Environment and Climate Change Canada
Environnement et Changement climatique Canada

ClimART

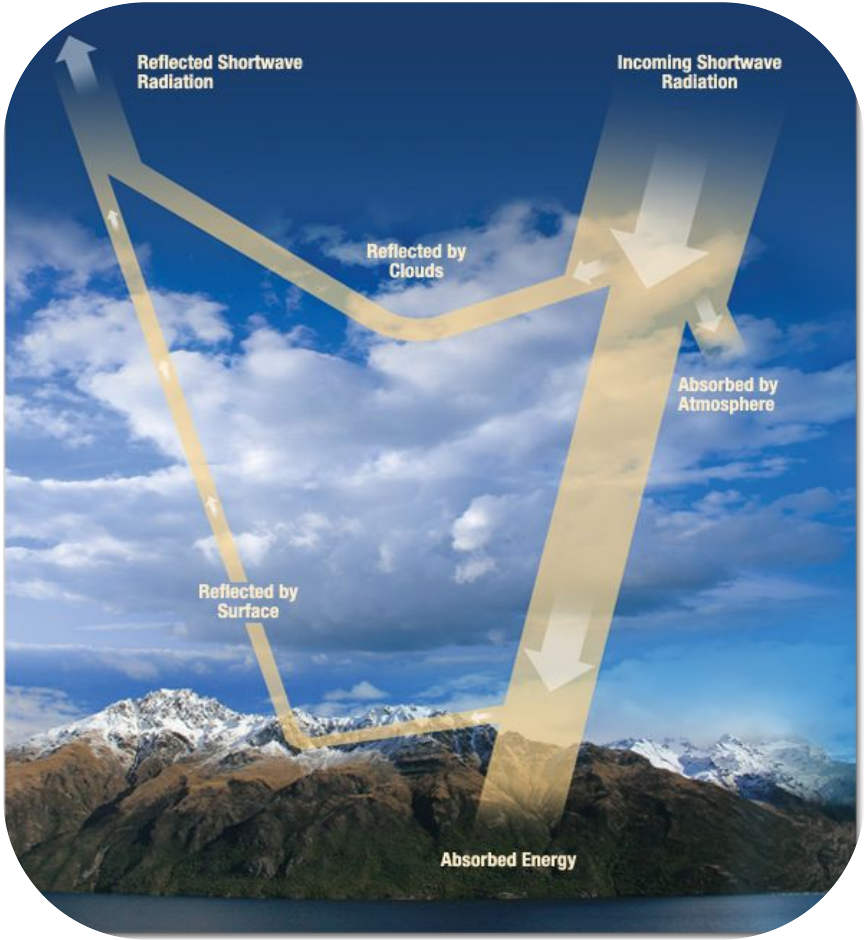
A Benchmark Dataset for Emulating Atmospheric Radiative Transfer in Weather and Climate Models

Salva Rühling Cachay*, Venkatesh Ramesh*, Jason N. S. Cole, Howard Barker, and David Rolnick.
In Thirty-fifth Conference on Neural Information Processing Systems (NeurIPS) Datasets and Benchmarks Track, 2021. URL
Code: https://github.com/RolnickLab/climart

Radiative transfer

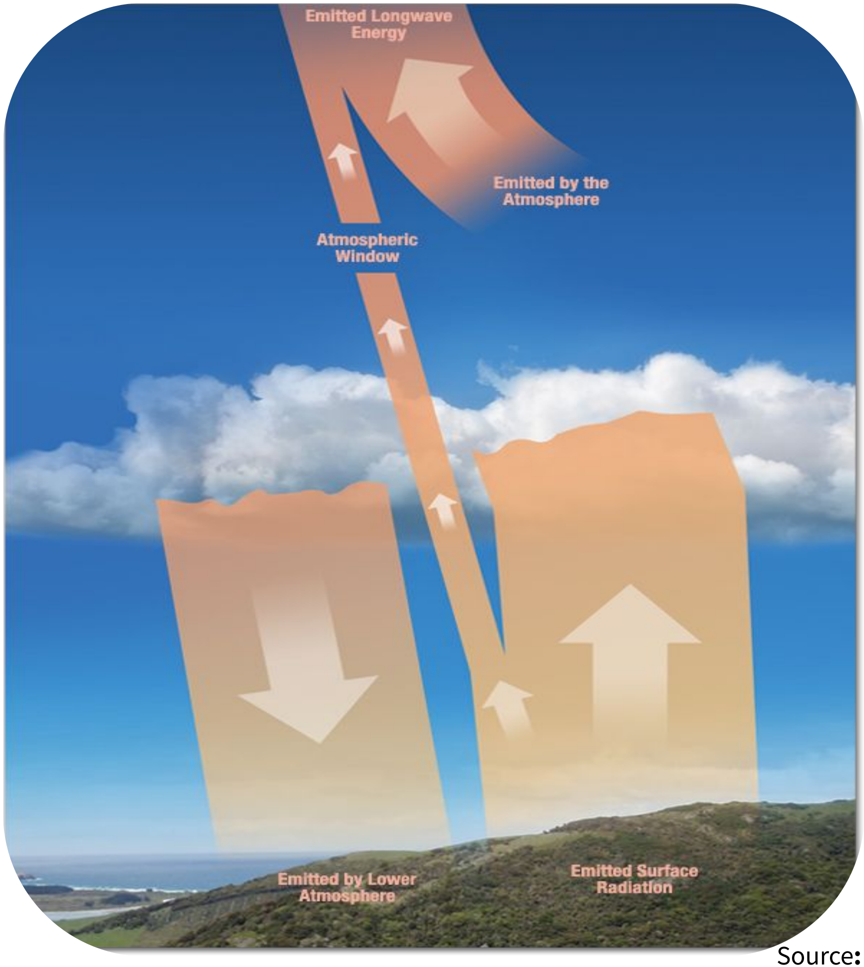
= Propagation of radiation

(through the atmosphere, in our case)



Shortwave radiation = emitted by the sun

Longwave radiation = emitted by the Earth



Source: **Nasa**

Goal: ***Speed-up computationally slow component of climate & weather models***

Why?

→ Allow for more simulations.

→ Improve simulations (e.g.: run at more simulation steps).

→ Run at higher spatial and/or temporal resolution.

<u>Goal</u>: ***Speed-up computationally slow component of climate & weather models***

Why?

→ Allow for more simulations.

→ Improve simulations (e.g.: run at every simulation step).

→ Run at higher spatial and/or temporal resolution.

→ Better understand & adapt to the impacts of climate change

→ Motivate stakeholders towards mitigating actions

ClimART dataset

Large-scale

> 10 million data points
& “ML-ready”

- Allow ML model failure analysis
- Standardize dataset, training setup (1979-2004), and evaluation (2007-14)

Comprehensive

Multiple data subsets with distributional shifts

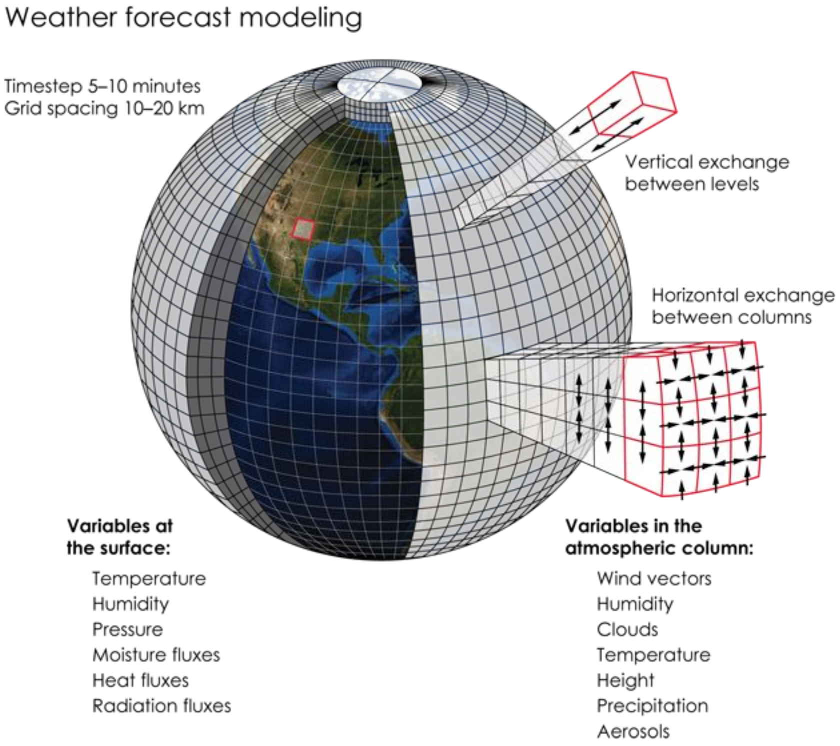
- Historical conditions (1850-52)
- Future conditions (2097-99)
- Anomalies due to volcanic eruptions (eg. Mt. Pinatubo, 1991)

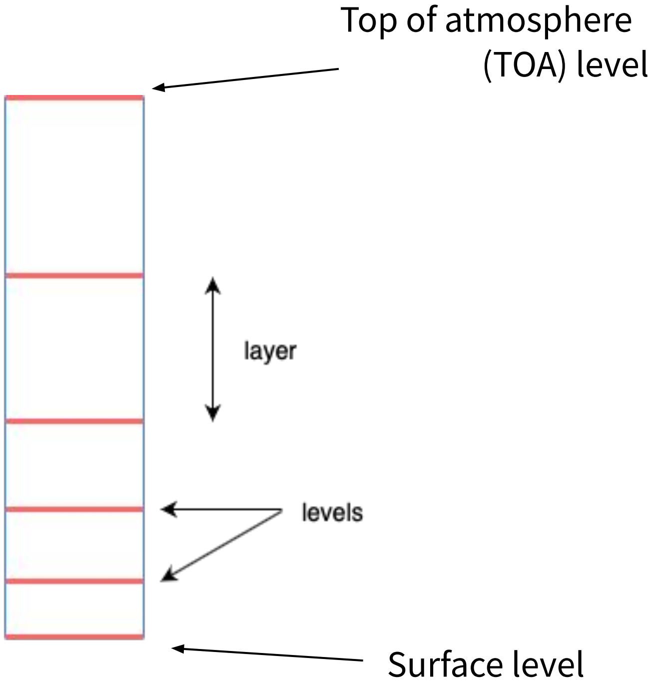
Challenging

Many promising directions for improving on our baselines

- Out-of-distribution generalization
- Complex underlying physics
- Accuracy ← → inference speed trade-off

Atmospheric Data Format





Lateral view of a profile/column

Source: SERC/Carleton College

ClimART Properties

→ Drawn from the Canadian Earth System Model (CanESM5)

- Physics RT model follows independent-column assumption
- Future and pre-industrial simulations based on CMIP6 scenarios
- 8192 geographical locations (horizontal discretization)
- 50 levels (vertical discretization)

Inputs

Non-spatial information

1D vertical profiles of the atmospheric state

Surface Variables	
emisrot	Surface emissivity for each surface tile
gtrot	Surface temperature for each surface tile
farerot	Fraction of grid of each surface tile
salbrot	All-sky surface albedo for each surface tile
csalrot	Clear-sky surface albedo for each surface tile
gtrow	Grid-mean surface temperature
pressg	Surface pressure

Layer/Level Variables	
shtj	Eta coordinate at layer interface
tfrow	Temperature at layer interfaces
shj	Eta coordinate at layer mid-point
dshj	Layer thickness in eta coordinate
dz	Geometric thickness of the layer
tlayer	Temperature at layer mid-point

Gas Variables	
ozphs	Ozone
qc	Water vapour
co2rox	CO2 concentration
ch4rox	CH4 (Methane) concentration
n2orox	N2O concentration
f11rox	CFC11 concentration
f12rox	CFC12 concentration

Potential targets

→ Pristine-sky (neither clouds nor aerosols) or clear-sky (aerosols, but no clouds) conditions

→ Long- and short-wave radiative fluxes as well as heating rates

Output Variables	
rldc	Downward thermal (longwave) flux profile
rluc	Upward thermal flux profile
rsdc	Downward solar (shortwave) flux profile
rsuc	Upward solar flux profile
hrsc	Solar heating rate profile
hrlc	Thermal heating rate profile

Experiments

Our baselines

As in prior work:

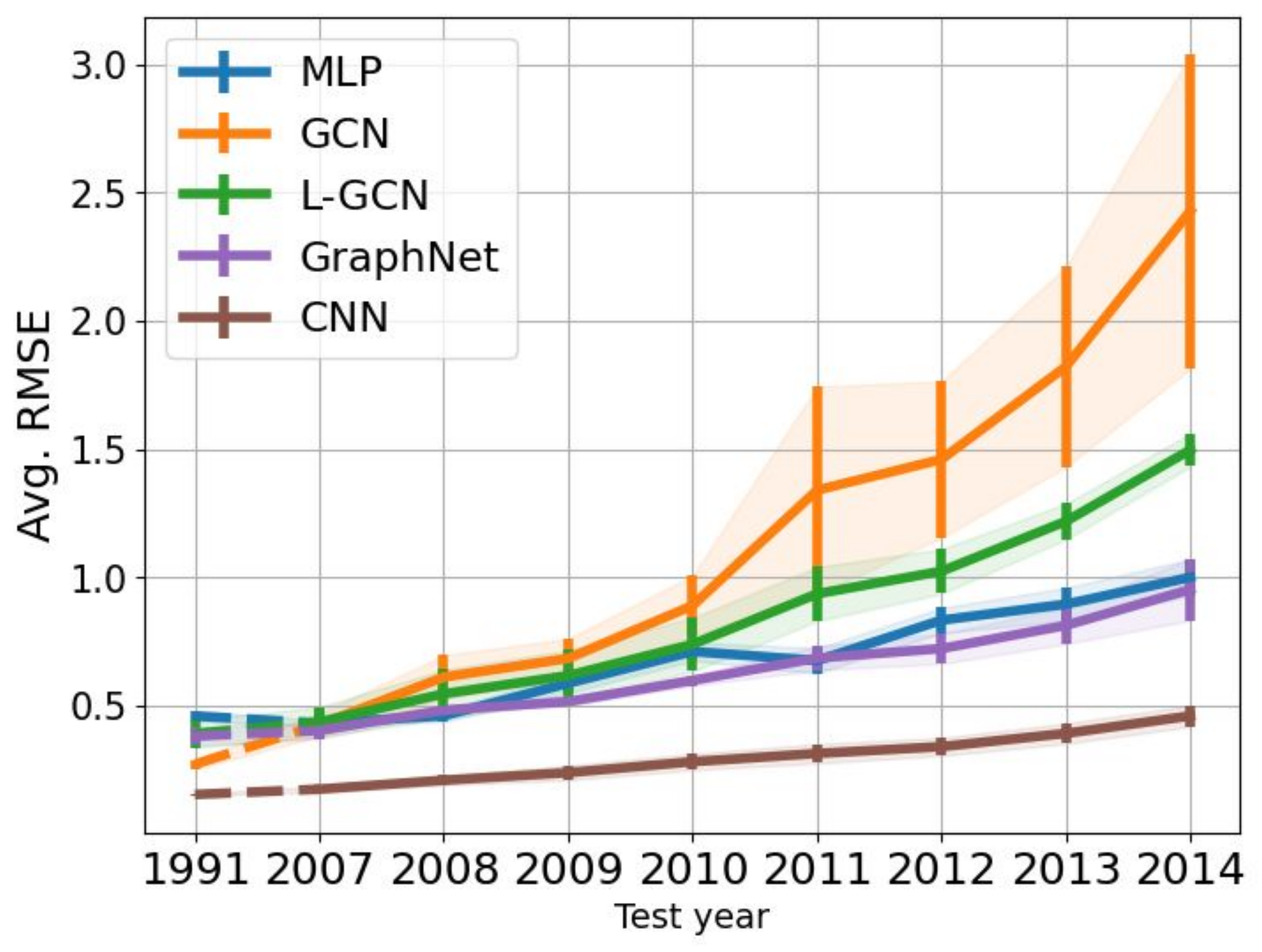
- Fully-connected net (MLP),

  as well as more structured models that we newly propose:

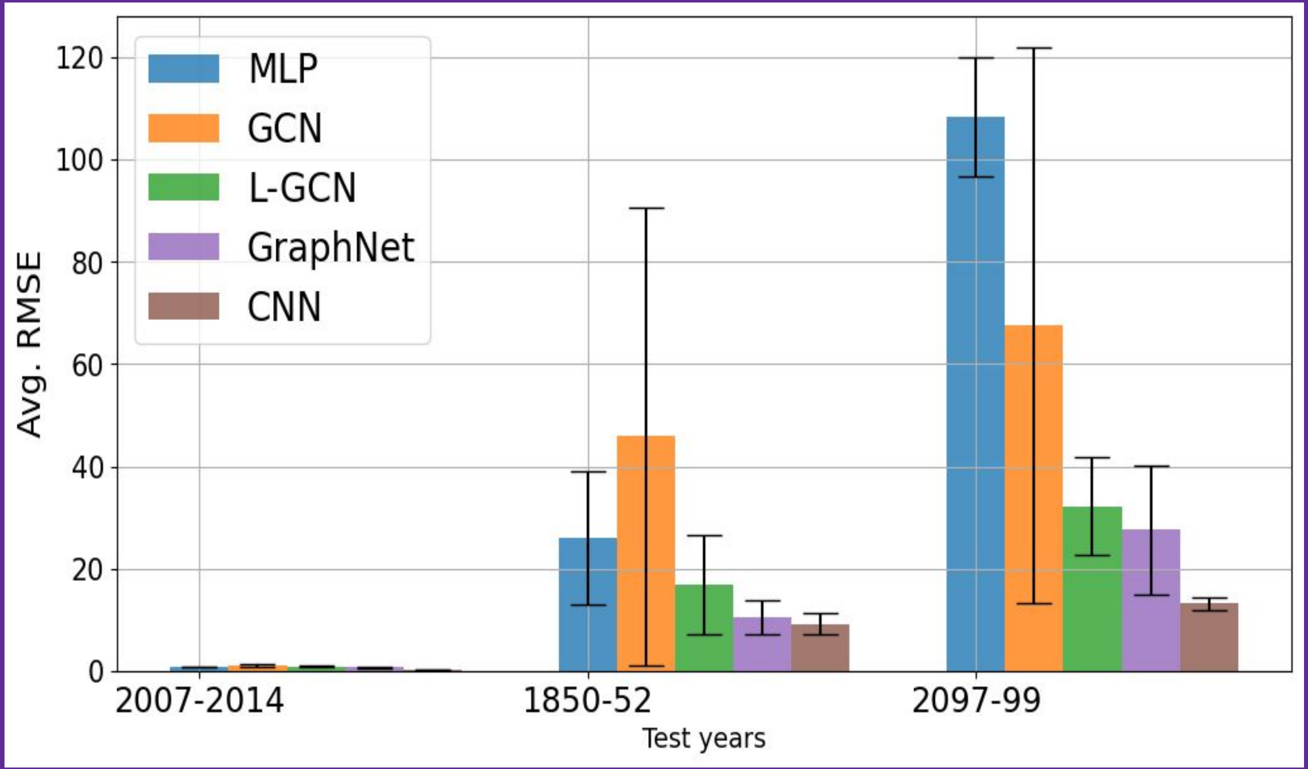
- Graph-based GCN and GraphNet
- Convolutional neural net (CNN)

Performance worsens as test year is farther away from training period

(1990, 1999, 2003)



Historical and future climate conditions pose a challenge



Important considerations

- Trade-off between accuracy/model complexity and speed
- Trained ML emulator needs to also be validated *on-line*, running jointly with the host weather/climate model
- Weather models require little random errors, but tolerate more bias errors
- Climate models require little bias errors, but tolerate more random errors

Thanks!

Salva Rühling Cachay*, Venkatesh Ramesh*, Jason N. S. Cole, Howard Barker, and David Rolnick.
ClimART: A benchmark dataset for emulating atmospheric radiative transfer in weather and climate models.
In Thirty-fifth Conference on Neural Information Processing Systems (NeurIPS) Datasets and Benchmarks Track, 2021.

Code: https://github.com/RolnickLab/climart

Paper: https://arxiv.org/abs/2111.14671

Contacts: salvaruehling@gmail.com and venka97@gmail.com