Tailor-made data-driven similarity theories for the temperature structure parameter C_T^2 in the lower atmospheric boundary layer

Maximilian Pierzyna (1), Rudolf Saathof (2), Sukanta Basu (3)(4)

m.pierzyna@tudelft.nl | mpier.eu

(1) Dept. of Geoscience and Remote Sensing, Delft University of Technology, Delft, NL

(2) Faculty of Aerospace Engineering, Delft University of Technology, Delft, NL

(3) Atmospheric Science Research Center, University at Albany, Albany, NY, USA

(4) Dept. of Environmental and Sustainable Engineering, University at Albany, Albany, NY, USA

What is Optical Turbulence (OT) and why does it matter?

Optical ground-based astronomy and future free-space optical communication (FSOC) suffer from **light getting distorted** when it propagates through the turbulent atmosphere. In astronomy, **turbulent fluctuations of the atmospheric refractive index, known as optical turbulence (OT),** cause blurry images and limit the detection of small objects. FSOC links, which use optical beams to transmit data instead of traditional radio waves, experience reduced data rates or even link interruptions due to laser beams getting distorted by OT (see Fig. 1).

Measuring OT needs expensive specialized instruments and simulating OT numerically is expensive due to high required spatial resolution to resolve the inertial subrange of turbulence. That is why, turbulence parameterizations, which estimate turbulent parameters from more readily accessible variables, have been actively researched for decades.

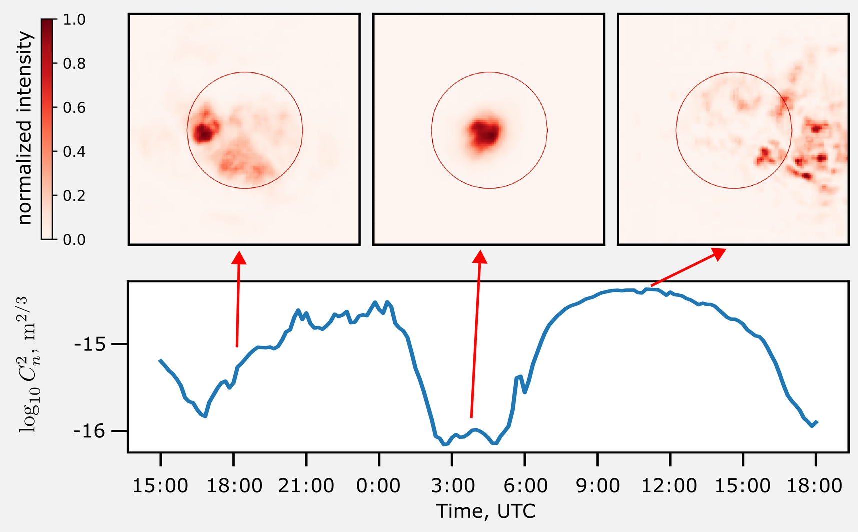


Figure 1: Effect of different strengths of optical turbulence on laser beam propagating through the atmosphere.

This poster presents how the **data-driven similarity theory framework Π-ML [PSB23] is used to obtain physically consistent non-dimensional scalings and parameterizations of OT** using only the variables available in specific use cases.

Direct and indirect measurements of optical turbulence strength C_n^2

As basis of our parameterizations, observed C_n^2 data is needed. C_n^2 can be obtained directly or indirectly, where **indirect observations from eddy-covariance systems are used in the present study.**

Direct: Scintillometer

Measures scintillation, i.e., fluctuations of the received laser intensity (path-averaged)

$$C_n^2 \sim \sigma^2_{\log(I)}$$

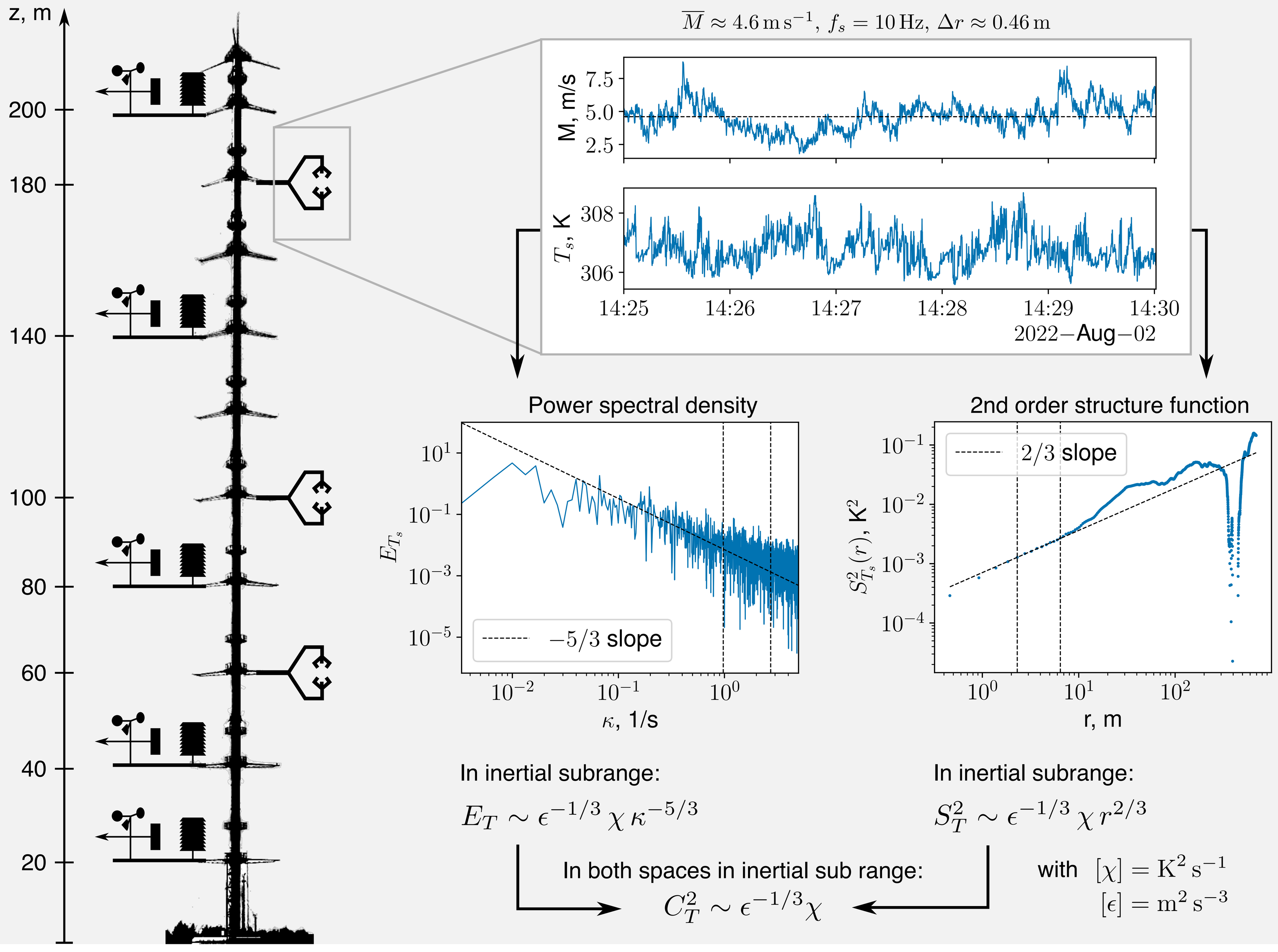
Indirect: Eddy-covariance system with post-processing

Measure fast (>10 Hz) wind, temperature, and humidity (not here) signals to determine structure function coefficients in inertial subrange.

$$C_n^2 = \frac{A_T^2}{\overline{T}^2} C_T^2 + \frac{A_T A_q}{\overline{T}\,\overline{q}} C_{Tq}^2 + \frac{A_q^2}{\overline{q}^2} C_q^2$$

Structure function approach used to obtain training data

The structure function coefficient for the indirect method is determined by fitting a linear function to the inertial subrange part of the structure function in log-log space. The 2[nd] order structure function of sonic temperature is defined as $S_T^2(r) = \langle [T_s(x) - T_s(x+r)]^2 \rangle$. Assuming frozen turbulence, the spatial increments are obtained from a time signal as $\Delta r \approx \overline{M}/f_s$.



Obtaining data-driven parameterizations using Π-ML [PSB23]

1. Collect relevant variables and their dimension

(a) traditional: gradients, fluxes, variances (5 min bins)

	Input features ($\boldsymbol{X}$)							Target (y)
	$\partial\overline{u}_i/\partial z$	$\partial\overline{\theta}/\partial z$	$(u_*)_L$	$(\overline{w'\theta'})_L$	$\left(\overline{u_i'^2}\right)_L$	$\left(\overline{\theta'^2}\right)_L$	z	C_T^2
K		1		1		2		2
L		−1	1	1	2		1	−2/3
T	−1		−1	−1	−2			

(b) practical: use downsampled temperature signal only

	Input features ($\boldsymbol{X}$)						Target (y)
	var$[\tilde{T}]$	skew$[\tilde{T}]$	mean$[\Delta_\tau\tilde{T}]$	var$[\Delta_\tau\tilde{T}]$	skew$[\Delta_\tau\tilde{T}]$	z	C_T^2
K	2	3	1	2	3		2
L						1	-2/3

with increment operator $\Delta_\tau\tilde{T} = \tilde{T}(t) - \tilde{T}(t+\tau),\ \tau \in [1,2,4]$

2. Systematically generate Π-sets using Buckingham's Π theorem

$$\Pi_1 = (S\,z)/u_* \qquad \Pi_4 = e/u_*^2$$

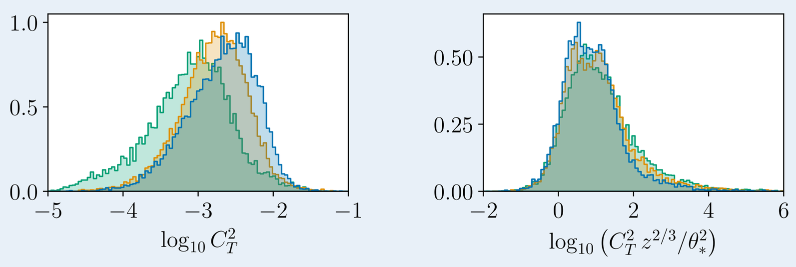
$$\Pi_2 = (\Gamma\,u_*\,z)/\overline{w'\theta'} \qquad \Pi_5 = (\sigma_\theta^2\,u_*^2)/\overline{w'\theta'} \qquad \Pi_y = \frac{C_T^2\,u_*^2\,z^{2/3}}{\left(\overline{w'\theta'}\right)^2}$$

$$\Pi_3 = \sigma_M^2/u_*^2 \qquad \Pi_6 = \sigma_w/u_*^2$$

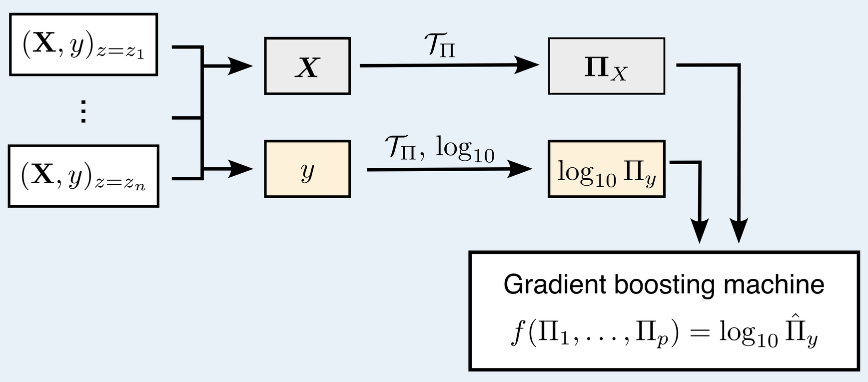
with $e = \left(\sigma_u^2 + \sigma_v^2 + \sigma_w^2\right)/2$ and $\sigma_M^2 = \sigma_u^2 + \sigma_v^2$

3. Train Gradient Boosting Machines (GBMs) for each Π-set

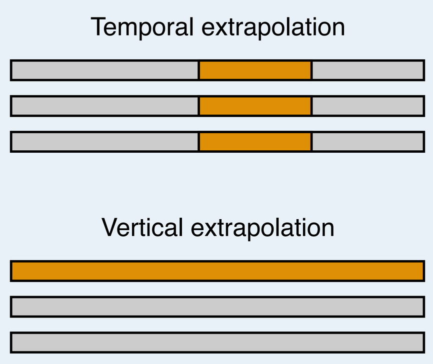
Overlapping histograms after scaling indicate good collapse of data in non-dimensional space:



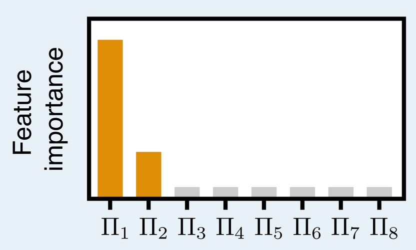
Training pipeline:



4. Evaluate each model in Π space and normal space



5. Select model and Π-set best balancing accuracy, complexity, and domain knowledge



Traditional parameterizations

For an in-depth intercomparison, see [PHSB24]

Monin-Obukhov similarity theory based [W71]...

... using **fluxes:** $C_T^2 = \left(-\overline{w'\theta'}/u_*\right)^2 z^{-2/3}\,g(\zeta)$

with stability-dependent similarity function $g(\zeta)$,

where $\zeta = z/L$ and $L = -u_*^3\,\overline{T}/\left(\kappa\,g\,\overline{w'\theta'}\right)$.

... using **gradients:** $C_T^2 = \Gamma^2 z^{4/3}\,f(\mathrm{Ri}_g)$

with stability-dependent similarity function $f(\mathrm{Ri}_g)$

MOST assumptions inherited → constant-flux layer!

Physics-based using variances [HB15]: $C_T^2 = 3.2\chi\epsilon^{-1/3}$

with $\epsilon = \frac{(2e)^{3/2}}{B_1 L_M}$, $\chi = \frac{(2e)^{1/2}}{B_2 L_M}\sigma_\theta^2$,

and master length scale [MY82, NN09, O19]

$$L_M = \left(\frac{1}{L_S} + \frac{1}{L_T} + \frac{1}{L_B}\right)^{-1}$$

L_M is difficult to determine experimentally

Results: tailor-made / use case-dependent parameterizations

Traditional: flux and variance-based parameterization

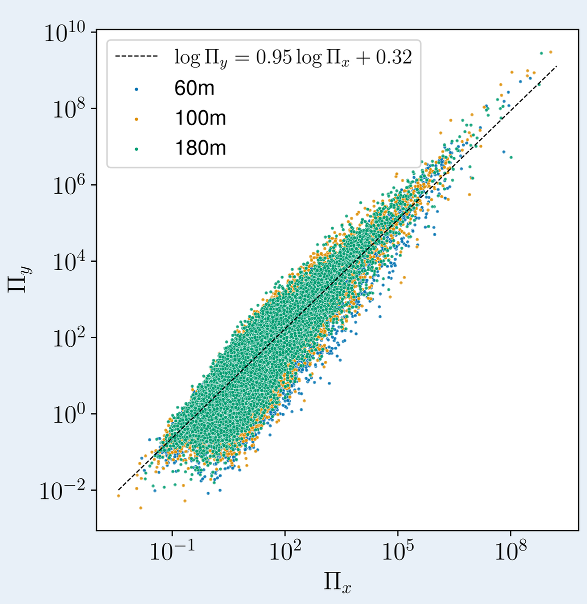


Figure 2: Demonstration of data collapsing in Π space for all levels with simple fitted linear model.

Use case: Outputs of **numerical simulations** with access to local fluxes and variances, e.g., large eddy simulations or mesoscale simulations with high-order closure.

Variables: local fluxes and variances

Result: $\Pi_x = \frac{\sigma^2_{\theta,L}}{\theta^2_{*,L}} \qquad \Pi_y = \frac{C_T^2\,z^{2/3}}{\theta^2_{*,L}} \qquad \theta_{*,L} = \frac{-\left(\overline{w'\theta'}\right)_L}{u_{*,L}}$

- Single flux- and variance-based Π expression dominates and is sufficient to collapse data reasonably well (cf. Fig 2).
- Groups containing gradients are discarded by GBM model.
- Scatter may point at missing length scale.
- Scaling is similar to W71 and consistent with theory ([BH22]: $C_T^2 \sim \sigma_\theta^2\,L^{-2/3}$, with $L \sim z$ in neutral conditions). Π groups containing more complex length scales are currently being investigated.

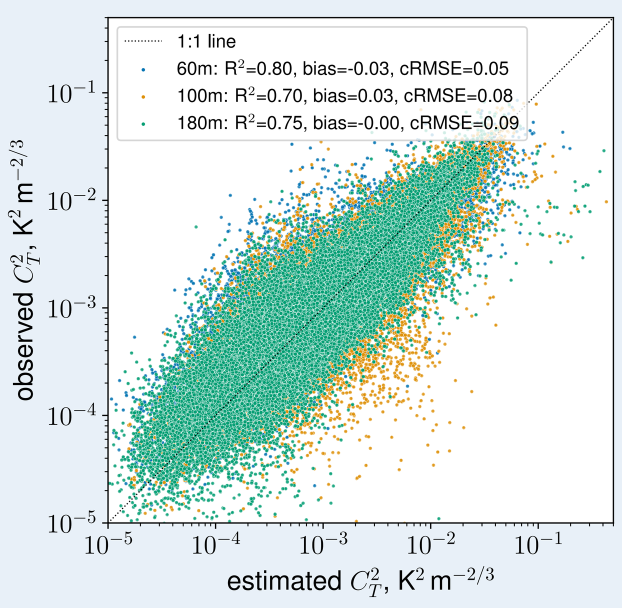


Figure 3: Vertical extrapolation performance of GBM models in log-space using flux-based Π set.

Practical: 1 Hz "temperature only" parameterization

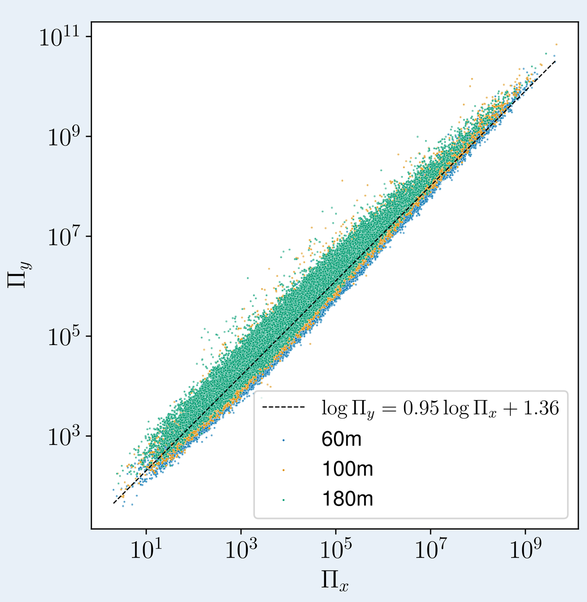


Figure 4: Demonstration of data collapsing in Π space for all levels with simple fitted linear model.

Use case: 1 Hz thermocouples during **field campaign** without wind measurements.

Variables: statistics of (downsampled) 1 Hz temperature time series $\tilde{T}(t)$ and temperature increment signal $\Delta_\tau\tilde{T}_s = \tilde{T}_s(t) - \tilde{T}_s(t+\tau)$ lagged by τ samples. Note that C_T^2 for training is still estimated from the 10 Hz signal.

Motivation: No expensive 3D sonic anemometer needed. No wind speed observations needed.

Result: $\Pi_x = \frac{\mathrm{var}\left[\Delta_1\tilde{T}_s\right]}{\left(\mathrm{mean}\left[\Delta_1\tilde{T}_s\right]\right)^2} \qquad \Pi_y = \frac{C_T^2\,z^{2/3}}{\left(\mathrm{mean}\left[\Delta_1\tilde{T}_s\right]\right)^2}$

- Single Π expression similar to traditional parameterization is sufficient to collapses data very well (cf. Fig. 4).
- Trained GBM models exhibit good but biased vertical extrapolation performance (cf. Fig. 5).

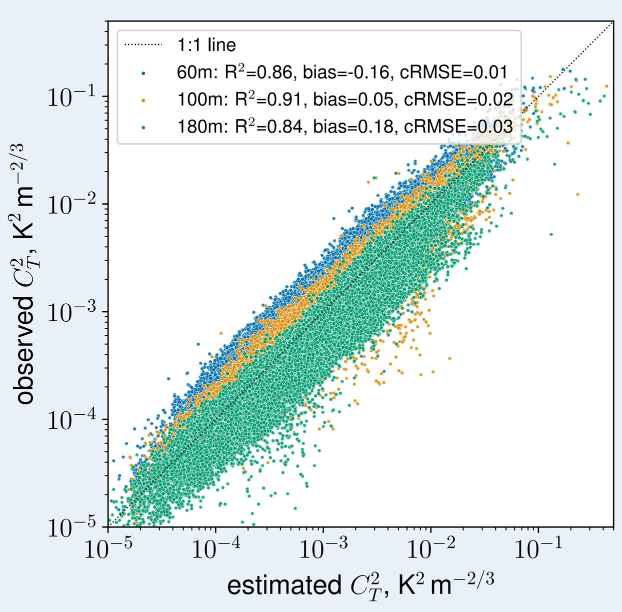


Figure 5: Vertical extrapolation performance of GBM models in log-space using temperature-only Π set.

Conclusions

- **Π-ML is a powerful tool to derive non-dimensional scalings for flow processes** using only the variables available → tailor-made
  - **First, expand complexity** of the fitting problem through generation of large number of physically motivated Π-groups and Π-sets.
  - Using observational data, **machine learning (ML) regression models are fit to assess the capability of each Π-set** to parameterize the target.
  - The Π-sets and corresponding fitted models are analyzed to **keep only moderately complex and well-performing combinations.**
- For C_T^2 parameterizations, simple enough Π spaces are found to replace **ML regression by a simple linear model**, ultimately, bringing the complexity back down to an "analytical" level.

Acknowledgments

MP thanks the Royal Netherlands Meteorological Institute (KNMI) for providing the sonic anemometer data used in this study. MP is financed by NWO TTW-Perspectief P19-13 "FREE — Optical Wireless Superhighways." SB is is grateful for financial support from the State University of New York's Empire Innovation Program.

References

[W71] Wyngaard, J. C., et al. "Behavior of the Refractive-Index-Structure Parameter near the Ground*." J. Opt. Soc. America, vol. 61, no. 12, Dec. 1971, pp. 1646–50. https://doi.org/10.1364/JOSA.61.001646.

[MY82] Mellor, George L., and Tetsuji Yamada. "Development of a Turbulence Closure Model for Geophysical Fluid Problems." Reviews of Geophysics, vol. 20, no. 4, 1982, pp. 851–75. Wiley Online Library, https://doi.org/10.1029/RG020i004p00851.

[NN09] Nakanishi, Mikio, and Hiroshi Niino. "Development of an Improved Turbulence Closure Model for the Atmospheric Boundary Layer." J. Met. Soc. Japan. Ser. II, vol. 87, no. 5, 2009, pp. 895–912. https://doi.org/10.2151/jmsj.87.895.

[HB15] He, P, and S Basu. Mesoscale Modeling of Optical Turbulence (Cn2) Utilizing a Novel Physically-Based Parameterization. SPIE. 2015, p. 96140K. https://doi.org/10.1117/12.2188227.

[O19] Olson, J. B., et al. A Description of the MYNN-EDMF Scheme and the Coupling to Other Components in WRF–ARW. Technical report, OAR GSD-61, National Ocean and Atmospheric Administration, 2019, https://doi.org/10.25923/n9wm-be49.

[BH22] Basu, S, and A. A. M. Holtslag. "Revisiting and Revising Tatarskii's Formulation for the Temperature Structure Parameter (CT2) in Atmospheric Flows." Env. Fluid Mech., vol. 22, no. 5, Oct. 2022, pp. 1107–19. https://doi.org/10.1007/s10652-022-09880-3.

[PSB23] Pierzyna, M, et al. "Π-ML: A Dimensional Analysis-Based Machine Learning Parameterization of Optical Turbulence in the Atmospheric Surface Layer." Opt. Let., vol. 48, no. 17, Sept. 2023, pp. 4484–87. https://doi.org/10.1364/OL.492652.

[PHSB24] Pierzyna, M, et al. "Intercomparison of Flux, Gradient, and Variance-Based Optical Turbulence (C_n^2) Parameterizations." App. Opt., vol. 63, no. 16, June 2024, pp. E107–19., https://doi.org/10.1364/AO.519942.