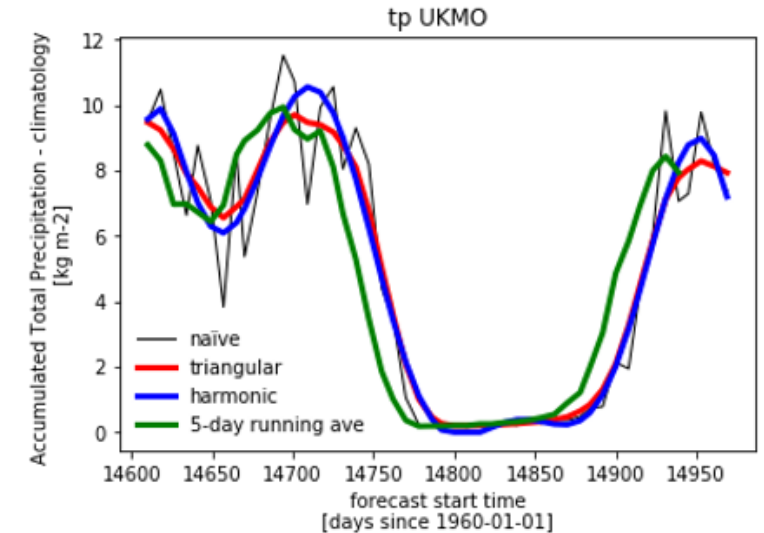
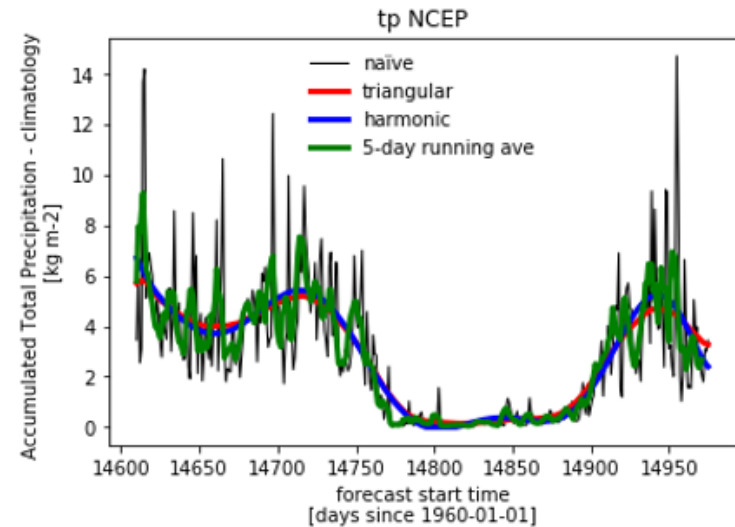
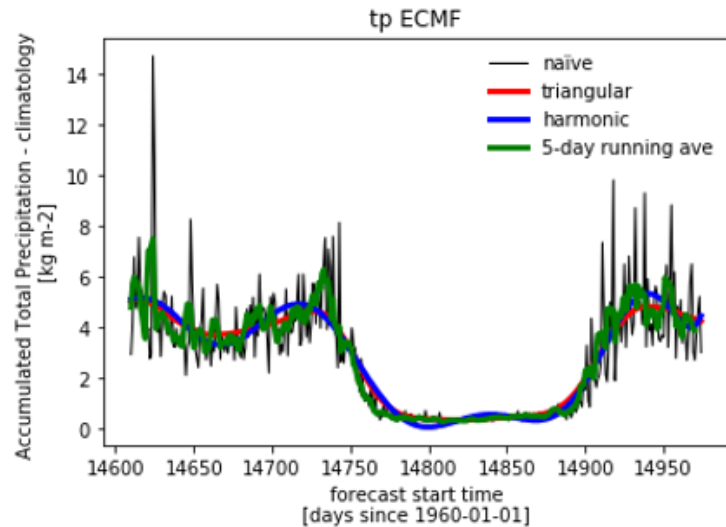
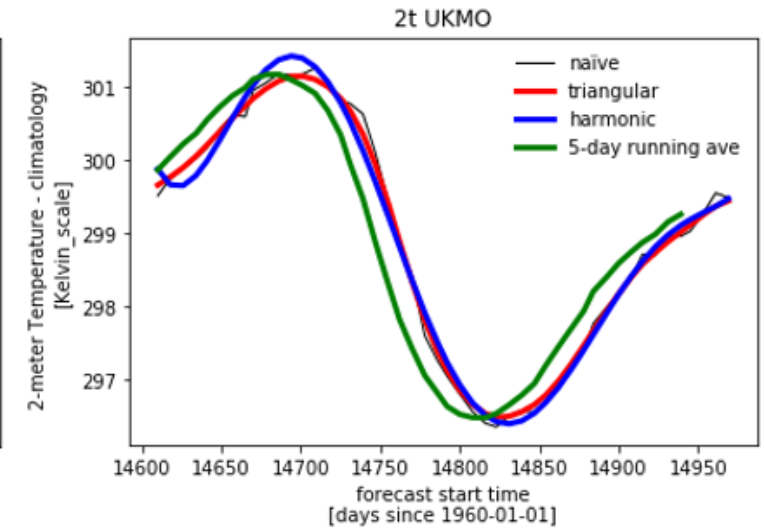
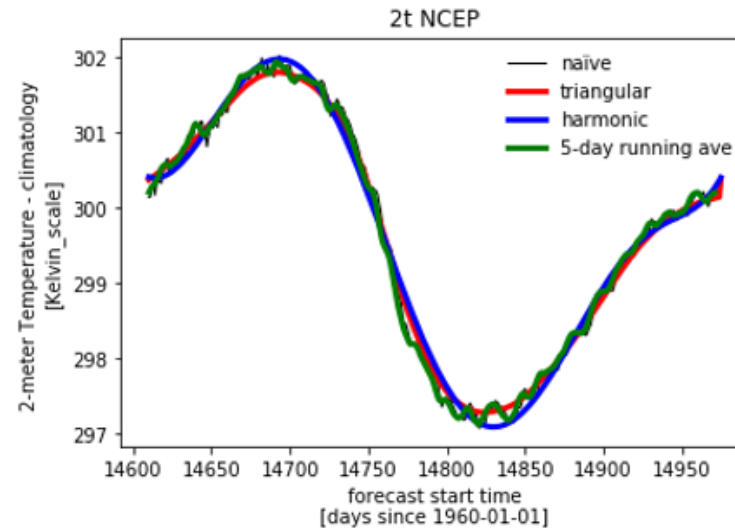
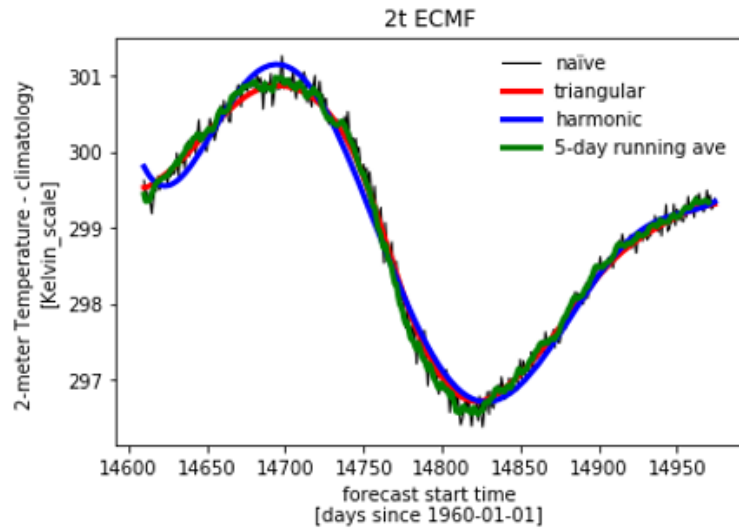
The figure displays three rows of maps, each showing the spatial distribution of a different metric over a geographic area (likely the Pacific Ocean region, bounded by 15°N to 6°S and 96°W to 54°W). The maps are arranged in a 3x4 grid, with columns representing different weeks (Week 1, Week 2, Week 3, Week 4) and rows representing different metrics (Spearman, 2AFC, RocAbove).

Spearman Correlation: The top row shows Spearman correlation maps for Weeks 1, 2, 3, and 4. A color bar below indicates the correlation scale from -1.00 (blue) to 1.00 (red), with intermediate values at -0.75, -0.50, -0.25, 0.00, 0.25, 0.50, and 0.75.

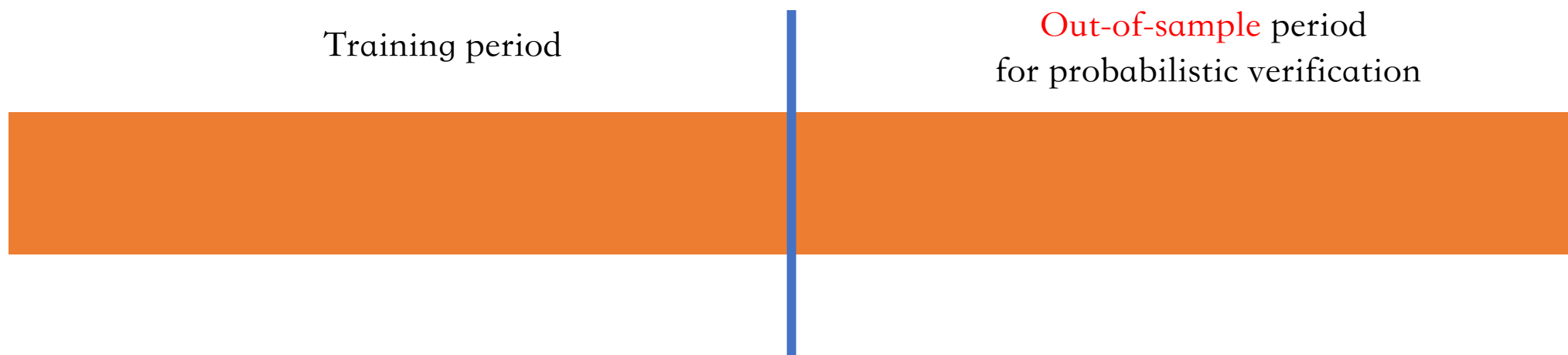
2AFC (%): The middle row shows 2AFC (%) maps for Weeks 1, 2, 3, and 4. A color bar below indicates the percentage scale from 0 (blue) to 100 (red), with intermediate values at 20, 40, 60, and 80.

RocAbove: The bottom row shows RocAbove maps for Weeks 1, 2, 3, and 4. The maps are displayed but lack a corresponding color bar.

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

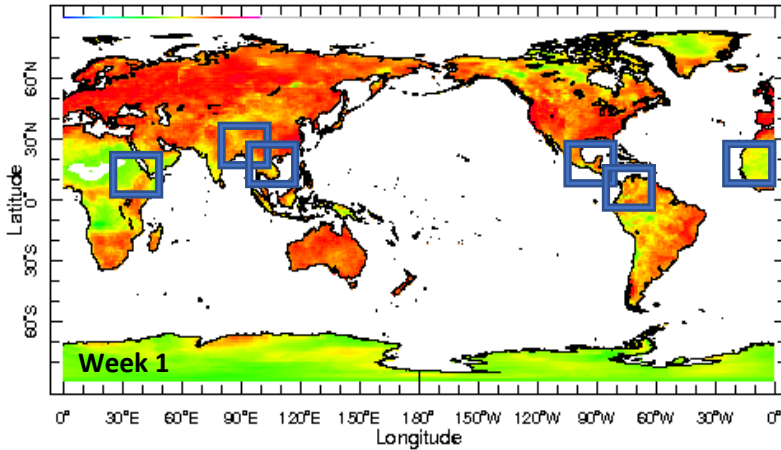
Half of the period is used for the probabilistic forecast verification

Outline

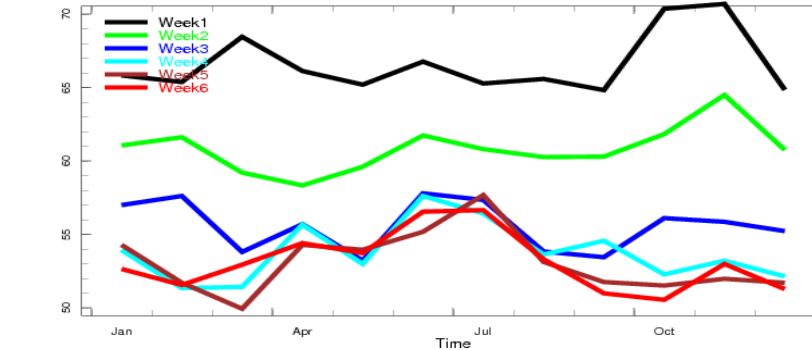


1. Introduction
2. On the seasonality of sub-seasonal skill
3. Can we improve sub-seasonal skill with calibration?
4. Concluding remarks and recommendations

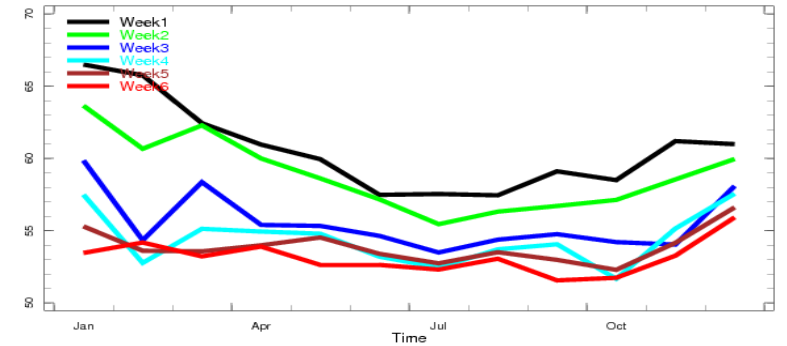
The Seasonality of Sub-Seasonal Skill



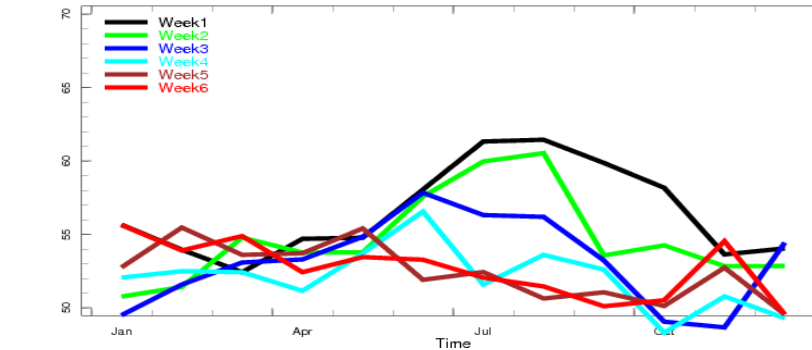
An “Obscurial”, or just subseasonal forecast skill?



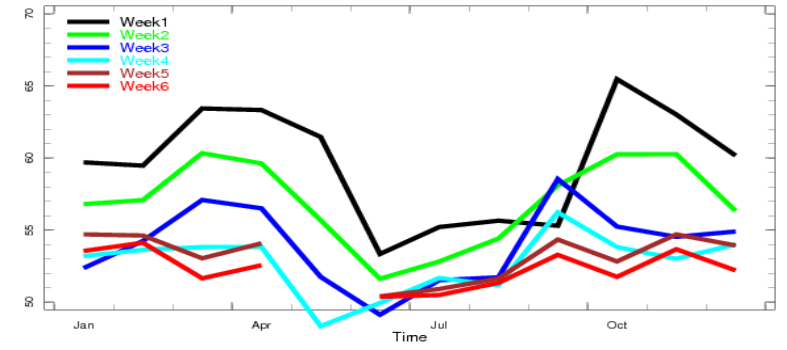
Guatemala - Deterministic 2AFC - Model: ECMWF (uncorrected)



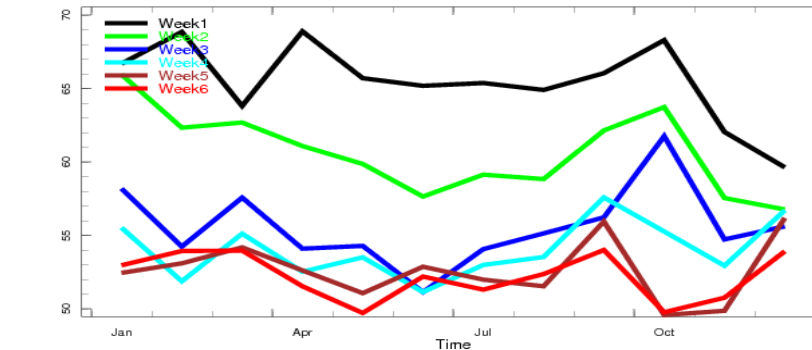
Continental Colombia - Deterministic 2AFC - Model: ECMWF (uncorrected)



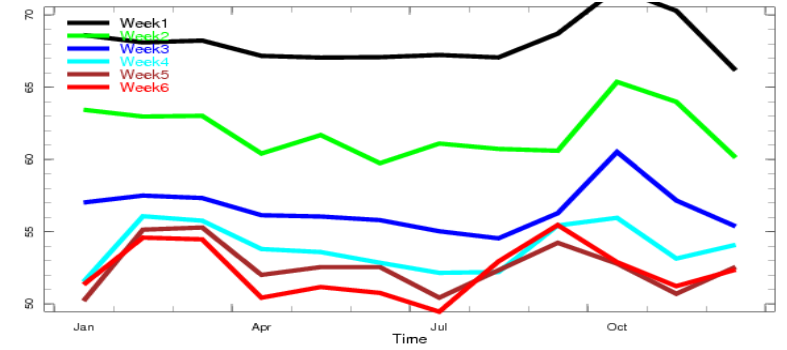
Senegal - Deterministic 2AFC - Model: ECMWF (uncorrected)



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



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

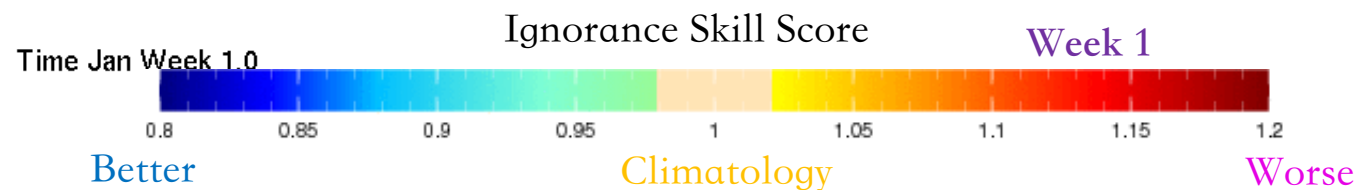
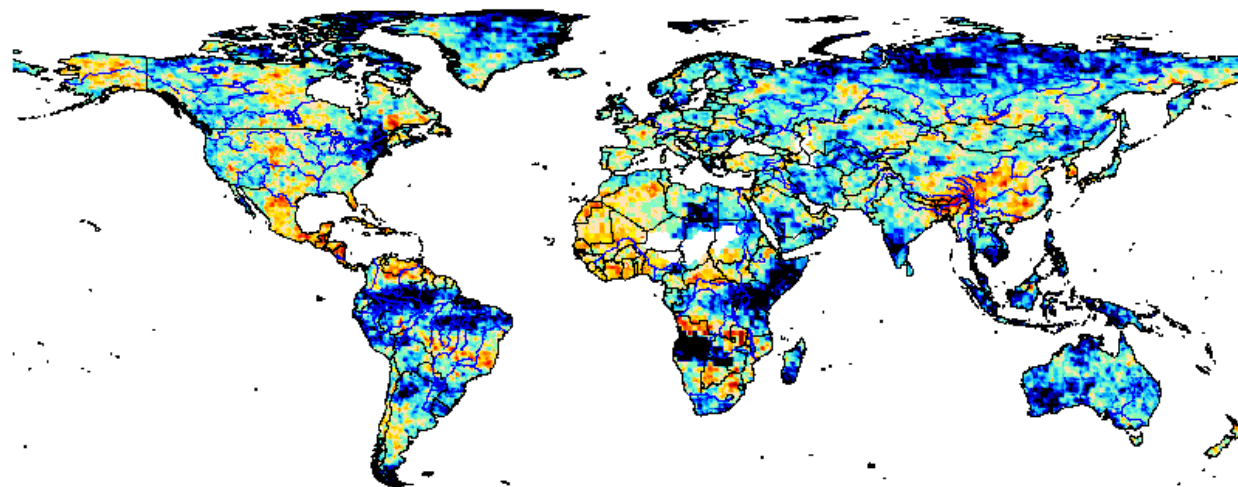
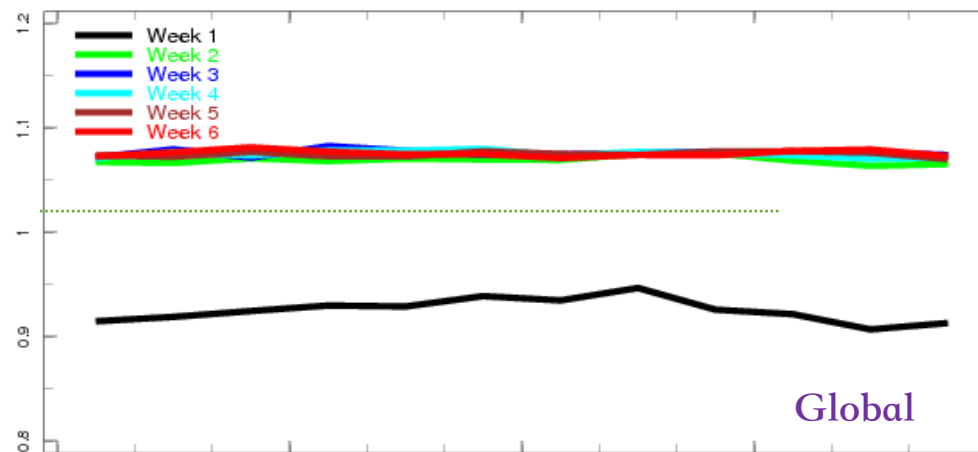


Vietnam - Deterministic 2AFC - Model: ECMWF (uncorrected)

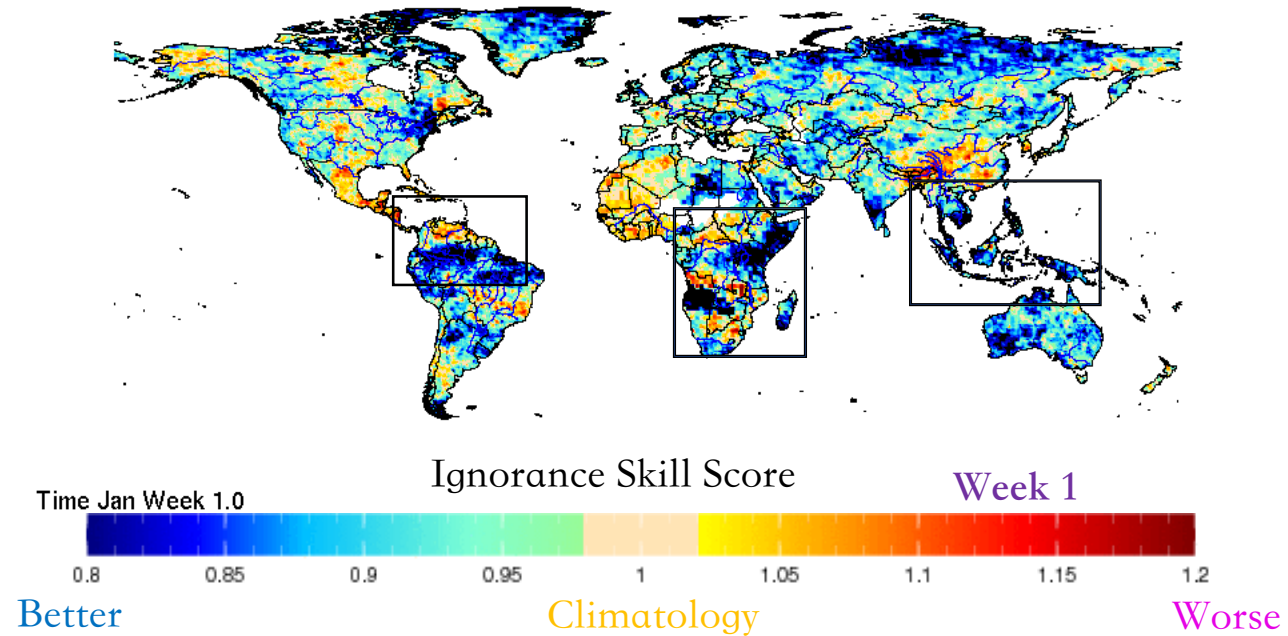
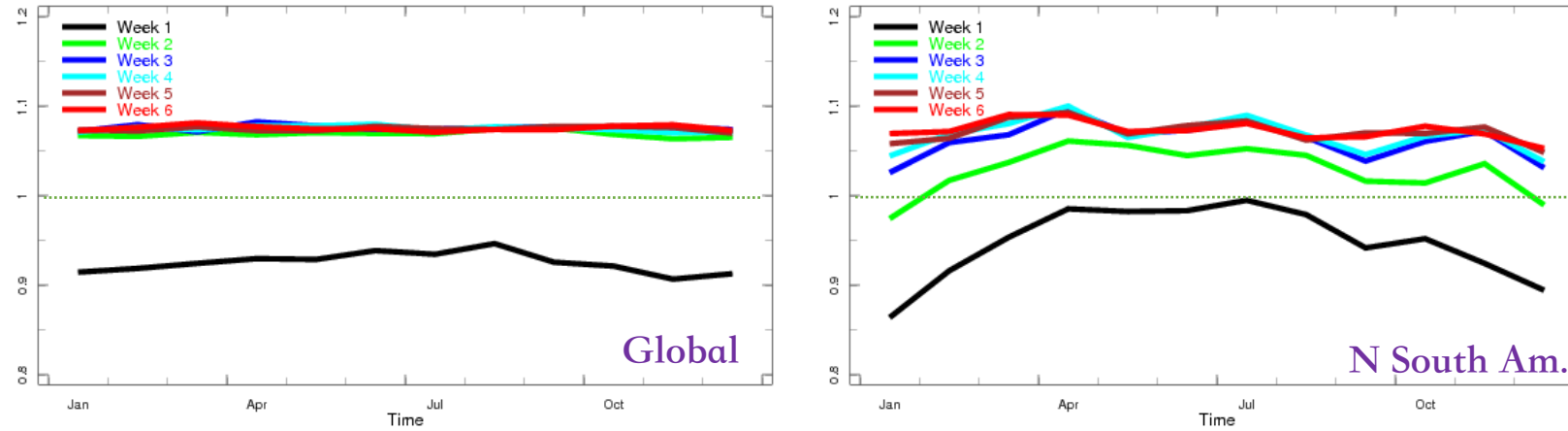
Seasonality of Sub-seasonal Skill

Skill Assessment

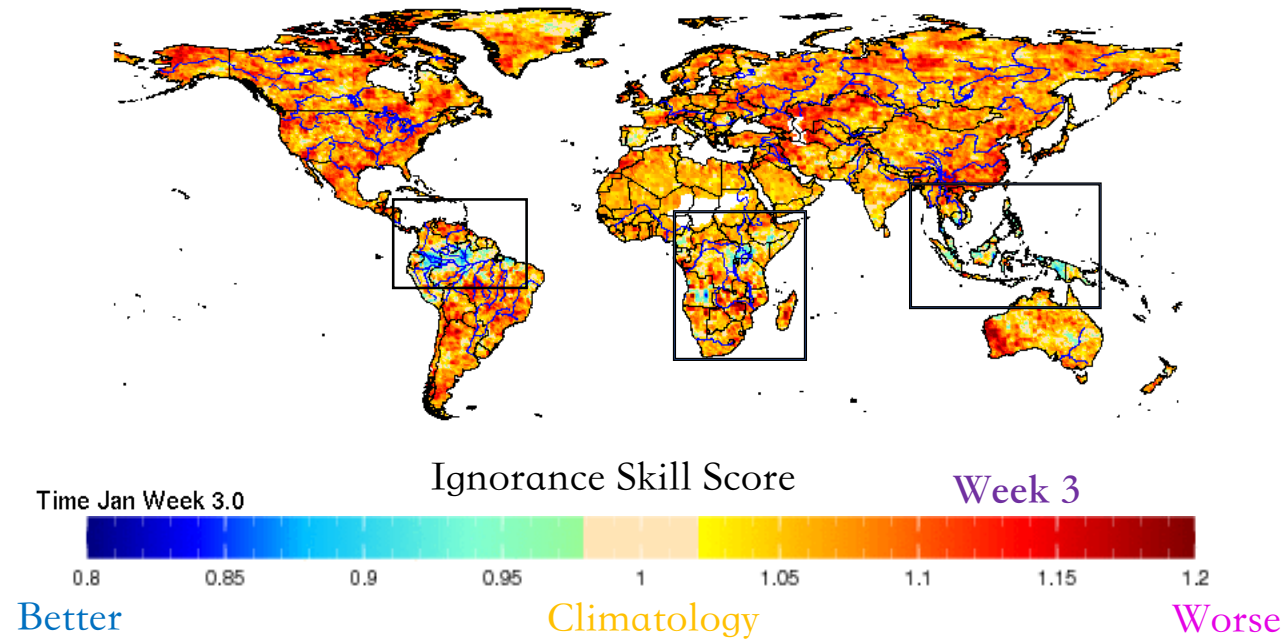
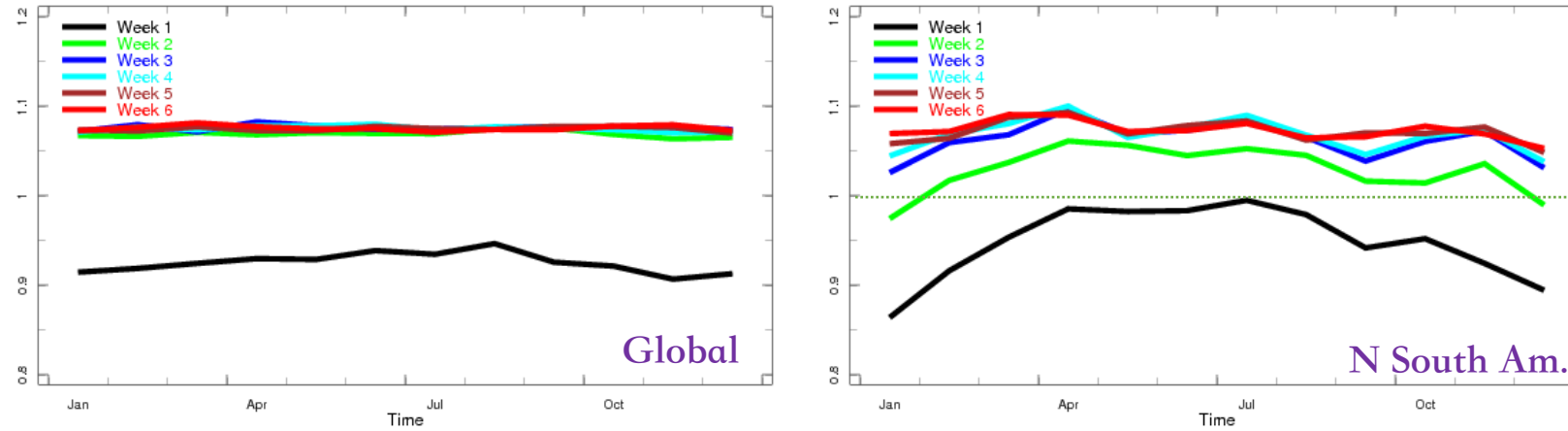
- Model: ECMWF
- Rainfall
- Probabilistic
- Hindcasts
- Obs: CPC Unified
- All initializations available per month
- Uncalibrated
- IGN, RPSS, Brier and decompositions, for Week 1-6



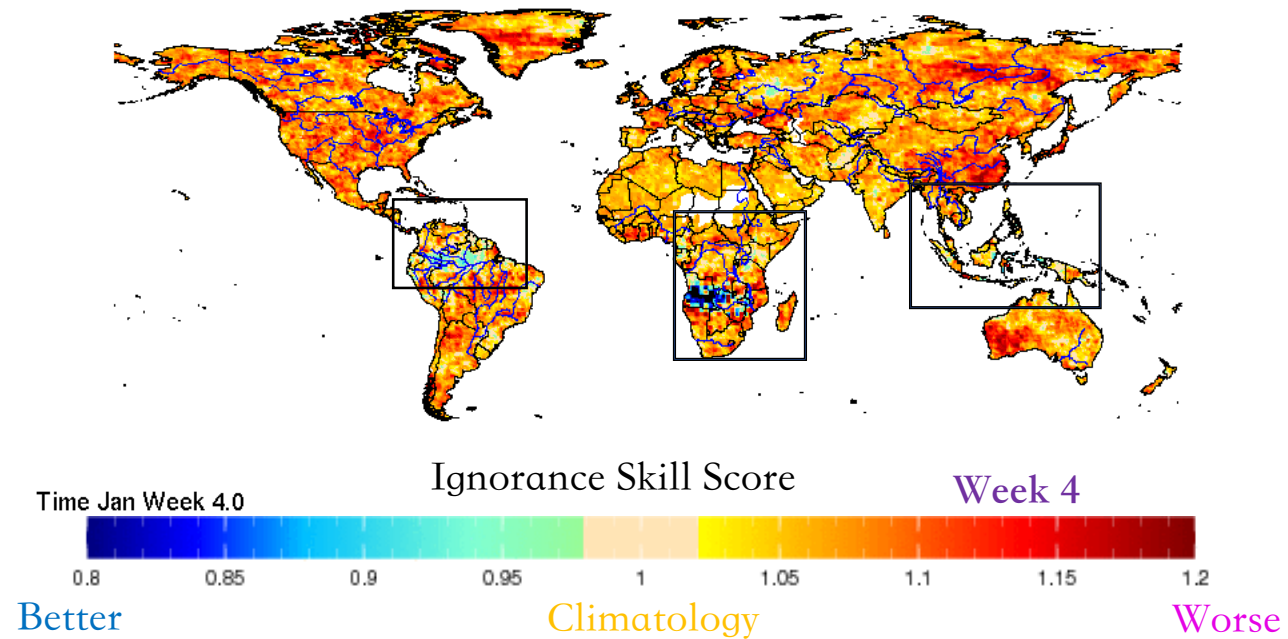
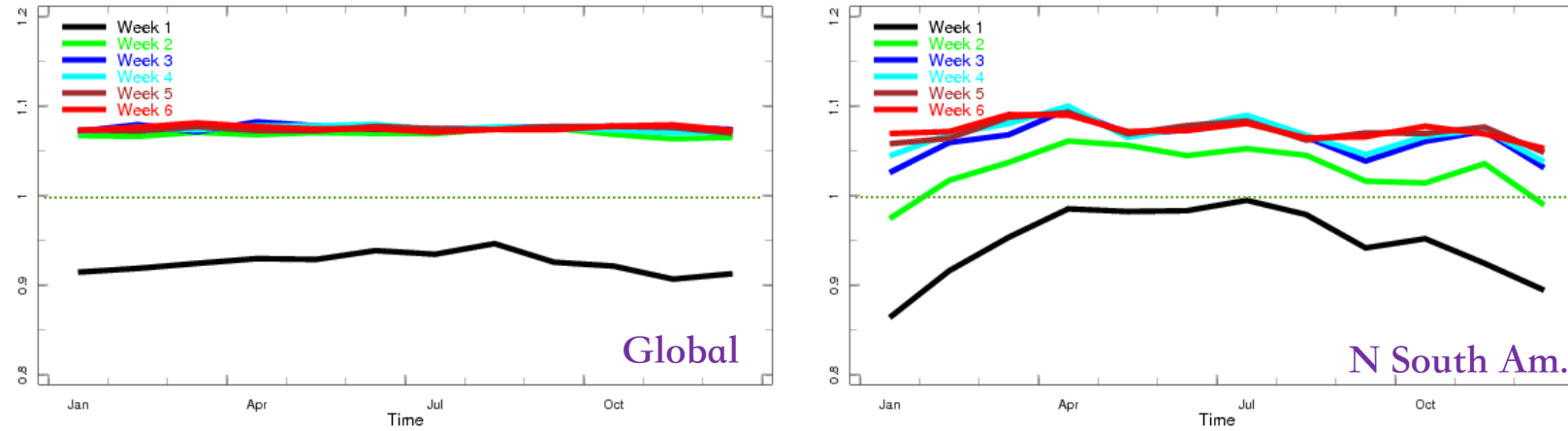
Seasonality of Sub-seasonal Skill



Seasonality of Sub-seasonal Skill



Seasonality of Sub-seasonal Skill



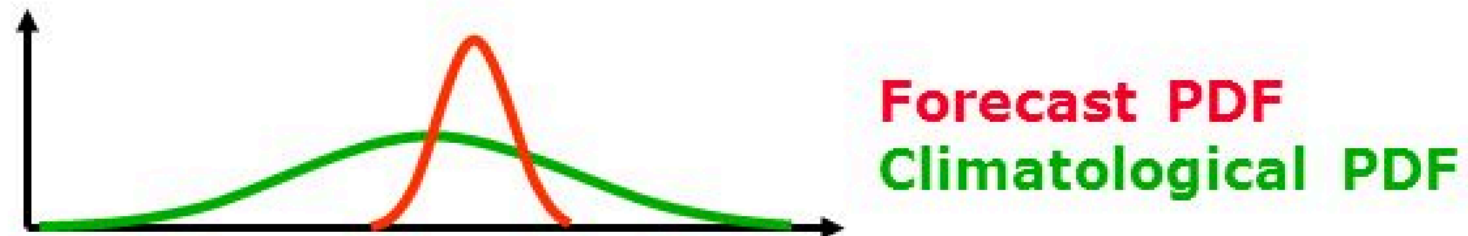
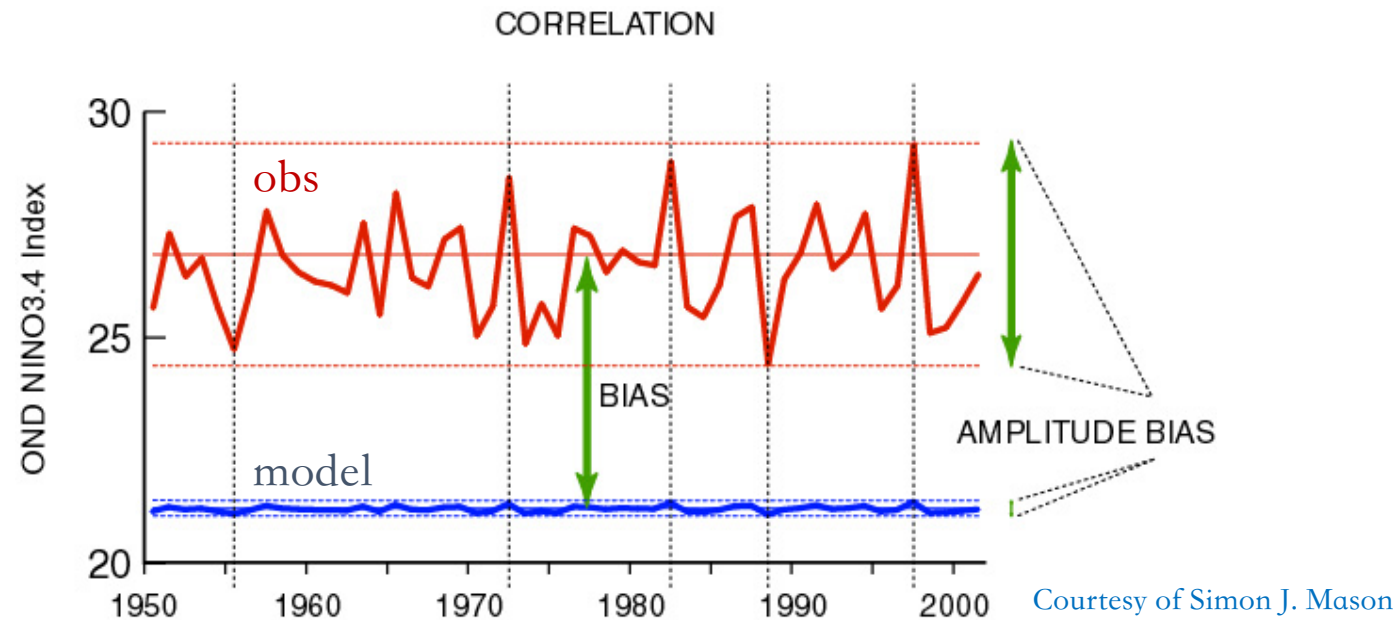
Outline



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!



Local Calibration

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

Homoscedastic XLR (Wilks, 2009)

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

Heteroscedastic XLR (Messner *et al*, 2014)

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

Particular case: HXLR becomes XLR

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

$$P(y < q | \mathbf{x}) = \frac{\exp(\mathbf{x}^T \boldsymbol{\beta})}{1 + \exp(\mathbf{x}^T \boldsymbol{\beta})} = \Lambda(\mathbf{x}^T \boldsymbol{\beta}), \quad \text{Logistic Regression}$$

$$P[g(y) < g(Q) | \mathbf{x}] = \Lambda \left[\frac{g(Q) - \mu}{\sigma} \right] \quad \text{Extended Logistic Regression}$$

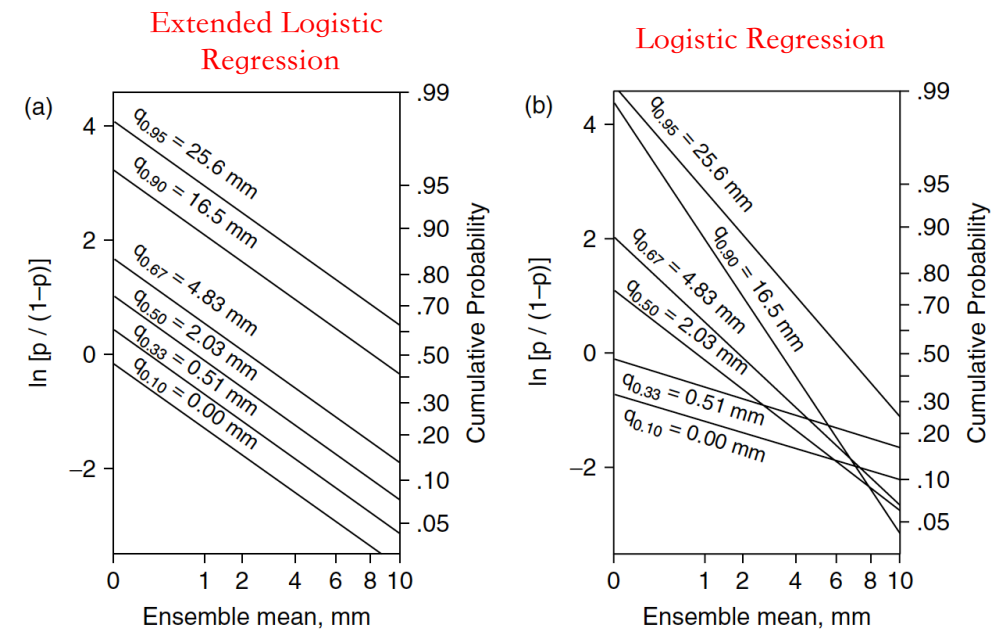


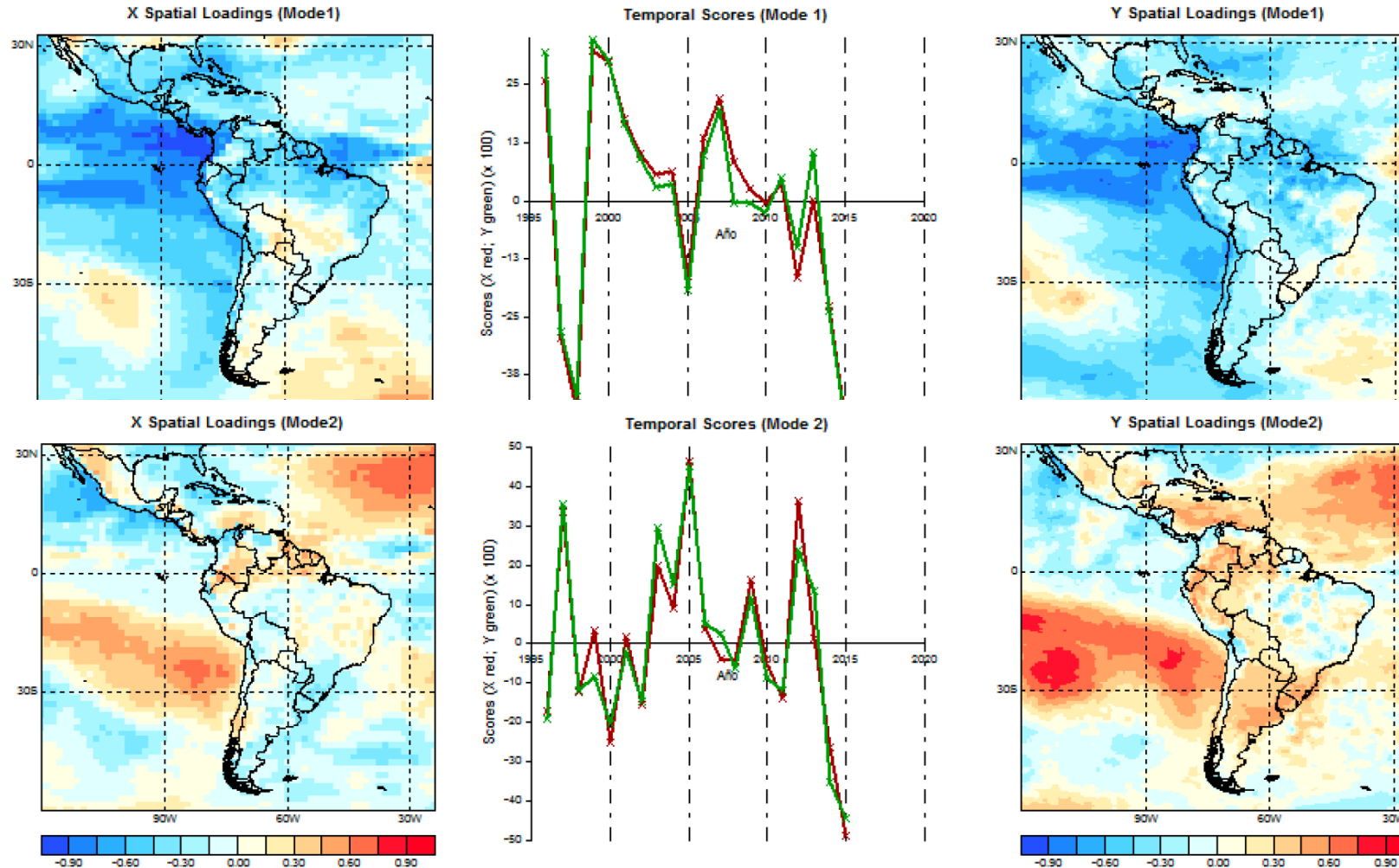
Figure 1. Logistic regressions plotted on the log-odds scale, for 28 November–2 December 2001, fitted using the full 25 year training length, for Minneapolis. Forecasts from Equation (6), evaluated at selected quantiles, are shown by the parallel lines in Figure 1(a), which cannot yield logically inconsistent 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

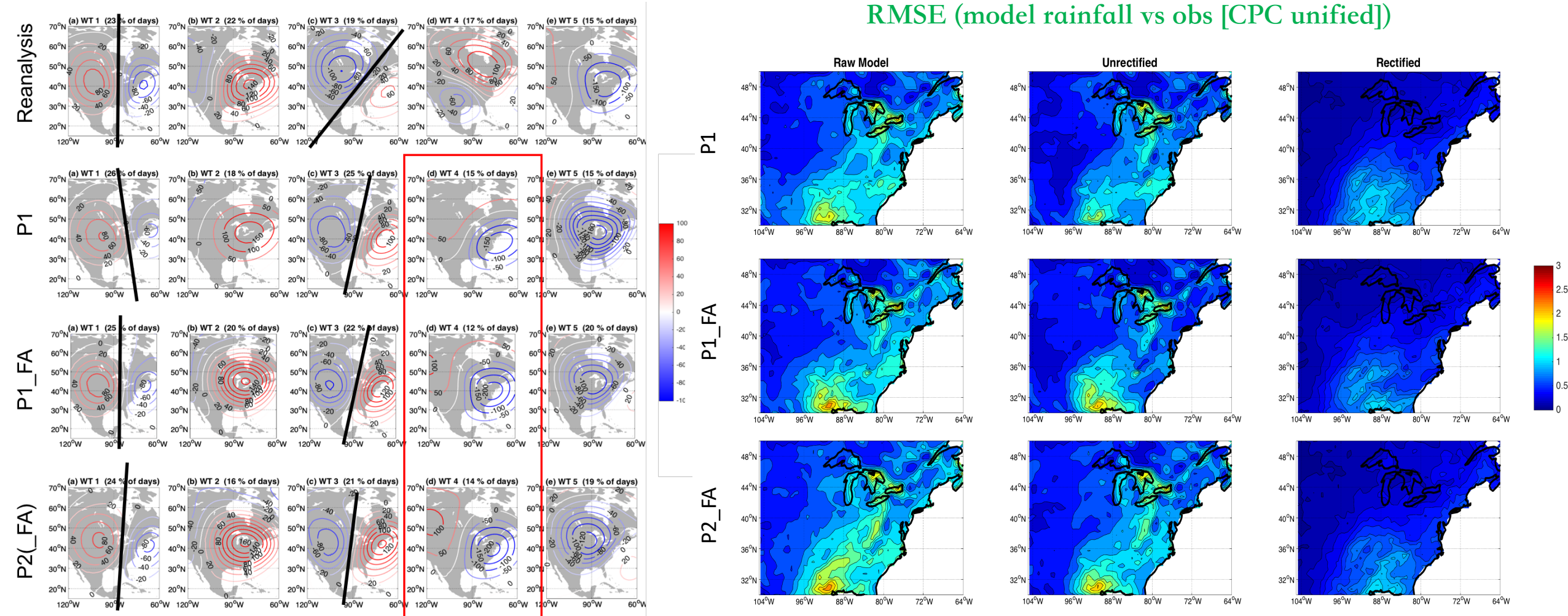
Obs



Pattern-based Calibration (WTs)

North America: z500 and rainfall examples (MAM)

RMSE (model rainfall vs obs [CPC unified])



Comparing calibration methods

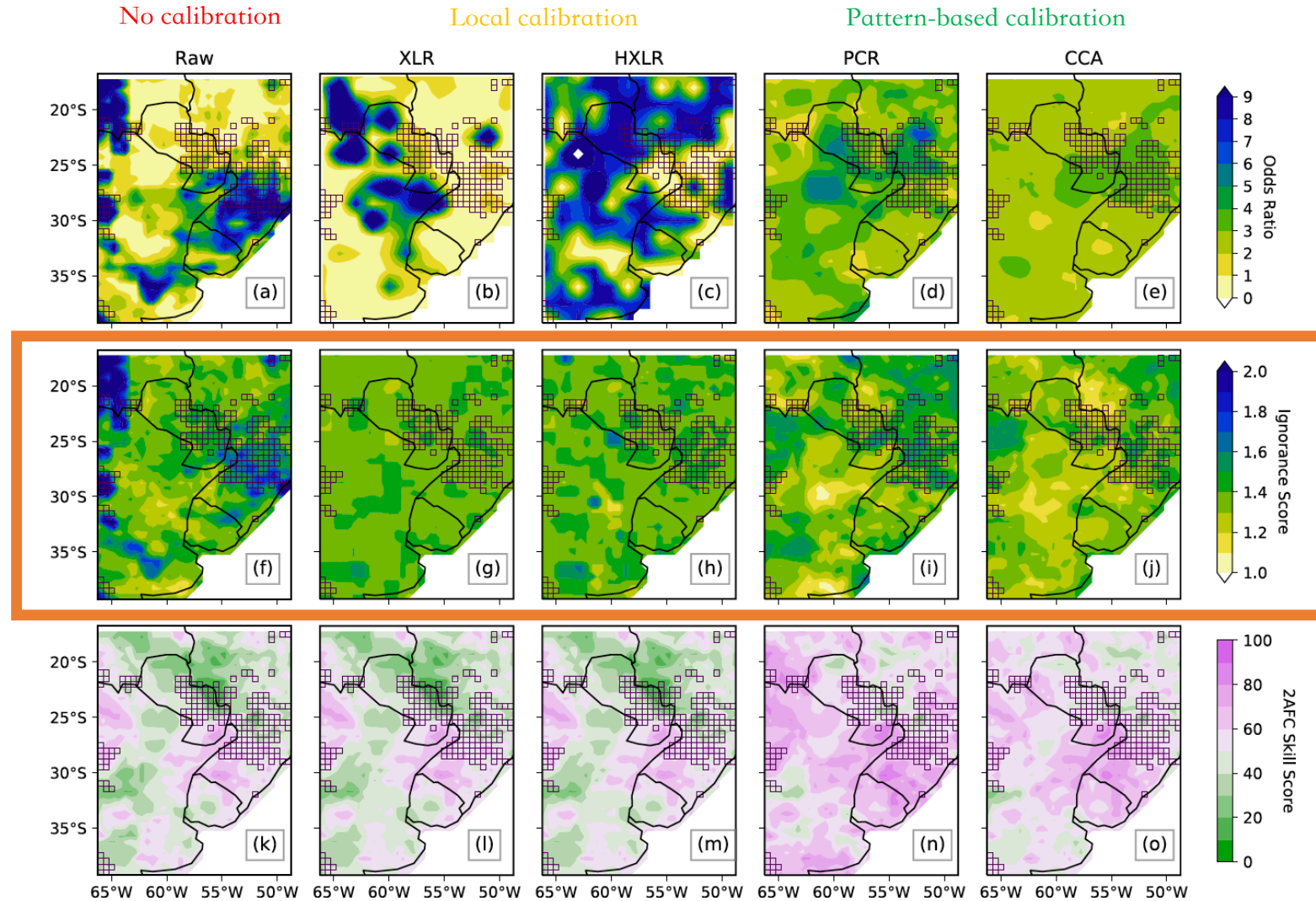


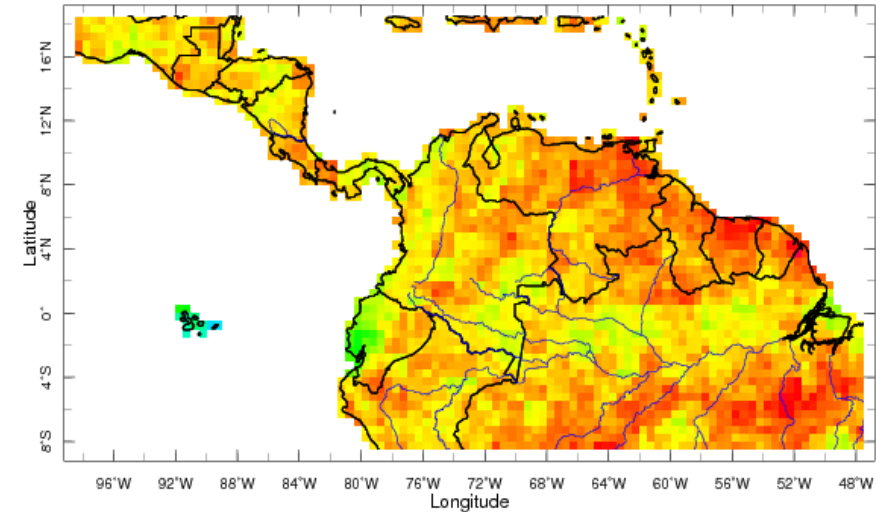
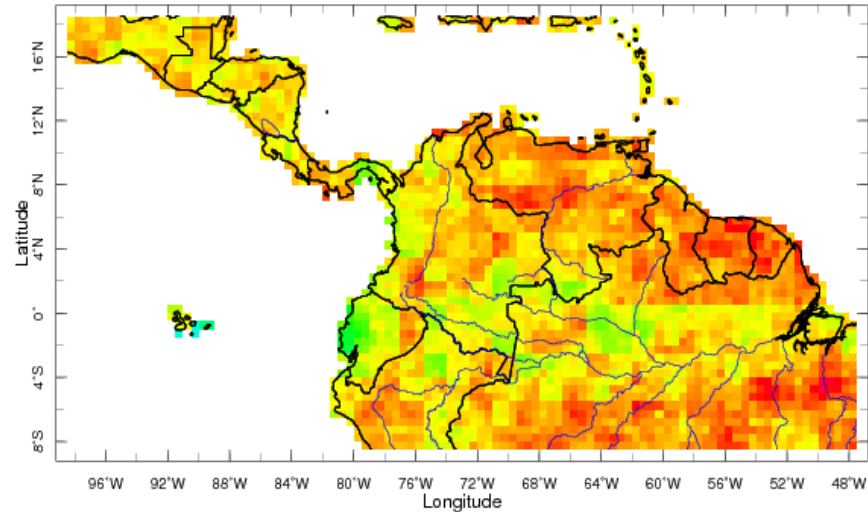
FIG. 10. Raw and MOS-adjusted S2S model forecasts and skill scores for the methods indicated in Table 1. (a)–(e) The heavy rainfall forecast for 1–7 Dec 2015 as odds, defined in Eq. (3) over the target domain. A value greater than 1 indicates that the model forecast greater-than-average odds of rainfall 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.

Initializations: April

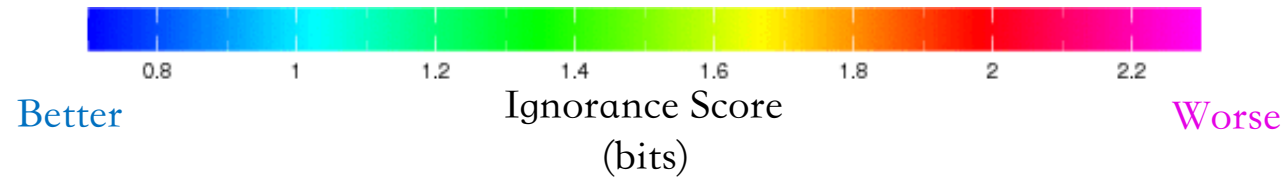
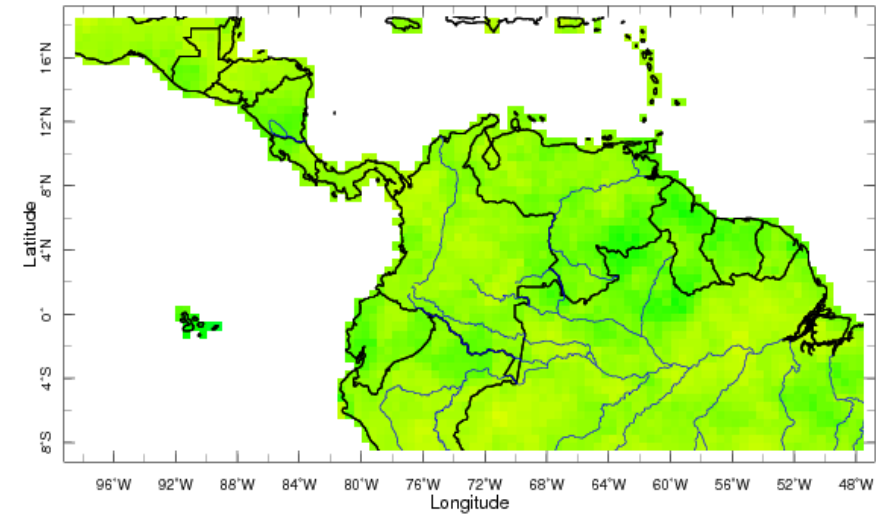
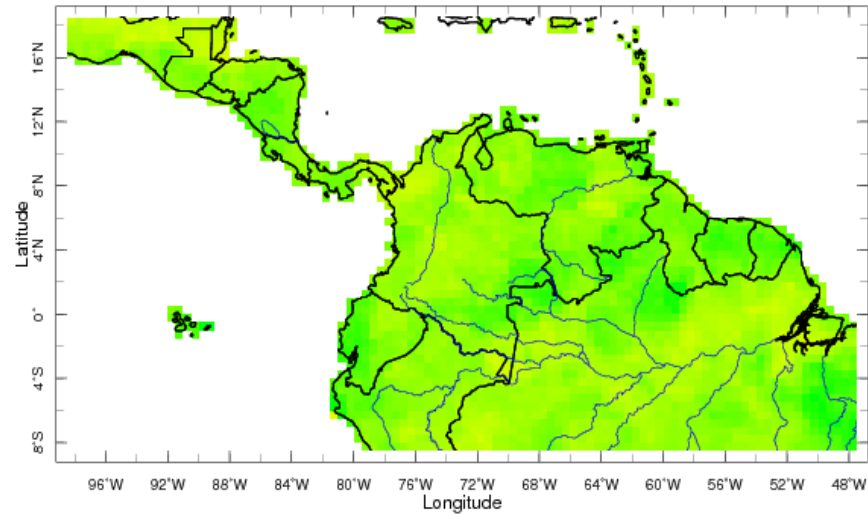
Week 3

Week 4

Uncalibrated



Calibrated (CCA)

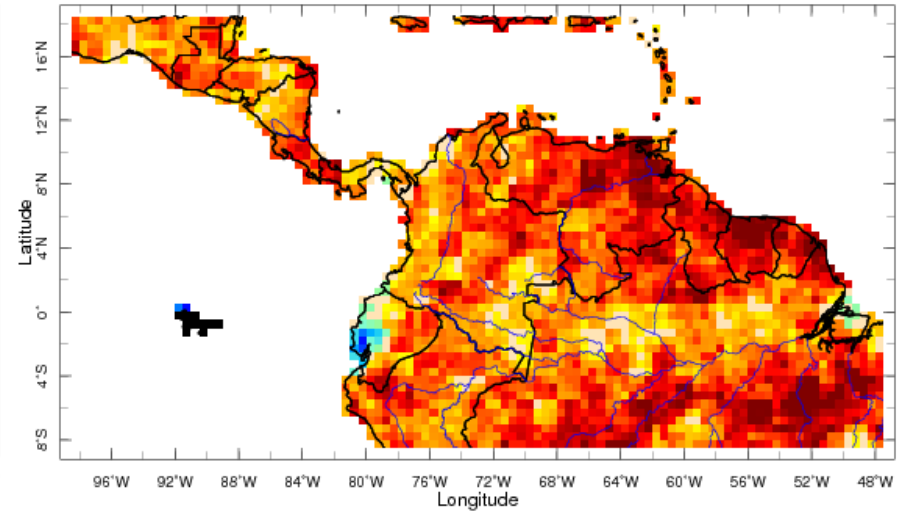
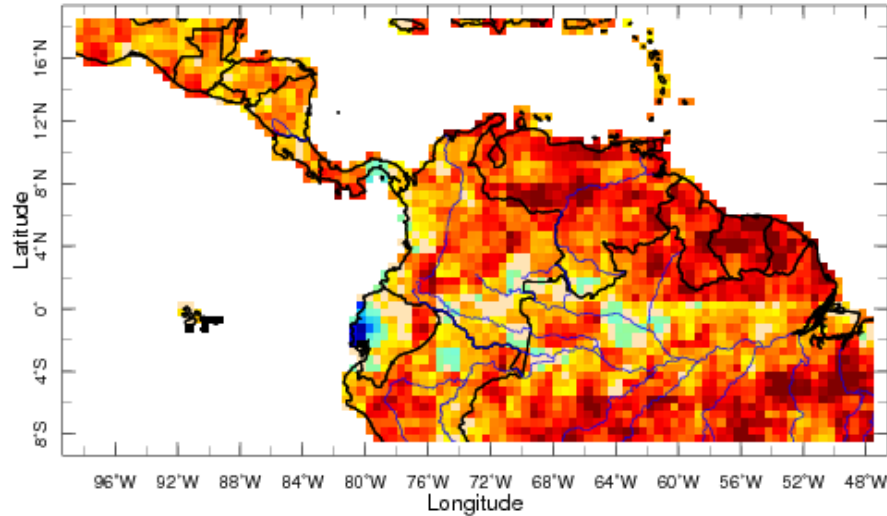


Initializations: April

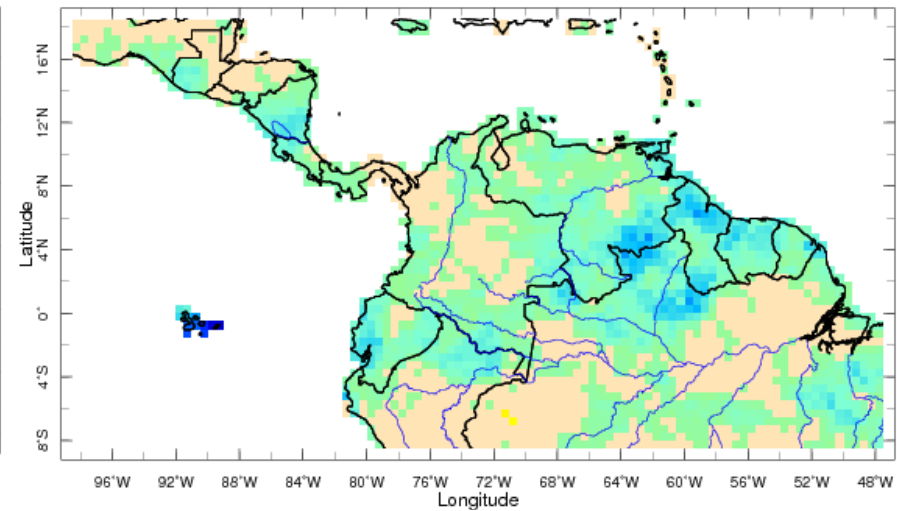
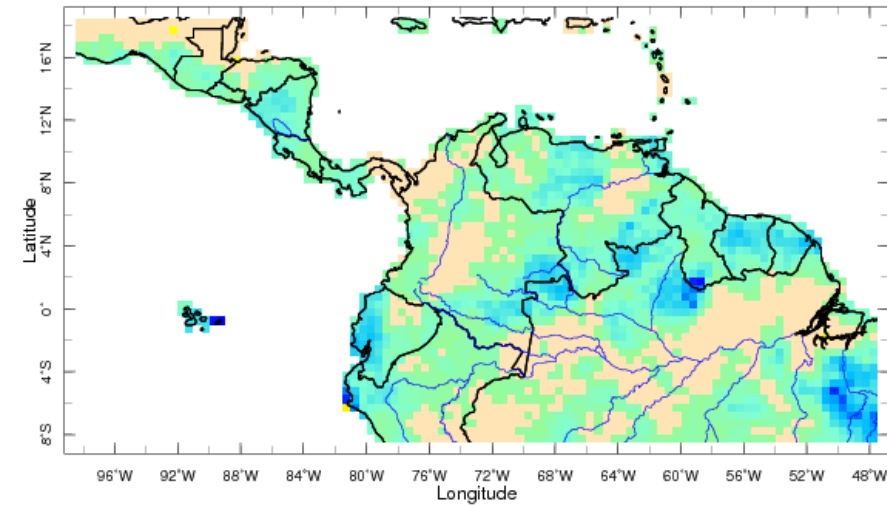
Week 3

Week 4

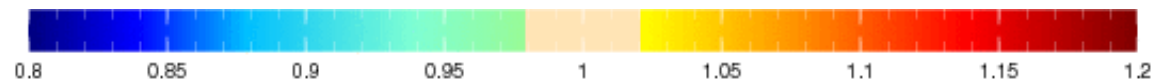
Uncalibrated



Calibrated (CCA)



Ignorance Skill Score



Better

Climatology

Worse

$$IGN = REL - RES + UNC$$

calibration

sharpness
(if reliable)

obs distribution

Outline

A stylized map of the North Atlantic and Europe serves as the background. The map uses a color-coded system: light blue for the ocean, yellow for landmasses, and various shades of red, orange, and green for specific regions, likely representing different climate zones or data sets. The map is oriented with North at the top.

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

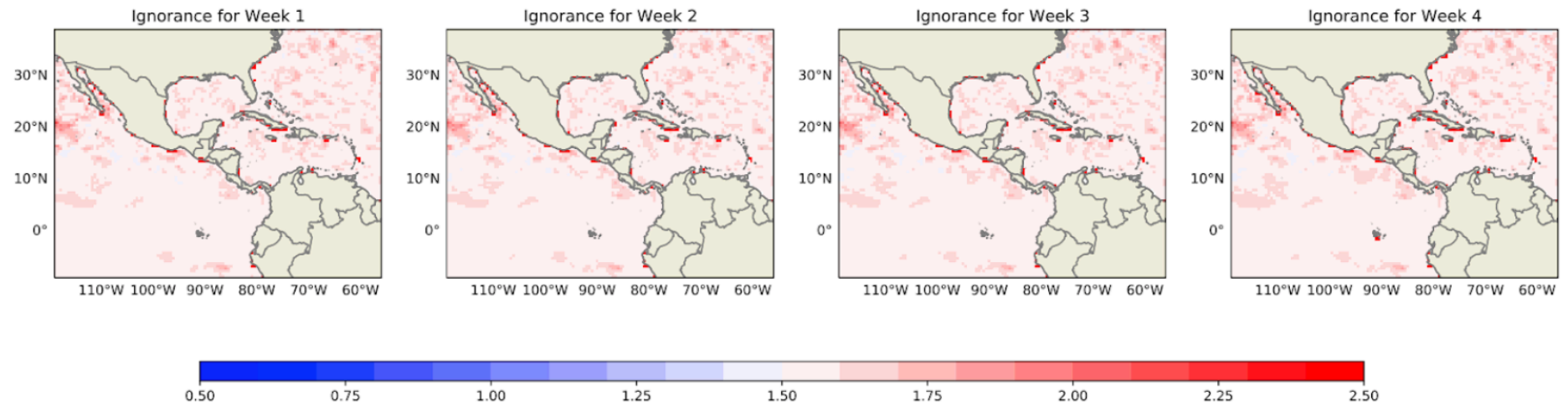
Concluding Remarks

- 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.

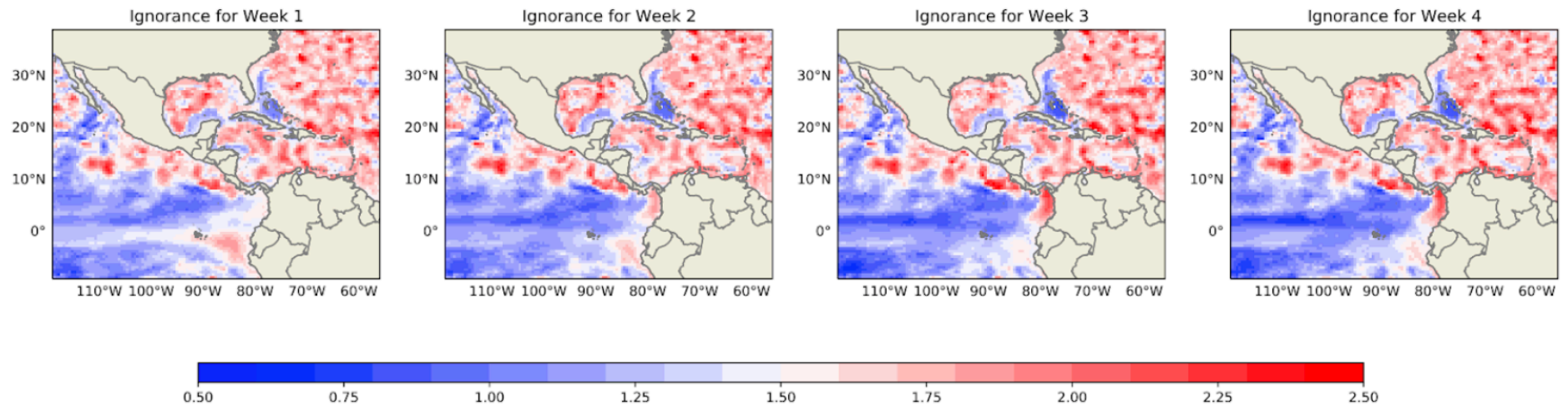
Sub-seasonal Forecast Skill: When, Where and How to Find It?

Ángel G. Muñoz

agmunoz@iri.columbia.edu



(a)



(b)

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

Campos et al (in prep)

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