

Towards a Data-driven Nowcasting of Severe Weather Based on Geostationary Data

Çağlar Küçük*, Apostolos Giannakos, Stefan Schneider, Alexander Jann

caglar.kucuk@zamg.ac.at
Zentralanstalt für Meteorologie und Geodynamik

EMS Annual Meeting – 6 Sept 2022

Outline

Introduction

Data & Methods
- Dataset
- Methodology

Results

Summary & Outlook

Outline

Introduction

Data & Methods
- Dataset
- Methodology

Results

Summary & Outlook

Motivation

- Nowcasting severe weather
  - Disaster management against extreme events like flash flood, storm, hail
  - Daily events with socio-economic impact
  - Weather-dependent decision-making

Motivation

- Nowcasting severe weather
  - Disaster management against extreme events like flash flood, storm, hail
  - Daily events with socio-economic impact
  - Weather-dependent decision-making
- Data-driven!
  - Stunning progress in machine learning
  - No shortage in data

Motivation

Meteosat Third Generation (MTG)

- High spatio-temporal resolution, up to:
  - 0.5 km in space
  - 2.5 minutes in time
- Improve resolution, accuracy and lead time of weather predictions
- Larger capacity to resolve convection in early stages

Motivation

Meteosat Third Generation (MTG)

- High spatio-temporal resolution, up to:
  - 0.5 km in space
  - 2.5 minutes in time
- Improve resolution, accuracy and lead time of weather predictions
- Larger capacity to resolve convection in early stages

→ Opportunities for data-driven nowcasting!

- EUMETSAT Research Fellowships

Outline

Introduction

Data & Methods
- Dataset
- Methodology

Results

Summary & Outlook

Storm Event ImageRy (SEVIR) Dataset

- Combines satellite (GOES) and ground-based radar (NEXRAD) data
- Similar sensors in GOES and MTG
- Vertically Integrated Liquid (VIL): Diagnostic for severe weather

→ 20 % overlap with storm events manually labelled by NOAA

Storm Event ImageRy (SEVIR) Dataset

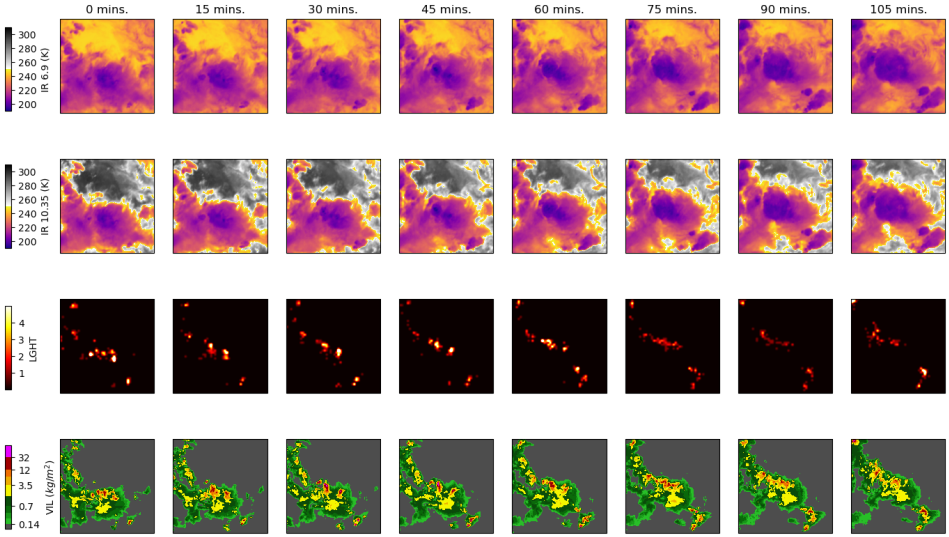
- Combines satellite (GOES) and ground-based radar (NEXRAD) data
- Similar sensors in GOES and MTG
- Vertically Integrated Liquid (VIL): Diagnostic for severe weather

$\rightarrow$ 20 % overlap with storm events manually labelled by NOAA

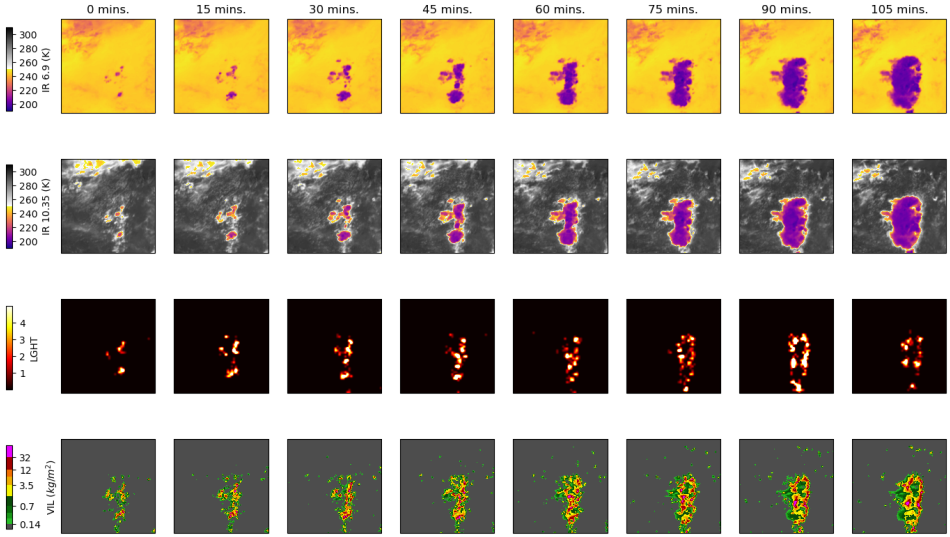
- 10k+ image sequences from 2018-2019 over Continental US:
  - 384x384 km
  - 4 hours
  - 5 minutes temp. res. (49 frames)

Variable	Spat. Res.
vis: GOES Visible channel	0.5 km
ir069: GOES infrared channel $(6.9\mu m)$	2 km
ir107: GOES infrared channel $(10.35\mu m)$	2 km
lght: GLM lightning flashes	8 km
vil: NEXRAD radar mosaic of VIL	1 km

SEVIR Dataset



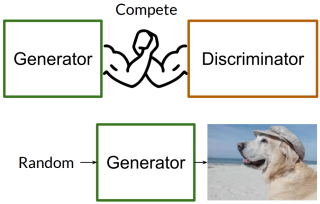
SEVIR Dataset



Generative Adversarial Networks (GAN)

→ GAN: Train an unsupervised method from competition of two models
  - Generator: Generates output similar to target
  - Discriminator: Judges whether output is real or fake
- Used in computer vision, text generation, downscaling atmospheric data, and also in Geosciences
- Very powerful, though difficult to train

→ Conditional GANs
  - Condition the output based on input
  - "Image-to-Image translation"
  - No temporal information...

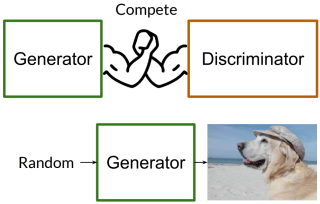


Taken from `https://www.deeplearning.ai/`

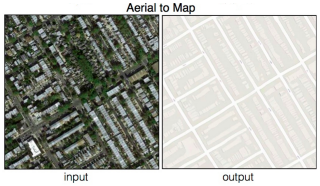
Generative Adversarial Networks (GAN)

→ GAN: Train an unsupervised method from competition of two models
  - Generator: Generates output similar to target
  - Discriminator: Judges whether output is real or fake
- Used in computer vision, text generation, downscaling atmospheric data, and also in Geosciences
- Very powerful, though difficult to train

→ Conditional GANs
  - Condition the output based on input
  - "Image-to-Image translation"
  - No temporal information...



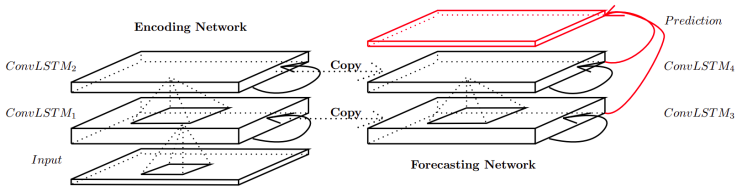
Taken from `https://www.deeplearning.ai/`



Taken from Isola et al., 2018

Convolutional LSTM

- Convolutional structures fed into LSTM
- Encoder-Forecaster structure

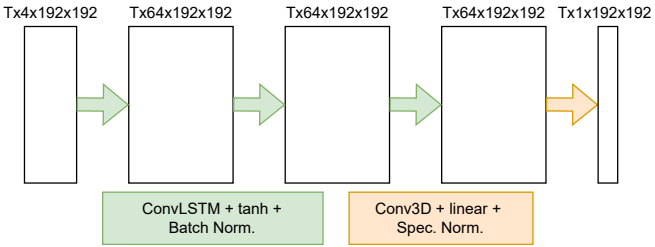


Taken from Shi et al., 2015

→ Use in the Generator of GAN

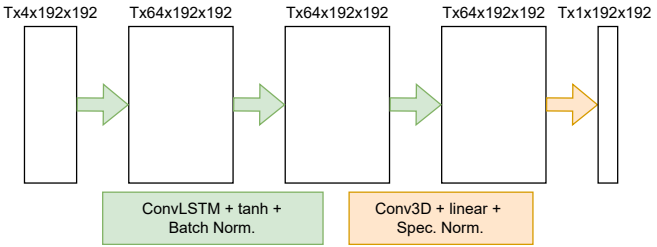
Model architecture

- Generator: 3 ConvLSTM + Conv3D
- Discriminator: *Time distributed* PatchGAN



Model architecture

- Generator: 3 ConvLSTM + Conv3D
- Discriminator: *Time distributed* PatchGAN



Training details

- Inputs: *T frames* of ir069, ir107, lght, vil (Past)
- Output: *T frames* of vil (Future)

- Loss function:
  - $L_{GAN} = L_{gen} + \lambda \times L1 + L_{disc}$ where $\lambda = 50$
  - L_{gen} and L_{disc}: Binary cross-entropy
  - $L1$ from errors b/w target & output via MAE, scaled by λ

Training details

- Inputs: *T frames* of ir069, ir107, lght, vil (Past)
- Output: *T frames* of vil (Future)

- Loss function:
  - $L_{GAN} = L_{gen} + \lambda \times L1 + L_{disc}$ where $\lambda = 50$
  - L_{gen} and L_{disc}: Binary cross-entropy
  - $L1$ from errors b/w target & output via MAE, scaled by λ
- Adam optimizer with $\alpha = 5 \times 10^{-5}$ and $\beta_1 = 0.5$

Training details

- Inputs: *T frames* of ir069, ir107, lght, vil (Past)
- Output: *T frames* of vil (Future)

- Loss function:
  - $L_{GAN} = L_{gen} + \lambda \times L1 + L_{disc}$ where $\lambda = 50$
  - L_{gen} and L_{disc}: Binary cross-entropy
  - $L1$ from errors b/w target & output via MAE, scaled by λ
- Adam optimizer with $\alpha = 5 \times 10^{-5}$ and $\beta_1 = 0.5$
- Update generator *twice* per training step

Training details

- Inputs: *T frames* of ir069, ir107, lght, vil (Past)
- Output: *T frames* of vil (Future)

- Loss function:
  - $L_{GAN} = L_{gen} + \lambda \times L1 + L_{disc}$ where $\lambda = 50$
  - L_{gen} and L_{disc}: Binary cross-entropy
  - $L1$ from errors b/w target & output via MAE, scaled by λ
- Adam optimizer with $\alpha = 5 \times 10^{-5}$ and $\beta_1 = 0.5$
- Update generator *twice* per training step
- 1 epoch (with a 48 GB vGPU in the European Weather Cloud):
  - 1-hour lead time: 135 mins
  - 2-hour lead time: 235 mins

Outline

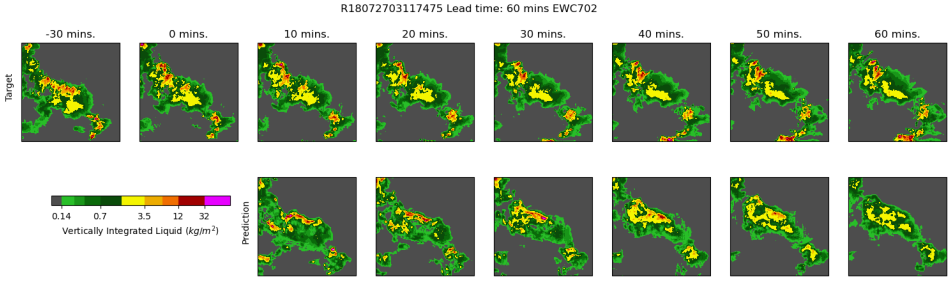
Introduction

Data & Methods
- Dataset
- Methodology

Results

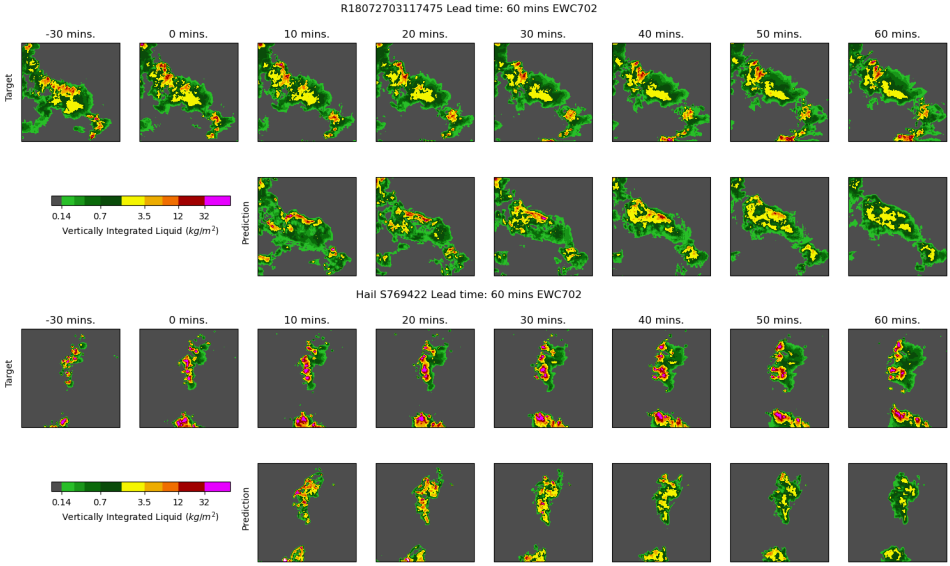
Summary & Outlook

1-hour lead time

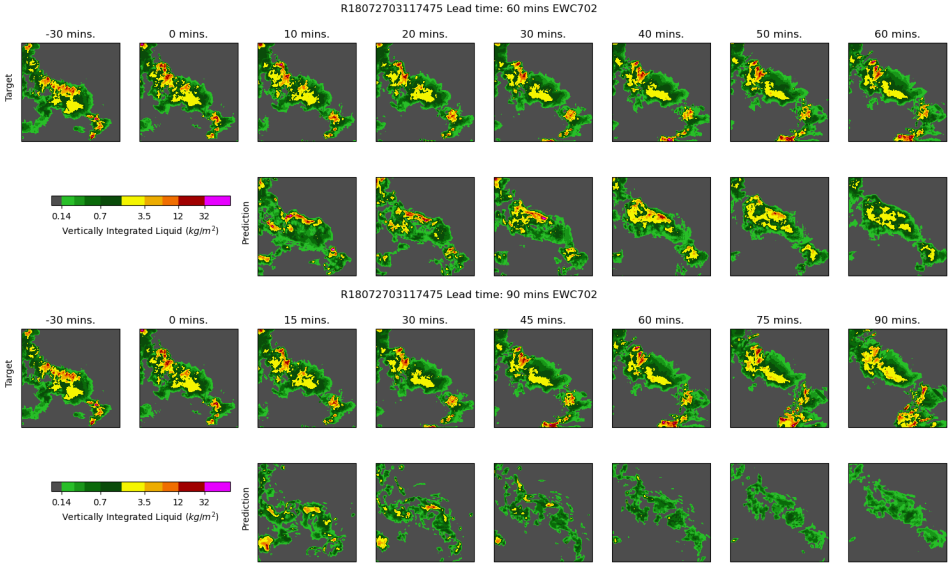


- Location is captured
- Extreme values are not
- Predictions smoothen out with time

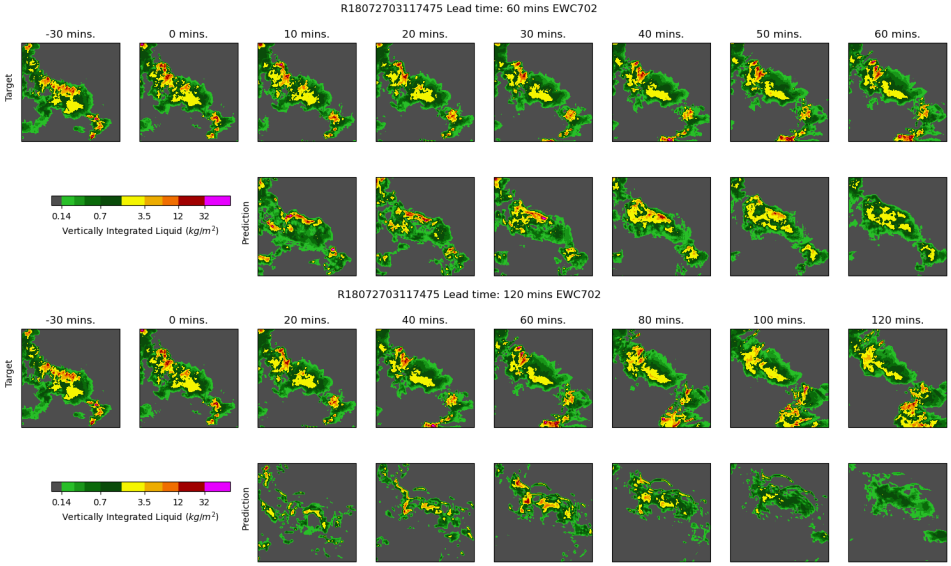
1-hour lead time



1- vs 1.5-hour lead time



1- vs 2-hour lead time



Evaluation

- Metric to use..
  - Critical Score Index, Heidke Skill Score, F1, ...
  - Fractional Skill Score
  - Structure–Amplitude-Location (SAL) score

Evaluation

- Metric to use..
  - Critical Score Index, Heidke Skill Score, F1, ...
  - Fractional Skill Score
  - Structure–Amplitude-Location (SAL) score

Threshold	Lead time	CSI	HSS	F1
0.14	60 mins.	0.26	0.27	0.42
0.27	60 mins.	0.18	0.19	0.31

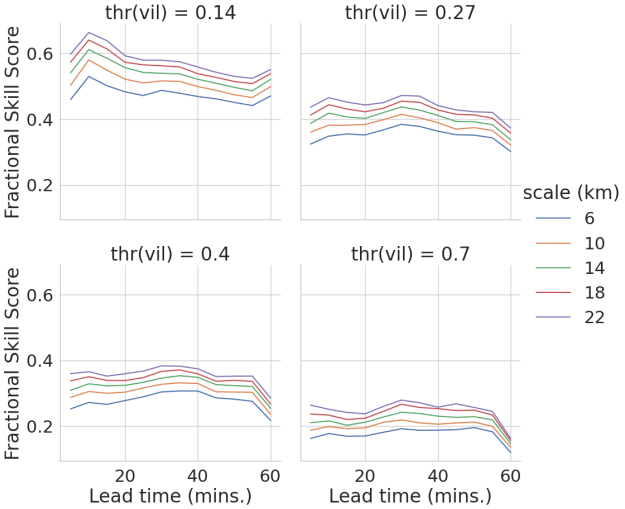
Evaluation

- Metric to use..
  - Critical Score Index, Heidke Skill Score, F1, ...
  - Fractional Skill Score
  - Structure–Amplitude-Location (SAL) score

Threshold	Lead time	CSI	HSS	F1
0.14	60 mins.	0.26	0.27	0.42
0.14	90 mins.	0.14	0.12	0.24
0.14	120 mins.	0.08	0.04	0.15
0.27	60 mins.	0.18	0.19	0.31
0.27	90 mins.	0.08	0.07	0.15
0.27	120 mins.	0.05	0.02	0.09

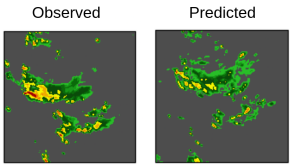
Fractional Skill Score

- Discrete, but spatially aware
- FSS $\sim$ 0.5 for 10 km scale, low threshold



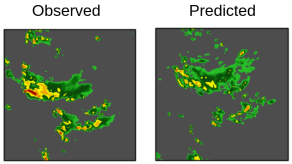
Structure-Amplitude-Location (SAL) Scores

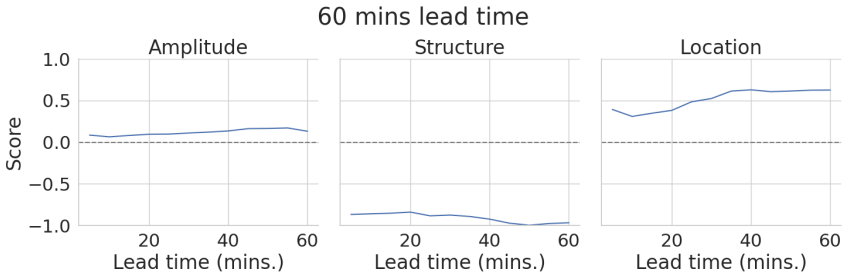
- Evaluates continuous variables
- Ideal forecast has 0 score
- Amplitude > 0: Overestimation in domain average
- Structure < 0: Pred. objects too small or peaked
- Location: Distances b/w center of mass per domain and per object



Structure-Amplitude-Location (SAL) Scores

- Evaluates continuous variables
- Ideal forecast has 0 score
- Amplitude > 0: Overestimation in domain average
- Structure < 0: Pred. objects too small or peaked
- Location: Distances b/w center of mass per domain and per object





Outline

Introduction

Data & Methods
- Dataset
- Methodology

Results

Summary & Outlook

Summary & Outlook

- Stable prediction performance over time
- Location & shape of events are captured
- Amplitude is not captured in large VIL values
- Room for improvement..

Summary & Outlook

- Stable prediction performance over time
- Location & shape of events are captured
- Amplitude is not captured in large VIL values
- Room for improvement..

- Modify loss function(s)
  - Make it more nowcasting relevant
  - Add skill scores as loss penalty
- Modify model architecture
  - Discrete values as model output
  - Autoregressive generator

- Thank you for your attention!
- Questions, comments..

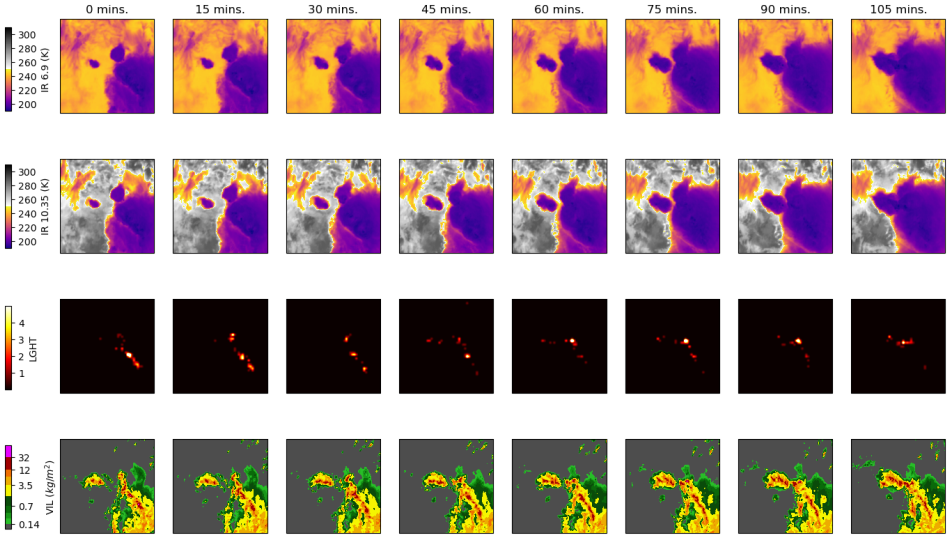
→ caglar.kucuk@zamg.ac.at

ÇK was supported in this work by the EUMETSAT within the Research Fellowship project "FUSEDCAST: a fused data approach to the nowcasting of severe weather"

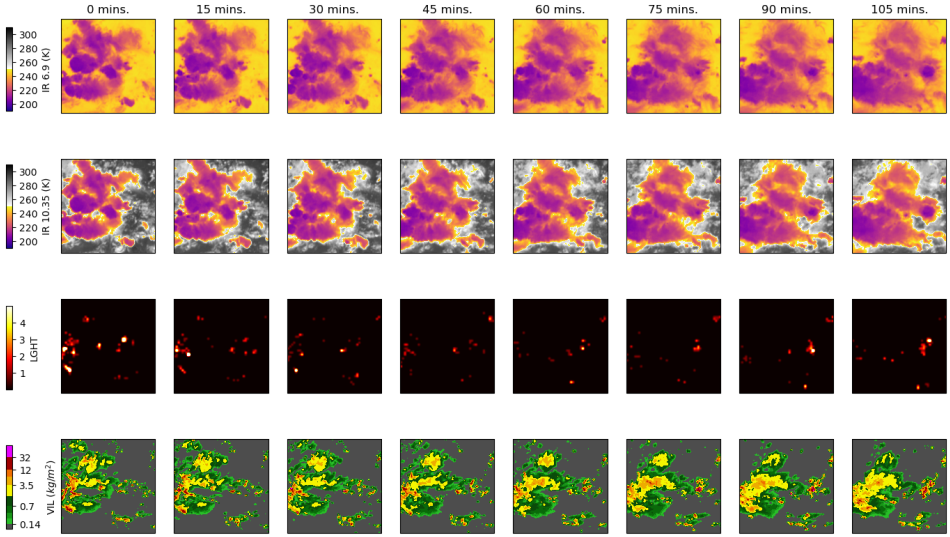
Experiments are conducted in the European Weather Cloud infrastructure

Supplementary slides below!

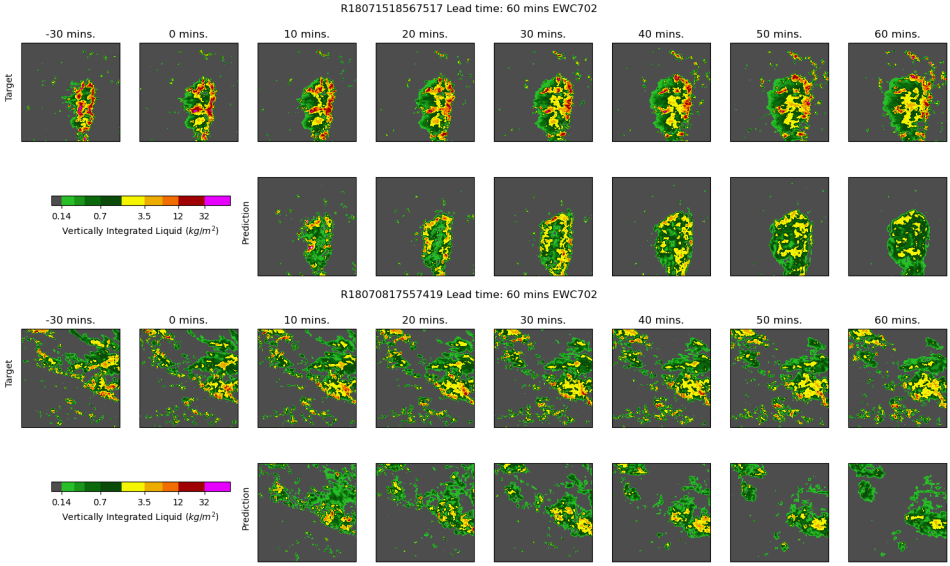
SEVIR Dataset



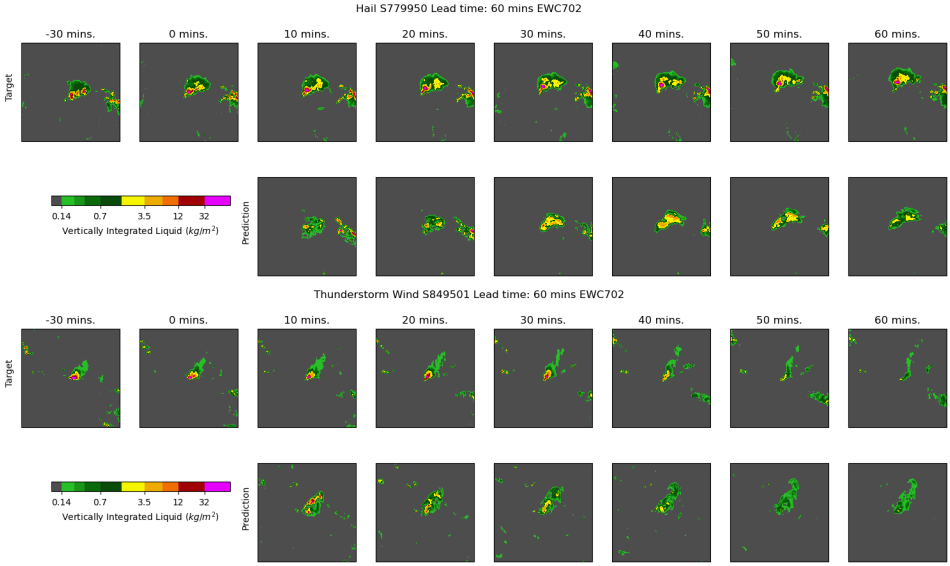
SEVIR Dataset



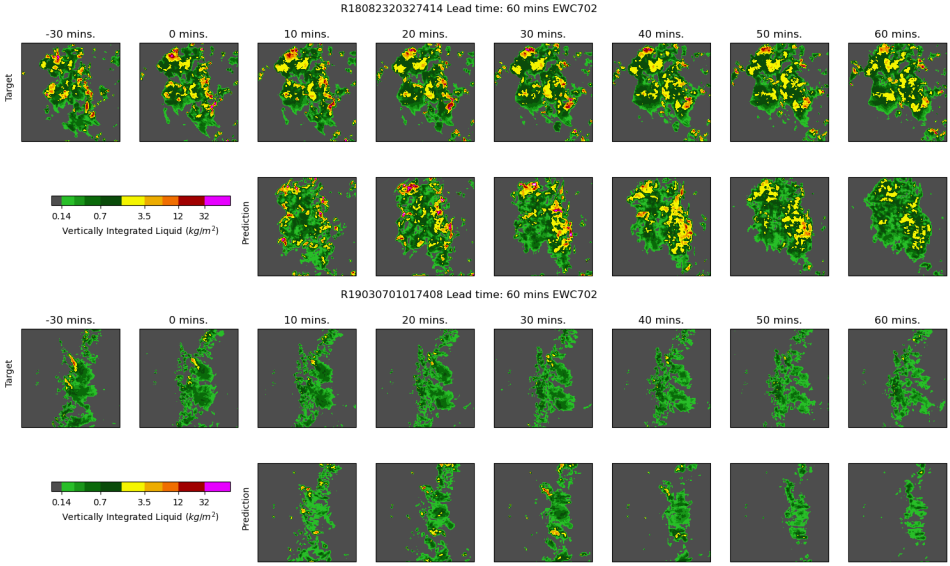
1-hour lead time



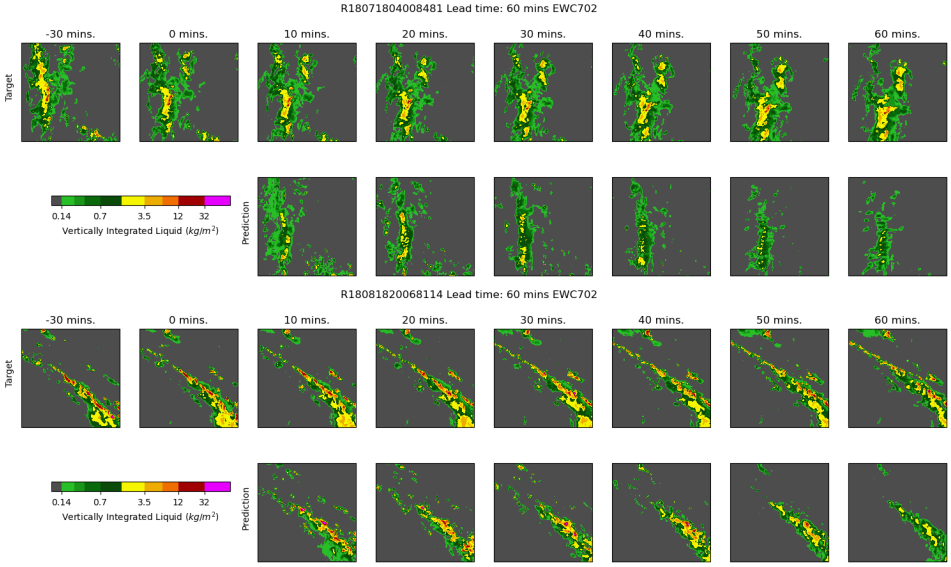
1-hour lead time



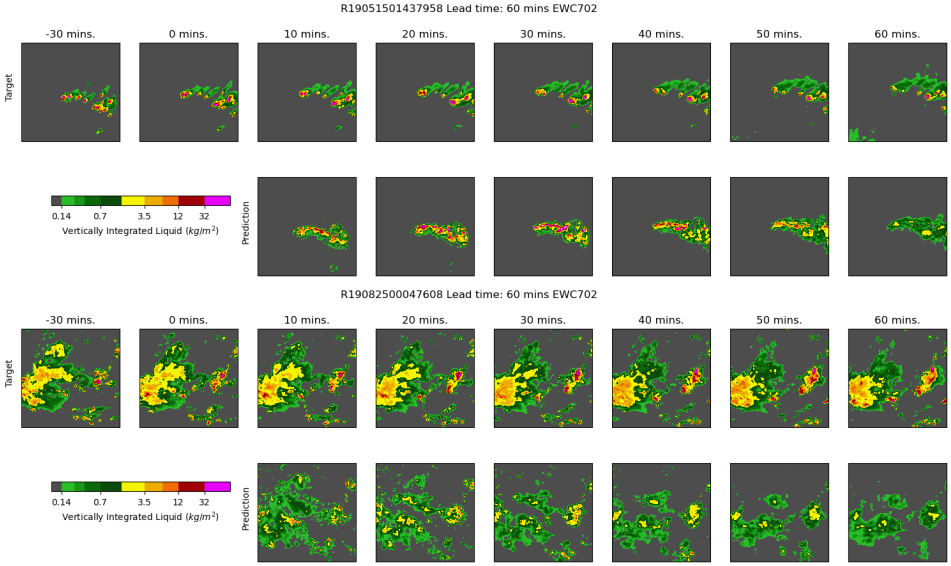
1-hour lead time



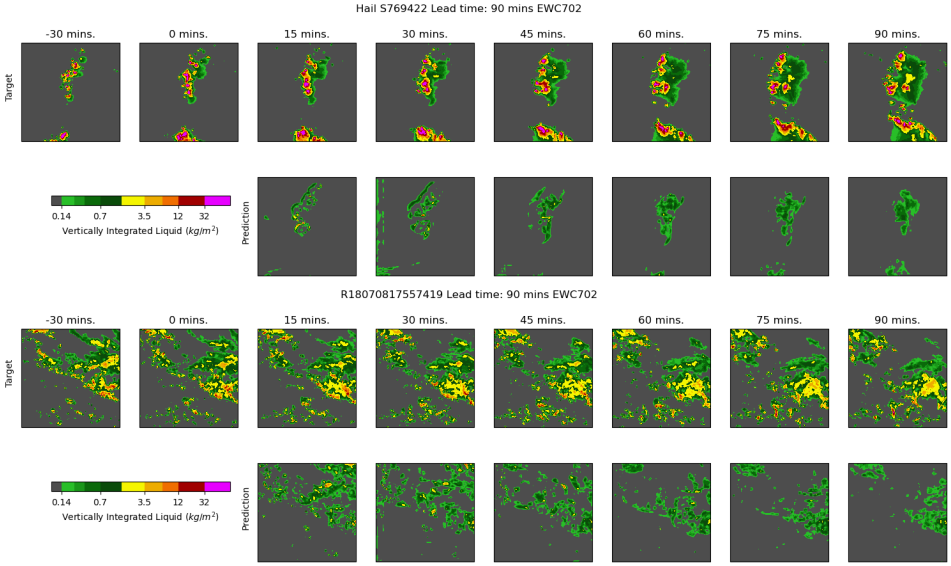
1-hour lead time



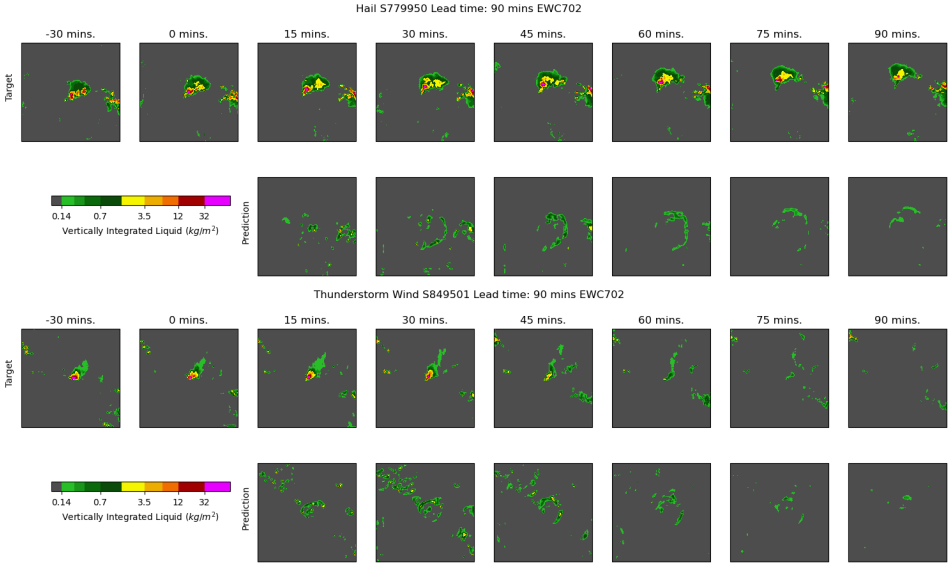
1-hour lead time



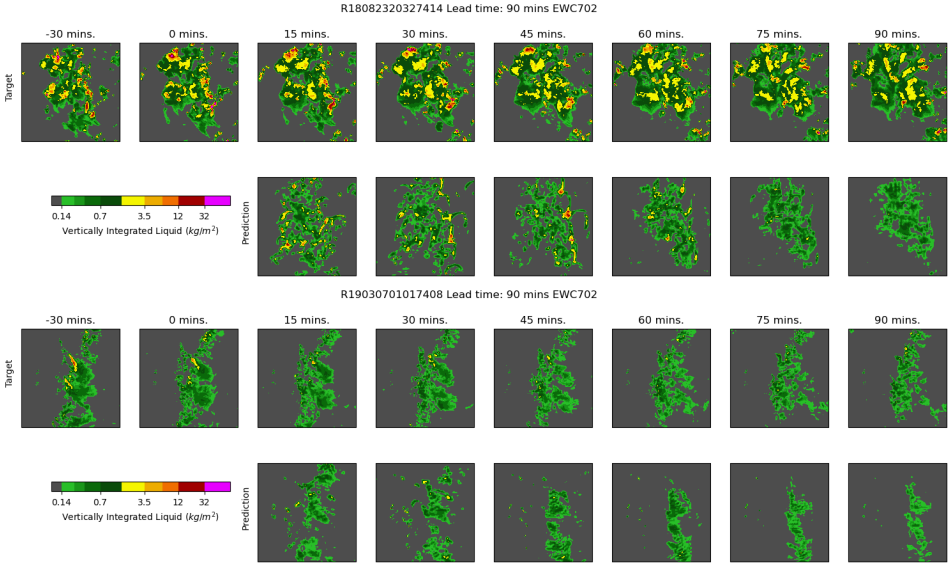
1.5-hour lead time



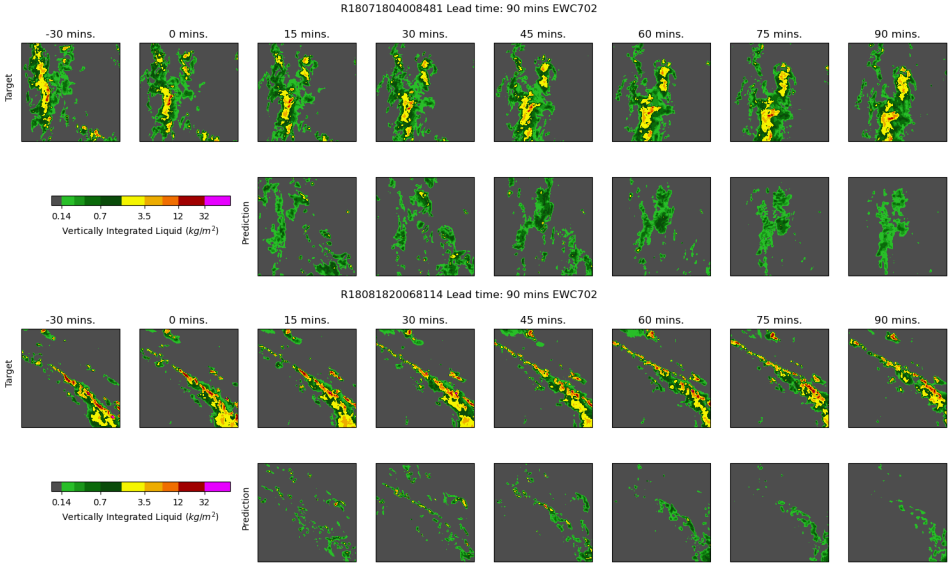
1.5-hour lead time



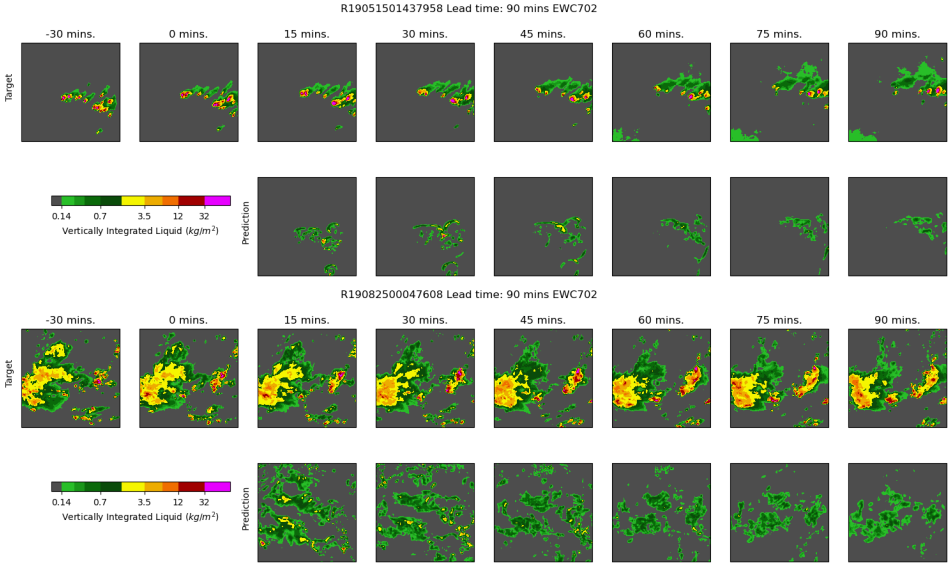
1.5-hour lead time



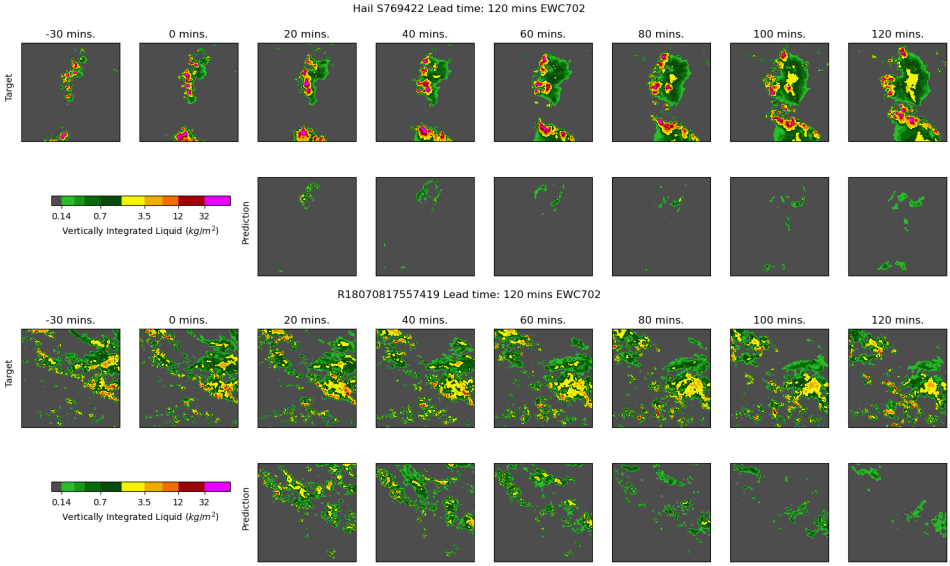
1.5-hour lead time



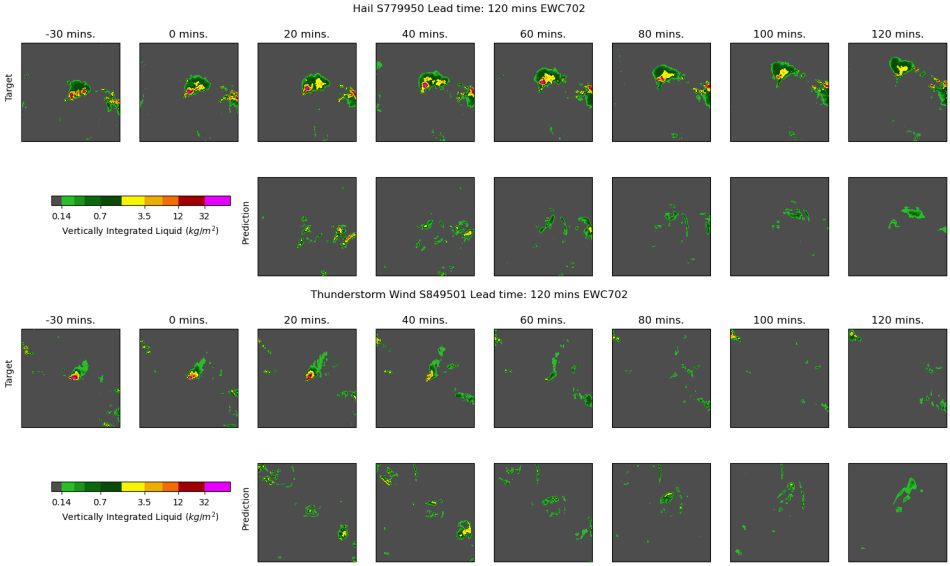
1.5-hour lead time



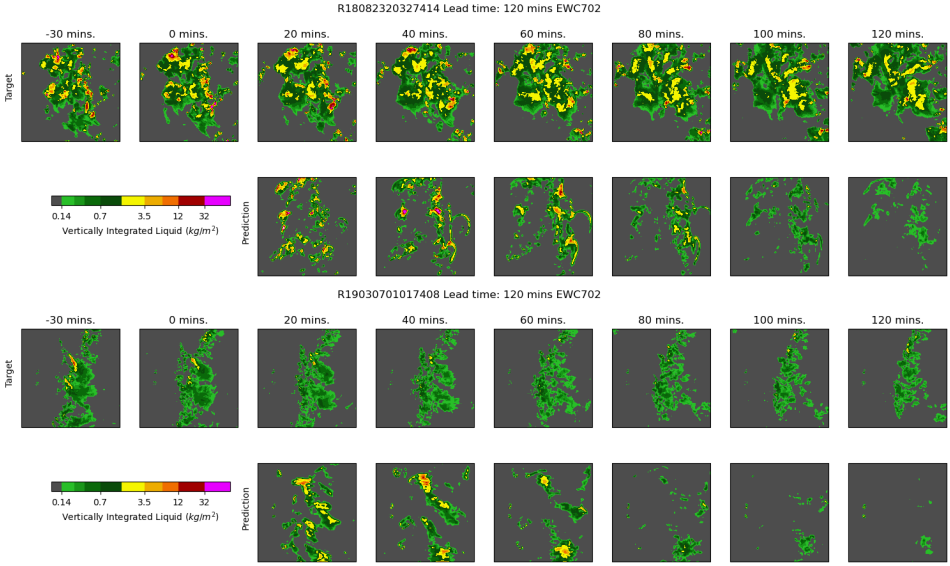
2-hours lead time



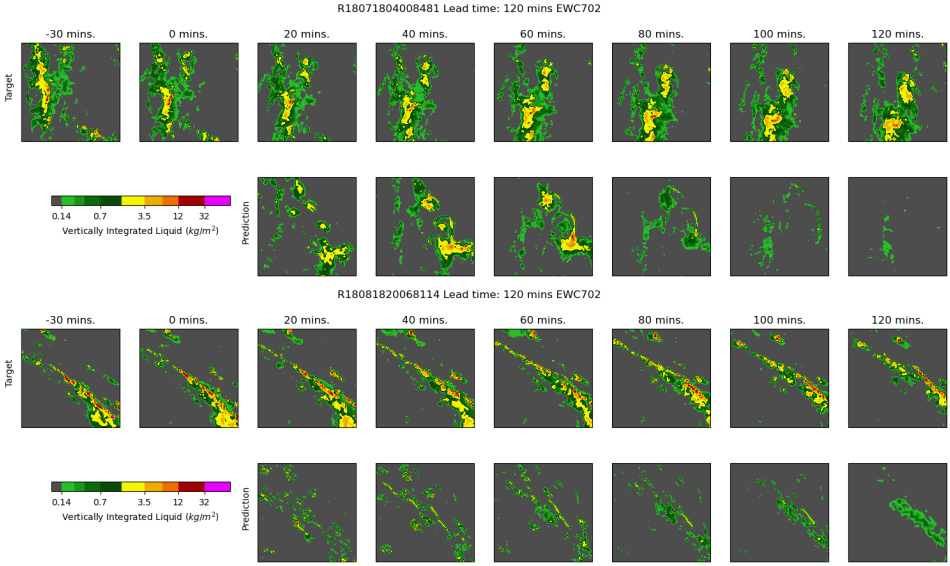
2-hours lead time



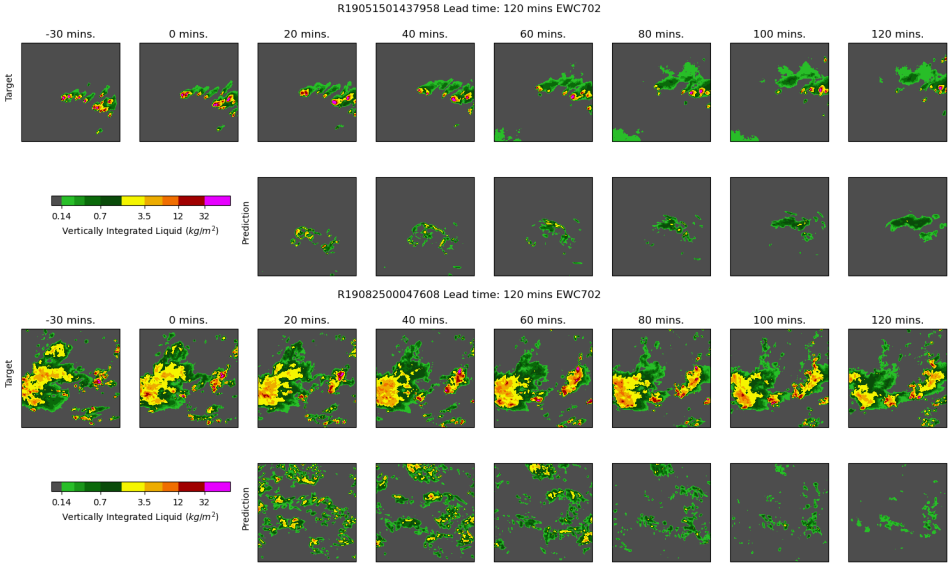
2-hours lead time



2-hours lead time

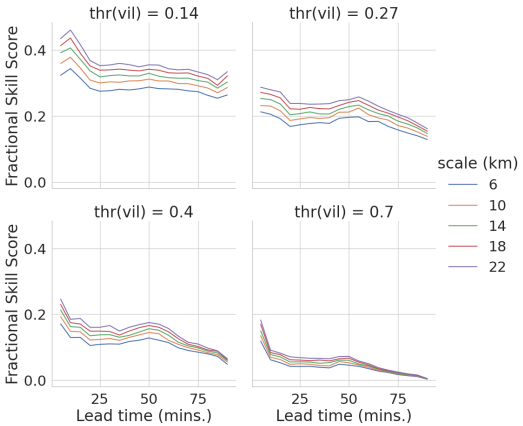


2-hours lead time

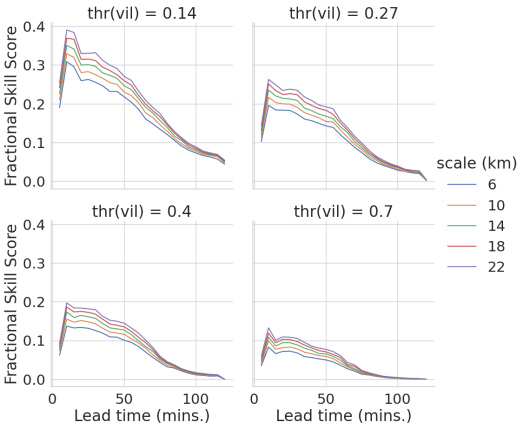


Fractional Skill Score – Different lead times

1.5-hour lead time:



2-hours lead time:



Structure-Amplitude-Location (SAL) Scores – Different lead times

