Extreme Precipitation Nowcasting Using Transformer-based Generative Models

Cristian Meo*, Ankush Roy*, Mircea Lica*,
Junzhe Yin, Zeineb Cher, Yanbo Wang, Ruben Imhoff,
Remko Uijlenhoet, Justin Dauwels

Motivation: Extreme precipitations

- 2 billion people affected by **floods** from 1998 to 2017 [1]
- Estimated **economic losses** of 656 billion USD [1]

[1] Imhoff, R. O., et al. "Spatial and temporal evaluation of radar rainfall nowcasting techniques on 1,533 events., 2020

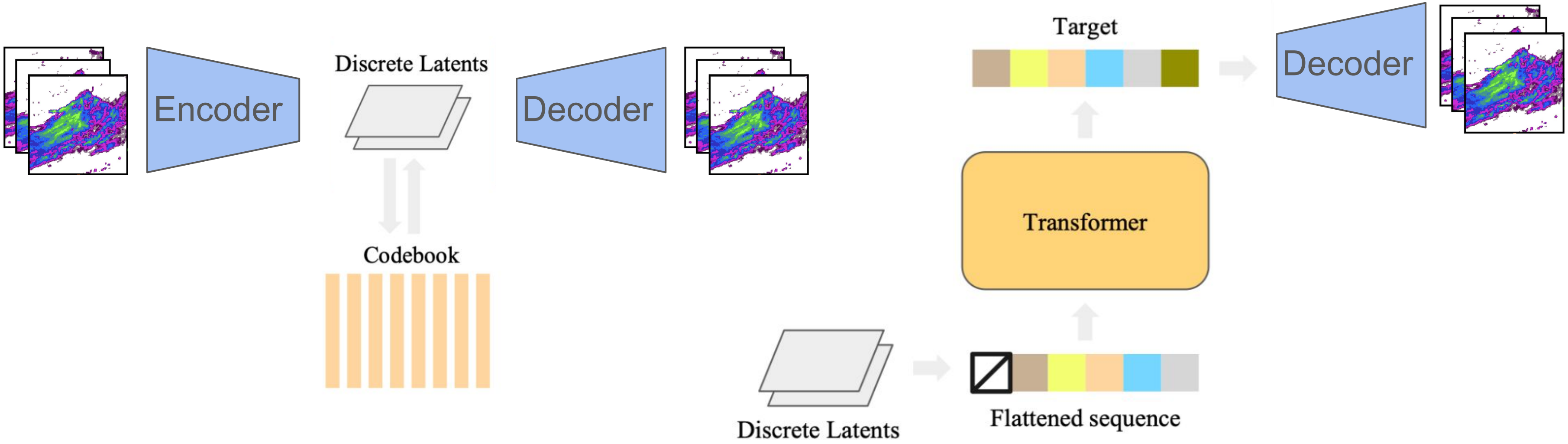
Precipitation Nowcasting

- Predict rainfall levels within 6-hours window
- Enables timely response to extreme meteorological events
- Allows for proactive disaster response and climate resilience strategies

Ravuri, Suman, et al. "Skilful precipitation nowcasting using deep generative models of radar.", 2021

Nowcasting as Video Prediction

Given a spatio-temporal sequence of N frames $\boldsymbol{x}_{\text{in}} \in \mathbb{R}^{N \times H \times W \times C}$, where *H*, *W* denote the spatial resolution and C represents the image channels or the different type of measurements (e.g., radar maps, heat maps, etc), the goal is to predict the next M frames $\boldsymbol{x}_{\text{out}} \in \mathbb{R}^{M \times H \times W \times C}$



Yan, Wilson, et al. "Videogpt: Video generation using vq-vae and transformers.", 2021

Deltares

Fallbacks of current methods

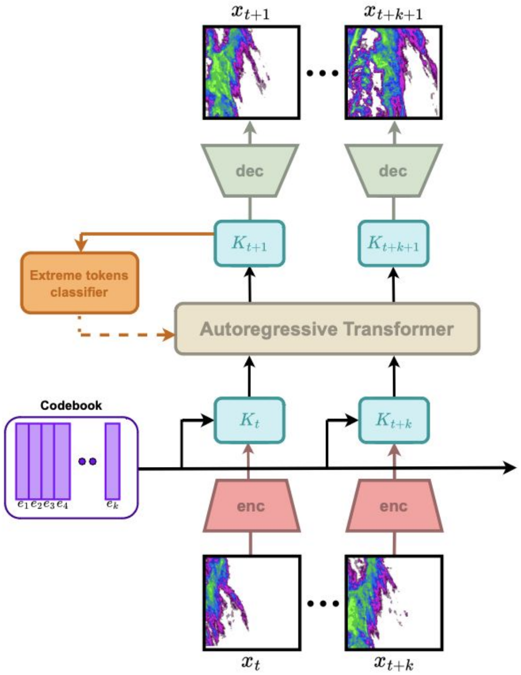
- Absent/Wrong **inductive bias** for extreme events [2, 2a]
- **Long generation times** which make nowcasting unfeasible [2]
- Lack of **temporal consistency** [3]

[2] Bi, Haoran, et al. "Nowcasting of Extreme Precipitation Using Deep Generative Models.", 2023
[2a] Ravuri, Suman, et al. "Skilful precipitation nowcasting using deep generative models of radar.", 2021
[3] Yan, Wilson, et al. "Temporally consistent transformers for video generation.", 2022

TU Delft

Our Approach: NowcastingGPT-EVL

- **Video prediction backbone** similar to the VideoGPT [4] framework
- **VQ-VAE** [5] as a feature extraction method
- Sequence modelling using **Autoregressive Transformer** [6]
- **Extreme Value Loss Regularization** (EVL) [7]

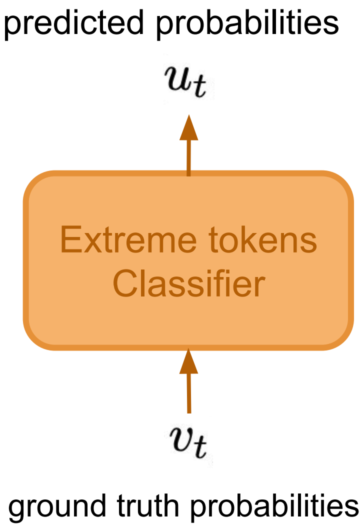


[4] Yan, Wilson, et al. "Videogpt: Video generation using vq-vae and transformers.", 2021
[5] Van Den Oord, Aaron, and Oriol Vinyals. "Neural discrete representation learning.", 2017
[6] Vaswani, Ashish, et al. "Attention is all you need.", 2017
[7] Ding, Daizong, et al. "Modeling extreme events in time series prediction.", 2019

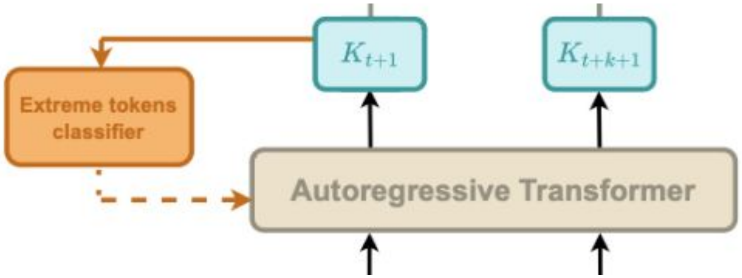
Deltares

Extreme Value Loss (EVL) + Autoregressive Transformer

- We employ a classifier that dynamically predicts whether tokens represents extreme or non extreme events.



$$\mathcal{L}_{\text{Transformer(NowcastingGPT-EVL)}} = \mathbb{E}_{x\sim p(x)}[-\log \prod_{i=1}^{N} p(\mathbf{s}_i|\mathbf{s}_{<i})] + \lambda[\text{EVL}(u_t, v_t)]$$



TU Delft

Extreme Value Loss (EVL) Regularization

- Extreme precipitation nowcasting deals with a severe imbalance in training data, which makes standard cross-entropy perform suboptimal.

- EVL is designed to solve the data imbalance between extreme and non-extreme cases in time series data.

$$\mathrm{EVL}(u_t, v_t) = -\beta_1 \left[1 - \frac{u_t}{\gamma}\right]^{\gamma} v_t \log(u_t) - \beta_0 \left[1 - \frac{1 - u_t}{\gamma}\right]^{\gamma} (1 - v_t) \log(1 - u_t),$$

Deltares

Extreme Value Loss (EVL) Regularization

- Extreme precipitation nowcasting deals with a severe imbalance in training data, which makes standard cross-entropy perform suboptimal.
- EVL is designed to solve the data imbalance between extreme and non-extreme cases in time series data.

ground truth probabilities

$$\mathrm{EVL}(u_t, v_t) = -\beta_1 \left[1 - \frac{u_t}{\gamma}\right]^{\gamma} v_t \log(u_t) - \beta_0 \left[1 - \frac{1 - u_t}{\gamma}\right]^{\gamma} (1 - v_t) \log(1 - u_t),$$

predicted probabilities

Deltares

Extreme Value Loss (EVL) Regularization

- Extreme precipitation nowcasting deals with a severe imbalance in training data, which makes standard cross-entropy perform suboptimal.

- EVL is designed to solve the data imbalance between extreme and non-extreme cases in time series data.

ground truth probabilities

$$\mathrm{EVL}(u_t, v_t) = -\beta_1 \left[1 - \frac{u_t}{\gamma}\right]^{\gamma} v_t \log(u_t) - \beta_0 \left[1 - \frac{1 - u_t}{\gamma}\right]^{\gamma} (1 - v_t) \log(1 - u_t),$$

predicted probabilities

Extreme

Not Extreme

TU Delft

Experimental Setup

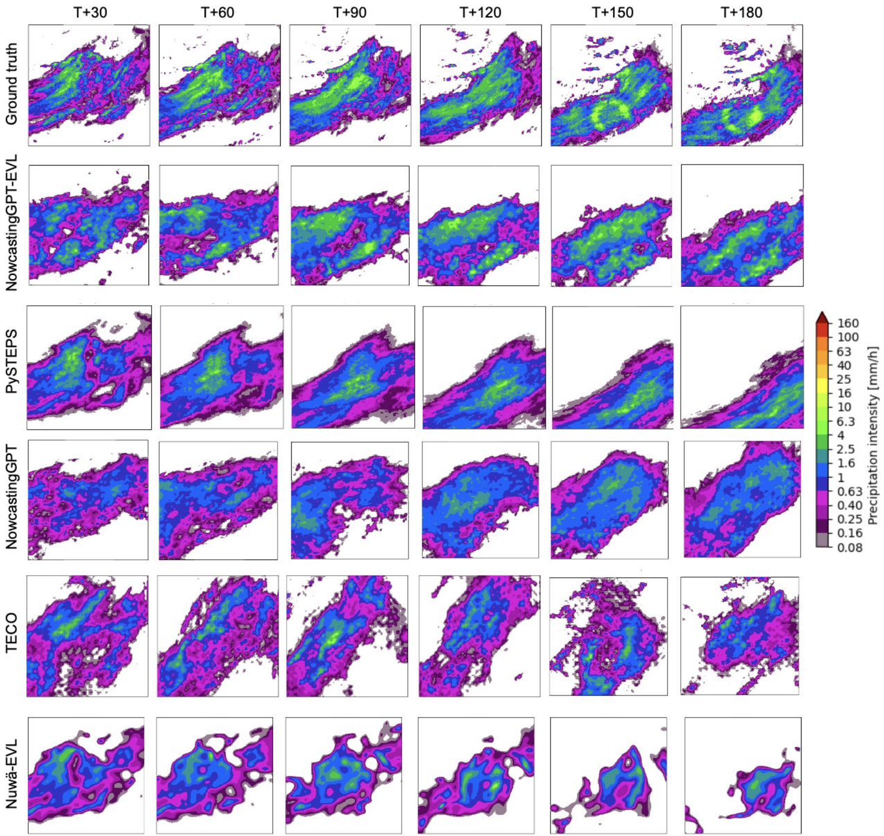
- Royal Netherlands Meteorological Institute (KNMI) dataset
- Predictions up to three hours into the future.
- Radar maps (256x256) between 2008 and 2021
- Catchment-level analysis for real world scenarios
- Baselines

Sluiter, R. *Interpolation Methods for the Climate Atlas. KNMI, 2009*

	[2]		[7]	[3]	
	Nuwä-EVL	NowcastingGPT	PySTEPS	TECO	NowcastingGPT-EVL
Number of parameters	772, 832 M	402, 735 M	-	165, 960 M	520, 374 M
Training time	672h	240h	-	155h	264h
Generation time	322.86s	38.90s	9.34s	0.51s	43.10s

[7] Pulkkinen, Seppo, et al. "Pysteps: An open-source Python library for probabilistic precipitation nowcasting (v1. 0).", 2019

Qualitative Results



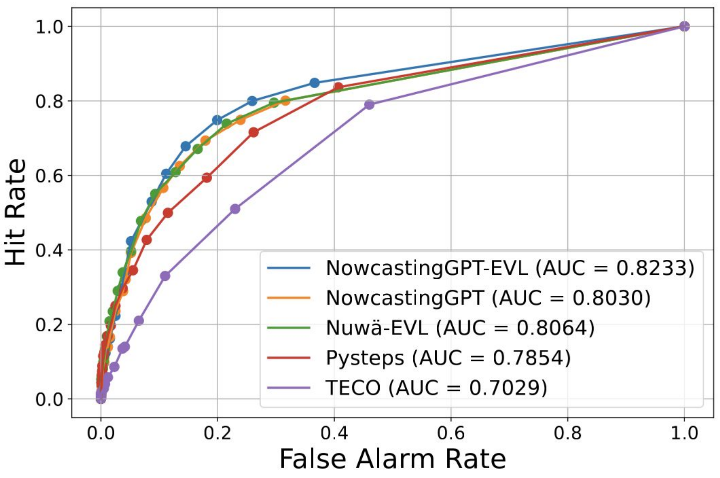
Deltares

Quantitative Results

	Nuwä-EVL	NowcastingGPT	PySTEPS	TECO	NowcastingGPT-EVL
PCC (↑)	0.15	0.20 ± 0.002	0.14	0.10 ± 0.002	**0.22** ± 0.002
MSE (↓)	4.85	3.60 ± 0.02	6.22	3.65 ± 0.008	**3.45** ± 0.02
MAE (↓)	1.00	0.72 ± 0.005	0.93	**0.68** ± 0.001	0.69 ± 0.005
CSI(1mm) (↑)	**0.23**	0.21 ± 0.002	0.21	0.07 ± 0.001	0.22 ± 0.002
CSI(2mm) (↑)	**0.13**	0.11 ± 0.001	0.12	0.03 ± 0.001	0.12 ± 0.001
CSI(8mm) (↑)	0.008	0.005 ± 0.0005	**0.01**	0.001 ± 0.0009	0.009 ± 0.0005
FAR(1mm) (↓)	0.61	0.59 ± 0.002	**0.55**	0.69 ± 0.002	0.59 ± 0.002
FAR(2mm) (↓)	0.76	0.71 ± 0.0007	**0.70**	0.78 ± 0.004	0.71 ± 0.0007
FAR(8mm) (↓)	0.85	0.59 ± 0.003	0.89	**0.49** ± 0.006	0.52 ± 0.003
FSS(1km) (↑)	0.35	0.49 ± 0.003	0.32	0.49 ± 0.003	**0.52** ± 0.003
FSS(10km) (↑)	0.42	0.55 ± 0.004	0.41	0.46 ± 0.003	**0.58** ± 0.004
FSS(20km) (↑)	0.48	0.59 ± 0.004	0.47	0.42 ± 0.003	**0.62** ± 0.004
FSS(30km) (↑)	0.52	0.62 ± 0.004	0.51	0.37 ± 0.002	**0.65** ± 0.004

TU Delft

Quantitative Results: ROC Curve



NowcastingGPT-EVL outperforms all other baselines, presenting the highest area under the curve.

Discussion

- The proposed model contains roughly 400M parameters.
- Training time is around two weeks.
- Extreme/Non-extreme tokens ground truth values are defined using a predefined threshold.
- The experimental setup contains only one dataset.

Conclusions

- Overall, the proposed NowcastingGPT-EVL, a video prediction model that incorporates a EVL loss regularizer, outperforms all related baselines in terms of nowcasting downstream task and visual fidelity metrics.

- According to the results, the proposed EVL regularization approach enhances the predictive capabilities of the underlying nowcasting model, allowing to represent and predict extreme events.

Deltares

Thank you for your attention!

For any question contact us at c.meo@tudelft.nl
Collaborations proposals are more than welcomed!

SCAN ME

Paper

SCAN ME

Code

TU Delft

Appendix

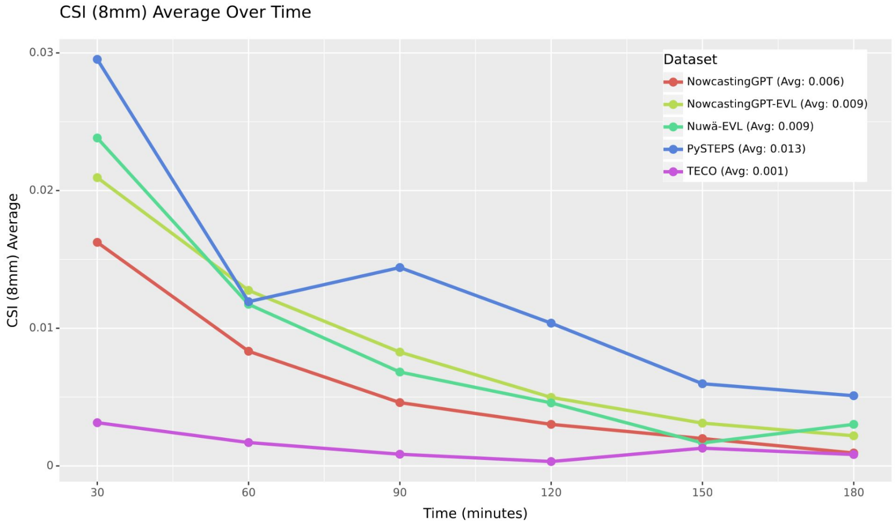
Fallbacks of current methods

- Absent/Wrong inductive bias for extreme events
- **Long generation times which make nowcasting unfeasible**
- Lack of temporal consistency

	Nuwä-EVL
Number of parameters	772, 832 M
Training time	672h
Generation time	322.86s

Fallbacks of current methods

- Wrong inductive bias for extreme events
- Long generation times which make nowcasting unfeasible
- **Lack of temporal consistency**



Temporal Consistency and Sampling Time

- Transformer-based nowcasting methods suffer from extremely slow generation time which makes nowcasting obsolete.

- Training these models require long training times in the order of weeks

Temporal Consistency and Sampling Time

- Transformer-based nowcasting methods suffer from extremely slow generation time which makes nowcasting obsolete.

- Training these models require long training times in the order of weeks

- Moreover, the quality of the generated samples degrades quickly lacking proper temporal consistency

Temporal Consistency and Sampling Time

In order to encourage faster and more efficient models, we benchmark TECO, an efficient transformer-based model that provides temporal consistency, while having a blazing fast sampling time and an order of magnitude less parameters than other models.

	Nuwä-EVL	NowcastingGPT	PySTEPS	TECO	NowcastingGPT-EVL
Number of parameters	772, 832 M	402, 735 M	-	165, 960 M	520, 374 M
Training time	672h	240h	-	155h	264h
Generation time	322.86s	38.90s	9.34s	0.51s	43.10s