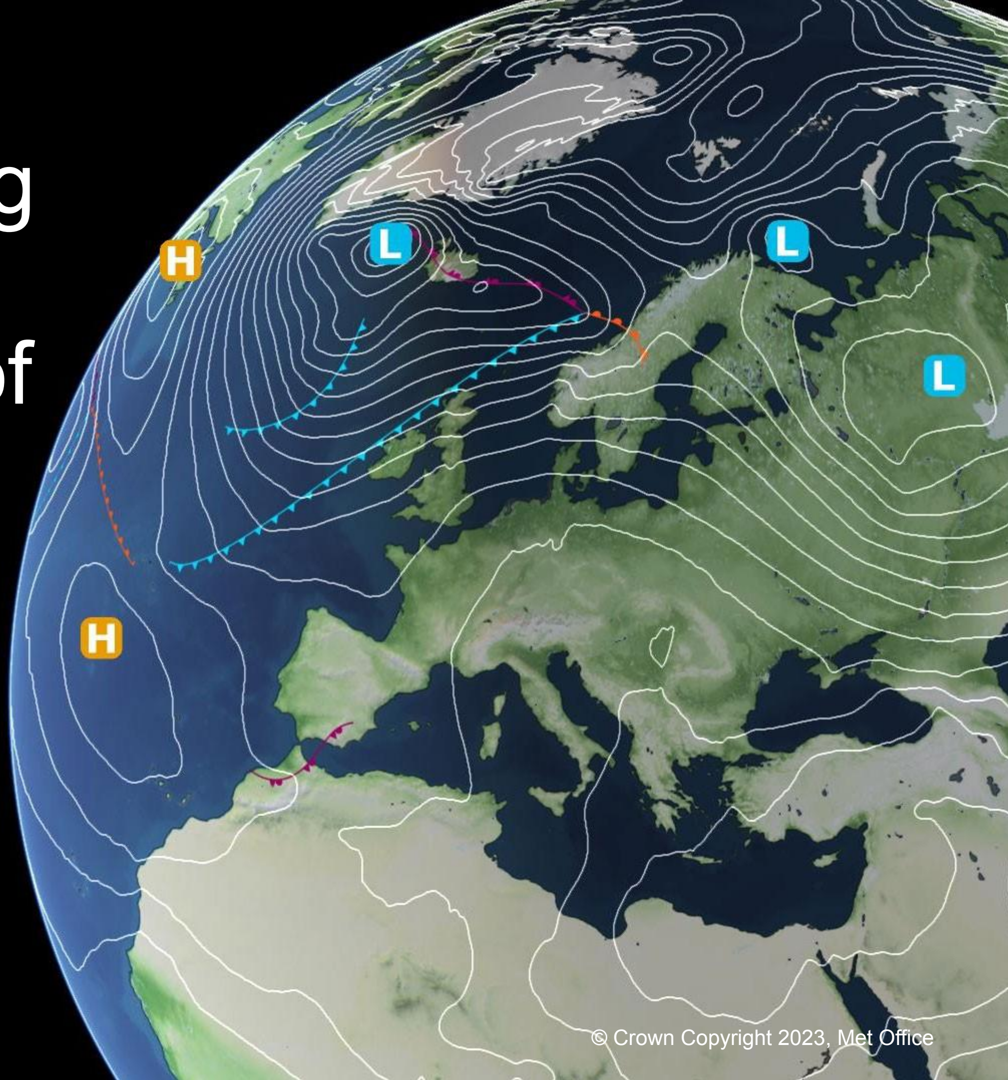
Met Office

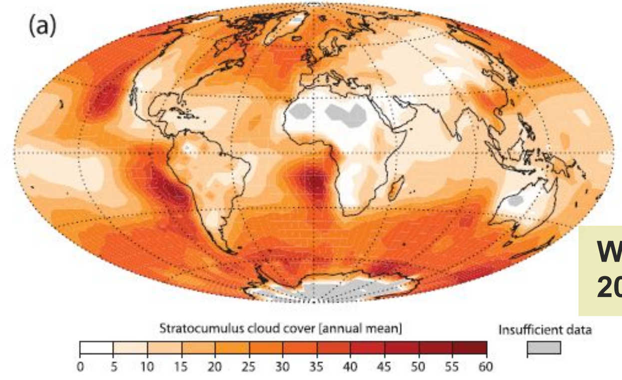
Can Machine Learning determine the ***CAUSATIVE*** effects of aerosol on clouds?

Daniel Grosvenor[1,2], Lukas Zipfel[3], Jan Cermak[3], Jane Mulcahy[1].

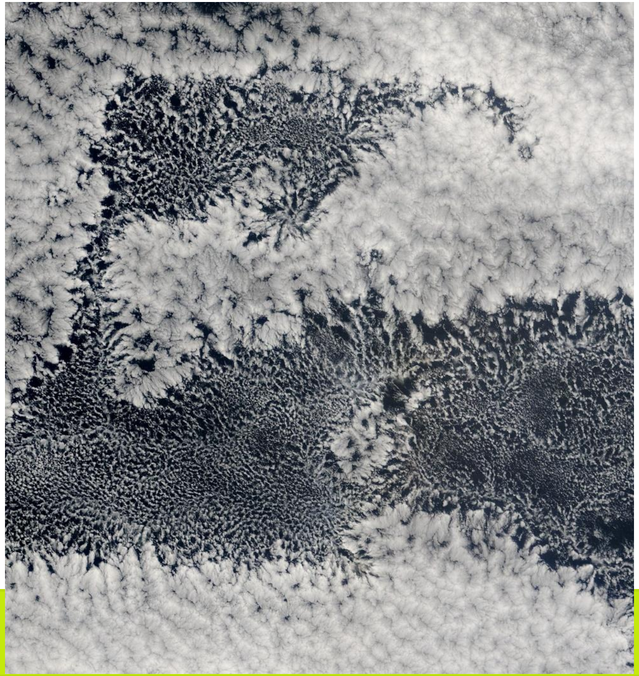
1. The Met Office, 2. University of Leeds (CEMAC), 3. Karlsruhe Institute of Technology (KIT).

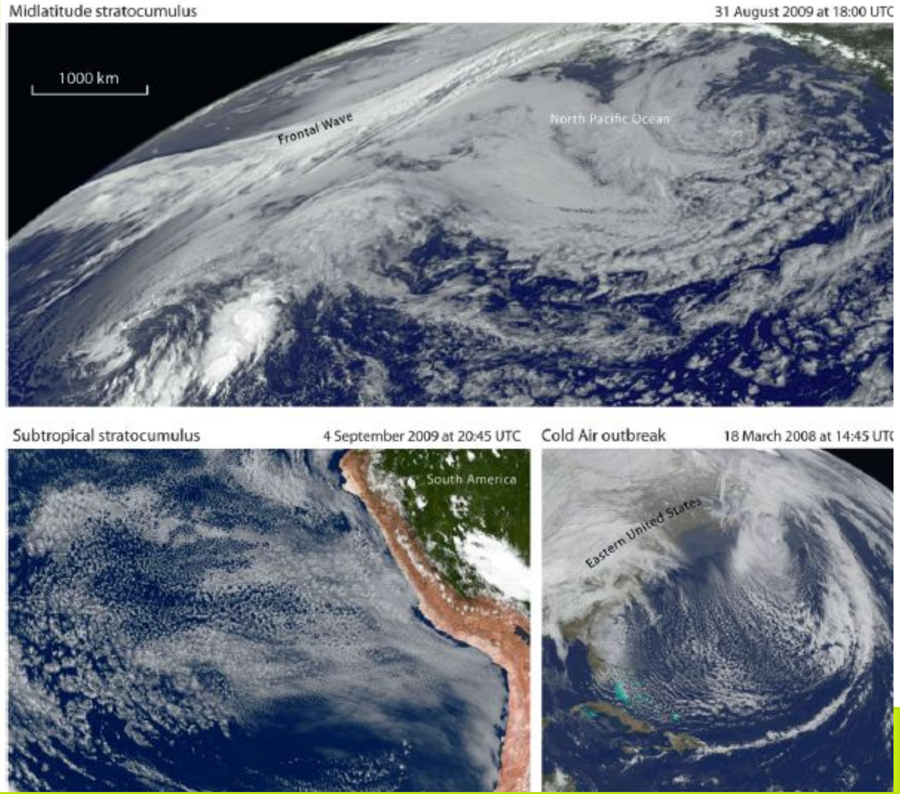


Stratocumulus clouds cover 20% of the Earth’s surface on average



Wood, (Mon. Wea. Rev., 2012) - Statocumulus Clouds





Stratocumulus clouds – reflecting sunlight & cooling the surface

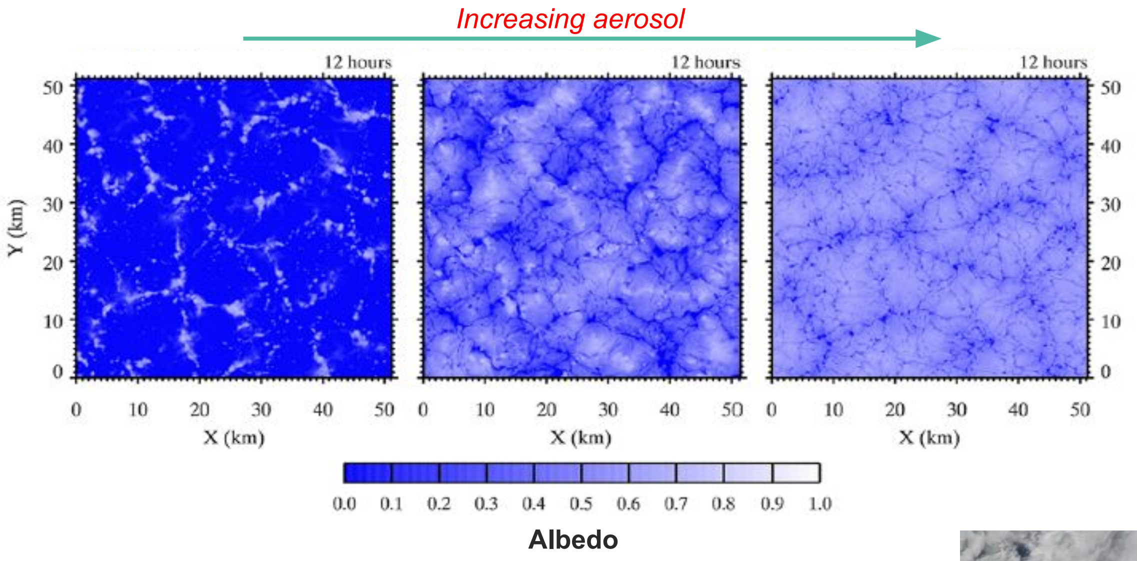
Stratocumulus clouds prevent sun from reaching the surface and cool the planet.

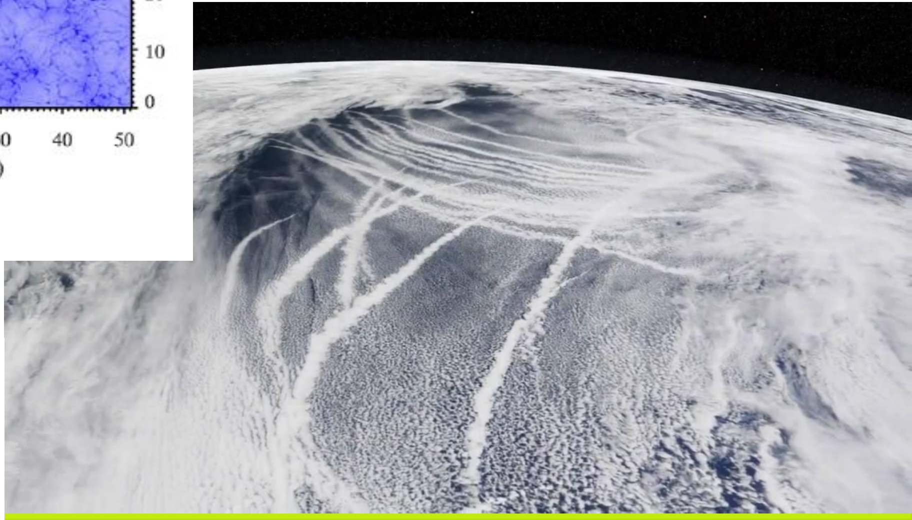
An approx. 10% increase in the albedo/reflectivity can offset the warming from a doubling of CO_2.
Aerosols can increase cloud reflectivity and coverage.

Aerosol effects on cloud fraction (cloud adjustment)

Yamaguchi and Feingold (ACP, 2015) On the relationship between open cellular convective cloud patterns and the spatial distribution of precipitation

- Cloud fraction changes have a large radiative impact

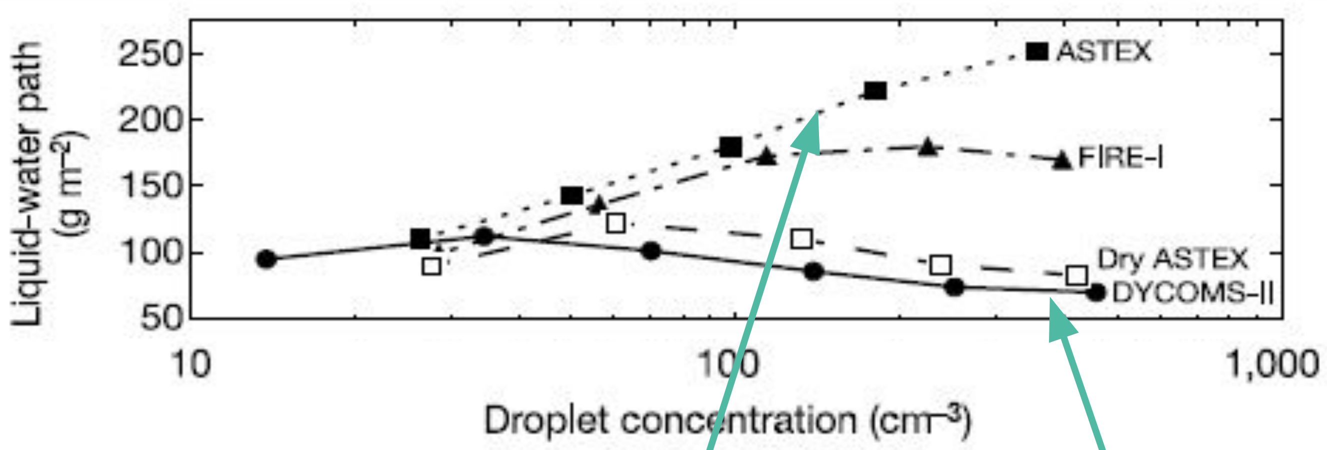




- High resolution (Δx=50-250m, Δz>5m) LES modelling suggests that aerosol can increase cloud fraction (due to precipitation suppression)

Aerosol effects on cloud thickness/Liquid Water Path (cloud adjustment)

Ackerman et al. (Nature, 2004) The impact of humidity above stratiform clouds on indirect aerosol climate forcing



Liquid Water Path (LWP) is the vertically integrated mass of cloud liquid water

If the zeros between clouds are included then it also depends on the cloud areal coverage:

LWP = LWP$_{in\text{-}cloud}$ x cloud fraction

- High resolution LES simulations of stratocumulus clouds

- **LWP increased with aerosols (droplet number) for precipitating clouds**
  - **Due to more smaller cloud droplets supressing precipitation**

LWP decreased with aerosols in non-precipitating clouds

- More aerosols caused more entrainment (reducing LWP)
- Entrainment is the mixing of dry air from above into the cloud

Attempts to observe Liquid Water Path vs droplet number with satellite data

Gryspeerdt et al. (ACP, 2019)
Constraining the aerosol influence on cloud liquid water path

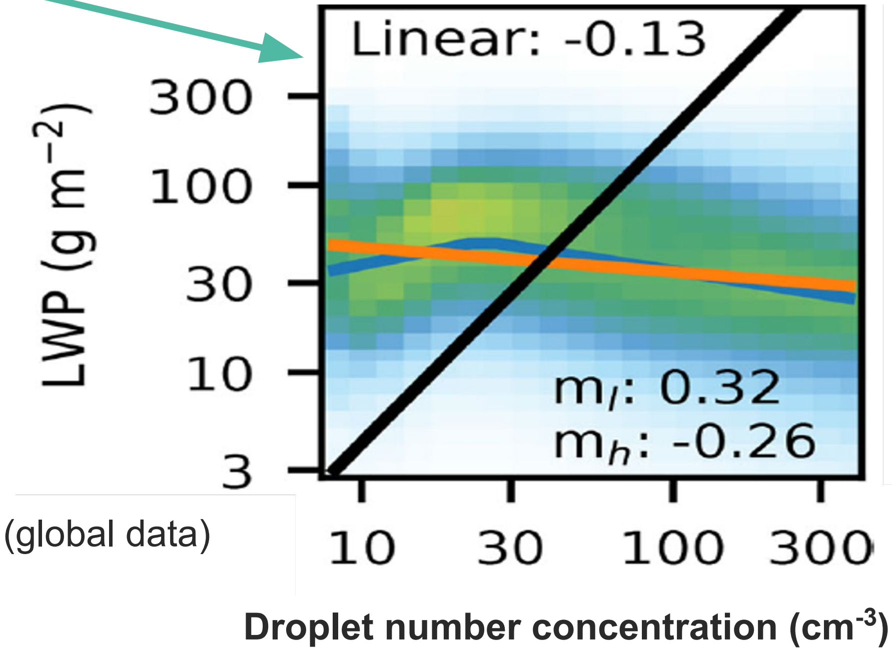
Attempts have been made to determine LWP vs cloud droplet number concentrations (N_d) using satellite data

The results show an "inverted V" shape, but with the peak LWP at an unexpectedly low N_d value of around 25 cm^{-3}.

Recent studies suggest that such curves are not the result of changes in aerosol.

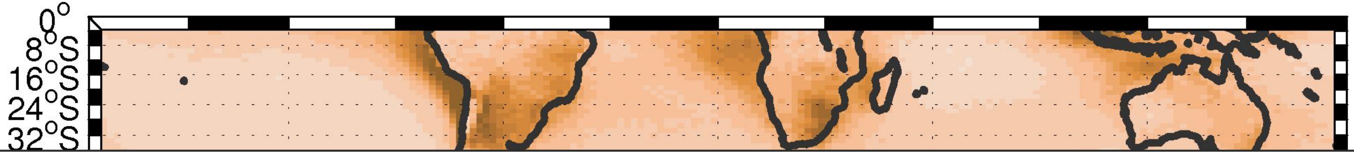
Rather due to confounding factors:

- **Natural co-variability (caused by e.g., meteorological effects)**
- **Instrument errors**
- **Arola, 2022; Mulmenstadt, 2024; Goren, 2024.**



Example of meteorological co-variability. Time-mean maps from UM model.

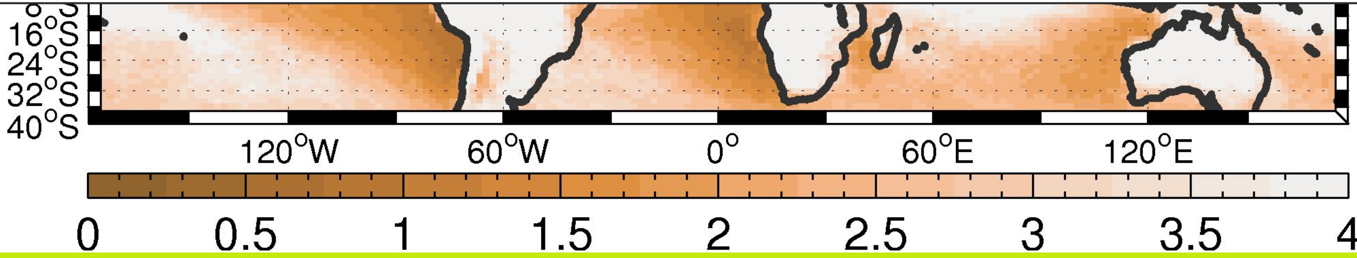
Cloud droplet concentration



Aerosol effects are difficult to observe hampering model testing

Aerosols are one of the leading causes of climate uncertainty

Cloud top height (km)



Can machine learning help to do better with the observations?

Met Office

Zipfel (2022) and Zipfel (2024)

Used Gradient Boosting Regression Tree (GBRT) machine learning techniques to *predict all-sky AMSR-E (satellite) LWP* values as a function of:-

Reanalysis meteorological variables.

MODIS droplet concentration

(20km grid boxes)

(Daily snapshots from polar orbiting satellites)

One year of data

(2010)

Met variables

Satellite cloud variables

Table 2. Overview of the variables used in the machine learning models.

Variable Name	Abbreviation	Origin
Temperature below cloud	T_{bc}	ERA5
Vertical velocity below cloud	w_{bc}	ERA5
Winds below cloud	u_{bc}/v_{bc}	ERA5
Winds above cloud	u_{ac}/v_{ac}	ERA5
Relative humidity below cloud	RH_{bc}	ERA5
Relative humidity above cloud	RH_{ac}	ERA5
Mean sea level pressure	MSL	ERA5
Sea surface temperature	SST	ERA5
Estimated inversion strength	EIS	ERA5
Cloud top height	CTH	CALIPSO
Precipitation fraction	PF	CloudSat
Cloud droplet number concentration	N_d	MODIS
Liquid water path	LWP	AMSR-E
Rain water content [1]	RWC	AMSR-E

[1] RWC not used to predict LWP.

Table 1. Locations of the study domains with the total number of observations (N) and the explained variability for the PF (R^2_{pf}) and LWP (R^2_{lwp}) models.

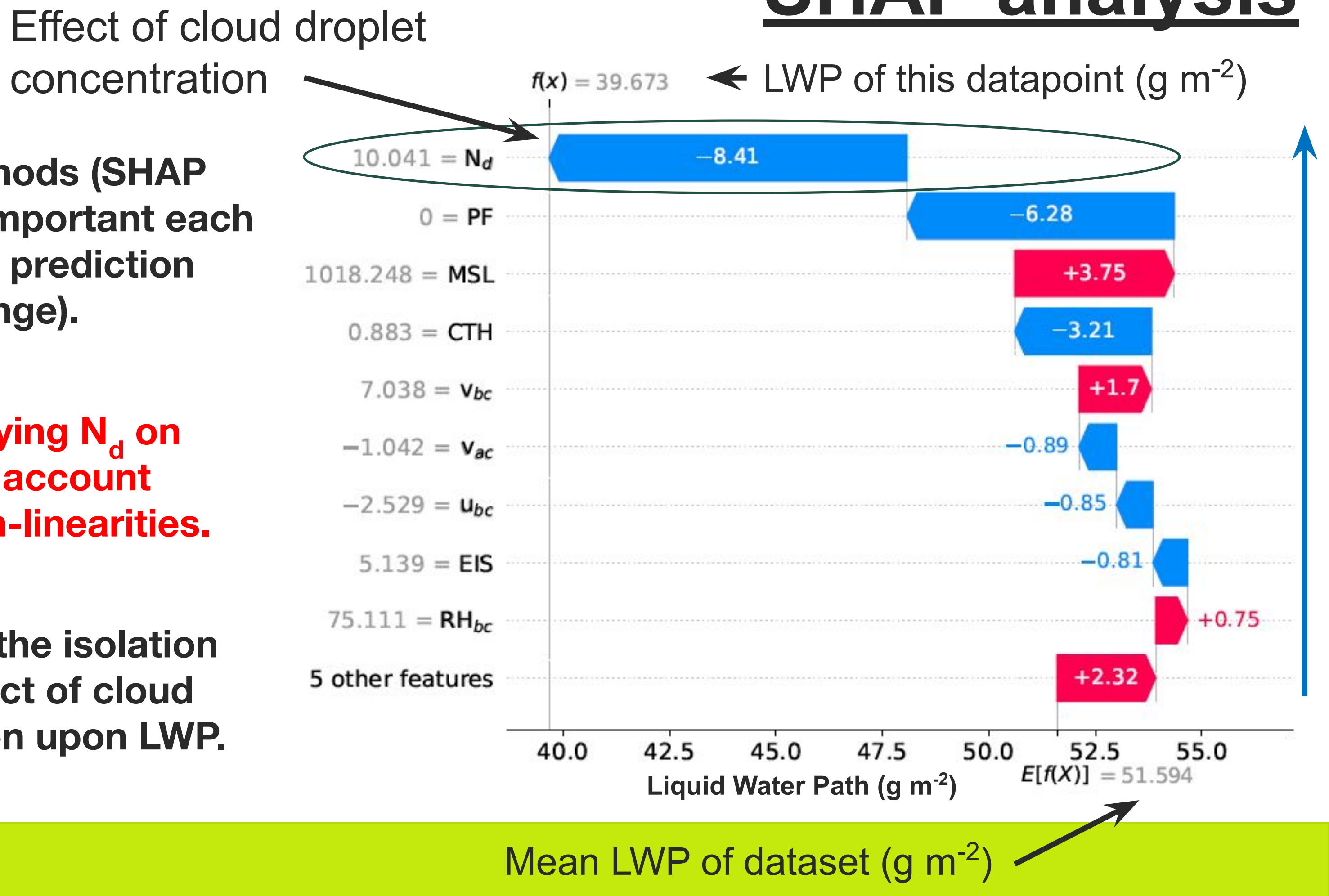
Region Name	Latitude	Longitude	N	R^2_{pf}	R^2_{lwp}
Australia	25° S–35° S	95° E–105° E	16,504	0.63	0.61
California	20° N–30° N	120° W–130° W	18,919	0.71	0.65
Canaries	15° N–25° N	25° W–35° W	8431	0.65	0.67
Namibia	10° S–20° S	0°–10° E	20,337	0.68	0.66
Peru	10° S–20° S	80° W–90° W	23,512	0.63	0.66

SHAP analysis

Explainable ML methods (SHAP values) tell us how important each variable is to the ML prediction (and the sign of change).

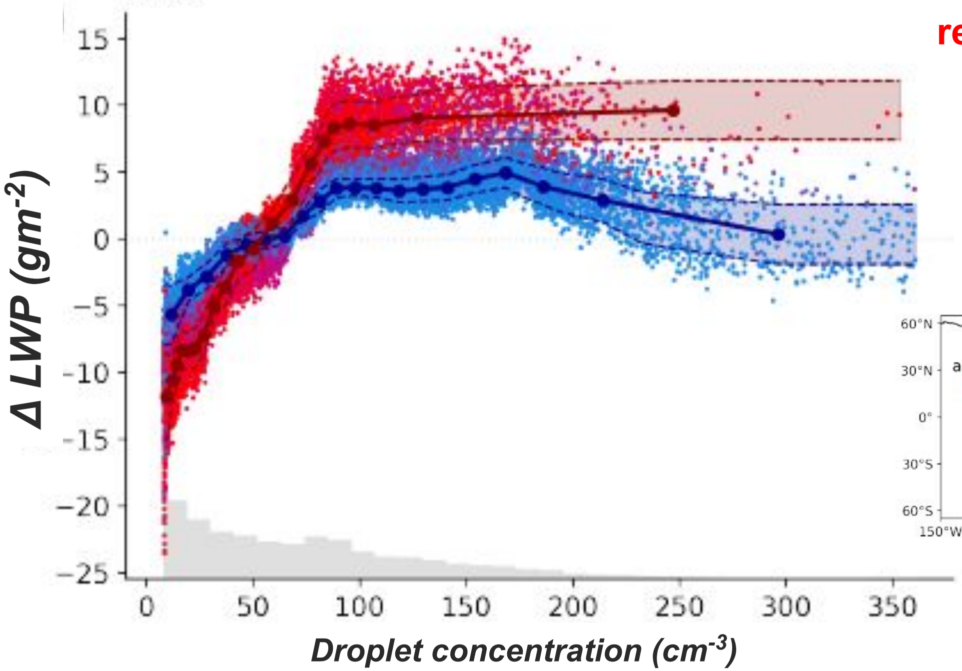
I.e., the effect of varying N_d on LWP, but taking into account co-variation and non-linearities.

Potentially allowing the isolation of the *causative* effect of cloud droplet concentration upon LWP.



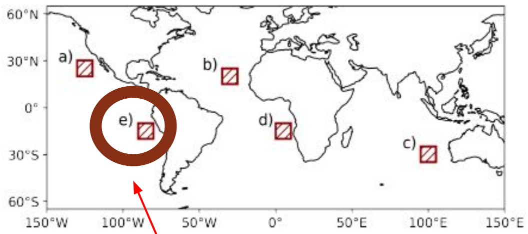
Can machine learning (ML) help to do better with the observations?

- **Plotting median SHAP value for the N_d effect on LWP binned by N_d.**
- **I.e., the LWP vs N_d relationship.**
- **Does this emulate what we would see in a model where we just change N_d and let everything else adjust?**



red: high precip quartile

blue: low precip quartile



Peru region used for model

Figure: Median contribution of N_d values to ML prediction (SHAP value) binned by N_d (Zipfel, 2024).

Applying ML to climate model data

- Zipfel (2022) and Zipfel (2024)
- Used same Gradient Boosting Regression Tree (GBRT) machine learning technique to ***predict all-sky AMSR-E (satellite) LWP*** values as a function of:-
  - *Meteorological variables*
  - *Droplet concentration*
  - *(140-200km grid boxes)*
  - *(Daily snapshots from model)*

Met variables

~~Satellite~~ model cloud variables

Table 2. Overview of the variables used in the machine learning models.

Variable Name	Abbreviation	Origin
Temperature below cloud	T_{bc}	Climate Model
Vertical velocity below cloud	w_{bc}	
Winds below cloud	u_{bc}/v_{bc}	
Winds above cloud	u_{ac}/v_{ac}	
Relative humidity below cloud	RH_{bc}	
Relative humidity above cloud	RH_{ac}	
Mean sea level pressure	MSL	
Sea surface temperature	SST	
Estimated inversion strength	EIS	
Cloud top height	CTH	
Precipitation fraction	PF	
Cloud droplet number concentration	N_d	
Liquid water path	LWP	
Rain water content [1]	RWC	

[1] RWC not used to predict LWP.

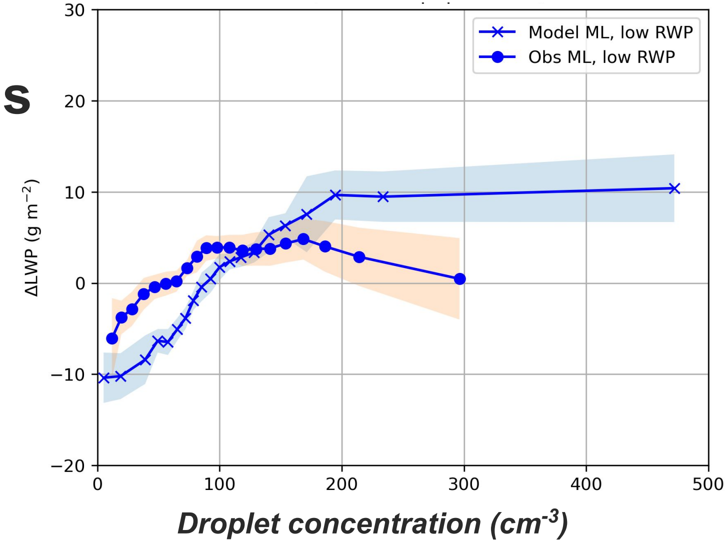
- Model: UK Earth System Model v1.1 (UKESM1.1)
  - Atmosphere only; prescribed sea-surface temperatures and sea-ice.
  - Temperature and winds nudged/relaxed to ERA5.

R^2 from ML model for LWP built from model snapshots = 0.73

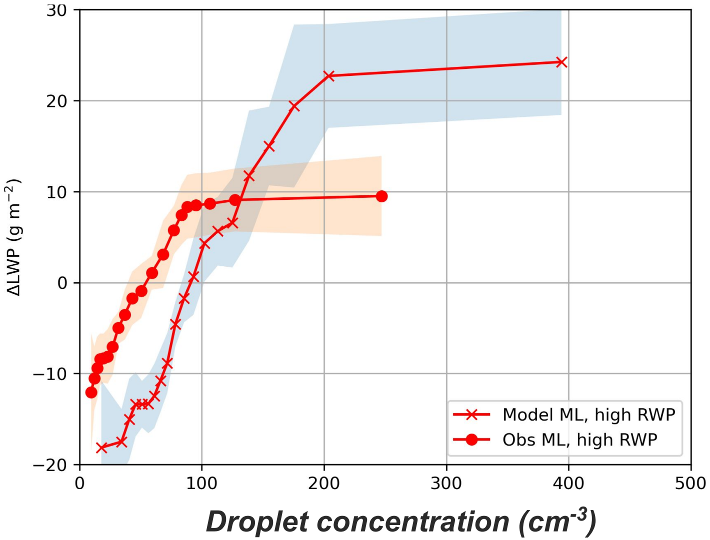
Using the result to evaluate the model vs obs

SHAP analysis

Low precipitation cases (0-25th percentile)



High precipitation cases (75-100th percentile)

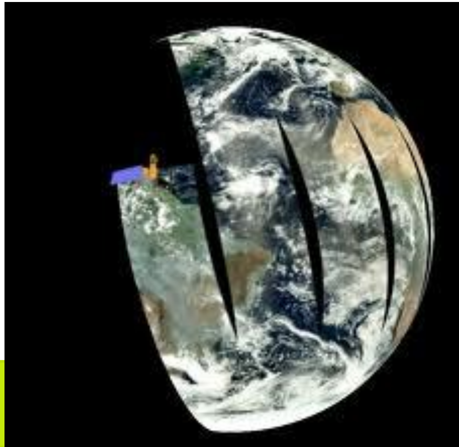


- **Low precipitation cases: Reasonable gradient at low droplet concs; poor agreement for higher values.**
  - ***Expected since model does not have the link between droplet number and entrainment.***
- **High precipitation cases: Good agreement (gradient) over a wider range; still issues at higher droplet concentrations.**

Met Office

But how well does the ML approach capture the causative effect of N_d on LWP?

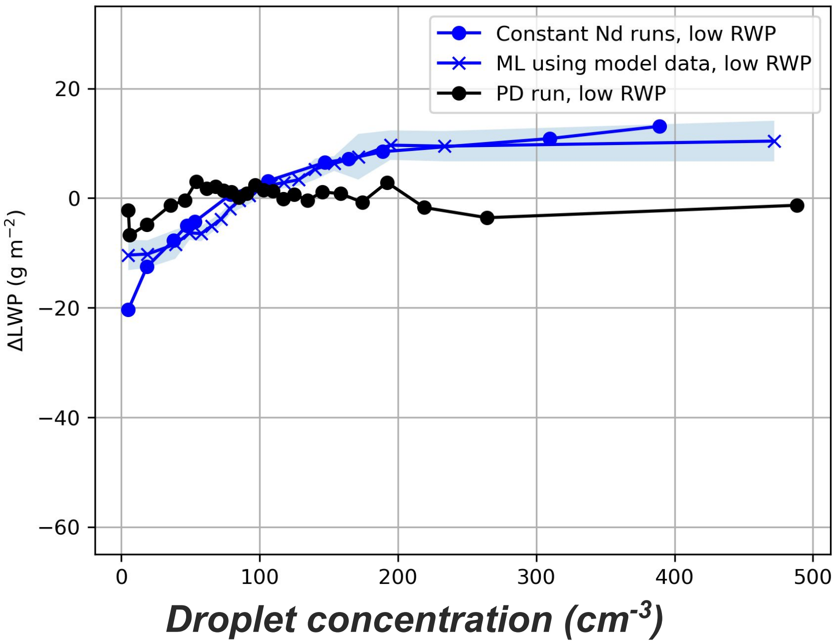
- Can use the model as a testbed for this as we can **vary droplet concentration (N_d) directly** and let the model and LWP respond.
- Then compare this to the ML approach trained using the model data (previous slide).
- This will give us confidence in the results from the observations.
  - Is there enough information in the daily snapshots?



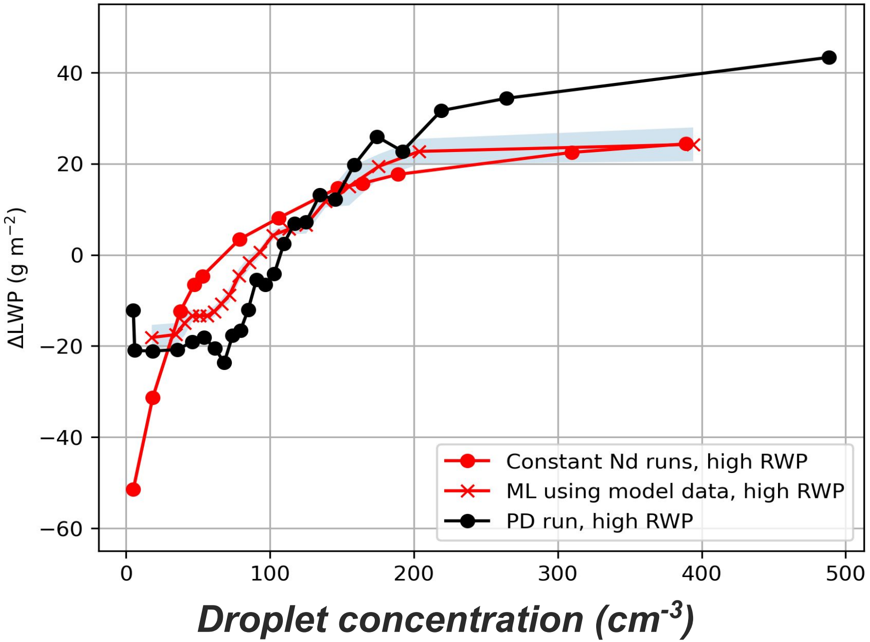
- Used idealised simulations for this with constant N_d applied.
- 12 simulations covering a range of $N_{d.}$
- Otherwise, the same setup as the original model (where N_d varies according to the aerosol).

Testing how well the ML approach captures the causative effect of Nd on LWP.

Low precipitation cases (0-25th percentile)



High precipitation cases (75-100th percentile)



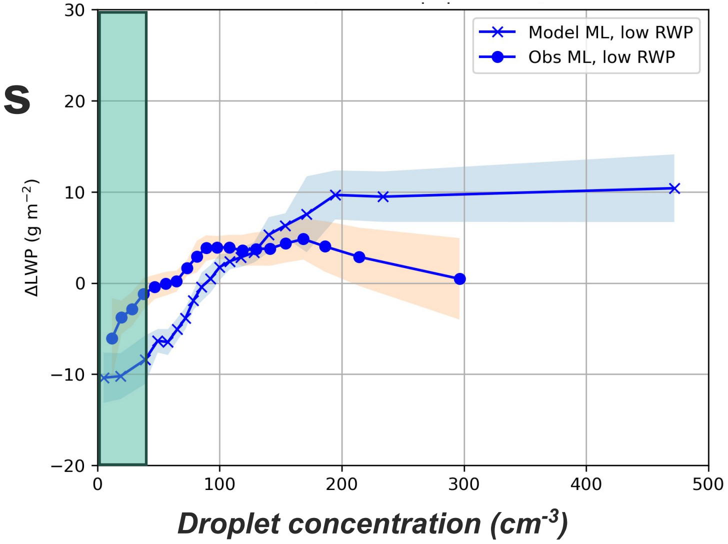
- It seems that the SHAP analysis is capturing the ***causative*** effect of droplet concentration on LWP quite well.
- Except at low droplet concentrations.
- Gives some confidence in the relationship obtained from observations and the resulting model evaluation.

Met Office

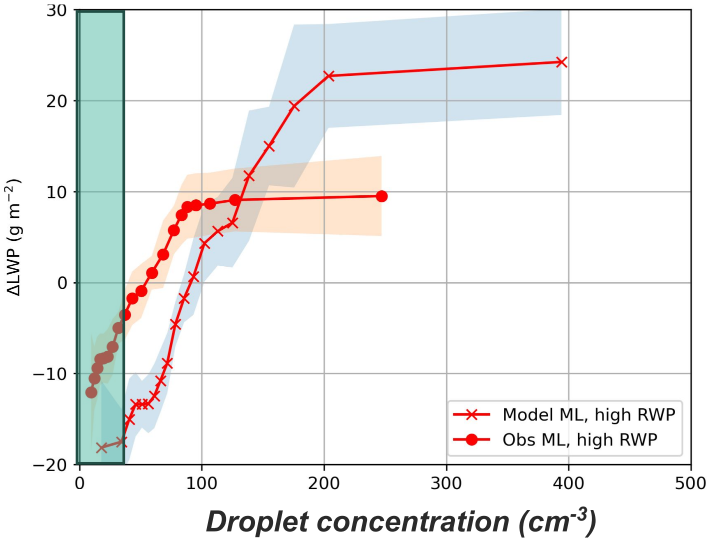
Using the result to evaluate the model vs obs

SHAP analysis

Low precipitation cases (0-25th percentile)



High precipitation cases (75-100th percentile)



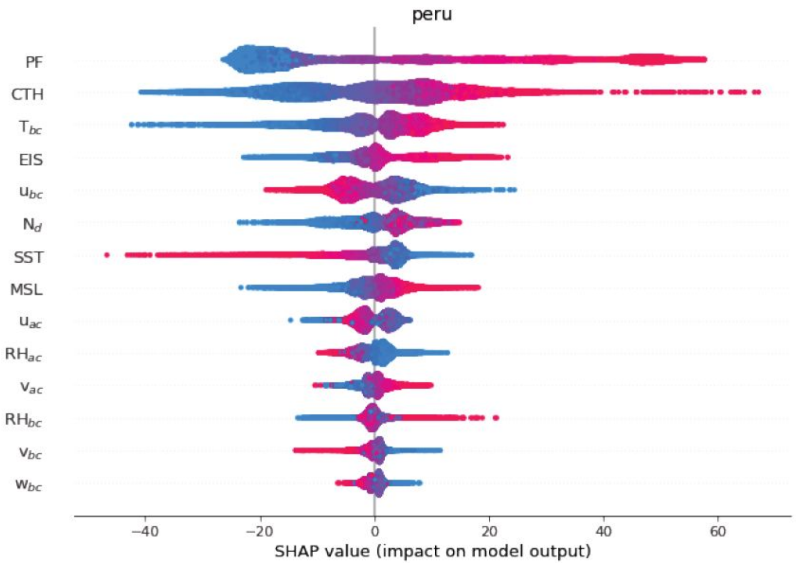
- **Low precipitation cases: Reasonable gradient at low droplet concs; poor agreement for higher values.**
  - ***Expected since model does not have the link between droplet number and entrainment.***
- **High precipitation cases: Good agreement (gradient) over a wider range; still issues at higher droplet concentrations.**

Met Office

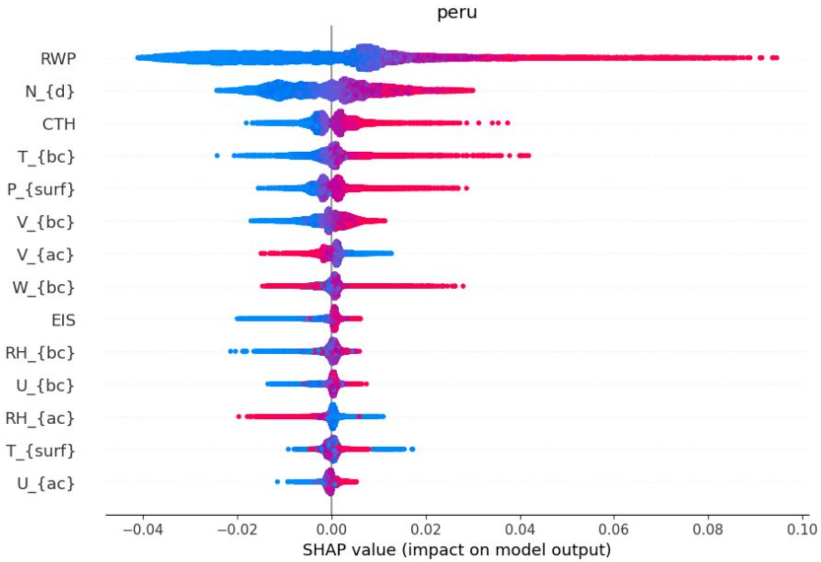
Conclusions

- Machine Learning techniques may be able to determine the ***causative*** relationship between Liquid Water Path (LWP; cloud thickness x fraction) effects and aerosols (droplet concentrations) ***from observations.***
- This could then be used to evaluate models.
- Doing this suggests that the UKESM climate model is ok in precipitating clouds (except at higher droplet concentration values).
- But bad for non-precipitating clouds.
  - Likely due to lack of link between entrainment & droplet concentration in model.
- However, is the ML relationship giving us the right answer?
- Tested this in "model world" where we can artificially vary droplet concentration / aerosol emissions.
- Suggests that the ML analysis does capture the causative effect of cloud droplet number concentration on Liquid Water Path, except at low droplet concentrations.
- Gives confidence in the model evaluation vs observations.

Observations

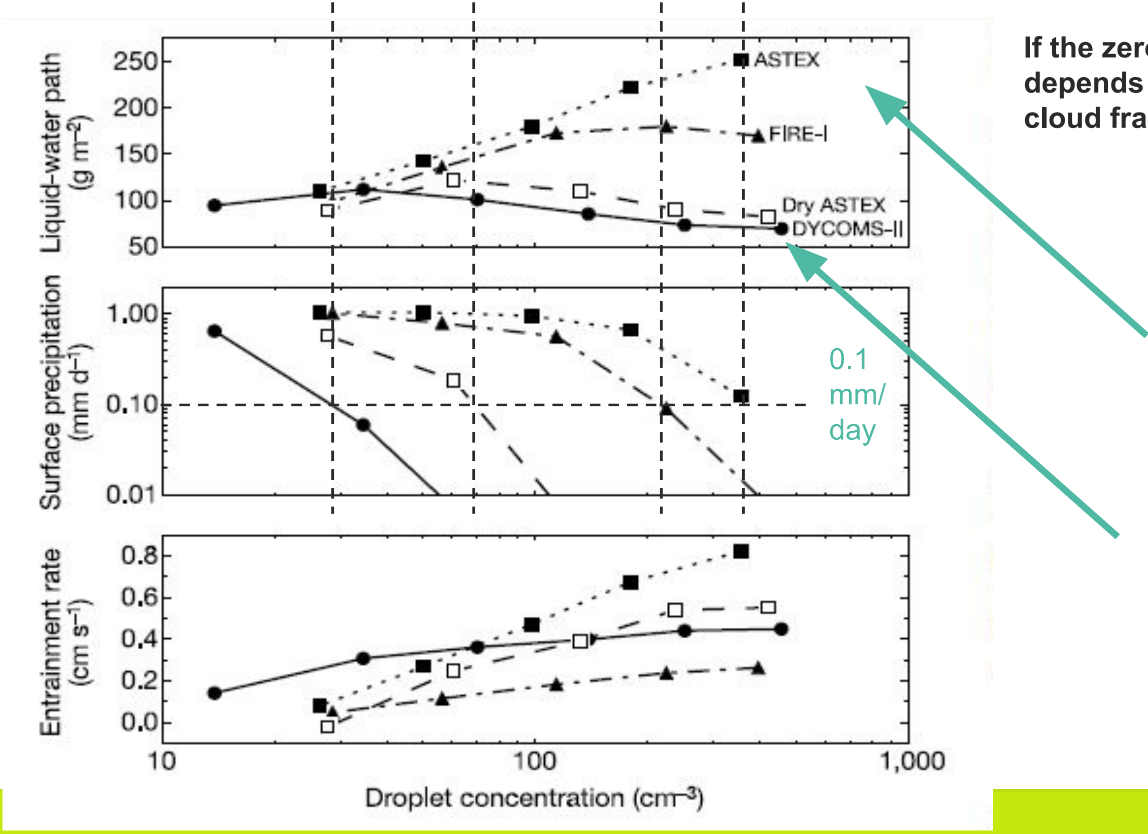


Model



Aerosol effects on cloud thickness/Liquid Water Path (cloud adjustment)

Ackerman et al. (Nature, 2004) The impact of humidity above stratiform clouds on indirect aerosol climate forcing



Liquid Water Path (LWP) is the vertically integrated mass of cloud liquid water

If the zeros between clouds are included then it also depends on the cloud areal coverage (LWP = $LWP_{in\text{-}cloud}$ x cloud fraction)

- High resolution LES simulations of stratocumulus clouds
- **LWP increased with aerosols (droplet number) for precipitating clouds**
  - **Due to more smaller cloud droplets supressing precipitation**
- But, LWP decreased with aerosols in non-precipitating clouds
  - More aerosols also caused more entrainment (reducing LWP)
  - Entrainment is the mixing of dry air from above cloud into the cloud

Met Office

Attempts to observe Liquid Water Path vs droplet number with satellite data

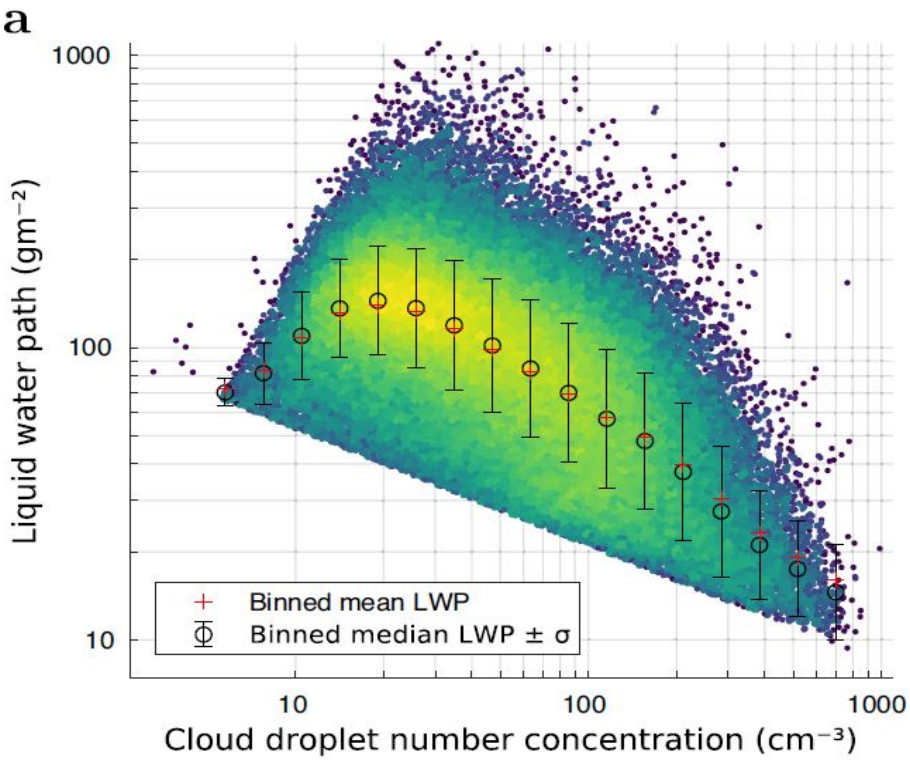
Attempts have been cloud droplet numb satellite data

The results show an the peak LWP at an around 25 cm^{-3}.

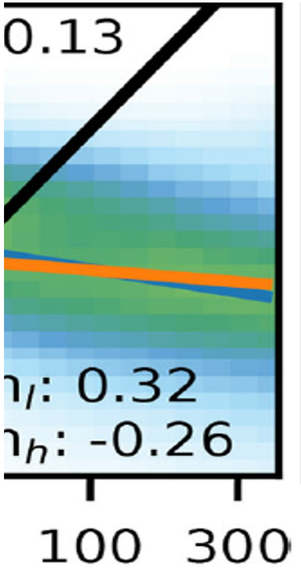
Recent studies sugg the result of change

Rather due to confo

- Natural co-var meteorologica
- Instrument err
- Arola, 2022; M



(ACP, 2019) aerosol influence ter path



Droplet number concentration (cm^{-3})