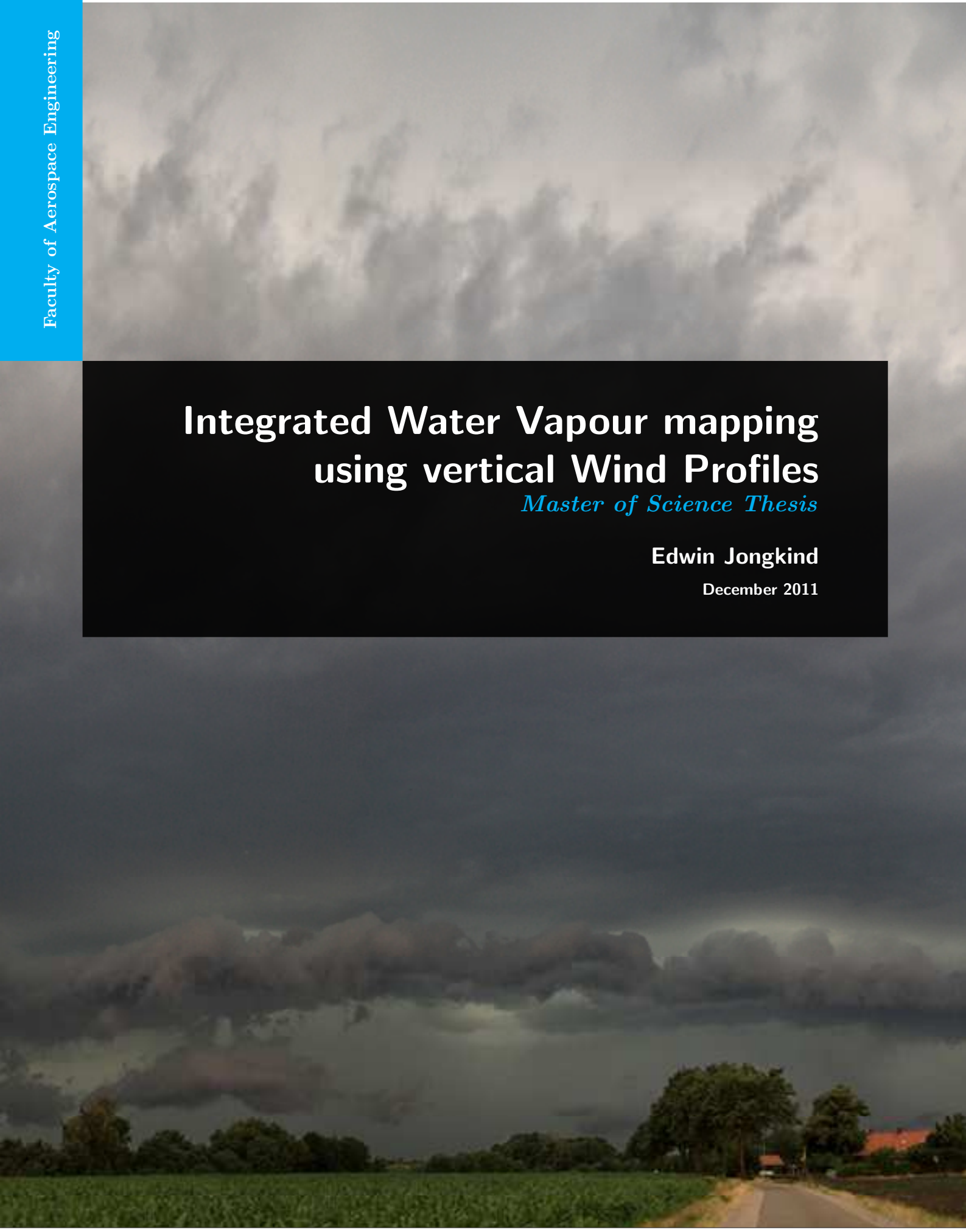
Faculty of Aerospace Engineering



Integrated Water Vapour mapping using vertical Wind Profiles

Master of Science Thesis

Edwin Jongkind

December 2011

TU Delft Delft University of Technology

Challenge the future

Integrated Water Vapour mapping using vertical Wind Profiles

Master of Science Thesis

For obtaining the degree of Master of Science in Aerospace Engineering
at Delft University of Technology

Edwin Jongkind

December 2011

Faculty of Aerospace Engineering · Delft University of Technology

The picture on the cover page shows a shelf cloud (a type of low horizontal cloud formation) in front of a Squall Line (line of severe thunderstorms). This cloud brought damaging winds and caused a lot of damage. (Photo by Jumbo0, nl.wikipedia.org/wiki/Shelf_cloud)

Delft University Of Technology
Department Of
Remote Sensing

The undersigned hereby certify that they have read and recommend to the Faculty of Aerospace Engineering for acceptance a thesis entitled **"Integrated Water Vapour mapping using vertical Wind Profiles"** by **Edwin Jongkind** in partial fulfillment of the requirements for the degree of **Master of Science**.

Dated: December 2011

Graduation professor:

prof.dr.ir. R.F. Hanssen (DUT)

Supervisors:

dr. R.C. Lindenbergh (DUT)

dr.ir. H. van der Marel (DUT)

dr. S. de Haan (KNMI)

Co-reader:

dr. S.R. de Roode (DUT)

TU Delft Delft University of Technology

Royal Netherlands Meteorological Institute
Ministry of Infrastructure and the Environment

Preface

During a project in my second year of the Bachelor of Aerospace Engineering I got interested in the subject of GPS and remote sensing, under the supervision of Hans van der Marel. When I finished most of my Master courses, I was searching for a graduation topic. As this graduation research would about a year, I was looking for a topic which would be interesting for everyday live. After reading the Master Thesis project presented by Roderik Lindenbergh, I immediately knew that I wanted this thesis to be my research topic for the coming period.

After one month of literature study, I was watching with some study mates '*Op volle kracht, vlieg 'm in de gracht!*', part of the sixty-fifth anniversary of the 'VSV Leonardo da Vinci' study society of Aerospace Engineering. Although it was rainy in the morning, in the afternoon we were enjoying the sun and the nice 'airplanes', made of wood and cardboard, held together with duct tape. Suddenly at around 17:00, the sky turned black and we had to hide from a thunderstorm, with hailstones as big as table tennis balls. Exactly this kind of short-term weather changes meteorologists want to predict better. Therefore, this date was chosen as a case study for this thesis.

During the thesis I was supervised by Roderik Lindenbergh and Hans van der Marel from the TU Delft, and Siebren de Haan from the Royal Netherlands Meteorological Institute (KNMI). Of course it was really nice to have such a number of visions on the subject. Sometimes however, it was hard to keep them all satisfied. Each one of them had their own view of the matter. Roderik as a mathematician an outsider for Aerospace Engineering, but therefore he had the best ideas about the algorithms. Hans, who knows everything about Global Navigation Satellite Systems, and although he started as a senior advisor, he was constantly very interested in the project. And finally of course, Siebren; Especially at the beginning he was very helpful with regard to obtaining data and to familiarize me with the weather and the KNMI itself. I want to thank all of them for their help during my graduation project.

Besides thanking my supervisors for their support, I want to thank my parents as well. Although my studies took me more than average time to finish, my parents supported me throughout the period. Fortunately my parents promoted me not only to study, but also to live! For that, I also want to thank all of my friends for making this possible!

Delft, The Netherlands Edwin Jongkind
December 2011

Abstract

Water vapour in the atmosphere is a key variable in the prediction of heavy rain and thunderstorms. Water vapour in the atmosphere is still badly modeled although more and more ground based, airborne, as well as spaceborne measurements are retrieved. The main input for current water vapour maps are Integrated Water Vapour (IWV) measurements obtained every 15 minutes from GPS ground stations. My graduation project focuses on the question whether the combination of water vapour measurements and wind measurements can increase the quality of these water vapour maps. The increase in quality is expected as wind is moving the water vapour, also called advecting. Therefore in addition, 3D wind data obtained from HIRLAM is incorporated in the map making. HIRLAM is the abbreviation for High Resolution Limited Area Model for Numerical Weather Prediction, and this model is created by a European consortium of meteorological institutes. A modified Kriging interpolation approach is used to combine both current measurements and measurements from the past advected by the wind. By using a case study it is investigated whether the IWV prediction in between the GPS ground stations could be improved. For this case study, and using this method, we can conclude that on average the IWV maps are not improved by incorporating advected measurements. However, there are also locations for which the IWV predictions are improved.

Samenvatting

Waterdamp in de atmosfeer is één van de belangrijkste variabelen voor het voorspellen van zwaar weer en onweersbuien. Waterdamp in de atmosfeer is nog altijd niet goed gemodelleerd, ondanks dat er steeds meer metingen beschikbaar zijn, zowel vanaf de grond, vanuit de lucht, als ook met behulp van satellieten. De belangrijkste input voor de huidige waterdampkaarten zijn geïntegreerde waterdamp (IWV) metingen die iedere 15 minuten worden verricht met behulp van GPS grondstations. Mijn onderzoek richt zich op de vraag of het combineren van waterdamp- en windmetingen de kwaliteit van deze waterdampkaarten kan verbeteren. Een verbetering zou verwacht worden, aangezien wind de waterdamp verplaatst. Daarom wordt bij het maken van de waterdampkaart, wind uit HIRLAM aan het model toegevoegd. HIRLAM is een model dat gebruikt wordt bij de weersverwachting voor de korte termijn en is ontwikkeld door een Europees consortium van meteorologische instituten. Een aangepast Kriging interpolatie wordt gebruik om de huidige metingen te combineren met metingen uit het verleden, die verplaatst worden door de wind. Door middel van een casestudy is het onderzocht of de voorspelde hoeveelheid waterdamp tussen de stations verbetert. Voor deze casestudy, gebruikmakend van deze methode, kunnen we concluderen dat de waterdamp kaarten niet verbeterd worden door het meenemen van metingen uit het verleden. Dit is echter gemiddeld gezien; er zijn ook gevallen aan te wijzen waar dit wel het geval is.

Contents

List of Figures

List of Tables

Nomenclature

Latin Symbols

AH	Absolute humidity	[g m^{-3}]
c_m	Variance of the measurement error	[kg^2 m^{-4}]
c_s	Spatial component of the nugget variance	[kg^2 m^{-4}]
C	Coriolis force acceleration	[m s^{-2}]
e	Vapour pressure	[hPa]
e_s	Saturation vapour pressure	[hPa]
f_m^t	Factor with which Kriging weight is multiplied	[.]
g_0	Standard acceleration due to free fall	[m s^{-2}]
G	Pressure-gradient acceleration	[m s^{-2}]
h	Distance between observations	[m]
h	Height	[m]
$k(t, \delta t)$	Relative weight change	[.]
L_0	Temperature lapse rate	[K m^{-1}]
L_i	Radiance measured in channel i	[W sr^{-1} m^{-1}]
m	Ratio of molar masses of water vapour and dry air	[.]
m_d	Dry air mass	[kg]
m_h	Niell hydrostatic mapping function	[.]
m_w	Niell wet mapping function	[.]
m_w	Water vapour mass	[kg]
M	Molar mass of the earth	[kg $mole^{-1}$]
n	Refractive index	[.]

N	Neutral atmospheric refraction	[.]
p_a	Total ground pressure	[hPa]
p_w	Partial pressure of water vapour	[hPa]
P	Pressure	[Pa]
P_0	Pressure at sea level	[Pa]
r	Correlation coefficient	[.]
R	Rainfall rate	[mm h^{-1}]
R_d	Specific gas constant for dry air	[J kg^{-1} K^{-1}]
R_{univ}	Universal constant	[J kg^{-1} K^{-1}]
R_v	Specific gas constant for water vapour	[J kg^{-1} K^{-1}]
RH	Relative humidity	[%]
s	Path from GPS satellite to receiver	[m]
t	Temperature	[°C]
t_d	Dew point temperature	[°C]
T	Temperature	[K]
T_0	Standard temperature at sea level	[K]
T_m	Mean temperature	[K]
T_s	Temperature at surface height	[K]
u	Wind speed in east-west direction	[m s^{-1}]
U	Wind	[m s^{-1}]
U_g	Geostrophic wind	[m s^{-1}]
v	Wind speed in north-south direction	[m s^{-1}]
w	Water vapour mixing ratio	[.]
W	Friction force acceleration	[m s^{-2}]
W	Water vapour column amount	[g m^{-2}]
x	Horizontal distance in x-direction	[m]
z	Value of the observation	[.]
z_{GPS}	Height of the GPS receiver	[m]
Z	Radar reflectivity	[mm^6 m^{-3}]
Z_w	Compressibility factor for water vapour	[.]

Greek Symbols

β	Elevation angle	[°]
γ_{ij}	Dissimilarity between x_i and x_j	[.]
ρ	Total density of dry air and water vapour	[kg m^{-3}]
ρ_w	Water vapour density	[kg m^{-3}]

$(\sigma_m^t)^2$	Variance of the prediction at location m with t minutes time difference	[kg^2 m^{-4}]
ϕ_{GPS}	Latitude of GPS receiver	[°]

Abbreviations

CAPPI Constant Altitude Plan Position Indicator
CDMA Code Division Multiple Access
DMSP Defense Meteorological Satellite Program
DWD Deutscher Wetterdienst
E-AMDAR EUMETNET Aircraft Meteorological Data Relay
E-ASAP EUMETNET Automated Shipboard Aerological Programme
E-GVAP EUMETNET GNSS Water Vapour Programme
E-SURFMAR EUMETNET Surface Marine Programme
E-WINPROF EUMETNET Wind Profiler
Envisat Environmental Satellite
EPS EUMETSAT Polar System
ESA European Space Agency
EUCOS EUMETNET Composite Observing System
EUMETNET European Meteorological Network
EUMETSAT European Organisation for the Exploitation of Meteorological Satellites
FOV Field Of View
GNSS Global Navigation Satellite System
GPS Global Positioning System
GRAS GNSS Receiver for Atmospheric Sounding
HIRLAM High Resolution Limited Area Model
HIRS High-resolution Infrared Radiation Sounder
HOV Hold Out Validation
HPBW Half-Power Beam Width
IASI Infrared Atmospheric Sounding Interferometer
InSAR Interferometric Synthetic Aperture Radar
IWV Integrated Water Vapour
KNMI Royal Netherlands Meteorological Institute
LIDAR Light Detection and Ranging
LOOCV Leave-One-Out Cross Validation
LWP Liquid Water Path
MATLAB MATrix LABoratory
MERIS Medium Resolution Imaging Spectrometer

ME Mean Error
MGRS Military Grid Reference System
MHS Microwave Humidity Sounder
MODIS Moderate Resolution Imaging Spectroradiometer
MSG Meteosat Second Generation
NAVSTAR Navigation Signal Timing And Ranging
OPERA Operational Programme for the Exchange of weather Radar information
PRN Pseudo Random Noise
PTU Pressure, Temperature, and Relative Humidity
Radar Radio detection and ranging
RMSE Root Mean Squared Error
RPG-HATPRO Radiometer Physics GmbH - Humidity And Temperature Profiler
SHD Slant Hydrostatic Delay
SODAR Sound Detection And Ranging
SSM/I Special Sensor Microwave Imager
STD Slant Total Delay
SWD Slant Wet Delay
TPW Total Precipitable Water
UPS Universal Polar Stereographic
UTM Universal Transverse Mercator
WMOID World Meteorologic Organization ID
WV Water Vapour
ZHD Zenith Hydrostatic Delay
ZTD Zenith Total Delay
ZWD Zenith Wet Delay

Chapter 1

Introduction

1-1 Background

Water vapour in the atmosphere is a key variable in the prediction of heavy rain and thunderstorms. Especially fast changing amounts can be a forerunner of heavy weather. For this reason, meteorological institutes are interested in the best possible models for water vapour in the atmosphere. However, water vapour in the atmosphere is still badly modeled although more and more ground based, airborne, as well as spaceborne measurements are retrieved. There is no single measurement technique which is able to measure an entire area with a high spatial and temporal resolution. Techniques with a high spatial resolution often have a bad temporal resolution, and vice versa. Also, they are often not able to give a vertical distribution of the water vapour.

Main input for current water vapour maps are Integrated Water Vapour (IWV) measurements obtained every 15 minutes from GPS ground stations. My graduation focuses on the question whether the combination of water vapour measurements and wind measurements can increase the quality of these water vapour maps. Such increase in quality is expected as wind is advecting the water vapour.

Therefore in addition, 3D wind data obtained from HIRLAM is incorporated in the map making. HIRLAM is the so-called High Resolution Limited Area Model for Numerical Weather Prediction, and is produced by a European consortium of meteorological institutes. A modified Kriging approach is used to combine both current measurements and measurements from the past moved by the wind.

The concept of using old data for current water vapour maps is not a complete new idea. Instead, the research carried out during this graduation is a follow-up of research described in among others (Elgered et al., 2005). The Dutch Royal Netherlands Meteorological Institute (KNMI) in De Bilt is one of the institutes involved with the study of water vapour in the atmosphere. This graduation research is therefore performed in cooperation with the KNMI. In the beginning of the graduation, De Bilt was the location were most of the time was spent. In this way data and knowledge was easy to obtain.

1-2 Research Objectives

The main topic for this thesis is 'Integrated Water Vapour (IWV) mapping using vertical Wind Profiles'. Water vapour is the gas phase of water. In the atmosphere there is advection of water vapour, movement of water vapour caused by wind. Vertically integrated atmospheric water vapour is measured in the Netherlands using ground based GPS measurements at about 40 stations. Vertically integrated means that only the total amount of water vapour at a location is known. So, there is no information about the distribution of the water vapour along the path. From these IWV measurements a 2-dimensional water vapour map is constructed by e.g. the KNMI, using various spatial interpolation techniques. In these computations, so far, no information is used about the transport, or advection, of water vapour by the wind.

The main research question to be answered in this graduation thesis is:

Will the 2D mapping of water vapour in the atmosphere improve if we include advection of water vapour into the model?

This question is subdivided into three subquestions:

- Divide the 2D water vapour mapping into several vertical layers, each layer containing part of the water vapour obtained from GPS measurements. The question here will be how to split up the atmosphere in layers and how to divide the water vapour among these layers.
- Use wind data to advect water vapour from previous measurements to obtain extra data points. Question here will be how long data points of the past should be used and how to weight them, which will result in the best IWV map.
- Use and compare different interpolation techniques to see which technique yields the best result.

These three subquestions are all part of a model which is used to generate an IWV map. The first subquestion is primarily depending on the source of the input data which is used to distribute the GPS IWV in the vertical direction. The second subquestion is the main focus of the thesis. Several options will be discussed and compared. At the beginning of the graduation the idea of the third subquestion was to compare different interpolation techniques. But due to time constraints, instead it was decided to compare different settings of the algorithm which is investigated during the thesis.

1-3 Methodology

The following data inputs are used for the Integrated Water Vapour mapping procedure that is developed in this thesis. The first one are the IWV values as measured by GPS ground stations. These IWV values are measured every 15 minutes. Second data input are the measurements from radiosondes. A radiosonde is launched only twice a day. The data from the radiosonde is used to divide the IWV values at the GPS ground stations among a number of layers. The last data input is wind from HIRLAM. HIRLAM is updated every 3 hours. The

wind from HIRLAM is used to advect the water vapour in each layer above each station to a next measurement epoch. Water vapour measurements can also be advected multiple epochs. The advected water vapour measurements provide extra data points. By interpolating both the current plus the advected measurements in each layer a water vapour map for each layer is obtained. By summing all water vapour maps, again a IWV map is obtained.

The read in of data files, the processing of the data in the model, and the visualization of the output have all been programmed with the use of MATLAB. MATLAB is a widely used mathematical software environment, which can be used for the programming of the algorithm as well providing functions for visualization. Other languages and software programs are available as well, but during the study of Aerospace Engineering most experience has been acquired with MATLAB.

1-4 Structure of the Report

Chapter 2 starts the thesis with a general background of water vapour and wind. It is described how these two arise in the atmosphere. Also an extensive overview is given of observing techniques which are available nowadays. This chapter is mainly based on the literature study which was carried out as part of the graduation.

Chapter 3 kicks off with a description of the main input data of GPS IWV measurements. It is followed by an introduction to the process of interpolation of the IWV measurements to obtain an IWV map. This is also an example of how water vapour maps are produced today.

In Chapter 4 the method is proposed to extend the interpolation described in Chapter 3, by including the advection of water vapour measurements.

Chapter 5 continues with a refinement of the method described in Chapter 4. The chapter introduces a method which makes a distinction between current and advected measurements in the IWV map computation.

In Chapter 6 the results of the proposed method are described. Using a case study it is validated if the new method gives an improvement to the IWV maps.

Chapter 7 finalizes the thesis by presenting the conclusions of the research. The chapter ends with the recommendations for further research.

Chapter 2

Measuring Water Vapour and Wind

This chapter gives a general background of water vapour and wind, how they behave in the atmosphere, and how they can be measured. Section 2-1 starts with a general background on water vapour. Section 2-2 continues with a background on wind. Section 2-3 concludes with observing techniques for water vapour and wind in the atmosphere.

2-1 Water Vapour

This chapter starts with a section about water vapour. In Subsection 2-1-1 some general properties of water and water vapour are given. Subsection 2-1-2 deals with the most common definitions and units used for expressing the amount of water vapour in air. Subsection 2-1-3 gives a description of the role and distribution of water vapour in the atmosphere.

2-1-1 General Properties

Like most materials, water exists in solid, liquid, and gas phase. Water on earth is mostly in its liquid phase, simply called *water*. Its solid state is called *ice* and its gas state is called *water vapour*. Water vapour can be generated by evaporation or boiling of water. Evaporation occurs at the surface of the liquid, whereas in boiling, which occurs at a temperature above the boiling point, bubbles are formed inside the liquid. When water evaporates it is continuously leaving a surface which is usually at normal earth atmospheric conditions. When water molecules evaporate it takes part of the heat from the surface, this is called evaporative cooling (Schroeder & Pribram, 1999). This is used in some warm countries to quickly cool streets, by just spraying water on them. Transpiration is also an example of evaporative cooling. Water molecules can also evaporate from an ice surface without going through the liquid phase. In this case the evaporation is called sublimation. When ice is directly produced from water vapour, frost is formed. When water vapour is transformed back into liquid water it is called condensation.

2-1-2 Definitions and Units

Several definitions are in use to describe the amount of water vapour in the air: water vapour density, water vapour mixing ratio, specific humidity, relative humidity, saturation water vapour pressure, Integrated Water Vapour (IWV), Total Precipitable Water (TPW), and even more. Some of the definitions describe the *local* water vapour amount, while others describe the *total vertically integrated* amount of water vapour.

For the local water vapour amount the following definitions are in use. Water vapour density is the amount of water in gas state in [g m^{-3}]. The water vapour mixing ratio w which is dimensionless, is the ratio between the mass of water vapour, m_w, and the mass of dry air, m_d, in a volume of air. Specific humidity is the amount of water vapour in grams of a total air volume with a mass of one kilogram [kg]. In a mixture of ideal gases, each gas has a partial pressure, which is the pressure the gas has if it would be alone in the volume. Water vapour pressure is the partial pressure of water vapour in the atmosphere [Pa]. The saturation water vapour pressure is the pressure of the air, under the condition that the water vapour evaporation in the air is in equilibrium with the condensation [Pa]. Then the air is saturated with water vapour. Relative humidity is the ratio between water vapour pressure and the saturation water vapour pressure expressed as a percentage [%]. At 100 % relative humidity the air is saturated with water vapour.

Parameters which describe the total vertically integrated amount of water vapour are Total Precipitable Water Vapour (TPW), (Total) Integrated Precipitable Water Vapour (IPW), and (Total) Integrated Water Vapour (IWV). These parameters are all similar and express the amount of water vapour in the vertical column from the measurement point to the top of the atmosphere. They are expressed in kilograms water vapour per square meter [kg m^{-2}]. An alternative is the total amount of water in the air condensed into a single layer, expressed in [mm]. Both can be used and indicate the same amount of water vapour (1 mm = 1 kg m^{-2}). IWV is mostly expressed in [kg m^{-2}], whereas definitions containing 'precipitable' are expressed in [mm].

2-1-3 Water Vapour in the Atmosphere

Nitrogen accounts for 78.08 volume percent, and Oxygen for 20.95 volume percent of the atmosphere at sea level. In this perspective, water vapour only take between 0.1–2.8 volume percent at sea level. However, vertically integrated over the entire atmosphere it varies between 6.5 and 180 kg m^{-2}, which is a significant amount. Through the absorption of infrared radiation, water vapour accounts for 60% of the greenhouse effect (Kiehl & Trenberth, 1997).

Because water has a large heat capacity, water in the oceans and water vapour in the air contain a lot of energy. Daily heating of the atmosphere is influenced by movement of water vapour in the air. This movement of water vapour can cause severe weather such as rain, hail, and thunderstorms.

Besides instruments measuring water vapour to predict weather, water vapour in the atmosphere disturbs measurements on other instruments. Especially unknown variations cause measurement inaccuracy and lower precision with respect to vacuum. GPS, Radar, radio telecommunication, and laser profiling are examples of techniques suffering from water vapour.

Water vapour is one of the driving forces of heat transfer in the atmosphere and plays a significant role in the development of clouds and precipitation. When water vapour condensates, water vapour is removed from the atmosphere and converted into liquid water (rain), producing heat. When water vapour is produced from evaporation of water or sublimating from ice, cooling occurs.

The density of air in the atmosphere decreases with height. This is due to decreasing hydrostatic pressure (mass of air above the measurement point) and decreasing gravitational force with increasing height. Humidity and temperature differences also cause density variations. The atmosphere is often divided in two ways. When the mixing of gases is considered, two layers are present. The lower part is the homosphere up to about 100 km where primarily turbulent mixing occurs. Here the gas composition is homogeneous, except for water vapour. At higher altitudes, in the heterosphere, gases are divided by molecular mass. Above 1000 km only Helium and Hydrogen are left. The transition zone between the homosphere and heterosphere is called the turbopause.

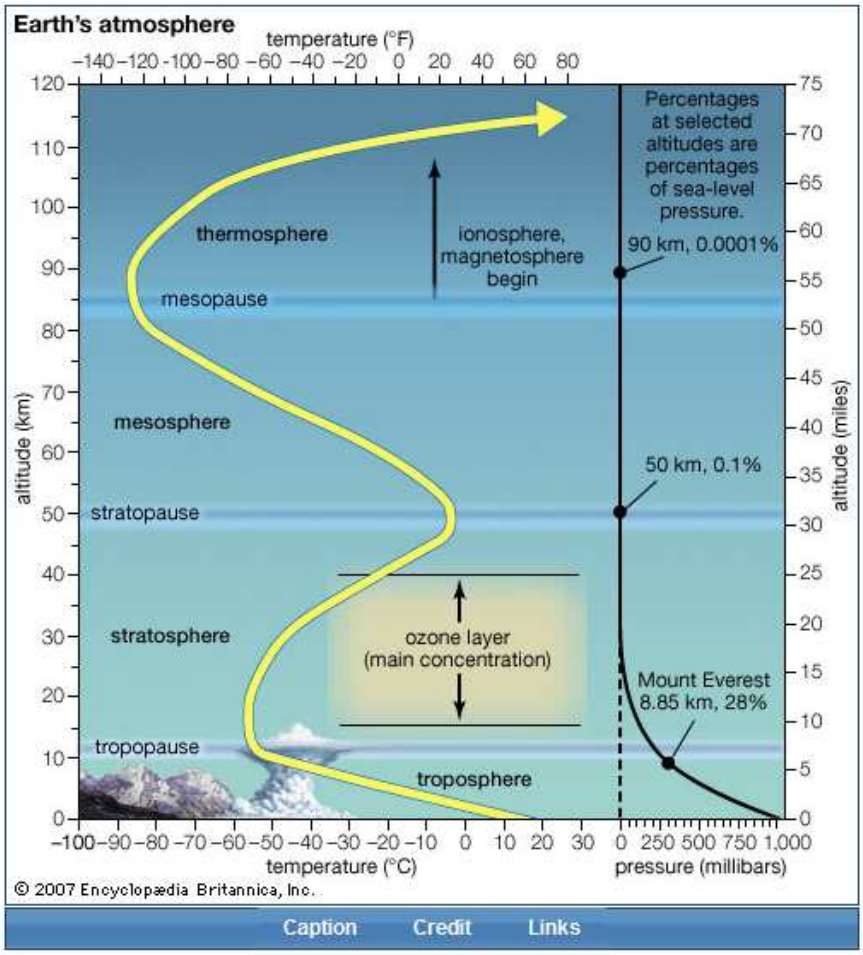


Figure 2-1: The layers of Earth's atmosphere. The yellow line shows the response of air temperature to increasing height. (Encyclopædia Britannica, 2010a)

Another division can be made by air temperature, which is visualized in Figure 2-1. The first layer, closest to the Earth's surface, is the troposphere. In this layer the air temperature decreases with height with about 6.5 °C per kilometer until it reaches the tropopause. The height of the tropopause depends on the location on earth and atmospheric turbulence. At the poles, where the vertical mixing is low, the tropopause reaches 8 km. When vertical mixing is high, for example at the equator, the tropopause can be as high as 18 km. In the lower part of the troposphere, the weather is highly influenced by local surface topography. This lower part is called the planetary boundary layer.
At the tropopause the temperature is almost constant with height. In the layer above the troposphere the temperature increases with increasing height. This layer is called the stratosphere and is located from the tropopause up to 50 km. On top of the stratosphere is the stratopause, located at about 45–50 km, where again the temperature is relatively constant with increasing height. Starting from the stratopause to about 85–90 km the temperature decreases with height, this layer is called the mesosphere. The mesopause separates the mesosphere from the next layer, the thermosphere. The thermosphere is the highest layer at which temperature is meaningful; above this layer, there are almost no collisions between molecules. The height of the top of the thermosphere depends on the solar activity. At low solar activity the top is at 250 km, at high solar activity the height can be up to 500 km. Temperatures in this layer vary between 500 and 2000 K. The lower part of the thermosphere is called the ionosphere. It extends from about 80 km and up to 600 km and contains a lot of charged particles, with a maximum at about 400 km. (Encyclopædia Britannica, 2010a)

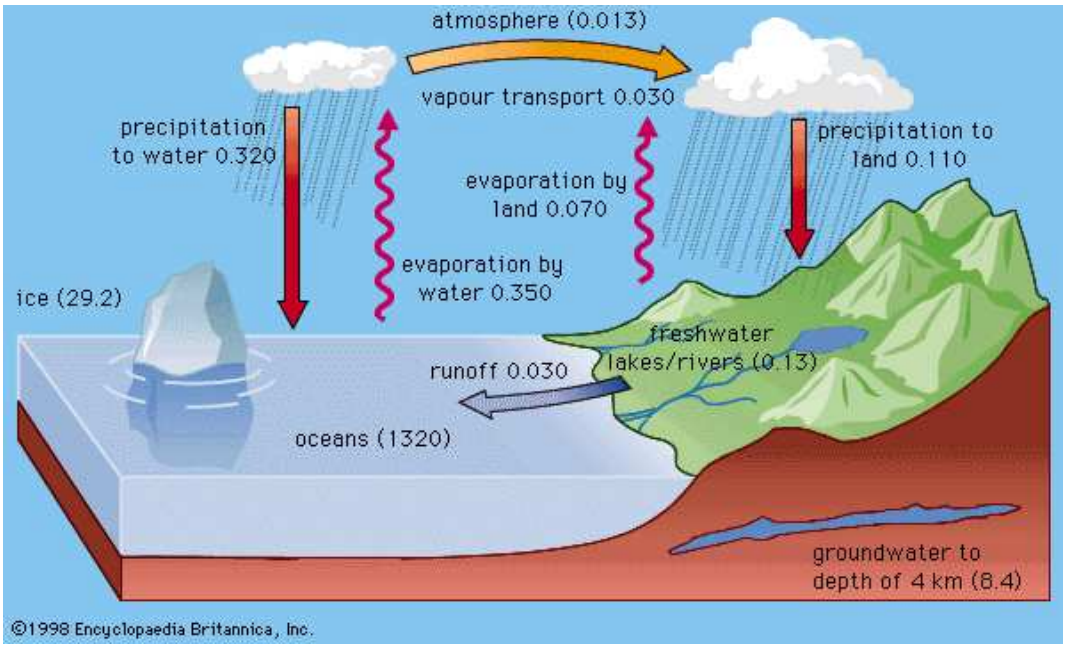


Figure 2-2: Hydrologic cycle: present-day surface. The numbers in parentheses refer to volumes of water in millions of cubic kilometres, and the fluxes adjacent to the arrows are in millions of cubic kilometres of water per year. (Encyclopædia Britannica, 2010b)

Figure 2-2 shows the general hydrological cycle. As can be seen in Figure 2-2 water vapour in the atmosphere is only one part of the total water cycle. About $0.420 \cdot 10^6$ km^3/year of water vapour evaporates from oceans and land. The mean amount of water vapour in

the atmosphere is about $0.013 \cdot 10^6$ km^3, and $0.030 \cdot 10^6$ km^3 is transported yearly from oceans to land trough the atmosphere. Yearly precipitation is about $0.430 \cdot 10^6$ km^3 of which $0.110 \cdot 10^6$ km^3 falls on land. From the precipitation on land $0.07 \cdot 10^6$ km^3 is evaporated back to the atmosphere. The average residence time of water vapour in the atmosphere is only 10 days. It can also be in its liquid state (rain and droplets in clouds) or solid state (snow, hail, or ice crystals in clouds). In these last two states water exists only for a relative short time. Water vapour is usually invisible, only when the amount of water vapour is very large it can be seen as a haze around the sun. When water droplets or ice crystals are just formed they are very small and can float freely. When concentration is increasing, clouds are formed, which are visible. When the droplets or crystals grow further to about 0.01 cm, rain or snow is formed. Generally, at high latitudes 1% of the water in the air is in liquid or solid state. At tropical regions this can be up to 5% or more. (Encyclopædia Britannica, 2010b; Haan, 2008)

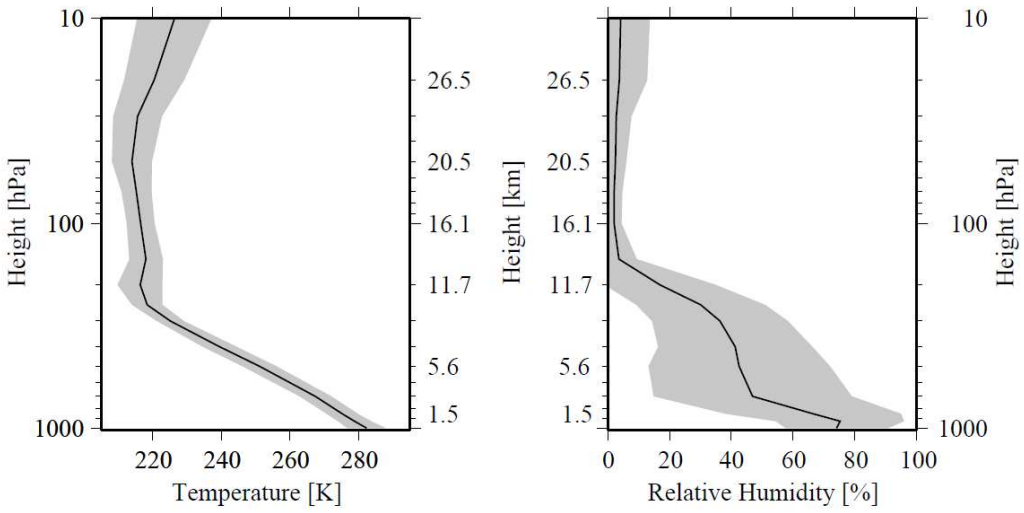


Figure 2-3: Annual mean temperature (left panel) and relative humidity (right panel) at significant levels as observed by radiosondes launched from De Bilt in 2002. The shaded area at each side of the mean indicates the standard deviation of temperature and relative humidity. (Haan, 2008)

In Figure 2-3, the annual mean temperature and relative humidity at De Bilt in 2002 are shown. The solid black line indicates the mean and the shaded area the variation. Going up in the atmosphere, the temperature decreases till about 11 km, till this height the troposphere is situated. In the troposphere the temperature decrease is constant with increasing height (temperature lapse rate is negative) and most of the water vapour is in this layer. After that the temperature is more or less constant to about 20 km, this part is called the tropopause. Higher, at the stratosphere, the temperature lapse rate is inverted. When air is getting colder it is able to contain less water vapour. Because of this above 15 km there is less than 0.1% water vapour of the concentration at the Earth's surface. This can be seen in Figure 2-4, where the lower 10 km of the atmosphere is visualized.

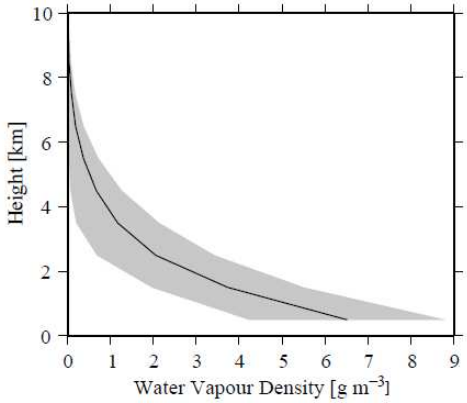


Figure 2-4: Annual mean and standard deviation of water vapour density as observed by radiosondes launched from De Bilt in 2002. The shaded area at each side indicates the variation of the water vapour density. (Haan, 2008)

2-2 Wind

In this section an introduction to wind is given. In Subsection 2-2-1 the main units are introduced. Subsection 2-2-2 continues with the main causes of wind. Subsection 2-2-3 closes this chapter with a description of global and macro scale patterns of wind.

2-2-1 Units

The International System of Units (SI) prescribes as unit for speed [m/s]. Wind speed however uses other units. In shipping and aviation knots is widely used, reason for this is that the speed of ships and aircraft themselves is also expressed primarily in knots. For the same reason [km/h] is used for describing wind speed for people driving cars. 1 knot equals 1 nautical mile (1852 m) an hour, which is approximately equal to 0.515 m/s. There are historical reasons for using this unit. Maximum wind speeds measured at the early days was about 50 m/s at the Earth's surface (lower part of boundary layer), so only two digits were necessary for daily use. Moreover, precision of most weather stations is about 0.5 m/s.

A widely used unit for wind speed is the Beaufort scale. This is a scale ranging from 0 to 12. The higher the number, the higher the wind speed. 0 indicates (almost) no wind (≤ 0.2 m/s), 6 is a strong breeze (10.8–13.8 m/s), 9 is a storm (20.8–24.4 m/s), and 12 is a hurricane (≥ 32.7 m/s). These scale units are averages over 10 minutes, so wind gusts are averaged out.

Wind direction is defined by the direction it is coming from, so *northerly* wind flows from north to south. The angle of the wind is defined positive clockwise with respect to North. This definition is opposite in comparison with water currents in the ocean. Several divisions exist for the wind into directions. Sometimes the course 16 divisions are used (N, NNE, NE, ENE, etc.). Also 30 degrees or even 10 degrees divisions are used. (Wieringa & Rijkoort, 1983)

2-2-2 Causes of Wind

Wind is air in movement caused by several forces acting on the air particles. These forces are caused by pressure differences, the rotation of earth (the wind direction is deflected except for easterly wind exactly at the equator), friction by the earth surface, and the centrifugal force.

The first cause of a force is a pressure difference. Horizontal temperature differences are the main cause of pressure differences. Pressure is visualized on weather maps by isobars, which are lines of equal pressure. The amplitude of the pressure is usually denoted in mbar (10^{-3} bar), which is 100 times the SI unit Pascal (Pa). Pressure is usually between 970 and 1030 mbar at sea level. The pressure change perpendicular to isobars in mbar per km is called the *pressure gradient*. This pressure gradient causes a force which is called the pressure-gradient force. The size of acceleration G due to this pressure-gradient force is given by

$$G = -\frac{1}{\rho}\frac{\mathrm{d}P}{\mathrm{d}x}, \tag{2-1}$$

in which P is the pressure in [Pa], ρ is the density of air in [kg m^{-3}], and x the horizontal distance along the x-axis in the direction of the pressure gradient.
The second cause of force differences is due to the rotation of the Earth. When wind is flowing on the northern hemisphere from south to north, the Earth rotates beneath it to the left. Seen from the earth, wind seems to be deflected to the right. On the southern hemisphere, seen from the earth, wind is deflected counterclockwise. This apparent deflection is caused by an apparent force called *Coriolis force*.
The third cause of force differences is friction. This force is approximately proportional to wind speed and is in opposite direction and zero when there is no wind. The size of the friction is depending on the surface, the height above the surface, and the wind speed itself. At higher altitudes there is less friction.

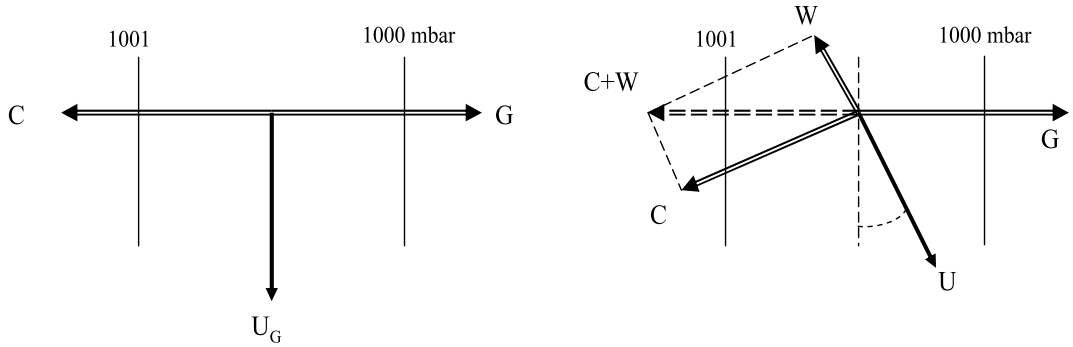


Figure 2-5: Formation of (geostrophic) wind U (U_g) from the pressure-gradient force acceleration G, Coriolis force acceleration C, and friction force acceleration W on the northern hemisphere. (Wieringa & Rijkoort, 1983)

As a result of the three forces mentioned before, the wind gets a speed and direction. The friction force is opposite with respect to the wind speed. The Coriolis force is perpendicular to the wind speed. At the northern hemisphere the Coriolis force is to the right of the wind speed and at the southern hemisphere this is to the left. When friction is assumed to be zero and there is an equilibrium condition, the Coriolis force is opposite to and of equal magnitude as the gradient force, and both are in a balance. This assumption is approximately correct

high in the atmosphere, away from the Earth's surface. In this case the wind direction is parallel to the isobars, so the pressure differences would stay the same. When there is no friction, the wind speed and direction are depending solely on the pressure gradient. This wind is called the *geostrophic wind*.
Of course some friction exists. Now G is counteracted by the resultant force of the Coriolis force plus the friction force and the Coriolis force is still perpendicular to the wind speed. This implies that the wind speed is not any longer parallel to the isobars, but they make a (small) angle with each other. In this case where friction is present, some wind is flowing from high pressure to low pressure. Both cases are visualized in Figure 2-5. Due to this friction force, the wind is rotating counterclockwise on the northern hemisphere with respect to the geostrophic wind. This is called a backing wind. If the wind is rotating clockwise with respect to the geostrophic wind, it is called a freeing wind. The last force is a result of the circular motion of the wind around high and low pressure areas. This force is called centrifugal force, which is a force away from the centre of rotation.

The lower part of the atmosphere in which the friction force plays a role is called the *planetary boundary layer* and reaches from the earth surface to about 1–1.5 km in the Netherlands. Above the planetary boundary layer, the friction has little impact and the wind is nearly geostrophic. In the upper part of the planetary boundary layers, from 1000 m to about 60 m, there is some weakening of the wind. In this layer the primary backing of the wind occurs. This layer is called the Ekman layer. The layer between the Earth's surface and this Ekman layer is called the surface layer. In this layer there is almost no backing of the wind, however the wind speed drops to zero at the earth surface. The height of the surface layer depends on the terrain (large city, suburbs, or open terrain). (Wieringa & Rijkoort, 1983)

2-2-3 Wind and Atmosphere

To a large extend temperature, pressure, and wind are influenced by the Sun. Because the Earth orbit around the Sun has a small ellipticity, the incoming radiation varies throughout the year. A more important cause of variation of the incoming radiation at the Earth's surface, is the local elevation angle of the Sun. This local elevation angle depends on the latitude, time of year, and time of day. More incoming radiation, means higher surface temperature, producing lower pressure. Because the incoming radiation is highest around the equator, a low pressure belt along the equator exists. At around 30 ° this low pressure belt causes a high pressure belt, which again causes a low pressure belt at around 45–60 °. The latitude where these belts are located, depend on the current axial tilt of the Earth's rotation axis with respect to the orbital plane, which is the main cause of the seasons. Wind in the atmosphere and currents in the ocean provide the heat transport between the different high and low pressure belts. In Figure 2-6 the global winds of the Earth can be seen.

Warm air from the equator and cold air from the poles meet each other at fronts. A cold front arises when cold air intrudes into warmer air. For a warm front this is the other way around. When warmer and lighter air intrudes into colder and heavier air it is forced up at both sides by the colder heavier air. This rising air is cooled down because of expansion and is able to contain less water. This creates clouds and the air is rising higher because of the condensation heat created by the condensation of the water vapour. This area is called a depression. In Figure 2-7 the arising of an occluded cyclone is visualized.

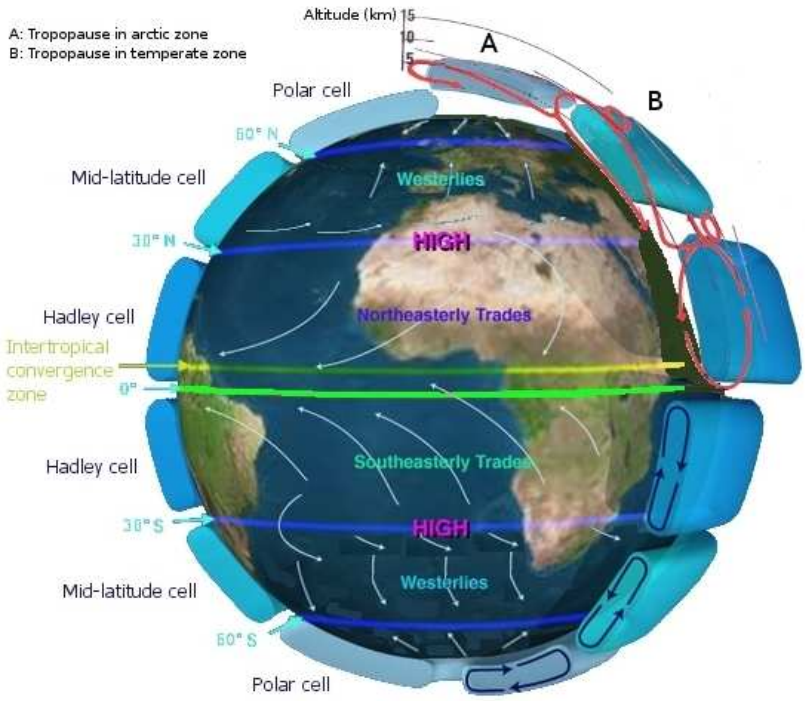


Figure 2-6: Global circulation of the Earth's atmosphere displaying Hadley cell, Ferrell cell and polar cell. (sealevel.jpl.nasa.gov)

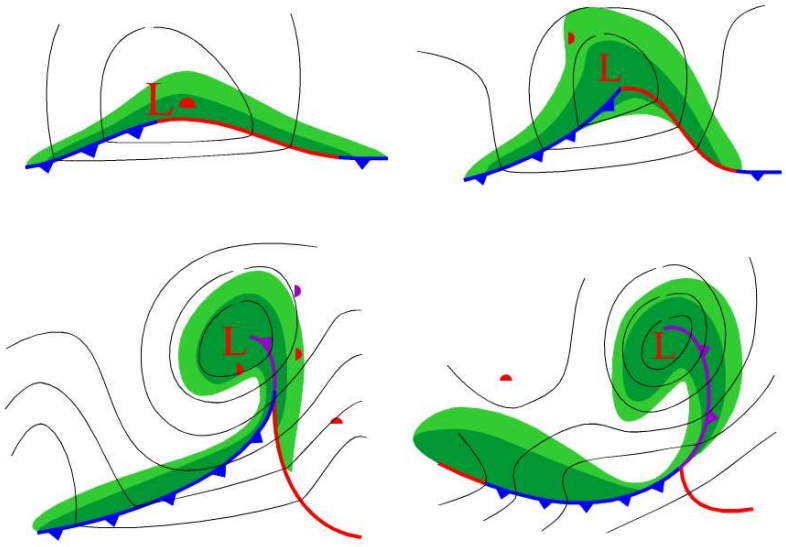


Figure 2-7: Arising of an occluded cyclone. In a stationary front a wave arises (upper left), the wave intensifies (upper right), an occluded front is formed because the cold front moves faster than the warm front (lower left), and finally no warm moist air moves into the occlusion an the low pressure system dissipates (lower right). (www.srh.weather.gov)

Table 2-1: Overview of different observation techniques and their properties. '−−' is bad, '++' is good, '0' is neutral.

Technique	Detects or calculates	Temporal resolution	Footprint	Coverage
Surface Measurements	Humidity, wind, temperature	++	−	−
Radiosonde	Humidity and wind profile	−−	++	−−
Ground Weather Radar	Cloud, droplets, precipitation, insects, velocity w.r.t. ground station	++	+	+
Wind Profilers	Vertical wind profile	++	++	−
Water Vapour Radiometer	Vertical Humidity profile	++	++	−
LIDAR	Clouds, aerosols, WV mixing ratio	++	++	−
AMDAR	Wind	+	+	0
GPS	IWV	+	+	−
Satellite	Humidity	−−	−	0

2-3 Observing Techniques

In this section several observation techniques for measuring water vapour and wind will be described. In Table 2-1 an overview is given of the observation techniques, including what they detect and an indication of temporal resolution, footprint, and coverage. A good temporal resolution means a frequent update. A good footprint indicates the measurement represents a small area. A bad coverage finally, means a lot of instruments are necessary to cover a large area. In Europe all observations are brought together by the European Network of Meteorological Services (EUMETNET) Composite Observing System (EUCOS). This program bundles all observation to create one big European observation network.

Measurements can be divided in three categories: in situ measurements, remote sensing from a satellite, and remote sensing from ground surface. Most important in situ observations of wind and water vapour are retrieved from the KNMI surface network, civil and military airports, radiosondes, AMDAR, ships, voluntary rain gauge network, several measurements at the measurement tower at Cabauw. Most important ground based remote sensing observations are from weather radar, wind profiler, LIDAR, microwave radiometer, and GPS. Most important satellite based remote sensing observations are part of EUMETSAT and NOAA.

In Subsection 2-3-1 the system of surface measurements is described. In Subsection 2-3-2 the system of radiosondes is presented. Subsection 2-3-3 explains the principles of a ground weather radar. Subsection 2-3-4 continues with the wind profiler technique. Subsection 2-3-5 continues with the water vapour radiometer. Subsection 2-3-6 describes the LIDAR measurement technique. Subsection 2-3-7 describes the system of measurements taken by instruments aboard aircraft. Subsection 2-3-8 explains how water vapour is measured by means of GPS receivers. Finally Subsection 2-3-9 describes some instruments aboard satellites.

2-3-1 Surface Measurements

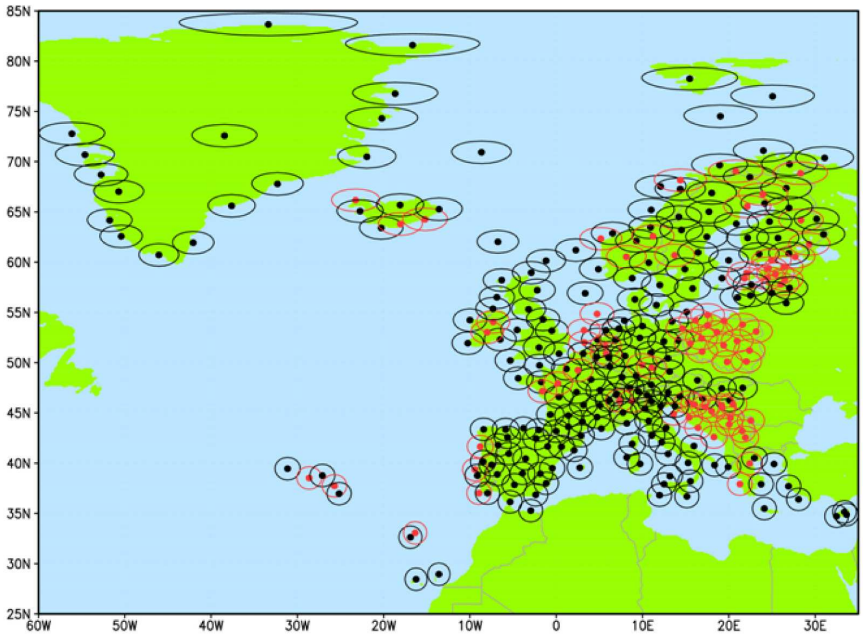


Figure 2-8: The updated EUCOS surface land station network 2010 consists of 268 stations. Black dots indicate EUCOS ground stations that were available until October 2009, red dots became available thereafter, rings around the stations indicate a 125 km range. (Kleinert, 2010)

As part of the EUCOS Operational Programme, a surface land station network is operating. Aim is that at each 250 km a surface station is situated that does all kind of observations. In this way macro scale weather patterns can be observed, also called synoptic scale weather. In 2010 the surface land network consists of 268 weather stations, which take each hour or each 3 hours measurements, depending if the station is human or automatically operated. Besides continental Europe, also some stations are located in sea, ocean, and Canada. In Figure 2-8 all surface stations in 2010 are presented.

On The Netherlands mainland stations are situated near Leeuwarden, Twente, Amsterdam Schiphol Airport, and Volkel. Also three platforms in the North Sea are maintained. Due to dense horizontal spacing of stations in the Southern of the Netherlands and Belgium, the station in Beek is declared as an additional station, and now act as a quality monitoring station (Kleinert, 2010). At the surface stations pressure, temperature, wind speed and direction, and specific humidity are measured. Also it is determined if it is raining. The amount of rain is not determined as this is obtained from the Ground Weather Radar.

Next to this fixed network of ground stations, EUMETNET is also working on a Surface Marine observation programme (E-SURFMAR). This is a collection of observation sites aboard Voluntary Observing Ships (both manned and automated observatories), drifting buoys, and moored buoys. Together they provide additional observations mainly of the local air pressure.

Wind speed on land is mainly measured by means of a cup anemometer (also called Robinson anemometer). This an instrument that consists of three or four half hollow spheres (cups) attached to a vertical axis, which rotates when wind is blowing against the half hollow spheres.

By measuring the rotation speed of the axis, the wind speed is determined. Using a weather vane, the wind direction is determined. The cup anemometer used by the KNMI has a measurement range of 0.5–50 m/s, a resolution of 0.1 m/s, an accuracy of 0.5 m/s, and measures the wind each second. The weather vane has a measurement range of 360°, a resolution of 1°, an accuracy of 3° and also measures each second. The measurements are average over 10 minutes to average out wind gusts. (KNMI, 2010b, Handboek waarnemingen)

Next to these automatic weather stations, a denser network in The Netherlands is available. This network consists of about 325 volunteers who make each morning measurements of precipitation of the last day. These measurements of the voluntary network is a good complement to the radar measurements.

2-3-2 Radiosonde

Figure 2-9: Balloon with an attached radiosonde together with a close-up of the Vaisala Radiosonde RS92. (Vaisala, 2006)

A measurement technique which provides measurements of water vapour and other variables in the lower and upper atmosphere is the radiosonde. This is a balloon with attached measurement devices that communicates to a groundstation via a radio link. The KNMI is launching a radiosonde from De Bilt twice a day. They use the Vaisala RS92 Radiosonde, see Figure 2-9. It is computing its location using a GPS receiver. It measures pressure, temperature, and relative humidity (PTU). Some radiosondes also measure Ozone concentration. The pressure is determined with a Silicon cell, consisting of a silicon diaphragm with a piezo-resistive strain gauge diffused to it. The temperature is measured by a capacitive wire. The relative humidity is determined by a thin-film capacitor with a heated twin-sensor design; while one sensor is measuring, the other is heated to prevent ice formation. The wind velocity is computed from the track of the radiosonde. Note that velocity is a vector containing both the amplitude of the vector, also known as speed, as well as the direction.

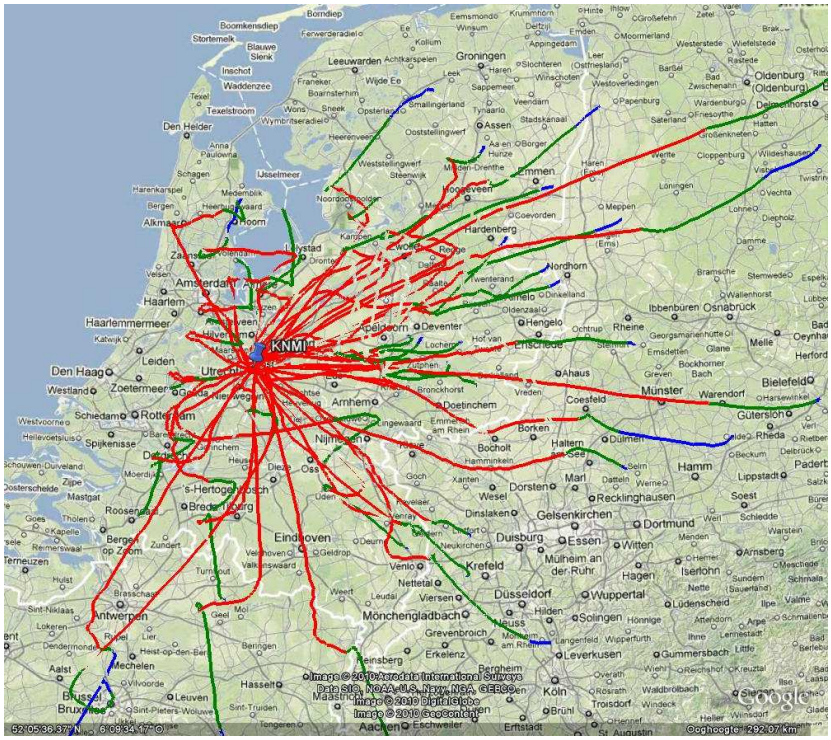


Figure 2-10: 2D graphic of radiosondes launched from De Bilt in 2009. Red indicates the radiosonde is going up, green indicates going down, and blue means the position was determined by using the last known position and velocity. (KNMI, 2010b)

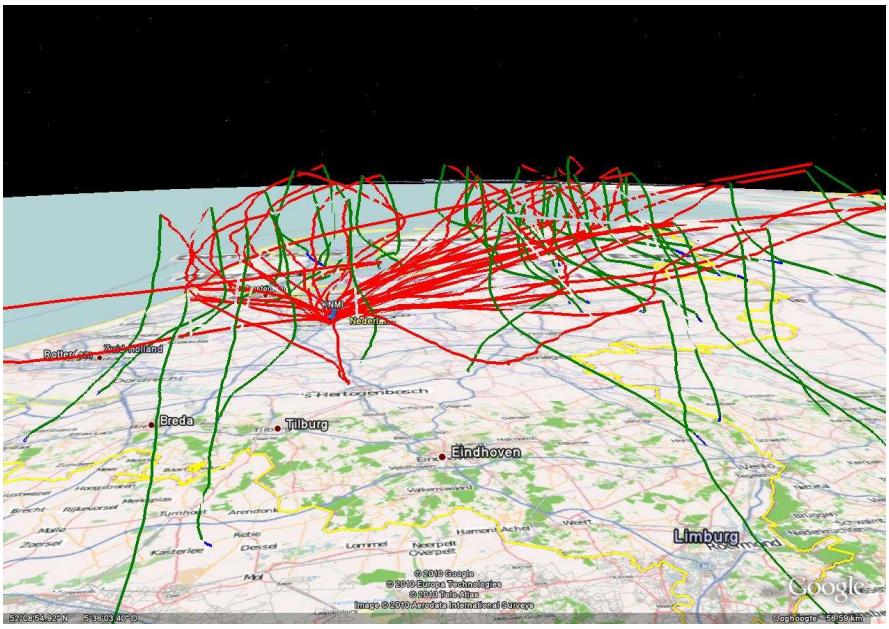


Figure 2-11: 3D graphic of radiosondes launched from De Bilt in 2009. Red indicates the radiosonde is going up, green indicates going down, and blue means the position was determined by using the last known position and velocity. (KNMI, 2010b)

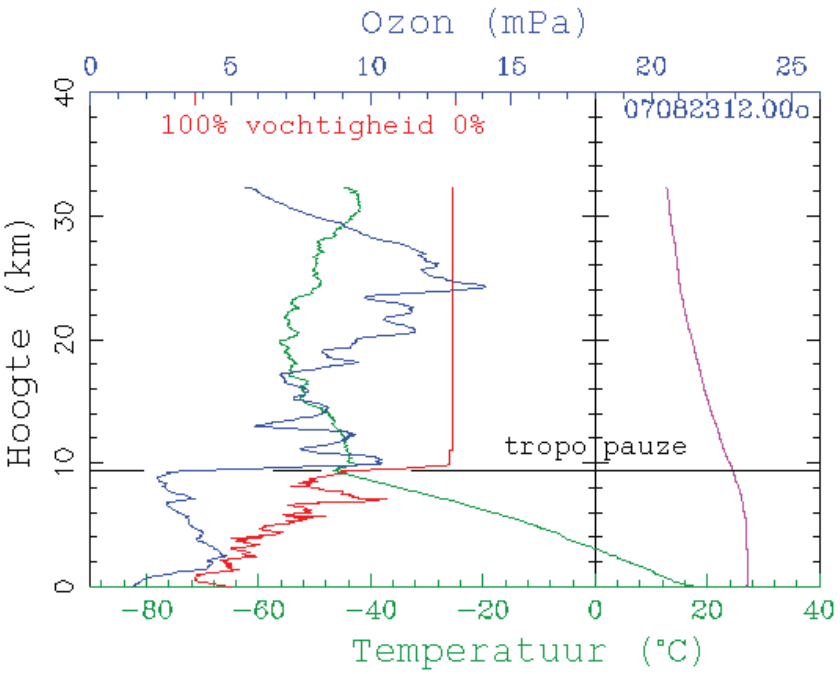


Figure 2-12: Measurement results of a radiosonde that landed in Voorburg, The Netherlands. Blue indicates Ozone (O_3) partial pressure, green the temperature, red the relative humidity in percentage, and magenta the internal temperature of the balloon. (KNMI, 2010b)

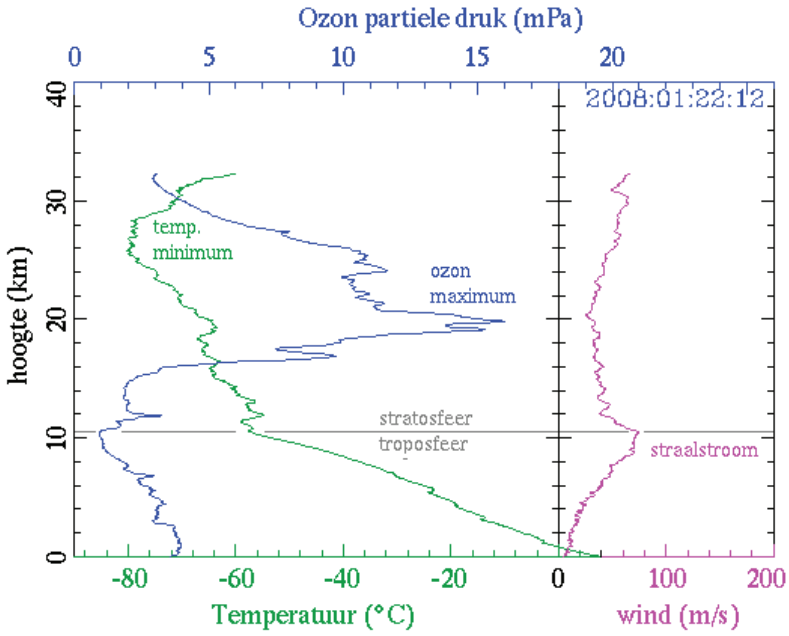


Figure 2-13: Measurement results of a Radiosonde that was launched at 22 January 2008. Blue indicates Ozone partial pressure, green the temperature, and magenta the wind velocity. (KNMI, 2010b)

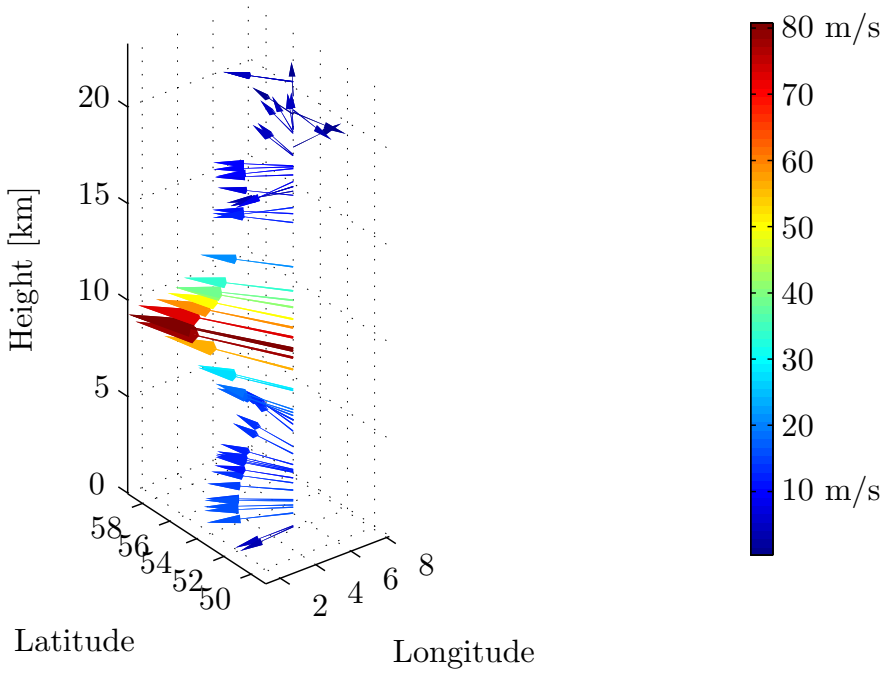


Figure 2-14: Vertical wind profile [m/s] measured by a radiosonde launched at 30 May 2010 from De Bilt, the arrow shows the direction where the wind is coming from.

In Figures 2-10 and 2-11 Google Earth pictures are shown with tracks of several radiosondes launched during 2009. The majority of tracks go from West to East, because of the prevailing wind-direction in The Netherlands. Most of the radiosondes reach a height between 20 and 30 km, while some of them travel up to several hundreds of kilometres. Results of radiosondes launched from De Bilt can be found in Figure 2-12 and 2-13. The RS92 radiosonde has a measurement range from 0% to 100% relative humidity, with a resolution of 1%, and an accuracy of 2%. Each measurement cycle takes 1 second and the measurement data is sent to the ground station with a 2400 bit/s downlink (Vaisala, 2006). In Figure 2-14 a wind profile obtained by a radiosonde is visualized.

The biggest drawback of using radiosondes is the coarse horizontal and temporal coverage, but still they are the main source of information about pressure, temperature, and humidity in the atmosphere.

EUMETNET also has a programme for automatically launching radio sondes from ships, called EUMETNET Automated Shipboard Aerological Programme (E-ASAP). At this moment 19 stations aboard ships are launching radiosondes. Due to bad weather or logistic problems there are only about 20–30 launches a month.

2-3-3 Ground Weather Radar

A Ground Weather Radar is a radar instrument mounted on a building or tower. It consists of a transmitter and receiver of microwave radiation (wavelength of mm–m) and a parabolic antenna. The antenna is scanning the surrounding atmosphere at relatively small angle intervals of e.g. 1 degree.

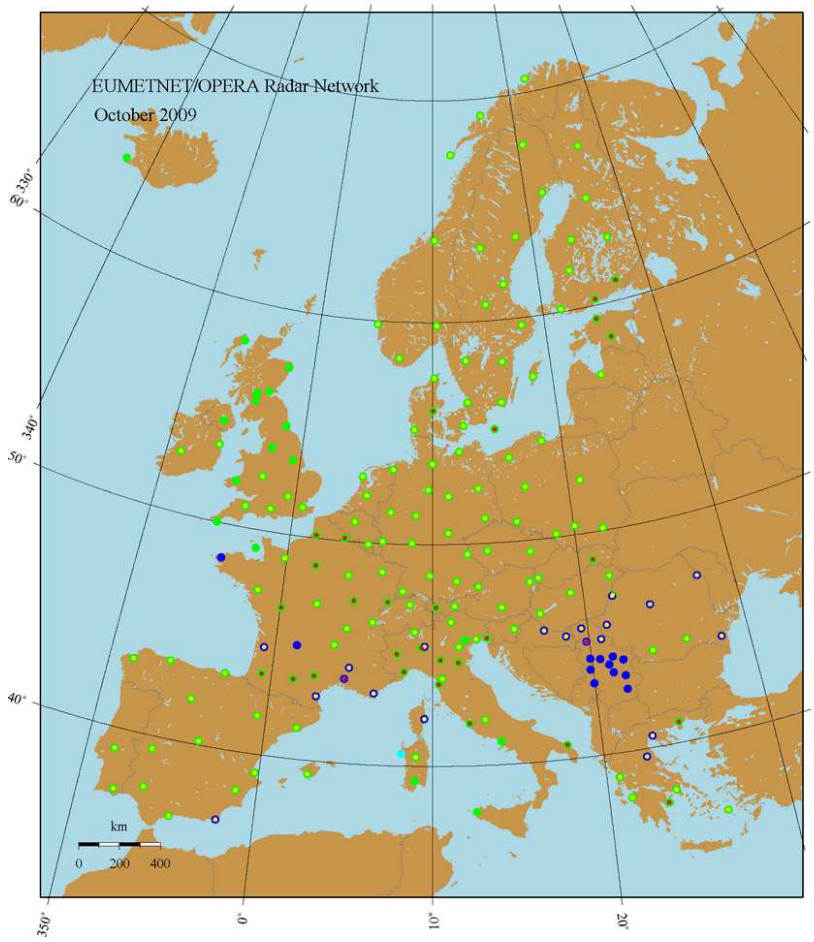


Figure 2-15: Map of Operational European Weather Radars in October 2009. Green indicates C-band, blue indicates S-band, cyan indicates X-band radar. A red centre indicates a Dual polarization radar and a yellow centre a Doppler radar. (KNMI, 2010b)

OPERA is the Operational Programme for the Exchange of weather Radar information. This is a five year programme of the European Network of Meteorological Services (EUMETNET) coordinated by the KNMI. In Figure 2-15 the weather radars of the OPERA network can be found. Primarily C-band radar is used and also some S-band radars. S-band has a lower frequency (2–4 GHz instead of 4–8 GHz) resulting in a lower attenuation (loss of intensity with distance) and its range is larger, but it is more expensive.

The KNMI is using two radars which scan 360 degrees in azimuth, and between 0.3 degree and 25 degrees elevation which is completed every 5 minutes. The wavelength used depends on the size of the particles to detect. To detect particles in the air, the radiation sent has to be reflected from the particles in all directions, this is called Rayleigh scattering. For a particle that can be assumed to be a Rayleigh scatterer, the wavelength of the transmitted radar wave has to be about 10 times the diameter of the particle. Raindrops have a maximum diameter

of about 6 mm. The KNMI is using 2 C-band weather radars, which have a wavelength of 5 cm. One of them is located at De Bilt, the other is located at Den Helder.

Another parameter which determines the performance of a radar is the pulse length and time between pulses. The pulse length is determining the volume of air scanned. The shorter the pulse length, the smaller the pulse volume and the higher the resolution, but the reflections are smaller and therefore harder to detect. The time between the pulses is used to detect reflections. If the pause between the pulses is too short, reflections from earlier sent pulses are received which is not desired. From the time difference between sending the pulse and receiving the reflection, the location of the scatterers in the volume can be determined. From reflectivity measurements the rainfall rate is computed using an empirical relationship between rainfall and reflectivity. The KNMI is using the following relationship:

$$Z = 200R^{1.6}, \tag{2-2}$$

where Z is the radar reflectivity and R is the rainfall rate. The values of the a-priori constants are characteristic for the local weather. For example for tropical regions other constants are used to describe the radar reflectivity as a function of the rainfall rate.

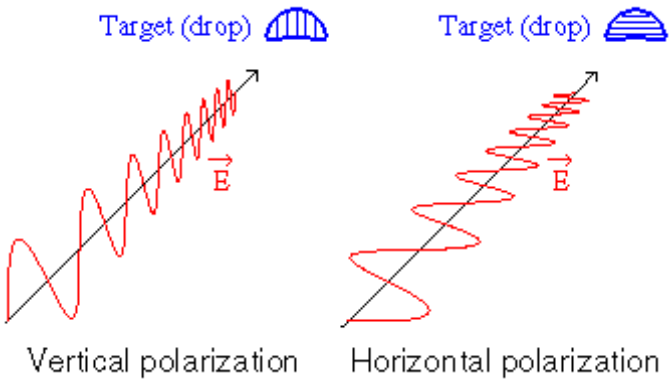


Figure 2-16: Horizontal and vertical polarization radiation. (Fabry, 2000)

There are several techniques used with radar for the detection of particles in the atmosphere. The first one is conventional radar, which sends a pulse and measures the travelling time of the pulse and the intensity of the pulse. The travelling time is used to calculate the position and the intensity the amount of precipitation. This kind of radar is not used any more at KNMI, Doppler radars are now used instead. They measure also the Doppler shift of the particles. This is used to calculate the velocity of the particles relative to the ground station. After the year 2000 also a newer technique became available. The polarization of a radar waves describe the orientation of the oscilations. Radar radiation can be horizontally or vertically polarized or a combination of them, see Figure 2-16. Usually only horizontally polarized microwave radiation is sent. By using both polarizations information on the kind of scatterer can be deduced. If the targets orientation is primarily horizontal, it will mostly react to horizontal polarized radiation and vice versa. By comparing the returns of both polarizations, one can determine the main orientation of the particles. In this way one can make a distinction between snowflakes, hailstones, and raindrops, (Fabry, 2000).

Data from the radar is visualized in several ways. From each pulse sent by the radar, reflections are coming back from different heights. Just viewing the return signal from one pulse is

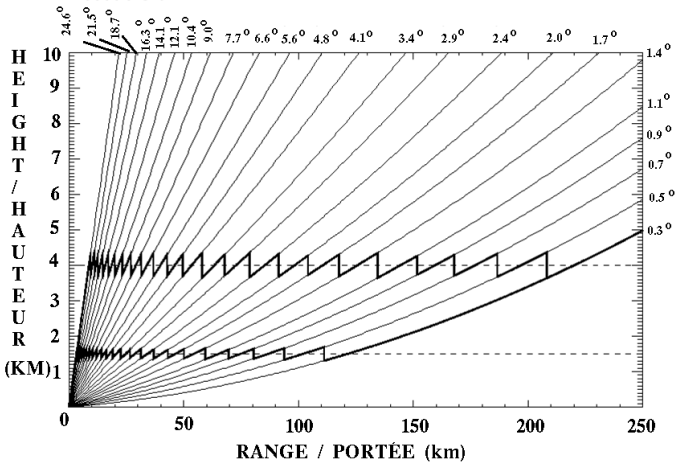


Figure 2-17: Trajectories at different angles of the radar beam. This diagram is showing angles used by the Canadian Meteorological Service of Environment Canada. (Wikipedia, 2009)

hard to interpret, because when going further away, the data it is reflecting is higher in the atmosphere. So for comparing different epochs or heights it is better to combine data from more pulses. This is visualized in Figure 2-17, the zigzag patterns represent data which is used to make at map between 1.5 and 4 km height. This kind of map is called a Constant Altitude Plan Position Indicator (CAPPI). Remind that rain on the CAPPI does not have to fall on the earth, it can evaporate in the air while falling. If this is visible it is called a virga.

There are some limitations to ground radars and the interpretation of the results of the measurements. First of all a radar beam can be blocked by e.g. mountains. Secondly, if a lot of radiation is absorbed or reflected close to the antenna, less radiation is available for scatterers further away (also called beam attenuation). When there is a clear sky early in the morning, there can be a strong temperature inversion, which is an increase of temperature with height (the temperature profile is inverted from its usual behaviour). When this happens the radar does not propagate further in the sky, but gets reflected down to earth. In this case a strong false signal is detected from the ground. This phenomenon is called Anomalous Propagation. For radar beams with low elevation angle also reflections from buildings and other low objects can return. This 'ground clutter' is more or less constant in time and can therefore be filtered out. (Environment Canada, 2006)

2-3-4 Wind Profilers

Another part of the EUCOS Operational Programme is the EUMETNET Wind Profiler programme (E-WINPROF). This is a network currently consisting of 26 European wind profiler radars. The KNMI operates one wind profiler at Cabauw. A wind profiler is used to measure the wind profile in vertical direction up to 30 km height, depending on the frequency of the radar used. Wind profilers with frequencies of about 50 MHz, 400 MHz, and 1 GHz exist. The one operating at Cabauw has a frequency of 1.29 GHz. The instrument sends a series of pulses and listens to the received echo. By measuring the Doppler shift, the radial velocity (velocity with respect to the instrument) of the particle is computed. For computing the 3 dimensional speed, at least three beams are needed. The main difference with a weather radar is that a

wind profiler is looking above with two side beams at North and West direction (15°–30° off zenith) a weather radar is able to scan 360° up to a few hundred kilometres range.

Another instrument which also measures wind profiles is a Doppler Sound Detection And Ranging (SODAR) instrument. The working principle is comparable to the (radar) wind profiler. Currently only at Schiphol airport a SODAR instrument is operational.

2-3-5 Water Vapour Radiometer

A radiometer is an instrument that measures incoming radiant energy. A water vapour radiometer measures the incoming radiation at two or more frequencies. For measuring IWV one of the channels is measuring in the water vapour channel (around 22.235 GHz) where the response to water vapour is nearly constant with height. The other channel has a higher frequency (around 60 GHz) and is measuring both the Liquid Water Path (LWP, total liquid water in a column) and IWV. By using multiple frequencies (multi-spectral) a vertical profile is obtained for water vapour. (Rose et al., 2005)

The KNMI is using the Humidity And Temperature Profiler from Radiometer Physics GmbH (RPG-HATPRO). Because the vertical resolution depends on the frequency, de vertical resolution decreases with height. For measurements between 0–2000 m the vertical resolution for tropospheric humidity profiles is 200 m, between 2000–5000 m this is 400 m, and 800 m is the resolution between 5000 m and 10000 m. The absolute humidity profile has an accuracy of 0.4 g/m^3 Root Mean Squared (RMS) and the relative humidity has an accuracy of 5% RMS. The IWV has an accuracy of 0.2 kg/m^2 RMS and 0.05 kg/m^2 noise in the signal. The optical resolution, in this case the angle of the Half-Power Beam Width (HPBW) at 22 GHz, is 3.5°. (Rose & Czekala, 2009)

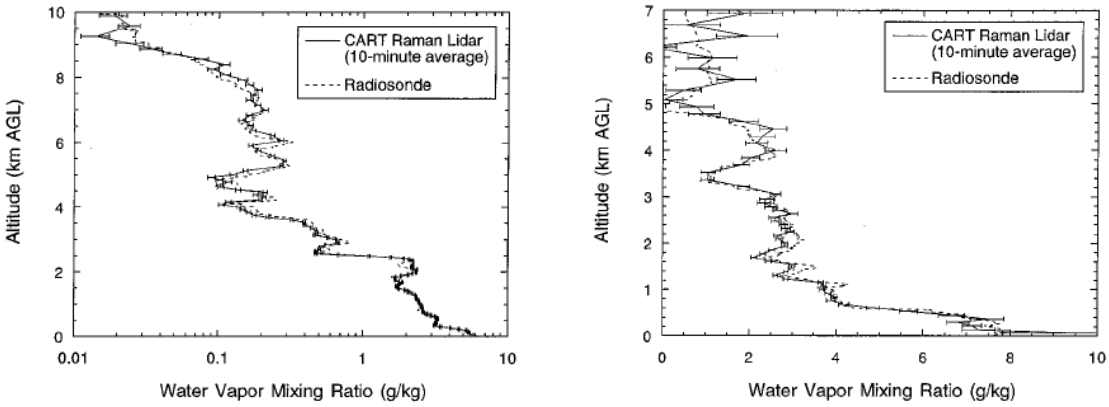
2-3-6 LIDAR

LIDAR is short for LIght Detection And Ranging. The technique is similar to radar (see section 2-3-3): the system sends a pulse with a certain wavelength and records the backscattered radiation from the atmosphere as a function of time. For LIDAR two different backscattered components are of importance. The first component is a strong elastically scattered one. This is the Rayleigh component. Elastically means that the molecules do not absorb energy and so the backscattered frequency is the same as the incoming frequency. This component is used for profiling clouds (consisting of ice and liquid water) and aerosols. The second component is inelastically scattered and is weaker. This component is called Raman scattering and is used for detection of water vapour. Because the molecules absorb or loose energy, the frequency is shifted (the energy of photons is proportional to the frequency). The size of the frequency shift is characteristic for the chemical composition of the particles in the atmosphere. By measuring both the frequency shift due to water vapour and nitrogen at the same time and taking the ratio, the resulting value is proportional to the water vapour mixing ratio:

$$w(z) \propto \frac{n_{\mathrm{wv}}(z)}{n_{\mathrm{nit}}(z)}, \tag{2-3}$$

in which $w(z)$ is the water vapour mixing ratio as a function of range z, $n_{\mathrm{wv}}(z)$ and $n_{\mathrm{nit}}(z)$ is the number density (concentration) of water vapour and nitrogen (N_2). For further details

see (Goldsmith et al., 1998). This system is able to operate day and night. The system continuously records profiles of the atmosphere at e.g. every minute. Once an hour LIDAR data are automatically transferred to a site data system.



(a) 6:30 a.m. at 28 September

(b) 12:30 p.m. at 9 September

Figure 2-18: Night-time (left) and daytime (right) profiles of water vapour recorded near Oklahoma. (Goldsmith et al., 1998)

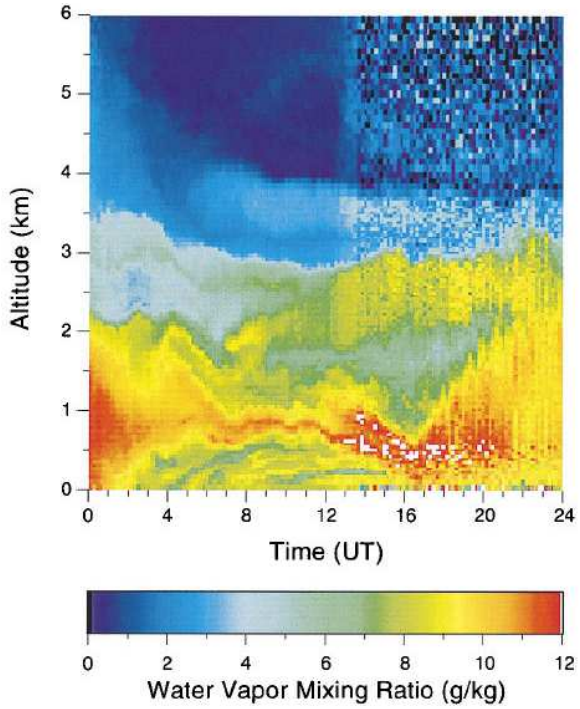


Figure 2-19: LIDAR false-colour image portraying the variation in the water vapour mixing ratio over the full 24-h period on 10 September 1996. Local time can be obtained by subtracting 5 h from UT. (Goldsmith et al., 1998)

In Figure 2-18 10-minutes averaged Raman LIDAR water vapour profiles are compared with radiosondes. In Figure 2-19 a 24-hour period in 1996 is depicted. Between 13 and 24 UT

(Universal Time, 8:00 a.m. to 7:00 p.m. local time) the figure is grainier. This is because this period is daytime, during which there is strong scattering due to solar radiation.

2-3-7 Aircraft

From 1980, measurements from sensors aboard several aircraft are used as meteorological observations. For each position during flight they record temperature, wind speed, and wind direction. In this way during take off and landing information is retrieved about the vertical variation of those variables. This information is sent by radio to UK Met Office. UK MET Office acts as the processing and web server. The Deutscher Wetterdienst (DWD) optimizes the data streams and provides a web portal, which is accessible for Meteorological institutes. In the EUCOS area about 40 airports deliver 3 hourly profiles of wind and temperature, in total about 780 profiles a day. This network of aircraft delivering meteorological measurements in Europe is called EUMETNET Aircraft Meteorological DAta Relay (E-AMDAR). Humidity sensors aboard aircraft are also in development. These sensors could partly replace expensive radiosondes. For the period 2009–2011 the extension of the E-AMDAR to new European Union countries is investigated. Also a conversion to true height in the measurements is introduced to make modeling of parameters easier. (KNMI, 2010b)

2-3-8 GPS

The Global Position System (GPS) is the main Global Navigation Satellite System (GNSS) now in use and is the only one currently fully operational (May 2010). Its original name is NAVSTAR GPS (Navigation Signal Timing And Ranging GPS). It is maintained by the US Air Force and consists of more than 30 satellites. The satellites are at a height of about 20,200 km in six orbital planes with an orbital period of about 11 hours and 58 minutes and at an inclination of 55°. Satellites at the moment are using 2 carier frequencies, 1575 MHz (L1) and 1227 MHz (L2). In the near future a third frequency will be transmitted to the users at 1176 MHz (L5). The satellites transmit a sequence of zeros and ones, also called a code. On the L1 frequency both a precise P(Y)-code and a less precise C/A-code are transmitted. On the L2 frequency only the P(Y)-code is present. Both codes are modulated on the carrier as Code Division Multiple Access (CDMA). The different satellites are distinguished by their different Pseudo Random Noise (PRN) codes.

Because the signal from the satellite to the receiver is travelling through the atmosphere which is not vacuum, the signals are delayed. This delay can be separated into ionospheric delay and neutral atmospheric delay. The ionospheric delay is caused by charged particles in the ionosphere. The neutral atmospheric delay is caused by neutral molecules and atoms in the troposphere and stratosphere (see also Subsection 2-1-3). The ionospheric delay can be mitigated using 2 frequencies, because this delay is frequency dependent.
The neutral delay in the line of sight is called the Slant Total Delay (STD)

$$\mathrm{STD} = \int (n-1)\mathrm{d}s = 10^{-6} \int_{sat}^{rec} N(s)\mathrm{d}s, \tag{2-4}$$

where n is the refractive index (ratio of speed of light in vacuum with respect to the medium), N is the neutral atmospheric refractivity, and s is the path from the GPS satellite to the

receiver. This neutral atmospheric refraction value is empirically related to meteorological variables by (Davis et al., 1985) as

$$N = k_1 R_d \rho + k_2' \frac{p_w}{T} Z_w^{-1} + k_3 \frac{p_w}{T^2} Z_w^{-1}, \tag{2-5}$$

where ρ is the total density of dry air and water vapour; R_d is a specific gas constant for dry air; $k_2' = k_2 - mk_1$ and m is the ratio of molar masses of water vapour and dry air. The constants k_1, k_2, and k_3 have the values $k_1 = 7.76 \cdot 10^{-3}$ [K/Pa], $k_2 = 7.04 \cdot 10^{-1}$ [K/Pa], and $k_3 = 3.74 \cdot 10^3$ [K^2/Pa], (Haan, 2008); p_w is the partial pressure of water vapour; T is the temperature in [K], and Z_w is the compressibility factor for water vapour, which is a measure of how much the thermodynamic properties of water vapour deviate from an ideal gas. Inserting Equation (2-5) into Equation (2-4) gives 3 terms to integrate. The first term is the Slant Hydrostatic Delay (SHD) and the second and third terms together are called the Slant Wet Delay (SWD). The hydrostatic delay depends on the total ground pressure, latitude, and height above the geoid. The values for the temperature, partial pressure, and total density depend on the height of the receiver above the geoid. Because the compressibility of water vapour is very close to 1, the integration of the SWD can be done by using profiles of temperature and relative humidity

$$\mathrm{SWD} = 10^{-6} \int_{sat}^{rec} (k_3 + k_2' T_m) \frac{p_w}{T^2} \mathrm{d}s, \tag{2-6}$$

where the first term of the wet delay is approximated by

$$\int_{sat}^{rec} \frac{p_w}{T} \mathrm{d}z = T_m \int_{sat}^{rec} \frac{p_w}{T^2} \mathrm{d}z, \tag{2-7}$$

were T_m is known as the "mean temperature". This mean temperature is approximated by a function of the temperature at the surface T_s (see Subssection 2-1-3) thus $k'(T_m) = k(T_s)$

For computation of vertically integrated water vapour, STD and SWD are converted to 90 ° elevation, called zenith direction (straight up). This is done using the Niell's hydrostatic m_h and wet mapping functions m_w as a function of elevation β. These functions were published by (Niell, 1996). The hydrostatic function covers the delay due to the neutral gasses and the non-dipole part of water vapour. The wet mapping function covers the delay due to the dipole part of water vapour. The resulting delays are the Zenith Total Delay (ZTD) and Zenith Wet Delay (ZWD). The ZWD can be computed from Equation (2-6) as

$$\mathrm{ZWD} = \frac{\mathrm{SWD}}{m_w(\beta)} = \frac{1}{m_w(\beta)} \underbrace{10^{-6}(k_2 R_v + k_1 R_d + R_v k_3 / T_m)}_{k'(T_m)} \int_z \rho_w \mathrm{d}z, \tag{2-8}$$

where R_v is the specific gas constant for water vapour and ρ_w is the water vapour density.

The Zenith Hydrostatic Delay (ZHD) is computed using the model (Saastamoinen, 1972)

$$\mathrm{ZHD} = 10^{-6} k_1 R_d \int_z \rho \mathrm{d}z \approx \frac{2.2768 \cdot 10^{-5} p_a}{1 - 2.66 \cdot 10^{-3} \cos(2\phi_{GPS}) - 2.8 \cdot 10^{-7} z_{GPS}} \doteq \mathrm{ZHD}_{Saas}, \tag{2-9}$$

where p_a is the total ground pressure, ϕ_{GPS} is the latitude of the GPS receiver, and z_{GPS} the height of the GPS receiver. The last term of Equation (2-8) is the IWV:

$$\mathrm{IWV} = \int_z \rho_w \mathrm{d}z \tag{2-10}$$

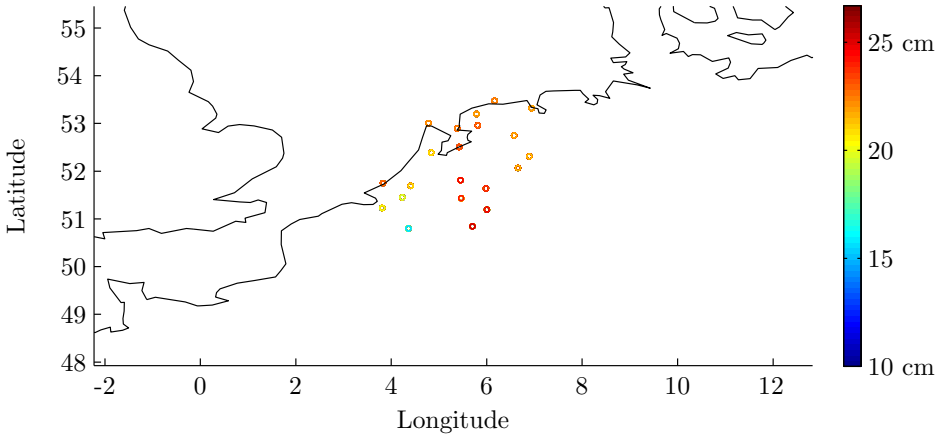


Figure 2-20: IWV in [cm] measured by 21 GPS ground receivers.

Using ZTD = ZWD + ZHD and combining Equations (2-9) and (2-10) gives for the computation of the IWV

$$\text{IWV} = \frac{1}{k(T_s)}\text{ZWD} = \frac{1}{k(T_s)}(\text{ZTD} - \text{ZHD}_{Saas}). \tag{2-11}$$

So either the IWV is calculated from measurements of ZWD or from measurements of ZTD and the Saastamoinen model approximation. The value of $k(T_s)$ is around 6.5 [kg m^{-3}].

The ZHD is the largest but most stable delay of the two parts of the ZTD and typically ranges from 2.0 to 2.3 m at sea level and the ZWD varies from approximately 0.05-0.06 m at the poles to 0.30 to 0.35 m in the tropics (Haase et al., 2003). In Europe the EUMETNET has set up the GNSS water vapour programme (E-GVAP). Currently about 400 stations are measuring water vapour and send these measurements to the UK Met Office. In the Netherlands about 40 stations are situated. Figure 2-20 displays a map of The Netherlands with IWV values computed at 22 stations at the 30th of May in 2010.

2-3-9 Satellite

One advantage of using satellites for measuring water vapour in the atmosphere is the larger area the satellite can cover if it is moving with respect to the earth. A polar orbit is an orbit where the satellite rotates around the earth with an angle with respect to the equator (inclination) of 90° or close to it. In this configuration the satellite passes above both poles and if the rotation of the earth has a different speed, the satellite is covering the whole earth. Drawback is the time between successive passes of the same spot on earth, which can be up to several weeks.

SSM/I

An example of a passive microwave sensor is the Special Sensor Microwave Imager (SSM/I). Passive indicates it is using the Sun radiation instead of sending radiation itself. SSM/I is using frequencies ranging from 19.35–85.5 GHz. The spatial resolution is 49 km × 63 km and therefore it can only be used for large scale variability. Moreover it is only capable of

measuring above water surfaces because the surface itself should not emit too much radiation. So also in case the Sun reflection on the ocean is seen at the imager (sun glitter) no data is available. An advantage of using a microwave sensor in space is that clouds do not affect the retrieval. Current SSM/I sensors are flown aboard satellites of the United States Air Force Defense Meteorological Satellite Program (DMSP). In Figure 2-21 the total water vapour column retrieved by the SSM/I sensor at 30 June 2010 can be seen.

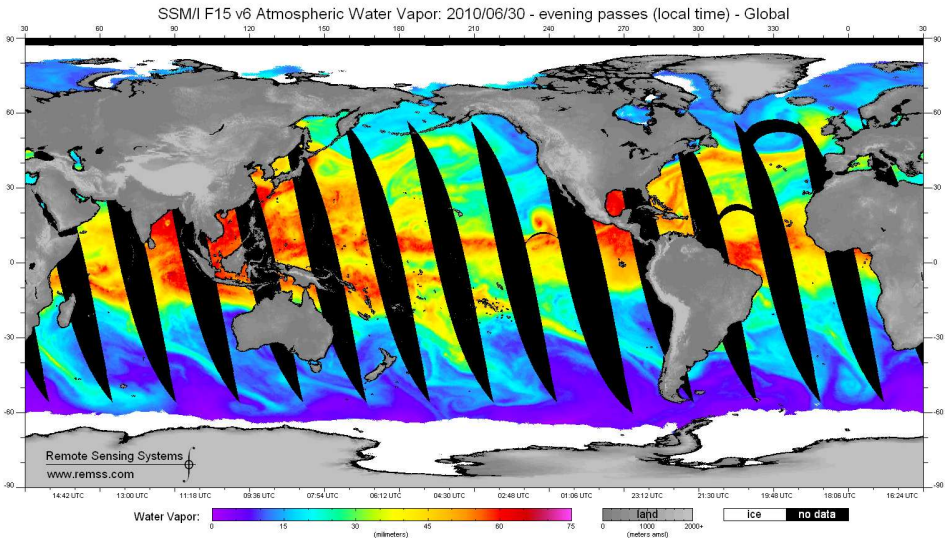
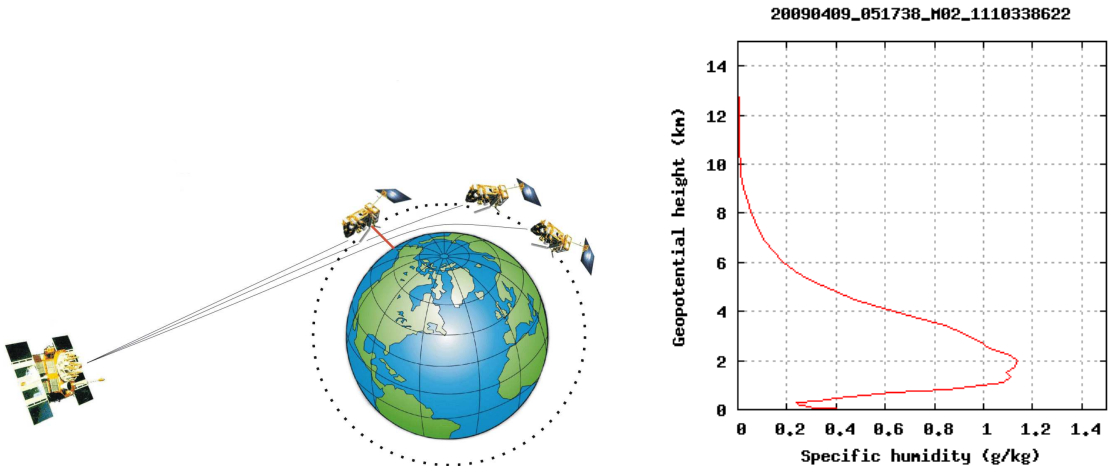


Figure 2-21: SSM/I Total water vapour column on 2010/06/30. (www.ssmi.com)

GRAS

The European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) is also flying satellites for measuring water vapour in the atmosphere. One of her programs is the EUMETSAT Polar System (EPS). It consists of a series of polar orbiting satellites, flying from 2006. The satellite series improve long term operational meteorological and environmental forecasting and global climate monitoring. The satellites carry several instruments for measuring water vapour and wind.

One of the sensors aboard the EPS is a GNSS Receiver for Atmospheric Sounding (GRAS). By measuring the Doppler shift of the received GNSS signal combined with data received by GNSS ground stations, atmospheric profiles of the humidity can be retrieved, see Figure 2-22(a). See Figure 2-22(b) for a typical humidity profile obtained by a GRAS sensor. Domain of the measurements is from ground to about 15 km (100 hPa) and it can measure up to 50 g/kg, with a precision of 0.001 g/kg and a vertical resolution of about 250–500 m which depends on the back-ground pressure level (The EUMETSAT Network of Satellite Application Facilities, 2010). GRAS is one of the many satellites using the radio occultation technique. At this moment MetOp, COSMIC, CHAMP, and GRACE-A are monitoring the quality of the GRAS instrument. A near real-time specific humidity profile product is currently in development.



(a) GRAS principle (b) GRAS humidity profile

Figure 2-22: GRAS principle and typical GRAS humidity profile, April 9 2009 05:17:38UT at 76.02N 15.28W. (The EUMETSAT Network of Satellite Application Facilities, 2010)

IASI

Another instrument of the EPS is the Infrared Atmospheric Sounding Interferometer (IASI). Interferometry means deriving properties of a medium based upon the pattern of two or more waves. In this case infrared light waves are used to deduce properties of the atmosphere to derive temperature and humidity profiles. The instrument is decomposing the thermal heat emission from the earth into different wavelengths, using a Michelson interferometer. A Michelson interferometer splits an incoming beam into two paths, reflects them on mirrors and combines them again to create an interference pattern. IASI is only able to measure under cloud free conditions because of the infrared radiation used, which is blocked by clouds. It has a an accuracy of 10%, a vertical resolution of 1–2 km in the lower troposphere, and a horizontal resolution of 25 km. (EUMETSAT, 2010)

HIRS

The EPS is also using a High-resolution Infrared Radiation Sounder series (HIRS). Latest version is the HIRS/4 which is a 20-channel infrared scanning radiometer. It provides total water vapour content and with an altitude of 837 km the measured area on earth is 10 km at nadir (exactly beneath). Drawback again for this instrument is it is only capable of measuring under cloud-free condition.

Interferometric Radar

Interferometric Synthetic Aperture Radar (InSAR) is a technique that uses (mostly space-borne) radar measurements. Interferometry uses two or more radar phase images. Synthetic aperture indicates the processing of radar data which creates a very narrow effective beam. The signal which is measured by InSAR consists of contributions of topography, deformations,

atmosphere, orbital effects, phase ambiguities, and noise. So one of the output of InSAR processing is an estimation of the difference in atmospheric delay in both images. Because this method is currently being developed it will not be treated any further. For more information see (Hanssen et al., 1999) and (Hanssen et al., 2006).

MHS

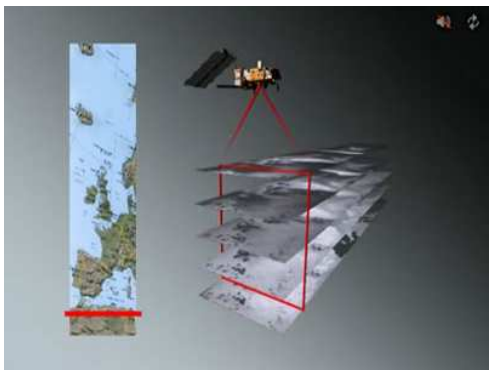


Figure 2-23: MHS provides atmospheric humidity at different altitudes layers in the atmosphere. (www.esa.int)

Another radiometer used by EPS is the Microwave Humidity Sounder (MHS), which is a cross-track scanning, five-channel microwave radiometer. It is using frequencies from 89 to 190 GHz for measuring atmospheric water vapour, including atmospheric ice, cloud cover, and precipitation. For measuring in the lowest layers in the atmosphere frequencies of 89 GHz and 157 GHz are used. 190 GHz and 183.3 GHz are used to measure water vapour at larger heights. The footprint of MHS at nadir is about 16 km. See Figure 2-23 for a picture of the MHS scanning several layers in the atmosphere (PDF readers can watch a movie by clicking on the picture)

MERIS

In 2002 the European Space Agency (ESA) launched the Environmental Satellite (Envisat). This is a polar-orbiting satellite, which observes the Earth's atmosphere, ocean, land, and ice. Its orbiting height is about 800 km with a repeat cycle of 35 days and an inclination of 98.55°. Because most of the instruments aboard have a wide swath, this configuration provides a complete coverage of the Earth within 1–3 days. One of the global mission objectives is to provide near real time data (from a few hours to one day from sensing) of atmospheric variables, like temperature, pressure, water vapour, and cloud top height. One of the instruments aboard Envisat is the Medium Resolution Imaging Spectrometer (MERIS), which measures among other parameters the water vapour total column content for all surfaces.

MERIS is a multi-spectral radiometer operating in the visible and infrared spectral range. It measures sun reflection in 15 bands between 390 nm to 1040 nm. Because it measures sun reflections, observation are limited to 43.5 minutes per orbit (100.59 minutes for a full orbit). One of the missions is to improve global monitoring of cloud properties and their processes. It provides data on cloud top height and optical thickness, water vapour column

content, and aerosol properties. By means of a push broom imaging spectrometer with a swath width of 1150 km, MERIS produces images of the Earth. For across track sampling it uses 5 cameras with a small overlap between them. For along track sampling it uses the motion of the satellite. See Figure 2-24 for an illustration of the Field of View (FOV) of the instrument.

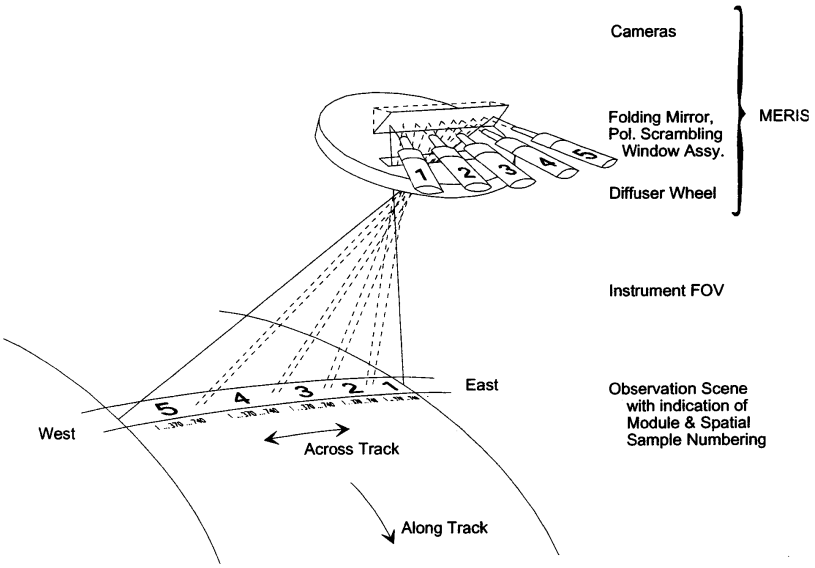


Figure 2-24: MERIS FOV, camera tracks, pixel enumeration and swath dimension. (envisat.esa.int)

MERIS has two options for imaging data. The first option is Full Resolution data with 300 m resolution at nadir. The second option is Reduced Resolution with 1200 m resolution at nadir. Full resolution data is used for land and coastal zones, reduced resolution is used for water surfaces. Reduced resolution data is acquired by combining 4×4 adjacent pixels in across-track and along-track direction. For measuring water vapour in the atmosphere band 14 (885 nm, outside the water vapour absorption region) and band 15 (900 nm, inside the water vapour absorption region) are used. An algorithm is estimating the amount of water vapour based on a differential absorption in these bands. The algorithm uses a second-degree polynomial of the logarithm of the two radiances at the top of the atmosphere. The general form of the algorithm used is

$$W = k_0 + k_1 \log(\frac{L_{Ch15}}{L_{Ch14}}) + k_2 \log^2(\frac{L_{Ch15}}{L_{Ch14}}), \tag{2-12}$$

where W is the column amount of water in [g m^{-2}], L_i is the radiance measured in channel 14 and 15 (Fischer & Bennartz, 1997). The constants k_0, k_1, and k_2 depend among others on the observation geometry. For measurements above ocean they also depend on the aerosol optical depth, which is derived from MERIS channels 9, 12, and 13. For measurements above land, also the surface pressure has to be taken into account. Also in case of sun glitter, other constants has to be used, because the radiation measured can be comparable to land surface. Above oceans the measurements show a bias of 0.05 g/m^2 and a root mean squared error (RMSE) of 0.49 g/m^2 with respect to measurements of the Envisat

Microwave Radiometer (MWR) instrument. Measurements above land show a bias with respect to GPS measurements (see Subsection 2-3-8) of 0.03 g/m^2 and a RMSE of 0.17 g/m^2. (MERIS Quality Working Group, 2006)

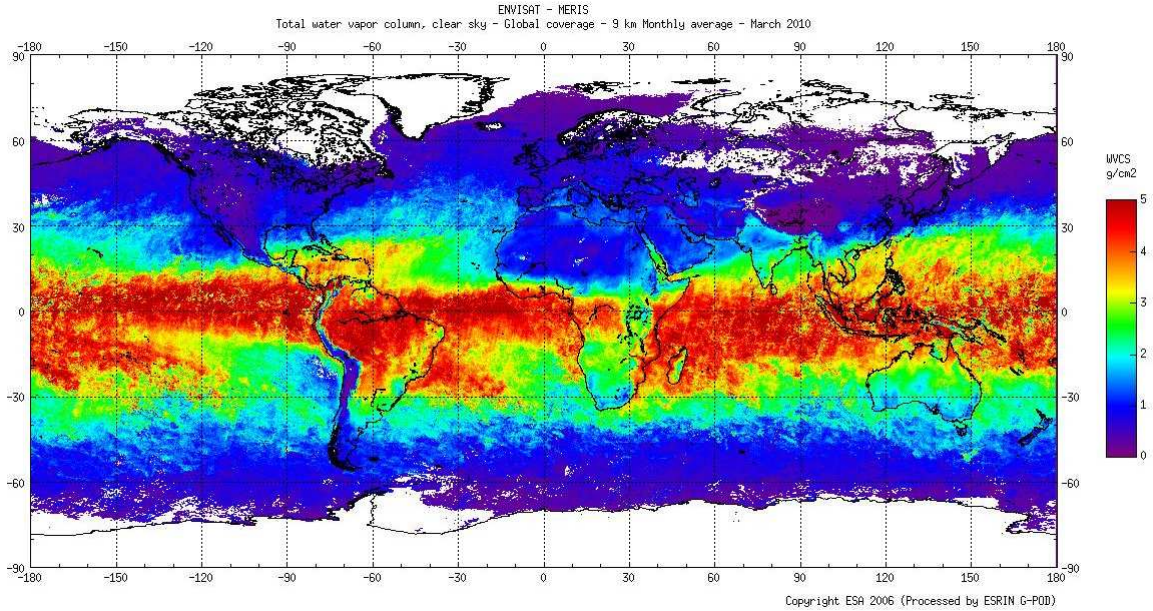


Figure 2-25: MERIS total water vapour column, clear sky.

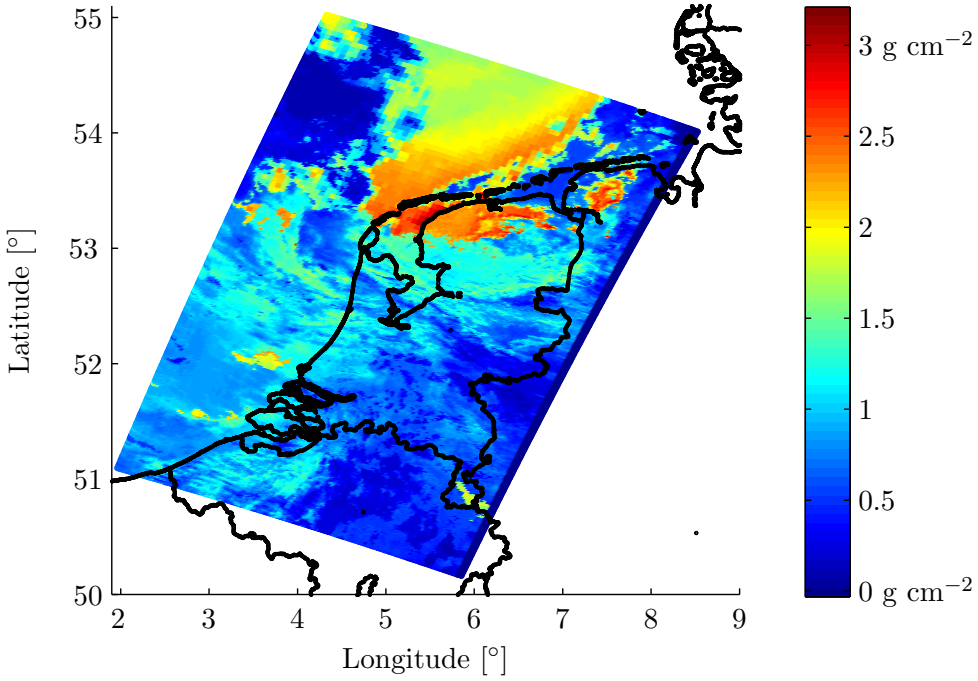


Figure 2-26: MERIS total water vapour column [g cm^{-2}] on 31 May 2010 at around 11:00 UTC, above land, water, or clouds.

In Figure 2-25 an example of MERIS total water vapour column over the entire Earth is visualized. In Figure 2-26 an example of MERIS total water vapour column at 13 May 2010 in the surrounding of The Netherlands is visualized. Other instruments aboard satellites which also measuring temperature, pressure, water vapour and wind are coexisting as well. Examples are the Moderate Resolution Imaging Spectroradiometer (MODIS), and Meteosat Second Generation (MSG). MODIS is an instrument aboard the Terra and Aqua polar orbiting satellites, whereas MSG refers to a series of geostationary satellites. Both have lower resolution than MERIS an for that reason, they will not be treated any further.

Chapter 3

Spatial Interpolation Of GPS Water Vapour

In this chapter the process of obtaining GPS IWV maps is described. The chapter starts with a description of the GPS IWV measurements in Section 3-1. In Section 3-2 it is described which method is used to interpolate between the GPS ground stations to obtain an integrated water vapour map.

3-1 GPS Integrated Water Vapour

Current input data for Integrated Water Vapour maps are the IWV values as computed by GPS ground stations in The Netherlands, see Subsection 2-3-8.. An example of such an IWV map can be found at the website of the KNMI (www.knmi.nl). The IWV values are computed every 15 minutes and sent to a central database. An Id consisting of 4 characters is defining the area in which a GPS station is located, e.g. 'RDAM' stands for the GPS receiver located in Rotterdam. Also, for each station the longitude and latitude in degrees are given for a more precise definition of the location. The time epoch for each measurement is given in YYYYMMDDHHmmss with the time in UTC. In this report all timestamps are in UTC,

Table 3-1: Preview of GPS IWV input data. GPS indicates the GPS ground station, the second row gives the date and time in UTC, ZTD is the Zenith Total Delay, err is the error in ZTD, ZWD is the Zenith Wet Delay, IWV is the Integrated Water Vapour value, and QC is a quality control flag. The longitude and latitude are in derees.

GPS	YYYYMMDDHHmmss	latitude	longitude	ZTD	err	ZWD	IWV	QC
AMST	20100530000000	52.38731	4.83562	2415.1	2.4	136.8	21.5	0
AMST	20100530001400	52.17980	5.95982	2414.5	2.6	136.2	21.4	0
AMST	20100530002900	52.17980	5.95982	2418.2	2.7	139.9	22.0	0

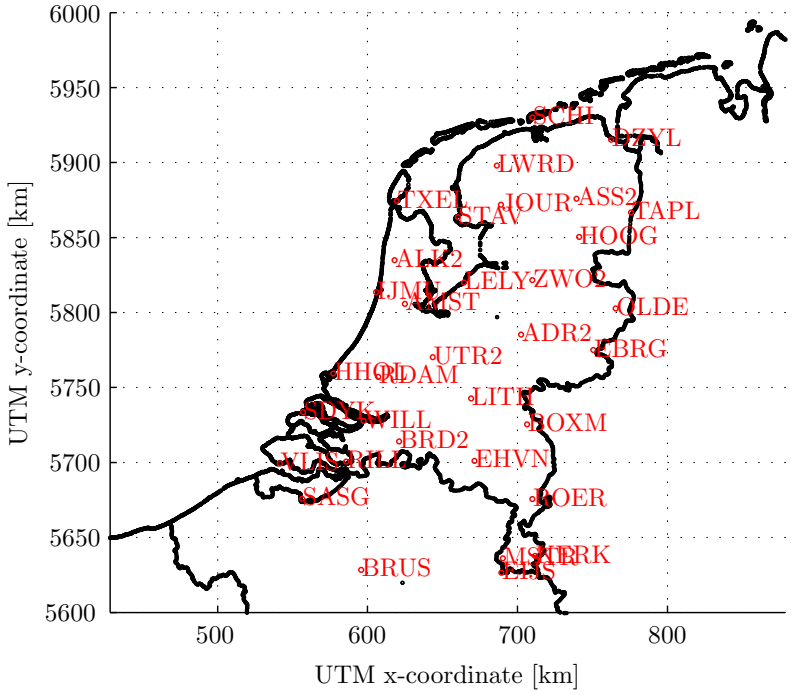


Figure 3-1: Location of GPS stations used in the processing

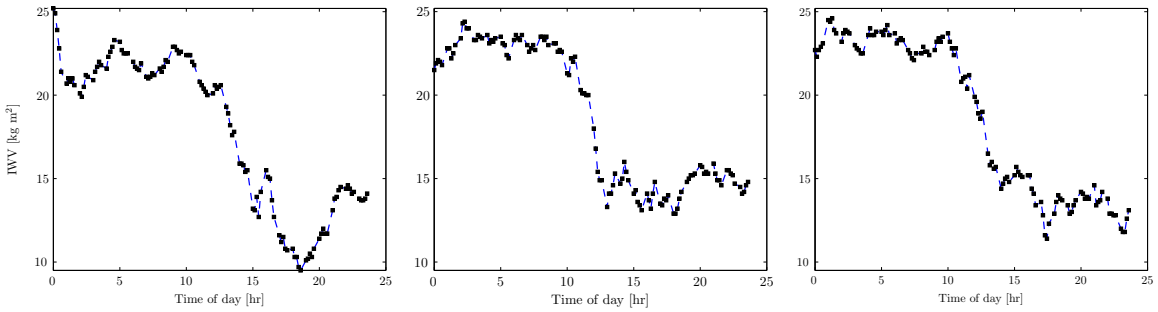
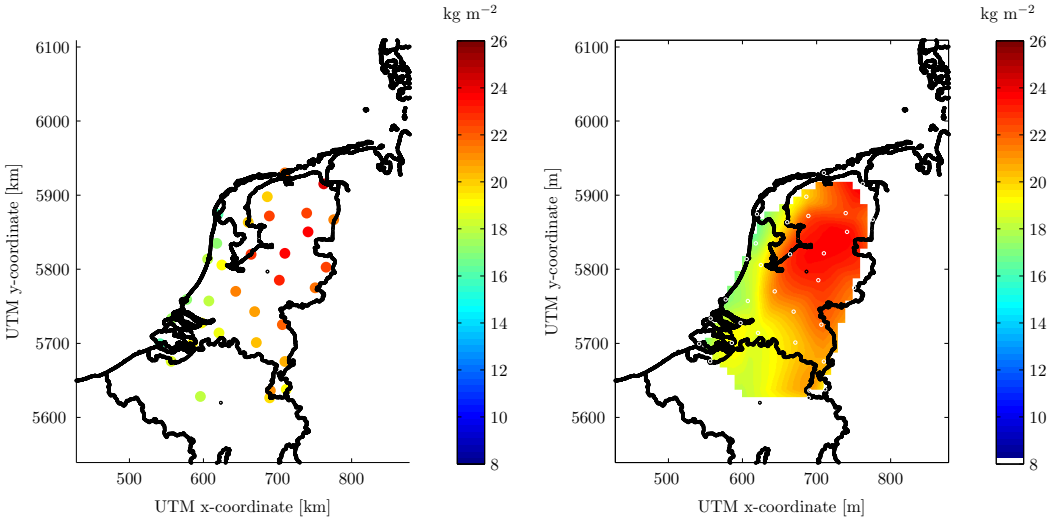


Figure 3-2: IWV as a function of time at 30 May 2010 for Eindhoven (a), IJmuiden (b), and Stavoren (c).

unless stated otherwise. Finally the IWV value is given in [kg m^{-2}] with one decimal number. The given IWV values are already converted to IWV values at sea level (antenna height of the ground stations is not zero and different for each station). Information about the stations and measurements is provided as an ASCII file. By means of a MATLAB function (*gps.m*) the information in the ASCII file is put into a *GPS.mat* file which can be imported by the main program. In Table 3-1 a sample of GPS IWV input data is displayed. In Figure 3-1 the location of all GPS stations used for processing are visualized. The x and y coordinates are in UTM zone 31U (see Section 4-1). In Figure 3-2 time series of one day of IWV measurements for three GPS ground stations are shown.

In Figure 3-3(a) the position of the GPS ground stations are visualized. The color of the dots



(a) Initial IWV values at the ground

(b) IWV map by cubic interpolation

Figure 3-3: (a) Initial IWV values from GPS measurements at the ground stations at 12:00 on 30 May 2010 and (b) an IWV map from those measurements obtained by cubic interpolation, both in [kg m^{-2}]

indicate the IWV value as calculated by the ground stations on 30 May 2010 at 12:00 UTC. The xy-plane of the figure is also the domain which is used in further processing. Figure 3-3(b) shows the water vapour map based on these measurements, obtained by a cubic interpolation (Christopher & Gold, 1993) to get a first impression of the distribution of water vapour over The Netherlands. Note that the small black dot just east of Brussels is a lake and not a ground station.

3-2 Kriging

From the IWV value at each station we want to create a two dimensional map. In Figure 3-3(b) a map was already shown which was created using cubic interpolation. This is triangle-based cubic interpolation, which produces a smooth surface. No values are generated outside the convex hull. The convex hull is the area which would be obtained if a virtual rope would be spanned around all the stations. Cubic interpolation respects the input values, which means the interpolated surface passes through all input values. The technique of cubic interpolation was only used at the beginning of the thesis for a quick representation.

In this thesis Kriging interpolation will be used. This method is used for several reasons. Among others, because it does not give large differences on short distance, which occurs for example when nearest neighbor interpolation is used. Another advantage of Kriging is that it is able to include an error description of the input measurement data. Another great feature of Kriging interpolation is the variance map created, which gives an indication of the accuracy of the interpolated results. Also it is pretty easy to set up a Kriging interpolation and a lot of packages are already available. Moreover Kriging is already known to the author, whereas e.g. 3DVAR (Barker et al., 2001) which is used a lot in the field of meteorology, is not. Of course

there are also disadvantages of Kriging. Some of them will be mentioned, when reviewing the obtained water vapour maps in Section 5-4 and Section 6-4.

Kriging is a technique based on the Master thesis of Daniel Gerhardus Krige. Krige used it to estimate a distance-weighted average of gold grades at the Witwatersrand reef complex in South Africa (Krige et al., 1989). It is a technique used to estimate the value at an unobserved location, based on surrounding observations. Besides giving an estimate for each location, Kriging also provides an error variance at each location. Several kinds of Kriging exist. Ordinary Kriging is mostly used. It assumes second order stationarity and an unknown constant mean. This means there is no drift and the variability (dissimilarity) between pairs of random variables stabilizes with increasing distance. Another assumption is that enough measurements are available to estimate the variogram. If this is not the case, the variogram will not represent the measurements very well, which results in a poorer interpolation.

3-2-1 Variogram

A variogram is a graph representing the dissimilarity between any two observations as a function of distance h between two observations. The dissimilarity γ_{ij} between observation points x_i and x_j is computed by

$$\gamma_{ij} = \frac{(z_i - z_j)^2}{2}, \tag{3-1}$$

where z_i and z_j are the values at those observations points. The distance between observations is computed in the xy-plane from the UTM coordinated.

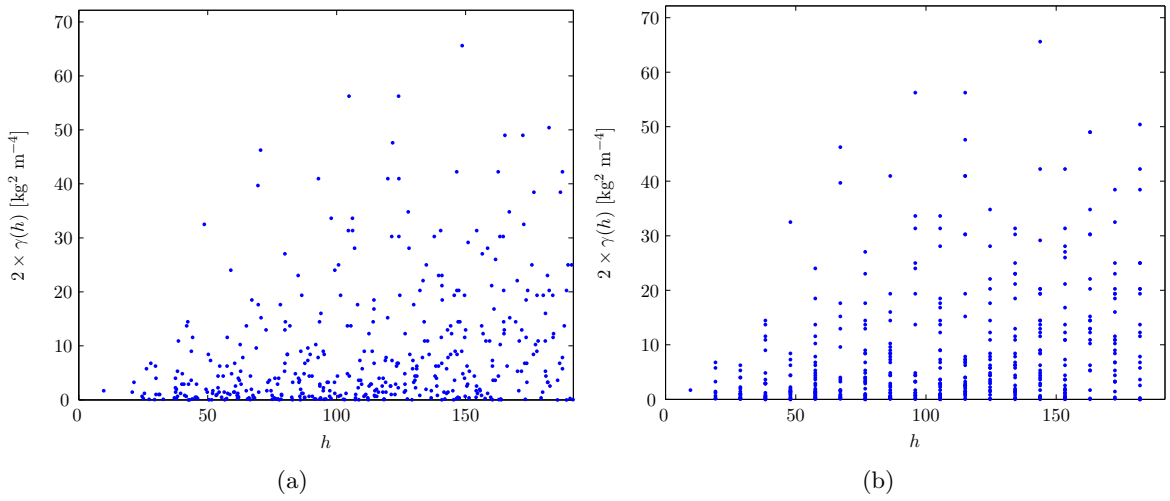


Figure 3-4: (a) Experimental dissimilarity cloud and (b) experimental dissimilarities grouped into 20 bins

In Figure 3-4(a) the dissimilarity cloud of the IWV measurements from Figure 3-3(a) is visualized. In Figure 3-4(b) the dissimilarities are grouped into 20 bins. Next step is to take the average in each bin. In Figure 3-5 the mean of each bin is plotted as a red square. The standard deviation σ in each bin is denoted by the total length of the blue vertical lines. Note the difference in y-axis scale compared to Figure 3-4(a) and 3-4(b). The largest distance

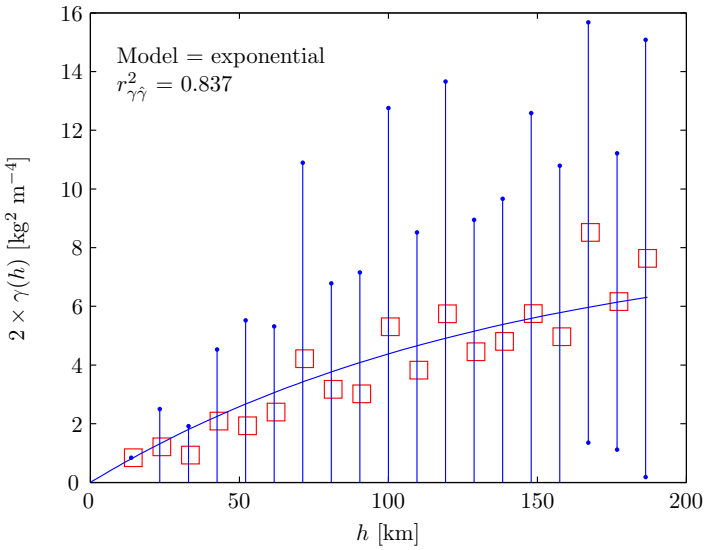


Figure 3-5: Theoretical (semi-)variogram based on an exponential model fit through the means of dissimilarities in each of the bin in Figure 3-4(b), the total length of the vertical lines denote the standard deviation in each bin

between two GPS ground stations used for processing is about 332 km. This is the distance between the ground station in Brussels and Delfzijl. However to limit the range over which dissimilarities are computed, the maximum distance is set to half of the maximum distance in the dataset. This is done to limit the influence of large differences in IWV on the other side of The Netherlands. E.g. if a real high water value is measured at the coast while there is a real low value at the inland, the dissimilarity between the two is high. That would generate a variogram which is higher for larger distance in Figure 3-5. That could have an effect on the whole prediction area. This is however also depending on the weighting function used to compute the best fitted function through the mean of the dissimilarities as a function of distance.

For computing the experimental variogram, an exponential best fit is computed through the mean of each bin. The exponential function model γ_{expo} as a function of distance h can be written as (Webster & Oliver, 2007)

$$\gamma_{expo}(h) = c_0 + c\left\{1 - e^{-\frac{h}{R}}\right\}, \tag{3-2}$$

where c is the maximum dissimilarity (sill), e^x is the exponential function, and R is the range at which dissimilarity stabilizes. The variance at zero distance is denoted by c_0, the nugget variance. The nugget variance consists of the limit of the spatial part of $\gamma_{expo}(h)$ as the variogram approaches zero distance, c_s, and the measurements error c_m. c_s is the small scale variability that is present at smaller than the shortest sampling interval distance. So $c_0 = c_s + c_m$ (Webster & Oliver, 2007; Franklin & Mills, 2007). The resulting function of the best fit (least squares) is the theoretical variogram function. This theoretical variogram is a stochastic model representing the spatial dependence of the measurements, in this case the water vapour measurements. Other functions could also be fitted through the mean in each bin, as long as the are positive definite. For example, spherical and Gaussian functions

can be used. Also linear models, nugget models, or combinations of models can be used. For processing of the GPS water vapour measurements, an exponential model is used. Both the spherical, Gaussian, and exponential model were compared. No combinations of models where examined. The exponential model results in the best fit of the model to the means of the binned dissimilarities. This is based on the correlation coefficient, r, which is obtained from the variance-covariance matrix (Webster & Oliver, 2007)

$$r^2_{\gamma\hat{\gamma}} = \frac{\text{cov}^2_{\gamma\hat{\gamma}}}{\sigma^2_{\hat{\gamma}}\sigma^2_{\gamma}}. \tag{3-3}$$

In this equation $\text{cov}_{\gamma\hat{\gamma}}$ is the covariance between experimental variogram values γ at the bin locations predicted by the variogram model and the estimated variogram values $\hat{\gamma}$ at those locations (the red squares), and σ^2_{γ} and $\sigma^2_{\hat{\gamma}}$ are the variances of γ and $\hat{\gamma}$. The value of r ranges between -1 and 1. If $r > 0$, large values of one variable matches with large values of the other variable. If the large values of one variable matches with small values of the other variable, $r < 0$. Ideally, $r^2_{\gamma\hat{\gamma}}$ should be 1.

For the data of 12:00, as also used in Figure 3-4(b), the variance covariance matrix resulting from the variogram fitting process, is given by,

$$\begin{pmatrix} \sigma^2_{\gamma} & \text{cov}_{\gamma\hat{\gamma}} \\ \text{cov}_{\gamma\hat{\gamma}} & \sigma^2_{\hat{\gamma}} \end{pmatrix} = \begin{pmatrix} 2.88 & 3.42 \\ 3.42 & 4.84 \end{pmatrix}. \tag{3-4}$$

This results in a correlation coefficient $r^2_{\gamma\hat{\gamma}}$ of 0.837. Trying other variogram models, resulted in correlation coefficients lower than 0.7.

3-2-2 Ordinary Kriging Equation and Interpolation

The ordinary Kriging equation system is given by

$$\begin{pmatrix} \gamma(x_1,x_1) & \cdots & \gamma(x_1,x_n) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(x_n,x_1) & \cdots & \gamma(x_n,x_n) & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} w_1 \\ \vdots \\ w_n \\ \mu \end{pmatrix} = \begin{pmatrix} \gamma(x_1,x^*) \\ \vdots \\ \gamma(x_n,x^*) \\ 1 \end{pmatrix} \tag{3-5}$$

In this equation, w_1 to w_n are the weights of the n observations, μ is the unknown assumed constant mean, and $\gamma(x_i, x_j)$ is the dissimilarity (of the fitted variogram) between observation points x_i and the prediction point x_j. $\gamma(x_i, x_j)$ is equal to γ_{ij} in Equation (3-1). The interpolation point is represented by x^*. The unknown constant mean μ is used as a Lagrange multiplier in the minimization of the Kriging error (Webster & Oliver, 2007; Wackernagel, 2003). This equation can be written as

$$\boldsymbol{C_n} \cdot \boldsymbol{w_n} = \boldsymbol{d_n}. \tag{3-6}$$

The matrix $\boldsymbol{C_n}$ is often called the redundancy matrix and $\boldsymbol{d_n}$ the proximity vector. As soon as the variogram is known, the redundancy matrix in Equation (3-5) can be filled. Then for each interpolation point the proximity vector $\boldsymbol{d_n}$ is filled with the distances to the observations. Solving for $\boldsymbol{w_n}$ gives the weights w_1 to w_n of all observations.

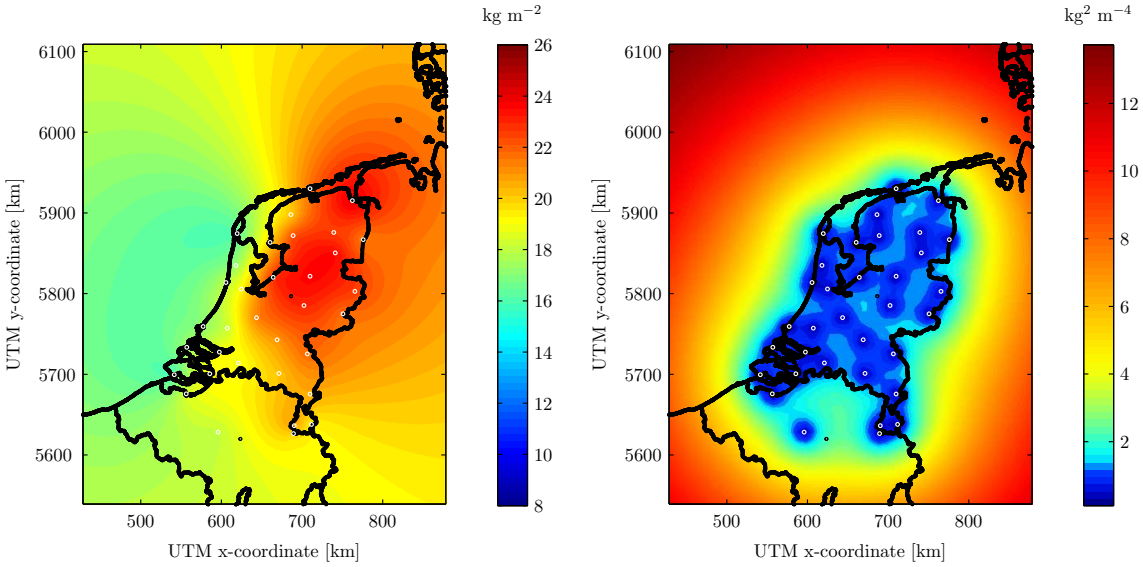


Figure 3-6: (a) IWV map from GPS measurements at 12:00 on 30 May 2010, based on Ordinary Kriging in [kg m^{-2}] and (b) the Kriging variance map for the Kriging interpolation in [kg^2 m^{-4}], the white circles are the GPS ground stations

Using the linear combination

$$\hat{Z}(x^*) = \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix}' \begin{pmatrix} Z(x_1) \\ \vdots \\ Z(x_n) \end{pmatrix}, \tag{3-7}$$

the estimate of the random field $Z(x)$ map at position x^* is retrieved, with $Z(x_1) \dots Z(x_n)$ the IWV values at the GPS ground stations. Figure 3-6(a) visualizes the result of a Kriging interpolation of the IWV measurements in Figure 3-3(a). From this result it is clear Kriging also extrapolates the observations outside the convex hull.

3-2-3 Error Variance and Measurement Error

The Kriging error variance at each interpolation location is computed by (Webster & Oliver, 2007, p.158-159, 180)

$$var\left(\hat{Z}(x^*) - Z(x^*)\right) = \begin{pmatrix} w_1 \\ \vdots \\ w_n \\ \mu \end{pmatrix}' \begin{pmatrix} \gamma(x_1, x^*) \\ \vdots \\ \gamma(x_n, x^*) \\ 1 \end{pmatrix} - c_m. \tag{3-8}$$

Here c_m is the variance of the measurement. Using this equation at each grid point, a map is obtained with for each point the Kriging variance. Figure 3-6(b) shows the Kriging variance for the same input data. As can be seen, the variance at the GPS ground stations is at

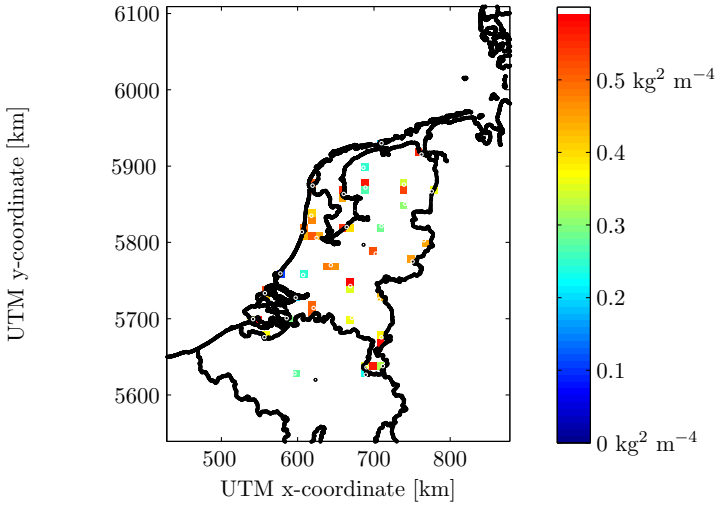


Figure 3-7: Kriging variances near the ground stations at 12:00 on 30 May 2010 in [kg^2 m^{-4}]

a minimum. The further away from the ground stations, the larger the Kriging variance becomes. It may look like from Figure 3-6(b), that the error variances at the stations are zero (besides the measurement error c_m). This is not the case, however, but this illusion is caused by the color range that is used for the figure. For that reason in Figure 3-7 the Kriging variances at the ground stations are plotted.

Say, the hypothetical assumption is made that the measurement error, c_m, is equal to 0 kg^2 m^{-4}. Then, in case an interpolation point would coincide with a ground station, the variance in Equation (3-8) is minimized when all weights for the data points are 0, except for the one at the ground station. The weight for that measurement would then be equal to 1. If we then substitute these weights into Equation (3-7), it is shown that Ordinary Kriging for points is an exact interpolator. However, in the processing here it is assumed that the IWV measurement error equals 1 kg^2 m^{-4}. Then the Kriging interpolation is smoothing the data. (Webster & Oliver, 2007)

3-3 Comparison with other Interpolation Techniques

3-3-1 Comparison of Cubic Interpolation and Kriging

In Figure 3-8 the water vapour maps based on cubic interpolation and Kriging interpolation are compared. The figures of the IWV map based on the cubic interpolation and the Kriging interpolation are repeated in Figure 3-8(a) and 3-8(b) for convenience. Most obvious difference is that the Kriging interpolation in Figure 3-8(b) also extrapolates outside the convex hull of GPS stations, whereas the cubic interpolation in Figure 3-8(a) only interpolates inside the convex hull. Also the IWV values as predicted by the two methods are different. In Figure 3-8(c) the difference between the two interpolation methods is shown. At the ground station locations the difference between the two is almost zero and in between the stations the difference is between -0.8 and 0.7 kg m^{-2}.

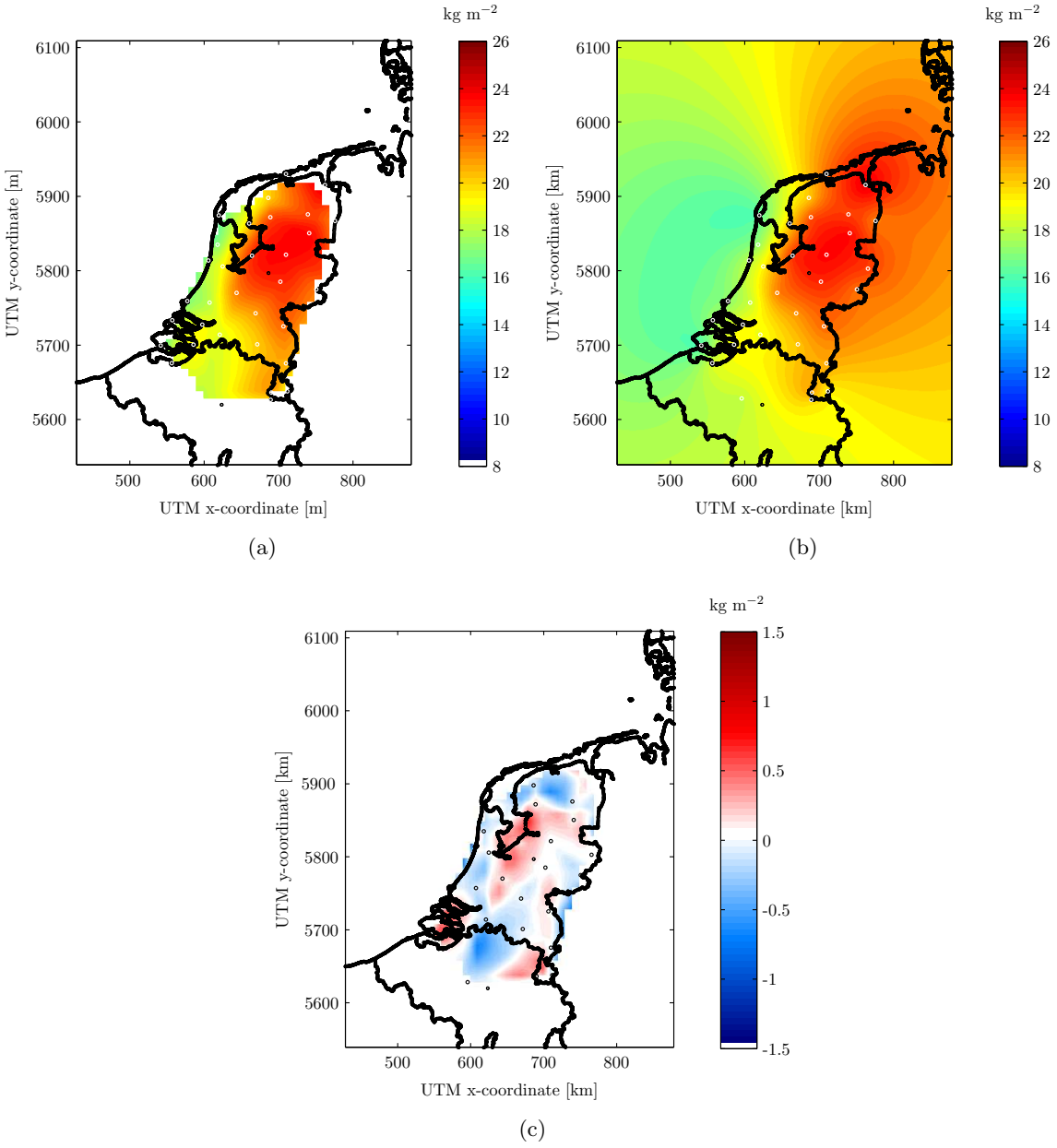


Figure 3-8: (a) IWV map based on cubic interpolation, (b) IWV map based on Kriging interpolation, and (c) Difference of IWV based on cubic interpolation minus Kriging interpolation. All figures use data from GPS measurements at 12:00 on 30 May 2010 and units are in [kg m^{-2}]

From the comparison it can be concluded that both methods are able to interpolate in between the stations. But because the Kriging interpolation can incorporate an error description of the input measurement data, can extrapolate outside the convex hull, and provides a variance map, the Kriging interpolation is preferred over the cubic interpolation.

3-3-2 Optimal Interpolation as used by the KNMI

At the KNMI already an experimental product is available which provides a real time GPS water vapour map. It uses optimal interpolation with the background being the previous IWV field. Optimal interpolation is a least squares method of data assimilation. In data assimilation one creates model initial conditions from observations and a first guess field, in this case the previous IWV measurements. For the optimal interpolation the following formula can be defined:

$$\boldsymbol{x_a} = \boldsymbol{x_b} + \boldsymbol{W}\boldsymbol{d}. \tag{3-9}$$

In this equation $\boldsymbol{x_a}$ is the analysis field of the variable(s) to be predicted, $\boldsymbol{x_b}$ is the background field, $\boldsymbol{W}$ is the gain or weight matrix which is a function of the errors of the observations and the background field, and $\boldsymbol{d}$ is the observational increment (difference between the observations and the background). (Miller, 1984; University, 2007)

Chapter 4

Extending the 2D Interpolation by Advected Measurements

In this chapter the model described in the previous chapter is extended by proposing a method to include advected measurements into the model. The chapter starts with a description of the algorithm and its settings in Section 4-1. The preprocessing steps which are required for the proposed method are described in Section 4-2. Finally, in Section 4-3 the algorithm is applied to input data and the result of each step is shown.

4-1 Algorithm Description and Settings

In this section the 2D interpolation is extending in the vertical. A method is proposed which is summarized in Figure 4-1. The method vertically divides the water vapour measurements at the stations. Radiosonde measurements are used to determine the vertical distribution of water vapour, see Section 4-2-1. Then above each ground station the wind at that height is used to advect the water vapour in that particular layer. Wind data from HIRLAM will be used, see Section 4-2-2. Because the wind is not the same in each layer, the advected measurements in each layer will not be at the same horizontal location. Then for each layer an interpolation is carried out using the current plus the advected measurements. For a description of this interpolation, see Section 4-3. After for each layer a 2D map has been computed, the resulting water vapour maps from each layer are added, which results in the integrated water vapour map.

First, the horizontal domain, maximum height and number of layers must be specified to the algorithm. The Horizontal domain must embody all GPS stations. The position of water vapour and wind data measurements are expressed in degrees north and east. But wind is expressed in [m s^{-1}], so all positions are converted to Universal Transverse Mercator (UTM) format. The Universal Transverse Mercator system (UTM) is used to compute planar coordinates. The UTM system is dividing the world into zones of 6 degrees in longitude. Each zone is a 2D Cartesian coordinate system. A further subdivision of each zone is into bands

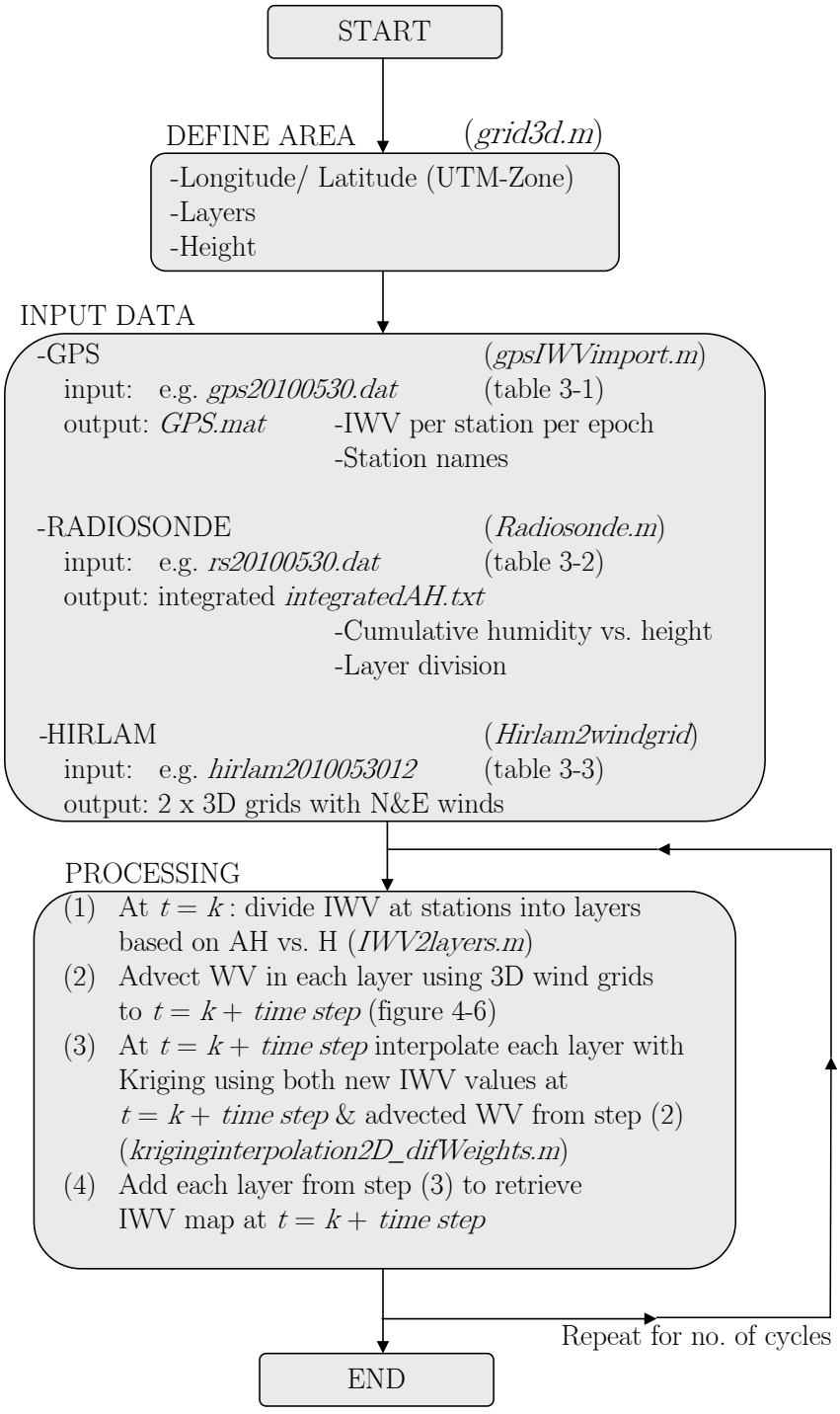


Figure 4-1: Flowchart of the Model to extend the 2D interpolation describing the various steps in the MATLAB script and functions.

Table 4-1: Sample of radiosonde input data. First column gives the data and time, 'wmoid' gives the launch site, 'lon' is the longitude in [°], 'lat' is the latitude in [°], 'pres' is the pressure in [Pa], 'temp' is the temperature in [K], 'dewp' is the dew point temperature in [K], 'wspd' is the windspeed in [m/s], and finally, 'dir' is the wind speed direction in [°] with respect to the north direction, counted clockwise positive.

YYYMMDDHHmm	wmoid	lon	lat	pres	temp	dewp	wspd	dir
201005301118	06260	5.18	52.10	100530.0	286.1	284.9	5.0	286.0
201005301118	06260	5.18	52.10	100000.0	285.5	284.2	2.6	283.0
201005301118	06260	5.18	52.10	95360.0	282.4	281.9	10.1	305.0
201005301118	06260	5.18	52.10	94410.0	281.9	281.4	10.3	305.0

of 8 degrees in latitude. This division is actually referred as the Military Grid Reference System (MGRS). The Transverse Mercator projections can be used between 80 degrees south and 84 degrees north. Outside this latitude band, the Universal Polar Stereographic (UPS) projections are used. Each zone is identified by a number followed by a letter. E.g. the UTM zone used for The Netherlands is *31U*. To make sure coordinates are always positive, the x- and y-coordinates for each zone have an offset. The UTM x-coordinate for the showcase ranges from about $4.3 \cdot 10^5$ m to $8.8 \cdot 10^5$ m and the y-coordinate from about $5.5 \cdot 10^6$ m to $6.1 \cdot 10^6$ m.

The maximum height is fixed at 8 km, because above this height almost no water vapour is present, see Subsection 2-1-3 and Section 3-1. In this thesis evenly distributed vertical layers of 1 km are used. An even distribution is used because of ease of programming and interpretation.

4-2 Preprocessing Data

In this chapter the remaining data inputs are described, which are used during processing of the model of Section 4-3. Subsection 4-2-1 is about the Radiosonde data. In Subsection 4-2-2 the wind data obtained from HIRLAM is described.

4-2-1 Radiosonde Data

Radiosonde data from De Bilt is used for modeling the vertical distribution of the water vapour (see also Subsection 2-3-2). This radiosonde, launched twice per day, takes measurements to about 30–35 km height. A part of the data retrieved with a radiosonde can be seen in Table 4-1. The file which can be read by the MATLAB script (*radiosonde.m*) consists of several columns. The first column is the time epoch in YYYMMDDHHmm, followed by a WMOID (World Meteorologic Organization ID). Radiosondes launched from De Bilt have WMOID '06260'. Beside the WMOID, the file contains the longitude and latitude of the launch site. The most important information however is in the next 5 columns. The pressure and temperature can be used to compute the barometric height using the following formula

(Standard Atmosphere et al., 1976):

$$h = \frac{T_0}{L_0}\left[\frac{P}{P_0}\right]^{-\frac{R_{univ}L_0}{g_0 M}} - \frac{T_0}{L_0} \tag{4-1}$$

In this formula h is the height in [m], T_0 is the standard temperature at sea level in [K], L_0 is the temperature lapse rate of -6.49 K^{-1}, P the pressure at height h in [Pa], P_0 the pressure at sea level in [Pa], R_{univ} is the universal constant in [J kg^{-1} K^{-1}], g_0 is the standard acceleration due to free fall in [m s^{-2}], and M is the molar mass of the earth in [kg $mole^{-1}$]. Actually this equation is only valid to about 11 km where the temperature lapse rate is set to -6.49 K/km. Above this height a slightly different equation should be used, using a temperature lapse rate of zero. But because above this height almost no water vapour is present, this is ignored for ease of programming. The result of this assumption is that the radiosonde appears to reach a lower altitude.

The dew point temperature can be used to compute the relative and absolute humidity. The relative humidity is computed with the MATLAB script *Td2RH.m*. First the saturation vapour pressure e_s is computed with (Haase et al., 2003)

$$e_s = 6.107 \cdot e^{(\frac{17.27 \cdot t}{t+237.3})}, \tag{4-2}$$

for saturating above water, where e is the mathematical constant and t is the temperature in [°C]. The vapour pressure e is computed with t equal to the dew point temperature t_d (Haase et al., 2003):

$$e = 6.107 \cdot e^{\frac{17.27 \cdot t_d}{t_d+237.3})}. \tag{4-3}$$

In this equation e on the left hand side is the vapour pressure and e in the radix at the right hand side is the mathematical constant (≈ 2.718). The ratio between the vapour pressure e and the saturation vapour pressure e_s gives the relative humidity RH in [%]

$$RH = \frac{e}{e_s} \cdot 100. \tag{4-4}$$

The absolute humidity is computed from the vapour pressure e using

$$AH = C \cdot \frac{e}{T}, \tag{4-5}$$

where AH is the absolute humidity in [g m^{-1}], T is the temperature in [K] and C is a constant equal to 216.679 g K J^{-1} (Haase et al., 2003). The wind speed and direction in the last 2 columns are not used, as this information is obtained from HIRLAM (see next subsection).

To summarize, radiosondes measure the pressure, temperature, and the dew point temperature. In Figure 4-2 measurements of a radiosonde launched from De Bilt are shown. The temperature is shown by the black squares and the dew point temperature by the blue asterisks. As the radiosonde rises, the pressure and temperature decrease until it reaches the top of the troposphere at 10–15 km (20 kPa, see also Subsection 2-1-3). Note that the y-axis of the pressure is inverted.

The pressure, temperature, and dew point temperature can be used to calculate the relative and absolute humidity, see Equation (4-2) to (4-4). In Figure 4-3 the absolute humidity is

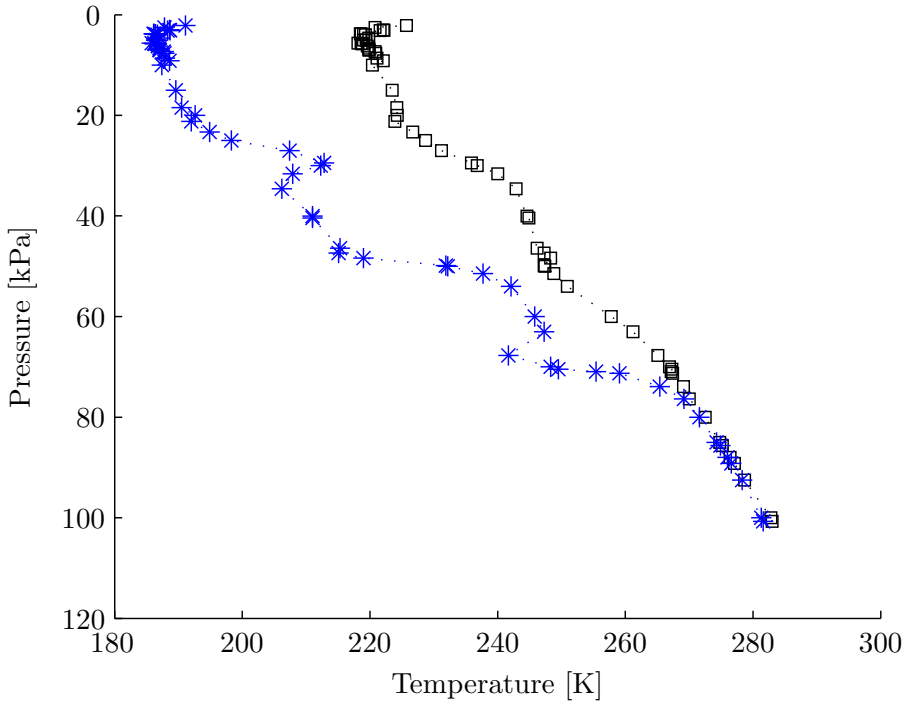


Figure 4-2: Temperature [K] (black squares) and Dew point temperature [K] (blue asterisks) as a function of pressure [Pa] from measurements from a radiosonde launched at De Bilt at 12:00 on 30 May 2010.

shown as a function of height. As can be seen, the absolute humidity decreases rapidly with increasing height. When integrating the absolute humidity over the height, the cumulative humidity as a function of height is obtained. The cumulative humidity as a function of height is visualized in Figure 4-4. Because of the rapid increase of cumulative humidity with height, both a linear (Figure 4-4(a)) and logarithmic scale (Figure 4-4(b)) are used for the horizontal axis.

In Figure 4-5 data from radiosondes launched during the month May 2010 have been visualized. On the lower scale the mean during the month is indicated with the green line with squares. Also on the lower scale are the 'mean minus standard deviation' (blue line with crosses) and 'mean plus standard deviation' (red line with plus signs). The coefficient of variation is given by the standard deviation divided by the mean. On the upper scale the coefficient of variation as a function of height is given by the black line. As can be seen the absolute humidity rapidly decreases. At a height of 8 km, the mean absolute humidity is only 0.04 g m^{-3} with a standard deviation of 0.03 g m^{-3}. For that reason the program uses data up to 8 km height.

Using the cumulative absolute humidity, the (total) integrated water vapour measurements of GPS stations are divided among vertical layers. The water vapour from 0 to 1 km height is assigned to the layer of the lowest 1 km. Water vapour from 1 to 2 km height is assigned to the second layer, and so on. In The Netherlands only 2 radiosondes a day are launched. Therefore the division of the integrated water vapour among the layers based on radiosonde measurements alone, is lagging behind up to 12 hours. Besides, for all GPS stations, the same

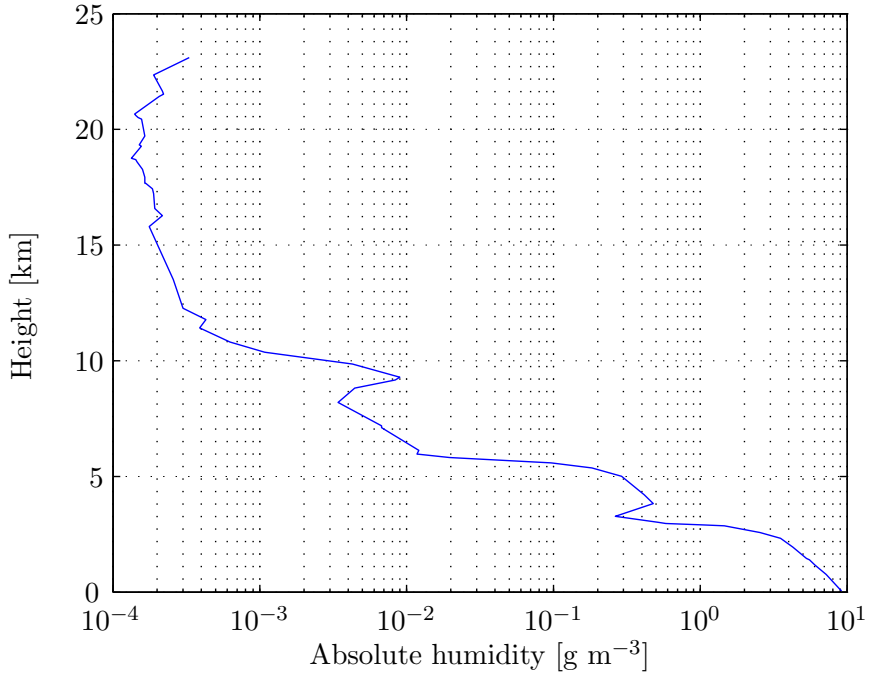


Figure 4-3: Absolute humidity as a function of height computed from measurements from a radiosonde launched at De Bilt at 12:00 on 30 May 2010.

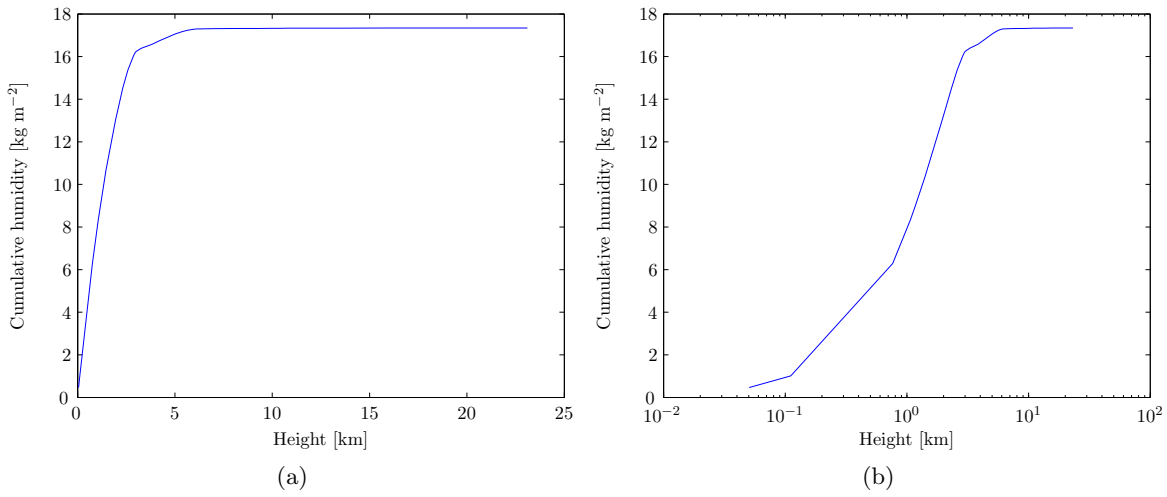


Figure 4-4: Cumulative humidity as a function of height retrieved from a radiosonde launched at De Bilt at 12:00 on 30 May 2010, both a linear scale (a) and a logarithmic scale (b) are used.

division is used. This is because only one radiosonde at a time is launched. This radiosonde however can travel sometimes throughout The Netherlands. This can be seen in Figures 2-10 and 2-11.

Besides dividing integrated water vapour values from GPS measurements among vertical

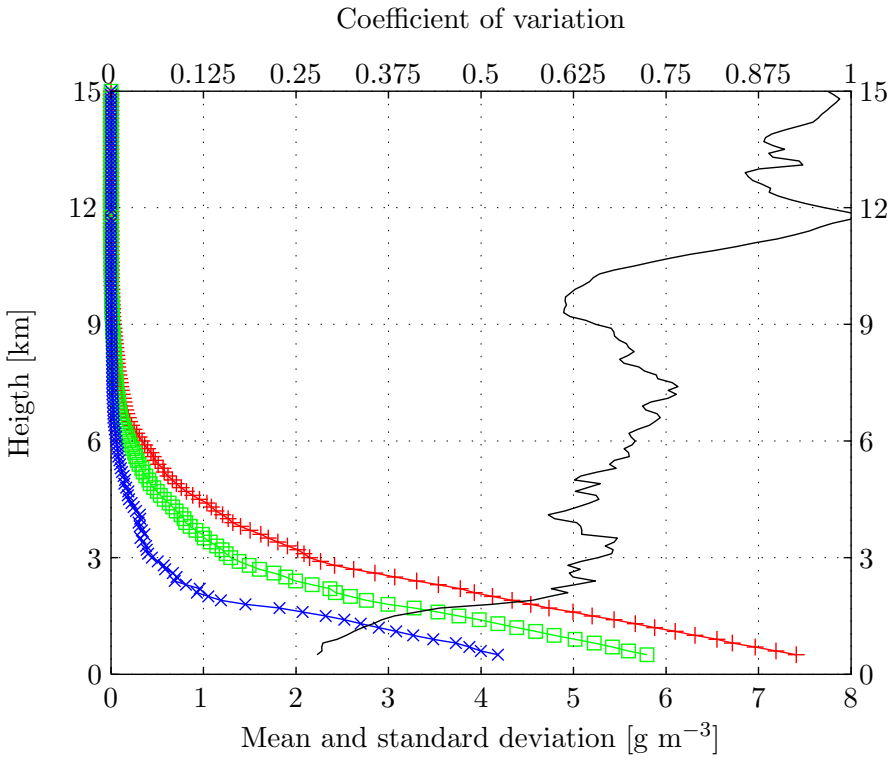


Figure 4-5: Absolute humidity as a function of height retrieved from radiosondes launched from De Bilt, during the month May 2010. On the lower scale: 'mean' (green line with squares), 'mean minus standard deviation' (blue line with crosses), 'mean plus standard deviation' (red line with plus signs); and on the upper scale: coefficient of variation (black line).

layers, and subsequently advect the water vapour in each layer, one could also use only 1 layer. In that case e.g. the wind velocity (speed and direction) at 1 km height could be used to advect the IWV values at each station. Because it is assumed that using multiple layers should give better results, in the processing 8 layers are used. Using only 1 layer however, is also included in the results in Chapter 6-4.

4-2-2 HIRLAM Wind Field

A group of smaller European meteorological institutes use the High Resolution Limited Area Model (HIRLAM) for the short term weather prediction. At the KNMI this model has been setup to cover Europe and part of the Atlantic Ocean. The standard grid size is 55 $\times$ 55 km and predictions are generated up to 48 hours into the future. Smaller grid sizes are also possible, down to 11 $\times$ 11 km. By zooming in to this smaller grid size, smaller disturbances can be predicted, at the cost of a shorter forecasting time down to 12 hours. The operational HIRLAM consists of 60 layers, with 3 in the lowest 200 m. For processing the data, the 11 $\times$ 11 km horizontal grid is used. HIRLAM is fed by measurements of among others wind, air pressure, precipitation, humidity, and temperature. (KNMI, 2010b)

A 3D wind grid is retrieved from HIRLAM, which is imported into the MATLAB program. This wind grid will be used to advect the water vapour. The data consists of 5 columns,

Table 4-2: Sample of HIRLAM wind input data. Longitude and latitude are in [°], height is in [m], and u and v are the wind speed components in north and east direction in [m s^{-1}].

longitude [°]	latitude [°]	height [m]	u [m s^{-1}]	v [m s^{-1}]
2.000	50.000	1000.00	13.17	-3.92
2.000	50.100	1000.00	12.98	-3.69
2.000	50.200	1000.00	12.71	-3.57
2.000	50.300	1000.00	12.36	-3.53

as can be seen in Table 4-2. The first 2 columns define the horizontal position of the grid points by the longitude and latitude in degrees. The third column defines the height of the current grid point in meters. The last 2 columns contain the north-south and east-west speed in meters per second in the current grid cell. A positive u means the east-west component of the wind blows from west to south and a positive v means the north-south component is from south to north.

4-3 Using Advected GPS IWV Measurements for the Current IWV Map

This section describes the steps of the model to use measurements of the past in the current water vapour map.

4-3-1 Divide IWV vertically using Radiosonde Data

The measurements of the radiosonde are used to divide the integrated water vapour measurements of the GPS ground stations among horizontal layers, compare Subsection 4-2-1. Although the radiosonde is not traveling straight up and it only measures variables at the location of the instrument, the vertical profile retrieved from the radiosonde is assumed to be representative for the whole area of interest. This means that far away from the radiosonde path, the vertical distribution of water vapour could potentially be quite different. Radiosondes are launched only twice a day from De Bilt, around midnight and noon. As a consequence the vertical distribution of water vapour, based on these radiosonde measurements, gets outdated in time. However, since there is no other data input available it is still used for the vertical division of water vapour.

Because we divide the IWV measurements into layers, we also have to divide the measurement error variance of 1 kg^2 m^{-4} into layers. So at each location where a GPS ground station is located we have a measurement error variance value for each of the vertical layers. For the values of the variance at each layer we use the same layer division as the one we used for the IWV division, based on the last radiosonde. The higher the altitude of the layer, the less water vapour is present, consequently the value of the variance is lower. According to the error propagation law (Teunissen, 2003) the variance of a finite sum of uncorrelated variables

is computed as the sum of variances of the variables,

$$\sigma^2_{\text{IWV}} = \sum_{i=1}^{\#\text{ of layers}} \sigma^2_{\text{WV}_i}. \tag{4-6}$$

In this equation σ^2_{IWV} is the total variance of 1 kg^2 m^{-4} and $\sigma^2_{\text{WV}_i}$ is the variance in layer i. To compute the variance in layer i the following equation is used

$$\sigma^2_{\text{WV}_i} = \sigma^2_{\text{IWV}} \times \frac{\text{WV}_i}{\text{IWV}}. \tag{4-7}$$

In this equation IWV is the (total) integrated water vapour as measured by the ground station and WV_i is the water vapour in layer i. This is only one option for assigning a variance to each layer. E.g. equal variances for all layers could also be used.

4-3-2 Advect WV in each Layer

Once the IWV above all stations is distributed among the layers, the wind speed and direction from the HIRLAM data file is used to advect the water vapour to a new location. It is assumed that the water vapour in the air moves with the wind velocity as given by the HIRLAM wind field. Because the wind speed and direction is not the same at different heights, the predicted position of the initial water vapour measurement shifted by the wind, varies with height. This is the reason that predicted IWV values cannot be computed by simple vertical addition of water vapour values. To solve this, for each horizontal layer an interpolation is done, again using Kriging.

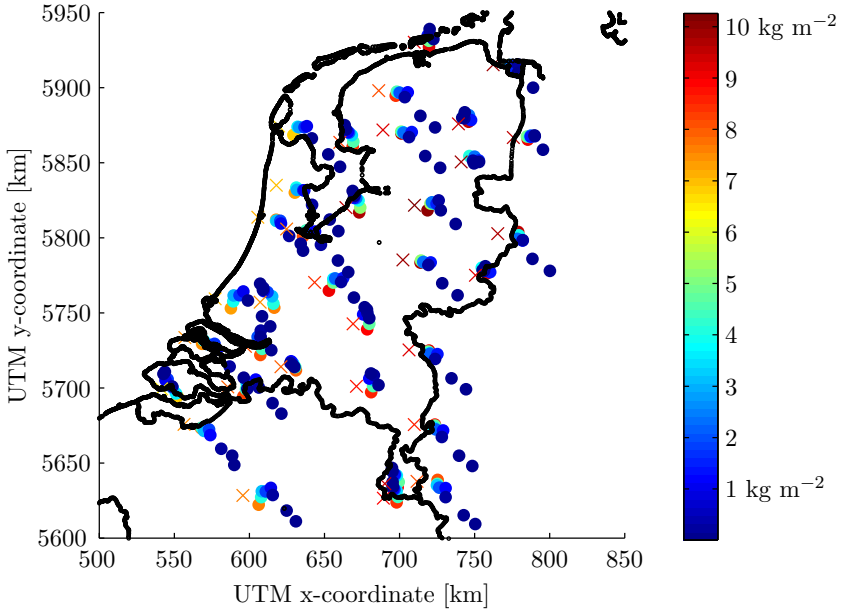


Figure 4-6: IWV measurements at 12:14 at the GPS ground stations (crosses) and WV predictions for eight vertical layers near the stations (dots). Color of the cross indicates the WV in the first layer and the color of the dots the WV in the first to eighth layer.

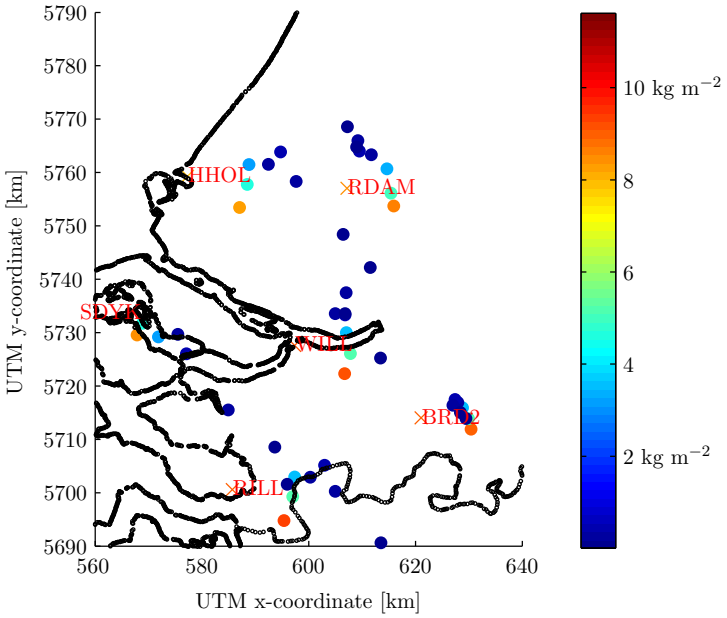


Figure 4-7: Zoomed version of Figure 4-6, focusing on the southwest of The Netherlands. Both IWV measurements at 12:14 at the GPS ground stations (crosses) and WV predictions for eight vertical layers near the stations (dots) are shown.

In Figure 4-6 both the IWV measurements at 12:14 (crosses) and WV predictions of 12:00 advected to the same time are shown (dots). The term WV is used to indicate the total water vapour at a position in only one layer, in contrary to IWV which is the water vapour at a position integrated over all layers. If we zoom in on the southwest of The Netherlands with the same data as in Figure 4-6, we get Figure 4-7. This figure clearly shows that the horizontal position of the predictions in the different vertical layers varies with height. For example, the location of the advected WV measurements from the station in Breda (BRD2) for higher altitude are close to each other, while the ones from Hoek van Holland (HHOL) and Willemstad (WILL) are more shifted for different layers. In Figure 4-8 the same data is used, but only the lowest vertical layer is shown. At the position of the GPS ground stations both the measurements at 12:00 (large dots), as well as the new measurements at 12:14 (small dots inside the large dots) are shown. To the East of the stations, the advected measurements are shown (also large dots). Remember that the WV at the location of the GPS ground station is the WV part in that vertical layer, based on the last known radiosonde measurement.

4-3-3 Interpolating Current and Advected WV Using Kriging

The next step in the model is to interpolate both the new WV values and the advected water vapour measurements in each layer. In Figure 4-9 the data in Figure 4-8 of both the measurements at 12:14 and the measurements of 12:00 advected to 12:14 are interpolated using Kriging. The determination of the variogram model is computed for each time epoch and for all layers again and again. By doing so, changes in weather circumstances could be better incorporated in the Kriging interpolation.

The title above Figure 4-9 starts with 12:14[15]. This means that the water vapour map is

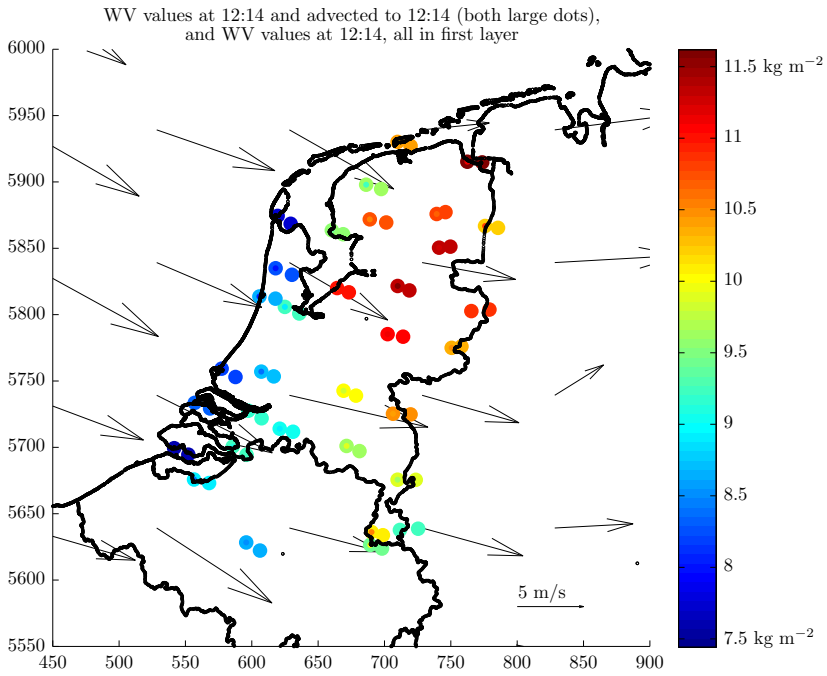


Figure 4-8: Original WV estimates at 12:00 and those estimates advected to 12:14 (large dots) and new WV estimates at 12:14 (small dots), all concerning the first layer. The arrows indicate the wind speed and direction at 1 km height

visualizing the water vapour map at 12:14. [15] means that for the interpolation, data is used from both current measurements as well as from 15 minutes ago. Further on in the report, e.g. 12:29[60] means that the map is visualizing (integrated) water vapour at 12:29, and the interpolation is based on data from 12:29 and from advected data from 15 minutes up to 60 minutes ago.

For the interpolation in Figure 4-9 the same weights are used for both the advected data as well as for the new data. As already mentioned, it is assumed that the GPS IWV measurements have an error variance of 1 $kg^2\ m^{-4}$ (nugget effect). This value is the total variance of the IWV value as measured by the ground station. In the figure the wind direction at 1 km height is shown by the black arrows. So the main direction of the wind is from west to east.

The wind which is used to advect the measurements at each layer, is located at the top of each layer. This could introduce an error in the position of the advected measurements. It is recommended to investigate what the optimal layer division and wind model are, that is, which choice would result in the best water vapour map (see also Section 7-3). In Figure 4-9 the Kriging variance (Equation (3-8)), is shown by the white contour lines, representing constant Kriging variances of 0.17 and 0.5 $kg^2\ m^{-4}$. On the east side of each station a slightly outstretched region is present, caused by the advected measurements due to the main wind direction from the west. For this and following figures, a mask has been used for which no data is shown on the maps. This mask prevents visualization of data which corresponds to high Kriging variances.

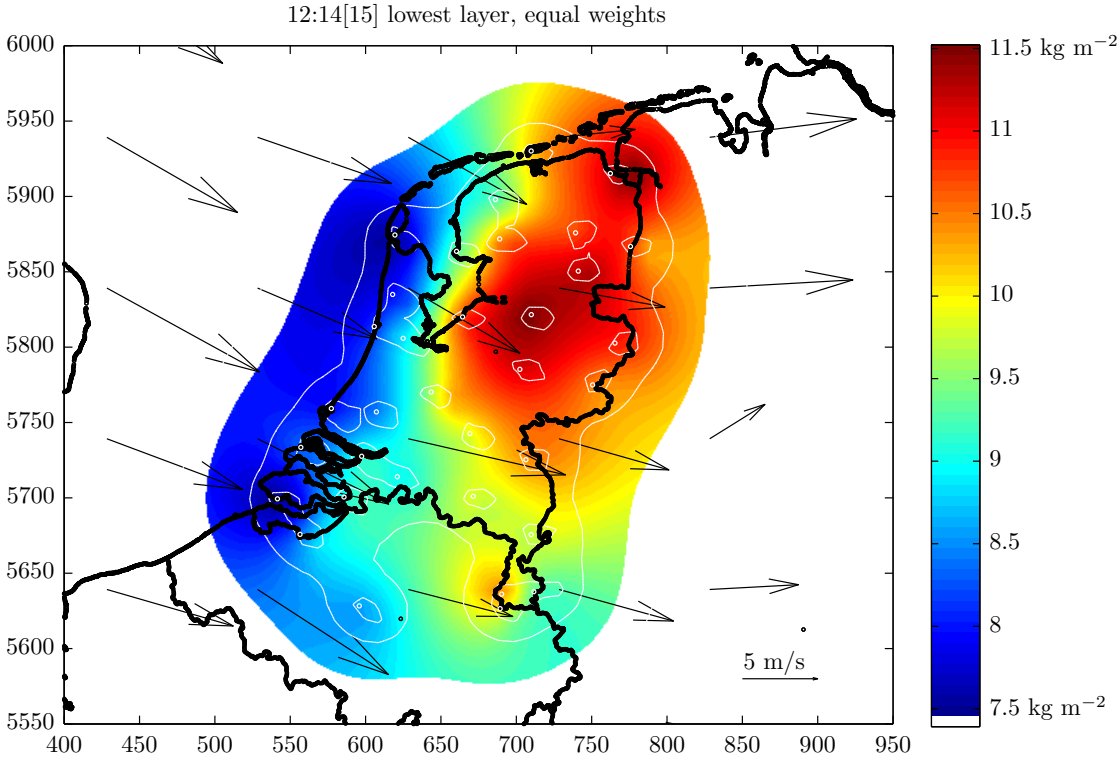


Figure 4-9: Water vapour map at 12:14 in [kg m^{-2}] and Kriging variances (white lines at 0.17 and 0.5 kg^2 m^{-4}) for the lowest vertical layer (between 0 and 1 km height), using both the measurements at 12:14 and the advected measurements from 12:00, with equal weight. The black arrows indicate the wind direction at 1 km height at 12:00.

4-4 Impact of Advected Measurements

Figure 4-10 the impact of using the advected measurements of 12:00 on the prediction at 12:14 is visualized. First the same data of Figure 3-6(a) is used in Figure 4-10(a). The only difference is the mask which has been applied to the data. In Figure 4-10(b) the total IWV for 12:14 is visualized. So in this figure the interpolated water vapour in each layer is added vertically. The ingredients for this figure are the current measurement at 12:14, and the measurements of 12:00 advected by the wind. When this figure is compared to Figure 4-10(a) where no advected measurements were used, we get a IWV difference map. This map is visualized in Figure 4-10(c). It can be clearly seen that in between the ground stations, positive as well as negative differences are present. Above the North Sea, west of The Netherlands, a large positive (red) area is present, whereas to the east of The Netherlands negative and positive differences (red and blue) areas alternate. This is caused by the advected measurements which are all advected by the wind to the east.

From these figures it can be concluded that adding advected measurements adds extra information to the IWV map. Whether this extra information is also valid will be explained later in Chapter 6. In the next chapter a method will be described which down weights advected

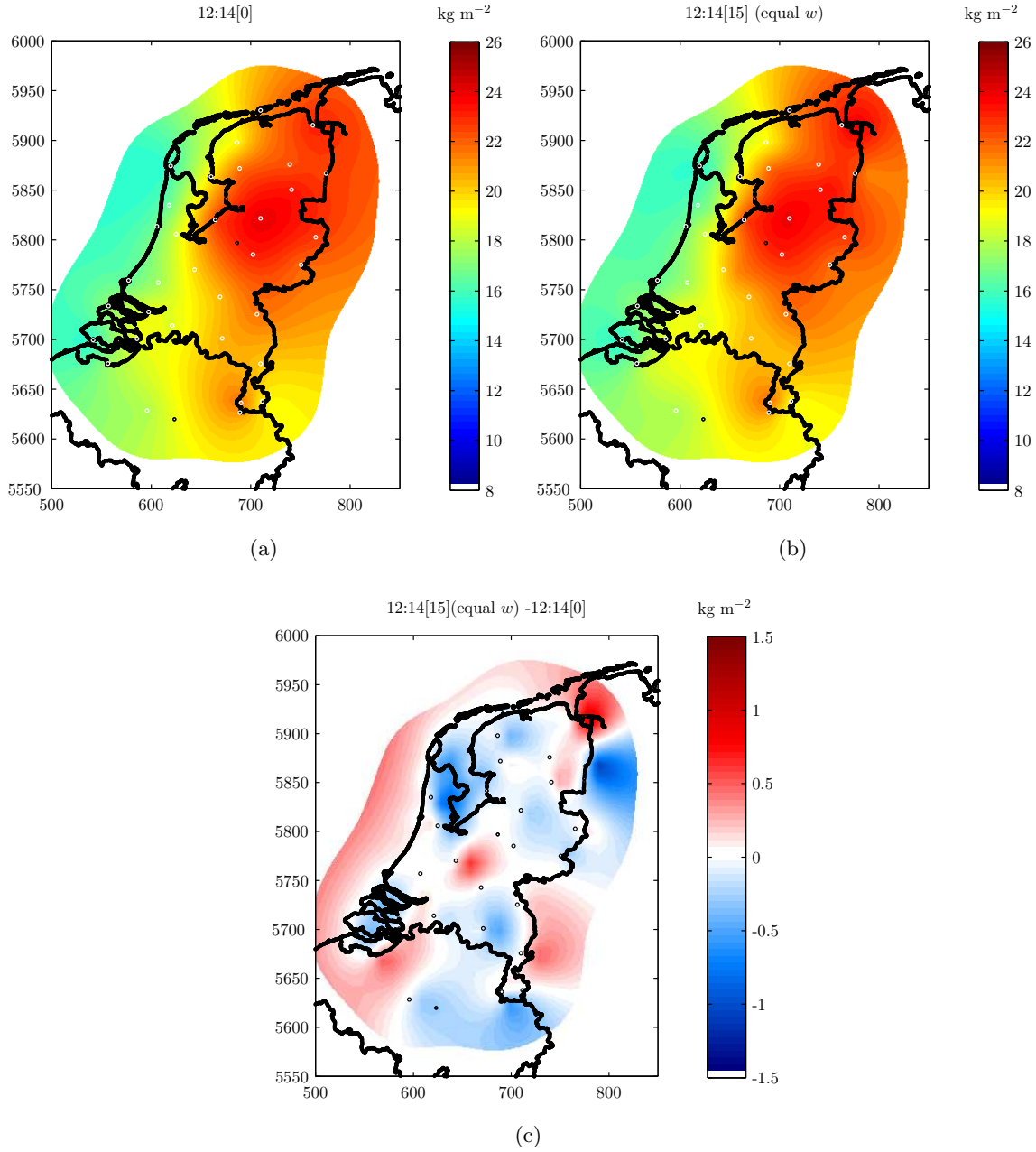


Figure 4-10: (a) IWV map at 12:14 using only current measurements of 12:14, (b) IWV map at 12:14 using both the current measurements of 12:14 and the advected measurements from 12:00, with equal weights, and (c) Differences of IWV at 12:14 between using both the advected and current measurements and using only current measurements. All units are in [kg m^{-2}]

measurements which are from a longer time ago.

Chapter 5

Down Weighting of Advected IWV Measurements

Because in general advected estimates of WV are less precise than new measurements, we should use higher weights for new measurements than for advected estimates. Also, advected estimates should become less precise if advected further into the future. Therefore for the Kriging interpolation lower weights should be assigned to predictions of advected WV values further into the future. Because there is no explicit formula known for downgrading Kriging weights as a function of time used for 2D spatial interpolation, it is assumed here that the inaccuracy resulting from extrapolating in time can be used to give an estimate for the downgraded weights given to advected measurements. See Chapter 7 for a more extensive discussion. Note that there is extensive literature available on using space-time models. These models use time correlation between successive time epochs for spatial interpolation (Kyriakidis & Journel, 1999). This is however beyond the scope of this thesis. So, in the model used in this thesis, interpolation in time and space is treated separately.

In this section it is discussed how the Kriging weights are modified. It starts with modifying the Kriging interpolation relation in Section 5-1. Then in Section 5-2, the IWV measurements at the ground stations are extrapolated in time using Kriging. Next, new Kriging weights are generated for the (I)WV interpolation in space in Section 5-3. Finally, the results are shown of using the new Kriging weights for interpolating in space in Section 5-4.

5-1 Modifying Kriging Interpolation Relation

In this section a modified Kriging interpolation equation will be derived. For the derivation both the current plus the advected measurements of 15 and 30 minutes ago will be used. For modifying the Kriging interpolation we reconsider Equation (3-7). In this equation the weights w_1 to w_n are all obtained from measurements of the function Z, in this case the IWV, at the same time. As the advected values are from the past, we have less confidence in these values. Therefore the weights are modified according to the timespan by which the

water vapour values are advected. So current measurements will have a higher weight than measurements of 15 minutes ago, and these measurements of 15 minutes ago will have a higher weight than 30 minutes ago. For this Equation (3-7) is slightly modified into

$$\hat{Z}(x^*) = \frac{1}{\lambda(t,s)} \cdot \begin{pmatrix} f^0 \cdot w_1^0 \\ \vdots \\ f^0 \cdot w_n^0 \\ f^{15} \cdot w_1^{15} \\ \vdots \\ f^{15} \cdot w_n^{15} \\ f^{30} \cdot w_1^{30} \\ \vdots \\ f^{30} \cdot w_n^{30} \end{pmatrix}' \begin{pmatrix} Z(x_1^0) \\ \vdots \\ Z(x_n^0) \\ Z(x_1^{15}) \\ \vdots \\ Z(x_n^{15}) \\ Z(x_1^{30}) \\ \vdots \\ Z(x_n^{30}) \end{pmatrix}. \tag{5-1}$$

In this equation, n is the number of measurement stations used. The fraction $1/\lambda(t,s)$ and the factor f^t on each row in the first vector on the right side are equal to 1 in this chapter. Both will be defined later on in Section 5-3. For example, w_1^{15} indicates the weight given to station number 1 from 15 minutes ago, which has a value of $Z(x_1^{15})$. This equation can easily be modified according to which measurements in time are used for the current estimation.

To determine the new fraction and factor in Equation (5-1) we use IWV Kriging extrapolating in time on the GPS water vapour time series at the three ground stations which were already shown in Figure 3-2. The Kriging variances for those predictions are used to obtain new Kriging weights.

5-2 IWV Kriging Extrapolation in Time

The script of the Kriging interpolation of the water vapour map has been modified to handle IWV measurements in time. This script can also be used to extrapolate IWV values into the future. In Figure 5-1(a) the experimental dissimilarity cloud for the GPS IWV measurements of Figure 3-2(a) is visualized. The dissimilarities are grouped into 50 bins. In Figure 5-1(b) the mean in each bin is plotted (red squares), together with a best fitted bilinear model. A bilinear model is used, because it gives a correlation coefficient (see Equation (3-3)) of 0.98 versus lower than 0.8 for other variogram models. Higher weight is given to measurements close by in time, using a Cressie weighting function (Webster & Oliver, 2007). Actually, only the beginning of Figure 5-1(b)is important, because measurements are advected for a limited amount of time.

This variogram model is used to interpolate and extrapolate the IWV measurements of Figure 3-2 in time. The result is shown in Figure 5-2(a). The black dots are the IWV measurements at the stations. The blue line is the Kriged IWV prediction. The IWV predictions could be further extrapolated. But as can be seen in Figure 5-2(b), where the Kriging variance versus time is plotted, the value of the accuracy rapidly increases for predictions further into the future. Note that in this figure the variance of the measurement error, c_m, of 1 kg^2 m^{-4} is

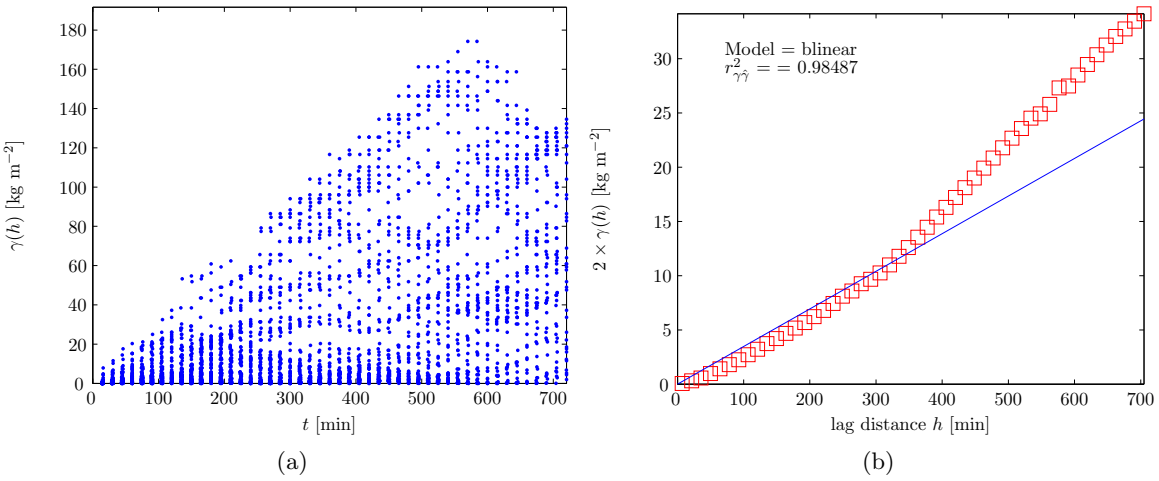


Figure 5-1: (a) Experimental dissimilarity cloud for a time series of IWV data of Eindhoven at 30 May, 2010 grouped into 50 bins and (b) Theoretical (semi-)variogram based on a bilinear model fit through the means of the dissimilarities in each bin in Figure 5-1(a).

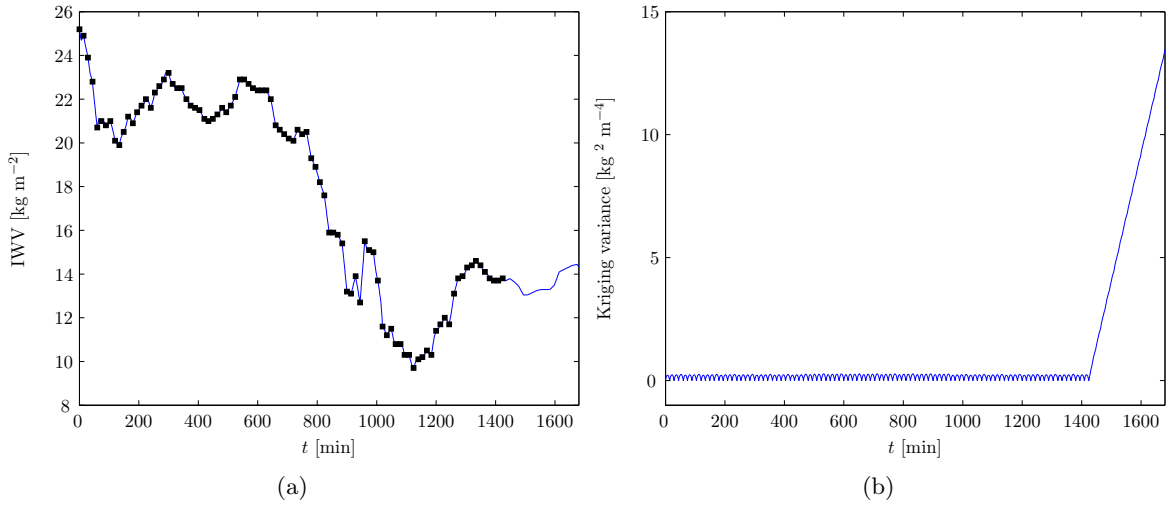


Figure 5-2: (a) IWV measurements of Eindhoven at 30 May, 2010 (black dots) and predictions in between measurements and into the future (blue line) and (b) Kriging variance belonging to the IWV predictions of Figure 5-2(a).

already subtracted, see Equation (3-8). The last known IWV measurement in Figure 5-2(a) is from $t = 1440$. In Figure 5-2(b) this is the last epoch at which the variance is small. The variance from the last known measurement into the future at each station is a linear function (the straight line from 1440 s in Figure 5-2(b)). For each station m, the variance $(\sigma_m^t)^2$, for future epochs can be calculated by

$$(\sigma_m^t)^2 = (\frac{\Delta\sigma_m^2}{15\text{min}}) \cdot 15t + (\sigma_m^{0\text{min}})^2 \qquad \text{for } t > 0, \tag{5-2}$$

where $t = 0$ corresponds to the time of the last measurement. In Table 5-1 the gradient of the linear function, $(\frac{\Delta\sigma_m^2}{15\text{min}})$, and the y-intercept, $(\sigma_{0\text{min}}^m)^2$, are shown for three stations. The

Table 5-1: Increase of Kriging Variances in [kg^2 m^{-4}] of IWV predictions using Kriging extrapolation on a time series of GPS ground station measurements, these values include the measurement error of 1 kg^2 m^{-4}.

	$(\sigma_m^{0\text{min}})^2$	$(\frac{\delta\sigma_m^2}{15\text{min}})$
Eindhoven	1.06	1.17
IJmuiden	1.05	0.974
Stavoren	1.05	0.969
Mean	1.06	1.04

values in the table include the variance of the measurement error. The last row shows the mean of the variance increase of these three stations.

Obviously, the variances increase rapidly with time. In Figure 3-7 it can be seen that the spatial component of the Kriging variance (c_s) around a ground station is not equal for all stations (see Figure 3-8). So the variance as a function of distance in space is not equal for different stations either. For example, if the water vapour in time is slowly varying, the variance in time will be different than in the case the water vapour is varying much. This is also true in case of one dimensional Kriging in time. So the variance as a function of time is also not equal for all stations. For further processing, the mean of the variances at these three stations is used. For a further development of the program, at each station IWV measurements at that particular ground station could be used for the determination of the variances, which could then better represent circumstances and weather conditions at that station. Other ways for downgrading advected measurements could also be investigated.

5-3 Determine new Weight Factors from Kriging Variances in Time

In this section the variances which are obtained from the Kriging interpolation in time in the previous section are used to down weight the Kriging weights for interpolation in space. This will be done by defining equations for the factor f^t and $\lambda(t,s)$ in Equation (5-1), which is repeated here for convenience:

$$\hat{Z}(x^*) = \frac{1}{\lambda(t,s)} \cdot \begin{pmatrix} f^0 \cdot w_1^0 \\ \vdots \\ f^0 \cdot w_n^0 \\ f^{15} \cdot w_1^{15} \\ \vdots \\ f^{15} \cdot w_n^{15} \\ f^{30} \cdot w_1^{30} \\ \vdots \\ f^{30} \cdot w_n^{30} \end{pmatrix}' \begin{pmatrix} Z(x_1^0) \\ \vdots \\ Z(x_n^0) \\ Z(x_1^{15}) \\ \vdots \\ Z(x_n^{15}) \\ Z(x_1^{30}) \\ \vdots \\ Z(x_n^{30}) \end{pmatrix}. \tag{5-3}$$

From Equation (5-2) and the values in Table 5-1 the prediction variance $(\sigma_m^t)^2$ for station m at time t can be calculated. Using this variance the factor f_m^t is defined as

$$f_m^t = \frac{1}{(\sigma_m^t)^2} \sum_{i=1}^{n} (\sigma_i^t)^2, \tag{5-4}$$

where $\sum_{i=1}^{n}(\sigma_i^t)^2$ is the sum of variances for all stations i and advected locations used in the interpolation with t minutes time difference. The time dependency in this formula is in the denominator, as well as in the summation. Note that the mean of the stations in Table 5-1 was used for further processing. This results in the same factor f_m^t for equal t which is denoted as f^t.

As an example assume 2 IWV measurements from the past are used for the current IWV map. E.g., at 11:29 an IWV map is computed from GPS IWV measurements from 11:29, 11:14, and from 11:00. Then the factors by which the Kriging interpolation weights are multiplied, are given by f_m^t in Table 5-2.

Table 5-2: Prediction variance factors f_m^t and relative weight factors $k(t, \delta t)$ in case of 2 epochs in the past are used for the current IWV map.

	0 minutes	15 minutes	30 minutes
f_m^t	4.930	3.017	2.147
$k(t, \delta t)$	0.488	0.299	0.213

To ensure all new weights '$f_m^t \cdot w_m^t$' sum up to 1, the factor $\lambda(t, s)$ is introduced which divides the new weights by their sum. So this factor $\lambda(t, s)$ is computed by

$$\lambda(t, s) = \sum_{t=0}^{30} \sum_{i=1}^{n} w_i^t f_i^t \tag{5-5}$$

Because for all stations the same downgrading is assumed, f_m^t is the same for each station m, and is denoted by f^t. Inserting f^t and $\lambda(t, s)$ into Equation (5-3) plus further simplifying results in the following Kriging estimation with downgraded weights for the advected measurements

$$\hat{Z}(x^*) = \frac{1}{\sum_{t=0}^{30} f^t \sum_{i=1}^{n} w_i^t} \cdot \begin{pmatrix} f^0 \cdot w_1^0 \\ \vdots \\ f^0 \cdot w_n^0 \\ f^{15} \cdot w_1^{15} \\ \vdots \\ f^{15} \cdot w_n^{15} \\ f^{30} \cdot w_1^{30} \\ \vdots \\ f^{30} \cdot w_n^{30} \end{pmatrix}' \begin{pmatrix} Z(x_1^0) \\ \vdots \\ Z(x_n^0) \\ Z(x_1^{15}) \\ \vdots \\ Z(x_n^{15}) \\ Z(x_1^{30}) \\ \vdots \\ Z(x_n^{30}) \end{pmatrix}. \tag{5-6}$$

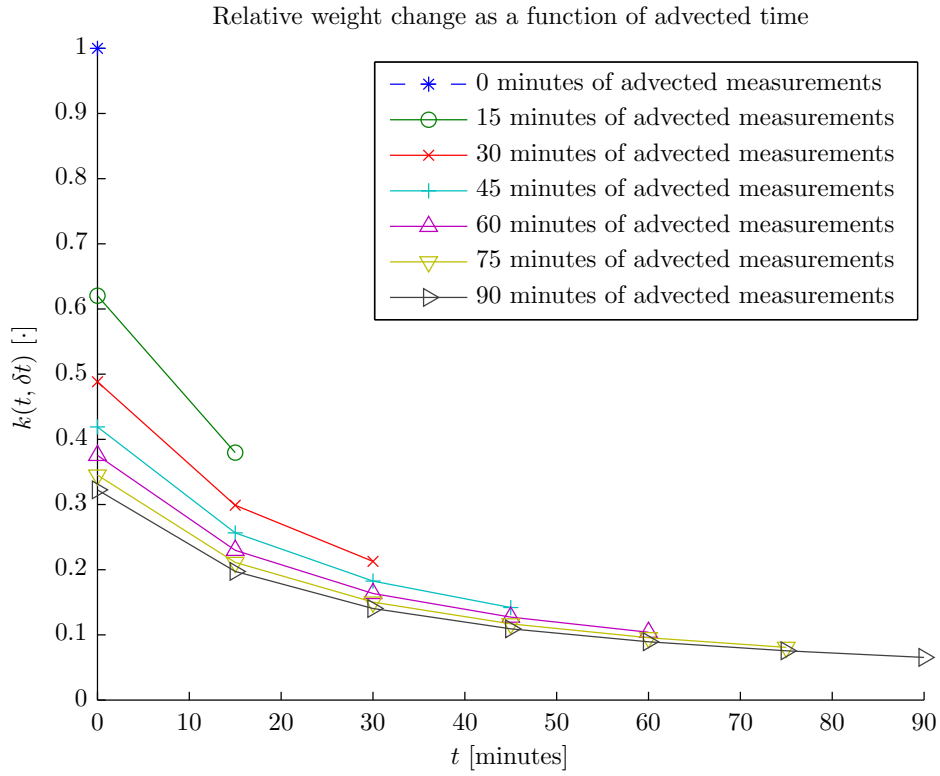
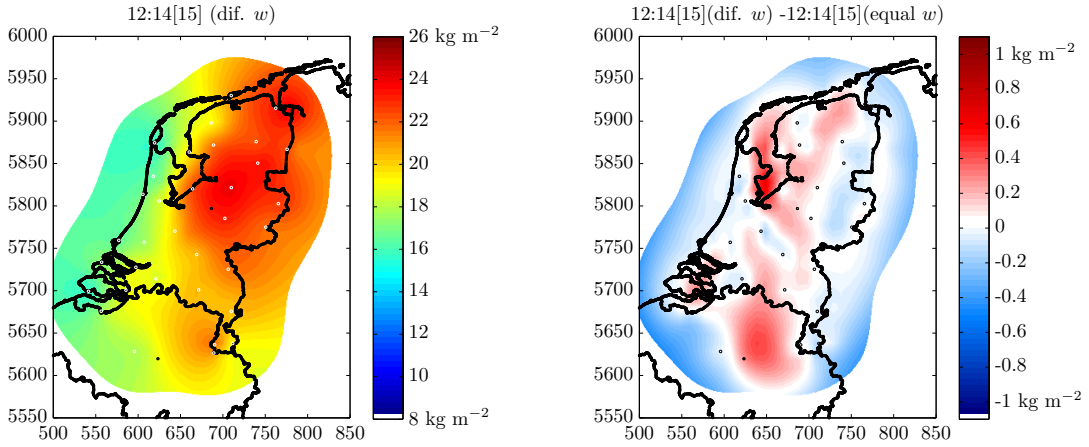


Figure 5-3: Relative weight change $k(\delta t)$ as a function of advected time shown for 0 up to 90 minutes of advected measurements.

To give an impression of the downgrading of advected measurements, $f^t / \sum_{t=0}^{30} f^t$ is denoted as the 'relative weight change' $k(t, \delta t)$ and is given in Table 5-2. δt is the total number of advected measurements and t is the time difference of the specific measurement. So, if 2 epochs from the past are used for the IWV map, the weights for the current measurements as computed by solving the original Ordinary Kriging Equation, are downgraded by a factor of 0.488.

To give an impression of the rate at which the weight changes, when using more advected measurements from the past, Figure 5-3 shows the 'relative weight changes' $k(t, \delta t)$ for 0 to 90 minutes of advected measurements. It is clear that using only current measurements for the IWV interpolation, the weights for these measurements are not modified. When using measurements of up to 90 minutes from the past the weights of the current measurements are decreased by a factor of 0.32.

The above mentioned equations are incorporated in the MATLAB script *kriging_difWeights.m*. In Figure 5-4(a) the result is shown for the Kriging interpolation at 12:14 by using current measurements at 12:14 and advected measurements of 12:00. Because this figure is again very similar to the one using the same weights (Figure 4-10(b)) the difference is shown in Figure 5-4(b). Again it is clear that in between the stations, the original IWV map is changed. Outside the convex hull of the ground stations, the difference is almost everywhere negative (blue). Apparently, the Kriged mean obtained using different weights is lower than in the case the weights are not downgraded.



(a) (b)

Figure 5-4: (a) IWV map at 12:14 in [kg m^{-2}] using both the current measurements of 12:14 and the advected measurements from 12:00, with different weights and (b) Differences of IWV at 12:14 in [kg m^{-2}] between using different weights (Figure 5-4(a)) and using equal weights for the advected and current measurements (Figure 4-10(b)).

This can be explained as follows, using lower weights for advected measurements, causes the IWV map to have lower values outside the convex hull. Inside the convex hull of the ground stations, both positive and negative differences are caused by using lower weights for advected measurements.

In Figure 5-5 the difference between Figure 5-4(a) is compared with Figure 3-6(a) where only the current measurements are used. The pattern in Figure 5-5 is clearly different than in Figure 4-10(c), which was also already predicted of course by Figure 5-4(b). So, at least in this particular data set, using different weights for advected and current measurements has its influence on the interpolated IWV map. In between the stations, again positive and negative differences are present. Again, Figure 5-5 shows a large positive area above the North Sea, and an alternating area to the east of the Netherlands. As already said, this is caused by the wind which is mainly from west to east.

5-4 Results from Using Different Weights

Up to now, all figures in this section use the current measurements plus one advected epoch from the past for the IWV map. The MATLAB script *grid3d.m* has been modified to use advected measurements from further back in the past. In Figure 5-6(a) the result is shown for the IWV map at 12:14 using both the current measurements at 12:14, as well as the advected measurements from 12:00 and 11:44. Compared with Figure 5-5 the pattern is slightly different and the absolute values of the differences are higher: darker blue and red areas are present in Figure 5-6(b).

To show how an interpolation is obtained from nearby measurements, Figure 5-7 shows nearby weights for an interpolation location south of Utrecht (black circle). Closest stations are Utrecht (UTR2), Lithoijen (LITH), Eindhoven (EHVN), Breda (BRD2), and Rotterdam

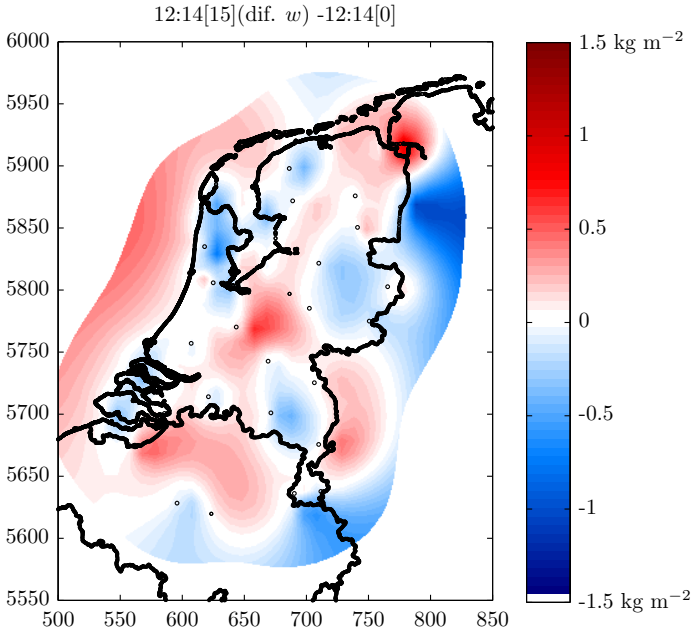
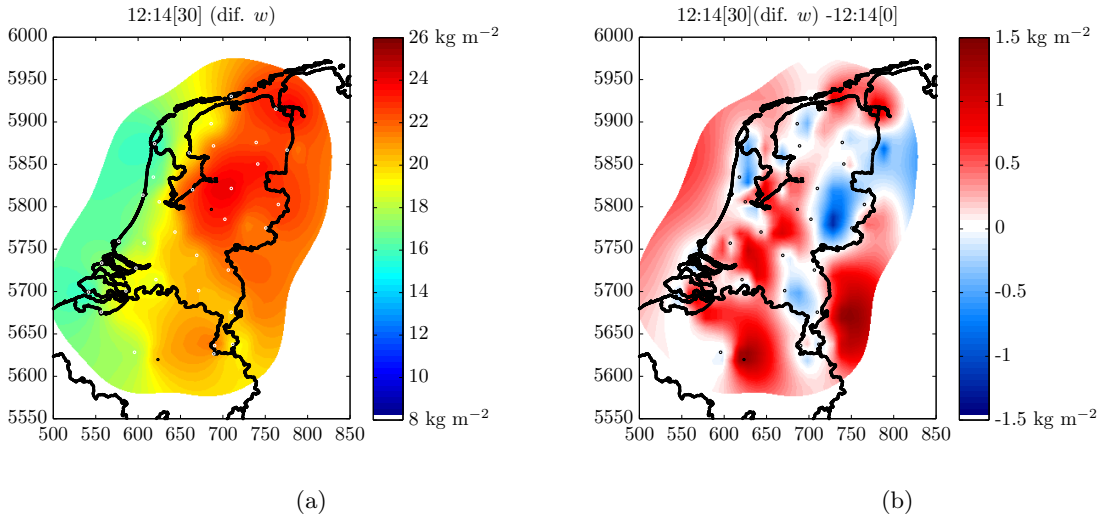


Figure 5-5: Difference of IWV at 12:14 in [kg m^{-2}] between using different weights for the advected and current measurements (Figure 5-4(a)), and using only current measurements (Figure 3-6(a)).



(a) (b)

Figure 5-6: (a) IWV map at 12:14 in [kg m^{-2}] using both the current measurements of 12:14 and the advected measurements from 12:00 and 11:44, with different weights and (b) Differences of IWV at 12:14 in [kg m^{-2}] between using different weights (Figure 5-6(a)) and using only current measurements (Figure 3-6(a)).

(RDAM). The station locations are marked with a black cross. At each station, the measurement of 12:14 are shown. Also to the east of each station, the advected measurements of 12:00 and 11:44 are shown. As can be seen, the advected WV values near each station slightly

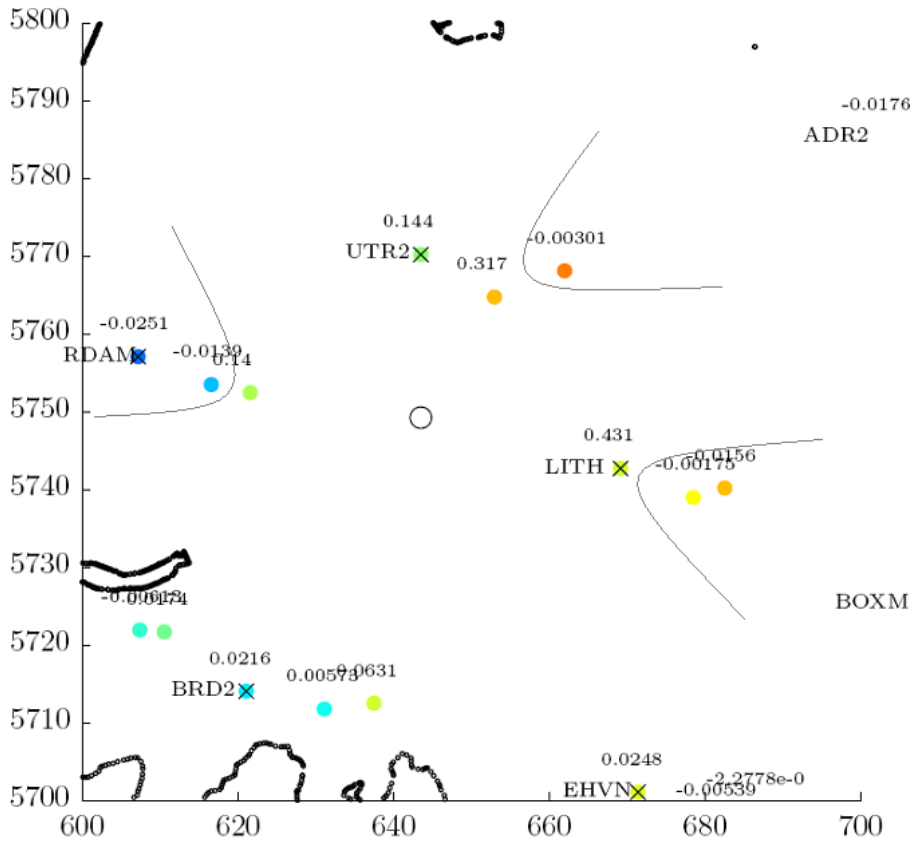


Figure 5-7: Map with weights for data points for a WV prediction in the lowest layer, south of Utrecht ground stations, prediction is performed at the black circle, a sketch of the screening effect is visualized by the black curved lines.

differ. Also the direction of the wind is slightly changed, because the three measurements are not on a straight line. This is caused by the HIRLAM wind data, which updates every three hours (KNMI, 2010b). For the interpolation location chosen, the highest weights are listed in Table 5-3.

Table 5-3: Kriging weights for a data point south of Utrecht.

Weight	Station	Time epoch measurement
0.431	LITH	12:14
0.317	UTR2	12:00
0.144	UTR2	12:14
0.140	RDAM	11:44
0.063	BRD2	11:44
0.025	EHVN	12:14

These weights add up to more than 1. Some weights of data points further away are negative, causing the sum of all weights to add up to 1 again. In this figure one can clearly see the screening effect of Kriging. The screening effect causes data points, which are 'in the

shadow' of other data points with respect to the estimation location, to get real small weights (Wackernagel, 2003). In Figure 5-7 this screening effect is visualized by the black curved lines near three data points. For example, the advected measurements east of LITH are assigned weights of -0.000175 and -0.0156. This is really low compared to the measurement of at station LITH, which has a weight of 0.431. So in this case the position of the measurements at station LITH causes that the advected measurements have almost no influence on the interpolation point.

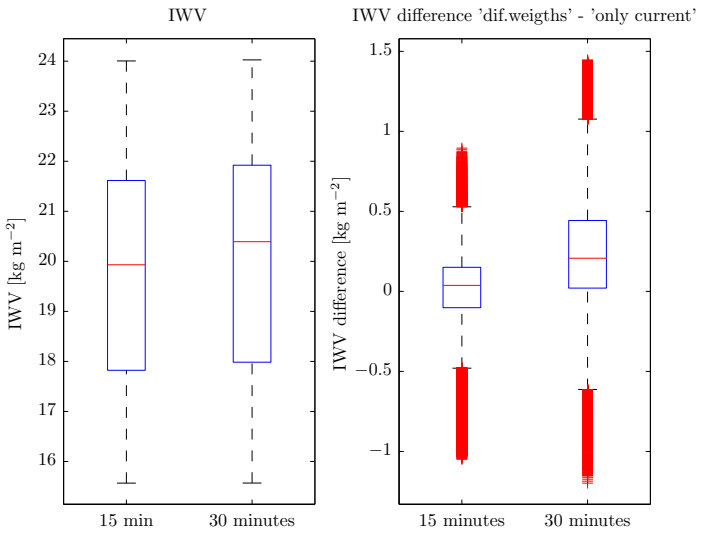


Figure 5-8: Boxplots showing the differences in IWV prediction when using current and advected measurements of 15 minutes ago on the one hand, and current and advected measurements of 15 & 30 minutes ago on the other hand. In the left part the IWV ranges is given and in the right part the IWV differences are visualized.

Another way to visualize differences caused by using advected measurements with respect to using only current measurements, is using a box plot. Figure 5-8 is an example of such a box plot. The left subfigure shows the IWV predictions of using current and advected measurements of 15 minutes ago (data of Figure 5-4(a)), and using current and advected measurements of 15 and 30 minutes ago (data of Figure 5-6(a)). The blue box indicates the upper and lower quartile of all data. The red horizontal line indicates the median of the data. The whiskers extend from the upper and lower quartile up to the minimum and maximum of the predictions, with a maximum length of 1.5 times the distance between the upper and lower quartile. All predictions that are not in between the whiskers is marked with a red plus sign. Using the advected measurements of 30 minutes ago, results in a slightly higher median. The right subfigure shows the predictions of the left subfigure minus the predictions which only uses current measurements. Also this box plot shows the higher data values when using also advected measurements of 11:44. Moreover, using in addition measurements of 30 minutes ago results in a wider range of differences.

In this chapter, only the differences in the IWV maps caused by using different weights for the advected measurements were presented. In the next chapter these differences are validated.

Chapter 6

Results and Validation

In Section 4-3 the process was described how IWV measurements from the past are used in the prediction of the current IWV map. It is clear that IWV maps are changing by including advected measurements from the past. In this chapter it is validated whether using advected measurements from the past in addition to the current measurements gives better results. Moreover it is investigated whether there exists an optimal duration with which IWV measurements are advected and which are used in the IWV map prediction.

This chapter starts with describing three methods of validation in Section 6-1. In Section 6-2 the diagnostic statistics used for the validation are described. In Section 6-3 a case study is described which is used for the validation. In Section 6-4, the results of the validation on the case study are described. This chapter is concluded with Section 6-5, which summarizes the results of the validation.

6-1 Methods of Validation

6-1-1 Test Set Cross Validation

One way of validation of a model is cross validation. In cross validation, part of the data is used for prediction (the training set), the remaining data is used for the validation (validation or testing set). One way of cross validation is the *test set method*, also called repeated *random sub-sampling validation*, or *hold out validation* (HOV). In the test set method part of the measurement data is excluded from the training set. Then the prediction from the training set is compared with the excluded validation set. It is an easy method, but if not much measurement data is available it does not perform very well, notably when even less data is used for the training. (Stone, 1974; Li et al., 2009)

6-1-2 Leave-One-Out Cross Validation

A more extensive method of cross validation is called *leave-one-out cross validation* (LOOCV). In LOOCV, one of the measurements at the ground stations is left out for the prediction of the

IWV map. Then the prediction at the ground station is compared with the measured values. The ground station used for this validation could be any ground station. With LOOCV, we repeat this validation for all stations. The mean of all differences between the predictions and original measurements at all measurement locations is used for the cross validation. Because the training process is repeated as many times as there are stations, this method of cross validation is really computer expensive. (Stone, 1974; Li et al., 2009)

6-1-3 K-fold Cross Validation

A trade-off between reliable validation and expensive validation from a computational point of view, is k-fold cross validation. In this case the dataset is split into k random subsets. Then one of these subsets is used for the validation set, the remaining subsets are used as the training set. This is repeated for all subsets, and the mean of all errors is used for the validation.

All of the afore mentioned cross validation techniques have their pros and cons. Because we only have 34 GPS ground stations the k-fold cross validation is not used as a validation technique. But since LOOCV takes a lot of time to process if all stations are used, the *test set method* is chosen. However, it is repeated for several ground stations, since the prediction accuracy will depend on the location of the ground station. This is because the measurements of the past are advected by the wind, and if no ground stations are upwind, there are no advected measurements for that position. For example, if the validation is performed at the location of Hoek van Holland (in the west of The Netherlands) we have no advected data, if the main wind direction is from the west.

6-2 Diagnostic Statistics

For the validation, multiple diagnostic statistics can be produced to determine how long measurements of the past should be advected to yield the best IWV values close to the original measurements. Two statistics are used in this chapter, the Mean Error and the Root Mean Squared Error.

6-2-1 Mean Error

A first diagnostic statistic is the Mean Deviation or Mean Error (ME). This is simply the difference between the prediction and the measurement, divided by the total number of epochs N,

$$\mathrm{ME} = \frac{1}{N}\sum_{i=1}^{N}\Big\{z(\boldsymbol{x_i}) - \hat{Z}(\boldsymbol{x_i})\Big\}. \tag{6-1}$$

In this equation $z(\boldsymbol{x_i})$ is the vector of the original time series of GPS IWV measurements at the ground station and $\hat{Z}(\boldsymbol{x_i})$ is the time series of IWV predicted at the location of that ground station. The mean error gives an indication of the extend to which the Kriging interpolation is underestimating or overestimating the predictions and thus is a measure of the precision. The precision is a measure of correctness (ESRI, 2011). Ideally it should be zero. (Webster & Oliver, 2007, p.191)

6-2-2 Root Mean Squared Error

The second diagnostic statistics is obtained when the square Root of the Mean Squared Error is taken (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left\{z(\boldsymbol{x_i}) - \hat{Z}(\boldsymbol{x_i})\right\}^2}. \tag{6-2}$$

The RMSE gives an indication of the average magnitude of the error and thus is a measure of the accuracy. The accuracy is a measure of correctness (ESRI, 2011). The RMSE is especially useful for detecting large errors. This is because the errors are squared before they are averaged. (*Eumetcal website*, 2011)

6-3 Case Study Description

The choice of using 30 May 2010 as a case study, was made in the beginning of the graduation. The purpose of this thesis is to investigate whether the IWV map improves by incorporating advected measurements. The possible improvement of water vapour maps will be especially useful in heavy weather situations. According to the weather prediction at 29 May 2010 for the next day:

"The morning starts cloudy with occasionally some rain. Gradually clouds break open. During the day storms are developing, inland also associated with lightning and possibly hail. The showers may be associated with gusts to around 70 km/hour, in the southeast to 80 km/hour. Initially the wind is south to southwest and moderate, 3 to 4 Bft., at sea powerful, 5 to 6 Bft. The wind turns west to northwest in the afternoon and increases in the late afternoon and evening to 6 to 7 Bft. at the west coast." (KNMI, 2010a)

And afterwards according to the weather description of the KNMI:

"On the 30th of May some showers were present, especially around the passage of a trough (rear of a depression). The showers were sometimes accompanied by lightning and hail." (KNMI, 2010b)

In Figure 6-1 a screen shot of the rain radar is shown at 17:00 local time (= UTC+2). The heavy rain is clearly visible in the east of The Netherlands.

The choice of using this particular date for the validation of my advection approach will have an effect on the validation results. This is because it will be hard to distinguish between errors in the model which result in a under or overestimation of the amount of water vapour, and between a change in water vapour because there were showers. It is therefore recommended to also validate the model with data from a calm day when no heavy showers were present.

6-4 Results of Validation

In this section the diagnostic statistics of Section 6-2 are used on the case study of 30 May 2010, for the validation of the model described in Section 4-3. Five stations are sequentially removed from the input data, namely Assen, Boxmeer, Breda, Utrecht, and Hoek van Holland. The

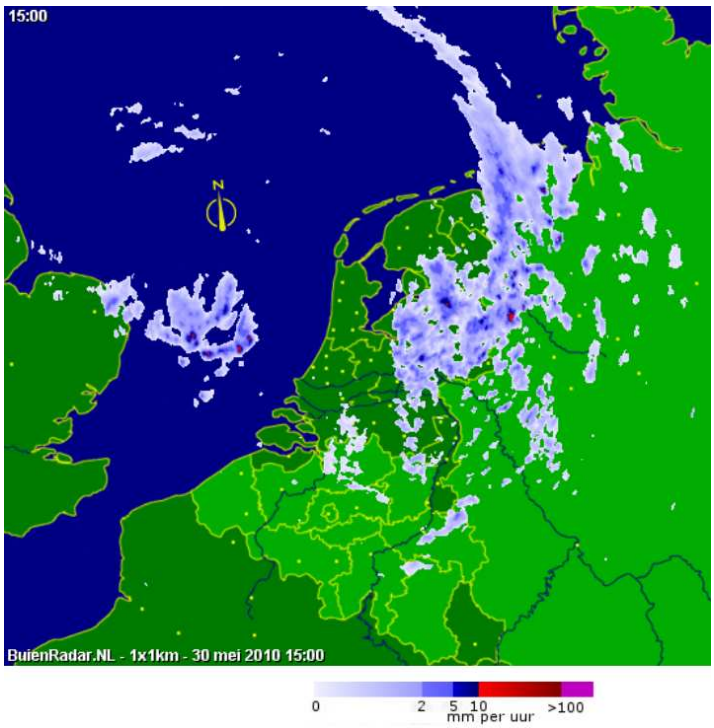


Figure 6-1: Rain radar image at 30 May 2010 at 15:00 local time (= UTC+2), showing heavy rain at the east of The Netherlands (buienradar.nl).

station which is removed from the GPS IWV data is then used for the validation. Assen, Boxmeer, Breda, and Utrecht are chosen because these stations are not located close to the west coast, where almost no advected measurements arrive. By choosing those four locations, it is expected that the advected measurements from other stations have an influence on the predicted IWV values. Hoek van Holland is also included as a validation location to confirm whether using epochs from the past does not influence the predictions at the west coast that much, when there is a dominant wind from the west.

For Assen, Breda, and Hoek van Holland figures will be shown for the IWV measurements and predictions as a function of time. In Appendix A larger figures can be found, as well as for two other locations, Boxmeer and Utrecht. Also the IWV differences of the predictions with respect to the measurements can be found in Appendix A. These figures give a clearer overview of the time evolution of these differences.

In Figure 6-2 the absolute value of the Mean Error and the Root Mean Squared Error are visualized for the five mentioned locations. The duration for which measurements are advected is on the x-axis, and the error is on the y-axis. For all stations discussed in this section, the errors are shown resulting from using different, as well as equal weights for current and advected measurements (see Subsection 4-3-3 and Chapter 5). In Table A-1 in Appendix A the values of the ME and the RMSE are shown, which are used to generate the two figures. In Figure A-11 and A-12 larger figures of the ME and RMSE as a function of advected duration can be found.

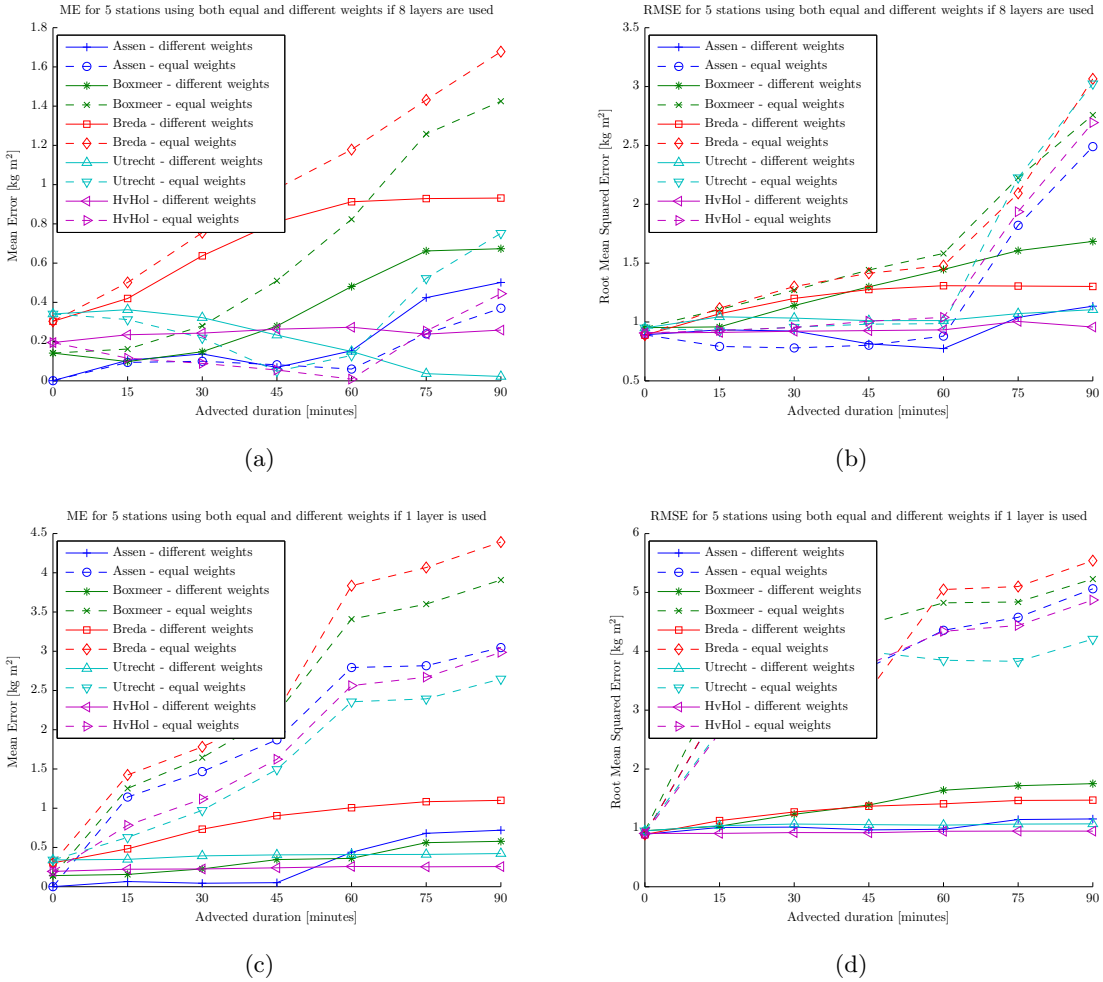


Figure 6-2: (a) & (c) Absolute value of the Mean Error, and (b) & (d) Root Mean Squared Error, as a function of advected duration for 5 stations using equal weights for current and advected measurements (dashed lines) and using lower weights for advected measurements (solid lines), the atmosphere is divided into 8 layers for (a) & (b), and into 1 layer for (c) & (d).

6-4-1 Assen

IWV prediction as a function of time for Assen

Assen is situated at the northeast of The Netherlands (see Figure 3-1). In Figure 6-3 the IWV measurements and predictions at the location of Assen are shown. See Figure A-1 in Appendix A for a larger version of the same figure. The original IWV measurements are indicated by black squares, connected with a thick black line. The predictions at the ground station location using only current measurements, are visualized by the dashed blue line with stars. The predictions using only current measurements, more or less follow the same pattern. However, there are also large differences. For example at 15:00 the difference is about 2.5 km m^{-2}. The other colored lines represent the predictions by using 15 up to 90 minutes advected measurements from the past. Some of these predictions follow the measurements closer. For example at 15:00, using 15, 30, 45, or 60 minutes of advected

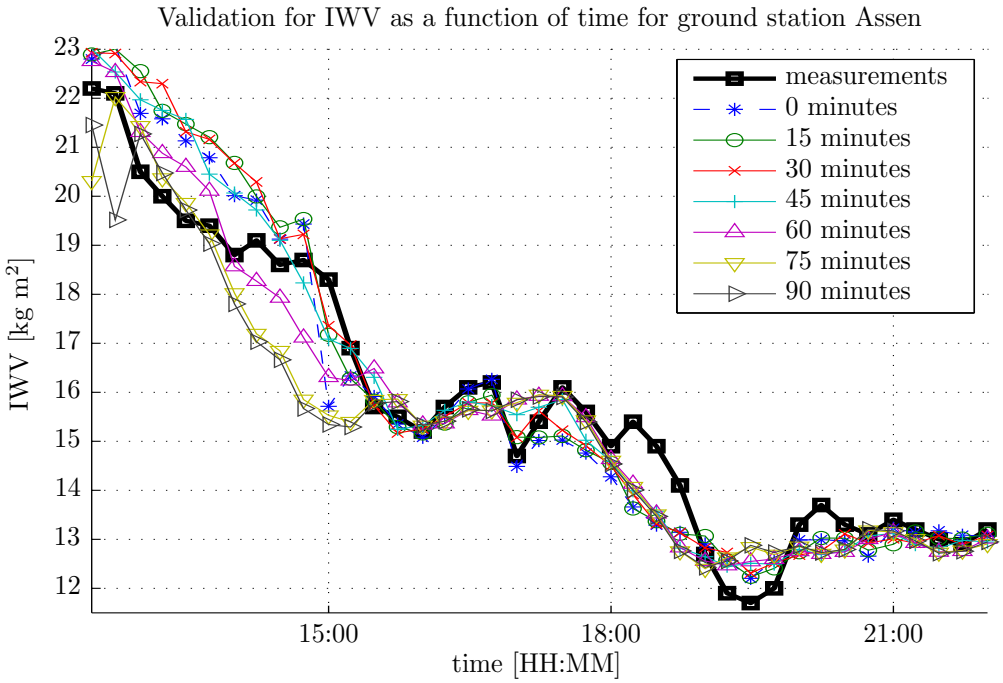


Figure 6-3: IWV as a function of time for Assen (ASS2), the black squares are the measurements at the ground station, the colored lines are the IWV values as predicted by the different Kriging interpolations using only current measurements (0 minutes, blue line) and using 1–6 advected measurements from 15–90 minutes from the past.

measurements results in a prediction closer to the measurements. Sometimes however, the predictions are getting worse. For instance at 16:44, all predictions using advected measurements, result in an underestimation of the IWV. Using advected measurements from more than 60 minutes ago, results in general in a worse prediction.

Since one of the main assumptions in our model for advection is that water vapour is conserved, we assume that on short time scales of the advection evaporation and condensation are not significant. However, we do not have a good means to verify this. What can be done instead is to have a look at the precipitation measured by weather radar. In Figure 6-4 the rainfall index weight level is shown for the province of Drenthe where Assen is located. The index is shown for the period between 12:30 and 22:00 of 30 May 2010. The rainfall index weight level is a measure of the rainfall intensity introduced by *Buienradar.nl*. It shows the relative amount of rainfall with respect to the theoretical maximum precipitation. Values around 20–30% already indicate heavy precipitation. The higher the index the more rainfall. (BuienRadar, 2010)

If we compare Figure 6-4 and Figure 6-3 it can be seen that high values of the rainfall index weight level for a particular epoch, in general coincide with large variability of the IWV predictions for that epoch.

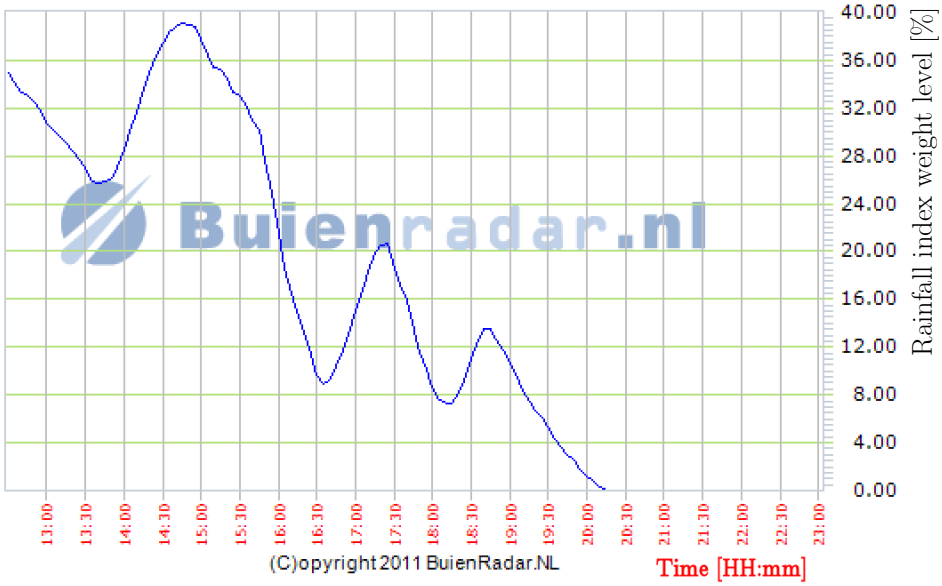


Figure 6-4: Rainfall index weight level for Drenthe, the Province where Assen is situated (buienradar.nl).

Mean Error of IWV prediction at Assen

In Figure 6-2(a) the absolute value of the Mean Error as a function of advected duration is shown. Using equal and modified weights, results in the same error when no advected measurements are used. This Mean Error is almost 0 kg m^{-2} for Assen. Using current plus advected measurements from 15 minutes ago, results in a difference of 0.10 kg m^{-2}. Using 30 minutes of advected data, the error slightly increases to 0.14 kg m^{-2}. Using 45 minutes of advected data however, the error decreases again to 0.07 kg m^{-2}. Using advected data of a longer time ago, gives a larger (absolute) difference. So, based on the Mean Error at station Assen, best predictions are obtained using advected measurements of up to 60 minutes from the past.

Root Mean Squared Error of IWV prediction at Assen

The Root Mean Squared Error, see Subsection 6-2-2, is shown in Figure 6-2(b). The RMSE of the predictions using only current measurements with respect to the measurements at Assen are shown again by the blue lines. Using only current measurements for the prediction of IWV at the location of Assen, results in a RMSE of 0.9 kg m^{-2} with respect to the measurements. If down weighted weights are used, using 15 and 30 minutes of advected data, results also in a RMSE of 0.9 kg m^{-2}. If however measurements of 45 and 60 minutes ago are used, the RMSE is decreased to 0.8 kg m^{-2}. When using more advected measurements the RMSE increases again. If the weights are not down weighted, the RMSE for 15 up to 45 minutes of advected measurements is at the level of 0.8 kg m^{-2}, and 0.9 kg m^{-2} if 60 minutes of advected measurements are used. If more advected measurements are used the RMSE rapidly increases. So if we want a RMSE as low as possible, best option is to use 45 or 60 minutes of advected

measurements with down weighted weights or 15 up to 45 minutes of advected measurements where the weights are not down weighted.

IWV prediction at Assen if 1 vertical layer in the atmosphere is used

In Subsection 4-3-1 it was chosen to divide the atmosphere in 8 vertical layers. To confirm whether this choice gives better results than using only 1 layer, in Table A-1 the Mean Error and Root Mean Squared Error are shown for the case where only 1 layer in the atmosphere is used. The wind at 1 km height is used to advect the measurements, because most of the water vapour is between 0 and 2 km height. The ME and RMSE as a function of advected duration are shown in Figure 6-2. Compared to using 8 layers in the atmosphere, the ME and RMSE when using different weights, are all slightly higher. If equal weights are used, the ME and RMSE for one layer are all clearly higher.

We now have a closer look at the prediction at Assen at 15:00, as a function of advected measurements using 1 layer. In Figure 6-5 the measurements closest tot Assen are visualized. The station locations are visualized by black crosses and to the left of each station the station name is given. The color of the circles indicate the IWV measurements in [kg m^{-2}]. Above each measurement the weight given to the particular measurement is indicated. The IWV prediction is performed for the location of Assen which is at the center of the figure. The original IWV measurement at Assen at 15:00 was 18.3 kg m^{-2} which is indicated by the orange-red color of the circle. The IWV prediction at each subfigure is indicated by the value above the location of Assen.

Although Figure 6-5 is regarding the IWV prediction when only 1 layer is used, it does show some problems which arise in general. For example, in Figure 6-5(d) the measurement at the station of Leeuwarden (LWRD) in the left of the figure, is assigned a weight of 0.01 whereas this is the current measurement at that station. The advected measurement of 60 minutes ago is assigned a weight of 0.155. This large difference in weights is caused by the screening effect of Kriging (see Section 5-4). Inverse distance weighting, possible also with modified weights for advected measurements, could give better results.

6-4-2 Breda

IWV as function of time for Breda

Breda is located at the south of The Netherlands. The measurements and predictions at this location are shown in Figure 6-6. See Figure A-5 for a larger version. For the location of Breda till about 15:30, all predictions are overestimating. This is the case if only current measurements are used, but even more if advected measurements are also used.

Mean Error of IWV prediction at Breda

Figure 6-2(a) shows that the Mean Error using only current measurements is about 0.3 kg m^{-2}. When advected measurements of 15 minutes ago with different weights for the advected measurements are used, this error increases to about 0.4 kg m^{-2}. The Mean Error keeps increasing up to an advected duration of 60 minutes. Using equal weights for current and advected measurements results in a higher Mean Error.

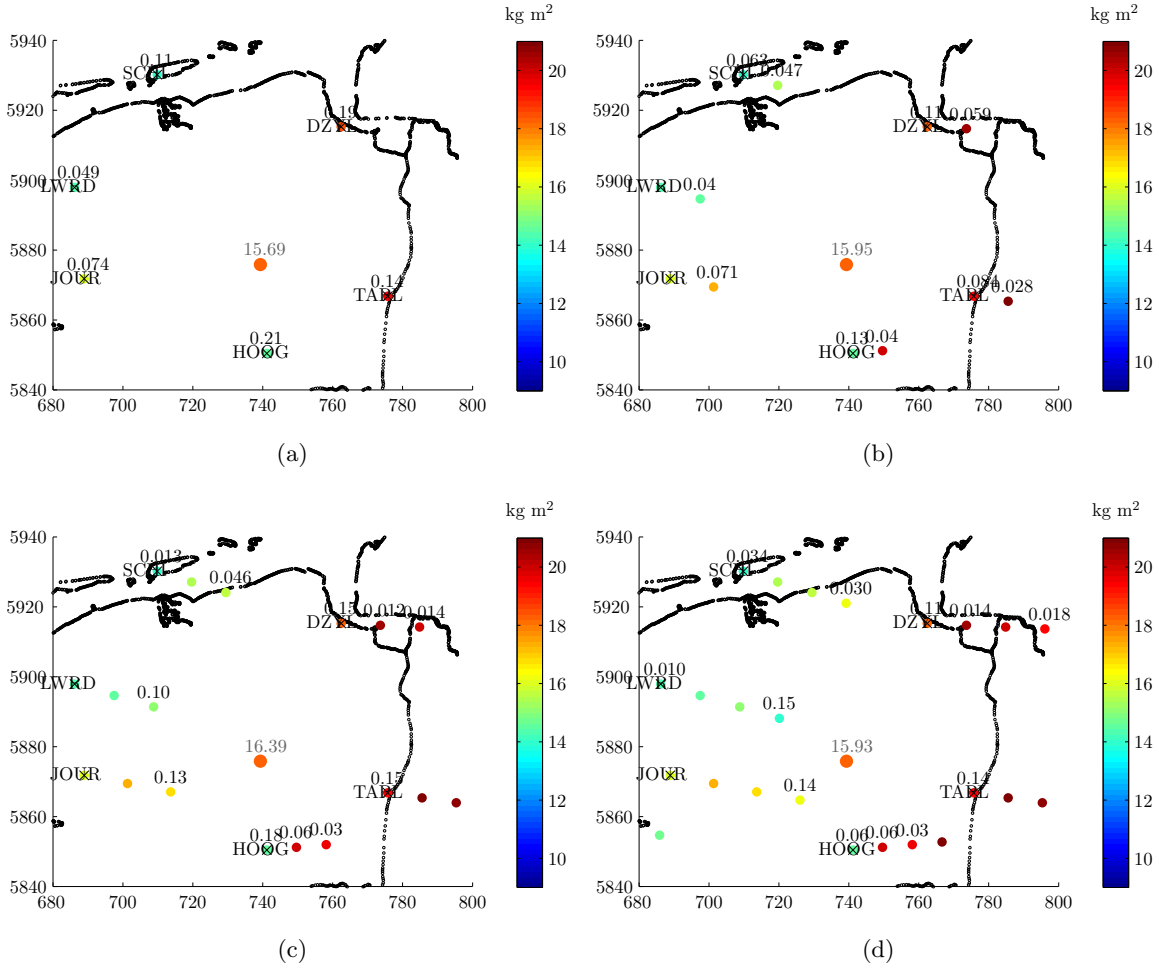


Figure 6-5: Weight numbers near Assen at 15:00 when 1 vertical layer in the atmosphere is used, using current and (a) no advected measurements, and advected measurements of (b) 15, (c) 15 & 30, and (d) 15, 30 & 45 minutes ago. The station locations are visualized by the black crosses. Near each station are the current and advected measurements in kg m^{-2} given by the color of the circle. The number above the measurements indicates the weight used for the prediction at the location of Assen. Weights with an absolute value less than 0.01 are omitted for clarity. The location of Assen is near the center of the figure. The original measurement at Assen is given by the color of the circle, while the predicted value is given above Assen

Root Mean Squared Error of IWV prediction at Breda

If we look at Figure 6-2(a), the RMSE is increasing from 0.9 kg m^{-2} when only current measurements are used, to about 1.3 kg m^{-2} when advected measurements from 60 minutes ago are used. For 75 and 90 minutes of advected measurements, the RMSE stays about equal. This is the case if different weights are used. If equal weights are used however, the RMSE is slightly higher for 15 to 60 minutes, and clearly higher for 75 and 90 minutes of advected measurements.

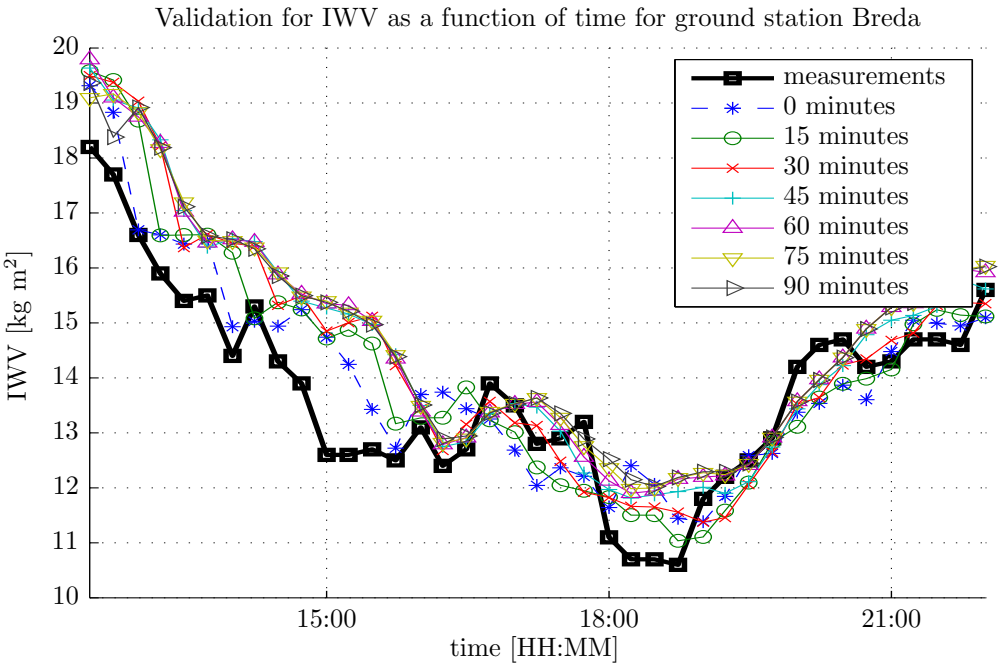


Figure 6-6: IWV as a function of time for Breda (BRD2), the black squares are the measurements at the ground station, the colored lines are the IWV values as predicted by the different Kriging interpolations using only current measurements (0 minutes, blue line) and using 1–6 advected measurements from 15–90 minutes from the past.

IWV prediction at Breda if 1 vertical layer in the atmosphere is used

From Table A-1 it can be seen that also for Breda holds that using 8 layers gives better results than using only 1 layer. When different weights are used for the advected measurements, the difference is only small. However, if equal weights are used in combination with 8 vertical layers, the errors are clearly lower. This is also visible in Figure 6-2.

Also for the prediction at Breda, we use 1 vertical layer in the atmosphere to get a better impression of the IWV prediction. In Figure 6-7, we see the closest data points used for the prediction at the location of Breda at 15:44. In Figure 6-7(a) we see that the prediction using only current measurements resulting in a value of 12.7 kg m^{-2}, is close to the original measurement of 12.5 kg m^{-2}. Using also the advected measurements of 15 minutes ago, the IWV prediction increases to 13.3 kg m^{-2}. If we look at Figure 6-7(b), this increase is primarily caused by the advected measurement of Willemstad (WILL). This advected measurement has a value of about 13.5 kg m^{-2} and an assigned weight of 0.44. In Figure 6-7(c), where the advected measurements of 30 minutes ago are also used, we see that the advected measurements of 30 minutes ago from Willemstad almost coincides with the location of station Breda. The yellow dot of the advected measurement is to the south-west of Breda and is almost invisible because it is behind the circle of the original measurement. The advected measurement has a value of 15 kg m^{-2} and an assigned weight of 0.651. The resulting IWV value of 14.4 kg m^{-2} is thus clearly an overestimation, compared to the actual IWV measurement at Breda.

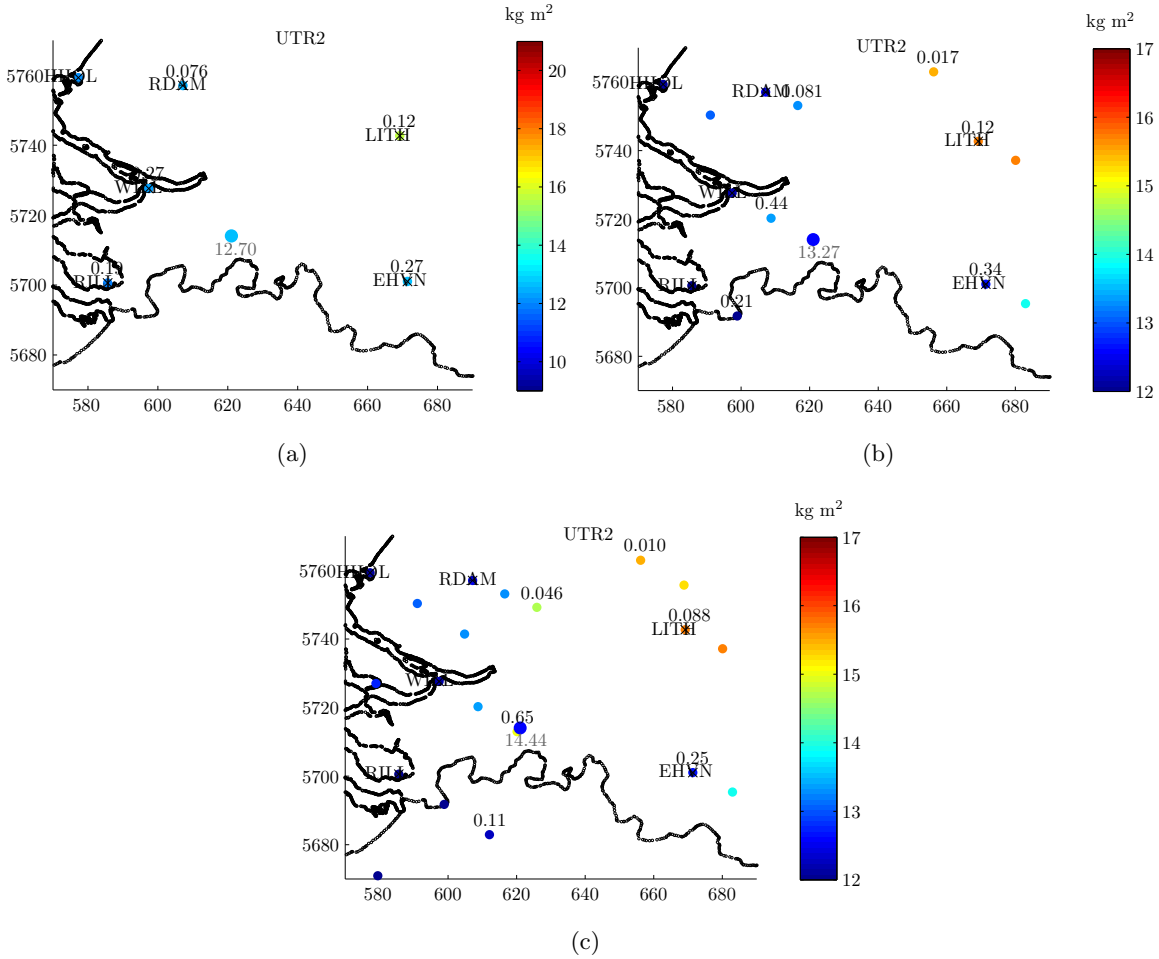


Figure 6-7: Weight numbers near Breda at 15:44 when 1 vertical layer in the atmosphere is used, using current and (a) no advected measurements, and advected measurements of (b) 15, and (c) 15 & 30 minutes ago. The station locations are visualized by the black crosses. Near each station are the current and advected measurements in kg m^{-2} given by the color of the circle. The number above the measurements indicate the weight used for the prediction at the location of Breda. Weights with an absolute value less than 0.01 are omitted for clarity. The location of Breda is near the center of the figure. The original measurement at Breda is given by the color of the circle, while the predicted value is beneath Breda.

6-4-3 Hoek van Holland

Under the circumstances of 30 May 2010, with a main wind direction from the west at the lower layers, using advected measurements does not influence the predictions at Hoek van Holland much. This is shown in Figure 6-8. It is clear that all predictions are very close to each other, and using advected measurements does not improve the predictions with respect to the original measurements.

When we look at Figure 6-2(a) the Mean Error for all predictions is low and is about 0.2 kg m^{-2}. If equal weights are used, the Mean Error is lower if advected measurement from 15–60 minutes ago are used. The RMSE, shown in Figure 6-2(b), indicates that using

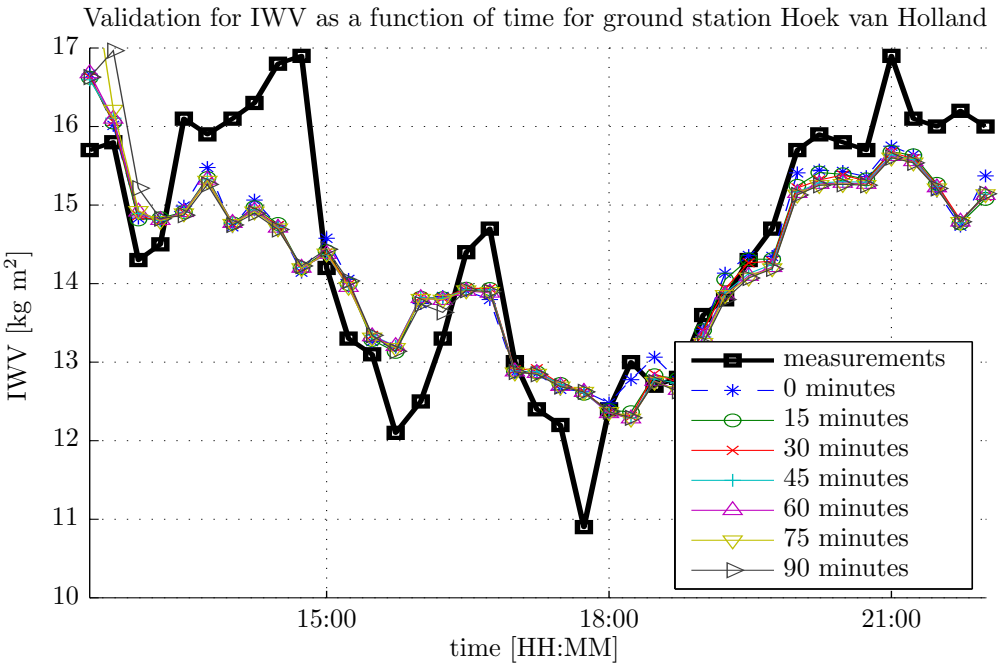


Figure 6-8: IWV as a function of time for Hoek van Holland (HHOL) the black squares are the measurements at the ground station, the colored lines are the IWV values as predicted by the different Kriging interpolations using only current measurements (0 minutes, blue line) and using 1–6 advected measurements from 15–90 minutes from the past.

advected measurements of up to 60 minutes does not change the RMSE. This is the case for both equal and different weights. If equal weights are used, the RMSE rapidly increases if measurements are advected more than 60 minutes.

So, it was assumed that for locations without measurements upstream, advection has no impact on the IWV prediction and this also shows in the result.

6-4-4 Boxmeer

Boxmeer is located at the east-southeast border of The Netherlands, near Nijmegen. A figure of the IWV values and differences as a function of time can be found in Figure A-3 and A-4.

The Mean Error as shown in Figure 6-2(a) suggests that using 15 minutes of advected measurements, gives the best results. But also using no or 30 minutes of advected measurements results in a low ME. Using more advected measurements increases the ME. Again, using different weights for the advected measurements, gives better results than using equal weights. The RMSE in Figure 6-2(b) gives equal results when using only current measurements, and using current measurements plus advected measurements from 15 minutes from the past. When using measurements from a longer time ago, the RMSE is increasing. Also for the location of Boxmeer, using different weights gives better results than using equal weights. This is especially the case if advected data of more than 60 minutes ago is used.

6-4-5 Utrecht

Centrally located in The Netherlands is the last validation location, ground station Utrecht. If we look at Figure A-7 and A-8 it is hard to determine if the predictions are getting better or worse using advected measurements. This can also be seen in Figure 6-2. If different weights are used, the Mean Error decreases if measurements of a longer time ago are used. However, the RMSE stays about constant if different weights are used. For equal weighs, the RMSE is rapidly increasing if the advection duration is longer than 60 minutes.

6-5 Summary

In this section the statistics of the previous section are summarized.

Table 6-1: Mean of the absolute value of the ME and mean of the RMSE as a function of advected epochs used for four model options, both in [kg m^{-2}]. The means are computed from the absolute values of the ME and RMSE for Assen, Breda, Hoek van Holland, Boxmeer, and Utrecht (see Table A-1).

Mean Error [kg m^{-2}]	0	15	30	45	60	75	90
8 layers, different weights	0.4	0.4	0.5	0.5	0.5	0.6	0.6
8 layers, equal weights	0.4	0.4	0.4	0.5	0.6	1.1	1.4
1 layers, different weights	0.4	0.4	0.5	0.6	0.6	0.7	0.7
1 layers, equal weights	0.4	1.4	1.7	2.3	3.3	3.5	3.8
Root Mean Squared Error [kg m^{-2}]	0	15	30	45	60	75	90
8 layers, different weights	0.9	1.0	1.0	1.1	1.1	1.2	1.2
8 layers, equal weights	0.9	1.0	1.1	1.1	1.2	2.1	2.8
1 layers, different weights	0.9	1.0	1.1	1.1	1.2	1.3	1.3
1 layers, equal weights	0.9	3.0	3.2	3.9	4.5	4.6	5.0

Figure 6-2 gave an overview of all validation results of the four models for all stations. From these figures it is hard to say which of the four models results in the best IWV map and how long measurements should be advected. Table 6-1 lists the means of the absolute value of the Mean Error and the mean of the Root Mean Squared Error of all model options. This table shows that on average the absolute value of the Mean Error and the Root Mean Squared Error are not decreased by using advected measurements. So using no advected measurements results in the best IWV map. However, the table contains more information.

An obvious conclusion which is drawn from the table is that using only 1 layer *and* using equal weights for the current and advected measurements, results in the worst IWV map. The results can easily be improved by either using lower weights for the advected measurements, or by dividing the atmosphere into 8 layers. Using 8 layers in the atmosphere especially improves the results if measurements are advected up to 60 minutes. If we want to use advected measurements of a longer time ago, using different weights further improves the results. Although the differences between the errors are very small between the first and the third option, better results are obtained when 8 layers and different weights are used.

For the division of the IWV values among the vertical layers, data from radiosondes are used. Because these are launched only twice a day, around midnight and noon, it was thought that this would cause degradation of the IWV prediction accuracy as a function of time, until data from a new radiosonde is available. But if the figures of the IWV differences as a function of time in Appendix A are investigated this is not the case. At least not for this particular case study.

Apart from these conclusions one has to bear in mind that the numbers in Table 6-1 are computed from the ME and the RMSE from the 5 stations mentioned before. So for each station there is a different optimal model, depending on the location and the weather circumstances.

Chapter 7

Conclusions and Recommendations

In this chapter we answer the main research questions as posed in Section 1-2. In Section 7-1 first a general discussion is given about the assumptions, limitations, and computational aspects. In Section 7-2 the conclusions are given. This chapter will be finished of with the recommendations in Section 7-3.

7-1 Discussion

In this section the assumptions made during the process are summarized. These include the assumptions made during preprocessing, as well as during the water vapour mapping itself.

7-1-1 Assumptions Preprocessing

For the preprocessing of the input data described in Section 3-1 and Section 4-2, the following assumptions were made:

- The original IWV measurement error for the Integrated Water Vapour value at the GPS ground station is assumed to be 1 kg m^{-2}.
- For the water vapour division from the radiosonde data, the absolute humidity as computed from the formulas in Subsection 4-2-1 is used. For the water vapour division from the radiosonde a maximum height of 8 km is used, because above this height almost no water vapour is present. Furthermore it is assumed that the vertical profile retrieved from the radiosonde is representative for the whole area used during processing and this profile is used for all further epochs, until a new profile is measured by a new radiosonde.
- It is assumed that the water vapour in the air moves with the same velocity as the wind velocity (speed and direction), given by the HIRLAM wind field. This wind data is updated every 3 hours. The wind velocity at 1 km intervals is used, starting from 1 km

height up to 8 km height. So for example, the water vapour between 0 and 1 km height is assumed to be transported by the wind given by the model at 1 km high. Furthermore it is assumed that the HIRLAM wind data does not contain any error for the speed and direction. An error in each of them would cause an error in the position of the advected measurements.

- UTM projection is used for the projection of all data. UTM zone *31U* is used. Main part of the data is inside this zone.

7-1-2 Assumptions Generating the IWV Map

For the processing described in Section 4-3 and Chapter 5, also several assumptions have been made. Some of them are general assumptions, some of them are regarding the program itself, and others are specific for Kriging:

- It is assumed that the advection of the water vapour can be modeled using only 8 horizontal layers of 1 km thick. Also no vertical exchange of water vapour is assumed between the different layers during the time of advection.
- Related to the previous item is that it is assumed that water vapour is preserved inside each layer during the time of advection. So also no precipitation or advection is allowed.
- The weighting function (Equation (5-4) and (5-6)) for current measurements and measurements advected by wind, is equal for all ground stations at all time. The weighting function is computed from GPS IWV data of May 2010 at four stations.
- Ordinary Kriging is used for the interpolation of the water vapour map. This results in the following assumptions and limitations:
    - Ordinary Kriging assumes intrinsic stationarity or wide sense stationarity of the field. This means that a dataset comes from a random process with a constant mean and a semi-variogram that depends only on the distance (and direction for an anisotropic variogram model) separating any two locations, (ESRI, 2011). The dataset does not come from a random process and also the water vapour in the air is not obliged to depend on the distance.
    - The other assumption of Ordinary Kriging is that it assumes an unknown constant mean of the water vapour in the spatial domain. The mean can vary for each time epoch, but at each time epoch this mean is constant. When for example, water vapour quickly drops at one station at the edge of the domain, this can cause a wrong mean and so could potentially cause an error in the whole spatial domain. A solution to overcome this problem could be to first remove a possible trend from the data and then use Kriging on top of this trend.
    - Also, if too few stations are used with Kriging, the computed variogram model will be a poor representation. This will result in a map which is a poor representation of the water vapour.

  These assumption are however necessary to use Kriging as an interpolation technique. Other interpolation techniques, also require assumptions.

- An *isotropic* variogram model is used for the Kriging process of the spatial interpolation. Depending on the weather conditions and the distributions of the ground stations however, using a *anisotropic* variogram could better reflect the water vapour change in different directions. Compare this to the example of small sand ripples on the beach formed by wind with respect to the large scale topography of the beach.

These assumptions will effect the results and maybe even the conclusions. Therefore, after the conclusions drawn in the next section, in Section 7-3 some recommendations will be given.

7-1-3 Computational Aspects

For the trade off between different models, the computational effort also has to be taken into account. For the processing of the case study an Intel Core 2 was used with 2GHZ clock rate and 2GB of physical memory. Data from 11:00 to 22:00 was used, and 33 ground stations were used for the generation of the IWV map. The 34th station was used for the validation. From the processing the following inferences were made:

- Using 8 layers, using advected measurements from only 15 minutes ago, and using different weights for the advected measurement, takes about 5 minutes of processing. The computation time increases by 3 to 5 minutes per extra advected epoch. Main cause of the increase of processing time, is the increase of workspace used by MATLAB, which increases because matrices were not removed from the workspace. The workspace of MATLAB is the memory used to store all variables.
- Difference in computation time between using 1 and 8 layers is almost negligible. Using 8 layers only takes more time to store, because for each time epoch, 16 matrices have to be stored, instead of only 2 (IWV and Kriging variance).
- Computing the IWV and Kriging variances, both using the same and different weights, using epochs from 0–90 minutes of the past, using 8 layers takes about 50 minutes of computational time for a period of 10 hours of predictions. This provides all information to validate the IWV map at one station, as shown in Section 6-4. The size of the workspace generated by MATLAB containing all variables and matrices of the predictions for a period of 10 hours is about 1GB.

So, we can conclude that increasing the model complexity does not really increase the computational heaviness, but does increases the amount of data generated. So near real time usage of the model is possible.

7-2 Conclusions

7-2-1 Main Conclusion

In Section 1-2 the main research question was defined as *'Will the 2D mapping of water vapour in the atmosphere improve if we include advection of water vapour into the model?'* From Section 6-5 we conclude that *on average*, this question is answered negatively, at least with the used settings. All possible options result in an increase for the Mean Error and the Root Mean Squared Error, for the five validation stations mentioned. Actually this conclusion is too straightforward and requires further clarification which is given in the next subsection. Also it has to be mentioned that only one day of data has been used, which also contains a period of heavy weather.

7-2-2 Other Conclusions

- The first subquestion was: *'Divide the 2D water vapour mapping into several vertical layers, each layer containing part of the water vapour obtained from GPS measurements. The question here will be how to split up the atmosphere in layers and how to divide the water vapour among these layers.'* Already in the beginning the choice was made to divide the atmosphere into layers of 1 km thick and up to 8 km height. The maximum height of 8 km probably was a good choice, because above this height almost no water vapour is present. However the main reason at the beginning of the thesis for choosing a layer thickness of 1 km was the ease of programming. Also in the beginning of the thesis it was thought that using 8 layers was sufficient, at least for starting of the development of the program. Only at the end of the thesis this choice was questioned again, but due to time constraint it was chosen not to use more layers. Instead it was chosen to look at the result if only 1 layer is used. In Section 6-5 both the results of using 8 layers and 1 layer were analyzed. From the validation results in Section 6-4 it is concluded that using eight layers gives better results than using one layer.

- The other part of the first subquestion was how to divide the IWV among the vertical layers. Radiosonde data was used to divide the IWV among the layers. These radiosondes are launched only twice a day. Therefore it was expected that the IWV prediction accuracy as function of time degrades. However, for the IWV predictions as a function of time at 30 May 2010, this is not the case. So using radiosonde data to divide the GPS IWV measurement among the layers is a suitable method.

- The second subquestion was *'Use wind data to advect water vapour from previous measurements to obtain extra data points. Question here will be how long data points of the past should be used and how to weight them, which will result in the best IWV map.'* The third subquestion was *'Use and compare different interpolation techniques to see which technique yields the best result.'* Because the third subquestion actually considers which algorithm settings results in the best interpolated map, it will be treated together with the second subquestion. This is the most tricky question, as there is no clear answer to this question, as there is no single option that results in an improvement of the IWV prediction *on average*. For some of the validation locations advecting measurements does not improve the predictions at the validation location at all. Some of them even

get worse, although this was mostly the case where the prediction using no advected measurement already gave a prediction close to the measurements. For other stations, advecting measurements did improve the prediction accuracy. However, the optimum duration of advection vary for all stations. Optimum advection duration for most stations is 15 minutes, but other stations have their optimum at 30 minutes up to even 90 minutes. It could be investigated if the optimum depends on the wind speed and direction and the location of other stations. But it is concluded anyway that for locations without measurements upstream, advection has no impact on the IWV prediction.

- Another important conclusion, is that assigning lower weights to advected measurements improves the water IWV maps when it is compared to the case where advected and current measurements have the same weight. This of course only applies to the case where measurements are advected. The increase in prediction accuracy is especially notable if only 1 layer is used. Kriging extrapolating in time at the stations is used for modifying the Kriging weights (Chapter 5). But the Kriging variance for epochs in time is growing very fast, resulting in low weights for advected epochs. When using more epochs of the past this decrease is even higher. This could cause that advected measurements are trusted too less. It should be noted that the method used for down weighting of advected measurements is one possibility, other ways for downgrading could also be investigated.
- At an early stage it was decided to use Kriging as the interpolation technique. Although this technique has several advantages as described in Section 3-2, perhaps this is not the most suitable interpolator for this application. Especially the screening effect is a major drawback of Kriging for this particular application. This screening effect is even enhanced if wind data is updated only every three hours. In that case the advected measurements near each station are all on a straight line. Other interpolation techniques could manage these kind of data better.

7-3 Recommendations

7-3-1 Main Recommendations

From the conclusions it is clear that choices made for the water vapour map generation should be re-evaluated. Other options, for example the number and division of layers in the atmosphere, could result in better water vapour maps. But the first and most important recommendation is that another case study has to be done. That case study should be a period without rain. If we would track a volume of air in case it is not raining, the amount of water vapour can only increase due to evaporation (and sublimation). The measured IWV value should then always be larger than or equal to the predicted value using advected measurements. If no clouds are present, MERIS satellite data can also be used for the validation. Another option could be to further investigate the current case study. Per station the rain data should be compared with the water vapour as a function of time.

Because the wind data is updated only every 3 hours, advected measurements are positioned in a line. The screening effect associated with the Kriging interpolation, can cause advected or current measurements to have almost no effect on the interpolation at certain regions.

It is therefore highly recommended to investigate other interpolation methods. An inverse distance weighted interpolation could be a good alternative. It is also possible to combined such a method with modified weights for current and advected measurements.

7-3-2 Other Recommendations

- HIRLAM consists of 60 horizontal layers with 3 layers in the lowest 200 m. During the thesis only 8 layers of 1 km are used. At lower altitudes the wind speed and direction can vary a lot, and because at lower altitudes most of the water vapour is present, a different position for advected measurements has a larger influence when compared to higher altitudes. Therefore it could be interesting to use more layers for low heights, where more water vapour is present. The question remains if using higher vertical wind sampling results in better results if the vertical distribution of water vapour is lagging behind because of the radiosonde data used.
- Also during processing the wind which is used for advecting the water vapour is at the top of each layer. This could cause that the position of the advected measurements is too far from the ground station, assuming a larger wind speed at higher altitudes. Therefore instead of using the wind velocity at the top of each layer, it is recommended to investigate whether using the velocity halfway between each layer results in better predictions.
- The determination of the variogram model is computed for each time epoch and for all layers again and again. By doing so, changes in weather circumstances could be better reflected in the Kriging interpolation. It could be investigated if this is really the case by testing different weather scenarios.
- For the water vapour map generation a constant number of advected measurements is used. If the wind speed is higher, advected measurements will be sooner at nearby stations. It could therefore be investigated if making the number of advected measurements dependent on the distance between the stations and the wind speed, results in better predictions.
- In addition it could be investigated if using space-time models could increase the accuracy of the model. See for example (Kyriakidis & Journel, 1999) and (Kerwin & Prince, 1999). Space-time models can be designed to incorporate the temporal correlation between a series of spatial functions. In that way changes in weather conditions as a function of time could be better represented within the model.

Bibliography

Barker, D., Huang, W., Guo, Y.-R., & Bourgeois, A. (2001). *A three-dimensional variational (3dvar) data assimilation system for use with mm5.* Technical documentation. Available from `http://www.mmm.ucar.edu/mm53dvar/docs/3DVARTechDoc.pdf`

BuienRadar. (2010). Website. Available from `buienradar.nl` (last visited November 2011)

Christopher, M., & Gold. (1993). Contouring: A guide to the analysis and display of spatial data: by david f. watson, 1992, computer methods in the geosciences, volume 10, pergamon press, oxford, 321p. *Computers & Geosciences*, *19*(10), 1571 - 1572. Available from `http://www.sciencedirect.com/science/article/pii/009830049390069H`

Davis, J. L., Herring, T. A., Shapiro, I. I., Rogers, A. E. E., & Elgered, G. (1985). Geodesy by radio interferometry: Effects of atmospheric modeling errors on estimates of baseline length. *Radio Sci.*, *20*(6), 1593–1607. Available from `http://dx.doi.org/10.1029/RS020i006p01593`

Elgered, G., Plag, H., Marel, H. Van der, Barlag, S., & Nash, J. (2005). *Exploitation of ground-based GPS for operational numerical weather prediction and climate applications* (Final Report No. COST Action 716). European cooperation in the field of scientific and technical research. Available from `http://bookshop.europa.eu/eubookshop/bookmarks.action?target=EUB:NOTICE:QSNA21639:EN:HTML&request_locale=EN`

Encyclopædia Britannica. (2010a). *Atmophere.* Encyclopædia Britannica Online. Available from `http://www.britannica.com/EBchecked/topic/41364/atmosphere` (Online; last visited 2010-04-19)

Encyclopædia Britannica. (2010b). *Hydrosphere.* Encyclopædia Britannica Online. Available from `http://www.britannica.com/EBchecked/topic/279025/hydrosphere` (Online; last visited 2010-03-19)

Environment Canada. (2006). *Common interpretation errors.* Website. Available from `http://www.msc-smc.ec.gc.ca/cd/factsheets/weather_radar/index_e.cfm` (Retrieved 2010-04-19)

ESRI. (2011). Available from http://support.esri.com/en/knowledgebase/Gisdictionary/browse

Eumetcal website. (2011). website. Available from http://www.eumetcal.org/resources/ukmeteocal/verification/www/english/msg/ver_cont_var/uos3/uos3_ko1.htm (Retrieved 2011-09-5)

Fabry, F. (2000). *Definition: dual-polarization.* Available from http://www.radar.mcgill.ca/define_dual_pol.html (Retrieved 2010-04-19)

Fischer, J., & Bennartz, R. (1997). *Algorithm theoretical basis document - retrieval of total water vapour content from meris measurements.* website. Available from http://envisat.esa.int/handbooks/meris/CNTR2-7.htm

Franklin, R., & Mills, A. (2007). *The spatial distribution of microbes in the environment.* Springer. Available from http://books.google.com/books?id=RG6i_CjYkZQC

Goldsmith, J., Blair, F., Bisson, S., & Turner, D. (1998). Turn-key Raman lidar for profiling atmospheric water vapor, clouds, and aerosols. *Appl. Opt*, *37*, 4979–4990. Available from http://www.opticsinfobase.org/abstract.cfm?URI=ao-37-21-4979

Haan, S. de. (2008). *Meteorological applications of a surface network of global positioning system receivers.* PhD thesis, Wageningen University, Wageningen. Available from http://www.knmi.nl/publications/fulltexts/thesis_sdehaan.pdf

Haase, J., Ge, M., Vedel, H., & Calais, E. (2003). Accuracy and variability of gps tropospheric delay measurements of water vapor in the western mediterranean. *Journal of Applied Meteorology*, *42*(11), 1547-1568. Available from http://journals.ametsoc.org/doi/abs/10.1175/1520-0450%282003%29042%3C1547%3AAAVOGT%3E2.0.CO%3B2

Hanssen, R. F., Ferretti, A., Grebenitcharsky, R., Kleyer, F., Elawar, A., & Bianchi, M. (2006). Atmospheric phase screen (APS) estimation and modeling for radar interferometry. In *Fourth international workshop on ERS/Envisat SAR interferometry, 'FRINGE05'* (p. 6 pp.). Frascati, Italy, 28 Nov-2 Dec 2005. Available from http://earth.esa.int/workshops/fringe2005/participants/727/pres_727_Hanssen.pdf

Hanssen, R. F., Weckwerth, T. M., Zebker, H. A., & Klees, R. (1999). High-resolution water vapor mapping from interferometric radar measurements. *Science*, *283*(5406), 1297-1299. Available from http://www.sciencemag.org/content/283/5406/1297.abstract

Kerwin, W., & Prince, J. (1999). The kriging update model and recursive space-time function estimation. *Signal Processing, IEEE Transactions on*, *47*(11), 2942 -2952.

Kiehl, J. T., & Trenberth, K. E. (1997). Earth's Annual Global Mean Energy Budget. *Bulletin of the American Meteorological Society*, *78*, 197-197. Available from http://journals.ametsoc.org/doi/abs/10.1175/1520-0477(1997)078%3C0197:EAGMEB%3E2.0.CO;2

Kleinert, T. (2010). *Eucos operational programme - synoptic surface land station network update.* website. Available from http://www.eucos.net/cln_007/nn_133514/EN/Home/networks/surface/EUCOS__update__surface__2010,templateId=raw,property=publicationFile.pdf/EUCOS_update_surface_2010.pdf (EUCOS/TSC/2010-001)

KNMI. (2010a). *Weather prediction 30-05-2010.* website. Available from http://www.meteoactueel.nl/cms/index.php?option=com_content&view=article&id=2702%3Aweersverwachting-30-05-2010&catid=107%3Aweersverwachtingen&Itemid=487 (last visited November 2011)

KNMI. (2010b). Available from www.knmi.nl

Krige, D., Guarascio, M., & Camisani-Calzolari, F. (1989). Early South African geostatistical techniques in today's perspective. *M. Armstrong (ed.), Geostatistics, volume I (Kluwer, Dordrecht)*, 1–19.

Kyriakidis, P. C., & Journel, A. G. (1999). Geostatistical Space–Time Models: A Review. *Mathematical Geology*, *31*(6), 651-684. Available from http://www.springerlink.com/content/h514k7v760045748/

Li, D., Shan, J., & Gong, J. (2009). *Geospatial technology for earth observation* (1st ed.). Springer Publishing Company, Incorporated.

EUMETSAT. (2010, July). *Iasi.* website. Available from http://www.eumetsat.int/Home/Main/Satellites/Metop/Instruments/SP_2010053151047495?l=en

MERIS Quality Working Group. (2006, March). *MERIS products quality status report.* website. Available from http://earth.eo.esa.int/pcs/envisat/meris/documentation/ (MEGS7.4 and IPF 5)

Miller, M. J. (1984). Numerical prediction and dynamic meteorology (2nd edition). g. j. haltiner and r. t. williams. j. wiley & sons ltd. 1983. i-xiii & pp. 477. *Quarterly Journal of the Royal Meteorological Society*, *110*(463), 280–280. Available from http://dx.doi.org/10.1002/qj.49711046321

Niell, A. E. (1996). Global mapping functions for the atmosphere delay at radio wavelengths. *J. Geophys. Res.*, *101*(B2), 3227–3246. Available from http://dx.doi.org/10.1029/95JB03048

Rose, T., Crewell, S., Löhnert, U., & Simmer, C. (2005). A network suitable microwave radiometer for operational monitoring of the cloudy atmosphere. *Atmospheric Research*, *75*(3), 183-200. Available from http://www.scopus.com/inward/record.url?eid=2-s2.0-19144370595&partnerID=40&md5=bf90a3b520b49122259f1695fe84c425

Rose, T., & Czekala, H. (2009, February). RPG-HATPRO, RPG-TEMPRO, RPG-HUMPRO (7.70 ed.) [Computer software manual]. 53340 Meckenheim, Germany. Available from http://www.radiometer-physics.de/docs/Manual_profiler.pdf (Operating Manual)

Saastamoinen, J. (1972). Atmospheric correction for the troposphere and stratosphere in radio ranging of satellites. *Geophys. Monogr. Series*, *15*, 247-521.

Schroeder, D., & Pribram, J. (1999). An introduction to thermal physics. *American Journal of Physics*, *67*, 1284.

Standard Atmosphere, U. S. C. on Extension to the, Oceanic, U. S. N., Administration, A., Aeronautics, U. S. N., Administration, S., & Force, U. S. A. (1976). *U.S. standard*

atmosphere, 1976. National Oceanic and Amospheric [sic] Administration : for sale by the Supt. of Docs., U.S. Govt. Print. Off. Available from http://books.google.nl/books?id=5WdTAAAAMAAJ

Stone, M. (1974). Cross-validation choice and assessment of statistical predictions (with discussion). *Journal of the Royal Statistical Society B*(36), 111-147.

Teunissen, P. (2003). *Adjustment theory: an introduction*. Delft Univ. Press. Available from http://books.google.nl/books?id=rcj7ewEACAAJ

The EUMETSAT Network of Satellite Application Facilities. (2010, July). *GRM-11: Offline specific humidity profile.* website. Available from http://www.grassaf.org/GRM-11.php

University, M. (2007). *Understanding the process of using observations to initialize weather prediction models.* Work sheet. Available from http://www.atmos.millersville.edu/~lead/Obs_to_Model_Home.html (Last visited 2011)

Vaisala. (2006). *Vaisala radiosonde RS92 brochure.* Helsinki, Finland. Available from www.vaisala.com

Wackernagel, H. (2003). *Multivariate geostatistics: an introduction with applications*. Springer. Available from http://books.google.com/books?id=Rhr7bgLWxx4C

Webster, R., & Oliver, M. A. (2007). *Geostatistics for environmental scientists* (Vol. 2; S. Senn, M. Scott, & V. Barnett, Eds.). John Wiley & Sons. Ltd. Available from http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0470028580.html

Wieringa, J., & Rijkoort, P. J. (1983). *Windklimaat van nederland*. Staatsuitgeverij. Available from www.knmi.nl/kd/achtergrondinformatie/Windklimaat_van_Nederland.pdf

Wikipedia. (2009). *Constant altitude plan position indicator — wikipedia, the free encyclopedia.* Available from http://en.wikipedia.org/w/index.php?title=Constant_Altitude_Plan_Position_Indicator&oldid=296496943 (Online; accessed 15-April-2010)

Appendix A

Validation with Ground Stations Removed from Measurements

On the following pages the IWV and IWV differences for the five validation stations are shown in more detail. Also the Mean Error and Root Mean Squared Error for the validation stations are shown. Finally the data which is used to obtain the figures of the Mean Error and Root Mean Squared Error is shown in a table.

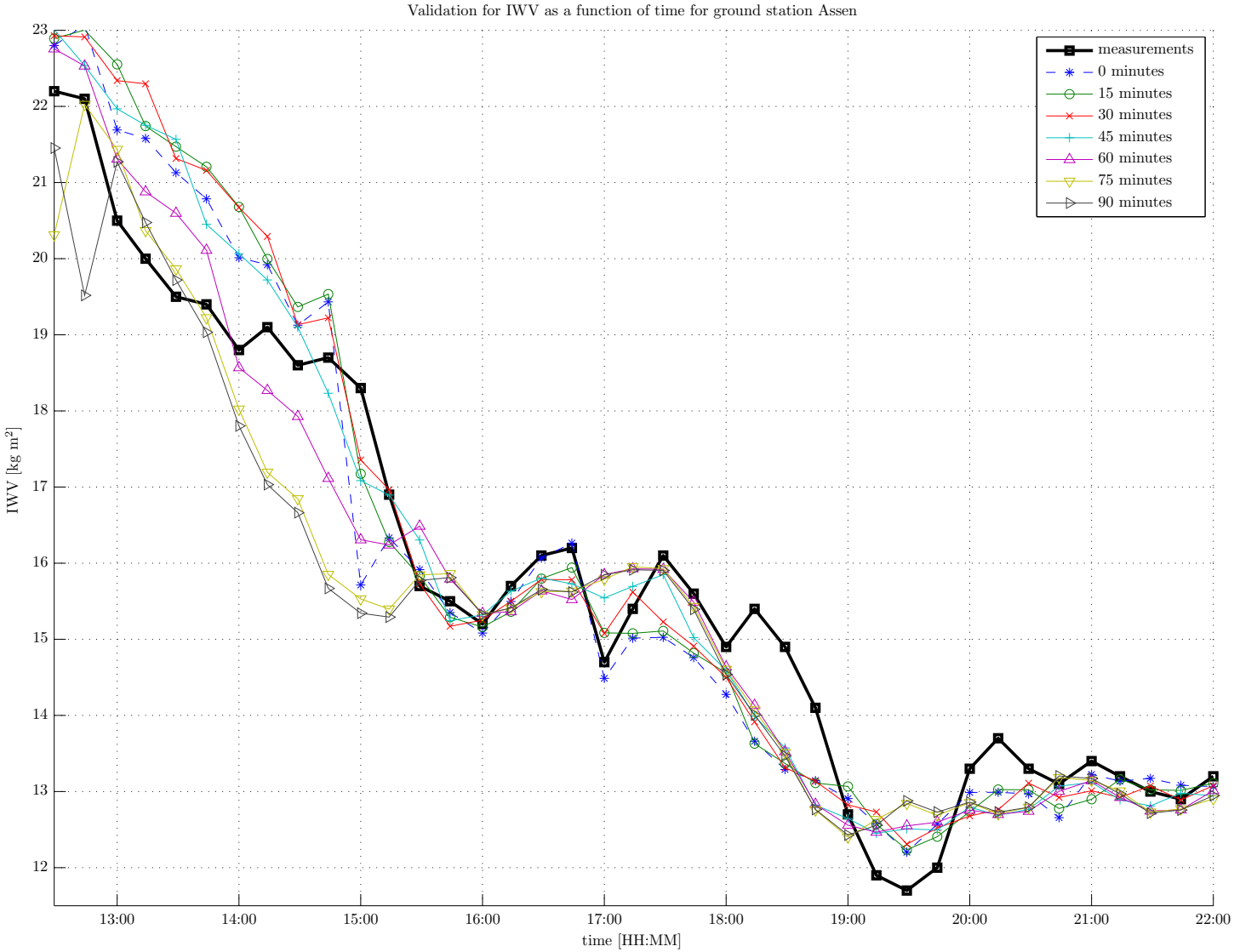


Figure A-1: IWV as a function of time for the location of GPS ground station Assen (ASS2), the black squares are the measurements at the ground station, the colored lines are the IWV values as predicted by the different Kriging interpolations using only current measurements (0 minutes, blue line) and using 1–6 advected measurements from 15–90 minutes from the past.

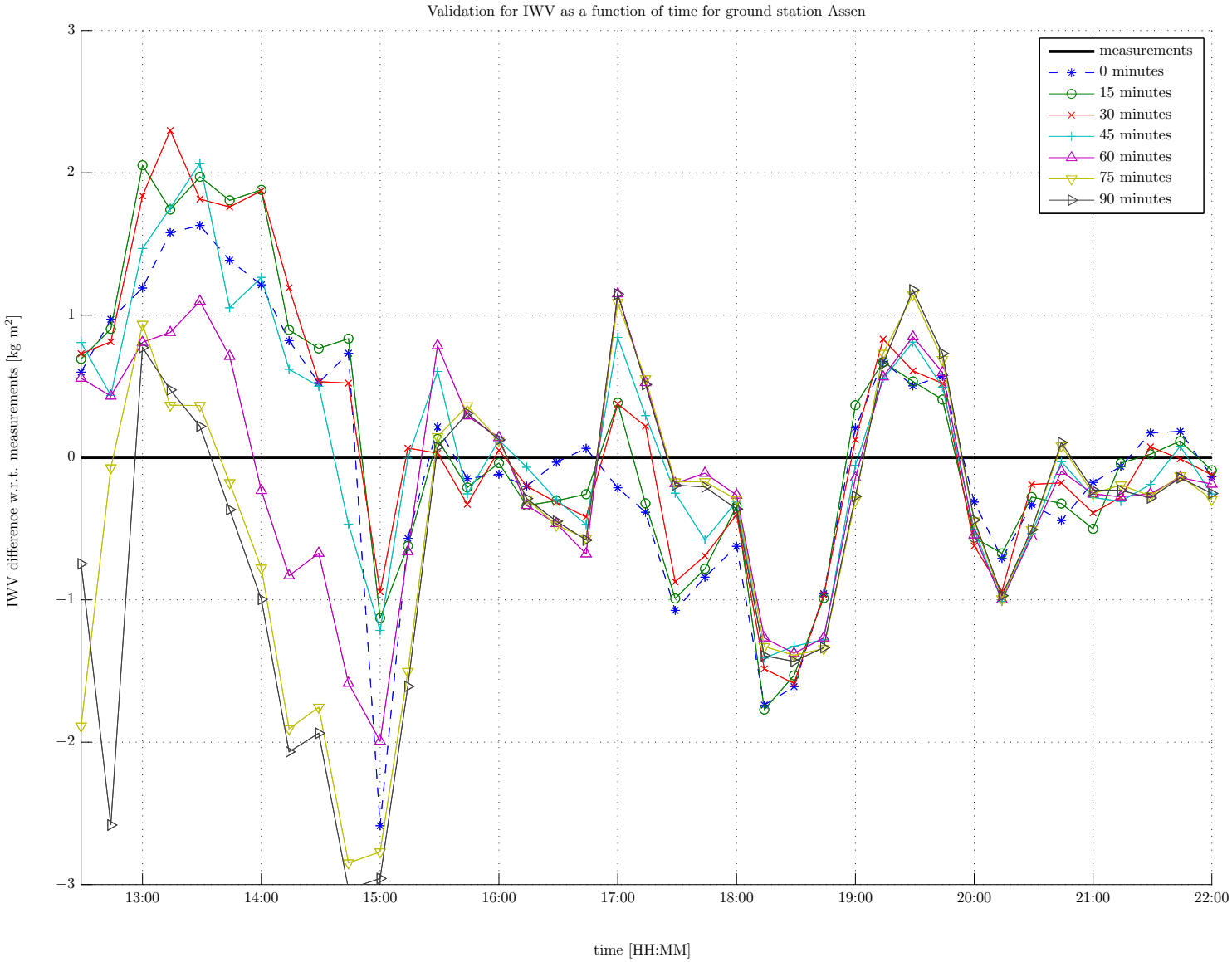


Figure A-2: IWV difference with respect to original measurements as a function of time for the location of GPS ground station Assen (ASS2), the colored lines are the IWV difference with respect to the GPS measurements as predicted by the different Kriging interpolations using only current measurements (0 minutes, blue line) and using 1–6 advected measurements from 15–90 minutes from the past.

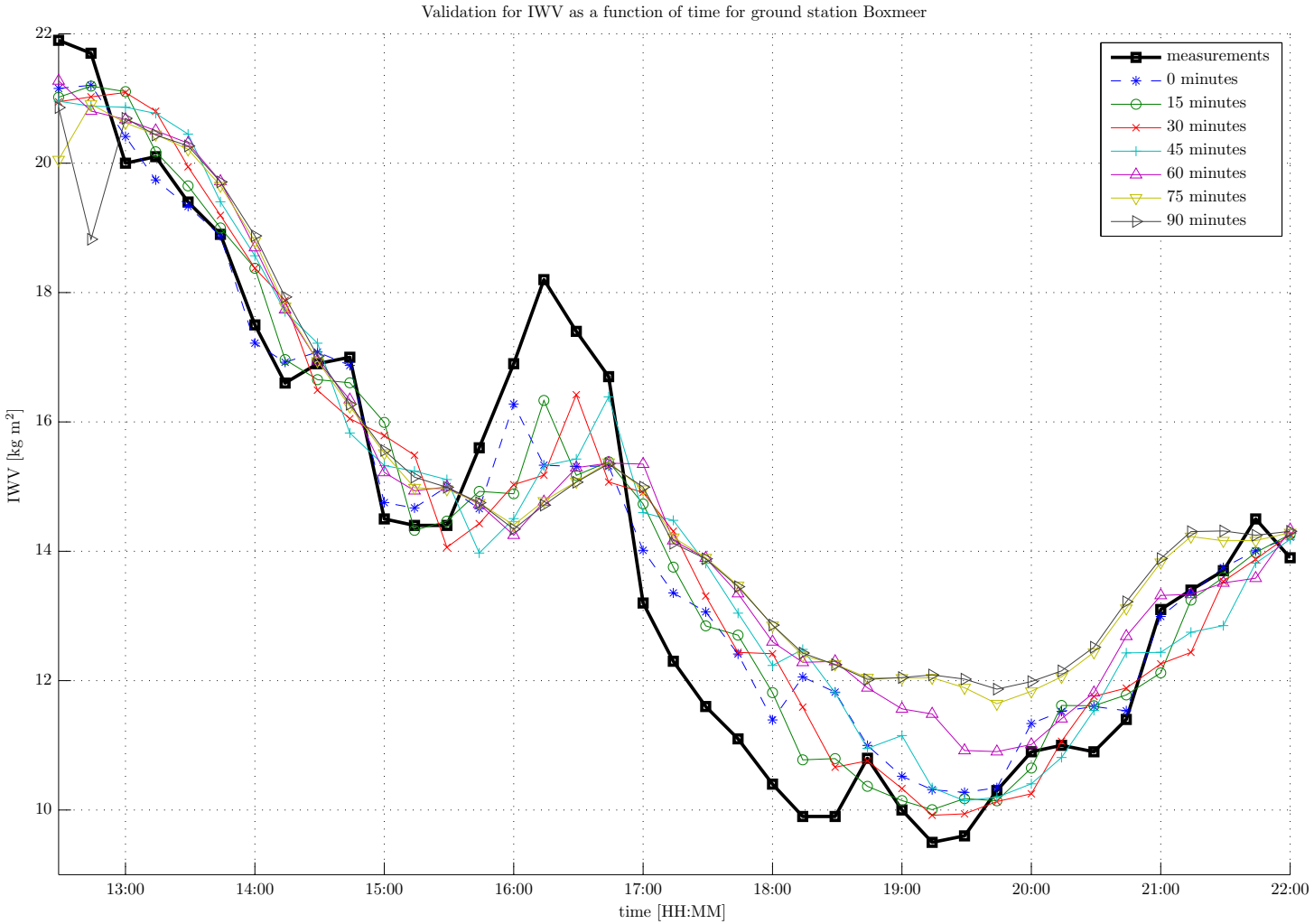


Figure A-3: IWV as a function of time for the location of GPS ground station Boxmeer (BOXM), the black squares are the measurements at the ground station, the colored lines are the IWV values as predicted by the different Kriging interpolations using only current measurements (0 minutes, blue line) and using 1–6 advected measurements from 15–90 minutes from the past.

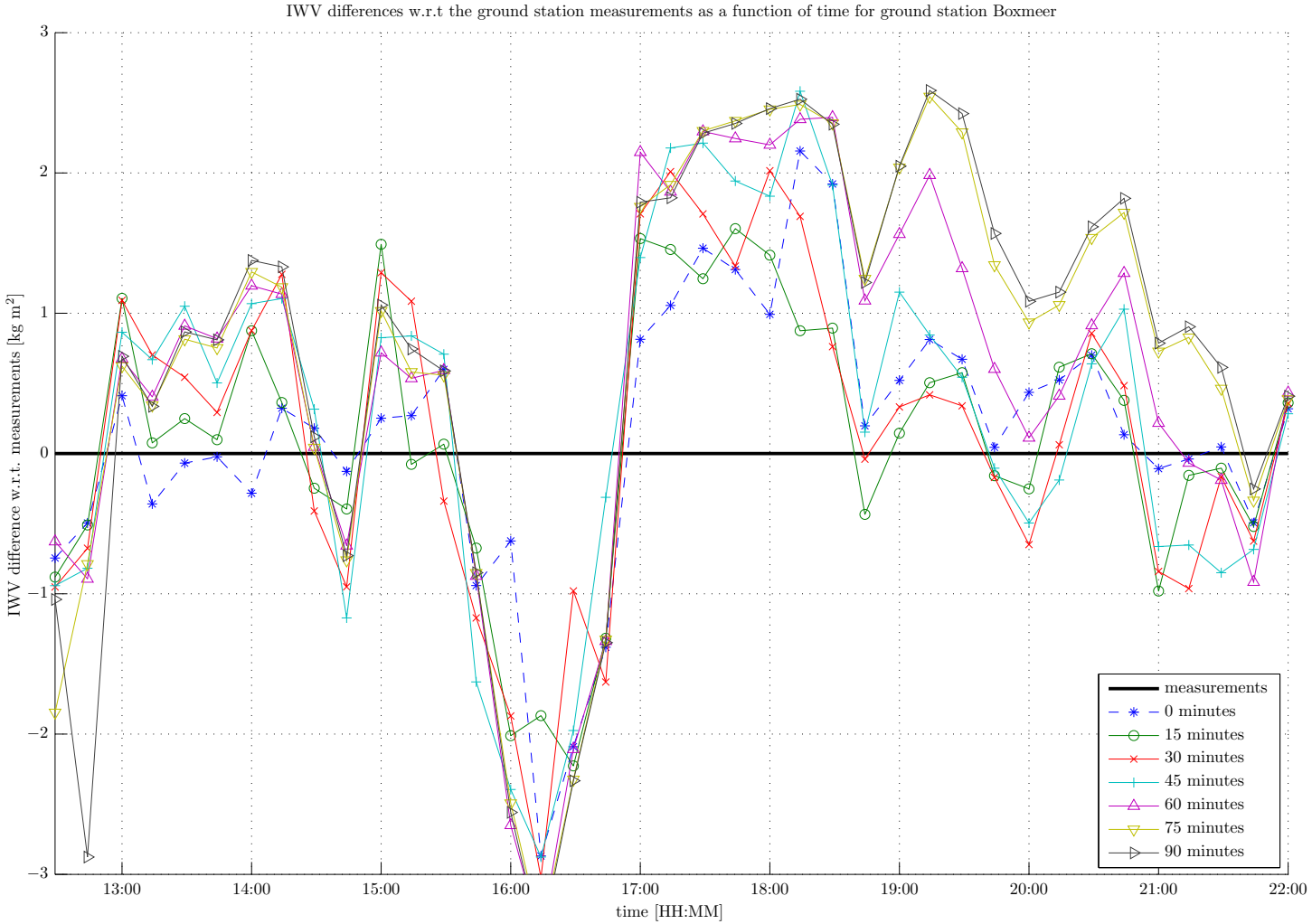


Figure A-4: IWV difference with respect to original measurements as a function of time for the location of GPS ground station Boxmeer (BOXM), the colored lines are the IWV difference with respect to the GPS measurements as predicted by the different Kriging interpolations using only current measurements (0 minutes, blue line) and using 1–6 advected measurements from 15–90 minutes from the past.

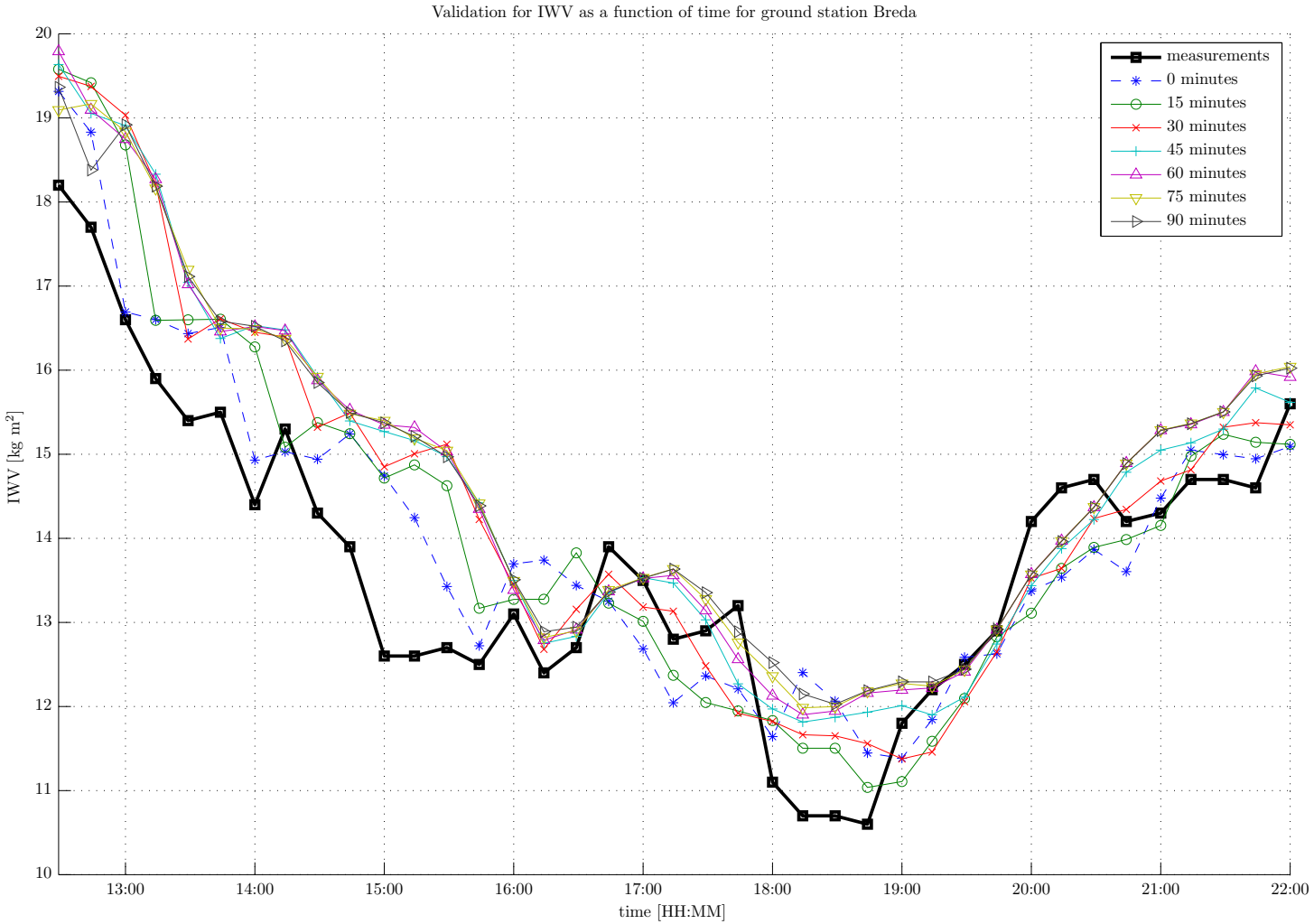


Figure A-5: IWV as a function of time for the location of GPS ground station Breda (BRD2), the black squares are the measurements at the ground station, the colored lines are the IWV values as predicted by the different Kriging interpolations using only current measurements (0 minutes, blue line) and using 1–6 advected measurements from 15–90 minutes from the past.

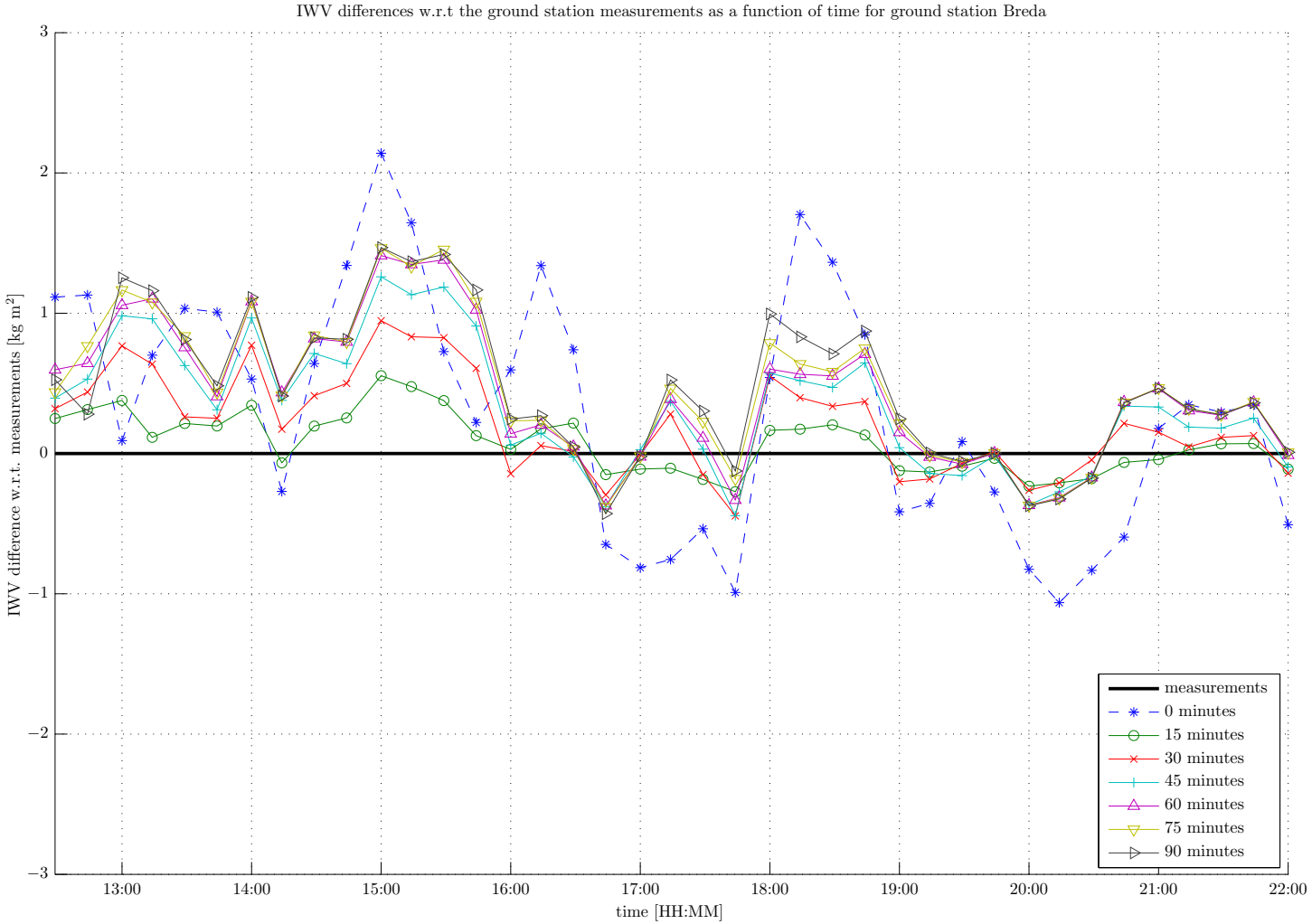


Figure A-6: IWV difference with respect to original measurements as a function of time for the location of GPS ground station Breda (BRD2), the colored lines are the IWV difference with respect to the GPS measurements as predicted by the different Kriging interpolations using only current measurements (0 minutes, blue line) and using 1–6 advected measurements from 15–90 minutes from the past.

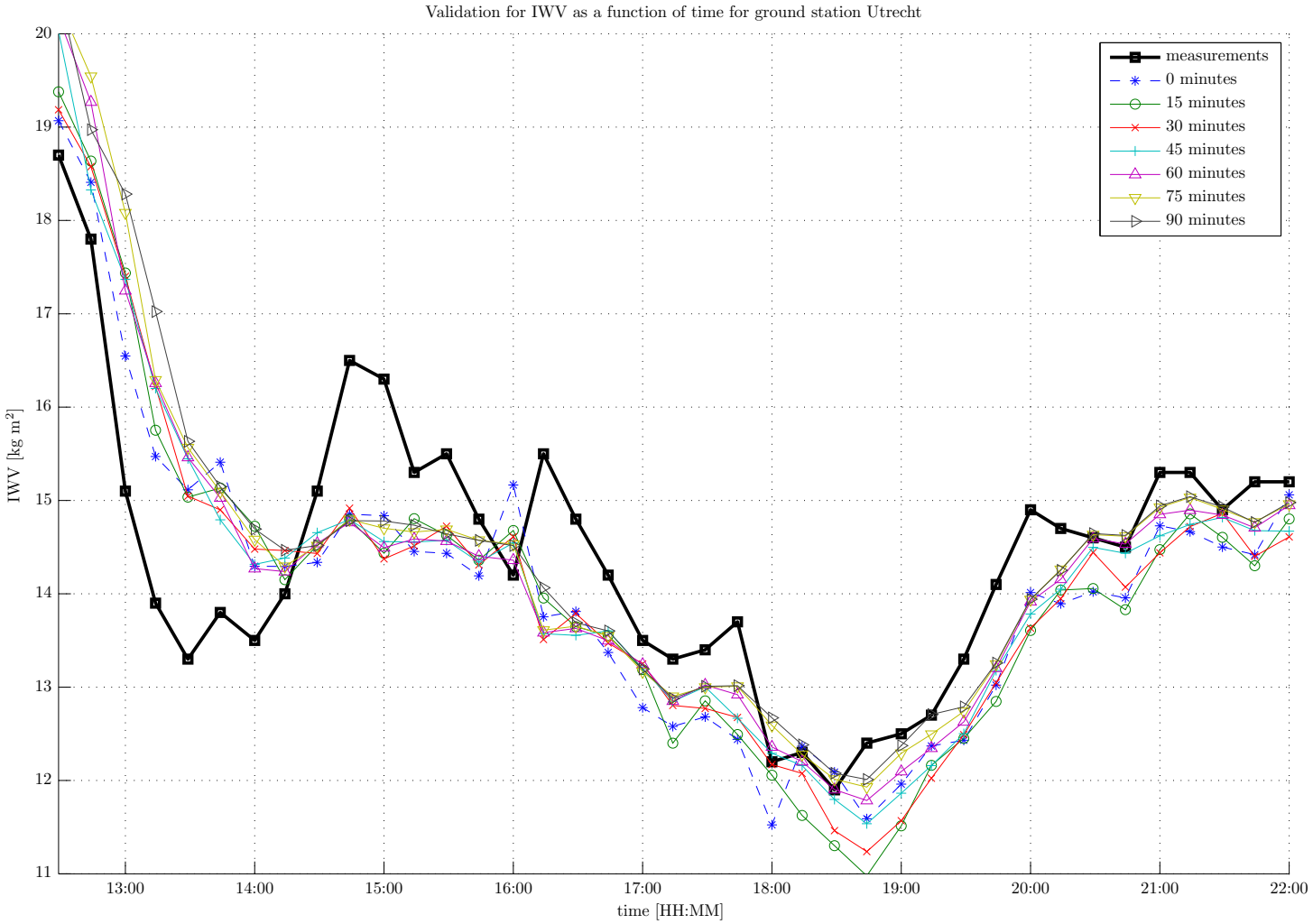


Figure A-7: IWV as a function of time for the location of GPS ground station Utrecht (UTR2), the black squares are the measurements at the ground station, the colored lines are the IWV values as predicted by the different Kriging interpolations using only current measurements (0 minutes, blue line) and using 1–6 advected measurements from 15–90 minutes from the past.

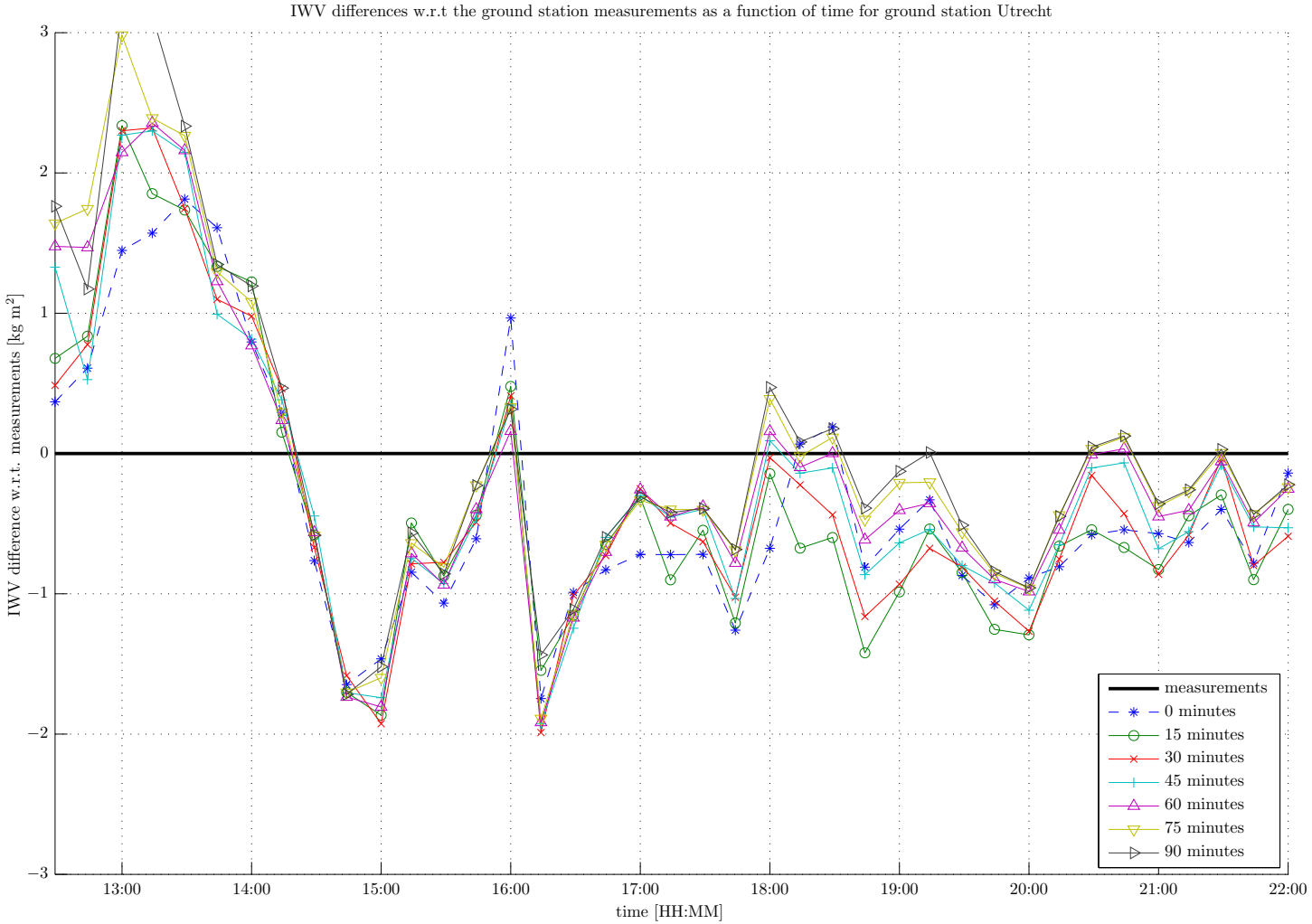


Figure A-8: IWV difference with respect to original measurements as a function of time for the location of GPS ground station Utrecht (UTR2), the colored lines are the IWV difference with respect to the GPS measurements as predicted by the different Kriging interpolations using only current measurements (0 minutes, blue line) and using 1–6 advected measurements from 15–90 minutes from the past.

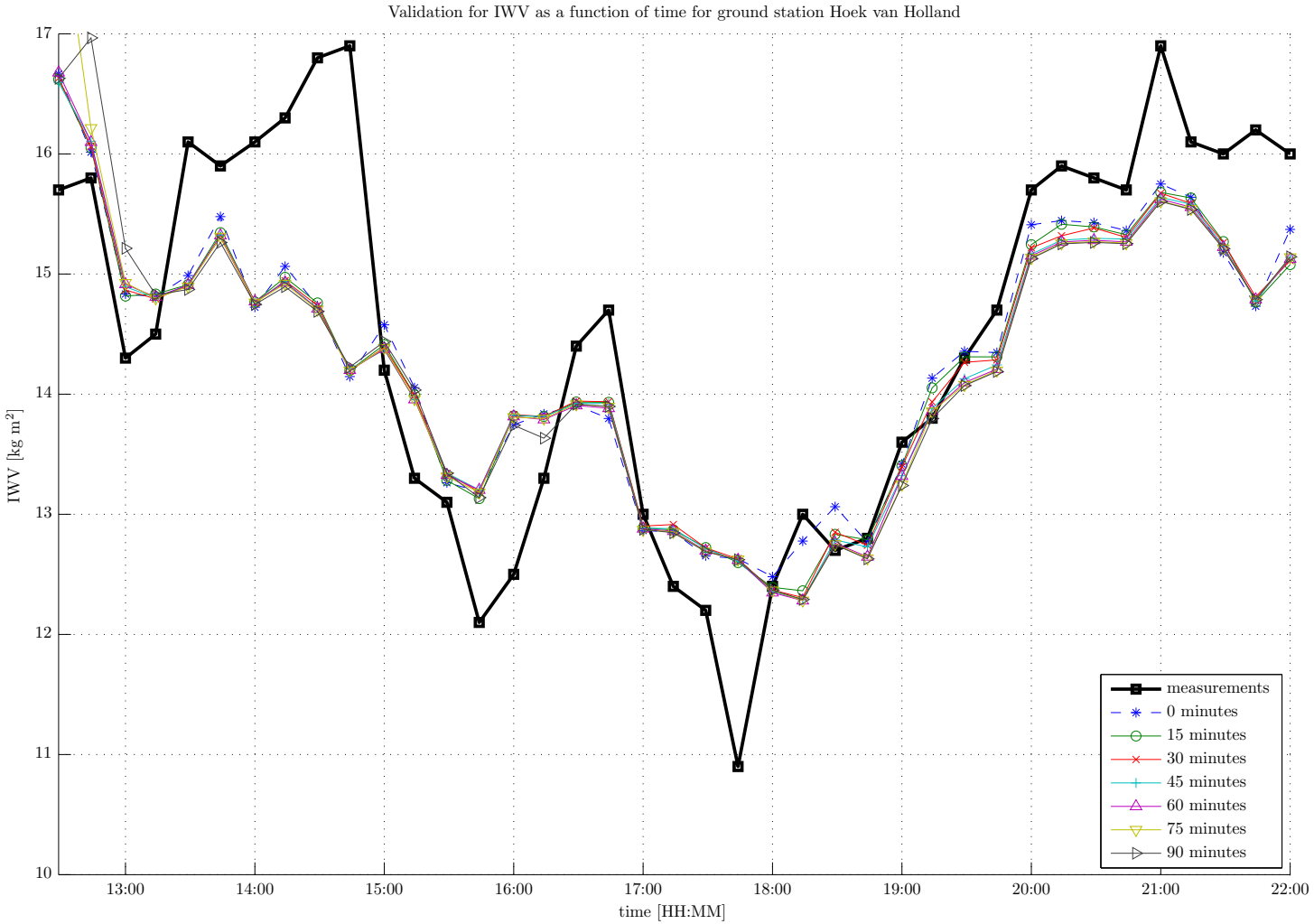


Figure A-9: IWV as a function of time for the location of GPS ground station Hoek van Holland (HHOL), the black squares are the measurements at the ground station, the colored lines are the IWV values as predicted by the different Kriging interpolations using only current measurements (0 minutes, blue line) and using 1–6 advected measurements from 15–90 minutes from the past.

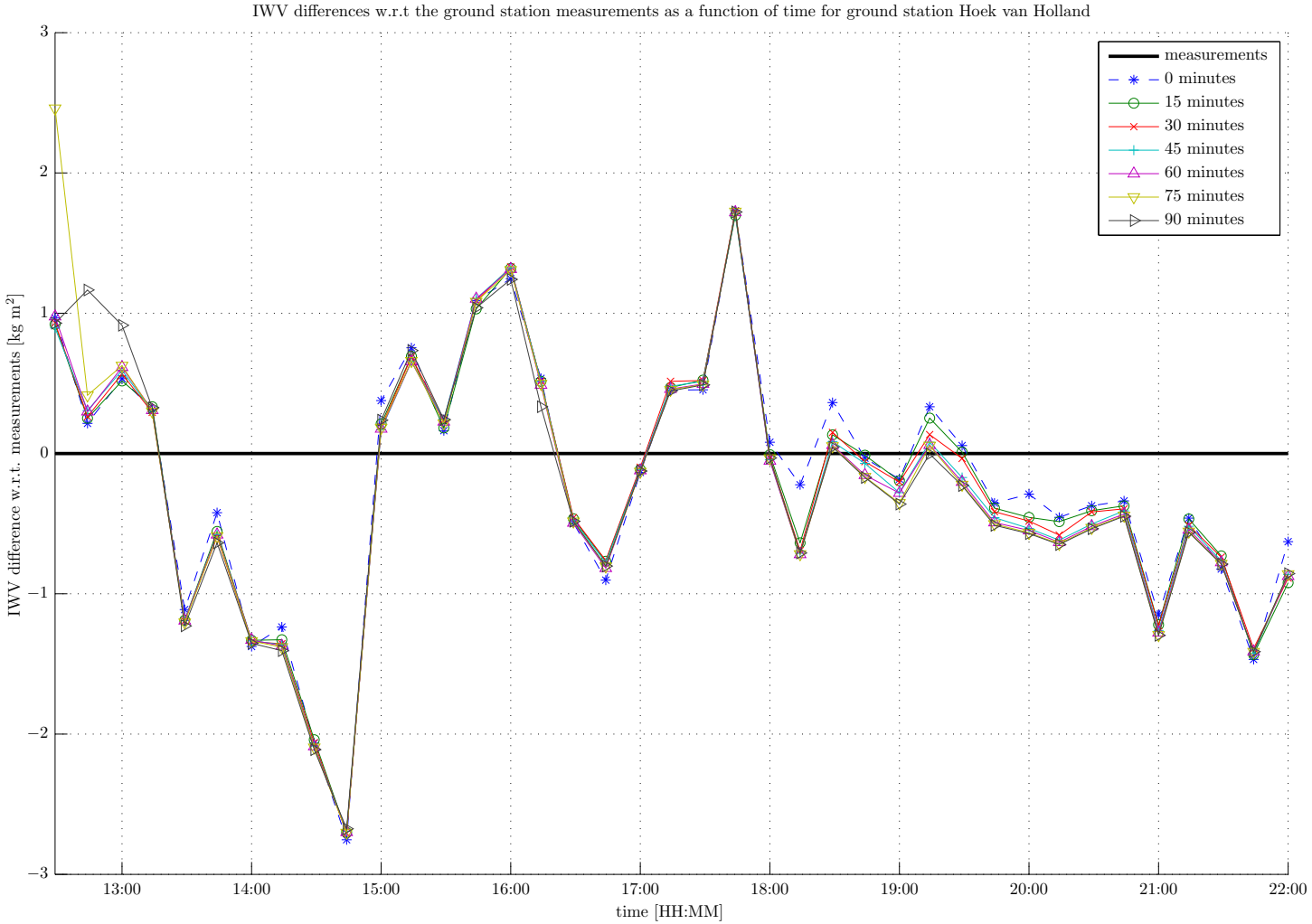


Figure A-10: IWV difference with respect to original measurements as a function of time for the location of GPS ground station Hoek van Holland (HHOL), the colored lines are the IWV difference with respect to the GPS measurements as predicted by the different Kriging interpolations using only current measurements (0 minutes, blue line) and using 1–6 advected measurements from 15–90 minutes from the past.

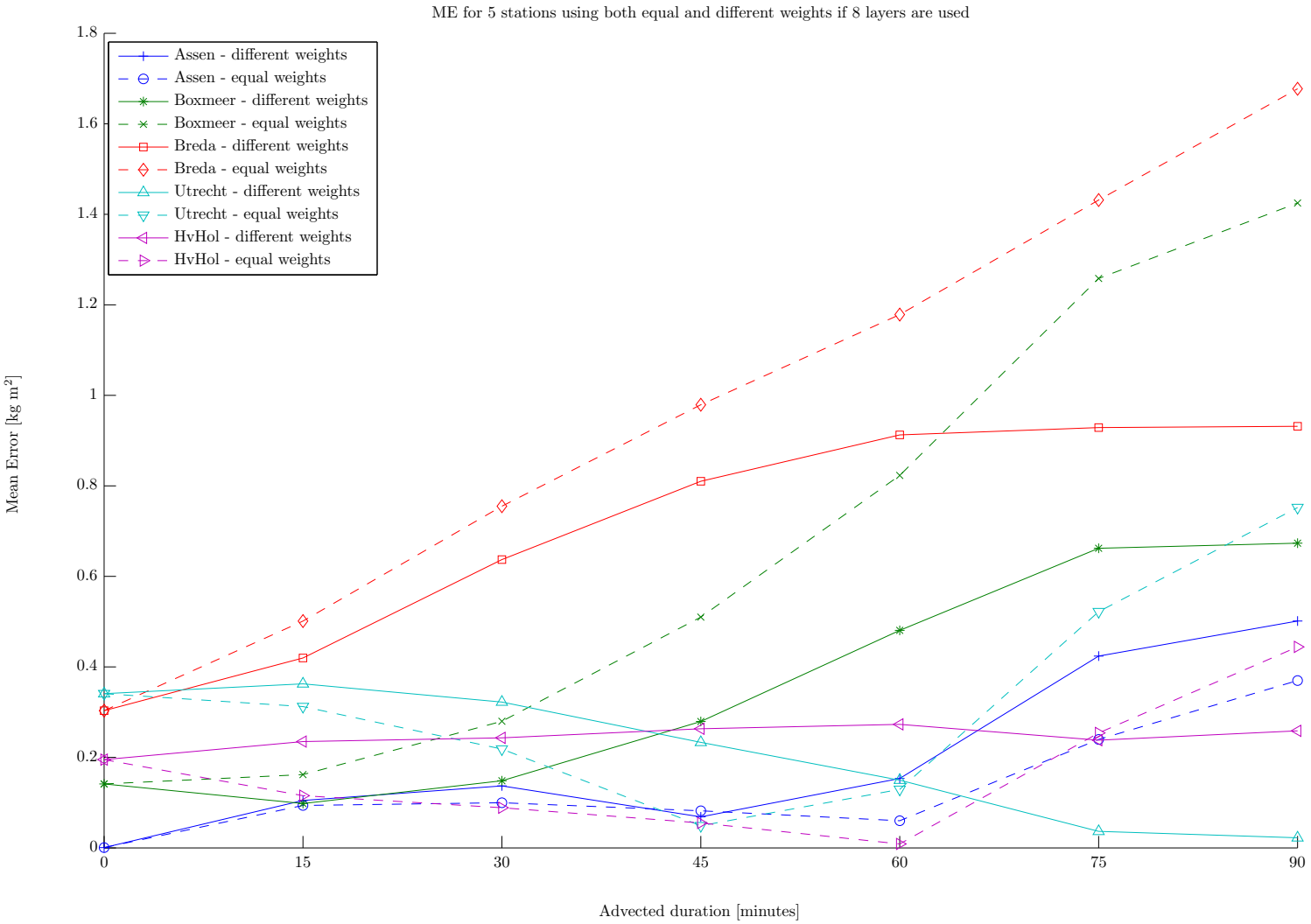


Figure A-11: Absolute value of the Mean Error as a function of advected duration for 5 stations using equal weights for current and advected measurements (dashed lines) and using lower weights for advected measurements (solid lines), the atmosphere is divided into 8 layers.

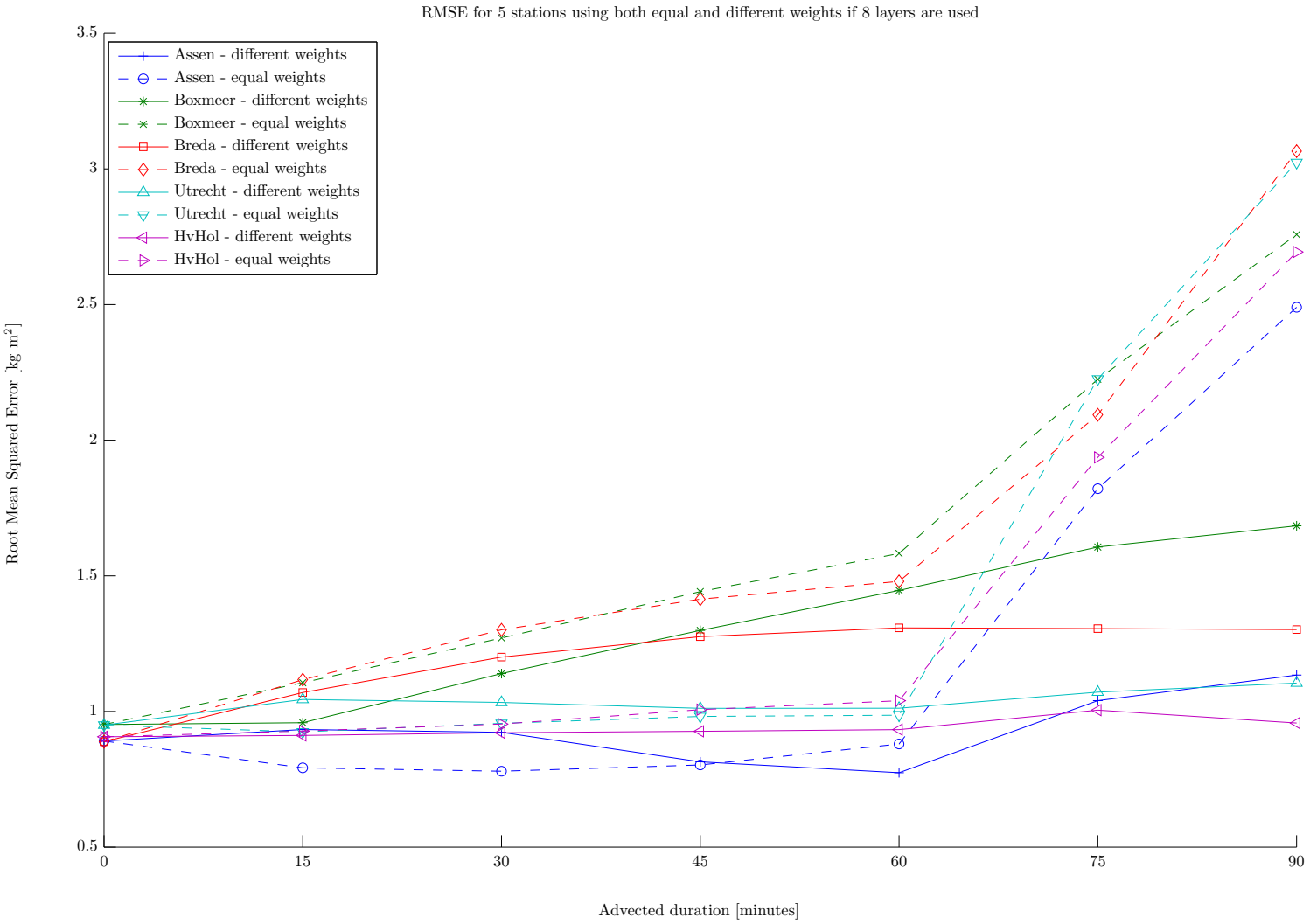


Figure A-12: Root Mean Squared Error as a function of advected duration for 5 stations using equal weights for current and advected measurements (dashed lines) and using lower weights for advected measurements (solid lines), the atmosphere is divided into 8 layers.

Table A-1: Mean Error (ME) and Root Mean Squared Error (RMSE) of the predictions with respect tot the measurements for Assen, Boxmeer, and Breda by using only current measurements (0 minutes advection) and using current and 15 up to 90 minutes advected measurements.

Advection	0	15	30	45	60	75	90
Assen:							
8l, difW							
ME	-0.000676	0.105	0.137	0.0686	-0.153	-0.424	-0.501
RMSE	0.891	0.934	0.923	0.814	0.774	1.04	1.13
8l, sameW							
ME	-0.000676	0.0938	0.0998	0.0819	0.0600	0.240	0.370
RMSE	0.891	0.792	0.780	0.803	0.880	1.82	2.49
1l, difW							
ME	-0.000678	0.0656	0.0440	-0.0514	-0.439	-0.680	-0.720
RMSE	0.891	1.01	1.01	0.966	0.977	1.14	1.15
1l, sameW							
ME	-0.000676	1.14	1.47	1.87	2.79	2.82	3.05
RMSE	0.891	3.05	2.98	3.72	4.36	4.58	5.06
Boxmeer:							
8l, difW							
ME	0.141	0.0981	0.148	0.279	0.481	0.662	0.673
RMSE	0.952	0.959	1.14	1.30	1.45	1.61	1.68
8l, sameW							
ME	0.141	0.162	0.280	0.510	0.823	1.26	1.43
RMSE	0.952	1.11	1.27	1.44	1.58	2.22	2.76
1l, difW							
ME	0.141	0.156	0.227	0.346	0.361	0.561	0.578
RMSE	0.952	1.03	1.23	1.39	1.64	1.72	1.75
1l, sameW							
ME	0.141	1.25	1.64	2.21	3.41	3.61	3.91
RMSE	0.952	3.46	3.64	4.47	4.82	4.84	5.23
Breda:							
8l, difW							
ME	0.303	0.420	0.637	0.810	0.913	0.929	0.932
RMSE	0.888	1.07	1.20	1.28	1.31	1.31	1.30
8l, sameW							
ME	0.303	0.501	0.755	0.979	1.18	1.43	1.68
RMSE	0.888	1.12	1.30	1.41	1.48	2.09	3.07
1l, difW							
ME	0.303	0.483	0.733	0.904	1.01	1.08	1.10
RMSE	0.889	1.12	1.27	1.37	1.41	1.47	1.47
1l, sameW							
ME	0.303	1.42	1.78	2.22	3.84	4.07	4.39
RMSE	0.889	3.06	3.19	3.32	5.05	5.10	5.54

Table A-1: Mean Error (ME) and Root Mean Squared Error (RMSE) of the predictions with respect tot the measurements for Utrecht and Hoek van Holland by using only current measurements (0 minutes advection) and using current and 15 up to 90 minutes advected measurements.

Advection	0	15	30	45	60	75	90
Utrecht:							
8l, difW							
ME	-0.341	-0.363	-0.322	-0.233	-0.150	-0.0366	0.0223
RMSE	0.950	1.04	1.03	1.01	1.01	1.07	1.10
8l, sameW							
ME	-0.341	-0.312	-0.218	-0.0489	0.129	0.522	0.752
RMSE	0.950	0.925	0.956	0.982	0.986	2.23	3.02
1l, difW							
ME	-0.341	-0.349	-0.393	-0.405	-0.409	-0.411	-0.423
RMSE	0.950	1.04	1.07	1.06	1.05	1.064	1.07
1l, sameW							
ME	-0.341	0.628	0.973	1.49	2.36	2.39	2.65
RMSE	0.950	2.65	2.96	3.99	3.85	3.83	4.21
Hoek v. Holland:							
8l, difW							
ME	-0.195	-0.235	-0.243	-0.263	-0.273	-0.238	-0.258
RMSE	0.907	0.912	0.922	0.927	0.933	1.00	0.957
8l, sameW							
ME	-0.195	-0.115	-0.0888	-0.0551	0.00871	0.254	0.444
RMSE	0.907	0.928	0.954	1.01	1.04	1.94	2.69
1l, difW							
ME	-0.195	-0.222	-0.226	-0.241	-0.257	-0.254	-0.256
RMSE	0.907	0.907	0.921	0.918	0.942	0.945	0.945
1l, sameW							
ME	-0.195	0.783	1.12	1.62	2.56	2.67	2.99
RMSE	0.907	2.62	2.99	3.78	4.34	4.44	4.87