RealPEP P2 QPN

Satellite observations of total column water vapor, clouds and convective intiation

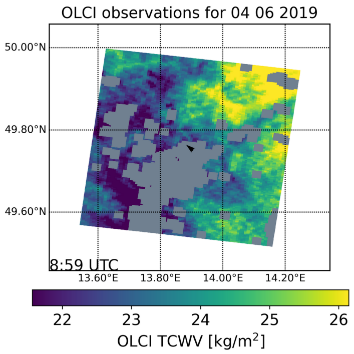
Cintia Carbajal Henken, Jan El Kassar, Rene Preusker

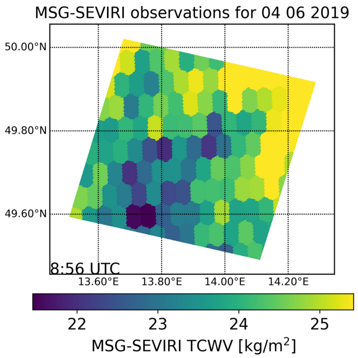
Status on most recent satellite products

Total column water vapor **TCWV** in clear-sky regions

- **OLCI** on polar-orbiting satellite Sentinel-3a/3b:
  - high precision 2d TCWV fields
  - high spatial resolution 300x300m^2
  - morning time snapshots (9-10 UTC)
  - processed 2016-2022

- **SEVIRI** on geostationary satellite MSG:
  - lower accuracy timeseries 2d TCWV fields
  - spatial resolution ~3x6km^2
  - every 15 min
  - processed 2016-2022





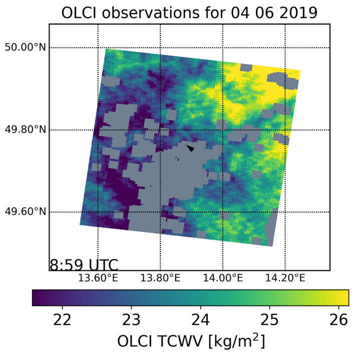
Freie Universität Berlin

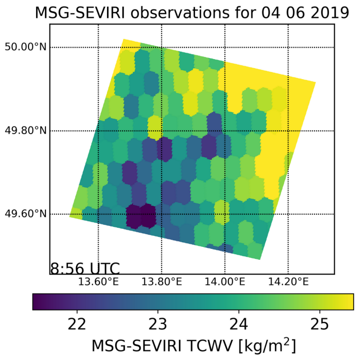
RealPEP

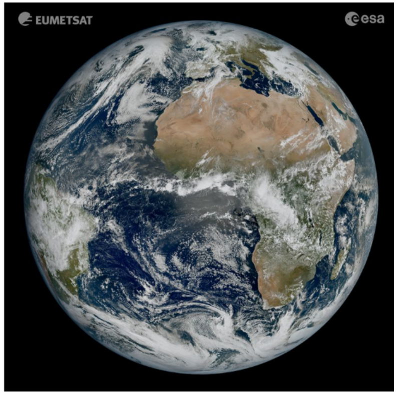
Status on most recent satellite products

Total column water vapor **TCWV** in clear-sky regions

- **OLCI** on polar-orbiting satellite Sentinel-3a/3b:
  - high precision 2d TCWV fields
  - high spatial resolution 300x300m^2
  - morning time snapshots (9-10 UTC)
  - processed 2016-2022
- **SEVIRI** on geostationary satellite MSG:
  - lower accuracy timeseries 2d TCWV fields
  - spatial resolution ~3x6km^2
  - every 15 min
  - processed 2016-2022
- **FCI** on geostationary satellite MTG:
  - launched, first (test) data expected soon
  - building retrieval framework NIR+TIR
  - spatial resolution ~1x2 km^2
  - Every 2.5-10 min



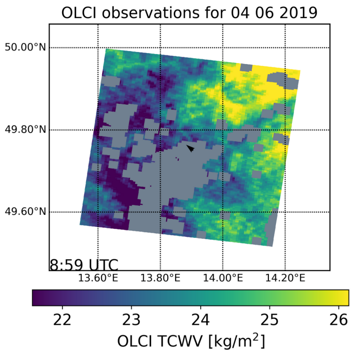


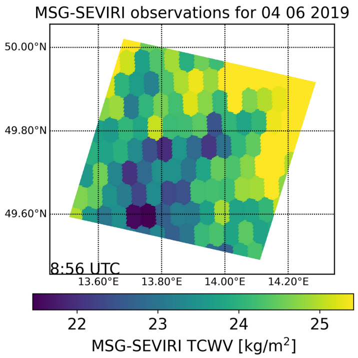


Status on most recent satellite products

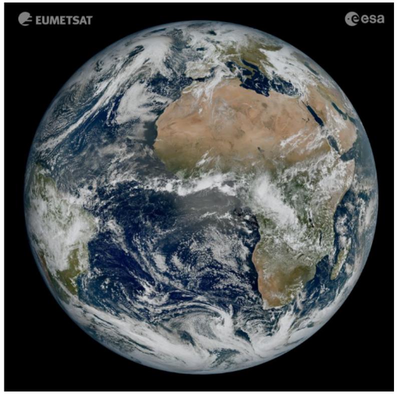
Total column water vapor **TCWV** in clear-sky regions

- **OLCI** on polar-orbiting satellite Sentinel-3a/3b:
  - high precision 2d TCWV fields
  - high spatial resolution 300x300m^2
  - morning time snapshots (9-10 UTC)
  - processed 2016-2022





- **SEVIRI** on geostationary satellite MSG:
  - lower accuracy timeseries 2d TCWV fields
  - spatial resolution ~3x6km^2
  - every 15 min
  - processed 2016-2022



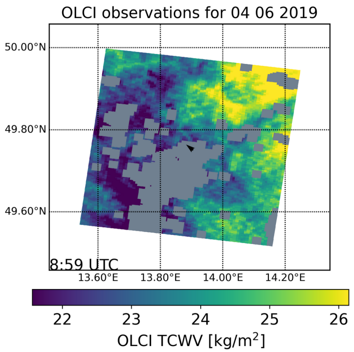
- **FCI** on geostationary satellite MTG:
  - launched, first (test) data expected soon
  - building retrieval framework NIR+TIR
  - spatial resolution ~1x2 km^2
  - Every 2.5-10 min

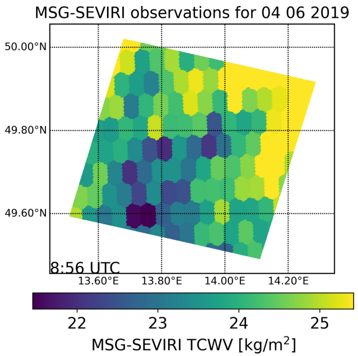
Exploit satellite-based TCWV to advance Convective Initiation (CI) detection

Status on most recent satellite products

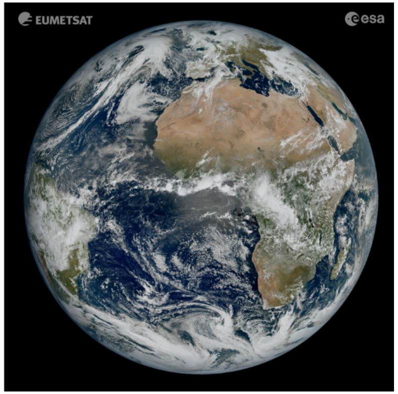
Total column water vapor **TCWV** in clear-sky regions

- **OLCI** on polar-orbiting satellite Sentinel-3a/3b:
  - high precision 2d TCWV fields
  - high spatial resolution 300x300m^2
  - morning time snapshots (9-10 UTC)
  - processed 2016-2022





- **SEVIRI** on geostationary satellite MSG:
  - lower accuracy timeseries 2d TCWV fields
  - spatial resolution ~3x6km^2
  - every 15 min
  - processed 2016-2022



- **FCI** on geostationary satellite MTG:
  - launched, first (test) data expected soon
  - building retrieval framework NIR+TIR
  - spatial resolution ~1x2 km^2
  - Every 2.5-10 min

Exploit satellite-based TCWV to advance Convective Initiation (CI) detection
→ information on potential CI can be used as a proxy
for future new (convective) cells in radar-based precipitation nowcasting

From clear-sky to heavy precipitation

TCWV → clear-sky CI → clouds → cloudy CI → thunderstorm → heavy precipitation

From clear-sky to heavy precipitation

TCWV → clear-sky CI → clouds → cloudy CI → thunderstorm → ~~heavy precipitation~~

From clear-sky to heavy precipitation

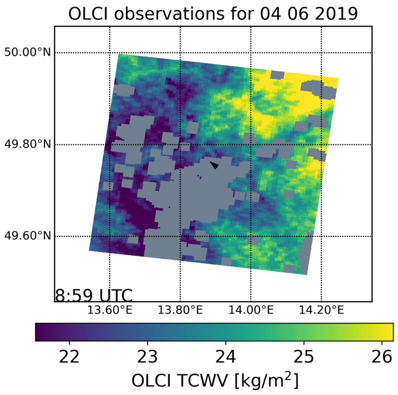
TCWV → clear-sky CI → clouds → cloudy CI → thunderstorm → ~~heavy precipitation~~

Clear-sky (pre-)convective environment → Deep, moist convection (RDT)
(heavy precipitation proxy?)

From clear-sky to heavy precipitation

TCWV → clear-sky CI → clouds → cloudy CI → thunderstorm → ~~heavy precipitation~~

Clear-sky (pre-)convective environment ➡ Deep, moist convection (RDT)
(heavy precipitation proxy?)



9-10 UTC/ 11-12 LT

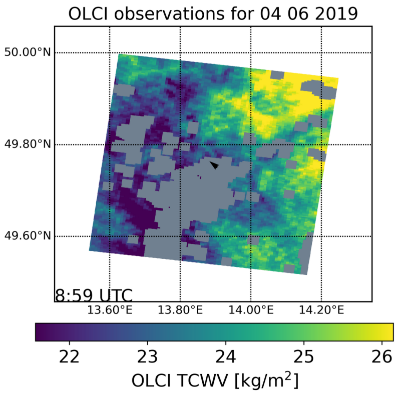
12-13 UTC/ 14-15 LT

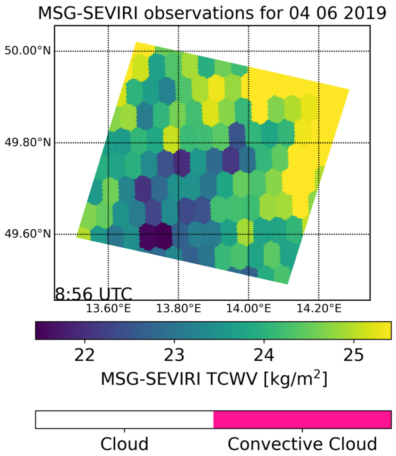
DFG RealPEP

From clear-sky to heavy precipitation

TCWV → clear-sky CI → clouds → cloudy CI → thunderstorm → ~~heavy precipitation~~

Clear-sky (pre-)convective environment ➡ Deep, moist convection (RDT)
(heavy precipitation proxy?)





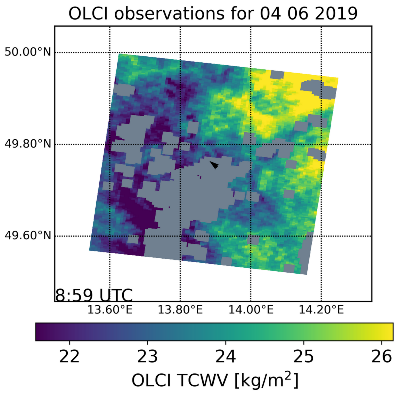
9-10 UTC/ 11-12 LT

From clear-sky to heavy precipitation

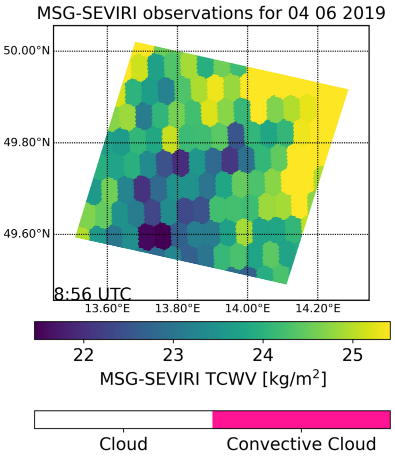
TCWV → clear-sky CI → clouds → cloudy CI → thunderstorm → ~~heavy precipitation~~

Clear-sky (pre-)convective environment ⟶ Deep, moist convection (RDT)
(heavy precipitation proxy?)

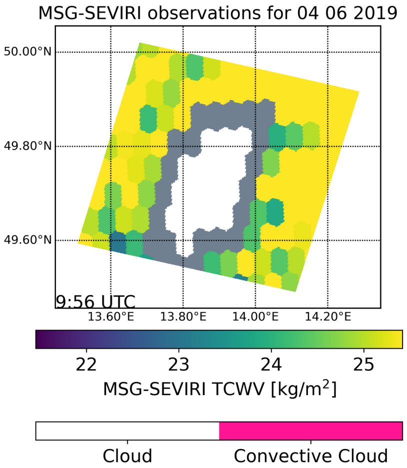
t_0



t_0



t_4



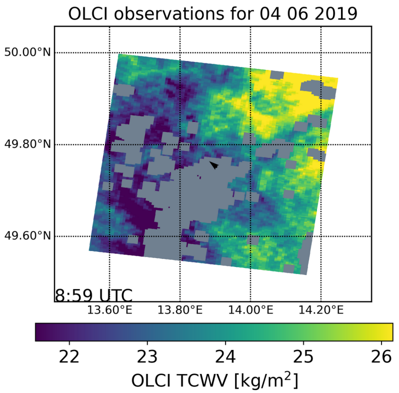
9-10 UTC/ 11-12 LT

From clear-sky to heavy precipitation

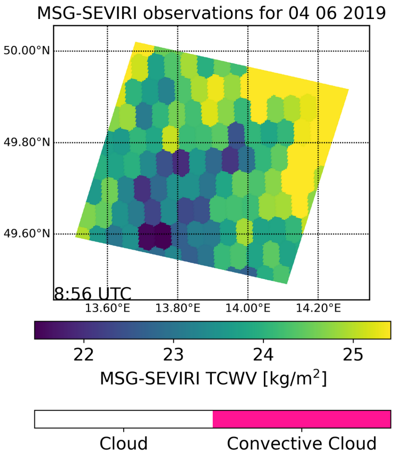
TCWV → clear-sky CI → clouds → cloudy CI → thunderstorm → ~~heavy precipitation~~

Clear-sky (pre-)convective environment → Deep, moist convection (RDT)
(heavy precipitation proxy?)

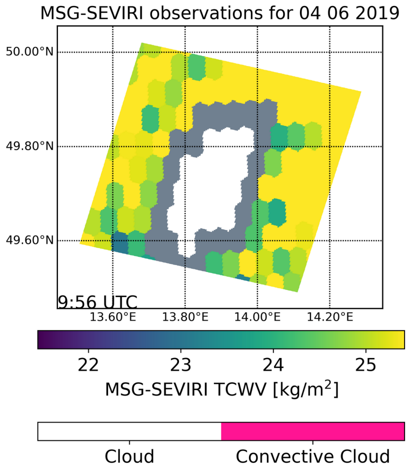
t_0



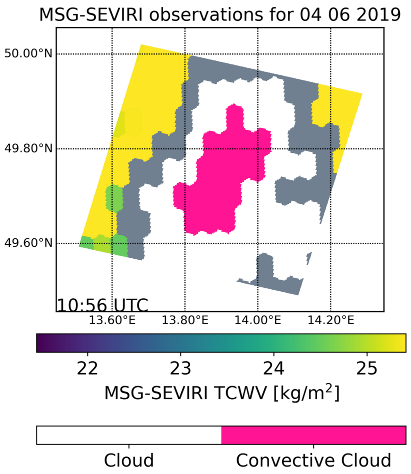
t_0



t_4



t_8



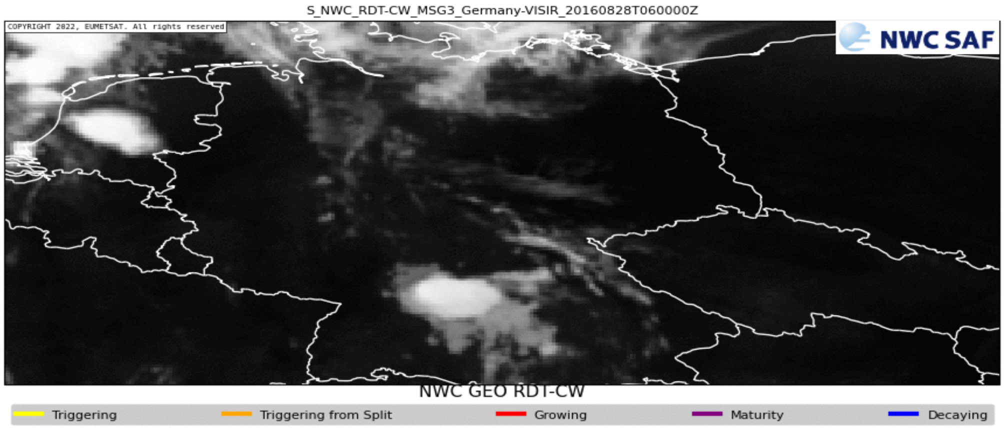
9-10 UTC/ 11-12 LT

12-13 UTC/ 14-15 LT

Convective cloud observations

Why the NWC-SAF RDT product?

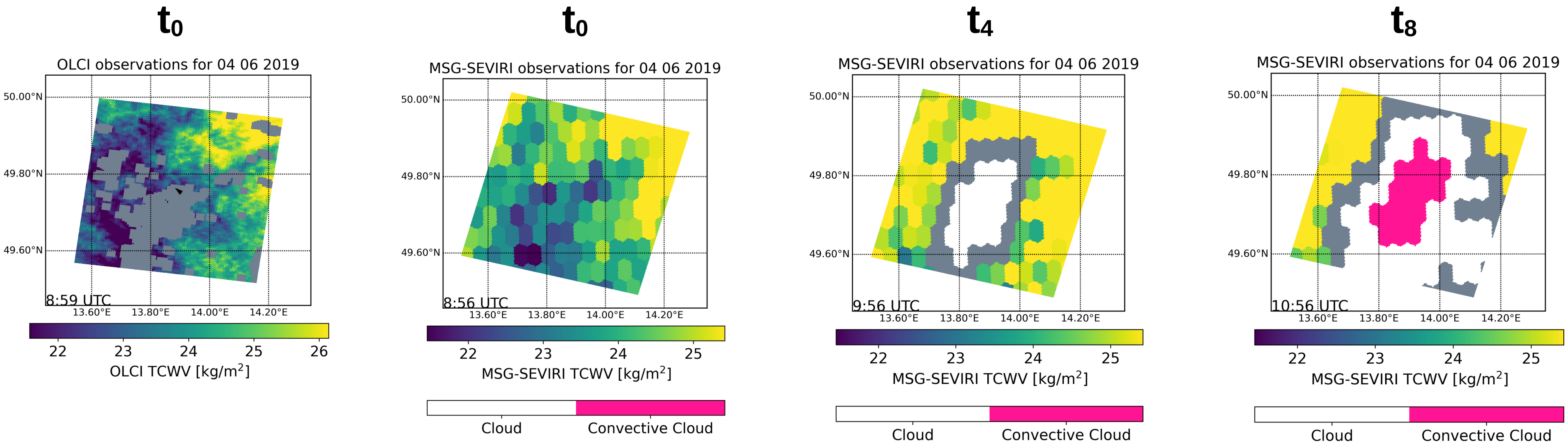
- freely available software, well established and supported
- SEVIRI L1 + ERA5 reanalyses: processed 2016-2022
- detection of (small) convective clouds
- tracking/monitoring of all detected convective systems:
  - life cycle/phase, duration, severity etc.
  - gravity lat/lon since detected as convective cloud



Match-up dataset

Constraints for non-convective/(pre-)convective environments

- clear-sky fraction through most of timeseries (CF > threshold)
- good quality TCWV pixels (outliers, buffer zone)
- no observed convection in hours before
- First observed convective cloud at least one hour after OLCI overpass
- no more than 3 hours after OLCI overpass
- short duration of RDT to avoid looking at advected convective system



9-10 UTC/ 11-12 LT

12-13 UTC/ 14-15 LT

Match-up dateset

Freie Universität Berlin

Constraints for non-convective/(pre-)convective environments

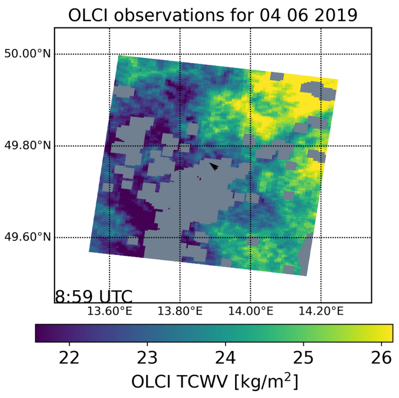
- clear-sky fraction through most of timeseries (CF > threshold)
- good quality TCWV pixels (outliers, buffer zone)
- no observed convection in hours before
- First observed convective cloud at least one hour after OLCI overpass
- no more than 3 hours after OLCI overpass
- short duration of RDT to avoid looking at advected convective system

Cases:

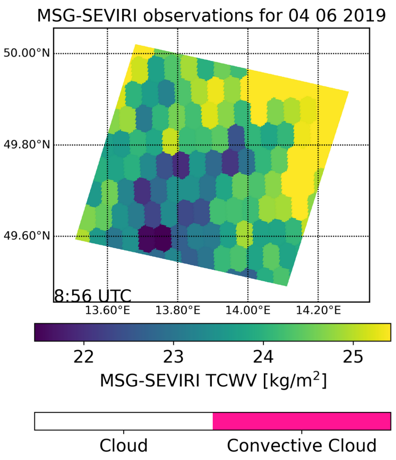
> 1000 non-convective/no RDT

~ 100 (pre-) convective/RDT

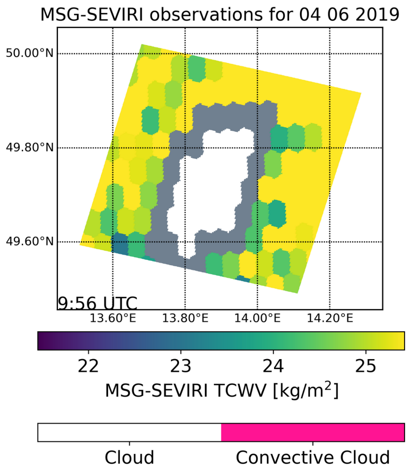
t_0



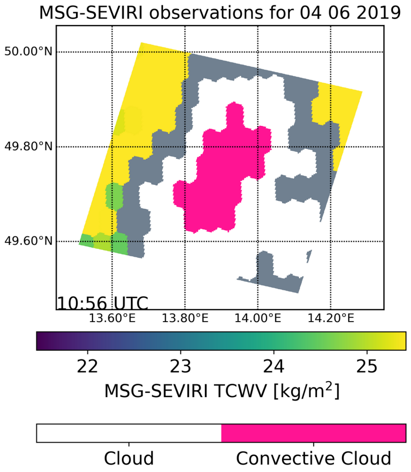
t_0



t_4



t_8



9-10 UTC/ 11-12 LT

12-13 UTC/ 14-15 LT

DFG RealPEP

TCWV-based metrics

Freie
Universität
Berlin

Characterizing non/pre-convective environments

Possible features/predictors for classification/prediction

Amount of TCWV (OLCI+SEVIRI):

- TCWV mean, std, percentiles (10, 50, 90)
- TCWV anomaly: mean, std, percentiles (10, 50, 90)

DFG RealPEP

TCWV-based metrics

Freie Universität Berlin

Characterizing non/pre-convective environments

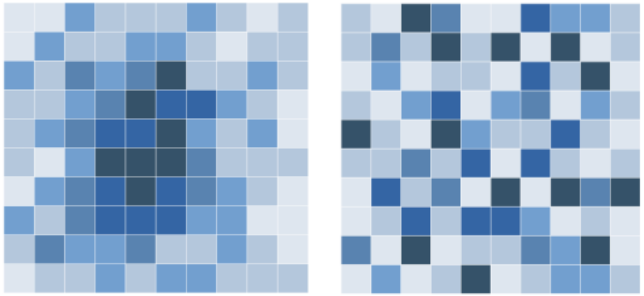
Possible features/predictors for classification/prediction

Amount of TCWV (OLCI+SEVIRI):

- TCWV mean, std, percentiles (10, 50, 90)
- TCWV anomaly: mean, std, percentiles (10, 50, 90)

TCWV spatial information (OLCI):
spatial auto-correlation and texture measures

- Grey Level Co-Occurrence Matrices
- Gradient
- Etc.



Measures of contrast, homogeneity/ correlation, orderliness:

- Varying pixel distances
- perpendicular/parallel to average BL wind direction (assymetry factor)

DFG RealPEP

TCWV-based metrics

Characterizing non/pre-convective environments

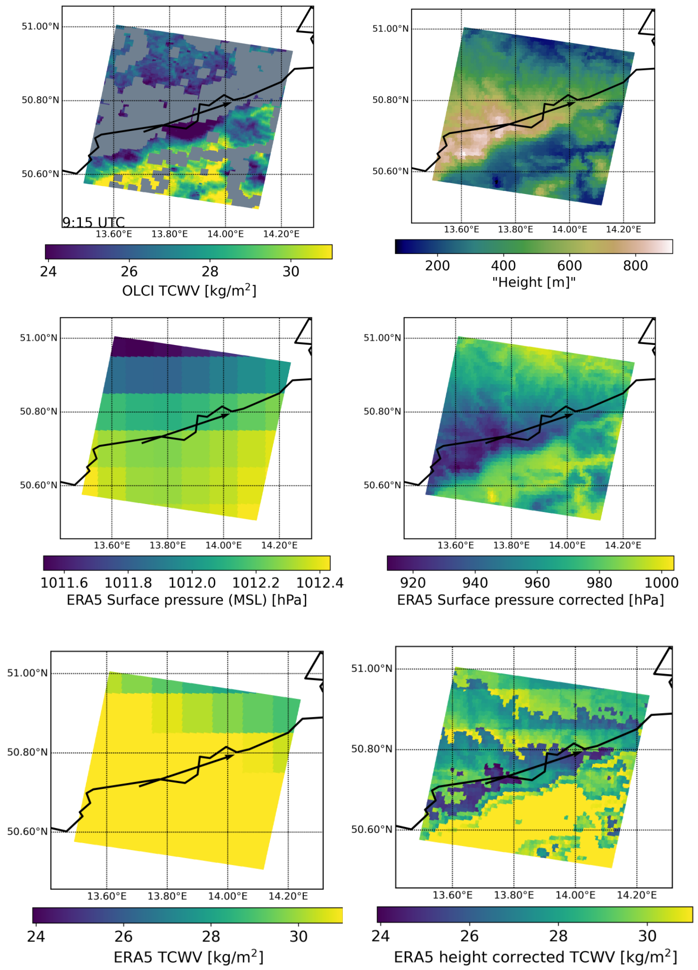
Possible features/predictors for classification/prediction

Amount of TCWV (OLCI+SEVIRI):
- TCWV mean, std, percentiles (10, 50, 90)
- TCWV anomaly: mean, std, percentiles (10, 50, 90)

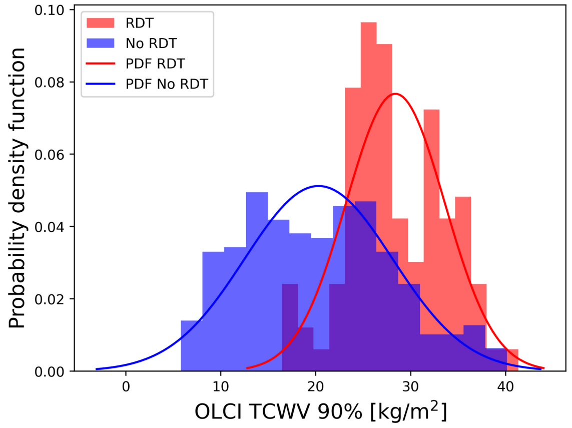
TCWV spatial information (OLCI):
spatial auto-correlation and texture measures
- Grey Level Co-Occurrence Matrices
- Gradient
- Etc.

TCWV temporal information (SEVIRI):
- Trends: SMA, CMA, relative, …
- Jumps

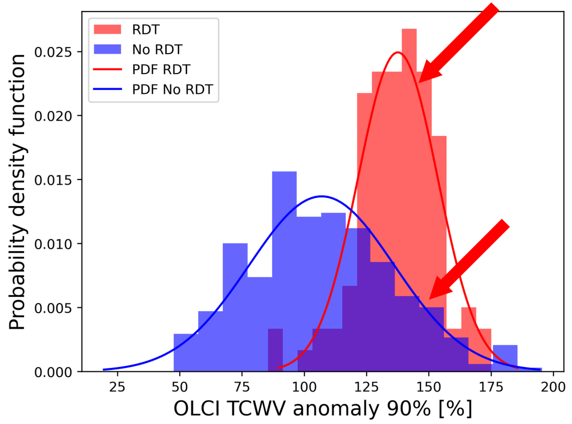
Model parameters (ERA5)
- T and T_{dew}
- Difference OLCI TCWV & ERA $TCWV_{corrected}$

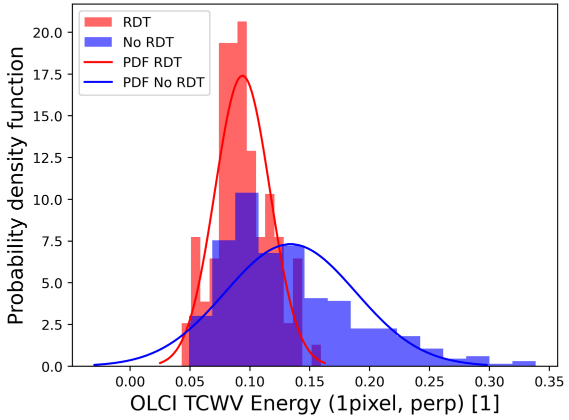


Likelihood functions/ PDFs

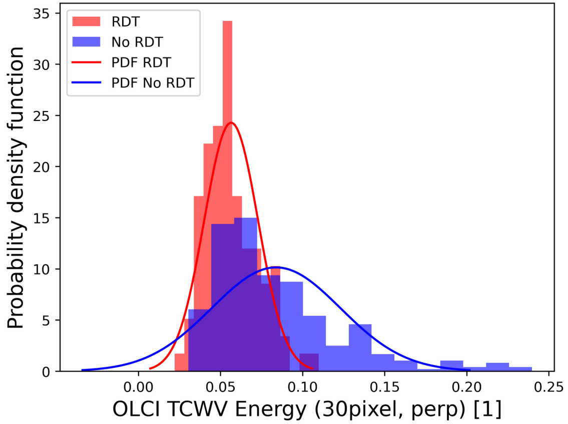


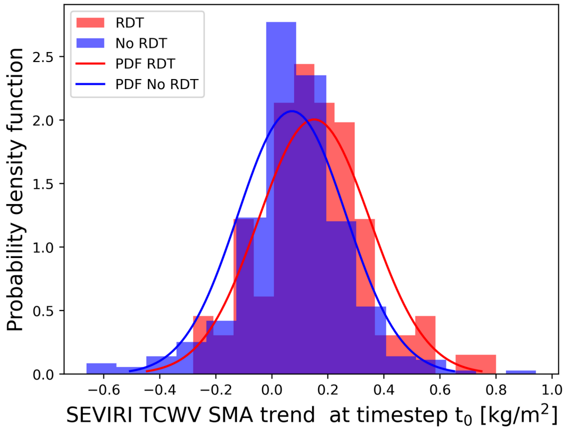
TCWV amount



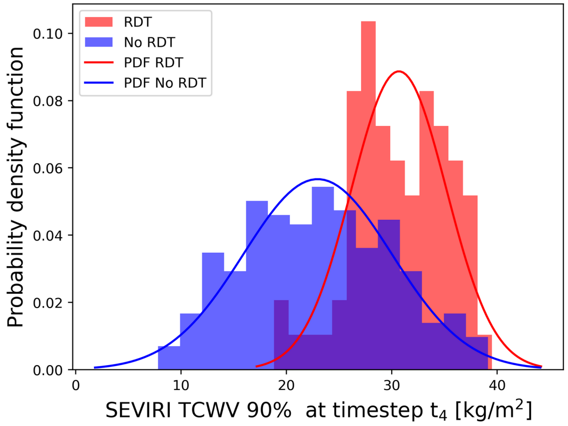


TCWV spatial information





TCWV temporal information



Conditional probabilities

I have **conditional probabilities/likelihoods** (from relative frequencies):

P(Data | RDT) = what is the probability of seeing Data if RDT development later on

Using a small set of features/predictors I want to compute:

P(RDT | Data) = what is the probability of observing RDT development later on given Data

Conditional probabilities

I have **conditional probabilities/likelihoods** (from relative frequencies):

P(Data | RDT) = what is the probability of seeing Data if RDT later on

Using a small set of features/predictors I want to compute:

P(RDT | Data) = what is the probability of observing RDT later on given Data

Use **Bayesian Theorem/Bayes'rule** to get from P(B|A) to P(A|B)

Powerful machine-learning classification tool, simple implementation, fast

Naive Bayesian approach

Chain rule of conditional probabilities
Features are assumed to be conditionally independent
D: Data

- **D1**: Amount of TCWV
- **D2**: TCWV spatial information
- **D3**: TCWV temporal information

H: Hypothesis RDT occurrence within X hours

$$P(H|D_1 \cap D_2 \cap D_3) = \frac{P(D_1|H) * P(D_2|H) * P(D_3|H) * P(H)}{P(D_1) * P(D_2) * P(D_3)}$$

Naive Bayesian approach

Freie Universität Berlin

Chain rule of conditional probabilities
Features are assumed to be conditionally independent
D: Data

- **D1**: Amount of TCWV
- **D2**: TCWV spatial information
- **D3**: TCWV temporal information

H: Hypothesis RDT occurrence within X hours

$$P(H|D_1 \cap D_2 \cap D_3) = \frac{P(D_1|H) * P(D_2|H) * P(D_3|H) * P(H)}{P(D_1) * P(D_2) * P(D_3)}$$

Bayesian update:
Posterior P at timestep *t* becomes prior P at timestep *t+1*

DFG RealPEP

Bayesian framework wrapper

For ff in feature_combinations:

> 1000 x

- select random feature set
- 1 TCWV amount, 1 TCWV spatial info, 1 TCWV temporal info

Bayesian framework wrapper

Freie Universität Berlin

For ff in feature_combinations:

> 1000 x
- select random feature set
- 1 TCWV amount, 1 TCWV spatial info, 1 TCWV temporal info

K-fold cross-validation
- divide samples in k groups to estimate skill of model on new data
- here dataset (1000+100) into 10 parts

10 x

For k in range(9):
- 9/10 for training → read Likelihood Tables for each SEVIRI time step

DFG RealPEP

Bayesian framework wrapper

> 1000 x **For ff in feature_combinations:**

- select random feature set
- 1 TCWV amount, 1 TCWV spatial info, 1 TCWV temporal info

K-fold cross-validation

- divide samples in k groups to estimate skill of model on new data
- here dataset (1000+100) into 10 parts

10 x **For k in range(9):**

- 9/10 for training → read Likelihood Tables for each SEVIRI time step

> 100 x **For dd in test_data:**

- 1/10 for testing

~ 4-12 x **For tt in seviri_timeseries:**

Bayesian framework

- read in prior and likelihood and apply Bayesian update
- store posterior probability timeseries

DFG RealPEP

Bayesian framework wrapper

For ff in feature_combinations:

- select random feature set
- 1 TCWV amount, 1 TCWV spatial info, 1 TCWV temporal info

> 1000 x

K-fold cross-validation

- divide samples in k groups to estimate skill of model on new data
- here dataset (1000+100) into 10 parts

10 x

For k in range(9):

- 9/10 for training → compute Likelihood functions for each SEVIRI time step

> 100 x

For dd in test_data:

- 1/10 for testing

~ 4-12 x

For tt in seviri_timeseries:

Bayesian framework

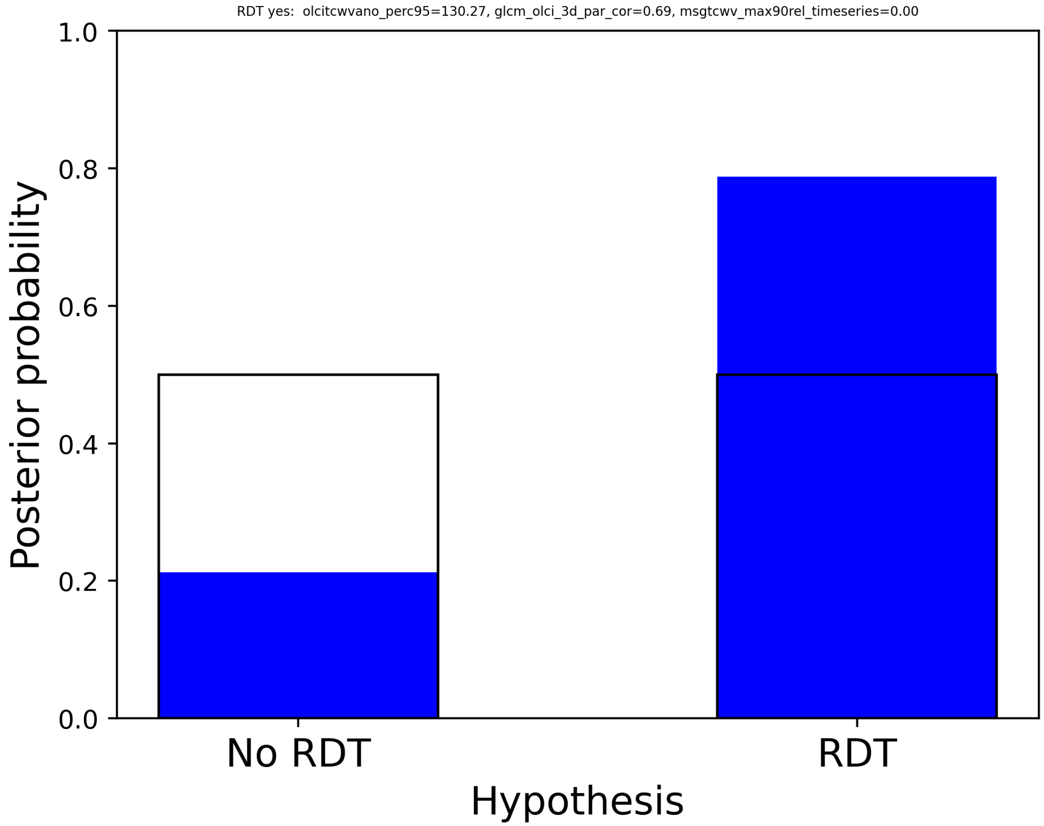
- read in prior and likelihood and apply Bayesian update
- store posterior probability timeseries

Compute and save skill scores for test data

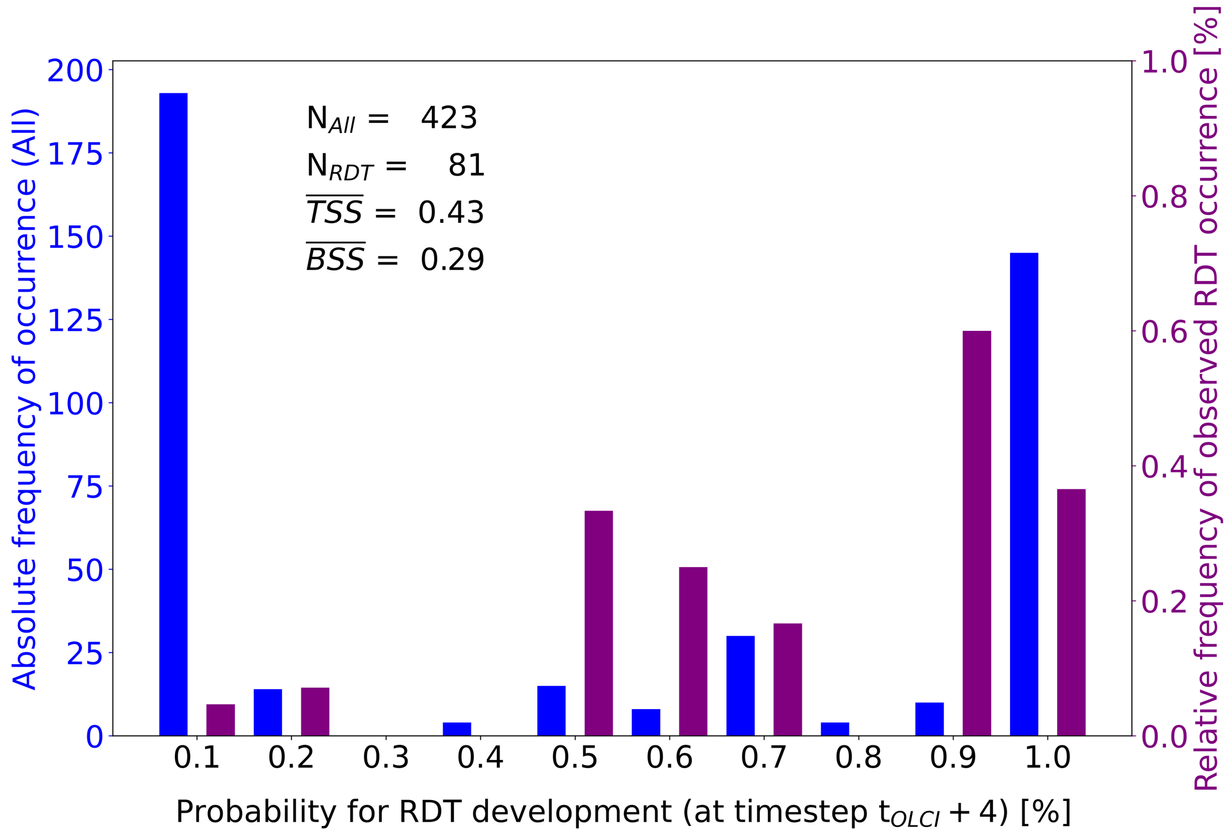
Compute and save mean skill scores for feature set

Assessment of best feature set performance

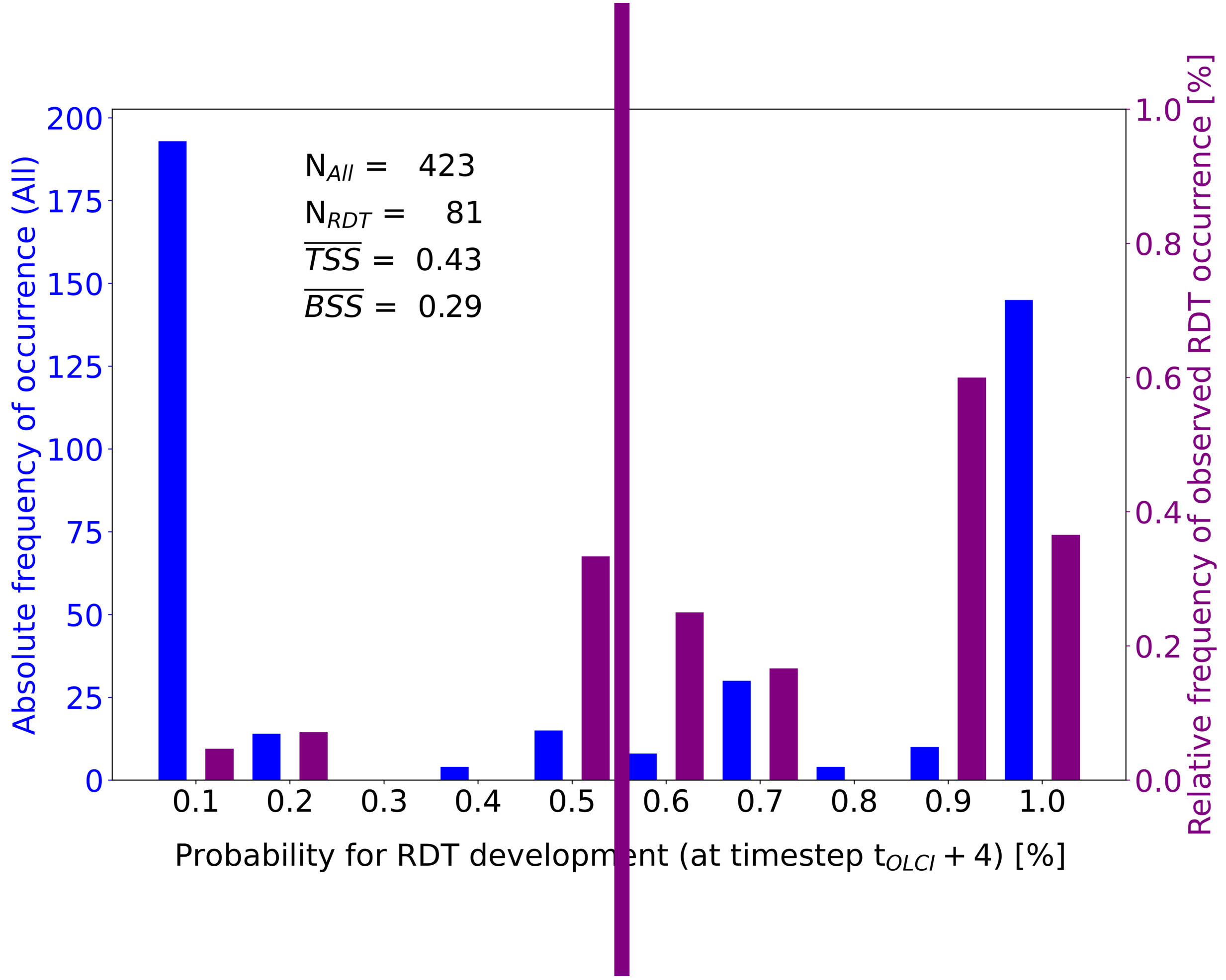
Posterior probability timeseries



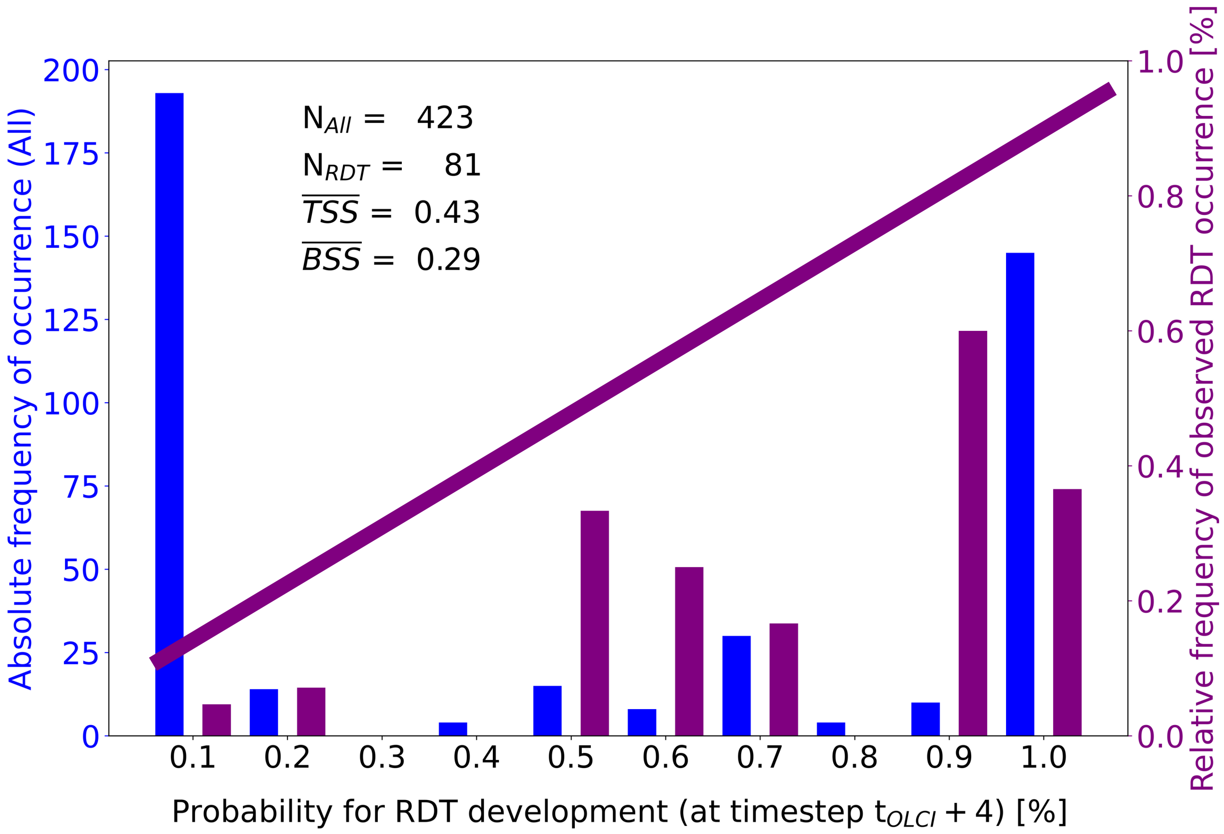
Evaluation example



Evaluation example



Evaluation example



Current status & Outlook

Freie Universität Berlin

Reconsidering constraints for match-up dataset
- Unbalanced dataset
- RDT vs. CI, new NWCSAF software version
- independent a-priori information (ERA5 stability indices)

Extend feature set
- Difference ERA5 TCWV forecast (height corrected) and OLCI TCWV
- Measure of (relative) water vapor amount in BL?
- Apply Kernel density estimator for PDFs

Evaluation performance
- Classification vs. probabilities
- Skill scores

Towards (Pred)RNN
- Match-up dataset of timeseries of TCWV fields + convective cloud information + metrics
- Python modules
- Merge satellite data with QPE data

DFG RealPEP

Thank you!