



# Verification of the ICON model with the GT4Py dycore - challenges and insights

Abishek Gopal, on behalf of EXCLAIM

9th Nov 2023 - NCAR Workshop on Correctness and Reproducibility

# Some context..



A unified next-generation global numerical weather prediction (NWP) and climate modeling system - originally developed by the Deutscher Wetterdienst (DWD) and the Max-Planck-Institute for Meteorology (MPI-M). Current partners: DWD, MPI-M, DKRZ, KIT, C2SM



EXCLAIM is an open ETHZ project that aims to develop an ICON-model based infrastructure that is capable of running kilometer-scale climate simulations at both regional and global scales.



GT4Py is a DSL compiler framework for weather and climate modeling, developed at CSCS. Stencil code is written in an abstracted Python based language and executed with high performance on different machine architectures, e.g. CPUs and GPUs.



# Acknowledgements

## Primary contributors towards the content of this talk

Christian Zeman<sup>1</sup>, Praveen Pothapakula<sup>1b</sup>, Anurag Dipankar<sup>1b</sup>, Christoph Müller<sup>3</sup>, Sebastian Schemm<sup>1</sup>

## Non-exhaustive list of EXCLAIM contributors past + present

N. Burgdorfer<sup>3</sup>, E. Cossevin<sup>3</sup>, T. Ehrenguber<sup>2</sup>, N. Farabullini<sup>1b</sup>, A. Gopal<sup>1b</sup>, L. Groner<sup>2</sup>, R. Häuselmann<sup>2</sup>, D. Hupp<sup>3</sup>, J. Jucker<sup>1b</sup>, A. Jocksch<sup>2</sup>, P. Kardos<sup>1b</sup>, S. Kellerhals<sup>1b</sup>, D. Leutwyler<sup>3</sup>, M. Luz<sup>1b</sup>, C. Müller<sup>3</sup>, C.R. Ong<sup>1b</sup>, C. Osuna<sup>3</sup>, E.G. Paredes<sup>1</sup>, P. Pothapakula<sup>1b</sup>, M. Röthlin<sup>3</sup>, W. Sawyer<sup>2</sup>, C. Schnadt Poberaj<sup>1b</sup>, F. Thaler<sup>2</sup>, H. Vogt<sup>2</sup>, B. Weber<sup>3</sup>, C. Zeman<sup>1</sup>, M. Bianco<sup>2</sup>, A. Dipankar<sup>1b</sup>, X. Lapillonne<sup>3</sup>

<sup>1</sup> Institute for Atmospheric and Climate Science (IAC), ETH Zurich

<sup>1b</sup> Center for Climate Systems Modeling (C2SM), ETH Zurich

<sup>2</sup> Swiss National Supercomputing Center (CSCS)

<sup>3</sup> Federal Office for Meteorology and Climatology, MeteoSwiss

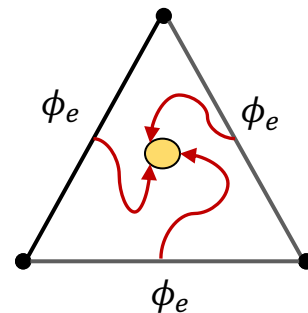
## Computational resources provided by CSCS

# GT4Py – A Domain Specific Language for Weather and Climate applications

```
!$ACC PARALLEL LOOP DEFAULT(PRESENT) GANG VECTOR COLLAPSE(2) ASYNC(1) IF(i_am_accel_node)
#ifdef __LOOP_EXCHANGE
DO jc = i_startidx, i_endidx
  DO jk = 1, nlev
#else
DO jk = 1, nlev
  DO jc = i_startidx, i_endidx
#endif
  z_temp(jc,jk,jb) =
    z_nabla2_e(ieidx(jc,jb,1),jk,ieblk(jc,jb,1))*p_int%geofac_div(jc,1,jb) + &
    z_nabla2_e(ieidx(jc,jb,2),jk,ieblk(jc,jb,2))*p_int%geofac_div(jc,2,jb) + &
    z_nabla2_e(ieidx(jc,jb,3),jk,ieblk(jc,jb,3))*p_int%geofac_div(jc,3,jb)
  ENDDO
ENDDO
!$ACC END PARALLEL LOOP
```

```
@field_operator
def _calculate_nabla2_of_theta(
    z_nabla2_e: Field[[EdgeDim, KDim], float],
    geofac_div: Field[[CEDim], float],
) -> Field[[CellDim, KDim], float]:
    z_temp = neighbor_sum(z_nabla2_e(C2E) * geofac_div(C2CE), axis=C2EDim)
    return z_temp
```

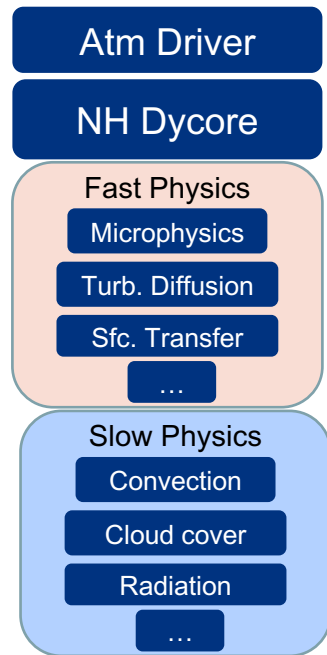
Hor. Divergence in ICON



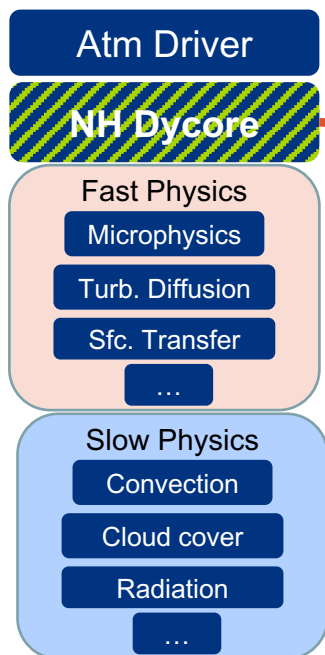
$$\tilde{\nabla}(\phi_e) = \frac{1}{a_c} \sum_{i=1}^3 l_p f_o \phi_e$$

# ICON and EXCLAIM development lines

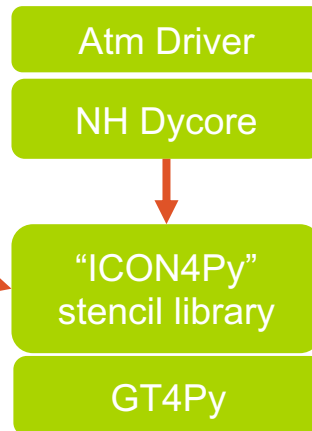
DWD ICON



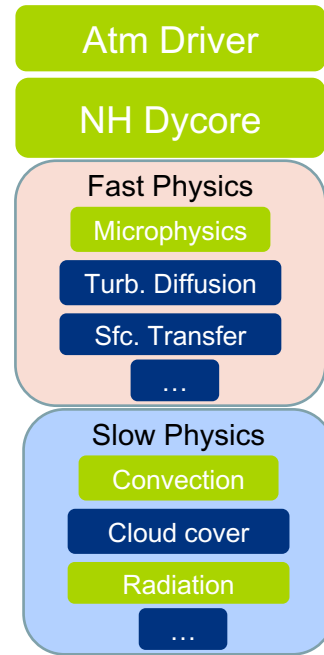
ICON-GT4Py dycore



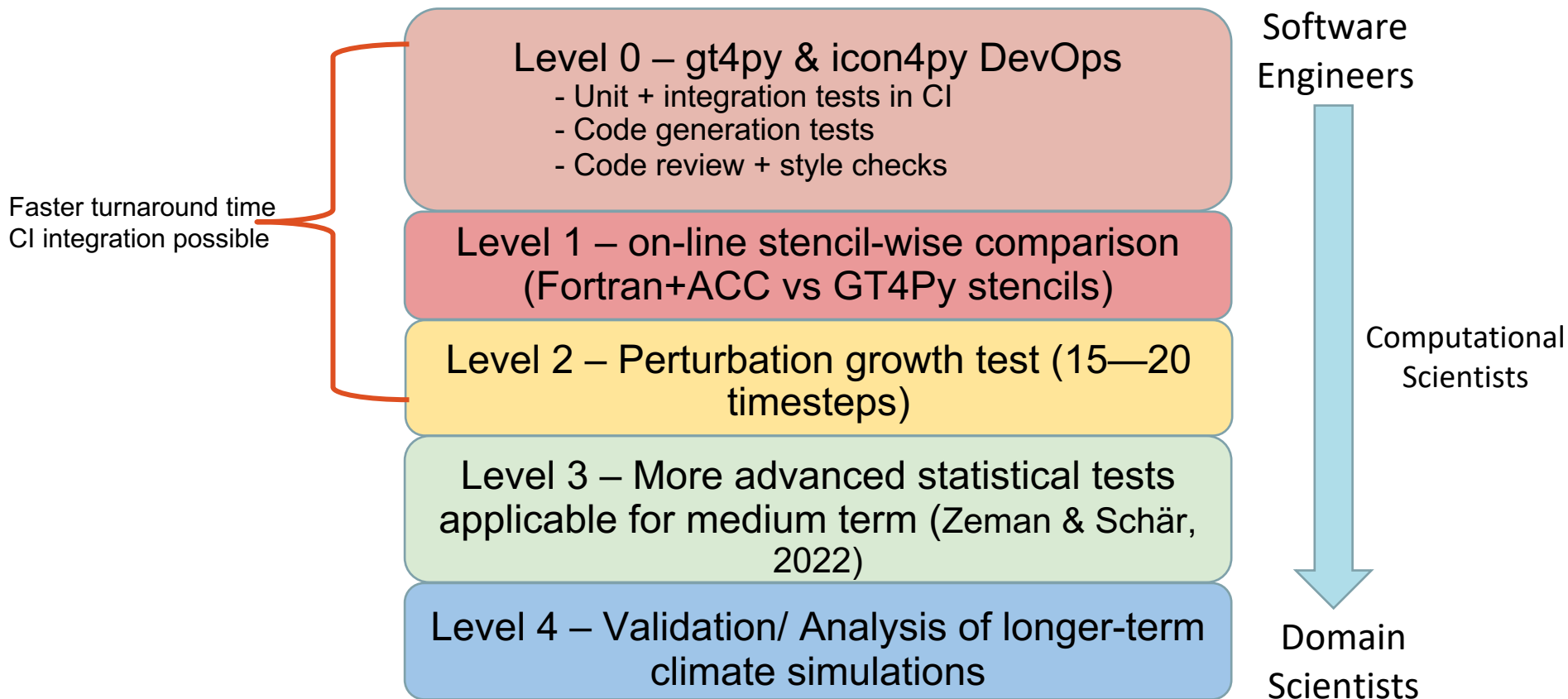
Python driver  
(Experimental)



Soon...



# Model Verification + Validation stack



# Some challenges

- **Separation of concerns** in the DSL context, between software engineers and domain scientists, can create gaps in understanding and model verification
  - Better communication between the groups is key!
  - Computational scientists and folks with a numerical methods background can help bridge some of this gap.
- Verification methods need to be **automatable & relatively cheap** to be useful to the CI workflow of a fast-paced software engineering effort
  - Compromise between the robustness of a method and the human/compute times
  - Methods adapted to CI can sometimes end up becoming black boxes that are not well understood
- The ICON consortium is large, and we have to work with what we have re. development and testing best practices

# Level 1 – Online verification of ICON-GT4Py dycore

```
!$DSL START STENCIL(name=calculate_nabla2_of_theta; z_nabla2_e=z_nabla2_e(:, :, 1); z_temp=z
!$DSL          geofac_div=p_int%geofac_div(:, :, 1); vertical_lower=1; vertical_upper=n
!$DSL          horizontal_lower=i_startidx; horizontal_upper=i_endidx)

!$ACC PARALLEL LOOP DEFAULT(PRESENT) GANG VECTOR COLLAPSE(2) ASYNC(1) IF(i_am_accel_node)
#ifdef __LOOP_EXCHANGE
DO jc = i_startidx, i_endidx
  DO jk = 1, nlev
#else
DO jk = 1, nlev
  DO jc = i_startidx, i_endidx
#endif
  z_temp(jc,jk,jb) =
    z_nabla2_e(ieidx(jc,jb,1),jk,ieblk(jc,jb,1))*p_int%geofac_div(jc,1,jb) + &
    z_nabla2_e(ieidx(jc,jb,2),jk,ieblk(jc,jb,2))*p_int%geofac_div(jc,2,jb) + &
    z_nabla2_e(ieidx(jc,jb,3),jk,ieblk(jc,jb,3))*p_int%geofac_div(jc,3,jb)
  ENDDO
ENDDO
!$ACC END PARALLEL LOOP

!$DSL END STENCIL(name=calculate_nabla2_of_theta)
```

## ICON-GT4Py Build Modes

### Verification mode:

- execute both DSL and Fortran + OpenACC versions
  - Compare DSL results with Fortran at runtime
  - Serialize fields if tolerance exceeded
- $$\text{abs}(a - b) \leq (atol + rtol * \text{abs}(b))$$

### Substitution mode:

- Only execute DSL version

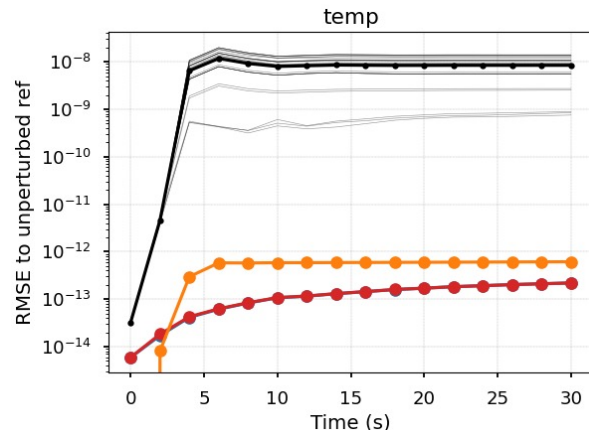
Stencil-wise verification framework developed or maintained by:

*Giacomo Serafini, Christoph Müller, Matthias Röthlin,  
Ben Weber, Sam Kellerhals, Daniel Hupp*



# Level 2 – Perturbation growth test

- Does growth in error of ported code fall within the expected error growth of initial perturbations in the reference code? \*
- We use **proptest**<sup>+</sup> integrated in our Jenkins CI
- Requires some domain knowledge and numerics to configure test for each scenario/case
- Physical parametrization makes this test less useful
  - Could combine with dycore-only tests to minimize false negatives
- Also sensitivity to timestep size and

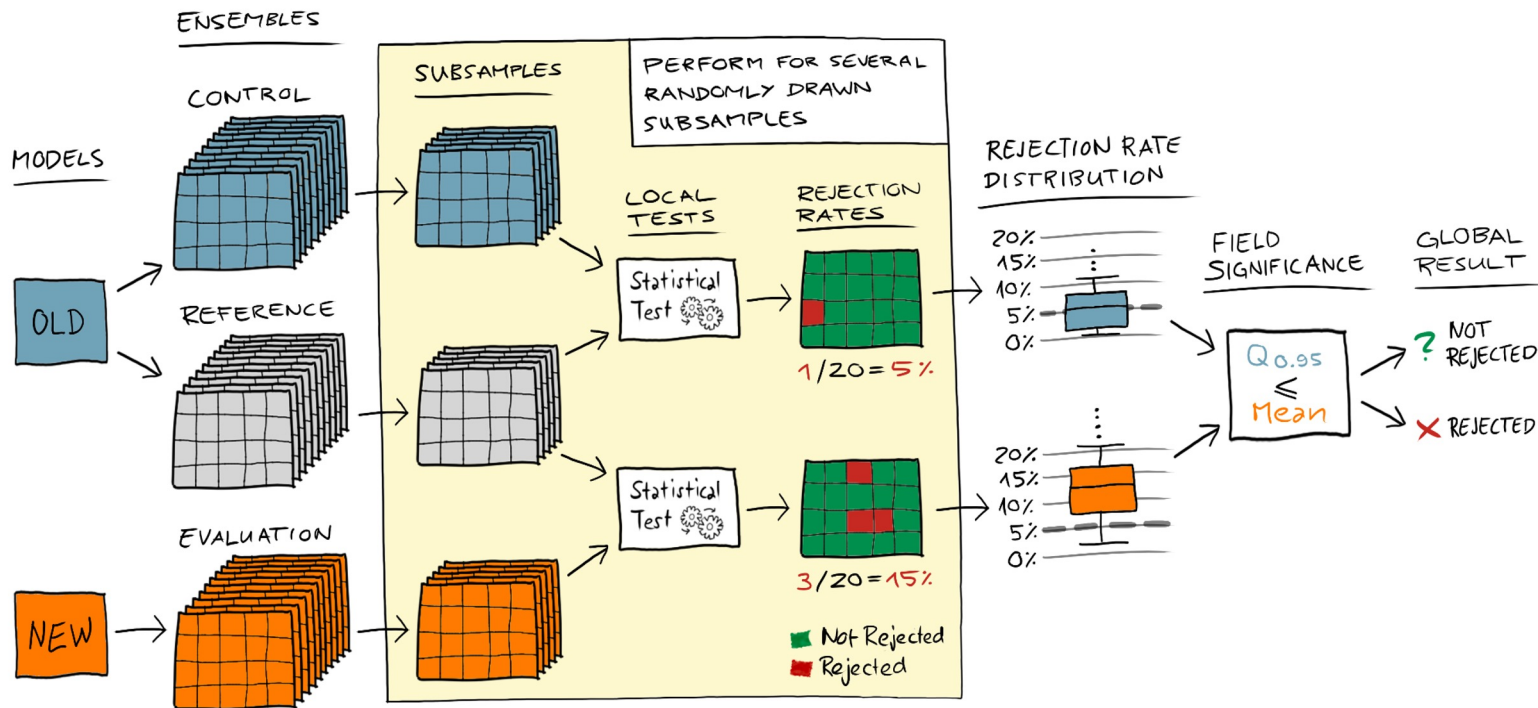


<https://github.com/MeteoSwiss/proptest>

Remo Dietlicher, Fabian Gessler  
Daniel Hupp, Marek Jacob

\*Rosinski, J. M., & Williamson, D. L. (1997). *The accumulation of rounding errors and port validation for global atmospheric models*. SIAM Journal on Scientific Computing, 18(2), 552-564.

# Level 3 – Ensemble-based statistical method

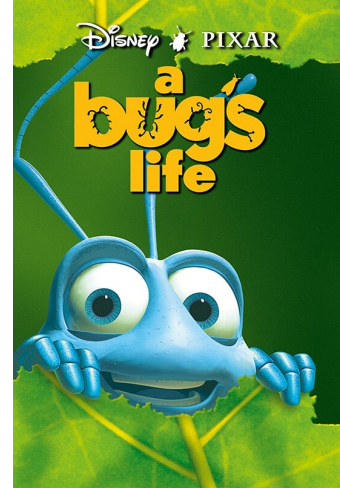


Zeman, C., & Schär, C. (2022). *An ensemble-based statistical methodology to detect differences in weather and climate model executables*. Geoscientific Model Development, 15(8), 3183-3203.

# A Case Study

## The Task

Verify our **ported ICON-GT4Py dycore** for an **aquaplanet** simulation, while also switching to a recently GPU-ported radiation library (ecRad)



## Complicating factors

- Needed to update ICON-GT4Py with upstream ICON to benefit from the newly ported ecRad
- Limited understanding of the perturbation growth test (among SW developers)
- Prior to this, we tested model with a Limited Area Model run (over Switzerland) for tests, which was using a different radiation module (RRTM)
- We couldn't switch ecRad, a standalone module, back to running on CPUs only

# Verification & debugging workflow

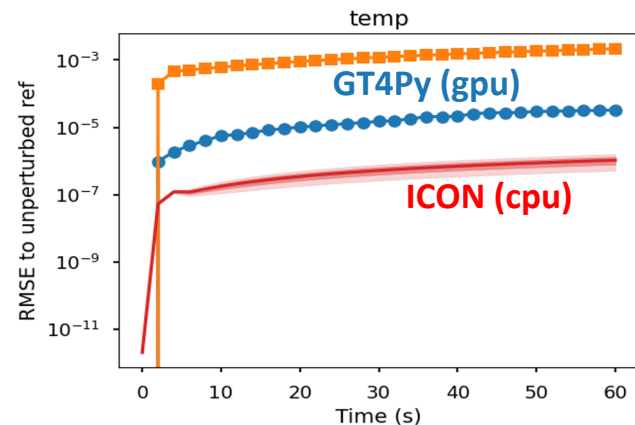
## Probstest & post hoc theorization

- Model runs from cold start failed probtest, but restarting from a 3-month spinup helped.
- Wrongly assumed the model run was correct, and just an issue with the cold initialization

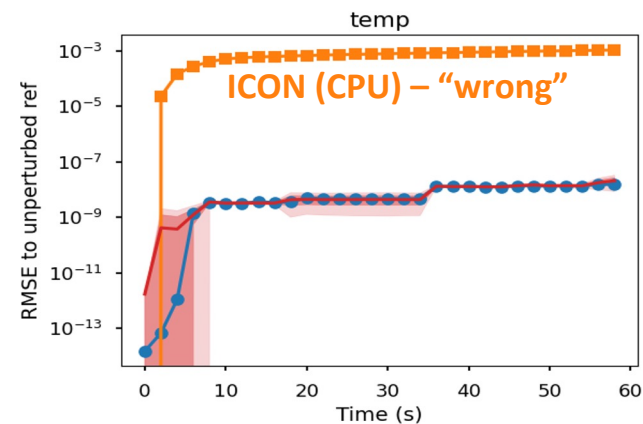
- Missed potential issues by not checking all quantities of interest

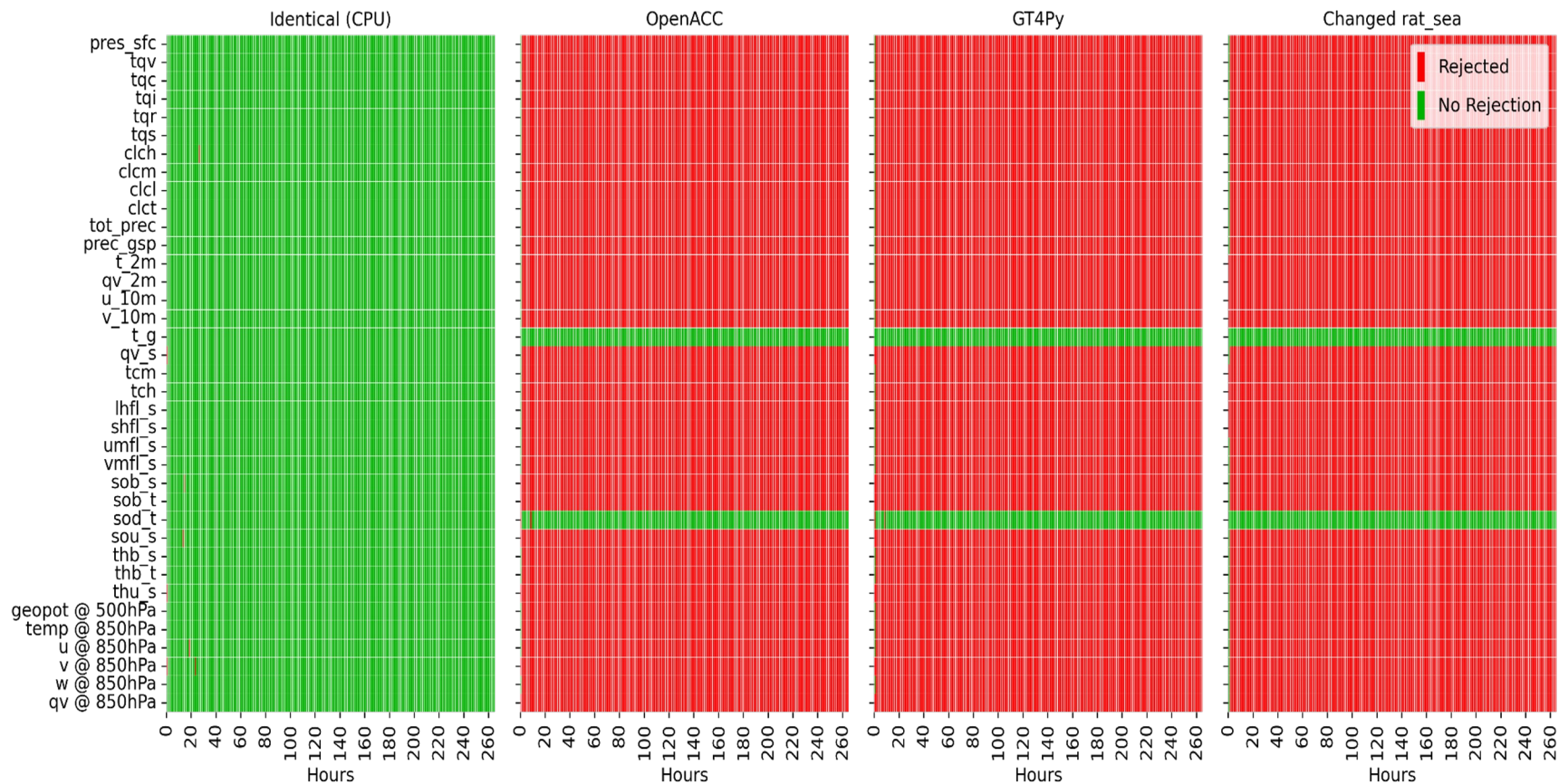
- The Zeman test provided the first indications that there was a potential bug

## From cold start



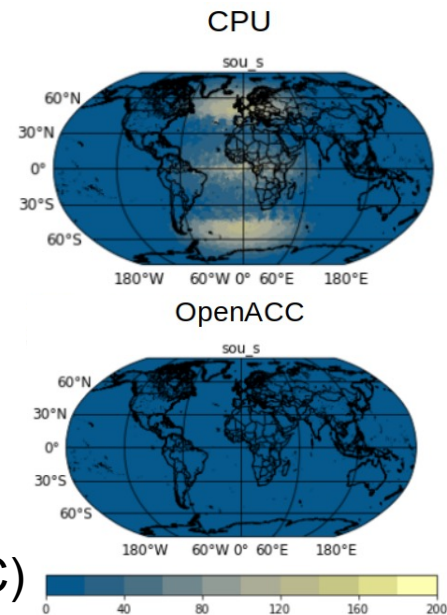
## After spin-up





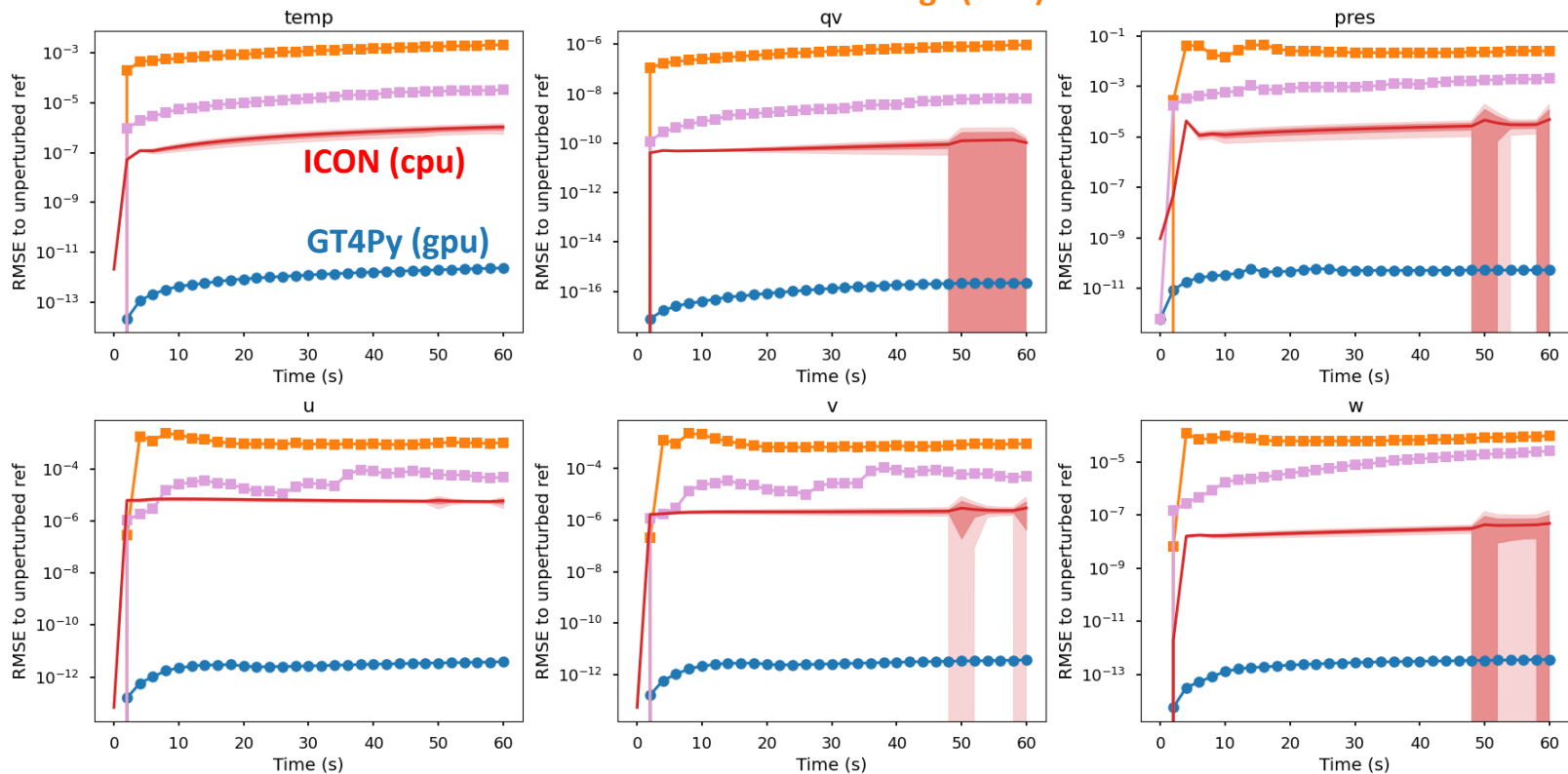
# Debugging process (contd.)

- Plotting further quantities revealed **sou\_s** (surface shortwave upward flux) was 0 in GPU runs
- Initial understanding: may just be an issue with the diagnostic variables, and not feeding back to prognostic state
- Further investigations revealed that ICON-GPU (OpenACC) was having the same issues
- Finally we discover **the bug**:
  - An interface between ICON and ecRad computes the surface albedo
  - A particular namelist option for using fixed surface albedo was not ported to GPU, and was missing a runtime exception usually present when namelist options are used before they are ported



# After bug fix

ICON "wrong" (CPU)





# After bug fix

