Physical-based probabilistic models for the instantaneous turbulent kinetic energy:modelling, calibration and validation

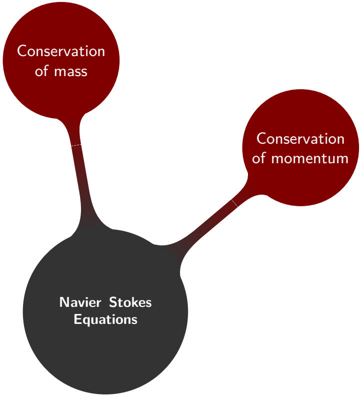
Kerlyns Martínez
Joint work with Mireille Bossy and Jean-François Jabir.

Seminar CALISTO.

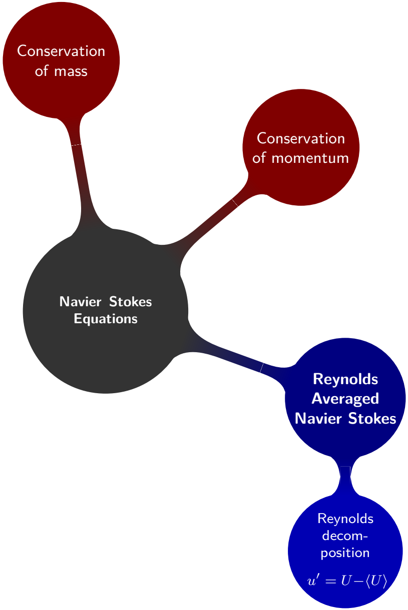
November 5th, 2020

Overview

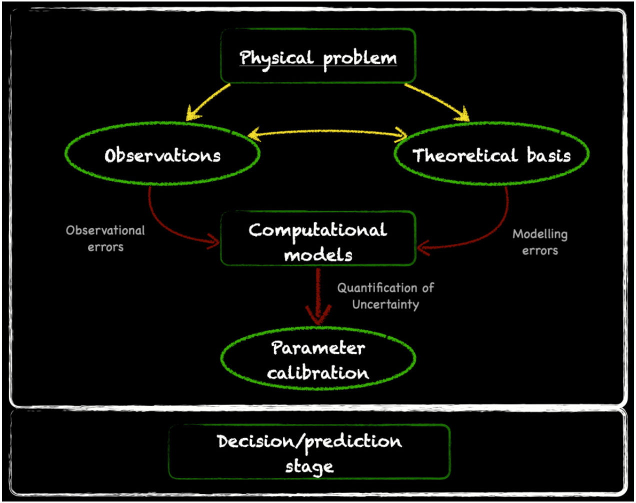
Introduction



Introduction

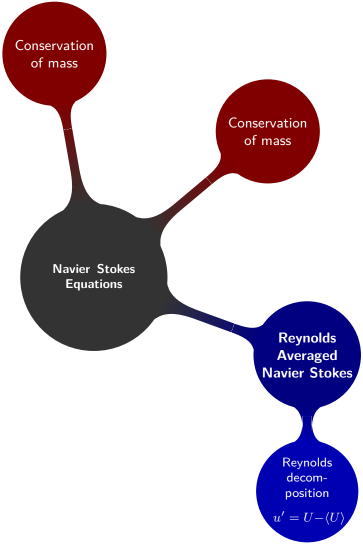


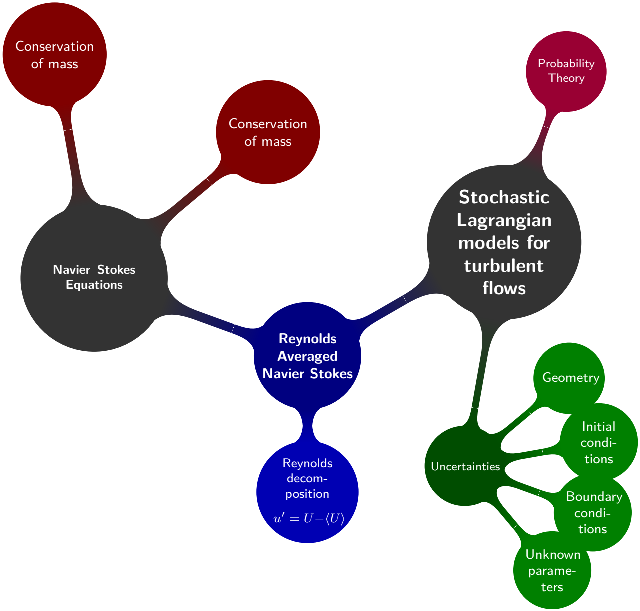
To produce accurate predictions of quantities of interest it is necessary to make a systematic treatment of uncertainties within the models and observations:



By means of a statistical model containing multiplicative inadequacy terms, Edeling and co-authors obtained estimates for the model error in RANS simulations based on Launder-Sharma turbulence closure relation, using a Bayesian calibration method employing measured boundary-layer velocity profiles.

The results suggest that the coefficients are not expected to be flow-independent.


Conservation of mass
Conservation of mass
Navier Stokes Equations
Reynolds Averaged Navier Stokes
Reynolds decom-position
$u' = U - \langle U \rangle$


Conservation of mass
Conservation of mass
Probability Theory
Stochastic Lagrangian models for turbulent flows
Navier Stokes Equations
Reynolds Averaged Navier Stokes
Reynolds decomposition
$u' = U - \langle U \rangle$
Uncertainties
Geometry
Initial conditions
Boundary conditions
Unknown parameters

Stochastic Lagrangian description of turbulent flows has been widely applied to complex industrial situations. For instance, in a meteorological context we have applications on the filtering of wind data, downscaling methods, simulation of fluid dynamics around windmills.

Following the ideas of Pope, we interpret the mean and the Reynolds stresses as the first and second moments of the Lagrangian PDF. Then, we follow fluid particles as they move through the flow considering the process $((X_t, U_t); 0 \leq t \leq T)$, and we write

$$\langle g(U) \rangle (t, x) \approx \mathbb{E}\left[g(U_t) \middle| \, X_t = x \right],$$

and quantities of interest can be identified.

In the case of stochastic Lagrangian models, the uncertainty quantification for model parameters arises naturally; and statistically reliable methods to estimate the parameters are available since (in principle) it is possible to construct the likelihood function (probability that the model, given the parameters, generates de observations).

Maximum likelihood estimators (MLE), pseudo-maximum likelihood estimators, quadratic variation estimators (QVE), Markov chain Monte Carlo (MCMC) methods, among others.

Data

The data used was obtained from the observation platform SIRTA. We have a family of time series containing different components of the wind, registered with a frequency of 10Hz during 24 hours spanned for all Wednesday of the year 2017. Is representative of the range of possible values for the temperature, degree of humidity, direction of the wind, intensity, and therefore a wide variety of wind profiles.

Wind measurements taken at a fixed point altitude of $x_{\mathtt{obs}} = 30[\mathtt{m}]$ with sonic anemometers.

Data

The data used was obtained from the observation platform SIRTA. We have a family of time series containing different components of the wind, registered with a frequency of 10Hz during 24 hours spanned for all Wednesday of the year 2017. Is representative of the range of possible values for the temperature, degree of humidity, direction of the wind, intensity, and therefore a wide variety of wind profiles.

Wind measurements taken at a fixed point altitude of $x_{\texttt{obs}} = 30[\mathtt{m}]$ with sonic anemometers.

We transform the set of velocity components into the observed instantaneous turbulent kinetic energy considering an ergodic mean (rather than the usual Monte Carlo approximation):

$$q_{t_k}^{\mathsf{obs}} \approx \sum_{i=1}^{3} \left(U_{t_k}^{i,obs} - \frac{1}{\zeta} \sum_{t_k - \zeta \leq s_k \leq t_k} U_{s_k}^{i,obs} \right)^2 ,$$

for ζ a time scale between 10 and 59 minutes.

An indicator that we could use to classify the observations in homogeneous regimes is the turbulence intensity (which is in fact connected with the wind production), defined as the quotient between the standard deviation of wind speed series and the mean velocity:

$$I_t := \frac{\sqrt{2k(t,x)}}{\sqrt{3}\|\langle U\rangle\|},$$

with k the TKE.

FIGURE: Observed wind speed (in blue), mean velocity (in blue scale) and turbulence intensity (in orange), measured during the period 4h - 20h.

An indicator that we could use to classify the observations in homogeneous regimes is the turbulence intensity (which is in fact connected with the wind production), defined as the quotient between the standard deviation of wind speed series and the mean velocity:

$$I_t := \frac{\sqrt{2k(t,x)}}{\sqrt{3}\|\langle U\rangle\|},$$

with k the TKE.

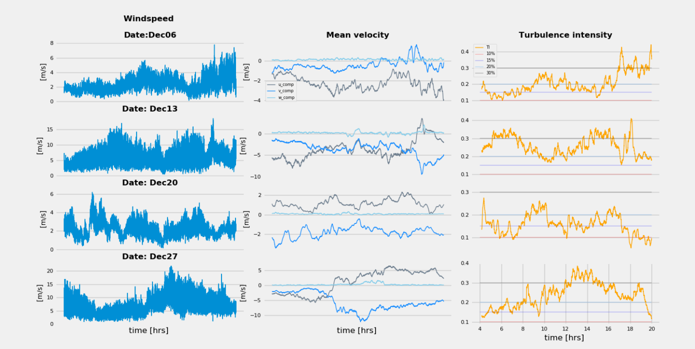


FIGURE: Observed wind speed (in blue), mean velocity (in blue scale) and turbulence intensity (in orange), measured during the period 4h - 20h.

In many physical applications we are interested in modelling characteristic components of the fluid at a fixed point in space.

- Baehr [2010]: construction of stochastic processes filtering corrupted measurements of a stochastic vector field along a random path. Method using a local model of the random medium to estimate locally the parameters and derive a dynamic model for mobile measurements signal. These ideas were applied to Doppler wind LIDAR observations.
- In our case, we have the data filtered at x_{obs} and the idea is to restore the stochastic modelling of the TKE at each time t. We do not have detailed information, thus we use a mesoscale approach.

The ideas presented here can be applied using different approaches, scales and context.

Stochastic Lagrangian models for turbulent flows

The simplest model for $(X_t, u'_t); 0 \leq t \leq T)$, consistent with the Navier Stokes equations, has been the generalized Langevin model:

$$\begin{cases} dX_t^{(i)} = U_t^{(i)} dt, \; 1 \leq i \leq 3, \\ du_t'^{(i)} = \sum_{j=1}^{3} G_{ij}(t, X_t) u_t'^{(j)} dt + \sqrt{C_0(t, X_t)\varepsilon(t, X_t)} dB_t^{(i)}, \end{cases} \tag{1}$$

for $i = 1, 2, 3$, with ε the dissipation rate of the turbulent kinetic energy

$$k(t, x) \quad = \frac{1}{2} \mathbb{E}\left[\left\| u'_t \right\|^2 | X_t = x \right]. \tag{2}$$

The non-dimensional coefficient $C_0(t, x)$ depends on the local values of the Reynolds stresses, the dissipation rate and the drag force $\frac{\partial \langle U^{(i)} \rangle}{\partial x_j}$.

- **Simplified Langevin Model (SLM):** The kinetic energy evolves correctly with homogeneous turbulence and the drift coefficient is isotropic. G is given as a return-to-mean force field and the velocity rewrites as an OU:

$$G_{ij}(t,x) = -\left(\frac{1}{2} + \frac{3}{4}C_0\right)\frac{\varepsilon(t,x)}{k(t,x)}\delta_{ij}. \tag{3}$$

At the Reynolds stress level, the SLM corresponds to a Rotta model with $C_R = \frac{3}{2}C_0 + 1$.

Closure models for G and ε

- **Simplified Langevin Model (SLM):** The kinetic energy evolves correctly with homogeneous turbulence and the drift coefficient is isotropic. G is given as a return-to-mean force field and the velocity rewrites as an OU:

$$G_{ij}(t,x) = -\left(\frac{1}{2} + \frac{3}{4}C_0\right)\frac{\varepsilon(t,x)}{k(t,x)}\delta_{ij}. \quad (3)$$

At the Reynolds stress level, the SLM corresponds to a Rotta model with $C_R = \frac{3}{2}C_0 + 1$.

1. **Mixing length parametrization:**

$$\varepsilon(t,x) = \frac{C_\varepsilon}{l_m}\left(k(t,x)\right)^{\frac{3}{2}}, \quad (4)$$

where $C_\varepsilon, l_m > 0$. The value of the mixing length l_m can be modelled proportional to z within the surface layer, and as a constant away from the surface layer.
We consider $l_m = \kappa z_{l_m}$, where κ is the Von Kármán constant and $z_{l_m} = 30[\mathtt{m}]$.

2. **Turbulent viscosity parametrization:** $\varepsilon(t,x) = \frac{C_\mu}{\nu_{turb}}\left(k(t,x)\right)^2$.

- **The isotropization-to-production (IP) model for homogeneous turbulence:**

$$\begin{aligned} G_{ij}(t,x) &= -\frac{C_R}{2}\frac{\varepsilon(t,x)}{k(t,x)}\delta_{ij} + C_2\frac{\partial\langle U^{(i)}\rangle}{\partial x_j}(t,x), \\ C_0\ \varepsilon(t,x) &= \frac{2}{3}\left(C_R\ \varepsilon(t,x) + C_2\ \mathscr{P}(t,x) - \varepsilon(t,x)\right), \end{aligned} \tag{5}$$

for $\mathscr{P}$ the turbulent production considering covariance terms $\langle u'^{(i)}u'^{(j)}\rangle$.
Here, C_0 is no more considered as a constant.

- **Elliptic blending model:** This closure model adds a wall effect (anisotropic effect) near the ground:

$$G_{ij}(t,x) = -\gamma_{i,j} - \frac{1}{2}\frac{\varepsilon(t,x)}{k(t,x)}\delta_{ij}, \qquad C_0\ \varepsilon(t,x) = \sum_{i,j=1}^{3}\frac{2}{3}(\gamma_{ij}(t,x))\langle u'^{(i)}u'^{(j)}\rangle(t,x) \tag{6}$$

where the tensor γ depends on the elliptic blending coefficient $\alpha(t,x)$, solution of the elliptic PDE:

$$L^2\nabla^2\alpha(t,x) - \alpha(t,x) = -\frac{1}{k(t,x)}.$$

Remark

The values of the constants C_0 and C_R might vary according to the closure model and the context. In general, the value $C_R = 1$ corresponds to no-return to isotropy, while values from 1.5 to 5.0 (C_0 between $1/3$ and $8/3$) have been suggested by different authors.

Hereafter, we focus on the SLM with mixing length closure parametrization, a model commonly used in Numerical Weather Prediction (NWP) solvers. Nonetheless, the modelling and methodology presented can be adapted to other closure models.

Reduced Lagrangian model for the instantaneous TKE

- Badosa et al. [2017], Murata et al. [2018]: Construction of a probabilistic forecast of solar irradiance, with a rather simple linear drift. Badosa and co-authors proposed an Itô process with diffusion $x^{\alpha}(1-x)^{\alpha}$ that results from a deterministic forecast and estimates the parameters involved by means of a variance-autocorrelation fitting.
- Arenas-López and Badaoui [2020] have proposed a data-driven OU process describing the wind speed on a scale of seconds.
- The Weibull distribution has been widely used in wind energy and other renewable energy sources where the main issue has been the estimation of the distribution coefficients.
- Benssousan and Brouste [2016] proposed a stochastic modelling of the squared norm of the wind velocity as a CIR process with coefficients to be calibrate.

The principle of our model is that an observer measures the velocity at each time t knowing the particle fluid position X_t equals $x_{\mathtt{obs}}$. Then, we consider

$$\langle U^{(i)}\rangle(t, x_{\mathtt{obs}}) = \mathbb{E}\left[U_t^{(i)} \middle| X_t = x_{\mathtt{obs}}\right],$$

and any of the Eulerian quantities involved in the model has to be understood as expectations of the corresponding Lagrangian quantities, knowing that the position is fixed at $x_{\mathtt{obs}}$. In particular, we get the relation

$$k(t, x_{\mathtt{obs}}) = \frac{1}{2}\mathbb{E}[\|u_t'\|^2].$$

Definition

We define the *instantaneous turbulent kinetic energy* as the stochastic process $(q_t;\ t \geq 0)$ given by

$$q_t := \|u_t'\|^2. \tag{7}$$

Within the real observations we noticed some external forces effect, producing some jumps (plus noise) in the wind velocity (and therefore in the kinetic energy). This behaviour evidence the need of a forcing term accounting for the turbulence production.

Using Itô formula, Levy's characterization theorem and a mixing length closure,

$$
\begin{aligned}
d\|u'_t\|^2 &= 2\sum_{i,j} u_t^{'(i)} G_{i,j} u_t^{'(j)} dt + 3C_0\varepsilon_t dt + 2\sqrt{C_0\varepsilon_t}\sum_i u_t^{'(i)} dB_t^{(i)} \\
&= \gamma_t dt - C_R \frac{\varepsilon_t}{k_t} q_t dt + 3C_0\varepsilon_t dt + 2\sqrt{C_0\varepsilon_t}\sqrt{q_t} dW_t,
\end{aligned}
$$

Hence, we deduce the CIR-type **mean-field TKE model**:

$$dq_t = \gamma dt - C_R \frac{C_\alpha}{\sqrt{2}} q_t \mathbb{E}^{1/2}[q_t] dt + 3C_0 \frac{C_\alpha}{2\sqrt{2}} \mathbb{E}^{3/2}[q_t] dt + \sqrt{\sqrt{2} C_0 C_\alpha} \mathbb{E}^{3/4}[q_t] \sqrt{q_t} dW_t, \tag{8}$$

with initial condition $q_0 = x$ and $C_\alpha := \frac{C_\varepsilon}{l_m}$.
The extra drift term γ accounting for the production of energy that, for simplicity, in this first part we consider as constant.

Hence, we deduce the CIR-type **mean-field TKE model**:

$$dq_t = \gamma dt - C_R \frac{C_\alpha}{\sqrt{2}} q_t \mathbb{E}^{1/2}[q_t]dt + 3C_0 \frac{C_\alpha}{2\sqrt{2}} \mathbb{E}^{3/2}[q_t]dt + \sqrt{\sqrt{2}C_0 C_\alpha \mathbb{E}^{3/4}[q_t]} \sqrt{q_t} dW_t, \tag{8}$$

with initial condition $q_0 = x$ and $C_\alpha := \frac{C_\varepsilon}{l_m}$.
The extra drift term γ accounting for the production of energy that, for simplicity, in this first part we consider as constant.

Wellposedness? Using the regularity of the map $t \mapsto \mathbb{E}[q_t]$ and classical stochastic analysis tools, we can prove (under some suitable assumptions on C_α, T, γ, C_0) the existence of a strong (positive) solution of the mean field model.

Remark

Formally, taking expectation on both sides of Eq. (8) we get the ODE

$$\frac{dk_t}{dt} = \left(\frac{\gamma}{2} - \varepsilon\right), \tag{9}$$

which corresponds to the classical equation with production term deduced from RANS equations with a $k - \epsilon$ closure model.

Remark

Formally, taking expectation on both sides of Eq. (8) we get the ODE

$$\frac{dk_t}{dt} = \left(\frac{\gamma}{2} - \varepsilon\right), \tag{9}$$

which corresponds to the classical equation with production term deduced from RANS equations with a $k - \epsilon$ closure model.

Using Itô formula to compute $\widetilde{k}_n(t) := \mathbb{E}[q_t^n]$, we have, for all $n \geq 1$:

$$\widetilde{k}_n(t) = \frac{1}{\mu_n(t)} \left\{ n \int_0^t \mu_n(s) \left[\left(n + \frac{1}{2}\right) C_0 \frac{C_\alpha}{\sqrt{2}} \widetilde{k}_1^{3/2}(s) + \gamma \right] \widetilde{k}_{n-1}(s) ds + x^n \mu_n(0) \right\}, \tag{10}$$

with $\mu_n(t) = \exp\left\{ n C_R \frac{C_\alpha}{\sqrt{2}} \int \widetilde{k}_1^{1/2}(t) \right\}$, and $\widetilde{k}_0(t) \equiv 1$.

In the particular case $\gamma = 0$:

$$\widetilde{k}_n(t) = \sum_{m=1}^{n} \beta_m(n, C_0, x) \left(x^{-1/2} + \frac{C_\alpha t}{2\sqrt{2}} \right)^{-2n-\alpha_m} \xrightarrow{t\to\infty} 0, \tag{11}$$

i.e. we conclude the dissipation of all the moments at large times.

Although $\mathbb{E}[q_t]$ cannot be written in analytic form when $\gamma \neq 0$, we can apply an isocline method to the ODE (9), concluding the boundedness of $\mathbb{E}[q_t]$ and its asymptotic behaviour at large times:

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

CIR MODEL FOR THE INSTANTANEOUS TKE

Although $\mathbb{E}[q_t]$ cannot be written in analytic form when $\gamma \neq 0$, we can apply an isocline method to the ODE (9), concluding the boundedness of $\mathbb{E}[q_t]$ and its asymptotic behaviour at large times:

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

Then,

$$\mathbb{E}[q_t] \approx \frac{1}{\tau}\int_0^\tau q_t dt = \left(\frac{\sqrt{2}\gamma[\tau]}{C_\alpha}\right)^{2/3} \tag{12}$$

INST. TKE AT THE EQUILIBRIUM

Substituting (12) in the mean-field SDE (8), we obtain the **CIR model for the instantaneous TKE**:

$$dq_t = \Theta(C_\alpha,\gamma)\left(\mu(C_\alpha,\gamma) - q_t\right)dt + \sigma(\gamma)\sqrt{q_t}dW_t, \quad q_0 = x, \tag{13}$$

where

$$\underbrace{\Theta(C_\alpha,\gamma)}_{\text{speed of adjst.}} = C_R\sqrt[3]{\frac{C_\alpha^2\gamma}{2}}, \qquad \underbrace{\mu(C_\alpha,\gamma)}_{\text{mean}}\Theta(C_\alpha,\gamma) = C_R\gamma, \qquad \sigma(\gamma) = \sqrt{2C_0\gamma}.$$

Remarkably, we recover the SDE form suggested by Bensoussan and Brouste! with parameters having an intrinsic physical meaning.

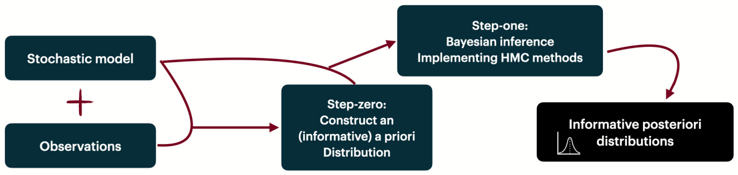
- Under the condition $2\Theta\mu \geq \sigma^2$ (always satisfied in our case) the state zero is exclude for the trajectories of the CIR process.
- Is an ergodic process with known stationary density and with a well-known explicit relation with a chi-square random variable.
- Explicit transition density in terms of Bessel functions.
- Explicit moments in terms of hypergeometric functions.

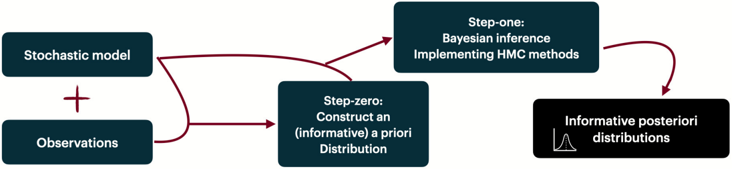
Calibration and analysis of the reduced model

Next step: infer on the possible values of the parameters in the model $\theta = (C_\alpha, \gamma)$ considering the observations. Here we consider the Kolmogorov constant as a prescribed value $C_0 = 1.9$, nonetheless, the methodology can be extend to the calibration of C_0.

- The main drawback with point-estimators is that we do not have access to the possible uncertainties.
- The main drawback with Bayesian approaches, where we construct a Markov chain that converge to the stationary distribution of the parameters, is that we need a first guess. **What about a guess for γ?**

To provide a reliable calibration with no external parameters, we propose a two-step method in which we construct the a priori distribution of the parameters through a learning stage (called step-zero) to then quantify the uncertainty of the parameters in the step-one


Stochastic model
+
Observations
Step-zero:
Construct an
(informative) a priori
Distribution
Step-one:
Bayesian inference
Implementing HMC methods
Informative posteriori
distributions



On the value of C_α

Recall that $C_\alpha = \dfrac{C_\epsilon}{l_m}$, with $l_m(z) = \kappa z_{l_m}$. Hence, considering the reference values for the Von-Kármán constant $\kappa \in [0.287, 0.615]$, and $C_\mu \in [0.054, 0.135]$ for $C_\epsilon = C_\mu^{3/4}$, we have

$$C_\alpha \in [0.0061, 0.0259].$$

For both stages of the calibration we use a numerical approximation for the solution of the model (13). For this, let $\Pi = \{t_0, t_1, \dots, t_N\}$ be a partition of the time interval $[0, T]$ (measure in seconds) with homogeneous time step $\Delta t = \frac{T}{N}$; and define the Symmetrized Euler scheme associated to the SDE (13) as:

$$\widehat{q}_{t_{n+1}} = \widehat{q}_{t_n} + \Theta(C_\alpha, \gamma)\left(\mu(C_\alpha, \gamma) - \widehat{q}_{t_n}\right)\Delta t + \sigma(\gamma)\sqrt{\widehat{q}_{t_n}}\left(W_{t_{n+1}} - W_{t_n}\right), \tag{14}$$

updated with $\widehat{q}_{t_n} = |\widehat{q}_{t_n}|$, for all $n = 0, \dots N-1$, with initial condition $\widehat{q}_0 = x$.

Converging weakly for $C_0 < 2$.

Step Zero: Prior calibration

Using this approach we do not need a priori distribution for the parameters θ but the joint density $p^{\theta}(\mathbf{q})$ associated with the model.

Assume we have a perfect model and independent observations (i.e. the SES (14) replicates perfectly the observed data). Notice that:

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

Step Zero: Prior calibration

Using this approach we do not need a priori distribution for the parameters θ but the joint density $p^{\theta}(\mathbf{q})$ associated with the model.

Assume we have a perfect model and independent observations (i.e. the SES (14) replicates perfectly the observed data). Notice that:

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

Considering $D \subset \mathbb{R}^+ \times \mathbb{R}^+$ as the compact set supporting the admissible values of the parameter vector, we compute the maximum pseudo-likelihood estimator as follows

$$\begin{aligned}
\widehat{\theta} &= \texttt{argmax}_{\theta\in D}\ \log p^{\theta}_{\Delta t}(\mathbf{q}) \\
&= \texttt{argmax}_{\theta\in D}\ \Bigg\{-\frac{N}{2}\log\gamma - \frac{\overline{M}_{2,-1}}{4C_0\Delta t\gamma} - \frac{C_R(\widehat{q}_T - x)}{2^{4/3}C_0}\left(\frac{C_\alpha}{\gamma}\right)^{2/3} \\
&\quad -\frac{C_R^2\Delta t\overline{M}_{0,1}}{C_0 2^{8/3}}\left(\frac{C_\alpha}{\gamma}\right)^{1/3} C_\alpha + \frac{C_R^2 N\Delta t}{C_0 2^{4/3}}\left(\frac{C_\alpha}{\gamma}\right)^{2/3}\gamma - \frac{C_R^2\Delta t\overline{M}_{0,-1}}{4C_0}\gamma + \texttt{Ctte}\Bigg\},
\end{aligned}$$

where we used the notation:

$$\widehat{M}_{m_1,m_2} = \frac{1}{N}\sum_{n=0}^{N-1}\left(\widehat{q}_{t_{n+1}} - |\widehat{q}_{t_n}|\right)^{m_1}|\widehat{q}_{t_n}|^{m_2}.$$

- We construct an estimator for γ by using the convergence of the quadratic variation of a diffusion process: quadratic variation estimator

$$\widehat{\gamma} = \frac{\widehat{M}_{2,0}}{2C_0 \Delta t \widehat{M}_{0,1}}, \tag{15}$$

satisfying $\widehat{\gamma} \xrightarrow{N\to\infty} \gamma$, in probability.

- We construct an estimator for γ by using the convergence of the quadratic variation of a diffusion process: quadratic variation estimator

$$\widehat{\gamma} = \frac{\widehat{M}_{2,0}}{2C_0 \Delta t \widehat{M}_{0,1}}, \tag{15}$$

satisfying $\widehat{\gamma} \xrightarrow{N\to\infty} \gamma$, in probability.

- Then, we compute the maximum likelihood estimator for C_α:

$$\widehat{C}_\alpha = \sqrt{2/\widehat{\gamma}} \left(\frac{\widehat{\gamma}\Delta t C_R - \widehat{M}_{1,0}}{\Delta t C_R \widehat{M}_{0,1}} \right)^{3/2}. \tag{16}$$

- We construct an estimator for γ by using the convergence of the quadratic variation of a diffusion process: quadratic variation estimator

$$\widehat{\gamma} = \frac{\widehat{M}_{2,0}}{2C_0 \Delta t \widehat{M}_{0,1}}, \tag{15}$$

satisfying $\widehat{\gamma} \xrightarrow{N\to\infty} \gamma$, in probability.

- Then, we compute the maximum likelihood estimator for C_α:

$$\widehat{C}_\alpha = \sqrt{2/\widehat{\gamma}} \left(\frac{\widehat{\gamma} \Delta t C_R - \widehat{M}_{1,0}}{\Delta t C_R \widehat{M}_{0,1}} \right)^{3/2}. \tag{16}$$

- We construct an a priori distribution for θ, sufficiently informative and independent of external information, through the set

$$\Delta := \{(\widehat{C}_\alpha(d),\ \widehat{\gamma}(d)) : \text{ for all Wednesday } d \text{ of the year 2017}\},$$

and define the truncated Gaussian a priori distributions:

$$\gamma \sim \mathscr{N}^+\left(\overline{\Gamma}, \mathbb{V}_\Gamma\right), \quad C_\alpha \sim \mathscr{N}^+\left(\overline{C}, \mathbb{V}_C\right), \tag{17}$$

where $\overline{\Gamma}$ and $\overline{C}$ denotes the empirical mean of the γ and C_α estimators, and similarly $\mathbb{V}_\Gamma$ and $\mathbb{V}_C$ their corresponding empirical variance.

Step One: Posterior calibration of reduced model

Here we update our initial guess on the values of the parameters with a statistical model considering any deviation as an observation error:

$$\mathbf{q}^{\text{obs}}(\theta) = \widehat{\mathbf{q}}(\theta) + \mathscr{E}, \tag{18}$$

where $\widehat{\mathbf{q}}(\theta)$ stands for the output of the numerical approximation of the TKE; $\mathscr{E}$ is a logistic random vector having zero mean and scale parameter s to be calibrate.

Step One: Posterior calibration of reduced model

Here we update our initial guess on the values of the parameters with a statistical model considering any deviation as an observation error:

$$\mathbf{q}^{\text{obs}}(\theta) = \widehat{\mathbf{q}}(\theta) + \mathscr{E}, \tag{18}$$

where $\widehat{\mathbf{q}}(\theta)$ stands for the output of the numerical approximation of the TKE; $\mathscr{E}$ is a logistic random vector having zero mean and scale parameter s to be calibrate.

Therefore, from Bayes theorem, we know that

$$\pi(\theta|\mathbf{q}^{\text{obs}}) \propto p(\mathbf{q}^{\text{obs}}|\theta)p_\theta(\theta). \tag{19}$$

Step One: Posterior calibration of reduced model

Here we update our initial guess on the values of the parameters with a statistical model considering any deviation as an observation error:

$$\mathbf{q}^{\text{obs}}(\theta) = \widehat{\mathbf{q}}(\theta) + \mathscr{E}, \tag{18}$$

where $\widehat{\mathbf{q}}(\theta)$ stands for the output of the numerical approximation of the TKE; $\mathscr{E}$ is a logistic random vector having zero mean and scale parameter s to be calibrate.

Therefore, from Bayes theorem, we know that

$$\pi(\theta|\mathbf{q}^{\text{obs}}) \propto p(\mathbf{q}^{\text{obs}}|\theta) p_\theta(\theta). \tag{19}$$

Metropolis-Hastings Algorithm: Used to sample from the posterior.
Start from an initial value θ_0, for the nth-iteration we proceed as follows,

1. Simulate $\widetilde{\theta} \sim \rho(\widetilde{\theta}|\theta_n)$, and $u \sim \mathscr{U}(0,1)$, where ρ is a proposed transition density.
2. Compute
$$a := \min\left\{1, \frac{p_\theta(\widetilde{\theta}) p(\mathbf{q}^{\text{obs}}|\widetilde{\theta}) \rho(\theta_n|\widetilde{\theta})}{p_\theta(\theta_n) p(\mathbf{q}^{\text{obs}}|\theta_n) \rho(\widetilde{\theta}|\theta_n)}\right\}. \tag{20}$$
3. If $u < a$, then $\theta_{n+1} = \widetilde{\theta}$ and we accept the simulated state.
Else, we reject and keep the previous state, i.e. $\theta_{n+1} = \theta_n$.

A common issue in the implementation of this method is the correct exploration of the state space in high dimension, **which in our case is crucial**.

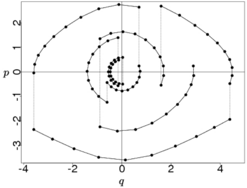
Hamiltonian Monte Carlo methods

Use Hamiltonian equations to mimic the dynamic of a particle following the contour with high probability mass by introducing a synthetic momentum variable p and the Hamiltonian

$$H(\theta, p) = K(\theta, p) \underbrace{- \log \pi(\theta|\mathbf{q}^{\text{obs}})}_{\text{potential}}, \qquad \left\{ \begin{array}{ll} \dfrac{d\theta}{dt} & = \dfrac{\partial H}{\partial p} \\ \dfrac{dp}{dt} & = -\dfrac{\partial H}{\partial \theta}. \end{array} \right.$$

Assuming constant energy, the evolution -in time- of the particle generates the contour of the target distribution π (obtained from the marginal distribution).

- Advantages: preservation of the volume, efficient exploration in high dimensional cases, reversibility of the dynamics.
- Disadvantage: the introduction of this auxiliary momentum duplicate the number of variables, and therefore the computational cost.



To propose a new state for the Markov chain, we project the trajectory back to the parameter space, and finally accept/reject the state by following a decision step similar to the one implemented in the Metropolis-Hasting algorithm.

The approximation of the Hamiltonian equations: Leapfrog method, which start with half step for the momentum variables, then do a full step for the position using the update momentum, and finally complete the remaining half step for the momentum.
Self-tuning variant of the leapfrog method: NUTS method (No-U-Turn Sampler).

Description of the Bayesian calibration

Recall that in the SDE (13):

C_α appears only in the drift.

γ appears in particular in the diffusion.

Description of the Bayesian calibration

Recall that in the SDE (13):
C_α appears only in the drift.
γ appears in particular in the diffusion.

Then, we balance the cost by playing with the observation frequency that serves in the method: $\xi_{C_\alpha} < \xi_\gamma < \frac{1}{10}[\texttt{sec}]$

1. *Estimate the observation error*: Compute the mean

$$\mathbb{E}[\mathscr{E}_t(d)] = \mathbb{E}[{q_t}^{\text{obs}} - \widehat{q}_t(\widehat{C}_\alpha(d), \widehat{\gamma}(d))],$$

   using a Monte Carlo method, and then approximate the scale $\widehat{s}^2 = 3\frac{\mathbb{V}[\mathscr{E}(d)]}{\pi^2}$.

Description of the Bayesian calibration

Recall that in the SDE (13):
C_α appears only in the drift.
γ appears in particular in the diffusion.

Then, we balance the cost by playing with the observation frequency that serves in the method: $\xi_{C_\alpha} < \xi_\gamma < \frac{1}{10}[\texttt{sec}]$

1. *Estimate the observation error*: Compute the mean

$$\mathbb{E}[\mathscr{E}_t(d)] = \mathbb{E}[{q_t}^{\text{obs}} - \widehat{q}_t(\widehat{C}_\alpha(d), \widehat{\gamma}(d))],$$

   using a Monte Carlo method, and then approximate the scale $\widehat{s}^2 = 3\frac{\mathbb{V}[\mathscr{E}(d)]}{\pi^2}$.

2. *Bayesian calibration of γ*: Split the signal considering the partition $[0,T] = \bigcup_{0\le i\le N} [T_i, T_{i+1}]$. Then, we estimate the density of γ_i considering the sub-signal $\mathbf{q}^{\text{obs}}\big|_{[T_i,T_{i+1}]}$, and therefore we allow for the parameter to variate in time.

   Here we use the statistical model

$$\mathbf{q}^{\text{obs}} = \widehat{\mathbf{q}}(\gamma_i(\omega), \widehat{C}_\alpha(d)) + \mathscr{E}(\omega),$$

   where $\gamma_i \sim \mathscr{N}^+\left(\overline{\Gamma}, \mathbb{V}_\Gamma\right)$ in (17), observation error $\mathscr{E} \sim \texttt{Logistic}(0,s)$ and scale s such that $\log(s) \sim \mathscr{N}(\widehat{s},1)$.

3. *Bayesian calibration of C_α:* From the family of Markov chains $(\{\gamma_{i,n}(d)\}_{n\geq 1}; i = 0, \ldots, N)$, we define the piecewise constant production term

$$\gamma_t = \sum_{i=0}^{N} \mathbb{E}[\gamma_i] \mathbf{1}_{[T_i, T_{i+1}]}(t),$$

and we apply a HMC method for the construction of the Markov chain associated with C_α with the statistical model:

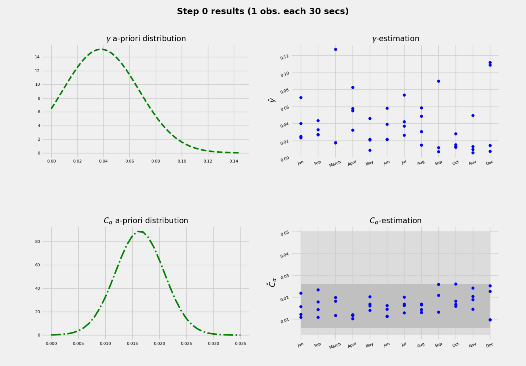
$$\mathbf{q}_t^{\mathtt{obs}} = \widehat{\mathbf{q}}_t(\gamma_t, C_\alpha(\omega)) + \mathscr{E}(\omega),$$

where $C_\alpha \sim \mathscr{N}^+ \left(\overline{C}, \mathbb{V}_C\right)$ in (17), and observation error $\mathscr{E}$ as in 2. In this case we consider $\mathbf{q}^{\mathtt{obs}}$ with frequency ξ_{C_α}.

Convergence diagnostic tests: Verify if the chain explores the state space thoroughly, convergence of the empirical mean, analyse if the simulated values are uncorrelated. We also can apply formal statistical diagnostic methods.

Results

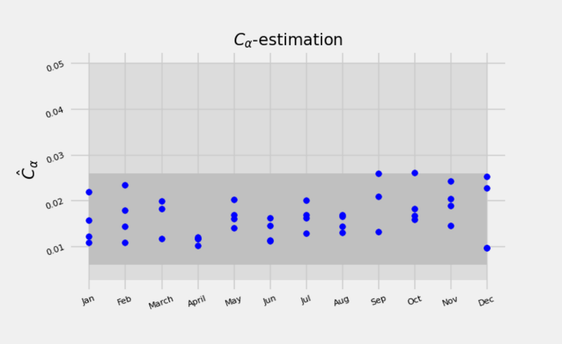
We have selected a window of 16 hours of observations for each day, between 4:00 am and 8:00 pm. The partition of each signal during this time will be done in 20-minute sub-signals, for a total of 48 sub-signals per day. The data frequencies were set as: $\xi_\gamma = 12[\texttt{obs/min}]$ and $\xi_{C_\alpha} = 2[\texttt{obs/min}]$.



In order to validate the results obtained from step zero, compare the quotient $\left(\frac{\sqrt{2}\widehat{\gamma}(d)}{\widehat{C}_\alpha(d)}\right)^{2/3}$ against the time-averaging $\frac{1}{\#\text{obs}}\sum_t \mathbf{q}_t^{\text{obs}}$, obtaining an absolute error of order -4.

Results

We have selected a window of 16 hours of observations for each day, between 4:00 am and 8:00 pm. The partition of each signal during this time will be done in 20-minute sub-signals, for a total of 48 sub-signals per day. The data frequencies were set as: $\xi_\gamma = 12[\mathtt{obs/min}]$ and $\xi_{C_\alpha} = 2[\mathtt{obs/min}]$.



In order to validate the results obtained from step zero, compare the quotient $\left(\frac{\sqrt{2}\widehat{\gamma}(d)}{\widehat{C}_\alpha(d)}\right)^{2/3}$ against the time-averaging $\frac{1}{\#\text{obs}}\sum_t \mathbf{q}_t^{\text{obs}}$, obtaining an absolute error of order -4.

Calibration of C_α

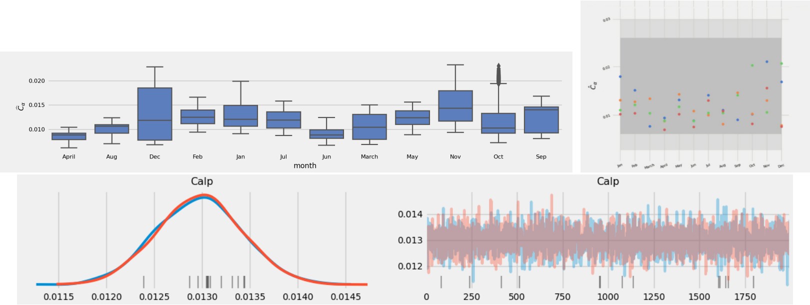


FIGURE: Bayesian calibration of C_α for all Wednesday of 2017. In the bottom figure, the exploration of the state space and posterior C_α distribution.

Calibration of γ

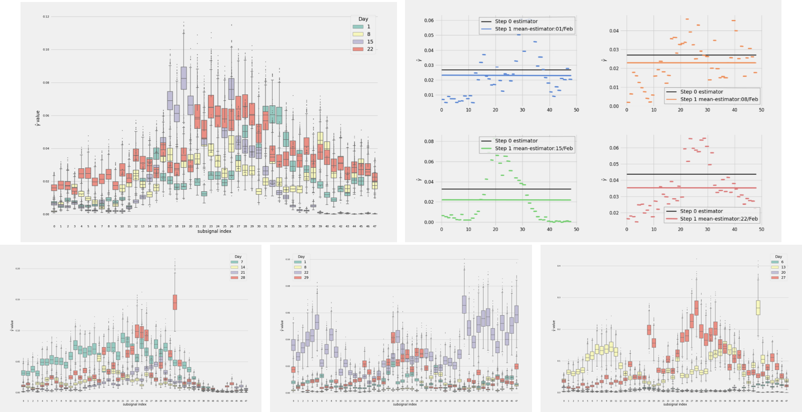
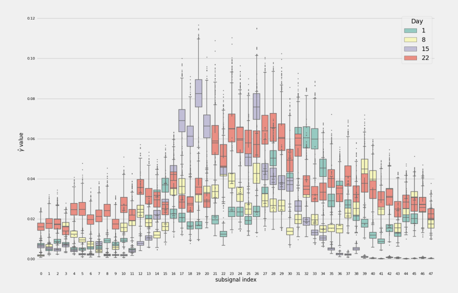
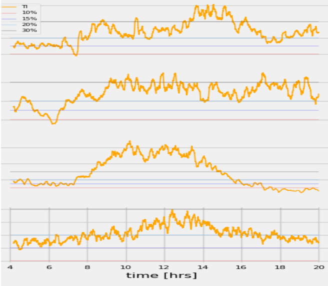


Figure: Blox plot for each 20minutes-length subsignal and comparison between step 0 and step 1 with mean posterior estimators (small line-segments) day-mean posterior estimator (solid color line) and prior estimator (black line).

An important observation to follow from the analysis of the calibration results is the connection between the production term γ and the turbulence intensity, not in magnitude but in in the evolution of the dynamic:

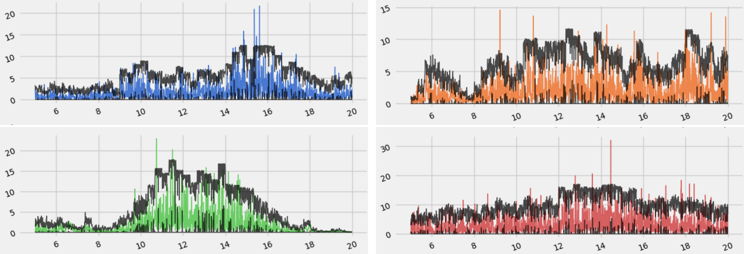




Once the calibration was performed, we fixed the inferred values $C_\alpha(d)$ and $\gamma_t(d) = \sum_{i=0}^{N} \mathbb{E}[\gamma_i(d)]\mathbf{1}_{[T_i,T_{i+1}]}(t)$, and construct the associated time-discretization for the instantaneous TKE as:

$$\widehat{q}_{t_{n+1}} = \widehat{q}_{t_n} + C_R\gamma_t(d)\Delta t - C_R\left(\frac{C_\alpha^2(d)\gamma_t(d)}{2}\right)^{1/3}\widehat{q}_{t_n}\Delta t + \sqrt{2C_0\gamma_t(d)}\sqrt{\widehat{q}_{t_n}}(W_{t_{t+1}} - W_{t_n}),$$

and compare the observations against a 95% confidence interval:



Relation between I_t and γ_t

Considering the kinetic energy in a stationary regime:

$$C_R\gamma_t - C_R\left(\frac{C_\alpha^2\gamma_t}{2}\right)^{1/3}\mathbb{E}[q_t] = 0,$$

from which we deduce the relation

$$\gamma_t = \frac{C_\alpha}{\sqrt{2}}|\langle U\rangle|^3 I_t^3 3^{3/2}. \tag{21}$$

Relation between I_t and γ_t

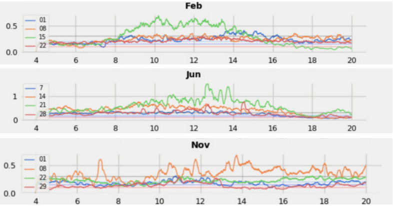
Considering the kinetic energy in a stationary regime:

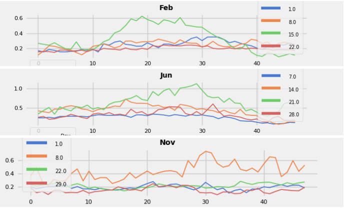
$$C_R \gamma_t - C_R \left(\frac{C_\alpha^2 \gamma_t}{2} \right)^{1/3} \mathbb{E}[q_t] = 0,$$

from which we deduce the relation

$$\gamma_t = \frac{C_\alpha}{\sqrt{2}} |\langle U \rangle|^3 I_t^3 3^{3/2}. \tag{21}$$

$$I_t = \frac{\sqrt{2k(t,x)}}{\sqrt{3}\|\langle U \rangle\|} \qquad \widehat{I}_t = \frac{1}{|\langle U \rangle|\sqrt{3}} \left(\frac{\sqrt{2}\gamma_t}{C_\alpha} \right)^{1/3}$$





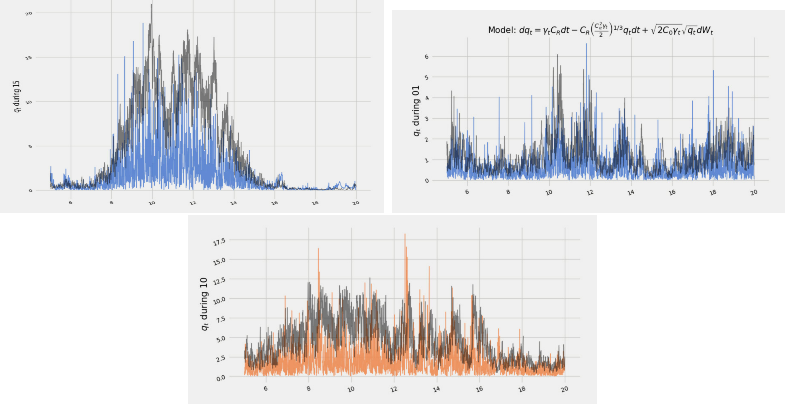


FIGURE: Prediction of the instantaneous TKE: construction of 95% confidence intervals from the CIR and γ_t given by (21). Observations were taken during February 15th, November 1rst and November 10th between 5:00 am and 8:00 pm (color plots).

References

S. Pope.
Lagrangian pdf methods for turbulent flows. 1994.

A. Durbin and C-G. Speziale.
Realizability of second-moment closure via stochastic analysis. 1994.

W. C. Miao.
Quadratic variation estimators for diffusion models in finance. 2004.

Bernardin, Bossy, Chauvin, Drobinski, Rousseau and Salameh.
Stochastic Downscaling Methods: Application to Wind Refinement, 2009.

C. Baehr.
Nonlinear Filtering for observations on a random vector field along a Random path, 2010.

Neal, Radford M and others.
MCMC using Hamiltonian dynamics, 2011.

W. Edeling and P. Cinnella and R. P. Dwight and H. Bij
Bayesian estimates of parameter variability in the $k-\varepsilon$ *turbulence model*, 2014.

A. Bensoussan, and A. Brouste.
Cox–Ingersoll–Ross model for wind speed modeling and forecasting, 2016.

J. Badosa, E. Gobet, M. Grangereau and D. Kim.
Day-ahead probabilistic forecast of solar irradiance: a Stochastic Differential Equation approach, 2017.

A. Murata, H. Ohtake, and T. Oozeki.
Modeling of uncertainty of solar irradiance forecasts on numerical weather predictions with the estimation of multiple confidence intervals, 2018.

K. Martínez.
PhD. Thesis, 2019.

J.P. Arenas-López, and M. Badaoui.
Stochastic modelling of wind speeds based on turbulence intensity, 2020.