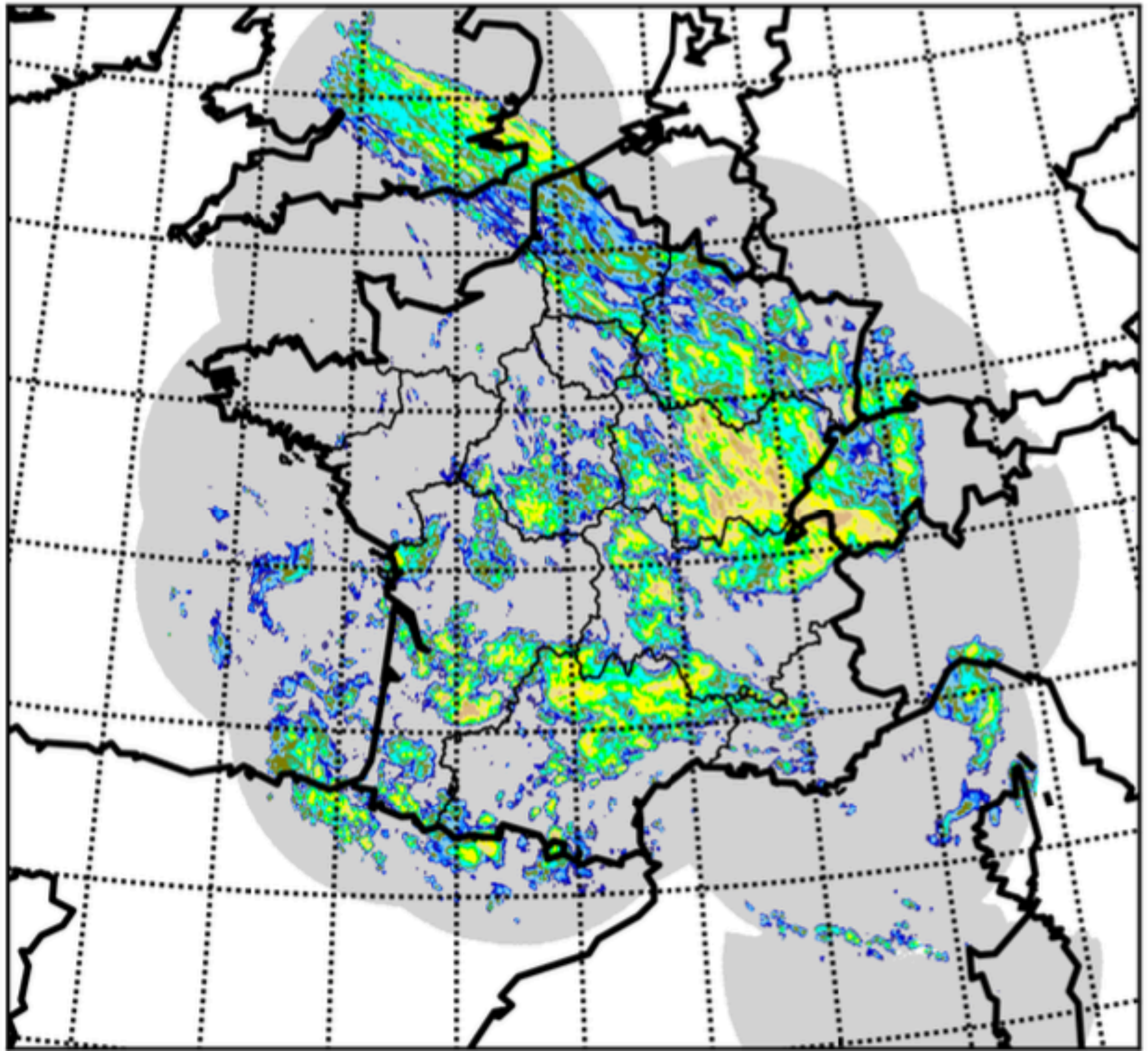
AI/DL and Nowcasting of Extremely Variable Fields

Daniel Schertzer, Hai Zhou,
Ioulia Tchiguirinskaia

École nationale des ponts et chaussées | Institut Polytechnique de Paris

USP-PPI workshop

111 March 2025

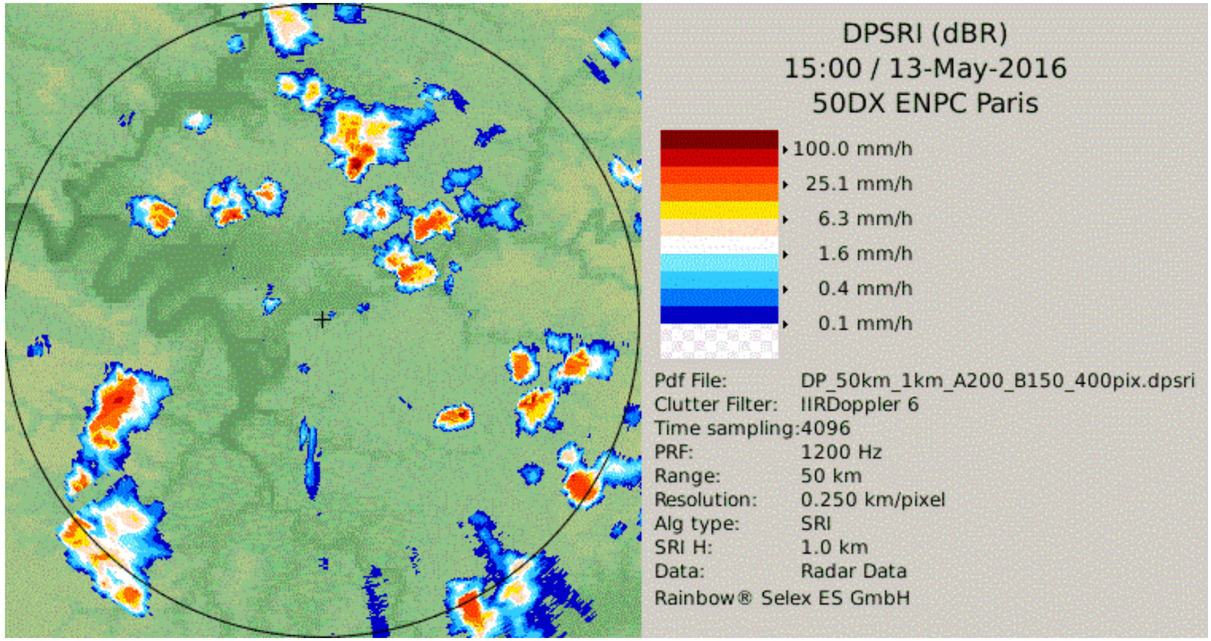


HM&Co

- **Hydrology Meteorology and Complexity**
  - geophysics/environment to be investigated as a complex system
    - HM&Co moto: **strong variability/heterogeneity** of natural and man-made environments **over a wide range of scales**
      - not only applications: this inspires/requires new complexity concepts and techniques
        - cascades
        - multifractals
        - complex networks
        - scaling anisotropy
      - more and more data of higher quality and resolution

Millenium problem of turbulence !

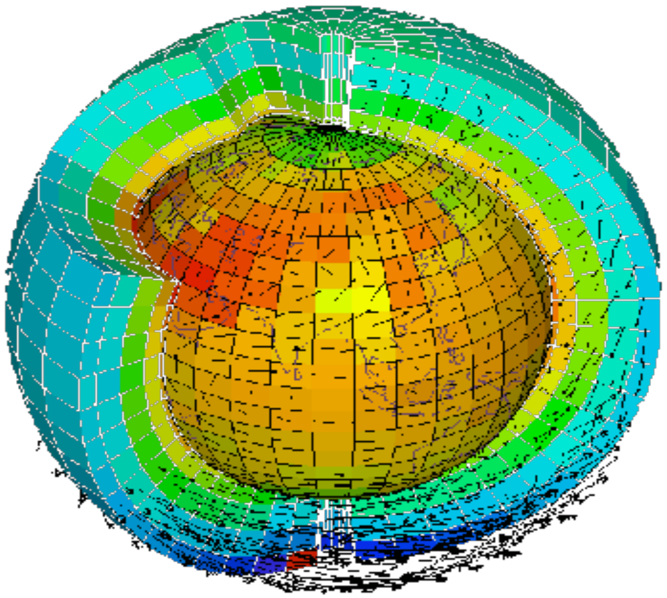
Art piece 'Windswept' (Ch. Sowers, 2012): 612 freely rotating wind direction indicators to help a large public to understand the complexity of environment near the Earth surface



Polarimetric radar observations of heavy rainfalls over Paris region during 2016 spring (100 m resolution):

- **heaviest rain cells** are much smaller than **moderate ones**
- **complex dynamics** of their aggregation into a large front

How many scales and voxels?



Computing brute force sufficient?
It requires N cubes of mm^3 to reach the ***viscous scale*** (≈1mm):
$N \approx 10^7(10^{10})^2 = 10^{27}$ >> $N_A = 10^{23}$
whereas $N_{effective} \approx 10^7 - 10^6$

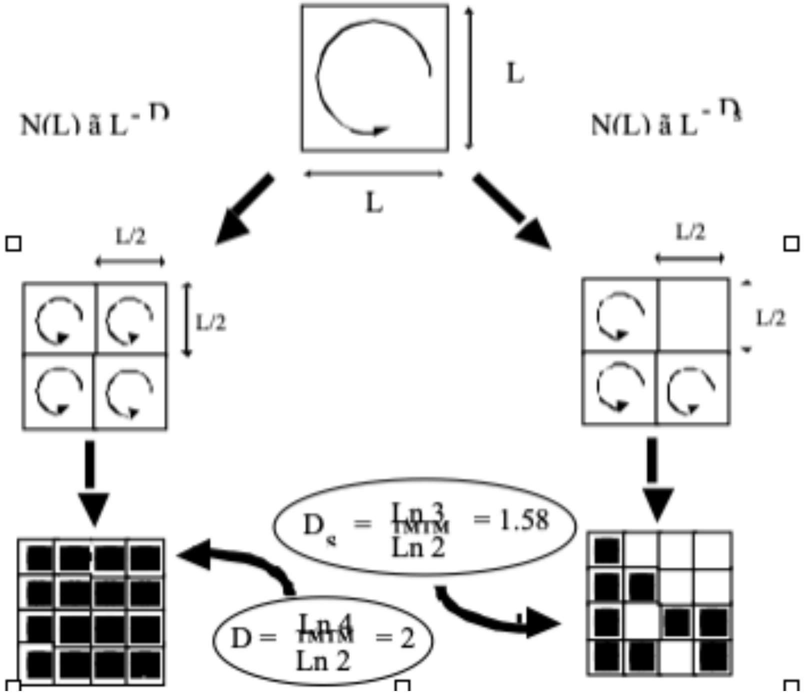
==> statistics or stochastics ?

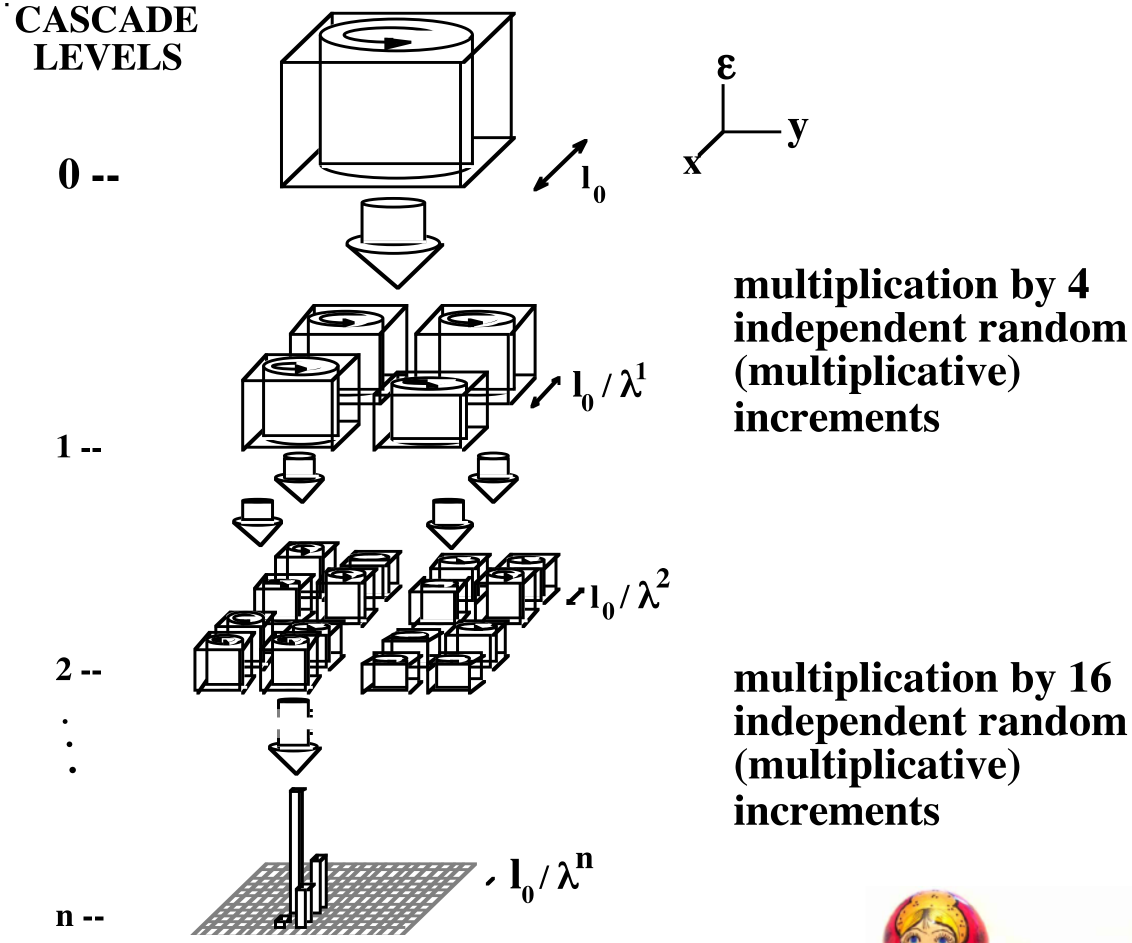
Phenomenology: multiplicative cascades

- Richardson's quatrain (1922) :

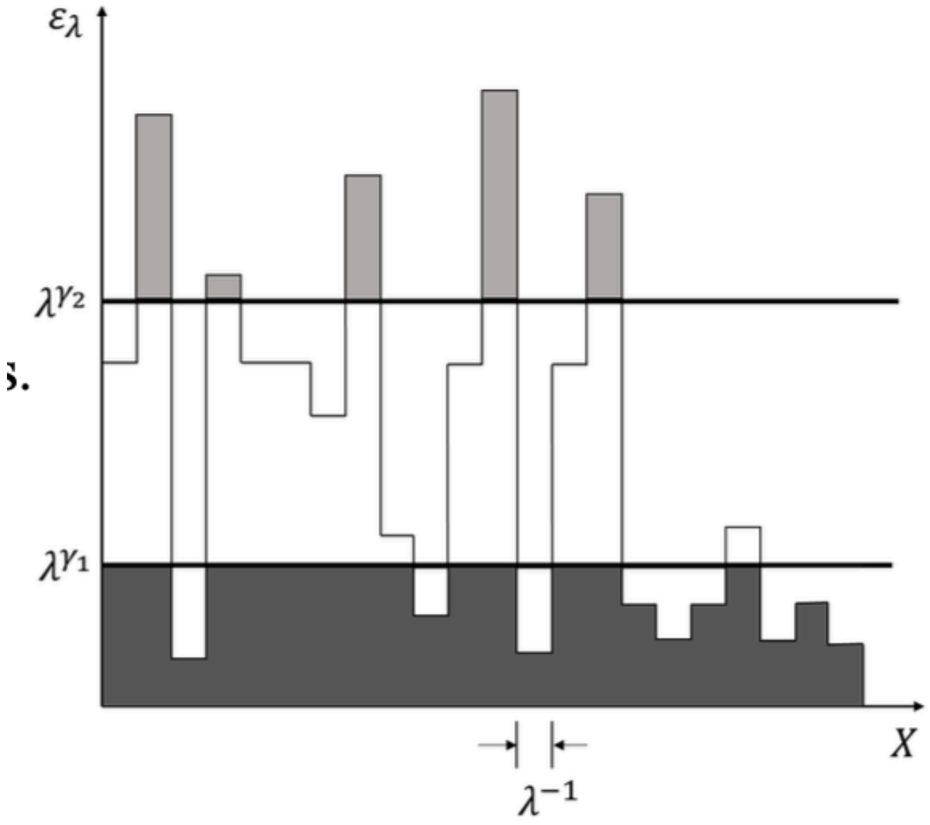
Big whirls have little whirls that feed on their velocit and little whirls have lesser have lesser whirls
and so on to viscosity…
in the molecular sense.

- discrete multiplicative cascade processes (Yaglom 1966, Mandelbrot 1974…)
- From dead/alive alternative (β-model) to weak/strong infinite hierarchy of intensities
- supported by an infinite hierarchy of fractals,
- i.e. these fields are in general **MULTIFRACTAL**
- however, multiplicative processes are not indispensable !





Universal Multifractals (UM)



A schematic illustration of a multifractal field analysed over a scale ratio λ, with two scaling thresholds, λ^{γ_1} and λ^{γ_2}, corresponding to two orders of singularity: $\gamma_2 > \gamma_1$.

- **Multifractals**: increasing **variability clusters** on smaller and smaller space-time fractions, in fact **fractal subsets**
  - => multi-scaling: $< \varepsilon_\lambda^q > \approx \lambda^{K(q)}$, the scaling function $K(q)$ is nonlinear for a wide range of resolutions $\lambda = L/\ell$
- **Universal Multifractals (UM)**: stable and attractive multifractal processes: $K(q) = \dfrac{C_1}{\alpha - 1}(q^\alpha - q)$
  - $0 < C_1$: mean intermittency, also the singularity of the mean field. The field is homogeneous for $C_1 = 0$
  - $0 \leq \alpha \leq 2$: multifractality index, measures the increase of the intermittency with deviation from the mean. The field is monofractal for $\alpha = 0$, lognormal for $\alpha = 2$.
- **Non-conservative fields**: $\varphi_\lambda =^d \varepsilon_\lambda \lambda^{-H}$ introduces the non-conservative parameter *H*, e.g. order of fractional integration/ derivation.

Varenna summer school (1983)

- primary version of the multifractal formalism of Parisi and Frisch (1985) presented at “Turbulence and Predictability in Geophysical Fluid Dynamics” organised by M. Ghil, R. Benzi et G. Parisi
- clustering of higher activity on smaller spacetime fractions
- but: “**Still the multifractal model appears to be somewhat more restrictive than Mandelbrot’s weighted-curdling model which does include the logrnormal case**”.
- the conference proceedings (1985) refers to S+L (1984)):
  - a small perturbation of the ß-model is no longer limited to a unique dimension ($\alpha-$model)
  - the divergence of higher order moments is rather generic in cascade models
  - the later introduces spurious scaling, an analytical approximation depending on a unique scaling exponent *H* and the critical order α was proposed:
  - it was shown to fit the experimental points from Anselmet et al. (1983), see fig. 1 with:

$$\xi(p) = pH + \theta(p-\alpha)(1-p/\alpha) \qquad H = 1/3, \alpha = 5, 5.5, 6$$

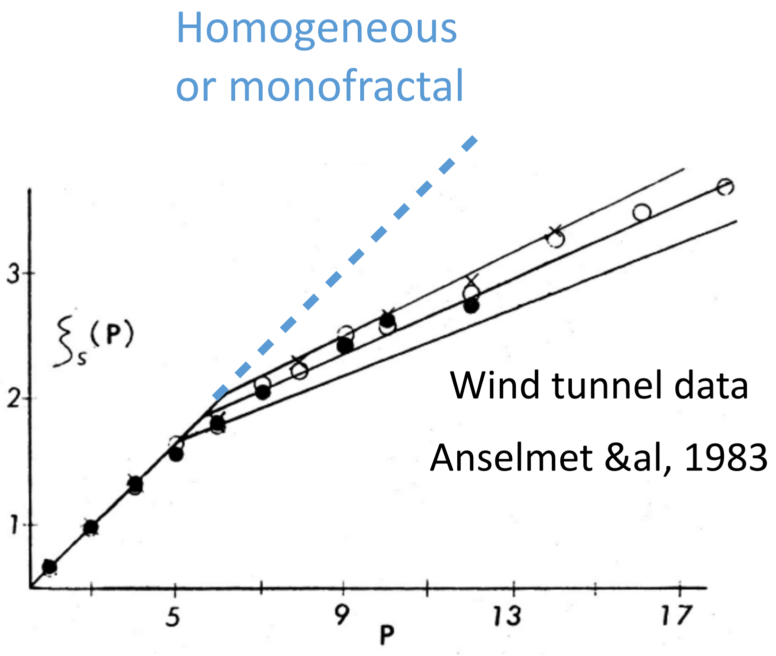
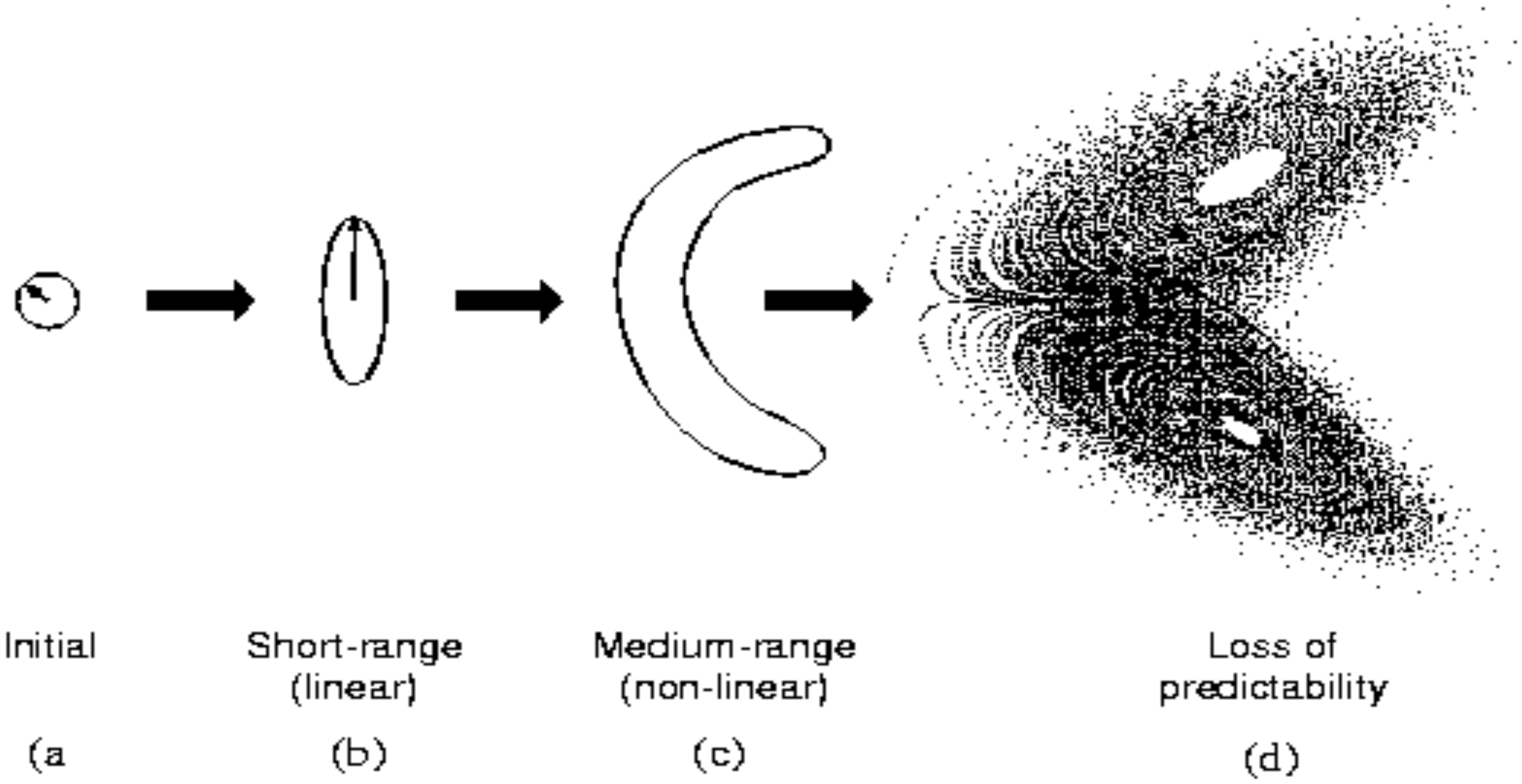


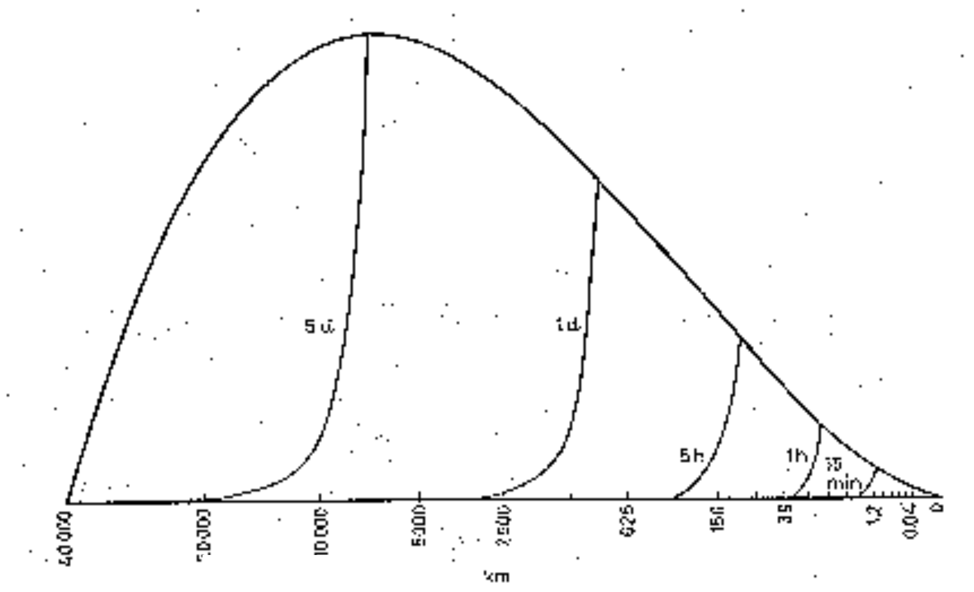
Fig.1 from S+L (1984)

Butterfly effect and ensemble predictions



Scheme of the evolution of the empirical pdf evolution of an Ensemble Prediction System (EPS), according to Palmer,1999: from the phase space region occupied by the initial ensemble (a), to (b) linear growth phase, to (c) nonlinear growth phase, to (d) loss of predictability

Spectral analysis of space-time predictability



Flux from correlated e^c to decorrelated energy e^Δ

$$\ell_c = 1/k_c \approx t^{3/2}$$

Lorenz (1969)
Leith and Kraichnan(1972)
Metais and Lesieur (1986)

$$e^c(\underline{x},t) = \underline{u}^2(\underline{x},t).\underline{u}^1(\underline{x},t)$$

$$e^\Delta(\underline{x},t) = \tfrac{1}{2}\left(\underline{u}^2(\underline{x},t) - \underline{u}^1(\underline{x},t)\right)^2$$

Similar results with turbulence phenomenology:

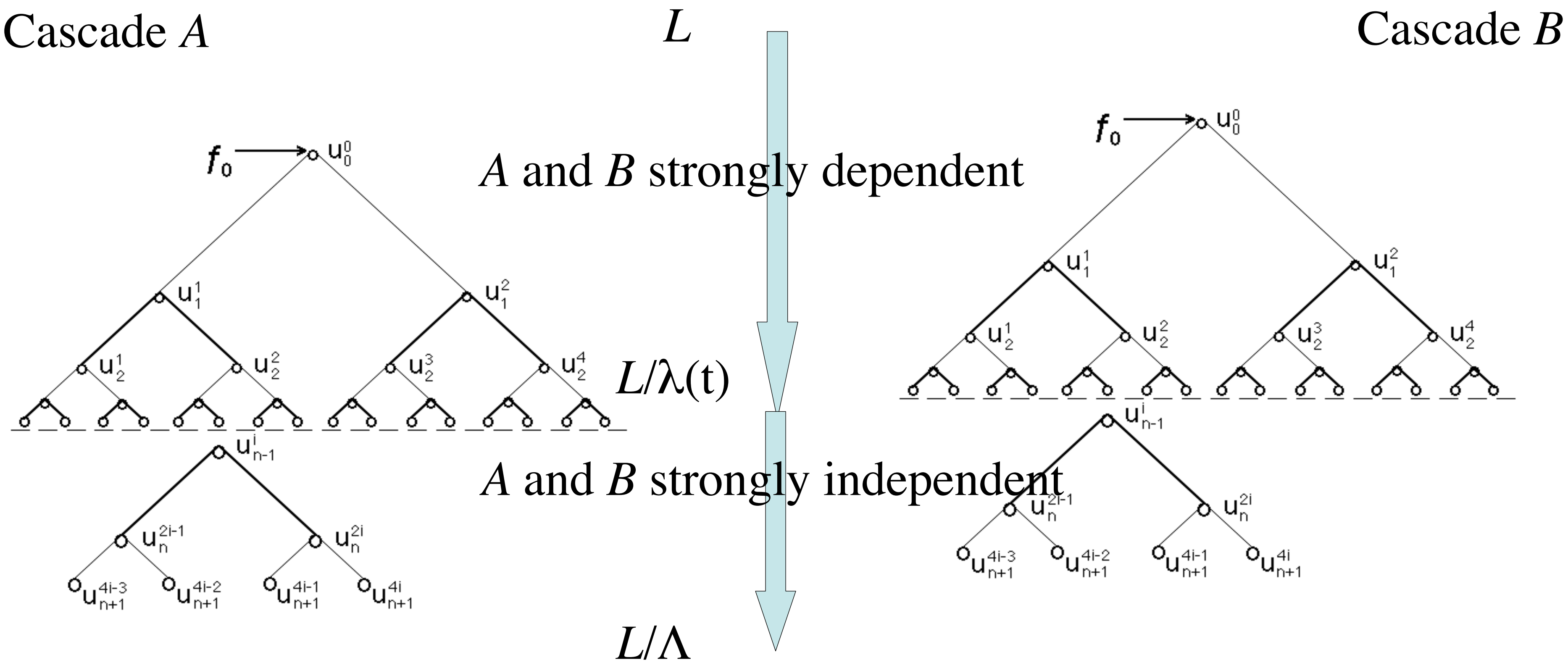
$$\ell \approx \bar{\varepsilon}^{1/2} t^{3/2};\ \bar{\varepsilon} \approx 10^{-3} m^2 s^{-3}, \eta = 10^{-3} m$$

Multifractal Predictability

Rough idea:

i) relaxation of (common) past structures ==> flux of the past

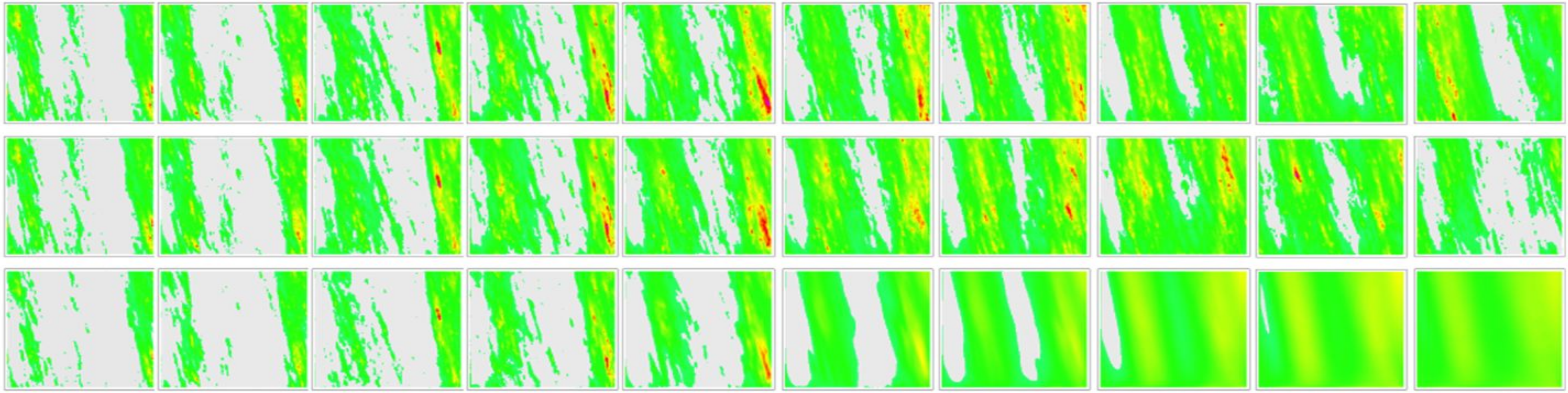
ii) (new) independent structures ==> flux of the future



Multifractal Predictability

Rain simulation (α=1.5, C_1=0.2, H=0.1 on log scale. Realizations A, B are identical until t=0, then they diverge.

Top: Realization A. Middle: Realization B. Bottom, forecast



•**Power law divergence** between the realizations A and B,
=> **irrelevance** of the **finite** dimensional **'LE + MET' scenario** !
•**Drastic loss of variability of forecast C** with deterministic sub-grid modeling (based on the conservation of the flux) => 'baby theorem': **stochastic sub-grid modeling** does much better than deterministic one!

(Schertzer and Lovejoy,Physica A 2004)

Rainfall time series prediction: VMD-RNN

Step1: **Divide the original rainfall time series** into a training set and a non-training set.

Step2: **Decompose the training set** into K sub-sequences using VMD.

Step3: **Sequentially append the non-training data** to the training set to generate new appended sequences and **repeat decompose** each append sequence into K sets of appended sub-sequences.

Step4: **Exact the last sample** of each set of appended sub-sequences as a non-training sample.

Step5: For each sub-sequences, **train four variant RNN models and tune hyperparameters** to find an ideal predicting model with optimal parameters.

Step6: For each sub-sequences, input testing samples into the correspond predicting models, and **obtain individual predicted results** $y_i(t)$.

Step7: Aggregate the predicting results of each sub-sequences to **generate the final predicted result** $y(t) = \sum_{i=k}^{K} y_i(t)$.

Step8: Use the framework of **UM to analyze** the predicted and actual time series in the testing samples.

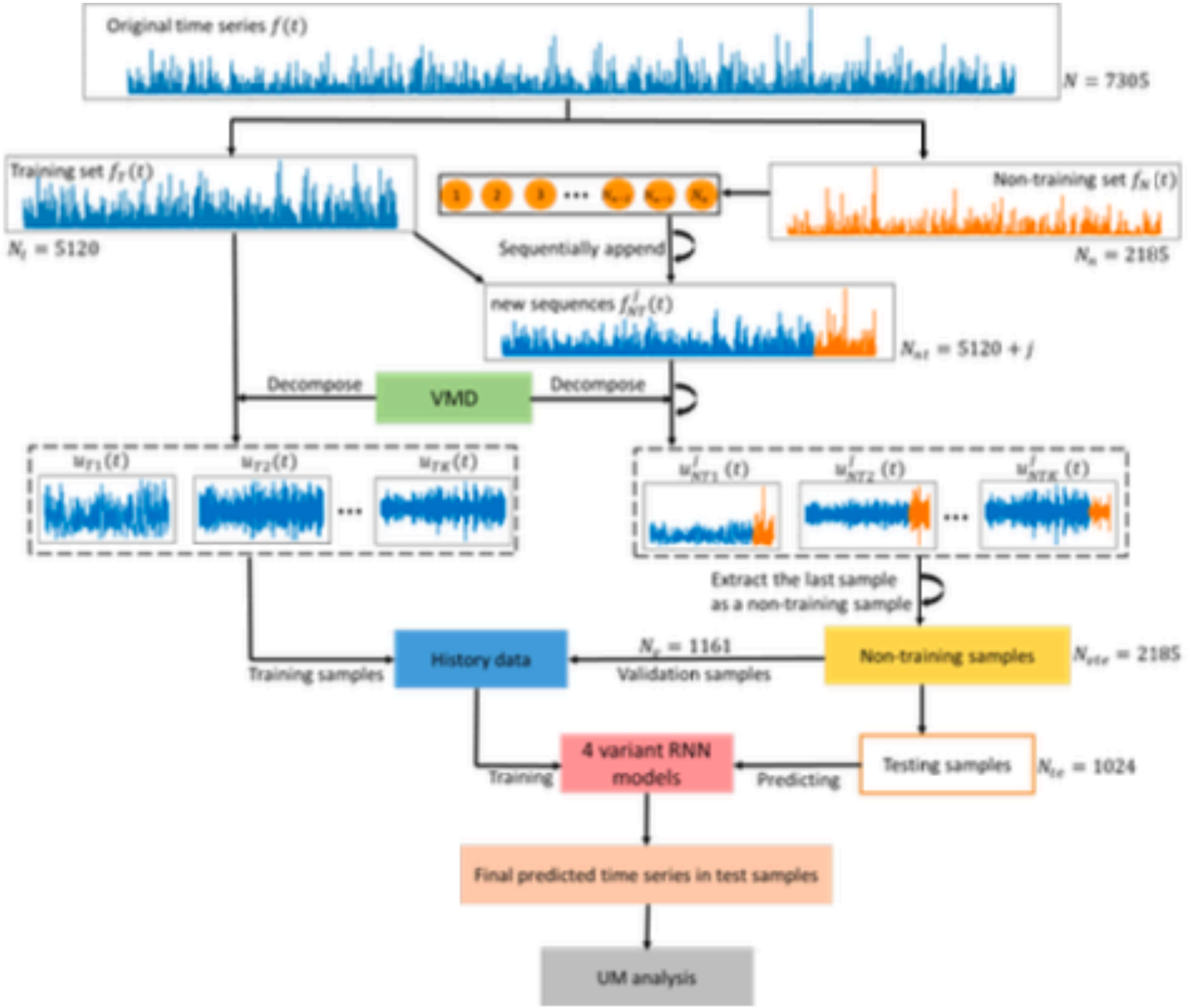


Figure 22: The process of the VMD-RNN model

(Zhou et al, 2022)

Rainfall time series prediction: VMD-RNN

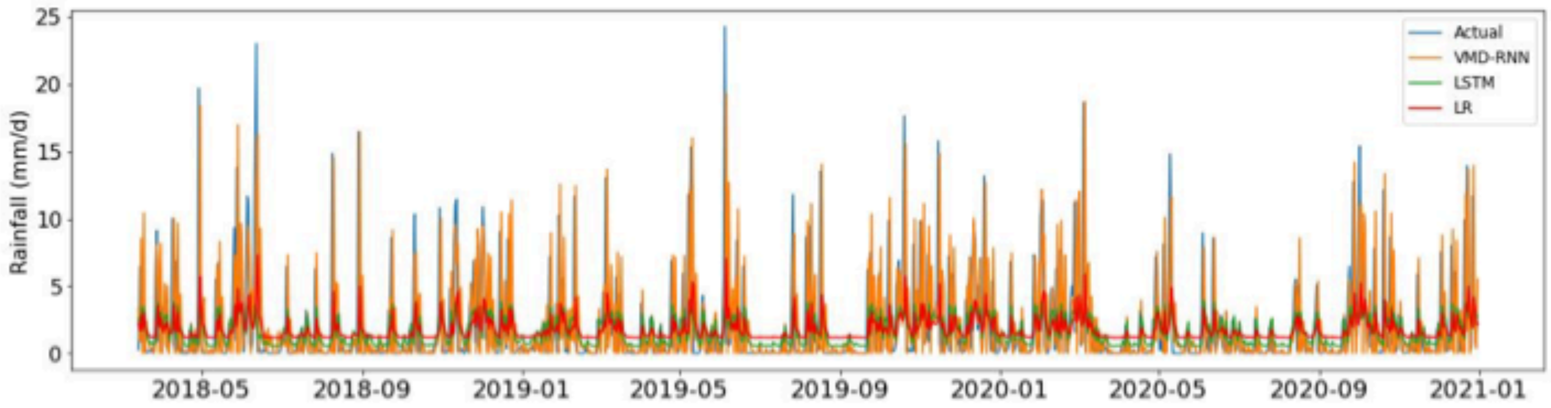


Figure 25: Predicted and actual daily time series in the testing set

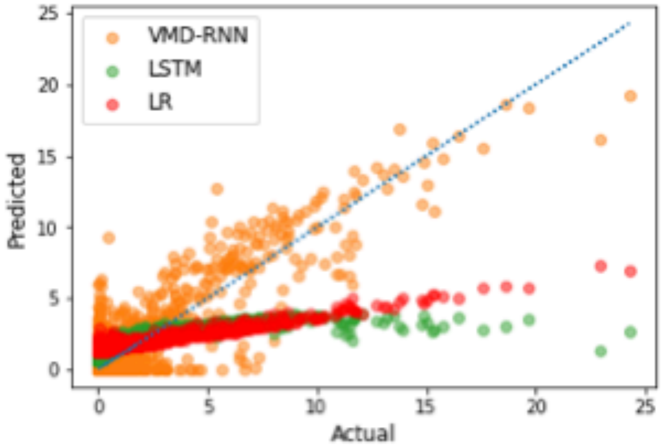


Figure 26: The comparison between predicted and actual daily rainfall values

Table 4: Prediction errors for daily time series in the testing set

	MAE	RMSE	MAPE
VMD-RNN	0.726	0.852	9.853
LSTM	6.825	2.612	10.475
LR	9.239	3.040	18.923

- VMD-RNN model has better performance in predicting high and low values, compared with the pure LSTM and linear regression
- VMD-RNN model has lowest prediction errors in MAE, RMSE, MAPE

Rainfall time series prediction: VMD-RNN

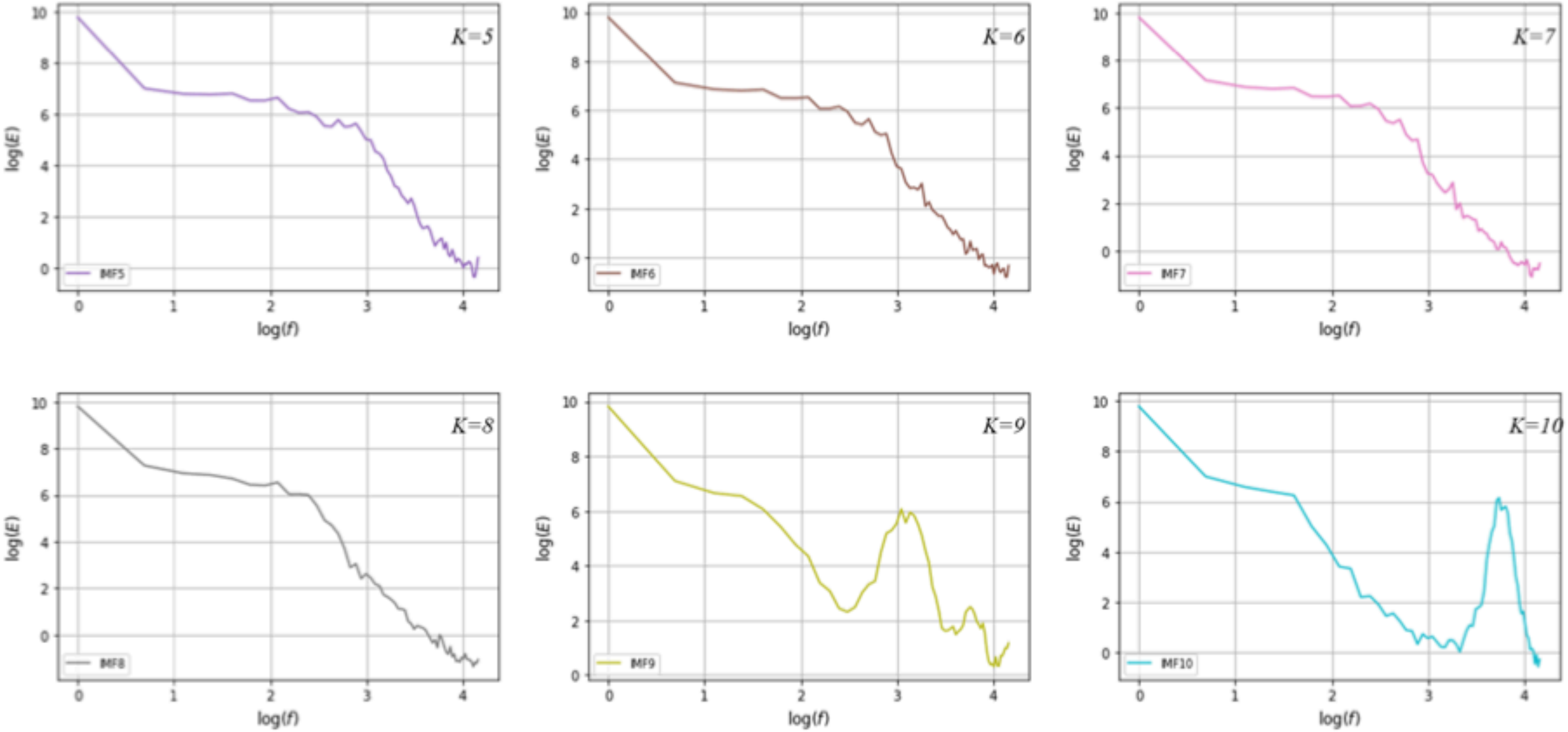


Figure 23: PSD of the corresponding last sub-sequence when K from 5 to 10

Methodologies: multifractal time series analyses

➢ Trace Moment (TM):

i. Calculate the empirical statistical moment $\langle \varepsilon_\lambda^q \rangle$
ii. Plot the logarithm of $\langle \varepsilon_\lambda^q \rangle$ versus the logarithm of λ
iii. Perform linear regression to obtain $K(q)$
iv. $K'(1) = C_1$ and $K''(1) = \alpha C_1$

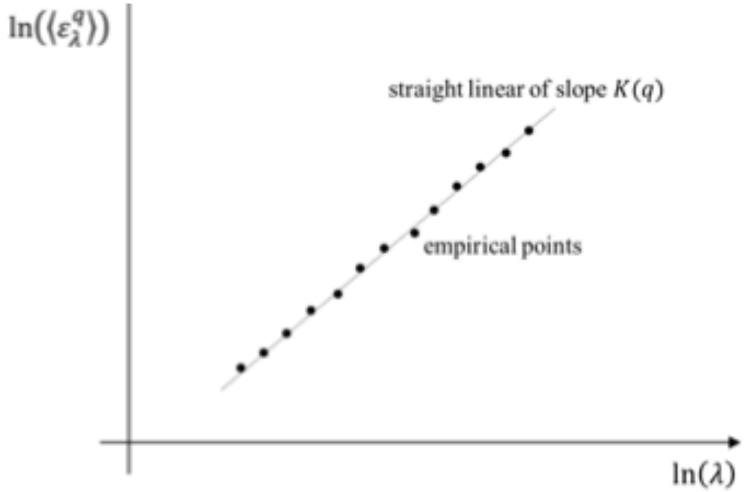


Figure 17: Illustration of the Trace Moment technique (Image source: Gires, 2012)

➢ Double Trace Moment (DTM):

i. $\varepsilon_\lambda^{(\eta)}$ is renormalized by upscaling the η-power of the field at maximum resolution
ii. $\left\langle \varepsilon_\lambda^{(\eta)q} \right\rangle \approx \lambda^{K(q,\eta)}, K(q,\eta) = \eta^\alpha K(q)$
iii. C_1 and α are obtained by the slope and intercept of the linear portion of the log-log plot of $K(q,\eta)$ vs η

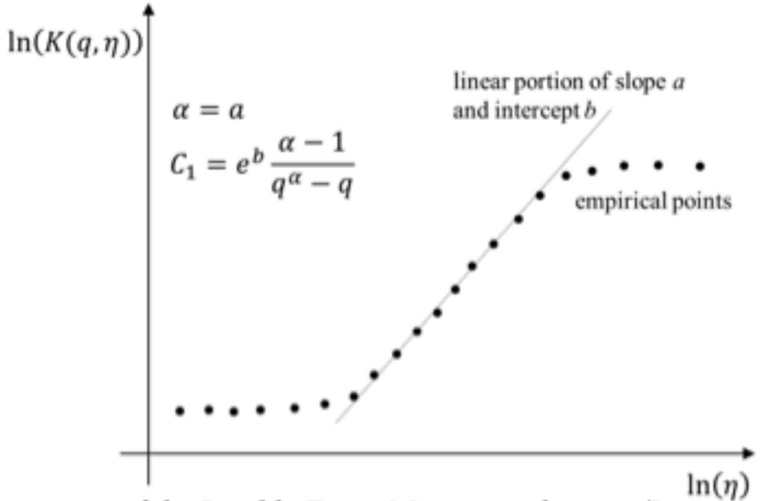


Figure 18: Illustration of the Double Trace Moment technique (Image source: Gires, 2012)

Rainfall time series prediction: VMD-RNN

actual time series

predicted time series by VMD-RNN

predicted time series by LSTM without decomposition

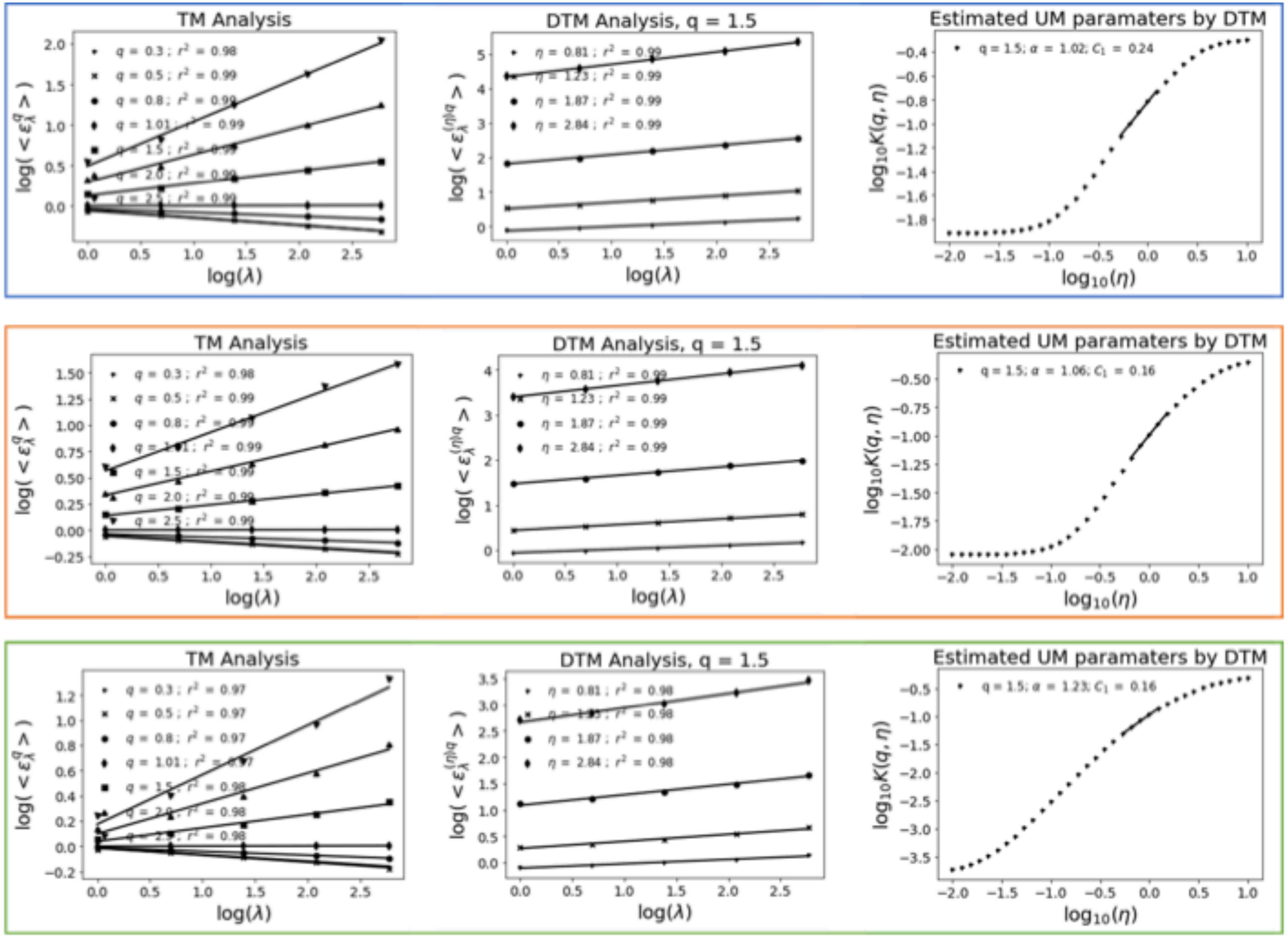
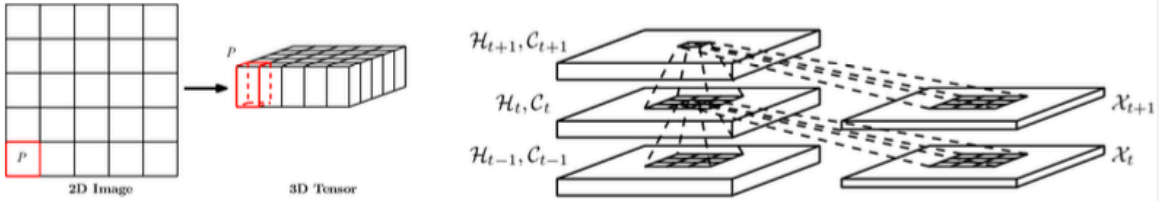
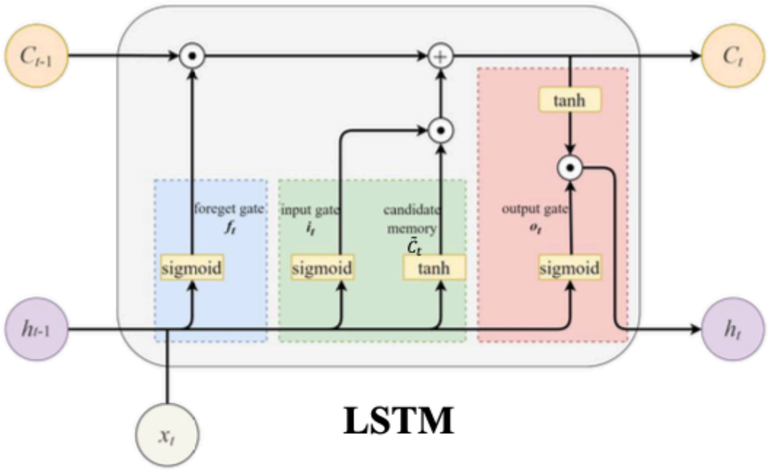


Figure 27: UM results for daily time series in the testing set

Methodologies: LSTM and ConvLSTM

spatio-temporal prediction:





LSTM

$$f_t = \sigma\left(W_{xf}\ x_t + W_{hf}\ h_{t-1} + b_f\right)$$

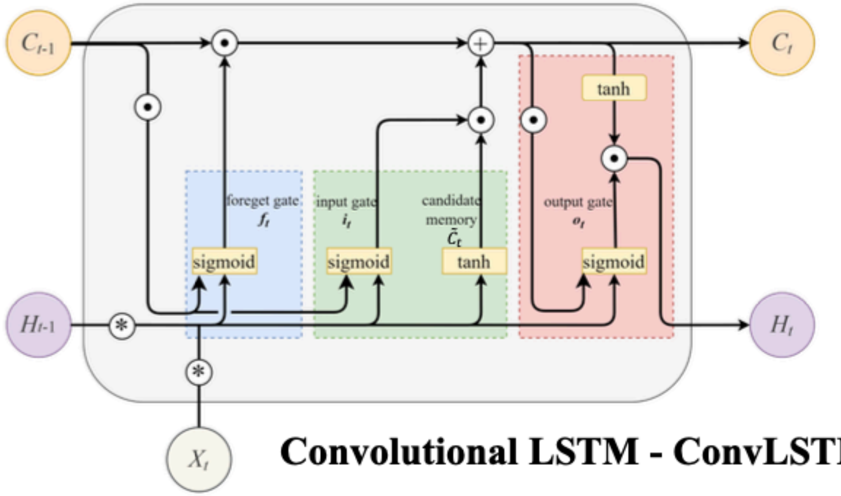
$$i_t = \sigma(W_{xi}\ x_t + W_{hi}\ h_{t-1} + b_i)$$

$$\tilde{C}_t = \tanh(W_{xc}\ x_t + W_{hc}\ h_{t-1} + b_c)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t$$

$$o_t = \sigma(W_{xo}\ x_t + W_{ho}\ h_{t-1} + b_o)$$

$$h_t = o_t \odot \tanh(C_t)$$



Convolutional LSTM - ConvLSTM

$$f_t = \sigma\left(W_{xf} * \mathcal{X}_t + W_{hf} * \mathcal{H}_{t-1} + W_{cf} \odot C_{t-1} + b_f\right)$$

$$i_t = \sigma(W_{xi} * \mathcal{X}_t + W_{hi} * \mathcal{H}_{t-1} + W_{ci} \odot C_{t-1} + b_i)$$

$$\tilde{C}_t = \tanh(W_{xc} * \mathcal{X}_t + W_{hc} * \mathcal{H}_{t-1} + b_c)$$

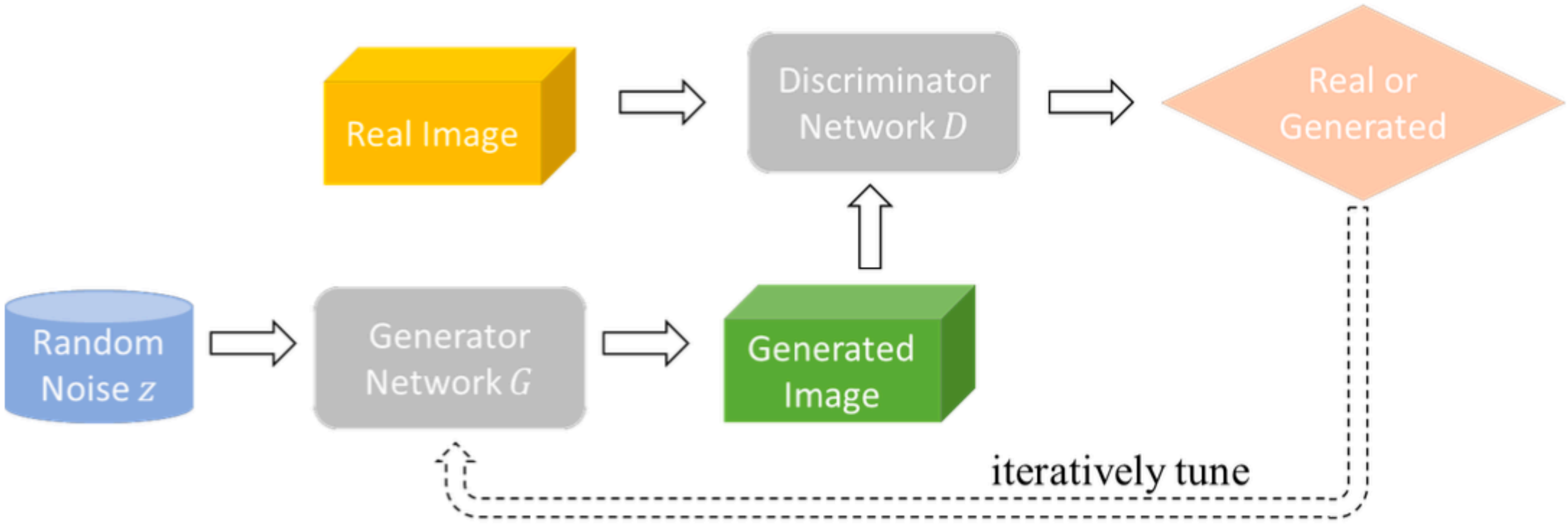
$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t$$

$$o_t = \sigma(W_{xo} * \mathcal{X}_t + W_{ho} * \mathcal{H}_{t-1} + W_{co} \odot C_t + b_o)$$

$$\mathcal{H}_t = o_t \odot \tanh(C_t)$$

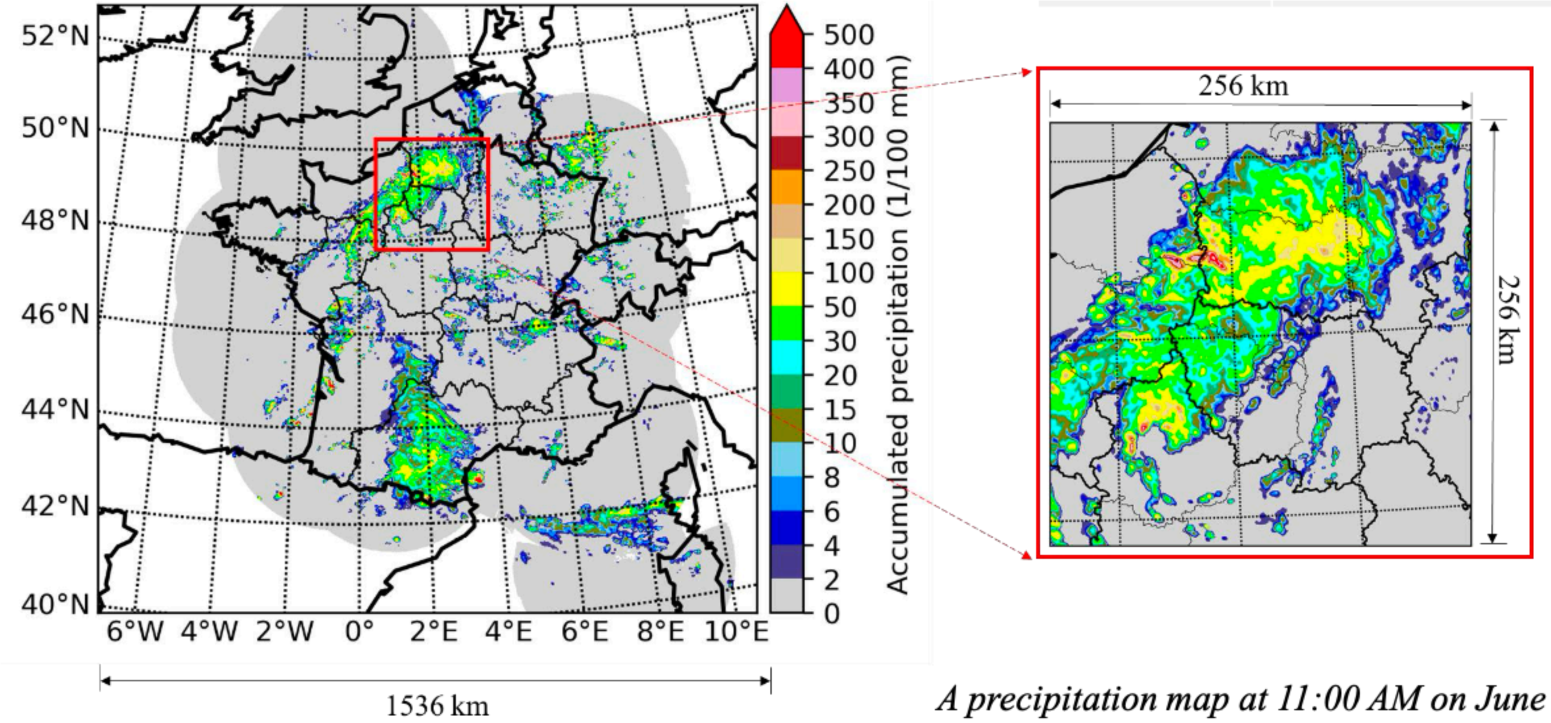
Methodologies: Generative Adversarial Networks (GAN) model

- **Generative:** generate synthetic data
- **Adversarial:** a generator and a discriminator compete against each other, zero-sum game
    - the generator produces samples and try to fool discriminator
    - the discriminator distinguishes real and generated data
- **Networks:** convolutional neural network, ConvLSTM, fully connected network



Precipitation nowcasting: dataset

- Data source: Météo-France[1]
- Accumulated precipitation (ACRR): the 5-minute rainfall in 1/100 mm
- Resolution: 5 min, 1 km



A precipitation map at 11:00 AM on June 11, 2018

Dataset	Period	Samples
Training	2018/01/01 - 2019/10/17	14014
Validation	2019/10/17 - 2019/12/24	3640

[1] thanks to Thibaut Montmerle and Yu Nan for suggestions and guidance

Precipitation nowcasting: UM-GAN model process

Step1: **Use universal multifractals** to obtain three parameters in space and time.

Step2: **Generate space-time data** with 18 timesteps by continuous-in-scale multiplicative cascades, then to pick the 2D noise data at the lead times at 5min, 30min and 60min.

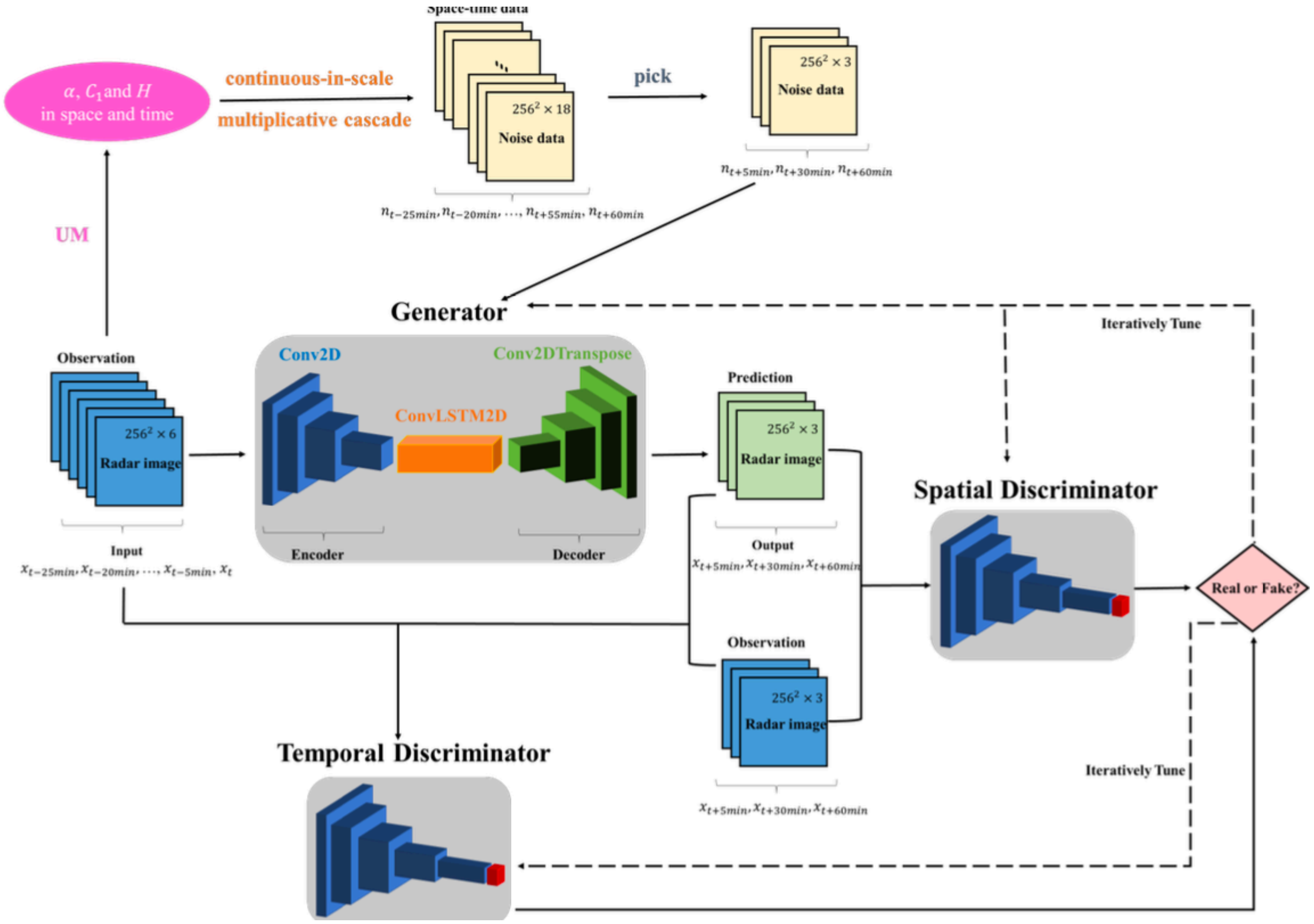
Step3: **Define UM-GAN architecture** based on the application of space-time nowcasting.

Step4: **Train the generator** to create predictions that try to fool two discriminators.

Step5: **Train the spatial discriminator and temporal discriminator** to distinguish real data from generated data in space and time.

Step 6: **Continue training** two discriminators and a generator alternately for multiple epochs.

Step7: **Save the best generator model** for creating predictions in the testing set.



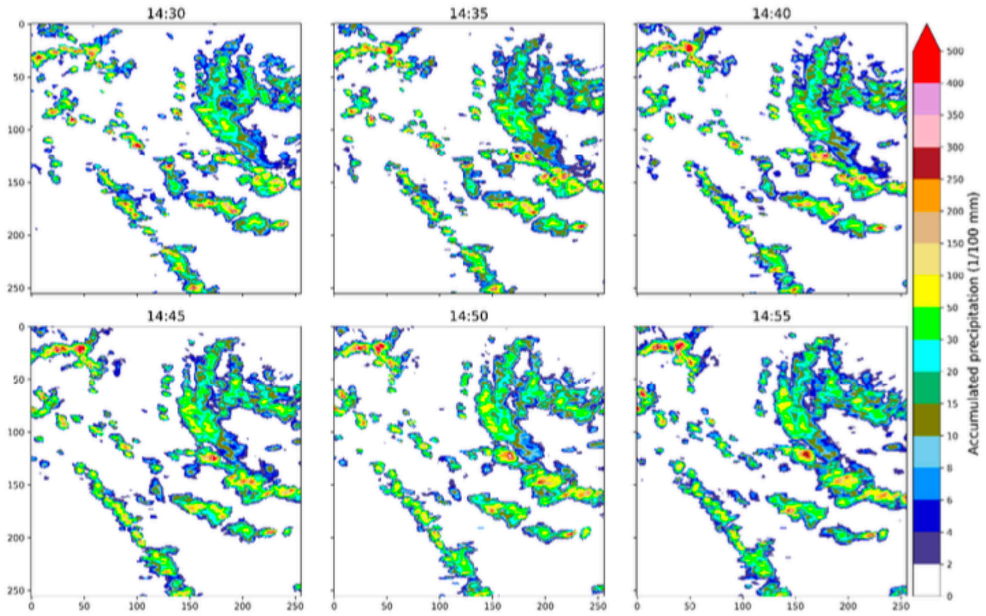
Precipitation nowcasting: results analysis– Event 12/06/2020

Description: the thunderstorm with moderate rain due to a jet stream

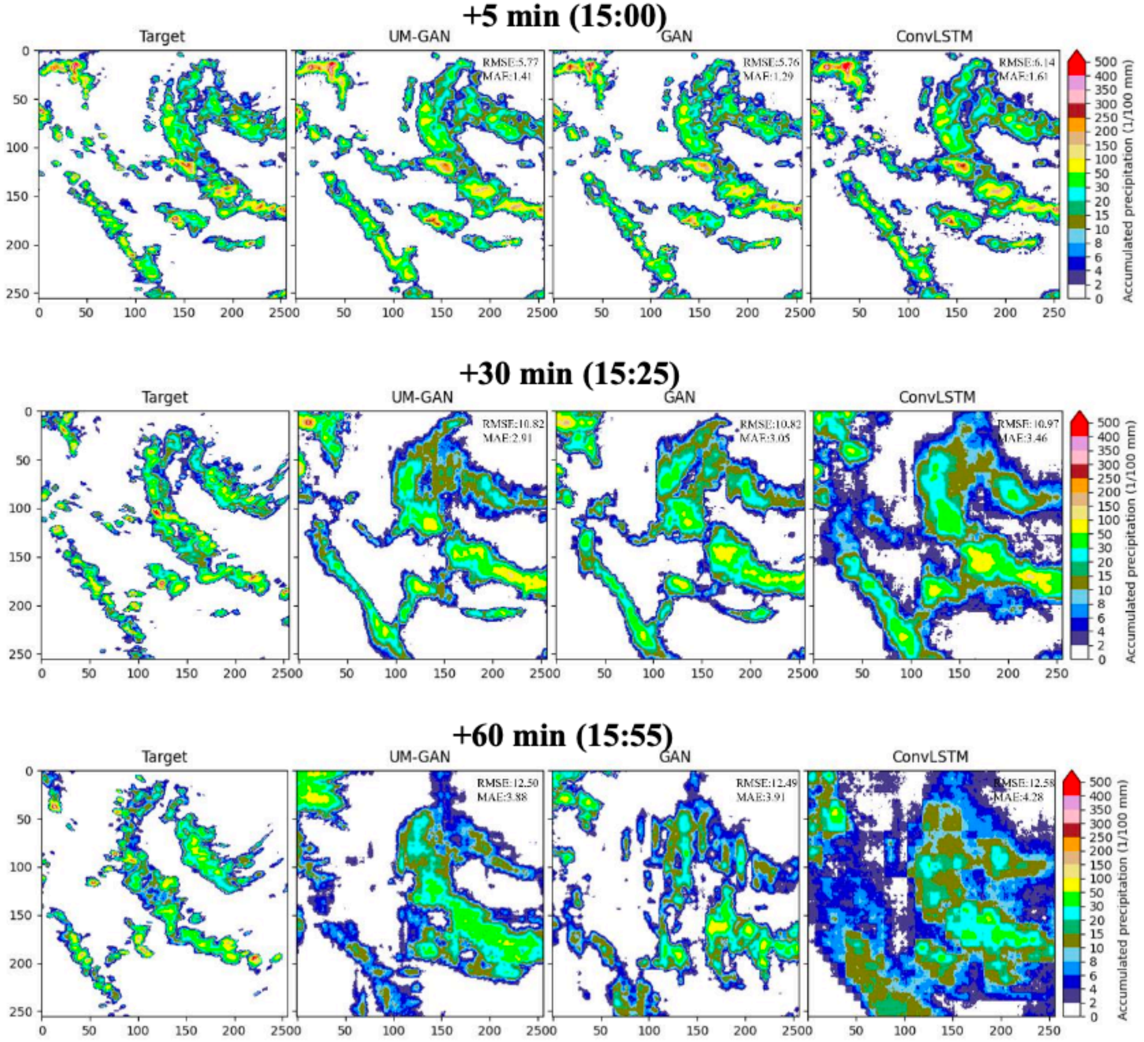
Input : previous 6 steps of historical data (14:30, 14:35, 14:40, 14:45, 14:50, 14:55); additional noise data generated by continuous-in-scale multiplicative cascades

Output: (15:00, 15:25, 15:55)

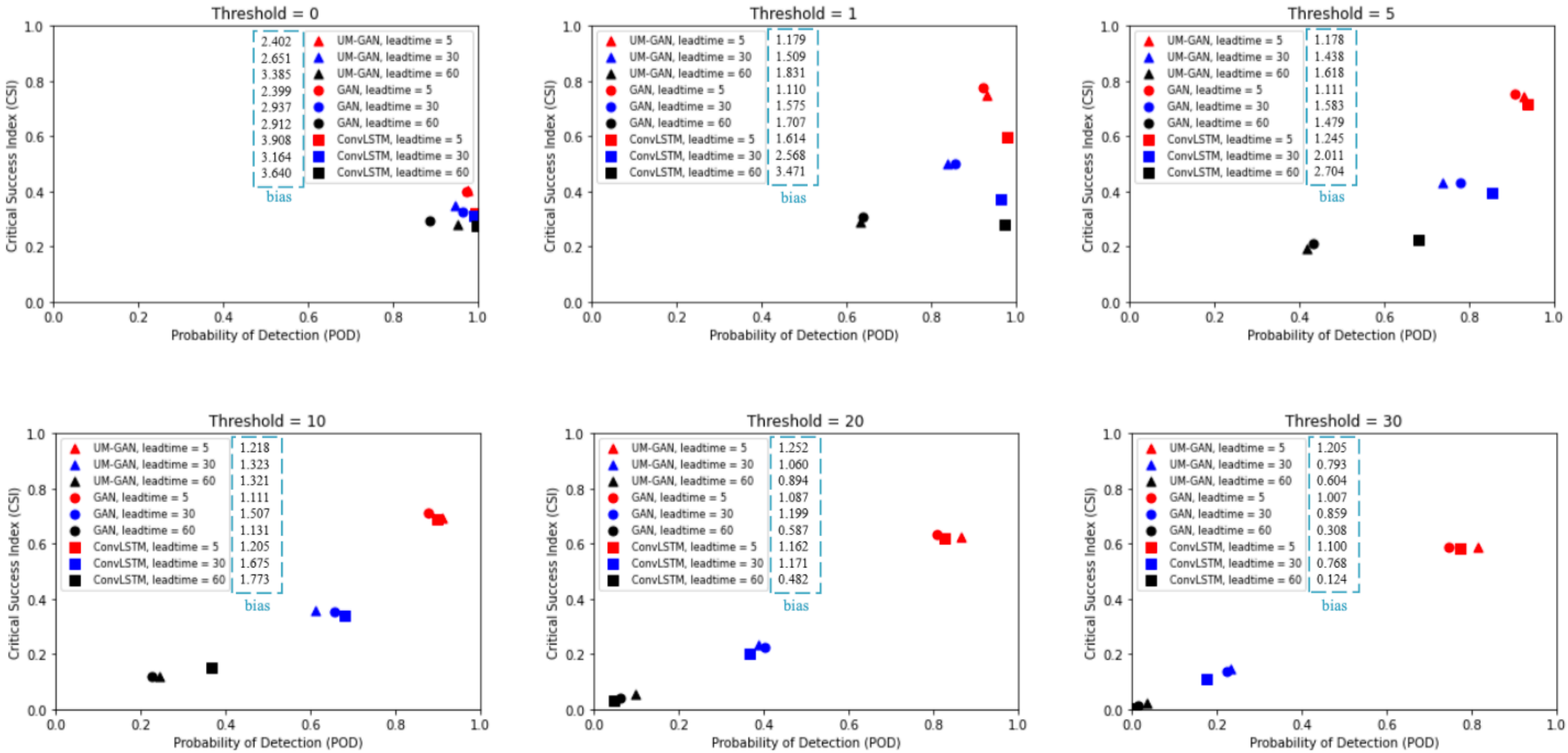
The historical data from six previous time steps



Precipitation nowcasts by three different models and targets at the lead times of 5, 30 and 60 minutes

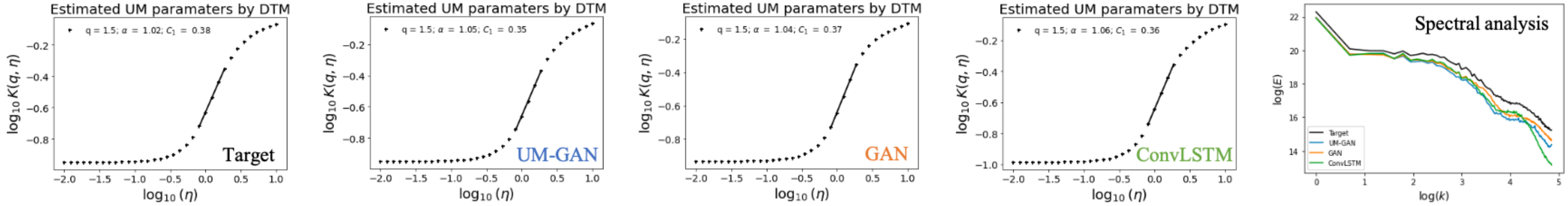


Precipitation nowcasting: Categorical scores – Event 12/06/2020

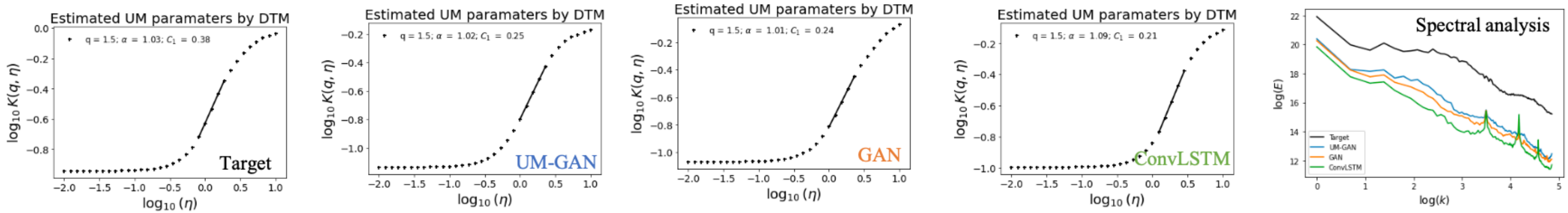


Precipitation nowcasting: UM results– Event 12/06/2020

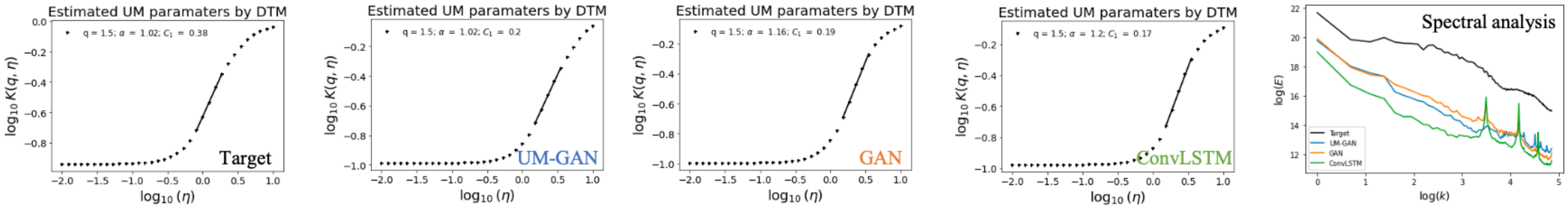
at 5-minute lead time



at 30-minute lead time



at 60-minute lead time

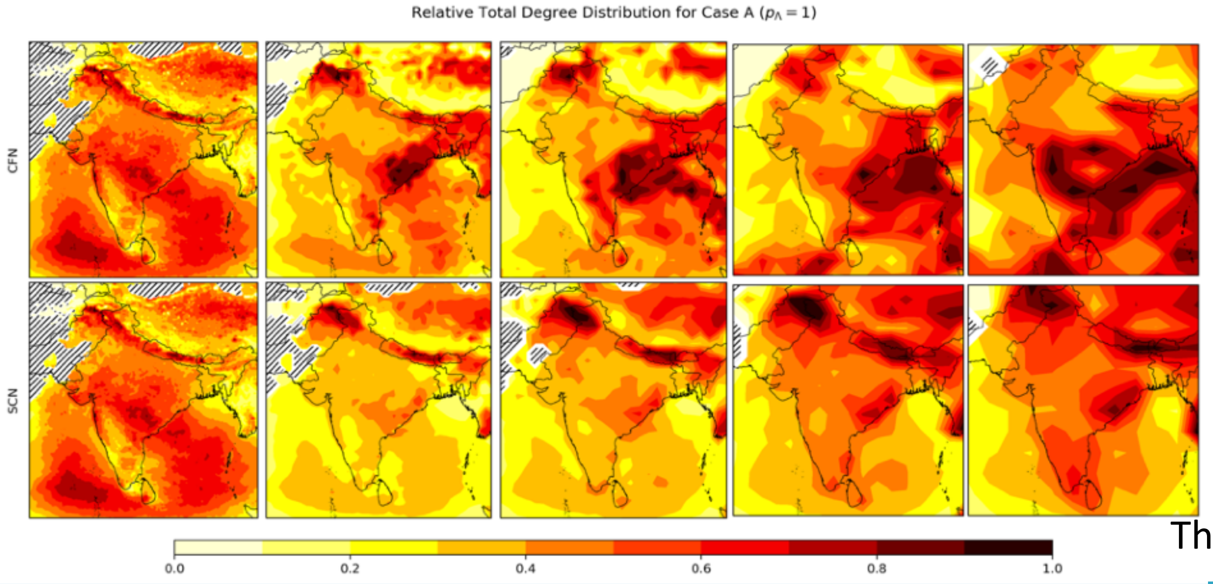


Conclusions and prospects

- **Conclusions**
  - VMD is effective for time series prediction (LSTM, GRU and bidirectional variants), but still underestimates extremes
  - GAN-based models have better MAE and RMSE scores, and higher prediction accuracy in comparison to the ConvLSTM without adversarial training or linear regression
  - stronger performance of UM-GAN in POD and CSI scores, and bias,particularly for thresholds of 10, 20, 30 (1/100 mm /5’)
  - UM parameters from ConvLSTM have a larger dispersion

Conclusions and prospects

- **Prospects**
  - increase accuracy of mean intermittency (*C_1*) for longer lead times
  - nowcasting of other geophysical fields
  - multifractal prediction vs. RNN prediction (beyond GAN)
  - neural networks and complex/climate networks
  - ensemble predictions, or probalistic versions



Thomas et al., EGU Letters. 2024