From the Mathematical Foundations of Turbulence to a Closure Model Alternative to Physics-Informed Approaches

Tsuyoshi Yoneda*†

1 Approximation

The Taylor series is an accurate approximation. A smooth function $f : \mathbb{R} \to \mathbb{R}$ is approximated near the origin by

$$f(x) = f(0) + f'(0)x + O(x^2)$$

where O denotes Landau notation. Here, we roughly write this approximation as

$$f(x) \sim f(0) + f'(0)x$$

This notation is commonly used in the field of turbulence physics. In mathematics, such notation cannnot be accepted. This gap arises because turbulence physics primarily deals with phenomena where it is often unclear whether an error estimate is even possible. In the next section, we will move on to more concrete discussions.

2 Kolmogorov's $-5/3$ power law

Let us derive Kolmogorov's $-5/3$ power law using " $\sim$ " introduced in the previous section. " $\sim$ " represents a rough approximation where both sides contain errors, and the error estimate cannot be performed rigorously. However, since we focus on the proportional relationship, we can apply multiplication and division on both sides.

For any vortex, let $\ell \in \mathbb{R}_{>0}$ be its characteristic length (we call it "length scale ℓ"), and let $t \in \mathbb{R}_{>0}$ be its rotation time, roughly speaking, the time it takes for the vortex to complete one rotation (or breakdown).

*Hitotsubashi University

†For the precise version of Section 1,2,7,8, see [15, 16], but written in Japanese.

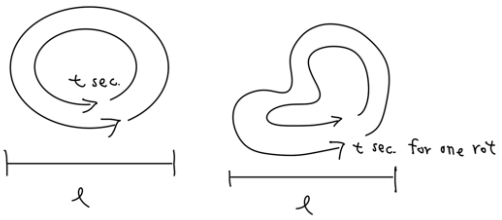

Figure 1: Image of vortex

We will provide the definition of vortex in detail later, but when deriving the $-5/3$ power law, the precise definition of vortex is actually unnecessary. This is a key point of the "dimensional analysis".

Since velocity is expressed as "length/time", $v \sim \ell/t$ is estimated as the representative velocity of vortex on the length scale ℓ. Since the square of velocity is energy, each vortex thus possesses v^2 energy (for simplicity, mass is always assumed to be 1). Here, let us assume **homogeneous and isotropic**. That is, the assumption that vortex of length scale ℓ are uniformly distributed in space and oriented with equal probability in any direction.

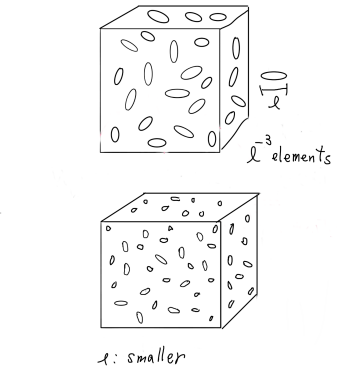

Figure 2: Image of homogeneous and isotropic

This assumption greatly simplifies the calculations. For example, consider homogeneous and isotropic turbulence in $\mathbb{T}^3 = \mathbb{R}^3/\mathbb{Z}^3$. In this case, the total energy (occupied by vortices of length scale ℓ) within a periodic box can be estimated as the volume of the domain itself (in this case 1) $\times v^2$, due to the

homogeneity that allows spatial variables to be neglected. Now, the time rate of change of energy at each scale is estimated as v^2/t.

Adopting the cascade picture, where vortices repeatedly split into smaller vortices (we justify this picture in Section 6), this time rate of change can be regarded as the "energy transfer rate".

Figure 3: Image of vortex breakdown

Therefore, let us denote this energy transfer rate by the symbol ϵ, and then we apply Kolmogorov's second hypothesis. That is, we assume that energy is transferred uniformly at any scale within the inertial range (i.e., ϵ does not depend on ℓ).

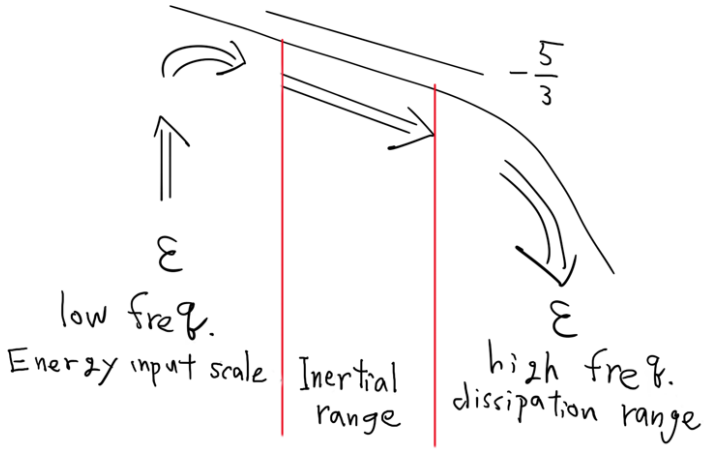

Figure 4: Image of inertial range in Fourier spectrum

Thus, we have

$$\epsilon \sim v^2/t \sim \ell^2/t^3.$$

Remark 1 *Moreover, the contrast between this similarity hypothesis and the self-similarity of the Navier-Stokes equations themselves (in function spaces such as $L^3(\mathbb{R}^3)$, $H^{1/2}(\mathbb{R}^3)$, etc.) is intriguing. Expressing turbulence similarity in functional spaces is corresponding to $C^{1/3}$ (appearing in the Onsager conjecture and the convex integration, see the pioneering work De Lellis-Székelyhidi[2] for example) and Constantin-E-Titi's $B^{1/3}_{3,\infty}$ [1]. See also Duchon-Robert [3].*

Then, the rotation time for vortices of length scale ℓ within the inertial range can be estimated as $t \sim (\ell^2/\epsilon)^{1/3}$. Furthermore, the energy spectrum $E(k)$ is estimated as $E(k) \sim k^{-1}v^2$, and $k \sim 1/\ell$, where k is called the wavenumber scale (these derivations will be explained in detail later). Consequently, we finally obtain

$$E(k) \sim \ell^3/t^2 \sim \epsilon^{2/3}\ell^{5/3} \sim \epsilon^{2/3}k^{-5/3}.$$

This is nothing more than Kolmogorov's $-5/3$ power law.

3 Three points of ambiguity for deriving $-5/3$ power law

The derivation of the power law mentioned earlier, contained three points of ambiguity:

- Definition of the energy spectrum,
- Definition of vortex,
- Definition of the energy transfer rate.

We will now clarify these three definitions. After understanding these three definitions, we will proceed to the true goal of this note: constructing a turbulence closure model.

4 Definition of Energy Spectrum

First, for a function $f : \mathbb{R}^3 \to \mathbb{R}$, the Fourier transform is defined as

$$\hat{f}(\xi) = \mathcal{F}[f(\cdot)](\xi) = \int_{\mathbb{R}^3} f(x)e^{ix\cdot\xi}dx,$$

and the energy E is defined as

$$E = \int_{\mathbb{R}^3} |u(x)|^2 dx = \int_{\mathbb{R}^3} |\hat{u}(\xi)|^2 d\xi.$$

Therefore, the goal is to obtain the following expression

$$E = \int_0^\infty E(k)dk$$

with an appropriate function $E(k)$, k is a positive real number, and this $E(k)$ defines the energy spectrum. By discretizing the positive real number k, we obtain the follwing:

$$\sum_{k\in(\Delta k)\mathbb{Z}_{\geq 0}} E(k)\Delta k \sim \sum_{k\in(\Delta k)\mathbb{Z}_{\geq 0}} \int_{k-\frac{\Delta k}{2}\leq|\xi|<k+\frac{\Delta k}{2}} |\hat{u}(\xi)|^2 d\xi.$$

Thus we have

$$E(k)\Delta k \sim \int_{k-\frac{\Delta k}{2}\le|\xi|<k+\frac{\Delta k}{2}} |\hat{u}(\xi)|^2 d\xi.$$

A vortex is **a broad object (a continuum)**, and thus it possesses a characteristic width (here essentially different from **vorticity**!!). Let us express the characteristic width at the wavenumber k of the vortex as $[k/\sqrt{\alpha}, \sqrt{\alpha}k]$, where $\alpha > 1$ represents the characteristic wavenumber width of the vortex. Roughly saying, the shape itself does not change much at any wavenumber scale (scaling invariant). With this in mind, substitute k into Δk, then we have the following proportional relation:

$$\left|\xi : k - \frac{\Delta k}{2} \le |\xi| < k + \frac{\Delta k}{2}\right| \sim \left|\xi : \frac{k}{\sqrt{\alpha}} \le |\xi| \le \sqrt{\alpha}k\right|.$$

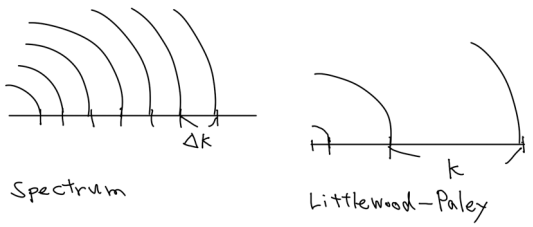

Figure 5: Image of spectrum

Therefore, the energy of vortices with wave number k is expressed as

$$v^2 \sim \int_{\mathbb{R}^3} |\hat{u}(\xi)|^2 \chi_{[\frac{k}{\sqrt{\alpha}}\le|\xi|\le\sqrt{\alpha}k]}(\xi) d\xi \sim kE(k).$$

Next, we show that $k \sim 1/\ell$. This is just the scale transformation of the Fourier transform. Specifically,

$$\int_{\mathbb{R}^3} f(\lambda x) e^{ix\cdot\xi} dx = \int_{\mathbb{R}^3} f(x) e^{i\lambda^{-1}x\cdot\xi} \lambda^{-3} dx \quad (\lambda > 0).$$

That is,

$$\mathcal{F}[f(\lambda\cdot)](\xi) = \lambda^{-3}\mathcal{F}[f(\cdot)](\lambda^{-1}\xi)$$

it follows (neglect a factor of λ^{-3}) that

$$k \sim 1/\ell.$$

As a mathematician, it seems better for us to regard that this $k \sim 1/\ell$ is rather defined by the above scale transformation calculation for any function f.

5 Definition of vortex

To define the scaled vortex, first, let us apply the scale decomposition to the velocity field u as follows:

$$\bar{u}_k(x) := \mathcal{F}_\xi^{-1}[\hat{u}(\xi)\chi_{[\frac{k}{\sqrt{\alpha}} \le |\xi| \le \sqrt{\alpha}k)}(\xi)](x),$$

where $\hat{\cdot}$ is the Fourier transform and $\mathcal{F}_\xi^{-1}$ is its inverse. Keep in mind that the decomposition of u in terms of $\bar{u}_k$

$$u = \sum_{k \in \alpha^{\mathbb{Z}}} \bar{u}_k$$

holds at least in the L^2 sense. $\alpha > 1$ represents the characteristic wavenumber width of vortices discussed in the previous section. Then we can characterize the **identified vortex**, more precisely, vortex axis ℓ_k (smooth curve) must satisfy the following condition (necessary condition):

$$\bar{u}_k(t, \Phi(t, \ell_k)) = 0, \quad \partial_t \Phi(t, x) = u(t, \Phi(t, x)).$$

Figure 6: Image of vorticity to vortex axis ℓ_k

To understand this condition, the crucial point (for me) is the following generalized vortex Rankine vortex as the typical example: For a smooth $\varphi : \mathbb{R} \to \mathbb{R}_{>0}$, let

$$\bar{u}_k(x) = \varphi(x_3) \times \begin{cases} \begin{pmatrix} x_2 \\ -x_1 \\ 0 \end{pmatrix} & (r \le 1/k), \\ \dfrac{1}{(rk)^2} \begin{pmatrix} x_2 \\ -x_1 \\ 0 \end{pmatrix} & (r > 1/k), \end{cases}$$

where $r = \sqrt{x_1^2 + x_2^2}$. Clearly, $\{x : x_1 = x_2 = 0\}$ is regarded as the vortex axis.

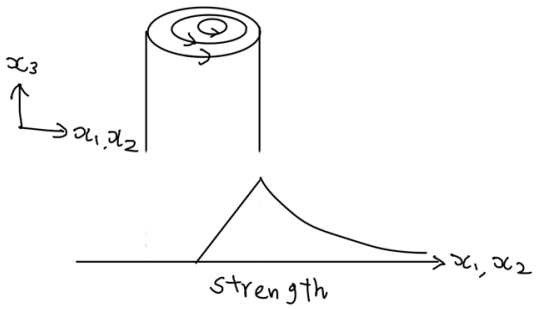

Figure 7: Image of Rankine vortex

Question 1 *Actually, this example is not accurate, since its Fourier transform is nonlocal. Find more appropriate function whose Fourier transform is local, and detect the corresponding vortex axis.*

But in reality, vortex axis is always slightly bent. In this case, manipulating the following pressure expression seems effective:

$$\bar{p}_k := (-\Delta)^{-1}\nabla \cdot (\bar{u}_k \cdot \nabla \bar{u}_k).$$

Remark 2 *Experimentally, manipulating this pressure expression goes well for detecting vortex axes, and we call it "low-pressure method". This detecting method is initiated and developed by Miura-Kida[9], Kida-Miura[7] and Trusuhashi-Goto-Oka-Y[13].*

For $x \in \mathbb{R}^3$, let the eigenvalues of the Hessian of $\bar{p}_k(x)$ be

$$\lambda_1(x) \geq \lambda_2(x) \geq \lambda_3(x)$$

and let the corresponding eigenvectors be $e_1(x), e_2(x), e_3(x)$. Since it is the Hessian, each eigenvector is orthogonal to the others. Since vortices are regions of low pressure with axis, it is natural to focus on the set (open set) of points x satisfying $\lambda_1 > 0$ and $\lambda_2 > 0$, i.e., where the first and second eigenvalues are positive which represent the low pressure region. In this case, one of the e_3 is a candidate for the vortex axis direction. To pick up the appropriate e_3, choose a reference curve $\zeta(r)$ such that $\partial_t \zeta \cdot e_3(\zeta) \neq 0$, and define a family of laminar surfaces orthogonal to e_3 as

$$\begin{aligned} D_r := \{&\eta(s_1, s_2) \in \mathbb{R}^3 : \eta(0,0) = \zeta(r), \\ &\partial_t \zeta \cdot e_3(\zeta(t)) \neq 0,\ \partial_{s_1}\eta \cdot e_3(\eta) = \partial_{s_2}\eta \cdot e_3(\eta) = 0, \quad \lambda_2(\eta) > 0\}. \end{aligned}$$

Note that $D_r \cap D_{r'} = \emptyset$ $(r \neq r')$. The vortex axis is obtained by finding the points on each D_r that achieve the local minimum of $\nabla \bar{p}_k \cdot e_3$ and smoothly connecting them along the parameter r of the reference curve.

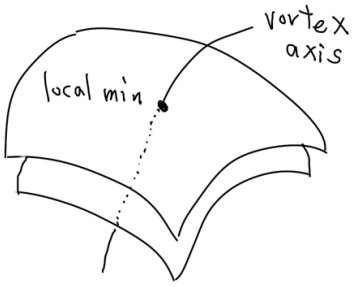

Figure 8: Image of D_t and vortex axis

Let $\bar{u}_k$ be the generalized Rankine vortex described before. By direct calculation, we obtain

$$\nabla \cdot (\bar{u}_k \cdot \nabla)\bar{u}_k = \varphi(x_3) \times \begin{cases} -2 & (r \le 1/k) \\ \dfrac{2}{k^4 r^4} & (r > 1/k). \end{cases}$$

This is rotationally symmetric about the axis $\{x : x_1 = x_2 = 0\}$, and since the integral kernel of $(-\Delta)^{-1}$ is also rotationally symmetric, it follows that on D_t, the point $x_1 = x_2 = 0$ is always a minimum value. It is easy to imagine that even if this vortex is slightly bent, the minimum point will likely remain near the axis (it may be interesting to formulate it an estimate the error in mathematics).

Recently, we figured out that this vortex axis definition represents the **multifractal nature of turbulence** (the paper is now in preparation). This, in turn, justifies the definition of the vortex axis here.

6 Spatial statistical analysis of energy transfer

If we focus on spatial statistical analysis, we easily conclude that, at each scale, energy is transferred to the next adjacent smaller scale, i.e., **vortices repeatedly split into smaller vortices**. To justify this picture in spatial statistics, we start from the following 3D-vorticity equations (omit the viscosity term):

$$\partial_t \omega + u \cdot \nabla \omega - \omega \cdot \nabla u = \nabla \times f, \quad \nabla \cdot \omega = 0,$$

where

$$\overline{(\nabla \times \hat{f})}_k = \begin{cases} \nabla \times \hat{f} & \text{in the energy input scale,} \\ 0 & \text{in the inertial range.} \end{cases}$$

By the statistically stationary, it is natural to assume that

$$\frac{d}{dt}\|\bar{\omega}_k(t)\|_{L^2}^2 = 0 \quad \text{for any} \quad k \quad \text{and} \quad t \ge 0,$$

and also assume the following approximation based on the previous section (vorticity to vortex axes: delta distribution supported on a curve):

$$\bar{\omega}_k(x)|_{x\in\mathbb{R}^3} \simeq \bar{\omega}_k(x)|_{x\in\ell_k}, \quad \ell_k \cap \ell_{k'} = \emptyset \quad (k \neq k').$$

Multiplying $\bar{\omega}_k$ on both sides of the above vorticity equations and applying integration by parts, then the convection term disappears. Then we have the following fundamental equation (divided by k^2, to adjust the dimension):

$$\sum_{k'} \frac{1}{k^2} \int \bar{\omega}_k \cdot \nabla \bar{u}_{k'} \cdot \bar{\omega}_k \simeq \begin{cases} \dfrac{1}{k^2} \displaystyle\int_{\mathbb{R}^3} \nabla \times f \cdot \bar{\omega}_k & \text{in the energy input scale,} \\ 0 & \text{in the inertial range} \end{cases}$$

for each k. Note that

$$u = \sum_{k'} \bar{u}_{k'}.$$

From this fundamental equation, we already have the numerical result (see [14, Figure 1]), which justifies the scale-by-scale energy forward cascade.

Question 2 *This argument does NOT work for the 2D case. The reason is that, since the convection term disappears, the energy transfer never occur due to the absence of the vortex stretching term, and it is in contradiction. We know there certainly exists cascade and inverse cascade in the 2D flow. Thus, energy cascade in 2D must NOT be based on the existence of broad objects (like identified vortices), which is different from the 3D case. From this insight, can we provide an interesting mathematical theorem?*

7 Spatiotemporal dynamics of energy transfer

This section provides the fundamental setting for considering the spatiotemporal dynamics of energy transfer. First, we describe the Euler equations as follows (for simplicity, we consider without external force case):

$$\partial_t u + (u \cdot \nabla)u = -\nabla p, \quad \nabla \cdot u = 0, \quad u|_{t=0} = u_0.$$

Here, we consistently assume that the Euler solutions $u \in L^2$ are sufficiently smooth, then we have the following energy balance:

$$\frac{d}{dt} \int_{\mathbb{R}^d} |u(t,x)|^2 dx = \frac{d}{dt} \|u(t)\|_{L^2}^2 = 0.$$

This expresses that the total energy (the L^2 norm) of the fluid motion does not change with time evolution. Keeping this in mind, we will examine what happens to the total energy of the mean flow in a short time. Here, we recall the Fourier transform and its inverse:

$$\hat{f}(\xi) = \int_{\mathbb{R}^d} f(x) e^{ix\cdot\xi} dx, \quad \mathcal{F}_\xi^{-1}[\hat{g}(\xi)](x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \hat{g}(\xi) e^{-ix\cdot\xi} d\xi.$$

We define the mean flow $\langle u \rangle$ and the turbulent fluctuation u' as follows. For simplicity, the wavenumber scale of the mean flow is set to 1.

$$\langle u(x) \rangle = \mathcal{F}_\xi^{-1}[\hat{u}(\xi)\chi_{\{|\xi|\leq 1\}}](x), \quad u'(x) = 1 - \langle u(x) \rangle.$$

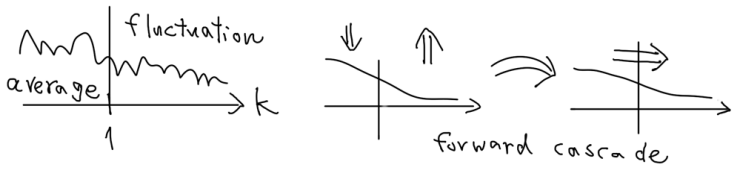

Figure 9: Image of rigorous forward cascade

Note that $\langle u \rangle$ can be re-expressed as a convolution with the sinc function. However, in the following discussion of Reynolds stresses, the convolution with the Gaussian function is more compatible than the convolution with the sinc function. Therefore, in what follows, we will proceed with the convolution with the Gaussian function. To derive the energy transfer rate, apply $\langle \cdot \rangle$ to both sides of the Navier-Stokes equations. This yields the following equation:

$$\partial_t \langle u \rangle + (\langle u \rangle \cdot \nabla)\langle u \rangle + \nabla \cdot \tau = -\nabla \langle p \rangle, \quad \nabla \cdot \langle u \rangle = 0.$$

Here, $\tau = \langle u \otimes u \rangle - \langle u \rangle \otimes \langle u \rangle$ is called the Reynolds stress. Note that this τ depends on the turbulent fluctuation u', while the other terms depend only on $\langle u \rangle$ and $\langle p \rangle$, i.e., they are expressed solely in terms of the mean flow (mean pressure). We emphasize that this u' is hidden in the original u.

Remark 3 *Expressing this τ using only $\langle u \rangle$ is a major goal of turbulence research. This is called the closure problem.*

To examine energy transfer, multiply both sides by $\langle u \rangle$ and take a spatial integral, then we have

$$\frac{d}{dt}\frac{1}{2}\int |\langle u(x,t) \rangle|^2 dx = \int \bar{S}(x,t) : \tau(x,t) dx,$$

where $A : B = \sum_{i,j=1}^{3} a_{ij} b_{ij}$ and $\bar{S}$ is called the rate-of-strain tensor and is defined as

$$\bar{S} := \frac{1}{2}\left[\nabla \langle u \rangle + (\nabla \langle u \rangle)^T\right].$$

Regarding the rate-of-strain tensor, it is useful to compare, for example, a typical two-dimensional hyperbolic flow $(-x_1, x_2)$ with a rotation $(x_2, -x_1)$ that is simply rigid body rotation. The rate-of-strain tensor for the former is calculated as

$$\bar{S} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}.$$

We immediately see that the eigenvalues and eigenvectors of this matrix represent the degree of deformation. In contrast, for the latter case,

$$\bar{S} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

which intuitively agrees with the absence of any deformation. Here, the expression on the right-hand side,

$$\int \bar{S}(t,x) : \tau(t,x)dx$$

represents the **energy transfer rate**. Recall that its dimension is $\sim v^2/t$.

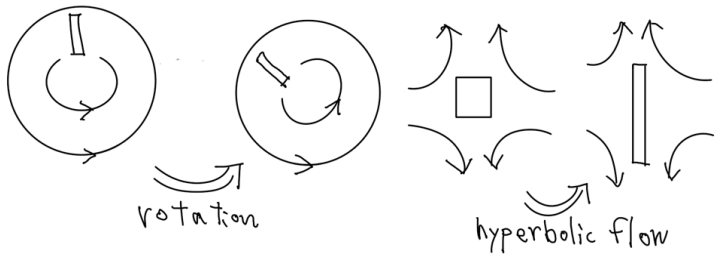

Figure 10: Image of rate of strain tensor

8 Derivation of a Classical Turbulence Model: The Smagorinsky Model

The most critical issue in turbulence research is the closure problem. It is no exaggeration to say that all turbulence research aims to solve this problem. More specifically,

Can τ be approximated using only the mean flow $\langle u \rangle$?

As a first step toward addressing this issue, we derive the well-known turbulence model, called "Smagorinsky model" (the calculation in this section is based on Eyink [4]). First, let us decompose τ. To do so, recall the following decomposition:

$$u = \sum_{k \in \alpha^{\mathbb{Z}}} \bar{u}_k.$$

Using this decomposition, τ becomes

$$\tau = \sum_{k \in \alpha^{\mathbb{Z}}} \sum_{k' \in \alpha^{\mathbb{Z}}} \langle \bar{u}_k \otimes \bar{u}_{k'} \rangle - \langle \bar{u}_k \rangle \otimes \langle \bar{u}_{k'} \rangle.$$

Assuming vortex stretching between scale ratios $\alpha \sim 1.7$ as in Y-Goto-Tsuruhashi [14], we have the following approximation:

$$\tau \sim \langle \bar{u}_\alpha \otimes \bar{u}_\alpha \rangle - \langle \bar{u}_\alpha \rangle \otimes \langle \bar{u}_\alpha \rangle.$$

Next, our aim is to apply a spatial approximation. For $\delta u := u(x+r) - u(x)$, then the above equation becomes

$$\langle \delta \bar{u}_\alpha \otimes \delta \bar{u}_\alpha \rangle - \langle \delta \bar{u}_\alpha \rangle \otimes \langle \delta \bar{u}_\alpha \rangle.$$

This follows from rigorous mathematical calculation using the fact that the integral of the Gaussian function is 1. That is, by definition,

$$\begin{aligned} & \langle \delta \bar{u}_\alpha \otimes \delta \bar{u}_\alpha \rangle - \langle \delta \bar{u}_\alpha \rangle \otimes \langle \delta \bar{u}_\alpha \rangle \\ = & \int_{\mathbb{R}^d} G(r)(\bar{u}_\alpha(x+r) - \bar{u}_\alpha(x)) \otimes (\bar{u}_\alpha(x+r) - \bar{u}_\alpha(x))dr \\ - & \int_{\mathbb{R}^d} G(r)(\bar{u}_\alpha(x+r) - \bar{u}_\alpha(x))dr \otimes \int_{\mathbb{R}^d} G(r)(\bar{u}_\alpha(x+r) - \bar{u}_\alpha(x))dr. \end{aligned}$$

Since canceling terms appear, taking them into account yields the final expression:

$$\begin{aligned} = & \int_{\mathbb{R}^d} G(r)\bar{u}_\alpha(x+r) \otimes \bar{u}_\alpha(x+r)dr \\ & - \int_{\mathbb{R}^d} G(r)\bar{u}_\alpha(x+r)dr \otimes \int_{\mathbb{R}^d} G(r)\bar{u}_\alpha(x+r)dr \\ = & \langle \bar{u}_\alpha \otimes \bar{u}_\alpha \rangle - \langle \bar{u}_\alpha \rangle \otimes \langle \bar{u}_\alpha \rangle. \end{aligned}$$

Now, let us apply the following mean value theorem:

$$\begin{aligned} \delta \bar{u}_\alpha & \sim \left(r_1 \frac{\partial}{\partial x_1} + r_2 \frac{\partial}{\partial x_2} + r_3 \frac{\partial}{\partial x_3} \right) \bar{u}_\alpha(x) \\ & = \left(r_1 \frac{\partial}{\partial x_1} + r_2 \frac{\partial}{\partial x_2} + r_3 \frac{\partial}{\partial x_3} \right) (\bar{u}_{\alpha,1}(x), \bar{u}_{\alpha,2}(x), \bar{u}_{\alpha,3}(x))^T \end{aligned}$$

(see Eyink [4], Section 3.2.2). Applying the mean value theorem reduces the calculation of τ to the following integral:

$$\begin{aligned} & \int_{\mathbb{R}^d} r_{i'} r_{j'} G(r)dr - \int_{\mathbb{R}^d} r_{i'} G(r)dr \int_{\mathbb{R}^d} r_{j'} G(r)dr \\ = & \int_{\mathbb{R}^d} r_{i'} r_{j'} G(r)dr \\ = & \begin{cases} 0 \quad (i' \neq j'), \\ \int r_{i'} r_{i'} G(r)dr = C > 0 \quad (i' = j'). \end{cases} \end{aligned}$$

By applying the above integral calculation (note that i, j and i', j' have different meanings), we obtain the following approximation:

$$\{\tau\}_{ij} \quad \sim \quad C\sum_{h=1}^{3}\frac{\partial\bar{u}_{\alpha,i}}{\partial x_h}\frac{\partial\bar{u}_{\alpha,j}}{\partial x_h}.$$

Here, we assume a typical Burgers vortex with the interior of the Rankine vortex for stretched fluctuation. More specifically, let

$$\bar{u}_\alpha(x) = \begin{pmatrix} x_2 \\ -x_1 \\ 0 \end{pmatrix}, \quad \langle u(x)\rangle = \begin{pmatrix} -\frac{x_1}{2} \\ -\frac{x_2}{2} \\ x_3 \end{pmatrix}.$$

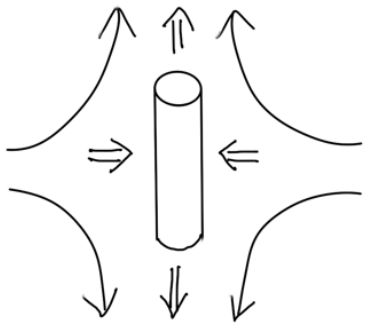

Figure 11: Image of this sernario

Then, we immediately have

$$\tau \sim C\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \bar{S} = \begin{pmatrix} -\frac{1}{2} & 0 & 0 \\ 0 & -\frac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Therefore, the energy transfer rate is given by

$$\int \bar{S} : \tau \sim -C.$$

Thus the energy of the mean flow decreases by the vortex stretching, and this is called "eddy viscosity". From this insight, we may accept the following well-known turbulence model, that is, Smagorinsky Model:

$$\tau_{ij} \sim -\nu_t\bar{S}_{ij}, \quad \nu_t \sim \ell^2\sqrt{\bar{S} : \bar{S}},$$

where ℓ is the grid size (recall that we set $\ell = 1$ in the previous section). In this case

$$\int \bar{S} : \tau \sim -\ell^2 \int |\bar{S} : \bar{S}|^{\frac{3}{2}}.$$

Note that, since $\epsilon \sim v^2/t \sim v^3/\ell$ and $|\bar{S} : \bar{S}|^{1/2} \sim v/\ell$, the dimension is correct.

Question 3 *Based on the mathematical calculation provided in this section, let us justify the Smagorinsky model more rigorously.*

9 A modern turbulence model based on machine learning, NOT physics-informed

In what follows, we drastically change the topic, not related to the classical physics anymore. Just I propose a totally new approach to construct a modern turbulence model. Without following the traditional argument of turbulence research up to the previous section, we rephrase the closure problem to the following:

Can a turbulence closure model with high prediction accuracy be constructed for the mean flow $\langle u \rangle$?

We point out a recent breakthrough regarding this closure problem. This breakthrough was induced by the Google Science Team in 2023 [8]. It involves constructing a machine learning model (NOT physics-informed anymore) for weather forecasting that surpasses conventional physical models. Again, the crucial point is that, the closure model was constructed directly from data, not from physical equations at all.

Here, we turn to a new discussion, which is completely different from the previous sections, and introduce the construction of a turbulence closure model using machine learning (NOT physics-informed). First, turbulence data can be specifically regarded as time-series data. For example, at N fixed observation points, three-directional wind velocity components generate $3 \times N$ (=: $u.shape[0]$) time series data.

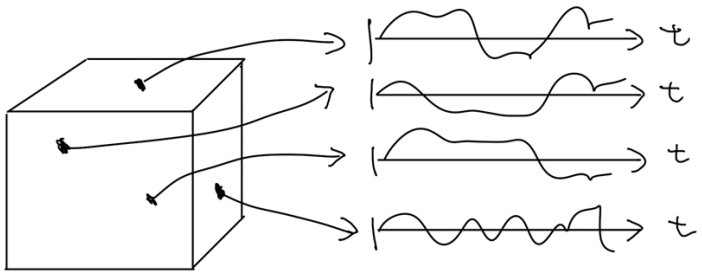

Figure 12: Image of generation of time series data

- Case 1: the set of time series data contains rich spatial information

  Whether an FFT can be applied in the spatial direction serves as a criterion. For example, in periodic box turbulence obtained from direct numerical computation, the corresponding data must contain abundant spatial information. In such cases, it is appropriate to apply the FFT (in the spatial direction) to the data and train a ML model on the time evolution of its amplitudes. The filter can be a low-pass or band-pass that directly cuts off high spatial frequencies. For the ML in this case, see Nakai-Saiki [10].

- Case 2: the set of time series data contains little spatial information

  If the spatial structure of the data is weak, or if spatial information is entirely absent, filters in the spatial direction may not be appropriate. Even in this case, filtering can still be applied in the time direction, provided that the underlying systems generating each time series are considered to be nearly identical.

- Case 3: The intermediate case

  Graph Neural Networks might be one of them?

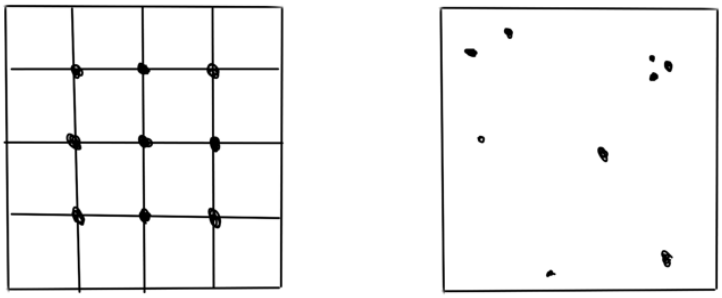

Figure 13: Left: suitable for FFT, RIGHT: try to apply GNN in the future

Here, we focus on the case 2, thus we apply a filter in the time direction.

Remark 4 *The filter stored on GitHub is the modern version as a data-driven filter. As for the algorithm of it, see Jinno-Mitsui-Nakai-Saiki-Y [6]. When using conventional filters (such as band-pass filter in the time direction), information from the future—which we should ideally predict—can sometimes be involved.*

We focus on one of the simplest ML, Reservoir Computing model (a type of RNN). Also here, we assume that **the underlying systems (for each time**

series) are identical. Under this assumption, it becomes unnecessary to construct separate RNNs for each time series data, i.e., a learning model where each different RNN interacts with others. Consequently, we only need to build a single, common RNN learning model applicable to all time series data. In what follows, let us explain this learning scheme in detail.

9.1 Overview of the Learning Scheme

In reality, the time direction is also discrete, so the obtained turbulence data can be written as:

$$\{u_{j,t}\}_{j\in[0,1,\cdots,u.shape[0]-1],t\in[0,1,\cdots,u.shape[1]-1]}.$$

The number of discrete data points is written as $u.shape[1]$, this notation comes from the Python code. Also, the number of time series data is written as $u.shape[0]$.

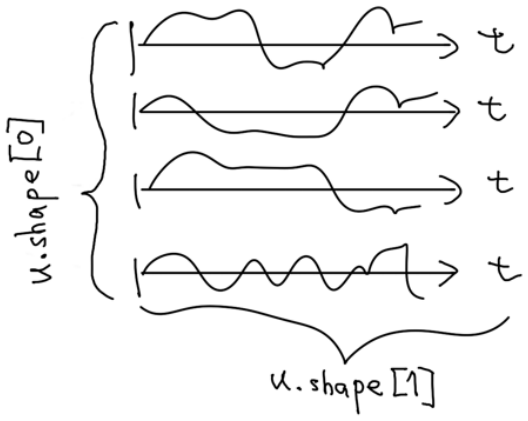

Figure 14: Image of data shape

Then, we rewrite this time series data in a form of dim-dimensional vector as the following (we say "delay coordinate", see [11] for example):

$$U_{j,t} = \begin{pmatrix} u_{j,t} \\ u_{j,t-lag} \\ \vdots \\ u_{j,t-(dim-1)lag} \end{pmatrix}.$$

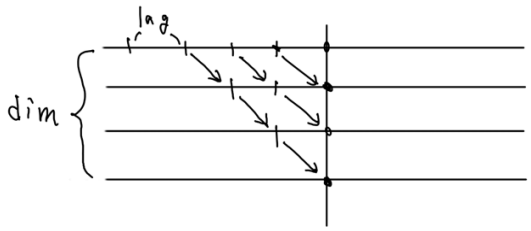

Figure 15: Image of $U_{j,t}$

First, let us consider the case where a distinct RNN is assigned to each time series data. Let M denote the number of learnable parameters. For a function $F_j : \mathbb{R}^M \times \mathbb{R}^{dim} \to \mathbb{R}^{dim}$ that depends on numerous parameters, the one-step-ahead prediction value $\widehat{U}_{j,t+1}$ is defined as

$$\widehat{U}_{j,t+1} = F_j\left(\{w_i\}_{i=0}^{M-1}, \{U_{j,t}\}_{j=0}^{u.shape[0]-1}\right)$$

where $\{w_i\}_{i=1}^{M}$ are called learnable parameters, corresponding to the reservoir's W_{in}, W, and W_{out} described in the next section. We then select the $\{w_i\}_i$ that minimizes the energy with respect to the actual data $U_{j,t+1}$. That is,

$$\sum_{j=0}^{u.shape[0]-1} \sum_{t=0}^{T_{train}-1} |\widehat{U}_{j,t+1} - U_{j,t+1}|^2$$

is minimized (we omit regularization term here). This is the broad framework of machine learning. As explained earlier, assuming "the identity of the underlying systems" is the key point in the turbulence model construction here. So, F_j can be unified into a single F independent of j, that is,

$$\widehat{U}_{j,t+1} = F\left(\{w_i\}_{i=0}^{M-1}, U_{j,t}\right).$$

This significantly simplifies the learning model. With this simplification, the next section explains reservoir computing.

10 Reservoir Computing

Although the previous sections focused on turbulence, the machine learning model introducing here is actually highly general and applicable to a wide range of problems. My research topics are actually forecasting of stock price data and weather data, focusing on small-size datasets (less than 10,000 samples). In this case we have a crucial question:

For such small-size datasets, is it really necessary to pursue the context-level dependencies that LSTMs and Transformers are good at?

To answer this question, please refer to my GitHub repository, **"compare-LSTM-GRU-Reservoir"** . My conclusion is that reservoir computing [5] performs best for the small-size datasets. First, let us consider the following matrices as trainable parameters:

- Input weight matrix: $W^{in} \in \mathbb{R}^{N_x \times dim}$
- Recurrent weight matrix: $W \in \mathbb{R}^{N_x \times N_x}$
- Output weight matrix: $W^{out} \in \mathbb{R}^{dim \times N_x}$

Here, substitute the data $U_{j,t} \in \mathbb{R}^{dim}$ and the hidden layer vector $x_{j,t} \in \mathbb{R}^{N_x}$ into the following recurrence relation [1] to output the next step's hidden layer vector $x_{j,t+1} \in \mathbb{R}^{N_x}$ and the predicted value $\widehat{U}_{j,t+1} \in \mathbb{R}^{dim}$:

$$(U_{j,t}, x_{j,t}) \mapsto (\widehat{U}_{j,t+1}, x_{j,t+1}),$$

$$\begin{cases} x_{j,t+1} &= (1-\alpha)x_{j,t} + \alpha \tanh(W^{in}U_{j,t} + Wx_{j,t}), \\ \widehat{U}_{j,t+1} &= W^{out}x_{j,t+1}. \end{cases}$$

$\alpha \in (0, 1]$ is called the leaking rate, which is similar to the well-known concept of residual connections. The learning scheme will be explained in detail starting in the next section. Referring to the conceptual diagram in the next section must help us understand more deeply.

10.1 Conceptual diagram of the learning scheme

The learning scheme introduced here involves two stages: model selection and generalization performance evaluation. The conceptual diagram is as follows (each of model selection and generalization performance evaluation includes a training phase):

[1] In the GitHub code, we also put "bias" in Reservoir.

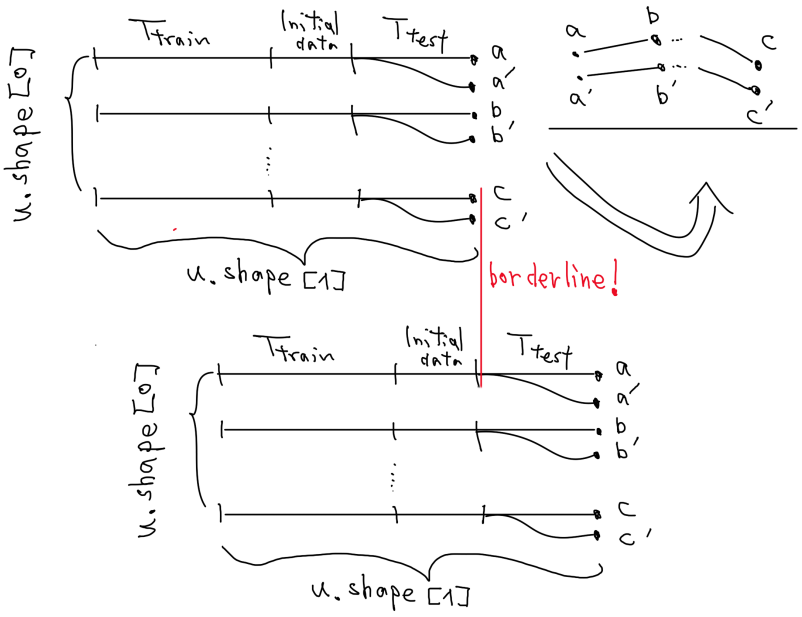

Figure 16: conceptual diagram of the learning scheme

10.2 Training phase

Using the least squares method, we select W^{out} that minimizes the following E.

$$E = \frac{1}{2} \sum_{j=0}^{u.shape[0]-1} \sum_{t=0}^{T_{train}-1} |U_{j,t} - W^{out} x_t|^2 + \frac{\beta}{2} \sum_{i=0}^{dim-1} \sum_{j=0}^{N_x-1} |W_{ij}^{out}|^2$$

Here, $\beta > 0$ is a regularization parameter to prevent overfitting, and T_{train} is the number of training data points. Due to convexity, this W^{out} can be solved analytically. A well-known expression exists, but details are omitted [2] .

10.3 Model selection phase

After selecting W^{out}, we generate forecast data using the recurrence relation. Here, we assume we are forecasting T_{test} steps ahead. First, the initial value for model selection is $U_{j,T_{train}}$. It is also important that $x_{j,T_{train}}$ is already obtained

[2]Recently, I employ Ohkubo-Inubushi's excellent method [12].

during the training phase (a sort of online). Based on these, the next step is estimated as follows:

$$\begin{cases} x_{j,T_{train}+1} &= (1-\alpha)x_{j,T_{train}} + \alpha\tanh(W^{in}U_{j,T_{train}} + Wx_{j,T_{train}}), \\ \widehat{U}_{j,T_{train}+1} &= W^{out}x_{j,T_{train}+1}. \end{cases}$$

The resulting $\widehat{U}_{j,t}, x_{j,t}$ are then sequentially fed into the following recurrence relation to generate the predicted data:

$$\begin{cases} x_{j,t+1} &= (1-\alpha)x_{j,t} + \alpha\tanh(W^{in}\widehat{U}_{j,t} + Wx_{j,t}), \\ \widehat{U}_{j,t+1} &= W^{out}x_{j,t+1} \end{cases}$$

for $t = T_{train}+1, \cdots, T_{train}+T_{test}$. Then, we implement **Bayesian learning (Optuna)** using the prediction data $\widehat{U}_{j,T_{train}+T_{test}}$. More specifically, we minimize the following MAE-based evaluation function:

$$\frac{1}{u.shape[0]}\sum_{j=0}^{u.shape[0]-1}|\widehat{U}_{j,T_{train}+T_{test}} - U_{j,T_{train}+T_{test}}|$$

in the selection of the following parameters:

$$\text{dim, lag, } \alpha\text{, } W\text{, } W^{in}.$$

10.4 Generalization Performance Evaluation Phase

After selecting dim, lag, α, W, and W^{in} in the previous section, we implement a new test to evaluate generalization performance. This process is largely the same as the model selection phase, but since the parameters differ slightly, we describe it again. First is the retraining phase. Using the least squares method again, we re-select W^{out} that minimizes the following E:

$$E = \frac{1}{2}\sum_{j=0}^{u.shape[0]-1}\sum_{t=T_{test}}^{T_{train}+T_{test}-1}|U_{j,t} - W^{out}x_{j,t}|^2 + \frac{\beta}{2}\sum_{i=0}^{dim-1}\sum_{j=0}^{N_x-1}|W_{ij}^{out}|^2.$$

Then, we evaluate the generalization performance from $T_{train}+T_{test}$. The initial value is $U_{j,T_{train}+T_{test}}$. It is also important that $x_{j,T_{train}+T_{test}}$ is already obtained during the retraining phase (a sort of online). Based on these, we estimate the next step as follows.

$$\begin{cases} x_{j,T_{train}+T_{test}+1} &= (1-\alpha)x_{j,T_{train}+T_{test}} \\ & \quad +\alpha\tanh(W^{in}U_{j,T_{train}+T_{test}} + Wx_{j,T_{train}+T_{test}}), \\ \widehat{U}_{j,T_{train}+T_{test}+1} &= W^{out}x_{j,T_{train}+T_{test}+1}. \end{cases}$$

The $(\widehat{U}_{j,t}, x_{j,t})$ obtained in this manner are then sequentially fed into the following recurrence relation to generate the prediction data:

$$\begin{cases} x_{j,t+1} &= (1-\alpha)x_{j,t} + \alpha\tanh(W^{in}\widehat{U}_{j,t} + Wx_{j,t}), \\ \widehat{U}_{j,t+1} &= W^{out}x_{j,t+1} \end{cases}$$

for $t = T_{train} + T_{test} + 1, \cdots, T_{train} + 2T_{test}$. Then, finally, the generalization performance is evaluated using the following MAE:

$$\frac{1}{u.shape[0]} \sum_{j=0}^{u.shape[0]-1} |\widehat{U}_{j,T_{train}+2T_{test}} - U_{j,T_{train}+2T_{test}}|.$$

11 Machine learning results

11.1 Wind speed in the Tokyo region

First, we applied this machine learning scheme to predict wind speed three hours ahead in the Tokyo region. The following is a comparable result from our study:

William Y.Y. Cheng and W. James Steenburgh, Evaluation of Surface Sensible Weather Forecasts by the WRF and the ETA Models over the Western United States, WRF/MM5 Users' Workshop - June 2005.

Although the data are not from the same location or period, the MAE for the 3-hour-ahead wind speed prediction may still be comparable.

- start date = "2024-Jan-01" end date = "2024-Feb-15"
- predict 3 hours ahead ($T_{test} = 3$)
- most recent 1,000 hours of data are used to train ($T_{train} = 1000$)
- $dim = 4$, $lag = 1$: fixed
- Thirty independent trials were performed using a rolling-window validation scheme.

Our result MAE: 2.01 m/s (data-driven filter [6] is applied: correlation 0.83)

Comparable result Cheng and Steenburgh (2005), using CIRP and WRF physical models: MAE 1.6 $\sim$ 1.9 m/s, but unclear whether a filter is applied...

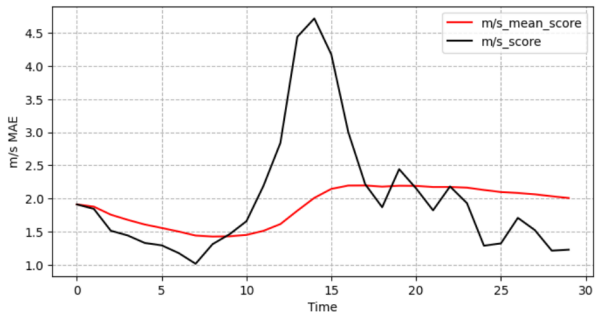

Figure 17: Our result: wind speed in the Tokyo region

11.2 Stock S&P500

We also applied this machine learning scheme to predict S&P500 that only the most recent 60 days of data are used to train. With this limited training data, online learning can be performed to predict whether the price of S&P500 will rise or fall tomorrow. The following is a comparable result from our study:

Gonzalo López Gil, Paul Duhamel-Sebline, Andrew McCarren, An Evaluation of Deep Learning Models for Stock Market Trend Prediction (2024)
https://arxiv.org/pdf/2408.12408v1

- start date = "2024-Jan-01" end date = "2025-May-10"
- predict 1 day ahead ($T_{test} = 1$)
- most recent 60 days of data are used to train ($T_{train} = 60$)
- $dim = 2$, $lag = 1$: fixed
- Thirty independent trials were performed using a rolling-window validation scheme.

Our result Accuracy 0.762, F1 0.712, Recall 0.721 (data-driven filter [6] is applied: correlation 0.985)

Comparable result Gil, Duhamel-Sebline, McCarren (2024), using xLSTM-TS: Accuracy 0.709, F1 0.730, Recall 0.768, wavelet filter is applied.

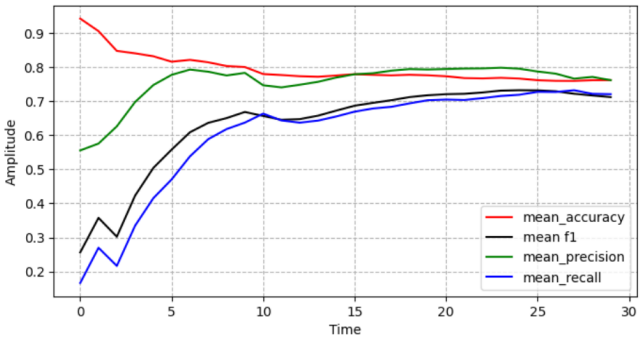

Figure 18: Our result: S&P500

Acknowledgments. I am grateful to Pritpal Matharu for his valuable comments, which helped improve this note. Research of TY was partly supported by the JSPS Grants-in-Aid for Scientific Research 24H00186. This is the lecture notes for the Oberwolfach Seminar "Phase Transitions and Turbulence Including Fluids," held on 19–24 October 2025.

References

[1] P. Constantin, W. E and E. S. Titi, Onsager's conjecture on the energy conservation for solutions of Euler's equation. Comm. Math. Phys. 165 (1994), 207–209.

[2] C. De Lellis and L. Székelyhidi, Dissipative Euler flows and Onsager's conjecture J. Eur. Math. Soc. (JEMS) 16 (2014), 1467–1505.

[3] J. Duchon and R. Robert, Inertial energy dissipation for weak solutions of incompressible Euler and Navier-Stokes equations. Nonlinearity 13 (2000), 249–255.

[4] G. L. Eyink, Multi-scale gradient expansion of the turbulent stress tensor, J. Fluid Mech., 549, (2006), 159–190.

[5] H. Jaeger, *The "echo state" approach to analysing and training recurrent neural networks-with an erratum note*, GMD Report, **148**, (2001).

[6] T. Jinno, T. Mitsui, K. Nakai, Y. Saiki and T. Yoneda, Long-term prediction of El Nino-Southern Oscillation using reservoir computing with data-driven realtime filter, Chaos: An Interdisciplinary J. Nonlinear Sci., 35 (2025) 053149.

[7] S. Kida and H. Miura, Swirl condition in low-pressure vortex. J. Phys. Soc. Jpn 67, (1998), 2166–2169.

[8] R. Lam , A. Sanchez-Gonzalez, M. Willson, P. Wirnsberger, M. Fortunato, F. Alet, S. Ravuri, T. Ewalds, Z. Eaton-Rosen, W. Hu, A. Merose, S. Hoyer, G. Holland, O. Vinyals, J. Stott, A. Pritzel, S. Mohamed and P. Battaglia, Learning skillful medium-range global weather forecasting, Science, 382, (2023), 1416-1421.

[9] H. Miura and S. Kida, Identification of tubular vortices in turbulence. J. Phys. Soc. Jpn 66, (1997), 1331–1334.

[10] K. Nakai and Y. Saiki, *Machine-learning inference of fluid variables from data using reservoir computing*, Phys. Rev. E, **98**, (2018) 023111.

[11] K. Nakai and Y. Saiki, *Machine-learning construction of a model for a macroscopic fluid variable using the delay-coordinate of a scalar observable*, Discr. Conti. Dyn. Sys.-S, **14**, (2021) 1079-1092.

[12] A. Ohkubo and M. Inubushi, *Reservoir computing with generalized readout based on generalized synchronization*, Scientific Reports, **14** (2024) 30918.

[13] T. Tsuruhashi, S. Goto, S. Oka and T. Yoneda, Self-similar hierarchy of coherent tubular vortices in turbulence, Phil. Trans. R. Soc., A, 380 (2022), 20210053.

[14] T. Yoneda, S. Goto and T. Tsuruhashi, Mathematical reformulation of the Kolmogorov-Richardson energy cascade in terms of vortex stretching, Nonlinearity 35 (2022), 1380–1401.

[15] T. Yoneda, 数理流体力学への招待, SGC ライブラリ（サイエンス社）2020 年 1 月

[16] T. Yoneda, 微積分から学ぶ深層ニューラルネットワークの各点収束構造, 不動点の世界，数理科学・2024 年 8 月号（サイエンス社）