AEGIS: DIFFERENTIABLE MARS CLIMATE MODEL WITH NEURAL CLOSURES

Anonymous authors
Paper under double-blind review

ABSTRACT

General circulation models (GCMs) are the primary tool for simulating planetary atmospheres. They play a vital role in understanding Mars's atmosphere, as forecasting its unique weather is mission-critical for operations such as entry, descent, and landing. Mars poses unusual challenges for these models, as observations are sparse compared to Earth. In addition, a thin CO_2 atmosphere alongside a radiatively active dust cycle creates a volatile atmosphere with large diurnal temperature swings and no true terrestrial analog for validation. Existing Mars GCMs, including the LMD PCM, the NASA Ames Mars GCM, and PlanetWRF, are mature and physically detailed but are implemented in legacy Fortran with finite-difference or finite-volume solvers, and they do not expose gradients for calibration or machine learning. Here we present AEGIS, a modular differentiable Mars climate model that couples Mars's unique atmospheric physics to the Dinosaur dynamical core, with interfaces for neural closures. We showcase stable ten-Mars-year simulations that reproduce the seasonal CO_2 cycle while conserving the total CO_2 inventory, capture realistic large-scale surface-temperature structure, and produce surface pressure that follows Mars Orbiter Laser Altimeter (MOLA) topography. Gradients through coupled trajectories agree with finite differences and support physical calibration and neural training. We compare with conventional GCMs, highlighting the framework's computational efficiency and differentiability.

1 INTRODUCTION

General circulation models (GCMs) connect the governing equations of atmospheric motion to the physical processes that shape planetary climate. Established Mars models include the LMD Mars GCM, the NASA Ames Mars GCM, and PlanetWRF (Forget et al., 1999b; Haberle et al., 1993; Richardson et al., 2007). On Mars, radiative heating by dust and CO_2, exchange with a thermally active surface, and seasonal condensation onto polar caps jointly shape circulation, pressure, and frost. A useful model must represent these interactions while allowing individual assumptions to be tested (Kahre & Haberle, 2010).

Traditional GCMs contain uncertain parameterization coefficients whose coupled effects can complicate tuning and permit compensating errors (Forget et al., 1999b). Parameter sweeps and ensemble calibration are often computationally expensive (McNeall et al., 2020). In addition, the forward-model interface does not generally expose gradients, and maintaining separate adjoint implementations incurs substantial development effort (Giering & Kaminski, 1998). Introducing a learned component can also require substantial changes to the coupling between numerical and physical routines. These challenges motivate models in which physical simulation and gradient optimization share the same implementation (Davenport et al., 2026).

Differentiable programming bridges this gap. When a simulation is implemented as differentiable numerical operations, automatic differentiation applies the chain rule through its computational graph to obtain derivatives of outputs with respect to inputs and parameters or learned weights (Baydin et al., 2018). This supports calibration, sensitivity analysis, inverse problems, and training over multiple time steps. In particular, a neural parameterization can be optimized for its effect on a coupled trajectory, accounting for feedback between the learned tendency and the resolved dynamics (Rasp, 2020; Kashyap et al., 2024). NeuralGCM demonstrates the potential of this approach for Earth's weather and climate by coupling learned parameterizations to a differentiable atmospheric dynamical

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

core (Kochkov et al., 2024). Its Dinosaur core solves the hydrostatic primitive equations using spectral horizontal coordinates and a sigma vertical coordinate. The JAX Circulation Model (JCM) uses the same core with an explicit suite of intermediate-complexity physical parameterizations, providing a setting for process-based studies and hybrid modeling for Earth's climate and atmosphere (Gelbrecht et al., 2023; Davenport et al., 2026). In a different physical domain, JAX MD illustrates how reusable numerical components and neural models can be composed within a common differentiable simulation environment (Schoenholz & Cubuk, 2020). AEGIS shares the Dinosaur backbone with JCM and NeuralGCM, but addresses a different physical regime and adds coupled state variables and operators needed for Mars.

In this work, we present AEGIS, a differentiable Mars Climate Model based on the Dinosaur Core. Unlike the Earth-focused configurations of its Dinosaur-based predecessors, AEGIS uses planetary radius, rotation, and thermodynamic scales at the dynamical interface. It adds Mars radiation and surface physics, evolves frost and regolith alongside the atmosphere, feeds CO_2 phase change into the dynamical core's prognostic log-pressure state, and projects the coupled atmosphere–frost inventory after each step. These additions make mass exchange and surface memory part of the differentiated trajectory.

We make three contributions:

1. We extend the Dinosaur core with planetary-unit rescaling, a coupled atmosphere–surface–frost–regolith state, CO_2 mass exchange into prognostic pressure, and an atmosphere–frost inventory projection.
2. We provide modular Mars radiation, surface, transport, and neural interfaces through which physical controls and learned components can enter the same coupled simulation.
3. We verify physical and neural gradients, evaluate Mars fields against reference products, and demonstrate calibration, year-long forward integration, and controlled neural experiments, and showcase them in Section 3.

2 METHODS

AEGIS couples a modular Mars physics suite to the Dinosaur spectral dynamical core, carrying the atmosphere, surface, CO_2 frost, and regolith in a single differentiable state (Section 2.1). Automatic differentiation through this coupled time step provides gradients for physical calibration and for training learned components that enter through a common neural interface (Section 2.2).

2.1 ATMOSPHERIC CONFIGURATIONS AND THE DYNAMICAL FRAMEWORK

AEGIS separates the description of a planetary atmosphere from the numerical machinery used to evolve it. A body configuration specifies radius a, gravity g, rotation rate Ω, specific gas constant R, heat capacity c_p, and reference temperature and pressure. We use Mars values for the planetary constants (NASA Goddard Space Flight Center, 2025). The dynamical equations consume these values through a common interface. Length and time are scaled by a and $(2\Omega)^{-1}$, respectively, so the nondimensional radius and rotation rate follow Dinosaur's conventions for each body, removing dependence on Earth-specific default scales while retaining the existing spectral operators.

The framework supports a hierarchy of atmospheric configurations. Dynamical configurations isolate resolved transport; physically forced configurations add radiation, surface exchange, mixing, and volatile reservoirs; hybrid configurations add a learned tendency or substitute a learned component at a selected physical interface. Complexity is therefore introduced through explicit components: a body definition establishes the mechanical and thermodynamic constants, while its radiation, condensables, and boundary conditions establish the atmospheric regime. The physically developed configuration evaluated here is Mars; extending to another atmosphere involves supplying the relevant process operators as well as its constants (Hu et al., 2025).

Dinosaur advances the hydrostatic primitive equations on a sphere (Kochkov et al., 2024; Davenport et al., 2026). The atmospheric state contains spectral vorticity and divergence, temperature relative to a reference profile, log surface pressure, optional transported tracers, and simulation time, on a vertical coordinate $\sigma = p/p_s$ that follows changes in atmospheric mass and the terrain-dependent lower boundary (Richardson & Newman, 2018). Spectral transform models using sigma coordinates

have a long history in Mars atmospheric modeling (Nayvelt et al., 1997). The main experiments use triangular truncation T21, a 64×32 longitude–latitude transform grid, and 12 equally spaced sigma layers.

The coupled state augments the atmospheric variables with surface temperature, pressure-equivalent CO_2 frost, and a vertical regolith temperature profile attached to their geographic columns. Writing the full state as $\boldsymbol{x}$, the coupled equations take the form

$$\frac{d\boldsymbol{x}}{dt} = \mathcal{D}_{\mathrm{imp}}(\boldsymbol{x}) + \mathcal{D}_{\mathrm{exp}}(\boldsymbol{x}) + \mathcal{P}_{\boldsymbol{\alpha}}(\boldsymbol{x}, t) + \mathcal{N}_{\boldsymbol{\theta}}(\boldsymbol{x}, t), \tag{1}$$

where $\mathcal{D}$ denotes resolved dynamics, $\mathcal{P}_{\boldsymbol{\alpha}}$ denotes physical parameterizations with parameters $\boldsymbol{\alpha}$, and $\mathcal{N}_{\boldsymbol{\theta}}$ denotes an optional neural tendency. A neural replacement for a physical component is composed inside $\mathcal{P}$ instead of being added a second time. Time integration uses Dinosaur's implicit–explicit SIL3 Runge–Kutta scheme (Whitaker & Kar, 2013), with linear gravity-wave terms treated implicitly and nonlinear dynamics and assembled physical tendencies on the explicit side. Planetary unit conversion, the augmented state, and the atmosphere–frost correction are implemented in AEGIS's surrounding operators.

The Mars operators in $\mathcal{P}_{\boldsymbol{\alpha}}$ span CO_2 and dust radiation, surface and regolith energy exchange, boundary-layer mixing and dry convection, and a CO_2 condensation cycle (Lacis & Oinas, 1991; Forget et al., 1999a; Haberle et al., 2019; Petrosyan et al., 2011). The last is the distinctive Mars coupling: surface condensation and sublimation transfer mass between the atmosphere and frost, feeding the prognostic log surface pressure,

$$\left.\frac{d \log p_s}{dt}\right|_{\mathrm{phase}} = -\frac{1}{p_s}\frac{dI}{dt}, \tag{2}$$

Here, p_s is the local surface pressure and I is the pressure-equivalent atmospheric mass inventory added by surface condensation–sublimation. Thus, Thus, $dI/dt > 0$ denotes net sublimation of surface CO_2 frost into the atmosphere and increases p_s, whereas $dI/dt < 0$ denotes net condensation from the atmosphere onto the surface and decreases p_s. The factor $1/p_s$ converts this pressure tendency to the tendency of the prognostic variable $\log p_s$. $I = g m_{\mathrm{ice}}$ is frost mass per unit area in pressure units, and a per-step projection matches the atmosphere-plus-frost inventory to its pre-step value while allowing seasonal exchange. Because CO_2 is both a radiative gas and a substantial mobile mass reservoir, carrying its phase change through the prognostic pressure state, alongside surface and subsurface temperatures (Kelly et al., 2006), makes this coupling part of the simulated system and its derivatives. The radiation, surface, boundary-layer, and CO_2-inventory operators are detailed in Appendix A.

2.2 Differentiation and neural components

Let $\Phi_{\boldsymbol{\alpha},\boldsymbol{\theta}}$ denote the operator that propagates a single time step, including the reservoir corrections. A trajectory is the repeated composition of $\boldsymbol{x}_{n+1} = \Phi_{\boldsymbol{\alpha},\boldsymbol{\theta}}(\boldsymbol{x}_n)$ while a loss function can be evaluated over an entire trajectory rollout as $\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\theta}) = \ell\big(\mathcal{H}(\boldsymbol{x}_1), \boldsymbol{y}_1, \mathcal{H}(\boldsymbol{x}_2), \boldsymbol{y}_2, \ldots, \mathcal{H}(\boldsymbol{x}_N), \boldsymbol{y}_N\big)$, where $\mathcal{H}$ extracts evaluated fields and $\boldsymbol{y}_n$ denotes the target data at step n. JAX differentiates this discrete numerical trajectory using automatic differentiation. We use forward-mode differentiation for the three physical controls and reverse-mode differentiation for neural weights; forward mode efficiently evaluates derivatives with respect to a small number of controls, whereas reverse mode is appropriate for a scalar loss with many parameters (Baydin et al., 2018; JAX Developers, a). Reverse-mode training through coupled rollouts, as in the heating experiment below, uses fixed-length scans with step rematerialization, which trades additional computation for reduced reverse-mode memory use by recomputing selected forward intermediates during the backward pass (JAX Developers, b).

Learned components are exposed through two adapters that share one interface: a model is initialized from a random key into a JAX parameter tree, input normalization is fitted on training data only, and a `bind` call returns a pure JAX function that is passed to the coupled equations when the time step is built. The residual adapter adds a bounded heating tendency computed from current-state column features: the temperature profile, surface temperature and pressure, lowest-layer winds, geographic coordinates, and insolation, 21 features at T21/L12, as the equation

$$\dot{T}_{\mathrm{NN}} = h_{\max} \tanh f_{\boldsymbol{\theta}}(\boldsymbol{z}), \tag{3}$$

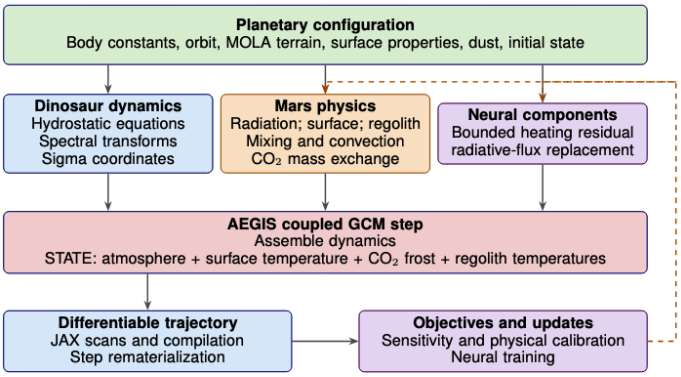


Figure 1: AEGIS architecture. Planetary constants and boundary data configure a shared GCM. Dinosaur supplies resolved dynamics; Mars physical operators supply column tendencies and surface evolution; neural components enter as bounded residuals or component replacements. The coupled time step advances atmosphere and surface reservoirs. Trajectory objectives connect the simulated fields to physical-parameter calibration and neural training through automatic differentiation. Dashed arrows denote this gradient-based update path.

where $f_{\boldsymbol{\theta}}$ is the network, $\boldsymbol{z}$ the normalized column features, and $h_{\max}$ a fixed heating bound, converts it to the dynamical core's tendency units, and adds it to the physical tendencies while leaving the surface reservoirs under physical control; a zero-initialized output layer reproduces the physical model exactly at the start of training. The replacement adapter is a drop-in radiation component: it predicts interface fluxes from the column temperature and pressure profiles, dust, sunlight, albedo, and emissivity, bounds shortwave fluxes by the incoming sunlight, keeps longwave fluxes positive, imposes the top and surface boundary conditions exactly, and computes heating from flux convergence, so the column energy budget closes for any network weights. Both adapters accept a custom network through an `apply(params, features)` function and check the vertical grid and radiation configuration when bound. Because a bound component is part of the coupled step, its parameters can be trained on local targets or by reverse-mode differentiation through rollouts, and they are saved together with their normalization and metadata in validated checkpoints. Training protocols for the reported experiments are given in Appendix B.

3 Results

3.1 Verified gradients through the coupled Mars model

We verify derivatives through both a scalar atmospheric diagnostic and the objectives used for physical calibration and neural fitting. The physical derivative experiment differentiates final area- and layer-mean atmospheric temperature with respect to the three physical controls. It compares automatic differentiation (AD) with centered finite differences (FD) over six rollout horizons, from 1 to 148 steps, corresponding to 300–44,400 s at a 300 s timestep. For each control and horizon, FD estimates use perturbations 10^{-3}, 10^{-4}, and 10^{-5}, yielding 54 comparisons through the coupled dynamics and Mars physics given by Table 1. All 54 physical-sensitivity checks pass the recorded criterion. At the longest horizon, $44{,}400$ s, the largest discrepancy is 3.56×10^{-7}. Agreement across perturbation sizes and integration horizons supports the use of the computed derivatives for physical-parameter sensitivity in these coupled simulations.

We extend this verification to multi-day rollouts and to several coupled diagnostics at once. Starting from a MY32, $L_s = 45^\circ$ ARCO-MACDA window, we compute the forward-mode Jacobian `jax.jacfwd` of global-mean surface temperature, surface pressure, and surface CO_2 frost with respect to the CO_2 longwave opacity multiplier β at its nominal value $\beta = 1$, through the complete coupled rollout at horizons of 0.25, 1, and 5 sols (74, 296, and 1480 steps). All nine derivatives agree with centered finite differences ($\epsilon = 10^{-3}$), with a maximum relative discrepancy of 3.99×10^{-6} (Fig. 2). The sensitivities are physically interpretable: increasing CO_2 longwave opacity warms the surface, with $\partial \bar{T}_s / \partial \beta$ growing from 0.488 to 0.785 and 0.945 K per unit β as the response

Table 1: Derivative checks through coupled AEGIS trajectories. All comparisons satisfy $|g_{\mathrm{AD}} - g_{\mathrm{FD}}| \leq 10^{-8} + 10^{-4}|g_{\mathrm{FD}}|$.

Category	Objective	Horizon	Checks	Passed
Physical sensitivity	Final mean atmospheric temperature	1–148 steps	54	54/54
Physical calibration	Scalar fitting objective	task-dependent	9	9/9
Neural fitting	Multi-field trajectory objective	25 steps	3	3/3
Multi-diagnostic sensitivity	Mean surface T, p_s, and CO_2 frost	74–1480 steps	9	9/9

accumulates, and the warmer surface suppresses condensation, so surface pressure rises while frost falls ($\partial \bar{p}_s / \partial \beta$ grows from 2.94×10^{-3} to 4.66×10^{-2} Pa). The pressure and frost derivatives are equal and opposite at every horizon, with their sum at most 9.3×10^{-11} Pa, so the gradient itself respects the atmosphere–frost mass inventory that the projection enforces on the state. A warm evaluation of the diagnostics and their Jacobian over the 5-sol rollout takes about 92 s on a single GPU, showing that exact sensitivities of the coupled mass exchange are available over multi-day horizons at modest cost.

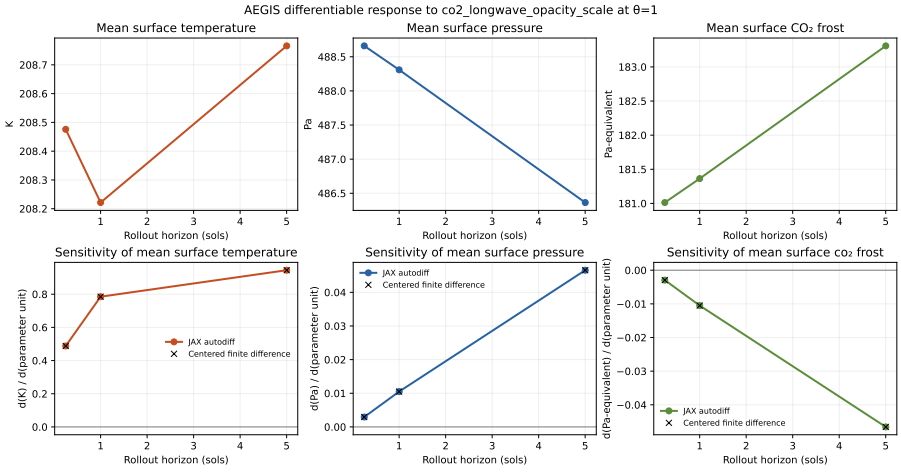


Figure 2: Differentiable response of AEGIS to the CO_2 longwave opacity multiplier β at $\beta = 1$, from a MY32, $L_s = 45°$ window. Top: global-mean surface temperature, surface pressure, and surface CO_2 frost at rollout horizons of 0.25, 1, and 5 sols. Bottom: their derivatives with respect to β computed by forward-mode automatic differentiation through the full coupled rollout (lines) and by centered finite differences (crosses). The pressure and frost sensitivities are equal and opposite, reflecting conservation of the atmosphere–frost CO_2 inventory.

Gradient-based calibration improves atmospheric temperature RMSE on each data split in the twelve-window pilot (Table 2). Training RMSE decreases from 2.45 to 2.38 K, validation RMSE from 2.49 to 2.44 K, and test RMSE from 2.46 to 2.40 K. The unrounded test reduction is about 0.06 K (2.4%), and all four test windows improve. This per-window consistency demonstrates that differentiating the coupled trajectory loss yields a reliable, direction-correct optimization signal across seasons.

3.2 Computational performance and long integrations

From our performance analysis, a full Mars year (668.6 sols; T21, 12 sigma layers, $\Delta t = 450$ s, float64) takes 131,900 steps and completes in 86.6 minutes on a single GPU, or about 463 simulated sols per wall-clock hour. This timing includes compilation, checkpointing, and diagnostics. The NASA Ames legacy Mars GCM documentation reports about 20 sols per hour for its tutorial configuration (60×36 grid, 24 layers, $\Delta t = 120$ s, hardware not specified), or roughly 33 h per Mars year. While resolution, physics, and hardware differ between the two models, AEGIS is significantly faster on a consumer GPU. All 68 saved states are finite, and the total atmosphere–frost CO_2 inventory is conserved to a maximum relative drift of 2.8×10^{-9}. The model reproduces the seasonal CO_2

Table 2: Twelve-window reanalysis pilot: atmospheric temperature RMSE (K) at a single lead of 7397.937 s. Each split contains four seasonally distributed windows, combined as the square root of their mean squared errors. The physical baseline is shared by the two fitting experiments.

Configuration	Training	Validation	Test
Physical baseline	2.45	2.49	2.46
Calibrated physics	2.38	2.44	2.40
Coupled neural heating	2.42	2.49	2.46

cycle: global-mean surface pressure falls to 485 Pa at $L_s \approx 146^\circ$ and peaks at 654 Pa at $L_s \approx 235^\circ$, a 30% annual range, while surface frost peaks at 185 Pa. The integration is stable across the physics configurations as well: all thirteen annual ablation runs complete with finite states and inventory drift below 8.6×10^{-9}. These runs remove boundary-layer mixing, convection, regolith conduction, or CO_2 exchange, weaken diffusion, halve the time step, or change each calibrated multiplier by $\pm 20\%$.

We also conduct a long-horizon stability analysis continuing with the same configuration ($\Delta t = 450$ s, float64), a ten-Mars-year run (6686 sols, 1,319,003 steps) completes in 11.4 h on one GPU (586 sols per hour, about 68 minutes per Mars year), with all 670 saved states finite and a maximum relative CO_2 inventory drift of 2.9×10^{-8}. By the tenth year the seasonal cycle repeats closely: global-mean surface pressure and surface temperature differing negligibly (0.16 Pa and 0.20 K) from year nine.

3.3 Seasonal comparisons with MOLA topography

The MOLA-based lower boundary produces a strong geographic structure in surface pressure (Smith et al., 2001; NASA Planetary Data System, n.d.). Elevated regions contain less atmospheric column mass, while low-lying terrain supports higher surface pressure. Across our twelve season–reference comparisons, surface-pressure pattern correlations range from 0.975 to 0.989 (Table 3). The model reproduces the terrain-driven hydrostatic pressure structure set by the MOLA lower boundary across all four seasons and all three reference products. Figure 3 shows the terrain context and the seasonal fields at $L_s = 135^\circ$.

Surface-temperature patterns also agree across the reference products, with correlations of 0.732–0.821. Surface frost provides a complementary diagnostic of the atmosphere–surface coupling. RMSE values are provided in Table 3. These fields help identify the processes and sampling choices that matter for subsequent calibration.

The seasonal comparison uses the documented sampling of each product: sparse model checkpoints within $\pm 5^\circ$ of the target season, five-sol Ames averages, daily ARCO-MACDA averages, and four-local-time MCD means. Wind levels and frost averaging periods also differ between products. We interpret these as model and reanalysis comparisons.

3.4 Physics ablations: process sensitivity

We experiment with an ablation suite that pairs the full Mars configuration with twelve process perturbations, such as removing boundary-layer mixing, dry convection, regolith conduction, or CO_2 exchange, CO_2 longwave opacity, and dust longwave opacity. Each is launched from the same initial state, so any divergence traces to the changed process, and we report the root-mean-square (RMS) separation from the paired full-physics trajectory. At a 44,400,s horizon, the perturbations separate into a well-resolved sensitivity ranking (Fig. 4): regolith conduction and boundary-layer mixing dominate, followed by dry convection and the surface-exchange multipliers, while the calibrated radiative opacity multipliers produce smaller but clearly resolved changes. Halving the time step changes temperature by only 0.0042 K, a numerical-consistency control confirming that the ranking reflects physics rather than discretization.

All thirteen configurations complete with finite states, demonstrating stable long-horizon integration; by that horizon the freely evolving trajectories reach chaotic saturation, so RMS separations converge and process sensitivity is read from the short horizon. These integrations also verify conservation: the full-physics year has a maximum relative CO_2 atmosphere–frost inventory drift of 4.28×10^{-9},

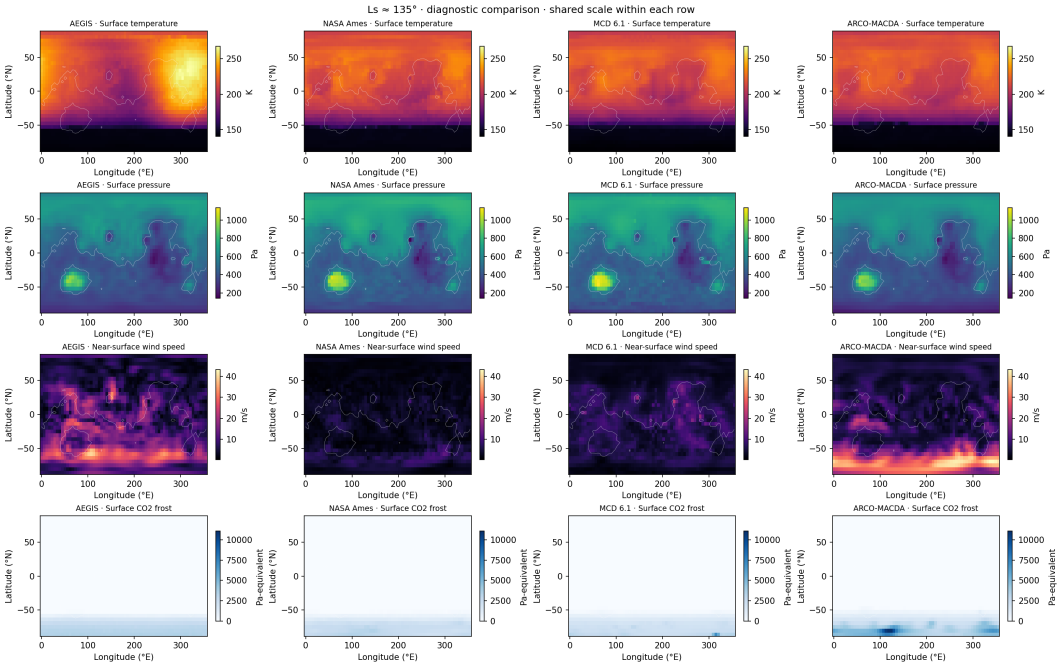


Figure 3: MOLA terrain on the comparison grid (top), and seasonal fields at $L_s = 135°$ (bottom). The seasonal panels compare the physical model with NASA Ames, MCD 6.1, and ARCO-MACDA for surface temperature, pressure, near-surface wind speed, and surface CO_2 frost.

Table 3: Seasonal reference comparisons. Entries are arithmetic means of the four seasonal RMSE values and pattern correlations; they are not pooled spatial RMSE. Pressure and temperature refer to the surface. The high pressure correlations largely reflect common terrain structure.

	Pressure		Temperature	
Reference	RMSE (Pa)	Correlation	RMSE (K)	Correlation
NASA Ames	52.37	0.9864	20.64	0.7889
MCD 6.1	57.18	0.9775	21.58	0.7872
ARCO-MACDA	40.72	0.9879	22.27	0.7715

confirming that the inventory projection holds total CO_2 mass to near machine precision over a Mars year while allowing seasonal exchange with the polar frost.

3.5 Neural components as coupled experiments

AEGIS exposes learned components through the same operator interface as its physical parameterizations. A neural module attaches at a chosen process either as a bounded additive residual or as a full component replacement. Because the module enters the coupled step, its effect propagates through the subsequent dynamics and physics, and it is trained by differentiating the resulting trajectory. We showcase two experiments using these interfaces exposed by the framework.

3.5.1 Learning radiative fluxes

We replace the compact physical radiation kernel with a structured learned flux operator that preserves the discrete radiative budget. Four-layer columns are first trained locally for 100 updates. Starting from this local fit, the operator is either fine-tuned through 12-step column trajectories for 20 updates or trained further on the local loss for the same wall-clock budget, which amounts to 234 updates. Table 4 compares the initial local fit, trajectory fine-tuning, and continued local training on interpolation and higher-dust extrapolation tests.

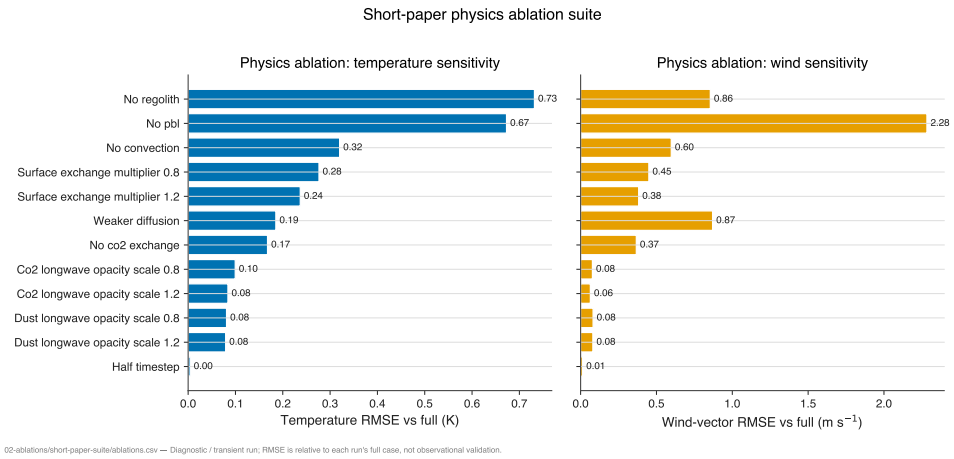


Figure 4: Sensitivity of the T21/L12 Mars configuration. Bars show RMS differences from the paired full-physics trajectory over five saved times through 44 400 s; the right panel combines zonal and meridional wind RMSE by quadrature. All cases begin from the same state.

Table 4: Atmospheric-temperature RMSE (K) after 48 coupled steps for learned radiation variants. All models use the same structured flux interface.

Training procedure	Interpolation	Higher dust
Initial local fit	0.620	1.799
Trajectory fine-tuning	0.566	1.653
Continued local fitting	**0.416**	**0.979**

Continued local fitting gives the lowest 48-step temperature error in this experiment, improving interpolation RMSE from 0.620 to 0.416 K and higher-dust extrapolation RMSE from 1.799 to 0.979 K. The learned flux representation retains the discrete radiative budget: the largest instantaneous closure residual across held-out local evaluations is $1.42 \times 10^{-14}\,\mathrm{W\,m^{-2}}$. All three variants also complete a two-step coupled T21 rollout. The shared framework permits comparisons among learned processes, physical constraints, and local versus coupled trajectory objectives.

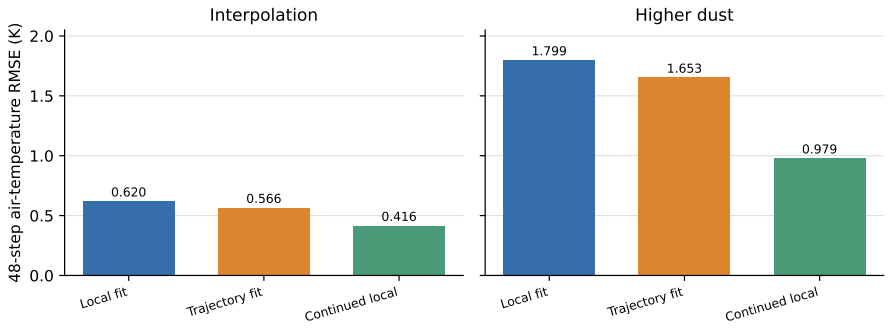


Figure 5: Archived neural-radiation experiment at a 48-step column rollout. Air-temperature RMSE is evaluated against the compact physical radiation kernel on held-out interpolation columns and higher-dust extrapolation columns (24 columns per split). The three bars compare the initial local fit, trajectory fine-tuning, and continued local fitting.

3.5.2 Recovering a physical parameter through a trajectory

The differentiable column system recovers a known 1.6 multiplier on a bulk sensible-heat conductance of $2\,\mathrm{W\,m^{-2}\,K^{-1}}$, using targets generated with fixed illumination, dust, and pressure per column.

After 120 updates through 24-step trajectories, initial multipliers of 0.5, 1.0, and 3.0 converge to 1.59950, 1.59901, and 1.59696, all within 0.20% of the target (Fig. 6). Derivative checks at multipliers 0.5, 1.0, and 2.5 agree with finite differences to relative discrepancies below 3.41×10^{-10}. A derivative-free scalar reference reaches 1.60000007 in nine objective evaluations. This controlled recovery establishes parameter identifiability and a gradient-based inverse workflow through the shared differentiable column operators.

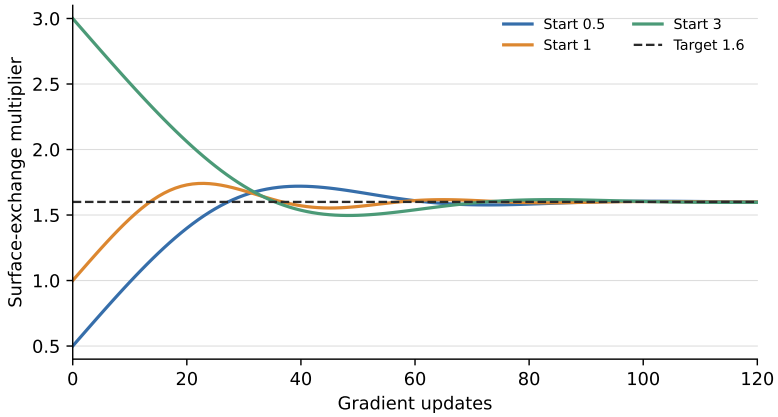


Figure 6: Physical inverse experiment from the neural-framework archive. Gradient-based optimization recovers a prescribed surface-exchange multiplier of 1.6 from three initial values over 120 updates. Each curve includes its initial value and the saved multiplier after every update. The parameter belongs to the controlled column model; the experiment tests the shared differentiable inference workflow.

4 Future Extensions

The present demonstrations span coarse-resolution physical simulations, coupled-heating seed, and structured neural component replacement, and they establish the differentiable physical and neural interfaces. The framework already supports the natural next steps: broader observational comparison, repeated neural experiments across seeds, and multi-year equilibration studies, each reusing the same coupled operators and trajectory objectives. The explicit planetary, radiation, and reservoir interfaces provide a route to experiments with changing Martian boundary conditions. Altered surface albedo, imposed heating, or specified changes in volatile inventory could be introduced through these interfaces and evaluated with spatial climate diagnostics. Differentiable objectives could then connect intervention parameters to quantities such as regional temperature, frost persistence, or atmospheric pressure. For larger changes in atmospheric composition and mass, the body thermodynamics, opacity data, condensable processes, and applicable dynamical assumptions would be extended together.

5 Conclusion

We have presented AEGIS, a differentiable Mars climate model that couples a modular Mars physics suite to the Dinosaur dynamical core and exposes interfaces for learned components. The model represents terrain, radiative forcing, surface and subsurface heat exchange, vertical transport, and the seasonal exchange of CO_2 between atmosphere and frost within a single differentiated trajectory. Its gradients agree with finite differences through coupled trajectories, and gradient-based calibration lowers held-out reanalysis temperature error in every test window. Integrations of up to ten Mars years remain stable, reproduce the seasonal CO_2 cycle and the terrain-driven surface-pressure structure, and conserve the total CO_2 inventory, with a Mars year completing in under 1.5 h on a single GPU. Neural heating and radiation interfaces make the same framework usable for training and comparing learned components. AEGIS thus provides a common foundation for physical sensitivity studies, gradient-based calibration, and the development of hybrid planetary climate models.

AI USE STATEMENT

We used generative AI tools in the following ways and have reviewed all AI-assisted work. We take responsibility for the final content of this work, including text, claims, and artifacts produced with the aid of generative AI.

- **To aid or polish writing.** AI tools helped draft, condense, and reword manuscript text, including section summaries, captions, and the abstract and conclusion. All AI-assisted text was checked by the authors against the reported results, and every number in the paper was traced to its archived result file.
- **For retrieval and discovery.** AI tools helped identify related work and check bibliography entries against their Crossref, DataCite, and arXiv records. Every citation was verified against its DOI or arXiv record, and incorrect entries were corrected or removed.
- **For research ideation or execution.** AI tools suggested candidate experiments and helped write and test analysis, plotting, and auxiliary experiment scripts. The core model was written by the authors. AI-written scripts were inspected, tested with smoke and gradient checks, and run by the authors. No simulation outputs or numerical results were produced or altered by AI tools; all reported values come from executed model runs archived with the code.

REFERENCES

Atilim Gunes Baydin, Barak A. Pearlmutter, Alexey Andreyevich Radul, and Jeffrey Mark Siskind. Automatic differentiation in machine learning: a survey. *Journal of Machine Learning Research*, 18(153):1–43, 2018. URL http://jmlr.org/papers/v18/17-468.html.

Ananyo Bhattacharya. ARCO-MACDA, 2026. URL https://huggingface.co/datasets/ananyo01/ARCO-MACDA. Dataset; accessed 23 September 2026; experiments pin archive revision 65a0bebd804b9c240752277e83f5737d58c6ee9c.

Ellen H. Davenport, J. Varan Madan, Rebecca Gjini, Jared Brzenski, Nick Ho, Tien-Yiao Hsu, Yueshan Liang, Zhixing Liu, Veeramakali Manivannan, Eric Pham, Rohith Vutukuru, Andrew I. L. Williams, Zhiqi Yang, Rose Yu, Nicholas J. Lutsko, Stephan Hoyer, and Duncan Watson-Parris. JCM v1.0: A Differentiable, Intermediate-Complexity Atmospheric Model. *EGUsphere*, 2026. doi: 10.5194/egusphere-2025-6266. URL https://doi.org/10.5194/egusphere-2025-6266. Preprint.

François Forget, François Hourdin, Robin Fournier, Christophe Hourdin, Olivier Talagrand, Mike Collins, Stephen R. Lewis, Peter L. Read, and Jean-Pierre Huot. Improved general circulation models of the martian atmosphere from the surface to above 80 km. *Journal of Geophysical Research: Planets*, 104(E10):24155–24175, 1999a. doi: 10.1029/1999JE001025.

Francois Forget, Frederic Hourdin, Richard Fournier, Christophe Hourdin, Olivier Talagrand, Matthew Collins, Stephen R. Lewis, Peter L. Read, and Jean-Paul Huot. Improved general circulation models of the Martian atmosphere from the surface to above 80 km. *Journal of Geophysical Research: Planets*, 104(E10):24155–24175, 1999b. doi: 10.1029/1999je001025. URL https://doi.org/10.1029/1999je001025.

Matthias Gelbrecht, Andrew White, Sebastian Bathiany, and Niklas Boers. Differentiable programming for earth system modeling. *Geoscientific Model Development*, 16(11):3123–3139, 2023. doi: 10.5194/gmd-16-3123-2023.

Ralf Giering and Thomas Kaminski. Recipes for adjoint code construction. *ACM Transactions on Mathematical Software*, 24(4):437–474, 1998. doi: 10.1145/293686.293695. URL https://doi.org/10.1145/293686.293695.

Robert M. Haberle, James B. Pollack, Jeffrey R. Barnes, Richard W. Zurek, Conway B. Leovy, James R. Murphy, Hilda Lee, and James Schaeffer. Mars atmospheric dynamics as simulated by the NASA Ames General Circulation Model: 1. The zonal-mean circulation. *Journal of Geophysical Research: Planets*, 98(E2):3093–3123, 1993. doi: 10.1029/92JE02946. URL https://doi.org/10.1029/92JE02946.

Robert M. Haberle, Melinda A. Kahre, Jeffery L. Hollingsworth, Franck Montmessin, R. John Wilson, Richard A. Urata, Amanda S. Brecht, Michael J. Wolff, Alexandre M. Kling, James R. Schaeffer, et al. Documentation of the NASA/Ames legacy Mars global climate model: Simulations of the present seasonal water cycle. *Icarus*, 333:130–164, 2019. doi: 10.1016/j.icarus.2019.03.026.

Zeyuan Hu, Akshay Subramaniam, Zhiming Kuang, Jerry Lin, Sungduk Yu, Walter M. Hannah, Noah D. Brenowitz, Josh Romero, and Michael S. Pritchard. Stable machine-learning parameterization of subgrid processes in a comprehensive atmospheric model learned from embedded convection-permitting simulations, 2025. URL https://arxiv.org/abs/2407.00124.

JAX Developers. Forward- and reverse-mode automatic differentiation in JAX. https://docs.jax.dev/en/latest/jacobian-vector-products.html, a. Accessed: 2026-09-24.

JAX Developers. Gradient checkpointing with jax.checkpoint (jax.remat). https://docs.jax.dev/en/latest/gradient-checkpointing.html, b. Accessed: 2026-09-24.

Melinda A. Kahre and Robert M. Haberle. Mars CO_2 cycle: Effects of airborne dust and polar cap ice emissivity. *Icarus*, 207(2):648–653, 2010. doi: 10.1016/j.icarus.2009.12.016.

Sameera S Kashyap, Raj Abhijit Dandekar, Rajat Dandekar, and Sreedath Panat. Modeling chaotic lorenz ode system using scientific machine learning, 2024. URL https://arxiv.org/abs/2410.06452.

N. J. Kelly, W. V. Boynton, K. Kerry, D. Hamara, D. Janes, R. C. Reedy, K. J. Kim, and R. M. Haberle. Seasonal polar carbon dioxide frost on mars: Co2 mass and columnar thickness distribution. *Journal of Geophysical Research: Planets*, 111(E3), 2006. doi: https://doi.org/10.1029/2006JE002678. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2006JE002678.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. arXiv:1412.6980, 2014. URL https://arxiv.org/abs/1412.6980.

Dmitrii Kochkov, Janni Yuval, Ian Langmore, Peter Norgaard, Jamie Smith, Griffin Mooers, Milan Klower, James Lottes, Stephan Rasp, Peter Duben, Sam Hatfield, Peter Battaglia, Alvaro Sanchez-Gonzalez, Matthew Willson, Michael P. Brenner, and Stephan Hoyer. Neural general circulation models for weather and climate. *Nature*, 632(8027):1060–1066, 2024. doi: 10.1038/s41586-024-07744-y. URL https://doi.org/10.1038/s41586-024-07744-y.

Andrew A. Lacis and Valdar Oinas. A description of the correlated k distribution method for modeling nongray gaseous absorption, thermal emission, and multiple scattering in vertically inhomogeneous atmospheres. *Journal of Geophysical Research: Atmospheres*, 96(D5):9027–9063, 1991. doi: 10.1029/90jd01945. URL https://doi.org/10.1029/90jd01945.

Doug McNeall, Jonny Williams, Richard Betts, Ben Booth, Peter Challenor, Peter Good, and Andy Wiltshire. Correcting a bias in a climate model with an augmented emulator. *Geoscientific Model Development*, 13(5):2487–2509, 2020. doi: 10.5194/gmd-13-2487-2020.

NASA Goddard Space Flight Center. Mars fact sheet. https://nssdc.gsfc.nasa.gov/planetary/factsheet/marsfact.html, 2025. NASA National Space Science Data Center. Accessed 2026-09-23.

NASA Planetary Data System. Mars Global Surveyor Mars Orbiter Laser Altimeter: Mission Experiment Gridded Data Records. MOLA MEGDR archive, n.d. URL https://pds-geosciences.wustl.edu/mgs/mgs-m-mola-5-megdr-l3-v1/. Terrain asset used here: meg004/megt90n000cb.img; checksum recorded with the experiment.

Lev Nayvelt, Peter J. Gierasch, and Kerry H. Cook. Modeling and Observations of Martian Stationary Waves. *Journal of the Atmospheric Sciences*, 54(8):986–1013, 1997. doi: 10.1175/1520-0469(1997)054⟨0986:MAOOMS⟩2.0.CO;2.

Arakel Petrosyan, Boris Galperin, Søren E. Larsen, Stephen R. Lewis, A. Määttänen, et al. The martian atmospheric boundary layer. *Reviews of Geophysics*, 49(3), 2011. doi: 10.1029/2010RG000351.

Stephan Rasp. Coupled online learning as a way to tackle instabilities and biases in neural network parameterizations: general algorithms and Lorenz 96 case study (v1.0). *Geoscientific Model Development*, 13(5):2185–2196, 2020. doi: 10.5194/gmd-13-2185-2020. URL https://doi.org/10.5194/gmd-13-2185-2020.

Mark I. Richardson and Claire E. Newman. On the relationship between surface pressure, terrain elevation, and air temperature. part i: The large diurnal surface pressure range at gale crater, mars and its origin due to lateral hydrostatic adjustment. *Planetary and Space Science*, 2018. URL https://api.semanticscholar.org/CorpusID:125650441.

Mark I. Richardson, Anthony D. Toigo, and Claire E. Newman. PlanetWRF: A general purpose, local to global numerical model for planetary atmospheric and climate dynamics. *Journal of Geophysical Research: Planets*, 112(E9), 2007. doi: 10.1029/2006JE002825. URL https://doi.org/10.1029/2006JE002825.

Samuel S. Schoenholz and Ekin D. Cubuk. JAX, M.D.: A framework for differentiable physics. In *Advances in Neural Information Processing Systems*, volume 33, 2020. URL https://arxiv.org/abs/1912.04232.

David E. Smith, Maria T. Zuber, Herbert V. Frey, James B. Garvin, James W. Head, Duane O. Muhleman, Gordon H. Pettengill, Roger J. Phillips, Sean C. Solomon, H. Jay Zwally, W. Bruce Banerdt, Thomas C. Duxbury, Matthew P. Golombek, Frank G. Lemoine, Gregory A. Neumann, David D. Rowlands, Oded Aharonson, Peter G. Ford, Anton B. Ivanov, Catherine L. Johnson, Patrick J. McGovern, James B. Abshire, Robert S. Afzal, and Xiaoli Sun. Mars Orbiter Laser Altimeter: Experiment summary after the first year of global mapping of Mars. *Journal of Geophysical Research: Planets*, 106(E10):23689–23722, 2001. doi: 10.1029/2000je001364. URL https://doi.org/10.1029/2000je001364.

Alex Valeanu, Kylash Rajendran, Luca Montabone, and Peter L. Read. Mars Analysis Correction Data Assimilation (MACDA): Atmospheric and surface fields produced from assimilated MGS/TES, ODY/THEMIS, and MRO/MCS observations, v2.0, 2026. URL https://catalogue.ceda.ac.uk/uuid/cd037a9ea387438fabf4d674dbe53088. Dataset; accessed 23 September 2026.

Jeffrey S. Whitaker and Sajal K. Kar. Implicit–Explicit Runge–Kutta Methods for Fast–Slow Wave Problems. *Monthly Weather Review*, 141(10):3426–3434, 2013. doi: 10.1175/mwr-d-13-00132.1. URL https://doi.org/10.1175/mwr-d-13-00132.1.

A MARS PHYSICS OPERATORS

MOLA elevation data provide the lower boundary (Smith et al., 2001; NASA Planetary Data System, n.d.). The raster is regridded onto the transform grid and converted into spectral orography. Prescribed dust fields can be supplied through seasonal or per-window adapters. These data remain separate from the numerical operators, making boundary-condition experiments straightforward.

Radiation combines CO_2 absorption and dust optical properties in a column flux calculation. The correlated-k approach represents nongray gaseous absorption using a reduced distribution of absorption coefficients (Lacis & Oinas, 1991). The main physical configuration uses bundled Ames-derived 12-band correlated-k tables, interpolated in temperature and log pressure, together with two-stream transfer. Solar and thermal fluxes are evaluated at layer interfaces, and atmospheric heating follows from their convergence. For layer k, with pressure thickness $\Delta p_k = p_s \Delta \sigma_k$, the atmospheric heat capacity per unit area is

$$C_k = \frac{c_p \Delta p_k}{g}, \qquad \left.\frac{dT_k}{dt}\right|_{\text{rad}} = \frac{Q_k}{C_k}, \tag{4}$$

where Q_k is radiative flux convergence. Computing atmospheric and surface heating from a common set of interface fluxes makes the column energy exchanges telescope to the top-of-atmosphere balance. This same construction is used for the neural radiation interface.

The surface temperature evolves from radiative, sensible, conductive, and latent heat exchange (Forget et al., 1999a; Haberle et al., 2019). A bulk aerodynamic sensible heat flux couples the surface

to the lowest atmospheric layer. The surface loss and atmospheric gain use the same flux, with atmospheric heating divided by the layer heat capacity. Surface stress transfers momentum, and Richardson-number-dependent vertical diffusion mixes momentum, heat, and tracers. The vertical diffusion solve uses a batched tridiagonal algorithm (Petrosyan et al., 2011). Dissipated resolved kinetic energy is returned as heat. Dry convective adjustment mixes unstable adjacent layers while conserving their sigma-weighted enthalpy. A 12-layer regolith conducts heat beneath the surface and supplies thermal memory across daily and seasonal forcing.

The CO_2 cycle couples temperature to atmospheric mass. Surface condensation releases latent heat and transfers mass from the atmosphere to frost; sublimation reverses the exchange. The local phase-change contributions satisfy

$$\left.\frac{dp_s}{dt}\right|_{\text{phase}} = -\frac{dI}{dt}, \qquad \left.\frac{d\log p_s}{dt}\right|_{\text{phase}} = -\frac{1}{p_s}\frac{dI}{dt}. \tag{5}$$

The frost temperature follows a pressure-dependent saturation relation, and an energy-limited option ties the exchange to the surface energy budget. After each complete time step, a projection repairs negative frost and transfers the corresponding deficit from the atmospheric column. A global pressure rescaling then matches the quadrature-integrated atmosphere-plus-frost inventory to its pre-step value, accounting for prescribed escape when enabled. This addresses the finite-step inventory error associated with advancing log pressure. Atmospheric pressure can still change seasonally as mass moves into and out of the surface reservoir.

B Calibration and neural-training protocol

Physical calibration and coupled heating are evaluated using twelve prespecified ARCO-MACDA reanalysis windows (Bhattacharya, 2026; Valeanu et al., 2026). Four windows from Mars years 24, 26, 29, and 31 are used for training; four from years 32–33 for validation; and four from years 34–35 for testing. Each split samples four seasonal quarters. Atmospheric initial conditions are constructed from the reference temperature, winds, and pressure. Surface and frost initialization use start-time fields, and the soil initialization uses preceding surface history. The native reference interval of 7397.937 s is integrated with 25 steps of 295.91748 s. Start-time dust and the orbital phase are matched for each window.

The training loss averages squared errors in temperature, zonal wind, and meridional wind, after scaling each field by its training-target standard deviation. It uses horizontal area weights, equal sigma-layer weights, and equal window weights. Feature normalization uses initial states from training data. Both fitting experiments perform 50 Adam updates (Kingma & Ba, 2014) with learning rate 10^{-3} and gradient-norm clipping at one. Validation selects the checkpoint, including the initial physical baseline among the candidates. Physical calibration adjusts CO_2 longwave opacity, dust longwave opacity, and surface exchange within bounds $[0.05, 2]$, $[0.05, 2]$, and $[0.25, 4]$. Neural fitting retains all three physical multipliers at one. Reported split RMSE is the square root of the mean per-window MSE, matching the figure-generation code.

C Additional seasonal reference comparisons

Figure 3 in the main text shows the seasonal fields at $L_s = 135°$. Figures 7 and 8 give the remaining perihelion-side and aphelion-side seasons, $L_s = 45°$ and $L_s = 315°$, under the same sampling conventions and the same set of reference products. The terrain-driven surface-pressure structure and the seasonal temperature and CO_2-frost patterns summarized in Table 3 are consistent across all evaluated seasons.

Each row of Figure 3 isolates one diagnostic, and the shared per-row color scale makes model–reference differences visible directly. Surface pressure (second row) is dominated by topography: low-lying basins hold more atmospheric column mass and appear as pressure maxima, while elevated terrain appears as minima, and the model reproduces this terrain-driven structure, which is shared by all three references. Surface temperature (first row) shows the daytime-weighted seasonal heating, with a cold winter pole. Near-surface wind speed (third row) is the most reference-dependent field, because the products differ in reported level and averaging; the model's organized near-surface jets

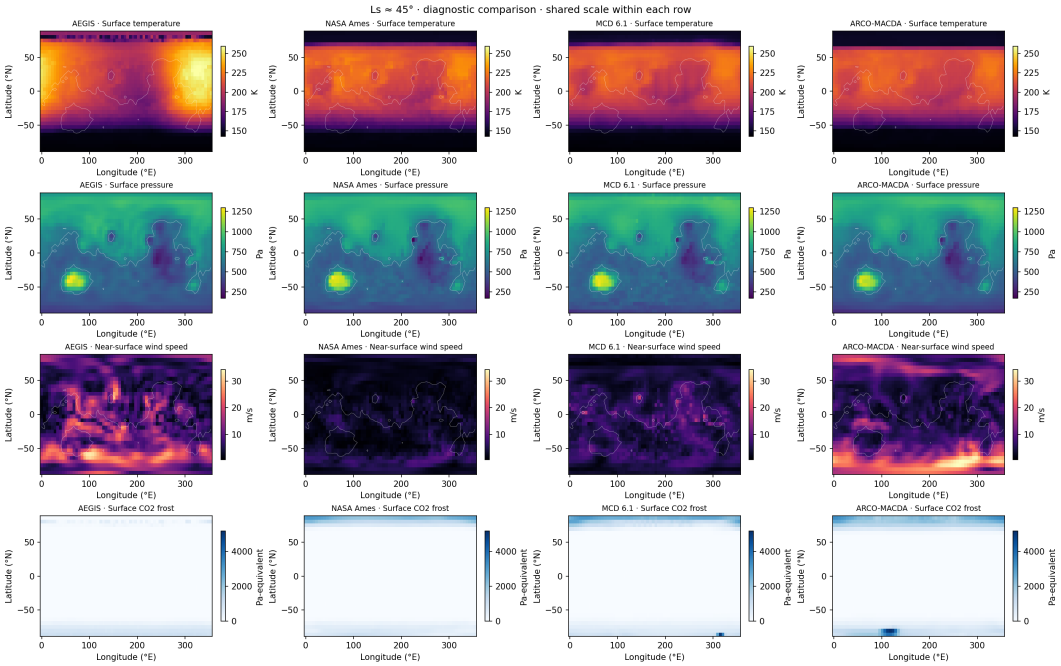


Figure 7: Seasonal fields at $L_s = 45°$. The panels compare the physical model with NASA Ames, MCD 6.1, and ARCO-MACDA for surface temperature, pressure, near-surface wind speed, and surface CO_2 frost, under the sampling conventions described in the main text.

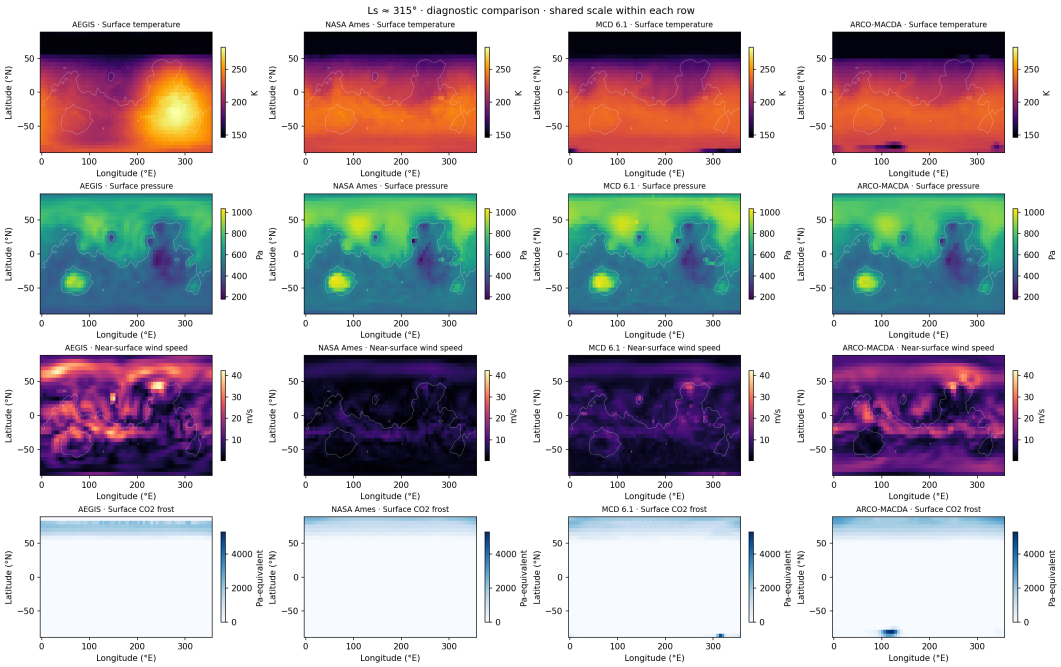


Figure 8: Seasonal fields at $L_s = 315°$. The panels compare the physical model with NASA Ames, MCD 6.1, and ARCO-MACDA for surface temperature, pressure, near-surface wind speed, and surface CO_2 frost, under the sampling conventions described in the main text.

are therefore compared against references with distinct sampling. Surface CO_2 frost (fourth row) concentrates near the seasonal winter cap and is a sensitive diagnostic of the atmosphere–surface

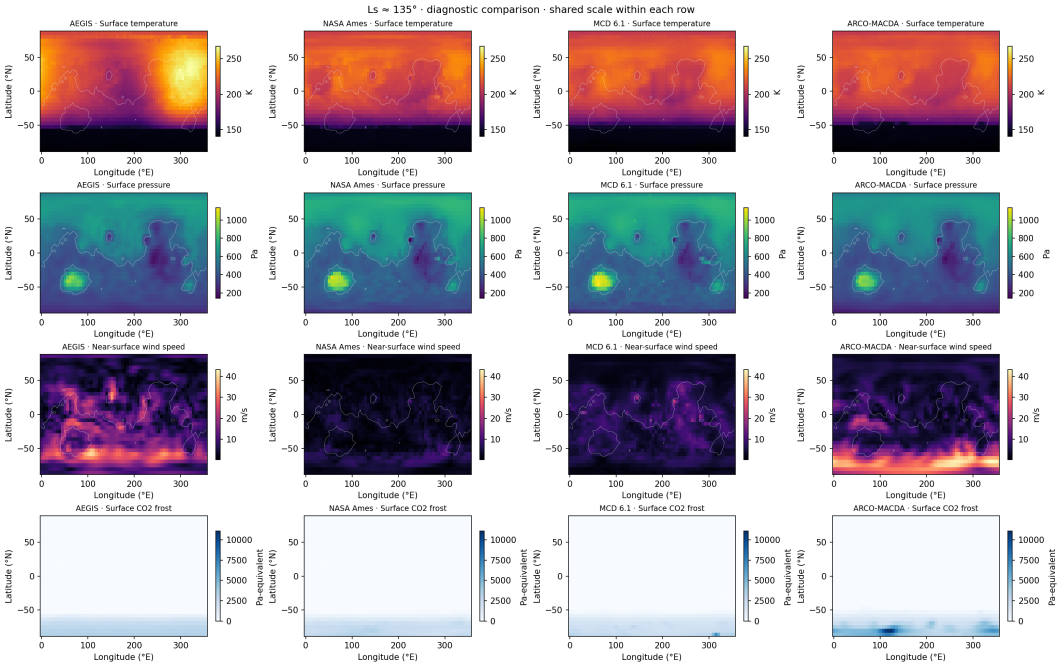


Figure 9: Seasonal diagnostic comparison at $L_s = 135°$, organized as a 4×4 grid. Each *row* is one surface field—surface temperature (K), surface pressure (Pa), near-surface wind speed ($\mathrm{m\,s^{-1}}$), and surface CO_2 frost (Pa-equivalent)—and each *column* is one source: the AEGIS physical model (leftmost) alongside the NASA Ames, MCD 6.1, and ARCO-MACDA reference products. All panels are longitude–latitude maps (0–360°E, 90°S–90°N) with the MOLA terrain contour overlaid for geographic reference, and the color scale is shared within each row so the model and the three references are directly comparable field by field. Reading across a row shows how closely the modeled field tracks each reference; reading down the leftmost column shows the modeled state across all four diagnostics. Table 3 summarizes the corresponding pattern correlations and RMSE.

mass exchange. These qualitative comparisons complement the quantitative pattern correlations and RMSE in Table 3.

Figure 3 in the main text shows the $L_s = 135°$ season. Figures 7 and 8 present the remaining evaluated seasons, $L_s = 45°$ and $L_s = 315°$, using the identical 4×4 layout: rows are the four surface fields, columns are the AEGIS model and the three reference products, panels are longitude–latitude maps with the MOLA terrain contour, and the color scale is shared within each row. The same reading applies—across a row for model–reference agreement, down the model column for the simulated state. Across all three evaluated seasons, the terrain-driven surface-pressure structure, the seasonal temperature pattern, and the winter-cap CO_2 frost distribution summarized in Table 3 remain consistent, while near-surface wind speed continues to show the largest spread between reference products.