Supplement of

Can machine learning improve the model representation of turbulent kinetic energy dissipation rate in the boundary layer for complex terrain?

Nicola Bodini et al.

Correspondence to: Nicola Bodini (nicola.bodini@nrel.gov)

The copyright of individual parts of the supplement might differ from the CC BY 4.0 License.
1. Proof on approximation for MYNN parameterization of TKE dissipation rate

Actual definition:
\[
\frac{1}{L_M} = \frac{1}{L_S} + \frac{1}{L_T} + \frac{1}{L_B}
\]

Approximation:
\[
\frac{1}{L_M} \approx \frac{1}{L_S}
\]

Thesis: the approximation will always overestimate $L_M$

Proof:
Let’s compare the right-hand sides of the definition of $L_M$ and the approximation followed in our analysis and suppose our thesis is true. Also, all length scales are strictly positive.

\[
\left( \frac{1}{L_S} + \frac{1}{L_T} + \frac{1}{L_B} \right)^{-1} < L_S
\]

\[
\frac{L_T L_B}{L_T L_B + L_S L_B + L_S L_T} < L_S
\]

\[
\frac{L_T L_B}{L_T L_B + L_S L_B + L_S L_T} < 1
\]

\[
L_T L_B < L_T L_B + L_S L_B + L_S L_T
\]

\[
0 < L_S L_B + L_S L_T
\]

Which is always true since all length scales are strictly positive.
2. Bias in MYNN parameterization using data from the three 100-m met towers

*Figure S1: Mean bias in the MYNN-parameterized log $\epsilon$ at different heights, as calculated from the sonic anemometers on the three 100-m towers at Perdigão.*
3. Input and output features distributions

Figure S2: Distributions of 30-minute average wind speed, log(TKE), log($u_*$), and $\text{sign}(z/L)\text{log}(1+\text{abs}(z/L))$ from the sonic anemometers at Perdigão.
Figure S3: Distributions of $\sigma(z_{terr})$, vegetation height, and 30-minute average $\log \epsilon$ from the sonic anemometers at Perdigão.
4. Feature importance as a function of height

Figure S4: Variability of the random forest feature importance with height, derived from several random forests, each trained and tested with data from all the sonic anemometers at a single height only.