Stochastic approach to model reduction in computational fluid mechanics. Application to wind time series

Mireille Bossy
Calisto team, Inria d'Université Côte d'Azur

Based on two joint works with
Kerlyns Martínez Rodríguez (University of Valparaíso)
and Jean-François Jabir (HSE University, Moscow)

CALISTO

Mascot-Num 2023
Le Croisic

Inria

Announcement

The 2024 annual meeting of MASCOT-NUM will be held in the are a of Nice-Sophia Antipolis, and will be organized by
Inria, Université Côte d'Azur LJAD, I3S, EURECOM & Mines ParisTech Sophia.

Save the date!

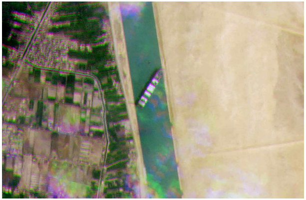
Local wind *variability* estimation is relevant in many situations

- For the risk of fatigue evaluation

  Pictures from vestas turbines, hub height of 140 m; wind speeds of about 14-18 m/s (ref: from a post on youtube)

- Air quality measurement uncertainty

- Need refined short term prediction of wind gust



the Ever-Given into the Suez canal

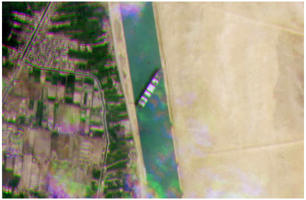
Local wind *variability* estimation is relevant in many situations

- For the risk of fatigue evaluation

  Pictures from vestas turbines, hub height of 140 m; wind speeds of about 14-18 m/s (ref: from a post on youtube)

- Air quality measurement uncertainty

- Need refined short term prediction of wind gust

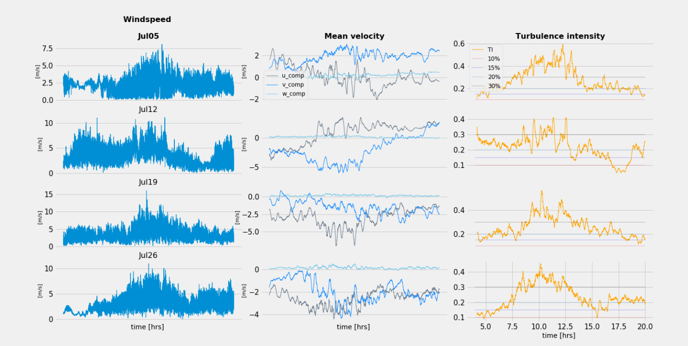


the Ever-Given into the Suez canal

- Common features: **near wall turbulence; presence of measurement points.**

Local wind as a time series – prediction issues

Wind velocity vector measured at a point, at discrete time
(with a frequency range from 1 Hz to 50 Hz or more; here t is incremented each 0.1 s.).



Measurements from a 30m height mast with sonic anemometer.

(SIRTA Observation platform, Palaiseau, France)

[Haeffelin et al., 2005].

$\|U_t^{\text{obs}}\|$
Instantaneous wind speed

$$\langle U_t^{\text{obs},i} \rangle = \frac{1}{\zeta} \sum_{t-\zeta \le s < t} U_s^{\text{obs},i}$$

Mean

$$I_t = \frac{\sqrt{\langle \|U_t^{\text{obs}} - \langle U_t^{\text{obs}} \rangle\|^2 \rangle}}{\sqrt{3}\, \|\langle U_{(\text{day})}^{\text{obs}} \rangle\|}$$

Turbulent intensity

$\langle U_t^{\text{obs}} \rangle$ is commonly compute by an average in time over an interval of 10 minutes to 60 minutes, corresponding to a minimum in the wind power spectral density.

Here the mean and intensity are plotted with the time-window $\zeta = 40$ minutes.

Several scales and methods

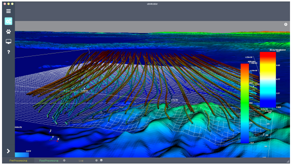
Persistence – Naive Predictor	$U(t+k) = U(t)$	very short term (seconds to 30 minutes)
Physical Approach	Global Forecasting, WRF, ...	for long term (one day to one week)
Statistical Approaches	ANN, TS-models	for short term (30 minutes to 6 hours)
Hybrid Structures	NWP + ANN, ...	medium and long term (6 hours to1 week)

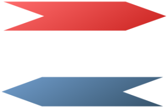
[Soman et al., 2010, Chang, 2014, Hanifi et al., 2020].

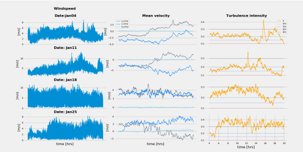
A double goal

(1) Propose a Times-series approach, based on SDEs derived from well established physical approaches (that are all including turbulence modelling) to predict the short term distribution of the turbulent velocity.

(2) Use wind observation as experiments allowing to quantify the uncertainty on supposed well known turbulence modelling parameters.





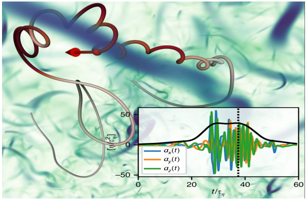


Fluid particles

taking the perspective of a 'air parcel', and given the flow field $\mathscr{U}(t, x)$, we consider parcel's state variables (x_f, U_f)

$$\frac{dx_f}{dt}(t) = U_f(t),$$
$$U_f(t) = \mathscr{U}(t, x_f(t)).$$

But how to get $\mathscr{U}(t, x_f(t))$?



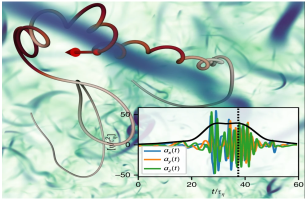
tracers trajectories in turbulence (borrowed from [Bentkamp et al., 2019]).

Fluid particles

taking the perspective of a 'air parcel', and given the flow field $\mathscr{U}(t,x)$, we consider parcel's state variables (x_f, U_f)

$$\frac{dx_f}{dt}(t) = U_f(t),$$
$$U_f(t) = \mathscr{U}(t, x_f(t)).$$

But how to get $\mathscr{U}(t, x_f(t))$?



tracers trajectories in turbulence
(borrowed from [Bentkamp et al., 2019]).

2023 is the bicentenary of Navier's work that led to the establishment of the master equations of fluid mechanics, known as the **Navier-Stokes equations**that governing $\mathscr{U}(t,x)$

$$\partial_t \mathscr{U}^{(i)} + \mathscr{U}^{(j)} \partial_{x_j} \mathscr{U}^{(i)} = \nu \triangle \mathscr{U}^{(i)} - \frac{1}{\varrho} \partial_{x_i} \mathscr{P}$$
$$\partial_{x_i} \mathscr{U}^{(i)} = 0$$

- **Direct Numerical Simulation** are from very expensive to totally prohibitive, as it requires a mesh below the Kolmogorov length scale η_K in $[50\,\mu\mathrm{m}, 1\mathrm{mm}]$ for most of industrial or environmental flows.
- **Averaged Navier Stokes equations**

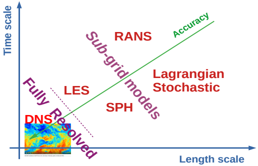
$$\mathscr{U}(t,x) \rightsquigarrow \langle \mathscr{U} \rangle (t,x) + \text{a model for the 2nd moments lost with the subscales}$$

Fluid particles

taking the perspective of a 'air parcel', and given the flow field $\mathscr{U}(t,x)$, we consider parcel's state variables (x_f, U_f)

$$\frac{dx_f}{dt}(t) = U_f(t),$$
$$U_f(t) = \mathscr{U}(t, x_f(t)).$$

But how to get $\mathscr{U}(t, x_f(t))$?



various averaged NS approaches for CFD.

- $\mathscr{U}(t,x) = \langle \mathscr{U} \rangle(t,x) +$ noise. Lagrangian modelling requires a model for the noise

$$\frac{dx_f}{dt}(t) = \langle \mathscr{U} \rangle_{\text{Ens, LES, RANS}}(t, x_f(t)) + u(t)$$

with $u(t)$ a random fluctuation of the Lagrangian mean velocity (Lagrangian Particle Dispersion Model (LPDM)).

Turbulent second order closure (see e.g. [Durbin and Speziale, 1994, Pope, 1994]).

Macroscopic random fluctuation, assuming decorrelation of time increments that lead to Gaussian fluctuation and 3D-Brownian motion B:

$$du^{(i)}(t) = -\frac{u^{(i)}(t)}{\tau_f}dt + (C_0\,\varepsilon)^{1/2} dB_t^{(i)}, \qquad \textit{(the simplest Langevin model)}$$

C_0 Kolmogorov constant and ε the dissipation rate of the mean kinetic energy required.

Stand alone Lagrangian stochastic model

Modelling consistency: the conditional mean field of air parcel velocity is the conditional expectation of its velocity

$$\langle U_f \rangle(t,x) = \mathbb{E}\Big[U_f(t)\, \underbrace{\big| x_f(t) = x \Big]}_{\text{conditionning}}$$

on $(\Omega, \mathscr{F}, \mathbb{P}, B)$, with (x_f, U_f) solution of a General Langevin Model:

$$dx_f(t) = U_f(t)\, dt,$$

$$dU_f^{(i)}(t) = -\partial_{x_i} \langle \mathscr{P} \rangle (t, x_f(t))dt + \left(G_{ij} \left(U_f^{(j)} - \langle U_f^{(j)} \rangle \right) \right) (t, x_f(t))dt + \sigma_{ij}(t, x_f(t))dB_t^{(i)}$$

B is a 3D-Brownian motion. (see e.g. [Durbin and Speziale, 1994, Pope, 2000, Minier and Peirano, 2001])

$$G_{ij} = -\frac{C_R}{2}\frac{\varepsilon}{\mathrm{k}}\delta_{ij} + C_2 \partial_j \langle U_f^{(i)} \rangle, \qquad \sigma_{ij} = \frac{2}{3}\left(C_R \varepsilon + C_2 \mathscr{P} - \varepsilon \right) \delta_{ij},$$

C_R is the Rotta constant, C_2 is the production isotropisation constant

Stand alone Lagrangian stochastic model

Modelling consistency: the conditional mean field of air parcel velocity is the conditional expectation of its velocity

$$\langle U_f \rangle(t,x) = \mathbb{E}\Big[U_f(t)\,\underbrace{|x_f(t) = x}_{\text{conditionning}}\Big]$$

on $(\Omega, \mathscr{F}, \mathbb{P}, B)$, with (x_f, U_f) solution of a General Langevin Model:

$$dx_f(t) = U_f(t)\,dt,$$
$$dU_f^{(i)}(t) = -\partial_{x_i}\langle \mathscr{P} \rangle(t, x_f(t))dt + \left(G_{ij}\left(U_f^{(j)} - \langle U_f^{(j)} \rangle\right)\right)(t, x_f(t))dt + \sigma_{ij}(t, x_f(t))dB_t^{(i)}$$

B is a 3D-Brownian motion. (see e.g. [Durbin and Speziale, 1994, Pope, 2000, Minier and Peirano, 2001])

$$G_{ij} = -\frac{C_R}{2}\frac{\varepsilon}{\mathrm{k}}\delta_{ij} + C_2 \partial_j \langle U_f^{(i)} \rangle, \qquad \sigma_{ij} = \frac{2}{3}\left(C_R\varepsilon + C_2\mathscr{P} - \varepsilon\right)\delta_{ij},$$

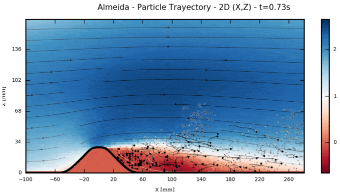
C_R is the Rotta constant, C_2 is the production isotropisation constant

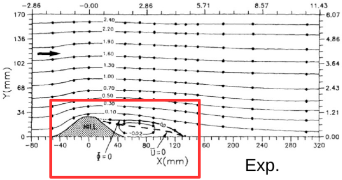
▶ The probabilistic model is a McKean Vlasov SDE, with

$$\langle \cdot \rangle(t,x) = \mathbb{E}[\,\cdot\,|x_f(t) = x]$$
$$u(t) = U_f(t) - \langle U_f \rangle(t, x_f(t)).$$

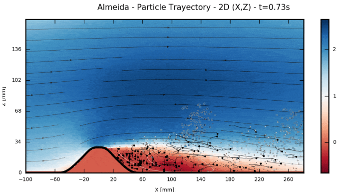
$\mathscr{P} = \frac{1}{2}\mathscr{P}_{ii}$, the turbulent production term $\mathscr{P}_{ij} := -(\langle u^{(i)}u^{(k)} \rangle \partial_k \langle U_f^{(i)} \rangle - \langle u^{(j)}u^{(k)} \rangle \partial_k \langle U_f^{(j)} \rangle$,
ε is closed with coherent parametrisation involving $\mathrm{k} = \frac{1}{2}\langle u^{(i)}u^{(i)} \rangle$.

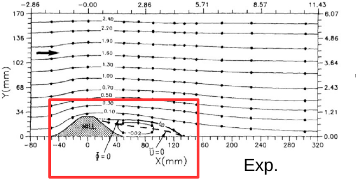
Numerical Stand alone Lagrangian stochastic model (in-house SDM code)





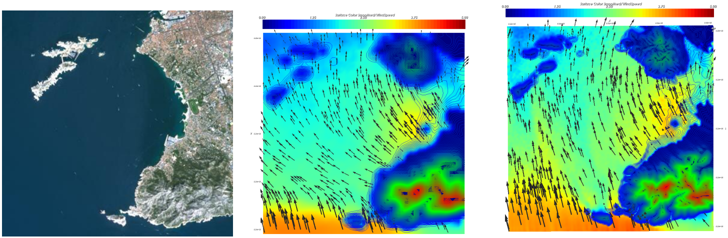
[Mokrani et al., 2019]





[Mokrani et al., 2019]

Very hight resolution simulation (downscaled from WRF)



Left: Numerical domain of the water body in WRF+SDM-WindPoS. Synchronised snapshot of the wind magnitude during the day on 24 April 2021 at the first height (10 m) of SDM-WindPoS. Middle: the subdomain is resolved to 150 m. Right: the resolution is 50 m.

Collaboration with SportRizer & Risk Weather Tech

From 3D+time Averaged Navier-Stokes equations to reduced 0D+time SDE

From the *Generalized Langevin Model* (GLM),

$$\begin{cases} dx_f^{(i)}(t) = U_f^{(i)}(t)dt,\ 1 \le i \le 3, \\ dU_f^{(i)}(t) = -\frac{1}{\rho}\partial_i \langle \mathscr{P} \rangle (t, x_f(t))\, dt + G_{ij}(t, x_f(t))\, \underbrace{\left(U_f^{(j)}(t) - \langle U_f^{(j)} \rangle (t, x_f(t))\right)}_{u(t) \text{ turb. velocity}}\, dt \\ \qquad\qquad + (C_0 \varepsilon)^{1/2}(t, x_f(t))\, dB_t^{(i)} \end{cases} \tag{1}$$

$\langle \mathscr{P} \rangle$ is the mean pressure

Fix $x_f(t) = x_{\mathrm{obs}}$,

$$u_t = U_f(t) - \langle U_f \rangle (t, x_{\mathrm{obs}}) \qquad \langle U_f^{(i)} \rangle (t, x_{\mathrm{obs}}) = \mathbb{E}[U_f^{(i)}(t) | x_f(t) = x_{\mathrm{obs}}],$$

Then the SDE for the instantaneous turbulent velocity $(u_t, t \ge 0)$ seen at x_{obs} is

$$du_t^{(i)} = G_{ij}(t, x_{\mathrm{obs}}) u_t^{(j)} dt + \sqrt{C_0(t, x_{\mathrm{obs}}) \varepsilon (t, x_{\mathrm{obs}})}\, dB_t^{(i)},$$

and its squared norm

$$d\|u_t\|^2 = -2u_t^{(j)} G_{ij}(t, x_{\mathrm{obs}}) u_t^{(j)} dt + 3(C_0\varepsilon)(t, x_{\mathrm{obs}}) dt + 2\sqrt{(C_0 \varepsilon)(t, x_{\mathrm{obs}})}\|u_t\|\, dW_t,$$

with the process $W_t = \sum_i \int_0^t \frac{u_s^{(i)}}{\|u_s\|}\, dB_s^{(i)}$ identifies as a one-dimensional Brownian motion.

Model hypothesis and turbulent closure

- We choose the isotropic (diagonal) structure for G_{ij}

$$G_{ij}(t,x) = -\frac{C_R}{2}\frac{\varepsilon}{\mathrm{k}}(t,x)\,\delta_{ij}, \qquad \text{where} \quad \mathrm{k}(t,x) = \tfrac{1}{2}\mathbb{E}[\|u_t\|^2 | x_f(t) = x] \tag{2}$$

corresponding to the Simplified Langevin model [Pope, 1985]. For consistency reason, C_0 is now $C_0 = \frac{2}{3}(C_R - 1)$.

Model hypothesis and turbulent closure

- We choose the isotropic (diagonal) structure for G_{ij}

$$G_{ij}(t,x) = -\frac{C_R}{2}\frac{\varepsilon}{\mathrm{k}}(t,x)\,\delta_{ij}, \qquad \text{where} \quad \mathrm{k}(t,x) = \tfrac{1}{2}\mathbb{E}[\|u_t\|^2 | x_f(t) = x] \tag{2}$$

corresponding to the Simplified Langevin model [Pope, 1985]. For consistency reason, C_0 is now $C_0 = \frac{2}{3}(C_R - 1)$.

- We choose a model for the dissipation rate of the kinetic energy $\varepsilon(t,x)$, classically used in the atmospheric boundary layer (ABL) :

$$\varepsilon(t,x) = \frac{C_\varepsilon}{\ell_{\mathrm{m}}}\mathrm{k}^{3/2}(t,x), \tag{3}$$

where C_ε is a constant, ℓ_{m} is a characteristic length scale called *m*ixing length. Near the ground, $\ell_{\mathrm{m}} = \kappa z$, where z denotes the distance to the wall from x, and κ is the Von Kármán constant [Cuxart et al., 2000, Drobinski et al., 2006].

Model hypothesis and turbulent closure

- We choose the isotropic (diagonal) structure for G_{ij}

$$G_{ij}(t,x) = -\frac{C_R}{2}\frac{\varepsilon}{\mathrm{k}}(t,x)\,\delta_{ij}, \qquad \text{where} \quad \mathrm{k}(t,x) = \tfrac{1}{2}\mathbb{E}[\|u_t\|^2 | x_f(t) = x] \tag{2}$$

  corresponding to the Simplified Langevin model [Pope, 1985]. For consistency reason, C_0 is now $C_0 = \frac{2}{3}(C_R - 1)$.
- We choose a model for the dissipation rate of the kinetic energy $\varepsilon(t,x)$, classically used in the atmospheric boundary layer (ABL) :

$$\varepsilon(t,x) = \frac{C_\varepsilon}{\ell_\mathrm{m}}\mathrm{k}^{3/2}(t,x), \tag{3}$$

  where C_ε is a constant, ℓ_m is a characteristic length scale called *m*ixing length. Near the ground, $\ell_\mathrm{m} = \kappa z$, where z denotes the distance to the wall from x, and κ is the Von Kármán constant [Cuxart et al., 2000, Drobinski et al., 2006].

C_0, C_R might vary according to the model and context : $C_R \in [1.5, 5]$, implying $C_0 \in [\frac{1}{3}, \frac{8}{3}]$ ⚠

We set $C_\alpha = \frac{C_\varepsilon}{\kappa z(x_{\text{obs}})}$, with $z(x_{\text{obs}}) = 30\text{m}$.

Von Kármán constant $\kappa \in [0.287, 0.615]$ [Edeling et al., 2014].

$C_\mu \in [0.054, 0.135]$ [Edeling et al., 2014], for $C_\varepsilon = C_\mu^{3/4}$.

⚠ We expect the values of C_α to be within the interval $[0.0061,\ 0.0259]$.

The reduced 0D+time SDEs

At x_{obs}, with notation $q_t = \|u_t\|^2$, such that $\mathrm{k}(t, x_{\text{obs}}) = \frac{1}{2}\mathbb{E}[q_t]$,
incorporating both $G_{ij} = -\frac{C_R}{2}\frac{\varepsilon}{\mathrm{k}}\delta_{ij}$ and the parametrisation of the dissipation $\varepsilon = C_\alpha \mathrm{k}^{\frac{3}{2}}$,
we end up with our physical-based model obtaining the following **CIR-type stochastic mean-field TKE model**:

$$dq_t = \gamma dt - C_R \frac{C_\alpha}{\sqrt{2}} q_t (\mathbb{E}[q_t])^{1/2} dt + 3C_0 \frac{C_\alpha}{2\sqrt{2}} (\mathbb{E}[q_t])^{3/2} dt + \sqrt{\sqrt{2} C_0 C_\alpha}\, (\mathbb{E}[q_t])^{3/4} \sqrt{q_t} dW_t, \tag{4}$$

$C_0 = \frac{2}{3}(C_R - 1)$; q_0 given.

The reduced 0D+time SDEs

At x_{obs}, with notation $q_t = \|u_t\|^2$, such that $\mathrm{k}(t, x_{\text{obs}}) = \frac{1}{2}\mathbb{E}[q_t]$,
incorporating both $G_{ij} = -\frac{C_R}{2}\frac{\varepsilon}{\mathrm{k}}\delta_{ij}$ and the parametrisation of the dissipation $\varepsilon = C_\alpha \mathrm{k}^{\frac{3}{2}}$,
we end up with our physical-based model obtaining the following **CIR-type stochastic mean-field TKE model**:

$$dq_t = \gamma dt - C_R \frac{C_\alpha}{\sqrt{2}} q_t (\mathbb{E}[q_t])^{1/2} dt + 3C_0 \frac{C_\alpha}{2\sqrt{2}} (\mathbb{E}[q_t])^{3/2} dt + \sqrt{\sqrt{2} C_0 C_\alpha}\, (\mathbb{E}[q_t])^{3/4} \sqrt{q_t}\, dW_t, \tag{4}$$

$C_0 = \frac{2}{3}(C_R - 1)$; q_0 given.

[Bossy et al., 2022]

(1) For positive parameters C_α, C_0, and γ, pathwise wellposedness holds for $(q_t;\ t \geq 0)$ solution of the McKean-Vlasov SDE (4).

(2) $\underline{C}\, \mathbb{1}_{\gamma > 0} \leq \sup_{t \geq 0} \mathbb{E}[q_t^p] \leq \overline{C}$, for all $p \geq 1$, for some constants $\overline{C}, \underline{C} > 0$.

(3) For $p = 1$,

$$q_0 \wedge \left(\frac{\sqrt{2}\gamma}{C_\alpha}\right)^{2/3} \leq \sup_{t \geq 0} \mathbb{E}[q_t] \leq q_0 \vee \left(\frac{\sqrt{2}\gamma}{C_\alpha}\right)^{2/3},$$

with long-time behaviour

$$\lim_{t \to +\infty} \mathbb{E}[q_t] = \left(\frac{\sqrt{2}\gamma}{C_\alpha}\right)^{2/3}.$$

using wellposedness condition for CIR processes with time dependant coefficients [Maghsoodi, 1996].

The calibration problem

A two steps strategy

1. Construct first a simple maximum likelihood estimator to recover the model constants from the data.
2. Next introduce a Bayesian estimation to quantify the uncertainty on theses parameters.

Difficulty : McKean Vlasov form of the model.

Two linearization tentatives :

- $\mathbb{E}[q_t] \rightsquigarrow q_t$

$$dq_t = \gamma dt - A q_t^{3/2} dt + B {q_t}^{5/4} dW_t$$

- $\mathbb{E}[q_t] \rightsquigarrow \mathbb{E}[q_\infty]$, explicitly known, leading to

$$dq_t = \Theta\left(\mu - q_t\right) dt + \sigma \sqrt{q_t} dW_t,$$

Linearisation $\mathbb{E}[q_t] \longrightarrow q_t$: Wellposedness and control of moments

$$dq_t = \gamma dt - \frac{C_\alpha}{\sqrt{2}} q_t^{3/2} dt + \sqrt{\sqrt{2} C_0 C_\alpha q_t}^{5/4} dW_t, \quad q_0 = |u_0'|^2$$

SDEs with superlinear growth coefficients :

$$dY_t = \gamma dt - B\, Y_t^{2\alpha-1} dt + \sigma Y_t^\alpha dW_t,\ Y_0 = y_0 > 0, \quad \text{with } \alpha > 1. \tag{5}$$

[Bossy et al., 2022]

Assume $\gamma, B \geq 0$. Then there exists a unique (strictly) positive strong solution Y to the SDE (5).

Moments : For p such that $0 \leq 2p \leq 1 + \frac{2B}{\sigma^2}$, $\quad \sup_{t\in[0,T]} \mathbb{E}\big[Y_t^{2p}\big] < +\infty$.

Exponential moments : for all $\mu \leq C(\alpha,\gamma)$, $\sup_{t\in[0,T]} \mathbb{E}\big[\exp\{\mu \int_0^t Y_s^{2\alpha-2} ds\}\big] < +\infty$.

⚠ Here $\frac{2B}{\sigma^2} = \frac{1}{C_0}$, so only $C_0 \in [\frac{1}{3}, 1]$ ensures that $\mathbb{E}[q_t^2]$ is finite.

- The (linear) model behave in time like the McKean Vlasov model (when $\gamma = 0$)
  $\lim_{t\to+\infty} \mathbb{E}[q_t] = 0, \quad \lim_{t\to+\infty} t^2 \mathbb{E}[q_t] = \frac{4}{C_\alpha^2 (C_0^2 + 3C_0 + 2)}$.
- The Euler scheme $q_{t_{n+1}} = q_{t_n} + B|q_{t_n}|^{3/2} + \sigma |q_{t_n}|^{5/4} (W_{t_{n+1}} - W_{t_n})$ is almost surely converging [Gyongy, 1998], allowing to propose a consistent Maximun Likelyhood estimator for (C_α, C_0, γ).

⚠ Falling in case of strong L^1 divergence of the Euler scheme [Hutzenthaler et al., 2010] for SDEs with super-linear growth condition.

Linearisation $\mathbb{E}[q_t] \longrightarrow \mathbb{E}[q_\infty] = \left(\frac{\sqrt{2}\gamma}{C_\alpha}\right)^{2/3}$

CIR model for the instantaneous TKE:

$$dq_t = \Theta(C_\alpha,\gamma)\,(\mu(C_\alpha,\gamma) - q_t)\,dt + \sigma(\gamma)\sqrt{q_t}dW_t, \quad q_0 \text{ given, and } \gamma > 0, \tag{6}$$

where $\Theta(C_\alpha,\gamma) = C_R(\frac{C_\alpha^2\gamma}{2})^{1/3}, \qquad \mu(C_\alpha,\gamma) = (\sqrt{2}\frac{\gamma}{C_\alpha})^{2/3}, \qquad \sigma(\gamma) = \sqrt{2C_0\gamma}.$

with physical parameters : $(\gamma, C_0, (C_R), C_\alpha)$

$\mathbb{P}(\inf\{t \geq 0, q_t = 0\} = +\infty) = 1 \Leftarrow 2\Theta(C_\alpha,\gamma)\mu(C_\alpha,\gamma) \geq \sigma^2(\gamma) \Leftarrow C_R \geq C_0$ structurally always satisfied.

CIR model for the instantaneous TKE:

$$dq_t = \Theta(C_\alpha,\gamma)\,(\mu(C_\alpha,\gamma) - q_t)\,dt + \sigma(\gamma)\sqrt{q_t}dW_t, \quad q_0 \text{ given, and } \gamma > 0, \tag{6}$$

where $\Theta(C_\alpha,\gamma) = C_R(\frac{C_\alpha^2\gamma}{2})^{1/3}, \qquad \mu(C_\alpha,\gamma) = (\sqrt{2}\frac{\gamma}{C_\alpha})^{2/3}, \qquad \sigma(\gamma) = \sqrt{2C_0\gamma}.$

with physical parameters : $(\gamma, C_0, (C_R), C_\alpha)$

$\mathbb{P}(\inf\{t \geq 0, q_t = 0\} = +\infty) = 1 \Leftarrow 2\Theta(C_\alpha,\gamma)\mu(C_\alpha,\gamma) \geq \sigma^2(\gamma) \Leftarrow C_R \geq C_0$ structurally always satisfied.

Symmetrized Euler scheme : for SDE (6), $t_n = n\Delta t$

$$\widehat{q}_{t_{n+1}} = |\widehat{q}_{t_n} + \Theta(C_\alpha,\gamma)\,(\mu(C_\alpha,\gamma) - \widehat{q}_{t_n})\,\Delta t + \sigma(\gamma)\sqrt{\widehat{q}_{t_n}}\,(W_{t_{n+1}} - W_{t_n})\,| \tag{7}$$

Equivalently

$$\widehat{q}_{t_{n+1}} \;\sim\; |\mathcal{N}\left(\widehat{q}_{t_n} + \Theta(C_\alpha,\gamma)(\mu(C_\alpha,\gamma) - \widehat{q}_{t_n})\Delta t,\; \sigma^2(\gamma)\widehat{q}_{t_n}\Delta t\right)|.$$

Under the assumption $C_0 < 2$, the scheme (7) converges in law with a rate one [Bossy and Diop, 2010].

We fix $C_0 = 1.9$ (in accordance with the literature).

$(\widehat{C}_\alpha, \widehat{\gamma})$ estimators

Considering the compact set $D \subset \mathbb{R}^+ \times \mathbb{R}^+$ supporting the admissible values of $\theta = (C_\alpha, \gamma)$, choosing Δt according to some data frequency, we compute the pseudo-maximum likelihood estimator

$$\widehat{\theta} = \arg\max_{\theta \in D} \ \log p^{\theta}_{\Delta t}(q_0^{\text{obs}}, \ldots, q_{t_N}^{\text{obs}}),$$

allowing for an explicit solution of the optimal pair $(\widehat{C}_\alpha, \widehat{\gamma})$

Define empirical moments

$$\widehat{M}_{m_1,m_2} = \frac{1}{N} \sum_{n=0}^{N-1} (q_{t_{n+1}}^{\text{obs}} - |q_{t_n}^{\text{obs}}|)^{m_1} |q_{t_n}^{\text{obs}}|^{m_2}. \tag{8}$$

Quadratic variation estimator for γ:

$$\widehat{\gamma} = \frac{\widehat{M}_{2,0}}{2C_0 \Delta t \widehat{M}_{0,1}}, \tag{9}$$

Pseudo-maximum likelihood estimator of C_α:

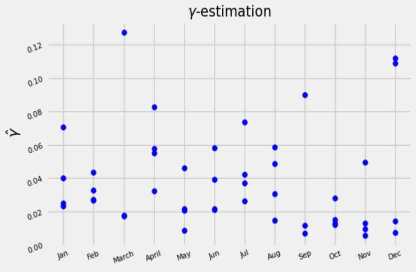
$$\widehat{C}_\alpha = \frac{\sqrt{2}}{\sqrt{\widehat{\gamma}}} \max\left\{ c_* \frac{\sqrt{\widehat{\gamma}}}{\sqrt{2}}, \left(\frac{\max\{\widehat{\gamma}\Delta t C_R - \widehat{M}_{1,0}\ , 0\}}{\widehat{M}_{0,1} \Delta t C_R} \right)^{3/2} \right\} \tag{10}$$

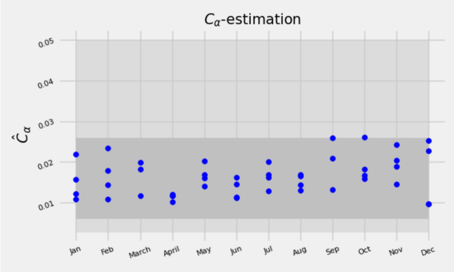
for some lower bound c_* to choose in $[0, 0.0061]$.

Step zero : Prior calibration with time depend regime

We end up with a family of estimators

$$\Sigma := \{(\widehat{C}_\alpha(d),\ \widehat{\gamma}(d)),\ \text{for } d \text{ in the selection of day-periods } \mathcal{S} \text{ in the year 2017}\},$$





Point-estimations for each Wednesday of 2017.
The dark grey area (right) is the reference interval compiling turbulence closure literature.

Priori distributions, defined as truncated Gaussian distributions:

$$\gamma \sim \mathcal{N}^+\left(\overline{\Gamma}(\mathcal{S}), \mathbb{V}_\Gamma(\mathcal{S})\right), \qquad C_\alpha \sim \mathcal{N}^+\left(\overline{C}(\mathcal{S}), \mathbb{V}_C(\mathcal{S})\right),$$

with $\overline{\Gamma}(\mathcal{S}), \overline{C}(\mathcal{S}), \mathbb{V}_\Gamma(\mathcal{S})$ and $\mathbb{V}_C(\mathcal{S})$ are the empirical means and variances over the days.

Step one: Posterior calibration

Goal : use Bayesian inference to extract precise information on the parameters distributions $\pi(\theta|\boldsymbol{q}^{\text{obs}})$ from the data $\boldsymbol{q}^{\text{obs}}$, with $\theta = (\gamma, C_\alpha)$,

$$\pi(\theta|\boldsymbol{q}^{\text{obs}}) = \frac{p(\boldsymbol{q}^{\text{obs}}|\theta)\, p_\theta(\theta)}{p(\boldsymbol{q}^{\text{obs}})}, \tag{11}$$

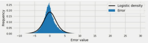
where

- $p(\cdot|\theta)$ the probability density of the model given the parameters (likelihood function, given)
- p_θ the prior distribution of θ (prior distribution, given),
- $p(\boldsymbol{q}^{\text{obs}})$ is the distribution of the observed data with

$$\boldsymbol{q}^{\text{obs}}(\theta) = \widehat{\boldsymbol{q}}(\theta) + \boldsymbol{\mathscr{E}}$$

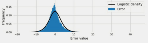
  - $\widehat{\boldsymbol{q}}(\theta)$ is the i.i.d random vector variable, with the equilibrium law (for a given θ) of the discrete-time model (7);
  - the random Error vector $\boldsymbol{\mathscr{E}} \sim$ centred logistic distribution and scale parameter to be estimated from the data. Choice made from a step0 analysis fitting the histogram of observation error distribution with a set of explicit mean-variance probability densities.

February 1st



February 8th

February 15th

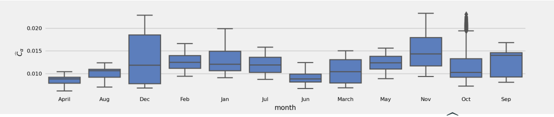


February 22th

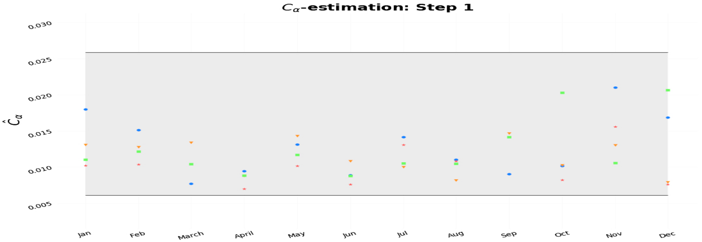
Method : Metropolis-Hasting algorithm and its Hamiltonian Monte Carlo (HMC) variant.

We have used the Python package PyMC3 [Salvatier et al., 2016] with the No U-Turn sampler[Gelman and Hoffman, 2014] method.

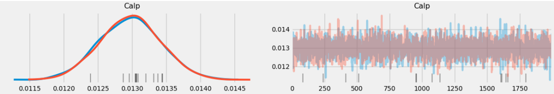
Posterior calibration on C_α



Box plots within month of posterior distribution of $\widehat{C}_\alpha$,
constructed from the Markov chains samples in a given month.



Posterior mean estimations of $\widehat{C}_\alpha(d)$ for each considered day

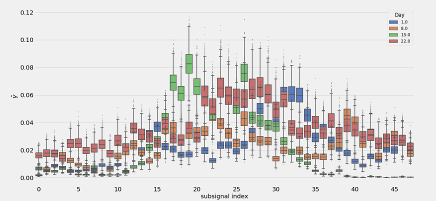


Two examples of Exploration of the state space with the Markov chain and posterior distribution

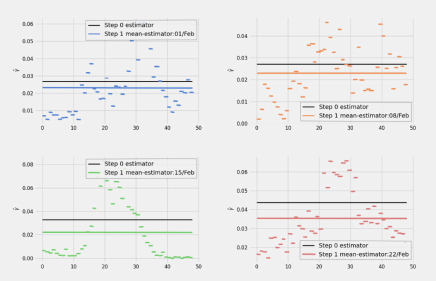
Posterior calibration on γ_t over days and over 20 minutes periods

Bayesian calibration of γ during February, 2017

Box plot of the $\gamma(d, i)$ for $i = 1, \ldots, 48$ for each 20 minutes-length sub-signal

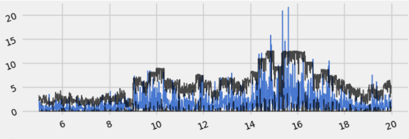


The $t \mapsto \bar{\gamma}_t$ obtained from Step one for the same four days; the horizontal lines are the level of the means over the period (the black line is the Step zero estimator).

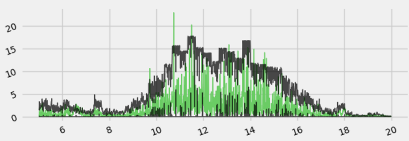


Validation of the calibration procedure against the observation

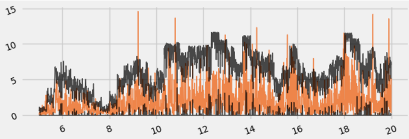
$\boldsymbol{q}^{\text{obs}}$ for $d =$ Feb 1st



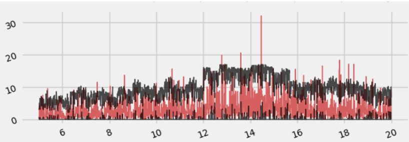
$\boldsymbol{q}^{\text{obs}}$ for $d =$ Feb 15



$\boldsymbol{q}^{\text{obs}}$ for $d =$ Feb 8th



$\boldsymbol{q}^{\text{obs}}$ for $d =$ Feb 22th



Instantaneous turbulent kinetic energy observed during February, 2017, between 5 a.m and 8 p.m (color plots) using the frequency of 1/30 s^{-1}

95% confidence interval (plotted in black) of (q_{t_n}, n) using the posterior mean values $\widehat{C}_\alpha(d)$ and the time dependent mean $\bar{\gamma}(t) = \sum_{i=0}^{S} \mathbb{E}[\gamma(d,i)]\mathbb{1}_{[T_i, T_{i+1}]}(t)$, with $\Delta t = 30s$.

Prediction of the confidence interval
The 10-minutes turbulence intensity as a substitute to the calibration of $(\bar{\gamma}_t)$

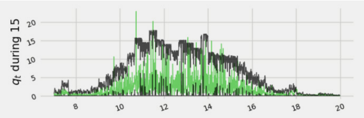
CIR model for the instantaneous TKE:

$$dq_t = \Theta(C_\alpha, \gamma(t))(\mu(C_\alpha, \gamma(t)) - q_t)\, dt + \sigma(\gamma(t))\sqrt{q_t} dW_t$$

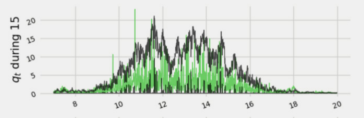
suggests that $\gamma(t) - \left(\frac{C_\alpha^2 \gamma(t)}{2}\right)^{1/3} \mathbb{E}[q_t] = 0$, from which we deduce the formal relation:

$$\gamma(t) = \frac{C_\alpha}{\sqrt{2}} \left(\sqrt{3} \|\langle U^{\text{obs}}_{(d)} \rangle\| I_t\right)^3 . \quad (12)$$

Putting the empirical $I_t = \sqrt{\langle \|U^{\text{obs}}_t - \langle U^{\text{obs}}_t \rangle\|^2 \rangle} / \sqrt{3} \|\langle U^{\text{obs}}_{(d)} \rangle\|$ in (q_{t_n}, n).



95% Confidence interval of (q_{t_n}, n) obtained using the posterior mean values $\widehat{C}_\alpha(d)$ and the time dependent mean $\bar{\gamma}(t) = \sum_{i=0}^{S} \mathbb{E}[\gamma(d,i)] \mathbb{1}_{[T_i, T_{i+1}]}(t)$, $\Delta t = 30s$.



Prediction of the 95% Confidence interval of (q_{t_n}, n) obtained by sampling the within-year posterior distribution of C_α and the time dependent mean $\bar{\gamma}(t)$ replaced by the turbulent intensity statistic through (12), $\Delta t = 30s$.

In green: the instantaneous turbulent kinetic energy observed during February 15th, 2017, between 7 a.m and 8 p.m using the frequency of 1/30 s.

As a conclusion (work in progress) : Gust modelling and intermitency

The wind gust speed $U_{\max}$ is defined as a short-duration maximum of the horizontal wind speed [Suomi and Vihma, 2018]

$$U_{\max}(\bar{t}) = \max\left\{U_f(s); s \in [\bar{t} - h^g, \bar{t}]\right\}.$$

(the choice of the gust duration h^g may vary with the activity sector).

Some predictive frameworks are ready to use, for example

$$U_{\max}(\bar{t}) = \langle U\rangle(\bar{t}) + \text{turbulent kinetic energy} \times g_x.$$

where, for a R level is fixed, the peak factor g_x, such that

$$\mathbb{P}(U_{\max}(\bar{t}) < g_x) = 1 - R$$

is given by some model formula (under assumption of stationary, and Gaussian behaviour of the acceleration) [Schreur and Geertsema, 2008].

Can we use the stochastic model to compute the probability of having a gust?

▶ within the CIR model

For a significance level R, we want to compute the probability

$$\mathbb{P}(q_t + \Delta t > g_x | q_t) = R$$

▶ Improvement of the model and noises

- Introduce acceleration :

$$dx_f(t) = U_f(t)\,dt,$$
$$dU_f^{(i)}(t) = -\partial_{x_i}\langle\mathscr{P}\rangle(t, x_f(t))dt + \left(G_{ij}\left(U_f^{(j)} - \langle U_f^{(j)}\rangle\right)\right)(t, x_f(t))dt + a(t)dt$$
$$da(t) = -\beta a(t)dt + \sigma_{ij}(t)dB_t^{(i)}$$

[Innocenti et al., 2020]

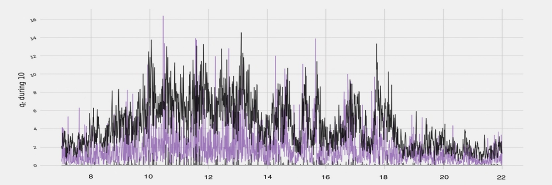
- Introduce intermittency : K62, the dissipation ε is a log-normal process.
  $\varepsilon_t = \mathbb{E}[\varepsilon_t]\exp(\sqrt{\lambda_I}X_t + \frac{\lambda_I}{2}\mathbb{E}[X_t^2])$
  with X_t
  - Ornstein-Uhlenbeck (K62)
  - Stochastic Voltera equation reaching log correlation [Letournel et al., 2021]

Some concluding remarks

Modelling of the distribution of the instantaneous wind speed for short term prediction purpose.

- Starting from the physical turbulent modelling, a 3d+time velocity field dynamics, we end up with several propositions of 0d+time stochastic models.
- The simplest proposition (CIR TKE)
  - allows to recover from measurements parameters values that fit the interval values given by the expert's knowledge, as a validation of the physical meaning of the model.
  - allows to give a good prediction of the 95% CI

    $\boldsymbol{q}^{\text{obs}}$ for $d = \texttt{November}$ 10th and its predicted 95% CI

- Still some need for stable explicit schemes for MLE prior calibration purpose
- Raising the still challenging question about MLE for McKean Vlasov
- Improving the modelling thorough uncertainty parameters inference.

Thank you for your attention !

[Bentkamp et al., 2019] Bentkamp, L., Lalescu, C. C., and Wilczek, M. (2019).
Persistent accelerations disentangle lagrangian turbulence.
Nature Communications.

[Bernardin et al., 2010] Bernardin, F., Bossy, M., Chauvin, C., Jabir, J., and Rousseau, A. (2010).
Stochastic Lagrangian Method for Downscaling Problems in Computational Fluid Dynamics.
ESAIM: M2AN, 44(5):885–920.

[Bossy and Diop, 2010] Bossy, M. and Diop, A. (2010).
Weak convergence analysis of the symmetrized Euler scheme for one dimensional SDEs with diffusion coefficient $|x|^\alpha, \alpha \in [\frac{1}{2}, 1)$.
Research report RR-5396, INRIA, arxiv:1508.04573.

[Bossy et al., 2018] Bossy, M., Dupré, A., Drobinski, P., Violeau, L., and Briard, C. (2018).
Stochastic Lagrangian approach for wind farm simulation.
In *Forecasting and Risk Management for Renewable Energy*, pages 45–71. Springer International Publishing.

[Bossy et al., 2013] Bossy, M., Fontbona, J., Jabin, P.-E., and Jabir, J.-F. (2013).
Local existence of analytical solutions to an incompressible lagrangian stochastic model in a periodic domain.
Communications in Partial Differential Equations, 38(7):1141–1182.

[Bossy et al., 2011] Bossy, M., Jabir, J., and Talay, D. (2011).
On conditional McKean Lagrangian stochastic models.
Probab. Theory Relat. Fields, 151:319–351.

[Bossy et al., 2022] Bossy, M., Jabir, J.-F., and Martínez Rodríguez, K. (2022).
Instantaneous turbulent kinetic energy modelling based on lagrangian stochastic approach in cfd and application to wind energy.
Journal of Computational Physics, page 110929.

[Chang, 2014] Chang, W. (2014).
A literature review of wind forecasting methods.
Journal of Power and Energy Engineering, 2(04):161–168.

[Cuxart et al., 2000] Cuxart, J., Bougeault, P., and Redelsperger, J. (2000).
A multiscale turbulence scheme apt for les and mesoscale modelling.
Quart. J. Roy. Meteor. Soc, 126:1–30.

[Drobinski et al., 2006] Drobinski, P., Redelsperger, J.-L., and Pietras, C. (2006).
Evaluation of a planetary boundary layer subgrid-scale model that accounts for near-surface turbulence anisotropy.
Geophysical research letters, 33(23).

[Durbin and Speziale, 1994] Durbin, P. A. and Speziale, C. G. (1994).
Realizability of second-moment closure via stochastic analysis.
Journal of Fluid Mechanics, 280:395–407.

[Edeling et al., 2014] Edeling, W., Cinnella, P., Dwight, R., and Bijl, H. (2014).
Bayesian estimates of parameter variability in the $k - \varepsilon$ turbulence model.
Journal of Computational Physics, 258(C):73–94.

[Gelman and Hoffman, 2014] Gelman, A. and Hoffman, M. (2014).
The No–U–Turn Sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo.
Journal of Machine Learning Research, 15:1593–1623.

[Gyongy, 1998] Gyongy, I. (1998).
A note on Euler's approximations.
Potential Anal., 8:205–216.

[Haeffelin et al., 2005] Haeffelin, M., Barthès, L., Bock, O., Boitel, C., Bony, S., Bouniol, D., Chepfer, H., Chiriaco, M., Cuesta, J., Delanoë, J., Drobinski, P., Dufresne, J.-L., Flamant, C., Grall, M., Hodzic, A., Hourdin, F., Lapouge, F., Lemaître, Y., Mathieu, A., Morille, Y., Naud, C., Noël, V., O'Hirok, W., Pelon, J., Pietras, C., Protat, A., Romand, B., Scialom, G., and Vautard, R. (2005).
SIRTA, a ground-based atmospheric observatory for cloud and aerosol research.
Annales Geophysicae, 23(2):253–275.

[Hanifi et al., 2020] Hanifi, S., Liu, X., Lin, Z., and Lotfian, S. (2020).
A critical review of wind power forecasting methods—past, present and future.
Energies, 13(15):3764.

[Hutzenthaler et al., 2010] Hutzenthaler, M., Jentzen, A., and Kloeden, P. (2010).
Strong and weak divergence in finite time of Euler's method for stochastic differential equations with non-globally Lipschitz continuous coefficients.
Proceedings of the Royal Society, 467:1563–1576.

[Innocenti et al., 2020] Innocenti, A., Mordant, N., Stelzenmuller, N., and Chibbaro, S. (2020).
Lagrangian stochastic modelling of acceleration in turbulent wall-bounded flows.
Journal of Fluid Mechanics, 892:A38.

[Letournel et al., 2021] Letournel, R., Goudenège, L., Zamansky, R., Vié, A., and Massot, M. (2021).
Reexamining the framework for intermittency in Lagrangian stochastic models for turbulent flows: A way to an original and versatile numerical approach.
Physical Review E, 104(1):015104.

[Maghsoodi, 1996] Maghsoodi, Y. (1996).
Solution of the extended CIR term structure and bond option valuation.
Mathematical Finance, 6(1):89–109.

[Minier and Peirano, 2001] Minier, J.-P. and Peirano, E. (2001).
The pdf approach to turbulent polydispersed two-phase flows.
Physics Reports, 352(1–3):1 – 214.

[Mokrani et al., 2019] Mokrani, C., Bossy, M., Di Iorio, M., and Rousseau, A. (2019).
Numerical Modelling of Hydrokinetic Turbines Immersed in Complex Topography using Non-Rotative Actuator Discs.
In *Three Years Promoting the Development of Marine Renewable Energy in Chile 2015 - 2018*. MERIC-Marine Energy and Innovation Center.

[Pope, 1985] Pope, S. (1985).
Pdf methods for turbulent reactive flows.
Progress in Energy and Combustion Science, 11(2):119 – 192.

[Pope, 1994] Pope, S. (1994).
Lagrangian pdf methods for turbulent flows.
Annu. Rev. Fluid Mech., 26:23–63.

[Pope, 2000] Pope, S. B. (2000).
***Turbulent flows*.**
Cambridge Univ. Press, Cambridge.

[Salvatier et al., 2016] Salvatier, J., Wiecki, T. V., and Fonnesbeck, C. (2016).
Probabilistic programming in Python using PyMC3.
PeerJ Computer Science, 2:e55.

[Schreur and Geertsema, 2008] Schreur, B. and Geertsema, G. (2008).
Theory for a tke based parameterization of wind gusts.

[Soman et al., 2010] Soman, S., Zareipour, H., Malik, O., and Mandal, P. (2010).
A review of wind power and wind speed forecasting methods with different time horizons.
In *North American Power Symposium 2010*, pages 1–8. IEEE.

[Suomi and Vihma, 2018] Suomi, I. and Vihma, T. (2018).
Wind gust measurement techniques-from traditional anemometry to new possibilities.
Sensors, 18(4 1300. 23).