Karlsruhe Institute of Technology

KIT-Campus Alpin
IMK-IFU: Atmosphärische Umweltforschung

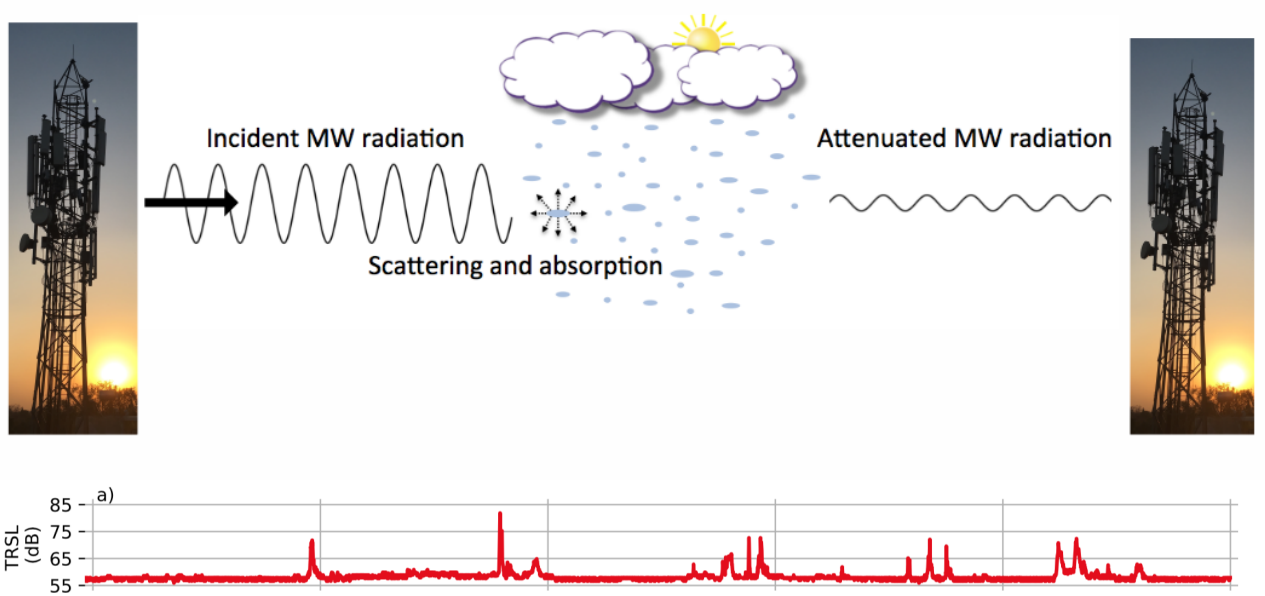
Big commercial microwave link data

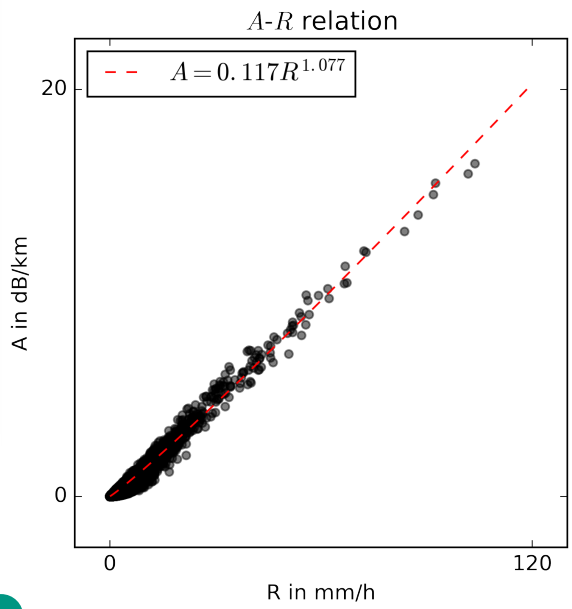
Detecting rain events with deep learning

Julius Polz, Christian Chwala, Maximilian Graf, Harald Kunstmann | May 1, 2020

KIT INSTITUTE OF ATMOSPHERIC ENVIRONMENTAL RESEARCH (IMK-IFU)

The theory





Commercial microwave links (CMLs)...

... are part of the cellular network.

→ Specific attenuation *A* is in a close to linear relationship with the path averaged rain rate *R*.

→ Monitoring of transmitted minus received signal level (TRSL) allows for rain rate estimation along the link path.

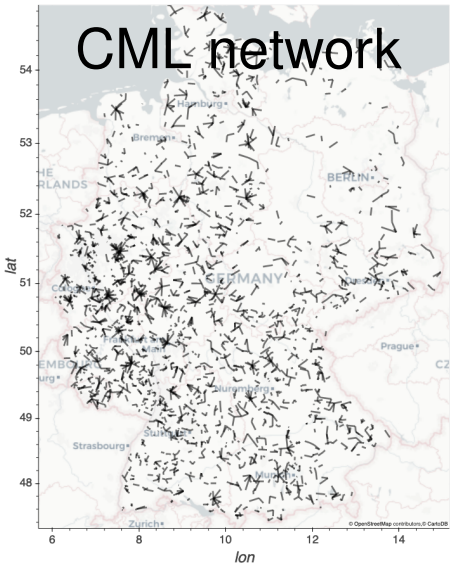
[mm/h]

A-R power law: $A = a\,R^b$

[dB/km]

Chwala & Kunstmann, 2019, WIRES Water

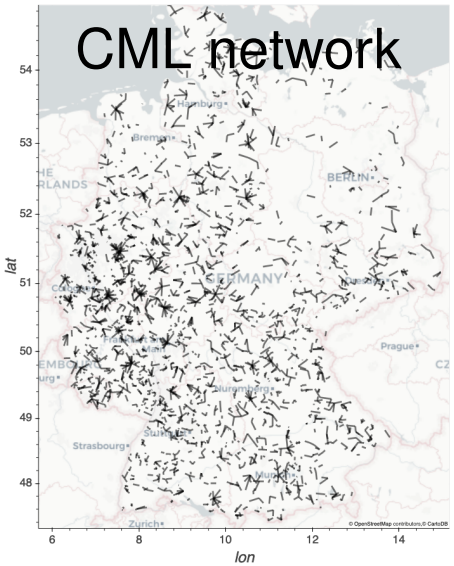
Why big data?



Spatial distribution

3904 CMLs all over Germany

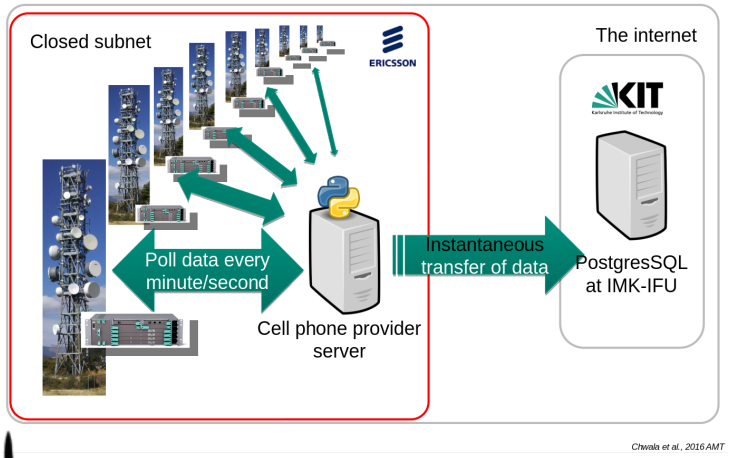
Why big data?



Spatial distribution

3904 CMLs all over Germany

Data acquisition

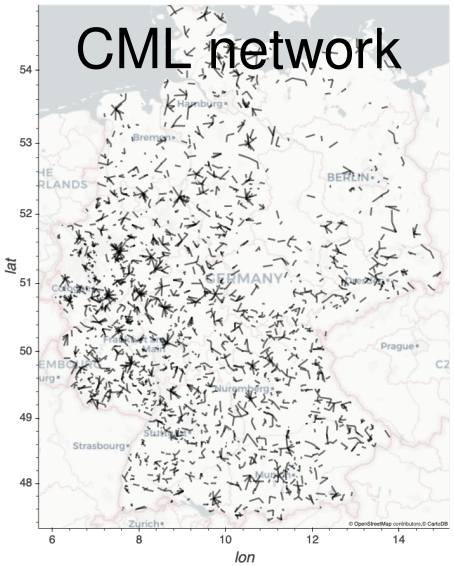


Temporal resolution and data availability

1 Minute resolution for 3 years at 97% availability

Why big data?

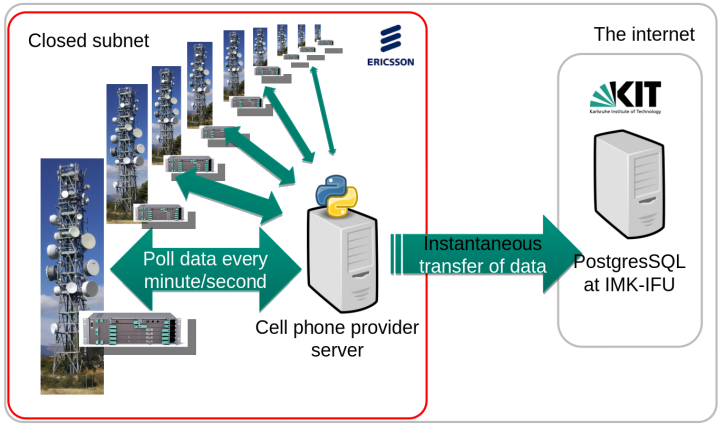
CML network



Spatial distribution

3904 CMLs all over Germany

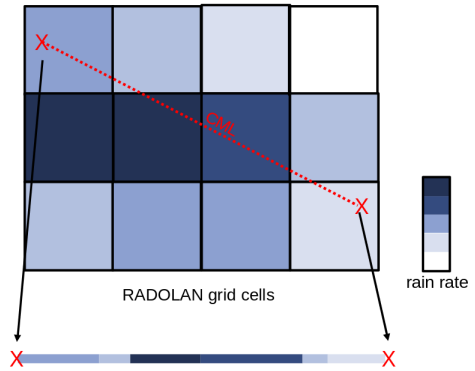
Data acquisition



Temporal resolution and data availability

1 Minute resolution for 3 years at 97% availability

Reference



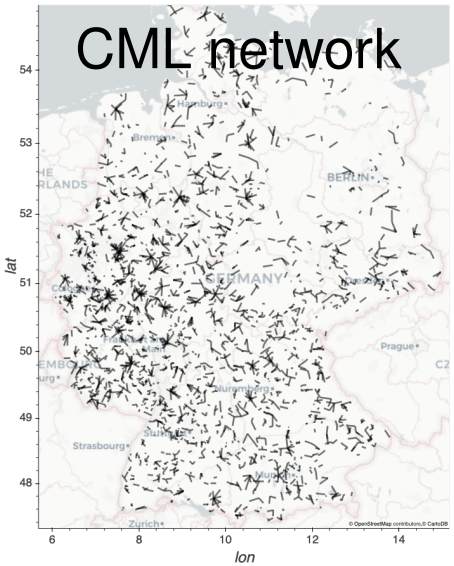
Path averaged rain rate by weighted length of intersects

Temporal aggregation for validation

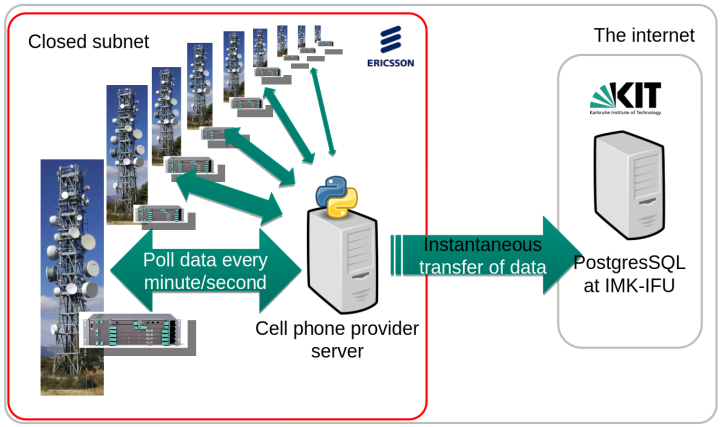
RADOLAN RW by DWD: Hourly gauge adjusted radar rainfall

Why big data?

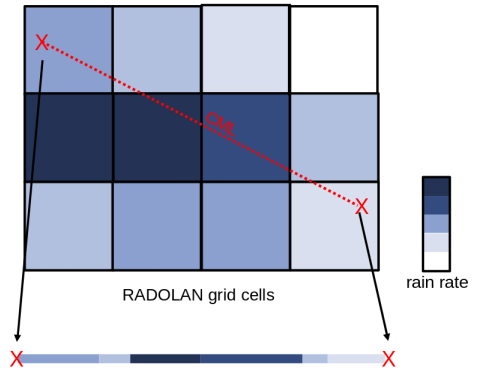
CML network



Data acquisition



Reference



Path averaged rain rate by weighted length of intersects

Spatial distribution

3904 CMLs all over Germany

Temporal resolution and data availability

1 Minute resolution for 3 years at 97% availability

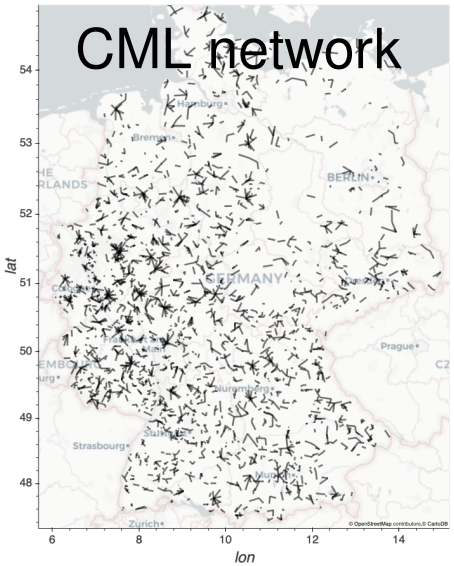
Temporal aggregation for validation

RADOLAN RW by DWD: Hourly gauge adjusted radar rainfall

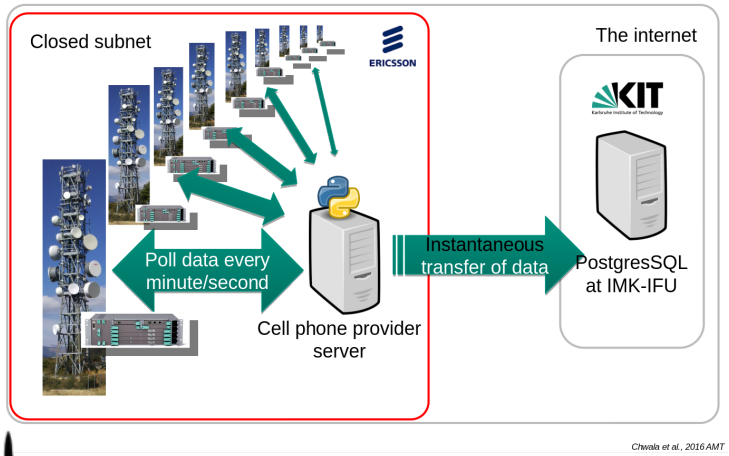
100.000 hours of CML time series

Why big data?

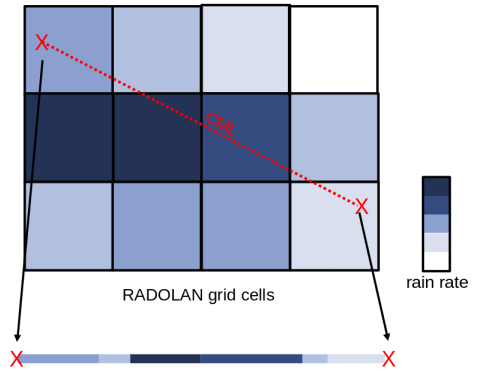
CML network



Data acquisition



Reference



Path averaged rain rate by weighted length of intersects

Spatial distribution

3904 CMLs all over Germany

Temporal resolution and data availability

1 Minute resolution for 3 years at 97% availability

Temporal aggregation for validation

RADOLAN RW by DWD: Hourly gauge adjusted radar rainfall

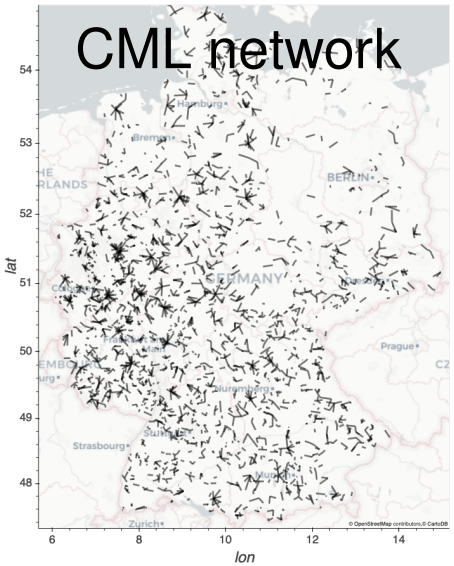
100.000 hours of CML time series

+

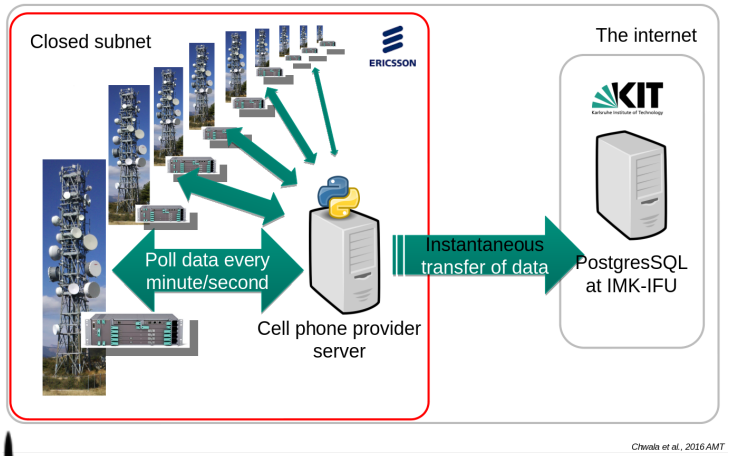
High variablity in data quality

Why big data?

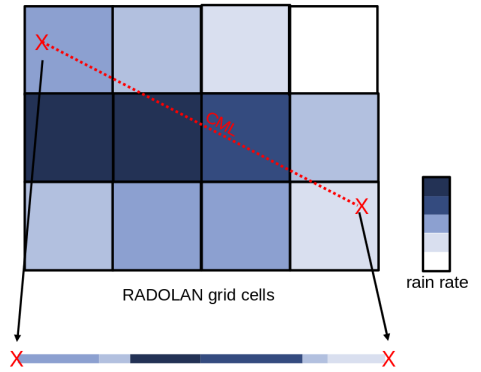
CML network



Data acquisition



Reference



Path averaged rain rate by weighted length of intersects

Spatial distribution

3904 CMLs all over Germany

Temporal resolution and data availability

1 Minute resolution for 3 years at 97% availability

Temporal aggregation for validation

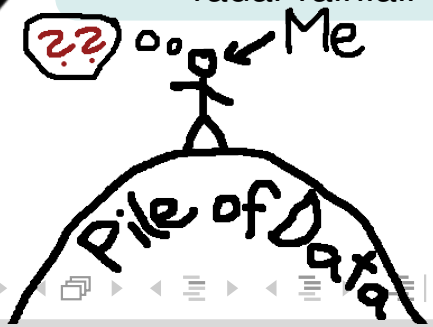
RADOLAN RW by DWD: Hourly gauge adjusted radar rainfall

100.000 hours of CML time series

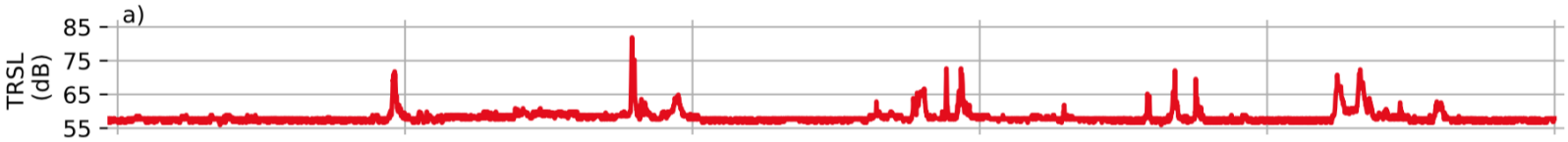
+

High variablity in data quality

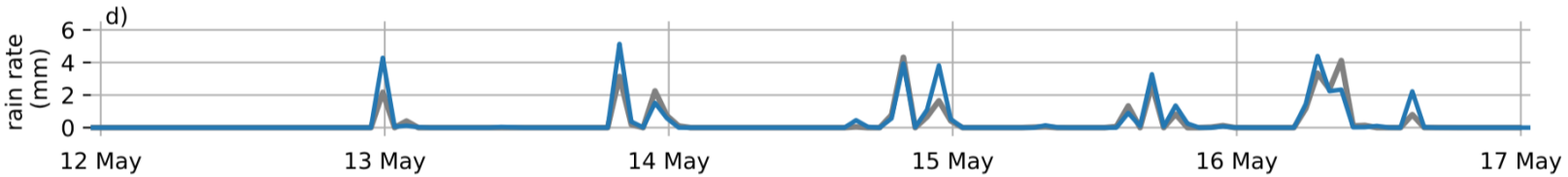
=



The processing



- ↓ Remove erroneous data
- ↓ Detect rain events
- ↓ Calculate attenuation from baseline level
- ↓ Compensate for wet antenna attenuation
- ↓ Derive rain rate

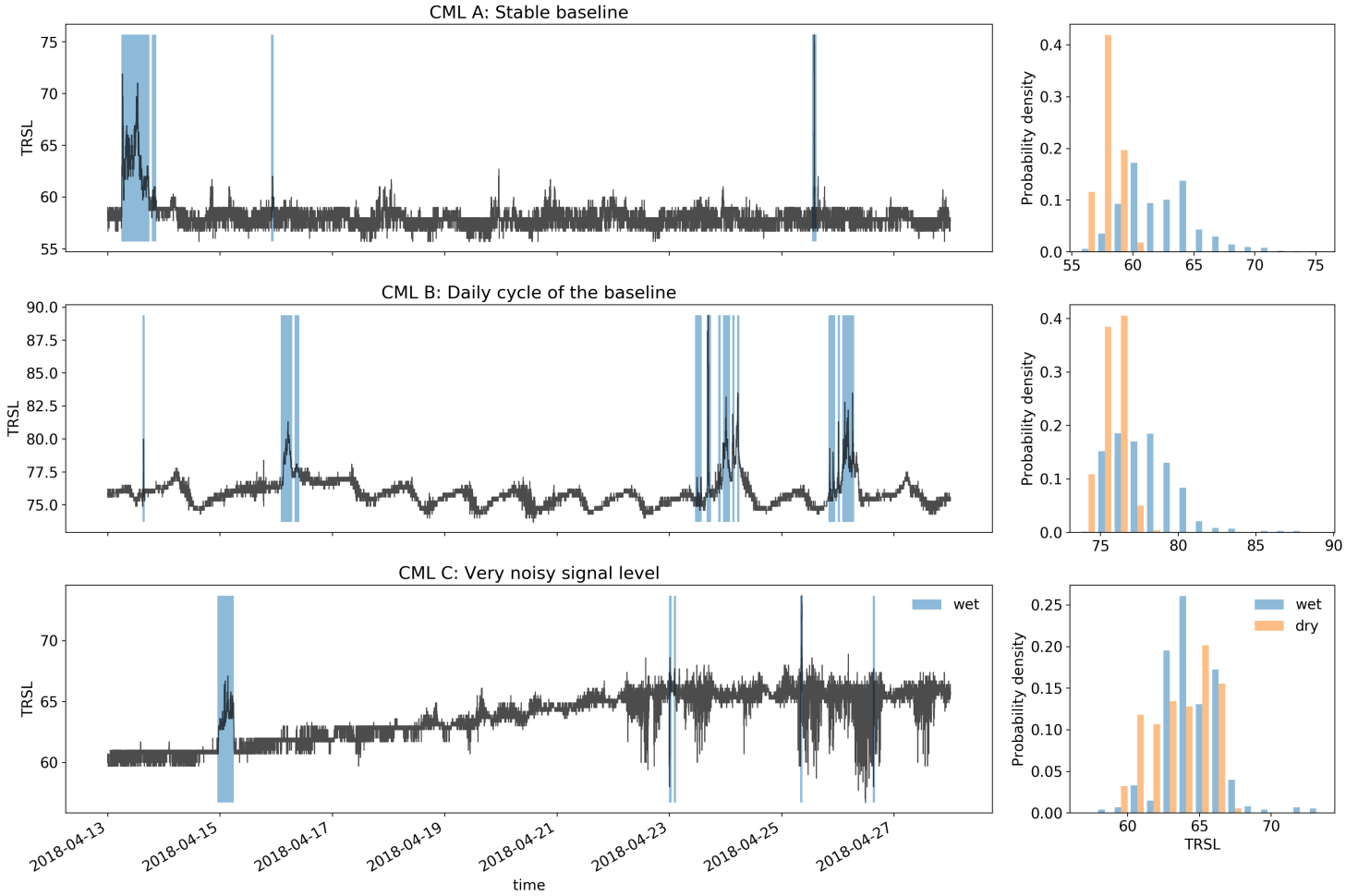


The challenge

CML signal levels can behave very differently and large fluctuations appear even during dry periods.

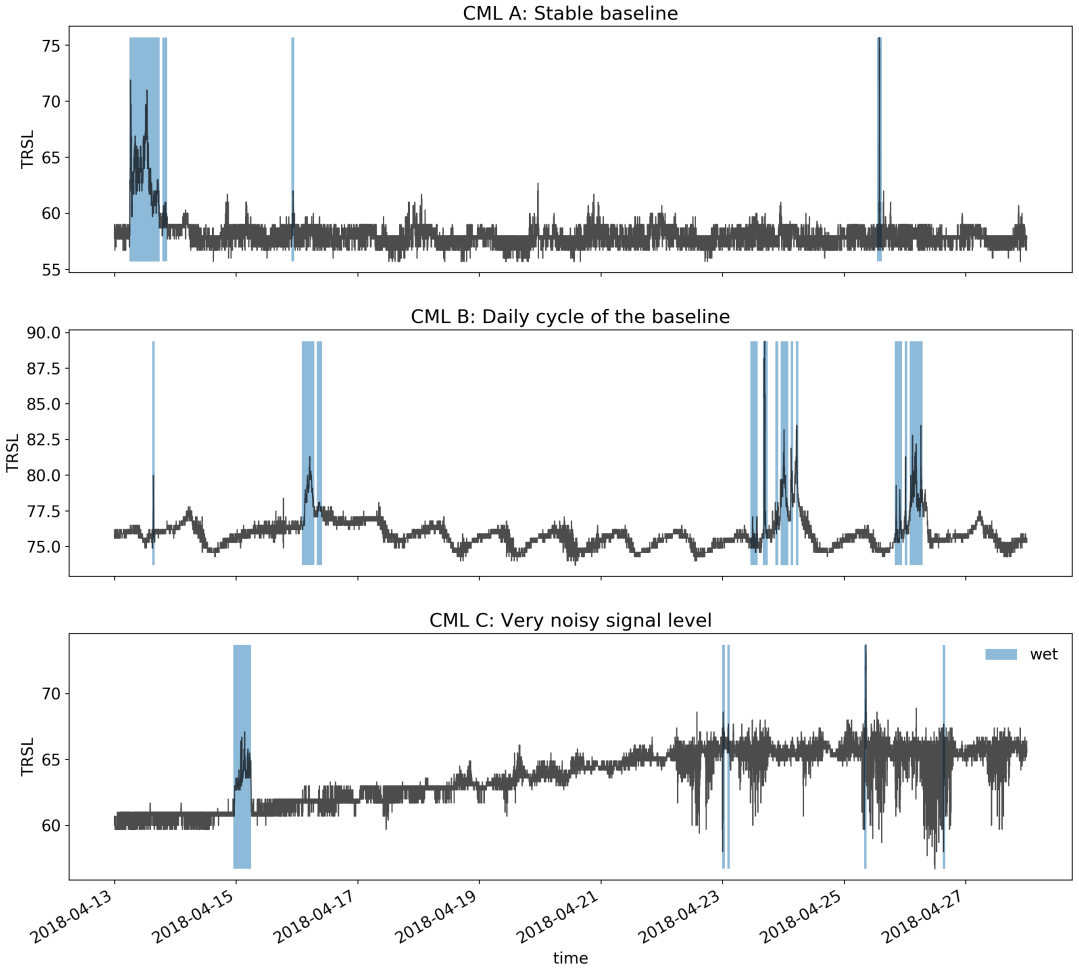
The challenge

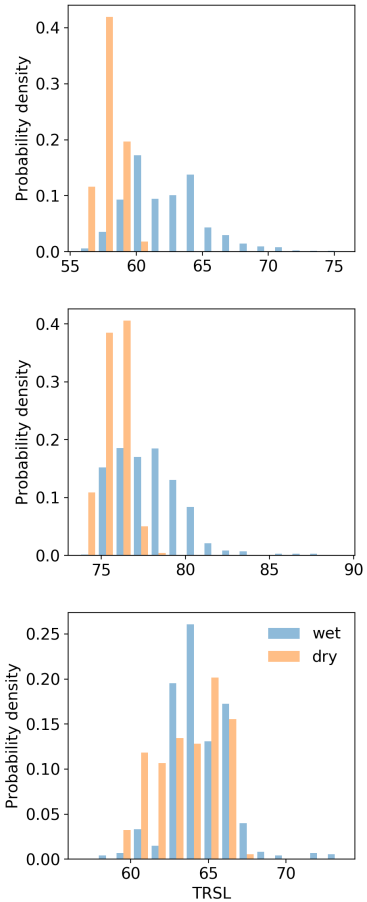
CML signal levels can behave very differently and large fluctuations appear even during dry periods.



The challenge

CML signal levels can behave very differently and large fluctuations appear even during dry periods.

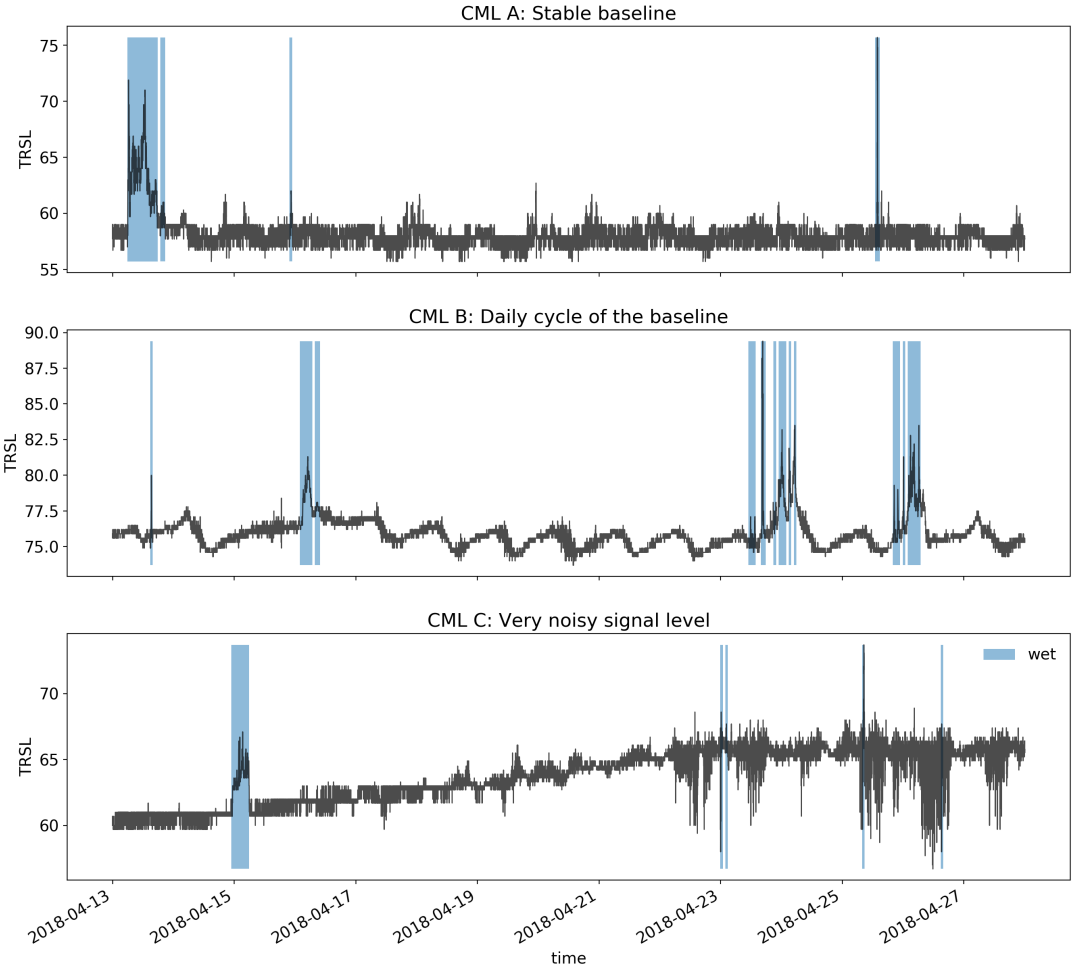


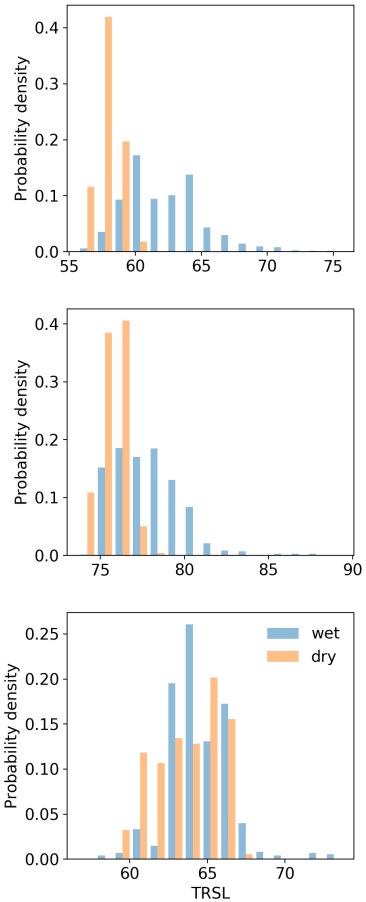


← Wet and dry periods nearly separable by TRSL values.

The challenge

CML signal levels can behave very differently and large fluctuations appear even during dry periods.



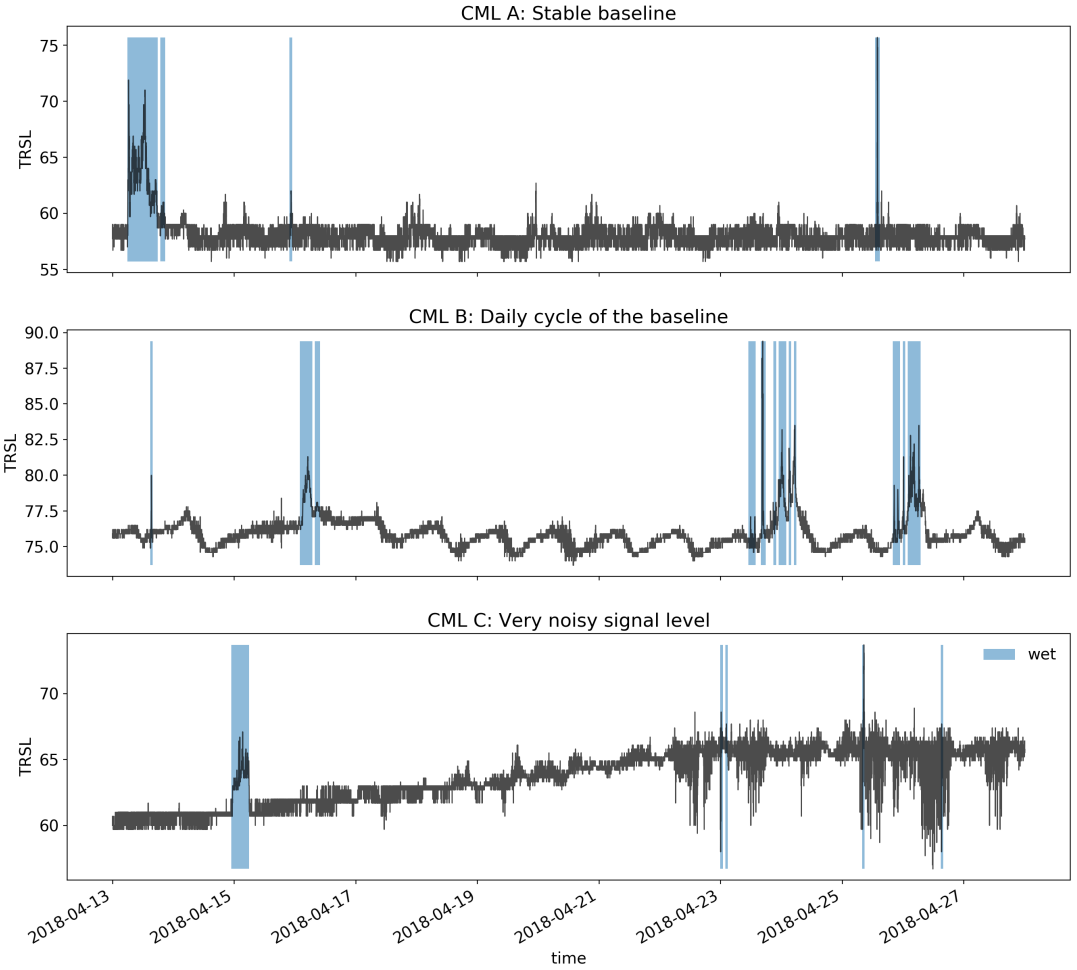


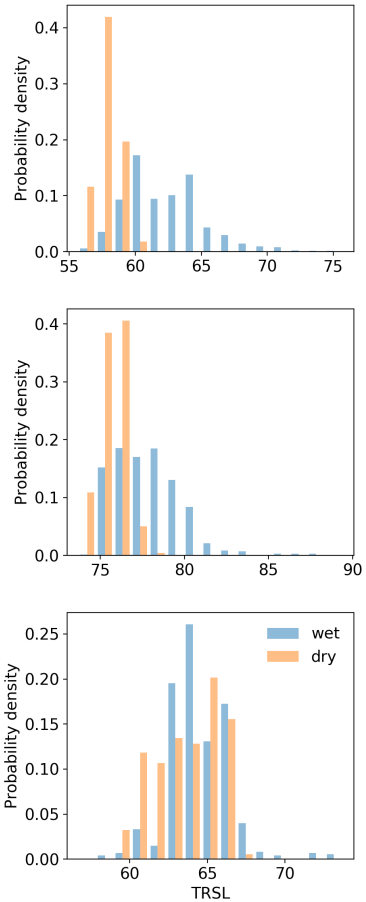
← Wet and dry periods nearly separable by TRSL values.

← Challenging daily fluctuations, but clearly visible rain events.

The challenge

CML signal levels can behave very differently and large fluctuations appear even during dry periods.





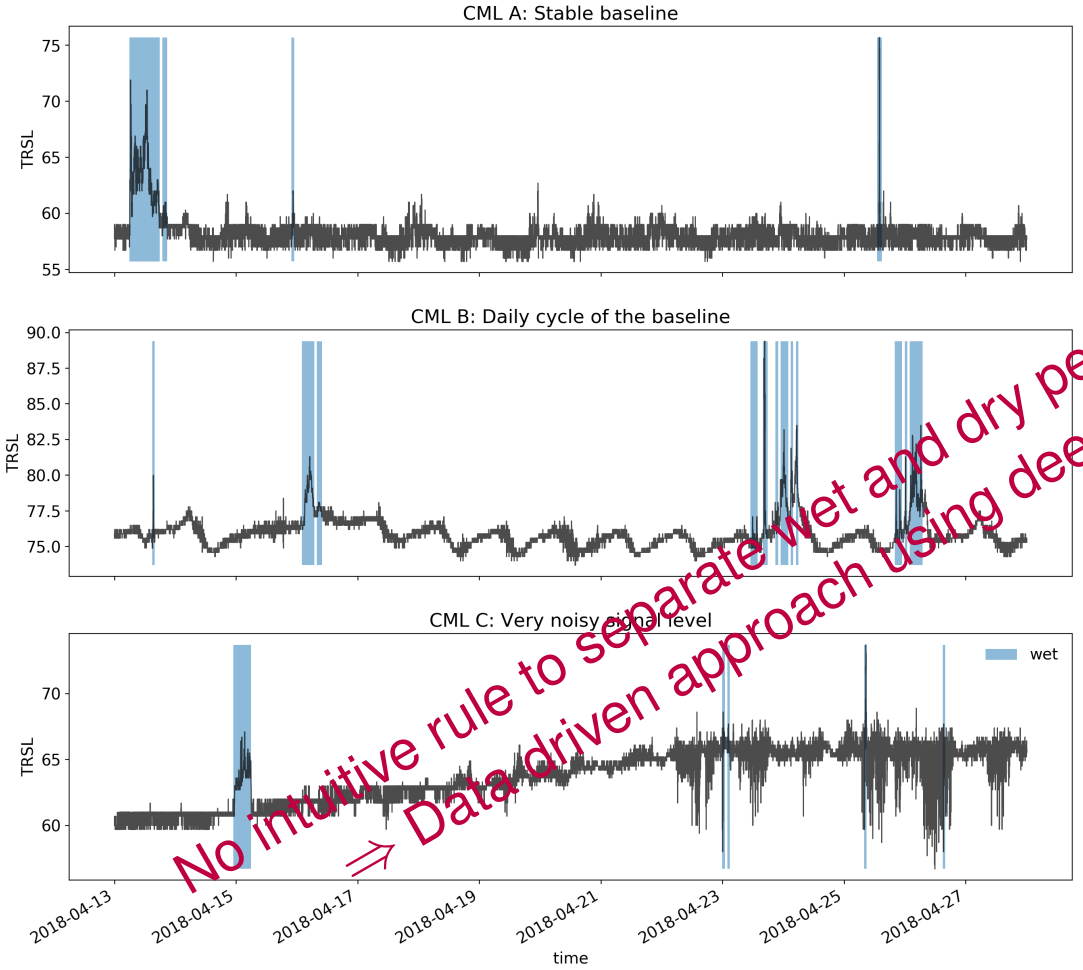
← Wet and dry periods nearly separable by TRSL values.

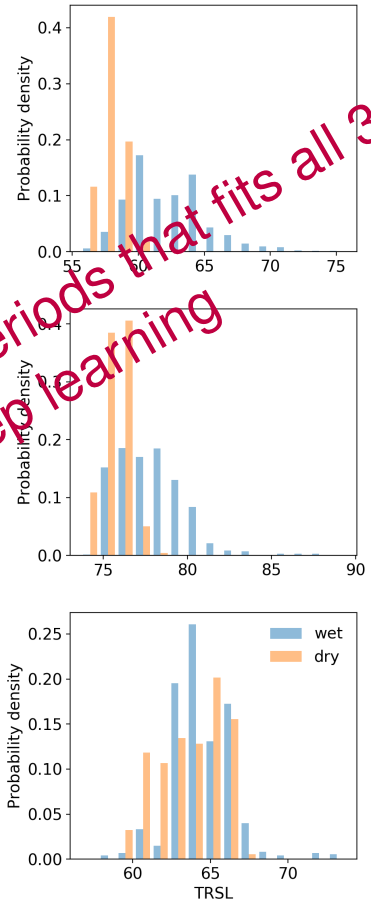
← Challenging daily fluctuations, but clearly visible rain events.

← Nearly arbitrary changes from clearly visible events to extremely noisy periods.

The challenge

CML signal levels can behave very differently and large fluctuations appear even during dry periods.





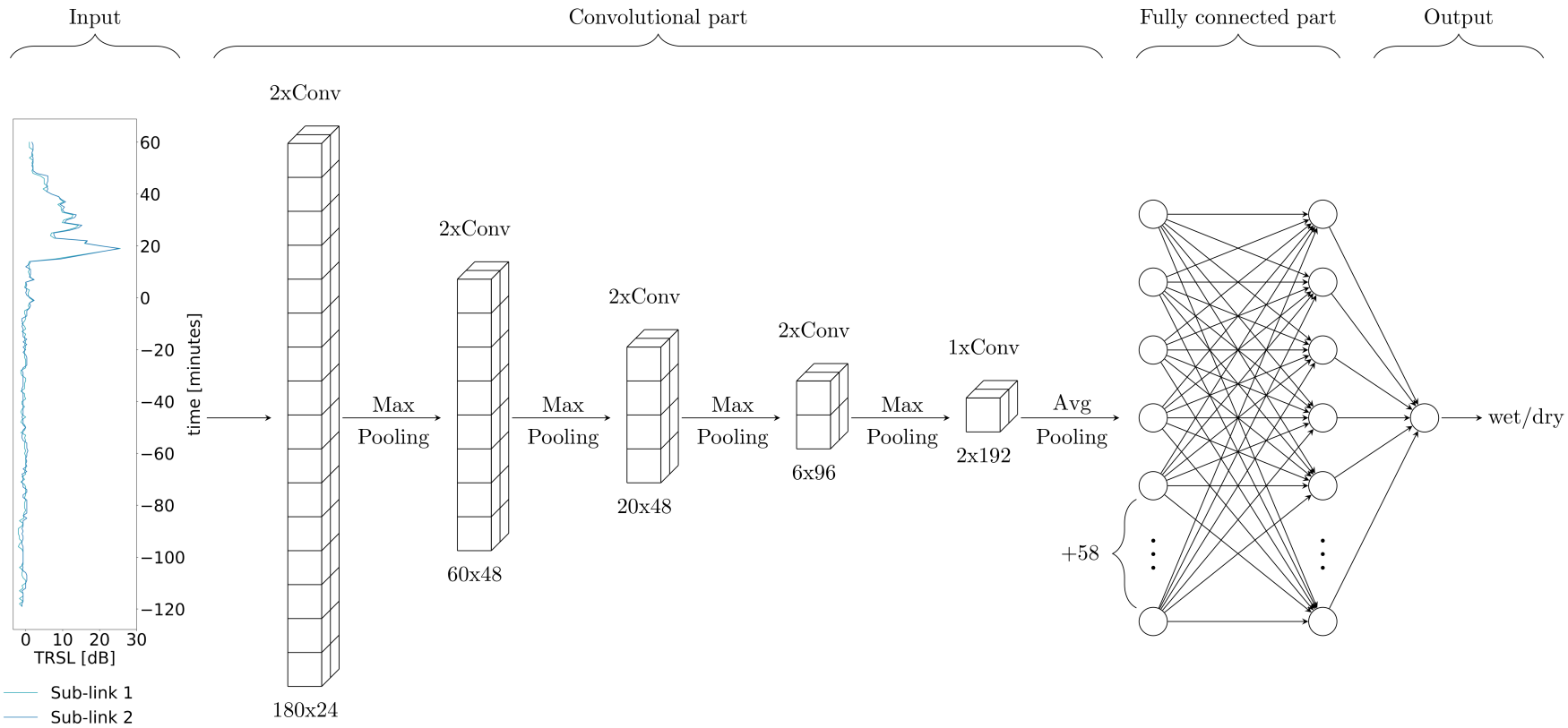
← Wet and dry periods nearly separable by TRSL values.

← Challenging daily fluctuations, but clearly visible rain events.

← Nearly arbitrary changes from clearly visible events to extremely noisy periods.

No intuitive rule to separate wet and dry periods that fits all 3904 CMLs...
⇒ Data driven approach using deep learning

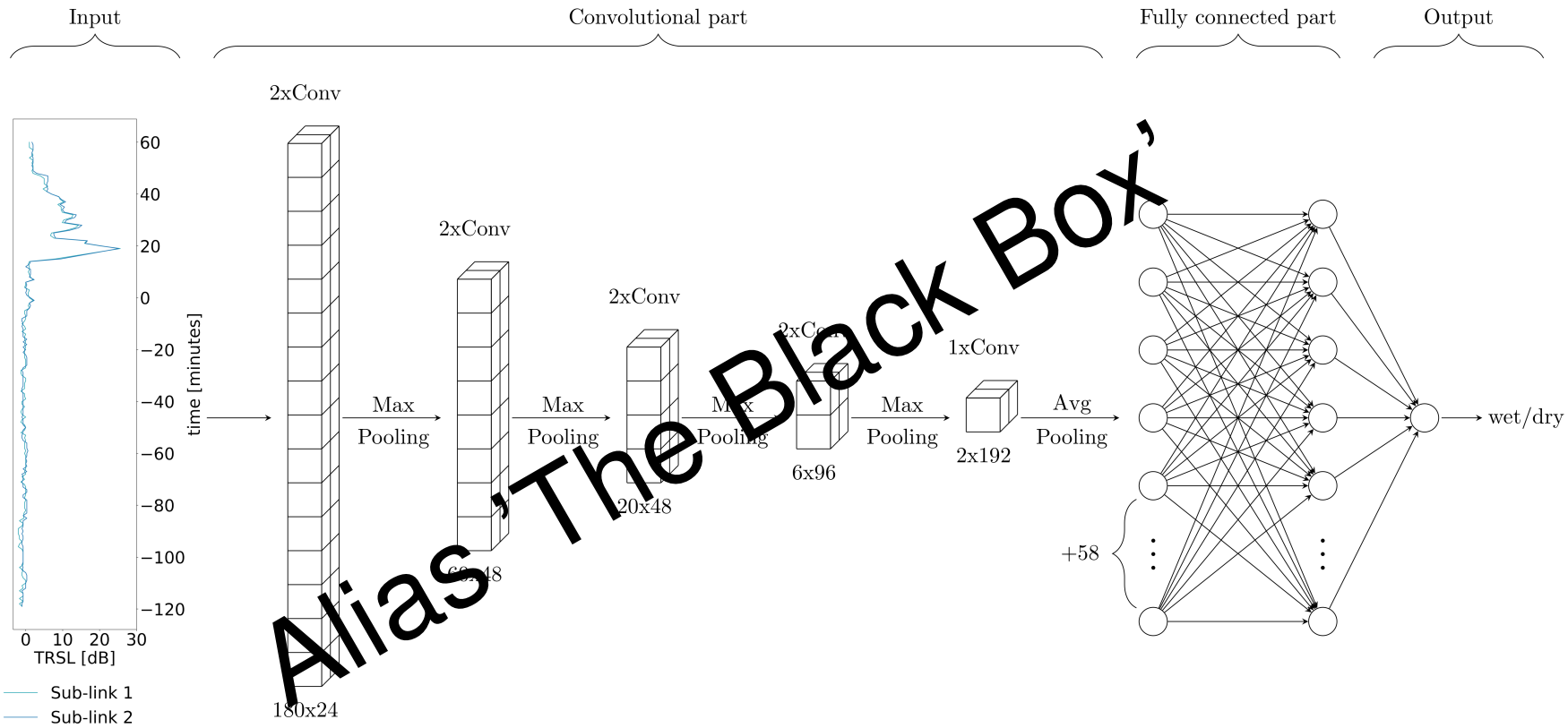
The deep learning approach



Our approach to separate wet and dry periods:

Rain event detection in commercial microwave link attenuation data with convolutional neural networks (CNNs) → `Paper under revision`

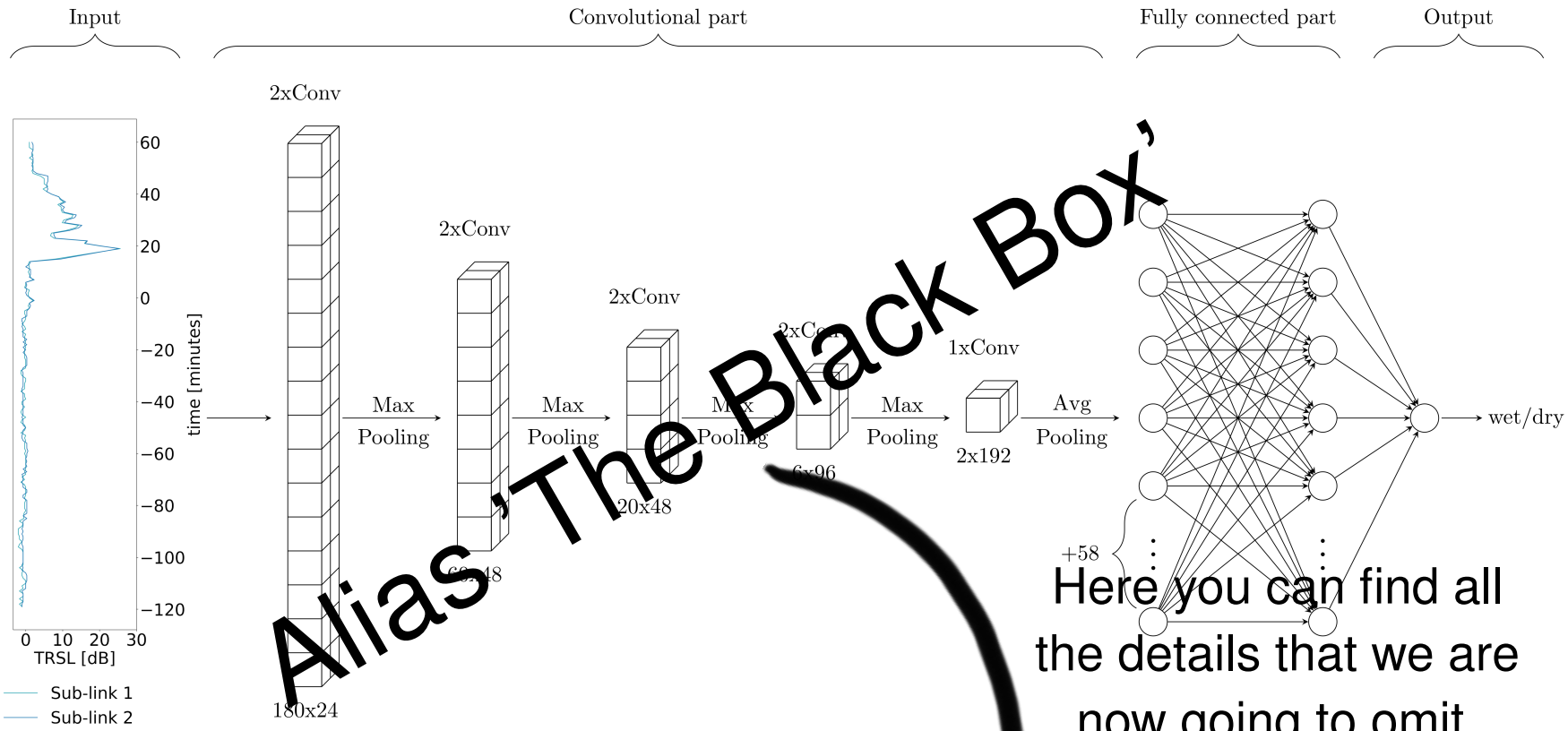
The deep learning approach



Our approach to separate wet and dry periods:

Rain event detection in commercial microwave link attenuation data with convolutional neural networks (CNNs) → `Paper under revision`

The deep learning approach

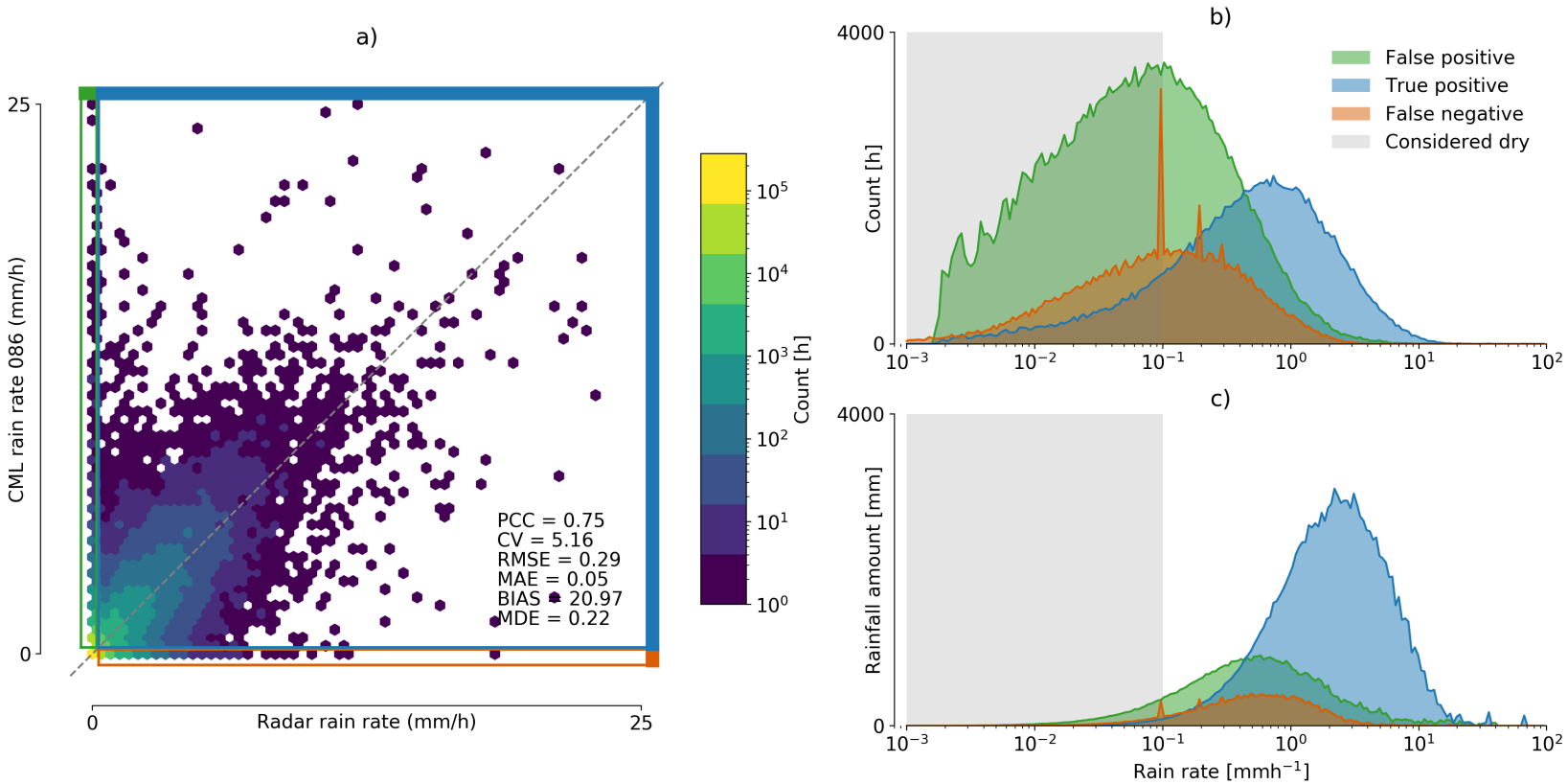


Our approach to separate wet and dry periods:

Rain event detection in commercial microwave link attenuation data with convolutional neural networks (CNNs) → `Paper under revision`

The reference event detection method

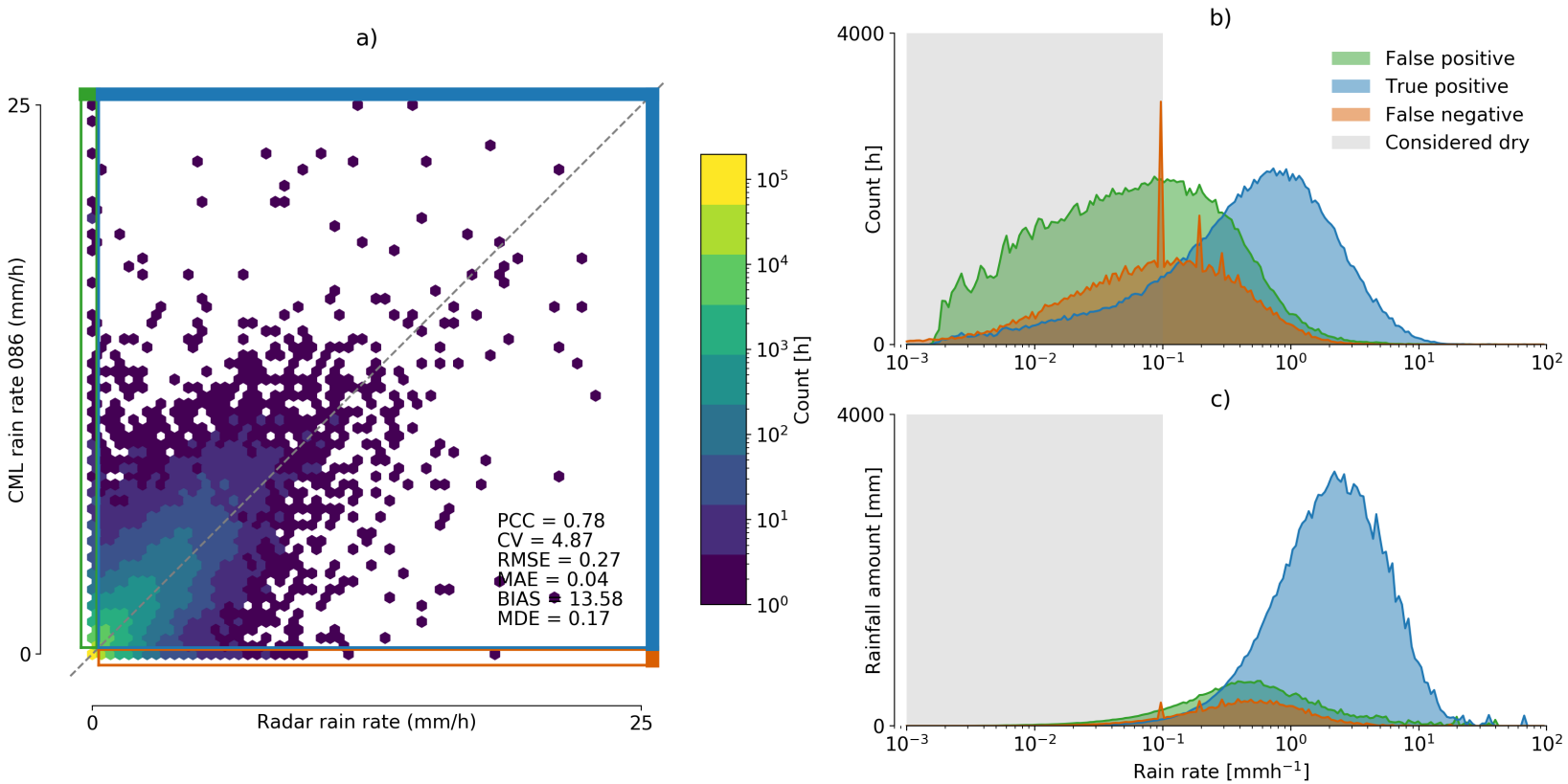
Graf et al. 2019 improved version of Schleiss and Berne 2010 refered to as Q80.



a) Hourly scatter density comparison of observed rain rates from 3904 CMLs and RADOLAN RW in April 2018
b) Histogram of the hourly rain rates derived from a)
c) Rainfall amount per Histogram bin in b)

The deep learning method

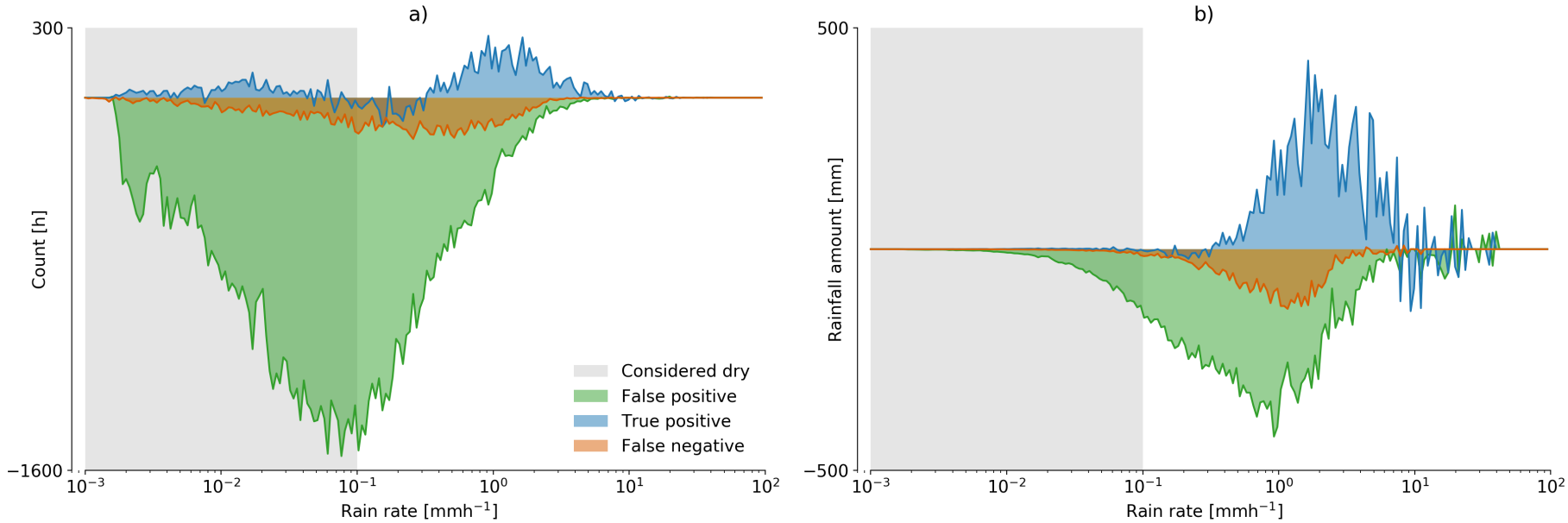
The same statistics for the CNN. All the technical details in the available preprint of Polz et al. 2019.



a) Hourly scatter density comparison of observed rain rates from 3904 CMLs and RADOLAN RW in April 2018
b) Histogram of the hourly rain rates derived from a)
c) Rainfall amount per histogram bin in b)

Improvement through the CNN

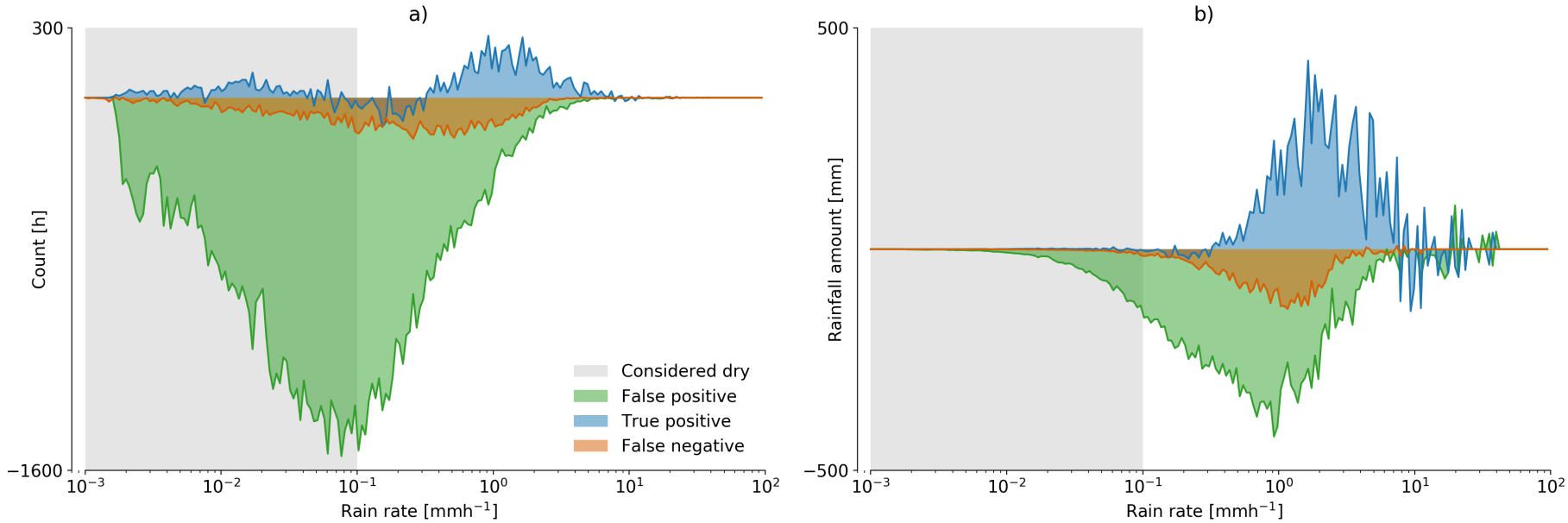
Difference in plots a) and b) from the previous slides
(numbers of Q80 subtracted by the numbers of the CNN)



a) Histogram of the difference in hourly rain rates
b) Rainfall amount per histogram bin in a)

Improvement through the CNN

Difference in plots a) and b) from the previous slides
(numbers of Q80 subtracted by the numbers of the CNN)



a) Histogram of the difference in hourly rain rates
b) Rainfall amount per histogram bin in a)

$\Rightarrow$ Reduction of falsely generated rainfall (green) by 40% while at the same time improving on True positive and False negative rates.

The end

Questions/Suggestions?

Ask me anything via julius.polz@kit.edu, on Twitter
or during the EGU 2020 live chat.

Interested in our open source model?

Get it at github.com.

Acknowledgements to

RealPEP

KIT-Campus Alpin
IMK-IFU: Atmosphärische Umweltforschung

Gefördert durch
DFG Deutsche Forschungsgemeinschaft

References

[1] Chwala, C. and Kunstmann, H.: Commercial microwave link networks for rainfall observation: Assessment of the current status and futurechallenges, Wiley Interdisciplinary Reviews: Water, 6, e1337, 2019.

[2] Chwala, C., Keis, F., and Kunstmann, H.: Real-time data acquisition of commercial microwave link networks for hydrometeorological applications, Atmos. Meas. Tech., 9, 991999, 2016.

[3] Graf, M., Chwala, C., Polz, J., and Kunstmann, H.: Rainfall estimation from a German-wide commercial microwave link network: Optimized processing and validation for one year of data, Hydrol. Earth Syst. Sci. Discuss., in review, 2019.

[4] Schleiss, M. and Berne, A.: Identification of Dry and Rainy Periods Using Telecommunication Microwave Links, IEEE Geoscience andRemote Sensing Letters, 7, 611615, 2010.

[5] Polz, J., Chwala, C., Graf, M., and Kunstmann, H.: Rain event detection in commercial microwave link attenuation data using convolutional neural networks, Atmos. Meas. Tech. Discuss., in review, 2019.