

Periodically activated physics-informed neural networks for assimilation tasks for three-dimensional Rayleigh–Bénard convection

Michael Mommert ^{a,*}, Robin Barta ^{a,b}, Christian Bauer ^a, Marie-Christine Volk ^a, Claus Wagner ^{a,b}

^a Department Ground Vehicles, Institute of Aerodynamics and Flow Technology, German Aerospace Center, Bunsenstr. 10, 37073, Göttingen, Germany

^b Institute of Thermodynamics and Fluid Mechanics, Technische Universität Ilmenau, 98684, Ilmenau, Germany

ARTICLE INFO

Keywords:

Rayleigh–Bénard convection
Physics-informed neural networks
Assimilation
Machine learning
Activation functions

ABSTRACT

We apply physics-informed neural networks to three-dimensional Rayleigh–Bénard convection in a cubic cell with a Rayleigh number of $Ra = 10^6$ and a Prandtl number of $Pr = 0.7$ to assimilate the velocity vector field from given temperature fields and vice versa. With the respective ground truth data provided by a direct numerical simulation, we are able to evaluate the performance of the different activation functions applied (sine, hyperbolic tangent and exponential linear unit) and different numbers of neurons (32, 64, 128, 256) for each of the five hidden layers of the multi-layer perceptron. The main result is that the use of a periodic activation function (sine) typically benefits the assimilation performance in terms of the analyzed metrics, correlation with the ground truth and mean average error. The higher quality of results from sine-activated physics-informed neural networks is also manifested in the probability density function and power spectra of the inferred velocity or temperature fields. Regarding the two assimilation directions, the assimilation of temperature fields based on velocities appears to be more challenging in the sense that it exhibits a sharper limit on the number of neurons below which viable assimilation results cannot be achieved.

1. Introduction

Turbulent thermal convection governs a wide variety of flows considered for solving engineering tasks (see e.g. Schmeling and Volkman [1]) and environmental issues (see e.g. Völker et al. [2]). Understanding and controlling these flows requires knowledge of the heat fluxes they induce. The acquisition of heat fluxes requires both velocity and temperature data. In terms of measurements, the simultaneous recording of both fields has been shown to be feasible in a laboratory environment (for recent examples see Mommert et al. [3], Käufer and Cierpka [4]). However, even on this scale, these measurements require a high level of experimental effort.

In this context, the approach of assimilating temperature fields from volumetric velocity measurements is an attractive option to avoid the significant increase in complexity that an additional temperature measurement represents. More specifically, the assimilation based on physics-informed neural networks (PINNs) is controlled by the underlying partial differential equations (PDEs) as part of the loss function for a network, and thus does not require additional data to train the models a priori. This approach also promises to be extensible to flows of greater complexity and scale beyond the laboratory by exploiting the potential of modern, highly parallelized computing architectures (Vinueza et al. [5]). Therefore, we pursue the development of a PINN-based

assimilation method that relies only on field data typically obtained from measurements (either just temperatures or just velocities).

The following are examples of applications of PINNs that are relevant to this application: One of the possible applications of PINNs is similar to that of a numerical solver, where the only data provided are the initial and boundary conditions. Such implementations are usually confined to two dimensions, such as the vortex shedding behind a circular object underlying the Navier–Stokes equations, as presented in the original PINN paper by Raissi et al. [6]. Other examples are neural solvers for the Poisson equation (Markidis [7]) or the compressible Euler equations (Wassing et al. [8]). In summary, these examples show that neural networks are able to approximate the solutions of the PDEs governing the respective flows by minimizing the losses representing the residuals of the PDEs.

Yet, a major advantage of PINNs is their ability to incorporate any kind of additional data that helps to find the desired solution more effectively, as is the case for the assimilation tasks at hand (Cai et al. [9]). An example of this is the use of a PINN to regularize the results of two-dimensional three-component particle image velocimetry data, shown by Hasanuzzaman et al. [10]. Furthermore, Di Leoni et al. [11] used velocity data from three-dimensional Lagrangian measurements to assimilate the pressure. The reconstruction of velocity and pressure

* Corresponding author.

E-mail address: michael.mommert@dlr.de (M. Mommert).

<https://doi.org/10.1016/j.compfluid.2024.106419>

Received 6 March 2024; Received in revised form 6 August 2024; Accepted 27 August 2024

Available online 30 August 2024

0045-7930/© 2024 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

fields was also achieved by Eusebi et al. [12], showing the reconstruction of the flow of a tropical cyclone from spatially sparse data. Another example is provided by Soto et al. [13], where the PINN regularizes the velocity fields of a preceding reconstruction of small time scales and provides additional pressure data. This set of examples demonstrates the basic ability of the PINN method to handle noisy or sparse data, which is required for assimilation tasks.

The case of Rayleigh–Bénard convection considered here is characterized by the transport of the scalar temperature. How data of a transported scalar can be leveraged to assimilate the velocity field was shown by Raissi et al. [14] for the three-dimensional flow in an intracranial aneurysm. Even more relevant for Rayleigh–Bénard convection, the temperature field was used by Cai et al. [15] to reconstruct three-dimensional velocity and pressure fields of the region above an espresso cup in which a tomographic background-oriented Schlieren measurement was applied. The same type of assimilation to provide temperature data for the training was done by Lucor et al. [16]. However, they still provided both velocities and temperatures as initial conditions, which would not be feasible in the context of avoiding the respective measurement effort. Another example of the assimilation of velocities is the work of Di Leoni et al. [17], who investigated the effect of sparsity on the provided temperature data in two-dimensional Rayleigh–Bénard convection. Cai et al. [18] also used local temperature measurements in combination with boundary conditions to recover the flow and thermal boundary conditions for either forced or mixed convective flows past cylinders.

The above-mentioned publications show that PINNs are usually used to assimilate flows based on temperature data, while modern measurement techniques are capable of providing highly resolved velocity data of thermally-driven convective flows. Therefore, we use the assimilation of velocity fields as a link to previous studies before investigating the temperature assimilation pursued, which further allows a comparison of both assimilation directions.

Also, a large part of the studies mentioned above deal with a limited amount of turbulence-induced complexity, as the variety of structure sizes also poses high requirements on PINNs to be able to reconstruct them. This issue is, for example, addressed by implementing periodic activation functions, typically sine (Sitzmann et al. [19]). Because of their relation to Fourier features (Tancik et al. [20]) and thus to the idea of decomposing the flow fields into Fourier modes, they seem predestined to allow neural networks to efficiently map complex flows. In the context of turbulent flows, periodically activated PINNs have been used by Angriman et al. [21] to reconstruct turbulence from underresolved flow fields by imposing information about higher order statistical moments.

To evaluate the performance of the PINNs, we base our investigation on ground truth data generated in a direct numerical simulation described in Section 2. Regarding turbulent Rayleigh–Bénard convection, the assimilation of velocity or temperature data becomes increasingly complex with increasing turbulence intensity, indicated by the dimensionless Rayleigh (Ra) and Prandtl (Pr) numbers. Therefore, we chose to study a case of Rayleigh–Bénard convection in a cubic cell at $Ra = 10^6$ and $Pr = 0.7$ with moderate turbulence. Since we investigate the influence of different activation functions on the ability of a neural network to map the complexity of turbulent flows, we also vary the number of neurons in each hidden layer of the PINN. Further details of the PINN and its training parameters are described in Section 3. The results of the PINNs with the different parameters and for both assimilation directions are then presented and discussed in Section 4.

2. Generation of ground truth data

The ground truth data are generated in a direct numerical simulation (DNS) of a turbulent Rayleigh–Bénard convection at $Pr = 0.7$ and $Ra = 10^6$ in a cubic domain, solving the transport equations for

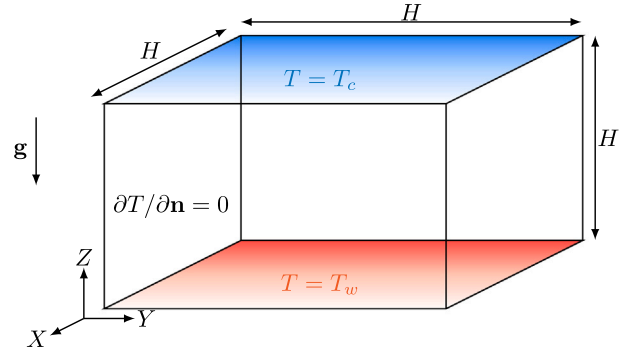


Fig. 1. Cubic Rayleigh–Bénard convection cell with the height H and the volume H^3 . All walls are no-slip boundaries; top and bottom walls are iso-thermal and side walls are adiabatic.

mass, momentum, and temperature for an incompressible fluid and the Boussinesq approximation,

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} = -\nabla p + \sqrt{\text{Pr}/\text{Ra}} \nabla^2 \mathbf{u} + T \mathbf{e}_z, \quad (1)$$

$$\frac{\partial T}{\partial t} + \mathbf{u} \cdot \nabla T = \sqrt{1/(\text{PrRa})} \nabla^2 T, \quad (2)$$

$$\nabla \cdot \mathbf{u} = 0, \quad (3)$$

where $\mathbf{u} = (u_x, u_y, u_z)$ is the velocity vector, p is the pressure, T is the temperature, and \mathbf{e}_z is the unit vector with respect to the vertical direction. The flow geometry, consisting of a cubic domain with a heated bottom plate ($T = T_w$), a cooled top plate ($T = T_c$), and adiabatic side walls, is displayed in Fig. 1.

Velocities are non-dimensionalized with the free-fall velocity $\hat{u}_{\text{ref}} = (\hat{\alpha} \hat{g} \Delta \hat{T} \hat{H}^3)^{1/2}$, spatial coordinates with the cell height $\hat{x}_{\text{ref}} = \hat{H}$, the time coordinate with the corresponding reference time $\hat{t}_{\text{ref}} = \hat{x}_{\text{ref}}/\hat{u}_{\text{ref}}$, and the pressure with the reference pressure $\hat{p}_{\text{ref}} = \hat{\rho} \hat{u}_{\text{ref}}^2$ with $\hat{\rho}$ being the fluid density, $\hat{\alpha}$ the thermal diffusivity, and \hat{g} the gravitational acceleration. The temperature is non-dimensionalized by $T = (\hat{T} - \hat{T}_0)/\Delta \hat{T}$ with $\Delta \hat{T} = \hat{T}_w - \hat{T}_c$ and $\hat{T}_0 = (\hat{T}_w + \hat{T}_c)/2$. No-slip and impermeability boundary conditions are applied to all walls. In addition, the top and bottom plates are assumed to be iso-thermal, whereas the side walls are modeled as adiabatic. After non-dimensionalization, the Eqs. (1)–(3) are discretized spatially and temporally with a fourth-order accurate finite volume scheme and a second-order accurate Euler–leapfrog time integration scheme, respectively. According to Wagner et al. [22], the temporal discretization of Eq. (1) for a leapfrog time step yields

$$\frac{1}{2\Delta t} (\mathbf{u}^{n+1} - \mathbf{u}^{n-1}) + \mathbf{u}^n \cdot \nabla \mathbf{u}^n = -\nabla p^n + \sqrt{\text{Pr}/\text{Ra}} \nabla^2 \mathbf{u}^{n-1} + T^n \mathbf{e}_z, \quad (4)$$

where n is the number of the time step, Δt is the temporal increment between two time steps, and δ_i^j is the Kronecker delta. To integrate Eq. (4) in time, a fractional step algorithm as the one introduced by Chorin [23,24] is applied. In a first step, an auxiliary velocity vector field \mathbf{u}^* is estimated from Eq. (4) neglecting the pressure term,

$$\frac{1}{2\Delta t} (\mathbf{u}^* - \mathbf{u}^{n-1}) + \mathbf{u}^n \cdot \nabla \mathbf{u}^n = \sqrt{\text{Pr}/\text{Ra}} \nabla^2 \mathbf{u}^{n-1} + T^n \mathbf{e}_z. \quad (5)$$

In the second step, the pressure Poisson equation is solved for the auxiliary field as follows:

$$\nabla^2 \phi^n = \nabla \cdot \mathbf{u}^*, \quad (6)$$

with $\phi^n = 2\Delta t p^n$ and the boundary condition $(\mathbf{n} \cdot \nabla \phi^n)|_{\partial \gamma} = 0$. Due to the inhomogeneity of the flow problem with respect to all three directions, the Poisson solver utilizes a separation of variables method as described by Shishkina et al. [25]. Finally, the velocity at time step $n+1$ is corrected via

$$\mathbf{u}^{n+1} = \mathbf{u}^* - \nabla \phi^n. \quad (7)$$

Table 1

Rayleigh–Bénard convection simulation case. Ra is the Rayleigh, Pr the Prandtl number. N_x , N_y and N_z are the number of grid points in x , y , and z direction, respectively. Δz is the grid spacing. N_{δ_θ} is the number of grid points in the thermal, N_{δ_u} in the kinetic boundary layer. The Nusselt number (see Eq. (14)) is computed a posteriori.

Ra	Pr	$N_x \times N_y \times N_z$	$\Delta z/H$	N_{δ_θ}	N_{δ_u}	Nu
10^6	0.7	$64 \times 64 \times 64$	0.0156	3	3	8.3

Regarding the spatial discretization, the numerical grid aims to fully resolve the smallest velocity and temperature scales, which are the Kolmogorov and Batchelor length scales

$$\eta_K = \left(\frac{\nu^3}{\varepsilon_u} \right) \text{ and} \quad (8)$$

$$\eta_B = \left(\frac{\nu \kappa^2}{\varepsilon_u} \right) = \eta_K \text{Pr}^{-1/2}, \quad (9)$$

respectively, with $\varepsilon_u = \nu \langle \partial u_i / \partial x_j \rangle$ the kinetic dissipation rate. Angle brackets indicate statistical averaging. For the present case of $\text{Pr} < 1$, the Kolmogorov length scale is smaller than the Batchelor length scale, and thus, more restrictive with respect to the grid resolution. According to Shishkina et al. [26], the minimum grid spacing in the bulk flow region can be estimated via the following approximation of the global Kolmogorov scale

$$h^{bulk}/H \leq \frac{\text{Pr}^{1/2}}{(\text{Nu} - 1)\text{Ra}^{1/4}}, \text{ for } \text{Pr} < 1, \quad (10)$$

whereas the minimum grid spacing in the thermal and viscous boundary layers is estimated based on the Prandtl–Blasius boundary layer theory,

$$h^{BL}/H \leq 2^{-3/2} a^{-1} \text{Nu}^{-3/2} \text{Pr}^{0.5355 - 0.033 \log \text{Pr}}, \text{ for } 3 \times 10^{-4} \leq \text{Pr} \leq 1, \quad (11)$$

with $a \approx 0.922$ [27]. Moreover, Shishkina et al. [26] recommend minimum numbers of nodes in the thermal and kinetic boundary layer of

$$N_{\delta_\theta}^{min} = \sqrt{2a} \text{Nu}^{1/2} \text{Pr}^{-0.5355 + 0.033 \log \text{Pr}}, \text{ for } 3 \times 10^{-4} \leq \text{Pr} \leq 1, \quad (12)$$

$$N_{\delta_u}^{min} = \sqrt{2a} \text{Nu}^{1/2} \text{Pr}^{-0.1785 + 0.011 \log \text{Pr}}, \text{ for } 3 \times 10^{-4} \leq \text{Pr} \leq 1. \quad (13)$$

With the Nusselt number derived a priori from the Grossmann–Lohse theory [28–31], a minimum grid spacing of 0.04 in the bulk flow region and 0.02 in the boundary layer is obtained. Additionally, the minimum number of nodes in the boundary layers — Eqs. (12) and (13) — is estimated to $N_{\delta_\theta}^{min} = 3$ and $N_{\delta_u}^{min} = 3$ for the thermal and kinetic boundary layer, respectively. Overall, an equidistant grid spacing of $\Delta z/H = 0.0156$ in all three dimensions is applied (see Table 1).

After an initial transient, when the average Nusselt number computed from the cooled and the heated plate,

$$\text{Nu} = -\frac{1}{2} \left(\left\langle \frac{dT}{dz} \right\rangle_{xyt} (z=0) + \left\langle \frac{dT}{dz} \right\rangle_{xyt} (z=H) \right), \quad (14)$$

has reached a statistically stationary state, instantaneous flow field realizations in the form of velocity, pressure, and temperature fields are written out every 0.05 dimensionless time units. The velocity fields are then interpolated from the staggered grid used in the DNS solver to a collocated grid with the pressure and temperature fields. Hereafter, these fields (denoted with $\bar{\cdot}$) serve as the ground truth and input data for the PINN.

3. PINN and training configuration

The following subsections are dedicated to the PINN approach and its application to assimilation. It covers the construction of the neural network 3.1, the data input 3.2 and additional details needed to configure the training process 3.3.

3.1. Architecture

To assimilate flow fields, such as velocity, temperature, or pressure fields, which are missing since they are not captured by a measurement, a multi-layer perceptron (MLP) is used. It is shown in Fig. 2 and serves as a universal function approximator for the u, v, w, T, p properties of the flow. The MLP consists of five fully connected hidden layers, each with a constant width of N_N neurons. This block of hidden layers is connected with linear scaling layers to the actual input and output to improve the condition of the network.

Each fully connected layer $\mathbf{g}^{(L+1)}$ is thereby configured as a linear combination, multiplying the weight matrix \mathbf{W} with the output \mathbf{g} of the previous layer L and adding the bias vector \mathbf{b} , which is then activated by a non-linear function σ :

$$\mathbf{g}^{(L+1)} = \sigma(\mathbf{W}^{(L+1)} \cdot \mathbf{g}^{(L)} + \mathbf{b}^{(L+1)}). \quad (15)$$

Overall this network can be viewed as a function

$$[u_x, u_y, u_z, T, p] = \mathbf{f}([t, x, y, z], \theta) \quad (16)$$

which is approximating solutions to the partial differential equations describing the investigated flow. Thereby, θ denotes the complete set of the trainable parameters of the network, i.e. the weights and biases of the layers.

Regarding the activation function, the hidden layer neurons are uniformly activated with either a sine as a periodic function, a hyperbolic tangent, or an exponential linear unit (ELU) [32] for comparison. However, the hyperbolic tangent function is used as the activation for the scaled output layer in any case due to its asymptotic behavior for large absolute inputs.

Scaling layers. Both scaling layers aim to condition the neural network in a way that its layer outputs are all scaled between -1 and 1 . Therefore, the already dimensionless input coordinates are scaled by

$$\zeta_{\text{scaled}} = \frac{2}{\zeta_{\text{max}} - \zeta_{\text{min}}} \zeta + \frac{\zeta_{\text{min}} + \zeta_{\text{max}}}{\zeta_{\text{min}} - \zeta_{\text{max}}}, \quad (17)$$

with $\zeta \in \{t, x, y, z\}$ and $\zeta_{\text{min}}, \zeta_{\text{max}}$ being the boundaries of the examined domain. This is especially necessary for the time coordinate, since we are examining relatively short intervals in the context of large DNS time coordinates.¹

For the output, only the temperatures are scaled with

$$T = T_{\text{scaled}}/2, \quad (18)$$

as -0.5 and 0.5 are the dimensionless temperatures of the bottom and top surface, which limit the range for the fluid. Since the velocity and pressure values do not have such clear limits, but are well distributed within $[-1, 1]$, no further scaling is applied to them.

3.2. Sampling

The sampling for the training process is conducted according to the DNS grid. To limit the computational effort, only 11 snapshots covering a time of $0.5 t_{\text{ff}}$ are processed. Together with the full DNS grid resolution of 64^3 used for the spatial sampling, this yields a number of approximately 2.9×10^6 data points. Thus, for each epoch, randomly composed mini-batches of size $N_b = 4 \times 10^3$ are provided for the individual training steps to benefit from the advantages of large batch sizes [33].

¹ The scaling can also be performed as a data preprocessing step. Here, we opted to implement it as part of the PINN architecture for compatibility reasons.

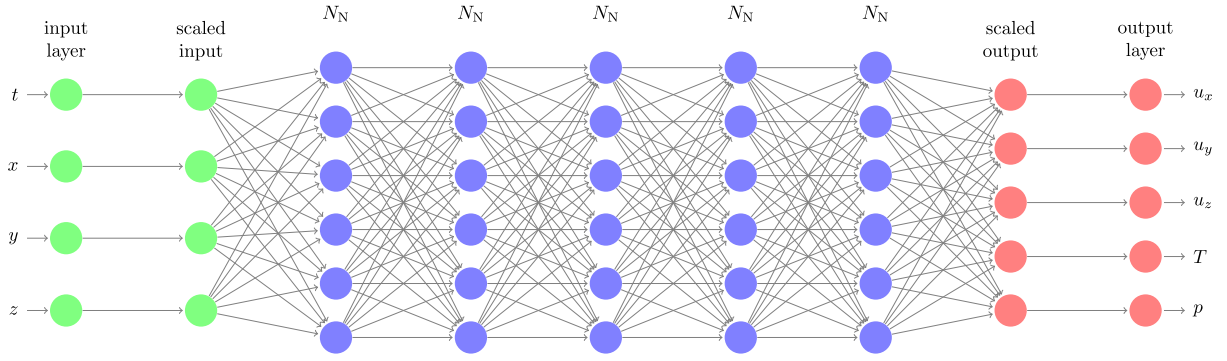


Fig. 2. Architecture of the MLP used as physics-informed neural network.

Jittery collocation. The collocation points of each training step, which are used to evaluate the PDEs, are based on the provided data points, but modified by some jitter δ_c . That means, in every training step, the dislocation δ_c obtained from a random uniform distribution within $[-\Delta_c/2, \Delta_c/2]$ is added to the coordinates of the collocation points, where Δ_c is the distance between two grid points for the coordinate ζ . This ensures that the PDEs are evaluated at a variety of positions between the original grid points over the course of the training, which is intended to prevent the formation of spurious subgrid-scale structures.

Boundary treatment. In this assimilation approach, the boundary conditions are already partially covered by the provided data, which is also sampled at the boundaries. To further support the assimilation, a separate boundary loss that covers the Dirichlet boundaries of the target fields has been implemented, i.e. the six walls for velocity assimilation or the constant temperature top and bottom plates for temperature assimilation.

These boundaries are sampled with a separate set of points. To account for the reduced density of data and collocation points used in the mini-batch of a training step, each boundary surface is represented by 8×8 spatial grid points for the same time instances provided by the data snapshots. Their spatio-temporal coordinates are then also modified by a jitter, as described above, to provide a dense boundary sampling by considering a large number of training steps. In this way, the network is provided with $\tilde{\mathbf{u}} = 0$ on all six faces of the cubic domain in the case of velocity assimilation and $\tilde{T} = \pm 0.5$ for the bottom and top face in the case of temperature assimilation.

3.3. Optimization

To perform the training, the PINN is implemented in the framework of Tensorflow [34] and Keras [35]. More specifically, the Adam optimizer [36] is used in the default configuration, except for a constant learning rate of 10^{-4} for a fixed number of 5000 epochs. The overall loss function \mathcal{L} used to train the network is a sum of several loss contributions described in more detail below. First, the data loss $\mathcal{L}_{\text{data}}$ is a mean-squared error loss of the provided data. That means, it is defined as $\mathcal{L}_{\text{data}} = \mathcal{L}_T$ in the case of velocity assimilation and $\mathcal{L}_{\text{data}} = \mathcal{L}_{\mathbf{u}}$ in the case of temperature assimilation:

$$\mathcal{L}_T = \frac{1}{N_b} \sum_{j=1}^{N_b} |T_j - \tilde{T}_j|^2 \quad (19)$$

$$\mathcal{L}_{\mathbf{u}} = \frac{1}{3N_b} \sum_{j=1}^{N_b} \|(\mathbf{u}_j - \tilde{\mathbf{u}}_j)^2\|_1 \quad (20)$$

Regarding the physics losses, the contributions of the residuals of the Navier–Stokes equations (\mathcal{L}_{NSE}), the convection–diffusion equation of the temperature (\mathcal{L}_{CDe}) and the continuity equation (\mathcal{L}_{Coe}) are calculated as shown below.

$$\mathcal{L}_{\text{NSE}} = \frac{1}{3N_b} \sum_{j=1}^{N_b} \left\| \left(\frac{\partial \mathbf{u}_j}{\partial t} + (\mathbf{u}_j \cdot \nabla) \mathbf{u}_j + \nabla p_j - \sqrt{\frac{\text{Pr}}{\text{Ra}}} (\nabla \cdot \nabla) \mathbf{u}_j - \mathbf{e}_z T_j \right) \right\|_1^2 \quad (21)$$

$$\mathcal{L}_{\text{CDe}} = \frac{1}{N_b} \sum_{j=1}^{N_b} \left| \frac{\partial T_j}{\partial t} + (\mathbf{u}_j \cdot \nabla) T_j - \sqrt{\frac{1}{\text{PrRa}}} (\nabla \cdot \nabla) T_j \right|^2 \quad (22)$$

$$\mathcal{L}_{\text{Coe}} = \frac{1}{N_b} \sum_{j=1}^{N_b} |\nabla \cdot \mathbf{u}_j|^2 \quad (23)$$

Finally, the boundary losses $\mathcal{L}_{\text{bounds}}$ are also data-based mean-squared error losses, and therefore, constructed as the data loss but for the respective other field and dedicated sampling points (cf. Section 3.2).

The computation of the physics loss terms as well as the gradients of the losses with respect to the trainable parameters is facilitated by automatic differentiation [37].

Pressure centering. An intricacy of finding solutions to the Navier–Stokes equations is the treatment of the pressure. Since the Boussinesq-approximated Navier–Stokes equations depend only on the pressure gradients, the absolute pressure level is not defined. This can lead to overall pressure levels close to -1 or 1 , which should be avoided due to the vanishing gradients caused by the activation of the scaled output layer, which is a hyperbolic tangent regardless of the case. Therefore, we introduce another loss term \mathcal{L}_{pc} , which centers the pressure values to variations around a level of 0:

$$\mathcal{L}_{\text{pc}} = \left| \frac{1}{N_b} \sum_{j=1}^{N_b} p_j \right| \quad (24)$$

Loss component weighting. Having different loss components means that the described training problem is based on a weighted-objective optimization. Therefore, weighting factors λ_i are introduced for each loss term \mathcal{L}_i to control their importance with respect to our reference loss contribution $\mathcal{L}_{\text{data}}$, i.e. $\lambda_{\text{data}} \equiv 1$. Since the physical scales contribute to the weighting of the different loss terms, we consider the use of non-dimensional equations as a first step in weighting the different contributions. Due to e.g. large gradients in the boundary layers, this per se does not guarantee PDE loss terms of an equal order of magnitude as the data losses. Yet, the order of magnitude of the bulk gradients is close to the dimensionless velocity, temperature, and pressure values for the moderate Ra and Pr numbers considered here. This implies that the meaning of the weights λ_i is close to one of importance in the present case, while for different dimensionless numbers, it may also include the compensation of the gradients within the flow.

Since adaptive weighting schemes are still under extensive investigation [38,39], this study of comparing different activations was conducted with a constant hierarchical weighting. Therefore, the λ_i are set in relation to $\mathcal{L}_{\text{data}}$, which was given the highest importance since the available data represent a direct access to the ground truth. This is especially the case for the DNS data set used here. For a possible application with data subject to uncertainty, this evaluation must certainly be lowered.

Second in the hierarchy is the loss of the PDE, which provides the essential information for the particular assimilation case. For the velocity

assimilation case, the transport of the scalar T carries information about the velocities. Therefore, the convection–diffusion equation of the temperature is used to infer u (cf. velocity measurements based on optical flow [40]). In the case of temperature assimilation, the information harnessed to infer the temperature is mainly in the buoyancy-driven acceleration of the fluid. Hence, the force balance of the Navier–Stokes equations is more important in this case. This means that $\lambda_{\text{CDe}} = 10^{-1}$ is assigned for velocity assimilation and $\lambda_{\text{NSE}} = 10^{-1}$ for temperature assimilation.

The next step in this hierarchy is occupied by the respective remaining loss contribution of those representing the Navier–Stokes and convection–diffusion equations. This means that $\lambda_{\text{NSE}} = 10^{-2}$ is assigned for the velocity assimilation and $\lambda_{\text{CDe}} = 10^{-2}$ for the temperature assimilation. Hyperparameter studies supporting this task-dependent relation between λ_{NSE} and λ_{CDe} can be found in Appendix B.

The lowest category is the same for both assimilation cases. It includes the loss contribution of the continuity equation (\mathcal{L}_{Coe}), as higher weights for this loss tend to smooth the velocity fields [16]. The boundary losses ($\mathcal{L}_{\text{bounds}}$) are also weighted in this lowest category to compensate for the fact that they are sampled separately in each training step. Finally, the pressure centering loss (\mathcal{L}_{pc}) is also set to the lowest weight, which is sufficient to keep the pressure at a reference level. Thus, the respective weights are set as follows $\lambda_{\text{Coe}} = \lambda_{\text{bounds}} = \lambda_{\text{pc}} = 10^{-3}$.

Initialization. When using periodic activation functions, the initialization of the weights is of paramount importance for a successful training result [19], since they determine the spatial and temporal wave numbers of the output of individual neurons. Therefore, we use the initialization of random weights from a uniform distribution in the interval $[-\sqrt{6/N^{(L-1)}}, \sqrt{6/N^{(L-1)}}]$ proposed by Sitzmann et al. [19], where $N^{(L-1)}$ is the number of neurons of the preceding layer.

As also discussed by Sitzmann et al. [19], the first hidden layer appears suitable to introduce higher spatial and/or temporal frequencies into the neural network. Thus, we introduce these higher frequencies or wave numbers by initializing the weights of the first layer with random numbers from an interval $[-w\sqrt{6/N^{(0)}}, w\sqrt{6/N^{(0)}}]$. Here, w serves as a factor to broaden the distribution of the initial weights and frequencies. Unless otherwise specified, $w = 4$ was used when sine activation functions were used. A parameter study on the influence of the factor w is also part of Appendix C.

3.4. Evaluation metrics

We use the DNS dataset as ground truth to monitor the evolution of the mean average error MAE_ξ and Pearson's correlation coefficient PCC_ξ with $\xi \in \{u_x, u_y, u_z, T\}$. Both are defined as follows:

$$\text{MAE}_\xi = \frac{1}{N_m} \sum_{k=1}^{N_m} |\xi_k - \tilde{\xi}_k| \quad (25)$$

$$\text{PCC}_\xi = \frac{N_m \sum_{k=1}^{N_m} \xi_k \tilde{\xi}_k - \sum_{k=1}^{N_m} \xi_k \sum_{k=1}^{N_m} \tilde{\xi}_k}{\sqrt{N_m \sum_{k=1}^{N_m} \xi_k^2 - \left(\sum_{k=1}^{N_m} \xi_k\right)^2} \sqrt{N_m \sum_{k=1}^{N_m} \tilde{\xi}_k^2 - \left(\sum_{k=1}^{N_m} \tilde{\xi}_k\right)^2}} \quad (26)$$

Here, N_m is the number of samples used for the calculation. For this purpose, every fourth grid point of each spatial direction was sampled for each snapshot.

4. Assimilation results

Since there are more studies on velocity assimilation in the literature, we will first discuss how the here presented method performs for the velocity assimilation in Section 4.1. After that, the inverse task of temperature assimilation and its characteristics are discussed in Section 4.2.

4.1. Velocity assimilation

Training process. Focusing on the influence of the applied activation functions, different numbers of neurons of each layer and the various activations were tested. For each combination, 3 training runs with different initialization were performed.

Fig. 3 shows the values of the validation metrics described in Section 3.4 obtained for the different training runs at the end of the 5000th training epoch. While the correlation coefficients (PCC) indicate how well the PINNs are able to model the existing flow structures, the additionally shown mean average errors (MAE) also reflect the reproduction of the correct amplitudes. The symbols represent the mean values of the three training runs, while the error bars display the respective minimum and maximum values.

Considering the correlation metric on the left side of Fig. 3, it is found that for all activation functions and network sizes N_N an almost perfect correlation of $\text{PCC}_T > 0.99$ is obtained for the fit of the provided temperature fields. On the other hand, the correlation of the assimilated velocity fields depends on N_N because there is an increase of the correlation with the number of neurons of the network for all activation functions.

Another trend shown by correlation values of the velocities in the different cases is that the correlation coefficients for the different components are always sorted from u_z with the highest values to u_x with the lowest values. This indicates the existence of characteristics inherent to the studied flow that make it easier to predict individual components, since the differences also occur in the two horizontal components.

To compare periodic and non-periodic activation functions, the correlation results for u_x with sine and hyperbolic tangent activations are also highlighted in the inserted plot. It shows that a periodic activation performs better than a non-periodic one for all network sizes. In particular, sine-activated networks with a layer width of 128 are reliable able to achieve correlations above 0.9 for all velocity components. It is also noteworthy that a sine-activated network with a layer width of 64 performs similarly to a hyperbolic tangent network with twice the number of neurons.

The results for the MAE metric displayed on the right side of Fig. 3 confirm the above findings with lower errors for an increased number of neurons and the same hierarchy of activation functions. In the best case, the MAE amounts to between 0.018 and 0.027 for the different velocity components. However, it is noteworthy that the fit error of the provided temperature field also depends on the varied parameters, although the correlation metric did not show large differences.

Also, doubling the number of neurons to 256 yields only marginal benefits in terms of the correlation performance. Based on this, the case of $N_N = 128$ will be investigated in more detail below as it represents the parsimonious case for this flow setup.

To examine these results in more detail, Fig. 4 shows the convergence plots of the discussed metrics (left and center) and a selection of loss terms (right) over the course of training. The lines represent the mean values obtained in three different runs. The respective envelopes of the minima and maxima are represented by the shaded areas. Thus, we can compare the periodic sine activation (darker colors) with the hyperbolic tangent activation (lighter colors), which represents the class of non-periodic activation functions, both for $N_N = 128$. Due to the random initialization, there are significant differences in the correlations metric of the velocities at the beginning of the training. However, these differences disappear during the first 100 epochs. Subsequently, they remain negligibly small for both activation functions.

Overall, the training process can be divided into several periods. During the first period, which lasts until about the 20th epoch, the training mainly optimizes the fit of the MLP to the provided temperature fields. In this phase, the physics losses are reduced due to a general attenuation of the initial gradients in the inferred fields. After that,

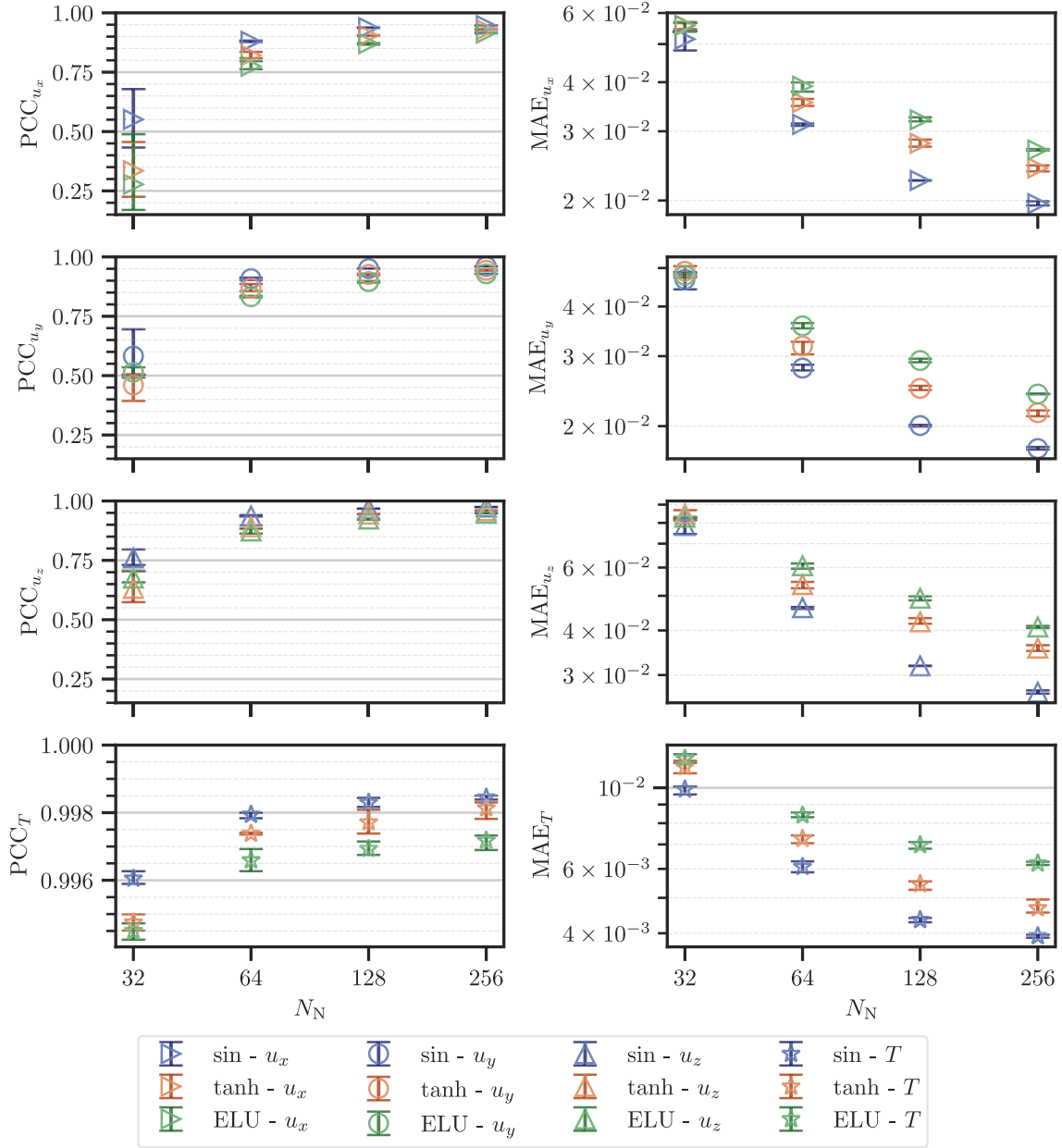


Fig. 3. Correlation (left) and MAE (right) metrics of the velocity and temperature fields (top to bottom) achieved after 5000 epochs of velocity-assimilation training with varying activation functions and sizes (N_N) of the hidden layers. The markers represent averages over multiple training runs and the error bars the respective minima and maxima.

the temperature fields are well enough represented by the MLP, so that the reduction of the physics losses starts to improve the inferred velocities. After a few hundred epochs, the correlation coefficient for the inferred velocity components starts to converge, meaning that the PINN has recovered the most salient velocity structures. This also marks the beginning of the decrease of the associated MAEs, showing that the physics loss terms are not only able to reconstruct the structures, but also push the inferred fields towards the quantitatively correct ones. From about epoch 1000 onward the MAEs and losses appear to flatten out, while they also begin to oscillate.

In summary, both sets of runs would yield negligible improvements in the validation metrics after the observed training period. The difference between the periodic and non-periodic activation functions is that the sine activation is able to fit the provided temperature field faster (except for the first 10 epochs) and better. We attribute this to the periodic nature of the sine, which provides more complexity for a

given number of neurons. This is beneficial for the inferred fields in two ways: First, because they also require a certain amount of complexity provided by the MLP to be reconstructed, and second, as a better fit to the given temperature fields yields more accurate gradients to the physics losses.

Inferred fields. Fig. 5 shows a visual comparison of the ground truth fields (top row) with the inferred fields (middle row) in a vertical cross-section of the domain at $x \approx 0.5$, obtained with sine activation and $N_N = 128$. Additionally, the absolute differences between the top and middle fields are shown in the bottom row. While the differences (bottom row) between the temperature fields of the PINN and the ground truth are close to zero, the velocity differences are clearly visible.

However, looking at the structures in the top and middle rows, it is found that they are qualitatively similar, although the amplitudes

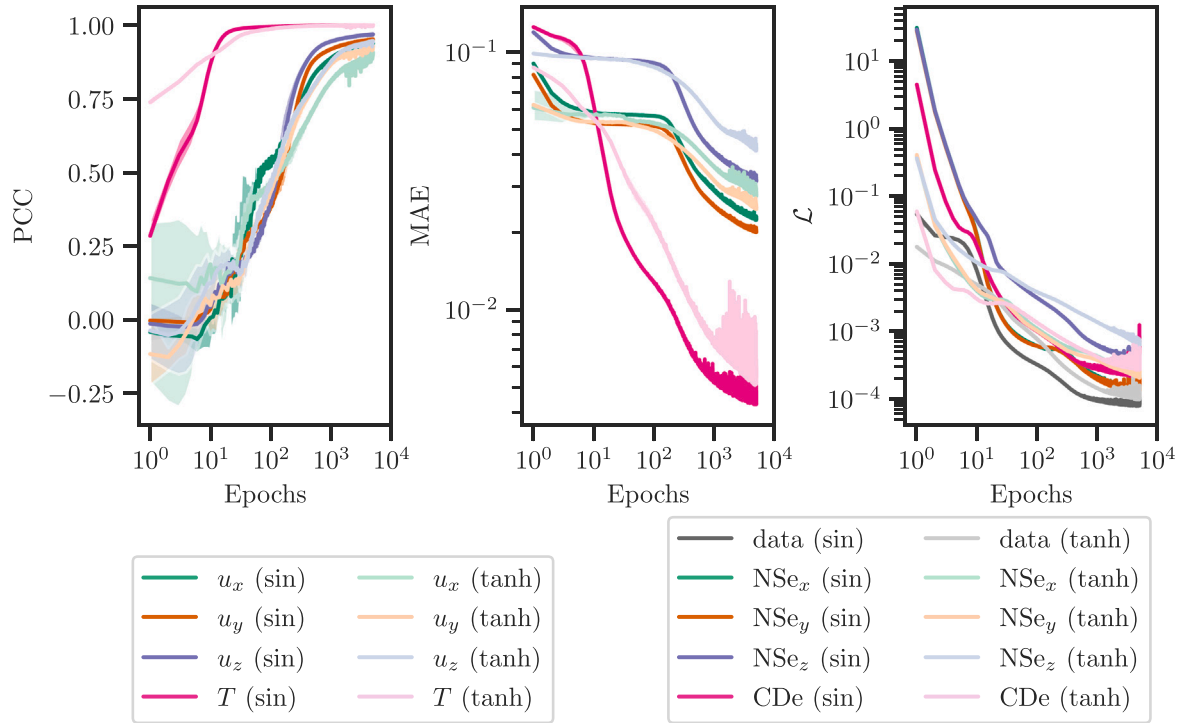


Fig. 4. Evolution of the correlation (left), MAE (center) and losses (left) during the training. Darker lines represent the sine-activated cases and lighter lines hyperbolic tangent-activated ones. Each solid line describes the mean of the single velocity-assimilation training runs at $N_N = 128$, while the respective transparent envelope is bound by the minimum and maximum values.

are different. This is confirmed by the plots of the absolute differences, where the largest deviations occur at the same positions as the high velocity magnitudes in the ground truth data set. To further investigate these differences in velocities between the PINN reconstruction and the ground truth, we subsequently analyze how well the PINN is able to reproduce extreme values and small structures.

The reproduction of extreme values is considered by analyzing the global histograms of vertical velocity (u_z), temperature (T) and vertical convective heat transport ($u_z T$) displayed in Fig. 6. While the gray lines represent the ground truth data, the colored lines correspond to the differently activated PINN models for $N_N = 128$.

First, we consider the temperature data displayed in Fig. 6(b) as it concerns the data provided to train the PINN. Besides a central peak, the distribution of the ground truth temperature also shows additional peaks at $T = \pm 0.5$, which refer to the constant temperature boundary conditions. This distribution is closely reproduced by the PINN models regardless of the activation function, except for the regions near $T = \pm 0.5$. There, all cases show a shift of the extreme boundary values towards zero, showing that the extreme gradients within the boundary layer are difficult to model with a PINN, even if the corresponding data are available.

Regrading u_z (Fig. 6(a)), the histograms of all PINN models are significantly narrowed, i.e. they underestimate the occurrence of extreme values. This is consistent with the lack of amplitude visible in the center sections presented in Fig. 5. However, this effect is less pronounced for the sine activation than for the non-periodic functions, demonstrating that the PINN's ability to model the amplitudes of the velocity vector field benefits from a periodic activation.

In addition, the PDFs of the convective heat fluxes $u_z T$ are displayed in Fig. 6(c). There, the distributions of the PINN-modeled heat fluxes display the same narrowed characteristic as for the vertical velocity component. However, the main differences between the various activation functions occur for the positive tail of the distribution, which constitutes the strong positive skew of the ground truth. Again, in this

regard, the sine activated PINNs perform best again, as they model the right tail the closest.

To test the ability to reproduce the turbulence structure in the assimilated velocity field, power density spectra $\phi_{u_z u_z}(k_\zeta)$ are shown in Fig. 7 in a central horizontal plane $z_p = 0.5$ (top) and a parallel plane near the heated bottom plate $z_p = 0.02$ (bottom). The spectra also distinguish between the wave numbers k_x (left) and k_y (right) in the two planar directions. For each of these positions, the figure displays the spatial spectra evaluated with the DNS data and the inferred fields using sine and hyperbolic tangent activations at $N_N = 128$. Note that these results are obtained by applying Fast Fourier Transforms (FFTs) to non-periodic signals, which means that they are affected by spectral leakage. This is particularly noticeable for some spectra that flatten out at high wave numbers. This effect is mitigated by the no-slip boundary conditions, which impose something like a natural window function. However, since this leakage is present in both the ground truth and inferred field data, the comparison is still possible.

In this comparison, the spectra of the PINN-generated fields exhibit lower power densities overall, regardless of plane, direction, and wave number. Of the two activation functions shown, the sine activation generally exhibits smaller differences to the ground truth of the spectral power densities than the hyperbolic tangent activation, confirming the observations made with respect to Figs. 5 and 6.

However, it cannot be observed that the gap of the power density spectra between the PINN models and the ground truth would grow towards higher wave numbers k . This means that the impression of the inferred velocity fields in Fig. 5 is not to be understood as a blurring, as this would result in small scales being suppressed more than the larger scales. On the contrary, the softer appearance of the inferred fields is caused by a general suppression of the amplitudes. For the plane at $z_p = 0.02$, the gap between ground truth and inferred vertical velocity actually closes for large wave numbers. This may be caused by an underrepresentation of the large scales in the inferred vertical velocity u_z , while the existing smaller structures are still well represented.

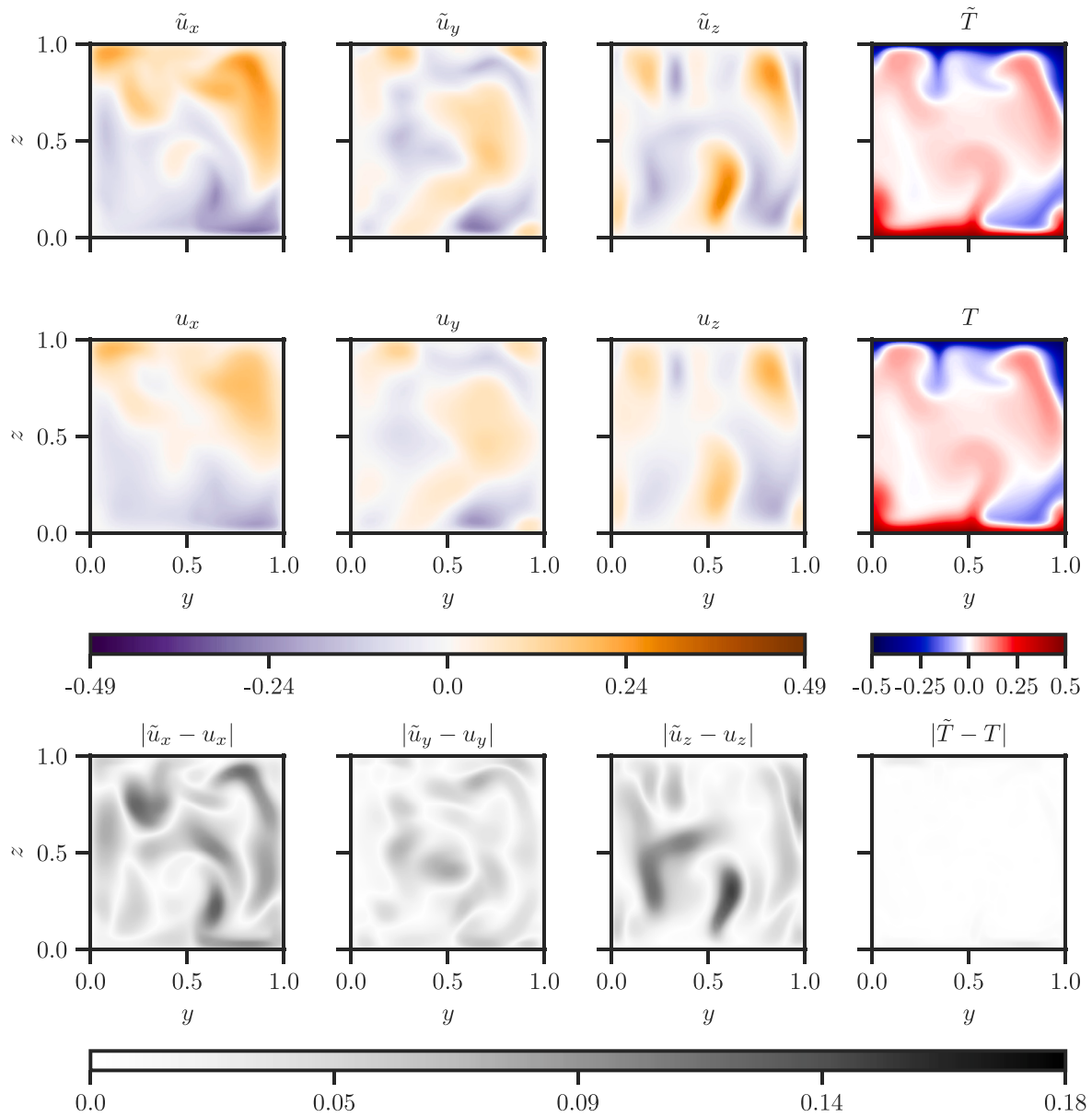


Fig. 5. Comparison of the ground truth (top row) and PINN-generated (middle row) fields of the velocity components u_x , u_y , u_z and the temperature T in a central vertical cross-section for velocity assimilation with sine activation and $N_N = 128$. The bottom row comprises the contour plots of the respective absolute differences.

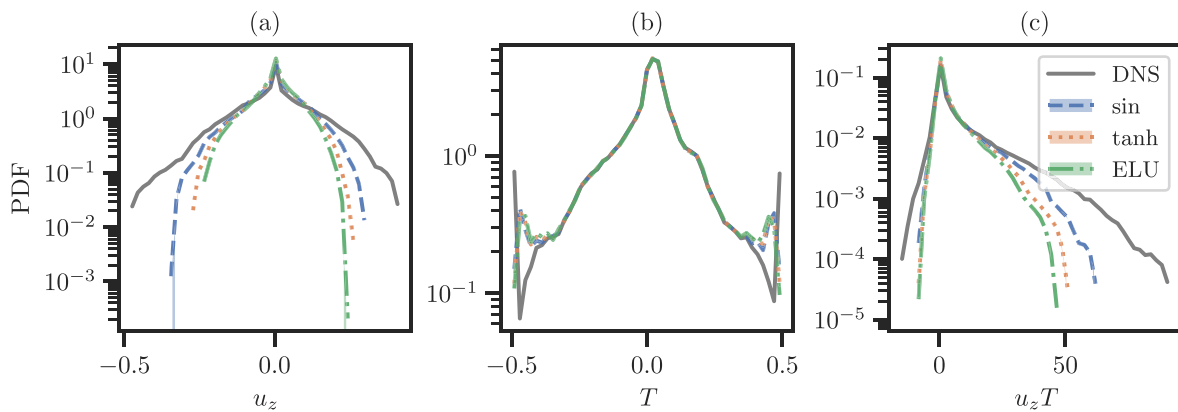


Fig. 6. Global histograms of the vertical velocity component u_z (a), the temperature T (b) and the vertical convective heat flux $u_z T$ (c) for the velocity assimilation cases with $N_N = 128$. The lines represent an average of the respective training runs, while the envelopes are bound by the minimum and maximum values.

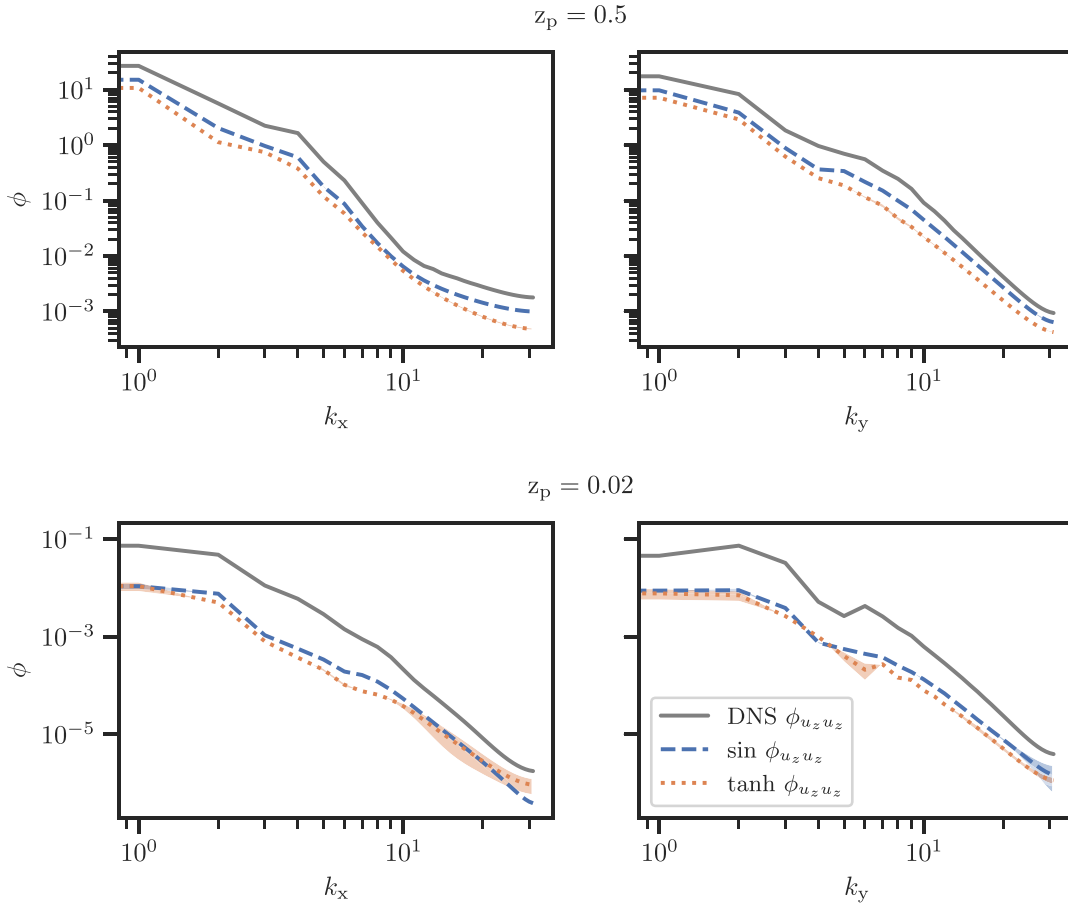


Fig. 7. Comparison of the spectra in x - and y -directions of the u_z field of the ground truth (gray) and sine (blue) and hyperbolic tangent (orange) activation with $N_N = 128$. The top row comprises the spectra in the central horizontal plane and the bottom row for a plane close to the boundary, while the columns comprise the different directions k_x (left) and k_y (right). The lines display the average over the runs while the envelopes are bound by the respective minima and maxima.

With the PINN, which minimizes the losses in form of the residuum of the Navier–Stokes equations, the pressure fields are also obtained during the assimilation. We have chosen to omit their investigation in the main part of this paper in order to keep the focus on temperature and velocity, which contribute to the convective heat transport. For insights into the inferred pressure fields, we refer the reader to [Appendix D](#).

4.2. Temperature assimilation

Training process. Following the structure of the discussion of the above section, which deals with velocity assimilation, an overview of the performance of the temperature assimilation with different activation functions and neuron numbers N_N is shown in [Fig. 8](#). Here, all combinations of neuron number² and activation function were analyzed based on three training runs. The only exception is the best case, $N_N = 128$ with sine activation, where five training runs were performed. The displayed markers are again averages over the training runs, and the error bars represent the respective minima and maxima.

Overall, the two plots display the trend that the correlation coefficients ([Fig. 8](#), left side) increase and the absolute errors decrease ([Fig. 8](#), right side) with increasing neuron number as it was already

observed for the velocity assimilation. To investigate the characteristics of the temperature assimilation task, we first discuss the correlation coefficients in more detail. Starting with $N_N = 32$, the achieved correlations are between 0.5 and 0.7, which is rather low. At this stage, the sine activated neural network has the lowest correlation coefficient when comparing the activation functions, which means that it performs the worst in this configuration, while ELU performs the best.

For networks twice the size ($N_N = 64$), the increase in correlation values is not as great as for the velocity assimilation. More strikingly, the variations between the different runs of both sine and ELU activations increase with this doubling of N_N .

For the next doubling of N_N to 128, there is a transition where the PINN models reach correlations of $\rho > 0.8$ for all activation functions with sine now giving the best results with $\rho > 0.94$. Also, the initialization-induced variations of the results decrease again. To highlight this transition to viable inference results, the metrics for individual intermediate training runs for $N_N \in \{74, 84, 97, 111\}$ have been added to the plots, where they show the sine activation surpassing the other activation functions for the temperature metrics. The next doubling to $N_N = 256$ offers only small improvements, confirming that the relevant transition takes place between 64 and 128 neurons per hidden layer. As for the velocity assimilation, this renders $N_N = 128$ as parsimonious for the given task, so the detailed analysis is performed for this neuron number.

Regarding the fit of the provided velocity fields, correlation coefficients close to 1 are obtained in all cases. However, the associated MAE values ([Fig. 8](#), right side) decrease with increasing neurons. Regardless of the number of neurons, the lowest absolute errors are again obtained

² We limit our variation of the neuron number to changes of the width of the hidden layers to keep the investigations concise. Another option would be a variation of the number of hidden layers. A brief overview of the effects of this hyperparameter can be found in [Appendix A](#).

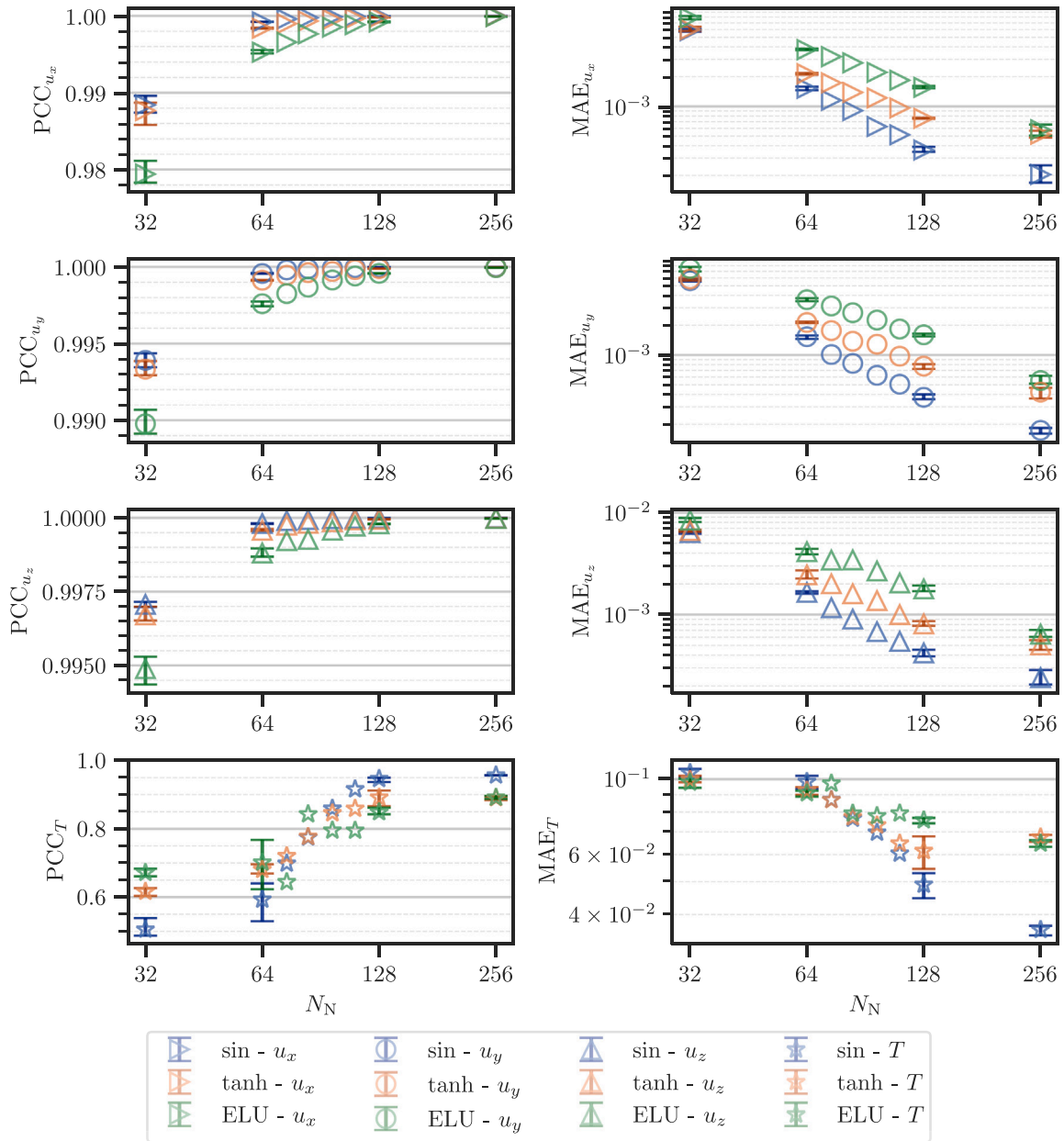


Fig. 8. Correlation (left) and MAE (right) metrics of the velocity and temperature fields (top to bottom) achieved after 5000 epochs of temperature-assimilation training with varying activation functions and sizes (N_N) of the hidden layers. The markers represent averages over multiple training runs and the error bars the respective minima and maxima.

with sine activation, confirming similar observations for the velocity assimilation.

Compared to the qualitative differences of the velocity errors obtained for the different activation functions, the differences between activations for the assimilated temperature fields are smaller (in terms of logarithmic scaling), but at a higher level overall. However, they follow the same trends as the correlation coefficients with the sine activation reaching $MAE_T \approx 0.049$ at $N_N = 128$ and $MAE_T \approx 0.036$ at $N_N = 256$.

To investigate the performance of the training processes for the different activation functions, Fig. 9 shows the correlation and MAE metrics along with selected losses for the training process of test cases with sine (darker) and hyperbolic tangent (lighter) activation at $N_N = 128$. The lines represent averaged values over the five or three test runs, respectively, while the shaded area represents the respective envelope between minima and maxima.

As a first observation, these envelopes cover only insignificantly small regions with the exception of the correlation metric (left side of Fig. 9) of the inferred temperatures. There, variations of up to 0.2 can occur between individual training runs. However, this envelope still collapses towards the end of the training at 5000 epochs, meaning that the final results do not depend on a favorable random initialization.

Regarding the correlation progress of the training with respect to the fields fitted to the ground truth data — here, the velocity fields — the sine activation produces high ($PCC_u > 0.9$) correlation values significantly faster than the hyperbolic tangent activation. This confirms the finding from the velocity assimilation, where sine-activated PINNs are also faster in fitting the provided data.

Next, the evolution of the correlation coefficients for the assimilated temperature fields is discussed: They initially grow strongly to $PCC_T \approx 0.4$ for both activation functions considered. After that, the increase of the correlation value is slower, before the values grow faster again (on

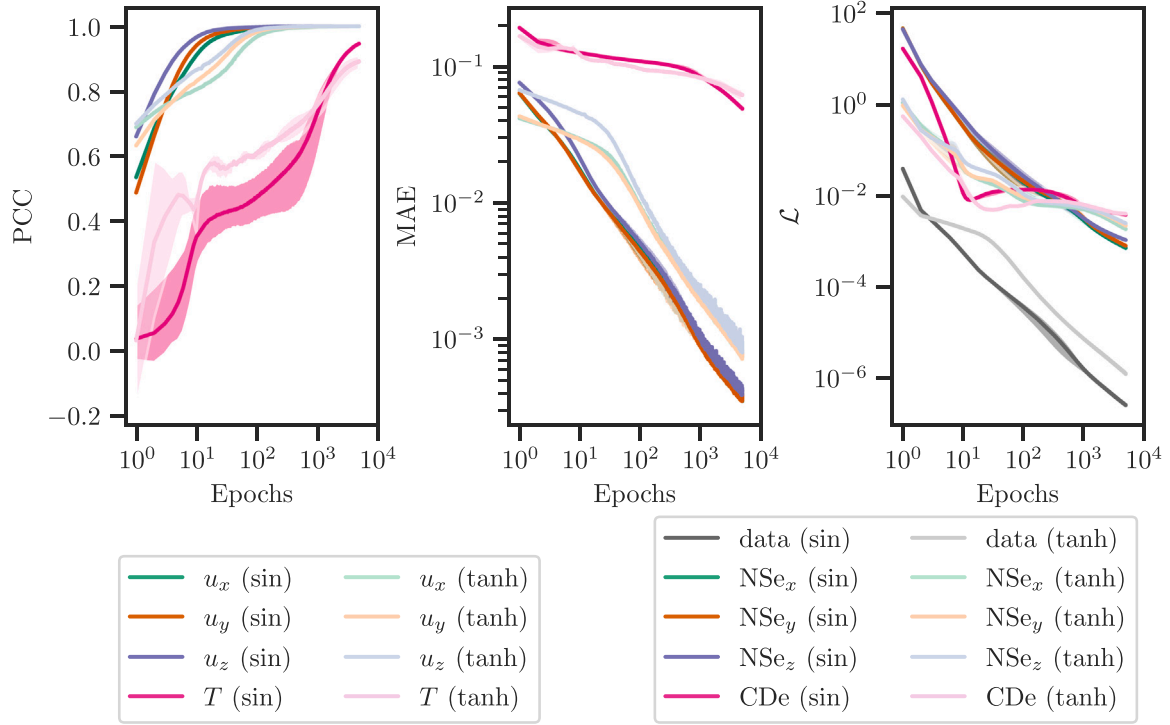


Fig. 9. Evolution of the correlation (left), MAE (center) and losses (left) during the training. Darker lines represent the sine-activated cases and lighter lines hyperbolic tangent-activated ones. Each solid line describes the mean of the single temperature-assimilation training runs at $N_N = 128$, while the respective transparent envelope is bound by the minimum and maximum values.

a logarithmic epoch scale) at about 1000 epochs. This second growth period is significantly stronger for the sine activation, i.e. it surpasses the performance of the hyperbolic tangent during this period.

These two periods of faster growth separated by a slower one are also obtained for the MAE metrics (center plot of Fig. 9) of the assimilated temperatures.

Considering the velocity MAEs and the losses (right side of Fig. 9), decays are observed towards the end of the training, while for the velocity assimilation they already started to level off after about 1000 epochs. This means that PINNs are harder to train for temperature assimilation than for velocity assimilation because the 5000 epochs are not sufficient to achieve a leveling out as in the velocity assimilation case. A fixed limit of 5000 epochs is still justifiable, since the logarithmic epoch axis still means strongly diminishing returns for further training, especially regarding the correlation values.

Inferred fields. The comparison between exemplary fields generated by the sine-activated PINN at $N_N = 128$ and the ground truth is displayed in Fig. 10. It includes velocity and temperature fields in a vertical central section for the ground truth (top row) and the PINN (middle row) as well as the resulting absolute differences (bottom row).

Similar to the case of velocity assimilation, the fit of the fields for which data were provided to train the PINN shows hardly any noticeable absolute errors. However, this is not the case for the inferred temperature field. While the PINN is able to reproduce all significantly warm or cold structures, such as the boundary layers and the plumes originating from them, it particularly struggles in the bulk region. This especially applies to the region of $0.25 < z < 0.5$ for the displayed vertical center section. Since the actual temperature amplitudes are smaller in this region, so is the resulting buoyancy acceleration. This finally makes an accurate reconstruction of the temperature more difficult.

To investigate how the described differences between the ground truth and the inferred fields manifest themselves in the flow statistics, the global histograms of u_z , T and $u_z T$ are compared for the different activation types at $N_N = 128$ in Fig. 11.

As expected from the results discussed above, the differently activated PINNs all manage to reproduce the distribution of the vertical velocity, shown in Fig. 11(a), without significant differences.

Regarding the histograms of the inferred T fields shown in Fig. 11(b), the differently activated PINNs also follow the expectations based on the velocity assimilation and are not able to cover the whole width of the temperature range, leading to slightly higher values of the probability density function for moderate temperature amplitudes. While the sine and hyperbolic tangent activations perform almost equally, the ELU activation is able to cover the widest temperature range. Yet, it also underestimates the occurrence of temperatures close to $T = 0$.

The histogram of $u_z T$ in Fig. 11(c) shows the characteristic right-skewed distribution of convective heat fluxes for the ground truth and the investigated PINN models. While the ELU-activated PINN was able to produce the widest temperature distribution, it also exhibits the largest deviation of the tested activations for the negative tail of $u_z T$, where all activations overestimate the occurrence of locally negative convective heat fluxes. We attribute this to an inherent mismatch between the large temperature amplitudes and the corresponding vertical velocities. In contrast, both tails of the histogram $u_z T$ are close to the ground truth for the sine activation. The hyperbolic tangent activation shows similar deviations to the ELU activation, but with more underestimation of positive values and less overestimation of negative values. This illustrates that the distributions of u_z and T alone are not sufficient to evaluate the quality of the predictions.

Overall, the inferred convective heat flux histogram is in much better agreement with the ground truth for the temperature assimilation

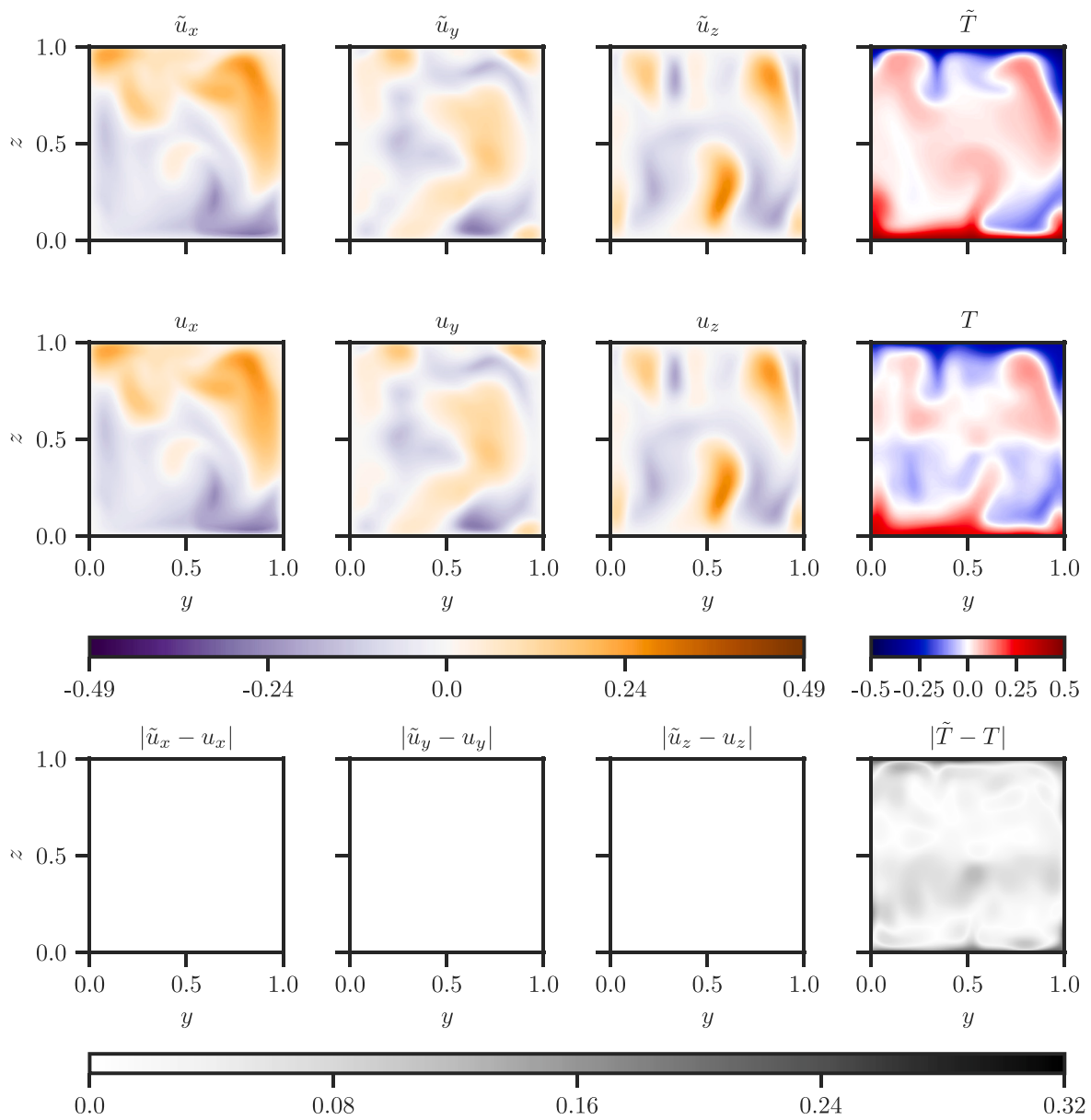


Fig. 10. Comparison of the ground truth (top row) and PINN-generated (middle row) fields of the velocity components u_x , u_y , u_z and the temperature T in a central vertical cross-section for temperature assimilation with sine activation and $N_N = 128$. The bottom row comprises the contour plots of the respective absolute differences.

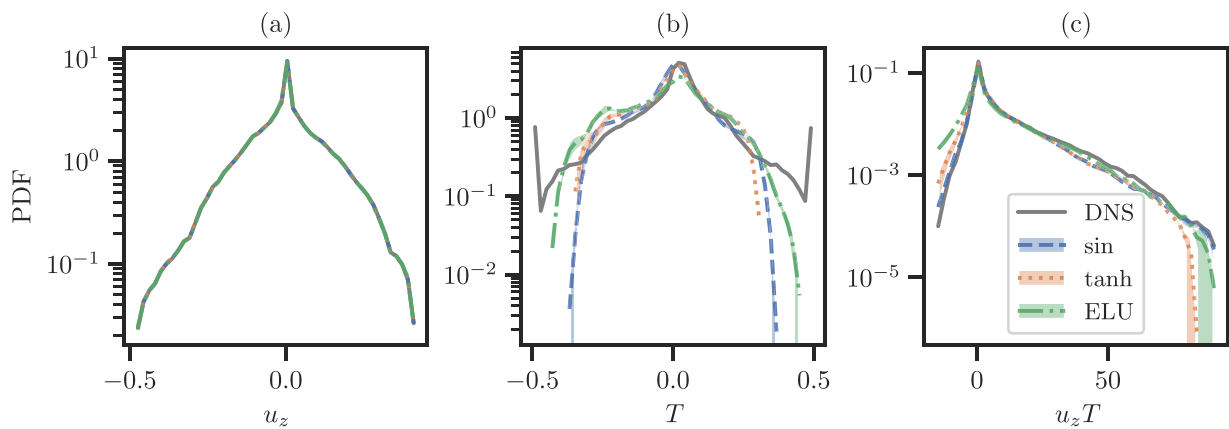


Fig. 11. Global histograms of the vertical velocity component u_z (a), the temperature T (b) and the vertical convective heat flux $u_z T$ (c) for the temperature assimilation cases with $N_N = 128$. The lines represent an average of the respective training runs, while the envelopes are bound by the minimum and maximum values.

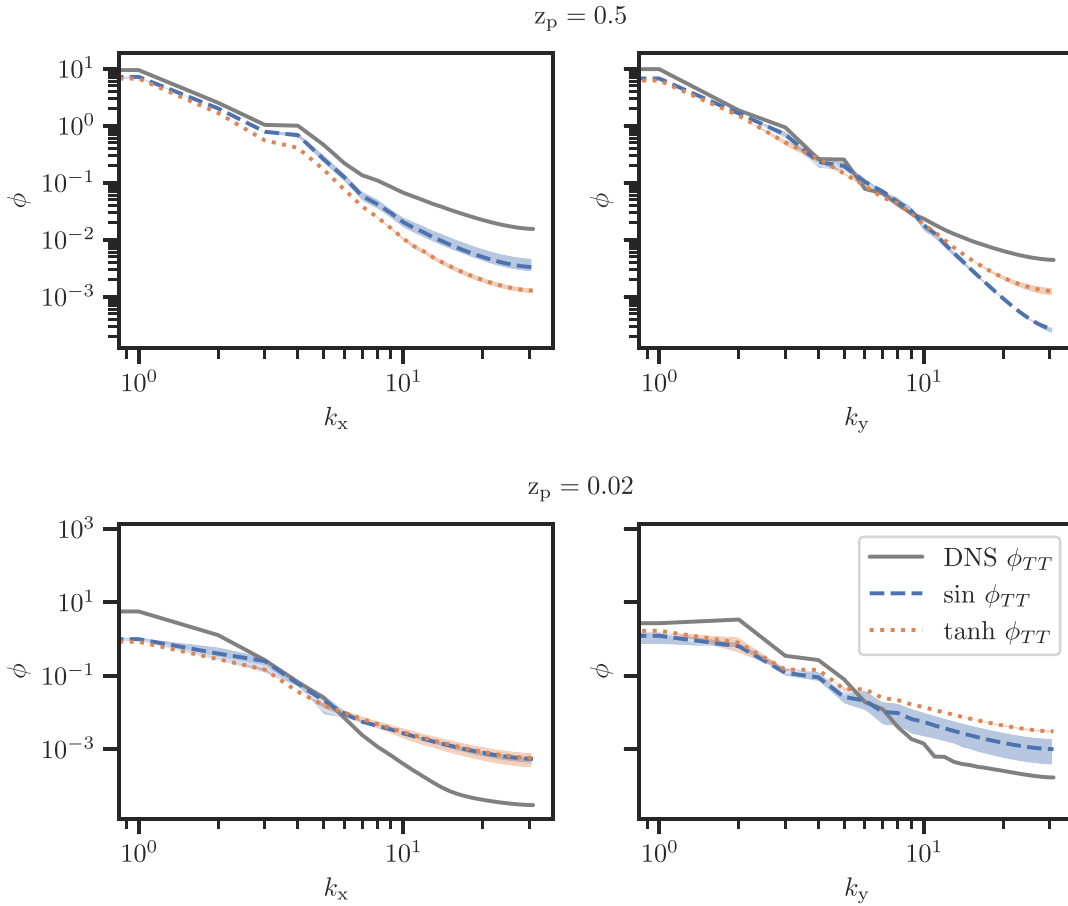


Fig. 12. Comparison of the spatial spectra of the ground truth (gray) and the inferred temperature fields for sine (blue) and hyperbolic tangent (orange) activation. The top row comprises the spectra at the central horizontal plane and the bottom row for one close to the boundary, while the columns comprise the different directions k_x (left) and k_y (right). Data shown for the training runs with $N_N = 128$. The lines display the average over the runs while the envelopes are bound by the respective minima and maxima.

than for the velocity assimilation. This is due to the fact that the largest temperature deviations occur in regions where the convective heat flux is low, namely the bulk region with low temperature amplitudes and the boundary layers with low vertical velocities.

The power density spectra of the temperature fields are shown in Fig. 12 for the x (left) and y (right) directions in the planes $z_p = 0.5$ (top) and $z_p = 0.02$ (bottom), to investigate the ability of the differently activated PINNs to assimilate structures of different sizes. Following the systematics of the comparable Fig. 7, the sine and hyperbolic tangent activations represent both periodic and non-periodic activations for the displayed cases of $N_N = 128$.

In the bulk flow at $z_p = 0.5$, the inferred temperature fields of both activations, sine and hyperbolic tangent, show that the gap of spectral power density widens towards higher wave number, implying a deficiency in the reproduction of smaller structures. Since none of the power densities predicted with the activation functions agrees well with the ground truth in either direction, there is no clear advantage for one of the two activations. A possible explanation for this shortcoming could be numerical effects that add to the thermal diffusivity. However, the power density spectra for the vicinity of the bottom plate ($z_p = 0.02$) question this explanation, as they both exhibit an overestimation of power densities for large wave numbers. This translates to an overestimation of the temperature variation on the small scales while the

variations on the larger scales are still underestimated. Therefore, we explain this by smaller structures leaking in from the region farther from the bottom plate, while large scale structures are still suppressed because the PINNs fall short on reproducing the steep gradients of the boundary layer.

5. Conclusion and outlook

The assimilation of both velocity and temperature has been pursued for a case of cubic Rayleigh–Bénard convection at $Ra = 10^6$ and $Pr = 0.7$. The results show that the PINN approach with a periodic activation function is advantageous for the assimilation in both directions compared to other widely used activation functions, which is primarily expressed by the high correlation coefficients of the inferred fields with the ground truth. We attribute the advantages of a sine activation to its ability to express the structures of the flow. An example of this is vortices, which are ubiquitous in turbulent flows and often appear as counter-rotating pairs or even clusters. To model just the multiple sign changes of a given velocity component of such structures, several monotonically activated neurons are required, whereas a single periodically activated neuron may be sufficient. Furthermore, sine functions are resilient against the vanishing gradient problem [19], since their derivative is close to zero only in small, periodically occurring regions.

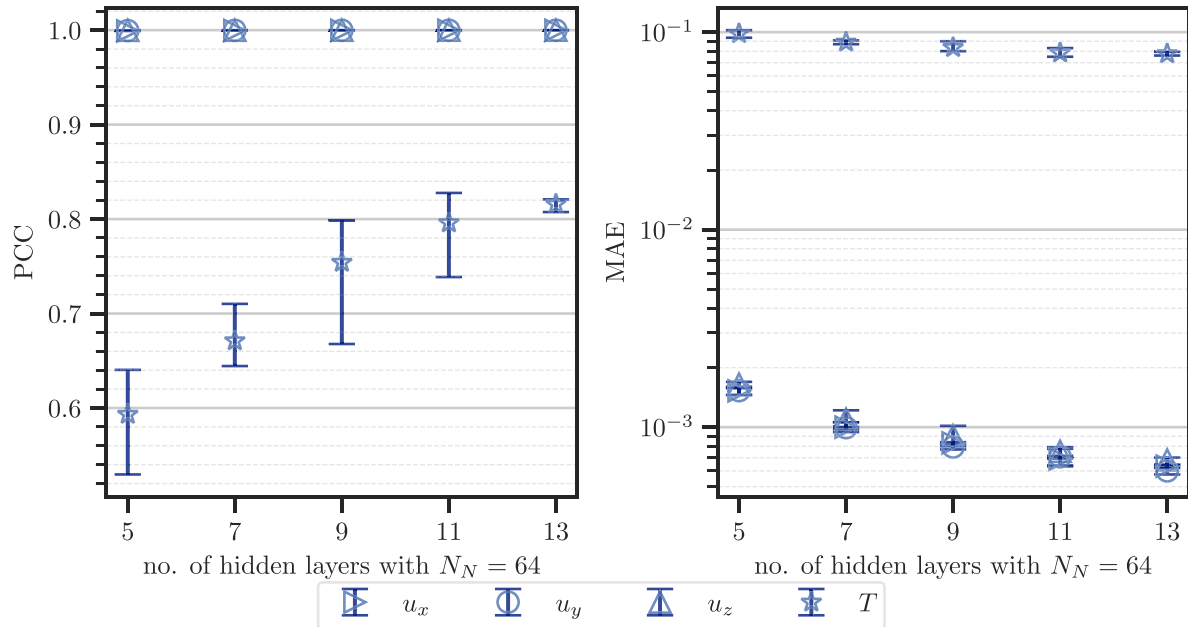


Fig. A.13. Correlation (left) and MAE (right) metrics of the velocity and temperature fields achieved after 5000 epochs of temperature-assimilation training for $N_N = 64$ with a varying number of hidden layers for the sine activation. The markers represent averages over multiple training runs and the error bars the respective minima and maxima.

ELU or tanh, on the other hand, produce vanishing gradients for the negative tail or both tails of the respective functions.

This means that a periodic activation with sine functions eases the tension between computational speed, cost and accuracy, as it typically produces higher quality results than the other activation functions investigated, hyperbolic tangent and ELU, for the same network size.

However, the inferred fields still exhibit some deviations regarding the absolute values. Further increasing the size of the MLP architecture mitigates this problem, as our results show decreasing errors for larger neural network layers.

The presented investigations also revealed a limitation regarding the benefits of a sine activation. We found that, for temperature assimilation, sine activations were only beneficial for sufficiently large network sizes.

Overall, we found that the direction of assimilation, generating velocity fields from temperature data or vice versa, has further significant influences on the results: Compared to velocity assimilation, temperature assimilation exhibited a critical range of the number of neurons of each hidden layer, which has to be exceeded to obtain viable results. Furthermore, the convergence of the training is shifted to a larger number of epochs for the temperature assimilation. Due to the complexity of the task, there are several possible reasons for this: It could be caused by the different weighting of the PDEs in action, the overall conditioning of the PDE-terms that are crucial for the inference, or more mundane facts, e.g. that the MLPs have three fields to fit and two to infer in the temperature assimilation case, while the ratio is one to four for the velocity assimilation. Since the temperature assimilation is more challenging, the deviations of the inferred fields from the ground truth are visually more prominent. However, the convective heat flux statistics are predicted significantly better than for the velocity assimilation, as its regions prone to error have less influence on the statistics.

Future challenges include pushing for higher Ra and Pr numbers, for which it may not be feasible to provide fully resolved fields. With respect to the necessary expressiveness for the temperature assimilation, $N_N = 128$ also proved to be close to the lower threshold

for obtaining viable results, which we expect to increase for higher dimensionless numbers. At the same time, further features such as von Neumann boundary conditions or geometrically more complex boundary shapes should be implemented to improve the assimilation results by providing more information about the flow, which is typically available. Besides that, network architecture, loss formulations, and the training process prove to be areas with a vast scope for design and thus also optimization potential. Relevant examples of such optimizations have been presented only recently by Toscano et al. [41] and promise efficiency improvements.

CRediT authorship contribution statement

Michael Mommert: Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Robin Barta:** Visualization, Software, Methodology. **Christian Bauer:** Writing – review & editing, Writing – original draft, Supervision, Resources, Data curation. **Marie-Christine Volk:** Writing – review & editing, Visualization, Formal analysis. **Claus Wagner:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The authors gratefully acknowledge the scientific support and HPC resources provided by the German Aerospace Center (DLR). The HPC system CARA is partially funded by “Saxon State Ministry for Economic

Table B.2

Correlation Performance of the different assimilation tasks for a sine activated PINN with $N_N = 128$ after 5000 epochs averaged over the respective number of runs (three except for the temperature assimilation with the corresponding weight set with five runs).

	u assimilation				T assimilation			
	PCC_{u_x}	PCC_{u_y}	PCC_{u_z}	PCC_T	PCC_{u_x}	PCC_{u_y}	PCC_{u_z}	PCC_T
weight set u	0.938	0.951	0.968	0.998	1.000	1.000	1.000	0.856
weight set T	0.892	0.886	0.929	0.999	1.000	1.000	1.000	0.946

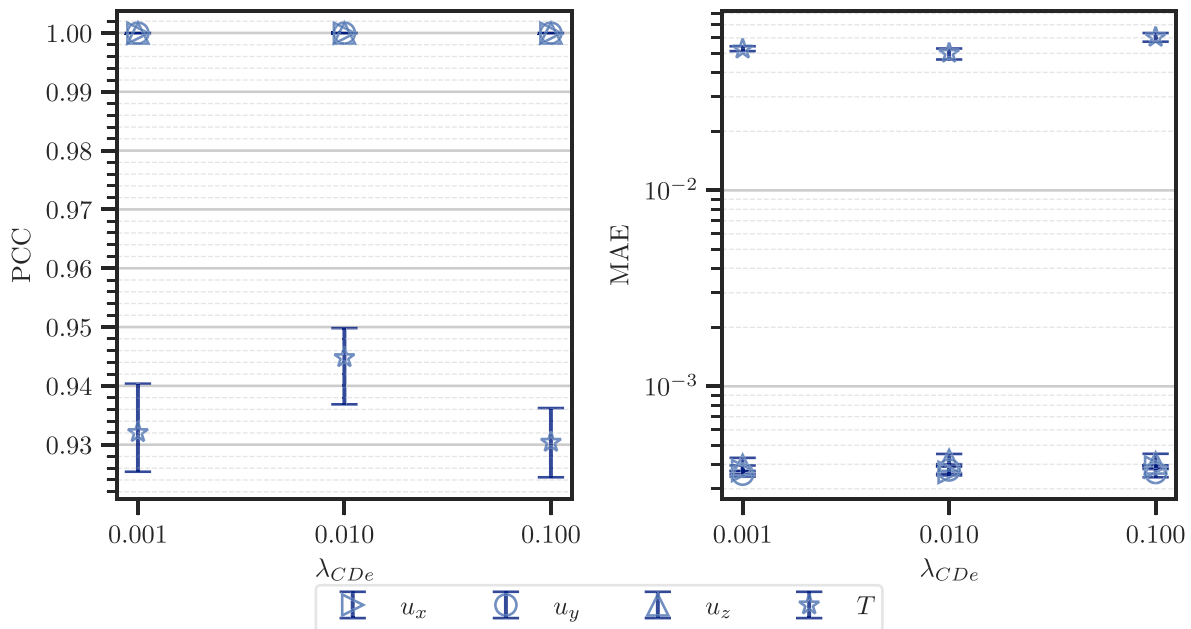


Fig. B.14. Correlation (left) and MAE (right) metrics of the velocity and temperature fields achieved after 5000 epochs of temperature-assimilation training for $N_N = 128$ with a varying convection–diffusion loss weight λ_{CDe} for the sine activation. The markers represent averages over multiple training runs and the error bars the respective minima and maxima.

Affairs, Labour and Transport” and “Federal Ministry for Economic Affairs and Climate Action”.

Appendix A. Hyperparameter study — number of hidden layers

While in the main part of the paper the layer width N_N is varied to adjust the complexity of the neural network for different tasks and activation functions, this appendix aims to give an outlook on the variation of the number of hidden layers. For this purpose, we test the temperature assimilation for sine activated PINNs with a constant layer width of $N_N = 64$, which was below the threshold for obtaining usable results with five hidden layers of neurons. Hence, three test runs each were performed for increasing numbers of hidden layers up to 13. As expected, the results displayed in Fig. A.13 show an increase in assimilation performance for larger numbers of hidden layers. Yet, the correlation values of a PINN of 5 hidden layers with $N_N = 128$ are not reached by the configuration of 13 hidden layers with $N_N = 64$. Overall, the determination of an optimal aspect ratio for the MLP architecture requires further investigation, which should be part of future studies.

Appendix B. Hyperparameter study — selected loss weights λ_i

In Section 3.3 paragraph on weighting the loss components, two different sets of loss weights λ_i are introduced. Specifically, there is one set for velocity assimilation ($\lambda_{NSe} = 10^{-2}$, $\lambda_{CDe} = 10^{-1}$, $\lambda_{Coe} = \lambda_{bounds} = \lambda_{pc} = 10^{-3}$) and one for temperature assimilation ($\lambda_{NSe} = 10^{-1}$, $\lambda_{CDe} = 10^{-2}$, $\lambda_{Coe} = \lambda_{bounds} = \lambda_{pc} = 10^{-3}$). To underpin the distinction between a convection–diffusion driven velocity assimilation and a Navier–Stokes driven temperature assimilation, we also performed three training runs with the opposite set of weights. The results in Table B.2 show that distinguishing between the different assimilation mechanisms is indeed advantageous for the assimilation tasks.

To provide more detail on how the convection–diffusion equation of the temperature affects the performance of the temperature assimilation, a small hyperparameter study of λ_{CDe} is provided by Fig. B.14. For this study, the remaining set of weights is kept constant at $\lambda_{NSe} = 10^{-1}$ and $\lambda_{Coe} = \lambda_{bounds} = \lambda_{pc} = 10^{-3}$. Besides the existing 5 runs for $\lambda_{CDe} = 10^{-2}$ from the main investigation, 3 additional runs were performed for $\lambda_{CDe} = 10^{-3}$ and $\lambda_{CDe} = 10^{-1}$. Overall the final correlation coefficients and absolute errors achieved by the training exhibit

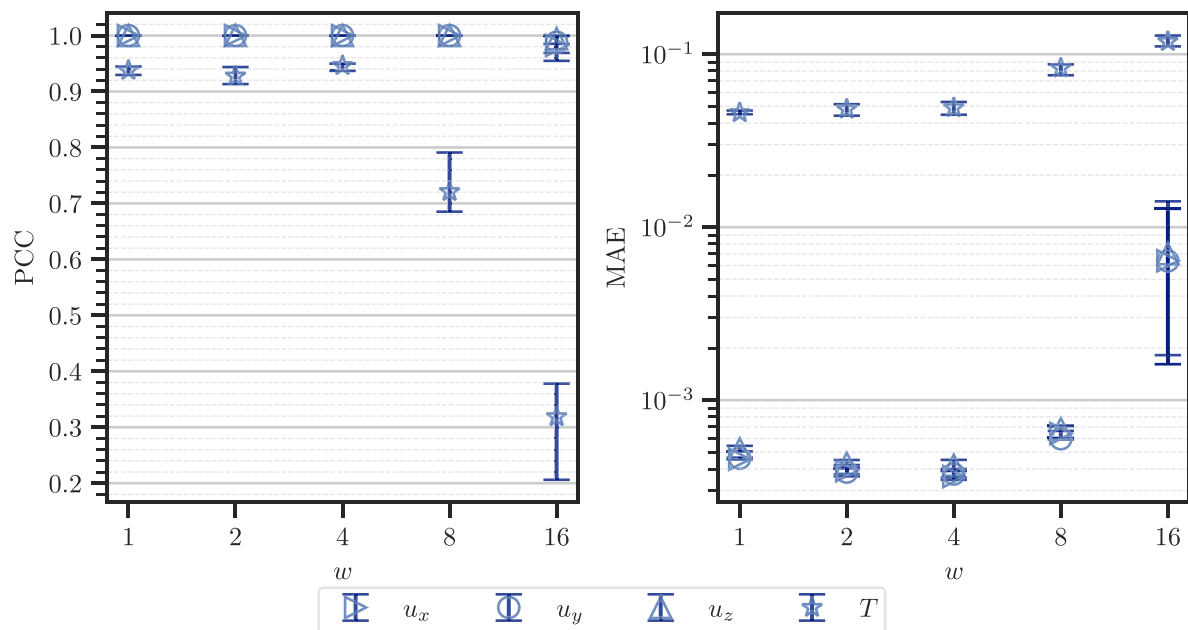


Fig. C.15. Correlation (left) and MAE (right) metrics of the velocity and temperature fields achieved after 5000 epochs of temperature-assimilation training for $N_N = 128$ with a varying initialization parameter w for the sine activation. The markers represent averages over multiple training runs and the error bars the respective minima and maxima.

only small variations, although the order of magnitude of λ_{CDE} was varied. This confirms, that the loss of the convection–diffusion equation is less relevant for the performance of the temperature assimilation than the Navier–Stokes loss. Furthermore, the decrease in assimilation performance is slightly more pronounced for increased λ_{CDE} . A possible reason for this behavior could be the existence of a trivial solution $T = \text{const.}$ for the convection–diffusion equation. A similar behavior can be observed for the continuity equation with the trivial solution $u = 0$ which leads to decaying PINN performances if weighted to strong [16].

Appendix C. Hyperparameter study — initialization factor w

As mentioned in Section 3.3, the sine activation also allows for different initialization strategies. Here we varied the prefactor w of the range of the random uniformly distributed initial weights of the first hidden layer. Fig. C.15 shows the assimilation performance of a sine-activated PINN with $N_N = 128$ for $w \in \{1, 2, 4, 8, 16\}$ in the style of Fig. 8, for which three training runs with different random initialization were carried out for each $w \neq 4$.

It shows that the sine-activated PINN achieves the best correlation results of the inferred T field and the lowest MAE of the fitted velocity fields for $w = 4$. However, the lowest MAE of the assimilated temperatures was obtained for $w = 1$. Overall, the performance of the PINN varies only slightly for $1 \leq w \leq 4$, while it degrades significantly for higher values. For $w = 16$, even the correlation coefficients of the fitted velocity components start to deteriorate.

A possible explanation for the poor performance of large w is that they introduce unphysically high spatio-temporal wave numbers into the fields represented by the MLP. The high wave numbers are not regularized by the data loss terms, since a fit to data with lower wave numbers also works on the basis of the excessively high wave numbers. This means that the wave numbers of neurons, that were initialized non-physically high, are not being reduced during training. This means

that these neurons will either be suppressed by the physics losses during training, causing them to die, or they will persistently disturb the physics represented by the PINN.

Therefore, we would rather recommend the use of $w = 1$, if a hyperparameter study of w is not feasible, since its critical value will certainly dependent on the investigated flow.

Appendix D. Pressure fields

Velocity assimilation. Fig. D.16 displays the comparison between the ground truth pressure field including its spatial gradients and the corresponding inferred fields. It shows that the assimilated pressures are as good as the inferred velocity fields, since they exhibit the same characteristics:

All structures, such as the depression at $y \approx 0.6, z \approx 0.2$ and the rises near the boundaries, are represented. However, the PINN falls short in reconstructing the amplitudes. This shows that, in this case, the velocities are mainly inferred by the convection–diffusion equation and then the loss of the Navier–Stokes equations acts to reconstruct a pressure (gradient) field according to the inferred velocities.

Temperature assimilation. The comparison of the pressure field and its gradients for the temperature assimilation case is displayed in Fig. D.17. Unlike velocity assimilation, in this case the pressure can be assimilated directly from the fitted velocity fields. This leads to a better agreement of the inferred pressure fields with the ground truth than for velocity assimilation. The largest remaining deviations occur for $\partial_z p$ in the same region as the largest bulk deviation of the inferred temperatures. This shows that the assimilation of both pressure and temperature poses a complication for the PINN approach. This means that supporting one field with additional information, e.g. by implementing pressure taps in an experimental context, should improve the inference performance of both fields.

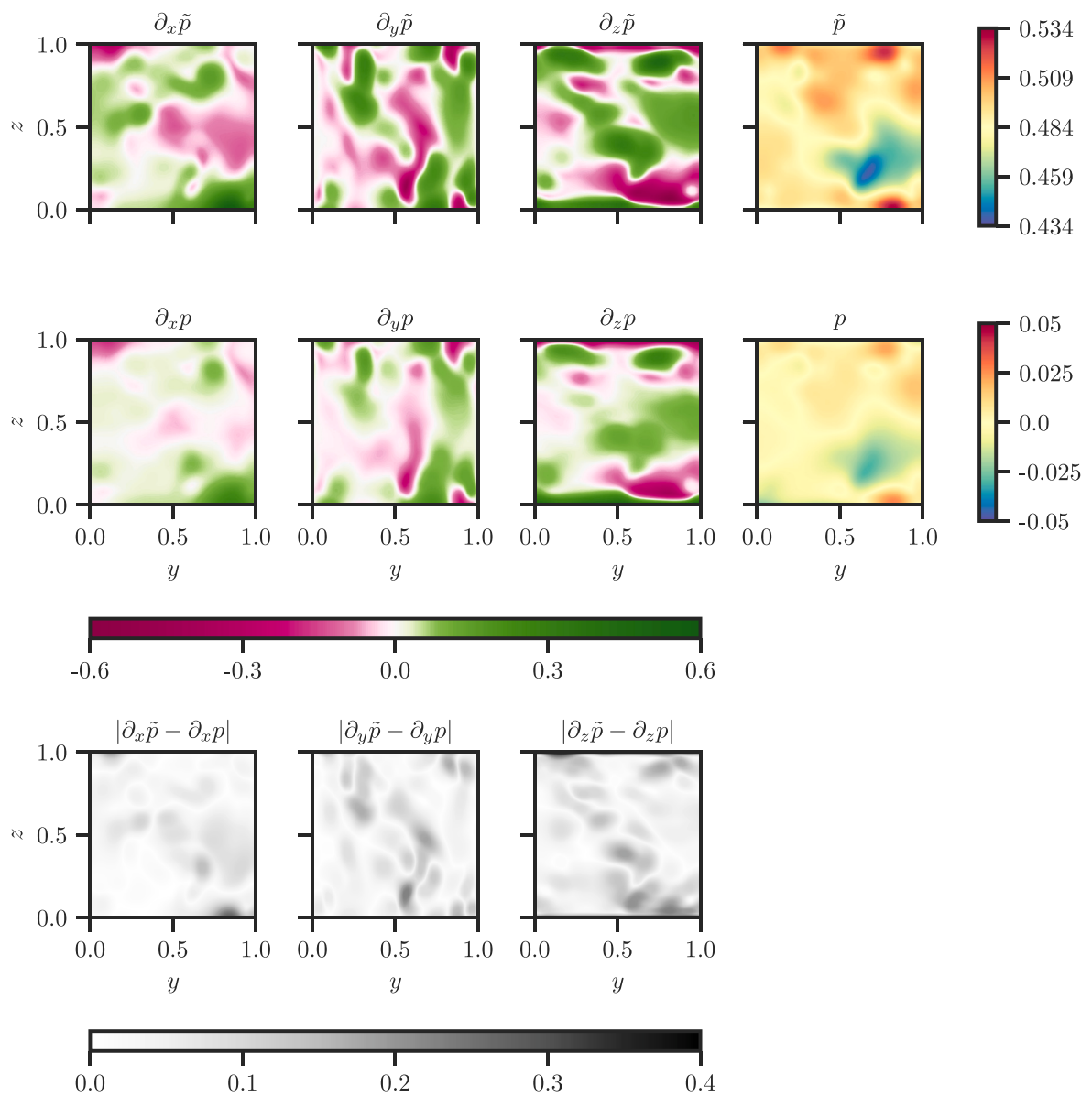


Fig. D.16. Comparison of the ground truth (top row) and PINN-generated (middle row) fields of the pressure and its spatial gradients in a central vertical cross-section for velocity assimilation with sine activation and $N_N = 128$. The bottom row comprises the contour plots of the gradients' absolute differences.

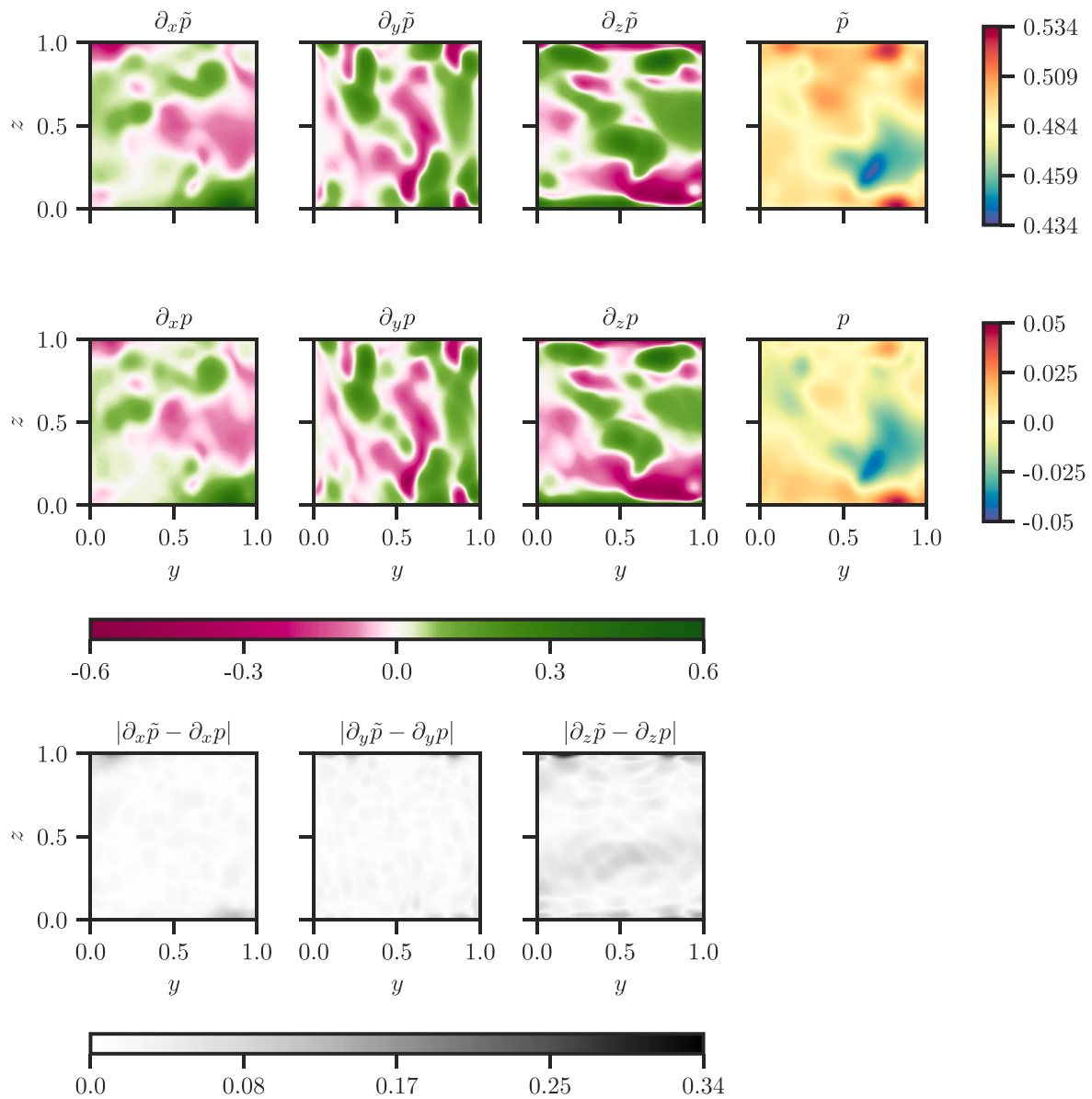


Fig. D.17. Comparison of the ground truth (top row) and PINN-generated (middle row) fields of the pressure and its spatial gradients in a central vertical cross-section for temperature assimilation with sine activation and $N_N = 128$. The bottom row comprises the contour plots of the gradients' absolute differences.

References

- [1] Schmeling Daniel, Volkmannd André. On the experimental investigation of novel low-momentum ventilation concepts for cooling operation in a train compartment. *Build Environ* 2020;182:107116.
- [2] Völker Christoph, Wallace Douglas WR, Wolf-Gladrow Dieter A. On the role of heat fluxes in the uptake of anthropogenic carbon in the North Atlantic. *Glob Biogeochem Cycles* 2002;16(4).
- [3] Mommert Michael, Niehaus Konstantin, Schiepel Daniel, Schmeling Daniel, Wagner Claus. Measurement of the turbulent heat fluxes in mixed convection using combined stereoscopic PIV and PIT. *Exp Fluids* 2023;64(6).
- [4] Käufer T, Cierpka C. Volumetric Lagrangian temperature and velocity measurements with thermochromic liquid crystals. *Meas Sci Technol* 2023;35(3):035301.
- [5] Vinuesa Ricardo, Brunton Steven L, McKeon Beverley J. The transformative potential of machine learning for experiments in fluid mechanics. *Nat Rev Phys* 2023.
- [6] Raissi M, Perdikaris P, Karniadakis GE. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J Comput Phys* 2019;378:686–707.
- [7] Markidis Stefano. The old and the new: Can physics-informed deep-learning replace traditional linear solvers? *Front Big Data* 2021;4.
- [8] Wassing Simon, Langer Stefan, Bekemeyer Philipp. Physics-informed neural networks for parametric compressible Euler equations. *Comput & Fluids* 2024;270:106164.
- [9] Cai Shengze, Mao Zhiping, Wang Zhicheng, Yin Minglang, Karniadakis George Em. Physics-informed neural networks (PINNs) for fluid mechanics: a review. *Acta Mech Sin* 2021;37(12):1727–38.
- [10] Hasanuzzaman Gazi, Eivazi Hamidreza, Merbold Sebastian, Egbers Christoph, Vinuesa Ricardo. Enhancement of PIV measurements via physics-informed neural networks. *Meas Sci Technol* 2023;34(4):044002.
- [11] Di Leoni Patricio Clark, Agarwal Karuna, Zaki Tamer A, Meneveau Charles, Katz Joseph. Reconstructing turbulent velocity and pressure fields from under-resolved noisy particle tracks using physics-informed neural networks. *Exp Fluids* 2023;64(5).
- [12] Eusebi Ryan, Vecchi Gabriel A, Lai Ching-Yao, Tong Mingjing. Realistic tropical cyclone wind and pressure fields can be reconstructed from sparse data using deep learning. *Commun Earth Environ* 2024;5(1).
- [13] Soto Álvaro Moreno, Güemes Alejandro, Discetti Stefano. Complete flow characterization from snapshot PIV, fast probes and physics-informed neural networks. *Comput Methods Appl Mech Engrg* 2024;419:116652.
- [14] Raissi Maziar, Yazdani Alireza, Karniadakis George Em. Hidden fluid mechanics: Learning velocity and pressure fields from flow visualizations. *Science* 2020;367(6481):1026–30.
- [15] Cai Shengze, Wang Zhicheng, Fuest Frederik, Jeon Young Jin, Gray Callum, Karniadakis George Em. Flow over an espresso cup: inferring 3-D velocity and pressure fields from tomographic background oriented Schlieren via physics-informed neural networks. *J Fluid Mech* 2021;915.

- [16] Lucor Didier, Agrawal Atul, Sargent Anne. Simple computational strategies for more effective physics-informed neural networks modeling of turbulent natural convection. *J Comput Phys* 2022;456:111022.
- [17] Di Leoni Patricio Clark, Agasthya Lokahith, Buzzicotti Michele, Biferale Luca. Reconstructing Rayleigh–Bénard flows out of temperature-only measurements using Physics-Informed Neural Networks. *Eur Phys J E* 2023;46(3).
- [18] Cai Shengze, Wang Zhicheng, Wang Sifan, Perdikaris Paris, Karniadakis George Em. Physics-Informed Neural Networks for Heat Transfer Problems. *J Heat Transfer* 2021;143(6).
- [19] Sitzmann Vincent, Martel Julien NP, Bergman Alexander W, Lindell David B, Wetzstein Gordon. Implicit neural representations with periodic activation functions. 2020.
- [20] Tancik Matthew, Srinivasan Pratul P, Mildenhall Ben, Fridovich-Keil Sara, Raghavan Nithin, Singhal Utkarsh, et al. Fourier features let networks learn high frequency functions in low dimensional domains. 2020.
- [21] Angriman Sofia, Cobelli Pablo, Mininni Pablo D, Obligado Martín, Clark Di Leoni Patricio. Assimilation of statistical data into turbulent flows using physics-informed neural networks. *Eur Phys J E* 2023;46(3).
- [22] Wagner Claus, Friedrich Rainer, Narayanan Ranga. Comments on the numerical investigation of Rayleigh and Marangoni convection in a vertical circular cylinder. *Phys Fluids* 1994;6(4):1425–33.
- [23] Chorin Alexandre Joel. A numerical method for solving incompressible viscous flow problems. *J Comput Phys* 1967;2(1):12–26.
- [24] Chorin Alexandre Joel. Numerical solution of the Navier-Stokes equations. *Math Comp* 1968;22(104):745–62.
- [25] Shishkina Olga, Shishkin Andrei, Wagner Claus. Simulation of turbulent thermal convection in complicated domains. *J Comput Appl Math* 2009;226(2):336–44.
- [26] Shishkina Olga, Stevens Richard JAM, Grossmann Siegfried, Lohse Detlef. Boundary layer structure in turbulent thermal convection and its consequences for the required numerical resolution. *New J Phys* 2010;12(7):075022.
- [27] Stevens Richard JAM, van der Poel Erwin P, Grossmann Siegfried, Lohse Detlef. The unifying theory of scaling in thermal convection: the updated prefactors. *J Fluid Mech* 2013;730:295–308.
- [28] Grossmann Siegfried, Lohse Detlef. Scaling in thermal convection: a unifying theory. *J Fluid Mech* 2000;407:27–56.
- [29] Grossmann Siegfried, Lohse Detlef. Thermal convection for large Prandtl numbers. *Phys Rev Lett* 2001;86(15):3316–9.
- [30] Grossmann Siegfried, Lohse Detlef. Prandtl and Rayleigh number dependence of the Reynolds number in turbulent thermal convection. *Phys Rev E* 2002;66(1):016305.
- [31] Grossmann Siegfried, Lohse Detlef. Fluctuations in turbulent Rayleigh–Bénard convection: The role of plumes. *Phys Fluids* 2004;16(12):4462–72.
- [32] Clevert Djork-Arné, Unterthiner Thomas, Hochreiter Sepp. Fast and accurate deep network learning by exponential linear units (ELUs). 2015.
- [33] Sankaran Shyam, Wang Hanwen, Guilhoto Leonardo Ferreira, Perdikaris Paris. On the impact of larger batch size in the training of Physics Informed Neural Networks. In: *The symbiosis of deep learning and differential equations II*. 2022.
- [34] Abadi Martín, Agarwal Ashish, Barham Paul, Brevdo Eugene, Chen Zhifeng, Citro Craig, et al. *Tensorflow: Large-scale machine learning on heterogeneous distributed systems*. 2016.
- [35] Chollet François, et al. *Keras*.
- [36] Kingma Diederik P, Ba Jimmy. Adam: A method for stochastic optimization. 2014.
- [37] Baydin Atilim Gunes, Pearlmutter Barak A, Radul Alexey Andreyevich, Siskind Jeffrey Mark. Automatic differentiation in machine learning: a survey. *J Mach Learn Res* 2018;18(153):1–43.
- [38] Wang Sifan, Teng Yujun, Perdikaris Paris. Understanding and Mitigating Gradient Flow Pathologies in Physics-Informed Neural Networks. *SIAM J Sci Comput* 2021;43(5):A3055–81.
- [39] Wang Sifan, Sankaran Shyam, Wang Hanwen, Perdikaris Paris. *An Expert’s Guide to Training Physics-informed Neural Networks*. 2023.
- [40] Horn Berthold KP, Schunck Brian G. Determining optical flow. *Artificial Intelligence* 1981;17(1–3):185–203.
- [41] Toscano Juan Diego, Käufer Theo, Wang Zhibo, Maxey Martin, Cierpka Christian, Karniadakis George Em. Inferring turbulent velocity and temperature fields and their statistics from Lagrangian velocity measurements using physics-informed Kolmogorov-Arnold networks. 2024.