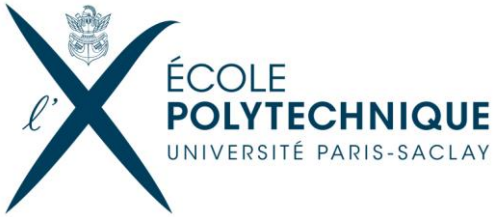
Interpretable Deep Learning for Probabilistic MJO Forecasting

A. Delaunay[1], H. M. Christensen[2]

[1] Ecole Polytechnique, Applied Mathematics, France
[2] University of Oxford, Atmospheric, Oceanic & Planetary Physics, UK

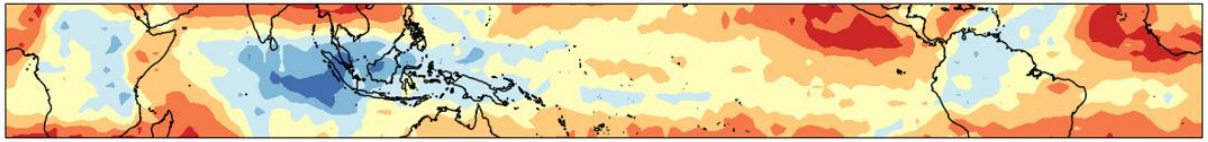
30/03/2022



Introduction

Introduction

- Madden Julian Oscillation
  - Anomalous precipitation and zonal winds area
  - Eastward propagation along the Equator, 30-60 days
  - Main source of variability on sub-seasonal timescales
  - Connections with extreme events



Introduction

Strong variability

- Probabilistic forecasts
- State-dependent and consistent distributions

MJO badly understood

- Interpretability of the network's behaviour
- Improve MJO knowledge on state-dependent predictability

Operational dynamical models don't fulfil completely these requirements

🎯 **Interpretable probabilistic deep learning model**

Plan

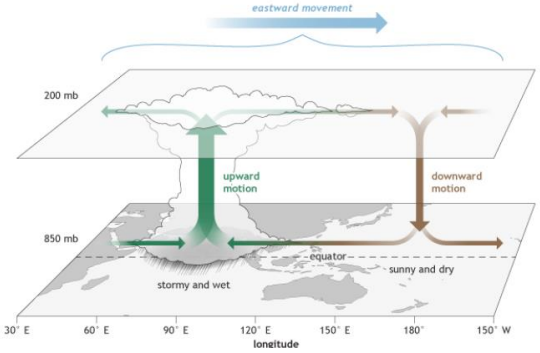
1. Data
2. Methods
3. Results

Data

Characterization of the MJO

Features

- U200, U850 = West-East wind at 200/850 hPa
- OLR = Outgoing Long-Wave Radiations



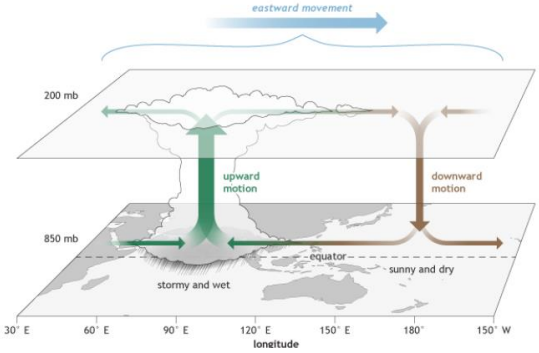
Characterization of the MJO

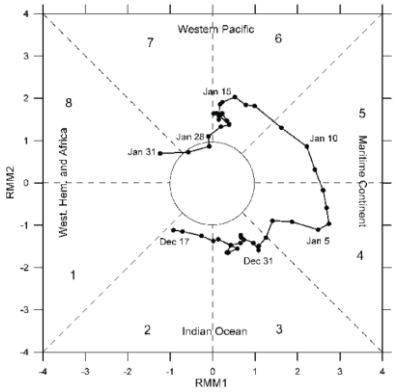
Features

- U200, U850 = West-East wind at 200/850 hPa
- OLR = Outgoing Long-Wave Radiations

1. Remove seasonal cycle (120 days mean, 3 first Fourier harmonics)
2. Compute Empirical Orthogonal Functions (EOFs)
3. Project features maps on EOFs 1,2 → RMM1, RMM2

Predict RMM1,2 (t+τ)





Dataset

- **Features**
  - Daily anomalies : OLR, U200, U850
  - Daily means : Sea Surface Temperature, Humidity (400 hPa), Geopotential (850 hPa), Downward Long-Wave Radiations

- **Characteristics**
  - ECMWF ERA5 – Reanalysis dataset
  - 0 – 360 °E, 20°S – 20°N, 2.5° x 2.5° grid
  - Train set : 1979 – 2011 (80%)
  - Test set : 2011 – 2019 (20%)

Methods

Probabilistic network design

Epistemic uncertainty

- Uncertainty on the weights θ : training on a sample X,Y

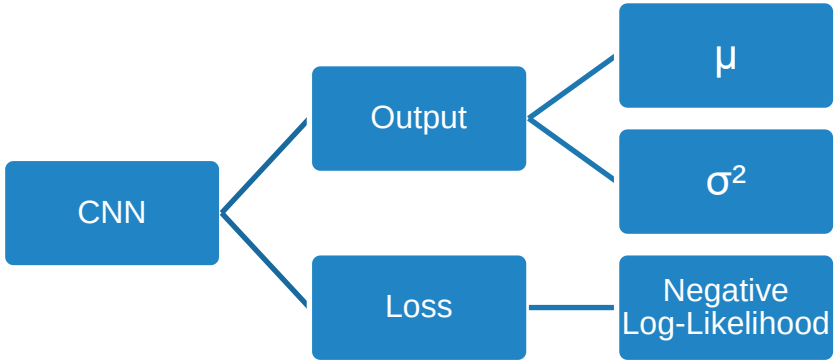
$$p(\mathbf{y}_{t+\tau} \mid \mathbf{x}_t, X, Y) = \int_\theta p(\mathbf{y}_{t+\tau} | \mathbf{x}_t, \theta) p(\theta \mid X, Y) d\theta$$

- Approximation of $p(\theta \mid X, Y)$ by a Bernoulli
- Monte-Carlo Dropout
  - Dropout at test time : M samples of θ
  - Integral approximated with Monte-Carlo

Probabilistic network design

Aleatoric uncertainty

- Uncertainty on the data (irreducible)
- Parametric assumption : bivariate Gaussian *with null correlation*



Probabilistic network design

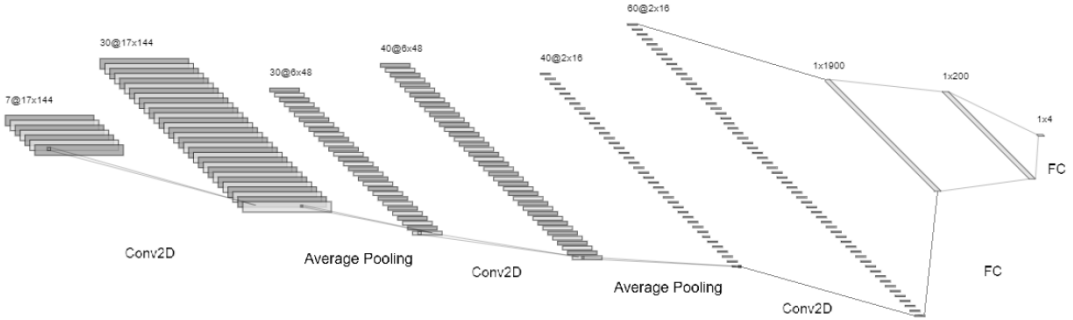
Total uncertainty

- Train one network
- Sample $\theta^{(i)}$ at test time using dropout: ensemble of $(\mu^{(i)}, \sigma^{2(i)})$
- Law of total variance : aleatoric + epistemic

$$\boldsymbol{\mu}_{t+\tau} = \frac{1}{M}\sum_i \boldsymbol{\mu}^{(i)}_{t+\tau}$$

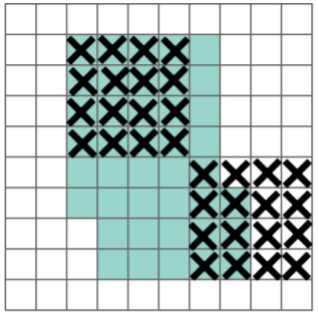
$$\boldsymbol{\sigma}^2_{tot\ t+\tau} = \frac{1}{M}\sum_i \boldsymbol{\sigma}^{(i)\ 2}_{a\ t+\tau} + \mathrm{Var}(\boldsymbol{\mu}^{(i)})$$

CNN architecture and training



CNN architecture

- Handling overfitting
  - L2 – regularization in the loss
  - Dropblock : dropout with correlation

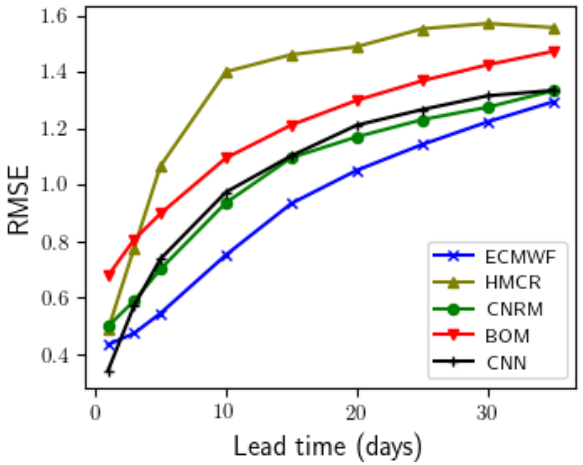


Results - Metrics

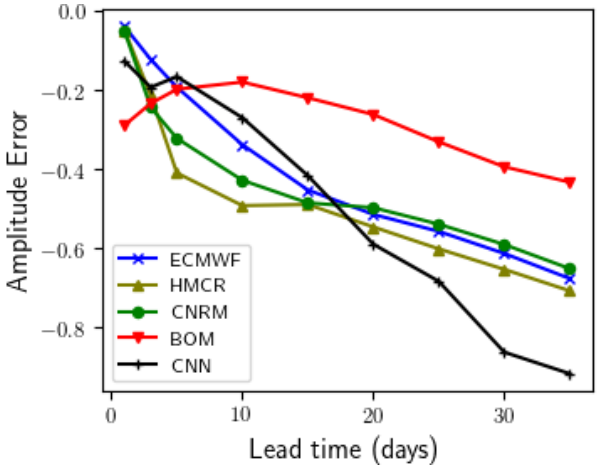
Deterministic metrics

Benchmark with existing baselines

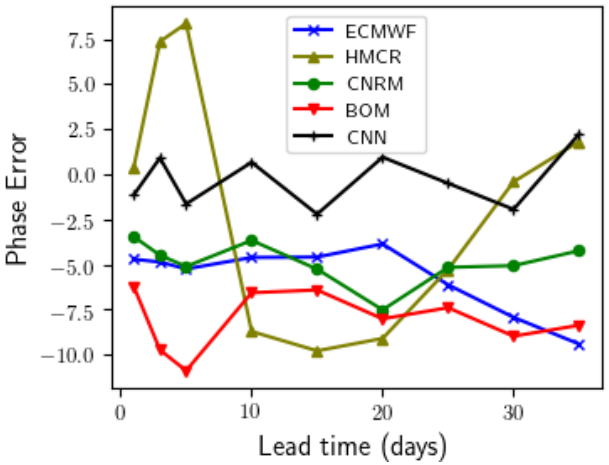
RMSE



Amplitude Error



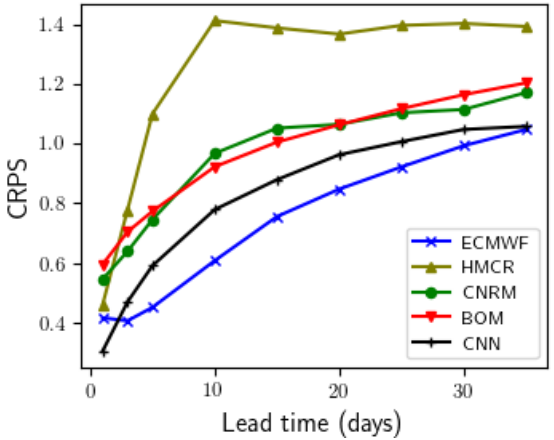
Phase Error



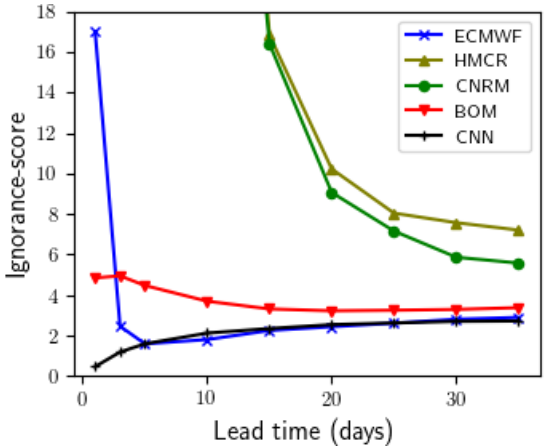
Probabilistic metrics

Summarize overall probabilistic performance

CRPS



Ignorance Score



- Overconfidence in the first days (except CNN)
- CNN outperforms overall
- Stability : uncertainty increases along with accuracy loss

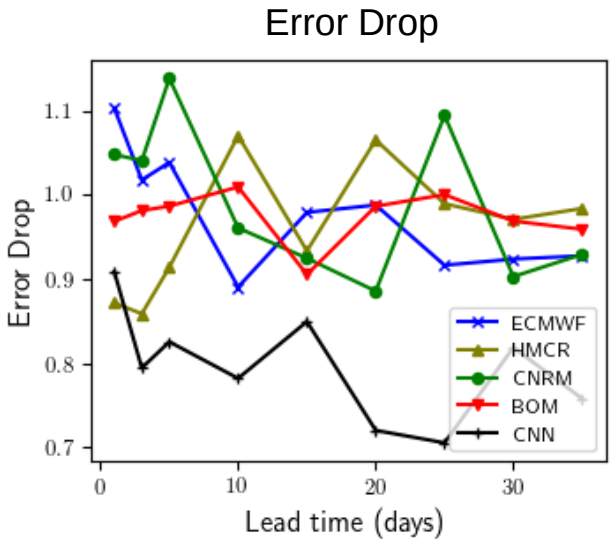
Probabilistic metrics – Ordering

Distinguish certain vs. uncertain days

1. Sort forecasts according to their predicted uncertainty
2. Gradually remove the α% most uncertain forecasts and compute RMSE of the 1 - α% remaining forecasts (confidence curve)
3. Error-Drop : ratio of confidence curve between last and first α quantiles

Probabilistic metrics – Ordering

Distinguish certain vs. uncertain days



CNN has the best uncertainty ordering

Results - Interpretation

Patternet

1. Train network, no dropout at test time
2. Compute attention vectors with statistical estimators (forward pass)
3. Backpropagate output signals by projecting onto the attention vectors (adapted from Kindermans (2017) for Leaky ReLU and Avg Pooling layers)
4. Plot signal attention maps for RMM1 and RMM2, take the absolute value

Maritime Continent

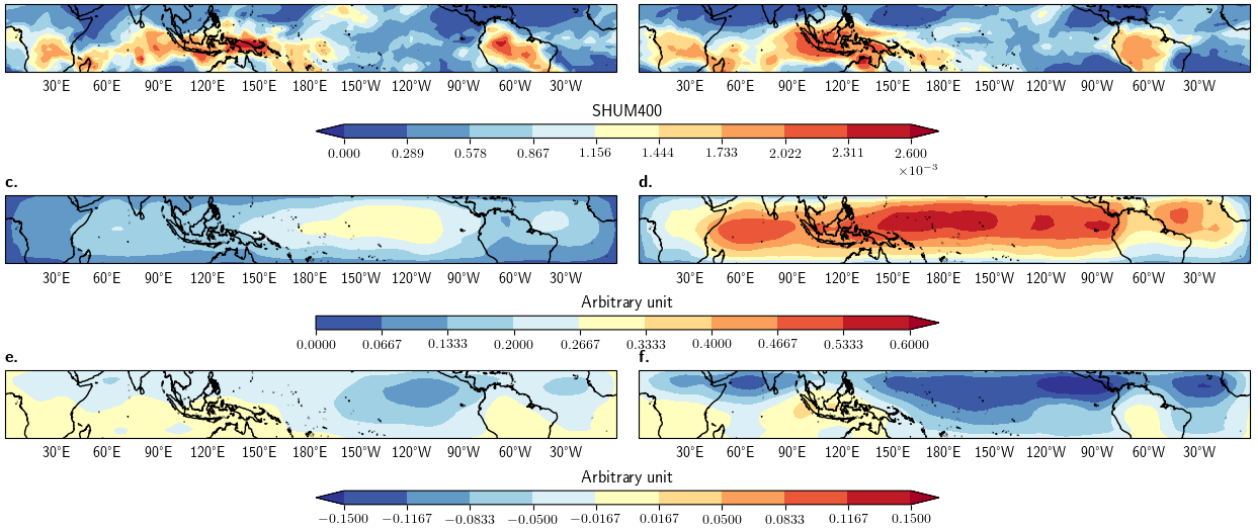
Consistency of CNN's behaviour with literature

Decaying (left) Vs. Propagating (right) events - SHUM400

Composites

Signal Means

Signal Anomalies



Maritime Continent

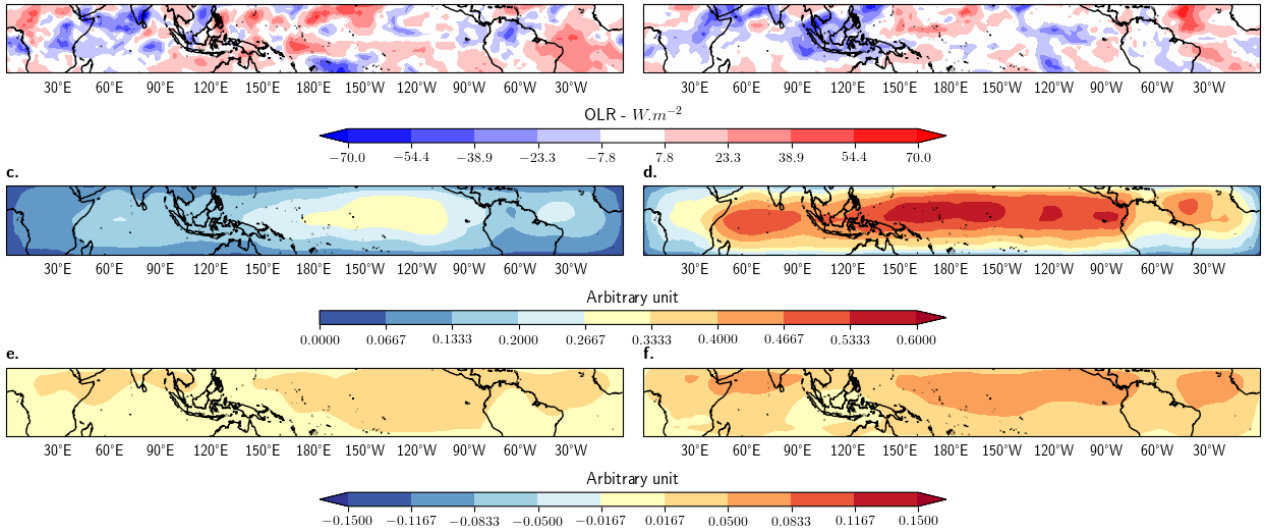
Consistency of CNN's behaviour with literature

Decaying (left) Vs. Propagating (right) events - OLR

Composites

Signal Means

Signal Anomalies

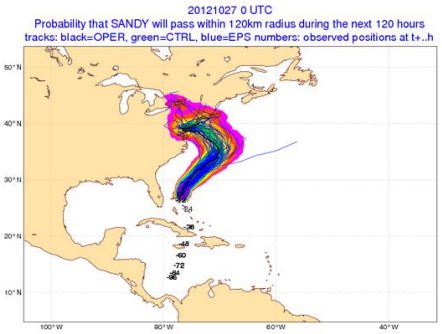


Uncertainty interpretation

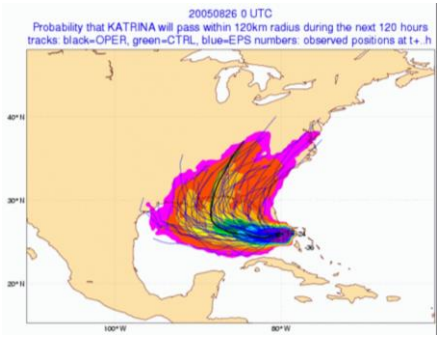
What makes an event predictable or not ?

- Hurricanes : example of a phenomenon with **state-dependent predictability**

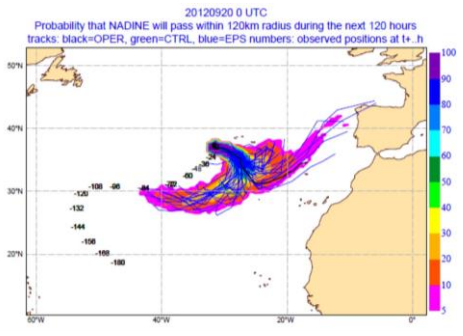
SANDY



KATRINA



NADINE

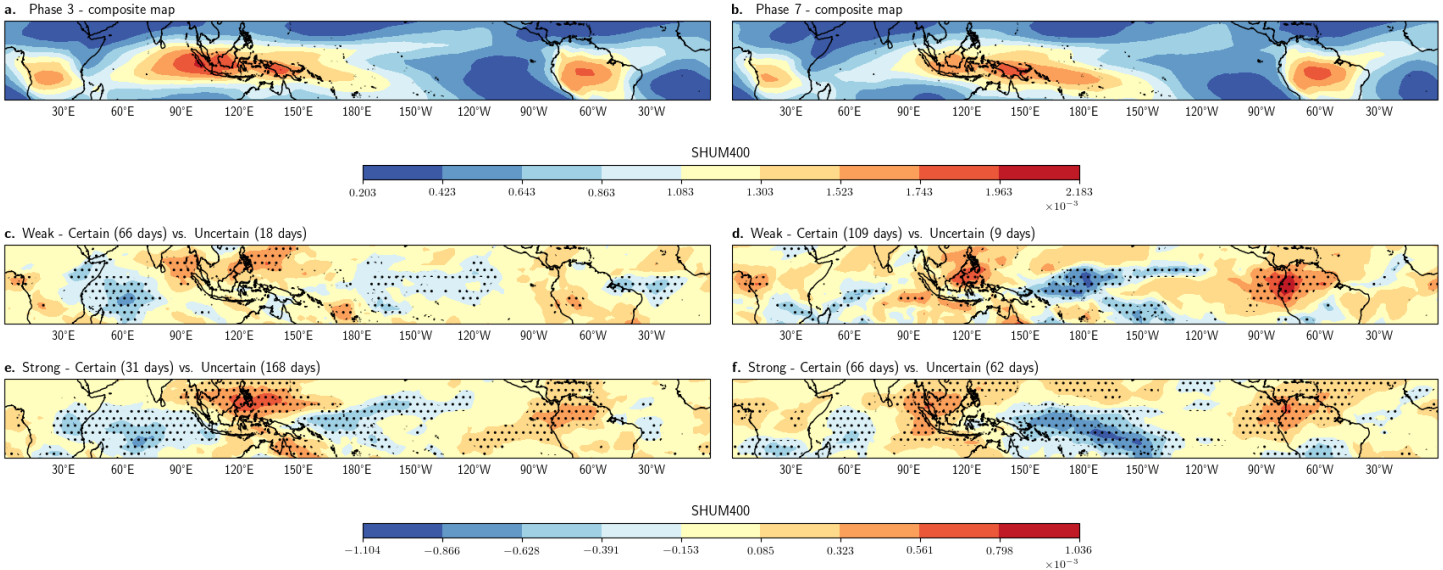


- **What about the Madden-Julian Oscillation ?**

Uncertainty interpretation

What makes an event predictable or not ?

SHUM400 Composites anomalies for certain & uncertain events
Initial phase 3 (left) & 7 (right)



Uncertainty interpretation

What makes an event predictable or not ?

Most uncertain events to end strong at day-10

- 80% (85%) of uncertain events in initial phase 3 (7) end strong
- 35% (40%) of certain events in initial phase 3 (7) end strong

Uncertainty interpretation

What makes an event predictable or not ?

Most uncertain events to end strong at day-10

- 80% (85%) of uncertain events in initial phase 3 (7) end strong
- 35% (40%) of certain events in initial phase 3 (7) end strong

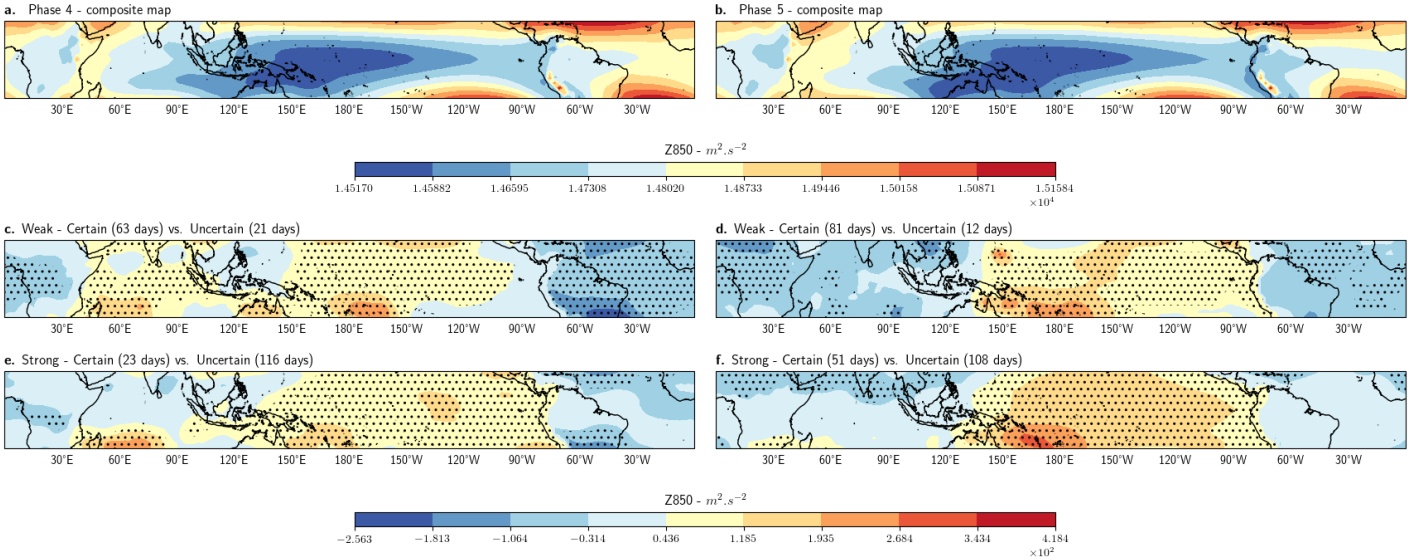
Moisture pattern : predictor of uncertainty or strength or both?

- Further stratification with initial and final RMM strengths
- Phase 3: pattern disappears: **Strength dominant factor**
- Phase 7: pattern persists : **Initially stronger MJO → Uncertainty**

Uncertainty interpretation

What makes an event predictable or not ?

Z850 Composites anomalies for certain & uncertain events
Initial phase 4 (left) & 5 (right)



Conclusion

Conclusion

- Our model can predict the MJO with state-dependent reliability
- Probabilistic distributions are more accurate
- Interpretability ensure the network's behaviour is consistent with literature
- Moisture pattern over the equator for phase 7 events lead to more uncertain forecasts
- Uncertainty is enhanced with a Z850 gradient over the Pacific