Using Artificial Intelligence and Machine Learning to Screen Dams For Seismic Risk

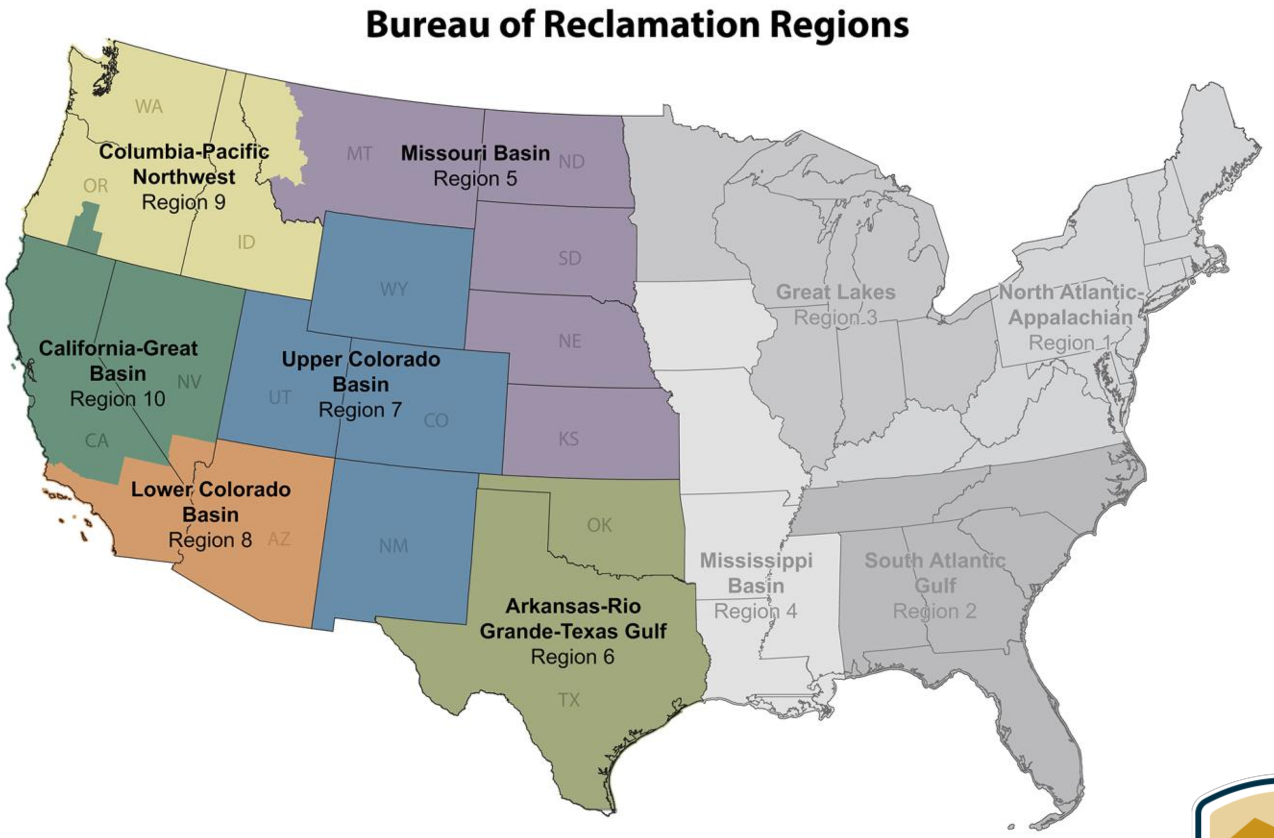
Cameron Carpenter
U.S. Bureau of Reclamation

Taiwan International Water Week

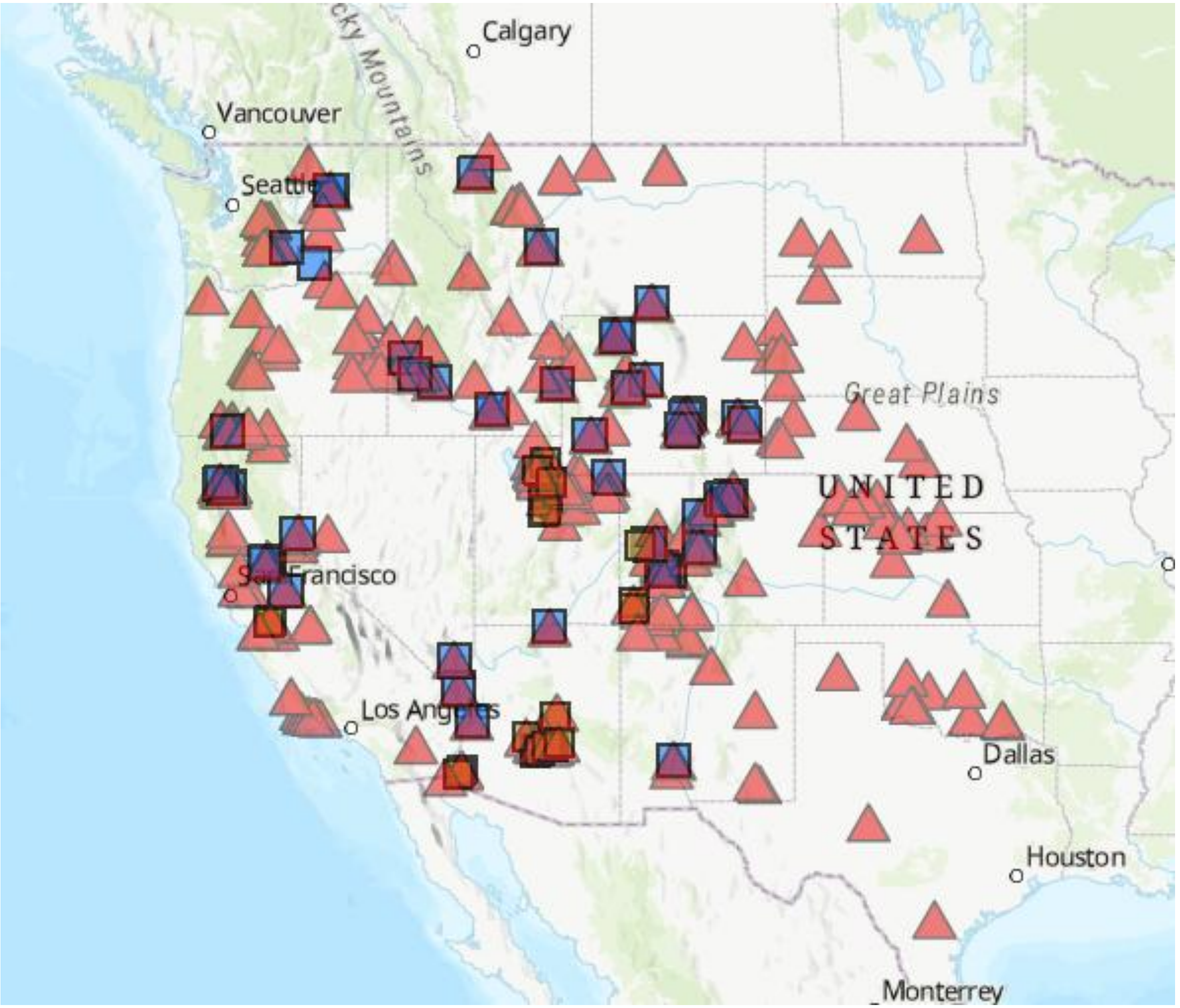
October 29, 2025

United States Bureau of Reclamation

- Largest wholesaler of water in the U.S.
- Second-largest producer of hydroelectric energy in the U.S.



Reclamation's Dam Inventory



Dams

High Hazard Potential

Hydropower Facilities

Reserved

Transferred

FY 2025 INVENTORY

Total Facilities – **242**

High Hazard Dams – **364**

Low Hazard Dams – **22**

FY 2025 ACTIVITIES

Comprehensive Reviews – **33**

Ongoing Issue Evaluations – **52**

Corrective Action Studies – **7** (includes 4 in Final Design)

Construction Projects – **5**

PROGRAM IMPACTS

$3.2-5.2 billion
Projected value of planned dam safety improvements

17
Dams being improved by the Dam Safety Program

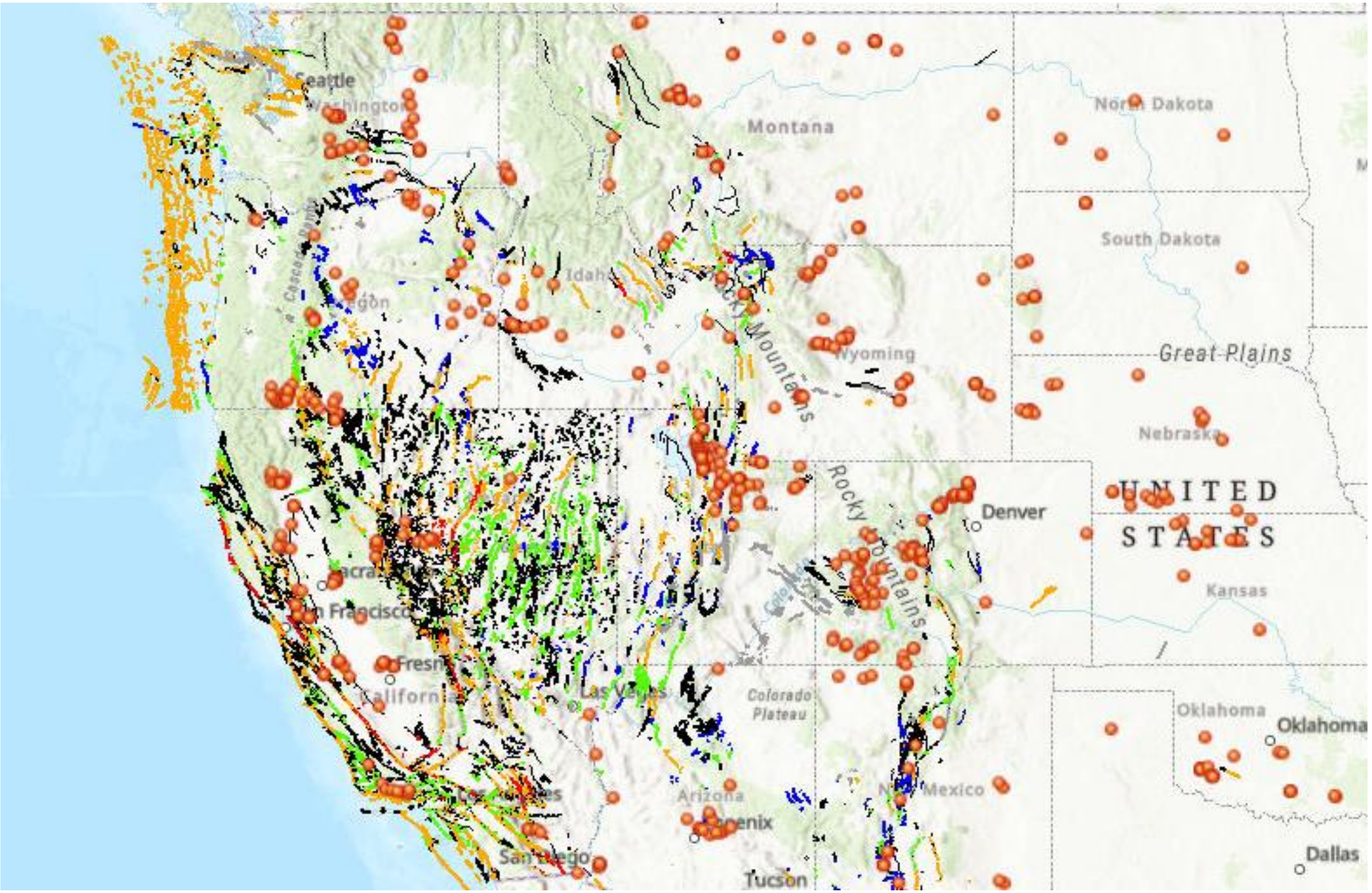
9
States benefiting from improved dams

400 thousand
People safer because of dam improvements

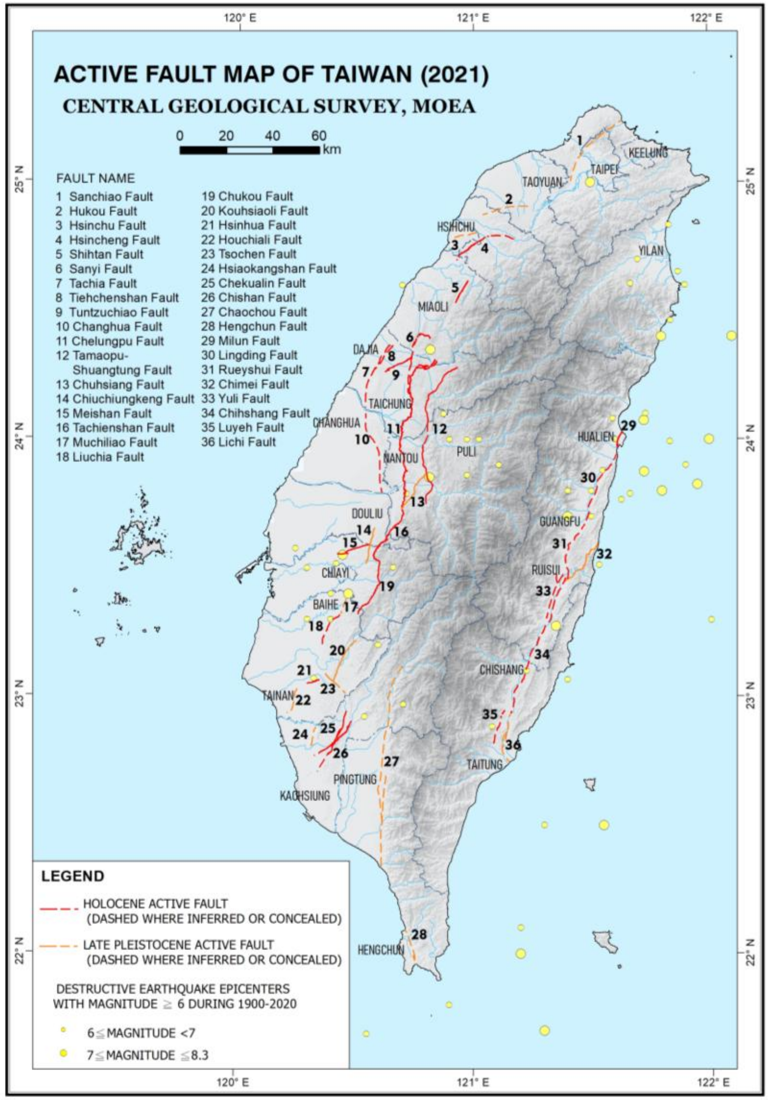
5 million
Acres of continued irrigation supply

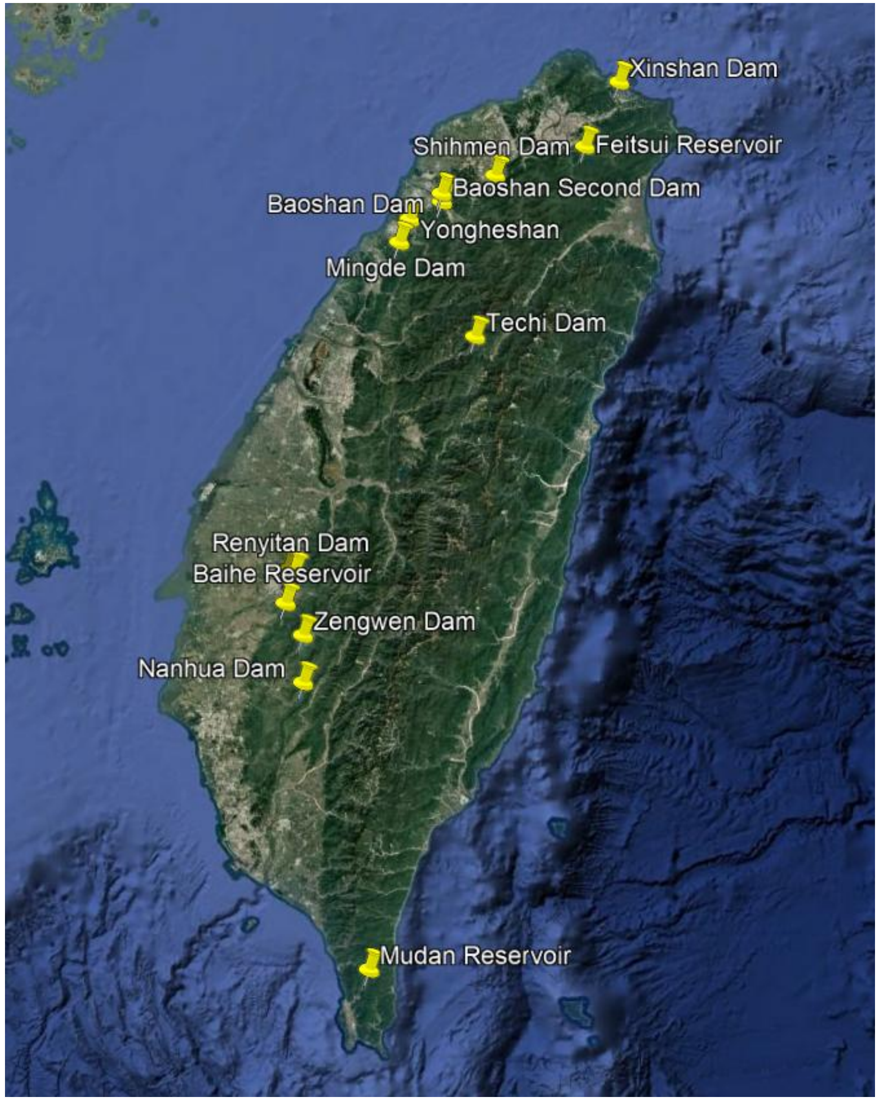
20+ million
People benefiting from continued water deliveries

Seismic Risk Across Reclamation's Portfolio



Seismic Risk In Taiwan





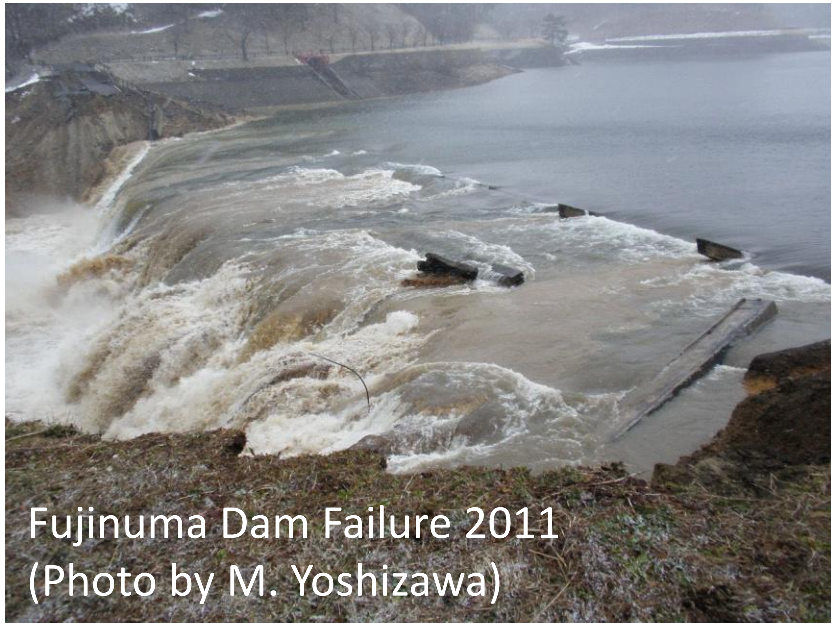
Source: Lin, Chii-Wen & Liu, Yen-Chiu & Chou, Ping-Shan & Lin, Yen-Hui. (2022). ACTIVE FAULT MAP OF TAIWAN, 2021.

Seismic Risk Factors at Reclamation Facilities

- Some Reclamation dams were constructed using hydraulic fill, but most of these have been fixed.
- High seismic risk areas:
  - Cascadia Subduction Zone
  - San Andreas Fault
- Seismic loading estimates are increasing with LiDAR and detailed satellite imagery finding new faults.

Seismic Failure of Dams

- Definition: Collapse or breach of a dam due to earthquake-induced forces.
- Consequences: Uncontrolled water release usually with little to no warning.
- Common Mechanisms:
  - Ground shaking–induced cracking
  - Foundation instability
  - Liquefaction of embankment materials



Fujinuma Dam Failure 2011
(Photo by M. Yoshizawa)

San Fernando Dam
(Source: CA DWR)

Understanding Seismic Risk

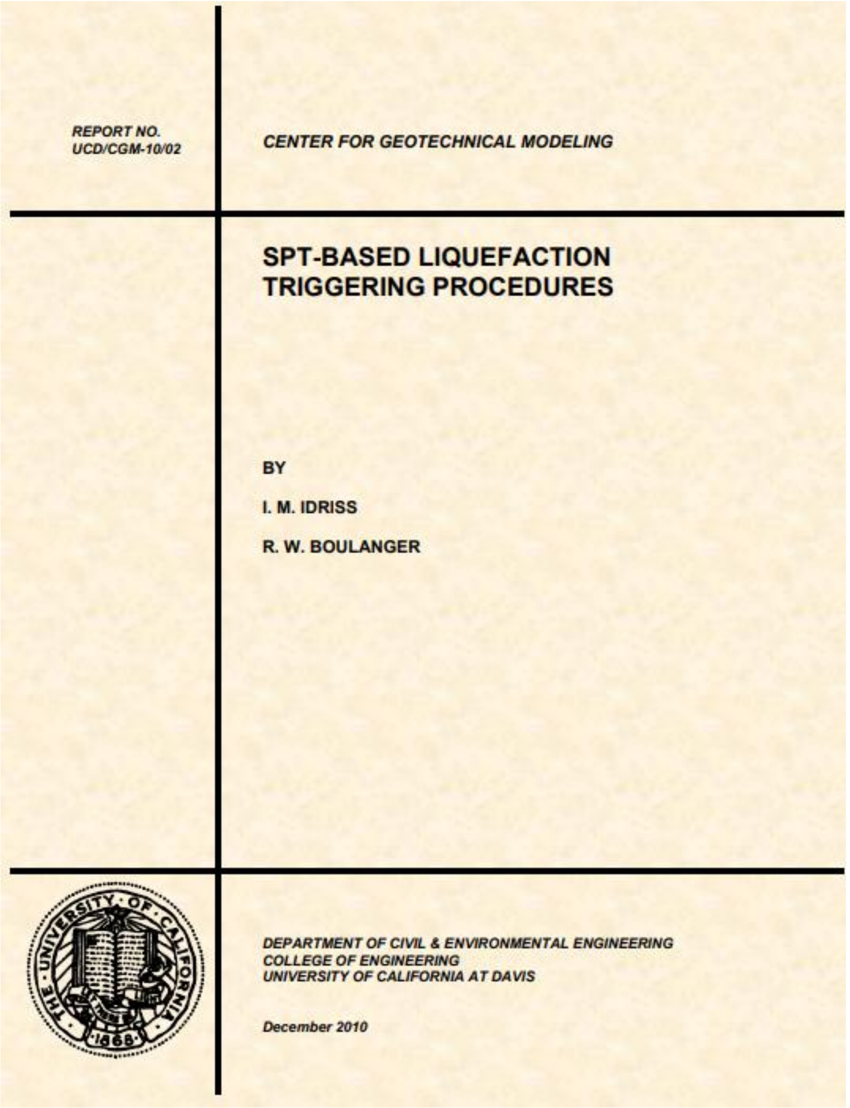
- We have many tools to help us understand our seismic risk:
  - Geotechnical Investigations
  - Seismic Hazard Analysis
  - Numerical Modeling Software
    - FLAC2D
  - Numerical tools
    - Idriss and Boulanger (SPT/CPT-based methods)
    - Seed and Idriss simplified procedure

AI/ML Tools to Understand Risk – Case Study

- Case Study 1: Re-Evaluating Idriss and Boulanger (2010)
  - Can we recreate this using machine learning?
- Case Study 2: Screen Dam Deformation Using Machine Learning
  - Can we screen existing dams or new designs for deformation risk using predetermined design events?

Case Study 1: Re-Evaluating Idriss and Boulanger (2010)

This paper presents an updated framework for evaluating the potential for soil liquefaction using the Standard Penetration Test (SPT). It refines the stress-based approach by incorporating an expanded case history database, re-examining key parameters (like overburden corrections and magnitude scaling), and introducing a probabilistic model for liquefaction triggering.


REPORT NO.
UCD/CGM-10/02

CENTER FOR GEOTECHNICAL MODELING

SPT-BASED LIQUEFACTION
TRIGGERING PROCEDURES

BY

I. M. IDRISS

R. W. BOULANGER

DEPARTMENT OF CIVIL & ENVIRONMENTAL ENGINEERING
COLLEGE OF ENGINEERING
UNIVERSITY OF CALIFORNIA AT DAVIS

December 2010

Idriss and Boulanger (2010)

<u>Corrected blow count:</u>

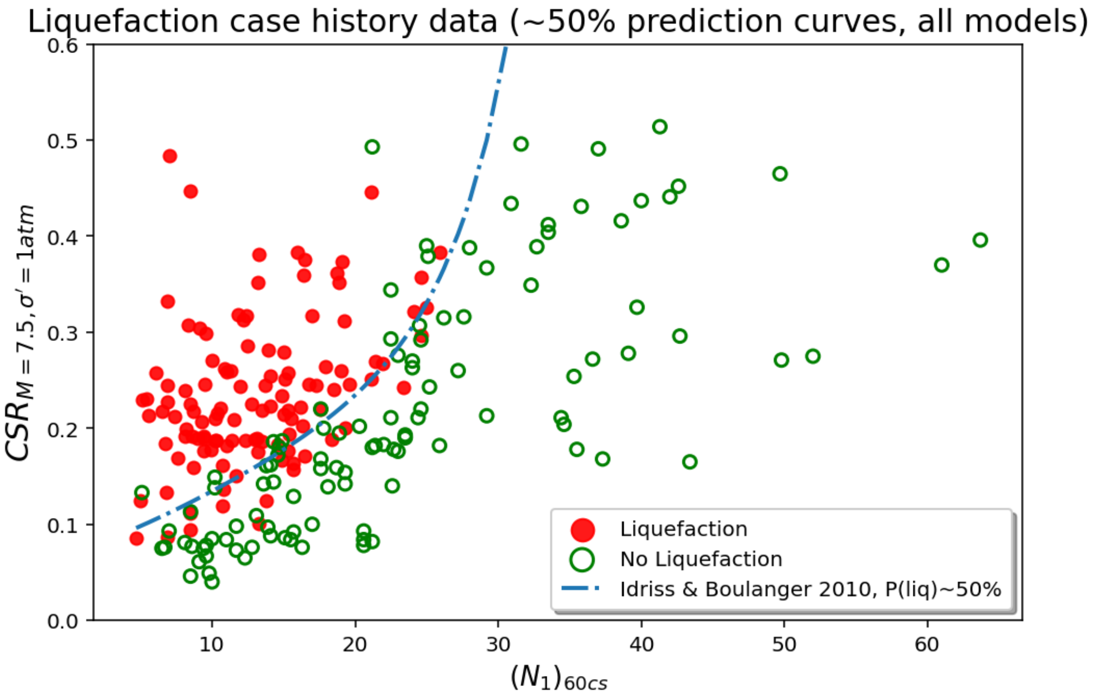
- A blow count is a physical test performed in the field on the soil that results in a measure of how much resistance the soil can withstand before failing.

- The corrected blow count means the test result was "corrected" to normalize the value for various factors such as depth of test, type of testing device used, etc.

<u>Adjusted cyclic stress ratio (CSR):</u>

- A cyclic stress ratio value is a measure of how much force will be applied to the soil at different depths during a seismic event.

- Adjusted CSR means the value was "adjusted" for factors such as magnitude of earthquake, etc.



Re-evaluating Idriss and Boulanger - Inputs

Two approaches were taken:

1.) Regularization of all potential inputs from Idress and Boulanger.

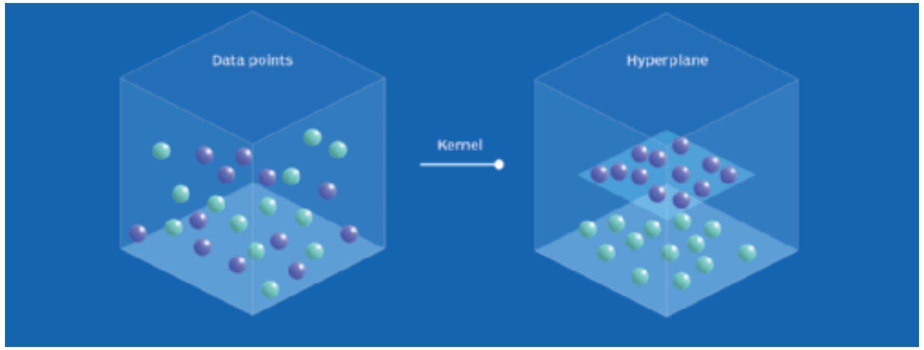
- All 21 features of a site were used as inputs into the ML models. But regularization forces the coefficients of the inputs that "matter less" to the problem to become 0, which means those inputs were essentially ignored in the ML model creation.

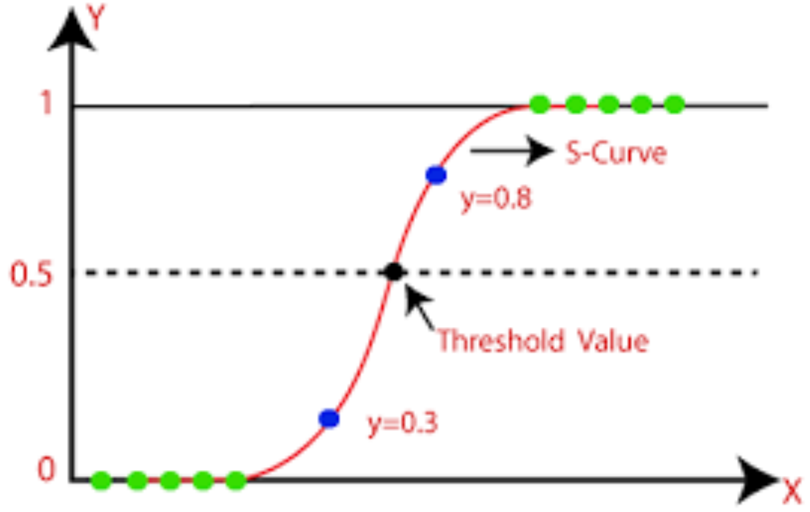
2.) Crowdsourced an answer from the Geotech industry.

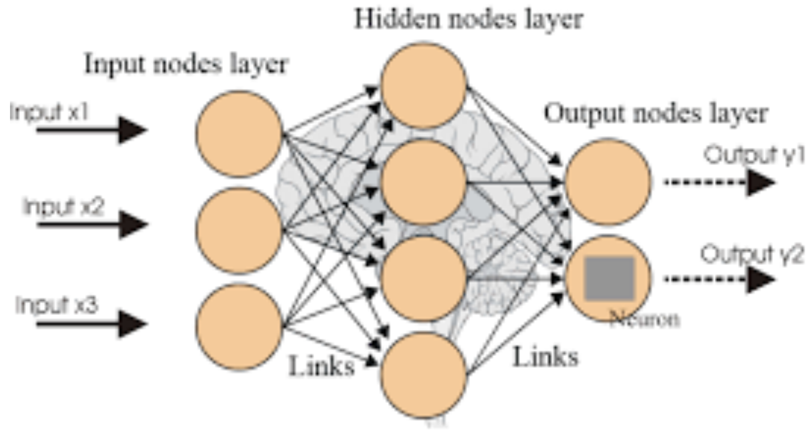
- Use only the equivalent CSR and corrected blow count. These are the 2 features accepted as sufficient to estimate liquefaction by the geotechnical industry.

Re-evaluating Idriss and Boulanger – Model Selection

- We used three types of models to re-evaluate the data:
  - Support Vector Machine (SVM)
  - Logistic regression (LR)
  - Neural Network (NN)



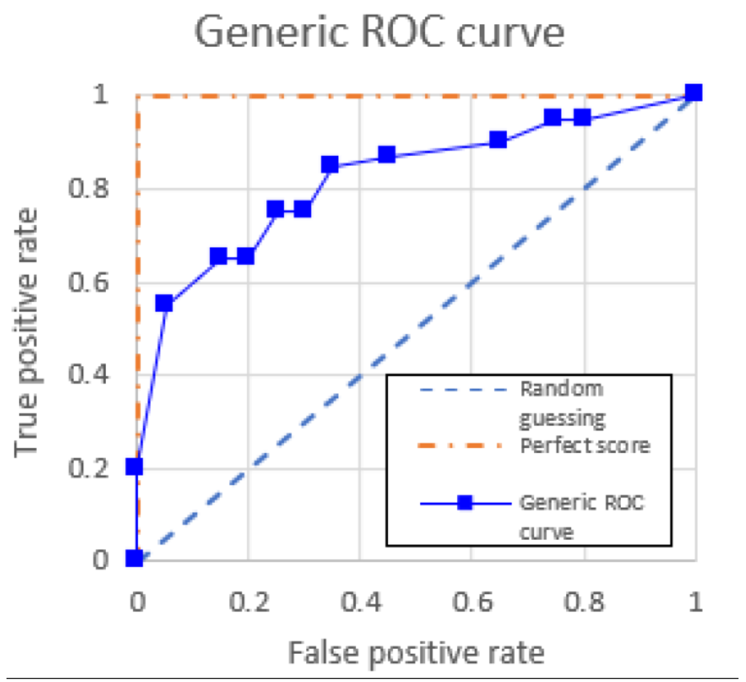




Re-evaluating Idriss and Boulanger – Model Assessment

- Model Assessment Metrics
  - Error: (FP+FN)/(total) - lower better
  - Recall: TP/(TP+FN) – high better
  - Receiver Operating Characteristic (ROC) Curve: TP vs FP
    - Area under the curve (AUC) – higher better

		Predicted class	
		Positive	Negative
Actual class	Positive	**True positives (TP)**	**False negatives (FN)**
	Negative	**False positives (FP)**	**True negatives (TN)**



Using ML to Re-evaluate Idriss and Boulanger – Results

- Model Assessment:
    1. I&B 2010 model is underperforming most models in all three metrics by a decent amount.
    2. “2 input” models outperforming “all 21” input models in all three metrics.

Error (lower is better):
LR 2 inputs = 8.7%
NN 2 inputs = 8.7%

NN all inputs = 10.9%
I&B 2 inputs = 13.2%
LR all inputs = 17.4%

Recall (higher is better):
LR 2 inputs = 95.7%
NN 2 inputs = 91.3%

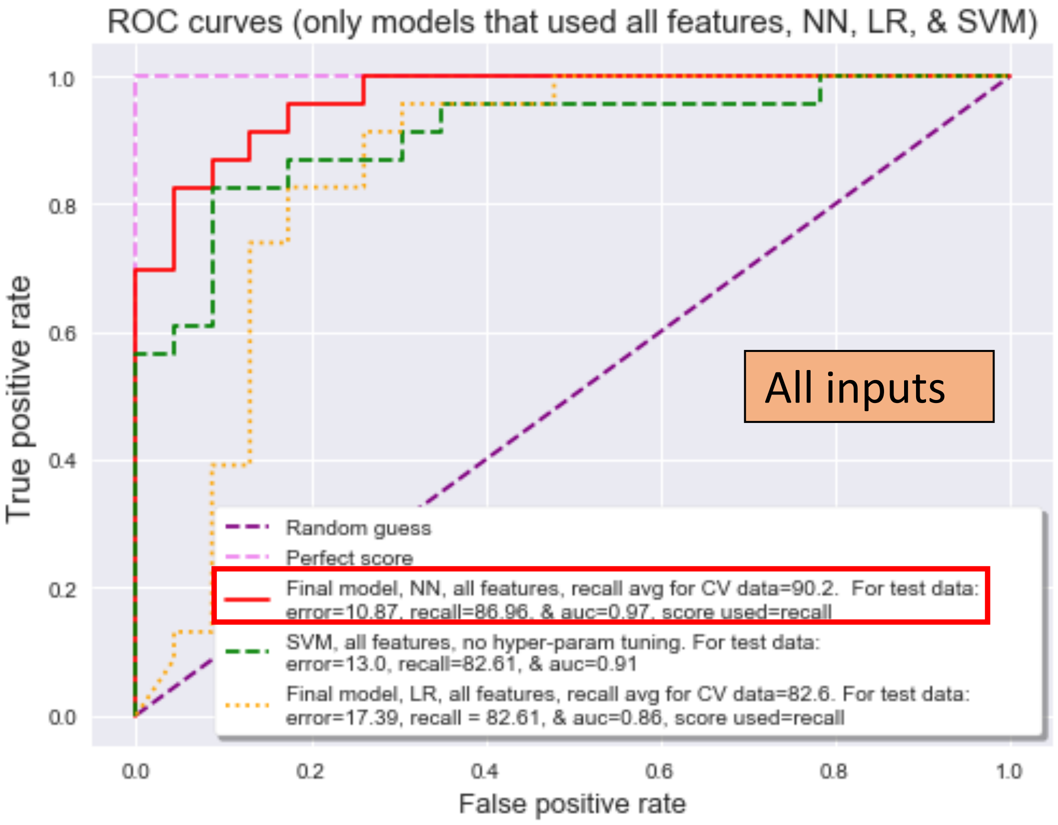
NN all inputs = 87%
I&B 2 inputs = 84.4%
LR all inputs = 82.6%

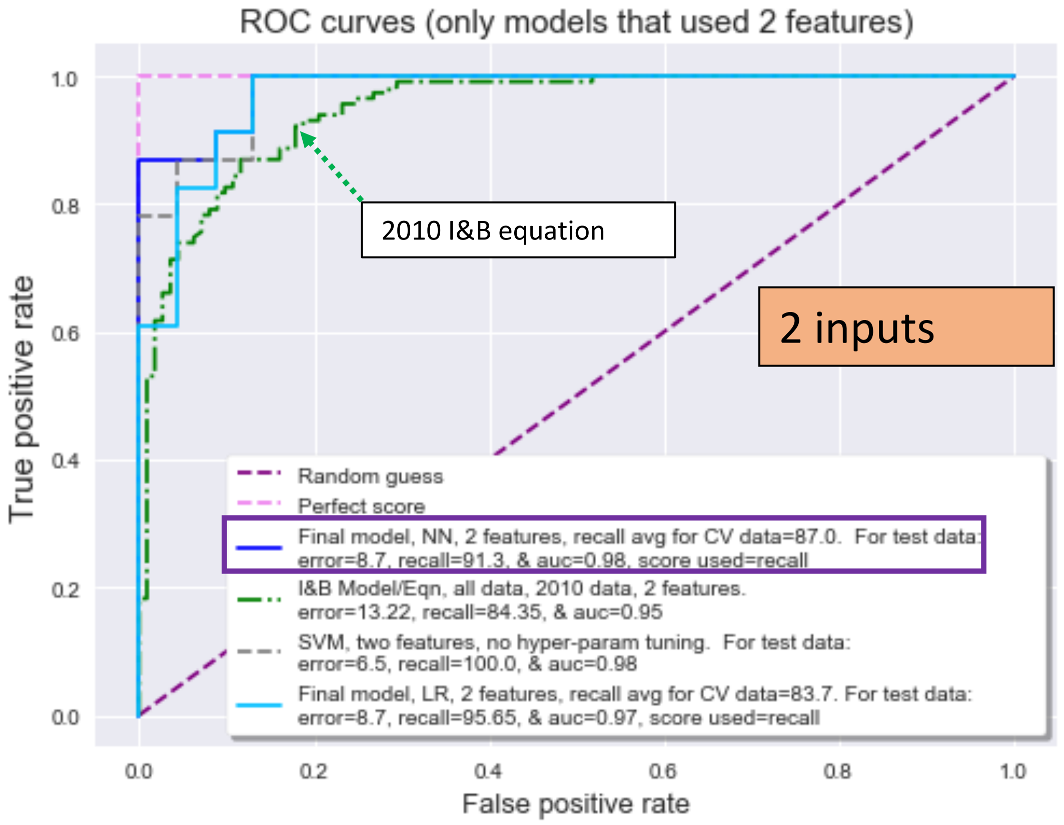
ROC AUC (higher is better):
NN 2 inputs = 98%
LR 2 inputs = 97%

NN all inputs = 97%
I&B 2 inputs = 95%
LR all inputs = 86%

Using ML to Re-evaluate Idriss and Boulanger – Results

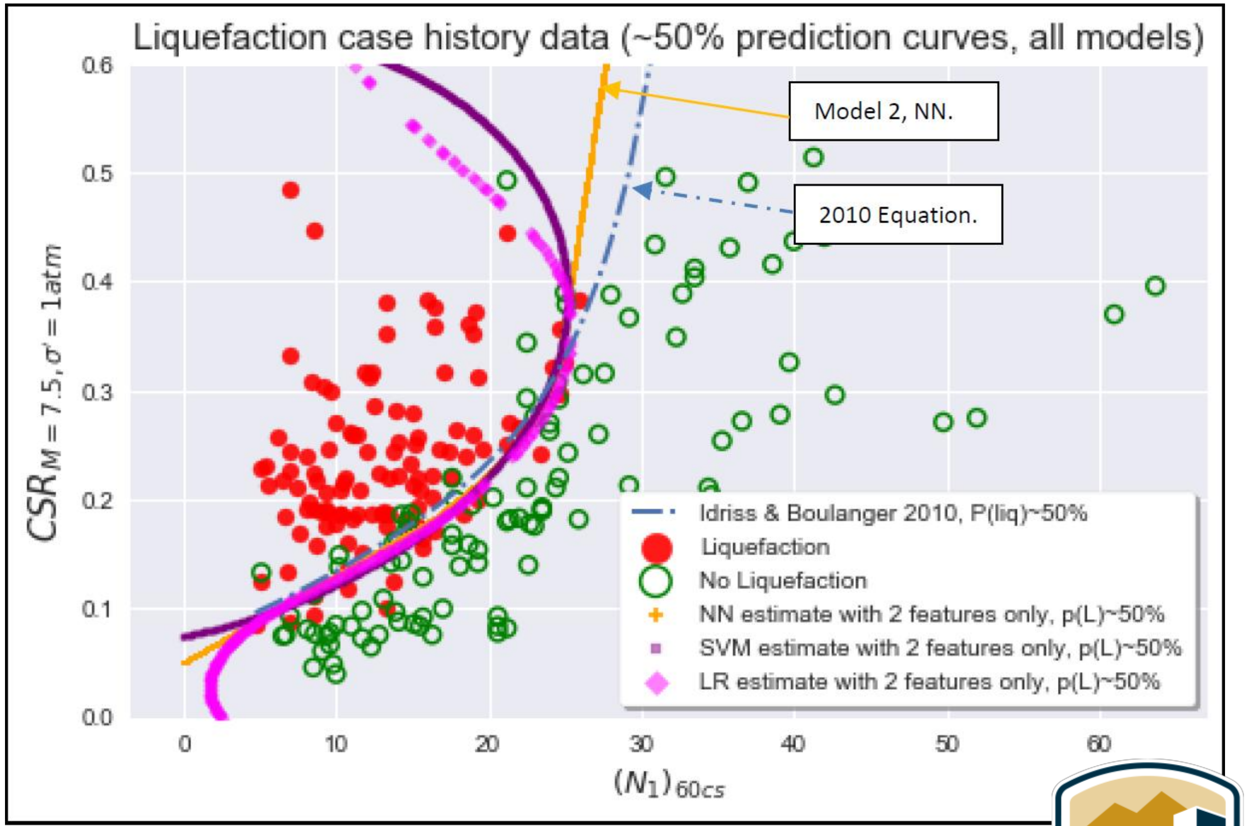
- Next step: Evaluate the ROC curve of all the models





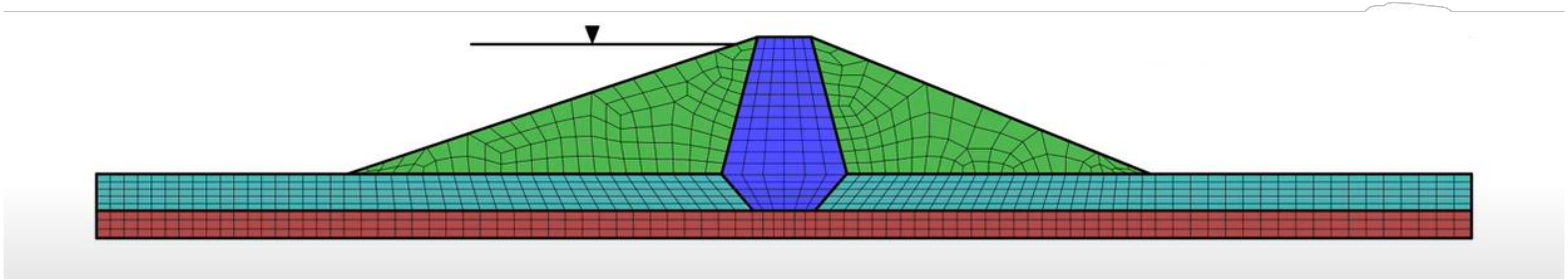
Using ML to Re-evaluate Idriss and Boulanger – Results

- More insight can be gained comparing the two-component models to the Idriss and Boulanger equation.
- Logistic regression and SVM models do not produce justifiable results from an engineering perspective.
- Highlights how even though the logistic regression produced the best assessment metrics, these models should still be reviewed by a subject matter expert (SME).



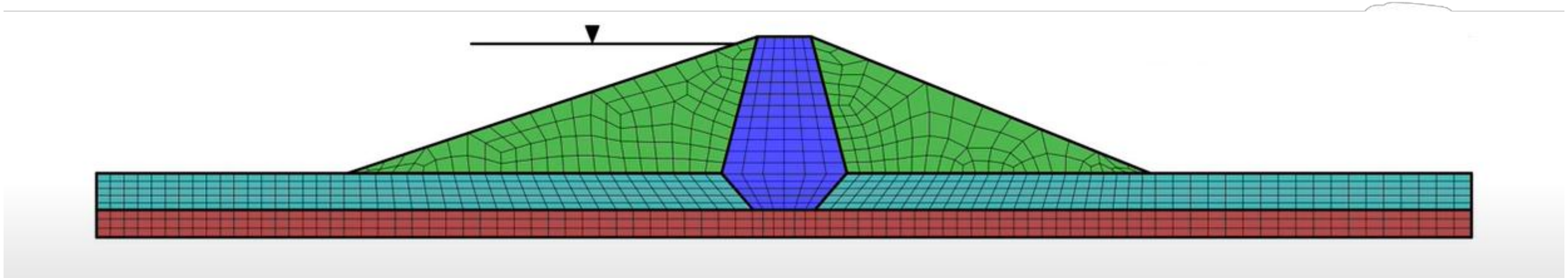
Case Study 2: Screen Dam Deformation Using Machine Learning

- While there are many steps in examining a dam's risk of deformation, one of the major steps is numerical modeling. At Reclamation, this is usually done using FLAC 2D.
- However, these models can take 20-100+ staff days to complete the modeling process to identify deformation depending on the state of the project



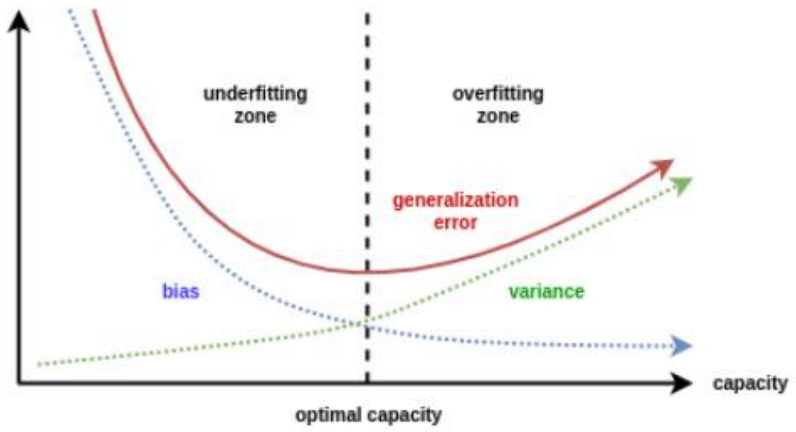
Screen Dam Deformation Using Machine Learning – FLAC2D Background

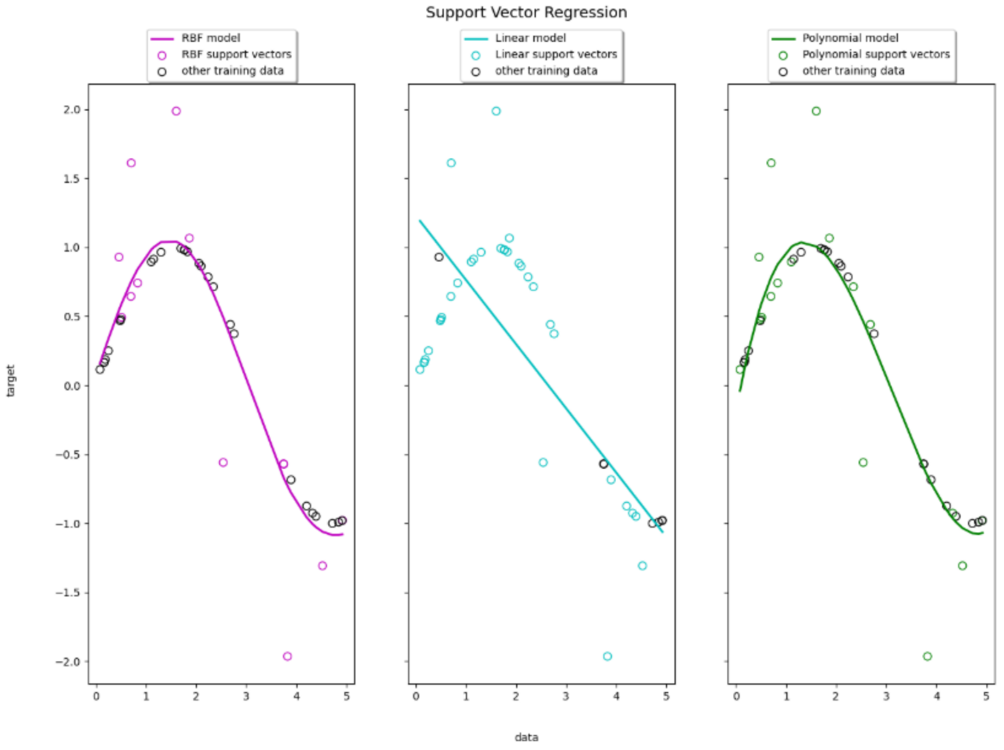
- FLAC2D (Fast Lagrangian Analysis of Continua in 2D) is a numerical modeling software based on the finite difference method (FDM).
- Soil strength modeled using Mohr-Coulomb.
- Cannot model 3D effects that could affect stability.
- Predicts seismic-induced deformations (e.g., crest settlement).



Screen Dam Deformation Using Machine Learning – Model Selection

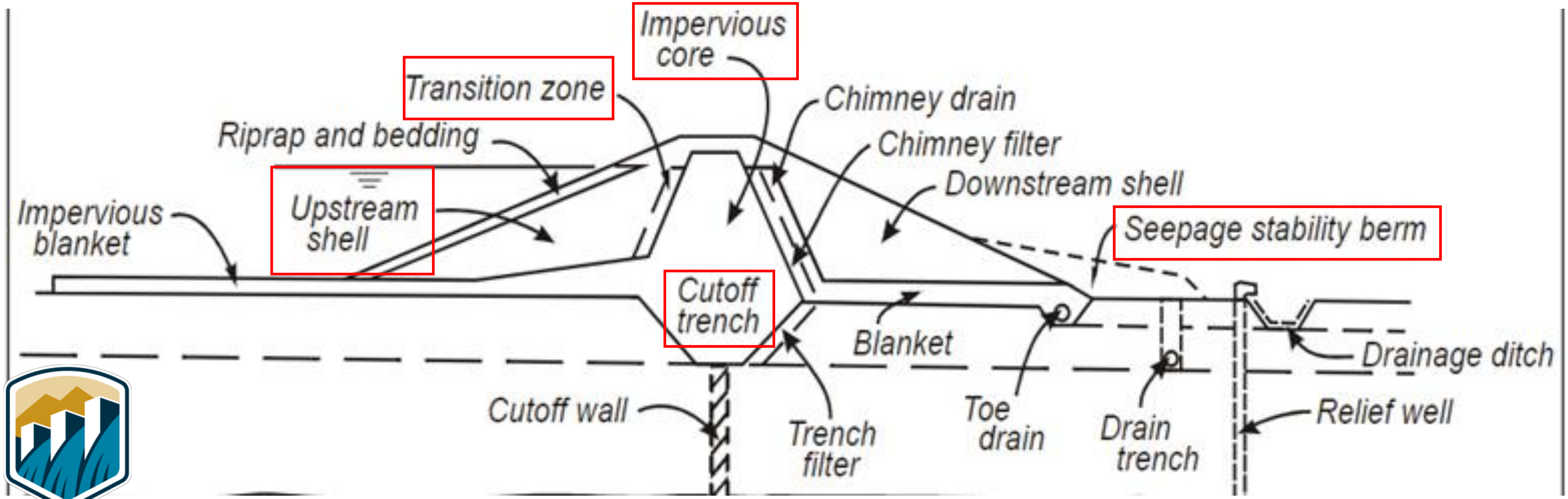
- We evaluated similar models to the previous case study but needed continuous outputs. The models evaluated:
  - Support Vector Regression (SVR)
  - Ridge/linear regression (Ridge)
  - Neural Network (NN)





Screen Dam Deformation Using Machine Learning – Parameter Selection

- Model inputs were based on input files from FLAC2D models with advice from Reclamation experts
  - Including dam geometric and soil strength parameters
  - Earthquake parameters
  - Target parameter of effective displacement

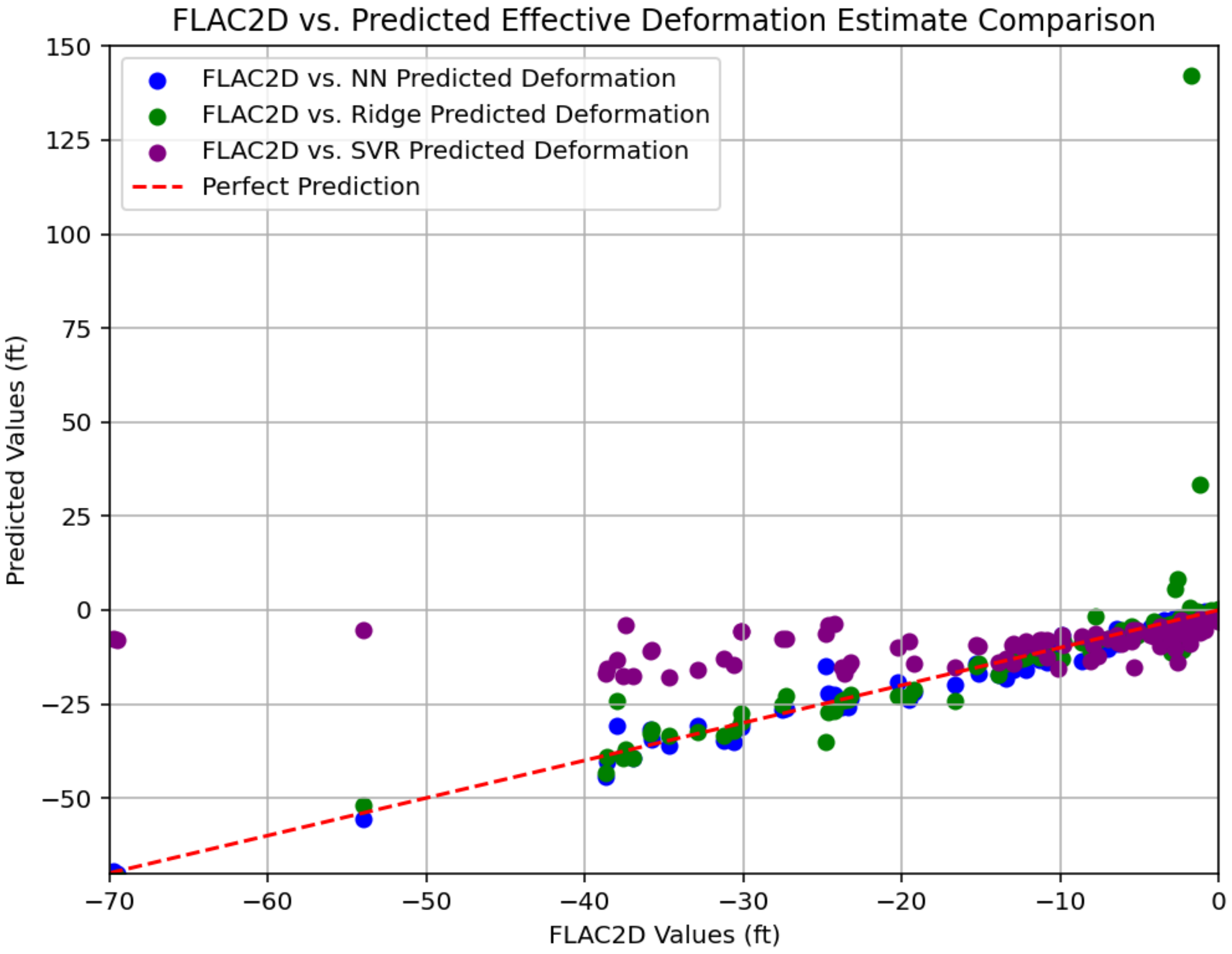


Screen Dam Deformation Using Machine Learning – Input Source

- Inputs came from approximately ~550 FLAC2D models input files
  - Only models that used Mohr-Coulomb for soil strength (cannot use new tools like PM4Sand/PM4Silt)
  - Models used were a mixture of dams in their current conditions and proposed modifications to dams.
  - The return period of earthquakes used in the models ranged from the 1,000- to 100,000-year earthquake.
  - Only models deemed to be of “high quality” by Reclamation SMEs were used.
  - 20% of the data was saved for testing

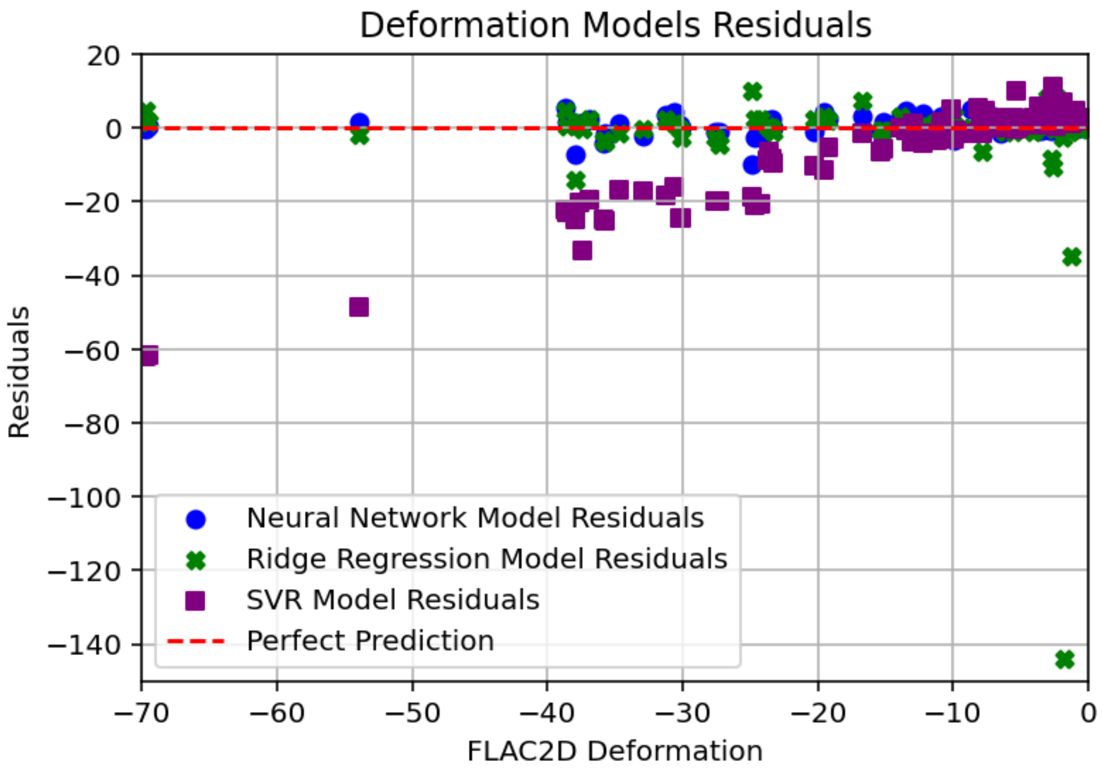
Screen Dam Deformation Using Machine Learning – Results

- Overall model performance using mean square error:
  - NN: 12
  - Ridge Regression: 244
  - SVR: 219



Screen Dam Deformation Using Machine Learning – Results

- SVR model generally predicts little deformation regardless of the inputs.
- Regression model produces better results but has substantial outliers.
  - Including impossible crest height increases
- NN generally provides best results.



Screen Dam Deformation Using Machine Learning – Discussion

- These results show that we can successfully estimate effective deformation being produced by FLAC2D model data.
- NN outperformed the other models in terms of mean squared error, but this may be due to some outliers produced by the ridge regression model.
- NN training took about 30 minutes.
- There are some caveats to this work:
    - It is a small dataset and due to needing for quality screening it is difficult to increase the dataset size.
    - Only can take inputs from one type of soil model, limiting the input pool.

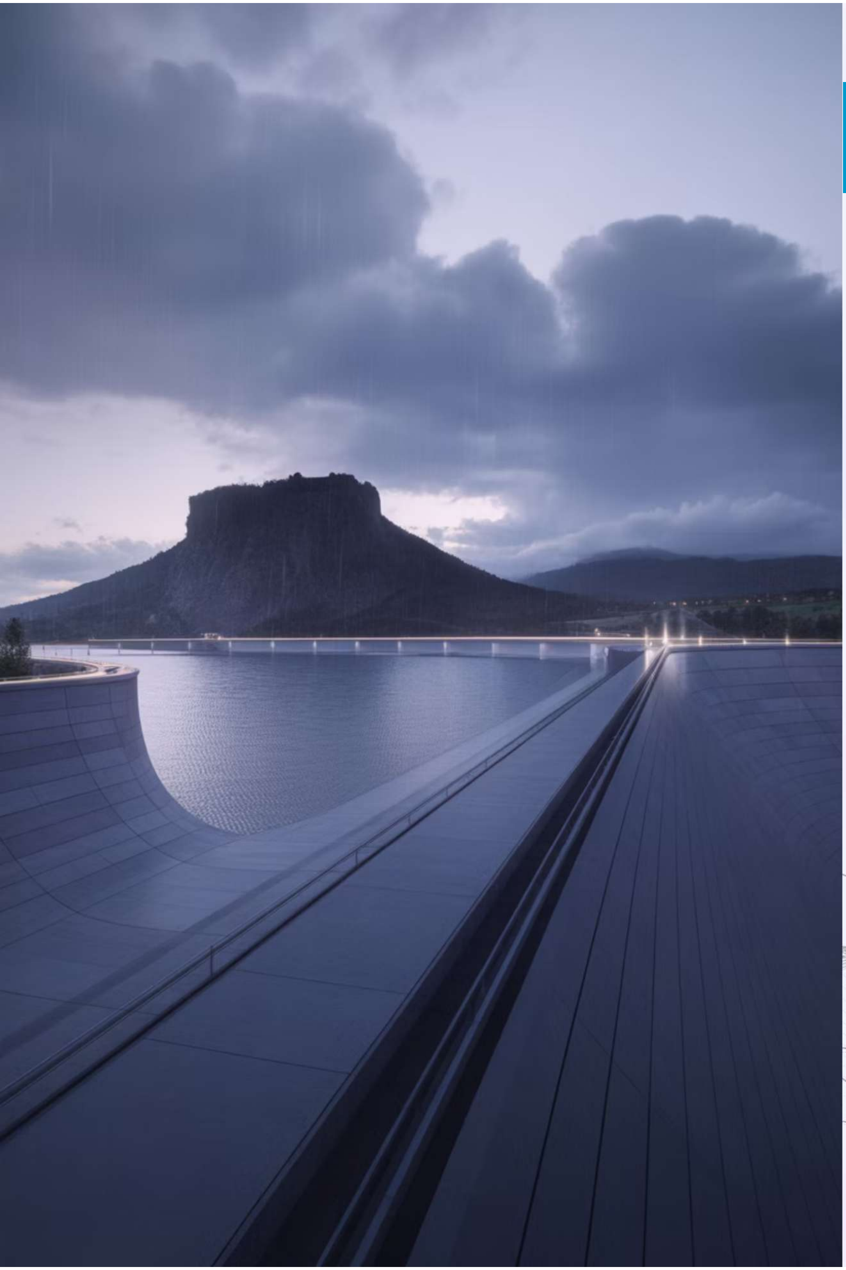
Conclusions

- Machine learning models can produce results similar to the "accepted" solution for liquefaction triggering with improved performance assessment metrics.
- A neural network was able to produce similar deformation results to FLAC2D, while requiring just minutes rather than multiple staff days. It has potential as a screening level analysis tool.
- Human input is incredibly important in the process; otherwise, you could end up with results that are not physically possible.

Thank you!

Cameron Carpenter, ccarpenter@usbr.gov
+1 303-445-3544

BUREAU OF
RECLAMATION



Hybrid AI-Hydrological Models for Full-Hydrograph Flood Forecasting in Reservoir Catchments

Prof. Li-Chiu Chang (張麗秋)

Dept of Water Resources and Environmental Engineering
Tamkang University

ICWE

Why Full-Hydrograph Forecasting Is Critical

Complete Flood Picture

Captures the entire flood event, from initial rise to final recession, providing critical timing and magnitude for reservoir operations.

Operational Precision

Supports accurate decisions on peak flows, timing, and recession characteristics for reservoir releases and storage.

Risk Mitigation

Enables proactive risk management, optimizing storage capacity and coordinating downstream flood protection.

The Critical Challenges of Flood Forecasting

◆Rainfall-Runoff Challenges

- ✓ Catchment heterogeneity
- ✓ Variable soil moisture conditions
- ✓ Spatial rainfall uncertainties
- ✓ Land use change impacts

◆Recession Flow Issues

- ✓ Non-linear storage-discharge relationships
- ✓ Baseflow separation difficulties
- ✓ Groundwater-surface water interaction

These challenges reduce forecasting accuracy, undermining flood risk management and reservoir operation effectiveness, especially during extreme events.

Bridging Research and Practical: Experience from Shihmen Reservoir

1 Data-Driven and Physically Interpretable

- ✓ **Extract key flood characteristic** from data
- ✓ Develop models that combine **data-driven learning** with **physical interpretability**, enhancing decision-makers' confidence.

2 Real-Time Data and Visualization

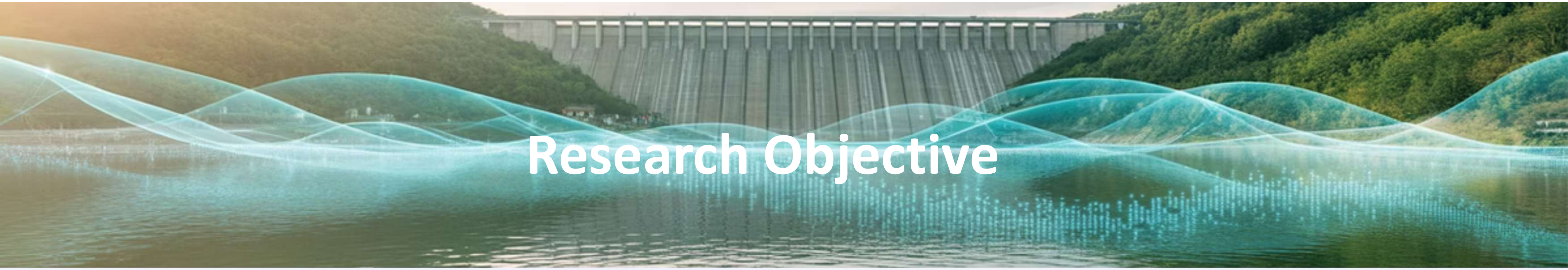
- ✓ Automate data processing and forecasting workflows.
- ✓ Provide **interactive and visualized outputs** to support rapid decision-making during flood events.

3 Human Capacity and Cross-Disciplinary Expertise

- ✓ Train professionals across hydrology, meteorology, and data science.
- ✓ Introduce **AI expertise** to build sustainable application capacity.

4 Innovation Driven by Practical Demands

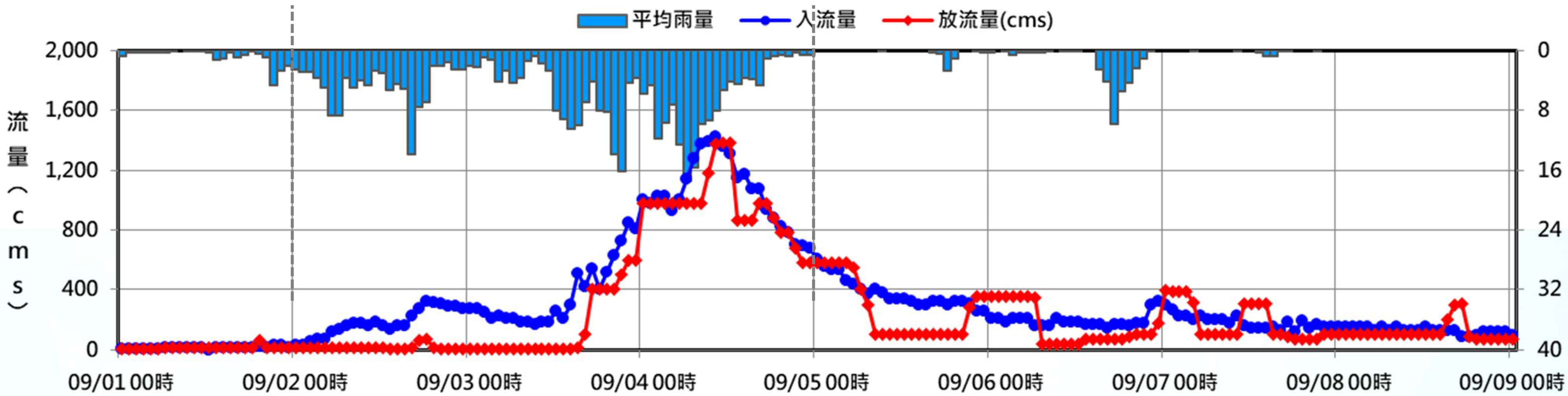
- ✓ Bridge gaps between **operational needs** and scientific research.
- ✓ Stimulate innovation in accuracy, stability, self-learning, and real-time performance.



Research Objective

Develop a hybrid framework **integrating AI and hydrological methods** for robust flood forecasting across the full hydrograph (rising, peak, and recession).

Theoretical vs. Practical Perspectives on Full-Hydrograph Forecasting



Pre-typhoon stage | Pre-peak stage | Post-Peak Stage

Theoretical flood forecasting domain (≈3 days)

Gap: time-scale difference

Practical need: full-hydrograph domain (≈1 week)

No-rain/ scattered rainfall stage

Hybrid Modeling Approach: Integrating AI and Hydrology

This study presents an innovative solution that combines the strengths of two complementary modeling paradigms: the AI-based RNARX (Rainfall-Runoff Autoregressive with Exogenous Inputs) model and the traditional hydrological storage function model.

RNARX → Rising & Peak

AI-driven, effective in rising & peak Periods

Storage Function → Recession

Physically-based, accurate in recession phase

Hybrid Integration → Full Hydrography

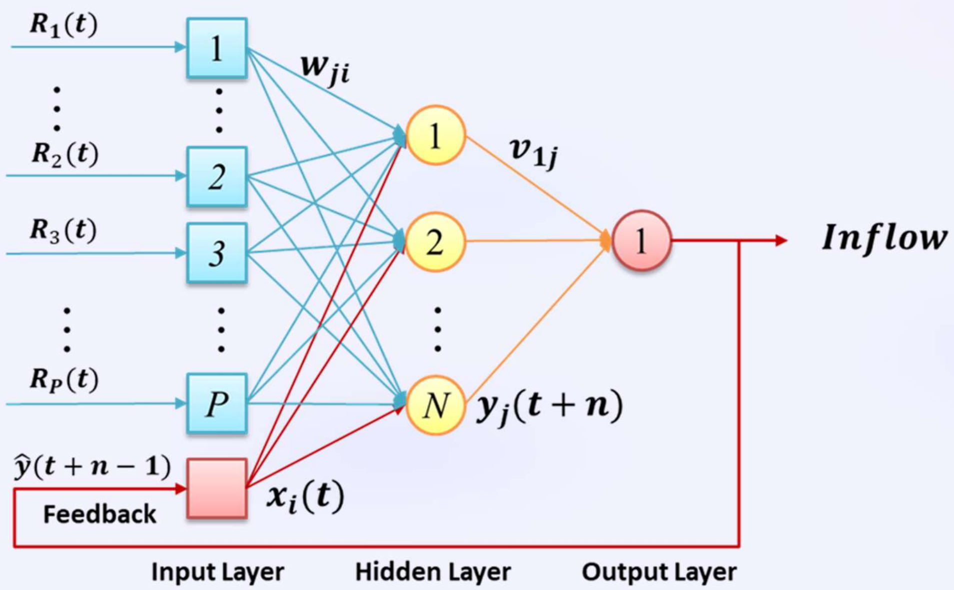
Seamless switching, robust across full hydrograph

This hybrid approach overcomes the limitations of each method, delivering robust full-hydrograph forecasting.

Understanding RNARX: The AI Component

Recurrent Nonlinear Autoregressive with Exogenous Inputs

- AI-driven rainfall-runoff model with a recurrent architecture using exogenous inputs (primarily rainfall)
- Learns nonlinear dynamics and adapts to catchment-specific responses; handles high-resolution time series.
- Best during rainfall & peak, capturing rapid rises and accurate peak flows.



Note: RNARX's strength is modeling fast flow changes driven by rainfall intensity—ideal for **peak-flow estimation**.

RNARX Performance Characteristics

Strengths

- High accuracy during intense rainfall events
- Captures rapid flow increases effectively
- Reliable peak-flow estimation
- Handles complex rainfall-runoff dynamics

Limitations

- Performance degrades when rainfall approaches zero
- Predicted flows drop too quickly post-rainfall
- Underestimates recession flows
- Struggles to sustain accuracy through baseflow periods

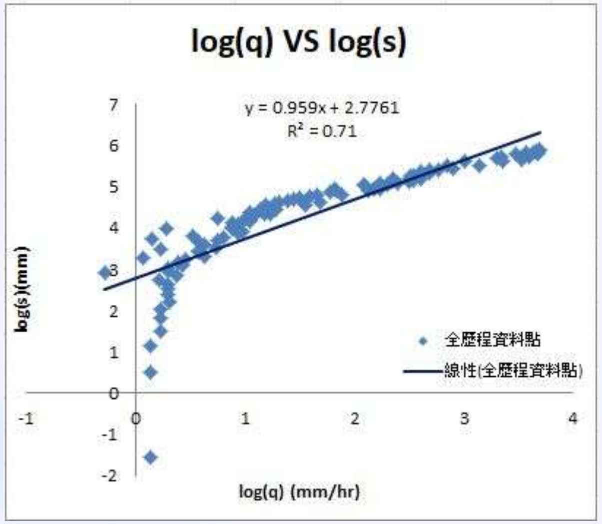
RNARX is powerful during rainfall and peak flows, but it reliance on rainfall input leads to significant errors during the recession phase.

The Storage Function Model: Hydrological Foundation

- **Conceptual hydrological model** linking streamflow to catchment storage
- Captures watershed storage characteristics, reflecting the interaction between surface water and groundwater
- **Fundamental principle**: Streamflow discharge rises as storage increases and gradually declines as storage depletes.
- **Operational relevance**: Widely applied in Japan and Taiwan for flood forecasting due to its physical interpretability and simple calibration.

$$S_t = KQ_{e,t}{}^{P}$$

s_t ： Water storage depth at time t (mm)
$Q_{e,t}$ ： Effective runoff depth at time t (mm/hr)
$K\&P$ ： Parameters of the Storage Function



Storage Function Model Performance

Strengths

- Excellent performance during the recession phase
- Accurately simulates gradual flow decline.
- Reflects catchment storage characteristic
- Physically-based parameter, simple to calibrate

Limitations

- Limited accuracy during the early rising phase
- Tends to overestimate flows under intense rainfall
- Significant errors in peak flow prediction
- Cannot capture rapid hydrological changes.

The Storage Function model is ideal for representing recession flows and storage processes, but it struggles with rapid runoff dynamics and high intensity rainfall.

Defining the Transition Criteria

Model Switching Logic

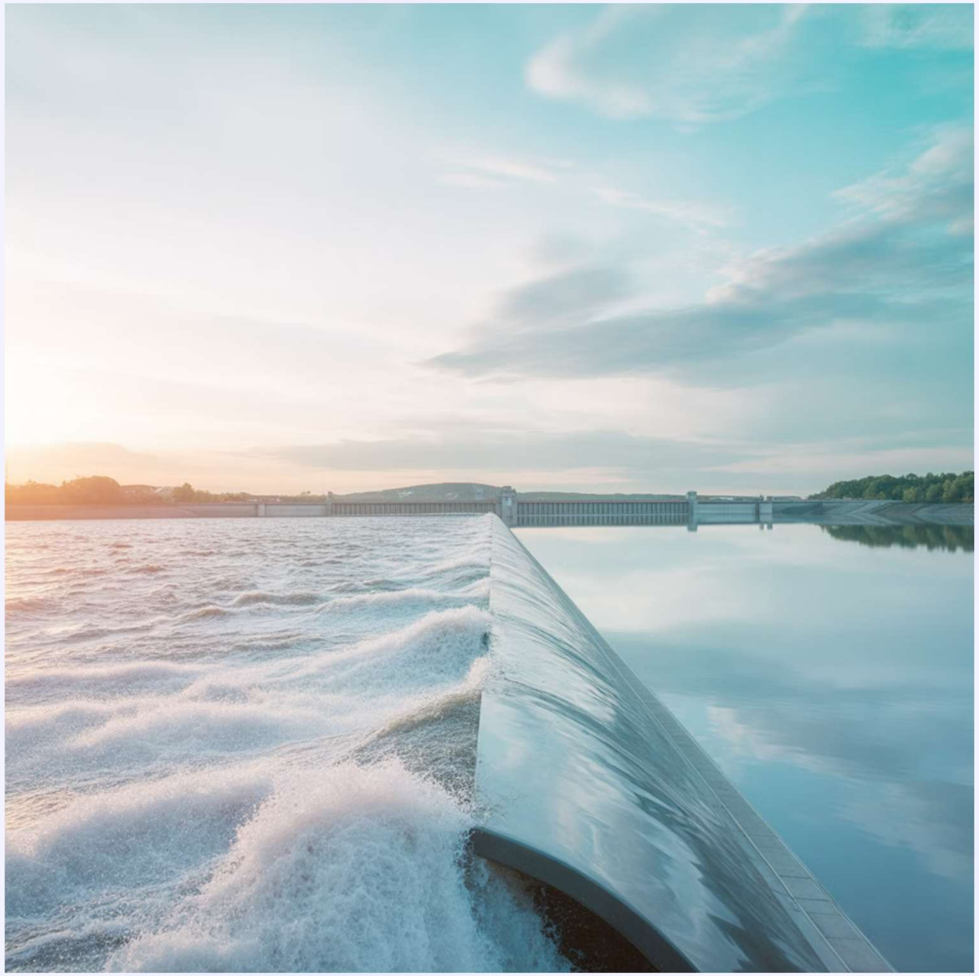
RNARX Active When:

- ✓ Rainfall intensity exceeds threshold
- ✓ Rising limb or peak flow stage
- ✓ Rapid flow variations observed

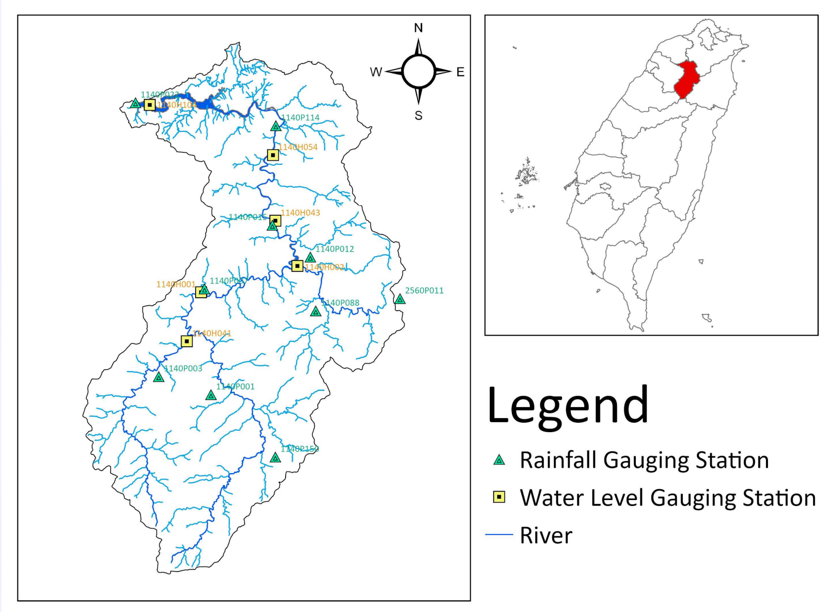
Storage Function Active When:

- ✓ Rainfall falls below threshold
- ✓ Recession phase dominates
- ✓ Baseflow conditions prevail

Rainfall intensity serves as **the key trigger** for switching, ensuring the hybrid model adapts dynamically across the full hydrograph.



Shihmen Reservoir Watershed



Location: Northern Taiwan

Area: 763.4 km^2

Effective Storage Capacity: 2.0526×10^8 m^3 (third largest in Taiwan)

Average Annual Rainfall: 2,280 mm (2019~2024)

Maximum Flood Peak: 8,594 cms (2004 Typhoon Aere)

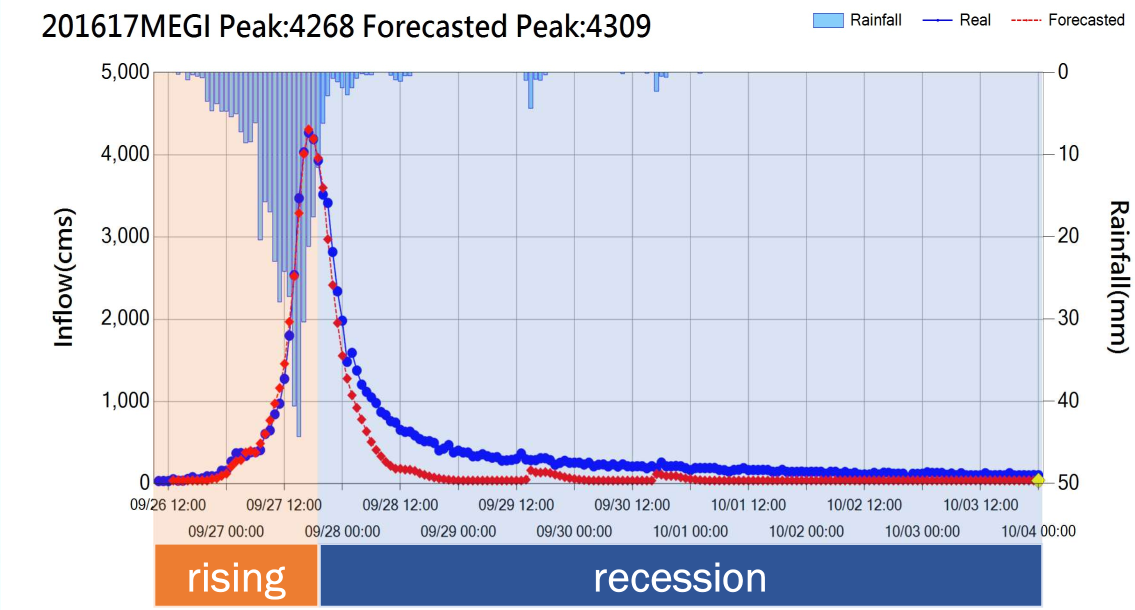
Limitations of Single-Model Approaches

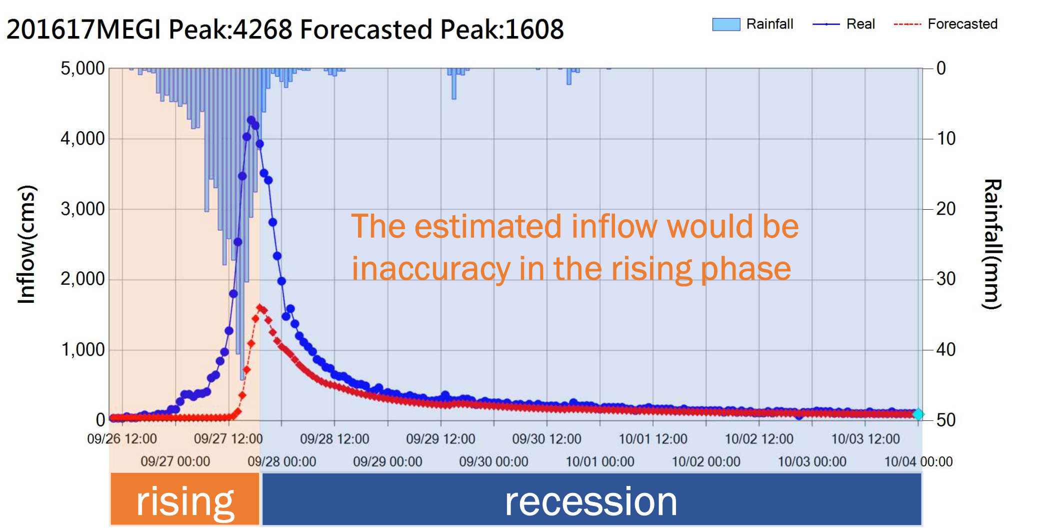
RNARX Model

- Performs well int the rising phase (rapid response to rainfall inputs)
- Tends to underestimate flow during the recession phase

Storage Function

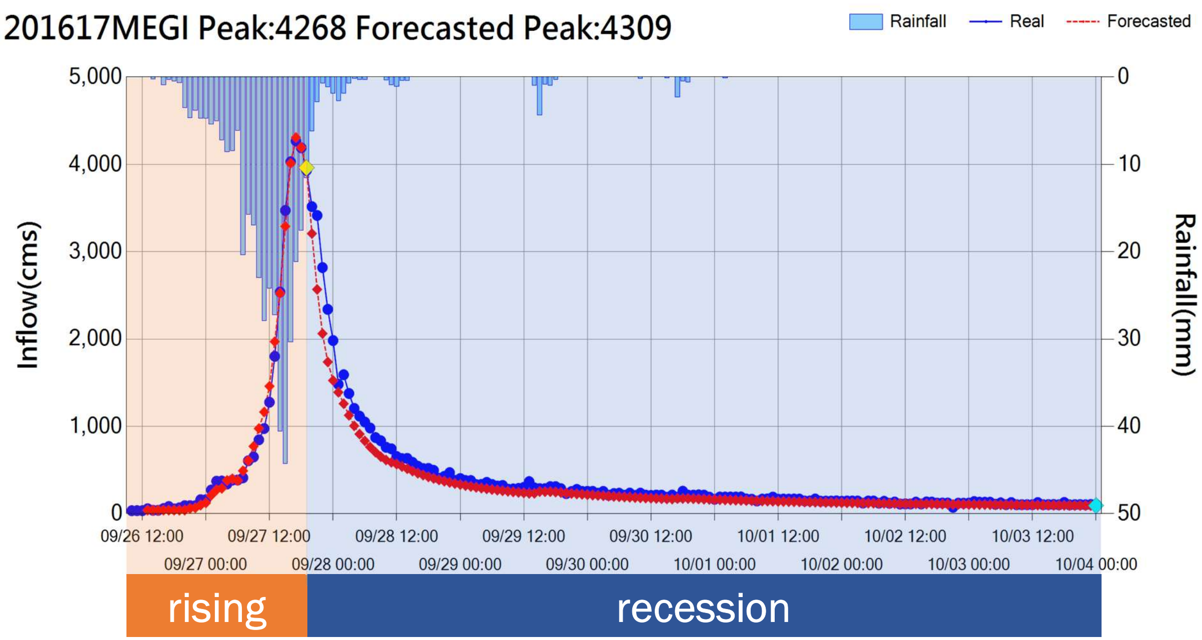
- Excels in the recession phase (capture gradual flow decline
- Tends to misrepresent flows during the rising phase and peak flow estimation



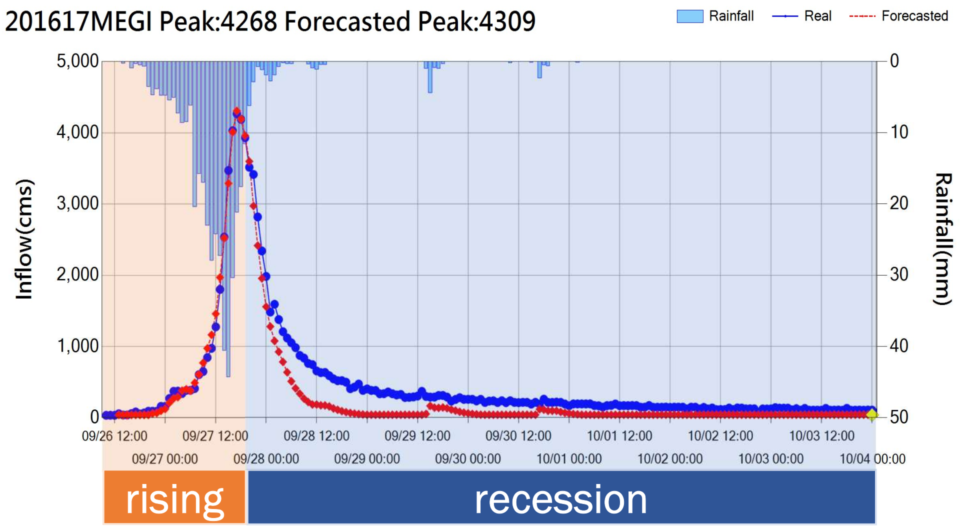


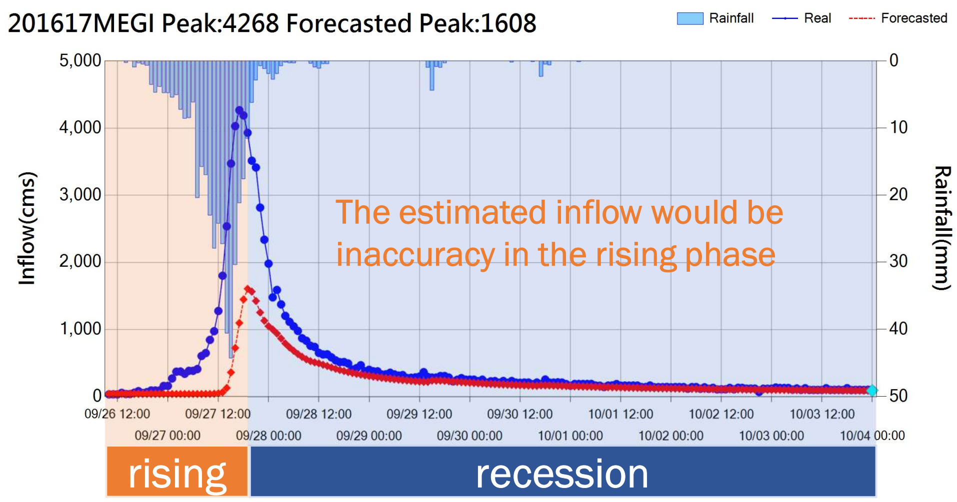
Hybrid Model of Full-Hydrograph Forecasting

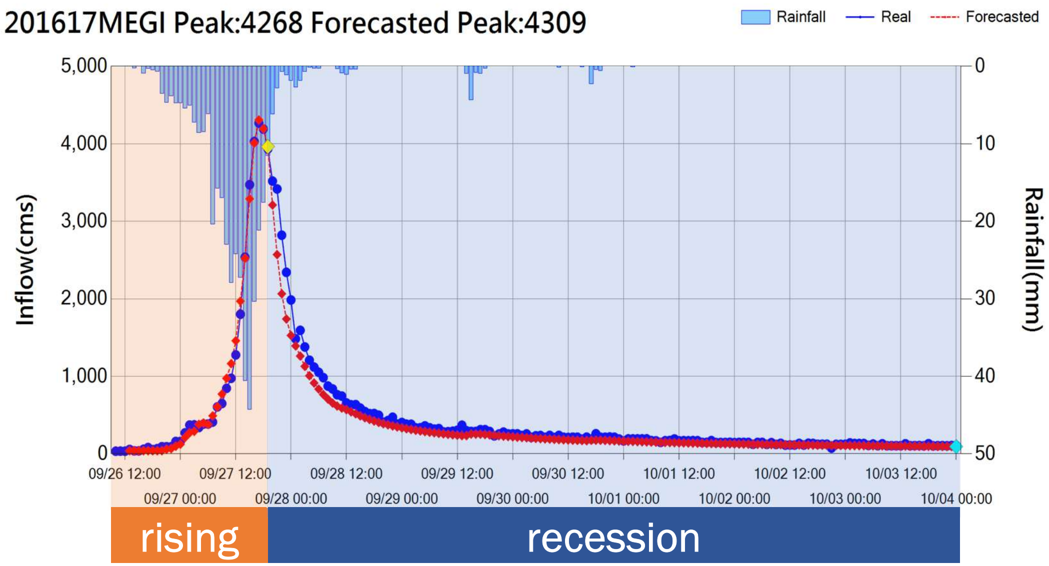
- ✓ Integrates RNARX and Storage Function
- ✓ RNARX active during rainfall and peak flows
- ✓ Storage Function active during recession phase
- ✓ Switching criteria based on rainfall intensity and flow stage
- ✓ Provides accurate prediction of the entire flood hydrograph



Hybrid Model Performance – Case Example

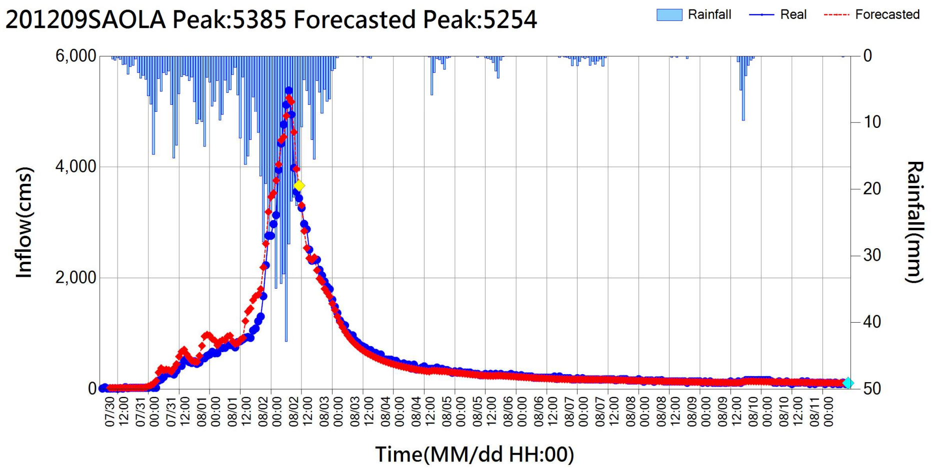


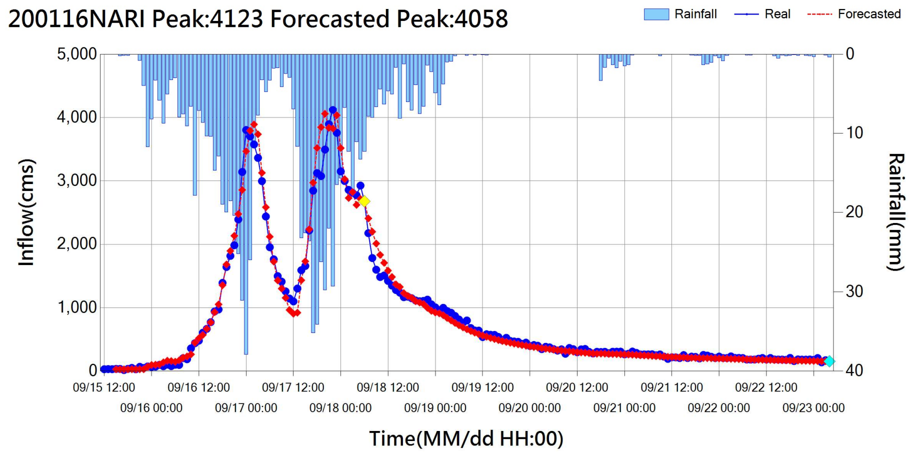




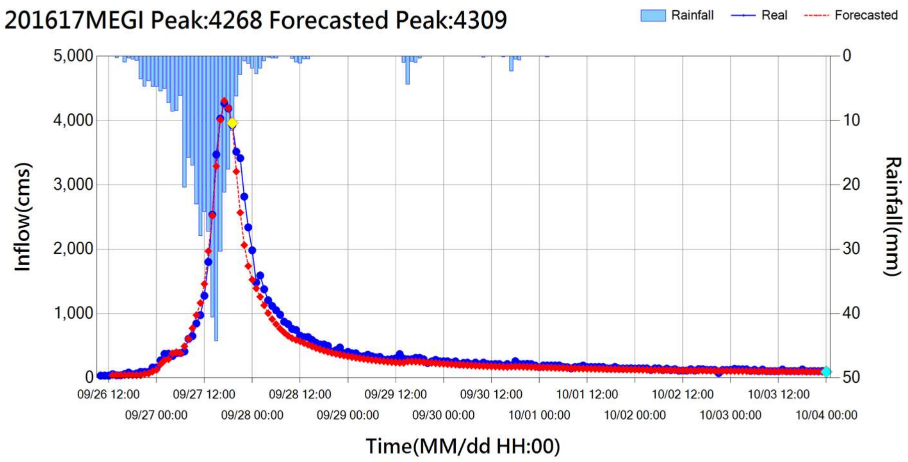
- ✓ **RNARX alone** → underestimates recession flows
- ✓ **Storage Function alone** → misrepresents rising phase and peak
- ✓ **Hybrid Model** → accurately captures both rising and recession phases
- ✓ Significant improvement in **full-hydrograph prediction**

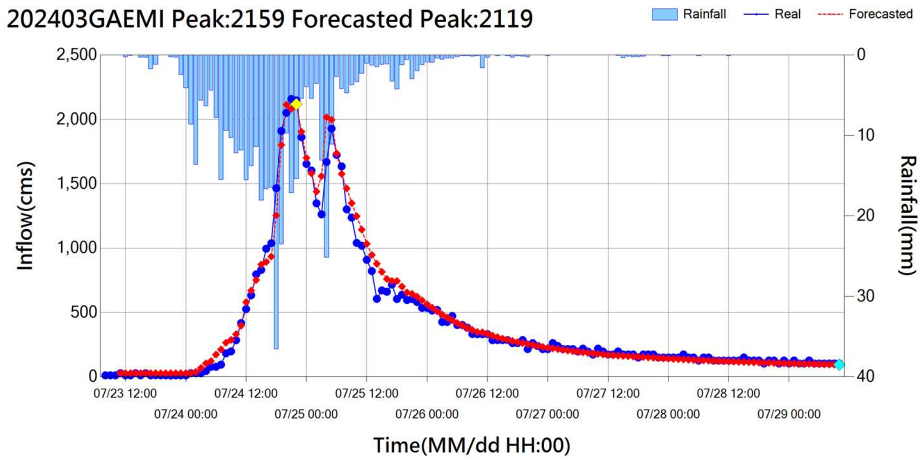
Consistency of Hybrid Model in Multiple Typhoon Events





Hybrid consistently outperforms RNARX and Storage





Consistency of Hybrid Model in Multiple Typhoon Events

- 4 typhoon cases: **Saola, Nari, Megi, Gaemi**
- Hybrid model consistently achieves:
- **Lowest RMSE**
- **Highest R^2**
- **Lowest MAE**
- Confirms robustness and general applicability

Typhoon	R-NARX			Storage Function			Hybrid Model		
	RMSE	R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE
201209 SAOLA	224	0.97	178	625	0.69	235	141	0.98	69
200116 NARI	281	0.94	199	356	0.91	124	132	0.99	74
201617 MEGI	72	0.82	59	158	0.97	104	126	0.98	64
202403 GAEMI	219	0.98	168	622	0.68	306	68	0.98	39

Operational Implementation Benefits

Extended Lead Time

Reliable forecasts up to 72 hours ahead, giving operators more time for proactive planning and coordination.

Precision Operations

Accurate hydrographs help optimize releases, preserve flood storage, and minimize unnecessary spills.

Enhanced Safety

Reliable recession forecasts reduce premature refilling and ensure sufficient capacity for subsequent storms.

Conclusion

- **Hybrid RNARX–Storage Function** approach integrates strengths of both models.
- Demonstrates **consistent improvement** in flood forecasting accuracy across multiple typhoon events.
- Provides a practical and reliable framework for **full-hydrograph inflow prediction**.
- Offers **clear operational value** for reservoir flood management in Taiwan and beyond.

AI in hydrometeorological applications
人工智慧於氣象水文之創新應用

Li-Pen Wang (lpwang@ntu.edu.tw)
National Taiwan University

International Forum on Smart Watershed, AI-Driven Innovation
Taiwan Int'l Water Week 2025
2025/10/29

NTU
CompHydroMet Lab

Li-Pen Wang

Education

- **PhD in Civil & Env. Engineering**, Imperial College London (2008-2012)
- **MSc in Computer Aided Engineering**, National Taiwan University (2003-2005)
- **BSc in Civil Engineering**, National Taiwan University (1999-2003)

Professional Experience

- **Associate Professor**, National Taiwan University (2025/08-now)
- **Expert Consultant**, IMF (2024-now), UK Met Office (2022-now)
- **Postdoctoral Researcher**, Imperial College London, KU Leuven
- **Head of CAT unit**, Microinsurance Catastrophe Risk Organisation
- **Founder/Director/Hydrometeorologist**, Rain++ Ltd.

Research Areas

- Hydrometeorology
- Remote Sensing
- Computational Statistics

Contents

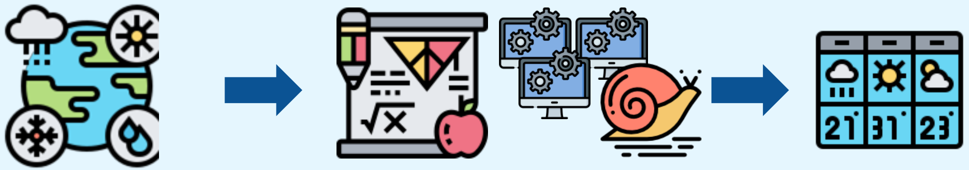
- AI in nowcasting isn’t new –so What’s it actually good at?
- Can we teach AI to say ‘I don’t know’?
- AI’s power of learning spatial and temporal features

Nowcasting

AI in nowcasting isn't new –so What's it actually good at?

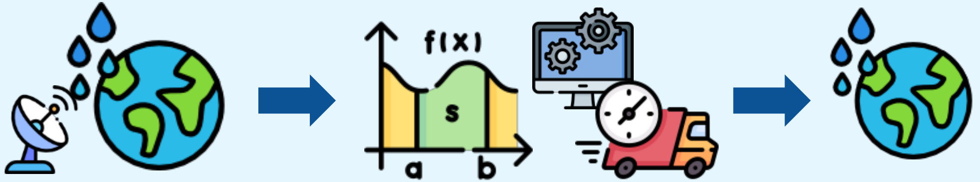
Radar-based nowcasting is effective and affordable

❑ Numerical weather prediction (NWP)

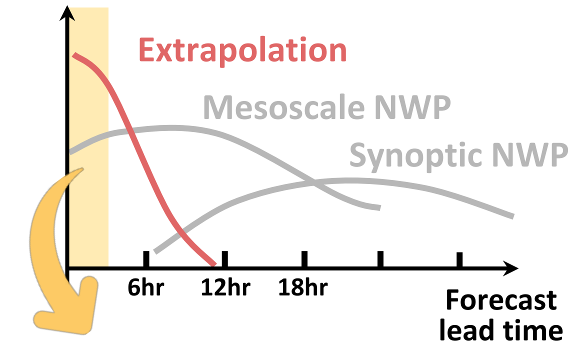


physical variables → mathematical models based on physical constaints → long-term forecasts

❑ Radar-based nowcasting



radar rainfall images → extrapolation from previous images → short-term forecasts



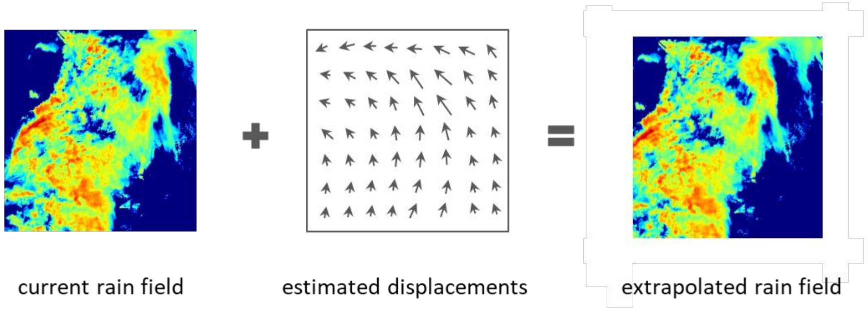
Target lead time of nowcasting

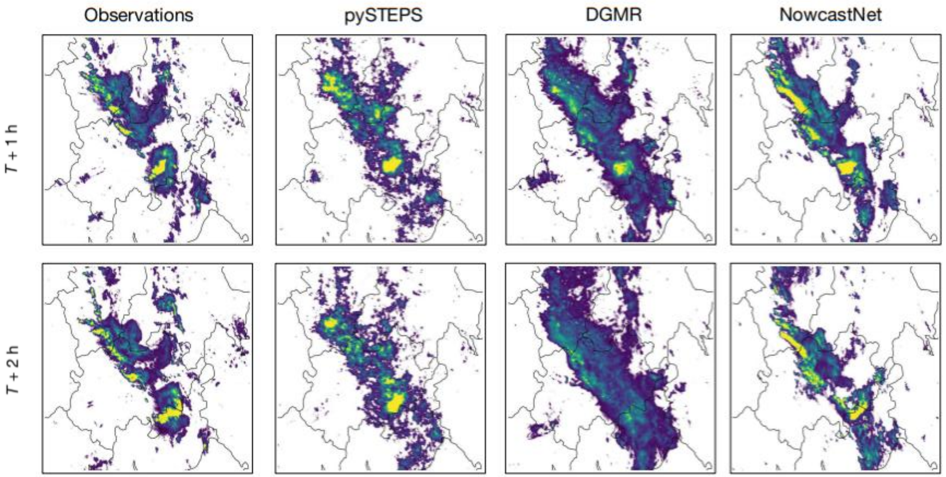
2~3 hours

For short forecast lead time,
radar-based nowcasting methods
are still more 'cost' effective than
NWP.

Field-based nowcasting

- **Advection nowcasting**
  - Extrapolates current radar rainfall fields into future frames using estimated displacements
- **STEPS**
  - State-of-the-art nowcasting model
  - **Optical flow** is employed for displacement estimation
  - Probabilistic nowcasting model
  - The spatial-temporal scaling relationship is explicitly modelled.

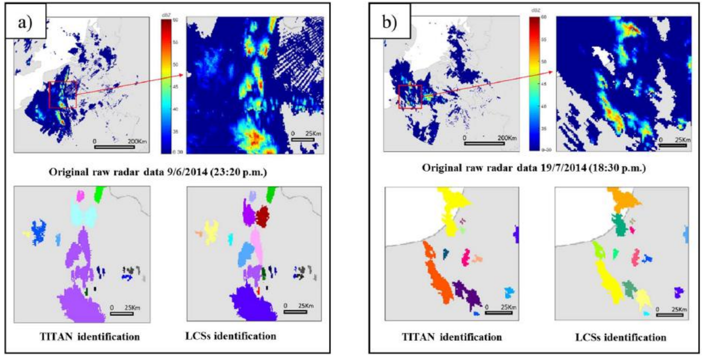


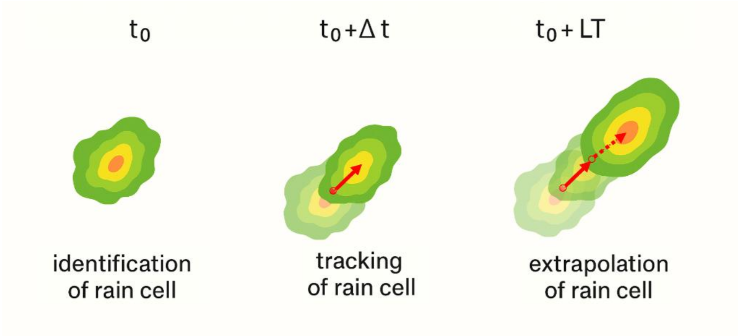


https://pysteps.github.io/

Object-based nowcasting

- **Object-based nowcasting**
  - Extrapolates the movement of identified rainfall cells
- **TITAN**
  - Most widely-used basis model
  - Storm identification
  - Temporal association of rainfall objects between successive time steps
  - Widely used for thunderstorm nowcasting



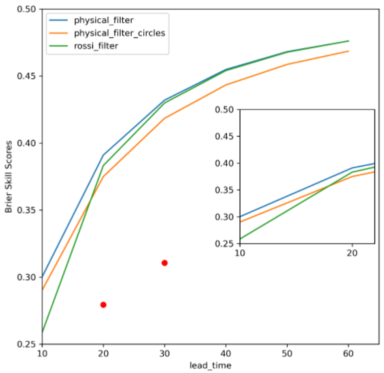


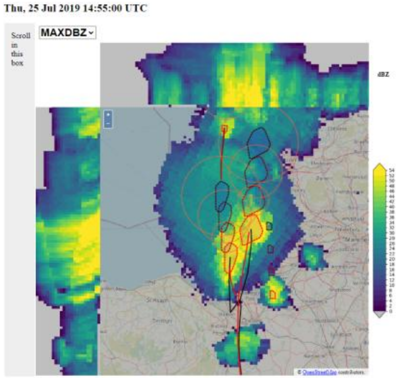
Challenges in modelling spatial-temporal rainfall process

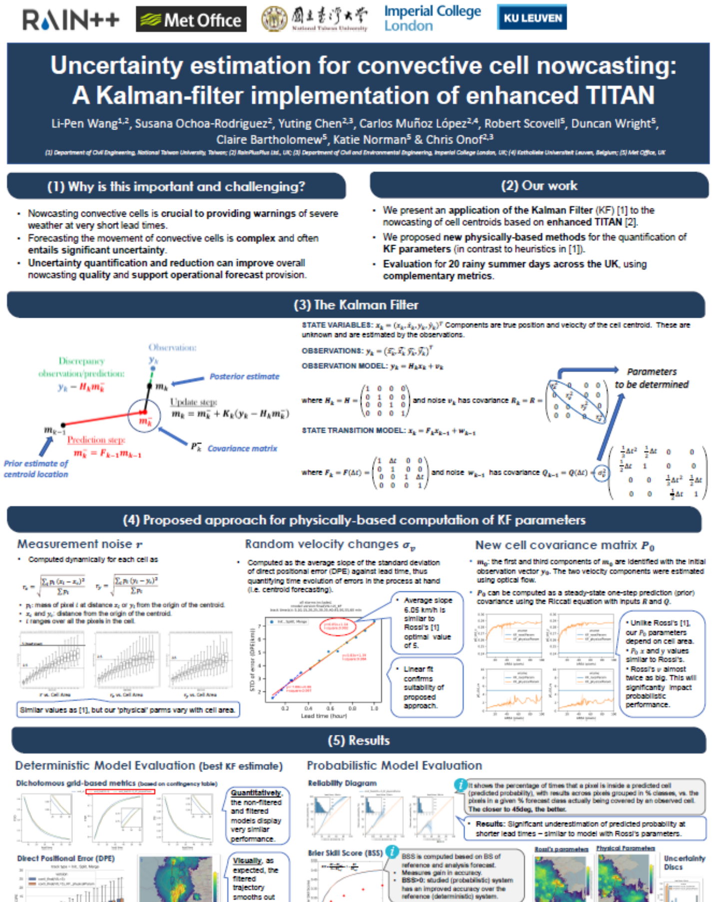
- Variations in rainfall = Advection + Evolution (in time)
- The variations in spatial and temporal features of rainfall are NOT independent from each other
- Preserving consistency across scales

A joint effort to develop an object-based probabilistic convective storm nowcasting system

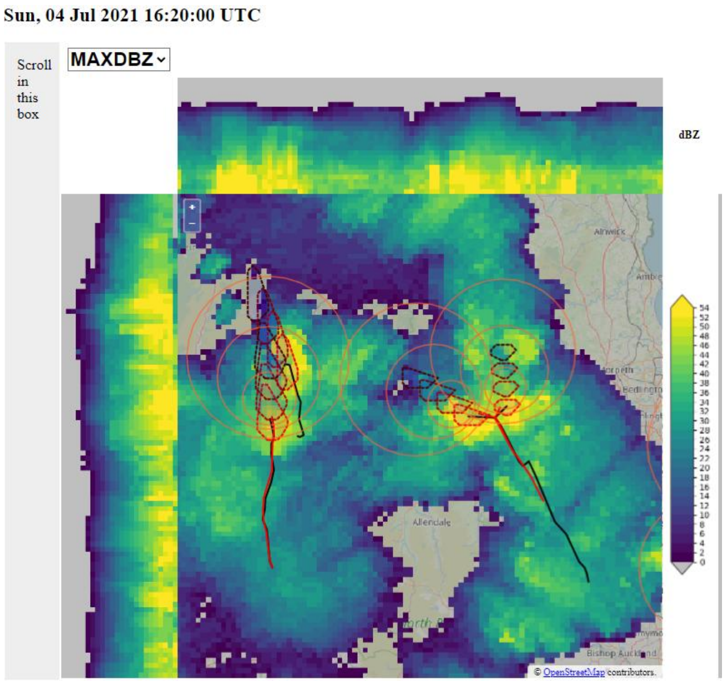
- A positional forecasting system based on Enhanced TITAN + Kalman filter, based on Rossi et al (2016)
- Physical-based computation of Kalman filter parameters
- More informative ensemble forecasts can be obtained at short lead time



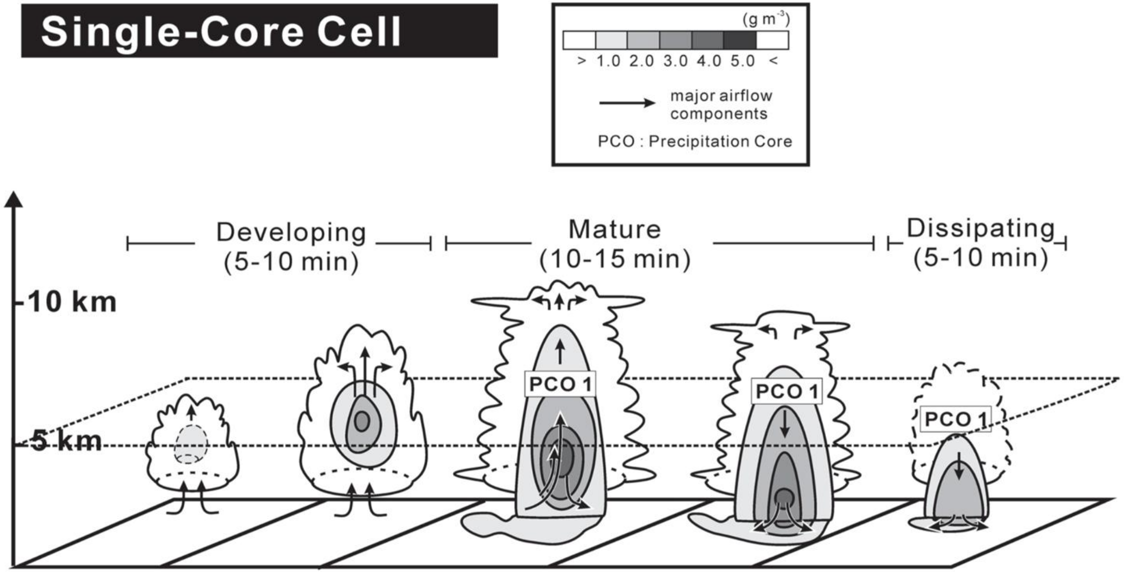




Advection is taken care of, *but what about rain cell evolution*?

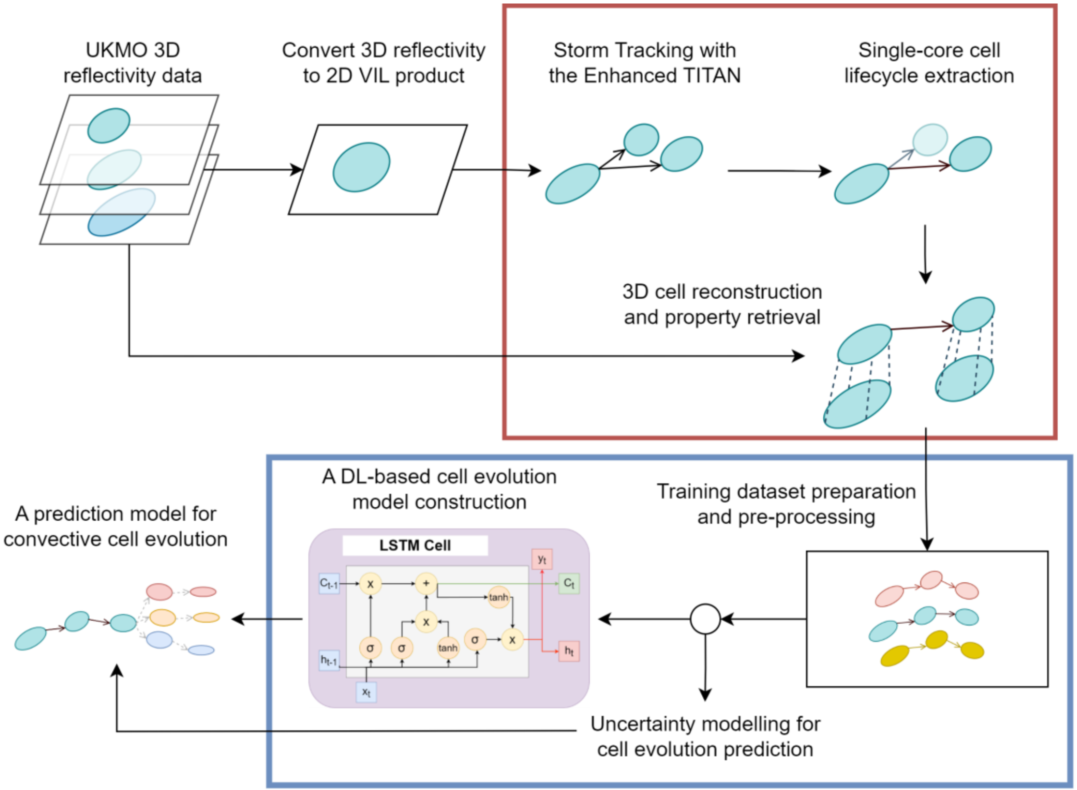


The mechanism of single-core cell evolution has been studied, but has not been incorporated into object-based nowcasting

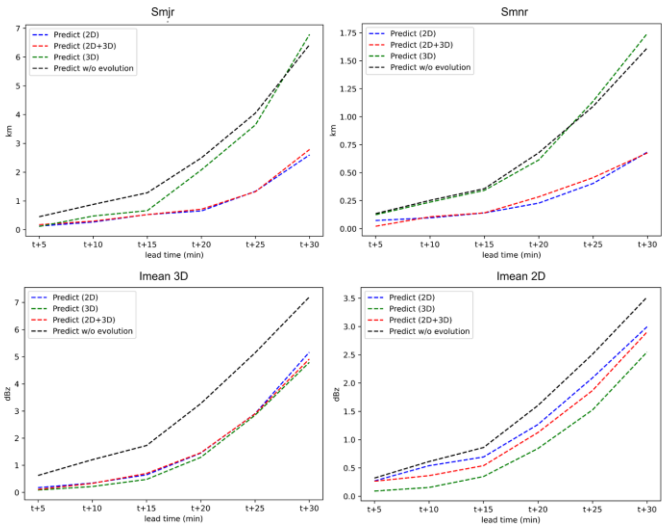


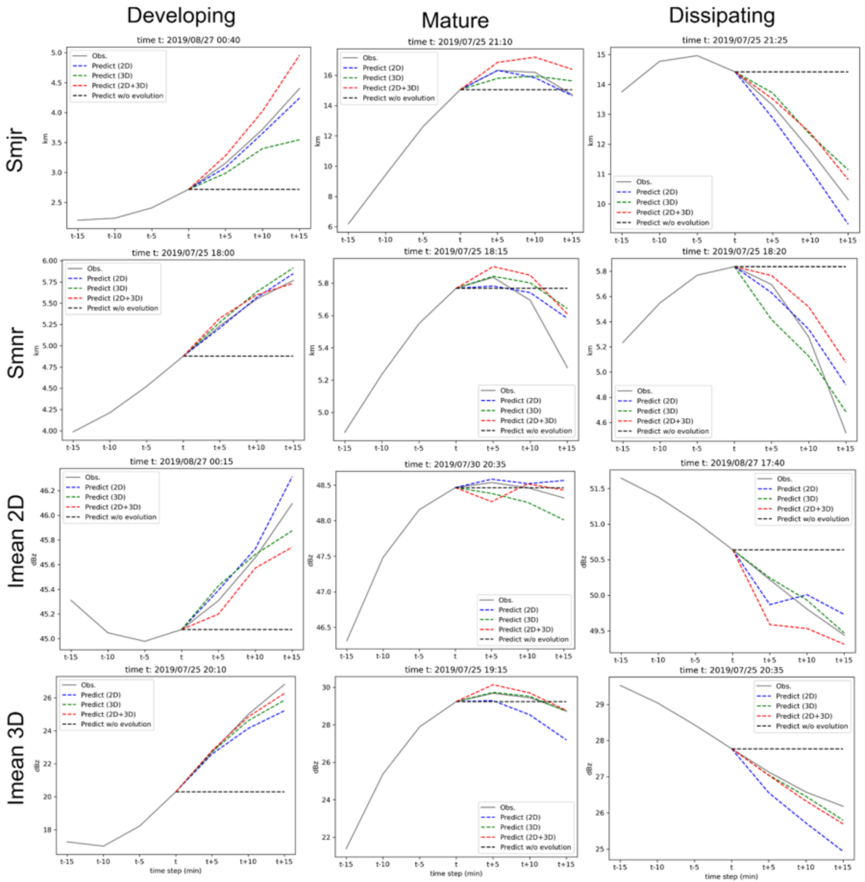
Our method

- Reconstruct single-core rain cell lifespans (in 3D)
- Train, test and validate DL-based cell evolution prediction model
- Model prediction errors



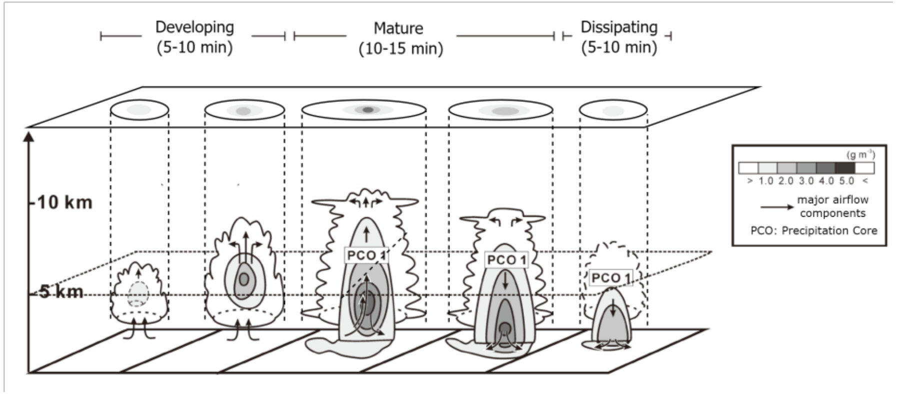
- A single LSTM model can well predict the evolution of rain cell properties at different life stages.
- Nowcasting intensity error (MAE) can be reduced by 50% at 15-min lead time.



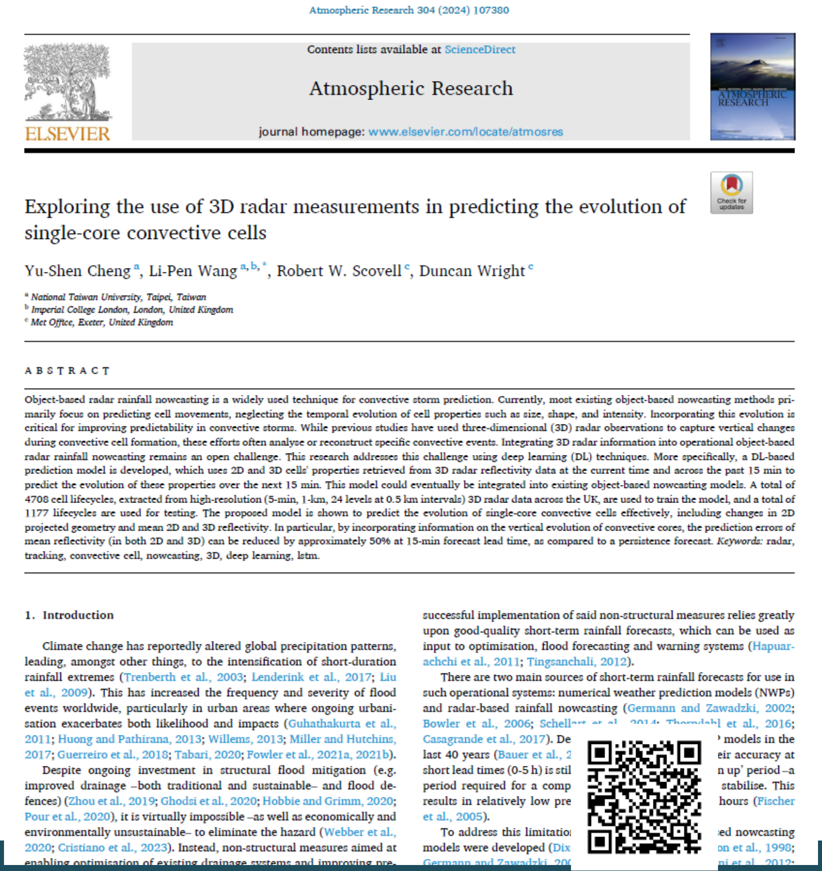


A deep-learning based model to predict convective cell lifecycles

- The altitude of convective cores found to be a good indicator of the underlying cell life stages



Adopted from Kim et al., 2012



Atmospheric Research 304 (2024) 107380

Contents lists available at ScienceDirect

Atmospheric Research

journal homepage: www.elsevier.com/locate/atmosres

ELSEVIER

Exploring the use of 3D radar measurements in predicting the evolution of single-core convective cells

Yu-Shen Cheng [a], Li-Pen Wang [a,b,*], Robert W. Scovell [c], Duncan Wright [c]

[a] National Taiwan University, Taipei, Taiwan
[b] Imperial College London, London, United Kingdom
[c] Met Office, Exeter, United Kingdom

ABSTRACT

Object-based radar rainfall nowcasting is a widely used technique for convective storm prediction. Currently, most existing object-based nowcasting methods primarily focus on predicting cell movements, neglecting the temporal evolution of cell properties such as size, shape, and intensity. Incorporating this evolution is critical for improving predictability in convective storms. While previous studies have used three-dimensional (3D) radar observations to capture vertical changes during convective cell formation, these efforts often analyse or reconstruct specific convective events. Integrating 3D radar information into operational object-based radar rainfall nowcasting remains an open challenge. This research addresses this challenge using deep learning (DL) techniques. More specifically, a DL-based prediction model is developed, which uses 2D and 3D cells' properties retrieved from 3D radar reflectivity data at the current time and across the past 15 min to predict the evolution of these properties over the next 15 min. This model could eventually be integrated into existing object-based nowcasting models. A total of 4708 cell lifecycles, extracted from high-resolution (5-min, 1-km, 24 levels at 0.5 km intervals) 3D radar data across the UK, are used to train the model, and a total of 1177 lifecycles are used for testing. The proposed model is shown to predict the evolution of single-core convective cells effectively, including changes in 2D projected geometry and mean 2D and 3D reflectivity. In particular, by incorporating information on the vertical evolution of convective cores, the prediction errors of mean reflectivity (in both 2D and 3D) can be reduced by approximately 50% at 15-min forecast lead time, as compared to a persistence forecast. *Keywords:* radar, tracking, convective cell, nowcasting, 3D, deep learning, lstm.

1. Introduction

Climate change has reportedly altered global precipitation patterns, leading, amongst other things, to the intensification of short-duration rainfall extremes (Trenberth et al., 2003; Lenderink et al., 2017; Liu et al., 2009). This has increased the frequency and severity of flood events worldwide, particularly in urban areas where ongoing urbanisation exacerbates both likelihood and impacts (Guhathakurta et al., 2011; Huong and Pathirana, 2013; Willems, 2013; Miller and Hutchins, 2017; Guerreiro et al., 2018; Tabari, 2020; Fowler et al., 2021a, 2021b).

Despite ongoing investment in structural flood mitigation (e.g. improved drainage –both traditional and sustainable– and flood defences) (Zhou et al., 2019; Ghodsi et al., 2020; Hobbie and Grimm, 2020; Pour et al., 2020), it is virtually impossible –as well as economically and environmentally unsustainable– to eliminate the hazard (Webber et al., 2020; Cristiano et al., 2023). Instead, non-structural measures aimed at enabling optimisation of existing drainage systems and improving pre-

successful implementation of said non-structural measures relies greatly upon good-quality short-term rainfall forecasts, which can be used as input to optimisation, flood forecasting and warning systems (Hapuarachchi et al., 2011; Tingsanchali, 2012).

There are two main sources of short-term rainfall forecasts for use in such operational systems: numerical weather prediction models (NWPs) and radar-based rainfall nowcasting (Germann and Zawadzki, 2002; Bowler et al., 2006; Schell... 2014; Thorndahl et al., 2016; Casagrande et al., 2017). De... models in the last 40 years (Bauer et al., 2... eir accuracy at short lead times (0-5 h) is stil... n up' period –a period required for a comp... stabilise. This results in relatively low pre... hours (Fischer et al., 2005).

To address this limitatio... ed nowcasting models were developed (Dix... on et al., 1998; Germann and Zawadzki, 200... ni et al., 2012;

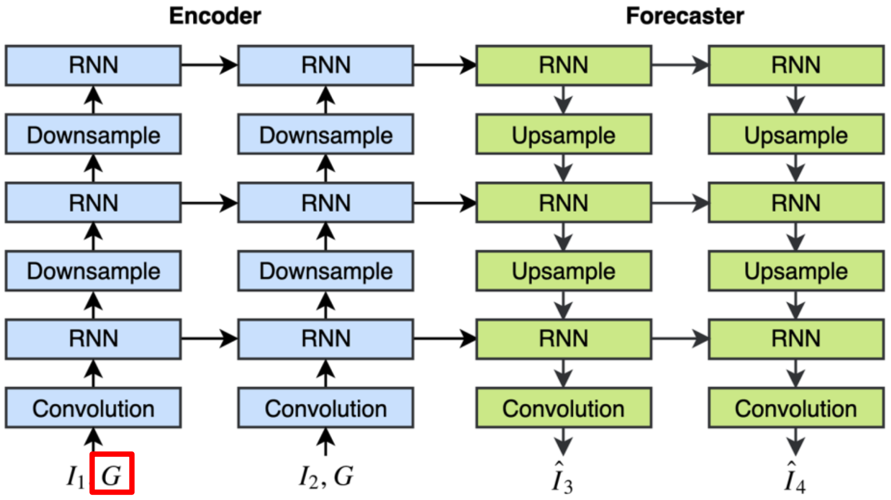
doi:10.1016/j.atmosres.2024.107380

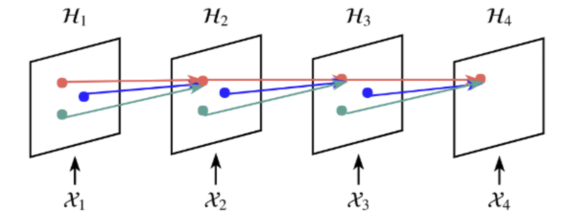
Category of recent AI-based field nowcasting methods

- Explicit separation of advection and evolution prediction (TrajGRU, NowcastNet)
- Focusing on spatial-temporal feature extraction (DGMR, MetNet)
- Diffusion models

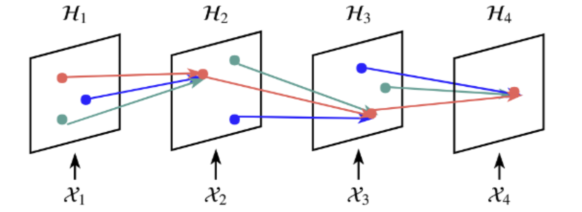
TrajGRU for precipitation nowcasting

- A classic encoding-forecasting framework
- Motion 'flow' is incorporated

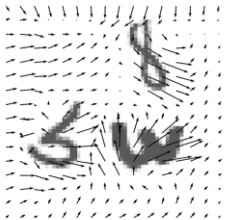




(a) For convolutional RNN, the recurrent connections are fixed over time.



(b) For trajectory RNN, the recurrent connections are dynamically determined.



TrajGRU is flexible and better at capturing complex spatiotemporal patterns, such as rotation and scaling

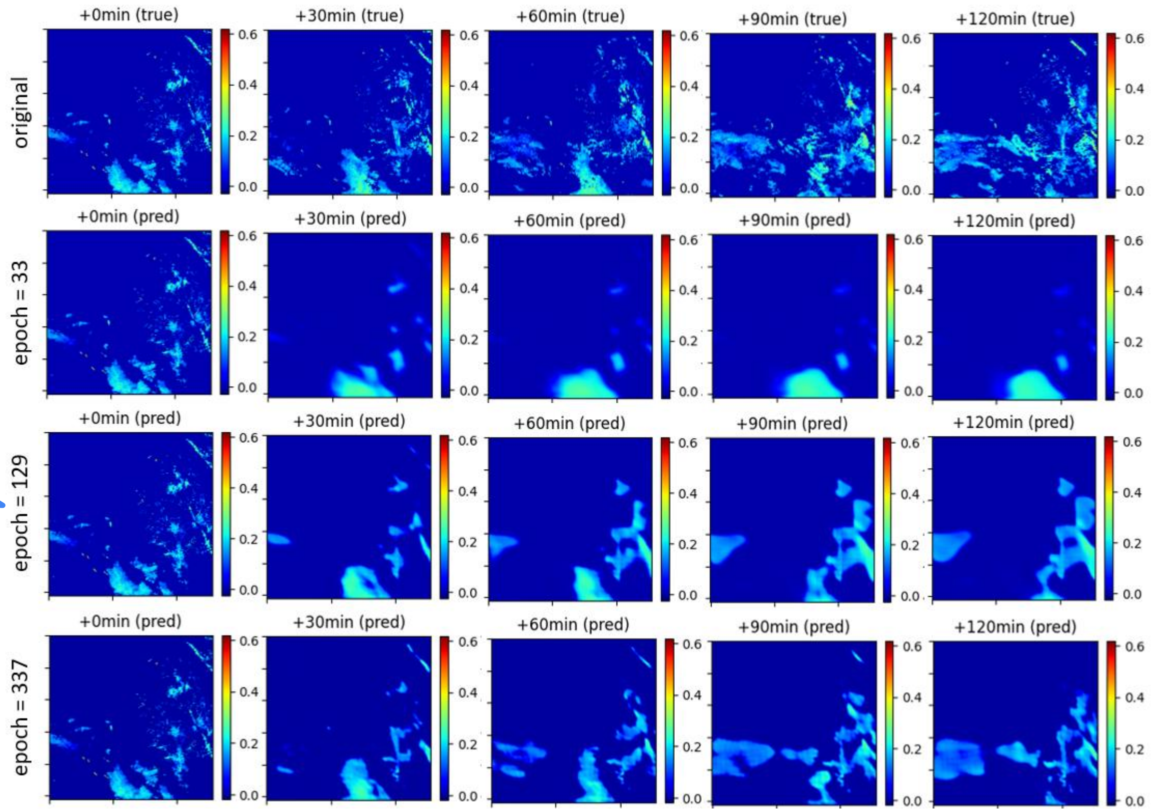
https://doi.org/10.48550/arXiv.1706.03458

A case of training TrajGRU with CWA 10-min radar data

In early training iterations, the rainfall field only 'evolves' – it doesn't move

Advection starts to be learnt at later iterations

AI appears to learn evolution first, then advection.



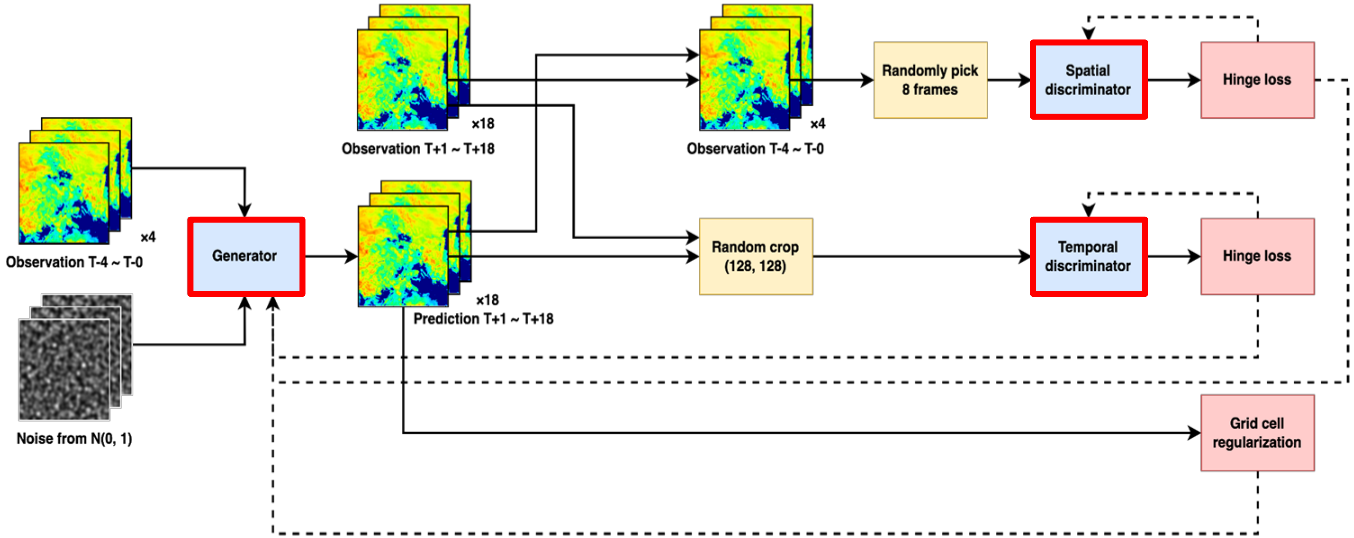
More detailed field reproduction

DGMR: A rainfall nowcasting model trained with the GAN technique

DGMR includes three main models, which are trained simultaneously via an adversarial process.

These models are:

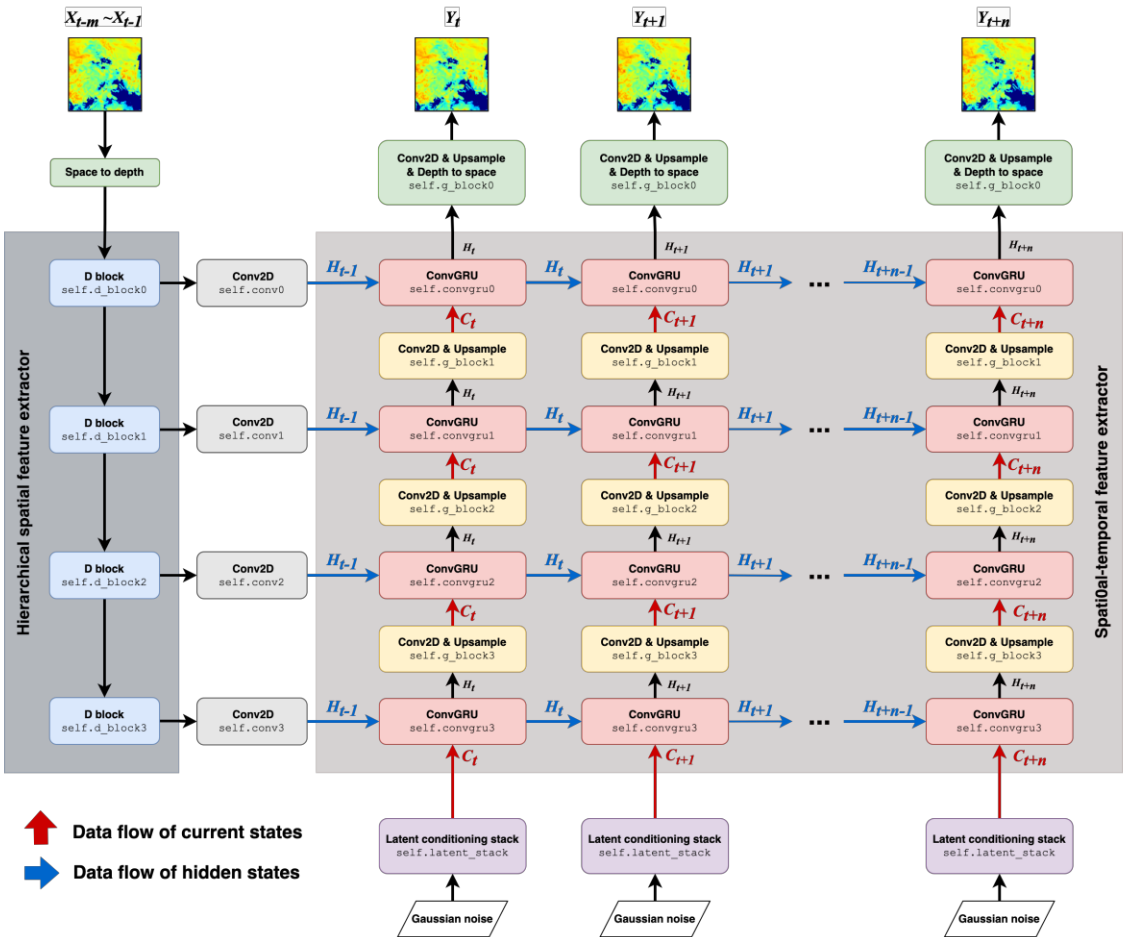
- **A Generator** that produces rainfall nowcasts.
- **Two Discriminators** that discriminate if the generated nowcasts are 'similar enough' to the ground truth in terms of their spatial and temporal features, respectively.



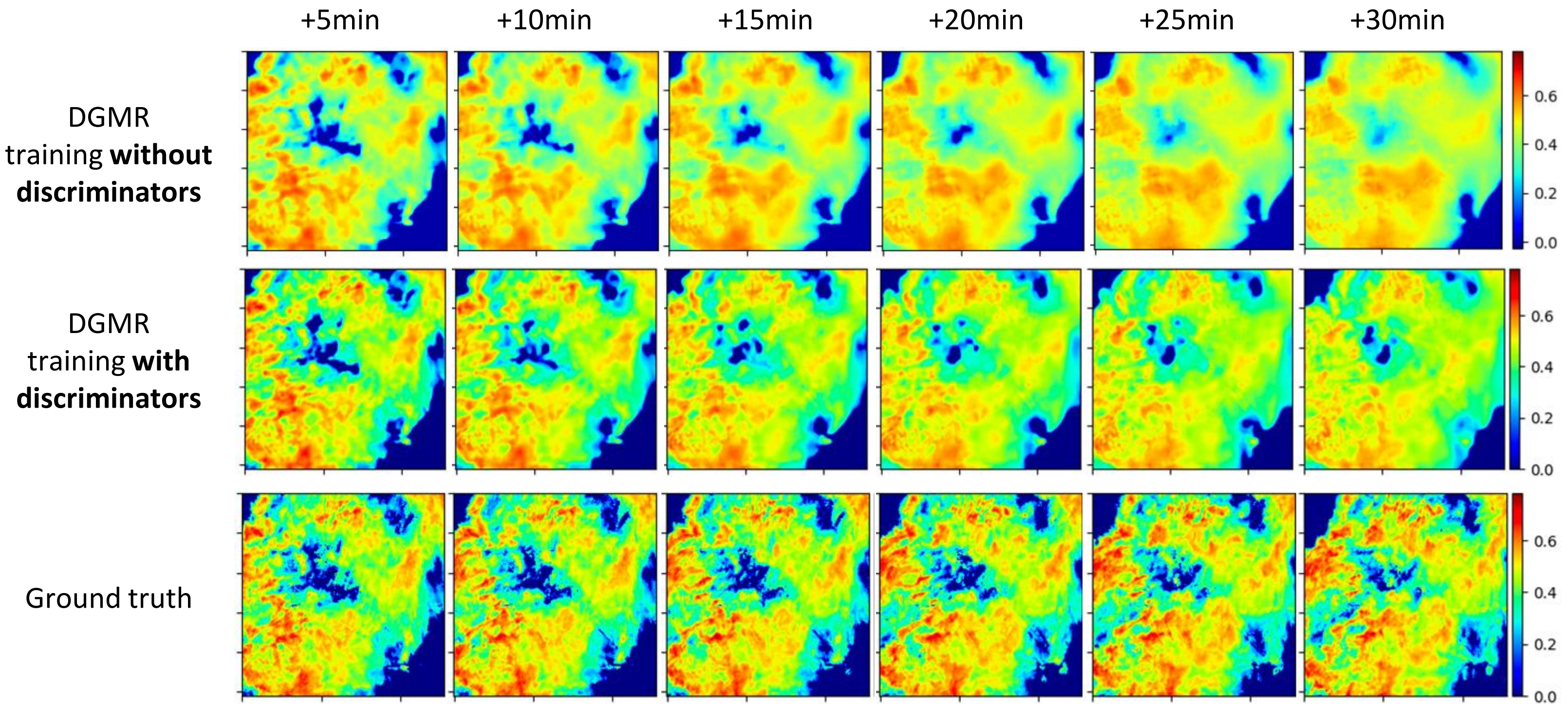
Our DGMR replica on Github!

https://doi.org/10.1038/s41586-021-03854-z

DGMR generator
=
Spatial feature extractor
+
Temporal feature extractor

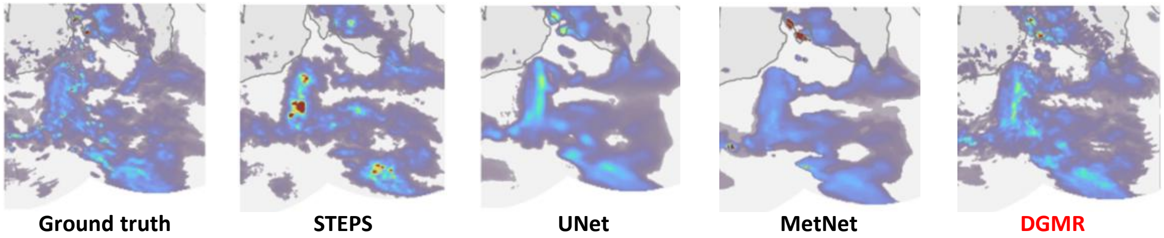


More details can be preserved in the predicted fields while GAN is used.



DeepMind’s introduction of GANs to rainfall nowcasting was a catalyst.

Since then, **GANs have been widely adopted for AI-based nowcasting**, valued for their ability to produce sharp, realistic forecasts.



Climate modelling

Can we teach AI to say ‘I don’t know’?

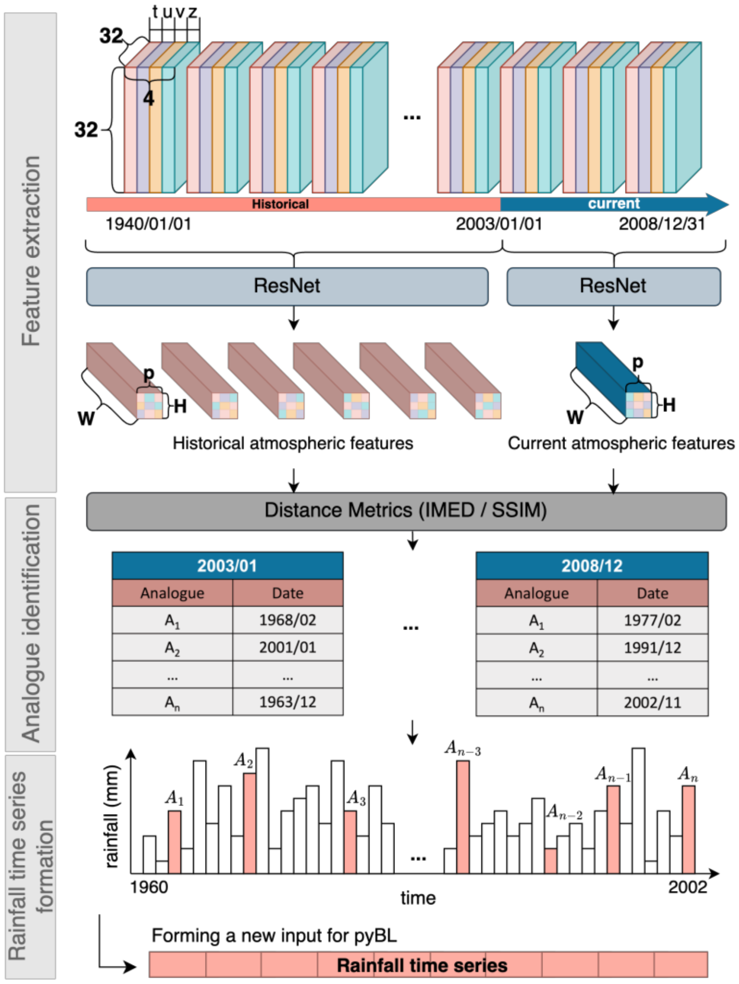
Climate risk modelling:

Understanding the key challenges and opportunities in creating climate transition pathways

Gap to be filled

	CATASTROPHE MODELS	CLIMATE MODELS	TRANSITION RISK MODELS
USE	To measure the impact or financial loss from physical risks and catastrophic events.	To understand the evolution of the system over different time scales (past, present and future).	To inform economic risks arising from the transition to a zero carbon economy.
INPUTS	Historical statistical distributions that describe physical hazards; do not explicitly consider future climate considerations.	Physical models that represent the Earth system and help to understand the evolution of the system over different time scales (past, present and future); do not measure the financial or economic impact of climate events.	Incorporates two different types of information: climate data that don't measure the financial and economic impacts of climate events, and economic data that leverage historical patterns to predict a future that will look different due to intensifying climate change impacts.
BENEFITS	Provides probabilities of extreme event occurrence assuming current climate conditions.	Can produce realistic future climate conditions.	Portrays plausible scenarios or pathways to transition the economy from a predominantly fossil fuel energy perspective to one incorporating new types of fuel sources.
LIMITATIONS	Only well developed for geographic areas and hazards where a large percentage of the population is insured against that hazard. They are less developed in geographies with a low amount of insurance coverage that could be susceptible to climate change.	Struggles to predict many of the extreme events that most impact the insurance industry (such as hurricanes and wildfires). These events occur on spatial scales that are too small to be "seen" in most climate models.	Risk of misinterpreting the output of the models when making portfolio-level decisions due to the highly simplified and backward looking representation of physical hazard impacts on the economy.

A climate-dependent rainfall time series modelling



https://www.researchsquare.com/article/rs-5382842/latest

analogue 1 of 2 noun

an·a·logue ˈa-nə-ˌlȯg -ˌläg

variants *or* **analog**

Synonyms of *analogue* >

1 **:** something that is similar or comparable to something else either in general or in some specific detail **:** something that is analogous to something else

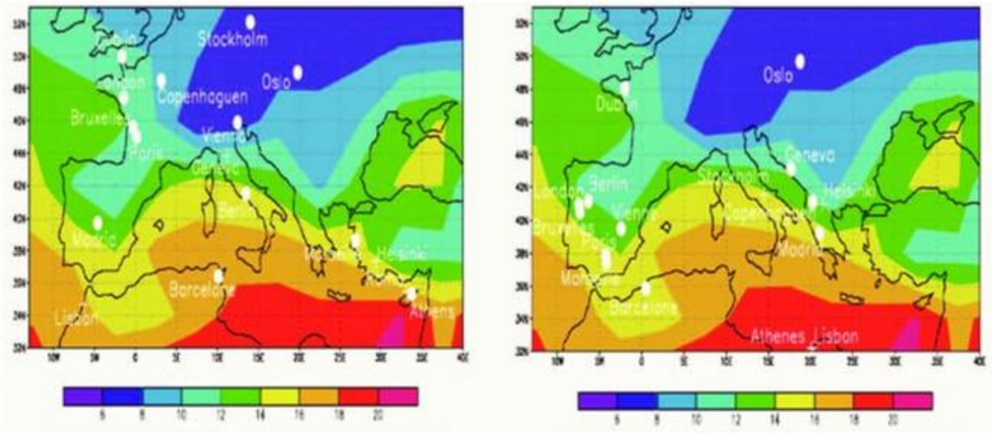
historical *analogues* to the current situation

What is Climate/weather analogue method?

- A method of examining today's forecast scenario and finding a day in the past when the weather condition looks very similar.

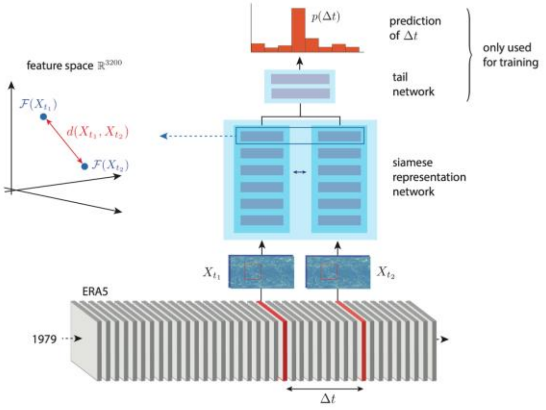
- Traditionally, the level of similarity was quantified by **computing the Euclidean Distance between key weather variables determined by the experts.**

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

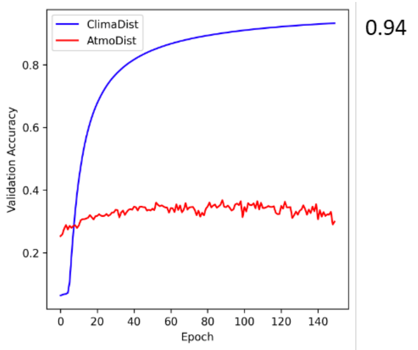


ClimaDist

- A deep-learning model (CNN) is used to capture weather dynamics, without the need for labelling.
- **Model outputs can be used to create a 'feature' to chracterise weather across a 30-day period, at 6-hour resolution, accounting for spatial-temporal variability and inter-dependence**.
- **The resulting features are used for analogue identification**.
- Ultimately, the analogue model can be applied e.g. for dynamic rainfall modelling -> finding 30-day rainfall patterns from a given set of climate variables.



(AtmoDist - Hoffmann and Lessig,2022)

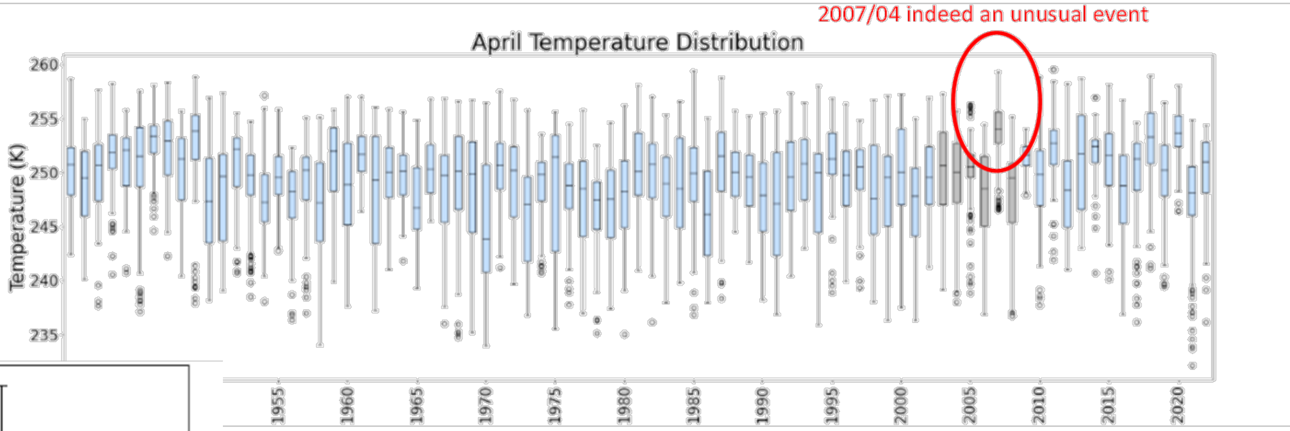


Dataset

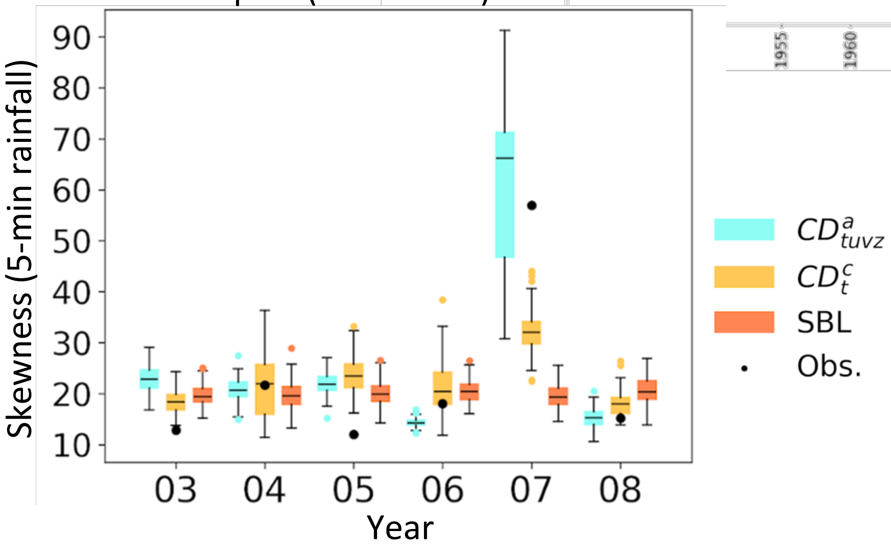
- ERA5 1940 – 2009
- Climate variates: Temperature, U and V-component of wind, geopotential height

Example Application – incorporating climate dynamics into rainfall modelling

Apr'2007 was an abnormally hot and dry month



All Aprils (2003-2008)



- Incorporating climate dynamics successfully captures extreme climate and rainfall behaviour!
- Stationary methods, without considering climate dynamics, cannot predict this type of extreme events – they are treated as outliers

Beyond Accuracy:
Teaching Models to Say "I Don't Know"

https://www.ettoday.net/dalemon/post/6299

The Overconfidence Trap

❌

This is a dog
Confidently Incorrect

✅

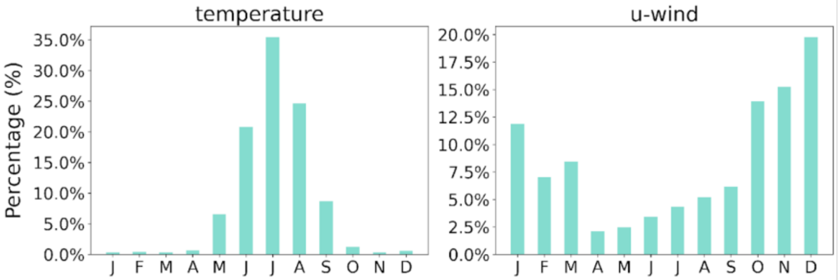
I think this is a dog, but I'm not sure

Solution: Introducing uncertainty assessment

Explainable AI (XAI) to understand ClimaDist

Temporal view (LRP)

- Tells us the decision flow of NN, accounting for seasonality

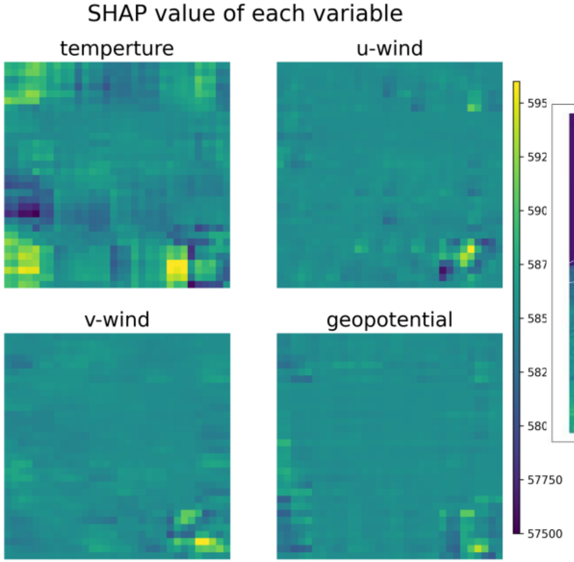


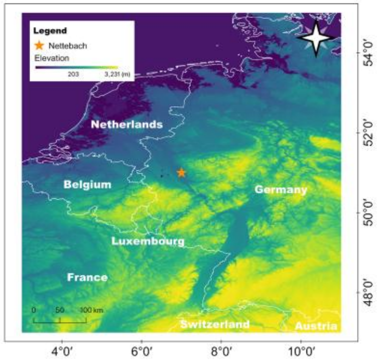
Temperature dominates in summer and is prioritised by the model

U–wind dominates in winter

Spatial view (SHAP)

- SHAP value tells us the importance of each variable at different locations





Evidential Deep Learning: Uncertainty Estimation

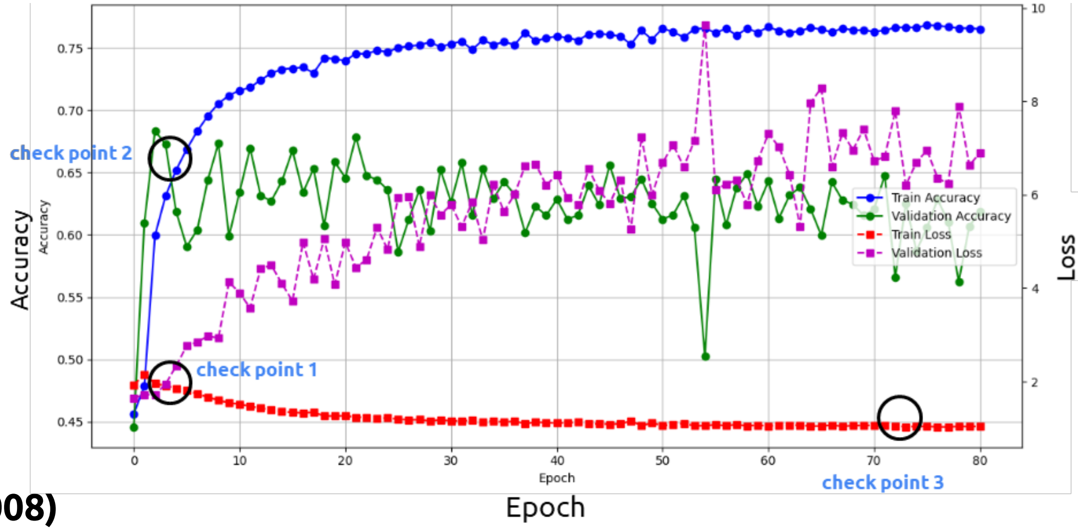
- A **Bayesian-inspired** ML approach that does not directly predict **class** probabilities but instead **predicts a distribution over these probabilities.**
- By modelling the output as a **Dirichlet distribution**, we can separately quantify two types of uncertainty
    - Dempster–Shafer Theory of Evidence: **DST** u
    - Law of Total Variance and Entropy (LoTV):
        - **Aleatoric (data)** uncertainty
        - **Epistemic (model)** uncertainty

Exercise: Training ClimaDist model with EDL

Three checkpoints (training targets) are saved from the same model architecture during training to serve as the basis for subsequent experiments

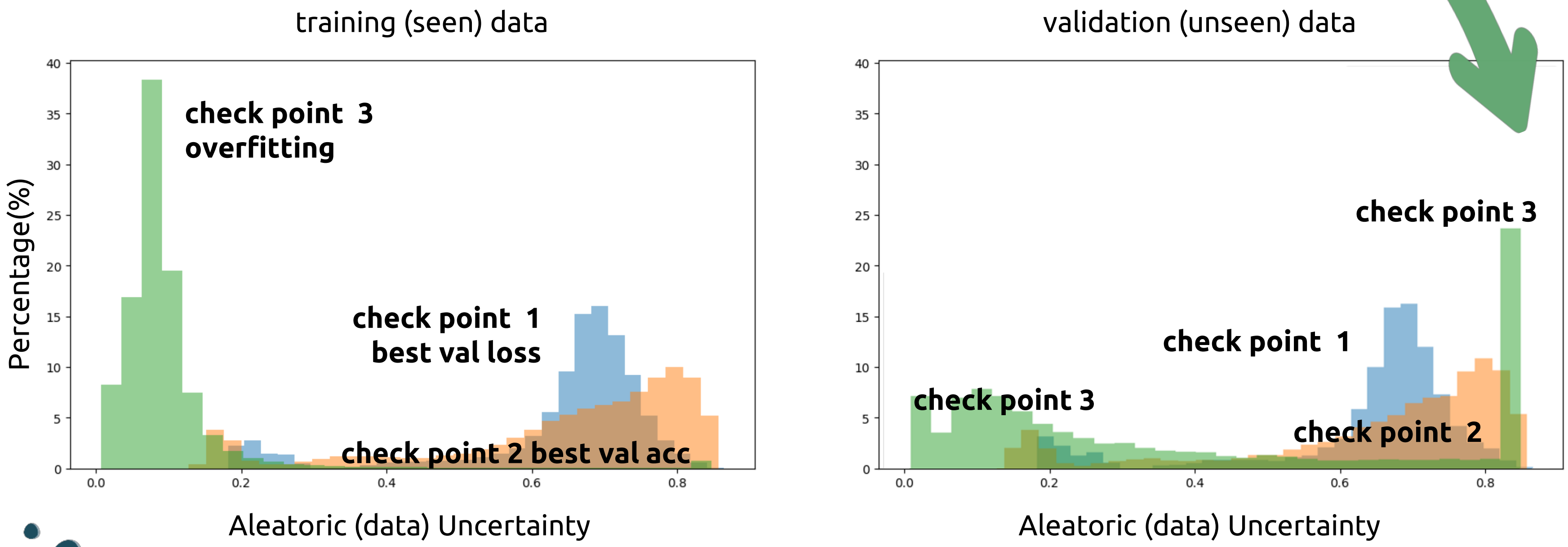
- ❑ **check point 1:**
  best validation loss
- ❑ **check point 2:**
  best validation accuracy
- ❑ **check point 3:**
  best training performance (overfitting)

➔ **Training: Seen data (1960 - 2003)**
➔ **validation: Unseen data (2004 - 2008)**



Distribution of Aleatoric (data) Uncertainty

- Overfitted models exhibit high data uncertainty on unseen data
- **Our EDL model is telling us ‘I haven’t seen this data’!**



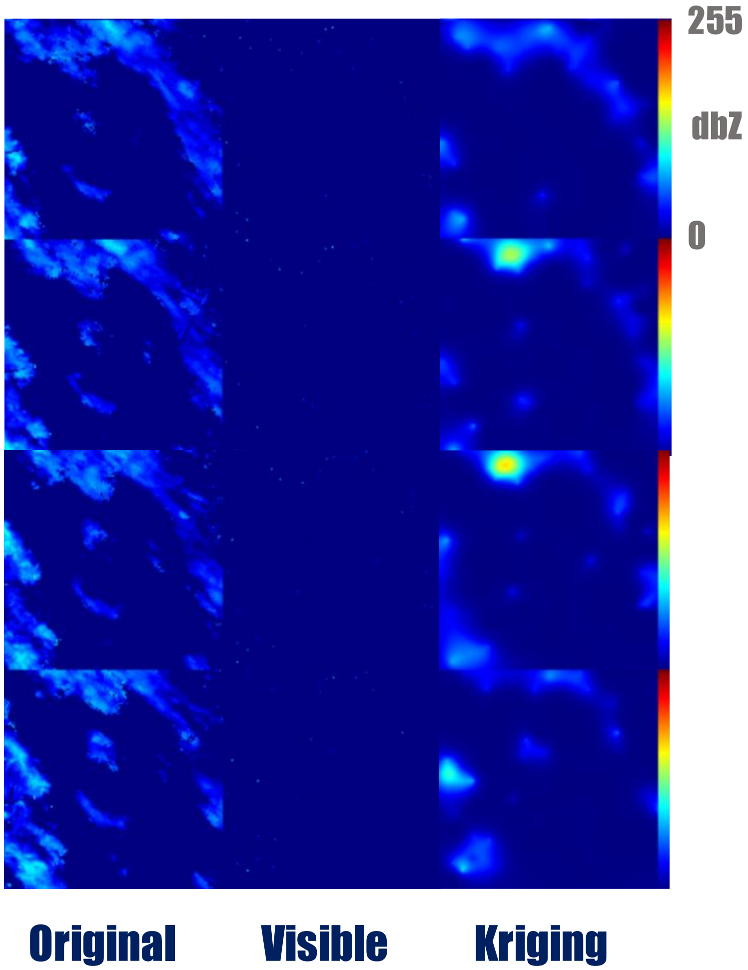
Spatial-temporal reconstruction of rainfall

AI’s power of learning spatial and temporal features

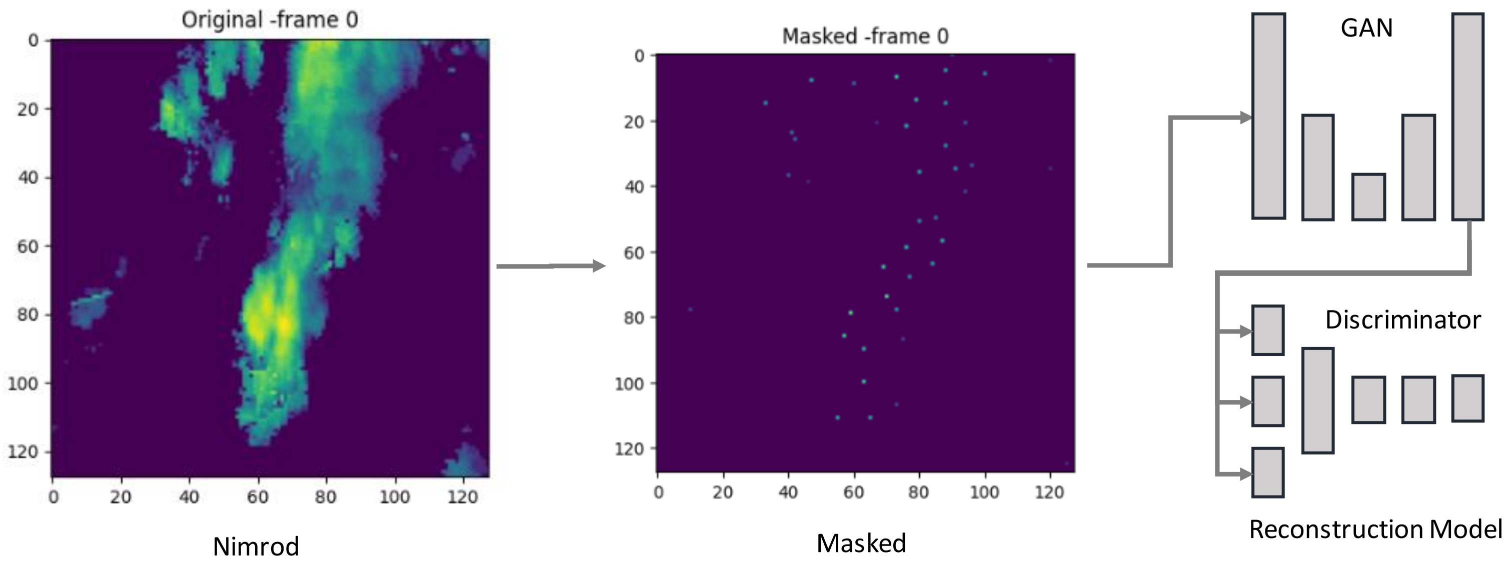
Kriging is a powerful tool to generate fields from point data, but...

- It is sensitive to the density and location of known data points
- The generated fields are of relatively low temporal consistency
- The generated fields are generally smooth, lacking detail

Can AI help us overcome these limitations?



Problem formulation in an AI style



Proposed model architecture

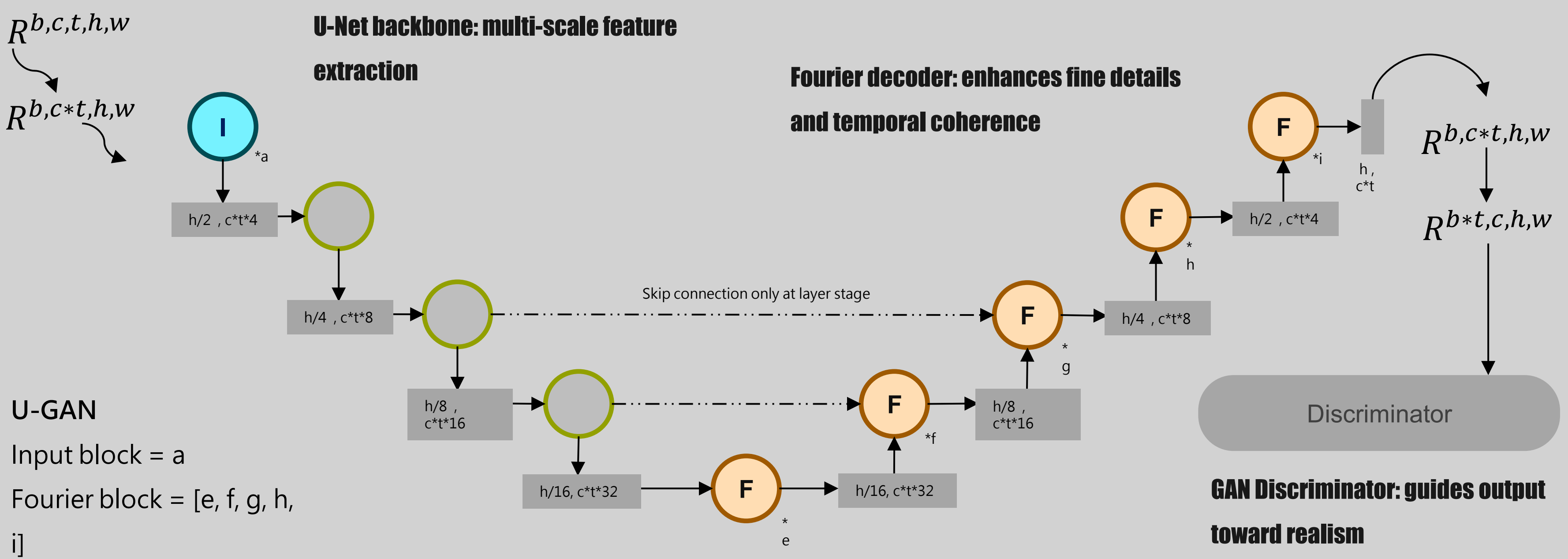
All Conv2d is 3*3

is Conv2d

is DN or UP

F = FD

I = IE



U-GAN

Input block = a

Fourier block = [e, f, g, h, i]

Key model components

I Input Block
: Interpolate coarse-scale field

G_{in}

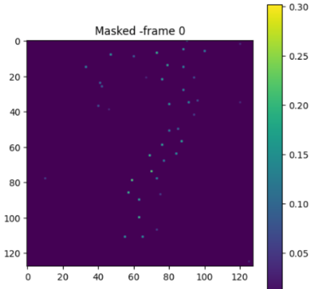
Temporal
- cat(s[1] ,s ,s [-1])
- x =(s[t] + s[t+1])/2

M-IDW

IDW , kernel [19, 5] , threshold [2,1]
For point with value in surrounding region 19x19 with at least 2 value then use IDW. Otherwise 5x5. p=2

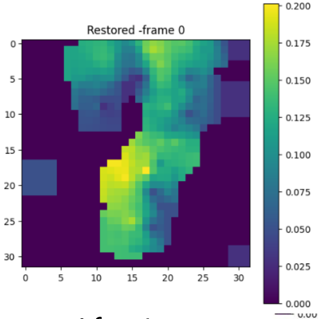
$$\bar{Z}(P) = \frac{\sum_{i=1}^{n} w_i Z(P_i)}{\sum_{i=1}^{n} w_i} \qquad w_i = \frac{1}{d^p}$$

1



Masked Frame

2



After Input Block

3

F Fourier Decoder
: Generate spatial detail and maintain temporal coherence

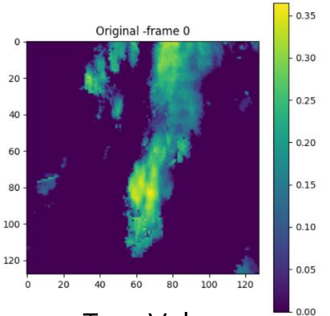
L_{in}

- Modulation
- Batchnorm
- Conv 3x3
- ReLU
- Conv 3x3
- Residual

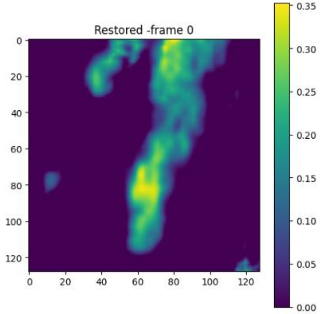
- Calculate_moments
- Mean,Var in 3x3 5x5
- Layernorm2d
- Conv 1x1
- sigmiod(m)*noise

- Fourier Transform
- Conv 1x1
- Fourier Transform
- Conv 3x3
- Fourier Inverse

4



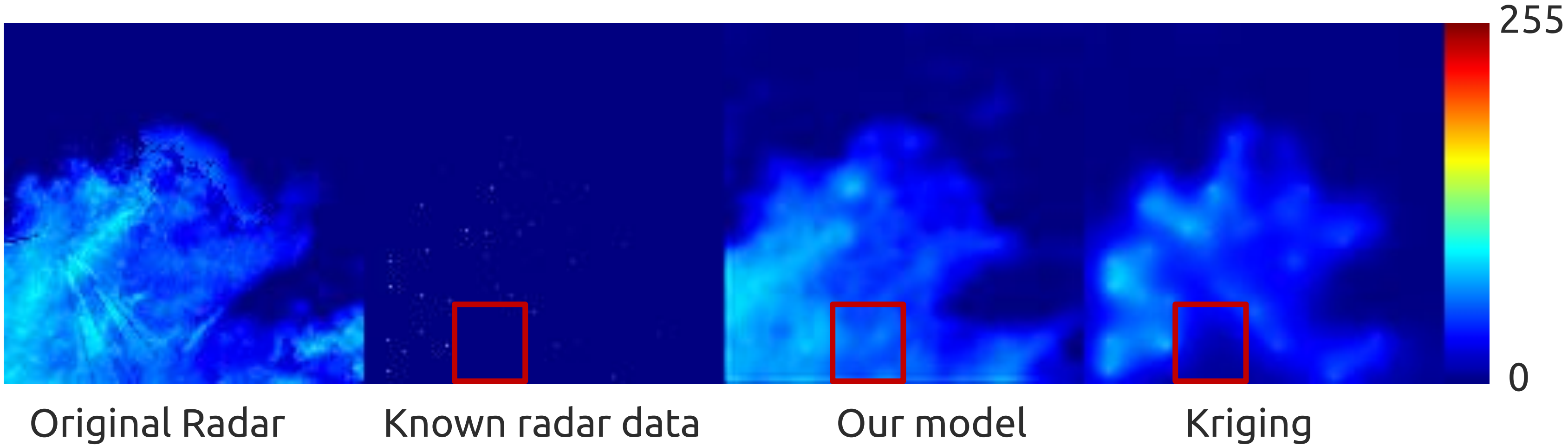
True Value



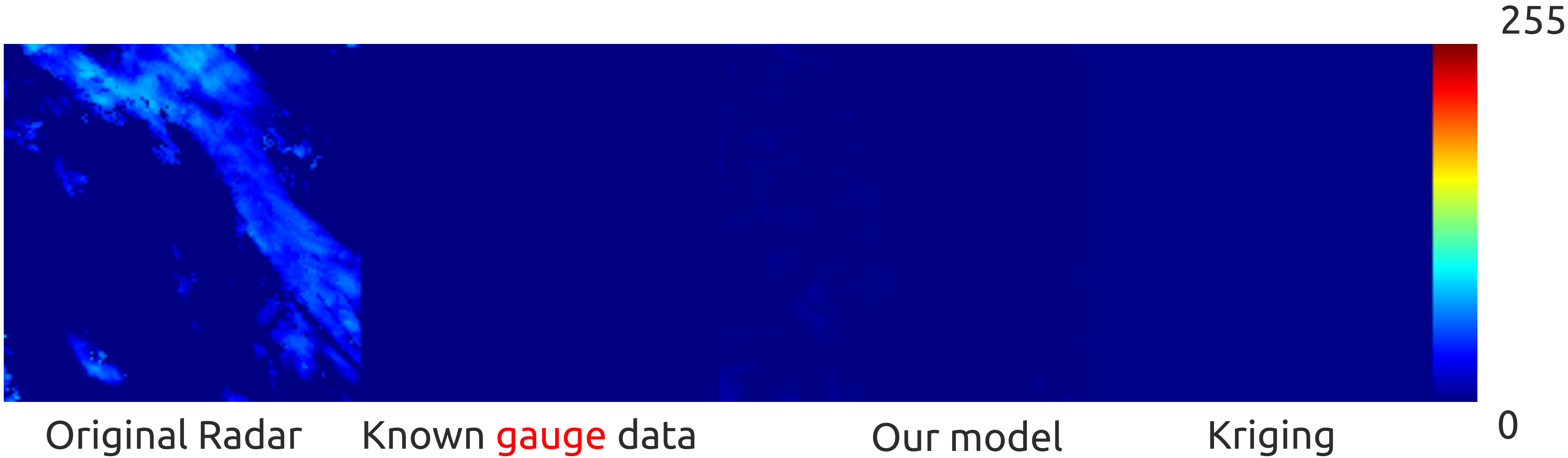
Model Output

Yes, with <1% known data, AI can well reconstruct the rainfall fields, where:

- More details are preserved
- Observed temporal consistency is maintained
- Less sensitive to gauge density -> see performance in low density areas



Extension: How if we replace the known radar data to real gauge data?



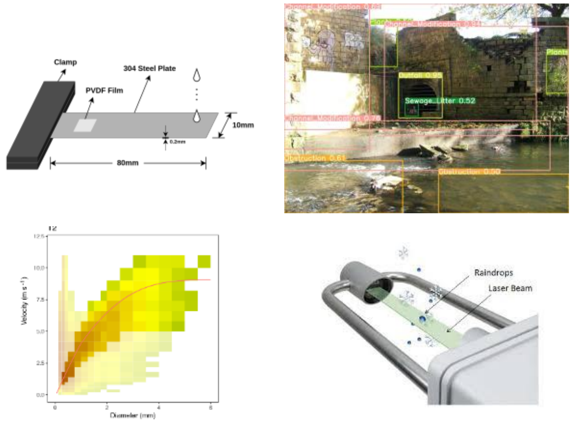
With real gauge data, kriging becomes even less realistic and displays lower temporal coherence

Concluding remarks

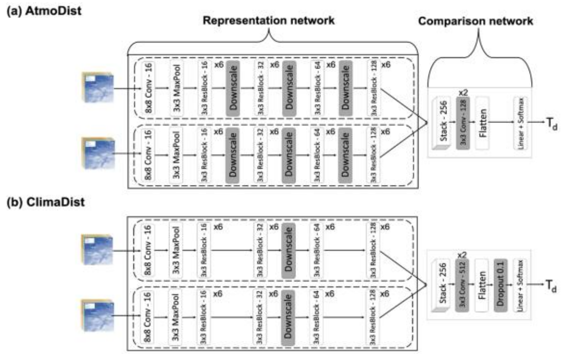
- Understanding physical processes is key to best use AI.
- We should take advantage of what (current) AI models are good at – in our applications:
  - Capturing spatial and temporal features
  - Faster in learning rainfall evolution than advection
- While pursuing high model accuracy, it is critical to assess model's behaviour and uncertainty
  - Explainable AI can help confirm if the model behaviour makes sense.
  - EDL helps quantify uncertainty and teaches AI models to indicate 'I don't know'.

Computational Hydrometeorology Lab

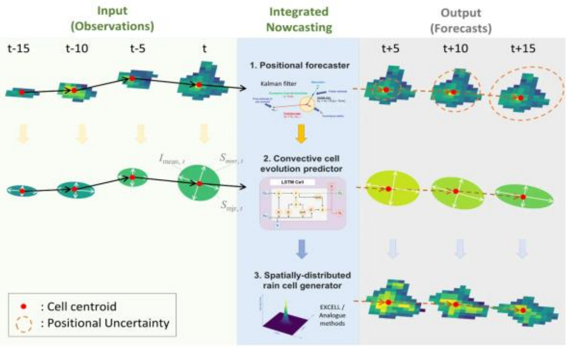
Unconventional environmental modelling



Modelling climate change impacts to rainfall at local scales



Modelling and nowcasting of convective storms



Learn more about CompHydroMetLab!

Visit our Github

Li-Pen Wang (lpwang@ntu.edu.tw)
https://wangup.caece.net/