UNIVERSITY OF TWENTE.

MODEL ASSIMILATION IN RIVER FLOW DATA PROCESSING

ESTIMATING DOMINANT SPATIAL SCALES IN CROSS-SECTIONAL VELOCITY DATA

R.T.J. VAN DONGEN

MARCH 2018

Additionally to the title and other information, the course of the meandering Mahakam River is presented on the frontpage and in particular the region around the measured river bend. Moreover, the coarse steps taken during the study are visualised. On the left a few measured velocity locations with an Acoustic Doppler Current Profiler are shown. In the middle, a model fit is shown conceptually, that tries to approximate measured velocities. On the right, an example of dominant scales, *i.e.*, wave forms, extracted from the model fit is shown.

MODEL ASSIMILATION IN RIVER FLOW DATA PROCESSING

ESTIMATING DOMINANT SPATIAL SCALES IN CROSS-SECTIONAL VELOCITY DATA

Author	R.T.J. (Rick) van Dongen
Graduation committee	Dr. Ir. P.C. Roos
	Associate Professor
	University of Twente
	Water Engineering and Management
	7500 AE Enschede, The Netherlands
	p.c.roos@utwente.nl
	Dr. Ir. B. Vermeulen
	Assistant Professor
	University of Twente
	Water Engineering and Management
	7500 AE Enschede, The Netherlands
	b.vermeulen@utwente.nl

Preface

This thesis is the final part of my study at the University of Twente, where I have done a masters in Civil Engineering and Management with specialisation in water. In this thesis, dominant spatial scales in cross-sectionally velocity data of a river are investigated with the help of Fourier transform. Furthermore, the extent to which the model complexity must increase to represent the velocity field adequately is studied. I have encountered some problems during the preparation of the thesis, but the support of my supervisors made sure that I can continue and have kept the focus.

I would like to thank Bart for the insightful sessions we have had and the support during the project that made me progress. We have met each other every week to discuss the model, the progress or other stuff, but you always made me feel welcome. I would also like to thank Pieter, for your enthusiasm during meetings, and for your view and ideas about the topic. Furthermore, I really liked the soccer games during lunch breaks with SouthWEMton. You have introduced me to the team and we have even played a match together.

In addition, I would like to thank Jon, Jurre and Anne-Lot for their feedback on the report, it was really helpful for me in improving my writing skills. Furthermore, your questions after reading gets me thinking.

Finally, I would like to thank my family, girlfriend, roommates and fellow students for their support the last couple of months.

Rick van Dongen

Enschede, 30 March, 2018

Abstract

Flow velocity measurements in rivers and coastal areas are increasingly carried out with Acoustic Doppler Current Profilers (ADCPs). This instrument collects velocity data at multiple locations in a cross-section over a certain period of time. Post-processing techniques of raw velocity data often involve temporal averaging and spatial smoothing. Smoothing and averaging windows are often chosen arbitrarily without a clear substantiation. The aim of this study is to identify dominant spatial patterns in river's cross-sectional velocity data in order to average and smooth the data with more certainty.

Spatial patterns are investigated by a method based on spectral analysis, which allows to identify the dominant scales. Higher order functions are progressively included to a base function for each velocity component, *i.e.,* increasing the truncation number, and fitted to the measured velocity data. This process is repeated until the residuals have no spatial structure. Velocity locations are transformed into a normalised coordinate system in order to conduct the model fit for multiple cross-sections. The method is applied to velocity measurements collected in a sharp river bend.

By analysing the available data set, weak flow is observed near the boundaries and strong flow in the center or slightly outwards of the center, but near the scour hole the flow recirculates at the outer sides causing an upstream flow locally. Water flows to the outer bend at the water surface and to the inner bend near the river bed. Dominant spatial scales can be observed from the computed amplitudes with Fourier transform, but the strongest amplitudes vary for different truncation scales. Most of the spatial structure in the residuals disappears after truncating at eight waveforms over local river width horizontally and almost eight waveforms over local water depth vertically. The method is particularly useful to represent the main flow pattern adequately with continuous functions, due to a relatively steady state during data collection.

Fourier transform are applied generally to analyse the presence of dominant modes in river flow data, but only with respect to time. Investigating the cross-sectional spatial velocity distribution with the help of Fourier transform provides insights on the dominant spatial scales, which are, however, linked to the created model and the applied set of base functions for the three velocity components. Further research is recommended on validating the model by using different instruments to measure velocity data and on including different of base functions.

Samenvatting

Stroomsnelheidsmetingen in rivieren en kustgebieden worden steeds vaker uitgevoerd met Acoustic Doppler Current Profilers (ADCP's). Dit instrument verzamelt snelheidsgegevens op meerdere locaties in een doorsnede over een bepaalde periode. Nabewerkingsmethoden van ruwe snelheidsgegevens omvatten vaak temporele middeling en het ruimtelijk glad strijken van gegevens. Afronding- en middelingsvensters worden vaak willekeurig gekozen zonder duidelijke onderbouwing. Het doel van deze studie is om dominante ruimtelijke patronen in de snelheidsgegevens van een dwarsdoorsnede in een rivier te identificeren, zodat de gegevens in de toekomst met meer zekerheid gemiddeld kunnen worden.

De ruimtelijke patronen worden onderzocht met een methode die gebaseerd is op spectrale analyse. Hiermee kunnen de dominante schalen worden geïdentificeerd. Hogere orde functies worden supergepositioneerd in een basisfunctie voor elke snelheidscomponent (met andere woorden: het truncatiegetal wordt verhoogd) en afgestemd op de gemeten snelheidsgegevens. Dit proces wordt herhaald totdat de residuen geen ruimtelijke structuur meer bevatten. Locaties waar snelheden zijn gemeten worden getransformeerd naar een genormaliseerd coördinatensysteem om het model geschikt te maken voor meerdere doorsneden. De methode wordt toegepast op snelheidsmetingen in een scherpe rivierbocht.

In de beschikbare data is een zwakke stroming is waargenomen dichtbij de randen en een sterke stroming in het middenboven of net iets naar de buitenbocht gelegen. Ter plaatse van de ontgrondingskuil circuleert de stroom aan de buitenzijden, waardoor het lokaal een stroomopwaartse stroming veroorzaakt. Water stroomt richting de buitenbocht aan het wateroppervlak en richting de binnenbocht over de rivierbodem. Dominante ruimtelijke schalen kunnen worden waargenomen met behulp van de berekende amplituden in de Fouriertransformatie, maar de sterkste amplitudes variëren voor verschillende truncatiegetallen. Het grootste deel van de ruimtelijke structuur in de residuen verdwijnt na het truncaren bij acht golfvormen over de lokale rivierbreedte (horizontaal) en bijna acht golfvormen over de lokale waterdiepte (verticaal). De methode is met name te gebruiken om de hoofdstroom te representeren met behulp van continue functies, omdat de snelheden zijn gemeten tijdens een relatief constante toestand van de rivierafvoer.

Fouriertransformatie wordt in het algemeen toegepast om de aanwezigheid van dominante perioden in stroomgegevens van rivieren te analyseren, maar tot dusver enkel met betrekking tot tijd. Het onderzoeken van de ruimtelijke snelheidsverdeling in een dwarsdoorsnede met behulp van Fouriertransformatie biedt inzichten in de dominante ruimtelijke schalen, die echter zijn gekoppeld aan het gemaakte model en de toegepaste set van basisfuncties voor de drie snelheidscomponenten. Verder onderzoek wordt aanbevolen Aanbevelingen voor verder onderzoek zijn verschillende instrumenten te gebruiken om stroomsnelheden te meten zodat het model kan worden gevalideerd en om verschillende basisfuncties op te nemen in de methode.

Contents

List of symbols

Symbol	Unit	Description
x	m	eastward coordinate
y	m	northward coordinate
z	m	upward coordinate
s	m	across cross-section coordinate
n	m	along cross-section coordinate, horizontally perpendicular to s
u_x	m s^{-1}	measured raw velocity in x-direction
u_y	m s^{-1}	measured raw velocity in y-direction
u_z	m s^{-1}	measured raw velocity in z-direction
$\hat{u}_s$	m s^{-1}	components in measured raw velocity that contributes to the longitudinal velocity in s-direction
$\hat{v}_n$	m s^{-1}	components in measured raw velocity that contributes to the lateral velocity in n-direction
$\tilde{u}$	m s^{-1}	modelled velocity component in (longitudinal) s-direction
$\tilde{v}$	m s^{-1}	modelled velocity component in (lateral) n-direction
$\tilde{w}$	m s^{-1}	modelled velocity component in (vertical) z-direction
$\vec{u}_i$	m s^{-1}	vectorial velocity
$\bar{u}$	m s^{-1}	mean longitudinal velocity (lateral: $\bar{v}$ and vertical: ($\bar{w}$)
N_{vec}	-	unit vector normal to the cross-section in xy-plane
T_{vec}	-	unit vector tangential to the cross-section in xy-plane
b_s	m	surface width of the river
b_m	m	measured width of the river
h	m	water depth
m	-	mode number of horizontal Fourier transform
n	-	mode number of vertical Fourier transform
$a_u^{m,n}$	m s^{-1}	amplitude that corresponds with certain modes m and n for longitudnal velocity fit
$a_v^{m,n}$	m s^{-1}	amplitude for lateral velocity fit
$a_w^{m,n}$	m s^{-1}	amplitude for vertical velocity fit
ζ	-	normalised river width
σ	-	normalised elevation above the river bed
ϵ	m s^{-1}	residual (velocity model error)

1 Introduction

This chapter is an introduction to the study about model assimilation in river flow data processing. Here, model assimilation can be described as the process of blending a data-driven model with measured velocity data in the cross-section of a river. The chapter is divided into four sections. The motivation, relevance and scope are described in §1.1. The problem is defined in §1.2. The research objective and research questions are formulated in §1.3 and in the last section a reading guide of the thesis is given.

1.1   Motivation

Measuring flow velocities and exploring the characteristics of complex flow patterns within rivers is important to improve understanding of the physical behaviour of these systems. Measurements provide input for flood- or low-flow forecasting, water quality assessment, river ecology and morphology, climate research, flood protection and hydropower generation (Adler and Nicodemus, 2001). Over the past decades, the Acoustic Doppler Current Profiler (ADCP[1]) has become standard for flow velocity measurements (*e.g.,* Dinehart and Burau, 2005; Parsons et al., 2013; Vermeulen et al., 2014b), because it collects velocity data at a high spatial and temporal resolution. An ADCP can carry out measurements from a moving vessel over certain period of time at multiple locations along its navigated track between the river banks. The raw velocity data should be post-processed to generate practicable output for managing, evaluating, analysing and displaying the three-dimensional velocity data (Parsons et al., 2013).

Data processing often involves several steps where raw velocity data will be averaged temporally and/or smoothed spatially. The process of averaging and smoothing is composed of assumptions and aggregation of data (*e.g.,* Kim et al., 2007, 2009; Le Bot et al., 2011; Parsons et al., 2013), which is motivated by expecting the presence of a certain spatial and temporal behaviour in the flow field. However, this process is often questionable and/or arbitrary, because the introduced uncertainties are not considered most of the time and the assumptions are not properly validated (Marsden and Ingram, 2004). Generally, the consequences of averaging windows in time and space on the output are obscure (Parsons et al., 2013; Vermeulen et al., 2014b).

[1] Acoustic Doppler Current Profiler; instrument that makes use of acoustic physics and small particles in the water column to measure three-dimensional velocity of the flowing water. More theoretical background of ADCPs is provided in §2.2.

In this study a method is investigated that might lead to a more grounded choice for spatial smoothing, which is based on spectral analysis[2]. Due to this spectral analysis it would be possible to identify the dominant spatial patterns in the cross-section of a river and to estimate the amount of variation contributed by each added sinusoid in the function. The method is based on and applied to velocity measurements that were collected with an ADCP at seven cross-sections in a sharp bend of the Mahakam River, Indonesia.

1.2   Problem definition

Velocity fields are computed and plotted in a (predefined) discrete mesh after processing the raw velocity data. Generally, processing involves temporal averaging and spatial smoothing of velocity data to reduce the local variability in the data, so that the main flow pattern can be clearly discerned. Temporal and spatial averaging can help in producing a composite representation of the cross-sectional flow field. However, averaging windows have often been chosen arbitrarily without a clear substantiation.

1.3   Research objective and questions

The aim of the study is to identify the dominant spatial patterns in a river cross-section, so that averaging and smoothing of data can be executed with more certainty. The objective will be achieved by analysing the spatial distribution of the three-dimensional velocity field in river cross-sections with the help of a data-driven model (based on spectral analysis), which is fed with raw velocity data. For that reason, the creation of a model which analyses flow patterns is a subgoal of this study. Actually, the model will "fit" a continuous function with the processed data, where the accuracy of the fit might improve by increasing the complexity of the function (*i.e.,* incorporate higher order functions progressively in the spectral analysis). However, model complexity should be minimised because of the risk of "overfitting". In this case, the model will be based on turbulence and secondary flow instead of the main flow pattern and might misses its purpose. The process of including higher order functions in the model will be repeated until the residuals[3] of each velocity component show no spatial structure. The objective is reached by answering the following research question:

> *To what extent must a data-driven model (based on spectral analysis) increase in its complexity, to adequately represent the three-dimensional spatial velocity distribution of a river's cross-section?*

Three subquestions are formulated in order to guide this thesis in providing an answer to the main research question.

1. *Which typical flow patterns can be observed by analysing the available data set having regard to the main flow pattern, secondary flow and spatial scales?*

2. *How can the dominant spatial scales in a river's cross-section be analysed and identified?*

3. *What are the spatial scales in an adequate representation of the cross-sectional flow field?*

1.4   Thesis outline

Theoretical background on flow characteristics in river systems, Acoustic Doppler Current Profilers, flow data processing and spectral analysis is provided in Chapter 2. Background on the study area and the data set is here given as well. The methodology to execute this research is described in Chapter 3, which structure is based on the research questions. The results are shown and described in Chapter 4, which treats the several research questions. Multiple discussion points are treated in Chapter 5. The conclusion is provided in Chapter 6 and recommendations for further research in Chapter 7.

2 Background information

This chapter is divided into five sections which provide necessary background information for clarification on topics treated in this thesis later on. Flow characteristics in river systems are described in §2.1. The operation method of the measuring instrument (ADCP) will be given in §2.2. Data processing techniques with their improvements in performance and limitations are discussed in §2.3. General background on spectral analysis is provided in §2.4. At last, the study area with the available data set is described in §2.5 in order to specify the application area of the executed research.

2.1 Flow characteristics of river systems

Rivers are important for navigation and agriculture for thousands of years. Furthermore, they (have) function(ed) as a defensive measure, water supply, disposing of waste and/or bathing. Besides traditional purposes, rivers serve a recreational purpose, cooling water and the generation of hydropower (Padmalal and Maya, 2014). Riverine environments are attractive for flora and fauna as well, because of the availability of water, fertility of the surrounding soils and the capacity of a river to transport sediments, nutrients, plants and animals (Vermeulen, 2014).

River management, and for that reason the measurement of water velocities, is important in order to retain the functions of the river. The velocity has influence on the dynamic behaviour of the river, such as their *e.g.,* patterns, sedimentation load and morphology. However, river flow velocities are highly dynamic, due to interacting processes with varying discharge and bed level changes caused by fluvial processes such as erosion and sedimentation. High velocities occur in upper regions where mountainous rivers are found. In the middle reaches, the discharge regime is calmer, and the lower region acts more like a transition zone towards the sea (Ribberink and Hulscher, 2012). The width-to-depth ratio (B/h) is usually of the order $10^2 - 10^3$ in lower reaches of a river, which means that a natural river is a very wide object in that region (Yalin, 1992). And usually, rivers are visualised in a distorted scale to clarify the features and processes in a cross-section.

In river systems the water flows dominantly in one direction (*i.e.,* downstream), which is called the longitudinal velocity component (*i.e.,* component in the direction of the cross-sectionally averaged flow). Flow velocities vary throughout the cross-section, because of local changing bed levels, gradients, roughness of the bed or obstructions. However, the velocity certainly tends to
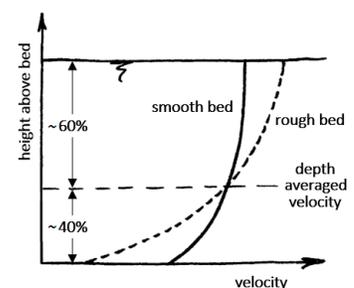

Figure 2.1: General vertical velocity profile of longitudinal component, adapted after Nortier and de Koning (1996).
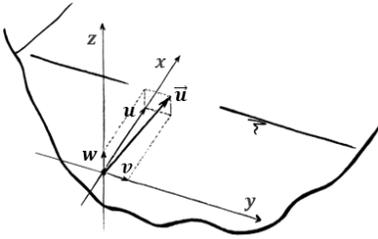
Figure 2.2: Vectorial velocity ($\vec{u}$) with decomposed velocity components (u, v and w).
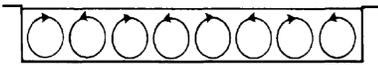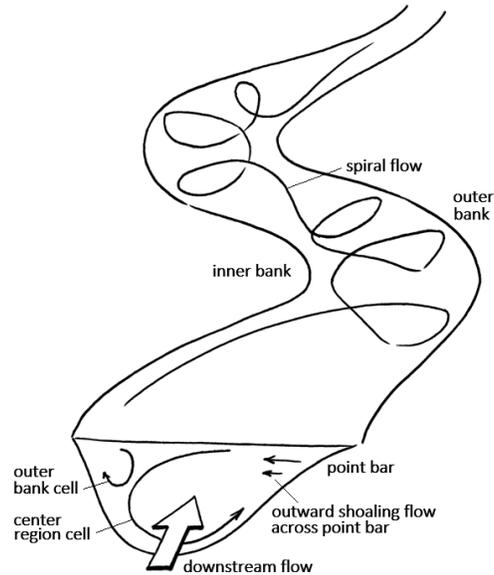
weaken near the solid surface and the strongest flow develops further away from these boundaries (see Figure 2.1, p.5). Each vectorial flow velocity ($\vec{u}$) consists of three velocity components u, v and w in x-, y- and z-direction respectively (Figure 2.2). Secondary flow is present and indicate the flow orthogonal to the cross-sectionally averaged flow, which is minor to the main flow velocity (u). This secondary flow can be distinguished in a lateral (*i.e.,* v, towards the banks) and a vertical (w) component.

Secondary flow can be caused by channel geometry and/or turbulence. Turbulence is generated as water flows along a solid surface or past an adjacent stream with a different velocity. The fluid particles move in irregular paths in turbulent flows (Robert, 2003). The geometry of the channel is mainly responsible for the interaction between the main flow, secondary flow and turbulence (Vermeulen, 2014). In addition, secondary circulation cells exist within straight wide-river flows. According to *e.g.,* Nezu et al. (1985); Kotsovinos (1988) the width of the cells is equal to the water depth (see Figure 2.3).

Figure 2.3: Secondary circulation cells in a wide channel, after Casey (1935).

Circulation in meander bends, often referred as spiral flow, is caused by a centrifugal force and pressure gradient forces. The centrifugal force pushes water to the outer bank, where higher velocities occur, and results in a tilted water surface level. Then a pressure force arises, which is equal and opposite to the mean centrifugal force and acts towards the inner bank attempting to balance the forces. However, the centrifugal and pressure gradient force are generally unbalanced locally (Allen, 1994; Powell, 1998). The flow is driven outwards near the water surface and inwards near the river bed (see Figure 2.4) resulting in a spiral motion of the flow (Robert, 2003; Vermeulen, 2014).

Figure 2.4: Spiral motion in meander bend, adapted after Vermeulen (2014).

Arbitrary cross-sections of a river can be parametrised by a few of parameters that frame the domain as a Cartesian coordinate system in the yz-plane (see Figure 2.5). The velocity field can be denoted by $\vec{u}(x, y, z, t)$ in the domain, which consists of three velocity components (*i.e.*, u, v and w) and may vary in the x-, y- and z-direction and with time t. However, it is assumed that the temporal variation of the flow is negligible during one measurement cycle due to relatively minor changes in flow characteristics in river systems for short periods of time (*i.e.*, neglecting the time derivative by taking $\partial/\partial t = 0$).
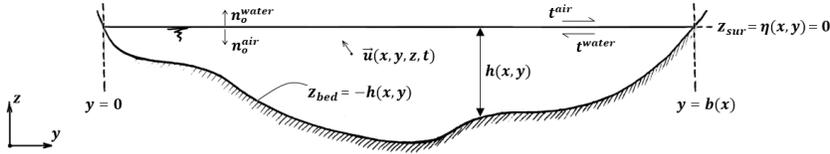

Figure 2.5: Arbitrary cross-section of a river and its parametrisation.

In Figure 2.5 represents z_{sur} the water surface level in which $\eta(x, y)$ is the free surface elevation at certain point x, y and taken as a reference level (*i.e.*, as zero). z_{bed} represents the river bed and is parametrised by the local water depth h at x, y below the water surface. Besides the upper and lower boundary of the domain, the flow is bounded at the sides where the water surface hits the edges of the river profile. These edges are denoted as $y = 0$ and $y = b(x)$ where $b(x)$ represents the surface width at x.

There are two boundaries at a river's cross-section for which boundary conditions (BCs) must be specified. One is located at the interface between the fluid and the solid surface and the other at the interface between two fluids (*i.e.*, air and water) as can be noticed from Figure 2.6. The lower BC at the river bed ($z = -h(x, y)$) can be described by:

$$\vec{u}_{bed} = \vec{0} \tag{2.1}$$

due to the no-slip condition, where $\vec{u}_{bed}$ is a vectorial velocity at the river bed.

The upper BC at the surface ($z = 0$) occurs at the interface between two fluids, which require to apply a kinematic and a dynamic boundary condition. At the free surface, the kinematic condition relates the motion of the free interface to the fluid velocities and the dynamic condition balances the forces (Heil, 2017). The water level is assumed to be steady and for that reason the kinematic BC can be described by:

$$\vec{u}_{sur} \cdot \vec{n}_o = 0 \tag{2.2}$$

where $\vec{u}_{sur}$ is the vectorial velocity at the free water surface, $\vec{n}_o$ is the outer unit normal on the free surface (see Figure 2.5).
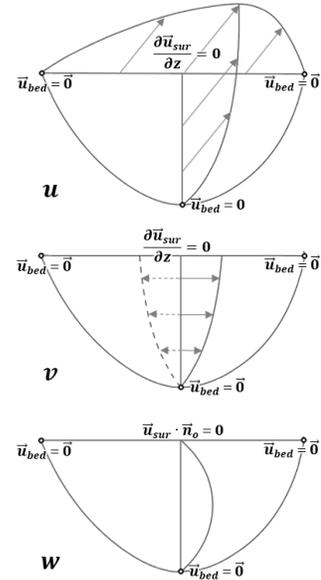

Figure 2.6: General river velocity profile with visualisation of boundary conditions for the three velocity components (u, v and w).

The dynamic BC requires stress continuity across the free surface, which separates the two fluids. The traction exerted by the air due to *e.g.,* wind onto the water surface t^{air} is equal and opposite to the traction exerted by the water surface on the air t^{water} (Heil, 2017). Stress continuity results in same stress τ in the two fluids at the boundary. Air can support no shear stress, since it is an inviscid fluid (*i.e.,* zero or very low viscosity) which results in zero shear stress $\tau = 0$ at the boundary (Morrison, 1998). Therefore, the following dynamic BC is obtained at the surface ($z = \eta(x, y) = 0$):

$$\frac{\partial \vec{u}_{sur}}{\partial z} = 0 \tag{2.3}$$

which suggests that the velocity is continuous from one fluid to another at the water surface boundary.

By summarising the equations above can be stated that the BC at the bed in Eq.(2.1), for which the no-slip condition is applied, indicates that each velocity component reduces to zero at the river bed. The BCs at the free water surface implies that there is no flow through the free water surface, see Eq.(2.2), but there might be a flow tangential to the surface. This tangential flow is equal to the velocity induced by the other fluid at the boundary, see Eq.(2.3).

2.2 Acoustic Doppler Current Profiler

During the mid-1970's, velocity in water flows was measured by an adapted Doppler speed log, which was the predecessor of the ADCP. This instrument, which was intended to measure the speed of ships, was redesigned to measure water velocity more accurately and allows measurement over a depth profile. This led to the first commercial ADCP in the late 1970's. In the years after, ADCPs were further improved for use in long-term, different ADCP models (*e.g.,* self-contained, vessel-mounted and direct-reading) and different frequencies ranging between 75 and 1,200 kHz (R.D.Instruments, 1996; Rowe and Young, 1979). Since the broadband[1] ADCP was developed back in 1992, the instrument has been increasingly used for measurements in shallower waters (Muste et al., 2004). And nowadays, the ADCP has become a standard for flow measurements in (large scale) water systems as mentioned in §1.1.

The instrument is named after Christian Johann Doppler, who discovered in 1842 the relation between the change in frequency of a source to the relative velocities of the source and the observer. He found that the frequency of a (sound) wave will increase as the source and the observer moves towards each other and decrease as they move away from each other (Simpson, 2001).

The transducer of an ADCP transmits sound pulses into the water column, generally along four beams. These pulses will be received by the suspended particles carried by the water and echoed back to the transducer, which results in a Doppler shift[2]. It is assumed that the particles travel with the same velocity

[1] The broadband method facilitates ADCPs to make use of the full signal bandwidth for velocity measurements. This provides more information to assess the velocity, which increases the accuracy and reduces the variance (R.D.Instruments, 1996).

[2] The change in frequency or wavelength when the observer moves towards or away from the source.

as the water. The velocity of the water can be computed with the help of the measured Doppler shift of particles in the water column. The ADCP divides the water column into equally spaced vertical segments, called bins (see Figure 2.7). The transmitted sound takes longer to travel back and forth when the particles are located further from the ADCP. The change in the travel time, the so-called propagation delay, corresponds to a change in distance. The travelled distance of a particle, and so the component of the fluid velocity along the beam path of that bin, can be determined if the propagation delay, speed of sound in water and time lag between two sound pulses are known (R.D.Instruments, 1996).

Transducers are mounted near the water surface and pointed downward for vessel-mounted ADCPs. At least three acoustic beams are necessary for computing three-dimensional water velocity. Generally, there are four independently working acoustic beams, which are angled 20-30° from the normal of the transducer assembly. The flow within the spread of the beams should be assumed as homogeneous in order to use multiple beams to obtain three-dimensional velocity in an ensemble[3]. The fourth beam can be used to evaluate the data quality and whether the assumption of horizontal homogeneity is reasonable with the help of the error velocity[4] (R.D.Instruments, 1996).

Several errors, problems or pitfalls might affect the output data during the operation of ADCP measurements, which may lead to incorrect estimates of the water velocity and directions when these are ignored. These can be caused by *e.g.,* instrument settings, experience of the operator, conditions during operation and/or by the measurement environment (Muste et al., 2004).

Besides the possibility of errors in ADCP data, there are some limitations as well. The suspended matter in the water system might affect the ability of making an accurate velocity measurement with an ADCP. On one hand, water can be too transparent (*i.e.,* no sediment detectable in water column), so the transmitted pulse is not reflected. On the other hand, the system might be too dense sedimented, which can cause inaccurate estimates of the water depth and invalid ship velocity measurements, or the signal might be weakened by the sediment, so the pulse is not received back by the transducer. The range of concentration in a system for appropriate measurements depends on sediment characteristics, water depth and instrument frequency (Mueller and Wagner, 2009).

Another limitation is the unmeasured area in the profile as the ADCP is not able to measure velocities at the water surface due to draft of the instrument and the required blanking distance. The blanking distance is the minimum distance that the sound pulse takes to travel from the transducer through the water to the suspended particles back and forth. First the transmitted pulse must be damped out, before it can received back. Several factors influencing the actual distance to the first measured bin, such as the speed of sound, operating mode, bin size, transmit frequency and beam angles (Simpson, 2001). Besides the unmeasured region near the surface, it cannot measure velocities near the bed
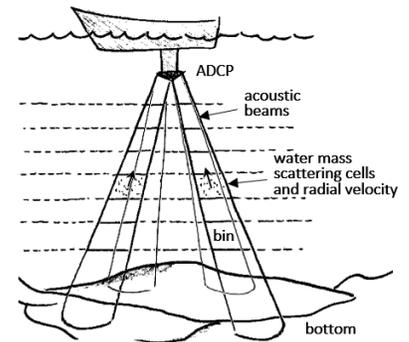

Figure 2.7: Acoustic beams of an ADCP, with bins and suspended particles, adapted after Simpson (2001).

[3] Averaged velocities from bins over water column for one single measurement.

[4] The difference between estimates of the velocity along the different beams.

due to side-lobe interference (see Figure 2.8). According to Simpson (2001) emits most transducers side lobes with an angle of 30-40° to the main beam. Sound intensity in side lobes is much lower than in the main beam. The energy of the backscattered signal from particles in the water column is relatively small compared to the energy transmitted. However, the river bed reflects a much higher percentage of the acoustic energy than the particles in the water column. This can cause errors in the measured Doppler shift, because particles in the main beam are at a point sufficiently close to the backscatter from the bed in the side lobe (Simpson and Oltmann, 1993; Mueller and Wagner, 2009).

Figure 2.8: ADCP beam pattern and unmeasured areas, adapted after Simpson (2001).
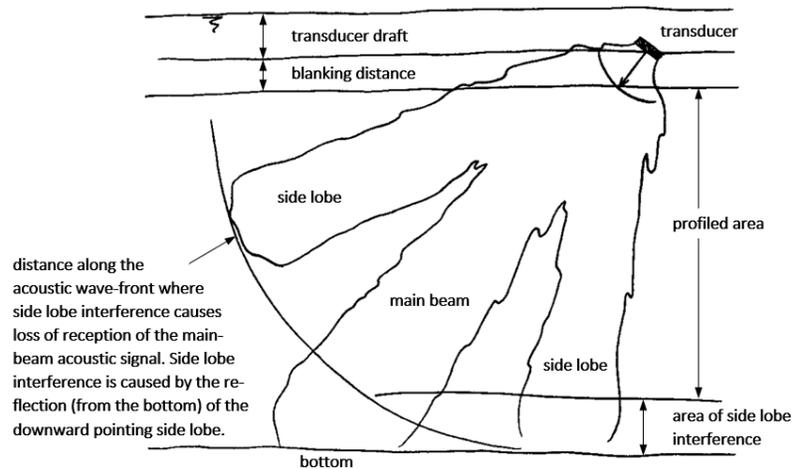
2.3   Flow data processing

[5] Generally, ADCP data includes velocity, echo intensity, correlation and percentage good according to R.D.Instruments (1996). However, here only the radial velocity (relative velocity along the acoustic beam) is meant with raw ADCP output data.

The raw ADCP output data[5] should be post-processed, which generally consists of the following steps (see Figure 2.9): averaging radial velocities, converting beam- into earth coordinates, detecting and modifying errors which can cause deviant velocities with respect to the actual velocities due to several error sources (Muste et al., 2004), determining the absolute water currents and computing the velocities in the unmeasured region (Snowbird, 2012). A quantitative data set in a certain coordinate system is obtained after processing the raw data, which can be used to interpret the data and visualise the flow field.
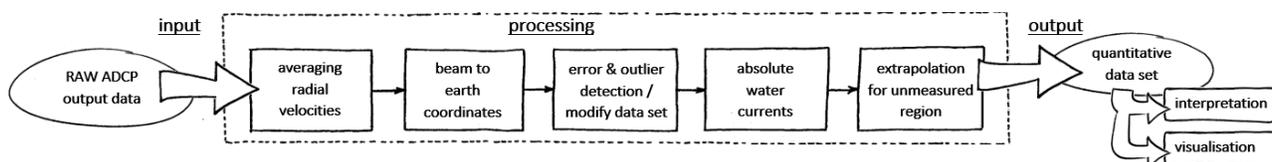

Figure 2.9: General steps in flow data processing.

Multiple techniques in flow data processing have been developed over the past decades, but some of them are already obsolete or not widely available *e.g.,* TRANSECT, MISSING LINK, ADCP toolbox and CASCADE (Adler and Nicodemus, 2001; Côté et al., 2011; Le Bot et al., 2011). Besides the fact that some of these packages became outdated, they were particularly usable for singular purposes (*e.g.,* 1D and single transects) and limited in their application. AdcpXP, VMS, ADCPtool and VMT are techniques which are still usable and applicable for post-processing the ADCP data (Kim et al., 2007, 2009; Steidl and Dorfmann, 2013; Parsons et al., 2013). For spatial averaging and smoothing the data, Parsons et al. (2013) suggest using a moving average in one or two dimensions (*i.e.,* planform or cross-section view respectively, see Figure 2.10) for which the window size can be controlled by the user. Smoothing can help in reducing the local variability in velocity data in order to discern overall patterns, especially for the secondary flow. Parsons et al. (2013) presents the differences in the resulting flow field without smoothing ($n_h = 0$, $n_v = 0$), light smoothing ($n_h = 1$, $n_v = 1$) and enhanced smoothing ($n_h = 8$, $n_v = 2$). Latter results in a 4 m smoothing window. However, even though the main pattern can be discerned more obvious, the window size is arbitrary.

However, all these techniques still assume flow homogeneity in a certain volume between the acoustic beams of the ADCP, which is often questionable (Marsden and Ingram, 2004). The divergence of the beams causes an increasing volume per bin where the flow is still considered as homogeneous. The quality of the measurements decreases (*i.e.,* reduction of variance or spatial filtering) when flow depth is increasing. Cross-sections should be measured multiple times back and forth in order to average the instantaneous velocities and assess the quality of the measurements. Generally, there is more spatial than temporal variation in the velocity field of a river's cross-section However, the measured track is arbitrary and should be averaged to one transect, which introduces uncertainties.

Recently, Vermeulen et al. (2014b) have developed a method where the influence of assuming homogeneous flow is minimised. This assumption is, as mentioned, needed for combining and averaging the radial velocity components in the bin that are collected by an ADCP. Conventional flow data processing collects radial velocity components simultaneously at a certain distance, *i.e.,* in a certain bin of each beam. The measured velocities are combined and averaged instantly (see Figure 2.11a). The proposed method of Vermeulen et al. (2014b) combines radial velocity components collected in a predefined cell instead of a certain volume between the acoustic beams (see Figure 2.11b). Here, measurements from different moments can be used to average the velocity. This reduces the volume in which the flow is assumed homogeneous from the distance between the beams to the size of a cell. However, even though this method provides for improved velocity estimates from moving-boat measurements, it does not conclude about the size of spatial averaging windows.
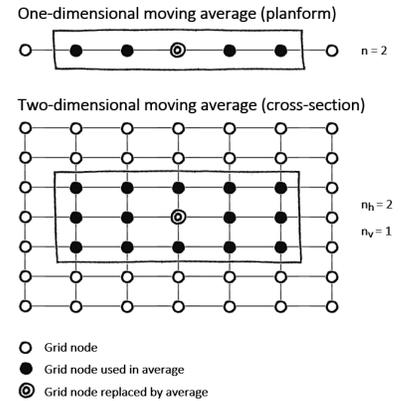

Figure 2.10: Schematic diagram of a spatial averaging procedure employed by VMT, after Parsons et al. (2013). Here, n_h and n_v represents the horizontal and vertical smoothing window size.
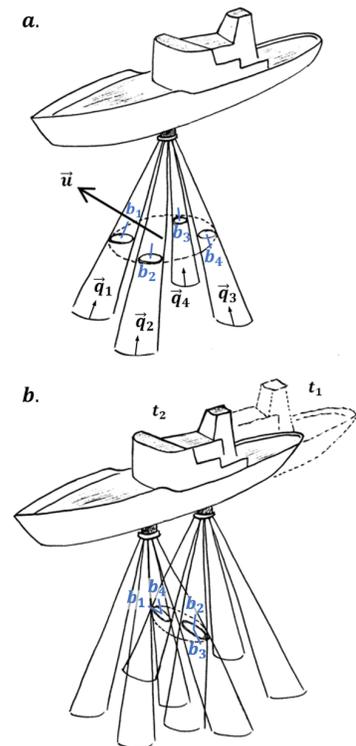

Figure 2.11: (a) Conventional flow data processing and (b) proposed method according to Vermeulen et al. (2014b).

2.4 Spectral analysis

There are several forms in spectral analysis, such as harmonic analysis, Fourier analysis and frequency analysis. The Fourier analysis is used in this study, which was provided by Jean Baptiste Fourier (1768-1830). He proved that any continuous, single valued function could be represented by a series of sinusoids. The continuous function can be a sequence of observations taken at equal intervals of time or distance, which results in time series or spatial series respectively. In this research we make use of spatial series. The method of Fourier analysis is provided in this section with the help of information found in Davis (2002).
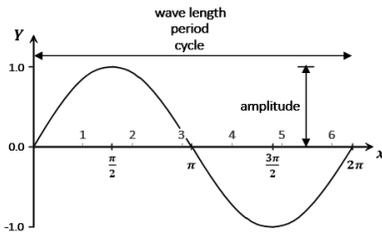
The curve of a sine wave oscillates between +1 and -1, with an equilibrium position of 0. The equation of the curve shown in Figure 2.12 is

$$Y = \sin x$$

where x is given in radians for $0 \leq x \leq 2\pi$. The oscillation, *i.e.,* amplitude, can be changed by multiplying $\sin x$ by any constant A

$$Y = A \sin x$$

Figure 2.12: Terms applied to a regularly repeating sine curve ($Y = \sin x$).

The distance between two similar points in the curve with the same slope is called the wave length, period or cycle. The frequency is the reciprocal of the wave length, being the number of waveforms, periods, or cycles that occur in some interval of distance. It can be changed by multiplying x by an integer k, which results in

$$Y_k = A_k \sin(kx)$$

where the amplitude is subscripted because it is associated with a specific number of waveforms, k, which is referred to as the mode number. Any series of spatial data can be represented as the sum of a series of sinusoids, resulting in the Fourier relationship:

$$Y = \sum_k \left[\, \alpha_k \cos(kx) + \beta_k \sin(kx) \,\right] \tag{2.4}$$

However, the Fourier transform is one-dimensional in this case. During the study a two-dimensional Fourier transform is applied to the measured velocity data in order to analyse the model fit of the cross-sectional velocity distribution with specific number of waveforms both vertically as horizontally. There are two trigonometric terms in Eq.(2.4) and each is multiplied by its own coefficients. These coefficients can be found by regression analysis. Note that spatial series are measured over a certain distance generally, but Eq.(2.4) is expressed in terms of x in radians. In this case x can be substituted by a scale over the total

spatial series. The two-dimensional general Fourier transform results in

$$Y(y,z) = \sum_{m}^{M} \sum_{n}^{N} \left[a_{m,n} \, \sin\left(m\pi \, \frac{y}{b_s}\right) \, \sin\left(n\pi \, \frac{z}{h(y)}\right) \right] \qquad (2.5)$$

where y is a certain velocity location horizontally for which m is the increment-ing index which influences the number of waveforms of width, z is a certain velocity location vertically for which n is the incrementing index which influ-ences the number of waveforms of depth, M and N are the upper bounds of the summation (*i.e.,* truncation numbers), b_s is total river width, $h(y)$ is water depth at point y. There is one Fourier coefficient $a_{m,n}$ for each combination of modes (m, n), which can be found by regression analysis. This equation is further elaborated and explained in §3.2 and applied to the different velocity components.

Quimpo (1967) has applied Fourier analysis to river flow data in order to identify the presence of significant harmonic components over time. Later on, other researches applied Fourier series for the simulation of river flow over time and is since decades a commonly used tool in hydrologic studies concerning periodicity (Tesfaye, 2005). Here, the periodic behaviour has been presented with Fourier functions that can be used for *e.g.,* analysis and design of water resource systems and river basin management (Saremi et al., 2011).

Even though Fourier transform over a spatial domain in analysing river flow fields is not shown before it seems applicable for estimating the three-dimensional flow velocity. Because typical flow patterns over river width and water depth show comparable shapes to parts of sinusoids. Moreover, the Fourier transform superpositions multiple trigonometric terms with specific amplitudes such that the velocity can be approached over the domain at a moment. The velocity might variate due to irregular bed elevation, gradients, roughness or obstructions. More mode numbers can be included to have a more detailed estimation of the flow field.

By representing the three-dimensional velocity field with series of sinusoids errors may occur resulting from the phenomena aliasing (as shown in Figure 2.13). There are insufficient samples present or taken to distinguish the high and low frequency, which lead to a distorted representation in this case. In fact, the measured velocity and the model fit with the Fourier transform show different velocities.
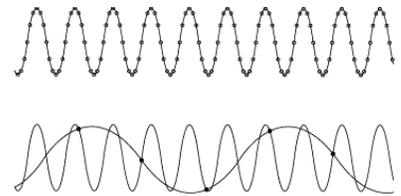

Figure 2.13: Principle of aliasing, where the upper signal is adequately sampled and the lower one is aliased due to undersampling.

2.5   Study area

The study focusses on the Mahakam River because of the available data set. The river is located in the Indonesian part of Borneo island, East Kalimantan and is the second longest in Kalimantan (see Figure 2.14).

Figure 2.14: Overview of the Mahakam River with the location of the measured bend, adapted after Pham Van et al. (2016).
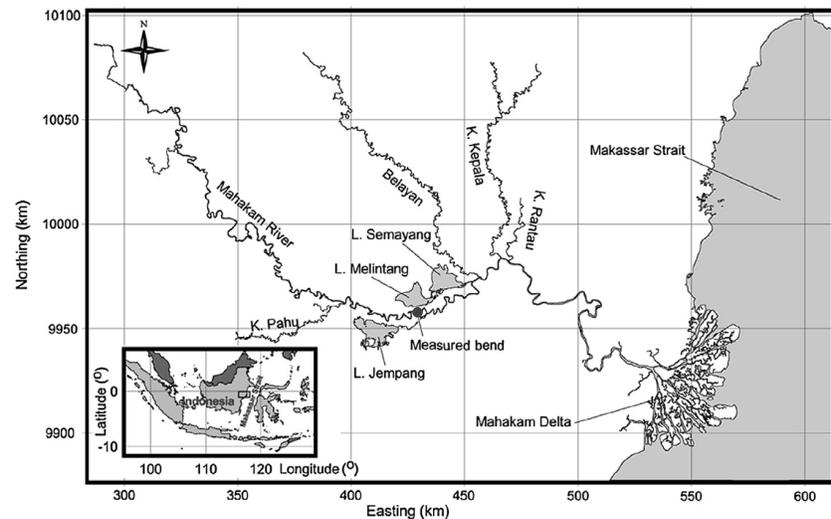
The catchment covers an area of about 75,000 km^2. The Mahakam River has a length of about 900 km with an annual river discharge that varies between 1,000 and 3,000 m^3 s^{-1} (Allen and Chambers, 1998). The river flows from the highlands of Borneo, through the Tertiary rocks of the Kutei Basin to the Mahakam delta and ends in the Makassar Strait. In the lower part of the river, the water level is influenced by high and low tide (ranging from 1.0 to 3.0 m) of the sea. This tidal influence extends upstream to the middle part of the river (Pham Van et al., 2016). This middle part is extremely flat, where several lakes are formed. The three main lakes are Lake Jempang, Lake Semayang and Lake Melintang (Vermeulen, 2014). These lakes function as buffer, with storage capacities up to 2.7 km^3 (depending on dry or rainy periods and the storage volume of the lake), and is used for fishing (Hidayat et al., 2011).
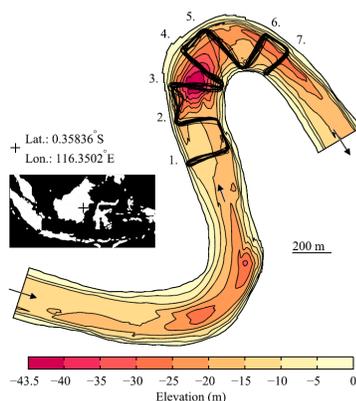
2.5.1   Measured river bend

This bend is located in the Mahakam lakes region and is according to Vermeulen et al. (2014a) representative for many sharp bends with deep scours found in this region. The mean of the measured width in the considered part is about 245 meters with an average depth of 15 meters. However, due to the presence of a scour hole, the water depth increased to more than 40 meters locally (see Figure 2.15). The cross-sectional area varies between 2,200 and 7,000 m^2. The water level remained constant during data collection and the average discharge was about 1,700 m^3 s^{-1} (Vermeulen et al., 2015).

Figure 2.15: Overview of sharp river bend with bed elevation, the measured track and flow direction, adapted after Vermeulen et al. (2015).

2.5.2   Data set

The available data set consists of measurements with a 1200 KHz (vessel-mounted) ADCP, that was collected in a sharp bend of the river on the 25th of August, 2009. Seven cross-sections were repeatedly monitored 16 times (on average, see Table 2.1) within this bend between 06:53am and 01:05pm (see Figure 2.15). The total length of the navigated track is estimated on 40 km with a duration of 6 hours and 12 minutes. The velocity of the vessel is determined on 1.75 m s^{-1} by assuming a constant velocity. From this can be approximated that a cross-section is measured in about 100-160 seconds depending on the measured width. The maximum period between two successive measuring cycles of a section amounts about 45 minutes.

The ADCP has four acoustic beams and had collected the velocity in 80 bins with a bin size of 50 cm. The blanking distance amounts 44 cm and measured velocities in the lowest 6% of the ensemble are ignored to account for side lobe interference. There are 15,783 measured ensembles in total. Of which 10,954 are located within the region of a transect. This region excludes the navigated distance between two measured sections and can filter for proximity to the average transect. The proximity represents the maximum permissible offset of data from the average. It is set as 0 m in this case, indicating that all measured data is included. Furthermore, outliers (*i.e.,* a velocity with a magnitude distant from other velocities) can be removed from the data set with the help of two built-in manners. In one way, outliers are determined with a value that represent the times that the residual in beam-velocity might exceed the median of all residuals to discard from the data set. In the other way, outliers are determined with a value that represent the times that the standard deviation of any parameter in a cell might excess the median of the standard deviation over all cells to be removed from the data set. The latter method is used during this study, in which the standard deviation of a measured parameter is allowed to exceed the standard deviation over all cells 6 times before it is treated as bad, which is the default value.

Table 2.1: Measured width (B_m), times measured (t) and number of ensembles for the seven transects (T).

T	b_m [m]	t	ensembles
1	240	17	1,776
2	215	17	1,491
3	280	17	1,879
4	285	15	1,524
5	280	16	1,665
6	210	16	1,356
7	180	16	1,263

3   Methodology

This chapter is divided into four sections in order to describe the steps that are taken to fulfil the research objective and provide answers to the research questions mentioned in §1.3. First, the method of determining the typical flow patterns in the available data set is clarified in §3.1. Second, the approach to set up a model that can be used to analyse and identify dominant spatial scales in cross-sectional velocity data is given in §3.2. Third, the method of determining spatial scales is explained in §3.3. Finally, the evaluation on the model fit and its performance on providing an adequate representation of the flow field is given in §3.4.

3.1   Identify typical flow patterns

The available data set, which is processed according the proposed method by Vermeulen et al. (2014b), will be analysed for typical flow patterns. The extent over which homogeneity is assumed is reduced strongly in this method by generating a mesh for each transect domain in which all measured velocities are stored and averaged afterwards. Moreover, instantaneous flow is filtered out by accumulating all the data first and computing averages. This results in the mean flow (in three dimensions) over all measurement cycles of a certain transect, from which the typical flow patterns can be analysed. Main flow patterns, secondary flow and spatial scales are observed and described qualitatively. These features are quantified with the help of defining the flow area that carries the strongest flow.

3.2   Model set up

The model should be set up with relevant data extracted from the data set. This relevant data consists of the data that is useful for the research, *i.e.,* measured velocities, and can be allocated to a particular transect. In addition, the data should be processed and transformed to a normalised coordinate system to become suitable for this analysis.

3.2.1   Data extraction

Raw velocity data will be extracted from the data set. Velocity data is measured along the track of the vessel (see Figure 3.1) and in that specific order. A virtual timeline of the measurement cycles is visualised in Figure 3.2, p.18. The vessel starts surveying transect 1 at 06:53am and navigates along the track

−43.5 −40 −35 −30 −25 −20 −15 −5   0
Bed elevation (m)

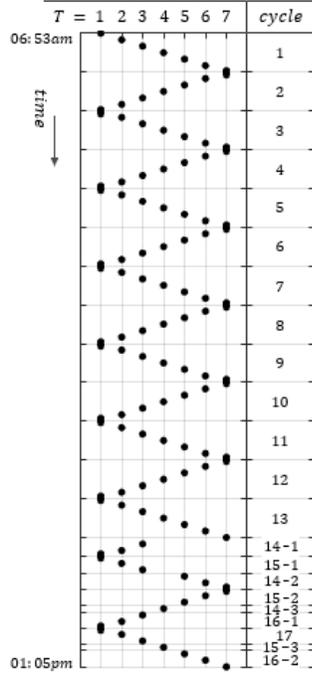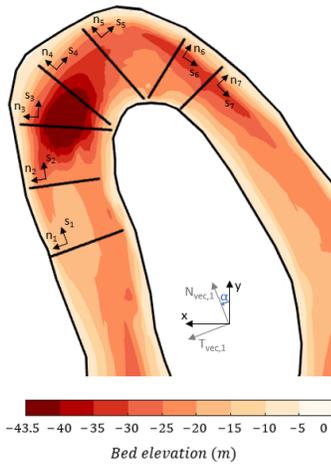Figure 3.1: Measured track of the vessel and removed parts in blue.

Figure 3.2: Measuring timeline of transects.

Figure 3.3: Average transect (nz-)planes with normal and tangential unit vectors of the transects in s- and n-direction, and an example of the principle of "rotation" for transect 1.

up to transect 7, where it turns back again along the same track to transect 1. All intermediate transects are measured consecutively, and so are the distances between two successive transects (blue in Figure 3.1, p.17). This process is repeated up to and including measurement cycle 13. Then a disturbance is visible for which the reason is not known. The vessel navigates with the help of barrels at transect ends. However, measured data in periods going from one transect to the next is not required, which will be removed. Ensembles located in the desired range of a particular transect are allocated to that transect. All data collected in the different measurement cycles are accumulated per transect. However, this can also be separated per cycle if desired. The measured data from the ADCP consists amongst others of a velocity component in x-, y- and z-direction and an error velocity.

3.2.2   Projection of data

Each measured point in the river has a certain location in earth coordinates (x and y) and is situated at a certain level below the water surface (z). The data is collected along the irregular track of the vessel in a time span of the measurement cycles. An average transect, *i.e., nz*-plane, is defined for all seven measured sections (Figure 3.3) with the help of all velocity locations in a referred section. Here, the n-coordinate is directed along the average transect orthogonal to the longitudinal direction and the z-coordinate is directed vertically. The velocity locations are projected to the averaged transect plane (in nz-coordinates) by an orthogonal translation.

Furthermore, the velocity data will be analysed per transect. For that reason, the velocity components in x- and y-direction must be broken down into a longitudinal and a lateral component for that specific transect, in s- and n-direction respectively (see Figure 3.3). The vertical velocity component remains unchanged and functions as it were like a rotation axis. Unit vectors normal and tangential to each transect (N_{vec} and T_{vec} respectively) will be computed. This provide the longitudinal and lateral direction of a transect in order to decompose measured velocities into the desired component.

$$N_{vec} = \begin{bmatrix} \sin \alpha, & \cos \alpha \end{bmatrix} \tag{3.1}$$

$$T_{vec} = \begin{bmatrix} -\cos \alpha, & \sin \alpha \end{bmatrix} \tag{3.2}$$

Here, α represents the angle between the x-axis and the average transect and between the angle orthogonal to the average transect and the y-axis (see Figure 3.3). The first term relates to the x-coordinate of that unit vector, second to the y-coordinate. Clockwise rotation results in a positive α and counter-clockwise in a negative α. The x- and y-coordinate of the unit vectors can be used to compute the velocity component in longitudinal ($\hat{u}_s$) and lateral ($\hat{v}_n$) direction

for that specific transect.

$$\hat{u}_s = N_{vec,x} \cdot u_x + N_{vec,y} \cdot u_y \qquad (3.3)$$

$$\hat{v}_n = T_{vec,x} \cdot u_x + T_{vec,y} \cdot u_y \qquad (3.4)$$

Here, $N_{vec,x}$ represents the x-coordinate of the unit vector normal to the transect and $N_{vec,y}$ the y-coordinate of that vector. Same for $T_{vec,x}$ and $T_{vec,y}$ that represents the x- and y-coordinate of the unit vector tangential to the transect. u_x is the measured raw velocity in x-direction and u_y the measured raw velocity in y-direction.

3.2.3   Transformation of arbitrary river cross-sections

Velocity locations are transformed to a normalised coordinate system in order to analyse (irregular) arbitrary cross-sections for spatial patterns in the three-dimensional velocity distribution. The normalised system that is being used, scaled the cross coordinate with width into ζ-coordinates and the vertical coordinate with water depth into σ-coordinates (see Figure 3.4).

The σ-coordinate system is commonly used for oceanography, meteorology and other fields where fluid dynamics are relevant (Janjic et al., 2010). The layers in the system follow the terrain by normalising the vertical coordinate by the fluid depth to smoothly incorporate the topography (Marshall et al., 2004). $\sigma = 0$ represents the river bed and $\sigma = 1$ the water surface.

ζ-coordinates are relative to the river width, which is ever changing along the river and with time due to $e.g.,$ high and low flows. The river width at the seven transects is determined with the help of x and y-coordinates of the river banks in that section. These coordinates are extracted from another source and the results are not by definition equal to the river width during collection of velocity data. However, this will not lead to problems directly, since the spectral analysis can still be executed. But it might generate wrong results, when the output is used to determine the total cross-sectional discharge for example. $\zeta = 0$ represents the inner edge of the river bend, and $\zeta = 1$ represents the outer edge. The normalised river width (ζ) is computed by.

$$\zeta = \frac{n + |d_{out}|}{b_s} \qquad (3.5)$$

Here, n is a certain point in lateral direction, $|d_{out}|$ is the absolute distance to the outer edge of the river bend from the centerline (CL) and b_s is the surface width of the river, which can be computed by the sum of the absolute distance to the outer edge $|d_{out}|$ and the distance to the inner edge of the river d_{in} both from centerline (see Figure 3.5). $n = d_{out}$ and $n = d_{in}$ results in $\zeta = 0$ and $\zeta = 1$ respectively.
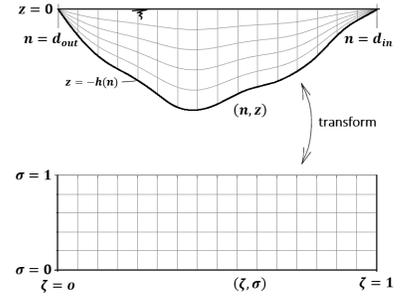

Figure 3.4: Transformation of nz-plane into σ- and ζ-coordinates to obtain the normalised (ζ,σ-)system.
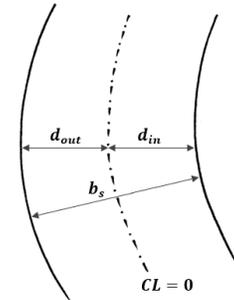

Figure 3.5: River bend with distance from centerline (CL) to outer bank (d_{out}) and from centerline to inner bank (d_{in}), which results in the water surface width (b_s) at a certain location.

The normalised elevation above river bed (σ) is computed by

$$\sigma = 1 + \frac{z}{h(n)} \tag{3.6}$$

Here, z is a certain point below the water surface (negative) in the water column and $h(n)$ is the total water depth at a certain point n, $z = 0$ and $z = -h(n)$ results in $\sigma = 1$ and $\sigma = 0$ respectively.

3.2.4 Fourier transform

Spatial patterns in the velocity distribution of several transects are investigated by a method based on Fourier transform. The velocity vector is decomposed into three velocity components (u, v and w), which are investigated separately by a two-dimensional Fourier transform. For each component a set of base function is formulated with the help of sinusoids. Within these equations higher order function are progressively included and fitted to all the data measured at a transect. The first four modes for u, v and w (with amplitude 1) are visualised separately in Figure 3.6 and 3.7.
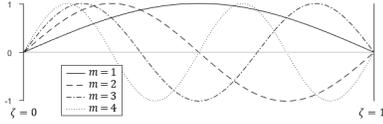

Figure 3.6: First four separate mode numbers (m) of the first part in Eq.(3.7), (3.8) and (3.9) over normalised river width for $\tilde{u}$, $\tilde{v}$ and $\tilde{w}$.
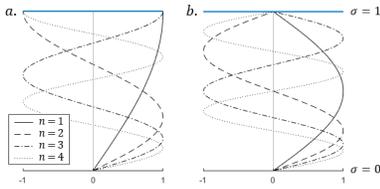

Figure 3.7: First four separate mode numbers (n) of (a) the second part in Eq.(3.7) and (3.8) over normalised water depth for $\tilde{u}$ and $\tilde{v}$ and (b) of second part in Eq.(3.9) for $\tilde{w}$.

$$\tilde{u}(\zeta,\sigma) = \sum_{m=1}^{M}\sum_{n=1}^{N} a_u^{m,n} \underbrace{\sin\left(m\pi\zeta\right)}_{(I)} \underbrace{\sin\left(\left(n-\tfrac{1}{2}\right)\pi\sigma\right)}_{(II)} \tag{3.7}$$

$$\tilde{v}(\zeta,\sigma) = \sum_{m=1}^{M}\sum_{n=1}^{N} a_v^{m,n} \underbrace{\sin\left(m\pi\zeta\right)}_{(I)} \underbrace{\sin\left(\left(n-\tfrac{1}{2}\right)\pi\sigma\right)}_{(II)} \tag{3.8}$$

$$\tilde{w}(\zeta,\sigma) = \sum_{m=1}^{M}\sum_{n=1}^{N} a_w^{m,n} \underbrace{\sin\left(m\pi\zeta\right)}_{(I)} \underbrace{\sin\left(n\pi\sigma\right)}_{(II)} \tag{3.9}$$

Here, $\tilde{u}(\zeta,\sigma)$ is the fitted longitudinal velocity component to $\hat{u}_s$ in Eq.(3.3), $\tilde{v}(\zeta,\sigma)$ the fitted lateral component to $\hat{v}_n$ in Eq.(3.4) and $\tilde{w}(\zeta,\sigma)$ the fitted vertical component to u_z in the $\zeta\sigma$-plane, i.e., normalised over river width and water depth. The velocity can be approximated by the sum of the product of sinusoids with specific mode numbers m and n. These modes affects the number of waveforms over normalised width and depth respectively and increase to truncation numbers M and N. The amplitudes $a_u^{m,n}$, $a_v^{m,n}$ and $a_w^{m,n}$ represents the maximum velocity magnitude [m s^{-1}] for the different components ($\tilde{u},\tilde{v},\tilde{w}$) and will change for each set of modes (m,n).

According the no-slip condition at solid surfaces, the velocity at the bed drops to zero, hence the form of the trigonometric terms in the base functions of Eq.(3.7), (3.8) and (3.9). The longitudinal and lateral component have a velocity at the water surface (Figure 3.7a), but the vertical component is there zero as well (Figure 3.7b). The sinusoids of all different modes are superpositioned

by taking the sum of modes up to and including truncation number M and N. The product of sinusoids results in a sort of density distribution within the normalised domain. Each velocity at a specific location in this normalised domain can be approximated with a certain summation of sinusoids and its corresponding Fourier coefficient, *i.e.,* amplitude. The amplitudes will be determined with the help of linear regression analysis[1] between the measured velocity data and the trigonometric terms. This analysis provides a least-squares solution that minimises the sum of squared residuals, which can be used as amplitude.

[1] Using the `lscov`-function in MATLAB, which returns the least-squares solution to a linear equation $A * x = B$, where A represents the trigonometric terms, B the measured velocity and x the amplitude.

3.3   Determine dominant spatial scales

The computed amplitudes in the model fit will be analysed for different set of modes, from which the contribution to the velocity fit is observed. In general, higher amplitude values contribute more to the model fit and the coarse main flow can be approximated with a few sinusoids. Including more modes will refine and reshape the model fit with relatively low amplitudes. The truncantion numbers are increased by $M = N$ in order to focus on increasing the numbers rather that cogitate about which specific combination of modes to analyse. Because there are many possible set of modes and too much to consider all of them. By analysing the amplitude values can be determined which are dominant for the three velocity components. The truncation number is increased to observe possible dominant spatial regions or scales in the set of modes for changing the upper bounds of the summation (M, N). Domination, *i.e.,* strongest amplitudes, for the three velocity components and between the different transects are compared to each other. Furthermore, the modes with dominant amplitudes have a certain number of waveforms relative to the normalised width and depth. This will be related to the actual river width and water depth in order to estimate the spatial scales.

3.4   Evaluate the model fit

The model fit on the measured velocity data is evaluated by observing its improvement over both normalised river width and water depth while truncation numbers are increased. In addition, the depth and width averaged velocity of the fit and measured data are compared to estimate the added value of including more modes. Moreover, the model fit will be evaluated with the help of an error function for each velocity component to determine the residuals. A residual is the value that remains after the estimated value is subtracted from the measured value. The function to compute residuals is generally formulated by

$$\epsilon_i = \hat{u}_i - \tilde{u}_i \qquad (3.10)$$

where, i is the index of observed velocity locations. $\hat{u}_i$ and $\tilde{u}_i$ represent measured and modelled longitudinal velocity respectively at a certain location i, which
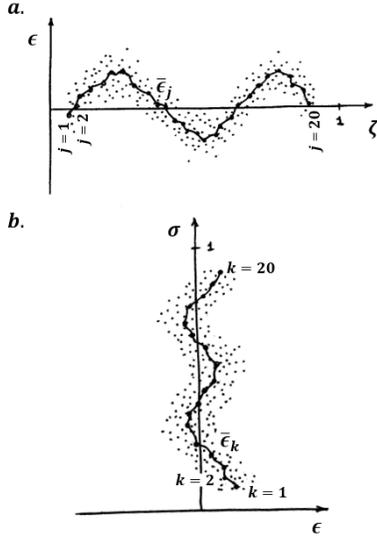
Figure 3.8: Residuals (ϵ) and $RMSE$ over both normalised (a) river width (ζ) and (b) water depth (σ).
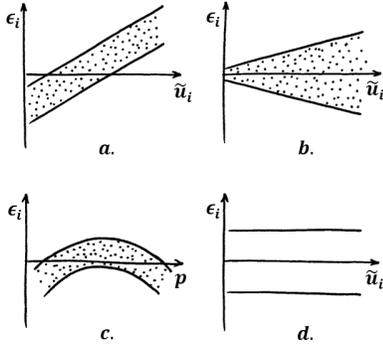

Figure 3.9: Anomalies in residuals, after Johnson and Wichern (2007)

results in the residual at that same location (ϵ_i). This function is also applicable for the lateral (v) and the vertical (w) velocity components.

The extent to which the model fit matches the measured velocity fit can be evaluated with the help of residuals, which can be analysed over both river width and water depth (see Figure 3.8). The residuals should be distributed around zero without spatial structure over the domain for a good model fit, which means that the residuals have no dependence on the predicted variable. However, within the residuals there might be some possible anomalies according to Johnson and Wichern (2007). The residuals may have dependence on the fitted velocities $\tilde{u}_i$ (see Figure 3.9a) when calculations are incorrect or a mean value has been omitted from the model. Another anomaly that may appear is a funnel shaped pattern (see Figure 3.9b), which shows large variability for large $\tilde{u}_i$ and small variability for small $\tilde{u}_i$. So, the variance of the error is not constant. It is also possible that a systematic pattern is observed as the residuals are plotted against a predictor variable p as shown in Figure 3.9c, which suggests the necessity of including more terms in the model. The residuals form a horizontal belt in Figure 3.9d, which is desired in this case because it indicates equal variances and no dependence on the fitted velocity $\tilde{u}_i$. The root mean squared error ($RMSE$) over normalised river width and water depth are computed in order to quantify the spatial structure in the residuals.

$$RMSE_\zeta = \sqrt{\frac{\sum \epsilon_j^2}{n_j}}, \quad RMSE_\sigma = \sqrt{\frac{\sum \epsilon_k^2}{n_k}} \tag{3.11}$$

Here, ϵ_j represents the local average of residuals for $j = 1, 2, \ldots, 20$ over normalised width and ϵ_k the local average of residuals for $k = 1, 2, \ldots, 20$ over normalised depth (see Figure 3.8). j and k are equivalent to about the river width divided by 10 and the water depth divided by 1, which provides sufficient insight in the spatial structure of residuals. Both are divided by the number of regions (n_j and n_k) where the local average is computed to average the sum of squared errors. The functions can be applied for the three velocity components (u, v and w). The outcome is in original units [m s^{-1}] since the root of MSE is taken.

In addition, the model performance will be analysed with the coefficient of determination (R^2) and the standard error of the regression (SER).

$$R^2 = 1 - \frac{\sum (\hat{u}_i - \tilde{u}_i)^2}{\sum (\hat{u}_i - \bar{u})^2} \tag{3.12}$$

$$SER = \sqrt{\frac{\sum (\tilde{u}_i - \hat{u}_i)^2}{n_{obs}}} \tag{3.13}$$

Here, i is the index of observed velocity locations. $\hat{u}_i$ represents a measured longitudinal velocity at a certain location i and $\tilde{u}_i$ is the fitted longitudinal

velocity for that same point. $\bar{u}$ is the mean of measured longitudinal velocities and n_{obs} are the number of observations. These functions are also applicable to the lateral (v) and vertical (w) velocity components. R^2 is in % and SER in m s^{-1}.

R^2 indicates the percentage of the variability of the model fit that is explained around its mean. Where 0% indicates that the model explains none of the variability and 100% indicates that the model explains all the variability of the response data around its mean. The latter implies that the regression line fits the measured data perfectly. However, a high value of R^2 is not necessarily good and a low value is not inherently bad. High R^2 values might show still a spatial pattern between residuals and the model fit, where including more terms is recommended. Some behaviour is difficult to predict, but a derived relative low value for R^2 may still help in draw conclusions (Frost, 2013). The standard error of the regression is analysed as well, which is another measure to estimate the accuracy of the model fit. It represents the average distance that the measured velocity data fall from the regression line and can be used to assess the precision of the model fit with the help of stating a confidence interval (Frost, 2014). Creating best values for R^2 and SER is not necessary but is been taken as an indication. The behaviour of the spatial pattern in residuals over normalised width and depth is most important.

Additionally, the velocity fit is interpolated between the velocity locations and extrapolated towards the boundaries to have an overview of the full flow field. This inter- and extrapolation of velocity is carried out with different truncation numbers. From this can be analysed which is best capable of describing the extended flow field. Finally, the model fit that is able to represent the flow field adequately will be transformed back to the original nz-coordinates. Its performance can be compared with the processed flow field according the method of Vermeulen et al. (2014b).

4 Results

The results are provided in the same order as the research questions, as formulated in §1.3. Typical flow patterns present in the available data of the measured river bend are described in §4.1. Next, the model set up is further clarified in §4.2. The dominant spatial scales are analysed and evaluated in §4.3 and the improvement of the model fit and its performance on representing an adequate flow field are evaluated in §4.4. In general, the results are provided by focussing on transect 1. As another transect behaves different or striking features are observed, will this be discussed additionally. Furthermore, for further clarifications and visualisations on the results is often referred to the appendices.

4.1   Typical flow patterns

The typical flow patterns, in the data processed by the proposed method of Vermeulen et al. (2014b), are analysed and described in order to attempt to quantify the present main flow (u), secondary flow (v, w) and spatial patterns. The three-dimensional flow throughout the river bend of the different transects is visualised in Figure 4.2, p.26. The inner edge of the bend is located at the right of the sections. Furthermore, the coloured shades in this figure indicate the longitudinal velocity component (u), which is pointing through the paper for positive velocity. Secondary flow is visualised with arrows, where the lateral (v) and vertical (w) components are combined. Highest velocities of the three components for all measured transects are provided next to Figure 4.2, p.26 and additionally the maximum and average water depth of each transect. Explanations of Vermeulen et al. (2015) are used additionally to describe the flow patterns. Characteristics of the flow and quantified spatial patterns are provided in Appendix A.

At transect 1 (Figure 4.2, p.26), the longitudinal flow shows a flow pattern with lower velocities near the boundaries and strong flow in the center, which can be interpreted as ordinary behaviour as it is mentioned in §2.1. However, the core of strong flow is separated by two regions of strongest flow, one near the left side and one near the right side of the cross-section. The profile is a bit deeper on the right side and is considered to be influenced by the upstream river bend that pushes strong flow to its outer bank, which causes local scouring at higher velocities (see Figure 4.1). Secondary flow is slightly visible at the river banks where it is directed towards the banks and pushed down towards the deeper regions. A small circulation cell is visible in the deepest part.
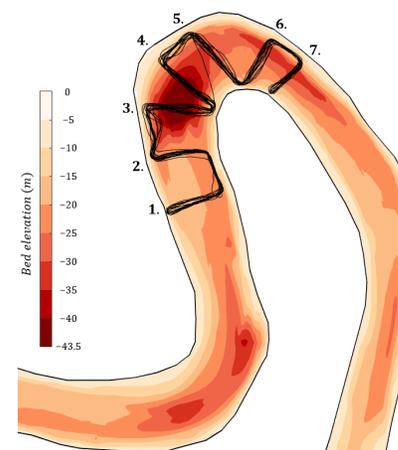

Figure 4.1: Measured track of the vessel with bed elevation and location of transects.

Just before the scour hole, at transect 2 (Figure 4.2), the core of high longitudinal flow is located on the right. The longitudinal velocity shows here an ordinary flow pattern as well. However, at the very outer bank a negative flow is perceived. This means that the water is flowing upstream locally, caused by the scour hole and the river bend. Secondary flow is visible, directed towards the lower middle, caused by the downward slope of the bed at this location and flowing of water towards the hole.

Figure 4.2: Flow patterns for the seven transects (T), corresponding with Figure 4.1, from available processed data set according to Vermeulen et al. (2014b) with longitudinal velocity in coloured shades and lateral and vertical velocity combined in vectors. All profiles are plotted with inner bank on the right side. This visualisation is adapted after Vermeulen et al. (2015).

T=1: $h_{max} = 15.74$ m, $h_{avg} = 11.12$ m, $u_{max} = 0.75$ m s^{-1}, $|v_{max}| = 0.10$ m s^{-1} and $|w_{max}| = 0.05$ m s^{-1}.

T=2: $h_{max} = 16.07$ m, $h_{avg} = 12.80$ m, $u_{max} = 0.82$ m s^{-1}, $|v_{max}| = 0.22$ m s^{-1} and $|w_{max}| = 0.08$ m s^{-1}.
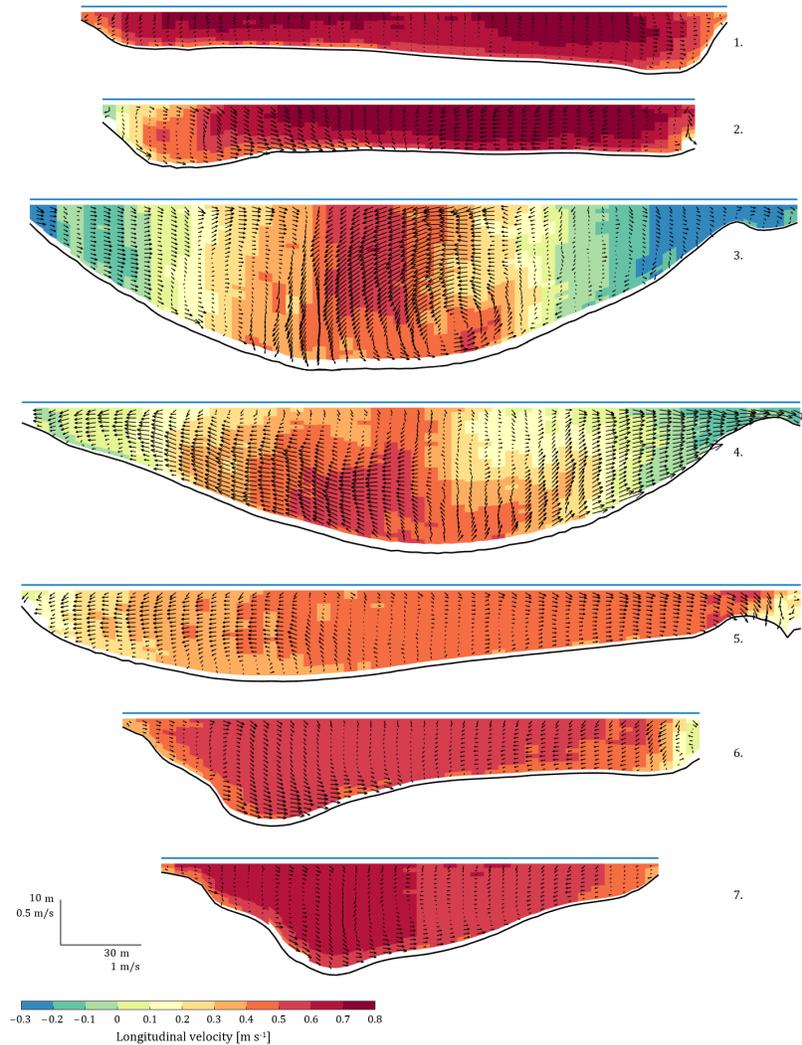
T=3: $h_{max} = 39.94$ m, $h_{avg} = 26.49$ m, $u_{max} = 0.60$ m s^{-1}, $|v_{max}| = 0.28$ m s^{-1} and $|w_{max}| = 0.12$ m s^{-1}.

T=4: $h_{max} = 35.29$ m, $h_{avg} = 22.25$ m, $u_{max} = 0.58$ m s^{-1}, $|v_{max}| = 0.38$ m s^{-1} and $|w_{max}| = 0.10$ m s^{-1}.

T=5: $h_{max} = 22.54$ m, $h_{avg} = 16.40$ m, $u_{max} = 0.54$ m s^{-1}, $|v_{max}| = 0.15$ m s^{-1} and $|w_{max}| = 0.10$ m s^{-1}.

T=6: $h_{max} = 26.36$ m, $h_{avg} = 16.56$ m, $u_{max} = 0.60$ m s^{-1}, $|v_{max}| = 0.21$ m s^{-1} and $|w_{max}| = 0.06$ m s^{-1}.

T=7: $h_{max} = 27.30$ m, $h_{avg} = 16.06$ m, $u_{max} = 0.68$ m s^{-1}, $|v_{max}| = 0.16$ m s^{-1} and $|w_{max}| = 0.07$ m s^{-1}.

At transect 3 (Figure 4.2), located at the scour hole, the longitudinal flow is concentrated at the center, with two recirculating zones next to it. About 75% of the cross-sectional area and about 50% of the measured river width is discharging downstream. The remain part flows upstream along the edges mainly caused by the sudden increase in the cross-sectional area that results in an adverse water surface gradient (Vermeulen et al., 2015). Vertical shear

layers separate the longitudinal flow and circulating flow. Secondary flow is clearly visible and is directed downwards to the scour hole. Near the center, a strong downward motion is visible where two circulation cells are present next to it. One secondary cell is observed near the bed at the inner bank side. The other one is weaker, counter-rotating, and more located to the water surface in the outer bend.

The area with strongest longitudinal flow at transect 4 (Figure 4.2, p.26) is located just slightly towards the outer bank of the river bend instead of the center, which is expected because of the outward shift due to spiral motion in river bends. The strongest flow is besides the horizontal outward shift also shifted downward, which is in fact not logical because the strongest flow appears far from solid surfaces generally, *i.e.,* the upper center. However, momentum is concentrated near the bed because of the strong downward flow at the scour and the upward sloping bed. For that reason, the core of the flow is located near the bed. About 90% of the cross-sectional area and about 75% of the measured river width is discharging downstream. The upstream flow is especially located at the inner bend. Secondary flow is directed outwards particularly and is separated in the center of the section. The inner half is directed to the inner bank into the outflow reach of the bend and the outer half is directed to the outer bank.

The longitudinal flow is nearly uniform over transect 5 (Figure 4.2, p.26). The secondary flow is directed outwards laterally, *i.e.,* inner half to the inner bank and outer half to the outer bank. Furthermore, the flow is directed downward at the inner bank due to impinging. More to the outer bank from the center and at the deepest location in the profile a circulation cell is present near the bed. According to Vermeulen et al. (2015) is the flow restored to a typical mild bend flow pattern at transect 6 and 7 (Figure 4.2, p.26), with deeper regions in the outer bend due to scouring at higher velocities.

The downward directed flow vectors in the topmost cells of the transects in Figure 4.2, p.26 are, according to Vermeulen et al. (2015), a measuring artifact probably caused by instrumental noise.

Spatial scales and patterns are hard to determine and describe uniformly because they are changing for each specific transect and the distinction is related to the used colour scale for the longitudinal velocity. However, spatial scales in the cross-sectional velocity data are attempting to quantify in Appendix A. It provides maximum and minimum observed longitudinal velocities (u_{max} and u_{min}) in the processed data and quantifies the spatial flow patterns that carries the strongest flow. Furthermore, the spatial scales in the lateral and vertical velocity components are provided. In general, the area that carries the strongest longitudinal flow is relatively large and located in the outer bend. However, the area of stronger flow is relatively small and located in the center at the scour hole (transect 3 and 4). In all transects is the strongest longitudinal flow present over almost whole water depth. Areas that contain strongest lateral and vertical flow are separated small regions instead of large continuous regions as is the

case for the longitudinal flow. Stronger lateral velocities occurs at more deeper locations and at the outer edges generally. Strongest vertical velocity occurs near the water surface but results from the measuring artifact. However, strong vertical velocity is also observed near the bed at deeper locations. Properties of raw velocity data are included in Appendix A in addition to the properties of processed data, since the model fit is based hereupon.

4.2   Model set up

Within this section, the model set up is shown for the first transect. Transect 1 is measured 17 times over a width of 240 meters and has a total of 1,776 allocated ensembles from the measured track. All measured velocities in these ensembles will be projected on the average transect. However, these velocities consist still of only components in x-, y-, and z-direction. The angle between the average transect and the x-axis is equal to the angle orthogonal to the average transect and the y-axis, and is represented by α. This angle is used in order to compute the velocity in the longitudinal (s) and lateral (n) direction. Given that $\alpha = -20.1239°$ for the first transect, the unit vectors can be computed with the help of Eq.(3.1) and (3.2), which are visualised in Figure 4.3. Next, the longitudinal and lateral velocity component are determined with Eq.(3.3) and (3.4). All measured velocity locations in the first transect are shown in Figure 4.4. From this figure can be noticed that the velocity locations are evenly distributed due to the bin size.
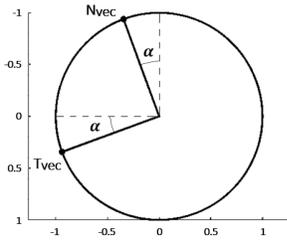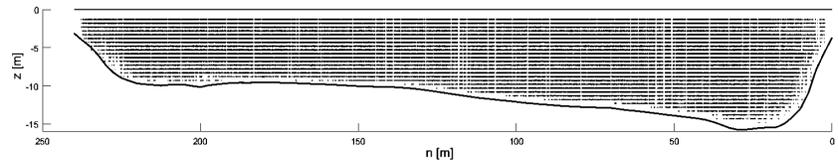

Figure 4.3: Unit circle with the normal unit vector (N_{vec}) and tangential unit vector (T_{vec}) for the first transect.

Figure 4.4: Measured velocity locations in the nz-plane (transect 1), for which the longitudinal, lateral and vertical velocity can be shown using an appropriate colour scale.

The river width at the water surface (b_s) will be determined to normalise the width and obtain the relative velocity locations, $i.e.,$ normalised lateral coordinate (ζ). The measured width (b_m) in transect 1 is 240 meters. By extrapolation of the transect to the river banks the actual surface width will be obtained. The left river bank is found at 145.59 meters from the centerline and the right bank at 131.08 meters from the centerline. This results in a surface width of 276.67 meters for the moment of measuring the river banks, which originates from another source, as mentioned. The normalised elevation above river bed, $i.e.,$ normalised vertical coordinate (σ), of the measured velocity locations are already present in the data set, which can be used. With the help of Eq.(3.5) and (3.6) the velocity locations are transformed into the normalised ζ, σ-system. All measured velocities locations in this normalised coordinate system for the first transect are plotted in Figure 4.5. Here, it becomes clear that the measured velocities near the bed, $i.e.,$ lowest 6%, is removed to account for side lobe interference. The data is "rounded" at the edges because the blanking
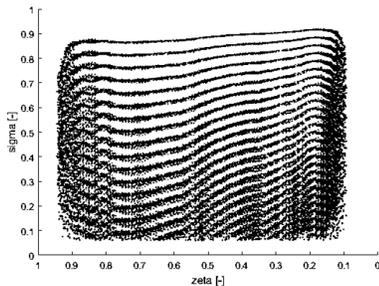

Figure 4.5: Normalised coordinates for all velocity locations in transect 1, without showing the actual magnitude of the measured velocity.

distance is increasing relatively for smaller water depths, *i.e.,* it becomes a bigger part of the water depth locally. Furthermore, the density of ensembles is higher at the edges due to manoeuvring reasons.

The normalised system is investigated for spectral behaviour by applying Eq.(3.7), (3.8) and (3.9) to approximate the longitudinal, lateral and vertical velocity component respectively. A general MATLAB script of this analysis is provided in Appendix B. The principle of adding more terms to the velocity fit is shown in Figure 4.6, over both the normalised river width and water depth. Note that all terms with different mode numbers (m and n) below the truncation numbers (M and N), and so all combinations of modes, have their own addition to the velocity fit. Velocity components are fitted to the measured velocity with more accuracy by increasing truncation numbers. It is possible to extrapolate the velocity through the normalised domain with this spectral analysis and magnitudes of velocity can for that reason be assessed near the actual measured locations as well. Amplitudes for the model fit are computed with the `lscov` function in MATLAB, which returns the least-squares solution of fitting the trigonometric term with the measured velocity. Truncation numbers forms the variable input and the computed amplitudes are the variable output of the model, because it will fit the measured velocity with the trigonometric terms (that change with mode numbers). However, the computed amplitudes are not final since they change with truncation numbers. The velocity is modelled more accurate by including more modes in the model fit because each combination reshapes the model fit, where the combination of all modes results in the least-squares solution. Furthermore, the maximum truncation number, and so the maximum complexity of the model fit, is different for each transect (see Table 4.1). For truncating above a certain number, the model indicates a warning with respect to rank deficiency. This warning is obtained as measured velocity data contain insufficient information to generate the desired model complexity.

Figure 4.6: Principle of increasing truncation numbers for longitudinal velocity fit. Measured velocity ($\hat{u}$) in grey and model fit ($\tilde{u}$) in red for (a) $M = N = 1$, (b) $M = N = 2$ and (c) $M = N = 3$.

Table 4.1: Maximum M and N per transect for truncating equally.

T	$M_{max} = N_{max}$
1	54
2	38
3	48
4	38
5	66
6	49
7	46

4.3   Dominant spatial scales

Truncation numbers are increased by taking $M = N$ to investigate spatial scales in the cross-sectional velocity data. First, the influence of truncation numbers on the dominant scales in the longitudinal velocity fit is analysed. Here, transect 1 is discussed in detail. Second, the dominant scales in the three velocity components and among the different transects are compared briefly. More visualisations to support and clarify this analysis are provided in Appendix C.

4.3.1   Influence of truncation numbers

The influence of increasing truncation numbers on the dominant spatial scales in cross-sectional velocity data is analysed because the amplitudes, and so the dominant scales, vary for different truncation numbers. Starting with $M = N = 1$ that represents the flow field generally, with zero velocity at the solid surface
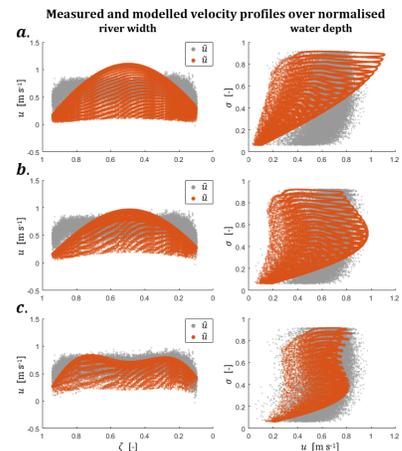
and high velocity far from these boundaries. The computed amplitude is 1.13 m s^{-1}, which is higher than the maximum measured longitudinal velocity but resulted from the best solution of the least-squares fit. Far from solid boundaries is the velocity overestimated by the model fit but underestimated near the solid boundaries over both normalised river width and water depth. Although the velocity field is modelled with more accuracy by increasing truncation numbers, the first mode adds most significant value to the model fit up to truncating at $M = N = 18$ (see Appendix C).

Specific dominant spatial scales include more a certain dominant region, which become visible by increasing the modes further. When the model fit is truncated at $M = N = 21$, two dominant regions are clearly visible around $m = 4$ and $m = 12$ (see Figure 4.7). These regions correspond to wave lengths of 0.5 and 0.1667 times the measured river width respectively. In other words, two and six numbers of waveforms over the width. Furthermore, it should be noticed that the first mode of m is still quite dominant as well and that the region around $m = 19$ adds a significant part to the model fit. These regions extend up to about $n = 10$ vertically, which corresponds to wave lengths of about 0.21 times the local water depth, so almost five numbers of waveforms over depth.

Four dominant regions are observed by truncating at $M = N = 36$ (see Figure 4.8), namely around $m = 4, 10, 16$ and 22. These mode numbers correspond to wave lengths of 0.5, 0.2, 0.1250 and 0.0909 times the measured river width. In addition to these four regions a fifth region is slightly visible around $m = 28$. Note that the steps between the mode numbers with a region of strong amplitudes are approximately equal, resulting in a repetitive pattern. However, the intermediate modes have less, but still significant, contribution to the resulting flow field. In general, these regions extend up to about $n = 13$ vertically, corresponding to wave lengths in the order of magnitude of 0.16 times the local water depth, *i.e.,* more than six numbers of waveforms over depth. Furthermore, it should be noticed that the values for the amplitudes in the dominant regions are exceptionally high compared with lower truncation numbers. This is caused by trying to fit to the relative high velocities near the bed. Since the longitudinal velocity profile over water depth is more *log*-shaped, it experiences a more sudden increase than a general sine-function. Higher amplitude values allow the sine-based model fit to adjust to this *log*-shaped pattern. The values next to a relatively high amplitude value show relatively low values to compensate and not overestimate the *log*-shaped flow.

Including modes up to $M = N = 52$ shows one clear region (see Figure 4.9), which is located around $m = 14$. This corresponds to seven waveforms over river width. It is mainly ranging down to $m = 5$ and up to $m = 25$. Vertically the region extends up to $n = 12$, corresponding to wave lengths of about 0.175 times the local water depth, *i.e.,* almost six numbers of waveforms. The values of the strongest amplitudes become here of the order 10^7.
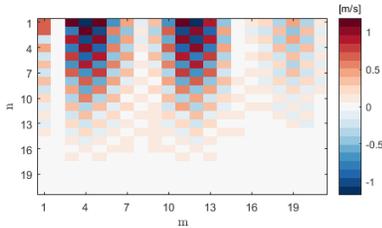

Figure 4.7: Values for amplitudes for the longitudinal velocity fit in transect 1, truncating at $M = N = 21$.
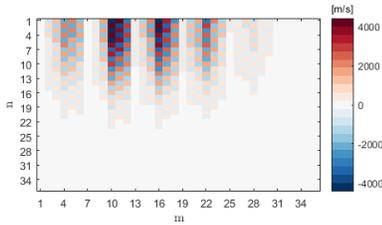

Figure 4.8: Values for amplitudes for the longitudinal velocity fit in transect 1, truncating at $M = N = 36$.
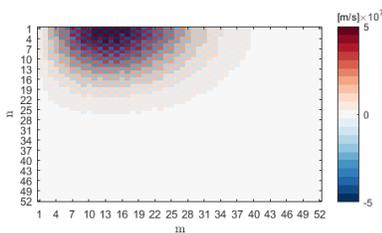

Figure 4.9: Values for amplitudes for the longitudinal velocity fit in transect 1, truncating at $M = N = 52$.

Dominant spatial scales are observed by analysing the longitudinal velocity data in the first transect with Fourier transform. However, as can be noticed from descriptions above, these dominant scales vary for different truncation numbers. In general, the modes which represent the domination are equivalent to:

- waveforms with a length of 0.1-0.5 times the local river width horizontally, *i.e.,* two to ten waveforms.
- waveforms with a length of 0.15-0.25 times the local water depth vertically, *i.e.,* about four to seven waveforms.

4.3.2   *Comparing velocity components and different transects*

Specific dominant scales for the lateral and vertical velocity fit are more difficult to distinguish, but dominant regions can be observed. For truncating at $M = N = 16$, the dominant regions over river width in the lateral velocity fit are located around $m = 1, 2, 4, 5, 9$ and 13 and extend vertically up to about $n = 9$ (see Figure 4.10a). The dominant mode in the vertical velocity fit is comparable to the longitudinal fit, namely $m = 1$. However, it extends up to about $n = 8$ vertically. In addition, more combination of modes have significant added value as can be noticed from Figure 4.10b.
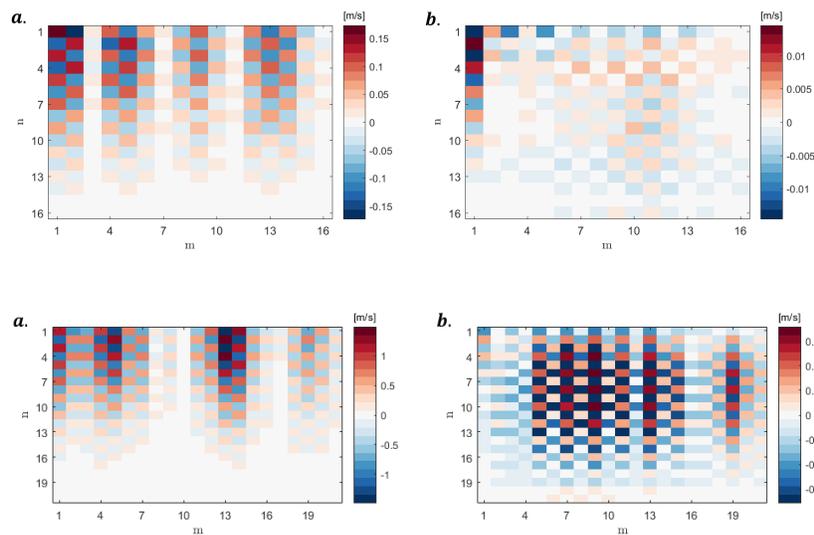

Figure 4.10: Values for amplitudes in transect 1 truncating at $M = N = 16$ in fitting (a) the lateral and (b) the vertical velocity component.
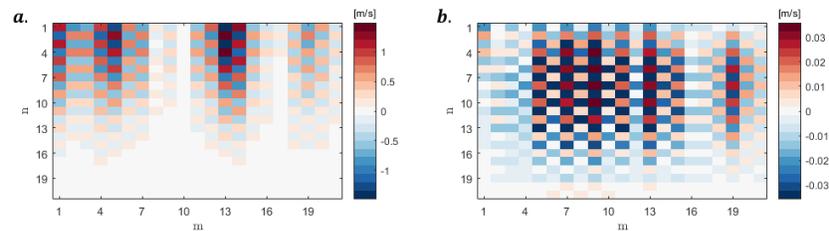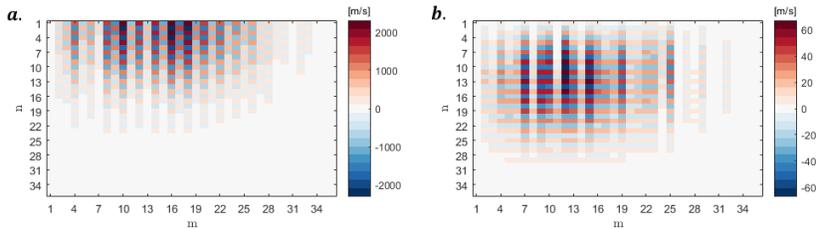

Figure 4.11: Values for amplitudes in transect 1 truncating at $M = N = 21$ in fitting (a) the lateral and (b) the vertical velocity component.

For truncating at $M = N = 21$ is the domination, *i.e.,* strongest amplitudes, in the lateral velocity fit somewhat comparable to the longitudinal fit (see Figure 4.11a). Dominant regions are located around $m = 1, 5, 13$ and 19 and extend up vertically to about $n = 10$. The dominant modes in the vertical velocity fit are more randomly distributed accross the domain. Actually, one dominant region is observed from Figure 4.11b. However, it can be noticed that odd modes add more value to the vertical velocity fit than even modes. The domination extends up to about $n = 17$.

Dominant regions for the lateral and vertical velocity fit are not comparable with the longitudinal velocity fit for truncating at $M = N = 36$. One region is observed for the lateral velocity fit around $m = 16$, which ranges down to $m = 4$ and up to $m = 25$. The region extends up to about $n = 13$ as can be noticed from Figure 4.12a. However, this region is not consecutively since some modes not really participate in the contribution to the model fit. In the vertical velocity fit is one dominant region observed between $m = 4$ and $m = 25$ (see Figure 4.12b). However, here is the presence of less participating modes in m to the model fit observed as well. Furthermore, the lower modes in n are not located in the dominant region. The dominant region ranges approximately between $n = 4$ and $n = 25$ vertically.

Figure 4.12: Values for amplitudes in transect 1 truncating at $M = N = 36$ in fitting (a) the lateral and (b) the vertical velocity component.

From Figure 4.13 can be noticed that a dominant region is present in both the lateral and vertical velocity fit for truncating at $M = N = 52$. Moreover, it is comparable with the regions in the longitudinal velocity fit because the domination is located in approximately the same region. Stronger amplitudes to model the lateral and vertical velocity component are observed for odd modes over m.

Figure 4.13: Values for amplitudes in transect 1 truncating at $M = N = 52$ in fitting (a) the lateral and (b) the vertical velocity component.
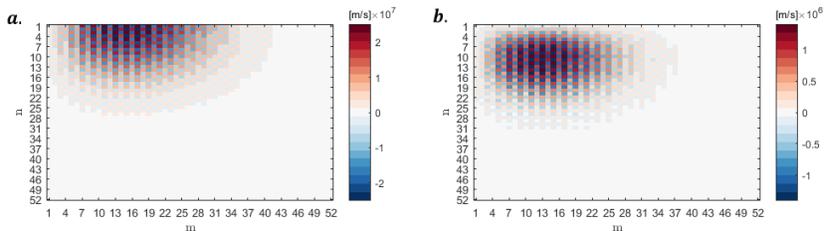
The location of the dominant regions in other transects are comparable to each other for relatively high truncation numbers in that transect. For lower truncation numbers (*e.g., $M = N = 16$ and $M = N = 21$*) the domination in the different transect varies significantly. In some transects is a repetitive pattern visible in others one clear region. However, the dominant spatial scales are comparable as in transect 1. So, waveforms with a length of 0.1-0.5 times the river width and 0.15 and 0.25 times the water depth generally. The fact that the dominant region in the vertical velocity fit (for relatively high truncation numbers) include higher modes for n and exclude to lower ones is visible throughout all transects (see Appendix C).

4.4   Performance of model fit

The model fit is evaluated for increasing truncation numbers. Interesting numbers of truncation for both M and N in the first transect for the longitudinal velocity fit are: 1, 4, 8, 10, 12, 16, 21, 36, and 52, which are included in Appendix D. These numbers are interesting because of their improvement in the model fit, presence of dominant scales and/or to show the principle of "overfitting". Features and behaviour in the longitudinal velocity fit for some of these truncation numbers are highlighted and explained in this section. Starting with describing the improvement of the model fit over normalised width and depth for increased truncation. Subsequently, the spatial structure and possible anomalies in the residuals are evaluated. The performance of the model fit for increased truncation numbers is evaluated with measures of accuracy. In the end, the velocity is interpolated between the measured locations and extrapolated to the boundaries.

4.4.1   Improvement of velocity fit over normalised width ζ

The truncation number is adapted in the model to evaluate the performance of modelling the three-dimensional velocity field. Here, only the longitudinal velocity fit is presented in order to support the results. Appendix D includes the lateral and vertical velocity fit as well. As can be noticed from Figure 4.14, the modelled longitudinal velocity approximates the measured velocity with more accuracy by increasing the truncation numbers. The depth averaged velocity over river width of the model fit ($\bar{u}_h$ of fit, red line in Figure 4.14) gets closer to the one from the measured velocity ($\bar{u}_h$ of meas., grey line). From the first plot in Figure 4.14 can be observed that the velocity is underestimated near the edges and overestimated in the center. As the truncation numbers are increased, the velocity is underestimated especially by the model fit. The underestimated part is in particular the presence of turbulence, with a magnitude in the order of 0.1-0.2 m s^{-1}, on top of the main flow. Increasing the truncation numbers in order to include relatively high mode numbers, so increasing to $e.g.$, $M = N = 52$, the velocity is been fitted for turbulence as well. This principle of "overfitting" can be noticed from the lowest plot in Figure 4.14. Here, many number of waveforms (with small wave lengths) over river width and water depth are included in the fit, which results in adaptations of the model fit to the measured velocity at small scales. As the truncation is adjusted from $M = N = 10$ to $M = N = 16$, the difference is hardly visible. Truncating between these modes seems to result in an adequate representation of the main flow.
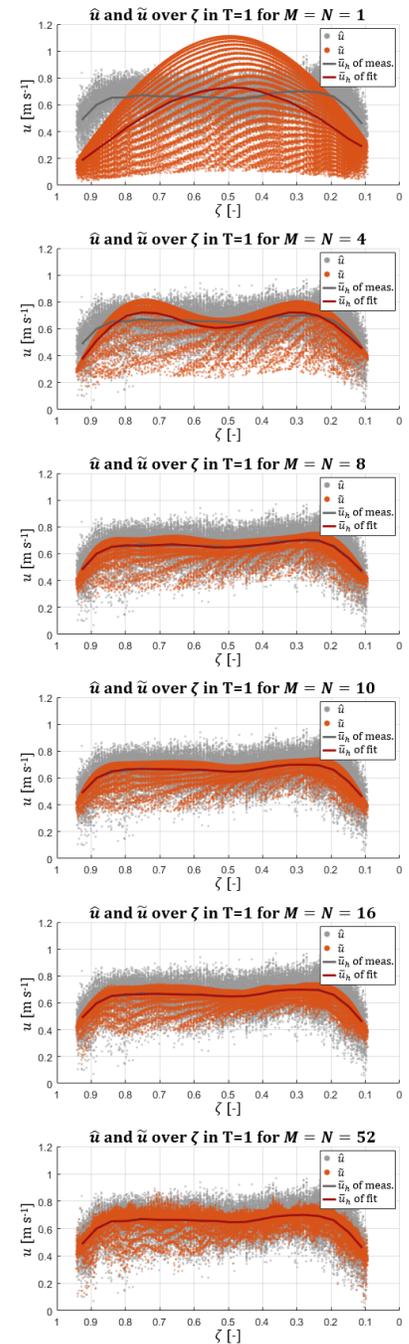

Figure 4.14: Measured and modelled longitudinal velocity over river width with depth averaged velocity over width for different truncation numbers M and N.
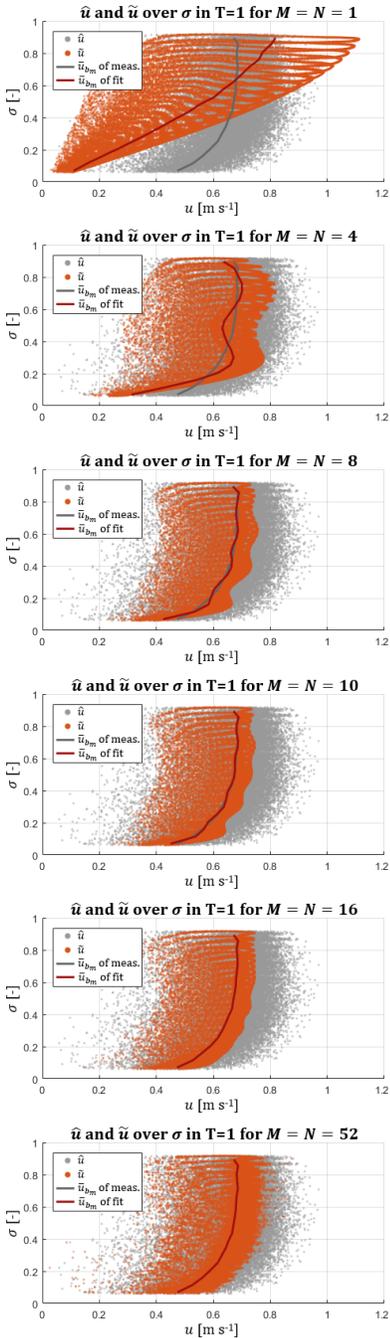
Figure 4.15: Measured and modelled longitudinal velocity over water depth with width averaged velocity over depth for different truncation numbers M and N.

Features over normalised river width of fitting the lateral and vertical velocity components, noticed from Appendix D, are described briefly below:

- Both components can be modelled with more accuracy by increasing the truncation numbers. However, both fits underestimate the actual velocity due to the presence of positive and negative flow in lateral and vertical direction, which results in small amplitude values.
- Both components are "flattened" in the model fit. Measured velocities variates more locally than the model fits, which complicates the recognisability. Velocity fields in the normalised coordinate system (in Appendix D) shows more recognisable flow patterns for increased truncation numbers, which are however still flattened.
- The depth averaged velocity of the lateral and vertical measured and fitted velocity becomes almost equal for truncating at $M = N = 16$.

4.4.2 Improvement of velocity fit over normalised depth σ

The longitudinal velocity fit over normalised depth is also underestimated near the boundary, *i.e.,* river bed, and overestimated at the top, *i.e.,* water surface, for the first mode. However, the fitted velocity approximates the measured velocity with more accuracy by including more modes as can be noticed from Figure 4.15. The width averaged velocity over water depth of the model fit ($\bar{u}_{b_m}$ of fit, red line in Figure 4.15) gets closer to the one from the measured velocity ($\bar{u}_{b_m}$ of meas., grey line). Furthermore, relatively high truncation numbers show here fitting to turbulence as well, which is not desired. Truncating at lower modes, such as 4, 8 and 10 for both M and N show oscillatory patterns for the velocity fit. The width averaged velocity of the fit oscillates around the width averaged velocity of measured data as well. This oscillatory pattern is largely disappeared for the velocity fit that is truncated at $M = N = 16$.

Features over normalised water depth of fitting the lateral and vertical velocity components, noticed from Appendix D, are described briefly below:

- Both components can be modelled with more accuracy by increasing the truncation numbers and are "flattened" in the model fit. Just as is the case as described for the improvement of the lateral and vertical velocity fit over normalised width.
- The width averaged velocity of the measured velocity and the fit over water depth for both the lateral and vertical component are becoming nearly equal relatively fast (for truncating at around $M = N = 10$).

4.4.3 Evaluation of residuals

Residuals are decreasing as truncation numbers increase, which can also indirectly be noticed from previous sections where the model fit improves for increasing truncation numbers. The residuals in the first transect for the longitudinal velocity fit with different truncation numbers are plotted back against the

normalised width and depth in Figure 4.16 and 4.17 in order to investigate the spatial structure in these residuals. Positive residuals indicate underestimated actual velocities by the model fit and negative values indicate overestimating. The density of the residuals, and so the main error region or its course, is visualised with the help of plotting with transparency. The local averages of the residuals over both normalised river width and water depth ($\bar{\epsilon}_j$ and $\bar{\epsilon}_k$ as described in §3.2.4) are plotted with the white line. Furthermore, the minimum and maximum values of the local average are indicated with red dots and the values are presented in the lower left corner of each plot. The root mean squared error over normalised width and depth ($RMSE_\zeta$ and $RMSE_\sigma$, which are used evaluating the model performance in next section) are presented in the lower right corner.

Figure 4.16: Residuals of longitudinal velocity fit over normalised width and depth for truncating at $M = N = 1, 4$ and 8.

Starting with $M = N = 1$, that is limited to model the flow field by only one trigonometric term for horizontal behaviour and one for the vertical behaviour. As can be noticed from Figure 4.16, there is spatial structure in the residuals of the longitudinal velocity fit both over normalised width as the depth. The underestimated velocities near the bed results in a sort of tail. The tail is the result for the lack of fitting to already relative high velocities near the bed, *i.e.,* the *log*-shaped flow, which indicates the presence of a relative smooth bed.

Relatively high errors are present, particularly at the edges of the measurement, for truncating at $M = N = 4$. There is spatial structure in the residuals over both normalised width and depth, which means that adding more modes in necessary for an adequate model fit of the velocity. The velocity fit for truncating at $M = N = 8$ shows already less spatial structure, but especially the tail in the lower region of normalised water depth is still rapidly recognised. By analysing further with increased truncation numbers, this tail near the bed remains visible, although to a smaller extent, up to $M = N = 14$ and is disappeared for $M = N = 16$ (see Figure 4.17). Disappearing of the tail indicates that the model is able to fit to the *log*-shaped flow pattern.

The residuals over normalised width are relatively low for truncating at $M = N = 16$ and higher, showing more variance at the edges of the measurement. Spatial structure is not noticed from the line that represents the local mean for truncating at $M = N = 16$ (see Figure 4.17). The residuals over normalised water depth show a clear vertical belt, indicates equal variances and no spatial structure. The added value of fitting with relative high mode numbers (*e.g.,* $M = N = 52$) is limited relative to the fit for truncating at $M = N = 16$ as can be noticed from Figure 4.17.
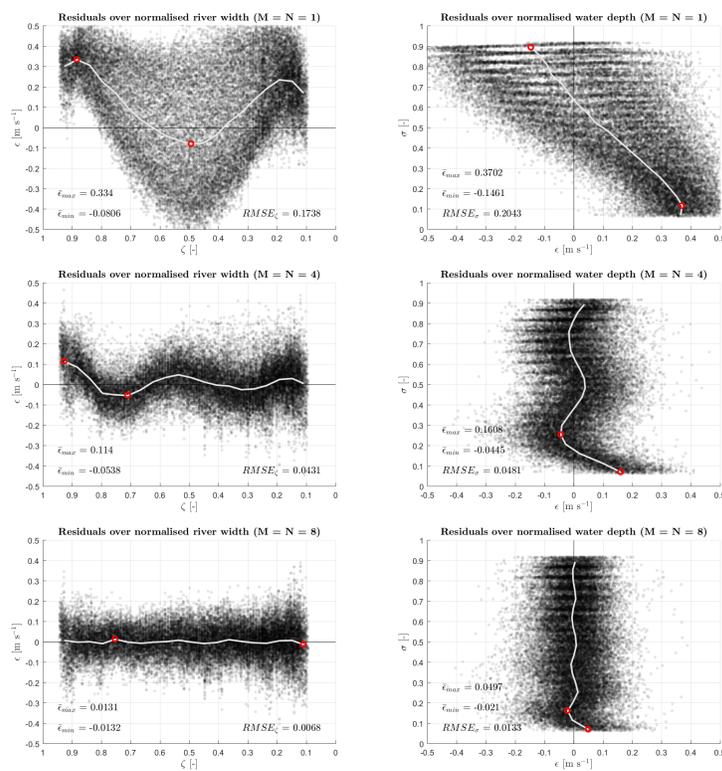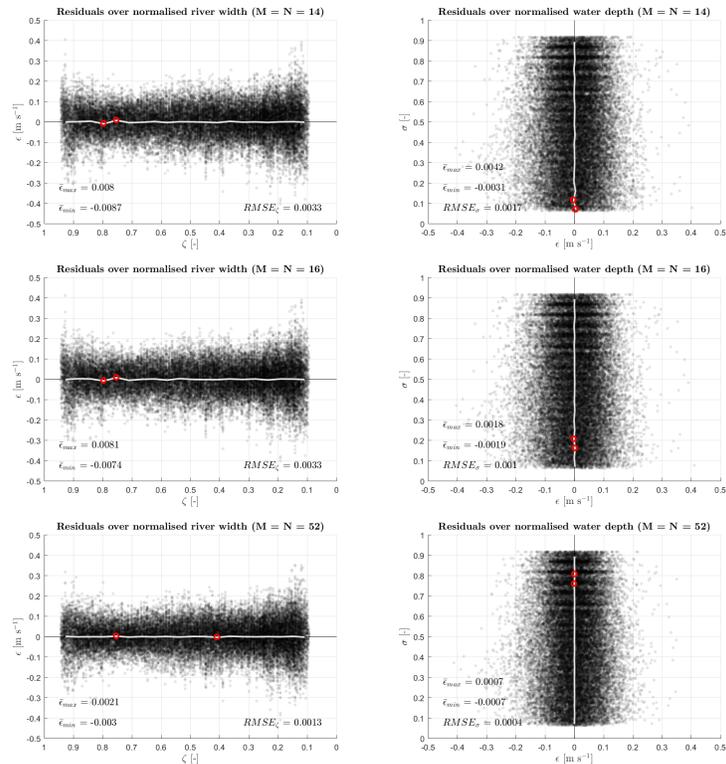
Figure 4.17: Residuals of longitudinal velocity fit over normalised width and depth for truncating at $M = N = 14, 16$ and 52.

Features in evaluating the residuals of the lateral and vertical velocity fit, noticed from Appendix D, are described briefly below:

- Approximately the same patterns in residuals for lateral velocity fit are observed as in the longitudinal fit, with higher variability at the edges of the measurement over normalised width. However, the tail and a clearly spatial structure are not present.
- The residuals in the vertical velocity fit are not clearly comparable to the residuals longitudinal and lateral fit. The variance in residuals over width is higher at the inner bend, but at the other edge it is more or less equal than the other parts.
- Generally, the spatial structure in the lateral and vertical velocity fit is disappeared before truncating it at $M = N = 16$.

In addition of evaluating residuals over width and depth, the residuals have been evaluated in order to detect possible anomalies. Hence, the residuals are plotted against the measured longitudinal velocity (first column) and the longitudinal velocity fit (second column) in Figure 4.18. The measured and modelled longitudinal velocity are plotted against each other as well in the third column of Figure 4.18, where the diagonal indicates a perfect match.
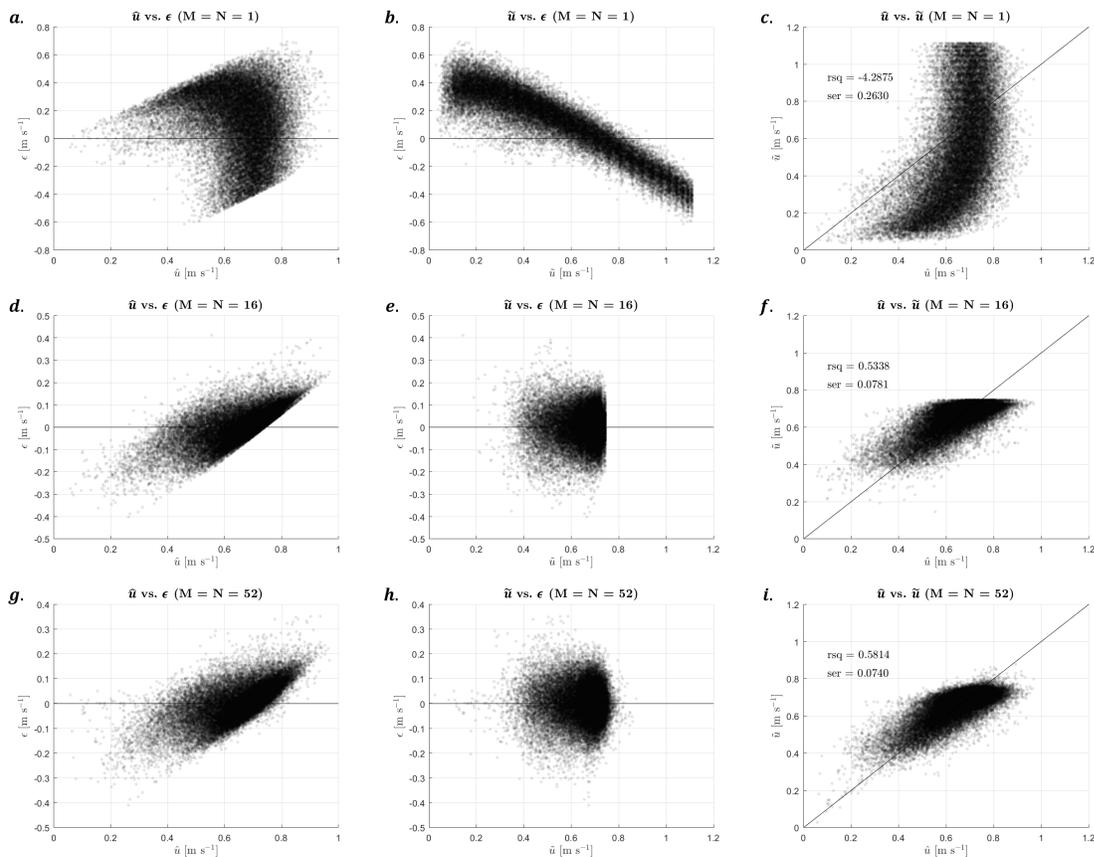

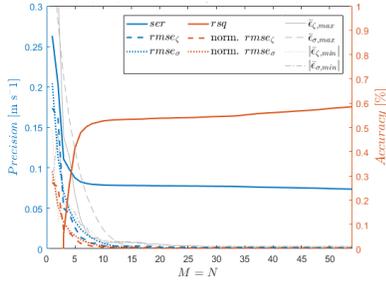Figure 4.18: Residual plots for truncating the longitudinal velocity fit at $M = N = 1, 16$ and 52.

Figure 4.19: Accuracy and precision of longitudinal velocity fit ($\tilde{u}$) in transect 1. *ser* is the standard error of regression (in blue), *rmse* the root mean squared error (in blue), which is normalised as well (in red), *rsq* is R^2 (in red). Maximum and minimum local averages of residuals ($\bar{\epsilon}_{\zeta/\sigma,max/min}$) are plotted in grey and pertain to the blue *y*-axis.
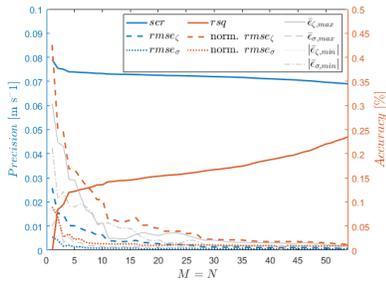

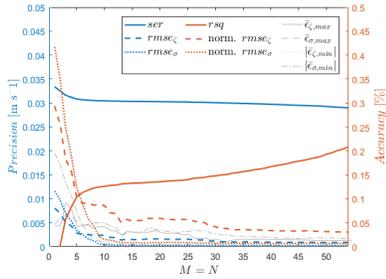Figure 4.20: Accuracy and precision of lateral velocity fit ($\tilde{v}$) in transect 1.

Figure 4.21: Accuracy and precision of vertical velocity fit ($\tilde{w}$) in transect 1.

From Figure 4.18 can be observed that the longitudinal velocity fit for truncating at $M = N = 1$ have a systematic pattern, which suggests that including more terms is necessary. Furthermore, the variability in residuals increases as the measured velocity increases (see Figure 4.18a). Low modelled velocity is underestimated, and higher velocities are overestimated as noticed from Figure 4.18b. Increasing truncation numbers to $M = N = 16$ results in overestimating low measured velocities by the model fit and underestimating high measured velocities in general (see Figure 4.18d). Most of the systematic pattern is disappeared, but the maximum possible modelled velocity is clearly visible (a sudden cut-off) in Figure 4.18e and f. Plotting the measured velocities against the modelled velocities results in a belt around the line of perfect match, which indicates equal variance. From Figure 4.18 can also be determined that the added value of increasing the truncation numbers in the velocity fit from $M = N = 16$ to $M = N = 52$ is relatively low.

4.4.4 Accuracy and precision of model fit

The course of accuracy and precision for modelling the different velocity components are shown in Figure 4.19, 4.20 and 4.21. Accuracy is in percentages and represents the degree to which the modelled velocity conforms to the measured velocity or becoming more accurate by increased truncation numbers. Precision represents the refinement in the model fit for increased truncation numbers in m s^{-1}. The computed coefficient of determination (R^2) indicates that the model is able to explain more than 58% of the variability of the response data around its mean for the longitudinal velocity fit in transect 1. The standard error of the regression (SER) is a measure of precision in the model fit. The average distance of the measured velocity data from the modelled velocity is $7.37 \cdot 10^{-2}$ m s^{-1} in the longitudinal velocity fit for truncating at $M = N = 54$. These are the highest truncation numbers before the model warns for rank deficiency. The root mean squared errors over normalised depth and width are decreasing relatively fast. The potential values (for truncating at $M = N = 54$) are shown in Table 4.2 and computed with the help of Eq.(3.11) mentioned in §3.2.3.

Table 4.2: Accuracy / precision of model fit in transect 1 for truncating at $M = N = 54$

Measure	u	v	w	Unit
R^2	58.52	23.65	21.04	[%]
SER	$7.37 \cdot 10^{-2}$	$6.90 \cdot 10^{-2}$	$2.90 \cdot 10^{-2}$	[m s^{-1}]
$RMSE_\zeta$	$1.36 \cdot 10^{-3}$	$7.13 \cdot 10^{-4}$	$8.33 \cdot 10^{-4}$	[m s^{-1}]
$RMSE_\sigma$	$3.67 \cdot 10^{-4}$	$6.26 \cdot 10^{-4}$	$1.84 \cdot 10^{-4}$	[m s^{-1}]

However, the actual values of *RMS E* are also normalised in order to account for its error relative to its own magnitude. These are computed by dividing $RMS E_\zeta$ by the depth averaged velocity in a specific direction over normalised river width and $RMS E_\sigma$ by the width averaged velocity in a specific direction over normalised water depth. These are plotted in red in Figure 4.19, 4.20 and 4.21, p.38. Note that the scales on the *y*-axes in these plots are changing for each velocity component.

From Figure 4.19, p.38 can be noticed that the measures of accuracy and precision for the longitudinal velocity fit has reached a sort of equilibrium state. Including more modes does not result in explaining more variance or more precise estimates. This point is here located between $M = N = 10$ and 20. An equilibrium state is not present within the lateral and vertical velocity fit, especially not for R^2 (see Figure 4.20 and 4.21, p.38). Including more modes seems to result in better estimates since the slope of R^2 increases at the end. The computed errors lowers as truncation numbers increase. However, for the errors, it seems that the equilibrium state is reached lately for higher truncation numbers (around $M = N = 40$). This can be coupled to higher variability in velocity locally, which cannot be smoothed due to its instantaneous sensitivity.

Potential values for measures of accuracy and precision are different for each transect, but the course of lines show comparable behaviour generally. In particular, for the longitudinal fit, where most of the variability is disappeared after about $M = N = 15$ and become not significant more accurate nor precise. By checking the accuracy and precision for the lateral and vertical velocity fit, other transects come to an equilibrium neither.

4.4.5 Inter- and extrapolation of velocity

The potential of this study is to interpolate the flow field between the measured velocity locations and extrapolate towards the boundaries. The flow field can be extended because the trigonometric terms are already determined, and the amplitudes are computed. So, the results from the Fourier transform can be used to fill the complete normalised domain with velocity data as shown in Figure 4.22.

Interpolation results in smooth patterns that seems recognisable compared with the described flow patterns in §4.1 for some set of truncation numbers. Especially for fits that are truncated between 4 and 10 for *M* and *N*. The extrapolated velocity in these cases show a general velocity pattern as well, because its magnitude decreases gradually towards the boundaries without strange features.

However, the interpolated velocity fluctuates between measured velocity locations for increased truncation, and in particular for modes with such a small wave length that (a half of) it fits in between two measured points. This interpolated velocity fluctuates especially vertically among points due to aliasing, which is already visible for truncating at $M = N = 21$ (see Figure 4.22). In
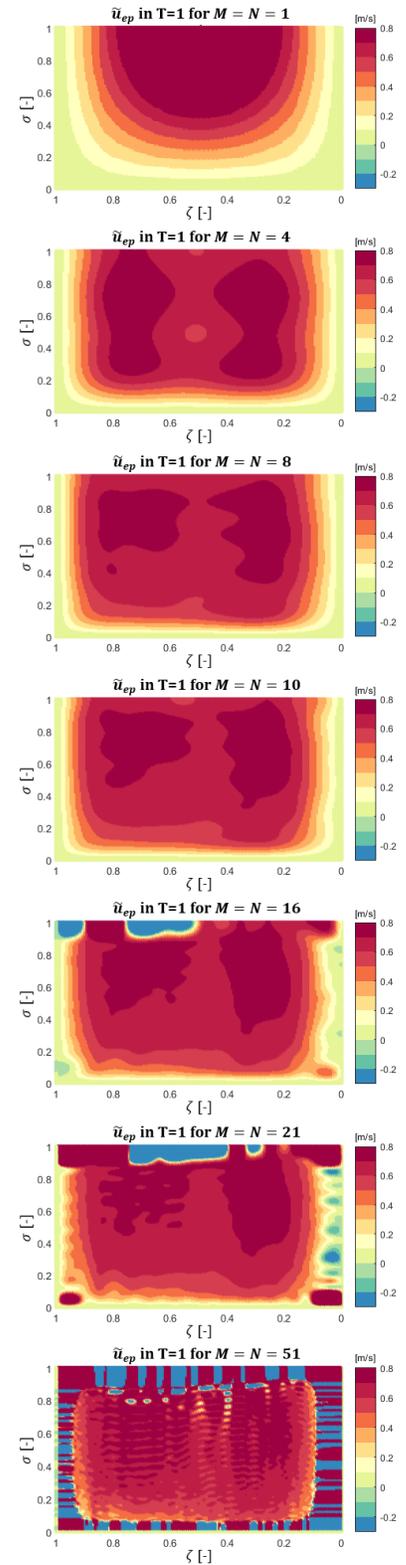

Figure 4.22: Normalised domain with extrapolated longitudinal velocity fit ($\tilde{u}_{ep}$) in transect 1 for truncating at 1, 4, 8, 10, 16, 21 and 52 for both *M* and *N*.

general, the measured locations show larger spacing vertically than horizontally due to the bin size of 50 cm. So, the vertical sampling distance between the velocity data is, in particular, too big for fitting with such high mode numbers.

Problems are encountered earlier in extrapolating the modelled velocities towards the boundaries (see Figure 4.22, p.39). This is due to the larger distance between the outermost measured velocity location and the boundary compared to the sampling distance. Truncating the model fit at $M = N = 10$ shows already a slightly weaker velocity region in a place (at the surface in the center) where it is not expected for unhindered flow. However, this difference is very minimal. But for truncating at even higher modes the extrapolated velocity shows enormous high and low magnitudes. Fitting with higher lengths of waveforms, *i.e.,* decreasing the number of waveforms in the model fit, offer better solutions for extrapolating the velocity.

4.4.6   *Transforming the model fit*

The normalised coordinate system will be transformed back to the *nz*-coordinates in order to compare the model fits with the processed data according the method of Vermeulen et al. (2014b). The processed velocity data is decomposed where each component is plotted with the use of colour shades in Figure 4.23, p.41. The decomposed velocity plots can be compared with the back transformed model fit in order to evaluate the features and patterns. In this section the transformation of the first transect is evaluated in detail. However, the third transect is also discussed since strong secondary flow occurs near the scour hole. Transforming the model fit from the normalised domain to the *nz*-plane results in Figure 4.24, p.41. It is chosen to truncate at mode 16 for both *M* and *N*. Over river width is the function truncated at $M = 16$ because the spatial structure is disappeared then. Over water depth is it truncated at $N = 16$ because oscillatory patterns are largely disappeared over normalised depth and the residuals show a clear vertical belt, so no spatial structure.

As mentioned in §4.1, the longitudinal velocity within the processed data shows a flow pattern with lower velocities at the edges and two cores of strong flow in the center (see Figure 4.23, p.41). Lateral flow is especially visible near the boundaries and is mainly directed towards the banks, (*i.e.,* outer part to outer bank and inner part to inner bank). Vertical flow is mainly downwards directed. The downward motion in the topmost cells is due to a measuring artifact. However, vertical flow is also present near the banks and at the deeper region. The processed data will be compared with the model fit that is truncated at $M = N = 16$.

The model fit shows comparable results as the processed data, with two cores of strong flow for the longitudinal velocity and decreasing in magnitude towards the boundaries. The lateral velocity is mainly directed towards the outer bank, and at the inner region a stream to the inner bank is observed. Lateral flow is overestimated at the deeper region in the model fit, which results in higher

velocities than the processed data locally. Vertically, the flow is downwards for the upper region, caused by a measuring artifact as mentioned. The flow patterns are pretty similar generally, even though processed data show coarse patterns where the model fit is more smoothed and generates a detailed flow field with the help of continuous functions.
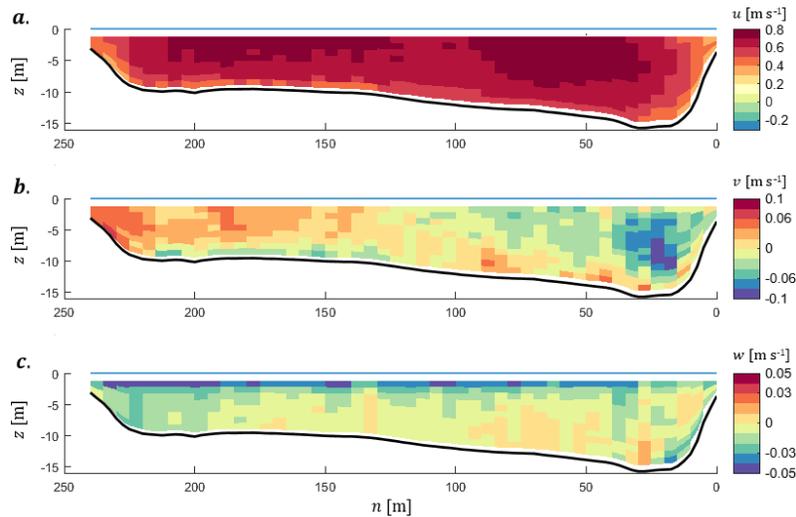
Figure 4.23: Decomposed velocity data of the first transect, which is processed according the proposed method of Vermeulen et al. (2014b). It shows the processed (a) longitudinal, (b) lateral and (c) vertical velocity.
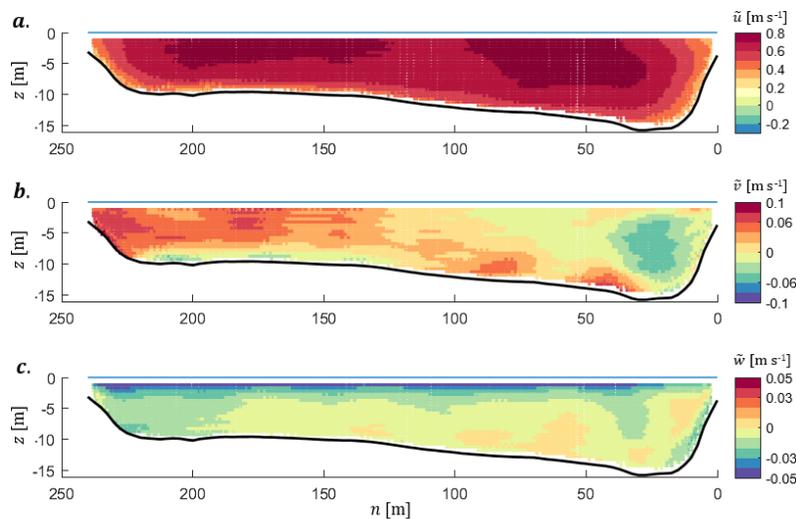
Figure 4.24: Modelled velocity in transect 1 for truncating at $M = N = 16$, where (a) shows the longitudinal velocity $\tilde{u}$, (b) the lateral velocity $\tilde{v}$ and (c) the vertical velocity $\tilde{w}$ in m s^{-1}.

Other transcects show comparable behaviour where the model fit is able to represent the flow field adequately for truncating at $M = N = 16$. Especially the longitudinal velocity can be approximated quite accurate. However, by evaluating other transects, and in particular one with relatively strong secondary flow (such as transect 3), the model encounters trouble in modelling the secondary flow, as can be noticed from Figure 4.25. The vertical velocity is modelled quite

accurate in the third transect for truncating at $M = N = 16$, but the pattern in the lateral velocity is disturbed and not comparable to the processed velocity. However, by analysing further increased truncation numbers, the lateral velocity in this transect is still difficult to model due to relatively large variability in the measured velocity data locally. The longitudinal velocity is overestimated in the center by the model fit but its pattern is almost similar to the processed velocity data.

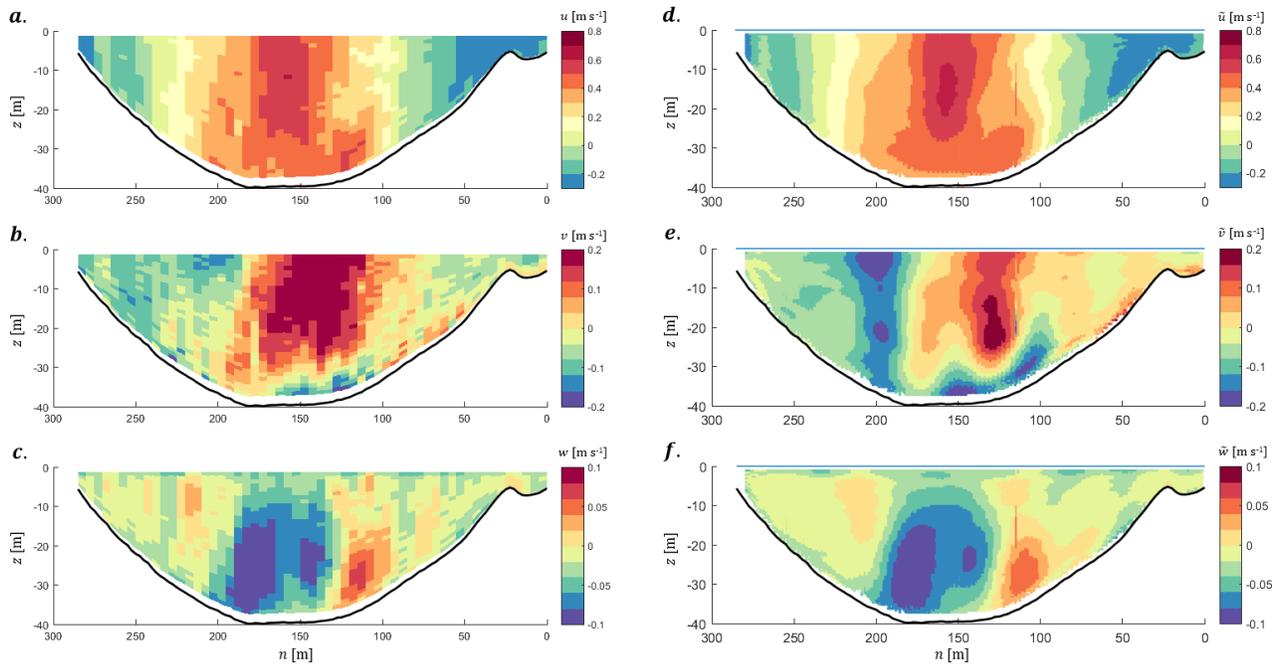Figure 4.25: Decomposed processed velocity data of the third transect, which shows the processed (a) longitudinal, (b) lateral and (c) vertical velocity on the left. The modelled velocity in transect 3 for truncating at $M = N = 16$ is shown on the right, where (d) shows the longitudinal, (e) the lateral and (f) the vertical velocity.

5 Discussion

Fourier analysis was already applied by Quimpo (1967) in order to identify the presence of significant harmonic components in river flow data. However, such as many other researches in the field of investigating and simulating river flow later on, the Fourier series were applied only with respect to time (Tesfaye, 2005; Saremi et al., 2011). New insights are obtained regarding dominant spatial scales in cross-sectional velocity data by applying a Fourier transform, that investigates the spatial velocity distribution. However, these insights are linked to the created model and the applied set of base functions for the velocity components.

Applying Fourier transform in this study has been chosen because the flow velocity field can be approximated by including higher order function progressively that is fitted to all the measured cross-sectional velocity data. However, velocity profiles over water depth are often more log-shaped, which are approximated by superpositioning sinusoids to refine the sine curve. The strength of the approach is having the possibility to include physical constraints within the base functions.

Only the truncation numbers and corresponding computed amplitudes for m and n are changing in this study, considering Eq.(5.1) that represent the velocity fit of the three components. In other words, the performance of the model fit, with respect to its accuracy, precision and usefulness is related to the set of modes that are taken into account and where it is truncated. However, some other parts in the functions might change as well. For example, the part that is influencing the considered number of waveforms over a certain direction, *i.e.,* the parts between brackets of the sinusoids. Currently, the no-slip condition is assumed at the river bed, which generates a zero velocity at the solid boundary for the three velocity components. At the water surface there is velocity present for the longitudinal and lateral velocity component, but not for the vertical velocity component.

$$\begin{pmatrix} \tilde{u}(\zeta,\sigma) \\ \tilde{v}(\zeta,\sigma) \\ \tilde{w}(\zeta,\sigma) \end{pmatrix} = \sum_{m=1}^{M} \sum_{n=1}^{N} \begin{pmatrix} a_u^{m,n} \\ a_v^{m,n} \\ a_w^{m,n} \end{pmatrix} \begin{pmatrix} sin\left(m\pi\zeta\right) \\ sin\left(m\pi\zeta\right) \\ sin\left(m\pi\zeta\right) \end{pmatrix} \begin{pmatrix} sin\left(\left(n-\frac{1}{2}\right)\pi\sigma\right) \\ sin\left(\left(n-\frac{1}{2}\right)\pi\sigma\right) \\ sin\left(n\pi\sigma\right) \end{pmatrix} \qquad (5.1)$$
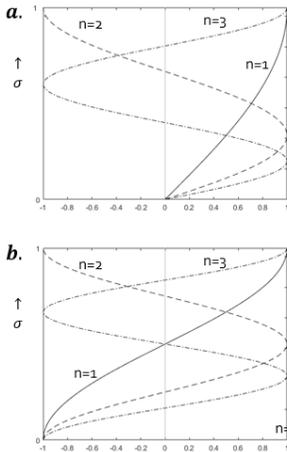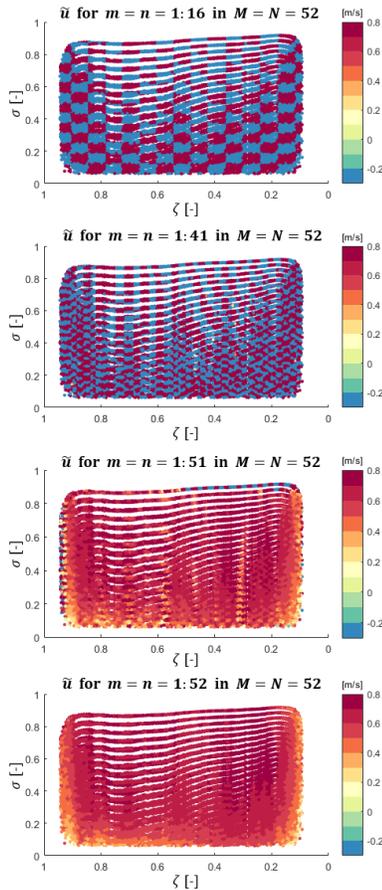
Figure 5.1: Sine- and cosine-based.

Figure 5.2: Normalised domain with model fit with parts of truncating at $M = N = 52$. These parts are ranging from 1 to 16, 41, 51 and 52 respectively.

Changing the part that influences the consideration of number of waveforms affect the maximum truncation numbers that can be analysed. Because the model cannot or hardly approximate measured velocities for an inadequate model set up due to insufficient information and link with the measured data. Current model set up can include highest mode numbers to model velocity.

In addition, the function is sine-based currently, which can also be changed to a cosine-based function. Here, the index is incrementing by one as well. However, it starts at zero ($n = 0$), which includes an average that can been taking into account for mode 0. This shows better results for truncating at lower mode numbers in general, because it is based on average measured velocity. However, besides the fact that this cosine-based function is not based on physical boundary conditions (see Figure 5.1), it results in the same potential of model performance by including higher modes due to phase shift.

Dominant spatial scales are observed with the current functions that are present in the model. One dominant region is visible for truncating the velocity fit at $M = N = 52$, which is located around $m = 14$ as mentioned. However, there is no adequate result obtained by trying to fit the velocity with only the amplitudes for the modes located in the dominant region (see Figure 5.2). The amplitudes computed outside the dominant region are lower than the ones inside the dominating zone. However, amplitudes outside domination are still adding a significant part to the accuracy of the velocity fit. The computed velocity over the normalised domain extends up to the order of magnitude of $1.5 \cdot 10^8$ m s^{-1} by considering the part 1:16 for both m and n, which is disturbed. However, the colour scale is limited to -0.3 and 0.8 m s^{-1} in Figure 5.2 for comparison reasons. The alternation between low and high values of amplitudes is clearly visible. Including almost all modes ($m = n = 1 : 51$) before the truncation of the velocity fit ($M = N = 52$), still differs significantly from the fit where all modes are included. The amplitudes are computed for a specific truncation numbers and are not interchangeably because the coefficients are not final. The amplitudes are computed with a least-square fit, where each specific combination of modes adds a significant part to the velocity fit.

An adequate representation of the flow field is obtained after truncating at $M = N = 16$, which corresponds to eight waveforms over river width. These waveforms have a length of approximately 30 meters in transect 1 and 2. So, half a wave length corresponds to the average water depth in these transects. This might be physically driven by the spatial scale of secondary circulation cells in a wide channel, as described in §2.1. However, it might also be coincidence since it is only based on two transects and holds for straight wide river flows. These transects have approximately a constant water depth over width. However, the water depth fluctuates more in other transects due to scouring by the spiral flow. Data collected in a river bend is not ideal to investigate this behaviour.

The method results in continuous functions that describe the velocity field in detail, which meets the requirement of no-slip at the boundaries. Furthermore, outcome of the study can be used for river management to compute the river discharge over the entire cross-section roughly but substantiated. It must be noted that the upper bound of the summation should not exceed $M = N = 8$, see §4.4.5 about inter- and extrapolation of the velocity fit. In addition, this study might provide input for executing ADCP measurements, in particular on the desired sample distance horizontally. For post-processing of the velocity data later on, the study allows to identify dominant spatial patterns that can help in providing properly smoothing windows in the spatial domain.

6 Conclusion

In this section conclusions are drawn based on the preliminary stated research questions. The goal of this research was to identify the dominant spatial patterns in a river cross-section, so that averaging and smoothing of data can be executed more certain in future. The main research question reads:

> *To what extent must a data-driven model (based on spectral analysis) increase in its complexity, to adequately represent the three-dimensional spatial velocity distribution of a river's cross-section?*

In order to provide an answer to this question the typical flow patterns in the available data sets are investigated. Next, a (data-driven) model is been set up to analyse the dominant spatial scales. Finally, the spatial scales and the influence of truncation numbers are evaluated. These steps correspond with the three subquestions in §1.3.

1. *Which typical flow patterns can be observed by analysing the available data set having regard to the main flow pattern, secondary flow and spatial scales?*

    The flow is weak near the boundaries and strong in the center or slightly outwards of the center in the cross-sections. However, near the scour hole the flow recirculates at the outer sides causing an upstream flow locally. Water flows to the outer bend at the surface and to the inner bend near the river bed. Strong vertical velocity is observed near the scour hole.

2. *How can the dominant spatial scales in a river's cross-section be analysed and identified?*

    The dominant spatial scales can be observed from the computed amplitudes in the model fit. An amplitude belongs to a certain combination of mode numbers that results from the least-squares solution, which fits the trigonometric term to the velocities measured in a cross-section.

3. *What are the spatial scales in an adequate representation of the cross-sectional flow field?*

The spatial scales change with considered truncation numbers in the model fit and are different for each transect. However, since an adequate representation is obtained after truncating at $M = N = 16$, the truncation numbers can be fixed. Nevertheless, relatively large differences are still observed in the dominant spatial scales among different transects. In general, there can be concluded that:

- The first mode ($m = 1$ and $n = 1$) add most significant value for the longitudinal velocity fit, which corresponds with waveforms with a length of 2 times the local river width and 4 times the local water depth.
- Dominant scales are mainly located in the region between $m = 1$ and $m = 13$ and between $n = 1$ and $n = 8$ for the lateral velocity fit. This corresponds with waveforms with a length of 0.15-2 times the local river width and 0.27-4 times the local water depth.
- For the vertical velocity fit are the dominant scales mainly located in the region between $m = 1$ and $m = 11$ and between $n = 1$ and $n = 11$ for transects with relatively strong vertical flow. This corresponds with waveforms with a length of 0.18-2 times the local river width and 0.19-4 times the local water depth. For transects without strong vertical flow, the dominant scale is an extension over m and/or n of the first mode, *i.e.,* $m = 1$ to $m = 10$ for $n = 1$ or vice versa.

The three-dimensional velocity distribution can be represented adequately as residuals have no spatial structure. Most of the spatial structure in residuals is disappeared after truncating at $M = N = 16$, which corresponds to eight waveforms over local river width horizontally and 7.75 waveforms the local water depth vertically.

7 Recommendations

In order to guide future research on investigating dominant spatial scales in cross-sectional velocity data, five recommendations are given. It concerns to investigate with another data set, possible modifications in the model and on the analysis afterwards.

The available data set consists of collected velocity data in a sharp river bend, which is challenging to model the three-dimensional velocity. However, it might provide general insights to validate the model with a (simple) straight river flow or flume that is measured with different instruments.

Current approach is based on Fourier transform, where the coefficients are computed with a linear regression analysis (*i.e.,* `lscov`-function in MATLAB). In theory, the model should be able to fit the velocity distribution without residuals left. However, the model complexity is limited at a certain point. The possibility of using other methods to compute the Fourier coefficients, in not evenly distributed data sets, should be investigated.

The study can be extended by including a weighted analysis. So that the big density of velocity data at the edges of the measurement is considered with a lesser degree, since the variability in residuals is large at the sides. The model can be disturbed by the fact that velocities with relatively large difference can be measured nearby each other or even at the very same location, which is in theory possible by measuring the same transect in multiple cycles.

Furthermore, future research can be done to extend the analysis by including different base functions where more physical constraints are included as well, such as continuity.

All combinations in the set of modes are considered in the velocity fit. However, during the investigation afterwards, only equal truncation numbers are considered, which is a research limitation. The case of $M = N = 16$ provide adequate results, but another, not necessarily equal, truncation for M and N might provide an adequate result as well. Further research on a sort of performance indicator is recommended.

Bibliography

Adler M. and Nicodemus U. A new computer model for the evaluation of data from acoustic Doppler current profilers (ADCP). *Physics and Chemistry of the Earth, Part C: Solar, Terrestrial & Planetary Science*, 26(10-12):711–715, 2001.

Allen G. and Chambers J. Sedimentation in the Modern and Miocene Mahakam Delta. *Indonesian Petroleum Association*, page p. 236, 1998.

Allen J. Fundamental properties of fluids and their relation to sediment transport processes. *Sediment transport and depositional processes*, pages 25–60, 1994.

Casey H. J. Uber Geschiebebewegung. Technical report, Berlin, 1935.

Côté J., Hotchkiss F., Martini M., Denham C., and Ramsey A. Acoustic Doppler Current Profiler Data Processing System Manual. Technical report, 2011.

Davis J. C. *Statistics and data analysis in geology*. John Wiley & Sons, Inc., 3rd ed. edition, 2002. ISBN 978-0-471-17275-8.

Dinehart R. and Burau J. Repeated surveys by acoustic Doppler current profiler for flow and sediment dynamics in a tidal river. *Journal of Hydrology*, 314(1-4):1–21, nov 2005.

Frost J. Regression Analysis: How Do I Interpret R-squared and Assess the Goodness-of-Fit?, 2013. URL `http://blog.minitab.com/blog/adventures-in-statistics-2/regression-analysis-how-do-i-interpret-r-squared-and-assess-the-goodness-of-fit`.

Frost J. Regression Analysis: How to Interpret S, the Standard Error of the Regression, 2014. URL `http://blog.minitab.com/blog/adventures-in-statistics-2/regression-analysis-how-to-interpret-s-the-standard-error-of-the-regression`.

Heil M. Chapter 4 - Boundary and initial conditions. In *Syllabus - Viscous Fluid Flow*. University of Manchester, 2017. URL `http://www.maths.manchester.ac.uk/{~}mheil/Lectures/Fluids/`.

Hidayat H., Vermeulen B., Sassi M. G., Torfs P. J. J. F., and Hoitink A. J. F. Discharge estimation in a backwater affected meandering river. *Hydrology and Earth System Sciences*, 15(8):2717–2728, aug 2011. ISSN 1607-7938.

Janjic Z., Gall R., and Pyle M. Scientific Documentation for the NMM Solver. *NATIONAL CENTER FOR ATMOSPHERIC RESEARCH*, NCAR/TN-4(NCAR TECHNICAL NOTE):54, 2010.

Johnson R. and Wichern D. *Applied Multivariate Statistical Analysis*. Pearson Prentice Hall, Upper Saddle River, New Jersey, sixth edit edition, 2007.

Kim D., Muste M., Weber L., and Asman R. Getting to Know AdcpXP (Adcp eXtended Processing) - a tutorial for using AdcpXP. *IIHR*, (Hydroscience & Engineering, University of Iowa):87, 2007.

Kim D., Mueller D., Winkler M., and Muste M. ADCP Velocity Mapping Software: A quick tutorial for using VMS. page 104, 2009.

Kotsovinos N. E. Secondary currents in straight channels. *Appl. Math. Modelling*, 12:1085–1088, 1988.

Le Bot P., Kermabon C., Lherminier P., and Gaillard F. Chaine Automatisée de Suivi des Courantomètres Acoustiques Doppler Embarqués. CASCADE V6.1 : Logiciel de validation et de visualisation des mesures ADCP de coque. *Rapport OPS/LPO 11-01*, page 93, 2011.

Marsden R. F. and Ingram R. G. Correcting for beam spread in acoustic Doppler current profiler measurements. *Journal of Atmospheric and Oceanic Technology*, 21(9):1491–1498, 2004.

Marshall J., Emanuel K., and Adcroft A. Atmospheric and Oceanic Modeling. *Massachusetts Institute of Technology*, 12.950(MIT OpenCourseWare, License: Creative Commons BY-NC-SA), 2004. URL `https://ocw.mit.edu`.

Morrison A. Common Boundary Conditions in Fluid Mechanics. (4):1998, 1998.

Mueller D. and Wagner C. Measuring discharge with acoustic Doppler current profilers from a moving boat. *U. S. Geological Survey Techniques and Methods*, (December):72, 2009.

Muste M., Yu K., and Spasojevic M. Practical aspects of ADCP data use for quantification of mean river flow characteristics; Part I: moving-vessel measurements. *Flow Measurement and Instrumentation*, 15(1):1–16, mar 2004.

Nezu I., Nakagawa H., and Tominaga A. Secondary Currents in a Straight Channel Flow and the Relation to Its Aspect Ratio. *Turbulent shear flow*, 4, 1985.

Nortier I. and de Koning P. *Toegepaste vloeistofmechanica - hydraulica voor waterbouwkundigen*. Wolters-Noordhoff, zevende dr edition, 1996.

Padmalal D. and Maya K. Rivers-Structure and Functions. In *Sand Mining - Environmental Impacts and Selected Case Studies*, chapter 2, page 162. Springer, 2014.

Parsons D. R., Jackson P. R., Czuba J. A., Engel F. L., Rhoads B. L., Oberg K. A., Best J. L., Mueller D. S., Johnson K. K., and Riley J. D. Velocity Mapping Toolbox (VMT): a processing and visualization suite for moving-vessel ADCP measurements. *Earth Surface Processes and Landforms*, 38(11): 1244–1260, sep 2013.

Pham Van C., Gourgue O., Sassi M., Hoitink A., Deleersnijder E., and Soares-Frazão S. Modelling fine-grained sediment transport in the Mahakam landâĂŞsea continuum, Indonesia. *Journal of Hydro-environment Research*, 13:103–120, dec 2016. ISSN 15706443.

Powell D. Patterns and processes of sediment sorting in gravel-bed rivers. *Progress in Physical Geography: Earth and Environment*, 22(1), 1998.

Quimpo R. G. Stochastic Model of Daily River Flow Sequences. *Hydrology Papers 18*, (18):37, 1967.

R.D.Instruments . Acoustic doppler current profilers principles of operation: A practical primer. *RD Instruments*, 00(January), 1996.

Ribberink J. and Hulscher S. RIVER DYNAMICS - I : Shallow- Water Flows. Technical Report March, 2012.

Robert A. *River Processes - An Introduction to Fluvial Dynamics*. Hodder Education, London, 2003.

Rowe F. and Young J. An Ocean Current Profiler Using Doppler Sonar. In *OCEANS '79*, pages 292–297. IEEE, 1979.

Saremi A., Pashaki, Karimi M. H., Sedghi H., Rouzbahani A., and Saremi A. Simulation of River Flow Using Fourier Series Models. *ICBEE*, 19:133–138, 2011.

Simpson M. Discharge measurements using a broad-band acoustic Doppler current profiler. *U.S. Geological Survey Open-File Report*, 01-01:134, 2001.

Simpson M. and Oltmann R. Discharge-measurement system using an acoustic Doppler current profiler with applications to large rivers and estuaries, 1993.

Snowbird U. Introduction to velocity mapping (USGS HydroAcoustics, short-course), 2012.

Steidl J. and Dorfmann C. ADCPtool - A postprocessing framework for ADCP measurements - Reference Manual. Technical report, 2013.

Tesfaye Y. G. *Seasonal Time Series Models and Their Application to the Modeling of River Flows*. PhD thesis, 2005.

Vermeulen B. *Rivers runnig deep - Complex flow and morphology in the Mahakam River, Indonesia*. PhD thesis, 2014.

Vermeulen B., Hoitink A. J. F., van Berkum S. W., and Hidayat H. Sharp bends associated with deep scours in a tropical river: The river Mahakam (East Kalimantan, Indonesia). *Journal of Geophysical Research: Earth Surface*, 119(7):1441–1454, jul 2014a.

Vermeulen B., Sassi M., and Hoitink A. Improved flow velocity estimates frommoving-boat ADCP measurements. *Water Resources Research*, 50:4186–4196, 2014b.

Vermeulen B., Hoitink A. J. F., and Labeur R. J. Flow structure caused by a local cross-sectional area increase and curvature in a sharp river bend. *Journal of Geophysical Research: Earth Surface*, 120(9):1771–1783, sep 2015.

Yalin M. *River Mechanics*. Pergamon Press Ltd., Oxford, 1st ed. edition, 1992. ISBN 0080401902.

Appendices

A Properties of velocity data

A.1 Processed velocity data

Properties of velocity data processed according the method of Vermeulen et al. (2014b).

T	b_m [m]	h_{avg} [m]	u_{max} [m s^{-1}]	u_{min} [m s^{-1}]	Condition of strongest 25%	$u_{25\%}$ in A [%]	$u_{25\%}$ over b_m [%]	$u_{25\%}$ over h_{avg} [%]
1	240	11.12	0.75	0.34	$u \geq 0.65$	61	81	75
2	215	12.80	0.82	-0.09	$u \geq 0.60$	66	77	90
3	280	26.49	0.60	-0.28	$u \geq 0.40$	28	19	95
4	285	22.25	0.58	-0.15	$u \geq 0.40$	33	26	90
5	280	16.40	0.54	0.04	$u \geq 0.40$	63	65	95
6	210	16.56	0.60	0.40	$u \geq 0.55$	23	28	50
7	180	16.06	0.68	0.35	$u \geq 0.60$	47	43	85

Table 1: Properties of longitudinal flow from processed data. The condition of strongest 25% of the longitudinal flow is computed by $u_{25\%} = 1/4\, u_{min} + 3/4\, u_{max}$, which is used to quantify the area (A), part of river width (b_m) and average water depth (h_{avg}) that carries the strongest flow.

T	v_{max} [m s^{-1}]	v_{min} [m s^{-1}]	Condition of strongest 25%	$v_{25\%}$ in A [%]	$v_{25\%}$ over b_m [%]	$v_{25\%}$ over h_{avg} [%]
1	0.02	-0.05	$v \geq 0.016$	27	24	30
2	0.03	-0.08	$v \geq 0.027$	30	29	32
3	0.07	-0.12	$v \geq 0.049$	26	36	45
4	0.10	-0.05	$v \geq 0.037$	30	42	33
5	0.04	-0.10	$v \geq 0.035$	7	10	27
6	0.03	-0.06	$v \geq 0.023$	31	35	29
7	0.02	-0.07	$v \geq 0.023$	22	25	22

Table 2: Properties of lateral flow from processed data. The condition of strongest 25% of the lateral flow is computed by $v_{25\%} = (1/2\, |v_{min}| + 1/2\, v_{max})/2$.

Table 3: Properties of vertical flow from processed data. The condition of strongest 25% of the vertical flow is computed by $w_{25\%} = (^1/_2\,|w_{min}| + ^1/_2\,w_{max})/2$.

T	v_{max} [m s^{-1}]	v_{min} [m s^{-1}]	Condition of strongest 25%	$w_{25\%}$ in A [%]	$w_{25\%}$ over b_m [%]	$w_{25\%}$ over h_{avg} [%]
1	0.06	-0.10	$w \geq 0.039$	19	25	36
2	0.14	-0.22	$w \geq 0.090$	18	20	31
3	0.28	-0.23	$w \geq 0.126$	21	19	42
4	0.22	-0.38	$w \geq 0.149$	29	29	51
5	0.15	-0.15	$w \geq 0.076$	38	39	57
6	0.12	-0.21	$w \geq 0.083$	17	42	29
7	0.09	-0.16	$w \geq 0.063$	17	34	21

A.2 Raw velocity data

The magnitudes of the raw velocity data are provided in this Appendix per component. First, the number of observations are shown with the number of ensembles measured per transect.

- $n_{obs} = 31{,}613$ over 1,776 ensembles in transect 1
- $n_{obs} = 31{,}753$ over 1,491 ensembles in transect 2
- $n_{obs} = 38{,}459$ over 1,879 ensembles in transect 3
- $n_{obs} = 60{,}738$ over 1,524 ensembles in transect 4
- $n_{obs} = 47{,}432$ over 1,665 ensembles in transect 5
- $n_{obs} = 36{,}575$ over 1,356 ensembles in transect 6
- $n_{obs} = 31{,}141$ over 1,263 ensembles in transect 7

Table 4: Here, T represents the transect number, $\hat{u}_{max}$ and $\hat{u}_{min}$ the maximum and minimum measured longitudinal velocity respectively. $\bar{u}$ is the mean of the velocity and $|\bar{u}|$ the mean of considering absolute velocities only. $\hat{u}_{var}$ is the variance and $\hat{u}_{std}$ the standard deviation of the rotated velocity all in [m s^{-1}].

T	$\hat{u}_{max}$	$\hat{u}_{min}$	$\bar{u}$	$|\bar{u}|$	$\hat{u}_{var}$	$\hat{u}_{std}$
1	0.97	-0.03	0.65	0.65	0.0131	0.1144
2	1.06	-0.25	0.60	0.60	0.0355	0.1884
3	1.78	-0.69	0.23	0.31	0.0835	0.2890
4	1.01	-0.63	0.29	0.32	0.0542	0.2328
5	0.91	-0.26	0.38	0.38	0.0135	0.1160
6	0.91	-0.17	0.51	0.51	0.0094	0.0971
7	0.95	-0.55	0.58	0.58	0.0094	0.0967

T	$\hat{v}_{max}$	$\hat{v}_{min}$	$\bar{v}$	$|\bar{v}|$	$\hat{v}_{var}$	$\hat{v}_{std}$
1	0.50	-0.46	0.01	0.06	0.0062	0.0789
2	0.79	-0.60	0.11	0.15	0.0175	0.1322
3	1.60	-0.90	-0.02	0.14	0.0325	0.1803
4	0.69	-0.67	0.02	0.15	0.0315	0.1775
5	0.36	-0.61	-0.14	0.16	0.0160	0.1263
6	0.61	-0.31	0.10	0.12	0.0092	0.0961
7	0.44	-0.40	0.07	0.09	0.0075	0.0865

Table 5: Same as Table 1, but now for lateral velocity ($\hat{v}$)

T	$\hat{w}_{max}$	$\hat{w}_{min}$	$\bar{w}$	$|\bar{w}|$	$\hat{w}_{var}$	$\hat{w}_{std}$
1	0.14	-0.16	-0.01	0.03	0.0011	0.0326
2	0.23	-0.22	-0.02	0.04	0.0016	0.0401
3	0.45	-0.35	-0.03	0.05	0.0038	0.0620
4	0.27	-0.23	0.02	0.05	0.0037	0.0606
5	0.18	-0.19	-0.005	0.03	0.0017	0.0414
6	0.12	-0.17	-0.02	0.03	0.0012	0.0352
7	0.13	-0.15	-0.01	0.03	0.0012	0.0344

Table 6: Same as Table 1, but now for vertical velocity ($\hat{w}$)

B MATLAB *script of model*

The general MATLAB script is shown in this appendix, where the velocity in *x*- and *y*-direction are already rotated into longitudinal and lateral velocity components respectively.

```matlab
% ** Script that analyses the longitudinal, lateral and vertical velocity component in the first transect (ct=1) **

clearvars, close all

% In this script the following tasks will be executed:
%  -1- Loading data                    ('raw','cumm','...')
%  -2- Set input arguments             (M & N)
%  -3- Find measured velocity locations (f_good)
%  -4- Store measured velocity         (Bu)
%  -5- Compute "fourier fits"          (Au)
%  -6- Determine amplitudes            (lscov)
%  -7- Compute fitted velocity         (u_fit)
%  -8- Determine error                 (err_u)
%  -9- Testing significance            (rsq, ser, eps_z eps_s)


%% Step 1 | loading data
load raw_corr 'raw'
load b_procvel.mat 'msh'

%  matrix of velocity per transect (bins*ensembles)
u = raw(1).snvel(:,:,1);                %measured u for transect 1
v = raw(1).snvel(:,:,2);                %measured v for transect 1
w = raw(1).vel(:,:,3);                  %measured w for transect 1

%create nondimensional domain for model-fit
zeta  = 0:0.005:1;
zeta = repmat(zeta,201,1);
sigma = 1:-0.005:0;
sigma = repmat(sigma,201,1)';


%% Step 2 | set input arguments
%  set truncation number
m = 1:3;                                %truncation number M, relates to river width
n = 1:3;                                %truncation number N, relates to water depth

i = combvec(m,n);                       %create matrix with all possible combination in set of modes


%% Step 3 | find measured velocity locations
%  find actual measured velocity locations and without side lobe interference
f_good_u = isfinite(raw(1).sigVel) & isfinite(raw(1).snvel(:,:,1)) & raw(1).sigVel > 0.06;  %find good within u
f_good_v = isfinite(raw(1).sigVel) & isfinite(raw(1).snvel(:,:,2)) & raw(1).sigVel > 0.06;  %find good within v
f_good_w = isfinite(raw(1).sigVel) & isfinite(raw(1).vel(:,:,3)) & raw(1).sigVel > 0.06;     %find good within w
```

```matlab
%% Step 4 | store measured velocity
Bu = u(f_good_u);    %for every u read further on, it can be substituted by v and w as well.


%% Step 5 | compute "trigonometric terms" (SINE-BASED)
%   Computes the trigonometric terms:
%   Au = sin(m*pi*zeta) * sin((n-1/2)*pi*sigma)
%   Av = sin(m*pi*zeta) * sin((n-1/2)*pi*sigma)
%   Aw = sin(m*pi*zeta) * sin(n*pi*sigma)

Au = zeros(length(Bu),length(m)*length(n)); %preallocation
for j=1:length(i)
if m==1 & n~=1
    Au(:,j) = sin(m*pi*raw(1).zetVel(f_good)) .* sin((i(2,j)*pi-(pi/2))*raw(1).sigVel(f_good));
elseif n==1 & m~=1
    Au(:,j) = sin(i(1,j)*pi*raw(1).zetVel(f_good)) .* sin((n*pi-(pi/2))*raw(1).sigVel(f_good));
elseif m==1 & n==1
    Au = sin(m*pi*raw(1).zetVel(f_good)) .* sin((n*pi-(pi/2))*raw(1).sigVel(f_good));
else
    Au(:,j) = sin(i(1,j)*pi*raw(1).zetVel(f_good_u)) .* sin((i(2,j)*pi-(pi/2))*raw(1).sigVel(f_good_u));
end
end


%   computes the trigonometric terms for whole domain (extrapolate):
for j=1:length(i)
    if  m==1 & n==1
        A_extrap(:,j) = sin(m*pi*zeta(:)) .* sin((n*pi-(pi/2))*sigma(:));
    else
        A_extrap(:,j) = sin(i(1,j)*pi*zeta(:)) .* sin((i(2,j)*pi-(pi/2))*sigma(:));
    end
end


%% Step 6 | compute amplitudes
%   Computes the amplitude (Fourier coefficient) with least-squares in regression (lscov)
a=zeros(length(i));                     %preallocation
for j=1:length(i)
    h = i(1,:)<=i(1,j) & i(2,:)<=i(2,j);    %take only the terms within the truncation, filter lower terms out
    a(h,j) = lscov(Au(:,h),Bu);             %compute amplitude with lscov function
                                            %    (least squares solution between Au(:,h) and Bu
end

if  m==1 & n==1
    a_mat=(a(1,1));                      %create a matrix for the first mode of M & N
else
    a_mat=reshape(a(:,end),m(end),n(end))';%create a matrix for the amplitudes of all combinations in set of modes
end



%% Step 7 | compute fitted velocity
%   The fitted velocity will be computed by: u_fit = a_u * sin(m*pi*zeta) * sin((n-1/2)*pi*sigma)
%                                            v_fit = a_v * sin(m*pi*zeta) * sin((n-1/2)*pi*sigma)
%                                            w_fit = a_w * sin(m*pi*zeta) * sin(n*pi*sigma)
u_fit = Au * a;                         %calculate the model-fit of velocity
u_extrap  = A_extrap * a(:,end);        %calculate the model-fit of velocity (extrapolated)


%% Step 8 | determine error
%   Calculates the residuals for all combinations of modes before M and N
err_u = (Bu - u_fit);
err_end = (Bu - u_fit(:,end));

% Find local mean in errors using for loop and linspace function
% z = zeta, s = sigma, 20 steps.
rz=linspace(min(raw(1).zetVel(:)),max(raw(1).zetVel(:)),21);  %range in zeta; linspace divides the range between
                                                        %     min. and max. zeta into 20 equal steps
rs=linspace(min(raw(1).sigVel(:)),max(raw(1).sigVel(:)),21);  %range in sigma; linspace divides the range between
                                                        %     min. and max. sigma into 20 equal steps


ind=zeros(length(Bu));                  %preallocation
ind=find(f_good_u==1);                  %index that create a logical matrix for f_good
err_mat=zeros(size(u));                 %matrix with zeros with the size of u [80x1776] -> for loop
for num=1:numel(err_end)
    err_mat(ind(num))=err_end(num);     %matrix with the actual and good fitted velocities
end
err_mat(err_mat==0) = NaN;              %replace zeros with NaN's
```

```matlab
for qz=1:length(rz)-1
    ind_mz{qz} = find(raw(1).zetVel >= rz(qz) & raw(1).zetVel < rz(qz+1));  %the brackets {} create a cell array for
indices located in argument
    err_mz(qz) = nanmean(err_mat(ind_mz{qz}));                               %local mean of error
    zet_mz(qz) = nanmean(raw(1).zetVel(ind_mz{qz}));                         %average of location in zeta
end


for qs=1:length(rs)-1
    ind_ms{qs} = find(raw(1).sigVel >= rs(qs) & raw(1).sigVel < rs(qs+1));  %the brackets {} create a cell array for
indices located in argument
    err_ms(qs) = nanmean(err_mat(ind_ms{qs}));                               %local mean of error
    sig_ms(qs) = nanmean(raw(1).sigVel(ind_ms{qs}));                         %average of location in sigma
end


% Find local mean in errors for set of modes using for loop and linspace function
err_allcell = num2cell(err_u,1);                        %all the errors for set of modes into cell array
err_allmat  = zeros(size(u));                           %matrix with zeros with the size of u [80x1776] -> for loop
err_allmat  = num2cell(err_allmat,[1 2]);               %matrix into cell array
err_allmat  = repmat(err_allmat,1,length(i));           %cell array for all modes

for j=1:length(i)
    for num=1:numel(err_u(:,end))
        err_allmat{1,j}(ind(num))=err_allcell{1,j}(num);%matrix for set of modes of actual & good fitted velocities
        err_allmat{1,j}(err_allmat{1,j}==0) = NaN;      %replace zeros with NaN's
    end
end


err_allmz   = zeros(size(err_mz));                      %matrix with zeros with the size of err_mz [1x20]
err_allmz   = num2cell(err_allmz,[1 2]);                %matrix into cell array
err_allmz   = repmat(err_allmz,1,length(i));            %cell array for all modes

for j=1:length(i)
    for qz=1:length(rz)-1
        ind_mz{qz} = find(raw(1).zetVel >= rz(qz) & raw(1).zetVel < rz(qz+1));  %the brackets {} create a cell array
                                                                                  for indices located in argument
        err_allmz{1,j}(qz) = nanmean(err_allmat{1,j}(ind_mz{qz}));               %local mean of error for set of
                                                                                  modes
        zet_mz(qz) = nanmean(raw(1).zetVel(ind_mz{qz}));                         %average of location in zeta
    end
end

err_allms   = zeros(size(err_mz));                      %matrix with zeros with the size of err_ms [1x20]
err_allms   = num2cell(err_allms,[1 2]);                %matrix into cell array
err_allms   = repmat(err_allms,1,length(i));            %cell array for all modes

for j=1:length(i)
    for qs=1:length(rs)-1
        ind_ms{qs} = find(raw(1).sigVel >= rs(qs) & raw(1).sigVel < rs(qs+1));  %the brackets {} create a cell array
                                                                                  for indices located in argument
        err_allms{1,j}(qs) = nanmean(err_allmat{1,j}(ind_ms{qs}));               %local mean of error for set of
                                                                                  modes
        zet_ms(qs) = nanmean(raw(1).sigVel(ind_mz{qs}));                         %average of location in zeta
    end
end


%% Step 9 | testing significance
%  R-squared
rsq = 1-sum(err_u.^2)./sum((Bu-mean(Bu)).^2);          %r-squared for last mode

rsq_mat=zeros(1,length(m)*length(n));                  %preallocation
for j=1:length(i)
    rsq_mat(j) = 1-sum(err_u(:,j).^2)./sum((Bu-mean(Bu)).^2); %computes rsq for each set of mode
end

if  m==1 & n==1
    rsq_mat=(rsq(1,1));
else
    rsq_mat=reshape(rsq_mat,m(end),n(end))';           %create matrix with values for rsq with size of M and N
end
```

```matlab
%  Standard error of regression
ser = sqrt(sum((Bu - u_fit(:,end).^2)/numel(Bu));          %ser for last mode

ser_mat=zeros(1,length(m)*length(n));                      %preallocation
for j=1:length(i)
    ser_mat(j) = sqrt(sum((Bu - u_fit(:,j).^2)/numel(Bu)); %computes ser for each set of mode
end


if  m==1 & n==1
    ser_mat=(ser(1,1));
else
    ser_mat=reshape(ser_mat,m(end),n(end))';               %create matrix with values for ser with size of M and N
end



% root mean squared error over zeta and sigma
rmse_z=sqrt(nansum((err_mz).^2)/length(err_mz));           %over zeta
rmse_s=sqrt(nansum((err_ms).^2)/length(err_ms));           %over sigma


for j=1:length(i)
    rmse_allz{1,j} = sqrt(nansum((err_allmz{1,j}).^2)/length(err_mz);
end

epz_mat=cell2mat(eps_allz);
if  m==1 & n==1
    epz_mat = (eps_z(1,1));
else
    epz_mat = reshape(epz_mat,m(end),n(end))';
end



for j=1:length(i)
    rmse_alls{1,j} = sqrt(nansum((err_allms{1,j}).^2)/length(err_ms);
end

eps_mat=cell2mat(eps_alls);
if  m==1 & n==1
    eps_mat = (eps_s(1,1));
else
    eps_mat = reshape(eps_mat,m(end),n(end))';
end
```

C Dominant spatial scales

C.1 Domination in transect 1

The values for amplitudes are shown for truncating at:

- $M = N = 1$
- $M = N = 3$
- $M = N = 5$
- $M = N = 9$
- $M = N = 13$
- $M = N = 18$
- $M = N = 21$
- $M = N = 36$
- $M = N = 52$

The dominant scatial scales for truncation numbers 1, 4, 8, 10, 12 and 16 are provided in Appendix D.
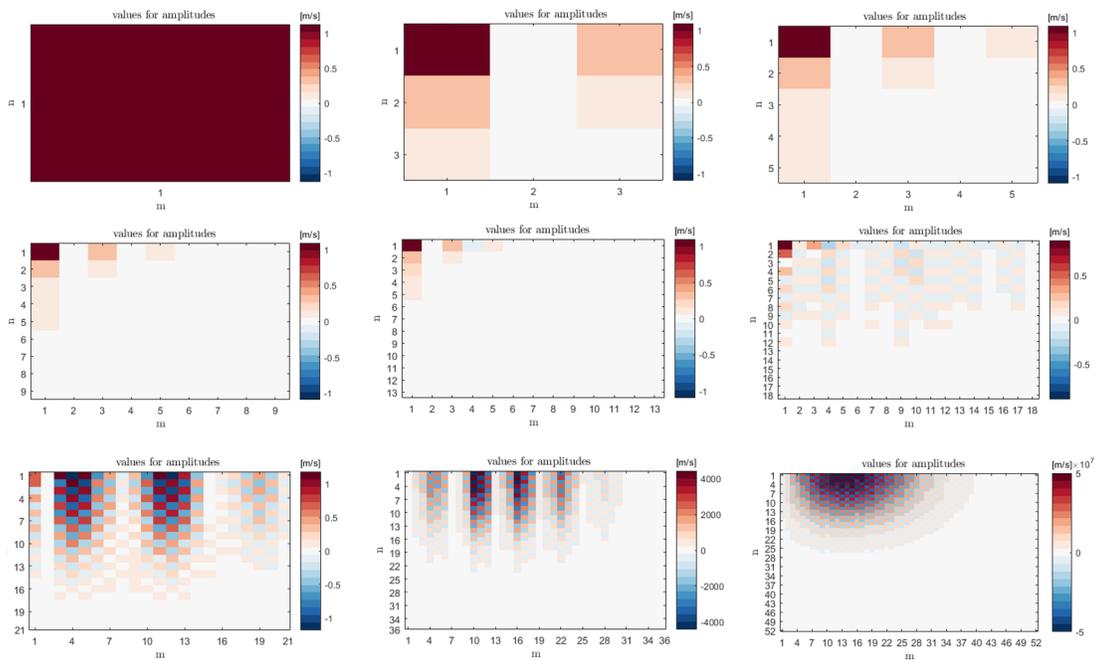
Figure 1: Values of amplitudes in $\tilde{u}$.
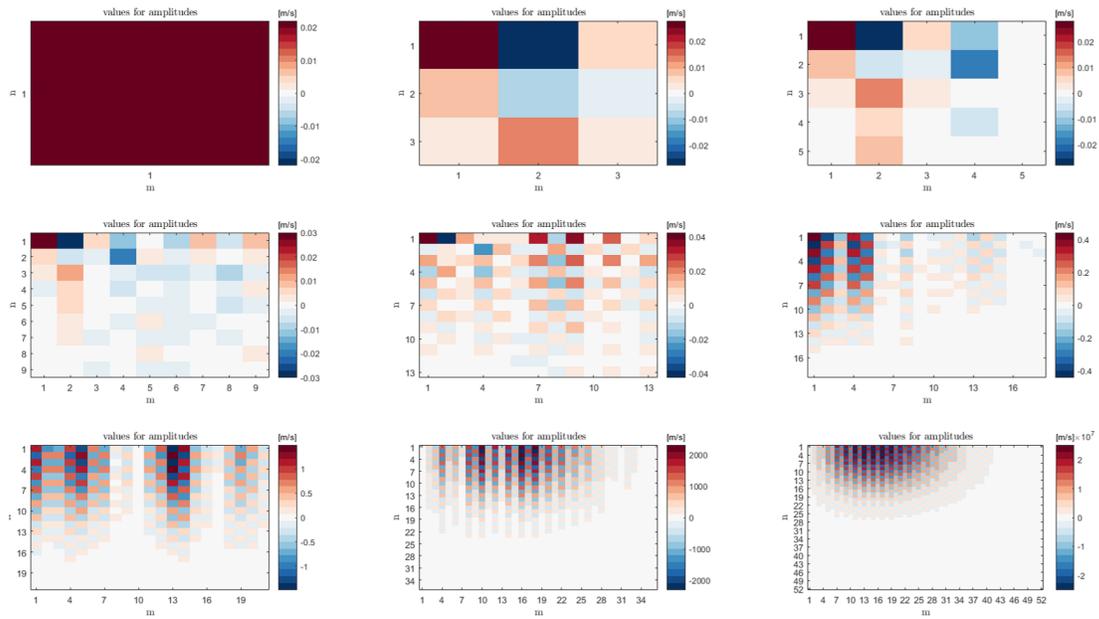
Figure 2: Values of amplitudes in $\tilde{v}$.
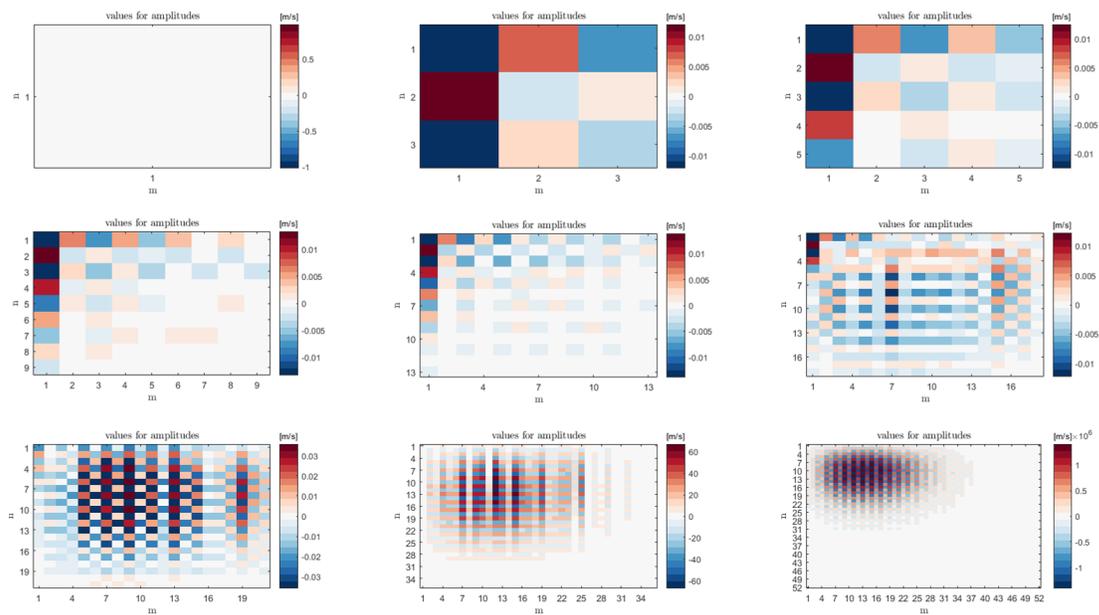
Figure 3: Values of amplitudes in $\tilde{w}$.

C.2 Domination in transects

For all transects the aabsolute values of amplitudes are plotted for truncating at $M = N = 3, 2136$ and 52.
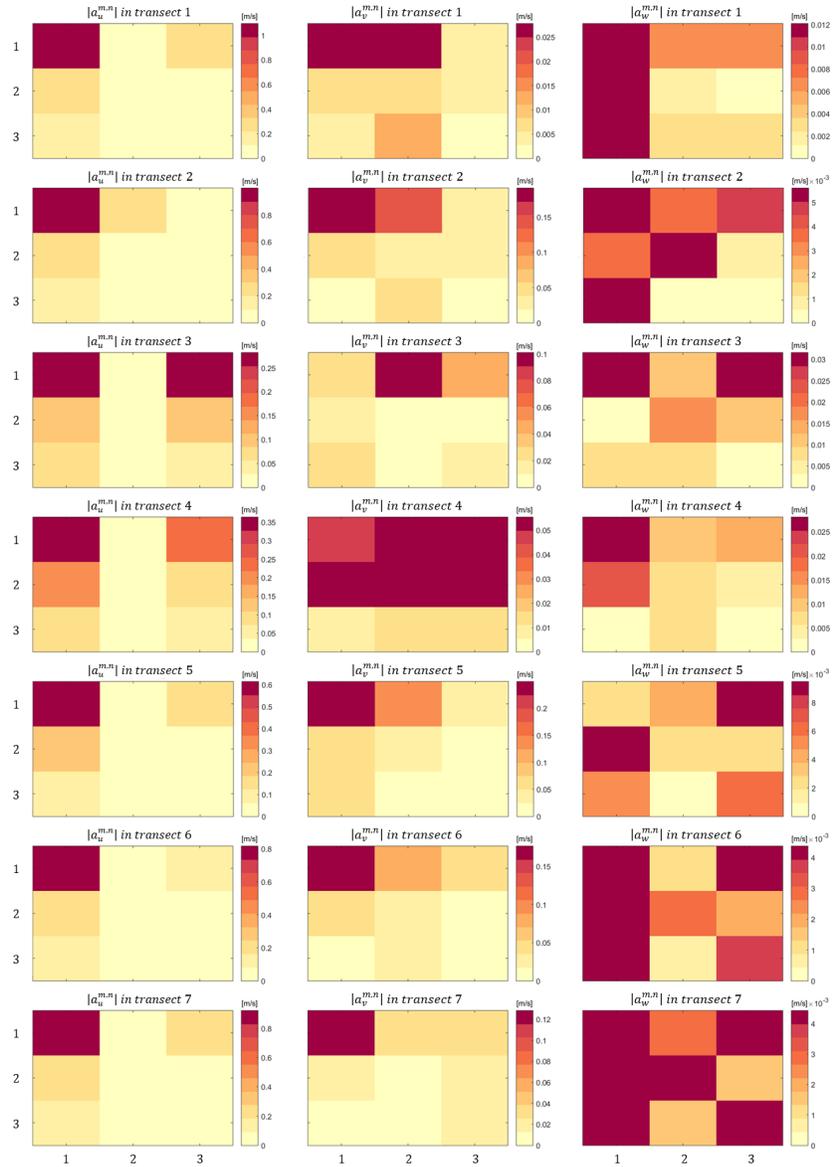
Figure 4: Absolute values of amplitudes in the seven transects for truncating at $M = N = 3$.
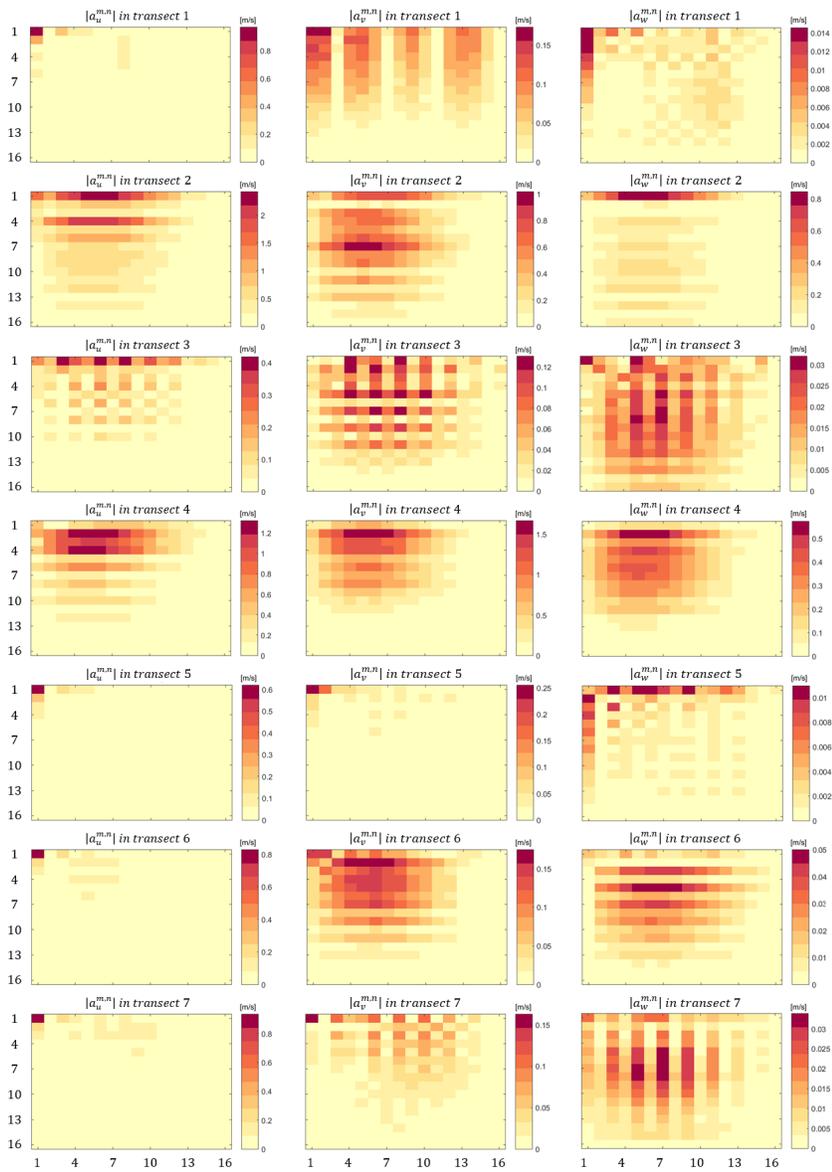
Figure 5: Absolute values of amplitudes in the seven transects for truncating at $M = N = 16$.
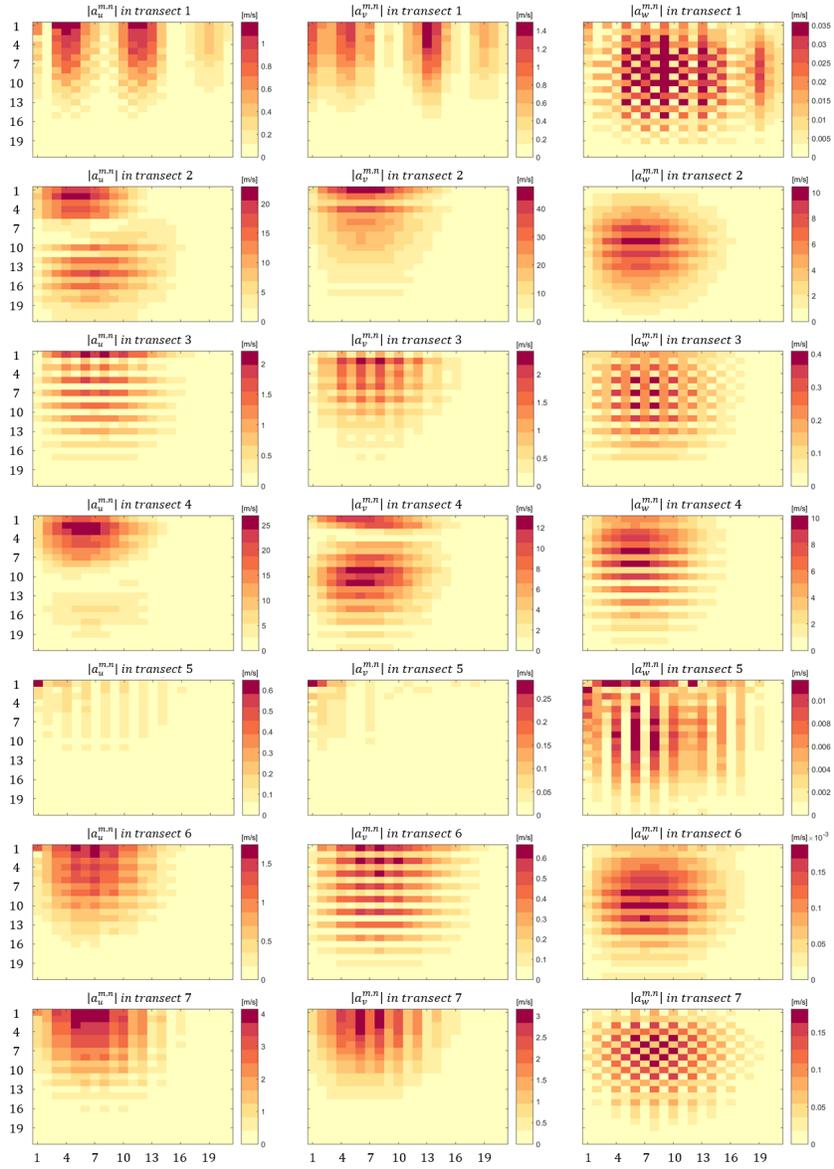
Figure 6: Absolute values of amplitudes in the seven transects for truncating at $M = N = 21$.
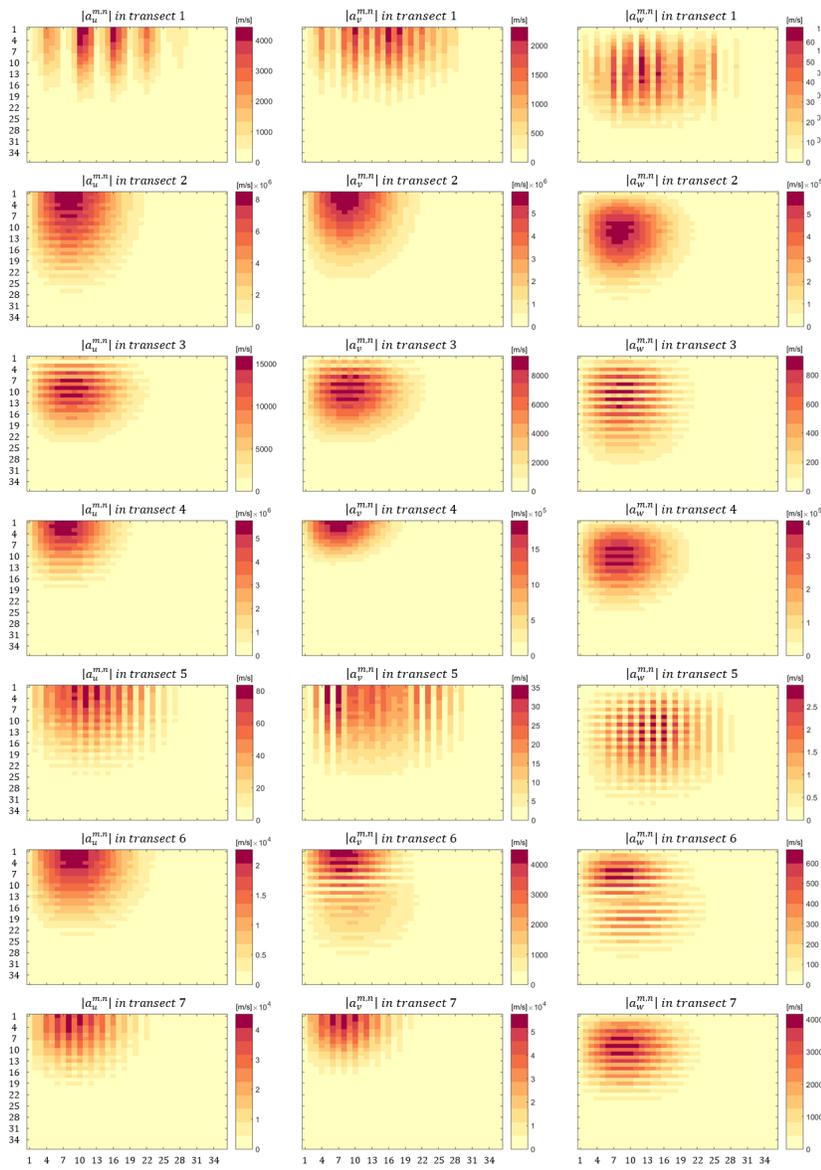
Figure 7: Absolute values of amplitudes in the seven transects for truncating at $M = N = 36$.
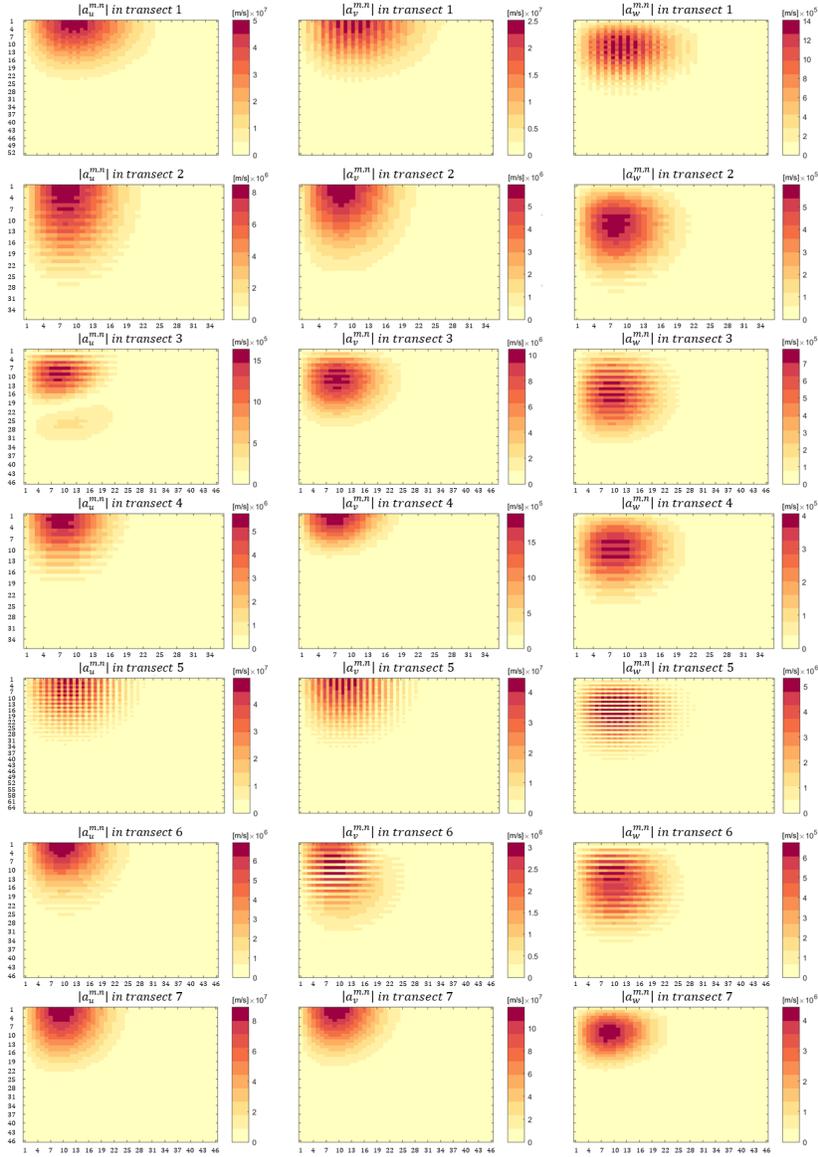
Figure 8: Absolute values of amplitudes in the seven transects for truncating at relatively high modes. However, these are changing for each transect.

D Analysis of model fit

The plots where the analysis is based on are shown for the longitudinal, lateral and veritcal veloctiy fits for the first transect in this appendix for truncations at:

- $M = 1$ and $N = 1$, which corresponds to 0.5 and 0.25 number of waveforms over width and depth respectively.

- $M = 4$ and $N = 4$, which corresponds to 2 and 1.75 number of waveforms over width and depth respectively.

- $M = 8$ and $N = 8$, which corresponds to 4 and 3.75 number of waveforms over width and depth respectively.

- $M = 10$ and $N = 10$, which corresponds to 5 and 4.75 number of waveforms over width and depth respectively.

- $M = 12$ and $N = 12$, which corresponds to 6 and 5.75 number of waveforms over width and depth respectively.

- $M = 16$ and $N = 16$, which corresponds to 8 and 7.75 number of waveforms over width and depth respectively.

- $M = 21$ and $N = 21$, which corresponds to 10.5 and 10.25 number of waveforms over width and depth respectively.

- $M = 36$ and $N = 36$, which corresponds to 18 and 17.75 number of waveforms over width and depth respectively.

- $M = 52$ and $N = 52$, which corresponds to 26 and 25.75 number of waveforms over width and depth respectively.
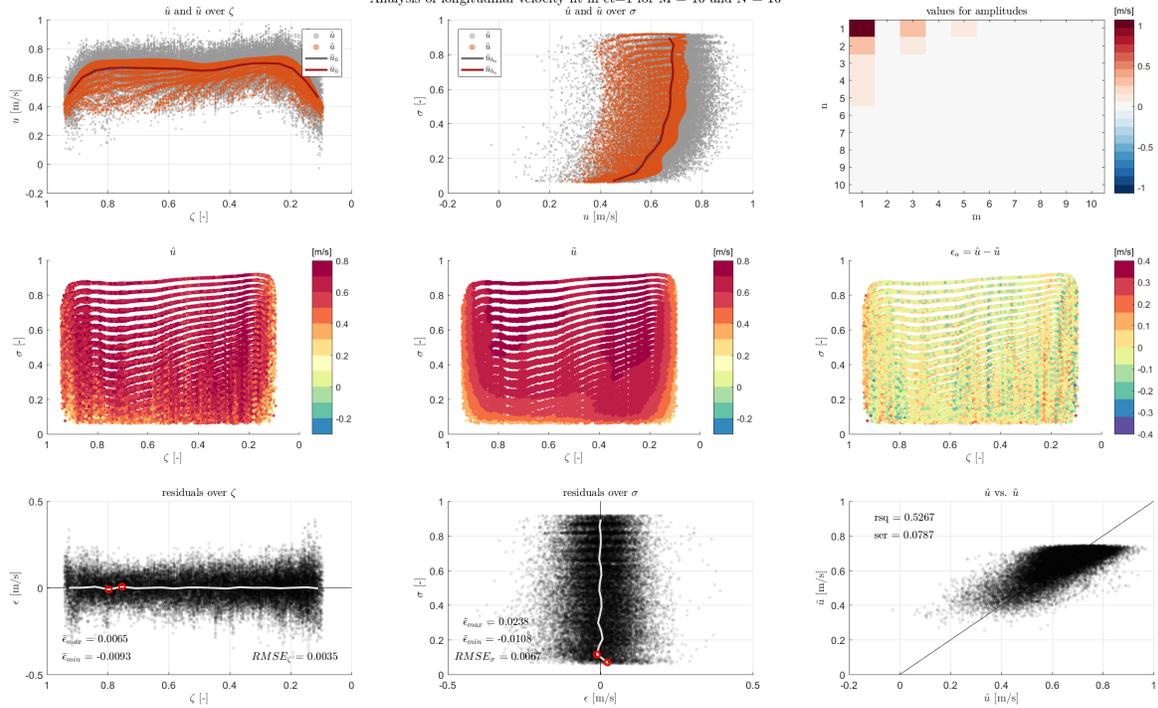
However, some of the transects are truncated earlier or later due to the maximum possible truncation. Besides that, there might some other interesting truncations present in the model fit. Analyses of the model fits in other transects are available on request, such as an analysis with desired specific truncation numbers.

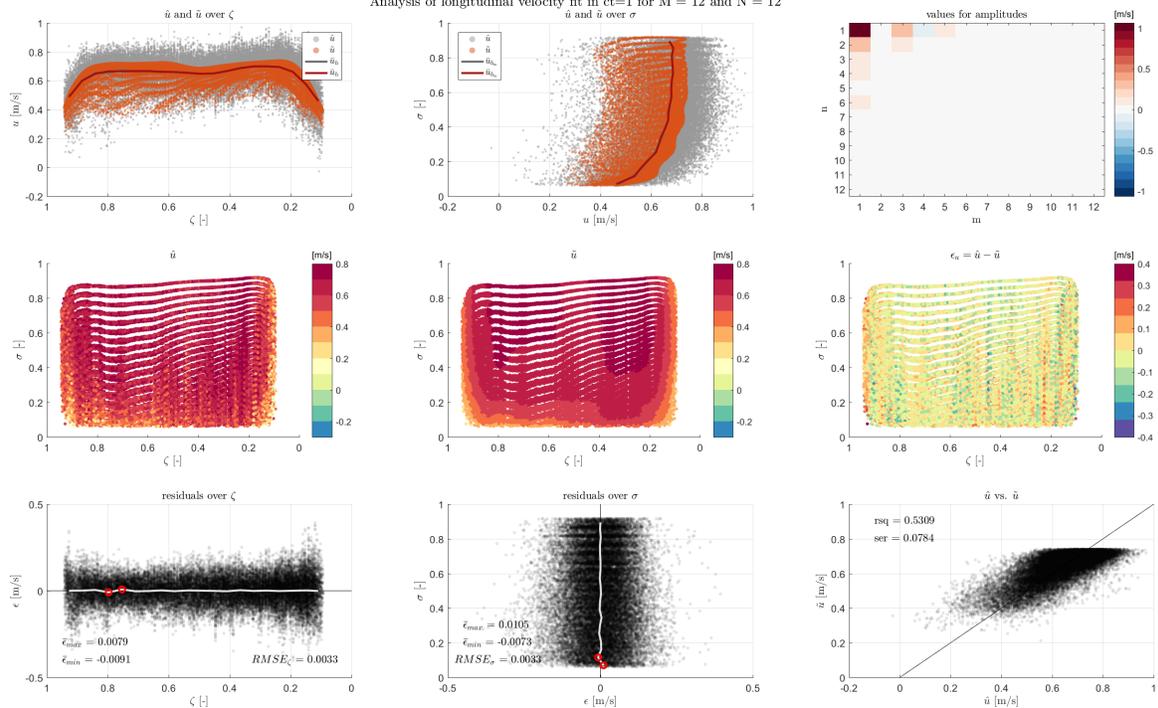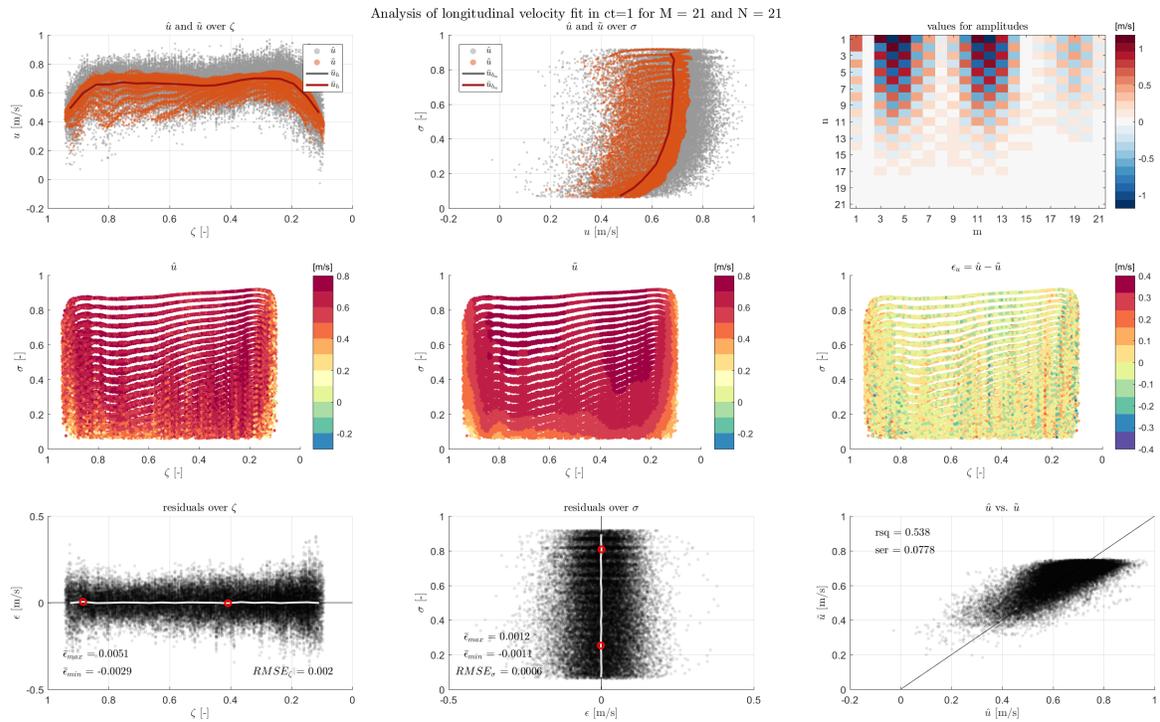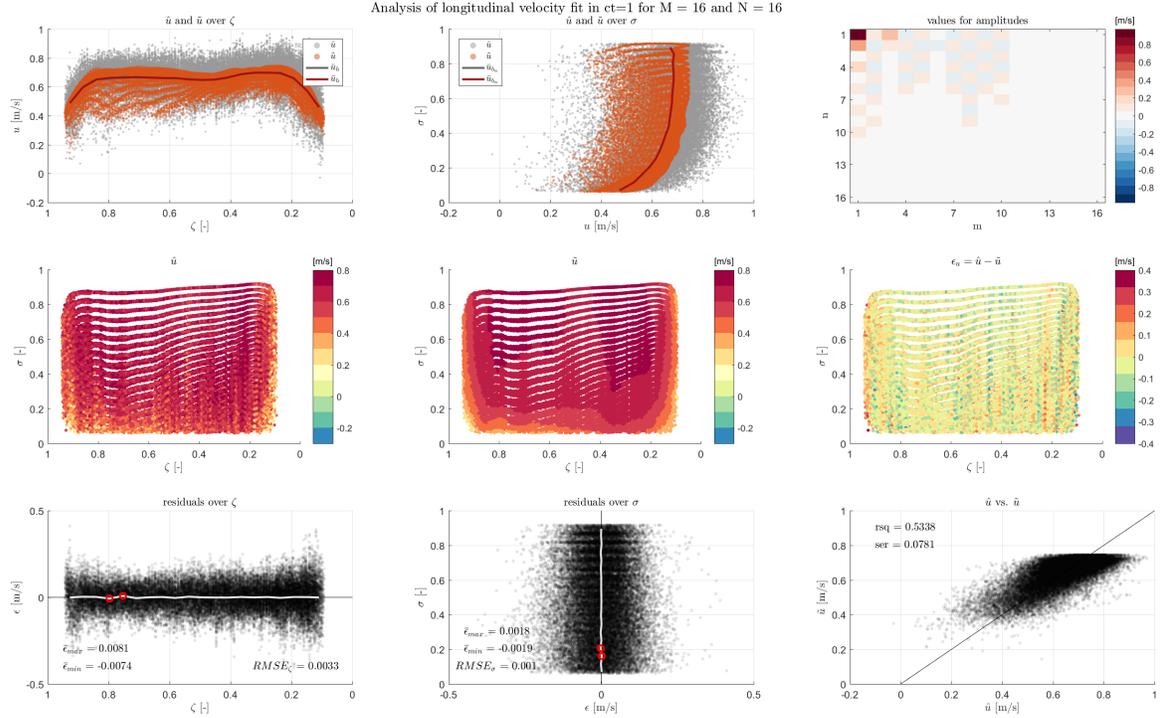## D.1 Analysis of longitudinal model fit in transect 1

Analysis of longitudinal velocity fit in ct=1 for M = 4 and N = 4

Analysis of longitudinal velocity fit in ct=1 for M = 8 and N = 8
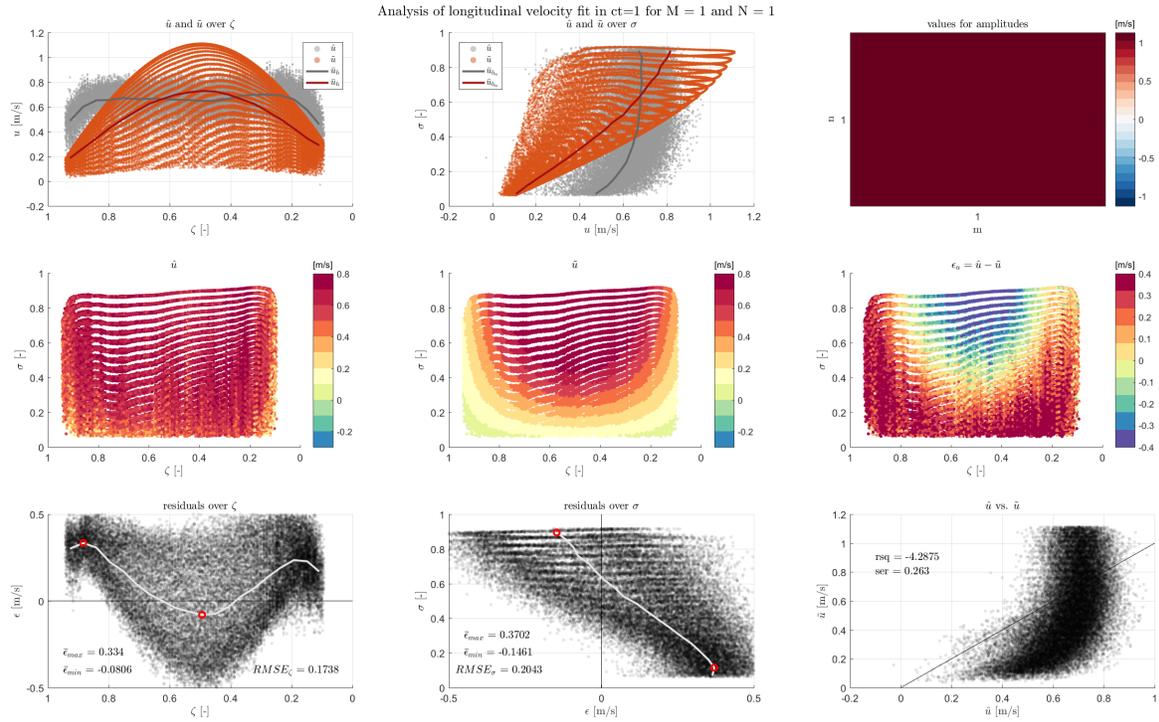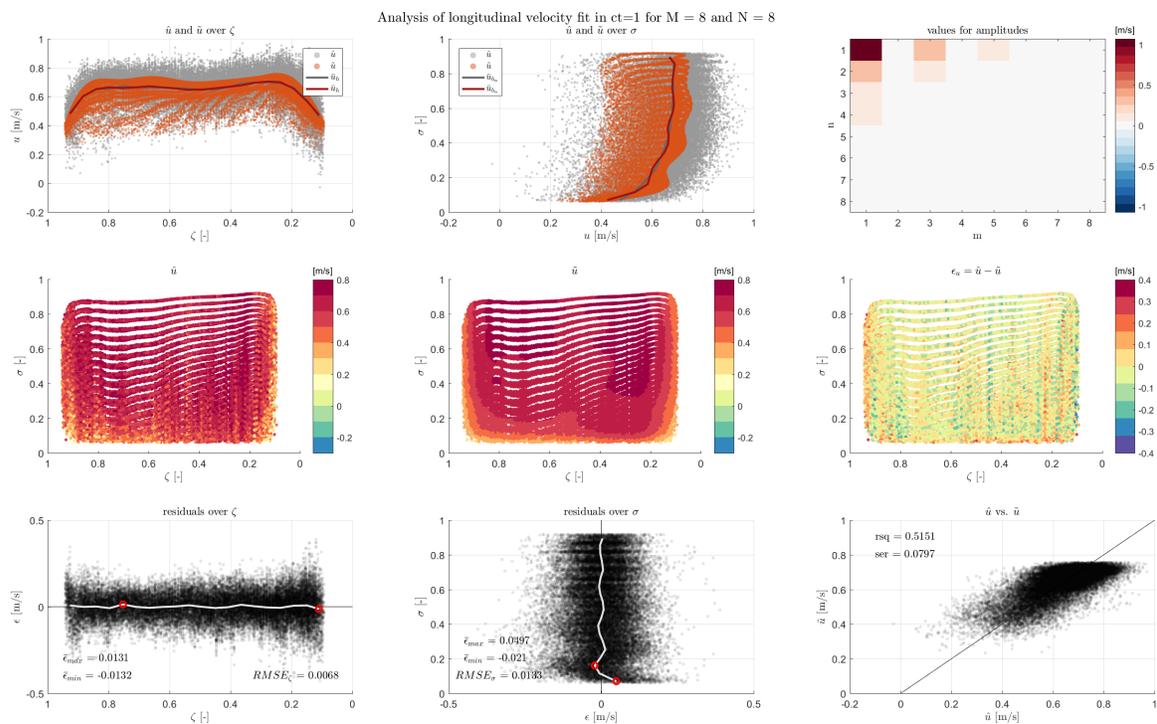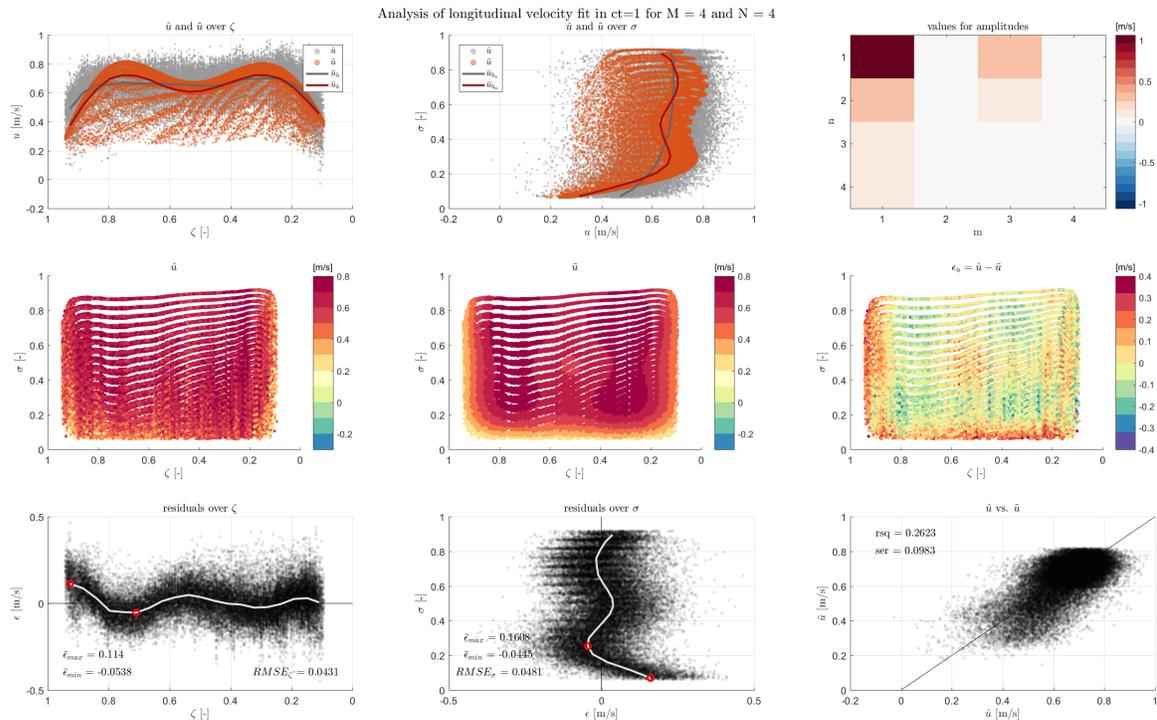
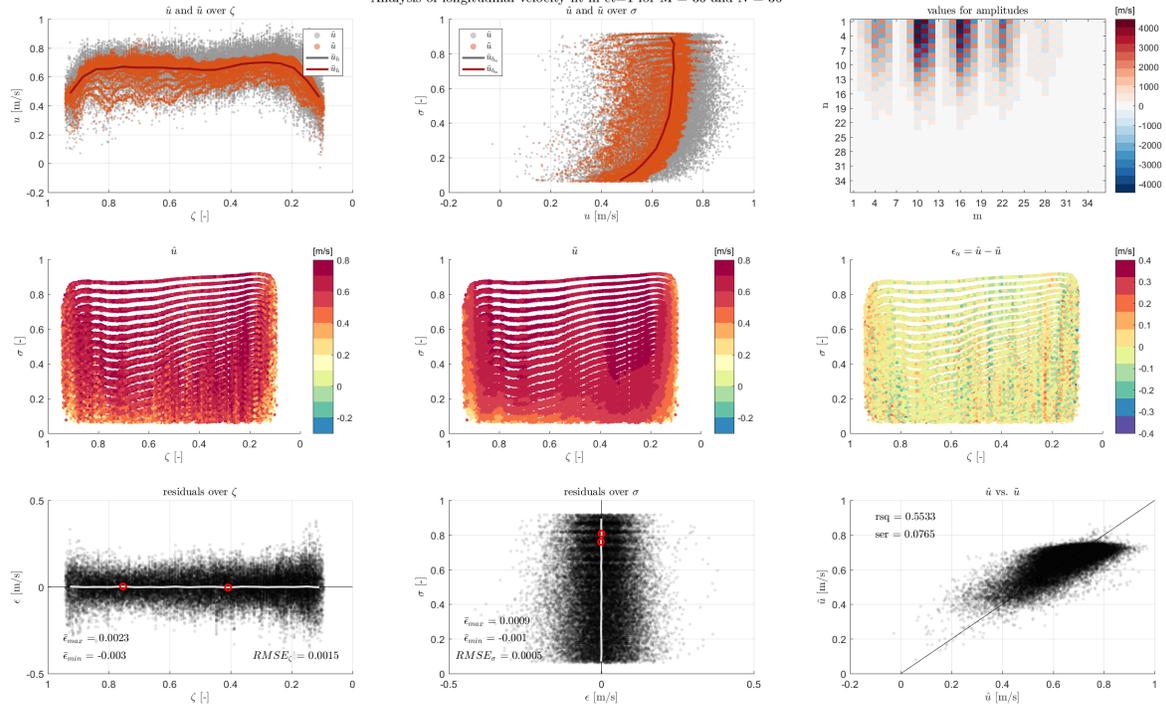Analysis of longitudinal velocity fit in ct=1 for M = 10 and N = 10

Analysis of longitudinal velocity fit in ct=1 for M = 12 and N = 12

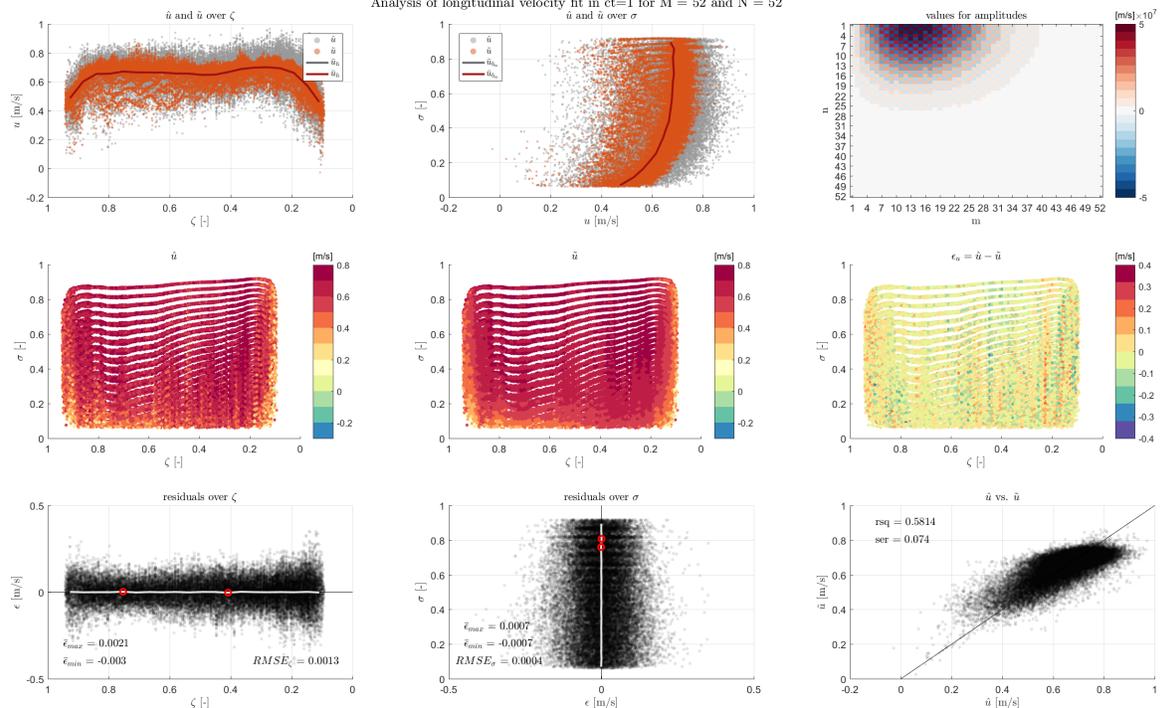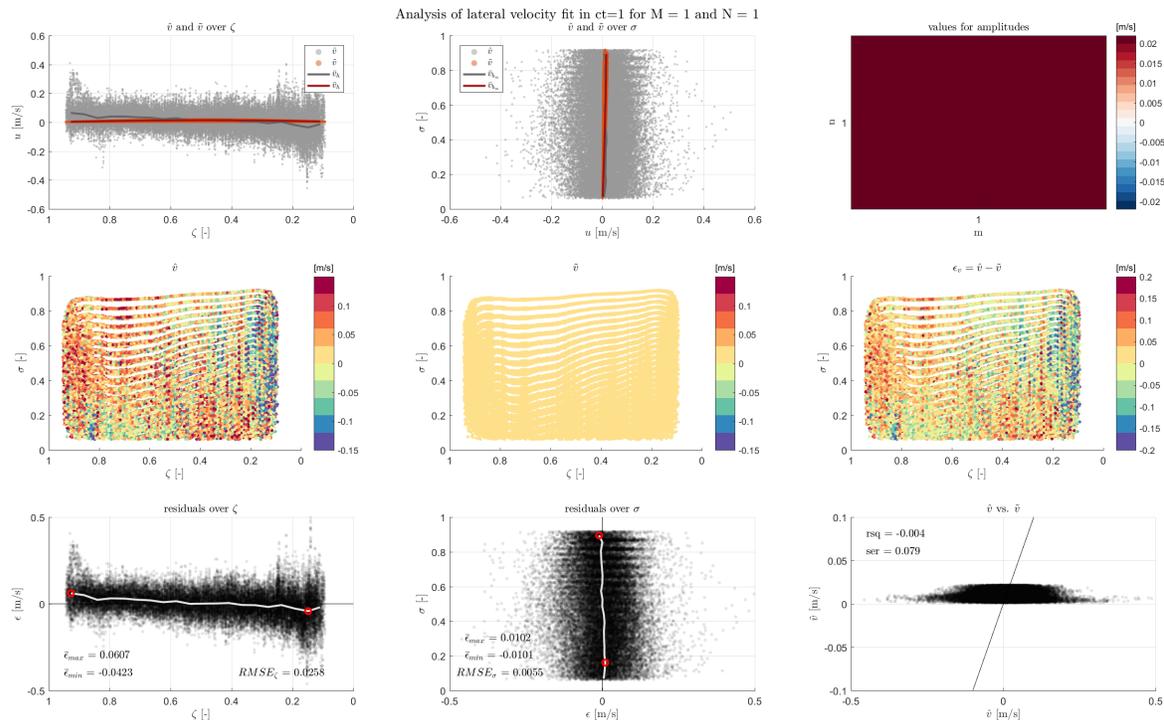Analysis of longitudinal velocity fit in ct=1 for M = 16 and N = 16

Analysis of longitudinal velocity fit in ct=1 for M = 21 and N = 21

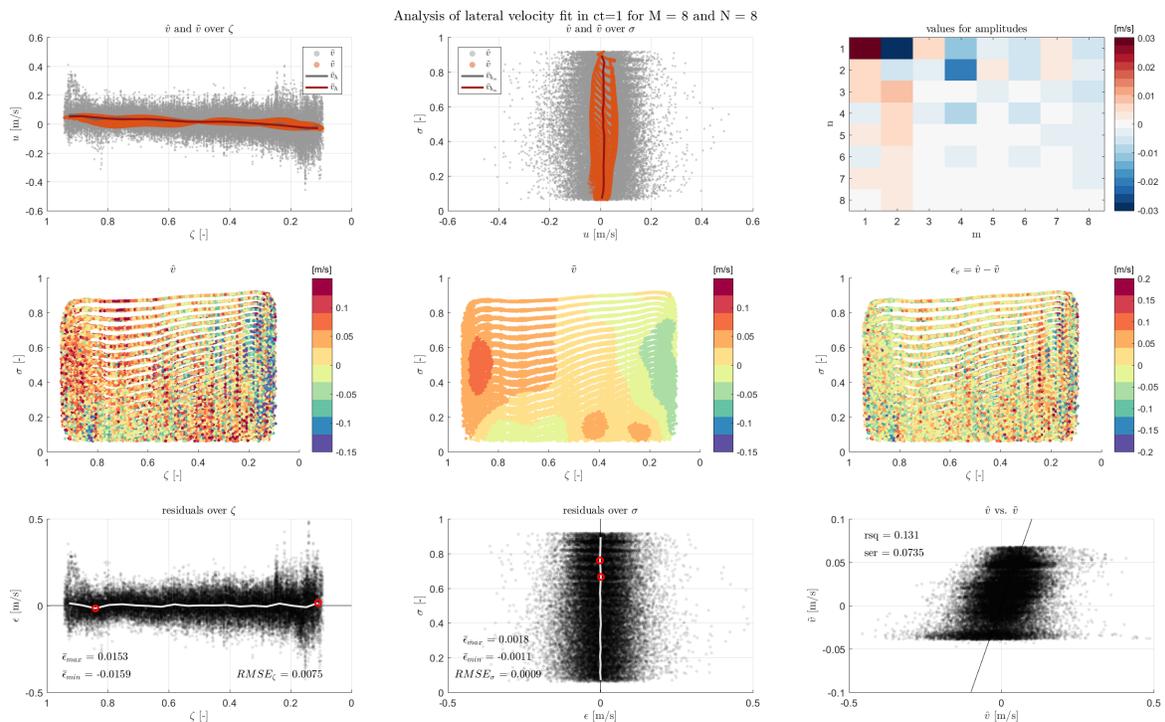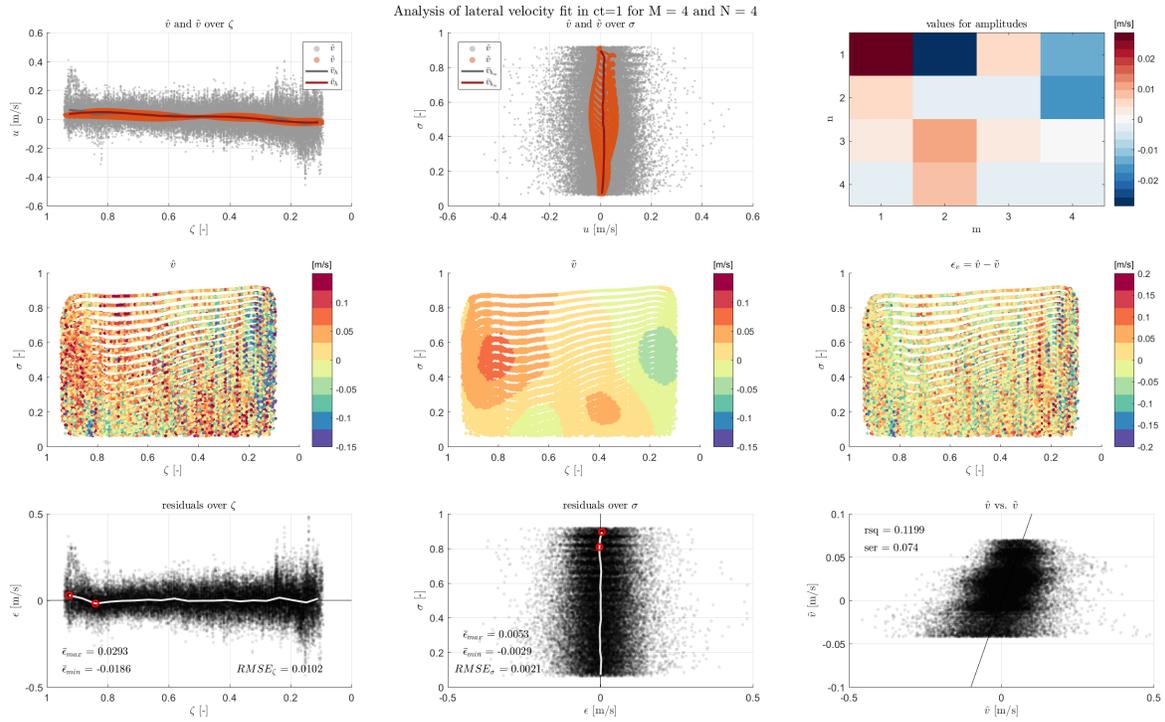Analysis of longitudinal velocity fit in ct=1 for M = 36 and N = 36

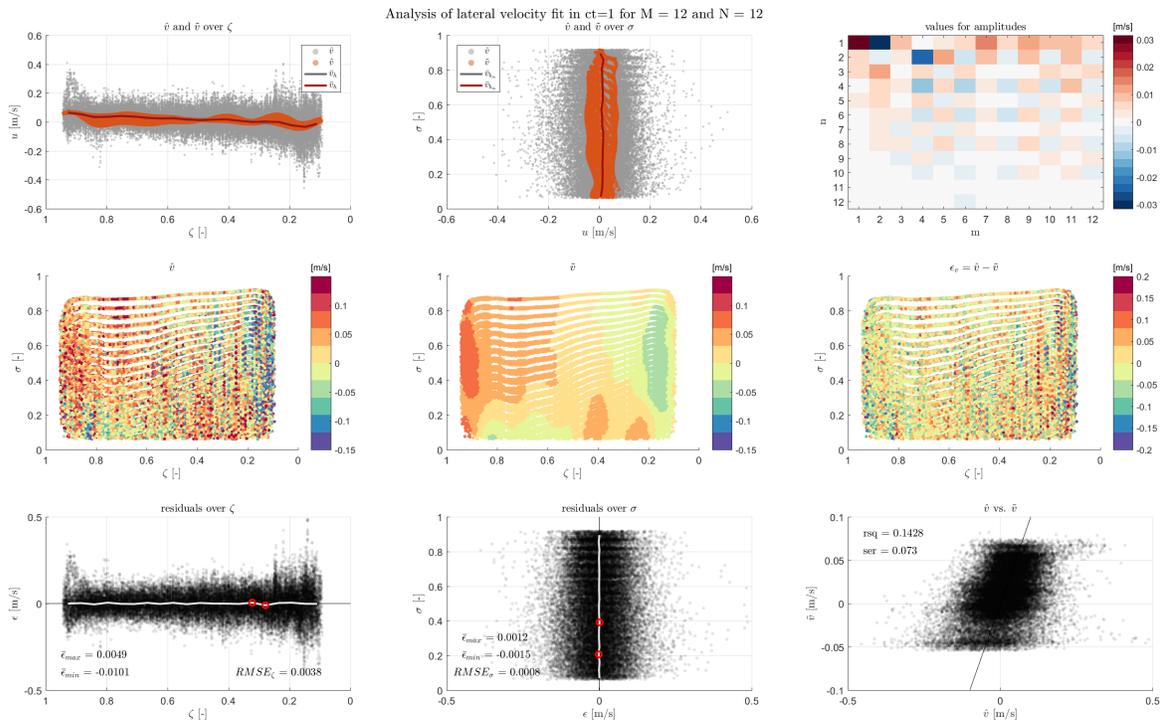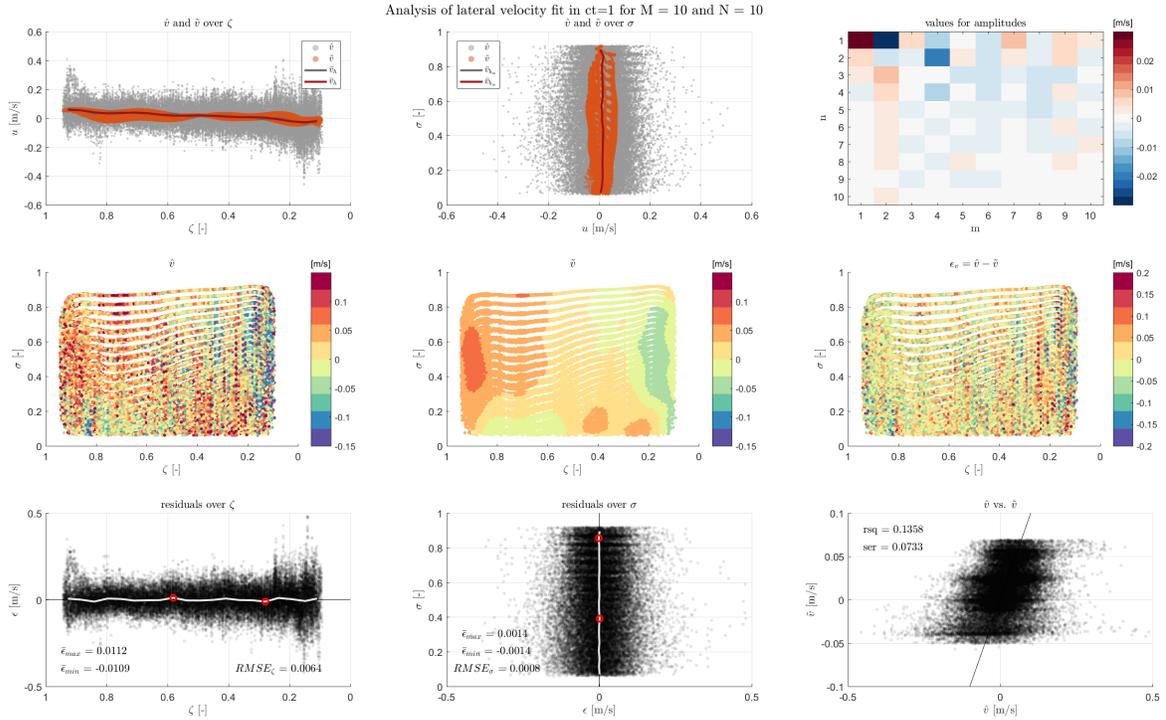Analysis of longitudinal velocity fit in ct=1 for M = 52 and N = 52

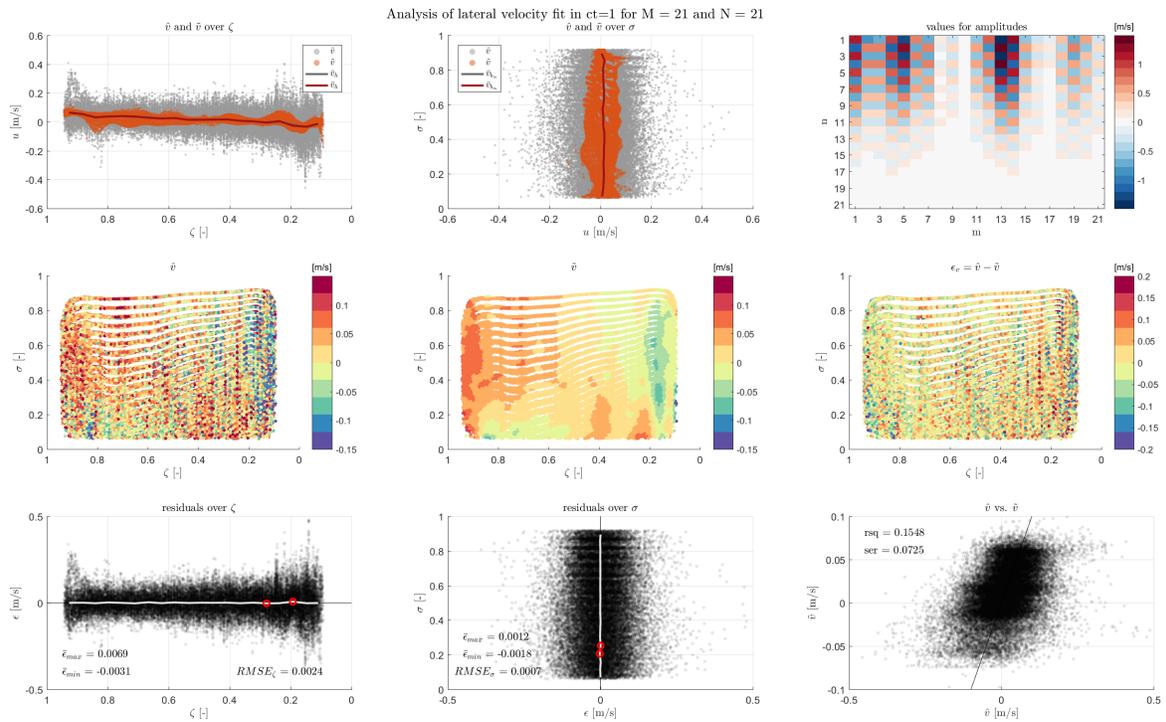## D.2 Analysis of lateral model fit in transect 1

Analysis of lateral velocity fit in ct=1 for M = 1 and N = 1

Analysis of lateral velocity fit in ct=1 for M = 4 and N = 4

Analysis of lateral velocity fit in ct=1 for M = 8 and N = 8

Analysis of lateral velocity fit in ct=1 for M = 10 and N = 10

Analysis of lateral velocity fit in ct=1 for M = 12 and N = 12

Analysis of lateral velocity fit in ct=1 for M = 16 and N = 16

Analysis of lateral velocity fit in ct=1 for M = 21 and N = 21

Analysis of lateral velocity fit in ct=1 for M = 36 and N = 36

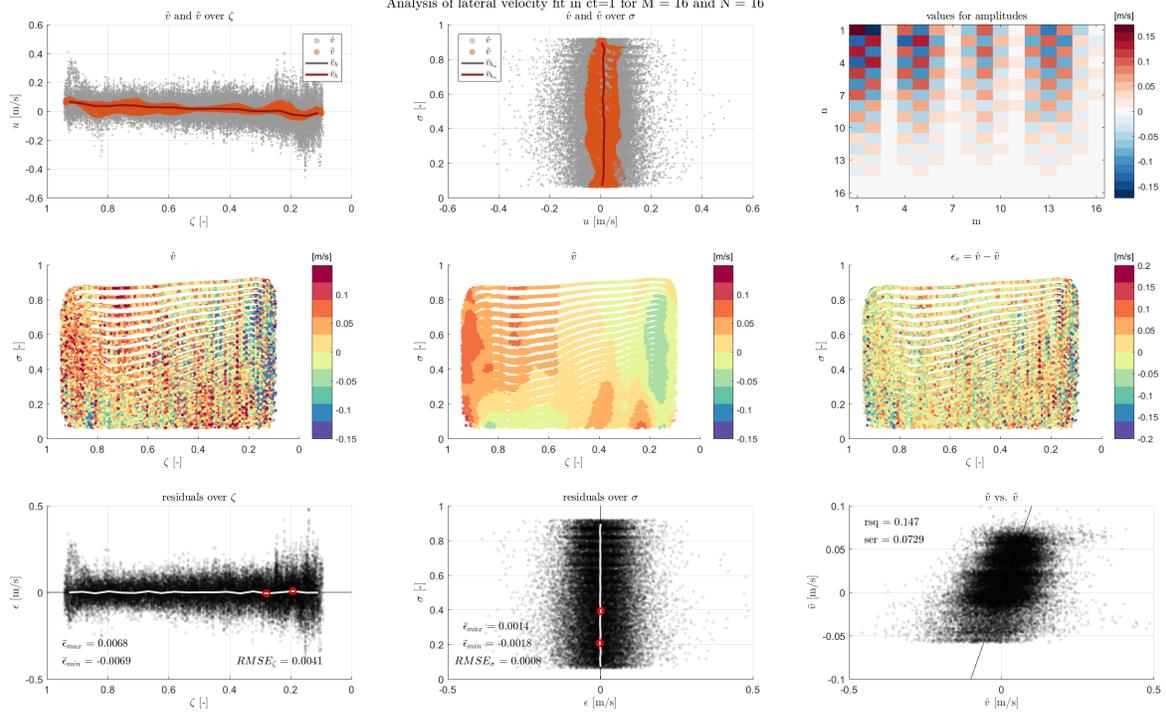

Analysis of lateral velocity fit in ct=1 for M = 52 and N = 52

D.3 Analysis of vertical model fit in transect 1

Analysis of vertical velocity fit in ct=1 for M = 4 and N = 4

Analysis of vertical velocity fit in ct=1 for M = 8 and N = 8

Analysis of vertical velocity fit in ct=1 for M = 10 and N = 10

Analysis of vertical velocity fit in ct=1 for M = 12 and N = 12

Analysis of vertical velocity fit in ct=1 for M = 16 and N = 16

Analysis of vertical velocity fit in ct=1 for M = 21 and N = 21

Analysis of vertical velocity fit in ct=1 for M = 36 and N = 36

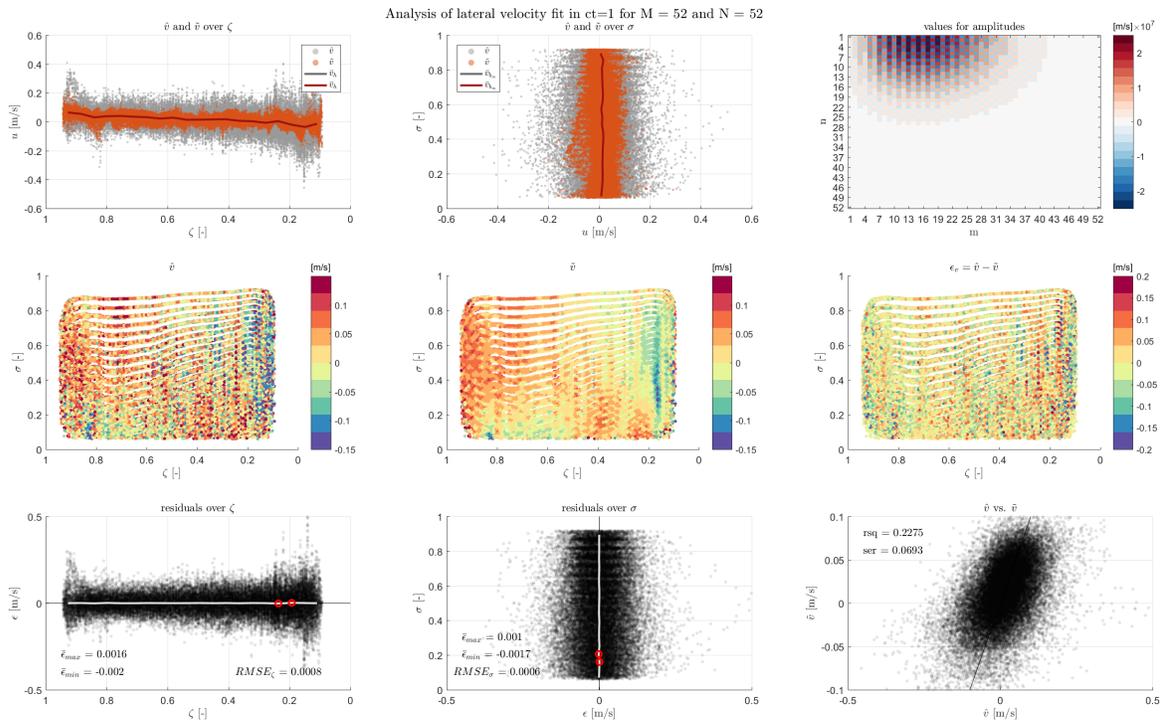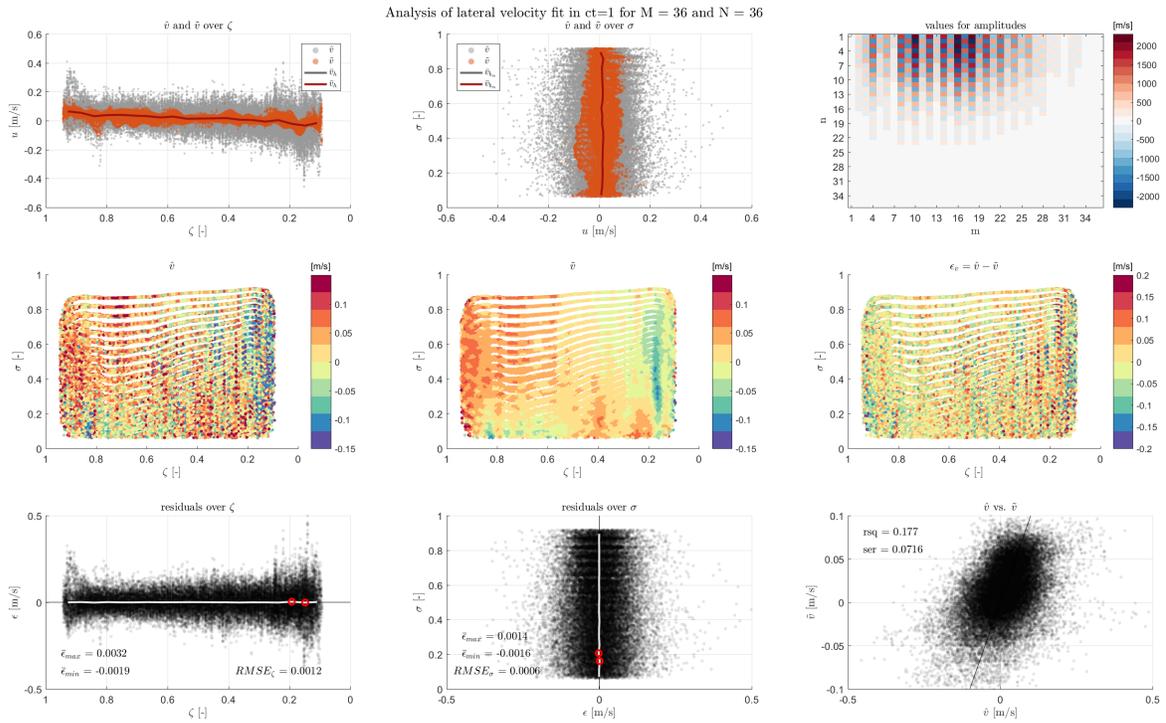Analysis of vertical velocity fit in ct=1 for M = 52 and N = 52