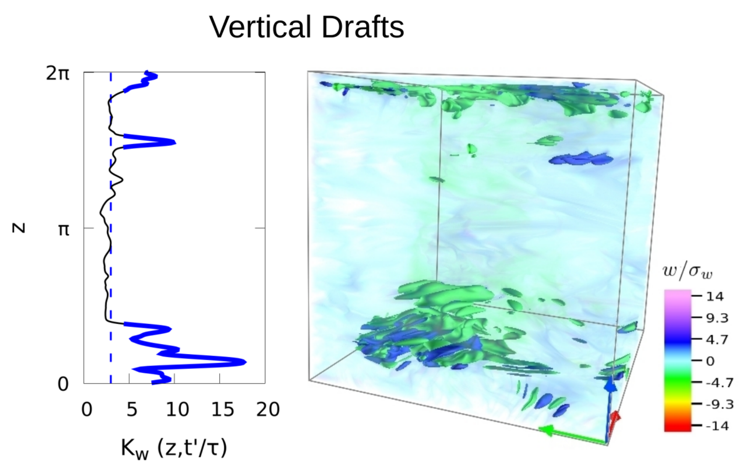
Physics-informed dimensionality reduction of direct numerical simulations of stratified turbulent flows using convolutional autoencoders

Raffaele Marino[1], **Raffaello Foldes**[1,3], Enrico Camporeale[3], Fabio Feraco[1,4] and Giovanni Lapenta[5]

(1) École Centrale de Lyon - CNRS, (2)University of L'Aquila, (3)University of Colorado, CIRES & NOAA SWPC, (4)University of Calabria, (5)Katholieke Universiteit Leuven

1.

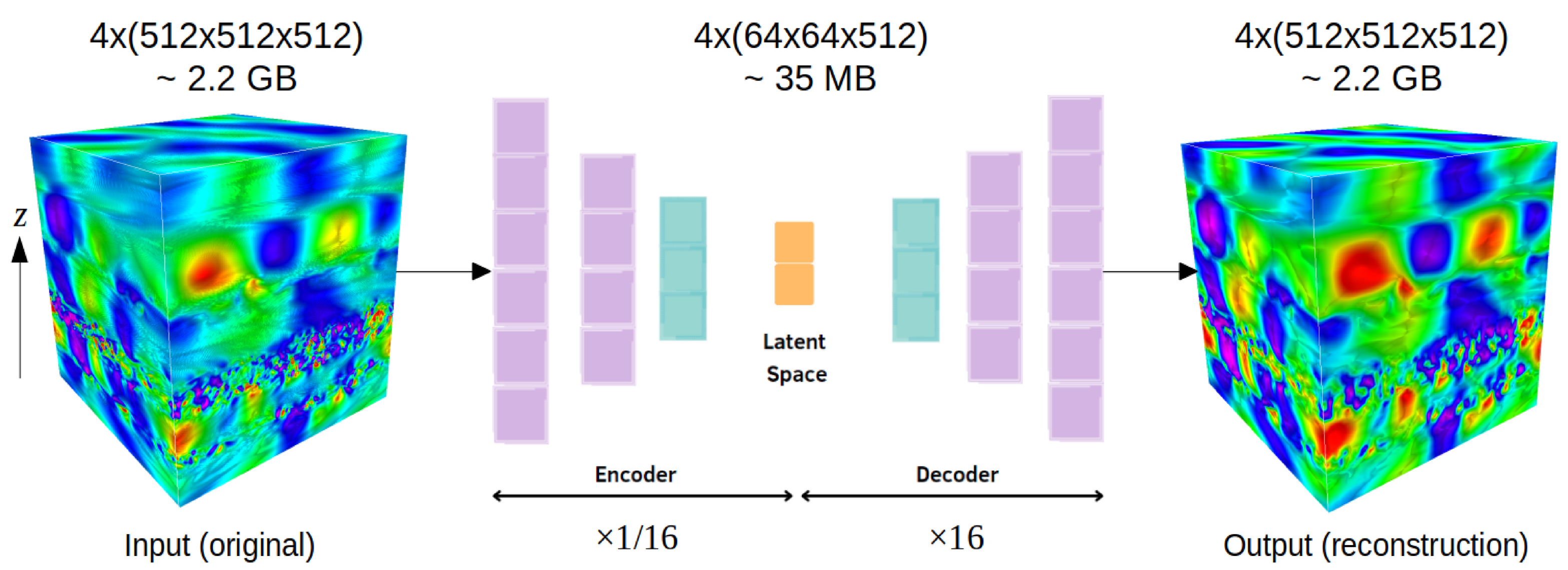
Stratified Turbulence Flows (STF) are particularly interesting for geophysical systems such as the Earth's upper atmosphere and oceans. They are described by the NSE in the *Boussinesq approximation.* STF are characterized by a strong **anisotropy** (gravity) and *extreme events* (**inhomogeneity**) which produce large values of the kurtosis of the vertical velocity K_w. The evolution of the velocity field (V_x, V_y, V_z) and the temperature fluctuations (θ) can be studied by means of high-resolution DNS that typically generate several TB of data to be stored.
Convolutional Autoencoder (CAE) can be useful to reduce the dimensionality of the output of DNS without losing crucial information about the system dynamic.



from Feraco et al., *Europhys. Lett.*, 2018

2.

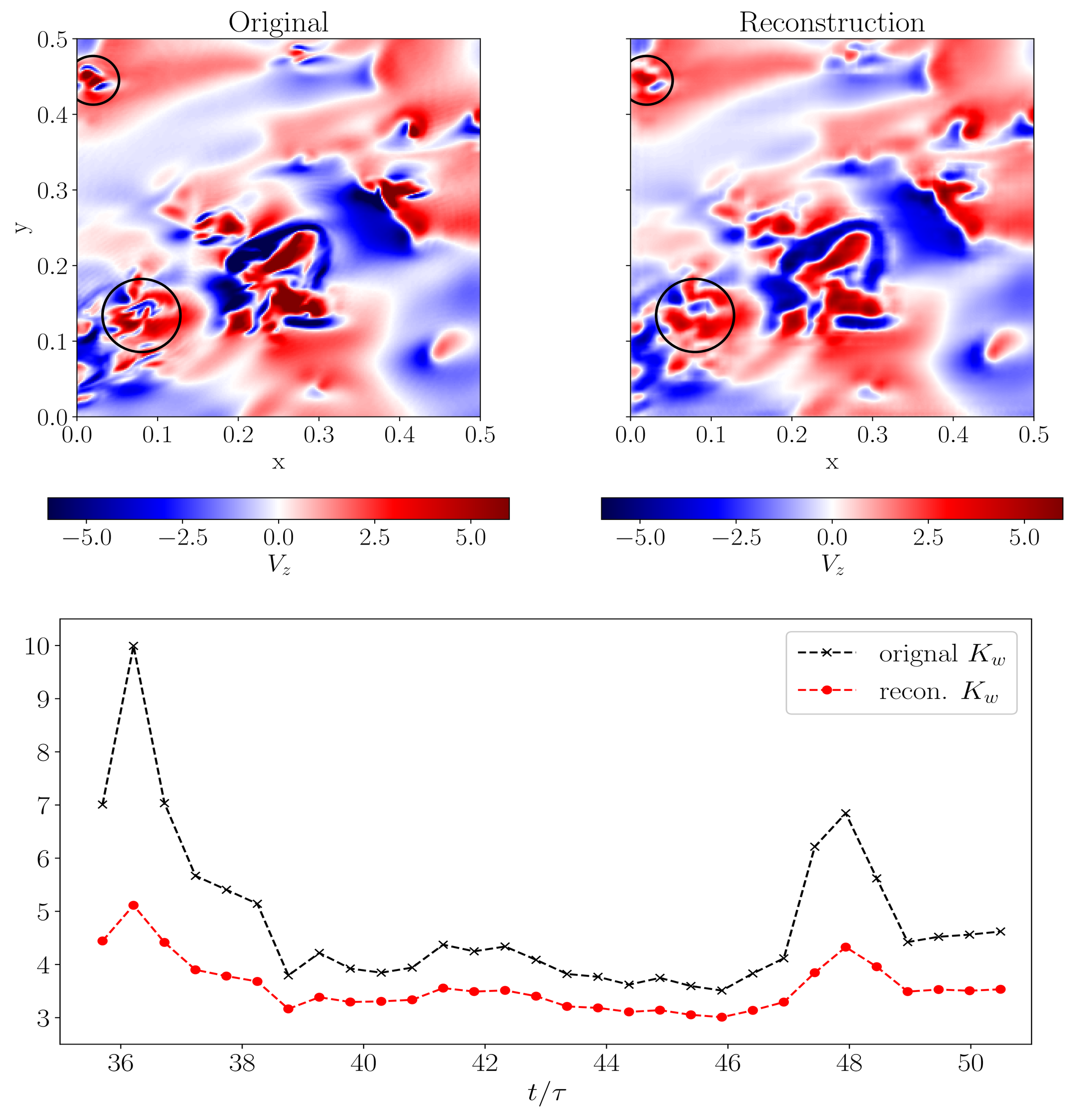
Convolutional Autoencoder (CAE) is a particular example of unsupervised convolutional neural network composed by two main elements: **encoder** and **decoder**. Autoencoders are usually adopted for *dimensionality reduction*, *noise reduction* and *anomaly detection*.
In case of dimensionality reduction, the encoder part reduces the initial dimensionality while the decoder reverses the process creating a reconstruction of the original input



The *anisotropy* of these simulations is addressed by working **plane by plane**, slicing the original 3D cube along the *z* axis. Then, the **input** is formed by a 4x(512x512) cube, where 4 is the number of the physical variables given from the DNS (V_x, V_y, V_z and θ).

3.

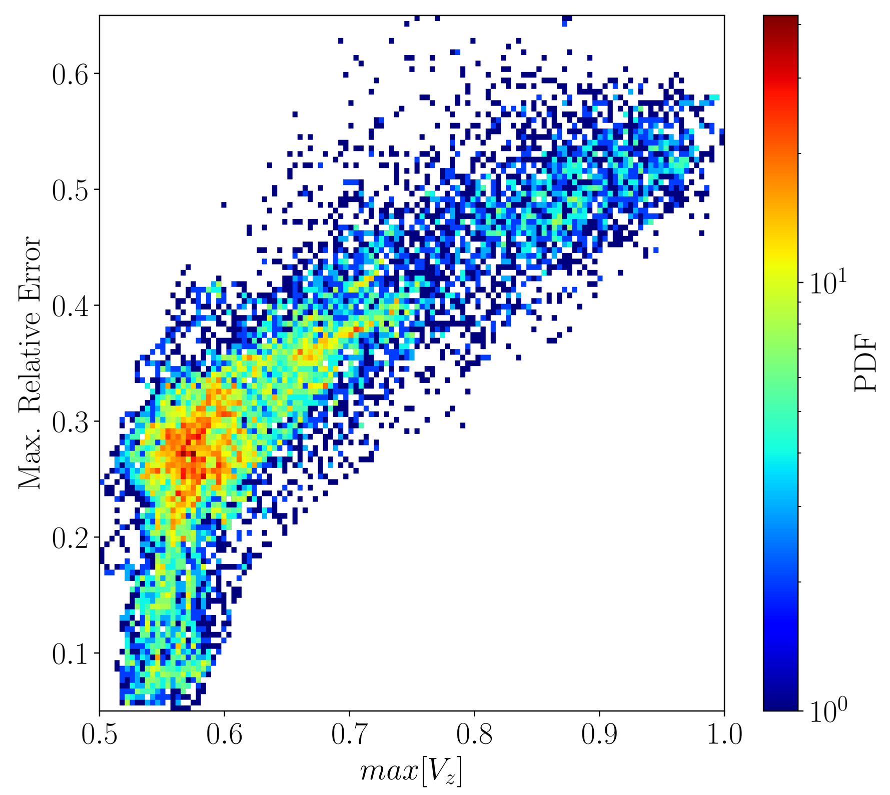
CAE is able to restore the original fields with good accuracy, *but* it has some difficulty with strong extreme events (black circles in figure below). Low order moments (average, standard deviaion and skewness) are well recovered by our model, but the forth-order moment (kurtosis) shows a higher reconstruction error meaning that we are missing information which describes the extreme events charachterizing the simulations. In particular the model does not completely recover the **intensity of extreme events.**



Upper panel: an example of a comparison of the vertical velocity V_z on a 2D plane (*x,y*) presenting several extreme events. The black circles indicate areas where there reconstruction misses extreme events. **Lower panel**: comparison of the original and reconstructed kurtosis K_w of V_z for the entire test set (~ 15000 samples). Simulation time on the horizontal axis.

The figure below shows a 2D histogram of the **maximum relative error** compared to the **maximum of the vertical velocity** $|V_z|$. Both quantities are computed plane by plane.

This figure shows that most of the points are in the region with $max|V_z| < 0.7$, however it is possible to observe several values corresponding to extreme events ($max|V_z| > 0.8$).
The relative error doubles when we consider extreme events, and starting from $max|V_z| > 0.55$ the relation shows a clear linear trend.



4.

We try to obtain an improvement on the reconstruction of extreme events by using **additional information** which can be derived from the physical fields and which have proved to be *statistically correlated* (references below) to the presence of such extreme events.

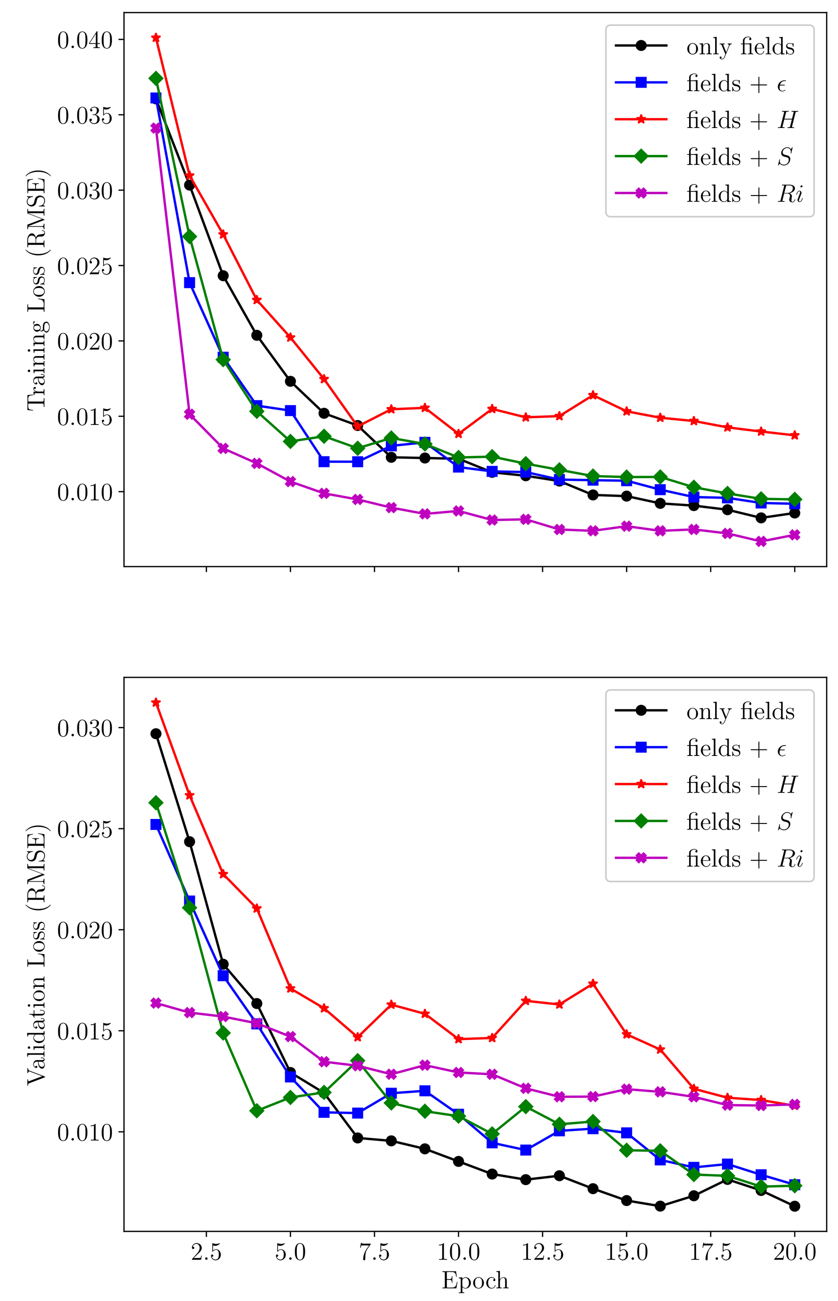
Dissipation: $\epsilon_V = \nu \langle (\nabla u)^2 \rangle$ [Marino et al., *Phys Rev. Fl.*, 2021 (submitted)]

Shear: $S = \langle \partial_z u_\perp \rangle$ [Feraco et al., *Phys. Fluids*, 2022a, (in preparation)]

Helicity: $\mathrm{H} = \overline{u} \cdot \overline{\omega}$ [Feraco et al., *Phys. Fluids*, 2022b, (in preparation)]

Richardson number: $Ri = N\,(N\ \ \partial_z \theta)/(\partial_z u_\perp)^2$ [Feraco et al., *EPL*, 2018]

Train/Validation loss defined as the **root mean squared error** (RMSE) obtained adding one additional information at time to the CAE. The RMSE is always computed over the 4 physical fields (V_x, V_y, V_z and θ), ignoring the reconstruction of the additional field.
Adding an additional field means that we need to slightly increase the number of weights of the network (~ 0.5%), however this little variations does not affect the performance of the various CAEs. This figure shows that the small-scale quantities (*Ri, S* and ε) perform better than the helicity *H* (large-scale). Hovewer, except for this latter, the other curves show more or less the same performance meaning that the additional field is not very helpful to improve the reconstruction of the physical fields.



References

[1] F. Feraco et al., *Vertical drafts and mixing in stratified turbulence: sharp transition with Froude Number*, Europhysic Letters, 2018

[2] R, Marino et al., *Turbulence generation by large-scale extreme vertical drafts and the modulation of local energy dissipation in stably stratified geophysical flows*, Phys. Rev. Lett., 2021 (under review)

[3] F. Feraco et al., *Extreme vertical drafts in sheared stratified geophysical flows*, Phys of Fluids, (in preparation) 2022a

[4] F. Feraco et al., *Local enhancement of the buoyancy flux and helicity depletion in stratified turbulence*, Phys of Fluids, (in preparation) 2022b