

Politecnico di Torino

Politecnico di Torino

Department of Applied Science and Technology (DISAT)

Master of Science's Degree in Physics of Complex Systems

Master's Degree Thesis

Understanding Turbulence via Machine Learning

External supervisor
Chibbaro Sergio

Internal supervisor
Braunstein Alfredo

Candidate
Del Nero Tommaso

2025/2026

Abstract

This work investigates machine learning methods applied to the analysis and modeling of turbulent shell models. Shell models represent a valid framework where data driven methodologies can be applied to turbulent data, given their ability to reproduce many fundamental properties of fully developed turbulence at a limited computational cost.

Three analyses are considered, after a brief summary of the necessary theoretical background. First, sparse regression techniques inspired by the SINDy algorithm are employed to recover governing equations from data. Second, causal relations between shells are investigated through perturbation-response analysis. A strong asymmetry propagation of disturbances is observed, reflecting the directionality of the turbulent energy cascade. Limitations of linear response theory predictions is also presented. Third, a Complemented Deep Reduced-Order Model is adapted for shell models learning in order to reconstruct the effects of unresolved scales.

Overall, the results illustrate how shell models represent a valid benchmark to develop data driven techniques for complex nonlinear dynamical systems.

1 Introduction

"Turbulence is the most important unsolved problem of classical physics." – R. Feynman

Everywhere we go we are surrounded by fluids; exceptions are rare and generally lie outside ordinary human experience.

A complete physical understanding of such fluids would be most useful for engineering and environmental applications. Still in classical fluid mechanics one problem remains open ever since the dawn the field, the problem of turbulent flow.

This work investigates the problem of turbulence through two complementary approaches. The first consists in a well known reduction in the complexity of fully developed turbulence via shell models. The second proposes machine learning techniques to analyze the resulting dynamics and extract physically meaningful information.

This thesis is organized as follows.

Chapter 1 in what follows, provides a conceptual introduction to turbulence and the reasons underlying its complexity.

Chapter 2 presents the phenomenology of Navier-Stokes turbulence, establishing the main characteristics of turbulence and familiarize the reader with its properties.

Chapter 3 introduces shell models, in particular the SABRA model, and their relation to the Navier-Stokes equations. It terminates with a description of the numerical methods and results utilized in the generation of the datasets, which will be used in the remainder of the thesis.

Chapter 4 investigates the learnability of the SABRA shell model from data-driven perspective. The governing equations are reconstructed through a sparse regression framework inspired by the SINDy algorithm. Physical effects and scaling laws are applied in the understanding of learning dynamics.

Chapter 5 investigates causal relationships between shell variables. Causal relationships are quantified through a response function, defined as the averaged response of the system to a stochastic perturbation.

Chapter 6 explores a Complemented Deep Reduced-Order Model (CD-ROM) for shell models. The proposed neural architecture is adapted to learn the influence of unresolved scales on the resolved dynamics, providing a first assessment of data-driven closure strategies for reduced-order modeling of turbulent systems.

The equations describing the behaviour of fluids are well known, they are the Navier-Stokes (NS) equations, derived in 1823:

$$\begin{aligned} \partial_t \boldsymbol{v} + (\boldsymbol{v} \cdot \nabla)\, \boldsymbol{v} &= -\nabla p + \nu \nabla^2 \boldsymbol{v} + \boldsymbol{f} \\ \nabla \cdot v &= 0. \end{aligned}$$

They comprise of a set of nonlinear partial differential equations, where the first equations describe the evolution in time of the velocity field $\boldsymbol{v}$ of the fluid and the last ensures the incompressibility of the fluid. Here p denotes the pressure field, ν is the dynamical viscosity, and $\boldsymbol{f}$ the external forcing. To each term in the equation there is an associated term describing an acting force on the velocity field, the pressure term $-\nabla p$, the viscous term $\nu \nabla^2 \boldsymbol{v}$ and the infamous non-linear convective term$(\boldsymbol{v} \cdot \nabla)\boldsymbol{v}$.

It must be said, before advancing too much, that more generally turbulence is a phenomenon characterized by two main properties: a chaotic behaviour, i.e. a strong sensibility to initial conditions, and a wide array of spatial and temporal

scales. As such with the term turbulence one indicates a set of phenomena that are much more general than the fluid flow. However given that historically the fluid-dynamic turbulence was the first observed and studied, and most importantly that turbulence exhibits some form of universal behaviour, in the following we shall limit our study only to it.

A linear system, in most cases is characterized by a single characteristic length (this can be guessed by the superposition principle, given that different processes at different scales act independently), as such there is a clear minimal unit of space (and as a consequence time) that allows to capture fully the physics. On the other hand when dealing with non-linear equations the processes at different scales get coupled with each other and so to capture the physics at a specific scale one must take into account the system at each scale. It is from this latter property of non-linear systems that chaos and multiscale behaviour emerges, and what makes turbulence such a difficult problem to deal with.

2 Phenomenology of Turbulence

Going into the depths of Navier-Stokes three-dimensional turbulence would be a rather difficult task and outside the scope of this work. This section is aimed at familiarizing the reader with the concept of turbulence and how it emerges, in addition to develop a common language for the main properties of turbulence. In the following the symbol "$\sim$" will be used, meaning *equal within an order unity constant*.

Going back to the Navier-Stokes equations:

$$\partial_t \boldsymbol{v} + (\boldsymbol{v} \cdot \nabla)\, \boldsymbol{v} = -\nabla p + \nu \nabla^2 \boldsymbol{v} + \boldsymbol{f} \tag{1}$$

$$\nabla \cdot v = 0 \tag{2}$$

We can then perform a rough study of the equations by fixing a length scale of interest ℓ. Disregarding the forcing term, since it is arbitrary and under control of the experimenter/simulator, and the pressure term, since the incompressibility of the fluid, expressed via the second equation, makes it removable [6], the evolution of the velocity field is driven by the convective and viscous terms, to which one can associate a characteristic intensity. We can thus obtain the following intensities via dimensional analysis:

$$\begin{aligned} \nu \nabla^2 \boldsymbol{v} &\sim \nu v_\ell / \ell^2 && \text{viscous term,} \\ (\boldsymbol{v} \cdot \nabla)\, \boldsymbol{v} &\sim v_\ell^2 / \ell && \text{convective term,} \end{aligned} \tag{3}$$

where v_ℓ is the characteristic velocity at scale ℓ, we'll come back to this in a moment to give a more rigorous definition.

Performing the ratio of the intensity of the convective term over the viscous one, gives the control parameter of the equations:

$$Re = \frac{\ell v_\ell}{\nu}, \tag{4}$$

which is the well known Reynolds number.

For a large value of Re one expects to find an evolution dominated by the convective term, and it is in such parameter region, dominated by nonlinear evolution, that turbulence develops.

As previously stated turbulence is a chaotic phenomenon of many interacting scales. To perform a complete study of we should take into account the many different scales and do so very precisely, this is untractable in the most general cases. To circumvent this problem we can utilize a probabilistic description of turbulence where instead of considering a particular evolution we are going to consider an ensemble of different turbulent flows and try to extract significant information [1]. In the following the averages performed on the ensemble are going to be indicated with "$\langle \cdot \rangle$".

A quantity that will come most useful in the future is the longitudinal velocity increment:

$$\delta \boldsymbol{v}_\parallel(\boldsymbol{r}, \boldsymbol{\ell}) = [\boldsymbol{v}(\boldsymbol{r}) - \boldsymbol{v}(\boldsymbol{r} + \boldsymbol{\ell})] \cdot \frac{\boldsymbol{\ell}}{|\boldsymbol{\ell}|}. \tag{5}$$

It measures how the velocity field, projected along a specific direction, varies along that same direction, measuring in some way how different points in the fluid are

[1] It might sound strange for a deterministic system to be described statistically, however we can try an justify this given the high sensibility of the system to initial conditions. This way our system becomes unpredictable exponentially fast and gives us a behaviour that is essentially probabilistic.

correlated with one another. Limiting ourselves to the simplest case of homogeneous and isotropic turbulence, there is no need to keep track of the direction of $\boldsymbol{r}$.

In this framework we can finally define precisely the previously cited v_ℓ as:

$$v_\ell = \sqrt{\langle \delta v_\parallel^2(\ell) \rangle}, \tag{6}$$

it is essentially a measure of how the velocity changes on the scale ℓ, i.e. the typical velocity of eddies of that same size.

Up to now we have defined the main terms to be used in this chapter, lastly before deriving some basic results we want to put ourselves in the parametric regime where a fully turbulent evolution is observed and other contributions are negligible, i.e. $Re \gg 1$ or $\nu \ll 1$, this condition is called *Fully Developed Turbulence*. This regime is the climax of analytical intractability and we must be aided by the addition of a further hypothesis: the energy dissipation remains finite in the limit of vanishing viscosity. It can be expressed mathematically as:

$$-\frac{dE}{dt} \xrightarrow[\nu \to 0]{} \varepsilon, \tag{7}$$

where ε is a finite and positive constant and E is the (purely kinetic) average energy and is defined as:

$$E = \frac{1}{2}\langle v^2 \rangle. \tag{8}$$

This assumption is strongly corroborated by experimental data and was used by A. Kolmogorov in his famous 1941 publication (K41) [11]. It consists essentially in assuming that lowering the dynamical viscosity will result in an increase of the gradients in the NS equations, keeping the viscous contribution more or less constant, being the only non conservative term in the equations. Indeed the eddies in a fluid are more convoluted and emerge on smaller scales the less viscous the medium is.

Finally we have all the tools to describe what is called the phenomenology of turbulence. We can, in this picture, consider a particular eddy of size l in our turbulent flow, and we can imagine that this singular eddy will get distorted with time. We can expect that the time necessary to observe a significant distortion would be inversely proportional to the average velocity difference observed at scale l, i.e. v_l, but also proportional to the eddy size l. We can thus estimate the eddy turnover time as:

$$t_\ell \sim \frac{\ell}{v_\ell}. \tag{9}$$

The distortion will bring some points of the velocity field closer and other further away contributing to the creation of larger and smaller eddies, and in turn modify the velocities and energies present at different scales. We can also imagine that such a distortion will affect only the scales closest to it, given that it will be very unlikely for the distorted eddy to evolve either in various very small eddies or organize with eddies of similar size in giving a very large one.

We can consider that on average the kinetic energy stored at a particular scale will follow:

$$E_\ell \sim v_\ell^2. \tag{10}$$

And consequentially it must vary significantly in a time scale of the eddy turnover time, as such the energy will flow through a particular scale as:

$$\Pi_\ell = \frac{dE_\ell}{dt} \sim \frac{v_\ell^2}{t_\ell} \sim \frac{v_\ell^3}{\ell}. \tag{11}$$

Now by the previous hypothesis of fixed energy dissipation, and assuming to be in a scale not affected by the forcing term, the energy flux must be fixed at any scale:

$$\Pi_\ell \sim \frac{v_\ell^3}{\ell} \sim \varepsilon \tag{12}$$

The energy can flow out from a specific scale in two ways, one is the non conservative viscous term, but given the limit of fully developed turbulence and its scaling of l^{-2} (3) it can observed only at extremely small scales, the other is the transfer of energy to another scale. Now given that the only way in which the turbulent process can globally dissipate energy is via the viscous interaction, we must imagine the turbulent process as a continuous cascade of energy from the integral scale where the forcing term dominates, to a scale where only the viscous term gives a significant contribution, called viscous scale.

Before proceeding we can already observe an interesting relation between the observed scale and the relative velocity, solving the similarity with respect to the velocity and since ε is a constant we have:

$$v_\ell \sim (\varepsilon\ell)^{1/3}. \tag{13}$$

This relation can be modified by squaring or cubing it revealing the *two-thirds law* and the *four-fifths law* scalings respectively. These relations can be derived more rigorously and are in their full form:

$$\langle \delta v_\parallel^2(\ell) \rangle = C\varepsilon^{2/3}\ell^{2/3} \qquad \text{two-thirds law,} \tag{14}$$

$$\langle \delta v_\parallel^3(\ell) \rangle = -\frac{4}{5}\varepsilon\ell \qquad \text{four-fifths law,} \tag{15}$$

where the scaling sign was dropped in favor of the equal sign for the correct expression of the two laws.

Finally we can obtain estimates of the extremal scales by confronting the scaling intensity of the different terms. Starting once again from the energy flux

$$\frac{dE}{dt} \sim \nu\frac{v_\ell^2}{\ell^2} \sim \varepsilon, \tag{16}$$

substituting equation (13) into the previous equation and solving for the length, we get the scale at which the viscous term dominates the energy transfer, i.e. what is called *Kolmogorov scale* η (called until now viscous scale):

$$\eta \sim \left(\frac{\nu^3}{\varepsilon}\right)^{\frac{1}{4}}. \tag{17}$$

Finally we can define the Kolmogorov time scale as the time scale of the processes of size η

$$t_\eta \sim \frac{\eta}{v_\eta} \sim \sqrt{\frac{\nu}{\varepsilon}}. \tag{18}$$

Having obtained an estimate of the minimal length scale of the turbulent process, and in turn also the minimal time scale, we can provide an estimate of the computational resources necessary to perform a complete simulation of a turbulent process.

Given a system of size ℓ_0, to keep into account all the relevant scales one must create a mesh of equally distanced points at distance η, as such for a given linear dimension:

$$\frac{\ell_0}{\eta} \sim \frac{\ell_0 \varepsilon^{1/4}}{\nu^{3/4}} \sim Re^{3/4}, \tag{19}$$

where the definition of the Reynolds number (4) and equation(13) were used. Generally for a three dimensional turbulent flow we must have a mesh where the number of points scales as $Re^{9/4}$. Considering also the time scale one can argue that the minimal timestep should be proportional to the mesh spacing, giving globally a scaling of the computational complexity of Re^3, which in the limit of fully developed turbulence is usually not tractable.

Before concluding this chapter a brief excursus on the concept of intermittency should be in order.

What was before derived via scaling relations in (13) is exact only in the case of the four-fifths law; a more precise derivation of the two-thirds law would reveal the need for further approximations, and indeed empirically this law is only a very good approximation.

As such, the scaling relation (13) is a handy tool to estimate some basic relations, but it cannot be applied too strictly. This limitation is an example of what is called intermittency, a property of turbulent flows characterized by strong fluctuations and the non-uniform distribution of dissipation across scales. In particular, velocity increments exhibit increasingly non-Gaussian statistics at small scales, leading to deviations from the simple self-similar Kolmogorov picture.

Instead of the dimensional prediction of the structure functions:

$$S_p(c) = \langle |\delta v(\ell)|^p \rangle \sim \ell^{p/3}, \tag{20}$$

one observes anomalous scaling exponents

$$S_p(\ell) \sim \ell^{\zeta_p} \tag{21}$$

where generally

$$\zeta_p \neq \frac{p}{3}. \tag{22}$$

The departure of the exponents ζ_p from the linear Kolmogorov prediction constitutes the signature of intermittency. Empirically, the exponents form a concave function of p, reflecting the increasing importance of rare and intense fluctuations at smaller scales.

Much more could be said for theoretical results in turbulence, but we'll stop here given that we are now equipped with the main results necessary in the understanding of shell models, which are introduced in the next chapter and will be the focus form now on.

3 Shell models

As it is usual in physics, to tackle a very difficult problem one starts from a simpler one which exhibits, sometimes partially, the same main properties. In the case of turbulence there are many such examples; in this work we'll treat shell models. In the hope that the more relevant statistical properties of turbulent flow would follow some form of universality, and as such independent from the details of the model utilized, shell models are an attempt to be as faithful as possible to the main properties of the NS equations but with the the number of degrees of freedom necessary to keep track of strongly reduced. Shell models have been extensively studied, and there would be much to say in their regard, this chapter focuses just on the introduction of a particular model, the *SABRA* [7] model, and how it can be related with the NS. For a more in depth look we refer to: [2], [3].

Generally a shell model can be understood as a scalar and spatially discrete version of the Navier-Stokes equations. The NS get decomposed in a set of N coupled differential equations, indexed by n, each of them describing the time evolution of the (complex) fluctuation of the velocity field ($u_n(t)$) at the spatial scale corresponding to a wavelength $k_n = k_0\lambda^n$, giving this way a logarithmic spacing between different spatial scales. The *SABRA* model is described by the following equations:

$$\frac{du_n}{dt} = F_{NL}(\mathbf{u}) - \nu k_n^2 u_n + f_n, \tag{23}$$

$$F_{NL} = ik_n \left(A\lambda u_{n+2}^* u_{n+1} + B u_{n+1} u_{n+-1} + C\frac{1}{\lambda} u_{n-1}^* u_{n-2} \right), \tag{24}$$

and the boundary conditions of $u_n = 0$ for $n > N$ or $n < 0$. The first equation contains the familiar forcing and viscous terms, that are exactly the same as in the case of NS equations but subjected to a Fourier transform. The last term F_{NL} is the non linear coupling between the different equations, which reproduces the non linear effects in the NS. In shell models the images invoked earlier of different length scales interacting with one another are rendered explicitly.

Lastly to fully define the model the parameters A, B, C need to be fixed, this can be done by continuing the parallel with the NS. In fluid dynamics in the absence of viscosity and forcing we are left with a system that conserves total energy and helicity, these quantities can be recasted in the shell model formalism (by taking into account dimensional consistency) as:

$$E = \sum_{n=0}^{N} |u_n|^2 \quad , \qquad H = \sum_{n=0}^{N} (-1)^n k_n |u_n|^2 \ , \tag{25}$$

and their conservation can be imposed, fixing the parameters to $A = 1, B = -1/2, C = -1/2$, for $\lambda = 2$.

Having fully introduced the model and how it is related to the NS, we can now discuss some of its main properties. First of all the reduced complexity is evident: the spatial degrees of freedom now scale as $\log(Re)$ instead of $Re^{9/4}$ allowing a simulation of many more spatial scales, while losing, of course, all geometrical information. Secondly, one can observe that the nonlinear term preserves the K41 scaling of (13), this can be proven by substituting $u_n \sim k_n^{-1/3}$ and with some abuse

of notation:

$$F_{NL}(u_n \sim k_n^{-1/3}) \sim k_n \left(A\lambda k_{n+2}^{-1/3} k_{n+1}^{-1/3} + B k_{n+1}^{-1/3} k_{n-1}^{-1/3} + C\frac{1}{\lambda} k_{n-1}^{-1/3} k_{n-2}^{-1/3} \right) \tag{26}$$

$$\sim k_n \left(A\lambda k_n^{-1/3} k_n^{-1/3} \lambda^{-2/3-1/3} + B k_n^{-1/3} k_n^{-1/3} + C\frac{1}{\lambda} k_n^{-1/3} k_n^{-1/3} \lambda^{+2/3+1/3} \right) \tag{27}$$

$$\sim k_n^{1/3}. \tag{28}$$

Considering that F_{NL} acts on the temporal derivative of u_n, we can compensate the scaling by introducing the turnover time ($t_n \sim k_n^{-2/3}$) obtaining finally that the application of the nonlinear term preserves the K41 scaling. Furthermore, the nonlinear term is built to be as simple as possible, composed of only three pairs of two-shell interactions and enforcing maximal locality in between shells. This construction provides a minimal model that preserves the conservation laws introduced above and is consistent with the symmetries of the Navier–Stokes equations (not discussed here). Lastly, another relevant property of shell models is their ability to generate intermittent dynamics, thereby breaking the K41 scaling, this will be illustrated in the next section through numerical results.

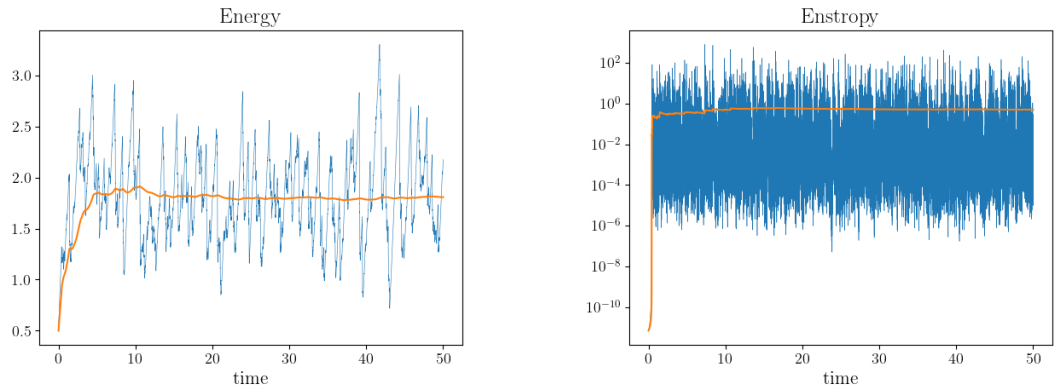
3.1 Numerical Simulation and Statistical Validation

In this section are presented the main statistical properties of shell models, obtained via numerical simulation. In this work two principal models were simulated, a first simpler model on which to apply the more complex studies in the following section, comprised of $N = 12$ shells and a viscosity of $\nu = 2 \cdot 10^{-5}$, and a second more complete model to better expose the statistical results in this section with $N = 32, \nu = 10^{-12}$. Results of the same analyses extended to the latter model are presented in the appendices.

The simulated system was integrated using a fourth order Runge-Kutta scheme. The first three shells were subjected to a random forcing, the intensity of which was proportional to the energy contained in the respective shell. In the same way the last three shells were chosen to lie in the dissipative range, fixing this way the viscosity, imposing a Kolmogorov scale of $k_\eta \sim \nu^{-3/4}\varepsilon^{1/4}$ (and the integral scale $\ell_0 = 1$). This setup allows for a broad inertial range, where turbulent dynamics are explicit. Furthermore to ensure the inclusion of all the relevant dynamics, and numerical stability, the integration time step was chosen to be one hundredth of the Kolmogorov time scale i.e. $dt = \frac{1}{100}\sqrt{\frac{\nu}{\varepsilon}}$, where here and in the following $\varepsilon = 1$.

A first characterization of the chaotic and intermittent nature of the system can be obtained by observing the time evolution of the total energy and enstrophy shown in figure 1. The signals exhibit strong fluctuations and rapid spiking, this behaviour is particular of intermittent dynamics, indicating that it cannot be characterized by a single self-similar scaling law.

From the previous considerations done on the nature of shell models we can expect to observe some of their statistical properties. In this section we try to show numerically that the SABRA model is capable of faithfully reproducing analytical properties of turbulence, considering the strong simplifications applied. Of particular interest is the main assumption of the Kolmogorov theory, i.e. the uniform transfer of energy in the inertial scale, and the result that follows from the previous assumption: the Four-Fifths law scaling. Regarding the former we can define in our model the



(a) (b)

Figure 1: Temporal evolution of (a) total energy $E(t)$ and (b) total enstrophy $\Omega(t)$, illustrating characteristic intermittent bursts. The average calculated in time is displayed via the orange line.

energy flux as [8]:

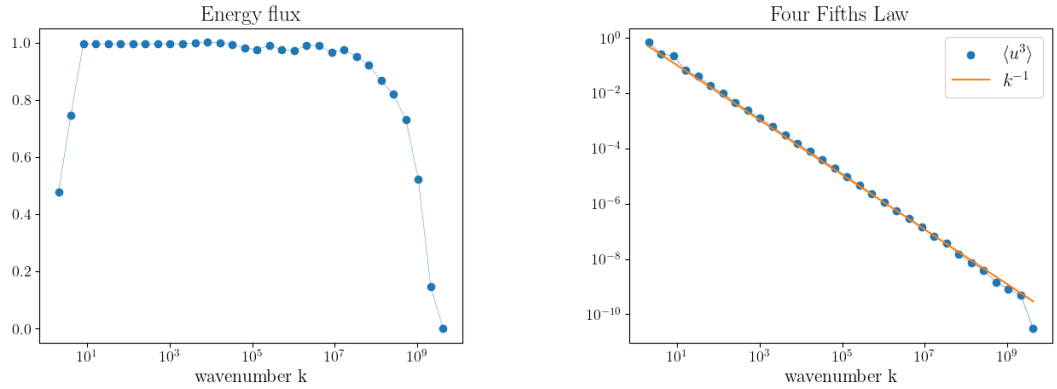
$$\Pi_n = \lambda \Im\{u_n u_{n+1} u^*_{n+2}\} + \frac{1}{\lambda} \Im\{u_{n-1} u_n u^*_{n+1}\}, \tag{29}$$

by performing the time average over it one observes that it is approximately constant throughout the inertial range.

Similarly to tackle the latter we can define the structure functions as:

$$S_p(k_n) = \langle |u_n|^p \rangle, \tag{30}$$

and considering the case $p = 3$. By studying how is scales with k_n, one can assess how closely it reproduces the predictions of the Four-Fifths law (15). Both results are shown in figure 2.

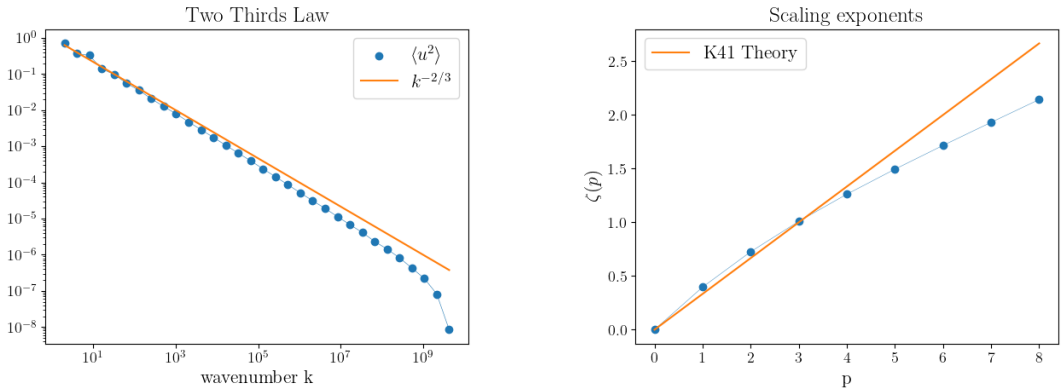


(a) (b)

Figure 2: Predictions from the K41 theory. (a) Constant energy flux across the inertial region. (b) Third-order structure function, demonstrating excellent agreement with the Kolmogorov four-fifths law.

On the other hand we can see how our model exhibits non-trivial properties due to intermittency. In particular, the Two-Thirds law (15) reveals itself as a good approximation of the scaling properties of the second order ($p = 2$) structure

function in the inertial range, but in the viscous range deviations become significant. More generally, figure 3 shows that the scaling behaviour of turbulent fluctuations departs from the K41 prediction.



(a) (b)

Figure 3: Deviations from K41 scaling predictions due to intermittency in the shell model. (a) The second-order structure function deviates from the classical two-thirds law, particularly when approaching the viscous dissipation range. (b) Anomalous scaling of the p-th order structure function exponents, which do not follow the linear K41 prediction.

4 Linear Regression

In many physical systems, one has no knowledge of the governing equations, on the other hand a great amount of data might be available. This is particularly relevant for shell models where the dynamics are themselves a simplification of turbulent flows and therefore cannot be derived directly from first principles.

To tackle this problem we present in this chapter a data-driven approach in which shell models can be learned. More specifically, having generated a dataset form numerical simulations, described in the previous section, we attempt to recover the dynamics equations using a method inspired by the Sparse Identification of Nonlinear Dynamics (SINDy) algorithm introduced by Brunton et al. [4].

SINDy algorithm consists in expressing the complex dynamics as a linear combination of nonlinear functions of the state variables stored in a dictionary $\mathbf{\Phi}(\mathbf{u}(t))$, where $\mathbf{u}(t) \in \mathbb{R}^N$ is the state vector of the system. As such the learning task is reduced to a linear regression problem in which the algorithm must estimate the weight matrix $\mathbf{W}$, assuming that the dictionary is opportunely built. The dynamics can thus be expressed as:

$$\frac{d}{dt}\mathbf{u}(t) - \mathbf{f}(t) = \mathbf{W}\mathbf{\Phi}(\mathbf{u}(t)), \tag{31}$$

where the forcing term $\mathbf{f}$ is known, one can justify this choice given that it is usually chosen by the experimenter, while the time derivative can be estimated from the data, provided that the sampling rate is sufficiently high.

The main obstacle in applying this algorithm lies in the construction of an appropriate dictionary of candidate nonlinear functions. Indeed in building it one must take into account both the expressivity of the functions basis and the limitations imposed by the available hardware means. If, for example, one would limit themselves to the case of polynomial basis functions one would find that the number of terms grows exponentially with the polynomial degree, quickly restricting the number of terms available. To simplify this task we take advantage of two characteristics that we want enforced in shell models. First, to keep the learned model comparable to the SABRA dynamics, we restrict ourselves to a dictionary containing only linear and quadratic terms. Secondly, we are interested in the few dominant contributions to the dynamics leading to a sparse weight matrix $\mathbf{W}$, we can be aided in this by the addition of a $L1$ regularizer, reducing the regression problem in a LASSO formulation. Both assumptions are motivated by the intrinsic simplicity of shell models. Indeed, if the observed dataset cannot be adequately represented within this restricted class of models, this may indicate either that the shell-model approximation itself is insufficient to capture the underlying dynamics or that a better fit would require a level of complexity that compromises the interpretability of the model. In both cases, the resulting description would fall outside the scope of the shell-model simplification.

4.1 Methods

We now present a formal description of the learning procedure followed by the main results. From the numerical simulation previously described we can construct a dataset of the form: $\mathcal{D} = \{\mathbf{x}_m, \mathbf{y}_m\}_{m=0\ldots M}$, where $\mathbf{x}_m$ denotes the state vector: $\mathbf{u}(m \cdot \delta t_{save})$, with δt_{save} denoting the time interval between two successive snapshots.

Similarly the target variable $\mathbf{y}_m$ is defined as:

$$\mathbf{y}_m = \frac{1}{\mathbf{k}} \left(\frac{d\mathbf{u}}{dt}(m \cdot \delta t_{save}) - \mathbf{f}(m \cdot \delta t_{save}) \right), \tag{32}$$

(in the equation above as in the remainder of this work, juxtaposed vectors without an intervening operator indicate component-wise multiplication) the presence at denominator of the wavenumber vector has the role of keeping all the coefficients of the nonlinear term of the same magnitude regardless of the shell considered.

The dictionary is then constructed as follows. At each time instant (m), we define the augmented state vector

$$\mathbf{U_m} = (\mathbf{u}_m, \mathbf{u}_m^*)^{\mathrm{T}}, \tag{33}$$

combining both the state variables and their complex conjugates.

The feature vector is then obtained by collecting both linear and quadratic observables,

$$\mathbf{\Phi_m} = (\mathrm{vec}(\mathbf{U}_m \otimes \mathbf{U}_m), \mathbf{u}_m)^{\mathrm{T}}, \tag{34}$$

where vec$(\cdot)$ denotes column-major vectorization of the outer product. Obtaining this way a feature vector of $(2N)^2 + N$ elements, of which only $\frac{(2N)(2N-1)}{2} + N$ are independent, in practice, this redundancy does not affect the regression procedure.

The dynamics are then identified by solving the regression problem

$$\mathbf{y}_m \simeq \mathbf{W}\mathbf{\Phi}(\mathbf{U}_m), \tag{35}$$

where $\mathbf{W}$ is the coefficient matrix.

The optimization problem can be thus fully specified by the choice of a loss function, in this case consisting of a mean squared error term and an L_1 regularization term:

$$\mathcal{L} = \frac{1}{M} \sum_{m=0}^{M} ||\mathbf{y}_m - \mathbf{W}\mathbf{\Phi}_m||_2^2 + \gamma \, ||\mathbf{W}||_1 . \tag{36}$$

The minimization of $\mathcal{L}$ was performed using stochastic gradient descent with the Adam optimizer.

4.2 Results

In this section are presented the results obtained through linear regression.

It is worth noting that the problem presented is relatively simple; it is a noiseless problem, and the equation of motion are fully known, as such the loss function is fully convex. Nevertheless, the Adam optimizer was deliberately selected to evaluate its potential for broader generalization. Since Adam introduces stochasticity in the learning procedure, small fluctuations are observed in this noiseless system. The role of the L_1 regularizes is to suppress these fluctuations (be it that they come form the optimizer as in this case or the use of a real dataset) and promote sparsity .

The regularization, however, introduces a trade-off, terms that have a true physical contribution to the dynamics, but that might not affect significantly the loss function, will get suppressed.

This trade-off becomes apparent when too large values of the γ term are chosen, in such cases the coefficients in the largest wavenumber shells are progressively driven to zero. To understand the origin of this effect, it is useful to analyze the scaling

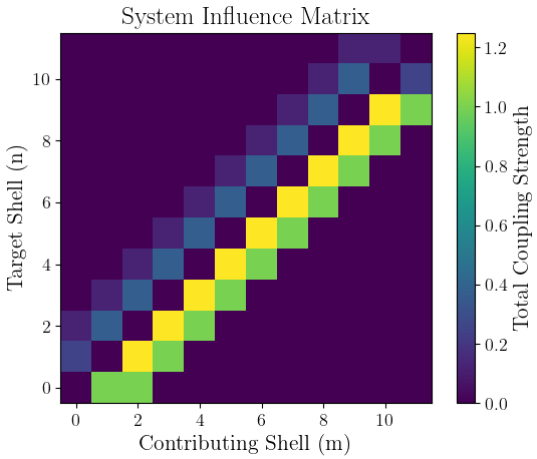


Figure 4: Magnitude of the true interaction matrix coefficients. For a given row (shell n), the matrix displays the intensity of the interaction coming from a column (shell m) within the dynamic equations. Notably, the strongest coupling occurs at shell $m = n + 1$, whereas the interaction with shell $m = n - 2$ is the weakest.

properties of the loss function by variation of the shell number. To do so we neglect the regularization term and rewrite the loss function explicitating the dependence on the error of the coefficients:

$$\mathcal{L} = \frac{1}{M} \sum_{m=1}^{M} \|\delta \mathbf{W}\, \mathbf{\Phi}_m\|_2^2 \,, \tag{37}$$

where

$$\delta \mathbf{W} = \mathbf{W} - \tilde{\mathbf{W}} \tag{38}$$

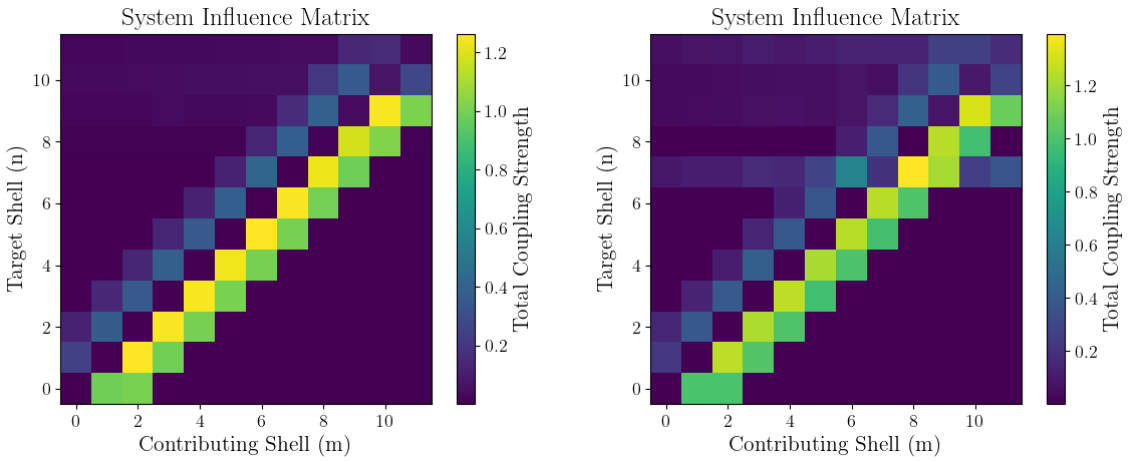
is the error between the learned ($\mathbf{W}$) and the true ($\tilde{\mathbf{W}}$) coefficient matrices.

To estimate the contribution of each shell, we approximate the average over the dataset using the K41 prediction for the shell amplitudes, neglecting shell-to-shell correlations and intermittency. Although this approximation is only qualitative, as in this case it neglects not only the intermittent properties of turbulence, but also the eventual correlations between different shell velocities, it is sufficient to determine qualitatively the scaling properties of the loss. Under these assumptions, the contribution associated with a single coefficient δW_{nm} scales as

$$\mathcal{L}_{nm} \sim (\delta W_{nm})^2 (k_n k_m)^{-2/3}. \tag{39}$$

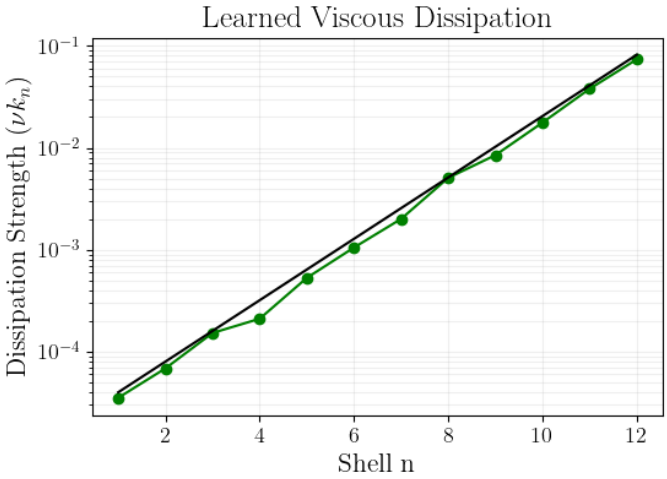
This result has an important consequence. Since the loss function gets a larger contribution by errors for the lowest (wavenumbers) shells we can expect that the learning is focused on those, neglecting the higher shells. As such adding back the regularizer one expects it to fully suppress these less relevant coefficients.

This scaling analysis naturally suggests modifying the loss function by assigning larger weights to the contributions of the high-wavenumber shells. A complete compensation of the theoretical scaling was found to overcorrect the problem, indicating that the actual regression errors do not follow the simple dimensional estimate exactly. The best performance was instead obtained by weighting the contribution of shell n proportionally to $k_n^{1/6}$. The resulting improvement is shown in figures 5 and 6.

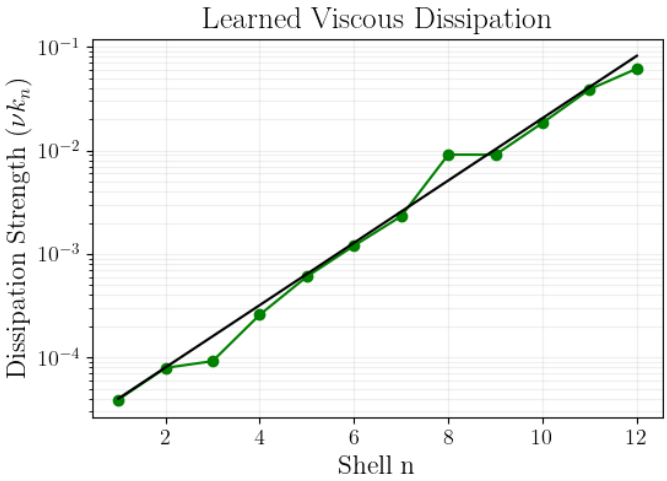


(a) (b)

Figure 5: Comparison between the interaction matrices obtained with (a) and without (b) scaling correction. The inclusion of scaling amplifies the error induced by noisy terms at the largest wavenumbers, which subsequently enables a more effective correction by the regularizer.



(a)



(b)

Figure 6: Comparison between the viscous coefficients obtained with (a) and without (b) the scaling correction, with true values indicated in black. Although the viscous terms are affected differently by the correction, the scaling significantly improves the estimation quality of the nonlinear coefficients. Consequently, this reduces the artificial compensation required by the viscous terms, yielding better results globally.

5 Causality

Other that recovering the governing equations, an important matter is the understanding and interpretation of the equations. In this section we perform a causal analysis following the framework of [1]. The concept of causality is difficult to define in the chaotic systems context, where the properties of the model are usually defined through statistical correlations. However statistical correlations do not express the directional relationship fundamental to causality; although two variables may be strongly correlated, such a correlation alone does not imply the existence of a causal interaction. The defining feature of causality is the temporal ordering between the perturbation and the response. Throughout this work we therefore identify causal relations with the *interventional* definition introduced in [5], i.e. the response of one shell following a controlled perturbation applied to another shell.

The Response of shell j to a perturbation in shell i can be quantified through the definition of a response function:

$$R_{i\to j}(t) = \left\langle \frac{\delta x_j(t)}{\delta x_i(0)} \right\rangle, \tag{40}$$

where the function $\delta x_j(t)$ is in turn defied as the difference between the perturbed trajectory $\tilde{\mathbf{x}}(t)$ and the unperturbed (parent) one $\mathbf{x}(t)$.

$$\delta \mathbf{x}(t) = \tilde{\mathbf{x}}(t) - \mathbf{x}(t). \tag{41}$$

5.1 Methods

Because the SABRA shell model is chaotic, the study of the response to a single perturbation would not be not representative of the underlying dynamics. As infinitesimal changes in the initial conditions produce exponentially diverging trajectories, in the same way the measured response is strongly dependent on the particular realization of the perturbation. To recover statistically meaningful response functions, an ensemble average is therefore required.

An ensemble of $S = 1000$ trajectories was generated by perturbing a single shell of the parent trajectory. The perturbation consisted of independent Gaussian random variables applied to the real and imaginary components of the shell velocity, with zero mean and standard deviation

$$\sigma = |u_n| \cdot 10^{-6}. \tag{42}$$

Unless otherwise stated, the perturbation was applied to the sixth shell of the 12-shell model.

figure 7 illustrates the evolution in time of the ensemble of trajectories and their separation from the parent solution. As expected for a chaotic dynamical system, the perturbation grows approximately exponentially, with a growth rate determined by the largest Lyapunov exponent.

To understand the importance of nonlinear effects, the measured response functions were then compared with the prediction of linear response theory [9]. Under the assumption of linear dynamics, the response function can be expressed as

$$R_{i\to j}(t) = \left(C_t C_0^{-1}\right)_{ji}. \tag{43}$$

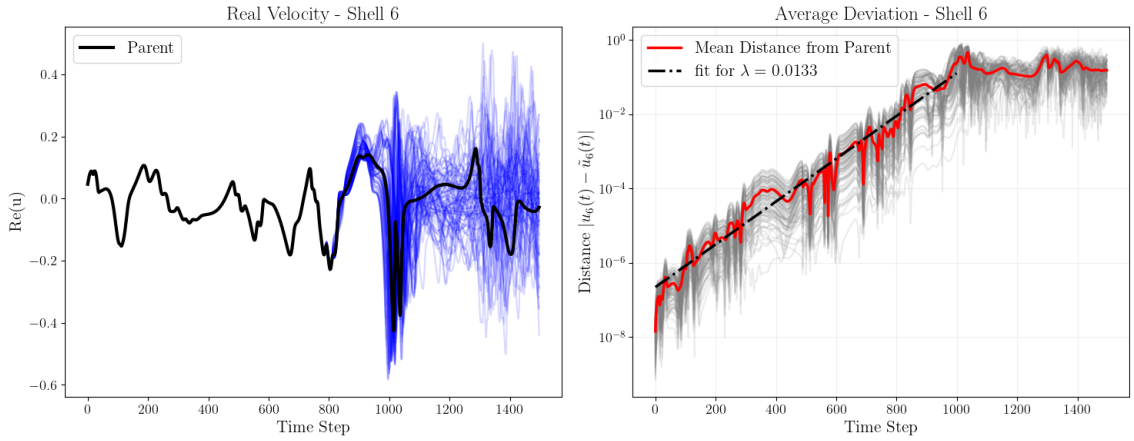
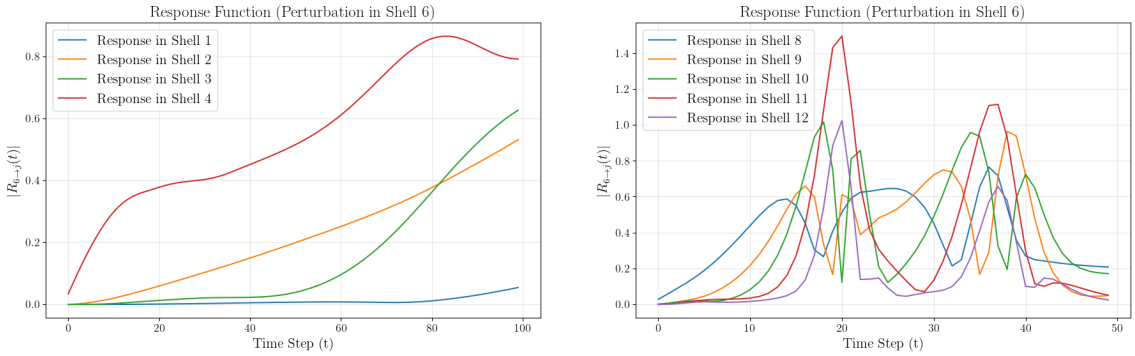


Figure 7: (Left) Real part of the trajectories sampled from the ensemble of the sixth shell, with the parent run highlighted. (Right) Corresponding deviations from the parent run. As is characteristic of chaotic dynamical systems, these deviations are well-fitted by an exponential growth function scaled by the maximum Lyapunov exponent λ.



(a) (b)

Figure 8: Response function magnitudes for (a) upstream and (b) downstream shells. Both cases exhibit clear ordering: closer shells respond earlier and with higher amplitudes. However, the response magnitude is highly asymmetric between the two regimes. The upstream time window is intentionally extended beyond the physical tracking domain to better visualize the growth of the weak signal.

5.2 Results

The first feature of the response functions that can be directly observed is the clear temporal ordering of the perturbation propagation. Indeed shells closer to the perturbed shell respond earlier with larger amplitudes, whereas distant shells exhibit a delayed response. We hypothesize that this behavior can be attributed to the locality of the interactions inherent to the shell model structure.

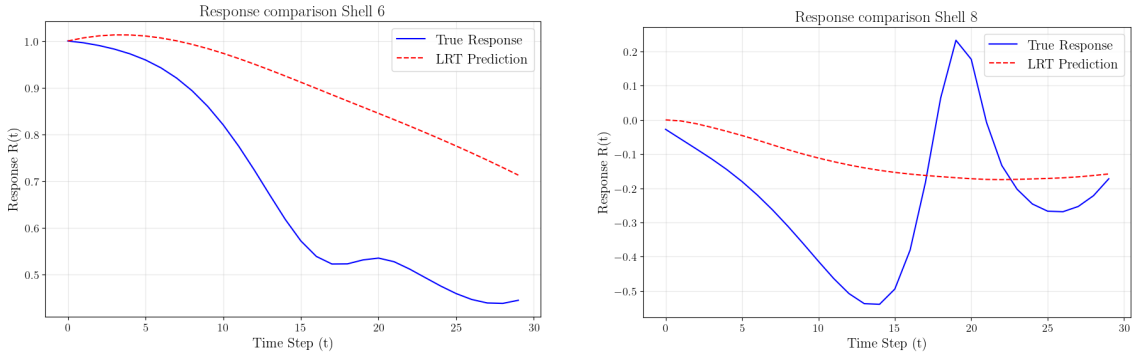
Moreover, a pronounced asymmetry is observed between upstream (high-to-low shell numbers) and downstream (low-to-high shell numbers) propagation; where upstream and downstream are defined relative to the direction of the mean energy flux. Perturbations propagate far more efficiently downstream, as it can be intuitively understood since eddies are more likely to breakdown than to form up. A simple study of the response half times, i.e. the time necessary for the response to reach a value of 1/2 for each shell, reveals that the downstream propagation time increases approximately linearly with shell number, indicating that the perturbation

requires nearly the same time to travel between successive neighboring shells.

An additional comment is necessary regarding the short time scale in 8, indeed the response functions can only be reliably estimated over a finite time interval. As the perturbed trajectories diverge exponentially from the parent solution, the variance of the estimator grows very rapidly in time. Thus, maintaining the same statistical accuracy would require a prohibitive number of realizations over the whole simulated time. More precisely following [1] it can be shown that the statistical error scales as

$$\text{err} \sim \frac{\exp(\lambda t)}{\sqrt{S}}, \tag{44}$$

where $\lambda = 0.0133$ denotes the maximum Lyapunov exponent.



(a) (b)

Figure 9: Comparison between true response function and linear response theory prediction in shell 6 (a) and shell 8 (b) to differentiate the particular case of the effect on the same shell and the effect on a different one.

Lastly, comparing the linear response theory prediction with the true response function in figure 9 one can observe that the the linear response is capable of reproducing the immediate properties of the full response, but it quickly deviates form the prediction. Highlighting once again that the system is strongly non linear and a more precise application of linear response theory would require the addition of some non trivial effective variables.

6 Complemented Deep - Reduced Order Model

The SABRA shell model, analyzed in different ways throughout this work, has been utilized as a reduced description of the Navier-Stokes equations, keeping only the essential equations describing the turbulent energy cascade and strongly reducing the number of degrees of freedom. A natural question consequently is how the model is affected by a further dimensional reduction. This question comes form the fact that many nonlinear systems are shown to exhibit an underlying simpler mechanism, limiting their evolution on a lower-dimensional manifold. Exploiting this property one could be able to reconstruct a simpler model.

There could be several motivations to study a particular ROM. A reduced model provides better interpretability of the underlying dynamics, restricting the analysis to a smaller region of phase space, and also dimensional reduction is useful when considering cases in which part of the system is unobservable or difficult to measure.

Furthermore the study of reduced-order modeling of turbulent flows is also closely related to engineering and geophysical applications. In practical simulations, such as Large Eddy Simulations (LES), computational and experimental limitations make it impossible to resolve all turbulent scales. Instead, the smallest scales are filtered out and their effect on the resolved dynamics must be modeled through suitable closure terms. The present work addresses the same fundamental problem within the simplified framework of shell models, providing a controlled environment for studying data-driven closure strategies.

In this chapter we present a reduction that is data driven, unlike the reduction form the NS equations to the SABRA model, which was guided by physical considerations and conservation laws. The applied reduction consists in retaining only the largest scales of the shell model. This is obtained by truncating the state vector after a particular shell number cutoff $N_{\text{cutoff}} < N$, discarding all higher-wavenumber shells.

Physically this approximation corresponds to neglecting the smallest space fluctuations. Since these high shells contain a small fraction of the total energy, their removal can also be justified by considering them the least relevant degrees of freedom from an energetic point of view. Moreover, when simulating shell models, this truncation is already implicit by imposing boundary conditions. Indeed a *complete* simulation would necessitate an infinite number of shells, but one can reproduce faithfully the statistical properties of turbulence with a few shells in the dissipative regime, given that they become exponentially less relevant.

A difficulty arises in the mathematical definition of the problem because the remaining shells equations are not dynamically closed. The discarded variables still appear through the nonlinear interactions, the simplest solution consists of replacing all discarded shell variables by zero, $u_n = 0, \quad n > N_{\text{cutoff}}$, which corresponds to the standard Galerkin projection. Although straightforward, this approximation fully ignores the influence of the discarded scales on the resolved dynamics. For a chaotic system a small modeling error grows rapidly in time, leading to trajectories that no longer follow the true dynamics.

It is worth emphasizing, however, that reproducing individual trajectories is not the ultimate objective. Instead, the relevant question is whether the reduced model reproduces the same invariant measure and therefore the same statistical observables as the original system.

The answer depends strongly on the chosen cutoff. As previously noted, eliminating partially the highly dissipative should not profoundly change significantly

the true properties of the model. However, removing the whole dissipation range would reduce the system to a shell model of an inviscid fluid.

6.1 Projection and Closure Problem

To formalize the previous discussion, let $\mathbf{z}$ be the state vector of the ROM. We can relate the different variables via:

$$\mathbf{z}(t) = \mathbf{V}\mathbf{u}(t), \tag{45}$$

where $\mathbf{V}$ is the projection operator. From which the associated projection operators can be defined

$$\mathbf{P} = \mathbf{V}^T\mathbf{V}, \quad \mathbf{Q} = \mathbb{1} - \mathbf{P}, \tag{46}$$

allowing to decompose the original state vector in the two components

$$\mathbf{u} = \mathbf{P}\mathbf{u} + \mathbf{Q}\mathbf{u}, \tag{47}$$

resolved and unresolved respectively. Allowing to write the dynamical equation of the reduced variables:

$$\frac{d\mathbf{z}}{dt} = \mathbf{V}g(\mathbf{u}) \tag{48}$$

where with $g(\cdot)$ we indicate the function describing the dynamics.

Equation (48) still depends on the complete state vector $\mathbf{u}$, consequently it does not describe a ROM, but only a projection of the dynamical equations. To build the Galerkin ROM an approximation is applied:

$$\frac{d\mathbf{z}}{dt} \simeq \mathbf{V}g(\mathbf{V}^{\mathrm{T}}\mathbf{z}). \tag{49}$$

It is thus necessary to understand the accuracy of the applied approximation, to do so it is useful to rewrite the dynamical equations (48) separating the contributions from the solved and unsolved terms:

$$\frac{d\mathbf{z}}{dt} = \mathbf{V}g(\mathbf{P}\mathbf{u} + \mathbf{Q}\mathbf{u}), \tag{50}$$

in this way, assuming $||\mathbf{Q}\mathbf{u}||_F \ll |\mathbf{P}\mathbf{u}||_F$, the dynamics can be expanded to first order around $\mathbf{P}\mathbf{u}(t)$:

$$\frac{d\mathbf{z}}{dt} \simeq \mathbf{V}g(\mathbf{P}\mathbf{u}) + \mathbf{V}J_{\mathbf{P}\mathbf{u}}\mathbf{Q}\mathbf{u}, \tag{51}$$

where $J_{\mathbf{P}\mathbf{u}}$ denotes the Jacobian of equations.

The first contribution depends only on the resolved variables and corresponds to the Galerkin dynamics. The second term accounts for the influence of the discarded degrees of freedom and will be referred to as the residual. For dissipative systems this approximation may still reproduce the dominant large-scale dynamics. However, for chaotic dynamical systems, the intricate coupling between different state variables generally causes these standard estimation techniques to fail and the reduced model may even converge to an entirely different invariant measure.

The objective of the present work is therefore to construct a model for the residual term rather than simply neglecting it.

6.2 Learning the Closure Term

Motivated by the Mori-Zwanzig formalism, the influence of the unresolved variables can be modeled through a memory term. Rather than storing the complete history of the trajectory, we introduce a continuous memory variable defined through an exponentially weighted average of past states,

$$\mathbf{y}_i(t) = \int_{-\inf}^{t} \exp(\lambda_\mathbf{i}(s-t))\mathbf{z}_i(s)\, ds, \tag{52}$$

where each exponent is chosen to be positive and large enough to ensure the convergence of the integral.

This particular choice of the temporal kernel was done because of its simplicity, indeed it is also possible to embed the memory variables in a Markovian representation obtained by introducing the auxiliary evolution equation:

$$\frac{d\mathbf{y}}{dt} = \mathbf{z}(t) - \lambda\mathbf{y}(t). \tag{53}$$

allowing the memory to be propagated without explicitly storing the entire trajectory. The resulting complemented reduced-order model takes the form

$$\frac{d\mathbf{z}}{dt} = g(\mathbf{z}(t)) + \mathcal{R}(\mathbf{y}(t)) \tag{54}$$

$$\frac{d\mathbf{y}}{dt} = L(\mathbf{z}(t)) - \lambda\mathbf{y}(t), \tag{55}$$

The first term represents the known projected dynamics, while the second acts as a learned closure correcting the effect of the unresolved variables. The lifting operator $L(\cdot)$ maps the reduced variables into a higher-dimensional latent space, increasing the expressivity of the memory representation. Both the closure operator $\tilde{\mathcal{R}}$, the lifting operator L, and the decay rates λ are learned during training.

6.3 Training Procedure and Numerical Results

The proposed Complemented Deep Reduced-Order Model follows the architecture introduced in [10], shown in figure 10.

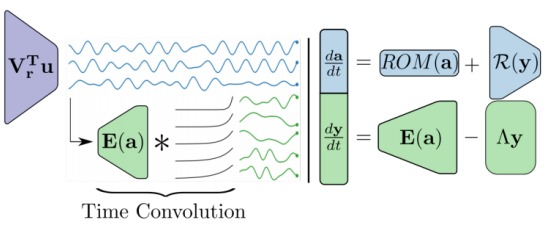


Figure 10: Scheme of the architecture utilized for the CD-ROM, figure reshapen and taken form [10] with permission?

The reduced dynamics are composed of two learned components: a residual function, responsible for reconstructing the influence of the discarded variables, and a lifting operator that embeds the reduced variables into a higher-dimensional latent space. Both functions are implemented as fully connected neural networks, with

SiLU (Sigmoid Linear Unit) activation function. Following the original implementation, the lifting operator receives as input both the reduced variables and the learned latent representation.

This global formulation allows the model to compensate for non-local interactions and structural anomalies characterized by significantly slower dynamical time scales.

In this context the residual should act only on the dynamical equations of shells $n = N_{\text{cutoff}} - 1$ and $n = N_{\text{cutoff}} - 1$, given that those are the only non-closed equations in the truncated system. However we extend the learned residual to act over the equations of every resolved shell. This way the model should be facilitated to compensate for systematic errors that might arise over slower dynamical time scales.

The training loss is defined as the mean squared error between the predicted and reference trajectories, with an additional term measuring also the difference between the predicted derivative and the one obtained via finite differences, this latter weighted at one-tenth of the former. To avoid divergence and metastable state in the loss function, a curriculum-learning technique is employed where the reference trajectory gets progressively longer in time. At each iteration, the loss function is evaluated on a validation trajectory. If the validation loss saturates for a chosen number of iterations, the training curriculum advances of another time step.

To simplify the learning problem, the external forcing is kept constant and applied only to the third shell. Finally, to enforce the same boundary conditions adopted by the Galerkin ROM, two zero-valued shells are added at both ends of the reduced system.

In this particular case the cutoff was fixed at $N_{\text{cutoff}} = 8$ removing fully the dissipative scale. It should also be emphasized that the present test case represents a relatively favorable setting for the proposed methodology. The truncation removes only four shells, all belonging to the dissipative range, where the unresolved dynamics are expected to be comparatively simple. Consequently, the residual to be learned remains relatively limited.

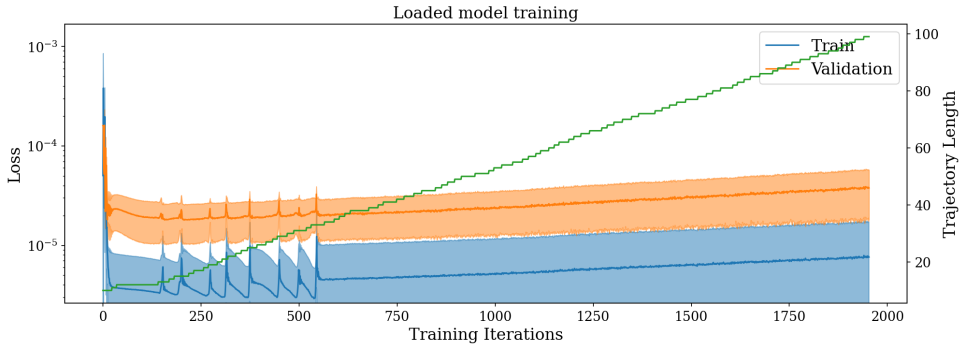


Figure 11: Loss function of the NN in time the green line represents the amount of time steps to learn in the loss. The width is given by the variance of the loss in time

figure 11 illustrates the evolution of the training and validation loss throughout the procedure. One can observe the presence of two regimes, separated in time. The first phase is characterized by an alternation of sudden spikes, given by the incremental step-ups in the curriculum-learning, followed by gradual decays, which indicate that the network is successfully adapting and learning the dynamics. In the second phase, however, that the average loss is observed to increases approximately linearly with the prediction horizon. This behavior indicates that the prediction error accumulates with time, suggesting that the learned model is capable of reproducing only the short term dynamics but progressively loses accuracy over longer

simulations.

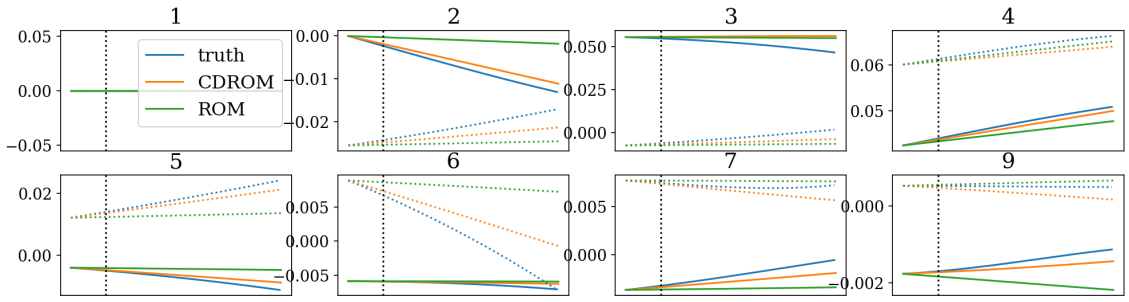


Figure 12: Comparison of trajectories between the true system (CD-ROM) and the reduced-order model (ROM). Solid lines denote the real components, while dotted lines represent the imaginary components of the fields.

The reconstructed trajectories are compared with both the full shell model and the Galerkin ROM in figure 12. The complemented ROM consistently reproduces the reference trajectory more closely than the standard Galerkin approximation. Although neither reduced model is capable of faithfully reproducing the complete trajectory over long times, the CD-ROM consistently remains closer to the reference dynamics than the Galerkin ROM.

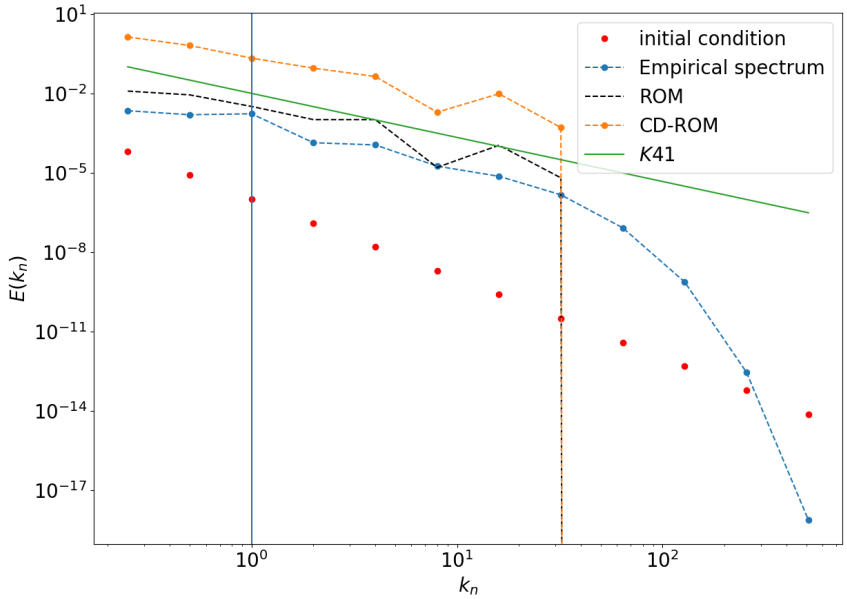


Figure 13: Second-order structure functions computed for the true model, the ROM, and the CD-ROM. The blue vertical line denotes the specific shell where the external forcing is applied. Statistical averages were obtained by initializing the system from a state strictly adhering to Kolmogorov 1941 (K41) scaling, to approximate a statistically stationary state from the beginning.

The statistical behavior of the models is shown in figure 13 through the second order structure functions. Both models systematically overestimate the structure functions, indicating that excessive energy remains stored in the resolved shells. An explanation can be made considering that the removed shells are almost all the shells belonging to the dissipative range. As such the truncation removes the dominant

mechanism responsible for energy dissipation while leaving unchanged the forcing, producing an accumulation of energy in the shells.

The previous results reveal the distinction between trajectory prediction and statistical prediction, in a learning context. The complemented ROM improves the short term evolution of the reduced system, demonstrating that the learned residual successfully reconstructs part of the missing dynamics. However, this improvement does not necessarily translate into a better reconstruction of the long term statistical properties.

This observations suggest that minimizing trajectory errors, over a limited time interval, might not be sufficient to recover the invariant measure of the original system. Instead, the closure must also reproduce the correct energy balance introduced by the unresolved scales. To improve the model the training objective could incorporate statistical quantities directly into the loss function, encouraging the network to reproduce the invariant measure. Furthermore, additional physical constraints could be introduced to enforce the correct energy balance between forcing and dissipation.

The implementation of [10] was modified to resolve shell-model dynamics and tested under different training conditions changing the value of N_{cutoff} and the hidden layer dimensions within the Lifter and Residual neural networks. Increasing the network capacity produced no significant improvement over certain values. The loss function exhibited more of the spiking for a larger model, but at the expense of both training and integration times. Reducing N_{cutoff} instead resulted frequently in numerical instability during integration.

Although the overall accuracy is still limited, these preliminary results are encouraging. Even in its current form, the complemented ROM consistently outperforms the standard Galerkin approximation in reproducing short term trajectories, indicating that the learned residual captures relevant information about the unresolved dynamics. The present case was intentionally performed under relatively favorable conditions, where only the dissipative shells are removed. A more stringent validation will therefore require larger shell models and a truncation deep in the inertial range. Such configurations are considerably closer to the coarse-graining problems encountered in practical turbulent flow simulations and will constitute the natural continuation of this work.

Future developments should also aim at enriching the loss function with statistical observables and/or physical constraints preserved over long times. In such a way we hope that the learned closure should reproduce better not only short term trajectories but also the invariant measure and long term energy balance of the full system.

7 Conclusions

The objective of this thesis was to investigate machine learning methodologies for the analysis of turbulent shell models. Shell models constitute a strong simplification of the NS equations, but they retain many of the characteristic features of fully developed turbulence, like chaos and intermittence. This makes them a good framework in which to benchmark and test machine learning techniques before extending them to more realistic turbulent systems.

The first implementation of the aforementioned techniques in this work was the identification of the governing equations of the SABRA shell model through a sparse regression approach inspired by the SINDy algorithm. Starting from an existing implementation, the numerical framework was substantially extended and adapted to perform a systematic investigation of the optimization procedure. Particular attention was put in understanding the learning procedure, in particular the scaling of the loss function with the shell wavenumber, in an effort to generalize the method to a real dataset. The scaling correction, while empirical in nature, was shown to improve the reconstruction of the sparse interaction, suggesting that a computationally light solution could be utilized in this context. However, the analysis highlighted limitations of the present approach, suggesting that behaviour typical of the highest shells, such as intermittency and/or shell correlations could also affect the learning dynamics.

The second part investigated causal relations through response theory. To estimate the response function a sampling from an ensemble of trajectories obtained from a parent one subjected to a gaussian perturbation. The numerical results revealed a pronounced asymmetry between upstream and downstream propagation, consistent with the directionality of the energy cascade. A confrontation with linear response theory prediction reveals that significant deviations rapidly emerge, reflecting the strongly nonlinear nature of the underlying dynamics.

Finally, a Complemented Deep-Reduced Order Model was adapted to the SABRA shell model to investigate data driven closure strategies for turbulent systems. Although the present implementation does not yet reproduce the long term statistical properties of the complete system, the closure learned consistently improves short term trajectory prediction with respect to the standard Galerkin reduced-order model. These preliminary results provide encouraging evidence that memory-based neural closures can recover the dynamics coming from unresolved scales.

The methodologies explored in this work are linked to one of the central challenges in the numerical simulation of turbulent flows, i.e. the construction of accurate closure models for large hydrodynamic systems. The numerical modeling of such systems has a wide range of applications, ranging form engineering to climate modeling, where the resolution of all the dynamically relevant scales is computationally prohibitive. From this perspective, shell models provide a controlled environment in which closure strategies can be developed and understood before being extended.

Several directions in which this work can be extended naturally emerge. On the sparse identification side, it would be interesting to investigate the learning dynamics under the addition of terms in the loss function enforcing physical properties, such as scale invariance or energy conservation. Furthermore a more in depth analysis of the loss function over the K41 could reveal more subtle dependencies of the scaling behaviour. For the reduced-order model, as previously anticipated, the next step could be adding in the learning objective statistical observables and physical constraints in order to allow for some form of long term learning. In parallel larger

systems and a truncation in the inertial range should be considered.

References

[1] L. Biferale et al. "Fluctuation-response relation in turbulent systems". In: *Phys. Rev. E* 65 (1 Dec. 2001), p. 016302. DOI: 10.1103/PhysRevE.65.016302. URL: https://link.aps.org/doi/10.1103/PhysRevE.65.016302.

[2] Luca Biferale. "SHELL MODELS OF ENERGY CASCADE IN TURBULENCE". In: *Annual Review of Fluid Mechanics* 35.Volume 35, 2003 (2003), pp. 441–468. ISSN: 1545-4479. DOI: https://doi.org/10.1146/annurev.fluid.35.101101.161122. URL: https://www.annualreviews.org/content/journals/10.1146/annurev.fluid.35.101101.161122.

[3] Tomas Bohr et al. *Dynamical Systems Approach to Turbulence*. Cambridge Nonlinear Science Series. Cambridge University Press, 1998.

[4] Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. "Discovering governing equations from data by sparse identification of nonlinear dynamical systems". In: *Proceedings of the National Academy of Sciences* 113.15 (2016), pp. 3932–3937. DOI: 10.1073/pnas.1517384113. eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.1517384113. URL: https://www.pnas.org/doi/abs/10.1073/pnas.1517384113.

[5] Sergio Chibbaro et al. "Building causation links in stochastic nonlinear systems from data". In: *Phys. Rev. E* 113 (4 Apr. 2026), p. 044114. DOI: 10.1103/kdyd-tc6n. URL: https://link.aps.org/doi/10.1103/kdyd-tc6n.

[6] Uriel Frisch. *Turbulence: The Legacy of A. N. Kolmogorov*. Cambridge University Press, 1995.

[7] Victor S. L'vov et al. "Improved shell model of turbulence". In: *Phys. Rev. E* 58 (2 Aug. 1998), pp. 1811–1822. DOI: 10.1103/PhysRevE.58.1811. URL: https://link.aps.org/doi/10.1103/PhysRevE.58.1811.

[8] Lorenzo Manfredini and Özgür D. Gürcan. *Nonlinear phase synchronization and the role of spacing in shell models*. 2026. DOI: https://doi.org/10.1103/2vxp-1k2t. arXiv: 2507.14142 [nlin.CD]. URL: https://arxiv.org/abs/2507.14142.

[9] Umberto Marconi et al. "Fluctuation-Dissipation: Response Theory in Statistical Physics". In: *Physics Reports* 461 (Apr. 2008). DOI: 10.1016/j.physrep.2008.02.002.

[10] Emmanuel Menier et al. "CD-ROM: Complemented Deep - Reduced order model". In: *Computer Methods in Applied Mechanics and Engineering* 410 (2023), p. 115985. ISSN: 0045-7825. DOI: https://doi.org/10.1016/j.cma.2023.115985. URL: https://www.sciencedirect.com/science/article/pii/S0045782523001081.

[11] Kolmogorov A. N. "The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers". In: *C. R. Acad. Sci. USSR* 30 (1941), p. 301. URL: https://cir.nii.ac.jp/crid/1573950398882283392.

Appendices

A Supplementary Regression Analysis

This appendix provides supplementary results to further clarify the structure of the linear regression framework. Specifically, we present the full weight matrices corresponding to the 12-shell system configuration.

The matrices in this section are structured as follows: to account for the commutativity of the product, redundant interaction terms have been removed from the basis. Consequently, the rows represent all unique, non-redundant two-shell interactions, while the columns correspond to the target shells whose dynamics are governed by these coefficients. These two-shell interactions are ordered following the convention established in Section 34, with the exception that the contributions of commuting terms are combined into the entry that appears first chronologically in the dictionary.

Structurally, the horizontal dotted lines denote a change in the first index of the interaction pair. The solid red line indicates the boundary below which only complex conjugate product terms (c.c. $\times$ c.c.) reside; as these terms do not physically couple to the forward dynamics, their coefficients are expected to be identically zero.

One can observe that the matrices seem almost noiseless indeed most of the problems observed for the interaction plots in section 4 come form the very small noisy terms.

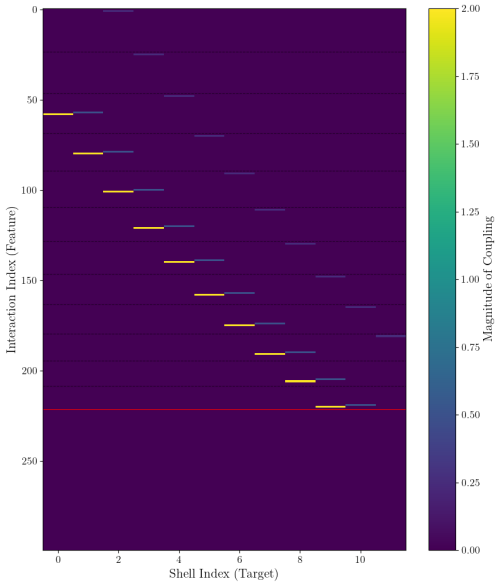
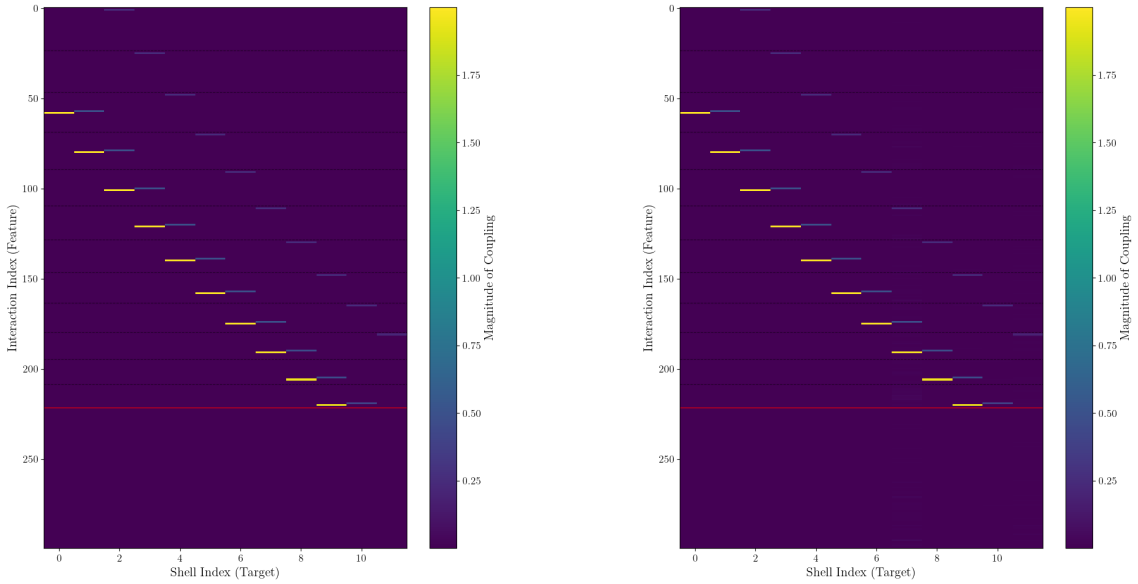


Figure 14: Magnitude of the true weights (a) and magnitude of the learned weights recovered via the LASSO optimization procedure (b).



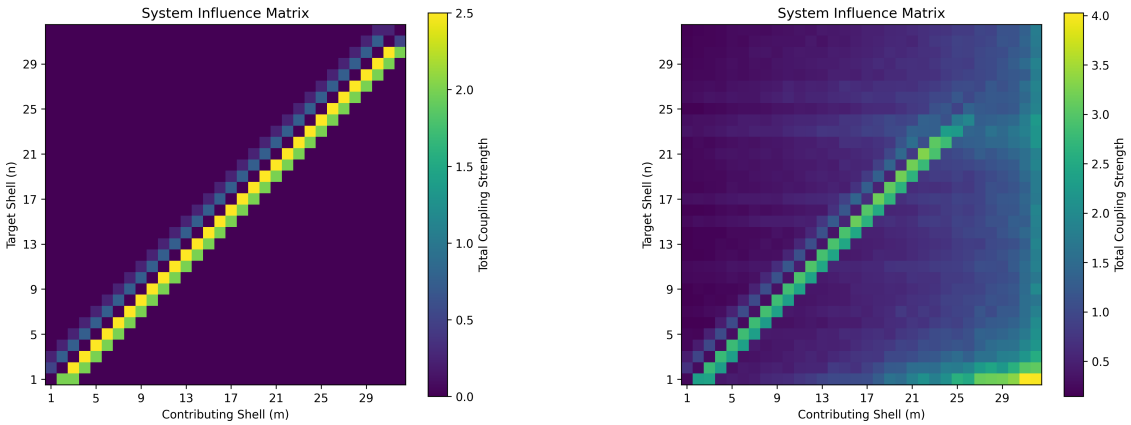
(a) (b)

Figure 15: Magnitude of the learned weights recovered via the LASSO optimization procedure and corrected scaling (a) and magnitude of the learned weights recovered via the LASSO optimization procedure (b).

B 32-shell model linear regression

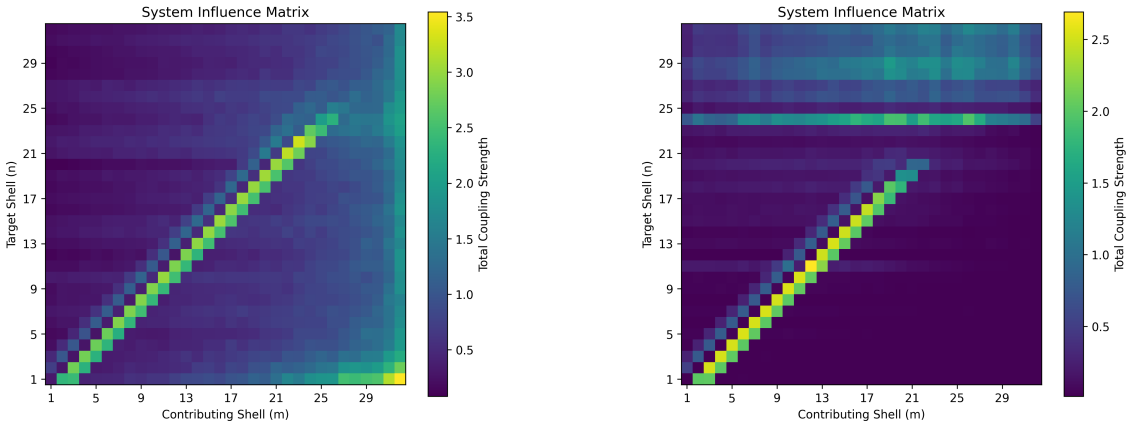
Results are presented in this appendix to provide a more complete picture of the limitations of the proposed sparse regression procedure when applied to larger shell systems.

To provide a more complete picture of the discussion presented in Section 4, we report in this appendix the results obtained by extending the proposed SINDy-like algorithm to a 32-shell model. The larger number of shells amplifies several of the effects discussed in the main text, making them easier to identify and analyze. Furthermore, we include results obtained both with and without the use of the L_1 regularization, allowing for a more direct comparison of the role played by sparsity in the reconstruction process.



(a) (b)

Figure 16: Interaction matrix of the ground truth (a) and linear regression without regularizer (b).



(a) (b)

Figure 17: Interaction matrix in the presence of the regularizer (a) and with the addition of a scaling (b).

figure 17(a) shows that, the L_1 regularization partially suppresses the spurious interactions introduced by the optimizer, still a significant number of nonphysical coefficients remain. Increasing the regularization strength further is not a viable solution, as it would also suppress physically relevant interactions, particularly those

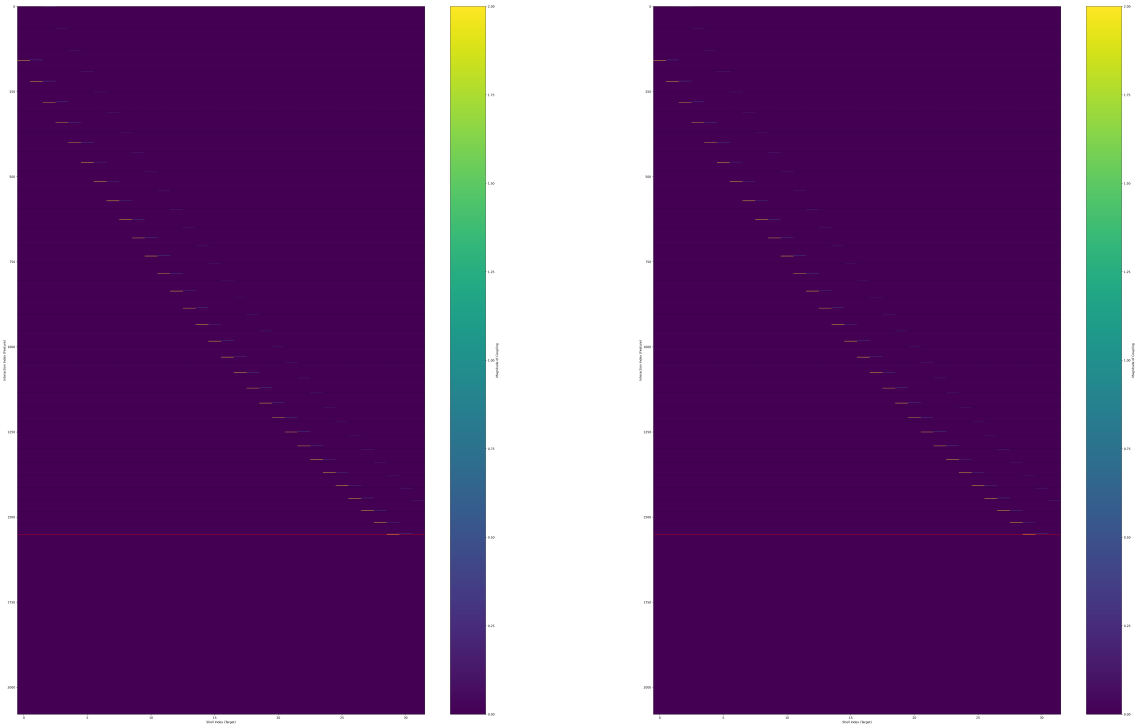
associated with the highest shell numbers, whose contribution to the loss function is intrinsically smaller.

Introducing the shell-dependent scaling discussed in Section 4 partially alleviates this issue. As shown in figure 17(b), the number of spurious interactions in the low-wavenumber region is substantially reduced. However, the correction overcompensates the contribution of the highest shells, causing the optimization to fit predominantly the stochastic fluctuations associated with those coefficients. As a consequence, the reconstructed interaction matrix becomes increasingly noisy in the high-wavenumber region.

The spatial distribution of the spurious coefficients also supports the scaling argument developed in Section 4. Most of the erroneous interactions concentrate in the upper-right corner of the interaction matrix, corresponding to interactions involving the highest shell numbers. According to the scaling derived in equation (39), these coefficients contribute the least to the loss function and are therefore only weakly constrained during optimization.

This effect is further amplified by the dimensionality of the parameter space. While the number of physically relevant coefficients increases only linearly with the number of shells $\mathcal{O}(N)$, the total number of possible interactions grows quadratically as $\mathcal{O}(N^2)$. Consequently, the stochastic fluctuations introduced by the Adam optimizer have a much larger space in which to generate spurious coefficients than physically meaningful ones, increasing the probability of converging to noisy solutions.

To provide a more complete comparison, the full learned weight matrices corresponding to the different optimization strategies are reported below.

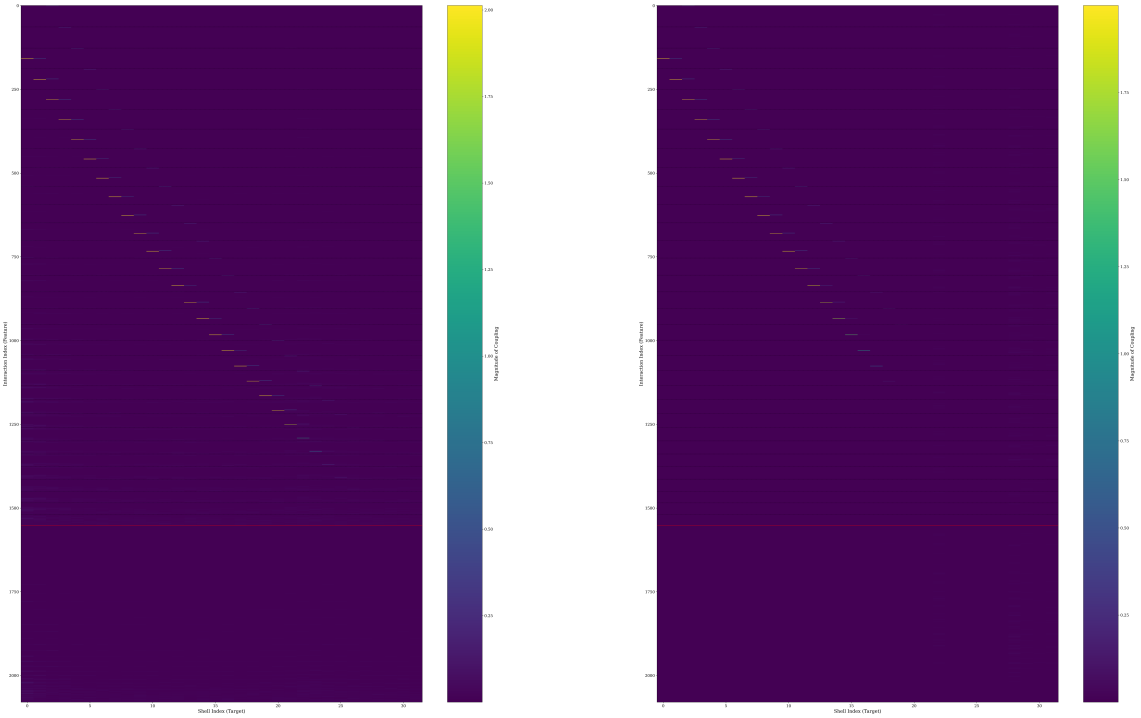


(a) (b)

Figure 18: Comparison of the weight matrices: (a) Ground truth versus (b) unregularized trivial linear regression.

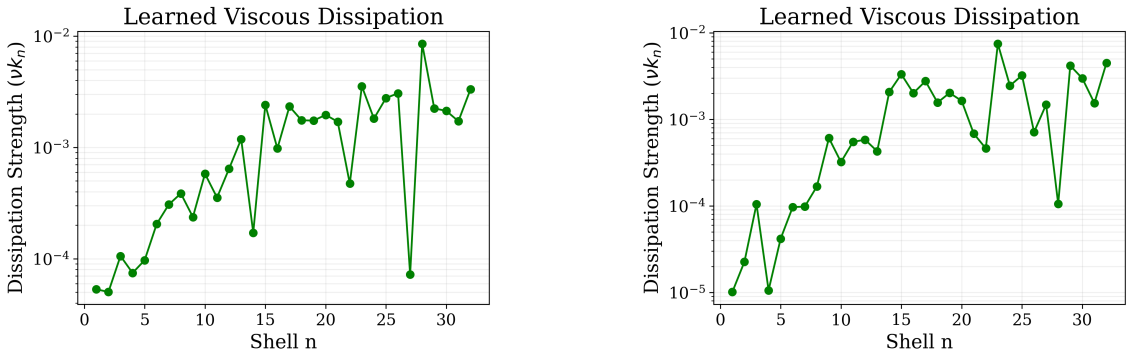
Although the weight matrices obtained through unregularized linear regression appear qualitatively similar to the ground truth (figure 18), this agreement is some-

what misleading. As shown in figure 16, the reconstructed interaction matrix contains a large number of spurious coefficients. While each of these coefficients has a relatively small magnitude, their cumulative contribution reproduces an interaction profile similar to the true one. The apparent agreement therefore arises from the superposition of many nonphysical interactions rather than from the correct identification of the sparse governing equations.



(a) (b)

Figure 19: Comparison of the weight matrices: (a) without and (b) with scaling correction.



(a) (b)

Figure 20: Learned viscous coefficients optimized (a) without and (b) with a scaling correction. While the qualitative scaling behavior remains partially identifiable, inaccuracies in training the nonlinear term appear to induce a strong numerical compensation effect. Consequently, the network introduces severe high-frequency oscillations into the viscous coefficients to counteract these unphysical model errors.

C 32-shell causality study

We include in this section the results obtained from the generalization of the causal study for a larger shell model. In this case the same statistics of $S = 1000$ samples were used. The perturbed shell was also in this case the sixth shell

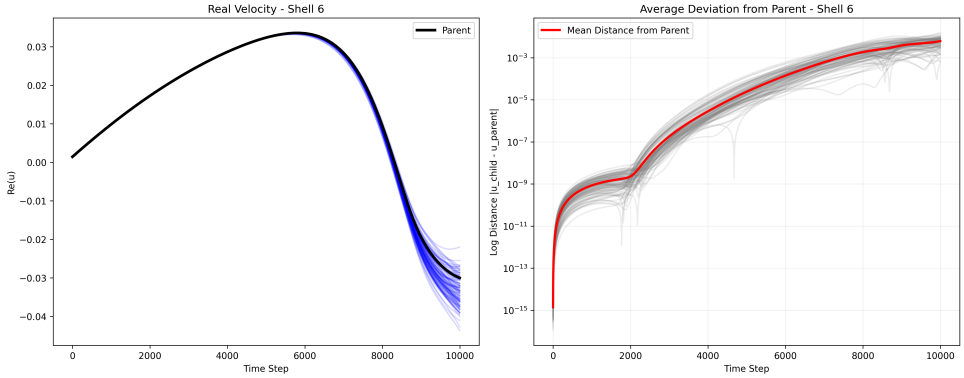


Figure 21: (Left) Real part of the trajectories sampled from the ensemble of the sixth shell, with the parent run highlighted. (Right) Corresponding deviations from the parent run.

In this case, because a larger shell model necessitates smaller time steps to satisfy numerical stability constraints, the simulated trajectories are significantly shorter, meaning we do not track them until the average distance saturates. This does not result in a significant loss of information, given the shcaling of the error.

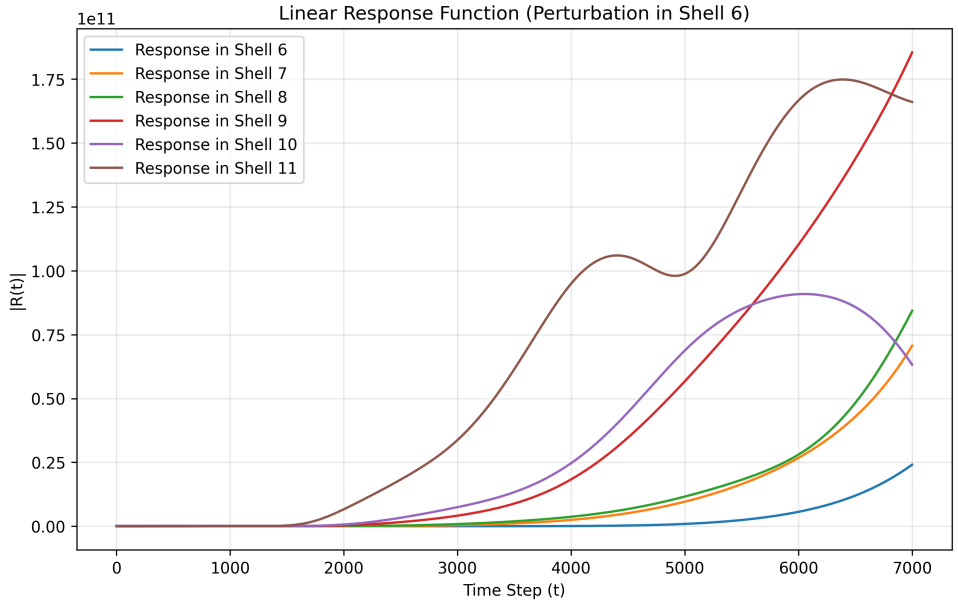
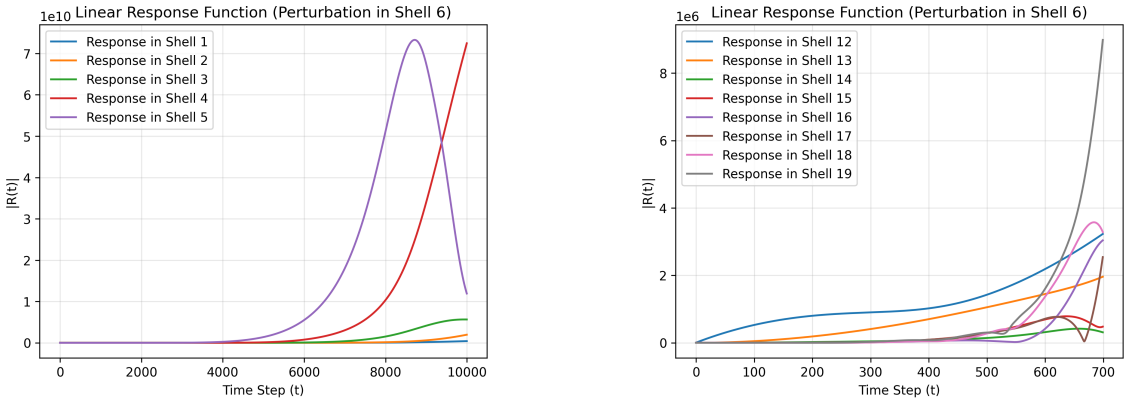


Figure 22: Response function of the first six shells downstream, there are many more time steps but it must be said that the single time step is 100 times smaller

The response of the immediate downstream shells resembles more closely the behavior described in Section 5.

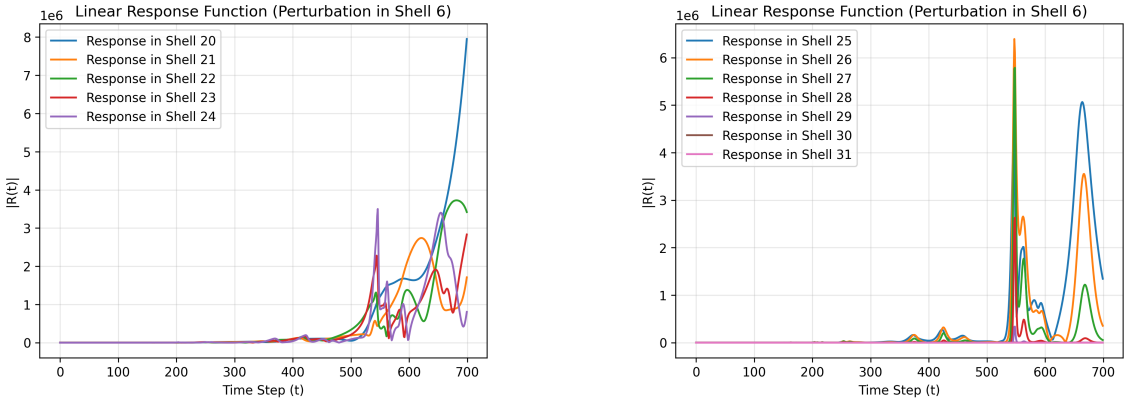
The response of the first few shells directly following the perturbed one is the one that follows more closely with the behavior described in Section 5. Where an ordering is present dependent on the distance from the perturbed shell.



(a) (b)

Figure 23: Response functions of the upstream shells (a) and the downstream shells spanning indices 12 to 19 (b).

On the other hand the behaviour shown in 23 , shows that the clear temporal asymmetry as shown previously is still present. Furthermore the response further downstream is much more chaotic and the spatio-temporal ordering gets progressively lost. This last result is better shown in the last shells in figure 24.



(a) (b)

Figure 24: Response functions of the upstream shells (a) and the downstream shells spanning indices 12 to 19 (b).