



INSIGHTS FROM IMAGE QUALITY ASSESSMENT

Dominique Brunet, Research Scientist
Gabrielle Gascon, Research Manager



MOTIVATION

Integrated gridded analysis for:

- **Total Precipitation**
- Precipitation Type
- Surface Temperature
- Surface Wind

Applications:

- **Verification**
- Nowcasting
- Post-processing
- Hydrology
- Land Surface Scheme

MOTIVATION

Integrated gridded analysis for:

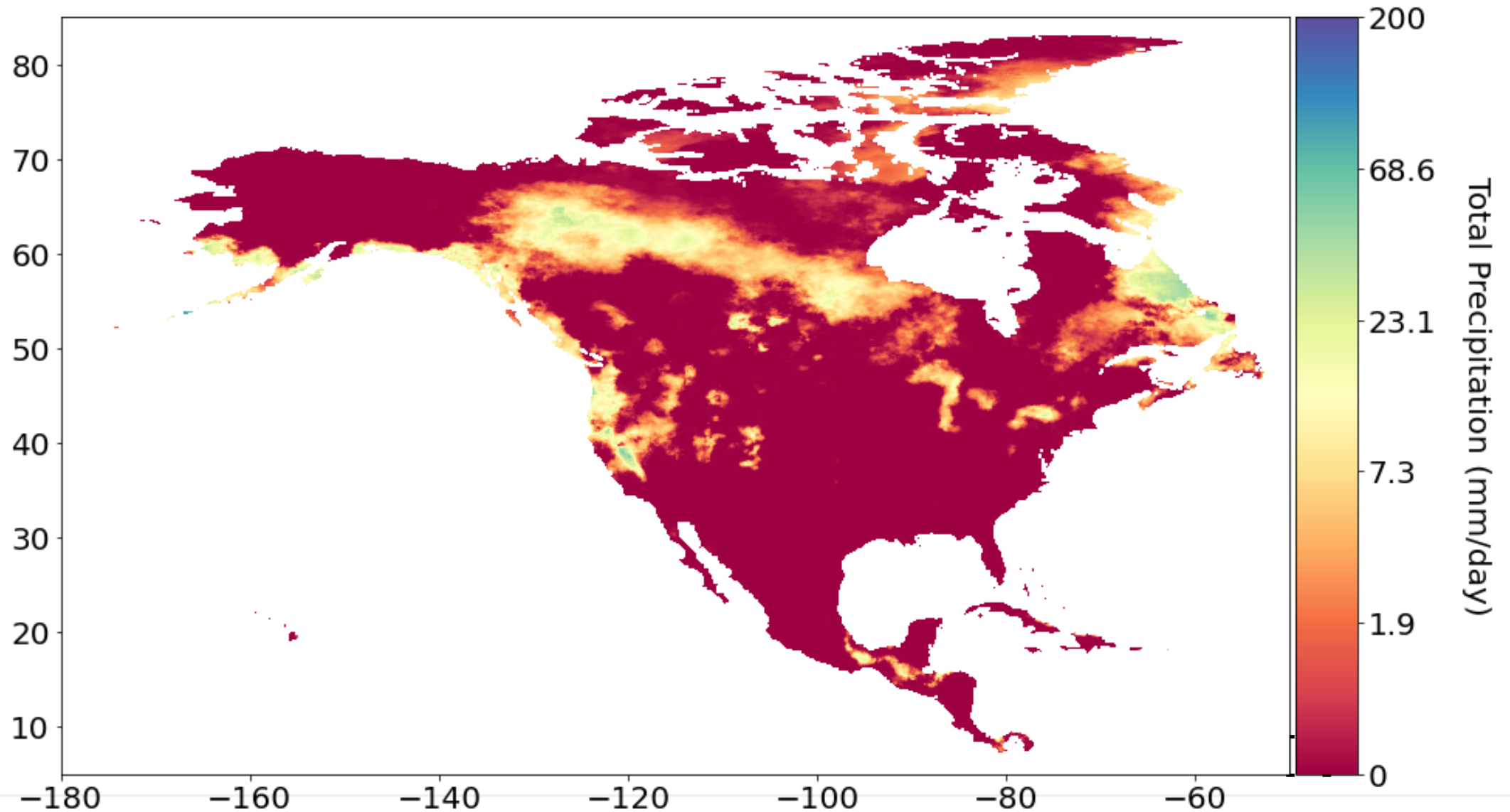
- **Total Precipitation**
- Precipitation Type
- Surface Temperature
- Surface Wind

Applications:

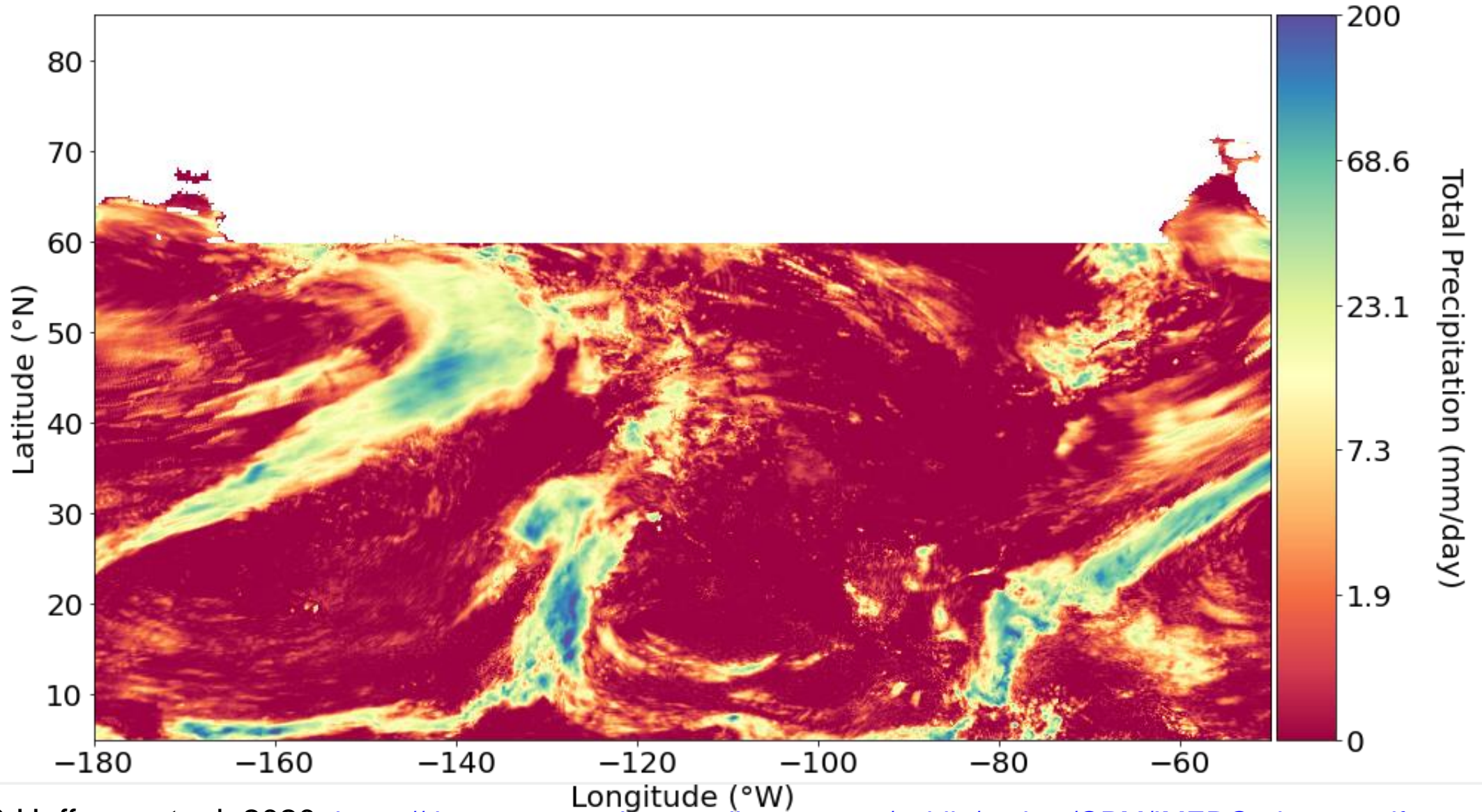
- **Verification**
- Nowcasting
- Post-processing
- Hydrology
- Land Surface Scheme

Training dataset for AI

REFERENCE DATA : EMDNA



SATELLITE VALIDATION: GPM-IMERG V06B



WHICH SCORE TO USE?

Domain field	Scores
Classical Verification	MSE, correlation
Hydrology	NSE [3], KGE [4, 5]
Spatial Verification	FSS [6]
Image Quality Assessment (IQA)	SSIM family [7, 8, 9]

[3] Nash and Sutcliffe 1970: [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6)

[4] Gupta et al., 2009: <https://doi.org/10.1016/j.jhydrol.2009.08.003>

[5] Kling et al. 2012: <https://doi.org/10.1016/j.jhydrol.2012.01.011>

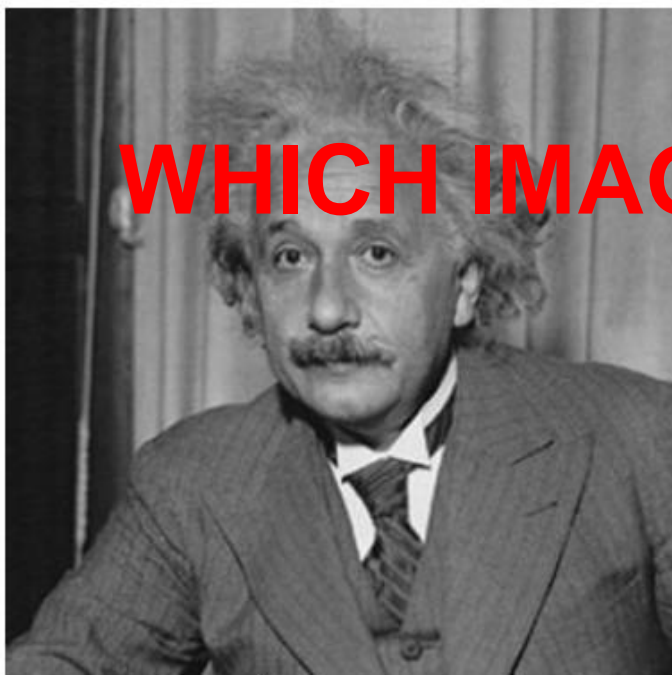
[6] Roberts and Lean, 2008: <https://doi.org/10.1175/2007MWR2123.1>

[7] Wang and Bovik, 2002: <https://doi.org/10.1109/97.995823>

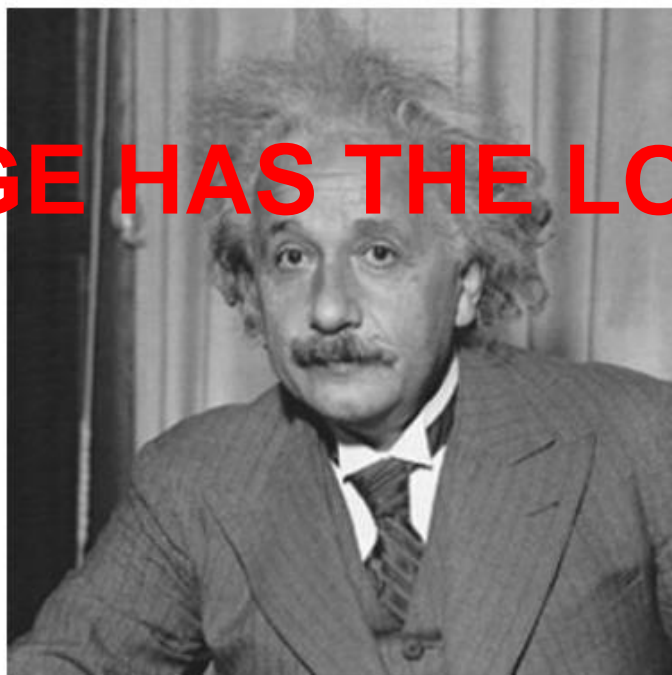
[8] Wang et al., 2004: <https://doi.org/10.1109/TIP.2003.819861>

[9] Brunet et al., 2012: <https://doi.org/10.1109/TIP.2011.2173206>

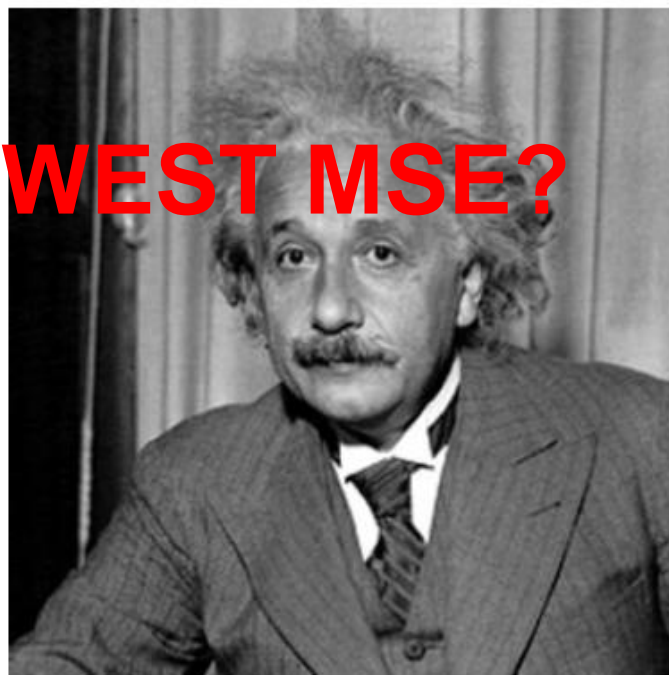
WHICH IMAGE HAS THE LOWEST MSE?



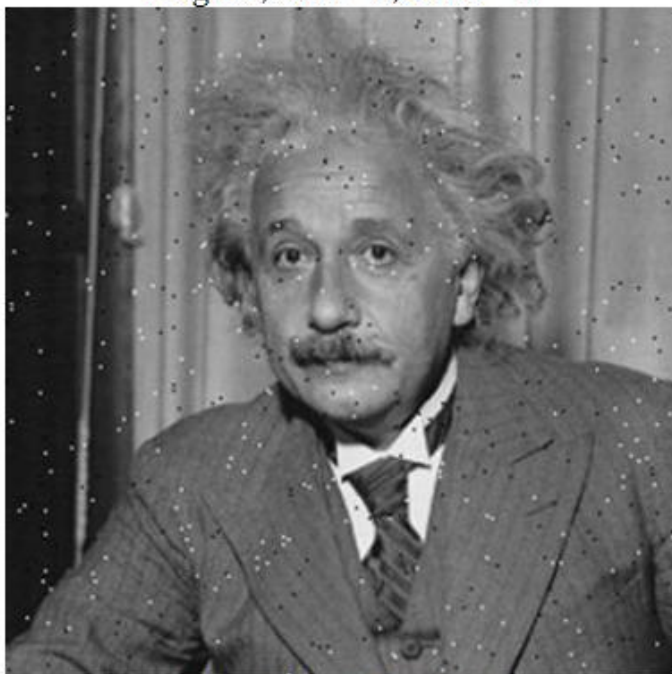
Original, MSE = 0; SSIM = 1



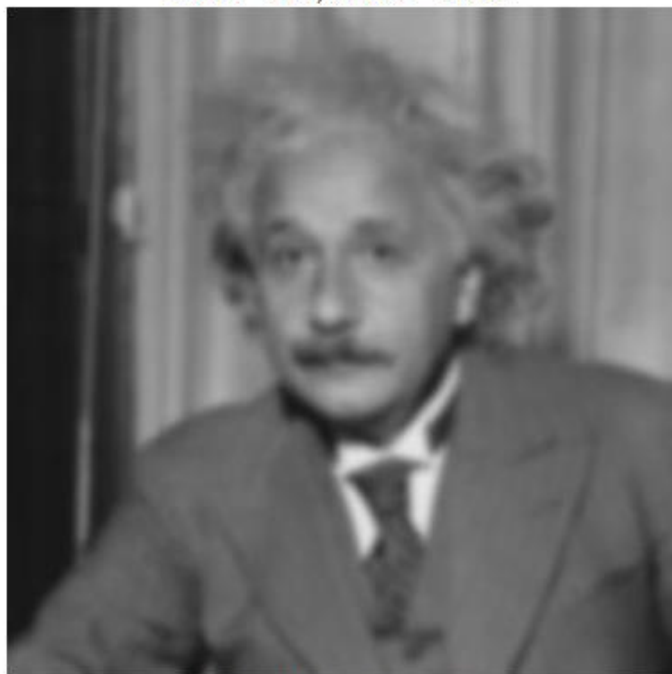
MSE = 144, SSIM = 0.988



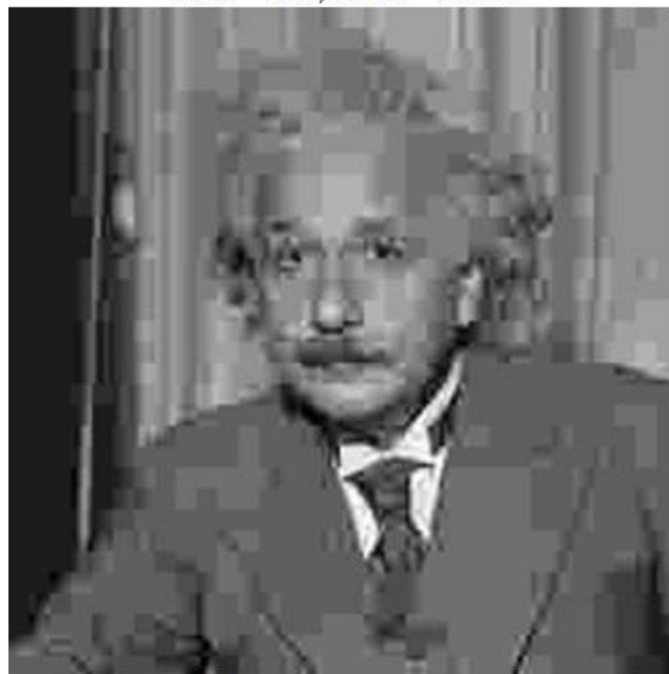
MSE = 144, SSIM = 0.913



MSE = 144, SSIM = 0.840

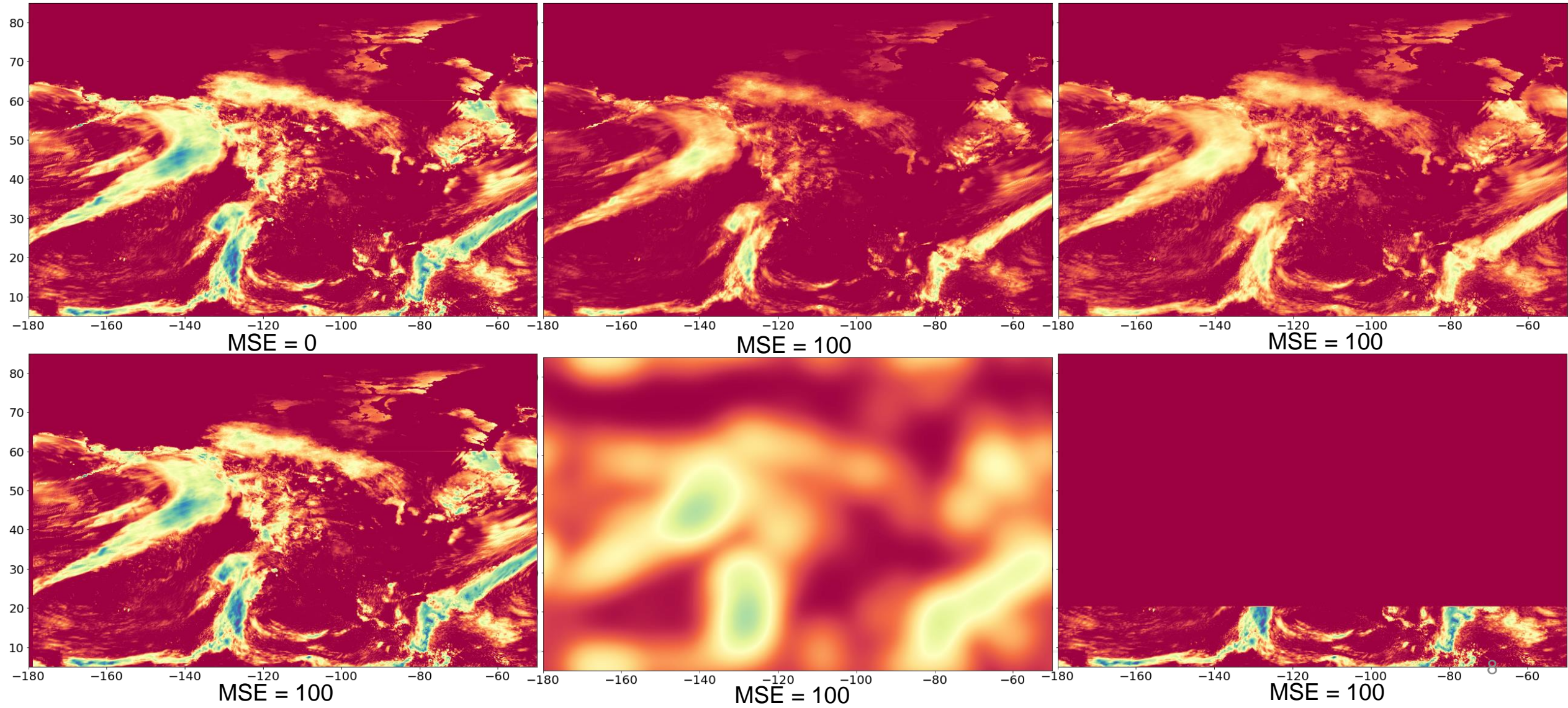


MSE = 144, SSIM = 0.694



MSE = 142, SSIM = 0.662

WHICH ANALYSIS HAS LOWEST MSE?



WISH LIST FOR VERIFICATION SCORES

- Error type decomposition
- Spatial/scale decomposition
- Distance-metric
- Robustness to noise/outliers
- Match subjective judgement of user

DECOMPOSITION OF VERIFICATION SCORES FOR DETERMINISTIC CONTINUOUS FORECASTS

Scores for observation-forecast pairs (x,y) decomposed into

- μ_x and μ_y : Means
- σ_x^2 and σ_y^2 : Variances
- $\sigma_{x,y}$: Covariance
- $r = \frac{\sigma_{x,y}}{\sigma_x\sigma_y}$: Correlation

e.g.
$$\text{MSE} = (\mu_x - \mu_y)^2 + \sigma_x^2 - 2\sigma_{x,y} + \sigma_y^2$$

DECOMPOSITION OF HYDROLOGICAL SCORES

Score	Decomposition
Nash-Sutcliffe Efficiency (NSE) a.k.a. Coefficient of determination (R^2)	$-\frac{(\mu_x - \mu_y)^2}{\sigma_x^2} - \frac{\sigma_y^2}{\sigma_x^2} + 2r \frac{\sigma_y}{\sigma_x}$
Kling-Gupta Efficiency (KGE)	$\left((r - 1)^2 + \left(\frac{\sigma_y}{\sigma_x} - 1 \right)^2 + \left(\frac{\mu_y}{\mu_x} - 1 \right)^2 \right)^{\frac{1}{2}}$
Modified KGE	$\left((r - 1)^2 + \left(\frac{\sigma_y/\mu_y}{\sigma_x/\mu_x} - 1 \right)^2 + \left(\frac{\mu_y}{\mu_x} - 1 \right)^2 \right)^{\frac{1}{2}}$

DECOMPOSITION OF IQA SCORES

Score	Decomposition
Universal Quality Index (UQI)	$\left(\frac{2\mu_x\mu_y}{\mu_x^2 + \mu_y^2} \right) \left(\frac{2\sigma_x\sigma_y}{\sigma_x^2 + \sigma_y^2} \right) r$
Structural Similarity (SSIM) index	$\left(\frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \right) \left(\frac{2\sigma_{x,y} + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \right)$
SSIM metric	$\sqrt{\frac{(\mu_x - \mu_y)^2}{\mu_x^2 + \mu_y^2 + C_1} + \frac{\sigma_x^2 + \sigma_y^2 - 2r\sigma_x\sigma_y}{\sigma_x^2 + \sigma_y^2 + C_2}}$

PROPERTIES OF VERIFICATION SCORES

Properties	NSE/KGE	UQI	SSIM index	SSIM metric
Error type decomposition	✓	✓	✓	✓
Robustness	X	X	✓	✓
Distance-metric	X	X	X	✓

RELATIVE VS ABSOLUTE ERROR

$$(\mu_x - \mu_y)^2 \text{ or } (\mu_y/\mu_x - 1)^2 ?$$

Absolute error :
deviation from mean

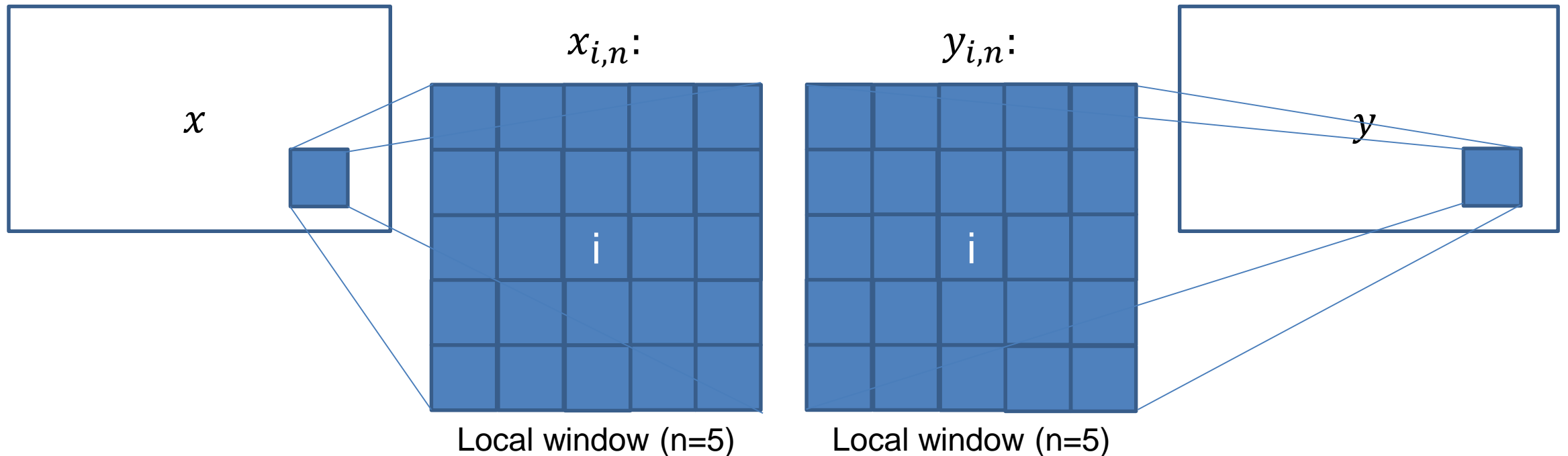
Relative error :
ratios

Temperature
Pressure

Precipitation amount
Wind speed

SPATIAL/SCALE DECOMPOSITION

$\mu_{x_{i,n}}, \mu_{y_{i,n}}, \sigma_{x_{i,n}}^2, \sigma_{y_{i,n}}^2, \sigma_{x_{i,n}, y_{i,n}}$: local means, variances and covariances at location i and scale n .



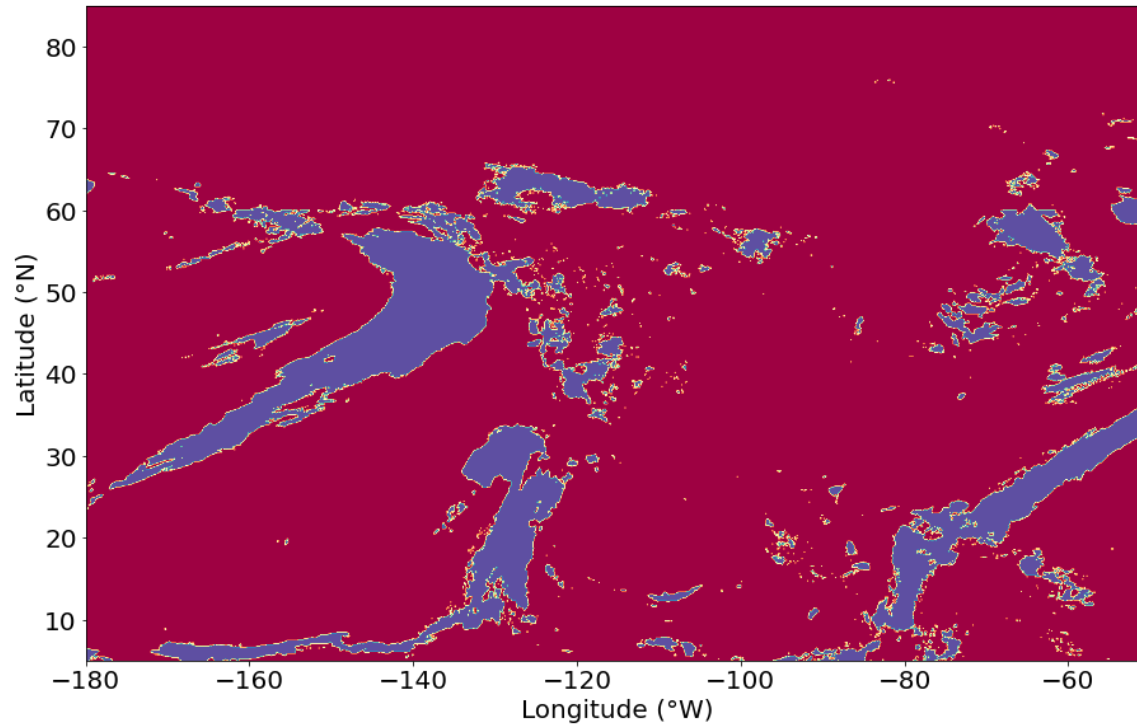
FRACTIONAL SKILL SCORE (FSS)

- Designed for binary forecasts.

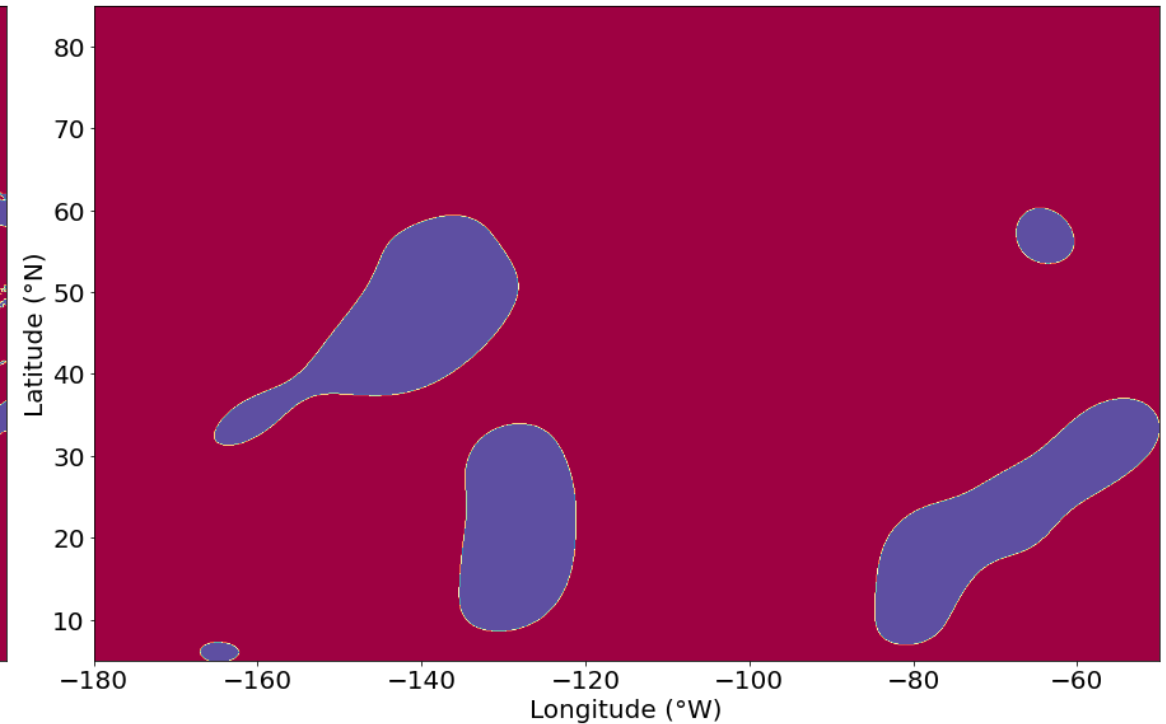
$$FSS_n = \frac{2 \sum_{i=1}^I \mu_{x_{i,n}} \mu_{y_{i,n}}}{\sum_{i=1}^I (\mu_{x_{i,n}}^2 + \mu_{y_{i,n}}^2)}$$

$$\text{Skillful when } FSS_n \geq \frac{1}{2} + \frac{1}{2} f_0 = \frac{1}{2} + \frac{1}{2} \mu_x$$

FRACTIONAL SKILL SCORE (FSS)



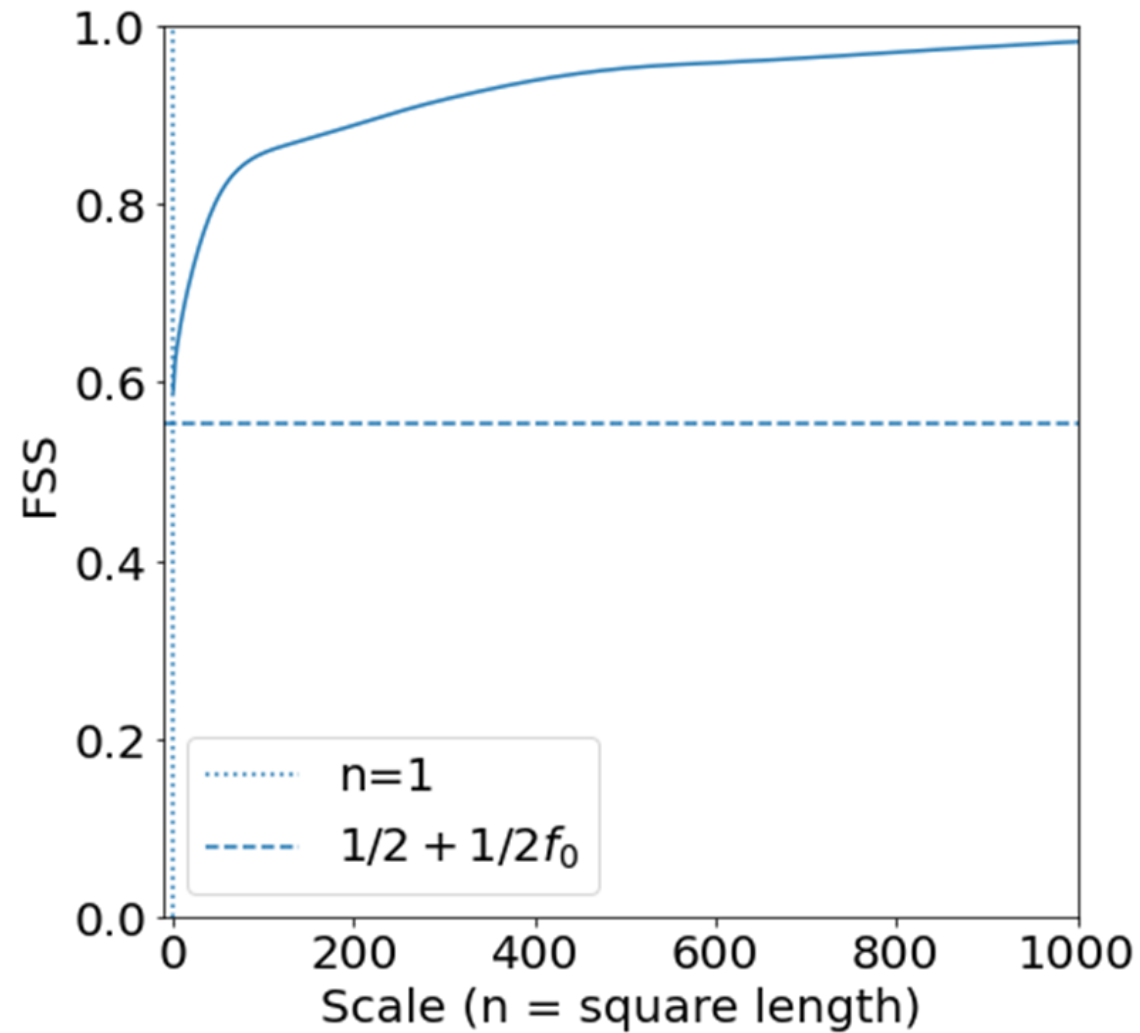
X: Reference
 $f_0 = 11\%$



Y: Forecast

Threshold: 10mm/day

SKILLFUL AT ALL SCALES (?)

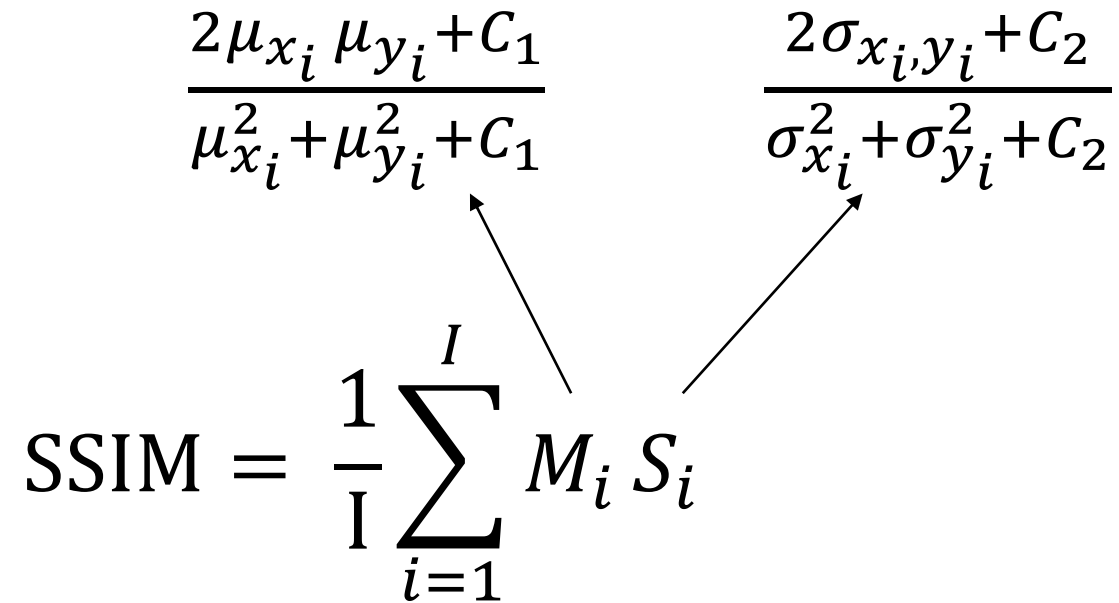


STRUCTURAL SIMILARITY (SSIM) FAMILY

- Weighted statistics instead of uniform window.

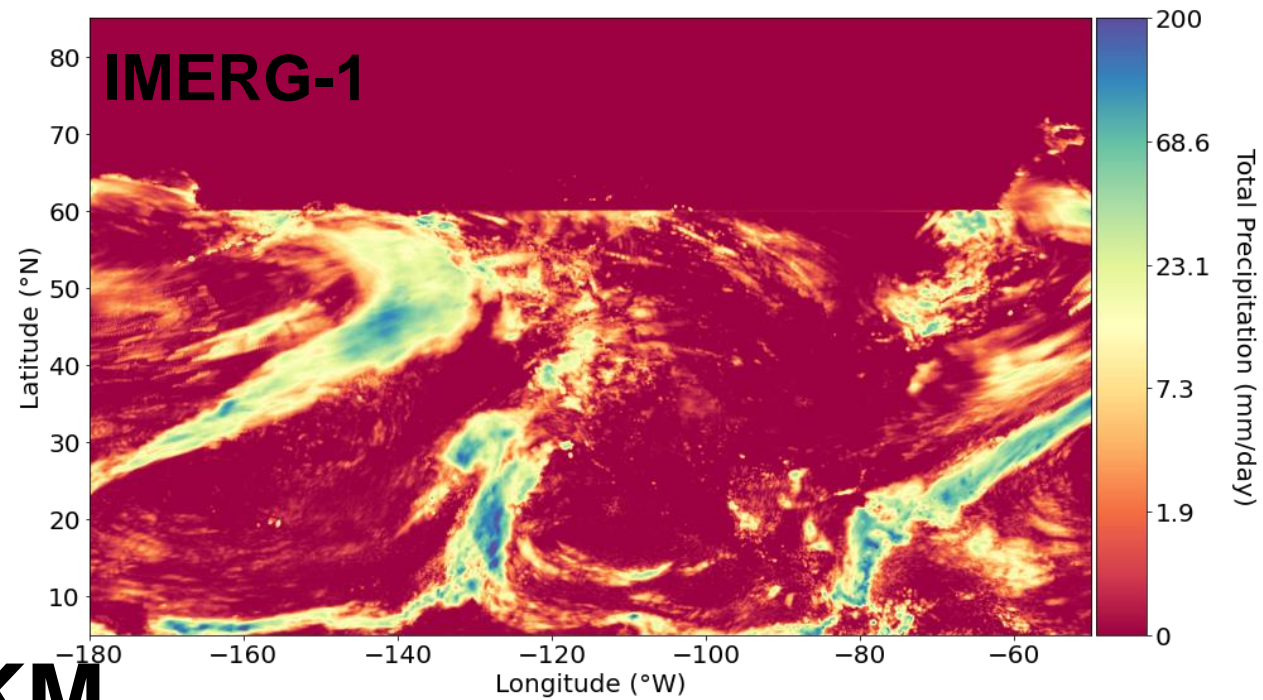
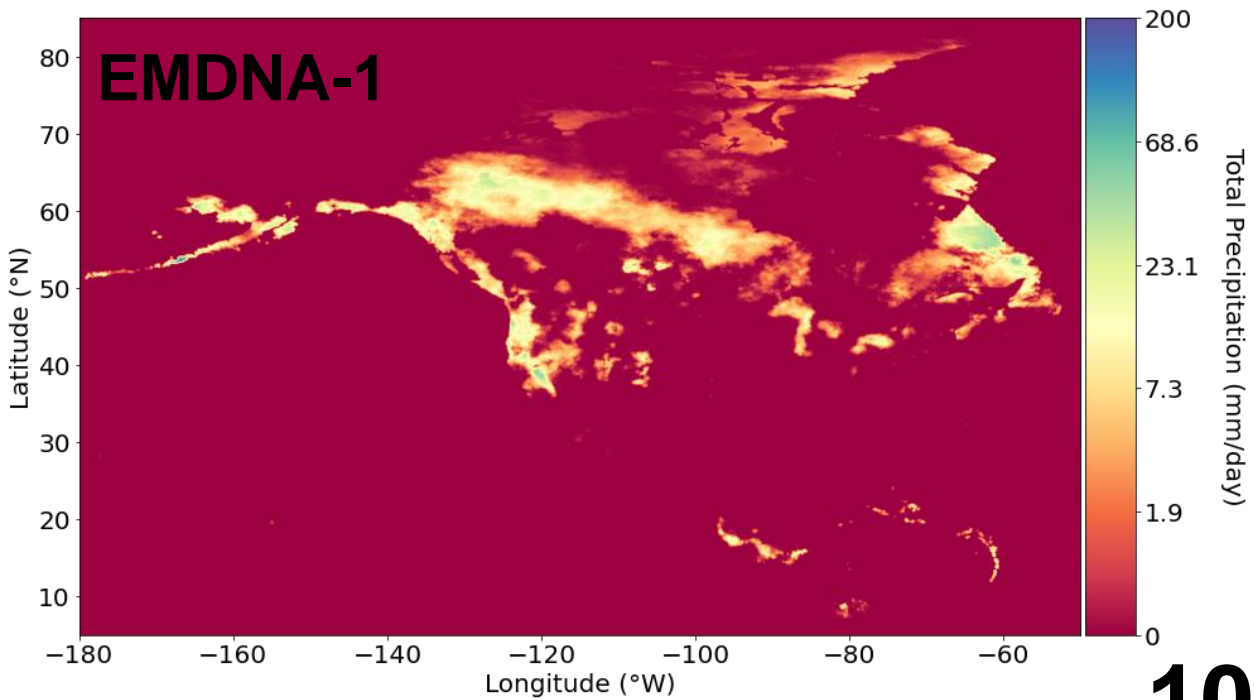
$$\text{SSIM} = \frac{1}{I} \sum_{i=1}^I M_i S_i$$

$\frac{2\mu_{x_i}\mu_{y_i}+C_1}{\mu_{x_i}^2+\mu_{y_i}^2+C_1}$ $\frac{2\sigma_{x_i,y_i}+C_2}{\sigma_{x_i}^2+\sigma_{y_i}^2+C_2}$

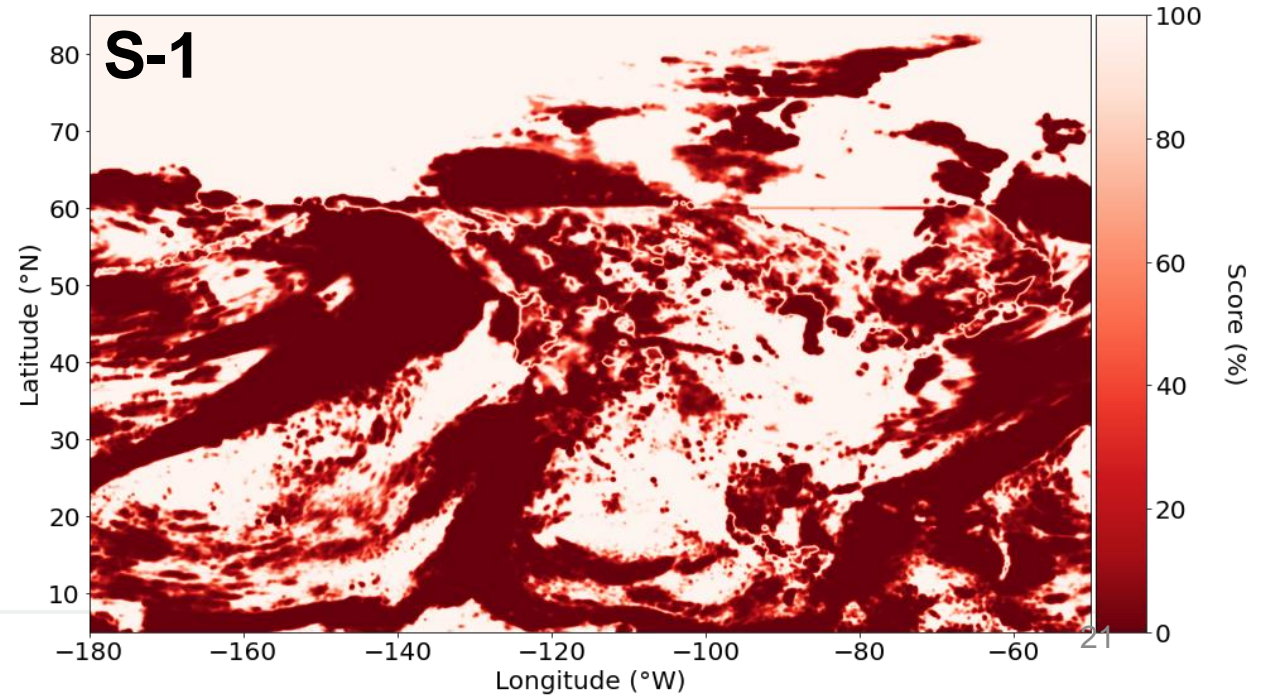
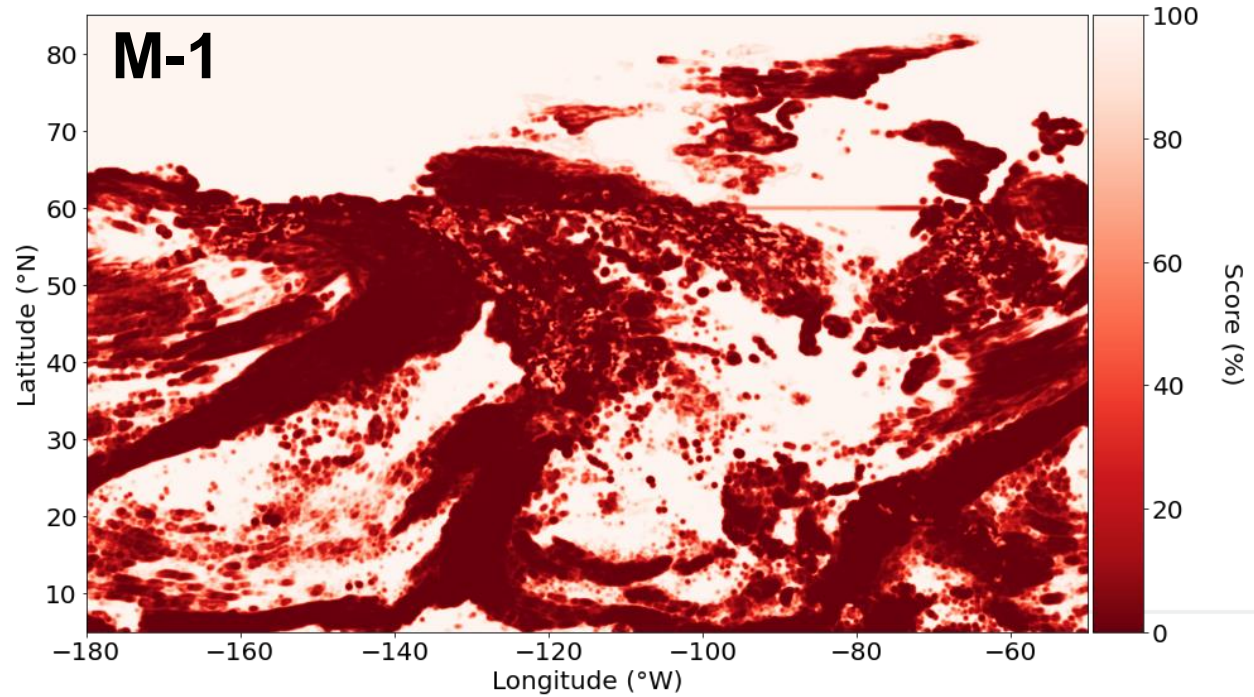


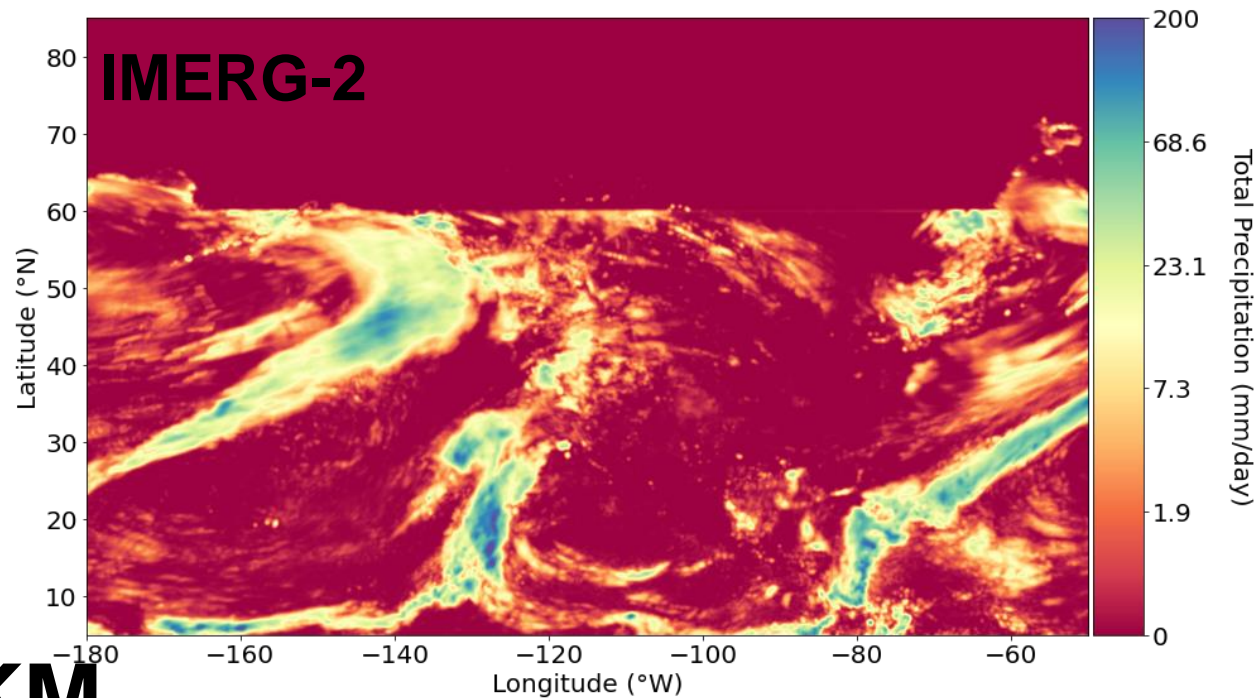
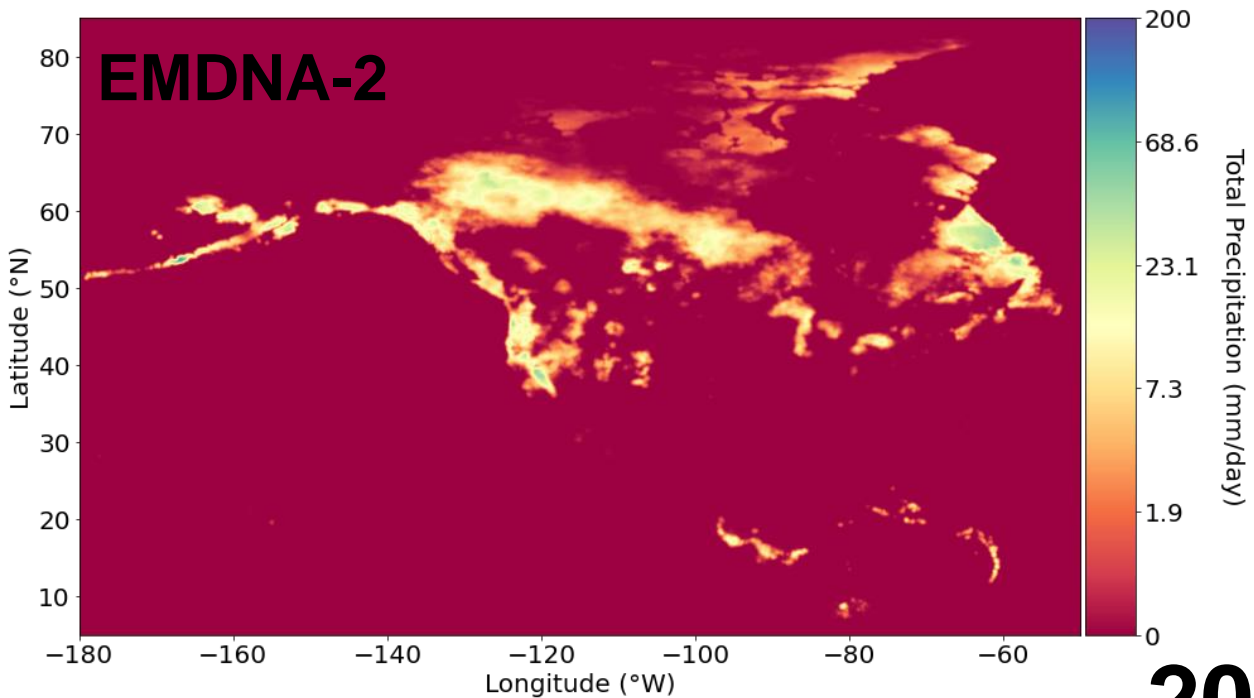
MULTI-SCALE SSIM

- Apply SSIM decomposition at multiple scales.

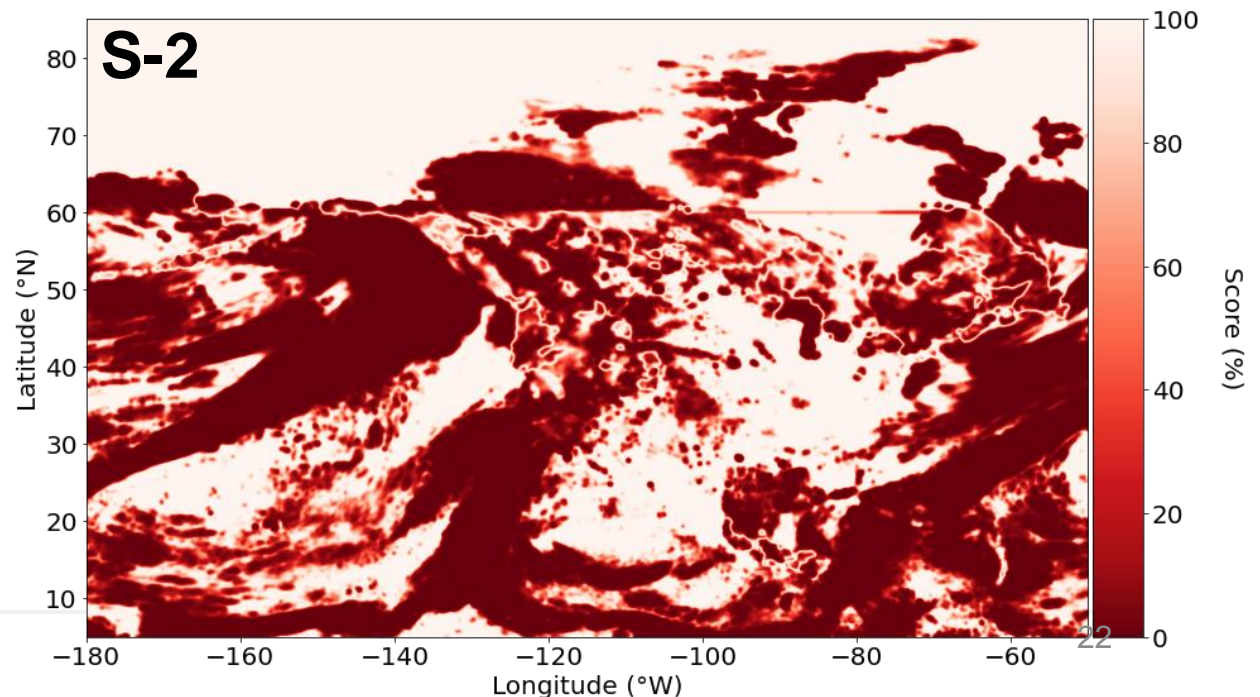
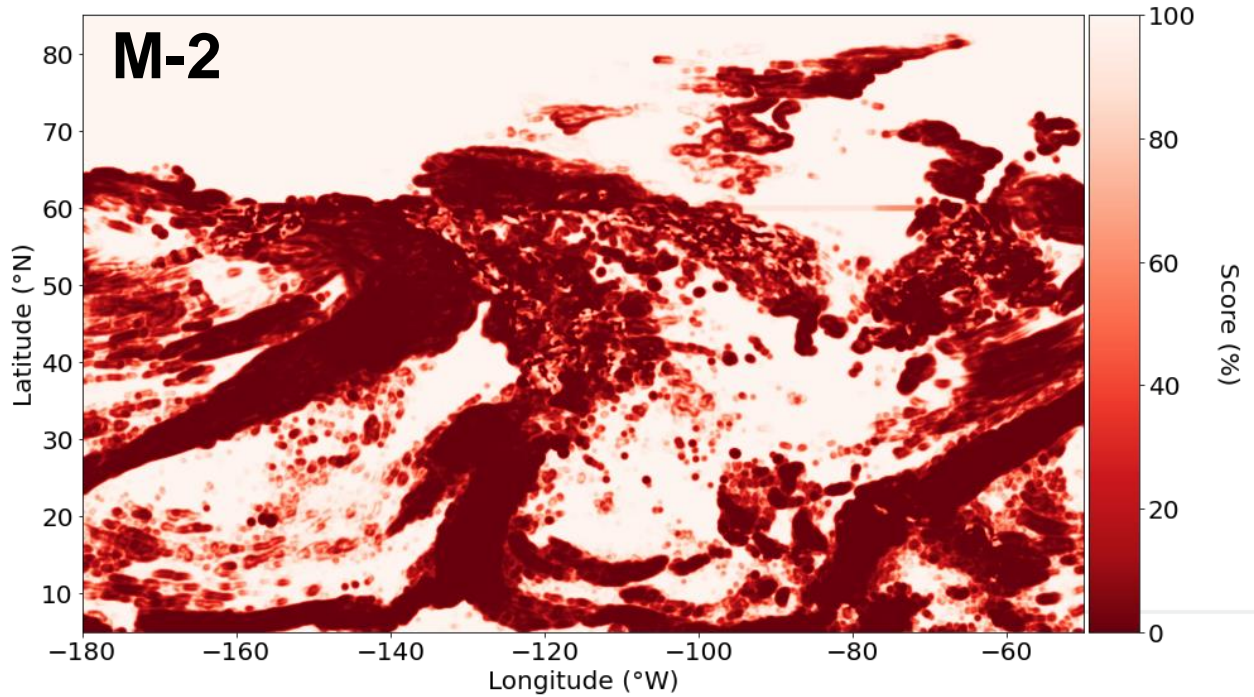


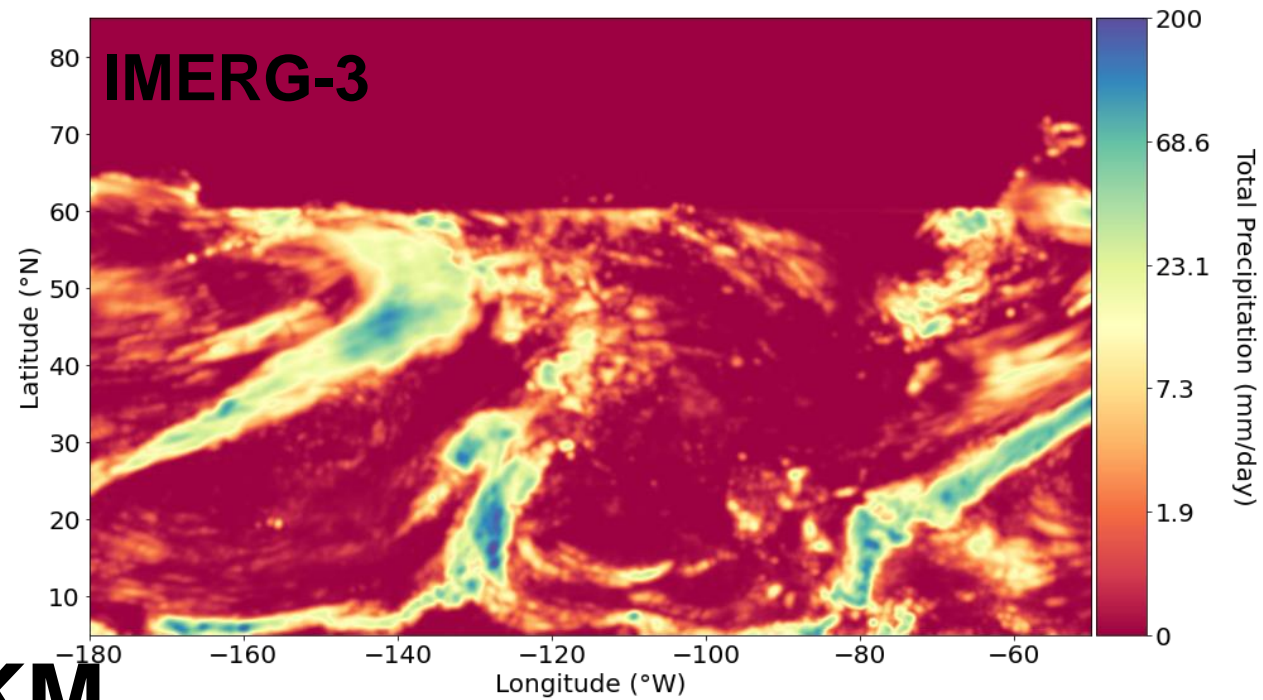
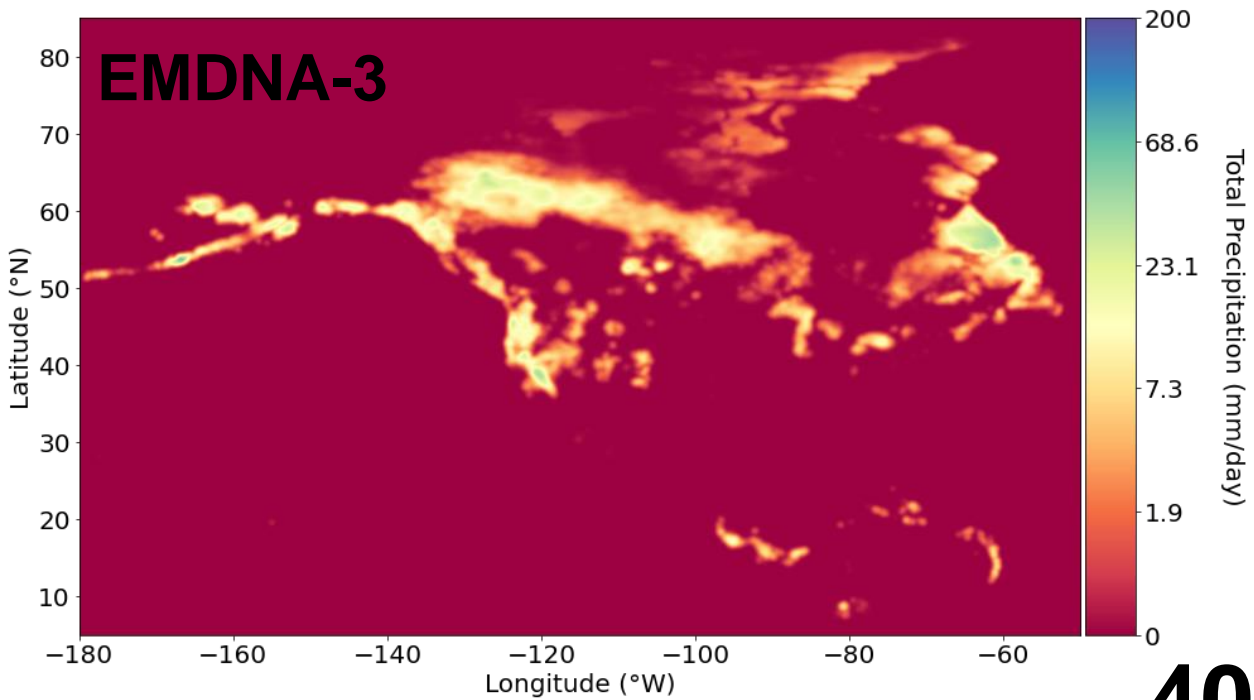
10KM



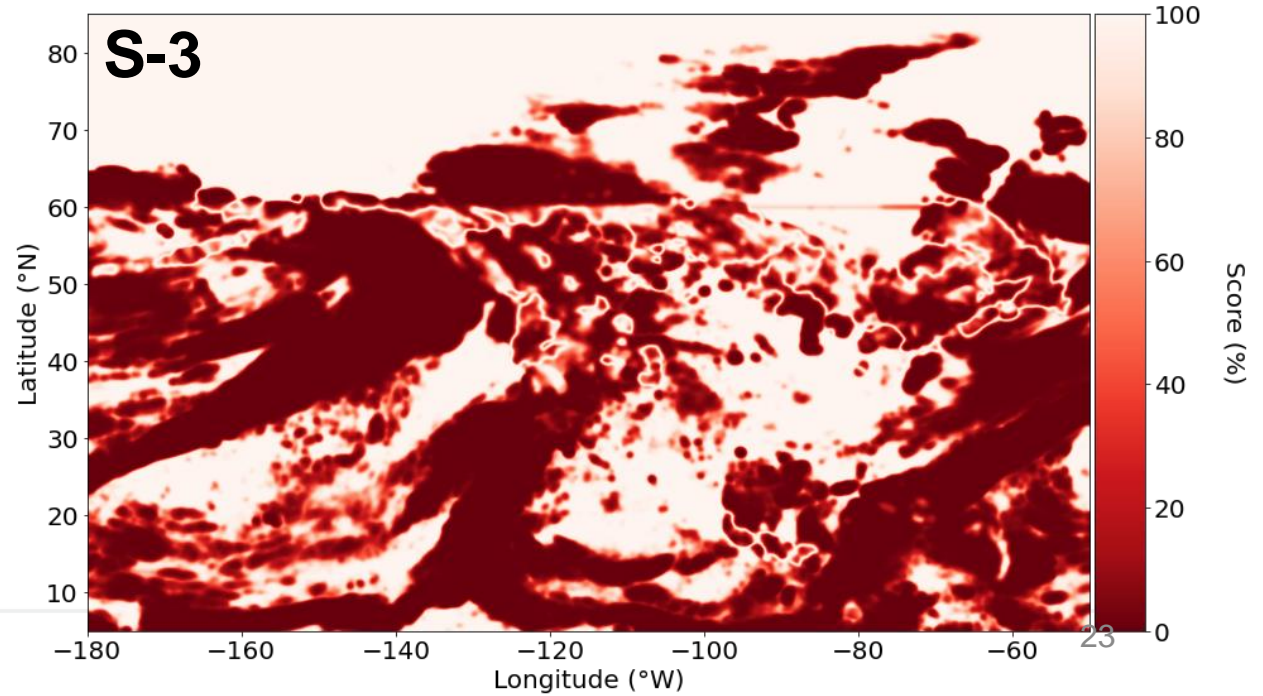
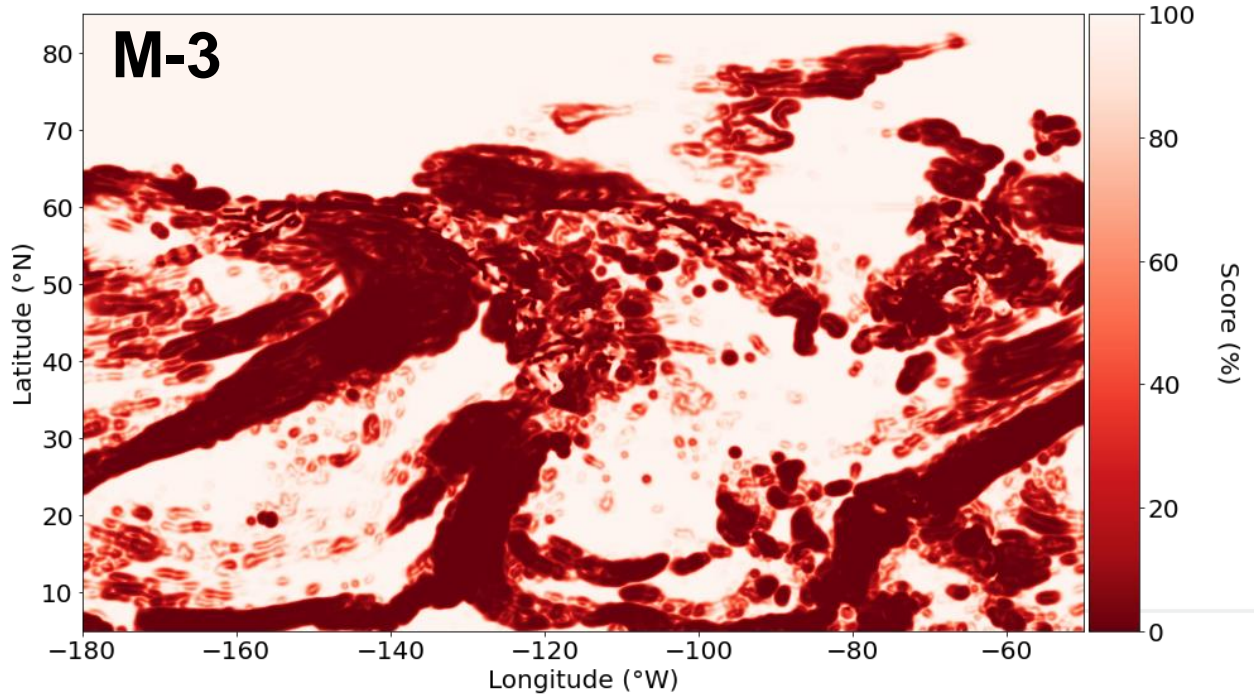


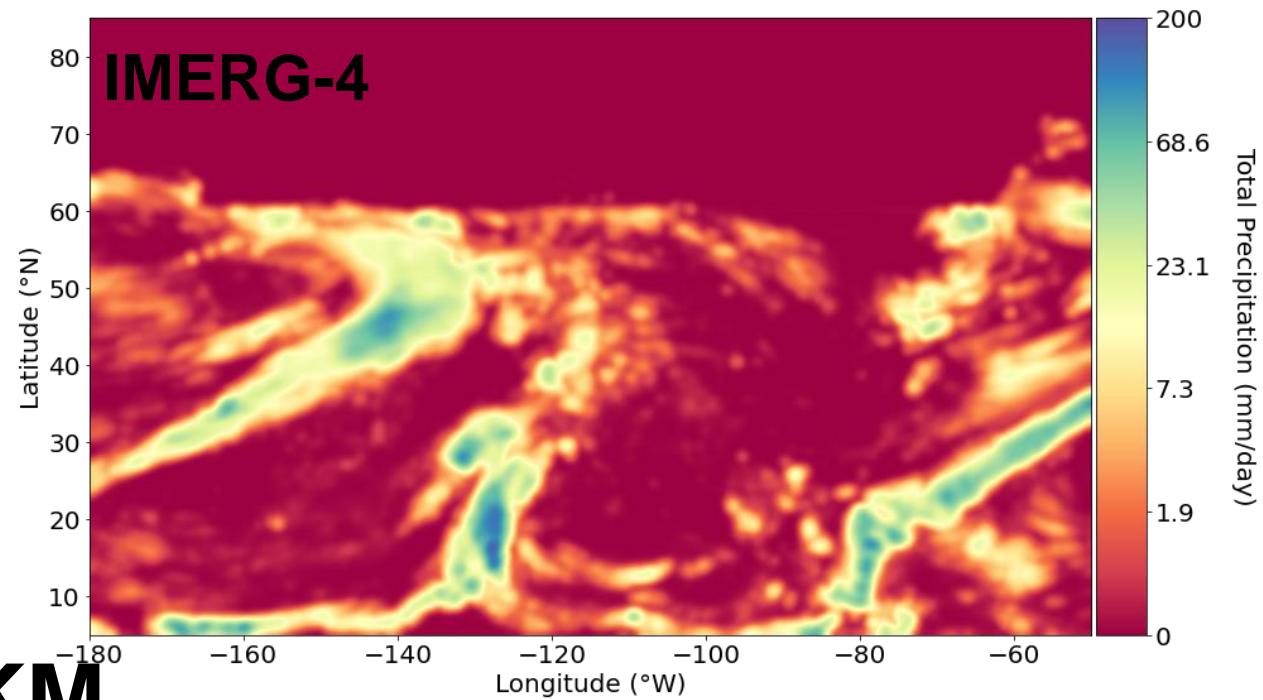
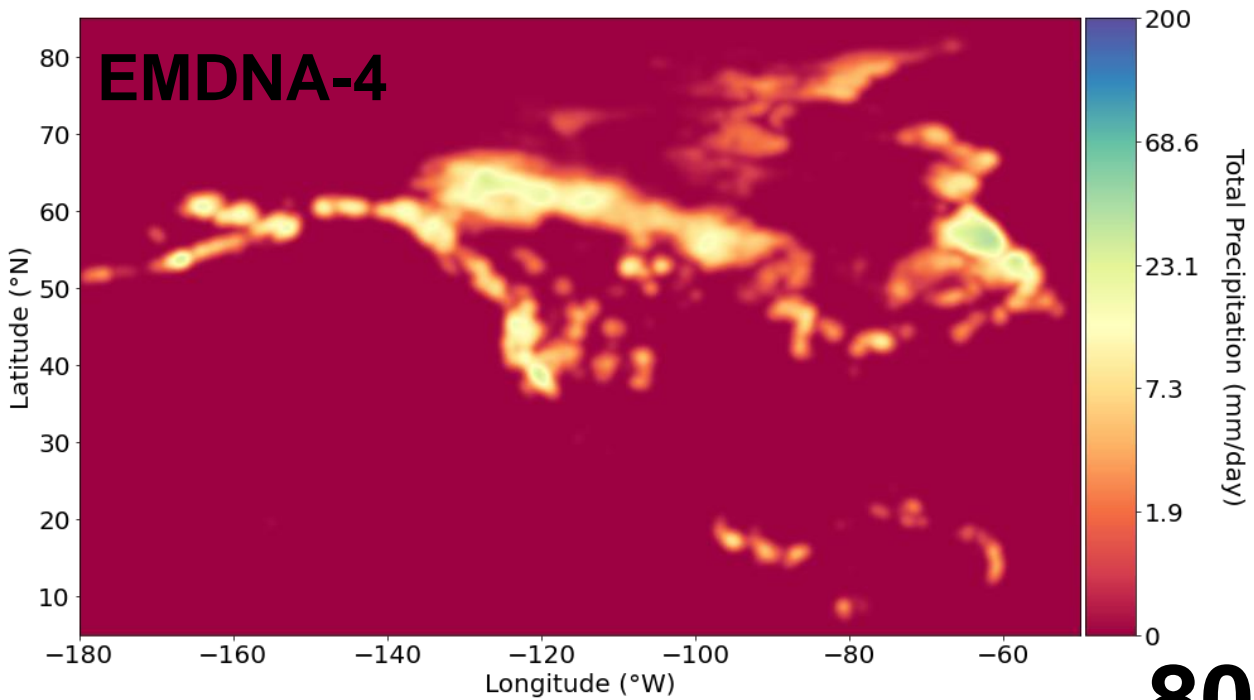
20KM



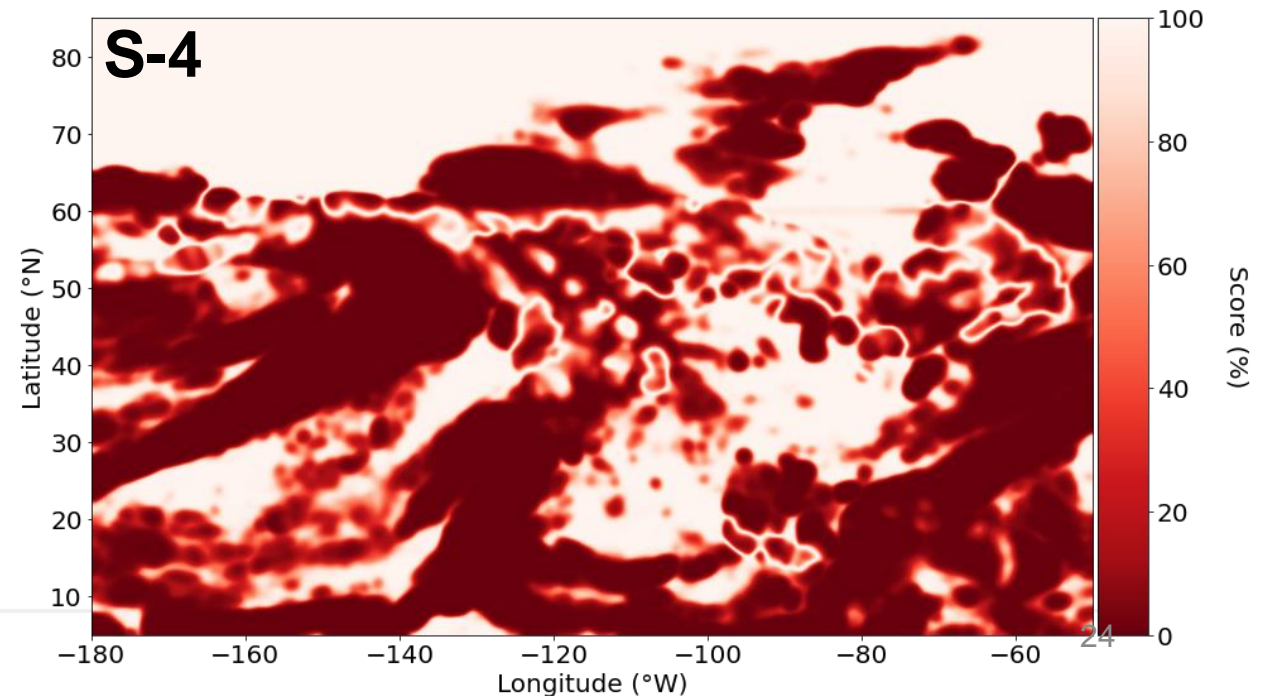
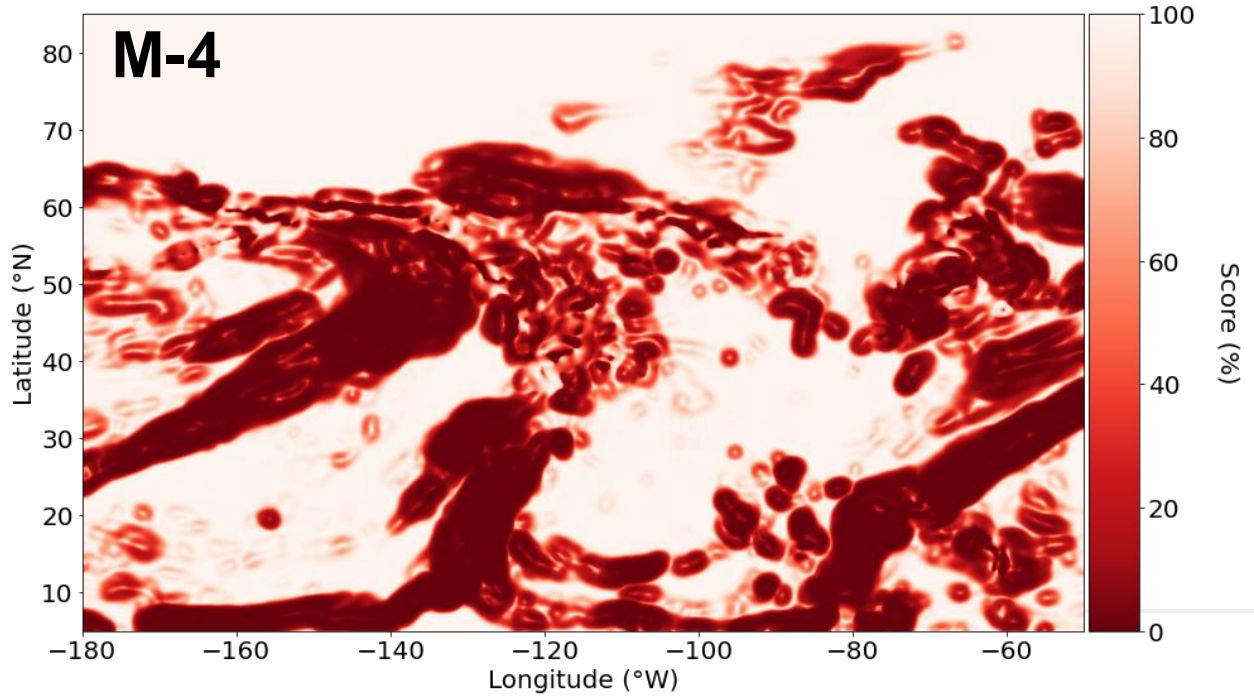


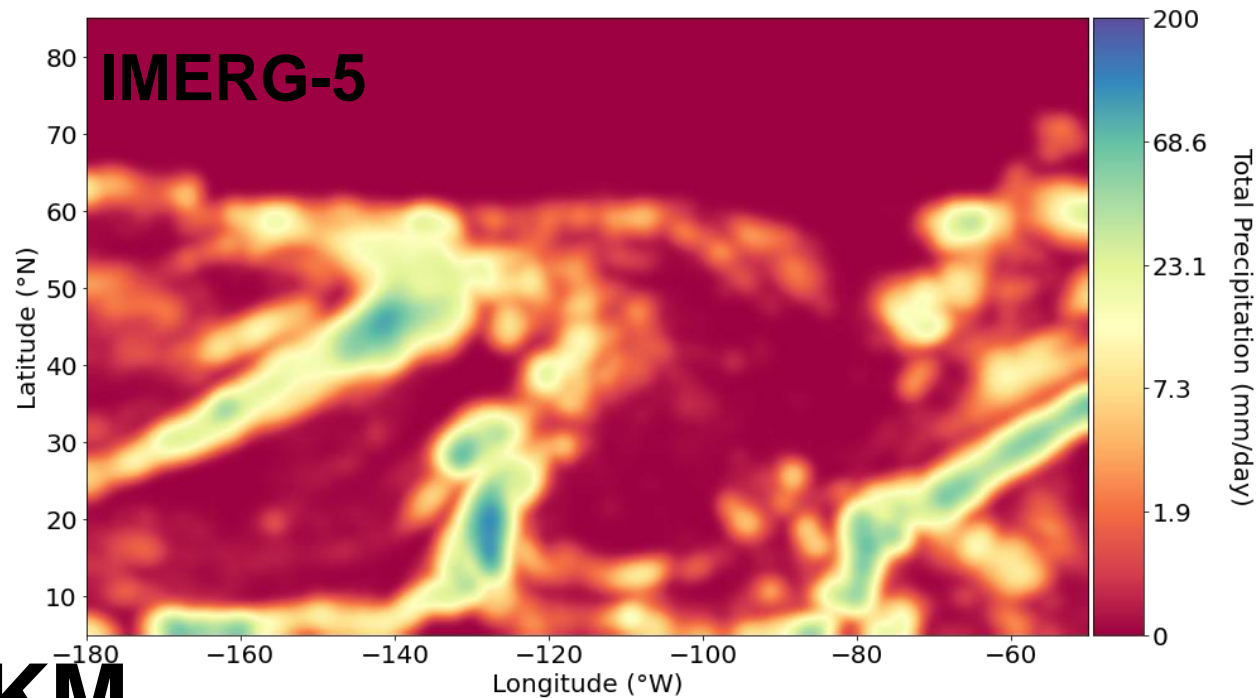
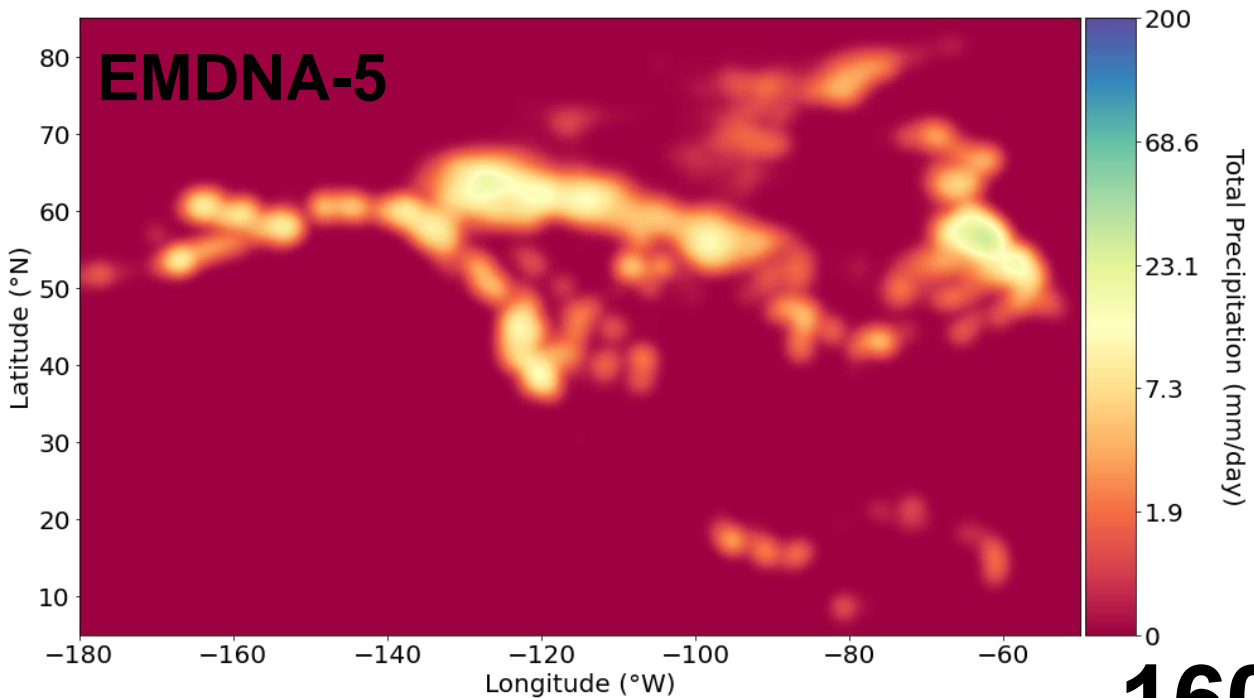
40KM



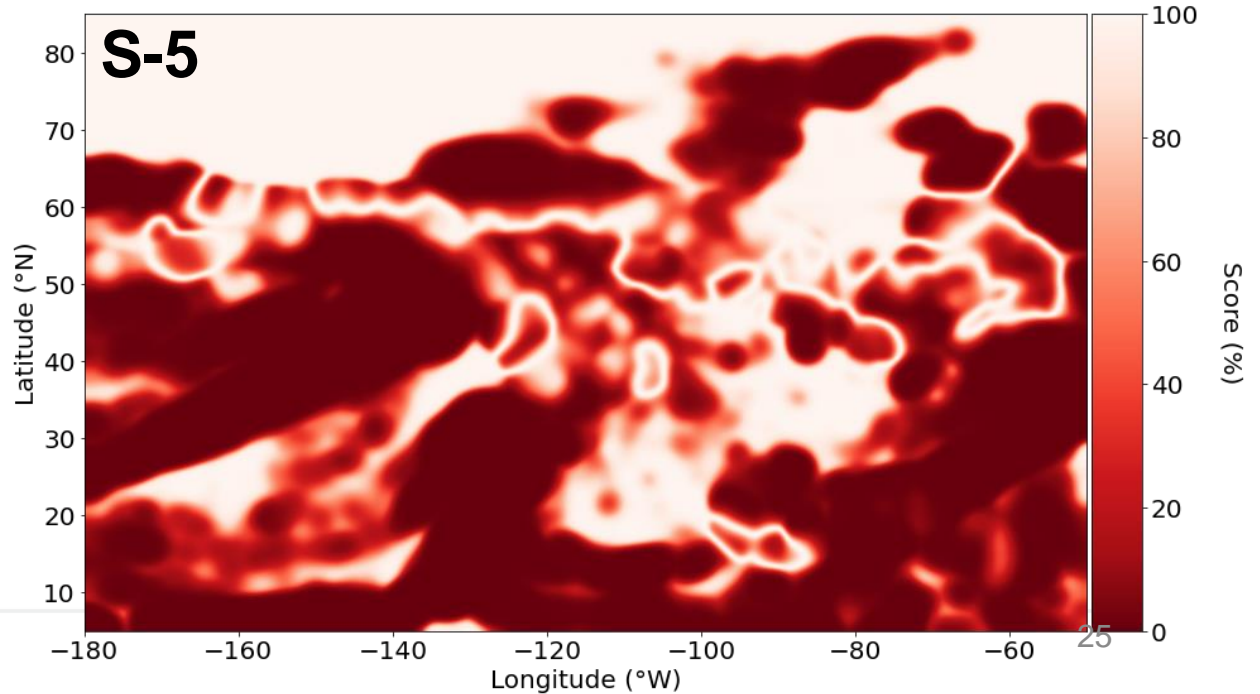
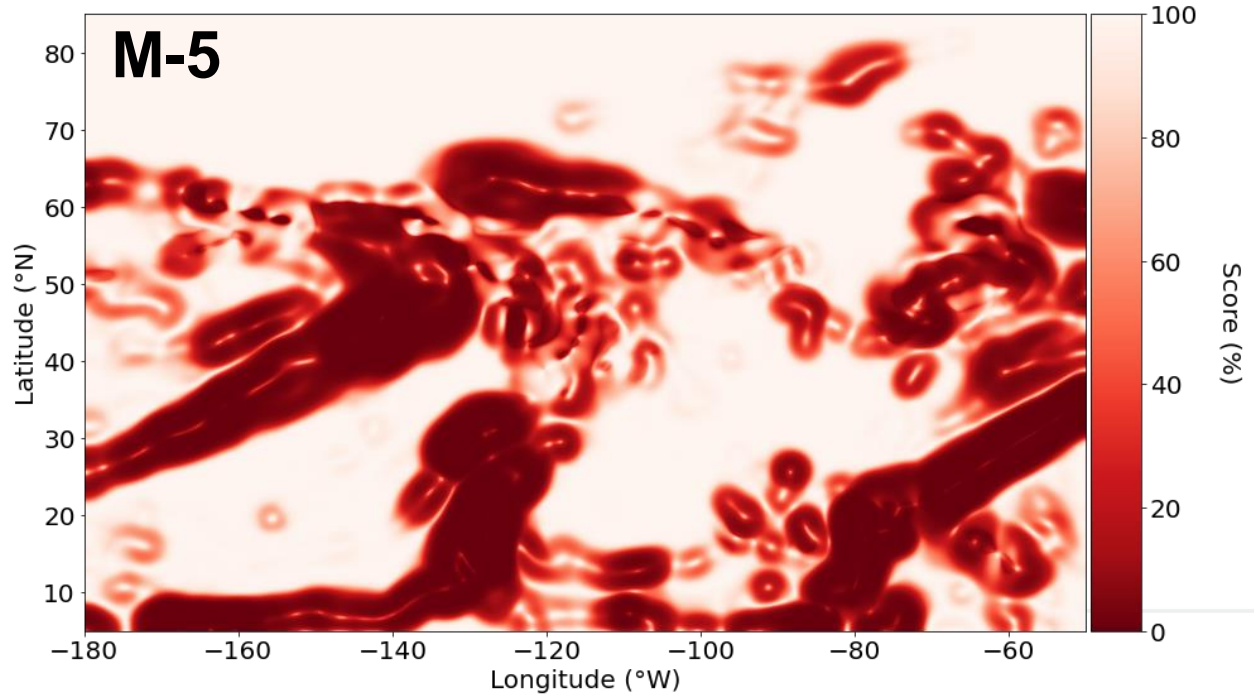


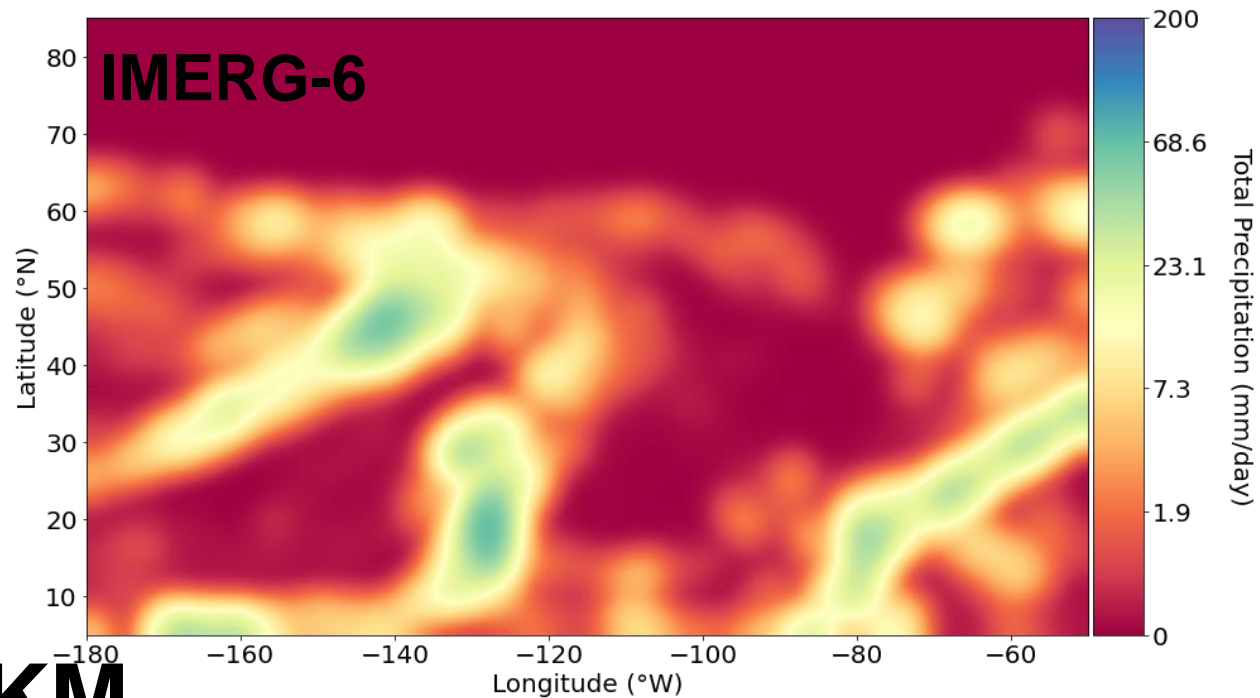
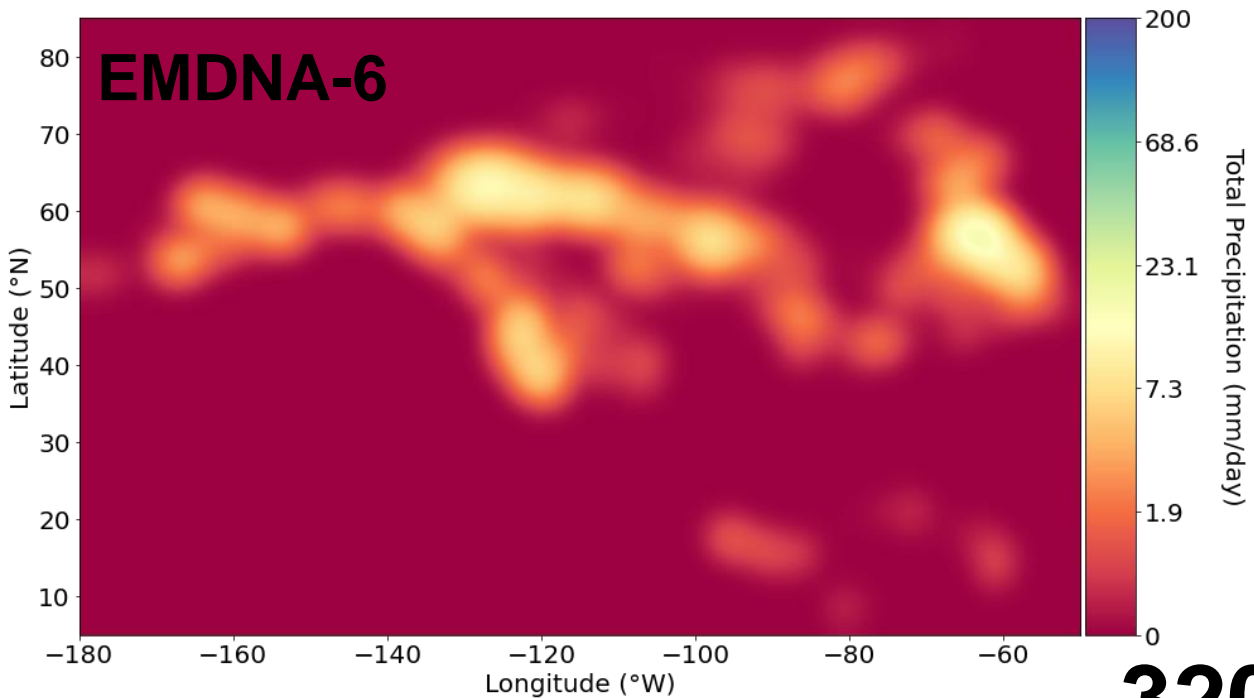
80KM



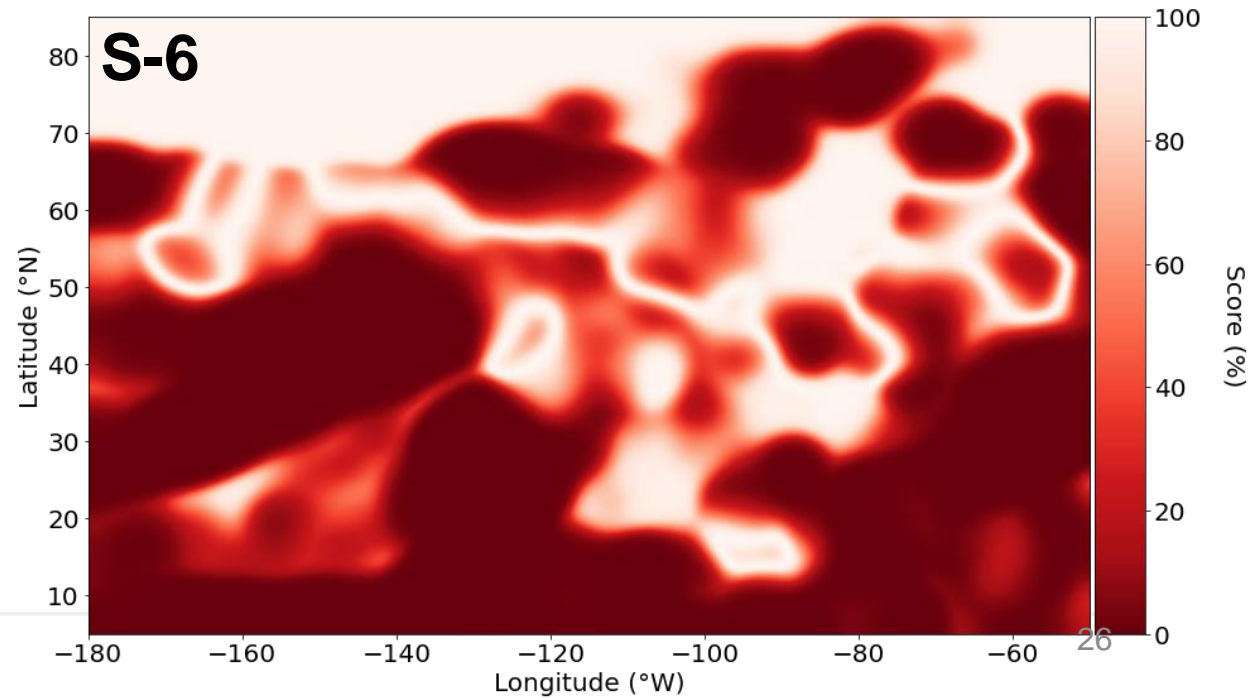
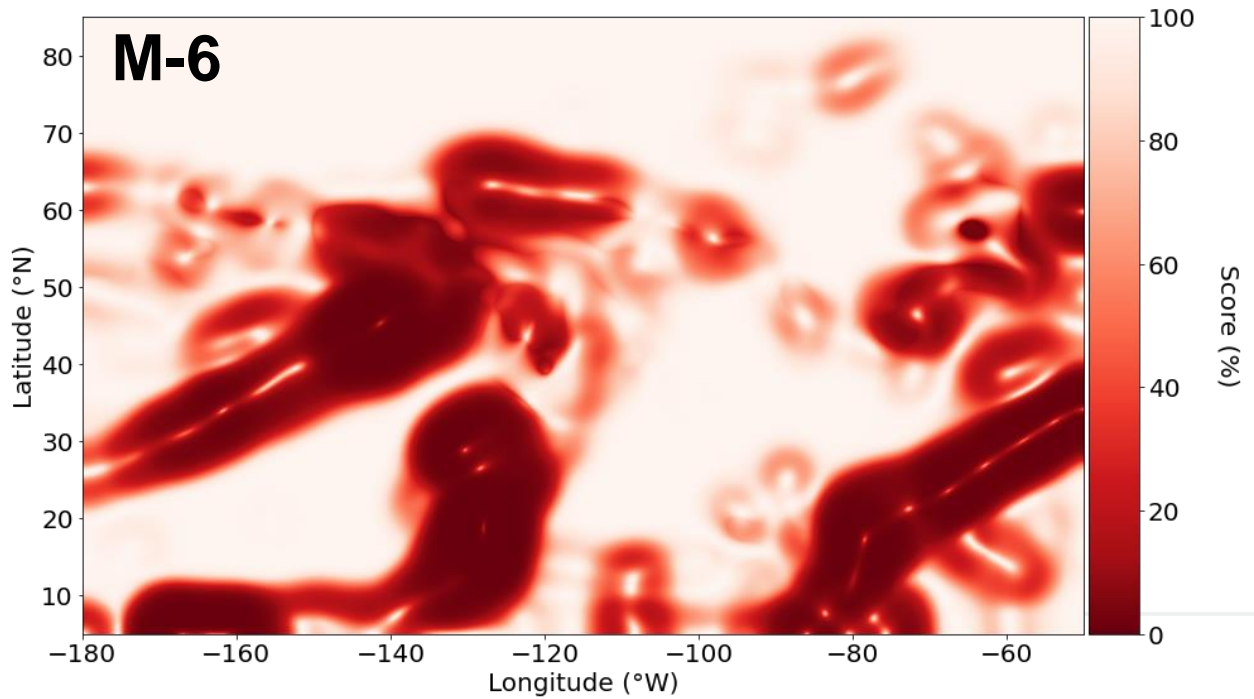


160KM

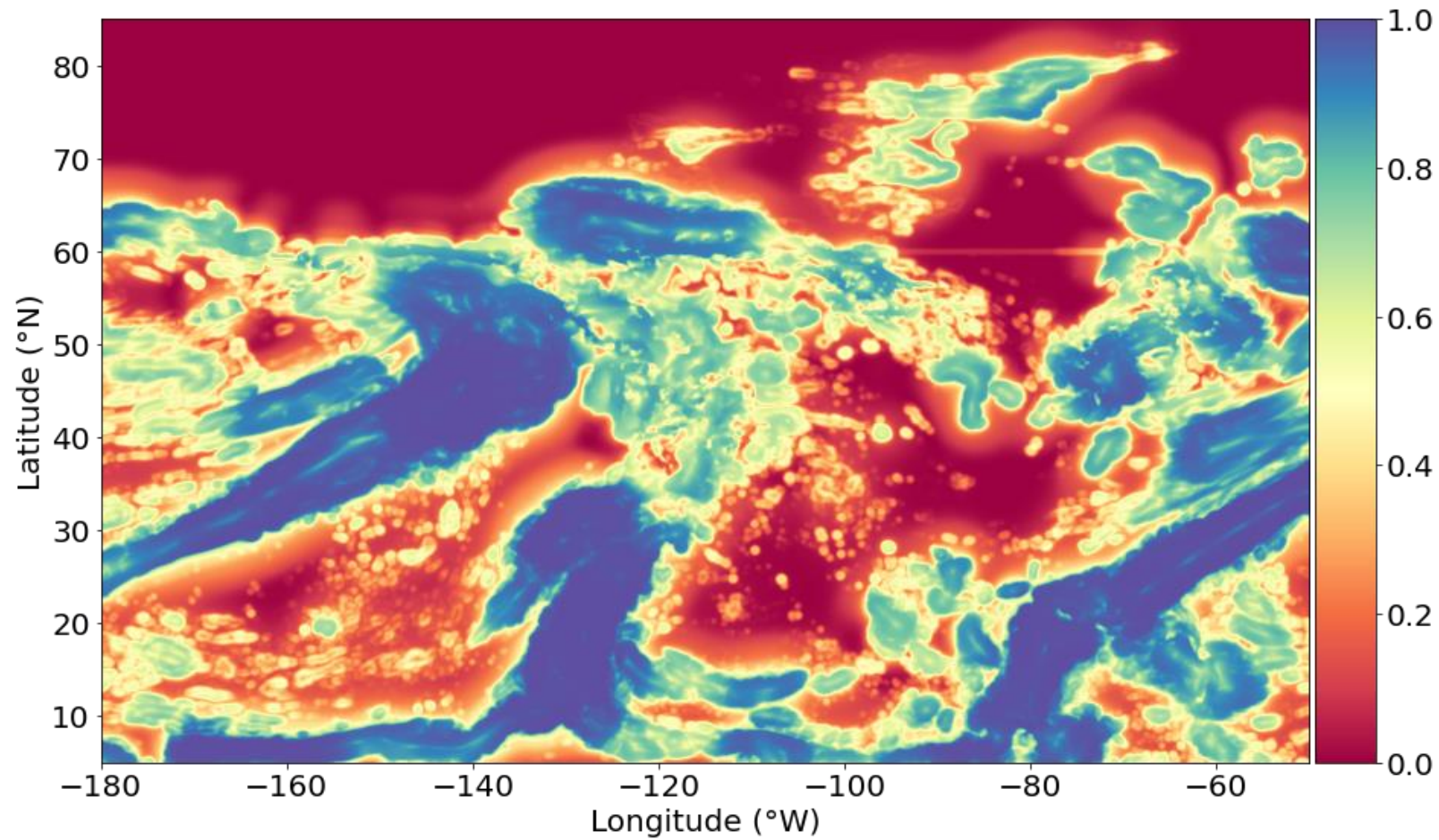




320KM



AGGREGATED SCORE



PARAMETERS TUNING

- How to choose parameters for weighting types of errors, scales and locations?

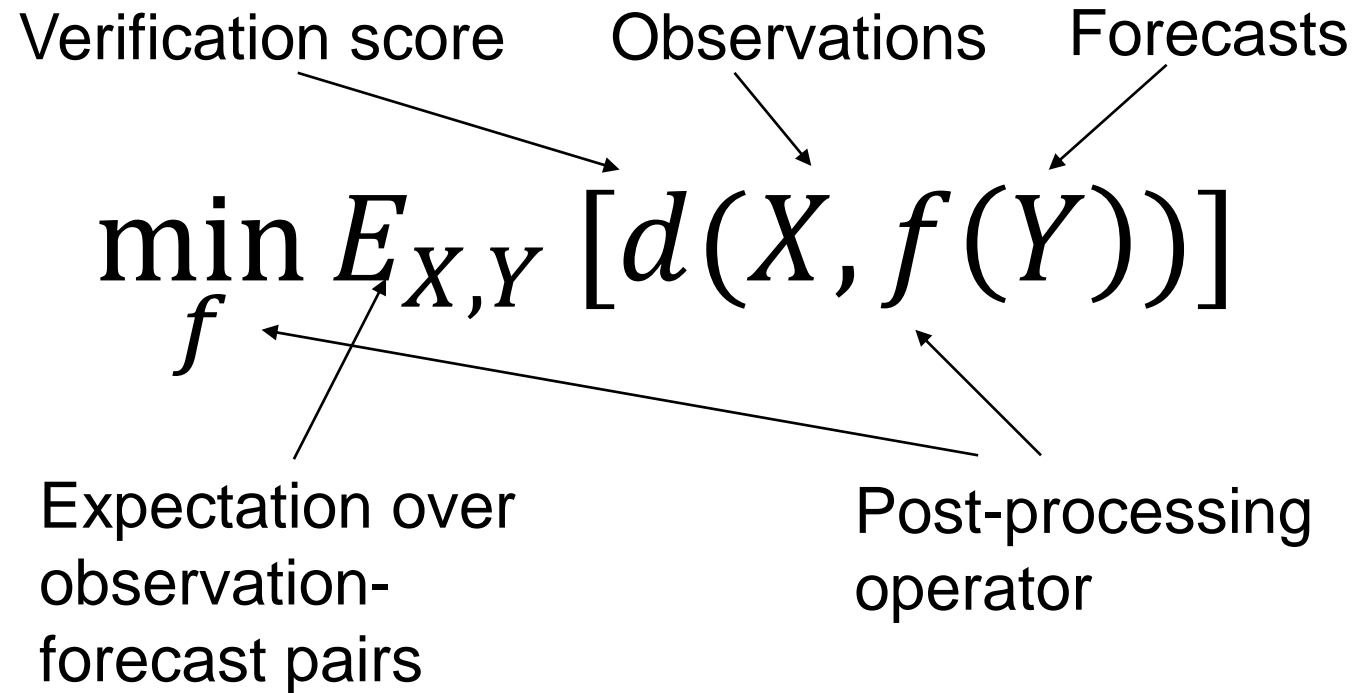
Meta-verification tuning problem:

- Fit objective score with subjective judgement

In IQA: psychophysical experiment

OPTIMIZATION

Minimum risk estimator:



EXPLICIT SPATIAL ERRORS

- Provided by metrics on motion vectors

NON-GAUSSIAN ERRORS

- Higher-moments: skewness, kurtosis

SUMMARY

Score decomposition framework:

- Multi-scale versions of scores such as MSE, NSE and KGE
- Comparison of IQA scores (SSIM) with FSS
- Tunable scores to match subjective assessment

What do we want to measure?