4-D Nowcasting of Clouds using Artificial Intelligence

David Ryglicki – MyRadar
Anthony Wimmers – UW-Madison/SSEC/CIMSS
Sarvesh Garimella – MyRadar

Funding provided by Office of Naval Research STTR Phase I
Solicitation Topic Number: N023-T025, Award Number: N6833523C0603

Acknowledgements: Jay Hoffman (SSEC/UW-Madison), George Greenwood (MyRadar), Annalisa Graziano (MyRadar), Daniel Hodyss (NRL-DC), Kevin Viner (NOAA EMC), Charlotte Cambier van Nooten (Radbound Univ. Nijmegen), Richard Peal (JMP)

Background and Motivation: Cloud Nowcasting

- What is a cloud!?
  - Can have a variety of definitions!
  - Optical depth?
  - Reflectance?
  - Radiance?
    - Brightness Temperatures
- For this project, two separate Proof-of-Concepts
  1. 1-D (height) Mapping of cloud profiles (SSEC)
     - **Vertical Cloud Profiler (VCP)**
  2. 3-D (time + horizontal) Generative AI Cloud Nowcasting (MyRadar)
     - **Deep Generative Modelling of Satellite (DGMS)**
- Pre-requisite: probabilistic capabilities
- Long-term (potential Phase 2), blend Nowcast with NWP
- Why GenAI?
  - Best short-term, high-res NWP model (HRRR) struggles with cloud nowcasting (Griffin et al. 2017)

Model Descriptions

VCP

- Goal: determine probability of cloud in vertical

Variable (units)	Source	Vertical dist.
Relative humidity (%)	GFS4 0.25° global model	GFS pressure levels
Cloud water content (kg/kg)	GFS4	GFS pressure levels
Vertical displacement from cloud top (km)	ACHA cloud retrieval	GFS pressure levels
Under cloud top (0 or 1)	ACHA cloud retrieval	GFS pressure levels
Optical depth (unitless)	ACHA cloud retrieval	Uniform through column

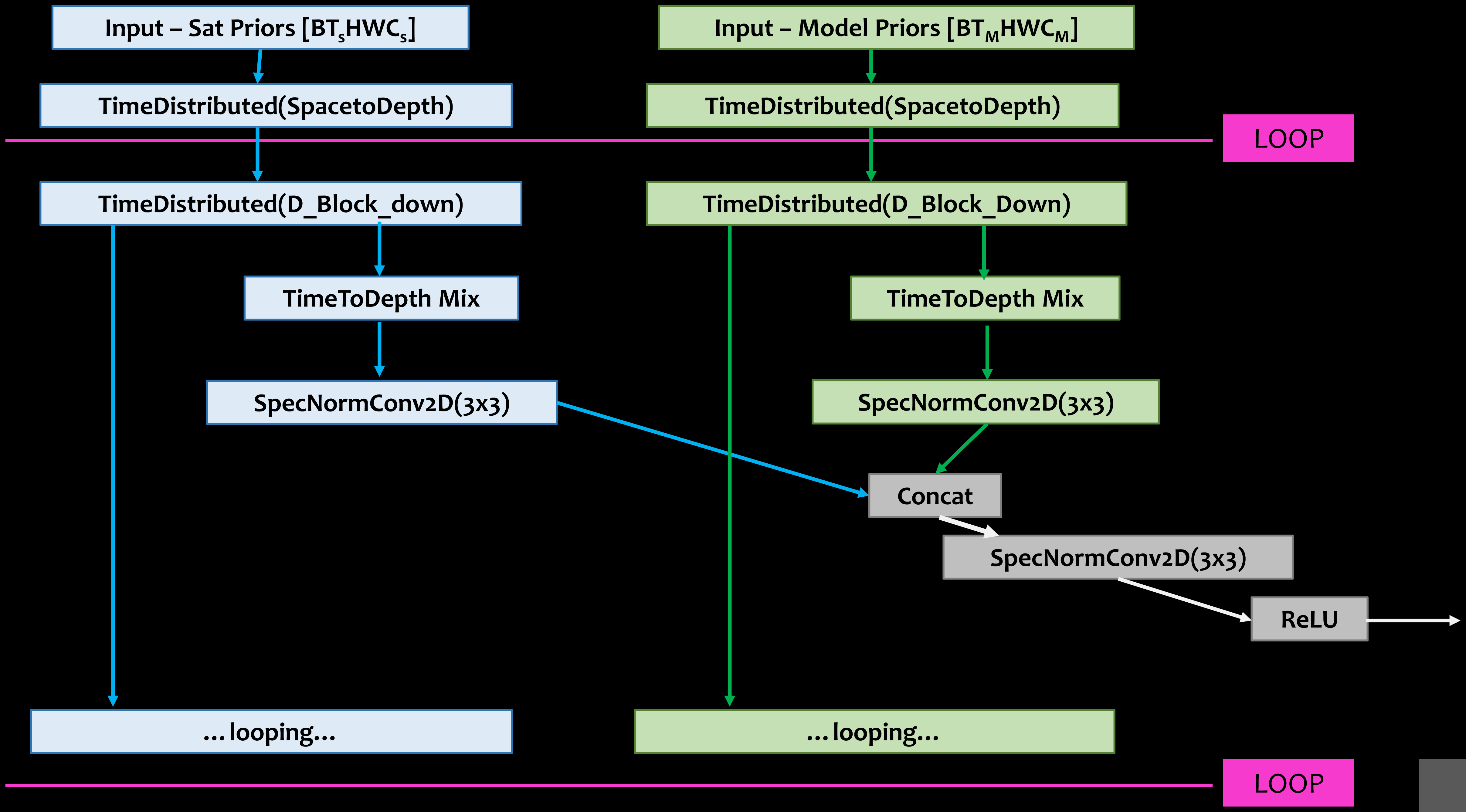
- ACHA: Algorithm Working Group (AWG) Cloud Height Algorithm
- CALIPSO data is extracted within 30 minutes of GFS 3-hourly output, interpolated to CALIPSO overpass
- Gradient-boosted tree method
- Output: probability of cloud (1-D, [Z]) at GFS levels from 50 hPa to 1000 hPa

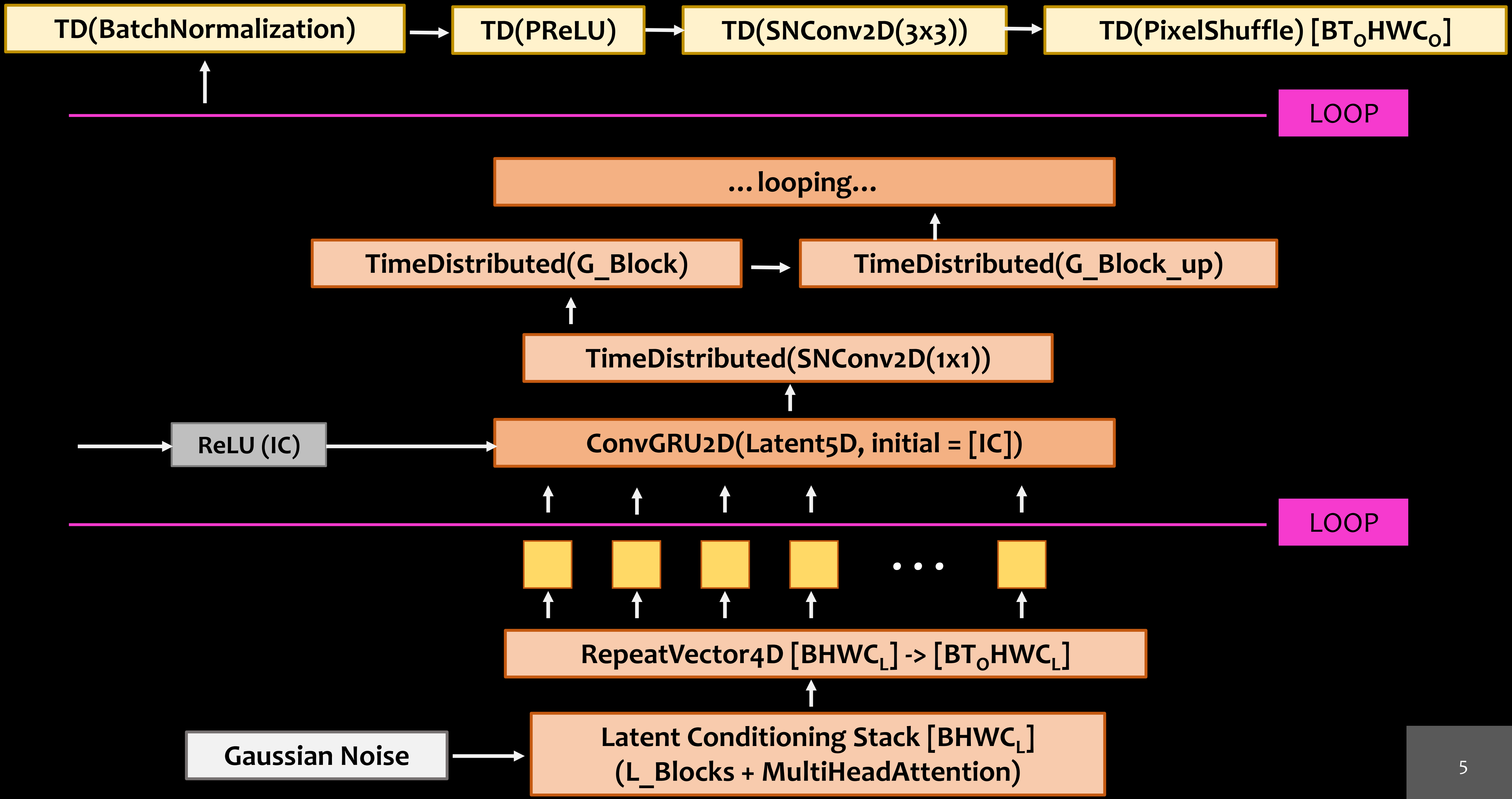
DGMS

- Goal: emulate multi-channel GOES-R imagery

Variable (units)	Source	Vertical dist.
Relative humidity (%)	GFS4 0.25°	GFS layers
Brunt-Vaisala frequency (s^{-1})	GFS4	GFS layers
Zonal wind ($m\ s^{-1}$)	GFS4	GFS layers
Meridional wind ($m\ s^{-1}$)	GFS4	GFS layers
Virtual temperature (K)	GFS4	GFS layers
Total precipitable water ($g\ cm^{-2}$)	GFS4	Single level
Convective available potential energy ($J\ kg^{-1}$)	GFS4	Single level
Convective inhibition ($J\ kg^{-1}$)	GFS4	Single level
T_B of GOES channels 8,9,10,11,12,13 (K)	GOES-16 ABI	Single level

- Modification of Deep Generative Modelling of Radar (DGMR; Ravuri et al. 2021)
  - One generator, two discriminators (Spatial and Temporal)
  - Recommended modifications by Cambier van Nooten et al. (2023); ReLU -> PReLU in decoder convolutions
- 4 prior GOES input times
- GFS data is hourly forecast/analysis closest to initialization time; interpolated to GOES grid (ESMF)
- Reduced model and dataset size to accommodate restricted computing resources (RTX Ada A6000)
  - Layers averaged according to UFS/GFS diagnostic clouds
- Output: multi-channel imagery sequence (3-D, [T, Y, X])

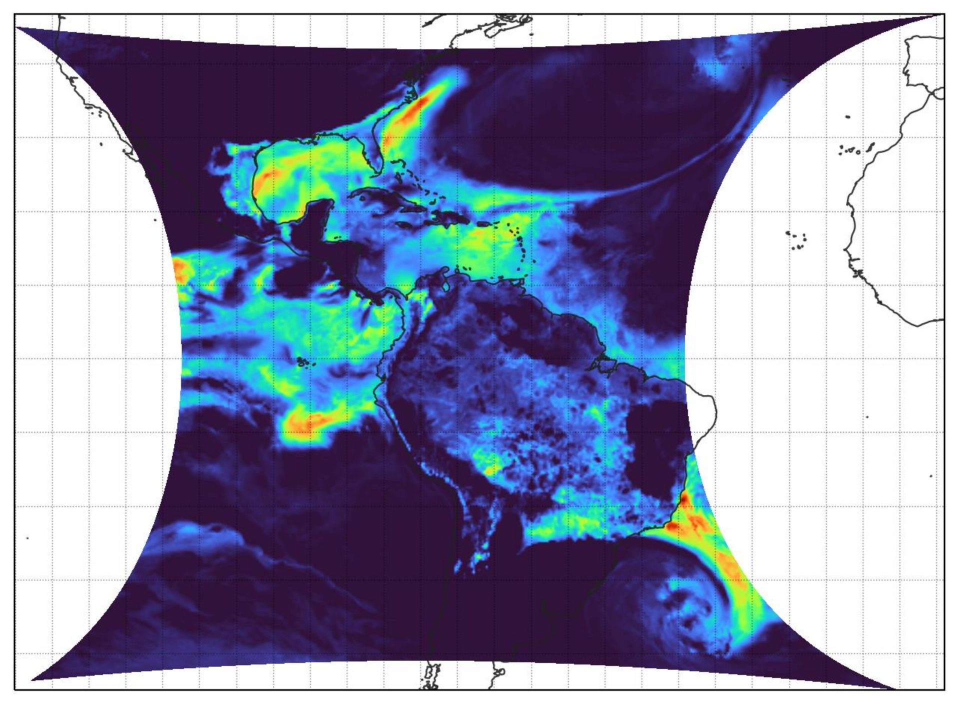

Input – Sat Priors [BT$_S$HWC$_S$]
TimeDistributed(SpacetoDepth)
Input – Model Priors [BT$_M$HWC$_M$]
TimeDistributed(SpacetoDepth)
LOOP
TimeDistributed(D_Block_down)
TimeDistributed(D_Block_Down)
TimeToDepth Mix
TimeToDepth Mix
SpecNormConv2D(3x3)
SpecNormConv2D(3x3)
Concat
SpecNormConv2D(3x3)
ReLU
…looping…
…looping…
LOOP


TD(BatchNormalization)
TD(PReLU)
TD(SNConv2D(3x3))
TD(PixelShuffle) [BT_OHWC_O]
LOOP
… looping…
TimeDistributed(G_Block)
TimeDistributed(G_Block_up)
TimeDistributed(SNConv2D(1x1))
ReLU (IC)
ConvGRU2D(Latent5D, initial = [IC])
LOOP
RepeatVector4D [$BHWC_L$] -> [BT_OHWC_L]
Gaussian Noise
Latent Conditioning Stack [$BHWC_L$]
(L_Blocks + MultiHeadAttention)

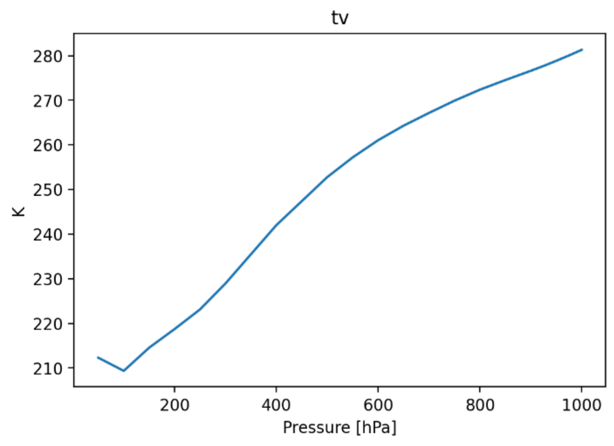
Data - Overview

- VCP
  - CALIPSO tracks from 2018
    - Vertical Feature Mask
    - January, April, July, October
      - 6-15 for training
      - 1-5 for validation
    - Within 60° of ABI zenith angle
    - Within 30 minutes of available GFS data
    - Vertical positions interpolated to GFS pressure levels
- DGMS
  - GOES data from April 2023
    - GFS hourly data interpolated to GOES subdomain (ESMF restriction)
      - Standardized by global mean/std. dev.
    - 256x256 subdomains extracted over water
    - ~299K total files after cleaning
      - Only selected files with >20% of C13 T_B < 263 K
        - Addresses "persistence" modeling
      - 56K files; 10K or 3K for training, 1K for testing

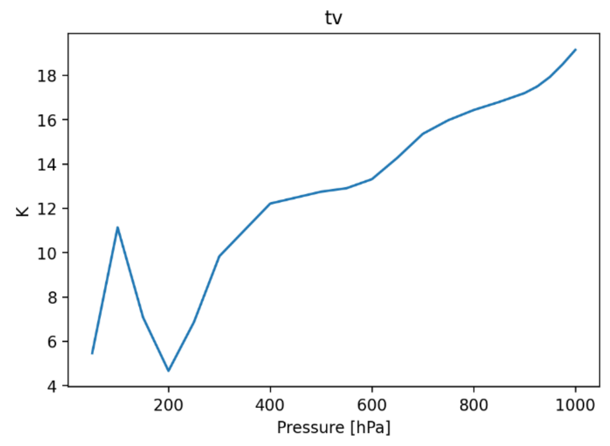
Interpolated CAPE



Mean T_v

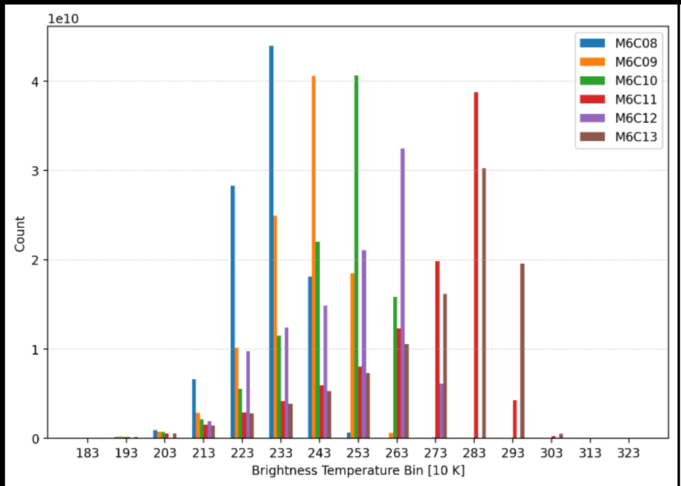


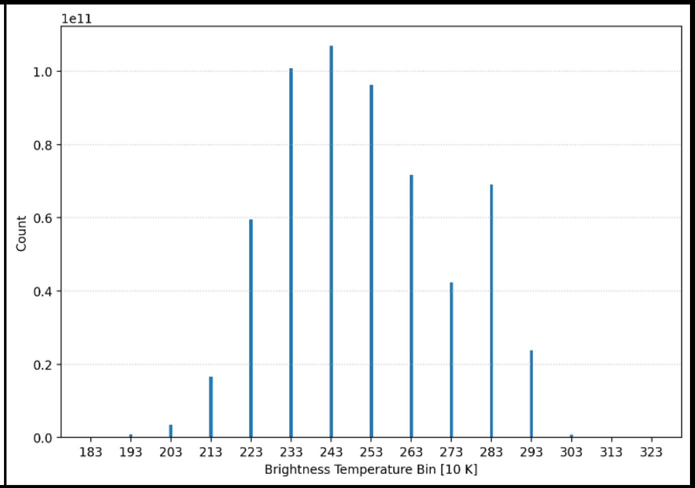
Std. Deviation T_v

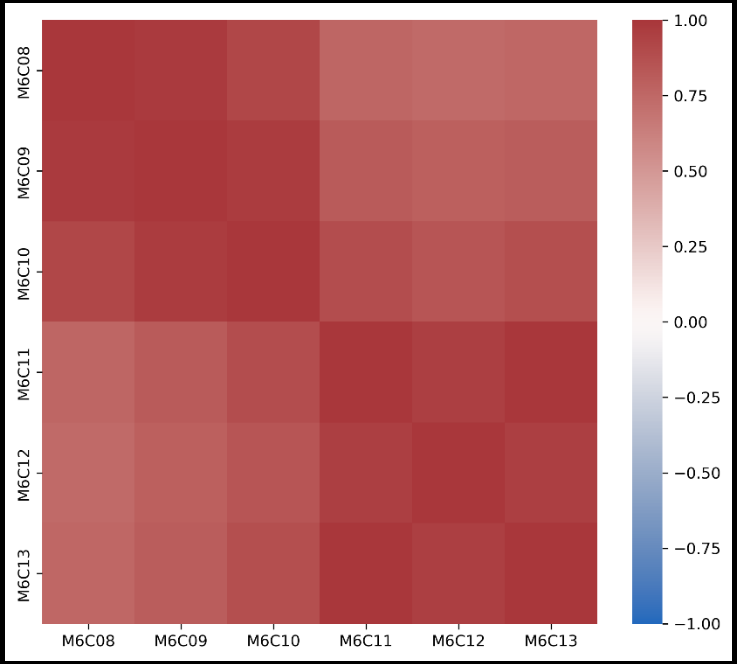


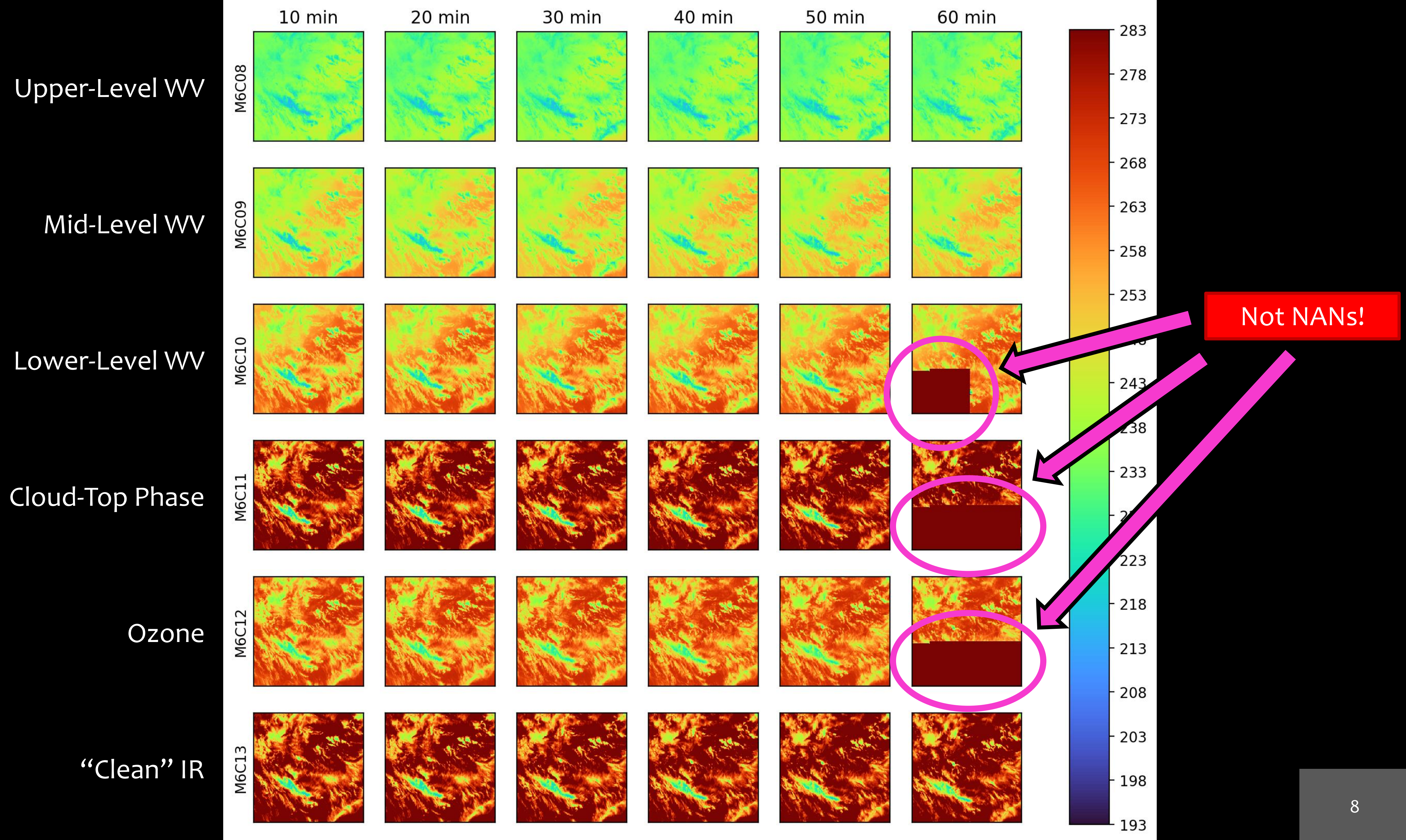
DGMS Satellite Data Processing

- Use of “Balanced” pixel losses
  - Shi et al. (2017) used step-wise weighting of rain rate
  - Hilburn et al. (2020) used inverse PDF of reflectivity for weighting
  - Ravuri et al. (2021) and Cambier van Nooten (2023) used linear weighting of rain rate
- “What is a cloud?”
- Satellite Brightness Temperatures: not an easy distribution to model parametrically!
  - Ideally, each channel should have its own PDF-based loss
  - For our purposes, focus on Channel 13 (“Clean” IR)
    - $e^{4.0}$ relationship
- 2 groups
  - Water Vapor
  - “Other” IR





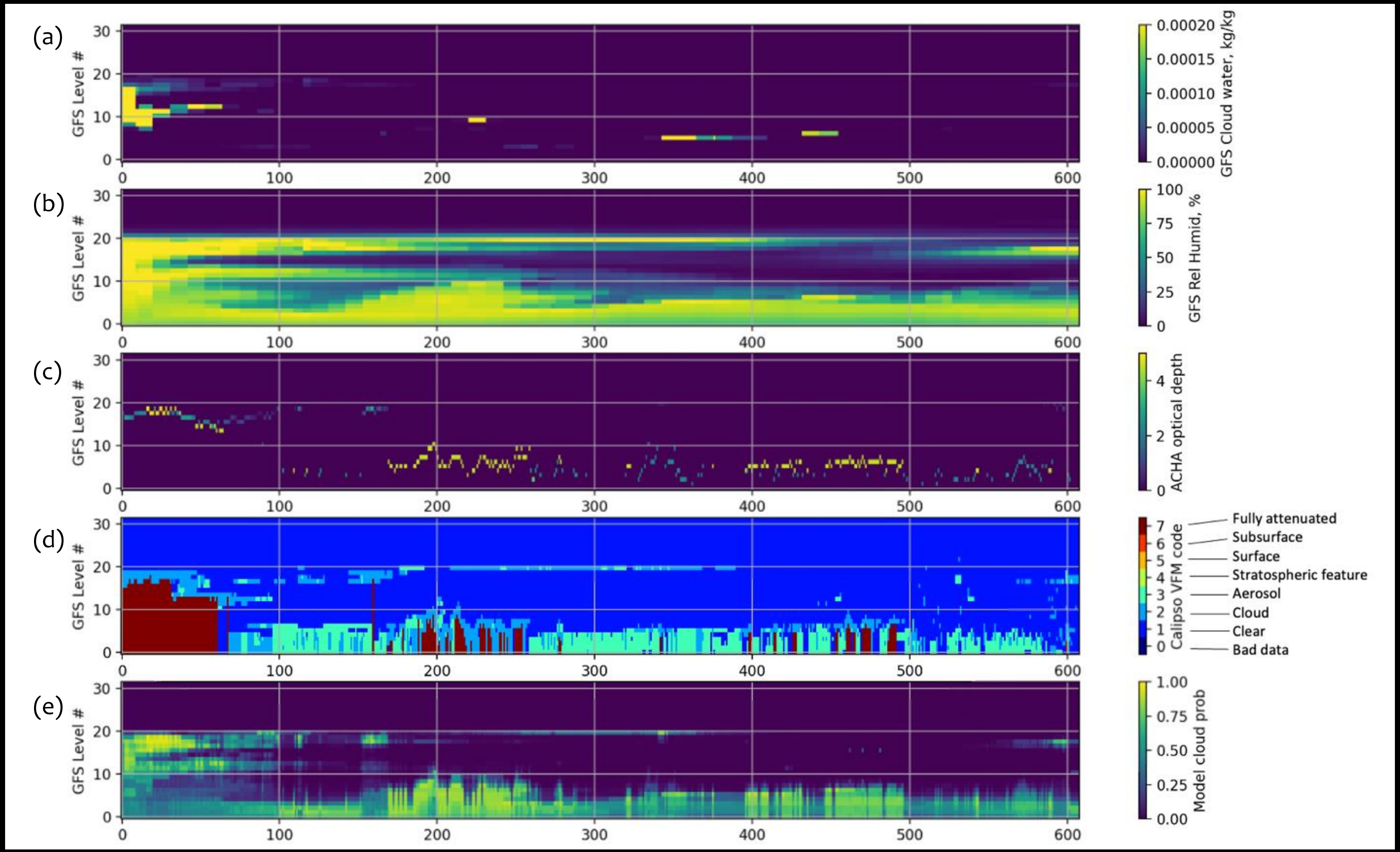




Upper-Level WV
Mid-Level WV
Lower-Level WV
Cloud-Top Phase
Ozone
"Clean" IR
10 min
20 min
30 min
40 min
50 min
60 min
M6C08
M6C09
M6C10
M6C11
M6C12
M6C13
283
278
273
268
263
258
253
243
238
233
223
218
213
208
203
198
193
Not NANs!

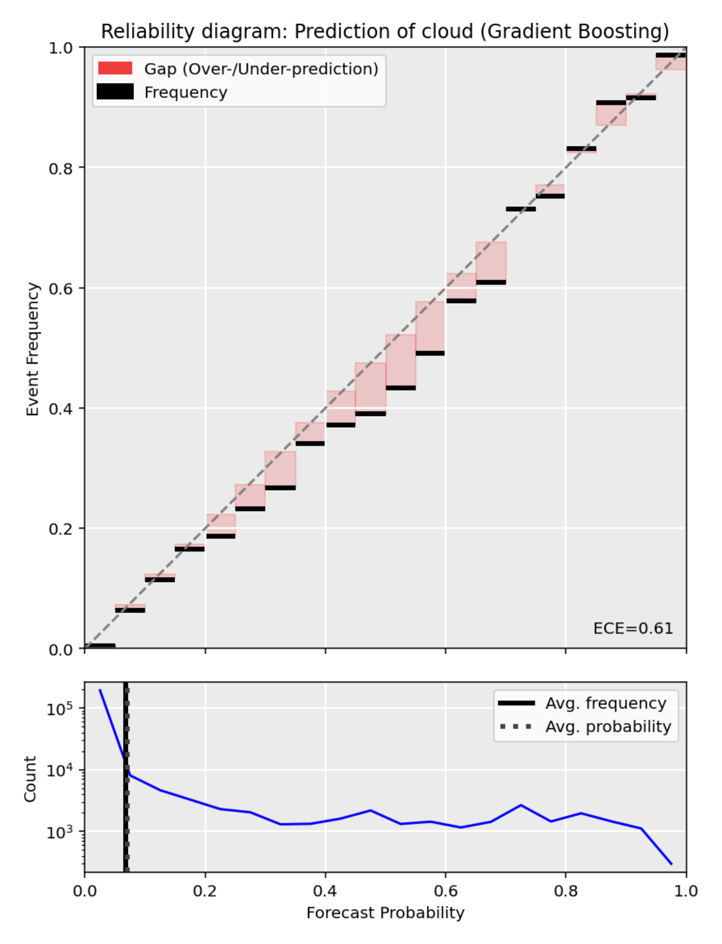
Results: VCP – Imagery

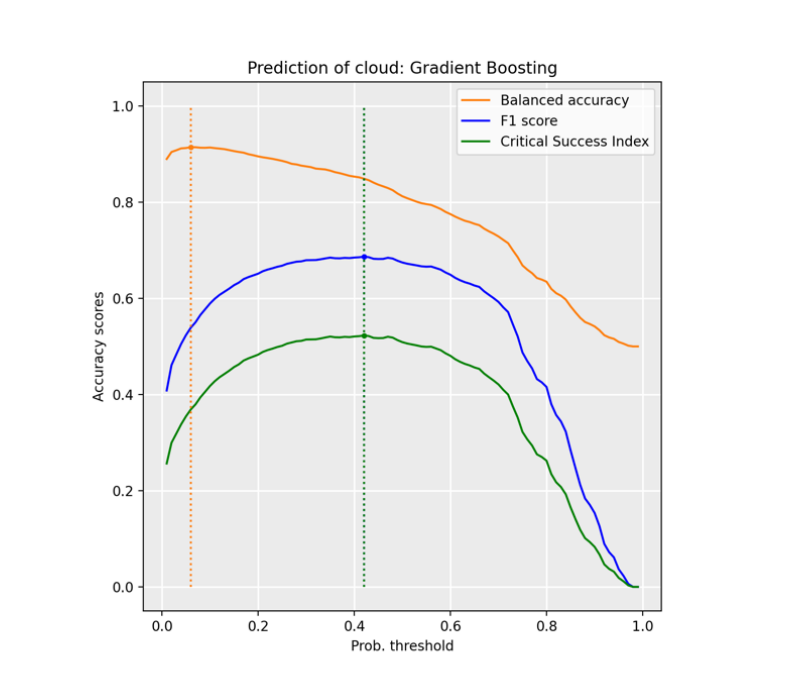
- Transect is ~3000 km, stepped every 5 km
- CALIPSO data is interpolated to GFS levels
  - CALIPSO cloud mask (truth) in (d); focus on “cloud” and “clear”
- Truth (d) indicates three kinds of clouds
  - High cirrus
  - Middle level
  - Lower-level/Boundary layer clouds
- GFS RH helps fill in gaps from cloud water



Results: VCP – Statistics

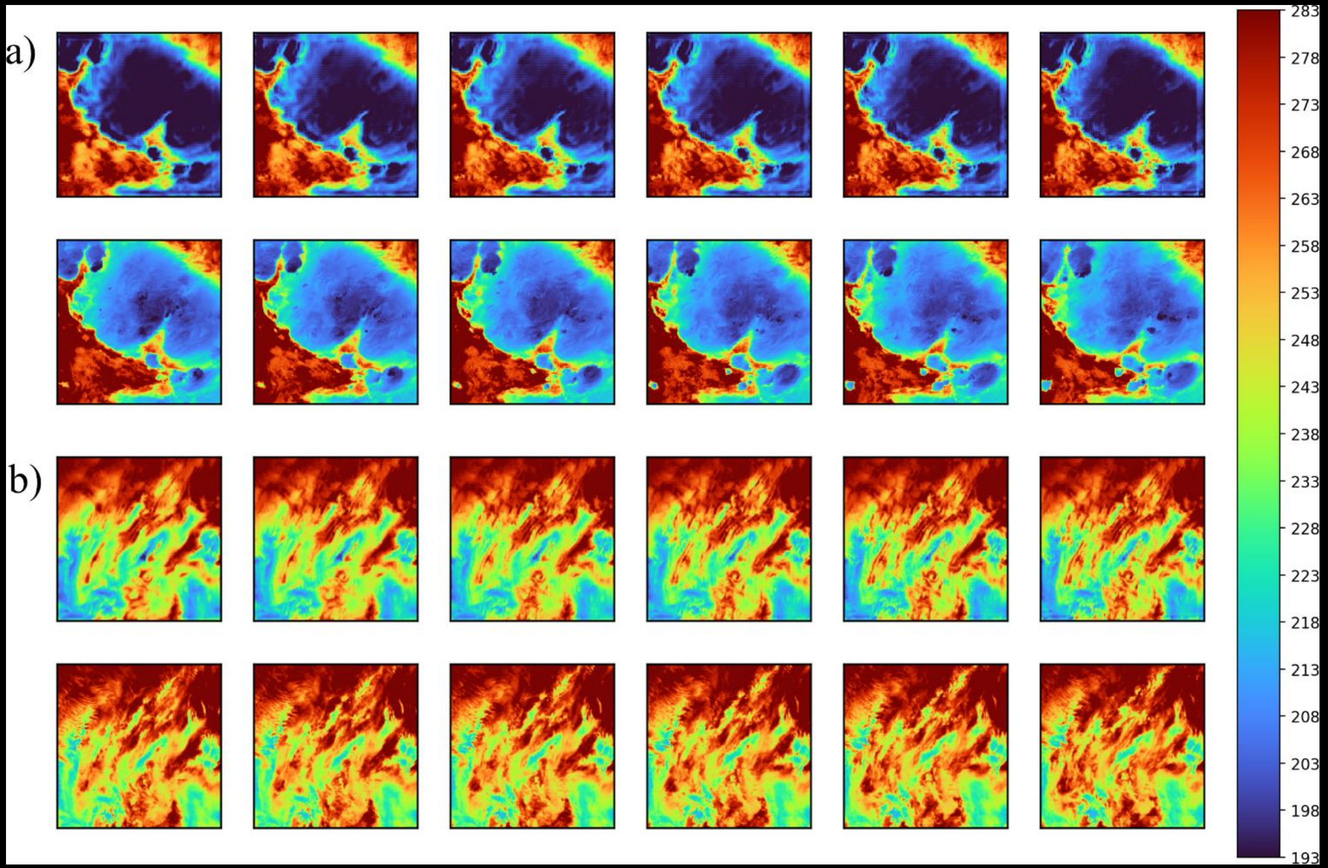
- Reliability diagram (left) indicates slight low bias (5%)
  - Most observations within no cloud bin
- Simplifying classes to cloud/no cloud, optimal detection threshold of 0.42
  - BA: 0.85
  - F1 Score: 0.69
  - CSI: 0.52





Results: DGMS – Imagery

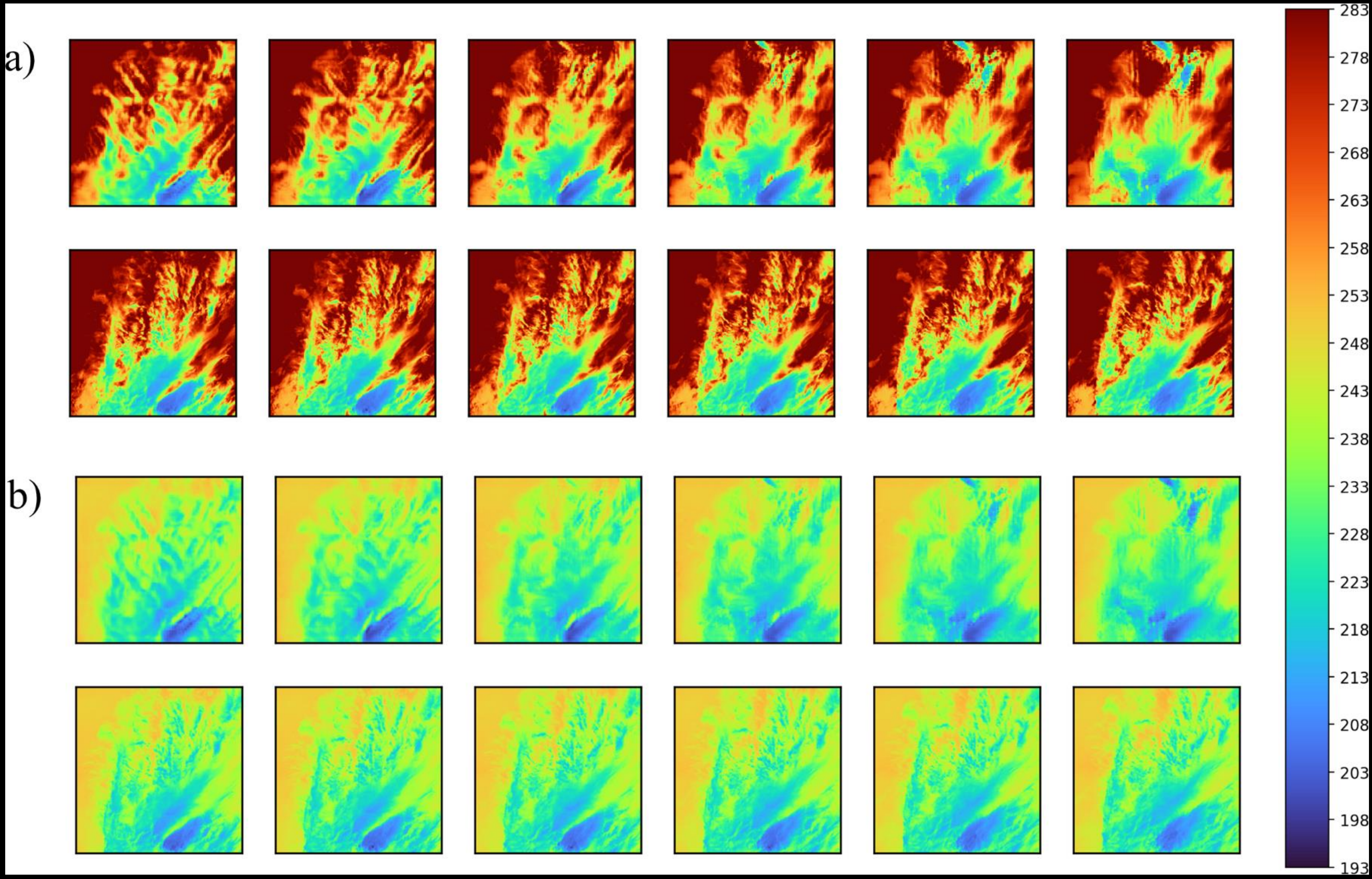
- Initially, only one channel (13)
  - Crisp results, cold bias, “stuck” on persistence
  - Including too many “clear” scenes restricted GRU
- Further refinements:
  1. Only selected images where 20% coverage of C13 T_B < -10 °C
  2. Add bias to all convolutions
     - Setting update_gate bias in ConvGRU2D to start as all predictor
  3. Tweaking of learning rates and weight on L1 + L2 loss
  4. Output all 6 GOES channels



Top row: DGMS
Bottom Row: Observations
10-minute interval

Results: DGMS – Imagery (cont'd)

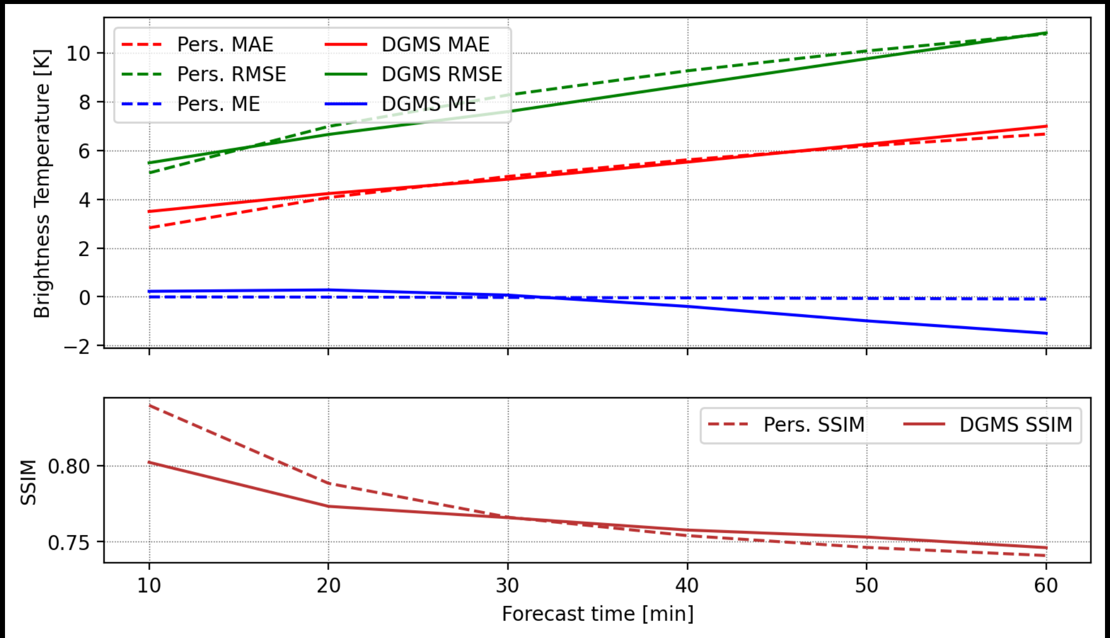
- After initial tests, migrated to multi-channel solutions
  - Multiple experiments, tuning parameters for stability
- Cross-hatching artifacts still remain, but broad structures exist
  - No evident bias
- Convective generation!
- Spatiotemporal consistency across the channels
  - (a) is C13, (b) is C9

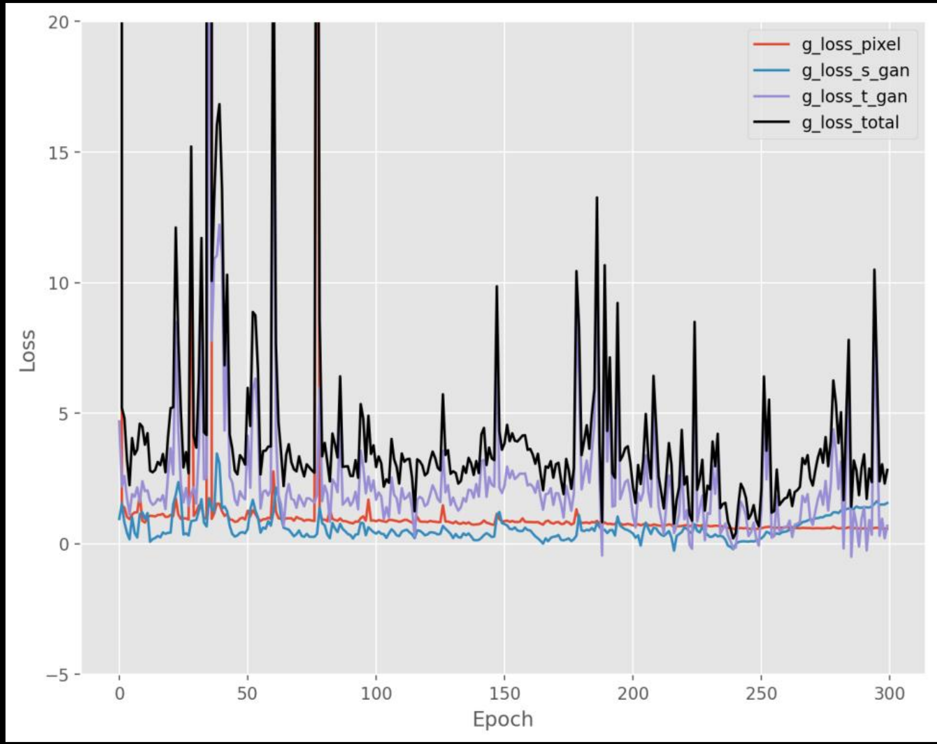


Top row: DGMS
Bottom Row: Observations
10-minute interval

Results: DGMS – Statistics

- Competitive results with persistence baseline
  - Slight cold bias in time
    - Disappears at a later epoch, but MAE/RMSE worse
- Temporal discriminator stability remains challenging
  - May need R1/R2 regularization or simply more filters/channels
  - Exponential Moving Average?





Summary & Next Steps

- Successfully created two component proof-of-concepts
  - Vertical Cloud Profiler (VCP): emulate CALIPSO data
  - Deep Generative Modelling of Satellite (DGMS): emulate multi-channel ABI sequence
- VCP
  - Expand dataset, possible inclusion of CloudSat (messier dataset)
  - Explore deep learning (ConvNets) to improve upon gradient-boosted trees
- DGMS
  - Expand dataset (convert to TFRecord), more ABI channels, more filters (more powerful hardware), EMA?
  - Temporal Discriminator needs to be addressed
    - R1/R2 regularization penalties (Mescheder et al. 2018), EMA?
  - Still some checkerboarding artifacts
    - Borrow some super-resolution post-pixel-shuffle convolution methodology (ESRGAN, Wang et al. 2018)
- Both!
  - Process channels 8-16 in DGMS ([T, Y, X]), process with ACHA algorithm, process with VCP ([Z]), blend with NWP output ([T, Z, Y, X])
    - Ablations on NWP input, GOES priors
  - Line of sight algorithm, 4-D visualization, integration into Navy systems

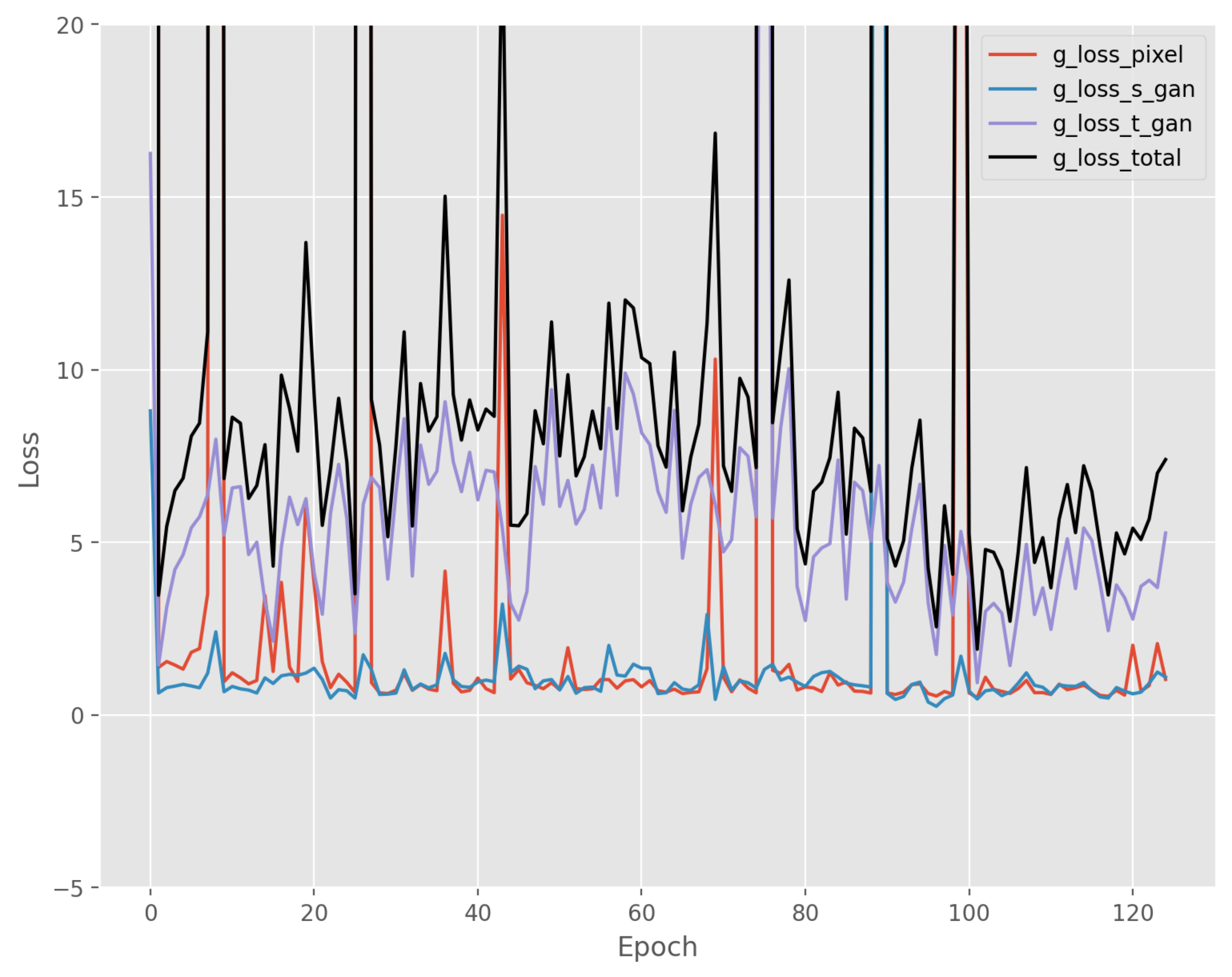
Thank you!

Results: VCP – Statistics

Statistic (aliases)	Formula	Value	Meaning (perfect score)
Brier Score	$\frac{1}{N}\sum_{t=1}^{N}(f_t - o_t)^2$	0.030	Mean squared error of the probabilistic prediction (0)
AUC	---	0.97	Area under the curve in Figure 5 (1)
Optimum threshold		0.42	Cloud probability threshold that optimizes the F1 and CSI scores
precision	$\frac{TP}{TP+FP}$	0.65	Fraction of cloud predictions that are correct (1)
recall (sensitivity, hit rate, true positive rate)	$\frac{TP}{TP+FN}$	0.73	Fraction of cloud events that are correctly predicted (1)
miss rate (false negative rate)	$\frac{FN}{TP+FN}$	0.27	Fraction of cloud events that are incorrectly predicted (0)
specificity (selectivity, true negative rate)	$\frac{TN}{TN+FP}$	0.972	Fraction of no cloud events that are correctly predicted (1)
fallout (false positive rate)	$\frac{FP}{TN+FP}$	0.028	Fraction of no cloud events that are incorrectly predicted (0)
Balanced accuracy	$\frac{1}{2}(\text{recall} + \text{specificity})$	0.85	Accuracy, assuming the costs of FN and FP are the same (1)
F1 score	$\frac{2TP}{2TP+FN+FP}$	0.69	The harmonic mean of precision and recall (1)
Critical Success Index (threat score)	$\frac{TP}{TP+FN+FP}$	0.52	Correctly predicted cloud out of all predictions and unpredicted cloud (1)

Model Details

- DGMR ported entirely to TF-Keras with custom layers
  - Serializable
  - 98% reproduced
    - For flexibility, model is initialized with arrays [*None, T, None, None, C*]
    - Problems with Attention Module in Latent Stack
      - Replaced DeepMind Attention with Keras MultiHeadAttention
- Why DGMR in Keras?
  - Flexibility!
  - DGMR’s latent processing at bottom of Res-U-net allows for generalization to larger domains*
  - Second paper by independent group (Cambier van Nooten et al. 2023)
  - Training is driven by YAML “namelist” files (WRF, MM5, CM1, etc.)
    - Rapid iteration; almost nothing is hard-coded
- Modifications
  - Filter sizes [48, 96, 192, 384, 768] changed to [32, 64, 128, 256, 512] where applicable
  - Batch size of 16 to 12
  - Decoder
    - Zero Padding replaced by Reflect Padding prior to convolutions, includes Residual Blocks
      - Combat edge effects
    - ReLUs replaced with PReLUs



g_loss_pixel
g_loss_s_gan
g_loss_t_gan
g_loss_total
Loss
Epoch
20
15
10
5
0
−5
0
20
40
60
80
100
120

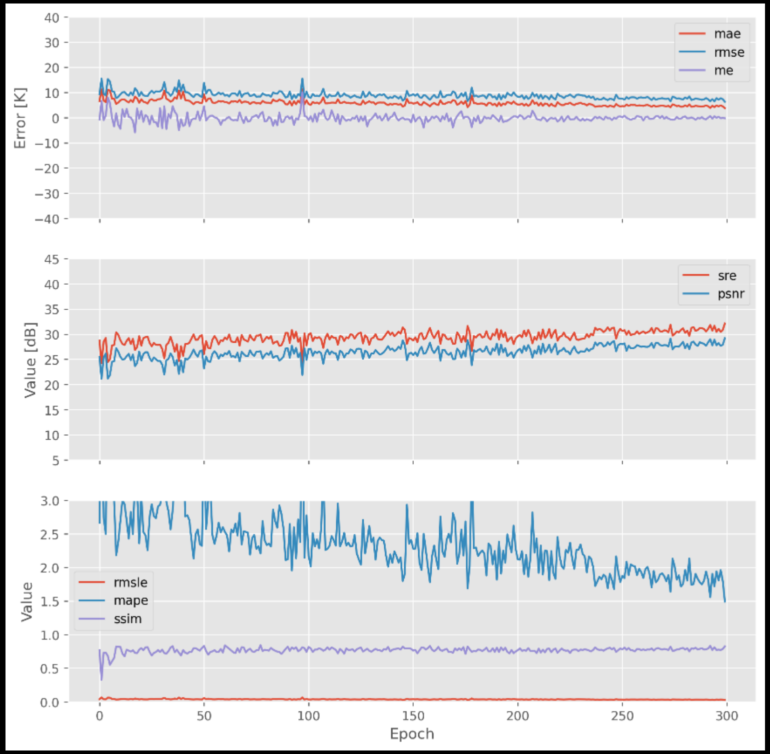
DGMS Model Description

- Basis for Cloud Nowcasting model is **Deep Generative Model of Radar (DGMR; Ravuri et al. 2021)**
  - Descendant of DVD-GAN (Brock et al. 2019)
  - 1 Generator, 2 Discriminators (Spatial and Temporal)
  - Hilburn Weighted [MAE + MSE] + Hinge Adversarial Losses (x2)
- Modifications necessary for computing limitations
  - 256 GB RAM, 48 GB VRAM (RTX Ada A6000)
    - Still not enough!
  - Same basic multi-grid (U-Net), predictor-corrector (ConvGRU2D) architecture
- Ported entirely to Keras (98% reproduced)
  - Filter sizes [48, 96, 192, 384, 768] changed to [32, 64, 128, 256, 512] where applicable
  - Batch size of 16 to 12
  - Custom DeepMind Attention replaced by Keras MultiHeadAttention
  - Decoder
    - Zero Padding replaced by Reflect Padding prior to convolutions, includes Residual Blocks
      - Combat edge effects
    - ReLUs replaced with PReLUs

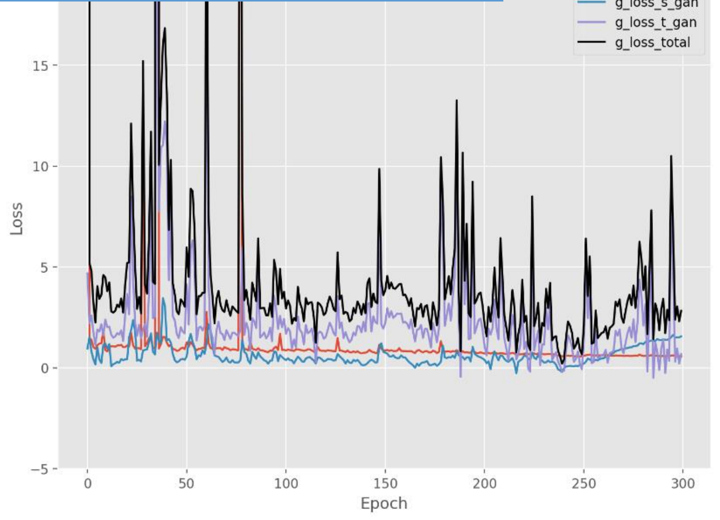
- Two separate input pathways
  - Satellite
    - GOES-East
    - ***Six Channels: 8, 9, 10, 11, 12, and 13***
      - Downstream ACHA/CLAVR-X processing
    - 4 prior images (10 minutes; Ravuri et al. 2021, Espeholt et al. 2020)
  - Model data
    - GFS
      - Variables informed by other cloud/convective studies (Nguyen et al. 2023, Ukkonen and Makela 2019, Kamangir et al. 2021)
      - Chose closest forecast or analysis time to Nowcasting initialization
    - ***CAPE, CIN, TPW***
    - Layer averages
      - ***N^2, T_v, U, V, RH***
      - Averaged over four cloud layers from GFS diagnostic output (boundary, low, mid, high)
    - 23 total variables ("channels")
- Output
  - 6 forecast lead times (10 minutes up to 1 hour) of six GOES channels

Results: DGMS – Statistics

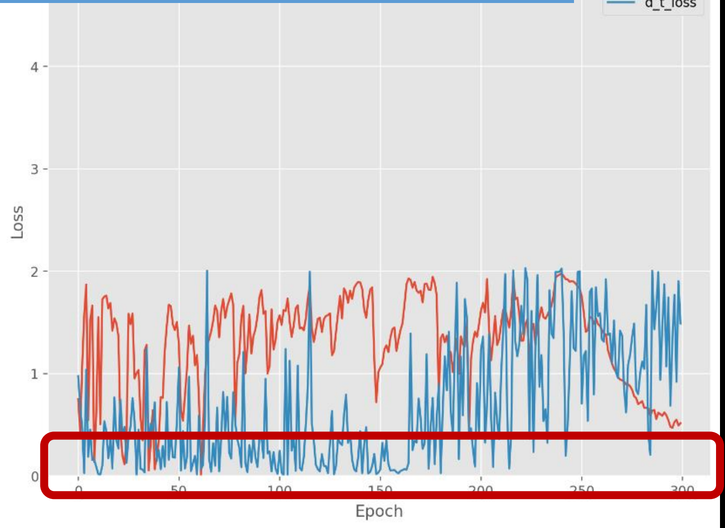
- MAE, RMSE, ME (bias), RMSLE, MAPE, SSIM, SRE (Lanaras et al. 2018), PSNR
  - Have separate metrics for forecast lead time, but not appropriate yet
- Bulk results indicate deceptive flatlining overall
- Temporal Discriminator appears to be source of problem
  - Adversarial loss function of discriminators is a hinge Wasserstein-style loss (Brock et al. 2019)
  - TD learns to separate too quickly, instability in scores
- Spatial Discriminator is operating as expected
- Instability in generator
  - Learning rate decay has stabilized the output somewhat



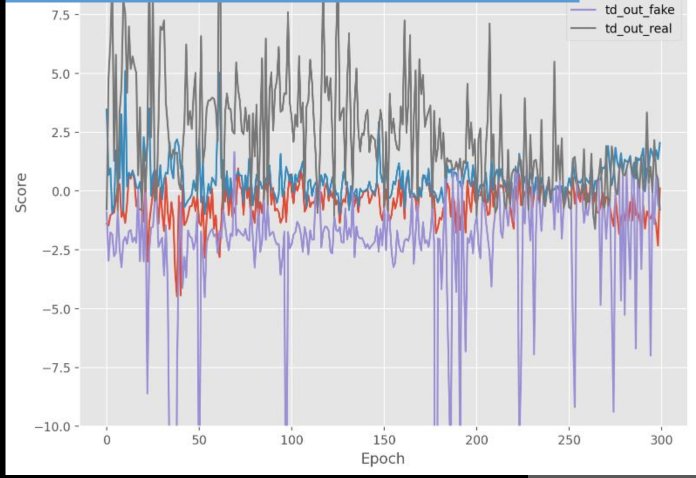
Generator Losses

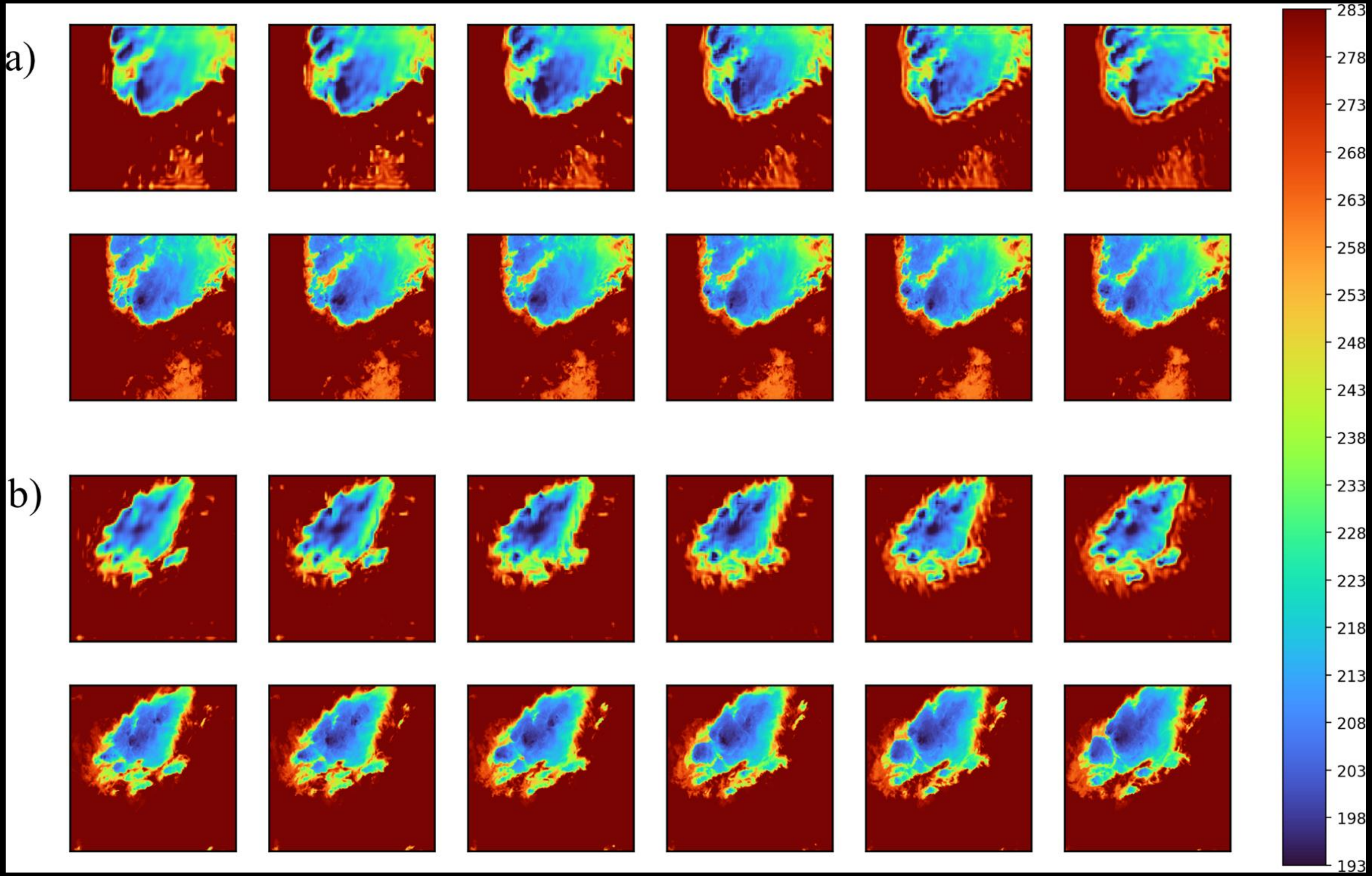


Discriminator Losses



Discriminator Scores




a)
b)
283
278
273
268
263
258
253
248
243
238
233
228
223
218
213
208
203
198
193