Thesis

Emilio Dorigatti

March 8, 2018

Summary

todo

Contents

1 Introduction **5**
1.1 Problem . . . 5
1.2 Research Question . . . 6
1.3 Ethics and Sustainability . . . 6
1.4 Outline . . . 6

2 Background **7**
2.1 Fluid Dynamics . . . 7
2.1.1 Laminar and Turbulent flow . . . 7
2.1.2 The Boundary Layer . . . 8
2.2 The Atmospheric Boundary Layer . . . 9
2.2.1 Surface Fluxes . . . 10
2.2.2 The Turbulence Kinetic Energy Budged . . . 11
2.2.3 Monin-Obukhov Similarity Theory . . . 13
2.3 Machine Learning . . . 14
2.3.1 Learning Theory . . . 14
2.3.2 Parameter Estimation for Regression . . . 14
2.3.3 Models . . . 16

3 Method **17**
3.1 Data Collection . . . 17
3.1.1 Wind Measurement . . . 17
3.1.2 Air Temperature Measurement . . . 18
3.1.3 Eddy Correlation . . . 18
3.1.4 Gap Filling . . . 19
3.2 Monin-Obukhov Similarity Theory . . . 19
3.2.1 Profile Computation . . . 20
3.2.2 Model Fitting . . . 20

4 Results **23**

5 Discussion **25**
5.1 Limitations and Future Work . . . 25

Chapter 1

Introduction

1.1 Problem

Climatology, or climate science, studies the long-term average weather conditions. In the last few decades, climate scientists found solid evidence of ongoing changes in Earth's climate, most notably, a general increase in average temperature; in the long run, this and other changes can potentially have a devastating impact on life on our planet. Regardless of the causal effect of human activities on it, having a solid understanding of the climate allows us to find the best way to mitigate its change, and possibly prevent it completely.

Climate science is an extremely difficult field, for a number of reasons. First, climate is evident, by definition, only over long periods of time and large distances, making the usual scientific approach of testing hypotheses by conducting experiments inapplicable; instead, climate scientists have to rely on historical data. Second, the atmosphere is a complex and chaotic system, which must be described through systems of nonlinear differential equation. They can be used to create numerical simulations, and the resulting predictions compared to historical data to assess the accuracy of the theory. Furthermore, the chaotic character of the atmosphere makes it impossible to study parts of it in isolation from others, as small scale phenomena affect large scale ones, and vice versa. In spite of this, it is useful to split the atmosphere in vertical layers and horizontal zones, in order to differentiate among conditions and phenomena typically occurring in one area or the other.

Usually, the troposphere is referred to as the lowest layer of the atmosphere, but it can actually be subdivided in more sub-layers, the lowest of which is the *atmospheric boundary layer*: it is the region of the atmosphere that is affected by the conditions of the surface. Most human activities happen in this layer, and it is responsible for a large part of the diffusion and transport of aerosol such as, among others, pollutants. Yet, the physics governing the atmospheric boundary layer is not fully understood, and the theory is lacking. One important issues in the study of the atmospheric boundary layer is the derivation of flux-profile relationships for wind and temperature: they essentially relate the transport of momentum and heat by the wind (flux) with the change of wind speed/temperature with altitude (profile). The state of the art relationships are defined by the Monin-Obukhov Similarity theory in terms of the instability parameter ξ, computed as the height above surface scaled by turbulence due to horizontal wind and vertical air movement due to variations in heat. Difficulties in measurements of relevant quantities make this theory accurate only up to 10-20%, and applicable in a restricted set of conditions.

In stark contrast to the traditional, top-down approach of science, recent developments in information technology made bottom-up approaches possible. In this new way of thinking, existing data is used to automatically infer the ”best” explanation for the measurements at hand, the underlying laws that originated that data. The field that makes this possible is called Machine Learning: it takes advantage of several methods coming from statistics, information theory, optimization theory, etc., to make computers learn from examples. Together with Natural Language Processing and Automated Planning, it is one of the three main branches of Artificial Intelligence, the sub-field of Computer Science that studies ways of making machines behave intelligently.

todo more fluff about ml ?

1.2 Research Question

With the availability of micro-meteorological data from specialized observation sites such as Cabauw, in the Netherlands, and the recent developments in Machine Learning, a natural question to ask is whether the Monin-Obukhov similarity theory can be improved. More specifically, we pose the following

Research Question: Is it possible to use the data from the Cesar database to improve the Monin-Obukhov model of the flux-profile relationships, by using more predictors besides the instability parameter?

A positive answer to this question would contribute to improve the quality of current large scale climate simulations. Since errors in the simulation accumulate over time, even slight improvements can greatly improve the accuracy of the results after years of simulated time.

1.3 Ethics and Sustainability

reproducibility: using open data, following standards of reproducible research (open source and jupyter notebooks)

1.4 Outline

todo

Chapter 2

Background

This chapter introduces the basic concepts the reader should be qualitatively familiar with, in order to understand the content of this thesis. It is assumed readers are already knowledgeable of simple mathematical concepts, such as calculus, linear algebra, statistics, and probability theory. Readers acquainted with the material should feel free to skip this chapter.

2.1 Fluid Dynamics

Fluid dynamics is the discipline that studies the flow of fluids; it has several branches that study different fluids, such as aerodynamics (the study of air motion) and hydrodynamics (the study of water motion). These disciplines are routinely applied when designing cars, airplanes, ships, pipelines, etc.

2.1.1 Laminar and Turbulent flow

There are two distinct ways in which particles in a fluid can move: laminar flow and turbulent flow. In the former, all the particles move orderly, perhaps with a different speed, but all in the same direction, whereas in the latter the movement of particles is highly chaotic and unpredictable, and tends to dive rise to eddies of varying sizes. People are most familiar with the distinction between the two through the smoke rising from a cigarette, which starts smooth and becomes turbulent shortly thereafter, as in figure 2.1. The kind of flow in under specific conditions can be predicted using the Reynolds number *Re*, which is the ratio between inertia forces, favoring turbulent flow, and viscosity forces, stabilizing the fluid towards laminar motion:

$$Re = \frac{\rho u L}{\mu} = \frac{uL}{\nu}$$

With ρ the density of the fluid, u its velocity, L a characteristic linear dimension of the system under consideration, μ and ν the kinematic and dynamic viscosity of the fluid. The viscosity describes, intuitively, how much the molecules of the fluid tend to stick together and resist motion by generating drag. For example, water has low viscosity, and honey has high viscosity.

Figure 2.1: Smoke from a cigarette, and the transition from laminar to turbulent flow.

Since turbulence is random, it is usually studied in terms of the statistical properties of physical quantities through the Reynolds decomposition; given a quantity $a(s,t)$ which varies in space and time, we can compute its average

$$\overline{a}(s) = \frac{1}{T}\int_{T_0}^{T_0+T} a(s,t)dt$$

and the deviation from the average

$$a'(s,t) = a(s,t) - \overline{a}(s)$$

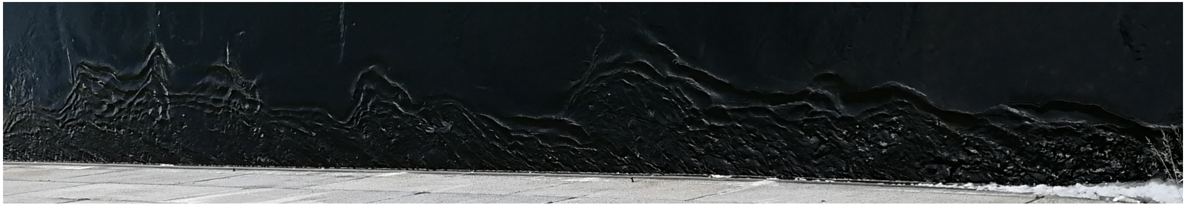
By definition, $\overline{a'} = 0$, which means that all the effects of turbulence are contained in a'. Common statistical properties such as variance and covariance are expressed respectively as $\overline{a'a'}$ and $\overline{a'b'}$.

2.1.2 The Boundary Layer

In the context of fluid dynamics, the boundary layer studies the behavior of a fluid when it is flowing close to a solid surface. Imagine a laminar flow close to a solid surface; because of viscosity, the molecules flowing near the surface move slower, and, in the limit, the velocity of the molecules in direct contact with the surface is 0 (this is called the *non-slip condition*). Thus, the velocity of the fluid increases smoothly, continuously and monotonously with the distance from the solid, until it reaches the *free-flow* velocity, after which it stays constant. The region close to the surface, where the fluid moves slower, is called the *boundary layer*, and is the region where the viscosity of the fluid influences its motion. Its height δ can be defined when the local velocity surpasses a certain threshold, such as 99% of the free-flow velocity.

The variation of velocity with distance from the surface, $\partial\overline{u}/\partial z$, is called *shear*, and, together with viscosity, determines the materialization of turbulence in the flow. Every layer of fluid is squeezed between a faster moving layer above and a slower moving below; in high shear conditions, this causes high stress on the particles, and prevents them from moving orderly, thus leading to turbulent motion. Figure 3.1 shows turbulence forming close to the

Figure 2.2: Turbulent boundary layer at the edge of a canal; the end of the BL is clearly visible, where the flow transitions from turbulent to laminar. Water flows from right to left.



wall in a canal, where the water flows from right to left. Viscosity and the no-slip condition prevent this phenomenon to arise in a region very close to the solid surface, called the *laminar (or viscuous) sub-layer*, where we still find laminar motion.

The strength of the turbulence is proportional to $u_{rms} = (\overline{u'^2})^{1/2}$, which is, in turn, proportional to the shear. Again, because of the no-slip condition, u_{rms} is zero at $z = 0$, increases in the laminar sub-layer, and decreases to 0 at the end of the boundary layer, assuming laminar flow outside of it. Higher free-stream velocity generates higher shear, more turbulence, and a thinner laminar sub-layer. The strength of turbulence can be written in units of velocity, resulting in the *friction velocity*, computed as $u_* = (\tau/\rho)^{1/2} = (\nu \cdot \partial\overline{u}/\partial z)^{1/2}$, where τ is the shear stress, ρ is the density of the fluid, $\nu = \mu/\rho$ is the kinematic viscosity, and μ the dynamic viscosity. Therefore, the friction velocity increases with shear and viscosity, and decreases with density; it is proportional to the free-stream velocity and the turbulence strength, and inversely proportional to the height of the laminar sub-layer.

The mean velocity $\overline{u}$ increases linearly within the laminar sub-layer, then logarithmically until the end of the boundary layer, thus the shear decreases further away from the surface. In the logarithmic sub-layer, the velocity is computed as $\overline{u}(z) = u_*(\log z - \log z_0)/\kappa$, where z_0 is the characteristic roughness of the surface, and κ is the von Karman's constant, whose value is around 0.4 [citation needed]. The characteristic roughness depends on the texture of the surface, and its relationship with the height δ_s of the laminar sub-layer; if the roughness scale is smaller than δ_s, the logarithmic velocity profile is not affected by the texture, because the laminar sub-layer completely covers the variations on the surface, and we have the so-called smooth turbulent flow. If, on the contrary, the bumps in the surface are larger than δ_s, the laminar sub-layer follow the profile of the surface, and the logarithmic velocity profile is altered depending on the texture, a regime called rough turbulent flow.

2.2 The Atmospheric Boundary Layer

The atmosphere is composed by air, which is behaves like fluid. Therefore, close to the Earth's surface, in the region called *atmospheric boundary layer*, we find the same effects described in the previous section. Additionally, there are other phenomena that complicate things further, such as the temperature of the surface, which changes widely from day to night and from Summer to Winter, the rotation of the Earth, the varying roughness of the surface, due to cities and mountains, etc. The effect of the surface on the first few hundred meters of the atmosphere is the main focus of *boundary layer meteorology*.

The height of the atmospheric boundary layer (hereafter abbreviated as ABL) typically varies between 100 and 1000 meters, highly depending on the conditions, and it is always turbulent. There are three main instabilities driving turbulence in the ABL:

- Shear instability: caused by shear, the mechanism described in the previous section. This happens at high Reynolds number, and, by using typical values for the ABL, we find *Re* well above 10^6.

- Kelvin-Helmholtz instability: occurs when there is a difference of density and velocity in different layers of flow. This is the mechanism that generates, for example, waves in ponds, lakes, and oceans.

- Rayleigh-Bernard instability: is caused by the decrease of potential density[1] with height, or, in other words, when warm fluid is below cold fluid; the warm fluid will rise, and the cold fluid will drop, a phenomenon called *convection*. During hot Summer days, the surface is much warmer than the air, thus the air close to the surface will heat and tend to rise.

Turbulence has the very important role of transport and mix of air properties, such as heat, moisture, particles, aerosols, etc. This is especially true in *unstable* conditions, when the air moving upwards (e.g. because it is warmer) is less dense than the air moving downwards; when the contrary happens, the ABL is called *stable*.

The ABL can be divided in two main sub-layers: the inner surface layer and the outer Ekmann layer. This distinction is mainly done based on the scale of the dominating turbulent eddies: they are much smaller than the height of the ABL in the surface layer, and of comparable size in the outer layer.

It is very important to have a macroscopic understanding of the turbulent processes in the ABL, because they happen at length and time scales too small to be simulated in global climate models. The process of expressing the result of turbulence as a function of large scale parameters is called parametrization; having realistic models is essential in order to conduct precise simulations of the global climate in the scale of tens or hundreds of years, because errors tend to accumulate and amplificate as the simulation goes on. Other fields that benefit from the study of the ABL are urban meteorology (interested in the dispersion of pollutants), agricultural meteorology (interested in the formation of frost and dew, the temperature of the soil, etc.), aviation (predict fog and strong winds), and so on.

2.2.1 Surface Fluxes

A flux measures the amount of a physical quantity that flows through a surface. In the context of boundary layer meteorology, we are interested in the flows through the surface of earth, because, through them, the surface and the atmosphere exchange energy; these fluxes are thus measured in $\mathrm{W\,m^{-2}}$. The main source of energy for the surface is long-wave radiation coming from the sun, and short-wave radiation coming from the atmosphere and the clouds. A small

[1] the potential density is the density that a parcel of air would attain if brought at a standard reference pressure adiabatically, i.e. disallowing exchanges of heat with its surroundings. Potential density is useful to compare densities irrespectively of pressure, i.e. altitude

amount of long-wave radiation is emitted from the surface, therefore let the net radiative flux be R, positive when the surface gains energy.

The main fluxes by which the surface loses energy to the atmosphere are called the turbulent flux of *sensible heat* H, also called kinematic heat flux, and the turbulent flux of *latent heat* λE, also called kinematic flux of water vapor/moisture. The difference between the two is that the former causes an actual change of temperature, whereas the latter does not affect temperature[2]. The main causes of sensible heat fluxes are conduction and convection, whereas the main cause of latent heat fluxes is water movement: condensation, evaporation, melting, etc.

The final flux of interest is the soil heat flux G, which is simply the heat "absorbed" by the surface and not given to the atmosphere. These four fluxes are linked by the surface energy balance equation:

$$R = H + \lambda E + G$$

which states that the total incoming energy R must be equal to the energy given back to the atmosphere $H + \lambda E$ (not counting long-wave radiation, which is accounted to in R) plus the energy absorbed by the surface G.

The turbulent fluxes H and λE are constant in the surface layer [todo add citation]. Experimentally, the energy balance is not always achieved [todo cite from cabauw] due to difficulty in measuring fluxes due to eddy correlation being inaccurate (verify this)

2.2.2 The Turbulence Kinetic Energy Budged

Kinetic energy is energy stored in form of movement: faster or heavier objects have more kinetic energy than slower or lighter ones. The Reynolds decomposition allows us to decompose the kinetic energy of turbulent flows in two terms: one caused by the mean flow, and one caused by turbulence. This decomposition can be justified by examining the temporal spectrum of kinetic energy, shown in figure 2.3. Four peaks are visible, corresponding to different sources of kinetic energy: turbulence, day-night cycle, westerlies[3], and seasons. Importantly, there are few sources of kinetic energy in the 30 minutes to one hour time scale; this so-called spectral gap allows us to separate between turbulence and other sources of fluctuations in the atmosphere.

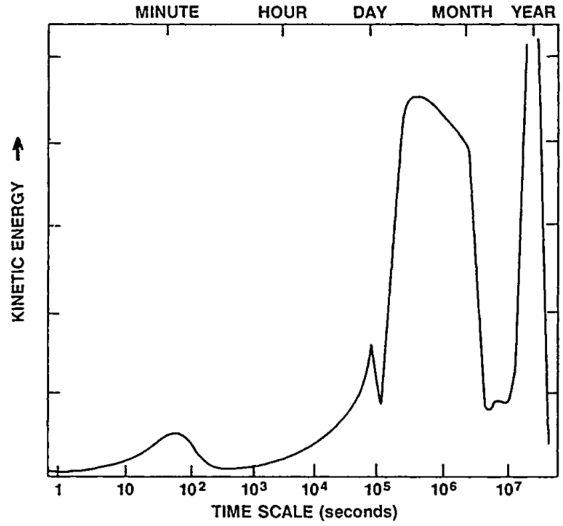
From now on, we will use a coordinate system with the x axis aligned to the average horizontal wind direction, the y axis perpendicular to it, and the z axis pointing away from the surface. Then, we will use the letters u, v and w to denote the components of the wind along the axes x, y and z respectively; clearly, $\overline{v} = 0$. Eddy fluxes can then be described in terms of covariances: let θ denote the potential temperature[4], then $\overline{w'\theta'}$ is the turbulent heat flux, i.e. the sensible heat flux in the vertical due to wind. Usually the ABL studied assuming

[2] imagine a pot of boiling water; selecting a higher temperature on the stove will not increase the temperature of water above 100 °C, but will make it boil faster. The additional heat introduced in the system is dissipated through increased evaporation

[3] winds blowing from the east towards the west in the mid-latitudes

[4] the potential temperature is final temperature after bringing a parcel of air to a standard pressure adiabatically, i.e. not allowing exchange of temperature with the surroundings. It is a useful mean to compare temperatures irrespectively of pressure, i.e. altitude

Figure 2.3: Change of atmospheric kinetic energy at different time-scales. The peaks in the scale of days and months and years are due to the day-night and Summer-Winter cycles, the peaks in the monthly scale are due to baroclinic instability in the mid-latitude westerlies, and the peaks at one minute are due to convection and atmospheric turbulence [1, 2]



homogeneous horizontal conditions, because they vary on a length scale larger than the height of the ABL. Because of this, the horizontal eddy correlations $\partial\overline{u'a'}/\partial x$ and $\partial\overline{v'a'}/\partial y$ are usually of negligible intensity, and are thus ignored. Note that this is not necessarily true when clouds are involved.

It is important to notice that turbulence is dissipative in nature. Consider a hot Summer day, where air is warmer close to the surface, and a circular eddy moving some air up and some down, so that the average motion is zero. The parcel of air moving up ($w' > 0$) ends up being warmer than its surroundings ($\theta' > 0$), while the one moving down ($w' < 0$) will be colder ($\theta' < 0$); the result is a new transport of heat through the eddy: $\overline{w'\theta'} > 0$. On the contrary, imagine a cold night, where the air close to the surface is colder; the same eddy would transport a colder parcel of air upwards, and a warmer one downwards. In both cases, the end result would be a net transport of heat without transport of mass. Because of the ??? law, the eddy must lose energy, and thus dissipate over time.

Since turbulence changes over time, we are more interested in the change of kinetic energy, the turbulent kinetic energy budget. A full derivation is out of the scope of this work, but its final form [3] can be derived from prime physical principles, resulting in

$$\frac{\partial \overline{e'^2}}{\partial t} = \underbrace{\overline{u'w'}\frac{\partial \bar{u}}{\partial z}}_{P} - \underbrace{\frac{g}{T}\overline{w'\theta'}}_{B} + \underbrace{\frac{\partial}{\partial z}\frac{\overline{w'e'^2}}{2}}_{T_t} + \underbrace{\frac{1}{\rho}\frac{\partial \overline{p'w'}}{\partial z}}_{T_p} + \varepsilon \tag{2.1}$$

Where $e'^2 = u'^2 + v'^2 + w'^2$. The P term is the production due to shear, B is the production due to buoyancy, T_t is the turbulent transport of TKE by large-scale eddies, T_p is the transport due to pressure, and ε is molecular dissipation due to viscosity. P and B are the most prominent terms, and the transport terms are close to zero in neutral conditions.

TODO elaborate on 3.4, 3.5 (reason why P and B have opposite sign), 3.6,

2.2.3 Monin-Obukhov Similarity Theory

One of the factors to distinguish laminar from turbulent flow is the length scale L of the system. This length scale for the ABL was derived by A. M. Obukhov in 1946, and forms the basis of the similarity theory. According to this theory, the normalized wind and temperature profiles can be expressed as an unique function of $\xi = z/L$:

$$L = -\frac{u_*^3}{\kappa \frac{g}{T_v} \frac{Q}{\rho c_\rho}} = -\frac{u_*^3 T_v}{\kappa g \overline{w' \theta_v}} \tag{2.2}$$

$$\frac{\partial \bar{u}}{\partial z} \frac{kz}{u_*} = \phi_m(\xi) \tag{2.3}$$

$$\frac{\partial \bar{\theta}}{\partial z} \frac{kz}{T_*} = \phi_h(\xi) \tag{2.4}$$

With

- $g = 9.81\,\mathrm{m\,s^{-2}}$ the acceleration due to Earth's gravity
- $\kappa = 0.4$ the von Karman constant
- T_v virtual temperature[5], obtained as

  $$T_v = T \frac{1 + r_v/\varepsilon}{1 + r_v} = T(1 + 0.61 \cdot q)$$

  Where r_v is the mixing ratio, $q = r_v/(1+r_v)$ the specific humidity, and ε is the ratio of the gas constants of dry air and water vapor, roughly 0.622.
- ρ the air density, computed from the pressure P and the specific gas constant for dry air $R = 287.058\,\mathrm{J\,kg^{-1}\,K}$ as

  $$\rho = \frac{P_0}{RT_v}$$
- c_ρ specific heat of dry air, $1005\,\mathrm{J\,kg^{-1}\,K^{-1}}$ at $300\,\mathrm{K}$
- Q the buoyancy flux, approximated by $H + 0.07\lambda E$ and measured in $\mathrm{W\,m^{-2}}$
- $\overline{w'\theta_v} = Q/\rho c_\rho$ the flux of virtual potential temperature, measured in $\mathrm{K\,m\,s^{-1}}$
- $T_* = -\overline{w'\theta}/u_*$

[5] potential temperature of dry air if it had the same density as moist air. It allows to use formulas for dry air when the air is not dry.

The stability parameter ξ is positive for stable conditions, where wind shear dominates the production of TKE, and negative for unstable conditions, where buoyancy is the main contributor to turbulence. It approaches 0 in the limit of neutral stratification (i.e. $\partial\overline{\theta}/\partial z = 0$), because the temperature flux goes to 0 causing L go to infinity.

The universal functions ϕ_m and ϕ_h must be determined experimentally. This is no easy task, and considerable effort has been devoted to it; one of the greatest difficulties lies in obtaining accurate and unbiased measurements, especially the fluxes. [hogstrom 1988] is a meta-study that aggregates and improves many previous results, and suggests the following expressions:

$$\phi_m(\xi) = \begin{cases} (1-19.3\xi)^{-1/4} & -2 < \xi < 0 \\ 1+6\xi & 0 < \xi < 1 \end{cases} \tag{2.5}$$

$$\phi_h(\xi) = \begin{cases} 0.95(1-11.6\xi)^{-1/2} & -2 < \xi < 0 \\ 0.95+8\xi & 0 < \xi < 1 \end{cases} \tag{2.6}$$

The Monin-Obukhov similarity theory is only applicable in the surface layer, at heights much larger than the aerodynamic roughness length, and with $|\xi| < 2$; even under ideal conditions, the predictions of this theory are accurate up to 10-20% [foken 2006].

2.3 Machine Learning

todo

2.3.1 Learning Theory

true risk vs empirical risk, kfold cv, nested cv

2.3.2 Parameter Estimation for Regression

A common assumption in the regression setting is that the observations are corrupted by additive Gaussian white noise, i.e. $t_n = f_n + \varepsilon_n$, where f_n is the ”true” value, and $\varepsilon_n \sim \mathcal{N}(0, \beta^{-1})$ is the noise. Let $y_n = y(\boldsymbol{x}_n; \boldsymbol{\theta})$ be the model's prediction for the sample $\boldsymbol{x}_n$, then we can write the probability of observing t_n, were $y_n = f_n$, as:

$$p(t_n|\boldsymbol{x_n}, \boldsymbol{\theta}, \beta) = \mathcal{N}(t_n|y(\boldsymbol{x_n}; \boldsymbol{\theta}), \beta^{-1}) \tag{2.7}$$

This probability is called *likelihood* of the observation t_n, under the model $y(\cdot;\cdot)$ with parameters $\boldsymbol{\theta}$. Since the training data is assumed to be independent and identically distributed, the likelihood of the whole training set $\boldsymbol{X} = [\boldsymbol{x}_1 \cdots \boldsymbol{x}_N]^\mathsf{T}$ is

$$p(\boldsymbol{t}|\boldsymbol{X}, \boldsymbol{\theta}, \beta) = \prod_{n=1}^{N} \mathcal{N}(t_n|y(\boldsymbol{x_n}; \boldsymbol{\theta}), \beta^{-1}) \tag{2.8}$$

Then, a sensible choice for the model parameters $\boldsymbol{\theta}$ is to choose the ones that give the largest dataset likelihood; this estimation method is called *maximum likelihood* (MLE). Note that, since the goal is to predict f_n, there is a high probability that the best parameters would *not* be the ones with maximum likelihood. When a model learns the noise in the training data, it cannot generalize well to new, unseen data, because the noise is random. This situation is known as *overfitting*, and tends to happen when the model is too complex relative to the amount of data available for training.

The risk of overfitting can be reduced with a number of *regularization* strategies. A widely used strategy consists in including a prior distribution on the parameters of the model, and maximizing their posterior distribution, computed using Bayes theorem:

$$\begin{aligned} p(\boldsymbol{\Theta}|\boldsymbol{t},\boldsymbol{X}) &= \frac{p(\boldsymbol{X}\boldsymbol{t}|\boldsymbol{\Theta})\cdot p(\boldsymbol{\Theta})}{p(\boldsymbol{X},\boldsymbol{t})} \\ &\propto p(\boldsymbol{t}|\boldsymbol{X},\boldsymbol{\Theta})\cdot p(\boldsymbol{X}|\boldsymbol{\Theta})\cdot p(\boldsymbol{\Theta}) \\ &\propto p(\boldsymbol{t}|\boldsymbol{X},\boldsymbol{\Theta})\cdot p(\boldsymbol{\Theta}) \end{aligned} \tag{2.9}$$

where we removed $p(\boldsymbol{X},\boldsymbol{t})$ and $p(\boldsymbol{X}|\boldsymbol{\Theta}) = p(\boldsymbol{X})$ because they are constant for a given dataset, and we are not interested in the exact probability, only where it reaches its maximum value. This parameter estimation procedure is called *maximum a posteriori* (MAP). Two commonly used prior distributions are the multivariate Laplace and the multivariate Normal, leading respectively to L1 and L2 regularization, when centered and symmetrical/spherical.

The maximization of the posterior can be done conveniently by maximizing its logarithm; this gives expressions that are easier to handle analytically, and give less numerical problems when computed. The MAP problem can be formulated as follows:

$$\boldsymbol{\Theta}^* = \underset{\boldsymbol{\Theta}}{\operatorname{argmax}}\ \log p(\boldsymbol{t}|\boldsymbol{X},\boldsymbol{\Theta}) + \log p(\boldsymbol{\Theta}) \tag{2.10}$$

If we assume a Gaussian likelihood like the one in equation 2.8, and a spherical Normal prior distribution $p(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{0},\lambda\boldsymbol{I})$, the MAP estimation of equation 2.9 becomes, after removing unnecessary constant terms,

$$\boldsymbol{\theta}^* = \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\ L(\boldsymbol{\theta}) = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \sum_{n=1}^{N} (y(\boldsymbol{x}_n,\theta) - t_n)^2 + \lambda\boldsymbol{\theta}^\intercal\boldsymbol{\theta} \tag{2.11}$$

Where L is the *loss function*, in this case the sum-of-squares error. It is customary to use the mean squared error instead of the sum of squares because it results in smaller numbers and is readily interpreted; being only a constant factor away from 2.11, it does not transcend the essence of maximum likelihood. The parameter β can be estimated as well in a similar way, if necessary, and a fully Bayesian treatment allows to estimate λ, too. Depending on the model $y(\boldsymbol{x}_n,\boldsymbol{\theta})$, 2.11 can be solved analytically to yield a closed-form solution for $\boldsymbol{\theta}$; this is the case for linear regression, where $y(\boldsymbol{x}_n,\boldsymbol{\theta}) = \boldsymbol{\theta}^\intercal\boldsymbol{x}_n$ and the optimal parameters are $\boldsymbol{\theta}^* = (\boldsymbol{X}^\intercal\boldsymbol{X} + \lambda\boldsymbol{I})^{-1}\boldsymbol{X}^\intercal\boldsymbol{t}$.

When this is not possible, a general approach called *gradient descent* is employed; the gradient of a function computed at a given location ”points” to the steepest direction where the function’s value increases. By repeatedly following the gradient, it is possible to reach a local maximum, and, in the opposite direction, a local minimum:

$$\boldsymbol{x}_{n+1} := \boldsymbol{x}_n - \eta\cdot\nabla f(\boldsymbol{x}_n) \tag{2.12}$$

the series $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ is guaranteed to converge, provided that the *learning rate* η goes to 0 as n goes to infinity [citation]. Vanilla gradient descent can be greatly improved with a number of techniques, such as momentum [citation], adaptive learning rate [citation], use of the Hessian [citation], and so on. The gradient of $L(\boldsymbol{\theta})$ is

$$\nabla_{\boldsymbol{\theta}} \log p(\boldsymbol{\theta}|\boldsymbol{t},\boldsymbol{X}) = 2\sum_{n=1}^{N} \big(y(\boldsymbol{x}_n;\theta) - \boldsymbol{t}_n\big) \cdot \nabla_{\boldsymbol{\theta}} y(\boldsymbol{x}_n;\theta) + 2\lambda\boldsymbol{\theta} \tag{2.13}$$

And its Hessian is

$$\nabla^2_{\boldsymbol{\theta}} \log p(\boldsymbol{\theta}|\boldsymbol{t},\boldsymbol{X}) = 2\sum_{n=1}^{N} \left[\big(\nabla_{\boldsymbol{\theta}} y(\boldsymbol{x}_n;\boldsymbol{\theta})\big)\big(\nabla_{\boldsymbol{\theta}} y(\boldsymbol{x}_n;\boldsymbol{\theta})\big)^{\intercal} + \big(y(\boldsymbol{x}_n;\boldsymbol{\theta}) - t_n\big)\nabla^2_{\boldsymbol{\theta}} y(\boldsymbol{x}_n;\theta)\right] + 2\lambda\boldsymbol{I} \tag{2.14}$$

2.3.3 Models

k-Nearest Neighbors

todo

Linear Regression

todo

Random Forests

todo

Neural Networks

todo

Chapter 3

Method

3.1 Data Collection

The Cabauw Experimental Site for Atmospheric Research[1] (Cesar) is a consortium formed by eight Dutch institutes and universities, which collaborate to operate and maintain an observatory for micro-meteorological conditions near the village of Cabauw, the Netherlands. The data collected characterizes the state of the atmosphere and the soil, and their interaction via radiation and surface fluxes.

The observatory is surrounded by fields and no urban agglomerations is present within 15 kilometers; the land is flat with changes of altitude within a few meters over 20 kilometers. The main mast is 213 meters high and offers measurement levels every 20 meters; at each level there are three booms of length 10 meters that allow observations unobstructed by the main mast. There are three additional smaller masts of height 10 and 20 meters located close to the main mast, in order to obtain undisturbed measurements at the lower levels, and facilities to perform soil and surface observations.

The main focus of this project is on wind and temperature profiles, and the turbulent fluxes of sensible and latent heat. Additional variables, such as temperature and humidity, are needed to compute quantities of interest, most importantly the Obukhov length, and as possible predictors for the universal functions. There is one measurement for each variable every ten minutes, and missing measurements are gap-filled with a number of techniques. The data collected is always visually validated by an operator, which marks suspect or invalid sections of data

The Cesar observatory provides full information regarding data collection[2], what follows is a brief summary of the most relevant sections.

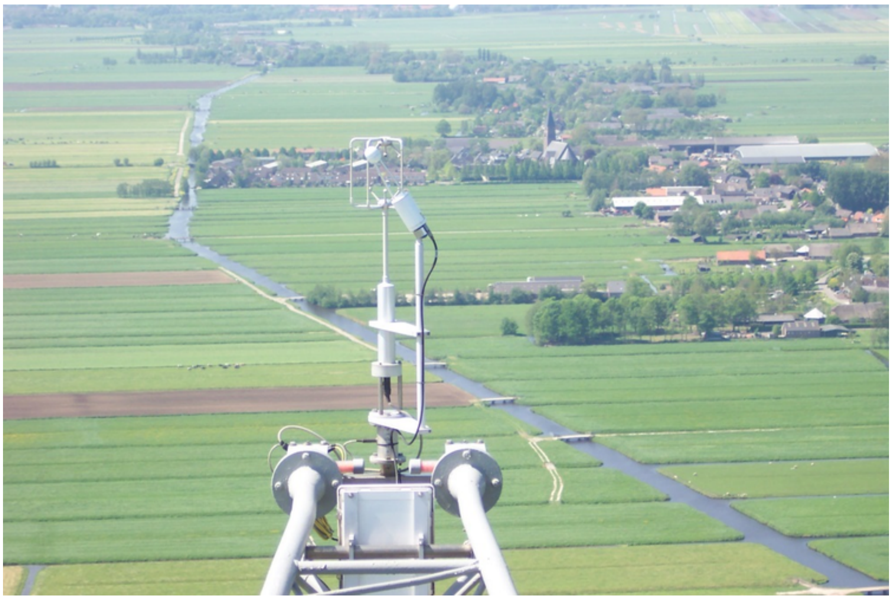
3.1.1 Wind Measurement

Wind speed and direction are measured at heights of 200, 140, 80, 40, 20 and 10 meters, in either two or all three booms available. The wind vane that measures direction has a resolution of 1.5°, and the cup anemometer that measures wind speed has an accuracy of the

[1] http://www.cesar-database.nl/

[2] http://projects.knmi.nl/cabauw/insitu/observations/documentation

Figure 3.1: Sonic anemometer to measure wind and optical open-path sensor to measure humidity at the 180m level of the Cabauw mast.



largest between 1% and $0.1\,\mathrm{m\,s^{-1}}$. The threshold sensitivity of both instruments is lower than $0.5\,\mathrm{m\,s^{-1}}$, and they report inaccurate measurements for up to $3\,\mathrm{m\,s^{-1}}$ [4].

For every ten minutes interval, the measurement comes from the instrument that is best exposed to the wind, and less affected by the obstruction caused by the mast. Corrections are then applied to the raw measurements to further attenuate the disturbance by the tower, following [5].

3.1.2 Air Temperature Measurement

Air temperature is measured at heights of 200, 140, 80, 40, 20, 10 and 1.5 meters. todo write more

3.1.3 Eddy Correlation

The eddy correlation technique is used to compute the turbulent surface fluxes of sensible and latent heat, as well as momentum and CO_2, starting from fluctuations in wind, temperature, and humidity.

These measurements are obtained with, respectively, a sonic anemometer, a sonic thermometer, and an optical open-path sensor. Sonic anemometers measure the wind speed by leveraging the fact that the speed of sound in free air is affected by the speed of the air itself; since the speed of sound is known, the wind speed can be easily recovered from the time a sound impulse takes to travel a short distance. By measuring the wind velocity along three orthogonal paths, the full wind vector can be recovered. Sonic thermometers work similarly, by leveraging the fact that the speed of sound is affected by the temperature of the medium it travels in. These instruments can take up to 100 measurements per second. Optical open-path sensors quantify the amount of water vapor and carbon dioxide in the air by emitting a ray of infrared light and measuring its intensity 10 to 20 centimeters further away. H_2O and CO_2 molecules in the air absorb electromagnetic radiation at known frequencies, thus the concentration of water vapor and carbon dioxide can be inferred by measuring the attenuation at these wavelengths.

The eddy correlation technique measures fluxes by computing their sample covariance with the vertical wind speed. Let w_t be the vertical wind speed at time t, then the turbulent vertical flux for the quantity a is computed as follows:

$$F_a = \frac{1}{T_2 - T_1} \sum_{t=T_1}^{T_2} (w_t - \overline{w})(a_t - \overline{a})$$

Where $\overline{w}$ and $\overline{a}$ are the averages of w_t and a_t for $T_1 \leq t \leq T_2$. The fluxes in the Cesar database are computed every ten minutes, with 10 measurements per second.

The eddy correlation technique is far from perfect, see [6] for a detailed summary of issues.

3.1.4 Gap Filling

With gap-filling, missing measurements are replaced by synthetic values. The gap-filling method depends on the missing parameter and the duration of the period where data is not available. There are two classes of parameters: forcing parameters, which include wind, temperature, specific humidity, incoming radiation and rain, and validation parameters, which include the surface fluxes, outgoing radiation, and friction velocity.

For less than two hours of missing measurements, both forcing and evaluation parameters are gap-filled by interpolation of nearby values. For longer periods, forcing parameters are derived by transforming measures obtained from the nearby site of De Bilt, which are themselves gap-filled. Evaluation parameters are computed with a vegetation model that uses the forcing parameters as input.

3.2 Monin-Obukhov Similarity Theory

This section describes how the data from the Cesar database is used to reproduce the Monin-Obukhov similarity theory. The flux of virtual potential temperature can be computed following the formulas in section 2.2.3, as the data contains all the necessary quantities; u_* is given, as well as the specific humidity and the pressure. The next section describes how the gradients are computed starting from the raw measurements, in order to obtain the complete flux-profile relationships.

Then, we describe how to use the data to obtain new coefficients for the universal functions. Even though there is agreement on form of their form, there is still debate on the exact values of the coefficients, with different experiments resulting in different values [citation]. In order to ensure a fair comparison with the results of this work, we fit the universal functions to the data from the Cesar database.

3.2.1 Profile Computation

The wind and temperature profiles are computed at 40, 20 and 10 meters, because the Monin-Obukhov similarity theory is only valid in the surface layer. It is very hard to know the exact height of the surface layer, because it depends on the weather and no exact formulas are known [citation], but it is usually assumed to be 10% as high as the boundary layer [citation]. Based on its typical heights [citation], the surface layer is often higher than 40 meters and lower than 80.

The gradient is computed with finite differencing using measurements from the neighboring levels. The bottom-most measurements are at 1.5 meters for the temperature, and the wind speed is 0 at z_0 meters. The characteristic roughness z_0 depends on the properties of the surface, and from the Cabauw observatory it is around 0.1 meters [citation]. Let z_i be the height in meters of level i and a_i the corresponding measurement of quantity a, $\Delta^+ a_i = a_{i+1} - a_i$ and $\Delta^- a_i = a_i - a_{i-1}$ and similarly for z. Then the gradient of a at level i is computed as the average of the top and bottom gradients, weighted by the distance between the levels:

$$\begin{aligned} \left.\frac{\partial a}{\partial z}\right|_{z=z_i} &= \frac{\Delta^+ a_i}{\Delta^+ z_i} \cdot \frac{\Delta^+ z_i}{\Delta^+ z_i + \Delta^- z_i} + \frac{\Delta^- a_i}{\Delta^- z_i} \cdot \frac{\Delta^- z_i}{\Delta^+ z_i + \Delta^- z_i} \\ &= \frac{\Delta^+ a_i + \Delta^- a_i}{\Delta^+ z_i + \Delta^- z_i} \\ &= \frac{a_{i+1} - a_{i-1}}{z_{i+1} - z_{i-1}} \end{aligned} \tag{3.1}$$

3.2.2 Model Fitting

The general form of the universal functions is:

$$\phi(\xi) = \begin{cases} a + b\xi & \xi \geq 0 \\ a(1 - c^2\xi)^d & \xi < 0 \end{cases} \tag{3.2}$$

Where a is close to 1, b is positive, and d is negative. Since d is negative, the base of the power must be positive, hence the squared c. Following the approach outlined in section 2.3.2, the coefficients a, b, c and d can be found by minimizing the L2-regularized mean squared error using equations 2.13 and 2.14; in this case, the parameter vector $\theta = [a, b, c, d]^\intercal$. The gradient of 3.2 is:

$$\nabla_\theta \phi(\xi)|_{\xi \geq 0} = [1, \xi, 0, 0]^\intercal \tag{3.3}$$

when ξ is non-negative, and

$$\nabla_\theta \phi(\xi)|_{\xi < 0} = \begin{bmatrix} \tau^d & 0 & -2acd\xi\tau^{d-1} & a\tau^d \ln\tau \end{bmatrix}^\intercal \tag{3.4}$$

Where $\tau = 1 - c^2\xi$. The Hessian of 3.2 when $\xi \geq 0$ is simply 0, because it is a linear function in all parameters, while, in the negative case, we have:

$$\nabla^2_\theta \phi(\xi)|_{\xi<0} = \begin{bmatrix} 0 & 0 & -2cd\xi\tau^{d-1} & \tau^d \ln\tau \\ 0 & 0 & 0 & 0 \\ -2cd\xi\tau^{d-1} & 0 & \frac{2ad\xi\tau^d}{c^4\xi^2+\tau}\left(2c^2d\xi - c^2\xi - 1\right) & -2ac\xi\tau^{d-1}\left(d\ln\tau + 1\right) \\ \tau^d\ln\tau & 0 & -2ac\xi\tau^{d-1}\left(d\ln\tau+1\right) & \tau^d\ln^2\tau \end{bmatrix} \tag{3.5}$$

Analytical computation of the Hessian allows us to use the Newton conjugate gradient descent algorithm, which provides super-linear convergence rate, unlike other conjugate gradient methods whose rate of convergence is only linear [7].

Chapter 4

Results

todo

Chapter 5

Discussion

todo

5.1 Limitations and Future Work

todo

Bibliography

[1] R. S. Scorer. Atmospheric data analysis, roger daley, cambridge atmospheric and space science series, cambridge university press, cambridge, 1991. no. of pages: xiv + 457. isbn 0521 382157. *International Journal of Climatology*, 12(7):763–764, 1992.

[2] N. K. VINNICHENKO. The kinetic energy spectrum in the free atmosphere1 second to 5 years. *Tellus*, 22(2):158–166, 1970.

[3] Ulf Högström. Review of some basic characteristics of the atmospheric surface layer. *Boundary-Layer Meteorology*, 78(3):215–246, Mar 1996.

[4] W.A.A. Monna. Comparative investigation of dynamic properties of some propeller vanes. Technical report, Koninklijk Nederlands Meteorologisch Instituut, 1978.

[5] H.R.A. Wessels. Distortion of the wind field by the cabauw meteorological tower. Technical report, Koninklijk Nederlands Meteorologisch Instituut, 1983.

[6] *Handbook of Micrometeorology: A Guide for Surface Flux Measurement and Analysis*. Kluwer Academic Publishers, 2004.

[7] *Numerical Optimization*. 1999.